ELSEVIER

The enhanced Sanov theorem and propagation of chaos

Jean-Dominique Deuschel[a], Peter K. Friz[a,b], Mario Maurelli[a,b,*], Martin Slowik[a]

[a] *Institut für Mathematik, Technische Universität Berlin, Germany*
[b] *Weierstraß–Institut für Angewandte Analysis und Stochastik, Berlin, Germany*

Abstract

We establish a Sanov type large deviation principle for an ensemble of interacting Brownian rough paths. As application a large deviations for the (k-layer, enhanced) empirical measure of weakly interacting diffusions is obtained. This in turn implies a propagation of chaos result in a space of rough paths and allows for a robust analysis of the particle system and its McKean–Vlasov type limit, as shown in two corollaries.
ⓒ 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction and main results

1.1. Large deviation and rough paths

The present paper is concerned with the intersection of large deviations, rough paths and (weakly) interacting diffusions. We note (i) that large deviations have been one of the first application areas of rough paths theory: indeed, following Ledoux et al. [15], a large deviation principle for Brownian motion and Lévy's area, scaled by ϵ and ϵ^2 respectively, in rough path topology, will yield immediately the Freidlin–Wentzell theory of large deviations for diffusions with small noise – it suffices to combine continuity of the Itô-map in rough path sense with the contraction principle of large deviation theory; [13]. (ii) The interplay of rough paths with

---

interacting stochastic differential equations was pioneered in [4]. This work, as well as the more recent [1], required in particular the development of a McKean–Vlasov theory in the context of random rough differential equations (which is not at all the aim of this paper). At last, (iii) large deviations for interacting diffusions is a huge field, a small selection of relevant references is given by [5–7,17,18].

In sense, we combine here aspects of all the afore-mentioned references. In particular, when compared to the many classical works (iii) an advantage of our approach is robustness: as soon as we have a LDP on a suitably enhanced space ("enhanced Sanov") – on which most stochastic operations of interest are continuous, the raison d'être of rough paths – basic facts of large deviation theory, such as contraction principle or Varadhan's lemma become directly applicable. On the contrary, stronger versions of contraction principles or Varadhan's lemma need suitable approximated continuity properties which must be checked case by case.

We briefly describe our main results. Let $\{B^i : i \in \mathbb{N}\}$ be a family of independent d-dimensional standard Brownian motions,[1] on a fixed filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_t, \mathbb{P})$. On a finite time-horizon, say $[0, T]$, we may regard them as $C([0, T]; \mathbb{R}^d)$-valued i.i.d. random variables. By a classical law of large numbers (LLN) argument (e.g. [10, Thm 11.4.1]) the *empirical measure*, L_n, a random measure on pathspace, converges to the d-dimensional Wiener measure $P^{\{d\}}$. More precisely, with probability one,[2]

$$L_n^B(\omega) := \frac{1}{n} \sum_{i=1}^n \delta_{B^i(\omega)} \longrightarrow P^{\{d\}} \qquad \text{as} \quad n \to \infty.$$

Sanov's theorem quantifies the speed of this convergence: for a measure Q on $C([0, T]; \mathbb{R}^d)$,

$$\mathbb{P}\big[L_n^B(\omega) \approx Q\big] \approx \exp\big(-n\, H(Q \mid P^{\{d\}})\big),$$

in the form of a large deviation principle [8,9], where H is the relative entropy. Now, for each $1 \le i, j \le n$, we introduce the $2d$-dimensional *double-layer* process $B^{\{2\};ij} \equiv B^{ij}$ as

$$B_t^{ij} := \big(B_t^i, B_t^j\big) \in \mathbb{R}^{2d}, \tag{1.1}$$

and then define the *enhanced double-layer* process $\mathbb{B}^{\{2\};ij} \equiv \mathbb{B}^{ij}$, with values in the space of $2d \times 2d$ matrices, as

$$\mathbb{B}_t^{ij} = \int_0^t B_s^{ij} \otimes \circ\, \mathrm{d}B_s^{ij}, \tag{1.2}$$

where $\circ$ denotes Stratonovich integration. Clearly, for any $i \neq j$, we have

$$\mathrm{Law}(B^{ij}) = \mathrm{Law}\big(B^{12}\big) = P^{\{2d\}},$$

where $P^{\{e\}}$ denotes *e-dimensional Wiener-measure*. We are interested in (the $G^2(\mathbb{R}^{2d})$-valued process)

$$\mathbf{B}^{\{2\};ij} \equiv \mathbf{B}^{ij} \equiv (B^{ij}, \mathbb{B}^{ij}) := ((B^i, B^j), \mathbb{B}^{ij}),$$

with law

$$\mathrm{Law}(\mathbf{B}^{ij}) = \mathrm{Law}(\mathbf{B}^{12}) = \mathrm{Law}(B^1, B^2, \mathbb{B}^{12}) =: \mathbf{P}^{\{2d\}},$$

---

[1] Later on, we shall allow for non-trivial $Law(B_0^i) \equiv \lambda$.

[2] We regard L_n^B and P^d as random variables with values in the (Polish) space $\mathcal{P}(C([0, T]; R^d))$, equipped with the C_b-weak topology.

where $\mathbf{P}$ denotes the *enhanced (e-dimensional) Wiener-measure*.[3]   For every n, define the *enhanced "double-layer" empirical measure* as

$$\mathbf{L}_n^{\mathbf{B};\{2\}}(\omega) \equiv \mathbf{L}_n^{\mathbf{B}}(\omega) \; := \; \frac{1}{n^2} \sum_{i,j=1}^{n} \delta_{(B^{ij}, \mathbb{B}^{ij})(\omega)} \; \equiv \; \frac{1}{n^2} \sum_{i,j=1}^{n} \delta_{\mathbf{B}^{ij}(\omega)}. \tag{1.3}$$

In order to extend the enhanced "double-layer" ($k = 2$) empirical measure to any $k \geq 3$, define the (kd-dimensional) *k-layer process* $B^{\{k\};i_1,\ldots,i_k} \equiv (B^{i_1}, \ldots, B^{i_k})$, its rough path lift $\mathbf{B}^{\{k\};i_1,\ldots,i_k}$, and then the *enhanced "k-layer" empirical measure* given by $\mathbf{L}_n^{\mathbf{B};\{k\}}(\omega) := n^{-k} \sum \delta_{\mathbf{B}^{\{k\};i_1,\ldots,i_k}(\omega)}$ with summation over all $1 \leq i_1, \ldots, i_k \leq n$. One may expect, as suggested by our notation, that, for any integer k,[4]

$$\mathbf{L}_n^{\mathbf{B};\{k\}}(\omega) \; \longrightarrow \; \mathbf{P}^{\{kd\}} \qquad \text{as} \quad n \to \infty. \tag{1.4}$$

This is indeed the case, however not a consequence of LLN, for even when $k = 2$ the $\{\mathbf{B}^{ij} : i, j = 1, \ldots, n\}$ are not independent. In fact, we shall study the speed of convergence around this limit: one of our main results is a large deviation principle for the law of $\mathbf{L}_n^{\mathbf{B};\{k\}}$, which gives (1.4), with convergence in probability (and a.s. by a Borel–Cantelli argument) with respect to the α-Hölder rough path topology, as a byproduct. Here and below $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$ denotes a space of rough paths (in notation of [12], some recalls below), a Polish space, elements of which are paths with values in the group $G^{(2)}(\mathbb{R}^e) \subset \mathbb{R}^e \otimes (\mathbb{R}^e)^{\otimes 2}$.

Theorem 1.1. *Fix α in $(1/3, 1/2)$. The sequence of laws $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}}) : n \in \mathbb{N}\}$ satisfies a large deviation principle on $\mathcal{P}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd}))$ endowed with the C_b-weak topology, with scale n and good rate function $\mathbf{I} : \mathcal{P}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd})) \to \mathbb{R} \cup \infty$ that is given by*

$$\mathbf{I}^{\{k\}}(\mu) \equiv \mathbf{I}(\mu) \; = \; \begin{cases} H(\mu \circ \pi_1^{-1} \mid P^{\{d\}}), & \text{if } \mu = F^{\{k\}}(\mu \circ \pi_1^{-1}), \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.5}$$

This LDP is also valid in a stronger ("modified Wasserstein") topology.

Here $\pi_1 : G^{(2)}(\mathbb{R}^{kd}) \cong G^{(2)}(\oplus_{i=1}^{k} \mathbb{R}^d) \to \mathbb{R}^d$ is given by the projection $((x^1, \ldots, x^k), \ldots) \mapsto x^1$. In particular, given a probability μ on $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd})$, the image measure $Q := (\pi_1)_* \mu \equiv \mu \circ \pi_1^{-1}$ is a measure on the classical Hölder space $C^{0,\alpha}([0, T], \mathbb{R}^d)$. Moreover, $H(.\,|P^{\{d\}})$ is the relative entropy and

$$F^{\{k\}} : Q \; \longmapsto \; Q^{\otimes k} \circ (S^{\{kd\}})^{-1} \tag{1.6}$$

defines a map $F^{\{k\}} : \mathcal{P}_1(C^{0,\alpha}([0, T]; \mathbb{R}^d)) \to \mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd}))$, where $S = S^{\{e\}}$ denotes the (measurable) lifting map of an e-dimensional path to a path with values in $G^2(\mathbb{R}^e)$.

The interest in a *modified Wasserstein* topology (on probability measures on the space of rough paths, Section 4 for details) is continuity of the map (here $k = 2$, but then trivially for $k \geq 2$ by projection)

$$\mu \; \longmapsto \; \int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} \int_0^T \bar{b}(X_t) \, \mathrm{d}\mathbf{X}_t \; \mu(\mathrm{d}\mathbf{X})$$

---

[3] That is, the law of e-dimension Brownian motion B and all its iterated integrals of the form $\int B^k \circ \mathrm{d}B^l$, $1 \leq k, l \leq e$.

[4] Again we regard $\mathbf{L}_n^{\mathbf{B};\{k\}}$ and $\mathbf{P}^{\{kd\}}$ as random variables with values in the (Polish) space $\mathcal{P}(C([0, T]; G^2(\mathbb{R}^{kd})))$, equipped with the C_b-weak topology.

for sufficiently nice $\bar{b}$. Indeed, combining Girsanov's theorem and Varadhan's lemma will then imply a LDP for the empirical measures, as $n \to \infty$, for the particle system given by[5]

$$
\begin{cases}
dX_t^{i,n} = \dfrac{1}{n} \sum_{j=1}^{n} b\big(X_t^{i,n}, X_t^{j,n}\big) \, dt + dB_t^i, & i = 1, \dots n, \\
\mathrm{Law}(X_0^{i,n}) = \lambda & \text{i.i.d.}
\end{cases}
\tag{1.7}
$$

In fact, our approach not only allows to recover the (known, see e.g. [2,7]) rate function for the large deviations of such a particle system, of the form $J_b(Q) = H(Q \,|\, \Phi(Q))$ cf. Sections 6 and 7 (where $\Phi = \Phi_b$ is introduced, such that fixed points of Φ are solutions to the martingale problem of the corresponding McKean–Vlasov equation with mean-field drift b), but it gives the LDP on the level of k-layer enhanced empirical measures. We shall see in two applications, namely Corollaries 1.5 and 1.6 below, how useful exactly this can be.

Theorem 1.2. *Assume that b is in $C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$, let $X^n = (X^{1,n}, \dots, X^{n,n})$ be the solution to the above system where the initial law satisfies a suitable exponential integrability condition (Condition (3.15)). Let $\mathbf{L}_n^{X,\{k\}}$ be the corresponding enhanced k-layer empirical measure, $k \geq 2$. Fix $\alpha \in (1/3, 1/2)$. Then the sequence of laws $\{\mathrm{Law}(\mathbf{L}_n^{X,\{k\}}) : n \in \mathbb{N}\}$ satisfies a large deviation principle on (a modified Wasserstein) space of probability measures on $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd})$ with scale n and good rate function $\mathbf{J}_b$ given by*

$$
\mathbf{J}_b^{\{k\}}(\mu) \equiv \mathbf{J}_b(\mu) =
\begin{cases}
H(\mu \circ \pi_1^{-1} \,|\, \Phi(\mu \circ \pi_1^{-1})), & \text{if } \mu = F^{\{k\}}(\mu \circ \pi_1^{-1}), \\
+\infty, & \text{otherwise.}
\end{cases}
\tag{1.8}
$$

A first consequence of this large deviation principle, together with the fact that the rate function has a unique zero, is a "law of large number" which already contains a remarkably strong form of *propagation of chaos* (POC), namely Theorem 1.4 below. Note that this result can also be recovered via classical Itô calculus (the reader can verify this as an exercise), nevertheless it illustrates well the extra information carried by the LDP above, moreover its Corollary 1.5 is another example of the combination of mean field and rough paths arguments. For context, we first give the classical form of POC. Let us also note there is much new interest in POC, with recent applications ranging from calibration methods in quantitative finance to the analysis of lithium-ion batteries.

Theorem 1.3 (*Classical POC, e.g. [17]*). *Let $\{\bar{X}^j : j \in \mathbb{N}\}$ be an i.i.d. realizations of the McKean–Vlasov diffusion $\bar{X}$ (see Section 7 for details). Then, for all $k \in \mathbb{N}$,*

$$
\mathrm{Law}\big(X^{1,n}, \dots, X^{k,n}\big) \xrightarrow[n \to \infty]{} \mathrm{Law}\big(\bar{X}^1, \dots, \bar{X}^k\big) = \mathrm{Law}(\bar{X})^{\otimes k},
\tag{1.9}
$$

as C_b-weak[6] convergence of probability measures on $\big(C([0, T], \mathbb{R}^d)\big)^{\times k} \cong C([0, T], \mathbb{R}^{kd})$ equipped with uniform topology.

In classical terminology [17] the law of $\big(X^{1,n}, \dots, X^{k,n}\big)$ is $\mathcal{Q}$-*chaotic*, where $\mathcal{Q} = \mathrm{Law}(\bar{X})$ is a probability measure on the (Polish) space $E = C([0, T], \mathbb{R}^d)$.

We now state the enhanced POC on the space of rough paths, that is paths with values in $G^2(\mathbb{R}^N)$ rather than $\mathbb{R}^N$. We insist that this is not just a form of the classical POC (a.k.a.

---

[5] Given a function $b: \mathbb{R}^{2d} \to \mathbb{R}^d$, we use the notation $(x^1, x^2) \in \mathbb{R}^d \times \mathbb{R}^d = \mathbb{R}^{2d}$ and we denote by $\bar{b}: \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ the function such that $\bar{b}(x^1, x^2)^1 = b(x^1, x^2)$ and $\bar{b}(x^1, x^2)^2 = 0$.

[6] Actually, in 1-Wasserstein sense ....

Q-chaos) in which $E = C([0, T], \mathbb{R}^d)$ is replaced by some other (Polish) space, which happens to be a space of rough paths. To wit, the limiting measure in our Theorem 1.4 below is *not* of product measure form, since it effectively tracks all areas between the particle trajectories (in the mean-field limit) which requires it to be a measure on the space of *geometric rough paths*

$$\mathcal{C}_g^{0,\alpha}\big([0, T], \mathbb{R}^{kd}\big) \;\cong\; C_g^{0,\alpha}\big([0, T], G^2(\mathbb{R}^{kd})\big),$$

which indeed offers enough room to capture

$$\int X^i \otimes \circ \, \mathrm{d}X^j \dots \quad \text{for } 1 \leq i, j \leq k$$

(the anti-symmetric part of which corresponds to the afore-mentioned areas). In contrast, a space of k rough paths over $\mathbb{R}^d$, say

$$\big(\mathbf{X}^1, \dots, \mathbf{X}^k\big) \;\in\; \mathcal{C}_g^{0,\alpha}\big([0, T], \mathbb{R}^d\big)^{\times k} \;\cong\; C_g^{0,\alpha}\big([0, T], \oplus_{i=1}^k G^2(\mathbb{R}^d)\big)$$

contains strictly less information as it contains, particle trajectories on $\mathbb{R}^d$ aside, only

$$\int X^i \otimes \circ \, \mathrm{d}X^i \quad \text{for } 1 \leq i \leq k$$

(and hence only the areas of each single d-dimensional particle trajectory).

This extra information contained in $\mathcal{C}_g^{0,\alpha}\big([0, T], \mathbb{R}^{kd}\big)$ makes a difference indeed when one is interested in subsequent analysis of this particle system, as we shall see in the corollary below. But first we state our enhanced POC. Recall that for an e-dimensional semi-martingale Z, its Stratonovich (level 2) lift is given by

$$S^{\{e\}}(Z) \;=\; \Big(Z^i \,:\, 1 \leq i \leq e; \int Z^i \otimes \circ \, \mathrm{d}Z^j \,:\, 1 \leq i, j \leq e\Big).$$

Theorem 1.4 (*Enhanced POC*). *Under the assumptions of the classical POC for all $k \in \mathbb{N}$*

$$\mathrm{Law}\big(S^{\{kd\}}(X^{1,n}, \dots, X^{k,n})\big) \;\underset{n\to\infty}{\longrightarrow}\; \mathrm{Law}\big(S^{\{kd\}}(\bar{X}^1, \dots, \bar{X}^k)\big), \tag{1.10}$$

as C_b-weak convergence of probability measures on $\mathcal{C}_g^{0,\alpha}([0, T], \mathbb{R}^{kd})$ equipped with α-Hölder geometric rough path topology.

We now illustrate the power of this new form of propagation of chaos. Recall that the solution flow to an SDE depends continuously on the driving noise in rough path topology (e.g. [13]). We then have immediately the following result, a direct proof of which would require substantial work.

Corollary 1.5. *Fix some $k \in \mathbb{N}$ and consider, for $n \geq k$, the solution flow $Y^n \equiv Y$ to*

$$\mathrm{d}Y_t \;=\; f_0(Y) \, \mathrm{d}t \,+\, \sum_{i=1}^k f_i(Y) \circ \mathrm{d}X^{i,n} \tag{1.11}$$

where the f_i's are C_b^3 vector fields on $\mathbb{R}^N$ in the case $d = 1$ (or $f_i \in C_b^3(\mathbb{R}^N; L(\mathbb{R}^d, \mathbb{R}^N))$ more generally). Then, (in the sense of flows, cf. [13], and $1/2^-$-Hölder on compacts in time)

$$\mathrm{Law}(Y^n) \;\underset{n\to\infty}{\longrightarrow}\; \mathrm{Law}(\bar{Y}), \tag{1.12}$$

where the weak limit flow is given by

$$\mathrm{d}\bar{Y}_t \;=\; f_0(\bar{Y}) \, \mathrm{d}t \,+\, \sum_{i=1}^k f_i(\bar{Y}) \circ \mathrm{d}\bar{X}^i. \tag{1.13}$$

We give now a second application of Theorem 1.2, which cannot be covered, to our understanding, by classical LDP results. This is a large deviation principle associated with SDEs driven by k-layer paths $(X^{i_1,n}, \ldots X^{i_k,n})$: we take, for $i_1, \ldots i_k$ in $\{1, \ldots n\}$, the SDE

$$dY_t^{i_1,\ldots i_k;n} = \sum_{j=1}^{k} f_j(Y^{i_1,\ldots i_k;n}) \circ dX^{i_j,n} \qquad (1.14)$$

with same initial condition $Y^{i_1,\ldots i_k;n} = y_0$; here $f_j : \mathbb{R}^d \to \mathbb{R}^m$, $j = 1, \ldots k$, are given C_b^3 vector fields. For this SDE we can consider the empirical measure

$$L_n^{Y;\{k\}} = \frac{1}{n^k} \sum_{(i_1,\ldots i_k)\in\{1,\ldots n\}^k} \delta_{Y^{i_1,\ldots i_k;n}}.$$

This empirical measure can be seen as a symmetrization of the system (1.11), as it tracks the positions of $Y^{i_1,\ldots i_k}$ discarding the particular choice of indices $i_1, \ldots i_n$. Now, as in the previous application, rough paths provide continuity of the solution map of this SDE with respect to the driving noise in rough path topology. Therefore contraction principle implies the following:

Corollary 1.6. *For any fixed* $1/3 < \beta < 1/2$*, the sequence* $\{Law(L_n^{Y;\{k\}})|n \in \mathbb{N}\}$ *satisfies a large deviation principle on* $\mathcal{P}(C^{0,\beta}([0,T];\mathbb{R}^m))$ *, endowed with the* C_0*-weak topology.*

The paper is organized as follows. In Section 1, after a brief introduction, we explain our main results. Section 2 is devoted to settle notation and some recalls on rough paths. In Section 3, we prove the enhanced Sanov theorem (Theorem 1.1) in the 1-Wasserstein metric, leaving the extension to the modified Wasserstein topology to Section 4. For notational simplicity we focus in Sections 3 and 4 on the two-layer case ($k = 2$). We explain in Section 5 how to extend this to general k (and so conclude with a full proof of Theorem 1.1). In Section 6, we introduce the n-particle system, more precisely a system of n weakly interacting diffusions, and prove a large deviation principle for the empirical measure, that is Theorem 1.3. At last, in Sections 7 and 8, we prove resp. the (enhanced k-layer) propagation of chaos property and the LDP for the system (1.14).

1.2. Relation to the work of Cass–Lyons [4]

We comment in some detail on the relation of our work to Cass–Lyons. In [4], the authors first and foremost establish a theory of mean-field RDEs (more precisely, [4, Theorem 4.9], rough differential equations with mean-field interaction in the drift term) for suitable classes of random rough paths $\mathbf{B}(\omega)$. When it comes to propagation of chaos (see [4, Section 5]) they are able to consider interacting particle dynamics of the form

$$d\mathbf{X}^{i,n}(\omega) = \frac{1}{n}\sum_{j=1}^{n} b(X_t^{i,n}(\omega), X_t^{j,n}(\omega))\, dt + \sigma(X_t^i(\omega))\, d\mathbf{B}^i(\omega), \qquad (1.15)$$

with i.i.d. initial data and driving noise, $(X_0^{i,n}, \mathbf{B}^i)$, and show [4, Theorem 5.2] that

$$\mathbf{L}_n^{\mathbf{X}} := \frac{1}{n}\sum_{j=1}^{n} \delta_{\mathbf{X}^{i,n}} \longrightarrow \mathrm{Law}(\bar{X}) \qquad \text{a.s.}$$

In the scale of k-layer enhanced empirical measure, $\mathbf{L}_n^{\mathbf{X}} \equiv \mathbf{L}_n^{\mathbf{X};\{k\}}|_{k=1}$. Furthermore, it is conjectured (see [4, page 25]) that their approach will be useful to establish Sanov-type theorem

à la Dawson–Gärtner for (1.15). Although related, our work is *not* a proof of this conjecture. That said, such a result will not imply our results. To be more specific, in our work no mean-field RDE theory is required, and in fact we have taken the noise to be additive Brownian noise, that is $dB^i(\omega)$ versus $\sigma(X_t^i(\omega)) \, dB^i(\omega)$. (We note that including non-interacting diffusion coefficients $dB^i \rightsquigarrow \sigma(X_t^i) \, dB^i$ would have been possible, as long as the Girsanov argument we use, cf. the proof of Theorem 6.1, remains feasible, which amounts to an ellipticity assumption on σ.) In the cases where our setting overlaps with [4], we indeed quantify the above with a large deviation principle, but then we also obtain (Theorem 1.2) a Sanov-type à la Dawson–Gärtner for the general k-layer enhanced empirical measure $\mathbf{L}_n^{\mathbf{X};\{k\}}$. This is in fact out of reach of [4] as can be trivially seen noting that $\mathbf{L}_n^{\mathbf{X};\{k\}}$ necessarily involves information of $(\mathbf{X}^{1,n}, \dots, \mathbf{X}^{n,n})$, and hence (take e.g. $b \equiv 0$) of $(\mathbf{B}^1, \dots, \mathbf{B}^n)$, as *joint* rough path, rather then a collection of n rough paths. But no such information is assumed in [4], making $\mathbf{L}_n^{\mathbf{X};\{k\}}$, $k \geq 2$, effectively an ill-defined object. In contrast, for us, by working directly with Brownian motion, we always have the Stratonovich lift at our disposal, so this is not an issue. For the same reason, our *robust* propagation of chaos (Theorem 1.4, and then e.g. Corollary 1.5) and the large deviation principle in Corollary 1.6 cannot possibly be obtained in the framework of Cass–Lyons. We finally note that forthcoming work of Bailleul–Catellier deals with Sanov-type theorem a la Dawson–Gärtner for (1.15), again in the spirit of Cass–Lyons.

2. Basic notation and results on rough paths

We introduce the space of rough paths and the space where our empirical measures live. Most of this section is taken from [12] or [13]. Before going into the theory, let us recall the basis of α-Hölder continuous functions. Given a Polish space (E, d) with a compatible structure of Lie group (it will be $\mathbb{R}^e$ or $G^2(\mathbb{R}^e)$) and given α in $(0, 1)$, we define the space $C^\alpha([0, T]; E)$ of the α-Hölder continuous paths from $[0, T]$ to E. This is a complete metric space, endowed with the distance

$$d_\alpha(\gamma^1, \gamma^2) = \sup_{t \in [0,T]} d(\gamma^1(t), \gamma^2(t)) + \sup_{s,t \in [0,T], s \neq t} \frac{d(\gamma^1(s)^{-1}\gamma^1(t), \gamma^2(s)^{-1}\gamma^2(t))}{|t - s|^\alpha}. \quad (2.1)$$

This space is not separable in general. However, the subspace $C^{0,\alpha}([0, T]; E)$ given by the closure, with respect of d_α, of the smooth (C^∞) paths is separable, hence Polish. Furthermore, for any $\beta > \alpha$, $C^\beta([0, T]; E)$ is included in $C^{0,\alpha}([0, T]; E)$ and the inclusion is compact.

When dealing with rough paths, we will always assume α in $(1/3, 1/2)$. An α-Hölder rough path on $\mathbb{R}^e$ is a triple $\mathbf{X} = (X_0, X, \mathbb{X})$, with X_0 point in $\mathbb{R}^e$, $X = (X_{s,t})_{s<t}$ two-index $\mathbb{R}^e$-valued map and $\mathbb{X} = (\mathbb{X}_{s,t})_{s<t}$ two-index $\mathbb{R}^{e \times e}$-valued map (we always suppose $0 \leq s, t \leq T$ when not specified), satisfying the following conditions (here $v \otimes w$ denotes the tensor product vw^T):

1. algebraic conditions (Chen's relation): for any $s < u < t$,

$$X_{s,t} = X_{s,u} + X_{u,t} \quad \text{and} \quad \mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}; \quad (2.2)$$

2. analytic conditions:

$$\sup_{0 \leq s < t \leq T} \frac{|X_{s,t}|}{|t - s|^\alpha} < \infty \quad \text{and} \quad \sup_{0 \leq s < t \leq T} \frac{|\mathbb{X}_{s,t}|}{|t - s|^{2\alpha}} < \infty. \quad (2.3)$$

Here X_0 represents the initial condition; it is not included in the standard definition (Definition 2.1 in [12], Chapter 2), but we need to keep track of it because we will work with paths starting from a generic probability measure (and not just from a single point). However, with some abuse

of notation, we will usually write $\mathbf{X} = (X, \mathbb{X})$, without X_0, when this is not relevant for our purposes, as for example when the initial point is fixed (this was the case for the main result 1.1).

The space of α-Hölder rough paths on $\mathbb{R}^e$ is denoted by $\mathcal{C}^\alpha([0, T]; \mathbb{R}^e)$. It is not a vector space (since the sum of two rough paths does not respect Chen's relation), but it is a complete metric space, endowed with the distance

$$
\begin{aligned}
\tilde{\rho}_\alpha(\mathbf{X}, \mathbf{Y}) &= |X_0 - Y_0| + \rho_\alpha(\mathbf{X}, \mathbf{Y}) \\
&= |X_0 - Y_0| + \sup_{0 \le s < t \le T} \frac{|X_{s,t} - Y_{s,t}|}{|t - s|^\alpha} + \sup_{0 \le s < t \le T} \frac{|\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}|}{|t - s|^{2\alpha}}.
\end{aligned}
$$

For convenience, we also introduce a "norm" on rough paths; this is actually not a norm, but it has some good homogeneity property. We define

$$
\|\mathbf{X}\|_\alpha = \sup_{0 \le s < t \le T} \frac{|X_{s,t}|}{|t - s|^\alpha} + \sup_{0 \le s < t \le T} \frac{|\mathbb{X}_{s,t}|^{1/2}}{|t - s|^\alpha}. \tag{2.4}
$$

A problem with the space $\mathcal{C}^\alpha([0, T]; \mathbb{R}^e)$ is that it is not separable. That is why we introduce also the space $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$ of geometric rough paths. This is the subspace of $\mathcal{C}^\alpha([0, T]; \mathbb{R}^e)$ obtained as the closure, with respect to the $\tilde{\rho}_\alpha$ distance, of the space of smooth $\mathbb{R}^e$-valued paths and their iterated integrals (see [12], Section 2.2). Now the space $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$, endowed with the distance $\tilde{\rho}_\alpha$ is a Polish space. This will be the space of interest for us.

The space of geometric rough paths has also the following geometrical interpretation (taken for example from [12, Section 2.3]): it can be identified with the space $C^{0,\alpha}([0, T]; G^2(\mathbb{R}^e))$ of the closure of smooth paths, with respect to the α-Hölder topology, over the (free step-2 nilpotent) Lie group $G^2(\mathbb{R}^e)$. In particular, we can consider the α-Hölder distance d_α associated with the (Carnot–Caratheodory) distance in $G^2(\mathbb{R}^e)$, as explained at the beginning of this section, and we have, for a constant $C > 0$,

$$
C^{-1}(|X_0| + \|\mathbf{X}\|_\alpha) \le d_\alpha(\mathbf{X}, 0) \le C(|X_0| + \|\mathbf{X}\|_\alpha) \tag{2.5}
$$

for every geometric rough path $\mathbf{X}$. We call this distance the homogeneous distance. Unless otherwise stated, we will always use the homogeneous distance for geometric rough paths. Notice however that, for the purpose of this paper, only the asymptotic behaviour of $d_\alpha(\mathbf{X}, 0)$, as $|X_0| + \|\mathbf{X}\|_\alpha \to \infty$, is of interest for us (see Appendix A and Section 4.1 on the link between this behaviour and the Wasserstein topology), therefore one can use $|X_0| + \|\mathbf{X}\|_\alpha$ instead of $d_\alpha(\mathbf{X}, 0)$.

A consequence of this geometrical interpretation is that, for any $\alpha < \beta$, we have the continuous embedding for spaces of rough paths,

$$
\mathcal{C}_g^\beta([0, T]; \mathbb{R}^e) \hookrightarrow \mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e) \hookrightarrow \mathcal{C}_g^\alpha([0, T]; \mathbb{R}^e), \tag{2.6}
$$

where the first embedding is compact.

A basic result in Lyons' rough paths theory is that, given a function f regular enough, the integral $\int_0^t f(Y) \, d\mathbf{Y}$ is well defined and continuous with respect to Y in the rough paths topology. We have (e.g. Theorem 4.4 in [12], Chapter 4):

Theorem 2.1. *Let f be a function in $C_b^2(\mathbb{R}^e)$ and let $\mathbf{X}$ be a geometric α-Hölder rough path on $\mathbb{R}^e$. Given a partition Δ of the interval $[0, T]$, define the approximated integral on Δ as*

$$
I_\Delta f(\mathbf{X}) = \sum_{[s,t] \in \Delta} f(X_s) X_{s,t} + Df(X_s) \mathbb{X}_{s,t}. \tag{2.7}
$$

Then, the limit

$$\int_0^T f(X) \, d\mathbf{X} := \lim_{n \to \infty} I_{\Delta_n} f(X) \tag{2.8}$$

exists for every sequence $(\Delta_n : n \in \mathbb{N})$ *with infinitesimal size* $|\Delta_n| = \sup_{[s,t] \in \Delta_n} (t - s)$ *and is independent of the sequence itself. Furthermore, the application*

$$\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e) \ni \mathbf{X} \mapsto \int_0^T f(X) \, d\mathbf{X} \in \mathbb{R} \tag{2.9}$$

is continuous and it holds, for some constant C_f *depending on* f,

$$\left| \int_0^T f(X) \, d\mathbf{X} \right| \le C_f \left(\|\mathbf{X}\|_\alpha \vee \|\mathbf{X}\|_\alpha^{1/\alpha} \right). \tag{2.10}$$

Recall that Theorem 1.1, through definition of $F^{\{k\}}$ given in (1.6), involves a (measurable) "rough path lifting map"

$$S \equiv S^{\{e\}} : C^{0,\alpha}([0, T], \mathbb{R}^e) \to \mathcal{C}_g^{0,\alpha}([0, T], \mathbb{R}^e). \tag{2.11}$$

Here is the precise definition. Consider piecewise linear approximation $\{X^k : k \in \mathbb{N}\}$ of X based on dyadic partitions $\Delta_k := \{[2^{-k}i, 2^{-k}(i+1)] : i \in \mathbb{N}\}, k \in \mathbb{N}$, and set

$$A_{s,t}^k \equiv \lim_{k \to \infty} \int_s^t X_{s,r}^k \otimes dX_r^k. \tag{2.12}$$

Whenever $S^k := (X^k, A^k)$ is Cauchy in α-Hölder rough path metric, set

$$S(X) \equiv (X, A(X)) := \lim_{k \to \infty} (X^k, A^k) \tag{2.13}$$

and zero elsewhere. By construction, $S(X)$ is in $\mathcal{C}^\alpha$ and actually in $\mathcal{C}_g^{0,\alpha}$ (since X^k is a Lipschitz path and so S^k is in $\mathcal{C}_g^{0,\alpha}$) and $X \mapsto S(X)$ is a well-defined measurable (but in general discontinuous!) map on path space.

We now recall the basic relations between rough and stochastic integration, see [12, Proposition 3.5, 3.6 and Corollary 5.2]. We allow B to start from a generic initial probability measure λ with finite second moment.

Proposition 2.2. *Let B be an e-dimensional standard Brownian motion over a filtered probability space* $(\Omega, (\mathcal{F}_t)_t, \mathbb{P})$, *with initial measure* λ *with finite second moment. For any* $i, j = 1, \ldots, e$, *let* $(\mathbb{B}_{s,t}^{\mathrm{Strat}})^{ij} = \int_0^t B_{s,r}^i \circ dB_r^j$ *be its Stratonovich iterated integral. Then,*

(i) $\mathbb{P}$-a.s., $\mathbf{B}^{\mathrm{Strat}} := (B, \mathbb{B}^{\mathrm{Strat}})$ *is a geometric α-Hölder rough path for any* $\alpha < 1/2$;

(ii) *there exists a null-set N with respect to the e-dimensional Wiener measure $P = P^{\{e\}}$ (and hence to every Q absolutely continuous with respect to $P^{\{e\}}$), such that, away from this null-set, S^k is a Cauchy sequence in the rough path metric and $S(B) = \mathbf{B} = (B, \mathbb{B})$ $\mathbb{P}$-a.s.*

Proposition 2.3. *Let B be as before and let f be a function in $C_b^2(\mathbb{R}^e)$. Then the Stratonovich integral*

$$\int_0^T f(B) \circ dB \tag{2.14}$$

and the rough integral

$$\int_0^T f(B)\, d\mathbf{B}^{\text{Strat}} \tag{2.15}$$

coincide $\mathbb{P}$-*a.s.*

3. The enhanced Sanov theorem

The main objective in this section is to prove an LDP for the enhanced empirical measures $\mathbf{L}_n^{\mathbf{B}} = \mathbf{L}_n^{\mathbf{B};\{k\}}$ in the 1-Wasserstein topology, in the double layer case ($k = 2$ will be fixed, and often omitted, throughout this section). For this purpose, consider a sequence of independent d-dimensional Brownian motions $\{B^i : i \in \mathbb{N}\}$ each starting with initial distribution λ, defined on some filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_t, \mathbb{P})$. In the sequel, for fixed $\alpha \in (1/3, 1/2)$, we use the convention to denote a generic measure on $C^{0,\alpha}([0, T]; \mathbb{R}^d)$ by Q, and we write $\mathbb{P}^Y$ to denote the law on this space of a process Y; $P^{\{d\}} = \mathbb{P}^B$ is the Wiener measure on $C^{0,\alpha}([0, T]; \mathbb{R}^d)$ with initial distribution λ unless differently specified.

The empirical measure L_n^B is defined as

$$L_n^B = \frac{1}{n} \sum_{i=1}^n \delta_{B_i}. \tag{3.1}$$

We use the 1-Wasserstein metric as the topology on the space of probability measures (with finite first moment) on the spaces C^α and $C_g^{0,\alpha}$. In this topology, all the maps of the form

$$\mu \mapsto \int \varphi\, d\mu, \tag{3.2}$$

for φ continuous with at most linear growth, are continuous; on the contrary, in the C_b-weak topology we could only allow for continuous bounded φ. The reason why we consider the 1-Wasserstein metric is mainly because it is more convenient in the proof: first it gives an easy-to-handle distance between probability measure, then it makes the map $C^{0,\alpha}(\mathbb{R}^e) \ni \mu \mapsto \int \mathbb{X}^{(m)} \mu(dX)$ (where $\mathbb{X}^{(m)}$ will be a suitable approximation of the stochastic integral $\int_0^t X_r \otimes \circ dX_r$) continuous for m fixed ($X \mapsto \mathbb{X}^{(m)}$ has linear growth with respect to d_α, so the C_b-weak topology would not fit into this scheme).

The section is organized as follows. We start with proving Sanov theorem in the 1-Wasserstein metric. Then, as an intermediate result, we prove an LDP for the double-layer empirical measures which is a consequence of Sanov theorem (in 1-Wasserstein metric) and the contraction principle. Finally, we show an LDP for the enhanced empirical measures, whose proof uses the idea for the double-layer empirical measures but exploits the extended contraction principle, together with approximation lemmata coming from rough paths theory.

Before going to the results, for completeness we recall the definition of LDP. Classical notions and results on large deviations can be found for example in [8] and in [9].

Definition 3.1. Let E be a regular Hausdorff topological space (endowed with its Borel σ-algebra) and let $I : E \to [0, +\infty]$ be a nonnegative lower semi-continuous function (i.e. such that $\{I \leq a\}$ is closed for every finite a). We say that a sequence $\{\mu_n : n \in \mathbb{N}\}$ of probability measures on E satisfies a large deviation principle (LDP) with scale n and rate function I if the following facts hold:

- $\liminf_n \frac{1}{n} \log \mu_n[G] \geq -\inf_G I$ for every open subset G of E;
- $\limsup_n \frac{1}{n} \log \mu_n[F] \leq -\inf_F I$ for every closed subset F of E.

We say that the rate function I is good if, for every $a \geq 0$, the set $\{I \leq a\}$ is compact.

3.1. Sanov theorem in 1-Wasserstein metric

We quickly review Sanov theorem in 1-Wasserstein metric on a general Polish space. A necessary and sufficient condition for Sanov theorem in p-Wasserstein metric was in fact given in [20], but as the argument is short we include it in a form convenient to us.

Given a Polish space (E, d_E), we denote by $\mathcal{P}_1(E)$ the space of probability measures on E with finite first moment, i.e. the probability measures μ satisfy $\int_E d_E(x, x_0)\,\mu(\mathrm{d}x) < +\infty$ for some (equivalently for all) $x_0 \in E$. It is a Polish space endowed with the 1-Wasserstein distance d_W, namely

$$d_W(\mu, \nu) \;=\; \inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int_{E \times E} d_E(x^1, x^2)\, \pi\left(\mathrm{d}(x^1, x^2)\right) \right\}, \tag{3.3}$$

where $\Gamma(\mu, \nu)$ is the set of all probability measures on $E \times E$ with the first marginal and the second marginal equal resp. to μ and ν. Whenever E is some (Polish) space of α-Hölder continuous (rough) paths, cf. beginning of Section 2, we shall write $d_{W,\alpha}$ for the corresponding 1-Wasserstein distance. Some basic facts on 1-Wasserstein metric will be specified later in the Appendix.

We also recall that the relative entropy between two probability measures μ and ν on F is defined as

$$H(\mu|\nu) \;=\; \begin{cases} \int_F \rho \log \rho\,\mathrm{d}\nu, & \text{if } \mu \ll \nu \text{ and } \dfrac{\mathrm{d}\mu}{\mathrm{d}\nu} = \rho, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.4}$$

Theorem 3.2 (*Sanov Theorem in Wasserstein Metric*). *Let E be a Polish space and let $(X^i)_i$ be a sequence of E-valued i.i.d. random variables, with law μ. Assume that μ satisfies the following condition: there exists a function $G : E \to [0, +\infty]$, with compact sublevel sets (in particular lower semi-continuous), with more than linear growth (i.e., for some x_0, $|G(x)|/d(x, x_0) \to +\infty$ as $d(x, x_0) \to +\infty$), such that*

$$\int_E e^G \,\mathrm{d}\mu \;<\; +\infty. \tag{3.5}$$

Then the sequence of laws of the empirical measures

$$L_n^X \;=\; \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \tag{3.6}$$

satisfies a large deviation principle on $\mathcal{P}_1(E)$, endowed with the 1-Wasserstein metric, with rate n and good rate function $H(\cdot|\mu)$.

This result differs from the classical Sanov theorem by the fact that it involves the 1-Wasserstein metric, while classical Sanov theorem involves C_b-weak topology. In this, the statement above is stronger, but does need the additional condition on the measure λ.

Remark 3.3. In the case $E = C^{0,\alpha}([0, T]; \mathbb{R}^d)$, $\alpha < 1/2$, the assumption above is satisfied by $\{B^i : i \in \mathbb{N}\}$ (independent Brownian motions starting from λ), if λ verifies Condition (3.15). Indeed one can take

$$G(\gamma) = c\left(\sup_{0 \le s < t \le T} \frac{d(\gamma(t), \gamma(s))}{|t - s|^\beta}\right)^{1+\varepsilon} + c|\gamma(0)|^{1+\varepsilon}, \tag{3.7}$$

where β is in $(\alpha, 1/2)$ and c, ε are the same of Condition (3.15). This G has compact sublevel sets and more than linear growth; Condition (3.5) is verified since ($B^1(x = 0)$ is the Brownian motion starting at 0)

$$\mathbb{E}\big[e^{G(B^1)}\big] = \mathbb{E}\big[\exp(c\|B^1(x = 0)\|_{C^\beta}^{1+\varepsilon})\big] \int_{\mathbb{R}^d} e^{c|x|^{1+\varepsilon}} \lambda(dx) < \infty, \tag{3.8}$$

by Condition (3.15) and exponential integrability of $c\|B^1\|_{C^\beta}^{1+\varepsilon}$ (a consequence for example of Corollary 13.15 in [13]).

Proof of Theorem 3.2. The assertion is a consequence of classical Sanov theorem (in the weak convergence topology, see for example [9, Theorem 3.2.17]) and the inverse contraction principle, see [8, Theorem 4.2.4], provided we prove exponential tightness, in 1-Wasserstein metric, of the laws of the empirical measures L_n^X. We need to prove that, for any $M > 0$, there exists a compact set $K = K_M$ in $\mathcal{P}_1(E)$ (with the 1-Wasserstein metric) such that

$$\limsup_n \frac{1}{n} \log \mu_n[K_M^c] < -M. \tag{3.9}$$

We take K_M as in Lemma A.3. By Markov inequality and i.i.d. hypothesis on X^i, for any C_M, we have

$$\mathbb{P}\big[L_n^X \in K_M^c\big] \le e^{-nC_M} \mathbb{E}\left[\exp\left(\int_E nG \, dL_n^X\right)\right] = e^{-nC_M} \mathbb{E}\big[\exp\big(G(X_1)\big)\big]^n. \tag{3.10}$$

The assumption implies that $A := \mathbb{E}[\exp(G(X_1))] < \infty$. Hence, by taking $C_M = M + \log A + 1$, we obtain (3.9) which completes the proof. $\square$

3.2. The LDP for the double-layer empirical measure

As a warm-up example, we investigate what happens with the double layer empirical measure

$$L_n^{B,\{2\}} = \frac{1}{n^2} \sum_{i,j=1}^n \delta_{(B^i, B^j)} \in \mathcal{P}_1(C^{0,\alpha}([0, T]; \mathbb{R}^{2d})), \tag{3.11}$$

where $\mathcal{P}_1(C^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$ denotes the space of probability measures on $C^{0,\alpha}([0, T]; \mathbb{R}^{2d})$ endowed with the 1-Wasserstein metric. In the following, we identify $C^{0,\alpha}([0, T]; \mathbb{R}^{2d})$ with $C^{0,\alpha}([0, T]; \mathbb{R}^d)^2$ (they are equivalent as metric spaces) and we call π_1 the canonical projection in $C^{0,\alpha}([0, T]; \mathbb{R}^d)^2$ on the first d components.

Lemma 3.4. *The double layer empirical measure $L_n^{B,\{2\}}$ is the image of the empirical measure L_n^B under the map $Q \mapsto Q \otimes Q$.*

Proof. Obvious via the identification $\delta_{(B^i, B^j)} = \delta_{B^i} \otimes \delta_{B^j}$. $\square$

Proposition 3.5. *The family* $\{\text{Law}(L_n^{B,\{2\}}) : n \in \mathbb{N}\}$ *satisfies a LDP on* $\mathcal{P}_1(C^{0,\alpha}([0,T]; \mathbb{R}^{2d}))$ *endowed with the 1-Wasserstein metric, with scale n and good rate function $I^{\{2\}}$, given by*

$$I(Q^{\{2\}}) = \begin{cases} H(Q \mid P^{\{d\}}), & \text{if } Q^{\{2\}} = (Q \otimes Q) \text{ with } Q = Q^{\{2\}} \circ \pi_1^{-1} \\ \infty, & \text{otherwise.} \end{cases} \tag{3.12}$$

Proof. The result is a consequence of Sanov theorem in the 1-Wasserstein metric 3.2 (together with Remark 3.3 for our context) and the contraction principle, cf. [8, Theorem 4.2.1], provided that the map

$$\mathcal{P}_1(C^{0,\alpha}([0,T]; \mathbb{R}^d)) \ni Q \longmapsto Q \otimes Q \in \mathcal{P}_1(C^{0,\alpha}([0,T]; \mathbb{R}^{2d})) \tag{3.13}$$

is continuous. This continuity result is provided in Lemma A.4 in the Appendix. □

3.3. The LDP for the enhanced empirical measure

We are ready to prove the large deviation result for sequence of the enhanced empirical measure $\{\mathbf{L}_n^{\mathbf{B}} : n \in \mathbb{N}\}$.

Probability measures on $\mathcal{C}_g^{0,\alpha}([0,T]; \mathbb{R}^{2d})$ are denoted by Greek letters μ or ν. Further, we write $d_{W,\alpha}$ to denote the 1-Wasserstein distance on $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0,T]; \mathbb{R}^{2d}))$. We call B^{ij} the path (B^i, B^j) and $\mathbf{B}^{ij} = (B^{ij}, \mathbb{B}^{ij})$ the corresponding rough paths. We define it as $\mathbf{B} = S(B)$ (this ensures we can apply the extended contraction principle on the whole space), but, as far as the law is concerned, it is equivalent to define $\mathbf{B}$ via Stratonovich integral (see the section on rough paths). The enhanced empirical measure $\mathbf{L}_n^B$ is defined as

$$\mathbf{L}_n^{\mathbf{B}} = \frac{1}{n^2} \sum_{i,j=1}^{n} \delta_{(B^{ij}, \mathbb{B}^{ij})}.$$

Recall the definition of S given in (2.13) and of $F : \mathcal{P}_1(C^{0,\alpha}([0,T]; \mathbb{R}^d)) \to \mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0,T]; \mathbb{R}^{2d}))$ (formula (1.6) in the case $k = 2$) as the map

$$F : Q \mapsto (Q \otimes Q) \circ S^{-1}. \tag{3.14}$$

Recall also the definition of the projection π_1 as $\pi_1(\mathbf{X}) = X^1$ for any element $\mathbf{X} = ((X^1, X^2), \mathbb{X})$ in $\mathcal{C}_g^{0,\alpha}([0,T]; \mathbb{R}^{2d})$.

Theorem 3.6. *Let $\{B^i : i \in \mathbb{N}\}$ be a family of independent d-dimensional Brownian motion, with initial measure λ and assume that there exists $c, \varepsilon > 0$ such that*

$$\int_{\mathbb{R}^d} e^{c|x|^{1+\varepsilon}} \lambda(dx) < \infty. \tag{3.15}$$

The family $\{\text{Law}(\mathbf{L}_n^{\mathbf{B}}) : n \in \mathbb{N}\}$ *satisfies a LDP on* $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0,T]; \mathbb{R}^{2d}))$ *endowed with the 1-Wasserstein metric, with scale n and good rate function $\mathbf{I}$ given by*

$$\mathbf{I}(\mu) = \begin{cases} H(\mu \circ \pi_1^{-1} \mid P^{\{d\}}), & \text{if } \mu = F(\mu \circ \pi_1^{-1}), \\ \infty, & \text{otherwise.} \end{cases} \tag{3.16}$$

The basic fact, that invites us to use the extended contraction principle, is the following lemma.

Lemma 3.7. *The enhanced empirical measure $\mathbf{L}_n^{\mathbf{B}}$ is a.s. the image of the (true) empirical measure L_n^B under the map $F : Q \mapsto (Q \otimes Q) \circ S^{-1}$.*

Proof. The image measure of L_n^B under F is given by

$$\frac{1}{n^2} \sum_{i,j=1}^{n} \delta_{S(B^{ij})}.$$ (3.17)

By [Proposition 2.2](#), the Stratonovich rough paths $\mathbf{B}^{ij}$ coincide a.s. with $S(B^{ij})$, hence the image measure of L_n^B under F coincides a.s. with $\mathbf{L}_n^{\mathbf{B}}$. $\quad\square$

In order to apply the extended contraction principle, we introduce a continuous approximation F_m to the map F, defined in this way. Given a continuous trajectory Y, we define its piecewise linear approximation $Y^{(m)}$ as

$$Y^{(m)}(t) \;=\; Y\left(\frac{[mt]}{m}\right) + m\left(Y\left(\frac{[mt]+1}{m}\right) - Y\left(\frac{[mt]}{m}\right)\right)\left(t - \frac{[mt]}{m}\right).$$

The iterated integral of $Y^{(m)}$ is classically defined as Riemann integral, precisely

$$\left(\mathbb{Y}_t^{(m)}\right)^{ij} \;=\; \int_0^t Y_s^{(m),i}\, \mathrm{d}Y_s^{(m),j}.$$

Now we set F_m as

$$F_m : \mathcal{P}_1\left(C^{0,\alpha}([0,T];\mathbb{R}^d)\right) \;\longrightarrow\; \mathcal{P}_1\left(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})\right),$$
$$Q \;\longmapsto\; \left(Q \otimes Q\right) \circ (S^{(m)})^{-1}$$

where

$$C^{0,\alpha}([0,T];\mathbb{R}^{2d}) \ni Y \;\longmapsto\; S^{(m)}(Y) := \left(Y^{(m)}, \mathbb{Y}^{(m)}\right)$$
$$:= S(Y^{(m)}) \;\in\; \mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d}).$$

Note that this $S^{(m)}$ is defined as S^k, but replacing the dyadic approximation with the approximation at step $1/m$. We denote by $\mathbf{L}_n^{\mathbf{B}^{(m)}}$ the enhanced empirical measure associated with $B^{(m)}$, namely $\mathbf{L}_n^{\mathbf{B}^{(m)}} = F_m(L_n^B)$. Notice that, for each m, $S^{(m)}$ is continuous with at most linear growth (this is due to the use of the homogeneous distance d_α) and the map $Q \mapsto Q \otimes Q$ is continuous with respect to the 1-Wasserstein metrics on $\mathcal{P}_1(C^{0,\alpha}([0,T],\mathbb{R}^d))$ and $\mathcal{P}_1(C^{0,\alpha}([0,T],\mathbb{R}^{2d}))$ ([Lemma A.4](#) in the [Appendix](#)). So F_m is continuous in the 1-Wasserstein metric (by [Corollary A.2](#) in the [Appendix](#)).

In the proceeding lemmata, we show that the approximation given by F_m is indeed exponentially good, in the sense of the extended contraction principle (as in [9, Lemma 2.1.4]). The main tool is the following lemma, which follows from [13] (see Corollary 13.21 and Exercise 13.22, a proof is given in the [Appendix](#)), which gives an exponential bound for the approximation.

Lemma 3.8. *Let $\mathbf{B}$ the Stratonovich enhanced Brownian motion on $\mathbb{R}^e$, let $\mathbf{B}^{(m)}$ be its piecewise linear approximation, defined as before. Fix $\alpha < 1/2$. Then, for every η in $(0, 1/2 - \alpha)$, there exists $c > 0$ such that*

$$\sup_{m \geq 1} \mathbb{E}\left[\exp\left(cm^{\eta/2}\, d_\alpha\left(\mathbf{B}, \mathbf{B}^{(m)}\right)\right)\right] \;<\; \infty.$$ (3.18)

As a first step, we establish the exponential tightness of the approximation $\mathbf{L}_n^{\mathbf{B}^{(m)}}$ of $\mathbf{L}_m^{\mathbf{B}}$.

Lemma 3.9. *For any $\delta > 0$, it holds*

$$\lim_{m \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[d_{W,\alpha}\left(\mathbf{L}_n^{\mathbf{B}}, \mathbf{L}_n^{\mathbf{B}^{(m)}}\right) > \delta\right] \;=\; -\infty.$$ (3.19)

Proof. Consider the coupling measure $\frac{1}{n^2}\sum_{i,j=1}^{n}\delta_{(\mathbb{B}^{ij},\mathbb{B}^{(m),ij})}$ with marginals $\mathbf{L}_n^{\mathbf{B}^{(m)}}$ and $\mathbf{L}_n^{\mathbf{B}}$. Then, in view of (3.3), we obtain that

$$d_{W,\alpha}\big(\mathbf{L}_n^{\mathbf{B}}, \mathbf{L}_n^{\mathbf{B}^{(m)}}\big) \leq \frac{1}{n^2}\sum_{i,j=1}^{n} d_\alpha(\mathbf{B}^{ij}, \mathbf{B}^{(m),ij}), \tag{3.20}$$

where we used the fact that the map $(\mathbf{X}, \mathbf{X}') \mapsto d_\alpha(\mathbf{X}, \mathbf{X}')$ is Lipschitz continuous. By means of Hoeffding's decomposition [14], the right-hand side of (3.20) can be rewritten as

$$\frac{1}{n(n-1)}\sum_{\substack{i,j=1 \\ i\neq j}}^{n} H_m(i,j) = \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\frac{1}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor} H_m\big(\sigma(2i-1), \sigma(2i)\big),$$

where $\mathcal{S}_n$ denotes the set of all permutations of $\{1, \ldots, n\}$ and

$$H_{m,n}(i,j) \equiv H_m(i,j) := \frac{n-1}{n} d_\alpha(\mathbf{B}^{ij}, \mathbf{B}^{(m),ij}) + \frac{1}{n} d_\alpha(\mathbf{B}^{ii}, \mathbf{B}^{(m),ii}).$$

Hence, an application of the Markov inequality and Jensen's inequality gives, for any $C > 0$ and any n and m,

$$\mathbb{P}\Big[d_{W,\alpha}\big(\mathbf{L}_n^{\mathbf{B}}, \mathbf{L}_n^{\mathbf{B}^{(m)}}\big) > \delta\Big] \leq \mathbb{P}\bigg[\frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\frac{1}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor} H_m\big(\sigma(2i-1), \sigma(2i)\big) > \delta\bigg]$$

$$\leq e^{-C\delta}\,\mathbb{E}\bigg[\exp\bigg(\frac{C}{n!}\sum_{\sigma\in\mathcal{S}_n}\frac{1}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor} H_m\big(\sigma(2i-1), \sigma(2i)\big)\bigg)\bigg]$$

$$\leq e^{-C\delta}\,\frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\mathbb{E}\bigg[\exp\bigg(\frac{C}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor} H_m\big(\sigma(2i-1), \sigma(2i)\big)\bigg)\bigg].$$

Here, we see the advantage of Hoeffding's decomposition: by using the mutual independence of $\{H\big(\sigma(2i-1), \sigma(2i)\big) : i = 1, \ldots, \lfloor n/2\rfloor\}$ we finally get that

$$\mathbb{P}\Big[d_{W,\alpha}\big(\mathbf{L}_n^{\mathbf{B}}, \mathbf{L}_n^{\mathbf{B}^{(m)}}\big) > \delta\Big] \leq e^{-C\delta}\,\mathbb{E}\bigg[\exp\bigg(\frac{C}{\lfloor n/2\rfloor} H_m(1,2)\bigg)\bigg]^{\lfloor n/2\rfloor}. \tag{3.21}$$

On the other hand, by choosing $C = cm^\eta n/(6(c'\vee 1))$ for some $c' < \infty$ such that $d_\alpha(\mathbf{B}^{11}, \mathbf{B}^{(m),11}) \leq c'\, d_\alpha(\mathbf{B}^1, \mathbf{B}^{(m),1})$, Lemma 3.8 implies that, for any $\eta \in (0, 1/2 - \alpha)$ and any $n \geq 2$,

$$\sup_{m\geq 1}\mathbb{E}\bigg[\exp\bigg(\frac{c}{6(c'\vee 1)}\frac{m^\eta n}{\lfloor n/2\rfloor} H_m(1,2)\bigg)\bigg]$$

$$\leq \sup_{m\geq 1}\mathbb{E}\Big[\exp\big(cm^\eta d_\alpha(\mathbf{B}^{12}, \mathbf{B}^{(m),12})\big)\Big]^{1/2}\mathbb{E}\Big[\exp\big(cm^\eta d_\alpha(\mathbf{B}^1, \mathbf{B}^{(m),1})\big)\Big]^{1/2} < \infty.$$

By combining this estimate with (3.21), the assertion follows. $\quad\square$

Lemma 3.10. *For every $a < \infty$, it holds*

$$\lim_{m\to\infty}\sup_{Q:H(Q|P^{\{d\}})\leq a} d_{W,\alpha}\big(F_m(Q), F(Q)\big) = 0. \tag{3.22}$$

Proof. Using the coupling $(Q \otimes Q) \circ (S^{(m)}(X), S(X))^{-1}$, we get

$$
\begin{aligned}
d_{W,\alpha}\big(F_m(Q), F(Q)\big) &\leq \int_{C^{0,\alpha}([0,T],\mathbb{R}^{2d})} d_\alpha\big(S^{(m)}(X), S(X)\big)\, Q \otimes Q(\mathrm{d}X) \\
&= \int_{C^{0,\alpha}([0,T],\mathbb{R}^{2d})} d_\alpha\big(S^{(m)}(X), S(X)\big) \\
&\quad \times \left(\frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}} \otimes \frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}} \right)(X)\, \big(P^{\{d\}} \otimes P^{\{d\}}\big)(\mathrm{d}X).
\end{aligned}
$$

The idea is the following: For any Q with bounded entropy, $\frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}} \otimes \frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}}$ has a uniform $L \log L$ bound with respect to the Wiener measure $P^{\{d\}}$. Hence, the lemma is proven if the norm of $d_\alpha(S^{(m)}(X), S(X))$ in the dual space of $L \log L$, again with respect to $P^{\{d\}}$, converges to 0. This convergence follows by an exponential control of $d_\alpha(S^{(m)}(X), f(X))$ under $P^{\{d\}}$, which is a consequence of Lemma 3.8.

To make this argument work, we use the theory of Orlicz space. Let $\Phi, \Psi : [0, \infty) \to [0, \infty)$ be a complementary Young pair of N-functions defined by

$$
\Phi(r) = \frac{1}{2} r^2\, 1_{r \leq 1} + \left(e^{r-1} - \frac{1}{2} \right) 1_{r>1}, \quad \text{and}
$$

$$
\Psi(r) = \frac{1}{2} r^2\, 1_{r \leq 1} + \left(r \log r + \frac{1}{2} \right) 1_{r>1}.
$$

Further, on a given measure space (Λ, Σ, μ), introduce for any $g, h : \Lambda \to [0, \infty)$ measurable

$$
\|g\|_{L_{\exp}} := \inf_{k>0} \left\{ \frac{1}{k} \left(1 + \int_\Lambda \Phi(kg)\, \mathrm{d}\mu \right) \right\} \quad \text{and}
$$

$$
\|h\|_{L \log L} := \inf_{k>0} \left\{ \frac{1}{k} \left(1 + \int_\Lambda \Psi(kh)\, \mathrm{d}\mu \right) \right\}.
$$

Then, the classical Orlicz–Birnbaum estimate, see [16, Section 3.3] implies that for any measurable, nonnegative functions g and h, it holds

$$
\int_\Lambda gh\, \mathrm{d}\mu \leq 4\, \|g\|_{L \log L}\, \|h\|_{L_{\exp}}. \tag{3.23}
$$

In particular, by using the explicit form of the Orlicz pair (Φ, Ψ) the following estimates hold for any measurable, nonnegative functions g, h and $k > 0$

$$
\int_\Lambda gh\, \mathrm{d}\mu \leq \frac{4}{k} \left(1 + \int_\Lambda \exp(kg)\, \mathrm{d}\mu \right) \left(2 + \int_\Lambda h \log h\,, \mathrm{d}\mu \right). \tag{3.24}
$$

By applying (3.24) with $\Lambda = C^{0,\alpha}(\mathbb{R}^{2d})$, $\mu = P^{\{d\}} \otimes P^{\{d\}}$, $g = d_\alpha(S^{(m)}(X), S(X))$, $h = \frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}} \otimes \frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}}$ we get

$$
d_{W,\alpha}\big(F_m(Q), F(Q)\big) \leq \frac{4}{k}(2 + 2a)\left(1 + \mathbb{E}\Big[\exp\big(k\, d_\alpha\big(\mathbf{B}^{12}, \mathbf{B}^{12,(m)}\big)\big) \Big] \right), \tag{3.25}
$$

where we used that $\int_\Lambda h \log h\, \mathrm{d}\mu = 2H(Q|P^{\{d\}}) \leq 2a$. Finally, by choosing $k = cm^{\eta/2}$, a further application of Lemma 3.8 yields

$$
d_{W,\alpha}\big(F_m(Q), F(Q)\big) \leq \frac{4}{cm^{\eta/2}}(2 + 2a)(1 + C),
$$

which completes the proof. $\square$

Proof of Theorem 3.6. By Sanov Theorem 3.2 and Remark 3.3, the extended contraction principle ([9], Lemma 2.1.4) together with Lemmas 3.9 and 3.10 shows that $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{B}^{(m)}}) : n \in \mathbb{N}\}$ satisfies an LDP with scale n and good rate function given by

$$\mu \mapsto \inf\Big\{H(Q \mid P)\Big|Q \in \mathcal{P}_1\big(C^{0,\alpha}([0, T], \mathbb{R}^d)\big) \ \text{ and } \ F(Q) = \mu\Big\}.$$

It is easy to see that this rate function coincides with the $\mathbf{I}$ defined in Theorem 3.6. $\square$

We close the section with the convergence (in probability) of the enhanced empirical measures, which follows from the LDP (as well known in large deviations theory).

Corollary 3.11. *The sequence of $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$-valued random variables $\{\mathbf{L}_n^{\mathbf{B}} : n \in \mathbb{N}\}$ converges in probability (and in law) to the constant random variable $\mathbf{P}^{\{2d\}}$, the enhancement of the 2d-Wiener measure, that is the law on $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d})$ of $(B^{12}, \mathbb{B}^{12})$.*

Proof. The result is a consequence of the LDP for the laws of $\mathbf{L}_n^{\mathbf{B}}$ and of the fact that the good rate function has a unique zero in $\mathbf{P}^{\{2d\}}$. $\square$

4. The modified Wasserstein space

As already mentioned in the Introduction, in view of our application (Corollary 6.2), we will have to deal with maps of the form

$$\mu \ \longmapsto \ \int_{\mathcal{C}_g^{0,\alpha}} \int_0^T f(X) \, \mathrm{d}\mathbf{X} \, \mu(\mathrm{d}\mathbf{X}) \tag{4.1}$$

and we would like these maps to be continuous (to apply standard tools of large deviations theory). On one side, we know that a map $\mu \mapsto \int G \, \mathrm{d}\mu$ is continuous in the 1-Wasserstein metric if G is continuous with at most linear growth. But on the other side, by Theorem 2.1, the rough path integral has a growth of order at most $1/\alpha$, in particular a more than linear growth (with respect to the homogeneous rough paths norm).[7] This creates a problem. Following [3], we introduce a new function N of $\mathbf{X}$ with good concentration properties (w.r.t. to Brownian rough paths) such that the rough integral has at most linear growth with respect to N. We then device a strengthened topology, on a restriction of the space $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e))$, which allows us to use as test functions also functions with linear growth with respect to such N.

In this new topology we prove the large deviation principle for the enhanced empirical measures, as a consequence of the LDP in the 1-Wasserstein metric, via inverse contraction principle. This amounts to verify exponential tightness in the new topology, which can be proved using again Hoeffding decomposition and also Gaussian estimates for Brownian rough paths.

Remark 4.1. One may ask why we do not take simply $N(\mathbf{X}) = \|X\|_\alpha^{1/\alpha}$, or allow for p-Wasserstein distance, for $p = 1/\alpha$. The reason is that, with this choice of N, we are not able to prove a Sanov-type theorem for the enhanced empirical measure. Actually, in [20], it is proved that a large deviation result in the p-Wasserstein distance does not hold for any $p > 2$ (and actually also for $p = 2$), as a consequence of the lack of exponential integrability of $\|X\|_\alpha^p$.

---

[7] The path-by-path estimate in Theorem 2.1 is optimal.

4.1. A modified Wasserstein topology

For the definition of N, consider the following sequence of stopping times: given $\mathbf{X}$ in $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$, we define

$$\tau_0^\alpha(\mathbf{X}) \ = \ 0, \qquad \tau_{i+1}^\alpha(\mathbf{X}) \ = \ \inf\big\{t > \tau_i^\alpha(\mathbf{X}) \ : \ \|\mathbf{X}\|_{(1/\alpha)-\mathrm{var},[\tau_i^\alpha(\mathbf{X}),t]} \geq 1\big\}, \quad i \in \mathbb{N}. \quad (4.2)$$

Here $\|X\|_{(1/\alpha)-\mathrm{var},[s,t]}$ is the $(1/\alpha)$-variation of $\mathbf{X}$, as group-valued path, in the interval $[s, t]$, see [13] for precise definition. What we need here is that the norm $\|X\|_{(1/\alpha)-\mathrm{var},[s,t]}$ is a continuous function of $\mathbf{X}$, in the space $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$, for fixed s, t, and it is independent of the initial datum X_0. Notice, that it is also a continuous function of s, t, for fixed $\mathbf{X}$, and it is monotone in s and t, in the sense that $\|X\|_{(1/\alpha)-\mathrm{var},[s',t']} \leq \|X\|_{(1/\alpha)-\mathrm{var},[s,t]}$ for any $s \leq s' < t' \leq t$. We define $N = N_\alpha$ as

$$N_\alpha(\mathbf{X}) \ := \ \sup\big\{i \in \mathbb{N} \ : \ \tau_i^\alpha(\mathbf{X}) < T\big\}. \qquad (4.3)$$

We omit α when not necessary. The following lower-semicontinuity property of N will be proved in the Appendix.

Lemma 4.2. *The function N is lower semi-continuous on $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$.*

The next lemma gives the desired sublinear growth of the rough integral in terms of N, see [12] for a proof.

Lemma 4.3. *Let f be a function in $C_b^2(\mathbb{R}^e)$ and let $\mathbf{X}$ be in $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$. Then it holds, for some constant C_f depending on f,*

$$\left| \int_0^T f(X)\, d\mathbf{X} \right| \ \leq \ C_f \left(1 + N(\mathbf{X})\right). \qquad (4.4)$$

Now we introduce a modified topology, on a restriction of $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e))$, in order to deal with functionals of the form $\mu \mapsto \int G\, d\mu$ for some continuous G with $G(\mathbf{X}) \leq C(1 + N(\mathbf{X}))$.

First, for given $\varepsilon > 0$, we introduce the space

$$\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}\big(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)\big)$$

$$:= \ \left\{\mu \in \mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)) : \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon}\, \mu(d\mathbf{X}) < \infty\right\} \quad (4.5)$$

of probability measures with finite $(|X_0| + \|\cdot\|_\alpha + N_\alpha)^{1+\varepsilon}$.

Definition 4.4. Let μ_n, $n \in \mathbb{N}$, μ be in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}\big(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)\big)$. We say that $\{\mu_n : n \in \mathbb{N}\}$ converges to μ in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology if the following two conditions hold:

1. $\{\mu_n : n \in \mathbb{N}\}$ converges to μ in the weak topology, i.e. wrt. any test function in $C_b(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e))$;
2. we have

$$\sup_{n \in \mathbb{N}} \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon}\, \mu_n(d\mathbf{X}) \ < \ \infty. \qquad (4.6)$$

We say that a subset C of $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e))$ is closed in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology if it is closed under convergence of sequences.

Due to (2.5), $|X_0| + \|\mathbf{X}\|_\alpha$ is equivalent to $d_\alpha(\mathbf{X}, 0)$. Thus, every sequence converging in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology converges also in the 1-Wasserstein metric (remind the characterization of 1-Wasserstein distance in Lemma A.1). Hence the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology is stronger than the 1-Wasserstein topology. Note that the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology might not be metrizable. However, the following result, proved in the Appendix, gives the properties needed for large deviations analysis.

Lemma 4.5. *The space* $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e))$ *, with the* $(\| \cdot \| + N)^{1+\varepsilon}$*-Wasserstein topology, is a regular Hausdorff space.*

Furthermore, since N is lower semi-continuous, by Corollary B.2 the functional

$$\mu \longmapsto \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon} \, \mu(\mathrm{d}\mathbf{X})$$

is sequentially lower semi-continuous with respect to the 1-Wasserstein topology: if $\{\mu_n : n \in \mathbb{N}\}$ converges to μ in the 1-Wasserstein or in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology, then

$$\int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon} \, \mu(\mathrm{d}\mathbf{X})$$

$$\leq \liminf_n \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon} \, \mu_n(\mathrm{d}\mathbf{X}). \tag{4.7}$$

Proposition 4.6. *Assume that* $\{\mu_n : n \in \mathbb{N}\}$ *converges to* μ *in the* $(\| \cdot \| + N)^{1+\varepsilon}$*-Wasserstein topology. Let* G *be a continuous function on* $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$ *such that* $G(\mathbf{X}) \leq C(1 + |X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))$*, as for example the rough integral. Then,*

$$\lim_{n \to \infty} \int_{\mathcal{C}_g^{0,\alpha}} G(\mathbf{X}) \, \mu_n(\mathrm{d}\mathbf{X}) = \int_{\mathcal{C}_g^{0,\alpha}} G(\mathbf{X}) \, \mu(\mathrm{d}\mathbf{X}). \tag{4.8}$$

Proof. For any m positive integer, let G_m be the continuous bounded function defined from G with truncation at level m, that is $G_m = G 1_{|G| \leq m} + m 1_{G > m} - m 1_{G < -m}$. We have for every m, n,

$$\left| \int_{\mathcal{C}_g^{0,\alpha}} G \, \mathrm{d}(\mu_n - \mu) \right| \leq \left| \int_{\mathcal{C}_g^{0,\alpha}} G_m \, \mathrm{d}(\mu_n - \mu) \right| + \int_{\mathcal{C}_g^{0,\alpha}} |G_m - G| \, \mathrm{d}(\mu_n + \mu)$$

Notice that the condition $G(\mathbf{X}) \leq C(1 + |X_0| + \|\mathbf{X}\| + N(\mathbf{X}))$ implies (for m with $m/C > 4$) that

$$|G(\mathbf{X}) - G_m(\mathbf{X})| \leq 2C(|X_0| + \|\mathbf{X}\| + N(\mathbf{X})) 1_{\|\mathbf{X}\| + N(\mathbf{X}) > m/(2C)}.$$

So it holds for any n

$$\int_{\mathcal{C}_g^{0,\alpha}} |G_m - G| \, \mathrm{d}\mu_n \leq 2C \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\| + N(\mathbf{X})) 1_{\|\mathbf{X}\| + N(\mathbf{X}) > m/(2C)} \, \mathrm{d}\mu_n$$

$$\leq m^{-\varepsilon}(2C)^{1+\varepsilon} \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\| + N(\mathbf{X}))^{1+\varepsilon} \, \mathrm{d}\mu_n$$

$$\leq m^{-\varepsilon}(2C)^{1+\varepsilon} D,$$

where $D := \sup_n \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\| + N(\mathbf{X}))^{1+\varepsilon} \, \mu_n(\mathrm{d}\mathbf{X})$ is bounded by assumption. The same estimates hold also for μ in place of μ_n, by the lower semi-continuity property (4.7). Hence, for

any $\rho > 0$, we can find m_ρ such that

$$\int_{\mathcal{C}_g^{0,\alpha}} \left| G_{m_\rho} - G \right| \mathrm{d}(\mu_n + \mu) \; < \; \rho.$$

Fix such m_ρ. Since G_{m_ρ} is continuous bounded, there exists $n_\rho < \infty$ such that, for every $n \geq n_\rho$,

$$\left| \int_{\mathcal{C}_g^{0,\alpha}} G_{m_\rho} \, \mathrm{d}(\mu_n - \mu) \right| \; < \; \rho.$$

So we conclude that, for every $n \geq n_\rho$,

$$\left| \int_{\mathcal{C}_g^{0,\alpha}} G \, \mathrm{d}(\mu_n - \mu) \right| \; < \; 2\rho.$$

The proof is complete. $\quad\square$

We conclude this subsection with a lemma on compact sets on this space. This will be useful in view of exponential tightness on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^e))$. Recall that, given a topology τ (i.e. the set of all open sets), its restriction τ_A to a set A is given by $\{B \cap A : B \in \tau\}$.

Lemma 4.7.

1. *For any $R > 0$, the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology restricted on the set*

$$\bar{B}(R) \; = \; \left\{ \mu \in \mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}\left(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^e)\right) : \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \| \cdot \| + N)^{1+\varepsilon} \, \mathrm{d}\mu \leq R \right\} \tag{4.9}$$

coincides with the 1-Wasserstein topology restricted there.

2. *Let H be a subset of $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^e))$, which is compact in the 1-Wasserstein metric and is "bounded" in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^e))$, in the sense that*

$$\sup_H \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\| + N(\mathbf{X}))^{1+\varepsilon} \, \mu(\mathrm{d}\mathbf{X}) \; < \; \infty. \tag{4.10}$$

Then, H is compact in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^e))$ (with its $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology).

Proof. For the first part, every closed set in $\bar{B}(R)$ with respect to the (restricted) 1-Wasserstein topology is also closed with respect to the (restricted) $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology, this being stronger. Conversely, let C be a closed subset of $\bar{B}(R)$ in the restricted $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology; notice that C is closed also in the (not restricted) $(\|\cdot\|+N)^{1+\varepsilon}$-Wasserstein topology, since $\bar{B}(R)$ is closed in this topology (by the lower semicontinuity property (4.7)). Let $(\mu_n)_n$ be a sequence in C, converging to μ in the 1-Wasserstein metric. Since C is in $\bar{B}(R)$, the uniform bound

$$\sup_{n\in\mathbb{N}} \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\| + N(\mathbf{X}))^{1+\varepsilon} \, \mu_n(\mathrm{d}\mathbf{X}) \; < \; \infty \tag{4.11}$$

holds, hence μ_n converges to μ also in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology. Furthermore μ is also in $\bar{B}(R)$, by the lower semicontinuity property (4.7). Since C is closed in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology, μ must be in C and so C is closed also in the 1-Wasserstein topology. The first statement is proved.

The second part follows from the first one (as a general fact in topology), we give a proof for completeness. Let $(A_i)_{i \in I}$ be a family of open sets, in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology, whose union contains H and take $R > 0$ such that H is contained in $\bar{B}(R)$. Consider $\tilde{A}_i := A_i \cap \bar{B}(R)$, which are open sets in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology restricted on $\bar{B}(R)$. By the first statement, they are open also in the restricted 1-Wasserstein topology on $\bar{B}(R)$. That is, there exist B_i (subsets of $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha})$), open sets in the 1-Wasserstein topology, such that $\tilde{A}_i = B_i \cap \bar{B}(R)$. Actually, since $\bar{B}(R)$ is closed in every topology under consideration, one can choose $B_i = A_i \cup \mathcal{P}_1(\mathcal{C}_g^{0,\alpha}) \setminus \bar{B}(R)$. In particular $(B_i)_{i \in I}$ is a family of open sets, in the 1-Wasserstein metric, covering H. By the compactness of H in the 1-Wasserstein metric, we can extract a finite subset $\{i_1, \ldots, i_m\}$ of I such that $\bigcup_{1 \le k \le m} B_{i_k}$ contains H. Since $\tilde{A}_i = B_i \cap \bar{B}(R)$ and H is in $\bar{B}(R)$, also $\bigcup_{1 \le k \le m} A_{i_k}$ contains H. The proof is complete. $\square$

4.2. The LDP in the modified Wasserstein space

In this section we prove the LDP for the enhanced empirical space, in the stronger $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology, again for the double layer case ($k = 2$ will be fixed and often omitted in the notation). Recall the definition of F and S in (3.14), (2.13). Recall that the Brownian motions B^i (and their corresponding rough paths) start from measure λ satisfying (3.15). Here we assume ε to be the one appearing in condition (3.15). Mind that we need large deviation tools on a regular Hausdorff spaces (as in [8]) and not just on metric spaces.

Theorem 4.8.   *The sequence* $\mathrm{Law}(\mathbf{L}_n^{\mathbf{B}})_n$ *satisfies a LDP on* $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$ *(endowed with the* $(\| \cdot \| + N)^{1+\varepsilon}$-*Wasserstein topology) with scale n and good rate function*

$$\mathbf{I}(\mu) = \begin{cases} H(\mu \circ \pi_1^{-1} \mid P^{\{d\}}), & \text{if } \mu = F(\mu \circ \pi_1^{-1}), \\ \infty, & \text{otherwise.} \end{cases} \tag{4.12}$$

Recall that the strategy is to use the inverse contraction principle starting from the previous Theorem 3.6 and that, for this, we need the exponential tightness of the family $(\mathrm{Law}(\mathbf{L}_n^{\mathbf{B}}))_n$ in the $(\| \cdot \| + N)^{1+\varepsilon}$-Wasserstein topology.

The main tool is the following lemma, which follows for example from [12, Theorems 11.9 and 11.13], see also [3, Theorem 6.3] adapted to our case in the Appendix, and gives an exponential bound for $N(\mathbf{B})$.

Lemma 4.9.   *Let* $\mathbf{B}$ *be the Stratonovich enhanced Brownian motion on* $\mathbb{R}^e$, *with initial measure* $\tilde{\lambda}$ *satisfying condition* (3.15). *Then, for any* $\alpha < 1/2$, $\beta < 1/2$ *the random variables* $\|\mathbf{B}\|_\beta$ *and* $N_\alpha(\mathbf{B})$ *have Gaussian tails, in particular, for some $c > 0$,*

$$\mathbb{E}\left[\exp\big(c(|B_0| + \|\mathbf{B}\|_\beta + N_\alpha(\mathbf{B}))^{1+\varepsilon}\big)\right] < \infty. \tag{4.13}$$

The same result holds for $\mathbf{B}^{11}$ *on* $\mathbb{R}^{2d}$ *(where B^1 is a Brownian motion on* $\mathbb{R}^d$ *starting from λ and λ satisfies* (3.15)).

Here is the exponential tightness result:

Lemma 4.10.   *The sequence* $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{B}}) : n \in \mathbb{N}\}$ *is exponentially tight on* $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T];$ $\mathbb{R}^{2d}))$ *with respect to the* $(\| \cdot \| + N)^{1+\varepsilon}$-*Wasserstein topology.*

Proof. For any $M > 0$, we have to find a set K_M, compact in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology, such that

$$\limsup_n \frac{1}{n} \log \mathbb{P}\{\mathbf{L}_n^{\mathbf{B}} \in K_M^c\} < -M. \tag{4.14}$$

Our candidate for K_M is

$$K_M = \left\{ \mu \in \mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}} : \int_{\mathcal{C}_g^{0,\alpha}} G \, d\mu \leq M \right\},$$

where $G : \mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}) \to [0, +\infty]$ is defined by $G(\mathbf{X}) := (|X_0| + \|\mathbf{X}\|_\beta + N_\alpha(\mathbf{X}))^{1+\varepsilon}$ for some fixed β with $\alpha < \beta < 1/2$.

The compactness of K_M follows from Lemma 4.7, since K_M satisfies the hypotheses of that result. Indeed 1) K_M is "bounded" in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}$ (in the sense of (4.10)), since $\|\mathbf{X}\|_\alpha \leq C\|\mathbf{X}\|_\beta$ for some $C > 0$; 2) it is also compact in the 1-Wasserstein metric, as a consequence of Lemma A.3 (since G has more than linear growth, is lower semi-continuous and has pre-compact sublevel sets for the compact inclusion of $\mathcal{C}_g^\beta$ in $\mathcal{C}_g^{0,\alpha}$).

Now we verify (4.14). We use a strategy similar to that for Lemma 3.9. By Markov inequality, we have for any $C > 0$,

$$\mathbb{P}\big[\mathbf{L}_n^{\mathbf{B}} \in K_M^c\big] = \mathbb{P}\left[\int_{\mathcal{C}_g^{0,\alpha}} G \, d\mathbf{L}_n^{\mathbf{B}} > M \right]$$

$$\leq e^{-CM} \mathbb{E}\left[\exp\left(C \int_{\mathcal{C}_g^{0,\lambda}} G \, d\mathbf{L}_n^{\mathbf{B}} \right) \right] = e^{-CM} \mathbb{E}\left[\exp\left(\frac{C}{n^2} \sum_{i,j=1}^n G(\mathbf{B}^{ij}) \right) \right].$$

Exploiting Hoeffding decomposition as in the proof of Lemma 3.9, we get

$$\mathbb{P}\big[\mathbf{L}_n^{\mathbf{B}} \in K_M^c\big] \leq e^{-CM} \mathbb{E}\left[\exp\left(\frac{C}{[n/2]} H'(1, 2) \right) \right]^{[n/2]},$$

where now

$$H'(i, j) := \frac{n-1}{n} G(\mathbf{B}^{ij}) + \frac{1}{n} G(\mathbf{B}^{ii}).$$

By using that $H'(1, 2) \leq G(\mathbf{B}^{12}) + G(\mathbf{B}^{11})$ and applying Lemma 4.9 to $\mathbf{B}^{12}$ and to $\mathbf{B}^{11}$ (with initial measure $\lambda \otimes \lambda$), we get that $\mathbb{E}\big[\exp(cH'(1, 2))\big] =: D$ is finite for some constant $c > 0$. Hence, choosing $C = c[n/2]$, we get

$$\mathbb{P}\big[\mathbf{L}_n^{\mathbf{B}} \in K_M^c\big] \leq e^{-c[n/2]M} D^{[n/2]}.$$

From this (4.14) follows (up to choosing $K_{3M/c}$ instead of K_M). □

Proof of Theorem 4.8. The result follows applying the inverse contraction principle (in the version of [8, Theorem 4.2.4]), from the space $\mathcal{P}_1$ to the space $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}$ (with the identity map), having the LDP on the former space (Theorem 3.6) and the exponential tightness on the latter space. □

5. Extension to *k*-layer enhanced empirical measures

So far we have considered the double layer enhanced empirical measure. We now deal with the extension of the LDP to the *k*-layer enhanced empirical measure, namely

$$\mathbf{L}_n^{B,\{k\}} = \frac{1}{n^k} \sum_{i_1,\dots i_k=1}^{n} \delta_{S^{\{kd\}}(B^{i_1},\dots,B^{i_k})}. \tag{5.1}$$

Here is the extension of Theorem 3.6, that is Theorem 1.1 in the Introduction, for the 1-Wasserstein metric, extended to a general initial measure.

Theorem 5.1. *Let* $\{B^i : i \in \mathbb{N}\}$ *be a family of independent d-dimensional Brownian motion, with initial measure* λ *and assume* Condition (3.15) *for some* $c, \varepsilon > 0$. *The sequence* $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}}) : n \in \mathbb{N}\}$ *satisfies a LDP on* $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$ *endowed with the* 1-*Wasserstein metric, with scale n and good rate function* $\mathbf{I}^{\{k\}}$ *given by*

$$\mathbf{I}^{\{k\}}(\mu) = \begin{cases} H(\mu \circ \pi_1^{-1} \mid P^{\{d\}}), & \text{if } \mu = F^{\{k\}}(\mu \circ \pi_1^{-1}), \\ \infty, & \text{otherwise.} \end{cases} \tag{5.2}$$

The proof of this LDP goes like the proof in the double layer case ($k = 2$). We recall the main steps and the main changes.

First Lemma 3.7 is extended to the *k* layer case; the map $Q \mapsto Q^{\otimes k}$ is continuous by Lemma B.4 in the Appendix. Therefore the strategy is still to apply the extended contraction principle. For this we define the approximant $S^{(m);\{k\}}$ as for the double layer case but on $\mathbb{R}^{kd}$, the approximation $\mathbf{B}^{(m);\{k\}}$ of the enhancement of the *kd*-dimensional Brownian motion and the maps $F^{(m);\{kd\}}$ are defined correspondingly. More in general, the space $\mathbb{R}^{2d}$ must be replaced in all the arguments by $\mathbb{R}^{kd}$. Then we need to extend Lemmas 3.9 and 3.10 to the *k* layer case.

Lemma 5.2. *For any* $\delta > 0$, *it holds*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[d_{W,\alpha}\left(\mathbf{L}_n^{\mathbf{B};\{k\}}, \mathbf{L}_n^{\mathbf{B}^{(m)};\{k\}}\right) > \delta \right] = -\infty. \tag{5.3}$$

Proof. Consider the coupling measure $\frac{1}{n^k} \sum_{i_1,\dots,i_k=1}^{n} \delta_{(\mathbb{B}^{\{k\},i_1,\dots,i_k}, \mathbb{B}^{(m);\{k\},i_1,\dots,i_k})}$ with marginals $\mathbf{L}_n^{\mathbf{B}^{(m)};\{k\}}$ and $\mathbf{L}_n^{\mathbf{B};\{k\}}$. As in the double layer case, we obtain that

$$d_{W,\alpha}\left(\mathbf{L}_n^{\mathbf{B};\{k\}}, \mathbf{L}_n^{\mathbf{B}^{(m)};\{k\}}\right) \leq \frac{1}{n^k} \sum_{i_1,\dots i_k=1}^{n} d_\alpha(\mathbf{B}^{\{k\},i_1,\dots,i_k}, \mathbf{B}^{(m);\{k\},i_1,\dots,i_k})$$

$$\leq \frac{(n-k)!}{n!} \sum_{\substack{i_1,\dots,i_k=1 \\ \text{mutually distinct}}}^{n} H_{(m);\{k\}}(i_1,\dots,i_k), \tag{5.4}$$

where the second sum is obtained from the first one rearranging the terms with at least two equal indices. For instance, in the case $k = 3$, $H_{(m);\{3\}}$ spells out as

$$H_{(m);\{3\}}(i_1, i_2, i_3) := \frac{(n-1)(n-2)}{n^2} d_\alpha(\mathbf{B}^{\{3\},i_1,i_2,i_3}, \mathbf{B}^{(m);\{3\},i_1,i_2,i_3})$$

$$+ \frac{n-1}{n} \left(d_\alpha(\mathbf{B}^{\{3\},i_1,i_2,i_1}, \mathbf{B}^{(m);\{k\},\{3\},i_1,i_2,i_1}) \right.$$

$$\left. + d_\alpha(\mathbf{B}^{\{3\},i_1,i_1,i_3}, \mathbf{B}^{(m);\{3\},i_1,i_1,i_3}) \right.$$

$$+ d_\alpha(\mathbf{B}^{\{3\},i_1,i_2,i_2}, \mathbf{B}^{(m);\{k\},\{3\},i_1,i_2,i_2}))$$

$$+ \frac{1}{n^2} d_\alpha(\mathbf{B}^{\{3\},i_1,i_1,i_1}, \mathbf{B}^{(m);\{k\},\{3\},i_1,i_1,i_1}).$$

For a general k, $H_{(m);\{k\}}(i_1, \ldots i_k)$ can be written as

$$H_{(m);\{k\}}(i_1, \ldots, i_k) = \frac{n!}{(n-k)!\, n^k} d_\alpha(\mathbf{B}^{\{k\},i_1,\ldots,i_k}, \mathbf{B}^{(m);\{k\},i_1,\ldots i_k})$$

$$+ \frac{n!}{(n-k)!\, n^k} \sum_{j_1,\ldots,j_k=1}^{n} a_{j_1,\ldots,j_k} d_\alpha(\mathbf{B}^{\{k\},j_1,\ldots,j_k}, \mathbf{B}^{(m);\{k\},j_1,\ldots,j_k}), \quad (5.5)$$

where the sum in the second addend is over all $(j_1, \ldots, j_k)$, with at least one repetition of indices, such that, for any $l \in \{1, \ldots k\}$, there exists $l' \le l$ with $i_{l'} = j_l$, and the coefficient $a_{j_1,\ldots,j_k}$ is the inverse of a positive integer depending on the repetition of indices in $(j_1, \ldots, j_k)$. The only relevant fact is that the number of terms is independent of m and n (for fixed k) and the coefficients $a_{j_1,\ldots,j_k}$ are bounded by 1.

Now, we use Hoeffding's decomposition [14], for the general k layer case: the right-hand side of (5.4) can be rewritten as

$$\frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \frac{1}{\lfloor n/k \rfloor} \sum_{j=1}^{\lfloor n/k \rfloor} H_{(m),\{k\}}\big(\sigma(kj - (k-1)), \ldots, \sigma(kj)\big),$$

where $\mathcal{S}_n$ denotes the set of all permutations of $\{1, \ldots, n\}$. As in the double layer case, an application of the Markov inequality and Jensen's inequality gives, for any $C > 0$ and any n and m,

$$\mathbb{P}\Big[d_{W,\alpha}\big(\mathbf{L}_n^{\mathbf{B};\{k\}}, \mathbf{L}_n^{\mathbf{B}^{(m)};\{k\}}\big) > \delta\Big]$$

$$\le e^{-C\delta} \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \mathbb{E}\Big[\exp\Big(\frac{C}{\lfloor n/k \rfloor} \sum_{i=1}^{\lfloor n/k \rfloor} H_{(m);\{k\}}\big(\sigma(kj - (k-1)), \ldots, \sigma(kj)\big)\Big)\Big].$$

By using the mutual independence of $\{H\big(\sigma(2i-1), \sigma(2i)\big) : i = 1, \ldots, \lfloor n/2 \rfloor\}$, we finally get that

$$\mathbb{P}\Big[d_{W,\alpha}\big(\mathbf{L}_n^{\mathbf{B}}, \mathbf{L}_n^{\mathbf{B}^{(m)}}\big) > \delta\Big] \le e^{-C\delta} \mathbb{E}\Big[\exp\Big(\frac{C}{\lfloor n/k \rfloor} H_{(m);\{k\}}(1, \ldots k)\Big)\Big]^{\lfloor n/k \rfloor}. \quad (5.6)$$

On the other hand, using the equality (5.5), we obtain via Hölder inequality

$$\mathbb{E}\Big[\exp\Big(\frac{C}{\lfloor n/k \rfloor} H_{(m);\{k\}}(1, \ldots k)\Big)\Big]^{\lfloor n/k \rfloor}$$

$$\le \prod_{l=1}^{k} \mathbb{E}\Big[\exp\Big(\frac{c_l C}{\lfloor n/k \rfloor} d_\alpha(\mathbf{B}^{\{l\},i_1,\ldots i_l}, \mathbf{B}^{(m);\{l\},i_1,\ldots i_l})\Big)\Big]^{\lfloor c_l' n/k \rfloor}.$$

Note that the constant c_l, c_l' can depend on $l = 1, \ldots k$ but are independent of n and m, because of the aforementioned uniform bounds on the number of addends and on the coefficients in (5.5). Now, by choosing a suitable C, proportional to $m^\eta n$ and applying Lemma 3.8 to $\mathbf{B}^{(m);\{l\},i_1,\ldots i_l}$, $l = 1, \ldots k$, we get for a suitable constant $c > 0$, for any $\eta \in (0, 1/2 - \alpha)$ and any n large

enough,

$$\sup_{m\geq 1} \mathbb{E}\left[\exp\left(c\frac{m^\eta n}{\lfloor n/k\rfloor}\, H_{(m);\{k\}}(1,\ldots k)\right)\right] < \infty.$$

By combining this estimate with (5.6), the assertion follows. □

Lemma 5.3. *For every $a < \infty$, it holds*

$$\lim_{m\to\infty}\ \sup_{Q:H(Q|P^{\{d\}})\leq a}\ d_{W,\alpha}\big(F^{(m);\{k\}}(Q), F^{(m)}(Q)\big) = 0. \tag{5.7}$$

Proof. The proof of the lemma goes on as the proof of Lemma 3.10. The only changes are: $Q \otimes Q$ must be replaced by $Q^{\otimes k}$, and similarly for the density with respect to $P^{\otimes k}$, μ must be taken as $(P^{\{d\}})^{\otimes k}$, h as $(\frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}})^{\otimes k}$ and the estimate on $\int h \log h \, \mathrm{d}\mu$ becomes $\int h \log h \, \mathrm{d}\mu = kH(Q|P^{\{d\}}) \leq ka$. □

Proof of Theorem 5.1. As for the proof of the double layer case, Lemmata 5.2 and 5.3 allow to apply the extended contraction principle, which gives the desired result. □

We also have the convergence (in probability) of the enhanced k layer empirical measures, which follows again from the LDP.

Corollary 5.4. *The sequence of $\mathcal{P}_1(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$-valued random variables $\{\mathbf{L}_n^{\mathbf{B};\{k\}} : n \in \mathbb{N}\}$ converges in probability (and in law) to the constant random variable $\mathbf{P}^{\{kd\}}$, the enhancement of the kd-Wiener measure, that is the law on $\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})$ of $(B^{1\ldots k}, \mathbb{B}^{1\ldots k})$.*

As in the double layer case, the LDP can be extended to the modified Wasserstein topology on $\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})$. This completes the proof of Theorem 1.1 as stated in the Introduction.

Theorem 5.5. *The sequence $\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})_n$ satisfies a LDP on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$ (endowed with the $(\|\cdot\|+N)^{1+\varepsilon}$-Wasserstein topology) with scale n and good rate function*

$$\mathbf{I}^{\{k\}}(\mu) = \begin{cases} H(\mu\circ\pi_1^{-1} \mid P^{\{d\}}), & \text{if}\ \ \mu = F^{\{k\}}(\mu\circ\pi_1^{-1}),\\ \infty, & \text{otherwise.} \end{cases} \tag{5.8}$$

Proof. The proof is analogous to the proof of Theorem 4.8, we recall only the main points and changes. In all the arguments, $\mathbb{R}^{2d}$ must be replaced in all the arguments by $\mathbb{R}^{kd}$.

First, Lemma 4.9 is extended to $\mathbf{B}^{\{k\};i_1,\ldots,i_k}$ for any multi index $(i_1,\ldots,i_k)$, with a similar proof. Then, Lemma 4.10 is extended to the k layer case and gives the exponential tightness, in the modified Wasserstein topology, of $\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})_n$; the proof is similar to the proof of Lemma 4.10, using the Hoeffding decomposition for the general k layer case (as in the proof of Lemma 5.2) and Lemma 4.9 applied to $\mathbf{B}^{\{k\};1,\ldots,k}$ and to $\mathbf{B}^{\{k\};i_1,\ldots,i_k}$ with repetition of indices. The exponential tightness allows to conclude the LDP in the modified Wasserstein topology, by the inverse contraction principle. □

6. Large deviations for weakly interacting diffusions

We consider an interacting particle system of the following type:

$$
\begin{cases}
\mathrm{d}X_t^{i,n} = \dfrac{1}{n} \sum_{j=1}^{n} b\big(X_t^{i,n}, X_t^{j,n}\big)\,\mathrm{d}t + \mathrm{d}B_t^i, & i = 1, \ldots n, \\
\mathrm{Law}(X_0^{i,n}) = \lambda & \text{i.i.d.}
\end{cases}
\tag{6.1}
$$

Here $X^{i,n}$, $i = 1, \ldots, n$ are the unknown positions of the particles, each of them in $\mathbb{R}^d$, $b \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a given vector field, which we assume regular as much as needed (precisely $C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$), B^i are independent standard $\mathbb{R}^d$-valued Brownian motions, on a fixed filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_t, \mathbb{P})$, and λ is a given probability measure on $\mathbb{R}^d$ satisfying the exponential integrability condition (3.15) for some $c > 0$, $\varepsilon > 0$. We will omit the superscript n when not necessary. It is well known that the above system admits existence and strong uniqueness (i.e. uniqueness for fixed X_0 and B).

The object of interest is an empirical measure associated to this system. For this, let $X^n = (X^{1,n}, \ldots, X^{n,n})$ be the solution to the SDE (6.1). However, we will not, as is classical [17], study

$$
L_n^X = \frac{1}{n} \sum_{i=1}^{n} \delta_{X^{i,n}}
\tag{6.2}
$$

but instead, as $n \to \infty$, the *k-layer, enhanced* empirical measure $L_n^{\mathbf{X};\{k\}}$, defined in complete analogy to the Brownian motion setting. To wit, with $k = 2$ for notational simplicity only,

$$
\mathbf{L}_n^{\mathbf{X},\{2\}}(\omega) \equiv \mathbf{L}_n^{\mathbf{X}}(\omega) := \frac{1}{n^2} \sum_{i,j=1}^{n} \delta_{\mathbf{X}^{ij,n}},
\tag{6.3}
$$

where $\mathbf{X}^{ij,n} = (X^{ij,n}, \mathbb{X}^{ij,n})$ is the rough path on $\mathbb{R}^{2d}$ associated with $X^{ij,n} = (X^{i,n}, X^{j,n})$, defined by $\mathbf{X}^{ij,n} = S^{\{2d\}}(X^{ij,n})$. Clearly, $L_n^{\mathbf{X},\{2\}}(\omega)$ is a (random) measure on $\mathcal{C}_g^{0,\alpha}([0,T]; \mathbb{R}^{2d})$ and we define, on the space of such measures,[8]

$$
\begin{aligned}
\mathbf{K}_b(\mu) = & \int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} \int_0^T \bar{b}(X_t)\,\mathrm{d}\mathbf{X}_t\,\mu(\mathrm{d}\mathbf{X}) \\
& - \frac{1}{2} \int_0^T \int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} \mathrm{div}\bar{b}(X_t)\,\mu(\mathrm{d}\mathbf{X})\,\mathrm{d}t \\
& - \frac{1}{2} \int_0^T \int_{\mathcal{C}_g^{0,\alpha}(\mathbb{R}^{2d})} \left(\int_{\mathcal{C}_g^{0,\alpha}(\mathbb{R}^{2d})} \bar{b}(X_t^1, Y_t^2)\,\mu(\mathrm{d}\mathbf{Y}) \right)^2 \mu(\mathrm{d}\mathbf{X})\,\mathrm{d}t,
\end{aligned}
\tag{6.4}
$$

noting that $\mathbf{K}_b(\mu)$ is well-defined whenever $b \in C_b^2$ and $\mu \in \mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d}))$.

Call Π_2 the projection from $G^2(\mathbb{R}^{kd}) \to G^2(\mathbb{R}^{2d})$.[9] Given a measure μ on $\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})$, the image measure $(\Pi_2)_* \mu \equiv \mu \circ \Pi_2^{-1}$ is a measure on the space $\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})$ of 2-layer rough paths. As previously, $P^{\{d\}}$ is d-dimensional Wiener measure with λ initial distribution. N was introduced in (4.3).

---

[8] Given $b \colon \mathbb{R}^{2d} \to \mathbb{R}^d$, using notation $(x^1, x^2) \in \mathbb{R}^d \times \mathbb{R}^d = \mathbb{R}^{2d}$, we denote by $\bar{b} \colon \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ the function such that $\bar{b}(x^1, x^2)^1 = b(x^1, x^2)$ and $\bar{b}(x^1, x^2)^2 = 0$.

[9] Π_2 is the projection $(X^{12}, \ldots; \mathbb{X}^{12}, \ldots) \mapsto (X^{12}; \mathbb{X}^{12})$.

Theorem 6.1. *Assume that b is in $C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$, let $X^n = (X^{1,n}, \ldots, X^{n,n})$ be the solution to the system (6.1), with initial law λ satisfying Condition (3.15) for fixed $\varepsilon > 0$, and let $\mathbf{L}_n^{\mathbf{X},\{k\}}$ be the corresponding enhanced k-layer empirical measure, $k \geq 2$. Fix $\alpha \in (1/3, 1/2)$. Then, the sequence of laws $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{X}}) : n \in \mathbb{N}\}$ satisfies a large deviation principle on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd}))$ with scale n and good rate function $\mathbf{J}_b$ given by*

$$\mathbf{J}_b^{\{k\}}(\mu) \equiv \mathbf{J}_b(\mu) = \begin{cases} H(\mu \circ \pi_1^{-1} | P^{\{d\}}) - \mathbf{K}_b(\mu \circ \Pi_2^{-1}), & \text{if } \mu = F^{\{k\}}(\mu \circ \pi_1^{-1}), \\ \infty, & \text{otherwise.} \end{cases}$$

(6.5)

Proof. FIRST STEP: Enhanced Girsanov theorem. Let $X = X^n = (X^{1,n}, \ldots, X^{n,n})$ be the solution to the SDE (6.1) with Stratonovich lift $\mathbf{X} = S^{\{nd\}}(X)$. We prove that the law of $\mathbf{X}$ on $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{nd})$ is absolutely continuous with respect to the law of the enhanced nd-dimensional Brownian motion $\mathbf{B}$, with density given by $\exp(\rho_n(\mathbf{B}))$, where ρ_n is deterministically defined by (recall $b = b(x, y)$)

$$\rho_n(\mathbf{Y}) = \frac{1}{n} \sum_{i,j=1}^n \int_0^T \bar{b}(Y_t^{ij}) \cdot \mathrm{d}\mathbf{Y}_t^{ij} - \frac{1}{2n} \sum_{i,j=1}^n \int_0^T \mathrm{div}\bar{b}(Y_t^{ij}) \, \mathrm{d}t$$

$$- \frac{1}{2n} \sum_{i=1}^n \int_0^T \mathrm{div}_y b(Y_t^{ii}) \, \mathrm{d}t$$

$$- \frac{1}{2} \sum_{i=1}^n \int_0^T \left| \frac{1}{n} \sum_{j=1}^n \bar{b}(Y_t^{ij}) \right|^2 \mathrm{d}t.$$

(6.6)

Indeed, the classical Girsanov theorem applied to (6.1), combined with Proposition 2.3, gives that, for every ψ measurable bounded function on $C^\alpha([0, T]; \mathbb{R}^{nd})$,

$$\mathbb{E}[\psi(X)] = \mathbb{E}[e^{\rho_n(\mathbf{B})} \psi(B)].$$

By applying the previously obtained formula to $\psi(X) := \varphi(S^{\{nd\}}(X))$, where φ is any measurable bounded function on $\mathcal{C}^{0,\alpha}([0, T]; \mathbb{R}^{nd})$, we get

$$\mathbb{E}[\varphi(\mathbf{X})] = \mathbb{E}[e^{\rho_n(\mathbf{B})} \varphi(\mathbf{B})],$$

(6.7)

that is enhanced Girsanov theorem.

SECOND STEP: Density for the law of the enhanced empirical measures. First consider the double-layer case $k = 2$. We prove that on the space $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$ the law of the enhanced empirical measure $\mathbf{L}_n^{\mathbf{X}}$ is absolutely continuous with respect to the law of $\mathbf{L}_n^{\mathbf{B}}$, with density given by $\exp(n\mathbf{K}_b) \exp(\mathbf{K}_b')$ for a bounded function $\mathbf{K}_b'$ specified below. The main point is that

$$\rho_n(\mathbf{B}) = n \, \mathbf{K}_b(\mathbf{L}_n^{\mathbf{B}}) + \mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B}}) = n \, \mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{2\}}) + \mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B};\{2\}}),$$

(6.8)

where

$$\mathbf{K}_b'(\mu) = -\int_0^T \int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} \mathrm{div}_y b(\pi_1(\mathbf{Y}_t), \pi_1(\mathbf{Y}_t)) \, \mu(\mathrm{d}\mathbf{Y}) \, \mathrm{d}t.$$

This follows from formula (6.6), the structural reason being the mean field interaction. Now by Lemma B.4 in the Appendix (applied with $k = 2$) the enhanced empirical measure associated

with a rough path in $\mathbb{R}^{nd}$ is a continuous, in particular measurable function G_n of the rough path, that is $\mathbf{L}_n^{\mathbf{X}} = G_n(\mathbf{X})$, $\mathbf{L}_n^{\mathbf{B}} = G_n(\mathbf{B})$. So it is enough to apply formula (6.7) to $\varphi = \phi \circ G_n$, where ϕ is any measurable bounded function on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$, and to use (6.8).

In the case $k > 2$, on the space $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd}))$ the law of the enhanced empirical measure $\mathbf{L}_n^{\mathbf{X},\{k\}}$ has density (with respect to the law of $\mathbf{L}_n^{\mathbf{B},\{k\}}$ given by $\exp(n\mathbf{K}_b \circ \Pi_2) \exp(\mathbf{K}_b' \circ \Pi_2)$. Indeed,

$$(\Pi_2)_* \mathbf{L}_n^{\mathbf{B};\{k\}} = \mathbf{L}_n^{\mathbf{B};\{k\}} \circ \Pi_2^{-1} = \mathbf{L}_n^{\mathbf{B};\{2\}},$$

and therefore

$$\rho_n(\mathbf{B}) = n\,\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{k\}} \circ \Pi_2^{-1}) + \mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B};\{k\}} \circ \Pi_2^{-1}) = n\,\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{2\}}) + \mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B};\{2\}}).$$

We can conclude as in the double-layer case (applying Lemma B.4 to the general k layer case).

THIRD STEP: LDP for $Z_n^{-1} \exp(n\mathbf{K}_b \circ \Pi_2^{-1})\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})$ (and goodness of $\mathbf{J}_b$). We are ready to prove a large deviation principle for the family $Z_n^{-1} \exp(n\mathbf{K}_b)\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})$, where $Z_n = \mathbb{E}[\exp(n\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{2\}}))]$ is the usual renormalization constant. Indeed, the second step invites to apply Varadhan's lemma (Theorem B.3, which is an easy and well-known consequence of Varadhan's lemma in [8, Theorem 4.3.1]). We need to verify the hypotheses, namely, for $k = 2$, that $\mathbf{K}_b$ is a continuous function on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$ and that it holds, for some $\gamma > 1$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\big[\exp\big(n\gamma \mathbf{K}_b(\mathbf{L}_n^{\mathbf{B}})\big)\big] < \infty. \tag{6.9}$$

The hypotheses for general k follow from those for $k = 2$ (so we will fix and omit $k = 2$ in the argument below).

On the continuity of $\mathbf{K}_b$, it is easy to see that the deterministic integrals in formula (6.4) (i.e. the second and third addend) are continuous bounded functions of μ (they are actually continuous bounded functions of $Q = \mu \circ \pi_1^{-1}$ in the C_b-weak topology on $\mathcal{P}(C^{0,\alpha}([0, T]; \mathbb{R}^d)))$, so we concentrate on the term with the rough integral. By Theorem 2.1, the rough integral

$$I_b(\mathbf{X}) := \int_0^T \bar{b}(X)\,\mathrm{d}\mathbf{X} \tag{6.10}$$

is continuous on $\mathcal{C}_g^{0,\alpha}$ with at most linear growth with respect to N (by (4.4)). So, by Proposition 4.6, the term

$$\int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} \int_0^T \bar{b}(X_t)\,\mathrm{d}\mathbf{X}_t\,\mu(\mathrm{d}\mathbf{X}) = \int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} I_b\,\mathrm{d}\mu$$

is continuous on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{2d}))$. Hence $\mathbf{K}_b$ is continuous. Now we prove (6.9) with $\gamma = 2$. We use the fact that

$$M_t = \exp\left(2n\,\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B}}) - \sum_{i=1}^n \int_0^T \frac{2}{n^2} \left|\sum_{j=1}^n b(X^i - X^j)\right|^2 \mathrm{d}t\right)$$

is a martingale, as one can verify easily (and classically). Hence we have

$$\mathbb{E}\big[\exp\big(2n\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B}})\big)\big] = \mathbb{E}\left[M_T \exp\left(\sum_{i=1}^n \int_0^T \frac{2}{n^2} \left|\sum_{j=1}^n b(X^i - X^j)\right|^2 \mathrm{d}t\right)\right]$$

$$\leq \mathrm{e}^{2nT\|b\|_\infty} \mathbb{E}[M_T] = \mathrm{e}^{2nT\|b\|_\infty},$$

which implies (6.9). Hence, we can apply Varadhan's lemma and get the LDP for $\{Z_n^{-1} \exp(n\mathbf{K}_b \circ \Pi_2^{-1})\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}}) : n \in \mathbb{N}\}$ with rate function $\mathbf{J}_b$. Moreover $\mathbf{J}_b$ is good: this follows from exponential tightness of $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}}) : n \in \mathbb{N}\}$ and Varadhan's lemma (in the version B.3).

CONCLUSION. In order to conclude the LDP for $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{X};\{k\}}) : n \in \mathbb{N}\}$, note that

$$\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}}) : n \in \mathbb{N}\} = Z_n \exp(\mathbf{K}_b' \circ \Pi_2^{-1}) \left(Z_n^{-1} \exp(n\mathbf{K}_b \circ \Pi_2^{-1})\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})\right).$$

Therefore, for any Borel set A in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$, we have

$$\frac{1}{n} \log \mathbb{P}[\mathbf{L}_n^{\mathbf{B};\{k\}} \in A] \leq \limsup_n \frac{1}{n} \log\left(Z_n^{-1} \exp(n\mathbf{K}_b \circ \Pi_2^{-1})\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})(A)\right)$$
$$+ \limsup_n \frac{1}{n} \sup_{\mu \in A} \left|\mathbf{K}_b'(\Pi_2^{-1}(\mu))\right| + \limsup_n \frac{1}{n} \left|\log Z_n\right|.$$

Now, $\mathbf{K}_b'$ is bounded on the whole $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$. Moreover,

$$Z_n = \mathbb{E}\left[\exp(n\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{2\}}))\right] = \mathbb{E}\left[\exp(-\mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B};\{2\}})) \exp(n\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{2\}}) + \mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B};\{2\}}))\right],$$

since $\mathbb{E}[\exp(n\mathbf{K}_b(\mathbf{L}_n^{\mathbf{B};\{2\}}) + \mathbf{K}_b'(\mathbf{L}_n^{\mathbf{B};\{2\}}))] = 1$ (the exponential being a density) and $\mathbf{K}_b'$ is bounded from above and from below, $|\log Z_n|$ is bounded uniformly in n. Hence,

$$\frac{1}{n} \log \mathbb{P}[\mathbf{L}_n^{\mathbf{B};\{k\}} \in A] \leq \limsup_n \frac{1}{n} \log\left(Z_n^{-1} \exp(n\mathbf{K}_b \circ \Pi_2^{-1})\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}})(A)\right);$$

similarly (with reverse inequalities) for the $\liminf_n$. Then the LDP for $\{\mathrm{Law}(\mathbf{L}_n^{\mathbf{X};\{k\}}) : n \in \mathbb{N}\}$ follows from that for $\{Z_n^{-1} \exp(n\mathbf{K}_b \circ \Pi_2^{-1})\mathrm{Law}(\mathbf{L}_n^{\mathbf{B};\{k\}}) : n \in \mathbb{N}\}$. The proof is complete. □

We insist that it is crucial in the above proof to work with $k = 2$ (or more) layers, for otherwise the argument – based on continuity of $\mathbf{K}$ – fails. Theorem 6.1 implies immediately, of course, a (known, cf. [2,7]) LPD for the (1-layer, non-enhanced) empirical measure L_n^X as defined in (6.2): it suffices to apply the contraction principle, applied to the map

$$\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha}([0,t];\mathbb{R}^{2d})) \ni \nu \longmapsto \nu \circ \pi_1^{-1} \in \mathcal{P}_1(C^{0,\alpha}([0,T];\mathbb{R}^d)),$$

with resulting (good) rate function

$$J_b(Q) := \inf\{\mathbf{J}_b(\mu) : \mu \circ \pi_1^{-1} = Q\}.$$

The (only) purpose of the following corollary is to re-express this rate function in more familiar terms of stochastic analysis. To this end, we define, for any measure Q on $C^{0,\alpha}([0,T];\mathbb{R}^d)$ which makes the coordinate process, and then also the doubled coordinate process $X = (X^1, X^2)$ under $Q \otimes Q$, a Wiener process plus a square integrable (in time and Ω) drift (this happens when $H(Q|P^{\{d\}}) < \infty$, see the proof of Corollary 6.2),

$$K_b(Q) := \int_{C^{0,\alpha}([0,T];\mathbb{R}^{2d})} \int_0^T \bar{b}(X_t) \circ \mathrm{d}X_t \, (Q \otimes Q)(\mathrm{d}X)$$
$$- \frac{1}{2} \int_0^T \int_{C^{0,\alpha}([0,T];\mathbb{R}^{2d})} \mathrm{div}\,\bar{b}(X_t)\,(Q \otimes Q)(\mathrm{d}X)\mathrm{d}t$$
$$- \frac{1}{2} \int_0^T \int_{C^{0,\alpha}([0,T];\mathbb{R}^d)} \left|\int_{C^{0,\alpha}([0,T];\mathbb{R}^d)} \bar{b}(Y_t, Z_t)\,Q(\mathrm{d}Z)\right|^2 Q(\mathrm{d}Y)\,\mathrm{d}t. \qquad (6.11)$$

Note the last two summands (integrals against $\mathrm{d}t$) are finite under our assumptions on b.

Corollary 6.2. *Under the assumptions of Theorem 6.1, the sequence of laws $\{\mathrm{Law}(L_n^X) : n \in \mathbb{N}\}$ satisfies an LDP on $\mathcal{P}_1(C^{0,\alpha}([0, T]; \mathbb{R}^d))$ with scale n and good rate function J_b given by*

$$J_b(Q) = H(Q \mid P^{\{d\}}) - K_b(Q), \tag{6.12}$$

with the understanding that the right-hand side above is $+\infty$ whenever $H(Q \mid P^{\{d\}}) = +\infty$.

Proof. Consider a measure Q with $H(Q, P^{\{d\}}) = \infty$. We need to show that $\inf\{\mathbf{J}_b(\mu) : \mu \circ \pi_1^{-1} = Q\} = \infty$, that is, $\mathbf{J}_b(\mu) = \infty$ whenever μ projects to Q. By looking at the definition of $\mathbf{J}_b$, there is nothing to show unless $\mu = F^{\{2\}}(\mu \circ \pi_1^{-1}) = F^{\{2\}}(Q)$. But in this case $\mathbf{J}_b(\mu) = H(Q|P^{\{d\}}) - \mathbf{K}_b(\mu) = \infty$, as desired.

We now consider a measure Q with $H(Q, P^{\{d\}}) < \infty$. We have to show that

$$H(Q \mid P^{\{d\}}) - K_b(Q) = \inf\{\mathbf{J}_b(\mu) : \mu \circ \pi_1^{-1} = Q\}.$$

In fact, from the very definition of $\mathbf{J}_b$, we have $\mathbf{J}_b(\mu) = \infty$ unless $\mu = F^{\{2\}}(Q)$. This measure $\mu = F^{\{2\}}(Q)$ satisfies $Q = \mu \circ \pi_1^{-1}$: by Proposition 2.2, denoting $Y = (Y^1, Y^2)$ the canonical process on $C^{0,\alpha}([0, T]; \mathbb{R}^{2d})$, $P^{\{d\}} \otimes P^{\{d\}} \simeq P^{\{2d\}}$-a.s., and so $Q \otimes Q$-a.s., $S^{m,\{2\}}(Y)$ converges to $S^{\{2\}}(Y) = ((Y^1, Y^2), \mathbb{Y})$, in particular $\pi_1(S(Y)) = Y^1$ $Q \otimes Q$-a.s. and so $F^{\{2\}}(Q) \circ \pi_1^{-1} = Q$. For such a μ, we have $\mathbf{J}_b(\mu) = H(\mu \circ \pi_1^{-1}|P^{\{d\}}) - \mathbf{K}_b(\mu) = H(Q|P^{\{d\}}) - \mathbf{K}_2(F^{\{2\}}(Q))$. Thus, it only remains to see that

$$K_b(Q) = \mathbf{K}_b(F^{\{2\}}(Q)).$$

Since $H(Q, P^{\{d\}}) < \infty$, by a classical result (see for example [11, Section II Remark 1.3]), there exists an adapted process g such that $W_t = X_t - X_0 - \int_0^t g_r \, dr$ is a Wiener process under Q and, denoting by ν the marginal of Q at time 0, it holds

$$H(Q \mid P) = H(\nu \mid \lambda) + \frac{1}{2} \mathrm{E}^Q\left[\int_0^t |g_r|^2 \, dr\right]. \tag{6.13}$$

In particular we can define $\int_s^t \bar{b}(X_r) \circ dX_r$, which appears in the definition of K_b (so $K_b(X)$ makes sense), as a Stratonovich integral under $Q \otimes Q$ or equivalently under $P \otimes P$, and by Proposition 2.3 this integral coincides $P \otimes P$-a.s. (and so $Q \otimes Q$-a.s.) with the rough integral $\int_0^T \bar{b}(X_t) \, d\mathbf{X}_t$ in the definition of $\mathbf{K}_b$. Therefore,

$$\int_{C_g^{0,\alpha}([0,T];\mathbb{R}^{2d})} \int_0^T \bar{b}(X_t) \, d\mathbf{X}_t \, \mu(d\mathbf{X}) = \int_{C^{0,\alpha}([0,T];\mathbb{R}^{2d})} \int_0^T \bar{b}(X_t) \circ dX_t \, (Q \otimes Q)(dX),$$

i.e. the first addend in the definitions of $K_b(Q)$ and $\mathbf{K}_b(\mu)$ coincide. The other addends also coincide, as easily verified (they are classical integrals). Therefore $K_b(Q) = \mathbf{K}_b(F^{\{2\}}(Q))$ as desired. □

The above discussion has another useful consequence.

Lemma 6.3. *The rate function given in Theorem 6.1 satisfies*

$$\mathbf{J}_b^{\{k\}}(\mu) = J_b(Q) \tag{6.14}$$

whenever $\mu = F^{\{k\}}(Q)$ and infinite otherwise.

7. Application 1: Robust propagation of chaos

It is an elementary fact of large deviations theory, that a LDP at scale n with good rate function, which has a *single* zero, implies a (weak and in fact – thanks to Borel–Cantelli – strong) law of large numbers. We now give different representations of the rate functions obtained in the last section, which will allow to "see" the single zero. This requires us to consider the following mean field (McKean–Vlasov) SDE on $\mathbb{R}^d$

$$\begin{cases} d\bar{X}_t = (b * u_t)(\bar{X}_t)\,dt + d\bar{B}_t \\ \quad u_t = \text{Law}(\bar{X}_t) \\ \quad \lambda = \text{Law}(\bar{X}_0) \end{cases} \tag{7.1}$$

where

$$(b * u_t)(x) := \int_{\mathbb{R}^d} b(x, y)\, u_t(dy). \tag{7.2}$$

The law $\mathbb{P}^{\bar{X}}$ of the solution $\bar{X}$ can be seen as fixed point of the map Φ defined in this way: for any probability measure Q on $C^{0,\alpha}([0, T], \mathbb{R}^d)$, calling Q_t the marginal of Q at time t, $\Phi(Q)$ is the law of the solution to the SDE

$$\begin{cases} d\bar{Y}_t = (b * Q_t)(\bar{Y}_t)\,dt + d\bar{B}_t \\ \quad \lambda = \text{Law}(\bar{X}_0) \end{cases} \tag{7.3}$$

(for given X_0 and Q, this SDE has a pathwise-unique solution).

Lemma 7.1.

(i) *("one-layer, non-enhanced") The zeros of J_b are precisely fixed points of Φ, as is seen from*

$$J_b(Q) = H(Q \mid \Phi(Q)).$$

(ii) *The zeros of $\mathbf{J}_b^{\{k\}}$ are precisely the image under $F^{\{k\}}$ of fixed points of Φ, as is seen from Lemma 6.3.*

Proof. (i) Indeed, by Girsanov theorem, $\Phi(Q)$ is absolutely continuous with respect to $P^{\{d\}}$, with density satisfying

$$\begin{aligned} \log \frac{d\Phi(Q)}{dP^{\{d\}}}(X) = {} & \int_{C^{0,\alpha}([0,T];\mathbb{R}^d)} \int_0^T \bar{b}(X_t, Z_t) \circ dX_t\, Q(dZ) \\ & - \frac{1}{2} \int_0^T \int_{C^{0,\alpha}([0,T];\mathbb{R}^d)} \text{div}\bar{b}(X_t, Z_t)\, Q(dZ)\, dt \\ & - \frac{1}{2} \int_0^T \left| \int_{C^{0,\alpha}([0,T];\mathbb{R}^d)} \bar{b}(X_t, Z_t)\, Q(dZ) \right|^2 dt \end{aligned}$$

(where we have used stochastic Fubini theorem for exchanging stochastic integration and integration in Q in the first term). Notice that, for Q absolutely continuous with respect to P^d,

$$K_b(Q) = \int_{C^{0,\alpha}([0,T];\mathbb{R}^d)} \log \frac{d\Phi(Q)}{dP^{\{d\}}}(X)\, Q(dX),$$

so we have

$$J_b(Q) \;=\; \int_{\mathcal{C}^{0,\alpha}([0,T];\mathbb{R}^d)} \log \frac{\mathrm{d}Q}{\mathrm{d}P^{\{d\}}} \,\mathrm{d}Q \;+\; \int_{\mathcal{C}^{0,\alpha}([0,T];\mathbb{R}^d)} \log \frac{\mathrm{d}P^{\{d\}}}{\mathrm{d}\Phi(Q)} \,\mathrm{d}Q \;=\; H(Q \mid \Phi(Q)).$$

(ii) The statement follows from Part (i) and Lemma 6.3. $\quad\square$

The previous lemma applies nicely in view of the following result (well-known, see e.g. [17, Theorem 1.1 in Chapter 1]).

Proposition 7.2. *For b as in Theorem 6.1, there is a unique strong solution to* (7.1) *and its law* Law$(\bar{X})$ *is the unique fixed point of* Φ.

We hence know that J_b, although not necessarily convex, has exactly one zero, given by $\mathbb{P}^{\bar{X}} = \mathrm{Law}(\bar{X})$. Similarly, and more importantly, $\mathbf{J}_b^k$ has exactly one zero given by

$$F^{\{k\}}(\mathrm{Law}(\bar{X})) \;=\; (\mathrm{Law}(\bar{X})^{\otimes k}) \circ (S^{\{kd\}})^{-1}.$$

This law is, of course, nothing else that the law of the Stratonovich lift of k IID copies $\bar{X}^1, \ldots, \bar{X}^k$ of the McKean–Vlasov diffusion $\bar{X}$. We can now deduce the enhanced propagation of chaos result as stated in the introduction.

Theorem 7.3. *Under the assumptions of Theorem 6.1 (that is, $b \in C_b^2$) and for all integer k,*

$$\mathrm{Law}\big(S^{\{kd\}}(X^{1,n}, \ldots, X^{k,n})\big) \xrightarrow[n\to\infty]{} \mathrm{Law}\big(S^{\{kd\}}(\bar{X}^1, \ldots, \bar{X}^k)\big), \tag{7.4}$$

as C_b-weak convergence of probability measures on $\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})$ equipped with α-Hölder rough path topology.

Proof. Theorem 6.1 gives us a LDP that quantifies the convergence (in the C_b-weak topology on $\mathcal{P}(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$) in probability

$$\frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^{n} \delta_{S(X^{i_1,n}(\omega),\ldots,X^{i_k,n}(\omega))} \xrightarrow[n\to\infty]{} \mathrm{Law}\big(S_2(\bar{X}^1, \ldots, \bar{X}^k)\big).$$

This convergence follows (by standard reasoning in large deviations on metric spaces) from Theorem 6.1 and Lemma 7.1 which identifies the law of $S_2(\bar{X}^1, \ldots, \bar{X}^k)$ as the unique zero of the rate function. Testing against

$$\varphi \;\in\; C_b\big(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})\big)$$

we get convergence (in probability, as $n \to \infty$)

$$\frac{1}{n^k} \sum_{i_1,\ldots,i_k=1} \varphi\big(S_2(X^{i_1,n}(\omega), \ldots, X^{i_k,n}(\omega))\big) \xrightarrow[n\to\infty]{} \big\langle \varphi, \mathrm{Law}\big(S_2(\bar{X}^1, \ldots, \bar{X}^k)\big)\big\rangle$$
$$= \mathbb{E}\big[\varphi\big(S_2(\bar{X}^1, \ldots, \bar{X}^k)\big)\big].$$

Now take expectation $\mathbb{E}[\cdot]$ on both sides. By using the boundedness of φ,

$$\frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^n \mathbb{E}\big[\varphi\big(S_2(X^{i_1,n}(\omega),\ldots,X^{i_k,n}(\omega)))\big)\big]$$

$$= \frac{1}{n^k} \sum_{\substack{i_1,\ldots,i_k=1 \\ \text{mutually distinct}}}^n \mathbb{E}\big[\varphi\big(S_2(X^{i_1,n}(\omega),\ldots,X^{i_k,n}(\omega)))\big)\big] + \mathcal{O}(1/n)$$

$$= \frac{n!}{n^k(n-k)!}\, \mathbb{E}\big[\varphi\big(S_2(X^{1,n}(\omega),\ldots,X^{k,n}(\omega)))\big)\big] + \mathcal{O}(1/n),$$

we have

$$\mathbb{E}\big[\varphi\big(S_2(X^{1,n}(\omega),\ldots,X^{k,n}(\omega)))\big)\big] \xrightarrow[n\to\infty]{} \mathbb{E}\big[\varphi\big(S_2(\bar{X}^1,\ldots,\bar{X}^k))\big)\big].$$

Since $\varphi \in C_b(\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}))$ is arbitrary, we proved (in rough path topology!), that

$$S_2\big(X^{1,n}(\omega),\ldots,X^{k,n}(\omega)\big) \implies S_2\big(\bar{X}^1,\ldots,\bar{X}^k\big) \qquad \text{as } n\to\infty. \quad \square$$

Remark 7.4. As already noted in the Introduction, this enhanced propagation of chaos is also a consequence of classical propagation of chaos and classical Itô calculus, applying Itô formula to $\int_s^t X_{s,r}^{i,n} \circ \mathrm{d}X_r^{i,n}$. We leave the computations as exercise.

8. Application 2: An LDP for SDEs driven by *k*-layer noises

We start recalling the notation. We fix k in $\mathbb{N}$ and, for a multi-index I in $\{1,\ldots n\}^k$, we use the notation I_j for the jth component of I. We denote by $X^{I,n} = X^{\{k\};I,n}$ the vector $(X^{I_1,n},\ldots X^{I_k,n})$. We take $f_j : \mathbb{R}^d \to \mathbb{R}^m$, $j = 1,\ldots k$, given C_b^3 vector fields. We consider the following family of SDEs on $\mathbb{R}^m$ driven by $X^{i,n}$, parameterized by multi-indices I in $\{1,\ldots n\}^k$:

$$\mathrm{d}Y_t^{I,n} = \sum_{j=1}^k f_j(Y_t^{I,n}) \circ \mathrm{d}X_t^{I_j,n}, \qquad Y_0^{I,n} = y_0, \qquad (8.1)$$

where y_0 is a point in $\mathbb{R}^m$ independent of I and n (however more general choices of initial data should be possible). We call

$$L_n^{Y;\{k\}} = \frac{1}{n^k} \sum_{I \in \{1,\ldots n\}^k} \delta_{Y^{I,n}};$$

it is a random variable with values in $\mathcal{P}(C^{0,\beta}([0,T];\mathbb{R}^m))$, for any $1/3 < \beta < 1/2$.

By rough paths theory, precisely Theorems 8.4 and 8.5 in [12], there exists a (unique) continuous function $\varphi : \mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd}) \to C^{0,\beta}([0,T];\mathbb{R}^m)$ such that, for every I and every n, $Y^{I,n} = \varphi(S^{kd}(X^{I,n}))$ (actually φ is locally Lipschitz continuous). This brings to the following LDP, as recalled in the introduction:

Corollary 8.1. *Fix $1/3 < \beta < 1/2$. The sequence $\{Law(L_n^{Y;\{k\}}) : n \in \mathbb{N}\}$ satisfies a large deviation principle on $\mathcal{P}(C^{0,\beta}([0,T];\mathbb{R}^m))$, endowed with the C_0-weak topology, with scale n and good rate function given by*

$$J^Y(Q) = \inf\{\mathbf{J}^{\{k\}}(\mu) : Q = \mu \circ \varphi^{-1}\}.$$

Proof. We have

$$L_n^{Y;\{k\}} = \mathbf{L}_n^{\mathbf{X};\{k\}} \circ \varphi^{-1}.$$

as it can be easily verified by testing the two measures with a function ψ in $C_b(C^{0,\beta}([0, T]; \mathbb{R}^m))$. In particular $L_n^{Y;\{k\}}$ is the image of $\mathbf{L}_n^{\mathbf{X};\{k\}}$ under the map $F : \mathcal{P}(C_g^{0,\alpha}([0, T]; \mathbb{R}^{kd})) \to \mathcal{P}(C^{0,\beta}([0, T]; \mathbb{R}^m))$, defined by $F(Q) = Q \circ \varphi^{-1}$, which is continuous between the C_0-weak topologies. We then conclude by Theorem 6.1 via contraction principle. $\square$

Acknowledgements

Appendix A. Basic facts on 1-Wasserstein metric

Let (F, d_F) be a Polish space. We denote by $\mathcal{P}_1(F)$ the space of probability measures on F with finite first moment. It is a Polish space endowed with the 1-Wasserstein distance d_W, namely

$$d_W(\mu, \nu) = \inf_{\pi \in \Gamma(\mu, \nu)} \int_{F \times F} d_F(x^1, x^2) \, \pi(\mathrm{d}(x^1, x^2)) \tag{A.1}$$

where $\Gamma(\mu, \nu)$ is the set of all probability measures on $F \times F$ with the first marginal and the second marginal equal resp. to μ and ν (such measures are sometimes called transportation plans). When $F = C^{0,\alpha}([0, T]; E)$ (for some Polish space E), we use the notation $d_{W,\alpha}$ for the 1-Wasserstein distance associated with the α-Hölder distance on $C^{0,\alpha}([0, T]; E)$.

We recall the following characterization of convergence in the 1-Wasserstein metric, stated in [19, Definition 6.8]. Here and in the following, we say that a map $\varphi : F \to F'$ (F, F' being Polish spaces) has at most linear growth if there exist $x_0 \in F$, $y_0 \in F'$ and $C \geq 0$ such that, for every x in F

$$d_{F'}(\varphi(x), y_0) \leq C \left(1 + d_F(x, x_0)\right). \tag{A.2}$$

It is easy to see that this property is equivalent to the following fact: for any x_0 in F, y_0 in F', there exists $C \geq 0$ such that, for every x in F, (A.2) holds.

Lemma A.1. *The following facts are equivalent:*

- $\mu_n \to \mu$ *in 1-Wasserstein distance;*
- $\int_F \varphi(x) \mu_n(\mathrm{d}x) \to \int_F \varphi(x) \mu(\mathrm{d}x)$ *for any function continuous $\varphi : F \to \mathbb{R}$ with at most linear growth;*
- $\mu_n \rightharpoonup \mu$ *and there exists $x_0 \in F$ such that, for any $\eta > 0$, there exists $R > 0$ verifying*

$$\sup_{n \geq 1} \int_{\{d(\cdot, x_0) > R\}} d(x, x_0) \mu_n(\mathrm{d}x) < \eta. \tag{A.3}$$

As a consequence, we have the following Corollary.

Corollary A.2. *Let $h : F \to F'$ be a continuous map (F, F' being Polish spaces) with at most linear growth. Then, the corresponding map at the level of measures, namely $\mathcal{P}_1(F) \ni \mu \mapsto \mu \circ h^{-1} \in \mathcal{P}_1(F')$, is continuous in the 1-Wasserstein metric.*

Proof. Using the equivalence above, it is enough to verify that, for any sequence $\{\mu_n : n \in \mathbb{N}\}$ converging to μ in $\mathcal{P}_1(F)$, for any continuous function $\varphi : F' \to \mathbb{R}$ with at most linear growth, $\int_F \varphi(h(x)) \, \mu_n(\mathrm{d}x) \to \int_F \varphi(h(x)) \, \mu(\mathrm{d}x)$. Now, since h is continuous with at most linear growth, also $\varphi \circ h$ is continuous with at most linear growth, hence the convergence above holds. $\square$

The following lemma provides a wide class of compact sets in the 1-Wasserstein metric.

Lemma A.3. *Let G be a function $G : F \to [0, \infty]$, with compact sublevel sets and with more than linear growth. Define the set*

$$K_M := \left\{ \nu \in \mathcal{P}_1(F) : \int_F G \, \mathrm{d}\nu \leq M \right\}. \tag{A.4}$$

Then K_M is compact (in the 1-Wasserstein metric).

Proof. We prove sequential compactness (which is equivalent to compactness for metric spaces). Let $\{\nu_n : n \in \mathbb{N}\}$ be a sequence of measures in K_M, we will prove that ν_n is tight and that there exists $x_0 \in F$ such that, for every $\eta > 0$, there exists $R > 0$ verifying

$$\sup_{n \geq 1} \int_{\{d(\cdot, x_0) > R\}} d(x, x_0) \, \nu_n(\mathrm{d}x) \ < \ \eta. \tag{A.5}$$

These two conditions imply the existence of a subsequence $\{\nu_{n_k} : k \in \mathbb{N}\}$ converging to some measure ν in $\mathcal{P}_1(F)$ in the 1-Wasserstein metric; it is easy to prove that ν is still in K_M (since the functional $\nu \to \int_F G \, \mathrm{d}\nu$ is lower semi-continuous by [Corollary B.2](#)), so that K_M is compact.

For tightness, we use the compact sublevel sets property of G: for every $\delta > 0$, the set $\{G \leq \delta\}$ is compact and, by Markov inequality, we have, for any n,

$$\nu_n \big[G > \delta^{-1} \big] \ \leq \ \delta \int_F G \, \mathrm{d}\nu_n \ \leq \ \delta M. \tag{A.6}$$

This proves tightness.

For [(A.5)](#), we use the more than linear growth property of G: for some $x_0 \in F$, for any $\eta > 0$, there exists $R > 0$ such that $d(x, x_0)/G(x) < \eta$. Hence, for any n,

$$\int_{\{d(\cdot, x_0) > R\}} d(x, x_0) \, \nu_n(\mathrm{d}x) \ \leq \ \eta \int_{\{d(\cdot, x_0) > R\}} G(x) \, \nu_n(\mathrm{d}x) \ \leq \ \eta M. \tag{A.7}$$

This proves [(A.5)](#) (up to choosing a different R). The lemma is proved. $\square$

We conclude this section with a result on the continuity of the doubling map for measures under the 1-Wasserstein metric. Recall that, if (F, d) is a Polish space, then $(F^2, d^{\{2\}})$ is a Polish space as well, where $d^{\{2\}}((x, y), (x', y'))^2 = d(x, x')^2 + d(y, y')^2$; similarly, for any $k \geq 2$, $(F^k, d^{\{k\}})$ is a Polish space as well, where $d^{\{k\}}((x_1, \ldots, x_k), (x'_1, \ldots, x'_k))^2 = d(x_1, x'_1)^2 + \ldots + d(x_k, x'_k)^2$.

Lemma A.4. *Let F be a Polish space, $k \geq 2$ integer. Then the map*

$$\mathcal{P}_1(F) \ni \mu \ \longmapsto \ \mu^{\otimes k} \in \mathcal{P}_1(F^k) \tag{A.8}$$

is continuous (where $\mathcal{P}_1(F, d)$, $\mathcal{P}_1(F^k, d^{\{k\}})$ are endowed with the 1-Wasserstein distance induced by d and $d^{\{k\}}$, respectively).

Proof. We start with the case $k = 2$. Let μ, ν be two probability measures in $\mathcal{P}_1(F)$. Let π in $\Gamma(\mu, \nu)$ be an admissible plan between μ and ν, namely a probability measure on $F \times F$ with first marginal μ and second marginal ν. Then, an admissible plan $\pi^{\{2\}}$ on $F^2 \times F^2$ between $\mu \otimes \mu$ and $\nu \otimes \nu$ is built from Π as follows: identifying $F^2 \times F^2$ with F^4 and calling q_j, $j = 1, \ldots, 4$, the canonical projections, $\pi^{\{2\}}$ is the unique measure on F^4 such that, under $\pi^{\{2\}}$, (q_1, q_3) and (q_2, q_4) are i.i.d. with distribution π. Indeed, with this definition, $\pi^{\{2\}} \circ (q_1, q_2)^{-1} = \pi \circ (q_1)^{-1} \otimes \pi \circ (q_2)^{-1} = \mu \otimes \mu$ and similarly $\pi^{\{2\}} \circ (q_3, q_4)^{-1} = \nu \otimes \nu$, so $\pi^{\{2\}}$ is in $\Gamma(\mu \otimes \mu, \nu \otimes \nu)$. Now we have

$$
\begin{aligned}
d_W(\mu \otimes \mu, \nu \otimes \nu) &= \inf_{\xi \in \Gamma(\mu \otimes \mu, \nu \otimes \nu)} \int_{F \times F} d^{\{2\}}((x^1, y^1), (x^2, y^2)) \, \xi\big(d((x^1, y^1), (x^2, y^2))\big) \\
&\leq \inf_{\pi \in \Gamma(\mu, \nu)} \int_{F^2 \times F^2} d^{\{2\}}((x^1, y^1), (x^2, y^2)) \, \pi^{\{2\}}\big(d((x^1, y^1), (x^2, y^2))\big) \\
&\leq \inf_{\pi \in \Gamma(\mu, \nu)} \int_{F^2 \times F^2} \big(d(x^1, x^2) + d(y^1, y^2)\big) \, \pi^{\{2\}}\big(d((x^1, y^1), (x^2, y^2))\big) \\
&\leq \inf_{\pi \in \Gamma(\mu, \nu)} \int_{F \times F} d(x^1, x^2) \, \pi\big(d(x^1, x^2)\big) \\
&\quad + \int_{F \times F} d(y^1, y^2) \, \pi\big(d(y^1, y^2)\big) = 2 d_W(\mu, \nu),
\end{aligned}
$$

where in the second inequality we used the simple estimate $d^{\{2\}}((x^1, y^1), (x^2, y^2)) \leq d(x^1, x^2) + d(y^1, y^2)$ and in the third inequality we used the fact that $(x^1, x^2) = (q_1, q_3)$ and $(y^1, y^2) = (q_2, q_4)$ are distributed according to π. The estimate above implies immediately continuity (and even Lipschitz continuity) for $k = 2$.

In the case $k = 2^h$ for some positive integer h, it is enough to note that $\mu \mapsto \mu^{\otimes 2^h}$ is the h-times iteration of the map $\mu \mapsto \mu^{\otimes 2}$. In the case k general, the measure $\mu^{\otimes k}$ is obtained projecting the measure $\mu^{\otimes 2^h}$ on the first k components, for some h with $k \leq 2^h$, so continuity of $\mu \mapsto \mu^{\otimes k}$ follows. $\quad\square$

Appendix B. Technical results and proofs

We start with a known result on lower semi-continuous functions, that we use at least twice in the paper.

Lemma B.1. *Let (E, d) be a metric space. Any lower semi-continuous function $f : E \to (-\infty, \infty]$, bounded from below, is the pointwise supremum of an increasing sequence of continuous (actually Lipschitz) maps.*

Proof. If f is identically $+\infty$, then it is enough to take $f_k \equiv k$ as Lipschitz approximants. Hence, we consider f assuming at least one finite value. We define $\{f_k : k \in \mathbb{N}\}$ as the lower envelope of f, namely

$$
f_k(x) = \inf_{y \in E}\big\{f(y) + k \, d(x, y)\big\}. \tag{B.1}
$$

Since f is bounded from below and not identically $+\infty$, f_k is a real-valued function. The sequence f_k is increasing and, for every k, x, we have $f_k(x) \leq f(x)$ (by choosing $y = x$ in (B.1)). Moreover, for each k, f_k is Lipschitz continuous: for every y, $|(f(y) + k d(x, y))$

$-(f(y) + kd(x', y)) \leq kd(x, x')$ and therefore $|f_k(x) - f_k(x')| \leq kd(x, x')$. We are left to prove the pointwise convergence of f_k to f.

We start with proving convergence on the points x with $f(x)$ finite. Fix $\varepsilon > 0$ and, for every k, take a point x_k such that $f(x_k) + kd(x, x_k) < f_k(x) + \varepsilon$. The sequence $\{x_k : k \in \mathbb{N}\}$ converges to x: indeed $kd(x, x_k) \leq f_k(x) + \varepsilon + (\inf(f))^- \leq f(x) + \varepsilon + (\inf(f))^-$ for every k. Therefore, by lower semi-continuity,

$$f(x) \leq \liminf_{k \to \infty} f(x_k) \leq \liminf_{k \to \infty} f_k(x) + \varepsilon. \tag{B.2}$$

By the arbitrariness of ε, we conclude $f(x) = \lim_{k \to \infty} f_k(x)$.

For the case $f(x) = +\infty$, fix $N > 0$, by lower semi-continuity, there exists $\delta > 0$ such that $f > N$ on $B(x, \delta)$. Therefore $f_k(x) \geq N + k\delta$ and so $\{f_k(x) : k \in \mathbb{N}\}$ converges to $+\infty = f(x)$. The proof is complete. $\square$

Corollary B.2. *Let (E, d) be a metric space and let $f : E \to (-\infty, \infty]$ be lower semi-continuous, bounded from below. Then, for every sequence $\{\mu_n : n \in \mathbb{N}\}$ in $\mathcal{P}(E)$, converging $C_b(E)$-weakly to μ in $\mathcal{P}(E)$, it holds*

$$\int_E f(x)\, \mu(\mathrm{d}x) \leq \liminf_{n \to \infty} \int_E f(x)\, \mu_n(\mathrm{d}x). \tag{B.3}$$

Proof. The previous lemma gives that $f = \sup_{k \geq 1} f_k$, where $\{f_k : k \in \mathbb{N}\}$ is an increasing sequence of continuous functions. We can assume, possibly replacing f_k with $f_k \wedge k$, that f_k is bounded for every k. By monotone convergence theorem, we have for every ν in $\mathcal{P}(E)$

$$\int_E f(x)\, \nu(\mathrm{d}x) = \sup_{k \geq 1} \int_E f_k(x)\, \nu(\mathrm{d}x). \tag{B.4}$$

So the function $\nu \mapsto \int_E f(x)\, \nu(\mathrm{d}x)$ is the supremum of a family of continuous functions in the $C_b(E)$-weak topology, therefore, by a standard argument, it is sequentially lower semi-continuous in that topology. $\square$

Here is the version of Varadhan's lemma we need.

Theorem B.3 (*Varadhan Lemma*). *Let E be a regular Hausdorff space. Suppose that $\{\mu_n : n \in \mathbb{N}\}$ is a sequence of probability measures on E satisfying a large deviation principle with scale n and good rate function I. Let $\varphi : E \to \mathbb{R}$ be a continuous function such that*

$$\limsup_{n \to \infty} \frac{1}{n} \log \int_E \exp(n\gamma\varphi)\, \mathrm{d}\mu_n < \infty \tag{B.5}$$

for some $\gamma > 1$. For any n, let ν_n be the probability measure having density $Z_n^{-1} e^{n\varphi}$ with respect to μ_n (Z_n being the normalization constant). Then the sequence $\{\nu_n : n \in \mathbb{N}\}$ satisfies a large deviation principle with scale n and rate function $J = I - \varphi - \inf_E(I - \varphi)$. It also holds

$$\lim_{n \to \infty} \frac{1}{n} \log Z_n = \inf_E(I - \varphi). \tag{B.6}$$

In particular, if $Z_n = 1$ for each n (i.e. if $e^{n\varphi}\mu_n$ is a probability measure), then $J = I - \varphi$. Furthermore, if $\{\mu_n : n \in \mathbb{N}\}$ is exponentially tight, then so is $\{\nu_n : n \in \mathbb{N}\}$ and the rate function J is good.

Proof. Apart for the last sentence, the statement is a simple consequence of Varadhan lemma in [8, Theorem 4.3.1, Lemma 4.3.4 and Lemma 4.3.6]. The goodness of J follows by [8, Lemma!1.2.18], if we have exponential tightness for $\{\nu_n : n \in \mathbb{N}\}$. Since $\{\mu_n : n \in \mathbb{N}\}$ is exponentially tight, for any $M > 0$, there exists K_M compact set such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n [K_M^c] < -M. \tag{B.7}$$

We have

$$\nu_n [K_M^c] = \frac{1}{Z_n} \int_E 1_{K_M^c} \, \mathrm{e}^{n\varphi} \, \mathrm{d}\mu_n \leq \frac{1}{Z_n} \mu_n [K_M^c]^{1-1/\gamma} \left(\int_E \mathrm{e}^{n\gamma\varphi} \, \mathrm{d}\mu_n \right)^{1/\gamma}. \tag{B.8}$$

Now, using the assumption (B.5), we easily get that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n [K_M^c] < -C(M-1) - \inf_E (I - \varphi) \tag{B.9}$$

for some constant $C > 0$. The proof is complete. $\quad\square$

We prove now the lower-semi-continuity of N_α.

Proof of Lemma 4.2. Notice first that, for any i,

$$\{N \leq i\} = \{\tau_{i+1} \geq T\} \tag{B.10}$$

so that lower semi-continuity of N follows from upper semi-continuity of τ_i, for any i, which we now aim to prove. We must show that, for any i in $\mathbb{N}$, for any $t > 0$,

$$\{\tau_i \geq t\} = \left\{ X \in \mathcal{C}_g^{0,\alpha} : \|X\|_{(1/\alpha)-var,[\tau_{i-1}(\mathbf{X}),t]} \leq 1 \right\} =: A_i(t) \tag{B.11}$$

is a closed set. We use induction on i. For $i = 1$, since $\tau_0 = 0$, closedness follows from continuity of the $(1/\alpha) - var$ norm (with respect to $\mathbf{X}$). For the passage from i to $i + 1$, take $\{\mathbf{X}^m : m \in \mathbb{N}\}$ sequence in $A_{i+1}(t)$ converging to some $\mathbf{X}$ in $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$, we must prove that $\mathbf{X}$ belongs to $A_{i+1}(t)$. By upper semi-continuity of τ_i (inductive hypothesis), we have that $\tau_i(\mathbf{X}) \geq \limsup_{m \to \infty} \tau_i(\mathbf{X}^m)$, so, for any $\delta > 0$, the interval $[\tau_i(\mathbf{X}) + \delta, t]$ is contained in $[\tau_i(\mathbf{X}^m), t]$ for m large enough. So, for any $\delta > 0$, by continuity and monotonicity properties of the $(1/\alpha) - var$ norm, we have

$$\|X\|_{(1/\alpha)-var,[\tau_i(\mathbf{X})+\delta,t]} \leq \limsup_{m \to \infty} \|X\|_{(1/\alpha)-var,[\tau_i(\mathbf{X}^m),t]} \leq 1. \tag{B.12}$$

By arbitrariness of $\delta > 0$ and again by continuity of the norm, we get that $\|X\|_{(1/\alpha)-var,[\tau_i(\mathbf{X}),t]} \leq 1$, that is $\mathbf{X}$ belongs to $A_{i+1}(t)$. The proof is complete. $\quad\square$

Now we prove Lemma 4.5.

Proof of Lemma 4.5. The Hausdorff property follows from the fact that the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology is stronger than the 1-Wasserstein metric (which is an Hausdorff space).

As for the regularity property, we prove it by embedding this space into a topological vector space (which is regular). Precisely, let V be the space of finite signed measures ν on $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e)$, with finite $(|X_0| + \|\cdot\| + N)^{1+\varepsilon}$ moment, i.e.

$$\int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon} \, |\nu|(\mathrm{d}\mathbf{X}) < \infty, \tag{B.13}$$

where $|\nu|$ denotes the total variation measure of ν. We say that a sequence $(\nu_n)_n$ converges to ν in V, in the $(\|\cdot\| + N)^{1+\varepsilon}$-sense, if:

1. $\{\nu_n : n\}$ converges to ν in the C_b-weak topology, i.e. against any test function in $C_b(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^e))$;
2. we have

$$\sup_{n \geq 1} \int_{\mathcal{C}_g^{0,\alpha}} (|X_0| + \|\mathbf{X}\|_\alpha + N_\alpha(\mathbf{X}))^{1+\varepsilon} \, |\nu_n|(d\mathbf{X}) \; < \; \infty. \tag{B.14}$$

This defines a topology on V which we call $(\|\cdot\| + N)^{1+\varepsilon}$ signed topology (or just signed topology). It is easy to see that this topology is Hausdorff and that it makes the operations $V \times V \ni (\nu_1, \nu_2) \mapsto \nu_1 + \nu_2 \in V$, $\mathbb{R} \times V \ni (\alpha, \nu) \mapsto \alpha\nu \in V$ continuous; so V is a Hausdorff topological vector space with the $(\|\cdot\| + N)^{1+\varepsilon}$ signed topology. As a general result in topology, any Hausdorff topological vector space is regular, so V is regular.

It is also easy to see that $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$ is closed in V and that the $(\|\cdot\| + N)^{1+\varepsilon}$ signed topology induces the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology on $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$: any subset in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$ which is closed in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology is also closed in the signed topology and, vice versa, the intersection of any closed (in the signed topology) subset of V with $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$ is closed in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology.

This allows to prove that $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$ is regular (with the original $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology). Indeed, let μ be in $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$ and let C be a closed set in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology. Since C is closed also in the signed topology, then there exist A, B disjoint subset of V, open in the signed topology, such that $\mu \in A$ and $C \subseteq B$. Hence, calling A', resp. B' the intersection of A, resp. B, with $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$, then A', B' are two disjoint subsets of $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$, open in the $(\|\cdot\| + N)^{1+\varepsilon}$-Wasserstein topology, with $\mu \in A'$ and $C \subseteq B'$. This proves regularity of $\mathcal{P}_{(\|\cdot\|+N)^{1+\varepsilon}}(\mathcal{C}_g^{0,\alpha})$. The proof is complete. $\quad\square$

Here we prove that the enhanced empirical measure associated with a rough path in $\mathbb{R}^{nd}$ is a continuous function (in the modified Wasserstein topology) of the rough path itself.

Lemma B.4. *Fix n and k (with $n \geq k$). The map $G_n : \mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{nd}) \to \mathcal{P}_{(\|\cdot\|_\alpha + N_\alpha)^{1+\varepsilon}}$ $(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd}))$ given by*

$$G_n(\mathbf{X}) \; = \; \mathbf{L}_n^{\mathbf{X},\{k\}} \; = \; \frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^{n} \delta_{\mathbf{X}^{\{k\};i_1,\ldots,i_k}}$$

is continuous (in particular measurable).

Proof. Let $\{\mathbf{X}^m : m \in \mathbb{N}\}$ be a sequence of nd-dimensional geometric rough paths, converging to $\mathbf{X}$ in $\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{nd})$ (as $m \to +\infty$), we have to prove that $\mathbf{L}_n^{\mathbf{X}^m,\{k\}}$ converges to $\mathbf{L}_n^{\mathbf{X},\{k\}}$ in the modified Wasserstein topology. We start proving convergence in the C_b-weak topology. For any φ in $C_b(\mathcal{C}_g^{0,\alpha}([0, T]; \mathbb{R}^{kd}))$, we have

$$\int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})} \varphi \, d\mathbf{L}_n^{\mathbf{X}^m,\{k\}} \; = \; \frac{1}{n^k} \sum_{i_1,\ldots,i_k=1}^{n} \varphi(\mathbf{X}^{m,\{k\};i_1,\ldots,i_k}),$$

so convergence of $\int \varphi \, d\mathbf{L}_n^{\mathbf{X}^m,\{k\}}$ to $\int \varphi \, d\mathbf{L}_n^{\mathbf{X},\{k\}}$ follows from continuity of φ (and of the projections on the $(i_1, \ldots i_k)$ components).

To conclude, we have to prove that

$$\sup_{m\geq 1} \int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})} \left(|Y_0| + \|Y\|_\alpha + N_\alpha(Y)\right)^{1+\epsilon} \mathbf{L}_n^{\mathbf{X}^m,\{k\}}(d\mathbf{Y}) \ < \ \infty.$$

For this, we remind that, for any geometric rough path Y, $N_\alpha(\mathbf{Y}) \leq \|\mathbf{Y}\|_{(1/\alpha)-\text{var},[0,T]}^{1/\alpha}$ (see for example [12], Section 11.2.3) and $\|\mathbf{Y}\|_{(1/\alpha)-\text{var},[0,T]} \leq C\|\mathbf{Y}\|_\alpha$ for some constant C (as easily verified). Therefore (using also that $\|\mathbf{Y}^{\{k\};i_1,\dots i_k}\|_\alpha \leq C\|\mathbf{Y}\|_\alpha$ for some C), we get

$$\int_{\mathcal{C}_g^{0,\alpha}([0,T];\mathbb{R}^{kd})} \left(|Y_0| + \|Y\|_\alpha + N_\alpha(Y)\right)^{1+\epsilon} \mathbf{L}_n^{\mathbf{X}^m,\{k\}}(d\mathbf{Y})$$

$$= \frac{1}{n^k} \sum_{i_1,\dots,i_k=1}^{n} \left(|X_0^{m,\{k\};i_1,\dots,i_k}| + \|\mathbf{X}^{m,\{k\};i_1,\dots,i_k}\|_\alpha + N_\alpha(\mathbf{X}^{m,\{k\};i_1,\dots i_k})\right)^{1+\epsilon}$$

$$\leq C\left(|X_0^m| + \|\mathbf{X}^m\|_\alpha + \|\mathbf{X}^m\|_\alpha^{1/\alpha}\right)^{1+\epsilon}$$

and the RHS above is uniformly bounded in m. The proof is complete. $\quad\square$

Finally we prove Lemma 3.8, starting from Corollary 13.22 in [13], following Exercise 13.22 there, and Lemma 4.9, starting from Theorems 11.9 and 11.13 in [12] (see also [3], Theorem 6.3).

Proof of Lemma 3.8. Corollary 13.22 in [13] applies clearly also to Brownian rough path starting from any initial measure (since $\mathbf{B}$ and $\mathbf{B}^{(m)}$ start from the same point) and gives the existence of a constant $C > 0$ such that, for every $q \geq 1$, for every m, it holds

$$\mathbb{E}\left[d_\alpha(\mathbf{B}^{(m)}, \mathbf{B})^q\right] \ \leq \ (Cq^{1/2}m^{-\eta/2})^q.$$

From this we get the following estimate on the exponential of the distance above: for any $\rho > 0$,

$$\mathbb{E}\left[\exp(\rho d_\alpha(\mathbf{B}^{(m)}, \mathbf{B}))\right] \ = \ 1 + \sum_{q=1}^\infty \frac{\rho^q \mathbb{E}\left[d_\alpha(\mathbf{B}^{(m)}, \mathbf{B})^q\right]}{q!} \ \leq \ 1 + \sum_{q=1}^\infty \frac{(\rho C q^{1/2}m^{-\eta/2})^q}{q!}.$$

Using the elementary estimate $q^q \leq e^{q-1}q!$ (which can be easily proved by induction on q), we have

$$\mathbb{E}\left[\exp(\rho d_\alpha(\mathbf{B}^{(m)}, \mathbf{B}))\right] \ \leq \ 1 + \sum_{q=1}^\infty (e\rho C m^{-\eta/2})^q.$$

So, taking $\rho = m^{\eta/2}/(2eC)$, we get that this series converges. Hence,

$$\mathbb{E}\left[\exp\left((2eC)^{-1}m^{\eta/2}d_\alpha(\mathbf{B}^{(m)}, \mathbf{B})\right)\right] \ < \ \infty. \tag{B.15}$$

The proof is complete. [Notice that some estimates were not optimal: in fact the result holds also for $d_\alpha(\mathbf{B}^{(m)}, \mathbf{B})^2$ replacing $d_\alpha(\mathbf{B}^{(m)}, \mathbf{B})$.] $\quad\square$

Proof of Lemma 4.9. Notice that (for $\varepsilon < 1$, using independence of the initial datum and the increments of Brownian motion)

$$\mathbb{E}\left[\exp\left(c(|B_0| + \|\mathbf{B}\|_\beta + N_\alpha(\mathbf{B}))^{1+\varepsilon}\right)\right]$$
$$\leq \ \mathbb{E}\left[e^{2c|B_0|}\right] \mathbb{E}\left[\exp\left(2c(\|\mathbf{B}\|_\beta + N_\alpha(\mathbf{B}))^{1+\varepsilon}\right)\right]. \tag{B.16}$$

Now, $\mathbb{E}\big[\exp\big(2c(\|\mathbf{B}\|_\beta + N_\alpha(\mathbf{B}))^{1+\varepsilon}\big)\big]$ is finite (actually for every $c > 0$), as proved in Theorems 11.9 and 11.13 in [12]; $\mathbb{E}\big[e^{2c|B_0|}\big] = \int_{\mathbb{R}^e} e^{2cx}\,\tilde{\lambda}(\mathrm{d}x)$ is finite because of the exponential integrability condition (3.15) (replacing c with $2c$). The same proof applies to $\mathbf{B}^{11}$ (and to $\mathbf{B}^{\{k\};i_1,\dots,i_k}$ for any multi-index $(i_1,\dots,i_k)$ also with repetition of indices). $\square$

References

[1] I. Bailleul, Flows driven by rough paths, Rev. Mat. Iberoam. (ISSN: 0213-2230) 31 (3) (2015) 901–934 MR 3420480.

[2] A. Budhiraja, P. Dupuis, M. Fischer, Large deviation properties of weakly interacting processes via weak convergence methods, Ann. Probab. (ISSN: 0091-1798) 40 (1) (2012) 74–102 MR 2917767.

[3] T. Cass, C. Litterer, T. Lyons, Integrability and tail estimates for gaussian rough differential equations, Ann. Probab. (ISSN: 0091-1798) 41 (4) (2013) 3026–3050 MR 3112937.

[4] T. Cass, T. Lyons, Evolving communities with individual preferences, Proc. Lond. Math. Soc. (3) (ISSN: 0024-6115) 110 (1) (2015) 83–107 MR 3299600. http://dx.doi.org/10.1112/plms/pdu040.

[5] P. Dai Pra, F. den Hollander, McKean-Vlasov limit for interacting random processes in random media, J. Stat. Phys. (ISSN: 0022-4715) 84 (3–4) (1996) 735–772 MR 1400186 (97f:60208). http://dx.doi.org/10.1007/BF02179656.

[6] D.A. Dawson, J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions, Stochastics (ISSN: 0090-9491) 20 (4) (1987) 247–308 MR 885876 (89c:60092). http://dx.doi.org/10.1080/17442508708833446.

[7] P. Del Moral, T. Zajic, A note on the Laplace-Varadhan integral lemma, Bernoulli (ISSN: 1350-7265) 9 (1) (2003) 49–65 MR 1963672 (2004a:60061). http://dx.doi.org/10.3150/bj/1068129010.

[8] A. Dembo, O. Zeitouni, Large deviations techniques and applications, in: Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, ISBN: 978-3-642-03310-0, 2010 Corrected reprint of the second (1998) edition, MR 2571413 (2011b:60094). http://dx.doi.org/10.1007/978-3-642-03311-7.

[9] J.-D. Deuschel, D.W. Stroock, Large deviations, in: Pure and Applied Mathematics, vol. 137, Academic Press, Inc., Boston, MA, ISBN: 0-12-213150-9, 1989 MR 997938 (90h:60026).

[10] R.M. Dudley, Real analysis and probability, in: Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, ISBN: 0-521-00754-2, 2002 Revised reprint of the 1989 original, MR 1932358 (2003h:60001). http://dx.doi.org/10.1017/CBO9780511755347.

[11] H. Föllmer, Random fields and diffusion processes, in: École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87, in: Lecture Notes in Math., vol. 1362, Springer, Berlin, 1988, pp. 101–203 MR 983373 (90f:60099). http://dx.doi.org/10.1007/BFb0086180.

[12] P.K. Friz, M. Hairer, A course on rough paths, in: Universitext, Springer, Cham, 2014 With an introduction to regularity structures, MR 3289027, ;. http://dx.doi.org/10.1007/978-3-319-08332-2.

[13] P.K. Friz, N.B. Victoir, Multidimensional stochastic processes as rough paths, in: Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge University Press, Cambridge, ISBN: 978-0-521-87607-0, 2010 Theory and applications, MR 2604669 (2012e:60001). http://dx.doi.org/10.1017/CBO9780511845079.

[14] Wassily Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. (ISSN: 0162-1459) 58 (1963) 13–30 MR 0144363 (26 #1908).

[15] M. Ledoux, Z. Qian, T. Zhang, Large deviations and support theorem for diffusion processes via rough paths, Stochastic Process. Appl. (ISSN: 0304-4149) 102 (2) (2002) 265–283 MR 1935127 (2003m:60152). http://dx.doi.org/10.1016/S0304-4149(02)00176-X.

[16] M.M. Rao, Z.D. Ren, Theory of Orlicz spaces, in: Monographs and Textbooks in Pure and Applied Mathematics, vol. 146, Marcel Dekker, Inc., New York, ISBN: 0-8247-8478-2, 1991 MR 1113700 (92e:46059). http://dx.doi.org/10.1080/03601239109372748.

[17] A.-S. Sznitman, Topics in propagation of chaos, in: ÉCole D'ÉTÉ De ProbabilitÉS De Saint-FLour XIX—1989, in: Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 165–251 MR 1108185 (93b:60179). http://dx.doi.org/10.1007/BFb0085169.

[18] H. Tanaka, Limit theorems for certain diffusion processes with interaction, in: Stochastic Analysis (Katata/KYoto, 1982), in: North-Holland Math. Library, vol. 32, North-Holland, Amsterdam, 1984, pp. 469–488 MR 780770 (86i:60204). http://dx.doi.org/10.1016/S0924-6509(08)70405-7.

[19] C. Villani, Optimal transport, in: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, ISBN: 978-3-540-71049-3, 2009 Old and new, MR 2459454 (2010f:49001). http://dx.doi.org/10.1007/978-3-540-71050-9.

[20] R. Wang, X. Wang, L. Wu, Sanov's theorem in the wasserstein distance: a necessary and sufficient condition, Statist. Probab. Lett. (ISSN: 0167-7152) 80 (5–6) (2010) 505–512 MR 2593592 (2010k:60106). http://dx.doi.org/10.1016/j.spl.2009.12.003.