COUPLING IN THE HEISENBERG GROUP AND ITS APPLICATIONS TO GRADIENT ESTIMATES

SAYAN BANERJEE[⋆], MARIA GORDINA[†‡], AND PHANUEL MARIANO[†]

Abstract. We construct a non-Markovian coupling for hypoelliptic diffusions which are Brownian motions in the three-dimensional Heisenberg group. We then derive properties of this coupling such as estimates on the coupling rate, 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 Laplacian which is the generator of Brownian motion in the Heisenberg group.

Contents

1. Introduction

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 article, 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 a geometric setting of a sub-Riemannian manifold such as the Heisenberg group $\mathbb{H}^3$. Namely, we discuss couplings of two Markov processes having the same generator but starting from different points joining together (coupling) at some random time, and how these can be used to obtain total variation bounds and prove gradient estimates for harmonic functions on $\mathbb{H}^3$. Couplings have been an extremely useful tool in probability theory and has resulted in establishing deep connections between probability, analysis and geometry.

1991 *Mathematics Subject Classification.* Primary 58G32; Secondary 35H20, 60J60, 35R03.
Key words and phrases. coupling, hypoelliptic diffusion, Heisenberg group.
⋆ Research was supported in part by EPSRC Research Grant EP/K013939.
‡ Research was supported in part by the Simons Fellowship.
† Research was supported in part by NSF Grant DMS-1007496.

We start by providing some background on couplings and then on gradient estimates in our setting. 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.

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*

$$\mu\left\{\tau(X,Y) > t\right\} \geqslant ||\mathcal{L}(X_t) - \mathcal{L}(Y_t)||_{TV}, \tag{1.1}$$

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.

This, in turn, can be used to provide sharp rates of convergence of Markov processes to their respective stationary distributions, when they exist (see [28] 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 [16] 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 [33] and generalized to non-Markovian processes by Goldstein [14] and continuous time càdlàg processes by Sverchkov and Smirnov [35].

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, these couplings usually do not attain the optimal rates. In fact, it has been shown in [3] 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, [2] 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 this article: 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 article, 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 2. 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 [23] and rates of these couplings have been studied in [24]. However, the rates for the coupling we construct are much better. In fact, we show in Remark 3.2 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 3.7.

Now we would like to describe gradient estimates in 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 [10,34,37]) 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 [11] used the method of coupling two diffusion processes to obtain a similar gradient estimate for solutions to the equation

$$\frac{1}{2}\Delta u + Zu = 0 \tag{1.2}$$

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 [22]. In particular, M. Cranston proved the following gradient estimate.

Theorem 1.1 (Cranston). *Suppose (M, g) is a complete d-dimensional Riemannian manifold with distance ρ_M and assume $Ric_M \geqslant -Kg$. 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* (1.2) *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 (1.2) *is satisfied on M and u is bounded and positive, then*

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

Cranston's approach generalized the coupling of Brownian motions on manifolds of Kendall [21] 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 [12].

In this paper we consider the simplest sub-Riemannian manifold, the Heisenberg group $\mathbb{H}^3$ as a starting point of using couplings for proving gradient estimates in such a setting. As the generator of $\mathbb{H}^3$-valued Brownian motion is a hypoelliptic operator, functional inequalities for the corresponding harmonic functions or hypoelliptic heat kernels are much more challenging to prove. There was recent progress in using generalized curvature-dimension inequalities for such results (e.g. [1, 4, 5], as well as results in the spirit of optimal transport (e.g. [26]). The main point of the current paper is not whether a coupling can be constructed, as these have been known since [6], but rather finding a (necessarily non-Markovian) coupling that gives sharp total variation bounds and explicit gradient estimates. The properties of the coupling we construct in the current paper are crucial in this, and it is interesting to contrast this with optimality (or the lack of it) for the Kendall-Cranston coupling in the Riemannian manifolds as described in [25, 27].

The paper is organized as follows. Section 2 gives basics on sub-Riemannian manifolds and the Heisenberg group $\mathbb{H}^3$ including Brownian motion on $\mathbb{H}^3$. In Section 3 we construct the non-Markovian coupling of Brownian motions in $\mathbb{H}^3$, and describe its properties. Finally, in Section 4 we prove the gradient estimates for harmonic functions for the hypoelliptic Laplacian which is the generator of Brownian motion in the Heisenberg group.

2. Preliminaries

2.1. **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 [31].

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_pM 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

$$d_{CC}(x,y) = \inf \left\{ \int_0^1 \|\gamma'(t)\|_{\mathcal{H}}\, dt \quad \text{where } \gamma(0)=x, \gamma(1)=y, \gamma \text{ is a horizontal curve} \right\}, \tag{2.1}$$

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 [18, 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 [15].

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 $X \in \mathcal{H}$,

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

We define the length of the gradient as in [26]. For a function f on M, let

$$|\nabla_{\mathcal{H}} f|(x) := \lim_{r \downarrow 0} \sup_{0 < d_{CC}(x,\tilde{x}) \leqslant r} \left| \frac{f(x) - f(\tilde{x})}{d_{CC}(x, \tilde{x})} \right|, \tag{2.2}$$

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

2.2. **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

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

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

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

The horizontal distribution is defined by $\mathcal{H} = \operatorname{span}\{\mathcal{X}, \mathcal{Y}\}$ 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 [15, Example 6.1], the (sum of squares) operator

$$\Delta_{\mathcal{H}} = \mathcal{X}^2 + \mathcal{Y}^2 \tag{2.3}$$

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 [9] 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 $\mu(\theta)\left(x^2 + y^2\right) = |z|$ in the interval $[0, \pi)$ and $\mu(z) = \frac{z}{\sin^2 z} - \cot z$ and where

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

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 [8] 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

$$\rho(\mathbf{x}, \mathbf{y}) = \left((x - \tilde{x})^2 + (y - \tilde{y})^2 + |z - \tilde{z} + x\tilde{y} - y\tilde{x}|\right)^{\frac{1}{2}}. \tag{2.4}$$

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)

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

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

$$\begin{aligned} X_1(t) &= b_1 + B_1(t), \\ X_2(t) &= b_2 + B_2(t), \\ X_3(t) &= a + \int_0^t (B_1(s) + b_1) dB_2(s) - \int_0^t (B_2(s) + b_2) dB_1(s). \end{aligned} \tag{2.5}$$

3. Coupling results

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

$$\mathbf{X}_t = \left(B_1(t), B_2(t), a + \int_0^t B_1(s) dB_2(s) - \int_0^t B_2(s) dB_1(s) \right) \tag{3.1}$$

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

$$\begin{aligned} B(t) &= B^{\mathrm{br}}(t) + \frac{t}{T} G^{(T)} \\ \widetilde{B}(t) &= \widetilde{B}^{\mathrm{br}}(t) + \frac{t}{T} G^{(T)} \end{aligned} \tag{3.2}$$

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 [2]. 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 [2]).

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

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

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-Loève expansion which goes back to [20, 30] and for examples of such expansions see [36, p.21]. For the Brownian bridge we have

$$B^{\mathrm{br}}(t)=\sqrt{T}\sum_{k=1}^{\infty}Z_k\frac{\sqrt{2}\sin\left(\frac{k\pi t}{T}\right)}{k\pi}=\sqrt{T}\sum_{k=1}^{\infty}Z_k g_{T,k}(t) \tag{3.3}$$

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 $\{Z_k\}_{k\geqslant 1}$. We now state and prove the following lemmas.

Lemma 3.1. *There exists a non-Markovian coupling of the diffusions*

$$\begin{gathered}\left\{\left(B_1(t),B_2(t),a+\int_0^t B_2(s)dB_1(s)\right):t\geqslant 0\right\},\\ \left\{\left(\widetilde{B}_1(t),\widetilde{B}_2(t),\widetilde{a}+\int_0^t \widetilde{B}_2(s)d\widetilde{B}_1(s)\right):t\geqslant 0\right\},\\ B_1(0)=\widetilde{B}_1(0)=b_1,\ B_2(0)=\widetilde{B}_2(0)=b_2,\ \text{and } a>\widetilde{a},\end{gathered}$$

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

$$\left(\frac{B_1(t)}{r},\ \frac{B_2(t)}{r}, \frac{a+\int_0^t B_2(s)dB_1(s)}{r^2}\right) \stackrel{d}{=} \left(B_1'(t/r^2), B_2'(t/r^2), \frac{a}{r^2} + \int_0^{t/r^2} B_2'(s)dB_1'(s)\right), \tag{3.4}$$

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 (3.4) with $r = \sqrt{a - \widetilde{a}}$.

Let us divide the non-negative real line into intervals $\left[2^n - 1, 2^{n+1} - 1\right], n \geqslant 0$. 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 $t \in \left[2^n - 1, 2^{n+1} - 1\right]$. To this end, we will couple two Brownian bridges B^{br} and $\widetilde{B}^{\mathrm{br}}$ on $\left[2^n - 1, 2^{n+1} - 1\right]$, then sample an independent Gaussian random variable $G^{(2^n)}$ with mean zero, variance 2^n and finally use the recipe (3.2) to get the coupling of B_2 and $\widetilde{B}_2$ on $\left[2^n - 1, 2^{n+1} - 1\right]$.

Let $\left(Z_1^{(n)}, Z_2^{(n)}, \dots\right)$ and $\left(\widetilde{Z}_1^{(n)}, \widetilde{Z}_2^{(n)}, \dots\right)$ denote the Gaussian coefficients in the Karhunen-Loève expansion (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

$$\lambda_n(t) = \frac{2}{\pi} \int_{2^n - 1}^{t} \sqrt{2} \sin\left(\frac{\pi(s - 2^n + 1)}{2^n}\right) dB_1(s), 2^n - 1 \leqslant t \leqslant 2^{n+1} - 1. \tag{3.5}$$

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

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

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

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

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

$$I(t) - \widetilde{I}(t) = I(2^n - 1) - \widetilde{I}(2^n - 1) + W^{(n)}\left(2^n \wedge \sigma^{(n)}\right) \lambda_n(t), \tag{3.6}$$

for $t \in \left[2^n - 1, 2^{n+1} - 1\right]$. 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 $I(2^n - 1) - \widetilde{I}(2^n - 1) = 0$, 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 $t \geqslant 0$,

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

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

$$\mathbb{P}\left(\tau > 2^{n+1} - 1\right) \leqslant C \mathbb{E}\left[\frac{1}{\sqrt{T\left(2^{n+1} - 1\right)}}\right]. \tag{3.7}$$

Thus, we need to obtain an estimate for the right hand side in (3.7). 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 $re^{-r^2/2}dr$ 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 (3.7), 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. □

Remark 3.2. Under the hypothesis of Lemma 3.1, 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 [2, 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

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

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 3.1 gives a rate of t^{-1}.

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 3.3. *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*

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

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

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

where the second step follows from Chebyshev's inequality. From standard estimates for Brownian motion, $\mathbb{P}(\tau_z < \tau_0) = b/z$ and $\mathbb{E}[\tau_0 \wedge \tau_z] = b(z-b) \leqslant bz$. 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 $z = \sqrt{y}$. □

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

$$A(t) = (a - \widetilde{a}) + \left(\int_0^t B_1(s)dB_2(s) - \int_0^t B_2(s)dB_1(s)\right) - \left(\int_0^t \widetilde{B}_1(s)d\widetilde{B}_2(s) - \int_0^t \widetilde{B}_2(s)d\widetilde{B}_1(s)\right) + B_1(t)\widetilde{B}_2(t) - B_2(t)\widetilde{B}_1(t). \tag{3.8}$$

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,

$$A(t) - A(0) = -2\int_0^t B_2(s)dB_1(s) + 2\int_0^t \widetilde{B}_2(s)dB_1(s), \tag{3.9}$$

where

$$A(0) = a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1, \tag{3.10}$$

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 3.4. *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 (3.8) 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 $t \geqslant \max\left\{\left|b_2 - \widetilde{b}_2\right|^2, 2\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|\right\}$,*

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

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 $t > 0$,

$$\begin{aligned}
\mathbb{E}\left[\frac{|A(T_1)|}{t} \wedge 1\right] &\leqslant \sum_{k=0}^{\infty} \mathbb{E}\left[\frac{|A(T_1)|}{t} \wedge 1; 2^{-k-1}t < |A(T_1)| \leqslant 2^{-k}t\right] + \mathbb{P}(|A(T_1)| > t) \\
&\leqslant \sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(2^{-k-1}t < |A(T_1)| \leqslant 2^{-k}t\right) + \mathbb{P}(|A(T_1)| > t) \\
&\leqslant \sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right) + \mathbb{P}(|A(T_1)| > t). \qquad (3.11)
\end{aligned}$$

As B_2 and $\widetilde{B}_2$ are reflection coupled, we can rewrite (3.9) 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 3.3, for $t \geqslant \max\left\{\left|b_2 - \widetilde{b}_2\right|^2, 2|A(0)|\right\}$,

$$\begin{aligned}
\mathbb{P}(|A(T_1)| > t) &\leqslant \mathbb{P}(|A(T_1) - A(0)| > t - |A(0)|) \\
&\leqslant \mathbb{P}\left(|A(T_1) - A(0)| > \frac{t}{2}\right) \leqslant C\frac{\left|b_2 - \widetilde{b}_2\right|}{\sqrt{t}}. \qquad (3.12)
\end{aligned}$$

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

$$\begin{aligned}
&\sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right) \qquad (3.13) \\
&= \sum_{k:2^{-k-1}t \leqslant \max\left\{\left|b_2 - \widetilde{b}_2\right|^2, 2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right) \\
&\quad + \sum_{k:2^{-k-1}t > \max\left\{\left|b_2 - \widetilde{b}_2\right|^2, 2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right).
\end{aligned}$$

To estimate the first term on the right hand side of (3.13), 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,

$$
\begin{aligned}
(3.14)\qquad & \sum_{k:2^{-k-1}t\leqslant \max\left\{\left|b_2-\widetilde{b}_2\right|^2,2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(\left|A\left(T_1\right)\right|\geqslant 2^{-k-1}t\right) \\
&\leqslant \sum_{k=k_0}^{\infty} 2^{-k} = 2^{-k_0+1} = \frac{4}{t}2^{-k_0-1}t \leqslant \frac{4}{t}\max\left\{\left|b_2-\widetilde{b}_2\right|^2, 2\left|A(0)\right|\right\} \\
&\leqslant 8\left(\frac{\left|b_2-\widetilde{b}_2\right|^2}{t}+\frac{|A(0)|}{t}\right) \leqslant 8\left(\frac{\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}}+\frac{\left|a-\widetilde{a}+b_1\widetilde{b}_2-b_2\widetilde{b}_1\right|}{t}\right),
\end{aligned}
$$

where we used the facts that $\frac{\left|b_2-\widetilde{b}_2\right|^2}{t} \leqslant \frac{\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}}$ 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 (3.13), we use Lemma 3.3 to get

$$
\begin{aligned}
(3.15)\qquad & \sum_{k:2^{-k-1}t> \max\left\{\left|b_2-\widetilde{b}_2\right|^2,2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(\left|A\left(T_1\right)\right|\geqslant 2^{-k-1}t\right) \\
&\leqslant \frac{C}{\sqrt{t}} \sum_{k:2^{-k-1}t> \max\left\{\left|b_2-\widetilde{b}_2\right|^2,2|A(0)|\right\}} 2^{-k/2}\left|b_2-\widetilde{b}_2\right| \\
&\leqslant \frac{C\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}}\sum_{k=0}^{\infty}2^{-k/2} \leqslant C'\frac{\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}}.
\end{aligned}
$$

Using (3.14) and (3.15) in (3.13),

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

Using (3.12) and (3.16) in (3.11), we complete the proof of the lemma. □

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

Theorem 3.5. *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 (3.8). 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 (3.9). 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 3.1 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

$$\mathbb{P}\left(T_1 > t\right) \leqslant \frac{C\left|b_2 - \widetilde{b}_2\right|}{\sqrt{t}} \tag{3.17}$$

for $t \geqslant \left|b_2 - \widetilde{b}_2\right|^2$. By Lemma 3.1 and Lemma 3.4, for $t \geqslant \max\left\{\left|b_2 - \widetilde{b}_2\right|^2, 2\left|A(0)\right|\right\}$,

$$\begin{aligned}\mathbb{P}\left(\tau - T_1 > t\right) &\leqslant C\mathbb{E}\left[\frac{\left|A\left(T_1\right)\right|}{t} \wedge 1\right] \\ &\leqslant C\left(\frac{\left|b_2 - \widetilde{b}_2\right|}{\sqrt{t}} + \frac{\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|}{t}\right).\end{aligned} \tag{3.18}$$

Equations (3.17) and (3.18) together yield the required tail bound on the coupling time probability stated in the theorem. □

An interesting observation to note from Theorem 3.5 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 3.6. *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*

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

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

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 3.5 and the Aldous' inequality (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

$$p_t^{(u,v,w)}(x,y,z) = p_t^e(x-u, y-v, z-w-uy+vx), \text{ where } e=(0,0,0). \tag{3.19}$$

Then

$$\begin{aligned}
&d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) = \int_{\mathbb{R}^3}\left|p_t^{(b_1,b_2,a)}(x,y,z)-p_t^{(b_1,b_2,\widetilde{a})}(x,y,z)\right| dxdydz \\
&= \int_{\mathbb{R}^3} |p_t^e(x-b_1, y-b_2, z-a-b_1y+b_2x) \\
&-p_t^e(x-b_1, y-b_2, z-\widetilde{a}-b_1y+b_2x)| \, dxdydz \\
&= \int_{\mathbb{R}^3} |p_t^e(x,y,z-a)-p_t^e(x,y,z-\widetilde{a})| \, dxdydz \\
&\geqslant \int_{\mathbb{R}} |f_t(z-a)-f_t(z-\widetilde{a})| \, dz,
\end{aligned}$$

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

$$\begin{aligned} d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) &\geqslant \int_{\mathbb{R}} \left| f_1\left(z - \frac{a}{t}\right) - f_1\left(z - \frac{\widetilde{a}}{t}\right)\right| dz \\ &= \int_{\mathbb{R}} \left| f_1\left(z - \frac{a-\widetilde{a}}{t}\right) - f_1\left(z\right)\right| dz \\ &\geqslant \int_{|z|\geqslant 1} \left| f_1\left(z - \frac{a-\widetilde{a}}{t}\right) - f_1\left(z\right)\right| dz. \end{aligned}$$

The explicit form of f_1 is well-known (see, for example, [38] or [32, 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 $\frac{a-\widetilde{a}}{t} \leqslant \frac{1}{2}$,

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

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,

$$\begin{aligned} d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) &\geqslant \int_{|z|\geqslant 1} \left| f_1\left(z - \frac{a-\widetilde{a}}{t}\right) - f_1\left(z\right)\right| dz \\ &\geqslant \frac{|a-\widetilde{a}|}{t} \int_{|z|\geqslant 1} |f_1'(3z/2)| dz = C_2 \frac{|a-\widetilde{a}|}{t}, \end{aligned}$$

which completes the proof of the theorem. □

Several remarks are in order.

Remark 3.7. Theorem 3.6 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 [2, Definition 1] for the precise notion of efficiency.

Remark 3.8. 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 $0 \leqslant y \leqslant 1$

$$x + y \leqslant \sqrt{2}\left(x^2 + y\right)^{\frac{1}{2}}. \tag{3.20}$$

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 3.9. *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*

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

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

Proof. Since $t > 1$, then $\frac{1}{t} \leqslant \frac{1}{\sqrt{t}}$, so by Theorem 3.5

$$\begin{aligned}
\mathbb{P}(\tau > t) &\leqslant C\left(\frac{\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|}{\sqrt{t}} + \frac{\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|}{t}\right) \\
&\leqslant \frac{C}{\sqrt{t}}\left(\left|\mathbf{b} - \widetilde{\mathbf{b}}\right| + \left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|\right) \\
&\leqslant \frac{C}{\sqrt{t}}\left(\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|^2 + \left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|\right)^{\frac{1}{2}}
\end{aligned}$$

where we used (3.20) 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 (2.4). Recall from Section 2 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. □

Liouville type theorems have been known for the Heisenberg group and other types of Carnot groups (e.g. [7, Theorem 5.8.1]). Using the coupling we constructed, we derive a functional inequality (a form of which appeared as [1, Equation (24)]) which consequently gives us the Liouville property rather easily.

In the following, for any bounded measurable function $u : \mathbb{H}^3 \to \mathbb{R}$ 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 3.10. *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 $t \geqslant 1$*

$$\|\nabla_{\mathcal{H}} P_t u\|_{\infty} \leqslant \frac{C}{\sqrt{t}} \|u\|_{\infty}. \tag{3.21}$$

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 $\left(\mathbb{H}^3, d_{CC}\right)$ 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 3.5 and by Proposition 3.9, we get

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

Dividing by $d_{CC}\left(\left(b_1, b_2, a\right), \left(\widetilde{b_1, }\widetilde{b_2}, \widetilde{a}\right)\right)$ 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 (3.21).

Finally if $\Delta_{\mathcal{H}} u = 0$, then $P_t u = u$ for all $t \geqslant 0$. Taking $t \to \infty$ in (3.21), we get $\nabla_{\mathcal{H}} u \equiv 0$ and hence $u \in C^{\infty}(\mathbb{H}^3)$ is constant by [7, Proposition 1.5.6]. □

4. Gradient estimates

The goal of this section 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 3.5. For a set Q, define the exit time of a process $\mathbf{X}_t$ from this set by

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

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

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

Before we can formulate and prove the main results of this section, Theorems 4.3 and 4.4, we need two preliminary results. Lemma 4.1 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 [19, 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 4.1. *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 3.1, 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

$$\begin{aligned}
Y_1(t) &= \sum_{n=0}^{\infty} 2^{n/2} Z_1^{(n)} g_{n,1}((t-2^n+1)^+\wedge 2^n)\\
Y_2(t) &= \sum_{n=0}^{\infty} 2^{n/2}\left(\frac{(t-2^n+1)^+\wedge 2^n}{2^n} Z_0^{(n)} + \sum_{k=2}^{\infty} Z_k^{(n)} g_{n,k}((t-2^n+1)^+\wedge 2^n)\right)
\end{aligned} \tag{4.1}$$

with $g_{n,k}(t) = g_{2^n,k}(t)$ as defined in the Karhunen-Loève expansion (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 (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 3.1 (in particular, see (3.6)),

$$\mathbb{P}[\tau\in\{2^{n+1}-1 : n\geqslant 0\}] = 1. \tag{4.2}$$

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 (4.2). Note that for any $n \geqslant 0$,

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

Recall that

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

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 $n \geqslant 0$, $\{\tau > 2^{n+1} - 1\}$ 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 filtration and W is a Brownian motion (hence a continuous martingale) under it. From the representation (4.1), 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

$$\mathbb{E}\left(\sup_{t\leqslant\tau\wedge M}|Y_1(t)|\right)^4\leqslant\frac{64}{\pi^4}\mathbb{E}\left(\sup_{t\leqslant\tau\wedge M}|W(t)|\right)^4\leqslant C\mathbb{E}(\tau\wedge M)^2. \tag{4.3}$$

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 $T>0$,

$$\mathbb{E}\left(\sup_{t\leqslant T}|Y_2(t)|\right)^4\leqslant CT^2. \tag{4.4}$$

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 (4.4), 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. □

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

Lemma 4.2. *Under the coupling of Lemma 3.1, 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

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

From Lemma 3.1, 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. □

Let $D \subset \mathbb{H}^3$ be a domain. Later in Theorem 4.4 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 (2.4). 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 3.5. 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 4.3. *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 3.5, there exists a constant $C > 0$ that does not depend on $x, \widetilde{x}$ such that*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\mathbb{P}(\tau > \tau_Q(\mathbf{X})) &= \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}) \\ &\leqslant \mathbb{P}(T_1 > \tau_{Q_1^*}(\mathbf{X})) + \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})) \\ &\leqslant \mathbb{P}(T_1 > \tau_{Q_1^*}(\mathbf{X})) + \mathbb{P}(\sigma_{\delta_x^2/32} \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) \\ &\qquad + \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}).\end{aligned} \tag{4.5}$$

It follows from a computation involving standard Brownian estimates (see, for example, the proof of [12, Theorem 1]) that

$$\mathbb{P}(T_1 > \tau_{Q_1^*}(\mathbf{X})) \leqslant C\frac{|\mathbf{b} - \widetilde{\mathbf{b}}|}{\delta_x}. \tag{4.6}$$

To estimate the second term in (4.5), 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

$$\begin{aligned}&\mathbb{E}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t) - U(0)|\right)^2 \\ &\leqslant C\mathbb{E}\left(\int_0^{T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |\mathbf{B}(s) - \hat{\mathbf{b}}|^2 ds\right) \\ &\leqslant C\mathbb{E}\left(\int_0^{T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} \delta_x^2 ds\right) \\ &\leqslant C\delta_x^2\mathbb{E}(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})).\end{aligned}$$

We can again appeal to standard Brownian estimates (e.g. see the proof of [12, Theorem 1]) to see that

$$\mathbb{E}\left(T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))\right) \leqslant C\delta_x|\mathbf{b} - \hat{\mathbf{b}}|. \tag{4.7}$$

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 (3.20). 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

$$\mathbb{P}(\sigma_{\delta_x^2/32}\leqslant T_1\wedge\tau_{Q_1^*}(\mathbf{X}))\leqslant C\frac{(1+\delta_x)^3}{\delta_x^4}\rho(x,\widetilde{x}). \tag{4.8}$$

To estimate the last term in (4.5), we write

$$\begin{aligned}&\mathbb{P}(\tau>\tau_{Q_1}(\mathbf{X})\wedge\sigma_{\delta_x^2/16},T_1\leqslant\tau_{Q_1^*}(\mathbf{X})\wedge\sigma_{\delta_x^2/32})\leqslant\mathbb{P}(\tau-T_1>1)\\&\quad+\mathbb{P}(\tau>\tau_{Q_1}(\mathbf{X})\wedge\sigma_{\delta_x^2/16},T_1\leqslant\tau_{Q_1^*}(\mathbf{X})\wedge\sigma_{\delta_x^2/32},\tau-T_1\leqslant 1).\end{aligned} \tag{4.9}$$

By Lemma 3.1, 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 (3.8).

Applying Lemma 3.4 with $t=1$ and appealing to our assumption that $|\mathbf{b}-\widetilde{\mathbf{b}}|\leqslant 1$ and $|a-\widetilde{a}+b_1\widetilde{b}_2-b_2b_1|\leqslant 1/2$, we have

$$\mathbb{E}|A(T_1)\wedge 1|\leqslant C(|\mathbf{b}-\widetilde{\mathbf{b}}|+|a-\widetilde{a}+b_1\widetilde{b}_2-b_2b_1|)\leqslant C\rho(x,\widetilde{x}).$$

which gives

$$\mathbb{P}(\tau-T_1>1)\leqslant C\rho(x,\widetilde{x}). \tag{4.10}$$

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

$$
\begin{aligned}
\mathbb{P}(\tau > &\tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}, \tau - T_1 \leqslant 1) \\
&\leqslant \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| \geqslant \delta_x/16 \right) + \\
&\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_2(t) - B_2(T_1)| \geqslant \delta_x/16 \right) \\
&\quad + \mathbb{P}\Bigg(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U(t) - U(T_1)| \geqslant \delta_x^2/32, \\
&\qquad \sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \Bigg).
\end{aligned}
\tag{4.11}
$$

By the strong Markov property applied at T_1, along with parts (ii) and (iii) of Lemma 4.1 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 3.5 and Lemma 4.2 and Lemma 3.4, 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,

$$
\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_i(t) - B_i(T_1)| \geqslant \delta_x/16 \right) \leqslant C \frac{\rho(x, \widetilde{x})}{\delta_x^4}. \tag{4.12}
$$

for $i = 1, 2$. To handle the last term in (4.11), 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

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

Using this, we can bound the last term in (4.11) as

$$
\begin{aligned}
(4.13)\quad & \mathbb{P}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|U(t)-U(T_1)|\geqslant \delta_x^2/32,\\
&\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-B_1(T_1)|<\delta_x/16, T_1\leqslant \tau_{Q_1^*}(\mathbf{X})\Bigg)\\
&\leqslant \mathbb{P}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|U^*(t)-U^*(T_1)|\geqslant \delta_x^2/64\Bigg)\\
&+\mathbb{P}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-\hat{b}_1||B_2(t)-\hat{b}_2|\geqslant \delta_x^2/64,\\
&\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-B_1(T_1)|<\delta_x/16, T_1\leqslant \tau_{Q_1^*}(\mathbf{X})\Bigg).
\end{aligned}
$$

By conditioning at time T_1 and part (i) of Lemma 4.1, followed by applications of Lemma 4.2 and Lemma 3.4, we obtain

$$
\begin{aligned}
&\mathbb{E}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|U^*(t)-U^*(T_1)|\Bigg)^2\leqslant\\
&4\mathbb{E}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}\left|\int_{T_1}^t(B_2(s)-\hat{b}_2)dB_1(s)\right|\Bigg)^2\leqslant\\
&C\mathbb{E}((\tau-T_1)\wedge 1)^2\leqslant \mathbb{E}|A(T_1)\wedge 1|\leqslant C\rho(x,\widetilde{x}).
\end{aligned}
$$

Consequently, by the Chebyshev inequality

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

Moreover,

$$
\begin{aligned}
(4.15)\quad & \mathbb{P}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-\hat{b}_1||B_2(t)-\hat{b}_2|\geqslant \delta_x^2/64,\\
&\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-B_1(T_1)|<\delta_x/16, T_1\leqslant \tau_{Q_1^*}(\mathbf{X})\Bigg)\\
&\leqslant \mathbb{P}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_2(t)-\hat{b}_2|\geqslant \delta_x/8\Bigg)\\
&+\mathbb{P}\Bigg(\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-\hat{b}_1|\geqslant \delta_x/8,\\
&\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-B_1(T_1)|<\delta_x/16, T_1\leqslant \tau_{Q_1^*}(\mathbf{X})\Bigg).
\end{aligned}
$$

We use the fact $B_2(T_1) = \hat{b}_2$ and proceed exactly along the lines of the proof of (4.12) to obtain

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_2(t) - \hat{b}_2| \geqslant \delta_x/8\right) \leqslant C\frac{\rho(x,\widetilde{x})}{\delta_x^4}. \tag{4.16}$$

The second probability appearing on the right hand side of (4.15) can be bounded as follows

$$\begin{aligned}
&\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right. \\
&\qquad \left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right) \\
&\leqslant \mathbb{P}\left(\sup_{(T_1\wedge\tau_{Q_1^*}(\mathbf{X})) \leqslant t \leqslant (T_1\wedge\tau_{Q_1^*}(\mathbf{X}))+(\tau-(T_1\wedge\tau_{Q_1^*}(\mathbf{X})))\wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right. \\
&\qquad \left.\sup_{(T_1\wedge\tau_{Q_1^*}(\mathbf{X})) \leqslant t \leqslant (T_1\wedge\tau_{Q_1^*}(\mathbf{X}))+(\tau-(T_1\wedge\tau_{Q_1^*}(\mathbf{X})))\wedge 1} |B_1(t) - B_1(T_1\wedge\tau_{Q_1^*}(\mathbf{X}))| < \delta_x/16\right) \\
&\qquad\qquad \leqslant \mathbb{P}\left(|B_1(T_1\wedge\tau_{Q_1^*}(\mathbf{X})) - \hat{b}_1| > \delta_x/16\right).
\end{aligned} \tag{4.17}$$

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 (4.7), we get

$$\begin{aligned}
\mathbb{P}\left(|B_1(T_1\wedge\tau_{Q_1^*}(\mathbf{X})) - \hat{b}_1| > \delta_x/16\right) &\leqslant C\frac{\mathbb{E}|B_1(T_1\wedge\tau_{Q_1^*}(\mathbf{X})) - \hat{b}_1|^2}{\delta_x^2} \\
\leqslant C\frac{\mathbb{E}\sup_{0\leqslant t\leqslant T_1\wedge\tau_{Q_1^*}(\mathbf{X})} |B_1(t) - b_1|^2}{\delta_x^2} &\leqslant C\frac{\mathbb{E}(T_1\wedge\tau_{Q_1^*}(\mathbf{X}))}{\delta_x^2} \leqslant C\frac{|\mathbf{b}-\hat{\mathbf{b}}|}{\delta_x}.
\end{aligned}$$

Using this in (4.17),

$$\begin{aligned}
&\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right. \\
&\qquad \left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right) \leqslant C\frac{|\mathbf{b}-\hat{\mathbf{b}}|}{\delta_x}.
\end{aligned} \tag{4.18}$$

Using (4.16) and (4.18) in (4.15), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - \hat{b}_1||B_2(t) - \hat{b}_2| \geqslant \delta_x^2/64,\right. \\
&\qquad \left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right) \\
&\qquad\qquad \leqslant C\left(\frac{1}{\delta_x} + \frac{1}{\delta_x^4}\right)\rho(x,\widetilde{x}).
\end{aligned} \tag{4.19}$$

Finally, using (4.14) and (4.19) in (4.13),

$$\begin{aligned}(4.20)\quad \mathbb{P}\Bigg(&\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|U(t)-U(T_1)|\geqslant \delta_x^2/32,\\ &\sup_{T_1\leqslant t\leqslant T_1+(\tau-T_1)\wedge 1}|B_1(t)-B_1(T_1)|<\delta_x/16, T_1\leqslant \tau_{Q_1^*}(\mathbf{X})\Bigg)\\ &\leqslant C\left(\frac{1}{\delta_x}+\frac{1}{\delta_x^4}\right)\rho(x,\widetilde{x}).\end{aligned}$$

Using the estimates from (4.12) and (4.20) in (4.11), we get

$$\begin{aligned}(4.21)\quad &\mathbb{P}(\tau>\tau_{Q_1}(\mathbf{X})\wedge\sigma_{\delta_x^2/16}, T_1\leqslant\tau_{Q_1^*}(\mathbf{X})\wedge\sigma_{\delta_x^2/32}, \tau-T_1\leqslant 1)\\ &\leqslant C\left(\frac{1}{\delta_x}+\frac{1}{\delta_x^4}\right)\rho(x,\widetilde{x}).\end{aligned}$$

Using (4.10) and (4.21) in (4.9), we get

$$(4.22)\quad \mathbb{P}(\tau>\tau_{Q_1}(\mathbf{X})\wedge\sigma_{\delta_x^2/16}, T_1\leqslant\tau_{Q_1^*}(\mathbf{X})\wedge\sigma_{\delta_x^2/32})\leqslant C\left(1+\frac{1}{\delta_x}+\frac{1}{\delta_x^4}\right)\rho(x,\widetilde{x}).$$

Using the estimates (4.6), (4.8) and (4.22) in (4.5), we obtain

$$(4.23)\qquad \mathbb{P}(\tau>\tau_Q(\mathbf{X}))\leqslant C\left(1+\frac{1}{\delta_x}+\frac{1}{\delta_x^4}+\frac{(1+\delta_x)^3}{\delta_x^4}\right)\rho(x,\widetilde{x}).$$

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. □

The above theorem yields the gradient estimate formulated in Theorem 4.4. Before we can formulate our result, we explain the argument in the proof of [26, Proposition 4.1] that leads to (4.24).

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 (2.2). 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 [17, 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 [26, Lemma 2.1], so [17, Theorem 11.7] is applicable and we have that

$$(4.24)\qquad \|\nabla_{\mathcal{H}}u\|_{\mathcal{H}}\leqslant|\nabla_{\mathcal{H}}u|,$$

a.e. with respect to the Lebesgue measure.

Let $C\left(\overline{D}\right)$ be the space of functions that are continuous on the closure of the domain D. We also let $C^2\left(D\right)$ be the space of functions that are twice continuously differentiable in D.

Theorem 4.4. *Suppose $u \in C\left(\overline{D}\right) \cap C^2\left(D\right)$ such that $\Delta_{\mathcal{H}} u = 0$ 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 such that for every $x \in D$*

$$\left\|\nabla_{\mathcal{H}} u(x)\right\|_{\mathcal{H}} \leqslant \left|\nabla_{\mathcal{H}} u\right|(x) \leqslant C\left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1+\delta_x)^3}{\delta_x^4}\right) \underset{B(x,\alpha\delta_x)}{\operatorname{osc}} u. \tag{4.25}$$

Proof. It clearly suffices to consider the case $\alpha = 1$. Since u is continuous on $\overline{D}$, $\operatorname{osc}_{B(x,\delta_x)} u < \infty$. 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 $|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1| \leqslant 1/2$. Consider the coupling from Theorem 3.5 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 4.3 and the equivalence of the Carnot-Carathéodory metric d_{CC} and the pseudo-metric ρ, we have

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

Using the coupling from Theorem 3.5 and Itô's formula we have that

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

Since $u \in C\left(\overline{D}\right) \cap C^2\left(D\right)$ therefore (4.24) holds for every $x \in D$. Dividing out by $d_{CC}\left(x, \widetilde{x}\right)$ and using (4.24) we have that for every $x \in D$,

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

as needed. □

Corollary 4.5. *Let $u \in C\left(\overline{D}\right) \cap C^\infty\left(D\right)$ 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*

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

for every $x \in D$.

Proof. By [7, Corollary 5.7.3] we have the following Harnack inequality

$$\sup_{B(x,\alpha^*\delta_x)} u \leqslant C \inf_{B(x,\alpha^*\delta_x)} u \tag{4.26}$$

for $x \in D \subset \mathbb{H}^3$, where $\alpha^* \in (0,1]$, $C > 0$ are constants not depending on u, δ_x, x, D. Then Equations (4.25) and (4.26) give the desired result. □

We can use Corollary 4.5 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 4.5 and the Cheng-Yau inequality.

Corollary 4.6. *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*

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

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 $B(0,2)$. By Corollary 4.5

$$\frac{\|\nabla_{\mathcal{H}} u(x)\|_{\mathcal{H}}}{u(x)} \leq C' = C \sup_{x \in B(0,1)} \left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x} + \frac{(1+\delta_x)^3}{\delta_x^4} \right), \quad x \in B(0,1), \tag{4.27}$$

where C is the same constant as in Corollary 4.5. This implies that

$$\sup_{B(0,1),} \|\nabla_{\mathcal{H}} \log u\|_{\mathcal{H}} \leqslant C'. \tag{4.28}$$

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 (4.28) implies

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

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

5. Concluding remarks

Our work gives the first use of explicit non-Markovian coupling techniques to get geometric information in the sub-Riemannian setting. We would like to point out some potentially significant connections with a different approach to such a setting. K. Kuwada in [26] proved an important result on the duality of L^q-gradient estimates for the heat kernel of diffusions and their L^p-Wasserstein distances under the assumptions of volume doubling and a local Poincaré inequality, for any $p \in [1,\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$. Using this duality, he used the L^1-gradient estimate of the heat kernel for Brownian motion on the Heisenberg group obtained in [29] and [1] to derive L^∞-Wasserstein bounds. More precisely, he proved that if $d_W(x,y;t)$ denotes the L^∞-Wasserstein distance between the laws of Brownian motion on $\mathbb{H}^3$ starting from x and y at time $t > 0$, then

$$d_W(x,y;t) \leqslant K d_{CC}(x,y) \tag{5.1}$$

for some constant K that does not depend on x, y, t. The constant K is not known, the best estimate obtained so far is $K \geqslant \sqrt{2}$ (see [13]). Although we work with the total variation distance instead of the Wasserstein distance, Theorem 3.6 gives a better estimate of the distance between the laws of the two Brownian motions on

$\mathbb{H}^3$, as it not only captures the dependence on the starting points, but also gives the "*polynomial decay*" in time.

Our intention is to use the techniques developed in this article and in [2], to give a systematic way to explicitly construct non-Markovian couplings via spectral expansions, and connect it to the previous results on the heat kernels such as those in [13, 26, 29]. This will be addressed in future work.

Acknowledgement. S.B. is grateful for many helpful and motivating conversations with W. S. Kendall. The authors also thank Fabrice Baudoin for drawing our attention to Kuwada's results and the Cheng-Yau estimate in Corollary 4.6, and Iddo Ben-Ari for his interest in the subject and his insight into coupling techniques. We also thank the anonymous referee whose careful review and suggestions greatly improved the presentation of this paper.

References

[1] Dominique Bakry, Fabrice Baudoin, Michel Bonnefont, and Djalil Chafaï. On gradient bounds for the heat kernel on the Heisenberg group. *J. Funct. Anal.*, 255(8):1905–1938, 2008.

[2] Sayan Banerjee and Wilfrid S. Kendall. Coupling the Kolmogorov diffusion: maximality and efficiency considerations. *Adv. in Appl. Probab.*, 48(A):15–35, 2016.

[3] Sayan Banerjee and Wilfrid S. Kendall. Rigidity for Markovian maximal couplings of elliptic diffusions. *Probab. Theory Related Fields*, 168(1-2):55–112, 2017.

[4] Fabrice Baudoin, Michel Bonnefont, and Nicola Garofalo. A sub-Riemannian curvature-dimension inequality, volume doubling property and the Poincaré inequality. *Math. Ann.*, 358(3-4):833–860, 2014.

[5] 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(1):151–219, 2017.

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

[7] A. Bonfiglioli, E. Lanconelli, and F. Uguzzoni. *Stratified Lie groups and potential theory for their sub-Laplacians.* Springer Monographs in Mathematics. Springer, Berlin, 2007.

[8] Ovidiu Calin, Der-Chen Chang, and Peter Greiner. *Geometric analysis on the Heisenberg group and its generalizations*, volume 40 of *AMS/IP Studies in Advanced Mathematics.* American Mathematical Society, Providence, RI; International Press, Somerville, MA, 2007.

[9] 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(1):41–72, 2010.

[10] S. Y. Cheng and S. T. Yau. Differential equations on Riemannian manifolds and their geometric applications. *Comm. Pure Appl. Math.*, 28(3):333–354, 1975.

[11] M. Cranston. Gradient estimates on manifolds using coupling. *J. Funct. Anal.*, 99(1):110–124, 1991.

[12] M. Cranston. A probabilistic approach to gradient estimates. *Canad. Math. Bull.*, 35(1):46–55, 1992.

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

[14] Sheldon Goldstein. Maximal coupling. *Z. Wahrsch. Verw. Gebiete*, 46(2):193–204, 1978/79.

[15] Maria Gordina and Thomas Laetsch. Sub-Laplacians on Sub-Riemannian Manifolds. *Potential Anal.*, 44(4):811–837, 2016.

[16] David Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.

[17] Piotr Hajłasz and Pekka Koskela. Sobolev met Poincaré. *Mem. Amer. Math. Soc.*, 145(688):x+101, 2000.

[18] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[19] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1991.

[20] Kari Karhunen. über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 1947(37):79, 1947.
[21] Wilfrid S. Kendall. Nonnegative Ricci curvature and the Brownian coupling property. *Stochastics*, 19(1-2):111–129, 1986.
[22] Wilfrid S. Kendall. Coupled Brownian motions and partial domain monotonicity for the Neumann heat kernel. *J. Funct. Anal.*, 86(2):226–236, 1989.
[23] Wilfrid S. Kendall. Coupling all the Lévy stochastic areas of multidimensional Brownian motion. *Ann. Probab.*, 35(3):935–953, 2007.
[24] Wilfrid S. Kendall. Coupling time distribution asymptotics for some couplings of the Lévy stochastic area. In *Probability and mathematical genetics*, volume 378 of *London Math. Soc. Lecture Note Ser.*, pages 446–463. Cambridge Univ. Press, Cambridge, 2010.
[25] Kazumasa Kuwada. On uniqueness of maximal coupling for diffusion processes with a reflection. *J. Theoret. Probab.*, 20(4):935–957, 2007.
[26] Kazumasa Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.*, 258(11):3758–3774, 2010.
[27] Kazumasa Kuwada and Karl-Theodor Sturm. A counterexample for the optimality of Kendall-Cranston coupling. *Electron. Comm. Probab.*, 12:66–72 (electronic), 2007.
[28] 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.
[29] Hong-Quan Li. Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg. *J. Funct. Anal.*, 236(2):369–394, 2006.
[30] Michel Loève. Sur l'équivalence asymptotique des lois. *C. R. Acad. Sci. Paris*, 227:1335–1337, 1948.
[31] Richard Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.
[32] Daniel Neuenschwander. *Probabilities on the Heisenberg group*, volume 1630 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1996. Limit theorems and Brownian motion.
[33] J. W. Pitman. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(4):315–322, 1976.
[34] S. T. Yau R. Schoen. *Lectures on Differential Geometry*. International Press, Boston, 1994.
[35] M. Yu. Sverchkov and S. N. Smirnov. Maximal coupling for processes in $D[0,\infty]$. *Dokl. Akad. Nauk SSSR*, 311(5):1059–1061, 1990.
[36] 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).
[37] S.T. Yau. Harmonic functions on complete riemannian manifolds. *Comm. Pure Appl. Math.*, 28:201–228, 1975.
[38] Marc Yor. The laws of some Brownian functionals. In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, pages 1105–1112. Math. Soc. Japan, Tokyo, 1991.

⋆ Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599, U.S.A.
E-mail address: sayan@email.unc.edu

† Department of Mathematics, University of Connecticut, Storrs, CT 06269, U.S.A.
E-mail address: maria.gordina@uconn.edu

† Department of Mathematics, University of Connecticut, Storrs, CT 06269, U.S.A.
E-mail address: phanuel.mariano@uconn.edu