ELSEVIER Stochastic Processes and their Applications 102 (2002) 265–283
stochastic processes and their applications
www.elsevier.com/locate/spa

Large deviations and support theorem for diffusion processes via rough paths

M. Ledoux[a,*], Z. Qian[a], T. Zhang[b]

[a] *Institut de Mathématiques, Université Paul-Sabatier, F-31062 Toulouse, France*
[b] *Department of Mathematics, University of Manchester, Manchester M13 9PL, UK*

Received 14 November 2001; received in revised form 31 May 2002; accepted 31 May 2002

Abstract

We use the continuity theorem of Lyons for rough paths in the p-variation topology to produce an elementary approach to the large deviation principle and the support theorem for diffusion processes. The proofs reduce to establish the corresponding results for Brownian motion itself as a rough path in the p-variation topology, $2 < p < 3$, and the technical step is to handle the Lévy area in this respect. Some extensions and applications are discussed.

Keywords: Rough paths; Large deviation principle; Support theorem; Diffusion processes

1. Introduction

It is well known that one-dimensional diffusion processes satisfy a small time large deviation principle as continuous images (in the uniform topology) of the Wiener process. Due to the lack of continuity of the Itô map, Schilder's theorem and the contraction principle may not be used any more in higher dimension. To overcome this difficulty, Freidlin and Wentzell developed in the early 1970s refined techniques involving exponential continuity to reach the large deviation properties for small perturbations of diffusion processes (under mild regularity conditions on the coefficients), cf. Freidlin and Wentzell (1984), Dembo and Zeitouni (1990), Deuschel and Stroock (1989),

* Corresponding author.
E-mail addresses: ledoux@math.ups-tlse.fr (M. Ledoux), qian@math.ups-tlse.fr (Z. Qian), tzhang@maths.man.ac.uk (T. Zhang).

PII: S0304-4149(02)00176-X

etc. The philosophy of the results is that while the Itô map may not be continuous, everything actually works as if it were.

Recently, Lyons (1998) developed an integration theory along rough paths based on Taylor's expansion and the use of p-variation metrics and proved the striking result that diffusion processes are actually continuous images, with respect to the p-variation norm, $2 < p < 3$, not of the Brownian motion itself but of the (a) rough path consisting of the sample paths of Brownian motion and its Lévy area process. As a consequence, the large deviation contraction principle may still be used, however, at the level of rough paths. It is the purpose of this paper to establish the large deviation principle for the Brownian rough path in the p-variation topology. To this task, it will be an essential technical step to get estimates for Brownian motion together with its Lévy area. By the continuity theorem, we deduce then the large deviation properties of small perturbations of diffusion processes (in the same topology) by the contraction principle. Along the same line, we describe an analogous approach to the support theorem for diffusions, emphasizing thus, what we believe, entlighting applications of the rough path theory.

The concept of p-variation norm (or p-variation distance) of paths in Banach spaces has been present in the mathematical literature for many years. However, not like the supremum, Sobolev or Hölder norms, its significance has not been completely recognized until the very recent work by Lyons (1998). It is well known that solutions to ordinary differential equations are not in general continuous in the driving paths under the supremum norm or the Hölder norm. This has been one main obstruction to carry over the ODE approach to random driving paths like sample paths of Brownian motion. Instead, Itô established the stochastic calculus by using the martingale property of Brownian motion. It remained an open problem to establish some kind of "continuity theorem" (or universal limit theorem as Malliavin (1997) called it) for Itô's functionals until Lyons showed that the Itô functionals (as rough paths) are indeed continuous in p-variation topology.

Lyons's continuity theorem of Itô's functionals exhibits that the p-variation distance leads to a natural metric structure on the space of Itô's functionals. In this work we consider two questions about Brownian motion under the p-variation metric structure. The first one is the large deviation principle for the Brownian rough path consisting of Brownian motion together with its Lévy area process (see below for the definition). The second one is the support theorem for the same process again under the p-variation distance. Both are concerned with the asymptotic behavior of Brownian motion and its Lévy area. The main interest in this approach is that the corresponding properties then immediately extend to solutions of stochastic differential equations by the continuity theorem for rough paths in the p-variation topology, leading thus to a direct description of these classical results. Moreover, the approach may be potentially applied to large classes of stochastic processes beyond semimartingales (see e.g. Bass et al., 1998), even though we do not address applications of this type here.

In order to state the main results proved in this paper, we first recall several basic notions and fix the notations as well. We refer to Lyons (1998) and Lyons and Qian (to appear) for further details.

The p-variation norm, $p \geqslant 1$ (in the sense of Banach), of a path (with finite running time say $[0,1]$) $x = (x_t)_{t\in[0,1]}$ in a Banach space B is defined by

$$\|x\|_p = \left(\sup_D \sum_\ell |x_{t_\ell} - x_{t_{\ell-1}}|_B^p\right)^{1/p},$$

where the supremum runs over all finite partitions

$$D = \{0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = 1\}$$

of $[0,1]$.

The p-variation norm introduced in Lyons (1998) is the p-variation norm involving not only the path itself but the path together with its iterated integrals up to degree $[p]$ (integer part). In the present work, we only consider for simplicity paths in Euclidean spaces (for the Banach space case, see Ledoux et al., 2002). Let Δ denote the simplex $\{(s,t);\ 0 \leqslant s \leqslant t \leqslant 1\}$. A continuous map X from Δ to the truncated tensor algebra

$$T^{([p])}(\mathbb{R}^d) \equiv \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d)^{\otimes 2} \oplus \cdots \oplus (\mathbb{R}^d)^{\otimes [p]}$$

$(p \geqslant 1)$ is called a rough path in $T^{([p])}(\mathbb{R}^d)$ with roughness p if it is of finite p-variation and multiplicative with respect to the tensor product of $T^{([p])}(\mathbb{R}^d)$. In other words, if we set

$$X_{s,t} = (1, X^1_{s,t}, \ldots, X^{[p]}_{s,t}), \quad (s,t) \in \Delta,$$

then

$$\max_{1\leqslant j\leqslant [p]} \left(\sup_D \sum_\ell |X^j_{t_{\ell-1},t_\ell}|^{p/j}\right)^{j/p} < \infty$$

and

$$X_{s,t} = X_{s,u} \otimes X_{u,t} \quad \text{for all } (s,u),(u,t) \in \Delta.$$

The collection of all rough paths in $T^{([p])}(\mathbb{R}^d)$ is denoted by $\Omega_p(\mathbb{R}^d)$, which is a metric space under the p-variation distance

$$d_p(X,Y) = \max_{1\leqslant j\leqslant [p]} \left(\sup_D \sum_\ell |X^j_{t_{\ell-1},t_\ell} - Y^j_{t_{\ell-1},t_\ell}|^{p/j}\right)^{j/p}.$$

Typical examples of rough paths are the Lipschitz paths. Those are rough paths obtained from continuous paths with finite total variation in the following way. Suppose $x = (x_t)_{t\in[0,1]}$ is a continuous path in $\mathbb{R}^d$ with finite total variation. We may then construct a rough path $X = X_{s,t} = (1, X^1_{s,t}, \ldots, X^{[p]}_{s,t})$ by setting $X^j_{s,t}$ to be the jth iterated integral of x over the interval $[s,t]$,

$$X^j_{s,t} \equiv \int \cdots \int_{s<t_1<\cdots<t_j<t} \mathrm{d}x_{t_1} \otimes \cdots \otimes \mathrm{d}x_{t_j}.$$

We write

$$X = (1, X^1, \ldots, X^{[p]}) = F^{[p]}(x) = F(x)$$

to describe X as a function of x, or rather lying above x. All rough paths obtained in this way will be called smooth rough paths. The rough paths in the closure of all smooth rough paths under the p-variation metric are called geometric rough paths and will be denoted $G\Omega_p(\mathbb{R}^d)$ (cf. Lyons and Qian, to appear).

Interesting examples of geometric rough paths come from stochastic analysis. Let us describe one class of these examples. Let $w = (w_t)_{t\in[0,1]}$ be Brownian motion in $\mathbb{R}^d$ (starting from the origin). Denote by $w(m)$ its mth dyadic polygonal approximation defined by

$$w(m)_t = w_{(\ell-1)/2^m} + 2^m\left(t - \frac{\ell-1}{2^m}\right)(w_{\ell/2^m} - w_{(\ell-1)/2^m})$$

if $(\ell-1)/2^m \leqslant t \leqslant \ell/2^m$, $\ell = 1,\dots,2^m$. Denote by $W(m) = F(w(m))$ the associated smooth rough path. Since Brownian motion is of finite p-variation for every $p > 2$, it is necessary, but also enough, to consider the iterated integrals up to the degree $[p] = 2$. Then, it was shown in Sippilainen (1993), Lyons and Qian (to appear), that there is a geometric rough path W (of degree 2) such that $d_p(W(m), W) \to 0$ almost surely as $m \to \infty$. (This actually also follows from the bounds provided in this work.) This limit W is called the canonical (with respect to possibly alternate approximation procedures, cf. Lyons (1998), Lyons and Qian (to appear)) rough path associated with Brownian motion w. In particular, W^1 is the increment process, $W^1_{s,t} = w_t - w_s$, $s < t$ in $[0,1]$, and W^2 may be identified with the Lévy area process

$$\begin{aligned} W^2_{s,t} &= \int\int_{s<t_1<t_2<t} \circ \mathrm{d}w_{t_1} \circ \mathrm{d}w_{t_2} \\ &= \int_s^t (w_r - w_s)\circ \mathrm{d}w_r = \left(\int_s^t (w^i_r - w^i_s)\circ \mathrm{d}w^j_r\right)_{1\leqslant i,j\leqslant d}, \qquad s<t, \end{aligned}$$

where $w = (w^1,\dots,w^d) \in \mathbb{R}^d$ and where $\circ\mathrm{d}$ indicates Stratonovich's integration. We may thus make sense of $W = F^2(w) = F(w)$ even though w is not smooth. We only work below with the canonical rough path W above w generated by the dyadic polygonal approximations.

To formulate Schilder's theorem for the Brownian rough path in the p-variation topology, $2 < p < 3$, denote, for each $\varepsilon > 0$, by μ_ε the law of $F(\varepsilon w)$ on $G\Omega_p(\mathbb{R}^d)$. As a first main result, we will show that the family $(\mu_\varepsilon)_{\varepsilon>0}$ satisfies a large deviation principle as $\varepsilon \to 0$ in $G\Omega_p(\mathbb{R}^d)$ with the good rate function

$$I(X) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 |\dot{x}_t|^2\,\mathrm{d}t & \text{if } X = F(x) \text{ for some } x \in H, \\ +\infty & \text{if not,} \end{cases} \tag{1}$$

where H is the usual Cameron–Martin space on $\mathbb{R}^d$. To better understand this rate function, it might be worthwhile mentioning that any $x \in H$ gives rise to a geometric rough path $X = (1, X^1, X^2) = F(x)$ in $G\Omega_p(\mathbb{R}^d)$, $2 < p < 3$. By a large deviation principle, we understand that for any closed set Γ in $G\Omega_p(\mathbb{R}^d)$,

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(\Gamma) \leqslant -\inf_{X\in\Gamma} I(X)$$

while for every open set G,

$$\liminf_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(G) \geqslant -\inf_{X\in G} I(X).$$

The proof of this result we present follows the classical pattern. We work with the dyadic approximations $W(m)$ of the canonical rough path W that are easily shown to satisfy large deviation properties. We then only need to show that the $W(m)$'s are exponentially good approximations of W as rough paths in the p-variation topology which is performed by standard tools on each level. Since we work with the dyadic polygonal approximations $W(m)$ of W, the large deviation principle we obtain actually takes place in the subspace $\mathscr{D}_p(\mathbb{R}^d) \subset G\Omega_p(\mathbb{R}^d)$ which is the closure, under the d_p-metric, of the set $\mathscr{D}$ of dyadic rough paths of order $[p]=2$ (generated by curves defined as continuous broken lines over dyadic intervals of $[0,1]$). Note also in this respect that the dyadic polygonal approximation of any element x in the Cameron–Martin space H is shown below to converge in the d_p-metric to a canonical rough path $G\Omega_p(\mathbb{R}^d)$ that is easily identified, by the martingale convergence theorem, to the rough path defined from x as iterated integrals (other approximation procedures may however yield different limits). The second level involving Lévy area contributions requires some careful estimates. (A large deviation principle for Lévy area might have been studied earlier in the classical setting but we are not aware of any references.) We use to this task bounds on p-variation distances involving dyadic partitions and moment estimates to control the exponential approximations. While general tools from the theory of Gaussian quadratic forms may be used alternatively, we have tried to proceed at an elementary level to demonstrate the actual simplicity of the overall approach.

One aspect of this large deviation principle is that the p-variation topology is stronger than the uniform topology, so that this result covers in particular the classical Schilder theorem. But the significance of this large deviation principle, as announced, is now that many Itô's functionals are continuous (as rough paths) of the Brownian rough path with respect to the p-variation distance, and thus satisfy, by the standard contraction principle, a large deviation principle. Lyons's continuity theorem (cf. Lyons and Qian, to appear) for stochastic differential equations namely indicates that, given the stochastic differential equation

$$\mathrm{d}y_t = f_0(t,y_t)\,\mathrm{d}t + f(t,y_t)\circ \mathrm{d}w_t, \qquad y_0 = x_0, \tag{2}$$

under the assumption that f_0, f be C^3 with linear growth (in both space and time), the Itô map, defined on smooth paths, may be extended into a continuous map φ from $G\Omega_p(\mathbb{R}^d)$ into itself under the p-variation distance for any $2<p<3$ so that the (one) rough path Y above the solution y of (2) is the image $Y=\varphi(W)$ of the canonical Brownian rough path W. Therefore, as an immediate application of the contraction principle, the distributions v_ε of $\varphi(F(\varepsilon w))$, $\varepsilon>0$, satisfy a large deviation principle as $\varepsilon\to 0$ in $G\Omega_p(\mathbb{R}^d)$ with rate function $J(Y)=\inf\{I(X);\ Y=\varphi(X)\}$. By projection on the first level paths, we thus recover the classical Freidlin–Wentzell theorem for small perturbations of solutions of stochastic differential equations. The continuity theorem for rough paths taken for granted, the large deviation principle for diffusion processes thus follows from the large deviation principle for Brownian motion as a rough path in the p-variation metric. While the regularity conditions are somewhat stronger than the classical ones, the conclusion actually holds in a stronger topology. (It might be possible to approximate Lipschitz coefficients by smooth ones to fully recover the large deviation principle for solutions of stochastic differential equations. However, we did not investigate carefully this question.)

The support of the distribution of a diffusion process under the uniform norm was first described by Stroock and Varadhan (1972), and has been called since then the support theorem. A simple proof (even working for stronger topologies) may be found in Millet and Sanz-Solé (1994). In our rough path setting, we first describe the support of the distribution of the canonical rough path consisting of Brownian motion and its Lévy area process in the p-variation topology as the closure of all dyadic paths under the p-variation distance, $2 < p < 3$. While the support theorem for Brownian motion itself is classically described, again the lack of continuity of the Itô map does not allow us to deduce the corresponding conclusion for diffusions. However, the continuity theorem may be used again to transfer the support theorem for the geometric rough path associated with Brownian motion to the support theorem for solutions of stochastic differential equations (now with respect to the p-variation distance).

We will show that the support, in the p-variation topology, $2 < p < 3$, of the distribution of the canonical rough path W induced from Brownian motion w is the closure $\mathscr{D}_p(\mathbb{R}^d)$ of the set $\mathscr{D}$ of the dyadic rough paths (of order $[p]=2$). Equivalently, it is also the closure of the canonical rough paths from the Cameron–Martin space given as iterated integrals, but different approximation schemes may actually yield different, and larger, closures. Projections on the first level paths, however, all coincides. The proof involves some precise small ball lower bounds for both Gaussian linear and quadratic forms and requires the use of correlation inequalities. As a consequence, the support of the distribution of the diffusion Y of (2) as a rough path is the closure of $\varphi(\mathscr{D})$ in this topology.

The paper is organized as follows. In Section 2, we prove the exponential approximation result of the Brownian rough path by the dyadic paths to establish the announced large deviation principle. The corresponding arguments for the support theorem are developed in Section 3. One aspect of the techniques emphasized in this work is that they applied to more general stochastic processes than just Brownian motion. We discuss in the last section a variety of further examples and applications of interest. In particular, we describe a version of Strassen's law of the iterated logarithm in this context.

Throughout this paper we use t_ℓ^m to denote the dyadic rationals $\ell/2^m$, $\ell=0,1,\dots,2^m$, $m \geqslant 0$. We set $\Delta_\ell^m x = X^1_{t_{\ell-1}^m,t_\ell^m} = x_{t_\ell^m} - x_{t_{\ell-1}^m}$. We will always denote by $x(m)$ the mth dyadic approximation of a path x (in $\mathbb{R}^d$) and $X(m)_{s,t} = (1, X(m)^1_{s,t}, X(m)^2_{s,t})$, $(s,t)\in\Delta$, will be the smooth rough path of order 2 associated with $x(m)$. We only consider rough paths up to degree 2 (in most part of the paper x will be Brownian motion). Crucial to the approach is the description of the rough path $X(m)$ associated to the dyadic approximation $x(m)$ of a path x in the following way:

For $n \leqslant m$, $\ell = 1,\dots,2^n$,

$$X(m)^1_{t_{\ell-1}^n,t_\ell^n} = \Delta_\ell^n x \quad \text{for } n \leqslant m \tag{3}$$

while

$$X(m)^1_{t^n_{\ell-1},t^n_\ell} = 2^{m-n}\Delta^m_{k(n,m,\ell)}x \quad \text{for } n \geqslant m, \tag{4}$$

where, for $n \geqslant m$ and $\ell \geqslant 1$, $k(n,m,\ell)$ is the unique integer k among $1,\dots,2^m$ such that $t^m_{k-1} \leqslant t^n_{\ell-1} < t^n_\ell < t^m_k$. Concerning the second level path, $X(m)^2_{t^n_{\ell-1},t^n_\ell}$ splits, when $n \leqslant m$, into a symmetric and an anti-symmetric parts as

$$X(m)^2_{t^n_{\ell-1},t^n_\ell} = \frac{1}{2}\,\Delta^n_\ell x \otimes \Delta^n_\ell x + \frac{1}{2}\sum_{\substack{r,s=2^{m-n}(\ell-1)+1\\ r<s}}^{2^{m-n}\ell} (\Delta^m_r x \otimes \Delta^m_s x - \Delta^m_s x \otimes \Delta^m_r x), \tag{5}$$

$n \leqslant m$, $\ell = 1,\dots,2^n$, while

$$X(m)^2_{t^n_{\ell-1},t^n_\ell} = 2^{2(m-n)-1}(\Delta^m_{k(n,m,\ell)}x \otimes \Delta^m_{k(n,m,\ell)}x) \quad \text{for } n \geqslant m. \tag{6}$$

We denote by $|\cdot|$ Euclidean norm in $\mathbb{R}^d$ and Hilbert–Schmidt norm in $(\mathbb{R}^d)^{\otimes 2}$.

2. Large deviations

In this section, we establish the main result on large deviations of Brownian motion as a rough path in $G\Omega_p(\mathbb{R}^d)$ and deduce thus by the contraction principle small time large deviations for solutions of stochastic differential equations.

As described in the Introduction, we only work with canonical rough paths as limits of dyadic polygonal approximations. Let $2 < p < 3$ and recall μ_ε, $\varepsilon > 0$, the distribution on $\mathscr{D}_p(\mathbb{R}^d) \subset G\Omega_p(\mathbb{R}^d)$ of the canonical rough path $F(\varepsilon w)$ defined as the limit in the p-variation metric of the dyadic rough paths $F(\varepsilon w(m))$.

Theorem 1. *The family $(\mu_\varepsilon)_{\varepsilon>0}$ satisfies a large deviation principle in $G\Omega_p(\mathbb{R}^d)$, $2 < p < 3$, with the good rate function I of (1).*

The continuity theorem of Lyons (1998) and the contraction principle immediately yield the following corollary. Denote by ν_ε the distribution on $G\Omega_p(\mathbb{R}^d)$ of $\varphi(F(\varepsilon w))$ where $\varphi : G\Omega_p(\mathbb{R}^d) \to G\Omega_p(\mathbb{R}^d)$ is the (continuous) Itô map associated to the stochastic differential equation (2). In other words, ν_ε is the distribution of a rough path associated to the solution of

$$\mathrm{d}y^\varepsilon_t = f_0(t,y^\varepsilon_t)\,\mathrm{d}t + \varepsilon f(t,y^\varepsilon_t) \circ \mathrm{d}w_t, \qquad y^\varepsilon_0 = x_0.$$

With respect to the classical Freidlin–Wentzell result, note that the regularity assumptions on the coefficients are more stringent than the usual ones. On the other hand, the large deviation result holds in a stronger topology (and as a rough path, also includes the second level path).

Corollary 2. *The family $(\nu_\varepsilon)_{\varepsilon>0}$ satisfies a large deviation principle in $G\Omega_p(\mathbb{R}^d)$, $2 < p < 3$, with the good rate function $J(Y) = \inf\{I(X);\ Y = \varphi(X)\}$.*

The rest of this section is devoted to the proof of Theorem 1.

Proof of Theorem 1. It follows the standard lines of arguments for sample path large deviations (cf. Dembo and Zeitouni, 1990; Deuschel and Stroock, 1989; Varadhan, 1984). As explained in the introduction, we establish the large deviation principle in the subspace $\mathscr{D}_p(\mathbb{R}^d) \subset G\Omega_p(\mathbb{R}^d)$.

If $z=(z_1,\dots,z_{2^m})\in(\mathbb{R}^d)^{2^m}$, define the (smooth) dyadic polygonal path x^z on $[0,1]$ (starting from the origin) by

$$x_t^z = z_{\ell-1} + 2^m(t-t_{\ell-1}^m)(z_\ell - z_{\ell-1}) \quad \text{if } t_{\ell-1}^m \leqslant t \leqslant t_\ell^m$$

(with the convention $z_0=0$). Denote by X^z the rough path of order 2 associated to x^z and define then a map F_m from $(\mathbb{R}^d)^{2^m}$ to $\mathscr{D}_p(\mathbb{R}^d)$ by $F_m(z)=X^z=(1,(X^z)^1,(X^z)^2)$. It is an easy exercise to check that F_m is continuous for each fixed m.

As is classical, for each fixed m, the Gaussian random vectors

$$\varepsilon(w_{t_1^m},\dots,w_{t_{2^m-1}^m},w_1)$$

in $(\mathbb{R}^d)^{2^m}$ satisfy a large deviation principle as $\varepsilon\to 0$ with good rate function

$$\frac{1}{2}\sum_{\ell=1}^{2^m} 2^m|z_\ell - z_{\ell-1}|^2 \quad \text{for } z=(z_1,\dots,z_{2^m})\in(\mathbb{R}^d)^{2^m}, \quad z_0=0.$$

By the contraction principle, the distributions of

$$F_m(\varepsilon(w_{t_1^m},\dots,w_{t_{2^m-1}^m},w_1)) = F(\varepsilon w(m)) = (1,\varepsilon W(m)^1, \varepsilon^2 W(m)^2)$$

satisfy a large deviation principle as $\varepsilon\to 0$ on $\mathscr{D}_p(\mathbb{R}^d)$ with rate function

$$\begin{aligned} I_m(X) &= \frac{1}{2}\inf\left\{\sum_{\ell=1}^{2^m} 2^m|z_\ell - z_{\ell-1}|^2; z\in(\mathbb{R}^d)^{2^m}, X=F_m(z)\right\} \\ &= \frac{1}{2}\inf\left\{\int_0^1 |\dot{x}_t^z|^2\,\mathrm{d}t; z\in(\mathbb{R}^d)^{2^m}, X=F_m(z)\right\}. \end{aligned}$$

Note that $I_m=I$ on the mth dyadic polygonal paths. It is classical and easy to see that if x is a path on $[0,1]$, and if $x(m)$ is the mth dyadic polygonal approximation of x, then as $m\to\infty$,

$$\int_0^1 |\dot{x}(m)_t|^2\,\mathrm{d}t = \sum_{\ell=1}^{2^m} 2^m |x_{t_\ell^m} - x_{t_{\ell-1}^m}|^2 \to \int_0^1 |\dot{x}_t|^2\,\mathrm{d}t$$

if x is absolutely continuous and $+\infty$ if not. Now, according to the general principles of the large deviation theory (cf. Dembo and Zeitouni, 1990, p. 133; Deuschel and Stroock, 1989), in order to conclude the proof of Theorem 1, we are left to show that the rough paths $W(m)$ are exponentially good approximations of W in $\mathscr{D}_p(\mathbb{R}^d)\subset G\Omega_p(\mathbb{R}^d)$ in the sense that, for every $\delta>0$,

$$\lim_{m\to\infty}\limsup_{\varepsilon\to 0}\varepsilon^2\log\mathbb{P}(d_p(F(\varepsilon w(m)),F(\varepsilon w))>\delta) = -\infty, \tag{7}$$

and that, for any $\alpha > 0$,

$$\lim_{m\to\infty} \sup_{x;\int_0^1 |\dot{x}_t|^2\,\mathrm{d}t \leqslant \alpha} d_p(F(x(m)), F(x)) = 0. \tag{8}$$

Part of the proof of (8) is that the rough path $F(x)$ above any element x in the Cameron–Martin space H is defined as the limit of $F(x(m))$, $m \geqslant 1$, in the metric d_p. (It is not difficult to check, by the martingale convergence theorem, that this limit coincides with the rough path constructed from x as iterated integrals.) Since the I_m's are good rate functions, (8) in particular implies that I is a good rate function. To establish both (7) and (8), we make use of the following lemma put forward in Hambly and Lyons (1998) (see also Ledoux et al., 2002; Lyons and Qian, to appear) that provides a useful bound on the p-variation distance. If X and Y are rough paths of degree $[p]$, and $\gamma > p-1$, we agree to denote, for $j = 1, \ldots, [p]$,

$$D_{j,p}(X,Y) = D_{j,p}(X^j, Y^j) = \left(\sum_{n=1}^{\infty} n^{\gamma} \sum_{\ell=1}^{2^n} |X^j_{t^n_{\ell-1},t^n_\ell} - Y^j_{t^n_{\ell-1},t^n_\ell}|^{p/j} \right)^{j/p}$$

and write $D_{j,p}(X)$ for $D_{j,p}(X,0)$. The values of $2 < p < 3$ and $\gamma > p-1$ are fixed below.

Lemma 3. *Let $2 < p < 3$ and $\gamma > p-1$. There is a constant $C > 0$ only depending on p and γ such that for any rough paths X and Y of degree $[p] = 2$,*

$$d_p(X,Y) \leqslant C \max(D_{1,p}(X,Y), D_{1,p}(X,Y)(D_{1,p}(X) + D_{1,p}(Y)), D_{2,p}(X,Y)).$$

In order to prove (7), it will be enough to show that there is a sequence $c(m) \to 0$ such that for every $q \geqslant 1$ (or only $q > p$),

$$(\mathbb{E}(D_{1,p}(W(m), W)^q))^{1/q} \leqslant c(m)\sqrt{q} \tag{9}$$

and

$$(\mathbb{E}(D_{2,p}(W(m), W)^q))^{1/q} \leqslant c(m)q. \tag{10}$$

Indeed, by the Chebyshev inequality we then have, for $q = q(\varepsilon) = \varepsilon^{-2}$, and $j = 1, 2$,

$$\mathbb{P}(D_{j,p}(W(m), W) > \delta\varepsilon^{-j}) \leqslant (\delta^{-1}\varepsilon^j)^q c(m)^q q^{jq/2} \leqslant (\delta^{-1}c(m))^q$$

so that

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(D_{j,p}(W(m), W) > \delta\varepsilon^{-j}) \leqslant \log(\delta^{-1}c(m)).$$

Moreover, since $\mathbb{E}(D_{1,p}(W)^q)^{1/q} \leqslant C\sqrt{q}$ for some constant $C > 0$ only depending on p, γ and d, by the Cauchy–Schwarz and the triangle inequalities, for every q,

$$(\mathbb{E}([D_{1,p}(W(m), W)(D_{1,p}(W(m)) + D_{1,p}(W))]^q))^{1/q} \leqslant 2c(m)(2C + c(m))q$$

so that similarly

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(D_{1,p}(W(m),W)(D_{1,p}(W(m)) + D_{1,p}(W)) > \delta\varepsilon^{-2})$$

$$\leqslant \log(\delta^{-1} 2c(m)(2C + c(m))).$$

By Lemma 3 and the fact that $c(m) \to 0$, Property (7) readily follows.

We thus establish (9) and (10). We start with the first level $j = 1$. By (3) and (4),

$$D_{1,p}(W(m),W) = \left(\sum_{n=m+1}^{\infty} n^{\gamma} \sum_{\ell=1}^{2^n} |2^{m-n} \Delta^m_{k(n,m,\ell)} w - \Delta^n_\ell w|^p \right)^{1/p}.$$

By Hölder's inequality, for any numbers $a_n > 0$, and any $q > p$,

$$\mathbb{E}(D_{1,p}(W(m),W)^q) \leqslant A(m,q) \sum_{n=m+1}^{\infty} a_n^{q/p} \sum_{\ell=1}^{2^n} \mathbb{E}(|2^{m-n} \Delta^m_{k(n,m,\ell)} w - \Delta^n_\ell w|^q),$$

where

$$A(m,q) = \left(\sum_{n=m+1}^{\infty} 2^n \left(\frac{n^\gamma}{a_n} \right)^{q/(q-p)} \right)^{(q-p)/p}.$$

Thus,

$$\mathbb{E}(D_{1,p}(W(m),W)^q) \leqslant A(m,q)(2d)^q \mathbb{E}(|g|^q) \sum_{n=m+1}^{\infty} a_n^{q/p} 2^{-(q-2)n/2},$$

where g is a standard normal random variable. Choose then the sequence (a_n) such that

$$a_n^{q/p} 2^{-(q-2)n/2} = 2^{-\varepsilon q n},$$

where $0 < \varepsilon < \frac{1}{2} - 1/p$. We then have

$$A(m,q)^{1/q} = \left(\sum_{n=m+1}^{\infty} n^{\gamma q/(q-p)} 2^{-qp/(q-p)(1/2 - 1/p - \varepsilon)n} \right)^{(q-p)/qp}$$

from which it follows that $A(m,q)^{1/q}$ is uniformly bounded in m and q. Since $(\mathbb{E}(|g|^q))^{1/q} \sim \sqrt{q}$ as $q \to \infty$, it is then easy to conclude to the first assertion (9).

Although the basic principle is similar, the second level requires somewhat more delicate arguments since the non-linearity of the Lévy area produces a number of mixed terms. To this end, it would have been possible to invoke general integrability properties of Gaussian chaos (cf. e.g. Ledoux, 1996). However, we use an argument as simple and direct as possible. As for the first level, by Hölder's inequality, for every m and $q > 2p$,

$$\mathbb{E}(D_{2,p}(W(m),W)^q) \leqslant A(q) \sum_{n=1}^{\infty} a_n^{2q/p} \sum_{\ell=1}^{2^n} \mathbb{E}(|W(m)^2_{t^n_{\ell-1},t^n_\ell} - W^2_{t^n_{\ell-1},t^n_\ell}|^q),$$

where

$$A(q) = \left(\sum_{n=1}^{\infty} 2^n \left(\frac{n^\gamma}{a_n} \right)^{2q/(2q-p)} \right)^{(2q-p)/p}.$$

By (5), for $n \leqslant m$, $\ell = 1, \ldots, 2^n$,

$$W(m+1)^2_{t^n_{\ell-1}, t^n_\ell} - W(m)^2_{t^n_{\ell-1}, t^n_\ell}$$

$$= \frac{1}{2} \sum_{r=2^{m-n}(\ell-1)+1}^{2^{m-n}\ell} (\Delta^{m+1}_{2r-1} w \otimes \Delta^{m+1}_{2r} w - \Delta^{m+1}_{2r} w \otimes \Delta^{m+1}_{2r-1} w).$$

We then make use of the following elementary lemma (cf. Ledoux et al., 2002).

Lemma 4. *Let $G = (G_1, \ldots, G_N)$ and $G' = (G'_1, \ldots, G'_N)$, $N \geqslant 1$, be independent centered Gaussian random vectors in $(\mathbb{R}^d)^N$. There exists a numerical constant $C > 0$ such that for every $q \geqslant 1$,*

$$\left(\mathbb{E}\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^q \right) \right)^{1/q} \leqslant Cq \left(\mathbb{E}\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^2 \right) \right)^{1/2}.$$

In particular, if the random variables $G_1, \ldots, G_N, G'_1, \ldots, G'_N$ are independent and identically distributed,

$$\left(\mathbb{E}\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^q \right) \right)^{1/q} \leqslant Cq\sqrt{N} (\mathbb{E}(|G_1 \otimes G'_1|^2))^{1/2}.$$

Proof. We apply the equivalence of moments of norms of Gaussian random vectors (cf. Ledoux and Talagrand, 1991) together with Fubini's theorem. Writing $\mathbb{E}$ for expectation with respect to the sample $(G_1, \ldots, G_N)$, and $\mathbb{E}'$ with respect to $(G'_1, \ldots, G'_N)$, we may write

$$\mathbb{E}\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^q \right) = \mathbb{E}\mathbb{E}'\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^q \right).$$

Now, conditionally on G, the equivalence of moments of norms of Gaussian random vectors yields, for some numerical constant $C > 0$,

$$\left(\mathbb{E}'\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^q \right) \right)^{1/q} \leqslant C\sqrt{q} \left(\mathbb{E}'\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^2 \right) \right)^{1/2}.$$

Repeat now the argument for the L^2-norm

$$\left(\mathbb{E}'\left(\left| \sum_{i=1}^{N} G_i \otimes G'_i \right|^2 \right) \right)^{1/2}$$

of the Gaussian vector G. The conclusion follows. □

As a consequence of Lemma 4, for $n \leqslant m$, $\ell = 1, \ldots, 2^n$, and $q \geqslant 1$,

$$(\mathbb{E}(|W(m+1)^2_{t^n_{\ell-1},t^n_\ell} - W(m)^2_{t^n_{\ell-1},t^n_\ell}|^q))^{1/q} \leqslant Cq2^{-(m+n)/2},$$

where, here and below, $C > 0$ is a constant depending possibly only on d and p, and changing from line to line. By the triangle inequality and Fatou's lemma, we similarly have

$$(\mathbb{E}(|W(m)^2_{t^n_{\ell-1},t^n_\ell} - W^2_{t^n_{\ell-1},t^n_\ell}|^q))^{1/q} \leqslant Cq2^{-(m+n)/2}. \tag{11}$$

We turn to the case $n > m$. As a consequence of (6),

$$(\mathbb{E}(|W(m)^2_{t^n_{\ell-1},t^n_\ell}|^q))^{1/q} \leqslant Cq2^{-n}$$

for all $n \geqslant m$, $\ell = 1, \ldots, 2^n$. Hence, together with (11), we both have, for $\ell = 1, \ldots, 2^n$ and $q \geqslant 1$,

$$\sup_m(\mathbb{E}(|W(m)^2_{t^n_{\ell-1},t^n_\ell}|^q))^{1/q}, (\mathbb{E}(|W^2_{t^n_{\ell-1},t^n_\ell}|^q))^{1/q} \leqslant Cq2^{-n}, \tag{12}$$

where we recall that $C > 0$ only depends on d and p and may change from line to line. In particular,

$$(\mathbb{E}(|W(m)^2_{t^n_{\ell-1},t^n_\ell} - W^2_{t^n_{\ell-1},t^n_\ell}|^q))^{1/q} \leqslant Cq2^{-n} \tag{13}$$

for $n > m$, $\ell = 1, \ldots, 2^n$, and every $q \geqslant 1$.

Together with (11) and (13), we may conclude the proof of (7). We first have

$$\mathbb{E}(D_{2,p}(W(m), W)^q) \leqslant A(q)(Cq)^q \left(\sum_{n=1}^{m} a_n^{2q/p} 2^{n-q(m+n)/2} + \sum_{n=m+1}^{\infty} a_n^{2q/p} 2^{-(q-1)n} \right).$$

Choose now $0 < \eta < \frac{1}{2}$ such that $(1-\eta)p > 2$ and write

$$\sum_{n=1}^{m} a_n^{2q/p} 2^{n-q(m+n)/2} \leqslant 2^{-q\eta m} \sum_{n=1}^{\infty} a_n^{2q/p} 2^{-((1-\eta)q-1)n}.$$

Choose then the sequence (a_n) so that

$$a_n^{2q/p} 2^{-((1-\eta)q-1)n} = 2^{-2\varepsilon qn},$$

where $\varepsilon > 0$ is such that $0 < \varepsilon < (1-\eta)/2 - (1/p)$. It is then easy to check that $A(q) < \infty$ and moreover that $A(q)^{1/q} \leqslant C$ uniformly in q. On the other hand,

$$\sum_{n=m+1}^{\infty} a_n^{2q/p} 2^{-(q-1)n} = \sum_{n=m+1}^{\infty} 2^{-(2\varepsilon+\eta)qn} \leqslant C2^{-(2\varepsilon+\eta)qm}.$$

By (11) and (13), we thus conclude that

$$\mathbb{E}(D_{2,p}(W(m), W)^q) \leqslant A(q)(Cq)^q(2^{-q\eta m} + 2^{-(2\varepsilon+\eta)qm})$$

from which claim (10) follows. Hence (7) holds.

We are left with the proof of (8) that follows, at the level of Cameron–Martin paths, from the preceding argument for Brownian paths. By Lemma 3, it is enough to show that

$$\lim_{m\to\infty} \sup_{x;\int_0^1 |\dot{x}_t|^2\,\mathrm{d}t\leqslant\alpha} D_{j,p}(F(x(m)),F(x))=0 \tag{14}$$

for $j=1,2$. Observe first that if $\int_0^1 |\dot{x}(t)|^2\,\mathrm{d}t \leqslant \alpha$, then $|x_s - x_t| \leqslant (\alpha d|s-t|)^{1/2}$ for all $s,t\in[0,1]$. Hence, by (3) again, and with $X^1_{s,t}=x_t-x_s$,

$$\begin{aligned}
D_{1,p}(X(m)^1,X^1)^p &\leqslant \sum_{n=m+1}^{\infty} n^\gamma \sum_{\ell=1}^{2^n} |2^{m-n}\Delta^m_{k(n,m,\ell)}x - \Delta^n_\ell x|^p\\
&\leqslant C_p \sum_{n=m+1}^{\infty} n^\gamma \sum_{\ell=1}^{2^n} (2^{(m-n)p}2^{-mp/2}+2^{-pn/2})\\
&\leqslant C_p \sum_{n=m+1}^{\infty} n^\gamma 2^{-(p-2)n/2},
\end{aligned}$$

where $C_p>0$ only depends on α, p and d. Hence (14) holds for $j=1$.

We turn to $j=2$. Again, if $\int_0^1 |\dot{x}_t|^2\,\mathrm{d}t\leqslant\alpha$, we get as before from (5) that for $n\leqslant m$ and $\ell=1,\dots,2^n$,

$$\begin{aligned}
|X(m+1)^2_{t^n_{\ell-1},t^n_\ell} - X(m)^2_{t^n_{\ell-1},t^n_\ell}| &\leqslant \sum_{r=2^{m-n}(\ell-1)+1}^{2^{m-n}\ell} |\Delta^{m+1}_{2r-1}x|\,|\Delta^{m+1}_{2r}x|\\
&\leqslant (\alpha d 2^{-(m+1)})^{1/2} \sum_{r=2^{m-n}(\ell-1)+1}^{2^{m-n}\ell} |\Delta^{m+1}_{2r-1}x|\\
&\leqslant (\alpha d 2^{-(m+1)})^{1/2} \int_{t^n_{\ell-1}}^{t^n_\ell} |\dot{x}(t)|\,\mathrm{d}t\\
&\leqslant \alpha\sqrt{d}2^{-(m+n+1)/2}.
\end{aligned}$$

When $n\geqslant m$, we get from (6) that

$$|X(m+1)^2_{t^n_{\ell-1},t^n_\ell} - X(m)^2_{t^n_{\ell-1},t^n_\ell}| \leqslant 2\alpha\sqrt{d}2^{-n}.$$

We thus obtain that

$$\begin{aligned}
&D_{2,p}(X(m+1)^2,X(m)^2)^{p/2}\\
&\quad=\sum_{n=1}^{\infty} n^\gamma \sum_{\ell=1}^{2^n} |X(m+1)^2_{t^n_{\ell-1},t^n_\ell} - X(m)^2_{t^n_{\ell-1},t^n_\ell}|^{p/2}\\
&\quad\leqslant C_p\alpha\sqrt{d}\left(\sum_{n=1}^{m} n^\gamma \sum_{\ell=1}^{2^n} 2^{-p(m+n)/4} + \sum_{n=m+1}^{\infty} n^\gamma\, 2^{-(p-2)n/2}\right).
\end{aligned}$$

The left-hand side of the preceding inequality defines the general term of a convergent series (in m). Thus, by a Cauchy argument, we may define X^2 as the limit in the $p/2$-variation metric of the $X(m)^2$'s. Furthermore,

$$D_{2,p}(X(m)^2, X^2)^{p/2} \leqslant C_p \alpha \sqrt{d} \varepsilon_m$$

with $\varepsilon_m \to 0$ as $m \to \infty$. Hence, the canonical rough path of order 2 $X = (1, X^1, X^2)$ above any element x in H is well defined as the limit in the metric d_p of the dyadic polygonal rough paths $X(m) = (1, X(m)^1, X(m)^2)$, and (8) is satisfied. The proof of Theorem 1 is thus complete. □

3. The support theorem

In this section, we establish the support theorem for Brownian motion as a canonical rough path in $G\Omega_p(\mathbb{R}^d)$, $2 < p < 3$, and as a corollary, the support theorem for diffusion processes. Recall from the introduction that $\mathscr{D}$ denotes the collection of all dyadic rough paths (of order $[p] = 2$) and $\mathscr{D}_p(\mathbb{R}^d)$ its closure in the p-variation metric.

Theorem 5. *Let W be the canonical geometric rough path associated with Brownian motion. Then the support of the distribution of W is the closure of $\mathscr{D}$ in the p-variation topology*, $2 < p < 3$.

As a consequence of Lyons's continuity theorem (Lyons, 1998), we immediately get the following corollary to the support theorem for diffusions that includes, by projection on the first level, the Stroock–Varadhan result.

Corollary 6. *Consider the stochastic differential equation*

$$\mathrm{d}y_t = f_0(t, y_t)\,\mathrm{d}t + f(t, y_t) \circ \mathrm{d}w_t, \qquad y_0 = x_0,$$

driven by a Brownian motion w where f_0, f are C^3 with linear growth. Then the support of the distribution of the rough path $Y = \varphi(W)$ associated to the solution of this equation is the closure of $\varphi(\mathscr{D})$ under the p-variation distance, $2 < p < 3$, *where we recall that $\varphi : G\Omega_p(\mathbb{R}^d) \to G\Omega_p(\mathbb{R}^d)$ is the continuous extension of the Itô map to $G\Omega_p(\mathbb{R}^d)$.*

To prove Corollary 6, note that since the support of the distribution of W contains the closure of $\mathscr{D}$, by continuity of φ, the support of $Y = \varphi(W)$ contains the closure of $\varphi(\mathscr{D})$. Conversely, since W is the almost sure limit of the dyadic rough paths $W(m)$ in the p-variation metric, by continuity again $Y = \varphi(W)$ is the limit of the $\varphi(W(m))$'s so that the support of Y is included in the closure of $\varphi(\mathscr{D})$.

We thus concentrate on the proof of Theorem 5. While, again, the first level path corresponds to the classical argument, several technical details are required to properly handle the second level path.

Proof of Theorem 5. Since W is the almost sure limit in the p-variation metric, $2 < p < 3$, of the dyadic rough paths $W(m)$, W belongs almost surely to $\mathscr{D}_p(\mathbb{R}^d)$. Therefore, the theorem will be established once we show that for any $\varepsilon > 0$ and $U \in \mathscr{D}$,

$$\mathbb{P}(d_p(W, U) \leqslant \varepsilon) > 0. \tag{15}$$

Indeed, if the support of the distribution of W would not contain the closure $\mathscr{D}_p(\mathbb{R}^d)$ of $\mathscr{D}$, there would exist a closed set Γ with $\mathbb{P}(W \in \Gamma) = 1$ not containing $\mathscr{D}_p(\mathbb{R}^d)$, thus contradicting (15).

To establish (15), let us first deal with the case $U = 0$, and indicate then the modifications that have to be supplied after the use of the Cameron–Martin formula. By Lemma 3, it is enough to prove in this case that for every $\varepsilon > 0$,

$$\mathbb{P}(D_{j,p}(W) \leqslant \varepsilon, j = 1, 2) > 0, \tag{16}$$

where we recall that $D_{j,p}(X) = D_{j,p}(X, 0)$.

Let us illustrate first the idea of the proof by showing that

$$\mathbb{P}(D_{1,p}(W)^p \leqslant \varepsilon) > 0.$$

By definition of $D_{1,p}$,

$$\mathbb{P}(D_{1,p}(W)^p \leqslant \varepsilon) \geqslant \mathbb{P}\left(|W^1_{t^n_{\ell-1},t^n_\ell}| \leqslant \left(\frac{\varepsilon}{2^{n+1}n^{\gamma+2}}\right)^{1/p}, n \geqslant 1, \ell = 1, \dots, 2^n\right)$$

$$\geqslant \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcap_{\ell=1}^{2^n}\bigcap_{k=1}^{d}\left\{|\Delta^n_\ell w^k| \leqslant \frac{1}{\sqrt{d}}\left(\frac{\varepsilon}{2^{n+1}n^{\gamma+2}}\right)^{1/p}\right\}\right).$$

We make use of the Khatri–Sidak inequality (parts of the Gaussian correlation conjecture, see Li and Shao, 2001 and the references therein) that we state in the next lemma.

Lemma 7. *Let $(G_1, \dots, G_N)$ be a centered Gaussian random vector. For every $\lambda_1, \dots, \lambda_N \geqslant 0$,*

$$\mathbb{P}\left(\bigcap_{\alpha=1}^{N}\{|G_\alpha| \leqslant \lambda_\alpha\}\right) \geqslant \prod_{\alpha=1}^{N}\mathbb{P}(|G_\alpha| \leqslant \lambda_\alpha).$$

As a consequence of Lemma 7, denoting as before by g a standard normal variable,

$$\mathbb{P}(D_{1,p}(W)^p \leqslant \varepsilon) \geqslant \prod_{n=1}^{\infty}\prod_{\ell=1}^{2^n}\left(1 - \mathbb{P}\left(|g| > \frac{1}{\sqrt{d}}\left(\frac{\varepsilon}{2n^{\gamma+2}}\right)^{1/p}2^{(1/2-1/p)n}\right)\right)^d \tag{17}$$

which is easily seen to be positive since $p > 2$.

The full proof of (16) is similar but one has to take into account simultaneously the first and second levels. Moreover, to handle the second level by Lemma 7, we have rather to work (by continuity) with a dyadic approximation. As before

$$\mathbb{P}(D_{j,p}(W)^{p/j} \leqslant \varepsilon, j = 1, 2)$$

$$\geqslant \mathbb{P}\left(|W^j_{t^n_{\ell-1},t^n_\ell}| \leqslant \left(\frac{\varepsilon}{2^{n+1}n^{\gamma+2}}\right)^{j/p},\ j = 1, 2,\ n \geqslant 1,\ \ell = 1, \dots, 2^n\right). \tag{18}$$

As announced, by continuity in the d_p-metric, it will be enough to show that the right-hand side of (18) is strictly positive for the mth dyadic approximation $W(m)$ uniformly in m large enough. To this task, we again make use of decompositions (3)–(6). We thus get that

$$\mathbb{P}(D_{j,p}(W(m))^{p/j} \leqslant \varepsilon, j = 1,2) \geqslant \mathbb{P}(A(m) \cap B(m) \cap C(m)),$$

where

$$A(m) = \bigcap_{n=1}^{m} \bigcap_{\ell=1}^{2^n} \left\{ |\Delta_\ell^n w| \leqslant \left(\frac{\varepsilon}{2^{n+1} n^{\gamma+2}} \right)^{1/p} \right\},$$

$$B(m) = \bigcap_{n=1}^{m} \bigcap_{\ell=1}^{2^n} \left\{ \left| \sum_{\substack{r,s=2^{m-n}(\ell-1)+1 \\ r<s}}^{2^{m-n}\ell} (\Delta_r^m w \otimes \Delta_s^m w - \Delta_s^m w \otimes \Delta_r^m w) \right| \leqslant \left(\frac{\varepsilon}{2^{n+1} n^{\gamma+2}} \right)^{2/p} \right\}$$

and

$$C(m) = \bigcap_{n=m+1}^{\infty} \bigcap_{\ell=1}^{2^n} \left\{ |\Delta_{k(n,m,\ell)}^n w| \leqslant 2^{n-m} \left(\frac{\varepsilon}{2^{n+1} n^{\gamma+2}} \right)^{1/p} \right\}.$$

With respect to the first level paths, the main difficulty arises with the events $B(m)$. However, the antisymmetry of the summands in the definition of $B(m)$ reveals the lack of quadratic terms. In particular,

$$B(m) \supset \bigcap_{n=1}^{m} \bigcap_{\ell=1}^{2^n} \bigcap_{j \neq k=1}^{d} \left\{ \left| \sum_{\substack{r,s=2^{m-n}(\ell-1)+1 \\ r<s}}^{2^{m-n}\ell} \Delta_r^m w^j \otimes \Delta_s^m w^k \right| \leqslant \frac{1}{2d} \left(\frac{\varepsilon}{2^{n+1} n^{\gamma+2}} \right)^{2/p} \right\},$$

where $w = (w^1, \ldots, w^d)$. It makes thus possible to use the correlation inequality of Lemma 7 conditionally on the independent coordinates of the Brownian motion w. More precisely, successive conditional applications of Lemma 7 yield the following consequence.

Lemma 8. *Let $G^{(1)}, \ldots, G^{(d)}$ be independent Gaussian random vectors in $\mathbb{R}^N$ with respective coordinates $G^{(\gamma)} = (G_1^{(\gamma)}, \ldots, G_N^{(\gamma)})$, $\gamma = 1, \ldots, d$. Let furthermore I_β, $\beta = 1, \ldots, M$, be subsets of $\{1, \ldots, N\}^2$. Set, for $\gamma' \neq \gamma''$ and $\beta = 1, \ldots, M$, $Z_\beta^{(\gamma',\gamma'')} = \sum_{\delta,\varepsilon \in I_\beta} G_\delta^{(\gamma')} G_\varepsilon^{(\gamma'')}$, and $Z_\beta^{(\gamma',\gamma')} = 0$. Then, for every $\lambda_{\alpha,(\gamma)} \geqslant 0$ and $\mu_{\beta,(\gamma',\gamma'')} \geqslant 0$, $\alpha = 1, \ldots, N$, $\beta = 1, \ldots, M$, $\gamma, \gamma', \gamma'' = 1, \ldots, d$,*

$$\mathbb{P}\left(\bigcap_{\alpha=1}^{N} \bigcap_{\beta=1}^{M} \bigcap_{\gamma,\gamma',\gamma''=1}^{d} \{ |G_\alpha^{(\gamma)}| \leqslant \lambda_{\alpha,(\gamma)}, |Z_\beta^{(\gamma',\gamma'')}| \leqslant \mu_{\beta,(\gamma',\gamma'')} \} \right)$$

$$\geqslant \prod_{\alpha=1}^{N} \prod_{\beta=1}^{M} \prod_{\gamma,\gamma',\gamma''=1}^{d} \mathbb{P}(|G_\alpha^{(\gamma)}| \leqslant \lambda_{\alpha,(\gamma)}) \mathbb{P}(|Z_\beta^{(\gamma',\gamma'')}| \leqslant \mu_{\beta,(\gamma',\gamma'')}).$$

As a consequence of this lemma applied to the independent components $(w_1, \ldots, w_d)$ of w, $\mathbb{P}(A(m) \cap B(m) \cap C(m))$ is bounded below by the product $\Pi_1 \Pi_2 \Pi_3$ where

$$\Pi_1 = \prod_{n=1}^{m} \prod_{\ell=1}^{2^n} \left(1 - \mathbb{P}\left(|g| > \frac{1}{\sqrt{d}} \left(\frac{\varepsilon}{2n^{\gamma+2}}\right)^{1/p} 2^{(1/2-1/p)n}\right)\right)^d,$$

$$\Pi_2 = \prod_{n=1}^{m} \prod_{\ell=1}^{2^n} \left(1 - \mathbb{P}\left(\left| \sum_{\substack{r,s=2^{m-n}(\ell-1)+1 \\ r<s}}^{2^{m-n}\ell} \Delta_r^m w^1 \Delta_s^m w^2 \right| > \frac{1}{2d} \left(\frac{\varepsilon}{2^{n+1} n^{\gamma+2}}\right)^{2/p}\right)\right)^{d^2-d}$$

and

$$\Pi_3 = \prod_{n=m+1}^{\infty} \prod_{\ell=1}^{2^n} \left(1 - \mathbb{P}\left(|g| > \frac{1}{\sqrt{d}} \left(\frac{\varepsilon}{2n^{\gamma+2}}\right)^{1/p} 2^{(1-1/p)n-m/2}\right)\right)^d.$$

As in (17), both Π_1 and Π_3 are bounded below uniformly in m by a strictly positive number. Concerning Π_2, it follows from the moment inequalities of Lemma 4 that, for $n \leqslant m$, $\ell = 1, \ldots, 2^n$, and every $q \geqslant 1$,

$$\left(\mathbb{E}\left(\left| \sum_{\substack{r,s=2^{m-n}(\ell-1)+1 \\ r<s}}^{2^{m-n}\ell} \Delta_r^m w^1 \Delta_s^m w^2 \right|^q\right)\right)^{1/q} \leqslant Cq 2^{-n}$$

for some numerical constant $C > 0$. Therefore, for some $C' > 0$ and every $t > 0$,

$$\mathbb{P}\left(\left| \sum_{r<s=2^{m-n}(\ell-1)+1}^{2^{m-n}\ell} \Delta_r^m w^1 \otimes \Delta_s^m w^2 \right| > t\right) \leqslant C' \exp(-2^n t/C').$$

One then observes in the same way that Π_2 is bounded below uniformly in m by a strictly positive number.

This concludes the proof of (16), that is of the theorem when $U = 0$. For an arbitrary U in $\mathscr{D}$, the Cameron–Martin formula implies that

$$\mathbb{P}(D_{j,p}(W, U) \leqslant \varepsilon, j = 1, 2) \geqslant \int_E M \, \mathrm{d}\mathbb{P},$$

where M is the (strictly positive) Cameron–Martin density and where

$$E = \{D_{j,p}(W) \leqslant c\varepsilon, \ j = 1, 2, \ D_{2,p}(A) \leqslant c\varepsilon, \ D_{2,p}(B) \leqslant c\varepsilon\}.$$

Here $c > 0$ and A and B are the processes

$$A_{s,t} = \int_s^t (h_u - h_s) \otimes \mathrm{d}w_u, \qquad B_{s,t} = \int_s^t (w_u - w_s) \otimes \mathrm{d}h_u$$

that come from the Lévy area of the process $w + h$. Moreover,

$$D_{2,p}(A) = \left(\sum_{n=1}^{\infty} n^\gamma \sum_{\ell=1}^{2^n} |A_{t_{\ell-1}^n, t_\ell^n}|^{p/2}\right)^{2/p}$$

and similarly for $D_{2,p}(B)$. Now, one has just to repeat the previous argument. Since A and B are Gaussian processes continuous in the $(p/2)$-variation, although somewhat lengthy, the arguments are easily modified to handle this case completely similarly. The proof of Theorem 5 is thus complete. □

4. Extensions and applications

In this last section, we informally discuss several extensions and applications that may be obtained on the basis of the material of the preceding sections.

First, following the recent contribution (Ledoux et al., 2002), one may consider the case of Banach space valued Brownian motion. Essentially, the framework of exact tensor norms described in Ledoux et al. (2002), in order the Lévy area to exist may be used to formulate and establish Schilder's large deviation principle and the support theorem for the rough path of Banach space valued Brownian motion.

The arguments developed in this paper for large deviations apply in the same way to fractional Brownian motion provided its Hurst parameter is strictly greater than one-third thanks to the investigation (Coutin and Qian, 2002). Presumably, the support theorem would also hold, but this claim would certainly require extreme work.

At the borderline between large deviations and support theorems, Onsager–Machlup functionals provide another source of examples of interest to which the rough path technique may be useful (see Lyons and Zeitouni, 1999 for an early investigation).

A typical application (cf. Deuschel and Stroock, 1989) of Schilder's large deviation concerns Strassen's functional law of the iterated logarithm for Brownian motion. Again, on the basis of the results of Section 2, one can formulate a rough path version of Strassen's law in the p-variation topology, $2 < p < 3$. For $n \geqslant 3$, define

$$\xi_t^n = (2n \log\log n)^{-1/2} w_{nt}, \qquad t \in [0,1].$$

Denote by Ξ^n the canonical rough path associated to ξ^n and let K denote the set of (canonical) rough paths $X=(1,X^1,X^2)$ above x in the unit ball of the Cameron–Martin Hilbert H.

Theorem 9. *For every* p, $2 < p < 3$, *almost surely*,

$$\lim_{n\to\infty} d_p(\Xi^n, K) = 0$$

and the set of limit points of the sequence (Ξ^n) *in* $G\Omega_p(\mathbb{R}^d)$ *is equal to* K.

Acknowledgements

We thank T. Lyons for his encouragements and his suggestion to investigate the support theorem in this context, and all the comments on this work. Part of this work was done while the second-named author was visiting the Department of Mathematics at Cornell University.

References

Bass, R.F., Hambly, B.M., Lyons, T., 1998. Extending the Wong–Zakai theorem to reversible Markov processes. J. European Math. Soc., to appear.

Coutin, L., Qian, Z., 2002. Stochastic analysis, rough path analysis and fractional Brownian motion. Probab. Theory Relat. Fields 122, 108–140.

Dembo, A., Zeitouni, O., 1990. Large Deviations Techniques and Applications, 1st Edition. Jones & Bartlett Publishers, Boston, MA, 1990; 2nd Edition, Springer, Berlin, 1998.

Deuschel, J.-D., Stroock, D., 1989. Large Deviations. Academic Press, New York.

Freidlin, M., Wentzell, A., 1984. Random Perturbations of Dynamical Systems. Springer, Berlin.

Hambly, B.M., Lyons, T., 1998. Stochastic area for Brownian motion on the Sierpinsky gasket. Ann. Probab. 26, 132–148.

Ledoux, M., 1996. Isoperimetry and Gaussian analysis. Ecole d'Eté de Probabilités de St-Flour 1994, Lecture Notes in Mathematics, Vol. 1648. Springer, Berlin, pp. 165–294.

Ledoux, M., Talagrand, M., 1991. Probability in Banach Spaces. Springer, Berlin.

Ledoux, M., Lyons, T., Qian, Z., 2002. Lévy area of Wiener processes in Banach spaces. Ann. Probab. 30, 546–578.

Li, W., Shao, Q.-M., 2001. Gaussian processes: inequalities, small ball probabilities and applications. Handbook Statist. 19, 533–597.

Lyons, T., 1998. Differential equations driven by rough signals. Rev. Math. Iberoamericana 14, 215–310.

Lyons, T., Qian, Z. System Control and Rough Paths. Oxford University Press, Oxford, to appear.

Lyons, T., Zeitouni, O., 1999. Conditional exponential moments for iterated Wiener integrals, and applications to the Onsager–Machlup functional of diffusions. Ann. Probab. 27, 1738–1749.

Malliavin, P., 1997. Stochastic Analysis. Springer, Berlin.

Millet, A., Sanz-Solé, M., 1994. A simple proof of the support theorem for diffusion processes. Séminaire de Probabilités XXVIII. Lecture Notes in Mathematics, Vol. 1583. Springer, Berlin, pp. 36–48.

Sippilainen, E.-M., 1993. A pathwise view of solution of stochastic differential equations. Ph.D. Thesis, University of Edinburgh.

Stroock, D., Varadhan, S.R.S., 1972. On the support of diffusion processes with applications to the strong maximum principle. Proceedings of the Sixth Berkely Symposium Mathematical Statistics Probability III. University California Press, pp. 333–359.

Varadhan, S.R.S., 1984. Large Deviations and Applications. SIAM Publications, Philadelphia, PA.