Theoretical Computer Science 411 (2010) 2101–2115

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Analysis of particle interaction in particle swarm optimization

Ying-ping Chen *, Pei Jiang

Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

ARTICLE INFO

Keywords:
Particle swarm optimization
Particle interaction
Social-only model
Statistical interpretation
Theoretical analysis
Progress rate
Convergence
Sphere function

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 particle swarm optimization 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.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Particle swarm optimization (PSO), introduced by Kennedy and Eberhart [1] in 1995, was proposed based on an inspiration from the social behavior of insects or animals that the exchanging and sharing of information among a group of individuals benefit the group survival by improving the group capability of foraging. In the framework of 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 is affected by three factors: the inertia, personal experience (the cognitive part), and particle interaction (the social part).

Since its introduction, PSO has been empirically shown to be a very useful and effective optimization framework [2] for the easiness to implement and flexibility to use. Although PSO is widely applied in many research fields nowadays, the theoretical analysis on PSO is still quite limited. To the best of our knowledge, the first analysis was proposed by Kennedy [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 [7–11]. 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 common 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

* Corresponding author.
E-mail addresses: ypchen@nclab.tw (Y.-p. Chen), pjiang@nclab.tw (P. Jiang).

0304-3975/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2010.03.003

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 our 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 F : R^n → R)
    Initialize a swarm of m particles
    while the stopping criterion is not satisfied do
        Evaluate each particle
        for particle i, i = 1, 2, ..., m do                          ▷ Update the best positions
            if F(X_i) < F(Pb_i) then
                Pb_i ← X_i
                if F(Pb_i) < F(Nb) then
                    Nb ← Pb_i
                end if
            end if
        end for
        for particle i, i = 1, 2, ..., m do                          ▷ Generate the next generation
            V_i(t+1) ← wV_i(t) + C_p ⊗ (Pb_i − X_i) + C_n ⊗ (Nb − X_i)
            X_i(t+1) ← X_i(t) + 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. The notation $\otimes$ indicates component-by-component multiplication. According to Algorithm 1, we can see that a standard PSO system comprises the following two main operations regarding the 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. Since $\mathbf{Pb_i}$ and $\mathbf{Nb}$ exert gravity on other particles, they are referred to as *attractors* in this study.
(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. $\mathbf{C_p}$ and $\mathbf{C_n}$ are random vectors with each component sampled from uniform distributions $U(0, c_p)$ and $U(0, c_n)$ 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 observe, 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 of PSO

In spite of its importance, the effect of particle interaction in PSO is hardly investigated in the literature. Although there are a number of remarkable theoretical studies that bring insights into the properties and behavior of PSO conducted in the past, most of those studies are based on the assumption that the attractor is fixed, e.g., the trajectory analysis [4,5] mentioned in Section 1. Such a setting seems 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 of the whole system. The analysis on the overall behavior of a 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 a 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 a 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 of 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 of the entire swarm.

More specifically, the exact locations of particles are no longer traced but instead modeled and expressed by using a distribution $\theta(t)$ over $\mathbb{R}^n$. 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., 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

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

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

$$\begin{aligned} \theta(t+1) &\leftarrow \mathcal{T}_{\mathcal{P}}(\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}) ; \\ \mathcal{V}(t+1) &\leftarrow \mathcal{T}_{\mathcal{V}}(\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}; \mathbf{V_1}, \mathbf{V_2}, \ldots, \mathbf{V_m}) . \end{aligned}$$

By considering PSO 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 $\mathbf{V} \sim \mathcal{V}(t)$, where the notation " $\sim$ " indicates "is distributed according to," 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. Therefore, in the following, we will alter the notation θ to denote this compound distribution and parameterize it based on varied contexts.

The remaining questions would be what distribution is suitable for the description of a swarm without sacrificing too much essence of PSO and how to update the distribution as the search process proceeds. We can consider the random vector $\mathbf{X} \sim \theta(t)$, denote $\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu}$, and 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 $i \in \{1, 2, \ldots, s\}$. Each region is associated with a random variable of velocity $\mathbf{V_i} \sim \mathcal{V}(t)$. If one point $\mathbf{x_i}$ is respectively selected from each region R_i, when s is sufficiently large, the average behavior of a swarm can therefore be characterized by

$$\begin{aligned} \sum_{i=1}^{s} \frac{1}{s}(\mathbf{x_i} + \mathbf{V_i}) &= \sum_{i=1}^{s} \frac{1}{s}\mathbf{x_i} + \sum_{i=1}^{s} \frac{1}{s}\mathbf{V_i} \\ &\approx \boldsymbol{\mu} + \sum_{i=1}^{s} \frac{1}{s}\mathbf{V_i} , \end{aligned}$$

and each component of the term $\sum_{i=1}^{s}(1/s)\mathbf{V_i}$ can be approximated with a normal distribution according to the central limit theorem. Thus, 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. We can then reparameterize $\theta(t)$, the distribution of $\mathbf{c(t)} + \mathbf{Z}$, as $\theta(\mathbf{c(t)}, \sigma^2)$.

The update of distributions can now be 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 be described with the following algorithm:

Table 1
Average p-values of normality tests.

Swarm size	Normality tests		
	Shapiro–Wilk [17]	Anderson–Darling [18]	D'Agostino–Pearson [19]
10	0.3879	0.3621	0.3985
20	0.3257	0.2842	0.3393
30	0.2903	0.2518	0.2876

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
    $\mathbf{P}^* = \arg\min\{\mathcal{F}(\mathbf{P_1}), \mathcal{F}(\mathbf{P_2}), \ldots, \mathcal{F}(\mathbf{P_m})\}$
    for $i = 1, 2, \ldots, m$ **do**
      $\mathbf{P'_i} \leftarrow \mathbf{P_i} + \mathbf{C_i} \otimes (\mathbf{P}^* - \mathbf{P_i})$
    end for
    $\boldsymbol{\mu}_{\mathbf{t+1}} \leftarrow (\sum_{i=1}^{m} \mathbf{P'_i})/m$
    $\sigma^2_{t+1} \leftarrow \mathbf{MLE}(\mathbf{P'_1}, \mathbf{P'_2}, \ldots, \mathbf{P'_m})$
    $\theta \leftarrow \theta(\boldsymbol{\mu}_{\mathbf{t+1}}, \sigma^2_{t+1})$
    $t \leftarrow t + 1$
  end while
end procedure

In order to validate the utilization of normal distributions for describing swarms, we conducted three well-known normality tests: the Shapiro–Wilk test [17], the Anderson–Darling test [18], and the D'Agostino–Pearson test [19] on the social-only PSO on the sphere function. Table 1 displays the test results, which were obtained for 100 independent runs and 10 iterations in each run. The weight for the inertia is 0.73 and the acceleration coefficient is 1.49. Since all p-values of the three normality tests significantly surpass the conventional significance level 0.05, none of these tests are able to reject the null hypothesis. As a result, in this study, adopting the normal distribution as the description of swarms is an acceptable assumption.

In summary, the macrostate model transforms the detailed configuration of PSO into a corresponding stochastic representation embodied by normal distributions. As a consequence, the update of particles is simplified as the modification of the parameters of normal distributions. In each iteration, Algorithm 2 generates a swarm of particles by means of sampling from the current distribution, and thereafter, the distribution is updated according to particle interaction. In others words, a state of Algorithm 2 is a distribution, and the sampled swarm serves as a medium for state transition. In this manner, the analysis of the behavior of the entire swarm is thus reduced to the analysis of parameterized distributions. The inclusion of particle interaction into analysis supplies numerous facets of PSO typically absent in related theoretical studies, e.g., the progress rate and the influence of objective functions, because the restriction of fixed attractors makes objective functions irrelevant. Since the No-Free-Lunch theorem states that all optimization algorithms perform identically on average [20], the effectiveness of PSO can hardly be theoretically identified unless the scope of functions is specified.

In the remainder 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., [21]) 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 PSO 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 the present work facilitates the analysis of inter-particle effects and how these effects are accounted for the progress rate of a 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 equal 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, $\mathbf{c}(\mathbf{t}) = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$, we would like to calculate the probability, denoted as $B_k(\mathbf{o})$, that $\mathbf{c}(\mathbf{t}) + \mathbf{Z} \sim \theta$ is in an n-ball located at the origin with radius k, 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 [22]. We elaborate this property as follows. Given $Z_1, Z_2, \ldots, Z_n \sim N(0, \sigma^2)$, for all $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

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

where $d(\cdot, \cdot)$ denotes the Euclidean distance. It is obvious that the density at point $\mathbf{x}$ is determined by $d(\mathbf{x}, \mathbf{c}(\mathbf{t}))$, regardless of the direction in which $\mathbf{x}$ is relatively to $\mathbf{c}(\mathbf{t})$. Therefore, without loss of generality, we may assume that $\mathbf{c}(\mathbf{t})$ is on the first axis by conducting a coordinate transformation. Let $r := d(\mathbf{c}(\mathbf{t}), \mathbf{o}) \geq 0$. As a result, $\mathbf{c}(\mathbf{t})$ 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 $\mathbf{c}(\mathbf{t}) + \mathbf{Z} \sim \theta(r, \sigma^2) \in \mathbb{R}^n$, where $\mathbf{c}(\mathbf{t}) = (r, 0, 0, \ldots, 0)$ and all coordinates of $\mathbf{Z}$ are distributed according to $N(0, \sigma^2)$, the n-ball hitting probability

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

The analysis approach adopted in the present work is similar to that used by Beyer in 2001 [21]. The vector $\mathbf{Z}$ is decomposed into two orthogonal vectors: $Z_1\mathbf{e_1} = (Z_1, 0, 0, \ldots, 0)$ and $\mathbf{Z}' = (0, Z_2, Z_3, \ldots, Z_n)$. We can take a closer look at the n-ball hitting probability $H_B(k, \theta(r, \sigma^2))$:

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

The 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

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

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

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

As a result, we can get

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

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 [23, 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. 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,

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

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 $H_B(k,\theta(r,\sigma^2))$*).

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

Proof. Let $\mathbf{Y} := \mathbf{c}(\mathbf{t})+\mathbf{Z}$, where $\mathbf{c}(\mathbf{t}) = (r,0,0,\ldots,0)$, 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 $\|\mathbf{x}\| \le \sqrt{n}\|\mathbf{x}\|_\infty \le \sqrt{n}(k/\sqrt{n}) = k$, we can know that $\mathbf{x}\in B_k(\mathbf{o})$. Hence, $\mathcal{D}\subseteq B_k(\mathbf{o})$, and

$$\begin{aligned}
\mathrm{Prob}\left\{\mathbf{Y}\in B_k(\mathbf{o})\right\} \ge \mathrm{Prob}\left\{\mathbf{Y}\in\mathcal{D}\right\} &= \mathrm{Prob}\left\{-\frac{k}{\sqrt{n}}-r \le Z_1 \le \frac{k}{\sqrt{n}}-r\right\}\prod_{i=2}^{n}\mathrm{Prob}\left\{-\frac{k}{\sqrt{n}} \le Z_i \le \frac{k}{\sqrt{n}}\right\} \\
&= \left[\Phi\left(\frac{\frac{k}{\sqrt{n}}-r}{\sigma}\right)-\Phi\left(\frac{-\frac{k}{\sqrt{n}}-r}{\sigma}\right)\right]\left[\Phi\left(\frac{\frac{k}{\sqrt{n}}}{\sigma}\right)-\Phi\left(\frac{-\frac{k}{\sqrt{n}}}{\sigma}\right)\right]^{n-1} \\
&= \left[\Phi\left(\frac{r+\frac{k}{\sqrt{n}}}{\sigma}\right)-\Phi\left(\frac{r-\frac{k}{\sqrt{n}}}{\sigma}\right)\right]\left[1-2\Phi\left(\frac{-k}{\sqrt{n}\sigma}\right)\right]^{n-1}.\quad\square
\end{aligned}$$

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 the antiderivative is defined as $\psi(k) := \int_{t=0}^{k} \psi'(t)\,\mathrm{d}t$.

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

$$\begin{aligned} H_S(k, \theta(r, \sigma^2)) &:= \mathrm{Prob}\left\{\|\mathbf{c}(\mathbf{t}) + \mathbf{Z}\| = k\right\} \\ &= \mathrm{Prob}\left\{-k - r \le Z_1 \le k - r, W = k^2 - x^2\right\} \\ &= \int_{x=-k-r}^{k-r} p(Z_1, x) p(W, k^2 - x^2)\,\mathrm{d}x. \end{aligned}$$

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

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

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. Hence, as we gain understandings of the characteristics of attractors and particles, we may capture the stochastic behavior of the PSO system. More specifically, because the distance from the origin 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 PSO.

Given the center of a particle distribution $\mathbf{c}(\mathbf{t}) = (r, 0, \ldots, 0)$ 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 $\mathbf{c}(\mathbf{t}) + \mathbf{Z}$, the expected norm of particles can be defined as

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

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* $\mathbf{c}(\mathbf{t}) = (r, 0, 0, \ldots, 0)$ *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:

$$\begin{aligned} \overline{P} &= \mathrm{E}\left[\|\mathbf{c}(\mathbf{t}) + \mathbf{Z}\|\right] \\ &= \mathrm{E}\left[\sqrt{(Z_1 - r)^2 + \sum_{i=2}^{n} Z_i^2}\right] \\ &\le \sqrt{\mathrm{E}\left[(Z_1 - r)^2 + \sum_{i=2}^{n} Z_i^2\right]} \\ &= \sqrt{\mathrm{E}\left[r^2\right] - 2r\mathrm{E}[Z_1] + \sum_{i=1}^{n} \mathrm{E}\left[Z_i^2\right]} \\ &= \sqrt{r^2 + n\sigma^2}. \end{aligned}$$

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

The expected particle norm describes how close on average a swarm is to the global optimum, i.e., the origin, 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 traceable 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

$$\begin{aligned}\operatorname{Prob}\left\{P_{(1,m)} = x\right\} &= \operatorname{Prob}\left\{\bigcup_{i=1}^{m}\left[\{\|\mathbf{P_i}\| = x\}\bigcap\left(\bigcap_{j\in\{1,2,\ldots,m\}\setminus\{i\}}\{\|\mathbf{P_j}\| > x\}\right)\right]\right\} \\ &= \int_{x=-k-r}^{k-r}\binom{m}{1}H_S(x,\theta(r,\sigma^2))\left[1-H_B(x,\theta(r,\sigma^2))\right]^m \mathrm{d}x\,.\end{aligned}$$

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

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

As a result,

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

Lemma 6 causes no 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)

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

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, $\mathrm{E}[|X|]^r = r\int_0^\infty t^{r-1}\operatorname{Prob}\{|X| > t\}\,\mathrm{d}t$ with $r > 0$ [24]. Since $P_{(1,m)}$ is a non-negative random variable, by letting $r = 1$ we have

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

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

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

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

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

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 $\int_{x=0}^{h}\left[1-\psi'(h)\right]\,\mathrm{d}x$ 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 $h>r\sqrt{n}$,

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

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

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

By applying Bernoulli's inequality, we can get

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

The integration of the normal distribution is given in [25]. When $h\to\infty$, the term

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

vanishes. Thus, 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

$$\begin{aligned}
\left(-r\sqrt{n}+h\right)\Phi\left(\frac{r-\frac{h}{\sqrt{n}}}{\sigma}\right) &= \left(h-r\sqrt{n}\right)\left[1-\Phi\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)\right] \\
&\le \left(h-r\sqrt{n}\right)\cdot\phi\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)\cdot\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right)^{-1} \\
&= \left(\sigma\sqrt{n}\right)\cdot\phi\left(\frac{\frac{h}{\sqrt{n}}-r}{\sigma}\right) \\
&= 0 \qquad \text{as } h\to\infty\,.
\end{aligned}$$

Therefore, $G(h)$ is bounded from above. The proof is completed. □

3.3. Lower and upper bounds for the expected progress rate

After the work was done in the previous sections, the progress rate of the social-only model PSO can now be formally investigated under the proposed statistical interpretation. The term "progress rate" was introduced by Rechenberg in 1973 [26]. As the 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 the two distribution centers in successive time steps, because the distance to the optimum is the L^2 norm for the sphere function. Given the current center of distribution $\mathbf{c}(\mathbf{t}) = (r, 0, 0, \dots, 0)$ and a random vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ with $Z_1, Z_2, \dots, Z_n \sim N(0, \sigma^2)$, the m particles $\mathbf{P_1}, \mathbf{P_2}, \dots, \mathbf{P_m}$ are sampled as $\mathbf{c}(\mathbf{t}) + \mathbf{Z}$. Let $P_{(i,m)}$ denote the ith order statistic of $\|\mathbf{P_1}\|, \|\mathbf{P_2}\|, \dots, \|\mathbf{P_m}\|$. Let $\mathbf{P}^* := \arg\min\{\mathcal{F}(\mathbf{P_1}), \mathcal{F}(\mathbf{P_2}), \dots, \mathcal{F}(\mathbf{P_m})\}$. By definition, $\|\mathbf{P}^*\| = P_{(1,m)}$. According to the update rules described in Section 2.2, the updated position $\mathbf{P'_i}$ is computed as $\mathbf{P'_i} = \mathbf{P_i} + \mathbf{C_i} \otimes (\mathbf{P}^* - \mathbf{P_i})$, where each coordinate of $\mathbf{C_i}$ is distributed according to $U(0, c)$ with c being the coefficient representing the compound effect of both the inertia weight and the acceleration coefficient of 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 $\mathbf{c}(\mathbf{t+1})$ is the mean of $\mathbf{P'_1}, \mathbf{P'_2}, \dots, \mathbf{P'_m}$, i.e., $\mathbf{c}(\mathbf{t+1}) = (\sum_{i=1}^{m} \mathbf{P'_i})/m$.

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

The following theorem shows that, when $c \le 1/2$, 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 $\overline{P_{(1,m)}}$.

Lemma 10. *Suppose $\mathbf{C} = (C_1, C_2, \dots, C_n)$ is a random vector of $\mathbb{R}^n$ with i.i.d. components and $\mathbf{X}$ is a random vector of $\mathbb{R}^n$. If $\mathbf{C}$ and $\mathbf{X}$ are independent, then $E[\|\mathbf{C} \otimes \mathbf{X}\|] \le \sqrt{\mu'_2}\, E[\|\mathbf{X}\|]$, where μ'_2 is the second moment of C_i.*

Proof. For any fixed vector $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

$$\begin{aligned}
\mathrm{E}[\|\mathbf{C} \otimes \mathbf{x}\|] &= \mathrm{E}\left[\sqrt{\sum_{i=1:n} C_i^2 x_i^2}\right] \\
&\le \sqrt{\mathrm{E}\left[\sum_{i=1:n} C_i^2 x_i^2\right]} \\
&= \sqrt{\sum_{i=1:n} \mathrm{E}\left[C_i^2\right] x_i^2} \\
&= \sqrt{\mu'_2}\, \|x\| .
\end{aligned}$$

Since $\mathbf{C}$ and $\mathbf{X}$ are independent, by the law of total expectation conditional on $\mathbf{X}$, this lemma is proved. □

Theorem 11 (*Upper Bound for the Expected Norm of the Next Center*). *(1) $E[\|\mathbf{c}(\mathbf{t+1})\|] \le E[|1 - C|]\overline{P} + E[|C|]\overline{P_{(1,m)}}$; and (2) If $c \le 1/2$, $E[\|\mathbf{c}(\mathbf{t+1})\|] \le (1 - c)\overline{P} + c\overline{P_{(1,m)}}$; otherwise, $E[\|\mathbf{c}(\mathbf{t+1})\|] \le [(2c^2 - 2c + 1)/2c]\overline{P} + c\overline{P_{(1,m)}}$.*

Proof. This result is derived from the triangle inequality for L^2-norm and the previous lemma:

$$\begin{aligned}
\mathrm{E}[\|\mathbf{c}(\mathbf{t+1})\|] &= \mathrm{E}\left[\left\|\frac{\sum_{i=1}^{m}\left[\mathbf{P_i} + \mathbf{C_i} \otimes \left(\mathbf{P}^* - \mathbf{P_i}\right)\right]}{m}\right\|\right] \\
&= \left(\frac{1}{m}\right) \mathrm{E}\left[\left\|\sum_{i=1}^{m} (\mathbf{1} - \mathbf{C_i}) \otimes \mathbf{P_i} + m\mathbf{C_i} \otimes \mathbf{P}^*\right\|\right] \\
&\le \left(\frac{1}{m}\right)\left(\sum_{i=1}^{m} \mathrm{E}[\|(\mathbf{1} - \mathbf{C_i}) \otimes \mathbf{P_i}\|] + m\mathrm{E}\left[\|\mathbf{C_i} \otimes \mathbf{P}^*\|\right]\right) \\
&\le \left(c^2/3 - c + 1\right)^{1/2} \overline{P} + \left(c^2/3\right)^{1/2} \overline{P_{(1,m)}}. \quad \square
\end{aligned}$$

Corollary 12 (*Lower Bound for the Progress Rate*). $E[\Delta_t] \ge r - \left(c^2/3 - c + 1\right)^{1/2} \overline{P} - \left(c^2/3\right)^{1/2} \overline{P_{(1,m)}}$.

After the lower bound for $\mathrm{E}[\Delta_t]$ is established in Corollary 12, the next theorem sets a lower bound for $\mathrm{E}[\|\mathbf{c}(\mathbf{t+1})\|]$. An upper bound for $\mathrm{E}[\Delta_t]$ will be accordingly obtained as a corollary.

Theorem 13 (*Lower Bound for the Expected Norm of the Next Center*). *If* $c \leq 1$, $E\left[\|\mathbf{c(t+1)}\|\right] \geq r(1-\exp(-2n'[\Phi(-r/\sigma)]^m))$.

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

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

Substituting a with r,

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

Let the jth coordinate of $\mathbf{P_i}$, $\mathbf{C_i}$, and $\mathbf{c(t+1)}$ be P_{ij}, C_{ij}, and $c(t+1)_j$, respectively. If there exists a coordinate j such that $\min\{P_{1j}, P_{2j}, \ldots, P_{mj}\} \geq r$, then

$$\begin{aligned}
\|\mathbf{c(t+1)}\| &\geq |c(t+1)_j| \\
&= \left|\frac{\sum_{i=1}^{m}\left[P_{ij} + C_{ij}\left(P_j^* - P_{ij}\right)\right]}{m}\right| \\
&= \frac{\sum_{i=1}^{m}\left[\left(1-C_{ij}\right)P_{ij} + C_{ij}P_j^*\right]}{m} \\
&\geq \frac{\left[\left(1-C_{ij}\right)mr + C_{ij}mr\right]}{m} \\
&= r.
\end{aligned}$$

Similarly, $\max\{P_{1j}, P_{2j}, \ldots, P_{mj}\} \leq -r$ implies $\|\mathbf{c(t+1)}\| > r$. Let E_j^+ be the event that $\min\{P_{1j}, P_{2j}, \ldots, P_{mj}\} \geq r$ and E_j^- be the event that $\max\{P_{1j}, P_{2j}, \ldots, P_{mj}\} \leq -r$. Let $E_j := E_j^+ \bigcup E_j^-$ and $E := \bigcup_{j=1}^m E_j$, we have

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

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 $\text{Prob}\{E_i\} = 2\text{Prob}\left\{E_i^+\right\} = 2[1-\Phi(r/\sigma)]^m = 2[\Phi(-r/\sigma)]^m$. Let $q := 2[\Phi(-r/\sigma)]^m$ for convenience of notation. By 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,

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

Corollary 14 (*Upper Bound for the Progress Rate*). *If* $c < 1$, *then* $E\left[\Delta_t\right] \leq r\exp\left(-2n'[\Phi(-r/\sigma)]^m\right)$.

With Theorems 11 and 13, we established the upper and lower bounds of the expected particle norm. Accordingly, with Corollaries 12 and 14, we derived the lower and upper bounds of the expected progress rate of a 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 distribution center by means of the updated positions, and using the maximum likelihood estimation to calculate the distribution variance. The issues related to the centers of distributions have been addressed in Section 3. Thus, the part of variance is considered in this section. While the center of a 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, a swarm *converges* as the variance of the distribution reduces to zero. The word "converge" is not a unified term in the research domain of PSO [27, p. 132]. It has been used to describe the behavior of a swarm approaching the local optimum in some papers, while it simply indicates the phenomenon that a swarm of particles crowds into a specific point, sometimes called the *equilibrium*, not necessarily the local optimum, in the search space in other papers. Here in the present work, we adopt the latter definition. We concentrate on the condition under which a 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, a swarm in the 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 distribution center is denoted as $\mathbf{c(t+1)} = \bar{\mathbf{y}} := (\Sigma_{i=1}^{m}\mathbf{y_i})/m$. Let $\mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_m}$ be random vectors sampled from $\theta(\|\bar{\mathbf{y}}\|, \sigma_{t+1}^2)$. These vectors are n-dimensional random vectors centered at $\bar{\mathbf{y}}$, 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:

$$
\begin{aligned}
L(\sigma_{t+1}^2) &:= \prod_{i=1}^{m}\left(\frac{1}{\sqrt{2\pi}\,\sigma_{t+1}}\right)^{n}\exp\left(\frac{-d\,(\mathbf{y_i},\bar{\mathbf{y}})^2}{2\sigma_{t+1}^2}\right)\\
&= \left(\frac{1}{\sqrt{2\pi}\,\sigma_{t+1}}\right)^{mn}\exp\left(\frac{-\displaystyle\sum_{i=1}^{m} d\,(\mathbf{y_i},\bar{\mathbf{y}})^2}{2\sigma_{t+1}^2}\right)\\
&= K\sigma_{t+1}^{-mn}\exp\left(\frac{-R}{2\sigma_{t+1}^2}\right)\,,
\end{aligned}
$$

where

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

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 examined is R, the sum of square of the distance between each updated particle and the center. 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 $\mathbf{P_i} + \mathbf{C_i} \otimes (\mathbf{P}^* - \mathbf{P_i})$, where $\mathbf{P}^*$ is the attractor. Since $\mathbf{c(t+1)} = \sum_{i=1}^{m}\left[\mathbf{P_i} + \mathbf{C_i}\otimes(\mathbf{P}^* - \mathbf{P_i})\right]/m$, R, as a random variable, can be defined by $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_m}$ and $\mathbf{P}^*$:

$$
R = \sum_{i=1}^{m}\left\|\mathbf{P_i} + \mathbf{C_i}\otimes(\mathbf{P}^* - \mathbf{P_i}) - \frac{\displaystyle\sum_{j=1}^{m}\left(\mathbf{P_j} + \mathbf{C_i}\otimes(\mathbf{P}^* - \mathbf{P_j})\right)}{m}\right\|^2.
$$

Denoting $\mathbf{P_i}$'s and $\mathbf{P}^*$'s kth coordinate as P_{ik} and P_k^*, respectively, the expectation of R, $\mathrm{E}[R]$, can be derived in the following lemma:

Lemma 15. *Given the swarm size, m, and the variance of distribution at time t,* $\sigma_t^2 = \sigma^2$,

$$
E\left[\sigma_{t+1}^2\right] \le \frac{(m-1)\sigma^2}{12m}\left\{\left(5 + \frac{\sqrt{3(m-1)}}{n}\right)c^2 - 6c + 12\right\}.
$$

Proof. Defining R_j as

$$R_j := \sum_{i=1}^{m} \left\{ P_{ij} + C_{ij}\left(P_j^* - P_{ij}\right) - \frac{\sum_{k=1}^{m}\left(P_{kj} + C_{kj}(P_j^* - P_{kj})\right)}{m} \right\}^2$$

yields $R = \sum_{j=1:n} R_j$ and that we can obtain for $j > 1$,

$$\begin{aligned} \mathrm{E}\left[R_j\right] &= \left(\frac{m-1}{12m}\right)\left\{\sigma^2\left(4c^2 - 6c + 12\right) + c^2\mathrm{E}\left[\left(P_j^*\right)^2\right]\right\} \\ &\leq \left(\frac{m-1}{12m}\right)\left\{\sigma^2\left(5c^2 - 6c + 12\right)\right\} . \end{aligned}$$

The last inequality follows from the fact that the independence of coordinates implies $\mathrm{E}\left[\left(P_j^*\right)^2\right] \leq \sigma^2$. Moreover,

$$E[R_1] = \left(\frac{m-1}{12m}\right)\left\{\sigma^2\left(4c^2 - 6c + 12\right) + c^2\mathrm{E}\left[\left(P_1^* - r\right)^2\right]\right\} .$$

Since $\mathrm{E}\left[\left(P_1^* - r\right)^2\right]$ is less than or equal to the expected value of the extreme order statistics of $T_1^2, T_2^2, \ldots, T_m^2$, where $T_i \sim N(0, \sigma^2)$, by using the upper bound for the extreme order statistics [28],

$$\mathrm{E}\left[\left(P_1^* - r\right)^2\right] \leq \sigma^2\left(1 + \sqrt{3(m-1)}\right) .$$

As a consequence,

$$\mathrm{E}\left[\sigma_{t+1}^2\right] = \mathrm{E}\left[R\right]/(mn) \leq \frac{(m-1)\sigma^2}{12m}\left\{\left(5 + \frac{\sqrt{3(m-1)}}{n}\right)c^2 - 6c + 12\right\}. \quad \square$$

While Lemma 15 is under the assumption that σ_t^2 is given or more formally, the conditional expectation $\mathrm{E}\left[\sigma_{t+1}^2 | \sigma_t^2 = \sigma^2\right]$ is derived, the following theorem indicates the relationship between $\mathrm{E}\left[\sigma_t^2\right]$ and $\mathrm{E}\left[\sigma_{t+1}^2\right]$ and gives a sufficient and necessary condition that the sequence $\left\{\mathrm{E}\left[\sigma_t^2\right]\right\}$ converges to zero. Without loss of generality for the normal operation of PSO, we assume that $\mathrm{E}\left[\sigma_0^2\right] < \infty$.

Theorem 16 (*Convergence of the Expectation of Variance*). *Let $\kappa := \sqrt{3(m-1)}/n$. If c satisfies the condition that*

$$\frac{3 - \sqrt{9 + \frac{60+5\kappa}{m-1}}}{5+\kappa} < c < \frac{3 + \sqrt{9 + \frac{60+5\kappa}{m-1}}}{5+\kappa},$$

$\lim_{t\to\infty}\left\{E\left[\sigma_t^2\right]\right\} = 0$.

Proof. The law of total expectation and Lemma 15 imply that

$$\mathrm{E}\left[\sigma_{t+1}^2\right] \leq \frac{(m-1)}{12m}\left\{\left(5 + \frac{\sqrt{3(m-1)}}{n}\right)c^2 - 6c + 12\right\}\mathrm{E}\left[\sigma_t^2\right].$$

Therefore, $\left\{\mathrm{E}\left[\sigma_t^2\right]\right\}$ is upper-bounded by the geometric sequence with the first term $\mathrm{E}\left[\sigma_0^2\right]$ and the ratio

$$\frac{(m-1)}{12m}\left\{\left(5 + \frac{\sqrt{3(m-1)}}{n}\right)c^2 - 6c + 12\right\}.$$

By solving

$$\frac{(m-1)}{12m}\left\{\left(5 + \frac{\sqrt{3(m-1)}}{n}\right)c^2 - 6c + 12\right\} < 1,$$

the theorem is proved. $\square$

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 17 (*Convergence of Variance*). *If* $\lim_{t\to\infty}\left\{E\left[\sigma_t^2\right]\right\} = 0$, *then* $\lim_{t\to\infty} \sigma_t^2 \xrightarrow{p} 0$, *i.e., for every* $\epsilon > 0$ $\lim_{t\to\infty} \mathrm{Prob}\left\{\sigma_t^2 \geq \epsilon\right\} = 0$.

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

Theorem 16 and Corollary 17 indicate that as long as the specified condition is satisfied, a swarm will converge in probability. However, it must be noted that the acceleration coefficient, c, used in this study is the coefficient for the compound effect of both the inertia weight and the common 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 applied 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 essential 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 a swarm on the sphere function. By examining in detail the variance of the particle distribution, we further showed that under certain condition, a swarm will converge in probability due 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 theoretically analyzed in the literature.

With regard to the practical implications of this study, we demonstrated that the optimization process of PSO can be interpreted as the interplay between the attractor and the overall swarm, as shown in Theorem 11 that the expected norm of the next center is upper-bounded by a linear combination of $\overline{P}$ and $\overline{P_{(1,m)}}$ as well as that the acceleration coefficient is the weight balancing the effects of these two quantities. The major resistance in the optimization process of PSO on the sphere function is the number of dimensions, as it can be observed in Corollary 14 that the progress rate deteriorates drastically with respect to the number of dimensions. On the other hand, the swarm size is the primary factor counteracting the increasing dimensions, for the exploratory capability of the swarm is augmented in accordance with the number of particles. It is noteworthy that in a variety of theoretical studies on PSO, the effect of the objective function has been rarely taken into consideration due to the assumption of fixed attractors. By means of characterizing a swarm as a unity, the analysis of the influence of the objective function becomes possible.

With this study, we 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 of 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 for suggesting applicable 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.

Acknowledgements

The work was supported in part by the National Science Council of Taiwan under Grant NSC 98-2221-E-009-072. The authors are grateful to the National Center for High-performance Computing for computer time and facilities.

References

[1] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of 1995 IEEE International Conference on Neural Networks, 1995, pp. 1942–1948.
[2] Y. Shi, 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, C.K. Mohan, Analysis of a simple particle swarm optimization system, Intelligent Engineering Systems Through Artificial Neural Networks 8 (1998) 253–258.
[5] E. Ozcan, C. K. Mohan, Particle swarm optimization: surfing the waves, in: Proceedings of 1999 IEEE Congress on Evolutionary Computation, CEC 99, 1999, pp. 1939–1944.
[6] M. Clerc, J. Kennedy, The particle swarm-explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation 6 (1) (2002) 58–73.
[7] Y. Shi, 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. Thesis, University of Pretoria, 2002.
[9] K. Yasuda, A. Ide, 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, 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 85 (2003) 317–325.
[12] V. Kadirkamanathan, K. Selvarajah, P.J. Fleming, Stability analysis of the particle dynamics in particle swarm optimizer, IEEE Transactions on Evolutionary Computation 10 (3) (2006) 245–255.
[13] M. Jiang, Y. Luo, S. Yang, Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm, Information Processing Letters 102 (2007) 8–16.
[14] H. M. Emara, 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, K. M. Passino, Stability analysis of social foraging swarms, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics 34 (1) (2004) 539–555.
[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] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (complete samples), Biometrika 52 (3/4) (1965) 591–611.
[18] T.W. Anderson, D.A. Darling, Asymptotic theory of certain "goodness of fi" criteria based on stochastic processes, Annals of Mathematical Statistics 23 (2) (1952) 193–212.
[19] R. D'Agostino, E. S. Pearson, Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$, Biometrika 60 (3) (1973) 613–622.
[20] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation 4 (1997) 67–82.
[21] H.G. Beyer, The Theory of Evolution Strategies, Springer, 2001.
[22] J. Jägersküpper, Algorithmic analysis of a basic evolutionary algorithm for continuous optimization, Theoretical Computer Science 379 (3) (2007) 329–347.
[23] J.K. Patel, C.B. Read, Handbook of the Normal Distribution, 2nd ed., CRC Press, 1996.
[24] Y. S. Chow, H. Teicher, Probability Theory: independence, Interchangeability, Martingales, 3rd edition, Springer, 1997.
[25] D. B. Owen, A table of normal integrals, Communications in Statistics — Simulation and Computation B.9 (1980) 389–419.
[26] I. Rechenberg, Evolutionsstrategie — Optimierung Technischer Systeme Nach Prinzipien der Biologischen Evolution, Frommann-Holzboog, Stuttgart, Germany, 1973.
[27] A. Engelbrecht, Fundamentals of Computational Swarm Intelligence, John Wiley & Sons, 2005.
[28] D. Bertsimas, K. Natarajan, C.-P. Teo, Tight bounds on expected order statistics, Probability in the Engineering and Informational Sciences 20 (4) (2006) 667–686.