CONVERGENCE OF A FIRST-ORDER CONSENSUS-BASED GLOBAL OPTIMIZATION ALGORITHM

SEUNG-YEAL HA, SHI JIN, AND DOHEON KIM

ABSTRACT. Global optimization of a non-convex objective function often appears in large-scale machine-learning and artificial intelligence applications. Recently, consensus-based optimization (in short CBO) methods have been introduced as one of the gradient-free optimization methods. In this paper, we provide a convergence analysis for the first-order CBO method in [5]. Prior to the current work, the convergence study was carried out for CBO methods on corresponding mean-field limit, a Fokker-Planck equation, which does not imply the convergence of the CBO method *per se*. Based on the consensus estimate directly on the first-order CBO model, we provide a convergence analysis of the first-order CBO method [5] without resorting to the corresponding mean-field model. Our convergence analysis consists of two steps. In the first step, we show that the CBO model exhibits a global consensus time asymptotically for any initial data, and in the second step, we provide a sufficient condition on system parameters–which is dimension independent– and initial data which guarantee that the converged consensus state lies in a small neighborhood of the global minimum almost surely.

1. INTRODUCTION

Large-scale optimization problems often appear in machine learning and artificial intelligence (AI) applications, in which objective functions to be optimized are not necessarily convex nor regular enough, say $\mathcal{C}^1$ in general. Thus, one might not be able to use the standard stochastic gradient descent method. Alternatively, several gradient-free optimization methods based on collective dynamics are used in application domain, for example swarm intelligence methods [15, 28] such as particle swarm optimization (in short PSO) [8], simulated annealing method [16, 20], ant-colony algorithm [27], genetic algorithm [13] etc. A basic idea of these metaheuristic algorithm is to use collective behaviors of underlying sample points coupled with suitable stochastic components in the choice of system parameters. Despite of its usefulness, rigorous convergence analysis of such swarm intelligence algorithms is often missing.

In this paper we provide a convergence analysis for the CBO method proposed in [5]. To be more specific, let $X_t^k = (x_t^{k,1}, \cdots, x_t^{k,d}) \in \mathbb{R}^d$ be the coordinate process of the k-th sample point at time t. Suppose that one looks for a global minimum $X^* \in \mathbb{R}^d$ for a given objective function L:

$$X^* \in \operatorname{argmin}_{X \in \mathbb{R}^d} L(X),$$

where the objective function L may be neither convex nor smooth enough. In this situation, gradient-based optimization methods such as the stochastic gradient descent method [4] can not be

used as it is. In a recent work [5], the authors proposed the following variant of the CBO algorithm introduced in [6, 23, 25]:

$$(1.1) \quad \begin{cases} dX_t^i = -\lambda(X_t^i - \bar{X}_t^*)dt + \sigma \sum_{l=1}^{d}(x_t^{i,l} - \bar{x}_t^{*,l})dW_t^l e_l, \quad t > 0, \ i = 1, \cdots, N, \\ \bar{X}_t^* = (x_t^{*,1}, \cdots, x_t^{*,d}) := \dfrac{\sum_{l=1}^{N} X_t^l e^{-\beta L(X_t^l)}}{\sum_{l=1}^{N} e^{-\beta L(X_t^l)}}, \end{cases}$$

where λ and σ denote the drift rate and noise intensity, respectively, and $\beta > 0$ is a positive constant corresponding to the reciprocal of temperature in statistical physics. Here $\{e_l\}$ is the standard orthonormal basis in $\mathbb{R}^d$. The one-dimensional Brownian motions W_t^l are i.i.d. and satisfy the mean zero and covariance relations:

$$\mathbb{E}[W_t^l] = 0 \quad \text{for } l = 1, \cdots, d \quad \text{and} \quad \mathbb{E}[W_t^{l_1} W_t^{l_2}] = \delta_{l_1 l_2} t, \quad 1 \le l_1, l_2 \le d.$$

Note that system (1.1) is expected to have a local relaxation dynamics which means that X_t^i relaxes toward the local weighted average $\bar{X}_t^*$, and this local weighted average tends to the global consensus state. This algorithm is an improvement upon those proposed in [6, 23] in that it is more suitable for higher-dimensional optimization problems, since its convergence conditions on system parameters are expected to be *independent* of the dimensionality d due to the use of *componentwise* geometric Brownian motion.

We denote by Ω the underlying sample space for model (1.1), and assume that the objective function L is locally Lipschitz continuous. For the optimization algorithm (1.1), we are interested in the following two questions:

- (Question I): Does the N-state ensemble $\{X_t^i\}$ exhibit a global consensus? i.e., for a.s. $\omega \in \Omega$, is there a global consensus state $X_\infty(\omega) \in \mathbb{R}^d$ such that

$$\lim_{t \to \infty} |X_t^i(\omega) - X_\infty(\omega)| = 0, \quad i = 1, \cdots, N,$$

  where $|\cdot| := \|\cdot\|_{\ell^2}$ is the standard ℓ^2-norm in $\mathbb{R}^d$?
- (Question II): If the answer to the first problem is positive, then under what condition the consensus state is a good approximation of the global minimum X^* of L?

In [5], the authors conducted a convergence analysis via the Fokker-Planck equation, which can be deduced from (1.1) in the mean-field limit $N \to \infty$, and showed that the global consensus state lies in a $\mathcal{O}(1/\beta)$-neighborhood of the global minimum under suitable assumptions on system parameters which are *independent* of dimension d and initial data for $\beta >> 1$. Since the Fokker-Planck equation is *not* the original model, thus this convergence result, although sheds lights on the convergence property of the original CBO model, it does not imply the latter. The purpose of this article is to conduct the convergence analysis of (1.1) *directly*.

Toward this goal, we first rewrite the relaxation term $X_t^i - \bar{X}_t^*$ (1.1) in consensus form:

$$(1.2) \quad dX_t^i = \lambda \sum_{k=1}^{N} \psi_t^k (X_t^k - X_t^i)dt + \sigma \sum_{k=1}^{N} \sum_{l=1}^{d} \psi_t^k (x_t^{k,l} - x_t^{i,l})dW_t^l e_l, \quad t > 0,$$

where $\psi_t^k := \psi^k(\mathcal{X}, t)$ is the communication weight function:

$$(1.3) \quad \psi_t^k := \frac{e^{-\beta L(X_t^k)}}{\sum_{l=1}^{N} e^{-\beta L(X_t^l)}}, \quad t \ge 0, \quad k = 1, \cdots, N.$$

Next, we return to the discrete-time dynamics associated with the continuous model (1.2)-(1.3). We set a time-step $h := \Delta t$ and state at discrete time $t = nh$:

$$X_n^i := X_{nh}^i, \qquad x_n^{i,l} := x_{nh}^{i,l}, \qquad \psi_n^i := \psi_{nh}^i, \quad n = 0, 1, \cdots.$$

Then, the discrete consensus-based optimization model reads as follows:

(1.4)
$$\begin{cases} X_{n+1}^i = X_n^i + \lambda h \sum_{k=1}^{N} \psi_n^k (X_n^k - X_n^i) + \sigma \sqrt{h} \sum_{k=1}^{N} \sum_{l=1}^{d} \psi_n^k (x_n^{k,l} - x_n^{i,l}) Z_n^l e_l, \quad n \geq 0, \\ \psi_n^k := \dfrac{e^{-\beta L(X_n^k)}}{\sum_{i=1}^{N} e^{-\beta L(X_n^i)}}, \qquad i = 1, \cdots, N, \end{cases}$$

where the random variables $\{Z_n^l\}_{n,l}$ are i.i.d standard normal distributions with $Z_n^l \sim \mathcal{N}(0,1)$.

We summarize the two main results of this paper now. First, we are concerned with the emergence of global consensus to the continuous and discrete models (1.1) and (1.4), respectively. Since the analysis for the discrete model (1.4) is almost parallel to the analysis for the continuous one (1.1), we mainly focus on the continuous model in what follows. To motivate the dynamic properties of (1.1) or (1.2), we consider the deterministic counterpart:

$$\frac{dX_t^i}{dt} = \lambda \sum_{k=1}^{N} \psi_t^k (X_t^k - X_t^i).$$

In this case, it is easy to see that the time-dependent convex hull generated by N sample points $X_t^1, \cdots, X_t^N$ in $\mathbb{R}^d$ is contractive (see Lemma 2.1). Moreover, due to the special structure of the communication weight ψ_t^k:

(1.5)
$$(i)\ \psi_t^k \geq 0, \quad 1 \leq k \leq N, \quad \sum_{k=1}^{N} \psi_t^k = 1 \quad \text{for all } t \geq 0,$$

(ii) Dependence only on the state of source sample point (independent of i in (1.2)),

the difference $X_t^i - X_t^j$ satisfies a system of ordinary differential equations:

$$\frac{d}{dt}(X_t^i - X_t^j) = -\lambda(X_t^i - X_t^j), \quad t > 0,$$

which has the analytic solution:

$$(X^i - X^j)(t) = e^{-\lambda t}(X_0^i - X_0^j), \quad t \geq 0.$$

On the other hand, maximal and minimal values of the component state of X_t^i are monotone so that they converge to the same value. Hence, we can show that the state X_t^i tends to the unique global consensus state X_∞ independent of i for any initial state (see Theorem 3.1).

Next, we return to the stochastic model with $\sigma > 0$. In this case, due to the white noise effect, the convex hull spanned by the state vector X_t^i is not contractive any more. However, fortunately the l-th components of the relative state difference $x_t^{i,l} - x_t^{j,l}$ satisfies the geometric Brownian motion:

$$d(x_t^{i,l} - x_t^{j,l}) = -\lambda(x_t^{i,l} - x_t^{j,l})dt - \sigma(x_t^{i,l} - x_t^{j,l})dW_t^l, \quad t > 0.$$

Then we use stochastic calculus to get the exact solution:

$$x_t^{i,l} - x_t^{j,l} = (x_0^{i,l} - x_0^{j,l}) \exp\left[-\left(\lambda + \frac{\sigma^2}{2}\right)t + \sigma W_t^l \right], \quad t \geq 0.$$

This yields the almost sure convergence of the relative state differences:

$$\lim_{t \to \infty} |x_t^{i,l} - x_t^{j,l}| = 0, \quad \text{a.s.}$$

Similar analysis can be performed for the discrete algorithm (1.4) (see Theorem 3.3). Moreover, under the condition $2\lambda > \sigma^2$–which is *independent of the dimensionalizty d*, we can show that there exists a random vector X_∞ which is the almost-sure limit of the X_t^i's (see Lemma 4.1). Thus we answered the first posed question affirmatively for the continuous and discrete models.

Second, we deal with Question II on whether the global consensus state X_∞ is close to the global minimum of L or not. Under suitable assumptions on system parameters λ, σ and initial data such that $X_0^i \sim X^{in}$ for some random variable X^{in}, we derive

$$\mathbb{E}e^{-\beta L(X_\infty)} \geq \varepsilon \mathbb{E}e^{-\beta L(X^{in})}, \quad \text{or} \quad -\frac{1}{\beta} \log \mathbb{E}e^{-\beta L(X_\infty)} \leq -\frac{1}{\beta} \log \mathbb{E}e^{-\beta L(X^{in})} - \frac{1}{\beta} \log \varepsilon.$$

If the global minimizer X^* of L is contained in $\text{supp}(\text{law}(X^{in}))$, then Laplace's principle yields the desired estimate (see Theorem 4.1):

$$\text{ess inf}_{\omega \in \Omega} L(X^\infty(\omega)) \leq L_m + \mathcal{O}\left(\frac{1}{\beta}\right) \quad \text{for} \quad \omega \in \Omega, \ \beta \gg 1.$$

Collective behaviors of agent-based models have been a hot topic in applied mathematics, control theory and related areas in recent years, see for example several survey articles [1, 3, 7, 24, 26] and related literature [9, 10, 11, 12, 17, 18, 19, 21, 22].

The rest of this paper is organized as follows. In Section 2, we provide preliminary materials on the deterministic analogs of the continuous and discrete algorithms (1.1) and (1.4). In Section 3, we study the emergence of global consensus states for the continuous and discrete consensus models. In Section 4, we prove the convergence of the global consensus state toward the global minimum of L as $\beta \to \infty$ only for the continuous algorithm. The corresponding convergence analysis for the discrete model seems to be very challenging, thus will be left for a future work. In Section 5, we provide several numerical examples and compare them with our analytical results. Finally, Section 6 is devoted to a brief summary of our main results and discussion on some remaining problems to be investigated in future study.

Notation. For a random variable $Z \in \mathbb{R}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote its mean by $\mathbb{E}Z$ or $\mathbb{E}[Z]$ interchangeably, and the function space $\mathcal{C}_b^2(\mathbb{R}^d)$ denotes the collection of all $\mathcal{C}^2(\mathbb{R}^d)$ functions with bounded derivatives up to second-order.

2. Preliminaries

In this section, we provide several preliminaries on the deterministic counterparts to (1.2)-(1.3) and (1.4) with $\sigma = 0$, respectively.

Consider the continuous model:

$$(2.1) \qquad \begin{cases} \dfrac{dX_t^i}{dt} = \lambda \displaystyle\sum_{k=1}^N \psi_t^k (X_t^k - X_t^i), \quad t > 0, \\ \psi_t^k \geq 0, \quad \displaystyle\sum_{k=1}^N \psi_t^k = 1, \quad i = 1, \cdots, N, \end{cases}$$

and the discrete model:

$$
(2.2) \quad
\begin{cases}
X_{n+1}^i = X_n^i + \lambda h \sum_{k=1}^N \psi_n^k (X_n^k - X_n^i), & n = 0, 1, \cdots, \\
\psi_n^k \geq 0, \quad \sum_{k=1}^N \psi_n^k = 1, & i = 1, \cdots, N.
\end{cases}
$$

Next, we study basic properties of the deterministic models (2.1) and (2.2) before moving to the stochastic ones.

2.1. Deterministic continuous algorithm.

Let $\mathcal{X}_t := (X_t^1, \cdots, X_t^N) \in \mathbb{R}^{Nd}$ be a solution to (2.1). For $t > 0$ and $l \in \{1, \cdots, d\}$, we introduce two extreme functions $\underline{x}^l$, $\bar{x}^l$ and component-diameter functional $\mathcal{D}^l(\mathcal{X}_t)$:

$$
\underline{x}_t^l := \min_{1 \leq j \leq N} x_t^{j,l}, \qquad \bar{x}_t^l := \max_{1 \leq j \leq N} x_t^{j,l}, \qquad \mathcal{D}^l(\mathcal{X}_t) := \bar{x}_t^l - \underline{x}_t^l.
$$

Note that trajectories of $\underline{x}^l$ and $\bar{x}^l$ are Lipschitz continuous, thus they are differentiable almost everywhere in $t \in (0, \infty)$.

Lemma 2.1. *Let* $\mathcal{X}_t = (X_t^1, \cdots, X_t^N)$ *be a solution to* (2.1) *with the initial data* $\mathcal{X}_0$. *Then, the following assertions hold.*

(1) *The extreme functions* $\underline{x}^l$ *and* $\bar{x}^l$ *are monotonically increasing and decreasing, respectively:*

$$
\underline{x}_t^l \geq \underline{x}_s^l \quad and \quad \bar{x}_t^l \leq \bar{x}_s^l, \quad for \; t \geq s.
$$

(2) *The component diameter functional* $\mathcal{D}^l(\mathcal{X})$ *is non-increasing in* t:

$$
\mathcal{D}^l(\mathcal{X}_t) \leq \mathcal{D}^l(\mathcal{X}_0), \quad t \geq 0.
$$

Proof. (i) Note that each component of (2.1) satisfies the same form of equations. Thus, it suffices to check one particular component. Consider the l-th component of system (2.1):

$$
(2.3) \quad \frac{dx_t^{i,l}}{dt} = \lambda \sum_{k=1}^N \psi_t^k (x_t^{k,l} - x_t^{i,l}).
$$

Now, we choose extreme indices i_t and j_t such that

$$
x_t^{i_t,l} = \underline{x}_t^l \quad and \quad x_t^{j_t,l} = \bar{x}_t^l.
$$

• Case A (Increasing property of $\underline{x}_t^l$): At a differentiable point t of $\underline{x}_t^l$, it follows from (2.3) that

$$
(2.4) \quad \frac{dx_t^{i_t,l}}{dt} = \lambda \sum_{k=1}^N \psi_t^k (x_t^{k,l} - x_t^{i_t,l}) \geq 0,
$$

where we used

$$
x_t^{k,l} - x_t^{i_t,l} \geq 0.
$$

Then, the Lipschitz continuity of $\underline{x}_t^l$ and (2.4) imply the non-increasing property of the lower envelope $\underline{x}_t^l$.

• Case B (Decreasing property of $\bar{x}_t^l$): Similar to Case A, one has

$$
\frac{dx_t^{j_t,l}}{dt} = \lambda \sum_{k=1}^N \psi_t^k (x_t^{k,l} - x_t^{j_t,l}) \leq 0.
$$

Finally, one combines Case A and Case B to see the non-increasing property of $\mathcal{D}^l(\mathcal{X}_t)$: for $t \geq s$,

$$\mathcal{D}^l(\mathcal{X}_t) = \bar{x}_t^l - \underline{x}_t^l \leq \bar{x}_s^l - \underline{x}_s^l = \mathcal{D}^l(\mathcal{X}_s).$$

This yields the desired estimate.   □

Remark 2.1. *Below, we comment two remarks for Lemma 2.1.*

1. The result of Lemma 2.1 implies that the convex hull of the set $\{X_t^i\}$ is non-increasing along the flow (2.1). More precisely, set

$$C_t := convex\ hull\{X_t^1, \cdots, X_t^N\}.$$

Then, one has

$$C_t \subset C_s, \quad for\ t \geq s.$$

2. Using similar arguments, one can also show that the mixed norm $\|\mathcal{X}_t\|_{2,\infty}$:

$$\|\|\mathcal{X}_t\|\|_{2,\infty} := \max_{1 \leq i \leq N} \|X_t^i\|_{\ell^2}$$

is non-increasing in t:

$$\|\mathcal{X}_t\|_{2,\infty} \leq \|\mathcal{X}_s\|_{2,\infty}, \quad for\ t \geq s.$$

2.2. Deterministic discrete algorithm. Let $\mathcal{X}_n := (X_n^1, \cdots, X_n^N) \in \mathbb{R}^{Nd}$ be a state vector to system (2.2), and for $n \geq 0$ and $l \in \{1, \cdots, d\}$, set

$$\underline{x}_n^l := \min_{1 \leq i \leq N} x_n^{i,l} \quad and \quad \bar{x}_n^l := \max_{1 \leq i \leq N} x_n^{i,l}.$$

Then, similar to Lemma 2.1, one has the discrete counterpart for Lemma 2.1.

Lemma 2.2. *Let $\mathcal{X}_n = (X_n^1, \cdots, X_n^N)$ be a state to (2.2) with the initial data $\mathcal{X}_0$. Then, the following assertions hold.*

(1) For each $k \in \{1, \cdots, N\}$, $\underline{x}_n^l$ and $\bar{x}_n^l$ are monotonically increasing and decreasing, respectively:

$$\underline{x}_n^l \geq \underline{x}_m^l \quad and \quad \bar{x}_n^l \leq \bar{x}_m^l, \quad for\ n \geq m.$$

(2) For $l \in \{1, \cdots, N\}$, the component diameter functional $\mathcal{D}^l(\mathcal{X}_n)$ is non-increasing in n:

$$\mathcal{D}^l(\mathcal{X}_n) \leq \mathcal{D}^l(\mathcal{X}_0), \quad n \geq 0.$$

Proof. Basically, we use the same arguments as in Lemma 2.1.

(i) Let i and j be two indices such that

$$\underline{x}_n^l := x_n^{i,l} \quad and \quad \bar{x}_n^l := x_n^{j,l}.$$

Then for such i, by $\psi_n^k(x_n^{k,l} - x_n^{i,l}) \geq 0$, one has

$$x_{n+1}^{i,l} = x_n^{i,l} + \lambda h \sum_{k=1}^N \psi_n^k(x_n^{k,l} - x_n^{i,l}) \geq x_n^{i,l} = \underline{x}_n^l.$$

This implies

(2.5) $$\underline{x}_{n+1}^l \geq \underline{x}_n^l.$$

Hence $\underline{x}_n^l$ is non-decreasing in n. Similarly, one has

$$x_{n+1}^{j,l} = x_n^{j,l} + \lambda h \sum_{k=1}^N \psi_n^k(x_n^{k,l} - x_n^{j,l}) \leq x_n^{j,l} = \bar{x}_n^l.$$

This yields

(2.6) $$\bar{x}_{n+1}^l \leq \bar{x}_n^l.$$

(ii) We combine (2.5) and (2.6) to get

$$\mathcal{D}^l(\mathcal{X}_{n+1}) = \bar{x}_{n+1}^l - \underline{x}_{n+1}^l \leq \bar{x}_n^l - \underline{x}_n^l = \mathcal{D}^l(\mathcal{X}_n).$$

$\square$

Remark 2.2. *The result of Lemma 2.2 yields*

$$\min_{1\leq k\leq N} x_0^{k,l} \leq x_n^{i,l} \leq \max_{1\leq k\leq N} x_0^{k,l}, \quad i = 1, \cdots, N, \ n \geq 0.$$

3. Emergence of global consensus

In this section, we study the emergence of global consensus to systems (2.1) and (2.2) based on the following two steps:

- Step A: We first derive an explicit formula for the state difference $X_t^i - X_t^j$, and then by using this formula, we show that the relative state differences tend to zero exponentially fast.
- Step B: For each component, we show that the maximal and minimal values are monotonically decreasing and increasing respectively over time so that as time tends to infinity, all extremal states tend to the same value. Then, together with the result of Step A, we can see that all states converge to the same global consensus state X_∞ independent of particle number i.

3.1. **Stochastic continuous algorithm.** Consider the continuous algorithms for X_t^i and X_t^j:

(3.1) $$\begin{cases} dX_t^i = \lambda \sum_{k=1}^N \psi_t^k (X_t^k - X_t^i)dt + \sigma \sum_{k=1}^N \sum_{l=1}^d \psi_t^k (x_t^{k,l} - x_t^{i,l})dW_t^l e_l, \quad t > 0, \\ dX_t^j = \lambda \sum_{k=1}^N \psi_t^k (X_t^k - X_t^j)dt + \sigma \sum_{k=1}^N \sum_{l=1}^d \psi_t^k (x_t^{k,l} - x_t^{j,l})dW_t^l e_l, \quad t > 0, \end{cases}$$

subject to the initial data:

(3.2) $$X_t^i\big|_{t=0} = X_0^i, \quad X_t^j\big|_{t=0} = X_0^j.$$

First, we use the unit sum condition (1.5) of ψ_t^k's to get

(3.3) $$\sum_{k=1}^N \psi_t^k (X_t^k - X_t^i) - \sum_{k=1}^N \psi_t^k (X_t^k - X_t^j) = -\Big(\sum_{k=1}^N \psi_t^k\Big)(X_t^i - X_t^j) = -(X_t^i - X_t^j).$$

Note that the dependence on ψ_t^i disappears on the R.H.S. of (3.3). Thus, one uses (3.3) to see that $X_t^i - X_t^j$ satisfies

(3.4) $$d(X_t^i - X_t^j) = -\lambda(X_t^i - X_t^j)dt - \sigma \sum_{l=1}^d (x_t^{i,l} - x_t^{j,l})dW_t^l e_l.$$

Next, we provide the emergence of global consensus to the deterministic model (3.4) with $\sigma = 0$. For a given configuration process $\mathcal{X}_t$, we set

$$\mathcal{D}(\mathcal{X}_t) := \max_{1\leq i,j\leq N} |X_t^i - X_t^j|.$$

Theorem 3.1. *Let $\{X^i\}$ be a solution to (3.1) - (3.2) with $\sigma = 0$. Then, the following two assertions hold.*

(1) *The diameter $\mathcal{D}(\mathcal{X})$ decays to zero exponentially fast:*

$$\mathcal{D}(\mathcal{X}_t) \le e^{-\lambda t}\mathcal{D}(\mathcal{X}_0), \quad t \ge 0.$$

(2) *There exits a unique global consensus state $X_\infty = (x_\infty^1, \cdots, x_\infty^d) \in \mathbb{R}^d$ such that for all $i = 1, \cdots, N$,*

$$\lim_{t\to\infty} X_t^i = X_\infty.$$

Proof. (i) It follows from (3.4) that

$$\frac{d}{dt}(X_t^i - X_t^j) = -\lambda(X_t^i - X_t^j), \quad t > 0.$$

This yields

$$|X_t^i - X_t^j| = e^{-\lambda t}|X_0^i - X_0^j|, \quad t \ge 0.$$

Again, by taking maximum over all the indices i and j, one gets the desired exponential decay of $\mathcal{D}(\mathcal{X}_t)$.

(ii) For each $l \in \{1, \cdots, d\}$, we claim that the extreme functions $\bar{x}_t^l$ and $\underline{x}_t^l$ converge to the same value x_∞^l so that all the other state $x^{i,l}$ should converge to the same value x_∞^l, because the component diameter shrinks to zero asymptotically. In the course of proof of (i), we showed that $\bar{x}_t^l$ is non-increasing and bounded by $\bar{x}_0^l$. Thus, it should converge to $\bar{x}_\infty^l$. Similarly, $\underline{x}^l$ should converge to $\underline{x}_\infty^l$. Then, it is easy to see that $\bar{x}_\infty^l = \underline{x}_\infty^l$ due to the exponential decay of the component diameter. $\square$

Remark 3.1. *Note that the explicit form of ψ_t^i does not appear in the decay estimate of the diameter.*

On the other hand, it follows from (3.4) that $x_t^{ij,l} := x_t^{i,l} - x_t^{j,l}$ satisfies

$$(3.5) \qquad \begin{cases} dx_t^{ij,l} = -\lambda x_t^{ij,l}dt - \sigma x_t^{ij,l}dW_t^l, \quad t > 0, \\ x_t^{ij,l}\big|_{t=0} = x_0^i - x_0^j. \end{cases}$$

Now, we apply Ito's formula for $\ln x_t^{ij,l}$ using (3.5) to see

$$(3.6) \qquad d\ln x_t^{ij,l} = \frac{dx_t^{ij,l}}{x_t^{ij,l}} - \frac{1}{2(x_t^{ij,l})^2}dx_t^{ij,l} \cdot dx_t^{ij,l} = -\left(\lambda + \frac{\sigma^2}{2}\right)dt + \sigma dW_t^l.$$

Integrating the above relation (3.6) gives

$$(3.7) \qquad x_t^{ij,l} = x_0^{ij,l}\exp\left[-\left(\lambda + \frac{\sigma^2}{2}\right)t + \sigma W_t^l\right], \quad t \ge 0.$$

Then, the explicit formula (3.7) yields the following result.

Theorem 3.2. *Let $\{X_t^i\}$ be a solution process of (2.1). Then, for $i \ne j = 1, \cdots N$ and $k \in \{1, \cdots, d\}$,*

$$\lim_{t\to\infty} |x_t^{i,k} - x_t^{j,k}| = 0, \quad a.s. \quad and \quad \lim_{t\to\infty} \mathbb{P}\left(|x_t^{i,k} - x_t^{j,k}|^2 > \varepsilon\right) = 0, \quad for \ any \ \varepsilon > 0.$$

Proof. The proof is essentially the same as in Theorem 3.1 of [2]. However, for reader's convenience, we briefly sketch the proof here.

(i) Recall the law of iterated logrithm of the Brownian motion:

$$\limsup_{t \to \infty} \frac{|W_t^k|}{\sqrt{2t \log \log t}} = 1, \quad \text{a.s.},$$

and note that for $t \gg 1$, the linear negative term $-\left(\lambda + \frac{\sigma^2}{2}\right)t$ in t is certainly dominant compared to the Brownian term σW_t^k which grows with a rate of at most $t^{\frac{1}{2}+}$. Thus, the trajectory $|x_t^{i,k} - x_t^{j,k}|$ tends to zero almost surely as $t \to \infty$.

(ii) Since a.s. convergence implies convergence in probability, the result follows from (i). □

3.2. **Stochastic discrete algorithm.** Consider the discrete algorithms: for $i, j = 1, \cdots, N$,

$$(3.8) \quad \begin{cases} X_{n+1}^i = X_n^i + \lambda h \sum_{k=1}^{N} \psi_n^k(X_n^k - X_n^i) + \sigma\sqrt{h} \sum_{k=1}^{N} \sum_{l=1}^{d} \psi_n^k(x_n^{k,l} - x_n^{i,l})Z_n^l e_l, \\ X_{n+1}^j = X_n^j + \lambda h \sum_{k=1}^{N} \psi_n^k(X_n^k - X_n^j) + \sigma\sqrt{h} \sum_{k=1}^{N} \sum_{l=1}^{d} \psi_n^k(x_n^{k,l} - X_n^{j,l})Z_n^l e_l, \end{cases}$$

subject to the initial data:

$$(3.9) \quad X_n^i\big|_{n=0} = X_0^i, \quad X_n^j\big|_{n=0} = X_0^j.$$

Here the random variables $\{Z_n^l\}$ are i.i.d. and satisfy $Z_n^l \sim \mathcal{N}(0, 1)$.

Note that

$$\sum_{k=1}^{N} \psi_n^k(X_n^k - X_n^i) - \sum_{k=1}^{N} \psi_n^k(X_n^k - X_n^j) = -(X_n^i - X_n^j).$$

Thus, one has

$$(3.10) \quad x_{n+1}^{i,k} - x_{n+1}^{j,k} = \left(1 - \lambda h - \sigma\sqrt{h}Z_n^k\right)(x_n^{i,k} - x_n^{j,k}).$$

Based on the above explicit recursive relation, we have the following emergent dynamics.

Theorem 3.3. *Suppose that parameters satisfy*

$$\sigma = 0, \quad \lambda > 0 \quad and \quad 0 < h < \frac{1}{\lambda}.$$

Then, for any solution $\{X_n^i\}$ to (3.8) - (3.9), the following two assertions hold.
   (1) The diameter $\mathcal{D}(\mathcal{X}_n)$ decays to zero exponentially fast:

$$\mathcal{D}(\mathcal{X}_n) \le e^{-\lambda n h}\mathcal{D}(\mathcal{X}_0), \quad n = 0, 1, \cdots.$$

   (2) There exits a unique global consensus state $X_\infty = (x_\infty^1, \cdots, x_\infty^d) \in \mathbb{R}^d$ independent of i such that for all $i = 1, \cdots, N$,

$$\lim_{n \to \infty} X_n^i = X_\infty.$$

Proof. (i) It follows from (3.10) that

$$X_{n+1}^i - X_{n+1}^j = X_n^i - X_n^j - \lambda h \Big(\sum_{l=1}^N \psi_n^l \Big)(X_n^i - X_n^j) = \Big(1 - \lambda h\Big)(X_n^i - X_n^j).$$

This implies

$$X_n^i - X_n^j = \Big(1 - \lambda h\Big)^n (X_0^i - X_0^j).$$

Thus, we have the desired estimate:

$$|X_n^i - X_n^j| = \Big(1 - \lambda h\Big)^n |X_0^i - X_0^j| \le e^{-\lambda h n} |X_0^i - X_0^j|, \quad n = 0, 1, \cdots.$$

(ii) We use the same argument in Theorem 3.1 to get the desired convergence. $\qquad\square$

Now, we return to the stochastic version with $\sigma > 0$ in the following theorem.

Theorem 3.4. *Let $\{X_n^i\}$ be a solution to (3.8) - (3.9). Then, we have the following stochastic consensus:*

(1) *(Weak stochastic consensus I): Suppose that h and systems parameters satisfy*

$$\sigma > 0, \quad \lambda > 0 \quad and \quad 0 < h < \frac{1}{\lambda}.$$

Then, for $i, j = 1, \cdots, N$, one has

$$\Big|\mathbb{E}[X_n^i - X_n^j]\Big| \le e^{-\lambda n h}\Big|\mathbb{E}[X_0^i - X_0^j]\Big|.$$

(2) *(Weak stochastic consensus II): Suppose that h and systems parameters satisfy*

$$\sigma > 0, \quad 2\lambda > \sigma^2 \quad and \quad 0 < h < \frac{2\lambda - \sigma^2}{\lambda}.$$

Then, for $i, j = 1, \cdots, N$, one has the following exponential decay estimate:

$$\mathbb{E}|X_n^i - X_n^j|^2 \le e^{-n h m}\mathbb{E}|X_0^i - X_0^j|^2,$$

where $m = m(\lambda, h, \sigma)$ is defined as follows:

$$m(\lambda, h, \sigma) := 2\lambda - \lambda^2 h - \sigma^2.$$

(3) *(Strong stochastic consensus): Suppose that h and system parameters satisfy*

$$\sigma > 0, \quad 2\lambda > \sigma^2 \quad and \quad 0 < h < \frac{2\lambda - \sigma^2}{\lambda}.$$

Then, for $i, j = 1, \cdots, N$, one has

$$|X_n^i - X_n^j| \le e^{-n Y_n}|X_0^i - X_0^j|, \quad a.s. \ \omega \in \Omega,$$

where Y_n is a random variable satisfying

$$\lim_{n \to \infty} Y_n(\omega) = \frac{hm}{2} := \frac{h}{2}(2\lambda - \lambda^2 h - \sigma^2) > 0, \quad a.e. \ \omega \in \Omega.$$

Proof. (i) It follows from (3.10) that

(3.11)
$$\begin{aligned} &X_{n+1}^i - X_{n+1}^j \\ &= X_n^i - X_n^j - \lambda h(X_n^i - X_n^j) + \sigma\sqrt{h}(X_n^i - X_n^j)Z_n = \Big(1 - \lambda h + \sigma\sqrt{h}Z_n\Big)(X_n^i - X_n^j). \end{aligned}$$

Iterating the above recursive relation (3.11) gives

$$(3.12) \qquad X_n^i - X_n^j = \Pi_{\ell=0}^{n-1}\Big(1 - \underbrace{(\lambda h - \sigma\sqrt{h}Z_\ell)}_{=:\Delta_\ell}\Big)(X_0^i - X_0^j).$$

Finally, we take expectation and absolute value on both sides of (3.12) using the independence of Z_ℓ and $X_0^i - X_0^j$ to get

$$\Big|\mathbb{E}[X_n^i - X_n^j]\Big| = (1 - \lambda h)^n \Big|\mathbb{E}[X_0^i - X_0^j]\Big| \le e^{-\lambda n h}\Big|\mathbb{E}[X_0^i - X_0^j]\Big|.$$

(ii) We take the absolute value of (3.12) and square of it to find

$$(3.13) \qquad |X_n^i - X_n^j|^2 = \Pi_{\ell=0}^{n-1}(1 - \Delta_\ell)^2 |X_0^i - X_0^j|^2.$$

Taking expectation of the above relation and using the independence of Δ_ℓ and $|X_0^i - X_0^j|$, one gets

$$(3.14) \qquad \mathbb{E}|X_n^i - X_n^j|^2 = \Pi_{\ell=0}^{n-1}\mathbb{E}[(1 - \Delta_\ell)^2] \times \mathbb{E}[|X_0^i - X_0^j|^2].$$

On the other hand, since $\{\Delta_\ell\}$ are i.i.d. and for each $\ell = 0, \cdots, n-1$, one has

$$
\begin{aligned}
\mathbb{E}[(1 - \Delta_\ell)^2] &= 1 - \mathbb{E}[2\Delta_\ell - \Delta_\ell^2] \\
(3.15) \qquad &= 1 - \mathbb{E}\Big(2\lambda h - 2\sigma\sqrt{h}Z_\ell - (\lambda h)^2 + 2\sigma\lambda h^{\frac{3}{2}}Z_\ell - \sigma^2 h Z_\ell^2\Big) \\
&= 1 - 2\lambda h + \lambda^2 h^2 + \sigma^2 h = 1 - h m(\lambda, h, \sigma) \ge 0.
\end{aligned}
$$

Now, we combine (3.14) and (3.15) to get

$$\mathbb{E}|X_n^i - X_n^j|^2 = (1 - hm)^n \mathbb{E}|X_0^i - X_0^j|^2 \le e^{-mnh}\mathbb{E}|X_0^i - X_0^j|^2.$$

So a sufficient condition for the exponential decay of $\mathbb{E}|X_n^i - X_n^j|^2$ is

$$m(\lambda, h, \sigma) > 0 \quad \Longleftrightarrow \quad h < \frac{2\lambda - \sigma^2}{\lambda}.$$

(iii) It follows from (3.16) and the inequality:

$$(1 - \Delta_\ell)^2 = 1 - (2\Delta_\ell - \Delta_\ell^2) \le e^{-(2\Delta_\ell - \Delta_\ell^2)}$$

that

$$
\begin{aligned}
|X_n^i - X_n^j|^2 &= \Pi_{\ell=0}^{n-1}(1 - \Delta(\ell))^2 |X_0^i - X_0^j|^2 \le \Pi_{\ell=0}^{n-1} e^{-(2\Delta_\ell - \Delta_\ell^2)}|X_0^i - X_0^j|^2 \\
&= \exp\Big[-\sum_{\ell=0}^{n-1}\Delta_\ell(2 - \Delta_\ell)\Big]|X_0^i - X_0^j|^2 \\
(3.16) \qquad &= \exp\Big[-n \times \frac{1}{n}\sum_{\ell=0}^{n-1}\Delta_\ell(2 - \Delta_\ell)\Big]|X_0^i - X_0^j|^2.
\end{aligned}
$$

On the other hand, using (3.15) one has

$$\mathbb{E}[2\Delta_\ell - \Delta_\ell^2] = hm.$$

Next, we use the strong law of large numbers to see

$$Y_n := \frac{1}{n}\sum_{\ell=0}^{n-1}\Delta_\ell(2 - \Delta_\ell) \ \to\ \mathbb{E}\Big[\Delta_\ell(2 - \Delta_\ell)\Big] = hm \quad \text{as } n \to \infty \text{ a.s.}$$

where Y_n is also a random variable since Δ_ℓ is a random variable. In (3.16), we use the above convergence to see that

$$|X_n^i - X_n^j| \leq \exp\left[-n \times \frac{1}{2n} \sum_{\ell=0}^{n-1} \Delta_\ell(2 - \Delta_\ell)\right]|X_0^i - X_0^j|$$

and

$$\frac{1}{2n} \sum_{\ell=0}^{n-1} \Delta_\ell(2 - \Delta_\ell) \to \frac{hm}{2}, \quad n \longrightarrow \infty.$$

$\square$

Remark 3.2. *If one uses the result (ii) and the Cauchy-Schwarz inequality, one can obtain*

$$\mathbb{E}|X_n^i - X_n^j| \leq \sqrt{\mathbb{E}[|X_n^i - X_n^j|^2]} \leq (1 - hm)^{\frac{n}{2}}\sqrt{\mathbb{E}[|X_0^i - X_0^j|^2]}.$$

4. Convergence analysis for continuous algorithm

In this section, we provide a convergence analysis for the continuous CBO algorithm using Ito's calculus. In previous section, we showed that the continuous algorithm admits a global consensus for any initial data. Thus, the natural question is whether this global consensus is a global minimum of L or not. If the answer is affirmative, then under what condition such a coincidence will occur? This is the main concern of this section.

Recall that X_t^i satisfies

$$(4.1) \qquad dX_t^i = -\lambda(X_t^i - \bar{X}_t^*)dt + \sigma \sum_{l=1}^d (x_t^{i,l} - \bar{x}_t^{*,l})dW_t^l e_l,$$

and we introduce an ensemble average:

$$\bar{X}_t := \frac{1}{N} \sum_{i=1}^N X_t^i = (\bar{x}_t^1, \cdots, \bar{x}_t^d).$$

Next, we present three elementary lemmas to be crucially used in the proof of convergence analysis.

Lemma 4.1. *Let $\{X_t^i\}_{1 \leq i \leq N}$ be a solution to (4.1). Then, the following estimates hold almost surely.*

$$(i)\ |X_t^i - \bar{X}_t|^2 = \sum_{l=1}^d (x_0^{i,l} - \bar{x}_0^l)^2 \exp\left[-\left(2\lambda + \sigma^2\right)t + 2\sigma W_t^l\right].$$

$$(ii)\ |\bar{X}_t - \bar{X}_t^*|^2 \leq \max_{1 \leq i \leq N} |X_t^i - \bar{X}_t|^2.$$

$$(iii)\ \frac{1}{N} \sum_{i=1}^N |X_t^i - \bar{X}_t^*|^2 \leq 2 \sum_{l=1}^d \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2\right) \exp\left[-\left(2\lambda + \sigma^2\right)t + 2\sigma W_t^l\right].$$

Proof. (i) It follows from (4.1) that

$$(4.2) \qquad d\bar{X}_t = -\lambda(\bar{X}_t - \bar{X}_t^*)dt + \sigma \sum_{l=1}^d (\bar{x}_t^l - \bar{x}_t^{*,l})dW_t^l e_l.$$

We subtract (4.1) from (4.2) to obtain

$$(4.3) \qquad d(X_t^i - \bar{X}_t) = -\lambda(X_t^i - \bar{X}_t)dt + \sigma \sum_{l=1}^{d}(x_t^{i,l} - \bar{x}_t^l)dW_t^l e_l.$$

The l-th component of (4.3) implies

$$x_t^{i,l} - \bar{x}_t^l = (x_0^{i,l} - \bar{x}_0^l)\exp\left[-\left(\lambda + \frac{1}{2}\sigma^2\right)t + \sigma W_t^l\right].$$

This yields

$$|X_t^i - \bar{X}_t|^2 = \sum_{l=1}^{d}(x_0^{i,l} - \bar{x}_0^l)^2 \exp\left[-\left(2\lambda + \sigma^2\right)t + 2\sigma W_t^l\right].$$

(ii) We use the triangle inequality and the Cauchy-Schwarz inequality to get

$$|\bar{X}_t - \bar{X}_t^*|^2 = \left|\frac{\sum_{k=1}^{N} e^{-\beta L(X_t^k)}(\bar{X}_t - X_t^k)}{\sum_{k=1}^{N} e^{-\beta L(X_t^k)}}\right|^2 \leq \left[\frac{\sum_{k=1}^{N} e^{-\beta L(X_t^k)}|\bar{X}_t - X_t^k|}{\sum_{k=1}^{N} e^{-\beta L(X_t^k)}}\right]^2$$

$$\leq \frac{\sum_{k=1}^{N} e^{-\beta L(X_t^k)}|\bar{X}_t - X_t^k|^2}{\sum_{k=1}^{N} e^{-\beta L(X_t^k)}} \leq \max_{1 \leq k \leq N} |\bar{X}_t - X_t^k|^2.$$

(iii) Note that

$$\frac{1}{N}\sum_{i=1}^{N} |X_t^i - \bar{X}_t^*|^2 = \frac{1}{N}\sum_{i=1}^{N}\left(|X_t^i - \bar{X}_t|^2 + 2(X_t^i - \bar{X}_t)\cdot(\bar{X}_t - \bar{X}_t^*) + |\bar{X}_t - \bar{X}_t^*|^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}|X_t^i - \bar{X}_t|^2 + |\bar{X}_t - \bar{X}_t^*|^2 \leq 2\max_{1 \leq i \leq N}|X_t^i - \bar{X}_t|^2$$

$$(4.4) \qquad = 2\max_{1 \leq i \leq N}\left(\sum_{l=1}^{d}(x_0^{i,l} - \bar{x}_0^l)^2 \exp\left[-\left(2\lambda + \sigma^2\right)t + 2\sigma W_t^l\right]\right)$$

$$\leq 2\sum_{l=1}^{d}\left(\max_{1 \leq i \leq N}(x_0^{i,l} - \bar{x}_0^l)^2\right)\exp\left[-\left(2\lambda + \sigma^2\right)t + 2\sigma W_t^l\right],$$

where we used the inequalities from (i) and (ii). □

Lemma 4.2. *Let $\{X_t^i\}_{1 \leq i \leq N}$ be a solution to (4.1). Then, the following estimates hold.*

$$(i) \quad \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|X_t^i - \bar{X}_t^*|^2 \leq 2e^{-(2\lambda - \sigma^2)t}\sum_{l=1}^{d}\mathbb{E}\left[\max_{1 \leq i \leq N}(x_0^{i,l} - \bar{x}_0^l)^2\right].$$

(ii) If $2\lambda > \sigma^2$, then there exists a random vector X_∞ such that
$$\lim_{t \to \infty} X_t^i = X_\infty \ a.s., \ 1 \leq i \leq N.$$

Proof. (i) We take expectation on both sides of (4.4) to get

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}|X_t^i - \bar{X}_t^*|^2 \leq 2 \sum_{l=1}^{d} \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \mathbb{E} \exp\left[-\left(2\lambda + \sigma^2 \right) t + 2\sigma W_t^l \right]$$

$$\leq 2e^{-(2\lambda - \sigma^2)t} \sum_{l=1}^{d} \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right),$$

where we used

$$\exp\left(2\sigma W_t^l \right) \sim \text{Lognormal}(0, 4\sigma^2 t) \quad \Longrightarrow \quad \mathbb{E}\exp\left(2\sigma W_t^l \right) = \exp(2\sigma^2 t).$$

(ii) Note that equation (4.1) is equivalent to the following integral relation: for $i = 1, \cdots, N$ and $l = 1, \cdots, d$,

$$x_t^{i,l} = x_0^{i,l} - \lambda \int_0^t (x_s^{i,l} - \bar{x}_s^{*,l}) ds + \sigma \int_0^t (x_s^{i,l} - \bar{x}_s^{*,l}) dW_s^l =: x_0^{i,l} - \lambda \mathcal{I}_{11} + \sigma \mathcal{I}_{12}.$$

Next, we show the a.s. convergence $\mathcal{I}_{11}$ and $\mathcal{I}_{12}$ separately.

• Case A (Almost sure convergence of $\mathcal{I}_{11}$): By (iii), we have

$$|x_t^{i,l} - \bar{x}_t^{*,l}| \leq \sqrt{\sum_{i=1}^{N} |X_t^i - \bar{X}_t^*|^2} \leq \sqrt{2N \sum_{l=1}^{d} \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \exp\left[-\left(2\lambda + \sigma^2 \right) t + 2\sigma W_t^l \right]}.$$

This yields that there exist positive random functions $C_i = C_i(\omega)$, $i = 1, 2$ such that

$$|x_t^{i,l} - \bar{x}_t^{*,l}| \leq C_1 e^{-C_2 t}, \quad \text{a.s. } \omega \in \Omega,$$

where C_1 and C_2 are positive constants. We set

$$\mathcal{J}_{11} := \mathcal{I}_{11} - \int_0^t C_1 e^{-C_2 s} ds = \int_0^t \underbrace{\left(x_s^{i,l} - \bar{x}_s^{*,l} - C_1 e^{-C_2 s} \right)}_{\leq 0} ds.$$

Since the integrand is nonpositive a.s., $\mathcal{J}_{11}$ is non-increasing in t a.s.

On the other hand, note that

$$\mathcal{J}_{11} = \mathcal{I}_{11} - \frac{C_1}{C_2}(1 - e^{-C_2 t}) \geq \mathcal{I}_{11} - \frac{2C_1}{C_2} + \frac{C_1}{C_2}(1 - e^{-C_2 t})$$

$$= \mathcal{I}_{11} - \frac{2C_1}{C_2} + \int_0^t C_1 e^{-C_2 s} ds = -\frac{2C_1}{C_2} + \int_0^t \left(x_s^{i,l} - \bar{x}_s^{*,l} + C_1 e^{-C_2 s} \right) ds \geq -\frac{2C_1}{C_2}.$$

Since $\mathcal{J}_{11}$ is monotone decreasing and bounded below along sample paths, one has

$$\exists\, \alpha = \lim_{t \to \infty} \mathcal{J}_{11}(t) = \lim_{t \to \infty} \left(\mathcal{I}_{11} - \int_0^t C_1 e^{-C_2 s} ds \right), \quad \text{a.s.}$$

This implies

$$\lim_{t \to \infty} \mathcal{I}_{11} = \alpha + \frac{C_1}{C_2}, \quad \text{a.s.}$$

• Case B (Almost sure convergence of $\mathcal{I}_{12}$): Note that $\mathcal{I}_{12}$ is martingale and its $L^2(\Omega)$-norm is uniformly bounded in t:

$$
\mathbb{E}\left[\int_0^t \left(x_s^{i,l} - \bar{x}_s^{*,l}\right)dW_s^l\right]^2 = \mathbb{E}\int_0^t (x_s^{i,l} - \bar{x}_s^{*,l})^2 ds \leq \int_0^t \sum_{i=1}^N \mathbb{E}|X_s^i - \bar{X}_s^*|^2 ds
$$

$$
\leq 2N\left(\int_0^t e^{-(2\lambda-\sigma^2)s}ds\right)\sum_{l=1}^d \left(\mathbb{E}\max_{1\leq i\leq N}(x_0^{i,l} - \bar{x}_0^l)^2\right)
$$

$$
\leq \frac{2N}{2\lambda-\sigma^2}\sum_{l=1}^d \left(\mathbb{E}\max_{1\leq i\leq N}(x_0^{i,l} - \bar{x}_0^l)^2\right).
$$

In the second inequality we used (iv). Hence $\lim_{t\to\infty}\mathcal{I}_{12}$ exists a.s. Now we have shown that for each $i = 1,\cdots,N$, there exists some random variable X_∞^i such that

$$
\lim_{t\to\infty} X_t^i = X_\infty^i \quad \text{a.s.}
$$

Since for any $1 \leq i,j \leq N$,

$$
\lim_{t\to\infty}|X_t^i - X_t^j| = 0, \quad \text{a.s.}
$$

Hence, there exists X_∞ such that

$$
X_\infty^i = X_\infty^j =: X_\infty \quad \text{a.s.}
$$

$\square$

Lemma 4.3. *Let* $\{X_t^i\}_{1\leq i\leq N}$ *be a solution to (4.1). Then, the quadratic variation of* x_t^k *and* x_t^l *is given as follows.*

$$
dx_t^k \cdot dx_t^l = \begin{cases} \sigma^2|x_t^{i,k} - \bar{x}_t^{*,k}|^2 dt, & k = l, \\ 0, & k \neq l. \end{cases}
$$

Proof. It follows from (4.1) that the l-th component of X_t^i satisfies

(4.5) $$ dx_t^{i,l} = -\lambda(x_t^{i,l} - \bar{x}_t^{*,l})dt + \sigma(x_t^{i,l} - \bar{x}_t^{*,l})dW_t^l. $$

Now, we use the following quadratic variation relations:

$$
dt\cdot dt = 0, \quad dt\cdot dW_t^l = 0, \quad dW_t^l\cdot dt = 0, \quad dW_t^l\cdot dW_t^k = \delta_{lk}dt
$$

and (4.1) to see

$$
dx_t^k \cdot dx_t^l = \sigma^2\delta_{kl}(x_t^{i,k} - \bar{x}_t^{*,k})(x_t^{i,l} - \bar{x}_t^{*,l})dt.
$$

This certainly implies the desired estimate. $\square$

In what follows, we use a handy notation for partial derivatives:

$$
\partial_l := \frac{\partial}{\partial x^l}, \quad \partial_{kl}^2 := \frac{\partial^2}{\partial x^k \partial x^l}, \quad l,k = 1,\cdots,d.
$$

Let $L = L(x)$ be a C_b^2-objective function satisfying the following relations:

(4.6) $$ L_m := \inf_{x\in\mathbb{R}^d} L(x) > 0 \quad \text{and} \quad C_L := \max\left\{\sup_{x\in\mathbb{R}^d}\|\nabla^2 L(x)\|_2, \max_{1\leq l\leq d}\sup_{x\in\mathbb{R}^d}|\partial_l^2 L(x)|\right\} < \infty, $$

where $\|\cdot\|_2$ denotes the spectral norm. First, note that

$$
\begin{aligned}
(4.7) \qquad \partial_k\left(e^{-\beta L(X_t^i)}\right) &= -\beta e^{-\beta L(X_t^i)}\partial_k L(X_t^i), \\
\partial_{kl}^2\left(e^{-\beta L(X_t^i)}\right) &= \beta e^{-\beta L(X_t^i)}\left[\beta\partial_l L(X_t^i)\cdot\partial_k L(X_t^i) - \partial_{lk}^2 L(X_t^i)\right].
\end{aligned}
$$

Now, we apply Ito's formula to $\frac{1}{N}\sum_{i=1}^{N}e^{-\beta L(X_t^i)}$ using the relations (4.7) to get

$$
(4.8)
$$

$$
\begin{aligned}
&d\left(\frac{1}{N}\sum_{i=1}^{N}e^{-\beta L(X_t^i)}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{k=1}^{d}\partial_k\left(e^{-\beta L(X_t^i)}\right)dx_t^{i,k} + \frac{1}{2}\sum_{k,l=1}^{d}\partial_{kl}^2\left(e^{-\beta L(X_t^i)}\right)dx_t^{i,k}\cdot dx_t^{i,l}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[-\beta e^{-\beta L(X_t^i)}\sum_{k=1}^{d}\partial_k L(X_t^i)\left(-\lambda(x_t^{i,k}-\bar{x}_t^{*,k})dt + \sigma(x_t^{i,k}-\bar{x}_t^{*,k})dW_t^k\right)\right] \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\beta e^{-\beta L(X_t^i)}\left[\sum_{k,l=1}^{d}\left(\beta\partial_l L(X_t^i)\cdot\partial_k L(X_t^i) - \partial_{lk}^2 L(X_t^i)\right)\sigma^2\delta_{kl}(x_t^{i,k}-\bar{x}_t^{*,k})(x_t^{i,l}-\bar{x}_t^{*,l})dt\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\beta e^{-\beta L(X_t^i)}\nabla L(X_t^i)\cdot\left[\lambda(X_t^i-\bar{X}_t^*)dt - \sigma\sum_{k=1}^{d}(x_t^{i,k}-\bar{x}_t^{*,k})dW_t^k e_k\right] \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\left[\beta e^{-\beta L(X_t^i)}\sum_{k=1}^{d}\left(-\partial_{kk}L(X_t^i) + \beta(\partial_k L(X_t^i))^2\right)\frac{1}{2}\sigma^2(x_t^{i,k}-\bar{x}_t^{*,k})^2\right]dt.
\end{aligned}
$$

We take expectations on both sides of (4.8) to get

$$
\begin{aligned}
(4.9) \qquad &d\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}e^{-\beta L(X_t^i)}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\beta e^{-\beta L(X_t^i)}\nabla L(X_t^i)\cdot\lambda(X_t^i-\bar{X}_t^*)\right]dt \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\beta e^{-\beta L(X_t^i)}\sum_{k=1}^{d}\left(-\partial_{kk}L(X_t^i) + \beta(\partial_k L(X_t^i))^2\right)\frac{1}{2}\sigma^2(x_t^{i,k}-\bar{x}_t^{*,k})^2\right]dt \\
&=: \mathcal{I}_{21}dt + \mathcal{I}_{22}dt.
\end{aligned}
$$

Below, we estimate the terms $\mathcal{I}_{2i}$, $i=1,2$ as follows.

Lemma 4.4. *Let $\{X_t^i\}$ be a solution to (4.1). Then, the term $\mathcal{I}_{2i}$, $i=1,2$ satisfies*

$$
(i)\ \mathcal{I}_{21} \geq -\lambda C_L\beta e^{-\beta L_m}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|X_t^i-\bar{X}_t^*|^2;
$$

$$
(ii)\ \mathcal{I}_{22} \geq -\frac{1}{2}\sigma^2 C_L\beta e^{-\beta L_m}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|X_t^i-\bar{X}_t^*|^2.
$$

Proof. Below, we estimate $\mathcal{I}_{2i}$ separately.

- (Estimate of $\mathcal{I}_{21}$): First, we use definition of $\bar{X}_t^*$ to see

$$\left(\sum_{i=1}^N e^{-\beta L(X_t^i)}\right)\bar{X}_t^* = \sum_{i=1}^N e^{-\beta L(X_t^i)}X_t^i.$$

This yields

(4.10)
$$\sum_{i=1}^N e^{-\beta L(X_t^i)}\nabla L(\bar{X}_t^*)\cdot(\bar{X}_t^* - X_t^i) = 0.$$

Then, we use (4.6) and (4.10) to find

$$\mathcal{I}_{11} = \frac{\beta\lambda}{N}\sum_{i=1}^N \mathbb{E}\left[e^{-\beta L(X_t^i)}\nabla L(X_t^i)\cdot(X_t^i - \bar{X}_t^*)\right]dt$$

(4.11)
$$= \frac{\beta\lambda}{N}\sum_{i=1}^N \mathbb{E}\left[e^{-\beta L(X_t^i)}\big(\nabla L(X_t^i) - \nabla L(\bar{X}_t^*)\big)\cdot(X_t^i - \bar{X}_t^*)\right]$$

$$\geq -\lambda C_L\beta e^{-\beta L_m}\frac{1}{N}\sum_{i=1}^N \mathbb{E}|X_t^i - \bar{X}_t^*|^2.$$

- (Estimate of $\mathcal{I}_{22}$): By direct calculation, one has

$$\mathcal{I}_{12} = -\frac{\sigma^2\beta}{2N}\sum_{i=1}^N \mathbb{E}\left[e^{-\beta L(X_t^i)}\sum_{k=1}^d \partial_{kk}L(X_t^i)(x_t^{i,k} - \bar{x}_t^{*,k})^2\right]$$

(4.12)
$$\geq -\frac{1}{2}\sigma^2 C_L\beta e^{-\beta L_m}\frac{1}{N}\sum_{i=1}^N \mathbb{E}|X_t^i - \bar{X}_t^*|^2.$$

$\square$

Now, we are ready to provide the convergence result of the continuous CBO algorithm. Note that in [5], the $L^2(\Omega)$-limit of the stochastic process X_t was actually equal to some non-random $\tilde{x}\in\mathbb{R}^d$, but it is not the case for our N-particle model. This resulted in the statement of Theorem 4.1 slightly different from the analogous theorem (Theorem 3.1) in [5].

Theorem 4.1. *Suppose that λ, σ and $\{X_0^i\}$ satisfy*

$$2\lambda > \sigma^2, \quad X_0^i : i, i.d,, \quad X_0^i \sim X^{in} \text{ for some random variable } X^{in},$$

$$(1-\varepsilon)\mathbb{E}\left[e^{-\beta L(X^{in})}\right] \geq \frac{2\lambda+\sigma^2}{2\lambda-\sigma^2}C_L\beta e^{-\beta L_m}\sum_{l=1}^d \mathbb{E}\left[\max_{1\leq i\leq N}(x_0^{i,l} - \bar{x}_0^l)^2\right],$$

for some $0 < \varepsilon < 1$. Then, one has

$$\operatorname{ess\,inf}_{\omega\in\Omega} L(X^\infty(\omega)) \leq \operatorname{ess\,inf}_{\omega\in\Omega} L(X^{in}(\omega)) + \mathcal{O}\Big(\frac{1}{\beta}\Big), \quad \text{for } \beta \gg 1.$$

Consequently, if the global minimizer X^* of L is contained in $\operatorname{supp}\operatorname{law}(X^{in})$, then

$$\operatorname{ess\,inf}_{\omega\in\Omega} L(X^\infty(\omega)) \leq L_m + \mathcal{O}\Big(\frac{1}{\beta}\Big), \quad \text{for } \beta \gg 1.$$

Proof. In (4.9), we use (4.11), (4.12) and Lemma 4.1 (iii) to find

$$
(4.13)\quad
\begin{aligned}
\frac{d}{dt}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}e^{-\beta L(X_t^i)}\right) &\geq -\left(\lambda+\frac{1}{2}\sigma^2\right)C_L\beta e^{-\beta L_m}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|X_t^i-\bar{X}_t^*|^2 \\
&\geq -2\left(\lambda+\frac{1}{2}\sigma^2\right)C_L\beta e^{-\beta L_m}e^{-(2\lambda-\sigma^2)t}\sum_{l=1}^{d}\mathbb{E}\left[\max_{1\leq i\leq N}(x_0^{i,l}-\bar{x}_0^l)^2\right].
\end{aligned}
$$

Now integrating (4.13) in t gives

$$
(4.14)\quad
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}e^{-\beta L(X_t^i)} \\
&\geq \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}e^{-\beta L(X_0^i)}-2\left(\lambda+\frac{1}{2}\sigma^2\right)C_L\beta e^{-\beta L_m}\frac{1-e^{-(2\lambda-\sigma^2)t}}{2\lambda-\sigma^2}\sum_{l=1}^{d}\mathbb{E}\left[\max_{1\leq i\leq N}(x_0^{i,l}-\bar{x}_0^l)^2\right] \\
&\geq \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}e^{-\beta L(X_0^i)}-\frac{2\lambda+\sigma^2}{2\lambda-\sigma^2}C_L\beta e^{-\beta L_m}\sum_{l=1}^{d}\mathbb{E}\left[\max_{1\leq i\leq N}(x_0^{i,l}-\bar{x}_0^l)^2\right].
\end{aligned}
$$

Letting $t\to\infty$, and we use Lemma 4.2 (i) to find

$$
\mathbb{E}e^{-\beta L(X_\infty)}\geq \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}e^{-\beta L(X_0^i)}-\frac{2\lambda+\sigma^2}{2\lambda-\sigma^2}C_L\beta e^{-\beta L_m}\sum_{l=1}^{d}\mathbb{E}\left[\max_{1\leq i\leq N}(x_0^{i,l}-\bar{x}_0^l)^2\right]\geq \varepsilon\mathbb{E}e^{-\beta L(X^{in})},
$$

i.e.,

$$
-\frac{1}{\beta}\log\mathbb{E}e^{-\beta L(X^\infty)}\leq -\frac{1}{\beta}\log\mathbb{E}e^{-\beta L(X^{in})}-\frac{1}{\beta}\log\varepsilon.
$$

Now Laplace's principle implies

$$
\operatorname{ess\,inf}_{\omega\in\Omega}L(X^\infty(\omega))\leq \operatorname{ess\,inf}_{\omega\in\Omega}L(X^{in}(\omega))+O\!\left(\frac{1}{\beta}\right)\quad\text{for }\beta\gg 1.
$$

$\square$

5. NUMERICAL SIMULATIONS

In this section, we conduct several numerical tests to verify the results of the convergence analysis. For a numerical test, we use the Rastrigin function as in [5, 23] as the objective function:

$$
L(X)=\sum_{i=1}^{d}\left[(x^i-B)^2-10\cos(2\pi(x^i-B))+10\right]+C,
$$

where constants B and C are given by

$$
B:=\operatorname{argmin}L(X),\quad C:=\min L(X).
$$

Note that this function has a unique global minimizer, namely $X=(B,\cdots,B)\in\mathbb{R}^d$. However, it has many local minimizers as can be seen Figure 1 (see the graph for L is provided in Figure 1 with $d=2, B=C=0$).
For the initial data and system parameters, we choose $N=100$ points uniformly from the square $[-2,2]\times[-2,2]$, which includes a global minimum point, and use parameters:

$$
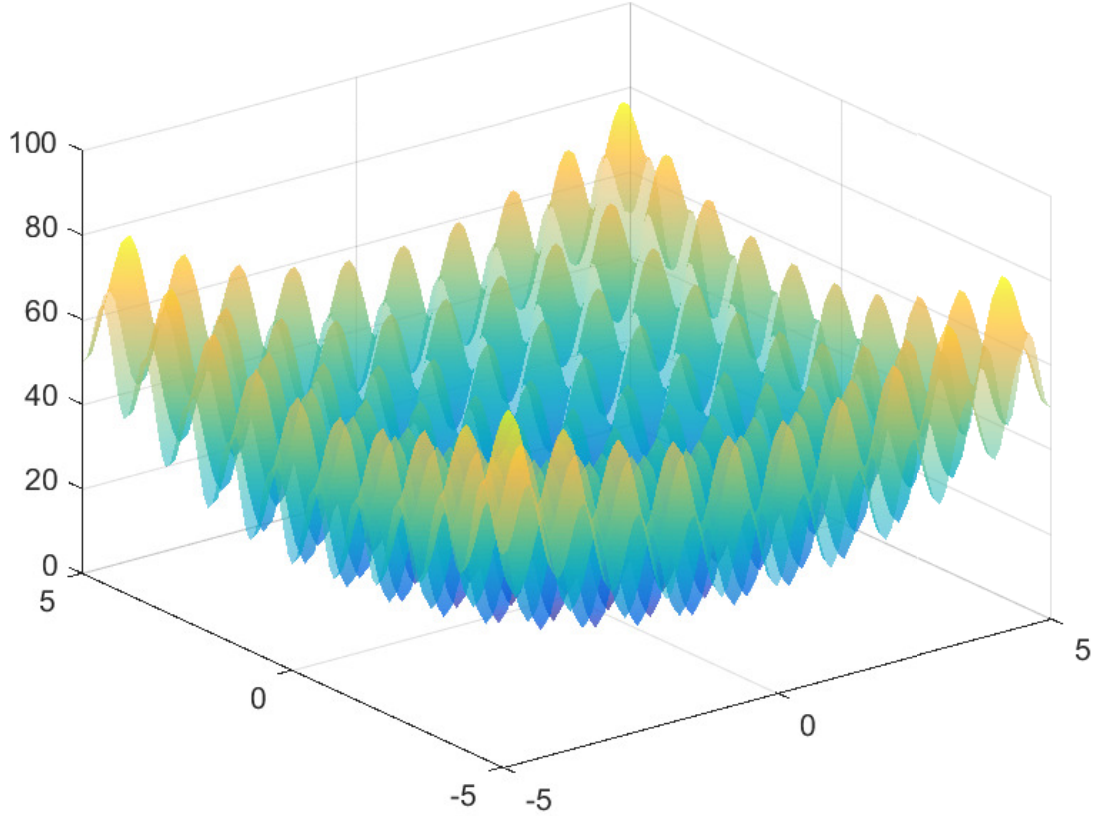\Delta t=h=0.01,\quad N=100,\quad \beta=10,\quad \lambda=1.
$$

FIGURE 1. A 3D plot of the Rastrigin function

For the same chosen initial data set as above, we perform simulations for $\sigma = 0, 1, 2$ and compare the results.

5.1. **Continuous algorithm.** For the simulations of the continuous algorithm, we use the following two-step numerical scheme:

$$\hat{X}_n^i = \bar{X}_n^* + (X_n^i - \bar{X}_n^*)e^{-\lambda h},$$

$$X_{n+1}^i = \hat{X}_n^i + \sigma\sqrt{h}\sum_{l=1}^{d}(\hat{x}_n^{i,l} - \bar{x}_n^{*,l})w_n^l e_l,$$

where w_n^l ($l = 1, \cdots, d, n = 0, 1, 2, \cdots$) are independent and follow the standard normal distribution, and Δt is the time step. In Section 3, we derived the following explicit formula for the continuous algorithm:

$$x_t^{i,l} - x_t^{j,l} = (x_0^{i,l} - x_0^{j,l})\exp\left[-\left(\lambda + \frac{1}{2}\sigma^2\right)t + \sigma W_t^l\right].$$

It is easy to see from the above formula that the particles will reach a global consensus almost surely if and only if the coupling strength and noise intensity satisfy

$$\lambda > 0 \quad \text{or} \quad |\sigma| > 0.$$

Note that for larger σ, the speed of consensus on average is faster. Note that these facts do not require the condition $2\lambda > \sigma^2$ as can be seen in convergence analysis in Section 4. One can observe this result numerically. In Figures 2,3 and 4, we plot the positions of the particles for $\sigma = 0, 1, 2$, respectively. Indeed, the particles seem to converge faster, as σ increases. In Figure 5, we plot a sample path of $\log |x_t^{1,1} - x_t^{2,1}|$ for $\sigma = 0, 1, 2$. As expected, the graph for $\sigma = 0$ is linear, and the function eventually decays faster for large σ.
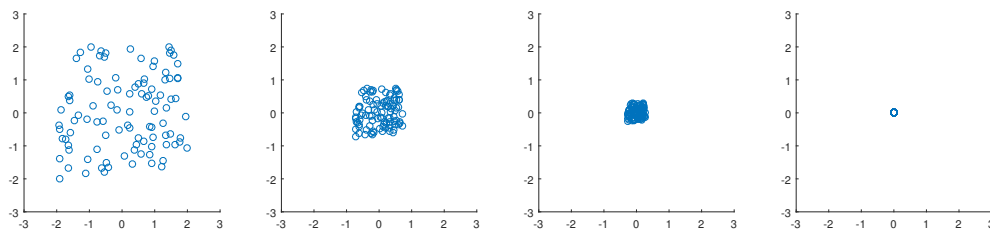

FIGURE 2. Temporal evolution of state configuration for $t = 0, 1, 2, 10$ ($\sigma = 0$).
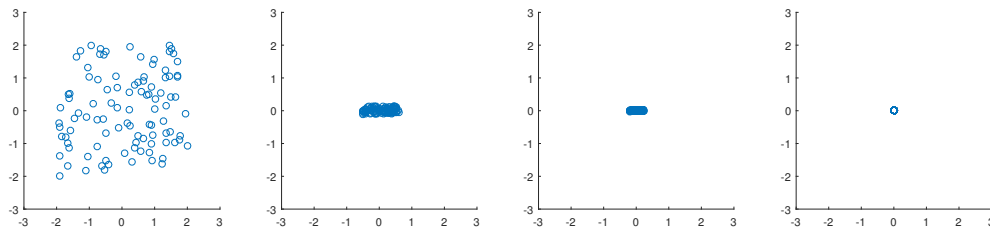

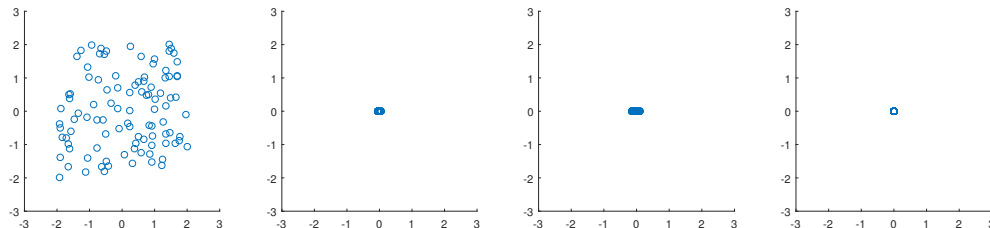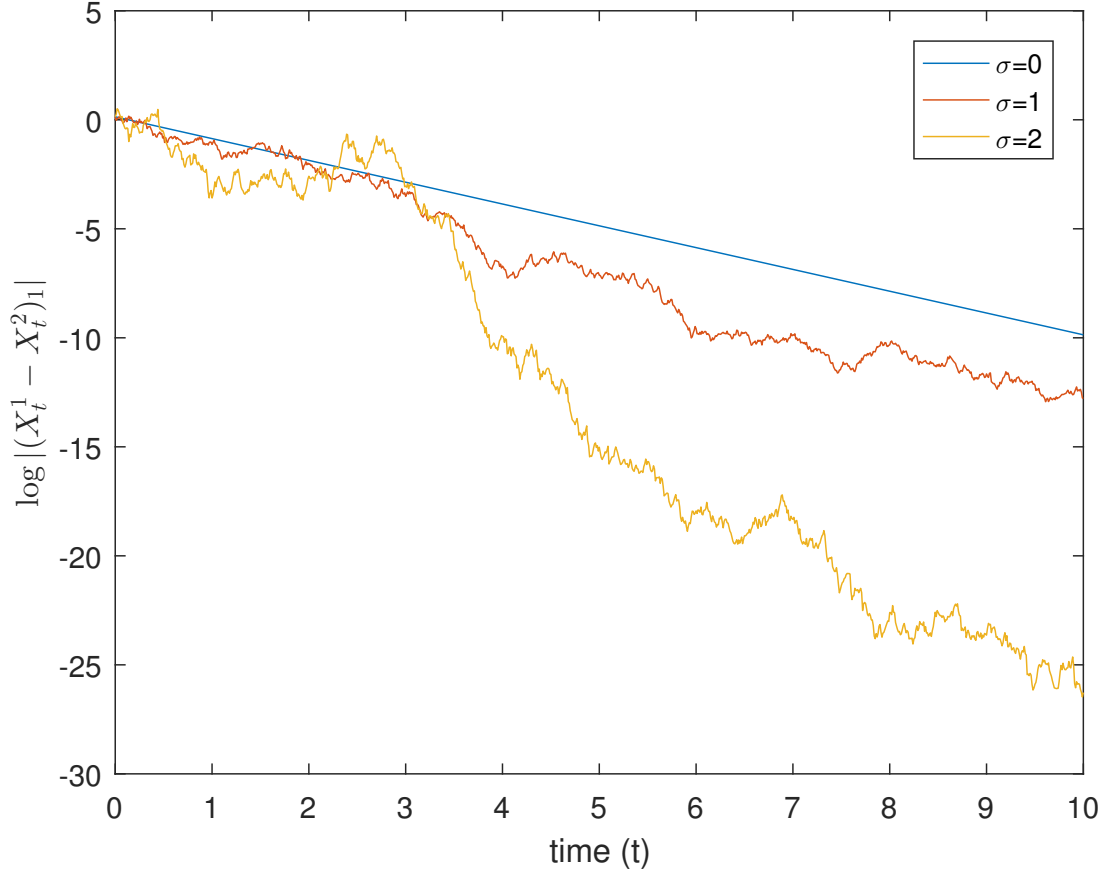FIGURE 3. Temporal evolution of state configuration for $t = 0, 1, 2, 10$ ($\sigma = 1$).

FIGURE 4. Temporal evolution of state configuration for $t = 0, 1, 2, 10$ ($\sigma = 2$).

FIGURE 5. Graph of $\log |x_t^{1,1} - x_t^{2,1}|$ for $\sigma = 0, 1, 2$.

5.2. **Discrete algorithm.** Recall the discrete algorithm:

$$(5.1) \quad \begin{cases} X_{n+1}^i = X_n^i + \lambda h \sum_{k=1}^N \psi_n^k (X_n^k - X_n^i) + \sigma \sqrt{h} \sum_{k=1}^N \sum_{l=1}^d \psi_n^k (x_n^{k,l} - x_n^{i,l}) Z_n^l e_l, \\ \psi_n^k := \dfrac{e^{-\beta L(X_n^k)}}{\sum_{i=1}^N e^{-\beta L(X_n^i)}}, \quad 1 \le k \le N, \quad n = 0, 1, \cdots . \end{cases}$$

In this subsection, we study the formation of global consensus for the discrete algorithm (5.1) numerically. In Theorem 3.4, we have shown that if $m := \lambda - \frac{\sigma^2}{2} - \frac{\lambda^2 h}{2} > 0$ then the quantity $\Delta_n^{ij} := |X_n^i - X_n^j|$ tends to zero almost surely, as $n \to \infty$ with a decay rate approximately $\exp(-\frac{1}{2}mnh)$. Figure 6 indicates that this result is not optimal. To compute $\mathbb{E}\left[\log |x_t^{1,1} - x_t^{2,1}| \right]$, we simulated 100 sample paths and then took average of those paths. Although $2\lambda < \sigma^2$ for $\sigma = 2$, Δ_n^{ij} converges in this case. Moreover, although the decay rate obtained in Theorem 3.4 decreases, as σ increases, one can see that the decay rate increases as σ increases.
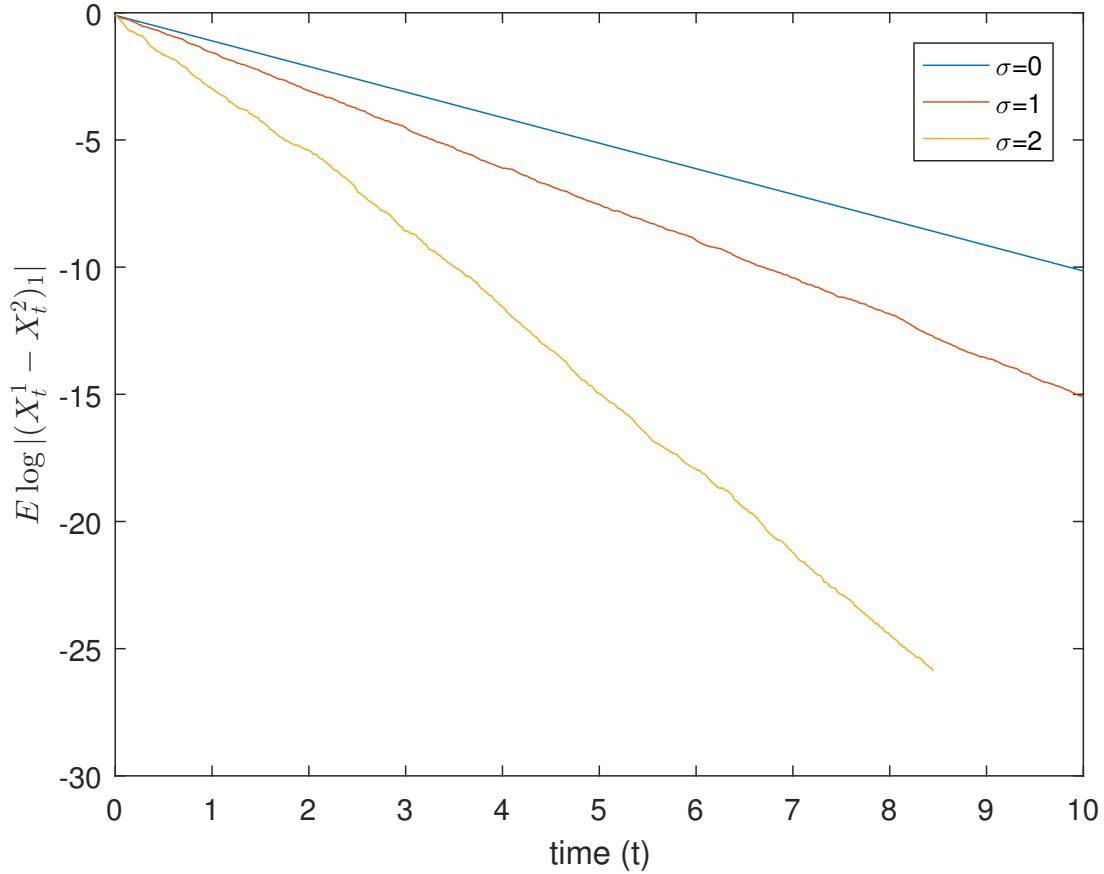
FIGURE 6. Graph of $\mathbb{E}\left[\log|x_n^{1,1} - x_n^{2,1}|\right]$ for $\sigma = 0, 1, 2$.

6. CONCLUSION

In this paper, we have provided a rigorous convergence analysis for the first-order consensus-based optimization algorithm. In [5], the convergence was understood using the corresponding mean-field limit, the Fokker-Planck equation. Thus, the convergence of the original CBO algorithm remains unresolved there. The main contribution of this work is to provide the convergence analysis directly on the CBO algorithm model at the particle level. After rewriting the given continuous optimization algorithm into a first-order consensus form, we use the detailed structure of the coupling term to derive an exact formula for the state differences. Our explicit formula shows that global consensus will emerge for any initial data, whereas in order to prove the convergence toward a global minimum, we need a sufficient condition–which is dimension independent– for systems parameters and initial data to show that the global consensus state tends to a global minimum, as the reciprocal of temperature tends to infinity using Laplace's principle. The emergence of global consensus will emerge for continuous and discrete algorithms. However, we can obtain the convergence analysis only for the continuous algorithm due to Ito's stochastic analysis for twice differentiable and bounded objective functions. In contrast, for discrete algorithm, we do not have available mathematical tools to derive convergence analysis at present.

There are several issues which we cannot deal with in this paper. To name a few, it will be interesting to relax the regularity of the objective function to the less regular objective function, at least continuous one. Finally, random batch methods were used to reduce the computational cost of the N-term summation in [5, 14] in which convergence remains as an open question.

Moreover, it will be interesting to see whether our presented analysis can be applied to other metaheuristic algorithms based on the swarm intelligence. These issues will be discussed in a future work.

References

[1] Acebron, J. A., Bonilla, L. L., Pérez Vicente, C. J. P., Ritort, F. and Spigler, R.: *The Kuramoto model: A simple paradigm for synchronization phenomena.* Rev. Mod. Phys. **77** (2005), 137-185.

[2] Ahn, S. M. and Ha, S.-Y.: *Stochastic flocking dynamics of the Cucker-Smale model with multiplicative white noises.* J. Math. Physics. **51** (2011), 103301.

[3] Albi, G., Bellomo, N., Fermo, L., Ha, S.-Y., Pareschi, L., Poyato, D. and Soler, J.: *Vehicular traffic, crowds, and swarms. On the kinetic theory approach towards research perspectives.* Math. Models Methods Appl. Sci. **29** (2019), 1901-2005.

[4] Bertsekas, D.: *Convex Analysis and Optimization.* Athena Scientific. 2003.

[5] Carrillo, J. A., Jin, S., Li, L. and Zhu, Y.: *A consensus-based global optimization method for high dimensional machine learning problems.* Submitted.

[6] Carrillo, J., Choi, Y.-P., Totzeck, C. and Tse, O.: *An analytical framework for consensus-based global optimization method.* Mathematical Models and Methods in Applied Sciences **28** (2018), 1037-1066.

[7] Choi, Y.-P., Ha, S.-Y. and Li, Z.: *Emergent dynamics of the Cucker-Smale flocking model and its variants.* In N. Bellomo, P. Degond, and E. Tadmor (Eds.), Active Particles Vol.I Theory, Models, Applications (tentative title), Series: Modeling and Simulation in Science and Technology, Birkhauser, Springer. 2017.

[8] Eberhart, R. and Kennedy, J.: *Particle swarm optimization.* Proceedings of the IEEE International Conference on Neural Networks **4** (1995), 1942-1948.

[9] Cucker, F. and Smale, S.: *On the mathematics of emergence.* Japanese Journal of Mathematics **2** (2007), 197-227.

[10] Fang, D., Ha, S.-Y. and Jin, S.: *Emergent behaviors of the Cucker-Smale ensemble under attractive-repulsive couplings and Rayleigh frictions.* Math. Models Methods Appl. Sci. **29** (2019), 1349-1385

[11] Ha, S.-Y. and Liu, J.-G.: *A simple proof of Cucker-Smale flocking dynamics and mean-field limit.* Commun. Math. Sci. **7** (2009), 297-325.

[12] Ha, S.-Y., Lee, K. and Levy, D.: *Emergence of time-asymptotic flocking in a stochastic Cucker-Smale system.* Commun. Math. Sci. **7** (2009), 453-469.

[13] Holland, J. H.: *Genetic algorithms.* Scientific American **267** (1992), 66-73.

[14] Jin, S. Li, L. and Liu, J.-G.: *Random batch methods(RBM) for interacting particle systems.* J. Comp. Phys., to appear. arXiv:1812.10575, 2018.

[15] Kennedy, J.: *Swarm intelligence.* Handbook of nature-inspired and innovative computing. Springer, 187-219 (2006).

[16] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P.: *Optimization by simulated annealing.* Science **220** (1983), 671-680.

[17] Kolokolnikov, T., Carrillo, J. A., Bertozzi, A., Fetecau, R. and Lewis, M.: *Emergent behavior in a multi-particle systems with non-local interactions.* Physica D **260** (2013), 1-4.

[18] Kuramoto, Y.: *Chemical oscillations, waves and turbulence.* Springer-Verlag, Berlin, 1984.

[19] Kuramoto, Y.: *International symposium on mathematical problems in mathematical physics.* Lecture Notes Theor. Phys. **30** (1975), 420.

[20] Laarhoven, P. J. M. van and Aarts, E. H. L.: *Simulated annealing: theory and applications.* D. Reidel Publishing Co., Dordrecht, 1987.

[21] Motsch, S. and Tadmor, E.: *Heterophilious dynamics enhances consensus.* SIAM. Rev. **56** (2014), 577-621.

[22] Peskin, C. S.: *Mathematical aspects of heart physiology.* Courant Institute of Mathematical Sciences, New York, 1975.

[23] Pinnau, R., Totzeck, C., Tse, O. and Martin, S.: *A consensus-based model for global optimization and its mean-field limit.* Math. Models Methods Appl. Sci. **27** (2017), 183-204.

[24] Pikovsky, A., Rosenblum, M. and Kurths, J.: *Synchronization: A universal concept in nonlinear sciences.* Cambridge University Press, Cambridge, 2001.

[25] Totzeck, C. Pinnau, R., Blauth, S. and Schotthófer, S.: *A numerical comparison of consensus-based global optimization to other particle-based global optimization scheme.* Proceedings in Applied Mathematics and Mechanics, **18**, 2018.
[26] Vicsek, T. and Zefeiris, A.: *Collective motion.* Phys. Rep. **517** (2012), 71-140.
[27] Yang, X.-S.: *Nature-inspired metaheuristic algorithms.* Luniver Press, 2010.
[28] Yang, X.-S., Deb, S., Zhao, Y.-X., Fong, S. and He, X.: *Swarm intelligence: past, present and future.* Soft Comput **22** (2018), 5923-5933.

(Seung-Yeal Ha)
Department of Mathematical Sciences and Research Institute of Mathematics,
Seoul National University, Seoul, 08826 and
Korea Institute for Advanced Study, Hoegiro 85, Seoul, 02455, Korea (Republic of)
    E-mail address: syha@snu.ac.kr

(Shi Jin)
School of Mathematical Sciences, MOE-LSC, and Institute of Natural Sciences,
Shanghai Jiao Tong University, Shanghai 200240, China
    E-mail address: shijin-m@sjtu.edu.cn

(Doheon Kim)
School of Mathematics, Korea Institute for Advanced Study,
Hoegiro 85, Seoul 02455, Korea (Republic of)
    E-mail address: doheonkim@kias.re.kr