

UNIFORM, LOCALIZED ASYMPTOTICS FOR SUB-RIEMANNIAN HEAT KERNELS AND DIFFUSIONS

ROBERT W. NEEL AND LUDOVIC SACCHELLI

ABSTRACT. We show how Molchanov's method provides a systematic approach to determining the small-time asymptotics of the heat kernel on a sub-Riemannian manifold away from any abnormal minimizers. The expansion is closely connected to the structure of the minimizing geodesics between two points. If the normal form of the exponential map at the minimal geodesics between two points is sufficiently explicit, a complete asymptotic expansion of the heat kernel can, in principle, be given. (But one can also exhibit metrics for which the exponential map is quite degenerate, so that even giving the leading term of the expansion seems doubtful.) In a different direction, we have uniform bounds on the heat kernel and its derivatives over any compact set with no abnormals, in an inherently local way.

The method extends naturally to logarithmic derivatives of the heat kernel, which are closely related to the law of large numbers for the corresponding bridge process. This allows for a general treatment of the small-time behavior of the bridge process on the cut locus, as well as the determination of the limiting measure of the bridge process if the normal form of the exponential map at the minimal geodesics between two points is sufficiently explicit. Further, we give an expression for the leading term of the logarithmic derivative of the heat kernel, of any order, as a cumulant of geometrically natural random variables and a characterization of the cut locus in terms of the blow-up of the logarithmic second derivative of the heat kernel. This method also provides uniform, local estimates of the logarithmic derivatives.

CONTENTS

1. Introduction	1
1.1. Sub-Laplacians and heat kernels	1
1.2. Asymptotics for the heat kernel	3
1.3. Asymptotics for log-derivatives and bridges	7
1.4. Acknowledgements	9
2. Localization and Léandre asymptotics	9
3. Ben Arous expansion theorem	18
3.1. Some properties of midpoint sets	18
3.2. Uniform Ben Arous expansions	20
3.3. Uniform universal bounds on the heat kernel	25
4. Complete expansions	28
4.1. A-type singularities	28
4.2. Morse-Bott case	30
4.3. Prescribed singularities	33
5. Logarithmic derivatives	42
5.1. Molchanov-type expansions of logarithmic derivatives	42
5.2. Characterizing the cut locus	47
5.3. Sheu-Hsu-Stroock-Turetsky type bounds	51
6. Law of large numbers	52
6.1. Extension to the cut locus	52
6.2. Real analytic methods	54

2010 *Mathematics Subject Classification.* Primary 58J65; Secondary 53C17 58J35 58K55.

Key words and phrases. sub-Riemannian, heat kernel, cut locus, small-time asymptotics, diffusion bridge measure.

6.3. LLN for A -type singularities	55
6.4. LLN for the Morse-Bott case	55
References	56

1. INTRODUCTION

Small-time heat kernel asymptotics, and a variety of related matters, have a long history and a substantial literature, as we (partially) outline below. The object of this paper is to give a systematic development on general sub-Riemannian manifolds of a method originally due to Molchanov [41], on Riemannian manifolds, to determine heat kernel asymptotics at points in the cut locus by “gluing together” the asymptotics at non-cut points, and to apply this method to a broad spectrum of asymptotic questions. Most of our results are new in the sub-Riemannian context, and a couple are also, to the best of our knowledge, new in the Riemannian case as well. In this section, we state many of our main results, to give an indication of their nature and range. However, a number of interesting results are left to the body of the paper, since including all of them here would be too unwieldy.

1.1. Sub-Laplacians and heat kernels. Let M be smooth (connected) manifold of dimension d , and let μ be a smooth volume on M . That is, μ is a measure on M such that, in any (smooth) local coordinates (u_1, \dots, u_d) on a coordinate patch U , $\mu|_U$ has smooth, non-vanishing density with respect to Lebesgue measure $du_1 \cdots du_d$ on U . The most efficient way to proceed is to introduce the sub-Laplacian and the sub-Riemannian metric together. We let Δ be a smooth, second-order differential operator on M such that any point is contained in a coordinate patch U on which

$$(1) \quad \Delta = \sum_{i=1}^k Z_i^2 + Z_0$$

where Z_0, Z_1, \dots, Z_k are smooth vector fields and Z_1, \dots, Z_k are bracket-generating (this is the strong Hörmander condition). In this situation, Δ induces a sub-Riemannian structure on M , which corresponds to the Carnot-Carathéodory distance in the older PDE literature. In the case where the distribution is of constant rank k on U , with $2 \leq k \leq d$, we can choose the Z_1, \dots, Z_k in Equation (1) to be an orthonormal basis for the distribution at each point of U . That is, the span of Z_1, \dots, Z_k gives the distribution at each point, and the orthonormality induces an inner product on the distribution. The formalism to accommodate the rank-varying case is more elaborate, and we refer to Chapter 3 of [2] for a rigorous treatment of the construction of a sub-Riemannian structure from the principal part of Δ . Nonetheless, for the purposes of this paper, the distribution and the inner product on it are not directly referenced; rather, the induced distance $d(\cdot, \cdot)$ and the structure of distance-minimizing curves are the central objects. In particular, the Chow-Rashevskii theorem shows that the distance between any two points of M is finite and (M, d) is a metric space such that the metric topology agrees with the topology of M . Moreover, one version of the Hopf-Rinow theorem for sub-Riemannian manifolds gives that if (M, d) is complete as a metric space, there exists a minimizing curve between any two points of M and that M is geodesically complete. (Again, we take [2] as the canonical reference for the basic results of sub-Riemannian geometry.)

We assume that M is complete (although our localization result allows us to treat some compact subsets of incomplete manifolds, as addressed below), and so in particular, M is a d -dimensional complete sub-Riemannian manifold equipped with a smooth volume and a sub-Laplacian, in the above sense. Turning back to the operator Δ , the corresponding diffusion X_t is given by the Stratonovich SDE

$$dX_t = \sqrt{2} \sum_{i=1}^k Z_i(X_t) \circ dW_t^i + Z_0(X_t) dt$$

where the W_t^i are independent, standard Brownian motions. Starting from $X_0 = x$, this diffusion has a transition density $p_t(x, y)$ with respect to μ , which is smooth for $(t, x, y) \in (0, \infty) \times M \times M$ by the

celebrated Hörmander theorem. (Note that we adopt the analyst's convention of using Δ rather than $(1/2)\Delta$ as the infinitesimal generator of X_t , and thus the SDE requires the $\sqrt{2}$ -factor. The difference simply amounts to rescaling t by a factor of 2.) We do not assume that $p_t(x, y)$ is symmetric, although that is an important special case, and we will refer to the situation when, in addition to the above, $p_t(x, y)$ is symmetric simply as the symmetric case. Note that this includes the most important type of sub-Laplacian in sub-Riemannian geometry, namely the case when Δ is given as the μ -divergence of the horizontal gradient (which generalizes the fact that on a Riemannian manifold, the Laplace-Beltrami operator can be written as the divergence of the gradient). The small-time asymptotics of $p_t(x, y)$ are a central topic for us.

We note that, in the Riemannian case, sub-Laplacians are exactly operators of the form $\Delta + Z_0$, where here Δ is the Laplace-Beltrami operator and Z_0 is a smooth vector field. Thus the heat kernels we consider in the Riemannian regime are more general than the standard heat kernel (corresponding to $Z_0 \equiv 0$). When giving results specifically for the Riemannian case, we will abuse the notation slightly and write the sub-Laplacian as $\Delta + Z_0$.

In sub-Riemannian geometry, extremal curves, by which we mean critical points of the distance function, can be normal or abnormal (or both). The Molchanov method is effective for pairs of points such that all minimizing geodesics are strongly normal (meaning no non-trivial subsegment is abnormal). In the properly sub-Riemannian situation, that is, not (locally) a Riemannian manifold, the trivial geodesic is always (non-strictly) abnormal, so the method does not apply on the diagonal. For broad classes of sub-Riemannian manifolds, such as contact manifolds, there are no non-trivial abnormals, in which case the diagonal is the only place where the method is ineffective. Of course, on-diagonal heat kernel asymptotics are a natural object of study (see [10], for a sub-Riemannian example), but this requires other approaches (frequently perturbation methods). Interpolating between the diagonal and off-diagonal asymptotics, say, to derive “good” uniform bounds on the heat kernel in small time, is in general a hard problem. For the case of the Heisenberg group and more general H-type groups, see [23] and [38]. In the Riemannian case, there are no abnormals, so the method applies everywhere, including the diagonal.

There are situations where, for more specialized sub-Riemannian structures, one can find expressions for the heat kernel that allow the small-time asymptotics to be extracted in an explicit way. For example, for left-invariant structures on Lie groups, generalized Fourier transforms can be used, as developed in [5]. The sub-Riemannian model spaces are especially well studied and have a large literature, but we mention [16] and [15] as two examples of explicit computation of the heat kernel and its small-time asymptotics on such spaces.

On the other hand, there are sub-Riemannian (and sub-Riemannian-adjacent) situations that go beyond this framework. For example, the Grushin plane is a sub-Riemannian structure, but the most natural Laplacian to put on it is only defined up to the singular set, and the first order term blows up as the singular set is approached. Thus this is not a smooth sub-Laplacian, and indeed, the Léandre asymptotics fail dramatically. For recent work in this direction on Grushin and related structures, see [21, 20, 26, 25].

1.2. Asymptotics for the heat kernel. In the first part of this paper, we rigorously establish the Molchanov method on general sub-Riemannian manifolds (not necessarily compact, and in principle not even complete, as explained below), show that it applies to derivatives of the heat kernel as well as the heat kernel itself, give the corresponding uniform bounds, and show that Molchanov's method supports complete asymptotic expansions for both the heat kernel and its derivatives. This direction itself requires several steps, and touches upon some related directions, which we now describe in more detail.

In what follows, (Z_1, \dots, Z_m) denote an arbitrary family of smooth vector fields on M such that at each point $x \in M$, $(Z_1(x), \dots, Z_m(x))$ spans the whole tangent space $T_x M$. We call multi-index any finite (and possibly empty) sequence of integers $\alpha \in \{1, \dots, m\}^k$, with $k \in \mathbb{N}$. Then for any smooth function $f : M \rightarrow \mathbb{R}$, we denote

$$Z^\alpha f = Z_{\alpha_k} \circ Z_{\alpha_{k-1}} \circ \dots \circ Z_{\alpha_1} f.$$

If $\alpha = \emptyset$, we intend that $Z^\alpha f = f$. For $g : M^2 \rightarrow \mathbb{R}$, $Z_x^\alpha g(x, y)$ and $Z_y^\alpha g(x, y)$ denote the derivatives with respect to the first and second space variable, respectively.

The Molchanov method has three ingredients. One is the Chapman-Kolmogorov equation (or the Markov property of the diffusion). The other two are a “coarse” estimate valid globally and a “fine” estimate valid away from the cut locus. In the sub-Riemannian case, the coarse estimate is essentially due to Léandre [36, 37]. For a sub-Riemannian structure on \mathbb{R}^d given by smooth, bounded vector fields with bounded derivatives of all orders, he proved that

$$\lim_{t \searrow 0} -4t \log p_t(x, y) = d^2(x, y)$$

$$\text{and } \limsup_{t \searrow 0} 4t \log \left(\left| \frac{\partial^\alpha}{\partial y^\alpha} p_t(x, y) \right| \right) \leq -d^2(x, y)$$

for any multi-index α , uniformly on compacts. (And note that these asymptotics holds without regard to abnormals or the cut locus.) The fine estimate is essentially due to Ben Arous [19]. For the same sub-Riemannian structures as Léandre, he proved that there are smooth functions $c_i(x, y)$ with $c_0(x, y) > 0$ such that, for any N ,

$$p_t(x, y) = t^{-d/2} e^{-\frac{d(x, y)^2}{4t}} \left(\sum_{k=0}^N c_k(x, y) t^k + t^{N+1} r_{N+1}(t, x, y) \right)$$

where r_{N+1} is an appropriate remainder term, uniformly on compact subsets of $M \times M$ that avoid the cut locus (and abnormals, including the diagonal). Further, this expansion can be differentiated as many times as desired in t , x , and y . Note that both Léandre and Ben Arous used the Euclidean volume to define their heat kernel, but it is an exercise in using the product rule to show that if either result holds for one smooth volume, then it holds for any smooth volume.

In reviewing the above results, not to mention those that follow, one might note that the first-order part (or sub-symbol) of the sub-Laplacian is relatively unimportant in the form of the expansion, as is the choice of sooth volume. Indeed, the distance function, and thus the minimal geodesics, cut locus, etc. depend solely on the principal symbol of the sub-Laplacian. The first-order term and the choice of volume only affect the constants c_k in the Ben Arous expansion, which are given by transport equations. This is unsurprising— if the first order part lies in the distribution, one can think of it as contributing a Girsanov factor, and more generally, its effect is negligible at distant points in small time. Similarly, changing the smooth volume multiplies p_t by a smooth, non-vanishing function. In the Riemannian case, where we write the operator as $\Delta + Z_0$ and use the Riemannian volume, the effect of Z_0 relative to the “standard” $Z_0 = 0$ case can be explicitly isolated as an action term, as can be found in the original paper of Molchanov [41] (and continuing into some of the other references mentioned). Here we follow Ben Arous and allow the c_i to account for matters.

Our first task is to establish Léandre asymptotics for p_t and its derivatives on a general sub-Riemannian manifold, along with natural localization results. These results go hand-in-hand. Indeed, the principle of “not feeling the boundary” was invoked in [41] (without proof). That the heat kernel on a general (complete) sub-Riemannian manifold satisfies the Léandre and Ben Arous asymptotics has been something of a folk theorem, alluded to in the literature used without elaboration in [13, 11, 12], for example. A general localization result was proven in [28], showing that any diffusion on a manifold satisfying Léandre asymptotics for p_t itself on compacts has the property that the asymptotics of p_t are local. A quick (one sentence) reference is made to Léandre’s result on sub-Riemannian manifolds, but one should not be too casual here. Proving that the Léandre asymptotics hold on a general manifold in the first place uses localization, so a careless approach ends up being circular. (The resolution is to build-up the result in stages, a version of which we carry out.) A similarly brief reference to localizing Léandre asymptotics (for p_t) on possibly incomplete sub-Riemannian manifolds is given in [32], which explicitly treats the Riemannian case. (Again, all the basic tools of the proof are present, such as exit time estimates for small balls, but no explicit argument is given. Also, the theorems are all stated for Riemannian manifolds, so the idea of extending to the sub-Riemannian case appears to have been somewhat buried. In fact, the idea of adapting the Riemannian arguments to the sub-Riemannian

case is already suggested by Azencott [6], but this preceded the work of Léandre.) Recently, Ballieul and Norris [9] gave a rigorous proof of the Léandre asymptotics for p_t itself on a possibly incomplete sub-Riemannian manifold. Their primary focus is working with incomplete manifolds, and especially establishing a best-possible localization result, in terms of the distance to infinity, under the additional assumption of a “sector condition.” For this reason, they employ considerable analytic machinery (such as volume doubling estimates, a local Poincaré inequality, a parabolic mean-value inequality, etc.), and it is not clear that these extend to derivatives of the heat kernel.

As indicated above, here we give an argument along the lines of [28], showing that the Léandre asymptotics are “self-localizing,” in the sense that establishing both localization and the Léandre asymptotics on a general sub-Riemannian manifold uses only the previous result of Léandre and basic stochastic considerations. Namely, we have

Theorem 1. *Let M be a complete sub-Riemannian manifold with a smooth volume μ and a smooth sub-Laplacian Δ , and let $p_t(x, y)$ be the corresponding heat kernel. Then for any compact $K \subset M$ and $\varepsilon > 0$, let K' be a compact such that $K \subset K'$ and $d(K, \partial K') > \frac{1}{2}\text{diam}(K) + \varepsilon$. If $p_t^{K'}$ is the heat kernel on K' with Dirichlet boundary conditions, then for any $x, y \in K$, and any multi-index α ,*

$$p_t(x, y) = p_t^{K'}(x, y) + O\left(\exp\left(-\frac{(\text{diam}(K) + \varepsilon)^2}{4t}\right)\right)$$

$$\text{and } \left|Z_y^\alpha p_t(x, y) - Z_y^\alpha p_t^{K'}(x, y)\right| = O\left(\exp\left(-\frac{(\text{diam}(K) + \varepsilon)^2}{4t}\right)\right),$$

where the “big- O ” terms are uniform in such x and y and the implicit constants can be taken to depend only on the restriction of the structure to a neighborhood of K' .

Also, for any $x, y \in K'$,

$$\limsup_{t \searrow 0} 4t \log p_t^{K'}(x, y) \leq -d^2(x, y)$$

uniformly in such x, y , and if $y \in K$, for any multi-index α ,

$$\limsup_{t \searrow 0} 4t \log \left(\left| Z_y^\alpha p_t^{K'}(x, y) \right| \right) \leq -d^2(x, y)$$

uniformly in such x, y .

Finally, for $x, y \in K$,

$$\lim_{t \searrow 0} 4t \log p_t^{K'}(x, y) = -d^2(x, y),$$

uniformly in x, y .

Note that, in contrast to the work just mentioned, we establish both localization and the Léandre asymptotics not only for p_t itself, but also for its derivatives (in y), thus extending Léandre’s original result fully. This is interesting in its own right, but moreover, having bounds on the derivatives of the heat kernel is needed to apply Molchanov’s method to heat kernel derivatives and also to study logarithmic derivatives of the heat kernel. While it is not our primary focus, localization implies that all of our results hold on an incomplete manifold under a standard condition to the distance to infinity (see Remark 13). Also, in light of the Ben Arous expansion one might wonder about taking derivatives in t and x as well. This is more complicated. Time derivatives are generally accessible by using the forward Kolmogorov (Fokker-Planck) equation to replace them with spatial derivatives. However, it turns out that, in the symmetric case, time derivatives can be controlled in a way that is compatible with the Molchanov method, and this allows for lower bounds on pure time derivatives, as we see in a moment. This is an interesting phenomenon for the most important special case, so we pursue it in what follows.

Thus, with Theorem 1 in hand, we similarly show that the Ben Arous asymptotics hold on a general sub-Riemannian manifold. Similar to the situation described above for the Léandre asymptotics, though less widely considered, this expansion, for the heat kernel itself, was assumed to generalize in earlier works, but here we make this rigorous. More precisely, we have the following.

Definition 2. A geodesic $\gamma : [0, T] \rightarrow M$ is said to be strongly normal if for every $[s, t] \in [0, T]$, $\gamma_{[s, t]}$ is not abnormal. Then the critical set \mathcal{C} in M^2 is the set of pairs of points (x, y) such that either

- There exists multiple length minimizing curves joining x and y .
- The unique geodesic joining x and y is conjugate.
- The unique geodesic joining x and y is not strongly normal. Crucially, if M is properly sub-Riemannian, this includes points on the diagonal, such that $x = y$, but not if M is Riemannian.

Theorem 3 (Uniform Ben Arous expansions). *For any multi-index α , and l a non-negative integer in the symmetric case and 0 otherwise, there exist sequences of smooth functions $c_k : M^2 \setminus \mathcal{C} \rightarrow \mathbb{R}$, $r_k : \mathbb{R}^+ \times M^2 \setminus \mathcal{C} \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, such that for all $n \in \mathbb{N}$, for all $(x, y) \in M^2 \setminus \mathcal{C}$, for all $t \in \mathbb{R}^+$*

$$\partial_t^l Z_y^\alpha p_t(x, y) = t^{-(|\alpha|+2l+d/2)} e^{-\frac{d(x, y)^2}{4t}} \left(\sum_{k=0}^N c_k(x, y) t^k + t^{n+1} r_{n+1}(t, x, y) \right).$$

For any compact $\mathcal{K} \subset M^2 \setminus \mathcal{C}$, l' which is any non-negative integer in the symmetric case and 0 otherwise, and any multi-index α' , there exists t_0 such that

$$\sup_{0 < t < t_0} \sup_{(x, y) \in \mathcal{K}} \left| \partial_t^{l'} Z_y^{\alpha'} r_{n+1}(t, x, y) \right| < \infty.$$

Additionally, if $\alpha = 0$, then $c_0(x, y) > 0$ on $M^2 \setminus \mathcal{C}$.

Note that this is not the complete generalization of the original Ben Arous expansion; here we don't allow derivatives the x or t variables in general, whereas that was allowed in [19]. Even in the symmetric case (where y -derivatives can also be replaced by x -derivatives by symmetry), we don't allow the mixing of x - and y -derivatives. Indeed, it isn't obvious whether or not one should expect such results without some global control of the geometry.

We next establish the general formula for the heat kernel asymptotics, valid at the non-abnormal cut locus, coming from Molchanov's method. The method is based on gluing together two copies of the Ben Arous expansion. For any two points $x, y \in M$, we denote by $\Gamma(x, y)$ the midpoint set of (x, y) , that is the set of points z that lay at the midpoint of length minimizing curves between x and y :

$$\Gamma(x, y) = \left\{ z \in M \mid d(x, z) = d(z, y) = \frac{d(x, y)}{2} \right\}.$$

For any $\varepsilon > 0$, we set

$$\Gamma_\varepsilon(x, y) = \left\{ z \in M \mid d(x, z) \leq \frac{d(x, y) + \varepsilon}{2} \text{ and } d(y, z) \leq \frac{d(x, y) + \varepsilon}{2} \right\}.$$

When the context is clear, we traditionally write Γ and Γ_ε instead of $\Gamma(x, y)$ and $\Gamma_\varepsilon(x, y)$. For any pair $(x, y) \in M^2$, we denote by

$$h_{x, y} = \frac{1}{2} (d(x, \cdot)^2 + d(y, \cdot)^2)$$

the *hinged energy functional*. Again, let l be any non-negative integer in the symmetric case and 0 in the general case, and let α be any multi-index. Now let $\Sigma^{l, \alpha} : \mathbb{R}^+ \times M^2 \setminus \mathcal{C} \rightarrow \mathbb{R}$ be the Taylor expansion type factor in the Ben Arous expansion of the heat kernel. That is, $\Sigma^{l, \alpha}$ is the smooth function such that

$$\Sigma_t^{l, \alpha}(x, y) = t^{|\alpha|+2l+d/2} e^{\frac{d(x, y)^2}{4t}} \partial_t^l Z_y^\alpha p_t(x, y).$$

Corollary 4. *Let \mathcal{K} be a compact subset of $M^2 \setminus \mathcal{D}$ such that all minimizers between pairs $(x, y) \in \mathcal{K}$ are strongly normal. Then for any $\varepsilon > 0$ small enough, we have uniformly on $\mathbb{R}^+ \times \mathcal{K}$, for all $(t, x, y) \in \mathbb{R}^+ \times \mathcal{K}$*

$$\partial_t^l Z_y^\alpha p_t(x, y) = \left(\frac{2}{t} \right)^{|\alpha|+2l+d} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x, y}(z)}{t}} \Phi(t/2, x, y, z) d\mu(z) + O\left(e^{-\frac{d(x, y)^2 + \varepsilon}{4t}} \right)$$

with

$$\Phi(t/2, x, y, z) = \sum_{j=0}^l \binom{l}{j} \Sigma_{t/2}^{j,0}(x, z) \Sigma_{t/2}^{k-j,\alpha}(x, z).$$

It turns out that, because the behavior of the exponential map away from the cut locus is qualitatively the same in sub-Riemannian as in Riemannian geometry, the resulting formula is structurally the same, giving the heat kernel via the Laplace asymptotics of a geometrically-motivated integral, namely a Laplace integral with phase $h_{x,y}$. The proof, however, requires working with sub-Riemannian formalism: accounting for the possibility of abnormal minimizers, defining the exponential map in terms of the Hamiltonian flow on the co-tangent space, and so on. This parallels the fact that the Ben Arous asymptotics on a sub-Riemannian manifold directly generalize the classical Minakshishundaram-Pleijel asymptotics on a Riemannian manifold, but requires additional work to prove. Given this, the asymptotics of heat kernel are determined by the theory of the asymptotics of Laplace integrals, which is a well-developed subject in its own right.

The application of this theory to Riemannian heat kernel asymptotics began with [41] and was systematically developed in [18], though only the leading term of the expansion for p_t itself was considered. In [13, 11, 12], the analogous sub-Riemannian situation was considered, including the relationship of the leading term to the geodesic geometry. Here we show that one can always bound the leading term, leading to two-sided estimates on the heat kernel itself, and to upper bounds on its derivatives, uniformly on compacts. These uniform bounds on the heat kernel, in the compact Riemannian case, are discussed in Chapter 5 of [30]. More recently, Ludewig [40] extended the upper bound to any number of simultaneous derivatives in x , y , and t for formally self-adjoint Laplace-type operators acting on a vector bundle over a compact Riemannian manifold. This was based on wave parametrix techniques, and, as above it isn't clear whether or not one should expect such results in a more general context. In the present situation, we have the following.

Proposition 5. *Let \mathcal{K} be a compact subset of $M^2 \setminus \mathcal{D}$ such that all minimizers between pairs $(x, y) \in \mathcal{K}$ are strongly normal. Then for l any non-negative integer in the symmetric case and 0 otherwise, and any multi-index α , there exists $C > 0$ such that for all $(x, y) \in \mathcal{K}$,*

$$\partial_t^l Z_y^\alpha p_t(x, y) \leq \frac{C}{t^{|\alpha|+2l} t^{d-1/2}} e^{-\frac{d(x,y)^2}{4t}}.$$

In the case $\alpha = 0$ there also exists $C' > 0$ such that for all $(x, y) \in \mathcal{K}$,

$$\frac{C'}{t^{2l} t^{d/2}} e^{-\frac{d(x,y)^2}{4t}} \leq \partial_t^l p_t(x, y).$$

More recently, there has been interest in “complete” expansions, meaning expansions to arbitrary order in t , not simply the leading term. For example, [34] uses a distributional form of Malliavin calculus, due to Watanabe, to give (potentially) complete expansions of the heat kernel on the sub-Riemannian cut locus, under some stochastically-motivated assumptions, while [39, 40] uses wave parametrix techniques to, among other things, give (potentially) complete expansions of the heat kernel of a self-adjoint Laplace-type operator acting on a vector bundle over a compact Riemannian manifold. Here, it is important to note that writing the heat kernel asymptotics in terms of a Laplace integral, which all of these approaches do, in one way or another, to obtain a complete expansion or even an exact leading term, is only possible if the asymptotics of the integral can be explicitly determined. If the hinged energy functional is real analytic in some coordinates, then the general theory of Laplace asymptotics, as developed by Arnold and collaborators [3, 4] guarantees an expansion in rational powers of t and integer powers of $\log t$. But in the general smooth case, one can have situations in which this theory does not apply, and where even the leading order appears not to be known. We observe that Corollary 4 supports complete expansions, and illustrate by giving them in two typical cases, when $h_{x,y}$ has A -type singularities and when it is Morse-Bott, in Sections 4.1 and 4.2.

In the Riemannian context, C. Bellaïche [18] discusses the possibility of non-analytic hinged energy functions and the resulting breakdown in computing the asymptotics of the Laplace integral. This

paper, however, seems not to be widely known, and the work on complete expansions just mentioned only explicitly considers the Morse-Bott case (as does the earlier work of [35]). Constructing a Riemannian metric realizing such non-analytic hinged energy functions (and more generally, arbitrary normal forms), was considered by A. Bellaïche [17], who stated the existence of such Riemannian metrics as a theorem and briefly sketched a construction. Here, we provide complete details of the proof and also consider the properly sub-Riemannian case. That is, in constructing a sub-Riemannian metric with prescribed singularities for the hinged energy function, one must take into account the constraint imposed by the distribution, and in principle one might wonder if that is an obstruction to constructing an arbitrary singularity. Given that there are an infinite number of possible growth vectors, addressing all of them is impractical, but we show that for contact manifolds (which is the most widely-studied class of sub-Riemannian manifolds), a similar construction is possible. (Other possible growth vectors are left to the interested and suitably intrepid reader.) These results are the content of Section 4.3. This indicates that properly sub-Riemannian manifolds exhibit the same diversity of possible singularities as Riemannian manifolds. (The generic situation in low dimensions is another story, for which one can see [11].)

1.3. Asymptotics for log-derivatives and bridges. In the second part of the paper, we turn to the asymptotics of logarithmic derivatives of the heat kernel and to the law of large numbers for the Brownian bridge, which are closely related and accessible to a natural modification of Molchanov's method. Probabilistically, this corresponds to considering the bridge process, rather than the underlying diffusion itself.

We express the leading term of the n th logarithmic derivative is given as an n th-order joint cumulant. In particular, we can define a family of probability measures m_t on Γ_ε in terms of a ratio of Laplace integrals, see Equation (20), which are sub-sequentially compact, see Theorem 36. In terms of the m_t , we have the following expression for the log-derivatives of p_t .

Theorem 6. *Let x and y be such that all minimal geodesics from x to y are strongly normal, and let Z^1, \dots, Z^N be smooth vector fields in a neighborhood of y (so that we understand that they act as differential operators in the y -variable). Then*

$$Z^N \dots Z^1 \log p_t(x, y) = \left(-\frac{1}{t}\right)^N \left\{ \kappa^{m_t} (d(\cdot, y) Z^1 d(\cdot, y), \dots, d(\cdot, y) Z^N d(\cdot, y)) + O(t) \right\},$$

where κ^{m_t} is the joint cumulant (of N random variables) with respect to m_t .

The logarithmic gradient and logarithmic Hessian for compact Riemannian manifolds were treated in [42], but the higher-order derivatives are new even in the Riemannian case. Moreover, we show that the (non-abnormal) cut locus is characterized by the blow up of the logarithmic Hessian, which was proven in the Riemannian case in [42]; we also note that the proof presented here is much improved over that of [42]. This is a differential analogue of the recent result of [14] showing that the cut locus is characterized by the square of the distance failing to be semi-convex. We have the following (see Section 5.2 for definitions and further details).

Corollary 7. *Let x and y be such that all minimal geodesics from x to y are strongly normal, and let \mathcal{Z} be a set of vector fields on a neighborhood of y which is C^1 -bounded and such that $\mathcal{Z}|_{T_y M}$ contains a neighborhood of the origin. Then $y \notin \text{Cut}(x)$ if and only if*

$$\limsup_{t \searrow 0} \left[\sup_{Z \in \mathcal{Z}} t |Z_y Z_y \log p_t(x, y)| \right] < \infty$$

and $y \in \text{Cut}(x)$ if and only if

$$\lim_{t \searrow 0} \left[\sup_{Z \in \mathcal{Z}} t Z_y Z_y \log p_t(x, y) \right] = \infty$$

Because of the uniformity of our approach (on compacts), as a consequence of the above, we obtain bounds on the logarithmic derivatives on compacts (disjoint from any abnormal), which in turn imply bounds on derivatives of the heat kernel itself. These bounds were proven in the case of compact

Riemannian manifolds by [29] and [44], via stochastic analysis. In the properly sub-Riemannian case, we our compact must avoid the diagonal, so the distance function doesn't explicitly appear; compare Theorem 41 with Equation (31) and Theorem 42.

Finally, we consider the small-time asymptotics of the bridge process, in particular, the law of large numbers. The law of large numbers in the sub-Riemannian case when there is a single minimizer between x and y was established in [9]; note that this includes the possibility that the minimizer is abnormal. (Such convergence to a point mass causes one to wonder about a central limit theorem result for the fluctuations, which they pursue in [8] for the case of a non-conjugate geodesic. For an on-diagonal central limit theorem, see [27].) The uniform version of this result serves as the basic ingredient in Molchanov's method, and we see that the small-time limit of the bridge process is also governed by the m_t . Let $\mu^{x,y,t}$ be the natural renormalized measure on pathspace of the bridge process from x to y in time t , and if m_0 is a probability measure on Γ , let \tilde{m}_0 denote its natural lift to pathspace (see Section 6 for details).

Theorem 8. *Let $x, y \in M$ be such that all minimizers from x to y are strongly normal. Then for any sequences of times $t_n \rightarrow 0$, μ^{x,y,t_n} converges if and only if m_{t_n} does, and if so, letting m_0 denote the limit of m_{t_n} , we have $\mu^{x,y,t_n} \rightarrow \tilde{m}_0$.*

The Riemannian version of this result (and the one below) was established in [31] using a large deviation principle. Similar large deviation principles were recently given for sub-Riemannian manifolds by Bailleul [7] and Inahama [33], but a large of large numbers was not addressed (beyond this single minimizer case as discussed above). Our approach circumvents direct use of large deviations.

In the real-analytic case, the support of the limiting measure arising in the law of large numbers can be described. In particular, in this case, one can quantify the degree of degeneracy of $h_{x,y}$ at any point $z \in \Gamma$, which by extension we think of as the degree of degeneracy of the exponential map. There is a closed, non-empty subset Γ^m of Γ which corresponds to those points of "maximum degeneracy." Then we have the following improvement to the law of large numbers.

Theorem 9. *Let x and y be such that all minimizers between them are strongly normal, and suppose that around any point of Γ there is a coordinate chart such that $h_{x,y}$ is real-analytic in these coordinates. (In particular, this holds if M itself is real-analytic.) Then m_t converges to a limit m_0 as $t \searrow 0$, and the support of m_0 is exactly Γ^m . Further, the bridge measure $\mu^{x,y,t}$ converges to \tilde{m}_0 as $t \searrow 0$.*

Finally, the limiting measure can be concretely determined in cases where the exponential map has a simple normal form, analogously to what we see for the asymptotic expansion of the heat kernel itself, and we treat the cases when $h_{x,y}$ has A -type singularities and when it is Morse-Bott in Sections 6.3 and 6.4.

1.4. Acknowledgements. The first author thanks Ismael Bailleul and Karen Habermann for helpful discussions about the law of large numbers and Ugo Boscaín for advice on sub-Riemannian technicalities (of which there are many). This work was partially supported by a grant from the Simons Foundation (#524713 to Robert Neel). This research was partially funded by the French Grant ANR ODISSE (ANR-19-CE48-0004-01).

2. LOCALIZATION AND LÉANDRE ASYMPTOTICS

For clarity in the arguments, we note that Léandre asymptotics for p_t means we can write

$$(2) \quad p_t(x, y) = \exp \left[-\frac{d^2(x, y) + a_{x,y}(t)}{4t} \right]$$

where $a_{x,y}(t)$ goes to 0 uniformly with t for any x, y in an appropriate compact. A similar bound applies to derivatives of p_t .

We first state two basic facts as lemmas.

Lemma 10. *For some n , consider the function $\sum_{i=1}^n \frac{d_i^2}{4t_i}$ on the “double simplex” $d_i \geq 0$, $\sum d_i = D$ and $t_i \geq 0$, $\sum t_i = T$. The minimum of this function is $\frac{D^2}{4T}$, achieved on the set*

$$\left\{ \frac{d_i}{t_i} = \frac{d_j}{t_j} \text{ for all } 1 \leq i, j \leq n \right\}.$$

In particular, if we fix the t_i , there is a unique choice of the d_i minimizing this function, and vice versa. Also, note the minimum does not depend on n .

The proof is an exercise in calculus, so we omit it. The next lemma is taken from [12]; see Lemma 21 and note that “geodesic” there refers to any length-minimizing curve and that $h_{x,y}(z) = \frac{1}{2} (d^2(x, z) + d^2(z, y))$.

Lemma 11. *For fixed x and y in M , consider $d^2(x, z) + d^2(z, y)$ as a function of $z \in M$. Then the function’s minimum is $\frac{1}{2}d^2(x, y)$, achieved exactly at points z with $d(x, z) = d(z, y) = \frac{1}{2}d(x, y)$. Further, such z are exactly the midpoints of length-minimizing curves from x to y , and thus there is always at least one z achieving this minimum.*

In what follows, we will work in the following situation (see Figure 1). Choose some $\varepsilon > 0$. Let K be a compact subset of M , let K' be a compact with $K \subset K'$ and $d(K, \partial K') > \frac{1}{2}\text{diam}(K) + \varepsilon$, and let K'' be a compact with $K' \subset K''$ and $d(K', \partial K'') > \varepsilon$. (This is always possible by the fact that M is complete. Note that we allow the possibility that $M = K = K' = K''$ if M itself is compact, in which case K has no boundary so that all of the distances above are infinite and the inequalities are trivially satisfied.) Here, K'' plays the role of K' in Theorem 1 (with ε replaced by 2ε), and we will prove the theorem with this change in notation.

Relative to such sets, we define the following families of stopping times for the diffusion X_t started from $x \in K'$. Let τ_1'' be the first time the process hits $\partial K''$, and let τ_1' be the first hitting time of $\partial K'$ after τ_1'' . We then continue recursively, with τ_i'' being the first hitting time of $\partial K''$ after τ_{i-1}'' and τ_i' the first hitting time of K' after τ_i'' . By path continuity, for any $t > 0$, only finitely many of the τ_i' and τ_i'' can be less than or equal to t . Next, choose any $\rho > 0$. Let $\tau_{0,1}$ be the first time X_t moves distance ρ from its starting point (that is, the first hitting time of the boundary of $B(x, \rho)$), before τ_1'' . Then let $\tau_{0,2}$ be the first time X_t moves distance ρ from $X_{\tau_{0,1}}$, and so on, with $\tau_{0,j}$ being the first time X_t moves distance ρ from $X_{\tau_{0,j-1}}$, until τ_1'' . Similarly, let $\tau_{1,1}$ be the first time after τ_1' the process moves distance ρ before τ_2'' , and so on, to define $\tau_{i,j}$ for all i and j as the j th time the process moves distance ρ since τ_i' before τ_{i+1}'' . Again by path continuity, for any $t > 0$, only finitely many of the $\tau_{i,j}$ can be less than or equal to t .

Similarly, let \mathcal{K} be compact, let \mathcal{K}' be a compact subset of K' with $\mathcal{K} \subset \mathcal{K}'$ and $d(\mathcal{K}, \partial \mathcal{K}') > \varepsilon'$. Then we let $\sigma_{0,1}$ be the first hitting time of \mathcal{K} before τ_1'' and $\sigma_{0,1}'$ the first hitting time of $\partial \mathcal{K}'$ after $\sigma_{0,1}$ and before τ_1'' . We continue recursively, with $\sigma_{0,i}$ being the first hitting time of \mathcal{K} after $\sigma_{0,i-1}'$ and before τ_1'' , and $\sigma_{0,i}'$ the first hitting time of $\partial \mathcal{K}'$ after $\sigma_{0,i}$ and before τ_1'' . Let $\sigma_{j,i}$ and $\sigma_{j,i}'$ be defined in the same way, except that they are between τ_j'' and τ_{j+1}'' .

The basic idea underlying the “self-localizing” property of the Léandre asymptotics is contained in the following lemma.

Lemma 12. *Let $K_0 \subset M$ be compact. Then there exists r such that for any $z \in K_0$, the Léandre asymptotics hold for $p_t^{\overline{B(z, r)}}(x, y)$ and its y -derivatives with $x, y \in B(z, \frac{r}{4})$, uniformly over all such x, y, z . Moreover, for such an r , choose $\rho \in (0, r)$. Let $\tau_{x, \rho}$ be the first exit time of X_t from $B(x; \rho)$, started from x , for $x \in K_0$. Then*

$$\limsup_{t \searrow 0} 4t \log (\mathbb{P}(\tau_{x, \rho} \leq t)) \leq -\rho^2,$$

uniformly over such x .

Proof. Choose $z \in K_0$ and let $r > 0$ be such that $B(z, r)$ is a topological ball and is contained in a coordinate patch around z where the sub-Riemannian structure is given by a collection of smooth vector fields (that we can assume are bounded along with all their derivatives, perhaps by taking a slightly

smaller coordinate patch). Then using a bump function, we can extend the sub-Riemannian structure on this coordinate patch to all of \mathbb{R}^k , so that it is given by a finite number of smooth vector fields that are bounded along with all of their derivatives. (Note that there is no assumption that the structure has constant rank, so there is no obstruction to performing this extension.) We also extend the smooth volume μ to all of \mathbb{R}^k , in such a way that the density of μ with respect to the Euclidean volume on \mathbb{R}^k is bounded along with all of its derivatives. The result is a sub-Riemannian structure, with sub-Laplacian and smooth volume, on \mathbb{R}^k that satisfies the hypotheses of 1, and such that $B^{\mathbb{R}^k}(0, r)$ is isometric to $B(z, r) \subset M$. Moreover, this structure on \mathbb{R}^k satisfies the assumptions of Léandre [36, 37] by construction, so that the Léandre asymptotics hold for this structure. For the moment, we will fix this sub-Riemannian structure on \mathbb{R}^k , for simplicity, so that, for example, $B^{\mathbb{R}^k}(0, r)$ above refers to the sub-Riemannian distance, not the usual Euclidean distance, and $p^{\mathbb{R}^k}(x, y)$ is the heat kernel relative to this sub-Riemannian structure. Then, in reference to the above notation, for some small $\varepsilon > 0$, we let K , K' , and K'' be the closures of $B(z, r/4)$, $B(z, r/2)$, and $B(z, r)$, respectively. We let $p_t^{K''}(x, y)$ for $x, y \in K''$ denote the heat kernel for K'' with Dirichlet boundary conditions, or equivalently the transition density for X_t killed at $\partial K''$.

Now we assume x and y are in K . Then we decompose the process according to whether or not it hits $\partial K''$ to write

$$\begin{aligned} p_t^{\mathbb{R}^k}(x, y) &= \mathbb{P}^x \left(X_t^{\mathbb{R}^k} \in dy \text{ and } \tau_1'' \geq t \right) + \mathbb{P}^x \left(X_t^{\mathbb{R}^k} \in dy \text{ and } \tau_1'' < t \right) \\ &= p_t^{K''}(x, y) + \int p_{t-\tau_1''}^{\mathbb{R}^k}(X_{\tau_1''}, y) \cdot \mathbf{1}_{\{\tau_1'' \leq t\}} d\mathbb{P}^x(\tau_1'', X_{\tau_1''}). \end{aligned}$$

(Here $X_t^{\mathbb{R}^k}$ refers to the associated diffusion process on \mathbb{R}^k , of course, and \mathbb{P}^x to the probability of the diffusion started from x .) In what follows, we adopt the convention that $p_{t-\tau}(x, y) = 0$ if $\tau > t$ to avoid the need to include various indicator functions in our path decompositions.

As noted, $p_t^{\mathbb{R}^k}(x, y)$ satisfies (2) uniformly on K'' . Then for the second term on the right-hand side, using that $d(X_{\tau_1''}, y) > \frac{3}{4}r$, we have

$$p_{t-\tau_1''}^{\mathbb{R}^k}(X_{\tau_1''}, y) \leq \exp \left[-\frac{(\frac{3}{4}r)^2 + a(t)}{4t} \right]$$

whenever $\tau_1'' \leq t$, independent of $X_{\tau_1''}$, y , and particular value of τ_1'' (and where $a(t) \rightarrow 0$). Because we're integrating with respect to a probability measure, the same estimate holds for the integral, so that

$$0 \leq p_t^{\mathbb{R}^k}(x, y) - p_t^{K''}(x, y) \leq \exp \left[-\frac{(\frac{3}{4}r)^2 + a(t)}{4t} \right]$$

uniformly for $x, y \in K$, for all small enough t . Since $p_t^{\mathbb{R}^k}(x, y)$ satisfies (2) and $d(x, y) \leq r/2$, this proves the result for $p_t^{K''} = \overline{p_t^{B(z, r)}}$ itself.

For the y -derivatives, first note that

$$\partial_y^\alpha p_t^{\mathbb{R}^k}(x, y) = \partial_y^\alpha p_t^{K''}(x, y) + \partial_y^\alpha \int p_{t-\tau_1''}^{\mathbb{R}^k}(X_{\tau_1''}, y) d\mathbb{P}^x(\tau_1'', X_{\tau_1''}).$$

for any multi-index α . Since $\partial_y^\alpha p_t^{\mathbb{R}^k}(x, y)$ also satisfies (2) uniformly (for all “sub-multi-indices” of α as well), we can move the derivative inside the integral, at which point the argument proceeds just as for p_t itself, to show that

$$(3) \quad \left| \partial_y^\alpha p_t^{\mathbb{R}^k}(x, y) - \partial_y^\alpha p_t^{K''}(x, y) \right| \leq \exp \left[-\frac{(\frac{3}{4}r)^2 + a(t)}{4t} \right].$$

Because $\partial_y^\alpha p_t^{\mathbb{R}^k}(x, y)$ satisfies (2) and $d(x, y) \leq r/2$, this completes the proof of the first part of the lemma.

Next, we consider the exit times. That Léandre asymptotics imply such bounds is fairly well known, but we give a short proof for completeness.

We begin with a preliminary estimate. We claim that for any $\varepsilon > 0$, there is $t_0 > 0$ such that

$$\int_{B(z, \varepsilon)} p_t(z, y) \mu(dy) > 1/2 \quad \text{for any } z \in K_0 \text{ and any } t \in (0, t_0).$$

To see this, note that p_t is smooth and converges to a point mass as $t \searrow 0$, so this holds at any z , and then uniformly in z by compactness. For the more stochastically-minded, we can instead observe that K_0 can be covered by finitely many smooth coordinate patches such that $B(z; 2\varepsilon)$ will be contained in a smooth coordinate patch, after possibly shrinking ε . Then the diffusion X_t started from x is given in these coordinates by a system of SDEs with smooth coefficients. It follows that there is some small time t_0 such that the probability that X_t started from z exits $B(z; \varepsilon)$ by time t_0 is less than $1/2$. Moreover, by smoothness this T can be found uniformly with respect to all such z in the coordinate patch, and thus by compactness, for any z .

Now choose $x \in K_0$ and $\rho < r$, and consider the heat kernel at time t on the annulus $A_\varepsilon(x; \rho) = \{d(z, x) \in [\rho - \varepsilon, \rho + \varepsilon]\}$ for some small $\varepsilon > 0$ and $t \leq t_0$ as above. If τ_ρ is the first hitting time of $\{z : d(x, z) = \rho\}$ for the diffusion X_t started from x , then by the strong Markov property, and the preliminary estimate, we have

$$\mathbb{P}^x(X_t \in A_\varepsilon(x; \rho)) = \int_{A_\varepsilon(x; \rho)} p_t(x, z) d\mu(z) > \frac{1}{2} \mathbb{P}^x(\tau_\rho \leq t).$$

This integral can be estimated uniformly in x by the Léandre estimate on the heat kernel. More precisely, for $\delta > 0$, after possibly making t_0 smaller,

$$\int_{A_\varepsilon(x; \rho)} p_t(x, z) d\mu(z) \leq \mu(A_\varepsilon(x; \rho)) \exp \left[-\frac{(\rho - \varepsilon)^2 - \delta}{4t} \right]$$

for any $t \leq t_0$ and for any $x \in K_0$. Since the measure of $A_\varepsilon(x; \rho)$ is uniformly bounded in x (by smoothness and compactness), we conclude that

$$\mathbb{P}^x(\tau_\rho \leq t) \leq C \exp \left[-\frac{(\rho - \varepsilon)^2 - \delta}{4t} \right]$$

for some $C > 0$ independent of x , for all $t \leq t_0$. Since ε and δ are (small and) arbitrary, standard algebraic manipulations then give

$$\limsup_{t \searrow 0} 4t \log(\mathbb{P}^x(\tau_\rho \leq t)) \leq -\rho^2,$$

uniformly for $x \in K_0$. □

We are now in a position to prove localization and the Léandre asymptotics, as given in Theorem 1.

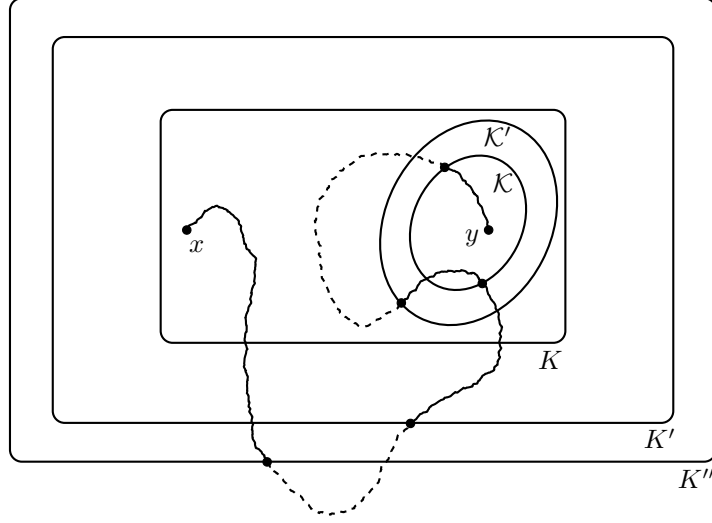
Proof of Theorem 1. In light of Lemma 12, we can choose $\rho > 0$ such that, for any $x \in K''$,

$$\limsup_{t \searrow 0} 4t \log(\mathbb{P}(\tau_{x, \rho} \leq t)) \leq -\rho^2,$$

uniformly over such x . Moreover, after possibly shrinking ρ , we can assume that there is some $\varepsilon' > 0$ and a finite number of pairs of sets $(\mathcal{K}_i, \mathcal{K}'_i)$ such that $\mathcal{K}_i \subset \mathcal{K}'_i \subset K''$ and $d(\mathcal{K}_i, (\mathcal{K}'_i)^c) > \varepsilon'$ for all i , every $y \in K'$ there is some i such that $B(y, 4\rho) \subset \mathcal{K}_i$ and Léandre asymptotics for $p_t^{\mathcal{K}'}(z, z')$ and its derivatives hold uniformly over all $z, z' \in K_i$.

Now choose $x, y \in K'$, and suppose that $d(x, y) > 2\rho$. Let m be the largest integer such that $m\rho < d(x, y) - \rho$, and note that m is at least 1. In reference to the notation introduced above, let \mathcal{K} be the closed ball around y of radius $d(x, y) - m\rho$ and observe that this radius is in the interval $(\rho, 2\rho]$. Further, because K'' has finite diameter, there is some N independent of x and y such that $m \leq N$. We \mathcal{K}' be the appropriate \mathcal{K}'_i containing y as above. If we further suppose that $d(y, \partial K') > 2\rho$, then we can write the heat kernel in terms of the following path decomposition,

$$p_t^M(x, y) = p_t^{K''}(x, y) + \mathbb{P}^x(X_t^M \in dy \text{ and } \tau_1'' < t),$$

FIGURE 1. A possible path from x to y , decomposed according to hitting times.

where

$$p_t^{K''}(x, y) = \sum_{i=1}^{\infty} \int p_{t-\sigma_{0,i}}^{K'}(X_{\sigma_{0,i}}, y) d\mathbb{P}(\sigma_{0,i}, X_{\sigma_{0,i}}) \text{ and}$$

$$\mathbb{P}^x(X_t^M \in dy \text{ and } \tau_1'' < t) = \sum_{j,i=1}^{\infty} \int p_{t-\sigma_{j,i}}^{K'}(X_{\sigma_{j,i}}, y) d\mathbb{P}(\sigma_{j,i}, X_{\sigma_{j,i}}).$$

Note that this is essentially a “two level” version of (13) in [9] (which is itself Bailleul and Norris’ summary of the main idea of Hsu’s approach in [32]), which accounts for the fact that we don’t yet have the Léandre asymptotics on compacts, but only on small enough balls.

We first consider $p_t^{K''}(x, y)$. Note that $\sigma_{0,1} \geq \tau_{0,m}$. We claim that for any $\delta > 0$, which we can and do assume is less than ρ^2 , there exists $t_0 > 0$ such that for $t < t_0$

$$\mathbb{P}^x[\tau_{0,m} \leq t] \leq \exp\left[-\frac{m^2\rho^2 - \delta}{4t}\right]$$

for all $x \in K'$ and all $m \in 1, 2, \dots, N$. We already know this is true for $m = 1$, and by the strong Markov property, $\tau_{0,i+1} - \tau_{0,i}$ satisfies the same bound as $\tau_{0,1}$. Now consider $\tau_{0,2}$ and any $\delta > 0$. Let $V_1(t)$ be the cdf of $\tau_{0,1}$. Then since, the probability of $\tau_{0,2} \leq t$, given that $\tau_{0,1} = s$, is bounded from above by $\exp\left[-\frac{\rho^2 - \delta}{4(t-s)}\right]$ for small t , uniformly in $X_{\tau_{0,1}}$, we have

$$\mathbb{P}^x[\tau_{0,2} \leq t] \leq \int_0^t \exp\left[-\frac{\rho^2 - \delta}{4(t-s)}\right] dV_1(s)$$

for all small enough t , uniformly in x , where the integral is understood as a Lebesgue-Stieljes integral. Now since V_1 is non-decreasing and has bounded variation and the integrand is non-increasing and continuous, has bounded variation, and is differentiable on $s \in (0, t)$, we have an integration by parts formula for the integral. The boundary terms vanish, since the integrand goes to zero as $s \nearrow t$ and $V_1(s)$ goes to zero as $s \searrow 0$, and we find

$$\mathbb{P}^x[\tau_{0,2} \leq t] \leq \int_0^t V_1(s) \frac{\rho^2 - \delta}{4(t-s)^2} \exp\left[-\frac{\rho^2 - \delta}{4(t-s)}\right] ds.$$

As long as t is small enough, we can absorb the $\frac{\rho^2 - \delta}{4(t-s)^2}$ factor into the exponential at the cost of replacing δ with 2δ . This plus the previous estimate for V_1 gives

$$\mathbb{P}^x[\tau_{0,2} \leq t] \leq \int_0^t \exp\left[-\frac{\rho^2 - \delta}{4s} - \frac{\rho^2 - 2\delta}{4(t-s)}\right] ds$$

uniformly in x , for all small enough t . Using Lemma 10 and the fact that δ is arbitrary, a naive estimate for the integral gives the result for $\tau_{0,2}$. From here, we can iterate the argument (finitely many times) to establish the claim for $\tau_{0,m}$ (with $m \leq N$), noting that estimates are valid as long as $\tau_{0,m}$ is finite, since by the definition of $\tau_{0,m}$ this means the process is still in K'' , and the estimates hold trivially on paths for which $\tau_{0,m}$ is infinite.

Next, we claim that, after possibly shrinking t_0 , for $t < t_0$,

$$\mathbb{P}^x[\sigma_{0,i} \leq t] \leq \left(\frac{1}{2}\right)^{i-1} \exp\left[-\frac{m^2 \rho^2 - \delta}{4t}\right]$$

for all x and y under consideration and all i , where m is understood as a function of y as above. The point is that we already have this for $i = 1$, and between $\sigma_{0,1}$ and $\sigma_{0,i}$, the particle must travel a distance of ε' at least $i - 1$ times. We have already seen in Lemma 12 that the probability of the particle traveling a fixed (small) distance in time t can be uniformly bounded from above. So as long as t_0 is small enough, that probability is less than $1/2$, and we have the above coarse estimate, for any i .

By our choice of \mathcal{K} , we have that, after possibly shrinking t_0 , for all $t < t_0$,

$$p_{t-\sigma_{0,i}}^{\mathcal{K}'}(X_{\sigma_{0,i}}, y) \leq \exp\left[-\frac{(d(x, y) - m\rho)^2 - \delta}{4(t - \sigma_{0,i})}\right]$$

for all x, y , and i under consideration. Since the cdf of σ is bounded as above, we again use integration by parts and Lemma 10 to see that, after adjusting δ and t_0 ,

$$\int p_{t-\sigma_{0,i}}^{\mathcal{K}'}(X_{\sigma_{0,i}}, y) d\mathbb{P}(\sigma_{0,i}, X_{\sigma_{0,i}}) \leq \left(\frac{1}{2}\right)^{i-1} \exp\left[-\frac{d^2(x, y) - \delta}{4t}\right]$$

for all $t < t_0$, for all x, y , and i . Since δ is arbitrary and the geometric series is summable, this establishes that

$$\limsup_{t \searrow 0} 4t \log p_t^{K''}(x, y) \leq -d^2(x, y).$$

Thus we have the upper bound for $p^{K''}$ we want, assuming in addition that $d(x, y) > 2\rho$, $d(y, \partial K') > 2\rho$ and $x \in K'$. We remove those restrictions next.

If $d(x, y) \leq 2\rho$, this upper bound follows by a straightforward modification of the above. Namely, now we set $\sigma_{0,1} = 0$ and have

$$p_t^{K''}(x, y) = p_t^{\mathcal{K}'}(x, y) + \sum_{i=2}^{\infty} \int p_{t-\sigma_{0,i}}^{\mathcal{K}'}(X_{\sigma_{0,i}}, y) d\mathbb{P}(\sigma_{0,i}, X_{\sigma_{0,i}}).$$

Then $p_t^{\mathcal{K}'}(x, y)$ satisfies the desired bound by Lemma 12, and the other terms are estimated as above to satisfy a smaller bound. For general $x, y \in K''$, take slightly larger compacts $K'_0 \subset K''_0$ with $K'' \subset K'_0$ and $d(K'', \partial K'_0) > 2\rho$, apply the previous with K'_0 and K''_0 in place of K' and K'' , and note that that we have the monotonicity property $p^{K''} \leq p^{K'_0}$.

Thus we have proven the upper bound for $p^{K''}$.

For $y \in K$, we extend the above argument to the derivatives of $p_t^{\mathcal{K}'}(x, y)$ exactly as in Lemma 12, noting that \mathcal{K}' is contained in K' , so there's no problem applying the derivatives to $p^{\mathcal{K}'}$ and using the Léandre estimate. (Indeed, this is why we assume $y \in K$. For the p_t itself, we have monotonicity under stopping the process at $\partial K''$, as just observed, but there's no guarantee that the derivatives are as well behaved under this extra stopping, if \mathcal{K}' were to intersect $\partial K''$.)

We turn to bounds on $\mathbb{P}^x(X_t^M \in dy \text{ and } \tau_1'' < t)$, which will give the desired localization. We now assume $x, y \in K$ (with no assumption on the distance between them). Now we can assume that $\rho < \varepsilon$,

and by considering $\tau_{0,m}$ for appropriate m , analogously to the above, we see that, for any $\delta > 0$, there exists t_0 such that

$$\mathbb{P}^x [\tau_1'' \leq t] \leq \exp \left[-\frac{(\frac{1}{2}\text{diam}(K) + 2\varepsilon)^2 - \delta}{4t} \right]$$

whenever $t < t_0$, for any $x \in K$. Since $\sigma_{1,1} - \tau_1''$ satisfies the same bound, arguing as above we see that, for $\delta > 0$ there exists t_0 such that

$$\mathbb{P}^x [\sigma_{j,i} \leq t] \leq \left(\frac{1}{2}\right)^{j-1} \left(\frac{1}{2}\right)^{i-1} \exp \left[-\frac{(\text{diam}(K) + \varepsilon)^2 - \delta}{4t} \right]$$

whenever $t < t_0$, for any $x, y \in K$. Using the obvious estimate for $p_{t-\sigma_{j,i}}^{K'}(X_{\sigma_{j,i}}, y)$, we then use integration by parts and summability of geometric series as above to conclude that

$$\mathbb{P}^x (X_t^M \in dy \text{ and } \tau_1'' < t) < \exp \left[-\frac{(\text{diam}(K) + \varepsilon)^2}{4t} \right]$$

for all sufficiently small t , uniformly in x and y . Again, derivatives are handled exactly as in Lemma 12. Note also that all of the estimates used above depend only on the behavior of X_t in a neighborhood of K'' , so that the implicit constants in the “big-O” terms also depend only on M restricted to a neighborhood of K'' .

At this point, we have proven everything in the theorem except for the lower bound on $p_t^{K''}(x, y)$ for $x, y \in K$.

We already know that the lower bound holds for $x, y \in K$ with $d(x, y) < \rho$ because $p_t^{K''}(x, y) \geq p_t^{K'}(x, y)$. Now take $\delta' \in (0, \rho/4)$, and suppose $d(x, y) < 2\rho - 2\delta'$. Recall the definition of $\Gamma = \Gamma(x, y)$ and of $\Gamma_{\delta'}$. Then Γ is compact and non-empty, and by the choice of K and K'' and the triangle inequality, $\Gamma_{\delta'} \subset K''$ (and we note for the future that this is true for any $x, y \in K$, not just those with $d(x, y) < 2\rho - 2\delta'$). By the Chapman-Kolmogorov inequality (which is another manifestation of the strong Markov property), and non-negativity of $p_t^{K''}$ we have

$$p_t^{K''}(x, y) \geq \int_{\Gamma_{\delta'}} p_{t/2}^{K''}(x, z) p_{t/2}^{K''}(z, y) d\mu(z).$$

By our choice of $\Gamma_{\delta'}$, the Léandre asymptotics hold for both factors in the integrand. Thus for any $\delta > 0$, we can find $t_0 > 0$ such that

$$p_t^{K''}(x, y) \geq \int_{\Gamma_{\delta'}} \exp \left[-\frac{d^2(x, z) + d^2(z, y) + \delta}{2t} \right] d\mu(z).$$

Since Γ is non-empty, for any $r' \in (0, \delta')$, $\Gamma_{\delta'}$ contains at least one ball of radius r' . Choose r' so that

$$d^2(x, z) + d^2(z, y) < \frac{1}{2} d^2(x, y) + \delta$$

for z in such a ball; this can be done uniformly in x, y, z by the triangle inequality and continuity of the distance. Because μ is smooth and K'' compact, the μ -measure of such a ball is uniformly bounded from below, say by $c > 0$. It follows that, for $t < t_0$,

$$p_t^{K''}(x, y) \geq c \exp \left[-\frac{d^2(x, y) + \delta}{4t} \right]$$

for all x, y under consideration. This proves the lower bound when $d(x, y) < 2\rho - 2\delta'$.

But we can iterate this argument any finite number of times, so that for any N , the lower bound holds uniformly for $x, y \in K$ such that $d(x, y) < 2N\rho - 2N\delta'$. But since K has finite diameter, for large enough N this is all $x, y \in K$, proving the theorem (after adjusting the notation to replace (K, K', K'') with (K, K')). \square

Remark 13. Since any compact K in a complete sub-Riemannian manifold is contained in some K' (possibly M itself, if M is compact) as in Theorem 1, it follows that the Léandre asymptotics (for p_t and its derivatives) hold uniformly on any K . Even if we allow M to be incomplete, as long as $d(K, \infty) > \frac{1}{2}\text{diam}(K)$, we can localize the analysis of heat kernel asymptotics to K , and in particular, the Léandre asymptotics hold on K . Thus, all of the results that follow, while stated for complete M , also hold for such K in incomplete M via a straightforward argument.

More generally, we can consider $(x, y) \in \mathcal{K}$ where $\mathcal{K} \subset M \times M$ is compact. Then the condition $d(K, \infty) > \frac{1}{2}\text{diam}(K)$ is replaced by $\inf_x d(x, \infty) + \inf_y d(y, \infty) > \sup_{(x, y)} d(x, y)$, with a superficial modification of the proof. Indeed, the above is just the symmetric case $\mathcal{K} = K \times K$. This implies that the Léandre asymptotics hold uniformly on any compact subset of the set $\{(x, y) : d(x, y) < d(x, \infty) + d(y, \infty)\}$, as proven in [9] for the sub-Riemannian heat kernel itself. This is the best possible such result in general, as shown in [32].

The question of when we can localize $p_t(x, y)$ to $\{z : d(x, z) + d(z, y) \leq d(x, y) + \varepsilon\}$ is much more subtle. As we will see below (starting Corollary 15 and continuing into Section 6), this is true if M is complete, and thus more generally on K as in Theorem 1. Better conditions were the primary focus of [9], where this result was established (for p_t itself) on incomplete sub-Riemannian manifolds under a sector condition that limits the amount of asymmetry of the sub-Laplacian. This allowed them to show that the Léandre asymptotics for p_t hold uniformly on any compact, with no assumption on the distance to infinity, under this condition.

Since the asymptotic results that follow inherently respect localization, the results for p_t itself can be generalized to incomplete manifolds satisfying the sector condition under the weaker constraint on the distance to infinity, using the work of Bailleul and Norris. Again, we don't explicitly state these extensions, leaving them to the interested reader.

Corollary 14. *Consider the same situation as in Theorem 1, but assume in addition that $p_t(x, y)$ is symmetric. Then for any multi-index α , and any non-negative integer l , the results of Theorem 1 hold with Z_y^α replaced by $Z_y^\alpha \partial_t^l$. Also, for $x \in K$ and $y \in K'$,*

$$\limsup_{t \searrow 0} 4t \log \left(\left| Z_x^\alpha \partial_t^l p_t^{K'}(x, y) \right| \right) \leq -d^2(x, y).$$

Proof. Note that by symmetry, $p_t(x, y)$ satisfies the heat equation $\partial_t f = \Delta f$ with respect to either spatial variable. Thus replacing ∂_t with Δ l times (recalling Equation (1)) and expanding, we see that there is some positive N and multi-indices $\alpha_1, \dots, \alpha_N$ such that

$$Z_y^\alpha \partial_t^l p_t(x, y) = \sum_{i=1}^N Z_y^{\alpha_i} p_t(x, y).$$

This the first claim follows by Theorem 1 and the observation that a sum of quantities satisfying the Léandre asymptotics also satisfies the Léandre asymptotics. If $x \in K$ and $y \in K'$, by symmetry we simply exchange the two variables, so that the result again follows from Theorem 1. \square

Recall

$$\Gamma_\varepsilon(x, y) = \left\{ z \in M \mid d(x, z) \leq \frac{d(x, y) + \varepsilon}{2} \text{ and } d(y, z) \leq \frac{d(x, y) + \varepsilon}{2} \right\}.$$

Now that Léandre asymptotics are proved, we have enough to show that the heat responsible for $p_t(x, y)$ is located near the midpoint set Γ at $t/2$.

Corollary 15. *Let \mathcal{K} be a compact subset of M^2 . Let l be a non-negative integer in the symmetric case and zero otherwise and α be a multi-index. For any $\varepsilon > 0$ small enough, we have uniformly on $\mathbb{R}^+ \times \mathcal{K}$, for all $(t, x, y) \in \mathbb{R}^+ \times \mathcal{K}$*

$$\partial_t^l Z_y^\alpha p_t(x, y) = \int_{\Gamma_\varepsilon} \partial_t^l Z_y^\alpha (p_{t/2}(x, z) p_{t/2}(z, y)) d\mu(z) + O\left(e^{-\frac{d(x, y)^2 + \varepsilon^2}{4t}}\right).$$

Proof. Let $\varepsilon > 0$ and let define $K \subset M$ the closure of the set of points z for which there exists $(x, y) \in M$ such that either $(x, z) \in \mathcal{K}$, $(z, y) \in K$ or $z \in \Gamma_\varepsilon(x, y)$ for $(x, y) \in \mathcal{K}$. K is naturally bounded and thus compact. As a consequence, Theorem 1 applies, as does Corollary 14 in the symmetric case. Let K' be a compact set such that $K \subset K'$, $d(K, \partial K') > 2\text{diam}(K)$, so that there exists $C > 0$ satisfying for all $x, y \in K$

$$\left| \partial_t^l Z_y^\alpha p_t(x, y) - \partial_t^l Z_y^\alpha p_t^{K'}(x, y) \right| \leq C \exp\left(-\frac{\text{diam}(K)^2}{2t}\right).$$

Then

$$\begin{aligned} \partial_t^l Z_y^\alpha p_t^{K'}(x, y) &= \partial_t^l Z_y^\alpha \int_{K'} \left(p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) \right) d\mu(z) \\ &= \int_{K'} \sum_{j=0}^l \binom{j}{l} \partial_t^j p_{t/2}^{K'}(x, z) \partial_t^{l-j} Z_y^\alpha p_{t/2}^{K'}(z, y) d\mu(z). \end{aligned}$$

Theorem 1 and Corollary 14 (in the symmetric case) imply that there exists $C > 0$ such that for all $x, y \in K$, $z \in K'$, and all integers j between 0 and l ,

$$\left| \partial_t^j p_{t/2}^{K'}(x, z) \right| \leq C \exp\left(-\frac{d(x, z)^2}{2t}\right)$$

and

$$\left| \partial_t^{l-j} Z_y^\alpha p_{t/2}^{K'}(z, y) \right| \leq C \exp\left(-\frac{d(z, y)^2}{2t}\right).$$

Hence

$$\left| \sum_{j=0}^l \binom{l}{j} \partial_t^j p_{t/2}^{K'}(x, z) \partial_t^{l-j} Z_y^\alpha p_{t/2}^{K'}(z, y) \right| \leq C^2 2^l \exp\left(-\frac{d(x, z)^2 + d(z, y)^2}{2t}\right).$$

If $d(x, z) > \frac{d(x, y) + \varepsilon}{2}$ then by triangular inequality, $d(z, y) > \frac{d(x, y) - \varepsilon}{2}$, so that

$$h_{x, y}(z) = \frac{1}{2} (d(x, z)^2 + d(z, y)^2) > \frac{1}{2} \left(\left(\frac{d(x, y) + \varepsilon}{2} \right)^2 + \left(\frac{d(x, y) - \varepsilon}{2} \right)^2 \right) = \frac{d(x, y)^2 + \varepsilon^2}{4}.$$

Of course, this observation holds by inverting the roles of x and y , so that for all $x, y \in K$, if $z \in K' \setminus \varepsilon$, then

$$\left| \sum_{j=0}^l \binom{l}{j} \partial_t^j p_{t/2}^{K'}(x, z) \partial_t^{l-j} Z_y^\alpha p_{t/2}^{K'}(z, y) \right| < C^2 2^l \exp\left(-\frac{d(x, y)^2 + \varepsilon^2}{4t}\right).$$

Integrated over K' , this yields

$$\begin{aligned} \int_{K'} \partial_t^l Z_y^\alpha \left(p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) \right) d\mu(z) &= \int_{\Gamma_\varepsilon} \partial_t^l Z_y^\alpha \left(p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) \right) d\mu(z) \\ &\quad + \int_{K' \setminus \Gamma_\varepsilon} \partial_t^l Z_y^\alpha \left(p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) \right) d\mu(z) \end{aligned}$$

where

$$\int_{K' \setminus \Gamma_\varepsilon} \partial_t^l Z_y^\alpha \left(p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) \right) d\mu(z) \leq \mu(K') C^2 2^l \exp\left(-\frac{d(x, y)^2 + \varepsilon^2}{4t}\right).$$

Now we impose a restriction: we only consider pairs of points $(x, y) \in \mathcal{K}$, so that pairs (x, z) and (z, y) , with $z \in \Gamma_\varepsilon(x, y)$ all belong to K^2 . Thanks to this restriction, we are able to say that for all $(x, y) \in \mathcal{K}$, and that there exists $C > 0$ such that for all $(x, y) \in \mathcal{K}$,

$$\int_{\Gamma_\varepsilon} \left[\partial_t^l Z_y^\alpha \left(p_{t/2}(x, z) p_{t/2}(z, y) \right) - \partial_t^l Z_y^\alpha \left(p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) \right) \right] d\mu(z) \leq C \mu(K') \exp\left(-\frac{\text{diam}(K)^2}{2t}\right)$$

(since $\mu(\Gamma_\varepsilon) \leq \mu(K')$).

As a conclusion, for all $(x, y) \in \mathcal{K}$,

$$\begin{aligned} \left| \partial_t^l Z_y^\alpha p_t(x, y) - \int_{\Gamma_\varepsilon} \partial_t^l Z_y^\alpha (p_{t/2}(x, z) p_{t/2}(z, y)) d\mu(z) \right| \leq \\ \left| \partial_t^l Z_y^\alpha p_t(x, y) - \partial_t^l Z_y^\alpha p_t^{K'}(x, y) \right| \\ + \int_{K' \setminus \Gamma} \partial_t^l Z_y^\alpha (p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y)) d\mu(z) \\ + \int_{\Gamma} \partial_t^l Z_y^\alpha (p_{t/2}^{K'}(x, z) p_{t/2}^{K'}(z, y) - p_{t/2}(x, z) p_{t/2}(z, y)) d\mu(z), \end{aligned}$$

so that

$$\begin{aligned} \left| \partial_t^l Z_y^\alpha p_t(x, y) - \int_{\Gamma_\varepsilon} \partial_t^l Z_y^\alpha (p_{t/2}(x, z) p_{t/2}(z, y)) d\mu(z) \right| \leq C e^{-\frac{\text{diam}(K)^2}{2t}} + C' e^{-\frac{d(x, y)^2 + \varepsilon^2}{4t}} C'' e^{-\frac{\text{diam}(K)^2}{2t}} \\ \leq C''' e^{-\frac{d(x, y)^2 + \varepsilon^2}{4t}}, \end{aligned}$$

which proves the corollary. \square

3. BEN AROUS EXPANSION THEOREM

In [19], Ben Arous gives full asymptotic expansions of the heat kernel in small time, for pairs of points away from the diagonal, the cut locus, or joined by strictly abnormal minimizers. In the rest of the paper, for any $x \in M$, $\text{Cut}(x)$ denotes the cut locus of x in M . Furthermore, we recall the critical set $\mathcal{C} \subset M \times M$ from Definition 2: the set of pairs of points (x, y) such that either $y \in \text{Cut}(x)$, $x \in \text{Cut}(y)$, $x = y$ in the non-Riemannian case (that is, \mathcal{C} contains the diagonal in the properly sub-Riemannian case), or such that a length minimizing curve from x to y is not strongly normal. Note that Ben Arous definition of cut locus includes points connected by an abnormal geodesic, which is not the convention we follow, hence the introduction of \mathcal{C} . We can describe Ben Arous results with the following definition, which will supply convenient terminology for this section and allow us to treat the symmetric and general cases in parallel.

Definition 16. We say that Ben Arous expansions hold uniformly on a compact subset $\mathcal{K} \subset M^2 \setminus \mathcal{C}$ if for any l which is a non-negative integer in the symmetric case and 0 in general and any multi-index α , we have the following.

There exists an open neighborhood \mathcal{O} of \mathcal{K} in $M^2 \setminus \mathcal{C}$, there exist sequences of smooth functions $c_k : \mathcal{O} \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, $r_k : \mathbb{R}^+ \times \mathcal{O} \rightarrow \mathbb{R}$, such that for all $N \in \mathbb{N}$, for all $(x, y) \in \mathcal{O}$, for all t small enough

$$\partial_t^l Z_y^\alpha p_t(x, y) = t^{-(|\alpha| + 2l + n/2)} e^{-\frac{d(x, y)^2}{4t}} \left(\sum_{k=0}^N c_k(x, y) t^k + t^{N+1} r_{N+1}(t, x, y) \right),$$

and, for l' which is any non-negative integer in the symmetric case and 0 otherwise, and any multi-index α' , there exists t_0 such that

$$\sup_{0 < t < t_0} \sup_{(x, y) \in \mathcal{K}} \left| \partial_t^{l'} Z_y^{\alpha'} r_{N+1}(t, x, y) \right| < \infty.$$

Additionally, if $\alpha = 0$, then $c_0(x, y) > 0$ on \mathcal{O} .

In particular, Theorems 3.1-3.4 in [19] imply that the heat kernel satisfies Definition 16 with $M = \mathbb{R}^d$ and \mathcal{K} any compact set in $\mathbb{R}^d \times \mathbb{R}^d \setminus \mathcal{C}$.

It has been widely accepted that Ben Arous expansions should hold uniformly on compact sets where no abnormal minimizers exist between two distinct points. Thanks to localization, we are able to prove this fact using Molchanov's method, and we then apply this result to give uniform universal bounds of the heat kernel on compact sets without abnormal minimizers.

Techniques in this section rely heavily on properties of midpoint sets, the set of points equidistant from two points. Hence we preface our work on Ben Arous expansions with some preliminary remarks and definitions that will appear in proofs throughout the section.

3.1. Some properties of midpoint sets. Recall that for any two points $x, y \in M$, we denote by $\Gamma(x, y)$ the midpoint set of (x, y) , that is the set of points z that lay at the midpoint of length minimizing curves between x and y :

$$\Gamma(x, y) = \left\{ z \in M \mid d(x, z) = d(z, y) = \frac{d(x, y)}{2} \right\}$$

and for any $\varepsilon > 0$, we set

$$\Gamma_\varepsilon(x, y) = \left\{ z \in M \mid d(x, z) \leq \frac{d(x, y) + \varepsilon}{2} \text{ and } d(y, z) \leq \frac{d(x, y) + \varepsilon}{2} \right\}.$$

For any $\eta \geq 0$, we denote by $\mathcal{D}(\eta)$ the subset of M^2

$$\mathcal{D}(\eta) = \{(x, y) \in M^2 \mid d(x, y) \leq \eta\}.$$

Then notice in particular that $\mathcal{D}(0) = \mathcal{D} \subset M \times M$ denotes the diagonal of M^2 .

Lemma 17. *Let \mathcal{K} be a compact subset of M^2 such that $\mathcal{K} \cap \mathcal{D} = \emptyset$ and no strictly abnormal minimizers exist between any pair $(x, y) \in \mathcal{K}$. Let*

$$\Gamma^l(\mathcal{K}) = \{(x, z) \mid \exists y \in M \text{ s.t. } (x, y) \in \mathcal{K}, z \in \Gamma(x, y)\},$$

$$\Gamma^r(\mathcal{K}) = \{(z, y) \mid \exists x \in M \text{ s.t. } (x, y) \in \mathcal{K}, z \in \Gamma(x, y)\}.$$

Then $\Gamma^l(\mathcal{K})$ and $\Gamma^r(\mathcal{K})$ are compact subsets of $M^2 \setminus \mathcal{C}$.

Proof. By symmetry of the definitions, we prove the statement for $\Gamma^l(\mathcal{K})$. The manifold M is geodesically complete, which implies that geodesics between pairs of points $(x, y) \in \mathcal{K}$ exist and all remain in a bounded set in M . As a consequence, $\Gamma^l(\mathcal{K})$ is well defined and is also bounded. By definition, a point in $(x, z) \in \Gamma^l(\mathcal{K})$ cannot be in \mathcal{C} since z is the midpoint of a strongly normal geodesic. Then what has to be shown is the closure of $\Gamma^l(\mathcal{K})$.

Let $(x_n, z_n)_{n \in \mathbb{N}} \in \Gamma^l(\mathcal{K})$, let $(x^*, z^*) \in M^2$ such that $(x_n, z_n) \rightarrow (x^*, z^*)$. Let us prove that there exists $y^* \in M$ such that $(x^*, y^*) \in \mathcal{K}$ and $z^* \in \Gamma(x^*, y^*)$. It is clear that for all $n \in \mathbb{N}$, there exists $y_n \in M$ such that $(x_n, y_n) \in \mathcal{K}$ and $z_n \in \Gamma(x_n, y_n)$. Up to extraction, (y_n) converges to a limit y^* that constitutes our candidate.

For all n , there exists a constant speed curve $\gamma_n : [0, 1] \rightarrow M$ such that $\gamma_n(0) = x_n$, $\gamma_n(1/2) = z_n$ and $\gamma_n(1) = y_n$.

Since γ_n is a length minimizing curve,

$$d(\gamma_n(s), \gamma_n(t)) = \int_s^t |\dot{\gamma}_n(\tau)| d\tau = \ell(\gamma_n)|t - s| \leq d(x_n, y_n)|t - s|.$$

Since \mathcal{K} is compact, this inequality implies that there exists $\kappa > 0$ such that $d(\gamma_n(s), \gamma_n(t)) \leq \kappa|t - s|$ for all n . Hence the set of maps $\{\gamma_n : [0, 1] \rightarrow M \mid n \in \mathbb{N}\}$ is pointwise bounded and equicontinuous. We can apply Arzelà–Ascoli Theorem to prove that up to extraction, the sequence (γ_n) uniformly converges to a limit curve $\gamma^* : [0, 1] \rightarrow M$. The curve γ^* is such that $\gamma^*(0) = x^*$, $\gamma^*(1/2) = z^*$ and $\gamma^*(1) = y^*$.

As a consequence, for instance, of [2, Corollary 3.42], the curve γ^* is an admissible length minimizing curve, thus proving that $(x^*, z^*) \in \Gamma^l(\mathcal{K})$. \square

The idea behind the introduction of the sets $\Gamma^l(\mathcal{K})$ and $\Gamma^r(\mathcal{K})$ is that Molchanov’s method separates heat kernels evaluated at pairs (x, y) in \mathcal{K} into products of heat kernels evaluated at pairs (x, z) and (z, y) with z in the neighborhood of $\Gamma(x, y)$. More can be said on such pairs using the following lemma.

Lemma 18. *Let \mathcal{K} be a compact subset of M^2 such that $\mathcal{K} \cap \mathcal{D} = \emptyset$ and only strongly normal minimizers exist between any pair $(x, y) \in \mathcal{K}$. For $\varepsilon > 0$, let*

$$\begin{aligned}\Gamma_\varepsilon^l(\mathcal{K}) &= \{(x, z) \mid \exists y \in M \text{ s.t. } (x, y) \in \mathcal{K}, z \in \Gamma_\varepsilon(x, y)\}, \\ \Gamma_\varepsilon^r(\mathcal{K}) &= \{(z, y) \mid \exists x \in M \text{ s.t. } (x, y) \in \mathcal{K}, z \in \Gamma_\varepsilon(x, y)\}.\end{aligned}$$

There exists $\varepsilon > 0$ such that

$$\Gamma_\varepsilon^l(\mathcal{K}) \cap \mathcal{C} = \emptyset \quad \text{and} \quad \Gamma_\varepsilon^r(\mathcal{K}) \cap \mathcal{C} = \emptyset.$$

Proof. We prove the statement by contradiction for $\Gamma_\varepsilon^l(\mathcal{K}) \cap \mathcal{C}$ only, the result then holds by symmetry. Assume that for all $\varepsilon > 0$, $\Gamma_\varepsilon^l(\mathcal{K}) \cap \mathcal{C} \neq \emptyset$. Then let $(x_n, y_n) \in \mathcal{K}$ be a minimizing sequence such that for all n , there exists $z_n \in M$ such that $d(x_n, z_n) < \frac{d(x_n, y_n) + 1/n}{2}$, $d(z_n, y_n) < \frac{d(x_n, y_n) + 1/n}{2}$ and $(x_n, z_n) \in \mathcal{C}$.

Since $(x_n, y_n) \in \mathcal{K}$, z_n belongs to a compact and, up to extraction, $x_n \rightarrow x^*$, $y_n \rightarrow y^*$, $z_n \rightarrow z^*$ with $(x^*, y^*) \in \mathcal{K}$. Furthermore, we have that $d(x^*, z^*) = d(z^*, y^*) = d(x^*, y^*)/2$, meaning that (x^*, z^*) belongs to $\Gamma^l(\mathcal{K})$. By closure of \mathcal{C} (see, e.g., [2, Proposition 8.76]), that implies that $\mathcal{C} \cap \Gamma^l(\mathcal{K}) \neq \emptyset$, which has been excluded by Lemma 17. \square

Let us introduce some useful compact sets of $M^2 \setminus \mathcal{C}$ and prove some of their properties relying on the same idea as Lemma 18. These sets are quite useful for the proof of Ben Arous expansions. For any compact set $\mathcal{K} \in M^2 \setminus \mathcal{C}$, for any $\varepsilon > 0, \eta > 0$, the set $\mathcal{U}_{\varepsilon, \eta}$ is the set of points $(x, y) \in M^2$ such that $d(x, y) \geq \eta$ and such that x and y both belong to a (closed) ε -tubular neighborhood of the same geodesic linking two points in \mathcal{K} .

These sets have the nice property that they remain compact while containing \mathcal{K} , and, more importantly, that the midpoint sets generated with the sets $\mathcal{U}_{\varepsilon, \eta}$ themselves are contained within another set of the same family (assuming ε is small enough). This is crucial to be able to apply Molchanov's method repeatedly.

Lemma 19. *Let \mathcal{K} be a compact subset of $M^2 \setminus \mathcal{C}$. For any $\varepsilon > 0, \eta > 0$, let $\mathcal{U}_{\varepsilon, \eta} \subset M^2$ be such that $(x, y) \in \mathcal{U}_{\varepsilon, \eta}$ if $d(x, y) \geq \eta$ and there exists $(x', y') \in \mathcal{K}$, with unique, strongly normal and non-conjugate minimizer $\gamma : [0, 1] \rightarrow M$, $\gamma(0) = x'$, $\gamma(1) = y'$, $s, t \in [0, 1]$ such that $d(x, \gamma(s)) \leq \varepsilon$, $d(y, \gamma(t)) \leq \varepsilon$.*

For ε small enough, $\mathcal{U}_{\varepsilon, \eta} \cap \mathcal{C} = \emptyset$. Furthermore, for any $\varepsilon_0 > 0, \eta > 0$, there exists $\varepsilon > 0$ such that

$$(4) \quad \Gamma_\varepsilon^l(\mathcal{U}_{\varepsilon, \eta}) \subset \mathcal{U}_{\varepsilon_0, \eta/3} \quad \text{and} \quad \Gamma_\varepsilon^r(\mathcal{U}_{\varepsilon, \eta}) \subset \mathcal{U}_{\varepsilon_0, \eta/3}.$$

Proof. Without loss of generality, we can assume η small enough to not have to account for the case $\mathcal{U}_{\varepsilon, \eta} = \emptyset$.

The idea to prove that $\mathcal{U}_{\varepsilon, \eta} \cap \mathcal{C} = \emptyset$ for $\varepsilon > 0$ is similar to the proof of Lemmas 17 and 18. The set $\mathcal{U}_{0, \eta}$, corresponds to a set of pairs of points that belong to the same strongly normal minimizer between pairs of points in \mathcal{K} , hence a subset of $M^2 \setminus \mathcal{C}$. It is also a compact subset of $M^2 \setminus \mathcal{C}$ by compactness of \mathcal{K} . Assuming that for all ε , $\mathcal{U}_{\varepsilon, \eta} \cap \mathcal{C} \neq \emptyset$ implies that $\mathcal{U}_{0, \eta} \cap \bar{\mathcal{C}} \neq \emptyset$, which is false since $\bar{\mathcal{C}} = \mathcal{C}$.

Consider the sequence $(x_n, z_n) \in \Gamma_{1/n}^l(\mathcal{U}_{1/n, \eta})$, there exists $y_n \in M$ such that $(x_n, y_n) \in \mathcal{U}_{1/n, \eta} \subset M^2 \setminus \mathcal{C}$. Up to extraction, $(x_n, y_n) \rightarrow (x^*, y^*) \in \mathcal{U}_{0, \eta}$, and $z_n \rightarrow z^* \in \Gamma(x^*, y^*)$. Since $(x^*, y^*) \in \mathcal{U}_{0, \eta}$, $\Gamma(x^*, y^*)$ is reduced to a point on the geodesic supporting x^* and y^* , so that $(x^*, z^*) \in \mathcal{U}_{0, \eta/2}$.

Assuming that for all $n \in \mathbb{N}$, $\Gamma_{1/n}^l(\mathcal{U}_{1/n, \eta}) \setminus \mathcal{U}_{\varepsilon_0, \eta/3} \neq \emptyset$ thus implies that $\mathcal{U}_{0, \eta/2} \cap \partial \mathcal{U}_{\varepsilon_0, \eta/3} \neq \emptyset$, which is false by definition. This proves (4). \square

3.2. Uniform Ben Arous expansions. Expansions of the heat kernel, as given by Ben Arous in [19], hold for sub-Riemannian distribution over \mathbb{R}^d . This means that we are able to write these expansions for pairs of points in a manifold as long as the points are close enough to appear in the domain of the same map. For pairs of points that are further apart, Molchanov's method and localization naturally shows that almost all information can be gathered from the heat kernel between the pair and small neighborhoods of the midpoints. Of course the neighborhoods are approximately halfway between the two original points. Using Laplace integral asymptotics to derive Ben Arous expansions from the integral, we are effectively increasing the possible distance between pairs points by a fixed rate (slightly

smaller than 2). Repeating this argument, we are able to prove that Ben Arous expansions hold for points arbitrarily far apart. Using compactness arguments, this allows to prove that the expansions hold uniformly on compact sets of the manifold.

The announced Theorem 3 can then be expressed as follows.

Proposition 20. *If \mathcal{K} is a compact set in $M^2 \setminus \mathcal{C}$, then Ben Arous expansions hold uniformly on \mathcal{K} . As a consequence, Theorem 3 holds.*

This statement is obtained as a consequence of three lemmas. First, the Ben Arous expansions hold uniformly for points that are close enough to each other as a consequence of the original Ben Arous expansion theorem. Second, we use Molchanov's technique to show that if the statement holds for pairs of points sufficiently close in a compact, then we can increase this maximal distance by shaving off an arbitrarily small neighborhood of the border. Finally, we tie things up by completing the statement on the derivatives of the remainders. The proof of Proposition 20 comes at the end of the section, as a conclusion of this sequence of lemmas.

Lemma 21. *Let $\mathcal{K} \subset M^2$ be a compact. There exists $\delta_0 > 0$ such that Ben Arous expansions hold uniformly on any compact set in $\mathcal{K}' \cap \mathcal{D}(\delta_0) \setminus \mathcal{C}$.*

Proof. Let $K \subset M$ be the compact set of points such that $x \in K$ if there exists $y \in K$ such that either $(x, y) \in \mathcal{K}$ or $(y, x) \in \mathcal{K}$.

For all $x \in K$, there exist $R_x > 0$ and an isometry $\zeta_x : B(x, R_x) \rightarrow \mathbb{R}^d$ that maps $B(x, R_x)$ to a neighborhood of 0 in \mathbb{R}^d . The family $(B(x, R_x/4))_{x \in K}$ is an open cover of K , so we can extract a finite collection $(x_i)_{1 \leq i \leq n}$, $R_i = R_{x_i}$, such that $K \subset \cup_{i=1}^n \bar{B}(x_i, R_i/4)$.

Let $\delta_0 = \min_i R_i/4$. For any $x \in K$ there exists an integer i , $1 \leq i \leq n$, such that $x \in B(x_i, R_i/4)$. For any $y \in M$ such that $d(x, y) \leq \delta_0$, $y \in B(x_i, R_i/2)$. Hence for all pairs $(x, y) \in K^2 \cap \{d \leq \delta_0\}$, there exists i , $1 \leq i \leq n$, such that $(x, y) \in \bar{B}(x_i, R_i/2) \subset B(x_i, R_i)$.

For all $1 \leq i \leq m$, let \tilde{p}_t^i denote the heat kernel on $\zeta_i(B(x_i, R_i))$. As a consequence of Theorem 15, there exists ε_i such that uniformly for all $(x, y) \in B(x_i, R_i/2) \cap \mathcal{D}(\delta_0)$,

$$p_t(x, y) = \tilde{p}_t^i(\zeta_i(x), \zeta_i(y)) + O\left(e^{-\frac{d(x, y)^2 + \varepsilon_i}{4t}}\right).$$

The same holds for all the time and spatial derivatives of p_t . Then for all $1 \leq i \leq m$, the Ben Arous expansion hold uniformly on $\bar{B}(x_i, R_i/2)^2 \cap \mathcal{D}(\delta_0)$ since they classically hold on the compact $\zeta_i(\bar{B}(x_i, R_i/2))^2$. By taking the maximum of the uniform bounds on each ball, and the shortest time intervals, we get that the Ben Arous expansion hold uniformly on any compact subset of $[\cup_{i=1}^m \bar{B}(x_i, R_i/2)]^2 \cap \mathcal{D}(\delta_0)$ that excludes \mathcal{C} . \square

We now prove that we can expand the domain on which Ben Arous expansions hold. However we only partially prove that fact at first; the bounds on the derivatives of the remainder will be proved in the next lemma. In the remainder of this section, we give the proofs assuming l is any non-negative integer. In the non-symmetric case, the results are given by taking $l = 0$. Indeed, this reflects the fact that there is no problem in taking derivatives of the Ben Arous expansion per se, the difficulties only arise due to the lack of a sufficient version of the localization results and Léandre asymptotics, as manifested in the proof of Corollary 15.

Lemma 22. *Let \mathcal{K} be a compact subset of $M^2 \setminus \mathcal{C}$. For any $\varepsilon > 0$, $\eta > 0$, let $\mathcal{U}_{\varepsilon, \eta}$ be as defined in Lemma 19, so that in particular $\mathcal{K} \subset \mathcal{U}_{\varepsilon, \eta} \subset M^2 \setminus \mathcal{C}$. We introduce a partial statement of Ben Arous expansions, $P(\varepsilon, \eta, \delta)$.*

$P(\varepsilon, \eta, \delta)$: Let $\mathcal{U} = \mathcal{U}_{\varepsilon, \eta} \cap \mathcal{D}(\delta)$ and let $\mathcal{O} \subset M^2 \setminus \mathcal{C}$ be an open neighborhood of \mathcal{U} . For all non-negative integer l and multi-index α , there exist sequences of smooth functions $c_k^{l, \alpha} : \mathcal{O} \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, $r_k^{l, \alpha} : \mathbb{R}^+ \times \mathcal{O} \rightarrow \mathbb{R}$, such that for all $n \in \mathbb{N}$, for all $(x, y) \in \mathcal{O}$, for all t small enough

$$\partial_t^l Z_y^\alpha p_t(x, y) = t^{-(|\alpha|+2l+d/2)} e^{-\frac{d(x, y)^2}{4t}} \left(\sum_{k=0}^n c_k^{l, \alpha}(x, y) t^k + t^{n+1} r_{n+1}^{l, \alpha}(t, x, y) \right),$$

and, furthermore, there exists t_0 such that

$$\sup_{0 < t < t_0} \sup_{(x, y) \in \mathcal{U}} \left| r_{n+1}^{l, \alpha}(t, x, y) \right| < \infty.$$

If there exists $\delta > 0$, ε_0 , η_0 such that $P(\varepsilon_0, \eta_0, \delta)$ holds true, then there exists $\varepsilon > 0$ such that $P(\varepsilon, 3\eta_0, 3\delta/2)$ also holds.

Proof. Step 1: localization. Let l be a non-negative integer, α be a multi-index. Let $(t, x, y) \in \mathbb{R}^+ \times \mathcal{U}_{\varepsilon_0, \eta_0}$. As a consequence of Léandre estimates, for any $\varepsilon > 0$ small enough, we have uniformly on $\mathcal{U}_{\varepsilon_0, \eta_0}$

$$\begin{aligned} \partial_t^l Z_y^\alpha p_t(x, y) &= \int_M \partial_t^l (p_{t/2}(x, z) Z_y^\alpha p_{t/2}(z, y)) d\mu(z) \\ &= \int_{\Gamma_\varepsilon} \partial_t^l (p_{t/2}(x, z) Z_y^\alpha p_{t/2}(z, y)) d\mu(z) + O\left(e^{-\frac{d(x, y)^2 + \varepsilon}{4t}}\right) \\ &= \sum_{j=0}^l \binom{l}{j} \int_{\Gamma_\varepsilon} \left(\partial_t^j p_{t/2}(x, z) \right) \left(\partial_t^{l-j} Z_y^\alpha p_{t/2}(z, y) \right) d\mu(z) + O\left(e^{-\frac{d(x, y)^2 + \varepsilon}{4t}}\right). \end{aligned}$$

Step 2: Ben Arous expansions on the midpoint set. By application of Lemma 19, there exists $\varepsilon > 0$ such that

$$\Gamma_\varepsilon^l(\mathcal{U}_{\varepsilon, 3\eta_0}) \subset \mathcal{U}_{\varepsilon_0, \eta_0} \quad \text{and} \quad \Gamma_\varepsilon^r(\mathcal{U}_{\varepsilon, 3\eta_0}) \subset \mathcal{U}_{\varepsilon_0, \eta_0}.$$

In the following, we denote $\mathcal{U} = \mathcal{U}_{\varepsilon, 3\eta_0} \cap \mathcal{D}(3\delta/2)$ and $\mathcal{U}' = \mathcal{U}_{\varepsilon_0, \eta_0} \cap \mathcal{D}(\delta)$. As long as $\varepsilon < \delta/4$, we still have

$$\Gamma_\varepsilon^l(\mathcal{U}) \subset \mathcal{U}' \quad \text{and} \quad \Gamma_\varepsilon^r(\mathcal{U}) \subset \mathcal{U}',$$

and we assume that $P(\varepsilon_0, \eta_0, \delta)$ holds on \mathcal{U}' . Furthermore, with \mathcal{O} the interior of $\mathcal{U}_{\varepsilon+\zeta, 3\eta_0-\zeta} \cap \mathcal{D}(3\delta/2+\zeta)$, we have for ζ small enough that $\mathcal{O} \subset M^2 \setminus \mathcal{C}$, $\mathcal{U} \subset \mathcal{O}$ and $\Gamma_\varepsilon^l(\mathcal{O}) \subset \mathcal{U}'$. We will now prove that $P(\varepsilon, 3\eta_0, 3\delta/2)$ holds for \mathcal{U} and \mathcal{O} .

Since $P(\varepsilon_0, \eta_0, \delta)$ holds, there exists an open set $\mathcal{O}' \subset M^2 \setminus \mathcal{C}$, $\mathcal{U}' \subset \mathcal{O}'$, and smooth functions $c_k^{l, \alpha} : \mathcal{O}' \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, $r_k^{l, \alpha} : \mathbb{R}^+ \times \mathcal{O}' \rightarrow \mathbb{R}$, satisfying $P(\varepsilon_0, \eta_0, \delta)$. For all $(x, y) \in \mathcal{O}'$, for all $t \in \mathbb{R}^+$, we denote

$$\Sigma_t^{l, \alpha}(x, y) = \sum_{k=0}^n c_k^{l, \alpha}(x, y) t^k + t^{n+1} r_{n+1}^{l, \alpha}(t, x, y).$$

There also exists t_0 such that

$$\sup_{0 < t < t_0} \sup_{(x, y) \in \mathcal{U}'} \left| r_{n+1}^{l, \alpha}(t, x, y) \right| < \infty.$$

Then, by construction, it uniformly holds for all $(x, y) \in \mathcal{O}$ that

$$\begin{aligned} (5) \quad t^{d/2+2l+|\alpha|} e^{\frac{d(x, y)^2}{4t}} \partial_t^l \partial_y^\alpha p_t(x, y) &= \\ &= \frac{2^{d+2l+|\alpha|}}{t^{d/2}} \sum_{j=0}^l \binom{l}{j} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x, y}(z) - d(x, y)^2/4}{t}} \Sigma_{t/2}^{j, 0}(x, z) \Sigma_{t/2}^{l-j, \alpha}(z, y) d\mu(z) + O\left(e^{-\frac{\varepsilon}{4t}}\right) \end{aligned}$$

with $h_{x, y}(z) = (d(x, z)^2 + d(z, y)^2)/2$.

Step 3: Cauchy product rearrangement of expansions. Each term in sum (5) will satisfy the same uniformity property, hence to alleviate the notations we will focus in the rest of the proof on one

specific index j , and write, for integers $0 \leq i \leq n$,

$$a_i = c_i^{j,0}, \quad b_i = c_i^{l-j,\alpha}, \quad r^a = r_{n+1}^{j,0}, \quad r^b = r_{n+1}^{l-j,\alpha}.$$

By rearranging terms in the sums, we have

$$(6) \quad \Sigma_{t/2}^{l,0}(x, z) \Sigma_{t/2}^{l-j,\alpha}(z, y) = \sum_{k=0}^n \left(\frac{t}{2}\right)^k \left(\sum_{i=0}^k a_i(x, z) b_{k-i}(z, y) \right) + \left(\frac{t}{2}\right)^{n+1} \Phi_{n+1}(t, x, y, z),$$

with the explicit remainder

$$\begin{aligned} \Phi_{n+1}(t, x, y, z) &= r^a(t/2, x, z) r^b(t/2, z, y) + \\ &\sum_{k=1}^{n+1} \left(\frac{t}{2}\right)^k \left(r^a(t/2, x, z) b_{n+1-k}(z, y) + \sum_{i=1}^{k-1} \left(a_{n+1-i}(x, z) b_{n+1-k+i}(z, y) \right) + a_{n+1-k}(x, z) r^b(t, z, y) \right). \end{aligned}$$

Step 4: Laplace integrals asymptotics. We can now compute asymptotic expansions of Laplace integrals, following [24]. We only consider pairs $(x, y) \in \mathcal{O} \subset M^2 \setminus \mathcal{C}$. For such pairs, Γ is reduced to a single point z_0 . Furthermore $h_{x,y}(z_0) = d(x, y)/4$ is the minimum of $h_{x,y}$ and $(x, y) \mapsto z_0(x, y)$ is smooth.

Let $\xi : M \rightarrow \mathbb{R}^d$ be a set of local coordinates on Γ_ε such that $h_{x,y}(z) - h_{x,y}(z_0) = |\xi|^2(z)$, with z_0 the minimum of $h_{x,y}$. (Notice that by smoothness of $h_{x,y}$ on $M^2 \setminus \mathcal{C}$, the family of local coordinates smoothly depends on (x, y) .) For any smooth function φ supported in Γ_ε ,

$$\int_M e^{-\frac{h_{x,y}(z) - h_{x,y}(z_0)}{t}} \varphi(z) d\mu(z) = \int_M e^{-|\xi|^2(z)/t} \varphi(z) d\mu(z) = \int_{\mathbb{R}^d} e^{-|u|^2/t} \frac{\varphi(\xi^{-1}(u))}{\det D\xi(\xi^{-1}(u))} du.$$

From [24, Equation (4.36)] and recognizing that $2^{d/2} \det D\xi(\xi^{-1}(u)) = \sqrt{\det \text{Hess}(h_{x,y}(\xi^{-1}(u)))}$, we get

$$(7) \quad \int_M e^{-\frac{h_{x,y}(z) - h_{x,y}(z_0)}{t}} \varphi(z) d\mu(z) \simeq \sum_{N=0}^{\infty} \frac{(2\pi t)^{d/2} t^N}{2^{2N}} \sum_{|\omega|=N} \frac{1}{\omega!} \partial^{2\omega} \Big|_{u=0} \frac{\varphi(\xi^{-1}(u))}{\sqrt{\det \text{Hess}(h_{x,y}(\xi^{-1}(u)))}}$$

where $\omega \in \mathbb{N}^d$ is a multi-index $(\omega_1, \dots, \omega_d)$ such that $|\omega| = \sum_{i=1}^d \omega_i$, $2\omega = (2\omega_1, \dots, 2\omega_d)$ and $\omega! = \prod_{i=1}^d \omega_i!$, and for $f : \mathbb{R} \rightarrow \mathbb{R}$ we write

$$f(t) \simeq \sum_{n=0}^{\infty} t^n f_n$$

if $f(t) = \sum_{n=0}^N t^n f_n + O(t^{N+1})$ for all $N > 0$. In particular if for all x in an open domain $f(t, x) \simeq \sum_{n=0}^{\infty} t^n f_n(x)$ and $f_n(x)$ is smooth for all n then f is actually smooth at $(0, x)$ (on the right in t), hence uniformity of the remainders. Notice also that Equation (7) holds for smooth compactly supported functions on \mathbb{R}^d but only the jets at z_0 appear in the expansion, hence the asymptotic expansion holds for evaluating integrals on Γ_ε .

As announced in step 2, we now prove that the elements we exhibited as building blocks of the expansion are indeed smooth on \mathcal{O} , and uniformly bounded in time and space on the compact \mathcal{U} .

Step 5: Remainder. First, we consider the remainder

$$\Psi_{n+1}(t, x, y) = t^{-d/2} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - h_{x,y}(z_0)}{t}} \Phi_{n+1}(t, x, y, z) d\mu(z).$$

It is a smooth function on $\mathbb{R}^+ \times \mathcal{O}$. Let us prove that it is uniformly bounded on \mathcal{U} .

As a consequence of the discussion in step 2, if $z \in \Gamma_\varepsilon(x, y)$, then $(x, z) \in \Gamma_\varepsilon^l(\mathcal{U})$ and $(z, y) \in \Gamma_\varepsilon^r(\mathcal{U})$, both subsets of \mathcal{U}' . Hence, as a consequence of $\mathbf{P}(\varepsilon_0, \eta_0, \delta)$, there exists t_{n+1} such that

$$\sup_{0 < t < t_{n+1}} \sup_{(x, z) \in \Gamma_\varepsilon^l(\mathcal{U})} |r^a(t, x, y)| < \infty \quad \text{and} \quad \sup_{0 < t < t_{n+1}} \sup_{(z, y) \in \Gamma_\varepsilon^r(\mathcal{U})} |r^b(t, x, y)| < \infty.$$

This implies that

$$\sup \{ |\Phi_{n+1}(t, x, y, z)| \mid t \in (0, t_{n+1}), (x, y) \in \mathcal{U}, z \in \Gamma_\varepsilon(x, y) \} < A < \infty.$$

Then, applying (7), we get that for all $(x, y) \in \mathcal{U}$, for all $0 < t < t_{n+1}$,

$$|\Psi_{n+1}(t, x, y)| \leq t^{-d/2} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - h_{x,y}(z_0)}{t}} A d\mu(z) = (2\pi)^{d/2} A + O(t)$$

Hence the boundedness of Ψ_{n+1} .

Step 6: summands. For $k \in \{0, \dots, n\}$, for $N \in \{k, \dots, n\}$, we denote by ψ_N^k the smooth function on \mathcal{O} such that

$$\psi_N^k(x, y) = \frac{(2\pi)^{d/2}}{2^{2(N-k)}} \sum_{|\omega|=N-k} \frac{2^{d/2}}{\omega!} \partial^{2\omega} \Big|_{z=0} \left(\frac{\sum_{i=0}^k a_i(x, \xi^{-1}(z)) a_{k-i}(\xi^{-1}(z), y)}{\sqrt{\det \text{Hess}(h_{x,y}(\xi^{-1}(z)))}} \right).$$

Following Equation (7), for all $(t, x, y) \in \mathbb{R}^+ \times \mathcal{O}$,

$$t^{k-d/2} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - h_{x,y}(z_0)}{t}} \left(\sum_{i=0}^k a_i(x, z) a_{k-i}(z, y) \right) dz = \sum_{N=k}^n t^N \psi_N^k(x, y) + t^{n+1} \Psi_k(t, x, y)$$

where $\Psi_k(t, x, y)$ is a smooth function on $\mathbb{R}^+ \times \mathcal{O}$ and there exists $t_k > 0$ such that

$$\sup_{0 < t < t_k} \sup_{(x,y) \in \mathcal{U}} |\Psi_k(t, x, y)| < \infty.$$

Then, plugging these sums in (6) yields

$$t^{-d/2} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - d(x,y)^2/4}{t}} \Sigma_{t/2}^{j,0}(x, z) \Sigma_{t/2}^{l-j,\alpha}(z, y) dz = \sum_{N=0}^n t^N \left(\sum_{k=0}^N \psi_N^k(x, y) \right) + t^{n+1} \sum_{k=0}^{n+1} \Psi_k(t, x, y).$$

By construction, the remainder $\sum_{k=0}^{n+1} \Psi_k(t, x, y)$ is uniform.

Applying this reasoning for each term in the sum (5) yields that $P(\varepsilon, 3\eta_0, 3\delta/2)$ holds. \square

Lemma 23. *If $P(\varepsilon, \eta, \delta)$ holds then all the derivatives in (t, y) of the remainders are also uniformly bounded.*

Proof. Let $\mathcal{U} = U_{\varepsilon, \eta} \cap \mathcal{D}(\delta)$, $\mathcal{O} \subset M^2 \setminus \mathcal{C}$ be an open neighborhood of \mathcal{U} and let $\psi(t, x, y) : \mathbb{R}^+ \times \mathcal{O} \rightarrow \mathbb{R}$ be such that there exist sequences of smooth functions $a_k : \mathcal{O} \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, $\rho_k : \mathbb{R}^+ \times \mathcal{O} \rightarrow \mathbb{R}$, such that

$$\psi(t, x, y) = t^{-d/2} e^{-\frac{d(x,y)^2}{4t}} \left(\sum_{k=0}^n a_k(x, y) t^k + t^{n+1} \rho_{n+1}(t, x, y) \right),$$

and there exists $t_0 > 0$ such that $n \in \mathbb{N}$,

$$\sup_{0 < t < t_0} \sup_{(x,y) \in \mathcal{U}} |\rho_n(t, x, y)| < \infty.$$

Assume there also exist sequences of smooth functions $b_k : \mathcal{O} \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, $\bar{\rho}_k : \mathbb{R}^+ \times \mathcal{O} \rightarrow \mathbb{R}$, such that

$$Z_y^i \psi(t, x, y) = t^{-d/2-1} e^{-\frac{d(x,y)^2}{4t}} \left(\sum_{k=0}^n b_k(x, y) t^k + t^{n+1} \bar{\rho}_{n+1}(t, x, y) \right)$$

and for all $n \in \mathbb{N}$,

$$\sup_{0 < t < t_0} \sup_{(x,y) \in \mathcal{U}} |\bar{\rho}_n(t, x, y)| < \infty.$$

Then

$$\sup_{0 < t < t_0} \sup_{(x,y) \in \mathcal{U}} |Z_y^i \rho_n(t, x, y)| < \infty.$$

Indeed, we have

$$Z_y^i \left(t^{d/2} e^{\frac{d(x,y)^2}{4t}} \psi(t, x, y) \right) = \sum_{k=0}^n t^k Z_y^i a_k(x, y) + t^{n+1} Z_y^i \rho_{n+1}(t, x, y).$$

On the other hand, by pushing the expansion to one more order,

$$\begin{aligned} Z_y^i \left(t^{d/2} e^{\frac{d(x,y)^2}{4t}} \psi(t, x, y) \right) &= t^{d/2} e^{\frac{d(x,y)^2}{4t}} \left(\frac{Z_y^i d(x, y)^2}{4t} \psi(t, x, y) + Z_y^i \psi(t, x, y) \right) \\ &= \sum_{k=0}^{n+1} t^{k-1} \left(\frac{Z_y^i d(x, y)^2}{4} a_k(x, y) + b_k(x, y) \right) \\ &\quad + t^{n+1} \left(\frac{Z_y^i d(x, y)^2}{4} \rho_{n+2}(t, x, y) + \bar{\rho}_{n+2}(t, x, y) \right) \end{aligned}$$

Other than compatibility conditions such as $a_0(x, y) Z_y^i d(x, y)^2 = -4b_0(x, y)$, we have

$$Z_y^i \rho_{n+1}(t, x, y) = \frac{Z_y^i d(x, y)^2}{4} \rho_{n+2}(t, x, y) + \bar{\rho}_{n+2}(t, x, y)$$

Similar expressions can be derived for derivatives with respect to t following the same reasoning. Chaining these arguments for both variables in all orders implies the statement. \square

As a conclusion to the section, we can finally prove Proposition 20.

Proof of Proposition 20. Let $\bar{\eta} = 1/2 \min_{x,y \in \mathcal{K}} d(x, y)$, and let $\bar{\delta} = 2 \max_{x,y \in \mathcal{K}} d(x, y)$. We prove that there exists $\varepsilon > 0$ such that $P(\varepsilon, \bar{\eta}, \bar{\delta})$ holds, with P introduced in Lemma 22. Once this is proved, Lemma 23 then implies Proposition 20.

Consider the set $\mathcal{U}_{1,0} \subset M^2$, the set of pairs (x, y) such that there exists a strongly normal length minimizing curve $\gamma : [0, 1] \rightarrow M$, with $(\gamma(0), \gamma(1)) \in \mathcal{K}$, and $d(\gamma, x) \leq 1$, $d(\gamma, y) \leq 1$. It is a compact set, hence Lemma 21 applies: there exists $\delta_0 > 0$ such that for any compact set contained in $\mathcal{U}_{1,0} \cap \mathcal{C}$, Ben Arous expansions hold uniformly. As a consequence, for any η , $P(1, \eta, \delta_0)$ holds.

Let $m \in \mathbb{N}$ be such that $(3/2)^m \delta_0 \geq \bar{\delta}$, and let $\eta_0 = \bar{\eta}/3^m$. We have that $P(1, \eta_0, \delta_0)$ holds. Applying Lemma 22 m times yields that there exists $\varepsilon > 0$ such that $P(\varepsilon, 3^m \eta_1, (3/2)^m \delta_0)$ holds. Consequently, $P(\varepsilon, \bar{\eta}, \bar{\delta})$ also holds and we have proved the statement.

If Ben Arous expansions hold uniformly for any $\mathcal{K} \in M^2 \setminus \mathcal{C}$, then Theorem 3. The issue is the existence and smoothness of the functions $c_k^{l,\alpha}$, $r_k^{l,\alpha}$ on the full set $M^2 \setminus \mathcal{C}$, for all $k, l \in \mathbb{N}$, α multi-index. However by filling $M^2 \setminus \mathcal{C}$ with compacts, since the functions give an expansion of the heat kernel, we finally get the statement. \square

3.3. Uniform universal bounds on the heat kernel. A natural application of Ben Arous expansions are a priori universal bounds on the heat kernel, that come as a direct consequence of Molchanov method in form of a Laplace integral. These are stated in Proposition 5, which we prove in a moment. In the case of sub-Riemannian manifolds, these estimates were initially proved in [12]. The approach for the proof is similar, however we extend this result by showing the existence of uniform bounds on compact subsets where no two distinct points are joined by abnormal minimizers. Lower bounds do not hold for spatial derivatives due to the necessity of non-vanishing terms in the Ben Arous expansion, which is only guaranteed for time derivatives of the kernel.

Proof of Proposition 5. By compactness of the sets involved, it is sufficient to show the estimates holds on all elements of a finite cover of \mathcal{K} . Most of the reasoning present in the proof focuses on proving the statement for elements $(x, y) \in \mathcal{K}$ such that x belongs to an open ball on which the cotangent bundle has been trivialized. Since we can cover the compact $\pi_x(\mathcal{K})$ with a finite number of neighborhoods on which the fiber bundle can be trivialized, this implies the statement.

For all $x \in M$, let us denote by $\text{Exp}_x : T_x^*M \rightarrow M$ the sub-Riemannian exponential (at time 1). Let $(x_0, y_0) \in \mathcal{K}$. There exists $\eta > 0$ such that TM can be trivialized on $B(x_0, 2\eta)$, that is $T^*M \simeq \mathbb{R}^d \times M$. We affix on \mathbb{R}^d a Euclidean structure $|\cdot|$.

Step 1: preliminary bounds on integrals over Γ_ε .

First, let us check that the set

$$V_{x_0}^\eta = \{(x, p) \in \bar{B}(x_0, \eta) \times \mathbb{R}^d \mid \exists y \in M \text{ s.t. } (x, y) \in \mathcal{K}, \text{Exp}_x(p) \in \Gamma_\varepsilon(x, y)\}.$$

is compact by showing that any sequence (x_n, p_n) in $V_{x_0}^\eta$ converges up to extraction.

By Lemmas 17 and 18, there exists $\varepsilon_0 > 0$ such that if $\varepsilon \leq \varepsilon_0$, $\Gamma_\varepsilon^l(\mathcal{K})$, is a compact subset of $M^2 \setminus \mathcal{C}$. This implies that up to extraction, $(x_n, \text{Exp}_{x_n}(p_n))$ converges to $(x^*, z^*) \in \Gamma_\varepsilon^l(\mathcal{K})$, that is, there exists $y^* \in M$ such $(x^*, y^*) \in \mathcal{K}$ and $z^* \in \Gamma_\varepsilon(x, y)$. Furthermore, $x^* \in \bar{B}(x_0, \eta)$ by closure of the set.

Let $\gamma_n(s) = \text{Exp}_{x_n}(sp_n)$. The family (γ_n) is a family of length-minimizing curves, it converges to a length minimizing curve that we denote γ^* . We also know that $\gamma^*(0) = x^*$ and $\gamma^*(1) = z^*$. Since there exists no abnormal minimizers, there exists p^* such that $\gamma^*(s) = \text{Exp}_{x^*}(sp^*)$.

This proves that $V_{x_0}^\eta$ is compact, and further implies that Exp_x is a diffeomorphism on the set of p such that $(x, p) \in V_{x_0}^\eta$ (since $(x, \text{Exp}_x(p))$ avoids \mathcal{C})

Let us denote by $\lambda_{\mathbb{R}^d}$ the Lebesgue measure associated to the Euclidean structure on the trivialized fibers of T^*M . For all $x \in \bar{B}(x_0, \eta)$, let V_x be the set of coverctors p such that $(x, p) \in V_{x_0}^\eta$. The exponential at x is a diffeomorphism from V_x onto its image, hence there exists a function ν_x that is the density of $(\text{Exp}_x^{-1})_*\mu$ with respect to the measure $\lambda_{\mathbb{R}^d}$ on V_x .

Furthermore, since the exponential is smooth in x and p over $V_{x_0}^\eta$, $(x, p) \rightarrow \nu_x(p)$ is at least continuous, which implies by compactness that

$$\underline{\nu} = \inf_{(x,p) \in V_{x_0}^\eta} \nu_x(p) > 0, \quad \text{and} \quad \bar{\nu} = \sup_{(x,p) \in V_{x_0}^\eta} \nu_x(p) < \infty.$$

Then for any smooth function $\varphi : M \rightarrow \mathbb{R}$, any $(x, y) \in \mathcal{K}$, $x \in \bar{B}(x_0, \eta)$,

$$(8) \quad \underline{\nu} \int_{V_x} \varphi(\text{Exp}_x(p)) \lambda_{\mathbb{R}^d}(dp) \leq \int_{\Gamma_\varepsilon} \varphi(z) \mu(dz) \leq \bar{\nu} \int_{V_x} \varphi(\text{Exp}_x(p)) \lambda_{\mathbb{R}^d}(dp).$$

Step 2: upper bounds on the kernel. As a consequence of uniform Léandre estimates, for all non-negative integer l and multi-index α , there exists ε_1 such that for all $\varepsilon < \varepsilon_1$, uniformly in $(x, y) \in \mathcal{K}$, we have

$$e^{\frac{d(x,y)^2}{4t}} \partial_t^l Z_y^\alpha p_t(x, y) = e^{\frac{d(x,y)^2}{4t}} \sum_{j=0}^l \binom{l}{j} \int_{\Gamma_\varepsilon} \left(\partial_t^j p_{t/2}(x, z) \right) \left(\partial_t^{l-j} Z_y^\alpha p_{t/2}(z, y) \right) \mu(dz) + O\left(e^{-\frac{\varepsilon}{4t}}\right).$$

Then we can apply Ben Arous expansion at the lowest order. Since $\Gamma_\varepsilon^l(\mathcal{K})$ and $\Gamma_\varepsilon^r(\mathcal{K})$ are both compact subsets of $M \setminus \mathcal{C}$, there exist $\bar{c} > 0$, $t_{l,\alpha} > 0$ such that for all $0 < t < t_{l,\alpha}$, for all $(x, y) \in \mathcal{K}$, for all $z \in \Gamma_\varepsilon$, for all $j \in \mathbb{N}$, $j \leq l$,

$$\left| \partial_t^j p_{t/2}(x, z) \right| \leq \bar{c} \left(\frac{t}{2} \right)^{-\frac{d}{2}-2j} e^{-\frac{d(x,z)^2}{2t}},$$

$$\left| \partial_t^{l-j} Z_y^\alpha p_{t/2}(z, y) \right| \leq \bar{c} \left(\frac{t}{2} \right)^{-\frac{d}{2}-2(l-j)-|\alpha|} e^{-\frac{d(z,y)^2}{2t}}.$$

Hence, uniformly in $(x, y) \in \mathcal{K}$, for all $\varepsilon < \varepsilon_1$,

$$e^{\frac{d(x,y)^2}{4t}} \partial_t^l Z_y^\alpha p_t(x, y) \leq \bar{c}^2 \frac{2^{d+3l+|\alpha|}}{t^{2l+|\alpha|} t^d} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - d(x,y)^2/4}{t}} \mu(dz) + O\left(e^{-\frac{\varepsilon}{4t}}\right).$$

Recall that for all points $z \in M$, $h_{x,y}(z) \geq \frac{d(x,y)^2}{4}$, the minimum being reached on points of Γ . Following [12], the triangular inequality implies

$$\begin{aligned} h_{x,y}(z) &= \frac{1}{2} (d(x,z)^2 + d(z,y)^2) \\ &\geq \frac{1}{2} (d(x,z)^2 + (d(x,y) - d(x,z))^2) \\ &\geq d(x,z)^2 - d(x,z)d(x,y) + \frac{d(x,y)^2}{2} \\ &\geq \left(d(x,z) - \frac{d(x,y)}{2} \right)^2 + \frac{d(x,y)^2}{4} \end{aligned}$$

With $H : TM \rightarrow \mathbb{R}$ the sub-Riemannian Hamiltonian. The homogeneity property of the Hamiltonian implies, for all positive s ,

$$H(x, sp) = s^2 H(x, p).$$

In particular, for any $z \in \Gamma_\varepsilon$, there exists a unique $p \in V_x$ such that $\text{Exp}_x(p) = z$ and we have $d(x, z) = \sqrt{2H(x, p)}$. Furthermore, with the map

$$\Phi_x : p \mapsto \left(\sqrt{2H(x, p)}, H \left(x, \frac{p}{\sqrt{2H(x, p)}} \right) \right)$$

we separate the space $\{H \neq 0\}$ into $\mathbb{R}^+ \times \{H = 1/2\}$.

It is clear that the inverse of Φ_x is simply

$$(\sigma, q) \mapsto \sigma q.$$

Hence, denoting $\lambda^{1/2}$ the trace of the Lebesgue measure on $\{H = 1/2\}$, for any smooth function $\varphi : T_x^*M \rightarrow \mathbb{R}$, and V a compact in $\{H \neq 0\}$, and

$$\int_V \varphi(p) \lambda(dp) = \int_{\Phi_x(V)} \varphi(\sigma q) \sigma^{d-1} d\sigma \lambda^{1/2}(dp) = \int_{\{H=1/2\}} \int_{\mathbb{R}^+} \mathbf{1}_{\Phi_x(V)}(\sigma, q) \varphi(\sigma q) \sigma^{d-1} d\sigma \lambda^{1/2}(dp).$$

By the lower bound on $h_{x,y}$, we have

$$\int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - d(x,y)^2/4}{t}} \mu(dz) \leq \int_{\Gamma_\varepsilon} e^{-\frac{1}{t} \left(d(x,z) - \frac{d(x,y)}{2} \right)^2} \mu(dz).$$

Using the upper bound (8),

$$\begin{aligned} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - d(x,y)^2/4}{t}} \mu(dz) &\leq \bar{\nu} \int_{V_x} e^{-\frac{1}{t} \left(\sqrt{2H(x,p)} - \frac{d(x,y)}{2} \right)^2} \lambda(dp) \\ &\leq \bar{\nu} \int_{\{H=1/2\}} \int_{\mathbb{R}^+} \mathbf{1}_{\Phi_x(V)}(\sigma, q) e^{-\frac{1}{t} \left(\sigma - \frac{d(x,y)}{2} \right)^2} \sigma^{d-1} d\sigma \lambda^{1/2}(dp) \end{aligned}$$

To conclude, we use again that $V_{x_0}^\eta$ is a compact set. By continuity,

$$\underline{\sigma} = \inf_{V_{x_0}^\eta} \sqrt{2H(x, p)} > 0 \quad \text{and} \quad \bar{\sigma} = \sup_{V_{x_0}^\eta} \sqrt{2H(x, p)} < \infty.$$

Also, $\underline{\sigma} < \frac{d(x,y)}{2} < \bar{\sigma}$. Furthermore, there exists $C > 0$ such that for all x such that $(x, p) \in V_{x_0}^\eta$,

$$\lambda^{1/2}(\{q \mid \sigma q \in V_x\}) < C.$$

Indeed, the opposite would imply that there exists a sequence $(p_n, x_n) \in V_{x_0}^\eta$ such that $\frac{p_n}{\sqrt{2H(x_n, p_n)}} \rightarrow$

∞ . Since $\sqrt{2H}$ is bounded on $V_{x_0}^\eta$ and $V_{x_0}^\eta$ is compact, this is impossible.

Thus,

$$\int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - d(x,y)^2/4}{t}} \mu(dz) \leq C \bar{\nu} \bar{\sigma}^{d-1} \int_{\underline{\sigma}}^{\bar{\sigma}} e^{-\frac{1}{t} \left(\sigma - \frac{d(x,y)}{2} \right)^2} d\sigma.$$

Then, as a classical application of Laplace integral asymptotics,

$$\int_{\bar{\sigma}}^{\sigma} e^{-\frac{1}{t}(\sigma - \frac{d(x,y)}{2})^2} d\sigma \leq 2\sqrt{2\pi t}.$$

Hence

$$e^{\frac{d(x,y)^2}{4t}} \partial_t^l Z_y^\alpha p_t(x, y) \leq (\bar{c}^2 + 1) \frac{2^{d+3l+|\alpha|+1} \sqrt{2\pi}}{t^{2l+|\alpha|} t^{d-1/2}} C \bar{\nu} \bar{\sigma}^{d-1}.$$

which proves the stated upper bound.

Step 3: lower bounds on the kernel. In that case we follow the same reasoning, but it takes less work to be able to give an upper bound of the hinged energy functional.

Notice there exists $\rho > 0$ such that for any $(x, y) \in \mathcal{K}$ such that $x \in \bar{B}(x_0, \eta)$, and p_0 such that $\text{Exp}_x(p_0) \in \Gamma$, $B(p_0, \rho) \subset V_x$. Then, for any for any smooth $\varphi : M \rightarrow \mathbb{R}$, and any such p_0 ,

$$\int_{\Gamma_\varepsilon} \varphi(z) \mu(dz) \geq \nu \int_{V_x} \varphi(\text{Exp}_x(p)) \lambda_{\mathbb{R}^d}(dp) \geq \nu \int_{B(p_0, \rho)} \varphi(\text{Exp}_x(p)) \lambda_{\mathbb{R}^d}(dp).$$

On $B(p_0, \rho)$, we can use a d -dimensional Taylor expansion,

$$|h_{x,y} \circ \text{Exp}_x(p) - h_{x,y} \circ \text{Exp}_x(p_0)| \leq \sup_{p_1 \in B(p_0, \rho)} \|D^2 h_{x,y} \circ \text{Exp}_x(p_1)\| |p - p_0|^2.$$

Again compactness implies that $\kappa = \sup \{\|D^2 h_{x,y} \circ \text{Exp}_x(p)\| \mid (x, p) \in V_{x_0}^\eta, (x, y) \in \mathcal{K}, p \in V_x\}$ is finite and thus

$$h_{x,y}(p) \leq \frac{d(x, y)^2}{4} + \kappa |p - p_0|^2.$$

Once again, by uniform Léandre estimates, there exists ε_1 such that for all $\varepsilon < \varepsilon_1$, uniformly in $(x, y) \in \mathcal{K}$,

$$e^{\frac{d(x,y)^2}{4t}} \partial_t^l p_t(x, y) = e^{\frac{d(x,y)^2}{4t}} \sum_{j=0}^l \binom{l}{j} \int_{\Gamma_\varepsilon} \left(\partial_t^j p_{t/2}(x, z) \right) \left(\partial_t^{l-j} p_{t/2}(z, y) \right) \mu(dz) + O(e^{-\frac{\varepsilon}{4t}}).$$

We apply Ben Arous expansions at the lowest order. Since we don't have any spatial derivatives, the first term in the expansion never vanishes outside of the cut locus. Hence, since $\Gamma_\varepsilon^l(\mathcal{K})$ and $\Gamma_\varepsilon^r(\mathcal{K})$ are compact subsets of $M \setminus \mathcal{C}$, there exist $\underline{c} > 0$, $t_0 > 0$ such that for all $0 < t < t_0$, for all $(x, y) \in \mathcal{K}$, for all $z \in \Gamma_\varepsilon$, for all integer $0 \leq j \leq l$,

$$\partial_t^j p_{t/2}(x, z) \partial_t^{l-j} p_{t/2}(z, y) \geq 2\underline{c} \left(\frac{t}{2}\right)^{-\frac{d}{2}-2j} e^{-\frac{d(x,z)^2}{2t}} \left(\frac{t}{2}\right)^{-\frac{d}{2}-2(l-j)} e^{-\frac{d(z,y)^2}{2t}}.$$

Consequently

$$\begin{aligned} e^{\frac{d(x,y)^2}{4t}} \partial_t^l p_t(x, y) &\geq \frac{2^{d+2l}\underline{c}}{t^{2l}t^d} \int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z) - d(x,y)^2/4}{t}} \mu(dz) \\ &\geq \frac{2^{d+2l}\underline{c}}{t^{2l}t^d} \int_{B(p_0, \rho)} e^{-\kappa \frac{|p-p_0|^2}{t}} dp. \end{aligned}$$

As a classical application of Laplace integrals asymptotics,

$$\int_{B(p_0, \rho)} e^{-\frac{|p-p_0|^2}{2t}} dp \geq \frac{(\pi t)^{d/2}}{2}.$$

Hence

$$e^{\frac{d(x,y)^2}{4t}} \partial_t^l p_t(x, y) \geq \frac{2^{2l+d}\pi^{d/2}\underline{c}}{t^{2l}t^{d/2}}.$$

Which concludes the proof. \square

4. COMPLETE EXPANSIONS

As we aim to illustrate in this section, applying the Molchanov method allows to translate information on the jets of the hinged energy functional on the midpoint set to complete expansions of the heat kernel and its derivatives, while simultaneously sidestepping heavier methods. Here we are able to give proofs for complete expansions for some well known singular cases: conjugate minimizing curves of type *A*, and Morse-Bott conjugacy.

One critical point is that the complete expansions draw information from jets of the hinged energy functional. We show in Theorem 26 that basically any smooth non-negative function can be realized as a hinged energy functional between two points of a Riemannian manifold. This points towards the idea that full expansions should not always be accessible.

4.1. A-type singularities. For some points in the cut locus, it is still possible to give a precise enough expansion of the heat kernel. In particular, we consider here the case where a pair of points $x, y \in M$ are connected by a unique geodesic that is conjugate.

If we assume that y is a singular value of Exp_x , the sub-Riemannian exponential at x , and, furthermore, that Exp_x has a A_n singularity, with $n > 0$, at a preimage of y , then n has to be odd for the normal extremal joining x to y to be minimizing (see Figure 2. Indeed in that case (see, e.g., [11]), the hinged energy functional has the normal form

$$(9) \quad h_{x,y}(z_1, \dots, z_n) = \frac{d(x, y)}{4} + z_d^{n+1} + \sum_{i=1}^{d-1} z_i^2.$$

This fact yields the following expansion.

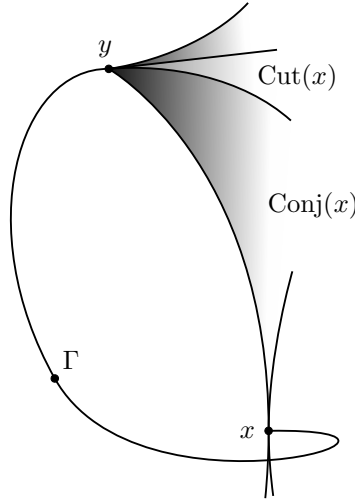


FIGURE 2. A minimizing conjugate curve typically appears at the boundary of the cut, where the sub-Riemannian exponential degenerates. A particular example where such a situation occurs correspond to points on a non-degenerate caustic of 3D contact sub-Riemannian manifolds. Indeed for generic x in such a manifold and y at the boundary of the cut (and at least sufficiently near x), the point y belongs to a cuspidal fold of the conjugate locus corresponding to an A_3 singularity (see, e.g., [1, 22]). The geodesic linking x and y is unique and Γ reduced to a point.

Proposition 24. *Let x and y be two points of a sub-Riemannian manifold such that the unique length minimizing curve joining x to y is strongly normal and a conjugate curve of type A_{2p-1} , $p \in \mathbb{N}$, $p \geq 2$. Then if l is any non-negative integer in the symmetric case and 0 otherwise, and α is any multi-index, there exists a sequence of real numbers $(\nu_k)_{k \in \mathbb{N}}$ and a sequence of functions $(\rho_k)_{k \in \mathbb{N}}$ such that for all $n \in \mathbb{N}$,*

$$\partial_t^l Z_y^\alpha p_t(x, y) = t^{-(|\alpha|+2l+\frac{d+1}{2})+\frac{1}{2p}} e^{-\frac{d(x,y)^2}{4t}} \left(\sum_{k=0}^n \nu_k t^{k/p} + t^{\frac{n+1}{p}} \rho_{n+1}(t) \right),$$

and there exists $t_0 > 0$ such that

$$\sup_{(0, t_0)} |\rho_{n+1}(t)| < \infty.$$

Proof. We give the proof in the case $l = 0$, $\alpha = 0$. The other cases can be straightforwardly deduced from this proof and the proof of Lemma 22, where the method by distribution of derivatives is explicated.

Under the assumptions of the theorem, the normal geodesic joining x and y is such that the hinged energy near the midpoint of the geodesic $z_0(x, y)$ is given in coordinates by $\frac{d(x, y)}{4} + \Phi \circ \xi$, with $\Phi(z) = z_d^{2p} + \sum_{i=1}^{d-1} z_i^2$ and ξ a local diffeomorphism centered at z_0 .

For any smooth function $\phi(z)$,

$$\int_{\mathbb{R}^d} e^{-\Phi \circ \xi(z)/t} \phi(z) dz = \int_{\mathbb{R}^d} e^{-\Phi(z)/t} \phi(\xi^{-1}(z)) \det(D\xi^{-1}(z)) dz.$$

Following [24, Equation (4.61)], computing Cauchy product of expansions in the z_d direction and the (z_1, \dots, z_{d-1}) directions,

$$e^{-\Phi(z)/t} \simeq \sum_{n=0}^{\infty} \sum_{k=0}^n \left(\frac{\Gamma\left(\frac{2k+1}{2p}\right) \partial_{z_d}^{2k} \delta_0}{p(2k)!} t^{\frac{2k+1}{2p}} \right) \left(\frac{\pi^{\frac{d-1}{2}}}{2^{2(n-k)}} t^{\frac{d-1}{2}+n-k} \sum_{\substack{\omega \in \mathbb{N}^{d-1} \\ |\omega|=n-k}} \frac{\partial_{(x_1, \dots, x_{d-1})}^{2\omega} \delta_0}{\omega!} \right)$$

We rearrange the terms so that the k index completes the multi-index $\omega \in \mathbb{N}^{d-1}$ such that $|\omega| = n-k$, into a $\omega' \in \mathbb{N}^d$ multi-index such that $|\omega'| = n$. Hence (dropping ω' in favor of ω again)

$$e^{-\Phi(z)/t} \simeq \sum_{n=0}^{\infty} \sum_{\substack{\omega \in \mathbb{N}^d \\ |\omega|=n}} \left(\frac{\pi^{\frac{d-1}{2}} \Gamma\left(\frac{2\omega_d+1}{2p}\right) \omega_d!}{2^{2(n-\omega_d)} p(2\omega_d)!} \right) t^{\frac{2\omega_d+1}{2p} + \frac{d-1}{2} + n - \omega_d} \frac{\partial_z^{2\omega} \delta_0}{\omega!}$$

As a consequence,

$$\int_{\mathbb{R}^d} e^{-\Phi \circ \xi(z)/t} \phi(z) dz \simeq \sum_{n=0}^{\infty} \sum_{\substack{\omega \in \mathbb{N}^d \\ |\omega|=n}} \left(\frac{\pi^{\frac{d-1}{2}} \Gamma\left(\frac{2\omega_d+1}{2p}\right) \omega_d!}{2^{2(n-\omega_d)} p(2\omega_d)! \omega!} \right) t^{\frac{2\omega_d+1}{2p} + \frac{d-1}{2} + n - \omega_d} \partial_z^2|_{z=0} [\phi(\xi^{-1}(z)) \det(D\xi^{-1}(z))].$$

Then we apply this statement to

$$e^{\frac{d(x,y)^2}{4t}} p_t(x, y) = \int_{\Gamma_\varepsilon} e^{\frac{d(x,y)^2}{4t}} p_{t/2}(x, z) p_{t/2}(z, y) dz + O(e^{-\frac{\varepsilon}{4t}}).$$

Similarly to the computations in the proof of Lemma 22, z being outside of the cut locus of both x and y , the heat kernel product decomposes as

$$p_{t/2}(x, z) p_{t/2}(z, y) = e^{-h_{x,y}(z)/t} \left(\frac{t}{2} \right)^{-d} \left[\sum_{k=0}^n \left(\frac{t}{2} \right)^k \sum_{j=0}^k c_j(x, z) c_{k-j}(z, y) + t^{n+1} R_n(t, x, y, z) \right],$$

with $z \mapsto c_k(x, z)$ and $z \mapsto c_k(z, y)$ smooth functions of z on Γ_ε , and $(t, z) \mapsto R_n(t, x, y, z)$ smooth function of (t, z) on $\mathbb{R}^+ \times \Gamma_\varepsilon$. Furthermore, there exists $t_0 > 0$ such that $\sup_{(0, t_0)} \sup_{\Gamma_\varepsilon} |R_n(t, x, y, z)| = A < \infty$.

Then, for all $N \in \mathbb{N}$, by noticing that $k + \frac{\omega_d}{p} + n - \omega_d \leq \frac{N}{p}$ if and only if $k \leq \lfloor N/p \rfloor$ and $n \leq N - kp$, since $0 \leq \omega_d \leq n$, we have

$$t^{\frac{d+1}{2} - \frac{1}{2p}} e^{\frac{d(x,y)^2}{4t}} p_t(x, y) = \sum_{k=0}^{\lfloor N/p \rfloor} \sum_{n=0}^{N-kp} \sum_{\substack{\omega \in \mathbb{N}^d \\ |\omega|=n}} \sigma_{k,n,\omega}(x, y) t^{k + \frac{\omega_d}{p} + n - \omega_d} + t^{\frac{N+1}{p}} \rho_{\frac{N+1}{p}}(t, x, y).$$

with

$$\sigma_{k,n,\omega}(x, y) = \left(\frac{\pi^{\frac{d-1}{2}} \Gamma\left(\frac{2\omega_d+1}{2p}\right) \omega_d!}{2^{2(n-\omega_d)+k-d} p (2\omega_d)! \omega!} \right) \sum_{j=0}^k \partial_z^{2\omega} \big|_{z=0} [c_j(x, \xi^{-1}(z)) c_{k-j}(\xi^{-1}(z), y) \det(D\xi^{-1}(z))]$$

and $\sup_{0, t_0} |\rho_{\frac{N+1}{p}}(t, x, y)| < \infty$. \square

4.2. Morse-Bott case. An interesting example of a point in the cut is the Morse-Bott case (so called because it corresponds to $h_{x,y}$ being a Morse-Bott function), where the set of geodesics between two points x and y is a continuous family, such that the midpoint set becomes a submanifold in M (of constant dimension) and the Hessian is non-degenerate in the normal directions. We follow [12, 11] for the definition of such a pair of points. All the elements necessary to give a full expansion of the heat kernel in that situation are present in [12] but that particular goal was not pursued. Here we apply Molchanov's method to obtain the full expansion. It should be noted that original methods for obtaining full expansions have been developed in [34] and [39, 40], where this particular example is treated. One observation that can be made from our technique is that, although these new approaches offer promising steps towards the construction of expansions of heat kernels in various situations, it doesn't appear necessary to introduce an original theory to treat this particular case in our sub-Riemannian situation.

Denoting $\Lambda_x = \{p \in T_x^* M \mid H(p, x) = 1/2\}$ and

$$L = \{p \in \Lambda_x \mid \text{Exp}_x(p, d(x, y)) = y\},$$

we assume that for a specific pair $(x, y) \in M^2$:

- All minimizers from x to y are strongly normal, hence given by the exponential map applies to elements of L ,
- L is a dimension r submanifold of Λ_x ,
- For every $p \in L$, we have $\dim \ker D_{(p, d(x, y))} \text{Exp}_x = r$.

(See, for instance, Figure 3 for an example of a such situation. See also [13] for another example where this type of cut points play an essential role.) Under these assumptions, Γ is a compact submanifold of M and it is proved in [12] that there exists a collection $(U_i)_{1 \leq i \leq N}$ of open sets such that for ε small enough, $\Gamma_\varepsilon \subset \cup_{i=1}^N U_i$, and on each U_i , there exists a set of coordinates $u : M \rightarrow \mathbb{R}^d$ such that for all $z \in U_i$

$$\Gamma \cap U_i = \{u_{r+1} = \dots = u_d = 0\}$$

and

$$h_{x,y}(z) = \frac{d(x,y)^2}{4} + \sum_{i=r+1}^d u_i^2.$$

Furthermore, there exists a partition of unity $(\varphi_i)_{1 \leq i \leq N}$, such that $\varphi_i|_{U_i^c} = 0$ and for all $z \in \Gamma_\varepsilon$,

$$\sum_{i=1}^N \varphi_i(z) = 1,$$

and on each U_i ,

$$\varphi_i(u_1, \dots, u_d) = \varphi_i(u_1, \dots, u_r, 0, \dots, 0).$$

In the described situation, we have the following.

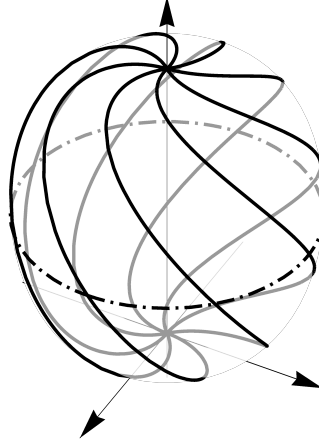


FIGURE 3. Any point of the cut locus in the Heisenberg group satisfies the definition of the Morse-Bott case. The set of all geodesics emanating from $(0, 0, 0)$ and becoming cut at some point $(0, 0, h)$, $h > 0$ form a sphere-like shape with rotational symmetry around the z -axis. When considering the pair $(0, 0, 0), (0, 0, h)$, the midpoint set is the equator of the sphere, a circle sitting at altitude $h/2$. Other examples include the Riemannian spheres of dimension at least 2.

Proposition 25. *For any l which is a non-negative integer in the symmetric case and 0 otherwise, and any multi-index α , there exists a sequence of real numbers $(\nu_k)_{k \in \mathbb{N}}$ and a sequence of functions $(\rho_k)_{k \in \mathbb{N}}$ such that for all $n \in \mathbb{N}$,*

$$\partial_t^l Z_y^\alpha p_t(x, y) = t^{-(|\alpha|+2l+\frac{d+r}{2})} e^{-\frac{d(x,y)^2}{4t}} \left(\sum_{k=0}^n \nu_k t^k + t^{n+1} \rho_{n+1}(t) \right),$$

and there exists $t_0 > 0$ such that

$$\sup_{(0, t_0)} |\rho_{n+1}(t)| < \infty.$$

Proof. The proof follows the same classical reasoning of multiplication of series, along with a Fubini theorem argument. Again, we give the proof for the case $\alpha = 0, l = 0$.

Then for $\varepsilon > 0$ small enough, we have the approximation

$$\begin{aligned} p_t(x, y) &= \int_M p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) \\ &= \int_{\Gamma_\varepsilon} p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) + O\left(e^{-\frac{d(x,y)^2 + \varepsilon}{4t}}\right) \end{aligned}$$

Then, applying our assumptions,

$$\begin{aligned} \int_{\Gamma_\varepsilon} p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) &= \int_{\Gamma_\varepsilon} \sum_{i=1}^N \varphi_i(z) p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) \\ &= \sum_{i=1}^N \int_{U_i} \varphi_i(z) p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) \end{aligned}$$

We can now use coordinates on each U_i so that (since $\varphi_i(z(u)) = 0$ outside of U_i)

$$\int_{U_i} \varphi_i(z) p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) = \int_{\Gamma \times (-\varepsilon, \varepsilon)^{d-r}} \varphi_i(u) p_{t/2}(x, z(u)) p_{t/2}(z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}}(u) du_1 \dots du_d.$$

As a consequence of Lemma 18, the set

$$\mathcal{K} = \{(x, z) \mid z \in \bar{\Gamma}_{\varepsilon'}\} \cup \{(z, y) \mid z \in \bar{\Gamma}_{\varepsilon'}\}$$

is a compact subset of $M^2 \setminus \mathcal{C}$ for ε' small enough. We can assume ε small enough so that the image of $\Gamma \times (-\varepsilon, \varepsilon)^{d-r}$ in the coordinate charts belongs to one such $\bar{\Gamma}_{\varepsilon'}$.

This implies that Theorem 3 holds for \mathcal{K} . Then there exists an open set $\mathcal{O} \subset M^2 \setminus \mathcal{C}$, $\mathcal{K} \subset \mathcal{O}$, and sequences of smooth functions $c_k : \mathcal{O} \rightarrow \mathbb{R}$, $r_k : \mathbb{R}^+ \times \mathcal{O} \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, such that for all $n \in \mathbb{N}$,

$$p_{t/2}(x, z)p_{t/2}(z, y) = e^{-\frac{h_{x,y}(z)}{4t}} \left(\frac{t}{2}\right)^{-d} \left(\sum_{k=1}^n c_k(x, z)t^k + t^{n+1}r_{n+1}(t, x, z)\right) \left(\sum_{k=1}^n c_k(z, y)t^k + t^{n+1}r_{n+1}(t, z, y)\right)$$

On one hand, as always,

$$\begin{aligned} \left(\sum_{k=0}^n c_k(x, z)t^k + t^{n+1}r_{n+1}(t, x, z)\right) \left(\sum_{k=0}^n c_k(z, y)t^k + t^{n+1}r_{n+1}(t, z, y)\right) = \\ \sum_{k=0}^n \psi_k(x, z, y)t^k + t^{n+1}R_{n+1}(t, x, z, y) \end{aligned}$$

with

$$\psi_k(x, z, y) = \sum_{j=0}^k c_j(x, z)c_{k-j}(z, y)$$

and, uniformly in space variables,

$$\sup_{t \in (0, t_0)} |R_{n+1}(t, x, z, y)| < \infty.$$

On the other hand, we have in coordinates

$$-\frac{h_{x,y}(z(u))}{4t} = -\frac{d(x, y)}{4t} - \frac{1}{t} \sum_{i=r+1}^d u_i^2.$$

Hence the separation

$$\begin{aligned} e^{\frac{d(x,y)}{4t}} \left(\frac{t}{2}\right)^d \int_{U_i} \varphi_i(z) p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) = \\ \sum_{k=0}^n t^k \int_{\Gamma \cap U_i} \varphi_i(u_1, \dots, u_r) \left(\int_{(-\varepsilon, \varepsilon)^{d-r}} e^{\frac{1}{t} \sum_{i=r+1}^d u_i^2} \psi_k(x, z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}}(u) du_{r+1} \dots du_d \right) du_1 \dots du_r \\ + t^{n+1} \int_{\Gamma \cap U_i} \varphi_i(u_1, \dots, u_r) \left(\int_{(-\varepsilon, \varepsilon)^{d-r}} e^{\frac{1}{t} \sum_{i=r+1}^d u_i^2} R_{n+1}(t, x, z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}}(u) du_{r+1} \dots du_d \right) du_1 \dots du_r \end{aligned}$$

As always, following [24], we have the expansion

$$e^{\frac{1}{t} \sum_{i=r+1}^d u_i^2} \simeq \sum_{j=0}^{\infty} \left(\frac{\pi^{\frac{d-r}{2}}}{2^{2j}} t^{\frac{d-r}{2}+j} \sum_{\substack{\alpha \in \mathbb{N}^{d-r} \\ |\alpha|=j}} \frac{\partial_{(u_r, \dots, u_d)}^{2\alpha} \delta_0}{\alpha!} \right),$$

so that

$$\begin{aligned} \int_{(-\varepsilon, \varepsilon)^{d-r}} e^{\frac{1}{t} \sum_{i=r+1}^d u_i^2} \psi_k(x, z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}}(u) du_{r+1} \dots du_d = \\ \pi^{\frac{d-r}{2}} t^{\frac{d-r}{2}} \sum_{j=0}^{n-k} \frac{t^j}{2^{2j}} \sum_{\substack{\alpha \in \mathbb{N}^{d-r} \\ |\alpha|=j}} \frac{\partial_{(u_r, \dots, u_d)}^{2\alpha}}{\alpha!} \left(\psi_k(x, z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}} \right) \Big|_{u_{r+1}=\dots=u_d=0} \\ + t^{n+1-k} \rho^k(t, x, y, u_1, \dots, u_r). \end{aligned}$$

Again here, uniformly in space variables,

$$\sup_{t \in (0, t_0)} |\rho^k(t, x, y, u_1, \dots, u_r)| < \infty.$$

Hence

$$\begin{aligned} t^k \int_{\Gamma \cap U_i} \varphi_i(u_1, \dots, u_r) \left(\int_{(-\varepsilon, \varepsilon)^{d-r}} e^{\frac{1}{t} \sum_{i=r+1}^d u_i^2} \psi_k(x, z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}}(u) du_{r+1} \dots du_d \right) du_1 \dots du_r = \\ \pi^{\frac{d-r}{2}} t^{\frac{d-r}{2}} \sum_{j=k}^n t^k \int_{\Gamma \cap U_i} \left(\varphi_i(u) \sum_{\substack{\alpha \in \mathbb{N}^{d-r} \\ |\alpha|=j}} \frac{\partial_{(u_r, \dots, u_d)}^{2\alpha}}{2^{2(j-k)} \alpha!} \left(\psi_k(x, z(u), y) \frac{d\mu}{d\lambda_{\mathbb{R}^d}} \right) \right) \Big|_{u_{r+1}=\dots=u_d=0} du_1 \dots du_r. \end{aligned}$$

By summing all elements together, this proves the statement. \square

4.3. Prescribed singularities. We wish to discuss situations where an explicit expansion of the Laplace integral for the small-time asymptotics of the heat kernel appears not be known. The first step is to show that essentially any possible phase function h for a Laplace integral can be realized as the hinged energy functional between two points of a manifold. We first restrict our attention to Riemannian metrics. As already noted, the Molchanov method doesn't distinguish between Riemannian and (properly) sub-Riemannian metrics, and it is simpler to give explicit constructions of Riemannian metrics.

Since the pair of points we consider will be fixed, in this section we use (q_1, q_2) rather than (x, y) to free the notation.

Theorem 26. *Let M be a smooth manifold of dimension d (with $d \geq 2$), q_1, q_2 in M such that $q_1 \neq q_2$, and a and σ be positive real numbers. Let h be a smooth, real-valued function in a neighborhood of $B_0^{d-1}(a) \subset \mathbb{R}^{d-1}$ such that $h(0, \dots, 0) = 0$, non-negative on $B_0^{d-1}(a)$, and positive on $\partial B_0^{d-1}(a)$. Then there exists a Riemannian metric g on M such that $\Gamma = \Gamma(x, y)$ is contained in a coordinate patch*

$$(u_1, \dots, u_d) : U \rightarrow B_0^{d-1}(a) \times (-\delta, \delta)$$

such that

$$h_{q_1, q_2}|_N = \frac{\sigma^2}{4} + h(u_1, \dots, u_{n-1}) + u_d^2$$

for some neighborhood N of Γ (thus Γ is given by the zero level set of h in the hyperplane $\{u_d = 0\}$, and $d(q_1, q_2) = \sigma$). Further, we have the heat kernel representation

$$(10) \quad p_t(q_1, q_2) = \frac{1}{t^d} e^{-\frac{d^2(q_1, q_2)}{4t}} \int_{B_0^{d-1}(a) \times (-\varepsilon, \varepsilon)} \Phi(t, u) e^{-\frac{h(u_1, \dots, u_{d-1}) + u_d^2}{t}} du_1 \dots du_d + O\left(e^{-\frac{d^2(q_1, q_2) + c}{4t}}\right).$$

for some positive ε and a smooth prefactor function Φ over $\mathbb{R}^+ \times B_0^{d-1}(a) \times (-\varepsilon, \varepsilon)$, smoothly extendable and positive at $t = 0$.

By rescaling, there is no loss of generality in assuming that the distance between q_1 and q_2 is prescribed to be 2, which is the same as prescribing $\sigma = 1$.

Let (z_1, \dots, z_d) be the standard Euclidean coordinates on \mathbb{R}^n , and let g_E be the Euclidean metric. We will identify q_1 with $(0, \dots, 0, 1)$, q_2 with $(0, \dots, 0, -1)$, and $B_0^{d-1}(a)$ with the corresponding subset of the hyperplane $\{z_d = 0\}$. (In particular, this will end up being compatible with the notation used in the theorem). We will use V^+ (respectively V^-) to denote a neighborhood of $\overline{B_0^{d-1}(a)}$ large enough to contain q_1 (respectively q_2), with further properties of V^- and V^+ to be specified later. If $V = V^- \cup V^+$, the main work of the proof is to construct a metric on V which gives a distance function with the desired properties.

Lemma 27. *Let ξ be a smooth non-negative function on a neighborhood of $\overline{B_0^{d-1}(a)}$, $a > 0$, everywhere less than $1/8$, with all of its derivatives bounded, and ξ bounded from below by a positive constant outside of $B_0^{d-1}(a)$. Under assumptions of Theorem 26, there exist V^+ a neighborhood of $\overline{B_0^{d-1}(a)} \times \{0\} \cup \{q_1\}$ and a (smooth) metric on V^+ such that the graph of ξ in $B_0^{d-1}(a) \times [0, 1/8]$ is a subset of the sphere of radius 1 around q_1 , with none of the minimal geodesics from q_1 to this graph conjugate, and such that the metric agrees with g_E on a neighborhood of*

$$\{z \in B_0^{d-1}(a) \times [0, 1/8] \mid 0 \leq z_d \leq \xi(z_1, \dots, z_{d-1})\}.$$

Proof. Let

$$G = \{z \in B_0^{d-1}(a+1) \times \mathbb{R} \mid z_d = \xi(z_1, \dots, z_{d-1})\}$$

denote the portion of the graph of ξ on a neighborhood of $B_0^{d-1}(a) \times [0, 1/8]$. The graph is a smooth hypersurface, and as a consequence of the bounded derivatives property, there is a tubular neighborhood U (of diameter $\eta > 0$) of G on which its normal lines do not develop focal (or conjugate) points, and any (smooth) coordinates on G extend to smooth coordinates on U by including the signed distance to G as the first coordinate. In what follows, we use the phrases “above G ” and “below G ” in the natural way to describe the regions on which z_d is larger or smaller than ξ , respectively, and similarly for other sets in place of G , when it makes sense.

Let $(\theta_1, \dots, \theta_{d-1})$ be coordinates on the (open) lower hemisphere of the unit tangent sphere at q_1 . They are assumed to be centered at the south pole. The lower hemisphere of the unit tangent sphere maps diffeomorphically to the hyperplane $\{z_d = 0\}$ by following the Euclidean rays from q_1 . Furthermore, lifting (z_1, \dots, z_{d-1}) to the graph gives coordinates on G . Thus, by composition, $(\theta_1, \dots, \theta_{d-1})$ gives coordinates on G , centered at the origin. Next, let ρ be the signed distance from G , with the sign chosen so that ρ is positive below G ; then $(\rho, \theta_1, \dots, \theta_{d-1})$ gives coordinates on U . For future use, let Θ denote the open subset of \mathbb{S}^{d-1} for which $(\theta_1, \dots, \theta_{d-1})$ gives coordinates on $G \cap (B_0^{d-1}(a) \times (-1/4, 1/4))$. Now if we write the Euclidean metric on U in these coordinates, it is given by the matrix

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \langle \partial_{\theta_1}, \partial_{\theta_1} \rangle_{g_E} & \cdots & \langle \partial_{\theta_{d-1}}, \partial_{\theta_1} \rangle_{g_E} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \langle \partial_{\theta_1}, \partial_{\theta_{d-1}} \rangle_{g_E} & \cdots & \langle \partial_{\theta_{d-1}}, \partial_{\theta_{d-1}} \rangle_{g_E} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & [\langle \partial_{\theta_i}, \partial_{\theta_j} \rangle_{g_E}]_{1 \leq i, j \leq d-1} \end{bmatrix},$$

where the last expression is understood in terms of the 1×1 and $(d-1) \times (d-1)$ diagonal block decomposition of this matrix. Note that the $\langle \partial_{\theta_i}, \partial_{\theta_j} \rangle_{g_E}$ are smooth, positive functions on U for all $1 \leq i, j \leq d-1$.

We can now describe the metric g on V^+ that we’re looking for. We will give g on part of V^+ including U and the region above U in polar coordinates around q_1 ; that is, we will specify the $\langle \partial_{\theta_i}, \partial_{\theta_j} \rangle_{g_E}$, which determines the metric (since all inner products with ∂_r are determined by the condition of being polar coordinates). In a ball around q_1 , of Euclidean radius $1/8$, let g agree with the Euclidean metric, so that the coordinate singularity at q_1 , which is $r = 0$, is the usual one from polar coordinates on \mathbb{R}^n and the metric is in fact smooth there. For $(\theta_1, \dots, \theta_{d-1}) \in \Theta$ and $r \in (1-\eta, 1+\eta)$, we let

$$\langle \partial_{\theta_i}, \partial_{\theta_j} \rangle_g(r, \theta_1, \dots, \theta_{d-1}) = \langle \partial_{\theta_i}, \partial_{\theta_j} \rangle_{g_E}(\rho = r-1, \theta_1, \dots, \theta_{d-1}).$$

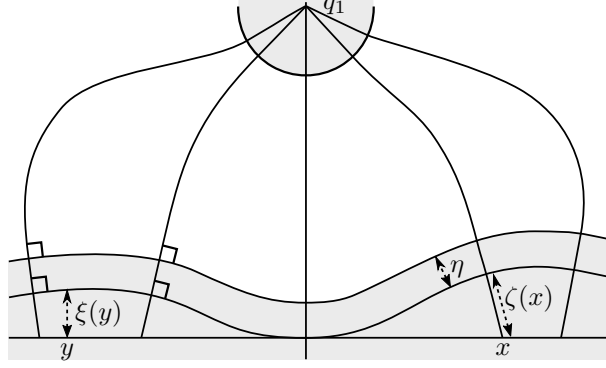


FIGURE 4. Schematic representation of the constructed geodesic front radiating from q_1 at time 1. Sections of the picture in grey correspond to regions where g agrees with g_E .

For $r \in (1/8, 1 - \eta)$ and $(\theta_1, \dots, \theta_{d-1}) \in \Theta$, we let $\langle \partial_{\theta_i}, \partial_{\theta_j} \rangle_g$ be some smooth, positive interpolation between the values we just fixed. This gives a metric on the part of V^+ including U and the region above U , but these coordinates may not extend to a neighborhood of $B_0^{d-1}(a)$ (because U might lie above $\{z_d = 0\}$ away from the zeroes of u). Nonetheless, if we put the Euclidean metric on the region below U , then the metric extends, because the transition map from the polar coordinates $(r, \theta_1, \dots, \theta_{d-1})$ to the Cartesian coordinates (z_1, \dots, z_d) is an isometry by construction.

In particular, we have given a metric g on a neighborhood of the origin in $T_{q_1}M$ (which includes $(0, 1 + \eta) \times \Theta$) along with an isometry from a subset of that neighborhood to a neighborhood of $B_0^{d-1}(a) \subset \mathbb{R}^n$ (with the Euclidean metric) that includes U , such that $G \cap (B_0^{d-1}(a) \times (-1/4, 1/4))$ is an open subset of the g -sphere of radius 1 around q_1 . Moreover, we did this by deforming the metric in between a small ball around q_1 and U so that Euclidean rays from q_1 “matched up” to the normal lines to G after passing through this “in between” region. \square

For the given hinged energy functional in Theorem 26, we now extrapolate a possible front at time 1 emanating from q_1 in accordance with the construction of Lemma 27.

Let $\tilde{\Gamma} = \{h = 0\} \subset B_0^{d-1}(a)$. For all $R \geq 0$, we denote by $N_0(R) \subset \mathbb{R}^{d-1}$ the set

$$N_0(R) = \tilde{\Gamma} + B_0^{d-1}(R).$$

Let ζ be the non-negative real valued smooth function in a bounded neighborhood \mathcal{V} of $\overline{B_0^{d-1}(a)} \subset \mathbb{R}^{d-1}$ such that

$$\zeta(x) = \sqrt{1 + h(x)} - 1.$$

The function ζ is introduced to allow the construction of h_{q_1, q_2} on the plane $z_d = 0$.

Lemma 28. *There exist neighborhoods $N_0 \subset B_0^{d-1}(a)$ and $\tilde{N}_0 \subset B_0^{d-1}(a)$ of $\tilde{\Gamma}$ and a function $\xi : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ that satisfy the following geometric property. In Cartesian coordinates (z_1, \dots, z_d) , for all $y \in \tilde{N}_0$, the normal line to the graph of ξ at $(y, \xi(y))$ crosses the plane $\{z_d = 0\}$ at a point $(x, 0)$, $x \in N_0$, such that the Euclidean distance between $(y, \xi(y))$ and $(x, 0)$ is $\zeta(x)$.*

Furthermore, ξ is a smooth non-negative function, everywhere less than $1/8$, with all of its derivatives bounded, and bounded from below by a positive constant outside of $B_0^{d-1}(a)$.

Proof. Let $\psi : \mathcal{V} \rightarrow \mathbb{R}^{d-1}$ be defined by $\psi(x) = x - \zeta(x)\nabla\zeta(x)$. For all $x \in \mathcal{V}$,

$$D\psi(x) = \text{id}_{\mathbb{R}^{d-1}} - \nabla\zeta(x) \cdot \nabla\zeta(x)^* - \zeta(x)\text{Hess}\zeta(x).$$

Since $h \geq 0$ and h vanishes on $\tilde{\Gamma}$, ∇h and $\nabla\zeta$ also vanish on $\tilde{\Gamma}$. Thus on $\tilde{\Gamma}$, $D\psi(x) = \text{id}_{\mathbb{R}^{d-1}}$. Since ψ is smooth, $D\psi(x)$ is uniformly continuous on the compact $\overline{B_0^{d-1}(a)}$ and there exists R_0 small enough such that $D\psi$ is invertible on $N_0(R_0) \subset \overline{B_0^{d-1}(a)}$.

In addition, there exists $R_1 \in (0, R_0)$ such that ψ is a diffeomorphism from $N_0(R_1)$ onto its image. This is shown by contradiction: assume for any $R > 0$ there exists a pair $(x, y) \in N_0(R)$ such that $x \neq y$ but $f(x) = f(y)$. Then let's define for each integer $n > 0$ such a pair $(x_n, y_n) \in N_0(1/n)$. Since the sequences evolve in the compact set $\overline{B_0^{d-1}(a)}$, they are convergent up to extraction. Furthermore, since $\bar{\Gamma}$ is a compact set, the only possible attractors for x_n and y_n belong to $\bar{\Gamma}$. Hence there exists $\tilde{x}, \tilde{y} \in \bar{\Gamma}$ such that $x_n \rightarrow \tilde{x}$, $y_n \rightarrow \tilde{y}$. Since ψ is continuous, $\psi(x_n) = \psi(y_n)$, we conclude that $\tilde{x} = \psi(\tilde{x}) = \psi(\tilde{y}) = \tilde{y}$. Hence $x_n - y_n \rightarrow 0$. This allows to conclude: indeed ψ is a local diffeomorphism, hence the compact set $\overline{N_0(R_0)}$ can be finitely covered with open balls on which the restriction of ψ is a diffeomorphism onto its image. There must exist one such open ball containing both x_n and y_n for n large enough (since $x_n, y_n \rightarrow \tilde{x}$). This imposes that $x_n = y_n$, which is a contradiction.

Let us pick $R_2 \in (0, R_1)$ small enough so that we also have that $\psi(N_0(R_2)) \subset B_0^{d-1}(a)$, and $|\nabla \zeta| < 1$, $\zeta < 1/16$ on $N_0(R_2)$.

Let $r \in (0, R_2)$. We can set the map $\xi_0 : \psi(\overline{N_0(r)}) \rightarrow \mathbb{R}$ such that

$$\xi_0(y) = \zeta(\psi^{-1}(y)) \sqrt{1 - |\nabla \zeta(\psi^{-1}(y))|^2}.$$

By definition, ξ_0 is the restriction of a C^∞ map on $\psi(N_0(R_2))$ to the closed set $\psi(\overline{N_0(r)})$. Hence ξ_0 automatically satisfies the Whitney compatibility condition from Whitney extension theorem and as a result can be extended to a smooth function with domain \mathbb{R}^{d-1} . Using smooth cut-off functions, we can ensure the existence of such an extension that has all of its derivatives bounded, and is bounded from below by a positive constant outside of $B_0^{d-1}(a)$, and since $\zeta < 1/16$ on $N_0(R_2)$, is strictly smaller than $1/8$. We pick one such extension as function ξ , $N_0 = N_0(r)$ and $\tilde{N}_0 = \psi(N_0(r))$.

Now let us check that the exhibited function ξ satisfies the stated geometric property. This statement is equivalent to

$$(\psi(x), \xi(\psi(x))) - (x, 0) = \zeta(x) \frac{(-\nabla \xi(\psi(x)), 1)}{\sqrt{1 + |\nabla \xi(\psi(x))|^2}}, \quad \forall x \in N_0.$$

This is translated to the pair of equations

$$(11) \quad \zeta(x) = \xi(\psi(x)) \sqrt{1 + |\nabla \xi(\psi(x))|^2},$$

$$(12) \quad x = \psi(x) + \frac{\zeta(x) \nabla \xi(\psi(x))}{\sqrt{1 + |\nabla \xi(\psi(x))|^2}}.$$

From the definition of ξ , we have for all $x \in N_0$

$$(13) \quad \xi(x - \zeta(x) \nabla \zeta(x)) = \zeta(x) \sqrt{1 - |\nabla \zeta(x)|^2}.$$

Differentiating the left-hand side,

$$\nabla(\xi(x - \zeta(x) \nabla \zeta(x))) = (\text{id} - \nabla \zeta(x) \cdot \nabla \zeta(x)^* - \zeta(x) \text{Hess} \zeta(x)) \cdot \nabla \xi(\psi(x)).$$

Differentiating the right-hand side,

$$\nabla \left(\zeta(x) \sqrt{1 - |\nabla \zeta(x)|^2} \right) = \frac{\nabla \zeta(x)}{\sqrt{1 - |\nabla \zeta(x)|^2}} (1 - |\nabla \zeta(x)|^2) - \zeta(x) \frac{\text{Hess} \zeta(x) \cdot \nabla \zeta(x)}{\sqrt{1 - |\nabla \zeta(x)|^2}}.$$

For any vector $v \in \mathbb{R}^n$, denoting v^* its transpose, we have the identity $v|v|^2 = v \cdot (v^* \cdot v) = (v \cdot v^*) \cdot v$. Hence

$$\nabla \left(\zeta(x) \sqrt{1 - |\nabla \zeta(x)|^2} \right) = (\text{id} - \nabla \zeta(x) \cdot \nabla \zeta(x)^* - \zeta(x) \text{Hess} \zeta(x)) \cdot \frac{\nabla \zeta(x)}{\sqrt{1 - |\nabla \zeta(x)|^2}}.$$

The radius r has been chosen so that $\text{id} - \nabla \zeta(x) \cdot \nabla \zeta(x)^* - \zeta(x) \text{Hess} \zeta(x)$ is invertible. Therefore (13) implies after differentiation

$$\nabla \xi(\psi(x)) = \frac{\nabla \zeta(x)}{\sqrt{1 - |\nabla \zeta(x)|^2}}$$

and, equivalently,

$$\nabla \zeta(x) = \frac{\nabla \xi(\psi(x))}{\sqrt{1 + |\nabla \xi(\psi(x))|^2}}.$$

Since $x = \psi(x) + \zeta(x)\nabla \zeta(x)$ by definition of ψ , we then have (12).

Likewise,

$$\xi(\psi(x)) = \zeta(x)\sqrt{1 - |\nabla \zeta(x)|^2} = \zeta(x)\sqrt{1 - \frac{|\nabla \xi(\psi(x))|^2}{1 + |\nabla \xi(\psi(x))|^2}} = \frac{\zeta(x)}{\sqrt{1 + |\nabla \xi(\psi(x))|^2}},$$

which implies (11), and concludes the proof of the lemma. \square

We are now able to give a proof of Theorem 26.

Proof of Theorem 26. Let $\xi : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ be as in the statement of Lemma 28. By application of Lemma 27, there exist V^+ a neighborhood of $B_0^{d-1}(a) \times \{0\} \cup \{q_1\}$ and a (smooth) metric on V^+ such that the graph of ξ in $B_0^{d-1}(a) \times [0, 1/8]$ is a subset of the sphere of radius 1 around q_1 . Furthermore, the metric agrees with g_E on a neighborhood of

$$\{z \in B_0^{d-1}(a) \times [0, 1/8] \mid 0 \leq z_d \leq \xi(z_1, \dots, z_{d-1})\}.$$

Now reflect V^+ around $\{z_d = 0\}$ to get V^- (and note that q_2 is the image of q_1) and reflect g to get a metric on V^- such that $-G$ (which denotes the graph of $-\xi$) is an open subset of the sphere of radius 1 around q_2 in this metric. Note that this metric on V^- is compatible with g because, on a neighborhood of $B_0^{d-1}(a)$, they are both isometric to the Euclidean metric (and thus reflection induces a valid transition function), and thus we can extend g to V^- via reflection.

In summary, we have built a metric g on a neighborhood V of $B_0^{d-1}(a) \times \{0\} \cup \{q_1\} \cup \{q_2\}$ such that

$$d_g(q_1, (y, \xi(y))) = d_g(q_2, (y, -\xi(y))) = 1 \quad \forall y \in B_0^{d-1}(a),$$

and g coincides with g_E on a neighborhood of

$$\{z \in B_0^n(a) \mid |z_d| \leq \xi(z_1, \dots, z_{d-1})\}.$$

Let N_0 be as in the statement of Lemma 28. Then for any $x \in N_0$, we have

$$d_g(q_1, (x, 0)) = 1 + \zeta(x).$$

Indeed this is a consequence of the geometric property in Lemma 28. Since g is flat on a neighborhood of $\{z \in B_0^{d-1}(a) \mid |z_d| \leq \xi(z_1, \dots, z_{d-1})\}$,

$$d_g(q_1, (x, 0)) = d_g(q_1, (y, \xi(y))) + d_g((y, \xi(y)), (x, 0)).$$

(See Figure 4.) Likewise

$$d_g(q_2, (x, 0)) = 1 + \zeta(x).$$

Hence for all $x \in N_0$,

$$h_{q_1, q_2}(x, 0) = \frac{1}{2} (d(q_1, (x, 0))^2 + d(q_2, (x, 0))^2) = (1 + \zeta(x))^2 = 1 + h(x).$$

We can now work on extending h_{q_1, q_2} to a neighborhood of Γ in \mathbb{R}^d . Let $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^{d-1}$ such that $\pi(z_1, \dots, z_d) = (z_1, \dots, z_{d-1})$. To prove the statement, we show that there exists $r > 0, \varepsilon > 0$ such that on $N_0(r) \times (-\varepsilon, \varepsilon)$, the map defined by

$$v(z) = \begin{cases} +\sqrt{h_{q_1, q_2}(z) - h_{q_1, q_2}(\pi(z), 0)} & \text{if } z_d \geq 0, \\ -\sqrt{h_{q_1, q_2}(z) - h_{q_1, q_2}(\pi(z), 0)} & \text{if } z_d < 0 \end{cases}$$

is smooth and $z \mapsto v(x, z)$ is a diffeomorphism for each $z \in N_0(r) \times (-\varepsilon, \varepsilon)$. As a result, $\Phi : z \mapsto (\pi(z), v(z))$ is a diffeomorphism and h_{q_1, q_2} is right equivalent to

$$u \mapsto 1 + h(u_1, \dots, u_{d-1}) + u_d^2, \quad \forall u \in \Phi(N_0(r) \times (-\varepsilon, \varepsilon)).$$

By symmetry of the metric with respect to the hyperplane $\{z = 0\}$,

$$h_{q_1, q_2}(z) = \frac{1}{2} (d_g(q_1, (\pi(z), z_d))^2 + d_g(q_1, (\pi(z), -z_d))^2), \quad \forall z \in V.$$

Thus by symmetry, on $B_0^{d-1}(a)$,

$$\left. \frac{\partial h_{q_1, q_2}}{\partial z_n} \right|_{z_d=0} = 0.$$

Furthermore,

$$\left. \frac{\partial^2 h_{q_1, q_2}}{\partial z_d^2} \right|_{z_d=0} = 2 \left(\frac{\partial d_g(q_1, \cdot)}{\partial z_d} \right)^2 \Big|_{z_d=0} + 2d_g(q_1, \cdot) \left. \frac{\partial^2 d_g(q_1, \cdot)}{\partial z_n^2} \right|_{z_d=0}.$$

Notice that if $x_0 \in \tilde{\Gamma}$, then the geodesic joining q_1 to q_2 passing through $(x_0, 0)$ is supported near $(x_0, 0)$ by the straight line $\{x = x_0\}$. This implies that $\left. \frac{\partial d_g(q_1, \cdot)}{\partial z_n} \right|_{z_n=0} = 1$ and $\left. \frac{\partial^2 d_g(q_1, \cdot)}{\partial z_n^2} \right|_{z_n=0} = 0$. Hence

$$\frac{\partial^2 h_{q_1, q_2}}{\partial z_n^2}(x_0, z) = 2.$$

This allows to apply Malgrange preparation theorem: there exists $\alpha : \mathbb{R}^d \rightarrow \mathbb{R}$, smooth, such that

$$h_{q_1, q_2}(z) = \alpha(z)z_n^2 + h_{q_1, q_2}(\pi(z), 0)$$

and $\alpha(x_0, 0) = 1$ for all $x_0 \in \tilde{\Gamma}$.

The function α admits a uniform positive lower bound on a sufficiently small neighborhood of $\tilde{\Gamma} \times \{0\}$, hence, up to reducing r and ε , $\sqrt{\alpha}$ is a smooth function on this neighborhood. As a consequence,

$$v(z) = z_d \sqrt{\alpha(z)}$$

and is a smooth function. This implies furthermore that $z \mapsto (\pi(z), v(z))$ is a diffeomorphism on $N_0(r) \times (-\varepsilon, \varepsilon)$ since $\partial_{z_d} u(x_0, 0) = \sqrt{\alpha(x_0)} > 0$ for all $x_0 \in \tilde{\Gamma}$.

Notice that for any $z \in B_0^d(a)$ such that $0 < |z_d| < \zeta(\pi(z))$, $h_{q_1, q_2}(z) > 2$. Hence $\Gamma = \tilde{\Gamma} \times \{0\}$ and $N = N_0(r) \times (-\varepsilon, \varepsilon)$ is a neighborhood of Γ . Furthermore, we have proved that

$$h_{q_1, q_2}(\Phi^{-1}(u_1, \dots, u_n)) = 1 + h(u_1, \dots, u_{d-1}) + u_d^2,$$

for all $u \in \Phi(N)$. Once this fact is proved, what remains to be shown is the shape of the heat kernel. However this is a direct application of Corollary 4, hence the statement. \square

Remark 29. Our treatment of prescribing singularities for the hinged energy function in the Riemannian case appears local; for example, the case of antipodal points on the standard sphere goes beyond the framework of Theorem 26. However, that is essentially the only situation not included in the theorem. To be more precise, for fixed points q_1 and q_2 , Γ can be identified as a subset of the sphere of radius $d(q_1, q_2)/2$ in $T_{q_1}M$. If Γ is the entire sphere, then necessarily we have the Morse-Bott case as covered by Proposition 25. Otherwise, since Γ is closed, for a point q on the sphere not in Γ , it has a neighborhood which is not in Γ , and stereographic projection around q maps Γ to a subset of $B_0^{d-1}(a)$ for some $a > 0$. Thus every case in which Γ is not the entire sphere of radius $d(q_1, q_2)/2$ in $T_{q_1}M$ can be realized as in Theorem 26.

We follow on the preceding proof by showing it can be extended to construct prescribed singularities for the hinged energy function also on contact sub-Riemannian structures, and we consider this to be sufficient for this line of inquiry.

Theorem 30. *Let M be a $2d+1$ -dimensional contact manifold, let a and σ be positive real numbers, and let h be a smooth, real-valued function in a neighborhood of $\overline{B_0^{2d-1}(a)} \subset \mathbb{R}^{d-1}$ such that $h(0, \dots, 0) = 0$, non-negative on $\overline{B_0^{2d-1}(a)}$, and positive on $\partial B_0^{2d-1}(a)$. Then there exists a sub-Riemannian metric on*

M (compatible with the contact structure), and some points q_1 and q_2 such that $\Gamma = \Gamma(q_1, q_2)$ is contained in a coordinate patch

$$(u_1, \dots, u_{2d+1}) : U \rightarrow B_0^{2d-1}(a) \times (-\delta, \delta) \times (-\delta, \delta)$$

such that

$$h_{q_1, q_2}|_N = \frac{\sigma^2}{4} + h(u_1, \dots, u_{2n-1}) + u_{2n}^2 + u_{2n+1}^2$$

and the analogue of (10) holds.

Proof. By the Darboux theorem, any point has a neighborhood that is contactomorphic to the standard contact structure. Thus we may take N to be a neighborhood for the origin in \mathbb{R}^{2d+1} with the standard contact structure. Moreover, by rescaling, we can take $q_1 = (-1, \dots, 0)$, $q_2 = (1, 0, \dots, 0)$, and N a ball around the origin of Euclidean radius 3. Also recall that contact sub-Riemannian structures don't admit non-trivial abnormal, so we don't need to worry that the metric we construct will have any.

It's convenient to use more standard notation for our coordinates, so let $(v_1, w_1, \dots, v_d, w_d, u)$ be coordinates on $(\mathbb{R}^2)^d \times \mathbb{R}$. Then every admissible curve is given as the Legendre lift of a curve in $(\mathbb{R}^2)^d$. In particular, let $\tilde{\gamma}(t) = (v_1(t), w_1(t), \dots, v_d(t), w_d(t))$ be a curve in $(\mathbb{R}^2)^d$, let

$$A_i(t) = \int_0^t v_i(t) dw_i - w_i(t) dv_i$$

be twice the enclosed signed area of the projection to the i th \mathbb{R}^2 factor, and let $u(t) = \sum_{i=1}^d A_i(t)$. Then $\gamma(t) = (\tilde{\gamma}(t), u(t)) = (v_1(t), w_1(t), \dots, v_d(t), w_d(t), u(t))$ is the lift of $\tilde{\gamma}(t)$.

Moreover, given a Riemannian metric \tilde{g} on $(\mathbb{R}^2)^d$, it lifts to a sub-Riemannian metric g on the contact structure, which is invariant under translation in the u -direction and which has the property that the length of any admissible curve γ is the Riemannian length of its projection $\tilde{\gamma}$ (with respect to \tilde{g} , of course). By the previous theorem, we can choose \tilde{g} such that, if $h_{x,y}$ is the hinged energy function on $(\mathbb{R}^2)^d$ with respect to \tilde{g} , then \tilde{h}_{q_1, q_2} has normal form

$$\frac{\sigma^2}{4} + h(u_1, \dots, u_{2d-1}) + u_{2d}^2.$$

Also, recall that the metric is symmetric under reflection in the v_1 axis ($v_1 \mapsto -v_1$), and thus the midpoint set $\tilde{\Gamma}$ is contained in the hyperplane $\{v_1 = 0\}$.

Now consider the corresponding sub-Riemannian lifted metric g and associated hinged energy function h_{q_1, q_2} — we claim that h has the desired normal form. First, consider a point

$$z = (v_1, w_1, \dots, v_d, w_d, u) \in (\mathbb{R}^2)^d \times \mathbb{R},$$

and let $\pi(z) = (v_1, w_1, \dots, v_d, w_d) \in (\mathbb{R}^2)^d$ be the projection. Then letting d and \tilde{d} denote the distance functions on $(\mathbb{R}^2)^d \times \mathbb{R}$ and $(\mathbb{R}^2)^d$, respectively, we see that $d(q_1, z) \geq \tilde{d}(q_1, \pi(z))$, with equality if and only if there is a minimizing geodesic $\tilde{\gamma}$ from q_1 to $\pi(z)$ such that the endpoint of the lift γ is z (that is, if and only if there is a minimizing geodesic that encloses the “right” signed area).

Take $\tilde{z} \in \tilde{\Gamma}$. We know that there is a unique (and non-conjugate) minimizing geodesic $\tilde{\gamma}$ from q_1 to \tilde{z} , and thus there is a unique z such that $\pi(z) = \tilde{z}$ and $h(q_1, z) = \tilde{h}(q_1, \tilde{z})$; we write this z as $(\tilde{z}, \bar{u}(\tilde{z}))$. Further, by the reflection symmetry of the metric, the minimal geodesic from q_2 to \tilde{z} is given by the reflection of $\tilde{\gamma}$ (under the map $v_1 \mapsto -v_1$), and thus $(\tilde{z}, \bar{u}(\tilde{z}))$ is also the unique z such that $\pi(z) = \tilde{z}$ and $h(q_2, z) = \tilde{h}(q_2, \tilde{z})$. It follows that Γ is the lift of $\tilde{\Gamma}$ under the map $\tilde{z} \mapsto (\tilde{z}, \bar{u}(\tilde{z}))$ (which is well defined on $\tilde{\Gamma}$), and that $h(\Gamma) = \tilde{h}(\tilde{\Gamma})$. Moreover, we know that there is a neighborhood of $\tilde{\Gamma}$ such that every point is joined to q_1 by a unique, non-conjugate minimizing geodesic. If we let U be the intersection of this neighborhood with the hyperplane $\{v_1 = 0\} \subset (\mathbb{R}^2)^d$, then the map $\tilde{z} \mapsto (\tilde{z}, \bar{u}(\tilde{z}))$ extends to U , by the same argument. Further, by the smoothness of the exponential map (and of the enclosed area as a function of the curve) this is a smooth embedding of U into the hyperplane $\{v_1 = 0\} \subset (\mathbb{R}^2)^d \times \mathbb{R}$ such that $h((\tilde{z}, \bar{u}(\tilde{z}))) = \tilde{h}(\tilde{z})$, for any $\tilde{z} \in U$. Denote this embedding by \bar{U} .

We are now in a position to show that h has the desired normal form. First, restricting our attention to \bar{U} , it follows from the above that there exist coordinates on \bar{U} such that

$$h_{q_1, q_2}|_{\bar{U}} = \frac{\sigma^2}{4} + h(u_1, \dots, u_{2d-1})$$

(that is, $h|_{\bar{U}}$ has the same “normal form” as $\tilde{h}|_U$). If we show that the Hessian of h on the normal bundle of \bar{U} is non-degenerate, which is 2-dimensional and spanned by ∂_{v_1} and ∂_u , then the Malgrange preparation theorem (or parametrized Morse lemma) for smooth functions implies that we can find coordinates in which h has the desired expression on all of N . Consider the Hessian (as a quadratic form), at a point $z_0 \in \Gamma$, along a vector $\alpha\partial_u + \beta\partial_{v_1}$. If $\beta \neq 0$, then because $d(x, z) \geq \tilde{d}(x, \pi(z))$ and the Hessian of $\tilde{d}(x, \pi(z_0))$ along $\beta\partial_{v_1}$ is positive, the Hessian of h is also positive. So it remains only to show that the Hessian along ∂_u is positive.

Again consider $z_0 = (v_1, w_1, \dots, v_d, w_d, u_0) \in \Gamma$, and let $z_s = (v_1, w_1, \dots, v_d, w_d, u_0 + s)$. Let $\tilde{\gamma}$ be the unique, non-conjugate minimal geodesic from q_1 to $\pi(z_0)$ and γ_0 its lift. Now let γ_s be the unique, non-conjugate sub-Riemannian minimal geodesic from q_1 to z_s , and let $\tilde{\gamma}_s$ be its projection (this is well defined for s near 0 because the complement of the cut locus is open, in both Riemann and sub-Riemannian geometry, and in particular, the relevant exponential maps are local diffeomorphisms). Then $\tilde{\gamma}_s$ is the unique shortest curve from q_1 to $\pi(z_0)$ (with respect to the Riemannian metric on $(\mathbb{R}^2)^d$), subject to the constraint that the endpoint lifts to z_0 (that is, subject to the constraint that it encloses the right area, making it the solution to the appropriate Dido problem). Further, $\tilde{\gamma}_s$ is a one-parameter family of proper deformations of $\tilde{\gamma}$ (meaning the endpoints are kept fixed), and the variation field at $s = 0$ is non-trivial because the enclosed area in changing to first-order. But, by the classical theory of the second variation of energy near a minimizing, non-conjugate Riemannian geodesic, this means that the second derivative of the length of $\tilde{\gamma}_s$ is positive at $s = 0$. Since this length is also $d(q_1, z_s)$, and since $d(q_2, z_s) = d(q_1, z, s)$ by symmetry, it follows that $\frac{\partial^2}{\partial s^2} h(z_s) > 0$. Recalling the definition of z_s , this completes the construction of the metric. \square

From here, the heat kernel representation follows as before.

One virtue of Theorem 26 is that many of the real-analytic normal forms appearing in [4] and corresponding to local minima can be realized as h_{q_1, q_2} on some Riemannian manifold M of a high enough dimension to support the normal form in question. The only restriction is that one needs the geodesic direction to be separate from the others. The corresponding Laplace asymptotic expansions can be realized as heat kernel asymptotics on such manifolds, which means that there are cases when the heat kernel asymptotics contain powers of $\log t$, for instance. Indeed, we have the following corollary.

Corollary 31. *For any integers $d \geq 2$, $p \geq 1$, and $0 \leq k \leq d - 2$, there exists a smooth Riemannian manifold M of dimension d , and q_1, q_2 in M , $q_1 \neq q_2$, such that for some $C \neq 0$,*

$$p_t(q_1, q_2) = e^{-\frac{d^2(q_1, q_2)}{4t}} t^{\frac{1}{2} + \frac{1}{2p} - d} \log(t)^k (C + o(1)).$$

Proof. This is a matter of applying Theorem 26 to the right function h .

From [4, Theorems 7.3-7.4], we have that for any smooth non-negative function $\phi : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$, positive at 0, we have the following Laplace integral asymptotics near $t = 0$:

$$(14) \quad \int_{\mathbb{R}^k} \exp\left(\frac{u_1^{2p} \cdots u_{k+1}^{2p}}{t}\right) \phi(u_1, \dots, u_{k+1}) du_1 \cdots du_{k+1} = t^{1/2p} \log(t)^k (C + o(1)).$$

(With C a non-zero constant on the only condition that $\phi(0) \neq 0$.)

Let $h : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ be defined by

$$h(u_1, \dots, u_{d-1}) = u_1^{2p} \cdots u_{k+1}^{2p} + \chi(u_1, \dots, u_{d-1}).$$

Here $\chi : \mathbb{R}^{d-1} \rightarrow [0, 1]$ is a smooth function, equal to 0 on $B_0^{d-1}(a)$, and equal to 1 on the complement of $B_0^{d-1}(a+1)$, for some $a > 0$.

Thus there exists a smooth Riemannian manifold M of dimension d , q_1, q_2 in M such that $q_1 \neq q_2$, and

$$(15) \quad p_t(q_1, q_2) = \frac{1}{t^d} e^{-\frac{d^2(q_1, q_2)}{4t}} \int_{(-\varepsilon, \varepsilon)^d} \Phi(t, u) e^{-\frac{h(u_1, \dots, u_{d-1}) + u_d^2}{t}} du_1 \cdots du_d + O\left(e^{-\frac{d^2(q_1, q_2) + c}{4t}}\right).$$

for some positive ε and a smooth prefactor function Φ over $\mathbb{R}^+ \times (-\varepsilon, \varepsilon)^d$, smoothly extendable and positive at $t = 0$.

For ε small enough, if $\sum_{i=1}^{d-1} u_i^2 < \varepsilon^2$, $h(u_1, \dots, u_{d-1}) = u_1^{2p} \cdots u_{k+1}^{2p}$. Thus equation (15) implies

$$(16) \quad p_t(q_1, q_2) = \frac{1}{t^d} e^{-\frac{d^2(q_1, q_2)}{4t}} \int_{(-\varepsilon, \varepsilon)^d} (\psi_0(u) + t\psi_1(t, u)) e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_d + O\left(e^{-\frac{d^2(q_1, q_2) + c}{4t}}\right)$$

with $\psi_0 : (-\varepsilon, \varepsilon)^d \rightarrow \mathbb{R}$, positive at 0 and smooth, and $\psi_1 : \mathbb{R}^+ \times (-\varepsilon, \varepsilon)^d \rightarrow \mathbb{R}$, smoothly extendable at $t = 0$.

Then

$$\left| \int_{(-\varepsilon, \varepsilon)^d} t\psi_1(t, u) e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_d \right| \leq Ct \int_{(-\varepsilon, \varepsilon)^d} e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_d.$$

From the formula (14), we have that for some $C \neq 0$,

$$\int_{(-\varepsilon, \varepsilon)^d} e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_{k+1} du_d = t^{1/2} t^{1/2p} \log(t)^k (C + o(1)).$$

Likewise

$$\begin{aligned} \int_{(-\varepsilon, \varepsilon)^d} \psi_0(u) e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_d = \\ \int_{(-\varepsilon, \varepsilon)^{k+2}} \left(\int_{(-\varepsilon, \varepsilon)^{d-k-2}} \psi_0(u) du_{k+2} \cdots du_{d-1} \right) e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_{k+1} du_d. \end{aligned}$$

Then $\Psi_0 : (-\varepsilon, \varepsilon)^{k+2} \rightarrow \mathbb{R}$ given by $\Psi_0(u_1, \dots, u_{k+1}, u_d) = \int_{(-\varepsilon, \varepsilon)^{d-k-2}} \psi_0(u) du_{k+2} \cdots du_{d-1}$ is a smooth positive function and

$$\int_{(-\varepsilon, \varepsilon)^{k+2}} \Psi_0(u) e^{-\frac{u_1^{2p} \cdots u_{k+1}^{2p} + u_d^2}{t}} du_1 \cdots du_{k+1} du_d = t^{1/2} t^{1/2p} \log(t)^k (C + o(1)).$$

Putting all three parts of (16) together, we get the asymptotic expansion first term:

$$p_t(q_1, q_2) = e^{-\frac{d^2(q_1, q_2)}{4t}} t^{\frac{1}{2} + \frac{1}{2p} - d} \log(t)^k (C + o(1)).$$

□

Theorem 26 also allows to go beyond functions admitting an analytic normal form, such as present in [4]. In that case, an asymptotic expansion of the Laplace integral in the theorem is not accessible by the methods of [4], and moreover, appear not to be known. To illustrate, we offer the following examples.

Example 32. Let

$$h(u_1) = \begin{cases} e^{-1/u_1^2} & \text{for } u_1 \neq 0 \\ 0 & \text{for } u_1 = 0 \end{cases}$$

on $(-\varepsilon, \varepsilon) \subset \mathbb{R}$. Then it's well known that h satisfies the hypotheses of Theorem 26.

Example 33. Let $g(\theta)$ be a smooth function on \mathbb{S}^1 which is equal to θ^2 near $\theta = 0$ and is strictly positive elsewhere. Then in polar coordinates on \mathbb{R}^2 , let

$$h(r, \theta) = g(\theta)(r-1)^2 + (r-1)^4$$

near the circle $\{r = 1\}$ in \mathbb{R}^2 , and extended to be greater than some $\varepsilon > 0$ elsewhere. This gives a situation where $\Gamma = \mathbb{S}^1$ and where h_{q_1, q_2} is locally Morse-Bott away from $\theta = 0$, but where the Hessian in the normal direction degenerates as θ approaches 0. Thus, the usual Morse-Bott expansion of Section 4.2 does not apply. Of course, the $(r - 1)^4$ can be replaced by $(r - 1)^{2k}$ for any positive integer k , or even by

$$\begin{cases} e^{-1/(r-1)^2} & \text{for } r \neq 1 \\ 0 & \text{for } r = 1 \end{cases}$$

to produce other examples in a similar vein, and similarly, $g(\theta)$ can have behavior near $\theta = 0$ modeled on any even power of θ or on e^{-1/θ^2} .

Example 34. Let $h(u_1)$ be a smooth, non-negative function with zeroes at $\pm \frac{1}{n}$ for all positive integers n and at 0. The existence of such functions is well-known, and while the Hessian can be made non-degenerate at all of the $\pm \frac{1}{n}$ (although it need not be), h necessarily vanishes to all order at 0. Moreover, in this case, Γ is not a union of smooth submanifolds (the condition to respect the submanifold topology is not satisfied at 0).

5. LOGARITHMIC DERIVATIVES

5.1. Molchanov-type expansions of logarithmic derivatives. We start by introducing an alternative representation of Molchanov method that will be useful in following computations. This is a direct consequence of Léandre estimates coupled with Ben Arous expansions on compact sets with no abnormal geodesics (Proposition 20).

Lemma 35 (Folding the remainder). *Let $\Sigma : \mathbb{R}^+ \times M^2 \setminus \mathcal{C} \rightarrow \mathbb{R}$ denote the smooth function such that*

$$\Sigma_t(x, y) = t^{d/2} e^{\frac{d(x, y)^2}{4t}} p_t(x, y).$$

Let \mathcal{K} be a compact subset of $M^2 \setminus \mathcal{D}$ such that all minimizers between pairs $(x, y) \in \mathcal{K}$ are strongly normal. For $\varepsilon, t_0 > 0$ small enough, let Ω be an open neighborhood of the set

$$\{(t, x, y, z) \in (0, t_0) \times \mathcal{K} \times M \mid (x, y) \in \mathcal{K}, z \in \Gamma_\varepsilon(x, y)\},$$

such that $\bar{\Omega} \subset [0, t_0] \times (M^2 \setminus \mathcal{D}) \times M$ and ε is assumed small enough so that (x, z) and (z, y) avoid \mathcal{C} .

Suppose we are in the symmetric case. Then there exists a continuous map $\bar{\Sigma} : \Omega \rightarrow \mathbb{R}$, smooth as a map of (t, y) , such that for all $(x, y) \in \mathcal{K}$, for all $t < t_0$,

$$p_t(x, y) = \int_{\Gamma_\varepsilon} \left(\frac{2}{t}\right)^d e^{-\frac{h_{x, y}(z)}{4t}} \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) d\mu(z)$$

and for all $l \in \mathbb{N}$ and α multi-index, there exists $C > 0$ such that for all $(x, y) \in \mathcal{K}$,

$$\partial_t^l Z_y^\alpha [\bar{\Sigma}_t^x(z, y) - \Sigma_t(z, y)] \leq C e^{-\frac{\varepsilon^2}{8t}}.$$

In particular, for all $l \in \mathbb{N}$ and α multi-index, for all $t, x, y, z \in \Omega$,

$$\partial_t^l Z_y^\alpha \big|_{t=0} \bar{\Sigma}_t^x(z, y) = \partial_t^l Z_y^\alpha \big|_{t=0} \Sigma_t(z, y).$$

In the general (non-symmetric) case, all of the above holds with $l = 0$.

Proof. We write the proof of the symmetric case; setting $l = 0$ gives the proof in the general case. We set

$$\zeta_t^x(y) = \int_{\Gamma_\varepsilon} e^{-\frac{h_{x, y}(z)}{t}} \Sigma_{t/2}(x, z) d\mu(z).$$

By definition, ζ is continuous on Ω and smooth with respect to (t, y) . Furthermore, following Laplace integrals asymptotics, the strategy given in the proof of Proposition 5 yields for all $l \in \mathbb{N}$ and α multi-index, the existence of a constant $C > 0$ such that on Ω

$$(17) \quad \partial_t^l Z_y^\alpha \zeta_t^x(y) \leq \frac{C e^{-\frac{d(x, y)}{4t}}}{t^{2l+|\alpha|+d-1/2}}$$

Furthermore, when $l = 0$, $\alpha = 0$,

$$(18) \quad \frac{e^{-\frac{d(x,y)}{4t}}}{Ct^{d/2}} \leq \zeta_t^x(y) \leq \frac{Ce^{-\frac{d(x,y)}{4t}}}{t^{d-1/2}}.$$

Now set

$$R_t^x(y) = \left(\frac{t}{2}\right)^d \int_{M \setminus \Gamma_\varepsilon} p_{t/2}(x, z) p_{t/2}(z, y) d\mu(y).$$

By Corollary 15, for all $l \in \mathbb{N}$ and α multi-index, there exists C such that for all $(t, x, y, z) \in \Omega$,

$$(19) \quad \partial_t Z_y^\alpha R_t^x(y) \leq Ce^{-\frac{d(x,y)^2}{4t}} e^{-\frac{\varepsilon^2}{4t}}$$

Furthermore, R is continuous on Ω and smooth as a function of (t, y) .

We now pick ψ to be

$$\bar{\Sigma}_t^x(z, y) = \Sigma_t(z, y) + \frac{R_t^x(y)}{\zeta_t^x(y)}.$$

Then

$$\begin{aligned} \int_{\Gamma_\varepsilon} \left(\frac{2}{t}\right)^d e^{-\frac{h_{x,y}(z)}{4t}} \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) d\mu(z) &= \int_{\Gamma_\varepsilon} p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) \\ &\quad + \underbrace{\int_{\Gamma_\varepsilon} e^{-\frac{h_{x,y}(z)^2}{t}} \Sigma_{t/2}(x, z) d\mu(z)}_{=\zeta_t^x(y)} \frac{R_t^x(y)}{\zeta_t^x(y)} \left(\frac{2}{t}\right)^d. \end{aligned}$$

By construction, this equation simplifies to

$$\begin{aligned} \int_{\Gamma_\varepsilon} \left(\frac{2}{t}\right)^d e^{-\frac{h_{x,y}(z)}{4t}} \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) d\mu(z) &= \\ \int_{\Gamma_\varepsilon} p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) + \int_{M \setminus \Gamma_\varepsilon} p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) &= \\ = \int_M p_{t/2}(x, z) p_{t/2}(z, y) d\mu(z) = p_t(x, y). \end{aligned}$$

To conclude the proof, we only have to combine (17), (18) and (19) to check that for all $l \in \mathbb{N}$ and α multi-index, there exists $C > 0$, and $m \in M$ such that

$$\partial_t^l Z_y^\alpha \frac{R_t^x(y)}{\zeta_t^x(y)} \leq C \frac{e^{-\frac{\varepsilon^2}{4t}}}{t^m} \leq C' e^{-\frac{\varepsilon^2}{8t}}.$$

□

Let K be a compact of M such that no two disjoint points are connected by an abnormal geodesic, and let $x, y, z \in K$. Let Z be a smooth vector field in a neighborhood of y , and let Z_y denote Z acting in the y -variable. Assuming z and y are not cut, we know that Ben Arous expansion applies and can be differentiated, so that

$$Z_y p_t(z, y) = -\frac{d(z, y)}{2t} \cdot Z_y d(z, y) \cdot p_t(z, y) + Z_y (\log \Sigma_t(z, y)) p_t(z, y).$$

Let $t, \varepsilon > 0$ be small enough to apply Lemma 35. Then the first logarithmic derivative (whenever x and y are connected only by strongly normal geodesics) is given by

$$\begin{aligned} Z_y \log p_t(x, y) &= \frac{Z_y p_t(x, y)}{p_t(x, y)} \\ &= \frac{\int_{\Gamma_\varepsilon} \left[-\frac{d(z, y)}{t} Z_y d(z, y) + Z_y \left(\log \bar{\Sigma}_{t/2}^x(z, y) \right) \right] \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) e^{-\frac{d^2(x, z) + d^2(z, y)}{2t}} d\mu(z)}{\int_{\Gamma_\varepsilon} \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) e^{-\frac{d^2(x, z) + d^2(z, y)}{2t}} d\mu(z)} \end{aligned}$$

Recall that since Γ_ε is a bounded subset of a smooth manifold, a probability measure on Γ_ε is determined by its integrals against smooth functions (on M), and also that weak convergence of probability measures on Γ_ε can be characterized by the convergence of their integrals against smooth functions. Then if we define the probability measure m_t (for $t > 0$) by

$$(20) \quad \mathbb{E}^{m_t}[f] = \frac{\int_{\Gamma_\varepsilon} f(z) \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) e^{-\frac{d^2(x, z) + d^2(z, y)}{2t}} d\mu(z)}{\int_{\Gamma_\varepsilon} \Sigma_{t/2}(x, z) \bar{\Sigma}_{t/2}^x(z, y) e^{-\frac{d^2(x, z) + d^2(z, y)}{2t}} d\mu(z)},$$

for any smooth f , then m_t is supported on Γ_ε and absolutely continuous with respect to μ , and we have that

$$(21) \quad Z_y \log p_t(x, y) = \mathbb{E}^{m_t} \left[-\frac{d(\cdot, y)}{t} Z_y d(\cdot, y) + Z_y \left(\log \bar{\Sigma}_{t/2}^x(\cdot, y) \right) \right].$$

Further, m_t (and in particular $\bar{\Sigma}^x$) is defined so as to make this an equality, but we are interested in asymptotic behavior. To this end, observe that we can write $\Sigma_{t/2}(x, z)$ and $\bar{\Sigma}_{t/2}^x(z, y)$ as $c_0(x, z) + O(t)$ and $c_0(z, y) + O(t)$, and recall that the c_0 are smooth and strictly positive. We see that, if $m_{t_n} \rightarrow m_0$ for some sequence of times $t_n \searrow 0$ and some probability measure m_0 , then, for any smooth f , we have

$$(22) \quad \frac{\int_{\Gamma_\varepsilon} f(z) c_0(x, z) c_0(z, y) e^{-\frac{d^2(x, z) + d^2(z, y)}{2t_n}} d\mu(z)}{\int_{\Gamma_\varepsilon} c_0(x, z) c_0(z, y) e^{-\frac{d^2(x, z) + d^2(z, y)}{2t_n}} d\mu(z)} \rightarrow \mathbb{E}^{m_0}[f],$$

and conversely, if there is some m_0 and some sequence of times t_n such that (22) holds for all smooth f , then $m_{t_n} \rightarrow m_0$.

Continuing with log-derivatives, if Z' is another smooth vector field in a neighborhood of y , we see that

$$Z'_y Z_y (\log p_t(z, y)) = \frac{Z'_y Z_y p_t(z, y)}{p_t(z, y)} - Z'_y (\log p_t(z, y)) \cdot Z_y (\log p_t(z, y)),$$

and moreover (writing $Z' = Z'_y$ and $Z = Z_y$ to unburden the notation)

$$\begin{aligned} Z' Z p_t(z, y) &= \left[-\frac{1}{2t} Z' d(z, y) \cdot Z d(z, y) - \frac{d(z, y)}{2t} Z' Z d(z, y) + \frac{Z' Z (\Sigma_t(z, y))}{\Sigma_t(z, y)} \right] p_t(z, y) \\ &\quad + \left[-\frac{d(z, y)}{2t} Z d(z, y) + Z (\log \Sigma_t(z, y)) \right] \cdot \left[-\frac{d(z, y)}{2t} Z d(z, y) + Z (\log \Sigma_t(z, y)) \right] p_t(z, y). \end{aligned}$$

Putting this together, we find that

$$(23) \quad \begin{aligned} Z' Z (\log p_t(x, y)) &= \mathbb{E}^{m_t} \left[-\frac{1}{t} Z' d(\cdot, y) \cdot Z d(\cdot, y) - \frac{d(\cdot, y)}{t} Z' Z d(\cdot, y) + \frac{Z' Z (\bar{\Sigma}_{t/2}^x(\cdot, y))}{\bar{\Sigma}_{t/2}^x(\cdot, y)} \right] \\ &\quad + \mathbb{E}^{m_t} \left[\left[-\frac{d(\cdot, y)}{t} Z d(\cdot, y) + Z (\log \bar{\Sigma}_{t/2}^x(\cdot, y)) \right] \cdot \left[-\frac{d(\cdot, y)}{t} Z' d(\cdot, y) + Z' (\log \bar{\Sigma}_{t/2}^x(\cdot, y)) \right] \right] \\ &\quad - \mathbb{E}^{m_t} \left[-\frac{d(\cdot, y)}{t} Z d(\cdot, y) + Z (\log \bar{\Sigma}_{t/2}^x(\cdot, y)) \right] \cdot \mathbb{E}^{m_t} \left[-\frac{d(\cdot, y)}{t} Z' d(\cdot, y) + Z' (\log \bar{\Sigma}_{t/2}^x(\cdot, y)) \right] \end{aligned}$$

Note (as already used above) that the Ben Arous asymptotics already give (subject to carrying out the relevant computations) the expansion of the log derivatives where that expansion is valid. In particular, if we multiply by t (which is already shown to be natural at the log-scale by the Leandre asymptotics), we have

$$(24) \quad \begin{aligned} \lim_{t \searrow 0} t \log p_t(x, y) &= -\frac{1}{4} d^2(x, y) \\ \lim_{t \searrow 0} t Z_y^\alpha \log p_t(x, y) &= -\frac{1}{4} Z_y^\alpha d^2(x, y) \end{aligned}$$

where Z^α is a multi-index derivative for a family (Z_1, \dots, Z_m) . Moreover, the first equation holds uniformly for all x and y in a compact, while the second holds uniformly, for any given Z^α , on a compact subset of M^2 that avoids the cut locus (including any abnormals and the sub-Riemannian diagonal). To see this, on such a set, the Ben Arous expansion gives that

$$t \log p_t(x, y) = -\frac{d}{2} t \log t - \frac{d^2(x, y)}{4} + t \log (c_0(x, y) + tR(t, x, y))$$

where R is some remainder function which is bounded along with all its derivatives as $t \rightarrow 0$. Since $c_0(x, y)$ is bounded, taking spatial derivatives gives

$$(25) \quad t Z_y^\alpha \log p_t(x, y) = -\frac{1}{4} Z_y^\alpha d^2(x, y) + t R'(t, x, y)$$

where R' is bounded along with all its derivatives as $t \rightarrow 0$.

However, the second equation in (24) need not hold on the cut locus. Instead, on the non-abnormal cut locus, we have, to leading order,

$$(26) \quad \begin{aligned} t \cdot Z_y \log p_t(x, y) &= \mathbb{E}^{m_t} [-d(\cdot, y) Z_y d(\cdot, y)] + O(t) \\ \text{and } t \cdot Z_y' Z_y \log p_t(x, y) &= \frac{1}{t} \left\{ \mathbb{E}^{m_t} [d^2(\cdot, y) Z_y d(\cdot, y) Z_y' d(\cdot, y)] \right. \\ &\quad \left. - \mathbb{E}^{m_t} [d(\cdot, y) Z_y d(\cdot, y)] \mathbb{E}^{m_t} [d(\cdot, y) Z_y' d(\cdot, y)] \right\} + O(1) \\ &= \frac{1}{t} \text{Cov}^{m_t} \left(d(\cdot, y) Z_y d(\cdot, y), d(\cdot, y) Z_y' d(\cdot, y) \right) + O(1). \end{aligned}$$

While the first two derivatives are the most interesting, one could, of course, compute the analogues of (21) and (23) for higher-order derivatives (see the proof of the following theorem). However, we content ourselves with an explicit expression for the leading order. Indeed, that is exactly Theorem 6, which we now prove.

Proof of Theorem 6. Faà di Bruno's formula implies that

$$(27) \quad Z^N \dots Z^1 \log p_t(x, y) = \sum_{\pi \in \Pi} \left(\frac{(-1)^{|\pi|-1} (|\pi| - 1)!}{p_t^{|\pi|}(x, y)} \prod_{B \in \pi} Z^B p_t(x, y) \right)$$

where the sum is over all partitions π of $\{N, N-1, \dots, 2, 1\}$, $|\pi|$ denotes the number of blocks in the partition π , the product is over all blocks B in π , and $Z^B p_t(x, y)$ means $Z^{k_m} \dots Z^{k_1} p_y(x, y)$ where $k_m > \dots > k_1$ are the elements of B . As above, we can use Molchanov's method to write derivatives of p_t as

$$Z^B p_t(x, y) = \int_{\Gamma_\varepsilon} \left(\frac{2}{t} \right)^d \Sigma_{t/2}(x, z) e^{-\frac{d^2(x, z)}{2t}} \cdot Z^B \left[e^{-\frac{d^2(z, y)}{2t}} \bar{\Sigma}_{t/2}^x(z, y) \right] d\mu(z).$$

Further, we see that

$$Z^B \left[e^{-\frac{d^2(z, y)}{2t}} \bar{\Sigma}_{t/2}^x(z, y) \right] = \sum_{I \subset B} Z^I \left[e^{-\frac{d^2(z, y)}{2t}} \right] \cdot Z^{I^c} \left[\bar{\Sigma}_{t/2}^x(z, y) \right],$$

where the sum is over all subsets I of B and I^c is the complement of I relative to B (if I is empty, we understand $Z^I \left[e^{-\frac{d^2(z, y)}{2t}} \right]$ to be $e^{-\frac{d^2(z, y)}{2t}}$ and similarly if I^c is empty). Finally, another application of Faà di Bruno's formula shows that

$$Z^I \left[e^{-\frac{d^2(z, y)}{2t}} \right] = e^{-\frac{d^2(z, y)}{2t}} \sum_{\pi' \in \Pi'} \left(-\frac{1}{2t} \right)^{|\pi'|} \prod_{B' \in \pi'} Z^{B'} [d^2(z, y)]$$

where the sum is over all partitions of I (and the vector fields are applied “in order” as above).

Combining the above is a bit messy. Nonetheless, we let $\pi(1), \dots, \pi(i), \dots$ enumerate the partitions of $\{N, N-1, \dots, 2, 1\}$, $B(j, i)$ enumerate the blocks of $\pi(i)$, $I(k, j, i)$ enumerate the subsets of

$B(j, i)$, $\pi'(\ell, k, j, i)$ enumerate the partitions of $I(k, j, i)$, and $B'(m, \ell, k, j, i)$ enumerate the blocks of $\pi'(\ell, k, j, i)$. Then if $c_i = (-1)^{|\pi_i|-1} (|\pi_i| - 1)!$, we have

$$(28) \quad Z^N \cdots Z^1 \log p_t(x, y) = \left(\sum_i c_i \right) \prod_j \sum_k \mathbb{E}^{m_t} \left[\frac{Z^{I^c(k, j, i)} \bar{\Sigma}_{t/2}^x(z, y)}{\bar{\Sigma}_{t/2}^x(z, y)} \cdot \sum_\ell \left(-\frac{1}{2t} \right)^{|\pi'_\ell|} \prod_m Z^{B'(m, \ell, k, j, i)} d^2(z, y) \right]$$

where if $I(k, j, i) = \emptyset$, the expectation is understood as simply $\mathbb{E}^{m_t} \left[Z^{I^c(k, j, i)} \bar{\Sigma}_{t/2}^x(z, y) / \bar{\Sigma}_{t/2}^x(z, y) \right]$ while if $I^c(k, j, i) = \emptyset$, we have $Z^{I^c(k, j, i)} \bar{\Sigma}_{t/2}^x(z, y) = Z^\emptyset \bar{\Sigma}_{t/2}^x(z, y) = \bar{\Sigma}_{t/2}^x(z, y)$.

Explicitly expanding this and collecting terms based on the power of $-1/(2t)$ is, fortunately, unnecessary. First, note that $\bar{\Sigma}_{t/2}^x(z, y)$ and $d^2(z, y)$ are both smooth on a neighborhood of Γ_ε (and $\bar{\Sigma}_{t/2}^x(z, y)$ is bounded from below by a positive constant), so that, after factoring out the $(-\frac{1}{2t})^{|\pi'_\ell|}$, all of the remaining expectations are finite, and moreover, bounded for all sufficiently small t solely in terms of bounds on $\bar{\Sigma}_{t/2}^x(z, y)$ and $d^2(z, y)$ and their first N derivatives (with respect to the Z^i). Further, we see that the largest power of $-1/(2t)$ we get in the expansion of the right-hand side of (28) is $(-\frac{1}{2t})^N$, which, for any given partition $\pi(i)$, occurs exactly when each $I(k, j, i) = B(j, i)$ and each $I(k, j, i)$ is partitioned into singletons. This gives

$$\begin{aligned} Z^N \cdots Z^1 \log p_t(x, y) &= \left(-\frac{1}{2t} \right)^N \sum_i c_i \prod_j \mathbb{E}^{m_t} \left[\prod_{k \in B(i, j)} Z^k d^2(z, y) \right] + O \left(\left(\frac{1}{2t} \right)^{N-1} \right) \\ &= \left(-\frac{1}{t} \right)^N \sum_i c_i \prod_j \mathbb{E}^{m_t} \left[\prod_{k \in B(i, j)} d(z, y) Z^k d(z, y) \right] + O \left(\left(\frac{1}{2t} \right)^{N-1} \right). \end{aligned}$$

Then the theorem follows after noting that the coefficient of $(-1/t)^N$ in this expression is exactly the formula for the joint cumulant of $d(z, y) Z^1 d(z, y), \dots, d(z, y) Z^N d(z, y)$ in terms of their joint (raw) moments. \square

Recall that the (first) cumulant of a single random variable is its expectation, while the cumulant of two random variables is their covariance (that is, $\kappa(X, Y) = \text{Cov}(X, Y)$), so that this generalizes the above results for $N = 1, 2$.

Theorem 36. *Let x and y be such that all minimal geodesics from x to y are strictly normal, and let $\{m_t : t \in (0, 1]\}$ be the family of probability measures defined by (20). Then this family is precompact in the topology of weak convergence, and in particular, for any sequence of times $t_n \rightarrow 0$, there is a subsequence $t_{n(i)}$ such that $m_{t_{n(i)}}$ converges weakly to a probability measure m_0 supported on Γ .*

Proof. By definition, the m_t are supported on $\bar{\Gamma}_\varepsilon$, which is compact (and which is thus a compact, separable metric space, when equipped with the metric inherited from M), so $\{m_t : t \in (0, 1]\}$ is tight. Thus the pre-compactness of the m_t (and resulting sequential compactness) follows from Prokhorov's theorem. Now let U_n be the subset of $\bar{\Gamma}_\varepsilon$ consisting of points x such that $d(\Gamma, x) > 1/n$. It's clear from Laplace asymptotics and (22) that, for any n , $m_t(U_n) \rightarrow 0$ as $t \rightarrow 0$ (indeed, this is implicit in the fact that all of the heat kernel asymptotics we've been considering are valid with respect to Γ_ε for any sufficiently small ε). Since U_n is open as a subset of $\bar{\Gamma}_\varepsilon$, the portmanteau theorem implies that $m_0(U_n) = 0$ for any limiting measure m_0 . Since n is arbitrary, this shows that $m_0(\Gamma) = 1$ for any limiting measure m_0 . \square

In the real-analytic case, one can say more, including that m_t converges. However, such results are most naturally discussed in connection with the bridge process, and we refer the reader to Theorem 9 for the convergence, and Sections 6.3 and 6.4 for the determination of m_0 in the A -type and Morse-Bott cases.

Observe that if $z \in \Gamma$, we have $d(z, y) = \frac{1}{2}d(x, y)$. Also, both $d(\cdot, y)$ and $Z_y d(\cdot, y)$, for any smooth vector field Z near y , are continuous, bounded functions on Γ_ϵ . Thus, since the cumulant can be written as a polynomial in products of such functions, if m_0 is a limiting measure and $t_n \searrow 0$ is a sequence of times corresponding to this m_0 , we have

$$(29) \quad \begin{aligned} \lim_{n \rightarrow \infty} t_n Z_y \log p_{t_n}(x, y) &= -\frac{1}{2}d(x, y) \mathbb{E}^{m_0} [Z_y d(\cdot, y)], \\ \lim_{n \rightarrow \infty} t_n^2 Z'_y Z_y \log p_{t_n}(x, y) &= \frac{d^2(x, y)}{4} \text{Cov}^{m_0} (Z_y d(\cdot, y), Z'_y d(\cdot, y)), \\ \text{and } \lim_{n \rightarrow \infty} t_n^N Z_y^N \cdots Z_y^1 \log p_{t_n}(x, y) &= \left(-\frac{d(x, y)}{2} \right)^N \kappa^{m_0} (Z_y^1 d(\cdot, y), \dots, Z_y^N d(\cdot, y)). \end{aligned}$$

Of course, if m_0 is a point mass, which is always the case if $y \notin \text{Cut}(x)$ and is possible also when $y \in \text{Cut}(x)$, then all of the cumulants after the first (the expectation) are zero. Nonetheless, the rate at which the variance goes to zero distinguishes the cut locus, as we now discuss.

5.2. Characterizing the cut locus. We know that if x and y are not cut, $t Z'_y Z_y \log p_t(x, y)$ converges as $t \searrow 0$. Our goal here is to prove that, conversely, if x and y are in the “non-abnormal” cut locus, then for any sequence of times going to zero, there is a subsequence t_n and a vector $Z \in T_y M$ such that $t_n Z_y^2 \log p_{t_n}(x, y)$ blows up at rate at least $t^{-1/2d}$. We begin with two preliminary lemmas.

Lemma 37. *Let x and y be such that all minimal geodesics from x to y are strictly normal. Then the map $\Gamma \rightarrow T_y^* M$ that takes $z \in \Gamma$ to $d_y d(z, y)$ (that is, the differential of $d(z, \cdot)$ at y) is a diffeomorphism onto its image.*

Proof. Recall that normal geodesics are given as the projections of curves in $T^* M$ under the Hamiltonian flow. In particular, let $e^{sH} : T^* M \rightarrow T^* M$ denote the time s Hamiltonian flow. Recall that $d(z, y)$ is constant for $z \in \Gamma$, so we can write this distance as $d(\Gamma, y)$. Since no point in Γ is in $\text{Cut}(y)$, for each $z \in \Gamma$ there is a unique $\lambda_z \in T_y^* M$ such that (the projection of) $e^{sH} \lambda_z$ for $s \in [0, d(\Gamma, y)]$ is the (unique) unit-speed minimal geodesic from y to z , and moreover, the map taking $z \in \Gamma$ to λ_z is a diffeomorphism onto its image. Since the Hamiltonian flow is reversible, for $z \in \Gamma$, the unique unit-speed minimal geodesic from z to y has terminal co-vector $-\lambda_z$. It follows from [2, Corollary 8.43] that for $z \in \Gamma$ we have $d_y d(z, y) = -\lambda_z$. This proves the lemma. \square

Lemma 38. *Let x and y be such that $y \in \text{Cut}(x)$ and all minimal geodesics from x to y are strongly normal. Suppose there is a sequence of times $t_n \searrow 0$ such that m_{t_n} converges to a point mass $m_0 = \delta_{z_0}$, for some $z_0 \in \Gamma$. Then x and y are conjugate along γ_{z_0} .*

Proof. Recalling that Γ parametrizes the minimal geodesics from x to y , [2, Theorem 8.72] implies that if z_0 is the only point in Γ , then γ_{z_0} is conjugate. Thus, we are left with the situation when there is at least one other point, which we denote w_0 , in Γ . Further, it is enough to show that if γ_{z_0} is not conjugate, then there is no sequence t_n such that m_{t_n} converges to δ_{z_0} . So assume that γ_{z_0} is not conjugate. Then there exist coordinates z_1, \dots, z_d defined on a neighborhood U of z_0 such that $h(z) = \sum_{i=1}^d z_i^2$ on U . Also, there exist coordinates w_1, \dots, w_d defined on a neighborhood V of w_0 such that $h(w) \leq \sum_{i=1}^d w_i^2$ on V , and we can assume that U and V are disjoint. Then we have that

$$\frac{m_t(V)}{m_t(U)} \geq \frac{\int_V (c_0(x, w) c_0(w, y) + O(t)) \frac{d\mu}{dw}(w) e^{-\sum_{i=1}^d w_i^2/t} dw}{\int_U (c_0(x, z) c_0(z, y) + O(t)) \frac{d\mu}{dz}(z) e^{-\sum_{i=1}^d z_i^2/t} dz}$$

and it follows from the basic Laplace asymptotics of [24] that the right-hand side is bounded from below by a positive constant as $t \rightarrow 0$. It follows that there is no sequence t_n such that m_{t_n} converges to δ_{z_0} , as desired. \square

We can now establish the basic estimate for the variance in (29) on the cut locus.

Theorem 39. *Let x and y be such that all minimal geodesics from x to y are strongly normal, and $y \in \text{Cut}(x)$. For any sequence of times going to 0, there is a subsequence t_n and a vector $Z \in T_y M$ such that, for any smooth extension of Z to a neighborhood of y ,*

$$\liminf_{n \rightarrow \infty} t_n^{-\frac{1}{2d}} [t_n Z_y Z_y \log p_{t_n}(x, y)] > 0,$$

and the value on the left-hand side depends only on Z and not on the choice of extension.

Proof. By Theorem 36, we know that for any sequence of times, after perhaps passing to a subsequence, the family m_{t_n} converges to a limiting probability measure, which we denote by m_0 , supported on Γ . Then, in light of (26), in order to prove Theorem 39, it is sufficient to show that there is some $Z \in T_y M$ such that

$$(30) \quad \liminf_{n \rightarrow \infty} t_n^{-\frac{1}{2d}} \text{Var}^{m_{t_n}}(Z_y d(\cdot, y)) > 0$$

(noticing also that the quantity on the left only depends on Z , and not the extension). There are two cases, depending on whether or not m_0 is a point mass, which we now treat. Further, in order to simplify the notation, we will simply write m_t and $t \rightarrow 0$ in place of m_{t_n} and $n \rightarrow \infty$, with the understanding that we always let t go to zero along an appropriate sequence of times, corresponding to m_0 .

Suppose m_0 is not a point mass (that is, it is not deterministic). Then 37 implies that the push-forward under the map $z \mapsto d_y d(z, y) \in T_y^* M$ is also not a point mass. Thus, because of the perfect pairing between $T_y M$ and $T_y^* M$, there exists some $Z \in T_y M$ such that the random variable $Z_y d(\cdot, y)$ is, under m_0 , not a.s./ constant, and thus, for this sequence of times and this Z ,

$$\liminf_{n \rightarrow \infty} \text{Var}^{m_{t_n}}(Z_y d(\cdot, y)) > 0,$$

which certainly implies (30).

The more interesting case is when m_0 is a point mass, which we now assume. In particular, we let $z_0 \in \Gamma$ be such that $m_0 = \delta_{z_0}$. By Lemma 38, we know that the minimal geodesic from x to y through z_0 is conjugate. Thus, by [11], we can choose coordinates (z_1, \dots, z_d) around z_0 so that $h(z) = h_{x,y}(z_1, \dots, z_d) \leq z_1^4 + \sum_{i=2}^d z_i^2$ on U , where $U \subset \mathbb{R}^d$ is a neighborhood of the origin contained in (the image of) Γ_ε . If we let $u(z) = c_0(x, z) c_0(z, y) \frac{d\mu}{dz}(z)$ on U , then u is a smooth, positive function on \bar{U} , so that it is bounded above and below by positive constants, say C and $1/C$ for some $C > 0$, and we have that

$$\begin{aligned} \phi_t(z) &= \frac{1}{\zeta(t)} \mathbf{1}_U(z) u(z) e^{-h(z)/t} dz \\ \text{and } \zeta(t) &= \int_U u(z) e^{-h(z)/t} dz \end{aligned}$$

is a family of probability densities (for $t > 0$) on \mathbb{R}^d supported on U . Let \tilde{m}_t be the probability measures determined by the densities $\phi_t(z)$ (and note that \tilde{m}_t is m_t conditioned to be in U).

We now show that we can restrict our attention to \tilde{m}_t . For fixed Z , for ease of notation, we temporarily let $f = Z d(\cdot, y)$ and $\alpha = \mathbb{E}^{m_t}[Z d(\cdot, y)]$. Then we have

$$\begin{aligned} \text{Var}^{m_t}(f) &= \mathbb{E}^{m_t}[(f - \alpha)^2] \\ &\geq \mathbb{E}^{m_t}[\mathbf{1}_U (f - \alpha)^2] \\ &= m_t(U) \mathbb{E}^{\tilde{m}_t}[(f - \alpha)^2] \\ &\geq m_t(U) \text{Var}^{\tilde{m}_t}(f), \end{aligned}$$

where we've used that $(f - \alpha)^2$ is non-negative and the variance of a random variable is the best L^2 -approximation by a constant. Since $m_t(U) \rightarrow 1$ by assumption, it is enough to show that (30) holds for \tilde{m}_t , rather than for m_t itself.

We now recall some basic facts about entropy. If φ is a probability density function on \mathbb{R}^d , we let

$$H(\varphi) = - \int_{\mathbb{R}^d} \varphi(x) \log \varphi(x) dx = \mathbb{E}^\varphi [-\log \varphi]$$

be the (differential) entropy. Further, let $Q(\varphi)$ be the covariance matrix of φ (or equivalently, the covariance of the identity function under the probability measure on \mathbb{R}^d determined by φ). Then we have the entropy inequality

$$H(\varphi) \leq \frac{1}{2} \log [(2\pi e)^d \det Q(\varphi)]$$

$$\text{which implies } \det Q(\varphi) \geq \frac{1}{(2\pi e)^d} e^{2H(\varphi)}.$$

(See [43] for instance.)

Returning to the case at hand, we have a one-parameter family of probability densities ϕ_t , given by the above, where we view U as a subset of \mathbb{R}^d . We can estimate the entropy of ϕ_t by

$$\begin{aligned} H(\phi_t) &= \mathbb{E}^{\phi_t} [-\log \phi_t] \\ &= \mathbb{E}^{\phi_t} \left[\frac{h(z)}{t} - \log(u(z)) \right] + \mathbb{E}^{\phi_t} [\log \zeta(t)] \\ &\geq \log \zeta(t) + \log C, \end{aligned}$$

where we've used that $h(z)$ is positive. Next, we have that

$$\zeta(t) \geq \int_U u(z) e^{-(z_1^4 + z_2^2 + \dots + z_d^2)/t} dz$$

and so Laplace integral asymptotics (see [24]) show that there is a positive constant C' such that

$$\log \zeta(t) \geq \log \left(C' \left(t^{1/4} + \prod_{i=2}^d t^{1/2} \right) \right) = \log \left(C' t^{\frac{d}{2} - \frac{1}{4}} \right)$$

for all sufficiently small t (so $\log \zeta(t)$ can go to $-\infty$, but at a controlled rate). Using this in the entropy inequality, we find that, for some constant $C'' > 0$,

$$\det Q(\phi_t) > C'' t^{d - \frac{1}{2}}.$$

Now $\det Q(\phi_t)$ is the product of the d eigenvalues of the covariance of ϕ_t , and it follows that, for all sufficiently small t , there is at least one eigenvalue which is greater than $C''' t^{1 - \frac{1}{2d}}$. Since we can choose the corresponding eigenvector to be a unit vector in the z_1, \dots, z_d coordinates, by compactness, there exists a linear random variable $v = c_1 z_1 + \dots + c_d z_d \in \mathbb{R}^d \simeq T_{z_0} M$ with $c_1^2 + \dots + c_d^2 = 1$ such that $\liminf_{t \rightarrow 0} t^{1 - \frac{1}{2d}} \text{Var}^{\tilde{m}_t}(v) > 0$. By Lemma 38, this implies that there is a vector $Z \in T_y M$ such that

$$\liminf_{n \rightarrow \infty} t_n^{-\frac{1}{2d}} \text{Var}^{m_{t_n}}(Z_y d(\cdot, y)) > 0,$$

which concludes the proof. \square

If we wish to consider sets of vector fields, then because we use Lie derivatives, we need to control the size of their derivatives as well as the size of the vectors themselves (this will be especially true in the next section). With this in mind, we say that a set \mathcal{Z} of (smooth) vector fields defined on a neighborhood of a compact set K is C^m -bounded on K if any $z \in K$ has a neighborhood U such that, for any system of coordinates on U , the C^m -norm of $Z \in \mathcal{Z}$, restricted to U and with respect to this system of coordinate, is uniformly bounded over \mathcal{Z} . Note that if the C^m -norm of $Z \in \mathcal{Z}$ restricted U is uniformly bounded with respect to one system of coordinates on U , then it is also uniformly bounded with respect to any other system of coordinates on U which extends to a neighborhood of \overline{U} .

We now have the natural context in which to state and prove the characterization of the (non-abnormal) sub-Riemannian cut locus, which was given as Corollary 7.

Proof of Corollary 7. If $y \notin \text{Cut}(x)$, then the C^1 -boundedness implies that $Z_y Z_y d^2(x, y)$ is uniformly bounded for $Z \in \mathcal{Z}$, which in light of (24), completes the proof in that case. If $y \in \text{Cut}(x)$, let t_n be any sequence of times going to 0. After possibly passing to a subsequence, the fact that $\mathcal{Z}|_{T_y M}$ contains a neighborhood of the origin means that there is some $c > 0$ (possibly 1) such that $cZ \in \mathcal{Z}$, where Z is the vector (field) from Theorem 39. By linearity of differentiation, we see that any sequence of times going to zero has a subsequence t_n such that

$$\lim_{n \rightarrow \infty} \left[\sup_{Z \in \mathcal{Z}} t_n Z_y Z_y \log p_{t_n}(x, y) \right] = \infty,$$

and this gives the desired result. \square

Finally, in the Riemannian case, there is no need to avoid abnormal minimizers, and we can work with covariant derivatives.

Corollary 40. *Let M be a (complete) Riemannian manifold, and x and y any two points in M . Then $y \notin \text{Cut}(x)$ if and only if*

$$\limsup_{t \searrow 0} \left[\sup_{\substack{Z \in T_y M \\ \|Z\|=1}} t |\nabla_{Z, Z}^2 \log p_t(x, y)| \right] < \infty,$$

and $y \in \text{Cut}(x)$ if and only if

$$\lim_{t \searrow 0} \left[\sup_{\substack{Z \in T_y M \\ \|Z\|=1}} t \nabla_{Z, Z}^2 \log p_t(x, y) \right] = \infty,$$

where ∇^2 is the (covariant) Hessian, acting on the y -variable.

5.3. Sheu-Hsu-Stroock-Turetsky type bounds. For a compact Riemannian manifold M , with the Riemannian volume and Laplace-Beltrami operator, a result of Stroock-Turetsky and Hsu, improving an earlier result of Sheu, is that, for each N , there exists a constant C_N depending on M and N such that, for all $(t, x, y) \in (0, 1] \times M \times M$,

$$(31) \quad |\nabla^N \log p_t(x, y)| \leq C_N \left(\frac{d(x, y)}{t} + \frac{1}{\sqrt{t}} \right)^N$$

which then implies that, for each N , there exists a constant D_N depending on M and N such that, for all $(t, x, y) \in (0, 1] \times M \times M$,

$$|\nabla^N p_t(x, y)| \leq D_N \left(\frac{d(x, y)}{t} + \frac{1}{\sqrt{t}} \right)^N p_t(x, y).$$

(In both cases, the differentiation is in the y -variable.)

Note that the $\frac{1}{\sqrt{t}}$ term is only relevant near the diagonal, since on any set where $d(x, y)$ is bounded from below by a positive constant, the $\frac{d(x, y)}{t}$ term dominates. On a (strictly) sub-Riemannian manifold, the diagonal is abnormal, and uniform bounds even for the heat kernel itself appear not to be generally known.

In light of this, we see that the natural generalization of to sub-Riemannian manifolds, using the Molchanov approach as above, is the following.

Theorem 41. *Let $\mathcal{K} \in M^2 \setminus \mathcal{D}$ be a compact subset such that all length minimizers between pairs $(x, y) \in \mathcal{K}$ are strongly normal. Let N be a positive integer and let \mathcal{Z} be a set of vector fields on a neighborhood of the projection of \mathcal{K} onto the second component which is C^{N-1} -bounded. Then there*

exist constants C_N and D_N , depending on M , \mathcal{K} , \mathcal{Z} , and N , such that, for all $t \in (0, 1]$ and $(x, y) \in \mathcal{K}$, and for all $Z^1, \dots, Z^N \in \mathcal{Z}$, we have

$$\begin{aligned} |Z^N \cdots Z^1 \log p_t(x, y)| &\leq \frac{C_N}{t^N} \\ \text{and} \quad |Z^N \cdots Z^1 p_t(x, y)| &\leq \frac{D_N}{t^N} p_t(x, y), \end{aligned}$$

where, as usual, the derivatives act on the y -variable.

(Note that, on a set \mathcal{K} as in the theorem, $d(x, y)$ is bounded from above and below by positive constants.)

Proof. As already noted in the proof of Theorem 6, equation (28) shows that $Z^N \cdots Z^1 \log p_t(x, y)$ can be expanded in powers of $1/t$, up to the N th power, with coefficients given in terms of (products of) the expectations of $\bar{\Sigma}_{t/2}^x(z, y)$ and $d^2(z, y)$ and their first N derivatives (with respect to the Z^i). Further, K is chosen so that, for small enough ε , $\bar{\Sigma}_{t/2}^x(z, y)$ and $d^2(z, y)$ are smooth on a neighborhood of the compact set

$$\Gamma^r(\mathcal{K}) = \{(z, y) \mid \exists x \in M \text{ s.t. } (x, y) \in \mathcal{K}, z \in \Gamma_\varepsilon(x, y)\}.$$

(See Lemma 18.) This, plus the definition of C^{N-1} -boundedness, implies that these expectations can be uniformly bounded over \mathcal{K} (recall that for each (x, y) , the corresponding probability measure is supported on $\Gamma_\varepsilon(x, y)$). This gives the result for $\log p_t(x, y)$. Then the result for p_t itself follows from this and Faà di Bruno's formula for the exponential. \square

The bounds on derivatives of p_t itself should be compared to those of Proposition 5. Indeed, using the upper bound on p_t from Proposition 5 in Theorem 41 implies the spatial derivative bounds in Proposition 5. At a single pair (x, y) , one should compare with Corollary C: Molchanov Laplace.

In contrast, on a Riemannian manifold, given any compact subset of the diagonal, there is a neighborhood on which the Ben Arous expansion holds, and can be differentiated. With this in mind, we have the following.

Theorem 42. *Let M be a (complete, but not necessarily compact) Riemannian manifold. Then for any compact set $K \subset M$ and any positive integer N , there exist positive constants δ , C_N and D_N , depending on K , M , and N , such that*

$$\begin{aligned} |\nabla \log p_t(x, y)| &\leq \begin{cases} C_1 \left(\frac{d(x, y)}{t} + 1 \right) & \text{for } d(x, y) \leq \delta \\ \frac{C_1}{t} & \text{for } d(x, y) > \delta \end{cases}, \\ |\nabla^N \log p_t(x, y)| &\leq \begin{cases} \frac{C_N}{t} & \text{for } d(x, y) \leq \delta \\ \frac{C_N}{t^N} & \text{for } d(x, y) > \delta \end{cases} \quad (\text{for } N \geq 2), \\ \text{and} \quad |\nabla^N p_t(x, y)| &\leq D_N \left(\frac{d(x, y)}{t} + \frac{1}{\sqrt{t}} \right)^N p_t(x, y). \end{aligned}$$

for all $(t, x, y) \in (0, 1] \times K \times K$.

Proof. For any m , covariant derivatives can be realized by C^m -bounded vector fields on K , so we can assume that such Z^i have been chosen. As already noted, we can find δ such that the Ben Arous expansion holds for $x, y \in K$ with $d(x, y) \leq \delta$. Thus the estimates for $\log p_t(x, y)$ when $d(x, y) \leq \delta$ follow from (25) and compactness. Then the estimates for $d(x, y) > \delta$ come from the previous theorem.

Again, the results for p_t itself follow. \square

Remark 43. The result for derivatives of $\log p_t(x, y)$ near the diagonal is stronger than that of (31). Nonetheless, for p_t itself, this is the correct form of the estimate, as can be seen by differentiating the Ben Arous expansion directly.

6. LAW OF LARGE NUMBERS

We finally turn to a more stochastic topic, the law of large numbers (LLN) for the bridge process associated to the diffusion X_t , which is essentially the “leading term” of the small-time asymptotics of the bridge process. We begin with the basic definitions needed to state and prove the results.

Let $\Omega_M^{[0,t]}$ be the space of continuous paths $\omega_\tau : [0, t] \rightarrow M$, for some $t \in (0, \infty)$. We define a metric

$$d_{\Omega_M^{[0,t]}}(\omega, \tilde{\omega}) = \frac{\sup_{0 \leq \tau \leq t} d(\omega_\tau, \tilde{\omega}_\tau)}{1 + \sup_{0 \leq \tau \leq t} d(\omega_\tau, \tilde{\omega}_\tau)}$$

on $\Omega_M^{[0,t]}$. This metric makes $\Omega_M^{[0,t]}$ into a Polish space and the topology is that of uniform convergence. We let $\Omega_M = \Omega_M^{[0,1]}$ and note that the map $\Omega_M^{[0,t]} \rightarrow \Omega_M$, $\omega_\tau \mapsto \omega_{\tau/t}$ is an isometry.

For x and y in M and $t \in (0, \infty)$, we can consider the bridge process $X_\tau^{x,y,t}$, which is the diffusion started from x and conditioned to be at y at time t . The law of this process is a probability measure $\hat{\mu}^{x,y,t}$ on $\Omega_M^{[0,t]}$. We let $\mu^{x,y,t}$ be the pushforward of $\hat{\mu}^{x,y,t}$ to Ω_M , that is, $\mu^{x,y,t}$ is the law of the bridge process rescaled to take unit time. For fixed x and y , this gives a family of probability measures on Ω_M , and we are interested in the weak convergence of these measures as $t \searrow 0$. A result determining such convergence is generally called a law of large numbers for the bridge process.

Again for fixed x and y , recall that Γ parametrizes the set of minimal geodesics from x to y . More precisely, for any $z \in \Gamma$, let γ_τ^z be the constant speed geodesic going from x to y in unit time, through z (so that $\gamma_{1/2}^z = z$). Then this gives an embedding of Γ into Ω_M , and we write the image as $\tilde{\Gamma}$.

6.1. Extension to the cut locus. Bailleul and Norris [9] proved the law of large numbers in the case when there is a single minimizer from x to y . In that case, if z is the unique point in Γ , $\mu^{x,y,t}$ converges (weakly) to a point mass at γ^z as $t \searrow 0$. The idea of the Molchanov technique in this context is to determine a law of large numbers on the (non-abnormal) cut locus by conditioning on the midpoint of the bridge and “gluing together” the result of Bailleul and Norris for the first and second halves of the path.

The connection with the previous heat kernel asymptotics is as follows. For each t , let ν_t be the pushforward of $\mu^{x,y,t}$ under the map $\Omega_M \rightarrow M$, $\omega_\tau \mapsto \omega_{1/2}$, so that ν_t is the distribution of the midpoint of the bridge process.

Lemma 44. *Let x and y be points of M such that all minimizers from x to y are strongly normal, let ν_t be as above, and let m_t be defined by (20). Then for a sequence of times t_n going to 0, ν_{t_n} converges if and only if m_{t_n} does, in which case they have the same limit.*

Proof. In terms of p_t , the density of $X_\tau^{x,y,t}$ at time $t/2$ is given by

$$\frac{d\nu_t}{d\mu}(z) = \frac{p_{t/2}(x, z)p_{t/2}(z, y)}{p_t(x, y)}.$$

But we know that the error induced by restricting z to Γ_ε and by replacing the heat kernel by the leading term of its expansion goes to 0 with t . Comparing with Equation (22), we see that for any smooth f , $\mathbb{E}^{m_t}[f] - \mathbb{E}^{\nu_t}[f] \rightarrow 0$ as $t \rightarrow 0$. But this proves the result. \square

Suppose that for some sequence of times $t_n \searrow 0$, $m_{t_n} \rightarrow m_0$ for some m_0 supported on Γ (recall that $\{m_t : t \in (0, 1]\}$ is subsequentially compact). Then m_0 maps to a probability measure \tilde{m}_0 on $\tilde{\Gamma}$ under the inclusion of Γ into Ω_M . (Of course, m_0 can be recovered from \tilde{m}_0 by the inverse map $\tilde{\Gamma} \rightarrow \Gamma$.)

It turns out that we need a uniform version of Bailleul and Norris’ law of large numbers. Namely, consider $z \in \Gamma_\varepsilon$, and let $g^z = g_\tau^z$ be the unique constant-speed minimal geodesic traveling from x to z in unit time. Then $\mu^{x,z,t}$ converges to the point mass at g^z . We require the additional fact that this convergence is uniform, in a suitable sense, over such z . This is only a slight extension of Bailleul and Norris’ work, so we simply indicate the argument. It is essentially a space-time version of an argument from [32].

Recall that Léandre asymptotics hold uniformly on some large compact containing x and Γ_ε . Let X_τ be the diffusion started from x . At $\tau = 0$, $X_0 = g_0^z = x$. For some small $\varepsilon' > 0$, let σ be the first time $d_M(X_\tau, g_\tau^z)$ hits ε' . We wish to decompose σ according to when it occurs and $d(X_\sigma, x)$ and $d(X_\sigma, z)$. So for some small $\delta' > 0$, partition $[0, t]$ by some sequence of reals $0 = s_0 < s_1 < \dots < s_N = t$ with $|s_{i+1} - s_i| < \delta'$, and similarly partition $[0, \infty)$ by some sequence of reals $0 = \rho_0 < \rho_1 < \dots < \rho_j < \dots$ with $|\rho_{j+1} - \rho_j| < \delta'$ and $\rho_j \rightarrow \infty$. Then let $\sigma_{i,j}$ be the first time $d_M(X_\tau, g_\tau^z)$ hits ε' such that $s_i \leq \sigma < s_{i+1}$ and $\rho_j \leq d(X_\sigma, x) < \rho_{j+1}$. Note that σ is the minimum of the $\sigma_{i,j}$ and that only finitely many $\sigma_{i,j}$ are necessary.

We can bound the contribution of paths with $\sigma_{i,j} \leq t$ to $p_t(x, z)$. For any $\delta > 0$, for small enough all t , the probability that $\sigma_{i,j} < t$ is less than

$$\exp \left[-\frac{\rho_j^2 - \delta}{4s_{i+1}} \right]$$

(this necessarily means $s_{i+1} < t$). We see this just as in the proof of Theorem 1 (it is an exit time, after all); indeed, the reason for introducing that $\sigma_{i,j}$ is that we cannot estimate the density of (X_σ, s) , but we can estimate this discretized version. If $\rho_j^2 \leq \delta$, which happens when $X_{\sigma_{i,j}}$ is close to x , this estimate is meaningless, but in that case, we just bound the probability by 1. Continuing, we have the Léandre estimate on the heat kernel, so that, for small enough t ,

$$p_{t-\sigma_{i,j}}(X_{\sigma_{i,j}}, z) < \exp \left[-\frac{d^2(X_{\sigma_{i,j}}, z) - \delta}{4(t - s_i)} \right].$$

Then we use integration by parts as in the proof of Theorem 1 to see that, as long as δ is small enough compared to ε' and δ' is small enough compared to δ and ε' , for each $\sigma_{i,j}$ the contribution to $p_t(x, y)$ from paths with $\sigma_{i,j} < t$ is less than

$$\exp \left[-\frac{\rho_j^2 - \delta}{4s_{i+1}} - \frac{d^2(X_{\sigma_{i,j}}, z) - \delta}{4(t - s_i)} \right]$$

whenever t is small enough. But σ is defined so that, for small enough δ , ε' , and δ' and considering Lemma 10, this is always “exponentially smaller” than $p_t(x, y)$, since for small enough t , $p_t(x, y) > Ct^{-d/2} \exp[-d^2(x, z)/(4t)]$ for some $C > 0$, by Theorem 3. Summing over the finitely many required i, j , we see that as $t \searrow 0$, the contribution to $p_t(x, y)$ of paths with $\sigma < t$ is exponentially smaller than $p_t(x, y)$.

Moreover, we see that this argument is uniform in z . That is, since $\bar{\Gamma}_\varepsilon$ is compact, the estimates on the probability that $\sigma_{i,j} < t$ and on the heat kernel are uniform over $z \in \Gamma_\varepsilon$. Also, because g^z varies smoothly in z (this is where we use the fact that $z \in \Gamma_\varepsilon$ is not in the cut locus of x , etc.) and the distance function is continuous, the geometry underlying the fact that $\sigma < t$ implies that paths are exponentially negligible (again referring to Lemma 10) is uniformly controlled for $z \in \Gamma_\varepsilon$. Recalling the definition of σ , this says that the $\mu^{x,z,t}$ -probability of an ε' -ball around g^z goes to 1 as t goes to 0, uniformly in z . This is the desired law of large numbers for $\mu^{x,z,t}$, uniformly in z . (And note that the same applies to $\mu^{z,y,t}$, uniformly in z .)

We can now prove our law of large numbers for the bridge process, as given in Theorem 8.

Proof of Theorem 8. Assume that m_{t_n} converges to m_0 . For simplicity of notation, we will assume that $m_t \rightarrow m_0$, with the general case following by restricting to a subsequence. We know that, under $\mu^{x,y,t}$, $\omega|_{[0,t/2]}$ and $\omega|_{[t/2,t]}$ are conditionally independent given $\omega_{t/2}$, by the Markov property. Thus, we can decompose (or disintegrate) $\mu^{x,y,t}$ by first drawing z from ν_t and then drawing $\omega|_{[0,t/2]}$ and $\omega|_{[t/2,t]}$ independently from $\mu^{x,z,t/2}$ and $\mu^{z,y,t/2}$, respectively.

Let $F : \Omega_M \rightarrow \mathbb{R}$ be Lipschitz continuous and bounded. Let $f : M \rightarrow \mathbb{R}$ be as follows. For any $z \in \Gamma_\varepsilon$, let γ^z be the (possibly) broken geodesic which travels the minimal geodesic from x to z at constant speed in time $1/2$, and then travels the minimal geodesic from z to y at constant speed in time $1/2$. Note that this is well defined and agrees with our earlier definition when $z \in \Gamma$. On Γ_ε , let $f(z) = F(\gamma^z)$; since this is bounded and continuous (recall that γ^z is continuous in z by the smoothness

of the exponential map), we can extend it to a bounded and continuous function on M in an arbitrary way. Now choose $\delta > 0$. By weak convergence of m_t and Lemma 44, for all small enough t , we have

$$|\mathbb{E}^{\nu_t} [f] - \mathbb{E}^{m_0} [f]| < \delta.$$

Next, by the uniform law of large numbers just discussed, applied to the “two halves” of γ^z (that is, for $\mu^{x,z,t/2}$ and $\mu^{z,y,t/2}$), and the fact that F is Lipschitz and bounded, we have that for all small enough t ,

$$\left| \mathbb{E}^{\mu^{x,z,t/2} \otimes \mu^{z,y,t/2}} [F] - f(z) \right| < \delta$$

uniformly for all $z \in \Gamma_\varepsilon$. But in light of the above decomposition of $\mu^{x,y,t}$, these inequalities imply that, for all sufficiently small t ,

$$\left| \mathbb{E}^{\mu^{x,y,t}} [F] - \mathbb{E}^{m_0} [f] \right| < 2\delta.$$

Recalling the definition of f and that δ is arbitrary, it follows from the portmanteau theorem that $\mu^{x,y,t} \rightarrow \tilde{m}_0$.

For the other direction, assume that μ^{x,y,t_n} converges to some μ_0 . Then ν_{t_n} converges to the pushforward of μ_0 under the evaluation at $\tau = 1/2$, and in particular, m_{t_n} also converges, by Lemma 44. Then applying the part of the theorem (just proven) when m_{t_n} converges to some m_0 shows that $\mu_0 = \tilde{m}_0$. This completes the proof. \square

6.2. Real analytic methods. We can give more precise results in the real-analytic case. This is essentially a direct application of results from Laplace asymptotics to the behavior of m_t . Hsu [31] already gave this application for Brownian motion on real-analytic Riemannian manifolds (having established Theorem 8 in the Riemannian case via large deviations), so we simply summarize his results, for completeness and to emphasize that they hold in the present sub-Riemannian context as well.

Suppose that every $z \in \Gamma$ is contained in a coordinate patch such that $h_{x,y}$ is real-analytic (in the coordinates— that is, there exists a local real-analytic stiffening of the structure with this property). Then for any $z \in \Gamma$, there is a rational $\alpha(z) \in [d/2, d - (1/2)]$, a non-negative integer $\beta(z)$, and an $r_0 > 0$ such that, for any open ball $B_z(r)$ around z with radius $r \in (0, r_0)$, we have

$$(32) \quad \int_{B_z(r)} e^{-\frac{h_{x,y}(u) - h_{x,y}(z)}{4t}} \mu(du) \sim \frac{C}{t^\alpha} \left(\log \frac{1}{t} \right)^\beta$$

where C is some positive constant depending on z and r . If we put the lexicographical order on $\mathbb{Q} \times \mathbb{Z}$, so that $(\alpha_1, \beta_1) < (\alpha_2, \beta_2)$ if either $\alpha_1 < \alpha_2$ or $\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$, then we see that $(\alpha(z_1), \beta(z_1)) < (\alpha(z_2), \beta(z_2))$ means that the integral (32) around z_2 dominates the integral around z_1 as $t \searrow 0$. Moreover, the resulting map $\Gamma \rightarrow \mathbb{Q} \times \mathbb{Z}$ is upper semi-continuous, and since Γ is compact, this means that $(\alpha(z), \beta(z))$ attains its maximum, which we denote (α_m, β_m) . We let

$$\Gamma_{x,y}^m = \Gamma^m = \{z \in \Gamma : (\alpha(z), \beta(z)) = (\alpha_m, \beta_m)\}$$

and note that Γ^m is a non-empty, closed subset of Γ (corresponding to geodesics of “maximal degeneracy”).

The significance of these considerations is given by Theorem 9, and we indicate its proof.

Proof of Theorem 9. The convergence of m_t to a measure with support Γ^m follows from the definition of m_t and the expansions (32); the details are given in the proofs of Theorems 4.1 and 4.2 in [31]. \square

Finally, we can identify Γ^m and m_0 more explicitly if we have more information on the normal form of $h_{x,y}$. We illustrate this with the two most important special cases.

6.3. LLN for A -type singularities. In parallel to Section 4, we give an explicit treatment of the asymptotics in two cases— the case when each minimal geodesic is A_n -conjugate (in this section) and the Morse-Bott case (in the next).

As usual, let x and y be distinct points such that every minimal geodesic from x to y is strongly normal. Further, we assume that there is some $\ell \in \{1, 3, 5, \dots\}$ such that for every $z \in \Gamma$, γ^z is A_m -conjugate for $1 \leq m \leq \ell$ and that there is at least one $z \in \Gamma$ for which γ_z is A_ℓ -conjugate. We refer to Section 4.1 for the relevant results about the normal form of $h_{x,y}$ and the resulting leading term in the expansion coming from each geodesic. In particular, let z_1, \dots, z_N be the points of Γ corresponding to A_ℓ -conjugate geodesics, and around each z_i let $(u_{i,1}, \dots, u_{i,d})$ be local coordinates diagonalizing $h_{x,y}$ as in Equation (9). Then m_t converges, and the limit is given by

$$m_0 = \frac{\sum_{i=1}^N c_0(x, z_i) c_0(z_i, y) \frac{d\mu}{d(u_{i,1}, \dots, u_{i,d})}(z_i) \cdot \delta_{z_i}}{\sum_{i=1}^N c_0(x, z_i) c_0(z_i, y) \frac{d\mu}{d(u_{i,1}, \dots, u_{i,d})}(z_i)}.$$

(To see this, integrate any smooth f against m_t and take the leading term of the resulting Laplace asymptotics.) In particular, $\Gamma^m = \{z_1, \dots, z_N\}$, and it may certainly be a proper subset of Γ .

However, note that if none of the minimal geodesics from x to y is conjugate (which in this terminology means being “ A_1 -conjugate” and implies that $h_{x,y} - d^2(x, y)/4$ can be written as a sum-of-squares around each z_i), then $\Gamma^m = \Gamma$. In particular, if M is a Riemannian manifold of non-positive sectional curvature, this is the only possibility. Indeed, in this case, the asymptotics of $p_t(x, y)$ can be written directly in terms of the Ben Arous expansion applied to the universal cover, and a slightly simpler formula for m_0 can be given; see Example 3.7 of [31] where the case of only non-conjugate geodesics is treated for a compact Riemannian manifold.

6.4. LLN for the Morse-Bott case. Again, we let x and y be distinct points such that every minimal geodesic from x to y is strongly normal, but now, as in Section 4.2, we assume that Γ is an r -dimensional submanifold (where necessarily we have $r < k$ and we recall that Γ is compact) and that the kernel of the differential of the exponential map has dimension r at γ^z for any $z \in \Gamma$. Then around any point of Γ we can find local coordinates (u_1, \dots, u_k) such that Γ is (locally) given by $u_{r+1} = \dots = u_k = 0$, (u_1, \dots, u_r) gives (local) coordinates on Γ , and

$$h_{x,y} = \frac{d^2(x, y)}{4} + u_{r+1}^2 + \dots + u_d^2.$$

In this case, another use of Laplace asymptotics to integrate any smooth f against m_t shows that m_t converges and m_0 has a smooth, non-vanishing density on Γ with respect to any local coordinates. Hence $\Gamma^m = \Gamma$. Moreover, the density of m_0 with respect to $du_1 \dots du_r$ as above can be written in terms of the density of μ , the Hessian of $h_{x,y}$ along the normal bundle over Γ , and the c_0 (see Section 3 of [13] for the basic framework of the computation), but the expression is messy and unenlightening, so we omit it. Instead, we note that the Morse-Bott case typically arises when M possess some rotational symmetry, in which case m_0 can be deduced via symmetry arguments. That is, let $\text{Iso}_{x,y}$ be the subgroup of the isometry group of M that fixes x and y , where isometries must also preserve μ and the sub-Laplacian. Suppose that $\text{Iso}_{x,y}$ acts transitively on Γ . Then m_0 must be the uniform probability measure on Γ , in the sense that m_0 is the unique probability measure on Γ invariant under the action of $\text{Iso}_{x,y}$.

For example, if M is the standard Riemannian sphere with the Laplace-Beltrami operator and Riemannian volume, and x and y are antipodal points, m_0 is the uniform probability measure on the equator, as observed in Example 3.6 of [31].

The natural sub-Riemannian analogue is the Heisenberg group. By symmetry, we can take x to be the origin (using \mathbb{R}^3 to give global coordinates, in the usual way). Then y is in the (non-abnormal) cut locus exactly when $y = (0, 0, h)$ for some $h \neq 0$, in which case Γ is a circle, invariant under rotation around the vertical axis (see Figure 3 again). We see that m_0 is the uniform probability measure on Γ .

REFERENCES

- [1] A. A. Agrachev, *Exponential mappings for contact sub-Riemannian structures*, J. Dynam. Control Systems **2** (1996), no. 3, 321–358. MR 1403262
- [2] Andrei Agrachev, Davide Barilari, and Ugo Boscain, *A comprehensive introduction to sub-Riemannian geometry*, Cambridge University Press, 2018.
- [3] V. I. Arnold, S. M. Guseĭn Zade, and A. N. Varchenko, *Singularities of differentiable maps. Vol. I*, Monographs in Mathematics, vol. 82, Birkhäuser Boston, Inc., Boston, MA, 1985, The classification of critical points, caustics and wave fronts, Translated from the Russian by Ian Porteous and Mark Reynolds. MR 777682
- [4] ———, *Singularities of differentiable maps. Vol. II*, Monographs in Mathematics, vol. 82, Birkhäuser Boston, Inc., Boston, MA, 1985, The classification of critical points, caustics and wave fronts, Translated from the Russian by Ian Porteous and Mark Reynolds. MR 777682
- [5] Malva Asaad and Maria Gordina, *Hypoelliptic heat kernels on nilpotent Lie groups*, Potential Anal. **45** (2016), no. 2, 355–386. MR 3518678
- [6] Robert Azencott, *Un problème posé par le passage des estimées locales aux estimées globales pour la densité d’une diffusion*, Asterisque (1981), no. 84, 131–150.
- [7] Ismaël Bailleul, *Large deviation principle for bridges of sub-Riemannian diffusion processes*, Séminaire de Probabilités XLVIII, Lecture Notes in Math., vol. 2168, Springer, Cham, 2016, pp. 189–198. MR 3618130
- [8] Ismael Bailleul, Laurent Mesnager, and James Norris, *Small-time fluctuations for the bridge of a sub-riemannian diffusion*, preprint, arXiv:1505.03464 (2015), To appear in Annales de l’École Normale Supérieure.
- [9] Ismael Bailleul and James Norris, *Diffusion in small time in incomplete sub-riemannian manifolds*, preprint, arXiv:1810.06328 (2018), To appear in Analysis & PDEs.
- [10] D. Barilari, *Trace heat kernel asymptotics in 3D contact sub-Riemannian geometry*, J. Math. Sci. (N.Y.) **195** (2013), no. 3, 391–411, Translation of Sovrem. Mat. Prilozh. No. 82 (2012). MR 3207127
- [11] Davide Barilari, Ugo Boscain, Grégoire Charlot, and Robert W. Neel, *On the heat diffusion for generic Riemannian and sub-Riemannian structures*, International Mathematics Research Notices **2017** (2016), no. 15, 4639–4672.
- [12] Davide Barilari, Ugo Boscain, and Robert W. Neel, *Small-time heat kernel asymptotics at the sub-Riemannian cut locus*, J. Differential Geom. **92** (2012), no. 3, 373–416. MR 3005058
- [13] ———, *Heat kernel asymptotics on sub-Riemannian manifolds with symmetries and applications to the bi-Heisenberg group*, Ann. Fac. Sci. Toulouse Math. (6) **28** (2019), no. 4, 707–732. MR 4045424
- [14] Davide Barilari and Luca Rizzi, *Sub-Riemannian interpolation inequalities*, Invent. Math. **215** (2019), no. 3, 977–1038. MR 3935035
- [15] Fabrice Baudoin and Michel Bonnefont, *The subelliptic heat kernel on $SU(2)$: representations, asymptotics and gradient bounds*, Math. Z. **263** (2009), no. 3, 647–672. MR 2545862
- [16] Fabrice Baudoin and Jing Wang, *The subelliptic heat kernel on the CR sphere*, Math. Z. **275** (2013), no. 1-2, 135–150. MR 3101801
- [17] Andre Bellaïche, *Propriétés extrémales des géodésiques*, Asterisque (1981), no. 84, 83–130.
- [18] Catherine Bellaïche, *Comportement asymptotique de $p(t, x, y)$ quand $t \rightarrow 0$ (points éloignés)*, Asterisque (1981), no. 84, 151–188.
- [19] G. Ben Arous, *Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus*, Ann. Sci. École Norm. Sup. (4) **21** (1988), no. 3, 307–331. MR MR974408 (89k:60087)
- [20] Ugo Boscain and Robert W. Neel, *Extensions of Brownian motion to a family of Grushin-type singularities*, Electron. Commun. Probab. **25** (2020), Paper No. 29, 12. MR 4089736
- [21] Ugo Boscain and Dario Prandi, *Self-adjoint extensions and stochastic completeness of the Laplace-Beltrami operator on conic and anticonic surfaces*, J. Differential Equations **260** (2016), no. 4, 3234–3269. MR 3434398
- [22] El-H. Ch. El-Alaoui, J.P.A. Gauthier, and I. Kupka, *Small sub-Riemannian balls on \mathbf{R}^3* , J. Dynam. Control Systems **2** (1996), no. 3, 359–421. MR 1403263
- [23] Nathaniel Eldredge, *Precise estimates for the subelliptic heat kernel on H -type groups*, J. Math. Pures Appl. (9) **92** (2009), no. 1, 52–85. MR 2541147
- [24] Ricardo Estrada and Ram P. Kanwal, *A distributional approach to asymptotics*, second ed., Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Boston Inc., Boston, MA, 2002, Theory and applications. MR 2002k:46096
- [25] Matteo Gallone and Alessandro Michelangeli, *Quantum particle across grushin singularity*, preprint, arXiv:2011.13712 (2020).
- [26] Matteo Gallone, Alessandro Michelangeli, and Eugenio Pozzoli, *Geometric confinement and dynamical transmission of a quantum particle in grushin cylinder*, preprint, arXiv:2003.07128 (2020).
- [27] Karen Habermann, *Small-time fluctuations for sub-Riemannian diffusion loops*, Probab. Theory Related Fields **171** (2018), no. 3-4, 617–652. MR 3827218
- [28] Elton P. Hsu, *On the principle of not feeling the boundary for diffusion processes*, J. London Math. Soc. (2) **51** (1995), no. 2, 373–382. MR 1325580
- [29] ———, *Estimates of derivatives of the heat kernel on a compact Riemannian manifold*, Proc. Amer. Math. Soc. **127** (1999), no. 12, 3739–3744. MR 1618694

- [30] ———, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics, vol. 38, American Mathematical Society, Providence, RI, 2002. MR 2003c:58026
- [31] Pei Hsu, *Brownian bridges on Riemannian manifolds*, Probab. Theory Related Fields **84** (1990), no. 1, 103–118. MR 1027823
- [32] ———, *Heat kernel on noncomplete manifolds*, Indiana Univ. Math. J. **39** (1990), no. 2, 431–442. MR 1089046
- [33] Yuzuru Inahama, *Large deviations for rough path lifts of Watanabe’s pullbacks of delta functions*, Int. Math. Res. Not. IMRN (2016), no. 20, 6378–6414. MR 3579967
- [34] Yuzuru Inahama and Setsuo Taniguchi, *Short time full asymptotic expansion of hypoelliptic heat kernel at the cut locus*, Forum Math. Sigma **5** (2017), Paper No. e16, 74. MR 3669328
- [35] Shigeo Kusuoka and Daniel W. Stroock, *Asymptotics of certain Wiener functionals with degenerate extrema*, Comm. Pure Appl. Math. **47** (1994), no. 4, 477–501. MR 1272385
- [36] Rémi Léandre, *Majoration en temps petit de la densité d’une diffusion dégénérée*, Probab. Theory Related Fields **74** (1987), no. 2, 289–294. MR 871256 (88c:60144)
- [37] ———, *Minoration en temps petit de la densité d’une diffusion dégénérée*, J. Funct. Anal. **74** (1987), no. 2, 399–414. MR 904825 (88k:60147)
- [38] Hong-Quan Li, *Estimations asymptotiques du noyau de la chaleur sur les groupes de Heisenberg*, C. R. Math. Acad. Sci. Paris **344** (2007), no. 8, 497–502. MR 2324485
- [39] Matthias Ludewig, *Heat kernel asymptotics, path integrals and infinite-dimensional determinants*, J. Geom. Phys. **131** (2018), 66–88. MR 3815228
- [40] ———, *Strong short-time asymptotics and convolution approximation of the heat kernel*, Ann. Global Anal. Geom. **55** (2019), no. 2, 371–394. MR 3923544
- [41] S. A. Molčanov, *Diffusion processes, and Riemannian geometry*, Uspehi Mat. Nauk **30** (1975), no. 1(181), 3–59. MR MR0413289 (54 #1404)
- [42] Robert Neel, *The small-time asymptotics of the heat kernel at the cut locus*, Comm. Anal. Geom. **15** (2007), no. 4, 845–890. MR MR2395259
- [43] O. Rioul, *Information theoretic proofs of entropy power inequalities*, IEEE Transactions on Information Theory **57** (2011), no. 1, 33–55.
- [44] Daniel W. Stroock and James Turetsky, *Upper bounds on derivatives of the logarithm of the heat kernel*, Comm. Anal. Geom. **6** (1998), no. 4, 669–685. MR 1664888

DEPARTMENT OF MATHEMATICS, CHANDLER-ULLMANN HALL, LEHIGH UNIVERSITY, BETHLEHEM, PENNSYLVANIA, USA

Email address: robert.neel@lehigh.edu

UNIV. LYON, UNIVERSITÉ CLAUDE BERNARD LYON 1, CNRS, LAGEPP UMR 5007, 43 BD DU 11 NOVEMBRE 1918, F-69100 VILLEURBANNE, FRANCE

Email address: ludovic.sacchelli@univ-lyon1.fr