A Large Deviation Principle for Martingales over Brownian Filtration

BY Z. QIAN AND C. XU
Mathematical Institute, University of Oxford

April 3, 2009

Abstract. In this article we establish a large deviation principle for the family \( \{ \nu_\varepsilon : \varepsilon \in (0, 1) \} \) of distributions of the scaled stochastic processes \( \{ P_{- \log \sqrt{\varepsilon} Z_t} \}_{t \leq 1} \), where \((Z_t)_{t \in [0, 1]}\) is a square-integrable martingale over Brownian filtration and \((P_t)_{t \geq 0}\) is the Ornstein-Uhlenbeck semigroup. The rate function is identified as well in terms of the Wiener-Itô chaos decomposition of the terminal value \( Z_1 \). The result is established by developing a continuity theorem for large deviations, together with two essential tools, the hypercontractivity of the Ornstein-Uhlenbeck semigroup and Lyons’ continuity theorem for solutions of Stratonovich type stochastic differential equations.

Key words: Brownian filtration, chaos decomposition, hypercontractivity, large deviation principle, martingales, rough path, Itô’s mapping

2000 Mathematics Subject Classification: 60F10, 60H10
1 Introduction

In 1938, H. Cramér [19] published a result on the probability of large deviations in the law of large numbers for sums of independent real random variables, and some years later, H. Chernoff [6] proved a general Cramér’s theorem. Cramér’s paper marked the beginning of the study of large deviations of distributions towards their limiting law. Schöffer [35], mainly developed from his Ph. D. thesis, proved a beautiful theorem for large deviations of Brownian motion, and thus opened the study of large deviations for probability measures on spaces of continuous paths. Schöffer’s analysis in [35] proved fundamental in dealing with functional integrations over function spaces. It took some years, however, in particular in the hands of Azencott [1], Donsker-Varadhan [38], [39], Freidlin-Ventcel [15], [37], Stroock [36], Deuschel-Stroock [8], Dembo-Zeitouni [7], Dupuis-Ellis [11] and etc. to turn the results of large deviations and the techniques developed to prove them into what nowadays we may call the theory of large deviations.

Large deviation principles have been established for a large class of distributions, mainly by exploiting Markov property or/and Gaussian nature of underlying stochastic processes, see [1], [2], [4], [7], [8], [9], [10], [13], [15], [25], [31], [32] etc. for a small sample.
Let $W^d_0 = C_0([0, 1]; R^d)$ be the Banach space of all continuous paths in $R^d$ started at 0 with running time $[0, 1]$, equipped with the uniform norm

$$||w|| = \sup_{t \in [0, 1]} |w(t)| \quad \forall w \in W^d_0.$$ 

Let $H = H^1_0([0, 1]; R^d)$ be the subspace of $h \in W^d_0$ such that its generalized derivative $\dot{h} \in L^2[0, 1]$. $H$ is a Hilbert space under the norm

$$||h||_{H^1} = \sqrt{\int_0^1 |\dot{h}(t)|^2 dt} \quad \forall h \in H.$$ 

Let $(w_t)_{t \geq 0}$ be the coordinate process on $W^d_0$

$$w_t(x) = x(t) \quad \forall x \in W^d_0 \text{ and } t \in [0, 1]$$

and $\mathcal{F}_t^0 = \sigma \{ w_s : s \leq t \}$ be the filtration generated by $(w_t)_{t \geq 0}$. Then $\mathcal{F}_1^0$ coincides with the Borel $\sigma$-algebra $\mathcal{B}(W^d_0)$ on $W^d_0$. The Wiener measure $P^w$ (see for example [23]), where the superscript $w$ is an attribute to Wiener who first constructed the law of Brownian motion as a measure on the space of continuous paths, is the unique probability on $(W^d_0, \mathcal{B}(W^d_0))$ such that the coordinate process $(w_t)$ is a Brownian motion started from 0. Another, but equivalent, description of $P^w$, is that $P^w$ is the unique probability on $(W^d_0, \mathcal{B}(W^d_0))$ with characteristic function

$$\int_{W^d_0} e^{\sqrt{-1}l(x)} P^w(dx) = e^{-\frac{1}{2}||l||^2_{H^1}} \quad \forall l \in (W^d_0)^*$$

where $(W^d_0)^*$ is the dual space of $W^d_0$, and the natural imbedding $(W^d_0)^* \hookrightarrow H \hookrightarrow W^d_0$ has been used.

The Hilbert space $H^1_0([0, 1]; R^d)$ is called the Cameron-Martin space, and the probability space $(W^d_0, \mathcal{F}_1, P^w)$ is called the Wiener space on $R^d$, where $\mathcal{F}_1$ is the completion of $\mathcal{F}_1^0$ under $P^w$, and $(\mathcal{F}_t)_{t \in [0, 1]}$ is the smallest $\sigma$-algebra containing $\mathcal{F}_t^0$ and the events in $\mathcal{F}_1$ with probability zero. $(\mathcal{F}_t)_{t \in [0, 1]}$ is the Brownian filtration.

For each $\varepsilon > 0$, $P^w_\varepsilon$ denotes the distribution of the scaled Brownian motion $(\sqrt{\varepsilon}w_t)_{0 \leq t \leq 1}$, that is, $P^w_\varepsilon$ is the probability measure on $(W^d_0, \mathcal{F}_1)$ such that

$$\int_{W^d_0} e^{\sqrt{-1}l(x)} P^w_\varepsilon(dx) = \int_{W^d_0} e^{\sqrt{-1}(\sqrt{\varepsilon}l(x))} P^w(dx)$$

$$= e^{-\frac{\varepsilon}{2}||l||^2_{H^1}} \quad \forall l \in (W^d_0)^*.$$
It is obvious that, as $\varepsilon \downarrow 0$, $P_{\varepsilon}^w$ approaches zero (the probability measure with the support containing only one path: $x(t) = 0$ for all $t$), at an exponential rate. The family of distributions, $\{P_{\varepsilon}^w : \varepsilon > 0\}$, satisfies the large deviation principle with rate function

$$I(h) = \begin{cases} \frac{1}{2} \int_0^1 |\dot{h}|^2(t)dt & \text{if } h \in H, \\ \infty & \text{otherwise}. \end{cases}$$

(1.1)

By a large deviation principle with rate function $I$, we mean that

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log P_{\varepsilon}^w(F) \leq -\inf_{w \in F} I(w)$$

(1.2)

and

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log P_{\varepsilon}^w(O) \geq -\inf_{w \in O} I(w)$$

(1.3)

for any closed subset $F$ and open set $O$. See [7], [8] for further information about the general theory of large deviations.

Brownian motion is a typical example among Markov processes and Gaussian processes, and is also a good example of continuous martingales. It is thus natural to seek for large deviation results for laws of properly scaled martingales. To the best knowledge of the present authors, it remains an open question whether a large deviation principle holds for martingales, see however [17], [18], [27] and the references therein for results on exponential tail estimates for martingales in discrete-time.

This article presents a solution of this problem: we are going to establish a large deviation principle for square-integrable martingales over the Brownian filtration.

Consider a square-integrable martingale $(Y_t)_{t \in [0,1]}$ (with initial zero) on $(W^d_0, \mathcal{F}_1, \mathcal{F}_t, P^w)$, then, by the martingale representation theorem (Theorem 3.5, page 201, [24]), $(Y_t)_{t \in [0,1]}$ is continuous and can be represented as an Itô integral against Brownian motion, i.e.

$$Y_t = \int_0^t f_s dw_s$$

where $(f_t)_{t \geq 0}$ is a predictable process on $(W^d_0, \mathcal{F}_1, \mathcal{F}_t, P^w)$. In particular, $(Y_t)_{t \geq 0}$ is a measurable function of Brownian motion $(w_t)_{t \geq 0}$.

It is obvious that the scaling $\sqrt{\varepsilon}$, which is correct for Brownian motion, does not apply to an arbitrary martingale. Consider the Wiener-Itô chaos decomposition (see [24], [10]) of a square-integrable martingale...
\( Y_t = P^w(Z_t|\mathcal{F}_t) \) with mean zero, where \( P^w(\cdot|\mathcal{F}_t) \) is the conditional expectation. Since \( Y_1 \in L^2(\mathcal{W}^d, \mathcal{F}_1, P^w) \), so that (for simplicity, let us consider the case that \( d = 1 \), but our arguments equally apply to higher dimensions)

\[
Y_1 = \sum_{k=1}^{\infty} \int_{0 < t_1 < \cdots < t_k < 1} f_k(t_1, \cdots, t_k) \, dw_{t_1} \cdots dw_{t_k}
\]  
(1.4)

where \( k \)-th term, a multiple Wiener-Itô integral, belongs to \( k \)-th Wiener chaos, and the integrands \( f_k \) are symmetric functions in \( L^2[0,1]^k \). Clearly

\[
Y_t = \sum_{k=1}^{\infty} \int_{0 < t_1 < \cdots < t_k < t} f_k(t_1, \cdots, t_k) \, dw_{t_1} \cdots dw_{t_k}.
\]

According to Schilder’s theorem, one simple while reasonable re-scaling for such martingale is multiplying \( k \)-th term in the decomposition by \( \varepsilon^{k/2} \). Therefore, one possible scaling for martingales should be \( P_{-\log \sqrt{\varepsilon} Y_t} \), where \( (P_t)_{t \geq 0} \) is the Ornstein-Uhlenbeck semigroup on \( (\mathcal{W}^d, \mathcal{F}_1, P^w) \). This is the first place the Ornstein-Uhlenbeck semigroup comes into our study.

Let us define the following mapping \( F : H^{1,0}_0([0,1]; \mathbb{R}^1) \rightarrow \mathcal{W}^d_0 \) by

\[
F(h)_t = \sum_{k=1}^{\infty} \int_{0 < t_1 < \cdots < t_k < t} f_k(t_1, \cdots, t_k) h(t_1) \cdots h(t_k) \, dt_1 \cdots dt_k.
\]

The main result of the paper is the following large deviation principle.

**Theorem 1.1** Let \( \xi \in L^2(\mathcal{W}^1, \mathcal{F}_1, P^w) \) which has the Wiener-Itô’s decomposition

\[
\xi = \sum_{k=1}^{\infty} \int_{0 < t_1 < \cdots < t_k < 1} f_k(t_1, \cdots, t_k) \, dw_{t_1} \cdots dw_{t_k}.
\]

Let \( Y_t = P^w(\xi|\mathcal{F}_t) \) for \( t \in [0,1] \), and \( \nu_\varepsilon \) be the distribution of \( (P_{-\log \sqrt{\varepsilon} Y_t})_{t \leq 1} \). Then \( \{\nu_\varepsilon : \varepsilon \in (0,1)\} \) satisfies the large deviation principle with rate function

\[
I'(w) = \inf \{ I(h) \mid h \in H \text{ such that } F(h) = w \},
\]

where \( I(h) = \frac{1}{2} ||h||_{H^1}^2 \), for \( h \in H \).

A special case of the above theorem, namely, for distributions of multiple Wiener-Itô’s integrals of the following form

\[
Y_t = \int_{0 < t_1 < \cdots < t_k < t} f(t_1, \cdots, t_k) \, dw_{t_1} \cdots dw_{t_k}
\]
has been established in M. Ledoux \[25\]. Nualart and et al. \[31\] extended to an even larger class of multiple Wiener-Itô's integrals on an abstract Wiener space. These authors used the same scaling \(\varepsilon_k^2\) but only for single iterated integrals, we however believe that their methods may be modified to develop a large deviation principle for finite sums of multiple Wiener-Itô integrals, or more precisely for the laws of martingales of the following form

\[
Y\varepsilon_t^k = \sum_{k=1}^{N} \varepsilon_k^2 \int_{0<t_1<\ldots<t_k<t} f_k(t_1, \ldots, t_k)dw_{t_1} \cdots dw_{t_k}.
\]

Our study is based on the following simple observation: the scaling for \(k\)-th term \(\varepsilon_k^2\) is itself sub-exponential for large \(k\), which ensures the laws of martingales

\[
Y\varepsilon_t^\infty = \sum_{k=1}^{\infty} \varepsilon_k^2 \int_{0<t_1<\ldots<t_k<t} f_k(t_1, \ldots, t_k)dw_{t_1} \cdots dw_{t_k},
\]

though the sum is infinite, remain to satisfy the large deviation principle.

In this stage we would like to describe the main steps of our proof of Theorem \[1.1\] which are necessary long and involve many technical issues. The first step is, of course, to approximate \(Y\varepsilon_t^k\) by a good family of martingales. More precisely, let \(\xi \in L^2(W_0^d, \mathcal{F}_1, P^w)\) have the decomposition

\[
\xi = \sum_{k=1}^{\infty} \int_{0<t_1<\ldots<t_k<1} f_k(t_1, \ldots, t_k)dw_{t_1} \cdots dw_{t_k}
\]

so that

\[
||\xi||_{L^2}^2 = \sum_{k=1}^{\infty} \frac{1}{k!} ||f_k||_{L^2[0,1]^k}^2 < \infty.
\]

For each \(n\) we may choose an \(N_n\) and

\[
\xi_n = \sum_{k=1}^{N_n} \int_{0<t_1<\ldots<t_k<1} f_k^n(t_1, \ldots, t_k)dw_{t_1} \cdots dw_{t_k}.
\]

such that \(\xi_n \to \xi\) in \(L^2(W_0^d, \mathcal{F}_1, P^w)\). The symmetric functions \(f_k^n\) may be chosen such that \(f_k^n \to f_k\) in \(L^2\) for each \(k\) as \(n \to \infty\). We can require that
all $f^n_k$ are smooth enough with bounded derivatives (up to order 4 is enough) on $[0,1]$, and moreover, we can assume that $f^n_k$ have a product form

$$f^n_k(t_1, \cdots, t_n) = \sum_{j_1, \cdots, j_k=1}^{N_n} C^{n,k}_{j_1, \cdots, j_k} f^k_j(t_1) \cdots f^k_{j_k}(t_k)$$

where $C^{n,k}_{j_1, \cdots, j_k}$ are constants and $N_n$ are natural numbers. Thanks to the hypercontractivity of the Ornstein–Uhlenbeck semigroup (see L. Gross [20] for more details), the corresponding martingales $Y(n)^\varepsilon_t = P_w(\xi_n | \mathcal{F}_t)$ converges to $Y^\varepsilon$ exponentially.

The next step is to show that $Y(n)^\varepsilon$ for each $n$ satisfies the large deviation principle, and to identify its rate function explicitly, which will be achieved by using Lyons’ continuity theorem ([28], see also [29], and excellent recent books [16], [30] etc.), Schilder’s large deviation principle in $p$-variation distance (see [26]) together with a simple application of Varadhan’s contraction principle. More precisely we demonstrate that, for each $n$, $Y(n)^\varepsilon$ may be realized (or more precisely lifted) as a continuous function on the space of geometric rough paths, with respect to a variation distance. This is the precise version of what belonging to the folklore that Startonovich’s integrals are continuous functions of Brownian motion paths. However we should emphasize that the continuity here must be understood in terms of Lyons’ $p$-variation distance, rather than the uniform norm, see Proposition 4.6 below for a more precise statement.

Nevertheless, it turns out that the rate function governing the large deviations of $\{Y(n)^\varepsilon : \varepsilon \in (0,1)\}$ is given by

$$I_n(w) = \inf \{I(h) : F_n(h) = w\}$$

where

$$F_n(h)_t = \sum_{k=1}^{N_n} \int_{0 < t_1 < \cdots < t_k < t} f^n_k(t_1, \cdots, t_k) \dot{h}(t_1) \cdots \dot{h}(t_k) dw_{t_1} \cdots dw_{t_k}.$$ 

It is easy to see that $F_n$ converges uniformly on any level set of $I$ with respect to the uniform norm, but in the uniform norm $F_n$ is not continuous from $H \subset W^d_0$ to $W^d_0$. Indeed there is no continuous extension of $F_n$ to the whole space $W^d_0$ in general. On the other hand, we may lift the mappings $F_n$ to the space of rough paths, that is, $F_n$ is continuous in the
$p$-variation distance, but then in general $F_n$ does not converge with respect to the $p$-variation metric as we do not have control over the derivatives of the integrands $f_k$ ($k = 1, 2, \cdots$). Therefore, the existed (extended or generalized) contraction principles, which require that $F_n$ are continuous and $F_n$ converges uniformly on level sets of $I$, do not apply to the present case to deduce a large deviation principle for $\{Y^\varepsilon : \varepsilon \in (0, 1)\}$.

The main technical tool established in Section 2, a continuity theorem for large deviations which we believe has independent interest by its own, however, allows us to prove the large deviation principle for $\{Y^\varepsilon : \varepsilon \in (0, 1)\}$. In Section 3, we show that the hypercontractivity of the Ornstein-Uhlenbeck operator allows us to establish the exponential tightness of the family of scaled martingales, which is one of the main ingredients in our proof of the main result. In Section 4, we construct the Itô-Lyons mappings associated with multiple Wiener-Itô integrals, which thus makes another key step towards the proof of Theorem 1.1. Finally in Section 5, we collect all technical estimates together to establish a large deviation principle for square-integrable martingales.

2 A continuity theorem for large deviations

An important method in the theory of large deviations is the contraction principle, formulated by S. R. S. Varadhan [39]. Suppose $\{Z^\varepsilon : \varepsilon \in (0, 1)\}$ is a family of random variables in a Polish space $E$ which satisfies the large deviation principle with a good rate function $I$, and suppose $F : E \to E'$ is a continuous mapping, where $E'$ is another Polish space, then $\{X^\varepsilon : \varepsilon \in (0, 1)\}$, where $X^\varepsilon = F(Z^\varepsilon)$, also satisfies the large deviation principle with rate function

$$I'(s') = \inf \{I(s) : s \in E \text{ such that } F(s) = s' \}$$

for any $s' \in E'$.

However, in stochastic analysis, we often deal with Wiener functionals, for example, strong solutions to stochastic differential equations, which are only measurable rather than continuous, the above contraction principle is not sufficient in applications. Different generalizations of the contraction principle, proposed by various authors over the past years, have been successfully applied to distributions of many interesting Wiener functionals. Among these generalizations, a typical one may be formulated as the following (see Theorem 4.2.23 in [7]). Suppose $F_n : E \to E'$ is a family of continuous mappings,
and \( \{X^\varepsilon : \varepsilon \in (0,1)\} \) is a family of random variables in \( E' \) on \((\Omega, \mathcal{F}, P)\), such that the continuous images \( F_n(Z^\varepsilon) \) approaches \( X^\varepsilon \) in probability at an exponential rate
\[
\lim_{n \to \infty} \varepsilon \log P \{ \rho(F_n(Z^\varepsilon), X^\varepsilon) \} = -\infty
\]
where \( \rho \) and \( \rho' \) are the distance functions on \( E \) and \( E' \) respectively. In addition, if \( F_n \) converges uniformly on any level set \( K_L \equiv \{ s : I(s) \leq L \} \), its limit is denoted by \( F \) (note that \( F \) is only well defined on the effective set \( H \equiv \{ s : I(s) < \infty \} \), but \( F \) is continuous on \( H \subset E' \)), then the distributions of \( \{X^\varepsilon : \varepsilon \in (0,1)\} \) obeys the large deviation principle with rate function
\[
I'(s') = \inf \{ I(s) : s \in H \text{ such that } F(s) = s' \}
\]
for any \( s' \in E' \).

In general we are interested in the following question. Suppose \( \{X^\varepsilon_n : \varepsilon \in (0,1)\} \) is a sequence of families of random variables in \( E \) on a probability space \((\Omega, \mathcal{F}, P)\) which converges to \( \{X^\varepsilon : \varepsilon \in (0,1)\} \) exponentially
\[
\lim_{n \to \infty} \varepsilon \log P \{ \rho(X^\varepsilon_n, X^\varepsilon) \} = -\infty
\]
Suppose for each \( n \), \( \{X^\varepsilon_n : \varepsilon \in (0,1)\} \) satisfies a large deviation principle with rate function \( I_n \). Then, according to Theorem 4.2.16, page 131, [7], the limiting distribution \( \mu^\varepsilon \) of \( X^\varepsilon \) satisfies a weak large deviation principle with rate function
\[
I_\infty(s') \equiv \sup_{\delta > 0} \lim_{n \to \infty} \inf_{s \in B(s'_0, \delta)} I_n(s)
\]
where \( B(s'_0, \delta) \) is open ball centered at \( s'_0 \) with radius \( \delta \). Furthermore, if in addition \( I_\infty \) is a good rate function and for any closed set \( S \) one has
\[
\inf_{s' \in S} I_\infty(s') \leq \limsup_{n \to \infty} \inf_{s \in S} I_n(s)
\]
then \( \{\mu^\varepsilon : \varepsilon \in (0,1)\} \) satisfies the large deviation principle with rate function \( I_\infty \).

However, in general, one can not deduce that a large deviation principle for limit processes \( \{X^\varepsilon : \varepsilon \in (0,1)\} \) under (2.3) alone.

In many applications, the rate functions \( I_n \) are often given as the images of a common good rate function \( I \) under some mappings \( F_n : H = \{ I < \infty \} \rightarrow \)
If \( F_n \) are not continuous with respect to the distance on \( E \), then the right-hand side of (2.4) is difficult to compute, and it is hard then to verify the condition (2.5). The main goal of this section is to provide useful sufficient conditions in this situation, such that the limiting processes \( \{ X^\varepsilon : \varepsilon \in (0, 1) \} \) satisfies a large deviation principle. To this end, we introduce a concept of rate-function mappings, see Definition 2.1 below, which are mappings sending a good rate function to another one.

More precisely, we handle the following situation. Suppose \( X^\varepsilon_n, X^\varepsilon \) are random variables in \( E \), \( X^\varepsilon_n \) converges to \( X^\varepsilon \) exponentially, i.e. (2.3) is satisfied, and for each \( n \), \( X^\varepsilon_n \) satisfies the large deviation principle with rate function given by

\[
I_n(s') = \inf \{ I(s) : F_n(s) = s' \}
\]

where \( I \) is a good rate function, and \( F_n \) are rate-function mappings, so that each \( I_n \) is a good rate function. In many interesting cases, \( F_n \) are not continuous in the topology on \( E \). To ensure a large deviation principle of limiting processes \( X^\varepsilon \) to hold, the main conditions we impose on the family of rate-function mappings \( \{ F_n \} \) are the followings. \( F_n \) converges uniformly on any level set \( \{ s : I(s) \leq L \} \), and all \( F_n \) are weakly continuous in a proper topology on the effective set \( H \equiv \{ s : I(s) < \infty \} \). For more details, see Theorem 2.5 below.

### 2.1 Rate-function mappings

Let \( E \) be a separable Banach space with its norm denoted by \( \| \cdot \| \). The induced distance function is denoted by \( \rho \), that is, \( \rho(s, s') = \| s - s' \| \). Let \( I \) be a good rate function on \( E \), that is, \( I : E \to [0, \infty] \) such that for each real number \( L \geq 0 \), its level set \( K_L = \{ s : I(s) \leq L \} \) is compact in \( E \).

Let \( H = \{ s : I(s) < \infty \} \) be the effective set of the definition of \( I \). We assume that

1. \( H \) is a dense vector subspace of \( E \), and there is a Hilbert norm \( \| \cdot \|_H \) on \( H \), such that \( (H, \| \cdot \|_H) \) is a Hilbert space.

2. For each real \( L \geq 0 \), \( K_L \) is weakly compact, bounded and closed in \((H, \| \cdot \|_H)\).

It is necessary that for all \( s \in H \), \( \| s \| \leq C \| s \|_H \) for some constant \( C \).
The aim of this part is to study a class of mappings $F : H \to E$ so that

$$I_F(s') = \inf \{ I(s) : s \in H \text{ such that } F(s) = s' \}$$

(2.6)

is again a good rate function on $E$.

**Definition 2.1** A mapping $F : H \to E$ is called a rate-function mapping, if the following conditions are satisfied.

1) $F$ is continuous with respect to the corresponding norms, i.e.

$$||F(s) - F(s')|| \to 0 \quad \text{as} \quad ||s - s'||_H \to 0.$$

Note that $F$ may be not continuous as a mapping $H \subset E$ to $E$ with respect to the norm $|| \cdot ||$.

2) $F : H \to E$ is weakly continuous on any $K_L$ in the following sense: if $s_k \to s$ weakly in $H$, where $s_k \in K_L$ (so that $s \in K_L$ as well), then $F(s_k) \to F(s)$ weakly in $E$.

3) For any $L \geq 0$, the range $F(K_L) = \{ s' : s \in K_L \text{ such that } F(s) = s' \}$ is compact in $(E, || \cdot ||)$.

The following is the main use of the concept of rate-function mappings.

**Proposition 2.2** If $F : H \to E$ is a rate-function mapping, then $I_F$ defined by (2.6) is a good rate function on $E$.

**Proof.** Let $K'_L = \{ s' \in E : I_F(s') \leq L \}$. We have to show that $K'_L$ is a compact subset of $E$. To this end, choose any sequence $\{ s'_n \} \subset K'_L$, such that $I_F(s'_n) \leq L$. Then there are $s_n \in H$ such that $F(s_n) = s'_n$ and $I(s_n) \leq L + \frac{1}{n}$. In particular $s_n \in K_{L+1}$, and $\{ s'_n \} \subset F(K_{L+1})$ which is compact in $E$. Therefore we may assume that $s'_n \to s'_0$ in $E$, otherwise consider a convergent subsequence instead. Since $\{ s_n \} \subset K_{L+1}$ so that we can extract a subsequence $s_{n_k} \to s_0$ weakly in $H$ as well as $s_{n_k} \to s_0$ in $E$. Since $F$ is weakly continuous on $K_{L+1}$, so that $F(s_{n_k}) \to F(s_0)$ weakly in $E$. Therefore we must have $F(s_0) = s'_0$ and, since $I$ is a good rate function, so that

$$I_F(s'_0) \leq I(s_0) \leq \lim_{n \to \infty} I(s_n) = L$$

which implies that $s'_0 \in K'_L$. Therefore $K'_L$ is compact. ■

The following proposition shows that the set of rate-function mappings is closed under the uniform convergence on level sets of $I$. 

11
Proposition 2.3 Let \( F_n : H \to E \) be a sequence of rate-function mappings. Suppose that \( F_n \) converges uniformly on \( K_L \) for any \( L \geq 0 \), and let \( F \) denote the limiting function. Then \( F \) is also a rate-function mapping.

**Proof.** As the uniform limit, \( F \) is clearly continuous from \((H, || \cdot ||_H)\) to \((E, || \cdot ||)\). To show that \( F \) is weakly continuous on \( K_L \), consider any \( s_k, s \in K_L \), \( s_k \to s \) weakly in \( H \). Since \( F_n \to F \) uniformly on \( K_L \), for every \( \epsilon > 0 \), there is an \( N_1 \) such that

\[
||F_n(s) - F(s)|| < \frac{\epsilon}{3} \quad \forall n \geq N_1 \quad \forall s \in K_L. \tag{2.7}
\]

Let \( \xi \in E^* \). Then

\[
|\langle \xi, F(s_k) - F(s) \rangle| \leq |\langle \xi, F(s_k) - F_n(s_k) \rangle| + |\langle \xi, F_n(s) - F(s) \rangle| + |\langle \xi, F_n(s_k) - F_n(s) \rangle| \\
\leq \frac{2\epsilon}{3} ||\xi||_{E^*} + ||\langle \xi, F_n(s_k) - F_n(s) \rangle|| \\
\to \frac{2\epsilon}{3} ||\xi||_{E^*} \quad \forall n \geq N_1
\]
as \( k \to \infty \), so that \( F \) is weakly continuous on \( K_L \).

Next we prove that \( F(K_L) \) is compact in \( E \). Consider any sequence \( \{s_k'\} \subset F(K_L) \), so that \( F(s_k) = s_k' \) for some \( s_k \in K_L \). For any \( \epsilon > 0 \), there is an \( N_1 \) such that \[(2.7)\] holds. Hence

\[
||F(s_k) - F(s_l)|| \leq \frac{2\epsilon}{3} + ||F_{N_1}(s_k) - F_{N_1}(s_l)|| \quad \forall k, l.
\]

Since \( F_{N_1}(K_L) \) is compact, we may assume that \( \{F_{N_1}(s_k)\} \) is convergent, so that there is an \( N_2 \) such that

\[
||F_{N_1}(s_k) - F_{N_1}(s_l)|| \leq \frac{\epsilon}{3} \quad \forall k, l \geq N_2
\]

and therefore

\[
||F(s_k) - F(s_l)|| \leq \frac{\epsilon}{3} \quad \forall k, l \geq N_1 \vee N_2.
\]

Hence \( F(s_k) \to s' \) in \( E \) for some \( s' \in E \). We need to show that \( s' \in F(K_L) \).

Since \( K_L \) is a compact subset of \( E \) and is weakly compact in \( H \), we may assume that \( \rho(s_k, s_0) \to 0 \) for some \( s_0 \in K_L \), and \( s_k \to s_0 \) weakly in \( H \) as
well, otherwise considering a convergent subsequence instead. By (2.7) we have
\[ ||F_n(s_k) - F(s_k)|| \leq \frac{\varepsilon}{3} \quad \forall n \geq N_1 \]
so that, for any \( \xi \in E^* \)
\[ |\langle \xi, F_n(s_k) - F(s_k) \rangle| \leq \frac{\varepsilon}{3} ||\xi||_{E^*} \quad \forall n \geq N_1. \]
Letting \( k \to \infty \), then \( \langle \xi, F_n(s_k) \rangle \to \langle \xi, F_n(s_0) \rangle \) and \( f(s_k) \to s' \) so that
\[ |\langle \xi, F_n(s_0) - s' \rangle| \leq \frac{\varepsilon}{3} ||\xi||_{E^*} \quad \forall n \geq N_1. \]
Letting \( n \to \infty \) in the above inequality, to obtain
\[ |\langle \xi, F(s_0) - s' \rangle| \leq \frac{\varepsilon}{3} ||\xi||_{E^*} \]
for any \( \varepsilon > 0 \) and \( \xi \in E^* \). Therefore we must have \( F(s_0) = s' \), so that \( s' \in F(K_L) \).

We end this sub-section by showing some examples of rate-function mappings.

**Proposition 2.4** Let \( E = C_0([0,1]; R^1) \) be the Banach space of all continuous paths in \( R^1 \) starting from zero, endowed with the uniform norm \( ||s|| = \sup_{t \in [0,1]} |s(t)| \), and \( H \) be the subspace of all paths \( s \in E \) which has a generalized derivative \( \dot{s} \in L^2[0,1] \), together with the Hilbert norm \( ||s||_H = \sqrt{\int_0^1 |\dot{s}(t)|^2 dt} \). Then \( H \) is a Hilbert space which is dense in \( E \). Let \( I(s) = \frac{1}{2} ||s||_H^2 \) if \( s \in H \), otherwise \( I(s) = \infty \). Then \( I \) is a good rate function on \( E \) with effective set \( H \).

Let \( f_n \in L^2[0,1]^n \) be symmetric functions \( (n = 1, 2, \cdots) \) such that
\[ \sum_{n=1}^{\infty} \frac{1}{n!} ||f_n||_{L^2[0,1]^n}^2 < \infty. \quad (2.8) \]
Define \( F_N \) and \( F : H \to E \) by
\[ F_N(h)_t = \sum_{n=1}^{N} \int_{0<t_1<\cdots<t_n<t} f_n(t_1, \cdots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n \quad (2.9) \]
and
\[ F(h) = \sum_{n=1}^{\infty} \int_{0 < t_1 < \cdots < t_n < t} f_n(t_1, \ldots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n \quad (2.10) \]
for \( t \in [0, 1] \), respectively. Then
\begin{enumerate}
  \item For any \( L \geq 0 \), \( F_N \) converges to \( F \) uniformly on \( K_L \) in \((E, ||\cdot||)\). That is
    \[ \sup_{h \in K_L} ||F_N(h) - F(h)|| \to 0 \quad \text{as} \quad N \to \infty. \quad (2.11) \]
  \item All \( F_N, F \) are rate-function mappings.
\end{enumerate}

**Proof.** Let \( K_L = \{ h \in E : I(h) \leq L \} \) be the level set of the rate function \( I \). First of all, we note that for each \( L \geq 0 \), \( K_L \) is a closed ball in \( H \), and therefore \( K_L \) is not only compact in \( E \) (by the Sobolev imbedding), \( K_L \) is also convex and bounded in \( H \), so that \( K_L \) is weakly compact in \( H \), according to Milman’s theorem and Theorem 1, page 126, [41].

It is easy to see that each \( F_N \) is continuous from \((H, ||\cdot||_H)\) to \((E, ||\cdot||)\).

Let us prove that \( F_N \to F \) uniformly on any \( K_L \). Let \( h \in K_L \). Then, by the Cauchy-Schwartz inequality
\[
\left| \sum_{n=N+1}^{\infty} \int_{0 < t_1 < \cdots < t_n < t} f_n(t_1, \ldots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n \right|
\leq \sqrt{ \sum_{n=N+1}^{\infty} \int_{0 < t_1 < \cdots < t_n < t} f_n(t_1, \ldots, t_n)^2 dt_1 \cdots dt_n } \times \sqrt{ \sum_{n=N+1}^{\infty} \int_{0 < t_1 < \cdots < t_n < t} |\dot{h}(t_1) \cdots \dot{h}(t_n)|^2 dt_1 \cdots dt_n }
= \sqrt{ \sum_{n=N+1}^{\infty} \frac{1}{n!} ||f_n||_{L^2[0,1]^n}^2 } \cdot \sqrt{ \sum_{n=N+1}^{\infty} \frac{1}{n!} ||\dot{h}||_{H^1}^2 n!^2 }
\leq \sqrt{ \sum_{n=N+1}^{\infty} \frac{(2L)^n}{n!} } \cdot \sqrt{ \sum_{n=N+1}^{\infty} \frac{1}{n!} ||f_n||_{L^2[0,1]^n}^2 n!^2 } \cdot \sqrt{ \sum_{n=N+1}^{\infty} \frac{1}{n!} ||\dot{h}||_{H^1}^2 n!^2 }
\]
where we have used the fact that, if \( g \) is a symmetric function on \([0, t]^k\), then
\[
\int_{0 < t_1 < \cdots < t_k < t} g(t_1, \ldots, t_k) dt_1 \cdots dt_k = \frac{1}{k!} \int_{[0,t]^k} g(t_1, \ldots, t_k) dt_1 \cdots dt_k
\]
as long as \( g \) is integrable.

Therefore

\[
\sup_{h \in K_L} \sup_{t \leq 1} |F(h)_t - F_N(h)_t| \to 0 \quad \text{as } N \to \infty
\]

which proves our claim.

Let us prove that \( F_N \) is weakly continuous on \( K_L \) as stated in the lemma. To show the weak continuity of \( F_N \), we only need to show the weak continuity of \( F_N \equiv \Psi \) which has a simple form, namely

\[
\Psi(h)_t = \int_{0<t_1<\cdots<t_n<t} f(t_1, \ldots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n
\]

where \( f \in L^2[0,1]^n \) which is symmetric, and has the following form

\[
f(t_1, \ldots, t_n) = \sum_{j_1, \ldots, j_n} C^{j_1, \ldots, j_n} f_{j_1}(t_1) \cdots f_{j_n}(t_n)
\]

where \( f_{j_k} \in L^2[0,1] \). Let \( h_k \to h \) weakly in \( H \), where \( h_k \in K_L \) (so that \( h \in K_L \)). Since \( f \) is symmetric, we thus have

\[
\Psi(h_k)_t - \Psi(h)_t = \frac{1}{n!} \int_{[0,t]^n} f(t_1, \ldots, t_n) \left( \dot{h}_k(t_1) \cdots \dot{h}_k(t_n) - \dot{h}(t_1) \cdots \dot{h}(t_n) \right) dt_1 \cdots dt_n
\]

\[
= \frac{1}{n!} \sum_{j_1, \ldots, j_n} C^{j_1, \ldots, j_n}
\]

\[
\times \int_{[0,t]^n} f_{j_1}(t_1) \cdots f_{j_n}(t_n) \left( \dot{h}_k(t_1) \cdots \dot{h}_k(t_n) - \dot{h}(t_1) \cdots \dot{h}(t_n) \right) dt_1 \cdots dt_n
\]

\[
= \frac{1}{n!} \sum_{j_1, \ldots, j_n} C^{j_1, \ldots, j_n} \sum_{l=1}^{n-1} \langle 1_{[0,t]} f_{j_l}, h_k \rangle \cdots \langle 1_{[0,t]} f_{j_{n-1}}, h_k \rangle \langle 1_{[0,t]} f_{j_n}, h_k - h \rangle
\]

\[
\times \langle 1_{[0,t]} f_{j_{n-j-1}}, h \rangle \cdots \langle 1_{[0,t]} f_{j_n}, h \rangle
\]

where \( \langle f, h \rangle = \int_0^1 f(t) \dot{h}(t) dt \), which yields that

\[
|\Psi(h_k)_t - \Psi(h)_t| \leq \frac{1}{n!} \sum_{j_1, \ldots, j_n} |C^{j_1, \ldots, j_n}| \left( \frac{2L}{\sqrt{2}} \right)^{n-1} \left\| f_{j_1} \right\|_{L^2[0,1]} \cdots \left\| f_{j_n} \right\|_{L^2[0,1]}
\]

\[
\times \left| \langle 1_{[0,t]} f_{j_{n-j}}, h_k - h \rangle \right|.
\]
Therefore $|\Psi(h_k)_t - \Psi(h)_t| \to 0$ for any $t \in [0, 1]$, as $h_k \to h$ weakly in $H$, and $\{\Psi(h_k)\}$ is bounded uniformly as $\{h_k\} \subset K_L$, so that $\Psi(h_k) \to \Psi(h)$ weakly in $E$.

Therefore all $F_N F$ are weakly continuous.

For any $h \in K_L$ and $[s, t] \subset [0, 1]$, we set

$$\Delta_{[s,t]}^n = [0, t]^n \setminus [0, s]^n.$$ 

Then

$$F(h)_t - F(h)_s = \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\Delta_{[s,t]}^n} f_n(t_1, \ldots, t_n) \hat{h}(t_1) \cdots \hat{h}(t_n) dt_1 \cdots dt_n$$

so that, by utilizing the Cauchy-Schwarz inequality

$$|F(h)_t - F(h)_s| \leq \sqrt{\sum_{n=1}^{\infty} \frac{1}{n!} \int_{\Delta_{[s,t]}^n} |f_n(t_1, \ldots, t_n)|^2 dt_1 \cdots dt_n} \sqrt{\sum_{n=1}^{\infty} \frac{1}{n!} ||h||_{L^2([0,1])}^2}$$

$$\leq e^L \sqrt{\sum_{n=1}^{\infty} \frac{1}{n!} \int_{\Delta_{[s,t]}^n} |f_n(t_1, \ldots, t_n)|^2 dt_1 \cdots dt_n}.$$ 

Since

$$\sum_{n=1}^{\infty} \frac{1}{n!} \int_{\Delta_{[s,t]}^n} |f_n(t_1, \ldots, t_n)|^2 dt_1 \cdots dt_n \leq \sum_{n=1}^{\infty} \frac{1}{n!} \int_{[0,1]^n} |f_n(t_1, \ldots, t_n)|^2 dt_1 \cdots dt_n \leq \sum_{n=1}^{\infty} \frac{1}{n!} \int_{[0,1]^n} |f_n(t_1, \ldots, t_n)|^2 dt_1 \cdots dt_n$$

and for each $n$, according to the Lebesgue theorem

$$\int_{\Delta_{[s,t]}^n} |f_n(t_1, \ldots, t_n)|^2 dt_1 \cdots dt_n \to 0 \text{ as } s \uparrow t,$$

therefore we can conclude that the functions in $F(K_L)$ are equi-continuous on $[0, 1]$, and are bounded in $E$:

$$|F(h)_t| \leq e^L \sqrt{\sum_{n=1}^{\infty} \frac{1}{n!} ||f||_{L^2([0,1])}^2} \quad \forall h \in K_L.$$
Therefore, according to Ascoli-Arzelà’s theorem (page 85, Section III-3, [31]), $F(K_L)$ is pre-compact. We now need to show that $F(K_L)$ is closed in $E$. Let $\{w_n\}$ be any sequence in $F(K_L)$ which converges to $w$ in $E$. Let $h_n \in K_L$ such that $F(h_n) = w_n$. $K_L$ is weakly compact in $H$, so let us assume that $h_n \rightharpoonup h$ weakly in $H$, otherwise consider a weakly convergent subsequence instead. Since $K_L$ is a closed and convex subset of $H$, so that $h \in K_L$. Therefore $F(h_n) = w_n$ weakly converges to $F(h)$ in $E$. We thus must have $F(h) = w$, so that $w \in F(K_L)$, and $F(K_L)$ is compact.

Of course, similar results hold in higher dimensions, where $E = C_0([0, 1]; R^d)$, $H = H^1_0([0, 1]; R^d)$ and the rate function $I(h) = \frac{1}{2} ||h||^2_{H^1}$.

### 2.2 Continuity of large deviations

We have thus developed necessary tools to formulate a continuity theorem for large deviation principles.

**Theorem 2.5** Suppose $H \subset E$ and $I$ is a good rate function satisfying the two conditions listed at the beginning of the last subsection 2.1. Let $F_n : H \to E$ be a sequence of rate-function mappings, and suppose that $F_n$ converges to $F$ uniformly on any level set $K_L = \{s : I(s) \leq L\}$. For each $n$, let $\{X_n^\varepsilon : \varepsilon \in (0, 1)\}$ (as well as $\{X^\varepsilon : \varepsilon \in (0, 1)\}$) be a family of random variables valued in $E$ on a complete probability space $(\Omega, \mathcal{F}, P)$. Suppose the following conditions are satisfied.

1) $\{X_n^\varepsilon : \varepsilon \in (0, 1)\}$ converges to $\{X^\varepsilon : \varepsilon \in (0, 1)\}$ exponentially: for any $\delta > 0$

$$\lim_{n \to \infty} \varepsilon \log P \{\rho(X_n^\varepsilon, X^\varepsilon) > \delta\} = -\infty$$

(2.12)

2) For each $n$, $\{X_n^\varepsilon : \varepsilon \in (0, 1)\}$ satisfies the large deviation principle with rate function $I_{F_n}$.

Then, the distribution family $\{\mu_\varepsilon : \varepsilon \in (0, 1)\}$ of the limiting process $\{X^\varepsilon : \varepsilon \in (0, 1)\}$ satisfies the large deviation principle with rate function $I_F$.

**Proof.** For simplicity, we use $I'_n$ to denote $I_{F_n}$ and $I'$ for $I_F$. Let $\rho$ be the distance function on $E$, i.e.

$$\rho(s, s') = ||s - s'|| \quad \forall s, s' \in E,$$

and for $s_0' \in E$ and $\delta > 0$, $B(s_0', \delta)$ denote the open ball in $E$ centered at $s_0'$ with radius $\delta$. 

17
Firstly we show the lower bound. Let $O$ be an open subset of $E$, we need to prove that
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \log \mu_\varepsilon (O) \geq - \inf_{s \in O} I'(s).
\] (2.13)
It is easy to see that we only need to show
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \log \mu_\varepsilon (B(s'_0, \delta)) \geq - \inf_{s' \in B(s'_0, \delta/2)} I'(s').
\] (2.14)
for any $s'_0 \in O$ and $\delta > 0$ such that $B(s'_0, \delta) \subset O$. We may assume that \( \inf_{s' \in B(s'_0, \delta/2)} I'(s') < \infty \), otherwise there is nothing to prove. By the triangle inequality one has for any $\lambda > 0$
\[
P \left\{ \rho(X^\varepsilon_n, s'_0) < \frac{\delta}{3} \right\} \leq P \{ \rho(X^\varepsilon_n, X^\varepsilon) > \lambda \}
+ P \left\{ \rho(X^\varepsilon, s'_0) < \lambda + \frac{\delta}{3} \right\}
\]
it follows that
\[
\log P \left\{ \rho(X^\varepsilon_n, s'_0) < \frac{2\delta}{3} \right\}
\leq \log 2 + \log \left\{ P \{ \rho(X^\varepsilon_n, X^\varepsilon) > \lambda \} \lor P \left\{ \rho(X^\varepsilon_n, s'_0) < \lambda + \frac{2\delta}{3} \right\} \right\}
\leq \log 2 + \log P \{ \rho(X^\varepsilon_n, X^\varepsilon) > \lambda \} \lor \log P \left\{ \rho(X^\varepsilon_n, s'_0) < \lambda + \frac{2\delta}{3} \right\}.
\]
Therefore
\[
\varepsilon \log P \left\{ \rho(X^\varepsilon_n, s'_0) < \frac{2\delta}{3} \right\}
\leq \varepsilon \log 2
+ \max \left\{ \varepsilon \log P \{ \rho(X^\varepsilon_n, X^\varepsilon) > \lambda \} ; \varepsilon \log P \left\{ \rho(X^\varepsilon_n, s'_0) < \lambda + \frac{2\delta}{3} \right\} \right\}
\]

hence
\[
\lim_{n \to \infty} \lim_{\varepsilon \downarrow 0} \varepsilon \log P \left\{ \rho(X^\varepsilon_n, s'_0) < \frac{2\delta}{3} \right\}
\leq \max \left\{ \lim_{n \to \infty} \lim_{\varepsilon \downarrow 0} \varepsilon \log P \{ \rho(X^\varepsilon_n, X^\varepsilon) > \lambda \} ; \lim_{\varepsilon \downarrow 0} \varepsilon \log P \left\{ \rho(X^\varepsilon_n, s'_0) < \lambda + \frac{2\delta}{3} \right\} \right\}
= \lim_{\varepsilon \downarrow 0} \varepsilon \log P \left\{ \rho(X^\varepsilon, s'_0) < \lambda + \frac{2\delta}{3} \right\}
\]
for any $\lambda > 0$, we have used the assumption that
\[
\lim_{n \to \infty} \lim_{\varepsilon \to 0} \log P \{ \rho(X_n^\varepsilon, X^\varepsilon) > \lambda \} = 0.
\]
On the other hand, as $\{X_n^\varepsilon : \varepsilon \in (0, 1)\}$ satisfies the large deviation principle with rate function $I_N'$, so that
\[
\lim_{\varepsilon \to 0} \varepsilon \log P \{ \rho(X_n^\varepsilon, s'_0) < \lambda + \frac{2\delta}{3} \} \geq - \lim_{n \to \infty} \inf_{s' \in B(s'_0, \frac{\delta}{12})} I_n'(s')
\]
and therefore
\[
\lim_{\varepsilon \to 0} \varepsilon \log P \{ \rho(X_n^\varepsilon, s'_0) < \lambda + \frac{2\delta}{3} \} \geq - \lim_{n \to \infty} \inf_{s' \in B(s'_0, \frac{\delta}{12})} I_n'(s') \quad (2.15)
\]
for any $\lambda > 0$.

According to the assumption that $\inf_{s' \in B(s'_0, \frac{\delta}{12})} I'(s') = M < \infty$. Since $I'$ is a good rate function, there is an $s'_1 \in B(s'_0, \frac{\delta}{12})$, such that $I'(s'_1) = M$. Since
\[
I'(s') = \inf \{ I(s) \mid s \in H \text{ such that } F(s) = s' \}
\]
and since $I$ is a good rate function on $E$, there is an $s_1 \in H$ such that $I(s_1) = M$ and $F(s_1) = s'_1$. Let $t'_n = F_n(s_1) \in E$. Then $\lim_{n \to \infty} t'_n = F(s_1) = s'_1$, so that for every $\alpha > 0$ there exists an $N_0$ such that
\[
t'_n \in B(s'_1, \alpha) \quad \forall n > N_0.
\]
Then for $n > N_0$ we have
\[
\inf_{s' \in B(s'_1, \alpha)} I_n'(s') \leq I_n'(t'_n) = \inf \{ I(s) : s \in H \text{ and } F_n(s) = t'_n \} \leq I(s_1).
\]
Choose $\alpha = \frac{1}{24} \delta$. Then
\[
\inf_{s' \in B(s'_0, \frac{\delta}{12})} I_n'(s') \leq I(s_1) = M
\]
and
\[
\inf_{s' \in B(s'_0, \frac{\delta}{2})} I'(s') = \inf_{s' \in B(s'_0, \frac{\delta}{2})} I'(s')
\]
so that
\[ \lim_{n \to \infty} \inf_{s' \in B(s_0', \frac{2\delta}{3})} I'_n(s') \leq \inf_{s' \in B(s_0', \frac{\delta}{2})} I'(s') \quad \forall n > N_0. \]

Hence
\[ \overline{\lim}_{\varepsilon \downarrow 0} \log P \left\{ \rho(X^\varepsilon, s_0') < \lambda + \frac{2\delta}{3} \right\} \geq - \inf_{s' \in B(s_0', \frac{\delta}{2})} I'(s') \quad (2.16) \]
for any \( \lambda > 0 \), which implies (2.14).

Now prove the upper bound: for any closed set \( S \) in \( E \),
\[ \overline{\lim}_{\varepsilon \downarrow 0} \log \mu_\varepsilon(S) \leq - \inf_{s' \in F} I'(s'). \quad (2.17) \]
For any \( \delta > 0 \), set
\[ F_\delta = \{ s' \in E \mid s'' \in S \text{ s.t. } \rho(s', s'') < \delta \}. \]
Then
\[ P \{ X^\varepsilon \in S \} \leq P \{ X_n^\varepsilon \in S_\delta \} + P \{ \rho(X^\varepsilon, X_n^\varepsilon) > \delta \} \]
and therefore
\[ \log P \{ X^\varepsilon \in S \} \leq \log 2 + \log [P \{ X_n^\varepsilon \in S_\delta \} \lor P \{ \rho(X^\varepsilon, X_n^\varepsilon) > \delta \}] \]
so that, for any \( \delta > 0 \)
\[ \overline{\lim}_{\varepsilon \downarrow 0} \log P \{ X^\varepsilon \in S \} \leq \overline{\lim}_{\varepsilon \downarrow 0} \log P \{ X_n^\varepsilon \in S_\delta \} \lor \overline{\lim}_{\varepsilon \downarrow 0} \log P \{ \rho(X^\varepsilon, X_n^\varepsilon) > \delta \}. \]
For any \( K > 0 \) there is a \( N_2 \) depending only on \( \delta \) and \( K \), such that
\[ \overline{\lim}_{\varepsilon \downarrow 0} \log P \{ \rho(X^\varepsilon, X_n^\varepsilon) > \delta \} \leq -K \quad \forall n > N_2. \]
On the other hand, \( \{ X_n^\varepsilon : \varepsilon \in (0, 1) \} \) satisfies the large deviation principle with rate function \( I'_n \), so that
\[ \overline{\lim}_{\varepsilon \downarrow 0} \log P \{ X^\varepsilon \in S \} \leq (-K) \lor \overline{\lim}_{\varepsilon \downarrow 0} \log P \{ X_n^\varepsilon \in S_\delta \} \]
\[ \leq \max \left\{ -\inf_{s' \in F} I'_n, -K \right\} \quad \forall n > N_2. \]
It follows that
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \log P \{ X^\varepsilon \in S \} \leq - \lim_{\delta \downarrow 0} \lim_{n \to \infty} \inf_{S^n} I'_n. \quad (2.18)
\]

Let us consider
\[
l = \lim_{\delta \downarrow 0} \lim_{n \to \infty} \inf_{S^n} I'_n.
\]

If \( l = \infty \), then
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \log P \{ X^\varepsilon \in S \} = -\infty \leq - \inf_{s' \in S} I'(s')
\]
so let us assume that \( l < \infty \). In this case we show that
\[
\inf_F I' \leq l = \lim_{\delta \downarrow 0} \lim_{n \to \infty} \inf_{S^n} I'_n. \quad (2.19)
\]

In this case, by definition of the multiple limits of the right-hand side of (2.19) we may choose a sequence \((s_m) \subset K_{t+1}\), and a subsequence \(n_m \to \infty\) such that
\[
f_{n_m}(s_m) = s'_m, \quad \rho(s'_m, S) \leq \frac{1}{m}
\]
and
\[
I(s_m) \leq l + \frac{1}{m}.
\]

Then \( \{s_m\} \subset K_{t+1} \). Since \( K_{t+1} \) is compact in \( E \), and weakly compact in \( H \), we can further assume that \( s_m \to s \) in \( K_{t+1} \) (in the distance \( \rho \)), and \( s_m \to s \) weakly in \( H \). Since \( I \) is lower semi-continuous, \( I(s) \leq l \).

For any \( \alpha > 0 \), there is a number \( N_3 \) such that
\[
\rho(F_{n_m}(s), F(s)) < \frac{\alpha}{2} \quad \forall s \in K_{t+1} \quad (2.20)
\]
for any \( m \geq N_3 \). In particular
\[
\rho(s'_m, F(s_m)) < \frac{\alpha}{2} \quad \forall m \geq N_3.
\]
However \( \{F(s_m) : m = 1, 2, \cdots\} \subset F(K_{t+1}) \) which is compact in \( E \). Therefore, if necessary by extracting a subsequence, we may assume \( \{F(s_m)\} \) converges in \( E \) to \( s' \). Hence, there is an \( N_4 \) such that
\[
\rho(F(s_m), s') < \frac{\alpha}{2} \quad \forall m \geq N_4.
\]
Therefore
\[\rho(s'_m, s') < \alpha \quad \forall m \geq N_3 \lor N_4.\]
That is, \(s'_m \rightarrow s'\) in \(E\), so that \(s' \in F\). On the other hand, \(s_m \rightarrow s\) weakly in \(H\), so that, as \(F\) is weakly continuous on \(K_{t+1}\), \(F(s_m) \rightarrow F(s)\) weakly in \(E\). We thus must have \(F(s) = s' \in F\) and \(I(s) \leq l\). Therefore \(\inf_F I' \leq l\) which completes the proof of (2.19). □

3 Hypercontractivity and martingales

Let us retain the notations we have established in Introduction. In particular, \((W_0^d, \mathcal{F}_1, P^w)\) is the Wiener space on \(R^d\). However, for simplicity, we may assume that \(d = 1\) without loss of generality.

If \(f \in L^2[0, 1]^n\) we use
\[J_n(f)_t = \int_{0<t_1<\cdots<t_n<t} f(t_1, \ldots, t_n) dw_{t_1} \cdots dw_{t_n}\]
to denote the multiple Wiener-Itô integral on \([0, t], t \in [0, 1]\). \(\{J_n(f)_t\}\) is a square-integrable martingale up to time 1. According to Wiener-Itô’s chaos decomposition ([24], [10]), if \(\xi \in L^2(W_0^d, \mathcal{F}_1, P^w)\), then
\[\xi = E\xi + \sum_{n=1}^{\infty} J_n(f_n)_1\]
for a sequence of symmetric functions \(f_n \in L^2[0, 1]^n\) and
\[||\xi - E\xi||_2^2 = \sum_{n=1}^{\infty} \frac{1}{n!} ||f_n||^2_{L^2[0, 1]^n}\]
where \(||\xi||_p\) denotes the \(L^p\)-norm of \(\xi\). The Ornstein-Uhlenbeck semigroup \((P_t)_{t \geq 0}\) is defined by
\[P_t\xi = E(\xi) + \sum_{n=1}^{\infty} e^{-nt} J_n(f_n)_1.\]

\((P_t)_{t \geq 0}\) is a symmetric diffusion semigroup on \(L^2(W_0^d, \mathcal{F}_1, P^w)\), which may be extended uniquely to a strongly continuous semigroup on \(L^p(W_0^d, \mathcal{F}_1, P^w)\) for every \(p \geq 1\).

The following hypercontractivity of the Ornstein-Uhlenbeck semigroup plays a major role in this paper.
Theorem 3.1 (L. Gross, Nelson [20]) The Ornstein-Uhlenbeck semigroup 
\((P_t)_{t \geq 0}\) possesses the hypercontractivity

\[ \|P_t \xi\|_{p(t)} \leq \|\xi\|_p \]

for all \(\xi \in L^2(W_0^d, F_1, P^w)\), \(p > 1\) and \(t > 0\), where \(p(t) = 1 + (p - 1)e^{2t}\).

As an application of the hypercontractivity, we present a proof of the following estimate, a well-known result in Gaussian analysis, which shows that tail behaviors of multiple Wiener-Itô integrals.

Proposition 3.2 Let \(\xi = I_n(f)\) where \(f \in L^2([0, 1]^n)\) for some \(n\). Let \(Y_t = P^w(\xi | F_t)\) for \(t \in [0, 1]\). Then for any \(\alpha < n/(2e)\)

\[ E \exp \left( \alpha \left| \frac{1}{\|\xi\|_2} \sup_{t \leq 1} Y_t \right|^{2/n} \right) \leq C_{\alpha, n} \]

(3.1)

where

\[ C_{\alpha, n} = 1 + 4e^\alpha + \frac{2e}{\sqrt{2\pi}} \sum_{k \geq n} \frac{1}{\sqrt{k}} \left( \frac{2\alpha e}{n} \right)^k. \]

Therefore, for any \(\delta > 0\)

\[ P^w \left\{ \left| \sup_{t \leq 1} Y_t \right| \geq \delta \right\} \leq C_{\alpha, n} \exp \left\{ -\alpha \frac{\delta^{2/n}}{\|\xi\|_2^{2/n}} \right\}. \]

(3.2)

Proof. Without losing generality, we may assume that \(\|\xi\|_2 = 1\). According to the hypercontractivity of \((P_t)\), \(P_t \xi = e^{-nt} \xi \in L^{p(t)}\) where \(p(t) = 1 + e^{2t}\), and

\[ \left( E |\xi|^{1+e^{2t}} \right)^{1/(1+e^{2t})} \leq e^{nt} \quad \forall t > 0 \]

that is for any \(p > 1\), \(\xi \in L^p(W_0^d, F_1, P^w)\), and

\[ E |\xi|^p \leq (p - 1)^{np/2} \quad \forall p > 1. \]

By Doob’s inequality, \(\sup_{t \leq 1} Y_t \in L^p(W_0^d, F_1, P^w)\) and

\[ E \left| \sup_{t \leq 1} Y_t \right|^p \leq \left( \frac{p}{p - 1} \right)^p (p - 1)^{np/2} \]

\[ < cp(p - 1)^{\frac{np}{p-1}} \]

23
for any $p > 1$. Since
\[ E \exp \left( \alpha \sup_{t \leq 1} Y_t^\theta \right) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} E \left| \sup_{t \leq 1} Y_t \right|^\theta \]
\[ = 1 + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} E \left| \sup_{t \leq 1} Y_t \right|^k + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} E \left| \sup_{t \leq 1} Y_t \right|^k \]
\[ \leq 1 + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \left( E \left| \sup_{t \leq 1} Y_t \right|^2 \right)^{\theta k/2} \]
\[ + e \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \frac{k\theta}{k\theta - 1} (k\theta - 1)^{nk\theta/2} \]
\[ \leq 1 + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} 4^{\theta k/2} + e \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \frac{k\theta}{k\theta - 1} (k\theta - 1)^{nk\theta/2}, \]
choosing $\theta = 2/n$, we thus have
\[ E \exp \left( \alpha \sup_{t \leq T} Y_t^{2/n} \right) \leq 1 + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} 4^{k/n} + 2e \sum_{k=0}^{\infty} \frac{k^k}{k!} \left( \frac{2\alpha}{n} \right)^k. \]

According to Stirling’s formula
\[ \frac{k^k}{k!} \leq \frac{1}{\sqrt{2\pi}} \frac{e^k}{\sqrt{k}} e^{-\frac{k}{2}} \leq \frac{1}{\sqrt{2\pi}} \frac{e^k}{\sqrt{k}} \]
(see page 52, W. Feller [14]) which follows that
\[ E \exp \left( \alpha \sup_{t \leq T} Y_t^{2/n} \right) \leq 1 + 4\alpha^2 + 2e \frac{\alpha^2}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{1}{k} \left( \frac{2\alpha}{n} \right)^k \]
the right-hand is finite if $\alpha < n/(2e)$. ■

**Proposition 3.3** If $\xi \in L^2 \left( W_0^d, F_1, P^w \right)$ and $Y_t = P^w (\xi | F_t)$, then for every $\varepsilon \in (0, 1)$ and $\delta > 0$
\[ P^w \left\{ \sup_{t \leq 1} \left| P_{-\log \sqrt{\varepsilon}} Y_t \right| \geq \delta \right\} \leq (1 + \varepsilon)^{1+\frac{1}{\tau}} \frac{||\xi||^{2+\frac{1}{\tau}}}{\delta^{1+\frac{1}{\tau}}}. \quad (3.3) \]
Proof. By the previous lemma, \( P_{-\log \sqrt{\varepsilon} \xi} \in L^{1+\frac{1}{\varepsilon}} (W^d_0, \mathcal{F}_1, P^w) \) for any \( \varepsilon \in (0, 1) \), thus, by Doob’s \( L^p \)-inequality

\[
E \left| \sup_{t \leq T} (P_{-\log \sqrt{\varepsilon} Y_t}) \right|^{1+\frac{1}{\varepsilon}} \leq (1 + \varepsilon)^{1+\frac{1}{\varepsilon}} E \left| P_{-\log \sqrt{\varepsilon} \xi} \right|^{1+\frac{1}{\varepsilon}} \leq (1 + \varepsilon)^{1+\frac{1}{\varepsilon}} \| \xi \|^{1+\frac{1}{2}}
\]

the second inequality follows from the hypercontractivity of the Ornstein-Uhlenbeck semigroup \((P_t)_{t \geq 0}\). Therefore

\[
P^w \left\{ \sup_{t \leq T} (P_{-\log \sqrt{\varepsilon} Y_t}) \geq \delta \right\} \leq \frac{1}{\delta^{1+\frac{1}{\varepsilon}}} E \left| \sup_{t \leq T} (P_{-\log \sqrt{\varepsilon} Y_t}) \right|^{1+\frac{1}{\varepsilon}} \leq (1 + \varepsilon)^{1+\frac{1}{\varepsilon}} \frac{\| \xi \|^{1+\frac{1}{2}}}{\delta^{1+\frac{1}{\varepsilon}}}.
\]

\[\blacksquare\]

4 Itô’s mappings defined by Itô’s multiple integrals

The large deviation principle for multiple Wiener-Itô integrals has been established in M. Ledoux [25], also in [31]. We believe their arguments, with a little bit of extra work, can equally apply to the case of finite sum of multiple Itô’s integrals. For completeness we however include a different proof, which we believe has independent interest by its own.

Our approach is to apply the contraction principle to the Itô-Lyons mappings on spaces of geometric rough paths. Not like the original Itô’s mappings defined by solving stochastic differential equations via Itô’s calculus, Itô-Lyons mappings will serve the same aim as that of Itô mappings, but in addition they are continuous with respect to variation distances. The main concept and the continuity result were established in an important work by T. Lyons [28] (see also [29], the excellent recent books [30], [16] etc), which says solutions to Stratonovich type stochastic differential equations are continuous functions of Brownian motion paths together with its Lévy area. A more precise statement, see items 1 and 2 in Theorem 4.1 below.

Lyons’ continuity theorem, or called the universal limit theorem as suggested by Malliavin, has been finding many applications in analyzing Wiener
functionals, for example, see the recent articles by Hambly and Lyons [21], Cass and Friz [5] and etc. The important fact here is that, the rough path analysis, as developed in [29], allows us more effectively to apply classical functional analytic techniques to stochastic analysis. The result in this section is another example of the power of this new analysis.

4.1 Schilder’s theorem in the $p$-variation topology

In M. Ledoux, Z. Qian and T. Zhang [26], a version of the large deviation principle of Schilder’s in the $p$-variation topology has been established, with which we will prove the large deviation principle for martingales.

Let $p \in (2, 3)$ be a fixed constant. Let $\mathbb{W}$ be the space of all continuous path $w \in \mathbb{W}_0^d$ which has finite total variations over $[0, 1]$:

$$
\sup_D \sum_l |w_{t_l} - w_{t_{l-1}}| < +\infty
$$

where $D$ runs over all finite partitions \{0 = t_0 < t_1 < \cdots < t_n = 1\} of the interval [0, 1]. For a path $w \in \mathbb{W}$ we may consider its increment $w_{s,t} = w_t - w_s$ and its Lévy area

$$
w^2_{s,t} = \int_{s<t_1<t_2<t} dw_{t_1} \otimes dw_{t_2}
$$

de ned via Riemann sum limits. $w^2$ can be considered as a $d \times d$ matrix-valued function on $\Delta \equiv \{(s, t) : 0 \leq s \leq t \leq 1\}$. Then define

$$
w_{s,t} = (1, w^1_{s,t}, w^2_{s,t}) \quad \text{if} \quad (s, t) \in \Delta
$$

and $w : (s, t) \in \Delta \to w_{s,t}$ which is called the rough path associated to $w \in \mathbb{W}$, a path of finite variations. The space of all such rough paths is denoted by $\mathbb{W}^\infty$ (and we may thus identify $\mathbb{W}$ with its “lift” $\mathbb{W}^\infty$), equipped with a natural metric $d_p$ (called the $p$-variation metric where $p \in (2, 3)$)

$$
d_p(w, y) = \sup_D \left( \sum_l |w^1_{t_{l-1},t_l} - y^1_{t_{l-1},t_l}|^p \right)^{\frac{1}{p}} + \sup_D \left( \sum_l |w^2_{t_{l-1},t_l} - y^2_{t_{l-1},t_l}|^{p/2} \right)^{\frac{2}{p}}.
$$

(4.1)

Since any $w \in H^1_0([0, 1]; R^d)$ has a finite variation on [0, 1], therefore the Cameron-Martin space $H^1_0([0, 1]; R^d)$ may be considered as a subspace of $\mathbb{W}^\infty$, hence of $\mathbb{W}^p$ to be introduced later on.
The completion of $\mathbb{W}^\infty$ under the $p$-variation metric $d_p$ is denoted by $\mathbb{W}^p$. T. Lyons [28] has established the following result. Consider the ordinary differential equation

$$dy^i_t = f^i_0(t, y_t) dt + \sum_{j=1}^d f^i_j(t, y_t) \circ dw^j_t, \quad y_0 = 0 \quad (4.2)$$

$i = 1, \ldots, m$, where we have used $\circ dw^j_t$ to denote the usual differential if $w$ is differentiable, to indicate (4.2) should be understood as Stratonovich stochastic differential equations for Brownian motion $w$. Both interpretation of (4.2) within the setting of rough path analysis.

If $f^i_j, f^i_0$ are $C^3_b$ functions, then $w \to y$ is continuous map from $\mathbb{W}^\infty$ into $\mathbb{W}^\infty$ under $p$-variation metric $d_p$ and therefore extended continuously to be a map from $\mathbb{W}^p$ into $\mathbb{W}^p$, called the Itô-Lyons mapping determined by (4.2).

This result, together with the following theorem proved in [29] and Ledoux, Qian and Zhang [26], can be used to establish large deviation principles for a large class of Itô’s functionals.

**Theorem 4.1** Let $p \in (2, 3)$ be a fixed a constant. Let $(\mathbb{W}^d_0, \mathcal{F}_1, P^w)$ be the $d$-dimensional Wiener space, so that its coordinate process $(w_t)_{t \in [0,1]}$ is an $\mathbb{R}^d$-valued Brownian motion. Let $2 < p < 3$ be a fixed constant. Set

$$w^1_{s,t} = w_t - w_s$$

and

$$w^2_{s,t} = \int_{s < t_1 < t_2 < t} \circ dw_{t_1} \otimes \circ dw_{t_2}$$

where $\circ d$ denotes the Stratonovich integration. Let $w_{s,t} = (1, w^1_{s,t}, w^2_{s,t})$. The law of $\{w_{s,t} : (s, t) \in \Delta\}$ is denoted by $\tilde{P}^w$ which is a probability measure on $(\mathbb{W}^p, \mathcal{B}(\mathbb{W}^p))$.

1. For any $w \in \mathbb{W}$ there is a unique solution $y$ of (4.2) which belongs to $\mathbb{W}$, denoted by $G(w)$. Their corresponding geometric rough paths are denoted by $w \in \mathbb{W}^\infty$ and $G(w) \in \mathbb{W}^\infty$. The mapping $G : w \to F(w)$ can be uniquely extended to be a continuous mapping from $(\mathbb{W}^p, d_p)$ to $(\mathbb{W}^p, d_p)$, denoted again by $G$, called the Itô-Lyons mapping defined by (4.2). Moreover, the projection to the first level path, $y_t = G^1(w)_{0,t}$ is a version of the strong solution of (4.2) on the probability space $(\mathbb{W}^p, \mathcal{B}(\mathbb{W}^p), \tilde{P}^w)$. The results remain true if all $f^i_j$ are linear in the space variables, with bounded derivatives in $t$. 27
2. We have 
\[ \tilde{P}^w \{ \Gamma(\varepsilon)w \in \mathbb{W}^p : \forall \varepsilon > 0 \} = 1 \]
where \( \Gamma(\varepsilon)w_{s,t} = (1, \sqrt{\varepsilon} w_{s,t}^1, \varepsilon w_{s,t}^2) \).

3. Let \( \tilde{P}^w_\varepsilon \) be the distribution of \( (\Gamma(\varepsilon)w_{s,t})_{0 \leq s \leq t \leq 1} \), a probability measure on \( (\mathbb{W}^p, \mathcal{B}(\mathbb{W}^p)) \). Then \( \{ \tilde{P}^w_\varepsilon : \varepsilon > 0 \} \) possesses the large deviation principle with respect to the topology induced by the \( p \)-variation metric, with rate function
\[ \phi(w) = \frac{1}{2} \int_0^1 |\dot{w}(t)|^2 dt, \]
if \( w \in \mathbb{W}^\infty \) such that its first level path \( w \in H^1_1([0,1]; \mathbb{R}^d) \), otherwise \( \phi(w) = \infty \).

The first item in the theorem is called the universal limit theorem of Lyons’, the second item says the Brownian motion may be lifted to geometric rough paths, and the last item is Schilder’s large deviation principle in the \( p \)-variation metric, proved in Ledoux-Qian-Zhang [26].

4.2 Several elementary facts

In this part we present some important facts about the relationship between multiple Wiener-Itô integrals and solutions of stochastic differential equations of Stratonovich type. To this end we need to introduce more notations.

If \( f \in L^2(\mathbb{R}^d) \), then \( J_n(f) = \{ J_n(f)_t \} \) is the process of \( n \)-th multiple Wiener-Itô integrals where
\[ J_n(f)_t = \int_{\theta < t_1 < \cdots < t_n < t} f(t_1, \cdots, t_n) dw_{t_1} \cdots dw_{t_n} \]
which is a martingale for \( n \geq 1 \).

It occurs in the computations below some “partial” multiple Wiener-Itô integrals which are no-longer martingales. Here is a typical example.

If \( f \) is a function of \( n \)-variables \( (t_1, \cdots, t_n) \), then for \( 1 \leq k \leq n \) we use \( f_{,k}(\cdot; t) \) to denote the function of \( (t_1, \cdots, t_k) \):
\[ f_{,k}(\cdot; t) : (t_1, \cdots, t_k) \rightarrow f(t_1, \cdots, t_k, t, \cdots, t). \]
Then \( f_n = f \). The following stochastic process

\[
J_k(f_{j,k}(\cdot; t))_t = \int_{0 < t_1 < \cdots < t_k < t} f(t_1, \cdots, t_k, \cdot, t, \cdot) \, dw_{t_1} \cdots dw_{t_n}
\]

is well-defined, for example, if \( f \) is differentiable in all variables.

In what follows, we always consider a function \( f \) of \( n \) variables in the order from left to right (i.e. we use the standard coordinate system in \( \mathbb{R}^n \)), and \( \nabla_j f \) denotes the partial derivative in the \( j \)-th coordinate, i.e. \( \frac{\partial}{\partial t_j} f \).

**Lemma 4.2** If \( f(t_1, \cdots, t_n) \) is smooth with bounded derivatives, then

\[
dJ_n(f)_t = J_{n-1}(f_{n-1}(\cdot; t))_t \circ dw_t - \frac{1}{2} J_{n-2}((\nabla_n f)_{n-2}(\cdot; t)) \, dt \tag{4.3}
\]

where \( \circ dw_t \) denotes the Stratonovich differential.

**Proof.** By definition

\[
J_n(f)_t = \int_0^t J_{n-1}(f_{n-1}(\cdot; s))_s \, dw_s.
\]

To simplify our proof, let \( Z_t = J_{n-1}(f_{n-1}(\cdot; t))_t \) so that \( J_n(f)_t = \int_0^t Z_s \, dw_s \).

Therefore

\[
J_n(f)_t = \int_0^t Z_s \circ dw_s - \frac{1}{2} \langle Z, w \rangle_t
\]

and we aim to compute the bracket process \( \langle Z, w \rangle_t \). To this end, we begin with the case that

\[
f_n(t_1, \cdots, t_{n-1}, t_n) = g_{n-1}(t_1, \cdots, t_{n-1}) g(t_n).
\]

Then, according to integration by parts

\[
Z_t = g(t) J_{n-1}(g_{n-1})_t
\]

\[
= \int_0^t g'(s) J_{n-1}(g_{n-1})_s \, ds + \int_0^t g'(s) dJ_{n-1}(g_{n-1})_s
\]

\[
= \int_0^t g'(s) J_{n-1}(g_{n-1})_s \, ds
\]

\[
+ \int_0^t g'(s) J_{n-2}(g_{n-1, n-2}(\cdot; s))_s \, dw_s
\]

29
which follows that

\[
(Z, w)_t = \int_0^t g'(s) J_{n-2}(g_{n-1;n-2}(\cdot; s)) s ds
\]

\[
= \int_0^t J_{n-2}(g_{n-1;n-2}(\cdot; s) g'(s)) s ds
\]

\[
= \int_0^t J_{n-2}(\nabla_n f)_{n-2}(\cdot, s) s ds.
\]

It is immediate that this equality holds for general \(f\), and thus proves the lemma. ■

**Lemma 4.3** Let \( f_n(t_1, \cdots, t_n) \) be a smooth symmetric function, let \( 1 \leq k \leq n \), and consider Itô’s multiple integral

\[
H_t = J_k(f_n; k; t)
\]

\[
= \int_{0 < t_1 < \cdots < t_k < t} f_n(t_1, \cdots, t_k, t_1, \cdots, t) dw_{t_1} \cdots dw_{t_k}.
\]

Then

\[
dH_t = \sum_{j=k+1}^n J_k(\nabla_j f_n; k; t) dt + J_{k-1}(f_{n;k-1}(\cdot; t)) \circ dw_t
\]

\[
- \frac{1}{2} J_{k-2}(\nabla_k f_n; k-2; t) dt.
\]

**Proof.** Let us consider the case that

\[
f_n(t_1, \cdots, t_k, t_{k+1}, \cdots, t_n) = g_k(t_1, \cdots, t_k) g_{k+1}(t_{k+1}) \cdots g_n(t_n)
\]

so that

\[
f_n(t_1, \cdots, t_k, t, \cdots, t) = g_k(t_1, \cdots, t_k) g(t)
\]

with

\[
g(t) = g_{k+1}(t) \cdots g_n(t).
\]
Then, by integration by parts,
\[
\begin{align*}
\frac{dH}{dt} &= g'(t)J_k(g_k)_t dt + g(t)J_k(g_k)_t \\
&= g'(t)J_k(g_k)_t dt + g(t)J_{k-1}(g_{k-1;}t)_t \circ dw_t \\
&\quad - \frac{1}{2}g(t)J_{k-2}((\nabla_k g_{k-2;} ; t))_t dt
\end{align*}
\]
\[
\begin{align*}
&= \sum_{j=k+1}^n J_k((\nabla_j f_n)_k (; t))_t dt + J_{k-1}(f_{n;k-1} ; t)_t \circ dw_t \\
&\quad - \frac{1}{2}J_{k-2}((\nabla_k f_n ; k-2 ; t))_t dt
\end{align*}
\]
which proves the lemma. □

### 4.3 Stochastic differential equations

Let \( \xi = \sum_{n=1}^N J_n(f_n)_t \in L^2(W^d_0, \mathcal{F}_1, P_w) \) (but in the following computations, we assume for simplicity that \( d = 1 \)) for some natural number \( N \) and smooth functions \( f_n \) on \([0,1]^n\) with bounded derivatives, and

\[
Y_\varepsilon = P_{-\log \varepsilon} Y_t = \sum_{n=1}^N \varepsilon^n J_n(f_n)_t, \quad t \in [0,1]. \quad (4.5)
\]

The aim of this section is to construct a continuous function \( F_\varepsilon \) on \((W^p, B(W^p), \tilde{P}_w)\) (where the space \( W^p \) of geometric rough paths is endowed with the \( p \)-variation distance), such that \( F_\varepsilon(\Gamma(\varepsilon)w) = Y_\varepsilon(w) \) almost surely.

To this end, we demonstrate that \( Y_\varepsilon \) is a part of the solution of a Stratonovich type stochastic differential equation, at least for good functions \( f_n \).

According to (4.4)

\[
\begin{align*}
\frac{dY_\varepsilon}{dt} &= \sqrt{\varepsilon} \sum_{n=1}^N \varepsilon^{n-1} J_{n-1}(f_{n-1} ; t)_t \circ dw_t \\
&\quad - \frac{1}{2} \varepsilon \sum_{n=1}^N \varepsilon^{n-2} J_{n-2}((\nabla_n f_n)_n-2 ; t)) dt,
\end{align*}
\]
and

\[ dJ_{n-1}(f_{n-1}; t) = \nabla_n [J_{n-1}(f_{n-1}); t)] dt + J_{n-2}(f_{n-2}; t) \circ dw_t - \frac{1}{2} J_{n-3}((\nabla_{n-1} f_{n-1}; n-3; t)) dt. \tag{4.6} \]

where

\[
\nabla_n [J_{n-1}(f_{n-1}; t)] = \left. \frac{\partial}{\partial t_n} \right|_{t_n=t} \int_{0<t_1<\cdots<t_{n-1}<t} f_n(t_1, \ldots, t_{n-1}, t_n) dw_{t_1} \cdots dw_{t_{n-1}}.
\]

Unfortunately it does not lead to a closed system of stochastic differential equations of Stratonovich type. Therefore we consider a special case in which each \( f_n \) is a linear combination of functions of product form. We need some more notations.

For \( \varepsilon \in (0, 1), n \in \mathbb{N} \), and \( \{g; f^1, \ldots, f^n\} \) a family of smooth functions on \([0, 1]\) with bounded derivatives, then we define

\[
Z^n_{t, \{g; f^1, \ldots, f^n\}} = \varepsilon^n g(t) \int_{0<t_1<\cdots<t_n<t} f^1(t_1) \cdots f^n(t_n) dw_{t_1} \cdots dw_{t_n} \tag{4.7}
\]

and

\[
X^n_{t, \{g; f^1, \ldots, f^n\}} = \begin{pmatrix} Z^n_{t, \{1; f^1, \ldots, f^n\}} \\ Z^n_{t, \{g; f^1, \ldots, f^n\}} \end{pmatrix}.
\]

Therefore

\[
Z^n_{t, \{1; f^1, \ldots, f^n\}} = \varepsilon^n \int_{0<t_1<\cdots<t_n<t} f^1(t_1) \cdots f^n(t_n) dw_{t_1} \cdots dw_{t_n},
\]

\[
Z^n_{t, \{g; f^1, \ldots, f^n\}} = g(t) Z^n_{t, \{1; f^1, \ldots, f^n\}}
\]

and \( X^n_{t, \{1; f^1, \ldots, f^n\}} \) contains just two identical copies of \( Z^n_{t, \{1; f^1, \ldots, f^n\}} \). We use the convention that \( Z^0, \{\cdot\} = 1 \) and \( Z^n, \{\cdot\} = 0 \) for \( n < 0 \).

**Lemma 4.4** The stochastic process \( X^n_{t, \{g; f^1, \ldots, f^n\}} \) satisfies the following recursion equations

\[
dX^n_{t, \{g; f^1, \ldots, f^n\}} = \varepsilon^n g'(t) E_{21} X^{n, \{g; f^1, \ldots, f^n\}} dt - \frac{1}{2} \varepsilon E_{12} X^{n-2, \{f_{n-1}; f^1, \ldots, f^n\}} dt
\]

\[
- \frac{1}{2} \varepsilon E_{22} X^{n-2, \{f_{n-1}; f^1, \ldots, f^n\}} dt
+ \sqrt{\varepsilon} \left( E_{12} X^{n-1, \{f_{n+1}; f^1, \ldots, f^n\}} + E_{22} X^{n-1, \{g; f^1, \ldots, f^n\}} \right) \circ dw_t, \tag{4.8}
\]

\[ 32 \]
where $E_{ij}$ is the $2 \times 2$ matrices with 1 at $(i, j)$ entry and other entries zero.

**Proof.** It follows from (4.4) that $Z^n_{\xi, \{g; f^1, \ldots, f^n\}}$ satisfies the following stochastic differential equation

\[
dZ^n_{\xi, \{g; f^1, \ldots, f^n\}} = \sqrt{\varepsilon} Z^n_{\xi-1, \{g f^n; f^1, \ldots, f^n-1\}} \circ dw_t + \varepsilon \frac{g'}{2} Z^n_{\xi, \{1; f^1, \ldots, f^n\}} dt
\]

and

\[
dZ^n_{\xi, \{1; f^1, \ldots, f^n\}} = \sqrt{\varepsilon} Z^n_{\xi-1, \{f^n; f^1, \ldots, f^n-1\}} \circ dw_t - \frac{1}{2} \varepsilon Z^n_{\xi-2, \{f^n-1 f^n; f^1, \ldots, f^n-2\}} dt
\]

which is equivalent to (4.8). ■

Now let us consider

\[
Y^\varepsilon_t = P_{\log \sqrt{\varepsilon} P^w}(\xi | F_t)
\]

where $\xi = \sum_{n=1}^N J_n(f_n) \in L^2(W_0, F_1, P^w)$ with each integrand $f_n(t_1, \ldots, t_n)$ has a product form, say

\[
f_n(t_1, \ldots, t_n) = \sum_{j_1, \ldots, j_n=1}^{N_n} C_{j_1, \ldots, j_n} f_{j_1}^n(t_1) \cdots f_{j_n}^n(t_n)
\]

where $C_{j_1, \ldots, j_n}$ are constants, $N_n$ is a natural number, and all $f_{j_k}^n$ are smooth functions with bounded derivatives.

In this case

\[
Y^\varepsilon_t = \sum_{n=1}^{N} \sum_{j_1, \ldots, j_n=1}^{N_n} C_{j_1, \ldots, j_n} \varepsilon^{\frac{3}{2}} \int_{0<t_1<\cdots<t_n<t} f_{j_1}^n(t_1) \cdots f_{j_n}^n(t_n) dw_{t_1} \cdots dw_{t_n}.
\]

(4.10)

We are going to show that $Y^\varepsilon_t$ is part of the solution to a Stratonovich type stochastic differential equation. More precisely, we are going to show that

\[
\left(Y^\varepsilon_t, X^n_{\xi, \{g; f^1, \ldots, f^n\}} \right)_{0 \leq j < k < n, 1 \leq n \leq N}
\]

is the unique strong solution to a system of stochastic differential equations of Stratonovich type, where the general term is given by

\[
X^n_{\xi, \{g; f^1, \ldots, f^n\}} = X^{n-k, \{g; f^1, \ldots, f^{n-k}\}}_t
\]

for $k = 1, \ldots, n-1, j = 0, \ldots, k-1$.
and \( g_{n,k,j} = \prod_{i=j}^{k-1} f_{n}^{f_{n-1}} \). The projection to the first component in (4.11), i.e.

\[
(y, y, (x_{n,k,j}^{n,k,j})_{0 \leq j < k < n, 1 \leq n \leq N}) \rightarrow y
\]

will be denoted by \( \pi_{1} \).

**Proposition 4.5** Let \( Z_t = (Y^\varepsilon_t, Y^{\varepsilon^2}_t) \) where \( (Y^\varepsilon_t)_{t \leq 1} \) be given by equation (4.10), and let

\[
X_t = \begin{pmatrix} Z_t, (X^{n,k,j}_t)_{0 \leq j < k < n, 1 \leq n \leq N} \end{pmatrix}
\]

Then the stochastic process \( X_t \) is the unique strong solution to the following system of Stratonovich stochastic differential equations

\[
dZ_t = \sqrt{\varepsilon} \sum_{n=1}^{N} \tilde{\sum}_{j_1, \ldots, j_n=1}^{\tilde{N}} C_{n}^{j_1 \cdots j_n} (E_{12} + E_{22}) X_{t}^{n,1,0} \circ dw_t
\]

\[
-\frac{1}{2} \varepsilon \sum_{n=1}^{N} \tilde{\sum}_{j_1, \ldots, j_n=1}^{\tilde{N}} C_{n}^{j_1 \cdots j_n} (E_{12} + E_{22}) X_{t}^{n,2,0} dt
\]

(4.12)

\[
dX_{t}^{n,k,j} = \sqrt{\varepsilon} \frac{a_k}{2} g_{n,j}^{n,k,j} (t) E_{21} X_{t}^{n,k,j} dt
\]

\[
+ \sqrt{\varepsilon} (E_{12} X_{t}^{n,k+1,k} + E_{22} X_{t}^{n,k+1,j}) \circ dw_t
\]

\[
- \frac{1}{2} \varepsilon (E_{12} X_{t}^{n,k+2,k} + E_{22} X_{t}^{n,k+2,j}) dt
\]

(4.13)

for \( 0 \leq j < k < n \leq N, X_{t}^{n,n,j} = 1 \) and \( X_{t}^{n,k,j} = 0 \) for any \( k > n \). The system (4.12,4.13) can be written into a compact form

\[
dX_{t} = \sqrt{\varepsilon} A(X_t) \circ dw_t + \varepsilon B(X_t) dt
\]

\[
+ \sum_{k=1}^{N-1} \varepsilon \frac{k}{2} C_k(t, X_t) dt
\]

(4.14)

where all \( A, B \) and \( C_k \) defined by (4.12,4.13) are linear in the space variable, with bounded derivatives in \( t \). \( \circ dw_t \) denotes the Stratonovich differential.

**Proof.** By definition

\[
Y_t^\varepsilon = \sum_{n=1}^{N} \tilde{\sum}_{j_1, \ldots, j_n=1}^{\tilde{N}} C_{n}^{j_1 \cdots j_n} Z_{t}^{n,j_1^{f_{n}}, \ldots, j_n^{f_{n}}}.
\]

(4.15)
Instead of considering $Y^\varepsilon$ we take two copies of the same equation, i.e. we consider
\[
Z^\varepsilon_t = \begin{pmatrix} Y^\varepsilon_t \\ Y^\varepsilon_t \end{pmatrix} = \sum_{n=1}^{N} \sum_{j_1, \ldots, j_n=1}^{\tilde N} C_{n}^{j_1 \cdots j_n} X^n_t f_1^{j_1} \cdots f_n^{j_n}
\]
so that
\[
dZ^\varepsilon_t = \sum_{n=1}^{N} \sum_{j_1, \ldots, j_n=1}^{\tilde N} C_{n}^{j_1 \cdots j_n} dX^n_t f_1^{j_1} \cdots f_n^{j_n}. \tag{4.16}
\]
Using (4.8) we obtain
\[
dX^n_t f_1^{j_1} \cdots f_n^{j_n} = \sqrt{\varepsilon} (E_{12} + E_{22}) X^{n-1}_t f_1^{j_1} \cdots f_n^{j_n-1} \circ dw_t - \frac{1}{2} \varepsilon (E_{12} + E_{22}) X^{n-2}_t f_1^{j_1} \cdots f_n^{j_n-2} dt. \tag{4.17}
\]
so that
\[
dZ^\varepsilon_t = \sqrt{\varepsilon} \sum_{n=1}^{N} \sum_{j_1, \ldots, j_n=1}^{\tilde N} C_{n}^{j_1 \cdots j_n} (E_{12} + E_{22}) X^{n-1}_t f_1^{j_1} \cdots f_n^{j_n-1} \circ dw_t - \frac{1}{2} \varepsilon \sum_{n=1}^{N} \sum_{j_1, \ldots, j_n=1}^{\tilde N} C_{n}^{j_1 \cdots j_n} (E_{12} + E_{22}) X^{n-2}_t f_1^{j_1} \cdots f_n^{j_n-2} dt. \tag{4.18}
\]
Now repeating the use of Lemma 4.3 we obtain
\[
dX^{n-1}_t f_1^{j_1} \cdots f_n^{j_n-1} = \begin{pmatrix} \varepsilon^{-1} d f_n^{j_n} - \frac{1}{2} \varepsilon E_{12} \varepsilon^{-1} d f_n^{j_n-1} - \frac{1}{2} \varepsilon E_{22} \varepsilon^{-1} d f_n^{j_n-2} \\ \end{pmatrix} dt - \frac{1}{2} \varepsilon E_{12} X^{n-3}_t f_1^{j_1} \cdots f_n^{j_n-3} dt + \sqrt{\varepsilon} \left( E_{12} X^{n-2}_t f_1^{j_1} \cdots f_n^{j_n-2} + E_{22} X^{n-2}_t f_1^{j_1} \cdots f_n^{j_n-2} \right) \circ dw_t,
\]
35
\[ \begin{align*}
&dX_t^{n-2,(f_n^{j_{n-1}},f_n^{j_{n-2}})} \\
&= \epsilon^{n-2} \frac{d(f_n^{j_{n-1}})}{dt} E_{21} X_t^{n-2,(f_n^{j_{n-1}},f_n^{j_{n-2}},...)} dt \\
&+ \sqrt{\epsilon} \left( E_{12} X_t^{n-3,(f_n^{j_{n-2}},f_n^{j_{n-3}})} + E_{22} X_t^{n-3,(f_n^{j_{n-2}},f_n^{j_{n-3}})} \right) \circ dw \\
&- \frac{1}{2} \epsilon E_{12} X_t^{n-4,(f_n^{j_{n-3}},f_n^{j_{n-2}},f_n^{j_{n-4}})} dt \\
&- \frac{1}{2} \epsilon E_{22} X_t^{n-4,(f_n^{j_{n-3}},f_n^{j_{n-2}},f_n^{j_{n-4}})} dt,
\end{align*} \]

and

\[ \begin{align*}
&dX_t^{n-2,(f_n^{j_{n-1}},f_n^{j_{n-2}})} \\
&= \epsilon^{n-2} \frac{d(f_n^{j_{n}} f_n^{j_{n-1}})}{dt} E_{21} X_t^{n-2,(f_n^{j_{n-1}},f_n^{j_{n-2}})} dt \\
&+ \sqrt{\epsilon} \left( E_{12} X_t^{n-3,(f_n^{j_{n-2}},f_n^{j_{n-3}})} + E_{22} X_t^{n-3,(f_n^{j_{n-2}},f_n^{j_{n-3}})} \right) \circ dw \\
&- \frac{1}{2} \epsilon E_{12} X_t^{n-4,(f_n^{j_{n-3}},f_n^{j_{n-2}},f_n^{j_{n-4}})} dt \\
&- \frac{1}{2} \epsilon E_{22} X_t^{n-4,(f_n^{j_{n-3}},f_n^{j_{n-2}},f_n^{j_{n-4}})} dt.
\end{align*} \]

and so on. The general term appearing in this system is

\[ X_t^{n,k,j} \equiv X_t^{n-k,(g_{n,k,j},f_{n}^{j_{n-k}})}, \quad k = 1, \ldots, n-1, j = 0, \ldots, k-1, \]

where \( g_{n,k,j} = \prod_{i=j}^{k-1} f_n^{j_{n-i}}, n \) runs through 1 up to \( N \). We have thus completed the proof. \( \blacksquare \)

### 4.4 Itô-Lyons mappings

Let \( \xi = \sum_{n=1}^{N} J_n(f_n) \) be given by (4.9) and use the notations in the previous sub-section. For each \( \delta \in (0,1) \), we consider the following differential equation (4.14)

\[ \begin{align*}
\frac{dX_t}{dt} &= A(X_t) \circ dw_t + \delta B(X_t) dt \\
&+ \sum_{k=1}^{N-1} \delta^{\frac{k}{2}} C_k(t, X_t) dt.
\end{align*} \quad (4.19) \]

36
on the rough path space $\mathbb{W}^p$, where $A, B$ and $C_k$ are given in Proposition 4.5. According to Theorem 4.1, the differential equation (4.19) defines an Itô-Lyons mapping $G^\delta$ which is continuous with respect to the $p$-variation topology. The projection of $G^\delta$ to the first component $Y$ in Proposition 4.5 of the first level path (the projection is denoted by $\pi_1$) is then denoted by $F^\delta$. That is $F^\delta(w)_t = \pi_1 \left( G^\delta(w)_{0,t} \right)$. We also consider the differential equation
\[ dX_t = A(X_t) \circ dw_t, \quad X_0 = 0 \]
whose corresponding Itô-Lyons mappings are denoted by $\tilde{G}$ and $\tilde{F}$ (i.e. $\tilde{G} = G^0$ and $\tilde{F} = F^0$).

Let us list some properties about $F^\delta$.

Recall that $\mathbb{W}^p$ is the space of all rough paths in $\mathbb{R}^d$ endowed with the $p$-variation metric, $\tilde{\mu}$ is the distribution of Brownian motion with its area process, $(\mathbb{W}_0, \mathcal{F}_1, P^w)$ is the Wiener space, and $\mathbb{W}_0^d$ equipped with the uniform norm. The natural projection $\pi : \mathbb{W}^p \to \mathbb{W}_0^d$ which takes $w = (1, w_{1}, w_{2})$ to its first level path $w : t \in [0, 1] \to w_{1,t}$ is continuous.

**Proposition 4.6**

1) For each $\delta \in (0, 1)$, $F^\delta : \mathbb{W}^p \to \mathbb{W}_0^d$ is continuous, where $\mathbb{W}^p$ is equipped with the $p$-variation metric, $\mathbb{W}_0^d$ endowed with the uniform norm.

2) If $h \in H^1_0([0, 1]; \mathbb{R}^d)$, then $F^\varepsilon(\Gamma(\varepsilon)h)_t = \pi_1(x_t)$ where $(x_t)$ is the unique solution to the ordinary differential equation
\[ dx_t = \sqrt{\varepsilon A(x_t)} dh_t + \varepsilon B(x_t) dt + \sum_{k=1}^{N-1} \varepsilon^{\frac{k}{2}} C_k(t, x_t) dt. \quad (4.20) \]

3) For every $\varepsilon \in (0, 1)$, we have
\[ \tilde{P}^w \left\{ w : F^\varepsilon(\Gamma(\varepsilon)w)_t = Y^\varepsilon_t(w) \quad \forall t \in [0, 1] \right\} = 1. \quad (4.21) \]
where $w = w_{0,t}$ is the first level path of $w = (1, w^1, w^2) \in \mathbb{W}^p$.

4) We have
\[ \tilde{P}^w \left\{ w : \tilde{F}(\Gamma(\varepsilon)w)_t = S^\varepsilon_t(w) \quad \forall t \in [0, 1] \right\} = 1. \quad (4.22) \]
where
\[ S^\varepsilon_t = \sum_{n=1}^{N} \varepsilon^{\frac{n}{2}} \int_0^t f_n(t_1, \cdots, t_n) \circ dw_{t_1} \cdots \circ dw_{t_n} \]
and, if $h \in \mathbb{W}^\infty$ such that $t \to h_t = h_{0,t}^1 \in H_0^1([0,1];R^d)$, then

$$
\tilde{F}(h) = \sum_{n=1}^N \int_0^t f_n(t_1, \cdots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n.
$$

**Proof.** The first claim and second claim follow from Lyons’ continuity theorem, Theorem 4.1. The last item comes from the fact that the terms involving vector fields $B$ and $C_k$ come from the correction terms from Ito integrals to Stratonovich’s integrals, therefore if we started with the multiple Stratonovich’s integrals (or ordinary integrals), all these terms disappeared. We thus completed the proof.

**Proposition 4.7** Let $\delta > 0$. Consider the solutions $(x_t^\varepsilon)_{t \geq 0}$ and $(y_t^\varepsilon)_{t \geq 0}$ be the solutions to Stratonovich differential equations

$$
\begin{align*}
\frac{dy_t}{dt} &= \sqrt{\varepsilon}A(y_t) \circ dw_t + \varepsilon B(y_t) dt \\
&\quad + \sum_{k=1}^{N-1} \varepsilon^{\frac{k}{2}} C_k(t, y_t) dt , \quad y_0 = 0
\end{align*}
$$

and

$$
\begin{align*}
\frac{dx_t}{dt} &= \sqrt{\varepsilon}A(x_t) \circ dw_t , \quad x_0 = 0
\end{align*}
$$

on $(W_0^d, F_1, P^w)$, respectively, where $A, B$ and $C_k$ are given in Proposition 4.6. Then

$$
\lim_{\varepsilon \to 0} \varepsilon \log P^w \left\{ \sup_{t \in [0,1]} |\pi_1(x_t^\varepsilon) - \pi_1(y_t^\varepsilon)| > \delta \right\} = -\infty.
$$

**Proof.** According to the definition our system

$$
\pi_1(y_t^\varepsilon) = \sum_{n=1}^N \varepsilon^{\frac{n}{2}} \int_0^t f_n(t_1, \cdots, t_n) dw_{t_1} \cdots dw_{t_n}
$$

and

$$
\pi_1(x_t^\varepsilon) = \sum_{n=1}^N \varepsilon^{\frac{n}{2}} \int_0^t f_n(t_1, \cdots, t_n) \circ dw_{t_1} \cdots \circ dw_{t_n}
$$
so that (for example, by applying Hu-Meyer formula \cite{22}, see also \cite{31})
\[
\lim_{\varepsilon \to 0} \varepsilon \log P^w \left\{ \sup_{t \in [0,1]} |\pi_1(x_\varepsilon^t) - \pi_1(y_\varepsilon^t)| > \delta \right\} = -\infty.
\]

\[\square\]

**Corollary 4.8** Let \(\nu_\varepsilon\) be the law of \((\pi_1(y_\varepsilon^t))_{t \in [0,1]}\). Then the family \(\{\nu_\varepsilon : \varepsilon \in (0,1)\}\) satisfies the large deviation principle with the rate function given by
\[
I'_N(w) = \inf \{ I(h) : h \in H \text{ s.t. } \Phi(h) = w \} \tag{4.23}
\]
where \(I(h) = \frac{1}{2} \int_0^1 |\dot{h}(t)|^2 dt\) for \(h \in H\), and
\[
\Phi(h)_t = \sum_{n=1}^N \int_0^t f_n(t_1, \ldots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n. \tag{4.24}
\]

**Proof.** Let \(\tilde{G}\) be the Itô-Lyons mapping determined by the differential equation
\[
dx = A(x_t) \circ dw_t, \quad x_0 = 0
\]
on \(\mathbb{W}^p\). Then \(\tilde{G} : \mathbb{W}^p \to \mathbb{W}^d\) is continuous. Define \(\tilde{F} : \mathbb{W}^p \to \mathbb{W}^d\) by
\[
\tilde{F}(w)_t = \pi_1 \left( \tilde{G}(w)_{0,t} \right)
\]
which is continuous, and moreover \(\tilde{F}(\Gamma(\varepsilon)w)_t = \pi_1(x_\varepsilon^t)\), and
\[
\tilde{F}(h)_t = \sum_{n=1}^N \int_0^t f_n(t_1, \ldots, t_n) \dot{h}(t_1) \cdots \dot{h}(t_n) dt_1 \cdots dt_n
\]
for any \(h \in \mathbb{W}^\infty\) such that \(h = \pi_1(h) \in H^1_1([0,1]; \mathbb{R}^d)\). It follows from Theorem 4.1 the distributions of \((\pi_1(x_\varepsilon^t))\) satisfy the large deviation principle with the good rate function \(I'_N\) defined by (4.23). Now, according to Theorem 4.2.13 on page 130, \cite{7} and Proposition 4.7 one may conclude that \(\{\nu_\varepsilon : \varepsilon \in (0,1)\}\) satisfies the large deviation principle with the same rate function. The proof is complete. \(\square\)
5 Large deviations for martingales

In this section we extend the large deviation principle to a general square-integrable martingale on \((W_0^d, \mathcal{F}_1, \mathcal{F}_t, P^w)\).

Let \(\xi \in L^2(W_0^d, \mathcal{F}_1, P^w)\) with mean zero, whose Wiener-Itô chaos decomposition \(\xi = \sum_{k=1}^{\infty} J_k(f_k)_1\), where \(f_k \in L^2[0, 1]^k\) for every \(k\) and

\[
\|\xi\|^2 = \sum_{n=1}^{\infty} \frac{1}{n!} \|f_k\|_{L^2[0, 1]^k}\.
\]

Let \(Y_t = P^w(\xi|\mathcal{F}_t) = \sum_{k=1}^{\infty} J_k(f_k)_t\), and for each \(\varepsilon \in (0, 1)\)

\[
Y_t^\varepsilon = P_{-\log \sqrt{\varepsilon}} Y_t = \sum_{k=1}^{\infty} \varepsilon^{k/2} J_k(f_k)_t \quad \forall t \in [0, 1].
\]

Let \(\nu_\varepsilon\) be the law of \((Y_t^\varepsilon)_{t \in [0,1]}\) which is a probability measure on \((W_0^d, \| \cdot \|)\).

In this section, we prove the main result, Theorem 1.1, that is, we show that \(\{\nu_\varepsilon: \varepsilon \in (0, 1)\}\) satisfies the large deviation principle on \((W_0^d, \| \cdot \|)\).

The idea, as we have mentioned, is to construct a sequence of exponential approximations to \((Y_t^\varepsilon)_{t \in [0,1]}\). For each natural number \(n\), there is a natural number \(N_n\) such that

\[
\sum_{k=1}^{\infty} \frac{1}{k!} \|f_k\|^2_{L^2[0, 1]^k} < \frac{1}{2n^2}.
\]

For each \(k = 1, \ldots, N_n\), choose a symmetric function \(f_k^n\) on \([0, 1]^k\) which has a product form

\[
f_k^n(t_1, \ldots, t_k) = \sum_{j_1, \ldots, j_n = 1}^{N_n, k} C_{j_1 \cdots j_k} f_k^j(t_1) \cdots f_k^j(t_k)
\]

where \(C_{j_1 \cdots j_k}\) are constants, \(N_n, k\) is a and all \(f_k^j\) are smooth functions on \([0, 1]\) with bounded derivatives, such that

\[
\|f_k - f_k^n\|^2_{L^2[0, 1]^k} < \frac{1}{2en^2} \quad \text{for } k = 1, \ldots, N_n.
\]

Define \(\xi_n = \sum_{k=1}^{N_n} I_k(f_k^n)_1\). Then

\[
\sum_{k=1}^{N_n} \frac{1}{k!} \|f_k - f_k^n\|^2_{L^2[0, 1]^k} < \frac{1}{2n^2}.
\]
so that
\[ ||\xi - \xi_n||^2_2 = \sum_{k=1}^{N_n} \frac{1}{k!} ||f_k - f^n_k||^2_{L^2[0,1]^k} + \sum_{k=N_n+1}^\infty \frac{1}{k!} ||f_k||^2_{L^2[0,1]^k} \]
\[ < \frac{1}{n^2} \]
which implies that \( \xi_n \to \xi \) in \( L^2(W^1_0, \mathcal{F}_1, P^\mu) \). We of course can choose \( N_n \) increasing in \( n \). It is obvious that for each \( k \), \( f^n_k \to f_k \) as \( n \to \infty \).

Let \( Y(n)_t = E^\mu(\xi_n|\mathcal{F}_t) \) and
\[ Y(n)^\varepsilon_t = P_{-\log \sqrt{\varepsilon}} Y(n)_t = \sum_{k=1}^{N_n} \varepsilon^{\frac{k}{2}} I_k(f^n_k)_t \]

Let \( \nu^\varepsilon_n \) denote the distribution of \( (Y(n)^\varepsilon)_t \).

Let \( A_n, B_n, C_{j,n} \) be the corresponding vector fields determined in Proposition 4.5 for each \( \xi_n \) in place of \( \xi \). Then \( Y(n)^\varepsilon_t = \pi_1(y^{\varepsilon,n}_t) \), where \( (y^{\varepsilon,n})_{t \geq 0} \) is the unique strong solution to
\[ dy_t = \sqrt{\varepsilon} A_n(y_t) \circ dw_t + \varepsilon B_n(y_t) dt \\
+ \sum_{j=1}^{N_n-1} \varepsilon^{\frac{j}{2}} C_{j,n}(t, y_t) dt, \quad y_0 = 0 \tag{5.1} \]
on \( (W^d_0, \mathcal{F}_1, P^\mu) \). Let \( X(n)^\varepsilon_t = \pi_1(x^{\varepsilon,n}_t) \) where \( (x^{\varepsilon,n})_{t \geq 0} \) is the unique strong solution to
\[ dx_t = \sqrt{\varepsilon} A_n(x_t) \circ dw_t, \quad x_0 = 0 \tag{5.2} \]

**Lemma 5.1** Both families \( \{(Y(n)^\varepsilon)_t : t \leq 1, \varepsilon \in (0,1)\}_{n=1,2,\ldots} \) and \( \{(X(n)^\varepsilon)_t : t \leq 1, \varepsilon \in (0,1)\}_{n=1,2,\ldots} \) converge to \( \{(Y^\varepsilon_t)_{t \leq 1} : \varepsilon \in (0,1)\} \) exponentially. That is, for each \( \delta > 0 \),
\[ \lim_{N \to \infty} \lim_{\varepsilon \to 0} \varepsilon \log P^\mu \left\{ \sup_{t \leq 1} |Y(n)^\varepsilon_t - Y^\varepsilon_t| \geq \delta \right\} = -\infty \tag{5.3} \]
and
\[ \lim_{N \to \infty} \lim_{\varepsilon \to 0} \varepsilon \log P^\mu \left\{ \sup_{t \leq 1} |X(n)^\varepsilon_t - Y^\varepsilon_t| \geq \delta \right\} = -\infty. \tag{5.4} \]
Proof. By Proposition 4.7, we only need to show (5.3). By Lemma 3.3, for any \( \delta > 0, \varepsilon \in (0, 1) \) we have

\[
P_w\left\{ \sup_{t \leq 1} |Y(n)^{\varepsilon}_t - Y^\varepsilon_t| \geq \delta \right\} \leq (1 + \varepsilon)^{1 + \frac{1}{2}} \frac{||\xi_n - \xi||_2^{1 + \frac{1}{2}}}{\delta^{1 + \frac{1}{2}}}
\]

so that

\[
\varepsilon \log P_w\left\{ \sup_{t \leq 1} |Y(n)^{\varepsilon}_t - Y^\varepsilon_t| \geq \delta \right\} \\
\leq \varepsilon \left(1 + \frac{1}{\varepsilon}\right) \log(1 + \varepsilon) - \varepsilon \left(1 + \frac{1}{\varepsilon}\right) \log \delta \\
+ \varepsilon \left(1 + \frac{1}{\varepsilon}\right) \log ||\xi_n - \xi||_2.
\]

Hence

\[
\lim_{\varepsilon \to 0} \varepsilon \log P_w\left\{ \sup_{t \leq 1} |Y(n)^{\varepsilon}_t - Y^\varepsilon_t| \geq \delta \right\} \\
\leq \log ||\xi_n - \xi||_2 - \log \delta
\]

and therefore

\[
\lim_{N \to \infty} \lim_{\varepsilon \to 0} \varepsilon \log P_w\left\{ \sup_{t \leq 1} |Y(n)^{\varepsilon}_t - Y^\varepsilon_t| \geq \delta \right\} = -\infty.
\]

To prove the large deviation principle for the limit distributions of \( (Y^\varepsilon_t)_{t \leq 1} \), one would attempt to apply an extended contraction principle (for example, Theorem 4.2.23, page 133, [7]) to the exponential approximations \( X(n)^\varepsilon \). Since \( \tilde{F}_n(w)_t = \pi_1(\tilde{G}_n(\varepsilon)(w)^1_{0,t}) \) is a version of \( X(n)^\varepsilon_t \), which approximate \( \{Y^\varepsilon : \varepsilon \in (0,1)\} \) exponentially, where \( \tilde{G}_n \) is the Itô-Lyons mapping on \( (\mathbb{W}^p, d_p) \) associated with

\[
dx_t = A_n(x_t) \circ dw_t, \quad x_0 = 0.
\]

The mapping \( F_n : (\mathbb{W}^p, d_p) \to (\mathbb{W}^d_0, ||\cdot||) \) is continuous, and the distribution family of \( \{X(n)^\varepsilon_{t \leq 1} : \varepsilon \in (0,1)\} \) satisfies the large deviation principle with rate function given by

\[
I_n'(w) = \inf \{I(h) \mid h \in H \text{ such that } F_n(h) = w\}
\]

42
where
\[ \tilde{F}_n(h)_t = \sum_{k=1}^{N_n} \int_0^t f_k^h(t_1, \ldots, t_k) \dot{h}(t_1) \cdots \dot{h}(t_k) dt_1 \cdots dt_k \quad \forall h \in H. \]

Therefore, if \( \tilde{F}_n \) were convergent uniformly (in p-variation distance) on any level set \( \{ I(h) \leq L \} \) uniformly, one could conclude the proof of Theorem 1.1.

However, unfortunately, as a matter of fact, \( \tilde{F}_n \) does not converge uniformly on \( \{ I(h) \leq L \} \) in p-variation metric \( d_p \) in general, which would require a control on the derivatives of \( f_k \). Thus, we can not prove our main theorem by simply appealing to a (generalized) contraction principle.

This is the reason why we develop the continuity theorem for large deviations, Theorem 2.5.

**Proof of Theorem 1.1** Let \( E = W_0^d \) with the uniform norm, \( H = H_0^1([0,1]; R^d) \) with the Sobolev norm \( || \cdot ||_{H^1} \). Then \( I \) is a good rate function with the effective set \( \{ I < \infty \} = H \). For each \( N \) consider the following mappings \( F_n : H \rightarrow E \), where
\[ F_n(h)_t = \sum_{k=1}^{N_n} \int_{0 < t_1 < \cdots < t_k < t} f_k^h(t_1, \ldots, t_k) \dot{h}(t_1) \cdots \dot{h}(t_k) dt_1 \cdots dt_k \quad t \in [0, 1] \]
and \( F : H \rightarrow E \) by
\[ F(h)_t = \sum_{k=1}^{\infty} \int_{0 < t_1 < \cdots < t_k < t} f_k(t_1, \ldots, t_k) \dot{h}(t_1) \cdots \dot{h}(t_k) dt_1 \cdots dt_k \quad t \in [0, 1]. \]

Then, according to Proposition 2.4, \( F_n, F \) are rate-function mappings. It is easy to see that \( F_n \rightarrow F \) uniformly on any level set \( K_L = \{ h : I(h) \leq L \} \) with respect to the uniform norm, thus, according to Corollary 4.3 for each \( n \), both the laws of \( \{ X(n)^\varepsilon : \varepsilon \in (0, 1) \} \) and \( \{ Y(n)^\varepsilon : \varepsilon \in (0, 1) \} \) satisfy the large deviation principle on \( (W_0^d, || \cdot ||) \) with the common rate function
\[ I'_n(s) = \inf \{ I(h) : h \in H \quad \text{s.t.} \quad F_n(h) = s \} \]
and \( \{ Y(n)^\varepsilon : \varepsilon \in (0, 1) \} \) goes to \( \{ Y^\varepsilon : \varepsilon \in (0, 1) \} \) exponentially, therefore, by Theorem 2.5 \( \{ Y^\varepsilon : \varepsilon \in (0, 1) \} \) satisfies the large deviation principle with rate function
\[ I'(s) = \inf \{ I(h) : h \in H \quad \text{s.t.} \quad F(h) = s \}. \]
Acknowledgments. The first author would like to thank Professor M. Ledoux for his comments, and to thank Professor Quansheng Liu for his references on large deviations for martingales in discrete-time. The research of the paper was partly supported by EPSRC grant EP/F029578/1.

References

[1] Azencott, R., Grandes déviations et applications. Ecole d’Été de Probabilités de Saint-Flour VIII; *Lecture Notes in Mathematics* **774**, Springer, Berlin (1980).

[2] Bismut, J.-M., Large Deviations and the Malliavin Calculus. Progress in Mathematics, Birkhauser Boston (1984).

[3] Borell, C., Tail Probabilities in Gauss Space. Vector Space Measures and Applications, Dublin, Springer Berlin-Heidelberg-New York. *Lecture Notes in Math.** **644**, pp.73-82, (1978).

[4] Bolthausen, E., On the probability of large deviations in Banach spaces. *Ann. Probab.** **12** (1984), 427-435.

[5] Cass, T. and Friz, P., Densities for rough differential equations under Hörmander’s condition. To appear in *Annals of Math.*

[6] Chernoff, H., A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** (1952), 493-507.

[7] Dembo, A. and Zeitouni, O., Large Deviations Techniques and Applications. *Springer-Verlag, New York*, (1998).

[8] Deuschel, J.-D. and Stroock, D. W., Large Deviations, AMS Chelsea Publishing (1989).

[9] Donsker, M. and Varadhan, S. R.S., Asymptotic evaluation of certain Markov process expectations for large time, I and III. *Comm. Pure Appl. Math.* **28** and **29**, pp. 1-47 and pp. 389-461 (1975) and (1976).

[10] Donsker, M. and Varadhan, S. R.S., Large deviations for stationary Gaussian processes. *Comm. Math. Phys.* **97**, 187-210 (1985).
[11] Dupuis, P. and Ellis, R., A weak Convergence Approach to the Theory of Large Deviations. J. Wiley (1997).

[12] Ellis, R. S., Large deviations for a general class of random vector. Ann.Probab., 12, 1-12 (1984).

[13] Ellis, R. S., Entropy, Large Deviations and Statistical Mechanics. Springer-Verlag, New York (1985).

[14] Feller, W., An Introduction to Probability Theory and its Applications. Third Edition. Volume 1, John Wiley and Sons, Inc. (1968).

[15] Freidlin, M. I. and Wentzell, A. D., Random Perturbations of Dynamical Systems. Springer-Verlag (1984).

[16] Friz, P. and Victoir, N., Multidimensional Stochastic Processes as Rough Paths. Cambridge University Press (2008).

[17] Grama I., On moderate deviations for martingales. Annals of Probability, Vol. 25, No. 1, pp. 152-183 (1997).

[18] Grama I. and Haeusler E., Large deviations for martingales via Cramer’s method. Stochastic Processes and its Applications. Vol. 85, pp. 279-293 (2000).

[19] Cramér, H., Sur un nouveau théorème-limite de la théorie des probabilités. Actualités Scientifiques et Industrielles 736 (1938), 5-23, Colloque consacré à la théorie des probabilités, Vol. 3, Hermann, Paris.

[20] Gross, L., Logarithmic Sobolev inequalities. Amer. J. Math. 97, 1061-1083 (1976).

[21] Hambly, B. and Lyons, T., Uniqueness for the signature of a path of bounded variation and the reduced path group. To appear in The Annals of Math.

[22] Hu, Y. Z. and Meyer, P. A., Sur les integrales multiples de Stratonovich. Seminaire de Probabilites XXII (L.N.Math.1321), J.Azema, P.A.Meyer and M.Yor, eds, Springer Berlin-Heidelberg-New York, 72-81, (1988).
[23] Ikeda, N. and Watanabe, S., Stochastic Differential Equations and Diffusion Processes. North-Holland Pub. Company, Amsterdam, Oxford, New York (1981).

[24] Itô, K., Multiple Wiener integral. J. Math. Soc. Japan 3, 157-169 (1951).

[25] Ledoux, M., A Note on Large Deviations for Wiener Chaos. Séminaire de Probabilité, XXIV (L.N. Math. 1426):1-14 (J. Azema, P.A. Meyer and M. Yor, eds), Springer, Berlin-Heidelberg-New York, (1990).

[26] Ledoux, M., Qian, Z. and Zhang, T., Large deviations and support theorem for diffusion processes via rough paths. Stoch. Proc. Appl., 102:265-283, (2002).

[27] Lesigne, E. and Volny, D., Large deviations for martingales. Stochastic Processes and their Applications 96 (2001).

[28] Lyons, T., Differential equations driven by rough signals. Rev. Math. Iberoamericana, 2(14):215-310, (1998).

[29] Lyons, T. and Qian, Z., System Control and Rough Paths. Oxford Science Publications, (2002).

[30] Lyons, T., Caruana, M. and Lévy, T., Differential Equations driven by Rough Paths. École d’Été de Probabilités de Saint-Flour XXXIV-2004, Lecture Notes in Math. 1908, Springer 2007.

[31] Mayer-Wolf, E., Nualart, D. and Peres-Abreu, V., Large Deviations for Multiple Wiener-Itô Integral Processes. Séminaire de Probability, Lecture Notes in Math., Springer, Berlin, XXVI:11–31, (1992).

[32] Millet, A. and Sanz, M., Large deviations for rough paths of the fractional Brownian motion. Annales de l’Institut Henri Poincare (B) Probability and Statistics., Volume 42, Issue 2, March-April 2006, Pages 245-271 (2006).

[33] Nualart, D. and Zakai, M., Multiple Wiener-Itô integrals possessing a continuous extension. Probab. Th. Rel. Fields, 85:131-145, (1990).

[34] Revuz, D. and Yor, M., Continuous Martingales and Brownian Motion. Thrid Edition. Springer, 1999.
[35] Schilder, M., Some asymptotics formular for Wiener integrals. *Trans. Amer. Math. Soc.*, 125:63-85, (1966).

[36] Stroock, D. W., An Introduction to the Theory of Large Deviations. Springer-Verlag (1984).

[37] Ventcel, A. D. and Freidlin, M. I., On small perturbations of dynamical systems. *Russian Math. Surveys* 25 (1970), 1-55.

[38] Varadhan, S. R. S., Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261-286, (1966).

[39] Varadhan, S. R. S., Large Deviations and Applications. *SIAM, Philadelphia*, (1984).

[40] Wiener, N., Generalized harmonic analysis. *Acta Math.* 55, 117-258 (1930).

[41] Yosida, K., Functional Analysis. Sixth Edition. Springer-Verlag (1980).

Z. Qian and C. Xu, Mathematical Institute, University of Oxford, 24 - 29 St. Giles', Oxford OX1 3LB

Email: qianz@maths.ox.ac.uk