Journal of Machine Learning Research 25 (2024) 1-76 Submitted 8/22; Revised 1/24; Published 7/24

Law of Large Numbers and Central Limit Theorem for Wide Two-layer Neural Networks: The Mini-Batch and Noisy Case

Arnaud Descours ARNAUD.DESCOURS@UCA.FR
Laboratoire de Mathématiques Blaise Pascal UMR 6620
Université Clermont Auvergne
Aubière, France

Arnaud Guillin ARNAUD.GUILLIN@UCA.FR
Laboratoire de Mathématiques Blaise Pascal UMR 6620 and Institut Universaire de France
Université Clermont Auvergne
Aubière, France

Manon Michel MANON.MICHEL@UCA.FR
CNRS, Laboratoire de Mathématiques Blaise Pascal UMR 6620
Université Clermont Auvergne
Aubière, France

Boris Nectoux BORIS.NECTOUX@UCA.FR
Laboratoire de Mathématiques Blaise Pascal UMR 6620
Université Clermont Auvergne
Aubière, France

Editor: Pradeep Ravikumar

Abstract

In this work, we consider a wide two-layer neural network and study the behavior of its empirical weights under a dynamics set by a stochastic gradient descent along the quadratic loss with mini-batches and noise. Our goal is to prove a trajectorial law of large number as well as a central limit theorem for their evolution. When the noise is scaling as $1/N^\beta$ and $1/2 < \beta \le \infty$, we rigorously derive and generalize the LLN obtained for example in (Chen et al., 2020; Mei et al., 2019; Sirignano and Spiliopoulos, 2020b). When $3/4 < \beta \le \infty$, we also generalize the CLT (see also (Sirignano and Spiliopoulos, 2020a)) and further exhibit the effect of mini-batching on the asymptotic variance which leads the fluctuations. The case $\beta = 3/4$ is trickier and we give an example showing the divergence with time of the variance thus establishing the instability of the predictions of the neural network in this case. It is illustrated by simple numerical examples.

Keywords: Machine learning; Neural networks; Law of large numbers; central limit theorem; Empirical measures; Particle systems; Mean field

Contents

1. Setting and main results

1.1 Introduction

Setting and purpose of this work. Thanks to their impressive results, deep learning techniques have nowadays become standard supervised learning methods in various fields of engineering or research (Goodfellow et al., 2016). A robust understanding of their behavior and efficiency is however still lacking and a large effort is put towards achieving mathematical foundations of empirical observations. Among this effort, the case of wide two-layer single network, and its connection with mean-field network, has particularly been fruitful, as considered for example in (Rotskoff and Vanden-Eijnden, 2018; Mei et al., 2018; Sirignano and Spiliopoulos, 2020b,a). In such setting, a convergence towards a limit PDE system can be established when the neuron numbers goes to infinity. The behavior in long time of this limit PDE may then give an easier framework to establish the convergence towards minimizers of the loss function of the neural network. Partial results can be found in this direction (Mei et al., 2018; Chizat and Bach, 2018) but as underlined in (E, 2020), a lot still remains to be understood and proved mathematically rigorously. In this context, our work is two-fold. First, we will concern ourselves with the mathematical justification of the law of large numbers and central limit theorems of the trajectory of the empirical measure of the weights, under the optimization by a stochastic gradient descent (SGD), with mini-batching and in the presence of noise with a range of scalings. Mini-batch SGD (Bottou et al., 2018) is widely used in machine learning since it allows for shorter training times thanks to parallelisation, while reducing the variance in SGD estimates. How to choose the optimal mini-batch size, and furthermore with theoretical guarantees, remains an active research line (Keskar et al., 2017; Smith and Le, 2018; Gower et al., 2019). Introducing noise in SGD, as considered in (Mei et al., 2018), can lead to better generalisation perfomance thanks to an improved ability to escape saddle points, as shown in (Jin et al., 2021). Note that this differs from the analysis approach consisting in directly modelizing the noise of SGD as for instance done in (Wu et al., 2020; Simsekli et al., 2019). Second we will do so by providing a rigorous framework which could be generalized to study overparametrized limit of other neural networks (e.g. deep ensemble, bayesian neural networks, ...). Thus, the benefit of the overparametrized limit and its convexification of the loss landscape through a non-linear PDE could lead in these different architectures to derivations of theoretical guarantees of convergence, while it remains hard to analyse these landscapes directly in the case of a finite number of neurons, even large.

Let us now precise the framework for this paper. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and $\mathcal{X}$ and $\mathcal{Y}$ be subsets of $\mathbf{R}^{\mathsf{n}}$ ($\mathsf{n} \geq 1$) and $\mathbf{R}$ respectively. In this work, we consider the following two-layer neural network

$$g_W^N(x) := \frac{1}{N}\sum_{i=1}^N \sigma_*(W^i, x), \tag{1}$$

where $x \in \mathcal{X}$ denotes the input data, $g_W^N(x) \in \mathbf{R}$ the output returned by the neural network, $\sigma_* : \mathbf{R}^d \times \mathcal{X} \to \mathbf{R}$ the activation function, $N \geq 1$ the number of neurons on the hidden layer, and $W = (W^1, \ldots, W^N) \in (\mathbf{R}^d)^N$ are the weights to optimize ($d \geq 1$). In the supervised learning setting, a data point $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is distributed according to $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, where $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ denotes the set of probability measures on $\mathcal{X} \times \mathcal{Y}$. Ideally, one chooses the weights $W = (W^1, \ldots, W^N)$ as a global minimizer of the risk $\mathbf{E}_\pi[\mathsf{L}(g_W^N(x), y)]$, where $\mathsf{L} : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is the so-called loss function ($\mathbf{E}_\pi$ stands for the expectation when $(x, y) \sim \pi$). In this work, we consider the square loss function out of simplicity, but other loss function or classification problem could be considered, namely:

$$\mathsf{L}(g_W^N(x), y) = \frac{1}{2}\left|g_W^N(x) - y\right|^2.$$

Since the risk can not be computed (because π is unknown), the parameters are usually learned by stochastic gradient descent. In this work, we consider the mini-batch setting with weak noise, which is defined as follows. First, for $k \geq 0$, consider $((x_k^n, y_k^n))_{n\geq 1}$ a sequence of random elements on $\mathcal{X} \times \mathcal{Y}$ (each (x_k^n, y_k^n) being distributed according to π), and N_k a random element with values in $\mathbf{N}^* = \{1, 2, 3, \ldots\}$. Then, the mini-batch B_k is defined by:

$$B_k = \{(x_k^1, y_k^1), \dots, (x_k^{N_k}, y_k^{N_k})\}, \text{where } |B_k| \text{ denotes the cardinality of } B_k.$$

In particular $|B_k| = N_k$. In addition, at each iteration of SGD, we add a Gaussian noise term, whose variance is scaled according to $N^{-2\beta}$, with $\beta > \frac{1}{2}$, hence qualified *weak*. Note that the case of Gaussian noise with $\beta = 1/2$ is addressed in (Mei et al., 2018) and could also be considered here in our setting, but with additional assumptions to integrate the noise in the limit process.

Thus, the SGD algorithm we consider is the following : for $k \geq 0$ and $i \in \{1, \dots, N\}$,

$$\begin{cases} W_{k+1}^i = W_k^i + \dfrac{\alpha}{N|B_k|} \displaystyle\sum_{(x,y)\in B_k} (y - g_{W_k}^N(x)) \nabla_W \sigma_*(W_k^i, x) + \dfrac{\varepsilon_k^i}{N^\beta}, \\ W_0^i \sim \mu_0, \end{cases} \tag{2}$$

where $\alpha > 0$, $\varepsilon_k^i \sim \mathcal{N}(0, I_d)$ and $\mu_0 \in \mathcal{P}(\mathbf{R}^d)$. The evolution of the weights is tracked through their empirical distribution ν_k^N (for $k \geq 0$) and its scaled version μ_t^N (for $t \in \mathbf{R}_+$), which are defined as follows:

$$\nu_k^N := \frac{1}{N} \sum_{i=1}^N \delta_{W_k^i} \quad \text{and} \quad \mu_t^N := \nu_{\lfloor Nt \rfloor}^N.$$

For an element $\mu \in \mathcal{M}_b(\mathbf{R}^d)$ (the space of bounded countably additive measures on $\mathbf{R}^d$), we use the notation

$$\langle f, \mu \rangle_{\mathrm{m}} = \int_{\mathbf{R}^d} f(w) \mu(\mathrm{d}w),$$

for any $f : \mathbf{R}^d \to \mathbf{R}$ such that $\int_{\mathbf{R}^d} f(w)\mu(\mathrm{d}w)$ exists. If no confusion is possible, we simply denote $\langle f, \mu \rangle_{\mathrm{m}}$ by $\langle f, \mu \rangle$. For instance, considering the neural network (1), we have, for any $x \in \mathcal{X}$,

$$g_{W_k}^N(x) = \frac{1}{N} \sum_{i=1}^N \sigma_*(W_k^i, x) = \langle \sigma_*(\cdot, x), \nu_k^N \rangle, \quad k \geq 0.$$

In this work, we prove that the the whole trajectory of the scaled empirical measures of the weights defined by (2) (namely $\{t \mapsto \mu_t^N, t \in \mathbf{R}_+\}_{N\geq 1}$) satisfies a law of large numbers and a central limit theorem, see respectively Theorem 1 and Theorem 2. We also exhibit a particular fluctuation behavior depending on the value of the parameter β ruling the weakness of the added noise.

Related works. Law of large numbers and central limits theorems have been obtained for several kinds of mean-field interacting particle systems in the mathematical literature, see for instance (Sznitman, 1991; Hitsuda and Mitoma, 1986; Fernandez and Méléard, 1997; Jourdain and Méléard, 1998; Delarue et al., 2019; Del Moral and Guionnet, 1999; Kurtz and Xiong, 2004) and references therein. When considering particle systems arising from the SGD-minimization problem in a two-layer neural network, we refer to (Mei et al., 2018) for a law of large numbers on the empirical measure at fixed times, see also (Mei et al., 2019). We also refer to (Rotskoff and Vanden-Eijnden, 2018) where conditions for global convergence of the GD on the ideal loss and of the SGD with mini-batches increasing in size with N, as well as the scaling of the error with the size of the network, are established from formal asymptotic arguments. Doing so, they also observe with increasing mini-batch size in the SGD the reduction of the variance of the process leading the fluctuations of the empirical measure of the weights (see (Rotskoff and Vanden-Eijnden, 2018, Arxiv-V2. Sec 3.3)), until the mini-batches are large enough to recover the situation of the idealized gradient descent (similar to an infinite batch), which leads to other order of fluctuations (see (Rotskoff and Vanden-Eijnden, 2018, Arxiv-V2. Prop 2.3)). We also refer to (Chen et al., 2020) for a similar line of work on the GD on the empirical loss. A law of large numbers and a central limit theorem on the whole trajectory of the empirical measure are also obtained in (Sirignano and Spiliopoulos, 2020b,a) for a standard SGD scheme. We also mention the work done in (De Bortoli et al., 2020) on propagation of chaos for SGD with different step-size schemes. In this work, and compared to the existing literature dealing with the SGD minimization problem in two-layer neural networks, we provide a rigorous proof with precise justifications of all steps of the existence of the limit PDE (in particular, uniqueness and relative compactness) in the law of large numbers as well as the limit process for the central limit theorem on the trajectory of the empirical measure. We also mention that many of our arguments used in this work truly differ from those used in (Sirignano and Spiliopoulos, 2020b,a) as several key proofs there were not fullly satisfactory. We thus provide a careful treatment of topological issues regarding relative compactness of μ^N, the uniqueness of the mean-field limit equation (10), as well as the convergence of the martingale process in the CLT which actually requires extra non trivial analysis to hold in the space $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$, see Section 3.4. Also, contrary to what is claimed in (Sirignano and Spiliopoulos, 2020b), the functional $\mathbf{\Lambda}_t^N[f]$ (see Section 2.2.3) is not continuous in the Skorohod topology, and for that reason extra arguments must be used to pass to the limit in the pre-limit equation. Such a rigorous derivation will be the basis for future works on deep ensembles or overparameterized bayesian neural networks. We furthermore do so in a more general variant of SGD with mini-batching of any size and weak noise (see Equation (2)). A noisy SGD was also considered in (Mei et al., 2018), corresponding to $\beta = 1/2$ in our setting, for which they obtain for the LLN a different limit PDE than in the non-noisy case (presence of an additionnal regularizing Laplacian term in the limit equation). While we could recover in a straightforward manner a trajectorial version of (Mei et al., 2018), we consider here out of concision the range $\beta > 1/2$, showing a single limit PDE for the LLN, and obtain a similar result for $\beta > 3/4$ for the CLT, while showing analytically for $\beta = 3/4$ and numerically for $\beta \leq 3/4$ a particular fluctuation behavior. Furthermore, we analytically show the expected reduction, with the mini-batch size, of the variance of the process leading the fluctuations of the weight empirical measure and numerically display the reduction of the global variance.

1.2 Main results

1.2.1 Notation and assumptions

Weighted Sobolev spaces. Following (Adams and Fournier, 2003, Chapter 3), we consider, for a function $g \in \mathcal{C}_c^\infty(\mathbf{R}^d)$ (the space of functions $g : \mathbf{R}^d \to \mathbf{R}$ of class $\mathcal{C}^\infty$ with compact support), the following norm, defined for $J \in \mathbf{N}$ and $b \geq 0$:

$$\|g\|_{\mathcal{H}^{J,b}} := \Big(\sum_{|k| \leq J} \int_{\mathbf{R}^d} \frac{|D^k g(x)|^2}{1 + |x|^{2b}} \mathrm{d}x \Big)^{1/2}.$$

Let $\mathcal{H}^{J,b}(\mathbf{R}^d)$ be the closure of the set $\mathcal{C}_c^\infty(\mathbf{R}^d)$ for this norm. The space $\mathcal{H}^{J,b}(\mathbf{R}^d)$ is a Hilbert space when endowed with the norm $\| \cdot \|_{\mathcal{H}^{J,b}}$. The associated scalar product on $\mathcal{H}^{J,b}(\mathbf{R}^d)$ will be denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}^{J,b}}$. We denote by $\mathcal{H}^{-J,b}(\mathbf{R}^d)$ its dual space. For an element $\Phi \in \mathcal{H}^{-J,b}(\mathbf{R}^d)$, we use the notation

$$\langle f, \Phi \rangle_{J,b} = \Phi[f], \ f \in \mathcal{H}^{J,b}(\mathbf{R}^d).$$

For ease of notation, and if no confusion is possible, we simply denote $\langle f, \Phi \rangle_{J,b}$ by $\langle f, \Phi \rangle$. Let us now define $\mathcal{C}^{J,b}(\mathbf{R}^d)$ as the space of functions $g : \mathbf{R}^d \to \mathbf{R}$ with continuous partial derivatives up to order $J \in \mathbf{N}$ such that

$$\text{for all } |k| \leq J, \ \lim_{|x| \to \infty} \frac{|D^k g(x)|}{1 + |x|^b} = 0.$$

This space is endowed with the norm

$$\|g\|_{\mathcal{C}^{J,b}} := \sum_{|k| \leq J} \sup_{x \in \mathbf{R}^d} \frac{|D^k g(x)|}{1 + |x|^b}.$$

We also introduce $\mathcal{C}_b(\mathbf{R}^d)$, the space of bounded continuous functions $g : \mathbf{R}^d \to \mathbf{R}$, endowed with the supremum norm. We also denote by $\mathcal{C}_b^\infty(\mathbf{R}^d)$ the space of smooth functions over $\mathbf{R}^d$ whose derivatives of all order are bounded. We have $\mathcal{C}_b^\infty(\mathbf{R}^d) \subset \mathcal{H}^{J,b}(\mathbf{R}^d)$ as soon as $b > d/2$ (more generally $x \in \mathbf{R}^d \mapsto (1 - \chi(x))|x|^a \in \mathcal{H}^{J,b}(\mathbf{R}^d)$ if $b - a > d/2$, where $\chi \in \mathcal{C}_c^\infty(\mathbf{R}^d, [0, 1])$ equals 1 near 0).

Weighted Sobolev embeddings. We recall that from (Fernandez and Méléard, 1997, Section 2),

$$\mathcal{H}^{\ell+j,a}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{j,a+b}(\mathbf{R}^d) \text{ when } \ell > d/2, \, b > d/2, \text{ and } a, j \geq 0 \tag{3}$$

where $\hookrightarrow_{\text{H.S.}}$ means that the embedding is of Hilbert-Schmidt type, and

$$\mathcal{H}^{\ell+j,a}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{j,a}(\mathbf{R}^d) \text{ when } \ell > d/2, \text{ and } a, j \geq 0. \tag{4}$$

We set

$$L = \lceil \frac{d}{2} \rceil + 3, \ \gamma = 4 \lceil \frac{d}{2} \rceil + 5, \text{ and } \gamma_* := \gamma + 1. \tag{5}$$

According to (4) and since $\gamma_* > \gamma$, it holds:

$$\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{2,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d). \tag{6}$$

We set throughout this work, for all $N \geq 1$:

$$\mu^N := \{t \mapsto \mu_t^N, t \in \mathbf{R}_+\}.$$

When E is a metric space, we denote by E' its dual and by $\mathcal{D}(\mathbf{R}_+, E)$ the set of càdlàg functions[1] from $\mathbf{R}_+$ to E (endowed with the Skorohod $J1$-topology, see (Billingsley, 1999; Ethier and Kurtz, 2009)). For $b \geq 0$ and for all $N \geq 1$, μ^N is a random element of $\mathcal{D}(\mathbf{R}_+, \mathcal{C}^{0,b}(\mathbf{R}^d)')$, and thus also of $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J,b}(\mathbf{R}^d))$, as soon as $J > d/2$ (by (4)).

Let for $k \geq 1$,

$$\mathcal{P}_k(\mathbf{R}^d) := \big\{\mu \in \mathcal{P}(\mathbf{R}^d),\ \int_{\mathbf{R}^d} |w|^k \mu(\mathrm{d}w) < +\infty\big\}, \tag{7}$$

which is endowed with the Wasserstein distance

$$\mathsf{W}_k(\mu,\nu) = \big[\inf\{\mathbf{E}[|X-Y|^k|],\ \mathbf{P}_X = \mu \text{ and } \mathbf{P}_Y = \nu\}\big]^{1/k}.$$

We refer for instance to (Santambrogio, 2015, Chapter 5) for more about these spaces. We recall that $\mathsf{W}_1(\mu,\nu) \leq \mathsf{W}_k(\mu,\nu)$ $(k \geq 1)$ and the dual formula for $\mathsf{W}_1(\mu,\nu)$:

$$\mathsf{W}_1(\mu,\nu) = \sup\Big\{\big|\int_{\mathbf{R}^d} f(w)\mathrm{d}\mu(w) - \int_{\mathbf{R}^d} f(w)\nu(\mathrm{d}w)\big|,\ \|f\|_{\mathrm{Lip}} \leq 1\Big\}. \tag{8}$$

Note also that for all $N \geq 1$, μ^N is a random element of $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_q(\mathbf{R}^d))$, for all $q \geq 0$.

Assumptions. For $N \geq 1$, we introduce the $\boldsymbol{\sigma}$-algebras,

$$\mathcal{F}_0^N = \boldsymbol{\sigma}\{\{W_0^i\}_{i=1}^N\} \text{ and, for } k \geq 1,\ \mathcal{F}_k^N = \boldsymbol{\sigma}\big\{W_0^i, \{B_j\}_{j=0}^{k-1}, \{\varepsilon_j^i\}_{j=0}^{k-1}, i \in \{1,\dots,N\}\big\}. \tag{9}$$

The main assumptions of this work are the following:

A1. For all $k, q \in \mathbf{N}$, $|B_q| \perp\!\!\!\perp ((x_k^n, y_k^n))_{n\geq 1}$. In addition, for all $k \in \mathbf{N}$, $\big(|B_k|, ((x_k^n, y_k^n))_{n\geq 1}\big) \perp\!\!\!\perp \mathcal{F}_k^N$.

A2. The activation function $\sigma_* : \mathbf{R}^d \times \mathcal{X} \to \mathbf{R}$ belongs to $\mathcal{C}_b^\infty(\mathbf{R}^d \times \mathcal{X})$.

A3. For all $\ell \neq k \in \mathbf{N}$, $((x_\ell^n, y_\ell^n))_{n\geq 1} \perp\!\!\!\perp ((x_k^n, y_k^n))_{n\geq 1}$. In addition, for all $k \in \mathbf{N}$, $((x_k^n, y_k^n))_{n\geq 1}$ is a sequence of i.i.d random variables from $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, and $\mathbf{E}[|y|^{16\gamma_*}]$ is finite.

A4. The randomly initialized parameters $\{W_0^i\}_{i=1}^N$ are i.i.d. with a distribution $\mu_0 \in \mathcal{P}(\mathbf{R}^d)$ such that $\mathbf{E}[|W_0^1|^{8\gamma_*}] < +\infty$.

A5. For all $k \in \mathbf{N}$ and $i \in \{1,\dots,N\}$, $\varepsilon_k^i \sim \mathcal{N}(0, I_d)$ and $\varepsilon_k^i \perp\!\!\!\perp \mathcal{F}_k^N$. In addition, for all $k, l \in \mathbf{N}$ and $i, j \in \{1,\dots,N\}$ such that $(i,k) \neq (j,l)$, $\varepsilon_k^i \perp\!\!\!\perp \varepsilon_l^j$.

1.2.2 Law of large numbers for the empirical measure

Statement of the law of large numbers. The first main result of this work is a law of large numbers for the trajectory of the scaled empirical measures.

1. where we recall that the set of càdlàg (for *continue à droite, limite à gauche*) functions is the space of functions that are everywhere right-continuous with left limits everywhere.

Theorem 1 *Let $\beta > 1/2$ and assume* **A1**-**A5**. *Then, the sequence $(\mu^N)_{N\geq 1}$ converges in probability to a deterministic element $\bar\mu$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. In addition, $\bar\mu \in \mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$ and it is the unique solution in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$ of the following measure-valued equation:*

$$\forall f \in \mathcal{C}_b^\infty(\mathbf{R}^d), t \in \mathbf{R}_+,$$
$$\langle f, \bar\mu_t\rangle = \langle f, \mu_0\rangle + \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \bar\mu_s\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \bar\mu_s\rangle\, \pi(\mathrm{d}x, \mathrm{d}y)\, \mathrm{d}s. \quad (10)$$

Corollary 1 *Assume* **A1**-**A5**. *Then, $\bar\mu \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$. In addition, $\bar\mu$ satisfies also* (10) *for test functions $f \in \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$.*

Proof [Proof of Corollary 1] Note first that by **A4**, $\mu_0 \in \mathcal{C}^{0,\gamma}(\mathbf{R}^d)' \hookrightarrow \mathcal{H}^{-L,\gamma}(\mathbf{R}^d)$ according to (6). By (10), **A2**, and **A3**, it holds for all $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$ and $0 \leq s \leq t \leq T$,

$$|\langle f, \bar\mu_t\rangle - \langle f, \bar\mu_s\rangle| \leq C|t-s|\|f\|_{\mathcal{C}^{1,\gamma}} \sup_{u\in[0,T]} |\langle 1 + |\cdot|^\gamma, \bar\mu_u\rangle|.$$

Note that $\sup_{u\in[0,T]} |\langle 1 + |\cdot|^\gamma, \bar\mu_u\rangle| < +\infty$ since $t \geq 0 \mapsto \langle 1 + |\cdot|^\gamma, \bar\mu_u\rangle \in \mathcal{D}(\mathbf{R}_+, \mathbf{R})$ (indeed this follows from the fact that $\bar\mu \in \mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ and (Villani, 2009, Theorem 6.9)). Thus, using (6), it holds $\bar\mu_t \in \mathcal{H}^{-L,\gamma}(\mathbf{R}^d)$ and $|\langle f, \bar\mu_t\rangle - \langle f, \bar\mu_s\rangle| \leq C|t-s|\|f\|_{\mathcal{H}^{L,\gamma}}$, proving the first claim in Corollary 1. The second claim in Corollary 1 is obtained by a density argument and the fact that $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{1,\gamma}(\mathbf{R}^d)$. ■

On the proof of Theorem 1. Theorem 1 is proved in Section 2. The proof strategy is the following. We first derive an identity satisfied by $(\mu^N)_{N\geq1}$, namely the pre-limit equation (25). This is done in Section 2.1. Then, we show in Section 2.2.1 that $(\mu^N)_{N\geq1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. To this end we use (Jakubowski, 1986, Theorem 4.6). The compact containment of $(\mu^N)_{N\geq1}$ relies on a characterization of the compact subsets of $\mathcal{P}_{\gamma_0}(\mathbf{R}^{d+1})$ (see Proposition 11) and moment estimates on $\{\theta_k^i, i \in \{1, \dots N\}\}_{k=0,\dots,\lfloor NT\rfloor}$ (see Lemma 8). We then use the pre-limit equation (25) to prove that any limit point of the sequence $(\mu^N)_{N\geq1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ satisfies (10). This requires to study the continuity property of the involved operator (namely $\Lambda_t[f]$, see Lemma 17). This is the purpose of Section 2.2.3, and more precisely of Proposition 18 there. With rough estimates on the jumps of the function $t \in \mathbf{R}_+ \mapsto \langle f, \mu_t^N\rangle$ (where f is uniformly Lipschitz over $\mathbf{R}^{d+1}$), we also prove in Section 2.2.2 that any limit point $(\mu^N)_{N\geq1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ belongs a.s. to $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. This is indeed needed since we then prove in Section 2.3.1 that (10) admits a unique solution in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. To prove that there is at most one solution to (10), we use arguments of (Piccoli et al., 2015) which are based on a representation formula for solution to measure-valued equations (Villani, 2003, Theorem 5.34) together with time estimates in Wasserstein distances between two solutions of (10) derived in (Piccoli and Rossi, 2016).

Remark 2 *In view of their proofs, Theorem 1 and Corollary 1 are still valid for $\gamma > \frac{d}{2}$ and $L > d/2 + 1$.*

Remark 3 *When $\beta = 1/2$, one can obtain a similar limit equation for $\bar\mu$, with an additionnal (regularizing) Laplacian term in the limit equation. To derive it, one should consider a*

Taylor expansion up to order 3 of the test function in the pre-limit equation (25)*. Let us mention that the case $\beta = 1/2$ is studied in (Mei et al., 2018) but only at fixed t. Straightforward application of our method would lead to a trajectorial version of (Mei et al., 2018, Theorem 3) which we leave to the reader for the sake of brevity. We also refer to (Chizat and Bach, 2018, 2020) for a gradient flow approach for trajectorial results.*

Remark 4 *Of course, one important question is the convergence of $\bar\mu_t$ in long time. It is not hard to see that the loss function decays (but not strictly a priori) along the training, i.e. with t. This asymptotic behavior of $\bar\mu_t$ as $t \to +\infty$ has been studied in (Mei et al., 2018, Theorem 7) or (Chizat and Bach, 2018) who give partial results in the case without noise. Roughly speaking, they prove that if it is known that $\bar\mu_t$ is converging in Wasserstein distance then it converges to the minimum of the loss function. It is however quite hard to prove such a convergence. We refer also to (E, 2020; Ma et al., 2020) for what remains to do in this direction which is clearly a difficult open problem. In the case with noise $\beta = 1/2$ then the situation is different as the limit PDE is a usual McKean-Vlasov diffusion and one can study the free energy and study convergence in long time (Mei et al., 2018, Theorem 4).*

Remark 5 *Let us now discuss the necessity of* **A2***, and more precisely how to go beyond this assumption to consider the ReLU function. This function does not satisfy* **A2** *because it has a linear growth at $+\infty$ and secondly because it is not differentiable (only) at* 0*. The linear growth of the ReLU function can be handled with our techniques by adapting the moment estimates of Lemmata 8 and 9. On the other hand, dealing with the non-differentiability of the ReLU function at 0 is more subtle: first, to give a meaning to the SGD algorithm* (2)*, one has to choose a particular subgradient of σ_*. Secondly, all the arguments carry on, up until the uniqueness of the limiting equation (where we consider σ_* as a smooth test function). However, one can use an alternative proof of the uniqueness, given in (Descours, 2023), which is based on a probabilistic approach and Villani's argument for continuity equation (Villani, 2009). We also refer to (Wojtowytsch, 2020) for convergence results in the ReLU case.*

1.2.3 Central limit theorem for the empirical measure

Fluctuation process and extra assumptions. Assume **A1-A5**. The fluctuation process is the process $\eta^N = \{t \mapsto \eta^N_t, t \in \mathbf{R}_+\}$ defined by:

$$\eta^N_t = \sqrt{N}(\mu^N_t - \bar\mu_t),\ N \geq 1,\ t \in \mathbf{R}_+, \tag{11}$$

where $\bar\mu = \{t \mapsto \bar\mu_t, t \in \mathbf{R}_+\}$ is the limit of $(\mu^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ (see Theorem 1). Let us introduce the following additional assumptions:

A6. The distribution $\mu_0 \in \mathcal{P}(\mathbf{R}^d)$ is compactly supported.

A7. $|B_k| \to |B_\infty|$ a.s. as $k \to \infty$.

Let

$$J_0 \geq 4\lceil \frac{d}{2} \rceil + 8 \text{ and } j_0 = \lceil \frac{d}{2} \rceil + 2. \tag{12}$$

For later purpose, we also set

$$J_1 = 2\lceil \frac{d}{2} \rceil + 4,\ j_1 = 3\lceil \frac{d}{2} \rceil + 4,\ J_2 = 3\lceil \frac{d}{2} \rceil + 6, \text{ and } j_2 = 2\lceil \frac{d}{2} \rceil + 3. \tag{13}$$

By (3), we have the following embeddings:

$$\begin{aligned} &\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{J_2,j_2}(\mathbf{R}^d),\ \mathcal{H}^{J_2,j_2}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{J_1+1,j_1}(\mathbf{R}^d),\\ &\qquad\qquad\qquad \mathcal{H}^{J_1,j_1}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{L,\gamma}(\mathbf{R}^d). \end{aligned} \tag{14}$$

G-process and the limit equation.

Definition 1 *We say that* $\mathscr{G} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ *is a* G-process *if for all* $k \geq 1$ *and* $f_1 \ldots, f_k \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$, $\{t \mapsto (\langle f_1, \mathscr{G}_t\rangle, \ldots, \langle f_k, \mathscr{G}_t\rangle)^T, t \in \mathbf{R}_+\} \in \mathcal{C}(\mathbf{R}_+, \mathbf{R}^k)$ *is a process with zero-mean, independent Gaussian increments (and thus a martingale), and with covariance structure given by: for all* $1 \leq i, j \leq k$ *and all* $0 \leq s \leq t$,

$$\operatorname{Cov}\big(\langle f_i, \mathscr{G}_t\rangle, \langle f_j, \mathscr{G}_s\rangle\big) = \alpha^2 \mathbf{E}\left[\frac{1}{|B_\infty|}\right] \int_0^s \operatorname{Cov}(\mathrm{Q}_v[f_i](x,y), \mathrm{Q}_v[f_j](x,y))\, \mathrm{d}v, \tag{15}$$

where $\mathrm{Q}_v[f](x,y) := (y - \langle \sigma_*(\cdot, x), \bar\mu_v\rangle)\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \bar\mu_v\rangle$ *for* $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ *and* $\bar\mu$ *is given by Theorem 1.*

Let us make some comments about Definition 1. The first one is that we have decided to call such a process G-process to ease the statement of the results. In addition, notice that $\mathrm{Q}_s[f](x,y)$ is well defined for $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d))$ (indeed for all $k \in \{1, \ldots, d\}$, $\partial_{e_k} f \in \mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d) \hookrightarrow \mathcal{H}^{L,\gamma}(\mathbf{R}^d))$ and $\bar\mu \in \mathcal{C}(\mathbf{R}_+, \in \mathcal{H}^{-L,\gamma}(\mathbf{R}^d)))$. Finally, we mention that by Proposition 34 below, the law of a process $\mu \in \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J,b}(\mathbf{R}^d))$ is fully determined by the family of laws of the processes $(\langle f_1, \mu\rangle, \ldots, \langle f_k, \mu\rangle)^T \in \mathcal{D}(\mathbf{R}_+, \mathbf{R})^k$, $k \geq 1$ and where $\{f_a\}_{a\geq 1}$ is an orthonormal basis $\mathcal{H}^{J,b}(\mathbf{R}^d)$.

For η a $\mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$-valued process and $\mathscr{G} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ a G-process (see Definition 1), define the following equation:

$$\begin{aligned} &\text{A.s. } \forall f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d), \forall t \in \mathbf{R}_+,\\ &\langle f, \eta_t\rangle - \langle f, \eta_0\rangle = \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \bar\mu_s\rangle)\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \eta_s\rangle \pi(\mathrm{d}x, \mathrm{d}y)\\ &\qquad\qquad - \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle \sigma_*(\cdot, x), \eta_s\rangle\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \bar\mu_s\rangle \pi(\mathrm{d}x, \mathrm{d}y) + \langle f, \mathscr{G}_t\rangle. \end{aligned} \tag{16}$$

Definition 2 *Let* ν *be a* $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$*-valued random variable. We say that a* $\mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$*-valued process* η *on a probability space is a* weak solution *of* (16) *with initial distribution* ν *if there exist a* G*-process* $\mathscr{G} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ *such that* (16) *holds and* $\eta_0 = \nu$ *in distribution. In addition, we say that* weak uniqueness *holds if for any weak two solutions* η^1 *and* η^2 *of* (16) *(possibly defined on two different probability spaces) with the same initial distributions, it holds* $\eta^1 = \eta^2$ *in distribution.*

The second main result of this work is a central limit theorem for the trajectory of the scaled empirical measures.

Theorem 2 *Let $\beta > 3/4$. Assume* **A1**-**A7**. *Then:*

1. *(Convergence) The sequence $(\eta^N)_{N\geq 1} \subset \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$ (see* (11)*) converges in distribution to a process $\eta^* \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$.*

2. *(Limit equation) The process η^* has the same distribution as the unique weak solution $\eta^\star$ of* (16) *with initial distribution ν_0 (see Definition 2), where ν_0 is the unique $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$-valued random variable such that for all $k \geq 1$ and $f_1 \dots, f_k \in \mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$,*
$$(\langle f_1, \nu_0\rangle, \dots, \langle f_k, \nu_0\rangle)^T \sim \mathcal{N}(0, \Gamma(f_1, \dots, f_k)),$$
where $\Gamma(f_1, \dots, f_k)$ is the covariance matrix of the vector $(f_1(W_0^1), \dots, f_k(W_0^1))^T$.

Remark 6 *By looking at the definition of the* G*-process and in particular its covariance* (15)*, one remarks the effect of mini-batching by the $|B_\infty|^{-1}$ prefactor, thus leading to a reduced variance of the* G*-process. Note that this is quite intricate to deduce proper information on the variance of the fluctuation process η, since the terms appearing in* (16) *are a priori dependent. Nonetheless, it will be shown through the numerical experiments of the next subsection that the variance of fluctuation process reduces when the size of the mini-batches increases (see in particular Figure 1).*

Theorem 2 is proved in Section 3, following inspiration from the previous works (Fernandez and Méléard, 1997; Jourdain and Méléard, 1998; Delarue et al., 2019). The starting point to prove Theorem 2, like in the current literature (Sirignano and Spiliopoulos, 2020a), consists in proving that $(\eta^N)_{N\geq 1} \subset \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ is relatively compact (see Propositions 24). We then prove that the whole sequence $(\eta^N)_{N\geq 1}$ converges in distribution to the unique weak solution of (16) in Section 3.5.

When $\beta = 3/4$, $(\eta^N)_{N\geq 1}$ is still relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$ (see Proposition 24) but the derivation of the limit equation satisfied by its limit points is more tricky. However, in a specific case (when $d = 1$ and the test function is $\mathsf{f}_2 : x \in \mathbf{R} \mapsto |x|^2$), Proposition 7 below suggests how the equation (16) might be perturbed, as shown numerically in Figure 2 and more precisely in the inset.

Proposition 7 *Let $\beta = 3/4$ and assume that conditions* **A1**-**A7** *hold. Let η be a limit point of $(\eta^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}))$ (see Proposition 24). Then, $\eta_0 = \nu_0$ in distribution (see Lemma 35), and there exist a $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$-valued process η^* and a* G*-process $\mathscr{G}^* \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}))$ such that $\eta = \eta_*$ in distribution, and a.s. for every $t \in \mathbf{R}_+$,*

$$\begin{aligned}\langle \mathsf{f}_2, \eta^*_t\rangle - \langle \mathsf{f}_2, \eta^*_0\rangle &= \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \overline{\mu}_s\rangle)\langle \nabla \mathsf{f}_2 \cdot \nabla\sigma_*(\cdot, x), \eta^*_s\rangle \pi(\mathrm{d}x, \mathrm{d}y) \\ &\quad - \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle \sigma_*(\cdot, x), \eta^*_s\rangle\langle \nabla \mathsf{f}_2 \cdot \nabla\sigma_*(\cdot, x), \overline{\mu}_s\rangle \pi(\mathrm{d}x, \mathrm{d}y) + \langle \mathsf{f}_2, \mathscr{G}^*_t\rangle \\ &\quad + t\mathbf{E}[\mathsf{f}_2(\varepsilon^1_1)]. \end{aligned} \tag{17}$$

1.3 Numerical Experiments

We now illustrate numerically the results derived in the previous sections. First, we consider a regression task on simulated data, based upon an example of (Mei et al., 2018). More precisely, we consider (1) with $\sigma_*(W^i, x) = f(W^i \cdot x)$ where

$$f(t) = \begin{cases} -2.5 \quad \text{if} \quad t \leq 0.5, \\ 10t - 7.5 \quad \text{if} \quad 0.5 \leq t \leq 1.5, \\ 7.5 \quad \text{if} \quad t \geq 1.5. \end{cases}$$

The distribution π of the data is defined as follows: with probability 1/2, $y = 1$ and $x \sim \mathcal{N}(0, (1+0.2)^2 I_d)$ and, with probability 1/2, $y = -1$ and $x \sim \mathcal{N}(0, (1-0.2)^2 I_d)$. This setting satisfies the assumptions of Theorems 1 and 2, except **A2**, due to the fact that f is not differentiable at $t = 0.5$ and $t = 1.5$ (see also Remark 5).

Then, we consider a typical classification task on the MNIST dataset. The neural network we consider is fully connected with one-hidden layer of N neurons and ReLU activation function[2]. The last layer is a softmax layer (we consider one-hot encoding and use Keras and Tensorflow librairies). Given a data $x \in \mathbf{R}^d$ ($d = 784$ here), the neural network returns $\hat{y} = \text{softmax}((W^{\mathbf{o},c} \cdot W^{\mathbf{h}}(x))_{c=0}^9)$ where $W^{\mathbf{h}}(x) = ((W^{\mathbf{h},i} \cdot x)_+)_{i=1}^N$ is the hidden layer ($W^{\mathbf{h},i} \in \mathbf{R}^d$ is the weight of the i-th neuron) and $W^{\mathbf{o},c} \in \mathbf{R}^N$ is the weight of the output layer corresponding to class c. The total number of trainable parameters is thus $dN + 10N$. The neural network is trained with respect to the categorical cross-entropy loss. This case is not covered by our mathematical analysis and the motivation here is to show numerical evidence that the variance reduction derived in Theorem 2 is still valid in this case.

Variance Reduction with increasing mini-batch size. We illustrate here that the variance of the limiting fluctuation process decreases with the mini-batch size, even though we only have a mathematical structure of the variance of the G-process (see (15) together with Remark 6). On both experiments, we consider a fixed mini-batch size during the training (i.e. $|B_k| = |B|$ for all $k \in \mathbf{N}$). We first consider the regression task. Consider $\mathsf{L} = 1000$ neural networks (initialized and trained independently) whose $N = 800$ initial neurons are drawn independently according to $\mu_0 = \mathcal{N}(0, \frac{0.8^2}{d} I_d)$. For each neural network, we run $k = 1000$ iterations of the SGD algorithm (2) and compute $\mathsf{m}_\ell := \langle \| \cdot \|_2, \mu_t^N \rangle = \frac{1}{N} \sum_{i=1}^N \|W_k^i\|_2$, where $\ell \in \{1, \dots, \mathsf{L}\}$, $t = k/N = 1.25$ and $\|w\|_2 := \sqrt{\sum_{j=1}^d w_j^2}$. Finally, we compute the empirical variance of this quantity, i.e.,

$$\mathsf{V} := \widehat{\text{Var}}(\mathsf{m}_1, \dots, \mathsf{m}_\mathsf{L}) = \frac{1}{\mathsf{L}} \sum_{\ell=1}^{\mathsf{L}} \Big(\mathsf{m}_\ell - \frac{1}{\mathsf{L}} \sum_{\ell'=1}^{\mathsf{L}} \mathsf{m}_{\ell'} \Big)^2.$$

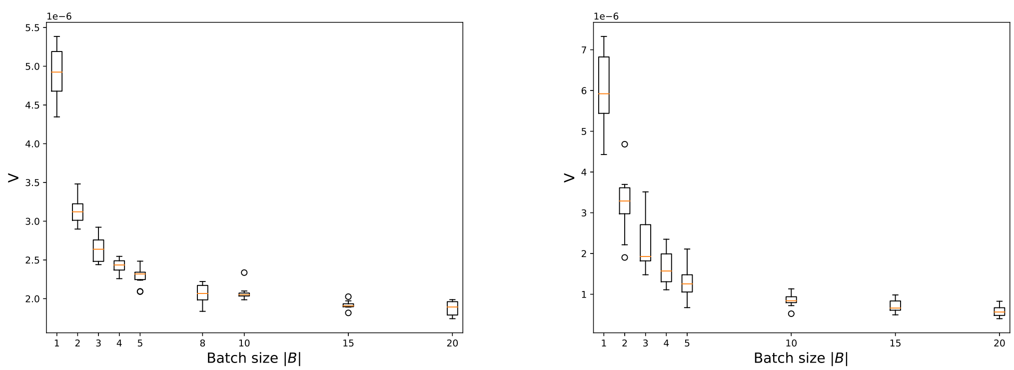
and display for different mini-batch sizes $|B|$ in Figure 1 the obtained boxplots from 10 samples of V. The other parameters are $d = 40$, $\alpha = 0.1$, $\beta = 1$, and the noise is $\varepsilon_k^i \sim \mathcal{N}(0, 0.01 I_d)$.

Second, we turn to the classification task. Consider $\mathsf{L} = 30$ neural networks (initialized and trained independently) with $N = 10000$ neurons on the hidden-layer, initially drawn

2. ReLU function $(\cdot)_+ \colon u \in \mathbf{R} \mapsto 0$ if $u < 0$, u if $u \geq 0$.

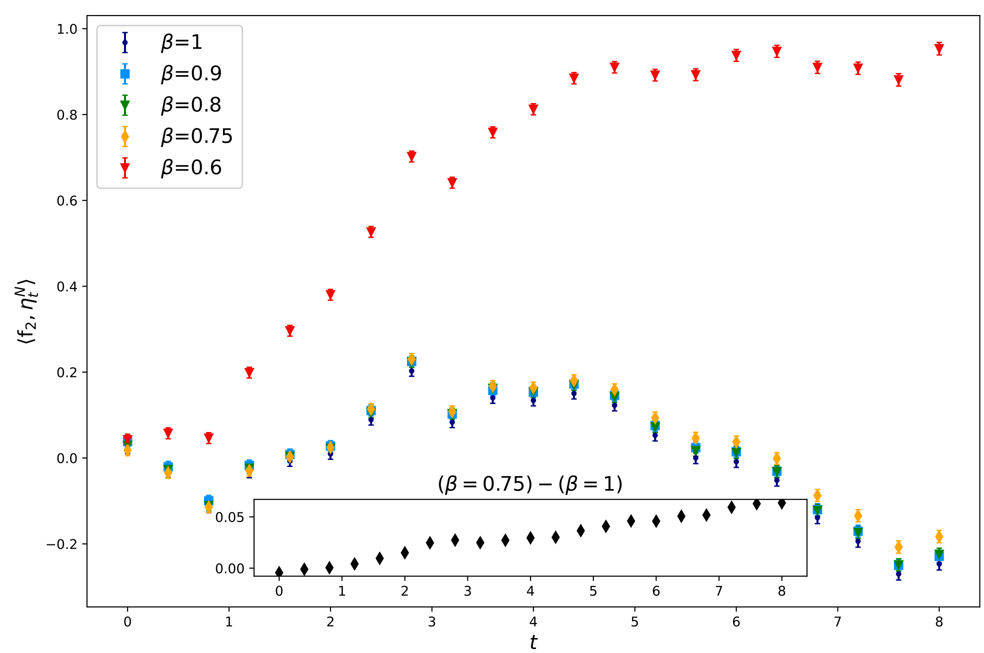
according to the He intialization $\mathcal{N}(0, \frac{2}{d} I_d)$, until iteration $k = 3000$ of the SGD algorithm ($t = k/N = 0.3$), and compute the mean of the weight of the output layer (each $W^{\mathbf{o},\mathbf{c}}$ is initially drawn according to the Xavier initialization $\mathcal{N}(0, \frac{2}{N+10} I_N)$) corresponding to class 0, *i.e.*, for each $\ell = 1, \ldots, \mathsf{L}$, we compute $\mathsf{m}_\ell := \frac{1}{N} \sum_{j=1}^N W_k^{\mathbf{o},0,j}$. Finally, we compute the empirical variance of this quantity, i.e., $\mathsf{V} = \widehat{\mathrm{Var}}(\mathsf{m}_1, \ldots, \mathsf{m}_\mathsf{L})$ and exhibit for different sizes $|B|$ the boxplots obtained with 10 samples of V in Figure 1.

Figure 1: Variance V reduction of the fluctuation process with increasing mini-batch size. **Left:** Regression task on simulated data. V is an empirical estimation from 1000 realisations of the variance of $\langle \| \cdot \|_2, \mu_t^N \rangle$, where $N = 800$ and $t = 1.25$. The other parameters are $d = 40$, $\alpha = 0.1$, $\beta = 1$, and the noise is $\varepsilon_k^i \sim \mathcal{N}(0, 0.01 I_d)$. The boxplots are obtained with 10 samples of V. **Right:** Classification task on MNIST dataset. V is an empirical estimation from 30 realisations of the variance of $\frac{1}{N} \sum_{j=1}^N W_k^{\mathbf{o},0,j}$, where $N = 10000$ and $k = 3000$ ($t = 0.3$). The boxplots are obtained with 10 samples of V.



Central Limit Theorem. We focus here on the regression task. For different values of β, we plot in Figure 2 $\langle \mathsf{f}_2, \eta_t^N \rangle$ for $0 \le t \le 8$ (recall $\mathsf{f}_2(x) = |x|^2$), to show the agreement of $\langle \mathsf{f}_2, \eta_t^N \rangle$ for different values of $\beta > 3/4$, corresponding to the regime of (16), and the divergence from it when $\beta \le 3/4$. For $\beta = 3/4$, we also illustrate the regime derived in Proposition 7. The parameters chosen are $d = |B| = 1$, $N = 20000$, $\alpha = 0.1$ and $\varepsilon_k^i \sim \mathcal{N}(0, 0.01)$. The procedure to obtain the plots is as follows. We first compute $\langle \mathsf{f}_2, \mu_t^N \rangle$ (we repeat this procedure 20000 times to get confidence intervals). Then, we approximate $\langle \mathsf{f}_2, \bar{\mu}_t \rangle$ by $\langle \mathsf{f}_2, \mu_t^{N'} \rangle$ where $N' = 250000$. On Figure 2, we plot $\sqrt{N}(\langle \mathsf{f}_2, \mu_t^N \rangle - \langle \mathsf{f}_2, \mu_t^{N'} \rangle) \simeq \langle \mathsf{f}_2, \eta_t^N \rangle$ as a function of t.

Figure 2: Time evolution of the fluctuation process for different values of β on the regression task, with $\mathsf{f}_2 : x \in \mathbf{R} \mapsto |x|^2$, $N = 20000$, $d = |B| = 1$, $\alpha = 0.1$ and $\varepsilon_k^i \sim \mathcal{N}(0, 0.01)$. Confidence intervals are obtained from 20000 realisations. The case $\beta > 3/4$ is driven by (16). The case $\beta = 3/4$ is driven by (17). The case $\beta < 3/4$ is not covered by our analysis. The inset exhibits the linear term in time appearing in (17).



2. Proof of Theorem 1

2.1 Pre-limit equation and remainder terms

In this section, we derive the so-called pre-limit equation (25). We then show that the remainder terms in this equation are negligible as $N \to +\infty$.

2.1.1 Pre-limit equation

In this section, we introduce several (random) operators acting on $\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. Using **A2** and **A3**, it is easy to check that all these operators belong a.s. to the dual of $\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. The duality bracket we use in this section then is the one for the duality in $\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. Let us consider $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. The Taylor-Lagrange formula yields, for $N \geq 1$ and $k \in \mathbf{N}$,

$$\begin{aligned}\langle f, \nu_{k+1}^N\rangle - \langle f, \nu_k^N\rangle &= \frac{1}{N}\sum_{i=1}^N f(W_{k+1}^i) - f(W_k^i)\\ &= \frac{1}{N}\sum_{i=1}^N \nabla f(W_k^i)\cdot(W_{k+1}^i - W_k^i)\\ &\quad + \frac{1}{2N}\sum_{i=1}^N (W_{k+1}^i - W_k^i)^T \nabla^2 f(\widehat{W}_k^i)(W_{k+1}^i - W_k^i),\end{aligned}$$

where, for all $i \in \{1, \dots, N\}$, $\widehat{W}_k^i \in (W_k^i, W_{k+1}^i)$. Using (2), we have

$$\begin{aligned}\langle f, \nu_{k+1}^N\rangle - \langle f, \nu_k^N\rangle &= \frac{1}{N}\sum_{i=1}^N \nabla f(W_k^i)\cdot\left[\frac{\alpha}{N|B_k|}\sum_{(x,y)\in B_k}(y - g_{W^k}^N(x))\nabla_W \sigma_*(W_k^i, x) + \frac{\varepsilon_k^i}{N^\beta}\right]\\ &\quad + \langle f, R_k^N\rangle\\ &= \frac{\alpha}{N|B_k|}\sum_{(x,y)\in B_k}(y - \langle\sigma_*(\cdot,x),\nu_k^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\nu_k^N\rangle\\ &\quad + \frac{1}{N^{1+\beta}}\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i + \langle f, R_k^N\rangle,\end{aligned} \tag{18}$$

where, for $N \geq 1$, $k \in \mathbf{N}$ and $i = 1, \dots, N$,

$$\langle f, R_k^N\rangle := \frac{1}{2N}\sum_{i=1}^N (W_{k+1}^i - W_k^i)^T \nabla^2 f(\widehat{W}_k^i)(W_{k+1}^i - W_k^i). \tag{19}$$

For $k \in \mathbf{N}$, we define:

$$\langle f, D_k^N\rangle := \frac{\alpha}{N}\int_{\mathcal{X}\times\mathcal{Y}}(y - \langle\sigma_*(\cdot,x),\nu_k^N\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\nu_k^N\rangle\pi(\mathrm{d}x,\mathrm{d}y), \tag{20}$$

$$\langle f, M_k^N\rangle := \frac{\alpha}{N|B_k|}\sum_{(x,y)\in B_k}(y - \langle\sigma_*(\cdot,x),\nu_k^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\nu_k^N\rangle - \langle f, D_k^N\rangle. \tag{21}$$

Equation (18) writes, for $k \in \mathbf{N}$,

$$\langle f, \nu_{k+1}^N \rangle - \langle f, \nu_k^N \rangle = \langle f, D_k^N \rangle + \langle f, M_k^N \rangle + \langle f, R_k^N \rangle + \frac{1}{N^{1+\beta}} \sum_{i=1}^N \nabla f(W_k^i) \cdot \varepsilon_k^i. \tag{22}$$

Define for $N \geq 1$ and $t \in \mathbf{R}_+$:

$$\langle f, D_t^N \rangle := \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, D_k^N \rangle \quad \text{and} \quad \langle f, M_t^N \rangle := \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, M_k^N \rangle, \tag{23}$$

with the convention that $\sum_0^{-1} = 0$ (which occurs if and only if $0 \leq t < 1/N$). It will be proved later that $\{t \mapsto \langle f, M_t^N \rangle, t \in \mathbf{R}_+\}$ is a martingale (see indeed Lemma 21), hence the notation. One has, for $t \in \mathbf{R}_+$,

$$\begin{aligned}
\langle f, D_t^N \rangle &= \sum_{k=0}^{\lfloor Nt \rfloor - 1} \int_{\frac{k}{N}}^{\frac{k+1}{N}} \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \nu_k^N \rangle) \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \nu_k^N \rangle \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s \\
&= \sum_{k=0}^{\lfloor Nt \rfloor - 1} \int_{\frac{k}{N}}^{\frac{k+1}{N}} \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N \rangle) \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \mu_s^N \rangle \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s \\
&= \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N \rangle) \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \mu_s^N \rangle \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s + \langle f, V_t^N \rangle,
\end{aligned}$$

where $\langle f, V_t^N \rangle$, for $t \in \mathbf{R}_+$:

$$\langle f, V_t^N \rangle := - \int_{\frac{\lfloor Nt \rfloor}{N}}^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N \rangle) \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \mu_s^N \rangle \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s. \tag{24}$$

Therefore, using (22), we obtain that the scaled empirical measure process $\{t \mapsto \mu_t^N, t \in \mathbf{R}_+\}$ satisfies the following pre-limit equation : for $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$, $N \geq 1$ and $t \in \mathbf{R}_+$,

$$\begin{aligned}
\langle f, \mu_t^N \rangle - \langle f, \mu_0^N \rangle &= \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, \nu_{k+1}^N \rangle - \langle f, \nu_k^N \rangle \\
&= \langle f, D_t^N \rangle + \langle f, M_t^N \rangle + \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, R_k^N \rangle + \frac{1}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^N \nabla f(W_k^i) \cdot \varepsilon_k^i \\
&= \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N \rangle) \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \mu_s^N \rangle \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s \\
&\quad + \langle f, M_t^N \rangle + \langle f, V_t^N \rangle + \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, R_k^N \rangle + \frac{1}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^N \nabla f(W_k^i) \cdot \varepsilon_k^i.
\end{aligned} \tag{25}$$

In the next section, we study the four last terms of (25).

2.1.2 The remainder terms in (25) are negligible

The aim of this section is to show that the last four terms of (25) vanish as $N \to +\infty$. This is the purpose of Lemma 9. The following result will be used several times in this work.

Lemma 8 *Let $\beta \geq 1/2$ and assume* **A1**-**A5**. *Then, for all $T > 0$, there exists a constant $C < +\infty$ such that for all $N \geq 1$, $i \in \{1, \ldots, N\}$ and $k \in \{0, \ldots, \lfloor NT \rfloor\}$,*

$$\mathbf{E}\left[|W_k^i|^{8\gamma_*}\right] \leq C.$$

Proof Let us recall the following convexity inequality : for $m, p \geq 1$ and $x_1, \ldots, x_p \in \mathbf{R}_+$,

$$\Big(\sum_{l=1}^{m} x_l\Big)^p \leq m^{p-1} \sum_{l=1}^{m} x_l^p. \tag{26}$$

Let $C > 0$ denotes a constant, independent of $i \in \{1, \ldots, N\}$ and $0 \leq k \leq \lfloor NT \rfloor$, which can change from one occurence to another. Set $p = 8\gamma_*$. For $i \in \{1, \ldots, N\}$ and $1 \leq k \leq \lfloor NT \rfloor$, we have, using (2) and **A2** :

$$|W_k^i| \leq |W_0^i| + \Big|\sum_{j=0}^{k-1} W_{j+1}^i - W_j^i\Big| \leq |W_0^i| + \frac{C}{N}\sum_{j=0}^{k-1}\frac{1}{|B_j|}\sum_{(x,y)\in B_j}(|y| + C) + \frac{1}{N^\beta}\Big|\sum_{j=0}^{k-1}\varepsilon_j^i\Big|.$$

Thus, by (26),

$$\begin{aligned}|W_k^i|^p &\leq C\Big[|W_0^i|^p + \frac{1}{N}\sum_{j=0}^{k-1}\frac{1}{|B_j|^p}\Big(\sum_{(x,y)\in B_j}(|y| + C)\Big)^p + \frac{1}{N^{p\beta}}\Big|\sum_{j=0}^{k-1}\varepsilon_j^i\Big|^p\Big]\\ &\leq C\Big[|W_0^i|^p + \frac{1}{N}\sum_{j=0}^{k-1}\frac{1}{|B_j|}\sum_{(x,y)\in B_j}(|y| + C)^p + \frac{1}{N^{p\beta}}\Big|\sum_{j=0}^{k-1}\varepsilon_j^i\Big|^p\Big].\end{aligned}$$

We have:

$$\mathbf{E}\Big[\frac{1}{|B_j|}\sum_{(x,y)\in B_j}(|y| + C)^p\Big] \leq C\Big[\mathbf{E}\Big[\frac{1}{|B_j|}\sum_{n=1}^{|B_j|}|y_j^n|^p\Big] + 1\Big],$$

and, using **A1** and (**A3**), it holds for $j \geq 0$:

$$\begin{aligned}\mathbf{E}\Big[\frac{1}{|B_j|}\sum_{n=1}^{|B_j|}|y_j^n|^p\Big] = \sum_{q=1}^{+\infty}\mathbf{E}\Big[\frac{\mathbf{1}_{|B_j|=q}}{q}\sum_{n=1}^{q}|y_j^n|^p\Big] &= \sum_{q=1}^{+\infty}\frac{1}{q}\sum_{n=1}^{q}\mathbf{E}\big[|y_j^n|^p\mathbf{1}_{|B_j|=q}\big]\\ &= \sum_{q=1}^{+\infty}\frac{1}{q}\sum_{n=1}^{q}\mathbf{E}\big[|y_j^n|^p\big]\,\mathbf{E}\big[\mathbf{1}_{|B_j|=q}\big]\\ &= \mathbf{E}\big[|y_1^1|^p\big] < +\infty.\end{aligned} \tag{27}$$

Thus, using the two previous inequalities, we deduce that:

$$\mathbf{E}\Big[\frac{1}{N}\sum_{j=0}^{k-1}\frac{1}{|B_j|}\sum_{(x,y)\in B_j}(|y|+C)^p\Big]\le C.$$

By **A4**, $\mathbf{E}\left[|W_0^i|^p\right]\le C$. In addition, we have that, for $i\in\{1,\dots,N\}$,

$$\Big|\sum_{j=0}^{k-1}\varepsilon_j^i\Big|^p\le C\sum_{l=1}^{d}\Big|\sum_{j=0}^{k-1}\varepsilon_j^{i,l}\Big|^p$$

Since we deal with the sum of centered independent Gaussian random variables, we have that, for all $i\in\{1,\dots,N\}$ and $l\in\{1,\dots,d\}$,

$$\mathbf{E}\Big[\Big|\sum_{j=0}^{k-1}\varepsilon_j^{i,l}\Big|^p\Big]\le Ck^{p/2}\le CN^{p/2}.$$

Putting all these inequalities together, we obtain that $\mathbf{E}\left[|W_k^i|^p\right]\le C\left[1+\frac{N^{p/2}}{N^{p\beta}}\right]\le C$ (recall $\beta\ge 1/2$). This concludes the proof of the lemma. ■

Lemma 9 *Let* $\beta\ge 1/2$ *and assume* **A1**-**A5**. *Then, for all* $T>0$ *there exists* $C<\infty$ *such that for all* $N\ge 1$ *and* $f\in\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$,

(i) $\max_{0\le k<\lfloor NT\rfloor}\mathbf{E}\left[|\langle f,R_k^N\rangle|\right]\le C\|f\|_{\mathcal{C}^{2,\gamma_*}}\left[\frac{1}{N^2}+\frac{1}{N^{2\beta}}\right]$.

(ii) $\sup_{t\in[0,T]}\mathbf{E}\left[|\langle f,V_t^N\rangle|\right]\le C\|f\|_{\mathcal{C}^{2,\gamma_*}}/N$.

(iii) $\sup_{t\in[0,T]}\mathbf{E}\left[|\langle f,M_t^N\rangle|^2\right]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}/N$.

(iv) $\sup_{t\in[0,T]}\mathbf{E}\left[\left|\frac{1}{N^{1+\beta}}\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^N\nabla f(W_k^i)\cdot\varepsilon_k^i\right|^2\right]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}/N^{2\beta}$.

Proof Let $T>0$ and $f\in\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. In what follows, $C>0$ is a constant, independent of $N\ge 1$, $t\in[0,T]$, f, and $k\in\{0,\dots,\lfloor NT\rfloor-1\}$, which can change from one line to another.

Proof of item *(i)*. For $k\in\{0,\dots,\lfloor NT\rfloor-1\}$, by (19), we have

$$|\langle f,R_k^N\rangle|\le\frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N|W_{k+1}^i-W_k^i|^2(1+|\widehat{W}_k^i|^{\gamma_*}).\tag{28}$$

On the other hand, by (2), we have:

$$|W_{k+1}^i-W_k^i|\le\frac{C}{N|B_k|}\sum_{(x,y)\in B_k}\big(|y|+|g_{W_k}^N(x)|\big)+\frac{|\varepsilon_k^i|}{N^\beta}.\tag{29}$$

By (26) and the triangle inequality, we deduce

$$|W^i_{k+1} - W^i_k|^2 \le C\Big[\frac{1}{N^2|B_k|}\sum_{(x,y)\in B_k}(|y|^2 + |g^N_{W_k}(x)|^2) + \frac{|\varepsilon^i_k|^2}{N^{2\beta}}\Big].$$

By definition of $\widehat{W}^i_k$, there exists $\alpha^i_k \in (0,1)$ such that $\widehat{W}^i_k = \alpha^i_k W^i_k + (1-\alpha^i_k)W^i_{k+1}$, leading, by (26), to $|\widehat{W}^i_k|^{\gamma_*} \le C\left[|W^i_k|^{\gamma_*} + |W^i_{k+1}|^{\gamma_*}\right]$. Therefore,

$$\begin{aligned}
&|W^i_{k+1} - W^i_k|^2(1 + |\widehat{W}^i_k|^{\gamma_*})\\
&\le C\Big[\frac{1}{N^2|B_k|}\sum_{(x,y)\in B_k}(|y|^2 + |g^N_{W_k}(x)|^2) + \frac{|\varepsilon^i_k|^2}{N^{2\beta}}\Big](1 + |W^i_k|^{\gamma_*} + |W^i_{k+1}|^{\gamma_*})\\
&\le \frac{C}{N^2}(1 + |W^i_k|^{\gamma_*} + |W^i_{k+1}|^{\gamma_*})^2 + \frac{C}{N^2|B_k|}\sum_{(x,y)\in B_k}(|y|^4 + |g_{W_k}(x)|^4)\\
&\quad + \frac{C}{N^{2\beta}}\left[|\varepsilon^i_k|^4 + (1 + |W^i_k|^{\gamma_*} + |W^i_{k+1}|^{\gamma_*})^2\right]. \qquad (30)
\end{aligned}$$

Plugging (30) in (28), we obtain

$$\begin{aligned}
|\langle f, R^N_k\rangle| \le \frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N\Big[&\frac{1}{N^2}(1 + |W^i_k|^{\gamma_*} + |W^i_{k+1}|^{\gamma_*})^2 + \frac{1}{N^2|B_k|}\sum_{(x,y)\in B_k}(|y|^4 + C)\\
&+ \frac{1}{N^{2\beta}}\left[|\varepsilon^i_k|^4 + (1 + |W^i_k|^{\gamma_*} + |W^i_{k+1}|^{\gamma_*})^2\right]\Big]. \qquad (31)
\end{aligned}$$

Finally, using Lemma 8, **A3**, and **A5**, one deduces that $\mathbf{E}\left[|\langle f, R^N_k\rangle|\right] \le C\|f\|_{\mathcal{C}^{2,\gamma_*}}(1/N^2 + 1/N^{2\beta})$. This proves item *(i)*.

Proof of item *(ii)*. Let $t\in[0,T]$. Since σ_* and all its derivatives are bounded (see **A2**), it holds for all $s\ge 0$:

$$|\langle \sigma_*(\cdot,x), \mu^N_s\rangle| = \Big|\frac{1}{N}\sum_{i=1}^N \sigma_*(W^i_{\lfloor Ns\rfloor}, x)\Big| \le C, \qquad (32)$$

and

$$\begin{aligned}
|\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \mu^N_s\rangle| &= \Big|\frac{1}{N}\sum_{i=1}^N \nabla f(W^i_{\lfloor Ns\rfloor})\cdot\nabla_W\sigma_*(W^i_{\lfloor Ns\rfloor}, x)\Big|\\
&\le \frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N(1 + |W^i_{\lfloor Ns\rfloor}|^{\gamma_*}). \qquad (33)
\end{aligned}$$

Notice that C above is also independent of $x\in\mathcal{X}$. Since $\mathbf{E}[|y|] < +\infty$ (see (**A3**)), we obtain

$$\Big|\int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle\sigma_*(\cdot,x), \mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x), \mu^N_s\rangle\pi(\mathrm{d}x,\mathrm{d}y)\Big| \le \frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N(1 + |W^i_{\lfloor Ns\rfloor}|^{\gamma_*}). \qquad (34)$$

Noticing that $s \in (\frac{\lfloor Nt \rfloor}{N}, t) \Rightarrow \lfloor Ns \rfloor = \lfloor Nt \rfloor$, we obtain (see (24))

$$|\langle f, V_t^N \rangle| \le \Big(t - \frac{\lfloor Nt \rfloor}{N} \Big) \frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N} \sum_{i=1}^N (1 + |W^i_{\lfloor Nt \rfloor}|^{\gamma_*}). \tag{35}$$

Then, by Lemma 8, $\mathbf{E}\left[|\langle f, V_t^N \rangle|\right] \le C\|f\|_{\mathcal{C}^{2,\gamma_*}}/N$. This proves item *(ii)*.

Proof of item *(iii)*. Let $t \in [0, T]$. Recall the definition of $\mathcal{F}_k^N$ in (9) and $\langle f, M_k^N \rangle$ in (21).

Step 1. In this step we prove that

$$\mathbf{E}\left[|\langle f, M_k^N \rangle|^2\right] \le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}/N^2. \tag{36}$$

With the same arguments as those used to get (32) and (33), we have

$$|\langle \sigma_*(\cdot, x), \nu_k^N \rangle| \le C \quad \text{and} \quad |\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \nu_k^N \rangle| \le \frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N} \sum_{i=1}^N (1 + |W_k^i|^{\gamma_*}). \tag{37}$$

Note that C above is also independent of $x \in \mathcal{X}$. By (26) and (37), we have:

$$\begin{aligned}
|\langle f, M_k^N \rangle|^2 &\le \frac{C}{N^2|B_k|} \sum_{(x,y)\in B_k} (y - \langle \sigma_*(\cdot, x), \nu_k^N \rangle)^2 \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \nu_k^N \rangle^2 + C|\langle f, D_k^N \rangle|^2 \\
&\le \frac{C}{N^2|B_k|} \sum_{(x,y)\in B_k} (|y|^2 + C) \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \nu_k^N \rangle^2 + C|\langle f, D_k^N \rangle|^2 \\
&\le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^3|B_k|} \sum_{(x,y)\in B_k} (|y|^2 + C) \sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2 + C|\langle f, D_k^N \rangle|^2.
\end{aligned} \tag{38}$$

On the other hand, it holds since $(W_k^1, \ldots, W_k^N)$ is $\mathcal{F}_k^N$-measurable and by **A1**:

$$\begin{aligned}
&\mathbf{E}\Big[\frac{1}{|B_k|} \sum_{(x,y)\in B_k} (|y|^2 + C) \sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2\Big] \\
&= \sum_{q\ge1} \frac{1}{q} \mathbf{E}\Big[\mathbf{1}_{|B_k|=q} \sum_{n=1}^q (|y_k^n|^2 + C) \sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2\Big] \\
&= \sum_{q\ge1} \frac{1}{q} \mathbf{E}\Big[\mathbf{1}_{|B_k|=q} \sum_{n=1}^q (|y_k^n|^2 + C)\Big] \mathbf{E}\Big[\sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2\Big] \\
&= \sum_{q\ge1} \frac{1}{q} \mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\Big] \mathbf{E}\Big[\sum_{n=1}^q (|y_k^n|^2 + C)\Big] \mathbf{E}\Big[\sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2\Big] \\
&= \sum_{q\ge1} \mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\Big] \mathbf{E}\Big[|y_1^1|^2 + C\Big] \mathbf{E}\Big[\sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2\Big] \\
&= \mathbf{E}\Big[|y_1^1|^2 + C\Big] \mathbf{E}\Big[\sum_{i=1}^N (1 + |W_k^i|^{\gamma_*})^2\Big] \le CN,
\end{aligned} \tag{39}$$

where we have used Lemma 8 and **A3** for the last inequality. Consequently, one has:

$$\mathbf{E}\left[|\langle f, M_k^N\rangle|^2\right] \leq \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^2} + C\mathbf{E}\left[\langle f, D_k^N\rangle^2\right]. \tag{40}$$

On the other hand, we easily obtain with similar arguments that

$$\mathbf{E}\left[|\langle f, D_k^N\rangle|^2\right] \leq C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}/N^2.$$

Together with (40), this ends the proof of (36).

Step 2. In this step we prove that for all $k \geq 0$:

$$\mathbf{E}\left[\langle f, M_k^N\rangle|\mathcal{F}_k^N\right] = 0. \tag{41}$$

For ease of notation, we set

$$\mathrm{Q}^N[f](x, y, \{W_k^i\}_{i=1,\dots,N}) = \big(y - \langle \sigma_*(\cdot, x), \nu_k^N\rangle\big)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \nu_k^N\rangle. \tag{42}$$

With this notation, we have (see (20)) $\langle f, D_k^N\rangle = \frac{\alpha}{N}\int_{\mathcal{X}\times\mathcal{Y}} \mathrm{Q}^N[f](x, y, \{W_k^i\}_{i=1,\dots,N})\pi(\mathrm{d}x, \mathrm{d}y)$ and $\langle f, M_k^N\rangle = \frac{\alpha}{N|B_k|}\sum_{(x,y)\in B_k} \mathrm{Q}^N[f](x, y, \{W_k^i\}_{i=1,\dots,N}) - \langle f, D_k^N\rangle$. It then holds:

$$\begin{aligned}
&\mathbf{E}\Big[\frac{1}{|B_k|}\sum_{(x,y)\in B_k} \mathrm{Q}^N[f](x, y, \{W_k^i\}_{i=1,\dots,N}) \,\big|\, \mathcal{F}_k^N\Big] \\
&= \mathbf{E}\Big[\frac{1}{|B_k|}\sum_{n=1}^{|B_k|} \mathrm{Q}^N[f](x_k^n, y_k^n, \{W_k^i\}_{i=1,\dots,N}) \,\big|\, \mathcal{F}_k^N\Big] \\
&= \sum_{q\geq 1}\frac{1}{q}\mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\sum_{n=1}^{q} \mathrm{Q}^N[f](x_k^n, y_k^n, \{W_k^i\}_{i=1,\dots,N}) \,\big|\, \mathcal{F}_k^N\Big].
\end{aligned}$$

Since $(W_k^1, \dots, W_k^N)$ is $\mathcal{F}_k^N$-measurable, $\big(|B_k|, ((x_k^n, y_k^n))_{n\geq 1}\big) \perp\!\!\!\perp \mathcal{F}_k^N$, and $|B_k| \perp\!\!\!\perp ((x_k^n, y_k^n))_{n\geq 1}$ (see **A1**), we deduce that

$$\begin{aligned}
&\mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\sum_{n=1}^{q} \mathrm{Q}^N[f](x_k^n, y_k^n, \{W_k^i\}_{i=1,\dots,N}) \,\big|\, \mathcal{F}_k^N\Big] \\
&= \mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\sum_{n=1}^{q} \mathrm{Q}^N[f](x_k^n, y_k^n, \{w_k^i\}_{i=1,\dots,N})\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i} \\
&= \mathbf{E}\big[\mathbf{1}_{|B_k|=q}\big]\mathbf{E}\Big[\sum_{n=1}^{q} \mathrm{Q}^N[f](x_k^n, y_k^n, \{w_k^i\}_{i=1,\dots,N})\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i} \\
&= q\mathbf{E}\big[\mathbf{1}_{|B_k|=q}\big]\mathbf{E}\Big[\mathrm{Q}^N[f](x_1^1, y_1^1, \{w_k^i\}_{i=1,\dots,N})\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i} \\
&= q\,\frac{N}{\alpha}\mathbf{E}\big[\mathbf{1}_{|B_k|=q}\big]\,\langle f, D_k^N\rangle,
\end{aligned}$$

where we have used **A3** to deduce the last two equalities. We have thus proved that

$$\mathbf{E}\Big[\frac{\alpha}{N|B_k|}\sum_{(x,y)\in B_k} \mathrm{Q}^N[f](x_k^n, y_k^n, \{W_k^i\}_{i=1,\dots,N}) \,\big|\, \mathcal{F}_k^N\Big] = \langle f, D_k^N\rangle.$$

Therefore, using in addition that $\mathbf{E}[\langle f, D_k^N\rangle|\,\mathcal{F}_k^N] = \langle f, D_k^N\rangle$ (because $(W_k^1,\dots,W_k^N)$ is $\mathcal{F}_k^N$-measurable), we finally deduce (41).

Step 3. We now end the proof of item *(iii)*. If $j>k$, $\langle f, M_k^N\rangle$ is $\mathcal{F}_j^N$-measurable (because $\langle f, M_k^N\rangle$ is $\mathcal{F}_{k+1}^N$-measurable). Then, using also (41), one obtains that for $j>k$:

$$\mathbf{E}\left[\langle f, M_k^N\rangle\langle f, M_j^N\rangle\right] = \mathbf{E}\left[\langle f, M_k^N\rangle \mathbf{E}\left[\langle f, M_j^N\rangle|\mathcal{F}_j^N\right]\right] = \mathbf{E}\left[\langle f, M_k^N\rangle\times 0\right] = 0. \quad (43)$$

We then have (see (23)):

$$\mathbf{E}\left[|\langle f, M_t^N\rangle|^2\right] = \sum_{k=0}^{\lfloor Nt\rfloor -1}\mathbf{E}\left[|\langle f, M_k^N\rangle|^2\right]. \quad (44)$$

Plugging (36) in (44) implies item *(iii)*.

Proof of item *(iv)*. Let $t\in[0,T]$. By Lemma 8, $\nabla f(W_k^i)\cdot\varepsilon_k^i$ is square-integrable for all $k\in\mathbf{N}$ and $i\in\{1,\dots,N\}$. From the equality

$$\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big] = \sum_{j,k=0}^{\lfloor Nt\rfloor -1}\sum_{i,\ell=1}^N \mathbf{E}\big[\nabla f(W_k^i)\cdot\varepsilon_k^i\,\nabla f(W_j^\ell)\cdot\varepsilon_j^\ell\big].$$

Recall that W_a^b is $\mathcal{F}_a^N$-measurable for all $a\in\mathbf{N}$ and $b\in\{1,\dots,N\}$, and that $\varepsilon_a^b \perp\!\!\!\perp \mathcal{F}_a^N$ (see **A5**). Let e_q denotes the q-th element of the canonical basis of $\mathbf{R}^d$ ($q\in\{1,\dots,d\}$). Assume that $0\le j<k\le\lfloor Nt\rfloor -1$. Then, ε_j^ℓ is $\mathcal{F}_k^N$-measurable, and it holds for all $i,\ell\in\{1,\dots,N\}$:

$$\mathbf{E}\big[\nabla f(W_k^i)\cdot\varepsilon_k^i\,\nabla f(W_j^\ell)\cdot\varepsilon_j^\ell\big] = \sum_{n,m=1}^d \mathbf{E}\Big[\partial_{e_n} f(W_k^i)\,\partial_{e_m} f(W_j^\ell)\,\varepsilon_j^\ell\cdot e_m\Big]\mathbf{E}\big[\varepsilon_k^i\cdot e_n\big] = 0. \quad (45)$$

because $\varepsilon_k^i\sim\mathcal{N}(0,I_d)$ (see **A5**). On the other hand, using **A5**, we have for all $0\le k\le\lfloor Nt\rfloor -1$ and when $i\neq\ell\in\{1,\dots,N\}$:

$$\mathbf{E}\big[\nabla f(W_k^i)\cdot\varepsilon_k^i\,\nabla f(W_k^\ell)\cdot\varepsilon_k^\ell\big] = \sum_{n,m=1}^d \mathbf{E}\Big[\,\partial_{e_n} f(W_k^i)\,\partial_{e_m} f(W_k^\ell)\,\Big]\mathbf{E}\big[\varepsilon_k^i\cdot e_n\big]\mathbf{E}\big[\varepsilon_k^\ell\cdot e_m\big] = 0. \quad (46)$$

Consequently, we have:

$$\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big] = \sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \mathbf{E}\big[|\nabla f(W_k^i)\cdot\varepsilon_k^i|^2\big].$$

Using the Cauchy-Schwarz inequality, we deduce, using also Lemma 8 and **A5**, that:

$$\begin{aligned}\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big] &\le \sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \mathbf{E}\left[|\nabla f(W_k^i)|^2\right]\mathbf{E}\left[|\varepsilon_k^i|^2\right]\\ &\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \mathbf{E}\left[(1+|W_k^i|^{\gamma_*})^2\right]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}N^2.\end{aligned} \quad (47)$$

This proves *(iv)*. The proof of Lemma 9 is complete. ■

We now want to pass to the limit in (25). To this end, we first prove that $(\mu^N)_{N\geq 1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$. This is the purpose of the following section.

2.2 Relative compactness in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ and convergence to the limit equation

In this section, we show that $(\mu^N)_{N\geq 1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Then, we prove that any limit point of $(\mu^N)_{N\geq 1}$ satisfies a.s. (10).

2.2.1 Relative compactness in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$

In this section we prove the following result.

Proposition 10 *Let $\beta \geq 1/2$ and assume that the conditions* **A1**-**A5** *hold. Then, $(\mu^N)_{N\geq 1}$ is relatively compact in in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$.*

We first recall the following standard result.

Proposition 11 *Let $q > p \geq 1$ and $C > 0$. The set $\mathscr{K}_C^q := \{\mu \in \mathcal{P}_p(\mathbf{R}^d), \int_{\mathbf{R}^d} |x|^q \mu(\mathrm{d}x) \leq C\}$ is compact.*

We have the following result.

Lemma 12 *Let $\beta \geq 1/2$ and assume that* **A1**-**A5** *hold. Then, for every $T > 0$, there exists $C > 0$ such that for all $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$,*

$$\sup_{N\geq 1} \mathbf{E}\Big[\sup_{t\in[0,T]} \langle f, \mu_t^N\rangle^2\Big] \leq C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}. \tag{48}$$

Proof Let $T > 0$ and $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. All along the proof, $C < \infty$ denotes a constant independent of $t \in [0,T]$, $N \geq 1$, $k \in \{0, \ldots, \lfloor Nt \rfloor\}$, $i \in \{1, \ldots, N\}$, and $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$, which can change from one occurence to another. From (25), we have:

$$\begin{aligned}\sup_{t\in[0,T]} \langle f, \mu_t^N\rangle^2 \leq C\Big[&\langle f, \mu_0^N\rangle^2 + \int_0^T \int_{\mathcal{X}\times\mathcal{Y}} (y - \langle \sigma_*(\cdot,x), \mu_s^N\rangle)^2 \langle \nabla f \cdot \nabla \sigma_*(\cdot,x), \mu_s^N\rangle^2 \pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s \\ &+ \sup_{t\in[0,T]} \langle f, M_t^N\rangle^2 + \sup_{t\in[0,T]} |\langle f, V_t^N\rangle|^2 + \sup_{t\in[0,T]} \Big|\sum_{k=0}^{\lfloor Nt\rfloor -1} \langle f, R_k^N\rangle\Big|^2 \\ &+ \frac{1}{N^{2+2\beta}} \sup_{t\in[0,T]} \Big|\sum_{k=0}^{\lfloor Nt\rfloor -1} \sum_{i=1}^N \nabla f(W_k^i) \cdot \varepsilon_k^i\Big|^2\Big].\end{aligned} \tag{49}$$

We now study each term of the right-hand side of (49). Let us deal with the first term in the right-hand side of (49). Using **A4** and (26), it holds:

$$\begin{aligned}\mathbf{E}\left[\langle f,\mu_0^N\rangle^2\right]=\mathbf{E}\left[\langle f,\nu_0^N\rangle^2\right]=\mathbf{E}\Big[\Big|\frac{1}{N}\sum_{i=1}^N f(W_0^i)\Big|^2\Big]&\le\frac{1}{N}\sum_{i=1}^N\mathbf{E}\left[|f(W_0^i)|^2\right]\\&\le\frac{\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N\mathbf{E}\left[(1+|W_0^i|^{\gamma_*})^2\right]\\&\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}.\end{aligned}$$

For the second term in the right-hand side of (49), we have since $\mathbf{E}[|y|^2]<+\infty$ (see (**A3**)) and using (32), (33), and Lemma 8:

$$\mathbf{E}\left[\int_0^T\int_{\mathcal{X}\times\mathcal{Y}}(y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle)^2\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\mu_s^N\rangle^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\right]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}.$$

Let us deal with the third term in the right-hand side of (49). By (23) and (26), we have, for $t\in[0,T]$,

$$\sup_{t\in[0,T]}|\langle f,M_t^N\rangle|^2\le\lfloor NT\rfloor\sum_{k=0}^{\lfloor NT\rfloor-1}\langle f,M_k^N\rangle^2.$$

Hence, using (36), we obtain that $\mathbf{E}\big[\sup_{t\in[0,T]}\langle f,M_t^N\rangle^2\big]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}$. Let us deal with the fourth term in the right-hand side of (49). From (26) and (35),

$$\sup_{t\in[0,T]}|\langle f,V_t^N\rangle|^2\le\frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^3}\sum_{i=1}^N\max_{0\le k\le\lfloor NT\rfloor}(1+|W_k^i|^{\gamma_*})^2,$$

which leads to

$$\begin{aligned}\mathbf{E}\Big[\sup_{t\in[0,T]}|\langle f,V_t^N\rangle|^2\Big]&\le\frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^3}\sum_{i=1}^N\mathbf{E}\left[\max_{0\le k\le\lfloor NT\rfloor}(1+|W_k^i|^{\gamma_*})^2\right]\\&\le\frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^3}\sum_{i=1}^N\sqrt{\sum_{k=0}^{\lfloor NT\rfloor}\mathbf{E}\left[(1+|W_k^i|^{\gamma_*})^4\right]}\le\frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^{3/2}}.\end{aligned}\tag{50}$$

Let us now consider the fifth term in the right-hand side of (49). From (31) and (26), we have

$$\begin{aligned}|\langle f,R_k^N\rangle|^2\le\frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N\Big[&\frac{1}{N^4}(1+|W_k^i|^{\gamma_*}+|W_{k+1}^i|^{\gamma_*})^4+\frac{1}{N^4|B_k|}\sum_{(x,y)\in B_k}(|y|^4+C)^2\\&+\frac{1}{N^{4\beta}}\left[|\varepsilon_k^i|^8+(1+|W_k^i|^{\gamma_*}+|W_{k+1}^i|^{\gamma_*})^4\right]\Big].\end{aligned}$$

Then, by **A3** and **A5** together with Lemma 8 and (27), we obtain

$$\mathbf{E}\left[|\langle f,R_k^N\rangle|^2\right]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\Big[\frac{1}{N^4}+\frac{1}{N^{4\beta}}\Big].\tag{51}$$

Therefore, using also (26), it holds:

$$\mathbf{E}\Big[\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f, R_k^N\rangle\Big|^2\Big] \le \lfloor NT\rfloor \sum_{k=0}^{\lfloor NT\rfloor-1}\mathbf{E}\left[|\langle f, R_k^N\rangle|^2\right] \le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\Big[\frac{1}{N^2}+\frac{N^2}{N^{4\beta}}\Big]. \quad (52)$$

Let us deal with the last term in the right-hand side of (49). Using the same arguments leading to (47) together with (26) and (46) we have

$$\begin{aligned}
\frac{1}{N^{2+2\beta}}\mathbf{E}\Big[\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big] &\le \frac{C}{N^{2+2\beta}}\mathbf{E}\Big[\sup_{t\in[0,T]}\lfloor Nt\rfloor\sum_{k=0}^{\lfloor Nt\rfloor-1}\Big|\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]\\
&\le \frac{C\lfloor NT\rfloor}{N^{2+2\beta}}\mathbf{E}\Big[\sum_{k=0}^{\lfloor NT\rfloor-1}\Big|\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]\\
&\le \frac{C}{N^{1+2\beta}}\sum_{k=0}^{\lfloor NT\rfloor-1}\mathbf{E}\Big[\Big|\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]\\
&\le \frac{C}{N^{1+2\beta}}\sum_{k=0}^{\lfloor NT\rfloor-1}\sum_{i=1}^{N}\mathbf{E}\big[|\nabla f(W_k^i)\cdot\varepsilon_k^i|^2\big]\\
&\le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^{2\beta-1}}. \quad (53)
\end{aligned}$$

Plugging all these previous bounds in (49), we obtain (recall that $\beta\ge 1/2$), for all $f\in\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$, $\mathbf{E}[\sup_{t\in[0,T]}\langle f,\mu_t^N\rangle^2]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}$. This proves (48) and ends the proof of the lemma. ■

Lemma 12 provides the following compact containment for $(\mu^N)_{N\ge1}$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))$.

Corollary 13 *Assume $\beta\ge1/2$ and* **A1**-**A5**. *Let $0<\epsilon<1$. For every $T>0$,*

$$\sup_{N\ge1}\mathbf{E}\Big[\sup_{t\in[0,T]}\int_{\mathbf{R}^d}|w|^{\gamma+\epsilon}\mu_t^N(\mathrm{d}w)\Big]<+\infty. \quad (54)$$

Proof Recall $\gamma_*-\gamma=1$. Thus, it holds $f: w\mapsto(1-\chi(w))|w|^{\gamma+\epsilon}\in\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$ since $\gamma_*>\gamma+\epsilon$. The result follows from Lemma 12. ■

The following result will also be needed.

Lemma 14 *Assume* **A1**-**A5** *and $\beta\ge1/2$. For all $T>0$, there exists $C>0$ such that for all $\delta>0$ and $0\le r<t\le T$ such that $t-r\le\delta$, one has for all $N\ge1$ and $f\in\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$:*

$$\mathbf{E}\left[|\langle f,\mu_t^N\rangle-\langle f,\mu_r^N\rangle|^2\right]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\left[\delta^2+\frac1N+\frac{1}{N^2}+(N\delta+1)\left[\frac{1}{N^4}+\frac{1}{N^{4\beta}}\right]+\frac{1}{N^{2\beta}}\right]. \quad (55)$$

Proof Let $\delta > 0$ and $0 \le r < t \le T$ such that $t - r \le \delta$. Let $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. In the following, $C > 0$ is a constant independent of t, r, δ, N, and f, which can change from one occurence to another. From (25), we have

$$\begin{aligned}\langle f, \mu_t^N\rangle - \langle f, \mu_r^N\rangle = &\int_r^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \mu_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s\\ &+ \langle f, M_t^N\rangle - \langle f, M_r^N\rangle + \langle f, V_t^N\rangle - \langle f, V_r^N\rangle + \sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1} \langle f, R_k^N\rangle\\ &+ \frac{1}{N^{1+\beta}} \sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1} \sum_{i=1}^N \nabla f(W_k^i)\cdot \varepsilon_k^i.\end{aligned}$$

Jensen's inequality provides

$$\begin{aligned}&|\langle f, \mu_t^N\rangle - \langle f, \mu_r^N\rangle|^2\\ &\le C\Big[(t-r)\int_r^t \int_{\mathcal{X}\times\mathcal{Y}} \big|(y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \mu_s^N\rangle\big|^2 \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s\\ &\quad + \big|\langle f, M_t^N\rangle - \langle f, M_r^N\rangle\big|^2 + \big|\langle f, V_t^N\rangle - \langle f, V_r^N\rangle\big|^2 + \Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1} \langle f, R_k^N\rangle\Big|^2\\ &\quad + \frac{1}{N^{2+2\beta}}\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1} \sum_{i=1}^N \nabla f(W_k^i)\cdot \varepsilon_k^i\Big|^2\Big].\end{aligned} \tag{56}$$

We now study each term of the right-hand side of (56). Let us consider the first term in the right-hand side of (56). From (32), (33) and (26), we have:

$$\begin{aligned}\mathbf{E}\left[|y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle|^2 |\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \mu_s^N\rangle|^2\right] &\le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N}\mathbf{E}\Big[(|y|^2 + C)\sum_{i=1}^N (1 + |W^i_{\lfloor Ns\rfloor}|^{\gamma_*})^2\Big]\\ &\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}},\end{aligned}$$

where the last inequality follows from **A3** and Lemma 8. We then have:

$$\begin{aligned}&\mathbf{E}\left[(t-r)\int_r^t \int_{\mathcal{X}\times\mathcal{Y}} \big((y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \mu_s^N\rangle\big)^2 \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s\right]\\ &\le C(t-r)^2\|f\|^2_{\mathcal{C}^{2,\gamma_*}} \le C\delta^2\|f\|^2_{\mathcal{C}^{2,\gamma_*}}.\end{aligned} \tag{57}$$

Let us consider the second term in the right-hand side of (56). From item *(iii)* of Lemma 9, we have

$$\mathbf{E}\left[\big(\langle f, M_t^N\rangle - \langle f, M_r^N\rangle\big)^2\right] \le 2\mathbf{E}\left[\langle f, M_t^N\rangle^2 + \langle f, M_r^N\rangle^2\right] \le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N}. \tag{58}$$

Let us consider the third term in the right-hand side of (56). From (35) and (26), we have

$$|\langle f, V_t^N\rangle|^2 \le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^3}\sum_{i=1}^N (1 + |W^i_{\lfloor Nt\rfloor}|^{\gamma_*})^2.$$

Therefore, by Lemma 8, we obtain that:

$$\mathbf{E}\big[\big|\langle f, V_t^N\rangle - \langle f, V_r^N\rangle\big|^2\big] \le 2\mathbf{E}\left[|\langle f, V_t^N\rangle|^2 + |\langle f, V_r^N\rangle|^2\right] \le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^2}. \tag{59}$$

Let us consider the fourth term in the right-hand side of (56). By (51),

$$\begin{aligned}\mathbf{E}\Big[\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\langle f, R_k^N\rangle\Big|^2\Big] &\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}(\lfloor Nt\rfloor - \lfloor Nr\rfloor)\left[\frac{1}{N^4}+\frac{1}{N^{4\beta}}\right]\\ &\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}(N\delta+1)\left[\frac{1}{N^4}+\frac{1}{N^{4\beta}}\right]. \end{aligned}\tag{60}$$

Let us consider the last term in the right-hand side of (56). By item *(iv)* in Lemma 9,

$$\begin{aligned}&\frac{1}{N^{2+2\beta}}\mathbf{E}\Big[\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]\\ &\le \frac{2}{N^{2+2\beta}}\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2+\Big|\sum_{k=0}^{\lfloor Nr\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big] \le \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^{2\beta}}. \end{aligned}\tag{61}$$

Using (57), (58), (59), (60), (61), and (56), we deduce (55). ■

We now collect the results of the previous lemmata to prove Proposition 10.

Proof [Proof of Proposition 10] To prove Proposition 10, we apply (Jakubowski, 1986, Theorem 4.6) with $E = \mathcal{P}_\gamma(\mathbf{R}^d)$ and $\mathbb{F} = \{\mathsf{V}_f, f \in \mathcal{C}_c^\infty(\mathbf{R}^d)\}$ where

$$\mathsf{V}_f : \nu \in \mathcal{P}_\gamma(\mathbf{R}^d) \mapsto \langle f, \nu\rangle.$$

The set $\mathbb{F}$ on $\mathcal{P}_\gamma(\mathbf{R}^d)$ satisfies Conditions (Jakubowski, 1986, (3.1) and (3.2) in Theorem 3.1). Condition (4.8) there is a consequence of Proposition 11, Corollary 13, together with Markov's inequality. We now prove that (Jakubowski, 1986, Condition (4.9)) is verified, i.e. let us show that all $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$, the sequence $(\langle f, \mu^N\rangle)_{N\ge1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathbf{R})$. To do so, it suffices to use Lemma 14 and Proposition 38 below (with $\mathcal{H}_1 = \mathcal{H}_2 = \mathbf{R}$ there). In conclusion, according to (Jakubowski, 1986, Theorem 4.6), $(\mu^N)_{N\ge1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. ■

2.2.2 Limit points in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ are continuous in time

In this section we show that any limit point of $(\mu^N)_{N\ge1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ belongs a.s. to $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$.

Proposition 15 *Let $\beta > 1/2$ and assume* **A1-A5**. *Consider $\mu^* \in \mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ a limit point of $(\mu^N)_{N\ge1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Then, a.s. $\mu^* \in \mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$.*

Proof Let N' be a subsequence such that in distribution $\mu^{N'} \to \mu^*$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Because $\mathsf{W}_1 \le \mathsf{W}_\gamma$, $\mu^{N'} \to \mu^*$ in distribution also in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. By (Jacod and Shiryaev, 1987, Proposition 3.26 in Chapter VI), $\mu^* \in \mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$ a.s. if for all $T > 0$, $\lim_{N\to+\infty} \mathbf{E}\big[\sup_{t\in[0,T]} \mathsf{W}_1(\mu^N_{t_-}, \mu^N_t)\big] = 0$. According to the duality formula (8), this is equivalent to

$$\lim_{N\to+\infty} \mathbf{E}\Big[\sup_{t\in[0,T]} \sup_{\|f\|_{\mathrm{Lip}}\le 1} |\langle f, \mu^N_{t_-}\rangle - \langle f, \mu^N_t\rangle|\Big] = 0. \tag{62}$$

Let $T > 0$ and consider a Lipschitz function $f : \mathbf{R}^d \to \mathbf{R}$ such that $\|f\|_{\mathrm{Lip}} \le 1$. One has that $\langle f, \mu^N_t\rangle = \langle f, \mu^N_0\rangle + \sum_{k=0}^{\lfloor Nt\rfloor-1} \langle f, \nu^N_{k+1}\rangle - \langle f, \nu^N_k\rangle$ (with the convention $\sum_0^{-1} = 0$). Therefore, the discontinuity points of $t \in [0,T] \mapsto \langle f, \mu^N_t\rangle$ are exactly $\{1/N, 2/N, \dots, \lfloor NT\rfloor/N\}$ and for all $t \in [0,T]$,

$$|\langle f, \mu^N_{t_-}\rangle - \langle f, \mu^N_t\rangle| \le \max_{k=0,\dots,\lfloor NT\rfloor-1} |\langle f, \nu^N_{k+1}\rangle - \langle f, \nu^N_k\rangle|. \tag{63}$$

Let $k \in \{0, \dots, \lfloor NT\rfloor - 1\}$. We have using (2) and **A2**:

$$|\langle f, \nu^N_{k+1}\rangle - \langle f, \nu^N_k\rangle| \le \frac{1}{N}\sum_{i=1}^N |W^i_{k+1} - W^i_k| \le \frac{C}{N}\sum_{i=1}^N \Big[\frac{1}{N|B_k|}\sum_{(x,y)\in B_k} (|y|+1) + \frac{|\varepsilon^i_k|}{N^\beta}\Big] =: \beta^N_k \tag{64}$$

Then, one deduces that

$$|\beta^N_k|^2 \le \frac{C}{N}\sum_{i=1}^N \Big[\frac{1}{N^2|B_k|}\sum_{(x,y)\in B_k} (|y|^2+1) + \frac{|\varepsilon^i_k|^2}{N^{2\beta}}\Big],$$

and hence that $\mathbf{E}[|\beta^N_k|^2] \le C(1/N^2 + 1/N^{2\beta})$ where $C > 0$ is independent of $N \ge 1$ and $k = 0, \dots, \lfloor NT\rfloor - 1$. Then, using (63) and(64),

$$\begin{aligned}\mathbf{E}\Big[\sup_{t\in[0,T]} \sup_{\|f\|_{\mathrm{Lip}}\le 1} |\langle f, \mu^N_{t_-}\rangle - \langle f, \mu^N_t\rangle|\Big] &\le \mathbf{E}\Big[\sup_{\|f\|_{\mathrm{Lip}}\le 1} \max_{k=0,\dots,\lfloor NT\rfloor-1} |\langle f, \nu^N_{k+1}\rangle - \langle f, \nu^N_k\rangle|\Big]\\ &\le \mathbf{E}\Big[\max_{k=0,\dots,\lfloor NT\rfloor-1} \beta^N_k\Big]\\ &\le \mathbf{E}\Big[\sqrt{\sum_{k=0}^{\lfloor NT\rfloor-1} |\beta^N_k|^2}\Big] \le \sqrt{\mathbf{E}\Big[\sum_{k=0}^{\lfloor NT\rfloor-1} |\beta^N_k|^2\Big]}\\ &\le C\big[1/\sqrt{N} + \sqrt{N/N^{2\beta}}\big].\end{aligned}$$

This proves (62) since $\beta > 1/2$. The proof of Proposition 15 is complete. ■

We end this section with the following result which will be used later in the proof of Theorem 2.

Lemma 16 *Let $\beta \geq 1/2$ and assume* **A1**-**A5**. *Then, for all $T > 0$ there exists $C > 0$,*

$$\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f, \mu_t^N - \mu_{t^-}^N\rangle^2\Big] \leq C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\Big[\frac{1}{N^{3/2}} + \sqrt{\frac{1}{N^7} + \frac{1}{N^{8\beta-1}}} + \frac{\sqrt{N}}{N^{2\beta}}\Big],\ \forall f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d). \tag{65}$$

Proof The arguments used in the proof of Proposition 15 are not sufficient to prove (65). We will rather use (25). Let $T > 0$ and $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$. In what follows, $C > 0$ is a constant, independent of $N \geq 1$, $k = 0, \dots \lfloor NT\rfloor - 1$, and f, which can change from one occurence to another. Recall that $t \in [0,T] \mapsto \langle f, \mu_t^N\rangle \in \mathbf{R}$ has $\lfloor NT\rfloor$ discontinuities, located at the points $\frac{1}{N}, \frac{2}{N}, \dots, \frac{\lfloor NT\rfloor}{N}$. In addition, from (25), (24), and (23), for $k \in \{1, \dots, \lfloor NT\rfloor\}$, its k-th discontinuity is equal to

$$\begin{aligned}\mathbf{d}_k^N[f] :=& \langle f, M_{k-1}^N\rangle + \int_{\frac{k-1}{N}}^{\frac{k}{N}}\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y - \langle\sigma_*(\cdot,x),\mu_s^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\mu_s^N\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\ &+ \langle f, R_{k-1}^N\rangle + \frac{1}{N^{1+\beta}}\sum_{i=1}^N \nabla f(W_{k-1}^i)\cdot\varepsilon_{k-1}^i.\end{aligned} \tag{66}$$

Thus,

$$\sup_{t\in[0,T]}|\langle f, \mu_t^N - \mu_{t^-}^N\rangle|^2 \leq \max\big\{|\mathbf{d}_{k+1}^N[f]|^2,\ 0 \leq k < \lfloor NT\rfloor\big\}. \tag{67}$$

By (38) and (26), it holds:

$$|\langle f, M_k^N\rangle|^4 \leq \frac{C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}}{N^5|B_k|}\sum_{(x,y)\in B_k}\sum_{i=1}^N (y^4 + C)(1 + |W_k^i|^{\gamma_*})^4.$$

Then, using Lemma 8, we have with the same computations as the one made in (39):

$$\mathbf{E}\left[|\langle f, M_k^N\rangle|^4\right] \leq \frac{C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}}{N^5}\sum_{i=1}^N \mathbf{E}\left[y^4 + C\right]\mathbf{E}\left[(1 + |W_k^i|^{\gamma_*})^4\right] \leq \frac{C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}}{N^4}. \tag{68}$$

Consequently, one has:

$$\mathbf{E}\left[\max_{0\leq k<\lfloor NT\rfloor}\langle f, M_k^N\rangle^2\right] \leq \Big|\sum_{k=0}^{\lfloor NT\rfloor-1}\mathbf{E}\left[\langle f, M_k^N\rangle^4\right]\Big|^{1/2} \leq \frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^{3/2}}. \tag{69}$$

By (32), (33) and since $\lfloor Ns\rfloor = k$ when $s\in[k/N,(k+1)/N]$, we have

$$\begin{aligned}
&\Big|\int_{\frac{k}{N}}^{\frac{k+1}{N}}\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\mu^N_s\rangle\,\pi(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\Big|\\
&\le\frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\int_{\mathcal{X}\times\mathcal{Y}}(|y|+C)\sum_{i=1}^N(1+|W^i_k|^{\gamma_*})\,\pi(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\
&=\frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N^2}\mathbf{E}[|y|+C]\sum_{i=1}^N(1+|W^i_k|^{\gamma_*})\\
&\le\frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N^2}\sum_{i=1}^N(1+|W^i_k|^{\gamma_*}).
\end{aligned}$$

By (26) and Lemma 8, it then holds:

$$\begin{aligned}
&\mathbf{E}\Big[\Big|\int_{\frac{k}{N}}^{\frac{k+1}{N}}\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\mu^N_s\rangle\,\pi(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\Big|^4\Big]\\
&\le\frac{C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}N^3}{N^8}\mathbf{E}\Big[\sum_{i=1}^N(1+|W^i_k|^{\gamma_*})^4\Big]\le\frac{C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}}{N^4}.
\end{aligned}$$

Thus, one has:

$$\begin{aligned}
&\mathbf{E}\Big[\max_{0\le k<\lfloor NT\rfloor}\Big|\int_{\frac{k}{N}}^{\frac{k+1}{N}}\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\mu^N_s\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\Big|^2\Big]\\
&\le\Big|\sum_{k=0}^{\lfloor NT\rfloor-1}\mathbf{E}\Big[\Big|\int_{\frac{k}{N}}^{\frac{k+1}{N}}\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\mu^N_s\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\Big|^4\Big]\Big|^{1/2}\\
&\le\frac{C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}}{N^{3/2}}.
\end{aligned}$$

On the other hand, from (31) and (26), we have

$$\begin{aligned}
|\langle f,R^N_k\rangle|^4\le\frac{C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N\Bigg[&\frac{1}{N^8}(1+|W^i_k|^{\gamma_*}+|W^i_{k+1}|^{\gamma_*})^8+\frac{1}{N^8|B_k|}\sum_{(x,y)\in B_k}(|y|^4+C)^4\\
&+\frac{C}{N^{8\beta}}\left[|\varepsilon^i_k|^{16}+(1+|W^i_k|^{\gamma}+|W^i_{k+1}|^{\gamma_*})^8\right]\Bigg].
\end{aligned}$$

Using Lemma 8, **A3**, and the same computations as those made in (27), we deduce that:

$$\mathbf{E}\left[|\langle f,R^N_k\rangle|^4\right]\le C\|f\|^4_{\mathcal{C}^{2,\gamma}}\big(1/N^8+1/N^{8\beta}\big).$$

Then, it holds:

$$\mathbf{E}\Big[\max_{0\le k<\lfloor NT\rfloor}|\langle f,R^N_k\rangle|^2\Big]\le\Big|\sum_{k=0}^{\lfloor NT\rfloor-1}\mathbf{E}\left[|\langle f,R^N_k\rangle|^4\right]\Big|^{1/2}\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\Big[\frac{1}{N^7}+\frac{1}{N^{8\beta-1}}\Big]^{1/2}.$$

By (26), **A5**, and Lemma 8,

$$\mathbf{E}\Big[\Big|\sum_{i=1}^N \nabla f(W_k^i)\cdot \varepsilon_k^i\Big|^4\Big] \le N^3\sum_{i=1}^N \mathbf{E}\left[|\nabla f(W_k^i)\cdot \varepsilon_k^i|^4\right]$$
$$\le N^3\|f\|^4_{\mathcal{C}^{2,\gamma_*}}\sum_{i=1}^N \mathbf{E}\left[(1+|W_k^i|^{\gamma_*})^4\right]\mathbf{E}\left[|\varepsilon_k^i|^4\right] \le C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}N^4.$$

Thus, one deduces that

$$\mathbf{E}\Big[\max_{0\le k<\lfloor NT\rfloor}\Big|\frac{1}{N^{1+\beta}}\sum_{i=1}^N \nabla f(W_k^i)\cdot \varepsilon_k^i\Big|^2\Big] \le \Big|\sum_{k=0}^{\lfloor NT\rfloor-1}\mathbf{E}\Big[\Big|\frac{1}{N^{1+\beta}}\sum_{i=1}^N \nabla f(W_k^i)\cdot \varepsilon_k^i\Big|^4\Big]\Big|^{1/2}$$
$$\le \Big|\frac{1}{N^{4+4\beta}}C\|f\|^4_{\mathcal{C}^{2,\gamma_*}}N^5\Big|^{1/2} \le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\frac{\sqrt{N}}{N^{2\beta}}.$$

Plugging all these previous bounds in (67) implies (65). ■

2.2.3 Convergence to the limit equation (10)

This section is devoted to prove Proposition 18 where we show that any limit point of $(\mu^N)_{N\ge1}$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))$ satisfies a.s. (10).

For $t\in\mathbf{R}_+$ and $f\in\mathcal{C}^{1,\gamma}(\mathbf{R}^d)$, we introduce the function $\mathbf{\Lambda}_t[f]:\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))\to\mathbf{R}$ defined by

$$\mathbf{\Lambda}_t[f]: m\mapsto \Big|\langle f,m_t\rangle-\langle f,\mu_0\rangle-\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),m_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),m_s\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\Big|.$$

To prove that any limit point of the sequence $(\mu^N)_{N\ge1}$ in the space $\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))$ satisfies (10), we study the continuity of the function $\mathbf{\Lambda}_t[f]$. This is the purpose of Lemma 17.

Lemma 17 *For any $t\in\mathbf{R}_+$ and $f\in\mathcal{C}^{1,\gamma}(\mathbf{R}^d)$, the function $\mathbf{\Lambda}_t[f]$ is well defined. In addition, let $(m^N)_{N\ge1}$ be such that $m^N\to m$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))$. Then, for all continuity points $t\in\mathbf{R}_+$ of m, $\mathbf{\Lambda}_t[f](m^N)\to\mathbf{\Lambda}_t[f](m)$ as $N\to+\infty$.*

Proof In the following $C>0$ is a constant independent of $f\in\mathcal{C}^{1,\gamma}(\mathbf{R}^d)$, $s\in[0,t]$, $x\in\mathcal{X}$, and $y\in\mathcal{Y}$, which can change from one occurence to another. By **A4**

$$|\langle f,\mu_0\rangle| = \Big|\int_{\mathbf{R}^d}\frac{f(w)}{1+|w|^\gamma}(1+|w|^\gamma)\mu_0(\mathrm{d}w)\Big| \le (1+\mathbf{E}[|W_0^1|^\gamma])\|f\|_{\mathcal{C}^{0,\gamma}}. \tag{70}$$

The following result (Villani, 2009, Theorem 6.9) will be used many times in the following:

$$\mu_n\to\mu \text{ in } \mathcal{P}_\gamma(\mathbf{R}^d) \text{ iff } \langle g,\mu_n\rangle\to\langle g,\mu\rangle \text{ for all } g:\mathbf{R}^d\to\mathbf{R} \text{ continuous s.t. } \frac{g}{1+|\cdot|^\gamma} \text{ is bounded.} \tag{71}$$

In particular $u \geq 0 \mapsto \langle 1+|\cdot|^\gamma, m_u\rangle \in \mathcal{D}(\mathbf{R}_+, \mathbf{R})$ and thus $\sup_{u\in[0,t]} |\langle 1+|\cdot|^\gamma, m_u\rangle| < +\infty$ for all $t \geq 0$. Define the function $\phi_s^{x,y}(m) = \alpha(y - \langle \sigma_*(\cdot, x), m_s\rangle)\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), m_s\rangle$. Using **A2**, one has for $s \in [0, t]$:

$$|\phi_s^{x,y}(m)| \leq C(1 + |y|) \sup_{u\in[0,t]} |\langle 1 + | \cdot |^\gamma, m_u\rangle| \|f\|_{\mathcal{C}^{1,\gamma}}. \tag{72}$$

Using also **A3**, this proves that $\mathbf{\Lambda}_t[f]$ is well defined.

Let us now consider $(m^N)_{N\geq 1}$ such that $m^N \to m$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Denote by $\mathcal{C}(m) \subset \mathbf{R}_+$ the set of continuity points of m. From (Ethier and Kurtz, 2009, Proposition 5.2 in Chapter 3), we have that for all $t \in \mathcal{C}(m)$, $m_t^N \to m_t$ in $\mathcal{P}_\gamma(\mathbf{R}^d)$, and thus, for all $t \in \mathcal{C}(m)$, according to (71),

$$\langle f, m_t^N\rangle \underset{N\to\infty}{\longrightarrow} \langle f, m_t\rangle.$$

For the same reasons, for all $s \in [0, t] \cap \mathcal{C}(m)$ and $x \in \mathcal{X}$,

$$\langle \sigma_*(\cdot, x), m_s^N\rangle \underset{N\to\infty}{\longrightarrow} \langle \sigma_*(\cdot, x), m_s\rangle \quad \text{and} \quad \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), m_s^N\rangle \underset{N\to\infty}{\longrightarrow} \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), m_s\rangle.$$

Since $\mathbf{R}_+\backslash\mathcal{C}(m)$ is at most countable (see (Ethier and Kurtz, 2009, Lemma 5.1 in Chapter 3)), it holds a.e. on $[0, t] \times \mathcal{X} \times \mathcal{Y}$, $\phi_s^{x,y}(m^N) \to \phi_s^{x,y}(m)$. Note that using (Ethier and Kurtz, 2009, Item (b) in Proposition 5.3 in Chapter 3) together with the triangular inequality:

$$\langle | \cdot |^\gamma, m_u^N\rangle = \mathsf{W}_\gamma(\delta_0, m_u^N) \leq [\mathsf{W}_\gamma(\delta_0, m_{\lambda_u^N}) + \mathsf{W}_\gamma(m_{\lambda_u^N}, m_u^N)]^\gamma,\ \lambda^N : \mathbf{R}_+ \to \mathbf{R}_+,\ u \geq 0,$$

one deduces that there exists $C > 0$, for all $N \geq 1$ and $s \in [0, t]$, $|\langle 1 + | \cdot |^\gamma, m_s^N\rangle| \leq C$. Together with (72), one has using the dominated convergence theorem,

$$\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \phi_s^{x,y}(m^N)\pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \to \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \phi_s^{x,y}(m)\pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s.$$

This proves the desired result. ∎

We are now in position to prove that any limit point of the sequence $(\mu^N)_{N\geq 1}$ in the space $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ satisfies (10).

Proposition 18 *Let $\beta > 1/2$ and assume* **A1-A5**. *Let μ^* be a limit point of $(\mu^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Then, a.s., μ^* satisfies* (10).

Proof Up to extracting a subsequence, we assume that in distribution $\mu^N \to \mu^*$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Let $t \in \mathbf{R}_+$ and $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$. By (25) and Lemma 9, we have:

$$\begin{aligned}
&\mathbf{E}\left[\mathbf{\Lambda}_t[f](\mu^N)\right]\\
&= \mathbf{E}\Big[\Big|\langle f, \mu_0^N\rangle - \langle f, \mu_0\rangle + \langle f, M_t^N\rangle + \langle f, V_t^N\rangle + \sum_{k=0}^{\lfloor Nt\rfloor-1} \langle f, R_k^N\rangle + \frac{1}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt\rfloor-1} \sum_{i=1}^N \nabla f(W_k^i) \cdot \varepsilon_k^i\Big|\Big]\\
&\leq \mathbf{E}[\,|\langle f, \mu_0^N\rangle - \langle f, \mu_0\rangle|\,] + \mathbf{E}[|\langle f, V_t^N\rangle|] + \sqrt{\mathbf{E}[\langle f, M_t^N\rangle^2]} + \mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1} \langle f, R_k^N\rangle\Big|\Big]\\
&\quad + \sqrt{\mathbf{E}\Big[\Big|\frac{1}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt\rfloor-1} \sum_{i=1}^N \nabla f(W_k^i) \cdot \varepsilon_k^i\Big|^2\Big]} \leq C\|f\|_{\mathcal{C}^{2,\gamma_*}}\Big[\frac{1}{\sqrt{N}} + \frac{1}{N} + \frac{1}{N^{2\beta-1}} + \frac{1}{N^\beta}\Big],
\end{aligned}$$

where the bound $\mathbf{E}[|\langle f, \mu_0^N\rangle - \langle f, \mu_0\rangle|] \le C\|f\|_{\mathcal{C}^{2,\gamma_*}}/\sqrt{N}$ follows from (70) and the fact that the initial coefficients are i.i.d. (see **A4**). Therefore, since $\beta > 1/2$, for all $t \in \mathbf{R}_+$ and $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$,

$$\lim_{N\to+\infty} \mathbf{E}\left[\mathbf{\Lambda}_t[f](\mu^N)\right] = 0. \tag{73}$$

By (Ethier and Kurtz, 2009, Lemma 7.7 in Chapter 3), the complementary of the set

$$\mathcal{C}(\mu^*) = \{t \ge 0,\, \mathbf{P}(\mu^*_{t^-} = \mu^*_t) = 1\}$$

is at most countable. Let $t_* \in \mathcal{C}(\mu^*)$. Denoting by $\mathsf{D}(\mathbf{\Lambda}_{t_*}[f])$ the set of discontinuity points of $\mathbf{\Lambda}_{t_*}[f]$, we recall that from Lemma 17, $m \notin \mathsf{D}(\mathbf{\Lambda}_{t_*}[f])$ if m is continuous at t_*. Then, we have:

$$\mathbf{P}\big(\mu^* \in \mathsf{D}(\mathbf{\Lambda}_{t_*}[f])\big) = 0.$$

By (Billingsley, 1999, Theorem 2.7), it then holds:

$$\lim_{N'\to+\infty} \mathbf{\Lambda}_{t_*}[f](\mu^N) = \mathbf{\Lambda}_{t_*}[f](\mu^*) \text{ in distribution, } \forall t_* \in \mathcal{C}(\mu^*). \tag{74}$$

By uniqueness of the limit in distribution, (73) and (74) imply that for all $t_* \in \mathcal{C}(\mu^*)$ and $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$, a.s. $\mathbf{\Lambda}_{t_*}[f](\mu^*) = 0$. It then remains to show that a.s. for all $t \in \mathbf{R}_+$ and $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$, $\mathbf{\Lambda}_t[f](\mu^*) = 0$. To do so we use a standard continuity argument.

First of all, for all $m \in \mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ and $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$, the function $t \in \mathbf{R}_+ \mapsto \mathbf{\Lambda}_t[f](m)$ is right-continuous. Moreover, there exists a countable subset $\mathcal{R}_{\mu^*}$ of $\mathcal{C}(\mu^*)$ such that for all $t \ge 0$ and $\epsilon > 0$, there exists $s \in \mathcal{R}_{\mu^*}$, $s \in [t, t+\varepsilon]$. Thus, for all $f \in \mathcal{C}_c^\infty(\mathbf{R}^d)$, it holds a.s. for all $t \in \mathbf{R}_+$ $\mathbf{\Lambda}_t[f](\mu^*) = 0$.

Secondly, for all $m \in \mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ and $t \ge 0$, using the dominated convergence theorem, the function $f \in \mathcal{H}^{L,\gamma}(\mathbf{R}^d) \mapsto \mathbf{\Lambda}_t[f](m)$ is continuous (because $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{1,\gamma}(\mathbf{R}^d)$, by (6)). Furthermore, $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ admits a dense and countable subset of elements in $\mathcal{C}_c^\infty(\mathbf{R}^d)$. Thus, a.s. for all $f \in \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ and all $t \in \mathbf{R}_+$ $\mathbf{\Lambda}_t[f](\mu^*) = 0$.

Note also that $\mathcal{C}_b^\infty(\mathbf{R}^d) \subset \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ since $2\gamma > d$. This ends the proof of the proposition.
■

Note that we have not used Proposition 15 in the proof of Proposition 18.

2.3 Uniqueness of the limit equation in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$ and proof of Theorem 1

2.3.1 Uniqueness of the limit equation in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$

Proposition 19 *There exists a unique solution to* (10) *in the space* $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$.

Proof We have already proved the existence. Let us now prove that there is at most one solution to (10) in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. The proof of the uniqueness of (10) relies on arguments developed in (Piccoli and Rossi, 2016; Piccoli et al., 2015) and is divided into several steps.

Step 1. Preliminary considerations.

If $\mu^\star$ is solution to (10), then for all $f \in \mathcal{C}_b^\infty(\mathbf{R}^d)$, $s \geq 0 \mapsto \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^\star\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \mu_s^\star\rangle \pi(\mathrm{d}x, \mathrm{d}y)$ is continuous (by the dominated convergence theorem). This implies that for all $f \in \mathcal{C}_b^\infty(\mathbf{R}^d)$ and $t \in \mathbf{R}_+$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle f, \mu_t^\star\rangle = \int_{\mathbf{R}^d} \nabla f(w) \cdot \mathbf{V}[\mu_t^\star](w)\mu_t^\star(\mathrm{d}w),$$

where $\mathbf{V} : \mu \in \mathcal{P}(\mathbf{R}^d) \mapsto \mathbf{V}[\mu]$ is defined by

$$\mathbf{V}[\mu] : w \in \mathbf{R}^d \mapsto \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu\rangle)\nabla\sigma_*(w, x)\pi(\mathrm{d}x, \mathrm{d}y) \in \mathbf{R}^d.$$

Adopting the terminology of (Santambrogio, 2015, Section 4.1.2), $\mu^\star$ is thus a *weak solution*[3] of the measure-valued equation

$$\begin{cases} \partial_t \mu_t^\star = \mathrm{div}\,(\mathbf{V}[\mu_t^\star]\mu_t^\star) \\ \mu_0^\star = \mu_0. \end{cases} \tag{75}$$

Therefore, to prove the uniqueness result in Proposition 19, it is enough to show that (75) has a unique weak solution in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. To this end, we consider two solutions $\mu^1 = \{t \mapsto \mu_t^1, t \geq 0\}$ and $\mu^2 = \{t \mapsto \mu_t^2, t \geq 0\}$ of (75) in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$, and we introduce the following mappings

$$\mathbf{v}^1 : (t, x) \in \mathbf{R}_+ \times \mathbf{R}^d \mapsto \mathbf{V}[\mu_t^1](x) \quad \text{and} \quad \mathbf{v}^2 : (t, x) \in \mathbf{R}_+ \times \mathbf{R}^d \mapsto \mathbf{V}[\mu_t^2](x).$$

Step 2. In this step, we prove some basic regularity properties of $\mathbf{V}$, $\mathbf{v}^1$, and $\mathbf{v}^2$.

Let us first prove that the velocity fields $\mathbf{v}^1$ and $\mathbf{v}^2$ are globally Lipschitz continuous over $\mathbf{R}_+ \times \mathbf{R}^d$. Let $\mu \in \{\mu^1, \mu^2\}$ and set $v(t, x) = \mathbf{V}[\mu_t](x)$. For $0 \leq s \leq t$ and $w_1, w_2 \in \mathbf{R}^d$, we have

$$|v(t, w_1) - v(s, w_2)| \leq |v(t, w_1) - v(t, w_2)| + |v(t, w_2) - v(s, w_2)|.$$

By **A2**, the function $w \mapsto \mathbf{V}[\mu](w)$ is smooth and $|\nabla\mathbf{V}[\mu](w)| \leq C$ for some $C > 0$ independent of μ and w. Thus, it holds

$$|v(t, w_1) - v(t, w_2)| = |\mathbf{V}[\mu_t](w_1) - \mathbf{V}[\mu_t](w_2)| \leq C|w_1 - w_2|,$$

for some $C > 0$ independent of t, w_1, and w_2. Secondly, for any $x \in \mathcal{X}$, considering (10) with $f = \sigma_*(\cdot, x)$, we obtain

$$\begin{aligned} |\langle \sigma_*(\cdot, x), \mu_s - \mu_t\rangle| &\leq \int_s^t \int_{\mathcal{X}\times\mathcal{Y}} \left|\alpha(y - \langle \sigma_*(\cdot, x'), \mu_r\rangle)\langle \nabla\sigma_*(\cdot, x) \cdot \nabla\sigma_*(\cdot, x'), \mu_r\rangle\right| \pi(\mathrm{d}x', \mathrm{d}y)\mathrm{d}r \\ &\leq C|t - s|, \end{aligned}$$

leading to

$$|v(t, w_2) - v(s, w_2)| = \left|\int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle \sigma_*(\cdot, x), \mu_s - \mu_t\rangle \nabla\sigma_*(w_2, x)\pi(\mathrm{d}x, \mathrm{d}y)\right| \leq C|t - s|.$$

3. We mention that according to (Santambrogio, 2015, Proposition 4.2), the two notions of solutions of (75) (namely the weak solution and the *distributional* solution) are equivalent.

Thus, there exists $C > 0$ such that for $0 \le s \le t$ and $w_1, w_2 \in \mathbf{R}^d$, $|v(t, w_1) - v(s, w_2)| \le C(|t-s| + |w_1 - w_2|)$, which proves that v is globally Lipschitz. Now we claim that there exists $L' > 0$ such that for every $\mu, \nu \in \mathcal{P}_1(\mathbf{R}^d)$,

$$\|\mathbf{V}[\mu] - \mathbf{V}[\nu]\|_\infty := \sup_{w \in \mathbf{R}^d} |\mathbf{V}[\mu](w) - \mathbf{V}[\nu](w)| \le L' \mathsf{W}_1(\mu, \nu). \tag{76}$$

By **A2**, there exists $C > 0$ such that for all $\mu, \nu \in \mathcal{P}_1(\mathbf{R}^d)$ and all $w \in \mathbf{R}^d$,

$$\begin{aligned} |\mathbf{V}[\mu](w) - \mathbf{V}[\nu](w)| &= \left| \int_{\mathcal{X} \times \mathcal{Y}} \alpha(\langle \sigma_*(\cdot, x), \nu \rangle - \langle \sigma_*(\cdot, x), \mu \rangle) \nabla_W \sigma_*(w, x) \pi(\mathrm{d}x, \mathrm{d}y) \right| \\ &\le C \int_{\mathcal{X} \times \mathcal{Y}} |\langle \sigma_*(\cdot, x), \nu \rangle - \langle \sigma_*(\cdot, x), \mu \rangle| \, \pi(\mathrm{d}x, \mathrm{d}y) \le C \mathsf{W}_1(\mu, \nu), \end{aligned}$$

where the last inequality is obtained by the Lipschitz continuity of $\sigma_*(\cdot, x)$ (which is uniform in $x \in \mathcal{X}$).

Step 3. End of the proof of Proposition 19.

Since v is globally Lipschitz, we can introduce the flows $(\phi_t^1)_{t \in [0,T]}$ and $(\phi_t^2)_{t \in [0,T]}$ with respect to μ^1 and μ^2. By (Villani, 2003, Theorem 5.34), one has

$$\mu_t^1 = \phi_t^1 \# \mu_0, \quad \mu_t^2 = \phi_t^2 \# \mu_0, \quad \forall t \ge 0. \tag{77}$$

The symbol # stands for the pushforward of a measure. Let $L > 0$ be a constant such that $|v_t^i(w_1) - v_t^i(w_2)| \le L|w_1 - w_2|$ for all $i = 1, 2$, $t \in \mathbf{R}_+$ and $w_1, w_2 \in \mathbf{R}^d$ (which exists by the previous step). Then by (Piccoli and Rossi, 2016, Proposition 4), it holds for all $\mu, \nu \in \mathcal{P}_1(\mathbf{R}^d)$,

$$\mathsf{W}_1(\phi_t^1 \# \mu, \phi_t^2 \# \nu) \le e^{Lt} \mathsf{W}_1(\mu, \nu) + \frac{e^{Lt} - 1}{L} \sup_{0 \le s \le t} \|v_s^1 - v_s^2\|_\infty. \tag{78}$$

We are now in position to prove that $\mu^1 = \mu^2$. We use the techniques introduced in (Piccoli et al., 2015). Let us now consider $T > 0$, and introduce

$$t_0 := \inf\{t \in [0, T], \, \mathsf{W}_1(\mu_t^1, \mu_t^2) \neq 0\}.$$

We shall prove that $t_0 = T$. Assume that $t_0 < T$. By (77) and (78), we have, for $0 \le s \le T - t_0$,

$$\mathsf{W}_1(\mu_{t_0+s}^1, \mu_{t_0+s}^2) \le e^{Ls} \mathsf{W}_1(\mu_{t_0}^1, \mu_{t_0}^2) + \frac{e^{Ls} - 1}{L} \sup_{t_0 \le \tau \le t_0 + s} \|v_\tau^1 - v_\tau^2\|_\infty$$

By continuity, $\mathsf{W}_1(\mu_{t_0}^1, \mu_{t_0}^2) = 0$. For s small enough such that $e^{Ls} - 1 < 2Ls$, we obtain, using (76),

$$\mathsf{W}_1(\mu_{t_0+s}^1, \mu_{t_0+s}^2) \le 2sL' \sup_{t_0 \le \tau \le t_0 + s} \mathsf{W}_1(\mu_\tau^1, \mu_\tau^2).$$

Then, for $0 \le s' \le s < \min(1/2L', T - t_0)$, applying the last inequality for s' gives

$$\mathsf{W}_1(\mu_{t_0+s'}^1, \mu_{t_0+s'}^2) < \sup_{t_0 \le \tau \le t_0 + s} \mathsf{W}_1(\mu_\tau^1, \mu_\tau^2),$$

which is not possible. Hence, $t_0 = T$, and again, by continuity, we conclude that $\mathsf{W}_1(\mu_t^1, \mu_t^2) = 0$, $\forall t \in [0,T]$. Therefore, $\mu^1 = \mu^2$. We have thus proved that (10) admits a unique solution in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$, which is the desired result. ■

2.3.2 End of the proof of Theorem 1

We can now prove Theorem 1.
Proof [Proof of Theorem 1] By Proposition 10, the sequence $(\mu^N)_{N\geq 1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Let $\bar{\mu}^1$ and $\bar{\mu}^2$ be two limit points of $(\mu^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Let $j \in \{1,2\}$. By Lemma 15, a.s. $\bar{\mu}^j \in \mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. According to Proposition 18, $\bar{\mu}^j$ satisfies a.s. (10). Let $\mu^\star \in \mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$ be the unique solution of (10) (see Proposition 19). Therefore, one has that a.s. $\bar{\mu}^j = \mu^\star$ in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. Note that this implies also that $\mu^\star \in \mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ and then that a.s. $\bar{\mu}^j = \mu^\star$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$. Therefore, $(\mu^N)_{N\geq 1}$ converges in distribution to $\mu^\star$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$ (and then the convergence holds in probability). This ends the proof of Theorem 1. ■

3. Proof of Theorem 2

In this section, we prove Theorem 2. Recall that $\bar{\mu} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$ is given by Corollary 1 and that the fluctuation process is defined by (see (11)):

$$\eta^N = \sqrt{N}(\mu^N - \bar{\mu}),\ N \geq 1.$$

Throughout this section, we assume that **A1-A7** hold.

3.1 Relative compactness of $(\eta^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$

To prove the relative compactness of $(\eta^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$, we will use Proposition 38 with $\mathcal{H}_1 = \mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$ and $\mathcal{H}_2 = \mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$. Mimicking the proof of (Sznitman, 1991, Theorem 1.1), there exists a unique, trajectorial and in law, solution of

$$\begin{cases} \mathrm{d}X_t = \alpha \int_{\mathcal{X}\times\mathcal{Y}} (y - \langle \sigma_*(\cdot, x), \hat{\mu}_t\rangle) \nabla_W \sigma_*(X_t, x) \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}t, \\ X_0 \sim \mu_0, \quad \hat{\mu}_t = \mathrm{Law}(X_t). \end{cases} \tag{79}$$

Denote by $\hat{\mu} \in \mathcal{P}(\mathcal{C}(\mathbf{R}_+, \mathbf{R}^d))$ this solution. The mapping $t \geq 0 \mapsto \hat{\mu}_t$ satisfies Equation (10). In addition, using **A2**, it is straightforward to show that the function $t \mapsto \hat{\mu}_t$ lies in $\mathcal{C}(\mathbf{R}_+, \mathcal{P}_1(\mathbf{R}^d))$. Since $\bar{\mu}$ is the unique solution of (10) (see Proposition 19), $\hat{\mu} = \bar{\mu}$. Therefore, we introduce, as it is customary, the particle system defined as follows: for $N \geq 1$, let $\bar{X}^i = \{t \mapsto \bar{X}^i_t, t \in \mathbf{R}_+\}$ ($i \in \{1, \ldots, N\}$) be the N independent processes satisfying:

$$\begin{cases} \bar{X}^i_t = W^i_0 + \int_0^t \alpha \int_{\mathcal{X}\times\mathcal{Y}} (y - \langle \sigma_*(\cdot, x), \bar{\mu}_s\rangle) \nabla_W \sigma_*(\bar{X}^i_s, x) \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s,\ t \in \mathbf{R}_+ \\ \bar{\mu}_t = \mathrm{Law}(\bar{X}^i_t). \end{cases} \tag{80}$$

We then introduce its empirical measure:

$$\bar{\mu}^N_t = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{X}^i_t},\ N \geq 1,\ t \in \mathbf{R}_+. \tag{81}$$

By **A2**, there exists $C_0 > 0$ such that a.s. for all $0 \le s \le t$, it holds for all $i \in \{1, \dots, N\}$:

$$|\bar X_t^i - \bar X_s^i| \le C_0(t-s). \tag{82}$$

In particular, $t \in \mathbf{R}_+ \mapsto \bar X_t^i \in \mathbf{R}^d$ is a.s. continuous for all $i \in \{1, \dots, N\}$. Because μ_0 is compactly supported (see indeed **A6**), one deduces that there exists $C > 0$ such that a.s. for all $T > 0$ and for all $i \in \{1, \dots, N\}$:

$$\sup_{t\in[0,T]} |\bar X_t^i| \le C(1+T). \tag{83}$$

Thus, for any $\beta \ge 0$, a.s. $\bar\mu^N \in \mathcal{C}(\mathbf{R}_+, \mathcal{C}^{1,\beta}(\mathbf{R}^d)')$. We now define, for $N \ge 1$,

$$\Upsilon^N := \sqrt{N}(\mu^N - \bar\mu^N) \ \text{ and } \ \Theta^N := \sqrt{N}(\bar\mu^N - \bar\mu). \tag{84}$$

It then holds:

$$\eta^N = \Upsilon^N + \Theta^N. \tag{85}$$

For all $N \ge 1$ and for any $\beta \ge 0$, since a.s. $\mu^N \in \mathcal{D}(\mathbf{R}_+, \mathcal{C}^{0,\beta}(\mathbf{R}^d)')$, one has

$$\text{a.s. } \Upsilon^N \in \mathcal{D}(\mathbf{R}_+, \mathcal{C}^{1,\beta}(\mathbf{R}^d)').$$

In particular a.s. $\Upsilon^N \in \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{1-J_1,j_1}(\mathbf{R}^d))$ because $\mathcal{H}^{J_1-1,j_1}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{1,j_1}(\mathbf{R}^d)$ (according to (4) and since $J_1 - 1 > d/2 + 1$). On the other hand, since $\bar\mu \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$ (see Corollary 1) and since a.s. $\bar\mu^N \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$, it holds for all $N \ge 1$:

$$\text{a.s. } \Theta^N \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d)). \tag{86}$$

We start with the following lemma.

Lemma 20 *Let $\beta \ge 3/4$ and assume* **A1**-**A7**. *Then, for all $T > 0$, we have*

$$\sup_{N\ge1}\sup_{t\in[0,T]} \mathbf{E}\big[\|\Theta_t^N\|^2_{\mathcal{H}^{-J_1,j_1}} + \|\Upsilon_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big] < +\infty.$$

In particular, $\sup_{N\ge1}\sup_{t\in[0,T]} \mathbf{E}[\|\eta_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}] < +\infty$.

Proof Let $T > 0$. In all this proof, $f \in \mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$ and $\{f_a\}_{a\ge1}$ is an orthonormal basis of $\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$. In the following, C denotes a constant independent of $N \ge 1$, $t \in [0,T]$, $(x,y) \in \mathcal{X} \times \mathcal{Y}$, and the test function f, which can change from one occurence to another. The proof is divided into two steps.

Step 1. Upper bound on $\mathbf{E}\left[\|\Theta_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right]$.

Since $\bar X_s^1, \dots, \bar X_s^N$ are i.i.d. with law $\bar\mu_s$ (see (80)), using (83), the fact that $\bar\mu \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$ (see Corollary 1), and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{0,\gamma}(\mathbf{R}^d)$, it holds for all $t \in [0,T]$:

$$\begin{aligned}
\mathbf{E}\left[\langle f, \Theta_t^N\rangle^2\right] = \mathbf{E}\Big[\Big\langle f, \sqrt{N}(\bar\mu_t^N - \bar\mu_t)\Big\rangle^2\Big] &= \mathbf{E}\Big[\Big|\frac{1}{\sqrt N}\sum_{i=1}^N \left[f(\bar X_t^i) - \langle f, \bar\mu_t\rangle\right]\Big|^2\Big]\\
&= \frac1N \sum_{i=1}^N \mathbf{E}\Big[\Big|f(\bar X_t^i) - \langle f, \bar\mu_t\rangle\Big|^2\Big]\\
&\le \frac2N \sum_{i=1}^N \mathbf{E}\big[|f(\bar X_t^i)|^2\big] + |\langle f, \bar\mu_t\rangle|^2 \le C\|f\|^2_{\mathcal{H}^{L,\gamma}}.
\end{aligned}$$

Taking $f = f_a$ in the previous inequality, summing over $a \in \mathbf{N}^*$, and using the fact that $\mathcal{H}^{J_1,j_1}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$, one deduces that:

$$\sup_{N\geq 1}\sup_{t\in[0,T]} \mathbf{E}\left[\|\Theta_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right] \leq C. \tag{87}$$

Step 2. Upper bound on $\mathbf{E}\left[\|\Upsilon_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right]$.

Recall that a.s. $\Upsilon^N \in \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{1-J_1,j_1}(\mathbf{R}^d))$. Thus, a.s. $\Upsilon^N \in \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_1,j_1}(\mathbf{R}^d))$. We then have for all $t \in \mathbf{R}_+$,

$$\|\Upsilon_t^N\|^2_{\mathcal{H}^{-J_1,j_1}} = \sum_{a=1}^{+\infty} \langle f_a, \Upsilon_t^N\rangle^2. \tag{88}$$

For all $i \in \{1,\ldots,N\}$, a.s. $t \in \mathbf{R}_+ \mapsto f(\bar X_t^i) \in \mathcal{C}^1(\mathbf{R}_+)$. Indeed, $f \in \mathcal{C}^1(\mathbf{R}^d)$ and a.s. $t \in \mathbf{R}_+ \mapsto \bar X_t^i$ (see (80)) is $\mathcal{C}^1$ (because a.s. $s \in \mathbf{R}_+ \mapsto \alpha \int_{\mathcal{X}\times\mathcal{Y}} (y - \langle \sigma_*(\cdot,x), \bar\mu_s\rangle) \nabla_W \sigma_*(\bar X_s^i, x)\pi(\mathrm{d}x,\mathrm{d}y)$ is continuous by the dominated convergence theorem). Therefore, it holds $\langle f, \bar\mu_t^N\rangle = \langle f, \bar\mu_0^N\rangle + \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\langle f, \bar\mu_s^N\rangle \mathrm{d}s$ and therefore,

$$\langle f, \bar\mu_t^N\rangle = \langle f, \bar\mu_0^N\rangle + \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot,x), \bar\mu_s\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \bar\mu_s^N\rangle \pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s. \tag{89}$$

Thus, by definition of Υ_t^N (see (84)) and using also (25), we have:

$$\begin{aligned}\langle f, \Upsilon_t^N\rangle &= \sqrt{N}\underbrace{\langle f, (\mu_0^N - \bar\mu_0^N)\rangle}_{=0}\\ &\quad + \sqrt{N}\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot,x), \mu_s^N\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \mu_s^N\rangle \pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\ &\quad - \sqrt{N}\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot,x), \bar\mu_s\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \bar\mu_s^N\rangle \pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\ &\quad + \sqrt{N}\langle f, M_t^N\rangle + \sqrt{N}\langle f, V_t^N\rangle + \sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor -1}\langle f, R_k^N\rangle + \frac{\sqrt{N}}{N^{1+\beta}}\sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i.\end{aligned} \tag{90}$$

Furthermore, it holds,

$$\begin{aligned}&\sqrt{N}(y - \langle \sigma_*(\cdot,x), \mu_s^N\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \mu_s^N\rangle\\ &= (y - \langle \sigma_*(\cdot,x), \mu_s^N\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \Upsilon_s^N\rangle - \langle \sigma_*(\cdot,x), \Upsilon_s^N\rangle\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \bar\mu_s^N\rangle\\ &\quad + \sqrt{N}(y - \langle \sigma_*(\cdot,x), \bar\mu_s\rangle - \langle \sigma_*(\cdot,x), \bar\mu_s^N - \bar\mu_s\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \bar\mu_s^N\rangle.\end{aligned} \tag{91}$$

Consequently, plugging (91) in (90), we obtain for $t \in \mathbf{R}_+$:

$$\begin{aligned}\langle f, \Upsilon_t^N\rangle &= \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \Upsilon_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \\ &\quad - \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle \sigma_*(\cdot, x), \Upsilon_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \bar\mu_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \\ &\quad - \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle \sigma_*(\cdot, x), \sqrt{N}(\bar\mu_s^N - \bar\mu_s)\rangle\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \bar\mu_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \\ &\quad + \sqrt{N}\langle f, M_t^N\rangle + \sqrt{N}\langle f, V_t^N\rangle + \sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor - 1}\langle f, R_k^N\rangle + \frac{\sqrt{N}}{N^{1+\beta}}\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i.\end{aligned} \tag{92}$$

By Lemma 41 (in Appendix B, see also Remark 42), one then has for all $t \in \mathbf{R}_+$:

$$\langle f, \Upsilon_t^N\rangle^2 \le \mathbf{A}_t^N[f] + \mathbf{B}_t^N[f], \tag{93}$$

where

$$\begin{aligned}\mathbf{A}_t^N[f] &= 2\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle f, \Upsilon_s^N\rangle(y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \Upsilon_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \\ &\quad - 2\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle f, \Upsilon_s^N\rangle\langle \sigma_*(\cdot, x), \Upsilon_s^N\rangle)\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \bar\mu_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \\ &\quad - 2\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle f, \Upsilon_s^N\rangle\langle \sigma_*(\cdot, x), \sqrt{N}(\bar\mu_s^N - \bar\mu_s)\rangle\langle \nabla f \cdot \nabla\sigma_*(\cdot, x), \bar\mu_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s \\ &=: \mathbf{I}_t^N[f] + \mathbf{J}_t^N[f] + \mathbf{K}_t^N[f],\end{aligned}$$

and

$$\begin{aligned}\mathbf{B}_t^N[f] &= \sum_{k=0}^{\lfloor Nt\rfloor-1}\Big[2\langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle\sqrt{N}\langle f, M_k^N\rangle + 4N\langle f, M_k^N\rangle^2\Big] \\ &\quad + \sum_{k=0}^{\lfloor Nt\rfloor-1}\Big[2\langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle\sqrt{N}\langle f, R_k^N\rangle + 4N|\langle f, R_k^N\rangle|^2\Big] \\ &\quad + \sum_{k=0}^{\lfloor Nt\rfloor-1}\Big[\frac{2\sqrt{N}}{N^{1+\beta}}\langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i + \frac{4}{N^{1+2\beta}}\Big|\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big] \\ &\quad + \sum_{k=0}^{\lfloor Nt\rfloor-1}\Big[2\langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle\mathbf{a}_k^N[f] + 4|\mathbf{a}_k^N[f]|^2\Big] - 2\sqrt{N}\int_0^t \langle f, \Upsilon_s^N\rangle\mathbf{L}_s^N[f]\mathrm{d}s,\end{aligned}$$

with

$$\mathbf{L}_s^N[f] = \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \mu_s^N\rangle)\langle \nabla f\cdot\nabla\sigma_*(\cdot, x), \mu_s^N\rangle\pi(\mathrm{d}x, \mathrm{d}y) \ \text{ and } \ \mathbf{a}_k^N[f] = \sqrt{N}\int_{\frac{k}{N}}^{\frac{k+1}{N}} \mathbf{L}_s^N[f]\mathrm{d}s. \tag{94}$$

Using (93) and (88),

$$\|\Upsilon_t^N\|^2_{\mathcal{H}^{-J_1,j_1}} \le \sum_{a=1}^{+\infty} \mathbf{A}_t^N[f_a] + \sum_{a=1}^{+\infty} \mathbf{B}_t^N[f_a]. \tag{95}$$

By Lemma 39, detailed in Appendix B, and since $\beta \ge 3/4$, one has for all $f \in \mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$ and $t \in \mathbf{R}_+$,

$$\mathbf{E}\big[\mathbf{B}_t^N[f]\big] \le C\big(\|f\|^2_{\mathcal{H}^{L,\gamma}} + \mathbf{E}\left[\int_0^t \langle f, \Upsilon_s^N\rangle^2 \mathrm{d}s\right]\big).$$

Thus, recalling $\mathcal{H}^{J_1,j_1}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ and (88), we obtain,

$$\sum_{a=1}^{+\infty} \mathbf{E}[\mathbf{B}_t^N[f_a]] \le C + C\int_0^t \mathbf{E}\big[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big]\mathrm{d}s. \tag{96}$$

Let us now provide a similar upper bound on $\sum_{a=1}^{+\infty} \mathbf{A}_t^N[f_a]$. By (83) and because $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$ (see (6)),

$$\begin{aligned}|\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \bar\mu_s^N\rangle| = \Big|\frac{1}{N}\sum_{i=1}^N \nabla f(\bar X_s^i)\cdot \nabla\sigma_*(\bar X_s^i, x)\Big| &\le \frac{C\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N (1+|\bar X_s^i|^{\gamma_*})\\ &\le C\|f\|_{\mathcal{H}^{L,\gamma}}.\end{aligned} \tag{97}$$

Then, using **A2** and since $j_1 > d/2$, it holds: we have using (97):

$$\begin{aligned}&-2\int_0^t\int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle f, \Upsilon_s^N\rangle\langle \sigma_*(\cdot,x), \Upsilon_s^N)\rangle\langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \bar\mu_s^N\rangle \pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\ &\le 4\alpha\int_0^t\int_{\mathcal{X}\times\mathcal{Y}} \big(\langle f,\Upsilon_s^N\rangle^2 + C\|f\|^2_{\mathcal{H}^{L,\gamma}}\|\sigma_*(W,x)\|^2_{\mathcal{H}^{J_1,j_1}}\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big)\,\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\ &\le C\int_0^t \big(\langle f,\Upsilon_s^N\rangle^2 + C\|f\|^2_{\mathcal{H}^{L,\gamma}}\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big)\,\mathrm{d}s.\end{aligned}$$

Thus, $\sum_{a=1}^{+\infty}\mathbf{E}[\mathbf{J}_t^N[f_a]] \le C\int_0^t \mathbf{E}\big[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big]\mathrm{d}s$. Let us now study $\sum_{a=1}^{+\infty}\mathbf{K}_t^N[f_a]$. Since $\bar X_s^1,\ldots,\bar X_s^N$ are i.i.d. with law $\bar\mu_s$ (see (80)) and because σ_* is bounded (see **A2**), we have:

$$\begin{aligned}\mathbf{E}\left[\big\langle \sigma_*(\cdot,x), \bar\mu_s^N - \bar\mu_s\big\rangle^2\right] &= \mathbf{E}\Big[\Big|\frac{1}{N}\sum_{i=1}^N \sigma_*(\bar X_s^i, x) - \langle\sigma_*(\cdot,x),\bar\mu_s\rangle\Big|^2\Big]\\ &= \frac{1}{N^2}\sum_{i=1}^N \mathbf{E}\left[(\sigma_*(\bar X_s^i,x) - \langle \sigma_*(\cdot,x),\bar\mu_s\rangle)^2\right] \le \frac{C}{N}.\end{aligned} \tag{98}$$

Hence, $\mathbf{E}[\langle \sigma_*(\cdot,x), \sqrt{N}(\bar\mu_s^N-\bar\mu_s)\rangle^2]\le C$. Thus, using in addition (97), one deduces that

$$\begin{aligned}
&\mathbf{E}\left[-2\int_0^t\int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle f,\Upsilon_s^N\rangle\langle \sigma_*(\cdot,x),\sqrt{N}(\bar\mu_s^N-\bar\mu_s)\rangle\langle \nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s^N\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\right]\\
&\le C\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\Big\{\mathbf{E}[\langle f,\Upsilon_s^N\rangle^2]+\mathbf{E}\left[\langle \sigma_*(\cdot,x),\sqrt{N}(\bar\mu_s^N-\bar\mu_s)\rangle^2\langle \nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s^N\rangle^2\right]\Big\}\,\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&\le C\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\Big\{\mathbf{E}[\langle f,\Upsilon_s^N\rangle^2]+\mathbf{E}\left[\langle \sigma_*(\cdot,x),\sqrt{N}(\bar\mu_s^N-\bar\mu_s)\rangle^2\right]\|f\|^2_{\mathcal{H}^{L,\gamma}}\Big\}\,\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&\le C\int_0^t\big\{\mathbf{E}[\langle f,\Upsilon_s^N\rangle^2]+\|f\|^2_{\mathcal{H}^{L,\gamma}}\big\}\,\mathrm{d}s.
\end{aligned}$$

Therefore, $\sum_{a=1}^{+\infty}\mathbf{E}[\mathbf{K}_t^N[f_a]]\le C+C\int_0^t\mathbf{E}\big[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big]\mathrm{d}s$. It remains to study $\sum_{a=1}^{+\infty}\mathbf{I}_t^N[f_a]$. To this end, for $x\in\mathcal{X}$, introduce the bounded linear operator

$$\mathsf{T}_x : f\in\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)\mapsto \nabla f\cdot\nabla\sigma_*(\cdot,x)\in\mathcal{H}^{J_1-1,j_1}(\mathbf{R}^d). \tag{99}$$

Then, one has:

$$\begin{aligned}
\sum_{a=1}^{+\infty}\mathbf{I}_t^N[f_a] &=\sum_{a=1}^{+\infty}2\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\alpha\langle f_a,\Upsilon_s^N\rangle(y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle)\langle \mathsf{T}_x f_a,\Upsilon_s^N\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&=2\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle)\sum_{a=1}^{+\infty}\langle f_a,\Upsilon_s^N\rangle\langle \mathsf{T}_x f_a,\Upsilon_s^N\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&=2\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle)\left\langle\Upsilon_s^N,\mathsf{T}_x^*\Upsilon_s^N\right\rangle_{-J_1,j_1}\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s.
\end{aligned}$$

Since σ_* is bounded, this implies that:

$$\begin{aligned}
\sum_{a=1}^{+\infty}\mathbf{E}[\mathbf{I}_t^N[f_a]] &\le C\int_0^t\mathbf{E}\Big[\int_{\mathcal{X}\times\mathcal{Y}}|y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle|\,\big|\langle\Upsilon_s^N,\mathsf{T}_x^*\Upsilon_s^N\rangle_{-J_1,j_1}\big|\pi(\mathrm{d}x,\mathrm{d}y)\Big]\mathrm{d}s\\
&\le C\int_0^t\mathbf{E}\Big[\int_{\mathcal{X}\times\mathcal{Y}}(|y|+C)\,|\langle\Upsilon_s^N,\mathsf{T}_x^*\Upsilon_s^N\rangle_{-J_1,j_1}|\pi(\mathrm{d}x,\mathrm{d}y)\Big]\mathrm{d}s
\end{aligned}$$

By Lemma 40, detailed in Appendix B, and since a.s. $\Upsilon^N\in\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{1-J_1,j_1}(\mathbf{R}^d))$, there exists $C>0$ such that for all $x\in\mathcal{X}$, $|\langle\Upsilon_s^N,\mathsf{T}_x^*\Upsilon_s^N\rangle_{\mathcal{H}^{-J_1,j_1}}|\le C\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}$. Hence, since $\mathbf{E}[|y|]<+\infty$, we deduce that:

$$\sum_{a=1}^{+\infty}\mathbf{E}[\mathbf{I}_t^N[f_a]]\le C\int_0^t\mathbf{E}\left[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right]\mathrm{d}s.$$

We have thus proved that $\sum_{a=1}^{+\infty}\mathbf{E}[\mathbf{A}_t^N[f_a]]\le C+C\int_0^t\mathbf{E}[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}]\mathrm{d}s$. In conclusion, using also (96) and (95), we have $\mathbf{E}[\|\Upsilon_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}]\le C+C\int_0^t\mathbf{E}[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}]\mathrm{d}s$. By Gronwall's Lemma, we get:

$$\sup_{N\ge1}\sup_{t\in[0,T]}\mathbf{E}[\|\Upsilon_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}]<+\infty.$$

Together with the first step, this ends the proof of the Lemma (recall the decomposition $\eta^N = \Upsilon^N + \Theta^N$, see (85)). ■

Lemma 21 *Assume* **A1**-**A7**. *Introduce the following* $\boldsymbol{\sigma}$*-algebra (see* (9)*):*

$$\mathfrak{F}_t^N := \mathcal{F}_{\lfloor Nt \rfloor}^N, \; t \in \mathbf{R}_+.$$

Then, for all $f \in \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$*, the two processes*

$$\Big\{ t \mapsto \sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^{N} \nabla f(W_k^i) \cdot \varepsilon_k^i, \; t \in \mathbf{R}_+ \Big\} \text{ and } \big\{ t \mapsto \langle f, M_t^N \rangle, t \in \mathbf{R}_+ \big\} \text{ are } \mathfrak{F}_t^N\text{-martingale.} \tag{100}$$

Proof Recall that by Lemma 8, the first process in (100) is integrable. By (36) and (23), the second process in (100) is integrable. For $0 \le s < t$, we write

$$\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^{N} \nabla f(W_k^i) \cdot \varepsilon_k^i | \mathfrak{F}_s^N \Big] = \sum_{k=0}^{\lfloor Ns \rfloor - 1} \sum_{i=1}^{N} \mathbf{E}\big[\nabla f(W_k^i) \cdot \varepsilon_k^i | \mathcal{F}_{\lfloor Ns \rfloor}^N \big] + \mathfrak{R}_{t,s}^N[f], \tag{101}$$

where

if $\lfloor Nt \rfloor = \lfloor Ns \rfloor$: $\mathfrak{R}_{t,s}^N[f] = 0$, and if $\lfloor Nt \rfloor > \lfloor Ns \rfloor$: $\mathfrak{R}_{t,s}^N[f] = \displaystyle\sum_{k=\lfloor Ns \rfloor}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^{N} \mathbf{E}[\nabla f(W_k^i) \cdot \varepsilon_k^i | \mathcal{F}_{\lfloor Ns \rfloor}^N]$.

When $\lfloor Nt \rfloor > \lfloor Ns \rfloor$, since the W_k^i's are $\mathcal{F}_k^N$-measurable (see (2) and (9)) and the ε_k^i's are centered and independent of $\mathcal{F}_k^N$ (see **A5**), we have, for $k \ge \lfloor Ns \rfloor$ and $i \in \{1, \dots, N\}$:

$$\mathbf{E}\big[\nabla f(W_k^i) \cdot \varepsilon_k^i | \mathcal{F}_{\lfloor Ns \rfloor}^N \big] = \mathbf{E}\Big[\nabla f(W_k^i) \cdot \mathbf{E}[\varepsilon_k^i | \mathcal{F}_k^N] \Big| \mathcal{F}_{\lfloor Ns \rfloor}^N \Big] = 0.$$

Hence, for all $0 \le s < t$, $\mathfrak{R}_{t,s}^N[f] = 0$. Furthermore, for $0 \le k \le \lfloor Ns \rfloor - 1$ and $i \in \{1, \dots, N\}$, $\nabla f(W_k^i) \cdot \varepsilon_k^i$ is $\mathcal{F}_{k+1}^N$-measurable and thus is $\mathcal{F}_{\lfloor Ns \rfloor}^N$-measurable. Therefore, (101) reduces to

$$\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^{N} \nabla f(W_k^i) \cdot \varepsilon_k^i \Big| \mathfrak{F}_s^N \Big] = \sum_{k=0}^{\lfloor Ns \rfloor - 1} \sum_{i=1}^{N} \nabla f(W_k^i) \cdot \varepsilon_k^i.$$

This proves that $\{t \mapsto \sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^{N} \nabla f(W_k^i) \cdot \varepsilon_k^i, \; t \in \mathbf{R}_+\}$ is a $\mathfrak{F}_t^N$-martingale. Let us now prove that the process $\big\{ t \mapsto \langle f, M_t^N \rangle, t \in \mathbf{R}_+ \big\}$ is a $\mathfrak{F}_t^N$-martingale (see (23)). We have, for $0 \le s < t$,

$$\mathbf{E}\big[\langle f, M_t^N \rangle | \mathfrak{F}_s^N \big] = \sum_{k=0}^{\lfloor Ns \rfloor - 1} \mathbf{E}\big[\langle f, M_k^N \rangle \big| \mathcal{F}_{\lfloor Ns \rfloor}^N \big] + \mathfrak{E}_{t,s}^N[f],$$

where

$$\text{if } \lfloor Nt\rfloor = \lfloor Ns\rfloor\text{: } \mathfrak{E}^N_{t,s}[f]=0, \text{ and if } \lfloor Nt\rfloor > \lfloor Ns\rfloor\text{: } \mathfrak{E}^N_{t,s}[f]=\sum_{k=\lfloor Ns\rfloor}^{\lfloor Nt\rfloor-1}\mathbf{E}\big[\langle f,M_k^N\rangle|\mathcal{F}^N_{\lfloor Ns\rfloor}\big].$$

When $\lfloor Nt\rfloor > \lfloor Ns\rfloor$, we have for $k\geq \lfloor Ns\rfloor$, by (41):

$$\mathbf{E}\big[\langle f,M_k^N\rangle|\mathcal{F}^N_{\lfloor Ns\rfloor}\big]=\mathbf{E}\Big[\mathbf{E}\big[\langle f,M_k^N\rangle|\mathcal{F}^N_k\big]\Big|\mathcal{F}^N_{\lfloor Ns\rfloor}\Big]=0.$$

Hence, for all $0\leq s<t$, $\mathfrak{E}^N_{t,s}[f]=0$. In addition, for $k\leq \lfloor Ns\rfloor-1$, $\langle f,M_k^N\rangle$ is $\mathcal{F}^N_{k+1}$-measurable and thus is $\mathcal{F}^N_{\lfloor Ns\rfloor}$-measurable. In conclusion, $\big\{t\mapsto\langle f,M_t^N\rangle, t\in\mathbf{R}_+\big\}$ is a $\mathfrak{F}^N_t$-martingale. This ends the proof of Lemma 21. ■

The following lemma provides the compact containment condition needed to prove that $(\eta^N)_{N\geq1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$.

Lemma 22 *Let $\beta\geq 3/4$ and assume* **A1**-**A7**. *Then, for all $T>0$,*

$$\sup_{N\geq1}\mathbf{E}\Big[\sup_{t\in[0,T]}\|\eta_t^N\|^2_{\mathcal{H}^{-J_2,j_2}}\Big]<+\infty. \tag{102}$$

Proof Let $T>0$. All along the proof, $C>0$ denotes a constant indendent of $t\in[0,T]$ and $N\geq1$, which can change from one occurence to another. Recall that $\eta^N=\Upsilon^N+\Theta^N$, see (85). By (92) and Jensen's inequality, it holds for $f\in\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$ (recall $\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)\hookrightarrow\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$, see (14)),

$$\begin{aligned}\sup_{t\in[0,T]}\langle f,\Upsilon_t^N\rangle^2\leq C\Bigg[&\int_0^T\int_{\mathcal{X}\times\mathcal{Y}}\big|(y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\Upsilon_s^N\rangle\big|^2\pi(\mathrm{d}s,\mathrm{d}y)\mathrm{d}s\\&+\int_0^T\int_{\mathcal{X}\times\mathcal{Y}}\big|\langle\sigma_*(\cdot,x),\Upsilon_s^N\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s^N\rangle\big|^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\&+\int_0^T\int_{\mathcal{X}\times\mathcal{Y}}\Big|\langle\sigma_*(\cdot,x),\sqrt{N}(\bar\mu_s^N-\bar\mu_s)\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s^N\rangle\Big|^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\&+N\sup_{t\in[0,T]}|\langle f,M_t^N\rangle|^2+N\sup_{t\in[0,T]}\big|\langle f,V_t^N\rangle\big|^2+N\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f,R_k^N\rangle\Big|^2\\&+\frac{1}{N^{1+2\beta}}\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^N\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Bigg].\end{aligned} \tag{103}$$

We now consider successively each term in the right-hand-side of (103). By Lemma 20, for $0\leq s\leq T$, we have using also **A2** and $\mathbf{E}[|y|^2]<+\infty$ (see **A3**):

$$\begin{aligned}&\mathbf{E}\Big[\int_{\mathcal{X}\times\mathcal{Y}}|y-\langle\sigma_*(\cdot,x),\mu_s^N\rangle|^2\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\Upsilon_s^N\rangle^2\pi(\mathrm{d}x,\mathrm{d}y)\Big]\\&\leq C\mathbf{E}\Big[\int_{\mathcal{X}\times\mathcal{Y}}(|y|^2+1)\|\nabla f\cdot\nabla\sigma_*(\cdot,x)\|_{\mathcal{H}^{J_1,j_1}}\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\pi(\mathrm{d}x,\mathrm{d}y)\Big]\\&\leq C\|f\|^2_{\mathcal{H}^{J_1+1,j_1}}\mathbf{E}[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}]\leq C\|f\|^2_{\mathcal{H}^{J_1+1,j_1}}.\end{aligned} \tag{104}$$

Let us now study the second term in (103). By (97) and since $\sup_{x\in\mathcal{X}}\|\sigma_*(\cdot,x)\|_{\mathcal{H}^{J_1,j_1}}<+\infty$ (because $j_1>d/2$ and $\sigma_*\in\mathcal{C}_b^\infty(\mathbf{R}^d\times\mathcal{X})$ by **A2**), for $0\le s\le T$,

$$\mathbf{E}\left[\left|\langle\sigma_*(\cdot,x),\Upsilon_s^N\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s^N\rangle\right|^2\right]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}\mathbf{E}\left[\|\Upsilon_s^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}. \quad (105)$$

Let us now consider the third term in (103). By (97) and (98), we have for $0\le s\le T$,

$$\mathbf{E}\Big[\langle\sigma_*(\cdot,x),\sqrt{N}(\bar\mu_s^N-\bar\mu_s)\rangle^2\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s^N\rangle^2\Big]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}. \quad (106)$$

Let us now deal with the fourth term in (103). By Lemma 21, we have using Doob's maximal inequality, $\mathbf{E}[\sup_{t\in[0,T]}\langle f,M_t^N\rangle^2]\le C\mathbf{E}[\langle f,M_T^N\rangle^2]$. Then by Lemma 9 and since $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$ (recall indeed (6)), we obtain

$$N\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f,M_t^N\rangle^2\Big]\le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}. \quad (107)$$

Using (50) and again (6), the fifth term in (103) satisfies:

$$N\mathbf{E}\Big[\sup_{t\in[0,T]}|\langle f,V_t^N\rangle|^2\Big]\le CN^{-1/2}\|f\|^2_{\mathcal{H}^{L,\gamma}}. \quad (108)$$

Using (52), the sixth term in (103) satisfies:

$$\mathbf{E}\Big[N\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f,R_k^N\rangle\Big|^2\Big]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}(1/N+N^3/N^{4\beta}).$$

Let us deal with the last term in the right-hand side of (103) for which we need a more accurate upper bound than (53). By Lemma 21 and Doob's maximal inequality, we have using (47) and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$,

$$\begin{aligned}\frac{1}{N^{1+2\beta}}\mathbf{E}\Big[\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]&\le\frac{C}{N^{1+2\beta}}\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor NT\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]\\&\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}N^{1-2\beta}.\end{aligned}$$

Collecting these bounds, we obtain, from (103), for $f\in\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$,

$$\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f,\Upsilon_t^N\rangle^2\Big]\le C\left(\|f\|^2_{\mathcal{H}^{J_1+1,j_1}}+\|f\|^2_{\mathcal{H}^{L,\gamma}}\right). \quad (109)$$

We now turn to the study of $\mathbf{E}[\sup_{t\in[0,T]}\langle f,\Theta_t^N\rangle^2]$ for $f\in\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$. By (89) and Corollary 1, we have, for all $t\in[0,T]$ (recall that $\Theta_t^N=\sqrt{N}(\bar\mu_t^N-\bar\mu_t)$, see (84)),

$$\langle f,\Theta_t^N\rangle=\langle f,\Theta_0^N\rangle+\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\alpha(y-\langle\sigma_*(\cdot,x),\bar\mu_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\Theta_s^N\rangle\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s. \quad (110)$$

By Jensen's inequality together with **A2** and (87), one has:

$$\begin{aligned}
&\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f,\Theta_t^N\rangle^2\Big]\\
&\le C\mathbf{E}\big[|\langle f,\Theta_0^N\rangle|^2\big]+C\int_0^T\int_{\mathcal{X}\times\mathcal{Y}}(|y|^2+1)\mathbf{E}\left[\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\Theta_s^N\rangle^2\right]\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&\le C\|f\|^2_{\mathcal{H}^{J_1,j_1}}\mathbf{E}\big[\|\Theta_0^N\|^2_{\mathcal{H}^{-J_1,j_1}}\big]\\
&\quad+C\int_0^T\int_{\mathcal{X}\times\mathcal{Y}}(|y|^2+1)\,\|\nabla f\cdot\nabla\sigma_*(\cdot,x)\|^2_{\mathcal{H}^{J_1,j_1}}\,\mathbf{E}\left[\|\Theta_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right]\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&\le C\|f\|^2_{\mathcal{H}^{J_1+1,j_1}}.
\end{aligned}\tag{111}$$

Let $\{f_a\}_{a\ge1}$ be an orthonormal basis of $\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$. Let us recall that (see (14)) $\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)\hookrightarrow_{\text{H.S.}}\mathcal{H}^{J_1+1,j_1}(\mathbf{R}^d)$ and $\mathcal{H}^{J_1+1,j_1}(\mathbf{R}^d)\hookrightarrow\mathcal{H}^{L,\gamma}(\mathbf{R}^d)$. Then, by (109) and (111), we obtain, since $\beta\ge 3/4$,

$$\begin{aligned}
\mathbf{E}\Big[\sup_{t\in[0,T]}\|\eta_t^N\|^2_{\mathcal{H}^{-J_2,j_2}}\Big]&=\mathbf{E}\Big[\sup_{t\in[0,T]}\sum_{a\ge1}\langle f_a,\eta_t^N\rangle^2\Big]\le\sum_{a\ge1}\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f_a,\eta_t^N\rangle^2\Big]\\
&\le 2\sum_{a\ge1}\Big(\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f_a,\Upsilon_t^N\rangle^2\Big]+\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f_a,\Theta_t^N\rangle^2\Big]\Big)\\
&\le C\sum_{a\ge1}\big(\|f_a\|^2_{\mathcal{H}^{J_1+1,j_1}}+\|f_a\|^2_{\mathcal{H}^{L,\gamma}}\big)\le C.
\end{aligned}$$

This conclude s the proof of the lemma. ■

The following result provides the regularity condition needed to prove that $(\eta^N)_{N\ge1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$.

Lemma 23 *Let $\beta\ge 3/4$ and assume* **A1**-**A7**. *Let $T>0$. Then, there exists $C>0$ such that for all $\delta>0$ and $0\le r<t\le T$ such that $t-r\le\delta$, one has*

$$\mathbf{E}\left[\big\|\eta_t^N-\eta_r^N\big\|^2_{\mathcal{H}^{-J_2,j_2}}\right]\le C\left[\delta^2+\frac{N\delta+1}{N}+\frac{1}{\sqrt{N}}+(N\delta+1)^2\Big(\frac{1}{N^3}+\frac{1}{N^{4\beta-1}}\Big)+\frac{N\delta+1}{N^{2\beta}}\right].\tag{112}$$

Proof Let $T>0$. In the following, $C>0$ is a constant independent of $\delta>0$, $0\le r<t\le T$, $N\ge1$, and $f\in\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$ which can change from one occurence to another. In what follows

$t-r\le\delta$. Recall $\eta^N=\Upsilon^N+\Theta^N$, see (85). Using (92) and the Jensen's inequality, one has:

$$\begin{aligned}
&\big|\langle f,\Upsilon^N_t\rangle-\langle f,\Upsilon^N_r\rangle\big|^2\\
&\le C\Big[(t-r)\int_r^t\int_{\mathcal{X}\times\mathcal{Y}}\big|(y-\langle\sigma_*(\cdot,x),\mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\Upsilon^N_s\rangle\big|^2\pi(\mathrm{d}s,\mathrm{d}y)\mathrm{d}s\\
&+(t-r)\int_r^t\int_{\mathcal{X}\times\mathcal{Y}}\big|\langle\sigma_*(\cdot,x),\Upsilon^N_s\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu^N_s\rangle\big|^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&+(t-r)\int_r^t\int_{\mathcal{X}\times\mathcal{Y}}\Big|\langle\sigma_*(\cdot,x),\sqrt{N}(\bar\mu^N_s-\bar\mu_s)\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu^N_s\rangle\Big|^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\
&+N|\langle f,M^N_t-M^N_r\rangle|^2+N\,\big|\langle f,V^N_t\rangle-\langle f,V^N_r\rangle\big|^2+N\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\langle f,R^N_k\rangle\Big|^2\\
&\quad+\frac{1}{N^{1+2\beta}}\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\sum_{i=1}^N\nabla f(W^i_k)\cdot\varepsilon^i_k\Big|^2\Big].
\end{aligned}\tag{113}$$

We now study each term of the right-hand side of (113). By (104), we bound the first term in (113) as follows:

$$\mathbf{E}\Big[(t-r)\int_r^t\int_{\mathcal{X}\times\mathcal{Y}}\big|(y-\langle\sigma_*(\cdot,x),\mu^N_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\Upsilon^N_s\rangle\big|^2\pi(\mathrm{d}s,\mathrm{d}y)\mathrm{d}s\Big]\le C\delta^2\|f\|^2_{\mathcal{H}^{J_1+1,j_1}}.$$

Using (105), we bound the second term of (113) as follows:

$$\mathbf{E}\Big[(t-r)\int_r^t\int_{\mathcal{X}\times\mathcal{Y}}\big|\langle\sigma_*(\cdot,x),\Upsilon^N_s\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu^N_s\rangle\big|^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\Big]\le C\delta^2\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

Using (106), we have the following bound on the third term of (113):

$$\mathbf{E}\Big[(t-r)\int_r^t\int_{\mathcal{X}\times\mathcal{Y}}\Big|\langle\sigma_*(\cdot,x),\sqrt{N}(\bar\mu^N_s-\bar\mu_s)\rangle\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu^N_s\rangle\Big|^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\Big]\le C\delta^2\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

In addition, we have, using (43), (36), and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$ (see (6)),

$$\begin{aligned}
N\mathbf{E}[|\langle f,M^N_t-M^N_r\rangle|^2]&=N\mathbf{E}\Big[\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\langle f,M^N_k\rangle\Big|^2\Big]=N\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\mathbf{E}\Big[\langle f,M^N_k\rangle^2\Big]\\
&\le CN(\lfloor Nt\rfloor-\lfloor Nr\rfloor)\|f\|^2_{\mathcal{C}^{2,\gamma_*}}/N^2\le C(N\delta+1)\|f\|^2_{\mathcal{H}^{L,\gamma}}/N.
\end{aligned}\tag{114}$$

The fifth term of (113) is bounded as follows using (108):

$$\mathbf{E}\left[N|\langle f,V^N_t\rangle-\langle f,V^N_r\rangle|^2\right]\le 2N\mathbf{E}[|\langle f,V^N_t\rangle|^2]+2N\mathbf{E}[|V^N_r[f]|^2]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}/\sqrt{N}.$$

Let us consider the sixth term in the right-hand side of (113). By (26), (51), and because $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$, we have that:

$$\begin{aligned}
\mathbf{E}\Big[N\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\langle f,R^N_k\rangle\Big|^2\Big]&\le N(\lfloor Nt\rfloor-\lfloor Nr\rfloor)\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\mathbf{E}\left[|\langle f,R^N_k\rangle|^2\right]\\
&\le CN(N\delta+1)^2\|f\|^2_{\mathcal{H}^{L,\gamma}}(1/N^4+1/N^{4\beta}).
\end{aligned}$$

Let us consider the last term in the right-hand side of (113). Using (45) and (46), we have:

$$\mathbf{E}\Big[\frac{1}{N^{1+2\beta}}\Big|\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}\nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\Big]=\frac{1}{N^{1+2\beta}}\sum_{k=\lfloor Nr\rfloor}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}\mathbf{E}\left[|\nabla f(W_k^i)\cdot\varepsilon_k^i|^2\right]$$
$$\leq\frac{1}{N^{1+2\beta}}\times C\|f\|^2_{\mathcal{H}^{L,\gamma}}N(N\delta+1)=\frac{C\|f\|^2_{\mathcal{H}^{L,\gamma}}}{N^{2\beta}}(N\delta+1).$$

Let $\{f_a\}_{a\geq1}$ be an orthonormal basis of $\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$. Gathering the previous bounds, we obtain, using also (14),

$$\mathbf{E}\big[\|\Upsilon_t^N-\Upsilon_r^N\|^2_{\mathcal{H}^{-J_2,j_2}}\big]=\sum_{a=1}^{+\infty}\mathbf{E}\left[\left|\langle f_a,\Upsilon_t^N\rangle-\langle f_a,\Upsilon_r^N\rangle\right|^2\right]$$
$$\leq C\left[\delta^2+\frac{N\delta+1}{N}+\frac{1}{\sqrt{N}}+(N\delta+1)^2\Big(\frac{1}{N^3}+\frac{1}{N^{4\beta-1}}\Big)+\frac{N\delta+1}{N^{2\beta}}\right].\tag{115}$$

By (110) and using the same arguments leading to (111), we obtain for $f\in\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$,

$$\mathbf{E}\left[|\langle f,\Theta_t^N\rangle-\langle f,\Theta_r^N\rangle|^2\right]\leq C\delta\int_r^t\mathbf{E}\left[\|\nabla f\|^2_{\mathcal{H}^{J_1,j_1}}\|\Theta_t^N\|^2_{\mathcal{H}^{-J_1,j_1}}\right]\mathrm{d}s\leq C\delta^2\|f\|^2_{\mathcal{H}^{J_1+1,j_1}}.$$

Considering an orthonormal basis of $\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$, and using the fact that $\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)\hookrightarrow_{\text{H.S.}}\mathcal{H}^{J_1+1,j_1}(\mathbf{R}^d)$, we obtain $\mathbf{E}[\|\Theta_t^N-\Theta_r^N\|^2_{\mathcal{H}^{-J_2,j_2}}]\leq C\delta^2$. Combining this result with (115), we obtain (112). This concludes the proof of the lemma. ■

Now, we collect the results of Lemmata 22 and 23 to prove the following result.

Proposition 24 *Let $\beta\geq 3/4$ and assume* **A1**-**A7**. *Then, the sequence $(\eta^N)_{N\geq1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$.*

Proof Recall $\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)\hookrightarrow_{\text{H.S.}}\mathcal{H}^{J_2,j_2}(\mathbf{R}^d)$ (by (14)). Using Markov's inequality, Lemma 22 implies item 1 in Proposition 38. In addition, according to Lemma 23, item 2 in Proposition 38 is satisfied. Consequently, $(\eta^N)_{N\geq1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$. The result follows from Proposition 38. ■

To prove Proposition 24, we mention that one could also have used (Jakubowski, 1986, Theorem 4.6) with $E=\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$ and $\mathbb{F}=\{\mathsf{L}_f, f\in\mathcal{C}_c^\infty(\mathbf{R}^d)\}$ where $\mathsf{L}_f:\Phi\in\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)\mapsto\langle f,\Phi\rangle_{J_0-1,j_0}$.

3.2 Relative compactness of $(\sqrt{N}M^N)_{N\geq1}$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$

We begin this section with the compact containment condition on the sequence $\{t\mapsto\sqrt{N}M_t^N, t\in\mathbf{R}_+\}_{N\geq1}$ (see (21) and (23)).

Lemma 25 *Let $\beta\geq 3/4$ and assume* **A1**-**A7**. *Then, for all $T>0$,*

$$\sup_{N\geq1}\mathbf{E}\Big[\sup_{t\in[0,T]}\big\|\sqrt{N}M_t^N\big\|^2_{\mathcal{H}^{-J_1,j_1}}\Big]<+\infty.$$

Proof Let $T>0$ and $f\in\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$. Then, according to (107), we have:

$$\mathbf{E}\Big[\sup_{t\in[0,T]}\langle f,\sqrt{N}M_t^N\rangle^2\Big]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

The proof of the lemma is complete considering an orthonormal basis $\{f_a\}_{a\ge1}$ of $\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)\hookrightarrow_{\text{H.S.}}\mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ (see (14)). ∎

Let us now turn to the regularity condition on the process $\{t\mapsto\sqrt{N}M_t^N, t\in\mathbf{R}_+\}_{N\ge1}$.

Lemma 26 *Let $\beta\ge3/4$ and assume* **A1**-**A7**. *Fix $T>0$. Then, there exists $C>0$ such that for all $\delta>0$ and $0\le r<t\le T$ such that $t-r\le\delta$, one has*

$$\mathbf{E}\Big[\Big\|\sqrt{N}M_t^N-\sqrt{N}M_r^N\Big\|^2_{\mathcal{H}^{-J_1,j_1}}\Big]\le C\frac{N\delta+1}{N}.$$

Proof This lemma is a direct consequence of (114) (which also holds for $f\in\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$) together with the embedding $\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)\hookrightarrow_{\text{H.S.}}\mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ (see (14)). ∎

Proposition 27 *Let $\beta\ge3/4$ and assume* **A1**-**A7**. *Then, the sequence $\{t\mapsto\sqrt{N}M_t^N, t\in\mathbf{R}_+\}_{N\ge1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$.*

Proof It is a direct consequence of Proposition 38, Lemmata 25 and 26, together with the embedding $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)\hookrightarrow_{\text{H.S.}}\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$ (see (14)). ∎

3.3 Regularity of the limit points

Lemma 28 *Let $\beta>3/4$ and assume* **A1**-**A7**. *Then, for all $T>0$,*

$$\lim_{N\to+\infty}\mathbf{E}\Big[\sup_{t\in[0,T]}\big\|\eta_t^N-\eta_{t^-}^N\big\|^2_{\mathcal{H}^{-J_0+1,j_0}}\Big]=0. \tag{116}$$

In particular, any limit point of $(\eta^N)_{N\ge1}$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$ belongs a.s. to $\mathcal{C}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$.

Proof Let $T>0$ and $f\in\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$. We have (see (85)):

$$\sup_{t\in[0,T]}\|\eta_t^N-\eta_{t^-}^N\|^2_{\mathcal{H}^{-J_0+1,j_0}}\le2\sup_{t\in[0,T]}\|\Upsilon_t^N-\Upsilon_{t^-}^N\|^2_{\mathcal{H}^{-J_0+1,j_0}}+2\sup_{t\in[0,T]}\|\Theta_t^N-\Theta_{t^-}^N\|^2_{\mathcal{H}^{-J_0+1,j_0}}.$$

According to (86) and since $\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)\hookrightarrow\mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ (see (14)), one has a.s. for all $t\in\mathbf{R}_+$ and $N\ge1$,

$$\|\Theta_t^N-\Theta_{t^-}^N\|_{\mathcal{H}^{-J_0+1,j_0}}=0.$$

Since a.s. $\bar\mu^N\in\mathcal{C}(\mathbf{R}_+,\mathcal{H}^{-L,\gamma}(\mathbf{R}^d))$, by definition of Υ^N (see (84)), it follows that a.s. for all $N\ge1$,

$$\sup_{t\in[0,T]}\langle f,\Upsilon_t^N-\Upsilon_{t^-}^N\rangle^2=N\sup_{t\in[0,T]}\langle f,\mu_t^N-\mu_{t^-}^N\rangle^2$$

From (65) and the fact that $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$, we obtain

$$\mathbf{E}\Big[\sup_{t\in[0,T]} \langle f, \Upsilon_t^N - \Upsilon_{t^-}^N\rangle^2\Big] \le C\|f\|^2_{\mathcal{H}^{L,\gamma}}\Big[\frac{1}{\sqrt{N}} + \sqrt{\frac{1}{N^5} + \frac{1}{N^{8\beta-3}}} + N^{\frac{3}{2}-2\beta}\Big].$$

Using $\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ (see (14)), we deduce that

$$\mathbf{E}\Big[\sup_{t\in[0,T]} \|\Upsilon_t^N - \Upsilon_{t^-}^N\|^2_{\mathcal{H}^{-J_0+1,j_0}}\Big] \le C\Big[\frac{1}{\sqrt{N}} + \sqrt{\frac{1}{N^5} + \frac{1}{N^{8\beta-3}}} + N^{\frac{3}{2}-2\beta}\Big].$$

Because $\beta > 3/4$, this ends the proof of (116). The second statement in Lemma 28 follows from Proposition 24, (116), and (Jacod and Shiryaev, 1987, Condition 3.28 in Proposition 3.26). The proof of Lemma 28 is complete. ■

Lemma 29 *Let $\beta > 3/4$ and assume* **A1**-**A7**. *Then, for all $T > 0$:*

$$\lim_{N\to+\infty} \mathbf{E}\Big[\sup_{t\in[0,T]} \|\sqrt{N}M_t^N - \sqrt{N}M_{t^-}^N\|^2_{\mathcal{H}^{-J_0,j_0}}\Big] = 0. \tag{117}$$

In particular, any limit point of $(\sqrt{N}M^N)_{N\ge1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ belongs a.s. to $\mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$.

Proof Let $T > 0$ and $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$. The function $t \in [0,T] \mapsto \langle f, \sqrt{N}M_t^N\rangle \in \mathbf{R}$ has $\lfloor NT\rfloor$ discontinuities, located at the times $\frac{1}{N}, \frac{2}{N}, \ldots, \frac{\lfloor NT\rfloor}{N}$. For $k \in \{1, \ldots, \lfloor NT\rfloor\}$, its k-th discontinuity is equal to $\sqrt{N}\langle f, M_{k-1}^N\rangle$. Thus,

$$\sup_{t\in[0,T]} \langle f, \sqrt{N}M_{t^-}^N - \sqrt{N}M_t^N\rangle^2 = N \max_{0\le k<\lfloor NT\rfloor} \langle f, M_k^N\rangle^2.$$

Then, using (69) and because $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$ (see indeed (6)),

$$\mathbf{E}[\sup_{t\in[0,T]} \langle f, \sqrt{N}M_{t^-}^N - \sqrt{N}M_t^N\rangle^2] \le C\|f\|^2_{\mathcal{C}^{2,\gamma_*}}/\sqrt{N} \le C\|f\|^2_{\mathcal{H}^{L,\gamma}}/\sqrt{N}. \tag{118}$$

Considering an orthonormal basis of $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ (by (14)), we obtain

$$\mathbf{E}\Big[\sup_{t\in[0,T]} \|\sqrt{N}M_t^N - \sqrt{N}M_{t^-}^N\|^2_{\mathcal{H}^{-J_0,j_0}}\Big] \le C/\sqrt{N},$$

for some $C > 0$ independent of $N \ge 1$ and f. This proves (117). The second statement in Lemma 29 is a consequence of Proposition 27, (117), and condition 3.28 of (Jacod and Shiryaev, 1987, Proposition 3.26). The proof of Lemma 29 is complete. ■

3.4 Convergence of $(\sqrt{N}M^N)_{N\geq 1}$ to a G-process

The aim of this section is to prove Proposition 32 below which states that $\{t\mapsto \sqrt{N}M_t^N, t\in \mathbf{R}_+\}_{N\geq 1}$ (see (21) and (23)) converges towards a G-process (see Definition 1). To this end, we first show the convergence of this process against test functions.

Proposition 30 *Let $\beta > 3/4$ and assume* **A1**-**A7**. *Then, for every $f\in \mathcal{C}^{2,\gamma}(\mathbf{R}^d)$ the sequence $\{t\mapsto \sqrt{N}\langle f, M_t^N\rangle, t\in \mathbf{R}_+\}_{N\geq 1}$ (see (23)) converges in distribution in $\mathcal{D}(\mathbf{R}_+,\mathbf{R})$ towards a process $X^f\in \mathcal{C}(\mathbf{R}_+,\mathbf{R})$ that has independent Gaussian increments. Moreover, for all $t\in\mathbf{R}_+$,*

$$\mathbf{E}[X_t^f]=0 \quad \textit{and} \quad \mathrm{Var}(X_t^f)=\alpha^2\mathbf{E}\left[\frac{1}{|B_\infty|}\right]\int_0^t \mathrm{Var}(\mathrm{Q}_s[f](x,y))\mathrm{d}s,$$

where we recall $\mathrm{Q}_s[f](x,y)=(y-\langle\sigma_*(\cdot,x),\bar\mu_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\bar\mu_s\rangle$ (see Definition 1).

Proof Let $f\in\mathcal{C}^{2,\gamma}(\mathbf{R}^d)$. Set for ease of notation

$$\mathfrak{m}_t^N[f]=\sqrt{N}\langle f, M_t^N\rangle.$$

To prove Proposition 30 we apply the martingale central limit theorem (Ethier and Kurtz, 2009, Theorem 7.1.4) to the sequence $\{t\mapsto \mathfrak{m}_t^N[f], t\in\mathbf{R}_+\}_{N\geq1}$. To this end, let $T>0$. Let us first show that Condition (a) in (Ethier and Kurtz, 2009, Theorem 7.1.4) holds. First of all, by (Ethier and Kurtz, 2009, Remark 7.1.5) and (23), the covariation matrix of $\mathfrak{m}_t^N[f]$ is $\mathfrak{a}_t^N[f]=N\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f, M_k^N\rangle^2$ and therefore $\mathfrak{a}_t^N[f]-\mathfrak{a}_s^N[f]\geq 0$ if $t\geq s$. On the other hand, by (118) and (6), we have:

$$\lim_{N\to\infty}\mathbf{E}\Big[\sup_{t\in[0,T]}\big|\mathfrak{m}_t^N[f]-\mathfrak{m}_{t^-}^N[f]\big|\Big]=0. \tag{119}$$

Thus Condition (a) in (Ethier and Kurtz, 2009, Theorem 7.1.4) holds. Let us now prove the last required condition in (Ethier and Kurtz, 2009, Theorem 7.1.4), namely that for all $t\in\mathbf{R}_+$, $\mathfrak{a}_t^N[f]\xrightarrow{p}\mathfrak{c}_t[f]$ where $\mathfrak{c}$ satisfies the assumptions of (Ethier and Kurtz, 2009, Theorem 7.1.1), i.e., $t\in\mathbf{R}_+\mapsto\mathfrak{c}_t[f]$ continuous, $\mathfrak{c}_0[f]=0$, and $\mathfrak{c}_t[f]-\mathfrak{c}_s[f]\geq0$ if $t\geq s$. Recall the definition of the $\boldsymbol{\sigma}$-algebra $\mathcal{F}_k^N$ in (9). For $t\in\mathbf{R}_+$,

$$\mathfrak{a}_t^N[f]=N\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}[\langle f,M_k^N\rangle^2|\mathcal{F}_k^N]+N\sum_{k=0}^{\lfloor Nt\rfloor-1}\left(\langle f,M_k^N\rangle^2-\mathbf{E}[\langle f,M_k^N\rangle^2|\mathcal{F}_k^N]\right). \tag{120}$$

We start by studying the first term in the right-hand side of (120). Recall that (see (42))

$$\mathrm{Q}^N[f](x,y,\{W_k^i\}_i)=(y-\langle\sigma_*(\cdot,x),\nu_k^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),\nu_k^N\rangle,$$

and set (see (20))

$$\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i):=\int_{\mathcal{X}\times\mathcal{Y}}\mathrm{Q}^N[f](x,y,\{W_k^i\}_i)\pi(\mathrm{d}x,\mathrm{d}y)=\frac{N}{\alpha}\langle f,D_k^N\rangle.$$

Using that $\big(|B_k|, ((x_k^n,y_k^n))_{n\geq1}\big) \perp\!\!\!\perp \mathcal{F}_k^N$, $|B_k| \perp\!\!\!\perp ((x_k^n,y_k^n))_{n\geq1}$ (see **A1**), and $(W_k^1,\dots,W_k^N)$ is $\mathcal{F}_k^N$-measurable, it holds:

$$\begin{aligned}
&\mathbf{E}\Big[\frac{1}{|B_k|^2}\sum_{1\leq n<m\leq|B_k|}\Big(\mathrm{Q}^N[f](x_k^n,y_k^n,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\Big)\\
&\quad\times\Big(\mathrm{Q}^N[f](x_k^m,y_k^m,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\Big)\Big|\mathcal{F}_k^N\Big]\\
&=\sum_{q\geq1}\frac{1}{q^2}\sum_{1\leq n<m\leq q}\mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\Big(\mathrm{Q}_k^N[f](x_k^n,y_k^n,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\Big)\\
&\quad\times\Big(\mathrm{Q}^N[f](x_k^m,y_k^m,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\Big)\Big|\mathcal{F}_k^N\Big]\\
&=\sum_{q\geq1}\frac{1}{q^2}\sum_{1\leq n<m\leq q}\mathbf{E}\Big[\mathbf{1}_{|B_k|=q}\Big(\mathrm{Q}^N[f](x_k^n,y_k^n,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\Big)\\
&\quad\times\Big(\mathrm{Q}^N[f](x_k^m,y_k^m,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\Big)\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i}\\
&=\sum_{q\geq1}\frac{1}{q^2}\sum_{1\leq n<m\leq q}\mathbf{E}[\mathbf{1}_{|B_k|=q}]\\
&\quad\times\mathbf{E}\Big[\Big(\mathrm{Q}^N[f](x_k^n,y_k^n,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\Big)\\
&\quad\times\Big(\mathrm{Q}^N[f](x_k^m,y_k^m,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\Big)\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i}\\
&=\sum_{q\geq1}\frac{1}{q^2}\sum_{1\leq n<m\leq q}\mathbf{E}[\mathbf{1}_{|B_k|=q}]\\
&\quad\times\mathbf{E}\Big[\mathrm{Q}^N[f](x_k^n,y_k^n,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i}\\
&\quad\times\mathbf{E}\Big[\mathrm{Q}^N[f](x_k^m,y_k^m,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i}\\
&=0.
\end{aligned}$$

where we have used **A3** at the two last equalities. We also have with the same arguments:

$$\begin{aligned}
&\mathbf{E}\Big[\frac{1}{|B_k|^2}\sum_{n=1}^{|B_k|}\big|\mathrm{Q}^N[f](x_k^n,y_k^n,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\big|^2\Big|\mathcal{F}_k^N\Big]\\
&=\sum_{q\geq1}\frac{1}{q^2}\sum_{n=1}^{q}\mathbf{E}[\mathbf{1}_{|B_k|=q}]\,\mathbf{E}\Big[\big|\mathrm{Q}^N[f](x_k^n,y_k^n,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\big|^2\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i}\\
&=\sum_{q\geq1}\frac{1}{q^2}\sum_{n=1}^{q}\mathbf{E}[\mathbf{1}_{|B_k|=q}]\mathbf{E}\Big[\big|\mathrm{Q}^N[f](x_1^1,y_1^1,\{w_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{w_k^i\}_i)\big|^2\Big]\Big|_{\{w_k^i\}_i=\{W_k^i\}_i}\\
&=\mathbf{E}\left[\frac{1}{|B_k|}\right]\mathrm{Var}_\pi\left(\mathrm{Q}^N[f](x,y,\{W_k^i\}_i)\right).
\end{aligned}$$

The notation Cov_π means that we consider the expectation only w.r.t. $(x,y)\sim\pi$ (see **A3**) . We then have, for $k\geq 0$ (see (21)),

$$\begin{aligned}\mathbf{E}[\langle f,M_k^N\rangle^2|\mathcal{F}_k^N] &= \frac{\alpha^2}{N^2}\mathbf{E}\Big[\Big|\frac{1}{|B_k|}\sum_{(x,y)\in B_k}\big[\mathrm{Q}^N[f](x,y,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\big]\Big|^2\Big|\mathcal{F}_k^N\Big]\\ &= \frac{\alpha^2}{N^2}\mathbf{E}\Big[\frac{1}{|B_k|^2}\sum_{n=1}^{|B_k|}\big|\mathrm{Q}^N[f](x_k^n,y_k^n,\{W_k^i\}_i)-\bar{\mathrm{Q}}^N[f](\{W_k^i\}_i)\big|^2\Big|\mathcal{F}_k^N\Big]\\ &= \frac{\alpha^2}{N^2}\mathbf{E}\left[\frac{1}{|B_k|}\right]\mathrm{Var}_\pi\left(\mathrm{Q}^N[f](x,y,\{W_k^i\}_i)\right).\end{aligned}$$

Then, one has:

$$\begin{aligned}N\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}[\langle f,M_k^N\rangle^2|\mathcal{F}_k^N] &= \alpha^2\sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\mathbf{E}\left[\frac{1}{|B_{\lfloor Ns\rfloor}|}\right]\mathrm{Var}_\pi\left(\mathrm{Q}^N[f](x,y,\{W_{\lfloor Ns\rfloor}^i\}_i)\right)\mathrm{d}s\\ &= \alpha^2\int_0^t\mathbf{E}\left[\frac{1}{|B_{\lfloor Ns\rfloor}|}\right]\mathrm{Var}_\pi\left(\mathrm{Q}^N[f](x,y,\{W_{\lfloor Ns\rfloor}^i\}_i)\right)\mathrm{d}s\\ &\quad-\alpha^2\int_{\frac{\lfloor Nt\rfloor}{N}}^t\mathbf{E}\left[\frac{1}{|B_{\lfloor Ns\rfloor}|}\right]\mathrm{Var}_\pi\left(\mathrm{Q}^N[f](x,y,\{W_{\lfloor Ns\rfloor}^i\}_i)\right)\mathrm{d}s.\end{aligned}\tag{121}$$

Using **A7**, a dominated convergence theorem, and the same arguments as those used in the proof of Lemma 17, we prove that if $m^N\to m$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))$, we have for all $f\in\mathcal{C}^{2,\gamma}(\mathbf{R}^d)$ and $t\in\mathbf{R}_+$, as $N\to+\infty$,

$$\begin{aligned}&\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\mathbf{E}\left[\frac{1}{|B_{\lfloor Ns\rfloor}|}\right]\Big[(y-\langle\sigma_*(\cdot,x),m_s^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),m_s^N\rangle\\ &\quad-\int_{\mathcal{X}\times\mathcal{Y}}(y'-\langle\sigma_*(\cdot,x'),m_s^N\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x'),m_s^N\rangle\pi(\mathrm{d}x',\mathrm{d}y')\Big]^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\\ &\to\int_0^t\int_{\mathcal{X}\times\mathcal{Y}}\mathbf{E}\left[\frac{1}{|B_\infty|}\right]\Big[(y-\langle\sigma_*(\cdot,x),m_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x),m_s\rangle\\ &\quad-\int_{\mathcal{X}\times\mathcal{Y}}(y'-\langle\sigma_*(\cdot,x'),m_s\rangle)\langle\nabla f\cdot\nabla\sigma_*(\cdot,x'),m_s\rangle\pi(\mathrm{d}x',\mathrm{d}y')\Big]^2\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s.\end{aligned}$$

Recall that by Theorem 1, $\mu^N\xrightarrow{p}\bar\mu$ in $\mathcal{D}(\mathbf{R}_+,\mathcal{P}_\gamma(\mathbf{R}^d))$. Therefore, using the continuous mapping theorem, we have for all $t\in\mathbf{R}_+$ and $f\in\mathcal{C}^{2,\gamma}(\mathbf{R}^d)$:

$$\begin{aligned}\alpha^2\int_0^t\mathbf{E}\left[\frac{1}{|B_{\lfloor Ns\rfloor}|}\right]\mathrm{Var}_\pi(\mathrm{Q}^N[f](x,y,\{W_{\lfloor Ns\rfloor}^i\}_i))\mathrm{d}s&\xrightarrow{p}\mathfrak{c}_t[f]\\ &:=\alpha^2\int_0^t\mathbf{E}\left[\frac{1}{|B_\infty|}\right]\mathrm{Var}(\mathrm{Q}_s[f](x,y))\mathrm{d}s.\end{aligned}$$

Note that $t\in\mathbf{R}_+\mapsto\mathfrak{c}_t[f]$ is locally Lipschitz continuous since for all $s\in[0,t]$,

$$\mathrm{Var}(\mathrm{Q}_s[f](x,y))\leq C\int_{\mathcal{X}\times Y}(|y|^2+1)\pi(\mathrm{d}x,\mathrm{d}y)\|f\|^2_{\mathcal{C}^{1,\gamma}}\sup_{s\in[0,t]}|\langle(1+|\cdot|^\gamma),\bar\mu_s\rangle|^2.$$

Let us now consider the second term in the right-hand side of (121). Using (37) and Lemma 8, $\mathbf{E}[|\mathrm{Q}^N[f](x,y,\{W_k^i\}_i)|^2] \le C\|f\|^2_{\mathcal{C}^{2,\gamma}}$. Consequently, it holds:

$$\mathbf{E}\Big[\Big|\alpha^2\int_{\frac{\lfloor Nt\rfloor}{N}}^t \mathbf{E}\Big[\frac{1}{|B_{\lfloor Ns\rfloor}|}\Big]\,\mathrm{Var}_\pi\left(\mathrm{Q}^N[f](x,y,\{W^i_{\lfloor Ns\rfloor}\}_i)\right)\mathrm{d}s\Big|\Big] \xrightarrow[N\to\infty]{} 0.$$

We have thus shown that

$$N\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}[\langle f, M_k^N\rangle^2|\mathcal{F}_k^N] \xrightarrow[N\to\infty]{p} \alpha^2\int_0^t \mathbf{E}\left[\frac{1}{|B_\infty|}\right]\mathrm{Var}(\mathrm{Q}_s[f](x,y))\mathrm{d}s.$$

At this point, the study of the first term in the right-hand side of (120) is complete. It remains to study the second term in the right-hand side of (120). Using (68), we obtain:

$$\begin{aligned}N^2\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f, M_k^N\rangle^2-\mathbf{E}[\langle f, M_k^N\rangle^2|\mathcal{F}_k^N]\Big|^2\Big] &= N^2\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}\Big[\Big|\langle f, M_k^N\rangle^2-\mathbf{E}[\langle f, M_k^N\rangle^2|\mathcal{F}_k^N]\Big|^2\Big]\\ &\le CN^2\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}[\langle f, M_k^N\rangle^4] \le CN^2\|f\|^4_{\mathcal{C}^{2,\gamma_*}}/N^3 \to 0.\end{aligned}$$

In conclusion, we have proved that for every $t\in\mathbf{R}_+$,

$$\mathfrak{a}_t^N[f] \xrightarrow{p} \mathfrak{c}_t[f] \text{ as } N\to+\infty. \tag{122}$$

By (Ethier and Kurtz, 2009, Theorem 7.1.4), the proof of Proposition 30 is complete.

■

Proposition 31 *Let $\beta > 3/4$ and assume* **A1**-**A7***. Consider a family $\mathscr{F} = \{f_a\}_{a\ge 1}$ of elements of $\mathcal{C}^{2,\gamma}(\mathbf{R}^d)$. Then, for $k\ge 1$, the sequence*

$$\{t\mapsto(\sqrt{N}\langle f_1, M_t^N\rangle,\dots,\sqrt{N}\langle f_k, M_t^N\rangle)^T, t\in\mathbf{R}_+\}_{N\ge1}$$

converges in distribution in $\mathcal{D}(\mathbf{R}_+,\mathbf{R}^k)$ towards a process $Y_k^{\mathscr{F}} = \{t\mapsto(Y_t^1,\dots,Y_t^k)^T, t\in\mathbf{R}_+\}\in\mathcal{C}(\mathbf{R}_+,\mathbf{R}^k)$ with zero-mean and independent Gaussian increments (which is thus a martingale). In addition, for all $0\le s\le t$,

$$\mathrm{Cov}(Y_t^i, Y_s^j) = \alpha^2\mathbf{E}\left[\frac{1}{|B_\infty|}\right]\int_0^s \mathrm{Cov}(\mathrm{Q}_v[f_i](x,y),\mathrm{Q}_v[f_j](x,y))\mathrm{d}v, \quad 1\le i,j\le k. \tag{123}$$

Notice that (123) is exactly (15).

Proof Set for ease of notation, $\mathscr{M}_t^N = (\sqrt{N}\langle f_1, M_t^N\rangle,\dots,\sqrt{N}\langle f_k, M_t^N\rangle)^T$, $t\in\mathbf{R}_+$. We have $\mathscr{M}_t^N=\sum_{q=0}^{\lfloor Nt\rfloor-1}\xi_q^N$, where $\xi_q^N = (\sqrt{N}\langle f_1, M_q^N\rangle,\dots,\sqrt{N}\langle f_k, M_q^N\rangle)^T$ (see indeed (23)). From (Ethier and Kurtz, 2009, Remark 7.1.5), the covariation matrix of $\mathscr{M}_t^N$ is

$$\mathscr{A}_t^N[f_1,\dots,f_k] := N\sum_{q=0}^{\lfloor Nt\rfloor-1}\xi_q^N(\xi_q^N)^T = N\sum_{q=0}^{\lfloor Nt\rfloor-1}(\langle f_i, M_q^N\rangle\langle f_j, M_q^N\rangle)_{i,j=1,\dots,k}.$$

If $t \geq s$, we have $\mathscr{A}_t^N[f_1,\dots,f_k] - \mathscr{A}_s^N[f_1,\dots,f_k] = N\sum_{q=\lfloor Ns\rfloor}^{\lfloor Nt\rfloor-1} \xi_q^T\xi_q \geq 0$. By (119), Condition (a) in (Ethier and Kurtz, 2009, Theorem 7.1.4) is satisfied for $\mathscr{M}^N$. Secondly, condition (1.19) in (Ethier and Kurtz, 2009, Theorem 7.1.4) is satisfied, using the decomposition

$$\begin{aligned}
&\mathscr{A}_t^N[f_1,\dots,f_k]_{i,j} = N\sum_{q=0}^{\lfloor Nt\rfloor-1}\langle f_i, M_q^N\rangle\langle f_j, M_q^N\rangle\\
&= N\sum_{q=0}^{\lfloor Nt\rfloor-1}\frac{1}{2}(\langle f_i+f_j, M_q^N\rangle^2 - \langle f_i, M_q^N\rangle^2 - \langle f_j, M_q^N\rangle^2)\\
&= \frac{1}{2}\big(\mathfrak{a}_t^N[f_i+f_j] - \mathfrak{a}_t^N[f_i] - \mathfrak{a}_t^N[f_j]\big)\\
&\xrightarrow{p} \frac{1}{2}\big(\mathfrak{c}_t[f_i+f_j] - \mathfrak{c}_t[f_i] - \mathfrak{c}_t[f_j]\big) \text{ (by (122))}\\
&= \alpha^2\int_0^t \mathbf{E}\left[\frac{1}{|B_\infty|}\right]\mathrm{Cov}(\mathrm{Q}_v[f_i](x,y), \mathrm{Q}_v[f_j](x,y))\mathrm{d}v := (\mathfrak{C}_t)_{i,j}.
\end{aligned}$$

It remains to check that $\mathfrak{C}$ satisfies the assumptions of (Ethier and Kurtz, 2009, Theorem 7.1.1). Clearly $\mathfrak{C}(0) = 0$ and $t \in \mathbf{R}_+ \mapsto \mathfrak{C}_t$ is continuous. In addition, if $0 \leq s \leq t$,

$$\begin{aligned}
\mathfrak{C}_t - \mathfrak{C}_s &= \alpha^2\Big(\int_s^t \mathbf{E}\left[\frac{1}{|B_\infty|}\right]\mathrm{Cov}(\mathrm{Q}_v[f_i](x,y), \mathrm{Q}_v[f_j](x,y))\mathrm{d}v\Big)_{i,j=1,\dots,k}\\
&= \alpha^2\int_s^t \mathbf{E}\left[\frac{1}{|B_\infty|}\right]\mathbf{E}[\Xi_v(x,y)\Xi_v^T(x,y)]\mathrm{d}v,
\end{aligned}$$

where $\Xi_v(x,y)_i = \mathrm{Q}_v[f_i](x,y) - \mathbf{E}[\mathrm{Q}_v[f_i](x,y)]$, $i \in \{1,\dots,k\}$. Thus, $\mathfrak{C}_t - \mathfrak{C}_s$ is symmetric and non negative definite. The proof of Proposition 31 is complete. ∎

Proposition 32 *Let $\beta > 3/4$ and assume that* **A1**-**A7** *hold. Then, the sequence $(\sqrt{N}M^N)_{N\geq 1}$ converges in distribution in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ to a* G*-process $\mathscr{G} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ (see Definition 1).*

To prove Proposition 32, we will prove that there is a unique limit point of the sequence $(\sqrt{N}M^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ (recall that this sequence is relatively compact in this space, see Proposition 27), so that the whole sequence converges in distribution. Proposition 31 will then imply that this unique limit point is a G-process. Before, we need to introduce some definitions. For a family $\mathscr{F} = \{f_a\}_{a\geq 1}$ of elements of $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$, we define, for $k \geq 1$, the projection

$$\pi_k^{\mathscr{F}} : \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d)) \to \mathcal{D}(\mathbf{R}_+, \mathbf{R})^k, \quad m \mapsto (\langle f_1, m\rangle, \dots, \langle f_k, m\rangle)^T.$$

The function $\pi_k^{\mathscr{F}}$ is continuous. In the following, $\mathscr{H} = \{h_a\}_{a\geq 1}$ is an orthonormal basis of $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$. Let d_R be a metric for the Skorohod topology on $\mathcal{D}(\mathbf{R}_+, \mathbf{R})$. Introduce the space $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$ defined as the set of sequences taking values in $\mathcal{D}(\mathbf{R}_+, \mathbf{R})$. We endow $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$ with the metric $\rho(u,v) = \sum_{a\geq 1} 2^{-a}\min(1, d_R(u_a, v_a))$. We consider on

$\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$ the topology associated with ρ. We have that $\rho(u^N, u) \to 0$ if and only if $d_R(u_a^N, u_a) \to 0$ for all $a \geq 1$. Notice with that with this metric ρ, $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$ is separable, since $\mathcal{D}(\mathbf{R}_+, \mathbf{R})$ is separable (Ethier and Kurtz, 2009, Theorem 3.5.6). We now define the map

$$\Pi : \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d)) \to \mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty, \quad m \mapsto (\langle h_a, m\rangle)_{a\geq 1}^T.$$

This map is injective (because $\mathscr{H}$ is a basis of $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$) and continuous. The map Π depends on the orthonormal basis $\mathscr{H}$ but, for ease of notation, we have omitted to write it. Finally, we introduce the continuous function

$$p_k : \mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty \to \mathcal{D}(\mathbf{R}_+, \mathbf{R})^k, \quad (m_a)_{a\geq 1}^T \mapsto (m_1, \ldots, m_k)^T.$$

It holds

$$\pi_k^{\mathscr{H}} = p_k \circ \Pi.$$

We now introduce the set

$$\mathcal{C} := \left\{p_k^{-1}(H), \quad H \in \mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^k),\ k \geq 1\right\} \subset \mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty,$$

where $\mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^k)$ denotes the Borel $\boldsymbol{\sigma}$-algebra of $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^k$. The continuity of p_k implies that $\mathcal{C} \subset \mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$. The following result shows that $\mathcal{C}$ is a separating class of $(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty, \mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty))$, where we recall that this means by definition that two probability measures on $\mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$ which agree on $\mathcal{C}$ necessarily agree on $\mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$.

Lemma 33 *The set $\mathcal{C}$ is a separating class of $(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty, \mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty))$.*

Proof We recall that any subset of $\mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$ which is a π-system (i.e. closed under finite intersection) and which generates the $\boldsymbol{\sigma}$-algebra $\mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$ is a separating class (see (Billingsley, 1999, Page 9)). Let us first prove that $\mathcal{C}$ is a π-system. Notice that it holds $p_k^{-1}(H) = p_{k+1}^{-1}(H \times \mathcal{D}(\mathbf{R}_+, \mathbf{R}))$. Thus, if A and $A' \in \mathcal{C}$ (write $A = p_k^{-1}(H)$ and $A' = p_{k'}^{-1}(H')$, and assume that $k' \geq k$), then $A \cap A' = p_{k'}^{-1}((H \times \mathcal{D}(\mathbf{R}_+, \mathbf{R}) \ldots \times \mathcal{D}(\mathbf{R}_+, \mathbf{R})) \cap H') \in \mathcal{C}$. Consequently $\mathcal{C}$ is a π-system. It remains to show that the $\boldsymbol{\sigma}$-algebra generated by $\mathcal{C}$ is equal to $\mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$. To prove it, it is sufficient to prove that any open set of $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$ is a countable union of sets in $\mathcal{C}$. Introduce, for $x \in \mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$, $k \geq 1$ and $\epsilon > 0$,

$$\mathcal{N}_{k,\epsilon}(x) := \{y \in \mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty;\ d_R(x_a, y_a) < \epsilon,\ 1 \leq a \leq k\} \subset \mathcal{C}. \tag{124}$$

By straightforward arguments $\mathcal{N}_{k,\epsilon}(x)$ is open in $(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty, \rho)$. Remark that for all $y \in \mathcal{N}_{k,\epsilon}(x)$, it holds $\rho(x, y) < \epsilon + 2^{-k}$. Given $r > 0$, choose $\epsilon > 0$ and $k \geq 1$ such that $\epsilon + 2^{-k} < r$. Then, $\mathcal{N}_{k,\epsilon}(x) \subset B_\rho(x, r)$ where $B_\rho(x, r)$ is the open ball of center x and radius r for the metric ρ of $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$. The space $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty$ is separable and we consider D a dense and countable subset of $(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty, \rho)$. Let $\mathcal{O}$ be an open subset of $(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty, \rho)$. We claim that

$$\mathcal{O} = \mathcal{N}_\mathcal{O} \text{ where } \mathcal{N}_\mathcal{O} := \bigcup_{\substack{x \in \mathrm{D} \cap \mathcal{O},\ k \geq 1 \\ \epsilon \in \mathbf{Q} \cap \mathbf{R}_+^*,\ \mathcal{N}_{k,\epsilon}(x) \subset \mathcal{O}}} \mathcal{N}_{k,\epsilon}(x).$$

We have $\mathcal{N}_\mathcal{O} \subset \mathcal{O}$. Let us now show that $\mathcal{O} \subset \mathcal{N}_\mathcal{O}$. To this end, pick $y \in \mathcal{O}$ and $r_0 > 0$ such that $B_\rho(y, r_0) \subset \mathcal{O}$. Choose $k_0 \geq 1$ such that

$$2^{-k_0} < \frac{r_0}{4}.$$

Consider, for $n \geq 1$, $x^n \in \mathrm{D}$ such that $\rho(y, x^n) < 1/n$. Choose $n \geq 1$ such that

$$\frac{1}{n} + \frac{r_0}{2} < r_0 \text{ and } \frac{2^{k_0}}{n} < \frac{r_0}{4}.$$

Since for all $a \geq 1$, $d_R(y_a, x_a^n) \to 0$ as $n \to +\infty$ (by definition of ρ), we choose if necessary $n \geq 1$ larger so that $\min(1, d_R(y_a, x_a^n)) = d_R(y_a, x_a^n)$ for all $a = 1, \ldots, k_0$. Finally, choose $\epsilon_{k_0,n} = 2^{k_0}/n \in \mathbf{Q}$. We have for all $a = 1, \ldots, k_0$, $d_R(x_n^a, y_a) \leq \rho(x^n, y)2^a \leq \rho(x^n, y)2^{k_0} < 2^{k_0}/n = \epsilon_{k_0,n}$, so that $y \in \mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n)$. It just remains to check that $\mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n) \subset \mathcal{O}$ to ensure that $\mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n) \subset \mathcal{N}_{\mathcal{O}}$. We have $\rho(y, x^n) < \epsilon_{k_0,n} + 2^{-k_0} < r_0/4 + r_0/4 = r_0/2$ and thus $\mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n) \subset B_\rho(x^n, r_0/2)$. Since $1/n + r_0/2 < r_0$, by triangular inequality, $B_\rho(x^n, r_0/2) \subset B_\rho(y, r_0) \subset \mathcal{O}$. Thus, $\mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n) \subset \mathcal{O}$, which proves that $\mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n) \subset \mathcal{N}_{\mathcal{O}}$. Consequently, we have proved that $y \in \mathcal{N}_{k_0,\epsilon_{k_0,n}}(x^n) \subset \mathcal{N}_{\mathcal{O}}$, and then that $\mathcal{O} \subset \mathcal{N}_{\mathcal{O}}$. Thus, $\mathcal{O} = \mathcal{N}_{\mathcal{O}}$. In conclusion, every open set $\mathcal{O}$ is a countable union of sets of the form (124). This implies that $\sigma(\mathcal{C}) = \mathcal{B}(\mathcal{D}(\mathbf{R}_+, \mathbf{R})^\infty)$ and therefore $\mathcal{C}$ is a separating class. ■

Proposition 34 *Let P, Q be two probability measures on $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$ such that $\pi_k^{\mathscr{H}} P = \pi_k^{\mathscr{H}} Q$ for all $k \geq 1$. Then, $P = Q$.*

Proof The equality $\pi_k^{\mathscr{H}} P = \pi_k^{\mathscr{H}} Q$ for all $k \geq 1$, writes $p_k \circ \Pi P = p_k \circ \Pi Q$. By Lemma 33, $\Pi P = \Pi Q$. Since Π is injective it admits a left inverse Π^{-1}, and therefore $P = Q$. The proof is complete. ■

We are now in position to prove Proposition 32.

Proof [Proof of Proposition 32] By Proposition 27, $(\sqrt{N} M^N)_{N \geq 1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$. Let $\mathscr{M}^*$ be one of its limit point. Let us show that $\mathscr{M}^*$ is independent of the extracted subsequence, say of N'. Since $\pi_k^{\mathscr{H}}$ is continuous for all $k \geq 1$, the continuous mapping theorem implies that in $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^k$,

$$\pi_k^{\mathscr{H}}(\sqrt{N'} M^{N'}) \to \pi_k^{\mathscr{H}}(\mathscr{M}^*) \text{ in distribution, as } N' \to +\infty.$$

For $k \geq 1$, introduce the following continuous and bijective mapping $\mathscr{Q}_k : \mathcal{D}(\mathbf{R}_+, \mathbf{R}^k) \to \mathcal{D}(\mathbf{R}_+, \mathbf{R})^k$ defined by: $m \mapsto (m \cdot e_1, \ldots, m \cdot e_k)^T$, where $\{e_1, \ldots, e_k\}$ denotes the canonical basis of $\mathbf{R}^k$. By Proposition 31 applied with $f_a = h_a$, $a \in \{1, \ldots, k\}$ (recall $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \subset \mathcal{C}^{2,\gamma}(\mathbf{R}^d)$), it holds in $\mathcal{D}(\mathbf{R}_+, \mathbf{R}^k)$,

$$\forall k \geq 1, \ \mathscr{Q}_k^{-1} \circ \pi_k^{\mathscr{H}}(\sqrt{N'} M^{N'}) \to Y_k^{\mathscr{H}} \text{ in distribution, as } N' \to +\infty.$$

Since $\mathscr{Q}_k$ is continuous, one then has in $\mathcal{D}(\mathbf{R}_+, \mathbf{R})^k$,

$$\forall k \geq 1, \ \pi_k^{\mathscr{H}}(\sqrt{N'} M^{N'}) \to \mathscr{Q}_k(Y_k^{\mathscr{H}}) \text{ in distribution, as } N' \to +\infty.$$

It follows that $\pi_k^{\mathscr{H}}(\mathscr{M}^*) = \mathscr{Q}_k(Y_k^{\mathscr{H}})$ in distribution. By Proposition 34, the distribution of $\mathscr{M}^*$ is fully determined by the collection of distributions of the processes $\pi_k^{\mathscr{H}}(\mathscr{M}^*)$, for $k \geq 1$. Thus, $\mathscr{M}^*$ is independent of the subsequence, and therefore the whole sequence $(\sqrt{N} M^N)_{N \geq 1}$ convergences to $\mathscr{M}^*$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$. By Lemma 29, $\mathscr{M}^* \in$

$\mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$. Let us now consider a family $\mathscr{F} = \{f_a\}_{a\geq 1}$ of elements of $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$. Since $\mathscr{D}_k$ and $\pi_k^{\mathscr{F}}$ are continuous, and by Proposition 31, one has that $\mathscr{Q}_k^{-1} \circ \pi_k^{\mathscr{F}}(\mathscr{M}^*) = Y_k^{\mathscr{F}} \in \mathcal{C}(\mathbf{R}_+, \mathbf{R}^k)$ in distribution. The proof of Proposition 32 is complete. ■

3.5 Limit points of $(\eta^N)_{N\geq 1}$ and end of the proof of Theorem 2

3.5.1 On the limit points of the sequence $(\eta^N, \sqrt{N}M^N)_{N\geq 1}$

Lemma 35 *Assume* **A1**-**A7**. *Then, the sequence $(\eta_0^N)_{N\geq 1}$ converges in distribution in $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$ towards a variable ν_0 which is the unique (in distribution) $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$-valued random variable such that for all $k\geq 1$ and $f_1\dots,f_k \in \mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$, $(\langle f_1,\nu_0\rangle,\dots,\langle f_k,\nu_0\rangle)^T \sim \mathcal{N}(0,\Gamma(f_1,\dots,f_k))$, where $\Gamma(f_1,\dots,f_k)$ is the covariance matrix of the vector $(f_1(W_0^1),\dots,f_k(W_0^1))^T$.*

Proof The sequence $(\eta_0^N)_{N\geq 1}$ is tight in $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$. Let $\mathscr{F} = \{f_a\}_{a\geq 1}$ be a family of elements of $\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$. Define, for $k\geq 1$, the projection

$$\mathscr{P}_k^{\mathscr{F}} : \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d) \to \mathbf{R}^k, \quad m \mapsto (\langle f_1, m\rangle, \dots, \langle f_k, m\rangle).$$

The map $\mathscr{P}_k^{\mathscr{F}}$ is continuous. By the standard vectorial central limit theorem, for $k\geq 1$, $\mathscr{P}_k^{\mathscr{F}}(\eta_0^N) \to \mathcal{N}(0,\Gamma(f_1,\dots,f_k))$ in distribution. In addition, we show with the same arguments as those used to prove Lemma 33 and Proposition 34, that when $\mathscr{F}$ is an orthonormal basis of $\mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$, the distribution of a $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$-valued random variable ν is fully determined by the collection of the distributions $\{\mathscr{P}_k^{\mathscr{F}}(\nu), k\geq 1\}$. Hence, $(\eta_0^N)_{N\geq 1}$ has a unique limit point ν_0 in distribution which is the unique (in distribution) $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$-valued random variable such that for all $k\geq 1$, $\mathscr{P}_k^{\mathscr{F}}(\nu_0) \sim \mathcal{N}(0,\Gamma(f_1,\dots,f_k))$. In particular, the whole sequence $(\eta_0^N)_{N\geq 1}$ converges in distribution towards ν_0. The proof of the lemma is complete. ■

Set

$$\mathscr{E} := \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)) \times \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d)). \tag{125}$$

According to Propositions 24 and 27, $(\eta^N, \sqrt{N}\, M^N)_{N\geq 1}$ is tight in $\mathscr{E}$. Let $(\eta^*, \mathscr{G}^*)$ be one of its limit point in $\mathscr{E}$. Along some subsequence, it holds:

$$(\eta^{N'}, \sqrt{N'}\, M^{N'}) \to (\eta^*, \mathscr{G}^*), \text{ as } N' \to +\infty.$$

Considering the marginal distributions, and according to Lemmata 28 and 29, it holds a.s.

$$\eta^* \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)) \text{ and } \mathscr{G}^* \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d)). \tag{126}$$

By uniqueness of the limit in distribution, using Lemma 35 (together with the fact that the projection $m \in \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)) \mapsto m_0 \in \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$ is continuous) and Proposition 32, it also holds:

$$\eta_0^* = \nu_0 \text{ and } \mathscr{G}^* = \mathscr{G}, \text{ in distribution.} \tag{127}$$

Proposition 36 *Let $\beta > 3/4$ and assume* **A1-A7***. Then, η^* is a weak solution of* (16) *(see Definition 2) with initial distribution ν_0 (see Lemma 35).*

Proof Let us introduce, for $\Phi \in \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$, $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$, and $s \geq 0$:

$$\mathbf{U}_s[f](\Phi) := \int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle \sigma_*(\cdot, x), \bar{\mu}_s\rangle)\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \Phi\rangle \pi(\mathrm{d}x, \mathrm{d}y) \tag{128}$$

and

$$\mathbf{V}_s[f](\Phi) := \int_{\mathcal{X}\times\mathcal{Y}} \langle \sigma_*(\cdot, x), \Phi\rangle\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \bar{\mu}_s\rangle \pi(\mathrm{d}x, \mathrm{d}y). \tag{129}$$

Recall $\eta_t^N = \sqrt{N}(\mu_t^N - \bar{\mu}_t)$ (see (11)). Using (25) and Corollary 1, it holds:

$$\langle f, \eta_t^N\rangle - \langle f, \eta_0^N\rangle - \int_0^t (\mathbf{U}_s[f](\eta_s^N) - \mathbf{V}_s[f](\eta_s^N))\mathrm{d}s - \langle f, \sqrt{N}M_t^N\rangle = -\mathbf{e}_t^N[f], \tag{130}$$

where

$$\begin{aligned}\mathbf{e}_t^N[f] := &\frac{1}{\sqrt{N}} \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha\langle \sigma_*(\cdot, x), \eta_s^N\rangle\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \eta_s^N\rangle \pi(\mathrm{d}x, \mathrm{d}y)\mathrm{d}s\\ &- \sqrt{N}\langle f, V_t^N\rangle - \sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor-1} \langle f, R_k^N\rangle - \frac{\sqrt{N}}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N} \nabla f(W_k^i)\cdot \varepsilon_k^i.\end{aligned}$$

In what follows, $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in \mathbf{R}_+$ are fixed.

Step 1. In this step we study the continuity of the mapping

$$\mathbf{B}_t[f] : m \in \mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)) \mapsto \langle f, m_t\rangle - \int_0^t (\mathbf{U}_s[f](m_s) - \mathbf{V}_s[f](m_s))\mathrm{d}s \in \mathbf{R}.$$

Let $(m^N)_{N\geq 1}$ such that $m^N \to m$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$. Recall that $\sup_{x\in\mathcal{X}} \|\sigma_*(\cdot, x)\|_{\mathcal{H}^{J_0-1,j_0}} < +\infty$ (by **A2** and because $j_0 > d/2$). Then, for all $N \geq 1$ and $s \in [0, t]$, it holds:

$$\begin{aligned}&\big|(y - \langle \sigma_*(\cdot, x), \bar{\mu}_s\rangle)\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), m_s^N\rangle\big|\\ &\leq C(|y| + 1)\|f\|_{\mathcal{H}^{J_0,j_0}} \sup_{N\geq 1}\sup_{s\in[0,t]} \|m_s^N\|_{\mathcal{H}^{-J_0+1,j_0}} < +\infty\end{aligned}$$

and since $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{1,\gamma}(\mathbf{R}^d)$,

$$\begin{aligned}\big|\langle \sigma_*(\cdot, x), m_s^N\rangle\langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \bar{\mu}_s\rangle\big| &\leq C \sup_{N\geq 1}\sup_{s\in[0,t]} \|m_s^N\|_{\mathcal{H}^{-J_0+1,j_0}}\\ &\quad\times \|f\|_{\mathcal{H}^{J_0,j_0}} \sup\nolimits_{s\in[0,t]} |\langle (1 + |\cdot|^\gamma), \bar{\mu}_s\rangle| < +\infty,\end{aligned}$$

for some $C > 0$ independent of $N \geq 1$, $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$, $s \geq 0$, and $(x, y) \in \mathcal{X}\times\mathcal{Y}$. With these two upper bounds, and using the same arguments as those used in the proof of Lemma 17, one deduces that for all continuity points $t \in \mathbf{R}_+$ of $\{t \mapsto m_t, t \in \mathbf{R}_+\}$, we have $\mathbf{B}_t[f](m^N) \to \mathbf{B}_t[f](m)$ as $N \to +\infty$. Consequently, using (126) and the continuous mapping theorem (Billingsley, 1999, Theorem 2.7), for all $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in \mathbf{R}_+$, it holds in distribution and as $N' \to +\infty$:

$$\mathbf{B}_t[f](\eta^{N'}) - \langle f, \eta_0^{N'} \rangle - \langle f, \sqrt{N'} M_t^{N'} \rangle \to \mathbf{B}_t[f](\eta^*) - \langle f, \eta_0^* \rangle - \langle f, \mathscr{G}_t^* \rangle. \tag{131}$$

Step 2. In this step we prove that for any $t \in \mathbf{R}_+$ and $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$:

$$\mathbf{E}\big[\big|\mathbf{e}_t^N[f]\big|\big] \to 0 \text{ as } N \to +\infty. \tag{132}$$

By (102), we have since $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \hookrightarrow \mathcal{H}^{J_2+1,j_2}(\mathbf{R}^d)$ (because $J_0 \geq J_2 + 1$ and $j_2 \geq j_0$, see (12) and (13)),

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\mathbf{E}\Big[\int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha|\langle \sigma_*(\cdot,x), \eta_s^N\rangle||\langle \nabla f \cdot \nabla\sigma_*(\cdot,x), \eta_s^N\rangle| \pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s\Big]\\
&\leq \frac{Ct\|\nabla f\|_{\mathcal{H}^{J_2,j_2}}}{\sqrt{N}}\mathbf{E}\Big[\sup_{t\in[0,T]} \|\eta_t^N\|^2_{\mathcal{H}^{-J_2,j_2}}\Big]\\
&\leq \frac{Ct\|f\|_{\mathcal{H}^{J_2+1,j_2}}}{\sqrt{N}} \to 0 \text{ as } N \to +\infty.
\end{aligned}$$

Using Lemma 9, we also have since $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$,

$$\begin{aligned}
&\mathbf{E}\Big[\Big|\sqrt{N}\langle f, V_t^N\rangle - \sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor -1} \langle f, R_k^N\rangle - \frac{\sqrt{N}}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt\rfloor -1}\sum_{i=1}^N \nabla f(W_k^i)\cdot \varepsilon_k^i\Big|\Big]\\
&\leq C\|f\|_{\mathcal{C}^{2,\gamma_*}}\Big[\frac{\sqrt{N}}{N} + \sqrt{N}N\big[\frac{1}{N^2} + \frac{1}{N^{2\beta}}\big] + \frac{\sqrt{N}}{N^\beta}\Big].
\end{aligned}$$

The right-hand-side of the previous term goes to 0 as $N \to +\infty$, since $\beta > 3/4$. This proves (132).

Step 3. End of the proof of Proposition 36. By (131), (132), and (130), we deduce that for all $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in \mathbf{R}^+$, it holds a.s. $\mathbf{B}_t[f](\eta^*) - \langle f, \eta_0^* \rangle - \langle f, \mathscr{G}_t^* \rangle = 0$. Since $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $\mathbf{R}_+$ are separable, we conclude by a standard continuity argument that a.s. for all $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in \mathbf{R}^+$, $\mathbf{B}_t[f](\eta^*) - \langle f, \eta_0^* \rangle - \langle f, \mathscr{G}_t^* \rangle = 0$. Hence, η^* is a weak solution of (16) (see Definition 2) with initial distribution ν_0 (see (127)). This ends the proof of Proposition 36. ■

Inspired by the proof of (Delarue et al., 2019, Corollary 5.7) (see also (Kurtz and Xiong, 2004)), to end the proof of Theorem 2, we will show that (16) has a unique strong solution. This is the purpose of the next section, where we also conclude the proof of Theorem 2.

3.5.2 Pathwise uniqueness

Proposition 37 *Let $\beta > 3/4$ and assume* **A1**-**A7**. *Then strong uniqueness holds for* (16), *namely, on a fixed probability space, given a $\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d)$-valued random variable ν and a* G*-process $\mathscr{G} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R}^d))$, there exists at most one $\mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$-valued process η solution to* (16) *with $\eta_0 = \nu$ almost surely.*

Proof By linearity of the involved operators in (16), it is enough to consider a $\mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$-valued process η solution to (16) when a.s. $\nu = 0$ and $\mathscr{G} = 0$, i.e., for every $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in \mathbf{R}_+$:

$$\begin{cases} \langle f, \eta_t \rangle - \int_0^t (\mathbf{U}_s[f](\eta_s) - \mathbf{V}_s[f](\eta_s))\mathrm{d}s = 0, \\ \langle f, \eta_0 \rangle = 0, \end{cases} \tag{133}$$

where $\mathbf{U}$ and $\mathbf{V}$ are defined respectively in (128) and (129). Pick $T > 0$. By (133), a.s. for all $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in [0,T]$, we have $\langle f, \eta_t\rangle^2 = 2\int_0^t (\mathbf{U}_s[f](\eta_s) - \mathbf{V}_s[f](\eta_s))\langle f, \eta_s\rangle \mathrm{d}s$. Recall $\sup_{s\in[0,T]}|\langle(1+|\cdot|^\gamma), \bar\mu_s\rangle| < +\infty$ (because $\bar\mu \in \mathcal{D}(\mathbf{R}_+, \mathcal{P}_\gamma(\mathbf{R}^d))$) and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{C}^{1,\gamma}(\mathbf{R}^d)$. Then, by Cauchy-Schwarz inequality, a.s. for all $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$ and $t \in [0,T]$:

$$\begin{aligned} &-\int_0^t \mathbf{V}_s[f](\eta_s)\langle f, \eta_s\rangle \mathrm{d}s \\ &\le \alpha \int_0^t \Big[\langle f, \eta_s\rangle^2 + \int_{\mathcal{X}\times\mathcal{Y}} \langle \sigma_*(\cdot,x), \eta_s\rangle^2 \langle \nabla f\cdot\nabla\sigma_*(\cdot,x), \bar\mu_s\rangle^2 \pi(\mathrm{d}x,\mathrm{d}y)\Big]\mathrm{d}s \\ &\le C\int_0^t \Big[\langle f, \eta_s\rangle^2 + \|\eta_s\|^2_{\mathcal{H}^{-J_0,j_0}}\|f\|^2_{\mathcal{H}^{L,\gamma}}\Big]\mathrm{d}s. \end{aligned} \tag{134}$$

Let $\{f_a\}_{a\ge1}$ be an orthonormal basis of $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$. Using the operator $\mathsf{T}_x : f \in \mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \mapsto \nabla f\cdot\nabla\sigma(\cdot,x) \in \mathcal{H}^{J_0-1,j_0}(\mathbf{R}^d)$ defined for all $x \in \mathcal{X}$ and Lemma 40, we have a.s. for all $t \in [0,T]$:

$$\begin{aligned} &\sum_{a\ge1}\int_0^t \mathbf{U}_s[f_a](\eta_s)\langle f_a, \eta_s\rangle \mathrm{d}s \\ &= \int_0^t\int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle\sigma_*(\cdot,x),\bar\mu_s\rangle)\Big(\sum_{a\ge1}\langle f_a,\eta_s\rangle\langle \mathsf{T}_x f_a, \eta_s\rangle\Big)\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s \\ &= \int_0^t\int_{\mathcal{X}\times\mathcal{Y}} \alpha(y - \langle\sigma_*(\cdot,x),\bar\mu_s\rangle)\langle \eta_s, \mathsf{T}_x^*\eta_s\rangle_{\mathcal{H}^{-J_0,j_0}}\pi(\mathrm{d}x,\mathrm{d}y)\mathrm{d}s \\ &\le C\int_0^t \|\eta_s\|^2_{\mathcal{H}^{-J_0,j_0}}\mathrm{d}s. \end{aligned} \tag{135}$$

Therefore, using the bounds (134) and (135), together with $\mathcal{H}^{J_0,j_0}(\mathbf{R}^d) \hookrightarrow_{\text{H.S.}} \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$, we have a.s. for all $t \in [0,T]$:

$$\|\eta_t\|^2_{\mathcal{H}^{-J_0,j_0}} = \sum_{a\ge1}\langle f_a, \eta_t\rangle^2 \le C\int_0^t \|\eta_s\|^2_{\mathcal{H}^{-J_0,j_0}}\mathrm{d}s.$$

By Gronwall's lemma, a.s. $\|\eta_t\|_{\mathcal{H}^{-J_0,j_0}} = 0$ for all $t \in [0,T]$. This concludes the proof of Proposition 37. ■

3.5.3 End of the proof of Theorem 2

Proof [Proof of Theorem 2]

Let $\ell \in \{1,2\}$ and N_ℓ be such that in distribution $\eta^{N_\ell} \to \eta^\ell$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$ (see Proposition 24). By Lemma 28, a.s. $\eta^\ell \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$. Consider now $(\eta^{\ell,*}, \mathscr{G}^{\ell,*})$ a limit point of $(\eta^{N_\ell}, \sqrt{N_\ell}\, M^{N_\ell})_{N_\ell \geq 1}$ in $\mathscr{E}$. Up to extracting a subsequence from N_ℓ, we assume that in distribution and as $N_\ell \to +\infty$,

$$(\eta^{N_\ell}, \sqrt{N_\ell}\, M^{N_\ell}) \to (\eta^{\ell,*}, \mathscr{G}^{\ell,*}) \text{ in } \mathscr{E}.$$

Considering the marginal distributions, we then have by uniqueness of the limit in distribution, for $\ell = 1,2$ (see also Proposition 32):

$$\eta^{\ell,*} = \eta^\ell, \text{ and } \mathscr{G}^{\ell,*} = \mathscr{G} \text{ in distribution.} \tag{136}$$

By Proposition 36, $\eta^{1,*}$ and $\eta^{2,*}$ are two weak solutions of (16) with initial distribution ν_0 (see also Lemma 35). Since strong uniqueness for (16) (see Proposition 37) implies weak uniqueness for (16), we deduce that $\eta^{1,*} = \eta^{2,*}$ in distribution. By (136), this implies that $\eta^1 = \eta^2$ in distribution and then, that the whole sequence $(\eta^N)_{N\geq 1}$ converges in distribution in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$. Denoting by η^* its limit, we have proved that η^* has the same distribution as the unique weak solution $\eta^\star$ of (16) with initial distribution ν_0. This concludes the proof of Theorem 2.

■

3.6 The case when $\beta = 3/4$

In this section, we assume that $d = 1$. Recall $\mathsf{f}_2 : x \in \mathbf{R} \mapsto |x|^2$, which belongs to $\mathcal{H}^{J_0,j_0}(\mathbf{R})$ because $j_0 - 2 = 3 - 2 > 1/2$ (see (12)).

Proof [Proof of Proposition 7] Assume $\beta = 3/4$. The proof of Proposition 7 is divided into two steps.

Step 1. Let $f \in \mathcal{H}^{J_0,j_0}(\mathbf{R})$. When $\beta = 3/4$, it appears that for non affine test functions f, the term $\langle f, R_k^N\rangle$ is not negligible any more. In this step we simply rewrite (130) by decomposing the term $\langle f, R_k^N\rangle$ into two terms: $\langle f, R_k^N\rangle = \langle f, \mathscr{R}_k^N\rangle + \langle f, \mathscr{B}_k^N\rangle$ where $\langle f, \mathscr{R}_k^N\rangle$ will be negligible and $\langle f, \mathscr{B}_k^N\rangle$ will not be negligible. More precisely, by (19) and (2), it holds

$$\begin{aligned}\langle f, R_k^N\rangle &= \frac{1}{2N}\sum_{i=1}^N (W_{k+1}^i - W_k^i)^2 f(\widehat{W}_k^i)\\ &= \frac{1}{2N}\sum_{i=1}^N \Big| \frac{\alpha}{N|B_k|} \sum_{(x,y)\in B_k} (y - g_{W_k}^N(x))\nabla_W \sigma_*(W_k^i, x) + \frac{\varepsilon_k^i}{N^{3/4}}\Big|^2 f''(\widehat{W}_k^i)\\ &= \langle f, \mathscr{R}_k^N\rangle + \langle f, \mathscr{B}_k^N\rangle,\end{aligned}$$

where

$$\langle f, \mathscr{R}_k^N \rangle = \frac{1}{2N} \sum_{i=1}^N \Big| \frac{\alpha}{N|B_k|} \sum_{(x,y)\in B_k} (y - g_{W_k}^N(x)) \nabla_W \sigma_*(W_k^i, x) \Big|^2 f''(\widehat{W}_k^i)$$
$$+ \frac{1}{N} \sum_{i=1}^N \frac{\alpha}{N|B_k|} \sum_{(x,y)\in B_k} (y - g_{W_k}^N(x)) \nabla_W \sigma_*(W_k^i, x) \frac{\varepsilon_k^i}{N^{3/4}} f''(\widehat{W}_k^i)$$

and

$$\langle f, \mathscr{B}_k^N \rangle = \frac{1}{2N^{5/2}} \sum_{i=1}^N |\varepsilon_k^i|^2 f''(\widehat{W}_k^i).$$

From (130), one then has:

$$\langle f, \eta_t^N \rangle - \langle f, \eta_0^N \rangle - \int_0^t (\mathbf{U}_s[f](\eta_s^N) - \mathbf{V}_s[f](\eta_s^N)) \mathrm{d}s - \langle f, \sqrt{N} M_t^N \rangle = -\tilde{\mathbf{e}}_t^N[f] + \sqrt{N} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, \mathscr{B}_k^N \rangle, \tag{137}$$

where

$$\tilde{\mathbf{e}}_t^N[f] := \frac{1}{\sqrt{N}} \int_0^t \int_{\mathcal{X}\times\mathcal{Y}} \alpha \langle \sigma_*(\cdot, x), \eta_s^N \rangle \langle \nabla f \cdot \nabla \sigma_*(\cdot, x), \eta_s^N \rangle \pi(\mathrm{d}x, \mathrm{d}y) \mathrm{d}s$$
$$- \sqrt{N} \langle f, V_t^N \rangle - \sqrt{N} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \langle f, \mathscr{R}_k^N \rangle - \frac{\sqrt{N}}{N^{1+\beta}} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \sum_{i=1}^N \nabla f(W_k^i)\, \varepsilon_k^i.$$

Step 2. Let η be a limit point of $(\eta^N)_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}^{-J_0+1,j_0}(\mathbf{R}))$. Let N' be such that in distribution $\eta^{N'} \to \eta$ as $N' \to +\infty$. In this step, we pass to the limit in (137) with the test function $\mathsf{f}_2 : x \in \mathbf{R} \mapsto |x|^2$. By Propositions 24 and 27, the sequence $(\eta^{N'}, \sqrt{N'} M^{N'})_{N'\geq 1}$ is tight in $\mathscr{E}$ (see (125)). Let $(\eta^*, \mathscr{G}^*)$ be one of its limit point in $\mathscr{E}$. Up to extracting a subsequence from N', it holds:

$$(\eta^{N'}, \sqrt{N'}\, M^{N'}) \to (\eta^*, \mathscr{G}^*), \text{ as } N' \to +\infty.$$

Considering the marginal distributions, it holds in distribution,

$$\eta^* = \eta \text{ and } \mathscr{G}^* = \mathscr{G} \in \mathcal{C}(\mathbf{R}_+, \mathcal{H}^{-J_0,j_0}(\mathbf{R})).$$

Introduce $\mathcal{C}(\eta^*) \subset \mathbf{R}_+$, whose complementary in $\mathbf{R}_+$ is at most countable, such that for all $u \in \mathcal{C}(\eta^*)$, $s \in \mathbf{R}_+ \mapsto \eta_s^* \in \mathcal{H}^{-J_0+1,j_0}(\mathbf{R})$ is a.s. continuous at u. Then, with the same arguments as those used to derive (131) and using also the fact that $0 \in \mathcal{C}(\eta^*)$, one has for all $t \in \mathcal{C}(\eta^*)$ and in distribution,

$$\mathbf{B}_t[f](\eta^{N'}) - \langle f, \eta_0^{N'} \rangle - \langle f, \sqrt{N'} M_t^{N'} \rangle \to \mathbf{B}_t[f](\eta^*) - \langle f, \eta_0^* \rangle - \langle f, \mathscr{G}_t^* \rangle \text{ as } N' \to +\infty. \tag{138}$$

Let us now deal with the two terms in the right-hand side of (137). Using (26), (27) and **A3**,

$$\begin{aligned}
&\mathbf{E}\Big[\frac{1}{|B_k|}\Big|\sum_{(x,y)\in B_k}(y-g^N_{W_k}(x))\nabla_W\sigma_*(W^i_k,x)\varepsilon^i_k\Big|\Big]\\
&\le 2\mathbf{E}\Big[\frac{1}{|B_k|}\sum_{(x,y)\in B_k}\big|(y-g^N_{W_k}(x))\nabla_W\sigma_*(W^i_k,x)\big|^2\Big]+2\mathbf{E}\Big[|\varepsilon^i_k|^2\Big]\\
&\le C\Big[\mathbf{E}\Big[\frac{1}{|B_k|}\sum_{(x,y)\in B_k}(|y|^2+1)\Big]+1\Big]\le C.
\end{aligned}$$

We now set $f=\mathsf{f}_2$. Then, we have $\mathbf{E}\big[|\langle \mathsf{f}_2,\mathscr{R}^N_k\rangle|\big]\le C(N^{-2}+N^{-7/4})$. Using also the lines below (132) and Lemma 9, it holds:

$$\mathbf{E}\big[|\tilde{\mathsf{e}}^N_t[\mathsf{f}_2]|\big]\le C\Big[\frac{1}{\sqrt{N}}+N^{3/2}\Big(\frac{1}{N^2}+\frac{1}{N^{7/4}}\Big)+\frac{1}{N^{1/4}}\Big]. \tag{139}$$

On the other hand, using **A5** and the law of large number, it holds a.s. as $N\to+\infty$,

$$\sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle \mathsf{f}_2,\mathscr{B}^N_k\rangle=\frac{1}{2N^2}\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}|\varepsilon^i_k|^2\mathsf{f}_2''(\widehat{W}^i_k)=\frac{1}{N^2}\sum_{k=0}^{\lfloor Nt\rfloor-1}\sum_{i=1}^{N}|\varepsilon^i_k|^2\to t\mathbf{E}[|\varepsilon^1_1|^2]. \tag{140}$$

Therefore, using (138),(139) and (140), it holds for all $t\in\mathcal{C}(\eta^*)$, a.s. $\mathbf{B}_t[\mathsf{f}_2](\eta^*)-\langle \mathsf{f}_2,\eta^*_0\rangle-\langle \mathsf{f}_2,\mathscr{G}^*_t\rangle=t\mathbf{E}[|\varepsilon^1_1|^2]$. The mapping[4] $s\in\mathbf{R}_+\mapsto\mathbf{B}_s[\mathsf{f}_2](\eta^*)$ is right continuous and $s\mapsto\langle \mathsf{f}_2,\mathscr{G}^*_s\rangle$ is continuous. By a standard continuity argument (the same as the one used in Proposition 18), it holds a.s. for all $t\in\mathbf{R}_+$, $\mathbf{B}_t[\mathsf{f}_2](\eta^*)-\langle \mathsf{f}_2,\eta^*_0\rangle-\langle \mathsf{f}_2,\mathscr{G}^*_t\rangle=t\mathbf{E}[|\varepsilon^1_1|^2]$. The proof of Proposition 7 is complete. ■

Acknowledgments

The authors are grateful to Benoît Bonnet for fruitful discussions about the proof of Proposition 19. A.D. is grateful for the support received from the Agence Nationale de la Recherche (ANR) of the French government through the program "Investissements d'Avenir" (16-IDEX-0001 CAP 20-25). A.G. is supported by the French ANR under the grant ANR-17-CE40-0030 (project *EFI*) and the Institut Universitaire de France. M.M. acknowledges the support of the the French ANR under the grant ANR-20-CE46-0007 (*SuSa* project). B.N. is supported by the grant IA20Nectoux from the Projet I-SITE Clermont CAP 20-25. The authors thank the referees as well as the Associate Editor for their fruitful remarks.

References

R. Adams and J. Fournier. *Sobolev Spaces*. Elsevier, 2003.

4. For all $m\in\mathcal{D}(\mathbf{R}_+,\mathcal{H}^{-J_0+1,j_0}(\mathbf{R}^d))$ and $f\in\mathcal{H}^{J_0,j_0}(\mathbf{R}^d)$, $t\in\mathbf{R}_+\mapsto\mathbf{B}_t[f](m)$ is right-continuous. This is clear since $t\mapsto\langle f,m_t\rangle$ is right-continuous, and because $s\mapsto\mathbf{U}_s[f](m_s)-\mathbf{V}_s[f](m_s)\in L^\infty_{loc}(\mathbf{R}_+)$.

P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, 2nd edition, 1999.

L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018. doi: 10.1137/16M1080173. URL https://doi.org/10.1137/16M1080173.

Z. Chen, G.M. Rotskoff, J. Bruna, and E. Vanden-Eijnden. A dynamical central limit theorem for shallow neural networks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 22217–22230. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/fc5b3186f1cf0daece964f78259b7ba0-Paper.pdf.

L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/a1afc58c6ca9540d057299ec3016d726-Paper.pdf.

L. Chizat and F. Bach. Implicit bias of gradient descent for wide two-layer neural networks trained with the logistic loss. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 1305–1338. PMLR, 09–12 Jul 2020. URL https://proceedings.mlr.press/v125/chizat20a.html.

V. De Bortoli, A. Durmus, X. Fontaine, and U. Simsekli. Quantitative propagation of chaos for SGD in wide neural networks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 278–288. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/02e74f10e0327ad868d138f2b4fdd6f0-Paper.pdf.

P. Del Moral and A. Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *The Annals of Applied Probability*, 9(2):275–297, 1999.

F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: a central limit theorem. *Electronic Journal of Probability*, 24:1–54, 2019.

A. Descours. *Vers une compréhension mathématiques des réseaux neuronaux profonds par une analyse champ moyen.* PhD thesis, Université Clermont Auvergne, 2023.

W. E. Machine learning and computational mathematics. *Communications in Computational Physics*, 28(5):1639–1670, 2020. ISSN 1815-2406. doi: 10.4208/cicp.oa-2020-0185. URL https://doi.org/10.4208/cicp.oa-2020-0185.

S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*, volume 282. John Wiley & Sons, 2009.

B. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the Mckean-Vlasov model. *Stochastic Processes and their Applications*, 71(1):33–53, 1997.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT press, 2016.

R.M. Gower, N. Loizou, X. Qian, A. Sailanbayev, E. Shulgin, and P. Richtárik. SGD: General analysis and improved rates. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5200–5209. PMLR, 09–15 Jun 2019. URL `https://proceedings.mlr.press/v97/qian19b.html`.

M. Hitsuda and I. Mitoma. Tightness problem and stochastic evolution equation arising from fluctuation phenomena for interacting diffusions. *Journal of Multivariate Analysis*, 19(2):311–328, 1986.

J. Jacod and A. Shiryaev. *Skorokhod Topology and Convergence of Processes*. Springer, 1987.

A. Jakubowski. On the skorokhod topology. In *Annales de l'IHP Probabilités et statistiques*, volume 22, pages 263–285, 1986.

C. Jin, P. Netrapalli, R. Ge, S. Kakade, and M. Jordan. On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points. *Journal of the ACM (JACM)*, 68(2):1–29, 2021.

B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 34(6):727–766, 1998.

N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. Tang. On large-batch training for deep learning: Generalization gap and sharp minima. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL `https://openreview.net/forum?id=H1oyRlYgg`.

T. Kurtz. Semigroups of conditioned shifts and approximation of Markov processes. *The Annals of Probability*, pages 618–642, 1975.

T. Kurtz and J. Xiong. A stochastic evolution equation arising from the fluctuations of a class of interacting particle systems. *Communications in Mathematical Sciences*, 2(3): 325–358, 2004.

C. Ma, S. Wojtowytsch, L. Wu, et al. Towards a mathematical understanding of neural network-based machine learning: what we know and what we don't. *CSIAM Transactions on Applied Mathematics*, 1(4):561–615, 2020.

S. Mei, A. Montanari, and P-M. Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671, 2018.

S. Mei, T. Misiakiewicz, and A. Montanari. Mean-field theory of two-layers neural networks: dimension-free bounds and kernel limit. In *Conference on Learning Theory*, pages 2388–2464. PMLR, 2019.

B. Piccoli and F. Rossi. On properties of the generalized Wasserstein distance. *Archive for Rational Mechanics and Analysis*, 222(3):1339–1365, 2016.

B. Piccoli, F. Rossi, and E. Trélat. Control to flocking of the kinetic Cucker–Smale model. *SIAM Journal on Mathematical Analysis*, 47(6):4685–4719, 2015.

G.M. Rotskoff and E. Vanden-Eijnden. Trainability and accuracy of neural networks: An interacting particle system approach. *Preprint arXiv:1805.00915, to appear in Comm. Pure App. Math.*, 2018.

F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 55. Springer, 2015.

U. Simsekli, M. Gurbuzbalaban, T. Nguyen, G. Richard, and L. Sagun. On the heavy-tailed theory of stochastic gradient descent for deep neural networks, 2019. URL `https://arxiv.org/abs/1912.00018`.

J. Sirignano and K. Spiliopoulos. Mean field analysis of neural networks: A central limit theorem. *Stochastic Processes and their Applications*, 130(3):1820–1852, 2020a.

J. Sirignano and K. Spiliopoulos. Mean field analysis of neural networks: A law of large numbers. *SIAM Journal on Applied Mathematics*, 80(2):725–752, 2020b.

S. L. Smith and Q. V. Le. A bayesian perspective on generalization and stochastic gradient descent. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL `https://openreview.net/forum?id=BJij4yg0Z`.

A-S. Sznitman. Topics in propagation of chaos. In *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, pages 165–251. Springer, 1991. ISBN 978-3-540-46319-1.

C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

S. Wojtowytsch. On the convergence of gradient descent training for two-layer relu-networks in the mean field regime. *arXiv preprint arXiv:2005.13530*, 2020.

J. Wu, W. Hu, H. Xiong, J. Huan, V. Braverman, and Z. Zhu. On the noisy gradient descent that generalizes as SGD. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 10367–10376. PMLR, 13–18 Jul 2020. doi: 10.48550/ARXIV.1906.07405. URL `https://proceedings.mlr.press/v119/wu20c.html`.

Appendix A. A note on relative compactness

In this section, we prove, in our Hilbert setting, that the condition (4.21) in (Kurtz, 1975, Theorem (4.20)) can be replaced by the slightly modified condition, namely the regularity condition of item 2 in Proposition 38 below.

In the following $\mathcal{H}_1$ and $\mathcal{H}_2$ are two Hilbert spaces (whose duals are respectively denoted by $\mathcal{H}_1^{-1}$ and $\mathcal{H}_2^{-1}$) such that $\mathcal{H}_1 \hookrightarrow_{\text{H.S.}} \mathcal{H}_2$.

Proposition 38 *Let $(\mu^N)_{N\geq 1} \subset \mathcal{D}(\mathbf{R}_+, \mathcal{H}_2^{-1})$ be a sequence of processes satisfying the following two conditions:*

1. *Compact containment condition : for every $T > 0$ and $\eta > 0$, there exists $C > 0$ such that*
$$\sup_{N\geq 1} \mathbf{P}\big(\sup_{t\in[0,T]} \|\mu_t^N\|^2_{\mathcal{H}_2^{-1}} > C\big) \leq \eta.$$
2. *Regularity condition : for every $\delta > 0$, $N \geq 1$, $0 \leq t \leq T$ and $0 \leq u \leq (T-t)\wedge\delta$, there exists $F_N(\delta) < \infty$ such that*
$$\mathbf{E}\left[\|\mu_{t+u}^N - \mu_t^N\|^2_{\mathcal{H}_2^{-1}}\right] \leq F_N(\delta),$$
with $\lim_{\delta\to 0} \limsup_{N\geq 1} F_N(\delta) = 0.$

Then, the sequence $(\mu^N)_{N\geq 1} \subset \mathcal{D}(\mathbf{R}_+, \mathcal{H}_1^{-1})$ is relatively compact.

Proof To prove this result, we follow the proof of (Kurtz, 1975, Theorem (4.20)). More precisely, we show that the assumptions of (Kurtz, 1975, Theorem 1) are satisfied (namely conditions (4.2) and (4.3) there), when $E = \mathcal{H}_1^{-1}$ there.

Step 1. The condition (4.2) in (Kurtz, 1975, Theorem 1) is satisfied (when $E = \mathcal{H}_1^{-1}$ there).

We have that $\mathcal{H}_1$ is compactly embedded in $\mathcal{H}_2$ (since a Hilbert-Schmidt embedding is compact). By Schauder's theorem, $\mathcal{H}_2^{-1}$ is compactly embedded in $\mathcal{H}_1^{-1}$. Thus, for all $C > 0$, the set $\{\phi \in \mathcal{H}_1^{-1},\ \|\phi\|_{\mathcal{H}_2^{-1}} \leq C\}$ is compact. Therefore, the condition (4.2) in (Kurtz, 1975) is satisfied.

Step 2. The condition (4.3) in (Kurtz, 1975, Theorem 1) is satisfied (when $E = \mathcal{H}_1^{-1}$ there).

By (Kurtz, 1975, Lemma 4.4), $\{t \mapsto \mu_t^N, t \in \mathbf{R}_+\}_{N\geq 1}$ is relatively compact in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}_1^{-1})$ if for all $\varepsilon > 0$, there exists a tight sequence $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1} \subset \mathcal{D}(\mathbf{R}_+, \mathcal{H}_1^{-1})$ which is ε-close to $\{t \mapsto \mu_t^N, t \in \mathbf{R}_+\}_{N\geq 1}$. Following (Kurtz, 1975), we define, for $\varepsilon > 0$, the sequence $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1} \subset \mathcal{D}(\mathbf{R}_+, \mathcal{H}_1^{-1})$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}_2^{-1})$ of pure jump processes as follows. Let us first introduce, for $N \geq 1$ and $\varepsilon > 0$, $\tau_0^{N,\varepsilon} := 0$ and, for $k > 0$:

$$\tau_k^{N,\varepsilon} := \inf\{t > \tau_{k-1}^{N,\varepsilon} : \|\mu_t^N - \mu^N_{\tau_{k-1}^{N,\varepsilon}}\|_{\mathcal{H}_2^{-1}} > \varepsilon\},$$
$$s_k^{N,\varepsilon} := \sup\{t < \tau_k^{N,\varepsilon} : \|\mu_t^N - \mu^N_{\tau_k^{N,\varepsilon}}\|_{\mathcal{H}_2^{-1}} \geq \varepsilon\}.$$

Then we define, for $\varepsilon > 0$,

$$\mu_t^{N,\varepsilon} := \begin{cases} \mu_0^N & \text{for} \quad t < \frac{1}{2}(s_1^{N,\varepsilon} + \tau_1^{N,\varepsilon}) \\ \mu^N_{\tau_k^{N,\varepsilon}} & \text{for } \frac{1}{2}(s_k^{N,\varepsilon} + \tau_k^{N,\varepsilon}) \leq t < \frac{1}{2}(s_{k+1}^{N,\varepsilon} + \tau_{k+1}^{N,\varepsilon}). \end{cases}$$

We claim that for any $\varepsilon > 0$, the sequence $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1}$ verifies condition (4.2) of (Kurtz, 1975, Theorem 4.1) when $E = \mathcal{H}_1^{-1}$ there. Indeed, by the discussion in the first step above, this follows from the compact containment condition verified by $\{t \mapsto \mu_t^N, t \in \mathbf{R}_+\}_{N\geq 1}$ in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}_2^{-1})$ (see item 1 in Proposition 38) together with the fact that $\sup_{t\in\mathbf{R}_+} \|\mu_t^N - \mu_t^{N,\varepsilon}\|_{\mathcal{H}_2^{-1}} \leq C_0\varepsilon$, where the constant $C_0 > 0$ is independent of N and ε.

It remains to prove that for any $\varepsilon > 0$, $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1}$ satisfies the condition (4.3) in (Kurtz, 1975, Theorem 4.1) when $E = \mathcal{H}_1^{-1}$ there (so that it will be tight for each $\varepsilon > 0$). Since $\mathcal{H}_2^{-1} \hookrightarrow \mathcal{H}_1^{-1}$, it is enough to show that for any $\varepsilon > 0$, $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1}$ satisfies the condition (4.3) in (Kurtz, 1975, Theorem 4.1) when $E = \mathcal{H}_2^{-1}$ there. By (Kurtz, 1975, Lemma 4.5) (and its note) and the construction of $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1}$, it is sufficient to bound $\delta \mapsto \mathbf{P}(\tau_1^{N,\varepsilon} \leq \delta)$ and $\delta \mapsto \mathbf{P}(\tau_{k+1}^{N,\varepsilon} - s_k^{N,\varepsilon} \leq \delta)$ by a function $\delta \mapsto G_N^\varepsilon(\delta)$ such that for every $\varepsilon > 0$,

$$\lim_{\delta\to 0} \limsup_N G_N^\varepsilon(\delta) = 0. \tag{141}$$

Let us prove that we can indeed bound these two terms by such $G_N^\varepsilon(\delta)$ satisfying (141). Recall that by item 2 in Proposition 38, for every $\delta > 0$, $N \geq 1$, $0 \leq t \leq T$ and $0 \leq u \leq (T-t) \wedge \delta$,

$$\mathbf{E}\left[\|\mu_{t+u}^N - \mu_t^N\|^2_{\mathcal{H}_2^{-1}}\right] \leq F_N(\delta), \text{ with } \lim_{\delta\to 0} \limsup_{N\geq 1} F_N(\delta) = 0.$$

Now, introduce as in (Kurtz, 1975), the distance r on $\mathcal{H}_2^{-1}$ defined by $r(\varphi^1, \varphi^2) = \min(1, \|\varphi^1 - \varphi^2\|_{\mathcal{H}_2^{-1}})$. We thus have:

$$\mathbf{E}\left[r\left(\mu_{t+u}^N, \mu_t^N\right)^2\right] \leq F_N(\delta), \tag{142}$$

On the one hand, we have, for $0 < \varepsilon < 1$,

$$\mathbf{P}(\tau_1^{N,\varepsilon} \leq \delta) = \mathbf{P}\big(\|\mu^N_{\tau_1^{N,\varepsilon}\wedge\delta} - \mu_0^N\|_{\mathcal{H}_2^{-1}} > \varepsilon\big) = \mathbf{P}\big(r\big(\mu^N_{\tau_1^{N,\varepsilon}\wedge\delta}, \mu_0^N\big) > \varepsilon\big) \leq \frac{1}{\varepsilon^2}\mathbf{E}\big[r\big(\mu^N_{\tau_1^{N,\varepsilon}\wedge\delta}, \mu_0^N\big)^2\big]. \tag{143}$$

On the other hand, we have, for $k \geq 1$ and $0 < \varepsilon < 1$,

$$\begin{aligned} \mathbf{P}(\tau_{k+1}^{N,\varepsilon} - s_k^{N,\varepsilon} \leq \delta) \leq \mathbf{P}(\tau_{k+1}^{N,\varepsilon} - \tau_k^{N,\varepsilon} \leq \delta) &= \mathbf{P}\big(r(\mu^N_{\tau_{k+1}^{N,\varepsilon}\wedge(\tau_k^{N,\varepsilon}+\delta)}, \mu^N_{\tau_k^{N,\varepsilon}}) > \varepsilon\big) \\ &\leq \frac{1}{\varepsilon^2}\mathbf{E}\big[r\big(\mu^N_{\tau_{k+1}^{N,\varepsilon}\wedge(\tau_k^{N,\varepsilon}+\delta)}, \mu^N_{\tau_k^{N,\varepsilon}}\big)^2\big]. \end{aligned} \tag{144}$$

From (142), and because the stopping times appearing in (143) and (144) can be approximated by sequences of decreasing discrete stopping times, we can indeed bound $\delta \mapsto \mathbf{P}(\tau_1^{N,\varepsilon} \leq \delta)$ and $\delta \mapsto \mathbf{P}(\tau_{k+1}^{N,\varepsilon} - s_k^{N,\varepsilon} \leq \delta)$ by $F_N(\delta)$ which satisfies (141). Consequently, for each $0 < \varepsilon < 1$, the condition (4.3) in (Kurtz, 1975, Theorem 4.1) is satisfied for the sequence $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1}$ when $E = \mathcal{H}_2^{-1}$ there. We can thus apply (Kurtz, 1975, Theorem 4.1) to $\{t \mapsto \mu_t^{N,\varepsilon}, t \in \mathbf{R}_+\}_{N\geq 1}$, which is therefore tight in $\mathcal{D}(\mathbf{R}_+, \mathcal{H}_1^{-1})$ for each $0 < \varepsilon < 1$. Using (Kurtz, 1975, Lemma 4.4) (with $E = \mathcal{H}_1^{-1}$ there), this concludes the proof of the lemma. ■

Appendix B. Technical lemmata

In this section we state and prove Lemma 39, Lemma 40 and Lemma 41.

Lemma 39 *Let $\beta \geq 1/2$ and assume* **A1**-**A7**. *Recall $\mathcal{H}^{J_1,j_1}(\mathbf{R}^d) \hookrightarrow \mathcal{H}^{L,\gamma}(\mathbf{R}^d)$ (see (14)). Then, for all $T > 0$, there exists $C > 0$ such that for all $f \in \mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$, $N \geq 1$, and $t \in [0,T]$, it holds:*

(i)

$$\mathbf{E}\left[\sum_{k=0}^{\lfloor Nt\rfloor-1} 2\langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle\sqrt{N}\langle f, M_k^N\rangle + 4N\langle f, M_k^N\rangle^2\right] \leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

(ii)

$$\mathbf{E}\left[\sum_{k=0}^{\lfloor Nt\rfloor-1} \frac{2\sqrt{N}}{N^{1+\beta}}\langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle \sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i + \frac{4}{N^{1+2\beta}}\Big|\sum_{i=1}^N \nabla f(W_k^i)\cdot\varepsilon_k^i\Big|^2\right]$$
$$\leq \frac{C}{N^{2\beta-1}}\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

(iii)

$$\mathbf{E}\left[\sum_{k=0}^{\lfloor Nt\rfloor-1} N|\langle f, R_k^N\rangle|^2\right] \leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}\left[\frac{1}{N^2} + \frac{N^2}{N^{4\beta}}\right].$$

(iv)

$$\mathbf{E}\left[\sum_{k=0}^{\lfloor Nt\rfloor-1} \langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle\sqrt{N}\langle f, R_k^N\rangle\right] \leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}\left[1 + \frac{1}{N} + \frac{N^3}{N^{4\beta}}\right] + \mathbf{E}\left[\int_0^t \langle f, \Upsilon_s^N\rangle^2 \mathrm{d}s\right].$$

(v)

$$\mathbf{E}\left[\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1} \langle f, \Upsilon^N_{\frac{k+1}{N}-}\rangle \mathbf{a}_k^N[f] - \sqrt{N}\int_0^t \langle f, \Upsilon_s^N\rangle \mathbf{L}_s^N[f]\mathrm{d}s\Big|\right] \leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

(vi)

$$\mathbf{E}\left[\sum_{k=0}^{\lfloor Nt\rfloor-1}|\mathbf{a}_k^N[f]|^2\right]\leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

Proof Let $T>0$ and $f\in\mathcal{H}^{J_1,j_1}(\mathbf{R}^d)$. In what follows, $C>0$ is a constant, independent of $N\geq 1$, $t\in[0,T]$, f, and $k\in\{0,\dots,\lfloor NT\rfloor-1\}$ which can change from one occurence to another. We recall that for $N\geq 1$ and $k\geq 1$, $\mathcal{F}_k^N$ is the $\boldsymbol{\sigma}$-algebra generated by $\{W_0^i\}_{i=1}^N$, B_j and $(\varepsilon_j^i)_{i=1}^N$ for $j=0,\dots,k-1$, and that $\mathcal{F}_0^N:=\boldsymbol{\sigma}\{\{W_0^i\}_{i=1}^N\}$, see (9). Recall also the definitions of M_k^N and R_k^N in (21) and (19) respectively, of $\mathbf{a}_s^N$ and $\mathbf{L}_s^N$ in (94), and that for $N\geq 1$ and $t\in\mathbf{R}_+$, $\Upsilon_t^N=\sqrt{N}(\mu_t^N-\bar\mu_t^N)$ (see also (81)). We start by proving item *(i)* in Lemma 39. For all $t\in[0,T]$, because for all $a\in\mathbf{N}$ and $b\in\{1,\dots,N\}$, W_a^b is $\mathcal{F}_a^N$-measurable and $\varepsilon_a^b \perp\!\!\!\perp \mathcal{F}_a^N$ (see **A5**) together with the fact that $\bar X_s^b$ is $\mathcal{F}_0^N$-measurable (for all $s\geq 0$), one has using also (41):

$$\begin{aligned}\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle\sqrt{N}\langle f,M_k^N\rangle\Big]&=N\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}\Big[\langle f,\mu^N_{\frac{k+1}{N}-}-\bar\mu^N_{\frac{k+1}{N}}\rangle\langle f,M_k^N\rangle\Big]\\&=N\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}\big[\langle f,\nu_k^N\rangle\mathbf{E}\left[\langle f,M_k^N\rangle\big|\,\mathcal{F}_k^N\right]\big]\\&\quad-N\sum_{k=0}^{\lfloor Nt\rfloor-1}\mathbf{E}\Big[\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle\mathbf{E}\Big[\langle f,M_k^N\rangle\big|\mathcal{F}_k^N\Big]\Big]=0.\end{aligned}\tag{145}$$

By (36) $\mathbf{E}\left[\langle f,M_k^N\rangle^2\right]\leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}/N^2$. Together with (145), we deduce item *(i)*.

Let us now prove item *(ii)*. We have, using **A5**,

$$\begin{aligned}\mathbf{E}\big[\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle\sum_{i=1}^N\nabla f(W_k^i)\cdot\varepsilon_k^i\big]&=\sqrt{N}\,\mathbf{E}\Big[(\langle f,\nu_k^N\rangle-\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle)\sum_{i=1}^N\nabla f(W_k^i)\cdot\varepsilon_k^i\Big]\\&=\sqrt{N}\sum_{i=1}^N\mathbf{E}\left[(\langle f,\nu_k^N\rangle-\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle)\nabla f(W_k^i)\right]\cdot\mathbf{E}[\varepsilon_k^i]=0.\end{aligned}$$

On the other hand, using (46), $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$ (see (6)), and the same arguments as those used in (47), it holds:

$$\mathbf{E}\big[\big|\sum_{i=1}^N\nabla f(W_k^i)\cdot\varepsilon_k^i\big|^2\big]=\sum_{i=1}^N\mathbf{E}\big[\big|\nabla f(W_k^i)\cdot\varepsilon_k^i\big|^2\big]\leq CN\|f\|^2_{\mathcal{C}^{2,\gamma_*}}\leq CN\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

This ends the proof of item *(ii)*. Item *(iii)* is a direct consequence of (51) and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$.

Let us now prove item *(iv)*. We have that

$$\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle\sqrt{N}\langle f,R_k^N\rangle\leq\sum_{k=0}^{\lfloor Nt\rfloor-1}\frac{1}{N}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2+\sum_{k=0}^{\lfloor Nt\rfloor-1}N^2|\langle f,R_k^N\rangle|^2.\tag{146}$$

On the one hand, by item *(iii)*,

$$\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1} N^2|\langle f, R_k^N\rangle|^2\Big] \le C\|f\|^2_{\mathcal{H}^{L,\gamma}}\Big[\frac{1}{N}+\frac{N^3}{N^{4\beta}}\Big]. \tag{147}$$

On the other hand, we have

$$\begin{aligned}\Big|\int_0^{\frac{\lfloor Nt\rfloor}{N}}\langle f,\Upsilon_s^N\rangle^2\mathrm{d}s-\sum_{k=0}^{\lfloor Nt\rfloor-1}\frac{1}{N}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big| &= \Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\Big(\langle f,\Upsilon_s^N\rangle^2-\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big)\mathrm{d}s\Big|\\ &\le \sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\Big|\langle f,\Upsilon_s^N\rangle^2-\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big|\mathrm{d}s.\end{aligned}$$

Let $0\le k<\lfloor Nt\rfloor$ and $s\in\big(\frac{k}{N},\frac{k+1}{N}\big)$. We have

$$\begin{aligned}\langle f,\Upsilon_s^N\rangle^2-\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2 &= N\left[\big(\langle f,\nu_k^N\rangle-\langle f,\bar\mu_s^N\rangle\big)^2-\big(\langle f,\nu_k^N\rangle-\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle\big)^2\right]\\ &= N\left[2\langle f,\nu_k^N\rangle(\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle-\langle f,\bar\mu_s^N\rangle)+\langle f,\bar\mu_s^N\rangle^2-\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle^2\right]\\ &= N\Big[2\langle f,\nu_k^N\rangle\Big(\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle-\langle f,\bar\mu_s^N\rangle\Big)\\ &\quad+\big(\langle f,\bar\mu_s^N\rangle-\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle\big)\big(\langle f,\bar\mu_s^N\rangle+\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle\big)\Big].\end{aligned} \tag{148}$$

It holds using (83) and that $|\bar X_{\frac{k+1}{N}}-\bar X_s|\le C/N$ (by (82)):

$$\begin{aligned}&|\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle-\langle f,\bar\mu_s^N\rangle| = \Big|\frac{1}{N}\sum_{i=1}^N f(\bar X^i_{\frac{k+1}{N}})-f(\bar X^i_s)\Big|\\ &\le \frac{1}{N}\Big[\sum_{i=1}^N|\bar X^i_{\frac{k+1}{N}}-\bar X^i_s||\nabla f(\bar X^i_s)|+C|\bar X^i_{\frac{k+1}{N}}-\bar X^i_s|^2\sup_{t\in(0,1)}|\nabla^2 f|(t\bar X^i_{\frac{k+1}{N}}+(1-t)\bar X^i_s)\Big]\\ &\le \frac{C}{N}\|f\|_{\mathcal{C}^{2,\gamma_*}}\sum_{i=1}^N|\bar X^i_{\frac{k+1}{N}}-\bar X^i_s|+|\bar X^i_{\frac{k+1}{N}}-\bar X^i_s|^2\le\frac{C}{N}\|f\|_{\mathcal{H}^{L,\gamma}}.\end{aligned} \tag{149}$$

Going back to (148), and using also (83), we have:

$$\begin{aligned}\Big|\langle f,\Upsilon_s^N\rangle^2-\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big| &\le C\|f\|_{\mathcal{H}^{L,\gamma}}\left(|\langle f,\nu_k^N\rangle|+\Big|\langle f,\bar\mu_s^N\rangle+\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle\Big|\right)\\ &\le C\|f\|_{\mathcal{H}^{L,\gamma}}\Big(\frac{\|f\|_{\mathcal{C}^{2,\gamma_*}}}{N}\sum_{i=1}^N(1+|W_k^i|^{\gamma_*})+C\|f\|_{\mathcal{C}^{2,\gamma_*}}\Big).\end{aligned}$$

Therefore, using Lemma 8, we have shown that

$$\begin{aligned}\mathbf{E}\Big[\Big|\int_0^{\frac{\lfloor Nt\rfloor}{N}}\langle f,\Upsilon_s^N\rangle^2\mathrm{d}s-\sum_{k=0}^{\lfloor Nt\rfloor-1}\frac{1}{N}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big|\Big] &\le \sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\mathbf{E}\Big[\Big|\langle f,\Upsilon_s^N\rangle^2-\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big|\Big]\mathrm{d}s\\ &\le C\|f\|^2_{\mathcal{H}^{L,\gamma}},\end{aligned}$$

so that

$$\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1}\frac{1}{N}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}+\mathbf{E}\Big[\int_0^{\frac{\lfloor Nt\rfloor}{N}}\langle f,\Upsilon^N_s\rangle^2\mathrm{d}s\Big].$$

On the other hand, using (48), (83), and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$,

$$\begin{aligned}\mathbf{E}[\langle f,\Upsilon^N_s\rangle^2]\le CN\Big[\mathbf{E}\big[\langle f,\mu^N_s\rangle^2\big]+\mathbf{E}\big[\langle f,\bar\mu^N_s\rangle^2\big]\Big]&\le CN\Big[\|f\|^2_{\mathcal{C}^{2,\gamma_*}}+\frac{1}{N^2}\mathbf{E}\big[\big|\sum_{i=1}^N f(\bar X^i_t)\big|^2\big]\Big]\\&\le CN\Big[\|f\|^2_{\mathcal{C}^{2,\gamma_*}}+\frac{1}{N}\mathbf{E}\big[\sum_{i=1}^N|f(\bar X^i_t)|^2\big]\Big]\\&\le CN\|f\|^2_{\mathcal{H}^{L,\gamma}}.\end{aligned}$$

Therefore, it holds: $\mathbf{E}[\int_{\frac{\lfloor Nt\rfloor}{N}}^t\langle f,\Upsilon^N_s\rangle^2\mathrm{d}s]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}$. Hence,

$$\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1}\frac{1}{N}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle^2\Big]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}+\mathbf{E}\Big[\int_0^t\langle f,\Upsilon^N_s\rangle^2\mathrm{d}s\Big]\tag{150}$$

Item *(iv)* is then a consequence of (146), (147), and (150).

Let us now prove item *(v)*. We have (see (94))

$$\begin{aligned}&\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle\mathbf{a}^N_k[f]-\sqrt{N}\int_0^{\frac{\lfloor Nt\rfloor}{N}}\langle f,\Upsilon^N_s\rangle\mathbf{L}^N_s[f]\mathrm{d}s\Big]\\&\qquad=\sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\mathbf{E}\Big[\Big(\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle-\langle f,\Upsilon^N_s\rangle\Big)\mathbf{L}^N_s[f]\Big]\mathrm{d}s.\end{aligned}$$

Using (149), for $s\in\big(\frac{k}{N},\frac{k+1}{N}\big)$, it holds:

$$\big|\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle-\langle f,\Upsilon^N_s\rangle\big|=\sqrt{N}\,\big|\langle f,\bar\mu^N_s\rangle-\langle f,\bar\mu^N_{\frac{k+1}{N}}\rangle\big|\le C\frac{\|f\|_{\mathcal{H}^{L,\gamma}}}{\sqrt{N}},$$

and using (34), Lemma 8, and $\mathcal{H}^{L,\gamma}(\mathbf{R}^d)\hookrightarrow\mathcal{C}^{2,\gamma_*}(\mathbf{R}^d)$, one deduces that:

$$\mathbf{E}\big[\big|\mathbf{L}^N_s[f]\big|^2\big]\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

Thus,

$$\begin{aligned}&\mathbf{E}\Big[\Big|\sum_{k=0}^{\lfloor Nt\rfloor-1}\langle f,\Upsilon^N_{\frac{k+1}{N}-}\rangle\mathbf{a}^N_k[f]-\sqrt{N}\int_0^{\frac{\lfloor Nt\rfloor}{N}}\langle f,\Upsilon^N_s\rangle\mathbf{L}^N_s[f]\mathrm{d}s\Big|\Big]\\&\le\sqrt{N}\sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}C\frac{\|f\|^2_{\mathcal{H}^{L,\gamma}}}{\sqrt{N}}\mathrm{d}s\le C\|f\|^2_{\mathcal{H}^{L,\gamma}}.\end{aligned}$$

In addition we have:

$$\mathbf{E}\Big[\sqrt{N}\Big|\int_{\frac{\lfloor Nt\rfloor}{N}}^{t}\langle f,\Upsilon_s^N\rangle \mathbf{L}_s^N[f]\mathrm{d}s\Big|\Big] \leq \sqrt{N}\int_{\frac{\lfloor Nt\rfloor}{N}}^{t}\sqrt{\mathbf{E}[|\langle f,\Upsilon_s^N\rangle|^2]}\sqrt{\mathbf{E}[|\mathbf{L}_s^N[f]|^2]}\mathrm{d}s$$
$$\leq C\|f\|^2_{\mathcal{H}^{L,\gamma}}.$$

We have thus proved item *(v)*.

Finally,

$$\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1}|\mathbf{a}_k^N[f]|^2\Big] = N\mathbf{E}\Big[\sum_{k=0}^{\lfloor Nt\rfloor-1}\Big|\int_{\frac{k}{N}}^{\frac{k+1}{N}}\mathbf{L}_s^N[f]\mathrm{d}s\Big|^2\Big] \leq \sum_{k=0}^{\lfloor Nt\rfloor-1}\int_{\frac{k}{N}}^{\frac{k+1}{N}}\mathbf{E}[|\mathbf{L}_s^N[f]|^2]\mathrm{d}s$$
$$\leq C\|f\|^2_{\mathcal{H}^{L,\gamma}},$$

which proves item *(vi)*. This ends the proof of the lemma. ■

Lemma 40 *Let* $J \geq 1$ *and* $\gamma \geq 0$. *For* $x \in \mathcal{X}$, *recall the definition of* T_x *(see* (99)*),* $\mathsf{T}_x : f \in \mathcal{H}^{J,\gamma}(\mathbf{R}^d) \mapsto \nabla f \cdot \nabla\sigma_*(\cdot,x) \in \mathcal{H}^{J-1,\gamma}(\mathbf{R}^d)$. *Then, there exists* $C > 0$ *such that for any* $\Upsilon \in \mathcal{H}^{-J+1,\gamma}(\mathbf{R}^d)$ *and* $x \in \mathcal{X}$,

$$|\langle\Upsilon, \mathsf{T}_x^*\Upsilon\rangle_{\mathcal{H}^{-J,\gamma}}| \leq C\|\Upsilon\|^2_{\mathcal{H}^{-J,\gamma}}. \tag{151}$$

This result is stronger than what one obtains with the Cauchy-Schwarz inequality. Indeed, the Cauchy-Schwarz inequality only implies

$$|\langle\Upsilon, \mathsf{T}_x^*\Upsilon\rangle_{\mathcal{H}^{-J,\gamma}}| \leq \|\Upsilon\|_{\mathcal{H}^{-J,\gamma}}\|\mathsf{T}_x^*\Upsilon\|_{\mathcal{H}^{-J,\gamma}} \leq C\|\Upsilon\|_{\mathcal{H}^{-J,\gamma}}\|\Upsilon\|_{\mathcal{H}^{-J+1,\gamma}}.$$

Let us mention that Lemma 40 extends (Sirignano and Spiliopoulos, 2020a, Lemma B.1) to the non compact and weighted case.

Proof Let $x \in \mathcal{X}$ and $\Upsilon \in \mathcal{H}^{-J+1,\gamma}(\mathbf{R}^d) \hookrightarrow \mathcal{H}^{-J,\gamma}(\mathbf{R}^d)$. By the Riesz representation theorem, there exists a unique $\Psi \in \mathcal{H}^{J,\gamma}(\mathbf{R}^d)$ such that,

$$\langle f,\Upsilon\rangle = \langle f,\Psi\rangle_{\mathcal{H}^{J,\gamma}}, \text{ for } f \in \mathcal{H}^{J,\gamma}(\mathbf{R}^d).$$

We set $F(\Upsilon) = \Psi$. The density of $\mathcal{C}_c^\infty(\mathbf{R}^d)$ in $\mathcal{H}^{J,\gamma}(\mathbf{R}^d)$ implies that $\{\Upsilon \in \mathcal{H}^{-J,\gamma}(\mathbf{R}^d) : F(\Upsilon) \in \mathcal{C}_c^\infty(\mathbf{R}^d)\}$ is dense in $\mathcal{H}^{-J,\gamma}(\mathbf{R}^d)$. It is thus sufficient to show (151) for Υ such that $\Psi = F(\Upsilon) \in \mathcal{C}_c^\infty(\mathbf{R}^d)$. We have

$$\langle\Upsilon, \mathsf{T}_x^*\Upsilon\rangle_{\mathcal{H}^{-J,\gamma}} = \langle\Psi, \mathsf{T}_x^*\Upsilon\rangle = \langle\mathsf{T}_x\Psi,\Upsilon\rangle = \langle\mathsf{T}_x\Psi,\Psi\rangle_{\mathcal{H}^{J,\gamma}}. \tag{152}$$

Let us prove that $|\langle\mathsf{T}_x\Psi,\Psi\rangle_{\mathcal{H}^{J,\gamma}}| \leq C\|\Psi\|^2_{\mathcal{H}^{J,\gamma}}$ for $\Psi \in \mathcal{C}_c^\infty(\mathbf{R}^d)$. By definition, we have

$$\langle\mathsf{T}_x\Psi,\Psi\rangle_{\mathcal{H}^{J,\gamma}} = \sum_{|k|\leq J}\int_{\mathbf{R}^d}\Big[D^k(\nabla\Psi(w)\cdot\nabla\sigma_*(w,x))D^k\Psi(w)\Big] \times \frac{1}{1+|w|^{2\gamma}}\mathrm{d}w.$$

In the previous sum, the only terms involving derivatives of Ψ of order greater than J are the terms for which $|k| = J$. Therefore, it is sufficient to only deal with such k. Pick a multi-index k such that $|k| = J$. For all $x \in \mathcal{X}$, we have

$$\begin{aligned}
&\int_{\mathbf{R}^d} \left[D^k(\nabla\Psi(w)\cdot\nabla\sigma_*(w,x))D^k\Psi(w)\right] \times \frac{\mathrm{d}w}{1+|w|^{2\gamma}}\\
&= \int_{\mathbf{R}^d} D^k\Big(\sum_{i=1}^d \partial_i\Psi(w)\partial_i\sigma_*(w,x)\Big) \times \frac{D^k\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w\\
&= \sum_{i=1}^d \int_{\mathbf{R}^d} D^k\left(\partial_i\Psi(w)\partial_i\sigma_*(w,x)\right) \times \frac{D^k\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w.
\end{aligned}$$

Let us consider the case when $i = 1$ and $k = (J, 0\ldots, 0)$. The other cases can be treated similarly. For all $x \in \mathcal{X}$,

$$\begin{aligned}
&\int_{\mathbf{R}^d} D^k\left(\partial_1\Psi(w)\partial_1\sigma_*(w,x)\right) \times \frac{D^k\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w\\
&= \int_{\mathbf{R}^d} \partial_1^J\left(\partial_1\Psi(w)\partial_1\sigma_*(w,x)\right) \times \frac{\partial_1^J\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w\\
&= \int_{\mathbf{R}^d} \partial_1^{J+1}\Psi(w)\partial_1\sigma_*(w,x) \times \frac{\partial_1^J\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w\\
&\quad + \sum_{j=0}^{J-1}\binom{J}{j}\int_{\mathbf{R}^d} \partial_1^{j+1}\Psi(w)\partial_1^{J-j+1}\sigma_*(w,x) \times \frac{\partial_1^J\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w. \qquad (153)
\end{aligned}$$

Since σ_* and all its derivatives are bounded, one has:

$$\sum_{j=0}^{J-1}\binom{J}{j}\int_{\mathbf{R}^d}\left|\partial_1^{j+1}\Psi(w)\partial_1^{J-j+1}\sigma_*(w,x) \times \frac{\partial_1^J\Psi(w)}{1+|w|^{2\gamma}}\right|\mathrm{d}w \le C\|\Psi\|_{\mathcal{H}^{J,\gamma}}.$$

Let us now deal with the first term in the right-hand side of (153). By Fubini's theorem :

$$\begin{aligned}
&\int_{\mathbf{R}^d} \partial_1^{J+1}\Psi(w)\partial_1\sigma_*(w,x) \times \frac{\partial_1^J\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w\\
&= \int_{\mathbf{R}^{d-1}}\int_{\mathbf{R}} \partial_1^{J+1}\Psi(z,w')\partial_1\sigma_*(z,w',x) \times \frac{\partial_1^J\Psi(z,w')}{1+|(z,w')|^{2\gamma}}\mathrm{d}z\mathrm{d}w'.
\end{aligned}$$

An integration by parts yields, for all $w' \in \mathbf{R}^{d-1}$, using that Ψ is compactly supported,

$$\begin{aligned}
&2\int_{\mathbf{R}} \partial_1^{J+1}\Psi(z,w')\partial_1\sigma_*(z,w',x) \times \frac{\partial_1^J\Psi(z,w')}{1+|(z,w')|^{2\gamma}}\mathrm{d}z\\
&= \left[|\partial_1^J\Psi(z,w')|^2\frac{\partial_1\sigma_*(z,w',x)}{1+|(z,w')|^{2\gamma}}\right]_{-\infty}^{+\infty} - \int_{\mathbf{R}} |\partial_1^J\Psi(z,w')|^2\partial_1\left(\frac{\partial_1\sigma_*(z,w',x)}{1+|(z,w')|^{2\gamma}}\right)\mathrm{d}z\\
&= \int_{\mathbf{R}} |\partial_1^J\Psi(z,w')|^2 \times \frac{2\gamma z|(z,w')|^{2\gamma-2}\partial_1\sigma_*(z,w',x) - \partial_1^2\sigma_*(z,w',x)(1+|(z,w')|^{2\gamma})}{(1+|(z,w')|^{2\gamma})^2}\mathrm{d}z.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\Big|\int_{\mathbf{R}^d} \partial_1^{J+1}\Psi(w)\partial_1\sigma_*(w,x) \times \frac{\partial_1^J\Psi(w)}{1+|w|^{2\gamma}}\mathrm{d}w\Big|\\
&= \frac{1}{2}\Big|\int_{\mathbf{R}^{d-1}}\int_{\mathbf{R}} |\partial_1^J\Psi(z,w')|^2\\
&\quad\times \frac{2\gamma w|(z,w')|^{2\gamma-2}\partial_1\sigma_*(z,w',x)-\partial_1^2\sigma_*(z,w',x)(1+|(z,w')|^{2\gamma})}{(1+|(z,w')|^{2\gamma})^2}\mathrm{d}z\mathrm{d}w'\Big|\\
&\le C\int_{\mathbf{R}^{d-1}}\int_{\mathbf{R}} \frac{|\partial_1^J\Psi(z,w')|^2}{1+|(z,w')|^{2\gamma}} \times \mathrm{d}z\mathrm{d}w' \le C\|\Psi\|^2_{\mathcal{H}^{J,\gamma}}.
\end{aligned}$$

To summarize, we have shown the existence of $C<\infty$ (independent of x) such that for any $\Psi \in \mathcal{C}_c^\infty(\mathbf{R}^d)$, $|\langle \mathsf{T}_x\Psi,\Psi\rangle_{\mathcal{H}^{J,\gamma}}| \le C\|\Psi\|^2_{\mathcal{H}^{J,\gamma}}$. Consequently, by (152), $|\langle \Upsilon, \mathsf{T}_x^*\Upsilon\rangle_{\mathcal{H}^{-J,\gamma}}| \le C\|\Upsilon\|^2_{\mathcal{H}^{J,\gamma}}$. This completes the proof of the lemma. ■

Lemma 41 *Let $N\ge 1$ and $f:\mathbf{R}_+\to\mathbf{R}$ be a piecewise continuous function whose jumps occur only at the times k/N, $k\ge 1$. Introduce the function $g:\mathbf{R}_+\to\mathbf{R}$ defined by, for all $t\ge 0$, $g(t)=\sum_{k=0}^{\lfloor Nt\rfloor-1}\alpha_k$ where for all $k\ge 0$, $\alpha_k\in\mathbf{R}$ (recall the convention $\sum_{k=0}^{-1}=0$). Set for $t\ge 0$,*

$$F(t)=\int_0^t f(s)\mathrm{d}s \ \text{and}\ \psi(t)=F(t)+g(t).$$

Then, for all $t\ge 0$,

$$\psi(t)^2 = 2\int_0^t f(s)\psi(s)\mathrm{d}s + \sum_{k=0}^{\lfloor Nt\rfloor-1}|\alpha_k|^2 + 2\sum_{k=0}^{\lfloor Nt\rfloor-1}\psi\Big(\frac{k+1}{N}^-\Big)\Big(\underbrace{g\Big(\frac{k+1}{N}\Big)-g\Big(\frac{k}{N}\Big)}_{=\alpha_k}\Big).$$

Proof For all $k\ge 0$ and $t\in[\frac{k}{N},\frac{k+1}{N})$, it holds $\psi(t)^2-\psi(\frac{k}{N})^2 = 2\int_{\frac{k}{N}}^t \psi'(s)\psi(s)\mathrm{d}s$. Letting $t\to\frac{k+1}{N}^-$, we obtain

$$\psi\Big(\frac{k+1}{N}^-\Big)^2-\psi\Big(\frac{k}{N}\Big)^2 = 2\int_{\frac{k}{N}}^{\frac{k+1}{N}}\psi'(s)\psi(s)\mathrm{d}s. \tag{154}$$

Since F is continuous and by definition of g, it holds $\psi(\frac{k+1}{N}^-)^2 = (F(\frac{k+1}{N})+g(\frac{k}{N}))^2$. Hence,

$$\psi\Big(\frac{k+1}{N}^-\Big)^2-\psi\Big(\frac{k}{N}\Big)^2 = F\Big(\frac{k+1}{N}\Big)^2-F\Big(\frac{k}{N}\Big)^2+2g\Big(\frac{k}{N}\Big)\Big(F\Big(\frac{k+1}{N}\Big)-F\Big(\frac{k}{N}\Big)\Big).$$

Therefore, (154) reads (using also that $g'(s)=0$ for all $s\in(\frac{k}{N},\frac{k+1}{N})$)

$$F\Big(\frac{k+1}{N}\Big)^2-F\Big(\frac{k}{N}\Big)^2+2g\Big(\frac{k}{N}\Big)\Big(F\Big(\frac{k+1}{N}\Big)-F\Big(\frac{k}{N}\Big)\Big) = 2\int_{\frac{k}{N}}^{\frac{k+1}{N}} f(s)\psi(s)\mathrm{d}s.$$

Now, for all $t \geq 0$, denoting $k = \lfloor Nt \rfloor$,

$$\begin{aligned}
2\int_0^t f(s)\psi(s)\mathrm{d}s &= \sum_{j=0}^{k-1} 2\int_{\frac{j}{N}}^{\frac{j+1}{N}} f(s)\psi(s)\mathrm{d}s + 2\int_{\frac{k}{N}}^{t} f(s)\psi(s)\mathrm{d}s \\
&= \sum_{j=0}^{k-1} F\Big(\frac{j+1}{N}\Big)^2 - F\Big(\frac{j}{N}\Big)^2 \\
&\quad + 2g\Big(\frac{j}{N}\Big)\Big(F\Big(\frac{j+1}{N}\Big) - F\Big(\frac{j}{N}\Big)\Big) + \psi(t)^2 - \psi\Big(\frac{k}{N}\Big)^2 \\
&= F\Big(\frac{k}{N}\Big)^2 + \sum_{j=0}^{k-1} 2g\Big(\frac{j}{N}\Big)\Big(F\Big(\frac{j+1}{N}\Big) - F\Big(\frac{j}{N}\Big)\Big) + \psi(t)^2 - \psi\Big(\frac{k}{N}\Big)^2 \\
&= F\Big(\frac{k}{N}\Big)^2 + \sum_{j=0}^{k-2} 2F\Big(\frac{j+1}{N}\Big)\Big(g\Big(\frac{j}{N}\Big) - g\Big(\frac{j+1}{N}\Big)\Big) \\
&\quad + 2g\Big(\frac{k-1}{N}\Big)F\Big(\frac{k}{N}\Big) + \psi(t)^2 - \psi\Big(\frac{k}{N}\Big)^2 \\
&= -g\Big(\frac{k}{N}\Big)^2 + \sum_{j=0}^{k-1} 2F\Big(\frac{j+1}{N}\Big)\Big(g\Big(\frac{j}{N}\Big) - g\Big(\frac{j+1}{N}\Big)\Big) + \psi(t)^2.
\end{aligned}$$

Hence,

$$\psi(t)^2 = 2\int_0^t f(s)\psi(s)\mathrm{d}s + g\Big(\frac{k}{N}\Big)^2 + 2\sum_{j=0}^{k-1} F\Big(\frac{j+1}{N}\Big)\Big(g\Big(\frac{j+1}{N}\Big) - g\Big(\frac{j}{N}\Big)\Big).$$

Using that $g(0) = 0$, one can write $g(\frac{k}{N})^2 = \sum_{j=0}^{k-1} |\alpha_k|^2 + 2\sum_{j=0}^{k-1}(g(\frac{j+1}{N}) - g(\frac{j}{N}))g(\frac{j}{N})$. This yields,

$$\begin{aligned}
\psi(t)^2 &= 2\int_0^t f(s)\psi(s)\mathrm{d}s + \sum_{j=0}^{k-1} |\alpha_k|^2 + 2\sum_{j=0}^{k-1} \Big(F\Big(\frac{j+1}{N}\Big) + g\Big(\frac{j}{N}\Big)\Big)\Big(g\Big(\frac{j+1}{N}\Big) - g\Big(\frac{j}{N}\Big)\Big) \\
&= 2\int_0^t f(s)\psi(s)\mathrm{d}s + \sum_{j=0}^{k-1} |\alpha_k|^2 + 2\sum_{j=0}^{k-1} \psi\Big(\frac{j+1}{N}^{-}\Big)\Big(g\Big(\frac{j+1}{N}\Big) - g\Big(\frac{j}{N}\Big)\Big),
\end{aligned}$$

which is the desired formula. ■

Remark 42 *Notice by Lemma 41 and* (26) *(with $m = 4$ there), if $\alpha_k = \alpha_k^1 + \alpha_k^2 + \alpha_k^3 + \alpha_k^4$, it holds*

$$\psi(t)^2 \leq 2\int_0^t f(s)\psi(s)\mathrm{d}s + 4\sum_{\ell=1}^{4} \sum_{k=0}^{\lfloor Nt \rfloor - 1} |\alpha_k^\ell|^2 + 2\sum_{\ell=1}^{4} \sum_{k=0}^{\lfloor Nt \rfloor - 1} \alpha_k^\ell \, \psi\Big(\frac{k+1}{N}^{-}\Big).$$