

CONVERGENCE AND ERROR ESTIMATES FOR TIME-DISCRETE CONSENSUS-BASED OPTIMIZATION ALGORITHMS

SEUNG-YEAL HA, SHI JIN, AND DOHEON KIM

ABSTRACT. We present convergence and error estimates of the time-discrete consensus-based optimization (CBO) algorithms proposed in [4] for general nonconvex functions. In authors' recent work [14], rigorous error analysis of the first-order consensus-based optimization algorithm proposed in [4] was studied at the particle level without resorting to the kinetic equation via a mean-field limit. However, the error analysis for the corresponding time-discrete algorithm was not done mainly due to lack of discrete analogue of Itô's stochastic calculus. In this paper, we provide a simple and elementary convergence and error analysis for a general time-discrete consensus-based optimization algorithm, which includes the three discrete algorithms in [4]. Our analysis provides numerical stability and convergence conditions for the three algorithms, as well as error estimates to the global minimum.

1. INTRODUCTION

The purpose of this work is to complete the convergence and error analysis for discrete consensus-based optimization algorithms introduced in [4, 14]. For modern machine learning methods, one needs to solve non-convex optimization problems in high dimensions. It is well-known that non-convex optimization problem is NP-hard. Usually a deterministic algorithm, such as the gradient descent method, will get stuck to local minima. In order to escape from local minima, or saddle points, one needs to introduce some numerical noises which allow the algorithms to escape from the local minima or saddle points. For this reason, stochastic optimizations—such as the stochastic gradient descent method, have been widely used in machine learning [3]. On the other hand, gradient-free algorithms, which do not need the gradient of objective functions, are attractive for problems with non-smooth objective functions or data-based optimization problems. Meta-heuristic stochastic optimization algorithms belong to the latter category, for example swarm intelligence methods [19, 32] such as particle swarm optimization (in short PSO) [10], simulated annealing method [20, 24], ant-colony algorithm [31], genetic algorithm [17] etc. A basic idea of these meta-heuristic algorithms is to use collective behaviors of underlying individual agents (or particles) coupled with suitable stochastic components in the choice of system parameters. Although each individual moves in some random fashion, one designs suitable communication functions between the particles such that collectively they exhibit some intelligent behavior, such as moving toward the global minimum. Although these algorithms are usually simple to implement and yields reasonably

Date: March 11, 2020.

2010 Mathematics Subject Classification. 37H10, 37N40, 90C26.

Key words and phrases. Consensus-based optimization, Gibb's distribution, global optimization, machine learning, objective function.

Acknowledgment. The work of S.-Y. Ha was supported by National Research Foundation of Korea (NRF-2020R1A2C3A01003881), and the work of S. Jin was supported by NSFC Grant Nos. 11871297 and 3157107, and the work of D. Kim was supported by a KIAS Individual Grant (MG073901) at Korea Institute for Advanced Study. The authors would like to thank to Dr. Dongnam Ko for his helpful comments on the Laplace principle.

good results with suitable choices of parameters, their rigorous convergence analysis were mostly open as far as the authors know.

Our main interest in this paper lies on the first-order consensus based optimization algorithm, first introduced in [27] and studied in [5], and then modified in [4] (see also [29] for comparisons between CBO algorithm and other heuristic algorithms based on collective dynamics). These are swarming intelligence models that can be proved to exhibit consensus approximating the global minimum of general nonconvex functions under suitable conditions on the parameters and initial data. We also refer to recent works on the consensus-based optimization algorithm on the sphere [12, 13].

We begin with the continuous optimization algorithm. Let $X_t^k = (x_t^{k,1}, \dots, x_t^{k,d}) \in \mathbb{R}^d$ be the coordinate of the k -th particle at time t , and $L = L(X)$, $X \in \mathbb{R}^d$ be a *non-convex* objective function to be minimized. Then, the main goal of optimization algorithms is to look for a global minimizer X^* of L in the search space (in our setting, the whole space) if it exists:

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

In a recent work [4], the authors proposed the following variant of the CBO algorithm introduced in [5, 27]:

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

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 assumed to be i.i.d. and satisfy the mean zero and covariance relations:

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

This model is an example of agent-based swarming models which have been studied intensively in recent years, see for example several survey articles [1, 2, 6, 28, 30] and related literature [8, 11, 15, 16, 21, 22, 23, 25, 26].

Next, we consider time-discrete analogue of (1.1). For this, we set

$$h := \Delta t, \quad X_n := X(nh), \quad n = 0, 1, \dots, \dots$$

Then the discrete scheme reads as follows:

$$(1.2) \quad \begin{cases} X_{n+1}^i = X_n^i - \gamma(X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) \eta_n^l e_l, & n \geq 0, i = 1, \dots, N, \\ \bar{X}_n^* = (x_n^{*,1}, \dots, x_n^{*,d}) := \frac{\sum_{j=1}^N X_n^j e^{-\beta L(X_n^j)}}{\sum_{j=1}^N e^{-\beta L(X_n^j)}}, \end{cases}$$

where the random variables $\{\eta_n^l\}_{n,l}$ are i.i.d. with

$$(1.3) \quad \mathbb{E}[\eta_n^l] = 0, \quad \mathbb{E}[(\eta_n^l)^2] = \zeta^2, \quad n = 1, \dots, \quad l = 1, \dots, d.$$

Note that the discrete process $\{\eta_n^l\}_{n,l}$ certainly includes the Gaussian noise process. In this sense, the discrete model (1.2) clearly generalizes the discrete model studied in [14].

In this paper, we are interested in the following issues for the discrete algorithm (1.2):

- (Question A): Does the N -state ensemble $\{X_n^i\}$ exhibit a global consensus? i.e., does
$$X_n^i - X_n^j \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad i, j = 1, \dots, N \quad \text{in suitable sense?}$$
- (Question B): If the answer to the first problem is positive, then under what conditions on system parameters and initial data, does there exist a global consensus state X_∞ such that
$$X_n^i \rightarrow X_\infty \quad \text{for all } i, \quad \text{as } n \rightarrow \infty, \quad \text{such that } L(X_\infty) \sim \min_X L(X).$$

In [14], the above two questions were answered positively for the continuous algorithm (1.1), whereas for the discrete algorithm (1.2), only the first question was discussed in the same paper. For a rigorous error analysis to the continuous algorithm, Itô's calculus was essentially used. The main reason that we did not cover the second question for the discrete algorithm is mainly due to the lack of the discrete analogue of Itô's stochastic calculus.

In this paper, we revisit the second question on the convergence of the discrete algorithm (1.2) and provide positive answer for the generalized discrete algorithm (1.2) which can cover several discrete algorithms proposed in [4] (see Section 2.1 for details). Our analysis also provides numerical stability conditions and error estimates or these algorithms

Next, we summarize our main results as follows. First, we provide several stochastic global consensus results in suitable sense. More precisely, if system parameters γ and ζ in (1.2) and (1.3) satisfy

$$(1.4) \quad |1 - \gamma| < 1, \quad 0 \leq \zeta \leq \infty,$$

then the expectation of $X_n^i - X_n^j$ tends to zero asymptotically:

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n^i - X_n^j] = 0, \quad \forall i, j = 1, \dots, N.$$

On the other hand, if system parameters satisfy a more restricted condition compared to (1.4):

$$(1.5) \quad (1 - \gamma)^2 + \zeta^2 < 1,$$

then L^2 -global consensus and almost-sure global consensus occur asymptotically:

$$\lim_{n \rightarrow \infty} \mathbb{E}|X_n^i - X_n^j|^2 = 0 \quad \text{and} \quad \mathbb{P}\left\{ \lim_{n \rightarrow \infty} |X_n^i - X_n^j| = 0 \right\} = 1, \quad \forall i, j = 1, \dots, N,$$

where $|\cdot|$ stands for the L^2 -norm of a vector.

Moreover, the above L^2 -global consensus also implies the L^1 -global consensus. We refer to Definition 2.1 and Theorem 2.1 for the definition of convergences and details.

Note that the above global consensus does not imply the existence of asymptotic consensus state X_∞ independent of i such that

$$(1.6) \quad X_n^i \rightarrow X_\infty \quad \text{in suitable sense,}$$

i.e., the process can fluctuate, but the relative distances tends to zero asymptotically, for example, the sample paths may tend to periodic orbit or limit cycle. Our second result provides the condition under which the process tends to a common fixed random variable X_∞ (see (1.6)). In

fact, under the same assumption (1.5), one can show that there exists a common constant state $X_\infty = (x_\infty^1, \dots, x_\infty^d)$ such that

$$\lim_{n \rightarrow \infty} X_n^i = X_\infty \quad \text{a.s., } 1 \leq i \leq N.$$

(see Theorem 3.1 for details).

Finally, our last result establishes the condition under which the asymptotic state X_∞ lies in a small neighborhood of a unique global minimum X_m for a large β . More precisely, under the assumption (1.5) and for a well-prepared initial random variable X_{in} such that $X_n^i \sim X_{in}$, we derive a key estimate (see Proposition 3.1):

$$\operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) \leq -\frac{1}{\beta} \log \mathbb{E} e^{-\beta L(X_{in})} - \frac{1}{\beta} \log \varepsilon.$$

Then, by Laplace's method ([18]), one can derive

$$\operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) \leq L(X_*) + \frac{d \log \beta}{2\beta} + E(\beta),$$

for some function $E(\beta) = O\left(\frac{1}{\beta}\right)$, $\beta \gg 1$. We refer to Remark A.1 for the intriguing relation between β and admissible reference random variable X_{in} .

The rest of this paper is organized as follows. In Section 2, we provide reductions of previously studied discrete algorithms to (1.2) - (1.3), and then study our first set of main results on the global consensus. In Section 3, we study the emergence of a global consensus state and provide error estimates toward the global minimum for (1.2). Finally, Section 4 is devoted to a brief summary of our main results and some remaining issues to be explored in a future work.

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.

2. EMERGENCE OF GLOBAL CONSENSUS

In this section, we first show that several discrete algorithms introduced in [4, 14] can be reduced to our generalized discrete algorithm (1.2) and then we provide a sufficient framework leading to the global consensus for the discrete optimization algorithm in terms of system parameters and initial data.

2.1. Discrete algorithms. In this subsection, we show how the previous discrete algorithms studied can be reduced as special cases for our general discrete model.

Let $X_n^k = (x_n^{k,1}, \dots, x_n^{k,d}) \in \mathbb{R}^d$ be the position of the k -th particle at time n ($1 \leq k \leq N$). Suppose that the function $L : \mathbb{R}^d \rightarrow \mathbb{R}$ has exactly one global minimum, and we consider the general discrete consensus-based optimization algorithm:

$$(2.1) \quad \begin{cases} X_{n+1}^i = X_n^i - \gamma(X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) \eta_n^l e_l, & n \geq 0, \quad i = 1, \dots, N, \\ \bar{X}_n^* = (x_n^{*,1}, \dots, x_n^{*,d}) := \frac{\sum_{j=1}^N X_n^j e^{-\beta L(X_n^j)}}{\sum_{j=1}^N e^{-\beta L(X_n^j)}}, \end{cases}$$

where $\{e_l\}$ is the standard orthonormal basis in \mathbb{R}^d and random variables $\{\eta_n^l\}_{n,l}$ are i.i.d with

$$(2.2) \quad \mathbb{E}[\eta_n^l] = 0 \quad \text{and} \quad \mathbb{E}|\eta_n^l|^2 = \zeta^2.$$

In the sequel, we consider the following three discrete algorithms.

- Model A: Consider the first-order Euler type discrete model in [14]:

$$X_{n+1}^i = X_n^i - \lambda h (X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) \sigma \sqrt{h} Z_n^l e_l, \quad n \geq 0, \quad i = 1, \dots, N,$$

where the random variables $\{Z_n^l\}_{n,l}$ are i.i.d standard normal distributions, i.e. $Z_n^l \sim \mathcal{N}(0, 1^2)$. If we set

$$(2.3) \quad \gamma := \lambda h \quad \text{and} \quad \eta_n^l := \sigma \sqrt{h} Z_n^l.$$

Then, the above setting clearly satisfies the relations (2.2) with $\zeta = \sigma \sqrt{h}$.

- Model B: Consider a predictor-corrector type discrete model in [4].

$$(2.4) \quad \begin{cases} \hat{X}_n^i = \bar{X}_n^* + e^{-\lambda h} (X_n^i - \bar{X}_n^*), \\ X_{n+1}^i = \hat{X}_n^i - \sum_{l=1}^d (\hat{x}_n^{i,l} - \bar{x}_n^{*,l}) \sigma \sqrt{h} Z_n^l e_l, \end{cases} \quad n \geq 0, \quad i = 1, \dots, N.$$

We substitute (2.4)₁ into (2.4)₂ and use an addition-subtraction trick to see that

$$X_{n+1}^i = X_n^i - (1 - e^{-\lambda h}) (X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) e^{-\lambda h} \sigma \sqrt{h} Z_n^l e_l, \quad n \geq 0, \quad i = 1, \dots, N.$$

If we set

$$(2.5) \quad \gamma := 1 - e^{-\lambda h} \quad \text{and} \quad \eta_n^l := e^{-\lambda h} \sigma \sqrt{h} Z_n^l,$$

then (2.4) reduces to the special case of (2.1) - (2.2) with $\zeta = e^{-\lambda h} \sigma \sqrt{h}$.

- Model C: Consider one of discrete optimization model proposed in [4]:

$$(2.6) \quad X_{n+1}^i = \bar{X}_n^* + \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) \left[\exp \left(- \left(\lambda + \frac{1}{2} \sigma^2 \right) h + \sigma \sqrt{h} Z_n^l \right) \right] e_l, \quad n \geq 0, \quad i = 1, \dots, N,$$

Again, the R.H.S. of (2.6) can be rewritten as

$$X_{n+1}^i = X_n^i - (1 - e^{-\lambda h}) (X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) e^{-\lambda h} \left[\exp \left(-\frac{1}{2} \sigma^2 h + \sigma \sqrt{h} Z_n^l \right) - 1 \right] e_l.$$

We set

$$(2.7) \quad \gamma := 1 - e^{-\lambda h} \quad \text{and} \quad \eta_n^l := e^{-\lambda h} \left[\exp \left(-\frac{1}{2} \sigma^2 h + \sigma \sqrt{h} Z_n^l \right) - 1 \right].$$

Then, we use the elementary facts [7]:

$$X \sim \text{Lognormal}(\alpha, \beta^2) \quad \Rightarrow \quad \mathbb{E}X = e^{\alpha + \frac{\beta^2}{2}} \quad \text{and} \quad \mathbb{E}X^2 = e^{2\alpha + 2\beta^2}$$

to see that (2.7) satisfies moment relations (2.2) with $\zeta = e^{-\lambda h} \sqrt{e^{\sigma^2 h} - 1}$.

2.2. Global consensus. In this subsection, we show that the global consensus for (2.1) occurs asymptotically for all initial data under suitable conditions on γ and ζ . First, we recall the concepts of L^p and almost sure global consensus in the following definition.

Definition 2.1. Let $\mathcal{X} = \{\mathcal{X}_n := (x_n^1, \dots, x_n^d)\}$ be a stochastic process, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space.

(1) The configuration process \mathcal{X} exhibits a global consensus in L^p with $p \geq 1$, if the following zero L^p -convergence holds:

$$\lim_{n \rightarrow +\infty} \mathbb{E}|X_n^i - X_n^j|^p = 0, \quad \forall i, j = 1, \dots, N.$$

(2) The configuration process \mathcal{X} exhibits a global consensus almost surely if for almost sure $\omega \in \Omega$ and $i, j = 1, \dots, N$, the sample path $X_n^i(\omega) - X_n^j(\omega)$ tends to zero asymptotically.

$$\mathbb{P}\left\{\lim_{n \rightarrow \infty} |X_n^i - X_n^j| = 0\right\} = 1, \quad \forall i, j = 1, \dots, N.$$

Remark 2.1. Recall that L^2 -convergence implies L^1 convergence on a probability space.

Now, we consider the relation for the process $X_n^i - X_n^j$. For this, we consider the discrete algorithms: for $i, j = 1, \dots, N$,

$$(2.8) \quad \begin{cases} X_{n+1}^i = X_n^i - \gamma(X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) \eta_n^l e_l, \\ X_{n+1}^j = X_n^j - \gamma(X_n^j - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{j,l} - \bar{x}_n^{*,l}) \eta_n^l e_l, \end{cases}$$

Then, it follows from (2.8) that $x_{n+1}^i - x_{n+1}^j$ satisfies

$$(2.9) \quad x_{n+1}^{i,l} - x_{n+1}^{j,l} = (1 - \gamma - \eta_n^l)(x_n^{i,l} - x_n^{j,l}).$$

In the following lemma, we provide several estimates for $x_n^{i,l} - x_n^{j,l}$.

Lemma 2.1. Let $\{\mathcal{X}_n\}$ be a solution process to (2.1). Then the following estimates hold.

- (i) $\mathbb{E}[X_n^i - X_n^j] = (1 - \gamma)^n \mathbb{E}[X_0^i - X_0^j]$.
- (ii) $\mathbb{E}|X_n^i - X_n^j|^2 = ((1 - \gamma)^2 + \zeta^2)^n \mathbb{E}|X_0^i - X_0^j|^2$.
- (iii) $|x_n^{i,l} - x_n^{j,l}|^2 \leq |x_0^{i,l} - x_0^{j,l}|^2 e^{-nY_n^l(\omega)}$, a.s. $\omega \in \Omega$,

where Y_n^l is a random variable satisfying

$$\lim_{n \rightarrow \infty} Y_n^l(\omega) = 2\gamma - \gamma^2 - \zeta^2, \quad a.s. \omega \in \Omega, \quad l = 1, \dots, d.$$

Proof. The estimates below for a special case (Model A) can be found in Theorem 3.4 in [14] and proofs are almost similar. However, for self-containedness of this paper, we present their proofs.

(i) It follows from the recursive relation (2.9) that

$$(2.10) \quad x_n^{i,l} - x_n^{j,l} = (x_0^{i,l} - x_0^{j,l}) \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l), \quad l = 1, \dots, d.$$

Now we take expectation on both sides of (2.10) using the independence of η_m^l and $x_0^{i,l} - x_0^{j,l}$ to get

$$\mathbb{E}[x_n^{i,l} - x_n^{j,l}] = (1 - \gamma)^n \mathbb{E}[x_0^{i,l} - x_0^{j,l}], \quad \text{i.e.,} \quad \mathbb{E}[X_n^i - X_n^j] = (1 - \gamma)^n \mathbb{E}[X_0^i - X_0^j].$$

(ii) We take the square of (2.10) to see

$$(2.11) \quad |x_n^{i,l} - x_n^{j,l}|^2 = |x_0^{i,l} - x_0^{j,l}|^2 \prod_{m=0}^{n-1} |1 - \gamma - \eta_m^l|^2, \quad l = 1, \dots, d.$$

Then, (1.3), (2.11), independence of η_m^l and $x_0^{i,l} - x_0^{j,l}$ yield

$$\mathbb{E}|x_n^{i,l} - x_n^{j,l}|^2 = ((1 - \gamma)^2 + \zeta^2)^n \mathbb{E}|x_0^{i,l} - x_0^{j,l}|^2,$$

i.e.,

$$\mathbb{E}|X_n^i - X_n^j|^2 = ((1 - \gamma)^2 + \zeta^2)^n \mathbb{E}|X_0^i - X_0^j|^2.$$

(iii) It follows from the inequality $e^x \geq 1 + x$, $x \in \mathbb{R}$ that

$$(2.12) \quad (1 - \gamma - \eta_m^l)^2 \leq e^{(1-\gamma-\eta_m^l)^2-1} = e^{-(\gamma+\eta_m^l)(2-\gamma-\eta_m^l)}.$$

Then, we use (2.11) and (2.12) to obtain

$$\begin{aligned} |x_n^{i,l} - x_n^{j,l}|^2 &\leq |x_0^{i,l} - x_0^{j,l}|^2 \prod_{m=0}^{n-1} e^{-(\gamma+\eta_m^l)(2-\gamma-\eta_m^l)} \\ &= |x_0^{i,l} - x_0^{j,l}|^2 \exp \left[-n \times \frac{1}{n} \sum_{m=0}^{n-1} (\gamma + \eta_m^l)(2 - \gamma - \eta_m^l) \right]. \end{aligned}$$

We set

$$Y_n := \frac{1}{n} \sum_{m=0}^{n-1} (\gamma + \eta_m^l)(2 - \gamma - \eta_m^l).$$

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

$$Y_n \longrightarrow \mathbb{E} \left[(\gamma + \eta_m^l)(2 - \gamma - \eta_m^l) \right] = 2\gamma - \gamma^2 - \zeta^2 \quad \text{a.s. as } n \rightarrow \infty.$$

□

As a direct application of Lemma 2.1, we have the following global consensus estimates.

Theorem 2.1. *Let $\{\mathcal{X}_n\}$ be a solution process to (2.1). Then, the following three global consensus results hold.*

(1) *Suppose that system parameters satisfy*

$$|\gamma - 1| < 1 \quad \text{and} \quad 0 \leq \zeta \leq \infty.$$

Then, $\mathbb{E}[X_n^i - X_n^j]$ tends to zero asymptotically:

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n^i - X_n^j] = 0, \quad \forall i, j = 1, \dots, N.$$

(2) *Suppose that system parameters γ and ζ satisfy*

$$(\gamma - 1)^2 + \zeta^2 < 1$$

then, L^2 and almost-sure global consensus emerge asymptotically: for a.s. $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \mathbb{E}|X_n^i - X_n^j|^2 = 0, \quad |x_n^{i,l} - x_n^{j,l}|^2 \leq |x_0^{i,l} - x_0^{j,l}|^2 e^{-nY_n^l(\omega)}, \quad i, j = 1, \dots, N, \quad l = 1, \dots, d,$$

where Y_n^l is a random variable satisfying

$$\lim_{n \rightarrow \infty} Y_n^l(\omega) = 1 - (\gamma - 1)^2 - \zeta^2 > 0, \quad \text{a.s. } \omega \in \Omega, \quad l = 1, \dots, d.$$

Proof. Since the proofs follow from Lemma 2.1 directly, we omit them here. \square

As a corollary of Theorem 2.1, one has the following global consensus for Model A - Model C discussed in previous subsection.

Corollary 2.1. *The following assertions hold.*

(1) *Suppose that system parameters satisfy*

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

then, Model A admits L^2 and almost sure global consensus.

(2) *Suppose that system parameters satisfy*

$$(1 + \sigma^2 h)e^{-2\lambda h} < 1,$$

then, Model B admits L^2 and almost sure global consensus.

(3) *Suppose that system parameters satisfy*

$$(2.13) \quad \lambda > \frac{\sigma^2}{2},$$

then, Model C admits L^2 and almost sure global consensus, for any $h > 0$.

Proof. (i) For Model A, we use relations $\gamma = \lambda h$, $\zeta = \sigma\sqrt{h}$ in (2.3) to see

$$2\gamma - \gamma^2 - \zeta^2 = 2\lambda h - (\lambda h)^2 - \sigma^2 h > 0, \quad \text{for } 0 < h < \frac{2\lambda - \sigma^2}{\lambda^2}.$$

Then, we use (3) in Theorem 2.1 to derive the desired estimate.

(2) For Model B, we use relations $\gamma = 1 - e^{-\lambda h}$, $\zeta = \sigma\sqrt{h}$ in (2.5) to see

$$2\gamma - \gamma^2 - \zeta^2 = 2(1 - e^{-\lambda h}) - (1 - e^{-\lambda h})^2 - \sigma^2 h e^{-2\lambda h} = 1 - (1 + \sigma^2 h)e^{-2\lambda h} > 0.$$

(3) For Model C, we use $\gamma = 1 - e^{-\lambda h}$, $\zeta = e^{-\lambda h}\sqrt{e^{\sigma^2 h} - 1}$ in (2.7) to get

$$2\gamma - \gamma^2 - \zeta^2 = 2(1 - e^{-\lambda h}) - (1 - e^{-\lambda h})^2 - (e^{\sigma^2 h} - 1)e^{-2\lambda h} = 1 - e^{(\sigma^2 - 2\lambda)h} > 0 \quad \Leftrightarrow \quad (2\lambda - \sigma^2)h > 0. \quad \square$$

Remark 2.2. *The above corollary provides the stability conditions for three algorithms. For Model C, the algorithm is unconditionally stable provided (2.13) holds, namely one can choose arbitrary $h > 0$. Model B is also unconditionally stable provided $\lambda \geq \frac{\sigma^2}{2}$ is satisfied, because*

$$e^{2\lambda h} > 1 + 2\lambda h \geq 1 + \sigma^2 h, \quad \text{for } h > 0.$$

Model A is conditionally stable.

3. CONVERGENCE ANALYSIS AND ERROR ESTIMATES

In this section, we provide a convergence analysis with error estimates for the discrete CBO algorithm. In previous section, we studied sufficient conditions leading to the global consensus which does not mean X_n^i tends to a common fixed state X_∞ . Thus, two main issues to be covered in this section are two-fold.

- (Q1): What is a sufficient framework leading to the common asymptotic state:

$$X_n^i(\beta) \rightarrow X_\infty(\beta), \quad \text{as } n \rightarrow \infty \text{ for all } i = 1, \dots, N?$$

- (Q2): If the above question is resolved, then how close is the asymptotic state X_∞ to the global minimum X_m of L if the latter exists?

3.1. Emergence of common consensus state. For the emergence of common consensus state, we first introduce an ensemble average:

$$\bar{X}_n := \frac{1}{N} \sum_{i=1}^N X_n^i = (\bar{x}_n^1, \dots, \bar{x}_n^d).$$

Before we present our second main result, we present two elementary lemmas to be crucially used in the proof of convergence analysis.

Lemma 3.1. *Let $\{X_n^i\}_{1 \leq i \leq N}$ be a solution to (2.1). Then, the following estimates hold almost surely.*

$$\begin{aligned} (i) \quad & |X_n^i - \bar{X}_n|^2 = \sum_{l=1}^d (x_0^{i,l} - \bar{x}_0^l)^2 \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l)^2. \\ (ii) \quad & |\bar{X}_n - \bar{X}_n^*|^2 \leq \max_{1 \leq i \leq N} |X_n^i - \bar{X}_n|^2. \\ (iii) \quad & \frac{1}{N} \sum_{i=1}^d |X_n^i - \bar{X}_n^*|^2 \leq 2 \sum_{l=1}^d \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l)^2. \end{aligned}$$

Proof. (i) It follows from (2.1) that

$$(3.1) \quad \bar{X}_{n+1} - \bar{X}_n = -\gamma(\bar{X}_n - \bar{X}_n^*) - \sum_{l=1}^d (\bar{x}_n^l - \bar{x}_n^{*,l}) \eta_n^l e_l.$$

We subtract (3.1) from (2.1) to obtain

$$(3.2) \quad (X_{n+1}^i - \bar{X}_{n+1}) - (X_n^i - \bar{X}_n) = -\gamma(X_n^i - \bar{X}_n) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^l) \eta_n^l e_l.$$

The l -th component of (3.2) implies

$$x_n^{i,l} - \bar{x}_n^l = (x_0^{i,l} - \bar{x}_0^l) \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l).$$

This yields

$$|X_n^i - \bar{X}_n|^2 = \sum_{l=1}^d (x_0^{i,l} - \bar{x}_0^l)^2 \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l)^2.$$

(ii) We apply the triangle inequality and the Cauchy-Schwarz inequality:

$$\begin{aligned} |\bar{X}_n - \bar{X}_n^*|^2 &= \left| \frac{\sum_{k=1}^N e^{-\beta L(X_n^k)} (\bar{X}_n - X_n^k)}{\sum_{k=1}^N e^{-\beta L(X_n^k)}} \right|^2 \leq \left[\frac{\sum_{k=1}^N e^{-\beta L(X_n^k)} |\bar{X}_n - X_n^k|}{\sum_{k=1}^N e^{-\beta L(X_n^k)}} \right]^2 \\ &\leq \frac{\sum_{k=1}^N e^{-\beta L(X_n^k)} |\bar{X}_n - X_n^k|^2}{\sum_{k=1}^N e^{-\beta L(X_n^k)}} \leq \max_{1 \leq k \leq N} |\bar{X}_n - X_n^k|^2. \end{aligned}$$

(iii) Note that

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^d |X_n^i - \bar{X}_n^*|^2 &= \frac{1}{N} \sum_{i=1}^d (|X_n^i - \bar{X}_n|^2 + 2(X_n^i - \bar{X}_n) \cdot (\bar{X}_n - \bar{X}_n^*) + |\bar{X}_n - \bar{X}_n^*|^2) \\
&= \frac{1}{N} \sum_{i=1}^d (|X_n^i - \bar{X}_n|^2 + |\bar{X}_n - \bar{X}_n^*|^2) \leq 2 \max_{1 \leq i \leq N} |X_n^i - \bar{X}_n|^2 \\
&= 2 \max_{1 \leq i \leq N} \left(\sum_{l=1}^d (x_0^{i,l} - \bar{x}_0^l)^2 \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l)^2 \right) \\
&\leq 2 \sum_{l=1}^d \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \prod_{m=0}^{n-1} (1 - \gamma - \eta_m^l)^2,
\end{aligned}$$

where we used the results from (i) and (ii). \square

Lemma 3.2. *Let $\{X_n^i\}_{1 \leq i \leq N}$ be a solution to (2.1). Then, one has*

$$\mathbb{E}|X_n^i - \bar{X}_n^*|^2 \leq 2((1 - \gamma)^2 + \zeta^2)^n \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right).$$

Proof. We take expectation on both sides of the inequality in Lemma 3.1 (iii) to get

$$\begin{aligned}
\mathbb{E}|X_n^i - \bar{X}_n^*|^2 &\leq 2 \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \prod_{m=0}^{n-1} \mathbb{E} (1 - \gamma - \eta_m^l)^2 \\
&\leq 2 \left((1 - \gamma)^2 + \zeta^2 \right)^n \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right).
\end{aligned}$$

\square

Next, we provide our second main result.

Theorem 3.1. *Suppose that system parameters satisfy*

$$(1 - \gamma)^2 + \zeta^2 < 1,$$

and let $\{X_n^i\}_{1 \leq i \leq N}$ be a solution to (2.1). Then, there exists a common constant state $X_\infty = (x_\infty^1, \dots, x_\infty^d)$ such that

$$\lim_{n \rightarrow \infty} X_n^i = X_\infty \text{ a.s., } 1 \leq i \leq N.$$

Proof. Note that summing the equation (2.1) over n yields the following relation: for $i = 1, \dots, N$ and $l = 1, \dots, d$,

$$x_n^{i,l} = x_0^{i,l} - \gamma \sum_{m=0}^{n-1} (x_m^{i,l} - \bar{x}_m^{*,l}) - \sum_{m=0}^{n-1} (x_m^{i,l} - \bar{x}_m^{*,l}) \eta_m^l =: x_0^{i,l} - \mathcal{I}_{31} - \mathcal{I}_{32}.$$

Next, we show the a.s. convergence of \mathcal{I}_{31} and \mathcal{I}_{32} separately.

- (Estimate of \mathcal{I}_{31}): By Lemma 3.1 (iii), we have

$$\begin{aligned}
 |x_m^{i,l} - \bar{x}_m^{*,l}| &\leq \sum_{i=1}^d |X_m^i - \bar{X}_m^*| \\
 &\leq \sqrt{2N \sum_{l=1}^d \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \prod_{p=0}^{m-1} (1 - \gamma - \eta_p^l)^2} \\
 &\leq \sqrt{2N \sum_{l=1}^d \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \prod_{p=0}^{m-1} \exp((1 - \gamma - \eta_p^l)^2 - 1)} \\
 &= \sqrt{2N \sum_{l=1}^d \left(\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \exp \left(\sum_{p=0}^{m-1} (-2\gamma + \gamma^2 - 2(1 - \gamma)\eta_p^l + (\eta_p^l)^2) \right)}.
 \end{aligned}$$

On the other hand, by the strong law of large numbers, one has

$$\begin{aligned}
 \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{p=0}^{m-1} \left(-2\gamma + \gamma^2 - 2(1 - \gamma)\eta_p^l + (\eta_p^l)^2 \right) \\
 = \mathbb{E} \left[-2\gamma + \gamma^2 - 2(1 - \gamma)\eta_p^l + (\eta_p^l)^2 \right] = -(2\gamma - \gamma^2 - \zeta^2), \quad \text{a.s.}
 \end{aligned}$$

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

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

where C_1 and C_2 are positive constants. We set

$$\mathcal{J}_{31} := \mathcal{I}_{31} - \gamma \sum_{m=0}^{n-1} C_1 e^{-C_2 m} = \gamma \sum_{m=0}^{n-1} \underbrace{(x_m^{i,l} - \bar{x}_m^{*,l} - C_1 e^{-C_2 m})}_{\leq 0}.$$

Since the summand is nonpositive a.s., \mathcal{J}_{31} is non-increasing in n a.s.

On the other hand, note that

$$\begin{aligned}
 \mathcal{J}_{31} &= \mathcal{I}_{31} + \gamma \sum_{m=0}^{n-1} C_1 e^{-C_2 m} - 2\gamma \sum_{m=0}^{n-1} C_1 e^{-C_2 m} \\
 &= \gamma \sum_{m=0}^{n-1} \underbrace{(x_m^{i,l} - \bar{x}_m^{*,l} + C_1 e^{-C_2 m})}_{\geq 0} - 2\gamma C_1 \frac{1 - e^{-C_2 n}}{1 - e^{-C_2}} \geq -\frac{2\gamma C_1}{1 - e^{-C_2}}.
 \end{aligned}$$

Since \mathcal{J}_{31} is monotone decreasing and bounded below along sample paths, one has

$$\exists \alpha = \lim_{n \rightarrow \infty} \mathcal{J}_{31}(n) = \lim_{n \rightarrow \infty} \left(\mathcal{I}_{31} - \gamma \sum_{m=0}^{n-1} C_1 e^{-C_2 m} \right), \quad \text{a.s.}$$

This implies

$$\lim_{n \rightarrow \infty} \mathcal{I}_{31} = \alpha + \frac{C_1 \gamma}{1 - e^{-C_2}}, \quad \text{a.s.}$$

- (Estimate of \mathcal{I}_{32}): Note that \mathcal{I}_{32} is martingale and its $L^2(\Omega)$ -norm is uniformly bounded in n :

$$\begin{aligned} \mathbb{E} \left[\sum_{m=0}^{n-1} (x_m^{i,l} - \bar{x}_m^{*,l}) \eta_m^l \right]^2 &= \zeta^2 \sum_{m=0}^{n-1} \mathbb{E} (x_m^{i,l} - \bar{x}_m^{*,l})^2 \\ &\leq 2N\zeta^2 \left(\sum_{m=0}^{n-1} ((1-\gamma)^2 + \zeta^2)^m \right) \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \\ &\leq \frac{2N\zeta^2}{2\gamma - \gamma^2 - \zeta^2} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right). \end{aligned}$$

In the first inequality we used (iii). Hence $\lim_{n \rightarrow \infty} \mathcal{I}_{32}$ exists a.s. Now we have shown that for each $i = 1, \dots, N$, there exists some random variable X_∞^i such that

$$\lim_{n \rightarrow \infty} X_n^i = X_\infty^i \quad \text{a.s.}$$

Recall that by Theorem 2.1, for any $1 \leq i, j \leq N$,

$$\lim_{n \rightarrow \infty} |X_n^i - X_n^j| = 0, \quad \text{a.s.}$$

Hence, there exists X_∞ such that

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

□

3.2. Error estimate. In this subsection, we study an error analysis of (2.1) toward the global minimum. Below, we present a sufficient framework (A1) – (A3) for error analysis in terms of the objective function L , global minimum point X_* and reference random variable X_{in} as follows:

- (A1): Let $L = L(x)$ be a C^2 -objective function satisfying the following relations:

$$L_m := \min_{x \in \mathbb{R}^d} L(x) > 0 \quad \text{and} \quad C_L := \sup_{x \in \mathbb{R}^d} \|\nabla^2 L(x)\|_2 < \infty,$$

where $\|\cdot\|_2$ denotes the spectral norm.

- (A2): Let X_* be the unique global minimum point of L in \mathbb{R}^d satisfying the local convexity relation:

$$\det(\nabla^2 L(X_*)) > 0.$$

- (A3): Let X_{in} be a reference random variable with a law which is absolutely continuous with respect to the Lebesgue measure, and let f be the probability density function of X_{in} satisfying the following conditions:

$$f \text{ is compactly supported, continuous at } X_*, \text{ and } f(X_*) > 0.$$

In the next theorem, we study how close is the common consensus state X_∞ to the global minimum X_* in suitable sense. Now, we are ready to provide an error analysis of the discrete CBO algorithm, which is analogous to Theorem 4.1 in [14] for the continuous case.

Theorem 3.2. *Suppose that the framework (A1) – (A3) holds, and system parameters β, γ, ζ and the initial data $\{X_0^i\}$ satisfy*

$$(3.3) \quad \begin{aligned} & \beta > 0, \quad (\gamma - 1)^2 + \zeta^2 < 1, \quad X_0^i : i, i.d., \quad X_0^i \sim X_{in}, \\ & (1 - \varepsilon) \mathbb{E} \left[e^{-\beta L(X_{in})} \right] \\ & \geq \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2)(\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{1 - e^{-[1 - (\gamma - 1)^2 - \zeta^2]}} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right), \end{aligned}$$

for some $0 < \varepsilon < 1$. Then for a solution $\{X_n^i\}_{1 \leq i \leq N}$ to (1.1), one has the following error estimate:

$$(3.4) \quad \left| \operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) - L(X_*) \right| \leq \frac{d \log \beta}{2\beta} + E(\beta),$$

for some function $E(\beta) = \mathcal{O}\left(\frac{1}{\beta}\right)$.

Remark 3.1. 1. Note that the third condition (3.3)₂:

$$(3.5) \quad (1 - \varepsilon) \mathbb{E} \left[e^{-\beta L(X_{in})} \right] \geq \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2)(\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{1 - e^{-[1 - (\gamma - 1)^2 - \zeta^2]}} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right),$$

does hold only for some intermediate β for given initial data satisfying

$$\sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) > 0.$$

This can be checked as follows. We multiply $e^{\beta L_m}$ on the both sides of (3.5) to get

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

By letting $\beta \rightarrow \infty$ in the above relation, we can see that L.H.S. is less than or equal to $1 - \varepsilon$, but R.H.S. goes to $+\infty$. Hence, the estimate (3.4) is only an error estimate for the discrete CBO algorithm for fixed β (indeed numerically one chooses a fixed β for computation), not viewed as a convergence for $\beta \rightarrow \infty$ (which would impose restrictive constratin on the initial data from (3.3)₂). In Appendix A, we provide an alternative error analysis for (1.1) without employing Laplace's principle.

2. In the sequel, we check how the conditions in (3.3) can be reinterpreted for Model A - Model C in Section 2.1:

◇ **Model A:** We use $(\lambda, \zeta) = (\lambda h, \sigma\sqrt{h})$ to rewrite (3.3) as

$$\begin{aligned} \beta > 0, \quad \lambda > \frac{\sigma^2}{2}, \quad 0 < h < \frac{2\lambda - \sigma^2}{\lambda^2}, \\ (1 - \varepsilon)\mathbb{E}\left[e^{-\beta L(X_{in})}\right] &\geq \frac{2C_L\sqrt{h(\sigma^2 + \lambda^2 h)[2 + h(\sigma^2 + \lambda^2 h - 2\lambda)]}\beta e^{-\beta L_m}}{1 - e^{-\lambda(2h - \lambda h^2 - \sigma^2)}} \\ &\quad \times \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right). \end{aligned}$$

◇ **Model B:** We use $(\lambda, \zeta) = (1 - e^{-\lambda h}, e^{-\lambda h}\sigma\sqrt{h})$ to rewrite (3.3) as

$$\begin{aligned} \beta > 0, \quad h > 0, \quad (1 + \sigma^2 h)e^{-2\lambda h} < 1, \\ (1 - \varepsilon)\mathbb{E}\left[e^{-\beta L(X_{in})}\right] &\geq \frac{2C_L\sqrt{[1 + e^{-\lambda h}(-2 + e^{-\lambda h}(1 + \sigma^2 h))][1 + e^{-2\lambda h}(1 + \sigma^2 h)]}\beta e^{-\beta L_m}}{1 - e^{-[1 - (1 + \sigma^2 h)e^{-2\lambda h}]}} \\ &\quad \times \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right). \end{aligned}$$

◇ **Model C:** We use $(\lambda, \zeta) = (1 - e^{-\lambda h}, e^{-\lambda h}\sqrt{e^{\sigma^2 h} - 1})$ to rewrite (3.3) as

$$\begin{aligned} \beta > 0, \quad h > 0, \quad \lambda > \frac{\sigma^2}{2}, \\ (1 - \varepsilon)\mathbb{E}\left[e^{-\beta L(X_{in})}\right] &\geq \frac{2C_L\sqrt{(1 + e^{h(\sigma^2 - 2h)})(1 + e^{-\lambda h}(e^{-\lambda h} + e^{\sigma^2 h} - 3))}\beta e^{-\beta L_m}}{1 - e^{-(1 - e^{(\sigma^2 - 2\lambda)h})}} \\ &\quad \times \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right). \end{aligned}$$

Before we present a proof of Theorem 3.2, we first briefly review Laplace's principle with a convergence rate.

• **(Laplace's principle with a rate for $\beta \gg 1$):** Note that under some suitable conditions on L and the law of X_{in} , we can apply Varadhan's lemma (cf. [9]) to get

$$(3.6) \quad \lim_{\beta \rightarrow \infty} \left(-\frac{1}{\beta} \log \mathbb{E} e^{-\beta L(X_{in})} \right) = \operatorname{ess\,inf}_{\omega \in \Omega} L(X_{in}).$$

However, as far as the authors know, standard proofs for Varadhan's lemma do not yield a convergence rate. Hence, we try a different approach for a possible convergence rate in (3.6).

Recall that the set $D \in \mathbb{R}^d$ is called a *d-dimensional closed domain* if it is a bounded finitely connected open set (in \mathbb{R}^d) plus its boundary, where the boundary is a Euclidean $(d - 1)$ -surface. One simple example of such D is a closed ball in \mathbb{R}^d . In the sequel, let D be a *d-dimensional closed domain*.

Lemma 3.3. [18] *Suppose that two functions $\phi : D \rightarrow \mathbb{R}$ and $h : D \rightarrow \mathbb{R}$ satisfy the following conditions:*

- (1) *h is positive and is of class C^2 .*
- (2) *ϕh^n is absolutely integrable over D , $n = 1, 2, \dots$.*

- (3) h has an absolutely maximum value at an interior point ξ of D and $\det(-\nabla^2 h(\xi)) > 0$.
(4) ϕ is continuous at ξ and $\phi(\xi) \neq 0$.

Then, we have

$$\int_D \phi(x)[h(x)]^n dx \sim \left(\frac{2\pi}{n}\right)^{\frac{d}{2}} \frac{\phi(\xi)[h(\xi)]^n}{\sqrt{\det(-\nabla^2 \log h(\xi))}}, \quad n \rightarrow \infty.$$

Proof. We refer to Lemma 1 in [18] for details. \square

Proposition 3.1. *Suppose that the law of X_{in} is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d and let f be a probability density function. Suppose that*

- (1) f is compactly supported.
(2) f is continuous at the global minimizer of L , namely X_* , and $f(X_*) > 0$.
(3) L is C^2 and $\det(\nabla^2 L(X_*)) > 0$.

Then, we have

$$-\frac{1}{\beta} \log \mathbb{E} e^{-\beta L(X_{in})} = L(X_*) + \frac{d \log \beta}{2 \beta} + \mathcal{O}\left(\frac{1}{\beta}\right), \quad \beta \rightarrow \infty.$$

Proof. Let D be any closed ball containing the support of f . We set

$$\phi(x) := f(x) \quad \text{and} \quad h(x) := e^{-L(x)}.$$

For the above pair (ϕ, h) , we will check that the assumptions in Lemma 3.3 hold. Note that

$$\begin{aligned} \det(-\nabla^2 h(X_*)) &= \det\left(e^{-L(X_*)}(\nabla^2 L(X_*) - (\nabla L \otimes \nabla L)(X_*))\right) \\ &= e^{-dL(X_*)} \det(\nabla^2 L(X_*)) > 0. \end{aligned}$$

The rest of the assumptions can be checked straightforwardly. By Lemma 3.3, we have

$$(3.7) \quad \lim_{\beta \rightarrow \infty} \frac{\sqrt{\det(\nabla^2 L(X_*))} \mathbb{E} e^{-\beta L(X_{in})}}{\left(\frac{2\pi}{\beta}\right)^{\frac{d}{2}} f(X_*) e^{-\beta L(X_*)}} = 1.$$

By taking logarithms on both sides of (3.7), we get

$$\log \mathbb{E} e^{-\beta L(X_{in})} - \frac{d}{2} \log\left(\frac{2\pi}{\beta}\right) - \log f(X_*) + \beta L(X_*) + \frac{1}{2} \log \det(\nabla^2 L(X_*)) = \mathcal{O}(1).$$

We multiply $-\frac{1}{\beta}$ on both sides to obtain

$$-\frac{1}{\beta} \log \mathbb{E} e^{-\beta L(X_{in})} = L(X_*) + \left(-\frac{d}{2} \log(2\pi) - \log f(X_*) + \frac{1}{2} \log \det(\nabla^2 L(X_*))\right) \frac{1}{\beta} + \frac{d \log \beta}{2 \beta} + \mathcal{O}\left(\frac{1}{\beta}\right).$$

Hence, one has the desired estimate:

$$-\frac{1}{\beta} \log \mathbb{E} e^{-\beta L(X_{in})} = L(X_*) + \frac{d \log \beta}{2 \beta} + \mathcal{O}\left(\frac{1}{\beta}\right),$$

for a sufficiently large β . \square

• **(Proof of Theorem 3.2):** First, we claim:

$$(3.8) \quad \operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) \leq -\frac{1}{\beta} \log \mathbb{E} e^{-\beta L(X_{in})} - \frac{1}{\beta} \log \varepsilon.$$

Note that

$$\begin{aligned}
(3.9) \quad & \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_{n+1}^i)} - \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \\
&= \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} (e^{-\beta(L(X_{n+1}^i) - L(X_n^i))} - 1) \\
&\geq \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} (-\beta) (L(X_{n+1}^i) - L(X_n^i)) \\
&= -\frac{\beta}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \nabla L(cX_{n+1}^i + (1-c)X_n^i) \cdot (X_{n+1}^i - X_n^i),
\end{aligned}$$

where we used a mean-value theorem with $c \in (0, 1)$. We use definition of \bar{X}_n^* to see

$$\left(\sum_{i=1}^N e^{-\beta L(X_n^i)} \right) \bar{X}_n^* = \sum_{i=1}^N e^{-\beta L(X_n^i)} X_n^i, \quad \left(\sum_{i=1}^N e^{-\beta L(X_n^i)} \right) \bar{x}_n^{*,l} = \sum_{i=1}^N e^{-\beta L(X_n^i)} x_n^{i,l}, \quad l = 1, \dots, d.$$

This yields

$$\begin{aligned}
(3.10) \quad & \sum_{i=1}^N e^{-\beta L(X_n^i)} \nabla L(\bar{X}_n^*) \cdot (X_{n+1}^i - X_n^i) \\
&= \sum_{i=1}^N e^{-\beta L(X_n^i)} \nabla L(\bar{X}_n^*) \cdot \left(-\gamma(X_n^i - \bar{X}_n^*) - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) \eta_n^l e_l \right) \\
&= \sum_{i=1}^N e^{-\beta L(X_n^i)} \nabla L(\bar{X}_n^*) \cdot \left(-\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l}) (\gamma + \eta_n^l) e_l \right) = 0.
\end{aligned}$$

Combining (3.9) and (3.10) gives

$$\begin{aligned}
(3.11) \quad & \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_{n+1}^i)} - \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \\
& \geq -\frac{\beta}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \nabla L(cX_{n+1}^i + (1-c)X_n^i) \cdot (X_{n+1}^i - X_n^i) \\
& = -\frac{\beta}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} (\nabla L(cX_{n+1}^i + (1-c)X_n^i) - \nabla L(\bar{X}_n^*)) \cdot (X_{n+1}^i - X_n^i) \\
& = -\frac{\beta}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \\
& \quad \times \left[\int_0^1 \nabla^2 L(s(cX_{n+1}^i + (1-c)X_n^i) + (1-s)\bar{X}_n^*) \cdot (cX_{n+1}^i + (1-c)X_n^i - \bar{X}_n^*) ds \cdot (X_{n+1}^i - X_n^i) \right] \\
& \geq -\frac{\beta}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \left[\int_0^1 \left\| \nabla^2 L(s(cX_{n+1}^i + (1-c)X_n^i) + (1-s)\bar{X}_n^*) \right\|_2 ds \right. \\
& \quad \left. \times |cX_{n+1}^i + (1-c)X_n^i - \bar{X}_n^*| |X_{n+1}^i - X_n^i| \right] \\
& \geq -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N |cX_{n+1}^i + (1-c)X_n^i - \bar{X}_n^*| |X_{n+1}^i - X_n^i| \\
& = -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \left| \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})(1 - c\gamma - c\eta_n^l) e_l \right| - \sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})(\gamma + \eta_n^l) e_l \\
& = -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \sqrt{\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (1 - c\gamma - c\eta_n^l)^2} \sqrt{\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (\gamma + \eta_n^l)^2}.
\end{aligned}$$

Set

$$g(r) := (1 - r\gamma - r\eta_n^l)^2, \quad 0 \leq r \leq 1.$$

Then, we use the convexity of g to find

$$(1 - c\gamma - c\eta_n^l)^2 \leq \sup_{0 \leq r \leq 1} g(r) = \max\{g(0), g(1)\} \leq g(0) + g(1) = 1 + (1 - \gamma - \eta_n^l)^2.$$

We substitute the above relation into (3.11) to obtain

$$\begin{aligned}
(3.12) \quad & \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_{n+1}^i)} - \frac{1}{N} \sum_{i=1}^N e^{-\beta L(X_n^i)} \\
& \geq -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \sqrt{\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (1 + (1 - \gamma - \eta_n^l)^2)} \sqrt{\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (\gamma + \eta_n^l)^2}.
\end{aligned}$$

Taking expectations on the both sides of (3.12) using the Cauchy-Schwarz inequality leads to (3.13)

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N \mathbb{E} e^{-\beta L(X_{n+1}^i)} - \frac{1}{N} \sum_{i=1}^N \mathbb{E} e^{-\beta L(X_n^i)} \\
& \geq -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \mathbb{E} \left[\sqrt{\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (1 + (1 - \gamma - \eta_n^l)^2)} \sqrt{\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (\gamma + \eta_n^l)^2} \right] \\
& \geq -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \sqrt{\mathbb{E} \left[\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (1 + (1 - \gamma - \eta_n^l)^2) \right]} \mathbb{E} \left[\sum_{l=1}^d (x_n^{i,l} - \bar{x}_n^{*,l})^2 (\gamma + \eta_n^l)^2 \right] \\
& = -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \sqrt{\left(\sum_{l=1}^d \mathbb{E} |x_n^{i,l} - \bar{x}_n^{*,l}|^2 \mathbb{E} [1 + (1 - \gamma - \eta_n^l)^2] \right) \left(\sum_{l=1}^d \mathbb{E} |x_n^{i,l} - \bar{x}_n^{*,l}|^2 \mathbb{E} |\gamma + \eta_n^l|^2 \right)} \\
& = -\frac{C_L \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \sqrt{\left((1 + (1 - \gamma)^2 + \zeta^2) \mathbb{E} |X_n^i - \bar{X}_n^*|^2 \right) \left((\gamma^2 + \zeta^2) \mathbb{E} |X_n^i - \bar{X}_n^*|^2 \right)} \\
& = -\frac{C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2) (\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{N} \sum_{i=1}^N \mathbb{E} |X_n^i - \bar{X}_n^*|^2 \\
& \geq -2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2) (\gamma^2 + \zeta^2)} e^{-(2\delta - \gamma^2 - \zeta^2)n} \beta e^{-\beta L_m} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right).
\end{aligned}$$

We sum up (3.13) over n to get

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \mathbb{E} e^{-\beta L(X_n^i)} & \geq \frac{1}{N} \sum_{i=1}^N \mathbb{E} e^{-\beta L(X_0^i)} \\
& - \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2) (\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{1 - e^{-(2\gamma - \gamma^2 - \zeta^2)}} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right).
\end{aligned}$$

Letting $n \rightarrow \infty$, we use Lemma 3.2 to find

$$\begin{aligned}
\mathbb{E} e^{-\beta L(X_\infty)} & \geq \frac{1}{N} \sum_{i=1}^N \mathbb{E} e^{-\beta L(X_0^i)} \\
& - \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2) (\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{1 - e^{-(2\gamma - \gamma^2 - \zeta^2)}} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \\
& \geq \varepsilon \mathbb{E} e^{-\beta L(X_{in})}.
\end{aligned}$$

Hence

$$e^{-\beta \operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty)} = \mathbb{E} e^{-\beta \operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty)} \geq \mathbb{E} e^{-\beta L(X_\infty)} \geq \varepsilon \mathbb{E} e^{-\beta L(X_{in})}.$$

By taking logarithm to the both sides of the above relation yields the desired estimate (3.8).

Next, we apply Proposition 3.1 to obtain

$$\operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) \leq L(X_*) + \frac{d \log \beta}{2 \beta} + E(\beta),$$

for some function $E(\beta) = O\left(\frac{1}{\beta}\right)$. This completes the proof of Theorem 3.2.

4. CONCLUSION

In this paper, we presented a convergence study with an error estimate for a time-discrete consensus-based non-convex optimization algorithm, which includes three discrete-in-time algorithms used in [4]. For the continuous algorithm introduced in [4], convergence and error analysis was studied by the authors recently using the detailed structure of the algorithm and Itô's stochastic analysis. However, for the discrete algorithm such analysis was not done in the aforementioned work [14]. The main reason for that was the lack of discrete analog of the Itô's stochastic analysis which has been crucially used in the study of the continuous algorithm. Our error analysis employed in this work is based on the elementary and simple probability arguments together with the detailed structure of nonlinear interaction terms, and consequently provides the stability condition, convergence and an error analysis toward the global minimum for all three time-discrete algorithms implemented in [4]. Our analysis shows that these algorithms, under certain conditions on the parameters and with sufficiently well-chosen initial data, do converge to a point close to the global minimum for moderately high dimensional problems.

APPENDIX A. ERROR ESTIMATE UNDER AN ALTERNATIVE FRAMEWORK

In this subsection, we provide an alternative result for an error estimate for time-discrete CBO scheme (1.1) without using Laplace's principle under a slightly different framework $(\mathcal{B}1) - (\mathcal{B}2)$. Below, we present a framework for the objective function L , global minimum point X_* and reference random variable X_{in} as follows:

- $(\mathcal{B}1)$: Let $L = L(x)$ be a C^2 -objective function satisfying the following relations:

$$L_m := \min_{x \in \mathbb{R}^d} L(x) > 0 \quad \text{and} \quad C_L := \sup_{x \in \mathbb{R}^d} \|\nabla^2 L(x)\|_2 < \infty,$$

where $\|\cdot\|_2$ denotes the spectral norm.

- $(\mathcal{B}2)$: Let X_{in} be a reference random variable associated with a law whose support \tilde{D} is compact and contains X_* .

Note that the condition $(\mathcal{B}1)$ is exactly the same as $(\mathcal{A}1)$, whereas the condition $(\mathcal{B}2)$ is different from $(\mathcal{A}3)$ in which the probability measures associated with X_{in} is absolutely continuous with respect to Lebesgue measure. Moreover, notice that this new framework does not need any condition on $\det(\nabla^2 L(X_*))$ as in $(\mathcal{A}2)$.

Next, we study how the common consensus state X_∞ is close to the global minimum X_* in suitable sense. Now, we are ready to provide an error estimate for the discrete CBO algorithm, which is analogous to Theorem 4.1 in [14] for the continuous case.

Theorem A.1. *Suppose that the framework $(\mathcal{B}1) - (\mathcal{B}2)$ holds, and parameters $\beta, \gamma, \zeta, \delta$ and the initial data $\{X_0^i\}$ satisfy*

$$(A.1) \quad \begin{aligned} & \beta > 0, \quad \delta > 0, \quad (\gamma - 1)^2 + \zeta^2 < 1, \quad X_0^i : i, i.d., \quad X_0^i \sim X_{in}, \quad \sup_{x \in \tilde{D}} L(x) - L_m < \delta, \\ & (1 - \varepsilon) \mathbb{E} \left[e^{-\beta L(X_{in})} \right] \\ & \geq \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2)(\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{1 - e^{-[1 - (\gamma - 1)^2 - \zeta^2]}} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right), \end{aligned}$$

for some $0 < \varepsilon < 1$. Then, for a solution $\{X_n^i\}_{1 \leq i \leq N}$ to (1.1),

$$(A.2) \quad \left| \operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) - L_m \right| \leq \delta + \left| \frac{\log \varepsilon}{\beta} \right|.$$

Remark A.1. *Before we provide a proof, we give several comments on the result of this theorem.*
1. *In the proof of Theorem A.1, we will first derive the estimate:*

$$(A.3) \quad \operatorname{ess\,inf}_{\omega \in \Omega} L(X_\infty) \leq \sup_{x \in \tilde{D}} L(x) - \frac{1}{\beta} \log \varepsilon.$$

2. *For any given $\delta > 0$, $\varepsilon > 0$ and $\beta > 0$, the conditions (A.1) can be attained with suitable X_{in} . To see this, we choose the law of X_{in} such that \tilde{D} is a small neighborhood of a global minimum X_* satisfying the following two relations:*

$$(A.4) \quad \sup_{x \in \tilde{D}} L(x) - L_m < \delta,$$

and

$$(A.5) \quad 1 - \varepsilon \geq \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2)(\gamma^2 + \zeta^2)} \beta e^{\beta(\sup_{x \in D} L(x) - L_m)}}{1 - e^{-[1 - (\gamma - 1)^2 - \zeta^2]}} (\text{diam}(\mathcal{R}_{in}))^2,$$

where $\mathcal{R}_{in} = [a_1, b_1] \times \cdots \times [a_d, b_d]$ is the smallest closed d -dimensional rectangle containing \tilde{D} so that

$$(A.6) \quad \mathbb{E} \left[\max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right] \leq |b_l - a_l|^2,$$

for each $\ell = 1, \dots, d$. Then, due to (A.6), the relation (A.5) implies the condition (A.1)₂. Hence, we can apply the estimate (A.3) and (A.4) to get the desired error estimate:

$$\text{essinf}_{\omega \in \Omega} L(X_\infty) \leq L_m + \left(\sup_{x \in D} L(x) - L_m \right) - \frac{1}{\beta} \log \varepsilon < L_m + \delta - \frac{\log \varepsilon}{\beta}.$$

Now we are ready to provide a proof of Theorem A.1. By the same way as in the proof of Theorem 3.2, we obtain

$$\begin{aligned} \mathbb{E} e^{-\beta L(X_\infty)} &\geq \frac{1}{N} \sum_{i=1}^N \mathbb{E} e^{-\beta L(X_0^i)} \\ &\quad - \frac{2C_L \sqrt{(1 + (1 - \gamma)^2 + \zeta^2)(\gamma^2 + \zeta^2)} \beta e^{-\beta L_m}}{1 - e^{-(2\gamma - \gamma^2 - \zeta^2)}} \sum_{l=1}^d \left(\mathbb{E} \max_{1 \leq i \leq N} (x_0^{i,l} - \bar{x}_0^l)^2 \right) \\ &\geq \varepsilon \mathbb{E} e^{-\beta L(X_{in})}. \end{aligned}$$

Hence one has

$$e^{-\beta \text{essinf}_{\omega \in \Omega} L(X_\infty)} = \mathbb{E} e^{-\beta \text{essinf}_{\omega \in \Omega} L(X_\infty)} \geq \mathbb{E} e^{-\beta L(X_\infty)} \geq \varepsilon \mathbb{E} e^{-\beta L(X_{in})} \geq \varepsilon e^{-\beta \sup_{x \in D} L(x)}.$$

Finally, we take logarithm to the both sides of the above relation to get the desired estimate (A.3).

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] 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.
- [3] Bertsekas, D.: *Convex analysis and optimization*. Athena Scientific. 2003.
- [4] Carrillo, J. A., Jin, S., Li, L. and Zhu, Y.: *A consensus-based global optimization method for high dimensional machine learning problems*. Submitted.
- [5] 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.
- [6] 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.
- [7] Crow, E. L. and Shimizu, K., eds.: *Lognormal distributions: theory and applications*. Statistics: Textbooks and Monographs, **88**, New York: Marcel-Dekker, Inc., 1988.
- [8] Cucker, F. and Smale, S.: *On the mathematics of emergence*. Japanese Journal of Mathematics **2** (2007), 197-227.
- [9] Dembo, A. and Zeitouni, O.: *Large deviations techniques and applications*. Springer-Verlag, Berlin, Heidelberg, second edition, 1998.

- [10] Eberhart, R. and Kennedy, J.: *Particle swarm optimization*. Proceedings of the IEEE International Conference on Neural Networks **4** (1995), 1942-1948.
- [11] 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.
- [12] Fornasier, M., Huang, H., Pareschi, L. and Sünnen, P.: *Consensus-based optimization on the sphere I: Well-posedness and mean-field limit*. Preprint. Available at arXiv:2001.110994v2.
- [13] Fornasier, M., Huang, H., Pareschi, L. and Sünnen, P.: *Consensus-based optimization on the sphere II: Convergence to global minimizer and machine learning*. Preprint. Available at arXiv:2001.11988v2.
- [14] Ha, S.-Y., Jin, S. and Kim, D.: *Convergence of a first-order consensus-based global optimization algorithm*. Submitted. arxiv: 1910.08239.
- [15] 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.
- [16] 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.
- [17] Holland, J. H.: *Genetic algorithms*. Scientific American **267** (1992), 66-73.
- [18] Hsu, L. C.: *A theorem on the asymptotic behavior of a multiple integral*. Duke Math. J. **15** (1948), 623-632.
- [19] Kennedy, J.: *Swarm intelligence*. Handbook of nature-inspired and innovative computing. Springer 2006, 187-219.
- [20] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P.: *Optimization by simulated annealing*. Science **220** (1983), 671-680.
- [21] 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.
- [22] Kuramoto, Y.: *Chemical oscillations, waves and turbulence*. Springer-Verlag, Berlin, 1984.
- [23] Kuramoto, Y.: *International symposium on mathematical problems in mathematical physics*. Lecture Notes Theor. Phys. **30** (1975), 420.
- [24] Laarhoven, P. J. M. van and Aarts, E. H. L.: *Simulated annealing: theory and applications*. D. Reidel Publishing Co., Dordrecht, 1987.
- [25] Motsch, S. and Tadmor, E.: *Heterophilious dynamics enhances consensus*. SIAM. Rev. **56** (2014), 577-621.
- [26] Peskin, C. S.: *Mathematical aspects of heart physiology*. Courant Institute of Mathematical Sciences, New York, 1975.
- [27] 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.
- [28] Pikovsky, A., Rosenblum, M. and Kurths, J.: *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press, Cambridge, 2001.
- [29] 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.
- [30] Vicsek, T. and Zefeiris, A.: *Collective motion*. Phys. Rep. **517** (2012), 71-140.
- [31] Yang, X.-S.: *Nature-inspired metaheuristic algorithms*. Luniver Press, 2010.
- [32] 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)
Email 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
Email address: shijin-m@sjtu.edu.cn

(Doheon Kim)

SCHOOL OF MATHEMATICS, KOREA INSTITUTE FOR ADVANCED STUDY,
 HOEGIRO 85, SEOUL 02455, KOREA (REPUBLIC OF)
Email address: doheonkim@kias.re.kr