The Annals of Applied Probability
2016, Vol. 26, No. 2, 892–916
DOI: 10.1214/15-AAP1107

QUANTITATIVE PROPAGATION OF CHAOS FOR GENERALIZED KAC PARTICLE SYSTEMS

BY ROBERTO CORTEZ[1] AND JOAQUIN FONTBONA[2]

Universidad de Chile

We study a class of one-dimensional particle systems with true (Bird type) binary interactions, which includes Kac's model of the Boltzmann equation and nonlinear equations for the evolution of wealth distribution arising in kinetic economic models. We obtain explicit rates of convergence for the Wasserstein distance between the law of the particles and their limiting law, which are linear in time and depend in a mild polynomial manner on the number of particles. The proof is based on a novel coupling between the particle system and a suitable system of nonindependent nonlinear processes, as well as on recent sharp estimates for empirical measures.

1. Introduction and main result.

1.1. *The kinetic equation.* We consider the collection $(P_t)_{t\geq 0}$ of probability measures on $\mathbb{R}$, solution of the following nonlinear kinetic-type equation:

$$\partial_t P_t = -P_t + \mathcal{Q}^+(P_t). \tag{1}$$

Here, $\mathcal{Q}^+$ is a generalized Wild convolution, which associates with every measure μ on $\mathbb{R}$ a new measure $\mathcal{Q}^+(\mu)$ given by

$$\int \phi(u)\mathcal{Q}^+(\mu)(du) = \iint \frac{1}{2}\mathbf{E}\big(\phi(Lu+Rv) + \phi(\tilde{L}v+\tilde{R}u)\big)\mu(dv)\mu(du), \tag{2}$$

for all bounded measurable functions ϕ, where $(L, R, \tilde{L}, \tilde{R})$ is a given random vector in $\mathbb{R}^4$ (with known distribution) and $\mathbf{E}$ denotes the expectation with respect to it.

Equations (1)–(2) describe the behavior of an infinite number of objects or "particles" subjected to binary interactions. The state of each particle is characterized by a scalar $u \in \mathbb{R}$, and $P_t(du)$ represents the proportion of particles in state u at time $t \geq 0$. The microscopic binary interactions, which occur randomly at constant rate, are heuristically described as follows: when a particle at state u interacts

Received June 2014; revised February 2015.

[1]Supported by Proyecto Mecesup UCH0607 Doctoral Fellowship.

[2]Supported by Fondecyt Grant 1110923, Basal-CONICYT Center for Mathematical Modeling (CMM), and Millenium Nucleus NC120062.

MSC2010 subject classifications. Primary 60K35; secondary 82C22, 82C40.

Key words and phrases. Propagation of chaos, Kac equation, wealth distribution equations, stochastic particle systems, Wasserstein distance, optimal coupling.

with a particle at state v, their states change according to the rule

$$(u, v) \mapsto (Lu + Rv, \tilde{L}v + \tilde{R}u). \tag{3}$$

This model is a generalization of Kac's one-dimensional simplification of the (more realistic) Boltzmann equation for a spatially homogeneous dilute gas in $\mathbb{R}^3$, in which the interacting objects represent actual physical particles. Specifically, in Kac's model introduced in Kac (1956), the state of a particle is its one-dimensional velocity, and the interactions correspond to random exchanges of velocities that occur at binary collisions that preserve kinetic energy, so that $L = \cos\theta = \tilde{L}$, $R = -\sin\theta = -\tilde{R}$, with θ randomly chosen in $[0, 2\pi)$. We refer the reader to Mischler and Mouhot (2013) and the references therein for historical background on Kac and Boltzmann's equations.

A further source of models of the type described by equations (1)–(2) is the kinetic description of the evolution of the wealth distribution in a simplified economy, studied, for instance, in Matthes and Toscani (2008) (see also the references therein). In that setting, the state of a particle represents the wealth of an economic agent, and the binary interactions correspond to trades or economic exchanges between them. Early versions of that model assumed

$$|L|^p + |\tilde{R}|^p = 1 \quad \text{a.s.}, \qquad |\tilde{L}|^p + |R|^p = 1 \quad \text{a.s.}, \tag{4}$$

for some $p \geq 1$ [notice that in Kac's model (4) is satisfied with $p = 2$]. In the case $p = 1$, for nonnegative L, R, $\tilde{L}$ and $\tilde{R}$, condition (4) can be seen as *exact* conservation of total wealth in each interaction. The weaker condition

$$\mathbf{E}\big(|L|^p + |\tilde{R}|^p\big) = 1, \qquad \mathbf{E}\big(|\tilde{L}|^p + |R|^p\big) = 1, \tag{5}$$

interpreted as conservation of wealth only *in the mean* (so that risky trades with possible gain or loss of total wealth in each interaction are allowed), has also been considered in order to obtain wider classes of equilibrium distributions for the nonlinear dynamics [see Matthes and Toscani (2008), Bassetti, Ladelli and Matthes (2011)].

1.2. *Particle system and propagation of chaos*. In order to rigorously justify the interpretation of the model (1)–(2) as representing the evolution of an infinite number of interacting particles or agents, one considers a finite system of N of such particles, which we denote $\mathbf{X}_t = (X_t^1, \dots, X_t^N)$, starting independently with common law P_0 and such that, at each binary interaction, the states of both involved particles are modified according to the rule (3). In the terminology of particle approximations of the Boltzmann equation, a particle system with such (true) binary interactions is called of *Bird type*, as opposed to particle systems of *Nanbu type*, in which only one particle changes its state after interaction with some other.

Specifically, the particle system $\mathbf{X}$ has infinitesimal generator

$$\mathcal{A}^N \phi(\mathbf{x}) = \frac{1}{2(N-1)} \sum_{i \neq j} \int_{\mathbb{R}^4} \big[\phi\big(\mathbf{x} + a_{ij}(\eta, x^i, x^j)\big) - \phi(\mathbf{x})\big] \Lambda(d\eta) \tag{6}$$

for all $\mathbf{x} = (x^1, \dots, x^N) \in \mathbb{R}^N$ and every test function ϕ on $\mathbb{R}^N$, where $\eta = (\xi, \zeta, \tilde{\xi}, \tilde{\zeta})$ denotes a generic point in $\mathbb{R}^4$, Λ is the joint law of $(L, R, \tilde{L}, \tilde{R})$ and $a_{ij}(\eta, u, v)$ is the vector of $\mathbb{R}^N$ whose ith and jth components are $(\xi - 1)u + \zeta v$ and $(\tilde{\xi} - 1)v + \tilde{\zeta} u$, respectively, and which is equal to 0 in the other components.

Convergence of such a particle system, more precisely of its empirical measures $\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$ toward the unique solution $(P_t)_{t\geq 0}$ of the nonlinear evolution (1) as N goes to infinity, has been studied in more general frameworks and from several points of view; see, for instance, Graham and Méléard (1997), Mischler and Mouhot (2013) and the references therein [in particular, well posedness of (1) is by now standard]. Since the particles are exchangeable, the convergence of the empirical measure to P_t for large N, as a random variable in the space of probability measures in $\mathbb{R}$ endowed with the weak topology, is equivalent to the property of *propagation of chaos* of $\mathbf{X}_t$ with respect to P_t [see Sznitman (1991) for background]: for every fixed $k \in \mathbb{N}$, the joint law of $X_t^1, \dots, X_t^k$ converges weakly to $P_t^{\otimes k}$ as N goes to ∞. That is, when N is large, any fixed number of particles of the system behaves at time t approximately like independent random variables of law P_t. This property was introduced and first established by Kac himself in Kac (1956) for the particle system bearing his name, and is nowadays known to hold for a large class of particle models, under general mild assumptions.

1.3. *Main result.* Typically, weak convergence results are not sufficiently informative, and one looks for more quantitative statements. In this article, we will study the Bird-type N-particle system $\mathbf{X} = (X^1, \dots, X^N)$ and its propagation of chaos property, in the cases $p = 1$ and $p = 2$. Our main goal is to obtain rates of convergence, as $N \to \infty$, for the Wasserstein distance between the empirical measure of the particle system at time t and its limiting law P_t, with explicit estimates on N and t that grow reasonably fast as functions of t.

Let $p \in \{1, 2\}$ be fixed. In the case $p = 2$, we will assume the additional condition $\mathbf{E}(LR + \tilde{L}\tilde{R}) = 0$, which is certainly satisfied in Kac's model. As a generalization of (5), we will work under the assumption

$$\frac{1}{2}\mathbf{E}\big(|L|^p + |R|^p + |\tilde{L}|^p + |\tilde{R}|^p\big) \leq 1. \tag{7}$$

With some abuse of language, for each value of $p \in \{1, 2\}$ we will say that the model is *inelastic* if the latter inequality is strict. In that case, the interaction between particles produce an average loss of energy when $p = 2$ [see, e.g., the inelastic Kac model in Pulvirenti and Toscani (2004)] or of "wealth" [in the context of Matthes and Toscani (2008)] when $p = 1$. Also, to avoid trivial situations, in all what follows we will assume that the model is nondegenerate, that is, $\mathbf{E}(|R| + |\tilde{R}|) > 0$; this means that the system produces at least some effective interactions.

Let us fix some notation. $\mathcal{P}(E)$ denotes the space of probability measures on the metric space E. For $\mathbf{x} \in \mathbb{R}^N$ and any $i = 1, \dots, N$ we define the empirical measures $\bar{\mathbf{x}} = \frac{1}{N}\sum_j \delta_{x^j}$ and $\bar{\mathbf{x}}^i = \frac{1}{N-1}\sum_{j\neq i} \delta_{x^j}$, both being elements of $\mathcal{P}(\mathbb{R})$. Define

$M_q(\mu) = \int |u|^q \mu(du)$ the absolute q-moment of $\mu \in \mathcal{P}(\mathbb{R})$. Given a random vector $\mathbf{Z}$ on $\mathbb{R}^N$, we denote its law by $\mathcal{L}(\mathbf{Z}) \in \mathcal{P}(\mathbb{R}^N)$, and the joint law of its first k components by $\mathcal{L}^k(\mathbf{Z}) \in \mathcal{P}(\mathbb{R}^k)$.

Recall that for $\mu, \nu \in \mathcal{P}(\mathbb{R}^k)$ their p-Wasserstein distance $\mathcal{W}_p(\mu, \nu)$ is defined to be the cost of the optimal transfer plan between μ and ν, that is,

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_\pi \int_{\mathbb{R}^k \times \mathbb{R}^k} d_{k,p}(\mathbf{x}, \mathbf{y})^p \pi(d\mathbf{x}, d\mathbf{y})\right)^{1/p} = \left(\inf_{\theta, \vartheta} \mathbb{E} d_{k,p}(\theta, \vartheta)^p\right)^{1/p},$$

where the first infimum is taken over all measures π on $\mathbb{R}^k \times \mathbb{R}^k$ with marginals μ and ν, and the second infimum is taken over all pairs of random vectors θ and ϑ such that $\mathcal{L}(\theta) = \mu$ and $\mathcal{L}(\vartheta) = \nu$ [see, e.g., Villani (2009) for background on Wasserstein distances]. We will use the *normalized* distance $d_{k,p}$ on $\mathbb{R}^k$ given by

$$d_{k,p}(\mathbf{x}, \mathbf{y}) = \left(\frac{1}{k} \sum_{i=1}^k |x^i - y^i|^p\right)^{1/p}, \tag{8}$$

which is natural when one cares about the dependence on the dimension.

In order to obtain good rates of convergence in N which moreover are well behaved with respect to t, it is convenient to introduce the concave function

$$\alpha_q = 1 - \tfrac{1}{2}\mathbf{E}\big(|L|^q + |R|^q + |\tilde{L}|^q + |\tilde{R}|^q\big) \qquad \forall q \geq 0.$$

We also define

$$q^* = \sup\{q : M_q(P_0) < \infty, \alpha_q > 0\}.$$

These objects play an important role in Matthes and Toscani (2008), since when $p = 1$ and q^* is nontrivial (i.e., $1 < q^* < \infty$), q^* corresponds to the Pareto index of the stationary distribution of P_t. More importantly, in the present context, the moments of order $q < q^*$ of P_t can be controlled uniformly in time (see Lemma 5 below). Assuming (7) and $M_p(P_0) < \infty$, the concavity of α_q implies that either $q^* \in [p, \infty]$ or $q^* = -\infty$. Also, define for all $q \in \{p\} \cup (p, q^*)$

$$\bar{\alpha}_{p,q} = \inf_{p \leq r \leq q} \alpha_r = \min(\alpha_p, \alpha_q).$$

Note that if $\alpha_p = 0$, then $\bar{\alpha}_{p,q} = 0$ for all such q, so this function is meaningful only in the case $\alpha_p > 0$, in which case it will be useful to obtain uniform (in time) estimates.

We are now ready to state our main theorem (see also Corollary 8 for a trajectorial result).

THEOREM 1. *Let $(P_t)_{t \geq 0}$ be the unique solution of (1) and let $\mathbf{X}$ be the particle system starting with law $P_0^{\otimes N}$ and with generator (6). For $p = 1$ or $p = 2$, assume $\alpha_p \geq 0$ and $M_p(P_0) < \infty$. If $p = 2$, assume also that $\mathbf{E}(LR + \tilde{L}\tilde{R}) = 0$ and $q^* > 2$. Then*:

- *for any* $q \in \{1\} \cup (1, q^*)$ *and any* $\gamma < (2 + 1/q)^{-1}$ *in the case* $p = 1$, *or*
- *for any* $q \in (2, q^*)$, $q \neq 4$ *and for* $\gamma = \min(1/3, \frac{q-2}{2q-2})$ *in the case* $p = 2$,

there exists a constant C, *depending on* p, q, γ *and some moments of* P_0 *and* $(L, R, \tilde{L}, \tilde{R})$ *of order at most* q, *such that*:

(i) *for all* $k \leq N$ *and for all* $t \geq 0$,

$$\mathcal{W}_p^p(\mathcal{L}^k(\mathbf{X}_t), P_t^{\otimes k}) \leq C\left(\frac{t(1+t)^{p-1}e^{-(p/q)\bar{\alpha}_{p,q}t}}{N^\gamma} + \mathbb{1}_{k\neq 1}\frac{k\min(1,t)e^{-\alpha_p t}}{N}\right),$$

(ii) *for all* $t \geq 0$,

$$\mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{X}}_t, P_t) \leq \frac{C(1+t)^p e^{-(p/q)\bar{\alpha}_{p,q}t}}{N^\gamma}.$$

REMARK 2. • The power γ in Theorem 1 is a consequence of using recently established sharp quantitative estimates in Wasserstein distance for the empirical measures of exchangeable or i.i.d. collections of random variables [which improve or extend a classical result in Rachev and Rüschendorf (1998)]. More specifically, the rate $N^{-\gamma}$ with $\gamma < (2 + 1/q)^{-1}$ in the case $p = 1$ comes from Theorem 1.2 of Hauray and Mischler (2014), whereas the value $\gamma = \min(1/3, \frac{q-2}{2q-2})$ in the case $p = 2$ comes from Theorem 1 of Fournier and Guillin (2013). On the other hand, the dependence on t results from our estimates, which rely on Gronwall's lemma.

• The restriction $q \neq 4$ in the case $p = 2$ comes from Theorem 1 of Fournier and Guillin (2013). As those authors mention, the case $q = 4$ would produce additional logarithmic terms, which in our case translate into a rate of order $N^{-1/3}$ times a logarithmic function of N.

• In the elastic case (i.e., $\alpha_p = 0 = \bar{\alpha}_{p,q}$), (i) and (ii) give estimates that grow linearly with time (in the case $p = 2$ both sides are squared). In the inelastic case, which corresponds to $\alpha_p, \bar{\alpha}_{p,q} > 0$, all estimates are uniform in time.

• From a physical point of view, it is interesting to consider models where infinitely many particles interact over finite time intervals, such as the Kac equation without cutoff. The techniques used in the proof of Theorem 1 can also be applied to cutoffed approximations of that equation and, in the case that [in the notation of Desvillettes, Graham and Méléard (1999)] the classical condition $\int_0^\pi \theta^2\beta(\theta)\,d\theta < \infty$ on the cross-section function $\beta : [-\pi, \pi] \to \mathbb{R}_+$ is satisfied, they yield a constant that does not depend on the cutoff parameter; see Remark 9.

1.4. *Particular cases and comparison with known results.*

1.4.1. *The Kac equation.* Note that if the stronger condition (4) is satisfied (or holds with $\leq$ instead of equality), then $|L|$, $|R|$, $|\tilde{L}|$ and $|\tilde{R}|$ are all ≤ 1 a.s., which implies that α_q is strictly increasing with q. Thus, $\bar{\alpha}_{p,q} = \alpha_p$ for $q \geq p$ and the value of q^* will depend only on the finiteness of the moments of P_0. In Kac's model, since (4) is satisfied for $p = 2$, if P_0 has finite moment of order $4 + \varepsilon$, then $q^* > 4$ and Theorem 1 gives

$$\mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{X}}_t, P_t) \leq \frac{C(1+t)^2}{N^{1/3}}.$$

Several similar results can be found in the literature. The closest one corresponds to quantitative rates for the Nanbu system associated with Kac's model, which are found, for instance, in the proof of Proposition 6.2 of Fournier and Godinho (2012). The authors state there a $\mathcal{W}_2^2$ convergence rate that also depends quadratically on t and is optimal on N, in the sense that it is equal to the $\mathcal{W}_2^2$ rate of convergence of the empirical measure of an i.i.d. sample toward their common law. The latter is of order $N^{-1/2}$, according to Theorem 1 of Fournier and Guillin (2013). Thus, the Bird-type particle system seems to produce a slower rate of convergence than the corresponding Nanbu-type system. An interesting question is whether this difference is a mere consequence of the techniques used in our proof (more specifically, some order is lost when one uses Lemma 7) or is intrinsically related to the type of binary interactions (Bird or Nanbu) in the system.

A similar result as the one of Fournier and Godinho (2012) can be found in Fournier and Mischler (2014) where, motivated by the numerical approximation of the Boltzmann equation for hard spheres, hard potentials and Maxwellian gases, a pathwise coupling argument was developed for Nanbu particle systems, which extends a coupling construction based on optimal transport developed in Fontbona, Guérin and Méléard (2009). That pathwise approach, however, does not readily extend to the particle systems of Bird type we are interested in, which in turn provide a physically more transparent description of the relevant interaction phenomena.

As for the Bird particle system, in Graham and Méléard (1997) the authors obtain an explicit rate in total variation distance on the path space, between the law of one particle and the law of the nonlinear process (to be introduced later). However, due to the generality of their hypotheses and the strong pathwise distance they use, the convergence rate depends exponentially on the length of the time interval that is considered. Similarly, in Theorem 4.3 of Desvillettes, Graham and Méléard (1999) the authors state a propagation of chaos result in $\mathcal{W}_2$ for the law at time t of one particle in the system with cutoff, toward the law P_t of the nonlinear dynamics without cutoff. Since some relations between N and the cutoff parameter must be satisfied when removing the latter, that result gives estimates that are logarithmic in N and grow exponentially with t.

On the other hand, the general theory developed in Mischler and Mouhot (2013) provides a framework and a methodology to establish quantitative (in t and N) propagation of chaos estimates which can be applied in the present framework. For instance, in their Theorem 5.2, a $\mathcal{W}_1$ estimate for the Boltzmann equation in the Maxwell molecules case is obtained, which is uniform in time and decays with N in a polynomial way [see also step 3 of the proof of Theorem 8 in Carrapatoso (2014a) for results in $\mathcal{W}_2$ distance]; we expect that similar bounds can be obtained with their techniques for the Kac model. The actual dependence on N they give seems however hard to trace in general, and we have not been able to deduce with their techniques an estimate in Wasserstein distance as sharp as ours in terms of N. Also, their approach does not provide any information on the way in which trajectories of particles get closer to those of the limiting processes. On the other side, unfortunately our techniques (ultimately relying on Gronwall's lemma) do not seem to yield uniform in time estimates for the elastic Kac equation, even if P_0 were compactly supported.

Finally, we observe that for the inelastic Kac model, $\bar{\alpha}_{2,q} = \alpha_2 > 0$ for all q, hence Theorem 1 does give a uniform-in-time rate in that case:

$$\mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{X}}_t, P_t) \leq \frac{Ce^{-(\alpha_2/(2+\varepsilon))t}}{N^{1/3}},$$

if P_0 is again assumed to have finite moment of order $q > 4$. This exponential decay is not surprising: when P_0 has finite moment of order 2 it is known that $M_2(P_t)$ decays exponentially fast; see Pulvirenti and Toscani (2004). Nevertheless, to our knowledge, our quantitative (in N) propagation of chaos result is new for the inelastic Kac model.

1.4.2. *Models for economic exchanges and wealth distribution.* Working with $p = 1$ as in Matthes and Toscani (2008) and assuming only that the first moment of P_0 is finite, Theorem 1 gives in the elastic case a propagation of chaos result for $\mathcal{W}_1$ of order almost $N^{-1/3}$, with estimates growing linearly with time. Any additional finite q-moment of P_0 (with $q < q^*$) can be used to improve the rate in N, up to almost $N^{-1/(2+1/q^*)}$. In the case of exact conservation of wealth [condition (4)] we have $q^* = \infty$ and we obtain a rate of $N^{-(1/2-\varepsilon)}$, which is almost optimal according to Theorem 1 of Fournier and Guillin (2013). To our knowledge, this is the first quantitative propagation of chaos result for kinetic equations modeling the evolution of wealth distribution.

1.5. *The nonlinear process and idea of the proof.* Following ideas pioneered by Tanaka in the case of the Boltzmann equation [see Tanaka (1978) and Tanaka (1978/1979)], it is also possible to establish the convergence of the *pathwise* law of a particle, to the law of some process obtained by the following construction: consider a Poisson point measure $\mathcal{M}$ on $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}$ with intensity $dt\bar{\Lambda}(d\xi, d\zeta)P_t(dv)$, where $\bar{\Lambda} = \frac{1}{2}(\mathcal{L}(L, R) + \mathcal{L}(\tilde{L}, \tilde{R}))$, and let $(V_t)_{t\geq 0}$ be the jump process on $\mathbb{R}$ defined as the unique solution starting with law P_0 of the stochastic equation

$$dV_t = \int_{\mathbb{R}^2}\int_{\mathbb{R}} [(\xi - 1)V_{t^-} + \zeta v]\mathcal{M}(dt, d\xi, d\zeta, dv). \tag{9}$$

It is not hard to see that such a jump process V exists, it is uniquely defined and it satisfies $\mathcal{L}(V_t) = P_t$ for all $t \geq 0$. We call P the pathwise law of V, and any process with law P is called a *nonlinear process*; it represents the trajectory of any fixed particle in the (infinite) population subjected to the random interactions described above in (3).

To prove our results, we will couple the Bird particle system $\mathbf{X}_t$ with a system $\mathbf{U}_t = (U_t^1, \dots, U_t^N)$ where each U^i is a copy of the nonlinear process V, constructed in such a way that it remains close to X^i. To achieve this, we will use techniques of optimal coupling inspired by those used in Fontbona, Guérin and Méléard (2009) and Fournier and Mischler (2014), in order to carefully choose the jumps of the nonlinear process U^i as similar as possible to those of the particle X^i. However, contrary to those papers which deal with Nanbu-type particle systems (in which each randomness source acts on the trajectory of only one of the particles), ensuring closeness of X^i and U^i simultaneously for all $i = 1, \dots, N$ will imply that the processes $U^1, \dots, U^N$ are *not* independent. Therefore, to obtain the desired estimates we will need, in a second step, to "decouple" the system $\mathbf{U}_t$ as N goes to infinity, which we will be able to do with estimates that are uniform in time; see Lemma 6 below.

Let us point out that the coupling construction we will introduce can in principle be replicated in higher dimensions, and with more general interaction rules, which is why we preferred to avoid the use of specific one-dimensional features in its construction; see, for instance, Remark 10. We thus expect these techniques to be applicable in physically more relevant situations, hopefully including (at least some instances of) the Boltzmann equation. Also, we think it should be possible to adapt this coupling construction in order to quantitatively study "Bird-type" Brownian particle approximations of a certain Gaussian white-noise driven nonlinear process, associated with the Landau equation arising in the grazing collisions limit of the Boltzmann equation. Such a process was studied in Funaki (1984) and Guérin (2003), and a particle approximation result with a "Nanbu type" Brownian particle system was proved in Fontbona, Guérin and Méléard (2009), by means of a coupling construction based on optimal transport. The corresponding particle system of Bird-type is studied in Carrapatoso (2014b) using the functional tools developed in Mischler and Mouhot (2013), but there seems to be so far no suitable coupling argument available in order to deal with such class of particle systems.[3]

1.6. *Plan of the paper.* In Section 2, we give the explicit construction of the particle system $\mathbf{X}_t$, and more importantly, we couple it with the system $\mathbf{U}_t = (U_t^1, \dots, U_t^N)$ of dependent nonlinear processes that we will use throughout the rest of this article. In Section 3, we prove Theorem 1. The proof of some intermediate lemmas, including statements of Section 2 and the "decoupling" of the process $\mathbf{U}_t$, is left for the final Section 4.

[3]When the present work was just finished, the authors learned from Nicolas Fournier that the latter question was currently being studied by him, François Bolley and Arnaud Guillin.

2. Coupling of the particle system and the nonlinear processes.

2.1. *The particle system.* Let us fix the number of particles $N \in \mathbb{N}$. Although most of the subsequent objects will depend on N, for notational simplicity we will not make this dependence explicit. We will define both the particle system $\mathbf{X}$ and the nonlinear processes $\mathbf{U}$ by means of integral equations driven by the same Poisson point measure. To this end, let us first introduce the function $\mathbf{i}:[0,N) \to \{1,\dots,N\}$ given by $\mathbf{i}(\rho) = \lfloor \rho \rfloor + 1$, and the set $\mathcal{C} \subseteq [0,N)^2$

$$\mathcal{C} = \{(\rho,\sigma) \in [0,N)^2 : \mathbf{i}(\rho) \neq \mathbf{i}(\sigma)\}.$$

Note that $|\mathcal{C}| = N(N-1)$. As in (6), denote $\eta = (\xi, \zeta, \tilde{\xi}, \tilde{\zeta})$ a generic point in $\mathbb{R}^4$ and $\Lambda = \mathcal{L}(L, R, \tilde{L}, \tilde{R})$. Now, let $\mathcal{N}(dt, d\eta, d\rho, d\sigma)$ be a Poisson point measure on $[0,\infty) \times \mathbb{R}^4 \times [0,N)^2$ with intensity

$$\frac{N}{2}\, dt\, \Lambda(d\eta)\, d\rho\, d\sigma \frac{1}{|\mathcal{C}|} \mathbb{1}_{\mathcal{C}}(\rho,\sigma) = \frac{1}{2(N-1)}\, dt\, \Lambda(d\eta)\, d\rho\, d\sigma\, \mathbb{1}_{\mathcal{C}}(\rho,\sigma).$$

In words, $\mathcal{N}$ picks atoms in $[0,\infty)$ at constant rate of $N/2$, and for each such atom it also independently samples a tuple $(\xi, \zeta, \tilde{\xi}, \tilde{\zeta})$ from Λ and a pair (ρ,σ) uniformly on $\mathcal{C}$. We will use (ρ,σ) to choose the indices of the particles that interact at each jump. Consider also N independent random variables $(X_0^1, \dots, X_0^N) =: \mathbf{X}_0$, independent from $\mathcal{N}$, each having distribution P_0. Finally, set $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ to be the complete right continuous filtration generated by $\mathbf{X}_0$ and $\mathcal{N}$. We denote $\mathbb{P}$ and $\mathbb{E}$ the probability and expectation in the corresponding probability space.

The particle system $\mathbf{X} = (X^1, \dots, X^N)$ is defined as the solution, starting from $\mathbf{X}_0$, of the following integral equation:

$$d\mathbf{X}_t = \int_{\mathbb{R}^4} \int_{[0,N)^2} \sum_{i,j=1}^{N} \mathbb{1}_{\{\mathbf{i}(\rho)=i, \mathbf{i}(\sigma)=j\}} a_{ij}(\eta, X_{t-}^i, X_{t-}^j) \mathcal{N}(dt, d\eta, d\rho, d\sigma). \tag{10}$$

[Recall that $a_{ij}(\eta, u, v)$ is the vector of $\mathbb{R}^N$ whose ith and jth components are $(\xi - 1)u + \zeta v$ and $(\tilde{\xi} - 1)v + \tilde{\zeta} u$, resp., and is equal to 0 in the other components]. Given the timely ordered atoms $(t_n, \eta_n, \rho_n, \sigma_n)_{n\geq 0}$ of $\mathcal{N}$ (i.e., $t_n \leq t_{n+1}$ for all $n \geq 0$), a solution of this equation can be constructed as follows: recursively define $\mathbf{X}_{t_n}$ as

$$X_{t_n}^{\ell} = \begin{cases} \xi_n X_{t_{n-1}}^i + \zeta_n X_{t_{n-1}}^j, & \ell = i, \\ \tilde{\xi}_n X_{t_{n-1}}^j + \tilde{\zeta}_n X_{t_{n-1}}^i, & \ell = j, \\ X_{t_{n-1}}^{\ell}, & \ell \neq i, j, \end{cases} \tag{11}$$

where $(i,j) = (\mathbf{i}(\rho_n), \mathbf{i}(\sigma_n))$, and set $\mathbf{X}_t = \mathbf{X}_{t_n}$ for all $t \in (t_n, t_{n+1})$. Uniqueness for (10) also holds, since there is no choice to make in this construction. It is straightforward to verify that $\mathbf{X}$ has generator (6).

Thus, the system $\mathbf{X}$ is what we want it to be: at rate $N/2$ we choose two distinct indices $i=\mathbf{i}(\rho)$ and $j=\mathbf{i}(\sigma)$, and then we update the particles X^i and X^j according to the rule described in (3). The fact that we use continuous variables (ρ,σ) to choose the indices (i,j) (instead of a discrete pair chosen uniformly from the set $\{1,\dots,N\}^2\setminus\{i=j\}$) will be crucial to define our system $\mathbf{U}$ of N nonlinear processes.

2.2. *Coupling with the nonlinear processes.* From (10), it follows that for any $i=1,\dots,N$, the process X^i satisfies

$$dX_t^i=\int_{\mathbb{R}^2}\int_{[0,N)}\big[(\xi-1)X_{t-}^i+\zeta X_{t-}^{\mathbf{i}(\tau)}\big]\mathcal{N}^i(dt,d\xi,d\zeta,d\tau), \tag{12}$$

where $\mathcal{N}^i$ is defined as

$$\begin{aligned}\mathcal{N}^i(dt,d\xi,d\zeta,d\tau)={}&\mathcal{N}\big(dt,\big(d\xi\times d\zeta\times\mathbb{R}^2\big),[i-1,i),d\tau\big)\\&+\mathcal{N}\big(dt,\big(\mathbb{R}^2\times d\xi\times d\zeta\big),d\tau,[i-1,i)\big).\end{aligned} \tag{13}$$

Clearly, $\mathcal{N}^i$ is a Poisson point measure on $[0,\infty)\times\mathbb{R}^2\times[0,N)$ with intensity

$$dt\,\bar{\Lambda}(d\xi,d\zeta)\frac{d\tau}{N-1}\mathbb{1}_{A^i}(\tau),$$

where $\bar{\Lambda}=\frac{1}{2}(\mathcal{L}(L,R)+\mathcal{L}(\tilde{L},\tilde{R}))$, and $A^i=[0,N)\setminus[i-1,i)$. In other words, $\mathcal{N}^i$ selects only the atoms of $\mathcal{N}$ that produce a jump of X^i, that is, the atoms in which $\mathbf{i}(\rho)=i$ or $\mathbf{i}(\sigma)=i$.

Let us examine the expression (12) in more detail. First, note that since τ is chosen uniformly in A^i, the variable $X_{t-}^{\mathbf{i}(\tau)}$ corresponds to a sample from the (random) probability measure $\bar{\mathbf{X}}_{t-}^i=\frac{1}{N-1}\sum_{j\neq i}\delta_{X_{t-}^j}$. Thus, from the point of view of the process X^i, the dynamics is as follows: at rate 1, a number $v=X_{t-}^{\mathbf{i}(\tau)}$ is sampled from the measure $\bar{\mathbf{X}}_{t-}^i$, and then the value of the process is updated according to the rule $X_{t-}^i\mapsto\xi X_{t-}^i+\zeta v$, where (ξ,ζ) is chosen with law $\bar{\Lambda}$.

Comparing (9) and (12), the key observation is the following: if for each jump time t one replaces $X_{t-}^{\mathbf{i}(\tau)}$ in (12) with a realization v of the law $P_t(dv)$, the resulting process has law P. In view of this, we would like to define the system of nonlinear processes $\mathbf{U}=(U^1,\dots,U^N)$ based on this idea, but using a realization of P_t that is optimally coupled to the realization $X_{t-}^{\mathbf{i}(\tau)}$ of the measure $\bar{\mathbf{X}}_{t-}^i$. In doing this, some measurability issues need to be taken into account.

LEMMA 3 (Coupling). *For every $p\ge 1$ and $i\in\{1,\dots,N\}$ there exist a measurable mapping $\Pi^i:\mathbb{R}_+\times\mathbb{R}^N\times A^i\to\mathbb{R}$, $(t,\mathbf{x},\tau)\mapsto\Pi_t^i(\mathbf{x},\tau)$, with the following property: for every $t\ge 0$ and $\mathbf{x}\in\mathbb{R}^N$, if τ is uniformly chosen from A^i, then the pair $(\Pi_t^i(\mathbf{x},\tau),x^{\mathbf{i}(\tau)})$ is an optimal coupling between P_t and $\bar{\mathbf{x}}^i=\frac{1}{N-1}\sum_{j\neq i}\delta_{x^j}$*

with respect to the cost function $c(u,v) = |u-v|^p$. *Moreover, if* $\mathbf{Y}$ *is any exchangeable random vector in* $\mathbb{R}^N$, *then* $\mathbb{E}\int_{j-1}^{j} \phi(\Pi_t^i(\mathbf{Y},\tau))\,d\tau = \langle P_t, \phi\rangle$ *for any* $j \in \{1,\dots,N\}$, $j \neq i$, *and any bounded measurable function* ϕ.

For simplicity, in our notation we have not made explicit the dependence of Π_t^i on p (however, see Remark 10). Now, we can define U^i as the solution of

$$dU_t^i = \int_{\mathbb{R}^2}\int_{[0,N)} [(\xi - 1)U_{t^-}^i + \zeta\, \Pi_t^i(\mathbf{X}_{t^-},\tau)]\mathcal{N}^i(dt, d\xi, d\zeta, d\tau), \tag{14}$$

where $\mathcal{N}^i$ is the same Poisson point measure as in (12). The proof of Lemma 3 will imply that the mapping $((t,\omega),\xi,\zeta,\tau) \mapsto (\xi-1)U_{t^-}^i(\omega) + \zeta\,\Pi_t^i(\mathbf{X}_{t^-}(\omega),\tau)$ above is measurable with respect to the product of the predictable sigma field [in (t,ω)] and the Borel sigma field of $\mathbb{R}^2 \times [0,N)$. This ensures that the integral in (14) has the usual properties of integrals with respect to Poisson point processes.

We summarize our construction in the following.

LEMMA 4. *Let* $p \geq 1$ *be fixed. For each* $i = 1,\dots,N$ *there is a unique solution* U^i *of* (14), *and it is a nonlinear process. Moreover, the collection* $(X^1,U^1),\dots,(X^N,U^N)$ *is exchangeable.*

Thus, the system $\mathbf{U} = (U^1,\dots,U^N)$ is indeed a tuple of N nonlinear processes. However, as we already mentioned, they are *not independent*, since $\mathcal{N}^i$ and $\mathcal{N}^j$ share a portion of $\mathcal{N}$, namely, the atoms of $\mathcal{N}$ whose coordinates (ρ,σ) lie in $[i-1,i)\times[j-1,j)$ or $[j-1,j)\times[i-1,i)$. In particular, whenever such an atom occurs the processes U^i and U^j jump simultaneously, using a single realization of $(L,R,\tilde{L},\tilde{R})$, and samples of P_t that also are correlated.

3. Proof of the main result.

Before proving our results, let us first state two lemmas that constitute our basic tools; they will be proven in Section 4. The first one provides uniform bounds for the moments of P_t; it can be seen as a version of Theorem 3.2 in Matthes and Toscani (2008).

LEMMA 5 (Moment bounds). *For* $p = 1$ *or* $p = 2$, *assume* $\alpha_p \geq 0$ *and* $M_p(P_0) < \infty$. *If* $p = 2$, *assume also that* $\mathbf{E}(LR + \tilde{L}\tilde{R}) = 0$. *Then for any* $q \in \{p\} \cup (p,q^*)$ *there exists a constant* C, *depending on* q *and some moments of* P_0 *and* $(L,R,\tilde{L},\tilde{R})$ *of order at most* q, *such that*

$$M_q(P_t) \leq Ce^{-\bar{\alpha}_{p,q}t} \qquad \forall t \geq 0.$$

The second lemma is fundamental in our developments since it decouples the nonindependent nonlinear processes uniformly in time, even in the case $\alpha_p = 0$:

LEMMA 6 (Decoupling). *For $p=1$ or $p=2$, assume $\alpha_p \geq 0$ and $M_p(P_0) < \infty$. If $p=2$, assume also that $\mathbf{E}(LR + \tilde{L}\tilde{R}) = 0$. Then there exists a constant C, depending only on the p-moment of P_0 and $(L, R, \tilde{L}, \tilde{R})$, such that for all $k = 2, \dots, N$ and $t \geq 0$,*

$$\mathcal{W}_p^p(\mathcal{L}^k(\mathbf{U}_t), P_t^{\otimes k}) \leq \frac{C(k-1)\min(1,t)e^{-\alpha_p t}}{N-1}.$$

PROOF OF THEOREM 1. Define the constants $\alpha_p^L = \frac{1}{2}\mathbf{E}(|L|^p + |\tilde{L}|^p)$ and $\alpha_p^R = \frac{1}{2}\mathbf{E}(|R|^p + |\tilde{R}|^p)$, so $\alpha_p = 1 - \alpha_p^L - \alpha_p^R$. We first treat the case $p=1$. Thus, we work with the processes U^i solution of (14) using the functions Π_t^i of Lemma 3 with $p=1$. Let us prove (i) first. We estimate the quantity $f_t = \mathbb{E}|X_t^1 - U_t^1|$ which provides an upper bound for $\mathcal{W}_1(\mathcal{L}^1(\mathbf{X}_t), P_t)$. Using (12) and (14), for all $0 \leq s \leq t$ we have

$$\begin{aligned}
&|X_t^1 - U_t^1| - |X_s^1 - U_s^1| \\
&\quad = \int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} \bigl(\bigl|\xi(X_{r-}^1 - U_{r-}^1) + \zeta(X_{r-}^{\mathbf{i}(\tau)} - \Pi_r^1(\mathbf{X}_{r-}, \tau))\bigr| \\
&\qquad\qquad - |X_{r-}^1 - U_{r-}^1|\bigr) \\
&\qquad\qquad \times \mathcal{N}^1(dr, d\xi, d\zeta, d\tau).
\end{aligned} \tag{15}$$

Recall that the intensity of $\mathcal{N}^1$ is $(N-1)^{-1}\, dt\, \bar{\Lambda}(d\xi, d\zeta)\, d\tau \mathbb{1}_{A^1}(\tau)$, where $\bar{\Lambda} = (\mathcal{L}(L,R) + \mathcal{L}(\tilde{L}, \tilde{R}))/2$. By the compensation formula, $t \mapsto f_t$ is absolutely continuous and we obtain

$$\begin{aligned}
f_t - f_s &\leq \mathbb{E}\int_s^t \int_{\mathbb{R}^2}\int_{A^1} \bigl((|\xi| - 1)|X_r^1 - U_r^1| + |\zeta|\,|X_r^{\mathbf{i}(\tau)} - \Pi_r^1(\mathbf{X}_r, \tau)|\bigr) \\
&\qquad \times \frac{d\tau}{N-1}\bar{\Lambda}(d\xi, d\zeta)\, dr \\
&= \mathbb{E}\int_s^t \bigl((\alpha_1^L - 1)|X_r^1 - U_r^1| + \alpha_1^R \mathcal{W}_1(\bar{\mathbf{X}}_r^1, P_r)\bigr)\, dr,
\end{aligned} \tag{16}$$

where in the last step we have used the fact that when τ is uniform in A^1, $(\Pi_s^1(\mathbf{x}, \tau), x^{\mathbf{i}(\tau)})$ is an optimal coupling between P_s and $\bar{\mathbf{x}}^1$. We deduce that for almost all $t \geq 0$

$$\partial_t f_t \leq -(1 - \alpha_1^L) f_t + \alpha_1^R \mathbb{E}\mathcal{W}_1(\bar{\mathbf{X}}_t^1, P_t). \tag{17}$$

Recall that $\bar{\mathbf{U}}_t^i = \frac{1}{N-1}\sum_{j \neq i} \delta_{U_t^j}$ for $i = 1, \dots, N$. The triangle inequality for $\mathcal{W}_1$ gives us

$$\begin{aligned}
\mathbb{E}\mathcal{W}_1(\bar{\mathbf{X}}_t^1, P_t) &\leq \mathbb{E}\mathcal{W}_1(\bar{\mathbf{X}}_t^1, \bar{\mathbf{U}}_t^1) + \mathbb{E}\mathcal{W}_1(\bar{\mathbf{U}}_t^1, P_t) \\
&\leq \mathbb{E}|X_t^1 - U_t^1| + \mathbb{E}\mathcal{W}_1(\bar{\mathbf{U}}_t^1, P_t),
\end{aligned} \tag{18}$$

where the last inequality comes from the fact that $(X_t^{\mathbf{i}(\tau)}, U_t^{\mathbf{i}(\tau)})$ is a coupling between $\bar{\mathbf{X}}_t^1$ and $\bar{\mathbf{U}}_t^1$ when τ is uniformly chosen in A^1, and from the exchangeability of $(X^i, U^i)_{i=1,\dots,N}$. Putting this together with (17), we obtain

$$\partial_t f_t \le -\alpha_1 f_t + \alpha_1^R \mathbb{E}\mathcal{W}_1(\bar{\mathbf{U}}_t^1, P_t). \tag{19}$$

Next, we need an estimate for $\mathbb{E}\mathcal{W}_1(\bar{\mathbf{U}}_t^1, P_t)$. Since the system $(U^2, \dots, U^N)$ is exchangeable, using a recent result [Theorem 1.2 of Hauray and Mischler (2014)], we obtain the following: for each $q > 0$ and each $\gamma < (2 + 1/q)^{-1}$, there exists a constant $C_{q,\gamma}$ such that

$$\mathbb{E}\mathcal{W}_1(\bar{\mathbf{U}}_t^1, P_t) \le C_{q,\gamma} M_q(P_t)^{1/q}\left(\mathcal{W}_1(\mathcal{L}(U_t^2, U_t^3), P_t^{\otimes 2}) + \frac{1}{N-1}\right)^{\gamma}. \tag{20}$$

Now, Lemma 6 in the case $p = 1$ and $k = 2$ implies $\mathcal{W}_1(\mathcal{L}(U_t^1, U_t^2), P_t^{\otimes 2}) \le C/N$, where C is some constant, which can change from line to line in what follows. From this, Lemma 5, and (19)–(20) we have $\partial_t f_t \le -\alpha_1 f_t + CN^{-\gamma} e^{-(1/q)\bar{\alpha}_{1,q} t}$, and then Gronwall's lemma yields

$$f_t \le \frac{C}{N^\gamma}\int_0^t e^{-\alpha_1(t-s)} e^{-(1/q)\bar{\alpha}_{1,q} s}\, ds,$$

since $f_0 = 0$. Bounding $e^{-\alpha_1(t-s)} \le e^{-(1/q)\bar{\alpha}_{1,q}(t-s)}$ gives (i) in the case $p = 1$ and $k = 1$. From this and Lemma 6, case $k \ge 2$ follows:

$$\begin{aligned} \mathcal{W}_1(\mathcal{L}^k(\mathbf{X}_t), P_t^{\otimes k}) &\le \mathcal{W}_1(\mathcal{L}^k(\mathbf{X}_t), \mathcal{L}^k(\mathbf{U}_t)) + \mathcal{W}_1(\mathcal{L}^k(\mathbf{U}_t), P_t^{\otimes k}) \\ &\le \mathbb{E}|X_t^1 - U_t^1| + \frac{Ck\min(1,t)e^{-\alpha_1 t}}{N}. \end{aligned} \tag{21}$$

We now prove (ii): as in (18) we have

$$\begin{aligned} \mathbb{E}\mathcal{W}_1(\bar{\mathbf{X}}_t, P_t) &\le \mathbb{E}|X_t^1 - U_t^1| + \mathbb{E}\mathcal{W}_1(\bar{\mathbf{U}}_t, P_t) \\ &\le \frac{Cte^{-(1/q)\bar{\alpha}_{1,q}t}}{N^\gamma} + \frac{Ce^{-(1/q)\bar{\alpha}_{1,q}t}}{N^\gamma}, \end{aligned}$$

where the last inequality comes from (i) in the case $k = 1$, and from (20) (with $\bar{\mathbf{U}}_t$ and N in place of $\bar{\mathbf{U}}_t^1$ and $N - 1$) together with Lemma 6 in the case $k = 2$. From the previous inequality, (ii) follows; moreover, the same estimate is also valid for $\mathbb{E}\mathcal{W}_1(\bar{\mathbf{X}}_t^1, P_t)$.

Now we treat the case $p = 2$. The proof is similar to the previous case, with adaptations where required. We work with the processes U^i solution of (14) using the functions Π_t^i of Lemma 3 with $p = 2$. As before, to prove the case $k = 1$ we want to estimate $f_t = \mathbb{E}(X_t^1 - U_t^1)^2$. We proceed as in (15): from (12) and (14), we have for all $0 \le s \le t$

$$\begin{aligned}
(X_t^1 - U_t^1)^2 &- (X_s^1 - U_s^1)^2 \\
&= \int_{(s,t]} \int_{\mathbb{R}^2} \int_{[0,N)} \big([\xi (X_{r-}^1 - U_{r-}^1) + \zeta (X_{r-}^{\mathbf{i}(\tau)} - \Pi_r^1(\mathbf{X}_{r-}, \tau))]^2 \\
&\qquad\qquad - [X_{r-}^1 - U_{r-}^1]^2\big) \\
&\qquad \times \mathcal{N}^1(dr, d\xi, d\zeta, d\tau) \\
&= \int_{(s,t]} \int_{\mathbb{R}^2} \int_{[0,N)} \big([\xi^2 - 1](X_{r-}^1 - U_{r-}^1)^2 + \zeta^2 (X_{r-}^{\mathbf{i}(\tau)} - \Pi_r^1(\mathbf{X}_{r-}, \tau))^2 \\
&\qquad\qquad + 2\xi\zeta (X_{r-}^1 - U_{r-}^1)(X_{r-}^{\mathbf{i}(\tau)} - \Pi_r^1(\mathbf{X}_{r-}, \tau))\big) \\
&\qquad \times \mathcal{N}^1(dr, d\xi, d\zeta, d\tau).
\end{aligned} \tag{22}$$

Taking expectations, the last term in the integral vanishes thanks to condition $\mathbf{E}(LR + \tilde{L}\tilde{R}) = 0$. As in (16)–(17), this yields

$$\partial_t f_t \le -(1 - \alpha_2^L) f_t + \alpha_2^R \mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{X}}_t^1, P_t). \tag{23}$$

Defining $g_t = \mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{U}}_t^1, P_t)$ and using the triangle inequality of $\mathcal{W}_2$ we have

$$\begin{aligned}
&\mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{X}}_t^1, P_t) \\
&\quad \le \mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{X}}_t^1, \bar{\mathbf{U}}_t^1) + 2\mathbb{E}\mathcal{W}_2(\bar{\mathbf{X}}_t^1, \bar{\mathbf{U}}_t^1)\mathcal{W}_2(\bar{\mathbf{U}}_t^1, P_t) + \mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{U}}_t^1, P_t) \\
&\quad \le f_t + 2 f_t^{1/2} g_t^{1/2} + g_t,
\end{aligned} \tag{24}$$

where in the last inequality the term f_t is obtained with the same argument as in (18), and the term $f_t^{1/2} g_t^{1/2}$ comes from the Cauchy–Schwarz inequality. From this and (23), we obtain

$$\partial_t f_t \le -\alpha_2 f_t + 2\alpha_2^R f_t^{1/2} g_t^{1/2} + \alpha_2^R g_t.$$

Using a version of Gronwall's lemma [see, e.g., Lemma 4.1.8 of Ambrosio, Gigli and Savaré (2008)] together with Jensen's inequality, we obtain

$$f_t \le \alpha_2^R e^{-\alpha_2 t} (2 + 8\alpha_2^R t) \int_0^t e^{\alpha_2 s} g_s \, ds. \tag{25}$$

Now, we need an estimate for $g_t = \mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{U}}_t^1, P_t)$. Unfortunately, we do not have at our disposal a result similar to (20), which is valid only for $\mathcal{W}_1$. To bypass this, we will make use of the following lemma (proved in Section 4); it has the spirit of (20) in the sense that it will allow us to work with $\mathcal{W}_2^2(\mathcal{L}^n(\mathbf{U}_t), P_t^{\otimes n})$ instead of $\mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{U}}_t^1, P_t)$, but at the price of the extra term $\varepsilon_{n,2}(P_t)$.

LEMMA 7. *Let* $\mathbf{Y} = (Y^1, \dots, Y^m)$ *be an exchangeable random vector, and let* μ *be a probability measure on* $\mathbb{R}$. *Then, for any* $p \geq 1$ *and* $n \leq m$, $n \in \mathbb{N}$, *we have*

$$\frac{1}{2^{p-1}}\mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Y}}, \mu) \leq \frac{kn}{m}\big(\mathcal{W}_p^p(\mathcal{L}^n(\mathbf{Y}), \mu^{\otimes n}) + \varepsilon_{n,p}(\mu)\big) + \frac{\ell}{m}\big(\mathcal{W}_p^p(\mathcal{L}^\ell(\mathbf{Y}), \mu^{\otimes \ell}) + \varepsilon_{\ell,p}(\mu)\big),$$

where k *and* ℓ *are the unique nonnegative integers satisfying* $m = kn + \ell$, *with* $\ell \leq n-1$. *Here,* $\varepsilon_{n,p}(\mu) := \mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Z}}, \mu)$, *where* $\mathbf{Z} = (Z^1, \dots, Z^n)$ *are i.i.d. and* μ *distributed.*

Note that $\mathcal{W}_2^2(\mathcal{L}^\ell(\mathbf{U}_t), P_t^{\otimes \ell}) + \varepsilon_{\ell,2}(P_t) \leq 8M_2(P_t)$. Using this lemma with $p = 2$, $m = N-1$, $\mathbf{Y} = (U_t^2, \dots, U_t^{N-1})$ and $\mu = P_t$, we obtain that for every $n \leq N-1$

$$\begin{aligned}\mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{U}}_t^1, P_t) &\leq \mathcal{W}_2^2(\mathcal{L}^n(\mathbf{U}_t), P_t^{\otimes n}) + \varepsilon_{n,p}(P_t) + \frac{n-1}{N-1}8M_2(P_t) \\ &\leq C\bigg(\frac{ne^{-\alpha_2 t}}{N} + \varepsilon_{n,2}(P_t)\bigg),\end{aligned}$$

where in the last inequality we have used Lemmas 5 and 6 with $p = 2$ and $k = n$; again C is some constant that can change from line to line. Putting this into (25) gives

$$f_t \leq C(1+t)\bigg(\frac{nte^{-\alpha_2 t}}{N} + \int_0^t e^{-\alpha_2(t-s)}\varepsilon_{n,2}(P_s)\,ds\bigg).$$

Given $q \in (2, q^*)$, $q \neq 4$, from Theorem 1 of Fournier and Guillin (2013) we know that $\varepsilon_{n,2}(P_t) \leq CM_q^{2/q}(P_t)n^{-\eta}$, where $\eta = \min(1/2, \frac{q-2}{q})$. Choosing $n = \lfloor N^{1/(1+\eta)} \rfloor$ and using Lemma 5 with $p = 2$ yields

$$f_t \leq C(1+t)\bigg(\frac{te^{-\alpha_2 t}}{N^\gamma} + \frac{1}{N^\gamma}\int_0^t e^{-\alpha_2(t-s)}e^{-(2/q)\bar{\alpha}_{2,q}s}\,ds\bigg),$$

where $\gamma = \eta/(1+\eta) = \min(1/3, \frac{q-2}{2q-2})$. Bounding $e^{-\alpha_2(t-s)} \leq e^{-(2/q)\bar{\alpha}_{2,q}(t-s)}$ gives (i) in the case $p = 2$ and $k = 1$. The case $k \geq 2$ follows as in (21).

Finally, (ii) in the case $p = 2$ follows from (24) with a similar argument as in the case $p = 1$. This completes the proof. $\square$

COROLLARY 8. *Under the same hypotheses and notation of Theorem* 1, *we have for all* $T \geq 0$,

$$\mathbb{E}\sup_{t \in [0,T]}|X_t^1 - U_t^1|^p \leq \frac{C}{N^\gamma}\int_0^T (1+t)^p e^{-(p/q)\bar{\alpha}_{p,q}t}\,dt.$$

PROOF. From (15), discarding the negative term in the integral, we have

$$\begin{aligned}\sup_{t\in[0,T]} |X_t^1 - U_t^1| \\ \leq \int_{(0,T]}\int_{\mathbb{R}^2}\int_{[0,N)} \big(|\xi||X_{t^-}^1 - U_{t^-}^1| + |\zeta||X_{t^-}^{\mathbf{i}(\tau)} - \Pi_t^1(\mathbf{X}_{t^-},\tau)|\big) \\ \times \mathcal{N}^1(dt, d\xi, d\zeta, d\tau).\end{aligned}$$

With the same argument that produced the term $\mathcal{W}_1(\bar{\mathbf{X}}_r^1, P_r)$ in (16), the conclusion follows taking expectations and using the previous estimates for $\mathbb{E}|X_t^1 - U_t^1|$ and $\mathbb{E}\mathcal{W}_1(\bar{\mathbf{X}}_t^1, P_t)$. This proves the case $p = 1$, and the case $p = 2$ follows from (22) with a similar argument. □

REMARK 9. To illustrate how our methods can indeed be used in noncutoff contexts, consider Kac's model: $L = \cos\theta = \tilde{L}$ and $R = -\sin\theta = -\tilde{R}$, where θ is chosen according to an even cross-section function $\beta : [-\pi, \pi] \to \mathbb{R}_+$ that possibly is singular at 0, but satisfies the classical condition $\int_0^\pi \theta^2\beta(\theta)\,d\theta < \infty$, see Desvillettes, Graham and Méléard (1999) for details. Define $\beta_\varepsilon(\theta) = \mathbb{1}_{|\theta|>\varepsilon}\beta(\theta)$ for a given cutoff level $\varepsilon > 0$, and associate with it the collection $(P_t^\varepsilon)_{t\geq 0}$ solving $\partial_t P_t^\varepsilon = \kappa_\varepsilon(-P_t^\varepsilon + \mathcal{Q}_\varepsilon^+(P_t^\varepsilon))$, where $\kappa_\varepsilon = \int_{-\pi}^\pi \beta_\varepsilon(\theta)\,d\theta$ and $\mathcal{Q}_\varepsilon^+$ is defined as

$$\begin{aligned}\int \phi(u)\mathcal{Q}_\varepsilon^+(\mu)(du) \\ = \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{-\pi}^{\pi} \phi(u\cos\theta - v\sin\theta)\frac{\beta_\varepsilon(\theta)\,d\theta}{\kappa_\varepsilon}\mu(dv)\mu(du).\end{aligned}$$

The particle system $\mathbf{X}^\varepsilon$ and nonlinear processes $\mathbf{U}^\varepsilon$ are constructed in a way similar as in (12) and (14) but now using a Poisson measure $\mathcal{N}^{\varepsilon,i}(dt, d\theta, d\tau)$ with intensity $(N-1)^{-1}\,dt\,\beta_\varepsilon(\theta)\,d\theta\,d\tau\mathbb{1}_{A^i}(\tau)$ and functions $\Pi_t^{\varepsilon,i}$ that couple optimally with P_t^ε instead of P_t. Note that:

- The even moments of P_t^ε are controlled uniformly in time and independently of ε [see, e.g., Lemma A.5 in Fournier and Godinho (2012) in the case of the noncutoff nonlinear process; also, an induction similar to the one used in the proof of Lemma 5 yields the desired uniform bounds for $M_q(P_t^\varepsilon)$ when q is even].
- The decoupling property of Lemma 6 is also valid for the system $\mathbf{U}_t^\varepsilon$, with constants independent of ε: in (31), all the terms involve either $1 - L$ or R^2, which correspond to $1 - \cos\theta$ and $\sin^2\theta$, respectively, both of order θ^2.
- In (25), the constant α_2^R corresponds to $\int_{-\pi}^{\pi} \sin^2\theta\beta_\varepsilon(\theta)$.

Thus, the argument can be replicated and the final constant will depend on $\int_0^\pi \theta^2\beta_\varepsilon(\theta)\,d\theta$, which remains bounded as we let the cutoff $\varepsilon \to 0$. Assuming, for instance, that $M_6(P_0) < \infty$, this yields a constant C independent of $\varepsilon > 0$ such that

$$\mathbb{E}\mathcal{W}_2^2(\bar{\mathbf{X}}_t^\varepsilon, P_t^\varepsilon) \leq \frac{C(1+t)^2}{N^{1/3}}.$$

However, we have not been able to obtain a trajectorial result in the noncutoff case: discarding the negative term in the integral of (22) produces the term $\int_0^\pi \cos^2\theta\beta_\varepsilon(\theta)\,d\theta$ which no longer stays bounded when $\varepsilon \to 0$.

4. Proof of intermediate lemmas.

PROOF OF LEMMA 3. For fixed $n \in \mathbb{N}$, given $\mathbf{y} = (y^1, \dots, y^n) \in \mathbb{R}^n$ recall that we write $\bar{\mathbf{y}} = \frac{1}{n}\sum_j \delta_{y^j}$. The mapping $(t, \mathbf{y}) \mapsto (P_t, \bar{\mathbf{y}})$ from $\mathbb{R}_+ \times \mathbb{R}^n$ to $\mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R})$ is continuous when $\mathcal{P}(\mathbb{R})$ is endowed with the weak topology (weak continuity of $t \mapsto P_t$ is clear from the pathwise properties of the nonlinear process). Thus, thanks to a measurable selection result [see, e.g., Corollary 5.22 of Villani (2009)], there exists a measurable mapping $(t, \mathbf{y}) \mapsto \pi_{t,\bar{\mathbf{y}}}$ such that $\pi_{t,\bar{\mathbf{y}}} \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ is an optimal transference plan between P_t and $\bar{\mathbf{y}}$. We now define

$$G(t, \mathbf{y}, B) = \frac{\pi_{t,\bar{\mathbf{y}}}(B \times \{y^1\})}{\pi_{t,\bar{\mathbf{y}}}(\mathbb{R} \times \{y^1\})} = \pi_{t,\bar{\mathbf{y}}}(B \times \{y^1\} | \mathbb{R} \times \{y^1\}),$$

for $t \geq 0$, $\mathbf{y} \in \mathbb{R}^n$ and any Borel set $B \subseteq \mathbb{R}$. We claim that G is a probability kernel from $\mathbb{R}_+ \times \mathbb{R}^n$ into $\mathbb{R}$. Indeed, it suffices to show that for every such B the mapping $(t, \mathbf{y}) \mapsto \pi_{t,\bar{\mathbf{y}}}(B \times \{y^1\})$ is measurable, which in turn follows from the measurability of $(t, \mathbf{y}) \mapsto (\pi_{t,\bar{\mathbf{y}}}, \mathbf{y})$ and the identity

$$\pi_{t,\bar{\mathbf{y}}}(B \times \{y^1\}) = \lim_{\varepsilon \to 0} \sum_{\ell \in \mathbb{N}} \pi_{t,\bar{\mathbf{y}}}(B \times D_\ell^\varepsilon) \mathbb{1}_{D_\ell^\varepsilon}(y^1),$$

where $(D_\ell^\varepsilon)_{\ell \in \mathbb{N}}$ is a measurable partition of $\mathbb{R}$ with $\operatorname{diam}(D_\ell^\varepsilon) \leq \varepsilon$.

Now, given $N \geq 1$, with the kernel G defined above for $n = N - 1$ we can associate a measurable mapping $g : \mathbb{R}_+ \times \mathbb{R}^{N-1} \times [0, 1] \to \mathbb{R}$ or randomization of G such that $g(t, \mathbf{y}, \theta)$ has distribution $G(t, \mathbf{y}, \cdot)$ whenever θ is a uniform random variable in $[0, 1]$ [see, e.g., Lemma 3.22 of Kallenberg (2002)]. For $\mathbf{x} \in \mathbb{R}^N$, we now put

$$\Pi_t^i(\mathbf{x}, \tau) = \sum_{j \neq i}^{N} \mathbb{1}_{\{\mathbf{i}(\tau) = j\}} g(t, \mathbf{x}^{(ij)}, \tau - \lfloor \tau \rfloor), \qquad \tau \in A^i, \tag{26}$$

where $\mathbf{x}^{(ij)} \in \mathbb{R}^{N-1}$ denotes the vector $\mathbf{x}$ with its i coordinate removed, the j coordinate in the first position, and the remaining coordinates in positions $2, \dots, N-1$ in increasing order. We now show that when τ is uniform in A^i, $\Pi_t^i(\mathbf{x}, \tau)$ and $x^{\mathbf{i}(\tau)}$

have joint distribution $\pi_{t,\bar{\mathbf{x}}^i}$. Denoting $\mathbf{P}^i$ the law of this random variable τ and using the fact that $g(t, \mathbf{x}^{(ij)}, \theta)$ has law $\pi_{t,\bar{\mathbf{x}}^i}(du \times \{x^j\}|\mathbb{R} \times \{x^j\})$ when θ is uniform in [0, 1], we have for every fixed measurable set $B \subseteq \mathbb{R}$ and every $j \neq i$:

$$
\begin{aligned}
&\mathbf{P}^i\big(\Pi_t^i(\mathbf{x}, \tau) \in B, x^{\mathbf{i}(\tau)} = x^j\big) \\
&\quad = \sum_{\ell : x^\ell = x^j, \ell \neq i} \int_{\ell-1}^{\ell} \mathbb{1}_B\big(g\big(t, \mathbf{x}^{(i\ell)}, \tau - \lfloor \tau \rfloor\big)\big) \frac{d\tau}{N-1} \\
&\quad = \frac{1}{N-1} \sum_{\ell : x^\ell = x^j, \ell \neq i} \frac{\pi_{t,\bar{\mathbf{x}}^i}(B \times \{x^\ell\})}{\pi_{t,\bar{\mathbf{x}}^i}(\mathbb{R} \times \{x^\ell\})} \\
&\quad = \frac{|\{\ell : x^\ell = x^j, \ell \neq i\}|}{(N-1)\pi_{t,\bar{\mathbf{x}}^i}(\mathbb{R} \times \{x^j\})} \pi_{t,\bar{\mathbf{x}}^i}\big(B \times \{x^j\}\big),
\end{aligned}
$$

where the quotient in the last line equals 1. This shows that $(\Pi_t^i(\mathbf{x}, \tau), x^{\mathbf{i}(\tau)})$ has distribution $\pi_{t,\bar{\mathbf{x}}^i}$ and completes the proof of the existence of Π^i.

It remains to show that $\mathbb{E} \int_{j-1}^{j} \phi(\Pi_t^i(\mathbf{Y}, \tau))\, d\tau = \langle P_t, \phi \rangle$ when $\mathbf{Y}$ is exchangeable, $j \neq i$ and ϕ is bounded and measurable. We get from (26) that

$$
\begin{aligned}
\int_{j-1}^{j} \phi\big(\Pi_t^i(\mathbf{Y}, \tau)\big)\, d\tau &= \int_0^1 \phi\big(g\big(t, \mathbf{Y}^{(ij)}, \tau\big)\big)\, d\tau \\
&= \int_{\mathbb{R}} \phi(u) \pi_{t,\bar{\mathbf{Y}}^i}\big(du \times \{Y^j\}|\mathbb{R} \times \{Y^j\}\big),
\end{aligned}
$$

where we have again used that $g(t, \mathbf{Y}^{(ij)}, \theta)$ has distribution $\pi_{t,\bar{\mathbf{Y}}^i}(du \times \{Y^j\}|\mathbb{R} \times \{Y^j\})$ when θ is uniform in [0, 1]. From the exchangeability of $\mathbf{Y}$, it is clear that the last expression has the same distribution, for all $j \neq i$. Thus, its expected value must be the same for all $j \neq i$, and since

$$
\langle P_t, \phi \rangle = \int_{A^i} \phi\big(\Pi_t^i(\mathbf{Y}, \tau)\big) \frac{d\tau}{N-1} = \sum_{j \neq i} \int_{j-1}^{j} \phi\big(\Pi_t^i(\mathbf{Y}, \tau)\big) \frac{d\tau}{N-1},
$$

the conclusion follows. □

REMARK 10. Since we are working on $\mathbb{R}$, the increasing coupling between P_t and $\bar{\mathbf{x}}^i$ is in fact an optimal coupling [see, e.g., Theorem 6.0.2 in Ambrosio, Gigli and Savaré (2008)], which allows for a simpler proof of Lemma 3. However, we opted to give a proof that remains valid on $\mathbb{R}^d$ with the hope that this coupling can be used in a more general setting.

PROOF OF LEMMA 4. Existence and uniqueness for (14) are obtained with a construction similar to (11). To show that U^i is a nonlinear process, define $\tilde{\mathcal{N}}^i(dt, d\xi, d\zeta, dv)$ to be the point measure on $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}$ with atoms $(t, \xi, \zeta, \Pi_t^i(\mathbf{X}_{t-}, \tau))$ for every atom (t, ξ, ζ, τ) of $\mathcal{N}^i$; since the dependence on $\mathbf{X}$ is predictable, one can use the compensation formula to compute the Laplace functional of $\tilde{\mathcal{N}}^i$ and conclude that $\tilde{\mathcal{N}}^i$ is a Poisson point measure with intensity $dt\,\bar{\Lambda}(d\xi, d\zeta)P_t(dv)$. Then (9) is satisfied for $V = U^i$ with $\mathcal{M} = \tilde{\mathcal{N}}^i$, implying that $\mathcal{L}(U^i) = P$. The collection $(X^1, U^1), \dots, (X^N, U^N)$ is obviously exchangeable. □

PROOF OF LEMMA 5. Call $h_t^q = \int |u|^q P_t(du)$. We first prove the statement for the case $p = 2$. Using (1)–(2) with $\phi = |\cdot|^2$ yields $\partial_t h_t^2 = -\alpha_2 h_t^2 + \mathbf{E}(LR + \tilde{L}\tilde{R})(h_t^1)^2$, and since $\mathbf{E}(LR + \tilde{L}\tilde{R}) = 0$ this implies $h_t^2 = h_0^2 e^{-\alpha_2 t}$. Assume now that $q \in (2, q^*)$ is an integer. Using (1)–(2) with $\phi = |\cdot|^q$, we have

$$\begin{aligned}\partial_t h_t^q &= -h_t^q + \frac{1}{2}\iint \mathbf{E}\big(|Lu + Rv|^q + |\tilde{L}v + \tilde{R}u|^q\big)P_t(du)P_t(dv)\\ &\le -\alpha_q h_t^q + \frac{1}{2}\sum_{i=1}^{q-1}\binom{q}{i} h_t^i h_t^{q-i}\mathbf{E}\big(|L|^i|R|^{q-i} + |\tilde{L}|^i|\tilde{R}|^{q-i}\big).\end{aligned} \tag{27}$$

Using loose bounds for $\binom{q}{i}$, we obtain

$$h_t^q \le h_0^q e^{-\alpha_q t} + C\sum_{i=1}^{q-1}\int_0^t e^{-\alpha_q(t-s)} h_s^i h_s^{q-i}\,ds,$$

where C is a constant that does not depend on t, and may change from line to line. We now apply induction: the case $q = 2$ was already proven, and for $q \in (2, q^*)$ integer, assuming the desired property for all integer in $\{2, \dots, q-1\}$ and using the bound $h_t^1 \le (h_t^2)^{1/2} \le Ce^{-\alpha_2 t/2}$, we obtain

$$\begin{aligned}h_t^q \le h_0^q e^{-\alpha_q t} &+ C\int_0^t e^{-\alpha_q(t-s)}e^{-\alpha_2 s/2}e^{-\bar{\alpha}_{2,q-1}s}\,ds\\ &+ C\sum_{i=2}^{q-2}\int_0^t e^{-\alpha_q(t-s)}e^{-\bar{\alpha}_{2,i}s}e^{-\bar{\alpha}_{2,q-i}s}\,ds.\end{aligned}$$

Note that $\alpha_q > 0$, since $2 < q < q^*$, and recall that $\bar{\alpha}_{2,q} := \inf_{2\le r\le q}\alpha_r = \min(\alpha_2, \alpha_q)$. Thus, if $\alpha_2 = 0$ then $\bar{\alpha}_{2,i} = \bar{\alpha}_{2,q-i} = \bar{\alpha}_{2,q} = 0$ and the last inequality yields $h_t^q \le h_0^q + C\int_0^t e^{-\alpha_q(t-s)}\,ds \le C$, as desired. On the other hand, if $\alpha_2 > 0$, we bound α_2, α_q, $\bar{\alpha}_{2,i}$ and $\bar{\alpha}_{2,q-i}$ from below by $\bar{\alpha}_{2,q} > 0$ and obtain $h_t^q \le h_0^q e^{-\bar{\alpha}_{2,q}t} + C\int_0^t e^{-\bar{\alpha}_{2,q}(t+s/2)}\,ds \le Ce^{-\bar{\alpha}_{2,q}t}$, which completes the induction and the proof in the case $p = 2$ and integer $q \in \{2\} \cup (2, q^*)$.

Assume now that $2 < q = m + \varepsilon < q^*$ with $m \in \{2, \dots\}$ and $\varepsilon \in (0, 1)$. Bounding $|x + y|^q \le (|x| + |y|)^m(|x|^\varepsilon + |y|^\varepsilon)$ in (27) and using the binomial theorem as before, we obtain

$$\begin{aligned}\partial_t h_t^q &= -h_t^q + \frac{1}{2}\sum_{i=0}^{m}\binom{m}{i}\iint \mathbf{E}\big((|Lu|^\varepsilon + |Rv|^\varepsilon)|Lu|^i|Rv|^{m-i} \\ &\qquad\qquad + (|\tilde{L}v|^\varepsilon + |\tilde{R}u|^\varepsilon)|\tilde{L}v|^i|\tilde{R}u|^{m-i}\big)P_t(du)P_t(dv) \\ &= -\alpha_q h_t^q + \sum_{i=0}^{m-1}\binom{m}{i}\mathbf{E}\big(|L|^{i+\varepsilon}|R|^{m-i} + |L|^{m-i}|R|^{i+\varepsilon} \\ &\qquad\qquad + |\tilde{L}|^{i+\varepsilon}|\tilde{R}|^{m-i} + |\tilde{L}|^{m-i}|\tilde{R}|^{i+\varepsilon}\big)h_t^{i+\varepsilon}h_t^{m-i},\end{aligned}$$

which yields

$$h_t^q \le h_0^q e^{-\alpha_q t} + C\sum_{i=0}^{m-1}\int_0^t e^{-\alpha_q(t-s)}h_s^{i+\varepsilon}h_s^{m-i}\,ds.$$

Note that $h_t^r \le (h_t^2)^{r/2} \le Ce^{-r\alpha_2 t/2}$ for $r \in (0,2)$. This and the fact that the property is true for the integers, allow us to use induction on m in a way similar as before, and complete the proof in the case $p = 2$.

A similar argument, with the induction starting at $q = 1$, proves the case $p = 1$. $\square$

PROOF OF LEMMA 6. Let us first prove the case $p = 1$. Given $k \in \{2, \dots, N\}$ fixed, we want to construct k independent nonlinear processes $V^1, \dots, V^k$ such that $\mathbb{E}|U_t^i - V_t^i|$ is small. To achieve this, we will decouple $U^1, \dots, U^k$ by replacing the shared atoms of $\mathcal{N}^1, \dots, \mathcal{N}^k$ with new, independent atoms. To this end, let $\mathcal{M}$ be an independent copy of $\mathcal{N}$ (also independent from $\mathbf{X}_0$), and define for each $i \in \{1, \dots, k\}$

$$\begin{aligned}&\mathcal{M}^i(dt, d\xi, d\zeta, d\tau) \\ &\quad= \mathcal{N}\big(dt, (d\xi \times d\zeta \times \mathbb{R}^2), [i-1, i), d\tau\big) \\ &\qquad + \mathcal{N}\big(dt, (\mathbb{R}^2 \times d\xi \times d\zeta), d\tau, [i-1, i)\big)\mathbb{1}_{[k,N)}(\tau) \\ &\qquad + \mathcal{M}\big(dt, (\mathbb{R}^2 \times d\xi \times d\zeta), d\tau, [i-1, i)\big)\mathbb{1}_{[0,k)}(\tau).\end{aligned} \tag{28}$$

Note that $\mathcal{M}^i$ is, like $\mathcal{N}^i$, a Poisson point measure on $\mathbb{R}_+ \times \mathbb{R}^2 \times [0, N)$ with intensity $(N-1)^{-1}\,dt\,\bar{\Lambda}(\xi,\zeta)\,d\tau\mathbb{1}_{A^i}(\tau)$, and that $\mathcal{M}^1, \dots, \mathcal{M}^k$ are independent. Following (14), we define V^i as the solution of

$$dV_t^i = \int_{\mathbb{R}^2}\int_{[0,N)}\big[(\xi - 1)V_{t-}^i + \zeta\,\Pi_t^i(\mathbf{X}_{t-}, \tau)\big]\mathcal{M}^i(dt, d\xi, d\zeta, d\tau), \tag{29}$$

with $V_0^i = U_0^i$. If we define $\tilde{\mathcal{M}}^i$ to be the point process in $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}$ with atoms $(t, \xi, \zeta, \Pi_t^i(\mathbf{X}_{t-}, \tau))$ for every atom (t, ξ, ζ, τ) of $\mathcal{M}^i$, it is clear that V^i

depends only on $\tilde{\mathcal{M}}^i$ and X_0^i. Since: (i) the dependence on $\mathbf{X}$ is predictable, (ii) the Poisson measures $\mathcal{M}^1, \dots, \mathcal{M}^k$ are independent and (iii) the τ-law of $\Pi_t^i(\mathbf{x}, \tau)$ is P_t for every $\mathbf{x} \in \mathbb{R}^N$, one can use the compensation formula to compute the joint Laplace functional of $\tilde{\mathcal{M}}^1, \dots, \tilde{\mathcal{M}}^k$ and conclude that they are independent Poisson point measures, all with intensity $dt\,\bar{\Lambda}(d\xi, d\zeta)P_t(dv)$. This shows that each V^i is a nonlinear process and that they are independent.

Consequently, we have

$$\mathcal{W}_1(\mathcal{L}^k(\mathbf{U}_t), P_t^{\otimes k}) \le \mathbb{E}\left(\frac{1}{k}\sum_{i=1}^k |U_t^i - V_t^i|\right) = \mathbb{E}|U_t^1 - V_t^1|,$$

where in the last step we used the fact that all the (U^i, V^i)'s have the same law. To estimate the last term $h_t = \mathbb{E}|U_t^1 - V_t^1|$, we proceed as in (15): from (13), (14), (28) and (29), we have for all $0 \le s \le t$:

$$h_t = h_s + \mathbb{E}\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} (J_r^1 + J_r^2 + J_r^3), \tag{30}$$

where J_r^1 is the term associated with the simultaneous jumps of U^1 and V^1, J_r^2 corresponds to the jumps of U^1 alone, and J_r^3 gives the jumps of V^1 alone. Specifically,

$$\begin{aligned}
J_r^1 &= (|\xi| - 1)|U_{r-}^1 - V_{r-}^1|\big(\mathcal{N}(dr, (d\xi \times d\zeta \times \mathbb{R}^2), [0,1), d\tau) \\
&\qquad + \mathcal{N}(dr, (\mathbb{R}^2 \times d\xi \times d\zeta), d\tau, [0,1))\mathbb{1}_{[k,N)}(\tau)\big), \\
J_r^2 &= \big(|\xi(U_{r-}^1 - V_{r-}^1) + \zeta\,\Pi_r^1(\mathbf{X}_{r-}, \tau) + (\xi - 1)V_{r-}^1| - |U_{r-}^1 - V_{r-}^1|\big) \\
&\quad \times \mathcal{N}(dr, (\mathbb{R}^2 \times d\xi \times d\zeta), d\tau, [0,1))\mathbb{1}_{[0,k)}(\tau), \\
J_r^3 &= \big(|\xi(U_{r-}^1 - V_{r-}^1) - \zeta\,\Pi_r^1(\mathbf{X}_{r-}, \tau) - (\xi - 1)U_{r-}^1| - |U_{r-}^1 - V_{r-}^1|\big) \\
&\quad \times \mathcal{M}(dr, (\mathbb{R}^2 \times d\xi \times d\zeta), d\tau, [0,1))\mathbb{1}_{[0,k)}(\tau).
\end{aligned}$$

Then

$$\mathbb{E}\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} J_r^1 = \left(\frac{1}{2}(\mathbf{E}|L| - 1) + \frac{1}{2}(\mathbf{E}|\tilde{L}| - 1)\frac{N-k}{N-1}\right)\int_s^t h_r\,dr.$$

Using the triangle inequality in the term J_r^2,

$$\begin{aligned}
&\mathbb{E}\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} J_r^2 \\
&\quad \le \mathbb{E}\int_s^t\int_1^k \big((\mathbf{E}|\tilde{L}| - 1)|U_r^1 - V_r^1| + \mathbf{E}|\tilde{R}||\Pi_r^1(\mathbf{X}_r, \tau)| + \mathbf{E}|\tilde{L} - 1||V_r^1|\big) \\
&\qquad\qquad \times \frac{dr\,d\tau}{2(N-1)}.
\end{aligned}$$

From Lemma 3, we know that $\mathbb{E}\int_{i-1}^{i}|\Pi_r^1(\mathbf{X}_r,\tau)|\,d\tau = M_1(P_r)$ for all $i=2,\dots,k$. Using that V_r^1 has law P_r, we obtain

$$\mathbb{E}\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} J_r^2 \\ \le \frac{k-1}{2(N-1)}\biggl(\bigl(\mathbf{E}|\tilde{L}|-1\bigr)\int_s^t h_r\,dr + \bigl(\mathbf{E}|\tilde{L}-1|+\mathbf{E}|\tilde{R}|\bigr)\int_s^t M_1(P_r)\,dr\biggr).$$

With a similar argument, the last inequality is also valid with J_r^3 in the left-hand side. Putting all this into (30), we have

$$h_t \le h_s - \biggl(1-\frac{1}{2}\mathbf{E}(|L|+|\tilde{L}|)+\frac{1}{2}(1-\mathbf{E}|\tilde{L}|)\frac{k-1}{N-1}\biggr)\int_s^t h_r\,dr \\ + \frac{(\mathbf{E}|\tilde{L}-1|+\mathbf{E}|\tilde{R}|)(k-1)}{N-1}\int_s^t M_1(P_r)\,dr.$$

Recall the constants $\alpha_1^L=\frac{1}{2}\mathbf{E}(|L|+|\tilde{L}|)$, $\alpha_1^R=\frac{1}{2}\mathbf{E}(|R|+|\tilde{R}|)$ and $\alpha_1 = 1-\alpha_1^L-\alpha_1^R$. Also, put $b=\frac{1}{2}(1-\mathbf{E}|\tilde{L}|)$, which can be assumed nonnegative without loss of generality [if not, exchange the roles of (L,R) and $(\tilde{L},\tilde{R})$]. From the previous inequality and from Lemma 5 in the case $q=1$, it follows that for almost all $t\ge 0$,

$$\partial_t h_t \le -\biggl(\alpha_1+\alpha_1^R+b\frac{k-1}{N-1}\biggr)h_t + \frac{C(k-1)e^{-\alpha_1 t}}{N-1},$$

and now Gronwall's lemma gives

$$h_t \le \frac{C(k-1)e^{-\alpha_1 t}}{(N-1)(\alpha_1^R+b((k-1)/(N-1)))}\bigl[1-e^{-(\alpha_1^R+b((k-1)/(N-1)))t}\bigr].$$

Using the inequality $1-e^{-x}\le x$, the desired result follows for the case $p=1$.

In the case $p=2$, we construct the system $V^1,\dots,V^k$ exactly as before, but using the functions Π_t^i provided by Lemma 3 with cost $|x-y|^2$. To obtain the desired inequality for $\mathcal{W}_2^2(\mathcal{L}^k(\mathbf{U}_t),P_t^{\otimes k})$, it suffices to work with $h_t=\mathbb{E}(U_t^1-V_t^1)^2$. We also have (30), where J_r^1, J_r^2 and J_r^3 now are given by

$$J_r^1 = (\xi^2-1)(U_{r-}^1-V_{r-}^1)^2\bigl(\mathcal{N}(dr,(d\xi\times d\zeta\times\mathbb{R}^2),[0,1),d\tau) \\ + \mathcal{N}(dr,(\mathbb{R}^2\times d\xi\times d\zeta),d\tau,[0,1))\mathbb{1}_{[k,N)}(\tau)\bigr),$$

$$J_r^2 = \bigl((\xi(U_{r-}^1-V_{r-}^1)+\zeta\Pi_r^1(\mathbf{X}_{r-},\tau)+(\xi-1)V_{r-}^1)^2-(U_{r-}^1-V_{r-}^1)^2\bigr) \\ \times\mathcal{N}(dr,(\mathbb{R}^2\times d\xi\times d\zeta),d\tau,[0,1))\mathbb{1}_{[0,k)}(\tau),$$

$$J_r^3 = \bigl((\xi(U_{r-}^1-V_{r-}^1)-\zeta\Pi_r^1(\mathbf{X}_{r-},\tau)-(\xi-1)U_{r-}^1)^2-(U_{r-}^1-V_{r-}^1)^2\bigr) \\ \times\mathcal{M}(dr,(\mathbb{R}^2\times d\xi\times d\zeta),d\tau,[0,1))\mathbb{1}_{[0,k)}(\tau).$$

Using that $\mathbb{E}\int_{i-1}^{i}\Pi_t^1(\mathbf{X}_t,\tau)^2\,d\tau = M_2(P_t)$ for all $i=2,\dots,k$, we obtain

$$\mathbb{E}\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} J_r^1 = \biggl(\frac{1}{2}(\mathbf{E}L^2-1)+\frac{1}{2}(\mathbf{E}\tilde{L}^2-1)\frac{N-k}{N-1}\biggr)\int_s^t h_r\,dr,$$

$$\begin{aligned}
\mathbb{E}&\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} J_r^2\\
&=\int_s^t\biggl((\mathbf{E}\tilde{L}^2-1)h_r+\mathbf{E}\tilde{R}^2M_2(P_r)+\mathbf{E}(\tilde{L}-1)^2M_2(P_r)\\
&\qquad+2\mathbf{E}(\tilde{L}\tilde{R})\mathbb{E}(U_r^1-V_r^1)\int_1^k\Pi_r^1(\mathbf{X}_r,\tau)\frac{d\tau}{k-1}\\
&\qquad+2\mathbf{E}(\tilde{L}(\tilde{L}-1))\mathbb{E}(U_r^1-V_r^1)V_r^1\\
&\qquad\qquad+2\mathbf{E}((\tilde{L}-1)\tilde{R})\mathbb{E}V_r^1\int_1^k\Pi_r^1(\mathbf{X}_r,\tau)\frac{d\tau}{k-1}\biggr)\frac{(k-1)\,dr}{2(N-1)},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}&\int_{(s,t]}\int_{\mathbb{R}^2}\int_{[0,N)} J_r^3\\
&=\int_s^t\biggl((\mathbf{E}\tilde{L}^2-1)h_r+\mathbf{E}\tilde{R}^2M_2(P_r)+\mathbf{E}(\tilde{L}-1)^2M_2(P_r)\\
&\qquad-2\mathbf{E}(\tilde{L}\tilde{R})\mathbb{E}(U_r^1-V_r^1)\int_1^k\Pi_r^1(\mathbf{X}_r,\tau)\frac{d\tau}{k-1}\\
&\qquad-2\mathbf{E}(\tilde{L}(\tilde{L}-1))\mathbb{E}(U_r^1-V_r^1)U_r^1\\
&\qquad\qquad+2\mathbf{E}((\tilde{L}-1)\tilde{R})\mathbb{E}U_r^1\int_1^k\Pi_r^1(\mathbf{X}_r,\tau)\frac{d\tau}{k-1}\biggr)\frac{(k-1)\,dr}{2(N-1)}.
\end{aligned}$$

From this and (30), we have for almost all $t\geq 0$

$$\begin{aligned}
\partial_t h_t &= -h_t\biggl(\biggl(1-\frac{1}{2}\mathbf{E}(L^2+\tilde{L}^2)\biggr)+\frac{k-1}{2(N-1)}\mathbf{E}(\tilde{L}-1)^2\biggr)\\
&\quad+M_2(P_t)\frac{k-1}{N-1}(\mathbf{E}(\tilde{L}-1)^2+\mathbf{E}\tilde{R}^2)\\
&\quad+\mathbf{E}((\tilde{L}-1)\tilde{R})\mathbb{E}(U_t^1+V_t^1)\int_1^k\Pi_t^1(\mathbf{X}_t,\tau)\frac{d\tau}{N-1}.
\end{aligned}\tag{31}$$

We also have $\int_{i-1}^{i}\mathbb{E}(U_t^1+V_t^1)\Pi_t^1(\mathbf{X}_t,\tau)\,d\tau\leq 2M_2(P_t)$ for all $i=2,\dots,k$, thanks to the Cauchy–Schwarz and Jensen inequalities. Recall the constants $\alpha_2^L=\frac{1}{2}\mathbf{E}(L^2+\tilde{L}^2)$, $\alpha_2^R=\frac{1}{2}\mathbf{E}(R^2+\tilde{R}^2)$, $\alpha_2=1-\alpha_2^L-\alpha_2^R$, and put $b=\frac{1}{2}\mathbf{E}(\tilde{L}-1)^2$. Using Lemma 5 in the case $p=q=2$, we thus obtain

$$\partial_t h_t\leq-\biggl(\alpha_2+\alpha_2^R+b\frac{k-1}{N-1}\biggr)h_t+\frac{C(k-1)M_2(P_0)e^{-\alpha_2 t}}{N-1},$$

and the conclusion follows from Gronwall's lemma as before. $\square$

PROOF OF LEMMA 7. For simplicity, we will prove only the case $\ell = 0$, that is, when n divides m. Let us arrange a vector $\mathbf{y} \in \mathbb{R}^m$ as a matrix with k rows and n columns, that is, $\mathbf{y} = (y^{ij})$, with $i = 1, \dots, k$, $j = 1, \dots, n$, and write $\mathbf{y}_i = (y^{i1}, \dots, y^{in})$ and $\bar{\mathbf{y}}_i = \frac{1}{n} \sum_{j=1}^n \delta_{y^{ij}}$. Let us couple $\mathbf{Y}$ with a random vector $\mathbf{Z} \in (\mathbb{R}^n)^k$ in such a way that each $(\mathbf{Y}_i, \mathbf{Z}_i)$ is an optimal coupling between $\mathcal{L}^n(\mathbf{Y})$ and $\mu^{\otimes n}$ [with respect to the cost function $d_{n,p}^p(\cdot, \cdot)$ of (8), as usual]. Using the latter, we have

$$\mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Y}}, \bar{\mathbf{Z}}) \le \frac{1}{k} \sum_{i=1}^k \mathbb{E} \frac{1}{n} \sum_{j=1}^n |Y^{ij} - Z^{ij}|^p = \mathcal{W}_p^p(\nu^n, \mu^{\otimes n}). \tag{32}$$

On the other hand, for each $i = 1, \dots, k$ there is a function $q^i : \mathbb{R}^m \times [0, 1] \to \mathbb{R}$ such that for all $\mathbf{z} \in \mathbb{R}^m$, the pair $(z^{i\mathbf{i}^n(\theta)}, q^i(\mathbf{z}, \theta))$ with $\mathbf{i}^n(\theta) = \mathbf{i}(n\theta) = \lfloor n\theta \rfloor + 1$, is an optimal coupling between $\bar{\mathbf{z}}_i$ and μ when θ is uniformly chosen in $[0, 1]$. Now we randomize the choice of i with a uniform variable $\vartheta \in [0, 1]$ independent of θ, so $z^{\mathbf{i}^k(\vartheta)\mathbf{i}^n(\theta)}$ and $q^{\mathbf{i}^k(\vartheta)}(\mathbf{z}, \theta)$ are (θ, ϑ)-realizations of $\bar{\mathbf{z}}$ and μ, respectively. Putting $\mathbf{Z}$ in place of $\mathbf{z}$, this construction gives

$$\begin{aligned}\mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Z}}, \mu) &\le \mathbb{E} \iint_{[0,1]^2} \left| Z^{\mathbf{i}^k(\vartheta)\mathbf{i}^n(\theta)} - q^{\mathbf{i}^k(\vartheta)}(\mathbf{Z}, \theta) \right|^p d\vartheta\, d\theta \\ &= \mathbb{E} \frac{1}{k} \sum_{i=1}^k \mathcal{W}_p^p(\bar{\mathbf{Z}}_i, \mu) = \varepsilon_{n,p}(\mu).\end{aligned}$$

[Recall that $\varepsilon_{n,p}(\mu) = \mathbb{E}\mathcal{W}_p^p(\frac{1}{n} \sum_i \delta_{\zeta^i}, \mu)$, with $\zeta^1, \dots, \zeta^n$ independent and μ-distributed]. With this and (32), we conclude in the case $\ell = 0$

$$\mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Y}}, \mu) \le \mathbb{E}\big(\mathcal{W}_p(\bar{\mathbf{Y}}, \bar{\mathbf{Z}}) + \mathcal{W}_p(\bar{\mathbf{Z}}, \mu)\big)^p \le 2^{p-1}\big(\mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Y}}, \bar{\mathbf{Z}}) + \mathbb{E}\mathcal{W}_p^p(\bar{\mathbf{Z}}, \mu)\big).$$

In the case $\ell > 0$, the construction is similar, but now $(\mathbf{Y}, \mathbf{Z})$ must include an additional optimal coupling between $\mathcal{L}^\ell(\mathbf{Y})$ and $\mu^{\otimes \ell}$, which gives the extra term. □

Acknowledgements. We thank two anonymous referees for carefully reading a former version of this work and for their questions and remarks that allowed us to improve its presentation.

REFERENCES

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. Birkhäuser, Basel. MR2401600

BASSETTI, F., LADELLI, L. and MATTHES, D. (2011). Central limit theorem for a class of one-dimensional kinetic equations. *Probab. Theory Related Fields* **150** 77–109. MR2800905

CARRAPATOSO, K. (2014a). Quantitative and qualitative Kac's chaos on the Boltzmann's sphere. Preprint. Available at arXiv:1205.1241.

CARRAPATOSO, K. (2014b). Propagation of chaos for the spatially homogeneous Landau equation for Maxwellian molecules. Preprint. Available at arXiv:1212.3724.

DESVILLETTES, L., GRAHAM, C. and MÉLÉARD, S. (1999). Probabilistic interpretation and numerical approximation of a Kac equation without cutoff. *Stochastic Process. Appl.* **84** 115–135. MR1720101

FONTBONA, J., GUÉRIN, H. and MÉLÉARD, S. (2009). Measurability of optimal transportation and convergence rate for Landau type interacting particle systems. *Probab. Theory Related Fields* **143** 329–351. MR2475665

FOURNIER, N. and GODINHO, D. (2012). Asymptotic of grazing collisions and particle approximation for the Kac equation without cutoff. *Comm. Math. Phys.* **316** 307–344. MR2993918

FOURNIER, N. and GUILLIN, A. (2013). On the rate of convergence in Wasserstein distance of the empirical measure. Preprint. Available at arXiv:1312.2128v1.

FOURNIER, N. and MISCHLER, S. (2014). Rate of convergence of the Nanbu particle system for hard potentials. Preprint. Available at arXiv:1302.5810.

FUNAKI, T. (1984). A certain class of diffusion processes associated with nonlinear parabolic equations. *Z. Wahrsch. Verw. Gebiete* **67** 331–348. MR0762085

GRAHAM, C. and MÉLÉARD, S. (1997). Stochastic particle approximations for generalized Boltzmann models and convergence estimates. *Ann. Probab.* **25** 115–132. MR1428502

GUÉRIN, H. (2003). Solving Landau equation for some soft potentials through a probabilistic approach. *Ann. Appl. Probab.* **13** 515–539. MR1970275

HAURAY, M. and MISCHLER, S. (2014). On Kac's chaos and related problems. *J. Funct. Anal.* **266** 6055–6157. MR3188710

KAC, M. (1956). Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. III* 171–197. Univ. California Press, Berkeley and Los Angeles. MR0084985

KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

MATTHES, D. and TOSCANI, G. (2008). On steady distributions of kinetic models of conservative economies. *J. Stat. Phys.* **130** 1087–1117. MR2379241

MISCHLER, S. and MOUHOT, C. (2013). Kac's program in kinetic theory. *Invent. Math.* **193** 1–147. MR3069113

PULVIRENTI, A. and TOSCANI, G. (2004). Asymptotic properties of the inelastic Kac model. *J. Stat. Phys.* **114** 1453–1480. MR2039485

RACHEV, S. T. and RÜSCHENDORF, L. (1998). *Mass Transportation Problems*: *Theory. Vol. I.* Springer, New York. MR1619170

SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185

TANAKA, H. (1978). On the uniqueness of Markov process associated with the Boltzmann equation of Maxwellian molecules. In *Proceedings of the International Symposium on Stochastic Differential Equations* (*Res. Inst. Math. Sci., Kyoto Univ., Kyoto*, 1976) 409–425. Wiley, New York. MR0536022

TANAKA, H. (1978/1979). Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete* **46** 67–105. MR0512334

VILLANI, C. (2009). *Optimal Transport. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

DEPARTMENT OF MATHEMATICAL ENGINEERING
AND CENTER FOR MATHEMATICAL MODELING
UMI(2807) UCHILE-CNRS
UNIVERSIDAD DE CHILE
CASILLA 170-3, CORREO 3
SANTIAGO-CHILE
E-MAIL: rcortez@dim.uchile.cl
fontbona@dim.uchile.cl