ENTROPIC CONVERGENCE OF RANDOM BATCH METHODS FOR INTERACTING PARTICLE DIFFUSION

DHEERAJ NAGARAJ

Abstract. We propose a co-variance corrected random batch method for interacting particle systems. By establishing a certain entropic central limit theorem, we provide entropic convergence guarantees for the law of the entire trajectories of all particles of the proposed method to the law of the trajectories of the discrete time interacting particle system whenever the batch size $B \gg (\alpha n)^{1/3}$ (where $n$ is the number of particles and $\alpha$ is the time discretization). This in turn implies that the outputs of these methods are nearly statistically indistinguishable when $B$ is even moderately large. Previous works mainly considered convergence in Wasserstein distance with required stringent assumptions on the potentials or the bounds had an exponential dependence on the time horizon. This work makes minimal assumptions on the interaction potentials and in particular establishes that even when the particle trajectories diverge to infinity, they do so in the same way for both the methods. Such guarantees are very useful in light of the recent advances in interacting particle based algorithms for sampling.

1. Introduction. Interacting particle systems arise ubiquitously in various fields of natural and social sciences, including Plasma Physics, Chemotaxis, molecular simulations, vortex models and group dynamics [19, 5, 16, 30, 31]. Here the system is made up of multiple agents which evolve in time by interacting with the environment and each other via the aggregation-diffusion equations [9]. These models give rise to rich dynamics, and can lead to various phenomenon like swarming [10, 6, 16, 33], flocking [22, 14, 1], chemotaxis [4, 23], vortex formation [28, 31, 12] and hence are widely used by the scientific community. This has also has generated significant interest from a mathematical perspective [3, 9, 4, 22] where precise behavior of popular models, large deviations principles and convergence to mean field limits have been studied. Inspired by these models, various optimization [8], sampling [7, 17, 29, 13] and optimal control [11, 13] methods have been studied. Many of the popular interactions between particles are singular like the Coulomb force [2] or the interactions in the Cucker-Smale model for flocking [14]. Therefore, it important to simulate these models on a large scale (many particles) in a computationally efficient manner with possibly singular interactions. The straight-forward method of accounting for each particle’s interaction with every other particle has the disadvantage of quadratic complexity in the number of particles, which can be computationally expensive. Therefore, in this work, we propose the co-variance corrected random batch method which has a much smaller computational complexity. Unlike previous works, we consider convergence guarantees with respect to the KL divergence and hence the total variation distance for the joint trajectories. Such guarantees are very important for sampling algorithms based on interacting particle systems [7, 17, 29, 13]. They also provide a way of establishing confidence intervals to the outcomes of simulations based on such dynamical systems.

1.1. Random Batch Methods:. Random batch methods are widely used in statistics and machine learning to improve the computational complexity of algorithms like gradient descent [20]. Here the basic idea is to use a random subset of the data in each step of to compute the gradient of the loss function, instead of using all the data points. This idea has also appeared in the context of stochastic Langevin gradient
descent (SLGD), which is widely used in Bayesian inference [34].

In the context of interacting particle systems, the random batch method means that during the time evolution of a given particle, we only consider its interaction with a random subset (or ‘random batch’) of the other particles chosen independently at random at each time step. This method was first proposed in [24] which also provided convergence guarantees to the computationally expensive method which considers interactions between all pairs of particles. Subsequent works [25, 26, 15, 27] extend these convergence guarantees to more general settings (and specialized settings). We mainly discuss the general results in [24, 25] which consider the convergence of random batch methods to the interacting particle method with respect to the Wasserstein metric. While this allows one to understand the closeness in terms of the Euclidean distance, the guarantees come with sometimes stringent assumptions which cannot be readily removed. For instance, [24] assumes that the confining potential is strongly convex and the interaction potentials are smooth enough, with the lipschitz constant being small enough compared to the strong-convexity parameter. While [26] relaxes the strong-convexity assumption, the convergence bounds suffer from an exponential dependence on time (due to control via. Gronwall’s lemma), which makes the bounds vacuous after a short duration of time. We note that such dependencies might be unavoidable with respect to the Wasserstein distance even in the discrete time case since the particles may start diverging to infinity at an exponential rate.

1.2. Covariance Corrected Random Batch Method. In this work, we propose a modification to the random batch method, called the co-variance corrected random batch method (CovRBM) where we correct for the excess noise introduced by the random batches. The analysis deviates from the prior works in that we consider statistical indistinguishability by showing convergence with respect to the KL divergence. Statistical indistinguishability is a powerful notion of convergence which can guarantee that no statistical test can distinguish the paths generated by CovRBM from those generated by the interacting particle method with non-vanishing probability. Such guarantees can be immensely useful to the practitioner - especially when such systems are used to sample from a distribution, or when inferences are to be made from computer simulations of physical systems.

Even though the Euclidean metric is important to the underlying systems, disregarding it to consider a statistical view allows us establish convergence under the mild assumption that the interaction forces are bounded and measurable, without any assumptions regarding regularity or strong convexity. Under these conditions, we establish entropic convergence (i.e, convergence in KL divergence) of the joint law of the paths given by the random batch method to the joint law of the paths given by the interacting particle method whenever the batch size $B$ is larger than $(\alpha n K)^{1/3}$ where $\alpha$ is the time discretization, $n$ is the number of particles and $K$ is the time horizon. In particular, our results show that even if the interacting particle method diverges to infinity exponentially fast, the CovRBM also diverges to infinity in a similar way with high probability. When the sets where the forces are unbounded is known to the user, we also derive methods to consider unbounded forces efficiently and with similar convergence guarantees.

On a technical level, the main observation behind our algorithm is that vanilla random batch method introduces noise in addition to the Brownian increments added in each step. Similar to the concept of bootstrap used in statistics [18], we can apply the central limit theorem to show that this additional noise looks approximately Gaussian. We construct CovRBM by estimating the covariance of this additional
noise efficiently and subtracting it from the diffusion term (i.e., the Brownian increments). By extending the Wasserstein CLT established by [35] to an entropic CLT under the current setting, we show that the total driving noise under CovRBM (that is Brownian increment plus the random batch noise) is nearly statistically indistinguishable from just the Brownian increments. Since the trajectories of the particles are a function of the driving noise, we can conclude that they too are nearly statistically indistinguishable.

Remark 1.1. Many of the prior works consider the random batch method with respect to the continuous time stochastic differential equation. The discrete time version is called the ‘forward Euler method’. In this work, we are concerned with convergence of the forward Euler random batch method to the forward Euler interacting particle system. However, we believe that considering the Wasserstein distance in this context leads to similar drawbacks as outlined above.

1.3. Organization. We introduce the problem setup and notation in Section 2. We then systematically derive CovRBM in Section 3. In Section 4, we state the main theoretical results and give the idea behind the proof and the proof of the main results in Section 5. We establish the important technical results in Sections 6 and 7. In particular, Section 7 shows a CLT in Wasserstein distance for empirical sums which are added to a (large) Gaussian random variable, which can be readily lifted into a CLT in KL divergence.

2. Notation and Problem Setup. We consider the discrete time aggregation-diffusion equations for multi-particle systems. Let $X_0^1,\ldots,X_0^n \in \mathbb{R}^d$ be the initial positions of $n$ particles. We allows these positions to be possibly random and drawn from an arbitrary initial distribution. We denote their positions at time $t \in \mathbb{N} \cup \{0\}$ by $X_i^t$ for $i \in [n]$. Let $Z_i^t$ be i.i.d. $\mathcal{N}(0,I_d)$ vectors independent of the initial positions. We fix some $\alpha > 0$ to be the time discretization parameter and $\sigma > 0$ to be the strength of diffusion.

\begin{equation}
X_{i+1}^t = X_i^t + \alpha g_i^t(X_i^t) + \alpha \sum_{j=1}^n K_{ij}^t(X_i^t, X_j^t) + \sqrt{\alpha} \sigma Z_i^t
\end{equation}

Where the ‘external force’ $g_i^t : \mathbb{R}^d \to \mathbb{R}^d \cup \{\infty\}$ and ‘interaction forces’ $K_{ij}^t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \cup \{\infty\}$ are measurable functions. We shall call the trajectories generated by the evolution equations above the interacting particle method.

Here the interactions $K_{ij}^t$ specify how the particle $i$ interacts with particle $j$ at time $t$. Popular choices for $g_i^t$ and $K_{ij}^t$ have the following gradient structure: $g_i^t = \nabla \phi(x_i^t)$ and $K_{ij}^t(x,y) = \nabla_x \psi(\|x-y\|)$, where $\phi$ is called the confining potential and $\psi$ is called the interaction potential (see [9]). This allows us to interpret the dynamics in Equation (2.1) as gradient flows in the space measures under certain mean-field limits. In this paper, however, we will work with the more general setup established above. Here, the evolution equations usually model the dynamics of some interacting particle system. We will consider the above dynamics from $t = 0$ to $t = T$ and define the ‘real time’ of the physical system at iteration $t$ is $\alpha t$. Therefore, we define the ‘real time’ horizon to be $K := \alpha T$ and give our convergence results with respect to the quantity $K$. Notice that a straightforward implementation of Equation (2.1) incurs a per-step computational complexity of $\Theta(n^2)$, which can be quite expensive. We will attempt to mitigate this drawback with CovRBM.
**Notation.** We will use the convention that $X^i_t = \infty$ for every $t > t_0$ and $i \in [n]$ if and only if for some $i_0, j_0 \in [n]$ either $|g_{i_0}^j(X_{t_0}^{j_0})| = \infty$ or $\|K_{i_0j_0}^{t_0} (X_{t_0}^{i_0}, X_{t_0}^{j_0})\| = \infty$. By $\text{KL}(\nu || \mu)$, we denote the KL-divergence of the probability measure $\nu$ with respect to $\mu$. Whenever $\mu$ and $\nu$ are probability measures over some Euclidean space $\mathbb{R}^k$ with finite second moments, we denote their 2-Wasserstein distance with respect to the Euclidean distance by $W_2(\mu, \nu)$. For random variable $X$ and a probability measure $\mu$, by $X \sim \mu$ we mean that the law of $X$ is $\mu$. For the sake of clarity, whenever $X \sim \mu$ and $Y \sim \nu$, we let $\text{KL}(Y || X)$ to mean $\text{KL}(\nu || \mu)$ and $\mathcal{W}_2(X, Y)$ to mean $\mathcal{W}_2(\mu, \nu)$.

Similarly, given a sigma algebra $\mathcal{G}$ over the same measure space as random variables $X, Y$, by $\text{KL}(X || Y | \mathcal{G})$ we denote $\text{KL}(\mathcal{L}(X | \mathcal{G}) || \mathcal{L}(Y | \mathcal{G}))$, which is a $\mathcal{G}$ measurable random variable. Whenever we say $X_t$, we mean the tuple $(X^1_t, \ldots, X^n_t)$. We denote the iterates of Equation (2.1) by $X_t$ and the iterates of CovRBM by $X_t$. For $S \subseteq [n]$, by $X^S_t$ we denote the collection $(X^i_t)_{i \in S}$.

**Remark 2.1.** Let $S_1, \ldots, S_L \subseteq 2^{[n]}$, we can consider a much more general version of the dynamics given in Equation (2.1) given as follows:

$$X^i_{t+1} = X^i_t + \alpha g^i_t(X^i_t) + \sum_{l=1}^L \frac{\alpha}{|S_l|} \sum_{S \in S_l} K^i_t(S) (X^i_t, X^S_t) + \sqrt{\alpha \sigma} Z^i_t$$

For instance, when we take $S_1$ to be the set of all subsets of $[n]$ with exactly one element, we recover Equation (2.1). However, this allows us to move beyond pairwise interactions. While we discuss only the case of pairwise interactions in the sequel for the ease of exposition, we note that the proofs follow for this general case in a straightforward manner.

3. **Co-Variance Corrected Random Batch Method (CovRBM).** We will now systematically derive CovRBM, starting from the random batch method [24]. Instead of accounting for the interaction of a given particle $i$ with every other particle $j$, the random batch method only includes a random sub-set of the interactions. Setting the batch size to be $B \in [n]$, at time step $t$ and for each particle $i$, we choose indices $I^1_t, \ldots, I^B_t$ i.i.d uniformly random from the set $[n]$, independent of everything else. We replace $\frac{1}{n} \sum_{j=1}^n K^i_t(X^j_t, X^i_t)$ in Equation (2.1) with its unbiased estimator $\hat{K}^i_t := \frac{1}{B} \sum_{k=1}^B K_{i}^{ij_{ik}} (X^i_t, X^{ij_{ik}}_t)$. There are two issues with this implementation of the random batch method, which we hope to fix with CovRBM:

1. The functions $K_{ij}^{ij_{ik}} (\cdot, \cdot)$ can have singularities which can make the unbiased estimator perform poorly.
2. The random batches introduce additional (non-Gaussian) noise to the dynamics, which can change the output distribution.

3.1. **Accounting for Singularities.** Suppose we are given the particle positions $X_1, \ldots, X^B_t$. We propose the following way of accounting for singularities: fix an $M > 0$ and let $S^j_t(M) \subseteq [n]$ be any set such that $\{j : ||K^{ij} (X^j_t, X^i_t)|| > M\} \subseteq S^j_t(M)$. Let $K^i_t(X_t) := \frac{1}{n} \sum_{j=1}^n K^{ij}_{ij_{ik}}(X^j_t, X^i_t)$.

Draw $I^1_t, \ldots, I^B_t$ i.i.d. uniformly at random from the set $[n] \setminus S^j_t(M)$ independent of everything else. We then set

$$\hat{K}^i_t(X_t) := \begin{cases} K^i_t & \text{if } |S^j_t(M)| \geq \frac{n}{2} \\ \frac{1}{B} \sum_{j \in S^j_t(M)} K^{ij}_{ij_{ik}} (X^i_t, X^j_t) + \frac{\alpha}{nB} \sum_{k=1}^B K^i_{i}^{ij_{ik}} (X^i_t, X^{ij_{ik}}_t) & \text{otherwise} \end{cases}$$

The naive way of finding the set $S^j_t(M)$ is to compute $K^{ij} (X^j_t, X^i_t)$ for every
We define \( \hat{\Sigma}_t \) as an unbiased estimator of \( \Sigma_t \) conditioned on \( X_1^t, \ldots, X_n^t \) be \( \Sigma_t \). To make the notation more suggestive, define:

\[
N_t^i := \hat{K}_t^i - K_t^i = \frac{1}{B} \sum_{k=1}^{B} \left[ K_t^{i_k} (X_i^t, X_i^{t_k}) - K_t^i \right]
\]

Here \( N_t^i \) has 0 mean and covariance \( \Sigma_t \) given the positions \( X_1^t, \ldots, X_n^t \). By Equation (3.2), we see that when \( X_1^t, \ldots, X_n^t \) are given, \( N_t^i \) is the average of \( B \) i.i.d. random variables. By the central limit theorem, this noise \( N_t^i \) has a law which is approximately \( \mathcal{N}(0, \Sigma_t) \) in an appropriate sense. If we write down the straight forward random batch method, the evolution equations become (i.e, replace \( K_t^i \) by \( \hat{K}_t^i \)), the evolution equations become:

\[
X_{i+1}^{t} = X_i^{t} + \alpha g_i(X_i^{t}) + \alpha \hat{K}_t^i(X_i) + \sqrt{\alpha \sigma} Z_i^t
\]

This changes the driving noise from \( \sqrt{\alpha \sigma} Z_i^t \) which has the law \( \mathcal{N}(0, \alpha \sigma^2 I) \) to \( \alpha N_i^t + \sqrt{\alpha \sigma} Z_i^t \) which approximately has the law \( \mathcal{N}(0, \alpha \sigma^2 I + \alpha^2 \Sigma_t) \). To correct for this error, we estimate \( \Sigma_t \) by \( \hat{\Sigma}_t \) in a computationally efficient way by using another random batch (see Section 3.4). We can use this estimate of co-variance \( \hat{\Sigma}_t \) to correct for the additional noise injected in each iteration. We therefore arrive at the variance corrected random batch iteration as follows whenever \( \sigma I \gtrsim \alpha \Sigma_t^i \):

\[
X_{i+1}^{t} = X_i^{t} + \alpha g_i(X_i^{t}) + \alpha \hat{K}_t^i(X_i) + \sqrt{\alpha} \sqrt{\sigma^2 I - \alpha \Sigma_t^i} Z_i^t
\]

We define \( \hat{Z}_t^i := \frac{\sqrt{\alpha}}{\sigma} N_i^t + \sqrt{I - \frac{\alpha}{\sigma^2} \Sigma_t^i} Z_i^t \). Notice that this approximately has the law \( \mathcal{N}(0, \alpha \sigma^2 + \alpha^2 (\Sigma_t^i - \Sigma_t^i)) \), which is closer to the desired law \( \mathcal{N}(0, \alpha \sigma^2 I) \).
3.3. Putting It All Together. We now define CovRBM below. For the rest of this paper, we will take $\hat{X}_l^i$ to be ite- rates of CovRBM. Given (possibly random) initial positions $X_0^i$, external force functions $g_l^i : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$, interaction force functions $K_t^{ij} : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$, diffusion strength $\sigma$, time discretization parameter $\alpha$, batch size $B$ and a covariance estimator $\hat{\Sigma}_l^i$ (see Section 3.4), we start with $\hat{X}_0^i = X_0^i$. For the sake of clarity, we will not explicitly consider the modification given in Section 3.1 but assume the uniform bound $\sup_{x,y \in \mathbb{R}^d} \|K_t^{ij}(x,y)\| \leq M$. However we can incorporate Equation (3.1) in a straightforward way and all the results hold in this case as well.

Let $Z_s^i$ for $s \in \mathbb{N} \cup \{0\}$, $i \in [n]$ be a sequence of i.i.d. $\mathcal{N}(0, I)$ random variables independent of $(\hat{X}_0^i)_{i \in [n]}$. We will now inductively define the trajectory $\hat{X}_t^i$ for $t \geq 1$. Define the sigma algebra $\mathcal{F}_0 = \sigma(\hat{X}_0^i : i \in [n])$ and $\mathcal{F}_t = \sigma(\hat{X}_0^i, Z_s^i, \hat{X}_{t+1}^i : i \in [n], 0 \leq s \leq t-1)$ whenever $t \geq 1$. To define $\hat{X}_{t+1}^i$, draw $I_{t}^{ik}$ for $i \in [n], k \in [B]$ i.i.d. uniformly at random over the set $[n]$ and independent of $\mathcal{F}_t$. We construct the estimator:

$$\hat{K}_t^i(\hat{X}_t) := \frac{1}{B} \sum_{k=1}^{B} K_t^{ik} (\hat{X}_t^i, \hat{X}_t^{ik})$$

Let the co-variance of $\hat{K}_t^i(\hat{X}_t)$, when conditioned on $\mathcal{F}_t$ be $\Sigma_t^i$. That is,

$$\Sigma_t^i := \mathbb{E} \left[ (\hat{K}_t^i(\hat{X}_t) - K_t^i(\hat{X}_t)) (\hat{K}_t^i(\hat{X}_t) - K_t^i(\hat{X}_t))^\top | \mathcal{F}_t \right].$$

Let $\hat{\Sigma}_t^i$ be the estimator for $\Sigma_t^i$ as defined in Section 3.2, constructed using one of the covariance estimators given in Section 3.4. Then, we construct $\hat{X}_{t+1}^i$ for each particle $i \in [n]$ as:

$$\hat{X}_{t+1}^i = \hat{X}_t^i + \alpha g_l^i(\hat{X}_t^i) + \alpha \hat{K}_t^i(\hat{X}_t) + \sqrt{\alpha} \sqrt{\sigma^2 I - \alpha \hat{\Sigma}_t^i} Z_t^i$$

Remark 3.1. In the case of the more general interactions given in Equation (2.2), we fix batch sizes $B_1, \ldots, B_L$, sample the random subsets $I_{t}^{ik} \sim \text{unif}(\mathcal{S}_t)$ for $k \in [B_l], l \in [L]$ and $i \in [n]$ independently at random. We then construct estimators $\hat{K}_t^{il}(\hat{X}_t) := \frac{1}{B_l} \sum_{k=1}^{B_l} K_{t}^{ik} (\hat{X}_t^i, \hat{X}_t^{ik})$ with conditional covariance $\Sigma_t^{il}$ and their estimators $\hat{\Sigma}_t^{il}$. The final time evolution equation becomes:

$$\hat{X}_{t+1}^i = \hat{X}_t^i + \alpha g_l^i(\hat{X}_t^i) + \alpha \sum_{l=1}^{L} \hat{K}_t^{il}(\hat{X}_t) + \sqrt{\alpha} \left( \sqrt{\sigma^2 I - \alpha \sum_{l=1}^{L} \hat{\Sigma}_t^{il}} \right) Z_t^i$$

When $\mathcal{S}_t = \binom{[n]}{l}$ is the set of all $n$-tuples, we notice that this improves the computational complexity from $O(n^2)$ to $O(nB)$.

3.4. Covariance Estimators. While we can plug in any estimator $\hat{\Sigma}_t^i$ for the conditional covariance $\Sigma_t^i$, we consider two specific cases here. The first one is the widely used empirical covariance estimator and the second one is a modification which is easier to analyze. The idea here is to draw another random batch of size $B'$ and obtain an unbiased estimator for $\Sigma_t^i$ when conditioned on $\mathcal{F}_t$. For the sake of clarity, we will denote $K_t^{ij}(\hat{X}_t^i, \hat{X}_t^j)$ by just $K_t^{ij}$. These can be easily modified in the presence of singularities as outlined in Section 3.1.
1. Let \( B' \in [n] \). Draw \( J_{t}^{i,k} \) for \( k \leq B' \), \( i \in [n] \) i.i.d. uniformly at random from the set \([n]\). Define the estimator for the mean:

\[
\hat{K}_{i}^{t} := \frac{1}{B'} \sum_{k=1}^{B'} K_{i}^{J_{t}^{i,k}}.
\]

We construct the sample covariance estimator:

\[
\hat{\Sigma}_{i}^{t} := \frac{1}{B(B' - 1)} \sum_{k=1}^{B'} \left( K_{i}^{J_{t}^{i,k}} - \hat{K}_{i}^{t} \right) \left( K_{i}^{J_{t}^{i,k}} - \hat{K}_{i}^{t} \right)^{T}.
\]

We will call this **Option I**.

2. Let \( B' \in [n] \). Draw \( J_{t}^{i,k}, J_{t}^{i,k} \) for \( k \leq B' \) i.i.d uniformly at random from the set \([n]\). Construct the estimator:

\[
\hat{\Sigma}_{i}^{t} := \frac{1}{2BB'} \sum_{k=1}^{B'} \left( K_{i}^{J_{t}^{i,k}} - K_{i}^{J_{t}^{i,k}} \right) \left( K_{i}^{J_{t}^{i,k}} - K_{i}^{J_{t}^{i,k}} \right)^{T}.
\]

We will call this **Option II**.

For the ease of exposition, we will only analyze **Option II** in detail. Lemma 3.2 below notes some important properties of the covariance estimator considered. We defer the proof to Section 6.

**Lemma 3.2.**

1. \( \text{Tr}(\hat{\Sigma}_{i}^{t}) \leq \frac{4M^{2}}{B} \)
2. For **Option I** and **Option II**, we have: \( \mathbb{E} \left[ \hat{\Sigma}_{i}^{t} | \mathcal{F}_{t} \right] = \Sigma_{i}^{t} \)
3. For **Option II**:

\[
\mathbb{E} \left[ \text{Tr}((\hat{\Sigma}_{i}^{t})^{2}) | \mathcal{F}_{t} \right] - \text{Tr}((\Sigma_{i}^{t})^{2}) \leq \frac{4M^{4}}{B^{2}B'}
\]

4. For **Option II** and any \( k \in \mathbb{N} \), we have:

\[
\text{Tr}((\hat{\Sigma}_{i}^{t})^{k}) \leq \frac{2^{k}M^{2k}}{B^{k}}
\]

**3.5. Data Structure to Handle Singularities at the Origin.** Suppose the singularities of \( K_{i}^{j}(x,y) \) (when \( i \neq j \)) occur when \( x = y \) and that \( \{ (x, y) : \| K_{i}^{j}(x,y) \| \geq M \} = \{(x, y) : \| x - y \| \leq \frac{1}{\beta} \} \) for some known \( \beta \in \mathbb{R}^{+} \) which is small enough. To compute the singularity aware random batch estimator proposed in Section 3.1, we want to efficiently compute a set \( S_{i}(M) \in [n] \) such that for any \( j \in [n] \) such that \( \| K_{i}^{j}(X_{i}^{t}, X_{i}^{t}) \| > M \), we have \( j \in S_{i}(M) \).

To this end, consider the partition of \( \mathbb{R}^{d} \) into cubes of the form \( B_{m} := \times_{i=1}^{d} \left[ \frac{m_{i}}{\beta}, \frac{m_{i} + 1}{\beta} \right) \) for some fixed \( \beta \in \mathbb{R}^{+} \) and \( m = (m_{1}, \ldots, m_{d}) \in \mathbb{Z}^{d} \). We fix a total order in \( \mathbb{Z}^{d} \) (say the dictionary order). At each time \( t \), for every \( i \in [n] \), we can efficiently find \( m_{i} \) such that \( X_{i}^{t} \in B_{m_{i}} \) with only \( O(nd) \) computations and construct the list \( (m_{i}, 1), \ldots, (m_{i}, n) \). We can sort this list in time \( O(dn \log n) \) with \( m_{i} \) along with the dictionary order as the criterion for sorting.

At time \( t \), we want to compute \( S_{i}(M) \). It is sufficient to set \( S_{i}(M) = \{ j : \| X_{i}^{t} - X_{i}^{j} \| \leq \frac{1}{\beta} \} \). Any particle within distance \( \frac{1}{\beta} \) from \( X_{i}^{t} \) must be in some cube \( B_{m} \) which is adjacent to \( B_{m_{i}} \) (There are \( 3^{d} \) such possibilities for \( m \). Since many
interesting cases have \( d \in \{2, 3\} \), this is still computationally efficient). We can query the existence of all these cubes in the sorted list via binary search in the sorted list constructed above, expending only \( O(d^3 \log n) \) computation time per particle. (Where \( d \log n \) is complexity of a single binary search and there are \( 3^d \) adjacent cubes to be queried.)

4. Assumptions and Results. In this work, the only assumption we make is that neither the particle trajectories given by the full dynamics (Equation (2.1)) nor the trajectories of CovRBM (Equation (3.6)) diverge to infinity in finite time. We do not assume that \( ||K_i^t||_\infty \leq M \), but use the modified dynamics as suggested in Section 3.1 along with the appropriately modified covariance estimator.

**Assumption 1.** The parameters of the dynamics are such that \( ||X_i^t|| < \infty \) and \( ||X_i^t|| < \infty \) almost surely for every \( i \in [n] \) and \( t \in \mathbb{N} \cup \{0\} \).

The proposition below shows that the assumption above is very weak. In fact, the assumption is automatically satisfied when the forces \( K_i^t \) and \( g_i^t \) do not have singularities and \( \alpha \) is small enough. For the sake of clarity, we will consider the domain of \( K_i^t \) to be \( \mathbb{R}^d \) instead of \( \mathbb{R}^{2d} \) since we intend to only consider its input as \((X_i^t, X_j^t)\). The proof follows from elementary measure theoretic arguments and we defer it to Section 6.

**Proposition 4.1.** Suppose the co-variance estimator is either from **Option I** or **Option II**. Define \( M_i^{ij} := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : ||K_i^{ij}(x, y)|| = \infty\} \) whenever \( i \neq j \), \( M_i^{ii} := \{x \in \mathbb{R}^d : ||K_i^{ii}(x, x)|| = \infty\} \) and \( G_i^t = \{x \in \mathbb{R}^d : ||g_i^t(x)|| = \infty\} \). Let \( \Lambda_k \) denote the Lebesgue measure over \( \mathbb{R}^k \). Assumption 1 holds if the following conditions hold:

1. \( \frac{4M_i^{ii}2M_i^{ij}}{\sigma^6 B^{3}} < 1 \) (under **Option I**) or \( \frac{2M_i^{ii}}{\sigma^6 B^{3}} < 1 \) (under **Option II**)
2. \((X_i^t, X_j^t) \in (M_i^{ij})^c \), \((X_i^t, X_j^t) \in (M_i^{ii})^c \) and \((X_i^t, X_j^t) \in (G_i^t)^c \) almost surely for every \( i, j \in [n] \) with \( i \neq j \).
3. \( \Lambda_{d}(M_i^{ij}) = 0 \) and \( \Lambda_{d}(G_i^t) = 0 \), \( \Lambda_{d}(M_i^{ii}) = 0 \) for every \( i, j, t \) with \( i \neq j \).

We now state the main result regarding entropic convergence:

**Theorem 4.2.** Let \((X_i^t)_{t \in [n], t \leq T} \) be the iterates of the interacting particle method (Equation (2.1)) and \((\tilde{X}_i^t)_{t \in [n], t \leq T} \) be the iterates of CovRBM with **Option II**. Suppose that \( \frac{M_i^{ii}}{\sigma^6 B^{3}} < 1 \), \( \frac{2M_i^{ii}}{\sigma^6 B^{3}} < 1 \) and that Assumption 1 holds. Then, for universal constants \( C_{\text{CLT}} \) and \( C_{\text{Cov}} \), we have:

\[
\text{KL} \left( (\tilde{X}_i^t)_{t \leq T, i \in [n]} \parallel (X_i^t)_{t \leq T, i \in [n]} \right) \leq \frac{C_{\text{CLT}}nTdM^6\alpha^3(1 + \log B)^2}{\sigma^6 B^3} + C_{\text{Cov}} \left[ \frac{M^4\alpha^2nT}{B^2B^\prime\sigma^4} + \frac{\alpha^3M^6nT}{\sigma^6 B^3} \right]
\]

\[
= \frac{C_{\text{CLT}}nKdM^6\alpha^2(1 + \log B)^2}{\sigma^6 B^3} + C_{\text{Cov}} \left[ \frac{M^4\alpha nK}{B^2B^\prime\sigma^4} + \frac{\alpha^3M^6nK}{\sigma^6 B^3} \right]
\]

(4.1)

**Remark 4.3.** Pinsker’s inequality states that the total variation distance \( \text{TV}(\mu, \nu) \leq \sqrt{2\text{KL}(\nu || \mu)} \). Therefore, whenever \( B = B' \) and \( B \gg (\alpha n)^{1/3} \), the output of CovRBM and the interacting particle method are nearly statistically indistinguishable.
Remark 4.4. For the case of Equation (2.2), we take the batch sizes $B_1, \ldots, B_L = B$ and replace $M$ in the conditions and bounds above by $ML$ to conclude similar bounds.

We provide the proof in Section 5. The proof has 3 main ingredients as outlined below.

Random Function Representation. The collection of trajectories of the particles up to step $T$ given by Equation (2.1) can be seen as a function of the input i.i.d. Gaussian random variables - $Z_1, \ldots, Z_T$ along with the initial conditions $X_0$. CovRBM process does the same updates but with randomness $\hat{Z}_1, \ldots, \hat{Z}_T$ as defined in Section 3.2. Under Assumption 1, we can find a measurable function $F_T: \mathbb{R}^{d \times (T+1) \times n} \rightarrow \mathbb{R}^{d \times T \times n}$ such that almost surely:

$$(X_i^t)_{1 \leq t \leq T, i \in [n]} = F_T \left( (Z_i^s)_{j \in [n], 0 \leq s < T}, (X_0^j)_{j \in [n]} \right)$$

Similarly, we have almost surely:

$$(\hat{X}_i^t)_{t \leq T, i \in [n]} = F_T \left( (\hat{Z}_i^s)_{j \in [n], 0 \leq s < T}, (X_0^j)_{j \in [n]} \right)$$

Such representations of Markov chain trajectories is usually called the random function representation.

Data Processing Inequality. Roughly speaking, due to the random function representation above, the data processing inequality implies that the trajectories $X_i^t$ and $\hat{X}_i^t$ are nearly statistically indistinguishable if the joint distribution $(Z_i^s)$ and the joint distribution $(\hat{Z}_i^s)$ for $i \in [n]$ and $0 \leq s < T$ are nearly statistically indistinguishable. We use the KL divergence to quantify ‘statistical indistinguishability’. Due to the chain rule (see Lemma 6.1), it is sufficient to show that the conditional law $Z_i^s | F_s$ is close in KL divergence to the law of $\hat{Z}_i^s$ for every $s$ and $i$. That is, we need to show an entropic CLT for $Z_i^s | F_s$.

Entropic CLT. Recall that $\hat{Z}_i^s = \sqrt{\sigma_i^s} N_i^s + \sqrt{I - \frac{\sigma_i^s}{\sigma_0^s}} \hat{V}_i^s Z_i^s$. When conditioned on $F_s$, $N_i^s$ is an average of i.i.d random variables and has conditional covariance $\Sigma_i^s$. Due to the central limit theorem ([35]), $N_i^s$ looks approximately Gaussian in the $W_2$ distance and hence we can expect $\sqrt{\sigma_i^s} N_i^s + \sqrt{I - \frac{\sigma_i^s}{\sigma_0^s}} \hat{V}_i^s Z_i^s$ to look approximately like $\mathcal{N}(0, I)$ with respect to KL divergence. In statistical terms, we are hiding a signal which looks approximately Gaussian in a noise which is exactly Gaussian - and this is hard to distinguish from a purely Gaussian noise. We decompose the KL divergence between $Z_i^s$ and $\hat{Z}_i^s | F_s$ into two parts:

1. Error arising due to the fact that $\hat{\Sigma}_i^s \neq \Sigma_i^s$. We will handle this using Lemma 3.2, which shows the closeness of $\hat{\Sigma}_i^s$ to $\Sigma_i^s$ along with the standard formula for KL divergence between two multi-variate Gaussian measures.
2. Error arising due to the fact that $N_i^s$ is approximately Gaussian but not exactly Gaussian. This can be handled by lifting the $W_2$ CLT shown by [35] into an entropic CLT via. a reverse T2 type inequality (Lemma 6.2).

5. Proof of Theorem 4.2.

Proof of Theorem 4.2. For the ease of notation, we will take the initial condition $X_0^j$ to be fixed constants. The proof for random initial positions follows by the convexity of KL divergence. By Assumption 1, we can find functions $f_i^t: \mathbb{R}^{d \times T \times n} \rightarrow \mathbb{R}^d$
such that \( X_t^i = f_t^i ((Z_t^j)_{j \in [n], 0 \leq s < t}) \). Therefore, we can find a measurable function \( F_T : \mathbb{R}^{d \times T \times n} \rightarrow \mathbb{R}^{d \times T \times n} \) such that almost surely:

\[
(X_t^i)_{1 \leq t \leq T, i \in [n]} = F_T ((Z_t^j)_{j \in [n], 0 \leq s < t})
\]

Similarly, we have almost surely:

\[
(\hat{X}_t^i)_{t \leq T, i \in [n]} = F_T ((\hat{Z}_t^j)_{j \in [n], 0 \leq s < t})
\]

By the data processing inequality for f-divergences, we must have:

\[
(5.1) \quad \text{KL} \left( (\hat{X}_t^i)_{t \leq T, i \in [n]} \middle| (X_t^i)_{t \leq T, i \in [n]} \right) \leq \text{KL} \left( (\hat{Z}_t^j)_{0 \leq t < T, i \in [n]} \middle| (Z_t^j)_{0 \leq t < T, i \in [n]} \right)
\]

Let \( \mu \) denote the standard normal distribution over \( \mathbb{R}^d \) and let \( \nu_t^i \) denote the law of \( \hat{Z}_t^i \). Recall the filtration \( (\mathcal{F}_t) \). It is clear that for a given \( t \), \( (\hat{Z}_t^j)_{i \in [n]} \) are conditionally independent when conditioned on \( \mathcal{F}_t \). By \( \nu_t^i (\cdot | \mathcal{F}_t) \), we denote the law of \( \hat{Z}_t^i \) conditioned on \( \mathcal{F}_t \). We apply Lemma 6.1 to conclude:

\[
(5.2) \quad \text{KL} \left( (\hat{Z}_t^j)_{0 \leq t < T, i \in [n]} \middle| (Z_t^j)_{0 \leq t < T, i \in [n]} \right) = \sum_{t=0}^{T-1} \sum_{i=1}^{n} \text{EKL} \left( \nu_t^i (\cdot | \mathcal{F}_t) \middle| \mu \right)
\]

Here, we have used the random function representation to argue that \( \mathcal{F}_t = \sigma (\hat{Z}_s : 0 \leq s \leq t - 1) \). We now consider another sequence of \( \sigma \)-algebras defined by \( \mathcal{G}_t := \mathcal{F}_t \vee \sigma (J_t^{ij}, j \in [B^t]) \), where \( J_t^{ij} \) are the additional random variables used in the estimator \( \hat{\Sigma}_t^i \). By Jensen’s inequality for the KL-divergence, we must have:

\[
(5.3) \quad \sum_{t=0}^{T-1} \sum_{i=1}^{n} \text{EKL} \left( \nu_t^i (\cdot | \mathcal{F}_t) \middle| \mu \right) \leq \sum_{t=0}^{T-1} \sum_{i=1}^{n} \text{EKL} \left( \nu_t^i (\cdot | \mathcal{G}_t) \middle| \mu \right)
\]

Now, by construction, note that \( \hat{Z}_t^i := \frac{\sqrt{n}}{\sigma} N_t^i + \sqrt{I - \frac{n \hat{\Sigma}_t^i}{\sigma^2}} Z_t^i \). Here, when conditioned on \( \mathcal{G}_t \), \( \hat{\Sigma}_t^i \) is a constant, \( Z_t^i \) is distributed as \( \mathcal{N}(0, I) \), and \( N_t^i \) is an average of \( n \) i.i.d. random variables. Furthermore, \( Z_t^i \) is independent of \( N_t^i \) conditioned on \( \mathcal{G}_t \). Therefore, we can apply Lemma 6.4 with \( \hat{\Sigma} \) replaced by \( \frac{\alpha \hat{\Sigma}_t^i}{\sigma} \), \( N \) replaced by \( \frac{\sqrt{n}}{\sigma} N_t^i \) and \( \Sigma \) replaced by \( \frac{\alpha}{\sigma} \Sigma_t^i \), where \( \Sigma_t^i = \frac{\alpha}{\sigma} \mathbb{E} \left[ N_t^i (N_t^i)^\top \bigg| \mathcal{F}_t \right] = \frac{\alpha}{\sigma} \mathbb{E} \left[ N_t^i (N_t^i)^\top \bigg| \mathcal{G}_t \right] \), and \( Z_2 \) is standard Gaussian independent of \( N_t^i \) and \( \mathcal{G}_t \) defined on the same measure space, to
conclude:

\[
\text{KL} (\nu_i^t(\cdot | G_t) || \mu) \leq 2W_2^2 \left( \frac{Z_2}{\sqrt{2}}, \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2 + \frac{\sqrt{\alpha}}{\sigma} N_i^t} \mid G_t \right) \\
+ 2\lambda_n \left( \frac{1}{\sqrt{2}} \frac{\alpha}{\sigma} \Sigma_Y \right) \text{KL} \left( \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2} \mid \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2} \mid G_t \right)
\]

\[
\leq 2W_2^2 \left( \frac{Z_2}{\sqrt{2}}, \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2 + \frac{\sqrt{\alpha}}{\sigma} N_i^t} \mid G_t \right) \\
+ 2\lambda_n \left( \frac{1}{\sqrt{2}} \frac{\alpha}{\sigma} \Sigma_Y \right) \text{KL} \left( \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2} \mid \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2} \mid G_t \right)
\]

(5.4)

Conditioned on \( G_t \), \( N_i^t \) is the average of \( B \) zero-mean i.i.d. random variables. That is, \( N_i^t = \frac{1}{B} \sum_{k=1}^B K_{i}^{t} \left( \hat{X}_i^t, \hat{X}_i^t \right) - E K_{i}^{t} \left( \hat{X}_i^t, \hat{X}_i^t \right), \) where \( \hat{X}_i^t \) are fixed and the randomness is only due to our choice of random indices \( I_t \). Taking \( Y_k = \frac{1}{\alpha \sqrt{B}} \left[ K_{i}^{t} \left( X_i^t, X_i^t \right) - E K_{i}^{t} \left( X_i^t, X_i^t \right) \right] \), we note that \( \| Y_k \| \leq M \frac{1}{\alpha \sqrt{B}} \) almost surely. Applying Theorem 7.1, we conclude that for some universal constant \( C_{\text{CLT}} \), almost surely:

\[
W_2^2 \left( Z_2, \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2 + \frac{\sqrt{\alpha}}{\sigma} N_i^t} \mid G_t \right) \leq C_{\text{CLT}} \frac{M^6 \alpha^3 d (1 + \log B)^2}{\sigma^6 B^4}
\]

(5.5)

By Lemma 6.5, we conclude:

\[
\text{KL} \left( \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2} \mid \sqrt{\frac{1}{T} I - \frac{\alpha}{\sigma} \Sigma_Y Z_2} \mid G_t \right)
\]

\[
= \mathbb{E} \text{Tr} \left( \left( I - 2 \frac{\alpha \Sigma_Y}{\sigma^2} \right)^{-1} \frac{\alpha \Sigma_Y}{\sigma^2} \right) + \sum_{k=1}^{\infty} 2^{k-1} \frac{1}{\alpha \sigma^2} \frac{\text{Tr}((\Sigma_i^k)^k) - \text{Tr}((\Sigma_i^k)^k)}{\sigma^{2k} k}
\]

\[
= \sum_{k=2}^{\infty} 2^{k-1} \frac{1}{\alpha \sigma^2} \frac{\text{Tr}((\Sigma_i^k)^k) - \text{Tr}((\Sigma_i^k)^k)}{\sigma^{2k} k}
\]

\[
= \frac{4 M^4 \alpha^2}{B^2 B' \sigma^4} + \sum_{k=3}^{\infty} \frac{32 \alpha^3 M^6}{3 \sigma^6 B^3 (1 - \frac{4 \alpha M^2}{\sigma B})} \leq C_{\text{cov}} \left[ \frac{M^4 \alpha^2}{B^2 B' \sigma^4} + \frac{\alpha^3 M^6}{3 \sigma^6 B^3} \right]
\]

(5.6)

Where \( C_{\text{cov}} \) is some universal constant. In the second step, we have used the fact that \( \Sigma_i^t \) is \( F_t \) measurable and \( \mathbb{E} \left[ \Sigma_i^t \mid F_t \right] = \Sigma_i^t \). Therefore the first term and the first term in
the infinite summation vanish. In the third step, we have used Lemma 3.2. In the last step, we have bounded the infinite series noting that \( \text{Tr}(\Sigma^k) \leq \frac{2^k M^{2k}}{D^k} \) (Lemma 3.2). We now use Equations (5.5) and (5.6) in Equation (5.4). This can be used to bound each term in Equation (5.3). Combined with Equations (5.1) and (5.2), we conclude the statement of the theorem.

\[ \square \]

6. Technical Lemmas. In this section, we will state and prove some important technical results which are used in the proof of Theorem 4.2. Lemma 6.1 below is a folklore result regarding chain rule for KL divergence. We refer to Lemma 4.18 in [32] for a more general statement of the result given below.

**Lemma 6.1.** Suppose \( \nu \) is a distribution over some Polish space \( \Xi^T \) and \( \mu \) be a product distribution over \( \Xi^T \) given as \( \mu = \otimes_{t=1}^T \mu_t \). Let \( \nu_t(\cdot | X_{<t}) \) denote the conditional distribution of the \( t \)-th co-ordinate conditioned on the co-ordinates \( 1, \ldots, t-1 \) (and the marginal of the first co-ordinate under \( \nu \) when \( t = 1 \)) and let \( \nu_{<t} \) denote the joint marginal law of the co-ordinates \( 1, \ldots, t-1 \) under the measure \( \nu \).

\[
\text{KL} (\nu || \mu) = \sum_{t=1}^T \mathbb{E}_{X_{<t} \sim \nu_{<t}} \text{KL} (\nu_t(\cdot | X_{<t}) || \mu_t)
\]

**Lemma 6.2.** Suppose \( Z \sim \mathcal{N}(0, \sigma^2 I) \), \( A, B \) are random variables independent of \( Z \). Then, \( \text{KL} (Z + A || Z + B) \leq \frac{1}{2\sigma^2} \mathcal{W}_2^2(A, B) \)

**Proof.** Suppose \( \Gamma \) be a \( \mathcal{W}_2 \) optimal coupling between the laws of \( A \) and \( B \). Let \( f : \mathbb{R}^d \rightarrow \mathbb{R} \) be defined by \( f(x) = \frac{1}{(2\pi \sigma^2)^{d/2}} \exp \left(-\frac{\|x-A\|^2}{2\sigma^2}\right) \). The density of \( Z + A \) with respect to the Lebesgue measure is given by \( P(x) = \mathbb{E}f(x-A) \) and similarly the density of \( Z + B \) is given by \( Q(x) = \mathbb{E}f(x-B) \). Therefore, the KL divergence can be written as:

\[
\text{KL} (Z + A || Z + B) = \int f(x-A) \log \left( \frac{\int f(x-A) d\Gamma(A \times B)}{\int f(x-B) d\Gamma(A \times B)} \right) d\Gamma(A \times B) dx
\]

\[
\leq \int f(x-A) \log \left( \frac{f(x-A)}{f(x-B)} \right) d\Gamma(A \times B) dx
\]

\[
= \frac{1}{2\sigma^2} \int f(x-A) \left[\|x-B\|^2 - \|x-A\|^2\right] d\Gamma(A \times B) dx
\]

\[
= \frac{1}{2\sigma^2} \int \|A-B\|^2 d\Gamma(A \times B) = \frac{1}{2\sigma^2} \mathcal{W}_2^2(A, B)
\]

(6.1)

The second step above follows from the log-sum inequality. The third step follows from the definition of \( f \). In the fourth step we have used Fubini’s theorem to integrate out \( x \), noting that \( f(x-A) \) is the density of a Gaussian with covariance \( \sigma^2 I \) and mean \( A \). We have finally used the fact that \( \Gamma \) is a \( \mathcal{W}_2 \) optimal coupling between \( A \) and \( B \).\( \square \)

The following lemma is a special case of tensorization of the \( \mathcal{T}_2 \) transport inequality. We refer to Proposition 1.8 in [21] for a full exposition.

**Lemma 6.3.** Suppose \( P \) is the law of \( \mathcal{N}(0, \Sigma) \) for some non-singular \( \Sigma \). Then for any probability measure \( Q \) over \( \mathbb{R}^d \), we have:

\[
\mathcal{W}_2^2(P, Q) \leq 2\lambda_{\max}(\Sigma) \text{KL} (Q || P)
\]

Consider two random variables \( Z, N \) with values in \( \mathbb{R}^d \) such that they are independent. Let \( Z \sim \mathcal{N}(0, I) \) and the covariance of \( N \), denoted by \( \Sigma \), satisfies \( \Sigma \prec \frac{1}{2} I \). Given any
PSD matrix $\mathbb{R}^{d \times d} \ni \Sigma \prec \frac{1}{2}$, we consider the KL divergence between the laws of the random variables $Z$ and $\hat{Z} := \sqrt{I - \Sigma}Z + N$. We have the following lemma:

**Lemma 6.4.** Suppose $Z_2 \sim \mathcal{N}(0, I)$ is independent of $N$. Then, 

$$ \text{KL} \left( \hat{Z} \parallel Z \right) \leq 2W_2^2 \left( \frac{Z_2}{\sqrt{2}} \sqrt{\frac{1}{2}I - \Sigma}Z_2 + N \right) $$

$$ + 4\lambda_{\max}(\frac{l}{2} - \Sigma)\text{KL} \left( \sqrt{\frac{1}{2}I - \Sigma}Z_2 \parallel \sqrt{\frac{1}{2}I - \Sigma}Z_2 \right) \tag{6.2} $$

**Proof.** If $Z_1, Z_2$ are i.i.d. isotropic Gaussians over $\mathbb{R}^d$, we can write: $Z = \frac{1}{\sqrt{2}}Z_1 + \frac{1}{\sqrt{2}}Z_2$ and $\hat{Z} = \frac{1}{\sqrt{2}}Z_1 + \sqrt{\frac{1}{2}I - \Sigma}Z_2 + N$, where $N$ is independent of both $Z_1$ and $Z_2$.

Applying Lemma 6.2 with $\sigma = \frac{1}{\sqrt{2}}$ and $Z$ replaced with $Z_1$, we have:

$$ \text{KL} \left( \hat{Z} \parallel Z \right) \leq W_2^2 \left( \frac{Z_2}{\sqrt{2}} \sqrt{\frac{1}{2}I - \Sigma}Z_2 + N \right) $$

$$ \leq 2W_2^2 \left( \frac{Z_2}{\sqrt{2}} \sqrt{\frac{1}{2}I - \Sigma}Z_2 + N \right) + 2W_2^2 \left( \sqrt{\frac{1}{2}I - \Sigma}Z_2 \parallel \sqrt{\frac{1}{2}I - \Sigma}Z_2 \right) $$

In the second step we have used the triangle inequality for $W_2$ along with the inequality $(x + y)^2 \leq 2x^2 + 2y^2$. The third step follows from the definition of $W_2$ by restricting to the optimal coupling between $\sqrt{I - \Sigma}Z_2$ and $\sqrt{I - \Sigma}Z_2$ only. The last step follows from an application of Lemma 6.3. \(\square\)

We will bound the first term with CLT for $W_2$ distance as given in Section 7. For the second term, we consider the explicit calculation below.

**Lemma 6.5.** 

$$ \text{KL} \left( \sqrt{\frac{1}{2}I - \Sigma}Z_2 \parallel \sqrt{\frac{1}{2}I - \Sigma}Z_2 \right) = \text{Tr} \left( (I - 2\Sigma)^{-1} (\Sigma - \hat{\Sigma}) \right) $$

$$ + \sum_{k=1}^{\infty} 2^{k-1} \left[ \frac{\text{Tr}(\hat{\Sigma}^k) - \text{Tr}(\Sigma^k)}{k} \right] \tag{6.3} $$

**Proof.** From standard formula for KL divergence between two multi-variate Gaussians, it follows that:

$$ \text{KL} \left( \sqrt{\frac{1}{2}I - \Sigma}Z_2 \parallel \sqrt{\frac{1}{2}I - \Sigma}Z_2 \right) = \frac{1}{2} \left[ \log \frac{\text{det}(\frac{1}{2}I - \Sigma)}{\text{det}(\frac{1}{2}I - \Sigma)} - d + \text{Tr} \left( (\frac{l}{2} - \Sigma)^{-1} (\frac{l}{2} - \hat{\Sigma}) \right) \right] $$

We first note that by basic algebraic manipulation,

$$ \text{Tr} \left( (\frac{l}{2} - \Sigma)^{-1} (\frac{l}{2} - \hat{\Sigma}) \right) - d = 2\text{Tr} \left( (I - 2\Sigma)^{-1} (\Sigma - \hat{\Sigma}) \right) $$
Now, consider $\log(\det(A))$ for a $d \times d$ PSD matrix $0 \prec A \prec I$. Taking $\lambda_1, \ldots, \lambda_d$ to be the eigenvalues of $A$, we have

$$\log(\det(A)) = \sum_{i=1}^{d} \log(\lambda_i) = \sum_{i=1}^{d} \log(1 + \lambda_i - 1) = \sum_{i=1}^{d} \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\lambda_i - 1)^k}{k}$$

(6.4)

$$= - \sum_{k=1}^{\infty} \text{Tr} \left[ (I - A)^k \right]$$

Therefore, $\log \frac{\det(\mathbf{I} - \mathbf{A})}{\det(\mathbf{I} - \mathbf{B})} = \log \frac{\det(\mathbf{I} - \mathbf{A})}{\det(\mathbf{I} - \mathbf{B})} = \sum_{k=1}^{\infty} \frac{1}{k} \left[ \text{Tr}(\mathbf{S}^k) - \text{Tr}(\mathbf{S}^k) \right]$, which proves the result. \hfill \square

6.1. Skipped Proofs.

Proof of Lemma 3.2. 1. From the definition of $\Sigma_i^t$, it is clear that:

$$\text{Tr}(\Sigma_i^t) = \frac{1}{B^2} \sum_{k,l=1}^{B} \mathbb{E} \left[ (K_i^{t,i} - K_i^t)\top (K_i^{t,i} - K_i^t) \mid \mathcal{F}_t \right]$$

$$= \frac{1}{B^2} \sum_{k=1}^{B} \mathbb{E} \left[ (K_i^{t,i} - K_i^t)\top (K_i^{t,i} - K_i^t) \mid \mathcal{F}_t \right]$$

$$= \frac{1}{B^2} \mathbb{E} \left[ (K_i^{t,i} - K_i^t)\top (K_i^{t,i} - K_i^t) \mid \mathcal{F}_t \right]$$

$$= \frac{1}{B^2} \mathbb{E} \left[ (K_i^{t,i})\top (K_i^{t,i}) \mid \mathcal{F}_t \right] - \frac{1}{B}(K_i^t)\top K_i^t$$

(6.5)

In the second step, we have used the fact that conditioned on $\mathcal{F}_t$, $K_i^{t,i}$ are i.i.d. random variables with mean $K_i^t$, and hence the terms where $k \neq l$ have zero expectation. In the third step, we have used the fact that all the terms in the summation in previous step are equal. In the fourth step, we have used the fact that $\|K_i^{t,i}\|^2 \leq M$ for every $j$, almost surely.

2. This follows from elementary calculations.

3. Under Option II, consider:

$$\langle \mathbf{\hat{S}}_i^t \rangle^2 = \frac{1}{4B^2(B')^2} \sum_{k=1}^{B'} \left( \mathbf{\hat{H}}_{i}^{t,k} \right)^2 + \frac{1}{4B^2(B')^2} \sum_{k,l \in [B'], k \neq l} \mathbf{\hat{H}}_{i}^{t,k} \mathbf{\hat{H}}_{i}^{t,l}$$

Where $\mathbf{\hat{H}}_{i}^{t,k} := \left( K_i^{t,i} - K_i^{t,j} \right) \left( K_i^{t,j} - K_i^{t,j} \right)\top$. Note that whenever $k \neq l$, by independent sampling we have: $\mathbb{E} \left[ \mathbf{\hat{H}}_{i}^{t,k} \mathbf{\hat{H}}_{i}^{t,l} \mid \mathcal{F}_t \right] = 4(\Sigma_i^t)^2$. Therefore, from the equation above, we conclude:

$$\mathbb{E} \left[ \langle \mathbf{\hat{S}}_i^t \rangle^2 \mid \mathcal{F}_t \right] - \langle \Sigma_i^t \rangle^2 = \frac{\mathbb{E} \left[ (\mathbf{\hat{H}}_{i}^{t,1})^2 \right]}{4B^2(B')} - \frac{(\Sigma_i^t)^2}{B^2B'}$$

(6.6)

Considering the trace on both sides and noting that $\text{Tr} \left( (\mathbf{\hat{H}}_{i}^{t,1})^2 \right) \leq 16M^4$, we conclude the result.
4. First, note that for any PSD matrix $A$ and $k \in \mathbb{N}$, $\text{Tr}(A^k) \leq \text{Tr}(A)^k$. Therefore, we conclude that:

$$\text{Tr}((\hat{\Sigma}_i^k)^k) \leq \text{Tr}(\hat{\Sigma}_i^k)^k \leq \frac{2^k M^{2k}}{B^k}$$

*Proof of Proposition 4.1.* First, we only consider the interacting particle method without random batches. We suppose the stochastic process $(X_t^i)_{t, i}$ given by Equation (2.1) is defined over some Polish probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Over this probability space, consider the events $C_t := \bigcup_{i,j \in [n]} \{(X_t^i, X_t^j) \in M_t^{ij}\} \cup \{X_t^i \in G_t^i \cup M_t^i\}$ and the event $D_t := \bigcup_{i \in [n]} \{|X_t^i| = \infty\}$. By the established convention for divergence, we must have: $\bigcup_{t=0}^{T} D_t := \bigcup_{t=0}^{T} C_t$ for every $T$.

Therefore, to show the claim, it is sufficient to show that $\mathbb{P}(C_t) = 0$ for every $t$. We will prove this by induction. Clearly, by the choice of initial conditions, $\mathbb{P}(C_0) = 0$. Now, suppose $\mathbb{P}(C_0) = \cdots = \mathbb{P}(C_{t-1}) = 0$. This implies $\mathbb{P}(D_{t-1}) = 0$.

$$\mathbb{P}(X_t^i \in G_t^i) = \mathbb{E}\mathbb{P}(X_t^i \in G_t^i \mid X_{t-1})$$

$$= \mathbb{E}\mathbb{P}(\sqrt{\alpha} \sigma Z_{t-1}^i \in G_t^i - X_t^i - \alpha g_{t-1}^i - \alpha K_{t-1}^i \mid X_{t-1}) \mathbb{1}(D_{t-1}^i)$$

$$= 0$$

Here, $K_t^i := \frac{1}{n} \sum_{j=1}^{n} K_t^{ij}(X_t^i, X_t^j)$ and $g_t^i := g_t^i(X_t^i)$. In the first step, we have expressed probability as the expectation of conditional probability. In the second step, we have used the fact that $D_{t-1}^i$ has measure 1 by induction hypothesis. In the third step, by $S - x$ for $S \subset \mathbb{R}^d$ and $x \in \mathbb{R}^d$ we mean $\{y - x : y \in S\}$. The third step follows from the fact that $Z_{t-1}^i$ is standard Gaussian independent of the sigma algebra $X_{t-1}$ and the Lebesgue measure of $G_t^i - X_t^i - \alpha g_{t-1}^i - \alpha K_{t-1}^i$ is 0. We add the factor $\mathbb{1}(D_{t-1}^i)$ to ensure that this set is well defined. A similar argument using the continuity of the Gaussian measure with respect to the Lebesgue measure shows that $\mathbb{P}((X_t^i, X_t^j) \in M_t^{ij}) = 0$, which implies that $\mathbb{P}(C_t) = 0$. By induction, we conclude that the interacting particle dynamics does not diverge to $\infty$ in finite time. The non-divergence of CovRBM, can be shown using similar continuity arguments once we note that $\frac{\alpha \sum K}{\sigma^2} \prec I$ almost surely whenever the first condition in the statement of the proposition holds. This ensures that conditioned on $X_{t-1}$, $\hat{Z}_t^i$ and $(\hat{Z}_t^i, \hat{Z}_t^j)$ have densities with respect to $\lambda_d$ and $\lambda_{2d}$ respectively, almost surely. This allows us use the same arguments as above to ensure the non-divergence of CovRBM in finite time.

7. **Improved Wasserstein CLT with Gaussian Convolution.** The reference [35] proves a high dimensional CLT with respect to $\mathcal{W}_2$ under the assumption of bounded i.i.d. random variables. In our case, stronger conditions are satisfied, which we can use to derive sharper bounds by adapting the proof. Since the adaptation is straightforward, we only give a brief sketch below for the sake of completeness.

Let $Z$ be a standard normal variable in $\mathbb{R}^d$ and let $Y_1, \ldots, Y_B$ be i.i.d. zero mean random vectors in $\mathbb{R}^d$, independent of $Z$ such that $\|Y_i\|_2 \leq \beta$ almost surely. Let $\Sigma_Y$ be the co-variance of $Y_1$. Define the random vector $\hat{Z} := \sqrt{I - \Sigma_Y} Z + \sum_{k=1}^{B} \frac{Y_k}{\sqrt{B}}$. Notice that $\hat{Z}$ can be written as:

$$\hat{Z} \overset{d}{=} \frac{1}{\sqrt{B}} \sum_{k=1}^{B} \sqrt{I - \Sigma_Y} Z_k + Y_k$$
Where \( Z_k \sim \mathcal{N}(0, I) \) are i.i.d and independent of \((Y_l)_{l=1}^B\). The goal of this section is to bound the quantity: \( W_2(Z, \hat{Z}) \). Without loss of generality, assume that \( \Sigma_Y \) is diagonal with entries \( \varsigma_i^2 \) for \( i \in [d] \).

**Theorem 7.1.** \( 5\beta^2 \leq 1, \sup_i 5\varsigma_i^2 d < 1 \). Then we have:

\[
W_2(Z, \hat{Z}) \leq 5\beta^3 \sqrt{d} (1 + \log B)
\]

In order to prove Theorem 7.1, we will first consider \( W_2 \left( Z, \sqrt{I - \Sigma_Y} Z_k + \frac{1}{\sqrt{k}} Y_k \right) \), analogous to Lemma 1.6 in \([35]\).

**Lemma 7.2.** Assume that \( 5\beta^2 \leq 1, \sup_i 5\varsigma_i^2 d < 1 \). Then, we must have:

\[
W_2 \left( Z_1, \sqrt{I - \frac{\Sigma_Y}{k}} Z_1 + \frac{1}{\sqrt{k}} Y_1 \right) \leq \frac{25d\beta^6}{k^3}
\]

**Proof of Theorem 7.1.** The proof proceeds similar to the proof of Theorem 1.1 in \([35]\) using Lemma 7.2 instead. First, by the property that a sum of independent Gaussian random variables is also a Gaussian random variable, we have:

\[
W_2 \left( \sum_{j=1}^k Z_j, \sum_{j=1}^{k-1} Z_j + \sqrt{I - \Sigma_Y} Z_k + Y_k \right) = W_2 \left( \sqrt{k} Z_1, \sqrt{k} \left( \sqrt{I - \frac{\Sigma_Y}{k}} Z_1 + Y_k \right) \right)
\]

\[
= \sqrt{k} W_2 \left( Z_1, \sqrt{I - \frac{\Sigma_Y}{k}} Z_1 + \frac{Y_k}{\sqrt{k}} \right) \leq \sqrt{\frac{25d\beta^6}{k^3}}
\]

(7.1)

In the third step, we have applied Lemma 7.2. Now, by triangle inequality, we must have:

\[
\sqrt{B} W_2 \left( Z, \hat{Z} \right) = W_2 \left( \sum_{j=1}^B Z_j, \sum_{j=1}^B \sqrt{I - \Sigma_Y} Z_j + Y_j \right)
\]

\[
\leq \sum_{k=1}^B W_2 \left( \sum_{j=1}^k Z_j + \sum_{j=k+1}^B \sqrt{I - \Sigma_Y} Z_j + Y_j, \sum_{j=1}^{k-1} Z_j + \sum_{j=k}^B \sqrt{I - \Sigma_Y} Z_j + Y_j \right)
\]

\[
\leq \sum_{k=1}^B W_2 \left( \sum_{j=1}^k Z_j + \sum_{j=1}^{k-1} Z_j + \sqrt{I - \Sigma_Y} Z_k + Y_k \right) \leq \sum_{k=1}^B \frac{5\beta^3 \sqrt{d}}{k}
\]

(7.2) \( \leq 5\beta^3 \sqrt{d} (1 + \log B) \)

In the third step, we have used the fact that \( W_2(Z + A, Z + B) \leq W_{A,B} \) whenever \( Z, A \) and \( Z, B \) are independent random variables.

We will dedicate the rest of this section to sketching the proof of Lemma 7.2, based on the proof of Lemma 1.6 in \([35]\).
7.1. Proving Lemma 7.2. Define $n_i := \frac{1}{\sqrt{n}}$. Suppose $Y, Z$ are independent random variables such that $Z \sim N(0, I)$ and $Y$ has the same distribution as any one of $Y_1, \ldots, Y_B$ used in the statement of Theorem 7.1. In this section only, for any vector $x \in \mathbb{R}^d$, we will let $x_i$ denote its component along the $i$-th standard basis vector. Notice that the $i$-th co-ordinate of $\sqrt{I - \frac{2}{\kappa} Z} + \frac{1}{\sqrt{k}} Y$ is given by $\sqrt{1 - \frac{1}{n_i} Z_i + \bar{Y}_i}$, where, $\bar{Y} := \frac{Y}{\sqrt{k}}$. Clearly, $\|\bar{Y}\| \leq \frac{\beta}{\sqrt{k}}$, a fact which we will use heavily below. We will also use the observation that $\sum_i \varsigma_i^2 \leq \beta^2$ and $\beta^2 n_j \geq k$.

Let $f(x) := \frac{r(x)}{\rho(x)}$ where $\tau$ is the density of $\sqrt{1 - \frac{2\varsigma}{\kappa} Z_1 + \bar{Y}_1}$ and $\rho$ is the density of $Z_1$. Proceeding similar to the proof of Lemma 4.1 in [35], but with $n$ replaced with co-ordinate dependent $n_i$ in our case, and $\sigma_i$ replaced with 1, we have:

**Lemma 7.3.** Suppose we have $\sup_i \varsigma_i^2 < 1$. Then:

$$\mathbb{E}[f(Z_1)^2] = \mathbb{E}_{\bar{Y}, \bar{Y}'} \left[ \exp \left( \sum_{i=1}^d \frac{2n_i^2 \bar{Y}_i \bar{Y}'_i - n_i \bar{Y}_i^2 - n_i (\bar{Y}_i')^2 + 1}{2(n_i^2 - 1)} - r(n_i) \right) \right]$$

Where $\bar{Y}'_i$ is an independent copy of $\bar{Y}$, and $r(n) := \frac{1}{2(n^2 - 1)} - \frac{1}{2} \log(1 + \frac{1}{n^2 - 1})$

Proceeding similarly, we let $Q_i := \frac{2n_i^2 \bar{Y}_i \bar{Y}'_i - n_i \bar{Y}_i^2 - n_i (\bar{Y}_i')^2 + 1}{2(n_i^2 - 1)} - r(n_i)$ and $Q := \sum_{i=1}^d Q_i$.

Let $f_{ij}(x)$ be the ratio $\frac{\tau_{ij}(x)}{\tau_{i}(x)}$, where $\tau_{ij}(x)$ denotes the marginal density under $\tau$ of all co-ordinates other than the $i$-th co-ordinate. The following lemma is a rewriting of Lemmas 4.4 and 4.5 in [35]

**Lemma 7.4.** $\sup_i 5\varsigma_i^2 d < 1$, $5\beta^2 \leq 1$. Then, the following bounds hold:

1. $|Q_i| \leq \frac{n_i^2 |Y_i||\bar{Y}'_i|}{n_i^2 - 1} + \frac{1}{2n_i}$, $|Q| \leq 1$, $|Q - Q_i| \leq 1$

2. $\mathbb{E}Q_i = -\frac{1}{2(n_i^2 - 1)} - r(n_i)$

3. $\mathbb{E}Q_i, Q_j \leq \frac{n_i^2 \delta_{ij}}{(n_i^2 - 1)^2} + \frac{n_i n_j \bar{Y}_i \bar{Y}'_j}{2(n_i^2 - 1)(n_j^2 - 1)} + \frac{1}{2(n_i^2 - 1)(n_j^2 - 1)}$

4. $\mathbb{E}Q_i^2 \leq \frac{2n_i^2 + n_i + 1}{2(n_i^2 - 1)^2}$

5. $\mathbb{E}(Q - Q_i)Q_i \leq \sup_j \frac{k(d - 1) + \beta^2 n_j}{2k(n_j^2 - 1)^2}$.

6. $\mathbb{E}Q^2 \leq \sup_j \frac{\beta^2 n_j d + 3n_j^2 kd}{2k(n_j^2 - 1)^2} \leq \sup_j \frac{2n_j^2 d}{(n_j^2 - 1)^2}$

**Proof.** Items 1 - 4 and item 6 can be shown by essentially the same methods as
in the original proof. For item 5, we have:

\[
\mathbb{E}(Q - Q_i)Q_i = \sum_{j \neq i} Q_j Q_i \leq \sup_{i} \frac{d - 1}{2(n_i^2 - 1)^2} + \sum_{j \neq i} \frac{n_i n_j \mathbb{E} \bar{Y}_i^2 \bar{Y}_j^2}{2(n_i^2 - 1)(n_j^2 - 1)}
\]

\[
\leq \sup_{i} \frac{d - 1}{2(n_i^2 - 1)^2} + \sup_{i} \frac{n_i \mathbb{E} \bar{Y}_i^2}{2(n_i^2 - 1)^2} \sum_{j \neq i} n_j \beta^2 \bar{Y}_j^2
\]

\[
\leq \sup_{i} \frac{d - 1}{2(n_i^2 - 1)^2} + \sup_{i} \frac{n_i \beta^2}{2k(n_i^2 - 1)^2} \bar{Y}_i^2
\]

\[
= \sup_{i} \frac{n_i \beta^2 + k(d - 1)}{2k(n_i^2 - 1)^2}
\]

(7.3)

In the last step we have used the fact that \(\mathbb{E} \bar{Y}_i^2 = \frac{1}{n_i}\).

The proof of Lemma 7.2 now follows by using the bounds established above along with the proof of Lemma 1.6 in [35] and by noting that \(\beta^2 n_i^2 \geq k\) for every \(i \in [d]\).

Acknowledgments. The author would like to thank Katy Craig and Matthew Jacobs for introducing them to this problem and for multiple lengthy and extremely helpful discussions.

REFERENCES

[1] Giacomo Albi and Lorenzo Pareschi. Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Modeling & Simulation*, 11(1):1–29, 2013.
[2] Grégoire Allaire, Jean-François Dufrevêche, Andro Mikelić, and Andrey Piatnitski. Asymptotic analysis of the poisson–boltzmann equation describing electrokinetics in porous media. *Nonlinearity*, 26(3):881, 2013.
[3] Julio Backhoff, Giovanni Conforti, Ivan Gentil, and Christian Léonard. The mean field schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probability Theory and Related Fields*, 178(1):475–530, 2020.
[4] Andrea L Bertozzi, John B Garnett, and Thomas Laurent. Characterization of radially symmetric finite time blowup in multidimensional aggregation equations. *SIAM Journal on Mathematical Analysis*, 44(2):651–681, 2012.
[5] Charles K Birdsall and A Bruce Langdon. *Plasma physics via computer simulation*. CRC press, 2018.
[6] Eric Carlen, Pierre Degond, and Bernt Wennberg. Kinetic limits for pair-interaction driven master equations and biological swarm models. *Mathematical Models and Methods in Applied Sciences*, 23(07):1339–1376, 2013.
[7] JA Carrillo, F Hoffmann, AM Stuart, and U Vaes. Consensus-based sampling. *Studies in Applied Mathematics*, 2021.
[8] José A Carrillo, Young-Pil Choi, Claudia Totzeck, and Oliver Tse. An analytical framework for consensus-based global optimization method. *Mathematical Models and Methods in Applied Sciences*, 28(06):1037–1066, 2018.
[9] José A Carrillo, Katy Craig, and Yao Yao. Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In *Active Particles, Volume 2*, pages 65–108. Springer, 2019.
[10] JA Carrillo de la Plata, L Pareschi, and M Zanella. Particle based gpc methods for mean-field models of swarming with uncertainty. *Communications in Computational Physics*, 25(2), 2019.
[11] Yongxin Chen. Density control of interacting agent systems. *arXiv preprint arXiv:2108.07342*, 2021.
[12] Alexandre Joel Chorin. Numerical study of slightly viscous flow. *Journal of fluid mechanics*, 57(4):785–796, 1973.
[13] Katy Craig, Karthik Elamvazuthi, Matt Haberland, and Olga Turanova. A blob method method for inhomogeneous diffusion with applications to multi-agent control and sampling. *arXiv preprint arXiv:2202.12927*, 2022.
[14] Felipe Cucker and Steve Smale. Emergent behavior in flocks. *IEEE Transactions on automatic control*, 52(5):852–862, 2007.
[15] Esther S Daus, Markus Fellner, and Ansgar Jüngel. Random-batch method for multi-species stochastic interacting particle systems. arXiv preprint arXiv:2109.01897, 2021.

[16] Pierre Degond, Jian-Guo Liu, and Robert L Pego. Coagulation–fragmentation model for animal group-size statistics. Journal of Nonlinear Science, 27(2):379–424, 2017.

[17] Andrew Duncan, Nikolas Nüsken, and Łukasz Szpruch. On the geometry of stein variational gradient descent. arXiv preprint arXiv:1912.00894, 2019.

[18] Bradley Efron and Robert J Tibshirani. An introduction to the bootstrap. CRC press, 1994.

[19] Daan Frenkel and Berend Smit. Understanding molecular simulation: from algorithms to applications, volume 1. Elsevier, 2001.

[20] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[21] Nathael Gozlan and Christian Léonard. Transport inequalities: a survey. Markov Processes and Related Fields, 16:635–736, 2010.

[22] Seung-Yeal Ha and Jian-Guo Liu. A simple proof of the cucker-smale flocking dynamics and mean-field limit. Communications in Mathematical Sciences, 7(2):297–325, 2009.

[23] Dirk Horstmann. From 1970 until present: the keller-segel model in chemotaxis and its consequences. ii, jahresber. Deutsch. Math.-Verein., 106:51–69, 2004.

[24] Shi Jin, Lei Li, and Jian-Guo Liu. Random batch methods (rbm) for interacting particle systems. Journal of Computational Physics, 400:108877, 2020.

[25] Shi Jin, Lei Li, and Jian-Guo Liu. Convergence of the random batch method for interacting particles with disparate species and weights. SIAM Journal on Numerical Analysis, 59(2):746–768, 2021.

[26] Shi Jin, Lei Li, Zhenli Xu, and Yue Zhao. A random batch ewald method for particle systems with coulomb interactions. SIAM Journal on Scientific Computing, 43(4):B937–B960, 2021.

[27] Dongnam Ko, Seung-Yeal Ha, Shi Jin, and Doheon Kim. Uniform error estimates for the random batch method to the first-order consensus models with antisymmetric interaction kernels. Studies in Applied Mathematics, 146(4):983–1022, 2021.

[28] Jian-Guo Liu and Zhouping Xin. Convergence of the point vortex method for 2-d vortex sheet. Mathematics of computation, 70(234):595–606, 2001.

[29] Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. Advances in neural information processing systems, 29, 2016.

[30] Sebastien Motsch and Eitan Tadmor. Heterophilious dynamics enhances consensus. SIAM review, 56(4):577–621, 2014.

[31] James A Sethian. A brief overview of vortex methods. Vortex methods and vortex motion, pages 1–32, 1991.

[32] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.

[33] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. Physical review letters, 75(6):1226, 1995.

[34] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In Proceedings of the 28th international conference on machine learning (ICML-11), pages 681–688, Citeseer, 2011.

[35] Alex Zhai. A high-dimensional clt in $W^2$ distance with near optimal convergence rate. Probability Theory and Related Fields, 170(3):821–845, 2018.