Analysis on the facet of particle interaction in particle swarm optimization

Ying-ping Chen
Pei Jiang

Natural Computing Laboratory (NCLab)
Department of Computer Science
National Chiao Tung University
329 Engineering Building C
1001 Ta Hsueh Road
HsinChu City 300, TAIWAN
http://nclab.tw/

Analysis on the facet of particle interaction in particle swarm optimization

Ying-ping Chen and Pei Jiang
Department of Computer Science
National Chiao Tung University
HsinChu City 300, Taiwan
{ypchen, pjiang}@nclab.tw

September 15, 2008

Abstract

In this paper, we analyze the behavior of particle swarm optimization (PSO) on the facet of particle interaction. We firstly propose a statistical interpretation of PSO in order to capture the stochastic behavior of the entire swarm. Based on the statistical interpretation, we investigate the effect of particle interaction by focusing on the social-only model and derive the upper and lower bounds of the expected particle norm. Accordingly, the lower and upper bounds of the expected progress rate on the sphere function are also obtained. Furthermore, the sufficient and necessary condition for the swarm to converge is derived to demonstrate the PSO convergence caused by the effect of particle interaction.

1 Introduction

Particle swarm optimization (PSO), introduced by Kennedy and Eberhart [1] in 1995, was proposed according to an inspiration from the social behavior of insects or animals that the exchanging and sharing of information among a group of individuals benefits the evolution of the species. In PSO, the insects or animals are considered as *particles* flying through the multi-dimensional search space and searching for the optimal position. The movement of particles are affected by three factors: the inertia, personal experience (the cognitive part), and particle interaction (the social part). PSO has been empirically shown that it is a very useful optimization framework [2] for the easiness to implement and flexibility to use.

Although PSO has been widely applied in many research fields since it was proposed, the theoretical analysis on PSO is still quite limited. To the best of our knowledge, the first analysis was proposed by Kenndy [3]. Particle trajectories for design choices were shown. Ozcan and Mohan [4, 5] assumed fixed attractors and constant coefficients to demonstrate the particle trajectory as a sinusoidal wave. With similar assumptions, Maurice and Kennedy [6] simplified PSO to a deterministic dynamical system and analyzed its stability. Such simplified, deterministic versions of PSO or similar systems, employing a single particle, fixed attractors, or constant coefficients, were analyzed by many researchers for stability, convergence, and parameter selection [7, 8, 9, 10, 11]. Kadirkamanathan, et. al. [12] and Jian, et. al. [13] started to consider the randomness in acceleration coefficients, but attractors were still fixed. Away from the normal PSO configuration, Emara and Fattah [14] as well as Gazi and Passino [15] analyzed PSO in a continuous time setting.

Most of the existing studies do not provide analysis on the facet of particle interaction, which is definitely an essential mechanism of PSO. In this paper, under more practical assumptions,

including multiple particles, unfixed attractors, and stochastic acceleration coefficients, we make the first attempt to analyze the effect of particle interaction. In particular, we consider the PSO system from a macrostate viewpoint, analyze the swarm behavior, and obtain theoretical results on the progress rate as well as the convergence criterion.

The paper is organized as follows. In section 2, we will describe the particle swarm optimization algorithm and propose the statistical interpretation. In section 3, we will analyze the mean positions of particles by considering the effect of particle interaction and derive the expected progress rate of the swarm on the sphere function. Next, we will look into the variance of the particle positions and show that the swarm will converge under certain condition in section 4. Finally, section 5 summarizes and concludes this paper.

2   PSO and Particle Interaction

In this section, we will firstly describe the standard PSO algorithm and then discuss the operations of PSO step by step, followed by the proposal of the statistical interpretation.

2.1   The Standard PSO algorithm

First of all, for easily making an abstraction of PSO based on statistics and probabilistic distributions, we restate the standard PSO system as the following algorithm:

Algorithm 1 (Standard PSO)**.**
  procedure STANDARD PSO(Objective function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$)
      Initialize a swarm of m particles
      while the stopping criterion is not satisfied **do**
         Evaluate each particle
         for particle i, $i = 1, 2, \ldots, m$ **do**   ▷ Update the best positions
            if $\mathcal{F}(\mathbf{X_i}) < \mathcal{F}(\mathbf{Pb_i})$ **then**
               $\mathbf{Pb_i} \leftarrow \mathbf{X_i}$
               if $\mathcal{F}(\mathbf{Pb_i}) < \mathcal{F}(\mathbf{Nb})$ **then**
                   $\mathbf{Nb} \leftarrow \mathbf{Pb_i}$
               end if
            end if
         end for
         for particle i, $i = 1, 2, \ldots, m$ **do**   ▷ Generate the next generation
            $\mathbf{V_i(t+1)} \leftarrow w\mathbf{V_i(t)} + C_p(\mathbf{Pb_i} - \mathbf{X_i}) + C_n(\mathbf{Nb} - \mathbf{X_i})$
            $\mathbf{X_i(t+1)} \leftarrow \mathbf{X_i(t)} + \mathbf{V_i(t+1)}$
         end for
      end while
  end procedure

Throughout this paper, boldface is used to distinguish vectors from scalars, and $\|\cdot\|$ denotes the L^2 norm of a vector. According to Algorithm 1, we can see that a standard PSO system comprises two main operations regarding information sharing and utilizing:

1. Updating attractors: Update the personal best position, $\mathbf{Pb_i}$, found by each particle, and the neighborhood best position, $\mathbf{Nb}$, found by any member within the neighborhood.

2. Updating particles: Update the velocities at time t by using a linear combination of the inertia, $\mathbf{V_i(t)}$, and the gravitation from the cognitive part, $\mathbf{Pb_i}$, and the social part, $\mathbf{Nb}$, respectively. w is the weight for the inertia and is usually a constant. C_p and C_n are

random variables sampled from uniform distribution $U(0, 2c_p)$ and $U(0, 2c_n)$ with $c_p > 0$ and $c_n > 0$ as acceleration coefficients. The position is then assigned according to the current position with application of the updated velocity.

As we can observed, the inherent characteristics of PSO – the interactions among particles – are implemented with the shared knowledge on the best position found by neighbors. When a particle within the neighborhood locates a position of an objective value which is better than $\mathcal{F}(\mathbf{Nb})$, the other particles will make corresponding adjustments and tend to go toward that position. Therefore, the neighborhood attractor can be viewed as a channel through which each particle can emulate the others, and the update of the neighborhood attractor can be considered as a signal urging the swarm to adjust their movements in order to respond to the new discovery in the search space.

2.2   A Macroscopic View on PSO

In spite of the importance, the effects of particle interaction in PSO are hardly investigated in the literature. Although there are a number of remarkable theoretical studies that bring insights into the properties and behaviors of PSO conducted in the past, most of those studies are based on the assumption that the attractor is fixed. It seems to be an inevitable path to simplify the PSO system to the extent that rigorous analysis can be done because the highly decentralized property of a particle swarm leads the system away from a unified depiction of the entire swarm. Each particle keeps its own position and memory, in the form of the inertia and the cognitive part, $\mathbf{Pb_i}$. In addition to the personal experience, the swarm also shares collective knowledge, $\mathbf{Nb}$, and any slight change in these quantities substantially defines a new state for the system. The analysis on the overall behavior of the swarm is thus beyond tractable due to the complication of state transition, and the simplification of invariant attractors becomes an unpleasant but necessary means that makes a particle able to be observed independently without the interference from the other factors of the entire swarm.

As a consequence, in order to take particle interaction into consideration in the theoretical analysis, an alternative interpretation of PSO that regards the swarm as a unity becomes necessary. With this point of view, the state of a PSO system should be considered as a measurement that reflects the overall behavior and characteristics of the swarm rather than as a detailed configuration directly related to each individual particle. For this purpose, the development of statistical mechanics may be a good example to learn from, especially the employment of statistical methods to bridge the macroscopic and microscopic descriptions. Accordingly, the state of the entire swarm can be considered as the "macrostate" – an abstraction from the detailed description of particles, i.e., the "microstate". Hence in the macrostate space, the precise configuration of particles are converted into a statistical abstraction and characterization toward the entire swarm.

More specifically, the exact locations of particles are not traced but modeled and expressed with a distribution $\theta(t)$ over $\mathbb{R}^n$, and the velocities on each dimension are viewed as a random vector $\mathcal{V}(t) \in \mathbb{R}^n$. To concentrate on the social behavior, i.e., the facet of particle interaction, we use the *social-only model* of PSO categorized by Kennedy [16], in which PSO works without the cognitive component, to make the system more concise. The swarm size m is considered as the number of realizations or samples of the distribution. As to the neighborhood attractor, since the geographic knowledge about the search space is embodied in the positional distribution, it can be viewed as the best observed value of the current time step. When the neighborhood attractor is determined, the social gravitation is also accordingly determined. Formally, each particle $\mathbf{P_i}$ is a random vector sampled from $\theta(t)$, and its velocity vector $\mathbf{V_i}$ is distributed as $\mathcal{V}(t)$. Since the neighborhood attractor is the best observed value, it can be defined as

$$\mathbf{P_a} := \arg\min\{\mathcal{F}(\mathbf{P_1}), \mathcal{F}(\mathbf{P_2}), \ldots, \mathcal{F}(\mathbf{P_m})\} \,,$$

and each particle $\mathbf{P_i}$ updates its position to $\mathbf{P_i} + w\mathbf{V_i} + C(\mathbf{P_a} - \mathbf{P_i})$. The distributions of the next time step $\theta(t + 1)$ and $\mathcal{V}(t + 1)$ are thus the statistical characterization, denoted as functions $\mathcal{T_P}$ and $\mathcal{T_V}$, of the observed values:

$$\theta(t + 1) \leftarrow \mathcal{T_P}(\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}) \, ;$$
$$\mathcal{V}(t + 1) \leftarrow \mathcal{T_V}(\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}; \mathbf{V_1}, \mathbf{V_2}, \ldots, \mathbf{V_m}) \, .$$

By considering the PSO system in this way, the search/optimization process is conducted through the repeated observations on the search space by realizing particles and modifying the distribution to accommodate the newly discovered results. Furthermore, going deeper into the notion of distribution, since the inertia weight w is usually a constant, $\mathcal{V}(t)$ can be considered redundant and may be removed because given two random vectors $\mathbf{X} \sim \theta(t)$ and $\mathbf{V} \sim \mathcal{V}(t)$, we can simply let $\widetilde{\theta}(t)$ be the distribution of $\mathbf{X}' := \mathbf{X} + w\mathbf{V}$ that includes the effects of both the position and the velocity.

The remaining questions would be what distribution is suitable for the description of the swarm without sacrificing too much essence of PSO and how to update the distribution as the search process proceeds. Consider the random vector $\mathbf{X} \sim \theta(t)$ and denote $\mathrm{E}\,[\mathbf{X}] = \boldsymbol{\mu}$. If we decompose the region

$$R := \{\mathbf{y} \in \mathbb{R}^n \mid \mathrm{Prob}\,\{\mathbf{X} = \mathbf{y}\} > 0\}$$

into s disjoint regions $R_1, R_2, \ldots, R_s$ such that $\mathrm{Prob}\,\{\mathbf{X} \in R_i\} = 1/s$ for all $i \in \{1, 2, \ldots, s\}$, and each region is associated with a random variable of velocity $\mathbf{V_i} \sim \mathcal{V}(t)$. Select a point $\mathbf{x_i}$ from each region R_i respectively; when s is sufficiently large, the average behavior of the swarm can thus be characterized by

$$\sum_{i=1}^{s} \frac{1}{s} (\mathbf{x_i} + \mathbf{V_i}) = \sum_{i=1}^{s} \frac{1}{s}\mathbf{x_i} + \sum_{i=1}^{s} \frac{1}{s}\mathbf{V_i}$$

$$\approx \boldsymbol{\mu} + \sum_{i=1}^{s} \frac{1}{s}\mathbf{V_i} \, ,$$

and each component of the term $\sum_{i=1}^{s}(1/s)\mathbf{V_i}$ can be approximated with a normal distribution by the central limit theorem. Therefore, as an attempt to characterize the overall behavior of a swarm, the normal distribution should be a reasonable starting point. It is assumed that, at time t, each particle is sampled from $\mathbf{c(t)} + \mathbf{Z}$, where $\mathbf{c(t)} \in \mathbb{R}^n$ is the center of distribution and $\mathbf{Z} \in \mathbb{R}^n$ is a random vector of which each coordinate is distributed according to $N(0, \sigma^2)$, where $N(0, \sigma^2)$ denotes the normal distribution with zero mean and variance σ^2. In this paper, $\phi(\cdot)$ and $\Phi(\cdot)$ are used as the probability density function (pdf) and the cumulative distribution function (cdf) of the standard normal distribution, respectively. The distribution of $\mathbf{c(t)} + \mathbf{Z}$ can then be expressed as $\theta(\mathbf{c(t)}, \sigma^2)$.

The update of distribution is now simplified into the modification of the mean and the variance. The mean is the arithmetic average of updated positions of particles, and the variance is estimated by a maximum likelihood estimation (MLE) which will be addressed later. Under such an interpretation, the PSO system can then be described with the following algorithm:

Algorithm 2 (Statistical interpretation of PSO)**.**
  procedure PSO(Objective function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$)
     Initialize θ
     while the stopping criterion is not satisfied **do**
       for $i = 1, 2, \ldots, m$ **do**
         $\mathbf{P_i} \sim \theta$

```
        end for
        P_a = arg min{F(P_1), F(P_2), ..., F(P_m)}
        for i = 1, 2, ..., m do
            P'_i ← P_i + C(P_a − P_i)
        end for
        μ_{t+1} ← (∑_{i=1}^m P'_i)/m
        σ²_{t+1} ← MLE(P'_1, P'_2, ..., P'_m)
        θ ← θ(μ_{t+1}, σ²_{t+1})
        t ← t + 1
    end while
end procedure
```

In the reminder of this paper, Algorithm 2 will be the study subject and be formally investigated on the sphere function, which is commonly adopted in the theoretical analysis of evolutionary algorithms (e.g., [17]) and can be formulated as

$$\mathcal{F}(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 \, ,$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

3 Progress Rate Analysis

The major benefit to develop and adopt the abstraction based on probabilistic distributions of the PSO system is that the mathematical model can be analyzed without the assumption of fixed attractors, because particles are in essence random vectors in the search space and consequently their behavior can be described and predicted in a statistical sense. In this section, we will demonstrate how the statistical interpretation of PSO proposed in this paper facilitates the analysis of inter-particle effects and how these effects are accounted for the progress rate of the entire swarm. We will begin with the n-ball hitting probability.

3.1 n-ball Hitting Probability

Given a distribution θ over $\mathbb{R}^n$, the term *n-ball hitting probability* refers to the probability that a random vector sampled from θ that "falls" into a specific n-dimensional ball. This probability is fundamental to the sphere model, because in the sphere model the objective function is simply the squared L^2 norm, and a subset of $\mathbb{R}^n$ constructed by collecting all the vectors with their norms bounded by a specific non-negative quantity forms an n-ball located at the origin with a radius defined by that non-negative quantity. Therefore, n-ball hitting probability is equal to the probability that the norm of a random vector is less than or equals to the radius; in other words, it is essentially the cumulative distribution function (cdf) of the norm of a random vector.

Given the center of distribution at time t, $\mathbf{c(t)} = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$, we would like to calculate the probability that $\mathbf{c(t)} + \mathbf{Z} \sim \theta$ is in an n-ball located at origin with radius k (denoted as $B_k(\mathbf{o})$), where $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n) \in \mathbb{R}^n$ is a random vector and each coordinate of $\mathbf{Z}$ is normally distributed. Since $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed (i.i.d.) random variables, $\mathbf{Z}$ is an isotropic random vector (i.e., all directions of $\mathbf{Z}$ are equally likely to occur) [18]. We elaborate this property as follows. Given $Z_1, Z_2, \ldots, Z_n \sim N(0, \sigma^2)$,

5

for all $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

$$
\begin{aligned}
\text{Prob}\{\mathbf{c(t)} + \mathbf{Z} = \mathbf{x}\} &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x_i - c_i)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-\sum_{i=1}^{n}(x_i - c_i)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-d(\mathbf{x}, \mathbf{c(t)})^2}{2\sigma^2}\right) ,
\end{aligned}
$$

where $d(\cdot, \cdot)$ is the Euclidean distance function. It is obvious that the density at point $\mathbf{x}$ is determined by $d(\mathbf{x}, \mathbf{c(t)})$, regardless of which direction $\mathbf{x}$ is in relatively to $\mathbf{c(t)}$. Therefore, without loss of generality, we may assume that $\mathbf{c(t)}$ is on the first axis by conducting a coordinate transformation. Let $r := d(\mathbf{c(t)}, \mathbf{o}) \geq 0$. As a result, $\mathbf{c(t)}$ can be expressed (after the coordinate transformation) as $(r, 0, 0, \ldots, 0)$, and the distribution is denoted as $\theta(r, \sigma^2)$. Now, the n-ball hitting probability can be formally defined as follows.

Definition 1. Given an n-ball $B_k(\mathbf{o}) \in \mathbb{R}^n$ and a random vector $\mathbf{c(t)} + \mathbf{Z} \sim \theta(r, \sigma^2) \in \mathbb{R}^n$, where $\mathbf{c(t)} = (r, 0, 0, \ldots, 0)$ and all coordinates of $\mathbf{Z}$ are distributed according to $N(0, \sigma^2)$, the n-ball hitting probability

$$
H_B(k, \theta(r, \sigma^2)) := \text{Prob}\{\mathbf{c(t)} + \mathbf{Z} \in B_k(\mathbf{o})\} .
$$

The analysis approach adopted in the present work is similar to that used by Beyer in 2001 [17]. The vector $\mathbf{Z}$ is decomposed into two orthogonal vectors: $Z_1\mathbf{e_1} = (Z_1, 0, 0, \ldots, 0)$ and $\mathbf{Z}' = (0, Z_2, Z_3, \ldots, Z_n)$. Let's take a closer look at the n-ball hitting probability $H_B(k, \theta(r, \sigma^2))$:

$$
\begin{aligned}
H_B(k, \theta(r, \sigma^2)) &= \text{Prob}\{\mathbf{c(t)} + \mathbf{Z} \in B_k(\mathbf{o})\} \\
&= \text{Prob}\{\|(r + Z_1)\mathbf{e_1} + \mathbf{Z}'\| \leq k\} \\
&= \text{Prob}\{(r + Z_1)^2 + \|\mathbf{Z}'\|^2 \leq k^2\} \\
&= \text{Prob}\{-k - r \leq Z_1 \leq k - r, \, 0 \leq \|\mathbf{Z}'\|^2 \leq k^2 - (r + Z_1)^2\}
\end{aligned}
$$

The above equation shows that the n-ball hitting probability is the joint distribution of Z_1 and $W := \|\mathbf{Z}'\|^2$. Since $Z_1 \sim N(0, \sigma^2)$, the probability density function can be expressed as

$$
p(Z_1, x) := \text{Prob}\{Z_1 = x\} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-x^2}{2\sigma^2}\right) ,
$$

and W is a chi-square random variable with $n' := n - 1$ degree of freedom:

$$
p(W, y) := \text{Prob}\{W = y\} = \frac{1}{\sigma^2} \frac{\left(\frac{y}{\sigma^2}\right)^{\frac{n'}{2} - 1} \exp\left(\frac{-y}{2\sigma^2}\right)}{2^{\frac{n'}{2}} \Gamma\left(\frac{n'}{2}\right)}
$$

As a result, we can get

$$
\begin{aligned}
H_B(k, \theta(r, \sigma^2)) &= \text{Prob}\left\{-k - r \leq Z_1 \leq k - r,\, 0 \leq \|\mathbf{Z}'\|^2 \leq k^2 - (r + Z_1)^2\right\} \\
&= \int_{x=-k-r}^{k-r} \int_{y=0}^{k^2-(x+r)^2} p(Z_1, x) p(W, y)\, dy\, dx \\
&= \int_{x=-k-r}^{k-r} p(Z_1, x) \int_{y=0}^{k^2-(x+r)^2} \frac{1}{\sigma^2} \frac{\left(\frac{y}{\sigma^2}\right)^{\frac{n'}{2}-1} \exp\left(\frac{-y}{2\sigma^2}\right)}{2^{\frac{n'}{2}} \Gamma\left(\frac{n'}{2}\right)}\, dy\, dx \\
(\text{let } u := y/\sigma^2) &= \int_{x=-k-r}^{k-r} p(Z_1, x) \int_{u=0}^{\frac{k^2-(x+r)^2}{\sigma^2}} \frac{u^{\frac{n'}{2}-1} \exp\left(\frac{-u}{2}\right)}{2^{\frac{n'}{2}} \Gamma\left(\frac{n'}{2}\right)}\, du\, dx \\
&= \int_{x=-k-r}^{k-r} p(Z_1, x) \mathcal{P}\left(\frac{n'}{2}, \frac{k^2 - (x+r)^2}{2\sigma^2}\right) dx \ ,
\end{aligned}
$$

where $\mathcal{P}(\cdot)$ is the regularized Gamma function.

Remark 2. If an asymptotic approximation is desired for the n-ball hitting probability, $H_B(k, \theta(r, \sigma^2))$, we can utilize the normal approximation to the regularized Gamma function ([19], chapter 7) as

$$
\mathcal{P}\left(\frac{n'}{2}, \frac{k^2 - (x+r)^2}{2\sigma^2}\right) \approx \Phi\left(\frac{1}{\sqrt{2n'}}\left[\frac{k^2 - (x+r)^2}{\sigma^2} - n'\right]\right) \ .
$$

For the asymptotic approximation, when n is sufficiently large, the term $(1/\sqrt{2n'})[k^2 - (x + r)^2]/\sigma^2$ vanishes, and thanks to the continuity of $\Phi(\cdot)$, we can obtain

$$
\mathcal{P}\left(\frac{n'}{2}, \frac{k^2 - (x+r)^2}{2\sigma^2}\right) \approx \Phi\left(-\sqrt{\frac{n'}{2}}\right)
$$

Hence,

$$
\begin{aligned}
H_B(k, \theta(r, \sigma^2)) &\approx \Phi\left(-\sqrt{\frac{n'}{2}}\right) \int_{x=-k-r}^{k-r} p(Z_1, x)\, dx \\
&= \Phi\left(-\sqrt{\frac{n'}{2}}\right)\left[\Phi\left(\frac{k-r}{\sigma}\right) - \Phi\left(\frac{-k-r}{\sigma}\right)\right] \\
&= \Phi\left(-\sqrt{\frac{n'}{2}}\right)\left[\Phi\left(\frac{r+k}{\sigma}\right) - \Phi\left(\frac{r-k}{\sigma}\right)\right] \ .
\end{aligned}
$$

In addition to the asymptotic properties of $H_B(k, \theta(r, \sigma^2))$, it would be helpful to derive a lower bound for $H_B(k, \theta(r, \sigma^2))$ to facilitate our analysis in the present work.

Lemma 3 (Lower bound for $H_B(k, \theta(r, \sigma^2))$).

$$
H_B(k, \theta(r, \sigma^2)) \geq \left[\Phi\left(\frac{r + \frac{k}{\sqrt{n}}}{\sigma}\right) - \Phi\left(\frac{r - \frac{k}{\sqrt{n}}}{\sigma}\right)\right]\left[1 - 2\Phi\left(\frac{-k}{\sqrt{n}\sigma}\right)\right]^{n-1} \ .
$$

Proof. Let $\mathbf{Y} := \mathbf{c(t)} + \mathbf{Z}$, where $\mathbf{c(t)} = (r, 0, 0, \ldots, 0)$, and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$. Let $\mathcal{D} := [-k/\sqrt{n}, k/\sqrt{n}]^n \subseteq R^n$. For all $\mathbf{x} \in \mathcal{D}$, because $\|\mathbf{x}\| \leq \sqrt{n}\|\mathbf{x}\|_\infty \leq \sqrt{n}(k/\sqrt{n}) = k$, we can know

that $\mathbf{x} \in B_k(\mathbf{o})$. Hence, $\mathcal{D} \subseteq B_k(\mathbf{o})$, and

$$\text{Prob}\{\mathbf{Y} \in B_k(\mathbf{o})\} \geq \text{Prob}\{\mathbf{Y} \in \mathcal{D}\}$$

$$= \text{Prob}\left\{-\frac{k}{\sqrt{n}} - r \leq Z_1 \leq \frac{k}{\sqrt{n}} - r\right\} \prod_{i=2}^{n} \text{Prob}\left\{-\frac{k}{\sqrt{n}} \leq Z_i \leq \frac{k}{\sqrt{n}}\right\}$$

$$= \left[\Phi\left(\frac{\frac{k}{\sqrt{n}} - r}{\sigma}\right) - \Phi\left(\frac{-\frac{k}{\sqrt{n}} - r}{\sigma}\right)\right]\left[\Phi\left(\frac{\frac{k}{\sqrt{n}}}{\sigma}\right) - \Phi\left(\frac{-\frac{k}{\sqrt{n}}}{\sigma}\right)\right]^{n-1}$$

$$= \left[\Phi\left(\frac{r + \frac{k}{\sqrt{n}}}{\sigma}\right) - \Phi\left(\frac{r - \frac{k}{\sqrt{n}}}{\sigma}\right)\right]\left[1 - 2\Phi\left(\frac{-k}{\sqrt{n}\sigma}\right)\right]^{n-1} .$$

$\square$

For the notational purpose, we let

$$\psi'(k) := \left[\Phi\left(\frac{r + \frac{k}{\sqrt{n}}}{\sigma}\right) - \Phi\left(\frac{r - \frac{k}{\sqrt{n}}}{\sigma}\right)\right]\left[1 - 2\Phi\left(\frac{-k}{\sqrt{n}\sigma}\right)\right]^{n-1} ,$$

and its antiderivative is defined as $\psi(k) := \int_{t=0}^{k} \psi'(t)\, dt$.

Remark 4. Similarly, we can also define the n-sphere hitting density $H_S(k, \theta(r, \sigma^2))$ for random vector $\mathbf{c(t)} + \mathbf{Z}$ as

$$H_S(k, \theta(r, \sigma^2)) := \text{Prob}\{\|\mathbf{c(t)} + \mathbf{Z}\| = k\}$$

$$= \text{Prob}\{-k - r \leq Z_1 \leq k - r, W = k^2 - x^2\}$$

$$= \int_{x=-k-r}^{k-r} p(Z_1, x)p(W, k^2 - x^2)\, dx .$$

Therefore, the n-ball hitting probability, $H_B(k, \theta(r, \sigma^2))$, as the cumulative function of $H_S(k, \theta(r, \sigma^2))$, can be alternatively defined as

$$\int_{y=0}^{k} \int_{x=-y-r}^{y-r} p(Z_1, x)p(W, y^2 - x^2)\, dx\, dy .$$

However, the density function $H_S(k, \theta(r, \sigma^2))$ serves no purpose other than a definition in the following analysis. We left it as a side note for completeness without further discussion.

3.2 Expected Particle Norm

The entire PSO system can be decomposed into two fundamental components: (1) the update of attractors to share and exchange information among particles, and (2) the update of particle positions through the interaction between particles and attractors. So, as we gain understanding of the characteristics of attractors and particles, we may capture the stochastic behavior of the PSO system. More specifically, because distance is the most important characteristic of the sphere model for its unimodality, in this section, we highlight the expected distance between particles and the global optimum. Given a probabilistic model according to which particles are distributed, we would like to know how close to the global optimum in expectation the sampled particles are. Since the global optimum is simply the origin in the sphere model, we concentrate on the L^2-norm of sampled particles. The expected norms of the attractor and of particles are examined, respectively. As the analysis proceeds, it can be shown that these two expectations influence the progress rate of the particle swarm.

Given the center of a particle distribution $\mathbf{c(t)} = (r, 0, \ldots, 0)$ and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ with $Z_1, Z_2, \ldots, Z_n \sim N(0, \sigma^2)$, suppose that there are m particles, $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}$, sampled as $\mathbf{c(t)} + \mathbf{Z}$, the expected norm of particles can be defined as

$$\overline{P} := \mathrm{E}\left[\|\mathbf{c(t)} + \mathbf{Z}\|\right] \,,$$

which can be considered as the mean solution quality of the current swarm on the sphere function. The following lemma gives an upper bound for $\overline{P}$.

Lemma 5 (Upper bound for the expected particle norm). *If* $\mathbf{c(t)} = (r, 0, 0, \ldots, 0)$ *and* $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ *with* $Z_1, Z_2, \ldots, Z_n \sim N(0, \sigma^2)$, $\overline{P} \le \sqrt{r^2 + n\sigma^2}$.

Proof. For all positive random variable X, since the square root is a concave function, we have $\mathrm{E}\left[\sqrt{X}\right] \le \sqrt{\mathrm{E}[X]}$ according to Jensen's inequality. By utilizing this property, we can have the following derivation:

$$\overline{P} = \mathrm{E}\left[\|\mathbf{c(t)} + \mathbf{Z}\|\right]$$

$$= \mathrm{E}\left[\sqrt{(Z_1 - r)^2 + \sum_{i=2}^{n} Z_i^2}\right]$$

$$\le \sqrt{\mathrm{E}\left[(Z_1 - r)^2 + \sum_{i=2}^{n} Z_i^2\right]}$$

$$= \sqrt{\left(\mathrm{E}[r^2] - 2r\mathrm{E}[Z_1] + \sum_{i=1}^{n} \mathrm{E}[Z_i^2]\right)}$$

$$= \sqrt{r^2 + n\sigma^2} \,,$$

Because $Z_i \sim N(0, \sigma^2)$, we have $\mathrm{E}[Z_i^2] = \sigma^2$ and $\mathrm{E}[Z_i] = 0$. An upper bound for the expected particle norm, $\overline{P}$, is therefore obtained. $\qquad\square$

The expected particle norm describes how close on average a swarm is to the global optimum of the sphere function. In order to capture the characteristic of the essential mechanism of PSO – particle interaction – we also need to investigate the attractor. As stated in the previous section, the attractor is the best observed value, i.e., in our case, the particle with the minimum objective value within the neighborhood in the current swarm. Under the adopted statistical interpretation of PSO, the expected minimum objective value of a swarm becomes tractable through order statistics, because particles are viewed as random vectors over $\mathbb{R}^n$.

Let $P_{(i,m)}$ denote the ith order statistic of $\|\mathbf{P_1}\|, \|\mathbf{P_2}\|, \ldots, \|\mathbf{P_m}\|$, e.g., $P_{(1,m)} = \min\{\|\mathbf{P_1}\|, \|\mathbf{P_2}\|, \ldots, \|\mathbf{P_m}\|\}$. Denoting the event $\|\mathbf{P_i}\| = x$ as $\{\|\mathbf{P_i}\| = x\}$, the density of $P_{(1,m)}$ at a non-negative real number x can be given as

$$\mathrm{Prob}\left\{P_{(1,m)} = x\right\}$$

$$= \mathrm{Prob}\left\{\bigcup_{i=1}^{m}\left[\{\|\mathbf{P_i}\| = x\} \bigcap \left(\bigcap_{j \in \{1,2,\ldots,m\}\setminus\{i\}} \{\|\mathbf{P_j}\| > x\}\right)\right]\right\}$$

$$= \int_{x=-k-r}^{k-r} \binom{m}{1} H_S(k, \theta(r, \sigma^2)) \left[1 - H_B(k, \theta(r, \sigma^2))\right]^m dx \,.$$

Denoting $\mathrm{E}\left[P_{(1,m)}\right]$ as $\overline{P_{(1,m)}}$, a naive upper bound for $\overline{P_{(1,m)}}$ is derived in the following lemma.

Lemma 6. $\overline{P_{(1,m)}} \le \overline{P}$

Proof. The general upper bound for the expected ith order statistic states

$$\overline{P_{(i,m)}} \le \overline{P} + (\mathrm{Var}\,[\|\mathbf{c(t)} + \mathbf{Z}\|])^{\frac{1}{2}} \sqrt{\frac{i-1}{m-i+1}} \;.$$

As a result,

$$\overline{P_{(1,m)}} \le \overline{P} + (\mathrm{Var}\,[\|\mathbf{c(t)} + \mathbf{Z}\|])^{\frac{1}{2}} \sqrt{\frac{1-1}{m-1+1}} = \overline{P}\;.$$

$\square$

The above lemma does not cause any surprise. The expected minimum particle norm is obviously less than or equal to the expected norm. However, inspired by Lemma 6, we can seek another upper bound for $\overline{P_{(1,m)}}$ by definition.

Lemma 7 (Upper bound for $\overline{P_{(1,m)}}$). *(1)*

$$\overline{P_{(1,m)}} = \int_{x=0}^{\infty} \left[1 - H_B(k, \theta(r, \sigma^2))\right]^m \, dx\;,$$

and (2)

$$\overline{P_{(1,m)}} \le \left(\lim_{h\to\infty} [h - \psi(h)]\right)^{\frac{m}{2}}\;.$$

Proof. (1) For any random variable X, $\mathrm{E}\,[|X|]^r = r \int_0^{\infty} t^{r-1}\mathrm{Prob}\,\{|X| > t\}\, dt$ with $r > 0$ [20]. Since $P_{(1,m)}$ is a non-negative random variable, by letting $r = 1$ we have

$$\begin{aligned}
\overline{P_{(1,m)}} &= \int_{x=0}^{\infty} \mathrm{Prob}\,\{P_{(1,m)} > x\}\, dx \\
&= \int_{x=0}^{\infty} \mathrm{Prob}\,\left\{\bigcap_{i=1}^{m} \{\|\mathbf{P_i}\| > x\}\right\}\, dx \\
&= \int_{x=0}^{\infty} \left[1 - H_B(k, \theta(r, \sigma^2))\right]^m \, dx\;.
\end{aligned}$$

(2) Based on the result of (1), we obtain

$$\overline{P_{(1,m)}} = \int_{x=0}^{\infty} \left[1 - H_B(k, \theta(r, \sigma^2))\right]^m \, dx \le \int_{x=0}^{\infty} \left[1 - \psi'(x)\right]^m \, dx\;.$$

By resorting to Hölder's inequality, we can move m outside of the integration to obtain a more comprehensible bound as

$$\begin{aligned}
\int_{x=0}^{\infty} \left[1 - \psi'(x)\right]^m \, dx &\le \left(\int_{x=0}^{\infty} \left[1 - \psi'(x)\right]^2 \, dx\right)^{\frac{m}{2}} \\
&\le \left(\int_{x=0}^{\infty} \left[1 - \psi'(x)\right] \, dx\right)^{\frac{m}{2}} \\
&= \left(\lim_{h\to\infty} [h - \psi(h)]\right)^{\frac{m}{2}}
\end{aligned}$$

The last equation follows from $[h - \psi(h)]|_{h=0} = 0$.

$\square$

Because this upper bound is presented in a limit form, a subsequent question would be whether or not it converges. The following theorem guarantees the convergence of the quantity.

Lemma 8. $(\lim_{h\to\infty}[h - \psi(h)])^{\frac{m}{2}}$ *is convergent.*

Proof. Denote $\int_{x=0}^{h} [1 - \psi'(h)]\, dx$ as $G(h)$. Since m is a constant, $(\lim_{h\to\infty}[h-\psi(h)])^{\frac{m}{2}}$ converges if $\lim_{h\to\infty} G(h)$ converges. $G(h)$ is incremental because $1 - \psi'(x)$ is always non-negative. Thus, it is sufficient to show that $G(h)$ is bounded from above. When $h > r\sqrt{n}$,

$$
\begin{aligned}
G(h) &= \int_{x=0}^{h} \left[1 - \psi'(h)\right]\, dx \\
&= \int_{x=0}^{h} \left(1 - \left[\Phi\left(\frac{r + \frac{x}{\sqrt{n}}}{\sigma}\right) - \Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]\left[1 - 2\Phi\left(\frac{-x}{\sqrt{n}\sigma}\right)\right]^{n-1}\right) dx \\
&\leq \int_{x=0}^{r\sqrt{n}} dx + \int_{x=r\sqrt{n}}^{h} \left(1 - \left[\Phi\left(\frac{r + \frac{x}{\sqrt{n}}}{\sigma}\right) - \Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]\left[1 - 2\Phi\left(\frac{-x}{\sqrt{n}\sigma}\right)\right]^{n-1}\right) dx \\
&\leq r\sqrt{n} + \int_{x=r\sqrt{n}}^{h} \left(1 - \left[\Phi\left(\frac{\frac{x}{\sqrt{n}} - r}{\sigma}\right) - \Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]\left[1 - 2\Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]^{n-1}\right) dx \\
&= r\sqrt{n} + \int_{x=r\sqrt{n}}^{h} \left(1 - \left[1 - 2\Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]^{n}\right) dx\, .
\end{aligned}
$$

When $x \geq r\sqrt{n}$,

$$
\Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right) \leq \frac{1}{2}\, .
$$

Applying Bernoulli's inequality, it yields

$$
\begin{aligned}
G(h) &\leq r\sqrt{n} + \int_{x=r\sqrt{n}}^{h} \left(1 - \left[1 - 2n\Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]\right) dx \\
&= r\sqrt{n} + 2n \int_{x=r\sqrt{n}}^{h} \Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right) dx \\
&= r\sqrt{n} + 2n \left[\left(-r\sqrt{n} + x\right)\Phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right) - \sigma\sqrt{n}\cdot\phi\left(\frac{r - \frac{x}{\sqrt{n}}}{\sigma}\right)\right]\Bigg|_{x=r\sqrt{n}}^{x=h}\, .
\end{aligned}
$$

The integration of the normal distribution is given in [21]. When $h \to \infty$, the term

$$
\sigma\sqrt{n}\cdot\phi\left(\frac{r - \frac{h}{\sqrt{n}}}{\sigma}\right)
$$

vanishes, so now we only need to show

$$
\lim_{h\to\infty}\left[\left(-r\sqrt{n} + h\right)\Phi\left(\frac{r - \frac{h}{\sqrt{n}}}{\sigma}\right)\right] < \infty\, .
$$

Here we apply Mill's ratio to replace $\Phi(\cdot)$ with $\phi(\cdot)$ and get

$$\left(-r\sqrt{n}+h\right)\Phi\left(\frac{r-\frac{h}{\sqrt{n}}}{\sigma}\right) = \left(h-r\sqrt{n}\right)\left[1-\Phi\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)\right]$$

$$\leq \left(h-r\sqrt{n}\right)\cdot\phi\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)\cdot\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)^{-1}$$

$$= \left(\sigma\sqrt{n}\right)\cdot\phi\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)$$

$$= 0, \text{ as } h\to\infty \ .$$

Therefore, $G(h)$ is bounded from above. The proof is completed. $\qquad\square$

3.3 Lower and Upper Bounds for the Expected Progress Rate

After the preparatory work was done in the previous sections, the progress rate the social-only model PSO can now be formally investigated under a statistical interpretation. The term "progress rate" was introduced by Rechenberg in 1973 [22]. As its name suggests, progress rate should be a quantity indicating how a particle swarm progresses, and hence in the present work, it is defined as the difference of the norms of two centers of distributions in successive time steps, because the distance to the optimum is the L^2 norm in the sphere function. Given the current center of distribution $\mathbf{c(t)} = (r,0,0,\ldots,0)$ and a random vector $\mathbf{Z} = (Z_1,Z_2,\ldots,Z_n)$ with $Z_1,Z_2,\ldots,Z_n \sim N(0,\sigma^2$, the m particles $\mathbf{P_1},\mathbf{P_2},\ldots,\mathbf{P_m}$ are sampled as $\mathbf{c(t)}+\mathbf{Z}$. Let $P_{(i,m)}$ denote the ith order statistic of $\|\mathbf{P_1}\|$, $\|\mathbf{P_2}\|$, ..., $\|\mathbf{P_m}\|$. Let $\mathbf{P_a} := \arg\min\{\mathcal{F}(\mathbf{P_1}),\mathcal{F}(\mathbf{P_2}),\ldots,\mathcal{F}(\mathbf{P_m})\}$. By definition, $\|\mathbf{P_a}\| = P_{(1,m)}$. According to the update rules described in section 2.2, the updated position $\mathbf{P'_i}$ is computed as $\mathbf{P'_i} = \mathbf{P_i}+C(\mathbf{P_a}-\mathbf{P_i})$, where $C \sim U(0,2c)$ with c being the coefficient representing the compound effect of both the inertia weight and the acceleration coefficient for the social part. For simplicity, we still call c the acceleration coefficient in this paper because the inertia weight is usually constant. The center of distribution in the next step $\mathbf{c(t+1)}$ is the mean of $\mathbf{P'_1},\mathbf{P'_2},\ldots,\mathbf{P'_m}$, i.e., $\mathbf{c(t+1)} = (\sum_{i=1}^{m}\mathbf{P'_i})/m$.

Definition 9. Given $\mathbf{c(t)} = (r,0,0,\ldots,0)$, the progress rate $\Delta_t := \|\mathbf{c(t+1)}\| - \|\mathbf{c(t)}\| = \|\mathbf{c(t+1)}\| - r$.

The following theorem shows that, when $c \leq 1/2$, the expected norm of the center of distribution in the next time step is bounded from above by a linear combination of the expected particle norm $\overline{P}$ and the expected minimum of the particle norm $\overline{P_{(1,m)}}$.

Theorem 10 (Upper bound for the expected norm of the next center). *(1)* $E[\|\mathbf{c(t+1)}\|] \leq E[|1-C|]\overline{P} + E[|C|]\overline{P_{(1,m)}}$; and *(2)* If $c \leq 1/2$, $E[\|\mathbf{c(t+1)}\|] \leq (1-c)\overline{P} + c\overline{P_{(1,m)}}$; otherwise, $E[\|\mathbf{c(t+1)}\|] \leq [(2c^2-2c+1)/2c]\overline{P} + c\overline{P_{(1,m)}}$.

Proof. (1) This result is derived from the the triangle inequality for L^2-norm:

$$\mathrm{E}\left[\|\mathbf{c(t+1)}\|\right] = \mathrm{E}\left[\|\frac{\sum_{i=1}^{m}\left[\mathbf{P_i} + C\left(\mathbf{P_a} - \mathbf{P_i}\right)\right]}{m}\|\right]$$

$$= \left(\frac{1}{m}\right)\mathrm{E}\left[\|\sum_{i=1}^{m}\left(1 - C\right)\mathbf{P_i} + mC\mathbf{P_a}\|\right]$$

$$\leq \left(\frac{1}{m}\right)\left(\sum_{i=1}^{m}\mathrm{E}\left[\|\left(1 - C\right)\mathbf{P_i}\|\right] + m\mathrm{E}\left[\|C\mathbf{P_a}\|\right]\right)$$

$$= \left(\frac{1}{m}\right)\left(m\mathrm{E}\left[|1 - C|\right]\overline{P} + m\mathrm{E}\left[|C|\right]\overline{P_{(1,m)}}\right)$$

$$= \mathrm{E}\left[|1 - C|\right]\overline{P} + \mathrm{E}\left[|C|\right]\overline{P_{(1,m)}}.$$

(2) When $c \leq 1/2$, $1 - C$ is a non-negative random variable. So, we have

$$\mathrm{E}\left[|1 - C|\right] = \mathrm{E}\left[1 - C\right] = 1 - c.$$

On the other hand, when $c > 1/2$,

$$\mathrm{E}\left[|1 - C|\right]$$
$$= \mathrm{Prob}\left\{1 - C \geq 0\right\}\mathrm{E}\left[1 - C|1 - C \geq 0\right]$$
$$+ \mathrm{Prob}\left\{1 - C < 0\right\}\mathrm{E}\left[C - 1|1 - C < 0\right]$$
$$= (1/2c)(1/2) + [(2c - 1)/2c][(2c - 1)/2]$$
$$= [(2c^2 - 2c + 1)/2c].$$

$\square$

Corollary 11 (Lower bound for the progress rate). *When $c \leq 1/2$, $E[\Delta_t] \geq r - (1-c)\overline{P} - c\overline{P_{(1,m)}}$; otherwise, $E[\Delta_t] \geq r - [(2c^2 - 2c + 1)/2c]\overline{P} - c\overline{P_{(1,m)}}$.*

After the lower bound for $\mathrm{E}\left[\Delta_t\right]$ is established in Corollary 11, the next theorem sets a lower bound for $\mathrm{E}\left[\|\mathbf{c(t+1)}\|\right]$. An upper bound for $\mathrm{E}\left[\Delta_t\right]$ will be accordingly yielded as a corollary.

Theorem 12 (Lower bound for the expected norm of the next center). *If $c \leq 1/2$, $E[\|\mathbf{c(t+1)}\|] \geq r\left(1 - \exp\left(-2n'[\Phi(-r/\sigma)]^m\right)\right)$.*

Proof. Since $\|\mathbf{c(t+1)}\|$ is a non-negative random variable, from the Markov's inequality, we have, for any positive number a,

$$\mathrm{Prob}\left\{\|\mathbf{c(t+1)}\| > a\right\} \leq a^{-1}\mathrm{E}\left[\|\mathbf{c(t+1)}\|\right].$$

Substituting a with r implies

$$r\,\mathrm{Prob}\left\{\|\mathbf{c(t+1)}\| > r\right\} \leq \mathrm{E}\left[\|\mathbf{c(t+1)}\|\right].$$

Let's take a closer look at $\mathrm{Prob}\left\{\|\mathbf{c(t+1)}\| > r\right\}$. Denoting the jth coordinate of $\mathbf{P_i}$ and $\mathbf{c(t+1)}$ as $P_i(j)$ and $c(t+1)(j)$, respectively. If there exists a coordinate j such that $\min\{P_1(j), P_2(j), \ldots, P_m(j)\} \geq r$, then

$$\|\mathbf{c(t+1)}\| \geq |c(t+1)(j)|$$

$$= \left|\frac{\sum_{i=1}^{m}\left[P_i(j) + C\left(P_a(j) - P_i(j)\right)\right]}{m}\right|$$

$$= \frac{\sum_{i=1}^{m}\left[\left(1 - C\right)P_i(j) + CP_a(j)\right]}{m}$$

$$\geq \frac{\left[\left(1 - C\right)mr + Cmr\right]}{m}$$

$$= r.$$

13

Similarly, $\max\{P_1(j), P_2(j), \ldots, P_m(j)\} \leq -r$ implies $\|\mathbf{c(t+1)}\| > r$. Let E_j^+ be the event that $\min\{P_1(j), P_2(j), \ldots, P_m(j)\} \geq r$ and E_j^- be the event that $\max\{P_1(j), P_2(j), \ldots, P_m(j)\} \leq -r$. Let $E_j := E_j^+ \bigcup E_j^-$ and $E := \bigcup_{j=1}^m E_j$, we have

$$
\begin{aligned}
\text{Prob}\{E\} &= \text{Prob}\left\{E \bigcap E_1^+\right\} + \text{Prob}\left\{E \bigcap (E_1^+)^c\right\} \\
&\geq \text{Prob}\{E_1^+\} + \text{Prob}\left\{\left(\bigcup_{i=2}^n E_i\right) \bigcap (E_1^+)^c\right\} \\
&= \text{Prob}\{E_1^+\} + (1 - \text{Prob}\{E_1^+\})\, \text{Prob}\left\{\bigcup_{i=2}^n E_i\right\} \ .
\end{aligned}
$$

Because $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}$ are i.i.d. and for each particle all of its coordinates other than the first one are identically distributed, for all $i > 1$ the symmetry and disjointness of E_i^+ and E_i^- imply that $\text{Prob}\{E_i\} = 2\text{Prob}\{E_i^+\} = 2[1 - \Phi(r/\sigma)]^m = 2[\Phi(-r/\sigma)]^m$. Let $q := 2[\Phi(-r/\sigma)]^m$ for convenience of notation. Using the inclusion-exclusion principle, we have

$$
\begin{aligned}
\text{Prob}\left\{\bigcup_{i=2}^n E_i\right\} &= \sum_{i=1}^{n'} \binom{n'}{i} q^i (-1)^{i+1} \\
&= 1 - \sum_{i=0}^{n'} \binom{n'}{i} (-q)^i \\
&= 1 - (1-q)^{n'} \\
&\geq 1 - \exp\left(-n'q\right) \ .
\end{aligned}
$$

As a result,

$$
\begin{aligned}
\text{E}\left[\|\mathbf{c(t+1)}\|\right] &\geq r \left(\text{Prob}\{E_1^+\} + \left(1 - \text{Prob}\{E_1^+\}\right)\left(1 - \exp\left(-n'q\right)\right)\right) \\
&\geq r \left(\text{Prob}\{E_1^+\} + 1 - \text{Prob}\{E_1^+\} - \exp\left(-n'q\right)\right) \\
&= r \left(1 - \exp\left(-2n' [\Phi(-r/\sigma)]^m\right)\right) \ .
\end{aligned}
$$

$\square$

Corollary 13 (Upper bound for the progress rate). *If $c < 1/2$, then $E[\Delta_t] \leq r \exp\left(-2n'[\Phi(-r/\sigma)]^m\right)$.*

With Theorems 10 and 12, we established the upper and lower bounds of the expected particle norm. Accordingly, with Corollaries 11 and 13, we derived the lower and upper bounds of the expected progress rate of the swarm in the social-only model. As aforementioned, by statistically interpreting the social-only model PSO, we can describe the "macrostate" of the particle swarm and therefore are able to analyze the stochastic behavior of PSO based on the facet of particle interaction.

4 Convergence Analysis

As stated in section 2.2, the transition from the current time step to the next time step consists of updating positions of particles, calculating the center of distribution by means of the updated positions, and using the maximum likelihood estimation to calculate the distribution variance. The issues related to the center of distribution have been addressed in section 3. Now, we move on to the part of variance. While the center of distribution can be viewed as the indication of the average quality of the swarm at a specific time step, the variance is a direct measurement

of convergence, because from the viewpoint of statistical interpretation, the swarm *converges* as the variance of distribution reduces to zero. The word "converge" is not a unified term in the research domain of PSO ([23], p. 132). It has been used to describe the behavior of the swarm approaching the local optimum in some papers, while it simply indicates the phenomenon that the swarm of particles crowds into a specific point, sometimes called the *equilibrium*, not necessarily the local optimum, in the search space. Here in the present work, we adopt the latter definition. We concentrate on the condition under which the swarm of particles may go into a stable state. We will demonstrate that if certain condition of the relationship between the swarm size and the acceleration coefficient is satisfied, the swarm in a social-only model does converge under the mechanism of particle interaction.

Given m observed vectors $\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m}$ that stand for the updated positions and the center of distribution denoted as $\mathbf{c(t+1)} = \overline{\mathbf{y}} := (\Sigma_{i=1}^m \mathbf{y_i})/m$, let $\mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_m}$ be random vectors sampled from $\theta(\|\overline{\mathbf{y}}\|, \sigma_{t+1}^2)$. They are n-dimensional random vectors centered at $\overline{\mathbf{y}}$ and the coordinate on each dimension is a random variable sampled from $N(0, \sigma_{t+1}^2)$, where σ_{t+1}^2 is the variance that we wish to estimate. In order to estimate the variance, the likelihood function of σ_{t+1}^2, $L(\sigma_{t+1}^2)$, can be defined as the joint probability:

$$L(\sigma_{t+1}^2) := \prod_{i=1}^m \left(\frac{1}{\sqrt{2\pi}\sigma_{t+1}} \right)^n \exp\left(\frac{-d\left(\mathbf{y_i}, \overline{\mathbf{y}}\right)^2}{2\sigma_{t+1}^2} \right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_{t+1}} \right)^{mn} \exp\left(\frac{-\sum_{i=1}^m d\left(\mathbf{y_i}, \overline{\mathbf{y}}\right)^2}{2\sigma_{t+1}^2} \right)$$

$$= K\sigma_{t+1}^{-mn} \exp\left(\frac{-R}{2\sigma_{t+1}^2} \right) ,$$

where

$$K := \left(\frac{1}{\sqrt{2\pi}} \right)^{mn} , \quad R := \sum_{i=1}^m d(\mathbf{y_i}, \overline{\mathbf{y}})^2 .$$

In order to get the σ_{t+1}^2 that maximizes $L(\sigma_{t+1}^2)$, we differentiate $L(\sigma_{t+1}^2)$ with respect to σ_{t+1}^2:

$$L'(\sigma_{t+1}^2) = -\frac{mn}{2} K \cdot \sigma_{t+1}^{-mn-2} \cdot \exp\left(\frac{-R}{2\sigma_{t+1}^2} \right) + \frac{R}{2} K \cdot \sigma_{t+1}^{-mn-4} \cdot \exp\left(\frac{-R}{2\sigma_{t+1}^2} \right) .$$

$L'(\sigma_{t+1}^2) = 0$ implies $\sigma_{t+1}^2 = R/(mn)$, and it is routine to check the maximality. Since both m and n are fixed, the only quantity needs to be taken care of is R, which the sum of square of the distance between each updated particle and the center. Now we look for the expectation of R. Recall that, given $\mathbf{c(t)} = (r, 0, 0, \ldots, 0)$ and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ with $Z_1, Z_2, \ldots, Z_n \sim N(0, \sigma_t^2)$, the m particles $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}$ are sampled from $\mathbf{c(t)} + \mathbf{Z}$, and the updated position is calculated as $\mathbf{P_i} + C(\mathbf{P_a} - \mathbf{P_i})$, where $\mathbf{P_a}$ is the attractor. Since $\mathbf{c(t+1)} = \sum_{i=1}^m [\mathbf{P_i} + C(\mathbf{P_a} - \mathbf{P_i})]/m$, R, as a random variable, can be defined by $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}$ and $\mathbf{P_a}$:

$$R = \sum_{i=1}^m \|\mathbf{P_i} + C(\mathbf{P_a} - \mathbf{P_i}) - \frac{\sum_{j=1}^m \left(\mathbf{P_j} + C(\mathbf{P_a} - \mathbf{P_j}) \right)}{m} \|^2 .$$

Denoting $\mathbf{P_i}$'s kth coordinate as $P_i(k)$, $\mathrm{E}\,[R]$ is derived in the following lemma:

Lemma 14. *Given the swarm size, m, and the variance of distribution at time t, $\sigma_t^2 = \sigma^2$, $E[R] = [(4/3)c^2 - 2c + 1]n\,(m-1)\,\sigma^2$ and $E\left[\sigma_{t+1}^2\right] = [(4/3)c^2 - 2c + 1][(m-1)/m]\sigma^2$.*

15

Proof.

$$E[R] = E\left[\sum_{i=1}^{m}\|\mathbf{P_i} + C(\mathbf{P_a} - \mathbf{P_i}) - \frac{\sum_{j=1}^{m}(\mathbf{P_j} + C(\mathbf{P_a} - \mathbf{P_j}))}{m}\|^2\right]$$

$$= \sum_{i=1}^{m}E\left[\|\frac{m(1-C)\mathbf{P_i} - (1-C)(\sum_{j=1}^{m}\mathbf{P_j})}{m}\|^2\right]$$

$$= \left(\frac{1}{m}\right)^2 E\left[(1-C)^2\right]\sum_{i=1}^{m}E\left[\|m\mathbf{P_i} - \left(\sum_{j=1}^{m}\mathbf{P_j}\right)\|^2\right]$$

$$= \left(\frac{1}{m}\right)^2 E\left[C^2 - 2C + 1\right]\sum_{i=1}^{m}\sum_{k=1}^{n}E\left[\left(mP_i(k) - \left(\sum_{j=1}^{m}P_j(k)\right)\right)^2\right]$$

$$= \left(\frac{1}{m}\right)^2\left(\frac{4}{3}c^2 - 2c + 1\right) \cdot$$

$$\sum_{i=1}^{m}\sum_{k=1}^{n}E\left[\left(m^2 P_i(k)^2 - 2mP_i(k)\left(\sum_{j=1}^{m}P_j(k)\right) + \left(\sum_{j=1}^{m}P_j(k)\right)^2\right)\right].$$

The independence of particles and their coordinates implies $E[P_i(k)P_j(k)] = E[P_i(k)]\,E[P_j(k)] = 0$ for all $i \neq j$ and $E\left[P_i(k)^2\right] = \sigma^2$ for all i. Hence,

$$E[R] = \left(\frac{1}{m}\right)^2\left(\frac{4}{3}c^2 - 2c + 1\right)mn\left(m^2\sigma^2 - 2m\sigma^2 + m\sigma^2\right)$$

$$= \left(\frac{4}{3}c^2 - 2c + 1\right)n\,(m-1)\,\sigma^2.$$

As a result,

$$E\left[\sigma_{t+1}^2\right] = \frac{E[R]}{mn} = \left(\frac{4}{3}c^2 - 2c + 1\right)\left(\frac{m-1}{m}\right)\sigma^2.$$

$\square$

While Lemma 14 is under the assumption that σ_t^2 is given (or more formally, the conditional expectation $E\left[\sigma_{t+1}^2|\sigma_t^2 = \sigma^2\right]$ is derived), the following theorem turns to the relationship between $E\left[\sigma_t^2\right]$ and $E\left[\sigma_{t+1}^2\right]$ and gives a sufficient and necessary condition that the sequence $\left\{E\left[\sigma_t^2\right]\right\}$ converges to zero. Without loss of generality for the normal operation of PSO, we assume that $E\left[\sigma_0^2\right] < \infty$.

Theorem 15 (Convergence of the expectation of variance). *(1)* $E\left[\sigma_{t+1}^2\right] = [(4/3)c^2 - 2c + 1][(m-1)/m]E\left[\sigma_t^2\right]$, *and (2)* $\lim_{t\to\infty}\left\{E\left[\sigma_t^2\right]\right\} = 0$ *if and only if* $[(4/3)c^2 - 2c + 1] \leq [m/(m-1)]$.

Proof. (1) The first part of the theorem can be proved by a direct application of the law of total expectation and Lemma 14.

$$E\left[\sigma_{t+1}^2\right] = \int_{\sigma^2 \in \mathbb{R}^+}E\left[\sigma_{t+1}^2|\sigma_t^2 = \sigma^2\right]\text{Prob}\left\{\sigma_t^2 = \sigma^2\right\}d\sigma^2$$

$$= \int_{\sigma^2 \in \mathbb{R}^+}\left(\frac{4}{3}c^2 - 2c + 1\right)\left(\frac{m-1}{m}\right)\sigma^2\,\text{Prob}\left\{\sigma_t^2 = \sigma^2\right\}d\sigma^2$$

$$= \left(\frac{4}{3}c^2 - 2c + 1\right)\left(\frac{m-1}{m}\right)\int_{\sigma^2 \in \mathbb{R}^+}\sigma^2\,\text{Prob}\left\{\sigma_t^2 = \sigma^2\right\}d\sigma^2$$

$$= \left(\frac{4}{3}c^2 - 2c + 1\right)\left(\frac{m-1}{m}\right)E\left[\sigma_t^2\right].$$

16

(2) As a consequence of (1), $\left\{\mathrm{E}\left[\sigma_t^2\right]\right\}$ forms a geometric sequence with ratio $[(4/3)c^2-2c+1][(m-1)/m]$. We then obtain $\lim_{t\to\infty}\left\{\mathrm{E}\left[\sigma_t^2\right]\right\}=0$ if and only if $\left|[(4/3)c^2-2c+1][(m-1)/m]\right|<1$, and $\left|[(4/3)c^2-2c+1][(m-1)/m]\right|<1$ if and only if $[(4/3)c^2-2c+1]\leq[m/(m-1)]$. $\square$

Since σ_t^2 takes the value on non-negative real numbers, the convergence of sequence $\left\{\mathrm{E}\left[\sigma_t^2\right]\right\}$ implies sequence $\left\{\sigma_t^2\right\}$ converges to zero *in probability*, as shown in the following corollary.

Corollary 16 (Convergence of variance). *If* $[(4/3)c^2-2c+1]\leq[m/(m-1)]$, *then* $\lim_{t\to\infty}\sigma_t^2\xrightarrow{p}0$, *i.e., for every* $\epsilon>0$ $\lim_{t\to\infty}Prob\left\{\sigma_t^2\geq\epsilon\right\}=0$.

Proof. Suppose for contradiction that there exist $\epsilon>0$ and $\delta>0$ such that, for all $N_0\in\mathbb{N}$, there exists an $N(N_0)>N_0$ with $\mathrm{Prob}\left\{\sigma_{N(N_0)}^2\geq\epsilon\right\}\geq\delta$. However, since $\mathrm{Prob}\left\{\sigma_{N(N_0)}^2\geq\epsilon\right\}\geq\delta$ implies $\mathrm{Prob}\left\{\mathrm{E}\left[\sigma_{N(N_0)}^2\right]\right\}\geq\epsilon\delta$, for all $N_0\in\mathbb{N}$ there exists an $N(N_0)>N_0$ such that $\mathrm{Prob}\left\{\mathrm{E}\left[\sigma_{N(N_0)}^2\right]\right\}\geq\epsilon\delta$, $\lim_{t\to\infty}\left\{\mathrm{E}\left[\sigma_t^2\right]\right\}=0$ is contradicted. $\square$

Theorem 15 and Corollary 16 indicate that as long as the specified condition is satisfied, the swarm will converge in probability. However, it must be noted that the acceleration coefficient, c, used here is the coefficient for the compound effect of both the inertia weight and the regular acceleration coefficient for the neighborhood (or global) best position as described in section 3.3. Therefore, further investigations are needed to gain understandings on the compound effect and clarify the relationship of these parameters such that the derived results in the present work can be used in practice.

5 Summary and Conclusions

In this study, we made the first attempt to analyze the behavior of particle swarm optimization on the facet of particle interaction. We firstly proposed a statistical interpretation of particle swarm optimization and modeled the PSO mechanisms with the operations on probabilistic distributions. In order to investigate the PSO behavior based on particle interaction, we focused on the social-only model of PSO, in which the personal experience of particles is ignored. From the viewpoint of macrostates, we obtained the lower and upper bounds of the expected progress rate for the swarm on the sphere function. By looking into the variance of the particle distribution, we further showed that under certain condition, the swarm will converge in probability thanks to the mechanism of particle interaction, i.e., exchanging and sharing information, which is commonly believed to be an essential mechanism of PSO but seldom analyzed theoretically in the past.

With this study, we wish to propose an alternative way to analyze particle swarm optimization from the viewpoint of macrostates instead of tracing the trajectory of each particle. The immediate follow-up work for this study includes the clarification of the compound effect of the inertia weight and the neighborhood acceleration coefficient for carrying over the theoretical results to practice and suggesting parameter settings. Moreover, tighter bounds may be derived to more accurately describe the behavior of PSO, and a complete PSO model may be considered instead of the social-only model adopted in the present work. Finally, in the long run, a unified behavioral model of PSO might be established by integrating the theoretical results from the two ends – macrostates and microstates – such that better, more robust optimization frameworks can be accordingly designed and developed.

Acknowledgments

References

[1] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of 1995 IEEE International Conference on Neural Networks*, 1995, pp. 1942–1948.

[2] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of 1999 IEEE Congress on Evolutionary Computation (CEC 99)*, 1999, pp. 1945–1950.

[3] J. Kennedy, "The behavior of particles," in *Proceedings of the 7th International Conference on Evolutionary Programming*, 1998, pp. 581–589.

[4] E. Ozcan and C. K. Mohan, "Analysis of a simple particle swarm optimization system," *Intelligent Engineering Systems Through Artificial Neural Networks*, vol. 8, pp. 253–258, 1998.

[5] ——, "Particle swarm optimization: surfing the waves," in *Proceedings of 1999 IEEE Congress on Evolutionary Computation (CEC 99)*, 1999, pp. 1939–1944.

[6] M. Clerc and J. Kennedy, "The particle swarm—explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.

[7] Y. Shi and R. C. Eberhart, "Parameter selection in particle swarm optimization," in *Proceedings of the 7th International Conference on Evolutionary Programming*, 1998, pp. 591–600.

[8] F. van den Bergh, "An analysis of particle swarm optimizers," Ph.D. dissertation, University of Pretoria, 2002.

[9] K. Yasuda, A. Ide, and N. Iwasaki, "Adaptive particle swarm optimization," in *Proceedings of 1999 IEEE International Conference on Systems, Man and Cybernetics*, 2003, pp. 1554–1559.

[10] Y.-L. Zheng, L.-H. Ma, L.-Y. Zhang, and J.-X. Qian, "On the convergence analysis and parameter selection in particle swarm optimization," in *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, 2003, pp. 1802–1807.

[11] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Information Processing Letters*, vol. 85, pp. 317–325, 2003.

[12] V. Kadirkamanathan, K. Selvarajah, and P. J. Fleming, "Stability analysis of the particle dynamics in particle swarm optimizer," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 245–255, 2006.

[13] M. Jiang, Y. Luo, and S. Yang, "Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm," *Information Processing Letters*, vol. 102, pp. 8–16, 2007.

[14] H. M. Emara and H. A. A. Fattah, "Continuous swarm optimization technique with stability analysis," in *Proceedings of the 2004 American Control Conference*, 2004, pp. 2811–2817.

[15] V. Gazi and K. M. Passino, "Stability analysis of social foraging swarms," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 34, no. 1, pp. 539–555, 2004.

[16] J. Kennedy, "The particle swarm: social adaptation of knowledge," in *Proceedings of the 1997 IEEE International Conference on Evolutionary Computation*, 1997, pp. 303–308.

[17] H. G. Beyer, *The Theory of Evolution Strategies.* Springer, 2001.

[18] J. Jägersküpper, "Algorithmic analysis of a basic evolutionary algorithm for continuous optimization," *Theoretical Computer Science*, vol. 379, no. 3, pp. 329–347, 2007.

[19] J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*, 2nd ed. CRC Press, 1996.

[20] Y. S. Chow and H. Teicher, *Probability theory: independence, interchangeability, martingales*, 3rd ed. Springer, 1997.

[21] D. B. Owen, "A table of normal integrals," *Communications in Statistics - Simulation and Computation*, vol. B.9, pp. 389–419, 1980.

[22] I. Rechenberg, *Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog, Stuttgart, Germany, 1973.

[23] A. Engelbrecht, *Fundamentals of Computational Swarm Intelligence.* John Wiley & Sons, 2005.