Probability Theory and Related Fields
https://doi.org/10.1007/s00440-023-01192-x

Sharp uniform-in-time propagation of chaos

Daniel Lacker[1] · **Luc Le Flem**[1]

Received: 3 June 2022 / Revised: 12 January 2023 / Accepted: 14 January 2023

Abstract
We prove the optimal rate of quantitative propagation of chaos, uniformly in time, for interacting diffusions. Our main examples are interactions governed by convex potentials and models on the torus with small interactions. We show that the distance between the k-particle marginal of the n-particle system and its limiting product measure is $O((k/n)^2)$, uniformly in time, with distance measured either by relative entropy, squared quadratic Wasserstein metric, or squared total variation. Our proof is based on an analysis of relative entropy through the BBGKY hierarchy, adapting prior work of the first author to the time-uniform case by means of log-Sobolev inequalities.

Mathematics Subject Classification Primary 82C22; Secondary 60H10

1 Introduction

This paper studies interacting diffusion processes in $\mathbb{R}^d$ of the form

$$dX_t^i = \left(b_0(t, X_t^i) + \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} b(t, X_t^i, X_t^j)\right)dt + \sigma dW_t^i, \qquad i = 1, \ldots, n, \tag{1.1}$$

D.L. is partially supported by the AFOSR Grant FA9550-19-1-0291 and the NSF CAREER award DMS-2045328. L.L. is partially supported by the same NSF award.

✉ Daniel Lacker
daniel.lacker@columbia.edu

Luc Le Flem
ll3240@columbia.edu

[1] Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, USA

Published online: 03 February 2023

where $W^1, \dots, W^n$ are independent d-dimensional Brownian motions and $\sigma > 0$ is scalar. Precise assumptions are deferred to Sect. 2. Typically, as $n \to \infty$, the limiting behavior of this n-particle system is described in terms of the McKean-Vlasov equation

$$dX_t = (b_0(t, X_t) + \langle \mu_t, b(t, X_t, \cdot)\rangle)\, dt + \sigma\, dW_t, \qquad \mu_t = \mathrm{Law}(X_t), \tag{1.2}$$

or the associated nonlinear Fokker–Planck equation

$$\partial_t \mu_t(x) = -\mathrm{div}_x\big((b_0(t, x) + \langle \mu_t, b(t, x, \cdot)\rangle)\mu_t(x)\big) + (\sigma^2/2)\Delta_x \mu_t(x), \tag{1.3}$$

where we abbreviate $\langle \mu_t, b(t, x, \cdot)\rangle = \int_{\mathbb{R}^d} b(t, x, y)\, \mu_t(dy)$. The large-$n$ limit is formalized by the notion of *propagation of chaos* originating in [37, 42]. Suppose for this introduction that the above equations are well-posed, with the law of $(X_t^1, \dots, X_t^n)$ being exchangeable at $t = 0$ and thus also at each $t > 0$. Let $P_t^{n,k} = \mathrm{Law}(X_t^1, \dots, X_t^k)$ denote the k-particle marginal. The problem of propagation of chaos is to show that $P_t^{n,k} \to \mu_t^{\otimes k}$ weakly as $n \to \infty$, for each $k \in \mathbb{N}$ and $t > 0$, where $\mu_t^{\otimes k}$ denotes the k-fold product measure (assuming that the same is true at $t = 0$, which is trivially the case if the initial positions are i.i.d.). This is well known to be equivalent to the statement that the empirical measure

$$L_t^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_t^i}$$

converges weakly in probability to μ_t as $n \to \infty$, for each $t > 0$.

Over several decades of intensive research, propagation of chaos has been rigorously justified in a wide variety of contexts which we do not attempt to list here, instead referring to the comprehensive recent surveys [12, 13] as well as [34]. While the earliest results were qualitative in nature, recent work has succeeded in *quantifying* propagation of chaos in various ways. Let us first discuss some typical quantitative results which hold on bounded time intervals. The classical synchronous coupling argument, going back to [50] in the case where (b_0, b) are Lipschitz, gives a convergence rate for $\mathbb{E}[\mathcal{W}_2^2(L_t^n, \mu_t)]$ of the same order as if L_t^n were an empirical measure of i.i.d. random variables, where $\mathcal{W}_2$ denotes the quadratic Wasserstein distance. While this estimate deteriorates with the dimension, the same coupling argument yields the dimension-free $\mathcal{W}_2^2(P_t^{n,k}, \mu_t^{\otimes k}) = O(k/n)$, and we note also the recent work [29] that gives dimension-free estimates for $\mathbb{E}[d^2(L_t^n, \mu_t)]$ for certain metrics d dominated by $\mathcal{W}_2$. A recently popular approach is to estimate the relative entropy $H(P_t^{n,n} \,|\, \mu_t^{\otimes n})$ of the full n-particle system; if it is shown to be bounded with respect to n, then the subadditivity of entropy yields $H(P_t^{n,k} \,|\, \mu_t^{\otimes k}) = O(k/n)$, or by Pinsker's inequality $\|P_t^{n,k} - \mu_t^{\otimes k}\|_{\mathrm{TV}}^2 = O(k/n)$. Relative entropy methods have gained popularity in part due to their ability to handle models with singular interactions relevant in physics, which were inaccessible by other methods [8, 33, 35].

The recent paper [38] of the first named author showed that the previous three estimates can actually be improved from $O(k/n)$ to $O((k/n)^2)$ in many cases, with a simple Gaussian example showing that the latter cannot be further improved, thus

obtaining the sharp rate of propagation of chaos. The broad class of models treated in [38] includes bounded or uniformly continuous interaction functions, among others, but nothing too singular. The novelty of the approach of [38], which is also based on relative entropy estimates, is its *local* character, estimating $H_t^k := H(P_t^{n,k} \mid \mu_t^{\otimes k})$ hierarchically in terms of H_t^{k+1} for each $k < n$ rather than estimating the *global* quantity H_t^n. Interestingly, both of the estimates $H_t^n = O(1)$ and $H_t^k = O((k/n)^2)$ are typically sharp, which means that something is lost in using subadditivity to pass from global to local bounds. Note by Pinsker's inequality that the latter estimate implies that $P_t^{n,k}$ has total variation distance at most $O(k/n)$ from the i.i.d. measure $\mu_t^{\otimes k}$, which is interesting to contrast with the well known result of Diaconis-Freedman [20, Theorem 13] about finite exchangeable measures on general spaces; the Diaconis-Freedman bound, which is sharp at its level of generality, implies merely that the total variation distance between $P_t^{n,k}$ and *some mixture* of i.i.d.'s is $O(k^2/n)$.

In this paper, we adapt the $O((k/n)^2)$ estimate of [38] to be *uniform in time*, under an additional assumption of a log-Sobolev inequality (LSI) for μ_t, uniformly in t. Before discussing the results in more detail, let us briefly review some typical prior uniform-in-time results, which were all obtained by *global* methods. First, it is an important and well known fact that McKean-Vlasov equations may admit multiple invariant measures even when the n-particle system is uniquely ergodic for each n, which precludes a uniform-in-time convergence; see [16, 31] for examples. Early results on uniform-in-time propagation of chaos appeared in [40, 41], based on the synchronous coupling approach combined with a uniform-in-n LSI for the n-particle system, in the case of uniformly convex potentials:

$$b_0(t,x) = -\nabla U(x), \quad b(t,x,y) = -\nabla W(x-y), \quad \nabla^2 U \succeq \alpha I, \ \nabla^2 W \succeq 0, \ \alpha > 0. \tag{1.4}$$

In particular, it is shown in [40, 41] that $\mathcal{W}_2^2(P_t^{n,k}, \mu_t^{\otimes k})$ and $H(P_t^{n,k} \mid \mu_t^{\otimes k})$ are both $O(k/n)$, uniformly in time. This approach was extended in [11] to relax the uniform convexity to mere convexity-at-infinity. More recent developments have managed to obtain similar rates under weaker convexity requirements, as long as the interaction strength is small or the temperature σ sufficiently large [17, 21, 47, 48]. A different but still global method is adopted in the recent paper [18] dealing with models on the torus $\mathbb{T}^d$ via analysis of a PDE ("master equation") set on $\mathcal{P}(\mathbb{T}^d)$, obtaining estimates like $\sup_{t\geq 0} |\mathbb{E}F(L_t^n) - F(\mu_t)| = O(1/n)$ for sufficiently smooth functionals F. The recent papers [26, 45] treat specific physically relevant models with singular interactions, again working globally and thus not obtaining our optimal rate.

A noteworthy source of recent applications beyond physics is the analysis of large neural networks trained by stochastic gradient descent [14, 43, 49], which is well approximated by particle systems of the form (1.1) and for which long-time behavior is particularly important. On a more mathematical note, uniform-in-time propagation of chaos can also be used to derive the rate of convergence of μ_t to equilibrium as $t \to \infty$, as in [11, 40, 41].

The main contribution of this paper is to identify the sharp rate of propagation of chaos, along with a methodology that is amenable to further development. On purely

qualitative grounds, though, we do not claim to cover any notable models for which propagation of chaos was not already known. As in [38], we state general results (Theorems 2.1 and 2.5) under somewhat abstract assumptions, which highlight the key a priori estimates one needs in order to implement our method. Most important among them are that μ_t obeys a transport-type inequality, which already appeared in [38], as well as a LSI, with constants uniform in time; the latter was not needed in the finite-time results of [38]. The main examples that we cover are convex-at-infinity potentials as in (1.4) with $\nabla^2 W$ also being bounded from above, and certain models on the torus with $b_0 \equiv 0$ and $b(t,x,y) = K(x-y)$ for a vector field K of small divergence.

The role of the LSI in our work warrants further discussion. Starting from [40], a common ingredient in a proof of uniform-in-time propagation of chaos is a uniform-in-n LSI for the invariant measure of the n-particle system. The remarkable recent result of [19, Theorem 3.7] shows that this uniform-in-n LSI implies uniform-in-time propagation of chaos, with minimal additional assumptions. See also the recent paper [28] on quantitative uniform-in-time propagation of chaos for kinetic models, which applies recent uniform LSIs for invariant measures developed in [27]. Based similarly on a LSI for $P_t^{n,n}$ uniform in n and t (plus other non-trivial assumptions), we obtain in Theorem 2.5 an estimate on the reversed entropy $H(\mu_t^{\otimes k} \mid P_t^{n,k}) = O((k/n)^2)$, and note that this optimal rate $O((k/n)^2)$ cannot be recovered from the aforementioned results. On the other hand, our first Theorem 2.1 is perhaps more novel in that it proves the optimal rate $H(P_t^{n,k} \mid \mu_t^{\otimes k}) = O((k/n)^2)$ under essentially no assumptions on the n-particle system itself, but instead we need a sufficiently high temperature and a LSI for the measure μ_t uniformly in t. See Theorem 2.1 for the general result, and Corollary 2.7 and 2.9 for the examples of convex interactions and small interactions on the torus, respectively. Among prior work, the closest to our approach appears to be [26], which derives the uniform-in-time estimate $H(P_t^{n,n} \mid \mu_t^{\otimes n}) = O(1)$ for certain singular models on the torus by using similarly a LSI for μ_t which is uniform in t; by subadditivity, they deduce $H(P_t^{n,k} \mid \mu_t^{\otimes k}) = O(k/n)$ in [26, Corollary 1].

Let us briefly highlight the main new ideas of the method. We employ a well known relative entropy calculation, formally stated as follows. Suppose $(\mu_t^i)_{t\geq 0}$ is a probability measure flow solving the Fokker–Planck equation

$$\partial_t \mu_t^i = -\mathrm{div}(b_t^i \mu_t^i) + (\sigma^2/2)\Delta \mu_t^i,$$

for some (time-dependent) vector field b_t^i, for $i = 1, 2$. Then

$$\frac{d}{dt} H(\mu_t^1 \mid \mu_t^2) = \int \left((b_t^1 - b_t^2) \cdot \nabla \log(d\mu_t^1/d\mu_t^2) - (\sigma^2/2) \left| \nabla \log(d\mu_t^1/d\mu_t^2) \right|^2 \right) d\mu_t^1. \tag{1.5}$$

See Lemma 3.1 for a rigorous version. For many purposes it is good enough to immediately bound the right-hand side of (1.5) by $(1/2\sigma^2)\|b_t^1 - b_t^2\|^2_{L^2(\mu_t^1)}$, which is actually precisely the time-derivative of the path-space relative entropy used in [38]. Here, we avoid completing the square, instead using the LSI in the natural way to take advantage of the final term in (1.5). More specifically, we apply (1.5) with b^1 and b^2 respectively

being the drifts of $P_t^{n,k}$ and $\mu_t^{\otimes k}$, the former identified using the BBGKY hierarchy. The $b^1 - b^2$ term is estimated similarly to [38], and we ultimately obtain a differential inequality, central to our approach, of the form

$$\frac{d}{dt} H_t^k \leq -c_1 H_t^k + c_2 \frac{k^3}{n^2} + c_3 k (H_t^{k+1} - H_t^k), \quad \text{where} \quad H_t^k := H(P_t^{n,k} \,|\, \mu_t^{\otimes k}), \tag{1.6}$$

for certain positive constants c_1, c_2, c_3 which do not depend on (n, k). The key difference with [38] is the first term, which stems from the LSI and provides the additional decay needed to obtain uniform-in-time bounds. Once (1.6) is established and Gronwall's inequality is applied, the remainder of the proof proceeds by iterating the resulting integral inequality from k to n, using the crude global estimate $\sup_t H_t^n = O(n)$ for the last step. Compared to the finite-time setting of [38], we face new difficulties in estimating the multiple integrals arising from this iteration in a sharp enough manner to produce the optimal exponent.

Two recent papers developed related quantitative estimates along the BBGKY hierarchy, using (weighted) L^p norms rather than relative entropy, for non-exchangeable models in [32] and for certain singular interactions in [7]. Lastly, the very recent [30] adapts the methods of [38] to handle certain singular interactions. These results are all on finite time horizons.

In Sect. 2 below we state precisely all of our main results. Sections 3 and 4 prove the two mains Theorems 2.1 and 2.5, respectively, and Sect. 5 proves the corollaries.

2 Main result and examples

The space of Borel probability measures on a metric space E is denoted $\mathcal{P}(E)$. We use the notation $\langle \mu, f \rangle = \int_E f \, d\mu$ for integration. For $\mu \in \mathcal{P}(E)$ and $k \in \mathbb{N}$, we write $\mu^{\otimes k}(dx_1, \ldots, dx_k) = \mu(dx_1) \cdots \mu(dx_k)$ for the product measure. We use the notation $x = (x_1, \ldots, x_k)$ for general element $x \in E^k$. For example, if $\varphi : E^k \mapsto \mathbb{R}$, $\varphi(x)$ and $\varphi(x_1, \ldots, x_k)$ denote the same quantity. Similarly, $\mu(dx)$ and $\mu(dx_1, \ldots, dx_k)$ are equivalent notations for a measure $\mu \in \mathcal{P}(E^k)$. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^k)$, we define the relative entropy and the Fisher Information between μ and ν respectively by

$$H(\nu \,|\, \mu) := \int_{\mathbb{R}^k} \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu, \qquad I(\nu \,|\, \mu) := \int_{\mathbb{R}^k} \left| \nabla \log \frac{d\nu}{d\mu} \right|^2 d\nu.$$

We set $H(\nu \,|\, \mu) := \infty$ when $\nu \not\ll \mu$, and similarly $I(\nu \,|\, \mu) := \infty$ when $\nu \not\ll \mu$ or $\nabla \log d\nu/d\mu$ does not exist in $L^2(\nu)$. Many measures on Euclidean space encountered in this paper are absolutely continuous with respect to Lebesgue measure. For such measures, we abuse notation by using the same letter to denote both a measure and its density, e.g., $\mu(dx) = \mu(x)dx$. Finally, we define $C_c^\infty(\mathbb{R}^k)$ to be the space of infinitely differentiable functions with compact support on $\mathbb{R}^k$.

2.1 General setup and main result

Fix $d, n \in \mathbb{N}$. The n-particle system of interest (1.1) is described by a weakly continuous flow of probability measures $(P_t^n)_{t\geq 0}$ on $(\mathbb{R}^d)^n$ satisfying the following Fokker–Planck equation, written in weak form: For every $\varphi \in C_c^\infty((\mathbb{R}^d)^n)$ and $t \geq 0$,

$$\begin{aligned}&\langle P_t^n - P_0^n, \varphi\rangle \\ &\quad = \int_0^t \int_{(\mathbb{R}^d)^n} \left[\sum_{i=1}^n \left(b_0(s, x_i) + \frac{1}{n-1}\sum_{j=1, j\neq i}^n b(s, x_i, x_j)\right)\cdot \nabla_{x_i}\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)\right] \\ &\quad P_s^n(dx)ds. \end{aligned} \tag{2.1}$$

The mean field limit is described by a continuous flow of probability measures $(\mu_t)_{t\geq 0}$ on $\mathbb{R}^d$ satisfying the following nonlinear Fokker–Planck equation, again written in weak form: For every $\varphi \in C_c^\infty(\mathbb{R}^d)$ and $t \geq 0$,

$$\langle \mu_t - \mu_0, \varphi\rangle = \int_0^t \int_{\mathbb{R}^d} \left[\left(b_0(s, x) + \langle \mu_s, b(s, x, \cdot)\rangle\right)\cdot \nabla\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)\right]\mu_s(dx)ds. \tag{2.2}$$

Recall that a function on a metric space is said to be *locally bounded* if its restriction to any bounded set is bounded. A probability measure on $(\mathbb{R}^d)^n$ is *exchangeable* if it is invariant under permutations of its n coordinates. Our first set of assumptions is technical in nature:

Assumption E

(E.1) We are given a scalar $\sigma > 0$, and Borel measurable functions $b_0 : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $b : [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, where b_0 is locally bounded.

(E.2) There exists a weak solution $(\mu_t)_{t\geq 0}$ to the nonlinear Fokker–Planck equation (2.2) such that $b(t, x, \cdot) \in L^1(\mu_t)$ for all (t, x), and $(t, x) \mapsto \langle \mu_t, b(t, x, \cdot)\rangle$ is locally bounded.

(E.3) There exists a weak solution $(P_t^n)_{t\geq 0}$ to the Fokker-Planck equation (2.1) such that P_t^n is exchangeable for each $t > 0$. Moreover, for each $p, T > 0$,

$$\int_0^T \int_{(\mathbb{R}^d)^n} \left(|b(t, x_1, x_2)|^p + |\langle \mu_t, b(t, x_1, \cdot)\rangle|^2\right) P_t^n(dx)dt < \infty,$$

$$\sup_{t\in[0,T]} \int_{(\mathbb{R}^d)^n} \left(|b_0(t, x_1)|^2 + |b(t, x_1, x_2)|^2\right) P_t^n(dx) < \infty.$$

Note that we prefer to *assume* the existence of P_t^n and μ_t, rather than placing assumptions on (b_0, b) which imply existence. The assumptions on (b_0, b) are thus mostly implicit, which makes our main result fairly general. Sections 2.3 and 2.4 give more concrete sufficient conditions, stated directly in terms of (b_0, b). The assumptions of local boundedness in (E.1,2) and the p-integrability condition in (E.3) are purely technical, serving only to justify a relative entropy estimate (see Lemma 3.1 below). The next set of assumptions is the more essential one.

Assumption A

(A.1) Log-Sobolev inequality (LSI): There exists a constant $\eta > 0$ such that

$$H(\nu \,|\, \mu_t) \le \eta I(\nu \,|\, \mu_t), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ t \ge 0. \tag{2.3}$$

(A.2) Transport-type inequality: There exists $\gamma > 0$ such that

$$|\langle \nu - \mu_t, b(t, x, \cdot)\rangle|^2 \le \gamma H(\nu \,|\, \mu_t), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ x \in \mathbb{R}^d,\ t \ge 0. \tag{2.4}$$

(A.3) L^2-boundedness: We have

$$M := \operatorname*{esssup}_{t \ge 0} \int_{(\mathbb{R}^d)^n} |b(t, x_1, x_2) - \langle \mu_t, b(t, x_1, \cdot)\rangle|^2 \, P_t^n(dx) < \infty. \tag{2.5}$$

Define $P_t^k \in \mathcal{P}((\mathbb{R}^d)^k)$ to be the k-particle marginal of P_t^n. That is,

$$\langle P_t^k, \varphi\rangle = \int_{(\mathbb{R}^d)^n} \varphi(x_1, \ldots, x_k) P_t^n(dx_1, \ldots, dx_n),$$

for bounded measurable $\varphi : (\mathbb{R}^d)^k \to \mathbb{R}$. For brevity, and because n can be considered as fixed in the following non-asymptotic results, we write P_t^k instead of $P_t^{n,k}$ as in the introduction. The following is our first main result.

Theorem 2.1 *Suppose Assumptions E and A hold. Let* $r_c := \frac{\sigma^4}{4\gamma\eta} - 1$.

(1) *Suppose that* $r_c > 1$ *and that there exists a constant* $C_0 > 0$ *such that*

$$H(P_0^k \,|\, \mu_0^{\otimes k}) \le C_0 (k/n)^2, \quad \text{for all } k = 1, \ldots, n.$$

Then there exists a constant $C > 0$ *depending only on* $(\sigma, \gamma, \eta, M, C_0)$ *such that*

$$H(P_T^k \,|\, \mu_T^{\otimes k}) \le C (k/n)^2, \quad \text{for all } T \ge 0,\ k = 1, \ldots, n.$$

(2) *Suppose that* $0 < r_c \le 1$ *and that for each* $0 < \epsilon_1 < \epsilon_2 < r_c$ *there exists a constant* $C_0^{\epsilon_1, \epsilon_2} > 0$ *such that*

$$H(P_0^k \,|\, \mu_0^{\otimes k}) \le C_0^{\epsilon_1, \epsilon_2} k^{1 + r_c - \epsilon_1} / n^{2r_c - \epsilon_2}, \quad \text{for all } k = 1, \ldots, n.$$

Then, for any $0 < \epsilon_1 < \epsilon_2 < r_c$, *there exists a constant* $C > 0$ *depending only on* $(\sigma, \gamma, \eta, M, \epsilon_1, \epsilon_2, C_0^{\epsilon_1, \epsilon_2}, C_0^{(\epsilon_2 - \epsilon_1)/2, \epsilon_2 - \epsilon_1})$ *such that:*

$$H(P_T^k \,|\, \mu_T^{\otimes k}) \le C k^{1 + r_c - \epsilon_1} / n^{2r_c - \epsilon_2}, \quad \text{for all } T \ge 0,\ k = 1, \ldots, n.$$

Note that Theorem 2.1 only applies at sufficiently *high temperature*, as it does not cover the case $\sigma^4 \leq 4\gamma\eta$ (equivalently, $r_c \leq 0$). For high enough temperature, $\sigma^4 > 8\gamma\eta$, Theorem 2.1(1) achieves the optimal order $(k/n)^2$. See [38, Example 2.8 and Section 3] for a simple example (with linear coefficients) showing that the exponent 2 cannot be improved. In the intermediate regime $4\gamma\eta < \sigma^4 \leq 8\gamma\eta$, we do not know if our rate is sharp.

The assumption $H(P_0^k \,|\, \mu^{\otimes k}) \leq C_0(k/n)^2$ holds trivially in the case of i.i.d. initial conditions $P_0^n = \mu_0^{\otimes n}$. See the main results of [39] for natural families of Gibbs measures for which this assumption holds non-trivially. The non-degenerate noise $\sigma > 0$ will ensure that the measures P_t^n and μ_t are absolutely continuous for each $t > 0$ (see Remark 3.2), but they need not be at $t = 0$; in particular, Theorem 2.1 can accommodate Dirac initial conditions.

Remark 2.2 For small temperature, uniform-in-time propagation of chaos can fail, for the simple reason that the mean field equation may admit multiple invariant measures which all satisfy a LSI. In other words, the presence of Assumptions A and E alone are not enough to guarantee uniform-in-time propagation of chaos, without an additional smallness condition. For a simple example, consider the Kuramoto model

$$dX_t^i = \frac{K}{n-1}\sum_{j=1}^{n} \sin(X_t^i - X_t^j)dt + dB_t^i, \tag{2.6}$$

where $K > 0$ is a constant, and particles take values in the circle $\mathbb{R}/2\pi\mathbb{Z} \cong [0, 2\pi]$. This n-particle system is uniquely ergodic. The uniform measure is always an invariant measure for the corresponding mean field model; i.e., $\mu_t(dx) = dx/2\pi$ for all t solves (2.2). In the supercritical case $K > 1$, the mean field limit admits an infinite set S_{MF} of invariant measures, obtained as the rotations of a common density which is bounded from above and below away from zero. See [4] for details. In particular, all invariant measures admit a LSI by the Holley-Stroock perturbation argument [3, Proposition 5.1.6]. One easily checks that Assumptions A and E hold when $\mu_t = \mu$ for all $t \geq 0$ for some $\mu \in S_{\mathrm{MF}}$, but uniform-in-time propagation of chaos cannot hold in the sense of Theorem 2.1 when $K > 1$. If it did, it would lead to the absurd conclusion that the 1-particle marginal of the unique invariant measure of the n-particle system converges to μ, for each $\mu \in S_{\mathrm{MF}}$. However, we highlight the remarkable recent results of [18, Section 4], which show that uniform-in-time propagation of chaos still holds modulo rotations, if one initializes away from the uniform measure which is unstable when $K > 1$.

Remark 2.3 The entropy bounds of Theorem 2.1 imply similar bounds in Wasserstein distance. To be precise, recall first the definition of the p-Wasserstein distance between two measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, for $p \geq 1$:

$$\mathcal{W}_p(\mu, \nu) = \inf_{\pi} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{1/p}, \tag{2.7}$$

where the infimum is over $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ and ν. By a famous theorem of Otto-Villani [44] (see also [25, Theorem 8.12]), the LSI (2.3) implies the quadratic transport inequality

$$\mathcal{W}_2^2(\nu, \mu_t) \le 4\eta H(\nu \,|\, \mu_t), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ t \ge 0. \tag{2.8}$$

The quadratic transport inequality tensorizes [25, Proposition 1.9], in the sense that

$$\mathcal{W}_2^2(\nu, \mu_t^{\otimes k}) \le 4\eta H(\nu \,|\, \mu_t^{\otimes k}), \quad \forall k \in \mathbb{N},\ \nu \in \mathcal{P}((\mathbb{R}^d)^k),\ t \ge 0.$$

In particular, in case (1) of Theorem 2.1, $\sup_{t\ge 0} \mathcal{W}_2(P_t^k, \mu_t^{\otimes k}) = O(k/n)$.

Remark 2.4 We have not optimized or reported a precise value of the constant C in Theorem 2.1, which is complicated and not very informative in general. However, in the case $\sigma^4 > 12\gamma\eta$, it is not difficult to track the constants in our proof to obtain

$$H(P_T^k \,|\, \mu_T^{\otimes k}) \le \big(C_1 + C_2 e^{-\sigma^2 T/24\eta}\big)(k/n)^2, \quad \forall T, k,$$
$$\text{where } C_1 = \frac{10000 M \sigma^4 \gamma^2 \eta}{(1 - 12\gamma\eta\sigma^{-4})^2}, \qquad C_2 = 1250\left(C_0 + \frac{\sqrt{\gamma M C_0}\eta\sigma^{-4}}{1 - 12\gamma\eta\sigma^{-4}}\right)\frac{\sigma^8}{\gamma^2\eta^2}.$$

This reveals that the term containing the constant C_0, which controls the initial entropy, decays as $T \to \infty$. Thus, as one would expect, the effect of the initial condition disappears over long time horizons. By being even more careful in the proof, we can obtain n-dependent constants $C_1(n)$ and $C_2(n)$ in place of C_1 and C_2, which are bounded in n and have the advantage of vanishing as either $\eta \to 0$, $M \to 0$, or $\sigma \to \infty$, for any fixed n (but not uniformly in n). Similar bounds are possible but much more involved in the remaining case $4\gamma\eta < \sigma^4 \le 12\gamma\eta$.

2.2 Reversing the relative entropy estimate

This section presents a similar result for $H(\mu_t^{\otimes k} \mid P_t^k)$. We need a similar set of assumptions as before, essentially with the roles of the measures P^n and μ inverted. We also add a simplifying assumption that the interaction function b is bounded, which is a major limitation; see Remark 2.6 for discussion.

Assumption R

(R.1) We are given a scalar $\sigma > 0$, and Borel measurable functions $b_0 : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $b : [0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, where b is bounded and b_0 is locally bounded.

(R.2) There exists a weak solution $(P_t^n)_{t\ge 0}$ to the Fokker–Planck equation (2.1) such that P_t^n is exchangeable for each $t > 0$.

(R.3) There exists a weak solution $(\mu_t)_{t\ge 0}$ to the nonlinear Fokker–Planck equation (2.2) with

$$\sup_{t\in[0,T]} \int_{\mathbb{R}^d} |b_0(t,x)|^2 \mu_t(dx) < \infty, \quad T > 0. \tag{2.9}$$

(R.4) Log-Sobolev inequality (LSI): There exists a constant $\eta > 0$ such that

$$H(\nu \mid P_t^n) \le \eta I(\nu \mid P_t^n), \quad \forall \nu \in \mathcal{P}((\mathbb{R}^d)^n),\ t \ge 0,\ n \in \mathbb{N}. \tag{2.10}$$

Theorem 2.5 *Suppose that Assumption R holds. Define*

$$p_c = \frac{\sigma^4}{8\eta \| |b|^2 \|_\infty}.$$

(1) *Suppose that $p_c > 2$ and that there exists a constant $C_0 > 0$ such that*

$$H(\mu_0^{\otimes k} \mid P_0^k) \le C_0 (k/n)^2, \quad \text{for all } k = 1, \dots, n.$$

Then there exists a constant $C > 0$ depending only on $(\sigma, \| |b|^2 \|_\infty, \eta, C_0)$ such that

$$H(\mu_T^{\otimes k} \mid P_T^k) \le C (k/n)^2, \quad \text{for all } T \ge 0,\ k = 1, \dots, n.$$

(2) *Suppose that $p_c \le 2$ and that there exist constants $C_0 > 0$ and $\epsilon \in (0, p_c)$ such that*

$$H(\mu_0^{\otimes k} \mid P_0^k) \le C_0 (k/n)^{p_c - \epsilon}, \quad \text{for all } k = 1, \dots, n.$$

Then there exists a constant $C > 0$ depending only on $(\sigma, \| |b|^2 \|_\infty, \eta, C_0, \epsilon)$ such that

$$H(\mu_T^{\otimes k} \mid P_T^k) \le C (k/n)^{p_c - \epsilon}, \quad \text{for all } T \ge 0,\ k = 1, \dots, n.$$

One can compare Theorems 2.5 with Theorem 2.1 by identifying $\gamma = 2\| |b|^2 \|_\infty$, noting that (2.4) holds with this constant by Pinsker's inequality. The advantage of Theorem 2.5 is that it applies no matter the value of $p_c = \sigma^4 / 4\gamma\eta = 1 + r_c$, whereas Theorem 2.1 applies only when $p_c > 1$. Moreover, in the range $1 < p_c \le 2$ the bound $O((k/n)^{p_c - \epsilon})$ obtained in Theorem 2.5 is better than the bound $O(k^{p_c - \epsilon_1} / n^{2(p_c - 1) - \epsilon_2})$ of Theorem 2.1. When $p_c > 2$, the bound of $O((k/n)^2)$ obtained in each theorem is the same. We do not know if the exponent $p_c - \epsilon$ is sharp in Theorem 2.5(2). Let us note, similarly to Remark 2.3, that the results of Theorem 2.5 imply the bound on Wasserstein distance, by marginalizing the LSI (2.10) and using the Otto-Villani theorem to get $\mathcal{W}_2^2(\mu_t^{\otimes k}, P_t^k) \le 4\eta H(\mu_t^{\otimes k} \mid P_t^k)$ for $t \ge 0$.

Remark 2.6 The assumption in Theorem 2.5 that b is bounded is difficult to relax. It could likely be generalized to the following analogue of Assumption (A.2):

$$\left| \langle \mu_t - P_{t,x}^{k+1|k}, b(t, x_1, \cdot) \rangle \right|^2 \le \gamma H\big(\mu_t \mid P_{t,x}^{k+1|k}\big),$$
$$\forall 1 \le k < n,\ x = (x_1, \dots, x_k) \in (\mathbb{R}^d)^k,\ t \ge 0,$$

where $P^{k+1|k}_{t,x}(dx_{k+1})$ denotes the conditional law of X^{k+1}_t given $(X^1_t, \dots, X^k_t) = x$ under P^n_t. These conditional measures do not seem tractable enough to enable a proof of functional inequalities of this form beyond the case of bounded b; cf. [38, Remark 4.11].

These main theorems give recipes for quantitative uniform-in-time propagation of chaos, and the rest of Sect. 2 describes concrete situations in which they apply. There are conceivably many situations in which these conditions can be checked, on a case-by-case basis. We give two somewhat general classes of examples below. The first deals with convex potentials, for which the well known Bakry-Emery framework yields log-Sobolev inequalities along dynamics. The second deals with a class of models set on the torus, where the LSI can be obtained using the Holley-Stroock perturbation lemma after showing that the density of μ_t is bounded from above and below away from zero uniformly in time.

2.3 Convex potentials

Our first Corollary 2.7 provides a sharper rate of convergence than was previously known for the extremely well-studied case of convex potentials. We impose similar assumptions to [11], albeit with more restrictions on the interaction potential W. We write $\succeq$ to denote positive definite (Loewner) order.

Assumption C Assume $b_0(t, x) = -\nabla U(x)$ and $b(t, x, y) = -\nabla W(x - y)$, where U and W are twice continuously differentiable functions satisfying the following:

(C.1) We have $\nabla^2(U + W) \succeq \alpha I$ for some $\alpha > 0$, and each function $\psi = U$ and $\psi = W$ is convex at infinity in the sense that there exist constants $c_1^\psi, c_2^\psi \geq 0$ for which

$$(x - y) \cdot (\nabla\psi(x) - \nabla\psi(y)) \geq c_1^\psi |x - y|^2 - c_2^\psi, \qquad \forall x, y \in \mathbb{R}^d.$$

We require *uniform* convexity at infinity for U, in the sense that $c_1^U > 0$.

(C.2) There exist $C_U, p_U > 0$ such that $|\nabla U(x)| \leq C_U(1 + |x|^{p_U})$ for all $x \in \mathbb{R}^d$.

(C.3) W is even and $\min(L, R) < \infty$, where we define $R := \||\nabla W|\|_\infty$ and $L = \sup_{x \neq y} |\nabla W(x) - \nabla W(y)|/|x - y|$. That is, ∇W is either bounded or Lipschitz (or both).

Corollary 2.7 *Suppose Assumption C holds. Let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ satisfy the LSI*

$$H(\nu \,|\, \mu_0) \leq (\eta_0/4) I(\nu \,|\, \mu_0), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),$$

for some $\eta_0 > 0$. Let $P^n_0 \in \mathcal{P}((\mathbb{R}^d)^n)$ be exchangeable and have finite moments of every order. Then there exists a unique solution $(P^n_t)_{t \geq 0}$ of the Fokker–Planck equation (2.1) starting from P^n_0, and there exists a solution $(\mu_t)_{t \geq 0}$ of the nonlinear Fokker–Planck equation (2.2) starting from μ_0, unique among the class of solutions satisfying $\int_0^T \int_{\mathbb{R}^d} |x|^p \mu_t(dx)dt < \infty$ for all $T, p > 0$. Moreover, the following hold:

(1) *Assumptions* A *and* E *are satisfied, with* $\eta = \max(\eta_0/4, \sigma^2/4\alpha)$, $\gamma = \min(4\eta L^2, 2R^2)$, *and* M *bounded by a finite constant depending only on* L, R, C_U, *and* $\int_{(\mathbb{R}^d)^n} |x_1|^2 P_0^n(dx)$. *In particular, Theorem 2.1 applies with*

$$r_c = \frac{\sigma^4}{4\gamma\eta} - 1 = \max\Big(\min\Big(\frac{\sigma^4}{\eta_0^2 L^2}, \frac{\alpha^2}{L^2}\Big), \min\Big(\frac{\sigma^4}{2\eta_0 R^2}, \frac{\sigma^2\alpha}{2R^2}\Big)\Big) - 1.$$

(2) *If* $R < \infty$ *and (C.1) holds with* $c_2^U = c_2^W = c_1^W = 0$ *and* $c_1^U = \alpha$ *(i.e.,* U *is* α*-uniformly convex and* W *is convex), then Assumption* R *is satisfied with* $\eta = \max(\eta_0/4, \sigma^2/4\alpha)$, *and Theorem 2.5 applies with*

$$p_c = \min\Big(\frac{\sigma^4}{2\eta_0 R^2}, \frac{\sigma^2\alpha}{2R^2}\Big).$$

Note that $\eta_0 = 0$ when μ_0 is a Dirac. In case (1), if $\eta_0 \le \sigma^2/\alpha$ and $R = \infty$, then $r_c = (\alpha/L)^2 - 1$ and the two ranges of r_c in Theorem 2.1 correspond to $\alpha > \sqrt{2}L$ and $L < \alpha \le \sqrt{2}L$, respectively. We are unable to treat the case $\alpha \le L$, instead requiring the convexity to be sufficiently stronger than the interaction strength in the sense that $\alpha > L$. This kind of assumption is common in the literature, more when dealing with non-convex cases [2, 18, 27, 39, 46]. On the other hand, in the case where ∇W is bounded, part (2) has no smallness constraint anymore, except that the optimal exponent of 2 is obtained only when $p_c \ge 2$, or $\sigma^2\alpha \ge 4R^2$.

There are two main limitations in Corollary 2.7. The first is that part (1) requires $\alpha > L$, as discussed just above. The second is that it requires ∇W to be globally Lipschitz or bounded, which in particular rules out relevant cases like $W(x) = |x|^3$. Neither of these assumptions is needed in order to obtain *qualitative* uniform-in-time propagation of chaos, which has been known since [11, 40].

Remark 2.8 Theorem 2.1 and Corollary 2.7 cover the Gaussian case, where $b_0(t, x) = -ax$ and $b(t, x, y) = b(y - x)$, for $x, y \in \mathbb{R}$, with $a, b > 0$. Suppose that $\sigma = 1$ and that $\mu_0 = \delta_0$. It was shown in [38, Example 2.8 and Section 3] that uniform-in-time propagation of chaos holds in this case at a rate of $O((k/n)^2)$, no matter the values of $a, b > 0$. Corollary 2.7 recovers this rate, but only when a/b is large enough. Indeed, this setting fits into Corollary 2.7(1), with $\eta_0 = 0$, $\alpha = a + b$, $L = b$ (and $R = \infty$), which yields the exponent $r_c = (1 + (a/b))^2 - 1$. We recover the optimal $O((k/n)^2)$ only when $r_c > 1$, or $a/b > \sqrt{2} - 1$. Corollary 2.7(1) still applies but yields a suboptimal exponent when $a/b \le \sqrt{2} - 1$.

We note also that [21] recently used coupling methods to obtain quantitative propagation of chaos even when $U + W$ is non-convex, but with a suboptimal rate compared to Corollary 2.7. Their assumptions are somewhat similar otherwise, assuming a smallness condition on the Lipschitz constant of ∇W, though they do not cover the case of non-Lipschitz ∇W.

2.4 Models on the torus

In this section, we present a class of models in which the state space $\mathbb{R}^d$ is replaced by the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, and we take $b_0 \equiv 0$ and $b(t, x, y) = K(x - y)$ for some vector field $K : \mathbb{R}^d \to \mathbb{R}^d$. The proofs of Theorems 2.1 and 2.5 adapt without change to the case of the torus. The McKean-Vlasov SDE and its corresponding PDE take the form

$$dX_t = K * \mu_t(X_t)dt + \sigma dB_t, \tag{2.11}$$

$$\partial_t \mu = -\mathrm{div}(\mu K * \mu) + \frac{\sigma^2}{2}\Delta\mu. \tag{2.12}$$

The n-particle Fokker–Planck equation becomes

$$\partial_t P_t^n(x) = -\sum_{i=1}^n \mathrm{div}_{x_i}\left(\frac{1}{n-1}\sum_{j\neq i} K(x_i - x_j)P_t^n(x)\right) + \frac{\sigma^2}{2}\Delta P_t^n(x). \tag{2.13}$$

We make the following assumptions:

Assumption T Assume that K is Lipschitz and also that the initial law μ_0 admits a smooth density satisfying the pointwise bound $\lambda^{-1} \leq \mu_0 \leq \lambda$, for some $\lambda \geq 1$.

Note since that $\mu \equiv 1$ solves the PDE (2.12), because $\mu K * \mu \equiv \int_{\mathbb{T}^d} K$ is a constant; that is, the uniform measure on $\mathbb{T}^d$ is invariant for (2.12). There may be additional invariant measures, in general. Our main result in this section gives a uniform-in-time propagation of chaos, with a sharp rate, for a sufficiently small mean field interaction K, which in particular rules out the existence of additional invariant measures:

Corollary 2.9 *Suppose Assumption* T *holds, and let* $P_0^n \in \mathcal{P}((\mathbb{T}^d)^n)$ *be arbitrary. Then there exists a unique weak solution* P^n *of* (2.13) *starting from* P_0^n*, and there exists a unique classical solution* μ *of* (2.12) *starting from* μ_0*. Assume that* div K *is small in the sense that*

$$\|\mathrm{div}\, K\|_\infty < \frac{\sigma^2\pi^2}{1 + 2\sqrt{2\log\lambda}} < \sigma^2\pi^2. \tag{2.14}$$

Then Assumptions A *and* E *hold, and Theorem* 2.1 *applies with*

$$r_c = \frac{\sigma^4(1 - 2r_0)}{2\lambda^2 \mathrm{diam}^2(K)} - 1,$$

where we define

$$r_0 := \frac{\|\mathrm{div}\, K\|_\infty\sqrt{2\log\lambda}}{\sigma^2\pi^2 - \|\mathrm{div}\, K\|_\infty}, \quad \mathrm{diam}(K) := \sup_{x,z\in\mathbb{T}^d} |K(x) - K(z)|.$$

Note that (2.14) ensures that $r_0 < 1/2$, so that $r_c > -1$. Theorem 2.1 only yields anything useful, of course, if $r_c > 0$, which can occur when some combination of the following effects are present: σ is large, or K has small oscillations or divergence, or λ is close to 1 which means that the initial law μ_0 is L^∞-close to the uniform distribution. Note when K is divergence-free the simplification $r_0 = 0$ occurs.

Example 2.10 Corollary 2.9 notably covers the example of the Kuramoto model (without disorder) discussed in Remark 2.2, in sufficiently subcritical regimes. We keep this discussion brief, referring to [1, 4] for background on this model. The Kuramoto model (2.6) was written for X_t^i taking values in $[0, 2\pi]$, and to fit it into Corollary 2.9 (set in $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z} \cong [0, 1]$) we simply rescale X_t^i to $X_t^i/2\pi$. This yields the parameters $\sigma = 1/2\pi$ and $K(x) = K \sin(2\pi x)/2\pi$. In Corollary 2.9, the condition 2.14 becomes

$$K < \frac{1}{4 + 8\sqrt{2 \log \lambda}} < \frac{1}{4},$$

and the constant r_c becomes

$$r_c = r_c(K, \lambda) := \frac{1 - 4K(1 + 2\sqrt{2 \log \lambda})}{32\pi^2 \lambda^2 K^2 (1 - 4K)} - 1.$$

We obtain quantitative uniform-in-time propagation of chaos when $K < K_c^0(\lambda)$, where we define $K_c^a(\lambda)$ for $a \geq 0$ as the (λ-dependent) value of K for which $r_c(K, \lambda) = a$. For $K < K_c^1(\lambda)$ we obtain the optimal exponent $O((k/n)^2)$ from Theorem 2.1. For example, when $\lambda = 1$ (meaning the initialization μ_0 is uniform), we have $K_c^0(1) = \sqrt{2}/8\pi$ and $K_c^1(1) = 1/8\pi$. We are not able to determine the sharp rate quantitative uniform-in-time propagation of chaos all the way to criticality, i.e., for all $K < 1$.

3 Proof of the main theorem

This section gives the proof of Theorem 2.1. We suppose that Assumptions E and A hold throughout this section. The central quantity of study will be the relative entropy

$$H_t^k := H(P_t^k \mid \mu_t^{\otimes k}), \quad t \geq 0, \ 1 \leq k \leq n.$$

3.1 The entropy estimate

The first step is to estimate the time derivative of H_t^k. We apply the following more general estimate between solutions of Fokker–Planck equations.

Lemma 3.1 *Let $d \in \mathbb{N}$ and $\sigma > 0$. For each $i = 1, 2$, let $b^i : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ be measurable, and assume we are given a continuous flow of probability measures*

$(\mu_t^i)_{t\geq 0}$ on $\mathbb{R}^d$ satisfying the weak Fokker–Planck equation

$$\langle \mu_t^i, \varphi\rangle = \langle \mu_0^i, \varphi\rangle + \int_0^t \int_{\mathbb{R}^d} \Big(b^i(s,x)\cdot\nabla\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)\Big)\mu_s^i(dx)ds,$$
$$\forall \varphi \in C_c^\infty(\mathbb{R}^d),\ t\geq 0,$$

as well as the following conditions:

(H.1) *The function b^2 is locally bounded.*

(H.2) *The function b^1 belongs to $L^p_{\mathrm{loc}}(\mu^1)$ for some $p > d+2$. That is, for each $T>0$ and each bounded Borel set $S\subset\mathbb{R}^d$ we have*

$$\int_0^T \int_S |b^1(t,x)|^p \mu_t^1(dx)dt < \infty.$$

(H.3) The following hold for each $T>0$:

$$\int_0^T \int_{\mathbb{R}^d} |b^2(t,x)|^2 \mu_t^1(dx)dt < \infty, \qquad \sup_{t\in[0,T]} \int_{\mathbb{R}^d} |b^1(t,x)|^2 \mu_t^1(dx) < \infty.$$

Then, it holds for each $t>s\geq 0$ that

$$H(\mu_t^1 \,|\, \mu_t^2) + \frac{\sigma^2}{4}\int_s^t I(\mu_u^1 \,|\, \mu_u^2)\,du \leq H(\mu_s^1 \,|\, \mu_s^2)$$
$$+ \frac{1}{\sigma^2}\int_s^t \int_{\mathbb{R}^d} \Big|b^1(u,x) - b^2(u,x)\Big|^2 \mu_u^1(dx)du.$$

In particular, if $H(\mu_0^1 \,|\, \mu_0^2) < \infty$, then $H(\mu_t^1 \,|\, \mu_t^2) < \infty$ for all $t>0$.

Estimates of this form might be considered folklore, with similar forms appearing in [8, 33, 35] under different assumptions. If one ignores questions of smoothness and integrability, the proof follows by applying the Fokker–Planck equation for μ^2 with the test function $\varphi = (\mu^1/\mu^2)\log(\mu^1/\mu^2)$. It takes some care, though, to make this rigorous. Lemma 3.1 follows almost immediately from the arguments in [6, Lemma 2.4], with their main results holding under the sole assumptions that b^1 and b^2 are Borel measurable and locally bounded. Local boundedness of b^1 turns out to be too restrictive for our application, where we wish to use $b^1 = \widehat{b^k}$ defined in (3.1) below; the issue is that local boundedness is not preserved under conditioning. Nonetheless, the assumptions of Lemma 3.1 are enough to ensure sufficient smoothness of μ^1, summarized in Remark 3.2 below, which enables the same arguments as in [6] to be carried out, after an additional mollification step in which b^1 is approximated by a locally bounded function. See Appendix A of the first arXiv version of the present paper for full details.

Remark 3.2 The assumptions (H.1,2) ensure that the measure $dt\,\mu_t^i(dx)$ admits a (Hölder) continuous density on $(0,\infty)\times\mathbb{R}^d$, which we denote by $\mu^i(t,x)$, for each

$i = 1, 2$. In addition, $\mu^i(t, \cdot) \in W^{p,1}(U)$ for every $p \geq 1$, every $t > 0$, and every bounded open set $U \subset \mathbb{R}^d$. See [5, Proposition 6.5.1]. Lastly, since b^2 is locally bounded, the continuous version of the density μ^2 is strictly positive on $(0, \infty) \times \mathbb{R}^d$; see [5, Example 8.3.8].

Remark 3.3 An easy adaptation of the proof of Lemma 3.1 gives the following more general estimate, for $0 < c < 1$:

$$H(\mu_t^1|\mu_t^2) + \frac{\sigma^2}{2}(1-c)\int_s^t I(\mu_u^1|\mu_u^2)du \leq H(\mu_s^1|\mu_s^2) + \frac{1}{2\sigma^2 c}\int_s^t \int_{\mathbb{R}^d} \left|b^1(u,x) - b^2(u,x)\right|^2 \mu_u^1(dx)du.$$

This allows one to trade off between the Fisher information and drift terms. Remarkably, it turns out in our context that the optimal choice is always $c = 1/2$.

3.2 The BBGKY hierarchy

We will apply Lemma 3.1 with P_t^k and $\mu_t^{\otimes k}$ in place of μ_t^1 and μ_t^2, respectively, but first we need to represent P^k and $\mu^{\otimes k}$ for each k as solutions of Fokker–Planck equations. For $\mu^{\otimes k}$ this is straightforward. For P^k this is accomplished using the so-called *BBGKY hierarchy*, which is well known, but we derive it in the lemma below for completeness; see [24, Section 1.5] for additional references and a derivation in the zero-noise case. First, note that the measure $dt\,P_t^n(dx)$ has a positive density on $(0, \infty) \times (\mathbb{R}^d)^n$; see Remark 3.2 and note that (H.2) holds because of the local boundedness of b_0 and because of Assumption (E.3). By marginalization, $dt\,P_t^k(dx)$ has a positive density on $(0, \infty) \times (\mathbb{R}^d)^k$. Recall in the following that we use the same letter to denote a measure and its density when it exists, e.g., $P_t^k(dx) = P_t^k(x)dx$.

Lemma 3.4 *Let $1 \leq k < n$. Define the conditional density*

$$P^{k+1|k}_{t,x_1,\ldots,x_k}(x_{k+1}) = \frac{P_t^{k+1}(x_1,\ldots,x_{k+1})}{P_t^k(x_1,\ldots,x_k)}, \quad t > 0,\ x_1,\ldots,x_{k+1} \in \mathbb{R}^d.$$

For $i = 1, \ldots, k$, define the functions $\widehat{b}_i^k : [0, \infty) \times (\mathbb{R}^d)^k \to \mathbb{R}^d$ by

$$\widehat{b}_i^k(t, x_1, \ldots, x_k) := \frac{1}{n-1}\sum_{j\neq i,\, j=1}^{k} b(t, x_i, x_j) + \frac{n-k}{n-1}\langle P^{k+1|k}_{t,x_1,\ldots,x_k}, b(t, x_i, \cdot)\rangle. \tag{3.1}$$

Then, for each $1 \leq k < n$, $\varphi \in C_c^\infty((\mathbb{R}^d)^k)$, and $t \geq 0$,

$$\langle P_t^k - P_0^k, \varphi\rangle = \int_0^t \int_{(\mathbb{R}^d)^k} \Big(\sum_{i=1}^k (b_0(s, x_i) + \widehat{b}_i^k(s, x)) \cdot \nabla_{x_i}\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)\Big) P_s^k(dx)ds.$$

Proof Apply the weak Fokker–Planck equation (2.1) to a test function $\varphi \in C_c^\infty((\mathbb{R}^d)^k)$ depending on only the first k variables to find

$$\begin{aligned}\langle P_t^k - P_0^k, \varphi\rangle = \sum_{i=1}^k \int_0^t \int_{(\mathbb{R}^d)^n} \Bigg(& b_0(s, x_i) \\ & + \frac{1}{n-1} \sum_{j=1, j\neq i}^n b(s, x_i, x_j) \Bigg) \cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^n(dx) ds \\ & + \frac{\sigma^2}{2} \int_0^t \int_{(\mathbb{R}^d)^n} \Delta\varphi(x_1, \ldots, x_k) P_s^n(dx) ds.\end{aligned}$$

Marginalizing yields

$$\begin{aligned}\int_{(\mathbb{R}^d)^n} b_0(s, x_i) \cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^n(dx) &= \int_{(\mathbb{R}^d)^k} b_0(s, x_i) \cdot \nabla_{x_i} \varphi(x)\, P_s^k(dx), \\ \int_{(\mathbb{R}^d)^n} \Delta\varphi(x_1, \ldots, x_k)\, P_s^n(dx) &= \int_{(\mathbb{R}^d)^k} \Delta\varphi(x)\, P_s^k(dx).\end{aligned}$$

For the interaction term, we compute

$$\begin{aligned}&\int_{(\mathbb{R}^d)^n} \frac{1}{n-1} \sum_{j=1, j\neq i}^n b(s, x_i, x_j) \cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^n(dx) \\ &= \frac{1}{n-1} \sum_{j=1, j\neq i}^k \int_{(\mathbb{R}^d)^n} b(s, x_i, x_j) \cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^n(dx) \\ &\quad + \frac{1}{n-1} \sum_{j=k+1}^n \int_{(\mathbb{R}^d)^n} b(s, x_i, x_j) \cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^n(dx) \qquad (3.2)\end{aligned}$$

We claim this is equal to $\int_{(\mathbb{R}^d)^k} \widehat{b}_i^k(t, x) \cdot \nabla_{x_i} \varphi(x)\, P_s^k(dx)$. Again by marginalizing, the first term on the right-hand of (3.2) can be simplified to

$$\frac{1}{n-1} \sum_{j=1, j\neq i}^k \int_{(\mathbb{R}^d)^k} b(s, x_i, x_j) \cdot \nabla_{x_i} \varphi(x)\, P_s^k(dx),$$

Using symmetry, the second term on the right-hand of (3.2) can be simplified to

$$\begin{aligned}&\frac{n-k}{n-1} \int_{(\mathbb{R}^d)^{k+1}} b(s, x_i, x_{k+1}) \cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^{k+1}(dx) \\ &\quad = \frac{n-k}{n-1} \int_{(\mathbb{R}^d)^k} \left[\int_{\mathbb{R}^d} b(s, x_i, x_{k+1})\, P_{s, x_1, \ldots, x_k}^{k+1|k}(dx_{k+1}) \right] \\ &\cdot \nabla_{x_i} \varphi(x_1, \ldots, x_k)\, P_s^k(dx_1, \ldots, dx_k).\end{aligned}$$

The expression in brackets is exactly $\langle P^{k+1|k}_{t,x_1,\dots,x_k}, b(t,x_i,\cdot)\rangle$, and the proof is complete. □

Remark 3.5 Alternatively, working at the level of the stochastic processes (1.1), the BBGKY hierarchy can be derived using the so-called *mimicking theorem* [9, Corollary 3.7].

3.3 Estimating H^k_t

We now apply Lemma 3.1 to estimate $H^k_t = H(P^k_t \,|\, \mu_t^{\otimes k})$. We know from Lemma 3.4 that P^k solves a Fokker–Planck equation with drift $(\widehat{b}^k_1, \dots, \widehat{b}^k_k)$. On the other hand, the nonlinear-Fokker–Planck equation (2.2) easily tensorizes to show that $\mu^{\otimes k}$ satisfies the following Fokker–Planck equation: for all $\varphi \in C^\infty_c((\mathbb{R}^d)^k)$ and $t \geq 0$,

$$\langle \mu_t^{\otimes k} - \mu_0^{\otimes k}, \varphi\rangle = \int_0^t \int_{(\mathbb{R}^d)^k} \Big(\sum_{i=1}^k \big(b_0(s,x_i) + \langle \mu_s, b(s,x_i,\cdot)\rangle\big) \cdot \nabla_{x_i}\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)\Big)\mu_s^{\otimes k}(dx)ds.$$

We now apply Lemma 3.1 to get, for all $t > s \geq 0$,

$$H^k_t - H^k_s + \frac{\sigma^2}{4}\int_s^t I(P^k_u \,|\, \mu_u^{\otimes k})\,du \leq \frac{1}{\sigma^2}\int_s^t \int_{(\mathbb{R}^d)^k} \sum_{i=1}^k \big|\widehat{b}^k_i(u,x) - \langle \mu_u, b(u,x_i,\cdot)\rangle\big|^2 P^k_u(dx)du. \tag{3.3}$$

Indeed, the assumptions of Lemma 3.1 are easily checked using Assumption E: We know by (E.1,2) that $b_0(t,x) + \langle \mu_t, b(t,x,\cdot)\rangle$ is locally bounded, so that (H.1) holds. The assumption (H.2) holds because b_0 is locally bounded, and because (E.3) gives

$$\int_0^T \int_{(\mathbb{R}^d)^k} |\widehat{b}^k_i(t,x)|^p\, P^k_t(dx)dt \leq \int_0^T \int_{(\mathbb{R}^d)^n} |b(t,x_1,x_2)|^p\, P^n_t(dx)dt < \infty,$$

for all $p, T > 0$ and $i = 1, \dots, k$ by Jensen's inequality and by exchangeability. The first part of (H.3) comes from the local boundedness of b_0 and the first part of (E.3). Similarly, the second part of (H.3) follows from the second part of (E.3). With (3.3) now justified, the integrability assumptions in E will henceforth play no role.

The following lemma summarizes the implications of (3.3), which will be used in multiple ways. Recall the constants (η, γ, M) defined in Assumption A.

Lemma 3.6 *Suppose that there exist constants $C_0 > 0$ and $p_1, p_2 \in (0,2]$ such that*

$$H^k_0 \leq C_0 k^{p_1}/n^{p_2}, \quad 1 \leq k \leq n.$$

Let $\delta > 0$, and define $Z := \sigma^2/4\eta$, $\tilde{\gamma} := \gamma(1+\delta)/\sigma^2$, and the quantities

$$\tilde{A}_k^\ell(t_k) := \left(\prod_{j=k}^{\ell} \tilde{\gamma} j\right) \int_0^{t_k} \int_0^{t_{k+1}} \cdots \int_0^{t_\ell} e^{-\sum_{j=k}^{\ell}(Z+\tilde{\gamma} j)(t_j - t_{j+1})} dt_{\ell+1} \ldots dt_{k+1}, \tag{3.4}$$

$$\tilde{B}_k^\ell(t_k) := \left(\prod_{j=k}^{\ell-1} \tilde{\gamma} j\right) \int_0^{t_k} \int_0^{t_{k+1}} \cdots \int_0^{t_{\ell-1}} e^{-(Z+\tilde{\gamma}\ell)t_\ell - \sum_{j=k}^{\ell-1}(Z+\tilde{\gamma} j)(t_j - t_{j+1})} dt_\ell \ldots dt_{k+1}, \tag{3.5}$$

for $t_k \geq 0$ and integers $\ell \geq k$, with the convention $\tilde{B}_k^k(t) := e^{-(Z+\tilde{\gamma}k)t}$. Then the following hold:

(1) *For each $T \geq 0$ and $1 \leq k < n$,*

$$H_T^n \leq C_0 n^{p_1 - p_2} e^{-\sigma^2 T/4\eta} + 4nM\eta\sigma^{-4}. \tag{3.6}$$

$$H_T^k \leq \frac{C_0}{n^{p_2}} \sum_{\ell=k}^{n-1} \ell^{p_1} \tilde{B}_k^\ell(T) + \frac{M}{\delta\gamma n^2} \sum_{\ell=k}^{n-1} \ell^2 \tilde{A}_k^\ell(T) + \tilde{A}_k^{n-1}(T) H_T^n. \tag{3.7}$$

(2) *If we assume additionally that there exist $C_1 > 0$ and $p \in [0, 2]$ such that*

$$\sup_{T \geq 0} H_T^3 \leq C_1/n^p, \tag{3.8}$$

then, by setting $C_2 = 2M + \sqrt{\gamma M C_1}$, it holds for each $T \geq 0$ and $1 \leq k < n$ that

$$H_T^n \leq C_0 n^{p_1 - p_2} e^{-\sigma^2 T/4\eta} + 4n^{1-\frac{p}{2}} C_2 \eta\sigma^{-4}, \tag{3.9}$$

$$H_T^k \leq \frac{C_0}{n^{p_2}} \sum_{\ell=k}^{n-1} \ell^{p_1} \tilde{B}_k^\ell(T) + \frac{C_2}{\delta\gamma n^2} \sum_{\ell=k}^{n-1} \ell^{2-\frac{p}{2}} \tilde{A}_k^\ell(T) + \tilde{A}_k^{n-1}(T) H_T^n. \tag{3.10}$$

Although we have no a priori information about H_T^3, the purpose of part (2) of Lemma 3.6 is as follows. In the case of $r_c > 2$ in Theorem 2.1, we use Lemma 3.6(1) on a first pass through the argument, which ultimately leads to a bound of $H_T^k = O(k^3/n^2)$. This is suboptimal in the exponent of k. However, it implies $H_T^3 = O(1/n^2)$, and we may then repeat the argument and apply the sharper Lemma 3.6(2) with $p = 2$ to reduce the exponent on k by one. This is carried out in Sect. 3.5 and is similar to an argument in [38]. The case where $0 < r_c < 2$ is trickier, because a first pass through the argument gives the worse bound $H_T^3 = O(1/n^{r_c})$. The idea is to then apply Lemma 3.6(2) inductively. The exponent in $H_T^3 = O(1/n^{r_c})$ improves after m iterations from r_c to $2\min(1, r_c(1-2^{-m}))$. If $r_c > 1$ then finitely many iterations yield the optimal exponent 2, whereas if $r_c \leq 1$ then the exponent merely approaches $2r_c$. (The constants grow with each iteration, so we may only iterate finitely many times.) This explains the exponent on n in Theorem 2.1(2). Lemma 3.9 below summarizes the

induction step, and Sects. 3.7 and 3.8 implement it to complete the proof of Theorem 2.1.

Proof of Lemma 3.6 We simplify the estimate (3.3), starting with the Fisher information term. The LSI (2.3) is well known to tensorize [3, Proposition 5.2.7]:

$$H(\nu \,|\, \mu_t^{\otimes k}) \le \eta I(\nu \,|\, \mu_t^{\otimes k}), \qquad \forall t \ge 0,\ 1 \le k \le n,\ \nu \in \mathcal{P}((\mathbb{R}^d)^k).$$

Use this in inequality (3.3) along with exchangeability to obtain

$$\begin{aligned} H_t^k - H_s^k &\le \frac{1}{\sigma^2} \int_s^t \int_{(\mathbb{R}^d)^k} \sum_{i=1}^k \left| \widehat{b}_i^k(u,x) - \langle \mu_u, b(u,x_i,\cdot)\rangle \right|^2 P_u^k(dx)du - \frac{\sigma^2}{4\eta} \int_s^t H_u^k\, du \\ &= \frac{k}{\sigma^2} \int_s^t \int_{(\mathbb{R}^d)^k} \left| \widehat{b}_1^k(u,x) - \langle \mu_u, b(u,x_1,\cdot)\rangle \right|^2 P_u^k(dx)du - \frac{\sigma^2}{4\eta} \int_s^t H_u^k\, du. \end{aligned} \tag{3.11}$$

Let $t > 0$. Using the definition of $\widehat{b}_1^k$ from Lemma 3.4, we have

$$\begin{aligned} &\int_{(\mathbb{R}^d)^k} \left| \widehat{b}_1^k(t,x) - \langle \mu_t, b(t,x_1,\cdot)\rangle \right|^2 P_t^k(dx) \\ &\quad = \int_{(\mathbb{R}^d)^k} \left| \frac{1}{n-1} \sum_{j=2}^k \big(b(t,x_1,x_j) - \langle \mu_t, b(t,x_1,\cdot)\rangle \big) \right. \\ &\qquad \left. + \frac{n-k}{n-1} \langle P_{t,x}^{k+1|k} - \mu_t, b(t,x_1,\cdot)\rangle \right|^2 P_t^k(dx). \end{aligned}$$

Note that the above is valid for any $1 \le k \le n$ with the convention that the second term inside the square equals 0 for $k = n$. Use the inequality $(x+y)^2 \le (1+\delta)x^2 + (1+\delta^{-1})y^2$, valid for any $\delta > 0$ and $x, y \in \mathbb{R}$, to bound this further by $(1+\delta^{-1})I + (1+\delta)II$, where we set

$$\begin{aligned} I &:= \int_{(\mathbb{R}^d)^k} \left| \frac{1}{n-1} \sum_{j=2}^k \big(b(t,x_1,x_j) - \langle \mu_t, b(t,x_1,\cdot)\rangle \big) \right|^2 P_t^k(dx), \\ II &:= \int_{(\mathbb{R}^d)^k} \left| \frac{n-k}{n-1} \langle P_{t,x}^{k+1|k} - \mu_t, b(t,x_1,\cdot)\rangle \right|^2 P_t^k(dx). \end{aligned}$$

Let us first investigate the second term, assuming $k < n$ since it vanishes otherwise. Discard the constant $(n-k)/(n-1) \le 1$, and use Assumption (A.2) to obtain

$$II \le \gamma \int_{(\mathbb{R}^d)^k} H\big(P_{t,x}^{k+1|k} \,\big|\, \mu_t\big)\, P_t^k(dx) = \gamma\big(H_t^{k+1} - H_t^k\big), \tag{3.12}$$

with the second identity being the so-called *chain rule* for relative entropy; indeed, this comes from the formula $P_t^{k+1}(x, x_{k+1}) = P_t^k(x) P_{t,x}^{k+1|k}(x_{k+1})$ and the simple

calculation

$$\begin{aligned}
&\int_{(\mathbb{R}^d)^k} H\big(P^{k+1|k}_{t,x} \,\big|\, \mu_t\big)\, P^k_t(x)dx \\
&= \int_{(\mathbb{R}^d)^k} \left[\int_{\mathbb{R}^d} \log \frac{P^{k+1|k}_{t,x}(x_{k+1})}{\mu_t(x_{k+1})}\, P^{k+1|k}_{t,x}(x_{k+1})\, dx_{k+1} \right] P^k_t(x)dx \\
&= \int_{(\mathbb{R}^d)^k} \int_{\mathbb{R}^d} \left[\log \frac{P^{k+1}(x, x_{k+1})}{\mu_t^{\otimes(k+1)}(x, x_{k+1})} - \log \frac{P^k(x)}{\mu_t^{\otimes k}(x)} \right] P^{k+1}_t(x, x_{k+1}) dx_{k+1} dx \\
&= H^{k+1}_t - H^k_t,
\end{aligned}$$

where we write (x, x_{k+1}) here for a typical element of $(\mathbb{R}^d)^{k+1} \cong (\mathbb{R}^d)^k \times \mathbb{R}^d$.

We next study the first term, I. For case (1) of the lemma, we use convexity of $|\cdot|^2$ to get

$$\begin{aligned}
I &\le \frac{k-1}{(n-1)^2} \int_{(\mathbb{R}^d)^k} \sum_{j=2}^{k} \big| \big(b(t, x_1, x_j) - \langle \mu_t, b(t, x_1, \cdot)\rangle\big) \big|^2 P^k_t(dx) \\
&\le \frac{(k-1)^2}{(n-1)^2} M \le \frac{k^2}{n^2} M, \qquad (3.13)
\end{aligned}$$

with the second step using exchangeability and the definition of M from (2.5). In case (2) of the lemma, when we have the assumption (3.8), we take a more refined approach. Start by expanding the square and using exchangeability to obtain

$$I = \frac{k-1}{(n-1)^2} I_{\text{diag}} + \frac{(k-1)(k-2)}{(n-1)^2} I_{\text{cross}},$$

where we define I_{diag} and I_{cross} by

$$\begin{aligned}
I_{\text{diag}} &:= \int_{(\mathbb{R}^d)^2} \big| b(t, x_1, x_2) - \langle \mu_t, b(t, x_1, \cdot)\rangle \big|^2 P^2_t(dx), \\
I_{\text{cross}} &:= \int_{(\mathbb{R}^d)^3} \big(b(t, x_1, x_2) - \langle \mu_t, b(t, x_1, \cdot)\rangle\big) \cdot \big(b(t, x_1, x_3) - \langle \mu_t, b(t, x_1, \cdot)\rangle\big) P^3_t(dx).
\end{aligned}$$

We have immediately $I_{\text{diag}} \le M$. Writing $P^3_t(dx, dx_3) = P^2_t(dx) P^{3|2}_{t,x}(dx_3)$, we have

$$\begin{aligned}
I_{\text{cross}} &= \int_{(\mathbb{R}^d)^2} \big(b(t, x_1, x_2) - \langle \mu_t, b(t, x_1, \cdot)\rangle\big) \cdot \Big\langle P^{3|2}_{t,x} - \mu_t, b(t, x_1, \cdot)\Big\rangle P^2_t(dx) \\
&\le \sqrt{M} \left(\int_{(\mathbb{R}^d)^2} \Big| \Big\langle P^{3|2}_{t,x} - \mu_t, b(t, x_1, \cdot)\Big\rangle \Big|^2 P^2_t(dx_1, dx_2) \right)^{1/2}
\end{aligned}$$

$$\leq \sqrt{\gamma M}\left(\int_{(\mathbb{R}^d)^2} H(P^{3|2}_{t,x} \,|\, \mu_t)\, P^2_t(dx)\right)^{1/2}$$
$$\leq \sqrt{\gamma M H^3_T}.$$

The first inequality comes from Cauchy-Schwarz, the second from the assumption (2.4), and the last from the chain rule for relative entropy. Using the assumption (3.8) on H^3_T, we thus obtain

$$I \leq \frac{k-1}{(n-1)^2} M + \sqrt{\gamma M C_1}\frac{(k-1)(k-2)}{(n-1)^2 n^{p/2}}.$$

Using $(k-1)/(n-1) \leq k/n$, $n-1 \geq n/2$, and $p \leq 2$, we have

$$I \leq 2M\frac{k}{n^2} + \sqrt{\gamma M C_1}\frac{k^2}{n^{2+\frac{p}{2}}} \leq C_2 \frac{k^{2-p/2}}{n^2}, \tag{3.14}$$

where we recall that $C_2 = 2M + \sqrt{\gamma M C_1}$. The rest of the proof is given for case (2), but recalling (3.13) we note that the same proof covers case (1) as long as we set $p = 0$ and $C_2 = M$.

Using these bounds on I and II, the strategy in this proof will be to bound H^k_t in terms of H^{k+1}_t, and we will need a bound on the final term H^n_t of the iteration. We thus first apply (3.11) with $k = n$ and bound the first term of the right-hand side using (3.14) (recalling that $II = 0$ for $k = n$). We obtain

$$H^n_t - H^n_s \leq \int_s^t n^{1-p/2}\sigma^{-2} C_2 du - \frac{\sigma^2}{4\eta}\int_s^t H^n_u du,$$

Using Gronwall's inequality (see Lemma A.1 for a precise form tailored to our case), and setting $Z = \sigma^2/4\eta$, we have

$$H^n_t \leq e^{-Zt} H^n_0 + \int_0^t n^{1-p/2} C_2 e^{Z(s-t)} ds \leq C_0 n^{p_1-p_2} e^{-Zt} + n^{1-p/2}\sigma^{-2} C_2/Z.$$

Let us now consider the case $1 \leq k < n$ and bound H^k in terms of H^{k+1}. Returning to (3.11) and using (3.14) and (3.12), we have, for $t > s \geq 0$,

$$H^k_t - H^k_s \leq \int_s^t \left[\frac{(1+\delta^{-1})k^{3-p/2}}{\sigma^2 n^2} C_2 + \frac{\gamma k(1+\delta)}{\sigma^2}\left(H^{k+1}_u - H^k_u\right) - \frac{\sigma^2}{4\eta} H^k_u\right] du.$$

The two H^k_u terms combine to become $-(Z + \tilde{\gamma} k) H^k_u$, where we recall $\tilde{\gamma} = \gamma(1+\delta)/\sigma^2$. Simplify further by noting that $\frac{1+\delta^{-1}}{\sigma^2} = \frac{\tilde{\gamma}}{\delta\gamma}$. Apply Gronwall's inequalty (again see Lemma A.1 for a precise form) to get

$$H^k_t \leq e^{-(Z+\tilde{\gamma}k)t} H^k_0 + \tilde{\gamma}\int_0^t e^{-(Z+k\tilde{\gamma})(t-s)}\left(\frac{k^{3-p/2}}{\delta\gamma n^2} C_2 + k H^{k+1}_s\right) ds,$$

Complete the proof via an elementary induction, iterating this inequality for $k+1, k+2, \ldots, n-1$, and then using the assumption on H_0^k. □

3.4 Estimating the iterated integrals

Here we prove two lemmas that allow us to estimate the iterated integrals (3.4) and (3.5) appearing in Lemma 3.6.

Lemma 3.7 *For any real $p, T \geq 0$ and integers $1 \leq k < n$,*

$$\sum_{\ell=k}^{n-1} \ell^p \tilde{B}_k^\ell(T) \leq 2k^p (1+p)^p e^{(\tilde{\gamma} p - Z)T}. \tag{3.15}$$

Proof First, we introduce a related iterated integral in which Z is removed from the exponent:

$$B_k^\ell(t_k) = \left(\prod_{j=k}^{\ell-1} \tilde{\gamma} j \right) \int_0^{t_k} \int_0^{t_{k+1}} \cdots \int_0^{t_{\ell-1}} e^{-\tilde{\gamma}\ell t_\ell - \sum_{j=k}^{\ell-1} \tilde{\gamma} j (t_j - t_{j+1})} dt_\ell \ldots dt_{k+1}, \tag{3.16}$$

for $t_k \geq 0$. Recognizing a telescoping sum in the exponent of (3.5), we find $\tilde{B}_k^\ell(T) = e^{-ZT} B_k^\ell(T)$. We thus focus on $B_k^\ell(T)$, which appeared also in [38, page 25]. Recognize first that $\tilde{\gamma}\ell B_k^\ell = h_k * h_{k+1} * \cdots * h_\ell$, where $h_j(t) = \tilde{\gamma} j e^{-\tilde{\gamma} j t} 1_{[0,\infty)}(t)$ is the exponential density with parameters $\tilde{\gamma} j$. Using the expression for the convolution found in [38, Proof of Lemma 5.2] (applied with $a = \tilde{\gamma} k$ and $b = \tilde{\gamma}$), we obtain

$$\tilde{\gamma}(k+\ell) B_k^{k+\ell}(t) = h_k * h_{k+1} * \cdots * h_{k+\ell}(t) = \tilde{\gamma} e^{-\tilde{\gamma} k t} \frac{(k+\ell)!}{\ell!(k-1)!} (1 - e^{-\tilde{\gamma} t})^\ell,$$

We can then rewrite

$$\sum_{\ell=k}^{\infty} \ell^p B_k^\ell(t) = \sum_{\ell=0}^{\infty} (k+\ell)^p B_k^{k+\ell}(t) = \frac{e^{-\tilde{\gamma} k t}}{(k-1)!} \sum_{\ell=0}^{\infty} (k+\ell)^{p-1} \frac{(k+\ell)!}{\ell!} (1 - e^{-\tilde{\gamma} t})^\ell. \tag{3.17}$$

To proceed, we will use the inequality

$$(k+\ell)^{p-1}(k+\ell)! \leq 2\Gamma(k+\ell+p), \tag{3.18}$$

where $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$ is the gamma function. To prove (3.18), we set $z = k+\ell$, so that $(k+\ell)^{p-1}(k+\ell)! = z^{p-1}\Gamma(z+1) = z^p \Gamma(z)$. In the case $p \geq 1$, a well known gamma ratio inequality [36, Theorem 1, (5)] states that $z^p \Gamma(z) \leq \Gamma(z+p)$, which

gives (3.18) without the factor of 2. In the case $0 \le p < 1$, Gautschi's inequality [23, inequality (7)] states that $(z+1)^{p-1} \le \Gamma(z+p)/\Gamma(z+1)$, which implies

$$(k+\ell)^{p-1}(k+\ell)! = z^{p-1}\Gamma(z+1) \le \left(\frac{z+1}{z}\right)^{1-p}\Gamma(z+p) \le 2^{1-p}\Gamma(z+p),$$

where the last inequality uses $z \ge 1$.

For $0 < y \le 1$, note by Fubini's theorem and the definition of the gamma function that

$$\sum_{\ell=0}^{\infty} \frac{\Gamma(k+\ell+p)}{\ell!}(1-y)^{\ell} = \int_0^{\infty} \sum_{\ell=0}^{\infty} \frac{1}{\ell!} u^{k+\ell+p-1} e^{-u}(1-y)^{\ell}\, du = \int_0^{\infty} u^{k+p-1} e^{-yu}\, du$$
$$= y^{-(k+p)}\Gamma(k+p).$$

Applying this with $y = e^{-\tilde{\gamma}t}$ and then (3.18) in (3.17) gives

$$\sum_{\ell=k}^{\infty} \ell^p B_k^{\ell}(t) \le \frac{2e^{-\tilde{\gamma}kt}}{(k-1)!} \sum_{\ell=0}^{\infty} \frac{\Gamma(k+\ell+p)}{\ell!}(1-e^{-\tilde{\gamma}t})^{\ell} = 2e^{\tilde{\gamma}rt}\frac{\Gamma(k+p)}{(k-1)!}.$$

Finally, another gamma ratio inequality [36, Theorem 2, (12)] yields $\Gamma(k+p) \le \Gamma(k)k(k+p)^{p-1} = k!(k+p)^{p-1}$, which implies

$$\frac{\Gamma(k+p)}{(k-1)!} \le k(k+p)^{p-1} \le (k+p)^p \le k^p(1+p)^p.$$

□

Lemma 3.8 *Denote* $\alpha = \frac{Z}{\tilde{\gamma}} = \frac{\sigma^4}{4\gamma\eta(1+\delta)}$. *Then, for integers* $\ell \ge k \ge 1$,

$$\sup_{t\ge 0} \tilde{A}_k^{\ell}(t) = \prod_{i=k}^{\ell} \frac{i}{i+\alpha} \le \left(\frac{k+\alpha}{\ell+1+\alpha}\right)^{\alpha} \le (1+\alpha)^{\alpha}\left(\frac{k}{\ell+1}\right)^{\alpha}, \quad 1 \le k \le \ell. \tag{3.19}$$

Moreover, for any $p > 0$ *with* $p \ne \alpha - 1$,

$$\left(1 + \frac{(1+\alpha)^{\alpha}}{|p-\alpha+1|}\right)^{-1} \frac{1}{n^2} \sum_{\ell=k}^{n-1} \ell^p \sup_{t\ge 0} \tilde{A}_k^{\ell}(t) \le \begin{cases} k^{\alpha}/n^{\alpha+1-p} & \text{if } p-\alpha > -1, \\ k^{p+1}/n^2 & \text{if } p-\alpha < -1. \end{cases} \tag{3.20}$$

Proof Let $\tilde{h}_j(t) = \tilde{\gamma} j e^{-(Z+\tilde{\gamma}j)t} 1_{[0,\infty)}(t)$ for each $j \in \mathbb{N}$. Note that $\tilde{A}_k^{\ell}$ and $\tilde{B}_k^{\ell}$ can be expressed as convolutions,

$$\tilde{A}_k^{\ell} = \tilde{h}_k * \cdots * \tilde{h}_{\ell} * 1_{[0,\infty)}, \quad \tilde{\gamma}\ell\tilde{B}_k^{\ell} = \tilde{h}_k * \cdots \tilde{h}_{\ell}.$$

It follows that $d\tilde{A}_k^\ell/dt = \tilde{\gamma}\ell\tilde{B}_k^\ell$. We observed in (3.16) and just after that $\tilde{\gamma}\ell\tilde{B}_k^\ell(t) = e^{-Zt}h_k * \cdots * h_\ell(t)$, where $h_j(t) = \tilde{\gamma} j e^{-\tilde{\gamma} jt}1_{[0,\infty)}(t)$ is an exponential probability density for each $j \in \mathbb{N}$. In other words,

$$\frac{d}{dt}\tilde{A}_k^\ell(t) = e^{-Zt}h_k * \cdots * h_\ell(t).$$

In particuar, $\tilde{A}_k^\ell$ is increasing. Consider independent exponential random variables $(E_i)_{i\in\mathbb{N}}$ with parameters $\tilde{\gamma}i$. The density of the sum $E_k + \ldots + E_\ell$ is precisely $h_k * \cdots * h_\ell$. By integrating from 0 and t, we compute

$$\sup_{t\geq 0}\tilde{A}_k^\ell(t) = \int_0^\infty e^{-Zt}h_k * \cdots * h_\ell(t)\,dt = \mathbb{E}\left[e^{-Z(E_k+\ldots+E_\ell)}\right] = \prod_{i=k}^{\ell}\frac{\tilde{\gamma}i}{\tilde{\gamma}i + Z},$$

which proves the first identity in (3.19). Use $\log x \leq x - 1$ for $x > 0$ to estimate

$$\sum_{i=k}^{\ell}\log\frac{i}{i+\alpha} \leq -\sum_{i=k}^{\ell}\frac{\alpha}{i+\alpha} \leq -\int_k^{\ell+1}\frac{\alpha}{x+\alpha}\,dx = \alpha\log\frac{k+\alpha}{\ell+1+\alpha},$$

and exponentiate to get the first inequality in (3.19). The last inequality of (3.19) follows from $\alpha > 0$ and $k + \alpha \leq k(1+\alpha)$. To prove (3.20), consider two cases:

(1) If $p - \alpha > -1$, we can bound $\sum_{\ell=k}^{n-1}\ell^{p-\alpha} \leq \int_0^n x^{p-\alpha}dx = n^{p-\alpha+1}/(p-\alpha+1)$. Thus,

$$\frac{1}{n^2}\sum_{\ell=k}^{n-1}\ell^p\sup_{t\geq 0}\tilde{A}_k^\ell(t) \leq \frac{(1+\alpha)^\alpha k^\alpha}{n^2}\sum_{\ell=k}^{n-1}\ell^{p-\alpha} \leq \frac{(1+\alpha)^\alpha}{p-\alpha+1}\frac{k^\alpha}{n^{1+\alpha-p}}.$$

(2) For $p - \alpha < -1$, we can bound $\sum_{\ell=k+1}^{n-1}\ell^{p-\alpha} \leq \int_k^\infty x^{p-\alpha}dx = k^{p-\alpha+1}/(\alpha - p - 1)$. Then,

$$\begin{aligned}\frac{1}{n^2}\sum_{\ell=k}^{n-1}\ell^p\sup_{t\geq 0}\tilde{A}_k^\ell(t) &= \frac{k^p}{n^2}\sup_{t\geq 0}\tilde{A}_k^k(t) + \frac{1}{n^2}\sum_{\ell=k+1}^{n-1}\ell^p\sup_{t\geq 0}\tilde{A}_k^\ell(t)\\ &\leq \frac{k^p}{n^2}\frac{k}{k+\alpha} + \frac{(1+\alpha)^\alpha k^\alpha}{n^2}\sum_{\ell=k+1}^{n-1}\ell^{p-\alpha}\\ &\leq \frac{k^{p+1}}{n^2} + \frac{(1+\alpha)^\alpha}{\alpha-p-1}\frac{k^{p+1}}{n^2}.\end{aligned}$$

□

3.5 Proof of Theorem 2.1, for $r_c > 2$

We first apply part (1) of Lemma 3.6. Combining (3.6) and (3.7) and the assumption $H_0^k \le C_0(k/n)^2$ yields

$$H_T^k \le \frac{C_0}{n^2}\sum_{\ell=k}^{n-1}\ell^2\tilde{B}_k^\ell(T) + \frac{M}{\delta\gamma n^2}\sum_{\ell=k}^{n-1}\ell^2\tilde{A}_k^\ell(T) + n\tilde{A}_k^{n-1}(T)\left(C_0e^{-ZT} + 4\eta M\sigma^{-4}\right).$$

Applying Lemma 3.8 for the last term, as well as Lemma 3.7 (with $p = 2$) for the first term,

$$H_T^k \le 18C_0\frac{k^2}{n^2}e^{(2\tilde{\gamma}-Z)T} + \frac{M}{\delta\gamma n^2}\sum_{\ell=k}^{n-1}\ell^2\tilde{A}_k^\ell(T) + (1+\alpha)^\alpha\frac{k^\alpha}{n^{\alpha-1}}\left(C_0e^{-ZT} + 4\eta M\sigma^{-4}\right). \tag{3.21}$$

In the following, we write C to denote a constant which can depend only on $(C_0, \gamma, \sigma, M, \eta)$, and may change from line to line but never depends on (n, k, T). Recall that $\alpha = \frac{Z}{\tilde{\gamma}} = \frac{\sigma^4}{4\gamma\eta(1+\delta)} = \frac{1+r_c}{1+\delta}$. Since $r_c > 2$, we may choose $\delta > 0$ such that $\alpha > 3$. Clearly $e^{(2\tilde{\gamma}-Z)T} = e^{\tilde{\gamma}(2-\alpha)T} \le 1$. Apply Lemma 3.8 (with $p = 2$) in the second term of (3.21), noting that $2 - \alpha < -1$, to get

$$H_T^k \le C\left(\frac{k^2}{n^2} + \frac{k^3}{n^2} + \frac{k^\alpha}{n^{\alpha-1}}\right).$$

Noting that $\alpha > 3$, this yields the suboptimal $H_T^k \le Ck^3/n^2$. This is where part (2) of Lemma 3.6 enters the picture, allows us to sharpen the k exponent.

Specifically, since we now know $H_T^3 \le C/n^2$, we can apply part (2) of Lemma 3.6, with $p = 2$, to get $H_T^n \le C$ and

$$H_T^k \le \frac{C_0}{n^2}\sum_{\ell=k}^{n-1}\ell^2\tilde{B}_k^\ell(T) + \frac{C}{n^2}\sum_{\ell=k}^{n-1}\ell\tilde{A}_k^\ell(T) + \tilde{A}_k^{n-1}(T)H_T^n.$$

The first term is bounded by $18C_0(k/n)^2$, by Lemma 3.7. For the third term, we use Lemma 3.8 to get $\tilde{A}_k^{n-1}(T)H_T^n \le C(k/n)^\alpha \le C(k/n)^2$ since $\alpha > 2$. For the middle term, we use Lemma 3.8 (with $p = 1$), noting that $1 - \alpha < -1$:

$$\frac{1}{n^2}\sum_{\ell=k}^{n-1}\ell\tilde{A}_k^\ell(T) \le \left(1 + \frac{(1+\alpha)^\alpha}{\alpha-3}\right)(k/n)^2.$$

This completes the proof of Theorem 2.1 in the case $r_c > 2$. □

3.6 An inductive lemma

In order to prove Theorem 2.1 in the remaining case $0 < r_c \le 2$, we first state a lemma which inductively applies part (2) of Lemma 3.6, as was described in the paragraph following the statement of the lemma. In the following, define

$$r_c := \frac{\sigma^4}{4\gamma\eta} - 1, \quad \text{and} \quad q_m := 2(1 - 2^{-m}), \ \text{for } m \in \mathbb{N},$$

and note that $q_{m+1} = 1 + q_m/2$.

Lemma 3.9 *Let $0 < r < 2$ satisfy $r < r_c$. Suppose $m \in \mathbb{N}$ is such that $rq_m < 2$. Suppose that there exists a constant C_0 such that $H_0^k \le C_0 k^{1+r}/n^{\min(2,2r)}$, for all $k = 1, \dots, n$. Suppose that $\sup_{T\ge 0} H_T^3 \le C_m n^{-rq_m}$ for some constant $C_m > 0$. Then, there exists a constant $C_m' > 0$ depending only on $(C_m, \sigma, \eta, \gamma, C_0, M)$ such that, by defining $C_{m+1} = \frac{C_m'}{(r_c - r)|2 - rq_{m+1}|}$, we have:*

(1) If $rq_{m+1} < 2$, then $\sup_{T\ge 0} H_T^3 \le C_{m+1} n^{-rq_{m+1}}$,
(2) If $rq_{m+1} > 2$, then $\sup_{T\ge 0} H_T^3 \le C_{m+1} n^{-2}$.

Proof Recall that $\alpha = \frac{Z}{\tilde\gamma} = \frac{\sigma^4}{4\gamma\eta(1+\delta)} = (1 + r_c)/(1 + \delta)$ where the parameter $\delta > 0$ can be chosen freely. Let $r^* = \min(2, 2r)$. Choose $\delta = (r_c - r)/(1 + r)$, so that $\alpha = 1 + r$. We apply Lemma 3.6 and deduce the existence of a constant $\tilde C_m > 0$, depending only on $(C_m, C_0, \gamma, \eta, \sigma, M)$ such that

$$H_T^n \le \tilde C_m (n^{1+r-r^*} + n^{1-\frac{rq_m}{2}}), \tag{3.22}$$

$$H_T^k \le \frac{C_0}{n^{r^*}} \sum_{\ell=k}^{n-1} \ell^{1+r} \tilde B_k^\ell(T) + \frac{\tilde C_m}{\delta n^2} \sum_{\ell=k}^{n-1} \ell^{2-\frac{rq_m}{2}} \tilde A_k^\ell(T) + \tilde A_k^{n-1}(T) H_T^n, \ 1 \le k < n, \tag{3.23}$$

for all $T \ge 0$. For the first term of the right hand side of (3.23), Lemma 3.7 yields

$$\sup_{T\ge 0} \frac{C_0}{n^{r^*}} \sum_{\ell=k}^{n-1} \ell^{1+r} \tilde B_k^\ell(T) \le \sup_{T\ge 0} 2C_0 \frac{k^{1+r}}{n^{r^*}} (2 + r)^{1+r} e^{(\tilde\gamma(1+r) - Z)T} \le 128 C_0 \frac{k^{1+r}}{n^{r^*}},$$

where the last step used $\tilde\gamma(1 + r) - Z = 0$, and also $r \le 2$ so that $2(2 + r)^{1+r} \le 128$. Since $r \le 2$, we have $1/\delta \le 3/(r_c - r)$, and so

$$H_T^k \le 128 C_0 \frac{k^{1+r}}{n^{r^*}} + \frac{3\tilde C_m}{(r_c - r)n^2} \sum_{\ell=k}^{n-1} \ell^{2-\frac{rq_m}{2}} \tilde A_k^\ell(T) + \tilde A_k^{n-1}(T) H_T^n. \tag{3.24}$$

We next wish to combine the exponents.

Suppose first that $rq_{m+1} < 2$. Then $2 - \frac{rq_m}{2} - \alpha > -1$, and Lemma 3.8 with $\alpha = 1 + r$ and $q = 2 - \frac{rq_m}{2}$ yields

$$H_T^k \le 128C_0 \frac{k^{1+r}}{n^{r^*}} + \frac{3\tilde{C}_m}{r_c - r}\left(1 + \frac{(2+r)^{1+r}}{2 - rq_m/2 - r}\right)\frac{k^{1+r}}{n^{2+r-(2-rq_m/2)}} + (2+r)^{1+r}\frac{k^{1+r}}{n^{1+r}}H_T^n$$
$$\le 128C_0 \frac{k^{1+r}}{n^{r^*}} + \frac{3\tilde{C}_m}{r_c - r}\left(1 + \frac{64}{2 - rq_{m+1}}\right)\frac{k^{1+r}}{n^{rq_{m+1}}} + 64\frac{k^{1+r}}{n^{1+r}}H_T^n,$$

where we used $q_{m+1} = 1 + q_m/2$ and $(2+r)^{1+r} \le 64$, the latter following from $r \le 2$. To simplify this further, we claim first that

$$1 + r - r^* < 1 - \frac{rq_m}{2}. \tag{3.25}$$

Indeed, if $r > 1$, then $r^* = 2$, and (3.25) follows from the assumed inequality $r + rq_m/2 = rq_{m+1} < 2$. If instead $r \le 1$, then $r^* = 2r$, and (3.25) follows simply from $q_m < 2$. Now, using (3.25), we see that (3.22) implies $H_T^n \le 2\tilde{C}_m n^{1-\frac{rq_m}{2}}$. Using again $q_{m+1} = 1 + q_m/2$, we deduce that $H_T^n/n^{1+r} \le 2\tilde{C}_m/n^{rq_{m+1}}$. Also, rearranging (3.25) shows that $r^* \ge rq_{m+1}$, so $n^{-r^*} \le n^{-rq_{m+1}}$. We can now simplify the first and last terms in (3.24) to get

$$H_T^k \le 128C_0 \frac{k^{1+r}}{n^{rq_{m+1}}} + \frac{3\tilde{C}_m}{r_c - r}\left(1 + \frac{64}{2 - rq_{m+1}}\right)\frac{k^{1+r}}{n^{rq_{m+1}}} + 128\tilde{C}_m\frac{k^{1+r}}{n^{rq_{m+1}}}.$$

Evaluating this for $k = 3$ and using $1 + r \le 3$ gives

$$H_T^3 n^{rq_{m+1}} \le 3456C_0 + \frac{81\tilde{C}_m}{r_c - r}\left(1 + \frac{64}{2 - rq_{m+1}}\right) + 3456\tilde{C}_m,$$

Using $r_c - r \le r_c$ and $2 - rq_{m+1} \le 2$, we may put with each term the same denominator $(r_c - r)(2 - rq_{m+1})$. This completes the proof in the case $rq_{m+1} < 2$.

If instead $rq_{m+1} > 2$, then $r > 1$ because $q_{m+1} < 2$, and thus $r^* = 2$. Moreover, $1 - \frac{rq_m}{2} < r - 1$, and thus $H_T^n \le 2\tilde{C}_m n^{r-1}$ by (3.22). Apply Lemma 3.8 to obtain this time

$$H_T^k \le 128C_0 \frac{k^{1+r}}{n^2} + \frac{3\tilde{C}_m}{r_c - r}\left(1 + \frac{(2+r)^{1+r}}{rq_{m+1} - 2}\right)\frac{k^{3-\frac{rq_m}{2}}}{n^2} + (2+r)^{1+r}\frac{k^{1+r}}{n^{1+r}}2\tilde{C}_m n^{r-1}.$$

Apply this with $k = 3$ to complete the proof. □

3.7 Proof of Theorem 2.1, for $1 < r_c \le 2$

We begin exactly as in Sect. 3.5, applying part (1) of Lemma 3.6 to deduce the inequality (3.21). We will again use C to denote a constant which depends only on $(C_0, \gamma, \sigma, M, \eta)$, and may change from line to line, but never depends on (n, k, T).

Recall that $\alpha = \frac{Z}{\tilde{\gamma}} = \frac{\sigma^4}{4\gamma\eta(1+\delta)} = \frac{1+r_c}{1+\delta}$, where $\delta > 0$ can be chosen freely. We choose here δ so that $2 < \alpha < 3$ and such that $r := \alpha - 1 \notin \{2/q_m : m \in \mathbb{N}\}$. Note then that $1 < r < r_c \leq 2$.

We simplify (3.21) in this case as follows. For the first term, note that $e^{(2\tilde{\gamma}-Z)T} = e^{\tilde{\gamma}(2-\alpha)T} \leq 1$. For the second and third terms, apply Lemma 3.8 (with $p = 2$), noting that $2 - \alpha > -1$:

$$H_T^k \leq C\left(\frac{k^2}{n^2} + \frac{k^{r+1}}{n^r} + \frac{k^{r+1}}{n^r}\right).$$

Setting $k = 3$ yields $H_T^3 \leq C/n^r = C/n^{rq_1}$. The idea is to now repeatedly apply Lemma 3.9, $m^* - 1$ times, where m^* is the smallest integer such that $rq_{m^*} > 2$, which exists because $\lim_{m\to\infty} rq_m = 2r > 2$. Note $rq_m \neq 2$ for all m, by design. This repeated application of Lemma 3.9 ultimately yields $H_T^3 \leq C/n^2$. We may now conclude exactly as in the last paragraph of Sect. 3.5, noting that $\alpha > 2$. □

3.8 Proof of Theorem 2.1, for $0 < r_c \leq 1$

Let $0 < \epsilon_1 < \epsilon_2 < r_c$. We first claim that we can find a constant C' depending only on $(\gamma, \sigma, M, \eta, \epsilon_1, \epsilon_2, C_0^{(\epsilon_2-\epsilon_1)/2, \epsilon_2-\epsilon_1})$, such that

$$\sup_{T\geq 0} H_T^3 \leq C'/n^{2(r_c-(\epsilon_2-\epsilon_1))}. \tag{3.26}$$

Abbreviate $\epsilon := \epsilon_2 - \epsilon_1$ and $r := r_c - \epsilon/2$, and choose $\delta > 0$ such that $\alpha = 1 + r \in (1, 2)$. Let $p_1 = 1 + r_c - \epsilon/2$ and $p_2 = 2r_c - \epsilon$. We have by assumption $H(P_0^k \mid \mu_0^{\otimes k}) \leq C_0^{\epsilon/2,\epsilon} k^{p_1}/n^{p_2}$. Apply Lemma 3.6(1), plugging (3.6) into (3.7) and using $p_1 - p_2 \leq 1$, we find

$$H_T^k \leq \frac{C_0^{\epsilon/2,\epsilon}}{n^{p_2}} \sum_{\ell=k}^{n-1} \ell^{p_1} \tilde{B}_k^\ell(T) + \frac{M}{\delta\gamma n^2} \sum_{\ell=k}^{n-1} \ell^2 \tilde{A}_k^\ell(T) + n\tilde{A}_k^{n-1}(T)\left(C_0 e^{-ZT} + 4\eta M\sigma^{-4}\right).$$

Apply Lemma 3.8 for the last term, as well as Lemma 3.7 for the first term:

$$H_T^k \leq 18C_0^{\epsilon/2,\epsilon} \frac{k^{p_1}}{n^{p_2}} e^{(\tilde{\gamma}p_1 - Z)T} + \frac{M}{\delta\gamma n^2} \sum_{\ell=k}^{n-1} \ell^2 \tilde{A}_k^\ell(T) + (1+\alpha)^\alpha \frac{k^\alpha}{n^{\alpha-1}}\left(C_0 e^{-ZT} + 4\eta M\sigma^{-4}\right), \tag{3.27}$$

where we used $p_1 \leq 2$ to get $2(1 + p_1)^{p_1} \leq 18$. Simplify the first term by noting that $p_1 = \alpha$ and so $e^{(\tilde{\gamma}p_1 - Z)T} = 1$. Apply Lemma 3.8 (with $p = 2$) to the second term in (3.27), noting that $2 - \alpha > -1$, to get

$$H_T^k \leq C'\left(\frac{k^{1+r}}{n^{2r}} + \frac{k^{1+r}}{n^r} + \frac{k^{1+r}}{n^r}\right),$$

with C' depending only on $(\gamma, \sigma, M, \eta, \epsilon, C_0^{\epsilon/2,\epsilon})$.

By evaluating the above for $k = 3$, we have $H_T^3 \le C'/n^r$. We may now repeatedly apply Lemma 3.9, as in Sect. 3.7, but this time we never reach the case where $rq_m \ge 2$ because $rq_m \uparrow 2r < 2r_c \le 2$ as $m \to \infty$. Instead, we choose m large enough so that $rq_m \ge 2r - \epsilon$ and deduce that $H_T^3 \le C/n^{2r-\epsilon} = C/n^{2r_c-2\epsilon}$, where $C > 0$ now depends only on $(\gamma, \sigma, M, \eta, \epsilon_1, \epsilon_2, C_0^{\epsilon/2,\epsilon})$. For the rest of this proof, $C > 0$ may additionally depend on $C_0^{\epsilon_1,\epsilon_2}$ and may change from line to line, but it will never depend on (n, k, T).

With (3.26) now established, let us now instead choose $\delta > 0$ (depending only on $\sigma, \gamma, \eta, \epsilon_1$) so that $\alpha = 1 + r_c - \epsilon_1$, and define $p_1 = 1 + r_c - \epsilon_1$ and $p_2 = 2r_c - \epsilon_2$. Then (3.27) again holds but with $C_0^{\epsilon/2,\epsilon}$ replaced by $C_0^{\epsilon_1,\epsilon_2}$. In light of (3.26), we may apply Lemma 3.6(2) with $p = 2(r_c - (\epsilon_2 - \epsilon_1))$ to get

$$H_T^n \le Cn^{1-r_c+\epsilon_2-\epsilon_1},$$
$$H_T^k \le C\frac{k^{1+r_c-\epsilon_1}}{n^{2r_c-\epsilon_2}} + \frac{C}{n^2}\sum_{\ell=k}^{n-1} \ell^{2-r_c+\epsilon_2-\epsilon_1}\tilde{A}_k^\ell(T) + \tilde{A}_k^{n-1}(T)H_T^n.$$

Apply Lemma 3.8 with $p = 2-r_c+\epsilon_2-\epsilon_1$, noting that then $p-\alpha = 1-2r_c+\epsilon_2 > -1$, to get

$$\frac{1}{n^2}\sum_{\ell=k}^{n-1} \ell^{2-r_c+\epsilon_2-\epsilon_1}\tilde{A}_k^\ell(T) \le C\frac{k^\alpha}{n^{\alpha+1-p}} = C\frac{k^{1+r_c-\epsilon_1}}{n^{2r_c-\epsilon_2}}.$$

Lemma 3.8 also yields

$$\tilde{A}_k^{n-1}(T)H_T^n \le C\frac{k^\alpha}{n^\alpha}n^{1-r_c+\epsilon_2-\epsilon_1} = C\frac{k^{1+r_c-\epsilon_1}}{n^{2r_c-\epsilon_2}}.$$

Put it together to complete the proof of this final case, and thus the theorem. □

4 Proof of for reversed entropy

In this section we prove Theorem 2.5. Define the reversed relative entropy

$$H_t^k := H(\mu_t^{\otimes k} \mid P_t^k), \quad t \ge 0,\ 1 \le k \le n.$$

By assumption, $H_0^k \le C_0(k/n)^p$ for some C_0 and $0 < p \le 2$, with the choice of p depending on which case in Theorem 2.5 we are considering. We make use again of the functions $\widehat{b}_i^k$ defined in (3.1). We first apply Lemma 3.1 to compute the time-derivative of H_t^k. It is evident that (H.1,2) hold because b is bounded and b_0 is locally bounded. Moreover, (H.3) is a direct consequence of (2.9). We thus obtain, for all $t > s \ge 0$,

$$H_t^k - H_s^k + \frac{\sigma^2}{4}\int_s^t I(\mu_u^{\otimes k}|P_u^k)du \le \frac{1}{\sigma^2}\int_s^t \int_{(\mathbb{R}^d)^k} \sum_{i=1}^k \big|\widehat{b}_i^k(u,x) - \langle \mu_u, b(u,x_i,\cdot)\rangle\big|^2 \mu_u^{\otimes k}(dx)du.$$

The LSI (2.10) marginalizes, in the sense that

$$H(\nu \mid P_t^k) \le \eta I(\nu \mid P_t^k), \quad \forall k = 1,\dots,n,\ \nu \in \mathcal{P}((\mathbb{R}^d)^k),\ t \ge 0.$$

Indeed, for a given k and $\nu \in \mathcal{P}((\mathbb{R}^d)^k)$, this follows by applying (2.10) to the measure $\tilde{\nu} \in \mathcal{P}((\mathbb{R}^d)^n)$ with density $d\tilde{\nu}/dP_t^n(x_1,\dots,x_n) = d\nu/dP_t^k(x_1,\dots,x_k)$. This yields

$$\begin{aligned} H_t^k - H_s^k + \frac{\sigma^2}{4\eta}\int_s^t H_u^k du &\le \frac{1}{\sigma^2}\int_s^t \int_{(\mathbb{R}^d)^k} \sum_{i=1}^k \big|\widehat{b}_i^k(u,x) - \langle \mu_u, b(u,x_i,\cdot)\rangle\big|^2 \mu_u^{\otimes k}(dx)du \\ &= \frac{k}{\sigma^2}\int_s^t \int_{(\mathbb{R}^d)^k} \big|\widehat{b}_1^k(u,x) - \langle \mu_u, b(u,x_1,\cdot)\rangle\big|^2 \mu_u^{\otimes k}(dx)du, \end{aligned} \tag{4.1}$$

where the second line comes from exchangeability. Use the definition of $\widehat{b}_1^k$ to write the inside of the time integral of right-hand side as

$$\begin{aligned} &\frac{k}{\sigma^2}\int_{(\mathbb{R}^d)^k} \bigg| \frac{1}{n-1}\sum_{j=2}^k \big(b(t,x_1,x_j) - \langle \mu_t, b(t,x_1,\cdot)\rangle\big) \\ &+ \frac{n-k}{n-1}\big\langle P_{t,x}^{k+1|k} - \mu_t, b(t,x_1,\cdot)\big\rangle\bigg|^2 \mu_t^{\otimes k}(dx). \end{aligned}$$

Let $\delta > 0$, to be chosen later, and use the inequality $(x+y)^2 \le (1+\delta)x^2 + (1+\delta^{-1})y^2$ to further bound this by

$$\frac{k}{\sigma^2}(1+\delta^{-1})I + \frac{k}{\sigma^2}(1+\delta)II,$$

where we set

$$\begin{aligned} I &:= \int_{(\mathbb{R}^d)^k} \bigg| \frac{1}{n-1}\sum_{j=2}^k \big(b(t,x_1,x_j) - \langle \mu_t, b(t,x_1,\cdot)\rangle\big)\bigg|^2 \mu_t^{\otimes k}(dx), \\ II &:= \int_{(\mathbb{R}^d)^k} \bigg| \frac{n-k}{n-1}\big\langle P_{t,x}^{k+1|k} - \mu_t, b(t,x_1,\cdot)\big\rangle\bigg|^2 \mu_t^{\otimes k}(dx), \end{aligned}$$

again with the convention that $II = 0$ for $k = n$. Expanding the square in I, the cross-terms vanish, and by exchangeability we have

$$I = \frac{k-1}{(n-1)^2} \int_{(\mathbb{R}^d)^2} \left| b(t, x_1, x_2) - \langle \mu_t, b(t, x_1, \cdot) \rangle \right|^2 \mu_t^{\otimes 2}(dx) \leq \frac{\gamma(k-1)}{2(n-1)^2}, \quad (4.2)$$

where we set $\gamma := 2\||b|^2\|_\infty$. We note here for future use that when $k = n$, we have $I \leq \gamma/2(n-1)$ and $II = 0$, so that (4.1) becomes

$$H_t^n - H_s^n + \frac{\sigma^2}{4\eta} \int_s^t H_u^n du \leq \frac{\gamma n}{2\sigma^2(n-1)} \leq \frac{\gamma}{\sigma^2}.$$

Using a form of Gronwall's inequality (see Lemma A.1), and setting $Z = \sigma^2/4\eta$, we have

$$H_t^n \leq e^{-Zt} H_0^n + \int_0^t \frac{\gamma}{\sigma^2} e^{Z(s-t)} ds \leq C_0 e^{-Zt} + \frac{4\gamma\eta}{\sigma^4}. \quad (4.3)$$

Next, for $k < n$, we bound II. Since b is bounded, Pinsker's inequality yields

$$\left| \left\langle P_{t,x}^{k+1|k} - \mu_t, b(t, x_1, \cdot) \right\rangle \right|^2 \leq \gamma H(\mu_t \mid P_{t,x}^{k+1|k}).$$

Using $(n-k)/(n-1) \leq 1$ and the chain rule for relative entropy (as in (3.12)), we get

$$II \leq \gamma \int_{(\mathbb{R}^d)^k} H(\mu_t \mid P_{t,x}^{k+1|k}) \, \mu_t^{\otimes k}(dx) = \gamma \left(H_t^{k+1} - H_t^k \right).$$

Adding this to (4.2) and returning to (4.1) yields

$$H_t^k - H_s^k \leq \int_s^t \left(\frac{k(k-1)\gamma(1+\delta^{-1})}{2(n-1)^2\sigma^2} + \tilde{\gamma} k H_u^{k+1} - (Z + \tilde{\gamma} k) H_u^k \right) du.$$

where we set $\tilde{\gamma} = \gamma(1+\delta)/\sigma^2$. The first term inside the integral is bounded by $\tilde{\gamma} k^2/\delta n^2$, which follows from using $(k-1)/(n-1) \leq k/n$ and $n \leq 2(n-1)$ as well as the identity $\frac{1+\delta^{-1}}{\sigma^2}\gamma = \frac{\tilde{\gamma}}{\delta}$. By Gronwall's inequality (as in Lemma A.1),

$$H_t^k \leq e^{-(Z+\tilde{\gamma}k)t} H_0^k + \int_0^t e^{-(Z+\tilde{\gamma}k)s} \left(\frac{\tilde{\gamma} k^2}{\delta n^2} + \tilde{\gamma} k H_s^{k+1} \right) ds.$$

By iterating for $\ell = k, k+1, \ldots, n$ as in the proof of Lemma 3.6, we have

$$H_T^k \leq \sum_{\ell=k}^{n-1} H_0^\ell \tilde{B}_k^\ell(T) + \frac{1}{\delta n^2} \sum_{\ell=k}^{n-1} \ell \tilde{A}_k^\ell(T) + \tilde{A}_k^{n-1}(T) H_T^n,$$

where $\tilde{A}_k^\ell$ and $\tilde{B}_k^\ell$ (and Z and $\tilde{\gamma}$) are defined as in Lemma 3.6. Plugging in (4.3) and the assumption $H_0^k \le C_0(k/n)^p$, we obtain

$$H_T^k \le \frac{C_0}{n^p}\sum_{\ell=k}^{n-1}\ell^p \tilde{B}_k^\ell(T) + \frac{1}{\delta n^2}\sum_{\ell=k}^{n-1}\ell\tilde{A}_k^\ell(T) + \tilde{A}_k^{n-1}(T)\left(C_0 e^{-Zt} + 4\gamma\eta\sigma^{-4}\right). \tag{4.4}$$

Let $\alpha = \frac{Z}{\tilde{\gamma}} = \frac{\sigma^4}{4\gamma\eta(1+\delta)}$ and apply Lemmas 3.7 and 3.8 in (4.4) to get

$$H_T^k \le \frac{2C_0k^p}{n^p}(1+p)^p e^{Z(\frac{p}{\alpha}-1)T} + \frac{1}{\delta n^2}\sum_{\ell=k}^{n-1}\ell\tilde{A}_k^\ell(T) + (1+\alpha)^\alpha\frac{k^\alpha}{n^\alpha}\left(C_0 + 4\gamma\eta\sigma^{-4}\right). \tag{4.5}$$

We now complete the proof in two cases separately.

First case: $p_c = \frac{\sigma^4}{4\gamma\eta} > 2$

In this case $p = 2$. Choose $\delta > 0$ so that $\alpha > 2$. Using Lemma 3.8,

$$\frac{1}{n^2}\sum_{\ell=k}^{n-1}\ell\tilde{A}_k^\ell(T) \le \left(1 + \frac{(1+\alpha)^\alpha}{\alpha-2}\right)\frac{k^2}{n^2}.$$

Using this in (4.5) and noting that $(k/n)^\alpha \le (k/n)^2$ completes the proof.

Second case: $p_c = \frac{\sigma^4}{4\gamma\eta} \le 2$

In this case $p = \frac{\sigma^4}{4\gamma\eta} - \epsilon$, where $\epsilon \in (0, p_c)$. Choose $\delta > 0$ so that $\alpha = p$. Since $\alpha < 2$, applying Lemma 3.8 yields

$$\frac{1}{n^2}\sum_{\ell=k}^{n-1}\ell\tilde{A}_k^\ell(T) \le \left(1 + \frac{(1+\alpha)^\alpha}{2-\alpha}\right)\frac{k^\alpha}{n^\alpha}.$$

Using this in (4.5) completes the proof. □

5 Proof of corollaries

5.1 Convex potentials: proof of Corollary 2.7

Here $b_0(t,x) = -\nabla U(x)$ and $b(t,x,y) = -\nabla W(x-y)$. We need to check that the conditions of Assumptions E and A hold. Note that the LSI for μ_0 is known to imply that μ_0 has finite moments of every order.

We start with Assumption E. Assumption (E.1) is immediate by continuity of $(\nabla U, \nabla W)$. Suppose that $L < \infty$. The existence and uniqueness of a weak solution $(\mu_t)_{t\geq 0}$ of the nonlinear Fokker–Planck equation follows from [11, Theorem 2.6], and since ∇W Lipschitz we may use the moment bounds of [11, Theorem 2.6] to deduce $\nabla W(x - \cdot) \in L^1(\mu_t)$ for all (t, x) as well as the continuity and thus local boundedness of $(t, x) \mapsto \langle \mu_t, \nabla W(x - \cdot)\rangle$. This shows Assumption (E.2). The existence of a unique (starting from the given initial law P_0^n) solution $(P_t^n)_{t\geq 0}$ to the Fokker–Planck equation required by Assumption (E.3) is justified in [11, Section 2]. Because ∇U and ∇W have polynomial growth in x, the integrability requirements of Assumption (E.3) are a simple consequences of the moments bounds for P_t^n and μ_t given in [11, Corollary 2.3, Theorem 2.6], which adapt easily to the case of non-i.i.d. initial conditions. If $L = \infty$ but $R < \infty$, we verify Assumption E in a similar fashion.

We now turn to checking Assumption A. Using the assumption $\nabla^2(U + W) \succeq \alpha I$, the (time-inhomogeneous) generator $\varphi \mapsto -\nabla\varphi \cdot \nabla U - \nabla\varphi \cdot \nabla W * \mu_t + \frac{\sigma^2}{2}\Delta\varphi$ of the Fokker–Planck equation satisfied by $(\mu_t)_{t\geq 0}$ obeys the curvature condition described in [40, Proposition 3.12] and [15, Theorem 4.1], and we may thus deduce from their arguments (see also [40, Corollary 3.7]) that μ_t satisfies a LSI with constant $\frac{\sigma^2}{4\alpha}(1 - e^{-4\alpha t/\sigma^2}) + \frac{\eta_0}{4}e^{-4\alpha t/\sigma^2}$. This shows that Assumption (A.1) holds with $\eta = \max(\eta_0/4, \sigma^2/4\alpha)$. For Assumption (A.2), first notice that if $R < \infty$, Pinsker's inequality gives us the required inequality with $\gamma = 2R^2$. In the case where $L < \infty$, recall from Remark 2.3 that μ_t obeys the quadratic transport inequality (2.8). Hence, since ∇W is L-Lipschitz, we use Kantorovich duality to deduce

$$\left|\langle \nu - \mu_t, \nabla W(x - \cdot)\rangle\right|^2 \leq L^2\mathcal{W}_1^2(\mu_t, \nu) \leq L^2\mathcal{W}_2^2(\mu_t, \nu) \leq 4\eta L^2 H(\nu \,|\, \mu_t). \quad (5.1)$$

This yields Assumption (A.2) with $\gamma = 4\eta L^2$. We finally check that the constant M of Assumption (A.3) is finite and bounded uniformly in n, which here is

$$M := \operatorname*{esssup}_{t\geq 0} \mathbb{E}\left|\nabla W(X_t^1 - X_t^2) - \langle \mu_t, \nabla W(X_t^1 - \cdot)\rangle\right|^2. \quad (5.2)$$

The case $R < \infty$ is evident. If $L < \infty$ then ∇W has linear growth, and we conclude from the second moment bounds of [11, Corollary 2.3, Proposition 2.7]. □

5.2 Models on the torus: Proof of Corollary 2.9

The weak well-posedness of the (linear) Fokker–Planck equation (2.13) is standard, as K is bounded and Lipschitz. The well-posedness of the SDE (2.11) follows by the standard argument of Sznitman [50, Chapter I.1], since K is Lipschitz. The well-posedness of the PDE (2.12) in the weak sense follows by the superposition principle [22, Theorem 2.6]. The fact that the solution is classical follows from standard results on linear Fokker–Planck equations $\partial_t\mu = -\mathrm{div}(\tilde{K}\mu) + (\sigma^2/2)\Delta\mu$, applied with the bounded Lipschitz drift $\tilde{K} = K * \mu$. Strict positivity follows from Harnack's inequality [5, Chapter 8].

The following lemma paves the way for a short proof of Corollary 2.9, given just after:

Lemma 5.1 *For each $t > 0$, the density of μ_t is C^2 and obeys the pointwise bound*

$$\frac{1}{\lambda e^{r_0}} \le \mu_t(x) \le \frac{\lambda}{1 - r_0 e^{r_0}}, \quad \text{where} \quad r_0 := \frac{\|\operatorname{div} K\|_\infty \sqrt{2\log\lambda}}{\sigma^2\pi^2 - \|\operatorname{div} K\|_\infty},$$

Moreover, it holds that $r_0 < 1/2$, and μ_t satisfies the LSI

$$H(\cdot \,|\, \mu_t) \le \eta I(\cdot \,|\, \mu_t), \quad \text{where} \quad \eta := \lambda^2/(1 - 2r_0).$$

Proof We note for future use that [10, Proposition 3.1] shows

$$H(\mu_t \,|\, \mathbf{1}) \le e^{-2ct} H(\mu_0 \,|\, \mathbf{1}), \quad \forall t \ge 0, \ \text{where } c := \sigma^2\pi^2 - \|\operatorname{div} K\|_\infty, \tag{5.3}$$

and where $\mathbf{1}$ denotes the uniform measure on $\mathbb{T}^d$. The precise setting of [10, Proposition 3.1] is somewhat different, as they take $K = \nabla W$ as a gradient field and have a sharper estimate involving only the "unstable part" ΔW_u in place of $\operatorname{div} K$ in the definition of c, but their proof works (and becomes slightly simpler) to yield (5.3). We next prove the pointwise bound on μ_t, following an idea from the proof of [26, Theorem 2]. Fix $T > 0$ and $x \in \mathbb{T}^d$, and let $(Y_t)_{t\in[0,T]}$ denote the unique strong solution of the SDE

$$dY_t = -K * \mu_{T-t}(Y_t)dt + \sigma dB_t, \quad Y_0 = x.$$

Using the PDE for μ and Itô's formula, then taking expectations, we have for $t \in [0, T]$

$$\mathbb{E}\mu_{T-t}(Y_t) = \mu_T(x) + \mathbb{E}\int_0^t \mu_{T-s}(Y_s)\operatorname{div} K * \mu_{T-s}(Y_s)\, ds. \tag{5.4}$$

Note if $\operatorname{div} K \equiv 0$, as in [26, Theorem 2], then this would immediately yield the desired conclusion $\mu_T(x) = \mathbb{E}\mu_0(Y_T) \in [\lambda^{-1}, \lambda]$. In general, we need an additional argument. Noting that $\operatorname{div} K * \mathbf{1} \equiv 0$, we have for any $u \in [0, T]$ that

$$\begin{aligned}\|\operatorname{div} K * \mu_u\|_\infty &\le \|\operatorname{div} K * (\mu_u - \mathbf{1})\|_\infty \le \|\operatorname{div} K\|_\infty \|\mu_u - \mathbf{1}\|_{\mathrm{TV}} \\ &\le \|\operatorname{div} K\|_\infty \sqrt{2H(\mu_u \,|\, \mathbf{1})} \le \sqrt{2\log\lambda}\|\operatorname{div} K\|_\infty e^{-cu},\end{aligned}$$

with the last step using (5.3) and $H(\mu_u \,|\, \mathbf{1}) \le \log\lambda$. Setting $a = \sqrt{2\log\lambda}\|\operatorname{div} K\|_\infty$, this implies

$$\mathbb{E}\mu_{T-t}(Y_t) \le \mu_T(x) + a\int_0^t e^{-cs}\mathbb{E}\mu_{T-s}(Y_s)\, ds.$$

By Gronwall's inequality, we obtain for all $t \in [0, T]$

$$\mathbb{E}\mu_{T-t}(Y_t) \le \mu_T(x)\exp\left(a\int_0^t e^{-cs} ds\right) \le \mu_T(x)e^{a/c}.$$

Setting $t = T$ yields $\mu_T(x) \geq e^{-a/c}\mathbb{E}\mu_0(Y_T) \geq e^{-a/c}\lambda^{-1}$, as well as by (5.4)

$$\mu_T(x) \leq \mathbb{E}\mu_0(Y_T) + a\mathbb{E}\int_0^T e^{-cs}\mu_{T-s}(Y_s)\,ds \leq \mathbb{E}\mu_0(Y_T) + \frac{a}{c}\mu_T(x)e^{a/c}.$$

Combining the last two inequalities, we find

$$e^{-a/c}\lambda^{-1} \leq \mu_T(x) \leq \mathbb{E}\mu_0(Y_T)(1-(a/c)e^{a/c})^{-1} \leq \lambda(1-(a/c)e^{a/c})^{-1}.$$

This yields the claimed bounds on the density μ_t. The assumption (2.14) ensures that $r_0 < 1/2$. Finally, the claimed LSI follows from the Holley-Stroock argument [3, Proposition 5.1.6], after noting that $\sup \mu_t / \inf \mu_t \leq \lambda^2 e^{r_0}/(1-r_0e^{r_0}) \leq \lambda^2/(1-2r_0)$ since $0 \leq r_0 < 1/2$. □

Proof of Corollary 2.9 Note that Lemma 5.1 justifies the LSI. By Pinsker's inequality, for any $y \in \mathbb{T}^d$ we have

$$\big|\langle \nu - \mu_t, K(x-\cdot)\rangle\big|^2 = \big|\langle \nu - \mu_t, K(x-\cdot) - y\rangle\big|^2 \leq 2\||K-y|^2\|_\infty H(\nu \,|\, \mu_t).$$

so we may take

$$\gamma = 2\inf_{y\in\mathbb{T}^d}\||K-y|^2\|_\infty \leq \frac{1}{2}\sup_{x,z\in\mathbb{T}^d}|K(x)-K(z)|^2 = \frac{1}{2}\mathrm{diam}^2(K).$$

Then $r_c = \frac{\sigma^4}{4\gamma\eta} - 1$ simplifies to the claimed value. □

Appendix A. A form of Gronwall's inequality

Lemma A.1 *Let $c > 0$. Let $(g_t)_{t\geq 0}$ be a non-negative measurable function, integrable on bounded sets. Let $(H_t)_{t\geq 0}$ be a non-negative measurable function such that $H_0 < \infty$. Suppose that*

$$H_t - H_s \leq \int_s^t (g_u - cH_u)\,du, \quad \textit{for all } t \geq s \geq 0. \tag{A.1}$$

Then

$$H_t \leq e^{-ct}H_0 + \int_0^t e^{-c(t-u)}g_u du, \quad \textit{for all } t \geq 0. \tag{A.2}$$

Proof Mollify g and H by a common smooth probability density of compact support. The mollifications satisfy (A.1). Applying the usual differential Gronwall's inequality leads to (A.2) for the mollifications. Taking L^1 limits to remove the mollification leads to the original claim (A.2), at least for *almost every* t. But the inequality (A.1) implies that $\limsup_{t\to s} H_t \leq H_s$ for every $s \geq 0$, and the claim is thus valid for *every* $t \geq 0$. □

References

1. Acebrón, J.A., Bonilla, L.L., Vicente, C.J.P., Ritort, F., Spigler, R.: The Kuramoto model: a simple paradigm for synchronization phenomena. Rev. Mod. Phys. **77**(1), 137 (2005)
2. Arnaudon, M., Del Moral, P.: A second order analysis of McKean–Vlasov semigroups. Ann. Appl. Probab. **30**(6), 2613–2664 (2020)
3. Bakry, D., Gentil, I., Ledoux, M.: Analysis and Geometry of Markov Diffusion Operators. Springer, Berlin (2014)
4. Bertini, L., Giacomin, G., Pakdaman, K.: Dynamical aspects of mean field plane rotators and the Kuramoto model. J. Stat. Phys. **138**(1), 270–290 (2010)
5. Bogachev, V.I., Krylov, N.V., Röckner, M., Shaposhnikov, S.V.: Fokker–Planck–Kolmogorov Equations, vol. 207. American Mathematical Society, Amsterdam (2015)
6. Bogachev, V.I., Röckner, M., Shaposhnikov, S.V.: Distances between transition probabilities of diffusions and applications to nonlinear Fokker–Planck–Kolmogorov equations. J. Funct. Anal. **271**(5), 1262–1300 (2016)
7. Bresch, D., Jabin, P.-E., Soler, J.: A new approach to the mean-field limit of Vlasov–Fokker–Planck equations, arXiv preprint arXiv:2203.15747 (2022)
8. Bresch, D., Jabin, P.-E., Wang, Z.: Mean-field limit and quantitative estimates with singular attractive kernels, arXiv preprint arXiv:2011.08022 (2020)
9. Brunick, B., Shreve, S.: Mimicking an Itô process by a solution of a stochastic differential equation. Ann. Appl. Probab. **23**(4), 1584–1628 (2013)
10. Carrillo, J.A., Gvalani, R.S., Pavliotis, G.A., Schlichting, A.: Long-time behaviour and phase transitions for the McKean–Vlasov equation on the torus. Arch. Ration. Mech. Anal. **235**(1), 635–690 (2020)
11. Cattiaux, P., Guillin, A., Malrieu, F.: Probabilistic approach for granular media equations in the non-uniformly convex case. Probab. Theory Relat. Fields **140**, 19–40 (2008)
12. Chaintron, L.-P., Diez, A.: Propagation of chaos: a review of models, methods and applications. II. Applications, arXiv preprint arXiv:2106.14812 (2021)
13. Chaintron, L.-P., Diez, A.: Propagation of chaos: a review of models, methods and applications. I. Models and methods, arXiv preprint arXiv:2203.00446 (2022)
14. Chizat, L., Bach, F.: On the global convergence of gradient descent for over-parameterized models using optimal transport. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp 3040–3050 (2018)
15. Collet, J.-F., Malrieu, F.: Logarithmic Sobolev inequalities for inhomogeneous Markov semigroups. ESAIM: Probab. Stat. **12**, 492–504 (2008)
16. Dawson, D.A.: Critical dynamics and fluctuations for a mean-field model of cooperative behavior. J. Stat. Phys. **31**(1), 29–85 (1983)
17. Del Moral, P., Tugaut, J.: Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions. Stoch. Anal. Appl. **37**(6), 909–935 (2019)
18. Delarue, F., Tse, A.: Uniform in time weak propagation of chaos on the torus, arXiv (2021)
19. Delgadino, M.G., Gvalani, R.S., Pavliotis, G.A., Smith, S.A.: Phase transitions, logarithmic Sobolev inequalities, and uniform-in-time propagation of chaos for weakly interacting diffusions, arXiv preprint arXiv:2112.06304 (2021)
20. Diaconis, P., Freedman, D.: Finite exchangeable sequences. Ann. Probab. **8**, 745–764 (1980)
21. Durmus, A., Eberle, A., Guillin, A., Zimmer, R.: An elementary approach to uniform in time propagation of chaos. Proc. Am. Math. Soc. **148**, 5387–5398 (2020)
22. Figalli, A.: Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. J. Funct. Anal. **254**(1), 109–153 (2008)
23. Gautschi, W.: Some elementary inequalities relating to the gamma and incomplete gamma function. J. Math. Phys. **38**, 77–81 (1959)
24. Golse, F.: On the dynamics of large particle systems in the mean field limit, Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, pp. 1–144. Springer, Berlin (2016)
25. Gozlan, N., Léonard, C.: Transport inequalities: A survey. Markov Process. Relat. Fields **16**, 635–736 (2010)
26. Guillin, A., Le Bris, P., Monmarché, P.: Uniform in time propagation of chaos for the 2D vortex model and other singular stochastic systems, arXiv preprint arXiv:2108.08675 (2021)
27. Guillin, A., Liu, W., Wu, L. Zhang, C.: Uniform poincaré and logarithmic Sobolev inequalities for mean field particles systems, arXiv preprint arXiv:1909.07051 (2019)

28. Guillin, A., Monmarche, P.: Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes. J. Stat. Phys. **185**(15), 1–20 (2021)
29. Han, J., Hu, R., Long, J.: A class of dimensionality-free metrics for the convergence of empirical measures, arXiv preprint arXiv:2104.12036 (2021)
30. Han, Y.: Entropic propagation of chaos for mean field diffusion with $l_t^q - l_x^p$ interaction, linear growth interaction and fractional noise, arXiv preprint arXiv:2205.02772 (2022)
31. Herrmann, S., Tugaut, J.: Non-uniqueness of stationary measures for self-stabilizing processes. Stoch. Process. Their Appl. **120**(7), 1215–1246 (2010)
32. Jabin, P.-E., Poyato, D., Soler, J.: Mean-field limit of non-exchangeable systems, arXiv preprint arXiv:2112.15406 (2021)
33. Jabin, P.-E., Wang, Z.: Mean field limit and propagation of chaos for Vlasov systems with bounded forces. J. Funct. Anal. **271**(12), 3588–3627 (2016)
34. Jabin, P.-E., Wang, Z.: Mean Field Limit for Stochastic Particle Systems, Active Particles, vol. 1, pp. 379–402. Springer, Berlin (2017)
35. Jabin, P.-E., Wang, Z.: Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. Invent. Math. **214**(1), 523–591 (2018)
36. Jameson, G.J.O.: Inequalities for gamma function ratios. Am. Math. Mon. **120**, 936–940 (2013)
37. Kac, M.: Foundations of kinetic theory. In: Proceedings of The third Berkeley symposium on mathematical statistics and probability, vol. 3, University of California Press Berkeley and Los Angeles, California, pp. 171–197 (1956)
38. Lacker, D.: Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions (2021, preprint)
39. Lacker, D.: Quantitative approximate independence for continuous mean field gibbs measures. (2021, preprint)
40. Malrieu, F.: Logarithmic sobolev inequalities for some nonlinear PDE's. Stoch. Process. Their Appl. **95**, 109–132 (2001)
41. Malrieu, F.: Convergence to equilibrium for granular media equations and their Euler schemes. Ann. Appl. Probab. **13**(2), 540–560 (2003)
42. McKean, H.P.: Propagation of chaos for a class of non-linear parabolic equations. In: Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967), pp 41–57 (1967)
43. Mei, S., Montanari, A., Nguyen, P.-M.: A mean field view of the landscape of two-layer neural networks. In: Proceedings of the National Academy of Sciences of the United States of America **115**(33) (2018)
44. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. **173**(2), 361–400 (2000)
45. Rosenzweig, M., Serfaty, S.: Global-in-time mean-field convergence for singular Riesz-type diffusive flows (2021). arXiv preprint arXiv:2108.09878
46. Salem, S.: A gradient flow approach of uniform in time propagation of chaos for particles in double a well confinement (2018). arXiv preprint arXiv:1810.08946
47. Salem, S.: A gradient flow approach of propagation of chaos. Discrete Contin. Dyn. Syst. **40**(10), 5729 (2020)
48. Salhi, J., MacLaurin, J., Toumi, S.: On uniform propagation of chaos. Stochastics **90**(1), 49–60 (2013)
49. Sirignano, J., Spiliopoulos, K.: Mean field analysis of neural networks: a law of large numbers. SIAM J. Appl. Math. **80**(2), 1 (2020). https://doi.org/10.1137/18M1192184
50. Sznitman, A.-S.: Topics in propagation of chaos, Ecole d'été de probabilités de Saint-Flour XIX-1989, pp. 165–251. Springer, Berlin (1991)