

# Heat trace asymptotics on equiregular sub-Riemannian manifolds

By Yuzuru INAHAMA and Setsuo TANIGUCHI

(Received Mar. 20, 2019)

**Abstract.** We study a “div-grad type” sub-Laplacian with respect to a smooth measure and its associated heat semigroup on a compact equiregular sub-Riemannian manifold. We prove a short time asymptotic expansion of the heat trace up to any order. Our main result holds true for any smooth measure on the manifold, but it has a spectral geometric meaning when Popp’s measure is considered. Our proof is probabilistic. In particular, we use Watanabe’s distributional Malliavin calculus.

## 1. Introduction and main result.

In Introduction of his textbook on sub-Riemannian geometry [29], Montgomery emphasized the importance of spectral geometric problems in sub-Riemannian geometry by asking “Can you ‘hear’ the subriemannian metric from the spectrum of its sublaplacian?” (Of course, this is a slight modification of Kac’s renowned question.) In the same paragraph, he also mentioned Malliavin calculus, which is a powerful infinite-dimensional functional analytic method for studying stochastic differential equations (SDEs) under the Hörmander condition on the coefficient vector fields.

However, there is no canonical choice of measure on a general sub-Riemannian manifold and hence no canonical choice of sub-Laplacian. Therefore, in order to pose spectral geometric questions, one should consider a subclass of sub-Riemannian manifolds. In this regard, the class of equiregular sub-Riemannian manifolds seems suitable for the following reason. As Montgomery himself proved in [29, Section 10.6], there exists a canonical smooth volume called Popp’s measure on an equiregular sub-Riemannian manifold. Popp’s measure is determined by the sub-Riemannian metric only.

In the present paper, we contribute to this topic by proving a short time asymptotic expansion of the heat trace up to an arbitrary order on a compact equiregular sub-Riemannian manifold. Our main tool is Watanabe’s distributional Malliavin calculus.

To state our main result, we start by recalling the definition of an equiregular sub-Riemannian manifold. Note that in many literatures an *equiregular* sub-Riemannian manifold is simply called *regular*.

We say that  $(M, \mathcal{D}, g)$  is a sub-Riemannian manifold if (i)  $M$  is a connected, smooth manifold of dimension  $d$ , (ii)  $\mathcal{D} \subset TM$ ,  $TM$  being the tangent bundle of  $M$ , is a smooth

---

2010 *Mathematics Subject Classification.* Primary 53C17; Secondary 60H07, 58J65, 35K08, 41A60.  
*Key Words and Phrases.* sub-Riemannian geometry, heat kernel, stochastic differential equation, Malliavin calculus, asymptotic expansion.

The first-named author is partially supported by JSPS KAKENHI Grant Number 15K04922, and the second-named author is partially supported by JSPS KAKENHI Grant Number 15K04931.

distribution of constant rank  $n$  ( $1 \leq n \leq d$ ) which satisfies the Hörmander condition at every  $x \in M$  and (iii)  $g = (g_x)_{x \in M}$ , where each  $g_x$  is an inner product on the fiber  $\mathcal{D}_x$ , and  $x \mapsto g_x$  is smooth as a function of  $x$ . (When there is no risk of confusion, we simply say that  $M$  is a sub-Riemannian manifold.)

The precise statement of the Hörmander condition on  $\mathcal{D}$  at  $x \in M$  is as follows: Define  $\mathcal{D}_0(x) = \{0\}$ ,  $\mathcal{D}_1(x) = \mathcal{D}(x)$  and

$$\mathcal{D}_k(x) = \text{linear span of } \left\{ \underbrace{[[[A_1, A_2], \dots], A_l]}_{(l-1)\text{brackets}}(x) \mid 1 \leq l \leq k, A_1, \dots, A_l \in C^\infty(M; \mathcal{D}) \right\}$$

for  $k \geq 2$ . Here,  $C^\infty(M; \mathcal{D})$  stands for the  $C^\infty$ -module of smooth sections of  $\mathcal{D}$  over  $M$ . We say that  $\mathcal{D}$  satisfies the Hörmander condition at  $x$  if there exists  $N = N(x)$  such that  $\mathcal{D}_N(x) = T_x M$ .

A sub-Riemannian manifold  $(M, \mathcal{D}, g)$  is said to be equiregular if  $\dim \mathcal{D}_k(x)$  is constant in  $x \in M$  for all  $k \geq 1$ . The smallest constant  $N_0$  such that  $\mathcal{D}_{N_0}(x) = T_x M$  is called the *step* of the Hörmander condition. In this case,  $\nu := \sum_{k=1}^{N_0} k(\dim \mathcal{D}_k(x) - \dim \mathcal{D}_{k-1}(x))$ , is also constant in  $x$  and equals the Hausdorff dimension of  $M$  equipped with the usual sub-Riemannian distance.

Now we define a “div-grad type” sub-Laplacian on a sub-Riemannian manifold  $M$ . Let  $\mu$  be a smooth volume on  $M$ , that is,  $\mu$  is a measure on  $M$  whose restriction to every local coordinate chart is written as a strictly positive smooth density function times the Lebesgue measure on the chart. In the equiregular case, the most important example of smooth volume is Popp’s measure introduced in [29, Section 10.6] (see also [3]) since Popp’s measure is determined solely by the equiregular sub-Riemannian structure.

We study the second-order differential operator of the form  $\Delta = \text{div}_\mu \nabla^\mathcal{D}$ , where  $\nabla^\mathcal{D}$  is the horizontal gradient in the direction of  $\mathcal{D}$  and  $\text{div}_\mu$  is the divergence with respect to  $\mu$ . (In our convention,  $\Delta$  is a non-positive operator.) By the way it is defined,  $\Delta$  with its domain being  $C_0^\infty(M)$  is clearly symmetric on  $L^2(\mu)$ . If  $M$  is compact, then  $\Delta$  is known to be essentially self-adjoint on  $C^\infty(M)$  and  $e^{t\Delta/2}$  is of trace class for every  $t > 0$ , where  $(e^{t\Delta/2})_{t \geq 0}$  is the heat semigroup associated with  $\Delta/2$ .

Now we are in a position to state our main result in this paper. The proof of this theorem is immediate from Theorem 6.1. As we have already mentioned, it has a spectral geometric meaning when  $\mu$  is Popp’s measure.

**THEOREM 1.1.** *Let  $M$  be a compact equiregular sub-Riemannian manifold of Hausdorff dimension  $\nu$  and let  $\mu$  be a smooth volume on  $M$ . Then, we have the following asymptotic expansion of the heat trace:*

$$\text{Trace}(e^{t\Delta/2}) \sim \frac{1}{t^{\nu/2}}(c_0 + c_1 t + c_2 t^2 + \dots) \quad \text{as } t \searrow 0 \tag{1.1}$$

for certain constants  $c_0 > 0$  and  $c_1, c_2, \dots \in \mathbb{R}$ .

Since the asymptotic expansion in Theorem 1.1 is up to an arbitrary order, we can prove meromorphic prolongation of the spectral zeta function associated with  $\Delta$  by a standard argument. Denote by  $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$  be all the eigenvalues of  $-\Delta$  in

increasing order with the multiplicities being counted and set

$$\zeta_{\Delta}(s) = \sum_{i=1}^{\infty} \lambda_i^{-s} \quad \left( s \in \mathbb{C}, \Re s > \frac{\nu}{2} \right).$$

By the Tauberian theorem, the series on the right-hand side absolutely converges and defines a holomorphic function on  $\{s \in \mathbb{C} \mid \Re s > \nu/2\}$ .

**COROLLARY 1.2.** *Let assumptions be the same as in Theorem 1.1. Then,  $\zeta_{\Delta}$  admits a meromorphic prolongation to the whole complex plane  $\mathbb{C}$ .*

To the best of our knowledge, Theorem 1.1 and Corollary 1.2 seem new for a general compact equiregular sub-Riemannian manifold. It should be noted, however, that the leading term of the asymptotics (1.1) is already known. (See [20], [28] for example. No explicit value of  $c_0$  is known in general.) For some concrete examples or relatively small classes of compact equiregular sub-Riemannian manifolds, the full asymptotic expansion (1.1) or the meromorphic extension of the spectral zeta function was proved. (See [4], [5], [7], [13], [32], [34] and references therein.) Most of such classes are subclasses of step-two or corank-one sub-Riemannian manifolds.

Our proof of Theorem 1.1 is based on Takanobu's result [35] on the short time asymptotic expansion of hypoelliptic heat kernels on the diagonal. A preceding work by Ben Arous [8] should also be mentioned. Both of [8], [35] are probabilistic and formulated on  $\mathbb{R}^d$ . Compared to [8], [35] has the following features: (i) The SDE has a drift term. Unlike most of the problems for SDEs, a drift term makes this kind of asymptotic quite complicated. (ii) The Hörmander condition is assumed only at the starting point. (iii) The asymptotics expansion takes place at the level of Watanabe distributions, which is stronger than an asymptotic expansion of the heat kernel. On the other hand, [8] proves a uniform asymptotic expansion of the heat kernel with respect to the starting point as it varies in a compact “equiregular” subset of  $\mathbb{R}^d$ . (It seems that Takanobu's motivation was to investigate what happens when the Hörmander condition is not so nice. He discovered that a pathological phenomenon happens when the condition is weak enough. Later, this phenomenon was further studied by Ben Arous and Léandre [9], [10].)

We first prove a uniform asymptotic expansion at the level of Watanabe distributions under the equiregular Hörmander condition for a driftless SDE on  $\mathbb{R}^d$  (Theorem 5.4). Although it is similar to the main results in [8], [35], this theorem, precisely speaking, is not included in these papers. We basically follow the argument in [35] to prove this theorem, but we believe that our proof is simpler and more readable for reasons that will be specified later (Remark 5.13).

Thanks to recent developments of the stochastic parallel transport on sub-Riemannian manifolds, we can construct the  $\Delta/2$ -diffusion process on  $M$  as a strong solution to an SDE of Eells–Elworthy–Malliavin type. Since the solution is non-degenerate in the sense of Malliavin calculus, a standard localization procedure for heat kernels works. Thus, our asymptotic problem on  $M$  reduces to one on  $\mathbb{R}^d$ . (The reason why it suffices to consider the driftless case in Theorem 5.4 is as follows. The SDE corresponding to  $\Delta/2$  on  $M$  and its localized version have a drift term, but it can be dealt with by Girsanov's

theorem fortunately. Hence, our asymptotic problem reduces to the driftless case.)

In our view, (possible) advantages of the probabilistic approach to analytic problems on sub-Riemannian manifolds are as follows. (For more information on this approach, see Thalmaier's recent survey [39].) Unlike in the elliptic (i.e., Riemannian) case, analytic methods (in particular, the theory of pseudo differential operators) does not work perfectly under a general bracket-generating condition (except for the corank-one or the step-two case). On the other hand, Malliavin calculus works under a general bracket-generating condition and the step of the condition does not really matter. Therefore, there seems to be a good chance that probability theory turns out to be more powerful than analysis at least for certain problems in sub-Riemannian geometry.

Merits of using Watanabe's version of Malliavin calculus in sub-Riemannian geometry could be as follows. First, it is probably the most powerful among a few versions of Malliavin calculus. In particular, it has a nice asymptotic theory. Second, it is highly self-contained. (For example, existence of the heat kernel can be shown within this theory and the heat kernel is expressed by a generalized Feynman–Kac formula. See Section 6.) This aspect of Watanabe's theory has not been paid much attention in the Riemannian case, probably because properties of many important objects on Riemannian manifolds were already obtained by analytic methods and one could just borrow them. On sub-Riemannian manifolds, however, analysis has not been fully developed. Hence, there is a possibility that the self-containedness will turn out to be of great advantage in the future development of this research topic.

The organization of this paper is as follows. In Section 2, a very brief review of Watanabe's distributional Malliavin calculus is given. In Section 3, the free nilpotent groups/algebras and canonical diffusion processes on them are introduced. These processes approximate the diffusion process we actually investigate. In Section 4, we summarize many non-trivial properties of vector fields on  $\mathbb{R}^d$  that satisfy the (equiregular) Hörmander condition. The main purpose of Section 5 is to present and prove our key theorem on  $\mathbb{R}^d$  (Theorem 5.4) by using Malliavin calculus. This theorem is a “uniform version” of the main result in [35] and can also be considered as a “Watanabe distribution version” of the main result in [8]. In Section 6, we prove our main theorem (Theorem 1.1) by showing a uniform asymptotic expansion of the heat kernel on a compact equiregular sub-Riemannian manifold  $M$ . By localization and Girsanov's theorem, the proof of this fact is reduced to that of the Euclidean case (Theorem 5.4). In Section 7, we give explicit expressions of the leading constants of the asymptotic expansions for some special examples of sub-Riemannian manifold.

In a paper of this kind, the term *distribution* may have three different meanings: (i) a subbundle of the tangent bundle of a manifold (e.g., Martinet distribution, contact distribution), (ii) a generalized function (e.g., Schwartz distribution, Watanabe distribution), (iii) a probability measure, in particular, the law of a random variable (e.g., normal distribution, chi-squared distribution). We use this term only for (i) and (ii) in this paper. Since (i) and (ii) are very different, we believe there is no risk of confusion.

## 2. Preliminaries from Malliavin calculus.

Let  $W = C_0([0, 1], \mathbb{R}^n)$  be the set of continuous functions from  $[0, 1]$  to  $\mathbb{R}^n$  which start at 0. This is equipped with the usual sup-norm. The  $n$ -dimensional Wiener measure on  $W$  is denoted by  $\mathbb{P}$ . We denote by

$$H = \left\{ h \in W \mid \text{absolutely continuous and } \|h\|_H^2 := \int_0^1 |h'_s|^2 ds < \infty \right\}$$

the Cameron–Martin subspace of  $W$ . The triple  $(W, H, \mathbb{P})$  is called the classical Wiener space. The canonical realization on  $W$  of  $n$ -dimensional Brownian motion is denoted by  $(w_t)_{0 \leq t \leq 1} = (w_t^1, \dots, w_t^n)_{0 \leq t \leq 1}$ .

We recall Watanabe’s theory of generalized Wiener functionals (i.e., Watanabe distributions) in Malliavin calculus. Most of the contents and the notations in this section are contained in Sections V.8–V.10, Ikeda and Watanabe [23] with trivial modifications. We also refer to Shigekawa [33], Nualart [30], Hu [22] and Matsumoto and Taniguchi [27].

The following are of particular importance in this paper:

(a) Basics of Sobolev spaces: We denote by  $\mathbb{D}_{p,r}(\mathcal{X})$  the Sobolev space of  $\mathcal{X}$ -valued (generalized) Wiener functionals, where  $p \in (1, \infty)$ ,  $r \in \mathbb{R}$ , and  $\mathcal{X}$  is a real separable Hilbert space. As usual, we will use the spaces  $\mathbb{D}_\infty(\mathcal{X}) = \bigcap_{k=1}^\infty \bigcap_{1 < p < \infty} \mathbb{D}_{p,k}(\mathcal{X})$ ,  $\tilde{\mathbb{D}}_\infty(\mathcal{X}) = \bigcap_{k=1}^\infty \bigcup_{1 < p < \infty} \mathbb{D}_{p,k}(\mathcal{X})$  of test functionals and the spaces  $\mathbb{D}_{-\infty}(\mathcal{X}) = \bigcup_{k=1}^\infty \bigcup_{1 < p < \infty} \mathbb{D}_{p,-k}(\mathcal{X})$ ,  $\tilde{\mathbb{D}}_{-\infty}(\mathcal{X}) = \bigcup_{k=1}^\infty \bigcap_{1 < p < \infty} \mathbb{D}_{p,-k}(\mathcal{X})$  of Watanabe distributions as in [23]. When  $\mathcal{X} = \mathbb{R}$ , we simply write  $\mathbb{D}_{p,r}$ , etc. The  $\mathbb{D}_{p,r}(\mathcal{X})$ -norm is denoted by  $\|\cdot\|_{p,r}$ . The precise definition of an asymptotic expansion up to any order in these spaces can be found in [23, Section V.9]. We denote by  $D$  the gradient operator ( $H$ -derivative) and by  $L = -D^*D$  the Ornstein–Uhlenbeck operator.

(b) Meyer’s equivalence of Sobolev norms: See [23, Theorem 8.4]. A stronger version can be found in [33, Theorem 4.6], [30, Theorem 1.5.1] or Bogachev [12, Theorem 5.7.1]. It states that the Sobolev norms  $\|F\|_{p,k} = \|(I - L)^{k/2}F\|_{L^p}$  and  $\|F\|_{L^p} + \|D^k F\|_{L^p}$  are equivalent for every  $k \in \mathbb{N}$  and  $1 < p < \infty$ .

(c) Watanabe’s pullback: For  $F = (F^1, \dots, F^m) \in \mathbb{D}_\infty(\mathbb{R}^m)$ , we denote by  $\sigma[F](w) = \sigma_F(w)$  the Malliavin covariance matrix of  $F$ , whose  $(i, j)$ -component is given by  $\sigma_F^{ij}(w) = \langle DF^i(w), DF^j(w) \rangle_{H^*}$ . Now we assume that  $F$  is non-degenerate in the sense of Malliavin, that is,  $(\det \sigma[F])^{-1} \in L^p$  for every  $1 < p < \infty$ .

Then, pullback  $T \circ F = T(F) \in \tilde{\mathbb{D}}_{-\infty}$  of a tempered Schwartz distribution  $T \in \mathcal{S}'(\mathbb{R}^m)$  on  $\mathbb{R}^m$  by a non-degenerate Wiener functional  $F \in \mathbb{D}_\infty(\mathbb{R}^m)$  is well-defined and has nice properties. (See [23, Section V.9].) The key to justify this pullback is an integration by parts formula in the sense of Malliavin calculus. (Its generalization is given in Item (d) below.)

(d) A generalized version of the integration by parts formula in the sense of Malliavin calculus for Watanabe distribution, which is given as follows (see [23, p.377]):

For a non-degenerate Wiener functional  $F = (F^1, \dots, F^m) \in \mathbb{D}_\infty(\mathbb{R}^m)$ , we denote by  $\gamma_F^{ij}(w)$  the  $(i, j)$ -component of the inverse matrix  $\sigma_F^{-1}$ . Note that  $\sigma_F^{ij} \in \mathbb{D}_\infty$  and

$D\gamma_F^{ij} = -\sum_{k,l} \gamma_F^{ik} (D\sigma_F^{kl}) \gamma_F^{lj}$ . Hence, derivatives of  $\gamma_F^{ij}$  can be written in terms of  $\gamma_F^{ij}$ 's and the derivatives of  $\sigma_F^{ij}$ 's. Suppose  $G \in \mathbb{D}_\infty$  and  $T \in \mathcal{S}'(\mathbb{R}^m)$ . Then, the following integration by parts holds:

$$\mathbb{E}[(\partial_i T \circ F)G] = \mathbb{E}[(T \circ F)\Phi_i(\cdot; G)] \quad (1 \leq i \leq m), \tag{2.1}$$

where  $\mathbb{E}$  stands for the generalized expectation and  $\Phi_i(w; G) \in \mathbb{D}_\infty$  is given by

$$\Phi_i(w; G) = \sum_{j=1}^m D^* \left( \gamma_F^{ij}(w) G(w) DF^j(w) \right). \tag{2.2}$$

(e) Watanabe's asymptotic expansion theorem is a key theorem in his distributional Malliavin calculus, which can be found in [23, Theorem 9.4, pp.387–388]. It can be summarized as follows:

Let  $F(\varepsilon, \cdot) \in \mathbb{D}_\infty(\mathbb{R}^m)$  for  $0 < \varepsilon \leq 1$ . We say  $F(\varepsilon, \cdot)$  is uniformly non-degenerate in the sense of Malliavin if

$$\sup_{0 < \varepsilon \leq 1} \|(\det \sigma[DF(\varepsilon, \cdot)])^{-1}\|_{L^p} < \infty \quad \text{for every } 1 < p < \infty.$$

Let us assume that  $F(\varepsilon, \cdot) \in \mathbb{D}_\infty(\mathbb{R}^m)$  ( $0 < \varepsilon \leq 1$ ) is uniformly non-degenerate in the sense of Malliavin and admits the following asymptotic expansion:

$$F(\varepsilon, \cdot) \sim f_0 + \varepsilon^1 f_1 + \varepsilon^2 f_2 + \dots \quad \text{in } \mathbb{D}_\infty(\mathbb{R}^m) \text{ as } \varepsilon \searrow 0$$

with  $f_j \in \mathbb{D}_\infty(\mathbb{R}^m)$  for all  $j \in \mathbb{N}$ . Then, for any  $T \in \mathcal{S}'(\mathbb{R}^m)$ ,  $T \circ F(\varepsilon, w)$  admits the following asymptotic expansion:

$$T \circ F(\varepsilon, \cdot) \sim \psi_0 + \varepsilon^1 \psi_1 + \varepsilon^2 \psi_2 + \dots \quad \text{in } \tilde{\mathbb{D}}_\infty \text{ as } \varepsilon \searrow 0, \tag{2.3}$$

where  $\psi_j \in \tilde{\mathbb{D}}_\infty$  is given by the formal Taylor expansion. (For example,  $\psi_0 = T(f_0)$  and  $\psi_1 = \sum_{i=1}^m f_1^i \cdot (\partial T / \partial x^i)(f_0)$ , etc.)

### 3. Free nilpotent Lie group and lift of Brownian motion.

In this section we introduce the free nilpotent Lie groups and algebras, following Friz–Victoir [16, Chapter 7]. The set of iterated integrals (i.e., multiple Wiener integrals) of Brownian motion becomes a left-invariant hypoelliptic diffusion process on this Lie group. See [40] for example. (According to [36], a similar fact also holds for the iterated integrals of a continuous local semimartingale.) The logarithm of this process will play a major role since it approximates the diffusion process under investigation in short times.

Let  $N \geq 1$ , which is the step of nilpotency. We denote by  $T^N(\mathbb{R}^n) = \bigoplus_{i=0}^N (\mathbb{R}^n)^{\otimes i}$  the truncated tensor algebra of step  $N$ , where  $(\mathbb{R}^n)^{\otimes 0} = \mathbb{R}$  by convention. The dilation by  $c \in \mathbb{R}$  is denoted by  $\Delta_c^N$ , that is,  $\Delta_c^N(1, a_1, \dots, a_N) = (1, c^1 a_1, \dots, c^N a_N)$ . For  $N \leq M$ ,  $\Pi_N^M$  denotes the canonical projection from  $T^M(\mathbb{R}^n)$  onto  $T^N(\mathbb{R}^n)$ .

We write  $\mathfrak{t}^N(\mathbb{R}^n) = \{0 + A \mid 0 \in \mathbb{R}, A \in \bigoplus_{i=1}^N (\mathbb{R}^n)^{\otimes i}\}$ . This is a Lie algebra under the bracket  $[A, B] := A \otimes B - B \otimes A$ . Then,  $1 + \mathfrak{t}^N(\mathbb{R}^n) = \{1 + A \mid 1 \in \mathbb{R}, A \in$

$\bigoplus_{i=1}^N (\mathbb{R}^n)^{\otimes i}$  is a Lie group. The unit element is denoted by  $\mathbf{1}$ . Here,  $\exp = \exp_N : \mathfrak{t}^N(\mathbb{R}^n) \rightarrow 1 + \mathfrak{t}^N(\mathbb{R}^n)$  is the exponential map defined in the usual way. Its inverse is the logarithm map  $\log = \log_N$ . By the correspondence  $1 + A \mapsto A \in \bigoplus_{i=1}^N (\mathbb{R}^n)^{\otimes i}$ ,  $1 + \mathfrak{t}^N(\mathbb{R}^n)$  is diffeomorphic to the linear space  $\bigoplus_{i=1}^N (\mathbb{R}^n)^{\otimes i} \cong \mathfrak{t}^N(\mathbb{R}^n)$ . This map gives a (global) chart on this group.

The free nilpotent Lie algebra of step  $N$  is denoted by  $\mathfrak{g}^N(\mathbb{R}^n)$ , which is a sub-Lie algebra of  $\mathfrak{t}^N(\mathbb{R}^n)$  generated by the elements of  $\mathbb{R}^n$ . More precisely,

$$\mathfrak{g}^N(\mathbb{R}^n) := \mathbb{R}^n \oplus [\mathbb{R}^n, \mathbb{R}^n] \oplus \cdots \oplus \underbrace{[[[\mathbb{R}^n, \mathbb{R}^n], \dots], \mathbb{R}^n]}_{(N-1)\text{brackets}}.$$

The set  $G^N(\mathbb{R}^n) = \exp(\mathfrak{g}^N(\mathbb{R}^n))$  is called the free nilpotent Lie group of step  $N$ . It is a sub-Lie group of  $1 + \mathfrak{t}^N(\mathbb{R}^n)$ . Note that  $\log : G^N(\mathbb{R}^n) \rightarrow \mathfrak{g}^N(\mathbb{R}^n)$  is a diffeomorphism and its inverse is the exponential map  $\exp : \mathfrak{g}^N(\mathbb{R}^n) \rightarrow G^N(\mathbb{R}^n)$ . Using this diffeomorphism, we can define a new group product on  $\mathfrak{g}^N(\mathbb{R}^n)$  by

$$A \times B := \log(\exp(A) \exp(B)) \quad (A, B \in \mathfrak{g}^N(\mathbb{R}^n)).$$

Thanks to the Baker–Campbell–Hausdorff formula, the right-hand side has an explicit expression:

$$\begin{aligned} \log(\exp(A) \exp(B)) &= A + B + \frac{1}{2}[A, B] \\ &\quad + \frac{1}{12}([A, [A, B]] + [B, [B, A]]) - \frac{1}{24}[B, [A, [A, B]]] + \cdots. \end{aligned}$$

Here, terms of degree greater than  $N$  should be neglected. This is in fact a finite sum due to nilpotency and hence is a well-defined Lie polynomial in  $A$  and  $B$ .

Now we fix some symbols for linear basis on free nilpotent Lie algebra and words. The canonical basis of  $\mathbb{R}^n$  is denoted by  $\{\mathbf{e}_i \mid 1 \leq i \leq n\}$ . Set  $\mathcal{I}(N) = \bigcup_{k=1}^N \{(i_1, \dots, i_k) \mid 1 \leq i_1, \dots, i_k \leq n\}$  for  $1 \leq N \leq \infty$ . This is the set of words of  $n$  letters with length at most  $N$ , where the length of a word is defined by  $|(i_1, \dots, i_k)| =: k$ . For  $I = (i_1, \dots, i_k) \in \mathcal{I}(\infty)$ , we set  $\mathbf{e}_I = \mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_k}$ . When  $k = 1$ , we will often write  $i_1$  for  $(i_1)$ . For  $I = (i_1, \dots, i_k) \in \mathcal{I}(\infty)$ , we define  $\mathbf{e}_{[I]}$  as follows:

$$\mathbf{e}_{[i_1]} := \mathbf{e}_{i_1}, \quad \mathbf{e}_{[i_1, \dots, i_k]} := [\mathbf{e}_{[i_1, \dots, i_{k-1}]}, \mathbf{e}_{i_k}] \quad (k \geq 2).$$

Here and in what follows, we write  $[i_1, \dots, i_k]$  for  $[(i_1, \dots, i_k)]$  for simplicity of notations.

Let  $\mathcal{G}(N) \subset \mathcal{I}(N)$  ( $N = 1, 2, \dots$ ) be such that  $\mathcal{G}(N) \subset \mathcal{G}(N+1)$  for all  $N \geq 1$  and  $\{\mathbf{e}_{[I]} \mid I \in \mathcal{G}(N)\}$  forms a linear basis of  $\mathfrak{g}^N(\mathbb{R}^n)$ . The choice of such  $\mathcal{G}(N)$  ( $N = 1, 2, \dots$ ) is not unique. We write  $\mathcal{G}(\infty) = \bigcup_{k=1}^{\infty} \mathcal{G}(k)$ . (For example, we can take  $\mathcal{G}(1) = \{(i) \mid 1 \leq i \leq n\}$  and  $\mathcal{G}(2) = \mathcal{G}(1) \cup \{(i, j) \mid 1 \leq i < j \leq n\}$ .)

Now we introduce vector fields on the Lie group and the Lie algebra. Note that  $\mathbf{e}_i \in \mathbb{R}^n \subset \mathfrak{t}^N(\mathbb{R}^n) \cong T_{\mathbf{0}}\mathfrak{t}^N(\mathbb{R}^n)$ . Here, since  $\mathfrak{t}^N(\mathbb{R}^n)$  is a linear space, it is identified with its tangent space at the origin. Since  $1 + \mathfrak{t}^N(\mathbb{R}^n)$  and  $G^N(\mathbb{R}^n)$  are submanifolds of a linear space  $T^N(\mathbb{R}^n)$ , their tangent space can naturally be identified with a linear subspace of  $T^N(\mathbb{R}^n)$ . By straightforward computation,

$$\exp_* \mathbf{e}_i := \left. \frac{d}{ds} \right|_{s=0} \exp(s\mathbf{e}_i) = \mathbf{e}_i \in T_{\mathbf{1}}(\mathfrak{g}^N(\mathbb{R}^n)) \cong \mathfrak{g}^N(\mathbb{R}^n).$$

Let  $Q_i^N$  be the unique left-invariant vector field on  $1 + \mathfrak{t}^N(\mathbb{R}^n)$  or on  $G^N(\mathbb{R}^n)$  such that  $Q_i^N(\mathbf{1}) = \mathbf{e}_i$  ( $1 \leq i \leq n$ ). More concretely,

$$Q_i^N(A) = \left. \frac{d}{ds} \right|_{s=0} A \otimes \exp(s\mathbf{e}_i) \quad (A \in 1 + \mathfrak{t}^N(\mathbb{R}^n)).$$

The above limit is taken in  $T^N(\mathbb{R}^n)$ . If we choose  $\{\mathbf{e}_I \mid I \in \mathcal{I}(N)\}$  as a basis of  $\mathfrak{t}^N(\mathbb{R}^n)$ , an element of this linear space can be expressed as  $(y^I)_{I \in \mathcal{I}(N)} \in \mathbb{R}^{\mathcal{I}(N)}$ . In this coordinate we have

$$Q_i^N(A) = \frac{\partial}{\partial y^i} + \sum_{(j_1, \dots, j_k) \in \mathcal{I}(N-1)} y^{(j_1, \dots, j_k)} \frac{\partial}{\partial y^{(j_1, \dots, j_k, i)}} \left( A = 1 + \sum_{I \in \mathcal{I}(N)} y^I \mathbf{e}_I \in 1 + \mathfrak{t}^N(\mathbb{R}^n) \right)$$

for  $N \geq 2$ . See [35, p.174]. As vector fields on  $G^N(\mathbb{R}^n)$ ,  $\{Q_i^N\}_{1 \leq i \leq n}$  satisfy Hörmander’s bracket-generating condition at  $\mathbf{1}$  and hence at every point in  $G^N(\mathbb{R}^n)$  by the left-invariance.

Define  $\hat{Q}_i^N = \log_* Q_i^N$ . Then,  $\{\hat{Q}_i^N\}_{1 \leq i \leq n}$  are smooth vector fields on  $\mathfrak{t}^N(\mathbb{R}^n)$  and satisfy the Hörmander condition as vector fields on  $\mathfrak{g}^N(\mathbb{R}^n)$  at every point in  $\mathfrak{g}^N(\mathbb{R}^n)$ . By way of construction, these are left-invariant with respect to the product  $\times$ . The Baker–Campbell–Hausdorff formula implies that, if we write

$$\hat{Q}_i^N(A) = \sum_{I \in \mathcal{G}(N)} (\hat{Q}_i^N)^I(A) \frac{\partial}{\partial y^I} \quad \left( A = \sum_{I \in \mathcal{G}(N)} y^I \mathbf{e}_{[I]} \in \mathfrak{g}^N(\mathbb{R}^n) \right),$$

then the coefficient  $(\hat{Q}_i^N)^I$  is actually a real-valued polynomial in  $(y^I)_{I \in \mathcal{G}(N)}$ .

If  $N = 3$  for example, we have for  $A = \sum_{I \in \mathcal{G}(3)} y^I \mathbf{e}_{[I]} \in \mathfrak{g}^3(\mathbb{R}^n)$  that

$$\hat{Q}_i^3(A) = \left. \frac{d}{ds} \right|_{s=0} A \times (s\mathbf{e}_i) = \mathbf{e}_i + \frac{1}{2}[A, \mathbf{e}_i] + \frac{1}{12}[A, [A, \mathbf{e}_i]],$$

which is a second order polynomial in  $(y^I)_{I \in \mathcal{G}(3)}$ . Here, the linear space  $\mathfrak{g}^3(\mathbb{R}^n)$  and its tangent space are identified in the usual way.

Consider the following ODE on  $G^N(\mathbb{R}^n)$  driven by an  $\mathbb{R}^n$ -valued Cameron–Martin path  $h \in H$ :

$$dy_t^N = \sum_{i=1}^n Q_i^N(y_t^N) dh_t^i \quad \text{with } y_0^N = \mathbf{1}. \tag{3.1}$$

It is well-known that a unique solution of (3.1) has the following explicit expression in the form of iterated path integrals (e.g., [16, Chapter 7]):

$$y_t^N = y_t^N(h) = \sum_{I \in \mathcal{I}(N)} h_t^I \mathbf{e}_I,$$

where we set

$$h_t^{(i_1)} := h_t^{i_1}, \quad h_t^{(i_1, \dots, i_k)} := \int_0^t h_s^{(i_1, \dots, i_{k-1})} dh_s^{i_k} \quad (k \geq 2).$$

In rough path theory,  $y^N$  is called the level  $N$  rough path lift of  $h$ . By (a trivial modification of) [16, Theorem 7.30], we have  $G^N(\mathbb{R}^n) = \{y_T^N(h) \mid h \in H\}$  for every  $T > 0$ .

The corresponding Stratonovich-type SDE on  $G^N(\mathbb{R}^n)$  driven by an  $n$ -dimensional Brownian motion  $w$  is as follows:

$$dY_t^N = \sum_{i=1}^n Q_i^N(Y_t^N) \circ dw_t^i \quad \text{with } Y_0^N = \mathbf{1}. \quad (3.2)$$

A unique solution of (3.2) has the following explicit expression in the form of iterated Stratonovich integrals:

$$Y_t^N = Y_t^N(w) = \sum_{I \in \mathcal{I}(N)} w_t^I \mathbf{e}_I, \quad \text{almost surely,}$$

where we set

$$w_t^{(i_1)} := w_t^{i_1}, \quad w_t^{(i_1, \dots, i_k)} := \int_0^t w_s^{(i_1, \dots, i_{k-1})} \circ dw_s^{i_k} \quad (k \geq 2).$$

In rough path theory,  $Y^N$  is called the level  $N$  rough path lift of  $w$  or Brownian rough path of level  $N$ . (In most of the cases in rough path theory,  $N = 2$ .)

Set  $U_t^N = \log Y_t^N$ . We can easily see that  $(U_t^N)$  is a diffusion process on  $\mathfrak{g}^N(\mathbb{R}^n)$  which satisfies the following Stratonovich-type SDE:

$$dU_t^N = \sum_{i=1}^n \hat{Q}_i^N(U_t^N) \circ dw_t^i \quad \text{with } U_0^N = \mathbf{0}. \quad (3.3)$$

Note that (i) the processes  $(\Delta_c^N U_t^N)$  and  $(U_{c^2 t}^N)$  have the same law for every  $c \in \mathbb{R}$  (i.e., the scaling property) and (ii)  $U_t^{N,I}(-w) = (-1)^{|I|} U_t^{N,I}(w)$  almost surely for every  $I \in \mathcal{G}(N)$ . One can show these facts by first showing the counterparts for  $(Y_t^N)$  and then taking the logarithm.

Since  $\{(U_t^N)_{t \geq 0} \mid N \geq 1\}$  are consistent with the projection system, that is,  $\Pi_N^M(U_t^M) = U_t^N$  for  $M \geq N$ , we have  $U_t^{N,I} = U_t^{M,I}$  if  $|I| \leq N \leq M$ . Therefore, we may and will simply write  $U_t^I$  for this object.

**REMARK 3.1.** Before we apply Malliavin calculus to (3.3), we make a comment on the regularity of smooth coefficient vector fields. A standard assumption requires that all the derivatives of the coefficients of  $(\hat{Q}_i^N)^I$  of order  $\geq 1$  be bounded. (However, this is not satisfied in our case.)

The main reason why this cannot be relaxed so easily is because a solution of the SDE may explode in finite time without this kind of assumption. However, if existence of a time-global solution is known for some reason, then it is enough to assume that all the

derivatives of the coefficients of  $(\hat{Q}_i^N)^I$  are of at most polynomial growth. Then, most of standard results in Malliavin calculus for SDEs still hold. (In our present case, the coefficients of  $(\hat{Q}_i^N)^I$  are literally polynomials, as we have seen.)

Precisely, it suffices to check that

$$\sup_{0 \leq t \leq 1} \left( \|U_t^N\|_{L^p} + \|\partial U_t^N\|_{L^p} + \|(\partial U_t^N)^{-1}\|_{L^p} \right) < \infty \quad (1 < p < \infty). \tag{3.4}$$

Here,  $\partial U^N$  is the Jacobian process (at  $\mathbf{0}$ ) associated with SDE (3.3) and takes values in  $GL(\mathfrak{g}^N(\mathbb{R}^n))$ . More explicitly, if we denote by  $U^N(t, A)$  the solution of SDE (3.3) which starts at  $A \in \mathfrak{g}^N(\mathbb{R}^n)$ , then  $\partial U_t^N := \nabla U^N(t, A)|_{A=\mathbf{0}}$ , where  $\nabla$  is the gradient operator with respect to  $A$ -variable on  $\mathfrak{g}^N(\mathbb{R}^n)$ .

The reason why this is sufficient is as follows: The higher order  $H$ -derivatives  $D^k U^N$  ( $k = 1, 2, \dots$ ) can be written as a stochastic integral which only involves  $w, U^N, \partial U^N, (\partial U^N)^{-1}$  and  $DU^N, D^2 U^N, \dots, D^{k-1} U^N$ . (See [33, Section 6.1] for example.) Due to this ‘‘triangular structure’’ of the integral expression, verifying (3.4) is enough.

Since  $U_t^N = \log Y_t^N$ , every component of  $U_t^N$  is a polynomial in  $w_t^{(i_1, \dots, i_k)}$  ( $1 \leq k \leq N$ ),  $L^p$ -norm of  $U_t^N$  clearly satisfies (3.4). By the left-invariance, we have  $U^N(t, A) = A \times U_t^N$ . Let  $\{e_{[I]} \mid I \in \mathcal{G}(N)\}$  be a basis of  $\mathfrak{g}^N(\mathbb{R}^n)$  and arrange them in increasing order of the step number. From the Baker–Campbell–Hausdorff formula and straightforward computation, we can see that  $\partial U_t^N$  is represented with respect to this basis by an lower triangular matrix with all the diagonal entries being 1. Other non-zero entries of this matrix are polynomials in  $w_t^{(i_1, \dots, i_k)}$  ( $1 \leq k \leq N - 1$ ). Therefore,  $L^p$ -norms of  $\partial U_t^N$  and its inverse satisfy (3.4).

Let  $\sigma[U_1^N] = (\langle DU_1^{N,I}, DU_1^{N,J} \rangle_H)_{I, J \in \mathcal{G}(N)}$  be the Malliavin covariance matrix of  $U_1$  and  $\lambda[U_1^N]$  its smallest eigenvalue. (This means that  $\mathfrak{g}^N(\mathbb{R}^n)$  is implicitly equipped with an inner product with respect to which  $\{e_{[I]} \mid I \in \mathcal{G}(N)\}$  forms an orthonormal basis.)

**PROPOSITION 3.2.** *Let the notations be as above. Then,  $\lambda[U_1^N] > 0$  almost surely and  $\lambda[U_1^N]^{-1} \in \bigcap_{1 < p < \infty} L^p$ . In particular,  $U_1^N$  is non-degenerate in the sense of Malliavin.*

**PROOF.** By using a standard stopping time argument, we only need information of the coefficient vector fields near the starting point  $\mathbf{0}$ . Then, this problem reduces to the one under the standard regularity assumption in Malliavin calculus presented in Remark 3.1 above. See Kusuoka–Stroock [26] for example. □

We need the following estimate of the exit probability: For every  $\kappa > 0$ , there exists positive constants  $C_{N, \kappa}, \hat{C}_{N, \kappa}$  such that

$$\mathbb{P} \left( \sup_{0 \leq s \leq t} |\Delta_\varepsilon^N U_s^N| > \kappa \right) \leq \hat{C}_{N, \kappa} \exp \left( -\frac{C_{N, \kappa}}{4t\varepsilon^2} \right) \quad (\text{if } 0 < t\varepsilon^2 < C_{N, \kappa}) \tag{3.5}$$

([35, p.181]). This follows from the scaling property for  $U^N$  and a standard argument for the exit probability for (local) semimartingales.

REMARK 3.3. Although  $N$  is an arbitrarily fixed number in this section, all the objects are in fact consistent with the system of projections  $\{\Pi_N^M\}_{M \geq N}$ . For example,  $\Pi_N^M(G^M(\mathbb{R}^n)) = G^N(\mathbb{R}^n)$ ,  $\Pi_N^M(\mathfrak{g}^M(\mathbb{R}^n)) = \mathfrak{g}^N(\mathbb{R}^n)$ ,  $\Pi_N^M \circ \exp_M = \exp_N \circ \Pi_N^M$ ,  $\Pi_N^M \circ \log_M = \log_N \circ \Pi_N^M$ ,  $(\Pi_N^M)_* Q_i^M = Q_i^N$ ,  $\Pi_N^M(U_t^M) = U_t^N$ , etc. The projections of course commute with the dilations, too. This consistency indicates that these objects actually live in the projective limit spaces, but we do not take this viewpoint in this paper.

REMARK 3.4. The Lie group product  $A \times B$  on  $\mathfrak{g}^N(\mathbb{R}^n)$  in [35] equals  $B \times A$  in the present paper. Concerning this, the vector field  $\hat{Q}_i^N$  is *left-invariant* here, not right-invariant as in [35]. This modification is only for the aesthetic reason and of no mathematical importance.

REMARK 3.5. In many literatures  $(\mathfrak{g}^N(\mathbb{R}^n), \times)$ , instead of  $G^N(\mathbb{R}^n)$ , is called the free nilpotent *group* of step  $N$ . See Cygan [14] for example.

#### 4. Vector fields on $\mathbb{R}^d$ .

In this section we discuss vector fields on  $\mathbb{R}^d$ . We fix notations and recall some basic facts for later use. There are no new results in this section. The set of all the vector fields on  $\mathbb{R}^d$  is denoted by  $\mathfrak{X}(\mathbb{R}^d)$ . An element of  $\mathbb{R}^d$  is denoted by  $x = (x^1, \dots, x^d)$  as usual. The set of all linear mappings from a vector space  $\mathcal{X}$  to another vector space  $\mathcal{Y}$  is denoted by  $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ .

For  $V \in \mathfrak{X}(\mathbb{R}^d)$ , we write  $V^i(x) = \langle dx^i, V(x) \rangle$  and hence  $V(x) = \sum_{i=1}^d V^i(x) (\partial/\partial x^i)$ . Note that a vector field is always regarded as a first order differential operator. The  $\mathbb{R}^d$ -valued function  $(V^1(x), \dots, V^d(x))^*$  is denoted by  $(V\text{Id})(x)$ , where  $\text{Id}$  stands for the identity map of  $\mathbb{R}^d$ . Here and in what follows, the superscript  $*$  stands for the transpose of a matrix.

For the rest of this section, let  $n \geq 1$  and  $V_1, \dots, V_n \in \mathfrak{X}(\mathbb{R}^d)$ . For  $I = (i_1, \dots, i_k) \in \mathcal{I}(\infty)$ , we set  $V_I = V_{i_1} V_{i_2} \cdots V_{i_k}$ , which is a differential operator of order  $n$ . We also set  $V_{[I]} \in \mathfrak{X}(\mathbb{R}^d)$  as follows:

$$V_{[i_1]} = V_{i_1}, \quad V_{[i_1, \dots, i_k]} = [V_{[i_1, \dots, i_{k-1}]}, V_{i_k}] \quad (k \geq 2).$$

The correspondence  $\mathbf{e}_{[I]} \mapsto V_{[I]}$  naturally extends to a Lie algebra homomorphism from the free Lie algebra generated by  $\mathbb{R}^d$  to  $\mathfrak{X}(\mathbb{R}^d)$ . In particular, every linear relation for  $\{\mathbf{e}_{[I]} \mid I \in \mathcal{I}(N)\}$  still holds for  $\{V_{[I]} \mid I \in \mathcal{I}(N)\}$ .

Now we give a simple lemma for later use. This lemma is essentially implied by [35, Corollary 2.3, Propositions 3.9 and 4.4]. Our proof below is, however, different from the one in [35] and more straightforward and algebraic.

LEMMA 4.1. *Let  $N \geq 1$ . Then, for every  $x \in \mathbb{R}^d$  and  $u = \sum_{J \in \mathcal{G}(N)} u^J \mathbf{e}_{[J]} \in \mathfrak{g}^N(\mathbb{R}^n)$ ,*

$$\sum_{I \in \mathcal{I}(N)} (V_I \text{Id})(x) \pi_I(\exp(u)) = \sum_{k=1}^N \frac{1}{k!} \sum_{|J_1| + \dots + |J_k| \leq N} (V_{[J_1]} \cdots V_{[J_k]} \text{Id})(x) u^{J_1} \cdots u^{J_k} \quad (4.1)$$

(the summation runs over all  $(J_1, \dots, J_k) \in \mathcal{G}(N)^k$  such that  $|J_1| + \dots + |J_k| \leq N$ ). Here,  $\pi_I$  is the linear functional on  $T^N(\mathbb{R}^n)$  that picks up the coefficient of  $\mathbf{e}_I$  and  $\exp$  is the exponential map from  $\mathfrak{g}^N(\mathbb{R}^n)$  to  $G^N(\mathbb{R}^n)$ .

PROOF. Let  $\alpha_J^K \in \mathbb{R}$  be such that

$$\mathbf{e}_{[J]} = \sum_{K \in \mathcal{I}(N)} \alpha_J^K \mathbf{e}_K \quad (J \in \mathcal{G}(N)).$$

Note that  $\alpha_J^K = 0$  if  $|J| \neq |K|$ . Then, it holds that

$$V_{[J]} = \sum_{K \in \mathcal{I}(N)} \alpha_J^K V_K \quad (J \in \mathcal{G}(N)). \tag{4.2}$$

The left-hand side of (4.1) is equal to

$$\sum_{I \in \mathcal{I}(N)} (V_I \text{Id})(x) \pi_I \left( \sum_{k=0}^N \frac{1}{k!} \left( \sum_{J \in \mathcal{G}(N)} u^J \sum_{K \in \mathcal{I}(N)} \alpha_J^K \mathbf{e}_K \right)^{\otimes k} \right),$$

which is a polynomial in  $u^I$ 's. Let us compute its  $k$ th order term. For  $k = 0$ , it vanishes since  $|I| \geq 1$ . For  $k = 1$ , we see from (4.2) that

$$\begin{aligned} \sum_{I \in \mathcal{I}(N)} (V_I \text{Id})(x) \pi_I \left( \sum_{J \in \mathcal{G}(N)} u^J \sum_{K \in \mathcal{I}(N)} \alpha_J^K \mathbf{e}_K \right) &= \sum_{J \in \mathcal{G}(N)} u^J \sum_{K \in \mathcal{I}(N)} \alpha_J^K (V_K \text{Id})(x) \\ &= \sum_{J \in \mathcal{G}(N)} u^J (V_{[J]} \text{Id})(x). \end{aligned}$$

For  $k \geq 2$ , the computation gets a little bit complicated. Let us consider the case  $k = 2$ . The concatenation of two words,  $K_1$  and  $K_2$ , is denoted by  $(K_1, K_2)$ . By summing over  $I$  first, we see that

$$\begin{aligned} \sum_{I \in \mathcal{I}(N)} (V_I \text{Id})(x) \pi_I \left( \sum_{J_1, J_2 \in \mathcal{G}(N)} u^{J_1} u^{J_2} \sum_{K_1, K_2 \in \mathcal{I}(N)} \alpha_{J_1}^{K_1} \alpha_{J_2}^{K_2} \mathbf{e}_{(K_1, K_2)} \right) \\ = \sum_{J_1, J_2 \in \mathcal{G}(N)} u^{J_1} u^{J_2} \sum_{K_1, K_2 \in \mathcal{I}(N), |K_1| + |K_2| \leq N} \alpha_{J_1}^{K_1} \alpha_{J_2}^{K_2} (V_{K_1} V_{K_2} \text{Id})(x). \end{aligned} \tag{4.3}$$

For any  $l, m \geq 1$  with  $l + m \leq N$ , we have by (4.2) that

$$\sum_{|K_1|=l} \sum_{|K_2|=m} \alpha_{J_1}^{K_1} \alpha_{J_2}^{K_2} (V_{K_1} V_{K_2} \text{Id})(x) = (V_{[J_1]} V_{[J_2]} \text{Id})(x) \delta_{|J_1|, l} \delta_{|J_2|, m}.$$

Hence, the left-hand side of (4.3) is equal to

$$\sum_{J_1, J_2 \in \mathcal{G}(N), |J_1| + |J_2| \leq N} u^{J_1} u^{J_2} (V_{[J_1]} V_{[J_2]} \text{Id})(x).$$

This proves the case for  $k = 2$ . We can prove the case  $k \geq 3$  essentially in the same way. Thus, we have shown (4.1).  $\square$

Next we give two types of bracket-generating condition for the vector fields. For  $x \in \mathbb{R}^d$  and  $k \geq 1$ , define  $\mathcal{A}_k(x)$  to be the linear span of  $\{V_{[I]}(x) \mid I \in \mathcal{I}(k)\}$  in  $T_x\mathbb{R}^d \cong \mathbb{R}^d$ . Note that it equals the linear span of  $\{V_{[I]}(x) \mid I \in \mathcal{G}(k)\}$ .

**(HC)<sub>x</sub>**: We say that  $\{V_1, \dots, V_n\}$  satisfies the Hörmander condition at  $x$  if there exists  $N \geq 1$  such that  $\mathcal{A}_N(x) = \mathbb{R}^d$ .

The smallest number  $N$  with this property is called the step of the Hörmander condition at  $x$  and denoted by  $N_0(x)$ . We set  $\nu(x) = \sum_{k=1}^{N_0(x)} k(\dim \mathcal{A}_k(x) - \dim \mathcal{A}_{k-1}(x))$  with  $\mathcal{A}_0(x) := \{0\}$  by convention.

**(ER)<sub>x</sub>**: We say that  $\{V_1, \dots, V_n\}$  satisfies the equiregular Hörmander condition on  $O \subset \mathbb{R}^d$  if (i) it satisfies **(HC)<sub>x</sub>** at every  $x \in O$  and (ii) for all  $k$ ,  $\dim \mathcal{A}_k(x)$  is constant in  $x \in O$ . If the equiregular Hörmander condition holds on some neighborhood of  $x$ , we simply say  $\{V_1, \dots, V_n\}$  satisfies the equiregular Hörmander condition *near*  $x$  and denote it by **(ER)<sub>x</sub>**.

Assume **(HC)<sub>x</sub>** at some  $x \in \mathbb{R}^d$ . Then, we can find  $\mathcal{H}(x) \subset \mathcal{G}(N_0(x))$  such that  $\#\mathcal{H}(x) = d$  and  $\mathcal{A}_k(x)$  equals the linear span of  $\{V_{[I]}(x) \mid I \in \mathcal{G}(k) \cap \mathcal{H}(x)\}$  for all  $k = 1, \dots, N_0(x)$ . Take  $J \in \mathcal{G}(N_0(x))$  and write  $V_{[J]}(x)$  as a unique linear combination of  $\{V_{[I]}(x)\}_{I \in \mathcal{H}(x)}$ :

$$V_{[J]}(x) = \sum_{I \in \mathcal{H}(x)} c_J^I(x) V_{[I]}(x).$$

Then, we can immediately see from the definition of  $\mathcal{H}(x)$  that  $c_J^I(x) = 0$  is  $|I| > |J|$ .

Now we assume **(ER)<sub>x<sub>0</sub></sub>** for  $x_0 \in \mathbb{R}^d$ . Then, on a certain neighborhood  $O$  of  $x_0$ , we can choose  $\mathcal{H}(x)$  independently from  $x$  and in that case we simply write  $\mathcal{H}$ . The linear subspace of  $\mathfrak{g}^{N_0}(\mathbb{R}^n)$  generated by  $\{\mathbf{e}_{[I]} \mid I \in \mathcal{H}\}$  is denoted by  $\mathbb{R}\langle \mathcal{H} \rangle$ . Likewise,  $N_0(x)$  and  $\nu(x)$  are independent of  $x \in O$  and denoted by  $N_0$  and  $\nu$ , respectively. We will fix such  $O$  for a while.

We introduce some linear maps for each  $x \in O$ . First, set  $B_{\mathcal{H}}(x) \in \mathcal{L}(\mathbb{R}\langle \mathcal{H} \rangle, \mathbb{R}^d)$  by

$$B_{\mathcal{H}}(x) = (V_{[I]}^i(x))_{1 \leq i \leq d, I \in \mathcal{H}}, \quad (4.4)$$

which is clearly invertible. Next, set  $\Gamma(x) = (\gamma_J^I(x))_{I \in \mathcal{H}, J \in \mathcal{G}(\infty)} \in \mathcal{L}(\mathfrak{g}^\infty(\mathbb{R}^n), \mathbb{R}\langle \mathcal{H} \rangle)$  by

$$\Gamma(x) = B_{\mathcal{H}}(x)^{-1} \cdot [(V_{[I]}^i(x))_{1 \leq i \leq d, I \in \mathcal{G}(\infty)}].$$

Here,  $\mathfrak{g}^\infty(\mathbb{R}^n) (\cong \mathbb{R}\langle \mathcal{G}(\infty) \rangle)$  is the free Lie algebra generated by  $\mathbb{R}^n$ . Then, from Lemma 4.2 below and the fact that  $c_J^I(x) = 0$  is  $|I| > |J|$  for  $I \in \mathcal{H}$  and  $J \in \mathcal{G}(N_0)$ , it follows that

$$\gamma_J^I(x) = \begin{cases} \delta_J^I & \text{if } J \in \mathcal{H}, \\ 0 & \text{if } J \in \mathcal{G}(N_0) \text{ and } |I| > |J|. \end{cases} \quad (4.5)$$

For  $N \geq N_0$ , we set  $\Gamma_N(x) = (\gamma_J^I(x))_{I \in \mathcal{H}, J \in \mathcal{G}(N)} \in \mathcal{L}(\mathfrak{g}^N(\mathbb{R}^n), \mathbb{R}\langle \mathcal{H} \rangle)$ . It immediately follows from (4.5) that  $\Gamma_N(x)\Gamma_N(x)^* \geq \text{Id}_{\mathcal{H}}$ .

Here, we give a simple lemma on linear algebra in a general setting.

LEMMA 4.2. *Suppose that  $\{\mathbf{b}_1, \dots, \mathbf{b}_d\}$  is a linear basis of  $\mathbb{R}^d$ . Let  $\mathbf{a}_1, \dots, \mathbf{a}_m$  ( $m \geq 1$ ) be given by linear combinations of  $\mathbf{b}_j$ 's as follows:*

$$\mathbf{a}_k = \sum_{j=1}^d c_k^j \mathbf{b}_j \quad (1 \leq k \leq m).$$

Set an invertible matrix  $B = [\mathbf{b}_1, \dots, \mathbf{b}_d]$  and a  $d \times m$  matrix  $C = (c_k^j)_{1 \leq j \leq d, 1 \leq k \leq m}$ . Then, we have

$$B^{-1}[\mathbf{b}_1, \dots, \mathbf{b}_d, \mathbf{a}_1, \dots, \mathbf{a}_m] = [\text{Id}_d | C]$$

as a  $d \times (d + m)$  matrix. Here,  $\text{Id}_d$  stands for the identity matrix of size  $d$ .

PROOF. The proof is immediate if we note that  $B^{-1}\mathbf{b}_i = \mathbf{e}_i$  for all  $i$ , where  $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$  is the canonical linear basis of  $\mathbb{R}^d$ . □

We fix a few more notations for  $N \geq 1$ . In this paragraph we do not assume **(HC)**, **(ER)** nor  $N \geq N_0(x)$ . Set  $B_N \in C^\infty(\mathbb{R}^d, \mathcal{L}(\mathfrak{g}^N(\mathbb{R}^n), \mathbb{R}^d))$  by

$$B_N(x) = (V_{[I]}^i(x))_{1 \leq i \leq d, I \in \mathcal{G}(N)}. \tag{4.6}$$

Next, define  $\mathbf{V}_{I_1, \dots, I_N} = (\mathbf{V}_{I_1, \dots, I_N}^{ij})_{1 \leq i, j \leq d} \in C^\infty(\mathbb{R}^d, \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d))$  for  $I_1, \dots, I_N \in \mathcal{I}(\infty)$  by

$$\mathbf{V}_{I_1, \dots, I_N}^{ij} = \frac{\partial}{\partial x^j} (V_{[I_1]} \cdots V_{[I_N]} x^i), \tag{4.7}$$

where  $x^i$  stands for the  $i$ th coordinate function  $x \mapsto x^i$  on  $\mathbb{R}^d$ . By convention we set  $\mathbf{V}_\emptyset = \text{Id}_d$ . It is obvious that

$$V_{[I_1]} \cdots V_{[I_N]} x^i = \sum_{j=1}^d V_{[I_1]}^j \mathbf{V}_{I_2, \dots, I_N}^{ij} \tag{4.8}$$

for  $N \geq 2$ . We also define  $M_N = (M_N^{ij})_{1 \leq i, j \leq d} \in C^\infty(\mathbb{R}^d \times \mathfrak{g}^N(\mathbb{R}^d), \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d))$  by

$$M_N^{ij}(x, u) = \delta_j^i + \sum_{k=1}^{N-1} \frac{1}{(k+1)!} \sum_{I_1, \dots, I_k \in \mathcal{G}(N)} \mathbf{V}_{I_1, \dots, I_k}^{ij}(x) u^{I_1} \cdots u^{I_k} \tag{4.9}$$

$$\left( u = \sum_{I \in \mathcal{G}(N)} u^I \mathbf{e}_{[I]} \right).$$

Finally, set  $F_N \in C^\infty(\mathbb{R}^d \times \mathfrak{g}^N(\mathbb{R}^d), \mathbb{R}^d)$  by

$$F_N(x, u) = M_N(x, u)B_N(x)u. \tag{4.10}$$

Let us assume  $(\mathbf{ER})_{x_0}$  again and that  $x$  is sufficiently close to  $x_0$  and  $N \geq N_0$ . It immediately follows that  $M_N(x, 0) = \text{Id}_d$  and  $(\partial_I F_N^i(x, 0))_{1 \leq i \leq d, I \in \mathcal{G}(N_0)} = B_{N_0}(x)$ . Here,  $\partial_I$  is a shorthand for  $\partial/\partial u^I$ . Therefore, there exist a neighborhood  $O_N$  of  $x_0$  and positive constants  $\kappa_N, r$  such that if  $|u| \leq \kappa_N$  and  $x \in O_N$ , then

$$\det M_N(x, u) \geq \frac{1}{2}, \quad M_N(x, u)^* M_N(x, u) \geq \frac{1}{2} \text{Id}_d \tag{4.11}$$

and

$$\begin{aligned} & (\partial_I F_N^i(x, u))_{1 \leq i \leq d, I \in \mathcal{G}(N_0)} [(\partial_I F_N^i(x, u))_{1 \leq i \leq d, I \in \mathcal{G}(N_0)}]^* \\ & \geq \frac{1}{2} B_{N_0}(x_0) B_{N_0}(x_0)^* \geq \frac{1}{2} B_{\mathcal{H}}(x_0) B_{\mathcal{H}}(x_0)^* \geq r \text{Id}_d. \end{aligned} \tag{4.12}$$

We continue to assume  $(\mathbf{ER})_{x_0}$  and let  $O_N$  be as above. We define four linear maps for  $N \geq N_0, 0 < \varepsilon \leq 1$  and  $x \in O_N$  as follows:

$$\tilde{\Gamma}_N^\varepsilon(x) = (\varepsilon^{|J|-|I|} \gamma_J^I(x))_{I \in \mathcal{H}, J \in \mathcal{G}(N)} \in \mathcal{L}(\mathfrak{g}^N(\mathbb{R}^n), \mathbb{R}\langle \mathcal{H} \rangle), \tag{4.13}$$

$$\tilde{\Gamma}_N^0(x) = (\delta_{|J|}^I \gamma_J^I(x))_{I \in \mathcal{H}, J \in \mathcal{G}(N)} \in \mathcal{L}(\mathfrak{g}^N(\mathbb{R}^n), \mathbb{R}\langle \mathcal{H} \rangle), \tag{4.14}$$

$$P_N = (\delta_J^J)_{I \in \mathcal{G}(N) \setminus \mathcal{H}, J \in \mathcal{G}(N)} \in \mathcal{L}(\mathfrak{g}^N(\mathbb{R}^n), \mathfrak{g}^N(\mathbb{R}^n)/\mathbb{R}\langle \mathcal{H} \rangle), \tag{4.15}$$

$$\Delta_\varepsilon^{\mathcal{H}} = (\varepsilon^{|J|} \delta_J^I)_{I \in \mathcal{H}, J \in \mathcal{H}} \in \mathcal{L}(\mathbb{R}\langle \mathcal{H} \rangle, \mathbb{R}\langle \mathcal{H} \rangle). \tag{4.16}$$

Note that  $\Delta_\varepsilon^{\mathcal{H}}$  is the dilation by  $\varepsilon$  restricted to  $\mathbb{R}\langle \mathcal{H} \rangle$  and that  $P_N$  is just the canonical projection. Via the inner product on  $\mathfrak{g}^N(\mathbb{R}^n)$ ,  $\mathfrak{g}^N(\mathbb{R}^n)/\mathbb{R}\langle \mathcal{H} \rangle$  is canonically identified with the orthogonal complement of  $\mathbb{R}\langle \mathcal{H} \rangle$  in  $\mathfrak{g}^N(\mathbb{R}^n)$ . In this way  $P_N$  can be regarded as the orthogonal projection. In fact, no negative power of  $\varepsilon$  is involved in the components of  $\tilde{\Gamma}_N^\varepsilon(x)$ , thanks to (4.5). By definition, we have  $\tilde{\Gamma}_N^0(x)u = \tilde{\Gamma}_{N_0}^0(x)\Pi_N^{N_0}u$  for all  $u \in \mathfrak{g}^N(\mathbb{R}^n)$  and  $N \geq N_0$ . The linear mapping

$$\begin{pmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{pmatrix} \in \mathcal{L}(\mathfrak{g}^N(\mathbb{R}^d), \mathfrak{g}^N(\mathbb{R}^d)) \tag{4.17}$$

will play an important role.

Now we give two simple lemmas for later use.

LEMMA 4.3. *Let the notations be as above and let  $N \geq N_0, 0 < \varepsilon \leq 1$ . Then, if we take  $O_N$  small enough, we have the following:*

- (1)  $\lim_{\varepsilon \searrow 0} \tilde{\Gamma}_N^\varepsilon(x) = \tilde{\Gamma}_N^0(x)$  uniformly in  $x \in O_N$ .
- (2)  $\Gamma_N(x)\Delta_\varepsilon^N = \Delta_\varepsilon^{\mathcal{H}}\tilde{\Gamma}_N^\varepsilon(x)$  for all  $x \in O_N$ .
- (3)  $\det \Delta_\varepsilon^{\mathcal{H}} = \varepsilon^\nu$ . In particular,  $\Delta_\varepsilon^{\mathcal{H}}$  is invertible.
- (4)  $\tilde{\Gamma}_N^0(x)\tilde{\Gamma}_N^0(x)^* \geq \text{Id}_{\mathbb{R}\langle \mathcal{H} \rangle}$  for all  $x \in O_N$ .

(5) *The linear mapping defined in (4.17) is invertible and there exists a positive constant  $r_N$  such that, for all  $x \in O_N$ ,*

$$\begin{pmatrix} \tilde{\Gamma}_N^0(x) \\ P_N \end{pmatrix} \begin{pmatrix} \tilde{\Gamma}_N^0(x) \\ P_N \end{pmatrix}^* \geq r_N \text{Id}_{\mathfrak{g}^N(\mathbb{R}^d)}.$$

(6) *There exists  $\varepsilon_0 = \varepsilon_0(N) \in (0, 1]$  such that, for all  $\varepsilon \in (0, \varepsilon_0]$  and  $x \in O_N$ ,*

$$\tilde{\Gamma}_N^\varepsilon(x) \tilde{\Gamma}_N^\varepsilon(x)^* \geq \frac{1}{2} \text{Id}_{\mathbb{R}\langle \mathcal{H} \rangle}, \quad \begin{pmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{pmatrix} \begin{pmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{pmatrix}^* \geq \frac{r_N}{2} \text{Id}_{\mathfrak{g}^N(\mathbb{R}^d)}.$$

PROOF. (2) and (3) are obvious. By (4.5) no component has a negative power of  $\varepsilon$ , from which (1) immediately follows. Noting that  $\tilde{\Gamma}_N^0(x) = [\text{Id}_{\mathbb{R}\langle \mathcal{H} \rangle} | *]$ , where  $*$  is a certain smooth function in  $x$ , we have (4). In a similar way, we show (5). Obviously,

$$\begin{pmatrix} \tilde{\Gamma}_N^0(x) \\ P_N \end{pmatrix} = \begin{pmatrix} \text{Id}_{\mathbb{R}\langle \mathcal{H} \rangle} & * \\ \mathbf{0} & \text{Id}_{\mathfrak{g}^N(\mathbb{R}^d)/\mathbb{R}\langle \mathcal{H} \rangle} \end{pmatrix}$$

is invertible for all  $x \in O_N$ . A standard compactness argument implies existence of such a positive constant  $r_N$ . (6) is immediate from (1), (4), (5). □

LEMMA 4.4. *Assume  $(\mathbf{ER})_{x_0}$  and  $N \geq N_0$ . For convenience we set  $Z_N^\varepsilon(x)$  to be either  $\tilde{\Gamma}_N^\varepsilon(x)U_1^N$  or*

$$\begin{pmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{pmatrix} U_1^N.$$

*Then, there exist a neighborhood  $O_N$  of  $x_0$  and  $\varepsilon_0 = \varepsilon_0(N) \in (0, 1]$  such that  $Z_N^\varepsilon(x)$  is non-degenerate in the sense of Malliavin uniformly in  $x \in O_N$  and  $\varepsilon \in (0, \varepsilon_0]$ , that is,*

$$\sup_{x \in O_N} \sup_{0 < \varepsilon \leq \varepsilon_0} \|\sigma[Z_N^\varepsilon(x)]^{-1}\|_{L^p} < \infty \quad (\text{for every } 1 < p < \infty).$$

PROOF. It is easy to see that

$$\sigma[\tilde{\Gamma}_N^\varepsilon(x)U_1^N] \geq \tilde{\Gamma}_N^\varepsilon(x)\sigma[U_1^N]\tilde{\Gamma}_N^\varepsilon(x)^* \geq \lambda[U_1^N] \cdot \tilde{\Gamma}_N^\varepsilon(x)\tilde{\Gamma}_N^\varepsilon(x)^*.$$

The other Wiener functional also satisfies a similar estimate. Then, this lemma easily follows from Lemma 4.3 (6), and Proposition 3.2. □

### 5. Stochastic differential equation on $\mathbb{R}^d$ .

For  $V_1, \dots, V_n \in \mathfrak{X}(\mathbb{R}^d)$  and  $0 < \varepsilon \leq 1$ , we consider the following Stratonovich-type SDE on  $\mathbb{R}^d$ :

$$dX^\varepsilon(t, x) = \varepsilon \sum_{i=1}^n V_i(X^\varepsilon(t, x)) \circ dw_t^i \quad \text{with } X^\varepsilon(0, x) = x. \tag{5.1}$$

When  $\varepsilon = 1$ , we simply write  $X(t, x)$  for  $X^\varepsilon(t, x)$ . By the well-known scaling property,  $(X^\varepsilon(t, x))_{t \geq 0}$  and  $(X(\varepsilon^2 t, x))_{t \geq 0}$  have the same law.

For SDE (5.1), we always assume that  $V_i^j := \langle dx^j, V_i \rangle$  has bounded derivatives of all order  $\geq 1$  ( $1 \leq i \leq n, 1 \leq j \leq d$ ). This is a standard assumption in Malliavin calculus.

The aim of this section is to prove that  $\delta_x(X^\varepsilon(1, x))$  admits an asymptotic expansion as  $\varepsilon \searrow 0$  in the space of Watanabe distributions uniformly in  $x$  under the equiregular Hörmander condition on the coefficient vector fields (Theorem 5.4). The expansion for each fixed  $x$  under the usual Hörmander condition was already proved in [35]. We carefully follow the argument in [35] and show the uniformity of the expansion under the equiregular condition. At the end of the section, we discuss the case of SDE with a nice drift term (Corollary 5.12).

Now we recall the stochastic Taylor expansion in  $\varepsilon$ . Note that (5.2)–(5.4) is an asymptotic expansion in  $\mathbb{D}_\infty$ -topology for each fixed  $x$  and  $t$ . The aim of the next proposition is to make sure the uniformity of the expansion in  $x$  as  $x$  varies in a compact subset.

PROPOSITION 5.1. *Let the notations be as above and let  $N \geq 1$ . Then, we have*

$$X^\varepsilon(t, x) = x + E_N^\varepsilon(t, x) + R_{N+1}^\varepsilon(t, x), \quad (5.2)$$

where we set

$$E_N^\varepsilon(t, x) = \sum_{I \in \mathcal{I}(N)} \varepsilon^{|I|} (V_I \text{Id})(x) w_t^I, \quad (5.3)$$

$$R_{N+1}^\varepsilon(t, x) = \varepsilon^{N+1} \sum_{I \in \mathcal{I}(N+1) \setminus \mathcal{I}(N)} \int_0^t \cdots \int_0^{t_3} \int_0^{t_2} (V_I \text{Id})(X^\varepsilon(t_1, x)) \circ dw_{t_1}^{i_1} \circ dw_{t_2}^{i_2} \cdots \circ dw_{t_{N+1}}^{i_{N+1}}. \quad (5.4)$$

Moreover, for each compact set  $K \subset \mathbb{R}^d$ , the asymptotic expansion (5.2) is uniform in  $(t, x) \in [0, 1] \times K$ , that is,

$$\sup_{0 \leq t \leq 1} \sup_{x \in K} \|R_{N+1}^\varepsilon(t, x)\|_{p,k} \leq C \varepsilon^{N+1} \quad (N \geq 1, 1 < p < \infty, k \in \mathbb{N}) \quad (5.5)$$

holds for all  $\varepsilon \in (0, 1]$ . Here,  $C = C(N, p, k)$  is a certain positive constant independent of  $\varepsilon$ .

PROOF. This is well-known at least when  $x$  is fixed. So we only give a sketch of proof so that one can see the uniformity in  $x$ . In this proof the positive constant  $C = C(N, p, k)$  may change from line to line.

Firstly, it is well-known that

$$\sup_{0 \leq t \leq 1} \sup_{x \in K} \|D^k X^\varepsilon(t, x)\|_{L^p} \leq C \varepsilon^k \quad (1 < p < \infty, k \in \mathbb{N}).$$

The proof is standard, although it is long and may not be so easy. Secondly, we can see

from (5.4) that

$$\sup_{0 \leq t \leq 1} \sup_{x \in K} \|R_{N+1}^\varepsilon(t, x)\|_{L^p} \leq C\varepsilon^{N+1} \quad (N \geq 1, 1 < p < \infty, k \in \mathbb{N}).$$

Thirdly,  $D^{N+1}E_N^\varepsilon(t, x) = 0$ .

Now we use the stronger form of Meyer’s inequality. If  $k \geq N + 1$ , then

$$\begin{aligned} \|R_{N+1}^\varepsilon(t, x)\|_{p,k} &\leq C(\|R_{N+1}^\varepsilon X^\varepsilon(t, x)\|_{L^p} + \|D^k R_{N+1}^\varepsilon(t, x)\|_{L^p}) \\ &\leq C(\|R_{N+1}^\varepsilon X^\varepsilon(t, x)\|_{L^p} + \|D^k X^\varepsilon(t, x)\|_{L^p}) \leq C\varepsilon^{N+1}. \end{aligned}$$

Since the Sobolev norm is increasing in  $k$ , we are done. □

We modify the stochastic Taylor expansion (5.2)–(5.4) for later use. The definition of  $F_N$  was given in (4.10).

PROPOSITION 5.2. *Let  $N \geq 1$ . Then, we have*

$$X^\varepsilon(t, x) = x + F_N(x, \Delta_\varepsilon^N U_t^N) + \hat{R}_{N+1}^\varepsilon(t, x). \tag{5.6}$$

Here we set

$$\begin{aligned} \hat{R}_{N+1}^\varepsilon(t, x) &= R_{N+1}^\varepsilon(t, x) \\ &\quad - \sum_{k=1}^N \frac{1}{k!} \sum_{|I_1|+\dots+|I_k|>N} \varepsilon^{|I_1|+\dots+|I_k|} (V_{[I_1]} \cdots V_{[I_k]} \text{Id})(x) U_t^{I_1} \cdots U_t^{I_k}, \end{aligned} \tag{5.7}$$

where the second summation runs over all  $(I_1, \dots, I_k) \in \mathcal{G}(N)^k$  such that  $|I_1| + \dots + |I_k| > N$ . Moreover, for each compact set  $K \subset \mathbb{R}^d$ ,  $\hat{R}_{N+1}^\varepsilon(t, x)$  satisfies the same estimate as in (5.5) for a different constant  $C > 0$ .

PROOF. The second assertion is almost obvious. Since it is immediate from (4.7)–(4.10) that

$$F_N(x, \Delta_\varepsilon^N U_t^N) = \sum_{k=1}^N \frac{1}{k!} \sum_{I_1, \dots, I_k \in \mathcal{G}(N)} \varepsilon^{|I_1|+\dots+|I_k|} (V_{[I_1]} \cdots V_{[I_k]} \text{Id})(x) U_t^{I_1} \cdots U_t^{I_k},$$

it is enough to see that

$$E_N^\varepsilon(t, x) = \sum_{k=1}^N \frac{1}{k!} \sum_{|I_1|+\dots+|I_k| \leq N} \varepsilon^{|I_1|+\dots+|I_k|} (V_{[I_1]} \cdots V_{[I_k]} \text{Id})(x) U_t^{I_1} \cdots U_t^{I_k}. \tag{5.8}$$

Here, the second summation runs over all  $(I_1, \dots, I_k) \in \mathcal{G}(N)^k$  such that  $|I_1| + \dots + |I_k| \leq N$ . Equality (5.8) immediately follows from Lemma 4.1 and the definition of  $U_t^N$ . □

Recall Kusuoka–Stroock’s estimate for Malliavin covariance matrix of  $X^\varepsilon(1, x)$  under the Hörmander condition at  $x_0$ . Our aim here is to make sure the estimate is uniform in

$x$  as it varies in a small neighborhood of  $x_0$ . Note that the equiregular condition is not needed here.

**PROPOSITION 5.3.** *Assume  $(\mathbf{HC})_{x_0}$  at  $x_0 \in \mathbb{R}^d$ . Then, there exist a neighborhood  $O$  of  $x_0$  and a positive constants  $M$  independent of  $p$ ,  $x$  and  $\varepsilon$  such that*

$$\sup_{x \in O} \sup_{0 < \varepsilon \leq 1} \varepsilon^M \|\det \sigma[X^\varepsilon(1, x)]^{-1}\|_{L^p} < \infty \quad \text{for every } p \in (1, \infty).$$

*In particular,  $X^\varepsilon(1, x)$  is non-degenerate in the sense of Malliavin for every  $\varepsilon \in (0, 1]$  and  $x \in O$ .*

**PROOF.** This is proved in [26, Theorem (2.17)].  $\square$

Due to the non-degeneracy of  $X^\varepsilon(1, x)$  in the sense of Malliavin, Watanabe's pullback of the delta function  $\delta_x(X^\varepsilon(1, x)) \in \tilde{\mathbb{D}}_{-\infty}$  is well-defined. Takanobu [35] showed that  $\delta_x(X^\varepsilon(1, x)) = \delta_0(X^\varepsilon(1, x) - x)$  admits an asymptotic expansion up to any order as  $\varepsilon \searrow 0$  in  $\tilde{\mathbb{D}}_{-\infty}$ -topology.

Now we present our main result in this section. This is a uniform version of Takanobu's main theorem in [35]. To prove the uniformity in the starting point  $x$ , we need to assume the equiregular Hörmander condition near  $x_0$ . (Note that the SDE in [35] has a drift term. On the other hand, the dependency on  $x$  is not studied in [35].)

**THEOREM 5.4.** *Assume  $(\mathbf{ER})_{x_0}$ . Then, there exists a decreasing sequence  $\{O_j\}_{j \geq 0}$  of neighborhoods of  $x_0$  such that the asymptotic expansion*

$$\delta_x(X^\varepsilon(1, x)) \sim \varepsilon^{-\nu} (\Theta_0(x) + \varepsilon \Theta_1(x) + \varepsilon^2 \Theta_2(x) + \dots) \quad \text{in } \tilde{\mathbb{D}}_{-\infty} \text{ as } \varepsilon \searrow 0 \quad (5.9)$$

*holds for every  $x \in O_0$  with the following properties: (i)  $\inf_{x \in O_0} \mathbb{E}[\Theta_0] > 0$ , (ii) for every  $j \geq 0$  there exists  $k = k(j) > 0$  such that*

$$\sup_{x \in O_j} \left\{ \|\Theta_j(x)\|_{p, -k} + \sup_{0 < \varepsilon \leq 1} \|\varepsilon^{-(j+1-\nu)} r_{j+1}^\varepsilon(x)\|_{p, -k} \right\} < \infty$$

*for all  $p \in (1, \infty)$ . Here, we set*

$$r_{j+1}^\varepsilon(x) = \delta_x(X^\varepsilon(1, x)) - \varepsilon^{-\nu} (\Theta_0(x) + \dots + \varepsilon^j \Theta_j(x)).$$

*Moreover,  $\Theta_{2j-1}(x; \cdot)$  is odd as a Wiener functional for every  $j \geq 1$  and  $x \in O_0$ , that is,  $\Theta_{2j-1}(x; -w) = -\Theta_{2j-1}(x; w)$ .*

**REMARK 5.5.** In fact,  $O_j \ni x \mapsto \Theta_j(x) \in \tilde{\mathbb{D}}_{-\infty}$  is continuous for every  $j$ . This follows from the uniformity of the asymptotic expansion (5.9) and continuity of  $x \mapsto \delta_x(X^\varepsilon(1, x)) = \delta_0(X^\varepsilon(1, x) - x) \in \tilde{\mathbb{D}}_{-\infty}$ . The latter, in turn, follows from Proposition 5.3 and the continuity of  $x \mapsto X^\varepsilon(1, x) - x \in \mathbb{D}_\infty$ .

The rest of this section is devoted to proving the above theorem. The neighborhoods  $O$  and  $O_j$ ,  $j \geq 0$ , may change from line to line.

We introduce a few functions for technical purposes. Take  $\psi \in C^\infty(\mathbb{R}, [0, 1])$  such that  $\psi(s) = 0$  if  $|s| \geq 1$  and  $\psi(s) = 1$  if  $|s| \leq 1/2$ . Take any  $\kappa > 0$  and set  $\psi_N(x) = \psi(x/(\kappa/2)^2)$  for  $N \geq N_0$ . Define a smooth Wiener function  $\chi_N^\varepsilon \in \mathbb{D}_\infty$  by  $\chi_N^\varepsilon = \psi_N(|\Delta_\varepsilon^N U_1^N|^2)$ .

LEMMA 5.6. *Assume  $(\mathbf{HC})_{x_0}$  and  $N \geq N_0$ . Then, there exist a positive constant  $k$  independent of  $N$  and a neighborhood  $O_N$  of  $x_0$  such that the following property holds: For every  $p \in (1, \infty)$ , there exist positive constants  $c_1$  and  $c_2$  independent of  $\varepsilon$  and  $x \in O_N$  such that*

$$\sup_{x \in O_N} \|\delta_x(X^\varepsilon(1, x)) - \chi_N^\varepsilon \cdot \delta_x(X^\varepsilon(1, x))\|_{p, -k} \leq c_1 e^{-c_2/\varepsilon^2} \quad \text{as } \varepsilon \searrow 0.$$

PROOF. We use [23, p.374, Formula (8.47)]: For every  $q, r \in (1, \infty)$  such that  $1/p := 1/q + 1/r < 1$  and every  $k \in \mathbb{N}$ , there exists a positive constant  $C_{q,r,k}$  such that

$$\|FG\|_{p, -k} \leq C_{q,r,k} \|F\|_{q,k} \|G\|_{r, -k} \quad (F \in \mathbb{D}_{q,k}, G \in \mathbb{D}_{r, -k}).$$

We use this formula with  $F = 1 - \chi_N^\varepsilon$  and  $G = \delta_x(X^\varepsilon(1, x))$ . By Proposition 5.3 and Watanabe’s pullback theorem, we can find  $k$  and  $M' > 0$  such that

$$\sup_{x \in O} \sup_{0 < \varepsilon \leq 1} \varepsilon^{M'} \|\delta_x(X^\varepsilon(1, x))\|_{r, -k} < \infty$$

for any  $r \in (1, \infty)$ . On the other hand, we can easily see from (3.5) that

$$\|1 - \chi_N^\varepsilon\|_{q,k} = O(e^{-C_{N,\kappa}/4\varepsilon^2}) \quad \text{as } \varepsilon \searrow 0$$

for every  $q \in (1, \infty)$  and  $k \in \mathbb{N}$ . This completes the proof. □

The following is a slight extension of [35, Lemma 5.8]. Under the equiregular Hörmander condition near  $x_0$ , we prove a uniform version of the lemma. Recall that, for a Wiener functional  $G$ ,  $\lambda[G]$  stands for the lowest eigenvalue of Malliavin covariance matrix  $\sigma[G]$ .

PROPOSITION 5.7. *Assume  $(\mathbf{ER})_{x_0}$  and let  $r > 0$  be the constant in (4.12) and  $N \geq N_0$ . Then, there exist a neighborhood  $O_N$  of  $x_0$  and  $\kappa_N > 0$  such that, for all  $\varepsilon \in (0, 1]$  and  $x \in O_N$ , it holds that*

$$\lambda[F_N(x, \Delta_\varepsilon^N U_1^N)] \geq r\varepsilon^{2N_0} \lambda[U_1^N] \quad \text{on } \{|\Delta_\varepsilon^N U_1^N| < \kappa_N\}. \tag{5.10}$$

*In particular, for  $\varepsilon \in (0, 1]$  and  $x \in O_N$ , we have  $\lambda[F_N(x, \Delta_\varepsilon^N U_1^N)] > 0$  almost surely on  $\{|\Delta_\varepsilon^N U_1^N| < \kappa_N\}$  and*

$$\mathbb{E}[\lambda[F_N(x, \Delta_\varepsilon^N U_1^N)]^{-p} ; |\Delta_\varepsilon^N U_1^N| < \kappa_N]^{1/p} \leq r^{-1} \varepsilon^{-2N_0} \|\lambda[U_1^N]^{-1}\|_{L^p} < \infty \tag{5.11}$$

*for every  $1 < p < \infty$ .*

PROOF. Inequality (5.11) is immediate from (5.10) and Proposition 3.2. We give a quick proof of (5.10). In the same way as in the proof of [35, Lemma 5.8], it holds that, for every  $z \in \mathbb{R}^d$ ,

$$\begin{aligned} & \langle \sigma[F_N(x, \Delta_\varepsilon^N U_1^N)]z, z \rangle \\ &= \sum_{I, J \in \mathcal{G}(N)} \sigma[U_1^N]^{IJ} \left( \sum_{i=1}^d z^i \varepsilon^{|I|} \partial_I F_N^i(x, \Delta_\varepsilon^N U_1^N) \right) \left( \sum_{j=1}^d z^j \varepsilon^{|J|} \partial_J F_N^j(x, \Delta_\varepsilon^N U_1^N) \right) \\ &\geq \lambda[U_1^N] \sum_{I \in \mathcal{G}(N)} \varepsilon^{2|I|} \left( \sum_{i=1}^d z^i \partial_I F_N^i(x, \Delta_\varepsilon^N U_1^N) \right)^2 \\ &\geq \lambda[U_1^N] \sum_{I \in \mathcal{G}(N_0)} \varepsilon^{2|I|} \left( \sum_{i=1}^d z^i \partial_I F_N^i(x, \Delta_\varepsilon^N U_1^N) \right)^2. \end{aligned}$$

The first equality is immediate from the chain rule for the  $H$ -derivative  $D$ . By (4.12), the right-hand side is bounded from below by

$$\varepsilon^{2N_0} \lambda[U_1^N] \cdot |[(\partial_I F_N^i(x, \Delta_\varepsilon^N U_1^N))_{1 \leq i \leq d, I \in \mathcal{G}(N_0)}]^* z|^2 \geq r \varepsilon^{2N_0} \lambda[U_1^N] |z|^2$$

uniformly in  $x \in O_N$ .  $\square$

In what follows, we choose  $\kappa_N > 0$  as in Proposition 5.7 and set  $\psi_N(x) = \psi(x/(\kappa_N/2)^2)$  and  $\chi_N^\varepsilon = \psi_N(|\Delta_\varepsilon^N U_1^N|^2)$  for  $N \geq N_0$ .

Since non-degeneracy of  $F_N(x, \Delta_\varepsilon^N U_1^N)$  is not known, we cannot use the standard version of Watanabe's pullback (see Item (c), Section 2) to justify  $\delta_0(F_N(x, \Delta_\varepsilon^N U_1^N))$ . However, thanks to Proposition 5.7 above, a modified version of Watanabe's pullback is available.

PROPOSITION 5.8. *Assume  $(\mathbf{ER})_{x_0}$ ,  $N \geq N_0$  and let  $O_N$  be as in Proposition 5.7. Fix any  $\varepsilon \in (0, 1]$  and  $x \in O_N$ . Then, the mapping  $\mathcal{S}(\mathbb{R}^d) \ni \phi \mapsto \chi_N^\varepsilon \cdot \phi(F_N(x, \Delta_\varepsilon^N U_1^N)) \in \mathbb{D}_\infty$  uniquely extends to a continuous linear mapping*

$$\mathcal{S}'(\mathbb{R}^d) \ni \Phi \mapsto \chi_N^\varepsilon \cdot \Phi(F_N(x, \Delta_\varepsilon^N U_1^N)) \in \tilde{\mathbb{D}}_{-\infty}.$$

PROOF. This fact is actually well-known to experts of Malliavin calculus. The key point is the integrability (5.11) in Proposition 5.7. For a detailed proof, see Yoshida [41].  $\square$

The next lemma states that  $\delta_x(X^\varepsilon(1, x))$  can be approximated by  $\delta_0(F_N(x, \Delta_\varepsilon^N U_1^N))$  uniformly in  $x$  if  $N$  is large enough. Therefore, the problem reduces to the expansion of the latter Watanabe distribution.

LEMMA 5.9. *Assume  $(\mathbf{ER})_{x_0}$ . Then, there exist  $k > 0$ , a sequence of  $\{O_N\}_{N \geq N_0}$  neighborhoods of  $x_0$  and a sequence  $\{l_N\}_{N \geq N_0}$  of real numbers diverging to  $+\infty$  such that, for every  $p \in (1, \infty)$  and  $N \geq N_0$ ,*

$$\sup_{x \in O_N} \|\delta_x(X^\varepsilon(1, x)) - \chi_N^\varepsilon \cdot \delta_0(F_N(x, \Delta_\varepsilon^N U_1^N))\|_{p, -k} = O(\varepsilon^{l_N}) \quad \text{as } \varepsilon \searrow 0.$$

PROOF. Due to Lemma 5.6, it is sufficient to show that

$$\sup_{x \in O_N} \|\chi_N^\varepsilon \cdot \delta_0(X^\varepsilon(1, x) - x) - \chi_N^\varepsilon \cdot \delta_0(F_N(x, \Delta_\varepsilon^N U_1^N))\|_{p, -k} = O(\varepsilon^{l_N}) \quad \text{as } \varepsilon \searrow 0. \tag{5.12}$$

As always, the key tool is the integration by parts formula for Watanabe distributions. We also use the estimates in Proposition 5.2 (ii), Proposition 5.3, Proposition 5.7, Proposition 5.8. In this proof we write  $A^\varepsilon = X^\varepsilon(1, x) - x$  and  $B^\varepsilon = F_N(x, \Delta_\varepsilon^N U_1^N)$  for simplicity.

First, we prove the case  $d = 1$  to observe what is happening. Set  $g(x) = x \vee 0$  for  $x \in \mathbb{R}$ . Then,  $g''(x) = \delta_0(x)$  in the distributional sense. Choose smooth functions  $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$  ( $i = 1, 2, 3$ ) so that  $\psi_1 = \psi$ ,  $\psi_i \equiv 1$  on the support of  $\psi_{i-1}$  ( $i = 2, 3$ ), and the support of  $\psi_3$  is contained in  $(-2, 2)$ . Set  $\chi_{N,i}^\varepsilon = \psi_i(|\Delta_\varepsilon^N U_1^N|^2 / (\kappa_N/2)^2)$ . Note that Proposition 5.8 still holds even if  $\chi_N^\varepsilon = \chi_{N,1}^\varepsilon$  is replaced by  $\chi_{N,2}^\varepsilon$  or  $\chi_{N,3}^\varepsilon$ . Note also that  $\|\chi_{N,i}^\varepsilon\|_{p,k}$  is bounded in  $\varepsilon$  for any  $1 < p < \infty$ ,  $k \geq 0$ ,  $1 \leq i \leq 3$ .

Take any  $G \in \mathbb{D}_\infty$ . By integration by parts formula and the way  $\psi_i$  ( $i = 1, 2, 3$ ) are defined, we have

$$\begin{aligned} \langle \chi_N^\varepsilon \cdot \delta_0(B^\varepsilon), G \rangle &= \langle \chi_{N,2}^\varepsilon \chi_{N,3}^\varepsilon g''(B^\varepsilon), \chi_{N,1}^\varepsilon G \rangle \\ &= \left\langle D[\chi_{N,2}^\varepsilon \chi_{N,3}^\varepsilon g'(B^\varepsilon)], \frac{DB^\varepsilon}{\|DB^\varepsilon\|_H^2} \chi_{N,1}^\varepsilon G \right\rangle \\ &= \left\langle \chi_{N,2}^\varepsilon \chi_{N,3}^\varepsilon g'(B^\varepsilon), D^* \left[ \frac{DB^\varepsilon}{\|DB^\varepsilon\|_H^2} \chi_{N,1}^\varepsilon G \right] \right\rangle, \end{aligned}$$

where  $D$  is the  $H$ -derivative (the gradient operator) and  $D^*$  is its adjoint. (Thanks to Proposition 5.7, the right-hand side is well-defined.) Note that  $\|DB^\varepsilon\|_H^2 = \det \sigma[B^\varepsilon]$  since  $d = 1$ . Therefore, the second component of the pairing on the right-hand side coincides at least formally with  $\Phi$  in (2.2) with  $m = 1$  and  $F$  and  $G$  being replaced by  $B^\varepsilon$  and  $\chi_{N,1}^\varepsilon G$ , respectively.

Using the formula again, we have

$$\begin{aligned} \langle \chi_N^\varepsilon \cdot \delta_0(B^\varepsilon), G \rangle &= \left\langle g(B^\varepsilon), \chi_{N,3}^\varepsilon D^* \left[ \frac{DB^\varepsilon}{\|DB^\varepsilon\|_H^2} \chi_{N,2}^\varepsilon D^* \left[ \frac{DB^\varepsilon}{\|DB^\varepsilon\|_H^2} \chi_{N,1}^\varepsilon G \right] \right] \right\rangle \\ &= \left\langle g(B^\varepsilon), D^* \left[ \frac{DB^\varepsilon}{\|DB^\varepsilon\|_H^2} \chi_{N,2}^\varepsilon D^* \left[ \frac{DB^\varepsilon}{\|DB^\varepsilon\|_H^2} \chi_{N,1}^\varepsilon G \right] \right] \right\rangle. \end{aligned}$$

This equation still holds for  $A^\varepsilon$  instead of  $B^\varepsilon$  for the same reason. Observe that on the right-hand side  $B^\varepsilon$  is plugged into a (Lipschitz) continuous function  $g$ , not a Schwartz distribution. Hence, the difference  $\|g(A^\varepsilon) - g(B^\varepsilon)\|_{L^p}$  is dominated by  $\|A^\varepsilon - B^\varepsilon\|_{L^p} = O(\varepsilon^{N+1})$ , where Proposition 5.2 (ii) is used.

By straight forward calculations, we can show the following estimate: There exist constants  $a \in \mathbb{N}$  (independent of  $\varepsilon, N, p, x$ ) and  $C_p > 0$  (independent of  $\varepsilon, N, x$ ) such that

$$\left| \langle \chi_N^\varepsilon \cdot \delta_0(A^\varepsilon) - \chi_N^\varepsilon \cdot \delta_0(B^\varepsilon), G \rangle \right| \leq C_p \|G\|_{q,2} \varepsilon^{N+1-a(M+N_0)} \tag{5.13}$$

for every  $p \in (1, \infty)$ , where  $1/p + 1/q = 1$  and every  $\varepsilon \in (0, 1]$  and  $x \in O$ . Here, we used Propositions 5.2 (ii), 5.3, 5.7, and 5.8 ( $M$  is the positive constant in Propositions 5.3). This implies (5.12) when  $d = 1$  with  $k = 2$  and  $l_N = N + 1 - a(M + N_0)$ .

The proof for  $d \geq 2$  is essentially the same in spirit, but the notations get quite complicated and we have to use the integration by parts formula many times ( $2d$ -times is enough). Note that the differentiability index  $-k$  in (5.12) is determined by this number and hence depends only on  $d$ .

Set  $g(x) = \prod_{i=1}^d (x_i \vee 0)$  for  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ . Then,  $(\partial_1^2 \cdots \partial_d^2 g)(x) = \delta_0(x)$  in the distributional sense. Choose smooth functions  $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$  ( $i \geq 0$ ) so that  $\psi_0 = \psi$ ,  $\psi_i \equiv 1$  on the support of  $\psi_{i-1}$  ( $i \geq 1$ ), and the support of  $\psi_i$  is contained in  $(-2, 2)$ . Set  $\chi_{N,i}^\varepsilon = \psi_i(|\Delta_\varepsilon^N U_1^N|^2 / (\kappa_N/2)^2)$ . For every  $i \geq 0$ , Proposition 5.8 still holds for  $\chi_{N,i}^\varepsilon$  and  $\|\chi_{N,i}^\varepsilon\|_{p,k}$  is bounded in  $\varepsilon$  for any  $1 < p < \infty$ ,  $k \geq 0$ .

For a multi-index  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ , set  $i_\alpha = \max\{i; \alpha_i \neq 0\}$  and define  $\alpha' = (\alpha_1 - \delta_{1i_\alpha}, \dots, \alpha_d - \delta_{di_\alpha})$ ,  $\delta_{ij}$  being Kronecker's delta. We define  $\Phi_{(\alpha)}$  with respect to  $B^\varepsilon$  as follows. If  $|\alpha| := \sum_{k=1}^d \alpha_k = 1$ , we set

$$\Phi_{(\alpha)}(\cdot; G) = \Phi_{i_\alpha}(\cdot; \chi_{N,1}^\varepsilon G).$$

Recall that  $\Phi_i$  is given in (2.2) with  $F$  being replaced by  $B^\varepsilon$ . Thanks to the ‘‘cutoff’’ functional  $\chi_{N,1}^\varepsilon$ ,  $\Phi_{i_\alpha}(\cdot; \chi_{N,1}^\varepsilon G)$  is well-defined, though  $B^\varepsilon$  is not non-degenerate in the standard sense of Malliavin calculus. If  $|\alpha| \geq 2$ , we set

$$\Phi_{(\alpha)}(\cdot; G) = \Phi_{i_\alpha}(\cdot; \chi_{N,|\alpha|}^\varepsilon \Phi_{(\alpha')}(\cdot; G)).$$

Using the integration by parts formula (2.1) repeatedly in the same way as above, we can show that

$$\langle \chi_N^\varepsilon \cdot \delta_0(B^\varepsilon), G \rangle = \langle g(B^\varepsilon), \Phi_{(\alpha)}(\cdot; G) \rangle \quad (5.14)$$

for every  $G \in \mathbb{D}_\infty$ , where  $\alpha = (2, 2, \dots, 2) \in \mathbb{N}^d$ . Note that (5.14) can be viewed as the definition of the Watanabe distribution  $\chi_N^\varepsilon \cdot \delta_0(B^\varepsilon)$ .

One can also define  $\Phi_{(\alpha)}(\cdot; G)$  for  $A^\varepsilon$  instead of  $B^\varepsilon$  in the same way. Then, (5.14) holds for  $A^\varepsilon$ , too. Once we have (5.14) for both  $B^\varepsilon$  and  $A^\varepsilon$ , it is straightforward to check that (5.13) also holds in the multi-dimensional case for with the differentiability index  $2d$  instead of 2 (and possibly different  $a$ ).  $\square$

In what follows we expand  $\chi_N^\varepsilon \cdot \delta_0(F_N(x, \Delta_\varepsilon^N U_1^N))$  for each fixed  $N \geq N_0$ . In the next lemma the same  $\varepsilon_0 = \varepsilon_0(N)$  as in Lemmas 4.3 and 4.4 will do. Note that (4.11) is implicitly used.

**LEMMA 5.10.** *Assume  $(\mathbf{ER})_{x_0}$  and  $N \geq N_0$ . Then, there exist a neighborhood  $O_N$  of  $x_0$  and  $\varepsilon_0 = \varepsilon_0(N) \in (0, 1]$  such that*

$$\chi_N^\varepsilon \cdot \delta_0(F_N(x, \Delta_\varepsilon^N U_1^N)) = \varepsilon^{-\nu} |\det B_{\mathcal{H}}(x)|^{-1} \frac{\chi_N^\varepsilon}{\det M_N(x, \Delta_\varepsilon^N U_1^N)} \cdot \delta_0(\tilde{\Gamma}_N^\varepsilon(x) U_1^N) \quad (5.15)$$

holds for all  $x \in O_N$  and  $\varepsilon \in (0, \varepsilon_0]$ . Here,  $\delta_0$  on the right-hand side is the delta function defined on  $\mathbb{R}(\mathcal{H})$ .

PROOF. We follow [35, pp.189–191]. Since (5.15) is an equality and have nothing to do with the uniformity in  $x$  of the asymptotic expansion, some parts of the proof here is not so detailed as the corresponding part in [35].

It is easy to see that

$$U_1^N = \begin{bmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{bmatrix} U_1^N = \begin{bmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\Gamma}_N^\varepsilon(x) U_1^N \\ P_N U_1^N \end{bmatrix}.$$

From Lemma 4.3 (2) and an obvious fact that  $\Gamma_N = B_{\mathcal{H}}^{-1} B_N$  we see easily that

$$F_N(x, \Delta_\varepsilon^N U_1^N) = M_N(x, \Delta_\varepsilon^N U_1^N) B_{\mathcal{H}}(x) \Delta_\varepsilon^{\mathcal{H}} \tilde{\Gamma}_N^\varepsilon(x) U_1^N.$$

Next, take a non-negative test function  $g$  on  $\mathbb{R}^d$  such that  $\int g = 1$  and set  $g_\kappa(x) = \kappa^{-d} g(x/\kappa)$  for  $\kappa > 0$ . Then,  $g_\kappa \rightarrow \delta_0$  in  $\mathcal{S}'(\mathbb{R}^d)$  as  $\kappa \rightarrow 0$ . By the modified version of Watanabe’s theory (Proposition 5.8),

$$\chi_N^\varepsilon \cdot g_\kappa(F_N(x, \Delta_\varepsilon^N U_1^N)) \rightarrow \chi_N^\varepsilon \cdot \delta_0(F_N(x, \Delta_\varepsilon^N U_1^N))$$

in  $\tilde{\mathbb{D}}_{-\infty}$  as  $\kappa \rightarrow 0$ .

Before we start computing this quantity, we set some notations for simplicity. Set  $T = \Delta_\varepsilon^N$ ,

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma}_N^\varepsilon(x) U_1^N \\ P_N U_1^N \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} \tilde{\Gamma}_N^\varepsilon(x) \\ P_N \end{bmatrix}, \quad \text{then} \quad U_1^N = C^{-1} \begin{bmatrix} V \\ W \end{bmatrix}.$$

Then, we have for every  $G \in \mathbb{D}_\infty$  that

$$\begin{aligned} & \mathbb{E}[G \chi_N^\varepsilon \cdot g_\kappa(F_N(x, \Delta_\varepsilon^N U_1^N))] \\ &= \mathbb{E}\left[G \psi_N \left( \left| TC^{-1} \begin{bmatrix} V \\ W \end{bmatrix} \right|^2 \right) g_\kappa \left( M_N \left( x, TC^{-1} \begin{bmatrix} V \\ W \end{bmatrix} \right) B_{\mathcal{H}}(x) \Delta_\varepsilon^{\mathcal{H}} V \right) \right] \\ &= \int_{\mathbb{R}(\mathcal{H})} dv \int_{\mathbb{R}(\mathcal{H})^\perp} dw \left\langle G, \delta_{(v,w)} \left( \begin{bmatrix} V \\ W \end{bmatrix} \right) \right\rangle \psi_N \left( \left| TC^{-1} \begin{bmatrix} v \\ w \end{bmatrix} \right|^2 \right) \\ & \quad \times g_\kappa \left( M_N \left( x, TC^{-1} \begin{bmatrix} v \\ w \end{bmatrix} \right) B_{\mathcal{H}}(x) \Delta_\varepsilon^{\mathcal{H}} v \right). \end{aligned}$$

(Since it is difficult to put a column vector as a subscript of  $\delta$ , we wrote  $\delta_{(v,w)\cdot}$ .) We change variables by  $v \mapsto (\Delta_\varepsilon^{\mathcal{H}})^{-1} \kappa v$  and use Lemma 4.3 (3). Then,

$$\begin{aligned} & \mathbb{E}[G \chi_N^\varepsilon \cdot g_\kappa(F_N(x, \Delta_\varepsilon^N U_1^N))] \\ &= \varepsilon^{-\nu} \int_{\mathbb{R}(\mathcal{H})} dv \int_{\mathbb{R}(\mathcal{H})^\perp} dw \left\langle G, \delta_{(\Delta_\varepsilon^{\mathcal{H}} \kappa v, w)} \left( \begin{bmatrix} V \\ W \end{bmatrix} \right) \right\rangle \end{aligned}$$

$$\times \psi_N \left( \left| TC^{-1} \begin{bmatrix} \Delta_\varepsilon^{\mathcal{H}} \kappa v \\ w \end{bmatrix} \right|^2 \right) g \left( M_N \left( x, TC^{-1} \begin{bmatrix} \Delta_\varepsilon^{\mathcal{H}} \kappa v \\ w \end{bmatrix} \right) B_{\mathcal{H}}(x)v \right).$$

Now, we use the dominated convergence theorem for  $dvdw$ -integration as  $\kappa \searrow 0$ . Due to (4.11), we can find a large constant  $R > 0$  independent of  $\kappa$  such that the integrand above is dominated by  $R \cdot \mathbf{1}_{\{|v| < R, |w| < R\}}$ . ( $R$  may depend on other parameters.) Letting  $\kappa \searrow 0$ , we have

$$\begin{aligned} \mathbb{E} [ G \chi_N^\varepsilon \cdot \delta_0 ( F_N ( x, \Delta_\varepsilon^N U_1^N ) ) ] &= \varepsilon^{-\nu} \int_{\mathbb{R} \langle \mathcal{H} \rangle} dv \int_{\mathbb{R} \langle \mathcal{H} \rangle^\perp} dw \left\langle G, \delta_{(0,w)} \left( \begin{bmatrix} V \\ W \end{bmatrix} \right) \right\rangle \\ &\quad \cdot \psi_N \left( \left| TC^{-1} \begin{bmatrix} 0 \\ w \end{bmatrix} \right|^2 \right) g \left( M_N \left( x, TC^{-1} \begin{bmatrix} 0 \\ w \end{bmatrix} \right) B_{\mathcal{H}}(x)v \right). \end{aligned}$$

Changing variables again by  $v \mapsto \{ M_N ( x, TC^{-1} \begin{bmatrix} 0 \\ w \end{bmatrix} ) B_{\mathcal{H}}(x) \}^{-1} v$ , we have

$$\begin{aligned} \mathbb{E} [ G \chi_N^\varepsilon \cdot \delta_0 ( F_N ( x, \Delta_\varepsilon^N U_1^N ) ) ] &= \varepsilon^{-\nu} |\det B_{\mathcal{H}}(x)|^{-1} \\ &\quad \times \left\langle G, \frac{\chi_N^\varepsilon}{\det M_N ( x, TU_1^N )} \int_{\mathbb{R} \langle \mathcal{H} \rangle^\perp} dw \delta_{(0,w)} \left( \begin{bmatrix} V \\ W \end{bmatrix} \right) \right\rangle. \end{aligned}$$

It is easy to see from Lemma 4.4 that

$$\int_{\mathbb{R} \langle \mathcal{H} \rangle^\perp} dw \delta_{(0,w)} \left( \begin{bmatrix} V \\ W \end{bmatrix} \right) = \delta_0(V).$$

This completes the proof. □

Now we are in a position to prove our main result in this section.

PROOF OF THEOREM 5.4. We expand the (generalized) Wiener functionals on the right-hand side of (5.15). First, note that

$$\Delta_\varepsilon^N U_1^N = \sum_{I \in \mathcal{G}(N)} \varepsilon^{|I|} U_1^I.$$

This is just a polynomial in  $\varepsilon$  whose coefficients belong to an inhomogeneous Wiener chaos.

By the choice of  $\psi$  and a routine argument, we have

$$\chi_N^\varepsilon = \psi \left( \frac{|\Delta_\varepsilon^N U_1^N|^2}{(\kappa_N/2)^2} \right) = 1 + O(\varepsilon^\infty) \quad \text{in } \mathbb{D}_\infty \tag{5.16}$$

as  $\varepsilon \searrow 0$ . Therefore, this is actually a dummy factor introduced for technical purposes and makes no contribution to the asymptotic expansion. Obviously,  $x$  is not involved in this functional.

By the definition in (4.13) and (a comment after that),  $\tilde{\Gamma}_N^\varepsilon(x) U_1^N$  is also a polynomial in  $\varepsilon$  that takes values in  $\mathcal{L}(\mathfrak{g}^N(\mathbb{R}^d), \mathbb{R} \langle \mathcal{H} \rangle)$  whose coefficients belong to an inhomogeneous

Wiener chaos. (Moreover, it depends smoothly in  $x$ .) Since this is uniformly non-degenerate (see Lemma 4.4), we can use the standard version of Watanabe’s theory (2.3) to obtain the following asymptotic expansion:

$$\delta_0(\tilde{\Gamma}_N^\varepsilon(x)U_1^N) = \mathcal{Y}_0^N(x) + \varepsilon\mathcal{Y}_1^N(x) + \varepsilon^2\mathcal{Y}_2^N(x) + \cdots \quad \text{in } \tilde{\mathbb{D}}_{-\infty} \quad (5.17)$$

as  $\varepsilon \searrow 0$ . Since Lemma 4.4 claims uniform dependency in  $x$ , this expansion is uniform in  $x \in O_N$ . By Lemma 4.3 (1) and a comment after (4.14),  $\mathcal{Y}_0^N(x) = \delta_0(\tilde{\Gamma}_N^0(x)U_1^N) = \delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0})$ .

By the explicit definition of  $M_N$  in (4.9) and the uniform lower bound of  $\det M_N$  in (4.11), we also obtain the following asymptotic expansion uniformly in  $x \in O_N$ :

$$\det M_N(x, \Delta_\varepsilon^N U_1^N)^{-1} = 1 + \varepsilon\mathcal{Z}_1^N(x) + \varepsilon^2\mathcal{Z}_2^N(x) + \cdots \quad \text{in } \mathbb{D}_\infty \quad (5.18)$$

as  $\varepsilon \searrow 0$ .

Take  $L > 0$  arbitrarily large. We will show that  $\delta_x(X^\varepsilon(1, x))$  admits an asymptotic expansion up to order  $L$  as  $\varepsilon \searrow 0$ . For this  $L$ , we choose  $N \geq N_0$  so that  $l_N \geq L + \nu + 1$ . Here,  $\{l_N\}$  is the diverging sequence given in Lemma 5.9. We also take  $O_N$  small enough so that all the previous results are available.

From Lemma 5.9, Lemma 5.10 and (5.16)–(5.18), we obtain the following asymptotics in  $\tilde{\mathbb{D}}_{-\infty}$  as  $\varepsilon \searrow 0$  uniformly in  $x \in O_N$  :

$$\begin{aligned} \delta_x(X^\varepsilon(1, x)) &= |\det B_{\mathcal{H}}(x)|^{-1}\varepsilon^{-\nu} \\ &\quad \times \{ \delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0}) + \varepsilon\Theta_1^N(x) + \cdots + \varepsilon^{L+\nu}\Theta_{L+\nu}^N(x) \} + O(\varepsilon^{L+1}) \end{aligned} \quad (5.19)$$

for some  $\Theta_j^N(x) \in \tilde{\mathbb{D}}_{-\infty}$  ( $1 \leq j \leq L + \nu$ ). Since the coefficients of an asymptotic expansion are uniquely determined,  $\Theta_j^N(x)$  is actually independent of the choice of  $N$ . This proves (5.9).

By a routine argument,  $\Theta_j^N(x; -w) = -\Theta_j^N(x; w)$  as a generalized Wiener functional if  $j$  is odd. This implies  $\mathbb{E}[\Theta_j^N(x)] = 0$  if  $j$  is odd.

Finally, we show  $\mathbb{E}[\delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0})] > 0$ . Recall that  $\tilde{\Gamma}_{N_0}^0(x) = [\text{Id}_{\mathbb{R}\langle\mathcal{H}\rangle} | *]$  is a (possibly non-orthogonal) projection from  $\mathfrak{g}^{N_0}(\mathbb{R}^d)$  to  $\mathbb{R}\langle\mathcal{H}\rangle$ . If we denote by  $q^{N_0}$  the smooth density of the law of  $U_1^{N_0}$  on  $\mathfrak{g}^{N_0}(\mathbb{R}^d)$ , then

$$\mathbb{E}[\delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0})] = K_x \int_{\ker \tilde{\Gamma}_{N_0}^0(x)} q^{N_0}(u) du,$$

where  $du$  is the Lebesgue measure on the subspace  $\ker \tilde{\Gamma}_{N_0}^0(x)$  and  $K_x > 0$  is a constant which depends on “the angle” of the kernel and the image of the projection  $\tilde{\Gamma}_{N_0}^0(x)$ . (If the projection is orthogonal, then  $K_x = 1$ .)

Since the everywhere positivity of  $q^{N_0}$  is shown in [35, p.202] or originally in Kunita [25], we have  $\mathbb{E}[\delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0})] > 0$  and the proof of Theorem 5.4 is done.  $\square$

REMARK 5.11. There is another way to prove that  $q^N$  is everywhere positive. Let  $u^N(h)$  be the solution of the skeleton ODE which corresponds to SDE (3.3) driven by  $h \in H$ . In other words,  $u^N(h) = \log y^N(h)$ . It is sufficient to show that, for every  $u \in \mathfrak{g}^N(\mathbb{R}^d)$ ,

there exists  $h \in H$  such that  $u_1^N(h) = u$  and the tangent map  $Du_1^N(h): H \rightarrow \mathfrak{g}^N(\mathbb{R}^d)$  is surjective. (See Aida–Kusuoka–Stroock [1] for example. See also Remark 3.1.)

Such an  $h$  can be found as follows. Take any Cameron–Martin path  $k: [0, 1/2] \rightarrow \mathbb{R}^n$  such that  $Du_{1/2}^N(k)$  is surjective (it does exist). Since  $G^N(\mathbb{R}^n) = \{y_{1/2}^N(h) \mid h \in H\}$ , there exists a Cameron–Martin path  $\hat{k}: [0, 1/2] \rightarrow \mathbb{R}^n$  such that  $u_{1/2}^N(\hat{k}) = u_{1/2}^N(k)^{-1} \times u$ . Then, the concatenated path  $k * \hat{k} \in H$  is the desired path. Here,  $k * \hat{k}$  is defined to be  $k$  on  $[0, 1/2]$  and  $\hat{k}(\cdot - 1/2) + k(1/2)$  on  $[1/2, 1]$ . Here, we used the left-invariance with respect to the product  $\times$ .

As a corollary of Theorem 5.4, we consider an SDE with drift instead of the driftless SDE (5.1) and prove an asymptotic expansion of the associated heat kernel under an assumption that the drift vector field  $V_0$  can be written as a linear combination of  $V_1, \dots, V_n$ . It is important that the leading positive constant in the expansion is independent of such  $V_0$ .

For  $V_0, V_1, \dots, V_n \in \mathfrak{X}(\mathbb{R}^d)$  and  $0 < \varepsilon \leq 1$ , we consider the following SDE on  $\mathbb{R}^d$ :

$$d\hat{X}^\varepsilon(t, x) = \varepsilon \sum_{i=1}^n V_i(\hat{X}^\varepsilon(t, x)) \circ dw_t^i + \varepsilon^2 V_0(\hat{X}^\varepsilon(t, x)) dt \quad \text{with } \hat{X}^\varepsilon(0, x) = x. \quad (5.20)$$

We continue to assume that  $V_i^j := \langle dx^j, V_i \rangle$  has bounded derivatives of all order  $\geq 1$  ( $0 \leq i \leq n, 1 \leq j \leq d$ ). By the scaling property,  $(\hat{X}^\varepsilon(t, x))_{t \geq 0}$  and  $(\hat{X}(\varepsilon^2 t, x))_{t \geq 0}$  have the same law. Here, we simply write  $\hat{X}(t, x)$  for  $\hat{X}^\varepsilon(t, x)$  when  $\varepsilon = 1$ . When it exists, we denote by  $p_t(x, x')$  the heat kernel associated with  $\hat{X}(t, x)$ , which is the density of the law of  $\hat{X}(t, x)$  with respect to the Lebesgue measure.

**COROLLARY 5.12.** *Let the notations be as above. Suppose  $(\mathbf{ER})_{x_0}$  for  $\{V_1, \dots, V_n\}$  at  $x_0 \in \mathbb{R}^d$ . Suppose also that there exist smooth, bounded functions  $a_1, \dots, a_n: \mathbb{R}^d \rightarrow \mathbb{R}$  with bounded derivatives of all order which satisfy that  $V_0(x) = \sum_{i=1}^n a_i(x) V_i(x)$  for every  $x \in \mathbb{R}^d$ .*

*Then, there exists a decreasing sequence  $\{O_j\}_{j \geq 0}$  of neighborhoods of  $x_0$  such that the asymptotic expansion*

$$p_t(x, x) \sim t^{-\nu/2} (c_0(x) + c_1(x)t + c_2(x)t^2 + \dots) \quad \text{as } t \searrow 0$$

*holds for every  $x \in O_0$  with the following properties: (i)  $\inf_{x \in O_0} c_0(x) > 0$ , (ii) for every  $j \geq 0$ ,*

$$\sup_{x \in O_j} \left\{ |c_j(x)| + \sup_{0 < t \leq 1} t^{(\nu/2) - j - 1} \left| p_t(x, x) - t^{-\nu/2} (c_0(x) + \dots + c_j(x)t^j) \right| \right\} < \infty.$$

*Moreover,  $c_0(x) = \mathbb{E}[\Theta_0(x)]$  and hence is independent of  $\{a_1, \dots, a_n\}$ .*

**PROOF.** We prove the corollary by combining the driftless case (Theorem 5.4) and Girsanov's theorem. Set

$$M_t^{\varepsilon,x} = \exp \left( \varepsilon \sum_{i=1}^n \int_0^t a_i(X^\varepsilon(s,x)) dw_s^i - \frac{\varepsilon^2}{2} \sum_{i=1}^n \int_0^t |a_i(X^\varepsilon(s,x))|^2 ds \right).$$

Since  $a_i$  ( $1 \leq i \leq n$ ) are bounded,  $t \mapsto M_t^{\varepsilon,x}$  is a true martingale. By the scaling property of Brownian motion and Girsanov’s theorem, we have

$$p_{\varepsilon^2}(x,x) = \mathbb{E}[\delta_x(\hat{X}(\varepsilon^2,x))] = \mathbb{E}[\delta_x(\hat{X}^\varepsilon(1,x))] = \mathbb{E}[M_1^{\varepsilon,x} \delta_x(X^\varepsilon(1,x))].$$

Since we have already seen in Theorem 5.4 that  $\delta_x(X^\varepsilon(1,x))$  admits an asymptotic expansion in  $\tilde{\mathbb{D}}_{-\infty}$ , it is sufficient to show that  $M_1^{\varepsilon,x}$  admits an asymptotic expansion in  $\mathbb{D}_\infty$  uniformly in  $x \in O_0$ .

By Proposition 5.1,  $X^\varepsilon(s,x)$  admits an asymptotic expansion in  $\mathbb{D}_\infty$  uniformly in  $s \in [0,1]$  and  $x \in O_0$ . Moreover, each term in the expansion is measurable with respect to  $\sigma(w_u \mid 0 \leq u \leq s)$ . Therefore,  $\sum_{i=1}^n \int_0^1 a_i(X^\varepsilon(s,x)) dw_s^i$  admits an asymptotic expansion in  $\mathbb{D}_\infty$  uniformly in  $x \in O_0$  and so does  $\sum_{i=1}^n \int_0^t |a_i(X^\varepsilon(s,x))|^2 ds$ . Since  $a_i$  ( $1 \leq i \leq n$ ) and their derivatives are all bounded, we can easily see that

$$M_1^{\varepsilon,x} \sim 1 + \varepsilon \Xi_1(x) + \varepsilon^2 \Xi_2(x) + \cdots \quad \text{in } \mathbb{D}_\infty \text{ as } \varepsilon \searrow 0 \text{ uniformly in } x \in O_0. \quad (5.21)$$

Moreover,  $\Xi_{2j-1}(x; \cdot)$  is odd as a Wiener functional for every  $j \geq 1$  and  $x \in O_0$ , that is,  $\Xi_{2j-1}(x; -w) = -\Xi_{2j-1}(x; w)$ .

By multiplying (5.9) and (5.21) and taking the generalized expectation, we have the desired expansion of  $p_{\varepsilon^2}(x,x)$ . Note that the odd-numbered terms in the expansion of  $M_1^{\varepsilon,x} \delta_x(X^\varepsilon(1,x))$  are also odd as generalized Wiener functionals and hence their generalized expectations vanish. Note also that since the leading term on the right-hand side of (5.21) is 1,  $c_0(x)$  does not depend on  $a_i$  ( $1 \leq i \leq n$ ).

It is a routine to check that  $x \mapsto \Xi_j(x)$  is continuous. By Remark 5.5, we can easily check the continuity of  $c_j(x)$  in  $x$ . Positivity of  $c_0(x)$  is immediate from Theorem 5.4.  $\square$

REMARK 5.13. Sections 3–5 basically follow its counterpart in [35]. However, we believe that our argument here is simpler and more readable for the following reasons. (i) Fortunately, it suffices to consider a driftless SDE (5.1) for our purpose. Hence, we need not use the “anisotropic dilation” on the tensor algebra. This simplifies our notations much. (ii) In [35] (originally in Yamato [40]) proofs of important properties of the free nilpotent groups/algebras are done via computations in the coordinates with respect to a linear basis  $\mathcal{G}(N)$ . This could be compared to doing all the differential geometric computation on a manifold via local coordinates and therefore does not provide a very clear view of what is going on. In recent developments of rough path theory and numerical analysis of SDEs, the study of the free nilpotent groups/algebras advanced much (cf. e.g., [16, Chapter 7] for a summary). It provides us a clear view of these objects and helps us simplify our argument. In particular, proofs via the flow of ODEs on the nilpotent Lie groups/algebra in [35] are replaced by (linear or Lie) algebraic proofs. (iii) Some non-trivial facts on Malliavin calculus are presented without proofs in [35]. We added proofs and explanations for non-experts.

## 6. On sub-Riemannian manifolds.

Let  $(M, \mathcal{D}, g)$  be a sub-Riemannian manifold as in Section 1; hence  $d, n, \nu$  and  $N_0$  are all as described there. In this section we prove the uniform asymptotic expansion of the heat kernel on  $M$  via localization method. We emphasize that our argument is almost purely probabilistic. Two key tools to achieve this goal are the stochastic parallel transport for the  $\Delta/2$ -diffusion process and Malliavin calculus for manifold-valued SDEs. The stochastic parallel transport, or the Eells–Elworthy–Malliavin method of constructing diffusion processes on a general sub-Riemannian manifolds was done in [18], [39]. Methods in these papers are slightly different and the latter is used in this paper. Malliavin calculus for manifold-valued SDEs was done in [37]. It was shown there that a solution to an SDE at a fixed time is non-degenerate in the sense of Malliavin under the *partial* Hörmander condition on the coefficient vector fields of the SDE.

We shall define a “div-grad type” sub-Laplacian. The horizontal gradient of  $f \in C^\infty(M)$  is defined to be the unique section  $\nabla^{\mathcal{D}} f \in C^\infty(M; \mathcal{D})$  such that

$$g(\nabla^{\mathcal{D}} f, A) = Af, \quad A \in C^\infty(M; \mathcal{D}).$$

Let  $A_1, \dots, A_k$  be a local orthonormal frame for  $\mathcal{D}$ , i.e., a family of local sections  $A_1, \dots, A_n \in C^\infty(U; \mathcal{D})$  over an open set  $U \subset M$  with  $g_x((A_i)_x, (A_j)_x) = \delta_{ij}$  for  $x \in U$  and  $1 \leq i, j \leq n$ . Then

$$\nabla^{\mathcal{D}} f = \sum_{i=1}^n (A_i f) A_i. \quad (6.1)$$

Take a smooth measure  $\mu$  on  $M$ . For  $A \in C^\infty(M; TM)$ , define its  $\mu$ -divergence  $\operatorname{div}_\mu A$  by

$$\int_M f(\operatorname{div}_\mu A) d\mu = - \int_M Af d\mu, \quad f \in C_0^\infty(M). \quad (6.2)$$

The sub-Laplacian associated with a positive volume form  $\mu$  is the second order differential operator given by

$$\Delta f = \operatorname{div}_\mu(\nabla^{\mathcal{D}} f), \quad f \in C^\infty(M).$$

In terms of a local orthonormal frame  $A_1, \dots, A_n$  for  $\mathcal{D}$ ,

$$\Delta = \sum_{i=1}^n \left( A_i^2 + \operatorname{div}_\mu(A_i) A_i \right). \quad (6.3)$$

In what follows, for the sake of simplicity, we assume that  $M$  is compact.

The goal of this section is to show

**THEOREM 6.1.** *Let  $(M, \mathcal{D}, g)$  and  $\mu$  be as above. Then, the following hold.*

(i) *There exists a diffusion process generated by  $\Delta/2$  and it possesses a transition density function  $p_t(x, y)$ , which is smooth in  $(t, x, y) \in (0, \infty) \times M \times M$ , with respect to  $\mu$ .*

(ii) Suppose  $M$  is equiregular. Then, the asymptotic expansion

$$p_t(x, x) \sim t^{-\nu/2} (c_0(x) + c_1(x)t + c_2(x)t^2 + \dots) \quad \text{as } t \searrow 0$$

holds for every  $x \in M$  with the following properties: (a)  $\inf_{x \in M} c_0(x) > 0$ , (b) for every  $j \geq 0$ ,

$$\sup_{x \in M} \left\{ |c_j(x)| + \sup_{0 < t \leq 1} t^{(\nu/2)-j-1} \left| p_t(x, x) - t^{-\nu/2} (c_0(x) + \dots + t^j c_j(x)) \right| \right\} < \infty.$$

We shall show the theorem by constructing the diffusion process via the Eells–Elworthy–Malliavin method modified for sub-Riemannian manifolds, and then applying Corollary 5.12. It should be noted that the method gives us strong solutions to stochastic differential equations, which enable us to treat systematically the assertions in the theorem together. In fact, to construct diffusion process, a weak solution is enough; since  $\Delta/2$  is smooth, the associated martingale problem is well-posed, and hence the diffusion process exists. In this case, by (6.3) and Hörmander’s theorem, one can prove the assertion (i) in the theorem, but proving the short time asymptotics is another matter.

Let

$$O(\mathcal{D})_x = \{u : \mathbb{R}^n \rightarrow \mathcal{D}_x \mid u \text{ is a bijective linear isometry}\} \quad \text{and} \quad O(\mathcal{D}) = \bigsqcup_{x \in M} O(\mathcal{D})_x$$

and define  $\pi : O(\mathcal{D}) \rightarrow M$  by  $\pi(u) = x$  for  $u \in O(\mathcal{D})_x, x \in M$ . Then,  $\pi : O(\mathcal{D}) \rightarrow M$  is an  $O(n)$ -principal bundle, where  $O(n)$  is the space of  $n \times n$  orthogonal matrices. To apply the Eells–Elworthy–Malliavin method to a sub-Riemannian manifold, we first recall the horizontal vector fields on  $O(\mathcal{D})$  (cf. [39]). To do this, let  $\nabla$  be a partial metric connection on  $(M, \mathcal{D}, g)$ ; that is,  $\nabla$  is a bilinear mapping

$$\nabla : C^\infty(M; \mathcal{D}) \times C^\infty(M; \mathcal{D}) \ni (A, B) \mapsto \nabla_A B \in C^\infty(M; \mathcal{D})$$

such that  $\nabla_A(fB) = f\nabla_A B + (Af)B$  for  $f \in C^\infty(M)$  and  $\nabla g = 0$ , where  $(\nabla_A g)(B, C) := \nabla_A g(B, C) - g(\nabla_A B, C) - g(B, \nabla_A C)$  for  $A, B, C \in C^\infty(M; \mathcal{D})$ . A typical example of partial metric connections is a restriction of Levi-Civita connection. In fact, let  $\tilde{g}$  be a Riemannian metric tensor on  $M$  and  $\tilde{\nabla}$  be its Levi-Civita connection. If  $\tilde{g}$  tames  $g$ , i.e.,  $\tilde{g}|_{\mathcal{D} \times \mathcal{D}} = g$ , then  $\nabla_A B = \text{pr}_{\mathcal{D}} \tilde{\nabla}_A B$ ,  $\text{pr}_{\mathcal{D}}$  being the projection onto  $\mathcal{D}$ , is a partial metric connection.

In terms of a local orthonormal frame  $A_1, \dots, A_n$  for  $\mathcal{D}$ , define  $\omega_i^j \in C^\infty(M; \mathcal{D}^*)$ , where  $\mathcal{D}^*$  is the dual subbundle of  $\mathcal{D}$ , by

$$\nabla A_i = \sum_{j=1}^n \omega_i^j A_j, \quad \text{i.e.,} \quad \nabla_B A_i = \sum_{j=1}^n \omega_i^j(B) A_j \quad \text{for } i = 1, \dots, n \text{ and } B \in C^\infty(M; \mathcal{D}).$$

Since  $\nabla g = 0$ ,  $\omega_i^i = -\omega_i^i, 1 \leq i, j \leq n$ . We now extend the partial connection form  $\omega = (\omega_i^j)$  to a smooth partial 1-form on  $O(\mathcal{D})$  with values in the Lie algebra  $\mathfrak{o}(n)$  of  $O(n)$ , say  $\omega$  again, given by

$$s^{-1}\omega s + s^{-1}ds,$$

where we have used the local trivialization  $M \times O(n)$  of  $O(\mathcal{D})$  and then  $s$  is the coordinate of  $O(n)$ ; more precisely,  $s^{-1}ds$  is the Maurer–Cartan form  $\theta$  given by  $\theta(X) = s^{-1}X$  for  $X \in T_s O(n)$  and  $s \in O(n)$ .

Define the horizontal subspace  $K_u \subset T_u O(\mathcal{D})$ ,  $u \in O(\mathcal{D})$ , by

$$K_u = \{A \in T_u O(\mathcal{D}) \mid (\pi_*)_u A \in \mathcal{D}_{\pi(u)}, \omega_u(A_u) = 0\}.$$

In terms of a local orthonormal frame  $A_1, \dots, A_n$  for  $\mathcal{D}$ , it holds

$$K_u = \left\{ \sum_{\alpha=1}^n a^\alpha A_\alpha - \sum_{p,q,r,s=1}^n \omega_{qr}^p e_s^q a^r \frac{\partial}{\partial e_s^p} \mid (a^1, \dots, a^n) \in \mathbb{R}^n \right\},$$

where

$$\omega_{qr}^p = \omega_q^p(A_r) = g(\nabla_{A_r} A_q, A_p), \quad 1 \leq p, q, r \leq n,$$

and  $(e_q^p)_{1 \leq p, q \leq n}$  stands for the matrix coordinate of  $O(n)$ . Then the horizontal lift  $\ell_u : \mathcal{D}_{\pi(u)} \rightarrow K_u$  defined by

$$\ell_u \left( \sum_{i=1}^n a^i A_i \right) = \sum_{i=1}^n a^i A_i - \sum_{p,q,r,s=1}^n \omega_{qr}^p e_s^q a^r \frac{\partial}{\partial e_s^p}$$

is bijective.

Let  $\{\mathbf{e}_i \mid 1 \leq i \leq n\}$  be the canonical basis of  $\mathbb{R}^n$ . Define the canonical horizontal vector fields  $V_1, \dots, V_n$  on  $O(\mathcal{D})$  by

$$(L_i)_u = \ell_u(u\mathbf{e}_i), \quad 1 \leq i \leq n.$$

In terms of an orthonormal frame  $A_1, \dots, A_n$  for  $\mathcal{D}$ , it holds

$$L_i = \sum_{j=1}^n e_i^j A_j - \sum_{p,q,r,s=1}^n \omega_{qr}^p e_s^q e_i^r \frac{\partial}{\partial e_s^p}, \quad 1 \leq i \leq n. \quad (6.4)$$

The following lemma asserts that the operator  $(1/2) \sum_{i=1}^n L_i^2$  corresponds to another sub-Laplacian  $\Delta'$  given by

$$\Delta' f = \text{tr} \nabla df, \quad f \in C^\infty(M), \quad (6.5)$$

where the Hessian  $\nabla df$  is given by

$$(\nabla df)(A, B) = ABf - (\nabla_A B)f, \quad A, B \in C^\infty(M; \mathcal{D}), \quad (6.6)$$

and  $\text{tr} \nabla df$  is the trace at each point in  $M$  of the bilinear form  $(A, B) \mapsto (\nabla df)(A, B)$ .

LEMMA 6.2. For  $f \in C^\infty(M)$ , set  $\tilde{f} = f \circ \pi$ . Then

$$L_i L_j \tilde{f}(u) = (\nabla df)_{\pi(u)}(u\mathbf{e}_i, u\mathbf{e}_j), \quad u \in O(\mathcal{D}), \quad 1 \leq i, j \leq n.$$

In particular,

$$\frac{1}{2} \sum_{i=1}^n L_i^2 \tilde{f} = \frac{1}{2} (\Delta' f) \circ \pi.$$

PROOF. By (6.4), we have

$$L_j \tilde{f} = \sum_{p=1}^n e_j^p \widetilde{A_p f}.$$

Hence

$$L_i L_j \tilde{f} = \sum_{p,q=1}^n e_i^q e_j^p \widetilde{A_q A_p f} - \sum_{p,q,r=1}^n \omega_{qr}^p e_j^q e_i^r \widetilde{A_p f}.$$

Since

$$\nabla_{u\mathbf{e}_i}(u\mathbf{e}_j) = \nabla_{\sum_{p=1}^n e_i^p A_p} \left( \sum_{q=1}^n e_j^q A_q \right) = \sum_{p,q,r=1}^n e_i^r e_j^q \omega_{qr}^p A_p,$$

we obtain the first identity by (6.6). The second identity immediately follows from the first one. □

We are now ready to prove Theorem 6.1.

PROOF OF THEOREM 6.1. (i) Extend the metric  $g$  to a Riemannian metric  $\tilde{g}$  on  $M$ . Let  $\tilde{\nabla}$  be the Levi-Civita connection associated with  $\tilde{g}$ , and define the partial metric connection  $\nabla$  by  $\nabla_{AB} = \text{pr}_{\mathcal{D}} \tilde{\nabla}_{AB}$ . Denote by  $\Delta'$  the sub-Laplacian on  $M$  given by (6.5).

In a local orthonormal frame  $A_1, \dots, A_n$  for  $\mathcal{D}$ , it holds

$$\Delta' = \sum_{i=1}^n A_i^2 - \sum_{i=1}^n \left( \sum_{j=1}^n \omega_j^i(A_j) \right) A_i.$$

Hence  $N = (\Delta - \Delta')/2$  satisfies

$$N = \frac{1}{2} \sum_{i=1}^n \left\{ \text{div}_{\mu} A_i + \sum_{j=1}^n \omega_j^i(A_j) \right\} A_i.$$

In particular,  $N \in C^\infty(M; \mathcal{D})$ .

Set

$$L_0(u) = \ell_u(N_{\pi(u)}), \quad u \in O(\mathcal{D}).$$

Let  $(r(t))_{t \geq 0}$  be the unique solution to the stochastic differential equation on  $O(\mathcal{D})$

$$dr(t) = \sum_{i=1}^n L_i(r(t)) \circ dw^i(t) + L_0(r(t))dt, \quad r(0) = u \in O(\mathcal{D}),$$

and put  $X(t) = \pi(r(t))$ ,  $t \geq 0$ . Since  $\pi_* L_0 = N$ , by Lemma 6.2, the projected process  $(X(t))_{t \geq 0}$  solves the  $\Delta/2$ -martingale problem, i.e.,

$$\left( f(X(t)) - \int_0^t \frac{1}{2} \Delta f(X(s)) ds \right)_{t \geq 0}$$

is a martingale for any  $f \in C^\infty(M)$ . Thus the  $\Delta/2$ -diffusion process is realized as the projected process  $(X(t))_{t \geq 0}$ . In particular, the law of  $(X(t))_{t \geq 0}$  is independent of the choice of  $u \in \pi^{-1}(x)$ .

By the Hörmander condition at every  $x \in M$ ,  $\{L_1, \dots, L_n\}$  satisfies the partial Hörmander condition at every  $u \in O(\mathcal{D})$ , that is, the linear span of  $\{(\pi_*)_u L_{[I]}(u) \mid I \in \mathcal{I}(\infty)\}$  is equal to  $T_{\pi(u)}M$ . Then, we know from [37] the non-degeneracy of the Malliavin covariance of  $X(t)$  for all  $t > 0$  and  $r(0) = u \in O(\mathcal{D})$ . By the integration by parts formula for manifold-valued Wiener functionals, we obtain the existence of  $p_t(x, y)$  and the smoothness in  $y \in M$  (cf. [24]). The smoothness in  $(t, x) \in (0, \infty) \times M$  is obtained as an application of Itô's formula and the stochastic flow property of  $(r(t))_{t \geq 0}$ . By the way, the heat kernel admits the following explicit expression as in the Euclidean case:

$$p_t(x, y) = \mathbb{E}[\delta_y(X(t))],$$

where  $\delta_y$  is the delta function at  $y$  with respect to  $\mu$ .

(ii) Since  $M$  is compact, it suffices to show that for each  $x_0 \in M$ , there exists a decreasing sequence  $\{O_j\}_{j \geq 0}$  of neighborhoods of  $x_0$  such that the asymptotic expansion

$$p_t(x, x) \sim t^{-\nu/2} (c_0(x) + c_1(x)t + c_2(x)t^2 + \dots) \quad \text{as } t \searrow 0$$

holds for every  $x \in O_0$  with the following properties: (a)  $\inf_{x \in O_0} c_0(x) > 0$ , (b) for every  $j \geq 0$ ,

$$\sup_{x \in O_j} \left\{ |c_j(x)| + \sup_{0 < t \leq 1} t^{(\nu/2) - j - 1} \left| p_t(x, x) - t^{-\nu/2} (c_0(x) + \dots + t^j c_j(x)) \right| \right\} < \infty.$$

To do this, let  $U_1$  and  $U_2$  be open sets in  $M$  such that  $x_0 \in U_1$ ,  $\overline{U_1} \subset U_2$  and there exists a local orthonormal frame  $A_1, \dots, A_n$  for  $\mathcal{D}$  over  $U_2$ . Viewing  $U_2$  as a part of  $\mathbb{R}^d$ , we extend  $A_1, \dots, A_n$  on  $U_2$  to  $C_b^\infty$ -vector fields  $V_1, \dots, V_n$  on  $\mathbb{R}^d$ , respectively, and extend each  $(\operatorname{div}_\mu A_i)/2$  on  $U_2$  to  $a_i \in C_b^\infty(\mathbb{R}^d)$ . Let  $\tilde{p}_t(x, y)$  be the heat kernel with respect to the Lebesgue measure on  $\mathbb{R}^d$  associated with

$$\frac{1}{2} \sum_{i=1}^n V_i^2 + \sum_{i=1}^n a_i V_i.$$

Denote by  $(\tilde{X}(t, x))_{t \geq 0}$  the solution to the SDE

$$d\tilde{X}(t) = \sum_{i=1}^n V_i(\tilde{X}(t)) \circ dw^i + V_0(\tilde{X}(t))dt, \quad \tilde{X}(0) = x,$$

where  $V_0 = \sum_{i=1}^n a_i V_i$ . Then  $\tilde{p}_t(x, y)$  is the transition density function of  $(\tilde{X}(t, x))_{t \geq 0}$  with respect to the Lebesgue measure.

In repetition of the argument employed to show the estimation (10.57) in [23, p.421], we obtain positive constants  $c_1$  and  $c_2$  such that

$$\sup_{x, y \in U_1} |\rho(y)p_t(x, y) - \tilde{p}_t(x, y)| \leq c_1 e^{-c_2/t} \quad \text{as } t \searrow 0, \tag{6.7}$$

where  $d\mu(y) = \rho(y)dy^1 \cdots dy^d$ ,  $(y^1, \dots, y^d)$  is the local coordinates on  $U_1$  identified with that on  $\mathbb{R}^d$ .

Indeed, we can show (6.7) by combining the following two observations: (i)  $u_f(t, x) := \int_{U_1} \{\rho(y)p_t(x, y) - \tilde{p}_t(x, y)\}f(y)dy$ , where  $f \in C^\infty(\mathbb{R}^d)$  whose support is contained in  $U_1$ , satisfies the estimation  $|u_f(t, x)| \leq \sup_{s \in [0, t], z \in \partial U_2} |u_f(s, z)|$ ,  $\partial U_2$  being the boundary of  $U_2$ . (ii) There exist positive constants  $c_3$  and  $c_4$  such that

$$p_s(z, y) \vee \tilde{p}_s(z, y) \leq c_3 e^{-c_4/s}, \quad y \in U_1, z \in \partial U_2, s \in (0, 1]. \tag{6.8}$$

A rough sketch of proof of (6.8) is as follows. The non-degeneracy of the Malliavin covariances of  $X(s)$  and  $\tilde{X}(s)$  under the (partial) Hörmander condition enables us to use the integration by parts formula. So we can replace the delta functions in the Feynman–Kac type representation formulae for the heat kernels by continuous functions as in the proof of Lemma 5.10. Then, the exponential decay of exit times of semimartingales like (3.5) and Kusuoka–Stroock’s estimate like Proposition 5.3 for both  $X$  and  $\tilde{X}$  imply (6.8).

It immediately follows from the two observations that  $|u_f(t, x)| \leq c_3(1 + \|\rho\|_\infty)\|f\|_{L^1} e^{-c_4/t}$  for every  $t \in (0, 1]$  and  $x \in U_1$ . Letting  $f$  tend to  $\pm\delta_y$  ( $y \in U_1$ ), we prove (6.7).

One should note that the equiregular condition has not been used so far. Once (6.7) is obtained, the desired asymptotic expansion of  $p_t(x, x)$  follows from that of  $\tilde{p}_t(x, x)$ . Thus, the assertion (ii) follows by applying Corollary 5.12 to  $\tilde{p}_t(x, x)$ .  $\square$

### 7. Leading constant: Examples.

From the viewpoint of spectral geometry, it is very important to obtain an explicit expression of the leading constant of the asymptotic expansion of the heat trace. However, it seems quite difficult in general. Therefore, in this section we provide some examples for which the leading constant is explicitly computable by our method and we check that these leading constants coincide with known results.

Since we have already shown in Theorem 6.1 that the asymptotic expansion of the heat kernel is uniform in the space parameter  $x$ , we may compute the leading term of the asymptotics of  $p_t(x, x)$  in the most convenient way for each fixed  $x \in M$ .

We recall symbols and notations which will be used in subsequent examples. The dimension of the manifold  $M$  is  $d$  and the number of independent linear Brownian motion is  $n$ . For a given set of vector fields  $\{V_i \mid 1 \leq i \leq n\}$ , which are actually the coefficients of

the corresponding SDE,  $N_0$  stands for the step of the equiregular Hörmander condition. Matrices  $B_{\mathcal{H}}(x)$  and  $B_{N_0}(x)$  are defined in (4.4) and (4.6), respectively. Recall also that  $\Gamma_{N_0}(x) = (\gamma_J^I(x))_{I \in \mathcal{H}, J \in \mathcal{G}(N_0)}$  and  $\tilde{\Gamma}_{N_0}^0(x) = (\delta_{|J|}^{|I|} \gamma_J^I(x))_{I \in \mathcal{H}, J \in \mathcal{G}(N_0)}$  which is defined in (4.14). The leading constant of  $p_t(x, x)$  in the Euclidian case was shown in (5.19) to be

$$|\det B_{\mathcal{H}}(x)|^{-1} \mathbb{E}[\delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0})].$$

Here,  $(U_t^{N_0})_{t \geq 0}$  is the  $\mathfrak{g}^{N_0}(\mathbb{R}^n)$ -valued hypoelliptic diffusion process introduced in (3.3). In the manifold case, this constant should be adjusted by being divided by the density function as we discussed in (6.7).

**EXAMPLE 7.1** (The case of Riemannian manifold). Let  $M$  be a compact Riemannian manifold of dimension  $d$  with the Riemannian measure  $\mu$ . The div-grad type operator is the usual Laplace–Beltrami operator  $\Delta_M$ . In this case  $N_0 = 1$ ,  $d = n$ ,  $\mathcal{G}(1) = \mathcal{H} = \{(i) \mid 1 \leq i \leq d\}$ .

Take a coordinate chart  $(x^1, \dots, x^d)$ . We denote the metric tensor by  $G(x) := (g_{ij}(x))_{1 \leq i, j \leq d}$ . Then,  $\mu(dx) = \sqrt{\det G(x)} dx^1 \cdots dx^d$  on this chart. We write  $G(x)^{-1/2} = (\sigma^{ij}(x))_{1 \leq i, j \leq d}$  and set  $V_i(x) = \sum_j \sigma^{ij}(x)(\partial/\partial x^j)$  so that  $\{V_i \mid 1 \leq i \leq d\}$  is a local orthonormal frame. The Laplace–Beltrami operator is expressed as  $\Delta_M = \sum_{i=1}^d V_i^2 + (\text{a vector field})$ .

It is easy to see that  $B_{\mathcal{H}}(x) = B_{N_0}(x) = G(x)^{-1/2}$ ,  $\Gamma_{N_0}(x) = \tilde{\Gamma}_{N_0}^0(x) = \text{Id}$ . Hence, by adjusting the density function of  $\mu$  as in (6.7), we see that the leading constant in the asymptotics of  $p_t(x, x)$  equals

$$\frac{1}{\sqrt{\det G(x)}} |\det B_{\mathcal{H}}(x)|^{-1} \mathbb{E}[\delta_0(\tilde{\Gamma}_{N_0}^0(x)U_1^{N_0})] = \mathbb{E}[\delta_0(w_1^1, \dots, w_1^d)] = (2\pi)^{-d/2}.$$

In particular, we see that  $\text{Trace}(e^{-t\Delta_M/2}) \sim (2\pi t)^{-d/2} \mu(M)$  as  $t \searrow 0$ . Thus, we have recovered the well-known result in Riemannian geometry.

**EXAMPLE 7.2** (The case of 3D contact sub-Riemannian manifold). In this example, we calculate the leading constant for a three-dimensional contact sub-Riemannian manifold and check that it coincides with Barilari’s result in [2].

Let  $M$  be a compact sub-Riemannian manifold with  $\dim M = 3$  with a distribution  $\mathcal{H}$  of rank 2. We assume that  $\mathcal{K}$  is contact, namely, there exists a one-form  $\omega$  such that  $\omega \wedge d\omega$  vanishes nowhere. As a volume on  $M$ , we choose the following measure. Let  $\{V_1, V_2\}$  be a local orthonormal frame of  $\mathcal{K}$  on a coordinate chart and regard  $\lambda_1 \wedge \lambda_2 \wedge \lambda_3$  as a measure on the chart, where  $\{\lambda_1, \lambda_2, \lambda_3\}$  is the dual basis of  $\{V_1, V_2, [V_1, V_2]\}$ . This defines a measure  $\mu$  on  $M$ . Note that  $\mu$  is a constant multiple of Popp’s measure if we use the definition (or results) in [3]. In this case  $N_0 = 2$ ,  $d = 3$ ,  $n = 2$ ,  $\mathcal{G}(2) = \mathcal{H} = \{(1), (2), (2, 1)\}$  and the Hausdorff dimension is  $\nu = 4$ .

We use the normal coordinates for three-dimensional contact manifolds in the same way as in [2]. For every  $x \in M$ , we can find a local coordinate chart  $(u^1, u^2, u^3)$  and a local orthonormal frame  $\{V_1, V_2\}$  of  $\mathcal{K}$  on this chart such that  $x$  corresponds to  $0 \in \mathbb{R}^3$  and

$$\begin{aligned}
 V_1(u^1, u^2, u^3) &= \left( \frac{\partial}{\partial u^1} + \frac{u^2}{2} \frac{\partial}{\partial u^3} \right) + \beta u^2 \left( u^2 \frac{\partial}{\partial u^1} - u^1 \frac{\partial}{\partial u^2} \right) + \gamma u^2 \frac{\partial}{\partial u^3}, \\
 V_2(u^1, u^2, u^3) &= \left( \frac{\partial}{\partial u^2} - \frac{u^1}{2} \frac{\partial}{\partial u^3} \right) - \beta u^1 \left( u^2 \frac{\partial}{\partial u^1} - u^1 \frac{\partial}{\partial u^2} \right) + \gamma u^1 \frac{\partial}{\partial u^3},
 \end{aligned}$$

where  $\beta = \beta(u^1, u^2, u^3)$  and  $\gamma = \gamma(u^1, u^2, u^3)$  are certain smooth functions which vanish at 0. The sub-Laplacian can be written locally as  $\Delta = V_1^2 + V_2^2 + (\text{a section of } \mathcal{K})$ .

From these explicit expressions, we can easily see that the density  $\rho := d\mu/du^1 du^2 du^3$  satisfies  $\rho(0) = 1$ . Moreover,  $B_{\mathcal{H}}(0) = B_{N_0}(0) = \text{Id}$  and  $\Gamma_{N_0}(0) = \tilde{\Gamma}_{N_0}^0(0) = \text{Id}$ . Hence, the leading constant in the asymptotics of  $p_t(x_0, x_0)$  associated with  $\Delta/2$  equals

$$\frac{1}{\rho(0)} |\det B_{\mathcal{H}}(0)|^{-1} \mathbb{E}[\delta_0(\tilde{\Gamma}_{N_0}^0(0)U_1^{N_0})] = \mathbb{E}[\delta_{(0,0,0)}(w_1^1, w_1^2, S_1(w^1, w^2))], \tag{7.1}$$

where

$$S_t(w^1, w^2) = \frac{1}{2} \int_0^t (w_s^1 dw_s^2 - w_s^2 dw_s^1) = \frac{1}{2} \int_0^t (w_s^1 \circ dw_s^2 - w_s^2 \circ dw_s^1) \tag{7.2}$$

is Levy’s stochastic area of the two-dimensional Brownian motion. A well-known formula for Levy’s stochastic area (e.g., [27, Theorem 5.8.5, p.272]) states that

$$\mathbb{E}[\exp(\sqrt{-1} \lambda S_1(w^1, w^2)) \delta_{(0,0)}(w_1^1, w_1^2)] = \frac{1}{2\pi} \frac{\lambda/2}{\sinh(\lambda/2)} \quad (\lambda \in \mathbb{R}). \tag{7.3}$$

Then, we see that the right-hand side on (7.1) equals

$$\begin{aligned}
 \mathbb{E}[\delta_{(0,0)}(w_1^1, w_1^2) \delta_0(S_1(w^1, w^2))] &= \mathbb{E} \left[ \delta_{(0,0)}(w_1^1, w_1^2) \frac{1}{2\pi} \int_{\mathbb{R}} \exp(\sqrt{-1} \lambda S_1(w^1, w^2)) d\lambda \right] \\
 &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \frac{\lambda/2}{\sinh(\lambda/2)} d\lambda = \frac{1}{4}.
 \end{aligned}$$

For a proof of the last equality, see [4, Lemma A.2, p.260]. This constant 1/4 coincides with one in Theorem 1, [2]. (We need to replace  $t$  in [2] by  $t/2$  since the heat kernel in [2] is associated with  $\Delta$ , not  $\Delta/2$ .) In particular, we see that  $\text{Trace}(e^{-t\Delta/2}) \sim \mu(M)/4t^2$  as  $t \searrow 0$ .

**EXAMPLE 7.3.** To state the next example, we review strictly pseudoconvex CR manifolds. For details, see [15].

Let  $M$  be a CR-manifold, i.e.,  $M$  is a real smooth manifold with a complex subbundle  $T_{1,0}$  of the complexified tangent bundle  $\mathbb{C}TM$  such that

$$T_{1,0} \cap T_{0,1} = \{0\} \quad \text{and} \quad [T_{1,0}, T_{1,0}] \subset T_{1,0},$$

where  $T_{0,1} = \overline{T_{1,0}}$ . Suppose that the real dimension of  $M$  is  $2k + 1$  and the complex dimension of  $T_{1,0}$  is  $k$  ( $k \geq 1$ ).

There exists a real nowhere vanishing 1-form  $\theta$  which annihilates  $\mathcal{D} := \text{Re}(T_{1,0} \oplus T_{0,1})$ . The associated Levi form  $L_\theta$  is defined by

$$L_\theta(Z, W) = -\sqrt{-1} d\theta(Z, W), \quad Z, W \in C^\infty(M; T_{1,0} \oplus T_{0,1}).$$

We assume that  $M$  is strictly pseudoconvex, i.e.,  $L_\theta$  is positive definite.

There exists a unique real vector field  $T$ , called the characteristic direction, such that

$$\theta(T) = 1, \quad T \lrcorner d\theta = 0,$$

where  $T \lrcorner$  stands for the interior product by  $T$ . The Webster metric  $g_\theta$  on  $TM = \mathcal{D} \oplus \mathbb{R}T$  is defined by

$$g_\theta(X, Y) = d\theta(X, JY), \quad g_\theta(X, T) = 0, \quad g_\theta(T, T) = 1 \quad \text{for } X, Y \in C^\infty(M; \mathcal{D}),$$

where  $J$  is a linear mapping on  $T_{1,0} \oplus T_{0,1}$  such that  $J|_{T_{1,0}} = \sqrt{-1}$  and  $J|_{T_{0,1}} = -\sqrt{-1}$ .

In this example, let  $\mu$  be the Riemannian volume measure associated with  $g_\theta$  and consider  $\Delta$  associated with this  $\mu$ . It should be noted that  $\mu$  is a constant multiple of Popp's measure (cf. [3]). Moreover,  $\Delta$  is the standard sub-Laplacian on a CR-manifold, and coincides with  $\Delta'$ , which is constructed by using the Tanaka–Webster connection on  $M$  ([15, Section 2.1]).

To compute locally around fixed  $x \in M$ , we introduce the Folland–Stein normal coordinates, following [15, Section 3.2]. Let  $T_1, \dots, T_n$  be a local orthonormal frame on an open set  $U \subset M$  for  $T_{1,0}$  with respect to  $L_\theta$ , i.e., (i)  $T_\alpha \in C^\infty(U; T_{1,0})$  and (ii)  $L_\theta(T_\alpha, T_{\bar{\beta}}) = \delta_{\alpha\beta}$  for  $1 \leq \alpha, \beta \leq k$ , where  $T_{\bar{\beta}} = \overline{T_\beta}$ . Set  $X_\alpha = T_\alpha + T_{\bar{\alpha}}$  and  $Y_\alpha = \sqrt{-1}(T_{\bar{\alpha}} - T_\alpha)$ . Then

$$g_\theta(X_\alpha, X_\beta) = g_\theta(Y_\alpha, Y_\beta) = 2\delta_{\alpha\beta}, \quad g_\theta(X_\alpha, Y_\beta) = 0 \quad \text{for } 1 \leq \alpha, \beta \leq k.$$

There exists a coordinate chart  $u = (u^1, \dots, u^{2k+1})$ , called the Folland–Stein normal coordinates, such that  $x$  corresponds to  $0 \in \mathbb{R}^{2k+1}$ , and

$$\begin{cases} X_\alpha = \frac{\partial}{\partial u^{2\alpha-1}} + 2u^{2\alpha} \frac{\partial}{\partial u^{2k+1}} + \sum_{i=1}^{2k} O^1 \frac{\partial}{\partial u^i} + O^2 \frac{\partial}{\partial u^{2k+1}}, \\ Y_\alpha = \frac{\partial}{\partial u^{2\alpha}} - 2u^{2\alpha-1} \frac{\partial}{\partial u^{2k+1}} + \sum_{i=1}^{2k} O^1 \frac{\partial}{\partial u^i} + O^2 \frac{\partial}{\partial u^{2k+1}}, \\ T = \frac{\partial}{\partial u^{2k+1}} + \sum_{i=1}^{2k} O^1 \frac{\partial}{\partial u^i} + O^2 \frac{\partial}{\partial u^{2k+1}}, \end{cases} \quad (7.4)$$

where  $O^j$ ,  $j = 1, 2$ , stand for functions with the property that

$$O^j = O \left( \left( \sum_{i=1}^{2k} |u^i| \right)^j + |u^{2k+1}|^{j/2} \right).$$

By (7.4),

$$g_\theta \left( \left( \frac{\partial}{\partial u^i} \right)_u, \left( \frac{\partial}{\partial u^j} \right)_u \right) = 2\delta_{ij} + O^1, \quad 1 \leq i, j \leq 2k,$$

$$g_\theta \left( \left( \frac{\partial}{\partial u^p} \right)_u, \left( \frac{\partial}{\partial u^{2k+1}} \right)_u \right) = \delta_{p,2k+1} + O^1, \quad 1 \leq p \leq 2k + 1.$$

Thus  $\mu(du) = (2^k + O^1)du^1 \cdots du^{2k+1}$ . In particular, the density  $\rho = d\mu/du^1 \cdots du^{2k+1}$  satisfies  $\rho(0) = 2^k$ .

Let

$$V_{2\alpha-1} = \frac{1}{\sqrt{2}}X_\alpha, \quad V_{2\alpha} = \frac{1}{\sqrt{2}}Y_\alpha, \quad 1 \leq \alpha \leq k,$$

where, as in the proof of Theorem 6.1, we have extended  $X_\alpha$  and  $Y_\alpha$ ,  $1 \leq \alpha \leq k$ , to  $\mathbb{R}^{2k+1}$ , and used the same letters to indicate the extensions. Then what we need to investigate is the transition density function of the diffusion process generated by

$$\frac{1}{2} \sum_{i=1}^{2k} V_i^2 + \sum_{i=1}^{2k} a_i V_i,$$

where  $a_i = (\operatorname{div}_\mu V_i)/2$ ,  $1 \leq i \leq 2k$ . Moreover, by (7.4), it holds

$$V_i(x) = \frac{1}{\sqrt{2}} \left( \frac{\partial}{\partial u^i} \right)_0, \quad 1 \leq i \leq 2k, \tag{7.5}$$

$$[V_i, V_j](x) = -2 \left( \sum_{p=1}^k \delta_{i,2p-1} \delta_{j,2p} \right) \left( \frac{\partial}{\partial u^{2k+1}} \right)_0 + \sum_{p=1}^{2k} C_{ij}^p \left( \frac{\partial}{\partial u^p} \right)_0, \quad 1 \leq i < j \leq 2k \tag{7.6}$$

for some  $C_{ij}^p \in \mathbb{R}$ . Thus we are in the situation that  $d = 2k + 1$ ,  $n = 2k$ ,  $N_0(x) = 2$ , and  $\nu(x) = 2k + 2$ .

We now proceed to the computation of  $\tilde{\Gamma}_2^0(0)U_t^2$ . Let

$$\mathcal{G}(1) = \{(i) \mid 1 \leq i \leq 2k\}, \quad \mathcal{G}(2) = \mathcal{G}(1) \cup \{(i, j) \mid 1 \leq i < j \leq 2k\}.$$

Set  $\mathcal{H} = \{(1), \dots, (2k), (1, 2)\}$ . Then, by (7.5) and (7.6),

$$B_{\mathcal{H}}(0) = \left( \begin{array}{ccc|c} & & & C_{12}^1 \\ & & & \vdots \\ & 2^{-1/2}\operatorname{Id}_{2k} & & C_{12}^{2k} \\ \hline 0 & \cdots & 0 & -2 \end{array} \right),$$

where  $\operatorname{Id}_{2k}$  denotes the  $2k$ -dimensional identity matrix. Hence



To see the last identity, we have used (7.3). Hence the leading constant in the asymptotics of  $p_t(x, x)$  associated with  $\Delta/2$  equals

$$\frac{1}{\rho(0)} |\det B_{\mathcal{H}}(0)|^{-1} \mathbb{E}[\delta_0(\tilde{\Gamma}_2^0(0)U_1^2)] = \frac{1}{2} \frac{1}{2\pi} \int_{\mathbb{R}} \left( \frac{1}{2\pi} \frac{\lambda/2}{\sinh(\lambda/2)} \right)^k d\lambda.$$

The right-hand side is the heat kernel  $p_1(0, 0)$  associated with the sub-Laplacian on the Heisenberg group of dimension  $2k + 1$  (cf. [17, Théorème 1]).

Before providing our final example, we fix some notations. Let  $(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g)$  be such that (i)  $\mathfrak{k}_1 \oplus \mathfrak{k}_2$  is a finite-dimensional graded Lie algebra (with  $\mathfrak{k}_j = \{0\}$  for  $j \in \mathbb{Z} \setminus \{1, 2\}$ ) and  $g$  is an inner product on  $\mathfrak{k}_1$ . Two such  $(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g)$  and  $(\hat{\mathfrak{k}}_1 \oplus \hat{\mathfrak{k}}_2, \hat{g})$  are said to be *isometrically isomorphic* and denoted by  $(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g) \cong (\hat{\mathfrak{k}}_1 \oplus \hat{\mathfrak{k}}_2, \hat{g})$  if there exists an isomorphism  $\phi: \mathfrak{k}_1 \oplus \mathfrak{k}_2 \rightarrow \hat{\mathfrak{k}}_1 \oplus \hat{\mathfrak{k}}_2$  of graded Lie algebras whose restriction to  $\mathfrak{k}_1$  preserves the inner product. An isometrical isomorphism class in this sense is denoted by  $[(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g)]$ .

Let  $(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g)$  be as above and write  $n = \dim \mathfrak{k}_1$  and  $p = \dim \mathfrak{k}_2$ . An adapted basis of this Lie algebra is defined to be a linear basis  $\{v_1, \dots, v_n; z_1, \dots, z_p\}$  such that  $\{v_1, \dots, v_n\}$  is an orthonormal basis of  $(\mathfrak{k}_1, g)$  and  $\{z_1, \dots, z_p\}$  be a linear basis of  $\mathfrak{k}_2$ . Write

$$[v_i, v_j] = \sum_{k=1}^p C_{ij}^k z_k \quad (1 \leq i, j \leq n).$$

We call  $\{C_{ij}^k\}$  the structure constants with respect to this adapted basis. (Note that there are no other non-trivial Lie brackets.) It is easy to see that  $(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g) \cong (\hat{\mathfrak{k}}_1 \oplus \hat{\mathfrak{k}}_2, \hat{g})$  if and only if we can find an adapted basis of  $(\mathfrak{k}_1 \oplus \mathfrak{k}_2, g)$  and an adapted basis of  $(\hat{\mathfrak{k}}_1 \oplus \hat{\mathfrak{k}}_2, \hat{g})$  whose structure constants exactly coincide.

EXAMPLE 7.4. Let  $(M, \mathcal{D}, g)$  be a step-two equiregular compact sub-Riemannian manifold with  $\dim M = n + p$  and  $\text{rank } \mathcal{D} = n$  ( $n \geq 1, p \geq 1$ ) and let  $\mu$  be Popp’s measure. In this case, the Hausdorff dimension is  $\nu = n + 2p$ . By the equiregularity,  $\mathcal{D}_1(x) \oplus (\mathcal{D}_2(x)/\mathcal{D}_1(x))$  has a natural structure of graded Lie algebra. Clearly,  $N_0 = 2$  and we set

$$\mathcal{G}(1) = \{(i) \mid 1 \leq i \leq n\}, \quad \mathcal{G}(2) = \mathcal{G}(1) \cup \{(i, j) \mid 1 \leq i < j \leq n\}.$$

The aim of this example is to calculate the leading constant  $c_0(x)$  explicitly in a probabilistic way and show that it depends only on  $[(\mathcal{D}_1(x) \oplus (\mathcal{D}_2(x)/\mathcal{D}_1(x)), g_x)]$ . (More precisely, if  $(\hat{M}, \hat{\mathcal{D}}, \hat{g})$  is another such sub-Riemannian manifold and

$$(\mathcal{D}_1(x) \oplus (\mathcal{D}_2(x)/\mathcal{D}_1(x)), g_x) \cong (\hat{\mathcal{D}}_1(\hat{x}) \oplus (\hat{\mathcal{D}}_2(\hat{x})/\hat{\mathcal{D}}_1(\hat{x})), \hat{g}_{\hat{x}}) \tag{7.7}$$

holds for  $x \in M$  and  $\hat{x} \in \hat{M}$ , then  $c_0(x) = c_0(\hat{x})$  holds.)

To this end we use (a very special case of) Bianchini–Stefani’s adapted chart. As was already demonstrated in [19], [31], this chart looks quite useful for short time asymptotic problems on sub-Riemannian manifolds. Take  $x \in M$  arbitrarily. Then, by [11,

Corollary 3.1], there exists a local coordinate chart  $(u^1, \dots, u^{n+p})$  around  $x$  such that  $x$  corresponds to  $0 \in \mathbb{R}^{n+p}$  and  $\mathcal{D}_1(x)$  equals the linear span of  $\{(\partial/\partial u^1)_0, \dots, (\partial/\partial u^n)_0\}$ . Note that this equality holds only at  $x$  and such a chart is obviously not unique.

Take a local frame  $\{V_1, \dots, V_n, Z_1, \dots, Z_p\}$  of  $TM$  around  $x$  such that  $\{V_1, \dots, V_n\}$  forms a local orthonormal frame of  $\mathcal{D} = \mathcal{D}_1$ . Such a local frame is called an adapted frame. As usual the structure constants  $\{C_{ij}^k\}$  is defined by

$$[V_i, V_j](x) = \sum_{k=1}^p C_{ij}^k Z_k(x) \quad \text{mod } \mathcal{D}_1(x) \quad (1 \leq i, j \leq n).$$

The rank of a  $p \times n(n-1)/2$  matrix  $(C_{ij}^k)_{1 \leq k \leq p, (i,j) \in \mathcal{G}(2) \setminus \mathcal{G}(1)}$  is  $p$  due to the Hörmander condition at  $x$ . We will denote this matrix by  $C$  for simplicity.

Changing the coordinates of  $(u^1, \dots, u^n)$  and  $(u^{n+1}, \dots, u^{n+p})$  by linear maps, we may additionally assume that  $V_i(x) = (\partial/\partial u^i)_0$  for  $1 \leq i \leq n$  and  $Z_j(x) = (\partial/\partial u^{j+n})_0$  modulo  $\mathcal{D}_1(x)$  for  $1 \leq j \leq p$ . Then, it is obvious that

$$B_2(0) = \begin{pmatrix} \text{Id}_n & | & * \\ \hline 0_{p \times n} & | & C \end{pmatrix}.$$

Choose  $(i_1, j_1), \dots, (i_p, j_p) \in \mathcal{G}(2) \setminus \mathcal{G}(1)$  so that  $(C_{i_a j_a}^k)_{1 \leq k, a \leq p}$  is an invertible  $p \times p$  matrix and we set  $\mathcal{H} = \mathcal{G}(1) \cup \{(i_a, j_a) \mid 1 \leq a \leq p\}$ . Then, it is easy to see that

$$B_{\mathcal{H}}(0) \tilde{\Gamma}_2^0(0) = \begin{pmatrix} \text{Id}_n & | & 0_{p \times n(n-1)/2} \\ \hline 0_{p \times n} & | & C \end{pmatrix}.$$

According to [3], Popp’s measure can be computed from the structure constants for the local adapted frame as follows. Set  $\mathcal{C}^{kl} = \langle C_{\bullet, \star}^k, C_{\bullet, \star}^l \rangle_{HS}$ ,  $1 \leq k, l \leq p$ , where the right-hand side stands for the Hilbert–Schmidt inner product for  $n \times n$ -matrices. Then, we have  $\mu(d\theta) = \rho(u) du^1 \cdots du^{n+p}$  with  $\rho(0) = \{\det(\mathcal{C}^{kl})_{1 \leq k, l \leq p}\}^{-1/2}$ .

Combining these all, we see that

$$\begin{aligned} c_0(x) &= \frac{1}{\rho(0)} |\det B_{\mathcal{H}}(0)|^{-1} \mathbb{E}[\delta_0(\tilde{\Gamma}_2^0(0)U_1^2)] \\ &= \sqrt{\det(\mathcal{C}^{kl})} \mathbb{E}[\delta_0(B_{\mathcal{H}}(0)\tilde{\Gamma}_2^0(0)U_1^2)] \\ &= \sqrt{\det(\mathcal{C}^{kl})} \mathbb{E}\left[\delta_0(w_1^1, \dots, w_1^n) \cdot \delta_0\left(\sum_{i < j} C_{ij}^1 S_1^{ij}, \dots, \sum_{i < j} C_{ij}^p S_1^{ij}\right)\right], \end{aligned}$$

where we wrote  $S_t^{ij}$  for Lévy’s stochastic area  $S_t(w^i, w^j)$  defined by (7.2) for simplicity. The generalized expectation on the right-hand side is computed in Appendix. Thus, we obtain

$$c_0(x) = \sqrt{\det(\mathcal{C}^{kl})} \frac{1}{(2\pi)^{(n/2)+p}} \int_{\mathbb{R}^p} \left[ \det \left( \frac{\sinh(\sqrt{-1}(\zeta \cdot C)/2)}{\sqrt{-1}(\zeta \cdot C)/2} \right) \right]^{-1/2} d\zeta,$$

where for  $\zeta = (\zeta^1, \dots, \zeta^p) \in \mathbb{R}^p$ ,  $(\zeta \cdot C)$  is the  $n \times n$  skew symmetric matrix defined by

$$(\zeta \cdot C) = \left( \sum_{k=1}^p \zeta^k C_{ij}^k \right)_{1 \leq i, j \leq n}.$$

Note that  $c_0(x)$  depends only on the structure constants.

Finally, let us assume that  $\hat{x} \in \hat{M}$  satisfies (7.7). Then, we can find a local adapted frame  $\{\hat{V}_1, \dots, \hat{V}_n, \hat{Z}_1, \dots, \hat{Z}_p\}$  around  $\hat{x}$  which yields the same structure constants  $(C_{ij}^k)$ . By doing the same computation again, we see  $c_0(x) = c_0(\hat{x})$ .

**Appendix A. On step-two nilpotent Lie groups.**

In [17], Gaveau obtained explicit expressions for the heat kernel for the Heisenberg groups and the free nilpotent Lie groups of step-two. The heat kernels for all nilpotent Lie groups of step-two were obtained by Cygan ([14]). They used an analytic method. In this section, we recover such expressions by using a probabilistic method. Indeed, we shall obtain the expressions by using an explicit expression of a stochastic oscillatory integral with a quadratic Wiener functional as its phase function (cf. [38]). Lévy’s stochastic area defined in (7.2) is a typical example of such a quadratic Wiener functional.

We start this section with a preliminary observation on linear combinations of Lévy’s stochastic areas. For  $t \geq 0$ ,  $x \in \mathbb{R}^n$ ,  $1 \leq i, j \leq n$ , and an  $n \times n$  skew symmetric matrix  $\Xi = (\Xi^{ij})_{1 \leq i, j \leq n}$ , set

$$S_t^{ij}(x) = \int_0^t \{ (x + w_s^i) \circ dw_s^j - (x + w_s^j) \circ dw_s^i \} \quad \text{and} \quad S_t(\Xi; x) = \frac{1}{2} \sum_{1 \leq i < j \leq n} \Xi^{ij} S_t^{ij}(x).$$

Our first goal of this section is revisiting the following expression ([14], [17]) by using the computation of oscillatory integrals associated with quadratic Wiener functionals in [38]. This is a generalization of the famous formula (7.3) for Lévy’s stochastic area.

**THEOREM A.1.** *It holds that*

$$\begin{aligned} \mathbb{E} \left[ e^{\sqrt{-1} S_t(\Xi; x)} \delta_y(x + w_t) \right] &= \frac{1}{(2\pi t)^{n/2}} \left[ \det \left( \frac{\sinh(\sqrt{-1} t \Xi / 2)}{\sqrt{-1} t \Xi / 2} \right) \right]^{-1/2} \\ &\times \exp \left( -\frac{\sqrt{-1}}{2} \langle \Xi x, y \rangle_{\mathbb{R}^n} - \frac{1}{2t} \langle T(t; \Xi)^{-1}(y - x), (y - x) \rangle_{\mathbb{R}^n} \right), \quad y \in \mathbb{R}^n, \end{aligned} \tag{A.1}$$

where, for  $n \times n$  matrix  $B$ ,

$$\frac{\sinh(B)}{B} = \sum_{k=1}^{\infty} \frac{1}{(2k-1)!} B^{2k-2}, \quad \cosh(B) = \sum_{k=0}^{\infty} \frac{1}{(2k)!} B^{2k},$$

and

$$T(t; \Xi) = \frac{\sinh(\sqrt{-1} t \Xi / 2)}{\sqrt{-1} t \Xi / 2} (\cosh(\sqrt{-1} t \Xi / 2))^{-1}.$$

REMARK A.2. For an  $n \times n$  skew symmetric matrix  $B$ , take  $\lambda_1, \dots, \lambda_k \in \mathbb{R} \setminus \{0\}$  such that  $\pm\sqrt{-1}\lambda_1, \dots, \pm\sqrt{-1}\lambda_k, \underbrace{0, \dots, 0}_{n-2k}$  are its eigenvalues. Then

$$\det\left(\frac{\sinh(\sqrt{-1}B)}{\sqrt{-1}B}\right) = \prod_{i=1}^k \left(\frac{\sinh \lambda_i}{\lambda_i}\right)^2 \neq 0,$$

$$\det(\cosh(\sqrt{-1}B)) = \prod_{i=1}^k (\cosh \lambda_i)^2 \neq 0.$$

Thus, the reciprocal number and the inverse matrix appearing in (A.1) are both well-defined.

PROOF. If  $i \neq j$ , then  $w_s^i \circ dw_s^j = w_s^i dw_s^j$ . Since  $\Xi$  is skew symmetric,

$$S_t(\Xi; x) = \frac{1}{2} \int_0^t \langle (-\Xi)(x + w_s), dw_s \rangle_{\mathbb{R}^n} = \frac{1}{2} \int_0^t \langle (-\Xi)w_s, dw_s \rangle_{\mathbb{R}^n} - \frac{1}{2} \langle \Xi x, w_t \rangle_{\mathbb{R}^n}.$$

By the skew symmetry of  $\Xi$  again, we have

$$\begin{aligned} &\mathbb{E}[e^{\sqrt{-1}S_t(\Xi;x)} \delta_y(x + w_t)] \\ &= e^{-\sqrt{-1}\langle \Xi x, y \rangle / 2} \mathbb{E}\left[\exp\left(\frac{\sqrt{-1}}{2} \int_0^t \langle (-\Xi)w_s, dw_s \rangle\right) \delta_{y-x}(w_t)\right]. \end{aligned} \tag{A.2}$$

Thus it suffices to compute  $\mathbb{E}[e^{\sqrt{-1}S_t(\Xi;0)} \delta_y(w_t)]$ .

Applying [38, Corollary 1.1 and Example 4.2], we obtain

$$\begin{aligned} &\mathbb{E}[e^{\sqrt{-1}S_t(\Xi;0)} \delta_y(w_t)] \\ &= \frac{1}{\sqrt{\det A(0, t; \Xi)}} \frac{1}{(2\pi)^{n/2} \sqrt{C(t; \Xi)}} \exp\left(-\frac{1}{2} \langle C(t; \Xi)^{-1}y, y \rangle_{\mathbb{R}^n}\right) \end{aligned} \tag{A.3}$$

where

$$\begin{aligned} A(s, t; \Xi) &= \frac{1}{2} \{I + \exp(-\sqrt{-1}(s - t)\Xi)\}, \\ C(t; \Xi) &= \int_0^t (A(0, s; \Xi)^{-1})^* A(0, s; \Xi)^{-1} ds, \end{aligned}$$

and, for  $n \times n$ -matrix  $B$ ,  $\exp(B) = \sum_{k=0}^\infty (1/k!)B^k$  and  $B^*$  is the transposed matrix of  $B$ . It should be emphasized that the superscript  $*$  indicates just being transposed and no complex conjugate are taken even if  $B$  is a complex matrix. We shall compute  $A(s, t; \Xi)$  and  $C(t; \Xi)$ .

First rewrite as

$$A(s, t; \Xi) = \cosh\left(\frac{\sqrt{-1}}{2}(s - t)\Xi\right) \exp\left(-\frac{\sqrt{-1}}{2}(s - t)\Xi\right). \tag{A.4}$$

Since  $\Xi$  is skew symmetric,

$$\det \left( \exp \left( -\frac{\sqrt{-1}}{2}(s-t)\Xi \right) \right) = 1.$$

Thus we have

$$\det A(0, t; \Xi) = \det \left( \cosh \left( \frac{\sqrt{-1}t}{2}\Xi \right) \right). \tag{A.5}$$

Next, due to the skew symmetry of  $\Xi$  again, by (A.4), we have

$$A(s, t; \Xi)^* = \exp \left( \frac{\sqrt{-1}}{2}(s-t)\Xi \right) \cosh \left( \frac{\sqrt{-1}}{2}(s-t)\Xi \right).$$

In conjunction with (A.4) again, this implies

$$A(s, t; \Xi)A(s, t; \Xi)^* = \left( \cosh \left( \frac{\sqrt{-1}}{2}(s-t)\Xi \right) \right)^2.$$

Hence

$$(A(s, t; \Xi)^{-1})^* A(s, t; \Xi)^{-1} = \left( \cosh \left( \frac{\sqrt{-1}}{2}(s-t)\Xi \right) \right)^{-2}.$$

Recall that  $\sinh(B) = (\exp(B) - \exp(-B))/2$  and

$$\frac{d}{ds} s \frac{\sinh(sB)}{sB} (\cosh(sB))^{-1} = (\cosh(sB))^{-2}.$$

Plugging this into the definition of  $C(t; \Xi)$ , we obtain

$$C(t; \Xi) = t \frac{\sinh(\sqrt{-1}t\Xi/2)}{\sqrt{-1}t\Xi/2} \left( \cosh \left( \frac{\sqrt{-1}}{2}t\Xi \right) \right)^{-1}. \tag{A.6}$$

Plugging (A.5) and (A.6) into (A.3), we obtain

$$\begin{aligned} & \mathbb{E} [e^{\sqrt{-1} S_t(\Xi; 0)} \delta_y(w_t)] \\ &= \frac{1}{(2\pi t)^{n/2}} \left[ \det \left( \frac{\sinh(\sqrt{-1}t\Xi/2)}{\sqrt{-1}t\Xi/2} \right) \right]^{-1/2} \exp \left( -\frac{1}{2} \langle C(t; \Xi)^{-1} y, y \rangle_{\mathbb{R}^n} \right). \end{aligned}$$

Combined with (A.2), this implies the desired expression (A.1). □

REMARK A.3. Given  $\Theta = (\Theta_1, \dots, \Theta_d) \in C^\infty(\mathbb{R}^d; \mathbb{R}^d)$  whose derivatives of all order are at most polynomial growth, the Schrödinger operator  $L$  with the vector potential  $\Theta$  is given by

$$L = -\frac{1}{2} \sum_{\alpha=1}^d \left( \frac{\partial}{\partial x^\alpha} + \sqrt{-1} \Theta_\alpha \right)^2.$$

The heat kernel  $q_t(x, y)$  associated with  $L$  possesses a probabilistic expression as follows (for example, see [27, Theorem 5.5.7]).

$$q_t(x, y) = \mathbb{E}[e(t, x)\delta_y(x + w_t)],$$

where  $e(t, x)$  is given by

$$e(t, x) = \exp\left(\sqrt{-1} \sum_{i=1}^n \int_0^t \Theta_\alpha(x + w_s) \circ dw_s^i\right).$$

If  $\Theta(x) = -\Xi x/2$  for  $x \in \mathbb{R}^n$ , then  $e(t, x) = \exp(\sqrt{-1} S_t(\Xi; x))$  and hence the right-hand side of (A.1) gives an explicit expression of  $q_t(x, y)$ .

We now proceed to step-two nilpotent Lie groups. For this purpose, let  $G$  be a  $(n+p)$ -dimensional connected and simply connected step-two nilpotent Lie group with the Lie algebra  $\mathfrak{g}$ , where  $p$  is the dimension of  $[\mathfrak{g}, \mathfrak{g}]$ . Using the diffeomorphism  $\exp : \mathfrak{g} \rightarrow G$  and suitable bases of  $\{Z_k\}_{k=1}^p$  and  $\{X_i\}_{i=1}^n$  of  $[\mathfrak{g}, \mathfrak{g}]$  and its complement, respectively, we introduce a global coordinate  $(x, z) \in \mathbb{R}^{n+p}$  on  $G$ ; for  $g \in G$ ,  $g = \exp(\sum_{i=1}^n x^i X_i + \sum_{k=1}^p z^k Z_k)$ , where  $x = (x^1, \dots, x^n) \in \mathbb{R}^n$  and  $z = (z^1, \dots, z^p) \in \mathbb{R}^p$ . In terms of this coordinate, the product  $\times$  on  $G$  is given by

$$(x, z) \times (u, v) = \left(x + u, z + v + \frac{1}{2} \sum_{i,j=1}^n x_i u_j C_{ij}\right), \quad (\text{A.7})$$

where

$$[X_i, X_j] = \sum_{k=1}^p C_{ij}^k Z_k \quad \text{and} \quad C_{ij} = \begin{pmatrix} C_{ij}^1 \\ \vdots \\ C_{ij}^p \end{pmatrix} \in \mathbb{R}^p.$$

Let  $\tilde{X}_i$  and  $\tilde{Z}_k$  be the left-invariant vector fields associated with  $X_i$  and  $Z_k$ ,  $1 \leq i \leq n$  and  $1 \leq k \leq p$ , respectively. Set  $\mathbf{e}_i = (\delta_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n$  and  $\hat{\mathbf{e}}_k = (\delta_{ik})_{1 \leq i \leq p} \in \mathbb{R}^p$ . Since  $X_i = d/dt|_{t=0}(\mathbf{te}_i, 0)$  and  $Z_k = d/dt|_{t=0}(0, t\hat{\mathbf{e}}_k)$ , we have

$$\begin{aligned} \tilde{X}_i(x, z) &= \frac{d}{dt} \Big|_{t=0} (x, z) \times (\mathbf{te}_i, 0) = \left(\frac{\partial}{\partial x^i}\right)_x + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^p x_j C_{ji}^k \left(\frac{\partial}{\partial z^k}\right)_z, \\ \tilde{Z}_k(x, z) &= \frac{d}{dt} \Big|_{t=0} (x, z) \times (0, t\hat{\mathbf{e}}_k) = \left(\frac{\partial}{\partial z^k}\right)_z. \end{aligned}$$

This implies  $[\tilde{X}_i, \tilde{X}_j] = \sum_{k=1}^p C_{ij}^k \tilde{Z}_k$ , and hence  $\tilde{X}_1(x), \dots, \tilde{X}_n(x), \tilde{Z}_1(x), \dots, \tilde{Z}_p(x)$  spans  $T_x G$  for every  $x \in G$ . In particular,  $\tilde{X}_1, \dots, \tilde{X}_n$  satisfies the equiregular Hörmander condition at every  $x \in G$ . Then the heat equation associated with the second order differential operator

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^n \tilde{X}_i^2$$

possesses the heat kernel  $p_t((x_0, y_0), (x, z))$  with respect to the Lebesgue measure. Note that the Lebesgue measure is a Haar measure on  $G$ , because, by (A.7), the Jacobian determinant of the left translation is equal to 1. Moreover, by [3], it coincides with Popp’s measure multiplied by  $(\det((\sum_{i,j=1}^n C_{ij}^k C_{ij}^\ell)_{1 \leq k, \ell \leq p}))^{1/2}$ .

The diffusion process generated by  $\mathcal{L}$  is

$$\left( \left( x_0 + w_t, z_0 + \sum_{i < j} C_{ij} S_t^{ij}(x_0) \right) \right)_{t \geq 0}.$$

Due to the Hörmander condition, with the help of generalized Wiener functional, the heat kernel is represented as

$$p_t((x_0, z_0), (x, z)) = \mathbb{E} \left[ \delta_{(x,z)} \left( x_0 + w_t, z_0 + \sum_{i < j} C_{ij} S_t^{ij}(x_0) \right) \right].$$

By the left invariance of  $\tilde{X}_i, 1 \leq i \leq n$ , it holds

$$\left( x_0 + w_t, z_0 + \sum_{i < j} C_{ij} S_t^{ij}(x_0) \right) = (x_0, z_0) \times \left( w_t, \sum_{i < j} C_{ij} S_t^{ij}(x_0) \right).$$

Hence

$$p_t((x_0, z_0), (x, z)) = p_t((0, 0), (x_0, z_0)^{-1} \times (x, z)).$$

Thus, in what follows, we assume  $(x_0, z_0) = (0, 0)$ .

Using the Fourier transform of the Dirac measure, we have

$$p_t((0, 0), (x, z)) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} e^{-\sqrt{-1} \langle \zeta, z \rangle} \mathbb{E} [ e^{\sqrt{-1} S_t((\zeta \cdot C); 0)} \delta_x(w_t) ] d\zeta, \tag{A.8}$$

where for  $\zeta = (\zeta^1, \dots, \zeta^p) \in \mathbb{R}^p$ ,  $(\zeta \cdot C)$  is the  $n \times n$  skew symmetric matrix

$$(\zeta \cdot C) = \left( \sum_{k=1}^p \zeta^k C_{ij}^k \right)_{1 \leq i, j \leq n}.$$

If we define

$$\begin{aligned} \hat{p}_t((x, z)) &= \frac{1}{(2\pi t)^{(n/2)+p}} \int_{\mathbb{R}^p} \left[ \det \left( \frac{\sinh(\sqrt{-1}(\eta \cdot C)/2)}{\sqrt{-1}(\eta \cdot C)/2} \right) \right]^{-1/2} \\ &\quad \times \exp \left( -\frac{1}{t} \left\{ \langle \eta, z \rangle_{\mathbb{R}^p} + \frac{1}{2} \langle T(1; (\eta \cdot C))^{-1} x, x \rangle_{\mathbb{R}^n} \right\} \right) d\eta, \end{aligned}$$

then plugging Theorem A.1 into (A.8) and using the change variable  $\eta = t\zeta$ , we obtain

$$p_t((0, 0), (x, z)) = \widehat{p}_t((x, z)).$$

Summing up, we arrive at the following expression of the heat kernel, which was also shown in an analytical way in [6], [14].

THEOREM A.4. *The heat kernel associated with  $\mathcal{L}$  has the form*

$$p_t((x_0, z_0), (x, z)) = \widehat{p}_t((x_0, z_0)^{-1} \times (x, z)). \quad (\text{A.9})$$

## References

- [1] S. Aida, S. Kusuoka and D. Stroock, On the support of Wiener functionals, In: Asymptotic Problems in Probability Theory: Wiener Functionals and Asymptotics (Sanda/Kyoto, 1990), Pitman Res. Notes Math. Ser., **284**, Longman Sci. Tech., Harlow, 1993, 3–34.
- [2] D. Barilari, Trace heat kernel asymptotics in 3D contact sub-Riemannian geometry, Translation of Sovrem. Mat. Prilozh., **82** (2012), *J. Math. Sci. (N.Y.)*, **195** (2013), 391–411.
- [3] D. Barilari and L. Rizzi, A formula for Popp’s volume in sub-Riemannian geometry, *Anal. Geom. Metr. Spaces*, **1** (2013), 42–57.
- [4] W. Bauer, K. Furutani and C. Iwasaki, Spectral zeta function of the sub-Laplacian on two step nilmanifolds, *J. Math. Pures Appl. (9)*, **97** (2012), 242–261.
- [5] W. Bauer, K. Furutani and C. Iwasaki, Spectral zeta function on pseudo  $H$ -type nilmanifolds, *Indian J. Pure Appl. Math.*, **46** (2015), 539–582.
- [6] R. Beals, B. Gaveau and P. Greiner, The Green function of model step two hypoelliptic operators and the analysis of certain tangential Cauchy Riemannian complexes, *Adv. Math.*, **121** (1996), 288–345.
- [7] R. Beals, P. C. Greiner and N. K. Stanton, The heat equation on a CR manifold, *J. Differential Geom.*, **20** (1984), 343–387.
- [8] G. Ben Arous, Développement asymptotique du noyau de la chaleur hypoelliptique sur la diagonale, *Ann. Inst. Fourier (Grenoble)*, **39** (1989), 73–99.
- [9] G. Ben Arous and R. Léandre, Décroissance exponentielle du noyau de la chaleur sur la diagonale. I, *Probab. Theory Related Fields*, **90** (1991), 175–202.
- [10] G. Ben Arous and R. Léandre, Décroissance exponentielle du noyau de la chaleur sur la diagonale. II, *Probab. Theory Related Fields*, **90** (1991), 377–402.
- [11] R. M. Bianchini and G. Stefani, Graded approximations and controllability along a trajectory, *SIAM J. Control Optim.*, **28** (1990), 903–924.
- [12] V. I. Bogachev, Gaussian Measures, *Math. Surveys Monogr.*, **62**, Amer. Math. Soc., Providence, RI, 1998.
- [13] D.-C. Chang and S.-K. Yeung, Trace of heat kernel, spectral zeta function and isospectral problem for sub-Laplacians, *Sci. China Ser. A*, **52** (2009), 2570–2589.
- [14] J. Cygan, Heat kernels for class 2 nilpotent groups, *Studia Math.*, **64** (1979), 227–238.
- [15] S. Dragomir and G. Tomassini, Differential Geometry and Analysis on CR Manifolds, *Progr. Math.*, **246**, Birkhäuser Boston, Inc., Boston, MA, 2006.
- [16] P. K. Friz and N. B. Victoir, Multidimensional Stochastic Processes as Rough Paths, *Cambridge Stud. Adv. Math.*, **120**, Cambridge Univ. Press, Cambridge, 2010.
- [17] B. Gaveau, Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents, *Acta Math.*, **139** (1977), 95–153.
- [18] E. Grong and A. Thalmaier, Curvature-dimension inequalities on sub-Riemannian manifolds obtained from Riemannian foliations: part I, *Math. Z.*, **282** (2016), 99–130.
- [19] K. Habermann, Small-time fluctuations for sub-Riemannian diffusion loops, *Probab. Theory Related Fields*, **171** (2018), 617–652.
- [20] A. Hassannezhad and G. Kokarev, Sub-Laplacian eigenvalue bounds on sub-Riemannian manifolds, *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, **16** (2016), 1049–1092.
- [21] E. P. Hsu, Stochastic Analysis on Manifolds, *Grad. Stud. Math.*, **38**, Amer. Math. Soc., Providence, RI, 2002.

- [22] Y. Hu, Analysis on Gaussian spaces, [World Scientific](#), 2017.
- [23] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, second edition, North-Holland Math. Library, **24**, North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, 1989.
- [24] Y. Inahama and S. Taniguchi, Short time full asymptotic expansion of hypoelliptic heat kernel at the cut locus, [Forum Math. Sigma](#), **5** (2017), e16, 74 pp.
- [25] H. Kunita, Supports of diffusion processes and controllability problems, In: Proceedings of the International Symposium on Stochastic Differential Equations, Kyoto, 1976, Wiley, New York-Chichester-Brisbane, 1978, 163–185.
- [26] S. Kusuoka and D. W. Stroock, Applications of the Malliavin calculus. II, *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, **32** (1985), 1–76.
- [27] H. Matsumoto and S. Taniguchi, Stochastic Analysis: Itô and Malliavin Calculus in Tandem, [Cambridge Stud. Adv. Math., \*\*159\*\*, Cambridge Univ. Press, 2017.](#)
- [28] G. Métivier, Fonction spectrale et valeurs propres d’une classe d’opérateurs non elliptiques, [Comm. Partial Differential Equations](#), **1** (1976), 467–519.
- [29] R. Montgomery, A Tour of Subriemannian Geometries, Their Geodesics and Applications, [Math. Surveys Monogr., \*\*91\*\*, Amer. Math. Soc., Providence, RI, 2002.](#)
- [30] D. Nualart, The Malliavin Calculus and Related Topics, second edition, [Probab. Appl. \(N.Y.\)](#), Springer-Verlag, Berlin, 2006.
- [31] E. Paoli, Small time asymptotic on the diagonal for Hörmander’s type hypoelliptic operators, [J. Dyn. Control Syst., \*\*23\*\* \(2017\), 111–143.](#)
- [32] R. S. Ponge, Heisenberg calculus and spectral theory of hypoelliptic operators on Heisenberg manifolds, [Mem. Amer. Math. Soc., \*\*194\*\* \(2008\), viii+ 134 pp.](#)
- [33] I. Shigekawa, Stochastic Analysis, [Transl. Math. Monogr., \*\*224\*\*, Iwanami Series in Modern Math., Amer. Math. Soc., Providence, RI, 2004.](#)
- [34] N. K. Stanton, Spectral invariants of CR manifolds, [Michigan Math. J., \*\*36\*\* \(1989\), 267–288.](#)
- [35] S. Takanobu, Diagonal short time asymptotics of heat kernels for certain degenerate second order differential operators of Hörmander type, [Publ. Res. Inst. Math. Sci., \*\*24\*\* \(1988\), 169–203.](#)
- [36] S. Takanobu, Multiple stochastic integrals appearing in the stochastic Taylor expansions, [J. Math. Soc. Japan](#), **47** (1995), 67–92.
- [37] S. Taniguchi, Malliavin’s stochastic calculus of variations for manifold-valued Wiener functionals and its applications, [Z. Wahrsch. Verw. Gebiete](#), **65** (1983), 269–290.
- [38] S. Taniguchi, Stochastic oscillatory integrals with quadratic phase function and Jacobi equations, [Probab. Theory Relat. Fields](#), **114** (1999), 291–308.
- [39] A. Thalmaier, Geometry of subelliptic diffusions, In: Geometry, Analysis and Dynamics on sub-Riemannian Manifolds: Volume II, EMS Ser. Lect. Math., European Math. Soc., 2016, 85–169.
- [40] Y. Yamato, Stochastic differential equations and nilpotent Lie algebras, [Z. Wahrsch. Verw. Gebiete](#), **47** (1979), 213–229.
- [41] N. Yoshida, Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe, [Probab. Theory Related Fields](#), **92** (1992), 275–311.

Yuzuru INAHAMA

Faculty of Mathematics  
 Kyushu University  
 Motooka 744, Nishi-ku  
 Fukuoka 819-0395, Japan  
 E-mail: inahama@math.kyushu-u.ac.jp

Setsuo TANIGUCHI

Faculty of Arts and Science  
 Kyushu University  
 Motooka 744, Nishi-ku  
 Fukuoka 819-0395, Japan  
 E-mail: se2otngc@artsci.kyushu-u.ac.jp