

Exponential Stability of Finite- N Consensus-Based Optimization

S. Göttlich*, J. Heieck*, A. Neuenkirch*

October 22, 2025

Abstract

We study the finite-agent behavior of Consensus-Based Optimization (CBO), a recent metaheuristic for the global minimization of a function, that combines drift toward a consensus estimate with stochastic exploration. While previous analyses focus on asymptotic mean-field limits, we investigate the stability properties of CBO for finite population size N . Following a hierarchical approach, we first analyze a deterministic formulation of the algorithm and then extend our results to the fully stochastic setting governed by a system of stochastic differential equations. Our analysis reveals that essential stability properties, including almost sure and mean square exponential convergence, persist in both regimes and provides sharp quantitative estimates on the rates of convergence.

AMS Classification. 34D35, 37H30, 90C26, 93D05

Keywords: stability analysis, Lyapunov function, multi-agent system, consensus-based optimization, derivative-free optimization

1 Introduction

1.1 Outline

A central problem in applied mathematics is the global minimization of a potentially non-convex and non-smooth function $f : \mathbb{R}^D \rightarrow \mathbb{R}$, along with the search for its global minimizer $\mathbf{X}^* \in \mathbb{R}^D$. Computing f or \mathbf{X}^* is generally NP-hard, but practical instances often permit approximate solutions within reasonable accuracy and time.

Metaheuristics [3] have proven effective in tackling such optimization problems, orchestrating local improvements and global strategies through a mix of random and deterministic decisions. Prominent examples include Random Search [30], Particle Swarm Optimization (PSO) [21], Wind-Driven Optimization [2] and Ant Colony Optimization [11]. Despite their empirical success, many metaheuristics lack rigorous mathematical guarantees for global convergence. Nevertheless, for some of them, such as Random Search, there exist probabilistic guarantees for global convergence [32].

Recently, Consensus-Based Optimization (CBO) [29, 7] has emerged as a promising alternative, following the principles of metaheuristics while offering greater mathematical tractability. CBO models a finite set $X_t^1, \dots, X_t^N \in \mathbb{R}^D$ of $N \in \mathbb{N}$ agents as stochastic processes in time $t \geq 0$ exploring the domain and forming a consensus. Their dynamics are driven by two competing forces: a drift term, which pulls agents toward a weighted consensus estimate $\nu_f^\alpha(X_t)$ of \mathbf{X}^* given by

$$\nu_f^\alpha(X_t) = \sum_{m=1}^N X_t^m \frac{\exp(-\alpha f(X_t^m))}{\sum_{k=1}^N \exp(-\alpha f(X_t^k))} \quad (1.1)$$

with $\alpha > 0$, contracting their convex hull; and a diffusion term, modeled as scaled Brownian motion in \mathbb{R}^D , enhancing exploration. Mathematically, the evolution of each agent X_t^n is governed by the stochastic differential equation (SDE)

$$dX_t^n = -\lambda (X_t^n - \nu_f^\alpha(X_t)) dt + \sigma \|X_t^n - \nu_f^\alpha(X_t)\| dW_t^n, \quad n, \dots, N, \quad (1.2)$$

¹University of Mannheim, Department of Mathematics, 68131 Mannheim, Germany (goettlich@math.uni-mannheim.de, jacob.heieck@students.uni-mannheim.de, neuenkirch@uni-mannheim.de).

where $\lambda > 0$ controls the strength of the drift towards the weighted consensus point $\nu_f^\alpha(X_t)$, and $\sigma > 0$ scales the noise term modeled by D -dimensional independent Brownian motions W_t^n . This formulation ensures a balance between exploitation of promising regions and exploration of the search space, ultimately leading agents to a near-optimal global consensus.

CBO has been applied in several different problem settings. Examples include constrained optimization, such as model predictive control [6, 4] and optimization on manifolds like the sphere [14], multi-objective optimization [5], as well as training neural networks [8] and 3D human pose estimation in camera coordinates [23]. These examples show that CBO can be used to address a broad class of non-convex optimization problems.

The theoretical analysis of CBO has seen significant progress, particularly in the context of its mean-field limit. In [29], the authors establish that as the number of agents N tends to infinity and α approaches infinity (Laplacian principle [10]), the system converges arbitrarily close to the minimum. Furthermore, in [13], rigorous results are provided to demonstrate that under suitable assumptions, the mean-field formulation of CBO converges globally to the minimizer \mathbf{X}^* . These results offer deep insights into the large-population behavior of CBO but rely on asymptotic limits. Apart from the well-posedness of the finite-agent system, which has been established in [7], little is known about the theoretical properties of CBO for finite N and finite α . Some progress in this direction has been made for simplified variants of CBO where all agents are influenced by the same random fluctuations, effectively removing the individual stochasticity. While these works provide valuable insights into the convergence properties and error bounds [15, 16] of these simplified algorithms, they differ structurally from the classical CBO dynamics we study here, which model independent agent behavior more faithfully.

In this work, we take a next step towards further bridging this gap by analyzing the stability properties of the classical CBO for finite N and finite α . Our approach follows a hierarchical methodology, motivated by the existing work on deterministic particle-based systems. In the linear setting, convergence to consensus has been successfully shown for multi-agent systems [36], and strict Lyapunov functions have been established for consensus dynamics under directed connected graphs [28, 37]. These studies leverage a graph-theoretic interpretation of particle systems, wherein agents are represented as nodes and their interactions as weighted edges.

Building upon these insights, we adopt the following structured approach:

- First, in section 2 we study the deterministic system

$$dX_t^n = -\lambda (X_t^n - \nu_f^\alpha(X_t)) dt, \quad n = 1, \dots, N, \quad (1.3)$$

leveraging existing results from linear graph-theoretic consensus models to establish global exponential stability of the component-wise non-linear system (1.3) and its explicit Euler discretization for $N < \infty$ and $\alpha < \infty$.

- Finally, in section 3 we investigate the full system governed by a SDE

$$dZ_t^n = -\lambda (Z_t^n - \nu_f^\alpha(Z_t)) dt + \sigma (Z_t^n - \nu_f^\alpha(Z_t)) \circ dW_t^n, \quad n = 1, \dots, N, \quad (1.4)$$

which incorporates both drift and diffusion effects. Here, \circ denotes the element-wise (Hadamard) product [20, Definition 7.5.1]. We rigorously establish both almost sure and mean square exponential stability of the system (1.4) and its Euler-Maruyama discretization under suitable conditions on the parameters. In our analysis, we focus on the anisotropic diffusion term $\sigma(Z_t^n - \nu_f^\alpha(Z_t))$ [12], rather than the more general formulation $\sigma\|Z_t^n - \nu_f^\alpha(Z_t)\|$, where $\|\cdot\|$ typically denotes the Euclidean norm $\|\cdot\|_2$ [7, 13]. We restrict ourselves to the diffusion term $\sigma(Z_t^n - \nu_f^\alpha(Z_t))$ since the graphical interpretation of the system is sufficient to reflect its stability properties.

By systematically progressing from the deterministic case to the full stochastic system, we aim to gain a deeper understanding of CBO dynamics beyond the mean-field limit, providing new theoretical insights into its behavior for practical, finite-agent settings.

1.2 Notation

To establish a consistent notation for our analysis, we introduce the following definitions. We denote D as the dimension, and the collective state of all particles at time t by the vector $X_t = (X_t^1, \dots, X_t^N)^\top \in \mathbb{R}^{ND}$, where each particle X_t^n is represented as $X_t^n = (x_t^{n,1}, \dots, x_t^{n,D})^\top \in \mathbb{R}^D$, $n = 1, \dots, N$.

We use I_N to denote the $(N \times N)$ -identity matrix. Furthermore, we define $\mathbf{0}_N := (0, \dots, 0)^\top \in \mathbb{R}^N$ and $\mathbf{1}_N := (1, \dots, 1)^\top \in \mathbb{R}^N$ as the all-zero and all-one column vectors of size N , respectively. Unless stated otherwise, we denote by $\|\cdot\|$ the standard Euclidean norm.

2 Purely Deterministic System

We begin by examining the purely deterministic system (1.3). In this setting, the exploitative diffusion is neglected, focusing solely on drift forces that drive particles towards a consensus state. Before we will establish our stability results, we precisely define what we mean by stability in this context.

Intuitively, stability describes the persistence of solutions under small perturbations: if a system is slightly disturbed, it shall return to its original state. More formally, stability is often characterized through equilibrium points.

Definition 2.1 (Equilibrium Point [31, Definiton 3.2]). The state $X^* \in \mathbb{R}^N$ is said to be an *equilibrium point* of an ordinary differential equation (ODE)

$$dX_t = a(X_t)dt, \quad t \geq 0, \quad (2.1)$$

if once X_t is equal to X^* , it remains equal to X^* for all future times, i.e., X^* has to satisfy $a(X^*) = \mathbf{0}_N$.

In particular, we distinguish between stability and exponential stability.

Definition 2.2 (Stability for ODE [31, Definiton 3.3 and 3.6]). The equilibrium point $X^* = \mathbf{0}_N$ is said to be *stable* if, for any $R > 0$, there exists $r > 0$, such that if the initial state satisfies $\|X_0\| < r$, then $\|X_t\| < R$ for all $t \geq 0$. Otherwise, the equilibrium point is said to be *unstable*. If stability holds for any initial state X_0 , the equilibrium point is said to be *globally stable*.

Exponential stability strengthens the notion of stability.

Definition 2.3 (Exponential Stability for ODE [31, Definition 3.5 and 3.6]). The equilibrium point $X^* = \mathbf{0}_N$ is *exponentially stable* if there exist constants $c > 0$ and $\delta > 0$ and some $r > 0$ such that $\|X_0\| < r$ implies

$$\|X_t\| \leq c \|X_0\| e^{-\delta t}, \quad t \geq 0.$$

If the inequality holds for all $X_0 \in \mathbb{R}^N$, the equilibrium point is called *globally exponentially stable*.

Exponential stability not only ensures that solutions converge to the equilibrium, but also quantifies how fast this convergence occurs. More precisely, the distance to the equilibrium decays at least as fast as $e^{-\delta t}$ for some rate $\delta > 0$.

Notice, that both definitions apply specifically to the equilibrium point $X^* = \mathbf{0}_N$. The stability of other equilibrium points or entire sets of points can be analyzed through a suitable transformation of the state space.

To assess the stability properties of equilibrium points, one may employ a fundamental mathematical tool: Lyapunov functions. These functions provide a systematic approach to determining whether solutions remain close to an equilibrium or converge to it over time.

Definition 2.4 (Lyapunov functions for ODE [31, Definiton 3.8]). Let $V : \mathbb{R}^N \rightarrow \mathbb{R}$ be a function,

- (1) which is positive definite, i.e., we have $V(x) = 0$ for $x = \mathbf{0}_N$ and $V(x) > 0$ else,
- (2) which has continuous partial derivatives and
- (3) whose time derivative along any state trajectory of the system $dX_t = a(X_t)dt$ is negative semi-definite, i.e.,

$$\forall t \geq 0 : \quad \dot{V}(X_t) := \frac{dV(X_t)}{dt} = (\nabla V(X_t))^\top \frac{dX_t}{dt} = (\nabla V(X_t))^\top a(X_t) \leq 0.$$

Then V is said to be a *Lyapunov function* for the system (2.1).

The reformulation in Definition 2.4 follows directly from the chain rule. Here, we exploit the fact that V depends solely on the state X_t and not explicitly on time t .

The core idea is to construct a scalar function V that decreases along system trajectories. If such a function exists, it serves as a certificate for stability.

Theorem 2.5 (Stability for ODEs [31, Theorem 3.2]). Consider the system (2.1) and assume that $X^* = \mathbf{0}_N$ is an equilibrium point. If there exists a Lyapunov function V for this system, then $X^* = \mathbf{0}_N$ is globally stable.

If V has additional properties, we can obtain exponential stability.

Theorem 2.6 (Exponential Stability for ODE [27, Theorem 4.5]). Consider the system (2.1) and assume that $X^* = \mathbf{0}_N$ is an equilibrium point. Assume that $V : \mathbb{R}^N \rightarrow \mathbb{R}$ is Lyapunov function (for this system), which has the following additional property: there exist constants $c_1, c_2, c_3, c_4 > 0$ such that

- (i) $c_1 \|x\|^2 \leq V(x) \leq c_2 \|x\|^2$ for $x \in \mathbb{R}^N$,
- (ii) $\dot{V}(X_t) \leq -c_3 \|X_t\|^2$ for all $t \geq 0$,
- (iii) $\|\nabla V(x)\| \leq c_4 \|x\|$ for $x \in \mathbb{R}^N$.

Then the equilibrium point $X^* = \mathbf{0}_N$ is globally exponentially stable with $c = \sqrt{\frac{c_2}{c_1}}$ and $\delta = \frac{c_3}{2c_2}$.

Lyapunov functions are a classical concept for deriving (exponential) stability of ODEs, and we apply it to our dynamical system (1.3). During the course of our analysis it will turn out that an appropriate transformation of the state space strikingly clarifies the stability properties of the dynamics of (1.3).

2.1 System Description

The purely deterministic system is given by

$$dX_t^n = -\lambda (X_t^n - \nu_f^\alpha(X_t)) dt \quad (2.2)$$

for $n = 1, \dots, N$. By definition of the consensus point $\nu_f^\alpha(X_t)$ in (1.1) we have

$$dX_t^n = -\lambda \left(X_t^n - \sum_{m=1}^N a_m(X_t) X_t^m \right) dt$$

for $n = 1, \dots, N$ with weights

$$a_m(X_t) = \frac{\exp(-\alpha f(X_t^m))}{\sum_{k=1}^N \exp(-\alpha f(X_t^k))}.$$

Note that $a_m = a_m(X_t)$ depends non-linearly on the particle vector X_t . In the following, we will often drop this dependence for notational easiness.

Since

$$\sum_{m=1}^N a_m = \sum_{m=1}^N \frac{\exp(-\alpha f(X_t^m))}{\sum_{k=1}^N \exp(-\alpha f(X_t^k))} = 1,$$

all weights sum up to 1 and we can rewrite

$$dX_t^n = -\lambda \left(\sum_{\substack{m=1 \\ m \neq n}}^N a_m \cdot (X_t^n - X_t^m) \right) dt.$$

Considering all N equations simultaneously and sorting them by spatial dimension d yields

$$dX_t = d \begin{pmatrix} x_t^{1,1} \\ \vdots \\ x_t^{N,1} \\ \vdots \\ x_t^{1,D} \\ \vdots \\ x_t^{N,D} \end{pmatrix} = -\lambda \begin{pmatrix} \sum_{m=2}^N a_m (x_t^{1,1} - x_t^{m,1}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (x_t^{N,1} - x_t^{m,1}) \\ \vdots \\ \sum_{m=2}^N a_m (x_t^{1,D} - x_t^{m,D}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (x_t^{N,D} - x_t^{m,D}) \end{pmatrix} dt. \quad (2.3)$$

Given the weights a_1, \dots, a_N , the evolution of each dimension is therefore independent of the others, as it depends only on the interactions between particles within the same dimension. Moreover, the weights a_m are equal in all equations.

Thus, it will suffice to study the stability of the system in each dimension $d \in \{1, \dots, D\}$ on its own. If we consider the system

$$d \begin{pmatrix} x_t^{1,d} \\ \vdots \\ x_t^{N,d} \end{pmatrix} = -\lambda \begin{pmatrix} \sum_{m=2}^N a_m (x_t^{1,d} - x_t^{m,d}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (x_t^{N,d} - x_t^{m,d}) \end{pmatrix} dt \quad (2.4)$$

and introduce the notation $X_t^{\cdot,d} = (x_t^{1,d}, \dots, x_t^{N,d})^\top \in \mathbb{R}^N$, the right-hand side of the corresponding system simplifies to

$$-\lambda \hat{L}(X_t) X_t^{\cdot,d} dt,$$

where $\hat{L}(X_t)$ is defined as follows

$$\hat{L}(X_t) := \begin{pmatrix} 1 - a_1(X_t) & -a_2(X_t) & \cdots & -a_N(X_t) \\ -a_1(X_t) & 1 - a_2(X_t) & \cdots & -a_N(X_t) \\ \vdots & \vdots & \ddots & \vdots \\ -a_1(X_t) & -a_2(X_t) & \cdots & 1 - a_N(X_t) \end{pmatrix}. \quad (2.5)$$

The diagonal entries arise since the weights sum up to 1. This leads to the compact representation

$$dX_t = -\lambda L(X_t) X_t dt$$

of system (2.3) with

$$L(X_t) = \begin{pmatrix} \hat{L}(X_t) & \mathbf{0}_{N \times N} & \cdots & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \hat{L}(X_t) & \cdots & \mathbf{0}_{N \times N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} & \cdots & \hat{L}(X_t) \end{pmatrix} \in \mathbb{R}^{DN \times DN}. \quad (2.6)$$

Here, $\mathbf{0}_{N \times N}$ denotes the zero-matrix of dimension $N \times N$. This formulation emphasizes that each dimension d can be treated separately, while still preserving the component-wise non-linear interactions between particles through the weights a_1, \dots, a_N .

From a graph-theoretic perspective, the matrix $\hat{L}(X_t)$ can be seen as the Laplacian [28] of a weighted complete graph, where the nodes represent the particles $x_t^{n,d}$ and the edges carry the interaction weights $-a_m(X_t)$ or $1 - a_n(X_t)$. Each off-diagonal $-a_m(X_t)$ entry describes how strongly particle n is influenced by particle m , while the diagonal terms ensure that the total influence on each particle is balanced. The motion of particle n is exactly a weighted combination of the positions of all other particles, with higher weights meaning that n is pulled more strongly toward those particular particles.

This framework links to the classical theory of multi-agent consensus, where stability has been fully characterized in the linear case [36, 28, 37]. In contrast, the CBO system involves highly nonlinear weights $a_m(X_t)$, which require a refined analysis. The next section develops stability results for this nonlinear setting. In the following we will drop the dependence of $\hat{L} = \hat{L}(X_t)$ and $L = L(X_t)$ on X_t for notational simplicity.

2.2 Global Exponential Stability

In this chapter, we establish the (global exponential) stability of the deterministic system (2.2). As a first step, we determine the equilibrium points of said system. These correspond to solutions of the equation

$$-\lambda (X_t^n - \nu_f^\alpha(X_t)) = \mathbf{0}_N \quad \Leftrightarrow \quad X_t^n = \nu_f^\alpha(X_t)$$

for $n = 1, \dots, N$. Since the right-hand side depends on the entire particle configuration rather than on individual indices n , a necessary condition for an equilibrium is

$$X_t^1 = \dots = X_t^N. \quad (2.7)$$

To verify whether this condition is also sufficient, we assume that all particles take the same value, i.e., equation (2.7) is satisfied. Given that $\nu_f^\alpha(X_t)$ represents a weighted consensus point, we compute for any $n \in \{1, \dots, N\}$:

$$X_t^n = \sum_{m=1}^N \frac{X_t^m}{N} = \sum_{m=1}^N \frac{X_t^m \exp(-\alpha f(X_t^m))}{\sum_{k=1}^N \exp(-\alpha f(X_t^k))} = \sum_{m=1}^N \frac{X_t^m \exp(-\alpha f(X_t^m))}{\sum_{k=1}^N \exp(-\alpha f(X_t^k))} = \nu_f^\alpha(X_t). \quad (2.8)$$

This confirms that the equilibrium points are uniquely characterized by the condition $X_t^1 = \dots = X_t^N$. Thus, all particles must reach a common consensus value at equilibrium, which belongs to the set

$$\mathcal{G}^* = \{x \in \mathbb{R}^{N \times D} : x^{1,d} = \dots = x^{N,d}, d = 1, \dots, D\}. \quad (2.9)$$

Having identified the equilibrium points, we now turn our attention to their stability properties. To this end, we focus on the matrix L , which governs the dynamics of the system $dX_t = -\lambda L X_t dt$. As previously indicated each $d \in \{1, \dots, D\}$ will be treated independently. We will therefore analyze the stability of the subsystem

$$dX_t^{:,d} = -\lambda \hat{L} X_t^{:,d} dt \quad (2.10)$$

where the matrix \hat{L} is given by equation (2.5).

Our first objective is to understand the spectral properties of \hat{L} , as these determine the qualitative behavior of the system. More precisely, we will study

$$\dot{V}(X_t^{:,d}) = (\nabla V(X_t^{:,d}))^\top a(X_t^{:,d}) = (\nabla V(X_t^{:,d}))^\top (-\lambda \hat{L} X_t^{:,d}) = -\lambda (\nabla V(X_t^{:,d}))^\top \hat{L} X_t^{:,d} \quad (2.11)$$

for a suitable function $V : \mathbb{R}^d \rightarrow [0, \infty)$. For the inequality $\dot{V}(X_t^{:,d}) \leq 0$ to hold, we need to understand the eigenstructure of \hat{L} , which reveals in which directions particles contract and in which directions motion persists or stagnates.

In order to compute the eigenvalues of \hat{L} explicitly, we make use of the following result on rank-one matrix updates, which allows for efficient determinant calculations.

Lemma 2.7 (Rank-One Update [26, p. 475]). If $A \in \mathbb{R}^{N \times N}$ is non-singular, and if c, d are N -dimensional column vectors, then:

- (i) $\det(I_N + cd^\top) = 1 + d^\top c$,
- (ii) $\det(A + cd^\top) = \det(A) (1 + d^\top A^{-1}c)$.

We now apply this lemma to derive the eigenvalues of \hat{L} . Recall that $\mathbf{1}_N$ denotes the all-ones column vector, and define the weight vector

$$\ell := (a_1, \dots, a_N)^\top \in (0, 1)^N, \quad (2.12)$$

which collects the weights associated with each particle X_t^n , $n = 1, \dots, N$.

Lemma 2.8 (Eigenvalues of \hat{L}). The matrix $\hat{L} \in \mathbb{R}^{N \times N}$ has exactly one eigenvalue equal to zero, while the remaining $N - 1$ eigenvalues are equal to 1, that is,

$$\mu_1(\hat{L}) = 0, \quad \mu_2(\hat{L}) = \dots = \mu_N(\hat{L}) = 1.$$

A right-eigenvector corresponding to $\mu_1(\hat{L})$ is $\mathbf{1}_N$, a left-eigenvector is ℓ .

Proof. We begin by rewriting \hat{L} using a rank-one update involving $\mathbf{1}_N$ and $\ell = (a_1, \dots, a_N)^\top$:

$$\hat{L} = \begin{pmatrix} 1 - a_1 & -a_2 & \cdots & -a_N \\ -a_1 & 1 - a_2 & \cdots & -a_N \\ \vdots & \vdots & \ddots & \vdots \\ -a_1 & -a_2 & \cdots & 1 - a_N \end{pmatrix} = I_N - \begin{pmatrix} a_1 & a_2 & \cdots & a_N \\ a_1 & a_2 & \cdots & a_N \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_N \end{pmatrix} = I_N - \mathbf{1}_N \ell^\top.$$

To determine the eigenvalues, we compute the characteristic equation by evaluating $\det(\hat{L} - \mu I_N) = 0$. Using the rank-one update representation of \hat{L} and applying (ii) of Lemma 2.7 with $A = (1 - \mu)I_N$, $c = -\mathbf{1}_N$ and $d = \ell$, we obtain

$$\det(\hat{L} - \mu I_N) = \det((1 - \mu)I_N) (1 - \ell^\top ((1 - \mu)I_N)^{-1} \mathbf{1}_N) = (1 - \mu)^N \left(1 - \frac{1}{1 - \mu} \ell^\top \mathbf{1}_N\right).$$

Here, we use $A^{-1} = \frac{1}{1-\mu} I_N$ and $\det(A) = (1-\mu)^N$. Since all weights sum to one, we have

$$\ell^\top \mathbf{1}_N = \sum_{n=1}^N a_n = 1.$$

Substituting this into the equation gives

$$\det(\hat{L} - \mu I_N) = (1-\mu)^N \left(1 - \frac{1}{1-\mu}\right).$$

As a result, the determinant satisfies $\det(\hat{L} - \mu I_N) = 0$ if either

$$(1-\mu)^N = 0 \quad \text{or} \quad 1 - \frac{1}{1-\mu} = 0.$$

The second condition simplifies to $1-\mu = 1$, which holds for $\mu = 0$. The first condition implies $\mu = 1$. Thus, the eigenvalues of \hat{L} are

$$\mu_1(\hat{L}) = 0, \quad \mu_2(\hat{L}) = \dots = \mu_N(\hat{L}) = 1.$$

For the corresponding eigenvectors to $\mu_1(\hat{L}) = 0$ we compute

$$\hat{L} \mathbf{1}_N = (I_N - \mathbf{1}_N \ell^\top) \mathbf{1}_N = \mathbf{1}_N - \mathbf{1}_N \ell^\top \mathbf{1}_N = \mathbf{1}_N - \mathbf{1}_N = \mathbf{0}_N = 0 \cdot \mathbf{1}_N = \mu_1(\hat{L}) \cdot \mathbf{1}_N$$

for the right-eigenvector and

$$\hat{L}^\top \ell = (I_N - \mathbf{1}_N \ell^\top)^\top \ell = \ell - \ell \mathbf{1}_N^\top \ell = \ell - \ell = \mathbf{0}_N = 0 \cdot \ell = \mu_1(\hat{L}) \cdot \ell$$

for the left-eigenvector. Thus, we conclude that both $\mathbf{1}_N$ and ℓ are right- and left-eigenvectors corresponding to the eigenvalue $\mu_1(\hat{L}) = 0$. The key property used in this derivation is that the sum of all weights equals one. \square

Lemma 2.8 reveals that the matrix \hat{L} has a highly structured spectrum: it has exactly one eigenvalue equal to zero, while all remaining $N-1$ eigenvalues are equal to one. The right-eigenvector associated with the zero eigenvalue is $\mathbf{1}_N$, the all-ones vector.

By restricting condition (2.7) to a fixed dimension d , we find that all equilibrium points of the deterministic system $dX_t^{:,d} = -\lambda \hat{L} X_t^{:,d} dt$ satisfy

$$x_t^{1,d} = \dots = x_t^{N,d}, \tag{2.13}$$

i.e., all particle components agree. Thus, any equilibrium point lies in the one-dimensional subspace spanned by the eigenvector $\mathbf{1}_N$. This spectral structure admits a dynamical interpretation: the eigenspace associated with the zero eigenvalue corresponds to a direction in which the system does not evolve. Once all particles are aligned, the dynamics stops changing. In contrast, all directions orthogonal to $\mathbf{1}_N$ correspond to the eigenvalue one and are actively contracted by the dynamics. As a result, the system is continuously pushed toward the subspace

$$\mathcal{G}^{*,d} = \{x \in \mathbb{R}^N : x = \gamma \mathbf{1}_N, \gamma \in \mathbb{R}\}$$

of equilibrium points.

Motivated by these observations, we consider the function

$$V(X_t^{:,d}) = \left\| X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N \right\|^2, \tag{2.14}$$

where, $\bar{X}_t^{:,d}$ denotes the average particle value in dimension d , defined as

$$\bar{X}_t^{:,d} := \frac{1}{N} \mathbf{1}_N^\top X_t^{:,d} = \frac{1}{N} \sum_{n=1}^N x_t^{n,d}. \tag{2.15}$$

Since $\bar{X}_t^{:,d}$ is actually the orthogonal projection of $X_t^{:,d}$ onto $\mathcal{G}^{*,d}$, we also have

$$V(X_t^{:,d}) = \text{dist}(X_t^{:,d}; \mathcal{G}^{*,d})^2 := \inf_{y \in \mathcal{G}^{*,d}} \left\| X_t^{:,d} - y \right\|^2.$$

Thus, V measures the squared distance of the particle system to its closest equilibrium point. Clearly, the difference $X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N \in \mathbb{R}^N$ captures how far each individual particle deviates from the mean, so an alternative interpretation of V is that this function measures the squared distance of $X_t^{:,d}$ to its mean value. However, note that V is not a classical Lyapunov function, since that we have $V(x) = 0$ for all $x \in \mathcal{G}^{*,d}$.

Now define the auxiliary functions

$$g : \mathbb{R}^N \rightarrow \mathbb{R}, \quad g(x) = x^\top x, \quad h : \mathbb{R}^N \rightarrow \mathbb{R}^N, \quad h(x) = x - \frac{1}{N} \mathbf{1}_N^\top x \mathbf{1}_N,$$

which satisfy

$$\nabla g(x) = 2x, \quad Jh(x) = I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top$$

for $x \in \mathbb{R}^N$. The expression for the Jacobi-matrix of h follows from the observation that

$$\partial_{x_m}(h(x))_k = \partial_{x_m} \left(x_k - \frac{1}{N} \mathbf{1}_N^\top x \right) = \mathbf{1}_{\{m=k\}} - \frac{1}{N}$$

for $m, k = 1, \dots, N$. Here, $\mathbf{1}_{\{m=k\}}$ equals 1 when $m = k$ and 0 otherwise.

Thus, we have

$$V(x) = g(h(x))$$

and so the chain rule yields

$$\begin{aligned} \nabla V(x) &= \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right)^\top 2 \left(x - \frac{1}{N} \mathbf{1}_N^\top x \mathbf{1}_N \right) \\ &= 2 \left(x - \frac{1}{N} \mathbf{1}_N^\top x \mathbf{1}_N \right) - 2 \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top x + 2 \frac{1}{N^2} \mathbf{1}_N \mathbf{1}_N^\top x \mathbf{1}_N^\top \mathbf{1}_N \\ &= 2 \left(x - \frac{1}{N} \mathbf{1}_N^\top x \mathbf{1}_N \right). \end{aligned}$$

Here we have used that $x^\top \mathbf{1}_N = \mathbf{1}_N^\top x$ is a scalar and $\mathbf{1}_N^\top \mathbf{1}_N = N$.

Using this expression for ∇V in (2.11) we obtain

$$\dot{V}(X_t^{:,d}) = -2\lambda(X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N)^\top \hat{L} X_t^{:,d} \quad (2.16)$$

Now, we will first rewrite and simplify the expression on the right-hand side of the above equality. To this end, we use the representation $\hat{L} = I_N - \mathbf{1}_N \ell^\top$ established in the proof of Lemma 2.8. Moreover, we have $\bar{X}_t^{:,d} = \frac{1}{N} \mathbf{1}_N^\top X_t^{:,d}$, which yields

$$\hat{L}(\bar{X}_t^{:,d} \mathbf{1}_N) = (I_N - \mathbf{1}_N \ell^\top) \left(\frac{1}{N} \mathbf{1}_N^\top X_t^{:,d} \mathbf{1}_N \right) = \frac{1}{N} \mathbf{1}_N^\top X_t^{:,d} \mathbf{1}_N - \frac{1}{N} \mathbf{1}_N \ell^\top \mathbf{1}_N^\top X_t^{:,d} \mathbf{1}_N.$$

Since $\mathbf{1}_N^\top X_t^{:,d}$ is a scalar, we can rearrange terms and obtain

$$\hat{L}(\bar{X}_t^{:,d} \mathbf{1}_N) = \frac{1}{N} \mathbf{1}_N^\top X_t^{:,d} \mathbf{1}_N - \frac{1}{N} \mathbf{1}_N^\top X_t^{:,d} \mathbf{1}_N \cdot \ell^\top \mathbf{1}_N = 0, \quad (2.17)$$

where we used that the weights sum to one,

$$\ell^\top \mathbf{1}_N = \sum_{n=1}^N a_n = 1.$$

As a result, we can rewrite the total derivative as

$$\dot{V}(X_t^{:,d}) = -2\lambda(X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N)^\top \hat{L} X_t^{:,d} = -2\lambda(X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N)^\top \hat{L}(X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N),$$

since the additional term vanishes due to (2.17). For brevity, we define the error vector $E_t^{:,d} := X_t^{:,d} - \bar{X}_t^{:,d} \mathbf{1}_N$ and compute

$$\begin{aligned} \dot{V}(X_t^{:,d}) &= -2\lambda(E_t^{:,d})^\top \hat{L} E_t^{:,d} \\ &= -2\lambda(E_t^{:,d})^\top (I_N - \mathbf{1}_N \ell^\top) E_t^{:,d} \\ &= -2\lambda(E_t^{:,d})^\top E_t^{:,d} + 2\lambda(E_t^{:,d})^\top \mathbf{1}_N \ell^\top E_t^{:,d}. \end{aligned} \quad (2.18)$$

To simplify the second term, we substitute back $E_t^{\cdot,d} = X_t^{\cdot,d} - \bar{X}_t^{\cdot,d} \mathbf{1}_N$, use that $(X_t^{\cdot,d})^\top \mathbf{1}_N$ is a scalar and $\mathbf{1}_N^\top \mathbf{1}_N = N$:

$$\begin{aligned} (E_t^{\cdot,d})^\top \mathbf{1}_N \ell^\top &= \left(X_t^{\cdot,d} - \frac{1}{N} \mathbf{1}_N^\top X_t^{\cdot,d} \mathbf{1}_N \right)^\top \mathbf{1}_N \ell^\top \\ &= (X_t^{\cdot,d})^\top \mathbf{1}_N \ell^\top - \frac{1}{N} \mathbf{1}_N^\top (X_t^{\cdot,d})^\top \mathbf{1}_N \mathbf{1}_N \ell^\top \\ &= (X_t^{\cdot,d})^\top \mathbf{1}_N \ell^\top - \frac{N}{N} (X_t^{\cdot,d})^\top \mathbf{1}_N \ell^\top \\ &= \mathbf{0}_N^\top. \end{aligned}$$

Hence, the second term in (2.18) vanishes, and observing that $V(X_t^{\cdot,d}) = \|E_t^{\cdot,d}\|^2$ we obtain

$$\dot{V}(X_t^{\cdot,d}) = -2\lambda (E_t^{\cdot,d})^\top E_t^{\cdot,d} = -2\lambda \|E_t^{\cdot,d}\|^2 = -2\lambda V(X_t^{\cdot,d}).$$

Solving this linear ODE yields

$$V(X_t^{\cdot,d}) = e^{-2\lambda t} V(X_0^{\cdot,d}), \quad t \geq 0, \quad (2.19)$$

so this distance of the particle system to its subspace of equilibrium points decays exponentially with rate λ .

At this point, we require a change of coordinates, if we want to work in the setup of Definitions 2.2 and 2.3, since we have a whole subspace $\mathcal{G}^{\star,d}$ of equilibrium point instead of $x^{\star,d} = \mathbf{0}_N$. This is also reflected by the fact that V can not be a Lyapunov function, since $V(x) = 0$ for all $x \in \mathcal{G}^{\star,d}$.

From equation (2.19) we already know that

$$\|E_t^{\cdot,d}\|^2 = e^{-2\lambda t} \|E_0^{\cdot,d}\|^2, \quad t \geq 0, \quad (2.20)$$

so we will now turn to $E_t^{\cdot,d} = X_t^{\cdot,d} - \bar{X}_t^{\cdot,d} \mathbf{1}_N$, which is the dynamics of the particle system that is orthogonal to $\mathcal{G}^{\cdot,d}$. In fact, we have

$$E_t^{\cdot,d} = X_t^{\cdot,d} - \bar{X}_t^{\cdot,d} \mathbf{1}_N = X_t^{\cdot,d} - \frac{1}{N} \mathbf{1}_N^\top X_t^{\cdot,d} \mathbf{1}_N = \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right) X_t^{\cdot,d}$$

and so

$$\dot{E}_t^{\cdot,d} = -\lambda P \hat{L} X_t^{\cdot,d},$$

where

$$P = I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top$$

is the matrix which corresponds to the projection on the orthogonal complement of $\mathcal{G}^{\star,d}$. Since

$$P \mathbf{1}_N = \mathbf{1}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \mathbf{1}_N = \mathbf{0}_N, \quad (2.21)$$

we have

$$P \hat{L} = P - P \mathbf{1}_N \ell^\top = P. \quad (2.22)$$

Thus, the dynamics of $E_t^{\cdot,d}$ is given by the linear ODE

$$\dot{E}_t^{\cdot,d} = -\lambda E_t^{\cdot,d},$$

which strikingly clarifies the dynamics of the particle system. This ODE has the unique solution

$$E_t^{\cdot,d} = e^{-\lambda t} E_0^{\cdot,d}, \quad t \geq 0, \quad (2.23)$$

which is obviously exponentially stable with rate λ at the unique fixed point $E^{\star,d} = \mathbf{0}_N$ and we trivially recover equation (2.20).

Since equation (2.23) does not depend on the considered dimension d of the particles, we have obtained the following result for the original dynamics:

Theorem 2.9 (Global Exponential Stability of the Deterministic System). The deterministic system

$$dX_t = -\lambda L X_t dt$$

with $X_t = (x_t^{1,1}, \dots, x_t^{1,D}, \dots, x_t^{N,1}, \dots, x_t^{N,D})^\top \in \mathbb{R}^{DN}$ is **globally exponentially stable** with exact rate λ on the set of equilibrium points

$$\mathcal{G}^* = \{x \in \mathbb{R}^{N \times D} : x^{1,d} = \dots = x^{N,d}, d = 1, \dots, D\}.$$

2.3 Exponential Stability under Discretization

In practical applications, usually the explicit Euler scheme

$$\hat{X}_{(k+1)\Delta}^n = \hat{X}_{k\Delta}^n - \lambda \left(\hat{X}_{k\Delta}^n - \nu_f^\alpha(\hat{X}_{k\Delta}^n) \right) \Delta, \quad k = 0, 1, \dots \quad (2.24)$$

with stepsize $\Delta > 0$ is used for the simulation of

$$dX_t^n = -\lambda (X_t^n - \nu_f^\alpha(X_t^n)) dt$$

for $n = 1, \dots, N$. Proceeding as in Subsection 2.1 it will again suffice to study the system in each dimension $d \in \{1, \dots, D\}$ on its own. So, for $X_t^{:,d} = (x_t^{1,d}, \dots, x_t^{N,d})^\top \in \mathbb{R}^N$, the Euler-discretization of

$$dX_t^{:,d} = -\lambda \hat{L}(X_t) X_t^{:,d} dt,$$

reads as

$$\hat{X}_{(k+1)\Delta}^{:,d} = \hat{X}_{k\Delta}^{:,d} - \lambda \hat{L}(\hat{X}_{k\Delta}) \hat{X}_{k\Delta}^{:,d} \Delta, \quad k = 0, 1, \dots \quad (2.25)$$

where $\hat{L}(\hat{X}_{k\Delta})$, is defined as before, i.e., as

$$\hat{L}(\hat{X}_{k\Delta}) := \begin{pmatrix} 1 - a_1(\hat{X}_{k\Delta}) & -a_2(\hat{X}_{k\Delta}) & \cdots & -a_N(\hat{X}_{k\Delta}) \\ -a_1(\hat{X}_{k\Delta}) & 1 - a_2(\hat{X}_{k\Delta}) & \cdots & -a_N(\hat{X}_{k\Delta}) \\ \vdots & \vdots & \ddots & \vdots \\ -a_1(\hat{X}_{k\Delta}) & -a_2(\hat{X}_{k\Delta}) & \cdots & 1 - a_N(\hat{X}_{k\Delta}) \end{pmatrix}. \quad (2.26)$$

Now we analyze also the dynamics of the Euler scheme on the orthogonal complement of $\mathcal{G}^{*,d}$, that is of

$$\hat{E}_{k\Delta}^{:,d} = \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T \right) \hat{X}_{k\Delta}^{:,d}.$$

With

$$P = I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T$$

we have

$$\begin{aligned} \hat{E}_{(k+1)\Delta}^{:,d} &= P \hat{X}_{(k+1)\Delta}^{:,d} \\ &= P \hat{X}_{k\Delta}^{:,d} - \lambda P \hat{L}(\hat{X}_{k\Delta}) \hat{X}_{k\Delta}^{:,d} \Delta \\ &= \hat{E}_{k\Delta}^{:,d} - \lambda P \hat{X}_{k\Delta}^{:,d} \Delta + \lambda P \mathbf{1}_N \ell(\hat{X}_{k\Delta})^T \hat{X}_{k\Delta}^{:,d} \Delta \\ &= (1 - \lambda \Delta) \hat{E}_{k\Delta}^{:,d} \end{aligned}$$

where we have used again (2.21), i.e., $P \mathbf{1}_N = \mathbf{0}_N$. So, we obtain

$$\hat{E}_{k\Delta}^{:,d} = (1 - \lambda \Delta)^k \hat{E}_0^{:,d}, \quad k = 0, 1, \dots \quad (2.27)$$

Since we have

$$\ln(1 - x) \leq -x, \quad x \in [0, 1),$$

we obtain that

$$(1 - \lambda \Delta)^k = \exp(k \ln(1 - \lambda \Delta)) \leq \exp(-\lambda k \Delta)$$

under the usual stability property $\Delta \lambda < 1$ for the Euler scheme. This implies

$$\|\hat{E}_{k\Delta}^{:,d}\| \leq e^{-\lambda k \Delta} \|\hat{E}_0^{:,d}\|, \quad (2.28)$$

so the Euler scheme preserves the globally exponential stability with rate λ of the equilibrium points from Theorem 2.6 under $\lambda \Delta < 1$.

3 Stochastic System

We now turn to the fully stochastic formulation of CBO, which incorporates both deterministic drift and Brownian diffusion. To rigorously analyze stability in this setting, we extend the deterministic framework to stochastic differential equations of the form $dZ_t = a(Z_t)dt + b(Z_t)dW_t$. Our goal is to establish a result analogous to the global exponential stability previously shown in the deterministic setting. To this end, we introduce the necessary concepts for exponential stability.

Definition 3.1 (Equilibrium Point [31, Definiton 3.2]). The state $Z^* \in \mathbb{R}^N$ is said to be an *equilibrium point* of an stochastic differential equation (SDE)

$$dZ_t = a(Z_t)dt + b(Z_t)dW_t, \quad t \geq 0, \quad (3.1)$$

if once Z_t is equal to Z^* , it remains equal to Z^* for all future times, i.e., Z^* has to satisfy $a(Z^*) = \mathbf{0}_N$ and $b(Z^*) = \mathbf{0}_N$.

Definition 3.2 (Almost Sure Exponential Stability for SDE [25, Section 4, Definition 3.1]). The equilibrium point $Z^* = \mathbf{0}_N$ is said to be *almost surely globally exponentially stable* with rate δ if there exists a constant $\delta > 0$ such that for all $Z_0 \in \mathbb{R}^N$ we have

$$\mathbb{P} \left(\limsup_{t \rightarrow \infty} \frac{1}{t} \ln (\|Z_t\|) \leq -\delta \right) = 1.$$

In addition to this *path-wise* notion of stability, another fundamental concept is based on the behavior of the second moment. This leads to the idea of *mean square stability*, where one does not analyze individual trajectories but instead considers the expected squared norm of the solution.

Definition 3.3 (Mean Square Exponential Stability for SDE [25, Section 4, Definition 4.1]). The equilibrium point $Z^* = \mathbf{0}_N$ is said to be *mean square globally exponentially stable* with rate δ if there exist constants $c, \delta > 0$ such that for all $Z_0 \in \mathbb{R}^N$ we have

$$\mathbb{E} (\|Z_t\|^2) \leq ce^{-\delta t} \mathbb{E} (\|Z_0\|^2) \quad \text{for all } t \geq 0.$$

In Chapter 2.2, we employed Lyapunov functions to motivate a change of coordinates, specifically the orthogonal projection onto $\mathcal{G}^{*,d}$. This allowed us to exploit a well-chosen linearization of the transformed system to establish exponential stability without further relying on Lyapunov functions. In the stochastic setting, we proceed analogously by once again applying the same change of coordinates. This enables us to derive both almost sure exponential stability and mean square exponential stability. To this end, we reformulate our system in the same spirit as in the deterministic case.

3.1 System Description

The dynamics of the system are governed by the stochastic differential equation

$$dZ_t^n = -\lambda (Z_t^n - \nu_f^\alpha(Z_t)) dt + \sigma (Z_t^n - \nu_f^\alpha(Z_t)) \circ dW_t^n \quad (3.2)$$

for $n = 1, \dots, N$, where W_t^1, \dots, W_t^N are independent d -dimensional Brownian motions, $Z_t = (Z_t^1, \dots, Z_t^N)^\top$ represents the particle states and \circ denotes the element-wise (Hadamard) product.

As in the deterministic case, we can reformulate $Z_t^n - \nu_f^\alpha(Z_t)$ by expressing it in terms of pair-wise differences. This transformation can be applied to both the drift and the diffusion terms, leading to the equivalent representation

$$dZ_t^n = -\lambda \left(\sum_{\substack{m=1 \\ m \neq n}}^N a_m \cdot (Z_t^n - Z_t^m) \right) dt + \sigma \left(\sum_{\substack{m=1 \\ m \neq n}}^N a_m \cdot (Z_t^n - Z_t^m) \right) \circ dW_t^n \quad (3.3)$$

for $n = 1, \dots, N$ with $a_m = a_m(Z_t)$ as defined in section 2.1. Additionally, we define $Z_t^{:,d} := (z_t^{1,d}, \dots, z_t^{N,d})^\top$ to organize the system component-wise across dimensions. Sorting the system by dimension, we obtain

the following structured representation

$$dZ_t = d \begin{pmatrix} z_t^{1,1} \\ \vdots \\ z_t^{N,1} \\ \vdots \\ z_t^{1,D} \\ \vdots \\ z_t^{N,D} \end{pmatrix} = -\lambda \begin{pmatrix} \sum_{m=2}^N a_m (z_t^{1,1} - z_t^{m,1}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (z_t^{N,1} - z_t^{m,1}) \\ \vdots \\ \sum_{m=2}^N a_m (z_t^{1,D} - z_t^{m,D}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (z_t^{N,D} - z_t^{m,D}) \end{pmatrix} dt + \sigma \begin{pmatrix} \sum_{m=2}^N a_m (z_t^{1,1} - z_t^{m,1}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (z_t^{N,1} - z_t^{m,1}) \\ \vdots \\ \sum_{m=2}^N a_m (z_t^{1,D} - z_t^{m,D}) \\ \vdots \\ \sum_{m=1}^{N-1} a_m (z_t^{N,D} - z_t^{m,D}) \end{pmatrix} \circ d \begin{pmatrix} w_t^{1,1} \\ \vdots \\ w_t^{N,1} \\ \vdots \\ w_t^{1,D} \\ \vdots \\ w_t^{N,D} \end{pmatrix} \quad (3.4)$$

where $W_t^n = (w_t^{n,1}, \dots, w_t^{n,D})^\top$ captures the Brownian motion associated with each component.

Since all components of each Brownian motion are independent, the evolution of each dimension is independent of the others, given the weights a_1, \dots, a_N , in the same way as in the deterministic case. Consequently, it suffices once more to analyze stability separately for each dimension $d \in \{1, \dots, D\}$.

Using the notation from Section 2, the system dynamics for a single dimension d can be expressed compactly as

$$dZ_t^{:,d} = -\lambda \hat{L} Z_t^{:,d} dt + \sigma \hat{L} Z_t^{:,d} \circ dW_t^{:,d}, \quad (3.5)$$

where $W_t^{:,d} = (w_t^{1,d}, \dots, w_t^{N,d})^\top \in \mathbb{R}^N$ represents the Brownian motion in dimension d . Similarly, the full system in (3.4) can be rewritten in matrix form as

$$dZ_t = -\lambda L Z_t dt + \sigma L Z_t \circ dW_t \quad (3.6)$$

with $W_t = (W_t^1, \dots, W_t^N)^\top \in \mathbb{R}^{DN}$.

In the following, we extend the stability analysis to stochastic CBO dynamics. In particular, we will demonstrate that the system (3.6) is almost sure and mean square globally exponentially stable on the set

$$\mathcal{G}^* = \{z \in \mathbb{R}^{N \times D} : z^{1,d} = \dots = z^{N,d}, d = 1, \dots, D\}, \quad (3.7)$$

which characterizes the set of equilibrium points of the system (3.6). This follows directly from the same calculations as in (2.7) and (2.8), since both the drift term $a(X_t)$ and the diffusion term $b(X_t)$ share the same functional form, differing only by a scaling factor.

3.2 Global Exponential Stability

Since the equilibrium manifold \mathcal{G}^* remains unchanged in the stochastic setting, we employ the same change of coordinates as in the deterministic case. Furthermore, due to the independence of the Brownian motions across particles and dimensions, it suffices to analyze each dimension $d \in \{1, \dots, D\}$ separately.

We consider the projected dynamics orthogonal to the consensus manifold by

$$E_t^{:,d} = Z_t^{:,d} - \bar{Z}_t^{:,d} \mathbf{1}_N, \quad \text{where} \quad \bar{Z}_t^{:,d} = \frac{1}{N} \mathbf{1}_N^T Z_t^{:,d},$$

so that $E_t^{:,d}$ lies in the orthogonal complement of

$$\mathcal{G}^{*,d} := \{z \in \mathbb{R}^N : z = \gamma \mathbf{1}_N, \gamma \in \mathbb{R}\}.$$

Analogously to the deterministic case, this leads to the convenient representation

$$E_t^{:,d} = Z_t^{:,d} - \bar{Z}_t^{:,d} \mathbf{1}_N = \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T \right) Z_t^{:,d} = P Z_t^{:,d},$$

where P is the projection onto the orthogonal complement of the consensus space.

In dimension d , the original system evolves according to

$$dZ_t^{:,d} = -\lambda \hat{L} Z_t^{:,d} dt + \sigma \hat{L} Z_t^{:,d} \circ dW_t^{:,d}.$$

Since P is a constant matrix (independent of the particle positions), it can be pulled inside the Itô-differential. Applying this projection yields

$$dE_t^{:,d} = d(PZ_t^{:,d}) = -\lambda P\hat{L}Z_t^{:,d} dt + \sigma P\hat{L}Z_t^{:,d} \circ dW_t^{:,d}.$$

Using the identity $P\hat{L} = P$, see (2.22), the system simplifies to

$$dE_t^{:,d} = -\lambda PZ_t^{:,d} dt + \sigma PZ_t^{:,d} \circ dW_t^{:,d} = -\lambda E_t^{:,d} dt + \sigma E_t^{:,d} \circ dW_t^{:,d}, \quad (3.8)$$

which takes the form of a linear stochastic differential equation.

This linear system admits the unique explicit solution

$$E_t^{:,d} = \exp\left(-\left(\lambda + \frac{\sigma^2}{2}\right)tI + \sigma W_t^{:,d}\right) E_0^{:,d}, \quad t \geq 0,$$

and, componentwise for each particle $n \in \{1, \dots, N\}$,

$$E_t^{n,d} = \exp\left(-\left(\lambda + \frac{\sigma^2}{2}\right)t + \sigma W_t^n\right) E_0^{n,d}, \quad t \geq 0. \quad (3.9)$$

Applying the logarithm to the absolute value of (3.9) and dividing by $t > 0$ yields

$$\frac{1}{t} \ln(|E_t^{n,d}|) = -\left(\lambda + \frac{\sigma^2}{2}\right) + \sigma \frac{W_t^n}{t} + \frac{1}{t} \ln(|E_0^{n,d}|). \quad (3.10)$$

By the law of the iterated logarithm [25, Theorem 4.2], we have

$$\limsup_{t \rightarrow \infty} \frac{W_t^n}{\sqrt{2t \log(\log t)}} = 1 \quad \text{almost surely.}$$

In particular, this implies

$$\limsup_{t \rightarrow \infty} \frac{W_t^n}{t} = 0 \quad \text{almost surely.}$$

Moreover, since the last term in (3.10) satisfies

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \ln(|E_0^{n,d}|) = 0,$$

we may take the lim sup of (3.10) to conclude

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \ln(|E_t^{n,d}|) = -\left(\lambda + \frac{\sigma^2}{2}\right) \quad \text{almost surely.} \quad (3.11)$$

Due to the equivalence of norms in finite-dimensional spaces, there exist constants $c_1, c_2 > 0$ such that

$$c_1 \cdot \max_{n=1, \dots, N} |E_t^{n,d}| \leq \|E_t^{:,d}\| \leq c_2 \cdot \max_{n=1, \dots, N} |E_t^{n,d}|.$$

Taking logarithms and dividing by t gives

$$\frac{1}{t} \ln(c_1) + \frac{1}{t} \ln\left(\max_{n=1, \dots, N} |E_t^{n,d}|\right) \leq \frac{1}{t} \ln(\|E_t^{:,d}\|) \leq \frac{1}{t} \ln(c_2) + \frac{1}{t} \ln\left(\max_{n=1, \dots, N} |E_t^{n,d}|\right).$$

Passing to the limit superior as $t \rightarrow \infty$, and using (3.11), we obtain

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \ln(\|E_t^{:,d}\|) = \limsup_{t \rightarrow \infty} \frac{1}{t} \ln\left(\max_{n=1, \dots, N} |E_t^{n,d}|\right) = -\left(\lambda + \frac{\sigma^2}{2}\right) \quad \text{almost surely.} \quad (3.12)$$

This proves that the system (3.8) is almost surely globally exponentially stable on $\mathcal{G}^{*,d}$ with exponential rate

$$\delta = \lambda + \frac{\sigma^2}{2}.$$

Since this analysis holds for each dimension d independently, the result carries over to the full system, completing the proof of almost sure global exponential stability for the original dynamics.

In particular, this shows that almost sure convergence holds for all positive values of λ and σ . Moreover, even if the drift parameter λ is negative, and the deterministic ODE would therefore be unstable, the presence of sufficiently strong noise can compensate for this and still yield almost sure stability. This phenomenon is known as *stabilization by noise*, see, e.g., the pioneering result on page 211f in [22] as well as [1, 24].

Theorem 3.4 (Almost Sure Global Exponential Stability of the Stochastic System). The stochastic system

$$dZ_t = -\lambda LZ_t dt + \sigma LZ_t \circ dW_t$$

with $Z_t = (z_t^{1,1}, \dots, z_t^{1,D}, \dots, z_t^{N,1}, \dots, z_t^{N,D})^\top \in \mathbb{R}^{DN}$ is **almost surely globally exponentially stable** with exact rate $\lambda + \frac{\sigma^2}{2}$ on the set of equilibrium points

$$\mathcal{G}^* = \{z \in \mathbb{R}^{N \times D} : z^{1,d} = \dots = z^{N,d}, d = 1, \dots, D\}.$$

To establish not only almost sure exponential convergence but also convergence in mean square, we now consider the squared explicit solution (3.9) of the linear SDE in expectation:

$$\mathbb{E}(|E_t^{n,d}|^2) = \mathbb{E}\left(\exp\left(-2\left(\lambda + \frac{\sigma^2}{2}\right)t + 2\sigma W_t^n\right) |E_0^{n,d}|^2\right).$$

Since W_t^n is independent of the initial condition $E_0^{n,d}$, the expectation factorizes

$$\mathbb{E}(|E_t^{n,d}|^2) = \exp\left(-2\left(\lambda + \frac{\sigma^2}{2}\right)t\right) \mathbb{E}(\exp(2\sigma W_t^n)) \mathbb{E}(|E_0^{n,d}|^2).$$

Recalling that $W_t^n \sim \mathcal{N}(0, t)$, the random variable $2\sigma W_t^n$ is Gaussian with mean 0 and variance $4\sigma^2 t$. For a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, it is well known that

$$\mathbb{E}(e^{aX}) = \exp\left(a\mu + \frac{1}{2}a^2\sigma^2\right).$$

Applying this identity with $X = W_t^n$, $a = 2\sigma$, $\mu = 0$, and $\text{Var}(W_t^n) = t$, we obtain

$$\mathbb{E}(\exp(2\sigma W_t^n)) = \exp\left(\frac{1}{2}(2\sigma)^2 t\right) = \exp(2\sigma^2 t).$$

Hence, the mean squared value simplifies to

$$\begin{aligned} \mathbb{E}(|E_t^{n,d}|^2) &= \exp\left(-2\left(\lambda + \frac{\sigma^2}{2}\right)t\right) \exp(2\sigma^2 t) \mathbb{E}(|E_0^{n,d}|^2) \\ &= \exp\left(-2\left(\lambda - \frac{\sigma^2}{2}\right)t\right) \mathbb{E}(|E_0^{n,d}|^2). \end{aligned}$$

Since this relation holds for each $n = 1, \dots, N$, we can apply the linearity of expectation to obtain a bound for the full vector

$$\mathbb{E}(\|E_t^{:,d}\|^2) = \exp\left(-2\left(\lambda - \frac{\sigma^2}{2}\right)t\right) \mathbb{E}(\|E_0^{:,d}\|^2). \quad (3.13)$$

Therefore, the system (3.8) is mean square globally exponentially stable on $\mathcal{G}^{*,d}$, with stability constant $c = 1$ and exponential decay rate

$$\delta = 2\lambda - \sigma^2. \quad (3.14)$$

Since this analysis holds for each dimension d independently, the result carries over to the full system, completing the proof of mean square global exponential stability for the original dynamics.

Theorem 3.5 (Mean Square Global Exponential Stability of the Stochastic System). The stochastic system

$$dZ_t = -\lambda LZ_t dt + \sigma LZ_t \circ dW_t$$

with $Z_t = (z_t^{1,1}, \dots, z_t^{1,D}, \dots, z_t^{N,1}, \dots, z_t^{N,D})^\top \in \mathbb{R}^{DN}$ is **mean square globally exponentially stable** with constant $c = 1$ and exact rate $2\lambda - \sigma^2$ on the set of equilibrium points

$$\mathcal{G}^* = \{z \in \mathbb{R}^{N \times D} : z^{1,d} = \dots = z^{N,d}, d = 1, \dots, D\}.$$

In order for the system to be mean square exponentially stable, it is necessary that the decay rate $\delta = 2\lambda - \sigma^2$ is positive. This requires the condition

$$2\lambda > \sigma^2. \quad (3.15)$$

We emphasize that, independently of this condition, almost sure exponential convergence always holds for all values of $\lambda > 0$ and $\sigma \geq 0$, as shown earlier.

We note that the stability condition (3.15) and the exponential rate $2\lambda - \sigma^2$ obtained in Theorem 3.5 coincide with the findings of [12], where the anisotropic CBO dynamics were analyzed in the mean-field limit. In contrast, our approach neither requires structural assumptions on the objective function f nor asymptotic regimes such as $N \rightarrow \infty$ or $\alpha \rightarrow \infty$. Moreover, it yields the exact mean square decay rate, rather than an upper bound that can get arbitrarily close to $2\lambda - \sigma^2$. At the same time, our analysis is restricted to the dynamics orthogonal to the consensus manifold \mathcal{G}^* , establishing exponential convergence toward \mathcal{G}^* but not to a specific minimizer. In contrast, [12] succeeded in proving convergence to the global minimizer in the mean-field setting. A precise characterization of the limiting point within \mathcal{G}^* in our framework remains an open problem for future research.

3.3 Exponential Stability under Discretization

In most works on CBO, such as [29, 13, 15, 34], the Euler-Maruyama scheme [18] is employed to simulate the dynamics of the system

$$dZ_t^n = -\lambda (Z_t^n - \nu_f^\alpha(Z_t)) dt + \sigma (Z_t^n - \nu_f^\alpha(Z_t)) \circ dW_t^n, \quad n = 1, \dots, N.$$

The Euler-Maruyama scheme then provides the discretized approximation $\hat{Z}_{k\Delta}^n$ with time step size $\Delta > 0$, given by

$$\hat{Z}_{(k+1)\Delta}^n = \hat{Z}_{k\Delta}^n - \lambda (\hat{Z}_{k\Delta}^n - \nu_f^\alpha(\hat{Z}_{k\Delta})) \Delta + \sigma (\hat{Z}_{k\Delta}^n - \nu_f^\alpha(\hat{Z}_{k\Delta})) \circ \Delta W_k^n, \quad (3.16)$$

where \circ denotes the element-wise (Hadamard) product. Here we use the abbreviation $\Delta W_k^n = W_{(k+1)\Delta}^n - W_{k\Delta}^n$ for the increments of the Brownian motion W^n . Note that $\Delta W_k^n \sim \mathcal{N}(\mathbf{0}_D, \Delta \cdot I_D)$ and moreover the random vectors $\Delta W_0^n, \dots, \Delta W_k^n, \dots$ are independent. Thus, the Euler-Maruyama scheme combines a forward Euler step for the drift with a pointwise scaled Gaussian perturbation to model the stochastic diffusion component.

As in Subsection 3.1, it suffices to study the dynamics of each coordinate $d \in \{1, \dots, D\}$ separately. For $Z_t^{:,d} := (Z_t^{1,d}, \dots, Z_t^{N,d})^\top \in \mathbb{R}^N$ the stochastic system

$$dZ_t^{:,d} = -\lambda \hat{L}(Z_t) Z_t^{:,d} dt + \sigma \hat{L}(Z_t) Z_t^{:,d} \circ dW_t^{:,d}$$

can be approximated via the Euler-Maruyama scheme as

$$\hat{Z}_{(k+1)\Delta}^{:,d} = \hat{Z}_{k\Delta}^{:,d} - \lambda \hat{L}(\hat{Z}_{k\Delta}) \hat{Z}_{k\Delta}^{:,d} \Delta + \sigma \hat{L}(\hat{Z}_{k\Delta}) \hat{Z}_{k\Delta}^{:,d} \circ \Delta W_k^{:,d}, \quad k = 0, 1, \dots, \quad (3.17)$$

where $\Delta W_k^{:,d} \in \mathbb{R}^N$ are the increments of the Brownian motions which are acting on the dimension d of the particles. Here, the matrix $\hat{L}(\hat{Z}_{k\Delta}) \in \mathbb{R}^{N \times N}$ is given as in (2.26).

Analogously to the deterministic case, we now analyze the dynamics of the Euler-Maruyama scheme on the orthogonal complement of $\mathcal{G}^{*,d}$, that is

$$\hat{E}_{k\Delta}^{:,d} = \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right) \hat{Z}_{k\Delta}^{:,d} = P \hat{Z}_{k\Delta}^{:,d}.$$

We obtain the dynamics

$$\begin{aligned} \hat{E}_{(k+1)\Delta}^{:,d} &= P \hat{Z}_{(k+1)\Delta}^{:,d} \\ &= P \left(\hat{Z}_{k\Delta}^{:,d} - \lambda \hat{L}(\hat{Z}_{k\Delta}) \hat{Z}_{k\Delta}^{:,d} \Delta + \sigma \hat{L}(\hat{Z}_{k\Delta}) \hat{Z}_{k\Delta}^{:,d} \circ \Delta W_k^{:,d} \right) \\ &= \hat{E}_{k\Delta}^{:,d} - \lambda P \hat{L}(\hat{Z}_{k\Delta}) \hat{Z}_{k\Delta}^{:,d} \Delta + \sigma P \hat{L}(\hat{Z}_{k\Delta}) \hat{Z}_{k\Delta}^{:,d} \circ \Delta W_k^{:,d} \\ &= \hat{E}_{k\Delta}^{:,d} - \lambda P \hat{Z}_{k\Delta}^{:,d} \Delta + \sigma P \hat{Z}_{k\Delta}^{:,d} \circ \Delta W_k^{:,d}, \end{aligned}$$

where we have again applied the relation $P \hat{L}(\hat{Z}_{k\Delta}) = P$.

We now analyze the mean square stability of the Euler-Maruyama discretization of the orthogonal dynamics. From the recursion

$$\hat{E}_{(k+1)\Delta}^{:,d} = \left(\mathbf{1}_N - \lambda \Delta \mathbf{1}_N + \sigma \Delta W_k^{:,d} \right) \circ \hat{E}_{k\Delta}^{:,d}, \quad (3.18)$$

we compute the second moment. Since the Brownian increments $\Delta W_k^{:,d} \sim \mathcal{N}(\mathbf{0}_N, \Delta \cdot I_N)$ are independent of $\widehat{E}_{k\Delta}^{:,d}$, we obtain

$$\begin{aligned} \mathbb{E} \left(\|\widehat{E}_{(k+1)\Delta}^{:,d}\|^2 \right) &= \mathbb{E} \left(\left\| \left(\mathbf{1}_N - \lambda \Delta \mathbf{1}_N + \sigma \Delta W_k^{:,d} \right) \circ \widehat{E}_{k\Delta}^{:,d} \right\|^2 \right) \\ &= \sum_{n=1}^N \mathbb{E} \left((1 - \lambda \Delta + \sigma \Delta W_k^{n,d})^2 \right) \mathbb{E} \left((\widehat{E}_{k\Delta}^{n,d})^2 \right) \\ &= ((1 - \lambda \Delta)^2 + \sigma^2 \Delta) \mathbb{E} \left(\|\widehat{E}_{k\Delta}^{:,d}\|^2 \right), \end{aligned}$$

where we used that $\mathbb{E}(\Delta W_k^{n,d}) = 0$ and $\mathbb{E}((\Delta W_k^{n,d})^2) = \Delta$.

By iteration, we get

$$\mathbb{E} \left(\|\widehat{E}_{k\Delta}^{:,d}\|^2 \right) = ((1 - \lambda \Delta)^2 + \sigma^2 \Delta)^k \mathbb{E} \left(\|E_0^{:,d}\|^2 \right). \quad (3.19)$$

Since for $x \in [0, 1)$ it holds that $\ln(1 - x) \leq -x$, we estimate

$$\ln((1 - \lambda \Delta)^2 + \sigma^2 \Delta) = \ln(1 - (2\lambda \Delta - \lambda^2 \Delta^2 - \sigma^2 \Delta)) \leq -(2\lambda \Delta - \lambda^2 \Delta^2 - \sigma^2 \Delta).$$

Under the condition $2\lambda - \sigma^2 - \lambda^2 \Delta \in [0, 1)$, this implies the exponential decay bound

$$((1 - \lambda \Delta)^2 + \sigma^2 \Delta)^k \leq \exp(-k\Delta(2\lambda - \lambda^2 \Delta - \sigma^2)).$$

Hence, the Euler–Maruyama approximation $\widehat{E}_{k\Delta}^{:,d}$ preserves mean square global exponential stability of the equilibrium manifold with rate

$$\delta_{\text{EM}} := 2\lambda - \sigma^2 - \lambda^2 \Delta > 0, \quad (3.20)$$

provided the step size Δ satisfies

$$0 < \Delta < \frac{2\lambda - \sigma^2}{\lambda^2}. \quad (3.21)$$

More precisely, under this step size condition the Euler–Maruyama scheme satisfies

$$\mathbb{E} \left(\|\widehat{E}_{k\Delta}^{:,d}\|^2 \right) \leq \exp(-k\Delta(2\lambda - \lambda^2 \Delta - \sigma^2)) \mathbb{E} \left(\|E_0^{:,d}\|^2 \right), \quad (3.22)$$

which shows that the mean square distance to the equilibrium manifold decays exponentially at rate δ_{EM} .

To establish almost sure exponential convergence of the Euler–Maruyama approximation $\widehat{E}_{k\Delta}^{:,d}$, we first consider the evolution of its components. For each n , the recursion (3.18) reads

$$\widehat{E}_{(k+1)\Delta}^{n,d} = \left(1 - \lambda \Delta + \sigma \Delta W_k^{n,d} \right) \widehat{E}_{k\Delta}^{n,d}, \quad (3.23)$$

which can be iteratively expanded to yield the product representation

$$\widehat{E}_{k\Delta}^{n,d} = \widehat{E}_0^{n,d} \prod_{j=0}^{k-1} \left(1 - \lambda \Delta + \sigma \Delta W_j^{n,d} \right).$$

Taking logarithms and dividing by $k\Delta$, we obtain

$$\frac{1}{k\Delta} \ln \left(|\widehat{E}_{k\Delta}^{n,d}| \right) = \frac{1}{k\Delta} \ln \left(|\widehat{E}_0^{n,d}| \right) + \frac{1}{k\Delta} \sum_{j=0}^{k-1} \ln \left(|1 - \lambda \Delta + \sigma \Delta W_j^{n,d}| \right).$$

Since the increments $\Delta W_j^{n,d}$ are i.i.d $\mathcal{N}(0, \Delta)$, the strong law of large numbers implies

$$\lim_{k \rightarrow \infty} \frac{1}{k\Delta} \ln \left(|\widehat{E}_{k\Delta}^{n,d}| \right) = \frac{1}{\Delta} \cdot \mathbb{E} \left(\ln \left(|1 - \lambda \Delta + \sigma \sqrt{\Delta} Z| \right) \right) \quad \text{almost surely,}$$

where $Z \sim \mathcal{N}(0, 1)$.

As this asymptotic rate is identical for all components and independent of n , we may pass from the componentwise behavior to the full vector norm. Using norm equivalence in finite dimensions and an argument analogous to the transition from (3.11) to (3.12), we conclude

$$\lim_{k \rightarrow \infty} \frac{1}{k\Delta} \ln \left(\|\widehat{E}_{k\Delta}^{:,d}\| \right) = \frac{1}{\Delta} \cdot \mathbb{E} \left(\ln \left(|1 - \lambda\Delta + \sigma\sqrt{\Delta}Z| \right) \right) \quad \text{almost surely.} \quad (3.24)$$

Consequently, the Euler-Maruyama approximation $\widehat{E}_{k\Delta}^{:,d}$ converges almost surely exponentially to the equilibrium manifold if and only if the expectation on the right-hand side is negative. In that case, the almost sure exponential decay rate is given by

$$\delta_{\text{EM}} = -\frac{1}{\Delta} \cdot \mathbb{E} \left(\ln \left(|1 - \lambda\Delta + \sigma\sqrt{\Delta}Z| \right) \right). \quad (3.25)$$

In the proof of Theorem 4.3, [19] establishes the bound

$$\mathbb{E} \left[\ln \left(|1 - \lambda\Delta + \sigma\sqrt{\Delta}Z| \right) \right] \leq -\left(\lambda + \frac{\sigma^2}{2} \right) \Delta + C\Delta^2, \quad (3.26)$$

where $C = C(\lambda, \sigma) > 0$ is a constant depending on the drift and diffusion coefficients. Inserting this into the almost sure asymptotic expression derived above, we obtain the estimate

$$\lim_{k \rightarrow \infty} \frac{1}{k\Delta} \ln \left(\|\widehat{E}_{k\Delta}^{:,d}\| \right) \leq -\left(\lambda + \frac{\sigma^2}{2} - C\Delta \right) \quad \text{almost surely.} \quad (3.27)$$

Thus, provided that

$$0 < \Delta < \frac{2\lambda + \sigma^2}{2C}, \quad (3.28)$$

the Euler-Maruyama scheme exhibits almost sure exponential convergence to the equilibrium manifold with rate

$$\delta_{\text{EM}} = \lambda + \frac{\sigma^2}{2} - C\Delta. \quad (3.29)$$

4 Numerical Results

In this chapter, we present numerical experiments to study the convergence behavior of both the **deterministic system**

$$dX_t^n = -\lambda (X_t^n - \nu_f^\alpha(X_t)) dt, \quad n = 1, \dots, N, \quad (4.1)$$

and the **stochastic system**

$$dZ_t^n = -\lambda (Z_t^n - \nu_f^\alpha(Z_t)) dt + \sigma (Z_t^n - \nu_f^\alpha(Z_t)) \circ dW_t^n, \quad n = 1, \dots, N, \quad (4.2)$$

applied to the Rastrigin function [33, Optimization Test Problems]. This classical non-convex benchmark function is defined for $x = (x_1, \dots, x_D) \in \mathbb{R}^D$ as

$$f_{\text{Rastrigin}}(x) = 10D + \sum_{d=1}^D (x_d^2 - 10 \cos(2\pi x_d)), \quad (4.3)$$

with global minimizer $\mathbf{X}^* = \mathbf{0}_D$. Its landscape contains many regularly spaced local minima, which makes the problem difficult for optimization methods since they can easily get trapped. In the context of CBO, the Rastrigin function is therefore well suited to test the *exploration ability* of particle systems and has become a standard example in numerical studies [29, 13, 12].

Our implementation builds upon the Python-based CBO framework introduced in [35]. The complete source code, including all function definitions, numerical solvers, and visualization scripts, is available in the Git repository accompanying this paper [17].

Unless stated otherwise, the following settings are used throughout all experiments:

- **Dimension:** $D = 2$, allowing intuitive visualization of particle trajectories.
- **Number of particles:** $N = 100$, initialized from the uniform distribution $\mathcal{U}([-5, 5]^D)$.

- **Discretization:** We employ the Euler–Maruyama method [18] with a time step $\Delta = 0.05$.
- **Time horizon:** $T = 4.95$ corresponding to 100 time points (epochs).
- **Consensus parameters:** $\lambda = 1$, $\alpha = 1000$.

In the numerical experiments, we compare the observed convergence rates with the exponential stability results established in Theorems 3.4 and 3.5, which concern the continuous system. We do not refer to the discretized rates, since the additional term $\lambda^2\Delta$ in their decay estimate is negligible for the step size $\Delta = 0.05$ considered here.

4.1 Convergence Behavior of 2D-Rastrigin

We start by comparing the deterministic system (4.1) and the stochastic system (4.2) for two diffusion values,

$$\sigma = 1 \quad \text{and} \quad \sigma = \sqrt{2.1}.$$

The first satisfies the mean square stability condition (3.15) from Theorem 3.5,

$$2\lambda > \sigma^2,$$

while the second violates it. We note, that the almost sure convergence holds for any configuration of λ and σ , due to Theorem 2.9.

We visualize the trajectories of all $N = 100$ particles in the different system variants on the 2D-Rastrigin function. Each row in Figure 1 corresponds to one system. We show four snapshots per system: the initial state at time $t = 0$, an intermediate state after 15 time steps ($t = 0.75$), another after 40 steps ($t = 2$), and the final state after 100 steps ($t = 4.95$). In each plot, the current consensus point is indicated.

The systems are ordered by increasing stochasticity, starting with the purely deterministic case. This visualization allows us to qualitatively assess the influence of noise on the convergence behavior. All views are rendered from above; the global minimum $\mathbf{X}^* = (0, 0)$ lies at the center of each snapshot.

Figure 1 shows that all three systems eventually form consensus at the global minimum. The swarm of the purely deterministic system steadily contracts toward the minimum without significant deformation of the initial distribution. The roughly square-shaped initialization from the uniform distribution $\mathcal{U}([-5, 5]^2)$ is preserved over time.

In contrast, the stochastic system exhibits more exploration, particularly when the diffusion parameter $\sigma^2 = 2.1$. The particle distribution becomes visibly more dispersed at intermediate times (time = 0.75 and time = 2.0), indicating that more noise facilitates broader exploration of the domain. Still, the system converges to consensus by the final time point, since the system converges almost surely for every combination of λ and σ in accordance to Theorem 2.9.

Since all methods converge to the same minimum, we now focus on the dynamics of consensus formation, i.e., how effectively consensus emerges over time. To quantify this, we consider the expected squared distance from the consensus manifold \mathcal{G}^* , defined as

$$\mathbb{E}(\|E_t\|^2) := \sum_{d=1}^D \mathbb{E}(\|E_t^{:,d}\|^2) = \sum_{d=1}^D \mathbb{E}(\|PY_t^{:,d}\|^2),$$

where $P = I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$ and $Y_t = Z_t$ in the stochastic case or $Y_t = X_t$ in the deterministic case. Figure 3 shows the evolution of $\mathbb{E}(\|E_t\|^2)$ over time for the same three scenarios as in Figure 1. The expectation is approximated by a Monte-Carlo simulation with 1000 runs. For comparison, the dotted lines indicate the theoretical convergence rates from Theorem 3.5. Note that for the stochastic system with σ^2 violating the stability condition $2\lambda > \sigma^2$, no such rate is available.

The purely deterministic system exhibits perfectly smooth exponential convergence, precisely matching the theoretical rate from Theorem 3.5. Since no randomness is present, no fluctuations occur. In contrast, the stochastic system with $\sigma^2 = 1$ shows small fluctuations, yet due to the moderate diffusion the Monte-Carlo average over 1000 runs closely follows the expected exponential rate from Theorem 3.5, resulting in an almost smooth convergence. For the stochastic system with $\sigma^2 = 2.1$, where the stability condition $2\lambda > \sigma^2$ is violated, randomness dominates: even after averaging over 1000 runs, pronounced fluctuations remain. At first, the strong diffusion causes the expected distance to consensus to increase, since the system explores the domain more widely. Later, however, the distance decreases again, and

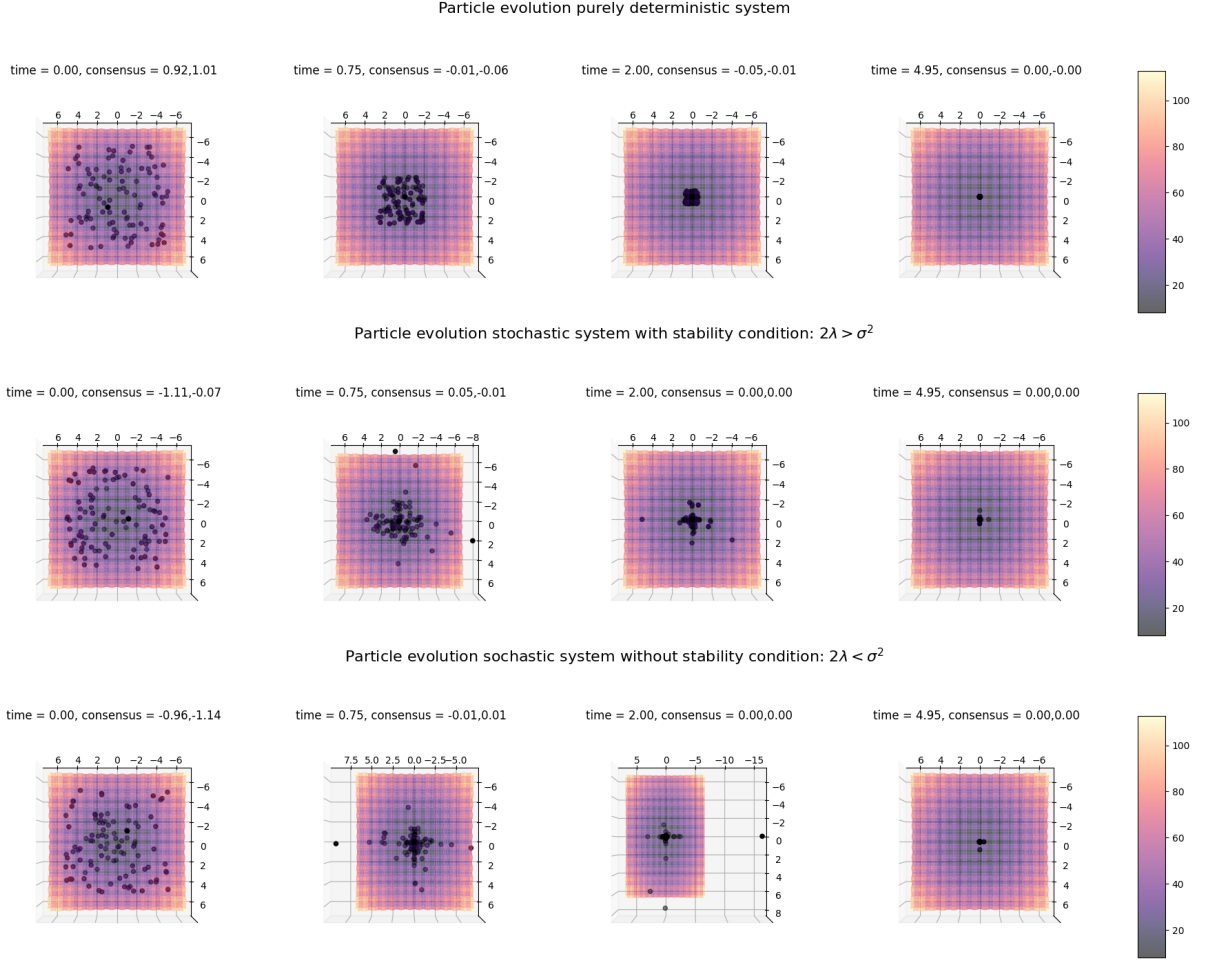


Figure 1: Evolution of $N = 100$ particles on the 2D-Rastrigin function for the purely deterministic system (4.1), the stochastic system (4.2) with $\sigma = 1$ (satisfying the stability condition $2\lambda > \sigma^2$), and the stochastic system (4.2) with $\sigma = \sqrt{2.1}$ (violating the stability condition $2\lambda > \sigma^2$). For each case, we display four snapshots over time: at initialization (time = 0), after 15 time steps (time = 0.75), after 40 time steps (time = 2), and at the end of the simulation (time = 4.95). All plots are rendered from a top-down perspective; the global minimum is located at $\mathbf{X}^* = (0, 0)$.

we still see convergence to consensus, but no longer at an exponential rate. This shows that large diffusion slows down convergence and prevents exponential stability, but consensus is still reached within the observed time horizon.

To investigate whether consensus formation behaves differently in the **isotropic** variant of CBO, we replace the anisotropic diffusion term $\sigma(Z_t^n - \nu_f^\alpha(Z_t))$ with the isotropic one $\sigma\|Z_t^n - \nu_f^\alpha(Z_t)\|_2$. We then repeat the simulations using $\sigma = \sqrt{0.5}$ and $\sigma = 1$ for the isotropic stochastic system. These values are motivated by the convergence analysis in [13], where in the mean-field limit ($N \rightarrow \infty$, $\alpha \rightarrow \infty$) it was shown that under the isotropic stability condition

$$2\lambda > D\sigma^2, \quad (4.4)$$

the system converges exponentially fast with rate $2\lambda - D\sigma^2$. According to this criterion, the purely deterministic system and the case $\sigma = \sqrt{0.5}$ are exponentially stable, while $\sigma = 1$ violates the condition. Consequently, in Figure 3, the expected convergence rates (dotted lines) are shown only for the first two cases.

As we can see in Figure 3, systems satisfying (4.4) converge exponentially, matching the rates established in [13]. The system with $\sigma = 1$, which violates the stability condition, exhibits larger deviations from consensus due to stronger exploration. Within the observed time horizon, this system shows no clear convergence toward consensus.

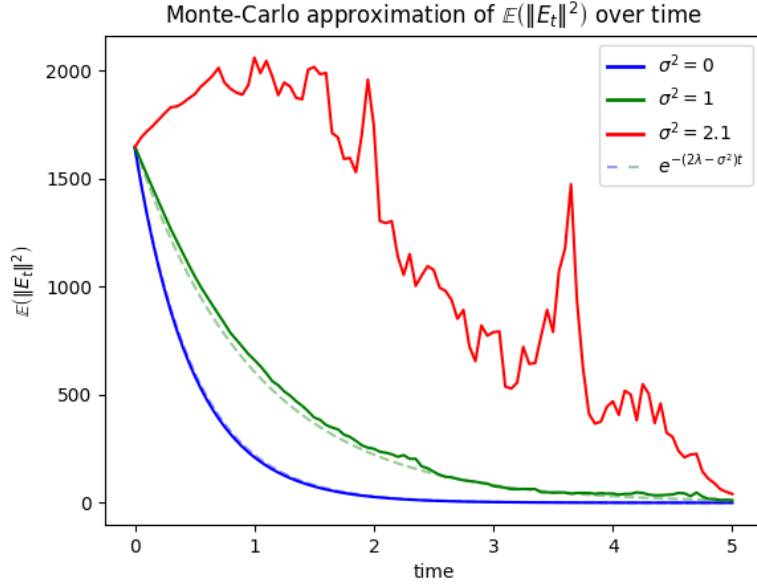


Figure 2: Monte-Carlo approximation over 1000 runs showing the evolution of the dynamics orthogonal to the consensus manifold \mathcal{G}^* for the purely deterministic system (4.1), the stochastic system (4.2) with $\sigma = 1$ (satisfying the stability condition $2\lambda > \sigma^2$), and the stochastic system (4.2) with $\sigma = \sqrt{2.1}$ (violating the stability condition). Dotted lines indicate the theoretical convergence rates from Theorem 3.5.

These results indicate that the anisotropic variant generally achieves faster convergence to the minimum, even for parameters that no longer guarantee exponential stability. They also confirm that the mean-field convergence rates from [13] are approximately preserved for finite $N \ll \infty$.

However, in the isotropic case, the diffusion term $\sigma \|Z_t^n - \nu_f^\alpha(Z_t)\|_2$ affects all dimensions jointly, depending on the positions of all particles. As a result, the approach used here for the anisotropic system cannot be directly applied, and extending the exponential stability analysis to the isotropic case is left for future work.

4.2 Sensitivity of the convergence behavior

We now examine how the parameters α , N , and D affect the convergence behavior to consensus. In the following, we vary each of the three parameters individually while keeping the others constant. All simulations are averages based on 1000 independent runs. Before averaging, we limit the values of each individual run to a maximum threshold, preventing extreme outliers, particularly in large-noise systems, from dominating the observed behavior. For each configuration, we display the evolution of the dynamics orthogonal to the consensus manifold over time.

Each figure is structured into three subplots corresponding to different system. Plot (a) displays the purely deterministic system. Plot (b) uses $\sigma = 1$, which satisfies the stability condition (3.15) from Theorem 3.5. Plot (c) employs $\sigma = \sqrt{2.1}$, which violates stability condition (3.15). Note that all three subplots share the same color scale to highlight differences in magnitude across the diffusion regimes.

We start by changing the value of α . It takes each value between 1 and 1001 with steps of 20 in between. Figure 4 shows the results.

In the deterministic setting (a), convergence to consensus appears smoothly and nearly identically across all values of α . This pattern persists in the large-diffusion cases (b) and (c), at least in expectation. The results suggest that the precise value of α has only a minor influence on the overall convergence behavior.

We repeat the same procedure, this time varying only the number of particles N from 10 to 1010 in steps of 20. The results are displayed in Figure 5.

In the purely deterministic system (a) and the system with small noise (b), increasing N primarily enlarges the initial value of $\mathbb{E}(\|E_t\|^2)$, so convergence remains exponential but slows down due to the higher starting values. In contrast, in the system with large noise (c), larger N not only raises the initial value of $\mathbb{E}(\|E_t\|^2)$ but also enhances exploration, causing the distance to consensus to increase

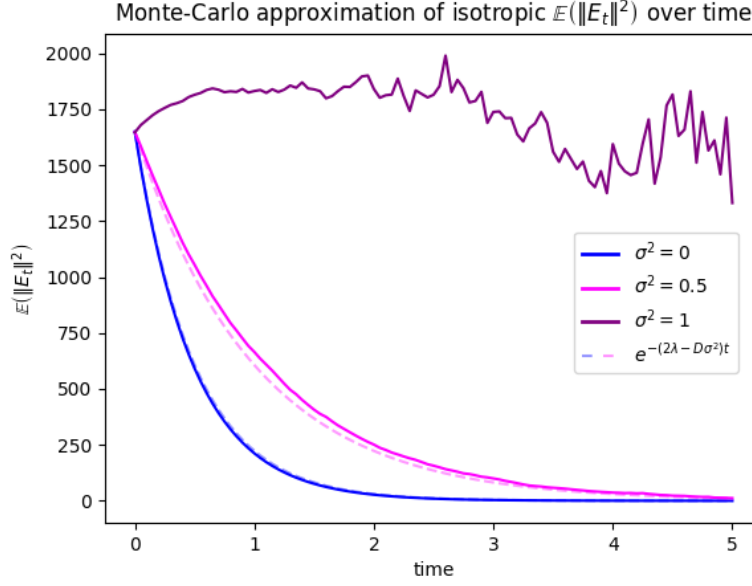


Figure 3: Monte-Carlo approximation over 1000 runs showing the evolution of the isotropic dynamics orthogonal to the consensus manifold \mathcal{G}^* for the purely deterministic system (4.1), the isotropic stochastic system with $\sigma = \sqrt{0.5}$ (satisfying the isotropic stability condition $2\lambda > D\sigma^2$), and the isotropic stochastic system with $\sigma = 1$ (violating the isotropic stability condition $2\lambda > D\sigma^2$). Dotted lines indicate the theoretical convergence rates from [13].

temporarily before eventually decreasing. This suggests that using fewer particles can be more efficient in such scenarios.

We now turn to the question of how the system behaves as the problem dimension D increases. To investigate this, we vary D from 1 to 201 in steps of 10.

All three systems show that increasing D raises the initial distance to consensus, and in highly stochastic settings also enhances exploration. This illustrates how both high dimensionality and strong stochasticity can slow down consensus formation, which is consistent with prior findings [9, 12].

We emphasize that our analysis focuses solely on **whether consensus is achieved**, not on the **quality of the consensus point** as an optimizer, which is often investigated in other works like [29, 14]. Our results show that small N and σ can lead to fast consensus formation, but this may result in convergence to suboptimal points. In general optimization tasks, a slow and exploratory phase can be beneficial, because the speed to consensus and solution quality do not necessarily coincide.

4.3 Convergence Behavior of other Benchmark functions

We also performed the same numerical simulations on two additional benchmark functions, which feature different landscape characteristics compared to the Rastrigin function.

Rosenbrock Function. The Rosenbrock function [33] is a smooth test function with a narrow, curved valley and a global minimum at $\mathbf{X}^* = \mathbf{1}_D$. It is given by

$$f_{\text{Rosenbrock}}(x) = \sum_{d=1}^{D-1} ((1 - x_d)^2 + 100(x_{d+1} - x_d^2)^2), \quad (4.5)$$

Within the CBO framework, it serves to evaluate the *convergence efficiency* of the particle system.

Discontinuous Integrand. The Discontinuous Integrand function [33] exhibits a steep exponential peak inside a hypercube and is zero elsewhere, creating a sharp discontinuity. This makes it suitable for

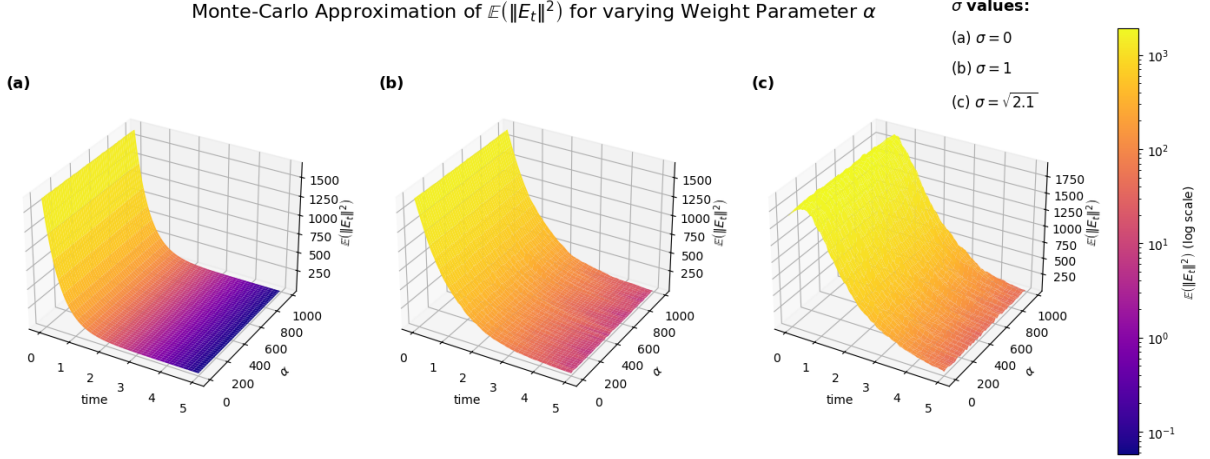


Figure 4: Monte-Carlo approximation over 1000 runs for varying α showing the evolution of the dynamics orthogonal to the consensus manifold \mathcal{G}^* for the (a) purely deterministic system (4.1), the (b) stochastic system (4.2) with $\sigma = 1$ (satisfying the stability condition $2\lambda > \sigma^2$), and the (c) stochastic system (4.2) with $\sigma = \sqrt{2.1}$ (violating the stability condition).

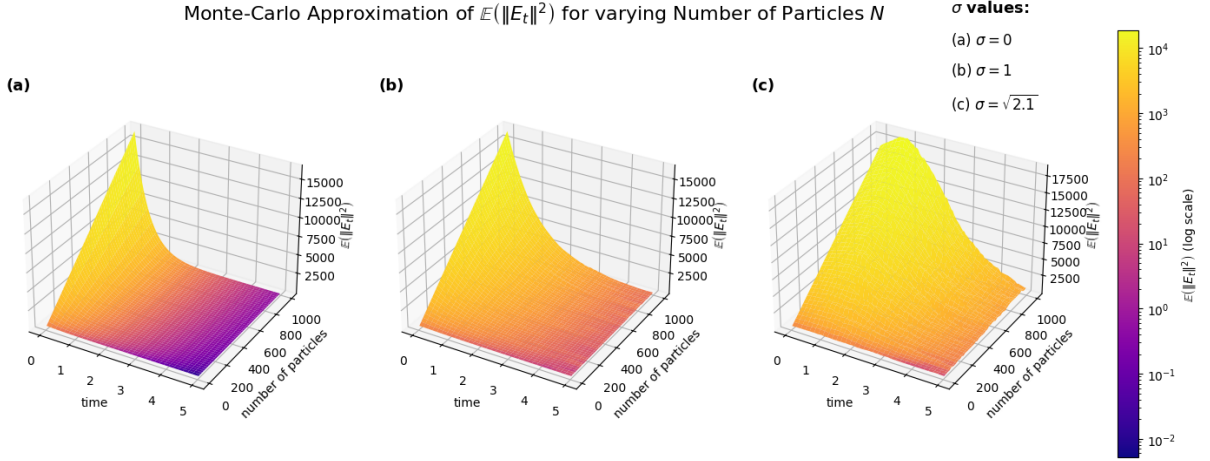


Figure 5: Monte-Carlo approximation over 1000 runs for varying number of particles N showing the evolution of the dynamics orthogonal to the consensus manifold \mathcal{G}^* for the (a) purely deterministic system (4.1), the (b) stochastic system (4.2) with $\sigma = 1$ (satisfying the stability condition $2\lambda > \sigma^2$), and the (c) stochastic system (4.2) with $\sigma = \sqrt{2.1}$ (violating the stability condition).

testing *stochastic navigation* capabilities of CBO particles in the presence of abrupt changes. It is defined as

$$f_{\text{disc}}(x) = \begin{cases} -\exp\left(\sum_{d=1}^D 5x_d\right), & \text{if } x_d \leq \frac{1}{2} \text{ for all } d = 1, \dots, D, \\ 0, & \text{otherwise.} \end{cases} \quad (4.6)$$

In both cases, our simulations revealed behavior largely consistent with the results observed for the Rastrigin function, confirming the robustness of the CBO dynamics across different types of objective landscapes.

Acknowledgments

This work was financially supported by the DFG Projects GO1920/11-1 and 12-1.

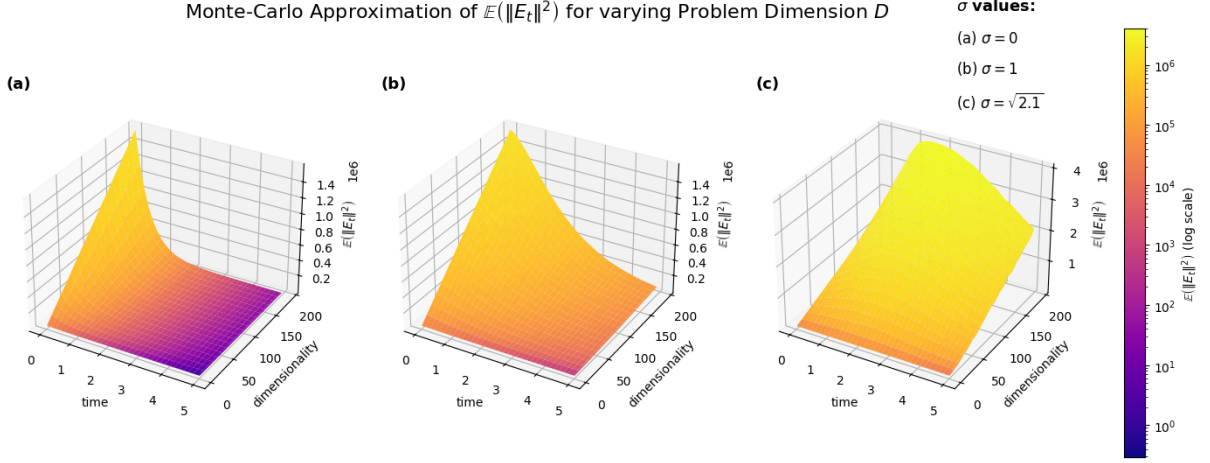


Figure 6: Monte-Carlo approximation over 1000 runs for varying D showing the evolution of the dynamics orthogonal to the consensus manifold \mathcal{G}^* for the (a) purely deterministic system (4.1), the (b) stochastic system (4.2) with $\sigma = 1$ (satisfying the stability condition $2\lambda > \sigma^2$), and the (c) stochastic system (4.2) with $\sigma = \sqrt{2.1}$ (violating the stability condition).

References

- [1] L. ARNOLD, H. CRAUEL, AND V. WIHSTUTZ, *Stabilization of linear systems by noise*, SIAM Journal on Control and Optimization, 21 (1983), pp. 451–461.
- [2] Z. BAYRAKTAR, M. KOMURCU, AND D. H. WERNER, *The wind driven optimization technique and its application in electromagnetics*, IEEE Transactions on Antennas and Propagation, 61 (2013), pp. 2745–2757.
- [3] C. BLUM AND A. ROLI, *Metaheuristics in combinatorial optimization: Overview and conceptual comparison*, ACM Comput. Surv., 35 (2001), pp. 268–308.
- [4] G. BORGHI AND M. HERTY, *Model predictive control strategies using consensus-based optimization*, 2023.
- [5] G. BORGHI, M. HERTY, AND L. PARESCHI, *An adaptive consensus based method for multi-objective optimization with uniform pareto front approximation*, Applied Mathematics & Optimization, 88 (2023), p. 58.
- [6] —, *Constrained consensus-based optimization*, SIAM Journal on Optimization, 33 (2023), pp. 211–236.
- [7] J. A. CARRILLO, Y.-P. CHOI, C. TOTZECK, AND O. TSE, *An analytical framework for consensus-based global optimization method*, Mathematical Models and Methods in Applied Sciences, 28 (2018), pp. 1037–1066.
- [8] J. CHEN, S. JIN, AND L. LYU, *A consensus-based global optimization method with adaptive momentum estimation*, 2020.
- [9] —, *A consensus-based global optimization method with adaptive momentum estimation*, 2020.
- [10] A. DEMBO AND O. ZEITOUNI, *Large deviations techniques and applications*, vol. 38, Springer Science & Business Media, 2009.
- [11] M. DORIGO AND C. BLUM, *Ant colony optimization theory: A survey*, Theoretical Computer Science, 344 (2005), pp. 243–278.
- [12] M. FORNASIER, T. KLOCK, AND K. RIEDL, *Convergence of Anisotropic Consensus-Based Optimization in Mean-Field Law*, Springer International Publishing, 2022, p. 738–754.
- [13] —, *Consensus-based optimization methods converge globally*, SIAM Journal on Optimization, 34 (2024), pp. 2973–3004.

- [14] M. FORNASIER, L. PARESCHI, H. HUANG, AND P. SÜNNEN, *Consensus-based optimization on the sphere: Convergence to global minimizers and machine learning*, Journal of Machine Learning Research, 22 (2021), pp. 1–55.
- [15] S.-Y. HA, S. JIN, AND D. KIM, *Convergence of a first-order consensus-based global optimization algorithm*, Mathematical Models and Methods in Applied Sciences, 30 (2020), pp. 2417–2444.
- [16] S.-Y. HA, S. JIN, AND D. KIM, *Convergence and error estimates for time-discrete consensus-based optimization algorithms*, Numerische Mathematik, 147 (2021), pp. 255–282.
- [17] J. HEIECK, *Repository for exponential stability of finite-n consensus-based optimization*. <https://github.com/jaheieck/Exponential-Stability-of-Finite-N-Consensus-Based-Optimization>, 2025. Accessed: 2025-10-22.
- [18] D. J. HIGHAM, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, 43 (2001), pp. 525–546.
- [19] D. J. HIGHAM, X. MAO, AND C. YUAN, *Almost sure and moment exponential stability in the numerical simulation of stochastic differential equations*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 592–609.
- [20] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 2nd ed., 2012.
- [21] J. KENNEDY AND R. EBERHART, *Particle swarm optimization*, in Proceedings of ICNN’95 - International Conference on Neural Networks, vol. 4, 1995, pp. 1942–1948 vol.4.
- [22] R. KHASMINSKII, *Stochastic stability of differential equations. With contributions by G. N. Milstein and M. B. Nevelson*, vol. 66 of Stoch. Model. Appl. Probab., Berlin: Springer, 2nd completely revised and enlarged ed., 2012.
- [23] D. C. LUVIZON, H. TABIA, AND D. PICARD, *Consensus-based optimization for 3d human pose estimation in camera coordinates*, 2021.
- [24] X. MAO, *Stochastic stabilization and destabilization*, Systems and Control Letters, 23 (1994), pp. 279–290.
- [25] ———, *Stochastic Differential Equations and Applications*, Horwood Publishing, Chichester, UK, 2 ed., 2007.
- [26] C. D. MEYER, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.
- [27] R. M. MURRAY, S. S. SASTRY, AND L. ZEXIANG, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Inc., USA, 1st ed., 1994.
- [28] E. PANTELEY, A. LORÍA, AND S. SUKUMAR, *Strict lyapunov functions for consensus under directed connected graphs*, in 2020 European Control Conference (ECC), 2020, pp. 1–6.
- [29] R. PINNAU, C. TOTZECK, O. TSE, AND S. MARTIN, *A consensus-based model for global optimization and its mean-field limit*, Mathematical Models and Methods in Applied Sciences, 27 (2017), p. 183–204.
- [30] L. A. RASTRIGIN, *The convergence of the random search method in the extremal control of a many-parameter system*, Automation and Remote Control, 24 (1963), pp. 1337–1342.
- [31] J.-J. E. SLOITINE AND W. LI, *Applied Nonlinear Control*, Prentice Hall, 1991.
- [32] F. J. SOLIS AND R. J.-B. WETS, *Minimization by random search techniques*, Mathematics of Operations Research, 6 (1981), pp. 19–30.
- [33] S. SURJANOVIC AND D. BINGHAM, *Virtual library of simulation experiments: Test functions and datasets*. Retrieved May 26, 2025, from <http://www.sfu.ca/~ssurjano>.
- [34] C. TOTZECK, *Trends in Consensus-Based Optimization*, Springer International Publishing, Cham, 2022, pp. 201–226.

- [35] I. TUKH AND K. RIEDL, *cbo-in-python*. <https://github.com/Igor-Tukh/cbo-in-python>, 2022.
- [36] D. WEBER, R. THEISEN, AND S. MOTSCH, *Deterministic versus stochastic consensus dynamics on graphs*, Journal of Statistical Physics, 176 (2019), pp. 40–68.
- [37] H. ZHANG, Z. LI, Z. QU, AND F. L. LEWIS, *On constructing lyapunov functions for multi-agent systems*, Automatica, 58 (2015), pp. 39–42.