HEAT KERNEL ASYMPTOTICS ON SUB-RIEMANNIAN MANIFOLDS WITH SYMMETRIES AND APPLICATIONS TO THE BI-HEISENBERG GROUP

Davide Barilari

IMJ-PRG, Université Paris Diderot, UMR CNRS 7586, Paris, France

davide.barilari@imj-prg.fr

Ugo Boscain

CNRS, CMAP Ecole Polytechnique, Paris, Equipe INRIA GECO Saclay-Île-de-France, Paris, France

ugo.boscain@polytechnique.edu

Robert W. Neel

Department of Mathematics, Lehigh University, Bethlehem, PA, USA

robert.neel@lehigh.edu

June 3, 2016

ABSTRACT. By adapting a technique of Molchanov, we obtain the heat kernel asymptotics at the sub-Riemannian cut locus, when the cut points are reached by an r-dimensional parametric family of optimal geodesics. We apply these results to the bi-Heisenberg group, that is, a nilpotent left-invariant sub-Riemannian structure on $\mathbb{R}^5$ depending on two real parameters α_1 and α_2. We develop some results about its geodesics and heat kernel associated to its sub-Laplacian and we illuminate some interesting geometric and analytic features appearing when one compares the isotropic ($\alpha_1 = \alpha_2$) and the non-isotropic cases ($\alpha_1 \neq \alpha_2$). In particular, we give the exact structure of the cut locus, and we get the complete small-time asymptotics for its heat kernel.

1. INTRODUCTION

The three-dimensional Heisenberg group is the simplest example of a sub-Riemannian manifold. It already possesses some features that are typical of more general sub-Riemannian structures: the Hausdorff dimension is bigger than the topological dimension, the cut locus and the conjugate locus starting from a point are adjacent to the point itself, spheres are not smooth, even for small radius, and the heat kernel small-time asymptotics on the diagonal are affected by the presence of the trivial abnormal minimizer (or, in other words, in sub-Riemannian geometry q is not a smooth point for the exponential map based at q). The Heisenberg group is also the nilpotent approximation of any three-dimensional contact structure.

The natural next example, that shares with the Heisenberg group many of these properties, is the sub-Riemannian structure on $\mathbb{R}^5$ (where we consider coordinates (x_1, y_1, x_2, y_2, z)) where the following vector fields

$$(1) \quad \begin{cases} X_1 = \partial_{x_1} - \dfrac{\alpha_1}{2} y_1 \partial_z, & Y_1 = \partial_{y_1} + \dfrac{\alpha_1}{2} x_1 \partial_z, \\ X_2 = \partial_{x_2} - \dfrac{\alpha_2}{2} y_2 \partial_z, & Y_2 = \partial_{y_2} + \dfrac{\alpha_2}{2} x_2 \partial_z. \end{cases} \quad , \quad \alpha_1, \alpha_2 \geq 0, \ (\alpha_1, \alpha_2) \neq (0, 0),$$

define an orthonormal frame. Usually the case $\alpha_1 = \alpha_2$ is called the five-dimensional Heisenberg group, while the case $\alpha_1 \neq \alpha_2$ is referred as the *non-isotropic* five-dimensional Heisenberg group, see [11]. In the following we refer to the structure defined by (1) as the *bi-Heisenberg group*.

When the bi-Heisenberg group is obtained as the nilpotent approximation of a sub-Riemannian structure with growth vector $(4, 5)$, then the parameters α_1 and α_2 keep track of some of the local structure of the original manifold. In other words, the presence of parameters is a consequence

of the fact that the tangent space (or more precisely the nilpotent approximation) of a structure with growth vector $(4,5)$ depends on the point. (See [2] for more details about the nilpotent approximation of a sub-Riemannian structure.)

Many features of the bi-Heisenberg group change when the two parameters become equal. The main purpose of this paper is to study these differences, in particular, as it concerns the structure of the cut locus and the behavior of the heat kernel. An additional important case is the one in which one of the two parameters vanishes. This is the case in which the bi-Heisenberg structure is non-contact but is diffeomorphic to the product $\mathbb{H} \times \mathbb{R}^2$, where $\mathbb{H}$ denotes the three-dimensional Heisenberg group. This case should be treated separately since abnormal minimizers are present.

The cut locus is obtained explicitly using known formulas for the cut time along each geodesics, see [11, 2]. The heat kernel asymptotics are obtained as a consequence of the main result of the paper, namely Theorem 16, which permits one to deduce the small-time heat kernel asymptotics at cut loci produced by symmetries of the exponential map. This result is obtained by refining a general method developed by Molchanov [18] and the authors in [9, 7] (note in particular that it is a more detailed, sub-Riemannian version of Case 3.3 in [18]). In the case of the bi-Heisenberg group, the cut points are also conjugate points and, as a consequence, one gets quite different asymptotic expansions in and out of the cut locus.

We note that an integral expression for the heat kernel of the bi-Heisenberg group is known, see [11]. From this expression is not difficult to obtain an asymptotic expansion on the z-axis, but the extraction of the asymptotics on points other than these seems more difficult. We also note that stochastic approaches to small-time heat kernel asymptotics are possible, as in the recent work of [17]. While stochastic methods are quite powerful and provide insight into the entire diffusion associated to the sub-Laplacian, they are also technically demanding. In contrast, the Molchanov-inspired approach of the present paper treats conjugacy in a relatively simple, finite-dimensional way, and this makes it well suited to deriving explicit and geometrically meaningful asymptotics.

1.1. **Structure of the paper.** Since the main result (Theorem 16) requires several preliminaries, we discuss first its application to the bi-Heisenberg group, in the next subsection.

After recalling basic definitions of sub-Riemannian geometry (Section 2), in Section 3 we state and prove Theorem 16. Section 4 is devoted to the explicit computation of the optimal synthesis on the bi-Heisenberg group, together with the explicit formulas for the cut locus. Finally in Section 5 we show that using the results of [11] one can recover the expansion we found on a particular subset of the cut locus.

1.2. **Description of the results for the bi-Heisenberg group.**

Theorem 1 (Structure of the cut locus)**.** *The cut locus* Cut_0 *starting from the origin for the bi-Heisenberg group (i.e., the sub-Riemannian metric on $\mathbb{R}^5$ for which (1) is an orthonormal frame) is characterized as follows:*

(i) *if $0 < \alpha_2 = \alpha_1$ then*
$$\mathrm{Cut}_0 = \{(0,0,0,0,z), \ z \in \mathbb{R} \setminus \{0\}\},$$

(ii) *if $0 < \alpha_2 < \alpha_1$ then*
$$\mathrm{Cut}_0 = \{(0,0,x_2,y_2,z), \ |z| \geq (x_2^2 + y_2^2)K(\alpha_1,\alpha_2), \ (x_2,y_2,z) \in \mathbb{R}^3 \setminus \{0\}\},$$
where
$$K(\alpha_1,\alpha_2) := \frac{\alpha_2^2}{8}\left(\frac{2\pi}{\alpha_1} - \frac{1}{\alpha_2}\sin\left(2\pi\frac{\alpha_2}{\alpha_1}\right)\right)\sin^{-2}\left(\pi\frac{\alpha_2}{\alpha_1}\right),$$

(iii) *if $0 = \alpha_2 < \alpha_1$ then*
$$\mathrm{Cut}_0 = \{(0,0,x_2,y_2,z), \ (x_2,y_2,z) \in \mathbb{R}^3, z \neq 0\}.$$

Moreover, in case (i) every point of the cut locus is reached by a family of optimal geodesics parametrized by S^3; in cases (ii)-(iii) every point of the cut locus is reached by a family of optimal geodesics parametrized by S^1. As a consequence all cut points are also conjugate points.

Notice that the theorem above covers all the possible cases, up to exchanging the role of the indexes 1 and 2, which simply gives an isometric structure.

Notice also that in case (i) the cut locus is a one-dimensional manifold; in cases (ii)-(iii) the cut locus is a 3-dimensional manifold. Moreover $K(\alpha_1, \alpha_2) \to +\infty$ for $\alpha_2 \to \alpha_1^-$ and $K(\alpha_1, \alpha_2) \to 0$ for $\alpha_2 \to 0^+$. Hence one can recover (i) and (iii) as limit cases of (ii). In all cases the cut locus is symmetric with respect to rotation along the z axes.

In the following d denotes the sub-Riemannian distance and $p_t(q_1, q_2)$ is the fundamental solution to the heat equation $\partial_t u = \Delta u$, where Δ is the sub-Laplacian defined on the bi-Heisenberg group (see Section 2 for precise definitions).

Theorem 2 (Asymptotics of the heat kernel)**.** *There exist positive constants C_i (depending on q) such that for $t \to 0$, the heat kernel for the bi-Heisenberg group satisfies*

(i) *On-diagonal asymptotics:*

$$p_t(0,0) = \frac{C_1 + O(t)}{t^3}.$$

(ii) *If $q \notin \mathrm{Cut}_0 \cup \{0\}$, then*

$$p_t(0,q) = \frac{C_2 + O(t)}{t^{5/2}} e^{-d^2(0,q)/4t}.$$

(iii) *If $q \in \mathrm{Cut}_0$ and*

    (a) $0 < \alpha_2 = \alpha_1$, *then*

$$p_t(0,q) = \frac{C_3 + O(t)}{t^4} e^{-d^2(0,q)/4t},$$

    (b) $0 \leq \alpha_2 < \alpha_1$, *then*

$$p_t(0,q) = \frac{C_4 + O(t)}{t^3} e^{-d^2(0,q)/4t}.$$

Case (i) is well known and can be obtained from the explicit expression of the heat kernel or applying the result of Ben Arous [12]. Notice that in this case one obtains $p_t(0,0) = t^{-Q/2}(C_1 + O(t))$ where $Q = 6$ is the Hausdorff dimension. Case (ii) is a consequence of a result of Ben Arous [13] and from the explicit expression of the cut locus (cf. Theorem 1). Case (iii) is a consequence of Theorem 16. As explained in the introduction, the case in which $\alpha_2 = 0$ should be treated separately as a consequence of the presence of abnormal extremals.

Remark 3. In Theorems 1 and 2 it is not restrictive to consider the origin as reference point, due to the left-invariance of the structure.

Remark 4. With the same techniques, Theorems 1 and 2 can be generalized to the case of nilpotent structure of step 2 on $\mathbb{R}^{2\ell+1}$ with distribution of codimension 1. Notice that these structures depend on ℓ real parameters $\alpha_1, \ldots, \alpha_\ell$. For an explicit description see [2] (cf. also [11]). In this case, the structure of the cut locus and the heat kernel asymptotics are determined by how many of these parameters coincide with the maximal one.

2. Basic definitions

In what follows M is a connected orientable smooth manifold of dimension $n \geq 3$.

Definition 5. A *sub-Riemannian structure* on M is a pair $(\mathcal{D}, g)$, where

(i) $\mathcal{D}$ is a smooth vector distribution of constant rank $k < n$ satisfying the *Hörmander condition*, i.e. a smooth map that associates to $q \in M$ a k-dimensional subspace $\mathcal{D}_q$ of $T_q M$ such that

$$\text{span}\{[X_1, [\ldots [X_{j-1}, X_j]]]_q \mid X_i \in \mathfrak{X}(\mathcal{D}), \ j \in \mathbb{N}\} = T_q M, \ \forall \, q \in M,$$

where $\mathfrak{X}(\mathcal{D})$ denotes the set of *horizontal smooth vector fields* on M, i.e.

$$\mathfrak{X}(\mathcal{D}) = \{X \in \text{Vec}(M) \mid X(q) \in \mathcal{D}_q \ \forall \, q \in M\}.$$

(ii) g_q is a Riemannian metric on $\mathcal{D}_q$ which is smooth as function of q. We denote the norm of a vector $v \in \mathcal{D}_q$ by $|v|_g = \sqrt{g_q(v, v)}$.

A Lipschitz continuous curve $\gamma : [0, T] \to M$ is said to be *horizontal* (or *admissible*) if

$$\dot{\gamma}(t) \in \mathcal{D}_{\gamma(t)} \qquad \text{for a.e. } t \in [0, T].$$

The *length* of an horizontal curve $\gamma : [0, T] \to M$ is

$$\ell(\gamma) = \int_0^T |\dot{\gamma}(t)|_g \ dt.$$

Notice that $\ell(\gamma)$ is invariant under time reparametrization of γ. The *sub-Riemannian (or Carnot-Carathéodory) distance* on M is the function

$$(2) \qquad d(q_0, q_1) = \inf\{\ell(\gamma) \mid \gamma(0) = q_0, \gamma(T) = q_1, \gamma \text{ horizontal}\}.$$

The hypothesis of connectedness of M and the Hörmander condition guarantee the finiteness and the continuity of $d : M \times M \to \mathbb{R}$ with respect to the topology of M (Chow-Rashevsky theorem, see for instance [6]). The function d defined in (2) gives to M the structure of a metric space (see [6]) compatible with the original topology of M. The sub-Riemannian manifold is said *complete* if (M, d) is complete as a metric space.

Locally, the pair $(\mathcal{D}, g)$ can be specified by assigning a set of k smooth vector fields spanning $\mathcal{D}$, being an orthonormal frame for g, i.e.

$$\mathcal{D}_q = \text{span}\{X_1(q), \ldots, X_k(q)\}, \qquad g_q(X_i(q), X_j(q)) = \delta_{ij}.$$

In this case, the set $\{X_1, \ldots, X_k\}$ is called a *local orthonormal frame* for the sub-Riemannian metric.

When a local orthonormal frame is defined globally, the sub-Riemannian manifold is called *trivializable*.

The sub-Riemannian metric can also be expressed locally in "control form" as follows. Fix $\{X_1, \ldots, X_k\}$ a local orthonormal frame and consider the control system:

$$\dot{q} = \sum_{i=1}^{k} u_i X_i(q).$$

The problem of finding the shortest curve minimizing that joins two fixed points q_0, $q_1 \in M$ is naturally formulated as the optimal control problem of minimizing

$$(3) \qquad \int_0^T \sqrt{\sum_{i=1}^{k} u_i^2(t)} \ dt \quad \text{subject to} \quad q(0) = q_0, \ q(T) = q_1.$$

A sub-Riemannian structure on M is said to be *left-invariant* if $M = G$, where G is a Lie group, and both $\mathcal{D}$ and g are left-invariant over G. Left-invariant sub-Riemannian manifolds are trivializable. An example of left-invariant sub-Riemannian manifold is provided by the bi-Heisenberg group presented in the introduction (cf. also Section 4).

Now we briefly recall some facts about sub-Riemannian geodesics. In particular, we define the sub-Riemannian exponential map.

Definition 6. A *geodesic* for a sub-Riemannian manifold $(M, \mathcal{D}, g)$ is an admissible curve $\gamma : [0, T] \to M$ such that $|\dot{\gamma}(t)|_g$ is constant and, for every $t \in [0, T]$ there exists $\varepsilon > 0$ such that the restriction $\gamma_{|[t-\varepsilon, t+\varepsilon]}$ is a minimizer of $\ell(\cdot)$. A geodesic satisfying $|\dot{\gamma}(t)|_g = 1$ is said to be parameterized by arclength.

A version of the Hopf-Rinow theorem (see [15, Chapter 2]) ensures that if the sub-Riemannian manifold is complete as metric space, then all geodesics are defined for every $t \geq 0$ and that for every two points there exists a minimizing geodesic connecting them.

Left invariant sub-Riemannian manifolds are always complete.

Trajectories minimizing the distance between two points are solutions of first-order necessary conditions for optimality, which in the case of sub-Riemannian geometry are described as follows.

Theorem 7. *Let $q(\cdot) : t \in [0, T] \mapsto q(t) \in M$ be a solution of the minimization problem (3) such that $|\dot{q}(t)|_g$ is constant, and let $u(\cdot)$ be the corresponding control. Then there exists a Lipschitz map $p(\cdot) : t \in [0, T] \mapsto p(t) \in T^*_{q(t)}M \setminus \{0\}$ such that one and only one of the following conditions holds:*

(i) $\dot{q} = \dfrac{\partial H}{\partial p}, \quad \dot{p} = -\dfrac{\partial H}{\partial q}, \quad u_i(t) = \langle p(t), X_i(q(t)) \rangle,$

   where $H(q, p) = \frac{1}{2} \sum_{i=1}^{k} \langle p, X_i(q) \rangle^2$.

(ii) $\dot{q} = \dfrac{\partial \mathcal{H}}{\partial p}, \quad \dot{p} = -\dfrac{\partial \mathcal{H}}{\partial q}, \quad 0 = \langle p(t), X_i(q(t)) \rangle,$

   where $\mathcal{H}(t, q, p) = \sum_{i=1}^{k} u_i(t) \langle p, X_i(q) \rangle$.

For an elementary proof of Theorem 7 see [3].

Remark 8. If $(q(\cdot), p(\cdot))$ is a solution of (i) (resp. (ii)) then it is called a *normal extremal* (resp. *abnormal extremal*). It is well known that if $(q(\cdot), p(\cdot))$ is a normal extremal then $q(\cdot)$ is a geodesic (see [3, 6]). This does not hold in general for abnormal extremals. An admissible trajectory $q(\cdot)$ can be at the same time normal and abnormal (corresponding to different covectors). If an admissible trajectory $q(\cdot)$ is normal but not abnormal, we say that it is *strictly normal*.

Definition 9. A minimizer $\gamma : [0, T] \to M$ is said to be *strongly normal* if for every $[t_1, t_2] \subset [0, T]$, $\gamma|_{[t_1, t_2]}$ is not an abnormal minimizer.

In the following we denote by $(q(t), p(t)) = e^{t\vec{H}}(q_0, p_0)$ the solution of (i) with initial condition $(q(0), p(0)) = (q_0, p_0)$. Moreover we denote by $\pi : T^*M \to M$ the canonical projection.

Normal extremals (starting from q_0) parametrized by arclength correspond to initial covectors $p_0 \in \Lambda_{q_0} := \{p_0 \in T^*_{q_0}M | H(q_0, p_0) = 1/2\}$.

Definition 10. Let $(M, \mathcal{D}, g)$ be a complete sub-Riemannian manifold and $q_0 \in M$. We define the *exponential map* starting from q_0 as

$$\exp_{q_0} : \Lambda_{q_0} \times \mathbb{R}^+ \to M, \qquad \exp_{q_0}(p_0, t) = \pi(e^{t\vec{H}}(q_0, p_0)).$$

Next, we recall the definition of cut and conjugate time.

Definition 11. Let $q_0 \in M$ and $\gamma(t)$ an arclength geodesic starting from q_0. The *cut time* for γ is $t_{\text{cut}}(\gamma) = \sup\{t > 0, \gamma|_{[0,t]}$ is optimal$\}$. The *cut locus* from q_0 is the set $\text{Cut}_{q_0} = \{\gamma(t_{\text{cut}}(\gamma)), \gamma$ arclength geodesic from $q_0\}$.

Definition 12. Let $q_0 \in M$ and $\gamma(t)$ a normal arclength geodesic starting from q_0 with initial covector p_0. Assume that γ is not abnormal.

(i) The *first conjugate time* of γ is $t_{\text{con}}(\gamma) = \min\{t > 0, (p_0, t)$ is a critical point of $\exp_{q_0}\}$.

(ii) The *first conjugate locus* from q_0 is the set

$$\text{Con}_{q_0} = \{\gamma(t_{\text{con}}(\gamma)), \gamma \text{ arclength geodesic from } q_0\}.$$

It is well known that, for a geodesic γ which is not abnormal, the cut time $t_* = t_{\text{cut}}(\gamma)$ is either equal to the conjugate time, or there exists another geodesic $\widetilde{\gamma}$ such that $\gamma(t_*) = \widetilde{\gamma}(t_*)$ (see for instance [3, 5]).

Remark 13. In sub-Riemannian geometry, the map $\exp_{q_0}$ is never a local diffeomorphism at $0 \in T_{q_0}^*M$. As a consequence one can show that the sub-Riemannian balls are never smooth and both the cut and the conjugate loci from q_0 are adjacent to the point q_0 itself (see [1, 3]).

2.1. The sub-Laplacian.

In this section we define the sub-Riemannian Laplacian on a sub-Riemannian manifold $(M, \mathcal{D}, g)$ endowed with a smooth volume μ.

The sub-Laplacian is the natural generalization of the Laplace-Beltrami operator defined on a Riemannian manifold, defined as the divergence of the gradient. The *horizontal gradient* is the unique operator $\nabla : \mathcal{C}^\infty(M) \to \mathfrak{X}(\mathcal{D})$ defined by

$$g_q(\nabla\varphi(q), v) = d\varphi_q(v), \quad \forall\, \varphi \in \mathcal{C}^\infty(M), q \in M, v \in \mathcal{D}_q.$$

By construction, it is a horizontal vector field. If $X_1, \ldots, X_k$ is a local orthonormal frame, it is easy to see that it is written as follows $\nabla\varphi = \sum_{i=1}^k X_i(\varphi)X_i$, where $X_i(\varphi)$ denotes the Lie derivative of φ in the direction of X_i.

The divergence of a vector field X with respect to a volume μ is the function $\text{div}_\mu X$ defined by the identity $L_X\mu = (\text{div}_\mu X)\mu$, where L_X stands for the Lie derivative with respect to X.

The sub-Laplacian associated with the sub-Riemannian structure, i.e. $\Delta\varphi = \text{div}_\mu(\nabla\varphi)$, is written in a local orthonormal frame $X_1, \ldots, X_k$ as follows

$$(4) \qquad \Delta = \sum_{i=1}^k X_i^2 + (\text{div}_\mu X_i)X_i.$$

Notice that Δ is always expressed as the sum of squares of the element of the orthonormal frame plus a first order term that belongs to the distribution and depends on the choice of the volume μ.

In what follows we denote p_t the heat kernel associated to the heat equation $\partial_t u = \Delta u$. Its existence, smoothness, symmetry, and positivity are guaranteed by classical results since the operator (4) is in divergence form and M is complete, see for instance [19].

Remark 14. In the case of a left-invariant structure on a Lie group (and in particular for a nilpotent structure), an intrinsic choice of volume is provided by the left Haar measure. When the Lie group is unimodular, which is the case in particular for nilpotent groups, the left Haar measure is both left and right invariant. Left invariant vector fields on a Lie group G are divergence free with respect to the right Haar measure. As a consequence the sub-Laplacian associated to the Haar measure has the form of "sum of squares" (see also [4])

$$\Delta = \sum_{i=1}^k X_i^2.$$

For unimodular left invariant structures one can show that the Haar measure coincides with the Popp volume, that is a well defined measure for equiregular sub-Riemannian manifolds (see [10]).

3. Small time asymptotic of the heat kernel

We begin by recalling the general framework for computing asymptotics developed in [9], which is built upon the basic insight of Molchanov [18]. Let M be a complete sub-Riemannian (or Riemannian) manifold of dimension n. Consider distinct points x and y such that every minimizing geodesic connecting them is strongly normal. (In this section, we will use x, y and z to denote points of M, to avoid the need for subscripts as in q_0, q_1, etc. Because we work on a general sub-Riemannian manifold, there should be no confusion with using these letters to denote Euclidean coordinates on $\mathbb{R}^5$, as when we specifically discuss the bi-Heisenberg group.) Let $\Gamma \subset M$ be the set of midpoints of minimizing geodesics from x to y (or from y to x, the situation is symmetric), and let $N(\Gamma)$ denote a neighborhood of Γ. Further, let $h_{x,y}(z) = \left[d^2(x,z) + d^2(z,y) \right]/2$; then $h_{x,y}$ attains its minimum of $d^2(x,y)/4$ exactly on Γ (see Lemma 21 of [9]). Because $h_{x,y}$ is continuous and sub-Riemannian balls are compact (by completeness of M), it follows that Γ is compact. Further, since strong normality is an open condition and Γ is a positive distance from x, y, and their cut loci, $h_{x,y}$ is smooth in a neighborhood of Γ. We assume that $N(\Gamma)$ is small enough so that its closure is positive distance from x, y, and their cut loci and that $h_{x,y}$ is smooth on its closure.

Continuing, suppose we give M a sub-Laplacian Δ coming from a smooth volume μ as in Section 2.1. Let $\Sigma \subset M \times M$ be the set of (x,y) such that $x \neq y$ and there exists a unique, strongly normal, non-conjugate minimizing geodesic connecting x and y. Then from the results contained in Ben Arous [12], it follows that for $(x,y) \in \Sigma$,

$$p_t(x,y) = \frac{1}{t^{n/2}} \exp \left(-\frac{d^2(x,y)}{4t} \right) (c_0(x,y) + O(t)) \quad \text{as } t \to 0,$$

where $c_0(x,y)$ is smooth and positive on Σ and the $O(t)$ is uniform over compact subsets of Σ.

We can now give the basic, general result for determining small-time asymptotics at the cut locus. The first part of Theorem 27 of [9] gives the following.

Theorem 15. *Let M be an n-dimensional (complete) sub-Riemannian manifold with smooth measure μ and corresponding heat kernel p_t, and let x and y be distinct points such that all minimal geodesics from x to y are strongly normal. Then with the above notation, for any sufficiently small $N(\Gamma)$, we have*

$$p_t(x,y) = \int_{N(\Gamma)} \left(\frac{2}{t} \right)^n e^{-h_{x,y}(z)/t} \left(c_0(x,z) c_0(z,y) + O(t) \right) \mu(dz),$$

where the $O(t)$ term is uniform over $N(\Gamma)$.

We wish to apply this theorem in concrete cases where we have an explicit expression for $h_{x,y}$, and possibly for c_0. For what follows, we will need the case when Γ is an embedded (necessarily compact) r-dimensional smooth submanifold of M and the Hessian of $h_{x,y}$ restricted to normal bundle of $T\Gamma$ in TM is non-degenerate (because in general there is no Riemannian metric on M, there is no notion of a canonical representative of the normal bundle as an orthogonal complement, but non-degeneracy of the Hessian is nonetheless well defined). In other words, $h_{x,y}$ is a Morse-Bott function with critical set Γ. Note that because the Hessian of $h_{x,y}$ at $z \in \Gamma$ is always non-degenerate along the minimal geodesic from x to y through z, we must have $0 \leq r < n$.

In this situation, for any local coordinates $(u_1, \ldots, u_r)$ on an open subset U of Γ, we can extend them to local coordinates $(u_1, \ldots, u_r, u_{r+1}, \ldots, u_n)$ on an open subset V of M such that Γ is locally given by $u_{r+1} = \cdots = u_n = 0$ and $U = V \cap \Gamma$. Further, perhaps by shrinking U and V, we can assume that these coordinates can be extended to a neighborhood of $\overline{V}$. We call such a

local coordinate system on M adapted to Γ. Then μ has a smooth density $F(u)$ with respect to $du_1 \cdots du_n$, and $F(u)$ along all with of its derivatives in the u_i are bounded on V. Next, note that for $z \in U$, $\partial_{u_i} \partial_{u_j} h_{x,y}(z) = 0$ if either $i \leq r$ or $j \leq r$. Then let $H^c h_{x,y}(z)$ be the $(n-r) \times (n-r)$ matrix with i, j entry $\partial_{u_{r+i}} \partial_{u_{r+j}} h_{x,y}(z)$ (this depends on the coordinate system u, but we supress that in the notation). Note that all the entries of $H^c h_{x,y}(z)$ along with all of their derivatives in the u_i are bounded on U. Now we claim that there is measure $\nu_\Gamma^{\mu, h_{x,y}}$ on Γ, induced by μ and the Morse-Bott function $h_{x,y}$, that is defined by having the following local expression in any local adapted coordinates,

$$\nu_\Gamma^{\mu, h_{x,y}} = \frac{F(u_1, \ldots, u_r, 0, \ldots, 0)}{\sqrt{\det H^c h_{x,y}(u_1, \ldots, u_r)}} du_1 \ldots du_r.$$

To see that this is a globally well-defined measure on Γ, first note that this expression depends only on the local coordinates $(u_1, \ldots, u_r)$ on Γ and not on the extension to adapted coordinates on M; more precisely, the factor of $F/\sqrt{\det H^c h_{x,y}}$ is invariant under change of fiber coordinates on the normal bundle. This is because if we change the coordinates $(u_{r+1}, \ldots, u_n)$ on any fixed fiber of the normal bundle to some other adapted choice, both F^Γ and $\sqrt{\det H^c h_{x,y}}$ are multiplied by the aboslute value of the determinant of the Jacobian, which clearly cancels. Further, if we change the coordinates $(u_1, \ldots, u_r)$ on Γ, it is clear that the given expression for $\nu_\Gamma^{\mu, h_{x,y}}$ transforms like a measure, and this is enough to establish the claim.

The motivation for introducing $\nu_\Gamma^{\mu, h_{x,y}}$ is that it encodes the relevant behavior of both μ and the Hessian of $h_{x,y}$ needed for the following asymptotic expansion.

Theorem 16. *Let M be an n-dimensional complete sub-Riemannian manifold provided with a smooth volume μ and let p_t be the heat kernel of the sub-Riemannian heat equation. Let x and y be distinct and assume that every optimal geodesic joining x to y is strongly normal. Define*

$$\mathcal{O} := \{\bar{p} \in \Lambda_x \mid \exp_x(\bar{p}, d(x, y)) = y\}$$

Assume that:

    (i) *$\mathcal{O}$ is a submanifold of Λ_x of dimension r.*
    (ii) *for every $\bar{p} \in \mathcal{O}$ we have $\dim \ker D_{\bar{p}, d(x,y)} \exp_x = r$.*

Then there exists a positive constant C such that

$$p_t(x, y) = \frac{C + O(t)}{t^{\frac{n+r}{2}}} e^{-d^2(x,y)/4t}, \qquad as\ t \to 0.$$

Further, $\exp_x(\cdot, d(x,y)/2)$ maps $\mathcal{O}$ diffeomorphically onto Γ, and the Hessian of $h_{x,y}$ restricted to the normal bundle of $T\Gamma$ in TM is non-degenerate, so that $\nu_\Gamma^{\mu, h_{x,y}}$ is defined as above. Finally, the constant C in the expansion is given by

$$C = 2^{(3n-r)/2} \pi^{(n-r)/2} \int_{z \in \Gamma} c_0(x, z) c_0(z, y) \nu_\Gamma^{\mu, h_{x,y}}(dz).$$

Proof. By its definition, Γ is the image of $\mathcal{O}$ under $\exp_x(\cdot, d(x,y)/2)$. Take $\bar{p} \in \mathcal{O}$. By the assumption of strong normality and basic properties of the exponential map, $\exp_x(\cdot, d(x,y)/2)$ is a diffeomorphism from a neighborhood of $\bar{p}$ in Λ_x to a neighborhood of $z_0 = \exp_x(\bar{p}, d(x,y)/2)$ in M. Thus Γ is locally (near z_0) an immersed submanifold of M of dimension r. Further, because the geodesics corresponding to covectors in Λ_x are optimal until time $d(x, y)$ and the cut locus of x is closed, $\exp_x(\bar{p}, d(x,y)/2)$ is injective on a neighborhood of $\mathcal{O}$. It follows that Γ is a compact embedded submanifold of M of dimension r, that $\exp_x(\bar{p}, d(x,y)/2)$ is a diffeomorphism from a neighborhood of $\mathcal{O}$ in Λ_x to a neighborhood of Γ in M that restricts to a diffeomorphism from $\mathcal{O}$ to Γ, and that $\mathcal{O}$ is compact.

Since $h_{x,y}$ is constant on Γ (see also Theorem 24 of [9]) and $h_{x,y}$ is smooth near Γ, the tangent space to Γ lies in the kernel of $\nabla^2 h_{x,y}$. Since this tangent space has dimension r, we see that this tangent space is in fact equal to the kernel of $\nabla^2 h_{x,y}$ (at every point of Γ). In particular, $\nabla^2 h_{x,y}$ is non-degenerate on the $(n-r)$-dimensional normal space to Γ. So we are in the situation discussed above, and we can define $\nu_\Gamma^{\mu,h_{x,y}}$.

Continuing, by general facts about compact submanifolds (like the tubular neighborhood theorem), it follows that we can find a sufficiently small $N(\Gamma)$ so that Theorem 15 holds for it and such that $N(\Gamma)$ is the union of a finite number of open sets $V_1, \ldots, V_k$ with the following properties:

- There are coordinates $(u_1^i, \ldots, u_n^i)$ on each V_i which form a set of coordinate charts for $N(\Gamma)$, and these coordinates can be extended to a neighborhood of $\overline{V_i}$.
- If $U_i = V_i \cap \Gamma$, then each U_i is nonempty and corresponds to the subset $\{u_{r+1}^i = \cdots = u_n^i = 0\}$ of V_i. In particular, $(u_1^i, \ldots, u_n^i)$ are local coordinates that are adapted to Γ, as above.
- Each V_i can be written, in the coordinates $(u_1^i, \ldots, u_n^i)$, as the product $U \times (-\varepsilon, \varepsilon)^{n-r}$ for some $\varepsilon > 0$.
- There exists a partition of unity $\varphi_1, \ldots, \varphi_k$ for Γ subordinate to the covering $U_1, \ldots, U_k$.

In the computations that follow, we drop the superscript i from the coordinates $(u_1^i, \ldots, u_n^i) = (u_1, \ldots, u_n)$ on each V_i, in order to lighten the notation. We can extend each φ to be a function on V_i by letting them be constant on the normal fibers (which are given by fixing $u_1, \ldots, u_r$); then $\sum_{i=1}^k \varphi_i = 1$ on all of $N(\Gamma)$. Now we can apply Theorem 15 and write the integral over $N(\Gamma)$ in terms of this system of coordinates charts to find

(5)
$$
\begin{aligned}
p_t(x,y) = \left(\frac{2}{t}\right)^n \sum_{i=1}^k \int_{U_i} &\left\{ \varphi_i(u_1, \ldots, u_r) \right. \\
&\times \int_{(-\varepsilon,\varepsilon)^{n-r}} e^{-h_{x,y}(u_1,\ldots,u_n)/t} \big(c_o(x, u_1, \ldots, u_n) c_0(u_1, \ldots, u_n, y) + O(t) \big) \\
&\left. \times F_i(u_1, \ldots, u_n) du_{r+1} \cdots du_n \right\} du_1 \cdots du_r
\end{aligned}
$$

where the $O(t)$ terms are uniform and F_i is the density of μ with respect to $du_1 \cdots du_n$. The asymptotic behavior of the inner integrals can be computed from the standard Laplace asymptotic formula (see page 198 of [16]), since the Hessian of $h_{x,y}$ in the normal directions is not degenerate, and this behavior is uniform by the smoothness and boundedness of everything involved. This gives

$$
\begin{aligned}
p_t(x,y) = \left(\frac{2}{t}\right)^n \sum_{i=1}^k \int_{U_i} &\varphi_i(u_1, \ldots, u_r) e^{-d^2(x,y)/4t} \left(2\pi t\right)^{(n-r)/2} \\
&\times (c_0|_\Gamma(x, u_1, \ldots, u_r) c_0|_\Gamma(u_1, \ldots, u_r, y) + O(t)) \nu_\Gamma^{\mu,h_{x,y}} (du_1 \cdots du_r).
\end{aligned}
$$

Using that the φ_i are a partition of unity and that the $O(t)$ term is uniform and thus its integral is also $O(t)$, we re-write this as

$$
p_t(x,y) = \frac{2^{(3n-r)/2} \pi^{(n-r)/2}}{t^{(n+r)/2}} e^{-d^2(x,y)/4t} \left[\int_{z\in\Gamma} c_0(x,z) c_0(z,y) \nu_\Gamma^{\mu,h_{x,y}} (dz) + O(t) \right].
$$

It only remains to see that the integral over Γ is positive (so that the C in the theorem is positive), but this follows immediately from the positivity of c_0 and the fact that $\nu_\Gamma^{\mu,h_{x,y}}$ clearly gives positive measure to every non-empty open (Borel) subset of Γ. $\qquad\square$

Remark 17. The statement of the theorem makes it clear that if one is only interested in the power of t appearing in the asymptotics of $p_t(x,y)$, which is the quantity that is most related to the

(conjugacy) structure of minimal geodesics from x to y, then only the properties (i) and (ii) of the exponential map are needed. It is only if one wants to determine the constant C that any specific control of μ, $c_0(\cdot, \cdot)$, or the Hessian of $h_{x,y}$ on the normal bundle are needed.

Remark 18. This theorem covers the most natural case, in which the dimension of $\mathcal{O}$ is equal to the number of independent directions in which the optimal geodesics are conjugate, i.e. to $\dim \ker D_{\bar{p}, d(x,y)} \exp_x$. Not only does this seem most common in situations of interest, but also there is a clear relationship between the geodesic structure and the behavior of $h_{x,y}$. In particular, the non-conjugacy in the normal directions, to $\mathcal{O}$ or equivalently Γ (in what follows we freely use the fact that the exponential map gives a diffeomorphism between a neighborhood of $\mathcal{O}$ and a neighborhood of Γ), corresponds exactly to the non-degeneracy of the Hessian of $h_{x,y}$ in those directions. More concretely, the non-degeneracy of the Hessian along the normal fibers means that we can find a further change of coordinates $(u_{r+1}, \ldots, u_n) \mapsto (\tilde{u}_{r+1}, \ldots, \tilde{u}_n)$ on each fiber such that $h_{x,y}$, restricted to the fiber, is given by the sum of squares $\tilde{u}_{r+1}^2 + \ldots + \tilde{u}_n^2$. Indeed, this normal form for $h_{x,y}$ is what underlies the asymptotic expansion of the "inner integrals" in (5).

This indicates that the above computation of the asymptotic expansion can be extended to other normal forms for $h_{x,y}$ along the normal fibers. To briefly illustrate this, suppose that we drop condition (ii) in Theorem 16 and instead assume that near each $z \in \Gamma$ we can find local coordinates such that $h_{x,y} = u_{r+1}^2 + \ldots + u_{n-1}^2 + u_n^4$. This implies that $\dim \ker D_{\bar{p}, d(x,y)} \exp_x = r + 1$ for $\bar{p} \in \mathcal{O}$. Moreover, as discussed in [7] (see especially Lemmas 27 and 29), if the exponential map restricted to a normal fiber has a singularity of type A_3 in the Arnol'd classification, then $h_{x,y}$ will have this normal form when restricted to the same normal fiber. So there is again a connection between the geodesic geometry and the behavior of $h_{x,y}$, even if it is not as clean. Nonetheless, once we assume this normal form for $h_{x,y}$, the argument proceeds in precisely the same way, except that the "inner integrals" in (5) now have asymptotics determined by this different local expression for $h_{x,y}$. Such an expansion is again well known, and completing the computation, one finds that

$$p_t(x, y) = \frac{C + O\left(\sqrt{t}\right)}{t^{\frac{n+r+1}{2} - \frac{1}{4}}} e^{-d^2(x,y)/4t}, \qquad \text{as } t \to 0.$$

Here C is again some positive constant, the precise value of which is given by integrating some local data (depending on c_0, the volume measure and behavior of $h_{x,y}$) over Γ.

4. The bi-Heisenberg group

Recall that the bi-Heisenberg group is the sub-Riemannian structure on $\mathbb{R}^5$ (where we consider coordinates (x_1, y_1, x_2, y_2, z)) where the following vector fields

$$\begin{cases} X_1 = \partial_{x_1} - \dfrac{\alpha_1}{2} y_1 \partial_z, & Y_1 = \partial_{y_1} + \dfrac{\alpha_1}{2} x_1 \partial_z, \\ X_2 = \partial_{x_2} - \dfrac{\alpha_2}{2} y_2 \partial_z, & Y_2 = \partial_{y_2} + \dfrac{\alpha_2}{2} x_2 \partial_z. \end{cases}, \qquad \alpha_1, \alpha_2 \geq 0, \ (\alpha_1, \alpha_2) \neq (0, 0),$$

define an orthonormal frame. If we set $Z = \partial_z$, it is easy to see that the only non trivial commutation relations are

$$[X_1, Y_1] = \alpha_1 Z, \qquad [X_2, Y_2] = \alpha_2 Z,$$

hence the structure is bracket generating. Notice that it is a nilpotent left-invariant structure (indeed it is a Carnot group). Moreover the structure is contact if and only if $\alpha_1, \alpha_2 > 0$. When one of the two parameters is zero the sub-Riemannian structure is the product $\mathbb{R}^2 \times \mathbb{H}$ where $\mathbb{H}$ is the 3-dimensional Heisenberg group.

4.1. Exponential map and synthesis.

In this section we compute explicitly the cut locus starting from the origin in the bi-Heisenberg group. In this section we assume $0 \leq \alpha_2 \leq \alpha_1$, treating the case $\alpha_2 = 0$ separately. The case $0 \leq \alpha_1 \leq \alpha_2$ can be obtained by exchanging the role of the indexes 1 and 2.

4.1.1. Contact case.

In the contact case (i.e., $\alpha_2 > 0$) there are no non-constant abnormal extremals. One then reduces the computations of the extremals (see Theorem 7) to the solution of the Hamiltonian system defined by the Hamiltonian

$$H(p,q) = \frac{1}{2} \sum_{i=1}^{2} \langle p, X_i(q) \rangle^2 + \langle p, Y_i(q) \rangle^2.$$

The arclength geodesics starting from the origin are parametrized by the initial covector $p_0 \in T_0^* M$ belonging to the level set $\Lambda_0 := H^{-1}(1/2) \cap T_0^* M$. If $(p_{x_1}, p_{y_1}, p_{x_2}, p_{y_2}, w)$ are the dual variables to (x_1, y_1, x_2, y_2, z) then

$$\Lambda_0 = \{ (p_{x_1}, p_{y_1}, p_{x_2}, p_{y_2}, w) \,|\, p_{x_1}^2 + p_{y_1}^2 + p_{x_2}^2 + p_{y_2}^2 = 1 \} \simeq S^3 \times \mathbb{R}$$

Performing the change of variable

$$p_{x_i} = -r_i \sin \theta_i, \qquad p_{y_i} = r_i \cos \theta_i, \qquad i = 1, 2,$$

we parametrize the set Λ_0 with $(r_1, r_2, \theta_1, \theta_2, w)$ such that $r_1, r_2 \geq 0, r_1^2 + r_2^2 = 1$, $\theta_1, \theta_2 \in S^1$, $w \in \mathbb{R}$. Solving the Hamiltonian system defined by H one can show that the arclength geodesic $\gamma(t) = (x_1(t), x_2(t), y_1(t), y_2(t), z(t))$ associated with an initial covector $p_0 = (r_1, r_2, \theta_1, \theta_2, w) \in \Lambda_0$ and $|w| \neq 0$, is described by the equations (we restrict to the case $w > 0$, the case $w < 0$ is analogous by symmetry)

$$
\begin{aligned}
x_i(t) &= \frac{r_i}{\alpha_i w} (\cos(\alpha_i w t + \theta_i) - \cos \theta_i), \\
(6) \qquad y_i(t) &= \frac{r_i}{\alpha_i w} (\sin(\alpha_i w t + \theta_i) - \sin \theta_i), \qquad i = 1, 2, \\
z(t) &= \frac{1}{2w^2} \left(wt - \sum_{i=1}^{2} \frac{r_i^2}{\alpha_i} \sin \alpha_i w t \right).
\end{aligned}
$$

If $w = 0$, geodesics are straight lines contained in the hyperplane $\{z = 0\}$

$$
\begin{aligned}
x_i(t) &= r_i t \cos \theta_i \\
y_i(t) &= r_i t \sin \theta_i, \\
z(t) &= 0.
\end{aligned}
$$

From equations (6) one easily shows that the projection of a non-horizontal extremal on every 2-plane (x_i, y_i) is a circle, with period T_i, radius ρ_i and center C_i defined by

$$T_i = \frac{2\pi}{\alpha_i w}, \qquad \rho_i = \frac{r_i}{\alpha_i w} \qquad C_i = -\frac{r_i}{\alpha_i w} (\cos \theta_i, \sin \theta_i), \qquad \forall\, i = 1, \ldots, \ell.$$

Moreover, generalizing the analogous property of the 3D Heisenberg group $\mathbb{H}$, one can recover that the z-component of the extremal at time t is the weighted sum (with coefficients α_i) of the areas spanned by the vectors $(x_i(t), y_i(t))$ in $\mathbb{R}^2$ (see also Figure 1).

We introduce the functions $\rho_i(t) = \sqrt{x_i(t)^2 + y_i(t)^2}$ that satisfy

$$(7) \qquad \rho_i(t) = \frac{2r_i}{\alpha_i |w|} \sin\left(\frac{\alpha_i w t}{2} \right) = r_i t \sin_c \left(\frac{\alpha_i w t}{2} \right), \qquad \text{where} \quad \sin_c(x) = \frac{\sin x}{x}.$$
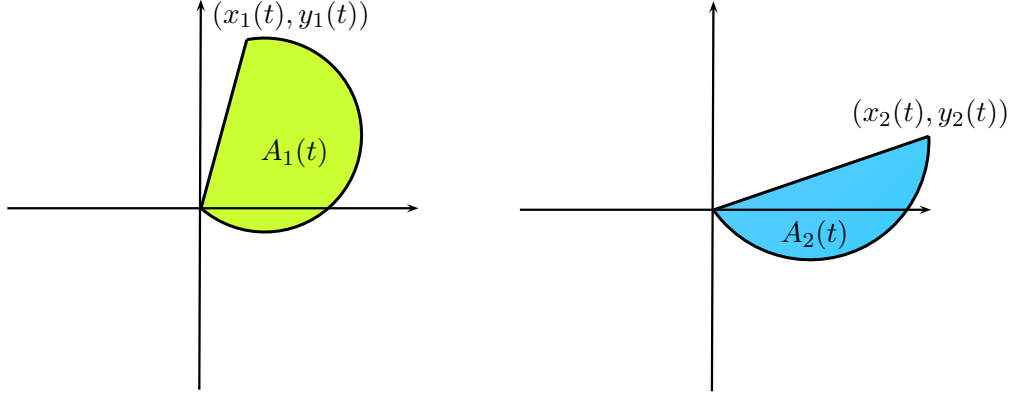
$$z(t) = \alpha_1 \mathrm{Area}(A_1(t)) + \alpha_2 \mathrm{Area}(A_2(t))$$

FIGURE 1. Projection of a non-horizontal geodesic: case $0 < \alpha_1 \leq \alpha_2$

4.1.2. *Non-contact case.* In the case when $\alpha_2 = 0$ the sub-Riemannian structure reduces to the product $\mathbb{H} \times \mathbb{R}^2$. The level set of the Hamiltonian in this case is again

$$\Lambda_0 = \{(p_{x_1}, p_{y_1}, p_{x_2}, p_{y_2}, w) \,|\, p_{x_1}^2 + p_{y_1}^2 + p_{x_2}^2 + p_{y_2}^2 = 1\} \simeq S^3 \times \mathbb{R}$$

Performing the change of variable

$$p_{x_i} = -r_i \sin \theta_i, \qquad p_{y_i} = r_i \cos \theta_i, \qquad i = 1, 2,$$

we parametrize the set Λ_0 with $(r_1, r_2, \theta_1, \theta_2, w)$ such that $r_1, r_2 \geq 0, r_1^2 + r_2^2 = 1$, $\theta_1, \theta_2 \in S^1$, $w \in \mathbb{R}$.

Again, solving the Hamiltonian system defined by H, one shows that the arclength geodesic $\gamma(t) = (x_1(t), x_2(t), y_1(t), y_2(t), z(t))$ associated with an initial covector $p_0 = (r_1, r_2, \theta_1, \theta_2, w) \in \Lambda_0$ and $|w| \neq 0$, are described as follows (we restrict to the case $w > 0$, the case $w < 0$ is analogous by symmetry): when $r_1 \neq 0$

$$x_1(t) = \frac{r_1}{\alpha_1 w}(\cos(\alpha_1 wt + \theta_1) - \cos \theta_1),$$

$$y_1(t) = \frac{r_1}{\alpha_1 w}(\sin(\alpha_1 wt + \theta_1) - \sin \theta_1),$$

$$x_2(t) = r_2 t \cos \theta_2, \quad y_2(t) = r_2 t \sin \theta_2,$$

$$z(t) = \frac{r_1^2}{2w^2}\left(wt - \frac{1}{\alpha_1}\sin(\alpha_1 wt)\right).$$

If $r_1 = 0$ (hence $r_2 = 1$) we have

$$x_1(t) = 0, \quad y_1(t) = 0,$$

$$x_2(t) = t \cos \theta_2, \quad y_2(t) = t \sin \theta_2,$$

$$z(t) = 0.$$

One can show that the trajectories corresponding to the case to $r_1 = 0$ are also abnormal. Finally, if $w = 0$, we have

$$x_1(t) = r_1 t \cos \theta_1, \quad y_1(t) = r_1 t \sin \theta_1,$$

$$x_2(t) = r_2 t \cos \theta_2, \quad y_2(t) = r_2 t \sin \theta_2,$$

$$z(t) = 0.$$

Let us denote the exponential map starting from the origin as the map

$$\exp_0 : \Lambda_0 \times \mathbb{R}^+ \to M, \qquad \exp_0(p_0, t) = \gamma(t),$$

where $\gamma(t) = (x_1(t), x_2(t), y_1(t), y_2(t), z(t))$ is the arclength geodesic associated with p_0.

The following lemma is proved in [2], see also [11].

Lemma 19. *An arclength geodesic* $\gamma(t) = \exp_0(p_0, t)$ *associated to* $p_0 = (r_1, r_2, \theta_1, \theta_2, w) \in \Lambda_0$ *is optimal up to its cut time*

$$t_{\mathrm{cut}}(\gamma) = \frac{2\pi}{|w| \max\{\alpha_1, \alpha_2\}}.$$

with the understanding $t_{\mathrm{cut}}(\gamma) = +\infty$ *when* $w = 0$. *Moreover the cut time coincides with the first conjugate time.*

Let us mention the following explicit formula for the distance from the origin to the vertical axis in terms of the parameters α_1, α_2, namely for every $\zeta = (0, 0, 0, 0, z)$ we have

$$(8) \qquad d(0, \zeta)^2 = \frac{4\pi |z|}{\max\{\alpha_1, \alpha_2\}}.$$

4.2. **Proof of Theorem 1.** By symmetry we will consider the case $w > 0$ and the cut locus will be the union of this set and its symmetric with respect to the hyperplane $\{z = 0\}$. (Recall the case $w = 0$ there is no cut locus along the geodesic.)

(i). In this case let $t_* = 2\pi/\alpha w$ be the cut time, where $\alpha := \alpha_1 = \alpha_2$. Substituting the cut time into the equations of the geodesic one gets that all the horizontal coordinates vanishes and $z(t_*) = \pi/\alpha w^2$. From this (i) is immediate. Notice moreover that the final point does not depend on $(r_1, r_2, \theta_1, \theta_2) \in S^3$, so we have a three-parameter family of optimal geodesics joining this point.

(ii). Since $\max\{\alpha_1, \alpha_2\} = \alpha_1$, the cut time is $t_* = 2\pi/\alpha_1 w$ and substituting the cut time into the equations we get

$$x_1(t_*) = y_1(t_*) = 0,$$

$$x_2(t_*) = \frac{r_2}{\alpha_2 w} \left(\cos\left(2\pi \frac{\alpha_2}{\alpha_1} + \theta_2 \right) - \cos\theta_2 \right),$$

$$y_2(t_*) = \frac{r_2}{\alpha_2 w} \left(\sin\left(2\pi \frac{\alpha_2}{\alpha_1} + \theta_2 \right) - \sin\theta_2 \right),$$

$$z(t_*) = \frac{1}{2w^2} \left(\frac{2\pi}{\alpha_1} - \frac{r_2^2}{\alpha_2} \sin\left(2\pi \frac{\alpha_2}{\alpha_1} \right) \right).$$

Notice that the set of cut points has rotational symmetry. Indeed the function ρ_2 describing the distance from the origin in the plane (x_2, y_2) (see (7))

$$\rho_2^2 = x_2^2 + y_2^2 = \frac{4r^2}{\alpha_2^2 w^2} \sin^2\left(\pi \frac{\alpha_2}{\alpha_1} \right)$$

does not depend on $\theta_2 \in S^1$. In other words the cut locus is a subset of the three-dimensional space $\{(x_2, y_2, z)\}$ which is symmetric with respect to the origin. This set is obtained by rotating with respect to the z axis the image of the map in the half-space $\{(\rho, z), \rho \geq 0\}$ (for simplicity we denote in the following by ρ, r respectively ρ_2 and r_2)

$$(9) \qquad (r, w) \mapsto \begin{cases} \rho = \dfrac{2r}{\alpha_2 w} \sin\left(\pi \dfrac{\alpha_2}{\alpha_1} \right) \\ z = \dfrac{1}{2w^2} \left(\dfrac{2\pi}{\alpha_1} - \dfrac{r^2}{\alpha_2} \sin\left(2\pi \dfrac{\alpha_2}{\alpha_1} \right) \right) \end{cases}, \qquad r \in [0, 1], w > 0.$$

13

From the equations above it is easy to get the relation

$$z = \Psi(\alpha_1, \alpha_2, r)\rho^2$$

where $\Psi(\alpha_1, \alpha_2, r)$ is a constant that depends only on r (and the parameters α_1, α_2) expressed by

$$\Psi(\alpha_1, \alpha_2, r) = \frac{\alpha_2^2}{8r^2} \sin^{-2}\left(\pi\frac{\alpha_1}{\alpha_2}\right)\left(\frac{2\pi}{\alpha_1} - \frac{r^2}{\alpha_2}\sin\left(2\pi\frac{\alpha_2}{\alpha_1}\right)\right).$$

In other words, for a fixed value of r, the point belongs to a parabola and (9) guarantees there exists a unique w associated to each point on this parabola. Moreover we have that

$$\lim_{r\to 0^+} \Psi(\alpha_1, \alpha_2, r) = +\infty, \qquad \lim_{r\to 1^-} \Psi(\alpha_1, \alpha_2, r) =: K(\alpha_1, \alpha_2),$$

and

$$\frac{\partial}{\partial r}\Psi(\alpha_1, \alpha_2, r) = -\frac{\pi\alpha_2^2 \csc^2(\pi\frac{\alpha_1}{\alpha_2})}{8\alpha_1 r^3} < 0, \qquad r \in [0, 1],$$

that ensure that the function (9) is injective on its image, which completes the proof of (ii).

(iii). This case is reduced to $\mathbb{R}^2$ times the cut locus of $\mathbb{H}$, that is the three dimensional space $\{(x_2, y_2, z), z \neq 0\}$.

Remark 20. Notice that the structure of the cut locus is continuous with respect to the parameters α_1, α_2 in the following sense

$$\lim_{\alpha_2 \to 0^+} \Psi(\alpha_1, \alpha_2, r) = 0, \qquad \lim_{\alpha_2 \to \alpha_1^-} \Psi(\alpha_1, \alpha_2, r) = +\infty.$$

This implies in fact that the cases (i) and (iii) can be recovered as the limit cases of (ii).

4.3. **Proof of Theorem 2.** Point (i) and (ii) of Theorem 2 follows from results of Ben Arous [12, 13].

Point (iii) of Theorem 2 can be obtained by applying Theorem 16 to the explicit structure of optimal geodesics described in the previous section. We have two different cases:

(iii.a). $0 < \alpha_2 = \alpha_1 =: \alpha$. In this case we proved that every point in the cut locus is reached by a three-parameter family of geodesics. More precisely, following the notation of Theorem 16, for every point $q \in \mathrm{Cut}_0 \setminus\{0\}$, we have $\mathcal{O} = S^3$. Moreover it is easy to see that for every $\bar{p} \in \mathcal{O}$ one has $D_{(\bar{p},d(0,q))}\exp_0$ has rank 2. Indeed the differential of the exponential map is never degenerate with respect to t and, for $t = d(0, q)$ fixed, one has (notice that $\bar{p} \in \mathcal{O}$ implies $w \neq 0$)

$$\exp_0(\bar{p}, t) = (0, 0, 0, 0, \pi/(\alpha w^2)).$$

Hence, applying Theorem 16 with $n = 5$ and $r = 3$, we have

$$p_t(0, q) = \frac{C + O(t)}{t^4}e^{-\frac{d^2(0,q)}{4t}}$$

(iii.b). $0 \leq \alpha_2 < \alpha_1$. Let us first consider the subcase $\alpha_2 > 0$. Then every point in the cut locus is reached by a one-parameter family of geodesics. More precisely, following the notation of Theorem 16, for every point $q \in \mathrm{Cut}_0 \setminus\{0\}$, we have $\mathcal{O} = S^1$. Moreover it is easy to see that for every $\bar{p} \in \mathcal{O}$ one has $D_{(\bar{p},d(0,q))}\exp_0$ has rank 4. Hence, applying Theorem 16 with $n = 5$ and $r = 1$, we have

$$p_t(0, q) = \frac{C + O(t)}{t^3}e^{-\frac{d^2(0,q)}{4t}}$$

When $\alpha_2 = 0$ the results of Section 3 cannot be applied due to the presence of abnormal minimizers. Nevertheless since the struture is the product of the Heisenberg group and $\mathbb{R}^2$ one easily gets that the heat kernel is the product of the heat kernel of the Heisenberg group and the heat kernel of $\mathbb{R}^2$. From the explicit expression one gets the result.

5. What one can get from the Greiner-Gaveau-Beals results

In this section we recall the expression of the heat kernel of the intrinsic sub-Laplacian associated with a 2-step nilpotent structure, that has been found in [11]. Then we rewrite it to have a convenient expression on the "vertical subspace".

Consider on $\mathbb{R}^n$ a 2-step nilpotent structure of rank $k < n$, where $X_1, \ldots, X_k$ is an orthonormal frame. Once a smooth complement $\mathcal{V}$ for the distribution is chosen (i.e. $T_q\mathbb{R}^n = \mathcal{D}_q \oplus \mathcal{V}_q$, for all $q \in \mathbb{R}^n$) we can complete an orthonormal frame to a global one $X_1, \ldots, X_k, Y_1, \ldots, Y_m$, where $m = n - k$ and $\mathcal{V}_q = \operatorname{span}_q\{Y_1, \ldots, Y_m\}$. Since the structure is nilpotent, we can assume that the only nontrivial commutation relations are

$$[X_i, X_j] = \sum_{h=1}^m b_{ij}^h Y_h,$$

where $B_1, \ldots, B_m$ defined by $B_h = (b_{ij}^h)$ are skew-symmetric matrices (see [8] for the role of these matrices in the exponential map).

Due to the group structure, the intrinsic sub-Laplacian takes the form of sum of squares $\Delta = \sum_{i=1}^k X_i^2$ (see Remark 14). The group structure also implies that the heat kernel is invariant with respect to the group operation, hence it is enough to consider the heat kernel $p_t(0, q)$ starting from the identity of the group, which we also denote $p_t(q)$. The heat kernel is written as follows (see again [11, 14])

$$p_t(q) = \frac{2}{(4\pi t)^{Q/2}} \int_{\mathbb{R}^m} V(B(\tau)) \exp\left(-\frac{W(B(\tau))x \cdot x}{4t}\right) \cos\left(\frac{z \cdot \tau}{t}\right) d\tau,$$

where $q = (x, z)$, $x \in \mathbb{R}^k, z \in \mathbb{R}^m$, and $B(\tau) := \sum_{i=1}^m \tau_i B_i$. Moreover $V : \mathbb{R}^{n \times n} \to \mathbb{C}$ and $W : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ are the matrix functions defined by

$$V(A) = \sqrt{\det\left(\frac{A}{\sin A}\right)}, \qquad W(A) = \frac{A}{\tan A}.$$

Here Q is the Hausdorff dimension of the sub-Riemannian structure.

Notice that (5) differs by some constant factors from the formulas contained in [11] since there the heat kernel is the solution of the equation $\partial_t u = \frac{1}{2}\Delta u$.

Remark 21. Assume that the real skew-symmetric matrix $B(\tau)$ is diagonalizable and denote by $\pm i\lambda_j(\tau)$, for $j = 1, \ldots \ell$, its non-zero eigenvalues. Then we have the formula for the expansion on the "vertical subspace" (i.e. where $x = 0$)

$$(10) \qquad p_t((0, z)) = \frac{2}{(4\pi t)^{Q/2}} \int_{\mathbb{R}^m} \prod_{j=1}^\ell \frac{\lambda_j(\tau)}{\sinh \lambda_j(\tau)} \cos\left(\frac{z \cdot \tau}{t}\right) d\tau.$$

For the bi-Heisenberg case, in which $m = 1$, we have that $B(\tau) = \tau B$ and $\operatorname{eig}(B(\tau)) = \{\pm i\alpha_1\tau, \pm i\alpha_2\tau\}$, and from (5) one gets

$$p_t(q) = \frac{2}{(4\pi t)^3} \int_{-\infty}^{\infty} \prod_{i=1}^2 \frac{\alpha_i\tau}{\sinh(\alpha_i\tau)} \exp\left(-\sum_{i=1}^2 \frac{x_i^2 + y_i^2}{4t} \frac{\alpha_i\tau}{\tanh(\alpha_i\tau)}\right) \cos\left(\frac{z\tau}{t}\right) d\tau,$$

where $p_t(q) := p_t(0, q)$ and $q = (x_1, x_2, y_1, y_2, z)$.

From this formula one can get the expansion of the heat kernel on the z-axis (that always lies in the cut locus by Theorem 1). We rewrite the expansion when $x_i = y_i = 0$ for $i = 1, 2$, i.e. at a

point $\zeta = (0, 0, 0, 0, z)$ (see also (10)), as

$$(11) \qquad p_t(\zeta) = \frac{2}{(4\pi t)^3} \int_{-\infty}^{\infty} \frac{\alpha_1 \tau}{\sinh(\alpha_1 \tau)} \frac{\alpha_2 \tau}{\sinh(\alpha_2 \tau)} \cos\left(\frac{z\tau}{t}\right) d\tau.$$

We show now that the different behavior of the asymptotics that we proved in the previous section can be recovered in the two cases when the parameters are equal or not.

(a) *Case $\alpha_1 = \alpha_2$.* It is enough to consider the case when $\alpha_1 = \alpha_2 = 1$.

In this case the integral (11) is explicitly computed by the formula

$$(12) \qquad p_t(\zeta) = \frac{2}{(4\pi t)^3} \int_{-\infty}^{\infty} \frac{\tau^2}{\sinh^2 \tau} \cos\left(\frac{z\tau}{t}\right) d\tau = \frac{\pi z \coth\left(\frac{\pi z}{2t}\right) - 2t}{32\pi t^4 \left(\cosh\left(\frac{\pi z}{t}\right) - 1\right)}.$$

Let us consider now the point ζ corresponding to $z = 1$. From (8) and (12) one gets

$$(13) \qquad p_t(\zeta) = \frac{1}{t^4} \exp\left(-\frac{d^2(0, \zeta)}{4t}\right) \varphi(t),$$

with φ a smooth function of t such that $\varphi(0) = 1/16$.

(b) *Case $\alpha_1 \neq \alpha_2$.* To simplify the expression of the heat kernel we consider the particular case when $\alpha_1 = 1, \alpha_2 = 1/2$. Also in this case we have the expression of the heat kernel on the vertical axis $\zeta = (0, 0, 0, 0, z)$ (again $z \neq 0$)

$$(14) \qquad p_t(\zeta) = \frac{\left(-8\pi z \sinh\left(\frac{\pi z}{t}\right) + 8t \cosh\left(\frac{\pi z}{t}\right) + \pi t \left(\cosh\left(\frac{2\pi z}{t}\right) - 3\right)\right)}{128\pi t^4 \cosh^3\left(\frac{\pi z}{t}\right)}.$$

If we consider the expansion at the point ζ corresponding to $z = 1$, using (8) and (14) we find

$$p_t(\zeta) = \frac{1}{t^3} \exp\left(-\frac{d^2(0, \zeta)}{4t}\right) \varphi(t),$$

with $\varphi(0) = 1/32$.

By homogeneity these expansion hold on any point of the z-axis.

Remark 22. As a byproduct of formula (13), we get that, in the symmetric case the maximal degeneration, at least in terms of the factor in front of the exponential, is not obtained on the diagonal. This is in contrast to what happens in the Heisenberg case.

REFERENCES

[1] A. Agrachev. Compactness for sub-Riemannian length-minimizers and subanalyticity. *Rend. Sem. Mat. Univ. Politec. Torino*, 56(4):1–12 (2001), 1998. Control theory and its applications (Grado, 1998).

[2] A. Agrachev, D. Barilari, and U. Boscain. On the Hausdorff volume in sub-Riemannian geometry. *Calc. Var. Partial Differential Equations*, 43(3-4):355–388, 2012.

[3] A. Agrachev, D. Barilari, and U. Boscain. Introduction to Riemannian and sub-Riemannian geometry (Lecture Notes), https://webusers.imj-prg.fr/~ davide.barilari/notes.php. Version April 17, 2016.

[4] A. Agrachev, U. Boscain, J.-P. Gauthier, and F. Rossi. The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups. *J. Funct. Anal.*, 256(8):2621–2655, 2009.

[5] A. A. Agrachev. Exponential mappings for contact sub-Riemannian structures. *J. Dynam. Control Systems*, 2(3):321–358, 1996.

[6] A. A. Agrachev and Y. L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.

[7] D. Barilari, U. Boscain, G. Charlot, and R. W. Neel. On the heat diffusion for generic riemannian and sub-riemannian structures. *to appear in Int. Math. Res. Not.*

[8] D. Barilari, U. Boscain, and J.-P. Gauthier. On 2-step, corank 2 sub-Riemannian metrics. *SIAM Journal of Control and Optimization*, 50(1):559–582, 2012.

[9] D. Barilari, U. Boscain, and R. W. Neel. Small-time heat kernel asymptotics at the sub-Riemannian cut locus. *J. Differential Geom.*, 92(3):373–416, 2012.

[10] D. Barilari and L. Rizzi. A formula for Popp's volume in sub-Riemannian geometry. *Anal. Geom. Metr. Spaces*, 1:42–57, 2013.

[11] R. Beals, B. Gaveau, and P. Greiner. The Green function of model step two hypoelliptic operators and the analysis of certain tangential Cauchy Riemann complexes. *Adv. Math.*, 121(2):288–345, 1996.

[12] G. Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus. *Ann. Sci. École Norm. Sup. (4)*, 21(3):307–331, 1988.

[13] G. Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique sur la diagonale. *Ann. Inst. Fourier (Grenoble)*, 39(1):73–99, 1989.

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

[15] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[16] R. Estrada and R. P. Kanwal. *A distributional approach to asymptotics*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston Inc., Boston, MA, second edition, 2002. Theory and applications.

[17] Y. Inahama and S. Taniguchi. Short time full asymptotic expansion of hypoelliptic heat kernel at the cut locus. *arXiv:1603.01386 [math.PR]*.

[18] S. A. Molčanov. Diffusion processes, and Riemannian geometry. *Uspehi Mat. Nauk*, 30(1(181)):3–59, 1975.

[19] R. S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.