On an Inversion Theorem for Stratonovich's Signatures of Multidimensional Diffusion Paths

X. Geng*and Z. Qian†

Abstract

In the present paper, we prove that with probability one, the Stratonovich sigatures of a multidimensional diffusion process (possibly degenerate) over $[0,1]$, which is the collection of all iterated Stratonovich's integrals of the diffusion process over $[0,1]$, determine the diffusion sample paths.

MSC: 60J60; 60G17; 60J45

Keywords: Hypoelliptic diffusions; Rough paths; Stratonovich's signatures

1 Introduction

Let X_t be an $\mathbb{R}^d$-valued continuous path over $[0,1]$ with bounded variation ($d \geqslant 2$). According to [8], [9], for $0 \leqslant s < t \leqslant 1$, we can define the sequence of iterated integrals

$$\mathbf{X}_{s,t} = (1, X_{s,t}^1, X_{s,t}^2, \cdots, X_{s,t}^n, \cdots),$$

where

$$X_{s,t}^n = \int_{s<u_1<\cdots<u_n<t} dX_{u_1} \otimes \cdots \otimes dX_{u_n}, \quad n \geqslant 1. \tag{1.1}$$

$X_{s,t}^n$ is regarded as an element in the tensor space $(\mathbb{R}^d)^{\otimes n} \cong \mathbb{R}^{nd}$ and $\mathbf{X}_{s,t}$ is hence an element in the tensor algebra

$$T^{(\infty)}(\mathbb{R}^d) = \oplus_{n=0}^{\infty} \mathbb{R}^{nd}.$$

$\mathbf{X}_{s,t}$ is multiplicative in the sense that it satisfies the following Chen's identity:

$$\mathbf{X}_{s,t} = \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}, \quad 0 \leqslant s < u < t \leqslant 1.$$

*Mathematical Institute, University of Oxford, Oxford OX2 6GG, England and the Oxford-Man Institute, University of Oxford, Oxford OX2 6ED, England.
Email: xi.geng@maths.ox.ac.uk.

†Exeter College, University of Oxford, Oxford OX1 3DP, England.
Email: qianz@maths.ox.ac.uk.

$\mathbf{X}_{s,t}$ is uniquely determined by the original path X_t; or intuitively speaking, the original path X_t contains all information about its differential dX_t. A remarkable consequence is that a theory of integration along X_t can be established in the sense of Riemann–Stieltjes, which leads to a theory of differential equations driven by X_t. Such a theory for paths with bounded variation is classical and well-studied.

If the path X_t is less regular, for example, X_t has finite p-variation for some $p > 1$, it may not be possible to establish an integration theory along X_t by using the information of the original path only. The fundamental reason is that the path X_t itself does not reveal enough information on its differential dX_t , which is essential to be fully understood if we want to develop an integration theory along X_t. As pointed out by T. Lyons in [7], for this purpose, together with the path itself, a finite sequence of iterated integrals up to level $[p]$ satisfying Chen's identity should be specified in advance. Such a finite sequence of iterated integrals

$$\mathbf{X}_{s,t} = (1, X^1_{s,t}, \cdots, X^{[p]}_{s,t})$$

is regarded as a multiplicative functional $\mathbf{X}$ from the simplex $\Delta = \{(s,t) : 0 \leqslant s \leqslant t \leqslant 1\}$ to the truncated tensor algebra

$$T^{([p])}(\mathbb{R}^d) = \oplus_{n=0}^{[p]} \mathbb{R}^{nd}.$$

$\mathbf{X}$ is called a rough path with roughness p. According to [7], $\mathbf{X}$ extends uniquely to a multiplicative functional from Δ to $T^{(\infty)}(\mathbb{R}^d)$. In the founding work of T. Lyons in [7], a general theory of integration and differential equations for rough paths was established.

For a rough path $\mathbf{X}$ with roughness p, the signature of $\mathbf{X}$ is defined as the formal sequence

$$S(\mathbf{X}) = \mathbf{X}_{0,1} = (1, X^1_{0,1}, \cdots, X^{[p]}_{0,1}, \cdots),$$

where for $n > [p]$, $X^n_{s,t}$ is the unique extension of $\mathbf{X}$ as mentioned before. The signature $S(\mathbf{X})$, proposed by K.T. Chen in [2] and T. Lyons in [7], can be regarded as the collection of overall information of any arbitrary level n about the rough path $\mathbf{X}$. It is of central interest and conjectured in the theory of rough paths that the signature $S(\mathbf{X})$ contains sufficient information to recover the path $\mathbf{X}$ completely. In the groundbreaking paper [3] by B. Hambly and T. Lyons, they proved that for a path X_t with bounded variation, the signature of X_t uniquely determines the path up to a tree-like equivalence. However, for paths with unbounded variation, very few results are available and it remains a lot of work to do.

In the work [6] by Y. Le Jan and Z. Qian, they considered the case of multi-dimensional Brownian motion and proved that for almost surely, the Brownian paths can be recovered by using the so-called Stratonovich's signature, which is defined via iterated Stratonovich's integrals of arbitrary orders. Since the Brownian paths are of unbounded variation and can be regarded as rough paths with roughness $p \in (2, 3)$, we may need to specify the second level in order to make sense in terms of rough paths. However, according to [5], [10], there is a canonical lifting of the Brownian paths to the second level by using dyadic approximations, which is called the Lévy's stochastic area process and it coincides exactly with the iterated Stratonovich's integral defined in the same way as (1.1). Such lifting is determined by the Brownian paths itself, and in [6] when regarding the Brownian motion as rough paths such lifting was used by the authors. Therefore, the recovery of Brownian motion as rough paths is essentially the recovery of the Brownian paths in terms of Stratonovich's signature.

In the present paper, we are going to generalize the result of Y. Le Jan and Z. Qian in [6] to the case of multidimensional diffusion processes (possibly degenerate). The main idea of the proof is similar to the case of Brownian motion, in which the authors used a specially designed approximation scheme and chose special differential 1-forms to define the so-called extended Stratonovich's signatures to recover the Brownian paths. However, there are several difficulties in the case of diffusion processes. Firstly, we need quantitative estimates for rare events of diffusion processes to prove a convergence result similar to the case of Brownian motion. In [6], the authors used the symmetry and explicit distribution of Brownian motion, which are not available in the case of diffusion processes and hence we need to proceed in a different way. Secondly, to construct special differential 1-forms, a quite special case of Hörmander's theorem was used to ensure the existence of density, in which the so-called Hörmander's condition was easily verified. In the case of diffusion processes, the construction of differential 1-forms is more complicated to ensure similar kind of hypoellipticity. Lastly, in the Brownian motion case, the Laplace operator is well-posed so that PDE methods could be applied to obtain a crucial estimate which enables us to relate the extended Stratonovich's signatures to the Brownian paths. However, for a general diffusion process, the generator L may not be well-posed any more (we don't impose uniform ellipticity assumption on L) and PDE methods may no longer apply (in fact, to ensure the application of PDE methods, rather technical assumptions should be imposed on the differential operator L and the domain if without uniform ellipticity). Therefore, we need a different approach to recover the diffusion paths by using extended Stratonovich's signatures.

2 Main result and idea of the proof

In this section, we are going to state our main result and illustrate the idea of the proof.

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let W_t be a d-dimensional Brownian motion on Ω. Consider an N-dimensional ($N \geqslant 2$) diffusion process X_t defined by the following SDE (possibly degenerate):

$$dX_t = \sum_{\alpha=1}^{d} V_\alpha(X_t) \circ dW_t^\alpha + V_0(X_t)dt \tag{2.1}$$

with $X_0 = 0$.

We are going to make the following three assumptions on the generating vector fields $\{V_1, \cdots, V_d; V_0\}$.

(A) $V_0, V_1, \cdots, V_d \in C_b^\infty(\mathbb{R}^N)$.

(B) For any $x \in \mathbb{R}^N$, Hörmander's condition (see [4]) holds at x in the sense that

$$V_1, \cdots, V_d,\ [V_\alpha, V_\beta], 0 \leqslant \alpha, \beta \leqslant d,\ [V_\alpha, [V_\beta, V_\gamma]], 0 \leqslant \alpha, \beta, \gamma \leqslant d,\ \cdots$$

generate the tangent space $T_x\mathbb{R}^N \cong \mathbb{R}^N$, where $[\cdot, \cdot]$ denotes the Lie bracket.

(C) There exists a positive orthonormal basis $\{e_1, \cdots, e_N\}$ of $\mathbb{R}^N$, such that for any $x \in \mathbb{R}^N$ and $i = 1, 2, \cdots, N$, $V_\alpha(x)$ is not perpendicular to e_i for some $\alpha = 1, 2, \cdots, d$.

Remark 2.1. Assumptions (A) and (B) are made to ensure the hypoellipticity of the generator

$$L = \frac{1}{2}\sum_{\alpha=1}^{d} V_\alpha^2 + V_0$$

of the diffusion process (2.1). Assumption (C) is made to ensure the escape condition and the non-tangential condition proposed in [1] hold on some domain of interest which is relatively small. Under these assumptions, we are able to apply results in [1] to obtain the existence of a continuous density function of the Poisson kernel for some domain of interest and a quantitative estimate on the density function, which are both crucial in the proof of our main result.

It should be pointed out that if the diffusion process (2.1) is nondegenerate, that is, if $\{V_1(x), \cdots, V_d(x)\}$ generate the tangent space $T_x\mathbb{R}^N \cong \mathbb{R}^N$ at each point $x \in \mathbb{R}^N$, then Assumptions (A), (B), (C) are all verified.

For $n \geqslant 1$, $j_1, \cdots, j_n \in \{1, 2, \cdots, N\}$, define the iterated Stratonovich's integral of order n:

$$[j_1, \cdots, j_n]_{s,t} = \int_{s<t_1<\cdots<t_n<t} \circ dX_{t_1}^{j_1} \circ dX_{t_2}^{j_2} \circ \cdots \circ dX_{t_n}^{j_n}, \quad 0 \leqslant s < t \leqslant 1.$$

Alternatively, $[j_1, \cdots, j_n]_{s,t}$ can be defined inductively by the following relation:

$$[j_1, \cdots, j_n]_{s,t} = \int_{s<u<t} [j_1, \cdots, j_{n-1}]_{s,u} \circ dX_u^{j_n}, \quad 0 \leqslant s < t \leqslant 1,$$

where $[j_1]_{s,t}$ is defined to be

$$[j_1]_{s,t} = \int_{s<u<t} \circ dX_u^{j_1} = X_t^{j_1} - X_s^{j_1}, \quad 0 \leqslant s < t \leqslant 1.$$

For convenience, if $n = 0$, we denote $[j_1, \cdots, j_n]_{s,t} = 1$. The family

$$\{[j_1, \cdots, j_n]_{0,1} :\ j_1, \cdots, j_n \in \{1, 2, \cdots, N\},\ n \geqslant 0\}$$

of iterated Stratonovich's integrals is called the *Stratonovich signature* of X_t over $[0, 1]$.

Let $\mathcal{F}_1$ be the completion of the σ-algebra generated by the diffusion process X_t over $[0,1]$, and let $\mathcal{G}_1$ be the completion of the σ-algebra generated by the Stratonovich's signature of X_t over $[0,1]$. More precisely,

$$\begin{aligned}\mathcal{F}_1 &= \overline{\sigma(X_t:\ 0\leqslant t\leqslant 1)},\\ \mathcal{G}_1 &= \overline{\sigma(\{[j_1,\cdots,j_n]_{0,1}:\ j_1,\cdots,j_n\in\{1,2,\cdots,N\},n\geqslant 0\})}.\end{aligned}$$

For the case of Brownian motion, it was proved by Y. Le Jan and Z. Qian in [6] that

$$\mathcal{F}_1=\mathcal{G}_1.$$

Such result for diffusion processes in our setting can be proved in the present paper. However, we are going to formulate the problem in a more illustrative way, which to some extend reveals how we can reconstruct the diffusion paths from the Stratonovich's signature over $[0,1]$ in a conceivable way.

First we need the following definition.

Definition 2.1. A *piecewise linear trajectory* (P.L.T.) $\mathcal{T}$ in $\mathbb{R}^N$ is a finite sequence of points in $\mathbb{R}^N$ (not necessarily all distinct). Here we always assume that the number of points in $\mathcal{T}$ is greater than one (if $\mathcal{T}$ consists of only one point x, we will regard $\mathcal{T}$as the finite sequence (x,x)). For a P.L.T. $\mathcal{T}$ in $\mathbb{R}^N$, the number of points in $\mathcal{T}$will be denoted by $|\mathcal{T}|$. If the points of $\mathcal{T}$ belongs to a subset $\Gamma\subset\mathbb{R}^N$, we say that $\mathcal{T}$ is a P.L.T. in Γ.

The reason why we use the notion "piecewise linear trajectory" is that when given $\mathcal{T}$, we actually think of $\mathcal{T}$ as a piecewise linear graph by connecting the points in $\mathcal{T}$ by line segments in order. Here we should point out that the order of points in $\mathcal{T}$ is rather important, and no parametrizations are involved.

Definition 2.2. For $n\geqslant 2$, a *parametrization* σ of order n is a partition of the time interval $[0,1]$ into $n-1$ nontrivial subintervals:

$$\sigma:\ 0=t_1<t_2<\cdots<t_{n-1}<t_n=1.$$

The space of all parametrizations of order n will be denoted by $\mathcal{P}_n$.

Let $\mathcal{T}$ be a P.L.T. in $\mathbb{R}^N$ and let σ be a parametrization of order $|\mathcal{T}|$. The piecewise linear path over $[0,1]$ defined by applying linear interpolation of $\mathcal{T}$ along the parametrization σ is denoted by $\mathcal{T}(t|\sigma)$.

Our formulation of the problem is related to a kind of convergence which is parametrization free. Therefore, we need the following definition of convergence in trajectory.

Definition 2.3. Let $(\gamma_t)_{0\leqslant t\leqslant 1}$ be a continuous path in $\mathbb{R}^N$. A sequence $\{\mathcal{T}^{(n)}\}$ of P.L.T.s is said to be *converging in trajectory* to $(\gamma_t)_{0\leqslant t\leqslant 1}$ if

$$\lim_{n\to\infty}\inf_{\sigma\in\mathcal{P}_{|\mathcal{T}^{(n)}|}}\sup_{0\leqslant t\leqslant 1}|\gamma_t-\mathcal{T}^{(n)}(t|\sigma)|=0.$$

Remark 2.2. Such kind of convergence modulo parametrization is similar to the notion of Fréchet distance, which was originally introduced by M. Fréchet in the study of shapes of geometric spaces.

Now we are in a position to state our main result.

Theorem 2.1. *Let $\mathcal{Z}$ be the space of P.L.T.s in $\mathbb{Z}^N$ equipped with the discrete σ-algebra. Then there exists a sequence $\{\mathcal{T}^{(n)}\}$ of $\mathcal{Z}$-valued $\mathcal{G}_1$-measurable random variables (random P.L.T.s), such that with probability one, $\frac{1}{n} \cdot \mathcal{T}^{(n)}$ converges in trajectory to the diffusion paths $(X_t)_{0\leqslant t\leqslant 1}$.*

It seems that the statement of Theorem 2.1 does not contain much information about the approximating sequence $\{\mathcal{T}^{(n)}\}$. However, when from the proof in the next section, we will see that $\mathcal{T}^{(n)}$ is constructed in a quite explicit way.

It should be pointed out that the result of Theorem 2.1 was already implicitly proved in [6] for the case of Brownian motion.

A direct consequence of Theorem 2.1 is the result based on Y. Le Jan and Z. Qian's formulation.

Theorem 2.2. $\mathcal{F}_1 = \mathcal{G}_1$.

Before proving our main result Theorem 2.1 in the next section, we first illustrate the idea and main steps of the proof.

We adopt the scheme and the key observation that the diffusion paths can be recovered by reading out the maximal sequence of well-chosen compactly supported differential 1-forms such that the iterated Stratonovich's integral of those 1-forms (extended Stratonovich's signature) along the diffusion paths over the duration of visiting their supports is nonzero , which were proposed in [6].

The idea of the proof of Theorem 2.1 is the following.

Firstly, decompose the Euclidean space $\mathbb{R}^N$ into disjoint small boxes and narrow tunnels. By recording the successive visit times of those small boxes, we can construct a piecewise linear approximation of the diffusion paths. A convergence theorem can be proved by developing certain types of estimates of rare events for the diffusion process. By enlarging the size of those small boxes a little bit (by a higher order infinitesimal relative to the size of boxes), we can similarly get another piecewise linear approximation also converging to the diffusion paths as the size of boxes goes to zero. Secondly, we construct a family of "special" differential 1-forms on $\mathbb{R}^N$ (depending on the size of boxes) in a way that for any larger box, we construct a 1-form supported in it such that it is highly non-degenerate on the inner smaller box. The crucial observation is that the Stratonovich's integral of any of those 1-forms along the diffusion paths over the duration of visit of its support is nonzero. It turns out that for a diffusion path, we can read out an associated unique maximal finite sequence of 1-forms (a P.L.T.) recording a sequence of boxes in order such that the iterated Stratonovich's integral of this sequence of 1-forms (extended Stratonovich's signature) along the diffusion path over the duration of visiting their supports is nonzero. It provides us with sufficient information to recover the diffusion path by taking limit in a reasonable way. This is due to the fact that based on our construction, we can prove that such a maximal sequence always "lies" between the two piecewise linear approximations constructed before, both of which converge to the diffusion path. Here we need to develop a kind of squeeze theorem for the type of convergence (convergence in trajectory in the setting of P.L.T.s defined as before) in our situation.

To carry out the above idea, we are going to establish the following three steps.

(1) Step one: proving a convergence result for the piecewise linear approximation based on successive visit times of small boxes.

The proof consists of two ingredients. The first one is a probabilistic estimate of the number of boxes visited over the time duration $[0, 1]$, which can be developed by using a random time change technique. It turns out that we can reduce to the Brownian motion case. The importance of such an estimate is that we can get an asymptotic rate of the probability that the number of boxes visited over $[0, 1]$ is quite large. The second one is the probabilistic estimate of the uniform distance between the piecewise linear approximation path and the original diffusion path, provided the number of boxes visited over $[0, 1]$ is fixed. This can be done by using the Strong Markov property and a quantitative result in [1] by G. Ben Arous, S. Kusuoka and D.W. Stroock, which gives us control on the density of the Poisson kernel of a given bounded domain in $\mathbb{R}^N$ and enables us to estimate the probability that the diffusion process travels through narrow tunnels. Combining the two ingredients, it is not hard to prove the convergence result by using the Borel-Cantelli's lemma via a subsequence.

(2) Step two: constructing special differential 1-forms and using extended Stratonovich's signatures.

For any larger box, we are going to construct a suitable differential 1-form supported in it and highly nondegenerate on the inner smaller box. The construction of such a differential 1-form can be reduced to the construction of a differential 1-form such that the generator of some associated SDE with dimension $N + 1$ is hypoelliptic on the support of the differential 1-form. The family of diffenrential 1-forms constructed in such a way will be used to construct extended Stratonovich's signatures, which in turn will be used to recover the diffusion paths as stated in the idea of the proof.

(3) Step three: proving a squeeze theorem for convergence in trajectory to recover the diffusion paths.

From the above two steps we constructed two sequences of piecewise linear approximations of the diffusion paths, and between which a sequence of P.L.T.s in terms of extended Stratonovich signatures. We will formulate the term "lying between" in a rigorous way in the setting of P.L.T.s and prove a squeeze theorem for convergence in trajectory which fits our situation. Here the squeeze theorem we are going to prove is not in the most general case (we need to make use of special parametrizations), so we need to modify the piecewise linear approximation associated to larger boxes to fit our case.

An advantage of using such a squeeze theorem is that we can get around the estimates based on potential theory and partial differential equations, which was used in [6] for the Laplace operator. In fact, for a general elliptic operator L, the associated partial differential equation may not be well-posed and the conditions to ensure a (regular) probabilistic representation of a solution is quite restrictive and technical.

3 Proof of the main result

In this section, we will give the detailed proof of our main result Theorem 2.1.

Recall that $\{X_t : t \geqslant 0\}$ is an N-dimensional diffusion process defined by the following SDE:

$$dX_t = \sum_{\alpha=1}^{d} V_\alpha(X_t) \circ dW_t^\alpha + V_0(X_t)dt$$

with $X_0 = 0$, in which the generating vector fields satisfy Assumptions (A), (B), (C).

In the following the coordinates of $x \in \mathbb{R}^N$ is taken with respect to the orthonormal basis given in Assumption (C).

3.1 Discretization and an approximation result

Similar to the idea of Y. Le Jan and Z. Qian, we first construct a suitable approximation scheme for the diffusion paths.

For convenience, a constant is called universal if it depends only on the generator L and the dimensions N, d. Moreover, sometimes we may use the same notation to denote universal constants coming out from estimates, although they may be different from line to line.

Let $0 < \varepsilon < 1$. For $z = (z_1, \cdots, z_N) \in \mathbb{Z}^N$, let H_z^ε be the N-cube in $\mathbb{R}^N$ defined by

$$H_z^\varepsilon = \{(x_1, \cdots, x_N) : \ \varepsilon z_i - \frac{\varepsilon - \varepsilon^\mu}{2} \leqslant x_i \leqslant \varepsilon z_i + \frac{\varepsilon - \varepsilon^\mu}{2}, \ i = 1, 2, \cdots, N\},$$

where μ is some universal constant to be chosen later on.

For technical reasons we assume that the boundary of H_z^ε is smoothed to the order of $\varepsilon^{2\mu}$. Such a smoothing procedure can be done in a simple geometric way, or by using standard mollifiers. In the case of $N = 2$, we can simply replace each corner of H_z^ε by a quarter of a circle with radius $\varepsilon^{2\mu}$. The space $\mathbb{R}^N$ is then divided into disjoint small boxes and narrow tunnels.

Now we are going to construct an approximation of diffusion paths X_t over the time duration $[0, 1]$.

Let $\tau_0^\varepsilon = 0$ and $\boldsymbol{m}_0^\varepsilon = (0, \cdots, 0)$. For $k \geqslant 1$, define

$$\tau_k^\varepsilon = \inf\{t > \tau_{k-1}^\varepsilon : \ X_t \in \bigcup_{z \neq \boldsymbol{m}_{k-1}^\varepsilon} H_z^\varepsilon\}.$$

If $\tau_k^\varepsilon < \infty$, define $\boldsymbol{m}_k^\varepsilon$ be the integer point in $\mathbb{Z}^N$ such that $X_{\tau_k^\varepsilon} \in H_{\boldsymbol{m}_k^\varepsilon}^\varepsilon$; if $\tau_k^\varepsilon = \infty$, define $\boldsymbol{m}_k^\varepsilon = \boldsymbol{m}_{k-1}^\varepsilon$. Intuitively, the sequence of hitting times $\{\tau_k^\varepsilon\}_{k=0}^\infty$

records the successive visit times of the small boxes and the sequence of integer points $\{\boldsymbol{m}_k^\varepsilon\}_{k=0}^\infty$ records the boxes visited by the diffusion paths in order (revisit of the same box before visiting other boxes doesn't count). Note that it is possible that $P(\tau_k^\varepsilon = \infty) > 0$ since with positive probability the process can always stay in narrow tunnels after leaving some box.

Let M_H^ε be the number of boxes visited by the diffusion paths over the time duration $[0,1]$. Formally,

$$M_H^\varepsilon = \inf\{k \geqslant 0 : \ \tau_{k+1}^\varepsilon > 1\}.$$

It follows from uniform continuity of the diffusion paths over $[0,1]$ that $M_H^\varepsilon < \infty$ for almost surely.

By a standard random time change argument, we can prove the following.

Lemma 3.1. *Let* $C = \max\{\|V_1\|_\infty, \cdots, \|V_d\|_\infty, \|V_0 + \frac{1}{2}\sum_{\alpha=1}^d \nabla_{V_\alpha} V_\alpha\|_\infty\}$. *Then for any* $k > \frac{2C}{\varepsilon^\mu}$,

$$P(M_H^\varepsilon = k) \leqslant 4Nke^{-\frac{\varepsilon^{2\mu}k}{8NdC^2}}.$$

Proof. For $k \geqslant 1$, it is obvious that

$$\begin{aligned}
P(M_H^\varepsilon = k) &= P(\tau_k^\varepsilon \leqslant 1, \ \tau_{k+1}^\varepsilon > 1) \\
&\leqslant P(\bigcup_{l=1}^k \{\tau_l^\varepsilon - \tau_{l-1}^\varepsilon \leqslant \frac{1}{k}, \ \tau_k^\varepsilon \leqslant 1\}) \\
&\leqslant \sum_{l=1}^k P(\tau_l^\varepsilon - \tau_{l-1}^\varepsilon \leqslant \frac{1}{k}, \ \tau_k^\varepsilon \leqslant 1) \\
&\leqslant \sum_{l=1}^k P(\sup_{0\leqslant t\leqslant 1/k} |X_{t+\tau_{l-1}^\varepsilon} - X_{\tau_{l-1}^\varepsilon}| \geqslant \varepsilon^\mu, \ \tau_{l-1}^\varepsilon < \infty),
\end{aligned}$$

where the last inequality comes from the fact that the distance between two different boxes is bounded from below by ε^μ. By the strong Markov property, it suffices to estimate

$$P(\sup_{0\leqslant t\leqslant 1/k} |X_t - x| \geqslant \varepsilon^\mu),$$

where X_t is the diffusion process defined by (2.1) starting at $x \in \mathbb{R}^N$.

By rewriting (2.1) in the sense of Itô, we have

$$\begin{cases} dX_t = \sum_{\alpha=1}^d V_\alpha(X_t) dW_t^\alpha + \widetilde{V_0}(X_t)dt, \\ X_0 = x, \end{cases}$$

where $\widetilde{V_0}=V_0+\frac{1}{2}\sum_{\alpha=1}^d \nabla_{V_\alpha}V_\alpha$. It follows that for $k>\frac{2C}{\varepsilon^\mu}$, we have

$$\begin{aligned}
&P(\sup_{0\leqslant t\leqslant 1/k}|X_t-x|\geqslant\varepsilon^\mu)\\
=&P(\sup_{0\leqslant t\leqslant 1/k}|\sum_{\alpha=1}^d\int_0^t V_\alpha(X_s)dW_s^\alpha+\int_0^t\widetilde{V_0}(X_s)ds|\geqslant\varepsilon^\mu)\\
\leqslant&P(\sup_{0\leqslant t\leqslant 1/k}|\sum_{\alpha=1}^d\int_0^t V_\alpha(X_s)dW_s^\alpha|\geqslant\frac{\varepsilon^\mu}{2})\\
\leqslant&\sum_{i=1}^N P(\sup_{0\leqslant t\leqslant 1/k}|\sum_{\alpha=1}^d\int_0^t V_\alpha^i(X_s)dW_s^\alpha|\geqslant\frac{\varepsilon^\mu}{2\sqrt{N}}).
\end{aligned}$$

By using a standard random time change technique and the inequality

$$\int_x^\infty\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt\leqslant e^{-\frac{1}{2}x^2},\quad x>0,$$

it is then easy to obtain that

$$P(\sup_{0\leqslant t\leqslant 1/k}|X_t-x|\geqslant\varepsilon^\mu)\leqslant 4Ne^{-\frac{\varepsilon^{2\mu}k}{8NdC^2}}.$$

Therefore, we have

$$P(M_H^\varepsilon=k)\leqslant 4Nke^{-\frac{\varepsilon^{2\mu}k}{8NdC^2}},\quad k>\frac{2C}{\varepsilon^\mu},$$

and the proof is complete. □

Now we define polygonal approximations of the diffusion paths through successive visits of those boxes. More precisely, if $M_H^\varepsilon=0$, define $X_t^\varepsilon\equiv(0,0,\cdots,0)$ on $[0,1]$; otherwise for $1\leqslant k\leqslant M_H^\varepsilon$, define

$$X_t^\varepsilon=\frac{\tau_k^\varepsilon-t}{\tau_k^\varepsilon-\tau_{k-1}^\varepsilon}\varepsilon\boldsymbol{m}_k^\varepsilon+\frac{t-\tau_{k-1}^\varepsilon}{\tau_k^\varepsilon-\tau_{k-1}^\varepsilon}\varepsilon\boldsymbol{m}_{k-1}^\varepsilon,\quad t\in[\tau_{k-1}^\varepsilon,\tau_k^\varepsilon],$$

and on $[\tau_{M_H^\varepsilon},1]$, define $X_t^\varepsilon\equiv\varepsilon\boldsymbol{m}_{M_H^\varepsilon}^\varepsilon$. Figure 1 illustrates the construction.

Now we have the following convergence result. The proof is developed for arbitrary dimension $N\geqslant 2$, but in the case of $N=2$ the idea is easier to visualize.

Proposition 3.1. *There exists a subsequence $\varepsilon_n\to 0$, such that with probability one, $(X_t^{\varepsilon_n})_{0\leqslant t\leqslant 1}$ converges uniformly to the diffusion paths $(X_t)_{0\leqslant t\leqslant 1}$ on $[0,1]$ as $n\to\infty$.*

Proof. We aim at estimating the following probability

$$P(\sup_{0\leqslant t\leqslant 1}|X_t^\varepsilon-X_t|>\lambda\varepsilon),$$

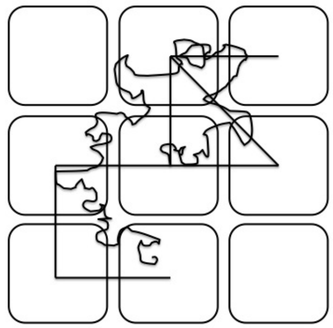

Figure 1: This figure illustrates the construction of the polygonal approximation of the diffusion path. Here the total number of boxes visited by the path in order is 8.

where λ is a large universal constant to be chosen later on. For convenience, we will assume that $\frac{\lambda}{12}$ is a positive integer.

For this purpose, let k be a large integer to be chosen later on (may depend on ε). It follows that

$$\begin{aligned}
&P(\sup_{0\leqslant t\leqslant 1}|X_t^\varepsilon - X_t| > \lambda\varepsilon) \\
\leqslant &\sum_{l=0}^{k} P(\sup_{0\leqslant t\leqslant 1}|X_t^\varepsilon - X_t| > \lambda\varepsilon, M_H^\varepsilon = l) + P(M_H^\varepsilon > k) \\
\leqslant &\sum_{l=0}^{k} P(\bigcup_{j=1}^{l}\{\sup_{\tau_{j-1}^\varepsilon\leqslant t\leqslant\tau_j^\varepsilon}|X_t^\varepsilon - X_t| > \lambda\varepsilon,\ M_H^\varepsilon = l\} \\
&\bigcup\{\sup_{\tau_l^\varepsilon\leqslant t\leqslant 1}|X_t^\varepsilon - X_t| > \lambda\varepsilon,\ M_H^\varepsilon = l\}) + P(M_H^\varepsilon > k)
\end{aligned}$$

$$\begin{aligned}
\leqslant &\sum_{l=0}^{k}[\sum_{j=1}^{l} P(\sup_{\tau_{j-1}^\varepsilon\leqslant t\leqslant\tau_j^\varepsilon}|X_t^\varepsilon - X_t| > \lambda\varepsilon,\ \tau_j^\varepsilon \leqslant 1) \\
&+ P(\sup_{\tau_l^\varepsilon\leqslant t\leqslant 1}|X_t^\varepsilon - X_t| > \lambda\varepsilon,\ M_H^\varepsilon = l)] + P(M_H^\varepsilon > k).
\end{aligned}$$

We first estimate $P(\sup_{\tau_{j-1}^\varepsilon\leqslant t\leqslant\tau_j^\varepsilon}|X_t^\varepsilon - X_t| > \lambda\varepsilon,\ \tau_j^\varepsilon \leqslant 1)$. The idea is the following: the event $\{\sup_{\tau_{j-1}^\varepsilon\leqslant t\leqslant\tau_j^\varepsilon}|X_t^\varepsilon - X_t| > \lambda\varepsilon,\ \tau_j^\varepsilon \leqslant 1\}$ implies that after time τ_{j-1}^ε, the process must have travelled through many narrow tunnels and spread far away from $H^\varepsilon_{\boldsymbol{m}^\varepsilon_{j-1}}$ by many boxes before visiting another box. Define σ_0 to be the first time after τ_{j-1}^ε that the process arrives at the entrance of some narrow tunnel which is far away from $H^{\varepsilon}_{\boldsymbol{m}^{\varepsilon}_{j-1}}$ with distance at least $\frac{\lambda}{6}\varepsilon$ without hitting any other boxes. For $1 \leqslant L \leqslant \lambda/12$, define σ_L to be the first time after σ_{L-1} that the process travels through a narrow tunnel without hitting any boxes other than $H^{\varepsilon}_{\boldsymbol{m}^{\varepsilon}_{j-1}}$ (define $\sigma_0 = \infty$ if there is no such arrival and $\sigma_L = \infty$ if there is no such travel through). It is easy to see that

$$\{\sup_{\tau^{\varepsilon}_{j-1} \leqslant t \leqslant \tau^{\varepsilon}_j} |X^{\varepsilon}_t - X_t| > \lambda\varepsilon,\ \tau^{\varepsilon}_j \leqslant 1\} \subset \{\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty\}.$$

Thus it suffices to estimate $P(\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty)$. This can be done by using the strong Markov property and a quantitative estimate for the Poisson kernel of some nice domain in [1]. In fact, by the strong Markov property,

$$\begin{aligned}
&P(\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty) \\
=&E[P(\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty | \mathcal{F}^X_{\sigma_{\lambda/12-1}}),\ \sigma_{\lambda/12-1} < \infty] \\
=&E[P^{X_{\sigma_{\lambda/12-1}}(\omega)}(\{\omega' :\ \omega' \in T(\omega)\}),\ \sigma_0(\omega) < \sigma_1(\omega) < \cdots < \sigma_{\lambda/12-1}(\omega) < \infty],
\end{aligned}$$

where $T(\omega)$ denotes the set of sample paths ω' of the diffusion process starting at $X_{\sigma_{\lambda/12-1}}(\omega)$ such that the first time of traveling through a narrow tunnel without hitting any boxes is finite. By the assumptions on the generating vector fields, the generator L and those small boxes H^{ε}_z verify the conditions of Lemma 2.6 in [1]. It follows from the lemma that the Poisson kernel $H(x, d\eta)$ of any small box has a continuous density $h(x, \eta)$ with respect to the normalized surface measure $d\eta$. Moreover, there are universal constants (in particular, not depending on ε) $K_0, \nu_0 > 0$, such that

$$|h(x, \eta)| \leqslant K_0 \cdot \mathrm{dist}(x, \partial G)/\mathrm{dist}(x, \eta)^{\nu_0},$$

for any x in the box and η on the boundary. Since traveling through narrow tunnels implies escaping through narrow windows of the boundary of some associated domain, it follows that on $\{\omega :\ \sigma_{\lambda/12-1}(\omega) < \infty\}$,

$$P^{X_{\sigma_{\lambda/12-1}}(\omega)}(\{\omega' :\ \omega' \in T(\omega)\}) \leqslant K(\varepsilon - \varepsilon^{\mu} - 2\varepsilon^{2\mu})^{1-\nu_0} \cdot \frac{\varepsilon^{\mu}}{\varepsilon},$$

for some universal constant $K > 0$. Now it is clear that if we choose μ to be universal and far greater than ν_0, then on $\{\omega :\ \sigma_{\lambda/12-1}(\omega) < \infty\}$, we have

$$P^{X_{\sigma_{\lambda/12-1}}(\omega)}(\{\omega' :\ \omega' \in T(\omega)\}) \leqslant K\varepsilon^{\mu-\nu_0},$$

for some universal constant $K > 0$. Therefore,

$$P(\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty) \leqslant K\varepsilon^{\mu-\nu_0} P(\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12-1} < \infty).$$

By induction, it is immediate that

$$\begin{aligned}
P(\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty) &\leqslant K^{\frac{\lambda}{12}} \varepsilon^{\frac{\lambda}{12}(\mu-\nu_0)} P(\sigma_0 < \infty) \\
&\leqslant K^{\frac{\lambda}{12}} \varepsilon^{\frac{\lambda}{12}(\mu-\nu_0)}.
\end{aligned}$$

Therefore, we arrive at

$$P(\sup_{\tau_{j-1}^{\varepsilon}\leqslant t\leqslant\tau_j^{\varepsilon}}|X_t^{\varepsilon}-X_t|>\lambda\varepsilon,\ \tau_j^{\varepsilon}\leqslant 1)\leqslant K^{\frac{\lambda}{12}}\varepsilon^{\frac{\lambda}{12}(\mu-\nu_0)}.$$

The estimate of $P(\sup_{\tau_l^{\varepsilon}\leqslant t\leqslant 1}|X_t^{\varepsilon}-X_t|>\lambda\varepsilon,\ M_H^{\varepsilon}=l)$ is exactly the same as above since on $\{M_H^{\varepsilon}=l\}$, there will be no visit of boxes other than $H^{\varepsilon}_{\boldsymbol{m}_l^{\varepsilon}}$ during $[\tau_l^{\varepsilon},1]$.

Now consider $P(M_H^{\varepsilon}>k)$. By Lemma 3.1, if $k>\frac{2C}{\varepsilon^{\mu}}$,

$$\begin{aligned}P(M_H^{\varepsilon}>k) &= \sum_{l=k+1}^{\infty}P(M_H^{\varepsilon}=l)\\ &\leqslant \sum_{l=k+1}^{\infty}4Nle^{-\frac{\varepsilon^{2\mu}l}{8NdC^2}}\\ &\leqslant \frac{4Ne^{-k\tilde{C}\varepsilon^{2\mu}}}{(1-e^{-\tilde{C}\varepsilon^{2\mu}})^2}+\frac{4Nke^{-k\tilde{C}\varepsilon^{2\mu}}}{1-e^{-\tilde{C}\varepsilon^{2\mu}}},\end{aligned}$$

where $\tilde{C}=\frac{1}{8NdC^2}$. Choose a universal constant $\gamma>>2\mu$, and let $k=[\frac{1}{\varepsilon^{\gamma}}]$ (when ε is small, the condition $k>\frac{2C}{\varepsilon^{\mu}}$ in Lemma 3.1 is satisfied). It follows that

$$P(M_H^{\varepsilon}>k)\leqslant C'(\frac{e^{-\frac{\tilde{C}}{\varepsilon^{\gamma-2\mu}}}}{(1-e^{-\tilde{C}\varepsilon^{2\mu}})^2}+\frac{1}{\varepsilon^{\gamma}}\frac{e^{-\frac{\tilde{C}}{\varepsilon^{\gamma-2\mu}}}}{1-e^{-\tilde{C}\varepsilon^{2\mu}}}),$$

where C' is a positive universal constant.

Combining with the estimates before, we arrive at

$$\begin{aligned}&P(\sup_{0\leqslant t\leqslant 1}|X_t^{\varepsilon}-X_t|>\lambda\varepsilon)\\ \leqslant &C'(K^{\frac{\lambda}{12}}\varepsilon^{\frac{\lambda}{12}(\mu-\nu_0)-2\gamma}+\frac{e^{-\frac{\tilde{C}}{\varepsilon^{\gamma-2\mu}}}}{(1-e^{-\tilde{C}\varepsilon^{2\mu}})^2}+\frac{1}{\varepsilon^{\gamma}}\frac{e^{-\frac{\tilde{C}}{\varepsilon^{\gamma-2\mu}}}}{1-e^{-\tilde{C}\varepsilon^{2\mu}}}).\end{aligned}\tag{3.1}$$

Finally, choose a positive universal integer λ such that

$$\lambda>\frac{24\gamma+24}{\mu-\nu_0}$$

and $\frac{\lambda}{12}$ is a positive integer. By taking $\varepsilon_n=1/n$, we have

$$\sum_{n=1}^{\infty}P(\sup_{0\leqslant t\leqslant 1}|X_t^{\varepsilon_n}-X_t|>\lambda\varepsilon_n)<\infty.$$

Borel-Cantelli's lemma then yields the desired result.

Now the proof is complete. □

Let $\mu' > \mu$ be another universal constant. Define $(V_z^\varepsilon, \zeta_k^\varepsilon, \boldsymbol{n}_k^\varepsilon, M_V^\varepsilon, \widetilde{X}^\varepsilon)$ in the same way as $(H_z^\varepsilon, \tau_k^\varepsilon, \boldsymbol{m}_k^\varepsilon, M_H^\varepsilon, X^\varepsilon)$ only with μ replaced by μ', then Proposition 3.1 is also true for $\widetilde{X}^\varepsilon$ (with $\varepsilon_n = \frac{1}{n}$ as in the proof of Proposition 3.1).

To complete the proof, it remains to trace the diffusion paths via extended Stratonovich's signatures "between" smaller boxes H_z^ε and larger boxes V_z^ε, and prove a squeeze theorem so that we are able to pass to the same limit X_t.

3.2 Constructing differential 1-forms and using extended Stratonovich's signatures

To trace the diffusion paths by using extended Stratonovich's signatures, we first need to construct suitable compactly supported differential 1-forms, such that the Stratonovich's integral of any such 1-form ϕ along the diffusion paths over the duration of visit of supp ϕ is nonzero with probability one.

To this end, it suffices to construct a suitable differential 1-form ϕ on $\mathbb{R}^N$ with compact support such that the family of vector fields on $\mathbb{R}^{N+1}$:

$$\{\begin{pmatrix} V_1 \\ \phi \cdot V_1 \end{pmatrix}, \cdots, \begin{pmatrix} V_d \\ \phi \cdot V_d \end{pmatrix}; \begin{pmatrix} V_0 \\ \phi \cdot V_0 \end{pmatrix}\}$$

satisfies Hörmander's condition on (supp ϕ) $\times \mathbb{R}^1$ so that the generator of the diffusion process on $\mathbb{R}^{N+1}$ defined by

$$\begin{cases} dX_t = \sum_{\alpha=1}^d V_\alpha(X_t) \circ dW_t^\alpha + V_0(X_t)dt, \\ dX_t^{N+1} = \phi(X_t) \circ dX_t, \end{cases}$$

is hypoelliptic on (supp ϕ) $\times \mathbb{R}^1$, which ensures the existence of smooth probability densities of certain Wiener functionals. Here and thereafter we use the geometric notation for convenience (so V_α is a regarded as a column vector and ϕ is regarded as a row vector in $\mathbb{R}^N$). In fact, if this is possible, then we can proceed in the same way as Lemma 2.1, Lemma 2.2 and Lemma 2.3 in [6] to show that the Stratonovich's integral of ϕ along the diffusion paths over the duration of visit of supp ϕ is nonzero with probability one, since starting from this point the proof relies only on the strong Markov property and again the results in [1], which hold true from our assumptions on the generating vector fields $\{V_1, \cdots, V_d; V_0\}$.

Now to make it more precise, for $z \in \mathbb{Z}^N$ and $\varepsilon > 0$, we are interested in constructing a differential 1-form ϕ_z^ε such that

$$H_z^\varepsilon \subset (\text{supp } \phi_z^\varepsilon)^\circ \subset \text{supp } \phi_z^\varepsilon \subset (V_z^\varepsilon)^\circ,$$

and ϕ_z^ε has the property mentioned before.

The following result gives the desired construction.

Proposition 3.2. *Assume that the family of vector field $\{V_1, \cdots, V_d; V_0\}$ satisfies Hörmander's condition at every point x in $\mathbb{R}^N$. Let G be a bounded domain in $\mathbb{R}^N$ and W be an open subdomain of G such that*

$$W \subset\subset G.$$

Let $\eta \in C_0^\infty(\mathbb{R}^N)$ be a cut-off function of $\overline{W}$, that is, $0 \leqslant \eta \leqslant 1$, $\eta \equiv 1$ on $\overline{W}$ and $\eta = 0$ outside a small neighborhood of $\overline{W}$. Then there exists $\Lambda > 0$, such that for any $\xi \in \mathbb{R}^N$ with $|\xi| > \Lambda$, if we define the differential 1-form ϕ on $\mathbb{R}^N$ by

$$\phi(x) = \eta(x) e^{-\frac{1}{2}|x-\xi|^2}(dx^1 + \cdots + dx^N), \tag{3.2}$$

and define the vector field $\widetilde{V_\alpha}$ on $\mathbb{R}^{N+1}$ (independent of x^{N+1}) by

$$\widetilde{V}_\alpha = \begin{pmatrix} V_\alpha \\ \phi \cdot V_\alpha \end{pmatrix}, \quad \alpha = 0, 1, \cdots d, \tag{3.3}$$

then the family of vector fields

$$\{\widetilde{V}_1, \cdots, \widetilde{V}_d; \widetilde{V}_0\}$$

satisfies Hörmander's condition at every point on $(\mathrm{supp}\ \phi) \times \mathbb{R}$. In other words, the differential operator $\widetilde{L}$ on $\mathbb{R}^{N+1}$ defined by

$$\widetilde{L} = \frac{1}{2}\sum_{\alpha=1}^{d} \widetilde{V}_\alpha^2 + \widetilde{V}_0 \tag{3.4}$$

is hypoelliptic on $(\mathrm{supp}\ \phi)\times\mathbb{R}$.

Proof. For a differential 1-form ϕ on $\mathbb{R}^N$ defined by (3.2), define the vector fields

$$\{\widetilde{V}_1, \cdots, \widetilde{V}_d; \widetilde{V}_0\}$$

on $\mathbb{R}^{N+1}$ by (3.3). Note that supp ϕ is independent of ξ, which will be denoted by K.

Let

$$\begin{aligned} \Theta_1 &= \{1, 2, \cdots, d\}; \\ \Theta_n &= \{(\alpha_1, \cdots, \alpha_n) : \ \alpha_i = 0, 1, \cdots, d\}, \ n \geqslant 2; \\ \Theta &= \bigcup_{n=1}^{\infty} \Theta_n. \end{aligned}$$

For $\theta = (\theta_1, \cdots, \theta_n) \in \Theta_n$, denote $|\theta| = n$, and we use the notation $V_{[\theta]}$ ($\widetilde{V}_{[\theta]}$, respectively) to denote $[V_{\theta_1}, [V_{\theta_2}, \cdots, [V_{\theta_{n-1}}, V_{\theta_n}]]]$ ($[\widetilde{V}_{\theta_1}, [\widetilde{V}_{\theta_2}, \cdots, [\widetilde{V}_{\theta_{n-1}}, \widetilde{V}_{\theta_n}]]]$, respectively).

We first prove that for any $\theta \in \Theta$, $\widetilde{V}_{[\theta]}$ can be written as

$$\widetilde{V}_{[\theta]} = \begin{pmatrix} V_{[\theta]} \\ g_{[\theta]} + \phi \cdot V_{[\theta]} \end{pmatrix}$$

for some $g_{[\theta]} \in C_b^\infty(\mathbb{R}^{N+1})$ independent of x^{N+1}. In fact, when $\theta \in \Theta_1$, it is just the definition of $\widetilde{V}_{[\theta]}$. Assume that it is true for any $\theta \in \Theta_n$. Let $\theta \in \Theta_{n+1}$, then there exists some $0 \leqslant \alpha \leqslant d$ and $\theta' \in \Theta_n$, such that

$$V_{[\theta]} = [V_\alpha, V_{[\theta']}], \ \widetilde{V}_{[\theta]} = [\widetilde{V}_\alpha, \widetilde{V}_{[\theta']}].$$

By the induction hypothesis, we have

$$\begin{aligned}
\widetilde{V}_{[\theta]} &= \left[\begin{pmatrix} V_\alpha \\ \phi \cdot V_\alpha \end{pmatrix}, \begin{pmatrix} V_{[\theta']} \\ g_{[\theta']} + \phi \cdot V_{[\theta']} \end{pmatrix}\right] \\
&= \begin{pmatrix} [V_\alpha, V_{[\theta']}] \\ \nabla^N (g_{[\theta']} + \phi \cdot V_{[\theta']}) \cdot V_\alpha - \nabla^N(\phi \cdot V_\alpha) \cdot V_{[\theta']} \end{pmatrix} \\
&= \begin{pmatrix} V_{[\theta]} \\ g_{[\theta]} + \phi \cdot V_{[\theta]} \end{pmatrix},
\end{aligned}$$

where

$$g_{[\theta]} = \nabla^N g_{[\theta']} \cdot V_\alpha + V_{[\theta']}^T \cdot \nabla^N \phi^T \cdot V_\alpha - V_\alpha^T \cdot \nabla^N \phi^T \cdot V_{[\theta']} \in C_b^\infty(\mathbb{R}^{N+1}),$$

which is independent of x^{N+1}. Here ∇^N denotes the gradient operator with respect to $x^{(N)} = (x^1, \cdots, x^N)$ and $(\cdot)^T$ denotes the transpose operator.

Now we are going to prove the result by a compactness argument.

A key observation is that for any fixed $\theta \in \Theta$, let $g_{[\theta]} \in C_b^\infty(\mathbb{R}^{N+1})$ be such that

$$\widetilde{V}_{[\theta]} = \begin{pmatrix} V_{[\theta]} \\ g_{[\theta]} + \phi \cdot V_{[\theta]} \end{pmatrix}$$

as in the previous discussion, then $g_{[\theta]}$ is of the form

$$g_{[\theta]}(x) = p_{[\theta]}(\xi; x) e^{-\frac{1}{2}|x-\xi|^2},$$

where $p_{[\theta]}(\xi; x)$ is a polynomial in $\xi = (\xi^1, \cdots, \xi^N)$ with C_b^∞ coefficients depending only on $x^{(N)}$. Here the degree of $p_{[\theta]}$ is at most $|\theta| - 1$. In other words,

$$p_{[\theta]}(\xi; x) = \sum_{j=0}^{|\theta|-1} \sum_{|\boldsymbol{\alpha}|=j} c_{\boldsymbol{\alpha}}(x^{(N)}) \xi^{\boldsymbol{\alpha}}.$$

Fix $x_0 = (x_0^{(N)}, x_0^{N+1}) \in K^\circ \times \mathbb{R}^1$, where $x_0^{(N)} = (x_0^1, \cdots, x_0^N) \in \mathbb{R}^N$. By the hypoellipticity of L and continuity, there exists a neighborhood $U \subset K^\circ$ of $x_0^{(N)}$ and $\theta^{(1)}, \cdots, \theta^{(N)} \in \Theta$, such that for any $x^{(N)} \in U$,

$$\{V_{[\theta^{(1)}]}(x^{(N)}), \cdots, V_{[\theta^{(N)}]}(x^{(N)})\}$$

constitutes a basis of $\mathbb{R}^N$. It follows that for any $x \in U \times \mathbb{R}^1$, the family of vectors in $\mathbb{R}^{N+1}$

$$\{\widetilde{V}_{[\theta^{(1)}]}(x), \cdots, \widetilde{V}_{[\theta^{(N)}]}(x)\}$$

generate an N-dimensional subspace of $\mathbb{R}^{N+1}$. Let $M = \max\{|\theta^{(1)}|, \cdots, |\theta^{(N)}|\}$. Again by the assumptions on L and continuity, it is possible to choose $\theta \in \Theta$ with $|\theta| > M$, such that

$$\text{degree } (p_{[\theta]}) > M \tag{3.5}$$

in some compact neighborhood $\overline{U_0} \subset U$ of $x_0^{(N)}$. In particular, the choice of θ and U_0 is independent of the ξ since the coefficents of $p_{[\theta]}$ are functions of x only.

Now we are going to show that there exists $\Lambda > 0$, such that when $\xi \in \mathbb{R}^N$ with $|\xi| > \Lambda$, the vector field $\widetilde{V}_{[\theta]}$ cannot be generated by $\{\widetilde{V}_{[\theta^{(1)}]}, \cdots, \widetilde{V}_{[\theta^{(N)}]}\}$ in $U_0 \times \mathbb{R}^1$, so that

$$\dim \operatorname{Span}\{\widetilde{V}_{[\theta^{(1)}]}, \cdots, \widetilde{V}_{[\theta^{(N)}]}, \widetilde{V}_{[\theta]}\} = N + 1,$$

which yields the hypoellipticity of $\widetilde{L}$ defined by 3.4 in $U_0 \times \mathbb{R}^1$.

To prove this, first notice that there exists $\lambda^i(x^{(N)}) \in C_b^\infty(U_0)$, such that

$$V_{[\theta]}(x^{(N)}) = \sum_{i=1}^{N} \lambda^i(x^{(N)}) V_{[\theta^{(i)}]}(x^{(N)}), \quad \text{for } x^{(N)} \in U_0.$$

Moreover, from (3.5) it is easy to see that there exists $\Lambda > 0$, such that

$$p_{[\theta]}(\xi; x) \neq \sum_{i=1}^{N} \lambda^i(x^{(N)}) p_{[\theta^{(i)}]}(\xi; x) \tag{3.6}$$

for $\xi \in \mathbb{R}^N$ with $|\xi| > \Lambda$ and $x \in U_0 \times \mathbb{R}^1$. If

$$\widetilde{V}_{[\theta]}(x_1) \in \operatorname{Span}\{\widetilde{V}_{[\theta^{(1)}]}(x_1), \cdots, \widetilde{V}_{[\theta^{(N)}]}(x_1)\}$$

for some $x_1 \in U_0 \times \mathbb{R}^1$, then we must have

$$\widetilde{V}_{[\theta]}(x_1) = \sum_{i=1}^{N} \lambda^i(x_1^{(N)}) \widetilde{V}_{[\theta^{(i)}]}(x_1).$$

It follows from simple calculation that

$$g_{[\theta]}(x_1) = \sum_{i=1}^{N} \lambda^i(x_1^{(N)}) g_{[\theta^{(i)}]}(x_1). \tag{3.7}$$

This is a contradiction to (3.6) when $|\xi| > \Lambda$. Therefore, $\widetilde{V}_{[\theta]}$ cannot be generated by $\{\widetilde{V}_{[\theta^{(1)}]}, \cdots, \widetilde{V}_{[\theta^{(N)}]}\}$ in $U_0 \times \mathbb{R}^1$ if we choose ξ with $|\xi| > \Lambda$ in the definition of ϕ.

The case when $x_0 \in \partial K \times \mathbb{R}^1$ can be proved in the same way by replacing U_0 with $U_0 \cap K$.

Finally, combining with the above local results and by the compactness of K, we are able to choose $\Lambda > 0$ (depending on K), such that for any $\xi \in \mathbb{R}^N$ with $|\xi| > \Lambda$, the differential operator $\widetilde{L}$ is hypoelliptic on $K \times \mathbb{R}^1$.

Now the proof is complete. $\square$

For $z \in \mathbb{Z}^N$ and $\varepsilon > 0$, by taking $W = H_z^\varepsilon$ and $G = V_z^\varepsilon$, we can construct a differential 1-form ϕ_z^ε supported in G according to Lemma 3.2 (just take some fixed admissible $\xi \in \mathbb{R}^N$ as in the lemma). By proceeding in the same way as in [6], we conclude that the Stratonovich's integral of ϕ_z^ε along the diffusion paths over the duration of visit of supp ϕ_z^ε is nonzero with probability one.

Now we are going to construct extended Stratonovich's signatures to trace the original diffusion paths by using these differential 1-forms ϕ_z^ε.

We first define extended Stratonovich's signatures formally.

For smooth differential forms $\psi^1, \cdots, \psi^k$ on $\mathbb{R}^N$, the iterated Stratonovich's integral $[\psi^1, \cdots, \psi^k]_{s,t}$ $(0 \leqslant s < t \leqslant 1)$ defined inductively by

$$[\psi^1, \cdots, \psi^k]_{s,t} = \int_{s<u<t} [\psi^1, \cdots, \psi^{k-1}]_{s,u} \psi^k(\circ dX_u),$$

where

$$[\psi^1]_{s,t} = \sum_{i=1}^N \int_{s<u<t} \psi_i^1(X_u) \circ dX_u^i,$$

is called an *extended Stratonovich's signature* of the diffusion process X_t.

The following lemma allows us to use extended Stratonovich's signatures for our study. The case of Brownian motion was proved in [6], but we can easily adopt the proof to the our case without changing anything (in fact, the proof does not rely on probabilistic features, but only on paths). Recall that $\mathcal{G}_1$ is the completion of the σ-algebra generated by the Stratonovich's signature of X_t over $[0, 1]$.

Lemma 3.2. *If $\psi^1, \cdots, \psi^k$ are smooth differential 1-forms on $\mathbb{R}^N$ with compact supports, then*

$$[\psi^1, \cdots, \psi^k]_{0,1}$$

is $\mathcal{G}_1$-measurable.

Proof. See [6], Lemma 1.3. □

For $m \geqslant 0$, let

$$\mathcal{W}_m = \{(z_0 = (0, \cdots, 0), z_1, \cdots, z_m) : \ z_i \in \mathbb{Z}^N, \ z_i \neq z_{i-1}, \ i = 1, 2, \cdots, m\}.$$

An element $(z_0, z_1, \cdots, z_m) \in \mathcal{W}_m$ is called an admissible word of length $m+1$. For $\varepsilon > 0$, define the $\mathcal{G}_1$-measurable random variable M^ε to be the supremum of those $m \geqslant 0$ such that

$$[\phi_{z_0}^\varepsilon, \phi_{z_1}^\varepsilon, \cdots, \phi_{z_m}^\varepsilon]_{0,1} \neq 0$$

for some admissible word $(z_0, z_1, \cdots, z_m) \in \mathcal{W}_m$. It follows that $M_H^\varepsilon \leqslant M^\varepsilon \leqslant M_V^\varepsilon$ for almost surely. For $m \geqslant 0$, $(z_0, z_1, \cdots, z_m) \in \mathcal{W}_m$, define

$$A_{m;(z_0, z_1, \cdots, z_m)}^\varepsilon = \{\omega : \ M^\varepsilon = m, \ [\phi_{z_0}^\varepsilon, \phi_{z_1}^\varepsilon, \cdots, \phi_{z_m}^\varepsilon]_{0,1} \neq 0\},$$

then $\{A^{\varepsilon}_{m;(z_0,z_1,\cdots,z_m)}: \ m \geqslant 0, \ (z_0, z_1, \cdots, z_m) \in \mathcal{W}_m\}$ are mutually disjoint $\mathcal{G}_1$-measurable sets whose union is the whole space Ω. See [6] for a more detailed discussion.

Let

$$\mathcal{W} = \bigcup_{m=0}^{\infty} \mathcal{W}_m$$

be the space of admissible words. For $\varepsilon > 0$, define the mapping $\mathcal{Y}^{\varepsilon}: \ \Omega \to \mathcal{W}$,

$$\mathcal{Y}^{\varepsilon}(\omega) = (z_0, z_1, \cdots, z_m),$$

if $(z_0, z_1, \cdots, z_m)$ is such that

$$\omega \in A^{\varepsilon}_{m;(z_0,z_1,\cdots,z_m)}.$$

It follows that $\mathcal{Y}^{\varepsilon}$ is well-defined and $\mathcal{G}_1$-measurable. Intuitively, $\mathcal{Y}^{\varepsilon}$ is the maximal admissible word such that the associated extended Stratonovich's signature is nonzero. It is natural to believe that $\mathcal{Y}^{\varepsilon}$ records a reasonable amount of information of the diffusion paths and as $\varepsilon \to 0$, it might be possible to recover the diffusion paths.

3.3 Completing the proof: a squeeze theorem for convergence in trajectory

In Section 2, we defined piecewise linear trajectories (P.L.T.s), parametrization of a P.L.T., and introduced the concept of convergence in trajectory. In this section, we are going to show that if $\mathcal{Y}^{\varepsilon}$ is regarded as a P.L.T. in $\mathbb{Z}^N$, then by taking $\varepsilon_n = \frac{1}{n}$, with probability one, $\varepsilon_n \cdot \mathcal{Y}^{\varepsilon_n}$ converges in trajectory to $(X_t)_{0 \leqslant t \leqslant 1}$, which completes the proof of our main theorem.

Recall that a P.L.T. $\mathcal{T}$ is essentially a finite sequence of points in $\mathbb{R}^N$ (not necessarily all distinct).

Definition 3.1. For a P.L.T. $\mathcal{T}$, $\mathcal{T}^-$ is denoted as the new P.L.T. by removing the last point of $\mathcal{T}$. Let $\mathcal{T}_1, \mathcal{T}_2$ be two P.L.T.s. $\mathcal{T}_1$ is called a *sub*-P.L.T. of $\mathcal{T}_2$ (denoted by $\mathcal{T}_1 \prec \mathcal{T}_2$) if $\mathcal{T}_1$ is a subsequence of $\mathcal{T}_2$.

By the convergence result and the construction of ϕ^{ε}_z in the last two subsections, if we denote $\mathcal{X}^{\varepsilon}$ (respectively, $\widetilde{\mathcal{X}}^{\varepsilon}$) as the associated P.L.T. of the piecewise linear path X^{ε} (respectively, $\widetilde{X}^{\varepsilon}$), then it is immediate that

$$(\mathcal{X}^{\varepsilon})^- \prec \varepsilon \cdot \mathcal{Y}^{\varepsilon} \prec (\widetilde{\mathcal{X}}^{\varepsilon})^-,$$

with probability one, and $\mathcal{X}^{\varepsilon_n}$ and $\widetilde{\mathcal{X}}^{\varepsilon_n}$ both converges in trajectory to $(X_t)_{0 \leqslant t \leqslant 1}$. Therefore, it is natural to claim a certain kind of squeeze theorem for convergence in trajectory so we may conclude that $\mathcal{Y}^{\varepsilon_n}$ also converges in trajectory to $(X_t)_{0 \leqslant t \leqslant 1}$ with probability one.

The following result is a squeeze theorem for convergence in trajectory we are looking for, which is sufficient for our use.

Proposition 3.3. *Assume that* $\{\mathcal{T}_1^{(n)}\}$, $\{\mathcal{T}_2^{(n)}\}$ *are two sequence of P.L.T.s such that:*

(1) the first points of $\mathcal{T}_1^{(n)}$ *and* $\mathcal{T}_2^{(n)}$ *are identical;*

(2) the last two points of $\mathcal{T}_i^{(n)}$ *are identical* $(i=1,2)$*.*

Let $\sigma_i^{(n)}$ *be a parametrization of* $\mathcal{T}_i^{(n)}$ $(i=1,2)$ *such that the partition points in* $\sigma_1^{(n)}$ *belong to the partition points in* $\sigma_2^{(n)}$ *and for any* $t<1$ *in* $\sigma_1^{(n)}$,

$$\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) = \mathcal{T}_2^{(n)}(t|\sigma_2^{(n)}).$$

(This assumption implies that $(\mathcal{T}_1^{(n)})^- \prec (\mathcal{T}_2^{(n)})^-$*.) Let* $\{\mathcal{T}^{(n)}\}$ *be a sequence of P.L.T.s such that*

$$(\mathcal{T}_1^{(n)})^- \prec \mathcal{T}^{(n)} \prec (\mathcal{T}_2^{(n)})^-,$$

and $(\gamma_t)_{0\leqslant t\leqslant 1}$ *be a continuous path in* $\mathbb{R}^N$ *such that*

$$\lim_{n\to\infty}\sup_{0\leqslant t\leqslant 1}|\mathcal{T}_i^{(n)}(t|\sigma_i^{(n)}) - \gamma_t| = 0, \quad i=1,2.$$

Then we can choose a parametrization $\sigma^{(n)}$ *of* $\mathcal{T}^{(n)}$*, such that*

$$\lim_{n\to\infty}\sup_{0\leqslant t\leqslant 1}|\mathcal{T}^{(n)}(t|\sigma^{(n)}) - \gamma_t| = 0.$$

In particular, $\mathcal{T}^{(n)}$ *converges in trajectory to* $(\gamma_t)_{0\leqslant t\leqslant 1}$*.*

Proof. For any $\varepsilon>0$, there exists $n_0>0$, such that for any $n>n_0$,

$$\sup_{0\leqslant t\leqslant 1}|\mathcal{T}_i^{(n)}(t|\sigma_i^{(n)}) - \gamma_t| < \varepsilon, \quad i=1,2. \tag{3.8}$$

On the other hand, it is obvious that we are able to construct a parametrization $\sigma^{(n)}$ of $\mathcal{T}^{(n)}$, such that:

(1) the partition points of $\sigma_1^{(n)}$ belong to the partition points in $\sigma^{(n)}$, and for any $t<1$ in $\sigma_1^{(n)}$,

$$\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) = \mathcal{T}^{(n)}(t|\sigma^{(n)});$$

(2) the partition points in $\sigma^{(n)}$ belong to the partition points in $\sigma_2^{(n)}$, and for any $t<1$ in $\sigma^{(n)}$,

$$\mathcal{T}^{(n)}(t|\sigma^{(n)}) = \mathcal{T}_2^{(n)}(t|\sigma_2^{(n)}).$$

Let t_n be the largest time spot in $\sigma_1^{(n)}$ such that $t_n<1$. Let $u_n<v_n$ be any consecutive time spots in $\sigma^{(n)}$, then on $[u_n,v_n]$ both $\mathcal{T}_1^{(n)}(\cdot|\sigma_1^{(n)})$ and $\mathcal{T}^{(n)}(\cdot|\sigma^{(n)})$ are linear. Therefore, by an elementary result on the comparison for linear paths, we have

$$\begin{aligned}&\sup_{u_n\leqslant t\leqslant v_n}|\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(t|\sigma^{(n)})| \\ \leqslant{}& \max\{|\mathcal{T}_1^{(n)}(u_n|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(u_n|\sigma^{(n)})|,\ |\mathcal{T}_1^{(n)}(v_n|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(v_n|\sigma^{(n)})|\}.\end{aligned} \tag{3.9}$$

If $v_n \leqslant t_n$, then

$$\mathcal{T}^{(n)}(u_n|\sigma^{(n)}) = \mathcal{T}_2^{(n)}(u_n|\sigma_2^{(n)}),\ \mathcal{T}^{(n)}(v_n|\sigma^{(n)}) = \mathcal{T}_2^{(n)}(v_n|\sigma_2^{(n)}).$$

It follows from (3.8) and (3.9) that

$$\sup_{u_n \leqslant t \leqslant v_n} |\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(t|\sigma^{(n)})| < 2\varepsilon, \quad n > n_0.$$

If $u_n \geqslant t_n$, since the last two points of $\mathcal{T}_1^{(n)}$ are identical (denoted by $x^{(n)}$), it follows that

$$\begin{aligned} &\sup_{u_n \leqslant t \leqslant v_n} |\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(t|\sigma^{(n)})| \\ &\leqslant \max\{|x^{(n)} - \mathcal{T}^{(n)}(u_n|\sigma^{(n)}),\ |x^{(n)} - \mathcal{T}^{(n)}(v_n|\sigma^{(n)})||\}. \end{aligned}$$

Obviously

$$\mathcal{T}^{(n)}(u_n|\sigma^{(n)}) = \mathcal{T}_2^{(n)}(u_n|\sigma_2^{(n)}).$$

But it may not be true for v_n since it is possible that $v_n = 1$. However, since $\mathcal{T}^{(n)} \prec \mathcal{T}_2^{(n)}$, there exists some $w_n > u_n$, such that

$$\mathcal{T}^{(n)}(v_n|\sigma^{(n)}) = \mathcal{T}_2^{(n)}(w_n|\sigma_2^{(n)}), \quad (w_n = v_n \text{ if } v_n < 1).$$

Due to the fact that $\mathcal{T}_1^{(n)} \equiv x^{(n)}$ on $[t_n, 1]$, we arrive again at

$$\sup_{u_n \leqslant t \leqslant v_n} |\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(t|\sigma^{(n)})| < 2\varepsilon, \quad n > n_0.$$

Consequently,

$$\sup_{0 \leqslant t \leqslant 1} |\mathcal{T}_1^{(n)}(t|\sigma_1^{(n)}) - \mathcal{T}^{(n)}(t|\sigma^{(n)})| < 2\varepsilon, \quad n > n_0.$$

It follows that

$$\lim_{n\to\infty} \sup_{0 \leqslant t \leqslant 1} |\mathcal{T}^{(n)}(t|\sigma^{(n)}) - \gamma_t| = 0,$$

and in particular, $\mathcal{T}^{(n)}$ converges in trajectory to $(\gamma_t)_{0 \leqslant t \leqslant 1}$.

Now the proof is complete. □

In order to apply Proposition 3.3, we are going to modify $\widetilde{\mathcal{X}}^\varepsilon$ and choose a suitable parametrization based on the one for $\widetilde{\mathcal{X}}^\varepsilon$ specified in Subsection 3.1, which is chosen according to the successive visit time of larger boxes for the diffusion paths (excluding revisit of the same box before visiting other boxes), so that the assumptions of Proposition 3.3 are all verified.

The method is the following. By using the notation in Section 3.1, if $(\zeta_k^\varepsilon, \tau_l^\varepsilon, \zeta_{k+1}^\varepsilon)$ is such that

$$\zeta_k^\varepsilon < \tau_l^\varepsilon < \zeta_{k+1}^\varepsilon \leqslant 1,$$

then we modify the linear path $\widetilde{X}^\varepsilon$ on $[\zeta_k^\varepsilon, \zeta_{k+1}^\varepsilon]$ to a new path such that it does not move during $[\zeta_k^\varepsilon, \tau_l^\varepsilon]$ and goes directly from its initial position at $t = \tau_l^\varepsilon$ to $\widetilde{X}^\varepsilon_{\zeta_{k+1}^\varepsilon}$ at $t = \zeta_{k+1}^\varepsilon$ with constant velocity. If

$$\zeta_k^\varepsilon < \tau_l^\varepsilon < 1 < \zeta_{k+1}^\varepsilon,$$

then we modify the linear path $\widetilde{X}^\varepsilon$ on $[\zeta_k^\varepsilon, 1]$ (in fact, $\widetilde{X}^\varepsilon$ remains still on $[\zeta_k^\varepsilon, 1]$) to a path such that during $[\zeta_k^\varepsilon, \tau_l^\varepsilon]$ and $[\tau_l^\varepsilon, 1]$ it remains still (equals $\widetilde{X}^\varepsilon_{\zeta_k^\varepsilon}$). It seems that such modification is trivial and does not change anything, but it does make a slight difference if we are using the associated P.L.T.. Let $\widehat{X}^\varepsilon$ be the modified piecewise linear path of $\widetilde{X}^\varepsilon$ and let $\widehat{\mathcal{X}}^\varepsilon$ be the associated P.L.T. of $\widehat{X}^\varepsilon$. If we can prove that $\widehat{X}^{\varepsilon_n}$ converges uniformly to $(X_t)_{0 \leqslant t \leqslant 1}$ with probability one, then all the assumptions in Proposition 3.3 for the triple sequence $\{(\mathcal{X}^{\varepsilon_n}, \varepsilon_n \cdot \mathcal{Y}^{\varepsilon_n}, \widehat{\mathcal{X}}^{\varepsilon_n})\}$ are verified, and we will complete the proof of Theorem 2.1. In fact, it is just a simple modification of the arguments in Subsection 3.1.

Lemma 3.3. *With probability one, $(\widehat{X}_t^{\varepsilon_n})_{0 \leqslant t \leqslant 1}$ converges uniformly to the diffusion paths $(X_t)_{0 \leqslant t \leqslant 1}$.*

Proof. As in the proof of Proposition 3.1, we need to estimate $P(\sup_{0 \leqslant t \leqslant 1} |\widehat{X}_t^\varepsilon - X_t| > \lambda\varepsilon)$ for some universal constant λ, which reduces to the estimation of $P(\sup_{\zeta_{j-1}^\varepsilon \leqslant t \leqslant \zeta_j^\varepsilon} |\widehat{X}_t^\varepsilon - X_t| > \lambda\varepsilon, \zeta_j^\varepsilon \leqslant 1)$, $P(\sup_{\zeta_l^\varepsilon \leqslant t \leqslant 1} |\widehat{X}_t^\varepsilon - X_t| > \lambda\varepsilon,\ M_V^\varepsilon = l)$ and $P(M_V^\varepsilon > k)$.

For the first quantity, from the definition of $\widehat{X}^\varepsilon$ we have

$$\widehat{X}^\varepsilon([\zeta_{j-1}^\varepsilon, \zeta_j^\varepsilon]) = \widetilde{X}^\varepsilon([\zeta_{j-1}^\varepsilon, \zeta_j^\varepsilon])$$

on $\{\zeta_j^\varepsilon \leqslant 1\}$, regardless of whether the path has visited the smaller box $H^\varepsilon_{\boldsymbol{n}^\varepsilon_{j-1}}$ during $[\zeta_{j-1}^\varepsilon, \zeta_j^\varepsilon]$. Therefore, the event $\{\sup_{\zeta_{j-1}^\varepsilon \leqslant t \leqslant \zeta_j^\varepsilon} |\widehat{X}_t^\varepsilon - X_t| > \lambda\varepsilon, \zeta_j^\varepsilon \leqslant 1\}$ again implies that during $[\zeta_{j-1}^\varepsilon, \zeta_j^\varepsilon]$, the path must have traveled through many narrow tunnels and spread far away from the box $V^\varepsilon_{\boldsymbol{n}^\varepsilon_{j-1}}$ before visiting another box. More precisely, again we have

$$\{\sup_{\zeta_{j-1}^\varepsilon \leqslant t \leqslant \zeta_j^\varepsilon} |\widehat{X}_t^\varepsilon - X_t| > \lambda\varepsilon, \zeta_j^\varepsilon \leqslant 1\} \subset \{\sigma_0 < \sigma_1 < \cdots < \sigma_{\lambda/12} < \infty\},$$

the same as in the proof of Proposition 3.1. Similar arguments apply to the estimation of the second quantity, and the third quantity has nothing to do with the polygonal approximation.

Therefore, we can apply exactly the same arguments as in the proof of Proposition 3.1 to concluded that

$$\sum_{n=1}^{\infty} P(\sup_{0 \leqslant t \leqslant 1} |\widehat{X}_t^\varepsilon - X_t| > \lambda\varepsilon_n) < \infty,$$

where λ is the universal constant chosen in that proof. □

Now the proof of Theorem 2.1 is complete.

Remark 3.1. From the proof of Theorem 2.1, it is not hard to see that the global assumption (C) on the generating vector fields can be weakened to a local one to some extend. In fact, the only property of the vector fields we've used from Assumption (C) is that at every point on the boundary of H_z^ε, the vector fields $V_1, \cdots, V_d$ do not generate a subspace of the tangent space at that point. Therefore, it suffices to assume that for each z and ε, there exists a small rotation O (an orthogonal transformation) such that after rotating the box H_z^ε by O with respect to its center, the vector fields do not generate a subspace of the tangent space at every point on the boundary. The smallness of the rotation O can be quantified as follows. If we let

$$\widetilde{H}_z^\varepsilon = \varepsilon z + O(H_z^\varepsilon - \varepsilon z)$$

be the rotated box, then O should satisfy the condition that for any $x = (x_1, \cdots, x_N) \in \widetilde{H}_z^\varepsilon$,

$$|x_i - \varepsilon z_i| < \frac{\varepsilon}{2}, \ \forall i = 1, \cdots, N.$$

This is to ensure that the geometric configuration, in particular the tunnel structure, is not damaged, so that the whole proof of Theorem 2.1 carries through in the same way.

Acknowledgement

The authors wish to thank the referees for their very careful reading on the manuscript and very useful suggestions. The authors are supported by the Oxford-Man Institute at University of Oxford.

References

[1] G. Ben Arous, S. Kusuoka, D.W. Stroock, The Poisson kernel for certain degenerate elliptic operators, *J. Funct. Anal.* 56 (2) (1984) 171-209.

[2] K.T. Chen, Iterated path integrals, *Bull. Amer. Math. Soc.* 83 (5) (1977) 831-879.

[3] B. Hambly, T. Lyons, Uniqueness for the signature of a path of bounded variation and the reduced path group, *Ann. of Math.* 171(1) (2010) 109-167.

[4] L. Hörmander, Hypoelliptic second order differential equations, *Acta Math.* 119 (1) (1967) 147-171.

[5] N. Ikeda, S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, 1989.

[6] Y. Le Jan, Z. Qian, Stratonovich's signatures of Brownian motion determine Brownian sample paths, *Probab. Theory Relat. Fields* 157 (2012) 440-454.

[7] T. Lyons, Differential equations driven by rough signals, *Rev. Mat. Iberoamericana* 14 (2) (1998) 215-310.

[8] T. Lyons, M. Caruana, T. Lévy, J. Picard, *Differential Equations Driven by Rough Paths*, *École d'été de probabilités de Saint-Flour XXXIV-2004*, Springer, 2007.

[9] T. Lyons, Z. Qian, *System Control and Rough Paths*, Oxford University Press, 2002.

[10] E. Wong, M. Zakai, On the relation between ordinary and stochastic differential equations, *Internat. J. Engrg. Sci.* 3 (2) (1965) 213-229.