The Annals of Applied Probability
2016, Vol. 26, No. 2, 1029–1081
DOI: 10.1214/15-AAP1111

A CONSISTENCY ESTIMATE FOR KAC'S MODEL OF ELASTIC COLLISIONS IN A DILUTE GAS

BY JAMES NORRIS[1]

University of Cambridge

An explicit estimate is derived for Kac's mean-field model of colliding hard spheres, which compares, in a Wasserstein distance, the empirical velocity distributions for two versions of the model based on different numbers of particles. For suitable initial data, with high probability, the two processes agree to within a tolerance of order $N^{-1/d}$, where N is the smaller particle number and d is the dimension, provided that $d \geq 3$. From this estimate we can deduce that the spatially homogeneous Boltzmann equation is well posed in a class of measure-valued processes and provides a good approximation to the Kac process when the number of particles is large. We also prove in an appendix a basic lemma on the total variation of time-integrals of time-dependent signed measures.

1. Kac process. Kac [8] proposed in 1954 a random process to model the dynamics of a dilute gas. The process models the velocities of N particles in $\mathbb{R}^d$ as they evolve under elastic collisions. The case $d = 3$ is of main interest, but we will allow any $d \geq 2$. Since no account is taken of particle positions, any physical justification for the model relies on assumptions of spatial homogeneity and rapid mixing. It is thus impossible to give a physical meaning to the number of particles N. Yet, on the mathematical side, we have to make a choice. Hence it is of interest to show consistency for sufficiently large values of N.

Kac's process depends on a choice of collision kernel B. This is a finite measurable kernel $B(v, d\sigma)$ on $\mathbb{R}^d \times S^{d-1}$ which is chosen to model physical characteristics of the gas. The collision kernel specifies the rate for collisions of pairs of particles with incoming relative velocity v and outgoing direction of separation σ. Since collisions are assumed to conserve momentum and energy, for a pair of particles with pre-collision velocities v and v_*, and hence relative velocity $v - v_*$, the post-collision velocities $v' = v'(v, v_*, \sigma)$ and $v'_* = v'_*(v, v_*, \sigma)$ are determined by the direction of separation through

$$v' + v'_* = v + v_*, \qquad v' - v'_* = |v - v_*|\sigma.$$

We will often write u for the direction of approach, given by $u = (v - v_*)/|v - v_*|$. We assume throughout that, for all $u \in S^{d-1}$, $B(u, \cdot)$ is a probability measure,

Received May 2014; revised March 2015.
[1]Supported by EPSRC Grant EP/103372X/1.
MSC2010 subject classifications. Primary 60J25; secondary 35Q20.
Key words and phrases. Kac process, law of large numbers, Wasserstein distance, Boltzmann equation.

supported on $S^{d-1} \setminus \{-u, u\}$, and that the following standard scaling and symmetry properties hold. For $\lambda \in [0, \infty)$ and $u \in S^{d-1}$, and for any isometry R of S^{d-1}, we have

$$B(\lambda u, \cdot) = \lambda B(u, \cdot), \qquad B(Ru, \cdot) = B(u, \cdot) \circ R^{-1}. \tag{1}$$

Our main results require further that the map $u \mapsto B(u, \cdot)$ is Lipschitz on S^{d-1} for the total variation norm on measures on S^{d-1}. Then there is a constant $\kappa \in [1, \infty)$ such that, for all $v, v' \in \mathbb{R}^d$,

$$\|B(v, \cdot) - B(v', \cdot)\| \le \kappa |v - v'|. \tag{2}$$

Here and throughout, we denote the total variation norm by $\|\cdot\|$. The Boltzmann sphere $\mathcal{S}$ is the set of probability measures μ on $\mathbb{R}^d$ such that[2]

$$\langle v, \mu\rangle = \int_{\mathbb{R}^d} v\mu(dv) = 0, \qquad \langle |v|^2, \mu\rangle = \int_{\mathbb{R}^d} |v|^2 \mu(dv) = 1.$$

For $N \in \mathbb{N}$, write $\mathcal{S}_N$ for the subset of $\mathcal{S}$ of normalized empirical measures of the form $N^{-1} \sum_{i=1}^N \delta_{v_i}$. The Kac process with collision kernel B and particle number N is the Markov chain in $\mathcal{S}_N$ with generator $\mathcal{G}$ given on bounded measurable functions F by

$$\mathcal{G}F(\mu) = N \int_{\mathbb{R}^d \times \mathbb{R}^d \times S^{d-1}} \{F(\mu^{v, v_*, \sigma}) - F(\mu)\} \mu(dv)\mu(dv_*) B(v - v_*, d\sigma),$$

where

$$\mu^{v, v_*, \sigma} = \mu + N^{-1}\{\delta_{v'} + \delta_{v'_*} - \delta_v - \delta_{v_*}\}.$$

The choice of state-space $\mathcal{S}_N$ is possible because in each collision the number of particles, the momentum $v + v_*$ and the energy $|v|^2 + |v_*|^2$ are conserved. There is no Kac process on $\mathcal{S}_1$ because this set is empty. For $N \ge 2$, the transition rates of the Kac process are bounded by $2N$ on $\mathcal{S}_N$. Hence, by the elementary theory of Markov chains, given any initial state $\mu_0^N \in \mathcal{S}_N$, there exists a Kac process $(\mu_t^N)_{t \ge 0}$ in $\mathcal{S}_N$ starting from μ_0^N, the law of this process is unique, and almost surely it takes only finitely many values in any compact time interval.

It is of special interest to model particles colliding as hard spheres. Under plausible physical assumptions, this leads, by a well-known calculation, to the choice of kernel $B(v, d\sigma) \propto |v| \sin^{3-d}(\theta/2)\, d\sigma$, where $\theta \in [0, \pi]$ is given by $\cos\theta = u \cdot \sigma$ and $d\sigma$ is the uniform distribution on S^{d-1}. It is straightforward to check that (1) and (2) hold in this case for all $d \ge 2$. Indeed, for $d = 3$, we can take $\kappa = 1$, and the dynamics have a particularly simple description: for every pair of particles (v, v_*), at rate $|v - v_*|/N$, consider the sphere with poles at v and v_*; choose randomly a new axis for the sphere, label the poles v' and v'_* and replace v and v_* by v' and v'_*.

[2]Here, on the left-hand side, and where convenient below, we use v to denote the identity function on $\mathbb{R}^d$.

Consider the set $\mathcal{F}$ of functions f on $\mathbb{R}^d$ such that

$$|\hat{f}(v)| \le 1, \qquad |\hat{f}(v) - \hat{f}(v')| \le |v - v'|$$

for all v, v', where[3]

$$\hat{f}(v) = f(v)/(1 + |v|^2).$$

Define a distance function W on $\mathcal{S}$ by

$$W(\mu, \nu) = \sup_{f \in \mathcal{F}} \langle f, \mu - \nu \rangle.$$

Then W makes $\mathcal{S}$ into a complete separable metric space. This is shown in Section 9, along with the convergence of a natural approximation scheme by random samples in $(\mathcal{S}, W)$. Our first main result is the following consistency estimate for Kac processes with different numbers of particles. We make no assumption on the joint law of the processes. They could, for example, be independent.

THEOREM 1.1. *Assume that the collision kernel B satisfies conditions* (1) *and* (2). *Let $\varepsilon \in (0, 1]$, $\lambda \in [1, \infty)$, $p \in (2, \infty)$ and $T \in [0, \infty)$. Then there exist constants $\alpha(d, p) > 0$ and $C(B, d, \varepsilon, \lambda, p, T) < \infty$ with the following property. Let $N, N' \in \mathbb{N}$ with $N \le N'$, and let $(\mu_t^N)_{t \ge 0}$ and $(\mu_t^{N'})_{t \ge 0}$ be Kac processes in $\mathcal{S}_N$ and $\mathcal{S}_{N'}$ such that*

$$\langle |v|^p, \mu_0^N \rangle \le \lambda, \qquad \langle |v|^p, \mu_0^{N'} \rangle \le \lambda. \tag{3}$$

Then, with probability exceeding $1 - \varepsilon$, for all $t \in [0, T]$, we have

$$W(\mu_t^N, \mu_t^{N'}) \le C\big(W(\mu_0^N, \mu_0^{N'}) + N^{-\alpha}\big).$$

We have not found a way to prove a similar estimate for $p = 2$. This is consistent with the current theory for the spatially homogeneous Boltzmann equation where also, for $p = 2$, there is no quantitative stability estimate. We can improve the rate of convergence at the cost of a stronger moment condition.

THEOREM 1.2. *Assume further that $(\mu_t^N)_{t \ge 0}$ and $(\mu_t^{N'})_{t \ge 0}$ are adapted as Markov processes to a common filtration. For $p > 8$ and $d \ge 3$, we can take $\alpha = 1/d$ in Theorem 1.1. Also, for $p > 8$ and $d = 2$, we can replace $N^{-\alpha}$ in Theorem 1.1 by $N^{-1/2} \log N$.*

The theorems could be considered as providing a measure of accuracy for a Monte Carlo scheme, using say N computational particles, for the evolution of a Kac process having a much larger number of particles N'.

[3]The notation is chosen as a reminder of the shape of the weight function $1/(1 + |v|^2)$.

The rates of convergence in Theorem 1.2 are known to be optimal for the convergence of sample empirical distributions in Wasserstein distance. Indeed, there is no discrete approximation scheme for a smooth measure which achieves a rate better than $N^{-1/d}$. So it seems unlikely that the rates in can be improved in this context. Our need for the condition $p > 8$ can be traced to the stochastic convolution estimates in Section 7. We show in Section 9 that, for laws in $\mathcal{S}$ having a finite pth moment, their sample empirical distributions converge in the metric W with optimal rates if $p > 3d/(d-1)$, but this can fail if $p < 3d/(d-1)$. This makes it plausible that some moment condition beyond $p > 2$ is necessary for the conclusions of Theorem 1.2, but we do not know whether this is so.

By combining Theorem 1.2 with Proposition 3.1 below, we obtain the following estimate.

THEOREM 1.3. *For $d \geq 3$, for all $\varepsilon \in (0, 1]$ and all $\tau, T \in (0, \infty)$ with $\tau \leq T$, there is a constant $C(B, d, \varepsilon, \tau, T) < \infty$ such that, for all $N, N' \in \mathbb{N}$ with $N \leq N'$ and any Kac processes $(\mu_t^N)_{t\geq 0}$ in $\mathcal{S}_N$ and $(\mu_t^{N'})_{t\geq 0}$ in $\mathcal{S}_{N'}$, with probability exceeding $1-\varepsilon$, for all $t \in [\tau, T]$, we have*

$$W(\mu_t^N, \mu_t^{N'}) \leq C\big(W(\mu_\tau^N, \mu_\tau^{N'}) + N^{-1/d}\big).$$

Note that τ can be arbitrarily small, and we obtain here the optimal rate $N^{-1/d}$ without the supplementary moment condition (3). Thus it is only for the initial evolution of the processes that consistency may rely on a such a moment condition.

We have avoided so far any mention of the Boltzmann equation, which classically is the starting point for kinetic theory. We shall show in our other main results, Theorem 10.1 and Corollaries 10.2 and 10.3, that the consistency estimate leads quickly to existence and uniqueness of measure solutions for the spatially homogeneous Boltzmann equation, and convergence to such solutions of the Kac process in the large N limit. Indeed, we obtain a more precise estimate of this convergence than was previously known. This was the original motivation for our work.

In the next two sections, we identify martingales of the Kac process, and we derive some moment estimates. The difference of two Kac processes, with the same collision kernel but different numbers of particles, satisfies a noisy version of a linearized Boltzmann equation. In Section 4 we develop a representation formula for solutions of this equation in terms of an auxiliary branching process, which we call the linearized Kac process. We use coupling arguments for this process to develop some estimates. The proof of Theorem 1.1 is given in Section 5.

We develop in Section 6 some further continuity estimates for the linearized Kac process, and in Section 7 some maximal inequalities for stochastic convolutions appearing in the representation formula. These are then used in Section 8 to prove Theorem 1.2. The relation of our estimates to prior work on the Kac process and the spatially homogeneous Boltzmann equation is discussed in Section 10. The final section is a self-contained appendix, proving a basic result on the evolution of signed measures, which is used in Sections 4 and 10.

2. Martingales of the Kac process. We compute the martingale decomposition for linear functions of the Kac process $(\mu_t^N)_{t\geq 0}$. Set $E=\mathbb{R}^d\times\mathbb{R}^d\times S^{d-1}\times(0,\infty)$. Denote by m the un-normalized empirical measure on E of the set of all random vectors (V,V_*,Σ,T) such that there is a collision at time T in the particle system $(\mu_t^N)_{t\geq 0}$ of a velocity pair (V,V_*) with direction of separation Σ. Denote by $\bar{m}$ the random measure on E given by

$$\bar{m}(dv,dv_*,d\sigma,dt)=N\mu_{t-}^N(dv)\mu_{t-}^N(dv_*)B(v-v_*,d\sigma)\,dt.$$

Define a random signed measure M^N on $(0,\infty)\times\mathbb{R}^d$ by specifying, for bounded measurable functions f on $(0,\infty)\times\mathbb{R}^d$, the integral[4]

$$\begin{aligned}M_t^{N,f}&=\int_0^t\langle f_s,dM_s^N\rangle=\int_{(0,t]\times\mathbb{R}^d}f(s,v)M^N(ds,dv)\\&=\frac{1}{N}\int_E\{f_s(v')+f_s(v_*')-f_s(v)-f_s(v_*)\}\\&\qquad\times 1_{(0,t]}(s)(m-\bar{m})(dv,dv_*,d\sigma,ds).\end{aligned}\tag{4}$$

Then, by standard results for Markov chains, the process $(M_t^{N,f})_{t\geq 0}$ is a martingale. We use the same notation also in the case where f has no dependence on the time parameter. Define for finite measures μ,ν on $\mathbb{R}^d$ a signed measure $Q(\mu,\nu)$ on $\mathbb{R}^d$ by specifying, for bounded measurable functions f of compact support in $\mathbb{R}^d$, the integral

$$\begin{aligned}\langle f,Q(\mu,\nu)\rangle=\int_{\mathbb{R}^d\times\mathbb{R}^d\times S^{d-1}}&\{f(v')+f(v_*')-f(v)-f(v_*)\}\\&\times\mu(dv)\nu(dv_*)B(v-v_*,d\sigma).\end{aligned}\tag{5}$$

Then the martingale decomposition for $(\langle f,\mu_t^N\rangle)_{t\geq 0}$ is given by

$$\langle f,\mu_t^N\rangle=\langle f,\mu_0^N\rangle+M_t^{N,f}+\int_0^t\langle f,Q(\mu_s^N,\mu_s^N)\rangle ds.\tag{6}$$

We note for later use the following estimates. First, by Doob's L^2-inequality,[5]

$$\begin{aligned}&\mathbb{E}\Big(\sup_{s\leq t}|M_s^{N,f}|^2\Big)\\&\quad\leq\frac{4}{N^2}\mathbb{E}\int_E\{f_s(v')+f_s(v_*')-f_s(v)-f_s(v_*)\}^2 1_{(0,t]}(s)\bar{m}(dv,dv_*,d\sigma,ds)\\&\quad\leq 128\|f\|_\infty t/N.\end{aligned}\tag{7}$$

[4]We will sometimes write f_s for $f(s,\cdot)$.

[5]For the same calculation in a general setting, see, for example, [2], Proposition 8.7.

Next, for the total variation measure $|M^N|$ of M^N, we have

$$
\begin{aligned}
&\mathbb{E}\int_{(0,t]\times\mathbb{R}^d}(1+|v|^2)\big|M^N(ds,dv)\big| \\
&\quad\le \mathbb{E}\int_E (4+2|v|^2+2|v_*|^2)1_{(0,t]}(s)(m+\bar{m})(dv,dv_*,d\sigma,ds) \\
&\quad= \mathbb{E}\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}(8+4|v|^2+4|v_*|^2)|v-v_*|\mu_s^N(dv)\mu_s^N(dv_*)\,ds \\
&\quad\le 24\mathbb{E}\int_0^t\langle 1+|v|^3,\mu_s^N\rangle\,ds.
\end{aligned}
\tag{8}
$$

We used $|v-v_*|\le|v|+|v_*|$ and the fact that $\mu_s^N\in\mathcal{S}$ for the second inequality. Finally, for any interval $(s,s']$ during which $(\mu_t^N)_{t\ge0}$ does not jump, there is no contribution to the left-hand side of (8) from m, so the same calculation yields the following pathwise estimate:

$$
\int_{(s,s']\times\mathbb{R}^d}(1+|v|^2)\big|M^N(dr,dv)\big|\le 12\int_s^{s'}\langle 1+|v|^3,\mu_r^N\rangle\,dr.
\tag{9}
$$

3. Moment estimates for the Kac process. We derive some moment inequalities for the Kac process, which we shall use later. The basic arguments are standard for the Boltzmann equation and are applied to the Kac process in [13], Lemma 5.4. We have quantified the moment-improving property and added some maximal inequalities. We begin with the Povzner inequality. For all $p\in(2,\infty)$, there is a constant $\beta(B,p)>0$ such that, for all $v,v_*\in\mathbb{R}^d$ and for $u=(v-v_*)/|v-v_*|$,

$$
\begin{aligned}
&\int_{S^{d-1}}\{|v'|^p+|v'_*|^p-|v|^p-|v_*|^p\}B(u,d\sigma) \\
&\qquad\le -\beta(|v|^p+|v_*|^p)+\beta^{-1}(|v||v_*|^{p-1}+|v|^{p-1}|v_*|).
\end{aligned}
\tag{10}
$$

Here is a proof for the class of collision kernels we consider. Note first that

$$
\begin{aligned}
|v'|^p+|v'_*|^p &\le (|v|^2+|v_*|^2)^{p/2} \\
&\le |v|^p+|v_*|^p+C(p)(|v||v_*|^{p-1}+|v|^{p-1}|v_*|).
\end{aligned}
\tag{11}
$$

It suffices by symmetry to consider the case $|v_*|\le|v|$. Set $y=|v-v_*|(u+\sigma)/2$, then $v'=v_*+y$ and $|y|^2=|v-v_*|^2t$, where $t=(1+u\cdot\sigma)/2$. Note that $t\in(0,1)$ for $B(u,\cdot)$-almost all σ. We use the inequalities $|v'|\le|y|+|v_*|$ and $|v-v_*|\le|v|+|v_*|$ to see that, for all $\delta\in(0,1]$,

$$
|v'|^2\le(1+\delta)|y|^2+(1+\delta^{-1})|v_*|^2\le(1+\delta)^2t|v|^2+2(1+\delta^{-1})|v_*|^2.
$$

From this inequality and a similar one for $|v'_*|^2$, we deduce that, for some $C(p) < \infty$,

$$|v'|^p \le (1+\delta)^{p+1} t^{p/2} |v|^p + C(p)\delta^{-1}|v_*|^p,$$
$$|v'_*|^p \le (1+\delta)^{p+1}(1-t)^{p/2}|v|^p + C(p)\delta^{-1}|v_*|^p.$$

Then

$$\begin{aligned} &|v'|^p + |v'_*|^p - |v|^p - |v_*|^p \\ &\quad \le -\beta(\delta,t)\big(|v|^p + |v_*|^p\big) + C(p)\delta^{-1}\big(|v||v_*|^{p-1} + |v|^{p-1}|v_*|\big), \end{aligned} \tag{12}$$

where $\beta(\delta,t) = (1-(1+\delta)^{p+1}(t^{p/2}+(1-t)^{p/2}))^+/2$. Set $\beta(\delta) = (\delta/C(p)) \wedge \int_{S^{d-1}} \beta(\delta,t) B(u,d\sigma)$. Then we obtain (10) for u with $\beta = \beta(\delta)$ by integrating (12). But $\beta(\delta)$ does not depend on u by the symmetry condition (1) and $\beta(\delta) > 0$ for all sufficiently small δ, so we are done.

PROPOSITION 3.1. *Let $(\mu_t^N)_{t\ge 0}$ be a Kac process with collision kernel B satisfying* (1). *Let $p \in [2,\infty)$ and $q \in (2,\infty)$ with $p \le q$. There exists a constant $C(B,p,q) < \infty$ such that, for all $t \ge 0$, we have*

$$\mathbb{E}\big(\langle |v|^q, \mu_t^N\rangle\big) \le C\big(1+t^{p-q}\big)\langle |v|^p, \mu_0^N\rangle. \tag{13}$$

Moreover, there is a constant $C(B,q) < \infty$ such that, for all $t \ge 0$,

$$\mathbb{E}\Big(\sup_{s\le t}\langle |v|^q, \mu_s^N\rangle\Big) \le (1+Ct)\langle |v|^q, \mu_0^N\rangle, \tag{14}$$

and there is a constant $C(B,p,q) < \infty$ such that, for all $t \ge 0$,

$$\mathbb{E}\Big(\sup_{s\le t}\langle 1+|v|^p, |\mu_s^N - \mu_0^N|\rangle\Big) \le C\big(t+t^{q-p}\big)\langle |v|^q, \mu_0^N\rangle. \tag{15}$$

PROOF. By the Povzner inequality, there are constants $\beta(B,q) > 0$ and $C(B,q) < \infty$ such that, for all $v, v_* \in \mathbb{R}^d$,

$$\begin{aligned} &\int_{S^{d-1}} \{|v'|^q + |v'_*|^q - |v|^q - |v_*|^q\} B(v-v_*, d\sigma) \\ &\quad \le -\beta|v-v_*|\big(|v|^q + |v_*|^q\big) + \beta^{-1}|v-v_*|\big(|v||v_*|^{q-1} + |v|^{q-1}|v_*|\big) \\ &\quad \le -\beta\big(|v|^{q+1} + |v_*|^{q+1}\big) + C\big(|v|^q\big(1+|v_*|\big) + \big(1+|v|\big)|v_*|^q\big). \end{aligned}$$

Set $f_q(t) = \mathbb{E}(\langle |v|^q, \mu_t^N\rangle)$ and $f_{q,p}(t) = f_q(t)/f_p^*$, where $f_p^* = \sup_{t\ge 0} f_p(t)$. Since $\langle |v|^q, \mu\rangle \le N^{q/2}$ for all $\mu \in \mathcal{S}$, we have $f_q(t) \le N^{q/2} < \infty$ for all t. The process $(\langle |v|^q, \mu_t^N\rangle)_{t\ge 0}$ makes jumps of size $\{|v'|^q + |v'_*|^q - |v|^q - |v_*|^q\}/N$ at rate $N\mu_{t-}^N(dv)\mu_{t-}^N(dv_*)B(v-v_*,d\sigma)\,dt$. Hence

$$\begin{aligned}
f_q(t) &= f_q(0) \\
&\quad + \mathbb{E}\int_0^t\int\{|v'|^q+|v_*'|^q-|v|^q-|v_*|^q\}\mu_s^N(dv)\mu_s^N(dv_*)B(v-v_*,d\sigma)\,ds \\
&\le f_q(0)-2\beta\int_0^t f_{q+1}(s)\,ds+2C\int_0^t f_q(s)\,ds.
\end{aligned}$$

By Hölder's inequality, we have $f_q(t)^{q-p+1}\le f_{q+1}(t)^{q-p}f_p(t)$, so we deduce that

$$f_{q,p}(t)\le f_{q,p}(0)-2\beta\int_0^t\big(f_{q,p}(s)\big)^{1+1/(q-p)}\,ds+2C\int_0^t f_{q,p}(s)\,ds$$

which implies by standard arguments that, for some $C(B,p,q)<\infty$ and all $t\ge 0$, we have

$$f_{q,p}(t)\le C\big(1+f_{q,p}(0)\wedge t^{p-q}\big). \tag{16}$$

Now $f_2^*=1$, so by taking $p=2$, we obtain (13), for the cases $p=2$ and $p=q$. In particular, this shows that $f_p^*\le C\langle|v|^p,\mu_0^N\rangle$ for all p, so (16) implies (13) also for $p\in(2,q)$.

Consider the process $(A_t)_{t\ge0}$ starting from 0 which jumps by $\{|v||v_*|^{q-1}+|v|^{q-1}|v_*|\}/N$ when $(\langle|v|^q,\mu_t^N\rangle)_{t\ge0}$ jumps by $\{|v'|^q+|v_*'|^q-|v|^q-|v_*|^q\}/N$. Then

$$\begin{aligned}
\mathbb{E}(A_t) &= \mathbb{E}\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}\{|v||v_*|^{q-1}+|v|^{q-1}|v_*|\}|v-v_*|\mu_s^N(dv)\mu_s^N(dv_*)\,ds \\
&\le 4\int_0^t f_q(s)\,ds.
\end{aligned}$$

Now $\sup_{s\le t}\langle|v|^q,\mu_s^N\rangle\le\langle|v|^q,\mu_0^N\rangle+C(q)A_t$ for all t, where $C(q)$ is the constant from (11). Hence we obtain (14) by taking expectations and using the case $p=q$ of (13) to estimate f_q.

The process $(\langle1+|v|^p,|\mu_t^N-\mu_0^N|\rangle)_{t\ge0}$ jumps by at most $\{4+|v'|^p+|v_*'|^p+|v|^p+|v_*|^p\}/N$ at each jump of $(\mu_t^N)_{t\ge0}$. Consider the process $(B_t)_{t\ge0}$ starting from 0 which jumps by $\{1+|v|^p+|v_*|^p\}/N$ at the same times. Then $\langle1+|v|^p,|\mu_s^N-\mu_0^N|\rangle\le 2^pB_t$ whenever $s\le t$ and

$$\begin{aligned}
\mathbb{E}(B_t) &= \mathbb{E}\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}\{1+|v|^p+|v_*|^p\}|v-v_*|\mu_s^N(dv)\mu_s^N(dv_*)\,ds \\
&\le 6\int_0^t f_{p+1}(s)\,ds.
\end{aligned}$$

So (15) follows from (13). $\square$

4. Linearized Kac process and representation formula. In this section we introduce a branching process of signed particles in $\mathbb{R}^d$ which may be considered as a linearization of the Kac process. A particular case of this process allows us to write a representation formula for the difference of two Kac processes $(\mu_t^N)_{t\geq 0}$ and $(\mu_t^{N'})_{t\geq 0}$. We use coupling arguments to obtain continuity estimates for the branching process, which are later used to control $\mu_t^N - \mu_t^{N'}$. The representation formula rests only on the fact that $(\mu_t^N - \mu_t^{N'})_{t\geq 0}$ solves the linear equation (25) below. It seems possible that the same conclusions can be reached by a direct analysis of this equation, but we have not done this.

The branching process will have "positive" and "negative" particles, making the following general notation convenient. Given a set V, we denote by V^* the signed space $V \times \{-1, 1\} = V^- \cup V^+$, by π the projection $V^* \to V$ and by $\pi_\pm$ the bijections $V^\pm \to V$. Note that $*$ does not signify the dual space. From now on, we set $V = \mathbb{R}^d$.

The data for our branching process are an initial time $s \in [0, \infty)$ and an initial type $v \in V^*$, together with a process $(\rho_t)_{t\geq 0}$ of measures on $\mathbb{R}^d$ such that, for all t,

$$\langle 1, \rho_t\rangle \leq 1, \qquad \langle |v|^2, \rho_t\rangle \leq 1. \tag{17}$$

The case $\rho_t = (\mu_t^N + \mu_t^{N'})/2$ will be of main interest later. Consider the continuous-time branching particle system[6] with types in V^* where each particle of type v in $V^\pm$, at rate $2\rho_t(dv_*)B(v - v_*, d\sigma)\,dt$ for $v_* \in \mathbb{R}^d$ and $\sigma \in S^{d-1}$, dies and is replaced by three particles $v'(v, v_*, \sigma)$ and $v'_*(v, v_*, \sigma)$ in $V^\pm$ and v_* in $V^\mp$. More properly, the rate is $2\rho_t(dv_*)B(\pi(v) - v_*, d\sigma)\,dt$ and the offspring are $(v'(\pi(v), v_*, \sigma), 1)$, $(v'_*(\pi(v), v_*, \sigma), 1)$ and $(v_*, -1)$ when $v \in V^+$, and $(v'(\pi(v), v_*, \sigma), -1)$, $(v'_*(\pi(v), v_*, \sigma), -1)$ and $(v_*, 1)$ when $v \in V^-$. We assume throughout that, for all $t \geq 0$,

$$\int_0^t \langle |v|^3, \rho_s\rangle\, ds < \infty. \tag{18}$$

We will show that (18) ensures there is no explosion; that is, the time T_n of the nth branching event tends to ∞ almost surely. So the process is well defined for all time by the specification of its branching rates, and consists at all times $t \geq s$

[6]The dynamics of the branching process can be motivated as follows. Fix a large integer N, and suppose that $(N\rho_t)_{t\geq 0}$ evolves as an unnormalized Kac process on N particles. Consider the perturbed process obtained by introducing one additional particle of velocity v at time s, where the pairwise collision rules are unchanged and where transitions are coupled as far as possible with the original. The discrepancy between the original and the perturbed systems will grow over time approximately as the branching process $(\Lambda_t^*)_{t\geq s}$, a "negative" particle in V^- corresponding to one present in the original system but removed by collision in the perturbed system. Formally, the approximation becomes exact as $N \to \infty$. We do not rely on this. The construction of $(\Lambda_t^*)_{t\geq s}$ does not require $(\rho_t)_{t\geq 0}$ to be a Kac process.

of a finite number of particles. Write $(\Lambda_t^*)_{t\ge s}$ for the associated process of unnormalized empirical measures on V^*. We call this process the *linearized Kac process in environment* $(\rho_t)_{t\ge 0}$ *starting from* v *at time* s.

Set $\Lambda_t = \Lambda_t^* \circ \pi^{-1}$. Then $(\Lambda_t)_{t\ge s}$ is itself the empirical process of a branching process in V, in which we forget the book-keeping exercise of giving a sign to each particle. Write $E_{(s,v)}$ for the expectation over $(\Lambda_t^*)_{t\ge s}$ to recall that $\Lambda_s^* = \delta_v$ and that this is not the full expectation in the case that $(\rho_t)_{t\ge 0}$ is itself random. Given an initial type $v \in \mathbb{R}^d$, without a sign, we will by default start the process $(\Lambda_t^*)_{t\ge s}$ with the positive type $(v, 1)$.

PROPOSITION 4.1. *There is almost surely no explosion in the branching construction described above. Moreover, for all* $p \in [2,\infty)$, *there is a constant* $c(p) < \infty$ *such that, for all* $v_0 \in \mathbb{R}^d$ *and all* $t \ge s$, *we have*

$$E_{(s,v_0)}\langle 1+|v|^p, \Lambda_t\rangle \le \left(1+|v_0|^p\right)\exp\left\{c(p)\int_s^t \langle 1+|v|^{p+1}, \rho_r\rangle\, dr\right\}.$$

In particular we can take $c(2) = 8$.

We will reserve the notation $c(p)$ for this constant throughout. We will also use throughout the notation

$$\tilde{\Lambda}_t = \Lambda_t^+ - \Lambda_t^-, \qquad \Lambda_t^\pm = \Lambda_t^* \circ \pi_\pm^{-1}.$$

Thus Λ_t^+ and Λ_t^- are random measures on $\mathbb{R}^d$, which are the empirical distributions of positive and negative particles, and $\tilde{\Lambda}_t$ is a random signed measure on $\mathbb{R}^d$. Note that $\Lambda_t = \Lambda_t^+ + \Lambda_t^-$. By Proposition 4.1, we can define, for any $s, t \ge 0$ with $s \le t$, a linear map E_{st} on the set of measurable functions of quadratic growth on $\mathbb{R}^d$ by

$$E_{st} f(v) = E_{(s,v)}\langle f, \tilde{\Lambda}_t\rangle.$$

Note that, by the Markov property, we have $E_{st}E_{tu} = E_{su}$. We will write f_{st} for $E_{st} f$ and sometimes just f_s when the value of t is understood. We will use the same notation for functions f of polynomial growth, whenever $(\rho_t)_{t\ge 0}$ has sufficient moments for this to make sense using Proposition 4.1. We base our main argument on the following representation formula, which is proved at the end of this section.

PROPOSITION 4.2. *In the case where* $\rho_t = (\mu_t^N + \mu_t^{N'})/2$ *for all* t, *we have*

$$\langle f, \mu_t^N - \mu_t^{N'}\rangle = \langle f_{0t}, \mu_0^N - \mu_0^{N'}\rangle + \int_0^t \langle f_{st}, dM_s^N\rangle - \int_0^t \langle f_{st}, dM_s^{N'}\rangle.$$

We will use the following two estimates expressing continuity of the linearized Kac process in its initial data. Write $\|f\|$ for the smallest constant such that $|\hat{f}(v)| \le \|f\|$ and $|\hat{f}(v) - \hat{f}(v')| \le \|f\||v - v'|$ for all $v, v' \in \mathbb{R}^d$. Thus $f \in \mathcal{F}$ if and only if $\|f\| \le 1$.

PROPOSITION 4.3. *Assume condition* (2). *Then*

$$\|E_{st}f\| \le 3\bigl(1+6\kappa(t-s)\bigr)\exp\left\{\int_s^t 8\langle 1+|v|^3,\rho_r\rangle\,dr\right\}\|f\|.$$

PROPOSITION 4.4. *For all* $v \in \mathbb{R}^d$ *and all* $s, s' \in [0,t]$ *with* $s \le s'$, *we have*

$$\begin{aligned}&\bigl|E_{st}f(v) - E_{s't}f(v)\bigr|\\&\quad\le 5\bigl(1+|v|^3\bigr)\exp\left\{\int_s^t 8\langle 1+|v|^3,\rho_r\rangle\,dr\right\}\|f\|\int_s^{s'}\langle 1+|v|^3,\rho_r\rangle\,dr.\end{aligned}$$

PROOF OF PROPOSITION 4.1. Consider first the case $p=2$ and $s=0$. Fix $v_0 \in \mathbb{R}^d$, and consider the branching particle system $(\Lambda_t)_{t<\zeta}$ starting from δ_{v_0} at time 0 and run up to explosion $\zeta = \sup_n T_n$. Note that, at a branching event with colliding particle velocity v_*, the total number of particles in the system increases by 2, and the total kinetic energy increases by $|v'|^2 + |v'_*|^2 + |v_*|^2 - |v|^2 = 2|v_*|^2$. Hence $\langle 1+|v|^2, \Lambda_t\rangle$ makes jumps of size $2(1+|v_*|^2)$ at rate $2|v-v_*|\Lambda_{t-}(dv)\rho_t(dv_*)\,dt$. Set $S_n = \inf\{t<\zeta : \langle 1+|v|^2,\Lambda_t\rangle \ge n\}$, and set

$$g(t) = E_{(0,v_0)}\langle 1+|v|^2, \Lambda_{t\wedge S_n}\rangle.$$

Note that $S_n \le T_n$. We use the estimate

$$\bigl(1+|v_*|^2\bigr)|v-v_*| \le 2\bigl(1+|v|^2\bigr)\bigl(1+|v_*|^3\bigr)$$

to see that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\bigl(1+|v_*|^2\bigr)|v-v_*|\Lambda_t(dv)\rho_t(dv_*) \le 2m_3(t)\langle 1+|v|^2,\Lambda_t\rangle,$$

where $m_3(t) = \langle 1+|v|^3, \rho_t\rangle$. Hence, by optional stopping, the process

$$\langle 1+|v|^2,\Lambda_{t\wedge S_n}\rangle - \int_0^{t\wedge S_n} 8m_3(s)\langle 1+|v|^2,\Lambda_s\rangle\,ds$$

is a supermartingale. On taking expectations, we obtain

$$\begin{aligned}g(t) &\le 1+|v_0|^2 + E_{(0,v_0)}\int_0^{t\wedge S_n} 8m_3(s)\langle 1+|v|^2,\Lambda_s\rangle\,ds\\&\le 1+|v_0|^2 + \int_0^t 8m_3(s)g(s)\,ds\end{aligned}$$

so $g(t) < \infty$ and then

$$g(t) \le \bigl(1+|v_0|^2\bigr)\exp\left\{\int_0^t 8m_3(s)\,ds\right\}.$$

The right-hand side does not depend on n, so we must have $S_n \to \infty$ almost surely as $n \to \infty$. Hence $T_n \to \infty$ almost surely, and the claimed estimate follows by monotone convergence.

For $p \in (2, \infty)$, there is a constant $C(p) < \infty$ such that

$$
\begin{aligned}
|v'|^p + |v'_*|^p + |v_*|^p - |v|^p &\le (|v|^2 + |v_*|^2)^{p/2} + |v_*|^p - |v|^p \\
&\le C(p)(|v|^{p-2}|v_*|^2 + |v_*|^p)
\end{aligned}
\tag{19}
$$

and then, for another constant $c(p) < \infty$,

$$
2|v - v_*|(2 + |v'|^p + |v'_*|^p + |v_*|^p - |v|^p) \le c(p)(1 + |v|^p)(1 + |v_*|^{p+1}). \tag{20}
$$

The argument used for $p = 2$ then gives the desired estimate in the case $s = 0$. The argument is the same for $s \ge 0$. $\square$

We now describe a coupling of linearized Kac processes starting from different initial velocities, constructed to branch at the same times and with the same sampled velocities v_* and angles σ, as far as possible. To simplify, we begin without the signs. Define sets

$$
V_0 = \mathbb{R}^d \times \mathbb{R}^d, \qquad V_1 = \mathbb{R}^d \times \{1\}, \qquad V_2 = \mathbb{R}^d \times \{2\} \tag{21}
$$

which we treat as disjoint. Consider the continuous-time branching process in $V_0 \cup V_1 \cup V_2$ with the following branching mechanism. For each particle (of type) $(v_1, v_2) \in V_0$, there are three possible transitions. First, at rate $2B(v_1 - v_*, d\sigma) \wedge B(v_2 - v_*, d\sigma)\rho_t(dv_*)\,dt$ for $v_* \in \mathbb{R}^d$ and $\sigma \in S^{d-1}$, the particle (v_1, v_2) dies and is replaced by three particles (v_*, v_*), (v'_1, v'_2) and (v'_{1*}, v'_{2*}) in V_0. Here we are writing v'_k for $v'(v_k, v_*, \sigma)$ and v'_{k*} for $v'_*(v_k, v_*, \sigma)$ for short. Call this a coupled transition. Second, at rate $2(B(v_1 - v_*, d\sigma) - B(v_2 - v_*, d\sigma))^+\rho_t(dv_*)\,dt$, the particle (v_1, v_2) dies and is replaced by four particles v_*, v'_1 and v'_{1*} in V_1 and v_2 in V_2. Third, at rate $2(B(v_1 - v_*, d\sigma) - B(v_2 - v_*, d\sigma))^-\rho_t(dv_*)\,dt$, the particle (v_1, v_2) dies and is replaced by v_*, v'_2 and v'_{2*} in V_2 and v_1 in V_1. The second and third will be called decoupling transitions. Finally, for $k = 1, 2$, each particle v_k in V_k, at rate $2B(v_k - v_*, d\sigma)\rho_t(dv_*)\,dt$, dies and is replaced by three particles, v_*, v'_k and v'_{k*} in V_k. It is easy to check, by the triangle inequality, that in each coupled transition, we have $|v'_1 - v'_2| \le |v_1 - v_2|$ and $|v'_{1*} - v'_{2*}| \le |v_1 - v_2|$.

Fix $v_1, v_2 \in \mathbb{R}^d$, and suppose we start with one particle $(v_1, v_2) \in V_0$ at time 0. Write $(\Gamma^0_t, \Gamma^1_t, \Gamma^2_t)_{t\ge0}$ for the empirical process of particle types on $V_0 \cup V_1 \cup V_2$. Then, inductively, Γ^0_t is supported on pairs (u_1, u_2) with $|u_1 - u_2| \le |v_1 - v_2|$. For $k = 1, 2$, write p_k for the projection to the kth component $V_0 \to \mathbb{R}^d$, and write π_k for the bijection $V_k \to \mathbb{R}^d$. Define a measure Λ^k_t on $\mathbb{R}^d$ by

$$
\Lambda^k_t = \Gamma^0_t \circ p_k^{-1} + \Gamma^k_t \circ \pi_k^{-1}.
$$

It is straightforward to check that $(\Lambda^1_t)_{t\ge0}$ and $(\Lambda^2_t)_{t\ge0}$ are copies of the Markov process $(\Lambda_t)_{t\ge0}$ starting from δ_{v_1} and δ_{v_2}, respectively.

For $k = 0, 1, 2$, consider the signed space $V^*_k = V^-_k \cup V^+_k = V_k \times \{-1, 1\}$. The process $(\Gamma^0_t, \Gamma^1_t, \Gamma^2_t)_{t\ge0}$ lifts in an obvious way to a branching process

$(\Gamma_t^{0,*}, \Gamma_t^{1,*}, \Gamma_t^{2,*})_{t\geq 0}$ in $V_0^* \cup V_1^* \cup V_2^*$ starting from $((v_1, v_2), 1)$ in V_0^+, where the "v_*" offspring switch signs, just as in $(\Lambda_t^*)_{t\geq 0}$. By lift we mean that $\Gamma_t = \Gamma_t^* \circ \pi^{-1}$ for the projection $\pi : V_k^* \to V_k$. We write $E_{(0,v_1,v_2)}$ for the expectation over this process. For $k = 1, 2$, set

$$\Lambda_t^{k,*} = \Gamma_t^{0,*} \circ p_k^{-1} + \Gamma_t^{k,*} \circ \pi_k^{-1}.$$

Then $(\Lambda_t^{k,*})_{t\geq 0}$ is a linearized Kac process with environment $(\rho_t)_{t\geq 0}$ starting from $(v_k, 1)$.

LEMMA 4.5. *Assume condition* (2). *Then*

$$\begin{aligned} &E_{(0,v_1,v_2)}\langle 1 + |v|^2, \Gamma_t^1 + \Gamma_t^2\rangle \\ &\qquad \leq 6\kappa t\bigl(2 + |v_1|^2 + |v_2|^2\bigr)|v_1 - v_2| \exp\biggl\{\int_0^t 8\langle 1 + |v|^3, \rho_s\rangle\, ds\biggr\}. \end{aligned}$$

Moreover, for all $p \in (2, \infty)$, there is a constant $C(p) < \infty$ such that

$$\begin{aligned} &E_{(0,v_1,v_2)}\langle 1 + |v|^p, \Gamma_t^1 + \Gamma_t^2\rangle \\ &\qquad \leq C(p)\kappa t\bigl(1 + |v_1|^p + |v_2|^p\bigr)|v_1 - v_2| \exp\biggl\{\int_0^t c(p)\langle 1 + |v|^{p+1}, \rho_s\rangle\, ds\biggr\}. \end{aligned}$$

PROOF. The decoupling transition $(u_1, u_2) \to (u_1', u_{1*}', v_*; u_2)$ occurs at rate

$$2\Gamma_{t-}^0\bigl(d(u_1, u_2)\bigr)\bigl(B(u_1 - v_*, d\sigma) - B(u_2 - v_*, d\sigma)\bigr)^+ \rho_t(dv_*)\, dt$$

and increases $\langle 1 + |v|^2, \Gamma_t^1 + \Gamma_t^2\rangle$ by $4 + 2|v_*|^2 + |u_1|^2 + |u_2|^2$. On adding the rate for the other decoupling transition $(u_1, u_2) \to (u_1; u_2', u_{2*}', v_*)$, we see that a decoupling transition which increases $\langle 1 + |v|^2, \Gamma_t^1 + \Gamma_t^2\rangle$ by $4 + 2|v_*|^2 + |u_1|^2 + |u_2|^2$ occurs at total rate

$$2\Gamma_{t-}^0\bigl(d(u_1, u_2)\bigr)\bigl\|B(u_1 - v_*, \cdot) - B(u_2 - v_*, \cdot)\bigr\|\rho_t(dv_*)\, dt.$$

By condition (2),

$$\bigl\|B(u_1 - v_*, \cdot) - B(u_2 - v_*, \cdot)\bigr\| \leq \kappa|u_1 - u_2| \leq \kappa|v_1 - v_2|$$

for all pairs (u_1, u_2) in the support of Γ_t^0 for all t. Hence the drift of $\langle 1 + |v|^2, \Gamma_t^1 + \Gamma_t^2\rangle$ due to decoupling transitions is no greater than

$$6\kappa|v_1 - v_2|\langle 2 + |u_1|^2 + |u_2|^2, \Gamma_{t-}^0\rangle.$$

On the other hand, by the same estimates used in Proposition 4.1, the drift of $\langle 1 + |v|^2, \Gamma_t^1 + \Gamma_t^2\rangle$ due to branching of uncoupled particles is no greater than

$$8m_3(t)\langle 1 + |v|^2, \Gamma_{t-}^1 + \Gamma_{t-}^2\rangle.$$

Hence the following process is a supermartingale:

$$\langle 1+|v|^2, \Gamma_t^1+\Gamma_t^2\rangle - 6\kappa|v_1-v_2|\int_0^t \langle 2+|u_1|^2+|u_2|^2, \Gamma_s^0\rangle\,ds$$
$$-\int_0^t 8m_3(s)\langle 1+|v|^2, \Gamma_s^1+\Gamma_s^2\rangle\,ds.$$

Set $g(t)=E_{(0,v_1,v_2)}(\langle 1+|v|^2, \Gamma_t^1+\Gamma_t^2\rangle)$. Since $\Gamma_t^0\circ p_k^{-1}\le \Lambda_t^k$, by Proposition 4.1

$$E_{(0,v_1,v_2)}\langle 2+|u_1|^2+|u_2|^2, \Gamma_t^0\rangle \le (2+|v_1|^2+|v_2|^2)\exp\left\{\int_0^t 8m_3(s)\,ds\right\}.$$

Then

$$g(t)\le 6\kappa(2+|v_1|^2+|v_2|^2)|v_1-v_2|\int_0^t \exp\left\{\int_0^s 8m_3(r)\,dr\right\}ds$$
$$+\int_0^t 8m_3(s)g(s)\,ds,$$

and the first of the claimed estimates follows by Gronwall's lemma. For $p>2$, a straightforward modification of this argument, using $|v'|^p+|v'_*|^p\le C(p)(|v|^p+|v_*|^p)$ and (20), leads to the second estimate. $\square$

PROOF OF PROPOSITION 4.3. For all $f\in\mathcal{F}$ and all $v, v'\in\mathbb{R}^d$, we have

$$|f(v)|\le 1+|v|^2, \qquad |f(v)-f(v')|\le (2+|v|^2+|v'|^2)|v-v'|.$$

To see the second inequality, note that

$$\begin{aligned}|f(v)-f(v')| &= |(1+|v|^2)\hat f(v)-(1+|v'|^2)\hat f(v')|\\ &\le (1+|v|^2)|\hat f(v)-\hat f(v')| + ||v|^2-|v'|^2||\hat f(v')|\\ &\le (1+|v|^2+|v|+|v'|)|v-v'|\end{aligned}$$

and then symmetrize. We write the proof for the case $\|f\|=1$ and $s=0$. Set $f_0=E_{0t}f$. By Proposition 4.1, for all $v\in\mathbb{R}^d$,

$$|f_0(v)|\le (1+|v|^2)\exp\left\{\int_0^t 8m_3(s)\,ds\right\}. \tag{22}$$

We have

$$f_0(v_1)-f_0(v_2)=E_{(0,v_1,v_2)}(\langle f\circ p_1 - f\circ p_2, \tilde\Gamma_t^0\rangle + \langle f\circ\pi_1, \tilde\Gamma_t^1\rangle - \langle f\circ\pi_2, \tilde\Gamma_t^2\rangle).$$

So, since $|u_1-u_2|\le|v_1-v_2|$ for all $(u_1,u_2)\in\operatorname{supp}\Gamma_t^0$,

$$|f_0(v_1)-f_0(v_2)|\le E_{(0,v_1,v_2)}(\langle 2+|u_1|^2+|u_2|^2, \Gamma_t^0\rangle|v_1-v_2| + \langle 1+|v|^2, \Gamma_t^1+\Gamma_t^2\rangle).$$

By Proposition 4.1,

$$
\begin{aligned}
E_{(0,v_1,v_2)}\langle 2+|u_1|^2+|u_2|^2, \Gamma_t^0\rangle
&\le E_{(0,v_1)}\langle 1+|v|^2, \Lambda_t\rangle + E_{(0,v_2)}\langle 1+|v|^2, \Lambda_t\rangle \\
&\le \bigl(2+|v_1|^2+|v_2|^2\bigr)\exp\biggl\{\int_0^t 8m_3(s)\,ds\biggr\}.
\end{aligned}
$$

We combine this with Lemma 4.5 to obtain

$$
\bigl|f_0(v_1)-f_0(v_2)\bigr| \le (1+6\kappa t)\bigl(2+|v_1|^2+|v_2|^2\bigr)|v_1-v_2|\exp\biggl\{\int_0^t 8m_3(s)\,ds\biggr\},
$$

which implies that

$$
\bigl|\hat{f}_0(v_1)-\hat{f}_0(v_2)\bigr| \le 3(1+6\kappa t)|v_1-v_2|\exp\biggl\{\int_0^t 8m_3(s)\,ds\biggr\}
$$

and in conjunction with (22) gives the claimed estimate. □

PROOF OF PROPOSITION 4.4. It will suffice to consider the case $\|f\|=1$. Write $f_s=E_{st}f$. Let $(\Lambda_t)_{t\ge s}$ and $(\Lambda_t')_{t\ge s'}$ be independent linearized Kac processes starting from δ_{v_0} at times s and s', respectively. Write T for the first branch time of $(\Lambda_t)_{t\ge s}$ and V_*, V', V_*' for the velocities of the new particles formed in $(\Lambda_t)_{t\ge s}$ at time T. By the Markov property of the branching process and using Proposition 4.1, on the event $\{T\le s'\}$,

$$
\begin{aligned}
&\bigl|E\bigl(\langle f,\tilde{\Lambda}_t-\tilde{\Lambda}_t'\rangle|T,V_*,V',V_*'\bigr)\bigr| \\
&\quad= \bigl|f_T\bigl(V'\bigr)+f_T\bigl(V_*'\bigr)-f_T(V_*)-f_{s'}(v_0)\bigr| \\
&\quad\le \bigl(4+|v_0|^2+|V_*|^2+\bigl|V'\bigr|^2+\bigl|V_*'\bigr|^2\bigr)\exp\biggl\{\int_s^t 8m_3(r)\,dr\biggr\}
\end{aligned}
$$

while

$$
E\bigl(\langle f,\tilde{\Lambda}_t-\tilde{\Lambda}_t'\rangle|T>s'\bigr)=0.
$$

Now $|V'|^2+|V_*'|^2=|v_0|^2+|V_*|^2$, so

$$
\begin{aligned}
\bigl|f_s(v_0)-f_{s'}(v_0)\bigr| &= \bigl|E\langle f,\tilde{\Lambda}_t-\tilde{\Lambda}_t'\rangle\bigr| \\
&\le E\bigl(\bigl(4+2|v_0|^2+2|V_*|^2\bigr)1_{\{T\le s'\}}\bigr)\exp\biggl\{\int_s^t 8m_3(r)\,dr\biggr\},
\end{aligned}
$$

and, using the inequality $|v-v_*|(4+2|v|^2+2|v_*|^2)\le 5(1+|v|^3)(|1+|v_*|^3)$, we have

$$
\begin{aligned}
&E\bigl(\bigl(4+2|v_0|^2+2|V_*|^2\bigr)1_{\{T\le s'\}}\bigr) \\
&\quad\le \int_s^{s'}\int_{\mathbb{R}^d}|v_0-v_*|\bigl(4+2|v_0|^2+2|v_*|^2\bigr)\rho_r(dv_*)\,dr \\
&\quad\le 5\bigl(1+|v_0|^3\bigr)\int_s^{s'} m_3(r)\,dr,
\end{aligned}
$$

so

$$\big|f_s(v_0) - f_{s'}(v_0)\big| \le 5\big(1 + |v_0|^3\big)\exp\left\{\int_s^t 8m_3(r)\,dr\right\}\int_s^{s'} m_3(r)\,dr. \qquad \square$$

PROOF OF PROPOSITION 4.2. Recall that now $\rho_t = (\mu_t^N + \mu_t^{N'})/2$ and $m_3(t) = \langle 1 + |v|^3, \mu_t^N + \mu_t^{N'}\rangle/2$. In particular $m_3(t) \le 1 + (N')^{3/2} < \infty$ for all t. Set $M = M^N - M^{N'}$, and write $M^{\pm}$ for the positive and negative parts of the signed measure M on $[0,\infty) \times \mathbb{R}^d$. Consider a branching particle system in V^*, with the same branching rules as $(\Lambda_t^*)_{t \ge s}$ above, but where, instead of starting with just one particle at time s, we initiate particles randomly in the system according to a Poisson random measure on $[0,\infty) \times V^*$ of intensity

$$\theta(dt, dv) = \begin{cases} \delta_0(dt)\mu_0^N(dv) + M^+(dt, dv), & \text{on } [0,\infty) \times V^+, \\ \delta_0(dt)\mu_0^{N'}(dv) + M^-(dt, dv), & \text{on } [0,\infty) \times V^-. \end{cases}$$

We use the same notation as above for the empirical measures associated to the branching process, and signify the new rule for initiating particles by writing now E for the expectation. Define, for $t \ge 0$, a signed measure λ_t on $\mathbb{R}^d$ by

$$\lambda_t = E(\tilde\Lambda_t) = \int_{[0,t]\times V} E_{(s,v)}(\tilde\Lambda_t)\theta(ds, dv). \tag{23}$$

Then, by Proposition 4.1,

$$\big\langle 1 + |v|^2, |\lambda_t|\big\rangle \le \exp\left\{\int_0^t 8m_3(s)\,ds\right\}\int_{[0,t]\times\mathbb{R}^d}\big(1 + |v|^2\big)\big|\theta(ds, dv)\big|$$

and, by estimate (8),

$$\begin{aligned}&\int_{[0,t]\times\mathbb{R}^d}\big(1 + |v|^2\big)\big|\theta(ds, dv)\big| \\ &\qquad \le \big\langle 1 + |v|^2, \mu_0^N + \mu_0^{N'}\big\rangle + \int_{[0,t]\times\mathbb{R}^d}\big(1 + |v|^2\big)\big|M(ds, dv)\big| < \infty.\end{aligned}$$

We see in particular that $\langle 1 + |v|^2, |\lambda_t|\rangle$ is bounded on compacts in t.

Under $E_{(s,v)}$, the pair of empirical processes of positive and negative particles $(\Lambda_t^+, \Lambda_t^-)_{t \ge s}$ evolves as a Markov chain, which makes jumps $(\delta_{v'} + \delta_{v_*'} - \delta_v, \delta_{v_*})$ at rate $2\Lambda_{t-}^+(dv)\rho_t(dv_*)B(v - v_*, d\sigma)\,dt$ and makes jumps $(\delta_{v_*}, \delta_{v'} + \delta_{v_*'} - \delta_v)$ at rate $2\Lambda_{t-}^-(dv)\rho_t(dv_*)B(v - v_*, d\sigma)\,dt$. So, using Proposition 4.1 for integrability, under $E_{(s,v)}$, for any bounded measurable function f, the following process is a martingale:

$$\langle f, \tilde\Lambda_t\rangle - \int_s^t \big\langle f, 2Q(\rho_r, \tilde\Lambda_r)\big\rangle\,dr, \qquad t \ge s.$$

Taking expectations and setting $f_{st}(v) = E_{st}f(v) = E_{(s,v)}\langle f, \tilde\Lambda_t\rangle$, we obtain

$$f_{st}(v) = f(v) + \int_s^t \big\langle f, 2Q\big(\rho_r, E_{(s,v)}(\tilde\Lambda_r)\big)\big\rangle\,dr. \tag{24}$$

Then

$$
\begin{aligned}
\langle f, \lambda_t \rangle &= \int_{[0,t]\times V} E_{(s,v)} \langle f, \tilde{\Lambda}_t \rangle \theta(ds, dv) \\
&= \langle f_{0t}, \mu_0^N - \mu_0^{N'} \rangle + \int_0^t \langle f_{st}, dM_s \rangle \\
&= \langle f, \mu_0^N - \mu_0^{N'} \rangle + \langle f, M_t \rangle + \int_0^t \langle f, 2Q(\rho_r, \lambda_r) \rangle \, dr.
\end{aligned}
$$

Here, we used (23) for the first equality, and for the third we substituted for f_{0t} and f_{st} using (24) and then rearranged the integrals using Fubini to make λ_r, as given by (23), appear on the inside. Since f is an arbitrary bounded measurable function, we have shown that

$$
\lambda_t = (\mu_0^N - \mu_0^{N'}) + M_t + \int_0^t 2Q(\rho_s, \lambda_s)\, ds. \tag{25}
$$

Note the estimate of total variation,

$$
\|Q(\rho_t, \lambda_t)\| \le 4 \int_{\mathbb{R}^d \times \mathbb{R}^d} |v - v_*| \rho_t(dv) |\lambda_t|(dv_*) \le 6\langle 1 + |v|^2, |\lambda_t| \rangle.
$$

For the second inequality, we used $\langle 1, \rho_t \rangle = \langle |v|^2, \rho_t \rangle = 1$ and $2|v - v_*| \le 2 + |v|^2 + |v_*|^2$. For any interval $(s, t]$ on which neither $(\mu_t^N)_{t\ge 0}$ nor $(\mu_t^{N'})_{t\ge 0}$ jump, using estimate (9), we deduce that

$$
\langle 1 + |v|^2, |M_t - M_s| \rangle \le 24 \int_s^t m_3(r)\, dr.
$$

On the other hand

$$
\left\langle 1 + |v|^2, \int_s^t |Q(\rho_r, \lambda_r)|\, dr \right\rangle \to 0
$$

as $t \downarrow s$, for all $s \ge 0$. Hence from equation (25) we deduce that $(1 + |v|^2)\lambda_t$ is right continuous in total variation.

Set $\lambda'_t = \mu_t^N - \mu_t^{N'}$, and note from (6) that $(\lambda'_t)_{t\ge 0}$ also satisfies (25). We subtract to see that $\delta_t = \lambda_t - \lambda'_t$ satisfies

$$
\delta_t = \int_0^t 2Q(\rho_s, \delta_s)\, ds.
$$

Set $\nu_t = 2Q(\rho_t, \delta_t)$. Then $\|\nu_t\| \le 12\langle 1 + |v|^2, |\delta_t| \rangle$ and, on any interval $(s, t]$ when neither $(\mu_t^N)_{t\ge 0}$ nor $(\mu_t^{N'})_{t\ge 0}$ jump,

$$
\|\nu_t - \nu_s\| = \|2Q(\rho_s, \lambda_t - \lambda_s)\| \le 12\langle 1 + |v|^2, |\lambda_t - \lambda_s| \rangle.
$$

The process of signed measures $(\nu_t)_{t\ge 0}$ is thus locally bounded and right continuous in total variation. Hence the measure $\int_0^T |\nu_t|\, dt + |\nu_T|$ is finite, and ν_t is absolutely continuous with respect to this measure for all $t \in [0, T]$.

We apply Lemma A.1, with $\mu_0 = 0$, to obtain a measurable map $\sigma : [0,\infty) \times V \to \{-1, 0, 1\}$ such that $\delta_t = \sigma_t|\delta_t|$ and $|\delta_t| = \int_0^t \sigma_s \nu_s \, ds$. Set $\check{\sigma}_s(v) = (1 + |v|^2)\sigma_s(v)$. Then

$$
\begin{aligned}
&\langle 1 + |v|^2, |\delta_t|\rangle \\
&\quad = \int_E 2\{\check{\sigma}_s(v') + \check{\sigma}_s(v'_*) - \check{\sigma}_s(v) - \check{\sigma}_s(v_*)\} \\
&\qquad \times 1_{(0,t]}(s) B(v - v_*, d\sigma)\rho_s(dv_*)\delta_s(dv)\, ds \\
&\quad \le \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} 4(1 + |v_*|^2)|v - v_*|\rho_s(dv_*)|\delta_s|(dv)\, ds \\
&\quad \le \int_0^t 4\langle 1 + |v|^2, |\delta_s|\rangle m_3(s)\, ds.
\end{aligned}
$$

But $\int_0^t m_3(s)\, ds < \infty$, so $\delta_t = 0$, for all t. □

5. Proof of Theorem 1.1. We will write the proof for $d \ge 3$, leaving the minor modifications necessary for $d = 2$ to the reader. Fix $p \in (2, \infty)$ and $\lambda \in [1, \infty)$. Suppose that $(\mu_t^N)_{t \ge 0}$ and $(\mu_t^{N'})_{t \ge 0}$ are Kac processes in $\mathcal{S}_N$ and $\mathcal{S}_{N'}$, respectively, with $\langle |v|^p, \mu_0^N\rangle \le \lambda$ and $\langle |v|^p, \mu_0^{N'}\rangle \le \lambda$. Set $\rho_t = (\mu_t^N + \mu_t^{N'})/2$. Fix $t \in [0, T]$ and a function $f_t \in \mathcal{F}$. Define a random function f on $[0, T] \times \mathbb{R}^d$ by

$$f(s, v) = f_s(v) = E_{(s \wedge t, v)}\langle f_t, \tilde{\Lambda}_t\rangle,$$

where $(\Lambda_t^*)_{t \ge s}$ is a linearized Kac process in environment $(\rho_t)_{t \ge 0}$. Note that we have extended f as a constant in time from t to T. We have, by Proposition 4.2,

$$\langle f_t, \mu_t^N - \mu_t^{N'}\rangle = \langle f_0, \mu_0^N - \mu_0^{N'}\rangle + \int_0^t \langle f_s, dM_s^N\rangle - \int_0^t \langle f_s, dM_s^{N'}\rangle. \tag{26}$$

Write $m_q(t) = \langle 1 + |v|^q, \rho_t\rangle$, as above. By Proposition 3.1, for $q < p + 1$, there is a constant $C(B, p, q) < \infty$ such that

$$\mathbb{E}\int_0^T m_q(s)\, ds \le C(T^{p+1-q} + T)\lambda. \tag{27}$$

Set

$$A = 3(1 + 6\kappa T)\exp\left\{\int_0^T 8m_3(s)\, ds\right\}. \tag{28}$$

By Proposition 4.3, for all $s \le t$, we have

$$\|f_s\| \le A \tag{29}$$

so

$$\langle f_0, \mu_0^N - \mu_0^{N'}\rangle \le AW(\mu_0^N, \mu_0^{N'}). \tag{30}$$

The main step of the proof is to bound the second and third terms on the right in (26), uniformly in $t \in [0, T]$ and $f_t \in \mathcal{F}$. We will derive estimates for the second term, which then apply also to the third, because $N \le N'$. The notation conceals the fact that the integrand f_s depends on the terminal time t. Worse, f_s depends on $(\mu_r^N + \mu_r^{N'})_{s\le r\le t}$, so is anticipating, and martingale estimates cannot be applied directly even at the individual time t.

For $p \ge 3$, set $\beta = 1$ and $Z = \sup_{t\in[0,T]} m_3(t)$. By Propositions 3.1 and 4.4, we have $\mathbb{E}(Z) \le 1 + (1 + CT)\lambda$ and, for all $v \in \mathbb{R}^d$ and $s, s' \in [0, T]$ with $s \le s'$,

$$|f_s(v) - f_{s'}(v)| \le A'(1 + |v|^3)(s' - s)^{\beta}, \tag{31}$$

where

$$A' = 5Z \exp\left\{\int_0^T 8m_3(s)\,ds\right\}. \tag{32}$$

For $p \in (2, 3)$, set $\beta = (p - 2)/2$, and set

$$Z = 2 \sup_{t\in[1,T]} m_3(t) + \sum_{\ell\in\mathbb{N}} 2^{(\beta-1)\ell+1}\beta^{-1} \sup_{t\in[2^{-\ell},2^{-\ell+1}]} m_3(t).$$

By Proposition 3.1, there is a constant $C(B, p) < \infty$ such that, for $t \le T$,

$$\mathbb{E}\Big(\sup_{s\in[t,T]} m_3(s)\Big) \le C(t^{p-3} \vee T)\lambda$$

so

$$\mathbb{E}(Z) \le C\lambda\left(T + \sum_{\ell\in\mathbb{N}} 2^{(\beta-1)\ell}2^{-\ell(p-3)}\right) = C\lambda(T + 1/(2^{\beta} - 1)). \tag{33}$$

Note that $m_3(t) \le (\beta t^{\beta-1} + 1)Z/2$ for all $t \le T$, so for $s \le s' \le T$ with $s' - s \le 1$,

$$\int_s^{s'} m_3(t)\,dt \le ((s' - s)^{\beta} + (s' - s))Z/2 \le (s' - s)^{\beta} Z.$$

Hence, by Proposition 4.4, (31) remains valid for $p \in (2, 3)$, provided $s' - s \le 1$ and β and Z have their new meanings.

Fix $r \in (0, 1]$ and $R \in [1, \infty)$ such that T/r and R/r are integers, and set $n = (T/r) \times (R/r)^d$. Set $B(R) = (-R, R]^d$. There exist $s_1, \dots, s_n \in [0, T)$ and $v_1, \dots, v_n \in B(R)$ such that $B_1 \cup \dots \cup B_n = (0, T] \times B(R)$, where $B_k = (s_k, v_k) + (0, r] \times (-r, r]^d$. Write

$$f = \sum_{k=1}^n a_k f^{(k)} + g,$$

where a_k is the average value of $\hat{f}$ on B_k and $f^{(k)}(s, v) = \check{1}_{B_k}(s, v) = (1 + |v|^2)1_{B_k}(s, v)$. Then

$$\int_0^t \langle f_s, dM_s^N\rangle = \sum_{k=1}^n a_k M_t^{(k)} + \int_0^t \langle g_s, dM_s^N\rangle, \tag{34}$$

where

$$M_t^{(k)} = \int_0^t \langle f_s^{(k)}, dM_s^N \rangle.$$

Now, by (29), for all k, we have $|a_k| \le A$ and, for $v, v' \in B_k$,

$$\big|\hat{f}_s(v) - \hat{f}_s(v')\big| \le A\big|v - v'\big| \le 2\sqrt{d}Ar.$$

By (31), we have, for $(s, v) \in B_k$,

$$\big(1 + |v|^2\big)\big|\hat{f}_s(v) - a_k(v)\big| \le A'\big(1 + |v|^3\big)r^\beta,$$

where $a_k(v)$ is the average value of $\hat{f}$ on $(s_k, s_k + r] \times \{v\}$. Hence, for $(s, v) \in B_k$,

$$\begin{aligned}\big|g_s(v)\big| &= \big(1 + |v|^2\big)\big|\hat{f}_s(v) - a_k(v) + a_k(v) - a_k\big| \\ &\le A'\big(1 + |v|^3\big)r^\beta + 2\sqrt{d}A\big(1 + |v|^2\big)r.\end{aligned}$$

On the other hand, $|g_s(v)| \le A(1 + |v|^2)$ for all $v \in \mathbb{R}^d \setminus B(R)$.

Set $Q_t = Q_t^N = \sum_{k=1}^n |M_t^{(k)}|^2$. Then

$$\left|\sum_{k=1}^n a_k M_t^{(k)}\right| \le A\sqrt{nQ_t}. \tag{35}$$

Fix $q \in (3, p + 1)$. Note that, for $s \in (0, T]$,

$$\begin{aligned}&\sum_{k=1}^n \big\{f_s^{(k)}(v') + f_s^{(k)}(v'_*) - f_s^{(k)}(v) - f_s^{(k)}(v_*)\big\}^2 \\ &\qquad \le 4\sum_{k=1}^n \big\{\check{\check{1}}_{B_k}(s, v') + \check{\check{1}}_{B_k}(s, v'_*) + \check{\check{1}}_{B_k}(s, v) + \check{\check{1}}_{B_k}(s, v_*)\big\} \\ &\qquad = 4\big\{\check{\check{1}}_{B(R)}(v') + \check{\check{1}}_{B(R)}(v'_*) + \check{\check{1}}_{B(R)}(v) + \check{\check{1}}_{B(R)}(v_*)\big\} \\ &\qquad \le CR^{(5-q)^+}\big(1 + |v|^{q-1} + |v_*|^{q-1}\big)\end{aligned}$$

for some constant $C < \infty$, depending only on d and q. So, by Doob's L^2-inequality,

$$\begin{aligned}\mathbb{E}\Big(\sup_{t\le T} Q_t\Big) &\le \frac{4}{N^2}\sum_{k=1}^n \mathbb{E}\int_E \big\{f_s^{(k)}(v') + f_s^{(k)}(v'_*) - f_s^{(k)}(v) - f_s^{(k)}(v_*)\big\}^2 \\ &\qquad\qquad \times 1_{(0,T]}(s)\bar{m}(dv, dv_*, d\sigma, ds) \\ &\le \frac{CR^{(5-q)^+}}{N}\mathbb{E}\int_0^T \langle |v|^q, \mu_s^N\rangle\, ds.\end{aligned} \tag{36}$$

On the other hand

$$\left|\int_0^t \langle g_s, dM_s^N\rangle\right| \le C\big(A'r^\beta R^{(4-q)^+} + Ar + AR^{3-q}\big)\int_0^T \langle 1 + |v|^{q-1}, \big|dM_s^N\big|\rangle \tag{37}$$

and

$$\begin{aligned}
&\mathbb{E}\int_0^T \langle 1+|v|^{q-1}, |dM_s^N|\rangle \\
&\quad \le \frac{2}{N}\mathbb{E}\int_E \{4+|v'|^{q-1}+|v'_*|^{q-1}+|v|^{q-1}+|v_*|^{q-1}\} \\
&\qquad\qquad \times 1_{(0,T]}(s)\bar{m}(dv,dv_*,d\sigma,ds) \qquad (38) \\
&\quad \le C\mathbb{E}\int_0^T \langle |v|^q, \mu_s^N\rangle ds.
\end{aligned}$$

We combine (26), (27), (28), (30), (32), (33), (34), (35), (36), (37) and (38) to see that, for all $\varepsilon \in (0,1]$, there is a constant $C < \infty$, depending only on $B, d, \varepsilon, \lambda, p, q$ and T, such that, for all $N, N' \in \mathbb{N}$ with $N \le N'$, with probability exceeding $1-\varepsilon$, we have, for all $t \in [0,T]$,

$$W(\mu_t^N, \mu_t^{N'}) \le C\big(W(\mu_0^N, \mu_0^{N'}) + \sqrt{R^{(5-q)^+}n/N} + rR^{(4-q)^+} + R^{3-q}\big).$$

An optimization over q, r and R now shows the existence of an $\alpha(d,p) > 0$ for which the estimate claimed in Theorem 1.1 holds.

For large p, the reader may check the optimization yields a value for $\alpha(d,p)$ close to $1/(d+3)$. The proof given can be varied by replacing the one-step discrete approximation by a chaining argument. See the proof of Proposition 9.2 for this idea in a simple context. This gives $\alpha(d,p) = 1/(d+1)$ for p sufficiently large. We omit the details because Theorem 1.2 gives a stronger result. Here $d+1$ is the dimension of space–time, reflecting the fact that we maximize over a class of functions on $[0,T] \times \mathbb{R}^d$. This is wasteful because, in fact, we only need to maximize over a certain process of functions $(f_s : s \in [0,t])$ associated to $t \in [0,T]$ and $f = f_t$ and then over a class of functions f on $\mathbb{R}^d$. In the next three sections, we exploit the structure of the process $(f_s : s \in [0,t])$ to obtain an improved bound.

6. Continuity of the linearized Kac process in its environment. We showed in Propositions 4.3 and 4.4 that the linearized Kac process is continuous in its initial data. For the proof of our main estimate with optimal rate $N^{-1/d}$, we will need also continuity in the environment. The following notation will be convenient. For $p \in [2,\infty)$ and a function f on $\mathbb{R}^d$, we will write $\hat{f}^{(p)}$ for the reweighted function $\hat{f}^{(p)}(v) = f(v)/(1+|v|^p)$ and write $\|f\|_{(p)}$ for the smallest constant such that, for all $v, v' \in \mathbb{R}^d$, we have

$$|\hat{f}^{(p)}(v)| \le \|f\|_{(p)}, \qquad |\hat{f}^{(p)}(v) - \hat{f}^{(p)}(v')| \le \|f\|_{(p)}|v-v'|.$$

Denote by $\mathcal{F}(p)$ the set of all functions f on $\mathbb{R}^d$ with $\|f\|_{(p)} \le 1$. We earlier wrote $\mathcal{F}$ for $\mathcal{F}(2)$ and $\|f\|$ for $\|f\|_{(2)}$. We will use the cases $p = 2$ and $p = 3$. Suppose that $(\rho_t^1)_{t\ge 0}$ and $(\rho_t^2)_{t\ge 0}$ are processes of measures on $\mathbb{R}^d$, both satisfying (17).

Given $t \geq 0$ and a function f of quadratic growth on $\mathbb{R}^d$, define for $s \in [0,t]$ and $v \in \mathbb{R}^d$, and for $j = 1, 2$,

$$E^j_{st} f(v) = E_{(s,v)}\langle f, \tilde{\Lambda}^j_t \rangle,$$

where $(\Lambda^{j,*}_t)_{t \geq s}$ is a linearized Kac process with environment $(\rho^j_t)_{t \geq 0}$ starting from v at time s. We will use the following notation:

$$d_p(t) = \langle 1 + |v|^p, |\rho^1_t - \rho^2_t| \rangle, \qquad \bar{m}_p(t) = \langle 1 + |v|^p, \rho^1_t + \rho^2_t \rangle.$$

PROPOSITION 6.1. *For all $p \in [2, \infty)$, there is a constant $C(p) < \infty$ with the following properties. Let f be a function on $\mathbb{R}^d$ with $|f(v)| \leq 1 + |v|^p$ for all v. Then, for all $s, t \geq 0$ with $s \leq t$ and all $v \in \mathbb{R}^d$, we have*

$$\begin{aligned} &\left|E^1_{st} f(v) - E^2_{st} f(v)\right| \\ &\quad \leq C(p)\left(1 + |v|^{p+1}\right) \exp\left\{C(p) \int_s^t \bar{m}_{p+2}(r)\, dr\right\} \int_s^t d_{p+1}(r)\, dr. \end{aligned}$$

Assume that the collision kernel satisfies condition (2). *Then $C(p)$ may be chosen so that*

$$\|E^1_{st} f - E^2_{st} f\|_{(p+1)} \leq C(p)\kappa \|f\|_{(p)} \exp\left\{C(p) \int_s^t \bar{m}_{p+2}(r)\, dr\right\} \int_s^t d_{p+1}(r)\, dr.$$

Our first step toward a proof of Proposition 6.1 is to describe a coupling of $(\Lambda^{1,*}_t)_{t \geq 0}$ and $(\Lambda^{2,*}_t)_{t \geq 0}$ when both process start from v at time 0. For this, we take as type space the set $\hat{V}_0 \cup \hat{V}_1 \cup \hat{V}_2$, where $\hat{V}_j = \mathbb{R}^d \times \{j\}$ for $j = 0, 1, 2$. Particles with types in $\hat{V}_0$ are called coupled, the others are uncoupled. Consider the branching process with the following branching transitions. For a particle v in $\hat{V}_0$, there are three possible transitions. First, at rate $2(\rho^1_t \wedge \rho^2_t)(dv_*) B(v - v_*, d\sigma)\, dt$, we replace v by three particles v_*, v' and v'_* in $\hat{V}_0$. Second, at rate $2(\rho^1_t - \rho^2_t)^+(dv_*) B(v - v_*, d\sigma)\, dt$, we replace v by three particles v_*, v' and v'_* in $\hat{V}_1$ and one particle v in $\hat{V}_2$. Third, at rate $2(\rho^2_t - \rho^1_t)^+(dv_*) B(v - v_*, d\sigma)\, dt$, we replace v by one particle v in $\hat{V}_1$ and three particles v_*, v' and v'_* in $\hat{V}_2$. The second and third are called uncoupling transitions. The transitions for uncoupled particles are as in the original branching process; that is, for $j = 1, 2$ and v in $\hat{V}_j$, at rate $2\rho^j_t(dv_*) B(v - v_*, d\sigma)\, dt$ we replace v by particles v_*, v' and v'_*, also in $\hat{V}_j$. For $v \in \mathbb{R}^d$ and $s \geq 0$, and for $j = 0, 1, 2$, write $\hat{\Gamma}^j_t$ for the un-normalized empirical distribution of particles in $\hat{V}_j$ when we initiate the branching process with a single particle v in $\hat{V}_0$ at time s. Define analogously the lifted processes $(\hat{\Gamma}^{j,*}_t)_{t \geq s}$ in $\hat{V}^*_j$. For $j = 1, 2$, set

$$\hat{\Lambda}^{j,*}_t = \hat{\Gamma}^{0,*}_t \circ \hat{\pi}_0^{-1} + \hat{\Gamma}^{j,*}_t \circ \hat{\pi}_j^{-1},$$

where $\hat{\pi}_j$ is the bijection $\hat{V}_j^* \to V^*$. It is straightforward to check that $(\hat{\Lambda}_t^{j,*})_{t\geq s}$ is a linearized Kac process with environment $(\rho_t^j)_{t\geq 0}$ starting from v in V^+ at time s. We have burdened the notation with hats so that we can later refer simultaneously to this coupling and to the coupling for two different starting points.

LEMMA 6.2. *For all $p \in [2, \infty)$, there is a constant $C(p) < \infty$ such that*

$$E_{(0,v_0)}\langle 1+|v|^p, \hat{\Gamma}_t^1 + \hat{\Gamma}_t^2\rangle \\ \leq C(p)\big(1+|v_0|^{p+1}\big)\exp\left\{\int_0^t C(p)\bar{m}_{p+2}(s)\,ds\right\}\int_0^t d_{p+1}(s)\,ds.$$

PROOF. The process $\langle 1+|v|^p, \hat{\Gamma}_t^1 + \hat{\Gamma}_t^2\rangle$ starts at 0 and makes jumps both at uncoupling transitions and due to the branching of uncoupled particles. Uncoupling transitions occur at rate $2B(v-v_*, d\sigma)\hat{\Gamma}_{t-}^0(dv)|\rho_t^1 - \rho_t^2|(dv_*)\,dt$ and result in jumps of $4+|v'|^p+|v_*'|^p+|v|^p+|v_*|^p$. Uncoupled particles branch at rate $2B(v-v_*, d\sigma)(\hat{\Gamma}_{t-}^1(dv)\rho_t^1(dv_*) + \hat{\Gamma}_{t-}^2(dv)\rho_t^2(dv_*))\,dt$ and result in jumps of $2+|v'|^p+|v_*'|^p+|v_*|^2-|v|^p$. We use the inequalities

$$4+\big|v'\big|^p+\big|v_*'\big|^p+|v|^p+|v_*|^p \leq C(p)\big(1+|v|^p+|v_*|^p\big)$$

and

$$\big(1+|v|^p+|v_*|^p\big)|v-v_*| \leq C(p)\big(1+|v|^{p+1}\big)\big(1+|v_*|^{p+1}\big)$$

to see that the drift of $\langle 1+|v|^p, \hat{\Gamma}_t^1+\hat{\Gamma}_t^2\rangle$ due to uncoupling transitions is no greater than $C(p)d_{p+1}(t)\langle 1+|v|^{p+1}, \hat{\Gamma}_{t-}^0\rangle$. On the other hand, inequalities (19) and (20) show that the drift of $\langle 1+|v|^p, \hat{\Gamma}_t^1+\hat{\Gamma}_t^2\rangle$ due to branching of uncoupled particles is no greater than $c(p)\bar{m}_{p+1}(t)\langle 1+|v|^p, \hat{\Gamma}_{t-}^1+\hat{\Gamma}_{t-}^2\rangle$. Hence the following process is a supermartingale:

$$\langle 1+|v|^p, \hat{\Gamma}_t^1+\hat{\Gamma}_t^2\rangle - C(p)\int_0^t \langle 1+|v|^{p+1}, \hat{\Gamma}_s^0\rangle d_{p+1}(s)\,ds \\ - c(p)\int_0^t \langle 1+|v|^p, \hat{\Gamma}_s^1+\hat{\Gamma}_s^2\rangle \bar{m}_{p+1}(s)\,ds.$$

On taking expectations, we obtain

$$g(t) \leq \int_0^t \big\{C(p)f_{p+1}(s)d_{p+1}(s) + c(p)\bar{m}_{p+1}(s)g(s)\big\}\,ds,$$

where $g(t) = E_{(0,v_0)}\langle 1+|v|^p, \hat{\Gamma}_t^1+\hat{\Gamma}_t^2\rangle$ and $f_p(t) = E_{(0,v_0)}\langle 1+|v|^p, \hat{\Gamma}_t^0\rangle$. By Proposition 4.1, we have

$$f_{p+1}(t) \leq \big(1+|v_0|^{p+1}\big)\exp\int_0^t c(p+1)\bar{m}_{p+2}(s)\,ds,$$

so, for some constant $C(p) < \infty$,

$$g(t) \leq C(p)\big(1 + |v_0|^{p+1}\big) \exp\left\{C(p) \int_0^t \bar{m}_{p+2}(r)\,dr\right\} \int_0^t d_{p+1}(s)\,ds. \qquad \square$$

The second ingredient needed for Proposition 6.1 is a coupling of four linearized Kac processes, with environments $(\rho_t^1)_{t\geq 0}$ and $(\rho_t^2)_{t\geq 0}$ and with starting points v_1 and v_2. We will specify this coupling in detail, at the cost of some heaviness of notation, the understanding of which may be guided by the thought that the coupling is the obvious one for typed branching at different rates and is an elaboration of the couplings described in the case of a single environment or a single starting point above. To define the coupled processes, we consider a type space which is the disjoint union of nine sets,

$$(V_{00} \cup V_{01} \cup V_{02}) \cup (V_{10} \cup V_{11} \cup V_{12}) \cup (V_{20} \cup V_{21} \cup V_{22}).$$

Here, for $k = 0, 1, 2$, we take $V_{0k} = V_k$ as at (21) and, for $j = 1, 2$, we take $V_{jk} = V_k \times \{j\}$. The first index refers to the environment, a 0 indicating a particle present in the branching process in both environments. The second index refers to the starting point.

The branching rules for a particle (v_1, v_2) in V_{00} are as follows. There are nine possible transitions. First, at rate $2(\rho_t^1 \wedge \rho_t^2)(dv_*)(B(v_1 - v_*, d\sigma) \wedge B(v_2 - v_*, d\sigma))\,dt$ (for all $v_* \in \mathbb{R}^d$ and all $\sigma \in S^{d-1}$), we replace (v_1, v_2) by three particles (v_*, v_*), (v_1', v_2') and (v_{1*}', v_{2*}') in V_{00}. As above, we are writing v_k' for $v'(v_k, v_*, \sigma)$ and v_{k*}' for $v_*'(v_k, v_*, \sigma)$. Second, at rate $2(\rho_t^1 - \rho_t^2)^+(dv_*)(B(v_1 - v_*, d\sigma) \wedge B(v_2 - v_*, d\sigma))\,dt$, we replace (v_1, v_2) by three particles (v_*, v_*), (v_1', v_2') and (v_{1*}', v_{2*}') in V_{10} and one particle (v_1, v_2) in V_{20}. Third, at rate $2(\rho_t^1 \wedge \rho_t^2)(dv_*)(B(v_1 - v_*, d\sigma) - B(v_2 - v_*, d\sigma))^+ dt$, we replace (v_1, v_2) by three particles v_*, v_1' and v_{1*}' in V_{01} and one particle v_2 in V_{02}. Fourth, at rate $2(\rho_t^1 - \rho_t^2)^+(B(v_1 - v_*, d\sigma) - B(v_2 - v_*, d\sigma))^+\,dt$, we replace (v_1, v_2) by three particles v_*, v_1' and v_{1*}' in V_{11}, one particle v_2 in V_{12} and one particle (v_1, v_2) in V_{20}. The second and third transitions each have an obvious counterpart by swapping 1 and 2, while there are three variants of the fourth transition by swapping 1 and 2 in the environment or in the collision intensity or in both.

On leaving V_{00}, either the coupling with respect to environment is broken, or that with respect to the starting points. This corresponds to transitions on the one hand to V_{1k} or V_{2k} for some k, or on the other hand to V_{j1} or V_{j2} for some j, respectively. Once the environment coupling is broken, a particle branches as in the starting point coupling, while if the starting point coupling is broken, a particle branches as in the environment coupling. Thus the transitions in V_{jk} for $j = 1, 2$ are as described above for V_k, while those in V_{jk} for $k = 1, 2$ are as described above for $\hat{V}_j$.

For $j, k = 0, 1, 2$, write $(\Xi_t^{jk})_{t\geq 0}$ for the empirical distribution of particles in V_{jk} when we initiate the branching process just described with a single particle

(v_1, v_2) in V_{00} at time 0. Write q_{jk} for the bijection $V_{jk} \to V_k$. For $k = 1, 2$, write p_{jk} for the projection $(v_1, v_2, j) \mapsto (v_k, j) : V_{j0} \to \hat{V}_j$, and write $\hat{q}_{jk}$ for the bijection $V_{jk} \to \hat{V}_j$. Note that $\hat{\pi}_j \circ p_{jk} = p_k \circ q_{j0}$ on V_{j0} and $\hat{\pi}_j \circ \hat{q}_{jk} = \pi_k \circ q_{jk}$ on V_{jk} for $j = 0, 1, 2$ and $k = 1, 2$. For $j, k = 1, 2$, set

$$\Gamma_t^{j0} = \Xi_t^{00} \circ q_{00}^{-1} + \Xi_t^{j0} \circ q_{j0}^{-1}, \qquad \Gamma_t^{jk} = \Xi_t^{0k} \circ q_{0k}^{-1} + \Xi_t^{jk} \circ q_{jk}^{-1}$$

and

$$\hat{\Gamma}_t^{0k} = \Xi_t^{00} \circ p_{0k}^{-1} + \Xi_t^{0k} \circ \hat{q}_{0k}^{-1}, \qquad \hat{\Gamma}_t^{jk} = \Xi_t^{j0} \circ p_{jk}^{-1} + \Xi_t^{jk} \circ \hat{q}_{jk}^{-1},$$

and set

$$\Lambda_t^{jk} = \Gamma_t^{j0} \circ \pi_k^{-1} + \Gamma_t^{jk} \circ p_k^{-1} = \hat{\Gamma}_t^{0k} \circ \hat{\pi}_0^{-1} + \hat{\Gamma}_t^{jk} \circ \hat{\pi}_j^{-1}.$$

It can be checked that $(\Lambda_t^{jk})_{t \geq 0}$ is a copy of the branching process $(\Lambda_t)_{t \geq 0}$ starting from v_k at time 0 in the environment $(\rho_t^j)_{t \geq 0}$. Moreover $(\Gamma_t^{j0}, \Gamma_t^{j1}, \Gamma_t^{j2})_{t \geq 0}$ is a copy of the starting point coupling in environment $(\rho_t^j)_{t \geq 0}$, and $(\hat{\Gamma}_t^{0k}, \hat{\Gamma}_t^{1k}, \hat{\Gamma}_t^{2k})_{t \geq 0}$ is a copy of the environment coupling with starting point v_k. As in the earlier constructions, we lift to processes $(\Xi_t^{jk,*})_{t \geq 0}$ in the signed spaces $V_{jk}^* = V_{jk}^- \cup V_{jk}^+ = V_{jk} \times \{-1, 1\}$, initiating with a particle (v_1, v_2) in V_{00}^+ and with the 'v_*' particles switching signs. Then, for $j, k = 1, 2$ the associated process $(\Lambda_t^{jk,*})_{t \geq 0}$ in V^* is a linearized Kac process with environment $(\rho_t^j)_{t \geq 0}$ starting from v_k.

LEMMA 6.3. *For all $p \in [2, \infty)$, there is a constant $C(p) < \infty$ such that*

$$\begin{aligned} &E_{(0, v_1, v_2)} \langle 1 + |v|^p, \Xi_t^{11} + \Xi_t^{12} + \Xi_t^{21} + \Xi_t^{22} \rangle \\ &\quad \leq C(p) \kappa \big(1 + |v_1|^{p+1} + |v_2|^{p+1}\big) |v_1 - v_2| \exp\left\{ C(p) \int_0^t \bar{m}_{p+2}(s)\, ds \right\} \\ &\qquad \times \int_0^t d_{p+1}(s)\, ds. \end{aligned}$$

PROOF. It will suffice by symmetry to consider $\langle 1 + |v|^p, \Xi_t^{11} \rangle$. The process $\langle 1 + |v|^p, \Xi_t^{11} \rangle$ makes jumps due to uncoupling transitions from V_{01} and V_{10} and also directly from V_{00}, and it makes further jumps due to the branching of particles in V_{11}. Jumps of $3 + |v_1'|^p + |v_{1*}'|^p + |v_*|^p$ occur at rate

$$\begin{aligned} &2B(v_1 - v_*, d\sigma) \Xi_{t-}^{01}(dv_1) \big(\rho_t^1 - \rho_t^2\big)^+ (dv_*)\, dt \\ &\quad + 2\big(B(v_1 - v_*, d\sigma) - B(v_2 - v_*, d\sigma)\big)^+ \Xi_{t-}^{10}(dv_1, dv_2) \rho_t^1(dv_*)\, dt \\ &\quad + 2\big(B(v_1 - v_*, d\sigma) - B(v_2 - v_*, d\sigma)\big)^+ \Xi_{t-}^{00}(dv_1, dv_2) \big(\rho_t^1 - \rho_t^2\big)^+ (dv_*)\, dt. \end{aligned}$$

Jumps of $1+|v_1|^p$ occur at rate

$$2B(v_1-v_*,d\sigma)\,\Xi^{01}_{t-}(dv_1)\big(\rho^1_t-\rho^2_t\big)^-(dv_*)\,dt$$
$$+2\big(B(v_1-v_*,d\sigma)-B(v_2-v_*,d\sigma)\big)^-\,\Xi^{10}_{t-}(dv_1,dv_2)\rho^1_t(dv_*)\,dt$$
$$+2\big(B(v_1-v_*,d\sigma)-B(v_2-v_*,d\sigma)\big)^-\,\Xi^{00}_{t-}(dv_1,dv_2)\big(\rho^1_t-\rho^2_t\big)^+(dv_*)\,dt.$$

Jumps of $2+|v'|^p+|v'_*|^p+|v_*|^p-|v|^p$ occur at rate

$$2B(v-v_*,d\sigma)\,\Xi^{11}_{t-}(dv)\rho^1_t(dv_*)\,dt.$$

Fix a starting point (v_1,v_2) in V_{00}. Recall that Ξ^{00}_t and Ξ^{10}_t are supported on pairs (u_1,u_2) with $|u_1-u_2|\le|v_1-v_2|$. We use the inequalities

$$3+\big|v'\big|^p+\big|v'_*\big|^p+|v_*|^p\le C(p)\big(1+|v|^p+|v_*|^p\big) \tag{39}$$

and

$$\big(1+|v|^p\big)|v-v_*|\le\big(1+|v|^p+|v_*|^p\big)|v-v_*|\le C(p)\big(1+|v|^{p+1}\big)\big(1+|v_*|^{p+1}\big)$$

to see that the drift of $\langle 1+|v|^p,\,\Xi^{11}_t\rangle$ due to uncoupling transitions from V_{01} is no greater than

$$C(p)\big\langle 1+|v|^{p+1},\,\Xi^{01}_{t-}\big\rangle d_{p+1}(t).$$

We use (39) and the inequalities

$$1+|v|^p\le 1+|v|^p+|v_*|^p\le\big(1+|v|^p\big)\big(1+|v_*|^p\big)$$

to see that the drift of $\langle 1+|v|^p,\,\Xi^{11}_t\rangle$ due to uncoupling transitions from V_{10} is no greater than

$$C(p)\kappa|v_1-v_2|\big\langle 1+|v|^p,\,\Xi^{10}_{t-}\circ p_{11}^{-1}\big\rangle\bar m_p(t)$$

while the drift of $\langle 1+|v|^p,\,\Xi^{11}_t\rangle$ due to uncoupling transitions from V_{00} is no greater than

$$C(p)\kappa|v_1-v_2|\big\langle 1+|v|^p,\,\Xi^{00}_{t-}\circ p_{01}^{-1}\big\rangle d_p(t).$$

Finally, by (19) and (20), the drift of $\langle 1+|v|^p,\,\Xi^{11}_t\rangle$ due to branching in V_{11} is no greater than

$$C(p)\big\langle 1+|v|^p,\,\Xi^{11}_{t-}\big\rangle\bar m_{p+1}(t).$$

Set

$$E_{p+2}(t)=C(p)\exp\left\{\int_0^t C(p)\bar m_{p+2}(s)\,ds\right\},$$

where $C(p)<\infty$ remains to be chosen. By Lemma 4.5, we can choose $C(p)$ so that

$$E_{(0,v_1,v_2)}\big\langle 1+|v|^{p+1},\,\Xi^{01}_t\big\rangle\le E_{(0,v_1,v_2)}\big\langle 1+|v|^{p+1},\,\Gamma^{11}_t\big\rangle$$
$$\le\kappa t\big(1+|v_1|^{p+1}+|v_2|^{p+1}\big)E_{p+2}(t)|v_1-v_2|.$$

By Lemma 6.2, we can choose $C(p)$ so that, moreover,

$$\begin{aligned}
&E_{(0,v_1,v_2)}\langle 1+|v|^p, \Xi_t^{10}\circ p_{11}^{-1}\rangle \\
&\quad\le E_{(0,v_1)}\langle 1+|v|^p, \hat\Gamma_t^{11}\rangle \\
&\quad\le \bigl(1+|v_1|^{p+1}\bigr)E_{p+2}(t)\int_0^t d_{p+1}(s)\,ds.
\end{aligned}$$

By Proposition 4.1, we can choose $C(p)$ so that, moreover,

$$E_{(0,v_1,v_2)}\langle 1+|v|^p, \Xi_t^{00}\circ p_{01}^{-1}\rangle \le E_{(0,v_1)}\langle 1+|v|^p, \Lambda_t^{11}\rangle \le \bigl(1+|v_1|^p\bigr)E_{p+2}(t).$$

Set $g(t)=E_{(0,v_1,v_2)}\langle 1+|v|^p, \Xi_t^{11}\rangle$. The three estimates just obtained give us control of the expected drift of $\langle 1+|v|^p, \Xi_t^{11}\rangle$, so we obtain a constant $C(p)<\infty$ such that

$$\begin{aligned}
g(t) &\le C(p)\kappa\bigl(1+|v_1|^{p+1}+|v_2|^{p+1}\bigr)|v_1-v_2|E_{p+2}(t)\int_0^t d_{p+1}(s)\,ds \\
&\quad + C(p)\int_0^t \bar m_{p+1}(s)g(s)\,ds,
\end{aligned}$$

which gives the claimed inequality by Gronwall's lemma. $\square$

PROOF OF PROPOSITION 6.1. It will suffice to consider the case where $s=0$. Set $f_0^j=E_{0t}^j f$. We use the coupling of linearized Kac processes for environments $(\rho_t^1)_{t\ge 0}$ and $(\rho_t^2)_{t\ge 0}$ described above. By Lemma 6.2,

$$\begin{aligned}
&\bigl|f_0^1(v)-f_0^2(v)\bigr| \\
&\quad= \bigl|E_{(0,v)}\langle f, \tilde\Lambda_t^1-\tilde\Lambda_t^2\rangle\bigr| \le E_{(0,v)}\langle |f|, \hat\Gamma_t^1+\hat\Gamma_t^2\rangle \le E_{(0,v)}\langle 1+|v|^p, \hat\Gamma_t^1+\hat\Gamma_t^2\rangle \qquad (40)\\
&\quad\le C(p)\bigl(1+|v|^{p+1}\bigr)\exp\biggl\{\int_0^t C(p)\bar m_{p+2}(s)\,ds\biggr\}\int_0^t d_{p+1}(s)\,ds.
\end{aligned}$$

Assume now that $f\in\mathcal F(p)$. Then

$$\bigl|f(v_1)-f(v_2)\bigr| \le p\bigl(1+|v_1|^p+|v_2|^p\bigr)|v_1-v_2|$$

for all $v_1,v_2\in\mathbb R^d$. On the other hand, if g satisfies this inequality, together with $|g(v)|\le 1+|v|^p$, then $\|g\|_{(p)}\le 3p$. We now use the coupling of four linearized Kac processes for the two environments and two starting points v_1 and v_2. For $j=1,2$, the measure $\Xi_t^{j0,*}$ is supported on pairs (u_1,u_2) with $|u_1-u_2|\le|v_1-v_2|$. So

$$\begin{aligned}
\langle f\circ p_1-f\circ p_2, \Xi_t^{j0,*}\rangle &\le p|v_1-v_2|\langle 1+|u_1|^p+|u_2|^p, \Xi_t^{j0}\rangle \\
&\le p|v_1-v_2|\langle 1+|v|^p, \hat\Gamma_t^{j1}+\hat\Gamma_t^{j2}\rangle.
\end{aligned}$$

By Lemmas 6.2 and 6.3,

$$\begin{aligned}(f_0^1 - f_0^2)(v_1) - (f_0^1 - f_0^2)(v_2) &= E_{(0,v_1,v_2)}\langle f, \tilde\Lambda_t^{11} - \tilde\Lambda_t^{12} - \tilde\Lambda_t^{21} + \tilde\Lambda_t^{22}\rangle \\ &= E_{(0,v_1,v_2)}\big(\langle f, \Xi_t^{11,*} - \Xi_t^{12,*} - \Xi_t^{21,*} + \Xi_t^{22,*}\rangle \\ &\qquad + \langle f\circ p_1 - f_2 \circ p_2, \Xi_t^{10,*} - \Xi_t^{20,*}\rangle\big) \\ &\le E_{(0,v_1,v_2)}\big(\langle 1+|v|^p, \Xi_t^{11} + \Xi_t^{12} + \Xi_t^{21} + \Xi_t^{22}\rangle \\ &\qquad + p|v_1 - v_2|\langle 1+|v|^p, \hat\Gamma_t^{11} + \hat\Gamma_t^{12} + \hat\Gamma_t^{21} + \hat\Gamma_t^{22}\rangle\big) \\ &\le C(p)\kappa\big(1+|v_1|^{p+1} + |v_2|^{p+1}\big) \\ &\quad \times |v_1 - v_2| \exp\left\{\int_0^t C(p)\bar m_{p+2}(s)\,ds\right\}\int_0^t d_{p+1}(s)\,ds.\end{aligned}$$

Here, there are no terms in $\Xi_t^{0k,*}$ for $k = 0, 1, 2$ because these are the empirical distributions of particles, or pairs of particles, with unbroken environment coupling, which cancel completely in the considered integral. On combining this estimate with (40), we deduce that

$$\|f_0^1 - f_0^2\|_{(p+1)} \le 3C(p)\kappa \exp\left\{\int_0^t C(p)\bar m_{p+2}(s)\,ds\right\}\int_0^t d_{p+1}(s)\,ds. \qquad \square$$

7. Maximal inequalities for stochastic convolutions. The key formula for our analysis is shown in Proposition 4.2. For all $t \ge 0$ and all functions f in our weighted Lipschitz class $\mathcal{F}$, we have

$$\langle f, \mu_t^N - \mu_t^{N'}\rangle = \langle f_{0t}, \mu_0^N - \mu_0^{N'}\rangle + \int_0^t \langle f_{st}, dM_s^N\rangle - \int_0^t \langle f_{st}, dM_s^{N'}\rangle,$$

where

$$f_{st}(v) = E_{st} f(v) = E_{(s,v)}\langle f, \tilde\Lambda_t\rangle$$

and where $(\Lambda_t^*)_{t\ge s}$ is the linearized Kac process in environment $((\mu_t^N + \mu_t^{N'})/2)_{t\ge 0}$ starting from v at time s.

The notion of stochastic convolution has been extensively studied in connection with infinite-dimensional stochastic evolution equations; see, for example, [3, 7]. The operator

$$f \mapsto \int_0^t \langle E_{st} f, dM_s^N\rangle$$

shares some features with stochastic convolutions, namely that $E_{st}E_{tu} = E_{su}$ for $s \le t \le u$ and that good estimates rely on exploiting martingale properties of the integrator. In this section, we prove a maximal inequality for this operator in Wasserstein norms, in the case where the environment $((\mu_t^N + \mu_t^{N'})/2)_{t\ge 0}$ is replaced by a nonrandom process $(\rho_t)_{t\geq 0}$. The proof of Proposition 9.2 below uses some of the same ideas in a simpler context.

We will use the following inequality for a function f on $\mathbb{R}^d$ which is Lipschitz of constant 1. For $B = [0, 2^{-k}]^d$, we have

$$|f(v) - \langle f\rangle_B| \leq 2^{-k}c_d, \qquad v \in B, \tag{41}$$

where $\langle f\rangle_B$ is the average value of f on B and where $c_d = \mathbb{E}|X|$, with X uniformly distributed on $[0, 1]^d$. To see this, set $Y = 2^{-k}X$, and note that $|f(v) - f(Y)| \leq |v - Y| \leq |Y|$ so $|f(v) - \langle f\rangle_B| = |\mathbb{E}(f(v) - f(Y))| \leq \mathbb{E}|Y| = 2^{-k}\mathbb{E}|X|$. By a similar calculation, we have also

$$|\langle f\rangle_B - \langle f\rangle_{2B}| \leq 2^{-k}c_d. \tag{42}$$

It is the scaling properties of inequalities (41) and (42) which will be critical for our argument, rather than the value of the constant c_d.

Let $(\rho_t)_{t\geq 0}$ be a nonrandom process[7] satisfying (17). Write, as above, $m_p(t) = \langle 1 + |v|^p, \rho_t\rangle$, and set

$$m^*(p) = \sup_{t\geq 0} m_p(t).$$

Then, by Proposition 4.1, for all $s \geq 0$ and all $v_0 \in \mathbb{R}^d$, and for the linearized Kac process $(\Lambda_t^*)_{t\geq s}$ in environment $(\rho_t)_{t\geq 0}$ starting from v_0 at time s, we have

$$E_{(s,v_0)}\langle 1 + |v|^p, \Lambda_t\rangle \leq (1 + |v_0|^p)e^{c(p)m^*(p+1)(t-s)}.$$

Thus, whenever $m^*(p + 1) < \infty$, we can define, for $s \leq t$ and $f \in \mathcal{F}(p)$,

$$f_{st}(v) = E_{st} f(v) = E_{(s,v)}\langle f, \tilde{\Lambda}_t\rangle.$$

PROPOSITION 7.1. *For all $d \geq 3$, $p \in [2, \infty)$ and all $\delta \in (0, 1]$, there is a constant $C(d, \delta, p) < \infty$ such that, for all $T \in [0, \infty)$, we have*

$$\begin{aligned} &\left\| \sup_{t\leq T} \sup_{f\in\mathcal{F}(p)} \int_0^t \langle f_{st}, dM_s^N\rangle \right\|_2 \\ &\quad \leq C\kappa N^{-1/d} e^{Cm^*(p+3+\delta)T} \left(\mathbb{E}\int_0^T \langle |v|^{2p+5+2\delta}, \mu_s^N\rangle\, ds\right)^{1/2}. \end{aligned} \tag{43}$$

The same inequality holds for $d = 2$ if we replace $N^{-1/d}$ by $N^{-1/2}\log N$.

Here we have written $\|\cdot\|_2$ for the norm in $L^2(\mathbb{P})$. This estimate will be applied in the next section, using the moment estimates derived in Section 3 to control

[7] We will in fact use only the case where $(\rho_t)_{t\geq 0}$ is constant.

the right-hand side. We will use also the following comparison estimate for two nonrandom processes $(\rho_t^1)_{t\geq 0}$ and $(\rho_t^2)_{t\geq 0}$ satisfying (17). Fix $p \in [2, \infty)$. Write

$$\bar{m}^*(p) = \sup_{t\geq 0}\langle 1 + |v|^p, \rho_t^1 + \rho_t^2\rangle.$$

We assume that $\bar{m}^*(p+1) < \infty$. For $j = 1, 2$ and $f \in \mathcal{F}(p)$, define for $s, t \geq 0$ with $s \leq t$

$$f_{st}^j(v) = E_{st}^j f(v) = E_{(s,v)}\langle f, \tilde{\Lambda}_t^j\rangle,$$

where $(\Lambda_t^{j,*})_{t\geq s}$ is a linearized Kac process in environment $(\rho_t^j)_{t\geq 0}$ starting from v.

PROPOSITION 7.2. *For all $d \geq 3$, $p \in [2, \infty)$ and all $\delta \in (0, 1]$, there is a constant $C(d, \delta, p) < \infty$ such that, for all $T \in [0, \infty)$, we have*

$$\begin{aligned}
&\left\|\sup_{t\leq T}\sup_{f\in\mathcal{F}(p)}\int_0^t \langle f_{st}^1 - f_{st}^2, dM_s^N\rangle\right\|_2 \\
&\quad\leq C\kappa N^{-1/d}\bar{m}^*(p+2+\delta) \qquad (44)\\
&\qquad\times T e^{C\bar{m}^*(p+3+\delta)T}\left(\mathbb{E}\int_0^T \langle |v|^{2p+5+2\delta}, \mu_s^N\rangle\, ds\right)^{1/2}.
\end{aligned}$$

The same inequality holds for $d = 2$ if we replace $N^{-1/d}$ by $N^{-1/2}\log N$.

A small variation of the following proofs would allow the insertion of a factor of $d^*(p+3+\delta)$ on the right in (44), where $d^*(p) = \sup_{t\geq 0}\langle 1 + |v|^p, |\rho_t^1 - \rho_t^2|\rangle$, at the cost of replacing p in all other terms on the right by $p+1$. We omit details as this variation is not needed for our main result.

PROOF OF PROPOSITION 7.1. Assume for now that $d \geq 3$. It will suffice to consider the case where $N \geq 2^{2d}$. We first prove a simpler estimate, where the function f_{st} is replaced by f_{sT} on the left-hand side. Set $L = \lfloor \log_2 N/d\rfloor$, and note that $L \geq 2$. For $k \in \mathbb{Z}$, set $B_k = (-2^k, 2^k]^d$. Set $A_0 = B_0$, and for $k \geq 1$, set $A_k = B_k \setminus B_{k-1}$. For $k \geq 1$ and any integer $\ell \geq 2$, there is a unique way to partition A_k by a set $\mathcal{P}_{k,\ell}$ of $2^{\ell d} - 2^{(\ell-1)d}$ translates of $B_{k-\ell}$. Also, there is a unique way to partition A_0 by a set $\mathcal{P}_{0,\ell}$ of $2^{\ell d}$ translates of $B_{-\ell}$. Fix $p \in [2, \infty)$ and $f \in \mathcal{F}(p)$. Then, for all $v, v' \in \mathbb{R}^d$, we have

$$|\hat{f}^{(p)}(v)| = |f(v)/(1 + |v|^p)| \leq 1, \qquad |\hat{f}^{(p)}(v) - \hat{f}^{(p)}(v')| \leq |v - v'|.$$

For $B \in \mathcal{P}_{k,2}$, set $a_B = \langle \hat{f}^{(p)}\rangle_B$, and note that $|a_B| \leq 1$. For $\ell \geq 3$ and $B \in \mathcal{P}_{k,\ell}$, set $a_B = \langle \hat{f}^{(p)}\rangle_B - \langle \hat{f}^{(p)}\rangle_{\pi(B)}$, where $\pi(B)$ is the unique element of $\mathcal{P}_{k,\ell-1}$ containing B, and note that $|a_B| \leq 2^{k-\ell+1}c_d$. Set $c'_d = 4 \vee (2c_d)$, then $|a_B| \leq 2^{k-\ell}c'_d$ for all $B \in \mathcal{P}_{k,\ell}$, for all $k \geq 0$ and all $\ell \geq 2$. Fix $\delta \in (0, 1]$, and for $B \in \mathcal{P}_{k,\ell}$, set

$$h^B(v) = 2^{(1+\delta)k}(1 + |v|^p)1_B(v).$$

Define a function g_k, supported on A_k, by

$$f 1_{A_k} = \sum_{\ell=2}^{L} \sum_{B \in \mathcal{P}_{k,\ell}} a_B \big(1 + |v|^p\big) 1_B(v) + g_k.$$

Fix $K \in \mathbb{N}$, and set $g = \sum_{k=0}^{K} g_k + f 1_{B_K^c}$. Note that $\hat{g}_k^{(p)} = \hat{f}^{(p)} - \langle \hat{f}^{(p)} \rangle_B$ on B for all $B \in \mathcal{P}_{k,L}$. For $v \in A_k$, we have $|v| \geq 2^{k-1}$, so

$$\big|\hat{g}_k^{(p)}(v)\big| \leq 2^{k+1-L} c_d \leq 2^{-L+2} c_d \big(1 + |v|\big).$$

For $v \in B_K^c$, we have $|v| \geq 2^{K-1}$, so $|\hat{f}^{(p)}(v)| \leq 2^{-K+1}|v|$. Set $c_d'' = (8c_d) \vee 4$. Then, for all $v \in \mathbb{R}^d$, we have

$$\big|g(v)\big| \leq \big\{2^{-L+2} c_d \big(1 + |v|\big) + 2^{-K+1}|v|\big\}\big(1 + |v|^p\big) \leq \big(2^{-K} + 2^{-L}\big) c_d'' \big(1 + |v|^{p+1}\big).$$

Now

$$f = \sum_{\ell=2}^{L} \sum_{k=0}^{K} \sum_{B \in \mathcal{P}_{k,\ell}} 2^{-(1+\delta)k} a_B h^B + g$$

so

$$\begin{aligned} &\int_0^t \langle f_{sT}, dM_s^N \rangle \\ &\quad = \sum_{\ell=2}^{L} \sum_{k=0}^{K} \sum_{B \in \mathcal{P}_{k,\ell}} 2^{-(1+\delta)k} a_B \int_0^t \langle h_{sT}^B, dM_s^N \rangle + \int_0^t \langle g_{sT}, dM_s^N \rangle, \end{aligned} \tag{45}$$

where $h_{sT}^B = E_{sT} h^B$ and $g_{sT} = E_{sT} g$. It will be convenient to set

$$E(p) = \exp\big\{c(p) m^*(p+1) T\big\}$$

and

$$c(d, \delta) = \big(1 - 2^{-2\delta}\big)^{-1/2} c_d', \qquad A = \big(2^{-K} + 2^{-L}\big) c_d'' E(p+1).$$

Note that $2^{-(1+\delta)k}|a_B| \leq 2^{-\ell - \delta k} c_d'$ for all $B \in \mathcal{P}_{k,\ell}$, and $\mathcal{P}_{k,\ell}$ has cardinality at most $2^{d\ell}$, so

$$\sum_{k=0}^{K} \sum_{B \in \mathcal{P}_{k,\ell}} \big(2^{-(1+\delta)k} a_B\big)^2 \leq 2^{(d-2)\ell} c(d, \delta)^2.$$

Also, by Proposition 4.1, for all $s \in [0, T]$, we have

$$\big|g_{sT}(v)\big| \leq A\big(1 + |v|^{p+1}\big).$$

We use Cauchy–Schwarz in (45) to obtain

$$
\begin{aligned}
&\sup_{t\le T}\sup_{f\in\mathcal{F}(p)}\int_0^t\langle f_{sT},dM_s^N\rangle\\
&\qquad\le\sum_{\ell=2}^{L}2^{(d/2-1)\ell}c(d,\delta)\left(\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}\sup_{t\le T}\left|\int_0^t\langle h_{sT}^B,dM_s^N\rangle\right|^2\right)^{1/2}\\
&\qquad\quad+A\int_0^T\langle 1+|v|^{p+1},|dM_s^N|\rangle.
\end{aligned}\tag{46}
$$

Set

$$
\begin{aligned}
h(v)&=\bigl(1+|v|^p\bigr)\sum_{k=0}^{K}2^{(1+\delta)k}1_{A_k}(v)\\
&=\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}h^B(v).
\end{aligned}
$$

Note that $1\vee|v|\ge 2^{k-1}$ for all $v\in A_k$ and all k. Set $q=p+1+\delta$ and $A'=8E(q)$. Then

$$
h(v)\le 2^{1+\delta}\bigl(1\vee|v|\bigr)^{1+\delta}\bigl(1+|v|^p\bigr)\le 8\bigl(1+|v|^q\bigr),
$$

so by Proposition 4.1,

$$
E_{(s,v)}\langle h,\Lambda_T\rangle\le A'\bigl(1+|v|^q\bigr).
$$

Note that $|h_{sT}^B(v)|\le E_{(s,v)}\langle h^B,\Lambda_T\rangle$, so

$$
\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}\bigl(h_{sT}^B(v)\bigr)^2\le\bigl(E_{(s,v)}\langle h,\Lambda_T\rangle\bigr)^2\le\bigl(A'\bigr)^2\bigl(1+|v|^q\bigr)^2.
$$

Hence, for some constant $C(q)<\infty$, we have

$$
\begin{aligned}
&\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}\bigl\{h_{sT}^B(v')+h_{sT}^B(v'_*)-h_{sT}^B(v)-h_{sT}^B(v_*)\bigr\}^2\\
&\quad\le 4\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}\bigl\{h_{sT}^B(v')^2+h_{sT}^B(v'_*)^2+h_{sT}^B(v)^2+h_{sT}^B(v_*)^2\bigr\}\\
&\quad\le 4\bigl(A'\bigr)^2E(q)^2\bigl\{\bigl(1+|v'|^q\bigr)^2+\bigl(1+|v'_*|^q\bigr)^2+\bigl(1+|v|^q\bigr)^2+\bigl(1+|v_*|^q\bigr)^2\bigr\}\\
&\quad\le C(q)\bigl(A'\bigr)^2\bigl(1+|v|^{2q}+|v_*|^{2q}\bigr).
\end{aligned}
$$

Then, by Doob's L^2-inequality,

$$\begin{aligned}
&\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}\mathbb{E}\Big(\sup_{t\le T}\Big|\int_0^t\langle h_{sT}^B, dM_s^N\rangle\Big|^2\Big)\\
&\quad=\frac{4}{N}\sum_{k=0}^{K}\sum_{B\in\mathcal{P}_{k,\ell}}\mathbb{E}\int_0^T\int\{h_{sT}^B(v')+h_{sT}^B(v'_*)-h_{sT}^B(v)-h_{sT}^B(v_*)\}^2\\
&\qquad\qquad\qquad\qquad\qquad\times B(v-v_*,d\sigma)\mu_{s-}^N(dv)\mu_{s-}^N(dv_*)\,ds\\
&\quad\le\frac{C(q)(A')^2}{N}\mathbb{E}\int_0^T\int\{1+|v|^{2q}+|v_*|^{2q}\}|v-v_*|\mu_{s-}^N(dv)\mu_{s-}^N(dv_*)\,ds\\
&\quad\le\frac{C(q)(A')^2}{N}\mathbb{E}\int_0^T\langle|v|^{2q+1},\mu_s^N\rangle ds.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\int_0^T\langle 1+|v|^{p+1},|dM_s^N|\rangle\\
&\qquad\le\int_0^T\int\{4+|v'|^{p+1}+|v'_*|^{p+1}+|v|^{p+1}+|v_*|^{p+1}\}\\
&\qquad\qquad\times(m+\bar m)(dv,dv_*,d\sigma,ds),
\end{aligned}$$

where the measures m and $\bar m$ are as defined in Section 2. We split the integral using $m+\bar m=(m-\bar m)+2\bar m$ and use the L^2-isometry for integrals with respect to the compensated measure $m-\bar m$ to obtain

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\int_0^T\langle 1+|v|^{p+1},|dM_s^N|\rangle\Big|^2\Big)\\
&\qquad\le C(p)\mathbb{E}\int_0^T\int\{1+|v|^{p+1}+|v_*|^{p+1}\}^2\,d\bar m\\
&\qquad\quad+C(p)\mathbb{E}\Big(\Big|\int_0^T\int\{1+|v|^{p+1}+|v_*|^{p+1}\}\,d\bar m\Big|^2\Big)\\
&\qquad\le C(p)\mathbb{E}\int_0^T\langle|v|^{2p+3},\mu_s^N\rangle ds+C(p)\mathbb{E}\Big|\int_0^T\langle|v|^{p+2},\mu_s^N\rangle ds\Big|^2\\
&\qquad\le C(p)(1+T)\mathbb{E}\int_0^T\langle|v|^{2p+3},\mu_s^N\rangle ds,
\end{aligned}\tag{47}$$

where the constant $C(p)<\infty$ varies from line to line. In the final inequality, we dealt with the second term on the right by writing $|v|^{p+2}=|v||v|^{p+1}$, applying Cauchy–Schwarz and then using the fact that $\langle 1,\mu_t^N\rangle=\langle|v|^2,\mu_t^N\rangle=1$. We take

L^2-norms in (46) to obtain

$$\begin{aligned}
&\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{sT}, dM_s^N \rangle \right\|_2 \\
&\quad \le \left(\sum_{\ell=2}^{L} 2^{(d/2-1)\ell} c(d,\delta) A' \left(\frac{C(q)}{N} \right)^{1/2} + A\big(C(p)(1+T)\big)^{1/2} \right) \\
&\qquad \times \left(\mathbb{E} \int_0^T \langle |v|^{2q+1}, \mu_s^N \rangle \, ds \right)^{1/2}.
\end{aligned}$$

Recall that $A = (2^{-K} + 2^{-L}) c''_d E(p+1)$, $A' = 8E(q)$, $L = \lfloor \log_2 N/d \rfloor$ and $q = p + 1 + \delta$. Note that $2^{(d/2-1)L} N^{-1/2} \le N^{-1/d}$ and $2^{-L} \le 2N^{-1/d}$. Hence, on letting $K \to \infty$, we deduce that, for some constant $C(d, \delta, p) < \infty$,

$$\begin{aligned}
&\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{sT}, dM_s^N \rangle \right\|_2 \\
&\quad \le C N^{-1/d} e^{Cm^*(p+2+\delta)T} \left(\mathbb{E} \int_0^T \langle |v|^{2p+3+2\delta}, \mu_s^N \rangle \, ds \right)^{1/2}.
\end{aligned} \tag{48}$$

This is not the inequality (43) we seek because f_{sT} rather than f_{st} appears on the right-hand side. However, it will prove to be a useful first step.

We now turn to the proof of (43). It will suffice to deal with the case where $T = 2^{-J_0}$ for some $J_0 \in \mathbb{Z}$. Set $\tau_j(t) = 2^{-j} \lceil 2^j t \rceil$. Then, for all $t \in (0, T]$, we have $\tau_{J_0}(t) = T$ so, for $J \ge J_0$ and $s \in [0, t]$,

$$f_{st} = f_{sT} + \sum_{j=J_0+1}^{J} (f_{s\tau_j(t)} - f_{s\tau_{j-1}(t)}) + (f_{st} - f_{s\tau_J(t)})$$

and hence

$$\begin{aligned}
&\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{st}, dM_s^N \rangle \right\|_2 \\
&\quad \le \left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{sT}, dM_s^N \rangle \right\|_2 \\
&\qquad + \sum_{j=J_0+1}^{J} \left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{s\tau_j(t)} - f_{s\tau_{j-1}(t)}, dM_s^N \rangle \right\|_2 \\
&\qquad + \left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{st} - f_{s\tau_J(t)}, dM_s^N \rangle \right\|_2.
\end{aligned} \tag{49}$$

Fix $j \ge J_0 + 1$ and, for $i = 0, 1, \dots, 2^j T$, set $t_i = i 2^{-j}$. Note that, for $t \in (t_{i-1}, t_i]$, we have

$$f_{s\tau_j(t)} - f_{s\tau_{j-1}(t)} = \begin{cases} f_{st_i} - f_{st_{i+1}}, & \text{if } i \text{ is odd}, \\ 0, & \text{if } i \text{ is even}. \end{cases}$$

Set $f^{(i)} = f - f_{t_i t_{i+1}}$. We can take $s = t_i$, $t = t_{i+1}$ and $\rho_r^1 = \rho_r$, $\rho_r^2 = 0$ for all r in Proposition 6.1 to obtain

$$\|f^{(i)}\|_{(p+1)} \le A'' 2^{-j},$$

where

$$A'' = C(p)\kappa e^{C(p)m^*(p+2)T} m^*(p+1).$$

For $s \in [0, t_i]$, set $f_s^{(i)} = E_{st_i} f^{(i)} = f_{st_i} - f_{st_{i+1}}$. Write

$$X_{ij} = \sup_{t\in(t_{i-1},t_i]} \sup_{f\in\mathcal{F}(p)} \int_0^t \langle f_s^{(i)}, dM_s^N\rangle$$

and note that

$$\sup_{t\le T} \sup_{f\in\mathcal{F}(p)} \int_0^t \langle f_{s\tau_j(t)} - f_{s\tau_{j-1}(t)}, dM_s^N\rangle \le \sup_{i\le 2^j T} X_{ij}.$$

Set

$$A''' = CA''N^{-1/d} e^{Cm^*(p+3+\delta)T} \left(\mathbb{E}\int_0^T \langle |v|^{2p+5+2\delta}, \mu_s^N\rangle\, ds\right)^{1/2},$$

where C is the constant $C(d, \delta, p+1)$ from (48). We replace T by t_i, p by $p+1$ and f by $f^{(i)}/\|f^{(i)}\|_{(p+1)}$ in (48) to see that

$$\|X_{ij}\|_2 \le 2^{-j} A'''.$$

Then

$$\sum_{j=J_0+1}^{J} \Big\| \sup_{i\le 2^j T} X_{ij} \Big\|_2 \le \sum_{j=J_0+1}^{J} (2^j T)^{1/2} 2^{-j} A''' \le 3TA''', \tag{50}$$

so there is a constant $C(d, \delta, p) < \infty$ such that

$$\begin{aligned} &\sum_{j=J_0+1}^{J} \left\| \sup_{t\le T} \sup_{f\in\mathcal{F}(p)} \int_0^t \langle f_{s\tau_j(t)} - f_{s\tau_{j-1}(t)}, dM_s^N\rangle \right\|_2 \\ &\qquad \le Cm^*(p+1)T\kappa N^{-1/d} e^{Cm^*(p+3+\delta)T} \left(\mathbb{E}\int_0^T \langle |v|^{2p+5+2\delta}, \mu_s^N\rangle\, ds\right)^{1/2}. \end{aligned} \tag{51}$$

Finally, we can take $t = \tau_J(t)$, $\rho_r^1 = \rho_r$ and $\rho_r^2 = \rho_r 1_{\{r\le t\}}$ for all r in Proposition 6.1 to obtain

$$\|f_{st} - f_{s\tau_J(t)}\|_{(p+1)} \le A'' 2^{-J}$$

for all $s \le t \le T$. Hence

$$\int_0^t \langle f_{st} - f_{s\tau_J(t)}, dM_s^N\rangle \le 2^{-J} A'' \mathbb{E}\int_0^T \langle 1 + |v|^{p+1}, |dM_s^N|\rangle,$$

so estimate (47) shows that, as $J \to \infty$,

$$\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle f_{st} - f_{s\tau_J(t)}, dM_s^N \rangle \right\|_2 \to 0.$$

Hence (43) follows from (48), (49) and (51). The proof is the same for $d = 2$ except that we get $N^{-1/2} \log_2 N$ in place of $N^{-1/d}$ in (48) and (51). $\square$

PROOF OF PROPOSITION 7.2. Fix $p \in [2, \infty)$ and $f \in \mathcal{F}(p)$. We follow the preceding proof to obtain, for $t \le T$,

$$\begin{aligned} &\int_0^t \langle \tilde{f}_{sT}, dM_s^N \rangle \\ &\quad = \sum_{\ell=2}^{L} \sum_{k=0}^{K} \sum_{B \in \mathcal{P}_{k,\ell}} 2^{-(1+\delta)k} a_B \int_0^t \langle \tilde{h}_{sT}^B, dM_s^N \rangle + \int_0^t \langle \tilde{g}_{sT}, dM_s^N \rangle, \end{aligned} \tag{52}$$

where $\tilde{f}_{sT} = (E_{sT}^1 - E_{sT}^2) f$, $\tilde{h}_{sT}^B = (E_{sT}^1 - E_{sT}^2) h^B$ and $\tilde{g}_{sT} = (E_{sT}^1 - E_{sT}^2) g$. By Proposition 6.1, we have

$$|\tilde{g}_{sT}(v)| \le \tilde{A}(1 + |v|^{p+2}),$$

where

$$\tilde{A} = (2^{-K} + 2^{-L}) c_d'' C T d^*(p+2) e^{C\bar{m}^*(p+3)T}$$

and $C = C(p+1) < \infty$. Note that

$$|\tilde{h}_{sT}^B(v)| = |E_{(s,v)} \langle h^B, \tilde{\Lambda}_T^1 - \tilde{\Lambda}_T^2 \rangle| \le E_{(s,v)} \langle h^B, \hat{\Gamma}_T^1 + \hat{\Gamma}_T^2 \rangle,$$

so by Lemma 6.2,

$$\sum_{k=0}^{K} \sum_{B \in \mathcal{P}_{k,\ell}} (\tilde{h}_{sT}^B(v_0))^2 \le 64 (E_{(s,v_0)} \langle 1 + |v|^q, \hat{\Gamma}_T^1 + \hat{\Gamma}_T^2 \rangle)^2 \le (\tilde{A}')^2 (1 + |v_0|^{q+1})^2,$$

where

$$\tilde{A}' = C T d^*(q+1) e^{C\bar{m}^*(q+2)T}$$

and $C = C(q+1) < \infty$. We continue to follow the steps of the preceding proof to arrive at

$$\begin{aligned} &\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle \tilde{f}_{sT}, dM_s^N \rangle \right\|_2 \\ &\quad \le \left(\sum_{\ell=2}^{L} 2^{(d/2-1)\ell} c(d,\delta) \tilde{A}' \left(\frac{C}{N} \right)^{1/2} + \tilde{A} (C(1+T))^{1/2} \right) \\ &\qquad \times \left(\mathbb{E} \int_0^T \langle |v|^{2q+3}, \mu_s^N \rangle \, ds \right)^{1/2}. \end{aligned}$$

Replace $\tilde{A}$, $\tilde{A}'$, L and q by their values and let $K \to \infty$ to deduce that, for some constant $C(d, \delta, p) < \infty$,

$$\begin{aligned}
&\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle \tilde{f}_{sT}, dM_s^N \rangle \right\|_2 \\
&\quad \le C T d^*(p + 2 + \delta) N^{-1/d} e^{C\bar{m}^*(p+3+\delta)T} \\
&\qquad \times \left(\mathbb{E} \int_0^T \langle |v|^{2p+5+2\delta}, \mu_s^N \rangle \, ds \right)^{1/2}.
\end{aligned} \tag{53}$$

Now

$$\begin{aligned}
&\left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle \tilde{f}_{st}, dM_s^N \rangle \right\|_2 \\
&\quad \le \left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle \tilde{f}_{sT}, dM_s^N \rangle \right\|_2 \\
&\qquad + \sum_{j=J_0+1}^{J} \left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle \tilde{f}_{s\tau_j(t)} - \tilde{f}_{s\tau_{j-1}(t)}, dM_s^N \rangle \right\|_2 \\
&\qquad + \left\| \sup_{t \le T} \sup_{f \in \mathcal{F}(p)} \int_0^t \langle \tilde{f}_{st} - \tilde{f}_{s\tau_J(t)}, dM_s^N \rangle \right\|_2,
\end{aligned} \tag{54}$$

and the final term tends to 0 as $J \to \infty$. We consider the case where $\rho_t^1 = \rho_t$ and $\rho_t^2 = 0$ for all t, from which the general case follows by the triangle inequality. Then $f_{st}^2 = f$ for all s and t, so $\tilde{f}_{s\tau_j(t)} - \tilde{f}_{s\tau_{j-1}(t)} = f_{s\tau_j(t)} - f_{s\tau_{j-1}(t)}$. We then use (53) for the first term on the right in (54), use (51) for the sum over j and let $J \to \infty$ to obtain the claimed estimate. □

8. Proof of Theorem 1.2. We seek to show that, for $p > 8$ and $\varepsilon > 0$, for $N \le N'$ and any two Kac processes $(\mu_t^N)_{t \ge 0}$ and $(\mu_t^{N'})_{t \ge 0}$ with collision kernel B, which are adapted to a common filtration $(\mathcal{F}_t)_{t \ge 0}$, with probability exceeding $1 - \varepsilon$, for all $t \in [0, T]$, we have

$$W(\mu_t^N, \mu_t^{N'}) = \sup_{f \in \mathcal{F}} \langle f, \mu_t^N - \mu_t^{N'} \rangle \le C\big(W(\mu_0^N, \mu_0^{N'}) + N^{-1/d}\big)$$

for some constant $C < \infty$ depending only on $B, d, \varepsilon, \lambda, p$ and T, where λ is an upper bound for $\langle |v|^p, \mu_0^N \rangle$ and $\langle |v|^p, \mu_0^{N'} \rangle$. Recall the representation formula of Proposition 4.2. For all $f \in \mathcal{F}$, we have

$$\langle f, \mu_t^N - \mu_t^{N'} \rangle = \langle f_{0t}, \mu_0^N - \mu_0^{N'} \rangle + \int_0^t \langle f_{st}, dM_s^N \rangle - \int_0^t \langle f_{st}, dM_s^{N'} \rangle,$$

where M^N is given by (4) and $f_{st}(v) = E_{st} f(v) = E_{(s,v)} \langle f, \tilde{\Lambda}_t \rangle$, with $(\Lambda_t^*)_{t \ge s}$ a linearized Kac process in environment $\rho_t = (\mu_t^N + \mu_t^{N'})/2$. We showed a suitable bound for $\langle f_{0t}, \mu_0^N - \mu_0^{N'}\rangle$ in Section 5. We now show that the stochastic convolution estimates just obtained allow us to control $\int_0^t \langle f_{st}, dM_s^N\rangle$ with rate $N^{-1/d}$, notwithstanding the fact that the functions f_{st} depend on the random environment $(\rho_r)_{r\in[s,t]}$ and therefore are anticipating.

It will suffice to consider the case where $p \in (8, 9]$ and $T = 2^{-J_0}$ for some $J_0 \in \mathbb{Z}$. Set $\delta = (p-8)/6$. Set $\sigma_j(t) = 2^{-j}\lfloor 2^j t\rfloor$, and note that $\sigma_{J_0}(t) = 0$ for all $t < T$. Set $\rho_t^j = \rho_{\sigma_j(t)}$, and define

$$E_{st}^j f(v) = E_{(s,v)}\langle f, \tilde{\Lambda}_t^j\rangle,$$

where $(\Lambda_t^{j,*})_{t\geq s}$ is a linearized Kac process in environment $(\rho_t^j)_{t\geq 0}$ starting from v. Then, for $t \leq T$ and $J \geq J_0$, we have

$$\begin{aligned}\int_0^t \langle E_{st} f, dM_s^N\rangle = \int_0^t \langle E_{st}^{J_0} f, dM_s^N\rangle + \sum_{j=J_0+1}^{J} \int_0^t \langle (E_{st}^j - E_{st}^{j-1}) f, dM_s^N\rangle \\ + \int_0^t \langle (E_{st} - E_{st}^J) f, dM_s^N\rangle.\end{aligned} \tag{55}$$

Note that $\rho_t^{J_0} = \rho_0$ for all $t < T$. Take $p = 2$ in Proposition 7.1 to see that, for some constant $C(d,\delta) < \infty$, we have

$$\begin{aligned}&\left\| \sup_{t\leq T}\sup_{f\in\mathcal{F}} \int_0^t \langle E_{st}^{J_0} f, dM_s^N\rangle\right\|_2 \\ &\quad\leq C\kappa N^{-1/d} e^{C\langle 1+|v|^{5+\delta},\rho_0\rangle T}\left(\mathbb{E}\int_0^T \langle 1+|v|^{9+2\delta}, \mu_s^N\rangle\, ds\right)^{1/2}.\end{aligned} \tag{56}$$

Fix j, and set $t_i = i2^{-j}$. Note that, for $t \in [t_i, t_{i+1})$,

$$\rho_t^j - \rho_t^{j-1} = \begin{cases} 0, & \text{if } i \text{ is even},\\ \rho_{t_i} - \rho_{t_{i-1}}, & \text{if } i \text{ is odd}.\end{cases}$$

We have

$$\begin{aligned}&\int_0^t \langle (E_{st}^j - E_{st}^{j-1}) f, dM_s^N\rangle \\ &\quad= \sum_{i=0}^{\lfloor 2^j t\rfloor - 1} \int_{t_i}^{t_{i+1}} \langle (E_{st}^j - E_{st}^{j-1}) f, dM_s^N\rangle + \int_{\sigma_j(t)}^t \langle (E_{st}^j - E_{st}^{j-1}) f, dM_s^N\rangle.\end{aligned}$$

For $s \leq t_{i+1} \leq t$, we have $E_{st}^j = E_{st_{i+1}}^j E_{t_{i+1}t}^j$ so, for all $f \in \mathcal{F}$,

$$\begin{aligned}&\int_0^t \langle (E_{st}^j - E_{st}^{j-1}) f, dM_s^N\rangle \\ &\quad\leq \sum_{i=0}^{\lfloor 2^j t\rfloor - 1} \| E_{t_{i+1}t}^j f - E_{t_{i+1}t}^{j-1} f\|_{(3)} \sup_{f\in\mathcal{F}(3)} \int_{t_i}^{t_{i+1}} \langle E_{st_{i+1}}^j f, dM_s^N\rangle\end{aligned}$$

$$+\sum_{i=0}^{\lfloor 2^j t\rfloor -1}\|E^{j-1}_{t_{i+1}t}f\|\sup_{f\in\mathcal{F}}\int_{t_i}^{t_{i+1}}\langle (E^j_{st_{i+1}}-E^{j-1}_{st_{i+1}})f, dM^N_s\rangle$$
$$+\int_{\sigma_j(t)}^{t}\langle (E^j_{st}-E^{j-1}_{st})f, dM^N_s\rangle.$$

Fix $A\geq 1$, and consider the event $\Omega_0=\Omega_1\cap\Omega_2\cap\Omega_3$, where

$$\Omega_1=\Big\{\sup_{t\leq T}\langle 1+|v|^{5+\delta},\rho_t\rangle\leq A\Big\},$$
$$\Omega_2=\Big\{\int_0^T\langle 1+|v|^3, |\rho^j_t-\rho^{j-1}_t|\rangle\,dt\leq A2^{-j(1-\delta)}\text{ for all } j\geq J_0+1\Big\},$$
$$\Omega_3=\Big\{\int_0^T\langle 1+|v|^3, |\rho^J_t-\rho_t|\rangle\,dt\leq A2^{-J}\Big\}.$$

By Proposition 3.1, there is an absolute constant $C<\infty$ such that

$$\mathbb{E}\Big(\sup_{t\leq T}\langle 1+|v|^{5+\delta},\rho_t\rangle\Big)\leq C\lambda(1+T),\qquad \mathbb{E}\big(\langle 1+|v|^3,|\rho_t-\rho_s|\rangle\big)\leq C\lambda|t-s|$$

so

$$\mathbb{E}\int_0^T\langle 1+|v|^3,|\rho^j_t-\rho^{j-1}_t|\rangle\,dt\leq CT\lambda 2^{-j},$$
$$\mathbb{E}\int_0^T\langle 1+|v|^3,|\rho^J_t-\rho_t|\rangle\,dt\leq CT\lambda 2^{-J}.$$

Hence

$$\mathbb{P}(\Omega\setminus\Omega_1)\leq C\lambda(1+T)A^{-1},\qquad \mathbb{P}(\Omega\setminus\Omega_3)\leq C\lambda TA^{-1}$$

and

$$\mathbb{P}(\Omega\setminus\Omega_2)\leq\sum_{j=J_0+1}^{\infty}CT\lambda A^{-1}2^{-j\delta}=CT^{1-\delta}\lambda A^{-1}(2^\delta-1)^{-1}.$$

Hence we can choose $A(\varepsilon,\lambda,p,T)<\infty$ such that $\mathbb{P}(\Omega_0)\geq 1-\varepsilon/2$. By Proposition 6.1, there is an absolute constant $C<\infty$ such that, for $f\in\mathcal{F}$ and $i\leq\sigma_j(t)-1$,

$$\|E^j_{t_{i+1}t}f-E^{j-1}_{t_{i+1}t}f\|_{(3)}$$
$$\leq C\kappa\exp\Big\{C\int_{t_{i+1}}^t\langle 1+|v|^4,\rho^j_s+\rho^{j-1}_s\rangle\,ds\Big\}\int_{t_{i+1}}^t\langle 1+|v|^3,|\rho^j_s-\rho^{j-1}_s|\rangle\,ds.$$

Also, by Proposition 4.1,

$$\|E^{j-1}_{t_{i+1}t}f\|\leq 3\big(1+6\kappa(t-t_{i+1})\big)\exp\Big\{8\int_{t_{i+1}}^t\langle 1+|v|^3,\rho^{j-1}_s\rangle\,ds\Big\}.$$

So, on Ω_0, for some absolute constant $C < \infty$, we have

$$\sup_{t\le T}\sup_{f\in\mathcal{F}}\int_0^t \langle (E_{st}^j - E_{st}^{j-1})f, dM_s^N\rangle$$
$$\le CA\kappa e^{CAT}\sum_{i=0}^{2^jT-1}\Big(2^{-j(1-\delta)}\sup_{f\in\mathcal{F}(3)}\int_{t_i}^{t_{i+1}}\langle E_{st_{i+1}}^j f, dM_s^N\rangle$$
$$+\sup_{t\in[t_i,t_{i+1}]}\sup_{f\in\mathcal{F}}\int_{t_i}^t \langle (E_{st}^j - E_{st}^{j-1})f, dM_s^N\rangle\Big).$$

Set $\mathcal{F}_t = \sigma\{\mu_s^N, \mu_s^{N'} : s \in [0,t]\}$. We apply Propositions 7.1 and 7.2 conditionally on $\mathcal{F}_{t_i}$ to obtain, for some constant $C(d,\delta) < \infty$,

$$\Big\| \sup_{f\in\mathcal{F}(3)}\int_{t_i}^t \langle E_{st}^j f, dM_s^N\rangle 1_{\{\langle 1+|v|^{5+\delta},\rho_{t_i}\rangle\le A\}}\Big\|_2$$
$$\le C\kappa e^{CA2^{-j}}N^{-1/d}\Big(\mathbb{E}\int_{t_i}^{t_{i+1}}\langle |v|^{9+2\delta}, \mu_s^N\rangle\, ds\Big)^{1/2}$$

and

$$\Big\| \sup_{t\in[t_i,t_{i+1}]}\sup_{f\in\mathcal{F}}\int_{t_i}^t \langle (E_{st}^j - E_{st}^{j-1})f, dM_s^N\rangle 1_{\{\langle 1+|v|^{5+\delta},\rho_{t_i}\rangle\le A\}}\Big\|_2$$
$$\le CA2^{-j}\kappa e^{CA2^{-j}}N^{-1/d}\Big(\mathbb{E}\int_{t_i}^{t_{i+1}}\langle |v|^{9+2\delta}, \mu_s^N\rangle\, ds\Big)^{1/2}.$$

By Proposition 3.1, there is a constant $C(B,p) < \infty$ such that

$$\mathbb{E}\int_{t_i}^{t_{i+1}}\langle |v|^{9+2\delta}, \mu_s^N\rangle\, ds \le \int_{t_i}^{t_{i+1}} C\lambda(1+t^{p-9-2\delta})\, dt = C\lambda(2^{-j}+2^{-4\delta j}/(4\delta)).$$

Hence, for constants $C(B,d,p) < \infty$,

$$\Big\|\sup_{t\le T}\sup_{f\in\mathcal{F}}\int_0^t \langle (E_{st}^j - E_{st}^{j-1})f, dM_s^N\rangle 1_{\Omega_0}\Big\|_2$$
$$\le CA^2\lambda^{1/2}\kappa^2 e^{CAT}2^jT(2^{-j(1-\delta)}+2^{-j})(2^{-j/2}+\delta^{-1/2}2^{-2j\delta})N^{-1/d} \tag{57}$$
$$\le A\lambda^{1/2}\kappa^2 e^{CAT}(2^{-j\delta}+2^{-j/2})N^{-1/d}.$$

Here, we absorbed $4\delta^{-1/2}CAT$ into e^{CAT} in the second inequality by changing the constant C. By Proposition 6.1, there is an absolute constant $C < \infty$ such that, for all $f \in \mathcal{F}$ and $s \le t \le T$,

$$|E_{st}^J f(v) - E_{st}f(v)|$$
$$\le C\kappa(1+|v|^3)\exp\Big\{C\int_0^T \langle 1+|v|^4, \rho_t^J + \rho_t\rangle\, dt\Big\}\int_s^t \langle 1+|v|^3, |\rho_r^J - \rho_r|\rangle\, dr.$$

So, on Ω_0, we have

$$\left|E^J_{st} f(v) - E_{st} f(v)\right| \le CA\kappa e^{CAT} 2^{-J}\left(1 + |v|^3\right),$$

and so as $J \to \infty$,

$$\begin{aligned}(58)\qquad &\left\|\sup_{t\le T}\sup_{f\in\mathcal{F}}\int_0^t \langle (E^J_{st} - E_{st}) f, dM^N_s\rangle 1_{\Omega_0}\right\|_2 \\ &\quad\le CA\kappa e^{CAT} 2^{-J}\left\|\int_0^T \langle 1 + |v|^3, |dM^N_s|\rangle\right\|_2 \to 0.\end{aligned}$$

Finally, we use estimates (56), (57) and (58) in (55) and let $J \to \infty$ to obtain a constant $C(B, d, \varepsilon, \lambda, p, T) < \infty$ such that

$$\left\|\sup_{t\le T}\sup_{f\in\mathcal{F}}\int_0^t \langle E_{st} f, dM^N_s\rangle 1_{\Omega_0}\right\|_2 \le CN^{-1/d}.$$

An analogous estimate holds for N', and Theorem 1.2 then follows by Chebyshev's inequality.

9. Properties of the distance function. Recall that $W : \mathcal{S} \times \mathcal{S} \to [0, 4]$ is defined by

$$W(\mu, \nu) = \sup_{f\in\mathcal{F}} \langle f, \mu - \nu\rangle,$$

where $\mathcal{F}$ is the set of functions f on $\mathbb{R}^d$ such that $|\hat{f}(v)| \le 1$ and $|\hat{f}(v) - \hat{f}(v')| \le |v - v'|$ for all v, v', where $\hat{f}(v) = f(v)/(1 + |v|^2)$.

PROPOSITION 9.1. *The metric space* $(\mathcal{S}, W)$ *is complete and separable.*

PROOF. Write $\mathcal{P}$ for the set of Borel probability measures on $\mathbb{R}^d$, and define $\Phi : \mathcal{S} \to \mathcal{P}$ by $\Phi(\mu)(dv) = \frac{1}{2}(1 + |v|^2)\mu(dv)$. Write W_1^ρ for the Wasserstein-1 metric on $\mathcal{P}$ associated with the bounded metric $\rho(v, v') = |v - v'| \wedge 2$ on $\mathbb{R}^d$. Then $W(\mu, \nu) = 2W_1^\rho(\Phi(\mu), \Phi(\nu))$ for all $\mu, \nu \in \mathcal{S}$. Now $(\mathcal{P}, W_1^\rho)$ is complete and separable, and $\Phi(\mathcal{S})$ is closed in $\mathcal{P}$ under W_1^ρ, so $(\mathcal{S}, W)$ is also complete and separable. □

We will prove two approximation schemes for a measure μ in the Boltzmann sphere $\mathcal{S}$, by empirical distributions of systems of N particles. The first uses the empirical distribution $\mu^N = \frac{1}{N}\sum_{i=1}^N \delta_{V_i}$ of a sample of N independent random variables $V_1, \dots, V_N$ with distribution μ. The convergence of μ^N to μ has been extensively investigated for standard Wasserstein distances; see [4] or [5]. We modify some of the simpler ideas from [5] to obtain estimates for the weighted Wasserstein distance W used in this paper. The sample empirical distribution μ^N is not, however, a random variable in the Boltzmann sphere. Set

$$\bar{V}_N = \frac{1}{N}\sum_{i=1}^N V_i, \qquad S_N = \frac{1}{N}\sum_{i=1}^N |V_i - \bar{V}|^2, \qquad \tilde{V}_i = S_N^{-1/2}(V_i - \bar{V}_N).$$

On the event $\{S_N > 0\}$, define the *rescaled empirical distribution* $\tilde{\mu}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{V}_i}$. On the event $\{S_N = 0\}$, we take $\tilde{\mu}^N$ to be some arbitrary element of $\mathcal{S}_N$. Then $\tilde{\mu}^N$ is a random variable in $\mathcal{S}_N$. We will quantify the convergence of $\tilde{\mu}^N$ to μ in weighted Wasserstein distance, using the convergence of μ^N as an intermediate step.

PROPOSITION 9.2. *For all $d \geq 3$ and all $\mu \in \mathcal{S}$, we have $\mathbb{E}(W(\mu^N, \mu)) \to 0$ as $N \to \infty$. Set*

$$\beta = \beta(p) = \begin{cases} (p-2)/(p+d), & \text{if } p \in \big(2, 3d/(d-1)\big), \\ 1/d, & \text{if } p \in [3d/(d-1), \infty). \end{cases}$$

For all $p \in (2,\infty) \setminus \{3d/(d-1)\}$, there is a constant $C(d,p) < \infty$ such that, for all $N \in \mathbb{N}$,

$$\mathbb{E}\big(W(\mu^N, \mu)\big) \leq C\langle |v|^p, \mu\rangle N^{-\beta}. \tag{59}$$

For $d = 2$ and $p \in (2,\infty) \setminus \{3d/(d-1)\}$, or for $d \geq 3$ and $p = 3d/(d-1)$, the same estimate holds with an additional factor of $\log(N+1)$ *on the right-hand side. In the case when both $d = 2$ and $p = 6$, the additional factor is squared.*

PROPOSITION 9.3. *The conclusions of Proposition 9.2 remain valid if μ^N is replaced by $\tilde{\mu}^N$ and β is replaced by $\tilde{\beta} = \beta \wedge ((p-2)^2/(3p-4))$. Moreover, the constant C may be chosen so that, for all $N \in \mathbb{N}$, there is an event $\Omega(1/4)$, of probability exceeding $1 - C\langle |v|^p, \mu\rangle N^{-(p/4)\wedge(p/2-1)}$, such that*

$$\mathbb{E}\big(\langle |v|^p, \tilde{\mu}^N\rangle 1_{\Omega(1/4)}\big) \leq C\langle |v|^p, \mu\rangle.$$

It is simple to check that $\tilde{\beta} = \beta$ whenever $d \geq 2$ and $p \geq 3$. The following example shows that the exponent $\beta(p)$ cannot be improved for $p \in (2, 3d/(d-1))$ and hence that the moment threshold $p = 3d/(d-1)$ for convergence with optimal rate $N^{-1/d}$ also cannot be improved. Fix $p > 2$ and $q > d + p$, and consider the measure $\mu(dv) = c1_{\{|v|>r\}}|v|^{-q}\,dv$ where c and r are determined so that $\mu \in \mathcal{S}$. Then $\langle |v|^p, \mu\rangle < \infty$. Define

$$f_N(v) = \big(\operatorname{dist}\big(v, \operatorname{supp}\mu^N\big) \wedge 1\big)\big(1 + |v|^2\big).$$

Then $f_N \in \mathcal{F}$, so

$$\mathbb{E}\big(W\big(\mu, \mu^N\big)\big) \geq \mathbb{E}\big(\big\langle f_N, \mu - \mu^N\big\rangle\big) = \mathbb{E}\big(\langle f_N, \mu\rangle\big).$$

There are constants $a < \infty$ and $r_0 \geq r$ such that $\mu(\{u \in \mathbb{R}^d : |u - v| \geq 1\}) \geq e^{-a|v|^{-q}}$ whenever $|v| \geq r_0$. Then $\operatorname{dist}(v, \operatorname{supp} \mu^N) \geq 1$ with probability at least $e^{-Na|v|^{-q}}$. Hence

$$\begin{aligned}\mathbb{E}(\langle f_N, \mu\rangle) &\geq c\sigma_{d-1} \int_{r_0}^{\infty} e^{-Nat^{-q}} t^{d+1-q}\, dt \\ &= c\sigma_{d-1} N^{-1+(d+2)/q} \int_{r_0 N^{-1/q}}^{\infty} e^{-as^{-q}} s^{d+1-q}\, ds.\end{aligned}$$

Consider the limit $q \to p + d$ in the case $p < 3d/(d-1)$. Then $1 - (d+2)/q \to (p-2)/(p+d) < 1/d$, so we have justified the claims made above.

PROOF OF PROPOSITION 9.2. The following estimate is known for the N-sample empirical distribution μ_0^N of a probability measure μ_0 supported on $B_0 = (-1, 1]^d$. For all $d \geq 3$, there is a constant $C(d) < \infty$ such that, for all $N \in \mathbb{N}$, we have

$$\mathbb{E}(W_1(\mu_0^N, \mu_0)) \leq C(d) N^{-1/d}. \tag{60}$$

Here, W_1 denotes the Wasserstein-1 distance for the Euclidean metric on $\mathbb{R}^d$. For completeness, and since it may be read as a warm-up for the proof of Proposition 7.1, we give a proof. Fix $L \in \mathbb{N}$. For $\ell = 0, 1, \ldots, L-1$, we can partition B_0 as a set $\mathcal{P}_\ell$ of $2^{\ell d}$ translates of $(-2^{-\ell}, 2^{-\ell}]^d$. Fix a function f on $\mathbb{R}^d$ with $f(0) = 0$ and $|f(v) - f(v')| \leq |v - v'|$ for all $v, v' \in B_0$. Then we can write

$$f = \sum_{\ell=0}^{L-1} \sum_{B \in \mathcal{P}_\ell} a_B 1_B + g,$$

where $a_{B_0} = \langle f\rangle_{B_0}$ and $a_B = \langle f\rangle_B - \langle f\rangle_{\pi(B)}$ for $B \in \mathcal{P}_\ell$ and $\ell \geq 1$. Here we have written $\langle f\rangle_B$ for the average of f over B and $\pi(B)$ for the unique element of $\mathcal{P}_{\ell-1}$ containing B. By (41) and (42), we have $|a_B| \leq 2^{-\ell+1} c_d$ for all $B \in \mathcal{P}_\ell$ and all ℓ, and $|g(v)| \leq 2^{-L+2} c_d$. So, by Cauchy–Schwarz,

$$\begin{aligned}\langle f, \mu_0^N - \mu_0\rangle &= \sum_{\ell=0}^{L-1} \sum_{B \in \mathcal{P}_\ell} a_B(\mu_0^N(B) - \mu_0(B)) + \langle g, \mu_0^N - \mu_0\rangle \\ &\leq 2c_d \sum_{\ell=0}^{L-1} 2^{(d-2)\ell/2} \Big(\sum_{B \in \mathcal{P}_\ell} (\mu_0^N(B) - \mu_0(B))^2\Big)^{1/2} + 8c_d 2^{-L}.\end{aligned}$$

The right-hand side does not depend on f, so it is an upper bound for $W_1(\mu_0^N, \mu_0)$, by duality. Note that $\operatorname{var}(\mu_0^N(B)) \leq \mu_0(B)/N$. Now take expectations and use Cauchy–Schwarz again to obtain

$$\mathbb{E}(W_1(\mu_0^N, \mu_0)) \leq 2c_d \sum_{\ell=0}^{L-1} 2^{(d-2)\ell/2} N^{-1/2} + 8c_d 2^{-L}.$$

We optimize at $L = \lceil \log_2(N+1)/d \rceil$ to obtain (60). The same argument produces $N^{-1/2} \log_2(N+1)$ on the right when $d = 2$.

Set $B_k = 2^k B_0$. Fix $K \in \mathbb{N} \cup \{\infty\}$, and partition $\mathbb{R}^d$ as $\bigcup_{k=0}^K A_k$, where $A_0 = B_0$, $A_k = B_k \setminus B_{k-1}$ for $1 \le k < K$, and $A_K = \mathbb{R}^d \setminus (\bigcup_{k=0}^{K-1} A_k)$. Set $p_k = \mu(A_k)$ and write μ_k for the conditional distribution of μ on A_k. Write N_k for the number of elements of the sample falling in A_k and write $\mu_k^{N_k}$ for the empirical distribution of this sub-sample. Set $\hat{p}_k = N_k/N$. Then

$$\mu = \sum_{k=0}^K p_k \mu_k, \qquad \mu^N = \sum_{k=0}^K \hat{p}_k \mu_k^{N_k}.$$

Fix a function f on $\mathbb{R}^d$ such that $|\hat{f}(v)| \le 1$ and $|\hat{f}(v) - \hat{f}(v')| \le |v - v'|$ for all v, v', where $\hat{f}(v) = f(v)/(1+|v|^2)$. Then, for all k and all $v, v' \in B_k$, we have

$$\begin{aligned} |f(v)| &\le 1 + d2^{2k}, \\ |f(v) - f(v')| &\le (2 + |v|^2 + |v'|^2)|v - v'| \le 2(1 + d2^{2k})|v - v'|. \end{aligned}$$

Hence

$$\begin{aligned} \langle f, \mu^N - \mu \rangle &= \sum_{k=0}^{K-1} \hat{p}_k \langle f, \mu_k^{N_k} - \mu_k \rangle + (\hat{p}_k - p_k)\langle f, \mu_k \rangle + \langle f, \hat{p}_K \mu_K^{N_K} - p_K \mu_K \rangle \\ &\le \sum_{k=0}^{K-1} (1 + d2^{2k})\{2\hat{p}_k W_1(\mu_k^{N_k}, \mu_k) + |\hat{p}_k - p_k|\} \\ &\quad + \langle (1+|v|^2)1_{A_K}, \hat{p}_K \mu_K^{N_K} + p_K \mu_K \rangle. \end{aligned}$$

Note the inequalities

$$\mathbb{E}|\hat{p}_k - p_k| \le (2p_k) \wedge (p_k/N)^{1/2} \le 2N^{-1/d} p_k^{1-1/d}, \mathbb{E}(\hat{p}_k^{1-1/d}) \le p_k^{1-1/d}.$$

Estimate (60) scales from B_0 to B_k to give, on the event $\{N_k \ge 1\}$,

$$\mathbb{E}(W_1(\mu_k^{N_k}, \mu_k)|N_k) \le 2^k C(d) N_k^{-1/d}.$$

Hence, on taking the supremum over f and then the expectation, we obtain

$$\mathbb{E}(W(\mu^N, \mu)) \le \sum_{k=0}^{K-1} 2^{k+2}(1 + d2^{2k}) C(d) N^{-1/d} p_k^{1-1/d} + 2\langle (1+|v|^2)1_{A_K}, \mu \rangle.$$

Since $\mu \in \mathcal{S}$, the final term on the right is small for large K, so $\mathbb{E}(W(\mu^N, \mu)) \to 0$ as $N \to \infty$. If $\langle |v|^p, \mu \rangle < \infty$ for some $p > 2$, we can control the right-hand side using the bounds

$$\sum_{k=1}^{K-1} 2^{p(k-1)} p_k \le \langle |v|^p, \mu \rangle, \langle (1+|v|^2)1_{A_K}, \mu \rangle \le 2^{-(K-1)(p-2)+1} \langle |v|^p, \mu \rangle.$$

Finally, we optimize at $K = \lceil \log_2(N+1)/(d+p) \rceil$ when $p < 3d/(d-1)$ and $K = \infty$ when $p > 3d/(d-1)$ to obtain the claimed estimate. $\square$

PROOF OF PROPOSITION 9.3. Set $Q_N = N^{-1} \sum_{i=1}^N |V_i|^2$. Fix $\delta \in (0, 1/4]$, and consider the event

$$\Omega(\delta) = \{|Q_N - 1| \le \delta \text{ and } |\bar{V}_N| \le \delta\}.$$

Note that $Q_N = S_N + |\bar{V}_N|^2$. On $\Omega(\delta)$, by some simple estimation, we have $|S_N^{-1/2} - 1| \le 4\delta$, so $|\tilde{V}_i - V_i| \le (4|V_i| + 2)\delta$. Hence, in particular, there is a constant $C(p) < \infty$ such that

$$\mathbb{E}(\langle |v|^p, \tilde{\mu}^N \rangle 1_{\Omega(1/4)}) \le C(p)\mathbb{E}(\langle |v|^p, \mu^N \rangle) = C(p)\langle |v|^p, \mu \rangle.$$

Now, for all $f \in \mathcal{F}$, we have

$$f(\tilde{V}_i) - f(V_i) \le (|\tilde{V}_i - V_i| \wedge 1)(2 + |\tilde{V}_i|^2 + |V_i|^2) \le 24((\delta + \delta|V_i|) \wedge 1)(1 + |V_i|^2)$$

and so

$$\langle f, \tilde{\mu}^N - \mu^N \rangle = \frac{1}{N} \sum_{i=1}^N (f(\tilde{V}_i) - f(V_i)) \le \frac{24}{N} \sum_{i=1}^N ((\delta + \delta|V_i|) \wedge 1)(1 + |V_i|^2).$$

Hence

$$\mathbb{E}(W(\tilde{\mu}^N, \mu^N) 1_{\Omega(\delta)}) \le 24\langle \{(\delta + \delta|v|) \wedge 1\}(1 + |v|^2)\mu \rangle \to 0 \tag{61}$$

as $\delta \to 0$.

Since $\langle v, \mu \rangle = 0$ and $\langle |v|^2, \mu \rangle = 1$, we have $\mathbb{P}(\Omega \setminus \Omega(\delta)) \to 0$ as $N \to \infty$ for all $\delta > 0$ by the weak law of large numbers. For $p \ge 2$, there is a constant $C < \infty$, depending only on d and p, such that $\mathbb{E}(|\bar{V}_N|^{p/2})^2 \le \mathbb{E}(|\bar{V}_N|^p) \le C\langle |v|^p, \mu \rangle N^{-p/2}$. Hence

$$\mathbb{P}(|\bar{V}_N| > \delta) \le C\langle |v|^p, \mu \rangle \delta^{-p/2} N^{-p/4}.$$

For $p \ge 4$, since $\langle |v|^2, \mu \rangle = 1$, C may be chosen so that also $\mathbb{E}(|Q_N - 1|^{p/2}) \le C\langle |v|^p, \mu \rangle N^{-p/4}$ and so

$$\mathbb{P}(|Q_N - 1| > \delta) \le C\langle |v|^p, \mu \rangle \delta^{-p/2} N^{-p/4}.$$

For $p \in (2, 4]$ we use a different estimate. Set $R = \sqrt{\delta N}$ and write $X_i = |V_i|^2 \wedge R$ and $\bar{X} = N^{-1} \sum_{i=1}^N X_i$ and $x = \mathbb{E}(X_1)$. Then $\mathbb{E}(X_1^2) \le \langle |v|^p, \mu \rangle R^{4-p}$ and

$$|x - 1| \le \mathbb{E}|\bar{X} - Q_N| \le \langle |v|^2 1_{\{|v| \ge R\}}, \mu \rangle \le \langle |v|^p, \mu \rangle R^{2-p}$$

so

$$\begin{aligned} \mathbb{P}(|Q_N - 1| > \delta) &\le \mathbb{P}(|Q_N - \bar{X}| > \delta/3) + \mathbb{P}(|\bar{X} - x| > \delta/3) + \mathbb{P}(|x - 1| > \delta/3) \\ &\le 12\langle |v|^p, \mu \rangle \delta^{-p/2} N^{-(p/2-1)}. \end{aligned}$$

Here, for the second inequality, we estimated the first term using Markov's inequality, the second using Chebyshev and noted that the third term vanishes except in cases where the final estimate exceeds 1. We combine these estimates to see that there is a constant $C < \infty$, depending only on d and p, such that, for all $\delta \in (0, 1/4]$,

$$\mathbb{P}(\Omega \setminus \Omega(\delta)) \le C\delta^{-p/2}\langle |v|^p, \mu\rangle N^{-(p/4)\wedge(p/2-1)}. \tag{62}$$

Now, from (59), (61) and (62), for all $p \in (2, \infty) \setminus \{3d/(d-1)\}$, all $\delta \in (0, 1/4]$ and all $N \in \mathbb{N}$, we have

$$\begin{aligned}\mathbb{E}(W(\tilde{\mu}^N, \mu)) &\le \mathbb{E}(W(\mu^N, \mu)) + \mathbb{E}(W(\tilde{\mu}^N, \mu^N)1_{\Omega(\delta)}) + 4\mathbb{P}(\Omega \setminus \Omega(\delta)) \\ &\le C(N^{-\beta} + \delta^{(p-2)\wedge 1} + \delta^{-p/2}N^{-(p/4)\wedge(p/2-1)})\langle |v|^p, \mu\rangle.\end{aligned}$$

Hence $\mathbb{E}(W(\tilde{\mu}^N, \mu)) \to 0$ as $N \to \infty$. Moreover, on optimizing over δ, the terms $\delta^{(p-2)\wedge 1}$ and $\delta^{-p/2}N^{-(p/4)\wedge(p/2-1)}$ can be absorbed in the term $N^{-\beta}$, except possibly when $p \in (2, 3)$, and in that case we can take $\delta = (1/4)N^{-(p-2)/(3p-4)}$ for the desired estimate. □

10. Spatially homogeneous Boltzmann equation. Given an initial state μ_0 in the Boltzmann sphere $\mathcal{S}$, one can ask whether there exists a process $(\mu_t)_{t\ge 0}$ in $\mathcal{S}$ such that, for all bounded measurable functions f of compact support in $\mathbb{R}^d$ and all $t \ge 0$,

$$\langle f, \mu_t\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q(\mu_s, \mu_s)\rangle\, ds. \tag{63}$$

Here Q is the Boltzmann operator, defined in equation (5). Such a process would then be called a measure solution of the spatially homogeneous Boltzmann equation. While the existence and uniqueness (in law) of the Kac process is elementary, the existence and uniqueness of measure solutions is a hard question, but one which, extending a long line of prior works, including [10, 14], has been positively answered by Lu and Mouhot [11], Theorem 1.5.

After Kac [8], important contributions to understanding the behavior of versions of the Kac process were made by McKean [12] and Tanaka [16, 17]. Sznitman [15] gave the first proof for hard spheres that the Kac process converges weakly to solutions of the Boltzmann equation. Mischler and Mouhot [13], Theorem 6.2, proved a quantitative refinement of Sznitman's result, using a Wasserstein distance on the laws of k-samples from the empirical distribution. In recent work, Fournier and Mischler [6] and Cortez and Fontbona [1] have proved Wasserstein estimates for some other particle systems associated to the spatially homogeneous Boltzmann equation.

Our consistency estimate allows a further strengthening of Sznitman's result. In the convergence theorem below, we obtain a pathwise estimate, expressed in terms of a Wasserstein distance on the empirical distribution itself, and we are able to show, under suitable moment conditions, that the rate of convergence is the optimal one for discrete approximations in Wasserstein distance. The convergence results of both Sznitman and Mischler–Mouhot are expressed in terms of propagation of chaos, while our estimate may be applied to any initial N-particle system. For $p \geq 2$, define

$$\mathcal{S}(p) = \{\mu \in \mathcal{S} : \langle |v|^p, \mu\rangle < \infty\},$$

and call a solution locally bounded in $\mathcal{S}(p)$ if $\langle |v|^p, \mu_t\rangle$ is bounded on compact time intervals. We know from [11], Theorem 1.5, that, for all $\mu_0 \in \mathcal{S}$, there is a unique solution $(\mu_t)_{t\geq 0}$ in $\mathcal{S}$ to (63). Sznitman's theorem assumes $\mu_0 \in \mathcal{S}(3)$. The convergence result of Mischler and Mouhot, which has good long-time properties, assumes compactly supported initial data or at least an exponential moment.

THEOREM 10.1. *Assume that the collision kernel B satisfies conditions* (1) *and* (2). *Let $\mu_0 \in \mathcal{S}(p)$ for some $p \in (2, \infty)$. Then there exists a unique locally bounded solution $(\mu_t)_{t\geq 0}$ to* (63) *in $\mathcal{S}(p)$ starting from μ_0. Let $\varepsilon \in (0, 1]$, $\lambda \geq \langle |v|^p, \mu_0\rangle$ and $T \in [0, \infty)$. Then there exists a constant $C(B, d, \varepsilon, \lambda, p, T) < \infty$ with the following property. For all $N \in \mathbb{N}$ and any Kac process $(\mu_t^N)_{t\geq 0}$ in $\mathcal{S}_N$ with $\langle |v|^p, \mu_0^N\rangle \leq \lambda$, with probability exceeding $1 - \varepsilon$, for all $t \in [0, T]$, we have*

$$W(\mu_t^N, \mu_t) \leq C(W(\mu_0^N, \mu_0) + N^{-\alpha}),$$

where $\alpha(d, p)$ is given in Theorem 1.1. *For $p > 8$ and $d \geq 3$, we can take $\alpha = 1/d$. For $p > 8$ and $d = 2$ the estimate holds with $N^{-\alpha}$ replaced by $N^{-1/2}\log N$.*

PROOF. We will prove the first assertion on existence and uniqueness for completeness, while noting, as discussed above, that a stronger statement is already known. Let $(V_i : i \in \mathbb{N})$ be a sequence of independent random variables in $\mathbb{R}^d$ of distribution μ_0. Write $\bar{V}_N$ for the sample mean and S_N for the sample variance of $V_1, \dots, V_N$. For each $N \in \mathbb{N}$, set

$$\nu_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{S_N^{-1/2}(V_i - \bar{V}_N)}$$

on the event $\{S_N > 0\}$, and set ν_0^N equal to some arbitrary element of $\mathcal{S}_N$ otherwise. Conditioning on ν_0^N, let $(\nu_t^N)_{t\geq 0}$ be a Kac process in $\mathcal{S}_N$ starting from ν_0^N. Choose sequences $(\varepsilon_k : k \in \mathbb{N})$ in $(0, 1]$ and $(T_k : k \in \mathbb{N})$ in $[0, \infty)$ such that $\sum_k \varepsilon_k < \infty$ and $T_k \to \infty$. By Proposition 9.3 and Theorem 1.1, there exists an increasing sequence $(N_k : k \in \mathbb{N})$ in $\mathbb{N}$ such that, for all $k \in \mathbb{N}$, with probability exceeding $1 - \varepsilon_k$,

$$\langle |v|^p, \nu_0^{N_k}\rangle \leq C\langle |v|^p, \mu_0\rangle, \qquad W(\nu_0^{N_k}, \mu_0) \leq C\langle |v|^p, \mu_0\rangle\varepsilon_k$$

and then for all $t \leq T_k$

$$W(\nu_t^{N_k}, \nu_t^{N_{k+1}}) \leq C(W(\nu_0^{N_k}, \nu_0^{N_{k+1}}) + \varepsilon_k). \tag{64}$$

By Borel–Cantelli, almost surely, these inequalities hold for all sufficiently large k, so the sequence $((\nu_t^{N_k})_{t\geq 0}: k \in \mathbb{N})$ is Cauchy in the Skorohod space $D([0,\infty), (\mathcal{S}, W))$, and hence converges, with limit $(\nu_t)_{t\geq 0}$ say, since $D([0,\infty), (\mathcal{S}, W))$ is complete.

By Fatou's lemma and the moment estimate (14),

$$\mathbb{E}\Big(\sup_{s\leq t}\langle |v|^p, \nu_s\rangle\Big) \leq \liminf_k \mathbb{E}\Big(\sup_{s\leq t}\langle |v|^p, \nu_s^{N_k}\rangle 1_{\{\langle |v|^p, \nu_0^{N_k}\rangle \leq C\langle |v|^p, \mu_0\rangle\}}\Big) < \infty,$$

so $(\nu_t)_{t\geq 0}$ is locally bounded in $\mathcal{S}(p)$ almost surely. Fix a function f on $\mathbb{R}^d$ satisfying $|f(v)| \leq 1$ and $|f(v) - f(v')| \leq |v - v'|$ for all $v, v' \in \mathbb{R}^d$. From (64), since $\|f\| \leq 2$, we see that $\langle f, \nu_t^{N_k}\rangle \to \langle f, \nu_t\rangle 0$ uniformly on compact time intervals almost surely. Consider the equation

$$\langle f, \nu_t^N\rangle = \langle f, \nu_0^N\rangle + M_t^{N,f} + \int_0^t \langle f, Q(\nu_s^N, \nu_s^N)\rangle\, ds$$

with $N = N_k$ in the limit $k \to \infty$. Estimate (7) implies that $M_t^{N_k,f} \to 0$ uniformly on compact time intervals in probability. Moreover,

$$\langle f, Q(\nu_t^{N_k}, \nu_t^{N_k})\rangle - \langle f, Q(\nu_t, \nu_t)\rangle = \langle g_t, \nu_t^{N_k} - \nu_t\rangle,$$

where

$$g_t(v) = \int_{\mathbb{R}^d \times S^{d-1}} \{f(v') + f(v_*') - f(v) - f(v_*)\} B(v - v_*, d\sigma)(\nu_t^{N_k} + \nu_t)(dv_*)$$

and, by some straightforward estimation, $\|g_t\| \leq \max\{16, 12 + 8\kappa\}$ for all $t \geq 0$. Hence, we can pass to the limit uniformly on compact time intervals in probability to obtain

$$\langle f, \nu_t\rangle = \langle f, \mu_0\rangle + \int_0^t \langle f, Q(\nu_s, \nu_s)\rangle\, ds$$

for all $t \geq 0$, almost surely. A separability argument shows that almost surely, this equation holds for all such functions f and all $t \geq 0$. So, almost surely, $(\nu_t)_{t\geq 0}$ is a solution, and in particular, a locally bounded solution in $\mathcal{S}(p)$ exists.

Now let $(\mu_t)_{t\geq 0}$ be any locally bounded solution in $\mathcal{S}(p)$ starting from μ_0, and let $(\mu_t^N)_{t\geq 0}$ be any Kac process in $\mathcal{S}_N$. Then

$$\mu_t^N - \mu_t = (\mu_0^N - \mu_0) + M_t^N + \int_0^t 2Q(\rho_s, \mu_s^N - \mu_s)\, ds,$$

where now $\rho_t = (\mu_t + \mu_t^N)/2$. The argument of Section 4 applies without essential change to show that, for all $t \geq 0$ and all functions f_t on $\mathbb{R}^d$, we have

$$\langle f_t, \mu_t^N - \mu_t\rangle = \langle f_0, \mu_0^N - \mu_0\rangle + \int_0^t \langle f_s, dM_s^N\rangle,$$

where $f_s(v) = E_{(s,v)}\langle f_t, \tilde{\Lambda}_t\rangle)$ and where $(\Lambda_t^*)_{t\geq s}$ is a linearized Kac process in environment $(\rho_t)_{t\geq 0}$. Next, the argument of Section 5 applies to show that, for all

$\varepsilon \in (0, 1]$ and all $T \in [0, \infty)$, for all $N \in \mathbb{N}$, with probability exceeding $1 - \varepsilon$, for all $t \le T$, we have

$$W(\mu_t^N, \mu_t) \le C\big(W(\mu_0^N, \mu_0) + N^{-\alpha(d,p)}\big), \tag{65}$$

where $C < \infty$ depends only on $B, d, \varepsilon, \lambda, p$ and T, where λ is an upper bound for $\langle |v|^p, \mu_0\rangle$ and $\langle |v|^p, \mu_0^N\rangle$. Convergence at rate $N^{-1/d}$ could be proved for $p > 8$ by checking that the arguments leading to the estimate for $W(\mu_t^N, \mu_t^{N'})$ apply also when $(\mu_t^{N'})_{t\ge 0}$ is replaced by $(\mu_t)_{t\ge 0}$. Alternatively, we can find N' so that $(N')^{-\alpha(d,p)} \le N^{-1/d}$ and, by Proposition 9.3, $W(\nu_0^{N'}, \mu_0) \le CN^{-1/d}$ with probability exceeding $1 - \varepsilon$. Then, by Theorem 1.1 and (65), with probability exceeding $1 - 3\varepsilon$, for all $t \le T$, we have

$$\begin{aligned} W(\mu_t^N, \mu_t) &\le W(\mu_t^N, \nu_t^{N'}) + W(\nu_t^{N'}, \mu_t) \\ &\le C\big(W(\mu_0^N, \nu_0^{N'}) + W(\nu_0^{N'}, \mu_0) + N^{-1/d} + (N')^{-\alpha(d,p)}\big) \\ &\le C\big(W(\mu_0^N, \mu_0) + 4N^{-1/d}\big). \end{aligned}$$

Finally, we can take $\mu_t^{N_k} = \nu_t^{N_k}$ and let $k \to \infty$ to see that $\mu_t = \nu_t$ for all $t \ge 0$, so $(\nu_t)_{t\ge 0}$ is the only solution which is locally bounded in $\mathcal{S}(p)$. □

We can combine Theorem 10.1 with Proposition 9.3 to obtain the following stochastic approximation for solutions to the spatially homogeneous Boltzmann equation.

COROLLARY 10.2. *Assume that the collision kernel B satisfies conditions* (1) *and* (2). *Let $\mu_0 \in \mathcal{S}(p)$ for some $p \in (2, \infty)$, and let $(\mu_t)_{t\ge 0}$ be the unique locally bounded solution to* (63) *in $\mathcal{S}(p)$ starting from μ_0. Write μ_0^N for the random variable in $\mathcal{S}_N$ constructed by sampling from μ_0 as in Proposition* 9.3, *and conditioning on μ_0^N, let $(\mu_t^N)_{t\ge 0}$ be a Kac process starting from μ_0^N. Then, for all $\varepsilon \in (0, 1]$, all $\lambda \ge \langle |v|^p, \mu_0\rangle$ and all $T \in [0, \infty)$, there are constants $\alpha(d, p) > 0$ and $C(B, d, \varepsilon, \lambda, p, T) < \infty$, such that with probability exceeding $1 - \varepsilon$, for all $t \le T$,*

$$W(\mu_t^N, \mu_t) \le CN^{-\alpha}.$$

For $p > 8$, we can take $\alpha = 1/d$ when $d \ge 3$, and the estimate holds with $N^{-1/2}\log N$ in place of $N^{-\alpha}$ when $d = 2$.

On the other hand, if one views the spatially homogeneous Boltzmann equation as a means to compute approximations to the Kac process, the following estimate provides a measure of accuracy for this procedure.

COROLLARY 10.3. *Assume that the collision kernel B satisfies conditions* (1) *and* (2). *Fix $d \ge 3$, $\varepsilon \in (0, 1]$ and $\tau, T \in (0, \infty)$ with $\tau \le T$. There is a constant*

$C < \infty$, depending only on B, d, ε, τ and T, with the following property. Let $N \in \mathbb{N}$ and let $(\mu_t^N)_{t\geq 0}$ be a Kac process in $\mathcal{S}_N$ with collision kernel B. Denote by $(\mu_t)_{t\geq\tau}$ the solution to the spatially homogeneous Boltzmann equation with collision kernel B starting from μ_τ^N at time τ. Then, with probability exceeding $1-\varepsilon$, for all $t \in [\tau, T]$, we have $W(\mu_t^N, \mu_t) \leq CN^{-1/d}$. The same holds for $d = 2$ if we replace $N^{-1/d}$ by $N^{-1/2}\log N$.

PROOF. Use (13) to find a constant $\lambda(B, \tau, \varepsilon) < \infty$ such that $\langle |v|^9, \mu_\tau^N\rangle \leq \lambda$ with probability exceeding $1 - \varepsilon/2$. Then apply Theorem 10.1 with $\varepsilon/2$ in place of ε to find the desired constant C. $\square$

APPENDIX

We state and prove a basic lemma on the time-evolution of signed measures, which allows us to control the evolution of the total variation when the signed measures are given by an integral over time. Let $(E, \mathcal{E})$ be a measurable space. Write $\mathcal{M}^+$ (resp., $\mathcal{M}$) for the set of finite measures (resp., signed measures of finite total variation) on $(E, \mathcal{E})$. For $\mu \in \mathcal{M}$, write $|\mu|$ for the associated total variation measure and $\|\mu\|$ for the total variation.

LEMMA A.1. *Assume that $(E, \mathcal{E})$ is separable. Let $T \in (0, \infty)$. Let $\mu_0 \in \mathcal{M}$ and $\lambda_0 \in \mathcal{M}^+$ be given, along with a measurable map $t \mapsto \nu_t : [0, T] \to \mathcal{M}$ such that ν_t is absolutely continuous with respect to λ_0 for all $t \in [0, T]$ and $\int_0^T \|\nu_t\|\, dt < \infty$. Set*

$$\mu_t = \mu_0 + \int_0^t \nu_s\, ds.$$

Then there exists a measurable map $\sigma : [0, T] \times E \to \{-1, 0, 1\}$ such that, for all $t \in [0, T]$, we have $\mu_t = \sigma_t|\mu_t|$ and

$$|\mu_t| = |\mu_0| + \int_0^t \sigma_s \nu_s\, ds.$$

A version of the lemma, without the hypothesis of separability and for the case where $t \mapsto \nu_t : [0, T] \to \mathcal{M}$ is continuous in total variation, was stated by Kolokoltsov in [9], Lemma A.1. The proof given in [9] contains a gap, which we have not been able to fill. The case where $(E, \mathcal{E})$ is $\mathbb{R}^d$ with its Borel σ-algebra and where $t \mapsto \nu_t : [0, T] \to \mathcal{M}$ is continuous in total variation, has been proved by Lu and Mouhot [11], Lemma 5.1. We will use a substantially different argument, which allows us to replace this hypothesis of continuity with the existence of a reference measure λ_0.

PROOF OF LEMMA A.1. There exists an increasing sequence $(\mathcal{E}_n : n \in \mathbb{N})$ of finite σ-algebras generating $\mathcal{E}$. Write $\mathcal{A}_n$ for the partition of E generating $\mathcal{E}_n$.

Consider the finite measure $\lambda = \lambda_0 + |\mu_0| + \int_0^T |\nu_t|\,dt$ on $(E, \mathcal{E})$. By scaling we reduce to the case where λ is a probability measure. For each $t \in [0, T]$, define $\mathcal{E}_n$-measurable functions α_t^n and β_t^n by on E by setting

$$\alpha_t^n(x) = \mu_t(A)/\lambda(A), \qquad \beta_t^n(x) = \nu_t(A)/\lambda(A)$$

if $x \in A$ for some $A \in \mathcal{A}_n$ with $\lambda(A) > 0$ and setting $\alpha_t^n(x) = \beta_t^n(x) = 0$ if there is no such A. Then, for all $x \in E$, the map $t \mapsto \beta_t^n(x)$ is integrable on $[0, T]$ and

$$\alpha_t^n(x) = \alpha_0^n(x) + \int_0^t \beta_s^n(x)\,ds.$$

For each $t \in [0, T]$, we have $|\mu_t| \le \lambda$ so $|\alpha_t^n| \le 1$ and $\alpha_t^n \lambda = \mu_t$ on $\mathcal{E}_n$. Moreover, the sequence $(\alpha_t^n : n \in \mathbb{N})$ is a λ-martingale in the filtration $(\mathcal{E}_n : n \in \mathbb{N})$. So, by the martingale convergence theorem, there exists $\tilde{\alpha}_t \in L^1(\lambda)$ such that $\alpha_t^n \to \tilde{\alpha}_t$ as $n \to \infty$, λ-almost everywhere and in $L^1(\lambda)$. Then $\tilde{\alpha}_t \lambda = \mu_t$ on $\bigcup_n \mathcal{E}_n$ and hence on $\mathcal{E}$ by uniqueness of extension.

For $\tau = \{t_0, \dots, t_N\} \subseteq [0, T]$ with $t_0 < \cdots < t_N$ and any function $(\alpha_t(x) : t \in [0, T], x \in E)$, define a function $|\alpha|_\tau$ on E by

$$|\alpha|_\tau = |\alpha_0| + \sum_{k=0}^{N-1} |\alpha_{t_{k+1}} - \alpha_{t_k}|.$$

Then, for all $A \in \mathcal{A}_n$, on A, we have

$$\lambda(A)\big|\alpha^n\big|_\tau = |\mu_0|(A) + \sum_{k=0}^{N-1} |\mu_{t_{k+1}} - \mu_{t_k}|(A) \le \lambda(A)$$

so $|\alpha^n|_\tau \le 1$ everywhere.

Fix $E_0 \in \mathcal{E}$ with $\lambda(E_0) = 1$ such that $\alpha_t^n(x) \to \tilde{\alpha}_t(x)$ as $n \to \infty$ for all $t \in [0, T] \cap (T\mathbb{Q})$ and all $x \in E_0$. Write $\mathcal{T}$ for the set of finite subsets of $[0, T] \cap (T\mathbb{Q})$. Then, for all $x \in E_0$ and $\tau \in \mathcal{T}$, we have $|\tilde{\alpha}|_\tau(x) \le 1$, so the map $t \mapsto \tilde{\alpha}_t(x) : [0, T] \cap (T\mathbb{Q}) \to [-1, 1]$ has total variation bounded by 1. Hence, for $x \in E_0$, we can define a càdlàg map $t \mapsto \alpha_t(x) : [0, T] \to [-1, 1]$ by

$$\alpha_t(x) = \begin{cases} \lim\limits_{s \to t, s \in (t, T) \cap (T\mathbb{Q})} \tilde{\alpha}_s(x), & t \in [0, T), \\ \tilde{\alpha}_T(x), & t = T. \end{cases}$$

For $x \in E \setminus E_0$, set $\alpha_t(x) = 0$ for all $t \in [0, T]$. We have $\alpha_T \lambda = \tilde{\alpha}_T \lambda = \mu_T$ as we showed above. For $t \in [0, T)$ and $s \in (t, T) \cap (T\mathbb{Q})$, we have in the limit $s \to t$

$$\|\alpha_t \lambda - \mu_t\| \le \|\alpha_t \lambda - \tilde{\alpha}_s \lambda\| + \|\mu_s - \mu_t\| \le \langle |\alpha_t - \tilde{\alpha}_s|, \lambda \rangle + \int_t^s \|\nu_r\|\,dr \to 0$$

so $\alpha_t \lambda = \mu_t$. Define $\sigma : [0, T] \times E \to \{-1, 0, 1\}$ by $\sigma_t(x) = \operatorname{sgn}(\alpha_t(x))$. Then σ is measurable and $\mu_t = \sigma_t |\mu_t|$ for all $t \in [0, T]$.

For any function ψ on $[-1, 1]$ with continuous bounded derivative, we have

$$\psi(\alpha_t^n(x)) = \psi(\alpha_0^n(x)) + \int_0^t \psi'(\alpha_s^n(x))\beta_s^n(x)\,ds$$

for all $t \in [0, T]$ and all $x \in E$. Since ν_t is absolutely continuous with respect to λ for all $t \in [0, T]$, we have on $\mathcal{E}_n$

$$\psi(\alpha_t^n)\lambda = \psi(\alpha_0^n)\lambda + \int_0^t \psi'(\alpha_s^n)\nu_s\,ds$$

for all $t \in [0, T]$. Since $\nu_s(dx)\,ds$ is absolutely continuous with respect to $\lambda(dx)\,ds$, we have $\alpha_s^n(x) \to \alpha_s(x)$ as $n \to \infty$ almost everywhere for $\nu_s(dx)\,ds$. Hence, on letting $n \to \infty$, we obtain on $\bigcup_n \mathcal{E}_n$

$$\psi(\alpha_t)\lambda = \psi(\alpha_0)\lambda + \int_0^t \psi'(\alpha_s)\nu_s\,ds$$

for all $t \in [0, T]$. The identity then holds on $\mathcal{E}$ by uniqueness of extension. Set $\psi_k(x) = \sqrt{x^2 + 1/k}$. Then $\psi_k(x) \to |x|$ and $\psi_k'(x) \to \operatorname{sgn}(x)$ as $k \to \infty$ for all $x \in [-1, 1]$. By dominated convergence, for all $A \in \mathcal{E}$ and all $t \in [0, T]$, we have

$$\langle \psi_k(\alpha_t)1_A, \lambda \rangle \to \langle |\alpha_t| 1_A, \lambda \rangle = |\mu_t|(A)$$

and

$$\int_0^t \langle \psi_k'(\alpha_s)1_A, \nu_s \rangle\,ds \to \int_0^t \langle \operatorname{sgn}(\alpha_s)1_A, \nu_s \rangle\,ds = \int_0^t \langle \sigma_s 1_A, \nu_s \rangle\,ds.$$

Hence, on taking $\psi = \psi_k$ above and letting $k \to \infty$, we obtain the desired identity. □

Acknowledgments. I am grateful to Clément Mouhot and Richard Nickl for several discussions in the course of this work, and to a helpful referee whose comments led to improvements in the paper.

REFERENCES

[1] CORTEZ, R. and FONTBONA, J. (2014). Quantitative propagation of chaos for generalized Kac particle systems. Available at arXiv:1406.2115.

[2] DARLING, R. W. R. and NORRIS, J. R. (2008). Differential equation approximations for Markov chains. *Probab. Surv.* **5** 37–79. MR2395153

[3] DA PRATO, G. and ZABCZYK, J. (1992). A note on stochastic convolution. *Stoch. Anal. Appl.* **10** 143–153. MR1154532

[4] DEREICH, S., SCHEUTZOW, M. and SCHOTTSTEDT, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. MR3127919

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

[6] FOURNIER, N. and MISCHLER, S. (2013). Rate of convergence of the Nanbu particle system for hard potentials. Available at arXiv:1302.5810.

[7] HAUSENBLAS, E. and SEIDLER, J. (2001). A note on maximal inequality for stochastic convolutions. *Czechoslovak Math. J.* **51(126)** 785–790. MR1864042

[8] 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. MR0084985

[9] KOLOKOLTSOV, V. N. (2006). Kinetic equations for the pure jump models of k-nary interacting particle systems. *Markov Process. Related Fields* **12** 95–138. MR2223422

[10] LU, X. (1999). Conservation of energy, entropy identity, and local stability for the spatially homogeneous Boltzmann equation. *J. Stat. Phys.* **96** 765–796. MR1716814

[11] LU, X. and MOUHOT, C. (2012). On measure solutions of the Boltzmann equation, Part I: Moment production and stability estimates. *J. Differential Equations* **252** 3305–3363. MR2871802

[12] MCKEAN, H. P. JR. (1966). Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Arch. Ration. Mech. Anal.* **21** 343–367. MR0214112

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

[14] MISCHLER, S. and WENNBERG, B. (1999). On the spatially homogeneous Boltzmann equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **16** 467–501. MR1697562

[15] SZNITMAN, A.-S. (1984). Équations de type de Boltzmann, spatialement homogènes. *Z. Wahrsch. Verw. Gebiete* **66** 559–592. MR0753814

[16] TANAKA, H. (1978/79). Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete* **46** 67–105. MR0512334

[17] TANAKA, H. (1983). Some probabilistic problems in the spatially homogeneous Boltzmann equation. In *Theory and Application of Random Fields* (*Bangalore*, 1982). *Lecture Notes in Control and Inform. Sci.* **49** 258–267. Springer, Berlin. MR0799949

STATISTICAL LABORATORY
CENTRE FOR MATHEMATICAL SCIENCES
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: J.R.Norris@statslab.cam.ac.uk