

Convergence rates for path-dependent mean field control: from discrete to continuous time

Mathieu Laurière* Mehdi Talbi†

Abstract

We establish quantitative convergence rates for finite-population approximations of mean field control problems. We first consider finite-horizon discrete-time problems on an unbounded state space, allowing for common noise and interactions through the joint empirical distribution of states and controls. Under Lipschitz regularity of the mean field value functions and uniform second-moment estimates, we derive explicit error bounds for the N -particle value functions. We also construct a path-dependent example with an explicit optimal feedback, showing that the dependence of the convergence exponent on the number of time steps is sharp. Finally, combining these estimates with a uniform time-discretisation argument, we establish a logarithmic convergence rate for continuous-time path-dependent McKean–Vlasov control problems with uncontrolled and possibly degenerate diffusion. Numerical experiments illustrate the dependence of the discrete-time rate on path dimension and the effect of common noise in a controlled Markovian model.

Keywords. Mean field control, convergence rates, path dependent control, euler scheme. **MSC Codes.** 49N80, 93E20, 60K35.

Contents

1	Introduction	2
2	Convergence rate for the discrete time problem	5
2.1	The mean field control problem in discrete time	5
2.2	The N -particle problem in discrete time	6
2.3	Convergence rate	7
2.4	Extension to the path-dependent case	9
2.5	An explicit feedback example with a full path-law rate	12
3	Convergence rate for continuous time problems	15
3.1	The path-dependent control problems	15
3.2	Discrete time approximations	16
3.3	Main result	17
3.4	The Markovian case	18
4	Numerical experiments	19
4.1	A controlled Markovian benchmark.	19
4.2	A controlled common-noise benchmark.	21
4.3	An explicit full-feedback path-law value gap.	22

*Shanghai Center for Data Science; NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai; NYU Shanghai, Shanghai, People’s Republic of China, mathieu.lauriere@nyu.edu

†Laboratoire de Probabilités, Statistiques et Modélisation, Université Paris-Cité, France, mtalbi@lpsm.paris

A	Technical Lemmata	27
B	Proofs for the explicit path-dependent example	35
B.1	Optimal feedback and convergence rate	35
B.2	Finite-sample bounds	36
C	Uniform time discretisation for path-dependent particle control	39
D	Proof of Lemma 3.2	44
D.1	The two randomised control problems	44
D.2	Equality of the value functions	45
D.3	Uniform regularity and moment estimates	47
D.4	Derivation of the particle rate	50

1 Introduction

Mean field control problems arise as limits of cooperative control problems for large systems of interacting agents, or alternatively as the problem of a central planner controlling a large system of interacting agents. The controls are chosen jointly to optimise a common objective, in contrast with mean field games (introduced by Lasry & Lions [19] and Huang, Caines & Malhamé [16]), where the limiting problem describes the equilibrium of a non-cooperative population. The mean field formulation replaces the empirical distribution of the particles by the law of a representative controlled process, conditional on the common noise when present. A natural question is then to quantify the error of this approximation at the level of the value functions. In this paper, we first study this question in discrete time, allowing for common noise and interactions through the joint distribution of states and controls. We then extend our results to the case of continuous-time control problems, including path-dependent dynamics and rewards.

The qualitative convergence of large population control problems towards their mean field counterparts has been established in fairly general settings. Lacker [18] develops a limit theory based on controlled martingale problems, including the convergence of nearly optimal controls. Djete, Possamai and Tan [11] establish a general convergence result with common noise, together with the equivalence between strong, weak and relaxed formulations. From a partial differential equation perspective, Talbi [23] proves convergence of finite-dimensional approximations of viscosity solutions on Wasserstein space, with an extension to path-dependent equations. These results provide a qualitative justification of particle approximations, but do not yield a convergence rate.

Quantitative results depend substantially on the regularity and structure of the control problem. On a finite state space, Cecchin [8] obtains an $N^{-1/2}$ rate for the value functions without a convexity assumption. For equations on Wasserstein space admitting a sufficiently smooth solution, Germain, Pham and Warin [15] derive an N^{-1} particle approximation error by backward stochastic differential equation techniques. Jackson and Lacker [17] obtain sharp quantitative comparisons in a displacement convex setting by studying the difference between full-information and distributed controls. Smoothness of the data does not, however, imply global smoothness of the mean field value function. Building on the regularity analysis of Cardaliaguet and Souganidis [5], Cardaliaguet, Jackson, Mimikos-Stamatopoulos and Souganidis [7] establish an N^{-1} rate on suitable compact subsets of the region of strong regularity.

Global estimates without assuming a smooth mean field value function have been developed along different lines. Cardaliaguet, Daudin, Jackson and Souganidis [6] prove an algebraic convergence rate in the presence of idiosyncratic and common noises, using uniform regularity estimates for the particle value functions and concentration inequalities. Daudin, Delarue and Jackson [10]

obtain more precise rates for non-degenerate Markovian problems on the torus through regularisation of the limiting value function: an $N^{-1/2}$ bound under stronger regularity assumptions on the data, and a bound of order $N^{-2/(3d+6)+\eta}$, for every $\eta > 0$, under their Wasserstein-Lipschitz and semiconcavity assumptions. Cecchin, Daudin, Jackson and Martini [9] extend this regularisation approach to common noise and possibly degenerate, state-dependent idiosyncratic diffusion. An alternative analytic approach, based on infimum and supremum convolutions and viscosity comparison on Wasserstein space, is developed by Bayraktar, Ekren and Zhang [2].

More recently, Munoz [22] obtains the empirical-measure rates $N^{-1/d}$ for $d \geq 3$ and $N^{-1/2}\sqrt{\log N}$ for $d = 2$ on the torus, with additive non-degenerate diffusion, a regular uniformly convex control cost and merely Wasserstein-Lipschitz mean field costs. The control-theoretic proof uses repeated optimal-transport recouplings and extends to additive common noise.

In discrete time, Gast and Gaujal [14] prove finite-horizon optimal-value convergence for populations of finite-state agents. Quantitative comparisons with common noise are developed by Motte and Pham [20; 21]. Their quantitative analysis concerns discounted infinite-horizon problems with common noise on compact state and action spaces. The value error is bounded by a power of an empirical-measure error, with an exponent reflecting the regularity of the value function. Their argument already relies on a comparison of the finite-population and mean field Bellman operators. Related path-dependent quantitative results are obtained by Bouchard and Tan [4] when only the drift is controlled. In their setting, mean field controls are adapted only to the common-noise filtration, and the finite-population controller applies an identical signal to all agents.

We also mention the contribution of Bokanowski, Chassagneux, Li and Reisinger [3, Theorems 3.6 and 5.2], who derive bounds for time discretisation and particle approximation, with a particle exponent depending on the dimension of the discretised path space. Their particle approximation uses the same feedback function for every particle, evaluated at that particle's initial state and past Brownian increments.

Our objective is to obtain finite-horizon estimates on an unbounded state space which remain quantitative when the state is enlarged to include the past trajectory. The main assumptions concern Wasserstein-Lipschitz continuity of the mean field value functions, continuity of the particle values, and uniform second-moment estimates. We do not assume differentiability or semiconcavity of the value functions, nor regularity of optimal feedbacks. A central point is that a propagation-of-chaos estimate for a fixed feedback does not suffice to compare the optimised values: the particle controller may use the information contained in the whole particle system. We address this difficulty directly at the level of the dynamic programming equations.

More precisely, let p be the number of time steps and let

$$\mu_{\mathbf{x}}^N := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \quad m_2^N(\mathbf{x}) := \frac{1}{N} \sum_{i=1}^N |x_i|^2.$$

In the Markovian setting, under an additional absolute-continuity assumption on the one-step conditional transition laws, Theorem 2.2 gives

$$|V_0^N(\mathbf{x}) - V_0(\mu_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x}))pR_{N,d}, \quad (1.1)$$

where V_0 and V_0^N denote respectively the value functions of the mean field and the finite population problems, and

$$R_{N,d} := \begin{cases} N^{-1/2}(\log(1+N))^{1/2}, & d = 1, \\ N^{-1/2}(\log(1+N))^{3/2}, & d = 2, \\ N^{-1/d}, & d \geq 3. \end{cases}$$

The constant C is independent of N and p when the regularity and moment constants are uniform in p . The estimate allows common noise and dependence on the joint state-control distribution.

The proof compares the two one-step optimisation problems at empirical measures. At a configuration with pairwise distinct positions, every particle action vector can be realised by a measurable individual-state feedback on that configuration. Conditionally on the current configuration and the next common-noise increment, the subsequent particle positions are independent, but need not have the same distribution. Their average conditional law coincides with the next mean field law. We therefore adapt the empirical Wasserstein estimates of Fournier and Guillin [13] to independent, non-identically distributed variables and use the Lipschitz continuity of the mean field value functions to control the one-step error. Iterating this comparison gives the factor p in (1.1). Continuity then extends the estimate to initial configurations with coinciding positions.

For path-dependent dynamics, the same argument is applied to the stopped processes. The main additional issue is to control the constants in the empirical-measure estimate as the dimension $d(p+1)$ of the discrete path space increases. Using the supremum norm on discrete trajectories and adapting the explicit bounds of Fournier [12], we obtain a constant independent of the number of dates. This yields, in Theorem 2.4, the bound

$$|V_0^N(\omega) - V_0(\mu_\omega^N)| \leq C(1 + m_2^N(\omega))pN^{-1/[d(p+1)]}.$$

The uniformity of C in p is essential for the continuous-time application, in which the number of dates must increase with N .

Section 2.5 presents a scalar path-dependent example with an explicit optimal feedback over the full mean field and particle control classes. For midpoint initial configurations, the value gap has order $N^{-1/(p+1)}$ for every fixed $p \geq 2$, attaining the exponent in Theorem 2.4 when $d = 1$. Section 4 compares Monte Carlo estimates of this gap with computable finite- N bounds and studies controlled Markovian models with and without common noise.

We finally derive a quantitative continuous-time population bound for path-dependent problems without common noise, with a controlled drift and an uncontrolled, possibly degenerate, diffusion. Under Lipschitz assumptions formulated with the L^2 -norm on paths and a closedness and convexity condition on the attainable drift-reward pairs, we adapt the mean field time-discretisation argument of [3] to the particle problem and obtain an error of order $h^{1/4}$ with a constant independent of N . We show that open-loop controls and suitable randomised feedback controls give the same value functions for the time-discretised problems, and establish the required regularity estimates uniformly in the grid. Combining these estimates gives an explicit bound in both N and p , from which we choose the mesh as a function of the population size. For $h = T/p$, the bound is

$$|V_0^N(\mathbf{x}) - V_0(\mu_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) \left(h^{1/4} + pN^{-1/[d(p+1)]} \right).$$

Figure 1 summarises the three comparisons leading to this bound. Here $U_0^{N,\pi}$ and U_0^π are the value functions of the particle and mean field problems after time discretisation on a uniform grid π of step $h = T/p$.

Choosing p of order $\log N / \log \log N$ with a suitable proportionality constant yields

$$|V_0^N(\mathbf{x}) - V_0(\mu_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) \left(\frac{\log \log N}{\log N} \right)^{1/4}$$

for sufficiently large N . This logarithmic estimate reflects the increasing dimension of the discretised path space. We do not claim that it is sharp.

The paper is organised as follows. Section 2 presents the discrete-time problems and proves the Markovian and path-dependent convergence estimates. Section 3 combines our discrete-time rates with time-discretisation error bounds to derive convergence rates for continuous-time problems, including the path-dependent case. Section 4 provides numerical illustrations of the value errors and of the dimension dependence of the empirical path-law approximation. The empirical-measure

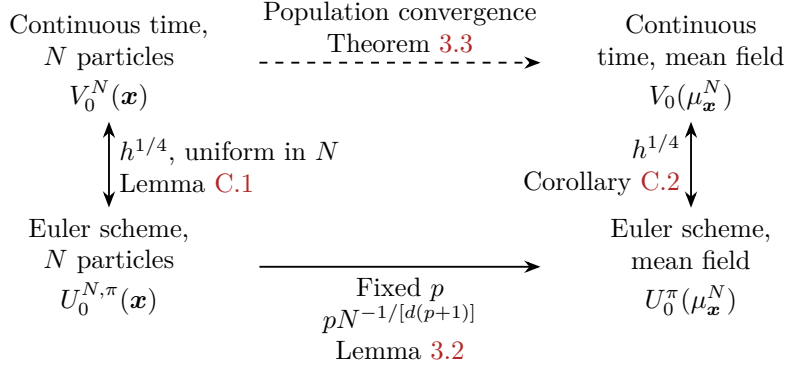


Figure 1: Comparison of the four optimal values for the path-dependent problem without common noise, with $h = T/p$. Solid arrows give error bounds with the factor $C(1 + m_2^N(\mathbf{x}))$ suppressed, where C is independent of N and p . The dashed arrow follows by combining these bounds and choosing $p = p(N)$ so that both error terms vanish.

estimates, the uniform time-discretisation argument, and the randomised Euler formulation are detailed in the appendices.

Notations. For a measurable space E , we denote by $\mathbb{L}^0(E, \mathbb{A})$ the set of measurable maps from E to $\mathbb{A}$. For $q \geq 1$, when (E, d) is a metric space, we denote by $\mathcal{P}_q(E)$ the q -Wasserstein space of probability measures on E , endowed with the distance:

$$\mathcal{W}_q(\mu, \nu) := \min_{\mathbb{Q} \in \Gamma(\mu, \nu)} \left(\int_{E \times E} d^q(x, y) \mathbb{Q}(dx, dy) \right)^{1/q},$$

with $\Gamma(\mu, \nu)$ the set of couplings of μ and ν . Given $\mathbf{x} := (x_1, \dots, x_N) \in E^N$, we denote by $\mu_{\mathbf{x}}^N$ the empirical measure on E defined by $\mu_{\mathbf{x}}^N := \frac{1}{N} \sum_{k=1}^N \delta_{x_k}$.

2 Convergence rate for the discrete time problem

2.1 The mean field control problem in discrete time

We fix a finite time horizon $p \in \mathbb{N}^*$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, endowed with a filtration $\mathbb{F} := (\mathcal{F}_n)_{n \in [p]}$. Let $\{\varepsilon_n\}_{n \in [p]^*}$ and $\{\varepsilon_n^0\}_{n \in [p]^*}$ be two independent sequences of i.i.d. random variables adapted to $\mathbb{F}$ and such that $(\varepsilon_{n+1}, \varepsilon_{n+1}^0)$ is independent of $\mathcal{F}_n$ for all $n \in [p-1]$. We denote by λ_ε the common distribution of the ε_n , $n \in [p]^*$.

Given $F : [p] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d \times \mathbb{A}) \times \mathbb{A} \times \mathbb{R}^K \times \mathbb{R}^{K'} \rightarrow \mathbb{R}^d$, we define the controlled process $X^\alpha : [p] \times \Omega \rightarrow \mathbb{R}^d$ by the dynamics:

$$X_{n+1}^\alpha = F_n(X_n^\alpha, \nu_n^{0,\alpha}, \alpha_n, \varepsilon_{n+1}, \varepsilon_{n+1}^0), \quad \nu_n^{0,\alpha} := \mathbb{P} \circ (X_n^\alpha, \alpha_n | \mathcal{F}_n^0)^{-1}, \quad (2.1)$$

where $X_0 \in \mathbb{L}^2(\Omega, \mathcal{F}_0, \mathbb{P})$, and $\mathcal{F}_n^0 := \sigma(\varepsilon_1^0, \dots, \varepsilon_n^0)$, $\mathcal{F}_0^0 := \{\emptyset, \Omega\}$. We also introduce:

$$\mu_n^{0,\alpha} := \mathbb{P} \circ (X_n^\alpha | \mathcal{F}_n^0)^{-1}, \quad n \in [p].$$

We denote by $\mathcal{A}$ the class of closed-loop controls of the form

$$\alpha_n = a_n(X_n^\alpha, \mu_n^{0,\alpha}), \quad a_n : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{A}, \quad n \in [p-1].$$

The law $\mu_n^{0,\alpha}$ is determined by the controls up to time $n-1$; it is therefore available before the choice of α_n .

Given Borel maps $f_n : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d \times \mathbb{A}) \times \mathbb{A} \rightarrow \mathbb{R}$, $n \in [p-1]$, and $g : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$, we consider the control problem

$$V_n(\mu) := \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\sum_{k=n}^{p-1} f_k(X_k^\alpha, \nu_k^{0,\alpha}, \alpha_k) + g(\mu_p^{0,\alpha}) \right], \quad \mu_n^{0,\alpha} = \mu \quad \text{a.s.} \quad (2.2)$$

The running rewards are assumed integrable under admissible controls, and all value functions considered below are finite.

Dynamic programming principle. We justify the existence of a dynamic programming principle for the control problem (2.2). Observe that the control problem (2.2) can be rewritten as the standard finite-horizon Markov Decision Process (MDP) problem:

$$V_0(\mu_0^\nu) = \sup_{(\nu_0, \dots, \nu_{p-1})} \mathbb{E} \left[\sum_{k=0}^{p-1} r_k(\nu_k) + g(\mu_p^\nu) \right], \quad (2.3)$$

with state process μ^ν driven by:

$$\begin{aligned} \mu_{n+1}^\nu &= \left((x, a, z) \mapsto F_n(x, \nu_n, a, z, \varepsilon_{n+1}^0) \right)_\# (\nu_n(dx, da) \otimes \lambda_\varepsilon(dz)) \\ &=: \Phi_n(\nu_n, \varepsilon_{n+1}^0), \end{aligned} \quad (2.4)$$

actions $\nu_n \in \Gamma(\mu_n^\nu)$, where $\Gamma(\mu) := \{(\text{Id}, a(\cdot, \mu))_\# \mu : a \in \mathbb{L}^0(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d), \mathbb{A})\}$, and

$$r_n(\nu) := \int_{\mathbb{R}^d \times \mathbb{A}} f_n(x, \nu, a) \nu(dx, da).$$

Then, under standard measurability and integrability assumptions, we have the Bellman equation for the MDP problem:

$$V_n(\mu) = \sup_{\nu \in \Gamma(\mu)} \left\{ r_n(\nu) + \mathbb{E} \left[V_{n+1}(\Phi_n(\nu, \varepsilon_{n+1}^0)) \right] \right\}.$$

By definition of $\Gamma(\mu)$, this yields immediately the dynamic programming principle:

$$V_n(\mu) = \sup_{a \in \mathbb{L}^0(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d), \mathbb{A})} \mathbb{E} \left[f_n(X_n^\alpha, \nu_n^\alpha, a(X_n, \mu)) + V_{n+1}(\Phi_n(\nu_n^\alpha, \varepsilon^0)) | X_n^\alpha, \varepsilon^0 \right], \quad (2.5)$$

where $\nu_n^\alpha = \mathbb{P} \circ (X_n^\alpha, a(X_n^\alpha, \mu) | \varepsilon^0)^{-1}$.

2.2 The N -particle problem in discrete time

For $N \in \mathbb{N}^*$, we define the N -particle problem:

$$V_n^N(\mathbf{x}) := \sup_{\alpha \in \mathcal{A}_N} \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \sum_{k=n}^{p-1} f_k(X_k^{i,\alpha}, \mu_{(X_k^\alpha, \alpha_k)}^N, \alpha_k^i) + g(\mu_{\mathbf{X}_p^\alpha}^N) \right], \quad (2.6)$$

where $\mathbf{X}^\alpha := (X^{\alpha,1}, \dots, X^{\alpha,N})$ is a solution to:

$$X_{m+1}^{\alpha,j} = F_m(X_m^{\alpha,j}, \mu_{(X_m^\alpha, \alpha_m)}^N, \alpha_m^j, \varepsilon_{m+1}^j, \varepsilon_{m+1}^0), \quad \mathbf{X}_n^\alpha = \mathbf{x}. \quad (2.7)$$

with $(\varepsilon^j)_{j \in [N]^*}$ i.i.d. copies of ε , independent of the common-noise sequence. Denote by $\mathcal{A}_N$ the set of Markovian feedback controls

$$\alpha_n = \mathbf{a}_n(\mathbf{X}_n^\alpha), \quad \mathbf{a}_n : (\mathbb{R}^d)^N \longrightarrow \mathbb{A}^N \quad \text{measurable}, \quad n \in [p-1].$$

2.3 Convergence rate

Assumption 2.1. (i) $\mathbf{x} \mapsto V_0^N(\mathbf{x})$ is continuous, and there exists a constant L such that:

$$\sup_{n \in [p]} |V_n(\mu) - V_n(\nu)| \leq L W_1(\mu, \nu) \quad \text{for all } \mu, \nu \in \mathcal{P}_1(\mathbb{R}^d).$$

(ii) There exists a constant $C \geq 0$, independent of N and p , such that, for every $N \geq 1$ and $\mathbf{x} \in (\mathbb{R}^d)^N$,

$$\sup_{\alpha \in \mathcal{A}_N} \frac{1}{N} \sum_{k=1}^N \mathbb{E} \left[\sup_{n \in [p]} |X_n^{k, \alpha}|^2 \right] \leq C(1 + m_2^N(\mathbf{x})),$$

where $m_2^N(\mathbf{x}) := N^{-1} \sum_{k=1}^N |x_k|^2$ and the particle system starts from $\mathbf{X}_0^\alpha = \mathbf{x}$.

(iii) For every n , every (x, μ, a, z) , the law of $F_n(x, \mu, a, \cdot, z)_\# \lambda_\varepsilon$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

Theorem 2.2. Let Assumption 2.1 hold. Then we have, for some constant $C \geq 0$:

$$|V_0(\mu_{\mathbf{x}}^N) - V_0^N(\mathbf{x})| \leq C(1 + m_2^N(\mathbf{x})) p R_N, \quad \text{for all } \mathbf{x} \in (\mathbb{R}^d)^N,$$

where:

$$R_N := \begin{cases} N^{-1/2}(\log(1+N))^{1/2} & \text{if } d = 1, \\ N^{-1/2}(\log(1+N))^{3/2} & \text{if } d = 2, \\ N^{-1/d} & \text{if } d \geq 3. \end{cases}$$

Proof. We write the proof for $f_n = 0$ to simplify the notation. Let us fix $\mathbf{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$. *Step 1:* We first assume that $\mathbf{x} \in \mathbb{X}_{\neq}^N := \{\mathbf{y} \in (\mathbb{R}^d)^N : y_i \neq y_j \text{ for all } i \neq j\}$. For each $\alpha \in \mathcal{A}_N$, we consider the particle system (2.7) starting from $\mathbf{X}_0^\alpha = \mathbf{x}$. Conditionally on $\sigma(\mathbf{X}_n^\alpha) \vee \mathcal{F}_{n+1}^0$, the positions at time $n+1$ are independent and their laws are absolutely continuous by Assumption 2.1(iii), implying that these positions coincide with probability zero. Thus, $\mathbf{X}_n^\alpha \in \mathbb{X}_{\neq}^N$ almost surely for all $n \in [p]$.

Now, for all $n \in [p]$ and $\mathbf{y} \in (\mathbb{R}^d)^N$, we define

$$\Delta_n^N(\mathbf{y}) := |V_n(\mu_{\mathbf{y}}^N) - V_n^N(\mathbf{y})|,$$

where V_n and V_n^N respectively correspond to the value functions (2.2) and (2.6) starting from time n . Let $n \in [p-1]$, and let α be an arbitrary admissible control prefix up to time $n-1$. For $\mathbf{a} \in \mathbb{A}^N$, we denote by $\alpha \oplus_n \mathbf{a}$ the process defined on $[n] \times \Omega$ by $(\alpha \oplus_n \mathbf{a})_m = \alpha_m \mathbf{1}_{m \leq n-1} + \mathbf{a} \mathbf{1}_{m=n}$. We also set:

$$\mathcal{G}_{n+1}^\alpha := \sigma(\mathbf{X}_n^\alpha) \vee \mathcal{F}_{n+1}^0.$$

Recall also the map Φ_n defined in (2.4). By (2.5) and the classical dynamic programming principle for (2.6), we have:

$$\begin{aligned} \Delta_n^N(\mathbf{X}_n^\alpha) &= \left| \sup_{a \in \mathbb{L}^0(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d), \mathbb{A})} \mathbb{E} \left[V_{n+1}(\Phi_n(\nu_n^a, \varepsilon_{n+1}^0)) \mid \mathbf{X}_n^\alpha \right] \right. \\ &\quad \left. - \sup_{\mathbf{a} \in \mathbb{A}^N} \mathbb{E} \left[V_{n+1}^N(\mathbf{X}_{n+1}^{\alpha \oplus_n \mathbf{a}}) \mid \mathbf{X}_n^\alpha \right] \right|, \end{aligned} \tag{2.8}$$

with, for $a \in \mathbb{L}^0(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d), \mathbb{A})$, $\nu_n^a := (\text{Id}, a(\cdot, \mu_{\mathbf{X}_n^\alpha}^N))_\# \mu_{\mathbf{X}_n^\alpha}^N$. We also write:

$$\mathbf{a}(\mathbf{y}) := (a(y_1, \mu_{\mathbf{y}}^N), \dots, a(y_N, \mu_{\mathbf{y}}^N)).$$

Then:

$$\nu_n^a = \frac{1}{N} \sum_{k=1}^N \delta_{(X_n^{k,\alpha}, a(X_n^{k,\alpha}, \mu_{\mathbf{X}_n^\alpha}^N))},$$

and the definition of Φ_n gives

$$\Phi_n(\nu_n^a, \varepsilon_{n+1}^0) = \frac{1}{N} \sum_{k=1}^N \mathbb{P} \circ \left(X_{n+1}^{k, \alpha \oplus_n a} \middle| \mathcal{G}_{n+1}^\alpha \right)^{-1}.$$

Since $\mathbf{X}_n^\alpha \in \mathbb{X}_{\neq}^N$ almost surely, every vector $(a_1, \dots, a_N) \in \mathbb{A}^N$ can be realised, at each fixed configuration, by setting $a(X_n^{k,\alpha}, \mu_{\mathbf{X}_n^\alpha}^N) = a_k$. Hence, setting

$$\mu_{n+1}^{0,\alpha,a} := \frac{1}{N} \sum_{k=1}^N \mathbb{P} \circ \left(X_{n+1}^{k, \alpha \oplus_n a} \middle| \mathcal{G}_{n+1}^\alpha \right)^{-1},$$

we may unify the two suprema in (2.8) and obtain:

$$\begin{aligned} \Delta_n^N(\mathbf{X}_n^\alpha) &\leq \sup_{a \in \mathbb{A}^N} \mathbb{E} \left[|V_{n+1}(\mu_{n+1}^{0,\alpha,a}) - V_{n+1}^N(\mathbf{X}_{n+1}^{\alpha \oplus_n a})| \middle| \mathbf{X}_n^\alpha \right] \\ &\leq \sup_{a \in \mathbb{A}^N} \mathbb{E} \left[L\mathcal{W}_1(\mu_{n+1}^{0,\alpha,a}, \mu_{\mathbf{X}_{n+1}^{\alpha \oplus_n a}}^N) + \Delta_{n+1}^N(\mathbf{X}_{n+1}^{\alpha \oplus_n a}) \middle| \mathbf{X}_n^\alpha \right], \end{aligned}$$

where we used Assumption 2.1(i). Conditionally on $\mathcal{G}_{n+1}^\alpha$, the next particle positions are independent, with average conditional law $\mu_{n+1}^{0,\alpha,a}$. Applying the second-moment estimate (A.1) of Lemma A.1, with $q = 1$ and $q' = 2$, and then the tower property, gives

$$\Delta_n^N(\mathbf{X}_n^\alpha) \leq \sup_{a \in \mathbb{A}^N} \mathbb{E} \left[CM_2^{1/2}(\mu_{n+1}^{0,\alpha,a}) R_N + \Delta_{n+1}^N(\mathbf{X}_{n+1}^{\alpha \oplus_n a}) \middle| \mathbf{X}_n^\alpha \right], \quad (2.9)$$

where:

$$M_2(\mu_{n+1}^{0,\alpha,a}) := \int_{\mathbb{R}^d} |x|^2 \mu_{n+1}^{0,\alpha,a}(dx) = \frac{1}{N} \sum_{k=1}^N \mathbb{E} \left[|X_{n+1}^{k, \alpha \oplus_n a}|^2 \middle| \mathcal{G}_{n+1}^\alpha \right].$$

Fix now $\delta > 0$. Let $\mathbf{a}_n^\delta : (\mathbb{R}^d)^N \rightarrow \mathbb{A}^N$ be a measurable map which is δ/p -optimal for the entire supremum in (2.9). Write $\mathbf{a}^\delta := \mathbf{a}_n^\delta(\mathbf{X}_n^\alpha)$. Taking expectations, we obtain

$$\mathbb{E}[\Delta_n^N(\mathbf{X}_n^\alpha)] \leq CR_N \mathbb{E}[M_2^{1/2}(\mu_{n+1}^{0,\alpha,\mathbf{a}^\delta})] + \mathbb{E}[\Delta_{n+1}^N(\mathbf{X}_{n+1}^{\alpha \oplus_n \mathbf{a}^\delta})] + \frac{\delta}{p}.$$

By Jensen's inequality, the tower property, and Assumption 2.1(ii), we obtain:

$$\begin{aligned} \mathbb{E}[M_2^{1/2}(\mu_{n+1}^{0,\alpha,\mathbf{a}^\delta})] &\leq \left(\frac{1}{N} \sum_{k=1}^N \mathbb{E}[|X_{n+1}^{k, \alpha \oplus_n \mathbf{a}^\delta}|^2] \right)^{1/2} \\ &\leq C(1 + m_2^N(\mathbf{x}))^{1/2} \leq C(1 + m_2^N(\mathbf{x})). \end{aligned}$$

where C denotes in each inequality a generic nonnegative constant, independent of p and N . Therefore,

$$\mathbb{E}[\Delta_n^N(\mathbf{X}_n^\alpha)] \leq C(1 + m_2^N(\mathbf{x})) R_N + \mathbb{E}[\Delta_{n+1}^N(\mathbf{X}_{n+1}^{\alpha \oplus_n \mathbf{a}^\delta})] + \frac{\delta}{p}. \quad (2.10)$$

Iterating this construction, and observing that $\Delta_p^N = 0$, we obtain by backward induction:

$$\Delta_0^N(\mathbf{x}) \leq C(1 + m_2^N(\mathbf{x}))pR_N + \sum_{n=0}^{p-1} \frac{\delta}{p} = C(1 + m_2^N(\mathbf{x}))pR_N + \delta.$$

Letting $\delta \rightarrow 0$ yields

$$\Delta_0^N(\mathbf{x}) \leq C(1 + m_2^N(\mathbf{x}))pR_N.$$

Step 2: We now assume $\mathbf{x} \notin \mathbb{X}_{\neq}^N$. Then, for η arbitrarily small, we may construct $\mathbf{x}^\eta$ such that $\mathbf{x}^\eta \in \mathbb{X}_{\neq}^N$ and $|\mathbf{x} - \mathbf{x}^\eta| \leq \eta$. Then, applying the result of Step 1, we have:

$$\Delta_0^N(\mathbf{x}^\eta) \leq C(1 + m_2^N(\mathbf{x}^\eta))pR_N.$$

Letting $\eta \rightarrow 0$ gives the desired inequality, as m_2^N and V_0^N are continuous, while $\mathcal{W}_1(\mu_{\mathbf{x}^\eta}^N, \mu_{\mathbf{x}}^N) \leq N^{-1} \sum_{k=1}^N |x_k^\eta - x_k| \rightarrow 0$ and V_0 is Lipschitz by Assumption 2.1(i). ■

2.4 Extension to the path-dependent case

We now extend our results to the case where the controlled dynamics are not Markovian. For $n \in [p]$, set

$$E_n := (\mathbb{R}^d)^{n+1}, \quad |\omega|_{\infty, n} := \max_{0 \leq k \leq n} |\omega_k|.$$

Denote by $\mathcal{W}_{1, \infty, n}$ the Wasserstein distance induced by $|\cdot|_{\infty, n}$ and, for $m \in \mathcal{P}_2(E_n)$, set

$$M_{2, \infty, n}(m) := \int_{E_n} |\omega|_{\infty, n}^2 m(d\omega).$$

For $\omega = (\omega^1, \dots, \omega^N) \in E_n^N$, define

$$m_2^N(\omega) := \frac{1}{N} \sum_{k=1}^N |\omega^k|_{\infty, n}^2 = M_{2, \infty, n}(\mu_\omega^N).$$

We also set

$$E_{n, \neq}^N := \left\{ (\omega^1, \dots, \omega^N) \in E_n^N : \omega^i \neq \omega^j \text{ for all } i \neq j \right\}.$$

For every $n \in [p-1]$, let $F_n : E_n \times \mathcal{P}_1(E_n \times \mathbb{A}) \times \mathbb{A} \times \mathbb{R}^K \times \mathbb{R}^{K'} \rightarrow \mathbb{R}^d$. Let X^α be defined by the dynamics:

$$X_{n+1}^\alpha = F_n(X_{\cdot \wedge n}^\alpha, \nu_n^{0, \alpha}, \alpha_n, \varepsilon_{n+1}, \varepsilon_{n+1}^0), \quad \nu_n^{0, \alpha} = \mathbb{P} \circ (X_{\cdot \wedge n}^\alpha, \alpha_n | \varepsilon_\cdot^0)^{-1}, \quad (2.11)$$

where $X_0 \in \mathbb{L}^2(\Omega, \mathcal{F}_0, \mathbb{P})$. Also set:

$$m_n^{0, \alpha} := \mathbb{P} \circ (X_{\cdot \wedge n}^\alpha | \varepsilon_\cdot^0)^{-1}, \quad n \in [p].$$

In this subsection, $\mathcal{A}$ denotes the class of feedback controls:

$$\alpha_n = a_n(X_{\cdot \wedge n}^\alpha, m_n^{0, \alpha}), \quad a_n : E_n \times \mathcal{P}_1(E_n) \longrightarrow \mathbb{A}, \quad n \in [p-1].$$

Given Borel running rewards $f_k : E_k \times \mathcal{P}_1(E_k \times \mathbb{A}) \times \mathbb{A} \rightarrow \mathbb{R}$ and $g : \mathcal{P}_1(E_p) \rightarrow \mathbb{R}$, the value function of the path-dependent mean field control problem is

$$V_n(m) := \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\sum_{k=n}^{p-1} f_k(X_{\cdot \wedge k}^\alpha, \nu_k^{0, \alpha}, \alpha_k) + g(m_p^{0, \alpha}) \right], \quad m_n^{0, \alpha} = m \quad \text{a.s.} \quad (2.12)$$

For $a \in \mathbb{L}^0(E_n \times \mathcal{P}_1(E_n), \mathbb{A})$, let

$$\begin{aligned}\nu^{m,a} &:= (\text{Id}, a(\cdot, m))_{\#} m, \\ \Phi_n(\nu^{m,a}, z^0) &:= \left((\omega, a, z) \mapsto \omega \oplus_{n+1} F_n(\omega, \nu^{m,a}, a, z, z^0) \right)_{\#} (\nu^{m,a}(d\omega, da) \otimes \lambda_{\varepsilon}(dz)),\end{aligned}$$

where $\omega \oplus_{n+1} x := (\omega_0, \dots, \omega_n, x) \in E_{n+1}$ for $\omega \in E_n$. Therefore:

$$m_{n+1}^{0,\alpha} = \Phi_n(\nu^{m,a}, \varepsilon_{n+1}^0),$$

and, as in the Markovian case, the following dynamic programming principle holds:

$$V_n(m) = \sup_{a \in \mathbb{L}^0(E_n \times \mathcal{P}_1(E_n), \mathbb{A})} \mathbb{E} \left[f_n(X_{\cdot \wedge n}^{\alpha}, \nu^{m,a}, a(X_{\cdot \wedge n}, m)) + V_{n+1}(\Phi_n(\nu^{m,a}, \varepsilon^0)) | X_{\cdot \wedge n}^{\alpha}, \varepsilon^0 \right]. \quad (2.13)$$

Again, we define the corresponding N -particle problem:

$$V_n^N(\omega) := \sup_{\alpha \in \mathcal{A}_N} \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \sum_{k=n}^{p-1} f_k(X_{\cdot \wedge k}^{i,\alpha}, \mu_{(\mathbf{X}_{\cdot \wedge k}^{\alpha}, \alpha_k)}^N, \alpha_k^i) + g(\mu_{\mathbf{X}^{\alpha}}^N) \right] \quad (2.14)$$

where $\mathbf{X}^{\alpha} := (X^{\alpha,1}, \dots, X^{\alpha,N})$ is a solution to:

$$X_{t+1}^{\alpha,j} = F_t(X_{\cdot \wedge t}^{\alpha,j}, \mu_{(\mathbf{X}_{\cdot \wedge t}^{\alpha}, \alpha_t)}^N, \alpha_t^j, \varepsilon_{t+1}^j, \varepsilon_{t+1}^0), \quad \mathbf{X}_{\cdot \wedge n}^{\alpha} = \omega. \quad (2.15)$$

with $(\varepsilon^j)_{j \in [N]^*}$ i.i.d. copies of ε . Again, $\mathcal{A}_N$ is the set of closed-loop controls: for each n , $\alpha_n = \mathbf{a}_n(\mathbf{X}_{\cdot \wedge n}^{\alpha})$, where $\mathbf{a}_n : E_n^N \rightarrow \mathbb{A}^N$ is measurable.

Assumption 2.3. (i) $\omega \mapsto V_0^N(\omega)$ is continuous, and there exists a constant L such that:

$$|V_n(m) - V_n(m')| \leq L W_{1,\infty,n}(m, m') \quad \text{for all } n \in [p], \quad m, m' \in \mathcal{P}_1(E_n).$$

(ii) There exists a constant $C \geq 0$, independent of N and p , such that, for every $n \in [p]$, $N \geq 1$ and $\omega = (\omega^1, \dots, \omega^N) \in E_n^N$,

$$\sup_{\alpha \in \mathcal{A}_N} \frac{1}{N} \sum_{k=1}^N \mathbb{E} \left[|X_{\cdot \wedge k}^{k,\alpha}|_{\infty,p}^2 \right] \leq C(1 + m_2^N(\omega)),$$

with $\mathbf{X}_{\cdot \wedge n}^{\alpha} = \omega$.

Note that no absolute-continuity assumption is needed here: distinct stopped paths remain distinct after extension, even if their current positions coincide.

We now present our convergence rate for the discrete time non-Markovian problem. We assume without loss of generality that $p \geq 2$.

Theorem 2.4. Let Assumption 2.3 hold. Then we have, for some constant $C \geq 0$, for all $\omega \in E_0^N$:

$$|V_0(m_{\omega}^N) - V_0^N(\omega)| \leq C(1 + m_2^N(\omega)) p N^{-\frac{1}{d(p+1)}}.$$

Proof. As in the proof of Theorem 2.2, we write the argument for $f_k = 0$. We fix $\omega = (\omega^1, \dots, \omega^N) \in E_{0,\neq}^N$. For every $n \in [p]$ and every $\tilde{\omega} \in E_n^N$, we define

$$\Delta_n^N(\tilde{\omega}) := |V_n(\mu_{\tilde{\omega}}^N) - V_n^N(\tilde{\omega})|.$$

For any particle control $\alpha \in \mathcal{A}_N$, we consider the particle system starting from $\mathbf{X}_{\cdot \wedge 0}^{\alpha} = \omega$. Since the initial paths $\omega^1, \dots, \omega^N$ are pairwise distinct, we have $\mathbf{X}_{\cdot \wedge n}^{\alpha} \in E_{n,\neq}^N$ for all $n \in [p]$, almost surely, as two distinct paths in E_n remain distinct after being extended to paths in E_{n+1} .

Let $n \in [p-1]$, and let α be an arbitrary admissible control prefix up to time $n-1$. Set

$$\mathcal{G}_{n+1}^\alpha := \sigma(\mathbf{X}_{\cdot \wedge n}^\alpha) \vee \mathcal{F}_{n+1}^0.$$

By (2.13) and the classical dynamic programming principle for (2.14), we have:

$$\Delta_n^N(\mathbf{X}_{\cdot \wedge n}^\alpha) = \left| \sup_{a \in \mathbb{L}^0(E_n \times \mathcal{P}_1(E_n), \mathbb{A})} \mathbb{E}[V_{n+1}(\Phi_n(\nu_n^a, \varepsilon_{n+1}^0)) \mid \mathbf{X}_{\cdot \wedge n}^\alpha] - \sup_{\mathbf{a} \in \mathbb{A}^N} \mathbb{E}[V_{n+1}^N(\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}}) \mid \mathbf{X}_{\cdot \wedge n}^\alpha] \right|, \quad (2.16)$$

where $\nu_n^a := (\text{Id}, a(\cdot, \mu_{\mathbf{X}_{\cdot \wedge n}^N}^N))_{\#} \mu_{\mathbf{X}_{\cdot \wedge n}^\alpha}^N$. For $a \in \mathbb{L}^0(E_n \times \mathcal{P}_1(E_n), \mathbb{A})$, set

$$\mathbf{a}(\omega) := (a(\omega^1, \mu_\omega^N), \dots, a(\omega^N, \mu_\omega^N)).$$

Then

$$\nu_n^a = \frac{1}{N} \sum_{k=1}^N \delta_{(X_{\cdot \wedge n}^{k, \alpha}, a(X_{\cdot \wedge n}^{k, \alpha}, \mu_{\mathbf{X}_{\cdot \wedge n}^N}^N))},$$

and

$$\Phi_n(\nu_n^a, \varepsilon_{n+1}^0) = \frac{1}{N} \sum_{k=1}^N \mathbb{P} \circ \left(X_{\cdot \wedge n+1}^{k, \alpha \oplus_n \mathbf{a}(\mathbf{X}_{\cdot \wedge n}^\alpha)} \mid \mathcal{G}_{n+1}^\alpha \right)^{-1}.$$

Since $\mathbf{X}_{\cdot \wedge n}^\alpha \in E_{n, \neq}^N$ almost surely, any prescribed vector of actions can be realised by a measurable a on the finite set of distinct stopped paths. Thus, at each time n , the two action sets in (2.16) may be identified. For $\mathbf{a} \in \mathbb{A}^N$, we write

$$m_{n+1}^{0, \alpha, \mathbf{a}} := \frac{1}{N} \sum_{k=1}^N \mathbb{P} \circ \left(X_{\cdot \wedge n+1}^{k, \alpha \oplus_n \mathbf{a}} \mid \mathcal{G}_{n+1}^\alpha \right)^{-1}.$$

It then follows from (2.16) that

$$\begin{aligned} \Delta_n^N(\mathbf{X}_{\cdot \wedge n}^\alpha) &\leq \sup_{\mathbf{a} \in \mathbb{A}^N} \mathbb{E} \left[\left| V_{n+1}(m_{n+1}^{0, \alpha, \mathbf{a}}) - V_{n+1}^N(\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}}) \right| \mid \mathbf{X}_{\cdot \wedge n}^\alpha \right] \\ &\leq \sup_{\mathbf{a} \in \mathbb{A}^N} \mathbb{E} \left[L\mathcal{W}_{1, \infty, n+1}(m_{n+1}^{0, \alpha, \mathbf{a}}, m_{\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}}}^N) + \Delta_{n+1}^N(\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}}) \mid \mathbf{X}_{\cdot \wedge n}^\alpha \right]. \end{aligned}$$

Set $R_N := N^{-\frac{1}{d(p+1)}}$. By Assumption 2.3(i) and Lemma A.2, applied conditionally on $\mathcal{G}_{n+1}^\alpha$, we obtain

$$\Delta_n^N(\mathbf{X}_{\cdot \wedge n}^\alpha) \leq \sup_{\mathbf{a} \in \mathbb{A}^N} \mathbb{E} \left[CM_{2, \infty, n+1}^{1/2}(m_{n+1}^{0, \alpha, \mathbf{a}}) R_N + \Delta_{n+1}^N(\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}}) \mid \mathbf{X}_{\cdot \wedge n}^\alpha \right], \quad (2.17)$$

where we used the fact that the constant-extension map

$$\iota_{n+1, p}(\omega_0, \dots, \omega_{n+1}) := (\omega_0, \dots, \omega_{n+1}, \omega_{n+1}, \dots, \omega_{n+1})$$

is an isometry from $(E_{n+1}, |\cdot|_{\infty, n+1})$ into $(E_p, |\cdot|_{\infty, p})$. Lemma A.2, applied to the push-forward conditional laws under $\iota_{n+1, p}$, therefore gives the rate $N^{-1/[d(p+1)]}$ with a constant depending only on the physical dimension d . Moreover,

$$M_{2, \infty, p}((\iota_{n+1, p})_{\#} m) = M_{2, \infty, n+1}(m), \quad m \in \mathcal{P}_2(E_{n+1}),$$

so the moment appearing in Lemma A.2 is exactly the one associated with $|\cdot|_{\infty, n+1}$. In particular, the constant C in (2.17) is independent of p .

Thus,

$$\begin{aligned} M_{2, \infty, n+1}(m_{n+1}^{0, \alpha, a}) &= \int_{E_{n+1}} |\omega|_{\infty, n+1}^2 m_{n+1}^{0, \alpha, a}(d\omega) \\ &= \frac{1}{N} \sum_{k=1}^N \mathbb{E}[|X_{\cdot \wedge n+1}^{k, \alpha \oplus_n a}|_{\infty, n+1}^2 \mid \mathcal{G}_{n+1}^\alpha]. \end{aligned}$$

Fix $\delta > 0$. By the measurable-selection convention stated above, choose a measurable map $\mathbf{a}_n^\delta : E_n^N \rightarrow \mathbb{A}^N$ which is δ/p -optimal for the entire supremum in (2.17). Setting $\mathbf{a}^\delta := \mathbf{a}_n^\delta(\mathbf{X}_{\cdot \wedge n}^\alpha)$ and taking expectations gives

$$\begin{aligned} \mathbb{E}[\Delta_n^N(\mathbf{X}_{\cdot \wedge n}^\alpha)] &\leq CR_N \mathbb{E}[M_{2, \infty, n+1}^{1/2}(m_{n+1}^{0, \alpha, \mathbf{a}^\delta})] \\ &\quad + \mathbb{E}[\Delta_{n+1}^N(\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}^\delta})] + \frac{\delta}{p}. \end{aligned}$$

By Jensen's inequality, the tower property, and Assumption 2.3(ii), we obtain:

$$\begin{aligned} \mathbb{E}[M_{2, \infty, n+1}^{1/2}(m_{n+1}^{0, \alpha, \mathbf{a}^\delta})] &\leq \left(\frac{1}{N} \sum_{k=1}^N \mathbb{E}[|X_{\cdot \wedge n+1}^{k, \alpha \oplus_n \mathbf{a}^\delta}|_{\infty, n+1}^2] \right)^{1/2} \\ &\leq C(1 + m_2^N(\omega))^{1/2} \leq C(1 + m_2^N(\omega)). \end{aligned}$$

Consequently,

$$\mathbb{E}[\Delta_n^N(\mathbf{X}_{\cdot \wedge n}^\alpha)] \leq C(1 + m_2^N(\omega))R_N + \mathbb{E}[\Delta_{n+1}^N(\mathbf{X}_{\cdot \wedge n+1}^{\alpha \oplus_n \mathbf{a}^\delta})] + \frac{\delta}{p}. \quad (2.18)$$

As in the proof of Theorem 2.2, using the fact that $\Delta_p^N = 0$, we obtain:

$$\Delta_0^N(\omega) \leq C(1 + m_2^N(\omega))pR_N + \sum_{n=0}^{p-1} \frac{\delta}{p} = C(1 + m_2^N(\omega))pR_N + \delta.$$

Letting $\delta \rightarrow 0$, we conclude that

$$|V_0(m_\omega^N) - V_0^N(\omega)| \leq C(1 + m_2^N(\omega))pN^{-\frac{1}{d(p+1)}}.$$

The extension to $\omega \notin E_{0, \neq}^N$ follows by an argument similar to that in Step 2 of the proof of Theorem 2.2. ■

2.5 An explicit feedback example with a full path-law rate

We construct a path-dependent model in the full control classes of (2.12) and (2.14), with an explicit optimal feedback and a value gap of order $N^{-1/(p+1)}$ for fixed p .

Fix $p \geq 2$, set $D = p + 1$, and write U_D for the uniform law on $[0, 1]^D$. In the notation of (2.11)–(2.12) and their N -particle counterparts (2.15)–(2.14), take $d = K = K' = 1$, the action space $\mathbb{A} = [-A, A]$, idiosyncratic noises $\varepsilon_{n+1} = \xi_{n+1} \sim U[0, 1]$ independent across dates and particles, and the degenerate common noise $\varepsilon_{n+1}^0 = 0$. Let $B_n, \phi_n : E_n \rightarrow \mathbb{R}$ be nonanticipative maps whose supremum norms and Lipschitz constants are bounded uniformly in n and p , with $|\phi_n| \leq A$. For

example, both maps may be nonzero multiples of $\tanh(\max_{0 \leq k \leq n} \omega_k)$. For $\omega \in E_n$, $\nu \in \mathcal{P}_1(E_n \times \mathbb{A})$, $a \in \mathbb{A}$, and $z, z^0 \in \mathbb{R}$, the transition map is

$$\begin{aligned} F_n(\omega, \nu, a, z, z^0) &:= B_n(\omega) + a + z, \\ X_{n+1}^\alpha &= F_n(X_{\cdot \wedge n}^\alpha, \nu_n^{0, \alpha}, \alpha_n, \xi_{n+1}, 0) = B_n(X_{\cdot \wedge n}^\alpha) + \alpha_n + \xi_{n+1}, \\ X_{n+1}^{i, \alpha} &= B_n(X_{\cdot \wedge n}^{i, \alpha}) + \alpha_n^i + \xi_{n+1}^i, \quad i = 1, \dots, N. \end{aligned} \quad (2.19)$$

For the terminal reward, use the fixed set of integer resolutions

$$\mathcal{S} := \{ \lceil 2^{k/4} \rceil : k = 0, 1, 2, \dots \} = \{1, 2, 3, 4, 5, 6, 7, 8, 10, \dots\}.$$

For $q \in \mathcal{S}$, let $\mathcal{Q}_q$ be the cubes of side $h_q = 1/q$ in the regular partition of $[0, 1]^D$. If $Q = \prod_{\ell=0}^p [b_\ell, b_\ell + h_q]$, define the tent function on $\mathbb{R}^D$ by

$$\psi_{q, Q}(z) := \left(\min_{0 \leq \ell \leq p} \{z_\ell - b_\ell, b_\ell + h_q - z_\ell\} \right)_+, \quad H_D(\nu) := \sup_{q \in \mathcal{S}} \sum_{Q \in \mathcal{Q}_q} \left| \int \psi_{q, Q} d(\nu - U_D) \right|.$$

The functional H_D compares the joint path law with U_D using smoothed histograms. The tents replace bin indicators and downweight paths near cell boundaries; absolute values prevent cancellation between cells, and the supremum selects the largest discrepancy. The resolution set $\mathcal{S}$ is independent of N , contains all dyadic powers, and includes intermediate scales. Each signed sum $\sum_{Q \in \mathcal{Q}_q} s_Q \psi_{q, Q}$, $s_Q \in \{-1, 1\}$, is 1-Lipschitz for the maximum norm: the tents have disjoint interiors and vanish on their boundaries. Consequently,

$$|H_D(\nu) - H_D(\nu')| \leq \mathcal{W}_{1, \infty, p}(\nu, \nu'), \quad 0 \leq H_D(\nu) \leq \mathcal{W}_{1, \infty, p}(\nu, U_D). \quad (2.20)$$

The integral of one tent under U_D is

$$\int \psi_{q, Q} dU_D = \frac{h_q^{D+1}}{2(D+1)}. \quad (2.21)$$

For $\omega \in E_p$, define its residual path $R(\omega) \in \mathbb{R}^D$ by

$$R_0(\omega) = \omega_0, \quad R_{n+1}(\omega) = \omega_{n+1} - B_n(\omega_{0:n}) - \phi_n(\omega_{0:n}).$$

The states X_n take values in $\mathbb{R}$. Under the feedback $\alpha_n = \phi_n(X_{\cdot \wedge n})$, the residual path is $R(X_{0:p}) = (X_0, \xi_1, \dots, \xi_p)$ and lies in $[0, 1]^D$ when $X_0 \in [0, 1]$. Other controls may move residuals outside this cube, where the tents vanish. Since the tents are defined on all of $\mathbb{R}^D$, $H_D(R_{\#} m)$ is well defined for every admissible path law. Choose $\lambda > 0$ and $\rho > \lambda$. For $\omega \in E_n$, $\nu \in \mathcal{P}_1(E_n \times \mathbb{A})$, $a \in \mathbb{A}$, and $m \in \mathcal{P}_1(E_p)$, the rewards in (2.12) and (2.14) are

$$f_n(\omega, \nu, a) = -\rho |a - \phi_n(\omega)|, \quad g(m) = -\lambda H_D(R_{\#} m). \quad (2.22)$$

This is a calibration problem with costly interventions: B_n is a history-based prediction, ϕ_n the preferred action, and the later coordinates of R are the innovations inferred after subtracting both. The terminal reward compares the joint law of the initial position and inferred shocks with independent uniforms, while the running reward charges deviations from ϕ_n . Under the preferred feedback and a uniform X_0 , the residual law is exactly U_D ; the midpoint grid used below approximates this initial law. A particle controller might improve empirical calibration by changing actions, but $\rho > \lambda$ makes the intervention cost dominant. This is a stylised model in which the law dependence is in the terminal reward, and the value gap isolates path-law sampling.

In the notation of the framework, with no common noise, $\nu_n^{0,\alpha} = \mathcal{L}(X_{\cdot \wedge n}^\alpha, \alpha_n)$ and $m_n^{0,\alpha} = \mathcal{L}(X_{\cdot \wedge n}^\alpha)$, and $m_p^{0,\alpha} = \mathcal{L}(X_{0:p}^\alpha)$ in the notation above. The admissible controls are:

$$\alpha_n = a_n(X_{\cdot \wedge n}^\alpha, m_n^{0,\alpha}), \quad \mathbf{a}_n = \mathbf{a}_n(\mathbf{X}_{\cdot \wedge n}^\alpha),$$

where $a_n : E_n \times \mathcal{P}_1(E_n) \rightarrow \mathbb{A}$ and $\mathbf{a}_n : E_n^N \rightarrow \mathbb{A}^N$ are measurable. In particular, $\mathbf{a}_n$ may use the whole stopped-path configuration. For an initial law $m \in \mathcal{P}_2(E_0)$ and initial configuration $\mathbf{x} \in E_0^N$, with $\mathcal{L}(X_0^\alpha) = m$ and $\mathbf{X}_0^\alpha = \mathbf{x}$, the two problems specialise to

$$\begin{aligned} V_0(m) &= \sup_{\alpha \in \mathcal{A}} \left\{ -\rho \mathbb{E} \sum_{n=0}^{p-1} |\alpha_n - \phi_n(X_{\cdot \wedge n}^\alpha)| - \lambda H_D(R_\# \mathcal{L}(X_{0:p}^\alpha)) \right\}, \\ V_0^N(\mathbf{x}) &= \sup_{\alpha \in \mathcal{A}_N} \mathbb{E} \left[-\frac{\rho}{N} \sum_{i=1}^N \sum_{n=0}^{p-1} |\alpha_n^i - \phi_n(X_{\cdot \wedge n}^{i,\alpha})| - \lambda H_D(R_\# m_{\mathbf{X}^\alpha}^N) \right]. \end{aligned} \quad (2.23)$$

Proposition 2.5 (Full feedback optimiser and path-law exponent). *For the full control problems (2.23), with data (2.19) and (2.22), the feedback*

$$\alpha_n^*(\omega, m) = \phi_n(\omega), \quad \mathbf{a}_n^*(\boldsymbol{\omega}) = (\phi_n(\omega^1), \dots, \phi_n(\omega^N))$$

is optimal in the full mean-field and N -particle control classes, for every admissible initial law and every initial particle configuration. Assumption 2.3 holds with constants independent of p and N .

In particular, take the deterministic initial configuration $x_i = (i - \frac{1}{2})/N$, $1 \leq i \leq N$, and set

$$\bar{\nu}_N := m_{\mathbf{x}}^N \otimes U[0, 1]^p, \quad \hat{\nu}_N := \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \xi_1^i, \dots, \xi_p^i)}.$$

Then the value gap of Theorem 2.4 is exactly

$$\Delta_N(p) := V_0(m_{\mathbf{x}}^N) - V_0^N(\mathbf{x}) = \lambda \{ \mathbb{E}[H_D(\hat{\nu}_N)] - H_D(\bar{\nu}_N) \} \geq 0. \quad (2.24)$$

For $D = p + 1 \geq 3$, define

$$C_D^{\text{dense}} := \frac{4}{\sqrt{2D(D+1)}(1 - 2^{1-D/2})} + \frac{16}{D}, \quad c_p := \frac{\lambda e^{-5/3}}{8(p+2)}.$$

Whenever $N^{p/(p+1)} \geq 2(p+2)e^{5/3}$,

$$c_p N^{-1/(p+1)} \leq \Delta_N(p) \leq \lambda C_D^{\text{dense}} N^{-1/(p+1)}. \quad (2.25)$$

The upper bound holds for every $N \geq 1$. Moreover, $C_D^{\text{dense}} \leq C_3^{\text{dense}}$ for all $D \geq 3$, so the upper constant can be chosen independently of p .

The proof is given in Appendix B.1.

Under the optimal feedback, the residual paths are $(x_i, \xi_1^i, \dots, \xi_p^i)$. For $N = q^D$ and $q \in \mathcal{S}$, each path cube of side $h = 1/q$ has q^{D-1} particles whose initial positions lie in its first-coordinate interval. Each places its remaining coordinates in the cube with probability $q^{-(D-1)}$. The expected occupancy is one, and the probability of being empty is

$$(1 - q^{-(D-1)})^{q^{D-1}} \longrightarrow e^{-1} \quad \text{as } q \rightarrow \infty.$$

Each empty cube contributes $h^{D+1}/[2(D+1)]$ to the discrepancy. Summing over the q^D cubes gives an expected contribution of order $h = N^{-1/D}$, explaining the exponent. Retaining $n + 1$ residual

coordinates gives q^{n+1} cells with average occupancy N/q^{n+1} ; each additional coordinate divides this occupancy by q . The target U_D is independent of N , and the midpoint initial law contributes an $O(N^{-1})$ correction by (B.2). For fixed p , the proposition bounds $N^{1/(p+1)}\Delta_N(p)$ above and away from zero for sufficiently large N . The maximising resolution may vary with N .

Corollary B.1 in Appendix B gives computable bounds on $\Delta_N(p)$ for every $N \geq 1$ and $p \geq 2$, using finitely many resolutions and a deterministic bound for the remaining ones. Figure 5 compares them with Monte Carlo estimates.

3 Convergence rate for continuous time problems

We combine the discrete-time particle convergence rates of Theorems 2.2 and 2.4 with time-discretisation error bounds to obtain quantitative convergence rates for continuous-time control problems.

3.1 The path-dependent control problems

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

$$|\omega|_2 := \left(\int_0^T |\omega_s|^2 ds \right)^{1/2}, \quad |\omega|_\infty := \sup_{s \in [0, T]} |\omega_s|,$$

and define the Wasserstein distances:

$$\begin{aligned} \mathcal{W}_{1,2}(m, m') &:= \inf_{\gamma \in \Pi(m, m')} \int_{\mathcal{X} \times \mathcal{X}} |\omega - \omega'|_2 \gamma(d\omega, d\omega'), \\ \mathcal{W}_{1,\infty}(m, m') &:= \inf_{\gamma \in \Pi(m, m')} \int_{\mathcal{X} \times \mathcal{X}} |\omega - \omega'|_\infty \gamma(d\omega, d\omega'), \end{aligned}$$

Unless stated differently, the measurable structure of $\mathcal{P}_1(\mathcal{X})$ is associated with $\mathcal{W}_{1,\infty}$. Also note that:

$$\mathcal{W}_{1,2}(m, m') \leq \sqrt{T} \mathcal{W}_{1,\infty}(m, m'), \quad \mathcal{W}_{1,2}^2(\mu_\omega^N, \mu_{\omega'}^N) \leq \frac{1}{N} \sum_{i=1}^N |\omega^i - (\omega')^i|_2^2. \quad (3.1)$$

Here $\mathcal{W}_{1,\infty}$ is induced by $|\cdot|_\infty$.

We work on a probability space carrying a d -dimensional Brownian motion W and an initial sigma-field $\mathcal{G}$, independent of W . For each initial law $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, we take a $\mathcal{G}$ -measurable random variable X_0 with law μ , and require that $\mathcal{G}$ also contain a uniform random variable independent of X_0 . We denote by $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ the usual augmentation of $\mathcal{G} \vee \sigma(W_s : 0 \leq s \leq t)$, and by $\mathcal{A}$ the set of $\mathbb{F}$ -progressively measurable $\mathbb{A}$ -valued controls.

For $b : [0, T] \times \mathcal{X} \times \mathcal{P}_1(\mathcal{X}) \times \mathbb{A} \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathcal{X} \times \mathcal{P}_1(\mathcal{X}) \rightarrow \mathcal{S}_d$, consider the dynamics

$$X_t^\alpha = X_0 + \int_0^t b_s(X_{\cdot \wedge s}^\alpha, m_s^\alpha, \alpha_s) ds + \int_0^t \sigma_s(X_{\cdot \wedge s}^\alpha, m_s^\alpha) dW_s, \quad m_s^\alpha := \mathbb{P} \circ (X_{\cdot \wedge s}^\alpha)^{-1}, \quad (3.2)$$

for $\alpha \in \mathcal{A}$. Given $f : [0, T] \times \mathcal{X} \times \mathcal{P}_1(\mathcal{X}) \times \mathbb{A} \rightarrow \mathbb{R}$ and $g : \mathcal{P}_1(\mathcal{X}) \rightarrow \mathbb{R}$, we define the value function:

$$V_0(\mu) := \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_0^T f_t(X_{\cdot \wedge t}^\alpha, m_t^\alpha, \alpha_t) dt + g(m_T^\alpha) \right], \quad \mu = \mathbb{P} \circ X_0^{-1}. \quad (3.3)$$

The corresponding N -particle problem is defined by dynamics and value functions:

$$X_t^{i,\alpha} = x_i + \int_0^t b_s(X_{\cdot \wedge s}^{i,\alpha}, m_{\mathbf{X} \cdot \wedge s}^N, \alpha_s^i) ds + \int_0^t \sigma_s(X_{\cdot \wedge s}^{i,\alpha}, m_{\mathbf{X} \cdot \wedge s}^N) dW_s^i, \quad i \in [N]^*, \quad (3.4)$$

$$V_0^N(\mathbf{x}) := \sup_{\alpha \in \mathcal{A}_N} \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \int_0^T f_t(X_{\cdot \wedge t}^{i,\alpha}, m_{\mathbf{X}_{\cdot \wedge t}^N}^\alpha, \alpha_t^i) dt + g(m_{\mathbf{X}^N}^\alpha) \right]. \quad (3.5)$$

Here the W^i are independent d -dimensional Brownian motions and $\mathcal{A}_N$ consists of open-loop controls adapted to their joint filtration.

Assumption 3.1. *We assume the following:*

(i) $\mathbb{A}$ is compact.

(ii) b is continuous in $a \in \mathbb{A}$. Moreover, there is a constant $L \geq 0$ such that, for all $t, t', \omega, \omega', m, m', a$:

$$\begin{aligned} & |b_t(\omega, m, a) - b_{t'}(\omega', m', a)| + |\sigma_t(\omega, m) - \sigma_{t'}(\omega', m')| \\ & \leq L(|t - t'|^{1/4} + |\omega_{\cdot \wedge t} - \omega'_{\cdot \wedge t'}|_2 + \mathcal{W}_{1,2}(m, m')), \end{aligned}$$

$$\text{and } \sup_{t \leq T, a \in \mathbb{A}} (|b_t(0, \delta_0, a)| + \|\sigma_t(0, \delta_0)\|_F) \leq L.$$

(iii) f is continuous in a . Moreover, there is a constant $L \geq 0$ such that, for all $t, t', \omega, \omega', m, m', a$:

$$\begin{aligned} & |f_t(\omega, m, a) - f_{t'}(\omega', m', a)| \leq L(|t - t'|^{1/4} + |\omega_{\cdot \wedge t} - \omega'_{\cdot \wedge t'}|_2 + \mathcal{W}_{1,2}(m, m')), \\ & |g(m) - g(m')| \leq L\mathcal{W}_{1,2}(m, m'), \end{aligned}$$

$$\text{and } \sup_{a \in \mathbb{A}} |f_0(0, \delta_0, a)| + |g(\delta_0)| \leq L;$$

(iv) For every (t, ω, m) , the set

$$\mathcal{K}(t, \omega, m) := \left\{ \begin{pmatrix} b_t(\omega, m, a) \\ f_t(\omega, m, a) - \eta \end{pmatrix} : a \in \mathbb{A}, \eta \geq 0 \right\} \quad (3.6)$$

is closed and convex in $\mathbb{R}^{d+1}$.

3.2 Discrete time approximations

Fix a uniform grid $\pi = \{t_k = kh : 0 \leq k \leq p\}$, with $h = |\pi| = T/p$. We denote by $\mathcal{A}^\pi$ the controls in $\mathcal{A}$ which are constant on each interval $[t_k, t_{k+1})$. For $n \in [p]$, let $\ell_{\pi,n} : E_n \rightarrow \mathcal{X}$ be the affine interpolation through $t_0, \dots, t_n$, extended constantly on $[t_n, T]$, and set $\ell_\pi = \ell_{\pi,p}$. These maps are also applied componentwise to vectors of paths. For $\alpha^\pi \in \mathcal{A}^\pi$, consider the dynamics:

$$\begin{aligned} X_{t_{k+1}}^{\pi,\alpha} &= X_{t_k}^{\pi,\alpha} + h b_{t_k}(\ell_{\pi,k}(X^{\pi,\alpha}), m_{t_k}^{\pi,\alpha}, \alpha_{t_k}^\pi) \\ &\quad + \sigma_{t_k}(\ell_{\pi,k}(X^{\pi,\alpha}), m_{t_k}^{\pi,\alpha}) \Delta W_{t_{k+1}}, \quad m_{t_k}^{\pi,\alpha} := \mathbb{P} \circ (\ell_{\pi,k}(X^{\pi,\alpha}))^{-1}, \end{aligned} \quad (3.7)$$

where $\Delta W_{t_{k+1}} = W_{t_{k+1}} - W_{t_k}$. We then define the control problem, for $m \in \mathcal{P}_1(E_n)$:

$$U_n^\pi(m) := \sup_{\alpha^\pi \in \mathcal{A}^\pi} \mathbb{E} \left[h \sum_{k=n}^{p-1} f_{t_k}(\ell_{\pi,k}(X^{\pi,\alpha}), m_{t_k}^{\pi,\alpha}, \alpha_{t_k}^\pi) + g(m_T^{\pi,\alpha}) \right], \quad m_{t_n}^{\pi,\alpha} = m. \quad (3.8)$$

For the particle problem, let $\mathcal{A}_N^\pi$ be the set of elements of $\mathcal{A}_N$ constant on each $[t_k, t_{k+1})$, $k = 0, \dots, p-1$. For $\alpha^\pi \in \mathcal{A}_N^\pi$, consider the dynamics:

$$\begin{aligned} X_{t_{k+1}}^{i,\pi,\alpha} &= X_{t_k}^{i,\pi,\alpha} + h b_{t_k}(\ell_{\pi,k}(X^{i,\pi,\alpha}), \mu_{\ell_{\pi,k}(X^{\pi,\alpha})}^N, \alpha_{t_k}^{i,\pi}) \\ &\quad + \sigma_{t_k}(\ell_{\pi,k}(X^{i,\pi,\alpha}), \mu_{\ell_{\pi,k}(X^{\pi,\alpha})}^N) \Delta W_{t_{k+1}}^i. \end{aligned} \quad (3.9)$$

Here $\mathbf{X}^{\pi,\alpha} = (X^{1,\pi,\alpha}, \dots, X^{N,\pi,\alpha})$. We then define, for $\omega \in E_n^N$, the control problem:

$$U_n^{N,\pi}(\omega) := \sup_{\alpha^\pi \in \mathcal{A}_N^\pi} \mathbb{E} \left[\frac{h}{N} \sum_{i=1}^N \sum_{k=n}^{p-1} f_{t_k}(\ell_{\pi,k}(X^{i,\pi,\alpha}), \mu_{\ell_{\pi,k}(\mathbf{X}^{\pi,\alpha})}^N, \alpha_{t_k}^{i,\pi}) + g(m_{\ell_\pi(\mathbf{X}^{\pi,\alpha})}^N) \right], \quad (3.10)$$

where $\mathbf{X}_{\wedge n}^{\pi,\alpha} = \omega$.

The problems (3.8) and (3.10) use open-loop controls. Proposition D.1 shows that their value functions coincide with those obtained using suitable randomised feedback controls. Together with the regularity estimates in Appendix D, this allows us to apply the discrete-time comparison of Theorem 2.4 and obtain the following bound, with a constant uniform in the grid.

Lemma 3.2 (Particle convergence for the discrete time problems). *Let Assumption 3.1 hold. There is a constant $C \geq 0$, depending only on T , d and the constants in that assumption, such that, for every $N \geq 1$, every uniform grid with $p \geq 2$ and $h = T/p \leq 1$, and every $\mathbf{x} \in (\mathbb{R}^d)^N$,*

$$|U_0^{N,\pi}(\mathbf{x}) - U_0^\pi(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) p N^{-1/[d(p+1)]}. \quad (3.11)$$

In particular, C is independent of N , p and $\mathbf{x}$.

3.3 Main result

Theorem 3.3. *Let Assumption 3.1 hold. For every $N \geq 1$, every uniform grid with $p \geq 2$ and $|\pi| \leq 1$, and every $\mathbf{x} \in (\mathbb{R}^d)^N$, we have, for a constant C independent of π , N and $\mathbf{x}$:*

$$|V_0^N(\mathbf{x}) - V_0(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) (h^{1/4} + p N^{-1/[d(p+1)]}). \quad (3.12)$$

In particular, for every $N \geq 3$:

$$|V_0^N(\mathbf{x}) - V_0(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) R_N, \quad R_N := \left(\frac{\log(\log(N))}{\log(N)} \right)^{1/4}. \quad (3.13)$$

Proof. Lemma 3.2 gives the population estimate (3.11) for the time-discretised problems.

Lemma C.1 and Corollary C.2 give, respectively,

$$\begin{aligned} |V_0^N(\mathbf{x}) - U_0^{N,\pi}(\mathbf{x})| &\leq C(1 + m_2^N(\mathbf{x})) h^{1/4}, \\ |V_0(\mu) - U_0^\pi(\mu)| &\leq C(1 + m_2(\mu))^{1/2} h^{1/4}, \quad \text{for all } \mu \in \mathcal{P}_2(\mathbb{R}^d). \end{aligned}$$

The triangle inequality proves (3.12). For large N , take

$$p_N = \left\lfloor \frac{\log N}{2d \log \log N} \right\rfloor.$$

Then $p_N \geq \max\{2, T\}$ for sufficiently large N ,

$$p_N^{-1/4} = O\left(\left(\frac{\log \log N}{\log N}\right)^{1/4}\right), \quad p_N N^{-1/[d(p_N+1)]} = O\left(\frac{1}{\log N \log \log N}\right).$$

This proves (3.13). ■

Remark 3.4. For fixed N , refining the time grid decreases $h^{1/4} = (T/p)^{1/4}$ but increases the population bound $p N^{-1/[d(p+1)]}$ in (3.12). The factor p accounts for the accumulation of one-step errors, while the exponent reflects the dimension $d(p+1)$ of the discrete path space. The choice $p = p_N$ in the proof tends to infinity slowly enough that both terms vanish. Along this choice, the population term is of smaller order than the time-discretisation term, giving the logarithmic upper bound in (3.13).

3.4 The Markovian case

We now discuss how our convergence rates improve under Markovian structure. Assume that

$$b_t(\omega, m, a) = b_t(\omega_t, m_t, a), \quad \sigma_t(\omega, m) = \sigma_t(\omega_t, m_t), \quad f_t(\omega, m, a) = f_t(\omega_t, m_t, a), \quad g(m) = g(m_T).$$

We assume the Markovian analogue of Assumption 3.1, with the path distances replaced by the Euclidean distance and $\mathcal{W}_1$, and with the time regularity strengthened to

$$\begin{aligned} & |b_t(x, \mu, a) - b_s(x', \mu', a)| + \|\sigma_t(x, \mu) - \sigma_s(x', \mu')\|_F \\ & + |f_t(x, \mu, a) - f_s(x', \mu', a)| \leq C(|t - s|^{1/2} + |x - x'| + \mathcal{W}_1(\mu, \mu')). \end{aligned}$$

The compactness and Filippov convexity assumptions are unchanged. We also assume that the Euler transition kernels are absolutely continuous; this holds, for instance, if

$$\sigma_t(x, \mu) \sigma_t(x, \mu)^\top \geq \lambda I_d$$

for some $\lambda > 0$.

There are two differences with the path-dependent case. First, the state space of the discrete-time problem is now $\mathbb{R}^d$, independently of the number of time steps. The randomisation argument of Appendix D, with the current state in place of the stopped path, therefore allows us to apply Theorem 2.2. Hence

$$|U_0^{N,\pi}(\mathbf{x}) - U_0^\pi(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x}))pR_{N,d},$$

where

$$R_{N,d} := \begin{cases} N^{-1/2}(\log(1+N))^{1/2}, & d = 1, \\ N^{-1/2}(\log(1+N))^{3/2}, & d = 2, \\ N^{-1/d}, & d \geq 3. \end{cases}$$

Second, the time-discretisation estimate can be improved from order $|\pi|^{1/4}$ to order $|\pi|^{1/2}$. Indeed, in the Markovian setting, the state comparison may be performed directly at the grid points. If

$$e_k := \frac{1}{N} \sum_{i=1}^N \mathbb{E}[|X_{t_k}^{i,\alpha} - \bar{X}_{t_k}^{i,\pi}|^2],$$

then the path-dependent argument of Appendix C leads to

$$e_{k+1} \leq (1 + C|\pi|)e_k + C(1 + m_2^N(\mathbf{x}))|\pi|^{3/2} + C|\pi|^2 \sum_{\ell=0}^k e_\ell.$$

In the Markovian case, this is replaced by

$$e_{k+1} \leq (1 + C|\pi|)e_k + C(1 + m_2^N(\mathbf{x}))|\pi|^2,$$

as there is no longer any error coming from the reconstruction of the past trajectory. Moreover, the integrated centred drift residuals arising from the Filippov projection can be kept as discrete martingale increments: their conditional means vanish and their second moments are of order $|\pi|^2$. Together with the 1/2-Hölder regularity in time, this gives the local error of order $|\pi|^2$ above. Discrete Gronwall therefore yields

$$\max_{0 \leq k \leq p} e_k \leq C(1 + m_2^N(\mathbf{x}))|\pi|.$$

The Lipschitz continuity of the running and terminal rewards then gives a value error of order $|\pi|^{1/2}$. Together with the $1/2$ -Hölder regularity in time and the Lipschitz assumptions, this yields, uniformly in N ,

$$|V_0^N(\mathbf{x}) - U_0^{N,\pi}(\mathbf{x})| + |V_0(m_{\mathbf{x}}^N) - U_0^\pi(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/2}.$$

Therefore, we obtain

$$|V_0^N(\mathbf{x}) - V_0(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) \left(|\pi|^{1/2} + \frac{R_{N,d}}{|\pi|} \right).$$

Choosing h of order $R_{N,d}^{2/3}$ gives

$$|V_0^N(\mathbf{x}) - V_0(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) \begin{cases} N^{-1/6}(\log(1+N))^{1/6}, & d = 1, \\ N^{-1/6}(\log(1+N))^{1/2}, & d = 2, \\ N^{-1/(3d)}, & d \geq 3. \end{cases}$$

For $d \geq 3$, the exponent $1/(3d)$ is reminiscent of the dimension-dependent rates obtained in the Markovian mean field control literature. In particular, Daudin, Delarue and Jackson [10] obtain, up to an arbitrarily small loss in the exponent, the rate $N^{-2/(3d+6)}$ in a more structured setting: the state space is the torus $\mathbb{T}^d$, the idiosyncratic volatility is constant and non-degenerate, the control enters additively in the drift, and the Hamiltonian is uniformly convex in the momentum variable, while the interaction costs are assumed Lipschitz and semiconcave with respect to the 1-Wasserstein distance. Their rate is sharper than the one obtained here. On the other hand, the present argument is set on $\mathbb{R}^d$, allows for a state- and measure-dependent uncontrolled volatility and for more general drift and reward functions, and follows directly from the discrete-time empirical-measure estimate combined with a uniform-in- N Euler approximation.

4 Numerical experiments

We compare the mean field and particle values in three controlled examples: a scalar Markovian model, its common-noise extension, and the path-dependent model of Proposition 2.5. The first two illustrate the convergence in Theorem 2.2. In the third, we vary the number of time steps to examine the dependence of the convergence rate on path dimension in Theorem 2.4. Table 1 collects the simulation settings.

4.1 A controlled Markovian benchmark.

Let $p \in \mathbb{N}^*$, $h = T/p$, and consider the scalar controlled particle dynamics

$$X_{k+1}^i = X_k^i + h \left(-\kappa X_k^i + \lambda \bar{X}_k + a_k^i \right) + \sigma \sqrt{h} \xi_{k+1}^i, \quad \bar{X}_k := \frac{1}{N} \sum_{i=1}^N X_k^i,$$

where $a_k^i \in [-A, A]$, $A > 0$, $\sigma > 0$, and the ξ_k^i are independent standard Gaussian variables. We maximise the reward

$$\mathbb{E} \left[-\frac{ch}{N} \sum_{k=0}^{p-1} \sum_{i=1}^N |a_k^i| - |\bar{X}_p - \theta| \right].$$

Here $c \geq 0$ is the intervention cost. We take $X_0^i = 0$ for every particle and the same initial state in the mean field problem, where the empirical mean is replaced by the law mean. For this model and its common-noise extension, we compute the value gap

$$\Delta_N := |V_0(\mu_{\mathbf{x}_0}^N) - V_0^N(\mathbf{x}_0)|, \quad \mathbf{x}_0 = (0, \dots, 0).$$

Write $\bar{a}_k := N^{-1} \sum_{i=1}^N a_k^i$ and $Z_{k+1} := N^{-1/2} \sum_{i=1}^N \xi_{k+1}^i$. The Z_k are independent standard Gaussian variables. The scalar reduction is exact for the full feedback control class: the mean dynamics depend on the actions only through their average, while $N^{-1} \sum_i |a_k^i| \geq |\bar{a}_k|$. Every feasible average action can be implemented by giving it to all particles, so a one-dimensional Bellman recursion computes the value of the full particle control problem. The mean field reduction follows by the same convexity argument. Writing $r := 1 + h(\lambda - \kappa)$, the mean field mean satisfies

$$m_{k+1} = rm_k + ha_k,$$

whereas the N -particle empirical mean satisfies

$$\bar{X}_{k+1} = r\bar{X}_k + h\bar{a}_k + \sigma\sqrt{\frac{h}{N}} Z_{k+1}.$$

The next innovation Z_{k+1} is independent of the full particle history up to time k . Backward induction in the scalar Bellman recursion therefore bounds the reward of every particle policy from above. Equality is attained by applying a maximising scalar feedback to all particles.

The controller regulates the population mean towards the terminal target, balancing the tracking error against the cost of intervention. For the chosen zero target, the mean field can stay exactly at the target without acting, whereas the finite population retains sampling fluctuations. The following calculation explains their effect on the optimal value.

Proposition 4.1 (Rate and scaling of the Markovian benchmark). *For $\theta = 0$, $c \geq 0$ and $\sigma > 0$, with all particles initially at zero, $V_0(\delta_0) = 0$ and, for every $N \geq 1$,*

$$\sigma\sqrt{\frac{2h}{\pi N}} \leq \Delta_N \leq \sigma\sqrt{\frac{2h}{\pi N} \sum_{j=0}^{p-1} r^{2j}}. \quad (4.1)$$

For fixed model parameters, $\sqrt{N}\Delta_N$ is nonincreasing in N and converges to a finite, strictly positive limit.

Proof. All rewards are nonpositive, and zero control keeps the mean field mean at zero, proving $V_0(\delta_0) = 0$. Conditional on the particle history at time $p-1$, the terminal empirical mean is $y + \sigma\sqrt{h/N} Z_p$, where $y = r\bar{X}_{p-1} + h\bar{a}_{p-1}$ is known and Z_p is an independent standard Gaussian. Symmetry and convexity show that $\mathbb{E}|y + \sigma\sqrt{h/N} Z_p|$ is minimised at $y = 0$. Nonnegativity of the running cost therefore gives the lower bound in (4.1), for every admissible feedback policy. Zero control gives a centred Gaussian terminal mean with variance $(\sigma^2 h/N) \sum_{j=0}^{p-1} r^{2j}$, proving the upper bound.

For the scaling assertion, define $\mathcal{J}_p(B)$ as the infimum of

$$\mathbb{E} \left[ch \sum_{k=0}^{p-1} |b_k| + |Y_p| \right], \quad Y_{k+1} = rY_k + hb_k + \sqrt{h} Z_{k+1}, \quad Y_0 = 0,$$

over scalar feedback controls with $|b_k| \leq B$, where the Z_k are independent standard Gaussians. The exact scalar reduction and the rescaling $Y_k = \sqrt{N}\bar{X}_k/\sigma$, $b_k = \sqrt{N}\bar{a}_k/\sigma$ give

$$\sqrt{N}\Delta_N = \sigma\mathcal{J}_p\left(\frac{A\sqrt{N}}{\sigma}\right). \quad (4.2)$$

The admissible control sets increase with B , so $\mathcal{J}_p(B)$ is nonincreasing. The positive lower bound and finite upper bound in (4.1) then give the stated limit. $\blacksquare$

We compute the scalar value functions on a state grid, using linear interpolation and analytic integration against the Gaussian transition. The optimal action moves the uncontrolled next mean towards an interval where zero control is optimal, subject to the action bound. Thus no action grid or numerical Gaussian quadrature is needed. In the left panel of Figure 2, the solid curve shows the resulting value gaps, and the markers evaluate the computed policies by independent Monte Carlo sampling. The remaining numerical approximation is the state grid. For $N \in \{16, 64, 512, 2048\}$, the finest grid comparisons change the gaps by less than 0.001%; doubling the state domain at fixed spacing changes them by less than 0.00001%.

These bounds establish the $N^{-1/2}$ exponent for this benchmark, without the logarithmic loss permitted by the general estimate of Theorem 2.2. The fitted slope -0.513 in Table 2 is consistent with (4.2): as N grows, the fixed action bound becomes less restrictive relative to the sampling fluctuations. Over the tested range, $\sqrt{N}\Delta_N$ decreases from approximately 0.304 to 0.282, while zero control gives the constant 0.6995. The feedback therefore reduces the total expected loss, including the running cost, by approximately 57–60% relative to zero control. The Monte Carlo estimates and dynamic-programming values differ by at most 0.05% in these runs.

4.2 A controlled common-noise benchmark.

We add common noise of amplitude $\sigma_0 \geq 0$ to the Markovian model, using independent standard Gaussian innovations ξ_k^0 shared by all particles and independent of the idiosyncratic noises:

$$X_{k+1}^i = X_k^i + h \left(-\kappa X_k^i + \lambda \bar{X}_k + a_k^i \right) + \sigma \sqrt{h} \xi_{k+1}^i + \sigma_0 \sqrt{h} \xi_{k+1}^0.$$

Writing M_k for the population mean conditional on the common noise, we obtain

$$M_{k+1} = rM_k + ha_k + \sigma_0 \sqrt{h} \xi_{k+1}^0,$$

whereas the particle empirical mean satisfies

$$\bar{X}_{k+1} = r\bar{X}_k + h\bar{a}_k + \sigma_0 \sqrt{h} \xi_{k+1}^0 + \sigma \sqrt{\frac{h}{N}} Z_{k+1}.$$

The reward and initial state are unchanged. The same scalar reduction applies, with noise amplitudes σ_0 for the mean field problem and $\sqrt{\sigma_0^2 + \sigma^2/N}$ for the particle problem. When $\sigma_0 = 0$, we recover the preceding Markovian model. The following estimate explains why positive common noise can give faster convergence.

Proposition 4.2 (Noise-amplitude comparison). *For $\nu \geq 0$, let $w_k^\nu(x)$ be the scalar value functions defined by*

$$\begin{aligned} w_p^\nu(x) &= -|x - \theta|, \\ w_k^\nu(x) &= \sup_{|a| \leq A} \left\{ -ch|a| + \mathbb{E} \left[w_{k+1}^\nu(rx + ha + \nu\sqrt{h}Z) \right] \right\}, \end{aligned} \tag{4.3}$$

where Z is standard Gaussian and the recursion holds for $k = p-1, \dots, 0$. For $\nu, \tilde{\nu} \geq 0$,

$$\sup_x |w_0^\nu(x) - w_0^{\tilde{\nu}}(x)| \leq \Gamma_p |\nu - \tilde{\nu}|, \quad \Gamma_p := \sqrt{\frac{2h}{\pi}} \sum_{j=0}^{p-1} |r|^j. \tag{4.4}$$

In particular, with $\nu_N := \sqrt{\sigma_0^2 + \sigma^2/N}$, the full value gap of the common-noise benchmark satisfies, whenever $\sigma_0 > 0$,

$$\Delta_N \leq \Gamma_p(\nu_N - \sigma_0) = \frac{\Gamma_p \sigma^2}{N(\nu_N + \sigma_0)} \leq \frac{\Gamma_p \sigma^2}{2\sigma_0 N}. \tag{4.5}$$

Proof. Using the same action and Gaussian innovation for two initial states in (4.3), backward induction gives

$$|w_k^\nu(x) - w_k^\nu(y)| \leq |r|^{p-k}|x - y|,$$

uniformly in ν . Now set $e_k := \sup_x |w_k^\nu(x) - \tilde{w}_k^\nu(x)|$, so $e_p = 0$. Compare the two Bellman expressions at the same state and action, and couple their transitions with the same standard Gaussian Z . The preceding Lipschitz estimate gives

$$e_k \leq e_{k+1} + |r|^{p-k-1}\sqrt{h}|\nu - \tilde{\nu}|\mathbb{E}|Z| = e_{k+1} + |r|^{p-k-1}\sqrt{\frac{2h}{\pi}}|\nu - \tilde{\nu}|.$$

The running costs cancel in this comparison. Taking the suprema over actions preserves the inequality, since the action sets coincide. Summing over k proves (4.4). The scalar reductions identify the mean field and particle values with $w_0^{\sigma_0}(0)$ and $w_0^{\nu_N}(0)$, respectively. Substituting these amplitudes and rationalising their difference proves (4.5). $\blacksquare$

The additional empirical noise has standard deviation of order $N^{-1/2}$, but in the presence of a fixed positive Gaussian background it changes the total scalar noise amplitude by only order N^{-1} . Proposition 4.2 transfers this smaller perturbation to the optimised value over all p steps. The bound is specific to this scalar reduction and is not uniform as $\sigma_0 \downarrow 0$.

For $p = 1$, $h = 1$ and $\theta = 0$, symmetry and convexity of the terminal loss, together with the nonnegative intervention cost, make zero control optimal. The two terminal means are centred Gaussians with standard deviations σ_0 and $\sqrt{\sigma_0^2 + \sigma^2/N}$, so

$$\Delta_N^{(1)} = \sqrt{\frac{2}{\pi}} \left(\sqrt{\sigma_0^2 + \frac{\sigma^2}{N}} - \sigma_0 \right).$$

For $\sigma_0 = 0$, this equals $\sigma\sqrt{2/(\pi N)}$. For fixed $\sigma_0 > 0$, it is $\sqrt{2/\pi}\sigma^2/(2\sigma_0 N) + O(N^{-2})$. In this one-step example, positive common noise spreads the terminal mean around the point 0, where the absolute-value loss is not differentiable. The transition between the two rates occurs around $N = \sigma^2/\sigma_0^2$. For $\sigma = 1$ and $\sigma_0 = 0.25, 0.5, 1$, these scales are 16, 4, 1, respectively.

For the numerical experiment, we compute both scalar value functions and evaluate their policies by Monte Carlo sampling. We use the same standard Gaussian innovations in the two simulations, multiplied by their respective noise amplitudes. This preserves both marginal laws and reduces the variance of their difference.

The right panel of Figure 2 shows the resulting gaps. For $\sigma_0 = 0.25, 0.5, 1$, the fitted slopes are $-0.979, -0.994, -0.998$, respectively (Table 2). The closer agreement with -1 at larger common-noise amplitudes is consistent with the one-step transition scales above. For fixed p and $\sigma_0 > 0$, (4.5) gives an $O(N^{-1})$ upper bound. The Monte Carlo gaps differ from those computed by the Bellman recursion by at most 0.39%.

4.3 An explicit full-feedback path-law value gap.

We now consider the path-dependent model of Proposition 2.5, with $\lambda = 1$ and midpoint initial positions. The proposition proves that $a_n^* = \phi_n(X_{0:n})$ is optimal over all adapted particle controls. We therefore estimate the value gap $\Delta_N(p)$ in (2.24) by simulating the corresponding residual paths. For each sample of N paths, the tent discrepancy is evaluated exactly at each visited resolution, and a deterministic tail bound certifies the supremum over the remaining scales. The correction $H_D(\bar{\nu}_N)$ is computed analytically.

Figure 3 uses $p \in \{2, 4, 6, 8\}$, $N = 2^7, \dots, 2^{18}$, and 50 independent repetitions per point. Fits over the full N range have slopes $-0.333, -0.198, -0.138, -0.102$, respectively, compared with the

Experiment	Main parameters	Reference and Monte Carlo setup
Controlled Markovian	$T = 1, p = 12, \kappa = 0.45, \lambda = 0.15, \sigma = 1, A = 1.2, c = 0.08, \theta = 0$	Scalar DP with 51,201 state points on $[-1.6, 1.6]$, analytic Gaussian integration and continuous actions; $N \in \{16, 32, \dots, 2048\}$; 5 seeds, 10^5 samples/seed
Controlled common noise	Same controlled model, with $\sigma_0 \in \{0, 0.25, 0.5, 1\}$	Scalar DP with 102,401 state points on $[-3.2, 3.2]$, analytic Gaussian integration and continuous actions; $N \in \{16, 32, \dots, 2048\}$; 5 seeds, 10^5 samples/seed
Full-feedback path-law	$\lambda = 1, p \in \{2, 4, 6, 8\}, D = p + 1$, midpoint initial grid	Analytic optimiser and midpoint-law correction; certified tent supremum; $N = 2^7, \dots, 2^{18}$; 50 repetitions per point

Table 1: Parameters for the three controlled experiments. The two Markovian finite- N references use scalar dynamic programming; the full-feedback path-law gap uses the analytic optimiser and a Monte Carlo expectation of the terminal discrepancy.

Experiment	Parameter	Reference slope	Monte Carlo slope
Controlled Markovian	–	–0.513	–0.513
Controlled common noise	$\sigma_0 = 0$	–0.513	–0.513
Controlled common noise	$\sigma_0 = 0.25$	–0.979	–0.979
Controlled common noise	$\sigma_0 = 0.5$	–0.994	–0.994
Controlled common noise	$\sigma_0 = 1$	–0.998	–0.998

Table 2: Least-squares slopes of the Markovian log–log value-gap plots over $N \in \{16, 32, \dots, 2048\}$. Reference slopes are computed from the solid curves in Figure 2; Monte Carlo slopes are computed from the markers.

fixed- p guides $-1/(p+1) = -0.333, -0.200, -0.143, -0.111$. For $p = 8$, the scaled means in the right panel range from approximately 0.062 to 0.071, and the fitted slope depends on the chosen range of N . These results illustrate the proved $N^{-1/(p+1)}$ rate for each fixed p .

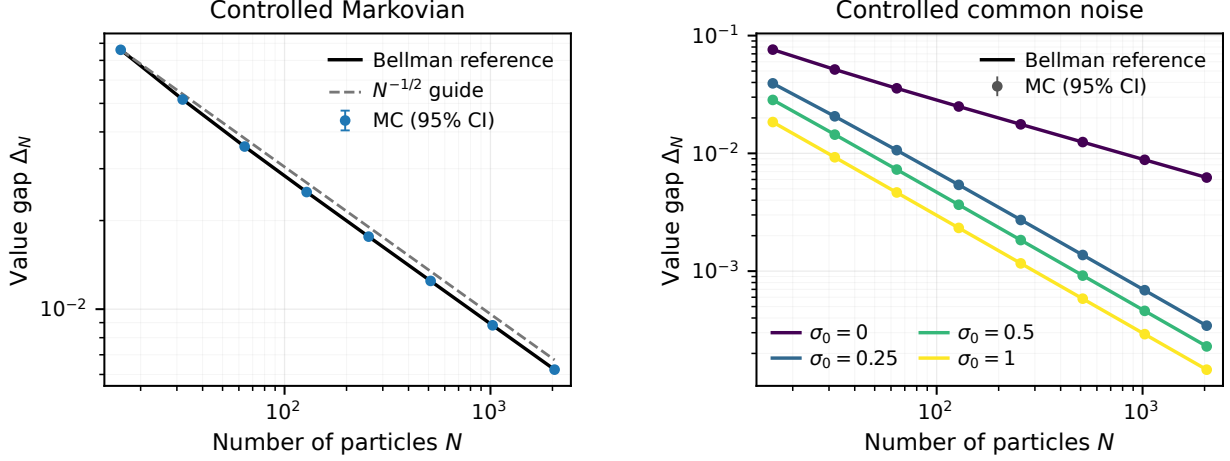
Figure 4 helps explain the variation of the scaled gap. At resolution q , the average number of particles per cell is N/q^D . Doubling q divides this number by 2^D , which is 512 when $p = 8$. The intermediate resolutions in $\mathcal{S}$ allow the maximising cell size to change in smaller increments as N grows.

The left panel shows these changes at $p = 8$. Across the 50 repetitions, $q = 2$ is selected at $N = 128$, $q = 3$ at $N = 2^9, \dots, 2^{12}$, $q = 4$ at $N = 2^{13}, \dots, 2^{16}$, and $q = 5$ at $N = 2^{17}, 2^{18}$. At $N = 256$, the first two resolutions split the repetitions almost evenly. The right panel shows that the sample standard deviation divided by the mean decreases with N .

Figure 5 compares the same 50-cloud means with the bounds of Corollary B.1. For $p = 8$, $N = 2^{18}$ and $\lambda = 1$, taking $M = 16$ ($q_M = 32, q_{M+1} = 39$), $R = 512$ and $s = 4800$ in (B.5) gives, rounding outwards,

$$0.01748608 \leq \Delta_N(8) \leq 0.01762813.$$

The sampled mean is 0.01762459: the lower and upper bounds are 0.79% below and 0.0201% above



(a) Controlled Markovian benchmark.

(b) Controlled common-noise benchmark.

Figure 2: Value gaps for the Markovian model (left) and its common-noise extension (right). Solid curves are computed by the scalar Bellman recursion; markers are Monte Carlo estimates obtained by evaluating the computed policies. Vertical bars show 95% Student confidence intervals based on five independent seeds and account only for sampling error. The dashed line in the left panel is an $N^{-1/2}$ rate guide.

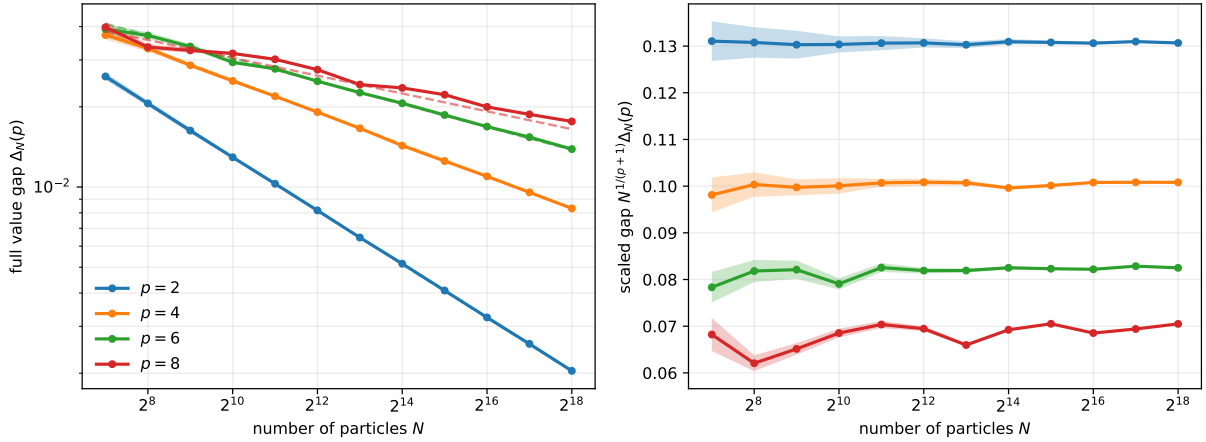


Figure 3: Full-feedback path-law value gap from Proposition 2.5. Solid curves are means over 50 independent repetitions per point, and shaded regions show mean plus or minus one sample standard deviation. Left: $\Delta_N(p)$, with separately anchored dashed $N^{-1/(p+1)}$ guides. Right: the same gaps multiplied by $N^{1/(p+1)}$. The payoff uses the fixed resolution set $\mathcal{S}$.

this mean. At the dominant resolution $q = 5$, the expected fraction of empty cells is 87.44%, and even the largest deficit threshold removes only about 0.67% of the cell side along each face. These features help explain the close bounds. At $N = 128$, the upper bound is 0.042741, rounded upwards, compared with a sampled mean of 0.039789; the finite- N envelope is less sharp at small N .

The bounds concern the exact expectation. At $N = 2^{14}$, the sampled mean exceeds the upper bound by 6.16×10^{-6} , or 0.54 estimated standard errors, consistent with sampling uncertainty. The shaded bands show sample standard deviations across clouds; the standard error of the 50-cloud mean is this standard deviation divided by $\sqrt{50}$.

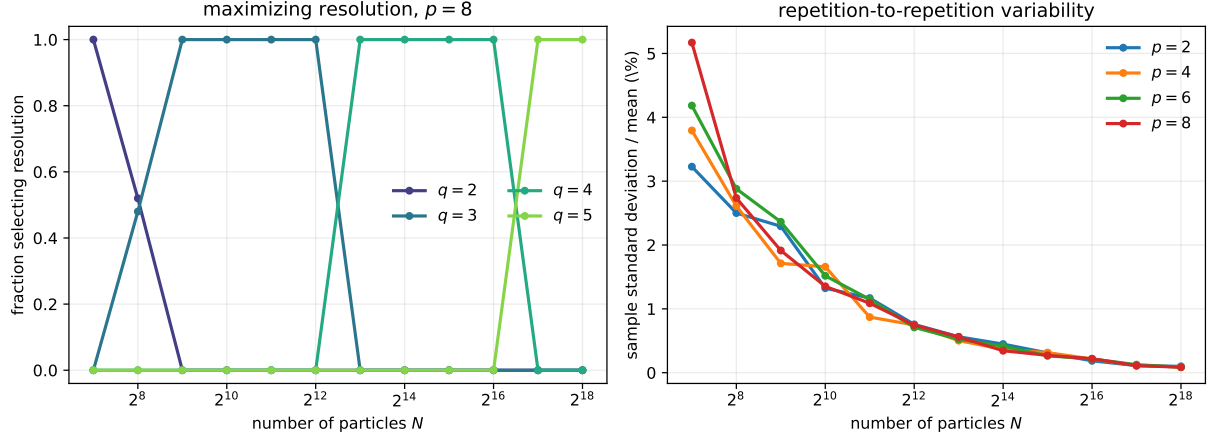


Figure 4: Selected resolutions and sampling variability for the denser-scale full-feedback experiment. Left: for $p = 8$, the fraction of 50 clouds at each N whose sample discrepancy is largest at $q = 2, 3, 4, 5$. The midpoint-law correction is the same across clouds at fixed N ; the selection refers to the sample payoff $H_9(\hat{\nu}_N)$. Right: sample standard deviation divided by mean value gap, in percent, for each p .

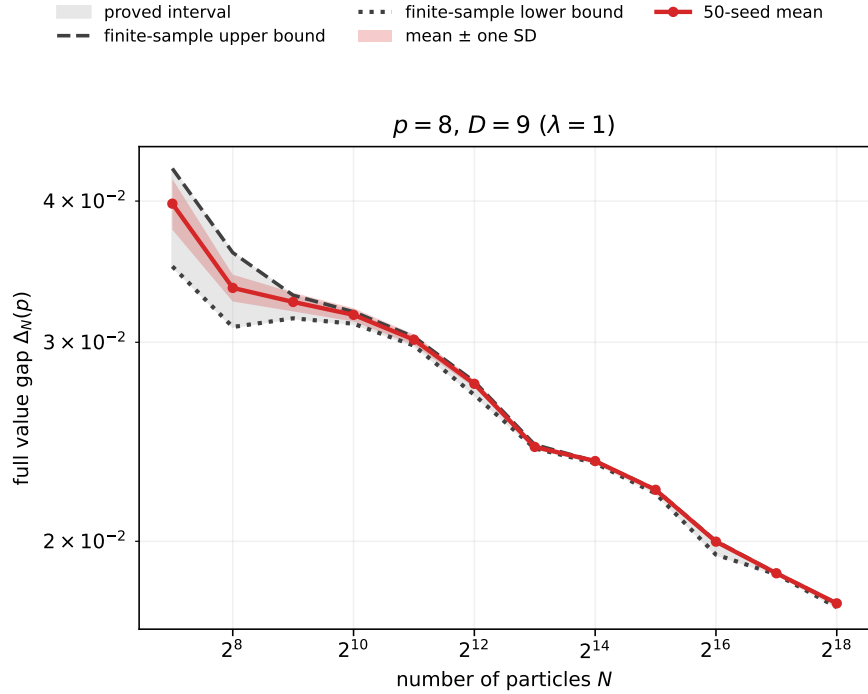


Figure 5: Finite- N bounds for the full-feedback path-law gap at $p = 8$ and $\lambda = 1$. The solid curve and shaded region are the same 50-cloud mean and one-sample-standard-deviation band as in Figure 3. The dotted lower and dashed upper curves are the proved bounds in (B.6); the grey region lies between them. At each N , the bounds use the initial segment $q_{M-1} < 8N^{1/9} \leq q_M$ of the fixed resolution set. The upper bound uses $R = 512$ in the proved finite sum (B.3) and a numerically selected $s > 0$ in (B.5).

Acknowledgements. Some of the experiments have used the HPC resources at NYU and NYU Shanghai.

References

- [1] Charalambos D Aliprantis and Kim C Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer, 2006.
- [2] Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Convergence rate of particle system for second-order PDEs on Wasserstein space. *SIAM Journal on Control and Optimization*, 63(3): 1768–1782, 2025.
- [3] Olivier Bokanowski, Jean-François Chassagneux, Xinyu Li, and Christoph Reisinger. Numerical approximation for path-dependent McKean–Vlasov control with non-asymptotic error estimates. *arXiv preprint arXiv:2606.27181*, 2026.
- [4] Bruno Bouchard and Xiaolu Tan. Limit theory for mean-field control problems with common noise adapted controls, 2025. Preprint, arXiv:2509.14734.
- [5] Pierre Cardaliaguet and Panagiotis E. Souganidis. Regularity of the value function and quantitative propagation of chaos for mean field control problems. *Nonlinear Differential Equations and Applications NoDEA*, 30(2):25, 2023. doi: 10.1007/s00030-022-00823-x.
- [6] Pierre Cardaliaguet, Samuel Daudin, Joe Jackson, and Panagiotis E. Souganidis. An algebraic convergence rate for the optimal control of McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 61(6):3341–3369, 2023. doi: 10.1137/22M1486789.
- [7] Pierre Cardaliaguet, Joe Jackson, Nikiforos Mimikos-Stamatopoulos, and Panagiotis E. Souganidis. Sharp convergence rates for mean field control in the region of strong regularity. *The Annals of Probability*, 54(4):1839–1876, 2026. doi: 10.1214/25-AOP1793.
- [8] Alekos Cecchin. Finite state N -agent and mean field control problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 27:31, 2021. doi: 10.1051/cocv/2021032.
- [9] Alekos Cecchin, Samuel Daudin, Joe Jackson, and Mattia Martini. Quantitative convergence for mean field control with common noise and degenerate idiosyncratic noise. *Electronic Journal of Probability*, 30:1–62, 2025. doi: 10.1214/25-EJP1435. Paper no. 173.
- [10] Samuel Daudin, François Delarue, and Joe Jackson. On the optimal rate for the convergence problem in mean field control. *Journal of Functional Analysis*, 287(12):110660, 2024. doi: 10.1016/j.jfa.2024.110660.
- [11] Mao Fabrice Djete, Dylan Possamaï, and Xiaolu Tan. McKean–Vlasov optimal control: Limit theory and equivalence between different formulations. *Mathematics of Operations Research*, 47(4):2891–2930, 2022. doi: 10.1287/moor.2021.1232.
- [12] Nicolas Fournier. Convergence of the empirical measure in expected Wasserstein distance: non-asymptotic explicit bounds in $\mathbb{R}^d$. *ESAIM: Probability and Statistics*, 27:749–775, 2023. doi: 10.1051/ps/2023011.
- [13] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3–4):707–738, 2015. doi: 10.1007/s00440-014-0583-7.

- [14] Nicolas Gast and Bruno Gaujal. A mean field approach for optimization in discrete time. *Discrete Event Dynamic Systems*, 21:63–101, 2011. doi: 10.1007/s10626-010-0094-3.
- [15] Maximilien Germain, Huy  n Pham, and Xavier Warin. Rate of convergence for particle approximation of PDEs in Wasserstein space. *Journal of Applied Probability*, 59(4):992–1008, 2022. doi: 10.1017/jpr.2021.102.
- [16] Minyi Huang, Roland P. Malham  , and Peter E. Caines. Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, 6(3):221–252, 2006. doi: 10.4310/CIS.2006.v6.n3.a5.
- [17] Joe Jackson and Daniel Lacker. Approximately optimal distributed stochastic controls beyond the mean field setting. *The Annals of Applied Probability*, 35(1):251–308, 2025. doi: 10.1214/24-AAP2114.
- [18] Daniel Lacker. Limit theory for controlled McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017. doi: 10.1137/16M1095895.
- [19] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.
- [20] M  d  ric Motte and Huy  n Pham. Mean-field Markov decision processes with common noise and open-loop controls. *The Annals of Applied Probability*, 32(2):1421–1458, 2022. doi: 10.1214/21-AAP1713.
- [21] M  d  ric Motte and Huy  n Pham. Quantitative propagation of chaos for mean field Markov decision process with common noise. *Electronic Journal of Probability*, 28:1–24, 2023. doi: 10.1214/23-EJP978.
- [22] Sebastian Munoz. The optimal rate of convergence in mean field control via recoupled shadow flows, 2026. Preprint, arXiv:2607.11062.
- [23] Mehdi Talbi. A finite-dimensional approximation for partial differential equations on Wasserstein space. *Stochastic Processes and their Applications*, 177:104445, 2024. doi: 10.1016/j.spa.2024.104445.

A Technical Lemmata

Lemma A.1 (An extension of Fournier and Guillin [13]). *Let $N \in \mathbb{N}^*$, $1 \leq q < q'$, and $\mathbf{Z} := (Z_1, \dots, Z_N)$ be a sequence of independent $\mathbb{R}^d$ -valued random variables such that $\mu_k := \mathbb{P} \circ Z_k^{-1} \in \mathcal{P}_{q'}(\mathbb{R}^d)$ for all $k \in [N]^*$. Then, denoting $\mu^N := \frac{1}{N} \sum_{k=1}^N \mu_k$ and $\mu_{\mathbf{Z}}^N := \frac{1}{N} \sum_{k=1}^N \delta_{Z_k}$:*

$$\mathbb{E}[\mathcal{W}_q(\mu^N, \mu_{\mathbf{Z}}^N)] \leq CM_{q'}^{1/q'}(\mu^N)R_N,$$

where $M_r(\nu) := \int_{\mathbb{R}^d} |x|^r \nu(dx)$, with C a constant independent of N and:

$$R_N := \begin{cases} N^{-1/2q} + N^{-(q'-q)/qq'} & \text{if } q > d/2 \text{ and } q' \neq 2q, \\ N^{-1/2q}(\log(1+N))^{1/q} + N^{-(q'-q)/qq'} & \text{if } q = d/2 \text{ and } q' \neq 2q, \\ N^{-1/d} + N^{-(q'-q)/qq'} & \text{if } 0 \leq q < d/2 \text{ and } q' \neq \frac{dq}{d-q}. \end{cases}$$

For the critical case $q = 1$ and $q' = 2$, we have:

$$\mathbb{E}[\mathcal{W}_1(\mu^N, \mu_{\mathbf{Z}}^N)] \leq CM_2^{1/2}(\mu^N) \begin{cases} N^{-1/2}(\log(1+N))^{1/2}, & d = 1, \\ N^{-1/2}(\log(1+N))^{3/2}, & d = 2, \\ N^{-1/d}, & d \geq 3, \end{cases} \quad (\text{A.1})$$

where C depends only on d . No moment of order greater than two is required.

The same bounds hold conditionally under conditional independence.

Proof. We follow the proof of [Fournier and Guillin \[13, Theorem 1\]](#). We first spell out the modification needed to pass from i.i.d. samples to independent, non-identically distributed samples, and then give the additional summation for $q = 1$, $q' = 2$. In that proof, after the standard dyadic decomposition of $\mathbb{R}^d$ into shells and cells, the only probabilistic input is the following estimate: for each cell A , one needs a bound on

$$\mathbb{E}[|\mu_{\mathbf{Z}}^N(A) - \mu^N(A)|].$$

We now prove that the same cell estimate as in the i.i.d. case remains valid with $\mu^N = N^{-1} \sum_{k=1}^N \mu_k$.

Let $A \subset \mathbb{R}^d$ be Borel and set $p_k := \mu_k(A)$. Then

$$\mu_{\mathbf{Z}}^N(A) = \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{\{Z_k \in A\}}, \quad \mu^N(A) = \frac{1}{N} \sum_{k=1}^N p_k.$$

The random variables $\mathbf{1}_{\{Z_k \in A\}}$ are independent Bernoulli random variables, hence

$$\mathbb{E}[\mu_{\mathbf{Z}}^N(A) - \mu^N(A)] = 0, \quad \text{Var}(\mu_{\mathbf{Z}}^N(A)) = \frac{1}{N^2} \sum_{k=1}^N p_k(1-p_k) \leq \frac{\mu^N(A)}{N}.$$

By Cauchy-Schwarz's inequality,

$$\mathbb{E}[|\mu_{\mathbf{Z}}^N(A) - \mu^N(A)|] \leq \left(\frac{\mu^N(A)}{N} \right)^{1/2}.$$

We also have the trivial bound

$$\mathbb{E}[|\mu_{\mathbf{Z}}^N(A) - \mu^N(A)|] \leq \mathbb{E}[\mu_{\mathbf{Z}}^N(A)] + \mu^N(A) = 2\mu^N(A).$$

Therefore

$$\mathbb{E}[|\mu_{\mathbf{Z}}^N(A) - \mu^N(A)|] \leq \min \left\{ 2\mu^N(A), \left(\frac{\mu^N(A)}{N} \right)^{1/2} \right\}.$$

This is precisely the binomial cell estimate used in [Fournier and Guillin \[13\]](#); the binomial structure is not used elsewhere in the dyadic summation. The remaining deterministic part of the proof is therefore unchanged after replacing the common law of the i.i.d. sample by μ^N . The hypotheses needed for this part are $1 \leq q < q'$ and $M_{q'}(\mu^N) < \infty$, both of which hold by assumption. The case $M_{q'}(\mu^N) = 0$ is immediate. Otherwise, scaling all variables by $M_{q'}(\mu^N)^{-1/q'}$ reduces the proof to $M_{q'}(\mu^N) = 1$. For the rescaled laws and the dyadic shells B_ℓ used there, the tail estimate is

$$\mu^N(B_\ell) \leq C2^{-q'\ell}.$$

Plugging the cell and tail estimates into the dyadic bound of [Fournier and Guillin](#), and then restoring the original scale, gives

$$\mathbb{E}[\mathcal{W}_q^q(\mu_{\mathbf{Z}}^N, \mu^N)] \leq CM_{q'}^{q/q'}(\mu^N) \begin{cases} N^{-1/2} + N^{-(q'-q)/q'} & \text{if } q > d/2, \\ N^{-1/2} \log(1+N) + N^{-(q'-q)/q'} & \text{if } q = d/2, \\ N^{-q/d} + N^{-(q'-q)/q'} & \text{if } q < d/2. \end{cases}$$

Since $q \geq 1$, Jensen's inequality gives

$$\mathbb{E}[\mathcal{W}_q(\mu_{\mathbf{Z}}^N, \mu^N)] \leq \left(\mathbb{E}[\mathcal{W}_q^q(\mu_{\mathbf{Z}}^N, \mu^N)] \right)^{1/q},$$

which yields the stated estimate for $\mathbb{E}[\mathcal{W}_q(\mu^N, \mu_{\mathbf{Z}}^N)]$.

We now focus on the case $q = 1$, $q' = 2$. For $d \geq 3$, the moment order two is strictly above the critical value $d/(d-1)$, so the preceding noncritical estimate yields (A.1). We now consider $d = 1, 2$. The case $M_2(\mu^N) = 0$ is immediate; otherwise, by scaling, we may assume that $M_2(\mu^N) = 1$. Set

$$B_0 := (-1, 1]^d, \quad B_j := (-2^j, 2^j]^d \setminus (-2^{j-1}, 2^{j-1}]^d, \quad j \geq 1, \quad u_j := \mu^N(B_j).$$

Since $|x| \geq 2^{j-1}$ on B_j for $j \geq 1$, we have

$$\sum_{j \geq 0} 2^{2j} u_j \leq u_0 + 4 \int_{\mathbb{R}^d} |x|^2 \mu^N(dx) \leq 5.$$

The dyadic transportation bound of [13, Lemmas 5 and 6], combined with the cell estimate above, gives

$$\mathbb{E}[\mathcal{W}_1(\mu_{\mathbf{Z}}^N, \mu^N)] \leq C \sum_{j \geq 0} 2^j S_d(u_j), \quad S_d(u) := \sum_{\ell \geq 0} 2^{-\ell} \min \left\{ 2u, 2^{d\ell/2} \sqrt{\frac{u}{N}} \right\}.$$

Indeed, at level ℓ each shell is partitioned into at most $2^{d\ell}$ cells, and Cauchy–Schwarz bounds the sum of the square roots of their masses by $2^{d\ell/2} \sqrt{u_j}$. For $u \in [0, 1]$, summation over ℓ yields

$$S_d(u) \leq C \min \left\{ u, \sqrt{\frac{u}{N}} h_d(Nu) \right\}, \quad h_1(t) := 1, \quad h_2(t) := \log(2 + t).$$

Indeed, the fact that $S_d(u) \leq Cu$ is immediate. For $d = 1$, the other bound proceeds from a geometric sum. For $d = 2$, take $L = \lceil \frac{1}{2} \log_2(2 + Nu) \rceil$ and split the sum at L to obtain

$$S_2(u) \leq (L + 1) \sqrt{\frac{u}{N}} + 2u 2^{-L} \leq C \sqrt{\frac{u}{N}} \log(2 + Nu).$$

Now set $K = \lceil \frac{1}{2} \log_2 N \rceil$. On the first $K + 1$ shells, Cauchy–Schwarz and the summed moment bound give

$$\sum_{j=0}^K 2^j \sqrt{u_j} \leq \sqrt{K+1} \left(\sum_{j=0}^K 2^{2j} u_j \right)^{1/2} \leq C \sqrt{\log(1 + N)}.$$

On the remaining shells,

$$\sum_{j>K} 2^j u_j \leq 2^{-K} \sum_{j>K} 2^{2j} u_j \leq CN^{-1/2}.$$

Since $u_j \leq 1$ and h_d is nondecreasing, it follows that

$$\mathbb{E}[\mathcal{W}_1(\mu_{\mathbf{Z}}^N, \mu^N)] \leq CN^{-1/2} \left(h_d(N) \sqrt{\log(1 + N)} + 1 \right).$$

This gives the factors $(\log(1 + N))^{1/2}$ for $d = 1$ and $(\log(1 + N))^{3/2}$ for $d = 2$. Restoring the original scale proves (A.1).

We next justify the conditional form used in the common-noise arguments. Let $\mathcal{G}$ be a sigma-field such that, conditionally on $\mathcal{G}$, the random variables $Z_1, \dots, Z_N$ are independent with conditional laws $\mu_1^{\mathcal{G}}, \dots, \mu_N^{\mathcal{G}}$. Since $(\mathbb{R}^d)^N$ is Polish, the vector $\mathbf{Z}$ has a regular conditional distribution given $\mathcal{G}$.

By conditional independence, this distribution equals $\bigotimes_{k=1}^N \mu_k^{\mathcal{G}}$ almost surely. Applying the preceding bounds to this product law, for almost every realisation, gives the same conditional estimates, including (A.1), with μ^N replaced by $N^{-1} \sum_{k=1}^N \mu_k^{\mathcal{G}}$ and with unchanged constants. $\blacksquare$

The following lemma refines the previous result for measures on discrete path spaces.

Lemma A.2 (A dimension-uniform empirical estimate on path spaces). *Let $p \geq 2$, and $Z_1, \dots, Z_N$ be independent E_p -valued random variables with respective laws $\mu_1, \dots, \mu_N \in \mathcal{P}_2(E_p)$, and set*

$$\mu^N := \frac{1}{N} \sum_{i=1}^N \mu_i, \quad \mu_{\mathbf{Z}}^N := \frac{1}{N} \sum_{i=1}^N \delta_{Z_i}.$$

Then there is a constant $C_d \geq 0$, depending only on d , such that

$$\mathbb{E} \left[\mathcal{W}_{1,\infty,p}(\mu^N, \mu_{\mathbf{Z}}^N) \right] \leq C_d M_{2,\infty,p}^{1/2}(\mu^N) N^{-\frac{1}{d(p+1)}}. \quad (\text{A.2})$$

In particular, C_d is independent of p , N , and the laws $(\mu_i)_{i \leq N}$.

The estimate also holds conditionally whenever the variables are conditionally independent.

Proof. We first work in $\mathbb{R}^D$, $D \geq 3$, with

$$|x|_{\max} := \max_{1 \leq j \leq D} |x_j|, \quad M_{2,\max}(\nu) := \int_{\mathbb{R}^D} |x|_{\max}^2 \nu(dx).$$

We shall prove that, for independent $\mathbb{R}^D$ -valued variables $Z_1, \dots, Z_N$,

$$\mathbb{E} \left[\mathcal{W}_{1,\max}(\mu^N, \mu_{\mathbf{Z}}^N) \right] \leq C M_{2,\max}^{1/2}(\mu^N) N^{-1/D}, \quad (\text{A.3})$$

where C is independent of D , N , and the individual laws. If $M_{2,\max}(\mu^N) = 0$, the result is immediate. Otherwise, by scaling, it suffices to consider

$$M_{2,\max}(\mu^N) = 1.$$

Step 1: The deterministic transport decomposition. As in [12, Section 3], set

$$B_0 := (-1, 1]^D, \quad B_j := (-2^j, 2^j]^D \setminus (-2^{j-1}, 2^{j-1}]^D, \quad j \geq 1.$$

Let $\mathcal{Q}_\ell$ be the dyadic partition of B_0 into $2^{D\ell}$ cubes of side length $2^{1-\ell}$. Thus $\mathcal{Q}_0 = \{B_0\}$, the partitions are nested, and

$$\#\mathcal{Q}_\ell = 2^{D\ell}, \quad \text{diam}_{\max}(C^\ell) = 2^{1-\ell}, \quad C^\ell \in \mathcal{Q}_\ell.$$

For $C^\ell \in \mathcal{Q}_\ell$, set

$$C_j^\ell := (2^j C^\ell) \cap B_j, \quad u_j := \mu^N(B_j), \quad v_j := \mu_{\mathbf{Z}}^N(B_j).$$

We use the convention $0/0 = 0$.

For each j , start with

$$t_{j,B_0} := u_j \wedge v_j.$$

Recursively, if $C^{\ell+1} \in \mathcal{Q}_{\ell+1}$ and $C^\ell \in \mathcal{Q}_\ell$ are such that $C^{\ell+1} \subset C^\ell$, set

$$t_{j,C^{\ell+1}} := t_{j,C^\ell} \left(\frac{\mu^N(C_j^{\ell+1})}{\mu^N(C_j^\ell)} \wedge \frac{\mu_{\mathbf{Z}}^N(C_j^{\ell+1})}{\mu_{\mathbf{Z}}^N(C_j^\ell)} \right).$$

The quantity t_{j,C^ℓ} is the mass kept together inside C_j^ℓ . In particular,

$$0 \leq t_{j,C^\ell} \leq \mu^N(C_j^\ell) \wedge \mu_{\mathbf{Z}}^N(C_j^\ell).$$

The mass separated between levels ℓ and $\ell + 1$ is

$$r_{j,\ell} := \frac{1}{2} \sum_{C^\ell \in \mathcal{Q}_\ell} t_{j,C^\ell} \sum_{\substack{C^{\ell+1} \in \mathcal{Q}_{\ell+1} \\ C^{\ell+1} \subset C^\ell}} \left| \frac{\mu^N(C_j^{\ell+1})}{\mu^N(C_j^\ell)} - \frac{\mu_{\mathbf{Z}}^N(C_j^{\ell+1})}{\mu_{\mathbf{Z}}^N(C_j^\ell)} \right|.$$

Since the child masses sum to the parent mass,

$$r_{j,\ell} = \sum_{F \in \mathcal{Q}_\ell} t_{j,F} - \sum_{C \in \mathcal{Q}_{\ell+1}} t_{j,C}.$$

Consequently, for every integer $K \geq 1$,

$$\sum_{\ell=0}^{K-1} r_{j,\ell} \leq u_j \wedge v_j. \quad (\text{A.4})$$

The coupling construction is the same as in [12, Proposition 3.2]. The mass $r_{j,\ell}$ is transported inside a cell of diameter at most $2^{j+1-\ell}$, while the mass remaining at level K is transported over a distance at most 2^{j+1-K} . The unmatched masses of the sets B_j contribute at most $\sum_j 2^j |u_j - v_j|$, by the triangle inequality through the origin. Hence

$$\begin{aligned} \mathcal{W}_{1,\max}(\mu^N, \mu_{\mathbf{Z}}^N) &\leq \sum_{j \geq 0} 2^j |u_j - v_j| + 2 \sum_{j \geq 0} 2^j \sum_{\ell=0}^{K-1} 2^{-\ell} r_{j,\ell} \\ &\quad + 2^{1-K} \sum_{j \geq 0} 2^j (u_j \wedge v_j). \end{aligned} \quad (\text{A.5})$$

No independence or identical-distribution assumption enters this construction.

Set

$$\rho_{j,\ell} := \mathbb{E}[r_{j,\ell}], \quad U_N := \sum_{j \geq 0} 2^j \mathbb{E}[|u_j - v_j|].$$

Taking expectations in (A.5) gives

$$\mathbb{E}[\mathcal{W}_{1,\max}(\mu^N, \mu_{\mathbf{Z}}^N)] \leq U_N + 2 \sum_{j \geq 0} 2^j \sum_{\ell=0}^{K-1} 2^{-\ell} \rho_{j,\ell} + 2^{1-K} \sum_{j \geq 0} 2^j u_j. \quad (\text{A.6})$$

Step 2: The probabilistic estimates. We now establish the counterparts of [12, (4.5), (4.7), and (4.6)]:

$$U_N \leq K_N, \quad (\text{A.7})$$

$$\rho_{j,\ell} \leq \frac{1}{2} \sum_{F \in \mathcal{Q}_\ell} \sum_{\substack{C \in \mathcal{Q}_{\ell+1} \\ C \subset F}} \mathbb{E} \left[\left| \mu_{\mathbf{Z}}^N(C_j) - \frac{\mu^N(C_j)}{\mu^N(F_j)} \mu_{\mathbf{Z}}^N(F_j) \right| \right], \quad (\text{A.8})$$

$$\rho_{j,\ell} \leq \frac{1}{2} \left(\frac{2^{D(\ell+1)} u_j}{N} \right)^{1/2}, \quad (\text{A.9})$$

where

$$K_N := \left(2(1 - u_0) \wedge \sqrt{\frac{1 - u_0}{N}} \right) + \sum_{j \geq 1} 2^j \left(2u_j \wedge \sqrt{\frac{u_j}{N}} \right).$$

We also have, directly from (A.4),

$$\sum_{\ell=0}^{K-1} \rho_{j,\ell} \leq u_j. \quad (\text{A.10})$$

To prove (A.7), let $A \subset \mathbb{R}^D$ be Borel, and write $p_i := \mu_i(A)$ and $\bar{p} := \mu^N(A)$. Independence gives

$$\begin{aligned} \mathbb{E}[\mu_{\mathbf{Z}}^N(A)] &= \bar{p}, \\ \text{Var}(\mu_{\mathbf{Z}}^N(A)) &= \frac{1}{N^2} \sum_{i=1}^N p_i(1-p_i) \\ &= \frac{1}{N} \left(\bar{p} - \frac{1}{N} \sum_{i=1}^N p_i^2 \right) \leq \frac{\bar{p}(1-\bar{p})}{N}. \end{aligned}$$

Together with the elementary bounds obtained from A and A^c , this yields

$$\mathbb{E}[|\mu_{\mathbf{Z}}^N(A) - \mu^N(A)|] \leq 2\bar{p} \wedge 2(1-\bar{p}) \wedge \left(\frac{\bar{p}(1-\bar{p})}{N} \right)^{1/2}. \quad (\text{A.11})$$

Applying this estimate to B_0 and to B_j , $j \geq 1$, proves (A.7).

Next, we prove (A.8). Let $C^\ell \in \mathcal{Q}_\ell$ and $C^{\ell+1} \in \mathcal{Q}_{\ell+1}$ be such that $C^{\ell+1} \subset C^\ell$. Since

$$t_{j,C^\ell} \leq \mu_{\mathbf{Z}}^N(C_j^\ell),$$

we have

$$t_{j,C^\ell} \left| \frac{\mu^N(C_j^{\ell+1})}{\mu^N(C_j^\ell)} - \frac{\mu_{\mathbf{Z}}^N(C_j^{\ell+1})}{\mu_{\mathbf{Z}}^N(C_j^\ell)} \right| \leq \left| \mu_{\mathbf{Z}}^N(C_j^{\ell+1}) - \frac{\mu^N(C_j^{\ell+1})}{\mu^N(C_j^\ell)} \mu_{\mathbf{Z}}^N(C_j^\ell) \right|.$$

Summing over $C^{\ell+1} \subset C^\ell$, then over $C^\ell \in \mathcal{Q}_\ell$, and taking expectations yields

$$\rho_{j,\ell} \leq \frac{1}{2} \sum_{C^\ell \in \mathcal{Q}_\ell} \sum_{\substack{C^{\ell+1} \in \mathcal{Q}_{\ell+1} \\ C^{\ell+1} \subset C^\ell}} \mathbb{E} \left[\left| \mu_{\mathbf{Z}}^N(C_j^{\ell+1}) - \frac{\mu^N(C_j^{\ell+1})}{\mu^N(C_j^\ell)} \mu_{\mathbf{Z}}^N(C_j^\ell) \right| \right].$$

It remains to prove (A.9). This is where we replace the conditional binomial argument of [12]. For Borel sets $A \subset B$ with $\mu^N(B) > 0$, set

$$\theta := \frac{\mu^N(A)}{\mu^N(B)}, \quad Y_i := \mathbf{1}_{\{Z_i \in A\}} - \theta \mathbf{1}_{\{Z_i \in B\}}.$$

The variables Y_i need not be individually centred, but

$$\sum_{i=1}^N \mathbb{E}[Y_i] = N(\mu^N(A) - \theta \mu^N(B)) = 0.$$

Moreover,

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}[Y_i^2] &= N(1-\theta)^2 \mu^N(A) + N\theta^2(\mu^N(B) - \mu^N(A)) \\ &= N\mu^N(B)\theta(1-\theta) \leq N\mu^N(A). \end{aligned}$$

Thus independence and the Cauchy–Schwarz inequality imply

$$\mathbb{E}[|\mu_{\mathbf{Z}}^N(A) - \theta \mu_{\mathbf{Z}}^N(B)|] = \frac{1}{N} \mathbb{E} \left[\left| \sum_{i=1}^N Y_i \right| \right]$$

$$\leq \frac{1}{N} \left(\sum_{i=1}^N \text{Var}(Y_i) \right)^{1/2} \leq \left(\frac{\mu^N(A)}{N} \right)^{1/2}. \quad (\text{A.12})$$

If $\mu^N(B) = 0$, both empirical masses vanish almost surely.

Applying (A.12) with $A = C_j$ and $B = F_j$ in (A.8), we obtain

$$\begin{aligned} \rho_{j,\ell} &\leq \frac{1}{2\sqrt{N}} \sum_{C \in \mathcal{Q}_{\ell+1}} \sqrt{\mu^N(C_j)} \\ &\leq \frac{1}{2\sqrt{N}} \left(\# \mathcal{Q}_{\ell+1} \sum_{C \in \mathcal{Q}_{\ell+1}} \mu^N(C_j) \right)^{1/2} \\ &= \frac{1}{2} \left(\frac{2^{D(\ell+1)} u_j}{N} \right)^{1/2}. \end{aligned}$$

This proves (A.9). The two variance computations above are the only probabilistic modifications needed for independent, non-identically distributed variables.

Step 3: Summation with a dimension-uniform constant. Rather than optimise the numerical constant as in [12], we sum the preceding bounds directly.

By (A.10) and (A.9),

$$\sum_{\ell=0}^{K-1} 2^{-\ell} \rho_{j,\ell} \leq S_D(u_j),$$

where, for $u \in [0, 1]$,

$$S_D(u) := \sum_{\ell \geq 0} 2^{-\ell} \left(u \wedge 2^{D(\ell+1)/2-1} \sqrt{\frac{u}{N}} \right).$$

We claim that

$$S_D(u) \leq C u^{1-1/D} N^{-1/D}, \quad D \geq 3, \quad (\text{A.13})$$

with an absolute constant C .

If $Nu \leq 1$, then

$$S_D(u) \leq 2u \leq 2u^{1-1/D} N^{-1/D}.$$

If $Nu > 1$, set

$$L := \left\lfloor \frac{\log_2(Nu)}{D} \right\rfloor.$$

Using the square-root bound for $\ell < L$ and the bound by u for $\ell \geq L$, we obtain

$$\begin{aligned} S_D(u) &\leq 2^{D/2-1} \sqrt{\frac{u}{N}} \sum_{\ell=0}^{L-1} 2^{(D/2-1)\ell} + u \sum_{\ell \geq L} 2^{-\ell} \\ &\leq \frac{1}{1-2^{1-D/2}} \sqrt{\frac{u}{N}} 2^{(D/2-1)L} + 2^{1-L} u \\ &\leq \left(\frac{1}{1-2^{-1/2}} + 4 \right) u^{1-1/D} N^{-1/D}. \end{aligned}$$

Here we used $D \geq 3$ and $2^{DL} \leq Nu < 2^{D(L+1)}$. This proves (A.13). The mass-fluctuation term has the same rate. Indeed, for $D \geq 3$,

$$2u \wedge \sqrt{\frac{u}{N}} \leq (2u)^{1-2/D} \left(\sqrt{\frac{u}{N}} \right)^{2/D} \leq 2u^{1-1/D} N^{-1/D}.$$

Therefore, since $N^{-1/2} \leq N^{-1/D}$,

$$K_N \leq N^{-1/D} \left(1 + 2 \sum_{j \geq 1} 2^j u_j^{1-1/D} \right).$$

Substituting these estimates into (A.6) gives

$$\begin{aligned} \mathbb{E}[\mathcal{W}_{1,\max}(\mu^N, \mu_{\mathbf{Z}}^N)] &\leq C N^{-1/D} \left(1 + \sum_{j \geq 0} 2^j u_j^{1-1/D} \right) \\ &\quad + 2^{1-K} \sum_{j \geq 0} 2^j u_j. \end{aligned} \tag{A.14}$$

The normalisation $M_{2,\max}(\mu^N) = 1$ implies

$$\sum_{j \geq 0} 2^{2j} u_j \leq u_0 + 4 \int_{\mathbb{R}^D} |x|_{\max}^2 \mu^N(dx) \leq 5.$$

Hölder's inequality, with conjugate exponents $D/(D-1)$ and D , now yields

$$\begin{aligned} \sum_{j \geq 0} 2^j u_j^{1-1/D} &= \sum_{j \geq 0} (2^{2j} u_j)^{1-1/D} 2^{-j(1-2/D)} \\ &\leq \left(\sum_{j \geq 0} 2^{2j} u_j \right)^{1-1/D} \left(\sum_{j \geq 0} 2^{-j(D-2)} \right)^{1/D} \\ &\leq 5^{1-1/D} (1 - 2^{-(D-2)})^{-1/D}, \end{aligned}$$

which is bounded uniformly in D . Note also that $\sum_j 2^j u_j < \infty$. Letting $K \rightarrow \infty$ in (A.14) then proves, for some constant C independent of D ,

$$\mathbb{E}[\mathcal{W}_{1,\max}(\mu^N, \mu_{\mathbf{Z}}^N)] \leq C N^{-1/D}$$

under the normalisation. Rescaling the variables proves (A.3).

Step 4: Path spaces and conditional estimates. Identify $E_p = (\mathbb{R}^d)^{p+1}$ with $\mathbb{R}^D$, where $D = d(p+1)$, and define

$$|\omega|_{\max} := \max_{\substack{0 \leq k \leq p \\ 1 \leq j \leq d}} |\omega_k^j|.$$

Since $|\omega|_{\max} \leq |\omega|_{\infty,p} \leq \sqrt{d} |\omega|_{\max}$, we have:

$$\mathcal{W}_{1,\infty,p}(\nu, \nu') \leq \sqrt{d} \mathcal{W}_{1,\max}(\nu, \nu'), \quad M_{2,\max}(\nu) \leq M_{2,\infty,p}(\nu).$$

Since $p \geq 2$ implies $D \geq 3$, (A.3) gives (A.2), with $C_d = C\sqrt{d}$.

Finally, suppose that $Z_1, \dots, Z_N$ are independent conditionally on a sigma-field $\mathcal{G}$. Let $\mu_i^{\mathcal{G}}$ be their regular conditional laws and set

$$\bar{\mu}^{\mathcal{G}} := \frac{1}{N} \sum_{i=1}^N \mu_i^{\mathcal{G}}.$$

The conditional law of $(Z_1, \dots, Z_N)$ is $\otimes_{i=1}^N \mu_i^{\mathcal{G}}$ almost surely. Applying the preceding deterministic-law estimate to this product law gives

$$\mathbb{E}[\mathcal{W}_{1,\infty,p}(\bar{\mu}^{\mathcal{G}}, \mu_{\mathbf{Z}}^N) \mid \mathcal{G}] \leq C_d M_{2,\infty,p}^{1/2}(\bar{\mu}^{\mathcal{G}}) N^{-1/[d(p+1)]},$$

with the same constant. This proves the conditional assertion. ■

B Proofs for the explicit path-dependent example

B.1 Optimal feedback and convergence rate

We use the model and notation of Section 2.5. We prove Proposition 2.5.

Proof. First, consider any admissible feedback policy, and couple its controlled paths with paths under $\mathbf{a}^*$ using identical initial states and innovations ξ_{n+1}^i . Put $\delta_n^i := a_n^i - \phi_n(X_{0:n}^i)$ along the arbitrary controlled path. By the definition of R , the corresponding residual paths are

$$R(X_{0:p}^i) = (x_i, \xi_1^i + \delta_0^i, \dots, \xi_p^i + \delta_{p-1}^i), \quad R(X_{0:p}^{i,*}) = (x_i, \xi_1^i, \dots, \xi_p^i).$$

The coupling by particle index and (2.20) give, pathwise, writing $m_{R(\mathbf{X})}^N = N^{-1} \sum_i \delta_{R(X_{0:p}^i)}$,

$$H_D(m_{R(\mathbf{X})}^N) \geq H_D(m_{R(\mathbf{X}^*)}^N) - \frac{1}{N} \sum_{i=1}^N \max_{0 \leq n < p} |\delta_n^i|.$$

Since $\rho > \lambda$ and $\max_n |\delta_n^i| \leq \sum_n |\delta_n^i|$, the reward of the arbitrary policy is no larger than that of $\mathbf{a}^*$ for each realisation. This establishes optimality even when every action depends on the full particle history. In the mean-field problem, coupling the two residual laws by the common initial state and innovations gives $\mathcal{W}_{1,\infty,p}(R_{\#}\mathcal{L}(X_{0:p}^a), R_{\#}\mathcal{L}(X_{0:p}^{a*})) \leq \mathbb{E}[\max_n |a_n - \phi_n(X_{0:n})|]$; the same penalty comparison proves mean-field optimality. Both arguments apply from any intermediate stopped path.

Under the optimal feedback, every future residual coordinate equals its innovation. The map R is Lipschitz on stopped paths with constant at most $1 + \sup_n \text{Lip}(B_n) + \sup_n \text{Lip}(\phi_n)$. Coupling two initial stopped-path laws and then using the same future innovations therefore shows that all V_n have a common $\mathcal{W}_{1,\infty,n}$ -Lipschitz constant, independent of p . Moreover, $|X_{n+1}| \leq 1 + A + \sup_n \|B_n\|_\infty$ under every policy, so the second-moment condition of Assumption 2.3(ii) holds uniformly in p . Finally, the explicit optimal-policy formula and (2.20) make V_0^N continuous in its initial configuration. This verifies Assumption 2.3.

For the specified $\mathbf{x}$, optimality gives $V_0(m_{\mathbf{x}}^N) = -\lambda H_D(\bar{\nu}_N)$ and $V_0^N(\mathbf{x}) = -\lambda \mathbb{E}[H_D(\hat{\nu}_N)]$. Convexity of H_D and $\mathbb{E}\hat{\nu}_N = \bar{\nu}_N$ prove the last inequality in (2.24). For the upper bound, write $S_q(\nu, \nu') := \sum_{Q \in \mathcal{Q}_q} |\int \psi_{q,Q} d(\nu - \nu')|$. The triangle inequality and the definition of H_D give

$$\Delta_N(p)/\lambda \leq \mathbb{E} \sup_{q \in \mathcal{S}} S_q(\hat{\nu}_N, \bar{\nu}_N) \leq \sum_{q \in \mathcal{S}} \mathbb{E} S_q(\hat{\nu}_N, \bar{\nu}_N).$$

At resolution q , put $h = 1/q$ and $K = h^{-D}$. Conditional on any fixed initial midpoint, the future coordinates are uniform, and integration of the minimum of their distances to the cell boundary gives

$$\mathbb{E} \sum_{Q \in \mathcal{Q}_q} \psi_{q,Q}(Z_i) \leq \frac{h}{2D}, \quad \mathbb{E} \sum_{Q \in \mathcal{Q}_q} \psi_{q,Q}(Z_i)^2 \leq \frac{h^2}{2D(D+1)}, \quad Z_i = (x_i, \xi_1^i, \dots, \xi_p^i).$$

Using independence and Cauchy–Schwarz for the K centred cell sums, and alternatively bounding their absolute values by the sums of their nonnegative terms, yields

$$\mathbb{E} S_q(\hat{\nu}_N, \bar{\nu}_N) \leq \min \left\{ \frac{q^{D/2-1}}{\sqrt{2D(D+1)N}}, \quad \frac{1}{Dq} \right\}.$$

Write $a = D/2 - 1 > 0$ and $j_0 = \lfloor (\log_2 N)/D \rfloor$. There are at most four distinct $q \in \mathcal{S}$ in each octave $[2^j, 2^{j+1})$. For $j < j_0$, bound q^a by $2^{(j+1)a}$ and sum the first term geometrically. For $j \geq j_0$, bound q^{-1} by 2^{-j} and sum the second term. Since $2^{j_0} \leq N^{1/D} < 2^{j_0+1}$, this gives the conservative bound

$$\Delta_N(p) \leq \lambda C_D^{\text{dense}} N^{-1/D}. \quad (\text{B.1})$$

Both terms in the definition of C_D^{dense} decrease with $D \geq 3$, so $C_D^{\text{dense}} \leq C_3^{\text{dense}}$, proving the stated uniformity in p .

For the lower bound, the midpoint coupling of m_x^N with $U[0, 1]$ and identical future innovations gives

$$H_D(\bar{\nu}_N) \leq \mathcal{W}_{1,\infty,p}(\bar{\nu}_N, U_D) \leq \frac{1}{4N}. \quad (\text{B.2})$$

Choose $j = \lceil (\log_2 N)/D \rceil$ and the admissible dyadic resolution $q = 2^j \in \mathcal{S}$. With $h = 1/q$, $N^{-1/D}/2 < h \leq N^{-1/D}$ and $Nh^D \leq 1$. Each cube of $\mathcal{Q}_q$ contains at most $Nh + 1$ of the fixed initial positions in its first coordinate. For any one such cube, the probability that none of its eligible particles has all $D - 1$ innovation coordinates in the cube is at least $(1 - h^{D-1})^{Nh+1}$. In fact, with $u = h^{D-1} \leq 1/4$, we have $(Nh + 1)u \leq Nh^D + h^{D-1} \leq 5/4$ and hence

$$(1 - u)^{Nh+1} \geq \exp\left(-\frac{(Nh + 1)u}{1 - u}\right) \geq e^{-5/3}.$$

Every empty cube contributes $h^{D+1}/[2(D+1)]$ to $H_D(\bar{\nu}_N)$ by (2.21). Summing over the h^{-D} cubes yields

$$\mathbb{E}H_D(\bar{\nu}_N) \geq \frac{h}{2(D+1)}(1 - h^{D-1})^{Nh+1} \geq \frac{e^{-5/3}}{4(D+1)}N^{-1/D}.$$

Subtracting (B.2) and using $N^{1-1/D} \geq 2(D+1)e^{5/3}$ proves the stated lower bound for $D = p+1 \geq 3$. $\blacksquare$

B.2 Finite-sample bounds

We retain the notation of Proposition 2.5. In particular, $D = p + 1$, $x_i = (i - \frac{1}{2})/N$, and $Z_i = (x_i, \xi_1^i, \dots, \xi_p^i)$, so that $\bar{\nu}_N = N^{-1} \sum_{i=1}^N \delta_{Z_i}$.

Corollary B.1 (Finite-sample bounds). *Write $\mathcal{S} = \{q_1 < q_2 < \dots\}$ and fix integers $M \geq 2$ and $R \geq 1$. For $q \geq 2$, set $a_q = \lfloor N/q \rfloor$, $r_q = N - a_q q$, and*

$$F_{N,p}(q) := \frac{(q - r_q)(1 - q^{-p})^{a_q} + r_q(1 - q^{-p})^{a_q+1}}{q}, \quad \ell_{N,p}(q) := \frac{F_{N,p}(q)}{(D+1)q} - \frac{1}{2N}.$$

Also define

$$b_N(q) := \min \left\{ \frac{q^{D/2-1}}{\sqrt{2D(D+1)N}}, \frac{1}{Dq} \right\}, \quad \tau_M := \frac{1/2 + 1/[2(D+1)]}{q_{M+1}}.$$

For $q \in \mathcal{S}$, set $\beta_q := q^{-(D+1)}/[2(D+1)]$. For $0 < t \leq \beta_q$, define

$$w_q(t) := (q^{-1} - 2Nt)_+, \quad G_{N,q}(t) := (1 - w_q(t)^{D-1})^{(Nw_q(t)-1)_+}.$$

Using R equally spaced deficit thresholds, define the finite sum

$$u_{N,R}(q) := \frac{1}{(D+1)qR} \sum_{r=1}^R G_{N,q}\left(\frac{r\beta_q}{R}\right), \quad \hat{b}_{N,R}(q) := \min\{b_N(q), u_{N,R}(q)\}. \quad (\text{B.3})$$

For the fluctuation bound, set

$$v_{N,q} := \frac{1}{2D(D+1)Nq^2}, \quad c_{N,q} := \frac{D+1}{2DNq},$$

and, for $s > 0$, put

$$K_{N,q}(s) := \min \left\{ \frac{s^2}{8Nq^2}, \frac{v_{N,q}}{c_{N,q}^2} (e^{sc_{N,q}} - 1 - sc_{N,q}) \right\}. \quad (\text{B.4})$$

Then, writing

$$U_{N,M,R}(s) := \frac{1}{s} \log \left(e^{s\tau_M} + \sum_{j=1}^M \exp \left\{ s\hat{b}_{N,R}(q_j) + K_{N,q_j}(s) \right\} \right), \quad (\text{B.5})$$

for every $N \geq 1$ and $p \geq 2$ we have

$$\lambda \max \left\{ 0, \max_{2 \leq j \leq M} \ell_{N,p}(q_j) \right\} \leq \Delta_N(p) \leq \lambda \inf_{s>0} U_{N,M,R}(s). \quad (\text{B.6})$$

The choices of M , R and s affect only the evaluation of these bounds; the payoff still uses every resolution in $\mathcal{S}$. Any fixed $s > 0$ in (B.5) gives a valid upper bound.

Proof. Lower bound. At a fixed resolution q , write $Y_Q = \int \psi_{q,Q} d\hat{\nu}_N$ and $b = \int \psi_{q,Q} dU_D$. The identity

$$\sum_Q |Y_Q - b| = \sum_Q (Y_Q - b) + 2 \sum_Q (b - Y_Q)_+$$

and $\left| \sum_Q \mathbb{E}(Y_Q - b) \right| \leq H_D(\bar{\nu}_N)$ give

$$\Delta_N(p)/\lambda \geq 2b \mathbb{E}[\text{number of empty cubes}] - 2H_D(\bar{\nu}_N).$$

Exactly r_q first-coordinate cells contain $a_q + 1$ initial midpoints, and the remaining $q - r_q$ contain a_q . Conditional on eligibility in such a cell, each particle chooses any specified future-coordinate cell with probability q^{-p} . Thus $F_{N,p}(q)$ is the expected fraction of empty cubes. Substituting the tent integral and (B.2) proves the lower bound.

Expected discrepancy at a fixed resolution. Put $A_q = \sum_{Q \in \mathcal{Q}_q} |\int \psi_{q,Q} d(\hat{\nu}_N - U_D)|$ and $X_q = A_q - H_D(\bar{\nu}_N)$. The proof of Proposition 2.5 gives $\mathbb{E}X_q \leq b_N(q)$, since $X_q \leq S_q(\hat{\nu}_N, \bar{\nu}_N)$. To obtain a second bound on this mean, use the identity above with $b = \beta_q$. For a threshold $0 < t \leq \beta_q$ with $w_q(t) > 0$, a particle in the concentric subcube of Q of side $w_q(t)$ satisfies $\psi_{q,Q}(Z_i) \geq Nt$. Its contribution alone makes $Y_Q \geq t$. Hence $Y_Q < t$ requires this central subcube to be empty. Its first-coordinate interval contains at least $(Nw_q(t) - 1)_+$ initial midpoints. Each eligible particle independently places its $D - 1$ future coordinates in the central intervals with probability $w_q(t)^{D-1}$. Therefore $\mathbb{P}(Y_Q < t) \leq G_{N,q}(t)$. If $w_q(t) = 0$, this inequality follows from $G_{N,q}(t) = 1$.

The deficit is the integral of its threshold indicators. Since $G_{N,q}$ is nondecreasing, the right Riemann sum bounds this integral from above:

$$\mathbb{E}(\beta_q - Y_Q)_+ = \int_0^{\beta_q} \mathbb{P}(Y_Q < t) dt \leq \int_0^{\beta_q} G_{N,q}(t) dt \leq \frac{\beta_q}{R} \sum_{r=1}^R G_{N,q}\left(\frac{r\beta_q}{R}\right).$$

Summing over all q^D cubes and using $\sum_Q \mathbb{E}(Y_Q - \beta_q) \leq H_D(\bar{\nu}_N)$ gives $\mathbb{E}X_q \leq u_{N,R}(q)$, and hence $\mathbb{E}X_q \leq \hat{b}_{N,R}(q)$. Every finite R gives a valid upper bound; $R = 1$ charges the full deficit whenever the smallest central subcube is empty. The finite sums converge to the integral as $R \rightarrow \infty$. Their excess over $2q^D \int_0^{\beta_q} G_{N,q}(t) dt$ is at most $1/[(D+1)qR]$.

Fluctuations at a fixed resolution. Replacing one independent residual sample changes X_q by at most $1/(Nq)$: the old and new points each contribute to at most one tent, of height at most $1/(2q)$. The bounded-differences exponential-moment estimate therefore gives, for $s > 0$,

$$\mathbb{E}e^{s(X_q - \mathbb{E}X_q)} \leq e^{s^2/(8Nq^2)}.$$

We also retain the smaller second moments of the tent contributions. Let $h = 1/q$ and $T_i := \sum_Q \psi_{q,Q}(Z_i)$. A point contributes to at most one tent, so $0 \leq T_i \leq h/2$. Uniformity of the $D-1$ future coordinates gives $\mathbb{P}(T_i > t) \leq (1 - 2qt)^{D-1}$ for $0 \leq t \leq h/2$. Integrating this tail, and then its product with $2t$, yields

$$\mathbb{E}T_i \leq \frac{h}{2D}, \quad \mathbb{E}T_i^2 \leq \frac{h^2}{2D(D+1)}. \quad (\text{B.7})$$

To use these moments for the nonlinear quantity A_q , reveal the independent samples one at a time. Write $\mathcal{G}_i := \sigma(Z_1, \dots, Z_i)$, with $\mathcal{G}_0$ trivial, and

$$d_i := \mathbb{E}[A_q \mid \mathcal{G}_i] - \mathbb{E}[A_q \mid \mathcal{G}_{i-1}].$$

Remove the i th atom without changing the denominator N , defining

$$A_q^{(-i)} := \sum_Q \left| \frac{1}{N} \sum_{j \neq i} \psi_{q,Q}(Z_j) - \beta_q \right|.$$

The triangle inequality gives $|A_q - A_q^{(-i)}| \leq T_i/N$. Independence gives $\mathbb{E}[A_q^{(-i)} \mid \mathcal{G}_i] = \mathbb{E}[A_q^{(-i)} \mid \mathcal{G}_{i-1}]$. Thus $g_i := \mathbb{E}[A_q - A_q^{(-i)} \mid \mathcal{G}_i]$ satisfies $d_i = g_i - \mathbb{E}[g_i \mid \mathcal{G}_{i-1}]$ and $|g_i| \leq T_i/N$. The conditional second moment of d_i is the conditional variance of g_i , which is bounded by $\mathbb{E}[g_i^2 \mid \mathcal{G}_{i-1}]$. Independence of Z_i from $\mathcal{G}_{i-1}$ and (B.7) therefore imply

$$\mathbb{E}[d_i^2 \mid \mathcal{G}_{i-1}] \leq \frac{\mathbb{E}T_i^2}{N^2} \leq \frac{v_{N,q}}{N}, \quad d_i \leq \frac{h/2 + \mathbb{E}T_i}{N} \leq c_{N,q}.$$

For $x \leq c$ and $s, c > 0$,

$$e^{sx} \leq 1 + sx + \frac{x^2}{c^2}(e^{sc} - 1 - sc),$$

as follows from $e^{sx} - 1 - sx = s^2 x^2 \int_0^1 (1-u)e^{sux} du$. Apply this inequality conditionally to the centred increments d_i , using the preceding bounds on d_i and its conditional second moment. Iterating over i and using $\sum_i d_i = X_q - \mathbb{E}X_q$ yields

$$\log \mathbb{E}e^{s(X_q - \mathbb{E}X_q)} \leq \frac{v_{N,q}}{c_{N,q}^2} (e^{sc_{N,q}} - 1 - sc_{N,q}).$$

Together with the bounded-differences estimate, this gives $\mathbb{E}e^{s(X_q - \mathbb{E}X_q)} \leq e^{K_{N,q}(s)}$.

Maximum over resolutions. Since $A_q \leq [1/2 + 1/(2(D+1))]/q$, all resolutions beyond q_M contribute at most τ_M . Applying the exponential-moment bound to $X_{q_1}, \dots, X_{q_M}$ and the constant τ_M yields

$$\mathbb{E} \max(X_{q_1}, \dots, X_{q_M}, \tau_M) \leq \frac{1}{s} \log \left(e^{s\tau_M} + \sum_{j=1}^M \mathbb{E}e^{sX_{q_j}} \right) \leq U_{N,M,R}(s).$$

The first inequality follows by bounding the exponential of the maximum by the sum of exponentials and applying Jensen's inequality to the logarithm. It requires no independence across resolutions. This maximum bounds $\Delta_N(p)/\lambda$ in expectation. Taking the infimum over $s > 0$ proves the upper bound. $\blacksquare$

The bounds have a direct occupancy interpretation. At a fixed resolution, the identity in the proof counts the empty-cell deficits twice: in expectation, their missing mass is accompanied by excess mass in other cells, up to the deterministic midpoint-law correction. The lower bound omits the deficits of occupied cells whose tent mass is below β_q , and the possible gain from choosing the resolution separately for each cloud. When the mean count per cell N/q^D is small, most cells are empty, and a single particle away from the cell faces already supplies the target tent mass. At

threshold t , the central subcube removes a strip of width Nt along each face. Integrating over thresholds measures the size of a possible deficit, rather than charging its full size whenever the smallest central subcube is empty. Several small particle contributions can also add up to the target mass; the upper bound remains conservative because it only uses the certificate supplied by one particle.

The fluctuation estimate improves for a different reason. The largest possible tent height is $h/2$, but a typical point in a high-dimensional cube is close to at least one face. The uniform future coordinates give moment bounds of order h/D and h^2/D^2 . The martingale argument transfers these moments to the discrepancy even though it is a nonlinear function of all particles. Thus the upper bound need not charge typical clouds for the largest possible change of one particle. These moments are analytic model bounds, independent of the Monte Carlo estimates. This improvement matters especially at small N : fluctuations are bounded at an $N^{-1/2}$ scale for each fixed resolution, faster than the principal $N^{-1/D}$ gap. Competition between resolutions can still contribute, and neither refinement asserts equality of the bounds and the value gap.

C Uniform time discretisation for path-dependent particle control

We use the setting and notation of Sections 3.1 and 3.2. This appendix establishes a time-discretisation estimate, uniform in N , for the particle problem (3.5) and its Euler approximation (3.10). The argument adapts that of Bokanowski et al. [3] to the N -particle framework, using the Filippov convexity condition in Assumption 3.1, without requiring time regularity of an optimal control.

Write $\mathbb{F}^N = (\mathcal{F}_t^N)_{t \leq T}$ for the augmented joint Brownian filtration underlying $\mathcal{A}_N$ in Section 3.1. For a fixed initial configuration $\mathbf{x} \in (\mathbb{R}^d)^N$, denote by $J^N(\mathbf{x}, \alpha)$ and $J^{N,\pi}(\mathbf{x}, \alpha^\pi)$ the expected rewards inside the suprema in (3.5) and (3.10) at $n = 0$, respectively. For a continuous path ω , we use the shorthand $\ell_{\pi,k}(\omega) := \ell_{\pi,k}(\omega_{t_0}, \dots, \omega_{t_k})$ for the interpolation map of Section 3.2.

Lemma C.1 (Uniform time-discretisation error). *Let Assumption 3.1 hold. Then there is a constant $C \geq 0$, depending only on T, d and L , such that, for every $N \geq 1$, every uniform grid π with $|\pi| \leq 1$, and every $\mathbf{x} \in (\mathbb{R}^d)^N$,*

$$\left| V_0^N(\mathbf{x}) - U_0^{N,\pi}(\mathbf{x}) \right| \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}. \quad (\text{C.1})$$

In particular, the constant is independent of the number of particles.

Proof. Throughout the proof, C denotes a constant which may change from line to line but is independent of N , π , and the controls. The empirical coupling estimate (3.1), together with the linear growth implied by Assumption 3.1 and standard SDE and Euler estimates, yields

$$\sup_{\alpha \in \mathcal{A}_N} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|X^{i,\alpha}|_\infty^2 \right] + \sup_{\alpha^\pi \in \mathcal{A}_N^\pi} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|\ell_\pi(X^{i,\pi,\alpha^\pi})|_\infty^2 \right] \leq C(1 + m_2^N(\mathbf{x})). \quad (\text{C.2})$$

Due to the normalisation in m_2^N and to (3.1), the constant C is indeed independent of N .

Step 1: From a continuous-time control to a grid control. Fix $\alpha \in \mathcal{A}_N$ and write $\mathbf{X} := \mathbf{X}^\alpha$. For every $k < p$ and $i \in [N]^*$, set

$$\omega_k^i := \ell_{\pi,k}(X^{i,\alpha}), \quad \omega_k := (\omega_k^1, \dots, \omega_k^N), \quad m_k := m_{\omega_k}^N.$$

Define the $\mathcal{F}_{t_k}^N$ -measurable random variables

$$B_k^i := \frac{1}{|\pi|} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} b_{t_k}(\omega_k^i, m_k, \alpha_t^i) dt \mid \mathcal{F}_{t_k}^N \right], \quad F_k^i := \frac{1}{|\pi|} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} f_{t_k}(\omega_k^i, m_k, \alpha_t^i) dt \mid \mathcal{F}_{t_k}^N \right].$$

For fixed (t_k, ω_k^i, m_k) , the random vector $(b_{t_k}(\omega_k^i, m_k, \alpha_t^i), f_{t_k}(\omega_k^i, m_k, \alpha_t^i))$ takes its values in the closed convex set $\mathcal{K}(t_k, \omega_k^i, m_k)$ by Assumption 3.1 (iv). Its time average belongs to this set by convexity, and its conditional expectation given $\mathcal{F}_{t_k}^N$ still belongs to it because this random closed convex set is $\mathcal{F}_{t_k}^N$ -measurable. After fixing $\mathcal{F}_{t_k}^N$ -measurable versions of B_k^i and F_k^i , define

$$\Phi_k^i(a) := \left| b_{t_k}(\omega_k^i, m_k, a) - B_k^i \right|^2 + \left(F_k^i - f_{t_k}(\omega_k^i, m_k, a) \right)_+^2, \quad a \in \mathbb{A}.$$

The map $(\omega, a) \mapsto \Phi_k^i(\omega, a)$ is $\mathcal{F}_{t_k}^N \otimes \mathcal{B}(\mathbb{A})$ -measurable and is continuous in a . Moreover, the preceding inclusion in $\mathcal{K}(t_k, \omega_k^i, m_k)$ implies that $\min_{a \in \mathbb{A}} \Phi_k^i(a) = 0$ almost surely. Hence, by the measurable maximum theorem [1, Theorem 18.19, p. 605], there exists an $\mathcal{F}_{t_k}^N$ -measurable minimiser $\bar{\alpha}_{t_k}^{i, \pi}$. Therefore,

$$B_k^i = b_{t_k}(\omega_k^i, m_k, \bar{\alpha}_{t_k}^{i, \pi}), \quad F_k^i \leq f_{t_k}(\omega_k^i, m_k, \bar{\alpha}_{t_k}^{i, \pi}). \quad (\text{C.3})$$

Applying this construction separately to each particle gives a grid control $\bar{\alpha}^\pi \in \mathcal{A}_N^\pi$. This coordinatewise construction is possible because the interaction involves only the empirical distribution of the state paths, which is frozen at time t_k in (C.3).

Let

$$\bar{\mathbf{X}}^\pi := \mathbf{X}^{\pi, \bar{\alpha}^\pi} = (\bar{X}^{1, \pi}, \dots, \bar{X}^{N, \pi})$$

be the Euler scheme (3.9) driven by $\bar{\alpha}^\pi$. We also introduce its continuous Euler–Maruyama interpolation. For $t \in [t_k, t_{k+1}]$, set

$$\tilde{X}_t^{i, \pi} := \bar{X}_{t_k}^{i, \pi} + b_{t_k}(\bar{\omega}_k^i, \bar{m}_k, \bar{\alpha}_{t_k}^{i, \pi})(t - t_k) + \sigma_{t_k}(\bar{\omega}_k^i, \bar{m}_k)(W_t^i - W_{t_k}^i),$$

where

$$\bar{\omega}_k^i := \ell_{\pi, k}(\bar{X}^{i, \pi}), \quad \bar{\omega}_k := (\bar{\omega}_k^1, \dots, \bar{\omega}_k^N), \quad \bar{m}_k := m_{\bar{\omega}_k}^N.$$

Then $\tilde{X}_{t_k}^{i, \pi} = \bar{X}_{t_k}^{i, \pi}$ for every k . Moreover, on each interval $[t_k, t_{k+1}]$, the difference between $\tilde{X}^{i, \pi}$ and the affine interpolation of the Euler values is a Brownian bridge multiplied by the frozen diffusion coefficient. It follows from (C.2) that

$$\sup_{t \leq T} \mathbb{E} \left[m_2^N(\tilde{\mathbf{X}}_t^\pi - \ell_\pi(\bar{\mathbf{X}}^\pi)_t) \right] + \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|\tilde{X}_t^{i, \pi} - \ell_\pi(\bar{X}^{i, \pi})|^2 \right] \leq C(1 + m_2^N(\mathbf{x}))|\pi|. \quad (\text{C.4})$$

We next compare $\mathbf{X}$ and $\tilde{\mathbf{X}}^\pi$. Put

$$e_k := \mathbb{E} \left[m_2^N(\mathbf{X}_{t_k} - \tilde{\mathbf{X}}_{t_k}^\pi) \right], \quad k \in [p].$$

The standard increment estimates and the definition of the affine interpolation imply

$$\sup_{t \in [t_k, t_{k+1}]} \mathbb{E} \left[m_2^N(\mathbf{X}_t - \mathbf{X}_{t_k}) + m_2^N(\tilde{\mathbf{X}}_t^\pi - \tilde{\mathbf{X}}_{t_k}^\pi) \right] \leq C(1 + m_2^N(\mathbf{x}))|\pi|, \quad (\text{C.5})$$

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|X_{\cdot \wedge t_k}^{i, \alpha} - \omega_k^i|_2^2 \right] \leq C(1 + m_2^N(\mathbf{x}))|\pi|, \quad (\text{C.6})$$

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|\omega_k^i - \bar{\omega}_k^i|_2^2 \right] \leq C \left(e_k + |\pi| \sum_{\ell=0}^k e_\ell \right). \quad (\text{C.7})$$

The term e_k in the last display comes from the constant extension of the stopped paths after t_k , while the sum controls their difference on the previous grid intervals.

For $t \in [t_k, t_{k+1}]$, set $\delta X_t^i := X_t^{i,\alpha} - \tilde{X}_t^{i,\pi}$. We decompose the drift difference as

$$\delta b_t^i = r_{k,t}^i + \beta_{k,t}^i + q_k^i,$$

where

$$\begin{aligned} r_{k,t}^i &:= b_t(X_{\cdot \wedge t}^{i,\alpha}, m_{X_{\cdot \wedge t}}^N, \alpha_t^i) - b_{t_k}(\omega_k^i, m_k, \alpha_t^i), \quad \beta_{k,t}^i := b_{t_k}(\omega_k^i, m_k, \alpha_t^i) - b_{t_k}(\omega_k^i, m_k, \bar{\alpha}_{t_k}^{i,\pi}), \\ q_k^i &:= b_{t_k}(\omega_k^i, m_k, \bar{\alpha}_{t_k}^{i,\pi}) - b_{t_k}(\bar{\omega}_k^i, \bar{m}_k, \bar{\alpha}_{t_k}^{i,\pi}). \end{aligned}$$

By (C.3),

$$\mathbb{E} \left[\int_{t_k}^{t_{k+1}} \beta_{k,t}^i dt \middle| \mathcal{F}_{t_k}^N \right] = 0. \quad (\text{C.8})$$

The diffusion difference is

$$\delta \sigma_t^i := \sigma_t(X_{\cdot \wedge t}^{i,\alpha}, m_{X_{\cdot \wedge t}}^N) - \sigma_{t_k}(\bar{\omega}_k^i, \bar{m}_k),$$

and is decomposed through the intermediate point (t_k, ω_k^i, m_k) .

Applying Itô's formula to $|\delta X_t^i|^2$, summing over i , and dividing by N gives

$$e_{k+1} = e_k + \frac{2}{N} \sum_{i=1}^N \mathbb{E} \int_{t_k}^{t_{k+1}} \delta X_t^i \cdot \delta b_t^i dt + \frac{1}{N} \sum_{i=1}^N \mathbb{E} \int_{t_k}^{t_{k+1}} \|\delta \sigma_t^i\|_F^2 dt. \quad (\text{C.9})$$

Assumption 3.1, (3.1), and (C.6)–(C.7) give, uniformly for $t \in [t_k, t_{k+1}]$,

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} [r_{k,t}^i]^2 \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/2}, \quad (\text{C.10})$$

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} [q_k^i]^2 \leq C \left(e_k + |\pi| \sum_{\ell=0}^k e_\ell \right), \quad (\text{C.11})$$

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} [\|\delta \sigma_t^i\|_F^2] \leq C \left((1 + m_2^N(\mathbf{x}))|\pi|^{1/2} + e_k + |\pi| \sum_{\ell=0}^k e_\ell \right). \quad (\text{C.12})$$

Indeed, the squared time increment in the first and third estimates is $|t - t_k|^{1/2}$, while the remaining terms are controlled by the L^2 path projection estimates above.

For the term containing $\beta_{k,t}^i$, write

$$\delta X_t^i = \delta X_{t_k}^i + (\delta X_t^i - \delta X_{t_k}^i).$$

The contribution of the first term vanishes by (C.8), since $\delta X_{t_k}^i$ is $\mathcal{F}_{t_k}^N$ -measurable. For the second one, Cauchy–Schwarz, (C.5), and the linear growth of b yield

$$\left| \frac{2}{N} \sum_{i=1}^N \mathbb{E} \int_{t_k}^{t_{k+1}} (\delta X_t^i - \delta X_{t_k}^i) \cdot \beta_{k,t}^i dt \right| \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{3/2}. \quad (\text{C.13})$$

Young's inequality, (C.5), (C.10), and (C.11) give

$$\frac{2}{N} \sum_{i=1}^N \mathbb{E} \int_{t_k}^{t_{k+1}} \delta X_t^i \cdot (r_{k,t}^i + q_k^i) dt \leq C|\pi|e_k + C(1 + m_2^N(\mathbf{x}))|\pi|^{3/2} \quad (\text{C.14})$$

$$+ C|\pi|^2 \sum_{\ell=0}^k e_\ell. \quad (\text{C.15})$$

Finally, integrating (C.12) over the interval gives the same right-hand side for the diffusion contribution in (C.9). Combining these estimates, we obtain

$$e_{k+1} \leq (1 + C|\pi|)e_k + C(1 + m_2^N(\mathbf{x}))|\pi|^{3/2} + C|\pi|^2 \sum_{\ell=0}^k e_\ell. \quad (\text{C.16})$$

For completeness, the exponent 3/2 comes from the time regularity $|t - t_k|^{1/4}$ after squaring and integrating over an interval of length $|\pi|$; the same order also bounds the product of the centred drift residual with the state increment on the interval.

Let $\bar{e}_k := \max_{0 \leq \ell \leq k} e_\ell$. Since $|\pi| \sum_{\ell=0}^k e_\ell \leq T\bar{e}_k$, the discrete Gronwall inequality applied to (C.16) gives

$$\max_{0 \leq k \leq p} e_k \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/2}. \quad (\text{C.17})$$

Combining (C.5), (C.4), and (C.17), we conclude that

$$\sup_{t \leq T} \mathbb{E} \left[m_2^N(\mathbf{X}_t - \ell_\pi(\bar{\mathbf{X}}^\pi)_t) \right] + \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|X^{i,\alpha} - \ell_\pi(\bar{X}^{i,\pi})|_2^2 \right] \quad (\text{C.18})$$

$$\leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/2}. \quad (\text{C.19})$$

We now compare the rewards. On each interval $[t_k, t_{k+1}]$, decompose

$$\begin{aligned} f_t(X_{\cdot \wedge t}^{i,\alpha}, m_{\mathbf{X}_{\cdot \wedge t}}^N, \alpha_t^i) - f_{t_k}(\bar{\omega}_k^i, \bar{m}_k, \bar{\alpha}_{t_k}^{i,\pi}) &= \left[f_t(X_{\cdot \wedge t}^{i,\alpha}, m_{\mathbf{X}_{\cdot \wedge t}}^N, \alpha_t^i) - f_{t_k}(\omega_k^i, m_k, \alpha_t^i) \right] \\ &\quad + \left[f_{t_k}(\omega_k^i, m_k, \alpha_t^i) - f_{t_k}(\omega_k^i, m_k, \bar{\alpha}_{t_k}^{i,\pi}) \right] \\ &\quad + \left[f_{t_k}(\omega_k^i, m_k, \bar{\alpha}_{t_k}^{i,\pi}) - f_{t_k}(\bar{\omega}_k^i, \bar{m}_k, \bar{\alpha}_{t_k}^{i,\pi}) \right]. \end{aligned}$$

By (C.3), the expectation of the time integral of the middle term is nonpositive. The first and third terms are bounded by global Lipschitz continuity in Assumption 3.1, (3.1), the reconstruction estimates, and (C.18). In particular, their integrated expected absolute values are bounded by $C(1 + m_2^N(\mathbf{x}))^{1/2}|\pi|^{1/4}$. We retain the coarser factor $1 + m_2^N(\mathbf{x})$ in the statement. Hence

$$\begin{aligned} &\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \int_0^T f_t(X_{\cdot \wedge t}^{i,\alpha}, m_{\mathbf{X}_{\cdot \wedge t}}^N, \alpha_t^i) dt \right] \\ &\leq \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{p-1} f_{t_k}(\bar{\omega}_k^i, \bar{m}_k, \bar{\alpha}_{t_k}^{i,\pi}) |\pi| \right] + C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}. \end{aligned} \quad (\text{C.20})$$

For the terminal reward, the first-order coupling gives explicitly

$$\mathbb{E} \left[|g(m_{\mathbf{X}}^N) - g(m_{\ell_\pi(\bar{\mathbf{X}}^\pi)}^N)| \right] \leq L \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|X^i - \ell_\pi(\bar{X}^{i,\pi})|_2^2 \right] \right)^{1/2}.$$

Together with (C.18), this yields

$$\mathbb{E} \left[|g(m_{\mathbf{X}}^N) - g(m_{\ell_\pi(\bar{\mathbf{X}}^\pi)}^N)| \right] \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}. \quad (\text{C.21})$$

Combining (C.20) and (C.21) yields

$$J^N(\mathbf{x}, \alpha) \leq J^{N,\pi}(\mathbf{x}, \bar{\alpha}^\pi) + C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}.$$

Taking the supremum over $\alpha \in \mathcal{A}_N$ gives

$$V_0^N(\mathbf{x}) \leq U_0^{N,\pi}(\mathbf{x}) + C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}. \quad (\text{C.22})$$

Step 2: From a grid control to a continuous-time control. Fix $\alpha^\pi \in \mathcal{A}_N^\pi \subset \mathcal{A}_N$. Let $\mathbf{X}$ be the solution of (3.4) under this control and write $\mathbf{X}^\pi := \mathbf{X}^{\pi, \alpha^\pi}$ for the Euler scheme (3.9). The preceding state comparison can be repeated with the same control in the two systems. The centred residual $\beta_{k,t}^i$ is then absent, so the argument is in fact simpler, and it gives

$$\begin{aligned} \sup_{t \leq T} \mathbb{E} \left[m_2^N(\mathbf{X}_t - \ell_\pi(\mathbf{X}^\pi)_t) \right] + \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|X^i - \ell_\pi(X^{i,\pi, \alpha^\pi})|_2^2 \right] \\ \leq C(1 + m_2^N(\mathbf{x}))|\pi|^{1/2}, \end{aligned}$$

uniformly over $\alpha^\pi \in \mathcal{A}_N^\pi$. The same running and terminal reward estimates, now used in absolute value, imply

$$J^{N,\pi}(\mathbf{x}, \alpha^\pi) \leq J^N(\mathbf{x}, \alpha^\pi) + C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}.$$

Taking the supremum over the grid controls gives

$$U_0^{N,\pi}(\mathbf{x}) \leq V_0^N(\mathbf{x}) + C(1 + m_2^N(\mathbf{x}))|\pi|^{1/4}. \quad (\text{C.23})$$

The conclusion follows from (C.22) and (C.23). ■

Corollary C.2 (Mean field time-discretisation error). *Under Assumption 3.1, the open-loop mean field values in (3.3) and (3.8) satisfy*

$$|V_0(\mu) - U_0^\pi(\mu)| \leq C(1 + m_2(\mu))^{1/2}|\pi|^{1/4}, \quad \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

with a constant independent of the grid.

Proof. The $|\pi|^{1/4}$ estimate is precisely the mean field Euler approximation result of [3, Theorem 3.6]. Their theorem is stated for a fixed initial law, so we only record the dependence on that law which is needed below. Under the globally Lipschitz assumptions used here, the moment estimates entering their proof give, uniformly over admissible controls,

$$\mathbb{E}[|X|_\infty^2] + \mathbb{E}[|\ell_\pi(X^\pi)|_\infty^2] \leq C(1 + m_2(\mu)).$$

Keeping this factor in their state-comparison argument yields a quadratic approximation error bounded by

$$C(1 + m_2(\mu))|\pi|^{1/2}.$$

Since our running and terminal rewards are globally Lipschitz for $|\cdot|_2$ and $\mathcal{W}_{1,2}$, respectively, and $\mathcal{W}_{1,2} \leq \mathcal{W}_{2,2}$, the reward comparison obtained by Cauchy–Schwarz is bounded by

$$C(1 + m_2(\mu))^{1/2}|\pi|^{1/4}.$$

Combining the two comparison inequalities established in [3, Theorem 3.6] provides the claimed estimate. ■

Remark C.3. (i) *The values in Lemma C.1 use open-loop controls. Their identification with particle path feedbacks is given in Proposition D.1. On the mean field side, the corresponding feedbacks select probability measures on $\mathbb{A}$; ordinary deterministic $\mathbb{A}$ -valued feedbacks are not identified with the open-loop class. This distinction does not enter the uniform-in- N time-discretisation estimate.*

- (ii) Corollary C.2 is the specialisation of [3, Theorem 3.6] needed here. The short argument following the statement only makes explicit the dependence of the constant on the initial second moment; no passage to the limit in the particle problem is used.
- (iii) The fact that σ is not controlled is essential for the simple convexity condition (3.6). The diffusion may nevertheless depend on time, on the stopped state path, and on its empirical distribution. If the diffusion were controlled, the Filippov condition and the projection argument would have to incorporate the covariance generated by the control.

D Proof of Lemma 3.2

Throughout this appendix, Assumption 3.1 holds, there is no common noise, and all mean field laws are unconditional. We use the open-loop information structure specified in Section 3.1. In particular, an independent private random seed is available at the initial date.

The plan of the proof is the following:

1. We define randomised counterparts to the discrete time problems (3.8) and (3.10).
2. We show that the value functions of the non-randomised and randomised problems are equal.
3. We show that Theorem 2.4 applies to the value functions of the randomised problems.

D.1 The two randomised control problems

For $k \in [p]$ and $\eta \in \mathcal{P}_1(E_k)$, write

$$\rho_k(\eta) := (\ell_{\pi,k})_{\#}\eta \in \mathcal{P}_1(\mathcal{X}), \quad \mathbf{Q} := \mathcal{P}(\mathbb{A}).$$

The space $\mathbf{Q}$ is equipped with the weak topology and is compact. Fix a Borel sampling map

$$\mathbf{S} : \mathbf{Q} \times [0, 1] \longrightarrow \mathbb{A}, \quad \mathbf{S}(q, U) \sim q \quad \text{when } U \sim \text{Unif}[0, 1].$$

Such a map exists for a standard Borel action space. For $\omega \in E_k$, $\eta \in \mathcal{P}_1(E_k)$, $a \in \mathbb{A}$ and $z \in \mathbb{R}^d$, set

$$F_k^\pi(\omega, \eta, a, z) := \omega_k + |\pi| b_{t_k}(\ell_{\pi,k}\omega, \rho_k(\eta), a) + \sigma_{t_k}(\ell_{\pi,k}\omega, \rho_k(\eta))z. \quad (\text{D.1})$$

The concatenation $(\omega, F_k^\pi(\omega, \eta, a, z))$ belongs to E_{k+1} . Gaussian increments below have law $\lambda_{|\pi|} := \mathcal{N}(0, |\pi|I_d)$.

The mean field problem. For an initial date n , the admissible randomised feedback policies are

$$\mathfrak{Q}_n^\pi := \left\{ \mathbf{q} = (q_k)_{k=n}^{p-1} : q_k : E_k \times \mathcal{P}_1(E_k) \longrightarrow \mathbf{Q} \text{ is Borel measurable} \right\}.$$

Thus, a feedback control is a function of both the path and its law. Fix $m \in \mathcal{P}_1(E_n)$ and $\mathbf{q} \in \mathfrak{Q}_n^\pi$. Assuming $(\Omega, \mathbb{P})$ is large enough, consider an E_n -valued random variable H_n such that $\mathbb{P} \circ H_n^{-1} = m$, as well as mutually independent pairs (Z_{k+1}, U_{k+1}) , $k = n, \dots, p-1$, with distribution $\lambda_{|\pi|} \otimes \text{Unif}[0, 1]$, independent of H_n . We then define the dynamics $(H)_{k=n, \dots, p}$:

$$\begin{aligned} \eta_k &:= \mathcal{L}(H_k), & Q_k &:= q_k(H_k, \eta_k), \quad a_k := \mathbf{S}(Q_k, U_{k+1}), \\ H_{k+1} &:= (H_k, F_k^\pi(H_k, \eta_k, a_k, Z_{k+1})), & \eta_n &= m. \end{aligned} \quad (\text{D.2})$$

The processes H and Q are adapted to the filtration $\hat{\mathbb{F}} = (\hat{\mathcal{F}}_k)_{k=n, \dots, p}$:

$$\hat{\mathcal{F}}_k = \sigma(H_n, (Z_j, U_j) : n < j \leq k).$$

The reward and value of this problem are then defined by:

$$\begin{aligned}\hat{J}_n^\pi(m, \mathbf{q}) &:= \mathbb{E} \left[|\pi| \sum_{k=n}^{p-1} \int_{\mathbb{A}} f_{t_k}(\ell_{\pi,k} H_k, \rho_k(\eta_k), a) Q_k(da) \right] + g(\rho_p(\eta_p)), \\ \hat{U}_n^\pi(m) &:= \sup_{\mathbf{q} \in \mathfrak{Q}_n^\pi} \hat{J}_n^\pi(m, \mathbf{q}).\end{aligned}\tag{D.3}$$

The particle problem. For $N \geq 1$, define

$$\mathfrak{Q}_n^{N,\pi} := \left\{ \mathbf{q}^N = (q_k^N)_{k=n}^{p-1} : q_k^N : E_k^N \longrightarrow \mathbb{Q}^N \text{ is Borel measurable} \right\}.$$

Given $\omega \in E_n^N$ and $\mathbf{q}^N \in \mathfrak{Q}_n^{N,\pi}$, take pairs (Z_{k+1}^i, U_{k+1}^i) with law $\lambda_{|\pi|} \otimes \text{Unif}[0, 1]$, independent across particles and dates, and define

$$\begin{aligned}\mathbf{H}_n &= \omega, & (Q_k^1, \dots, Q_k^N) &:= q_k^N(\mathbf{H}_k), & a_k^i &:= S(Q_k^i, U_{k+1}^i), \\ H_{k+1}^i &:= (H_k^i, F_k^\pi(H_k^i, \mu_{\mathbf{H}_k}^N, a_k^i, Z_{k+1}^i)),\end{aligned}\tag{D.4}$$

with $i \in [N]^*$. We define the filtration $\hat{\mathbb{F}}^N := (\hat{\mathcal{F}}_k^N)_{k=n, \dots, p}$:

$$\hat{\mathcal{F}}_k^N := \sigma((Z_j^i, U_j^i) : 1 \leq i \leq N, n < j \leq k).$$

Both processes $\mathbf{H}$ and $k \mapsto (Q_k^1, \dots, Q_k^N)$ are $\hat{\mathbb{F}}^N$ -adapted; the uniforms $(U_{k+1}^i)_{i \leq N}$ are then revealed before the Gaussian increments $(Z_{k+1}^i)_{i \leq N}$. We then define the reward and value function:

$$\begin{aligned}\hat{J}_n^{N,\pi}(\omega, \mathbf{q}^N) &:= \mathbb{E} \left[\frac{h}{N} \sum_{i=1}^N \sum_{k=n}^{p-1} \int_{\mathbb{A}} f_{t_k}(\ell_{\pi,k} H_k^i, \rho_k(\mu_{\mathbf{H}_k}^N), a) Q_k^i(da) + g(\rho_p(\mu_{\mathbf{H}_p}^N)) \right], \\ \hat{U}_n^{N,\pi}(\omega) &:= \sup_{\mathbf{q}^N \in \mathfrak{Q}_n^{N,\pi}} \hat{J}_n^{N,\pi}(\omega, \mathbf{q}^N).\end{aligned}\tag{D.5}$$

D.2 Equality of the value functions

Proposition D.1. For every $n \in [p]$, $m \in \mathcal{P}_1(E_n)$, $N \geq 1$ and $\omega \in E_n^N$,

$$\hat{U}_n^\pi(m) = U_n^\pi(m), \quad \hat{U}_n^{N,\pi}(\omega) = U_n^{N,\pi}(\omega).\tag{D.6}$$

Proof. Mean field problem. For $\eta \in \mathcal{P}_1(E_k)$, denote

$$\Gamma_k(\eta) := \left\{ \lambda \in \mathcal{P}_1(E_k \times \mathbb{A}) : (\text{pr}_{E_k})_\# \lambda = \eta \right\}.$$

Here $\text{pr}_{E_k} : E_k \times \mathbb{A} \rightarrow E_k$, $\text{pr}_{E_k}(\omega, a) = \omega$, is the history projection. For $\lambda \in \Gamma_k(\eta)$, define

$$\begin{aligned}\Psi_k^\pi(\lambda) &:= \int_{E_k \times \mathbb{A}} \int_{\mathbb{R}^d} \delta_{(\omega, F_k^\pi(\omega, \eta, a, z))} \lambda_h(dz) \lambda(d\omega, da), \\ r_k^\pi(\lambda) &:= |\pi| \int_{E_k \times \mathbb{A}} f_{t_k}(\ell_{\pi,k} \omega, \rho_k(\eta), a) \lambda(d\omega, da).\end{aligned}\tag{D.7}$$

The dynamic programming principle for the open-loop problem reads

$$U_k^\pi(\eta) = \sup_{\lambda \in \Gamma_k(\eta)} \{ r_k^\pi(\lambda) + U_{k+1}^\pi(\Psi_k^\pi(\lambda)) \},\tag{D.8}$$

with terminal condition $U_p^\pi(\eta) = g(\rho_p(\eta))$. Indeed, any $\alpha_{t_k}^\pi$ determines

$$\lambda := \mathbb{P} \circ (H_k, \alpha_{t_k}^\pi)^{-1} \in \Gamma_k(\eta), \quad \eta = \mathbb{P} \circ H_k^{-1}.$$

Conversely, let $\lambda \in \Gamma_k(\eta)$ and write, by disintegration,

$$\lambda(d\omega, da) = \eta(d\omega) q(\omega, da).$$

Let U_{k+1} be a uniform random variable, independent of the current history H_k , and measurable with respect to the private random seed available in the filtration $\mathbb{F}$. Define

$$\alpha_{t_k}^\pi := \mathcal{S}(q(H_k), U_{k+1}).$$

Since H_k and the private seed are available at time t_k , $\alpha_{t_k}^\pi$ is an admissible open-loop action. Moreover,

$$\mathbb{P} \circ (\alpha_{t_k}^\pi \mid H_k)^{-1} = q(H_k, \cdot),$$

and therefore

$$\mathcal{L}(H_k, \alpha_{t_k}^\pi)(d\omega, da) = \eta(d\omega) q(\omega, da) = \lambda(d\omega, da).$$

Thus every $\lambda \in \Gamma_k(\eta)$ can be realised by an admissible open-loop action. This establishes (D.8).

Now, the dynamic programming principle for the randomised problem provides:

$$\begin{aligned} \widehat{U}_k^\pi(\eta) = \sup_{q: E_k \rightarrow \mathcal{P}(\mathbb{A})} \left\{ |\pi| \int_{E_k} \int_{\mathbb{A}} f_{t_k}(\ell_{\pi, k} \omega, \rho_k(\eta), a) q(\omega, da) \eta(d\omega) \right. \\ \left. + \widehat{U}_{k+1}^\pi(\Psi_k^\pi(\eta(d\omega) q(\omega, da))) \right\}, \end{aligned} \quad (\text{D.9})$$

where the supremum is over Borel kernels. By disintegration,

$$\{\eta(d\omega) q(\omega, da) : q : E_k \rightarrow \mathcal{P}(\mathbb{A}) \text{ is Borel-measurable}\} = \Gamma_k(\eta).$$

Hence (D.8) and (D.9) have the same Bellman operator. Since their terminal values coincide, backward induction gives

$$\widehat{U}_k^\pi(\eta) = U_k^\pi(\eta), \quad k = n, \dots, p.$$

N-particle problem. We argue by backward induction. At time p , both values are equal to $g(\rho_p(m_\omega^N))$. Assume that

$$\widehat{U}_{k+1}^{N, \pi} = U_{k+1}^{N, \pi}.$$

Fix $\omega = (\omega^1, \dots, \omega^N) \in E_k^N$, set $\eta^N := m_\omega^N$, and, for $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{A}^N$ and $\mathbf{z} = (z_1, \dots, z_N) \in (\mathbb{R}^d)^N$, define

$$\omega^{\mathbf{a}, \mathbf{z}} := \left((\omega^i, F_k^\pi(\omega^i, \eta^N, a_i, z_i)) \right)_{i=1}^N.$$

By the classical dynamic programming principle for the finite-dimensional particle problem,

$$U_k^{N, \pi}(\omega) = \sup_{\mathbf{a} \in \mathbb{A}^N} \left\{ \frac{h}{N} \sum_{i=1}^N f_{t_k}(\ell_{\pi, k} \omega^i, \rho_k(\eta^N), a_i) + \int_{(\mathbb{R}^d)^N} U_{k+1}^{N, \pi}(\omega^{\mathbf{a}, \mathbf{z}}) \lambda_h^{\otimes N}(d\mathbf{z}) \right\}. \quad (\text{D.10})$$

On the other hand, the randomised Bellman equation gives

$$\begin{aligned} \widehat{U}_k^{N, \pi}(\omega) = \sup_{q \in \mathcal{P}(\mathbb{A})^N} \int_{\mathbb{A}^N} \left[\frac{h}{N} \sum_{i=1}^N f_{t_k}(\ell_{\pi, k} \omega^i, \rho_k(\eta^N), a_i) \right. \\ \left. + \int_{(\mathbb{R}^d)^N} U_{k+1}^{N, \pi}(\omega^{\mathbf{a}, \mathbf{z}}) \lambda_h^{\otimes N}(d\mathbf{z}) \right] \bigotimes_{i=1}^N q^i(da_i), \end{aligned}$$

where we used the induction hypothesis. Since the average of the integrand cannot exceed its supremum,

$$\sup_{\mathbf{q} \in \mathcal{P}(\mathbb{A})^N} \int_{\mathbb{A}^N} \Phi(\mathbf{a}) \bigotimes_{i=1}^N q^i(da_i) \leq \sup_{\mathbf{a} \in \mathbb{A}^N} \Phi(\mathbf{a}).$$

Conversely, for every $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{A}^N$, choosing $q^i = \delta_{a_i}$ yields the reverse inequality. Hence

$$\sup_{\mathbf{q} \in \mathcal{P}(\mathbb{A})^N} \int_{\mathbb{A}^N} \Phi(\mathbf{a}) \bigotimes_{i=1}^N q^i(da_i) = \sup_{\mathbf{a} \in \mathbb{A}^N} \Phi(\mathbf{a}).$$

Together with (D.10), this yields

$$\widehat{U}_k^{N,\pi}(\omega) = U_k^{N,\pi}(\omega),$$

and the conclusion follows by backward induction. ■

D.3 Uniform regularity and moment estimates

Recall that $\mathcal{W}_{1,\infty,n}$ is induced by $|\cdot|_{\infty,n}$ on E_n . The interpolation maps satisfy

$$|\ell_{\pi,n}\omega - \ell_{\pi,n}\omega'|_{\infty} \leq |\omega - \omega'|_{\infty,n}, \quad \mathcal{W}_{1,2}(\rho_n(m), \rho_n(m')) \leq \sqrt{T} \mathcal{W}_{1,\infty,n}(m, m'). \quad (\text{D.11})$$

These bounds are uniform in the number of dates; no inverse estimate from the L^2 path norm to the uniform norm is needed.

Lemma D.2. *There exists a constant C , independent of n, N and π , such that:*

$$|U_n^\pi(m) - U_n^\pi(m')| \leq C \mathcal{W}_{1,\infty,n}(m, m'), \quad \text{for all } m, m' \in \mathcal{P}_1(E_n), \quad (\text{D.12})$$

and

$$|U_n^{N,\pi}(\omega) - U_n^{N,\pi}(\omega')| \leq C \left(\frac{1}{N} \sum_{i=1}^N |\omega^i - (\omega')^i|_{\infty,n}^2 \right)^{1/2} \quad \text{for all } \omega, \omega' \in E_n^N. \quad (\text{D.13})$$

We also have the estimate:

$$\sup_{\alpha^\pi \in \mathcal{A}_N^\pi} \frac{1}{N} \sum_{i=1}^N \mathbb{E}[|X^{i,\pi,\alpha}|_{\infty,p}^2] + \sup_{\mathbf{q}^N \in \mathcal{Q}_n^{N,\pi}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}[|H_p^i|_{\infty,p}^2] \leq C(1 + m_2^N(\omega)), \quad (\text{D.14})$$

where $\mathbf{X}^{\pi,\alpha}$ is defined by (3.9) and $\mathbf{H}$ by (D.4).

Proof. We first establish (D.12). Fix $m, m' \in \mathcal{P}_1(E_n)$ and $\gamma \in \Pi(m, m')$. By richness of the initial sigma-field $\mathcal{G}$, we may assume without loss of generality that there exist $\mathcal{G}$ -measurable random variables ξ, ξ' such that $\mathbb{P} \circ (\xi, \xi')^{-1} = \gamma$, with $\mathcal{G}$ independent of the Brownian motion. For a given $\alpha \in \mathcal{A}^\pi$, we then denote by X and X' the dynamics (3.7) controlled by α and starting from ξ and ξ' , respectively. Then, for $k \geq n$, set

$$R_k := \max_{0 \leq j \leq k} |X_{t_j}^\pi - X_{t_j}'^\pi|, \quad e_k := \mathbb{E}[R_k],$$

and

$$d_k := \mathcal{W}_{1,2}(\mathcal{L}(\ell_{\pi,k}(X^\pi)), \mathcal{L}(\ell_{\pi,k}(X'^\pi))).$$

Since $(X_{t_0}^\pi, \dots, X_{t_n}^\pi) = \xi$ and $(X_{t_0}'^\pi, \dots, X_{t_n}'^\pi) = \xi'$,

$$e_n = \mathbb{E}[|\xi - \xi'|_{\infty,n}] =: D_\gamma.$$

Using the coupling given by the two processes themselves,

$$d_k \leq \mathbb{E}[|\ell_{\pi,k}(X^\pi) - \ell_{\pi,k}(X'^\pi)|_2] \leq \sqrt{T} e_k.$$

Since the two systems use the same open-loop control, their actions coincide at every grid date. Hence, writing

$$\begin{aligned} \delta b_k &:= b_{t_k}(\ell_{\pi,k}(X^\pi), \mathcal{L}(\ell_{\pi,k}(X^\pi)), \alpha_{t_k}^\pi) - b_{t_k}(\ell_{\pi,k}(X'^\pi), \mathcal{L}(\ell_{\pi,k}(X'^\pi)), \alpha_{t_k}^\pi), \\ \delta \sigma_k &:= \sigma_{t_k}(\ell_{\pi,k}(X^\pi), \mathcal{L}(\ell_{\pi,k}(X^\pi))) - \sigma_{t_k}(\ell_{\pi,k}(X'^\pi), \mathcal{L}(\ell_{\pi,k}(X'^\pi))), \end{aligned}$$

the Lipschitz assumptions give

$$|\delta b_k| + \|\delta \sigma_k\|_F \leq C(R_k + d_k).$$

For $k \geq n$,

$$X_{t_k}^\pi - X_{t_k}'^\pi = \xi_n - \xi_n' + h \sum_{j=n}^{k-1} \delta b_j + \sum_{j=n}^{k-1} \delta \sigma_j \Delta W_{j+1}.$$

Therefore, by the Burkholder–Davis–Gundy inequality,

$$e_k \leq D_\gamma + Ch \sum_{j=n}^{k-1} e_j + C \mathbb{E} \left[\left(h \sum_{j=n}^{k-1} R_j^2 \right)^{1/2} \right] + C \left(h \sum_{j=n}^{k-1} e_j^2 \right)^{1/2}.$$

Since $R_j \leq R_k$ and $e_j \leq e_k$ for $j \leq k$, Young's inequality and the monotonicity of $j \mapsto e_j$ give, for every $\theta > 0$,

$$\begin{aligned} \mathbb{E} \left[\left(h \sum_{j=n}^{k-1} R_j^2 \right)^{1/2} \right] &\leq \theta e_k + \frac{h}{4\theta} \sum_{j=n}^{k-1} e_j, \\ \left(h \sum_{j=n}^{k-1} e_j^2 \right)^{1/2} &\leq \theta e_k + \frac{h}{4\theta} \sum_{j=n}^{k-1} e_j. \end{aligned}$$

Choosing θ sufficiently small and absorbing the corresponding terms into the left-hand side yields

$$e_k \leq 2D_\gamma + Ch \sum_{j=n}^{k-1} e_j.$$

The discrete Gronwall inequality therefore provides

$$\mathbb{E} \left[\max_{0 \leq k \leq p} |X_{t_k}^\pi - X_{t_k}'^\pi| \right] \leq CD_\gamma, \tag{D.15}$$

where C is independent of the grid, the initial date, and the control. Then, the Lipschitz assumptions on f and g yield

$$\begin{aligned} &|J_n^\pi(m, \alpha^\pi) - J_n^\pi(m', \alpha^\pi)| \\ &\leq Lh \sum_{k=n}^{p-1} \left(\mathbb{E}[|\ell_{\pi,k}(X^\pi) - \ell_{\pi,k}(X'^\pi)|_2] + d_k \right) + Ld_p. \end{aligned}$$

Using $d_k \leq \sqrt{T} e_k$ and

$$|\ell_{\pi,k}(X^\pi) - \ell_{\pi,k}(X'^\pi)|_2 \leq \sqrt{T} R_k,$$

together with (D.15), we obtain

$$|J_n^\pi(m, \alpha^\pi) - J_n^\pi(m', \alpha^\pi)| \leq CD_\gamma.$$

The estimate is uniform in α^π , hence

$$|U_n^\pi(m) - U_n^\pi(m')| \leq CD_\gamma.$$

Taking the infimum over $\gamma \in \Pi(m, m')$ gives

$$|U_n^\pi(m) - U_n^\pi(m')| \leq C\mathcal{W}_{1,\infty,n}(m, m').$$

To prove (D.13), we use the same open-loop vector of controls and the same Brownian motions in both particle systems. The coupling with matching particle labels gives

$$\mathcal{W}_{1,2}^2(m_{\ell_{\pi,k}(\mathbf{X}^\pi)}^N, m_{\ell_{\pi,k}(\mathbf{X}'^\pi)}^N) \leq \frac{1}{N} \sum_{i=1}^N |\ell_{\pi,k}(X^{i,\pi}) - \ell_{\pi,k}(X'^{i,\pi})|_2^2.$$

Componentwise martingale estimates followed by averaging and Gronwall yield

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\max_{0 \leq k \leq p} |X_{t_k}^{i,\pi} - X_{t_k}'^{i,\pi}|^2 \right] \leq \frac{C}{N} \sum_{i=1}^N |\omega^i - (\omega')^i|_{\infty,n}^2.$$

Comparing the rewards and taking suprema over controls proves (D.13).

We now prove (D.14). We give the details for $\mathbf{X}^{\pi,\alpha}$; the case of H is handled similarly. Fix $\alpha^\pi \in \mathcal{A}_N^\pi$ and, for $k \geq n$, set

$$M_k := \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\max_{0 \leq j \leq k} |X_{t_j}^{i,\pi,\alpha}|^2 \right].$$

In particular, $M_n = m_2^N(\omega)$. For brevity, write

$$\begin{aligned} b_k^i &:= b_{t_k} \left(\ell_{\pi,k}(X^{i,\pi,\alpha}), m_{\ell_{\pi,k}(\mathbf{X}^{\pi,\alpha})}^N, \alpha_{t_k}^{i,\pi} \right), \\ \sigma_k^i &:= \sigma_{t_k} \left(\ell_{\pi,k}(X^{i,\pi,\alpha}), m_{\ell_{\pi,k}(\mathbf{X}^{\pi,\alpha})}^N \right). \end{aligned}$$

The Lipschitz assumptions imply the uniform linear growth estimate

$$|b_k^i|^2 + \|\sigma_k^i\|_F^2 \leq C \left(1 + |\ell_{\pi,k}(X^{i,\pi,\alpha})|_2^2 + \mathcal{W}_{1,2}^2(m_{\ell_{\pi,k}(\mathbf{X}^{\pi,\alpha})}^N, \delta_0) \right).$$

Now

$$|\ell_{\pi,k}(X^{i,\pi,\alpha})|_2^2 \leq T \max_{0 \leq j \leq k} |X_{t_j}^{i,\pi,\alpha}|^2,$$

while

$$\mathcal{W}_{1,2}^2(m_{\ell_{\pi,k}(\mathbf{X}^{\pi,\alpha})}^N, \delta_0) = \left(\frac{1}{N} \sum_{i=1}^N |\ell_{\pi,k}(X^{i,\pi,\alpha})|_2 \right)^2 \leq \frac{1}{N} \sum_{i=1}^N |\ell_{\pi,k}(X^{i,\pi,\alpha})|_2^2.$$

Consequently,

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|b_k^i|^2 + \|\sigma_k^i\|_F^2 \right] \leq C(1 + M_k). \quad (\text{D.16})$$

For $n \leq \ell \leq k$, the Euler recursion gives

$$X_{t_\ell}^{i,\pi,\alpha} = \omega_n^i + h \sum_{j=n}^{\ell-1} b_j^i + \sum_{j=n}^{\ell-1} \sigma_j^i \Delta W_{j+1}^i.$$

Hence, by Cauchy–Schwarz for the drift sum and the discrete martingale maximal inequality for the stochastic sum,

$$\begin{aligned} M_k &\leq C m_2^N(\omega) + Ch \sum_{j=n}^{k-1} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|b_j^i|^2 + \|\sigma_j^i\|_F^2 \right] \\ &\leq C(1 + m_2^N(\omega)) + Ch \sum_{j=n}^{k-1} M_j, \end{aligned}$$

where we used (D.16) and $h(k-n) \leq T$. Discrete Gronwall therefore yields

$$\sup_{\alpha^\pi \in \mathcal{A}_N^\pi} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[|X^{i,\pi,\alpha}|_{\infty,p}^2 \right] \leq C(1 + m_2^N(\omega)).$$

■

D.4 Derivation of the particle rate

We now put the explicitly defined randomised problems into the framework of Theorem 2.4. The action space is $\mathbf{Q} = \mathcal{P}(\mathbb{A})$ and the idiosyncratic noise is (Z_{k+1}, U_{k+1}) , with law $\lambda_h \otimes \text{Unif}[0, 1]$. There is no common noise. For $\nu \in \mathcal{P}_1(E_k \times \mathbf{Q})$, denote its history marginal by ν_{E_k} and define

$$\begin{aligned} \widehat{F}_k^\pi(\omega, \nu, q, (z, u)) &:= F_k^\pi(\omega, \nu_{E_k}, \mathbf{S}(q, u), z) \\ &= \omega_k + h b_{t_k}(\ell_{\pi,k}\omega, \rho_k(\nu_{E_k}), \mathbf{S}(q, u)) + \sigma_{t_k}(\ell_{\pi,k}\omega, \rho_k(\nu_{E_k}))z, \end{aligned} \quad (\text{D.17})$$

$$\widehat{f}_k^\pi(\omega, \nu, q) := h \int_{\mathbb{A}} f_{t_k}(\ell_{\pi,k}\omega, \rho_k(\nu_{E_k}), a) q(da), \quad g^\pi(\eta) := g(\rho_p(\eta)). \quad (\text{D.18})$$

For the mean field system, $\nu_k = \mathcal{L}(H_k, Q_k)$; for the particle system, $\nu_k^N = N^{-1} \sum_i \delta_{(H_k^i, Q_k^i)}$. Their history marginals are precisely η_k and $\mu_{\mathbf{H}_k}^N$. Thus these transition and reward maps define exactly the two values in (D.3) and (D.5). By Proposition D.1 and Lemma D.2, these values satisfy Assumption 2.3: the mean field values are uniformly Lipschitz on $\mathcal{P}_1(E_n)$, the particle values are continuous, and the particle second-moment bound is uniform in the grid. Theorem 2.4 therefore gives

$$|\widehat{U}_0^{N,\pi}(\mathbf{x}) - \widehat{U}_0^\pi(m_{\mathbf{x}}^N)| \leq C(1 + m_2^N(\mathbf{x})) p N^{-1/[d(p+1)]}.$$

Finally, (D.6) proves (3.11), completing the proof of Lemma 3.2.