Sharp Wasserstein Convergence Rates for Empirical Path Laws of Itô Processes*

Xihao He † Fengyi Yuan ‡

Abstract

We establish the sharp logarithmic order $(\log N)^{-1/2}$ for the expected p-Wasserstein distance, induced by the supremum norm, between the empirical law of N independent copies of a continuous Itô process and their common path law. We only assume that the initial condition and the drift and diffusion integrands are controlled by a time-uniform random upper bound with a finite ρ-moment for some $\rho > p \geq 1$. Under this assumption, we use an adaptive random time interval partition argument, which leads to a $(\log n)^{-1/2}$ functional quantization rate. A general transfer principle then converts the quantization estimate into a mean estimate and nonasymptotic deviation bounds for equal-weight empirical laws. Applications include empirical path-law estimates for path-dependent SDEs and a path-space propagation-of-chaos estimate for path-dependent McKean–Vlasov interacting particle systems.

MSC2010 subject classification: Primary 60B10; Secondary 49Q22, 60H10, 60K35.
Keywords: Empirical path laws, Wasserstein distance, Itô processes, functional quantization, sharp logarithmic convergence rates, concentration inequalities, path-dependent stochastic differential equations, McKean–Vlasov particle systems, propagation of chaos.

Contents

1 Introduction **2**
1.1 Related work 4

2 Main results **4**
2.1 Notation and conventions 4
2.2 Mean Wasserstein rate 6
2.3 Concentration inequality 7

3 Applications **8**
3.1 Convergence of empirical process law of SDE 8
3.2 Limit theory for path-dependent McKean–Vlasov SDEs 9

*Fengyi Yuan is supported by the Chinese University of Hong Kong (Shenzhen) start-up fund under UDF01004253.
†Department of Mathematics, University of Southern California. xihaohe@usc.edu
‡School of Science and Engineering, the Chinese University of Hong Kong (Shenzhen). yuanfengyi@cuhk.edu.cn

4 Technical results and proofs **10**
4.1 Proofs of the main results 10
4.1.1 Functional quantization with bounded initial value and characteristics . . 10
4.1.2 Localization for integrable characteristics 17
4.1.3 From functional quantization to empirical laws 20
4.2 Proof of Proposition 3.2 24
4.3 Proof of Proposition 3.4 26
4.3.1 Quantization of the nonlinear limit 26
4.3.2 Synchronous-coupling stability 28

1 Introduction

Let X be a continuous $\mathbb{R}^d$-valued Itô process on $[0,T]$, let $\mu := \mathcal{L}(X)$ be its law on $C([0,T];\mathbb{R}^d)$ equipped with $d_\infty(x,y) := \|x-y\|_\infty$, and let

$$\mu_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$$

be the empirical law of N independent copies of X. We study the rate at which μ_N converges to μ in the p-Wasserstein distance associated with d_∞. For Wiener measure, the empirical-measure estimates of Boissard and Le Gouic [3] and the quantization asymptotics of Dereich and Scheutzow [9] identify the scale $(\log N)^{-1/2}$. This raises two questions. Does the same scale persist for Itô processes? Can one also obtain quantitative deviations or concentration for the distance $\mathcal{W}_p(\mu_N,\mu)$?

We give an affirmative answer to both questions under a general assumption. Assumption 2.1 allows a random initial condition and general progressively measurable drift and diffusion integrands. The initial condition and both integrands are controlled by a time-uniform random upper bound Λ_X satisfying $\mathbb{E}\Lambda_X^\rho < \infty$ for some $\rho > p \geq 1$. Under this assumption, Theorem 2.1 establishes

$$\mathbb{E}\big[\mathcal{W}_p(\mu_N,\mu)\big] \leq C(\log N)^{-1/2}, \qquad N \geq 2.$$

The exponent $1/2$ cannot be improved over the class in Assumption 2.1. Indeed, Brownian motion belongs to this class and has a matching quantization lower bound, as recorded in Remark 2.3. We regard this as a worst-case lower bound, since degenerate and deterministic processes may converge faster.

The moment assumption also yields a nonasymptotic deviation estimate. Theorem 2.2 separates an exponential term for the truncated empirical law from a polynomial remainder determined by the available ρ-moment. Under the additional assumption that the path law satisfies a quadratic transport inequality, Proposition 2.5 gives exponential concentration around the same logarithmic mean scale. The transport assumption is specific to this transport-based concentration estimate and is not part of the general moment hypothesis.

We give two applications of the path-law estimate. First, Proposition 3.2 treats path-dependent SDEs with jointly Borel measurable coefficients of linear growth and an initial condition in L^ρ for some $\rho > p \geq 1$, whenever a continuous adapted solution exists. The diffusion coefficient may be unbounded or degenerate. Second, Proposition 3.4 combines the empirical estimate for independent copies of a McKean–Vlasov limit process with synchronous coupling. It gives the same $(\log N)^{-1/2}$ mean rate for path-space propagation of chaos in a path-dependent mean-field particle system. We only require the coefficients to be globally Lipschitz in the stopped path and the path law, and are uniformly bounded at the origin. The initial distribution is assumed to have an r-moment for some $r > p$, as specified in Assumption 3.3.

The proof our main results contains three key components: functional quantization for processes with uniformly bounded initial value and semimartingale characteristics, localization according to dyadic ranges of Λ_X, and a transfer from functional quantization to empirical measures.

The first ingredient, Theorem 4.1, gives a functional quantization result for general Itô processes. Its construction uses a dyadic partition of $[0, T]$ selected according to the realized path oscillations. Both the result and the proof seem to be new to the literature to the best of our knowledge, and are interesting in their own right. Suppose that a continuous semimartingale Y has its initial value, drift, and diffusion integrand bounded by a deterministic constant G. For $0 < \delta \leq G$, the construction recursively bisects a time interval only when the realized path deviates from its value at the left endpoint by more than δ. Conditional exponential martingale estimates show that deep refinements are sufficiently rare. Encoding the admissible tree shapes and quantized endpoint increments then produces a deterministic codebook $\mathcal{A}_\delta$ satisfying

$$\log |\mathcal{A}_\delta| \leq C \frac{G^2}{\delta^2}.$$

We also construct an event $\mathcal{G}_\delta$ by controlling the number of dyadic intervals, such that

$$\mathbb{P}(\mathcal{G}_\delta^c) \leq \exp(-cG^2/\delta^2), \qquad d_\infty(Y, \mathcal{A}_\delta) \leq C\delta \quad \text{on } \mathcal{G}_\delta.$$

A moment estimate controls the exceptional event, and the choice $\delta \asymp G(\log n)^{-1/2}$ gives the quantization rate $(\log n)^{-1/2}$.

The second ingredient removes the deterministic bound G. Proposition 4.4 partitions the probability space according to dyadic ranges of Λ_X. On each resulting event, we radially truncate the initial condition and the drift and diffusion integrands. The resulting process agrees with X on that event and satisfies the hypotheses of Theorem 4.1. A decreasing allocation of codepoints keeps the combined codebook within the prescribed cardinality. The strict gap $q < \rho$ makes the errors over these events summable and controls the remaining contribution from large values of Λ_X, preserving the exponent $1/2$ under our general assumption. For comparison, standard Hölder-ball localization arguments for empirical diffusion-path laws under stronger transport and exponential-integrability assumptions yield sub-critical logarithmic bounds of the form $(\log n)^{-1/2+\varepsilon}$ for every fixed $\varepsilon > 0$; see Dereich [7]. With the path-dependent dyadic partition and localization according to dyadic ranges of Λ_X, we reach the Brownian rate under the present (more general) finite-moment assumption and establish $e_{n,q}(X) \leq C(\log n)^{-1/2}$ for every $1 \leq q < \rho$, where $e_{n,q}$ is the smallest L^q approximation error over deterministic codebooks containing at most n paths.

The third ingredient addresses a structural gap between quantization and empirical sampling. Functional quantization uses a deterministic finite codebook whose cells generally have unequal masses, whereas an empirical measure has random support and assigns mass $1/N$ to every observation. Proposition 4.5 supplies a general transfer principle that requires only a quantization bound of order $(\log n)^{-1/2}$ and a ρ-moment. The argument truncates the law, projects both the truncated law and its empirical measure onto a common finite codebook, and estimates the resulting multinomial fluctuation of the cell masses. A bounded-difference estimate controls deviations of the truncated empirical Wasserstein functional, while the moment assumption controls the polynomial tail remainder. Balancing these terms retains the scale $(\log N)^{-1/2}$, and Corollary 4.6 gives the mean rate.

1.1 Related work

Several general theories relate empirical Wasserstein convergence to the geometry and tails of the underlying measure. Fournier and Guillin [10] establish nonasymptotic moment and concentration bounds in Euclidean spaces under moment assumptions. Boissard [2] obtains transport and entropy bounds for the 1-Wasserstein distance on Polish spaces and gives examples involving Gaussian measures and diffusion laws. Bolley [4] develops path-space concentration inequalities and applies them to empirical path measures of interacting diffusions. Through covering numbers, Boissard and Le Gouic [3] bound the expected p-Wasserstein distance and recover the order $(\log N)^{-1/2}$ for Wiener measure under the supremum norm. Weed and Bach [16] characterize rates on compact metric spaces through Wasserstein dimensions, while Lei [12] obtains rate-optimal bounds for classes of Gaussian measures on unbounded functional spaces. These results explain how infinite-dimensional geometry governs the empirical rate. They do not directly supply the process-specific quantization estimate of order $(\log n)^{-1/2}$ used here.

A complementary line of work studies the process-specific complexity through functional quantization. Dereich, Fehringer, Matoussi, and Scheutzow [8] relate quantization of Gaussian measures on Banach spaces to their small-ball probabilities. Dereich and Scheutzow [9] obtain high-resolution formulas for fractional Brownian motion under the supremum norm, which give the Brownian rate $(\log N)^{-1/2}$ when the Hurst parameter is $1/2$. For diffusion processes, Luschgy and Pagès [13] construct quantizers attaining the Brownian logarithmic rate for classes of one-dimensional diffusions and selected multidimensional models under their stated assumptions. Creutzig, Dereich, Müller-Gronbach, and Ritter [6] obtain the same order for autonomous multidimensional diffusions under smoother assumptions. Dereich [7] further establishes high-resolution asymptotic formulas under supremum-norm distortion and additional regularity assumptions, identifying the martingale component and its quadratic variation as the source of the leading functional complexity. These works obtain the Brownian order for Gaussian models or under additional structural and regularity assumptions. Their conclusions do not directly cover general progressively measurable integrands with a finite moment. There is also a structural difference between the two approximation problems. Functional quantization optimizes over deterministic codebooks with generally unequal cell masses, whereas empirical measures have random support and equal weights. The transfer principle developed here connects these two settings while retaining the order $(\log N)^{-1/2}$ and providing a deviation estimate.

The rest of the paper is organized as follows. Section 2 states the main mean and concentration results. Section 3 gives the applications to path-dependent SDEs and McKean–Vlasov particle systems. Section 4 develops the functional quantization and quantization-to-empirical arguments and proves the main results.

2 Main results

2.1 Notation and conventions

Throughout the paper, $T > 0$ is a fixed time horizon and $d, m \in \mathbb{N}$ are positive integers. For $k \in \mathbb{N}$, the Euclidean norm on $\mathbb{R}^k$ is denoted by $|\cdot|$. For a matrix $A \in \mathbb{R}^{d\times m}$, $\|A\|$ denotes its Hilbert–Schmidt norm.

We write

$$\mathcal{C}^d := C([0,T]; \mathbb{R}^d)$$

for the space of continuous $\mathbb{R}^d$-valued paths on $[0,T]$, equipped with the supremum norm and the associated metric

$$\|x\|_\infty := \sup_{0\le t\le T} |x_t|\,, \qquad d_\infty(x,y) := \|x-y\|_\infty\,, \qquad x,y\in\mathcal{C}^d.$$

The space $\mathcal{C}^d$ is always endowed with the Borel σ-field generated by d_∞. For $x\in\mathcal{C}^d$ and $t\in[0,T]$, the stopped path $x_{t\wedge\cdot}\in\mathcal{C}^d$ is defined by $x_{t\wedge\cdot}(s) := x_{t\wedge s}, 0\le s\le T$. The symbol 0 denotes either the zero vector or the zero path, according to context.

Let (E,d_E) be a Polish metric space. We denote by $\mathcal{P}(E)$ the set of Borel probability measures on E and, for $p\ge 1$, by

$$\mathcal{P}_p(E) := \left\{\mu\in\mathcal{P}(E) : \int_E d_E(x,x_0)^p\,\mu(dx) < \infty\right\},$$

where $x_0\in E$ is arbitrary. This definition does not depend on the choice of x_0. For $\mu,\nu\in\mathcal{P}(E)$, the set of their couplings is denoted by

$$\Pi(\mu,\nu) := \{\pi\in\mathcal{P}(E\times E) : \pi(\,\cdot\times E) = \mu,\ \pi(E\times\,\cdot\,) = \nu\}\,.$$

For $\eta,\zeta\in\mathcal{P}_p(\mathcal{C}^d)$, we use the distinct notation

$$\mathcal{W}_p(\eta,\zeta) := \left(\inf_{\pi\in\Pi(\eta,\zeta)} \int_{\mathcal{C}^d\times\mathcal{C}^d} \|x-y\|_\infty^p\ \pi(dx,dy)\right)^{1/p}$$

for the p-Wasserstein distance on path space.

If $F: E\to E'$ is measurable and $\mu\in\mathcal{P}(E)$, then $F_\#\mu := \mu\circ F^{-1}$ denotes the pushforward of μ under F. For an E-valued random variable Z, its law is denoted by

$$\mathcal{L}(Z) := \mathbb{P}\circ Z^{-1}.$$

The Dirac probability measure at x is denoted by δ_x.

Unless stated otherwise, all random variables and stochastic processes are defined on a filtered probability space

$$(\Omega,\mathcal{F},(\mathcal{F}_t)_{0\le t\le T},\mathbb{P})$$

satisfying the usual conditions and carrying an m-dimensional $(\mathcal{F}_t)$-Brownian motion W. For a continuous local martingale M, its quadratic variation is denoted by $\langle M\rangle$.

The following assumption packages the size of the initial condition and of the two Itô characteristics into a single random upper bound Λ_X. Conditional on an event of the form $\{\Lambda_X\le G\}$, the process lies in the deterministically bounded regime treated by Theorem 4.1. The moment condition on Λ_X is then used to aggregate these bounded regimes and to control the contribution from large values of Λ_X.

Assumption 2.1. *Fix $\rho>1$. Let X_0 be an $\mathcal{F}_0$-measurable $\mathbb{R}^d$-valued random variable, and let*

$$a:[0,T]\times\Omega\longrightarrow\mathbb{R}^d, \qquad H:[0,T]\times\Omega\longrightarrow\mathbb{R}^{d\times m}$$

be progressively measurable processes. We assume that the process X admits the Itô representation

$$X_t = X_0 + \int_0^t a_s\,ds + \int_0^t H_s\,dW_s, \qquad 0\le t\le T,$$

almost surely.

Define

$$\Lambda_X := 1 \vee |X_0| \vee \operatorname*{ess\,sup}_{0\le s\le T} |a_s| \vee \operatorname*{ess\,sup}_{0\le s\le T} \|H_s\|,$$

where the essential suprema are taken with respect to Lebesgue measure on $[0,T]$*. We assume that*

$$\mathbb{E}\Lambda_X^{\rho} < \infty.$$

Remark 2.2. (i) *Assumption 2.1 is a condition on the realized Itô characteristics, not on a particular coefficient representation. In particular,* a *and* H *may be path dependent, random, and merely progressively measurable. No continuity in the state variable, Markov property, ellipticity, or lower bound on* H *is required. The assumption also includes degenerate and deterministic processes. Accordingly, the optimality claim below concerns the worst-case exponent on subclasses with a fixed bound on* $\mathbb{E}\Lambda_X^{\rho}$*, rather than a matching lower bound for every admissible law.*

(ii) *In fact, the Itô process in Assumption 2.1 can be extended to continuous adapted processes* X *with* $\mathcal{L}(X)=\mu$ *such that*

$$X_t = X_0 + \int_0^t a_s \,\mathrm{d}s + M_t, \qquad 0 \le t \le T,$$

where a *is progressively measurable,* M *is a* d*-dimensional continuous local martingale with* $M_0 = 0$*, and*

$$\langle M\rangle_t = \int_0^t c_s \,\mathrm{d}s$$

for a predictable process c_s *with values in the nonnegative symmetric matrices. Correspondingly,*

$$\Lambda_X = 1 \vee |X_0| \vee \operatorname*{ess\,sup}_{0\le s\le T} |a_s| \vee \operatorname*{ess\,sup}_{0\le s\le T} \sqrt{Tr(c_s)}.$$

2.2 Mean Wasserstein rate

The first main theorem identifies the sampling rate in path space. Its significance is that the logarithmic exponent is the Brownian exponent $1/2$ throughout the full class of Assumption 2.1; the possibly unbounded and random size of the characteristics affects the constant but does not produce an ε-loss in the exponent.

Theorem 2.1 (Empirical path-law rate)**.** *Let* $\rho > p \ge 1$*. Under Assumption 2.1, let*

$$\mu := \mathcal{L}(X), \qquad \mu_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i},$$

where $X_1,\dots,X_N$ *are independent copies of* X*. Then there exists* $C = C\big(d,T,p,\rho,\mathbb{E}\Lambda_X^{\rho}\big) > 0$ *such that*

$$\mathbb{E}\big[\mathcal{W}_p(\mu_N,\mu)\big] \le C(\log N)^{-1/2}, \qquad N \ge 2.$$

Proof. See Subsection 4.1. □

Remark 2.3. (i) *The exponent* $1/2$ *cannot be improved over the class in Assumption 2.1. Indeed, one-dimensional Brownian motion belongs to this class. For every probability measure* ν *supported on at most* N *paths,*

$$\mathcal{W}_p(\mathcal{L}(W),\nu) \ge e_{N,p}(W), \tag{1}$$

and the supremum-norm quantization error of Brownian motion satisfies

$$e_{N,p}(W) \asymp (\log N)^{-1/2}$$

by Dereich and Scheutzow [9]. Applying (1) to each realization of the empirical measure gives the matching lower order. A matching lower bound need not hold for every individual diffusion because the assumptions allow degenerate and deterministic models.

(ii) *The condition $\rho > p$ is essential to our proofs. In the quantization argument it ensures summability of the errors over the dyadic ranges of Λ_X. In the empirical transfer argument it makes the truncation bias negligible. The logarithmic exponent $1/2$ comes from Theorem 4.1 and is therefore independent of the size of the moment gap.*

(iii) *The main difficulty lies in Theorem 4.1, since this error is the dominant rate, details are in Remark 4.2. The usual and most common idea in current literatures is to take use of α-Hölder continuity of SDE solutions to make time and space discretization with $\alpha < 1/2$. This will result in the rate of $C_\varepsilon(\log N)^{-1/2+\varepsilon}$, and C_ε will go to $+\infty$ when ε tends to 0, or consider nondegenerate or deterministic volatility so that one can transfer the process to a Brownian motion.*

2.3 Concentration inequality

The mean estimate describes the typical approximation scale but does not separate moderate fluctuations from rare observations with large supremum norm. The next theorem makes this separation explicit. The exponential term is produced by bounded differences after truncation, whereas the polynomial term records the cost of controlling the unbounded tails using only a finite ρ-moment.

Theorem 2.2. *Let $\rho > p \geq 1$, and let μ and μ_N be as in Theorem 2.1. Set*

$$\beta_{\rho,p} := \min\left\{\frac{\rho}{p} - 1, \frac{\rho}{2p}\right\}.$$

Under Assumption 2.1, there exist constants $c_p > 0$ and $C = C\big(d, T, p, \rho, \mathbb{E}\Lambda_X^\rho\big) > 0$ such that, for every $N \geq 2$ and $t > 0$,

$$\mathbb{P}\left(\mathcal{W}_p(\mu_N, \mu) > C(\log N)^{-1/2} + t\right) \leq \exp\left(-\frac{c_p N t^{2p}}{(\log N)^{p^2/(\rho-p)}}\right) + CN^{-\beta_{\rho,p}} t^{-\rho}. \tag{2}$$

After increasing C if necessary, one also has

$$\mathbb{P}\left(\mathcal{W}_p(\mu_N, \mu) > C(\log N)^{-1/2}\right) \leq CN^{-\beta_{\rho,p}} (\log N)^{\rho/2}, \qquad N \geq 2.$$

Proof. See Subsection 4.1. □

Remark 2.4. *The two terms in (2) have different origins and should not be conflated. The exponential term controls fluctuations of the truncated empirical measure on a bounded path-space ball. The polynomial term controls the empirical truncation error and is dictated by the available moment.*

For an exponential deviation around the mean, we record a standard transport-inequality formulation. It is the path-space counterpart of the approach in Boissard [2] and Bolley [4].

Proposition 2.5. *Let $\rho > p \geq 1$. Under Assumption 2.1, let μ and μ_N be as in Theorem 2.1. Suppose in addition that, for some $C_T > 0$, μ satisfies the quadratic transport inequality*

$$\mathcal{W}_2^2(\nu, \mu) \leq 2C_T \,\mathrm{H}(\nu \mid \mu)$$

for every probability measure ν on $\mathcal{C}^d$, where H denotes relative entropy. Then there exists $C = C\big(d, T, p, \rho, \mathbb{E}\Lambda_X^\rho\big) > 0$ such that, for every $t > 0$ and $N \geq 2$,

$$\mathbb{P}\left(\mathcal{W}_p(\mu_N, \mu) > C(\log N)^{-1/2} + t\right) \leq \exp\left(-\frac{N^{2/\max\{p,2\}}t^2}{2C_T}\right). \tag{3}$$

In particular, the exponent in (3) is $-Nt^2/(2C_T)$ when $1 \leq p \leq 2$.

Proof. See Subsection 4.1. $\square$

3 Applications

We now illustrate how the abstract results are used. The first application is a direct verification theorem for a single path-dependent SDE; the second is a stability argument for an interacting particle system, in which the i.i.d. empirical approximation of the nonlinear limit is propagated through the mean-field dynamics.

3.1 Convergence of empirical process law of SDE

The coefficient functions below may depend on the complete stopped path and are assumed only to be Borel measurable. Linear growth is sufficient because it yields a supremum-moment estimate for the solution and hence an integrable random bound on the realized drift and diffusion characteristics.

Assumption 3.1. *Fix $\rho > 1$. The maps*

$$b : [0, T] \times \mathcal{C}^d \longrightarrow \mathbb{R}^d, \qquad \sigma : [0, T] \times \mathcal{C}^d \longrightarrow \mathbb{R}^{d\times m}$$

are Borel measurable. There exists a constant $L \geq 1$ such that

$$|b(t, x)| + \|\sigma(t, x)\| \leq L(1 + \|x\|_\infty), \qquad (t, x) \in [0, T] \times \mathcal{C}^d.$$

The initial condition X_0 is $\mathcal{F}_0$-measurable and satisfies

$$\mathbb{E}|X_0|^\rho < \infty.$$

We assume that there exists an $\mathbb{R}^d$-valued continuous $(\mathcal{F}_t)$-adapted process X satisfying

$$X_t = X_0 + \int_0^t b(s, X_{s\wedge\cdot})\, ds + \int_0^t \sigma(s, X_{s\wedge\cdot})\, dW_s, \qquad 0 \leq t \leq T,$$

almost surely.

The next proposition shows that the abstract endpoint theorem is stable under this standard linear-growth formulation. No continuity or Lipschitz condition is needed for the convergence estimate itself once existence of a continuous adapted solution has been assumed.

Proposition 3.2. *Let $\rho > p \geq 1$. Under Assumption 3.1, let $X_1, \ldots, X_N$ be independent copies of X. Then there exists $C = C\big(d, T, p, \rho, L, \mathbb{E}\,|X_0|^\rho\big) > 0$ such that*

$$\mathbb{E}\left[\mathcal{W}_p\left(\frac{1}{N}\sum_{i=1}^N \delta_{X_i}, \mathcal{L}(X)\right)\right] \leq C(\log N)^{-1/2}, \qquad N \geq 2.$$

Proof. See Subsection 4.2. $\square$

3.2 Limit theory for path-dependent McKean–Vlasov SDEs

In the second application, we consider a path-dependent McKean–Vlasov SDE and its associated interacting particle system. In order to establish finite-particle convergence of this system, our path law convergence result in Section 2.2 is essential, as the marginal empirical law estimate on $\mathbb{R}^d$ from Fournier and Guillin [10] is not sufficient. The proof compares the interacting system with independent copies of the nonlinear limit. The empirical error of those copies is controlled by the preceding quantization theory, while the difference between the two systems is controlled by synchronous coupling and Grönwall stability.

Fix a time horizon $T > 0$, integers $d, m \geq 1$, and an exponent $p \geq 1$. Let

$$b : [0,T] \times \mathcal{C}^d \times \mathcal{P}_p(\mathcal{C}^d) \longrightarrow \mathbb{R}^d, \qquad \sigma : [0,T] \times \mathcal{C}^d \times \mathcal{P}_p(\mathcal{C}^d) \longrightarrow \mathbb{R}^{d\times m}$$

be jointly Borel measurable.

Assumption 3.3. *There are constants $L > 0$ and $r > p$ such that, for every $t \in [0,T]$, $x, y \in \mathcal{C}^d$, and $\nu, \eta \in \mathcal{P}_p(\mathcal{C}^d)$,*

$$|b(t,x,\nu) - b(t,y,\eta)| + \|\sigma(t,x,\nu) - \sigma(t,y,\eta)\| \leq L\big(\|x-y\|_\infty + \mathcal{W}_p(\nu,\eta)\big),$$

and

$$|b(t,\mathbf{0},\delta_{\mathbf{0}})| + \|\sigma(t,\mathbf{0},\delta_{\mathbf{0}})\| \leq L.$$

The initial random variable ξ is independent of the driving Brownian motion and satisfies

$$\mathbb{E}\,|\xi|^r < \infty.$$

Let Y be the unique strong solution of the path-dependent McKean–Vlasov stochastic differential equation

$$Y_t = \xi + \int_0^t b(s, Y_{s\wedge\cdot}, \mu_s)\,\mathrm{d}s + \int_0^t \sigma(s, Y_{s\wedge\cdot}, \mu_s)\,\mathrm{d}W_s, \qquad \mu_s := \mathcal{L}(Y_{s\wedge\cdot}), \qquad \mu := \mu_T.$$

For each $N \geq 2$, let $(\xi^i, W^i)_{1\leq i\leq N}$ be independent copies of (ξ, W), and let the interacting particles solve

$$X_t^{i,N} = \xi^i + \int_0^t b(s, X_{s\wedge\cdot}^{i,N}, \mu_s^N)\,\mathrm{d}s + \int_0^t \sigma(s, X_{s\wedge\cdot}^{i,N}, \mu_s^N)\,\mathrm{d}W_s^i,$$

where

$$\mu_s^N := \frac{1}{N}\sum_{j=1}^N \delta_{X_{s\wedge\cdot}^{j,N}} \in \mathcal{P}_p(\mathcal{C}^d), \qquad \mu^N := \mu_T^N.$$

The Lipschitz and origin bounds imply the linear-growth estimate

$$|b(t,x,\nu)| + \|\sigma(t,x,\nu)\| \leq L\big(1 + \|x\|_\infty + \mathcal{W}_p(\nu,\delta_{\mathbf{0}})\big).$$

Under these assumptions, the path-dependent McKean–Vlasov equation has a unique strong solution by the standard fixed-point argument on stopped-path law flows. The particle system also has a unique strong solution. Indeed, for two path configurations $(x^1,\dots,x^N)$ and $(y^1,\dots,y^N)$ in $(\mathcal{C}^d)^N$, matching paths with the same index gives, for every $t \in [0,T]$,

$$\mathcal{W}_p\left(\frac{1}{N}\sum_{j=1}^N \delta_{x_{t\wedge\cdot}^j}, \frac{1}{N}\sum_{j=1}^N \delta_{y_{t\wedge\cdot}^j}\right) \leq \left(\frac{1}{N}\sum_{j=1}^N \left\|x_{t\wedge\cdot}^j - y_{t\wedge\cdot}^j\right\|_\infty^p\right)^{1/p}.$$

Consequently, the particle coefficients are globally Lipschitz in the stopped path configuration, and the usual Picard iteration for functional SDEs applies.

Proposition 3.4. *Under Assumption 3.3, there is a constant*

$$C = C\big(p, r, d, m, T, L, \mathbb{E}\,|\xi|^r\big)$$

such that, for every $N \geq 2$,

$$\mathbb{E}\mathcal{W}_p(\mu^N, \mu) \leq C(\log N)^{-1/2}.$$

Remark 3.5. *The result can be extended to graphon mean field system with common noise in Bayraktar, He and Kim[1], which obtains convergence of process law without rate under similar general assumptions. The main difficulty of obtaining convergence rate results in that paper is the lack of techniques and results in this work.*

Proof. See Subsection 4.3. □

4 Technical results and proofs

4.1 Proofs of the main results

For a $\mathcal{C}^d$-valued random variable Z, $q \geq 1$, and $n \in \mathbb{N}$, define the n-point functional quantization error by

$$e_{n,q}(Z) := \inf_{\substack{\mathcal{A}\subset\mathcal{C}^d \\ 1\leq|\mathcal{A}|\leq n}} \big(\mathbb{E}\big[d_\infty(Z,\mathcal{A})^q\big]\big)^{1/q}, \qquad d_\infty(z,\mathcal{A}) := \min_{a\in\mathcal{A}} \|z-a\|_\infty.$$

The argument in this subsection is organized in three layers. First, we use a pathwise dyadic partition method to establish a critical functional quantization result for semimartingales whose initial value and characteristics are bounded. Second, localization according to dyadic levels of Λ_X removes this deterministic bound under Assumption 2.1. Third, the resulting pathwise quantization estimate is transferred to equal-weight empirical measures by truncation and finite-codebook sampling.

4.1.1 Functional quantization with bounded initial value and characteristics

The local martingale estimate below controls the conditional probability that a dyadic interval is split.

Lemma 4.1. *Let N be a real-valued continuous local martingale. Fix deterministic times $0 \leq u < v \leq T$, and suppose that, for some deterministic $V > 0$,*

$$\langle N\rangle_v - \langle N\rangle_u \leq V \qquad \textit{almost surely.}$$

Then, for every $x > 0$,

$$\mathbb{P}\left(\sup_{u\leq t\leq v} |N_t - N_u| \geq x \,\middle|\, \mathcal{F}_u\right) \leq 2\exp\left(-\frac{x^2}{2V}\right) \qquad \textit{almost surely.}$$

Proof. For $\lambda > 0$, define, for $u \leq t \leq v$,

$$Z_t^\lambda := \exp\left(\lambda(N_t - N_u) - \frac{\lambda^2}{2}(\langle N\rangle_t - \langle N\rangle_u)\right).$$

This is a positive local martingale starting from one at time u, hence it is a supermartingale. Let

$$\tau_+ := \inf\{t \in [u,v] : N_t - N_u \geq x\} \wedge v.$$

Continuity of N and the bracket bound imply that, on $\{\sup_{u\le t\le v}(N_t - N_u) \ge x\}$,

$$Z^\lambda_{\tau_+} \ge \exp\left(\lambda x - \frac{\lambda^2 V}{2}\right).$$

Conditional optional sampling therefore gives

$$1 \ge \mathbb{E}\left[Z^\lambda_{\tau_+} \mid \mathcal{F}_u\right] \ge \exp\left(\lambda x - \frac{\lambda^2 V}{2}\right)\mathbb{P}\left(\left\{\sup_{u\le t\le v}(N_t - N_u) \ge x\right\}\middle|\, \mathcal{F}_u\right).$$

Taking $\lambda = x/V$ yields

$$\mathbb{P}\left(\left\{\sup_{u\le t\le v}(N_t - N_u) \ge x\right\}\middle|\, \mathcal{F}_u\right) \le \exp\left(-\frac{x^2}{2V}\right).$$

Applying the same argument to $-N$ and taking a union bound proves the claim. □

Theorem 4.1. *Let $q \ge 1$, and let Y be an $\mathbb{R}^d$-valued continuous semimartingale of the form*

$$Y_t = Y_0 + \int_0^t a_s\,ds + \int_0^t H_s\,dW_s, \qquad 0 \le t \le T,$$

where a and H are progressively measurable and, for some $G \ge 1$,

$$|Y_0| \le G, \qquad |a_s| \le G, \qquad \|H_s\| \le G$$

almost surely, with the last two bounds holding for almost every s. There exist positive constants $c_{d,T}$ and $C_{d,T}$, depending only on d and T, such that for every $0 < \delta \le G$ there are a deterministic finite set $\mathcal{A}_\delta \subset \mathcal{C}^d$, containing the zero path, and an event $\mathcal{G}_\delta$ satisfying

$$\log|\mathcal{A}_\delta| \le C_{d,T}\frac{G^2}{\delta^2}, \qquad \mathbb{P}(\mathcal{G}_\delta^c) \le \exp\left(-c_{d,T}\frac{G^2}{\delta^2}\right),$$

and

$$d_\infty(Y, \mathcal{A}_\delta) \le C_{d,T}\delta \qquad \text{on } \mathcal{G}_\delta.$$

Consequently, there exists $C_{d,q,T} > 0$, depending only on d, q, T, such that

$$e_{n,q}(Y) \le C_{d,q,T}\,G(\log n)^{-1/2}, \qquad n \ge 2.$$

Remark 4.2. *The principal probabilistic difficulty is that the partition is random and path-dependent. Moreover, the split indicators are not independent, since the volatility H is allowed to be an arbitrary bounded progressively measurable process, and splits must be controlled simultaneously over infinitely many dyadic levels. The proof resolves the dependence within each level by ordering the intervals chronologically and iterating conditional exponential-moment bounds. Dependence between different levels is then handled by a generalized Hölder argument with suitably chosen exponents. This produces an exponential moment bound for the total number of possible splits and, consequently, an exponential tail for the size of the realized forest.*

Remark 4.3. *The exponent $1/2$ is consistent with the classical functional quantization theory for Wiener measure and Brownian diffusions. The connection between Gaussian quantization and small-ball probabilities is developed in [8], while precise high-resolution quantization and entropy-coding results for fractional Brownian motion are obtained in [9]. Constructive quantization* *schemes for Brownian diffusions are studied in [13], and process-specific asymptotic coding results under the supremum norm are proved in [7]; see also [6] for the closely related problems of infinite-dimensional quadrature and approximation of distributions. Compared with these results, the present theorem is more robust with respect to the underlying dynamics, i.e. it is d-dimensional, nonasymptotic, and allows random, path-dependent, progressively measurable coefficients, without Gaussianity, Markovianity, ellipticity, or spatial regularity assumptions.*

Proof strategy The main idea is an adaptive dyadic discretization of time. One starts with $n_0 \asymp_{d,T} G^2/\delta^2$ root intervals and recursively bisects an interval $I = [u, v]$ whenever

$$\sup_{t \in I} |Y_t - Y_u| > \delta.$$

The initial mesh is chosen so that the drift contributes at most a fixed fraction of δ on every possible dyadic interval. Hence a split can occur only if the martingale part makes an excursion of order δ. If an interval at depth ℓ has length h_ℓ, boundedness of H gives quadratic variation of order at most $G^2 h_\ell$, and a conditional martingale maximal inequality yields

$$\mathbb{P}(I \text{ is split} \mid \mathcal{F}_u) \leq \exp\left(-c_d \frac{\delta^2}{G^2 h_\ell}\right) \leq \exp(-c_d 2^\ell).$$

This is the probabilistic reason why the construction works. The number of potential intervals at depth ℓ grows like $n_0 2^\ell$, whereas the probability of a δ-oscillation decays like $\exp(-c_d 2^\ell)$. The latter decay dominates the combinatorial growth, so only $O_{d,T}(G^2/\delta^2)$ intervals are needed with exponentially high probability. This is also the origin of the Brownian scaling δ^{-2}, and hence of the exponent $1/2$ in the final logarithmic quantization rate.

Proof. Write

$$M_t := \int_0^t H_s \, dW_s$$

and fix $0 < \delta \leq G$. Except for the d-dependent constants defined explicitly below, the subscripts on constants denoted by C or c list all parameters on which they may depend. Their values may change from line to line, and they are independent of G and δ.

Step 0. Binary-tree convention. Set $\kappa_d := 4 + \frac{32d}{9} \log(2d)$, and choose

$$n_0 := \left\lceil \kappa_d (1+T) \frac{G^2}{\delta^2} \right\rceil.$$

Partition $[0, T]$ into n_0 intervals of equal length, ordered from left to right and viewed as the roots of dyadic binary trees. If $I = [u, v]$ is a node, its children are

$$I^0 := \left[u, \frac{u+v}{2}\right], \qquad I^1 := \left[\frac{u+v}{2}, v\right].$$

Let $\mathcal{D}_\ell$ be the collection of all potential nodes at depth ℓ below the roots. Thus every $I \in \mathcal{D}_\ell$ has length h_ℓ, where

$$|\mathcal{D}_\ell| = n_0 2^\ell, \qquad h_\ell := \frac{T}{n_0 2^\ell}.$$

A finite dyadic forest F is an ordered collection of n_0 finite full binary subtrees rooted at the initial intervals. Write $s(\mathsf{F})$ and $L(\mathsf{F})$ for its numbers of internal nodes(all the intervals that are split) and leaves(all the intervals that are not split). Then $L(\mathsf{F}) = n_0 + s(\mathsf{F})$, and F has $n_0 + 2s(\mathsf{F})$

nodes in total. Whenever F is fixed below, abbreviate $L = L(\mathsf{F})$ and enumerate its leaves from left to right as

$$[t_0, t_1], \dots, [t_{L-1}, t_L], \qquad 0 = t_0 < \cdots < t_L = T. \tag{4}$$

These endpoints are deterministic once F is fixed. With the cutoff s_δ to be specified in Step 1, let $\mathfrak{F}_\delta := \{\mathsf{F} : s(\mathsf{F}) \le s_\delta\}$ denote the set of all forests with at most s_δ internal nodes. Figure 1 below illustrates these conventions in the case $n_0 = 2$.

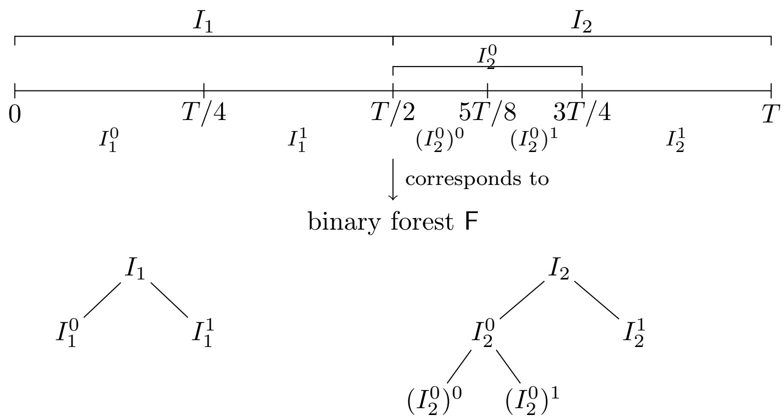

Figure 1: An $n_0 = 2$ example of the correspondence between a dyadic interval partition and an ordered binary forest.

Step 1. Count the internal nodes. Given a path of Y, we define a forest F as follows. Starting from the root intervals, recursively split a dyadic interval $I = [u, v]$ whenever

$$\sup_{t \in I} |Y_t - Y_u| > \delta. \tag{5}$$

The interval is retained as a leaf when the reverse inequality holds. After fixing a continuous version of Y, uniform continuity on $[0, T]$ shows that the resulting random forest F has finitely many nodes almost surely. Our choice of n_0, with $\kappa_d \ge 4$, gives

$$G h_\ell \le \frac{\delta}{4}.$$

Noting that

$$|Y_t - Y_s| \le G|t - s| + |M_t - M_s|, \qquad 0 \le s \le t \le T,$$

for any $I \in \mathcal{D}_\ell$ we have

$$\left\{ \sup_{t \in I} |Y_t - Y_u| > \delta \right\} \subset \left\{ \sup_{t \in I} |M_t - M_u| > \frac{3\delta}{4} \right\}. \tag{6}$$

For every coordinate i of M and every $t \in I$,

$$\langle M^i \rangle_t - \langle M^i \rangle_u \le G^2 (t - u) \le G^2 h_\ell.$$

The definition of n_0 also gives

$$\frac{\delta^2}{G^2 h_\ell} = \frac{\delta^2 n_0 2^\ell}{G^2 T} \ge \kappa_d \frac{1 + T}{T} 2^\ell \ge \kappa_d 2^\ell.$$

Applying Lemma 4.1 with $V = G^2 h_\ell$ yields

$$\mathbb{P}\left(\sup_{t\in I} |M_t^i - M_u^i| > x \,\middle|\, \mathcal{F}_u\right) \le 2\exp\left(-\frac{x^2}{2G^2 h_\ell}\right)$$

for every coordinate i. Set $c_d := 9/(8d)$. Hence, an application of (6), a union bound over the d coordinates, and the choice $x = 3\delta/(4\sqrt{d})$ give, for any $I \in \mathcal{D}_\ell$,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in I} |Y_t - Y_u| > \delta \,\middle|\, \mathcal{F}_u\right) &\le 2d\exp\left(-\frac{9\delta^2}{32dG^2 h_\ell}\right)\\ &\le 2d\exp\left(-\frac{9\kappa_d}{32d}2^\ell\right) \le \exp(-c_d 2^\ell).\end{aligned} \tag{7}$$

where the last inequality follows from the definition of κ_d and $2^\ell \ge 1$. Let N_ℓ count the potential intervals at depth ℓ that satisfy the splitting criterion, namely,

$$N_\ell := \sum_{I=[u,v]\in\mathcal{D}_\ell} \mathbf{1}_{\{\sup_{t\in I}|Y_t - Y_u|>\delta\}}.$$

Order the intervals in $\mathcal{D}_\ell$ chronologically. The event associated with each preceding interval is measurable at the left endpoint of the next interval. Iterating (7) therefore gives, for every $\lambda > 0$,

$$\mathbb{E}\exp(\lambda N_\ell) \le \Big[1 + \exp(-c_d 2^\ell)(e^\lambda - 1)\Big]^{n_0 2^\ell}.$$

Choose numbers

$$r_\ell := \frac{\pi^2}{6}(\ell+1)^2, \qquad \sum_{\ell=0}^{\infty}\frac{1}{r_\ell} = 1, \qquad K_d := \frac{4}{c_d e}\frac{\exp\left(\frac{8\pi^4}{9c_d} - \frac{c_d}{4}\right)}{1 - \exp(-c_d/4)}.$$

Generalized Hölder's inequality gives, for a finite J,

$$\log\mathbb{E}\exp\left(\sum_{\ell=0}^{J} N_\ell\right) \le \sum_{\ell=0}^{J}\frac{1}{r_\ell}\log\mathbb{E}\exp(r_\ell N_\ell) \le n_0\sum_{\ell=0}^{J}\frac{2^\ell}{r_\ell}\exp(-c_d 2^\ell)\big(e^{r_\ell} - 1\big) \le K_d n_0. \tag{8}$$

In this finite-level application, the reciprocal exponents have sum smaller than one. We add the constant random variable 1 with reciprocal exponent

$$1 - \sum_{\ell=0}^{J}\frac{1}{r_\ell}$$

to complete the family of Hölder exponents. To verify the bound by $K_d n_0$ in (8), note that $r_\ell \ge 1$ and $(\ell+1)^2 \le 8\,2^{\ell/2}$. Completing the square gives

$$\sum_{\ell=0}^{\infty}\frac{2^\ell}{r_\ell}\exp(-c_d 2^\ell)\big(e^{r_\ell} - 1\big) \le \exp\left(\frac{8\pi^4}{9c_d}\right)\sum_{\ell=0}^{\infty} 2^\ell\exp\left(-\frac{c_d}{2}2^\ell\right) \le K_d.$$

Indeed, the second inequality follows from $xe^{-c_d x/2} \le 4(c_d e)^{-1}e^{-c_d x/4}$ for $x > 0$ and $2^\ell \ge \ell + 1$. In particular, K_d depends only on d. Monotone convergence now gives

$$\mathbb{E}\exp\left(\sum_{\ell=0}^{\infty} N_\ell\right) \le e^{K_d n_0}. \tag{9}$$

Every internal node of F at depth ℓ is counted by N_ℓ, and hence

$$s(\mathsf{F}) \le \sum_{\ell=0}^{\infty} N_\ell. \tag{10}$$

Set $s_\delta := \lfloor (K_d+1)n_0 \rfloor$. Since $s(\mathsf{F})$ is integer-valued, the exponential Markov inequality, in conjunction with (9) and (10), gives

$$\mathbb{P}\big(s(\mathsf{F}) > s_\delta\big) = \mathbb{P}\big(s(\mathsf{F}) > (K_d+1)n_0\big) \le e^{-(K_d+1)n_0}\mathbb{E}e^{s(\mathsf{F})} \le e^{-n_0} \le \exp\left(-\frac{G^2}{\delta^2}\right). \tag{11}$$

Defining $\mathcal{G}_\delta := \{s(\mathsf{F}) \le s_\delta\}$, we have $\mathsf{F} \in \mathfrak{F}_\delta$ on $\mathcal{G}_\delta$.

Step 2. Construct the codebook. Define the coordinatewise lattice quantizer $Q_\delta : \mathbb{R}^d \to \delta\mathbb{Z}^d$ by

$$[Q_\delta(x)]^i := \delta \left\lceil \frac{x^i}{\delta} - \frac{1}{2} \right\rceil, \qquad 1 \le i \le d.$$

Thus ties are resolved toward the smaller lattice point. This quantizer satisfies

$$|Q_\delta(x) - x| \le \frac{\sqrt{d}}{2}\delta, \qquad |x^i - y^i| \le \delta \quad \Longrightarrow \quad [Q_\delta(x)]^i - [Q_\delta(y)]^i \in \delta\{-1,0,1\} \tag{12}$$

for every $1 \le i \le d$. Let

$$\Lambda_{\delta,G} := \delta\mathbb{Z}^d \cap \left[-G - \frac{\delta}{2}, G + \frac{\delta}{2}\right]^d.$$

For $\mathsf{F} \in \mathfrak{F}_\delta$, define the endpoint arrays by

$$\mathcal{Z}_{\mathsf{F}} := \left\{(z_0,\dots,z_L) \in (\delta\mathbb{Z}^d)^{L+1} : z_0 \in \Lambda_{\delta,G},\ z_j - z_{j-1} \in \delta\{-1,0,1\}^d \text{ for } 1 \le j \le L\right\}.$$

For $\boldsymbol{z} = (z_0,\dots,z_L) \in \mathcal{Z}_{\mathsf{F}}$, let $\mathsf{I}_{\mathsf{F}}[\boldsymbol{z}] \in \mathcal{C}^d$ be the path obtained by linear interpolation at the leaf endpoints. More precisely, for $t \in [t_{j-1}, t_j]$, set

$$\mathsf{I}_{\mathsf{F}}[\boldsymbol{z}](t) := \frac{t_j - t}{t_j - t_{j-1}} z_{j-1} + \frac{t - t_{j-1}}{t_j - t_{j-1}} z_j.$$

Define

$$\mathcal{A}_{\mathsf{F}} := \{\mathsf{I}_{\mathsf{F}}[\boldsymbol{z}] : \boldsymbol{z} \in \mathcal{Z}_{\mathsf{F}}\}, \qquad \mathcal{A}_\delta := \bigcup_{\mathsf{F} \in \mathfrak{F}_\delta} \mathcal{A}_{\mathsf{F}}.$$

In particular, the zero path belongs to every $\mathcal{A}_{\mathsf{F}}$ and hence to $\mathcal{A}_\delta$. The elementary count for ordered full binary forests gives

$$|\mathfrak{F}_\delta| \le \sum_{s=0}^{s_\delta} 2^{n_0+2s} \le 2^{n_0+2s_\delta+1}.$$

Since $G/\delta \ge 1$, we also have

$$|\Lambda_{\delta,G}| \le \left(4\frac{G}{\delta}\right)^d.$$

For every $\mathsf{F} \in \mathfrak{F}_\delta$, the identity $L(\mathsf{F}) = n_0 + s(\mathsf{F})$ gives $L(\mathsf{F}) \le n_0 + s_\delta$. Counting the initial lattice point and the successive leaf increments therefore yields

$$
\begin{aligned}
|\mathcal{A}_\delta| &\le |\mathfrak{F}_\delta| \left(4\frac{G}{\delta}\right)^d 3^{d(n_0+s_\delta)} \le 2^{n_0+2s_\delta+1} \left(4\frac{G}{\delta}\right)^d 3^{d(n_0+s_\delta)}, \\
\log|\mathcal{A}_\delta| &\le (n_0+2s_\delta+1)\log 2 + d\log\left(4\frac{G}{\delta}\right) + d(n_0+s_\delta)\log 3 \le C_d n_0 \le C_{d,T}\frac{G^2}{\delta^2}.
\end{aligned} \tag{13}
$$

Here we use $s_\delta \le (K_d+1)n_0$, $n_0 \ge (G/\delta)^2$, and the last inequality follows from the definition of n_0.

Step 3. Pathwise reconstruction. Fix $\omega \in \mathcal{G}_\delta$ and set $y := Y(\omega)$. Denote the realized forest simply by $\mathsf{F} \in \mathfrak{F}_\delta$. With the leaf partition in (4), set

$$
z_j := Q_\delta(y_{t_j}), \qquad 0 \le j \le L.
$$

For every leaf $[t_{j-1}, t_j]$, the reverse inequality in (5) gives

$$
\sup_{t\in[t_{j-1},t_j]} |y_t - y_{t_{j-1}}| \le \delta. \tag{14}
$$

The bound $|y_0| \le G$ and the adjacency property (12) therefore imply

$$
z_0 \in \Lambda_{\delta,G}, \qquad z_j - z_{j-1} \in \delta\{-1,0,1\}^d, \quad 1 \le j \le L.
$$

Thus $(z_0,\dots,z_L)$ belongs to $\mathcal{Z}_\mathsf{F}$, and the corresponding codeword $\widehat{y} := \mathsf{I}_\mathsf{F}[(z_0,\dots,z_L)]$ belongs to $\mathcal{A}_\mathsf{F} \subset \mathcal{A}_\delta$. For $t \in [t_{j-1}, t_j]$, write

$$
\theta := \frac{t - t_{j-1}}{t_j - t_{j-1}}.
$$

The leaf bound (14) and the quantization-error bound in (12) give

$$
|y_t - \widehat{y}_t| \le \left|y_t - (1-\theta)y_{t_{j-1}} - \theta y_{t_j}\right| + (1-\theta)|y_{t_{j-1}} - z_{j-1}| + \theta|y_{t_j} - z_j| \le \left(2 + \frac{\sqrt{d}}{2}\right)\delta.
$$

Consequently,

$$
d_\infty(Y(\omega), \mathcal{A}_\delta) \le \|y - \widehat{y}\|_\infty \le \left(2 + \frac{\sqrt{d}}{2}\right)\delta \qquad \text{for } \omega \in \mathcal{G}_\delta.
$$

Step 4. The L^q distortion. The Burkholder–Davis–Gundy inequality implies

$$
\left(\mathbb{E}\|Y\|_\infty^{2q}\right)^{1/(2q)} \le G + GT + C_q G\sqrt{T} \le C_{q,T} G. \tag{15}
$$

Since the zero path belongs to $\mathcal{A}_\delta$, Hölder's inequality and (11) yield

$$
\begin{aligned}
(\mathbb{E} d_\infty(Y, \mathcal{A}_\delta)^q)^{1/q} &\le C_d\delta + \left(\mathbb{E}\|Y\|_\infty^{2q}\right)^{1/(2q)} \mathbb{P}(\mathcal{G}_\delta^c)^{1/(2q)} \\
&\le C_d\delta + C_{q,T} G \exp\left(-\frac{G^2}{2q\delta^2}\right) \le C_{d,q,T}\delta.
\end{aligned} \tag{16}
$$

The final inequality follows from $G/\delta \ge 1$. For n larger than a threshold depending only on d and T, take

$$
\delta = C_{d,T} G (\log n)^{-1/2}
$$

where $C_{d,T}$ is from (13) and n is sufficiently large such that $\delta \le G$. Substituting this choice of δ into (16) gives

$$
e_{n,q}(Y) \le C_{d,q,T} G (\log n)^{-1/2}
$$

for all such n. The finitely many remaining values of n are absorbed into $C_{d,q,T}$ using (15). □

4.1.2 Localization for integrable characteristics

Theorem 4.1 cannot be applied directly under Assumption 2.1, because the initial value and integrands are bounded only by the integrable random variable Λ_X. The next proposition localizes according to dyadic levels of Λ_X and allocates codebooks of decreasing cardinality to increasingly rare events. The moment condition makes these errors summable without changing the rate.

Proposition 4.4. *Under Assumption 2.1, for every q satisfying $1 \le q < \rho$, there exists a constant $C = C\big(d, T, q, \rho, \mathbb{E}\Lambda_X^\rho\big) > 0$ such that*

$$e_{n,q}(X) \le C(\log n)^{-1/2}, \qquad n \ge 2.$$

Proof. Fix $q \in [1, \rho)$. Choose $\theta > 1$ sufficiently large that, with $\beta := 1 - \frac{1}{\theta}$, one has $\rho\beta > q$. Such a choice is possible because $q < \rho$. For instance, it is enough to take $\theta > \frac{\rho}{\rho - q}$.

Step 1. Supremum moment estimate. We first record the moment estimate that will also be used to control the contribution from large values of Λ_X. The Itô representation of X gives

$$\|X\|_\infty \le |X_0| + \int_0^T |a_s|\, ds + \sup_{0\le t\le T} \left| \int_0^t H_s\, dW_s \right|.$$

Since one has that

$$|X_0| \le \Lambda_X \qquad \text{and} \qquad \int_0^T |a_s|\, ds \le T\Lambda_X,$$

the Burkholder–Davis–Gundy inequality yields

$$\mathbb{E}\|X\|_\infty^\rho \le C_\rho \mathbb{E}|X_0|^\rho + C_\rho \mathbb{E}\left(\int_0^T |a_s|\, ds\right)^\rho + C_\rho \mathbb{E}\left(\int_0^T \|H_s\|^2\, ds\right)^{\rho/2} \le C_{\rho,T}\mathbb{E}\Lambda_X^\rho < \infty. \quad (17)$$

Step 2. Truncation. It remains to prove the quantization estimate. It is enough initially to consider sufficiently large n. Set

$$R_j := 2^{j+1}, \quad j \ge 0,$$

and decompose the probability space into the disjoint events

$$B_0 := \{\Lambda_X \le 2\}, \qquad B_j := \{2^j < \Lambda_X \le 2^{j+1}\}, \quad j \ge 1.$$

Since Λ_X is finite almost surely, these events form a partition of Ω up to a null set. For $R > 0$, let π_R denote radial projection onto the closed ball of radius R. We use the same notation for vectors and matrices, with the Euclidean norm for vectors and the Hilbert–Schmidt norm for matrices. Thus,

$$\pi_R(z) = \begin{cases} z, & \text{if } |z| \le R, \\ Rz/|z|, & \text{if } |z| > R, \end{cases}$$

with the corresponding definition for matrices. For every $j \ge 0$, define the continuous Itô process

$$Y_t^{(j)} = \pi_{R_j}(X_0) + \int_0^t \pi_{R_j}(a_s)\, ds + \int_0^t \pi_{R_j}(H_s)\, dW_s, \qquad 0 \le t \le T.$$

The initial condition and the two Itô characteristics of $Y^{(j)}$ satisfy

$$|Y_0^{(j)}| \le R_j, \qquad |\pi_{R_j}(a_s)| \le R_j, \qquad \|\pi_{R_j}(H_s)\| \le R_j$$

almost surely, with the last two inequalities holding for almost every $s \in [0,T]$. Consequently, $Y^{(j)}$ satisfies the assumptions of Theorem 4.1 with $G = R_j$. We next verify that $Y^{(j)}$ agrees with X on the event B_j. On B_j, we have

$$|X_0| \le R_j, \qquad |a_s| \le R_j, \qquad \|H_s\| \le R_j$$

for almost every $s \in [0,T]$. The finite-variation terms in the representations of X and $Y^{(j)}$ are thus equal on B_j. Moreover, the quadratic variation of the difference of their martingale terms is zero on B_j. It follows that

$$X = Y^{(j)} \qquad \text{on } B_j$$

as elements of $\mathcal{C}^d$, up to a null set.

Step 3. Quantization Error on $B_0, \dots, B_J$. We now choose the codebook sizes n_j for the truncated processes $Y^{(j)}$. Let

$$J := \left\lfloor \frac{\log n}{2} \right\rfloor, \qquad n_j := \left\lfloor \frac{1-e^{-1}}{4} n e^{-j} \right\rfloor, \quad 0 \le j \le J.$$

For every $0 \le j \le J$, one has

$$ne^{-j} \ge ne^{-J} \ge n^{1/2}.$$

It follows that, for all sufficiently large n,

$$n_j \ge 2 \qquad \text{and} \qquad \log n_j \ge \frac{\log n}{3}, \tag{18}$$

uniformly over $0 \le j \le J$. The total number of allocated codepoints satisfies

$$1 + \sum_{j=0}^{J} n_j \le 1 + \frac{1-e^{-1}}{4} n \sum_{j=0}^{\infty} e^{-j} = 1 + \frac{n}{4} \le n, \tag{19}$$

for all $n \ge 2$. Apply Theorem 4.1 to $Y^{(j)}$, with moment exponent θq and codebook size n_j. Since the bounding constant for the initial condition, drift, and diffusion coefficient is R_j, there exists a deterministic codebook $A_j \subset \mathcal{C}^d$ with $|A_j| \le n_j$ such that

$$\left(\mathbb{E} d_\infty\big(Y^{(j)}, A_j\big)^{\theta q}\right)^{1/(\theta q)} \le C_{d,\theta q,T} R_j (\log n_j)^{-1/2}. \tag{20}$$

Combining (20) with (18) gives

$$\left(\mathbb{E} d_\infty\big(Y^{(j)}, A_j\big)^{\theta q}\right)^{1/(\theta q)} \le C R_j (\log n)^{-1/2}, \qquad 0 \le j \le J,$$

where, throughout the remainder of the proof, C denotes a finite positive constant independent of n and j. Define the union codebook

$$A := \{0\} \cup \bigcup_{j=0}^{J} A_j.$$

By (19), $|A| \le n$. We first estimate the distortion on $\bigcup_{j=0}^{J} B_j$. Since $A_j \subset A$ and $X = Y^{(j)}$ on B_j, one has

$$\mathbb{E}\left[d_\infty(X,A)^q \mathbf{1}_{B_j}\right] \le \mathbb{E}\left[d_\infty\big(Y^{(j)}, A_j\big)^q \mathbf{1}_{B_j}\right].$$

Applying Hölder's inequality with conjugate exponents θ and $\theta/(\theta-1)$ gives

$$\begin{aligned}\mathbb{E}\left[d_\infty(X,A)^q \mathbf{1}_{B_j}\right] &\le \left(\mathbb{E} d_\infty\big(Y^{(j)},A_j\big)^{\theta q}\right)^{1/\theta} \mathbb{P}(B_j)^\beta \\ &= \left[\left(\mathbb{E} d_\infty\big(Y^{(j)},A_j\big)^{\theta q}\right)^{1/(\theta q)}\right]^q \mathbb{P}(B_j)^\beta \le C R_j^q (\log n)^{-q/2} \mathbb{P}(B_j)^\beta.\end{aligned}$$

For $j \ge 1$, Markov's inequality and the moment assumption on Λ_X yield

$$\mathbb{P}(B_j) \le \mathbb{P}(\Lambda_X > 2^j) \le 2^{-\rho j} \mathbb{E}\Lambda_X^\rho.$$

Since $R_j = 2^{j+1}$, it follows that

$$\sum_{j=0}^J R_j^q \mathbb{P}(B_j)^\beta \le 2^q + C\sum_{j=1}^J 2^{q(j+1)} 2^{-\rho\beta j} \le C + C\sum_{j=1}^\infty 2^{j(q-\rho\beta)}.$$

The final series converges because $\rho\beta > q$. We have consequently obtained

$$\sum_{j=0}^J \mathbb{E}\left[d_\infty(X,A)^q \mathbf{1}_{B_j}\right] \le C(\log n)^{-q/2}. \tag{21}$$

Step 4. Quantization Error on $\{\Lambda_X > 2^{J+1}\}$. It remains to estimate the distortion on $\{\Lambda_X > 2^{J+1}\}$. Since the zero path belongs to A,

$$d_\infty(X,A) \le \|X\|_\infty.$$

Up to a null set, the identity

$$\Omega \setminus \bigcup_{j=0}^J B_j = \{\Lambda_X > 2^{J+1}\}$$

holds. Hölder's inequality with conjugate exponents ρ/q and $\rho/(\rho-q)$ therefore gives

$$\mathbb{E}\left[d_\infty(X,A)^q \mathbf{1}_{\{\Lambda_X > 2^{J+1}\}}\right] \le \mathbb{E}\left[\|X\|_\infty^q \mathbf{1}_{\{\Lambda_X > 2^{J+1}\}}\right] \le (\mathbb{E}\|X\|_\infty^\rho)^{q/\rho}\, \mathbb{P}(\Lambda_X > 2^{J+1})^{1-q/\rho}. \tag{22}$$

Another application of Markov's inequality gives

$$\mathbb{P}(\Lambda_X > 2^{J+1}) \le 2^{-\rho(J+1)} \mathbb{E}\Lambda_X^\rho. \tag{23}$$

Substituting (23) and (17) into (22) yields

$$\mathbb{E}\left[d_\infty(X,A)^q \mathbf{1}_{\{\Lambda_X > 2^{J+1}\}}\right] \le C 2^{-(\rho-q)(J+1)}. \tag{24}$$

Because $J = \lfloor (\log n)/2 \rfloor$, the right-hand side is exponentially small in $\log n$. In particular, for all sufficiently large n,

$$2^{-(\rho-q)(J+1)} \le C e^{-c\log n} \le C(\log n)^{-q/2}. \tag{25}$$

Combining (21), (24), and (25) yields

$$\mathbb{E} d_\infty(X,A)^q \le C(\log n)^{-q/2}.$$

Since $|A| \le n$, the definition of the functional quantization error gives

$$e_{n,q}(X) \le (\mathbb{E} d_\infty(X,A)^q)^{1/q} \le C(\log n)^{-1/2}$$

for every sufficiently large n. For the finitely many remaining values of n, use the codebook consisting only of the zero path. Lyapunov's inequality and (17) give

$$e_{n,q}(X) \le (\mathbb{E}\|X\|_\infty^q)^{1/q} < \infty.$$

Increasing the constant to cover these finitely many values completes the proof. □

4.1.3 From functional quantization to empirical laws

Functional quantization alone does not control the empirical measure, because a deterministic quantizer and an equal-weight empirical sample have different support and weight structures. The following proposition supplies the missing transfer. Truncation confines the problem to a bounded ball, projection onto a finite codebook reduces the central part to multinomial sampling, and a final balance between bounded fluctuations and tail errors yields the deviation bound.

Proposition 4.5. *Let $p \geq 1$, and let Z be a $\mathcal{C}^d$-valued random variable with law η. Assume that, for some $\rho > p$ and $A > 0$,*

$$\mathbb{E}\|Z\|_\infty^\rho < \infty, \qquad e_{n,p}(Z) \leq A(\log n)^{-1/2}, \qquad n \geq 2.$$

Let $Z_1, \dots, Z_N$ be independent copies of Z, and set

$$\eta_N := \frac{1}{N}\sum_{i=1}^N \delta_{Z_i}.$$

Set

$$\beta_{\rho,p} := \min\left\{\frac{\rho}{p} - 1, \frac{\rho}{2p}\right\}.$$

There exist constants $C_{p,\rho,A,\eta} > 0$ and $c_p > 0$, where $C_{p,\rho,A,\eta}$ depends on η only through $\mathbb{E}\|Z\|_\infty^\rho$, such that, for every $N \geq 2$ and $t > 0$,

$$\mathbb{P}\left(\mathcal{W}_p(\eta_N, \eta) > C_{p,\rho,A,\eta}(\log N)^{-1/2} + t\right) \leq \exp\left(-\frac{c_p N t^{2p}}{(\log N)^{p^2/(\rho-p)}}\right) + C_{p,\rho,A,\eta} N^{-\beta_{\rho,p}} t^{-\rho}. \quad (26)$$

Proof. Except for constants displayed explicitly, the subscripts on constants denoted by C or c list all parameters on which they may depend. Their values may change from line to line, and they are independent of N and t.

Step 1. Truncation. Fix $N \geq 16$ and $t > 0$, and set

$$R := (\log N)^{p/[2(\rho-p)]}.$$

Let T_R be the radial truncation

$$T_R(z) := \begin{cases} z, & \|z\|_\infty \leq R, \\ \dfrac{R}{\|z\|_\infty} z, & \|z\|_\infty > R. \end{cases}$$

A direct comparison of the cases according to whether the two norms exceed R gives

$$\|T_R(x) - T_R(y)\|_\infty \leq \|x - y\|_\infty + \big|\|x\|_\infty - \|y\|_\infty\big| \leq 2\|x - y\|_\infty. \quad (27)$$

Set

$$\eta^R := (T_R)_\# \eta, \qquad \eta_N^R := (T_R)_\# \eta_N.$$

Let

$$\xi_i := \|Z_i - T_R(Z_i)\|_\infty^p = \big(\|Z_i\|_\infty - R\big)_+^p.$$

The moment assumption gives

$$\mathbb{E}\xi_1 \leq R^{-(\rho-p)}\mathbb{E}\|Z\|_\infty^\rho, \qquad \mathbb{E}\xi_1^{\rho/p} \leq \mathbb{E}\|Z\|_\infty^\rho. \quad (28)$$

The canonical couplings yield

$$\mathcal{W}_p^p(\eta, \eta^R) \le \mathbb{E}\xi_1, \qquad \mathcal{W}_p^p(\eta_N, \eta_N^R) \le \frac{1}{N}\sum_{i=1}^N \xi_i. \tag{29}$$

Since

$$\mathbb{E}|\xi_1 - \mathbb{E}\xi_1|^{\rho/p} \le C_{p,\rho}\mathbb{E}\|Z\|_\infty^\rho,$$

combining two classical results, one from von Bahr and Esseen [15, Theorem 2, p. 301] (for $p < \rho \le 2p$) and another from Rosenthal [14, Theorem 3, p. 279] (for $\rho > 2p$), give

$$\begin{aligned}\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N(\xi_i - \mathbb{E}\xi_1)\right|^{\rho/p} &\le C_{p,\rho}\mathbb{E}\|Z\|_\infty^\rho \begin{cases} N^{1-\rho/p}, & p < \rho \le 2p, \\ N^{1-\rho/p} + N^{-\rho/(2p)}, & \rho > 2p, \end{cases} \\ &\le C_{p,\rho}\mathbb{E}\|Z\|_\infty^\rho N^{-\beta_{\rho,p}}.\end{aligned}$$

The implication

$$\left\{\mathcal{W}_p(\eta_N, \eta_N^R) > (\mathbb{E}\xi_1)^{1/p} + \frac{t}{2}\right\} \subset \left\{\frac{1}{N}\sum_{i=1}^N(\xi_i - \mathbb{E}\xi_1) > \left(\frac{t}{2}\right)^p\right\}$$

follows from $(a+b)^p \ge a^p + b^p$ for $a, b \ge 0$. Markov's inequality therefore yields

$$\mathbb{P}\left(\mathcal{W}_p(\eta_N, \eta_N^R) > (\mathbb{E}\xi_1)^{1/p} + \frac{t}{2}\right) \le C_{p,\rho,\eta}N^{-\beta_{\rho,p}}t^{-\rho}. \tag{30}$$

Step 2. Projected quantization. Choose a deterministic codebook for Z with at most $\lfloor N^{1/2}\rfloor$ points and p-distortion at most $2A\big(\log\lfloor N^{1/2}\rfloor\big)^{-1/2}$. Projecting its points under T_R produces a codebook $\mathcal{A}$ with at most $\lfloor N^{1/2}\rfloor$ points and diameter at most $2R$. By (27),

$$(\mathbb{E}d_\infty(T_R(Z), \mathcal{A})^p)^{1/p} \le 4A\big(\log\lfloor N^{1/2}\rfloor\big)^{-1/2} \le C_pA(\log N)^{-1/2}.$$

Let $Q : \mathcal{C}^d \to \mathcal{A}$ be a measurable nearest-point projection, with ties resolved by a fixed deterministic rule, and set

$$\bar{\eta} := Q_\#\eta^R, \qquad \bar{\eta}_N := Q_\#\eta_N^R.$$

The population and samplewise nearest-point couplings give

$$\mathcal{W}_p^p(\eta^R, \bar{\eta}) \le \mathbb{E}d_\infty(T_R(Z), \mathcal{A})^p \le C_pA^p(\log N)^{-p/2}, \tag{31}$$

and

$$\mathbb{E}\mathcal{W}_p^p(\eta_N^R, \bar{\eta}_N) \le \frac{1}{N}\sum_{i=1}^N \mathbb{E}d_\infty(T_R(Z_i), \mathcal{A})^p \le C_pA^p(\log N)^{-p/2}. \tag{32}$$

Step 3. Bound between truncated laws. For $a \in \mathcal{A}$, the definition gives

$$\bar{\eta}_N(\{a\}) = N^{-1}\sum_{i=1}^N \mathbf{1}_{\{Q(T_R(Z_i))=a\}}.$$

Clearly, $\mathbf{1}_{\{Q(T_R(Z_i))=a\}}$ are independent Bernoulli random variables with success probability $\bar{\eta}(\{a\})$. Hence

$$\mathbb{E}\bar{\eta}_N(\{a\}) = \bar{\eta}(\{a\}), \qquad \mathrm{Var}\big(\bar{\eta}_N(\{a\})\big) = \frac{\bar{\eta}(\{a\})\big(1 - \bar{\eta}(\{a\})\big)}{N}.$$

Since $\mathrm{diam}(\mathcal{A}) \le 2R$, by the usual Wasserstein-TV inequality,

$$\begin{aligned}
\mathbb{E}\mathcal{W}_p^p(\bar{\eta}_N, \bar{\eta}) &\le \frac{(2R)^p}{2} \sum_{a\in\mathcal{A}} \mathbb{E}|\bar{\eta}_N(\{a\}) - \bar{\eta}(\{a\})| \\
&\le \frac{(2R)^p}{2\sqrt{N}} \sum_{a\in\mathcal{A}} \sqrt{\bar{\eta}(\{a\})(1-\bar{\eta}(\{a\}))} \\
&\le \frac{(2R)^p}{2\sqrt{N}} |\mathcal{A}|^{1/2} \left(\sum_{a\in\mathcal{A}} \bar{\eta}(\{a\})(1-\bar{\eta}(\{a\})) \right)^{1/2} \\
&\le \frac{(2R)^p}{2\sqrt{N}} N^{1/4} \\
&\le C_p R^p N^{-1/4}. \qquad (33)
\end{aligned}$$

The triangle inequality, together with (31), (32), and (33), yields

$$\begin{aligned}
\mathbb{E}\mathcal{W}_p^p(\eta_N^R, \eta^R) &\le 3^{p-1} \left[\mathbb{E}\mathcal{W}_p^p(\eta_N^R, \bar{\eta}_N) + \mathbb{E}\mathcal{W}_p^p(\bar{\eta}_N, \bar{\eta}) + \mathcal{W}_p^p(\bar{\eta}, \eta^R) \right] \\
&\le C_p \left[A^p (\log N)^{-p/2} + R^p N^{-1/4} \right]. \qquad (34)
\end{aligned}$$

For $\mathbf{z} = (z_1, \dots, z_N) \in (\mathcal{C}^d)^N$ with $\max_i \|z_i\|_\infty \le R$, set

$$\Phi(\mathbf{z}) := \mathcal{W}_p^p \left(\frac{1}{N} \sum_{i=1}^N \delta_{z_i}, \eta^R \right).$$

Suppose that $\mathbf{z}$ and $\mathbf{z}'$ are two such vectors that differ only in coordinate j. Let (X, Y) be an optimal coupling of $N^{-1} \sum_{i=1}^N \delta_{z_i}$ and η^R. For each distinct point x among $z_1, \dots, z_N$, write $I_x := \{i : z_i = x\}$. Attach to (X, Y) a random label I by setting

$$\mathbb{P}(I = i \mid X, Y) = \frac{\mathbf{1}_{\{z_i = X\}}}{|I_X|}, \qquad i = 1, \dots, N.$$

In other words, given $X = x$, the label I is chosen uniformly from I_x. Therefore

$$\mathbb{P}(I = i) = \mathbb{E}\big[\mathbb{P}(I = i \mid X, Y)\big] = \frac{\mathbb{P}(X = z_i)}{|I_{z_i}|} = \frac{|I_{z_i}|/N}{|I_{z_i}|} = \frac{1}{N}.$$

Thus I is uniform on $\{1, \dots, N\}$, and $X = z_I$, a.s.. The optimality of coupling (X, Y) gives

$$\Phi(\mathbf{z}) = \mathbb{E} d_\infty(X, Y)^p = \mathbb{E} d_\infty(z_I, Y)^p.$$

Now set $X' := z'_I$. Since I is uniform, X' has law $N^{-1} \sum_{i=1}^N \delta_{z'_i}$, while Y still has law η^R. Hence (X', Y) is a coupling of these two measures. Moreover, $X' = X$ on $\{I \ne j\}$, whereas $X = z_j$ and $X' = z'_j$ on $\{I = j\}$. Therefore

$$\begin{aligned}
\Phi(\mathbf{z}') &\le \mathbb{E} d_\infty(X', Y)^p = \mathbb{E} d_\infty(z'_I, Y)^p \\
&= \Phi(\mathbf{z}) + \mathbb{E}\left[\mathbf{1}_{\{I=j\}} \left(d_\infty(z'_j, Y)^p - d_\infty(z_j, Y)^p \right) \right] \\
&= \Phi(\mathbf{z}) + \frac{1}{N} \mathbb{E}\left[d_\infty(z'_j, Y)^p - d_\infty(z_j, Y)^p \,\middle|\, I = j \right] \\
&\le \Phi(\mathbf{z}) + \frac{(2R)^p}{N}.
\end{aligned}$$

Interchanging $\mathbf{z}$ and $\mathbf{z}'$ gives

$$|\Phi(\mathbf{z}) - \Phi(\mathbf{z}')| \leq \frac{(2R)^p}{N}.$$

The bounded-difference inequality of Boucheron, Lugosi, and Massart [5, Theorem 6.2], with coordinate bound $(2R)^p/N$ and deviation $(t/2)^p$, implies

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{W}_p(\eta_N^R, \eta^R) > \left(\mathbb{E}\mathcal{W}_p^p(\eta_N^R, \eta^R)\right)^{1/p} + \frac{t}{2}\right) \\
&\qquad \leq \mathbb{P}\left(\mathcal{W}_p^p(\eta_N^R, \eta^R) - \mathbb{E}\mathcal{W}_p^p(\eta_N^R, \eta^R) > \left(\frac{t}{2}\right)^p\right) \leq \exp\left(-\frac{c_p N t^{2p}}{R^{2p}}\right).
\end{aligned} \tag{35}$$

Step 4. Balance the parameters and complete the proof. Taking pth roots in (34) and using (28) gives

$$\left(\mathbb{E}\mathcal{W}_p^p(\eta_N^R, \eta^R)\right)^{1/p} + 2(\mathbb{E}\xi_1)^{1/p} \leq C_{p,\rho,A,\eta}\left[(\log N)^{-1/2} + RN^{-1/(4p)} + R^{-(\rho-p)/p}\right].$$

Here $R^{-(\rho-p)/p} = (\log N)^{-1/2}$, while

$$(\log N)^{1/2} RN^{-1/(4p)} = (\log N)^{1/2+p/[2(\rho-p)]} N^{-1/(4p)} \leq C_{p,\rho}$$

for $N \geq 16$. Noticing (29), outside the union of the events in (30) and (35), one has

$$\begin{aligned}
\mathcal{W}_p(\eta_N, \eta) &\leq \mathcal{W}_p(\eta_N, \eta_N^R) + \mathcal{W}_p(\eta_N^R, \eta^R) + \mathcal{W}_p(\eta^R, \eta) \\
&\leq \left(\mathbb{E}\mathcal{W}_p^p(\eta_N^R, \eta^R)\right)^{1/p} + 2(\mathbb{E}\xi_1)^{1/p} + t \leq C_{p,\rho,A,\eta}(\log N)^{-1/2} + t.
\end{aligned}$$

Finally, since $R^{2p} = (\log N)^{p^2/(\rho-p)}$, a union bound and the probability estimates in (30) and (35) prove (26) for $N \geq 16$.

For $2 \leq N < 16$, coupling both measures through $\delta_{\mathbf{0}}$ and using Jensen's inequality give $\mathbb{E}\mathcal{W}_p(\eta_N, \eta)^\rho \leq C_\rho \mathbb{E}\|Z\|_\infty^\rho$; Markov's inequality then extends (26) to this finite range after increasing $C_{p,\rho,A,\eta}$. □

Corollary 4.6. *Under the hypotheses of Proposition 4.5, there exists $C_{p,\rho,A,\eta} > 0$ such that*

$$\mathbb{E}\left[\mathcal{W}_p(\eta_N, \eta)\right] \leq \left(\mathbb{E}\left[\mathcal{W}_p(\eta_N, \eta)^p\right]\right)^{1/p} \leq C_{p,\rho,A,\eta}(\log N)^{-1/2}, \qquad N \geq 2.$$

Proof. For $N \geq 16$, retain the notation R, η^R, η_N^R, and ξ_i from the preceding proof. Minkowski's inequality and the canonical couplings in (29) give

$$\begin{aligned}
(\mathbb{E}\mathcal{W}_p(\eta_N, \eta)^p)^{1/p} &\leq \left(\mathbb{E}\mathcal{W}_p(\eta_N, \eta_N^R)^p\right)^{1/p} + \left(\mathbb{E}\mathcal{W}_p(\eta_N^R, \eta^R)^p\right)^{1/p} + \mathcal{W}_p(\eta^R, \eta) \\
&\leq 2(\mathbb{E}\xi_1)^{1/p} + \left(\mathbb{E}\mathcal{W}_p(\eta_N^R, \eta^R)^p\right)^{1/p} \\
&\leq C_{p,\rho,A,\eta}\left[(\log N)^{-1/2} + RN^{-1/(4p)} + R^{-(\rho-p)/p}\right] \\
&\leq C_{p,\rho,A,\eta}(\log N)^{-1/2}.
\end{aligned}$$

The third line follows from (28) and (34), while the last line follows from the choice of R and the calculation in Step 4.

For $2 \leq N < 16$, coupling both measures through $\delta_{\mathbf{0}}$ gives $\left(\mathbb{E}\mathcal{W}_p(\eta_N, \eta)^p\right)^{1/p} \leq 2\left(\mathbb{E}\|Z\|_\infty^p\right)^{1/p}$, which is absorbed into the claimed bound after increasing the constant. □

Proof of Theorem 2.1. The moment estimate (17) and Proposition 4.4, applied with $q = p$, verify the hypotheses of Corollary 4.6 with $Z = X$ and $\eta = \mu$. The dependence of the constants in these three results gives $C = C\big(d, T, p, \rho, \mathbb{E}\Lambda_X^\rho\big)$, as claimed. □

Proof of Theorem 2.2. The moment estimate (17) and Proposition 4.4, applied with $q = p$, verify the hypotheses of Proposition 4.5 with $Z = X$, $\eta = \mu$, and $\eta_N = \mu_N$. This proves (2) with $C = C\big(d, T, p, \rho, \mathbb{E}\Lambda_X^\rho\big)$.

Taking $t = (\log N)^{-1/2}$ in (2) changes the threshold to $(C+1)(\log N)^{-1/2}$. The exponential term is bounded by $CN^{-\beta_{\rho,p}}(\log N)^{\rho/2}$ uniformly over $N \geq 2$ after increasing C. Absorbing the additional one in the threshold and renaming the constant proves the remaining assertion. □

Proof of Proposition 2.5. The product law $\mu^{\otimes N}$ satisfies the same quadratic transport inequality for the product metric

$$\left(\sum_{i=1}^N d_\infty(x_i, x_i')^2\right)^{1/2}.$$

The reverse triangle inequality and the coupling of equally labelled atoms give

$$\begin{aligned}&\left|\mathcal{W}_p\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \mu\right) - \mathcal{W}_p\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i'}, \mu\right)\right| \\ &\qquad \leq \left(\frac{1}{N}\sum_{i=1}^N d_\infty(x_i, x_i')^p\right)^{1/p} \leq N^{-1/\max\{p,2\}}\left(\sum_{i=1}^N d_\infty(x_i, x_i')^2\right)^{1/2}.\end{aligned}$$

The product tensorization theorem and the corresponding mean-centered concentration inequality of Gozlan and Léonard [11, Theorem 4.12 and Lemma 6.1] therefore yield

$$\mathbb{P}\left(\mathcal{W}_p(\mu_N, \mu) > \mathbb{E}\mathcal{W}_p(\mu_N, \mu) + t\right) \leq \exp\left(-\frac{N^{2/\max\{p,2\}}t^2}{2C_T}\right).$$

Theorem 2.1 bounds the expectation by $C\big(d, T, p, \rho, \mathbb{E}\Lambda_X^\rho\big)(\log N)^{-1/2}$. □

4.2 Proof of Proposition 3.2

Proof. Since X is continuous and adapted, the $\mathcal{C}^d$-valued process $t \mapsto X_{t\wedge\cdot}$ is also adapted. Moreover, for almost every ω, the map

$$t \longmapsto X_{t\wedge\cdot}(\omega)$$

is continuous with respect to the uniform norm on $\mathcal{C}^d$. Hence the stopped-path process is progressively measurable. Since b and σ are Borel measurable, it follows that

$$a_t = b(t, X_{t\wedge\cdot}), \qquad H_t = \sigma(t, X_{t\wedge\cdot})$$

are progressively measurable. The equation in Assumption 3.1 can therefore be written as

$$X_t = X_0 + \int_0^t a_s\, ds + \int_0^t H_s\, dW_s, \qquad 0 \leq t \leq T. \tag{36}$$

It remains to verify the required integrability of the time-essential suprema. Define

$$X_t^* := \sup_{0\leq u\leq t} |X_u|.$$

We first prove that

$$\mathbb{E}\big[(X_T^*)^\rho\big] < \infty.$$

For $n \in \mathbb{N}$, let

$$\tau_n := \inf\{t \in [0,T] : X_t^* \geq n\} \wedge T, \qquad X_t^{*,n} := \sup_{0 \leq u \leq t} |X_{u \wedge \tau_n}|.$$

For $s \leq \tau_n$, the linear growth assumption gives

$$|a_s| + \|H_s\| \leq L\big(1 + \|X_{s\wedge\cdot}\|_\infty\big) = L(1 + X_s^*) = L(1 + X_s^{*,n}).$$

Consequently,

$$\mathbf{1}_{\{s \leq \tau_n\}}\big(|a_s| + \|H_s\|\big) \leq L(1 + X_s^{*,n}).$$

Using the stopped form of (36), the elementary inequality

$$|x + y + z|^\rho \leq C_\rho\big(|x|^\rho + |y|^\rho + |z|^\rho\big),$$

and the Burkholder–Davis–Gundy inequality, we obtain, for $0 \leq t \leq T$,

$$\begin{aligned}\mathbb{E}\big[(X_t^{*,n})^\rho\big] &\leq C_\rho \mathbb{E}|X_0|^\rho + C_\rho \mathbb{E}\left(\int_0^t \mathbf{1}_{\{s \leq \tau_n\}} |a_s|\, ds\right)^\rho + C_\rho \mathbb{E}\left[\sup_{0 \leq u \leq t}\left|\int_0^u \mathbf{1}_{\{s \leq \tau_n\}} H_s\, dW_s\right|^\rho\right] \\ &\leq C_\rho \mathbb{E}|X_0|^\rho + C_{\rho,L}\mathbb{E}\left(\int_0^t (1 + X_s^{*,n})\, ds\right)^\rho + C_{\rho,L}\mathbb{E}\left(\int_0^t (1 + X_s^{*,n})^2\, ds\right)^{\rho/2}. \quad (37)\end{aligned}$$

By Hölder's inequality,

$$\left(\int_0^t (1 + X_s^{*,n})\, ds\right)^\rho \leq t^{\rho-1}\int_0^t (1 + X_s^{*,n})^\rho\, ds. \quad (38)$$

Furthermore,

$$\left(\int_0^t (1 + X_s^{*,n})^2\, ds\right)^{\rho/2} \leq \left[(1 + X_t^{*,n})\int_0^t (1 + X_s^{*,n})\, ds\right]^{\rho/2}.$$

Thus, by Young's inequality, for every $\varepsilon > 0$,

$$\begin{aligned}\left(\int_0^t (1 + X_s^{*,n})^2\, ds\right)^{\rho/2} &\leq \varepsilon(1 + X_t^{*,n})^\rho + C_{\varepsilon,\rho}\left(\int_0^t (1 + X_s^{*,n})\, ds\right)^\rho \\ &\leq \varepsilon(1 + X_t^{*,n})^\rho + C_{\varepsilon,\rho,T}\int_0^t (1 + X_s^{*,n})^\rho\, ds. \quad (39)\end{aligned}$$

Combining (37), (38), and (39) with

$$(1 + x)^\rho \leq 2^{\rho-1}(1 + x^\rho), \qquad x \geq 0,$$

gives

$$\mathbb{E}\big[(X_t^{*,n})^\rho\big] \leq C_{\rho,L,T}\big(1 + \mathbb{E}|X_0|^\rho\big) + C_{\rho,L}\varepsilon\mathbb{E}\big[(X_t^{*,n})^\rho\big] + C_{\varepsilon,\rho,L,T}\int_0^t \mathbb{E}\big[(X_s^{*,n})^\rho\big]\, ds.$$

Choose $\varepsilon > 0$ sufficiently small so that the second term on the right-hand side can be absorbed into the left-hand side. We then obtain

$$\mathbb{E}\big[(X_t^{*,n})^\rho\big] \le C_{\rho,L,T}\big(1+\mathbb{E}|X_0|^\rho\big) + C_{\rho,L,T}\int_0^t \mathbb{E}\big[(X_s^{*,n})^\rho\big]\,ds.$$

Grönwall's lemma yields

$$\sup_{n\in\mathbb{N}} \mathbb{E}\big[(X_T^{*,n})^\rho\big] \le C_{\rho,L,T}\big(1+\mathbb{E}|X_0|^\rho\big) < \infty.$$

Since X has continuous paths on the compact interval $[0,T]$, we have $X_T^* < \infty$ almost surely and

$$X_T^{*,n} \longrightarrow X_T^* \qquad \text{almost surely.}$$

Fatou's lemma therefore gives

$$\mathbb{E}\big[(X_T^*)^\rho\big] \le \liminf_{n\to\infty} \mathbb{E}\big[(X_T^{*,n})^\rho\big] \le C_{\rho,L,T}\big(1+\mathbb{E}|X_0|^\rho\big).$$

Finally, the linear growth condition implies, for every $s \in [0,T]$,

$$|a_s| + \|H_s\| \le L\big(1+\|X_{s\wedge\cdot}\|_\infty\big) = L(1+X_s^*) \le L(1+X_T^*).$$

Hence

$$\operatorname*{ess\,sup}_{0\le s\le T} |a_s| \le L(1+X_T^*), \qquad \operatorname*{ess\,sup}_{0\le s\le T} \|H_s\| \le L(1+X_T^*).$$

Since $L \ge 1$ and $|X_0| \le X_T^*$, it follows that

$$\Lambda_X \le L(1+X_T^*).$$

Therefore,

$$\mathbb{E}\Lambda_X^\rho \le L^\rho 2^{\rho-1}\left(1+\mathbb{E}\big[(X_T^*)^\rho\big]\right) \le C_{\rho,L,T}\big(1+\mathbb{E}|X_0|^\rho\big).$$

Thus all the requirements of Assumption 2.1 are satisfied. The result and the stated dependence of the constant now follow from Theorem 2.1. □

4.3 Proof of Proposition 3.4

The proof has two components. The first verifies that the nonlinear limit satisfies the quantization estimate. The second couples each interacting particle with an independent nonlinear copy driven by the same initial condition and Brownian motion, and then propagates the i.i.d. empirical error through the Lipschitz dynamics.

4.3.1 Quantization of the nonlinear limit

Lemma 4.7. *Under Assumption 3.3,*

$$\mathbb{E}\,\|Y\|_\infty^r < \infty$$

and there is a constant C such that

$$e_{n,p}(Y) \le C(\log n)^{-1/2}, \qquad n \ge 2.$$

Proof. The linear-growth estimate in Assumption 3.3, together with the elementary inequality $(a+b+c)^r \le C_r(a^r+b^r+c^r)$, gives

$$|b(t,x,\nu)|^r + \|\sigma(t,x,\nu)\|^r \le C\big(1+\|x\|_\infty^r + \mathcal{W}_p(\nu,\delta_{\mathbf{0}})^r\big).$$

The standard localization argument used in the proof of Proposition 3.2, with Hölder's inequality for the drift term, the Burkholder–Davis–Gundy inequality for the martingale term, and Young's inequality when $r<2$, gives

$$\mathbb{E}\,\|Y_{t\wedge\cdot}\|_\infty^r \le C\big(1+\mathbb{E}\,|\xi|^r\big) + C\int_0^t \left(\mathbb{E}\,\|Y_{s\wedge\cdot}\|_\infty^r + \mathcal{W}_p(\mu_s,\delta_{\mathbf{0}})^r\right)\,\mathrm{d}s.$$

Because $r>p$ and $\mu_s = \mathcal{L}(Y_{s\wedge\cdot})$,

$$\mathcal{W}_p(\mu_s,\delta_{\mathbf{0}})^r = \left(\mathbb{E}\,\|Y_{s\wedge\cdot}\|_\infty^p\right)^{r/p} \le \mathbb{E}\,\|Y_{s\wedge\cdot}\|_\infty^r\,.$$

Gronwall's inequality therefore shows that

$$\mathbb{E}\,\|Y\|_\infty^r \le C\big(1+\mathbb{E}\,|\xi|^r\big).$$

The stopped-path law flow $t\mapsto\mu_t$ is deterministic. It is also continuous in $\mathcal{W}_p$, since the coupling provided by the same process gives

$$\mathcal{W}_p(\mu_t,\mu_s)^p \le \mathbb{E}\,\|Y_{t\wedge\cdot} - Y_{s\wedge\cdot}\|_\infty^p\,,$$

and the right-hand side converges to zero as $t\to s$ by path continuity and dominated convergence.

Set

$$a_t := b(t,Y_{t\wedge\cdot},\mu_t), \qquad H_t := \sigma(t,Y_{t\wedge\cdot},\mu_t).$$

The process $t\mapsto Y_{t\wedge\cdot}$ is progressively measurable as a $\mathcal{C}^d$-valued process, while $t\mapsto\mu_t$ is deterministic and continuous. Hence a and H are progressively measurable. Moreover,

$$\sup_{0\le t\le T}\mathcal{W}_p(\mu_t,\delta_{\mathbf{0}}) \le \left(\mathbb{E}\,\|Y\|_\infty^p\right)^{1/p} < \infty,$$

and the linear-growth estimate gives

$$|a_t| + \|H_t\| \le L\left(1+\|Y\|_\infty + \left(\mathbb{E}\,\|Y\|_\infty^p\right)^{1/p}\right), \qquad 0\le t\le T.$$

Consequently, with Λ_Y defined as in Assumption 2.1,

$$\Lambda_Y \le C\left(1+\|Y\|_\infty + \left(\mathbb{E}\,\|Y\|_\infty^p\right)^{1/p}\right),$$

and therefore

$$\mathbb{E}\Lambda_Y^r < \infty.$$

Thus Assumption 2.1 holds for Y with $\rho = r$. Proposition 4.4, applied with $q=p$, gives the claimed quantization estimate. $\square$

4.3.2 Synchronous-coupling stability

We now transfer the empirical approximation of the nonlinear law to the interacting system. The coupling separates the total error into an i.i.d. empirical term and a dynamical stability term. The latter is estimated in L^p; the treatment of the martingale integral differs according to whether $p \geq 2$ or $p < 2$, but both regimes close by Grönwall's inequality.

Proof of Proposition 3.4. For every $i \in \{1, \dots, N\}$, let Y^i solve the limit equation with the same initial value and the same Brownian motion as the ith interacting particle, i.e.,

$$Y_t^i = \xi^i + \int_0^t b(s, Y^i_{s\wedge\cdot}, \mu_s)\,\mathrm{d}s + \int_0^t \sigma(s, Y^i_{s\wedge\cdot}, \mu_s)\,\mathrm{d}W_s^i.$$

Because the stopped-path law flow $(\mu_s)_{0\leq s\leq T}$ is deterministic and the pairs (ξ^i, W^i) are independent, the processes $Y^1, \dots, Y^N$ are independent copies of Y. Introduce their empirical full path measure and empirical stopped-path measure by

$$\nu^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y^i}, \qquad \nu_s^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y^i_{s\wedge\cdot}}.$$

Lemma 4.7 and Corollary 4.6, applied with $\rho = r$, show that

$$a_N := \left(\mathbb{E}\mathcal{W}_p(\nu^N, \mu)^p\right)^{1/p} \leq C(\log N)^{-1/2}.$$

Set

$$\Delta_t^i := X_t^{i,N} - Y_t^i.$$

Subtracting the two stochastic equations gives

$$\Delta_t^i = \int_0^t \left[b(s, X^{i,N}_{s\wedge\cdot}, \mu_s^N) - b(s, Y^i_{s\wedge\cdot}, \mu_s)\right]\mathrm{d}s + \int_0^t \left[\sigma(s, X^{i,N}_{s\wedge\cdot}, \mu_s^N) - \sigma(s, Y^i_{s\wedge\cdot}, \mu_s)\right]\mathrm{d}W_s^i.$$

By Assumption 3.3, the absolute value of the drift difference and the Hilbert–Schmidt norm of the diffusion difference are both bounded, up to the factor L, by

$$\left\|X^{i,N}_{s\wedge\cdot} - Y^i_{s\wedge\cdot}\right\|_\infty + \mathcal{W}_p(\mu_s^N, \mu_s) = \sup_{0\leq u\leq s}\left|\Delta_u^i\right| + \mathcal{W}_p(\mu_s^N, \mu_s).$$

Synchronous coupling gives

$$\mathcal{W}_p(\mu_s^N, \nu_s^N)^p \leq \frac{1}{N}\sum_{j=1}^N \sup_{0\leq u\leq s}\left|\Delta_u^j\right|^p.$$

The contraction property of Wasserstein distance yields

$$\mathcal{W}_p(\nu_s^N, \mu_s) \leq \mathcal{W}_p(\nu^N, \mu).$$

Consequently, we have that for every $t \leq T$,

$$\mathcal{W}_p(\mu_s^N, \mu_s)^p \leq C\left[\frac{1}{N}\sum_{j=1}^N \sup_{0\leq u\leq s}\left|\Delta_u^j\right|^p + \mathcal{W}_p(\nu^N, \mu)^p\right], \tag{40}$$

and

$$\sup_{0\le s\le t} \mathcal{W}_p(\mu_s^N,\mu_s)^p \le C\left[\frac{1}{N}\sum_{j=1}^N \sup_{0\le u\le t}\left|\Delta_u^j\right|^p + \mathcal{W}_p(\nu^N,\mu)^p\right]. \tag{41}$$

Write

$$U_N(t) := \frac{1}{N}\sum_{i=1}^N \mathbb{E}\sup_{0\le u\le t}\left|\Delta_u^i\right|^p.$$

To proceed with the proof, we discuss two seperate cases by the value of p below.

Case $p \ge 2$. Hölder's inequality for the time integral and the Burkholder–Davis–Gundy inequality give, for $t \le T$,

$$\mathbb{E}\sup_{0\le u\le t}\left|\Delta_u^i\right|^p \le C\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\left|\Delta_u^i\right|^p + \mathcal{W}_p(\mu_s^N,\mu_s)^p\right] \mathrm{d}s.$$

Averaging over i and using (41) gives

$$U_N(t) \le C\int_0^t \big(U_N(s) + a_N^p\big)\,\mathrm{d}s.$$

Gronwall's inequality therefore yields $U_N(T) \le C a_N^p$.

Case $p \in [1,2)$. For notational convenience, set

$$h_s^i := \sup_{0\le u\le s}\left|\Delta_u^i\right| + \mathcal{W}_p(\mu_s^N,\mu_s).$$

Hölder's inequality for the drift integral and the Burkholder–Davis–Gundy inequality give

$$\mathbb{E}\sup_{0\le u\le t}\left|\Delta_u^i\right|^p \le C\int_0^t \mathbb{E}(h_s^i)^p\,\mathrm{d}s + C\mathbb{E}\left(\int_0^t (h_s^i)^2\,\mathrm{d}s\right)^{p/2}.$$

Since $2-p>0$, one has

$$\left(\int_0^t (h_s^i)^2\,\mathrm{d}s\right)^{p/2} \le \left(\sup_{0\le u\le t}(h_u^i)^p\right)^{(2-p)/2}\left(\int_0^t (h_s^i)^p\,\mathrm{d}s\right)^{p/2}.$$

Young's inequality with conjugate exponents $2/(2-p)$ and $2/p$ shows that, for every $\varepsilon>0$,

$$\left(\int_0^t (h_s^i)^2\,\mathrm{d}s\right)^{p/2} \le \varepsilon\sup_{0\le u\le t}(h_u^i)^p + C_{p,\varepsilon}\int_0^t (h_s^i)^p\,\mathrm{d}s.$$

It follows that

$$\mathbb{E}\sup_{0\le u\le t}\left|\Delta_u^i\right|^p \le \varepsilon\mathbb{E}\sup_{0\le u\le t}(h_u^i)^p + C_\varepsilon\int_0^t \mathbb{E}(h_s^i)^p\,\mathrm{d}s, \tag{42}$$

after rescaling ε to absorb the multiplicative constant.

Equations (41) and (40) imply, respectively,

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}\sup_{0\le s\le t}(h_s^i)^p \le C\big(U_N(t) + a_N^p\big), \tag{43}$$

and, for every $s \le t$,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(h_s^i)^p \le C\big(U_N(s)+a_N^p\big). \tag{44}$$

Averaging (42) over i and using (43) and (44) gives

$$U_N(t) \le C\varepsilon\big(U_N(t)+a_N^p\big)+C_\varepsilon\int_0^t\big(U_N(s)+a_N^p\big)\,\mathrm{d}s.$$

Choose $\varepsilon > 0$ sufficiently small that $C\varepsilon \le 1/2$. The term involving $U_N(t)$ can then be absorbed into the left-hand side, and Gronwall's inequality yields

$$U_N(T) \le Ca_N^p.$$

Together with the bound $U_N(T) \le Ca_N^p$ established for $p \ge 2$, the preceding estimate proves the same bound for every $p \ge 1$. Synchronous coupling thus gives

$$\big(\mathbb{E}\mathcal{W}_p(\mu^N,\nu^N)^p\big)^{1/p} \le U_N(T)^{1/p} \le Ca_N.$$

Finally, the triangle inequality for $\mathcal{W}_p$ and Minkowski's inequality imply

$$\big(\mathbb{E}\mathcal{W}_p(\mu^N,\mu)^p\big)^{1/p} \le \big(\mathbb{E}\mathcal{W}_p(\mu^N,\nu^N)^p\big)^{1/p}+\big(\mathbb{E}\mathcal{W}_p(\nu^N,\mu)^p\big)^{1/p} \le Ca_N \le C(\log N)^{-1/2}.$$

The stated mean estimate follows from Hölder's inequality. □

References

[1] E. Bayraktar, X. He, and D. Kim, Graphon particle systems with common noise, to appear in *Bernoulli*.

[2] E. Boissard, Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance, *Electron. J. Probab.* **16** (2011), 2296–2333.

[3] E. Boissard and T. Le Gouic, On the mean speed of convergence of empirical and occupation measures in Wasserstein distance, *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (2014), 539–563.

[4] F. Bolley, Quantitative concentration inequalities on sample path space for mean field interaction, *ESAIM Probab. Stat.* **14** (2010), 192–209.

[5] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, 2013.

[6] J. Creutzig, S. Dereich, T. Müller-Gronbach, and K. Ritter, Infinite-dimensional quadrature and approximation of distributions, *Found. Comput. Math.* **9** (2009), 391–429.

[7] S. Dereich, The coding complexity of diffusion processes under supremum norm distortion, *Stochastic Process. Appl.* **118** (2008), 917–937.

[8] S. Dereich, F. Fehringer, A. Matoussi, and M. Scheutzow, On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces, *J. Theoret. Probab.* **16** (2003), 249–265.

[9] S. Dereich and M. Scheutzow, High-resolution quantization and entropy coding for fractional Brownian motion, *Electron. J. Probab.* **11** (2006), 700–722.

[10] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, *Probab. Theory Related Fields* **162** (2015), 707–738.

[11] N. Gozlan and C. Léonard, A large deviation approach to some transportation cost inequalities, *Probab. Theory Related Fields* **139** (2007), 235–283.

[12] J. Lei, Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces, *Bernoulli* **26** (2020), 767–798.

[13] H. Luschgy and G. Pagès, Functional quantization of a class of Brownian diffusions: A constructive approach, *Stochastic Process. Appl.* **116** (2006), 310–336.

[14] H. P. Rosenthal, On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables, *Israel J. Math.* **8** (1970), 273–303.

[15] B. von Bahr and C.-G. Esseen, Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$, *Ann. Math. Statist.* **36** (1965), 299–303.

[16] J. Weed and F. Bach, Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance, *Bernoulli* **25** (2019), 2620–2648.