Order-2 Stability Analysis of Particle Swarm Optimization

Article *in* Evolutionary Computation · April 2014

DOI: 10.1162/EVCO_a_00129 · Source: PubMed

1 author:

Qunfeng Liu
Dongguan University of Technology

Order-2 Stability Analysis of Particle Swarm Optimization *

Qunfeng Liu liuqf@dgut.edu.cn
College of Computing, Dongguan University of Technology, Dongguan, 523808, China

Abstract

Several stability analyses and stable regions of particle swarm optimization (PSO) have been proposed before. The assumption of stagnation and different definitions of stability are adopted in these analyses. In this paper, the order-2 stability of PSO are analyzed based on a weak stagnation assumption. A new definition of stability is proposed and an order-2 stable region is obtained. Several existing stable analyses for canonical PSO are compared, especially their definitions of stability and the corresponding stable regions. It is shown that, the classical stagnation assumption is too strict and not necessary. Moreover, among all these definitions of stability, it is shown that our definition requires the weakest conditions, and additional conditions bring no benefit. Finally, numerical experiments are reported to show that the obtained stable region is meaningful. A new parameter combination of PSO is also shown to be good, even better than some known best parameter combinations.

Keywords

Particle swarm optimization, order-2 stability analysis, weak stagnation, order-2 stable region, parameter selection.

1 Introduction

Particle swarm optimization (PSO) is firstly proposed to model the intelligent behaviors of bird flocking (Kennedy and Eberhart, 1995), and is soon developed into a powerful optimization method for the following minimization (or maximization) optimization problem (Kennedy and Eberhart, 1995; Eberhart and Kennedy, 1995),

$$\min_x f(x), \quad s.t. \quad x \in \Omega \subseteq \mathbb{R}^n.$$

In the PSO algorithm, each particle "flies" from one position to another position, communicates information with other particles in its neighborhood, and then changes its velocity and position according to the following equations:

$$\boldsymbol{v}_{ij}(k+1) = \omega \boldsymbol{v}_{ij}(k) + C_{1,ij}(\boldsymbol{p}_{ij}(k) - \boldsymbol{x}_{ij}(k)) + C_{2,ij}(\boldsymbol{g}_{ij}(k) - \boldsymbol{x}_{ij}(k)), \tag{1a}$$

$$\boldsymbol{x}_{ij}(k+1) = \boldsymbol{x}_{ij}(k) + \boldsymbol{v}_{ij}(k+1). \tag{1b}$$

Here $\boldsymbol{x}_{ij}(k), \boldsymbol{v}_{ij}(k)$ are the position and the velocity of particle i on dimension j at iteration k, respectively, $i = 1, 2, ..., N, j = 1, ..., n, k = 1, 2, ...$, where N is the number of particles in the swarm. $C_{1,ij} \sim U(0, \phi_1), C_{2,ij} \sim U(0, \phi_2)$, and ω, ϕ_1, ϕ_2 are three important parameters in PSO. ω is the inertial weight, which determines the level of inertial;

*This work was supported by MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No.13YJC630095) and NSF of China (No.11271069).

ϕ_1 and ϕ_2 are often called the learning factors, which determine the amplitudes of affections come from personal best experiment and social best experiment, respectively.

$\boldsymbol{p}_i(k)$ and $\boldsymbol{g}_i(k)$ are the personal best and the neighborhood best of particle i, respectively, that is to say,

$$\boldsymbol{p}_i(k) = \arg\min_{0\le t\le k} f(\boldsymbol{x}_i(t)),\ \ \boldsymbol{g}_i(k) = \arg\min_{l\in N_i} f(\boldsymbol{p}_l(k)),$$

where N_i is the neighborhood of particle i which contains all the particles who interact with particle i. The identification of N_i is connected with the population topology of PSO. There are many different topologies in PSO, gbest topology and lbest topology are popular examples, see (Mendes, 2004) for more details.

Since its first proposal in 1995, PSO has been used across a wide range of applications, see (Poli, 2008) for an analysis of the publications on the applications of PSO. What is more, many variants of PSO have been proposed in order to improve performance of the original PSO algorithm for particular problems or general problems, such as (Clerc and Kennedy, 2002; Kennedy, 2003; Kennedy and Eberhart, 1997; Kennedy and Mendes, 2002; Mendes et al., 2002; 2003; Shi and Eberhart, 1998). See (Poli et al., 2007) for a comprehensive review of the PSO algorithm.

Although PSO has been enormously successful, there is still not enough comprehensive mathematical understanding about the general PSO with randomness. Among these theoretical issues of PSO, stability analysis or convergence analysis has been mostly discussed. Stability analysis of PSO discuss how different essential factors of PSO, especially the parameter combinations and network topology, affect the dynamic of particle swarm, and moreover, in what kind of conditions, particle swarm converges to some constant position. Most existing stability analyses focus on the dynamic behaviors or trajectories of particles, the stable region or convergence region and the sample distributions, see (Blackwell and Bratton, 2008; Blackwell, 2012; Clerc and Kennedy, 2002; Gazi, 2012; Jiang et al., 2007; Kadirkamanathan et al., 2006; Martínez and Gonzalo, 2008; Ozcan and Mohan, 1999; Poli, 2009; Trelea, 2003; van den Bergh and Engelbrecht, 2006) for examples.

In these analyses, the assumption of stagnation is often adopted. Such an assumption requires that each particle's personal best position (the swarm best position is also included) must be constant. It is clear that this assumption is not realistic at all.

Moreover, in these analyses, different definitions of stability are adopted. Some of them are often only suitable for constant parameters, for example (Clerc and Kennedy, 2002; van den Bergh and Engelbrecht, 2006). Some of them adopted conservative theories, for example, the Lyapunov theory adopted in (Gazi, 2012; Kadirkamanathan et al., 2006) is very conservative (Poli et al., 2007), and therefore only sufficient condition can be obtained (Kadirkamanathan et al., 2006). While some other of them are strict, for example, in (Poli, 2009) the (mixed) order-2 original moments are required to be constant over time, while in (Blackwell, 2012) all covariances are required to be constant over time. See Section 2 for more details.

In this paper, we provide a stability analysis of PSO based on the assumption of *weak* stagnation. The weak stagnation only requires that the position of the swarm best particle is constant, while all other particles's best positions are allowed to improve.

We point out that there are several advantages to adopt the weak stagnation assumption. First of all, we will show that the obtained stable region is not worse than the existing ones. This implies that the classical stagnation assumption is too strict and not necessary for stability analysis of PSO. Second, the weak stagnation assumption allows to classify the particles into four different types, in our viewpoint, this is very important for more comprehensive understanding about the particles' behaviors. For example, we can analyze the conditions needed for knowledge diffusion among different type of particles, and analyze the leading behavior of the best particle and the following behavior of other particles, etc.

Based on the weak stagnation assumption and a new definition of stability, we then pay much attention to derive an order-2 stable region. When $\phi_1 = \phi_2$, the obtained stable region is the same as those proposed by several other stable analyses. However, through comprehensive comparison of different definitions of stability, we will show that our conditions of stability are the weakest conditions among the existing stability analyses for canonical PSO and additional conditions bring no benefit.

Although our analyses are suitable for general PSO, we often provide specific results for special PSOs, especially the canonical PSO. In this paper, the canonical PSO means the PSO with parameter ω and $C_1 \sim U(0, \phi_1), C_2 \sim U(0, \phi_2)$. In this sense, both the PSO with inertia weight or the PSO with constriction are canonical PSOs.

Although a stable PSO, whose parameter combination lies inside the stable region, does not guarantee to be an efficient PSO, our experimental results show that, when the required accuracy is low, a stable PSO often performs better than an unstable PSO. In this sense, we say that the obtained stable region is meaningful. Our experimental results also show that, there exists at least one good parameter combination inside the stable region which is better than some known best parameter combinations on a large set of benchmark problems.

We organize the rest of this paper as follows. In section 2, we review some existing stability analyses of PSO, especially their definitions of stability and the resulted stable regions. In section 3, we propose the assumption of weak stagnation and provide some advantages of weak stagnation in the stability analysis of PSO. In section 4, we derive the stable region of PSO based on a new definition of stability. In section 5, we compare our stability analysis with existing stability analyses. In Section 6, we report some experimental results to show that our stable region is meaningful. In the final section, we give some conclusions and future works.

2 Existing stability analyses

In this section, we review several existing stability analyses of PSO. For our purpose, we mainly focus on their definitions of stability and the resulted stable regions.

Early analyses often assumed that randomness is not present, for example (van den Bergh and Engelbrecht, 2006; Clerc and Kennedy, 2002; Trelea, 2003). Thus a natural definition of stability is

$$\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{y}, \tag{2}$$

where $\boldsymbol{y}$ is a constant vector. Then through solving a deterministic difference equation (e.g., van den Bergh and Engelbrecht (2006)) or a deterministic dynamic system (e.g., Clerc and Kennedy (2002); Trelea (2003)), the following popular stable region

$$\omega \in (0, 1), \;\; \phi_1 + \phi_2 < 2(1 + \omega) \tag{3}$$

or

$$\omega \in (-1, 1), \;\; \phi_1 + \phi_2 < 2(1 + \omega) \tag{4}$$

were obtained. In the rest of this paper, we will not consider the region (3) partly because most literatures adopt $\omega \in (-1, 1)$ and partly for simplicity.

After obtained the stable region (4), there are two approaches to deal with the randomness of C_1, C_2. The first approach is simply replace C_1, C_2 with their maximal values ϕ_1 and ϕ_2, respectively, which results in the same stable region as (4). The second approach is replace C_1, C_2 with their expectations $\phi_1/2$ and $\phi_2/2$ respectively (e.g., Trelea (2003)), which results in the following different stable region

$$\omega \in (-1, 1), \ \ \phi_1 + \phi_2 < 4(1 + \omega). \tag{5}$$

However, when C_1 and C_2 are random variables, $\{\boldsymbol{x}(k)\}$ is a sequence of random variables, thus the definition of stability (2) almost makes no sense. When randomness presents, a popular choice is to let

$$\lim_{t\to\infty} E[\boldsymbol{x}(t)] = \boldsymbol{y}, \tag{6}$$

where $E(\boldsymbol{x})$ is the expectation of random variable $\boldsymbol{x}$, which results in the order-1 stability (Poli, 2009).

Definition 1. *If the condition (6) holds, then we say that PSO is order-1 stable and the corresponding parameter region is its order-1 stable region.*

Hence, both region (4) and region (5) satisfy the order-1 stability. Therefore, we call region (4) and region (5) the order-1 stable regions.

In order to deal with the randomness, in (Kadirkamanathan et al., 2006), the equilibrium of passive system is defined. Then the following sufficient but not necessary stable region

$$\begin{cases} \phi_1 + \phi_2 < 2(1 + \omega), & \omega \in (-1, 0] \\ \phi_1 + \phi_2 < \frac{2(1-\omega)^2}{1+\omega}, & \omega \in (0, 1) \end{cases} \tag{7}$$

is derived using Lyapunov stability analysis.

Unfortunately, Lyapunov theory is very conservative, and so the conditions (7) are extremely restrictive (Martínez and Gonzalo, 2008; Poli et al., 2007). Moreover, the region (7) does not include the best parameters found in PSO literatures (Martínez and Gonzalo, 2008).

Recently, in (Gazi, 2012), the stable region (7) is extended to the following larger region

$$\begin{cases} \phi_1 + \phi_2 < \frac{24(1+\omega)}{7}, & \omega \in (-1, 0] \\ \phi_1 + \phi_2 < \frac{24(1-\omega)^2}{7(1+\omega)}, & \omega \in (0, 1). \end{cases} \tag{8}$$

Because both region (7) and region (8) adopt the Lyapunov condition, we call them the Lyapunov stable regions for convenience.

It has been pointed out in (Jiang et al., 2007; Poli, 2009) that, order-1 stability is not enough to ensure convergence; the order-2 stability condition must also be satisfied to ensure the convergence of the variance or the standard deviation. However, Jiang et al. (2007) and Poli (2009) adopt different conditions to define order-2 stability.

In (Jiang et al., 2007), the following condition

$$\lim_{t\to\infty} E[\boldsymbol{x}(t) - \boldsymbol{y}]^2 = 0 \tag{9}$$

are adopted to ensure the order-2 stability, where $\boldsymbol{y} = \lim_{t\to\infty} E[\boldsymbol{x}(t)]$. This condition requires that $\boldsymbol{x}(t)$ converges in mean square to $\boldsymbol{y}$ (Taboga, 2012). Based on the conditions of convergence in mean square, the following region

$$\frac{5\phi - \sqrt{25\phi^2 - 336\phi + 576}}{24} < \omega < \frac{5\phi + \sqrt{25\phi^2 - 336\phi + 576}}{24} \tag{10}$$

is derived ($\phi_1 = \phi_2 = \phi$).

On the other hand, in (Poli, 2009), the following conditions

$$\lim_{t\to\infty} E[\boldsymbol{x}^2(t)] = \beta_0, \quad \lim_{t\to\infty} E[\boldsymbol{x}(t)\boldsymbol{x}(t-1)] = \beta_1 \tag{11}$$

are adopted to ensure the order-2 stability, where β_0 and β_1 are constant vectors which are not presented explicitly in (Poli, 2009). Then under the classical stagnation assumption and $\phi_1 = \phi_2 = \phi$, the following line

$$\phi = \frac{12(1-\omega^2)}{7-5\omega}$$

was obtained, on which the magnitude of the largest eigenvalue of an important matrix (M in (Poli, 2009)) is one. In other word, (Poli, 2009) presented implicitly an order-2 stable region

$$\phi < \frac{12(1-\omega^2)}{7-5\omega}. \tag{12}$$

If we reformulate (12) as follows

$$12\omega^2 - 5\phi\omega + 7\phi - 12 < 0,$$

then we obtain the same inequalities for ω as (10). Therefore, both (10) and (12) are the same stable region. For convenience, we call both region (10) and region (12) the order-2 stable regions.

Martínez and Gonzalo (2008) proposed a generalized continuous PSO, which extended the definitions of stability and the resulted stable regions proposed in (Trelea, 2003) and (Poli, 2009) to continuous cases.

Recently, a general stochastic difference equation (SDE) of PSO has been proposed in (Blackwell, 2012). The following conditions

$$E[\boldsymbol{x}(t)] = \boldsymbol{y} \tag{13}$$

and

$$E[\boldsymbol{x}(t) - \boldsymbol{y}][\boldsymbol{x}(t-k) - \boldsymbol{y}] = \gamma_k, \quad k = 0, 1, 2, ... \tag{14}$$

are adopted to ensure the order-1 stability and order-2 stability, where $\boldsymbol{y}$ and γ_k are constant vectors which are not presented explicitly. These conditions constitute the weak stationary condition of time series, which require both expectation and all covariances of $\boldsymbol{x}(t)$ are constant over time. For canonical PSO, the following stable region

$$\omega \in (-1, 1), \;\; \phi_1 + \phi_2 < \frac{24(1-\omega^2)}{7-5\omega}. \tag{15}$$

can be derived from the weak stationary condition. For convenience, we call region (15) the weak stationary stable region. It is clear when $\phi_1 = \phi_2$, this region is the same as the order-2 stable region (12).

3 Weak stagnation and its advantages

Existing stability analyses reviewed in the above section often adopted the assumption of stagnation, which requires all particles' best position remain constant. It is not a realistic assumption.

In this paper, we propose to adopt the weak stagnation assumption, whose definition is given as follows.

Definition 2. *If the whole swarm's best position $\mathbf{x}_i(K)$ remain constant since the K-th iteration until the $(K+M)$-th $(M \geq 3)$ iteration, then we say that PSO is in weak stagnation state during iteration $(K, K+M)$, and we call $\mathbf{x}_i(K)$ the stagnation point, where particle i is called the dominant particle during iteration $(K, K+M)$.*

The first advantage of the weak stagnation assumption is that it is much more realistic than the classical stagnation assumption. Since all particles' best positions are required to be constant, it may never happen in reality. However, in the weak stagnation assumption, only the global best (or swarm best) is required to be constant, while all other personal bests are allow to improve. Such a state is often reported in numerical experiments, for example (Evers, 2009). Moreover, if the global optimum is found, then the whole particle swarm stay in the weak stagnation state.

Moreover, the weak stagnation assumption allows to classify the particles into dominant particle and several type of nondominant particles. Such classification allows to analyze the conditions for knowledge sharing among these different type of particles, and analyze the behaviors of different type of particles.

In the rest of this section, we first provide the classification of particles and a simple condition for knowledge sharing among dominant particle and nondominant particle. Then we provide the leading behavior of the dominant particle, which is important for our later analysis.

3.1 Classification of particles and knowledge diffusion

During the weak stagnation state, we can classify all particles into the following four types:

- Type I: the dominant particle;
- Type II: particles which are informed by the dominant particle, i.e., particles whose neighborhood best particle is the dominant particle;
- Type III: neighborhood best particles except the dominant particle;
- Type IV: particles which are informed by the type III particles.

We use $T_i (i = 1, 2, 3, 4)$ to denote the sets of particles belong to these four types, respectively.

We note that the classification of particles is depend on the topology (social network). For example, for the gbest topology, there is no type III and type IV particles, so $T_3 = T_4 = \emptyset$. For other topologies except gbest, four types of particles are often exist.

Given any topology, the type I particle possesses the best experience or knowledge, the type II particles possess worse experiences, and the type IV particles (if exist) possess the worst experiences. In what follows, we discuss the conditions that the type IV particles can share the best experience with the type I particle. In this paper, if nondominant particles can share the best experience with the dominant particle, then we say that the best experience of the dominant particle is diffusible.

Theorem 1. *The best experience of the type I particle is diffusible, if and only if*

$$T_3 = \emptyset$$

or

$$T_2 \cap T_3 \neq \emptyset. \tag{16}$$

Proof. If $T_3 = \emptyset$, then there is only type I particle becomes the informer, i.e., the topology is gbest. In this situation, knowledge diffusion is direct.

If $T_3 \neq \emptyset$, then the topology must not be gbest. In order to guarantee the type IV particles share knowledge with the type I particle, T_2 must be intersect with T_3, i.e., $T_2 \cap T_3 \neq \emptyset$. In other words, after type I particle informs its knowledge to type II particles, the latter need to inform the knowledge to type IV particles. If there is no type II particle becomes the type III particle, then the knowledge diffusion is impossible. □

We note that, when $T_3 \neq \emptyset$, the condition (16) may hold for some iterations and not hold for some other iterations. In other words, knowledge diffuse from type I particle to type IV particles may be off and on during weak stagnation.

When the knowledge diffusion is off for some iteration, then all neighborhoods composed of type III and IV particles are fully independent on the best neighborhood composed of type I and II particles. Thus the diversity of particle swarm can be benefited from the off of knowledge diffusion.

3.2 Leading behavior of the dominant particle

From dynamic equations (1a) and (1b), we can obtain the following dynamic equation for particle i on dimension j

$$\boldsymbol{x}_{ij}(k+1) = \alpha_1 \boldsymbol{x}_{ij}(k) + \alpha_2 \boldsymbol{x}_{ij}(k-1) + \alpha_3 \boldsymbol{p}_{ij}(k) + \alpha_4 \boldsymbol{g}_{ij}(k), \tag{17}$$

where $i = 1, ..., N$, $j = 1, ..., n$ and

$$\begin{array}{rcl} \alpha_1 & = & 1 + \omega - C_{1,ij} - C_{2,ij}, \\ \alpha_2 & = & -\omega, \\ \alpha_3 & = & C_{1,ij}, \\ \alpha_4 & = & C_{2,ij}. \end{array} \tag{18}$$

We note that, here and after, $\alpha_1, \alpha_2, \alpha_3$ and α_4 are abbreviations of $\alpha_{1,ij}(k)$, $\alpha_{2,ij}(k)$, $\alpha_{3,ij}(k)$ and $\alpha_{4,ij}(k)$ for convenience, respectively. If more than one iteration are applied, we use $\alpha_1(k)$ to denote the value of α_1 at iteration k.

The stochastic difference equation (17) has also been used in several literatures in PSO, such as (Blackwell and Bratton, 2008; Jiang et al., 2007; Martínez and Gonzalo, 2008). In (Blackwell, 2012), a general stochastic difference equation is extended from (17).

Although PSO has a simple theoretical structure, the SDE (17) is not simple at all in that the global best $\boldsymbol{g}$ and the personal best $\boldsymbol{p}$ are dynamic. Moreover, the coefficients $\alpha_i (i = 1, 3, 4)$ are often random variables, which make the SDE (17) hard to solve.

From (18) we can see that $\alpha_i (i = 1, .., 4)$ always satisfy the following condition

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1. \tag{19}$$

This implies that the value of $\boldsymbol{x}_{ij}$ at iteration $k+1$ is always a random convex combination of its current value $\boldsymbol{x}_{ij}(k)$, its past value $\boldsymbol{x}_{ij}(k-1)$, its current personal best value $\boldsymbol{p}_{ij}(k)$ and its current neighborhood best value $\boldsymbol{g}_{ij}(k)$.

Condition (19) implies that $\sum_{i=1}^{4} \alpha_i - 1 = 0$. In another word, the sum of the coefficients of the SDE (17) always equals 0. This condition is helpful in solving the SDE (17).

In general cases, the index of the dominant particle is changing as the iteration k increasing. However, when the whole swarm is in weak stagnation state, the dominant particle will be unchanged.

Suppose that PSO is in weak stagnation state during iteration $(K, K+M)$, and particle d is the dominant particle. Then particle d's personal best equals its neighborhood best and equals the swarm's best. So we have $\boldsymbol{p}_d(k) = \boldsymbol{g}_d(k) = \boldsymbol{x}_d(K),\ k \in [K, K+M)$, where $\boldsymbol{x}_d(K)$ is the stagnation point. Thus for the dominant particle d, we have the following simpler SDE

$$\boldsymbol{x}_d(k+1) = \alpha_1 \boldsymbol{x}_d(k) + \alpha_2 \boldsymbol{x}_d(k-1) + \tilde{\alpha}_3 \boldsymbol{x}_d(K), \tag{20}$$

for $k \in [K, K+M)$, where $\tilde{\alpha}_3 = \alpha_3 + \alpha_4$. We note that, here and after, in order to ease notation the dimension index j has been dropped.

Because $\alpha_1, \tilde{\alpha}_3$ are random variables, we can not solve (20) by solving its characteristic equation. However, we can solve it by looking for the recursion of solutions.

Theorem 2. *Suppose that PSO is in weak stagnation during iteration $(K, K+M)$, and $\mathbf{x}_d(K)$ is the stagnation point, particle d is the dominant particle. Then the solutions of the SDE (20) satisfy the following formula*

$$\mathbf{x}_d(K+t) = \mathbf{x}_d(K) + R(t)(\mathbf{x}_d(K+1) - \mathbf{x}_d(K)),\ \ t \in [0, M), \tag{21}$$

where $R(t)$ satisfies the following recursion

$$R(t+1) = \alpha_1(K+t)R(t) - \omega R(t-1), \quad R(0) = 0,\ \ R(1) = 1. \tag{22}$$

Proof. We prove Theorem 2 by reduction. When $t = 0$ and $t = 1$, (21) implies that

$$\boldsymbol{x}_d(K) = \boldsymbol{x}_d(K).$$

$$\boldsymbol{x}_d(K+1) = \boldsymbol{x}_d(K) + (\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K)).$$

It is clearly true.

Suppose that (21) holds for $t = k-1$ and $t = k$, we have

$$\boldsymbol{x}_d(K+k-1) = \boldsymbol{x}_d(K) + R(k-1)(\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K)),$$

$$\boldsymbol{x}_d(K+k) = \boldsymbol{x}_d(K) + R(k)(\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K)),$$

Then we need to prove that (21) holds for $t = k+1$.

From (20), we have

$$\boldsymbol{x}_d(K+k+1) = \alpha_1(K+k+1)\boldsymbol{x}_d(K+k) + \alpha_2 \boldsymbol{x}_d(K+k-1) + \tilde{\alpha}_3 \boldsymbol{x}_d(K).$$

Combining with $\alpha_2 = -\omega$ and $\alpha_1 + \alpha_2 + \tilde{\alpha}_3 = 1$, we obtain

$$\boldsymbol{x}_d(K+k+1) = \boldsymbol{x}_d(K) + R(k+1)(\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K)),$$

where $R(k+1)$ satisfies $R(k+1) = \alpha_1(K+k+1)R(k) - \omega R(k-1)$. $\square$

If we let $k = K$ in (20), and combining with $\alpha_2 = -\omega$, we obtain

$$\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K) = \omega(\boldsymbol{x}_d(K) - \boldsymbol{x}_d(K-1)).$$

Combing with (21), we find that, any point sampled by the dominant particle during weak stagnation depends on not only the stagnation point $\boldsymbol{x}_d(K)$ but also $\boldsymbol{x}_d(K+1)$ or $\boldsymbol{x}_d(K-1)$.

From (21), we can draw the following conclusions about the leading behavior of the dominant particle.

- The leading behavior of the dominant particle is mainly controlled by $R(t)$, while $R(t)$ is mainly affected by the value of α_1.
- If $R(t)$ is the same on each dimension (e.g., parameters C_1, C_2 are constant), then all points sampled by the dominant particle will lie on a line identified by $\boldsymbol{x}_d(K+1)$ and $\boldsymbol{x}_d(K)$.
- In general cases, $R(t)$ is different on different dimension, hence the points sampled by the dominant particle distribute randomly around $\boldsymbol{x}_d(K)$. The exact distribution depend on α_1.

  We will analyze $R(t)$ further in next section.

4 Stability analysis of PSO

In this section, we first propose a new definition of order-2 stability, and then derive the order-2 stable region.

In the rest of this paper, the probability of a random event A is denoted as $P\{A\}$, and the variance of a random variable X is denoted as $D(X)$.

4.1 Definition of stability

Definition 3. *Suppose that the swarm is in weak stagnation state forever. Then PSO is said to be stable if*

$$\lim_{t\to\infty} D[\mathbf{x}_d(K+t)] = 0, \tag{23}$$

where $\mathbf{x}_d(K)$ is the stagnation point, and particle d is the dominant particle.

Remark 1. *Definition 3 implies implicitly that the stability of PSO is determined only by the dominant particle. Although for some iterations, $T_2 \cap T_3$ may equal empty set, thus there are some particles whose evolution are independent on the dominant particle. However, as long as PSO stays in the weak stagnation state for enough iterations (this will happen when the global optimum is found), all particles will be attracted by the dominant particle. Therefore, the implicit assumption is realistic, at least in the sense of limitation.*

Because the Definition 3 requires the variance converges to zero, it satisfies the order-2 stability. What is more, the following theorem shows that it also satisfies the order-1 stability.

Theorem 3. *The Definition 3 satisfies the order-1 stability. Especially, if $P\{C_1+C_2 \neq 0\} = 1$, i.e., $C_1 + C_2$ does not equal 0 almost surely, then*

$$\lim_{t\to\infty} E[\mathbf{x}_d(K+t)] = \mathbf{x}_d(K). \tag{24}$$

Proof. Since

$$D[\boldsymbol{x}_d(K+t)] = E[\boldsymbol{x}_d(K+t) - E(\boldsymbol{x}_d(K+t))]^2,$$

$\lim_{t\to\infty} D[\boldsymbol{x}_d(K+t)] = 0$ implies that $E[\boldsymbol{x}_d(K+t)]$ exists almost surely for all $t > 0$, and moreover, there exists a constant vector $\boldsymbol{y}$ such that $\boldsymbol{x}_d(K+t)$ converges to the constant vector $\boldsymbol{y}$ almost surely (Taboga, 2012). In other word, we have

$$P\{\lim_{t\to\infty} \boldsymbol{x}_d(K+t) = \boldsymbol{y}\} = 1,$$

therefore,

$$\lim_{t\to\infty} E[\boldsymbol{x}_d(K+t)] = \boldsymbol{y},$$

i.e., the Definition 3 satisfies the order-1 stability.

In the following, we will prove that $\boldsymbol{y}$ has to equal $\boldsymbol{x}_d(K)$. We prove with contradiction.

On one hand, $\boldsymbol{x}_d(K+t)$ converges to constant vector implies that, if we take limitation in equation (1a) (the i, j have been dropped for convenience), we obtain

$$\lim_{t\to\infty} \boldsymbol{v}_d(K+t) = \frac{(C_1+C_2)(\boldsymbol{x}_d(K)-\boldsymbol{y})}{1-\omega}.$$

On the other hand, if we take limitation in equation (1b), we obtain

$$\lim_{t\to\infty} \boldsymbol{v}_d(K+t) = 0.$$

Therefore, if $P\{C_1+C_2 \neq 0\} = 1$, then we have $\lim\limits_{t\to\infty} E[\boldsymbol{x}_d(K+t)] = \boldsymbol{y} = \boldsymbol{x}_d(K)$. $\square$

Hence, our definition of stability satisfies both the order-1 stability and the order-2 stability. Moreover, we have the following corollary, which can be derived straightforward from Theorem 2 and the equation (24).

Corollary 1. *Suppose that the swarm is in weak stagnation state forever and PSO is stable. If* $P\{C_1+C_2 \neq 0\} = 1$, *then*

$$\lim_{t\to\infty} E[R(t)] = 0. \tag{25}$$

The following lemma provides an easier way to verify whether PSO is stable or not.

Lemma 1. *Suppose that the swarm is in weak stagnation state forever. Then PSO is stable if and only if*

$$\lim_{t\to\infty} D[R(t)] = 0. \tag{26}$$

Proof. According to Theorem 2, during the weak stagnation state, the dominant particle has the following dynamic behavior

$$\boldsymbol{x}_d(K+t) = \boldsymbol{x}_d(K) + R(t)(\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K)).$$

Because the stagnation point $\boldsymbol{x}_d(K)$ and $\boldsymbol{x}_d(K+1)$ are constant,

$$D[\boldsymbol{x}_d(K+t)] = D[R(t)][\boldsymbol{x}_d(K+1) - \boldsymbol{x}_d(K)]^2.$$

Thus $\lim\limits_{t\to\infty} D[\boldsymbol{x}_d(K+t)] = 0$ is equivalent to $\lim\limits_{t\to\infty} D[R(t)] = 0$. $\square$

Since $E[R^2(t)] = D[R(t)] + [E(R(t))]^2$, the following corollary can be derived straightforward from Lemma 1 and Corollary 1.

Corollary 2. *Suppose that the swarm is in weak stagnation state forever. If* $P\{C_1+C_2 \neq 0\} = 1$, *then PSO is stable if and only if*

$$\lim_{t\to\infty} E[R^2(t)] = 0. \tag{27}$$

For canonical PSO and many of its variants, C_1+C_2 always greater than 0, so the condition $P\{C_1+C_2 \neq 0\} = 1$ always holds. Therefore, in the rest of this paper, we suppose that the following assumption holds.

Assumption 1. *The PSO parameters* C_1, C_2 *satisfy* $P\{C_1+C_2 \neq 0\} = 1$.

In the following subsections, we compute $E[R^2(t)]$ and then derive the stable region. First, denote

$$\mu = E(\alpha_1), \;\; \sigma^2 = D(\alpha_1). \tag{28}$$

Especially, for canonical PSO, we have

$$\mu = 1 + \omega - \frac{\phi_1 + \phi_2}{2}, \;\; \sigma^2 = \frac{\phi_1^2 + \phi_2^2}{12}.$$

4.2 Computing $E[R^2(t)]$

From (22) we have

$$R^2(t+1) = \alpha_1^2(K+t)R^2(t) + \omega^2 R^2(t-1) - 2\omega\alpha_1(K+t)R(t)R(t-1),$$

and

$$R(t+1)R(t) = \alpha_1(K+t)R^2(t) - \omega R(t)R(t-1).$$

So we have

$$E[R^2(t+1)] = (\mu^2+\sigma^2)E[R^2(t)] + \omega^2 E[R^2(t-1)] - 2\mu\omega E[R(t)R(t-1)], \tag{29}$$

and

$$E[R(t+1)R(t)] = \mu E[R^2(t)] - \omega E[R(t)R(t-1)]. \tag{30}$$

From (29), we can obtain $E[R(t)R(t-1)]$ and $E[R(t+1)R(t)]$, which can be substituted into (30). Finally, we have the following difference equation

$$E[R^2(t+2)] + (\omega-\mu^2-\sigma^2)E[R^2(t+1)] + (\omega\mu^2-\omega\sigma^2-\omega^2)E[R^2(t)] - \omega^3 E[R^2(t-1)] = 0, \tag{31}$$

and satisfies

$$E[R^2(0)] = 0, E[R^2(1)] = 1, E[R^2(2)] = \mu^2+\sigma^2. \tag{32}$$

The characteristic equation of (31) is

$$\Phi(r) = r^3 + (\omega-\mu^2-\sigma^2)r^2 + (\omega\mu^2-\omega\sigma^2-\omega^2)r - \omega^3 = 0, \tag{33}$$

which is the same as in (Jiang et al., 2007). Through solving the characteristic equation, we can obtain three solutions of this characteristic equation. In order to ensure $\lim_{t\to\infty} E[R^2(t)] = 0$, let the norm of all these three solutions less than 1, then we obtain the following theorem, see Appendix for its proof.

Theorem 4. *Suppose that $E[R^2(t)]$ is the solution of (31), then the sufficient and necessary condition for $\lim_{t\to\infty} E[R^2(t)] = 0$ is*

$$(1-\omega)\mu^2 + (1+\omega)\sigma^2 < (1+\omega)^2(1-\omega). \tag{34}$$

4.3 Stable region

From Corollary 2 and Theorem 4, we can easily obtain the stable region.

Theorem 5. *The PSO is stable if and only if the parameters ω, ϕ_1 and ϕ_2 satisfy the condition*

$$\begin{cases} \omega \in (-1,1) \\ (1-\omega)\mu^2 + (1+\omega)\sigma^2 < (1+\omega)^2(1-\omega), \end{cases} \tag{35}$$

where μ, σ^2 are defined in (28).

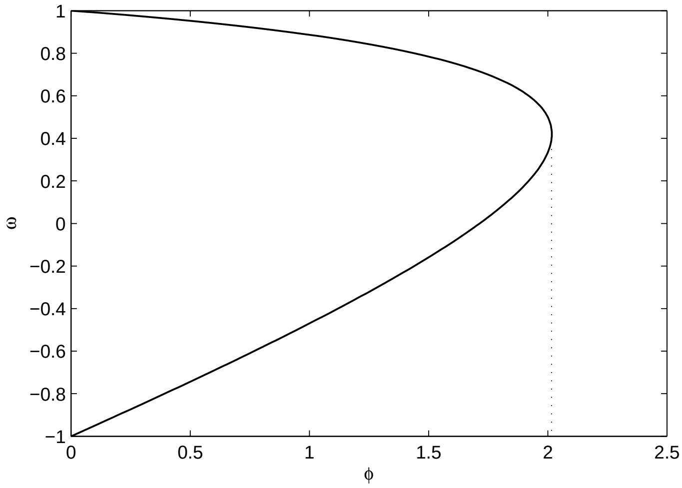

Figure 1: Stable region of canonical PSO ($\phi_1 = \phi_2 = \phi$).

Remark 2. *We note that the stable region (35) is suitable for any PSO whose SDE is (17), where C_1, C_2 are allow to be any random variables satisfying $P\{C_1 + C_2 \neq 0\} = 1$.*

A similar result has been proposed in (Blackwell, 2012) for more general PSOs. However, stricter conditions are needed to derive the equation (14) in (Blackwell, 2012). See Section 5 for more details.

Corollary 3. *The canonical PSO is stable if and only if the parameters satisfy*

$$\omega \in (-1, 1)$$

and

$$3(1-\omega)(\phi_1+\phi_2)^2 + (1+\omega)(\phi_1^2+\phi_2^2) - 12(1-\omega^2)(\phi_1+\phi_2) < 0.$$

Proof. Let

$$\mu = 1 + \omega - \frac{\phi_1+\phi_2}{2}, \quad \sigma^2 = \frac{\phi_1^2+\phi_2^2}{12},$$

we obtain the conditions straightforward. □

Corollary 4. *The canonical PSO with $\phi_1 = \phi_2 = \phi$ is stable if and only if*

$$\begin{cases} \omega \in (-1, 1) \\ \phi \in \left(0, \dfrac{12(1-\omega^2)}{7-5\omega}\right). \end{cases} \tag{36}$$

It is clear that the region (36) is the same as the order-2 stable regions (10), (12) and (15). Figure 1 shows the stable region (36). Any pair of parameters (ω, ϕ) in this region can guarantee the order-2 stability of PSO. What is more, ϕ achieves its maximal value 2.0170 when $\omega = 1.4 - \sqrt{0.96} \approx 0.42$. In our numerical experiments (see Section 6), the best parameter combination ($\omega = 0.42, \phi = 1.55$) lies on the line $\omega = 0.42$.

At the end of this section, we provide the following remark to show what kind of topology can be applied by our order-2 stability analysis.

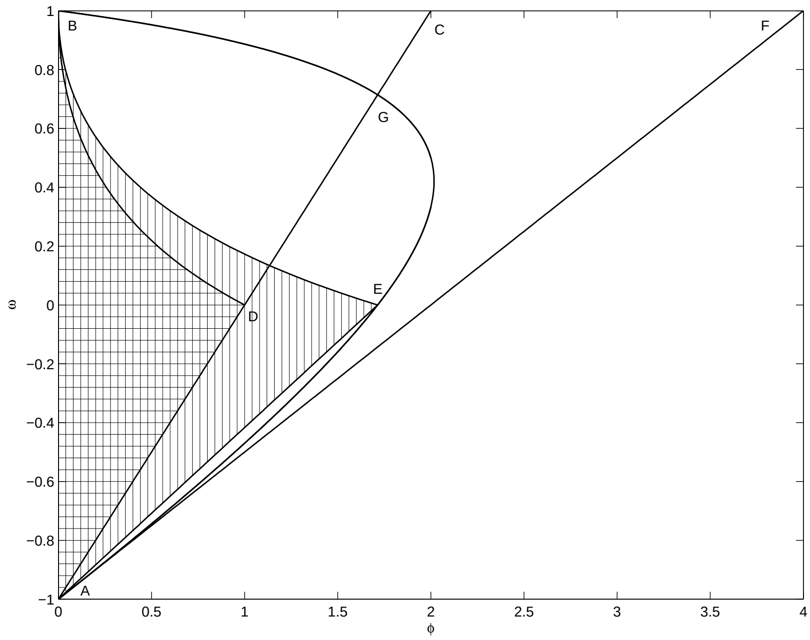

Figure 2: All stable regions discussed in this paper ($\phi_1 = \phi_2 = \phi$). The triangle region ABC and ABF are the order-1 stable regions; the concave triangle ABD (covered with horizontal lines) and ABE (covered with vertical lines) are the Lyapunov stable regions; the convex ABGE region is the order-2 stable region.

Remark 3. *Since the order-2 stability analysis proposed in this paper only need to analyze the leading behavior of the dominant particle, therefore it can be implied to any topology.*

In next section, we compare our stability analysis with some existing stability analyses mentioned in Section 2.

5 Comparison and discussion

In this section, we compare different definitions of stability mentioned in Section 2 and the stable regions resulted from these definitions. Since existing stability analyses are often implied to different PSO models, some of them are very general, for example in (Blackwell, 2012), it is difficult or unnecessary to compare them in general cases. In this section, we will focus on the canonical PSO with $\phi_1 = \phi_2 = \phi$.

First, we put all stable regions discussed in this paper together in Figure 2. In Figure 2, the triangle region ABC and ABF are the order-1 stable regions (4) and (5), respectively; the concave triangle ABD (covered with horizontal lines) and ABE (covered with vertical lines) are the Lyapunov stable regions (7) and (8), respectively; the convex ABGE region is the order-2 stable region (36). Not surprising, the order-2 stable region is larger than the Lyapunov stable regions but smaller than the order-1 stable region ABF.

Since it has been pointed out in (Jiang et al., 2007; Poli, 2009) that, the order-1 stability is not enough to ensure the convergence of PSO, we will focus on comparing the order-2 stable regions. Specifically, we mainly compare the order-2 stability analyses proposed in (Blackwell, 2012; Jiang et al., 2007; Poli, 2009) with the analysis proposed in this paper.

First of all, we must point out that, although these four stability analyses adopt different assumptions of stagnation and different definitions of stability, for canonical PSO with $\phi_1 = \phi_2 = \phi$, they bring the same order-2 stable region (36). This fact results in the following proposition straightforward.

Proposition 1. *The weak stagnation assumption is enough to derive an order-2 stable region for canonical PSO with $\phi_1 = \phi_2 = \phi$. In another word, the classical stagnation assumption is too strict and not necessary.*

In (Blackwell, 2012; Jiang et al., 2007; Poli, 2009), under the classical stagnation assumption, each particle behaves independently and each dimension is treated independently. Then their analyses were implied to any dimension of arbitrary particle. However, such a state maybe never happen, so the analyses based on the classical stagnation assumption are meaningless.

Fortunately, Proposition 1 shows that, similar analysis can still imply under the weak stagnation assumption, which is much more realistic. Specifically, under weak stagnation state, the dominant particle behaves independently, whose behavior can also result in an order-2 stable region. We note that, Jiang et al. (2007) also extended their analysis to a state where only the swarm best position is fixed while other particle's best positions are allowed to update.

Now we turn to compare the definitions of stability.

In (Poli, 2009), the conditions of stability include (6) and (11), which require not only expectation and variance but also the covariance $E[\boldsymbol{x}(t) - \boldsymbol{y}][\boldsymbol{x}(t-1) - \boldsymbol{y}]$ have limitations. In (Blackwell, 2012), all the covariances are required to keep constants over time (see (14)), which are stricter conditions. However, in both (Jiang et al., 2007) and this paper, there is no request about the covariances. Moreover, from our definition of stability and Theorem 3, the variance converges to zero can ensure not only the order-2 stability but also the order-1 stability. Therefore we have the following proposition.

Proposition 2. *Definition 3 is enough to derive an order-2 stable region for canonical PSO with $\phi_1 = \phi_2 = \phi$. In another word, let the variance of the dominant particle's position converges to zero is sufficient for deriving an order-2 stable region, and other conditions is not necessary.*

Then we turn to compare the definition of stability (see also (9)) proposed in (Jiang et al., 2007) with that proposed in this paper.

First, let $\boldsymbol{y} = \lim_{t\to\infty} E[\boldsymbol{x}(t)]$, then we have the following relationship

$$\begin{aligned} E[\boldsymbol{x}(t) - \boldsymbol{y}]^2 &= E\big[\boldsymbol{x}(t) - E[\boldsymbol{x}(t)] + E[\boldsymbol{x}(t)] - \boldsymbol{y}\big]^2 \\ &= D[\boldsymbol{x}(t)] + (\boldsymbol{y} - E[\boldsymbol{x}(t)])^2. \end{aligned}$$

This implies that $E[\boldsymbol{x}(t) - \boldsymbol{y}]^2 > D[\boldsymbol{x}(t)]$ often holds. However, in the limitation state, we have

$$\lim_{t\to\infty} E[\boldsymbol{x}(t) - \boldsymbol{y}]^2 = \lim_{t\to\infty} D[\boldsymbol{x}(t)].$$

Therefore, the definition of stability proposed in (Jiang et al., 2007) is equivalent with that proposed in this paper. So not surprisingly, they bring the same stable region.

However, since $E[\boldsymbol{x}(t) - \boldsymbol{y}]^2 > D[\boldsymbol{x}(t)]$ almost always holds, our definition can be regarded as an extension of the definition proposed in (Jiang et al., 2007). In order to

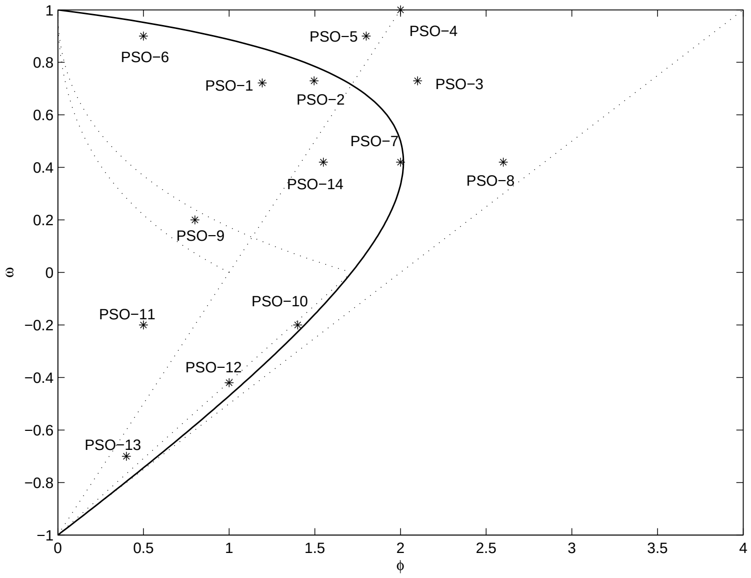

Figure 3: Distribution of 14 selected parameter combinations, 10 stable and 4 unstable. Each star denotes a parameter combination.

compute $E[\boldsymbol{x}(t) - \boldsymbol{y}]^2$, we need to compute $\lim_{t\to\infty} E[\boldsymbol{x}(t)]$ and $D[\boldsymbol{x}(t)]$. In this sense, our definition is easier to compute than that proposed in (Jiang et al., 2007). Moreover, our definition need not discuss the order-1 stability, which is guaranteed automatically (see Theorem 3).

The above comparisons show that, our definition of stability can be regarded as an extension of all the definitions of stability proposed in (Blackwell, 2012; Jiang et al., 2007; Poli, 2009). Although our definition adopts the weakest conditions of stability, for canonical PSO with $\phi_1 = \phi_2$, it is enough for deriving an order-2 stable region, additional conditions bring no benefit.

Remark 4. *Due to Theorem 5 and Remark 2, Proposition 1 and Proposition 2 can be extend to more general PSOs. In other words, the weak stagnation assumption and Definition 3 are enough to derive an order-2 stable region for more general PSOs.*

6 Experimental results

In this section, we report some experimental results. Our main purpose is to test

- whether a stable PSO performs better than an unstable PSO ? and
- is there any parameter combination that is better than some known best parameter combinations ?

6.1 Algorithm configurations and test problems

Specifically, we test 14 parameter combinations of the canonical PSO which adopt the gbest topology with 20 particles in the swarm. The only difference among them is that

Table 1: 14 parameter combinations of PSO tested in this paper.

algorithm	PSO-1	PSO-2	PSO-3	PSO-4	PSO-5	PSO-6	PSO-7
ω	$1/(2\ln 2)$	0.7298	0.7298	1	0.9	0.9	0.42
ϕ	$0.5+\ln 2$	1.49618	2.1	2	1.8	0.5	2
algorithm	PSO-8	PSO-9	PSO-10	PSO-11	PSO-12	PSO-13	PSO-14
ω	0.42	0.2	-0.2	-0.2	-0.42	-0.7	0.42
ϕ	2.6	0.8	1.4	0.5	1	0.4	1.55

Table 2: Information of the 50 CEC2005 test functions.

function name	d	minimal value	search space
CEC2005 F1	$10, 30$	-450	$[-100, 100]^d$
CEC2005 F2	$10, 30$	-450	$[-100, 100]^d$
CEC2005 F3	$10, 30$	-450	$[-100, 100]^d$
CEC2005 F4	$10, 30$	-450	$[-100, 100]^d$
CEC2005 F5	$10, 30$	-310	$[-100, 100]^d$
CEC2005 F6	$10, 30$	390	$[-100, 100]^d$
CEC2005 F7	$10, 30$	-180	$[0, 600]^d$
CEC2005 F8	$10, 30$	-140	$[-32, 32]^d$
CEC2005 F9	$10, 30$	-330	$[-5, 5]^d$
CEC2005 F10	$10, 30$	-330	$[-5, 5]^d$
CEC2005 F11	$10, 30$	90	$[-0.5, 0.5]^d$
CEC2005 F12	$10, 30$	-460	$[-100, 100]^d$
CEC2005 F13	$10, 30$	-130	$[-3, 1]^d$
CEC2005 F14	$10, 30$	-300	$[-100, 100]^d$
CEC2005 F15	$10, 30$	120	$[-5, 5]^d$
CEC2005 F16	$10, 30$	120	$[-5, 5]^d$
CEC2005 F17	$10, 30$	120	$[-5, 5]^d$
CEC2005 F18	$10, 30$	10	$[-5, 5]^d$
CEC2005 F19	$10, 30$	10	$[-5, 5]^d$
CEC2005 F20	$10, 30$	10	$[-5, 5]^d$
CEC2005 F21	$10, 30$	360	$[-5, 5]^d$
CEC2005 F22	$10, 30$	360	$[-5, 5]^d$
CEC2005 F23	$10, 30$	360	$[-5, 5]^d$
CEC2005 F24	$10, 30$	260	$[-5, 5]^d$
CEC2005 F25	$10, 30$	260	$[-2, 5]^d$

Table 3: Information of the 40 non-CEC2005 test functions.

function name (abbreviation)	d	minimal value	search space
Sphere (f26)	10, 30	0	$[-100, 100]^d$
Rastrigin (f27)	10, 30	0	$[-5.12, 5.12]^d$
Six Hump Camel Back (f28)	2	-1.031628	$[-5, 5]^d$
Step (f29)	10, 30	0	$[-100, 100]^d$
Rosenbrock (f30)	10, 30	0	$[-2, 2]^d$
Ackley (f31)	10, 30	0	$[-32, 32]^d$
Griewank (f32)	10, 30	0	$[-600, 600]^d$
Salomon (f33)	10, 30	0	$[-100, 100]^d$
Normalized Schwefel (f34)	10, 30	-418.9828872724338	$[-512, 512]^d$
Quartic (f35)	10, 30	0	$[-1.28, 1.28]^d$
Rotated hyper-ellipsoid (f36)	10, 30	0	$[-100, 100]^d$
Norwegian (f37)	10, 30	1	$[-1.1, 1.1]^d$
Alpine (f38)	10, 30	0	$[-10, 10]^d$
Branin (f39)	2	0.397887	$[-5, 15]^d$
Easom (f40)	2	-1	$[-100, 100]^d$
Goldstein Price (f41)	2	3	$[-2, 2]^d$
Shubert (f42)	2	-186.7309	$[-10, 10]^d$
Hartmann (f43)	3	-3.86278	$[0, 1]^d$
Shekel (f44)	4	-10.5364	$[0, 10]^d$
Levy (f45)	10, 30	0	$[-10, 10]^d$
Michalewicz (f46)	10	-9.66015	$[0, \pi]^d$
Shifted Griewank (f47)	10, 30	-180	$[-600, 600]^d$
Design of a Gear Train (f48)	4	$2.7e-12$	$[12, 60]^d$
Pressure Vessel (f49)	4	7197.72893	[1.125 0.625 0 0] $\times$[12.5 12.5 240 240]
Tripod (f50)	2	0	$[-100, 100]^d$
Compression Spring (f51)	3	2.6254214578	[1 0.6 0.207] $\times$[70 3 0.5]

they adopt different parameter combinations $(\omega, \phi_1 = \phi_2 = \phi)$. Table 1 shows the parameter combinations, whose distribution is illustrated in Figure 3.

Among these 14 parameter combinations, there are 4 unstable combinations (PSO-3, PSO-4, PSO-5, PSO-8) and 10 stable combinations. Among these stable combinations, PSO-2 adopts the parameter combination proposed in (Eberhart and Shi, 2000), and PSO-1 adopts the parameter combination proposed in the standard PSO (Clerc, 2011). Both combinations are popular in PSO literatures and can be regarded as two of the known best parameter combinations. Among the other 8 stable combinations, half of them (PSO-6, PSO-7, PSO-9, PSO-14) lie inside the region $\omega \in (0, 1)$ while another half of them (PSO-10, PSO-11, PSO-12, PSO-13) lie inside the region $\omega \in (-1, 0)$.

Our test set include 25 CEC2005 test functions and 26 other test functions. These 51 test functions are used to test the standard PSO 2011, whose codes are obtained from Professor Maurice Clerc. Table 2 and Table 3 show important information of the CEC2005 test functions and the non-CEC2005 test functions, respectively, where d is the dimension of test function. Totally, we test 90 functions.

6.2 Data profile technique

In order to compare the performance of different PSOs, we adopt the data profile technique (Moré and Wild, 2009) which is often used to compare the derivative free optimization (DFO) algorithms. Because PSO is also a DFO algorithm, the data profile technique is suitable and very convenient for our purpose.

Specifically, we use each PSO algorithm $s \in \mathcal{S}$ to solve each test problem $p \in \mathcal{P}$ for 50 runs, where $\mathcal{S}$ denotes the set of 14 PSO algorithms and $\mathcal{P}$ denotes the set of 90 test problems. In each run, the algorithm does not stop until 3100 function evaluations are consumed. We memorize the history of the found minimal function values. That is to say, in each run, we get a vector whose length is 3100, the k-th element is the minimal function value found during k function evaluations. After 50 runs are finished, we compute the component-wise average of the 50 vectors. Then the average vector can be regarded as a measure of the average behavior of the algorithm s on the problem p.

After all tests are finished, we obtain a $3100 \times 90 \times 14$ matrix. The data profile technique seeks to display these raw data properly.

In (Moré and Wild, 2009), the data profiles are defined as the following cumulative distribution function

$$d_s(\kappa) = \frac{1}{|\mathcal{P}|}\text{size}\left\{p \in \mathcal{P} : \frac{t_{p,s}}{n_p + 1} \leq \kappa\right\}, \tag{37}$$

where $|\mathcal{P}|$ denotes the cardinality of $\mathcal{P}$, n_p is the dimension of the problem p, and $t_{p,s}$ is the number of function evaluations needed to find a position x such that the following convergence condition holds

$$f(x_0) - f(x) \geq (1 - \tau)(f(x_0) - f_L). \tag{38}$$

Here x_0 is the initial best position of the whole swarm and it is fixed in each run for each algorithm; $\tau > 0$ is a tolerance and f_L is the smallest objective function value obtained by any algorithm within a given number $\mu_f = 3100$ of function evaluations. $t_{p.s} = \infty$ if the condition (38) is not satisfied after μ_f evaluations.

It can be seen from condition (38) that, tolerance τ determines the accuracy of solution x to f_L. Low tolerance τ means that a solution with high accuracy to f_L is required. In practice, $\tau = 10^{-k}, k = 1, 2, 3, 4$ are often used. Therefore, in this paper, we often say that the required accuracy is high when $\tau \leq 10^{-7}$ while low when $\tau \geq 10^{-1}$.

An advantage of data profile is that $d_s(\kappa)$ can be interpreted as the percentage of problems that can be solved with the equivalent of κ simplex gradient estimates. The reason is that $n_p + 1$ refers to the number of function evaluations needed to compute a one-side finite difference estimate of the gradient. The reason why we set $\mu_f = 3100$ is that it is enough to compute at least 100 simplex gradients for the largest dimension ($d = 30$) problem. The budget of 100 simplex gradients is proposed in (Moré and Wild, 2009).

In (Moré and Wild, 2009), the performance profile technique is also proposed to work with the data profile technique. The performance profile of an algorithm s is defined as the following cumulative distribution function

$$\rho_s(\alpha) = \frac{1}{|\mathcal{P}|}\text{size}\{p \in \mathcal{P} : r_{p,s} \leq \alpha\}, \tag{39}$$

where the performance ratio $r_{p,s}$ is defined by

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}, s \in \mathcal{S}\}}. \tag{40}$$

We can see that, $\rho_s(1)$ is the fraction of problems for which algorithm s performs the best, while for sufficient large α, $\rho_s(\alpha)$ is the fraction of problems solved by s. Moreover, $\rho_s(+\infty) = d_s(+\infty)$, which links the data profile and the performance profile.

6.3 Stability and efficiency

We now deal with the first question proposed in the beginning of this section.

Figure 4 and Figure 5 show the data profiles and the performance profiles when $\tau = 10^{-1}$, respectively. From Figure 4 and Figure 5 we can see clearly that, PSO-14, PSO-1, PSO-12, PSO-10, PSO-6 are the best PSO algorithms, they can solve at least 80% problems, while PSO-3, PSO-4, PSO-5, PSO-8 are the worst PSO algorithms, they only solve no more than 30% problems. It is interesting that, all the unstable PSOs perform worse than any stable PSO algorithm. For example, PSO-8 is the best among the unstable algorithms, however, it only solve about 27% problems. While even the worst stable algorithms (PSO-11) can solve about 54% problems. The performance difference between the worst stable PSO algorithm (PSO-11) and the best unstable PSO algorithm (PSO-8) is about $27\%(= 54\% - 27\%)$.

Now we increase the accuracy, i.e., decrease τ (see the convergence condition (38)). As the accuracy increases, the number of problems solved by PSO decreases, hence both the data profile and the performance profile become lower. For example, when $\tau = 10^{-7}$, many PSOs can solve no more than 10% problems. Therefore we only present the best five PSOs' profiles.

Figure 6 and Figure 7 show the data profiles and the performance profiles when $\tau = 10^{-7}$, respectively. From Figure 6 and Figure 7 we can see that, PSO-1, PSO-8, PSO-9, PSO-12 and PSO-14 are the best five PSOs, they can solve about 16%, 5%, 17%, 8% and 40% problems, respectively. It is interesting that, although PSO-8 is unstable and does not perform well when $\tau = 10^{-1}$, it performs better than 6 stable PSOs when $\tau = 10^{-7}$.

Therefore, based on our experimental results, we can say that, the answer to the first question proposed at the beginning of this section is "no". That is to say, we can not guarantee that a stable PSO always performs better than an unstable PSO. Actually, it is not surprising. The stable analysis of PSO only requires that, when the whole swarm can not find better swarm best position, then the particles need to converge to the swarm best position. Thus there is no guarantee that a stable PSO performs better than an unstable PSO.

However, our experimental results also show that, when the accuracy is low, for example $\tau = 10^{-1}$, all stable PSOs perform better than all unstable PSOs. In this sense, the stable region (36) is meaningful.

6.4 The "best" parameter settings

Then we turn to the second question proposed in the beginning of this section. We note that, in general, there is probably no best parameter setting for canonical PSO. Our purpose is just looking for the "best" parameter setting on the considered set of benchmark functions.

Our experimental results show that, the two PSOs (PSO-1 and PSO-2) which adopt two of the known best parameter combinations perform well, especially PSO-1. However, there exist parameter combinations that maybe better than $(\omega = 0.7298, \phi = 1.49618)$ and $(\omega = 1/(2\ln 2), \phi = 0.5 + \ln 2)$. For example, our experimental results show that PSO-14 performs much better than both PSO-1 and PSO-2, especially when the accuracy is high. The performance difference between PSO-14 and PSO-1 is about

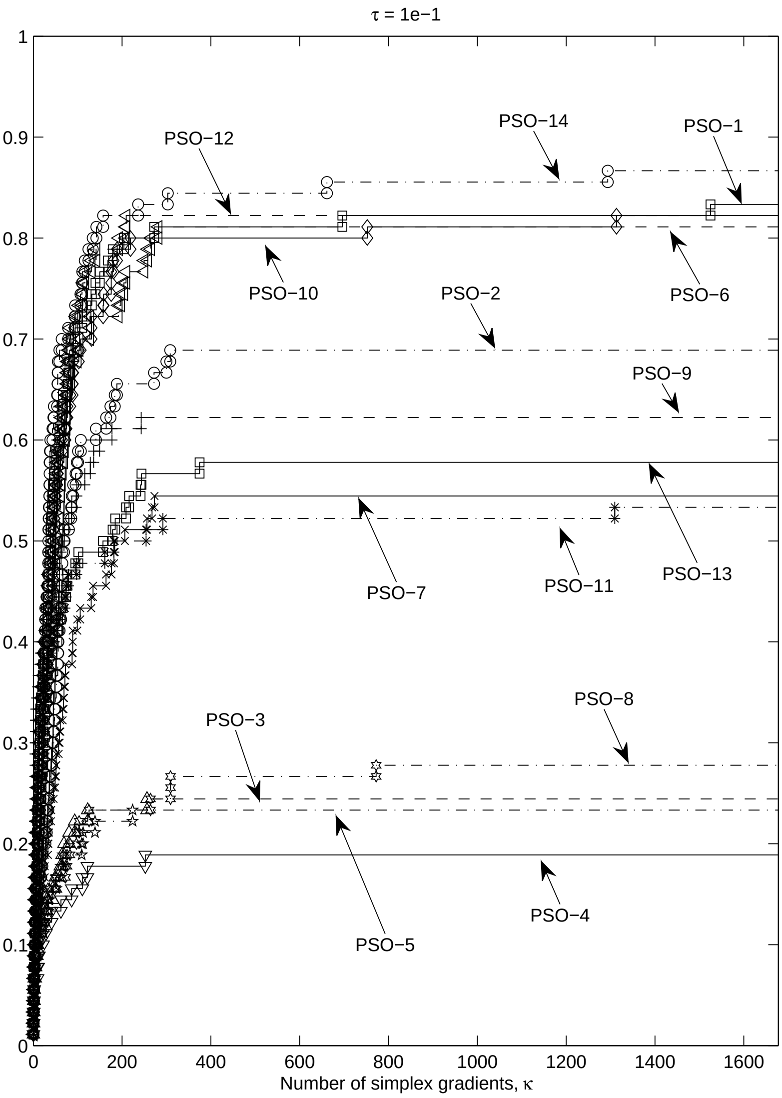

Figure 4: Data profiles ($\tau = 10^{-1}$).

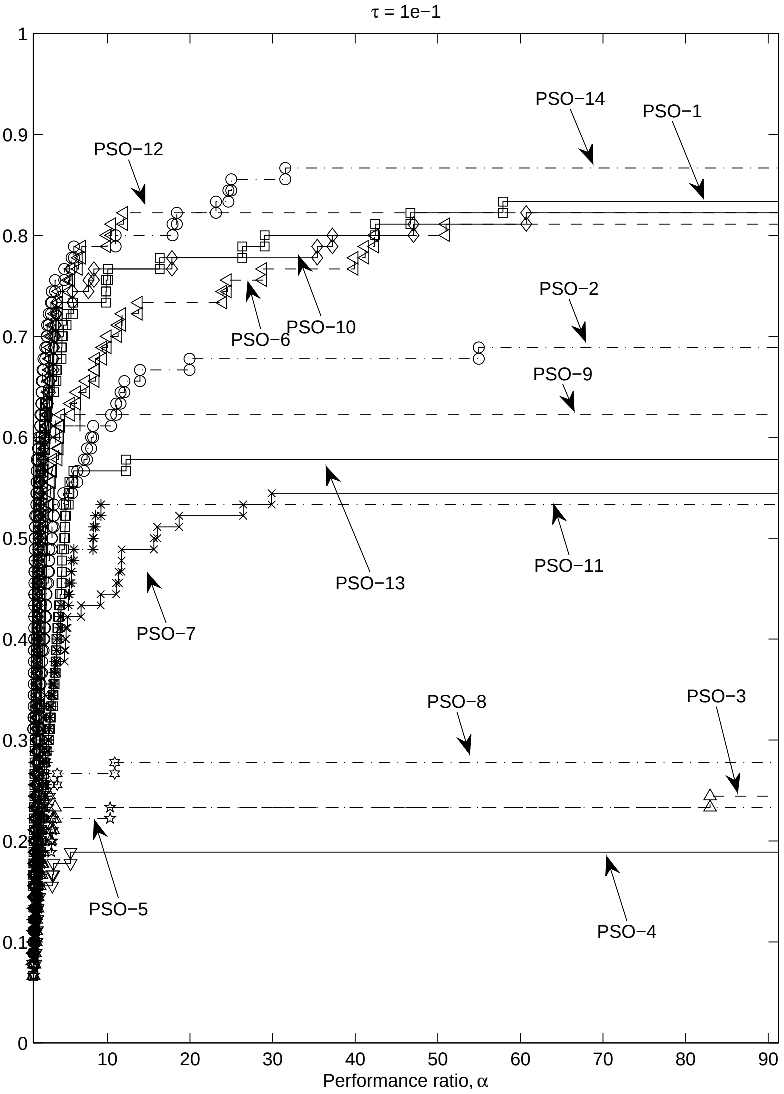

Figure 5: Performance profiles ($\tau = 10^{-1}$).

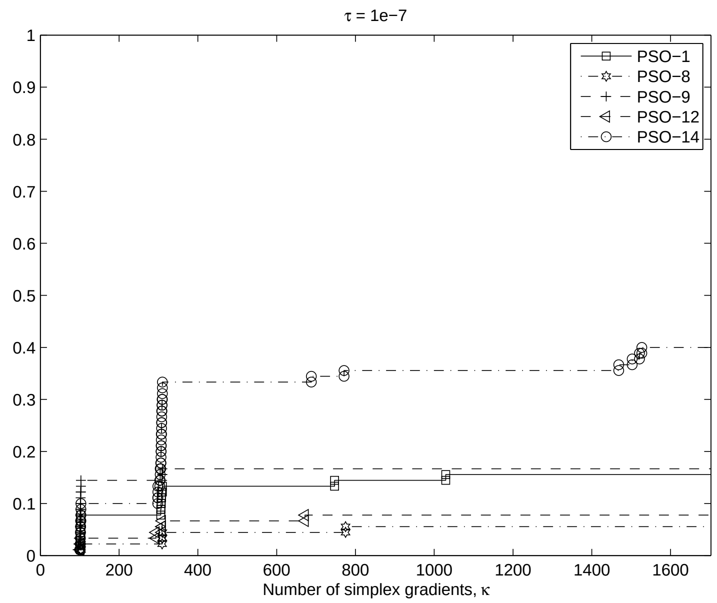

Figure 6: Data profiles ($\tau = 10^{-7}$). Only the best five PSOs' profiles are presented.

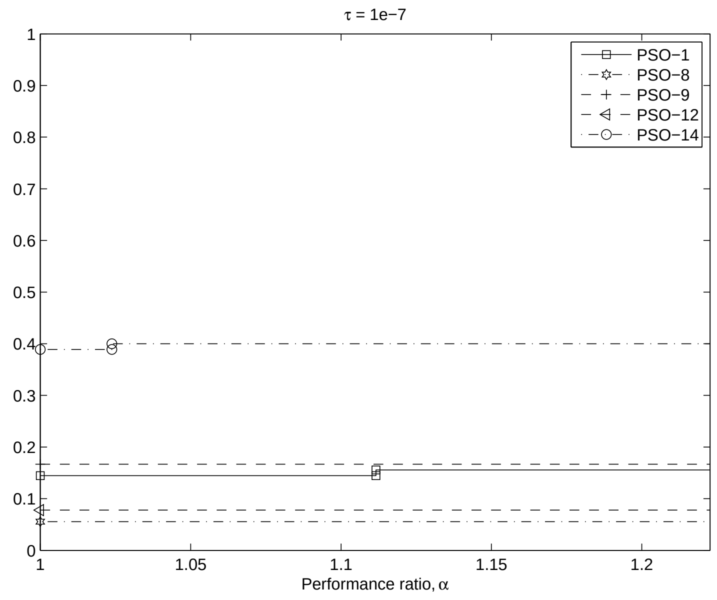

Figure 7: Performance profiles ($\tau = 10^{-7}$). Only the best five PSOs' profiles are presented.

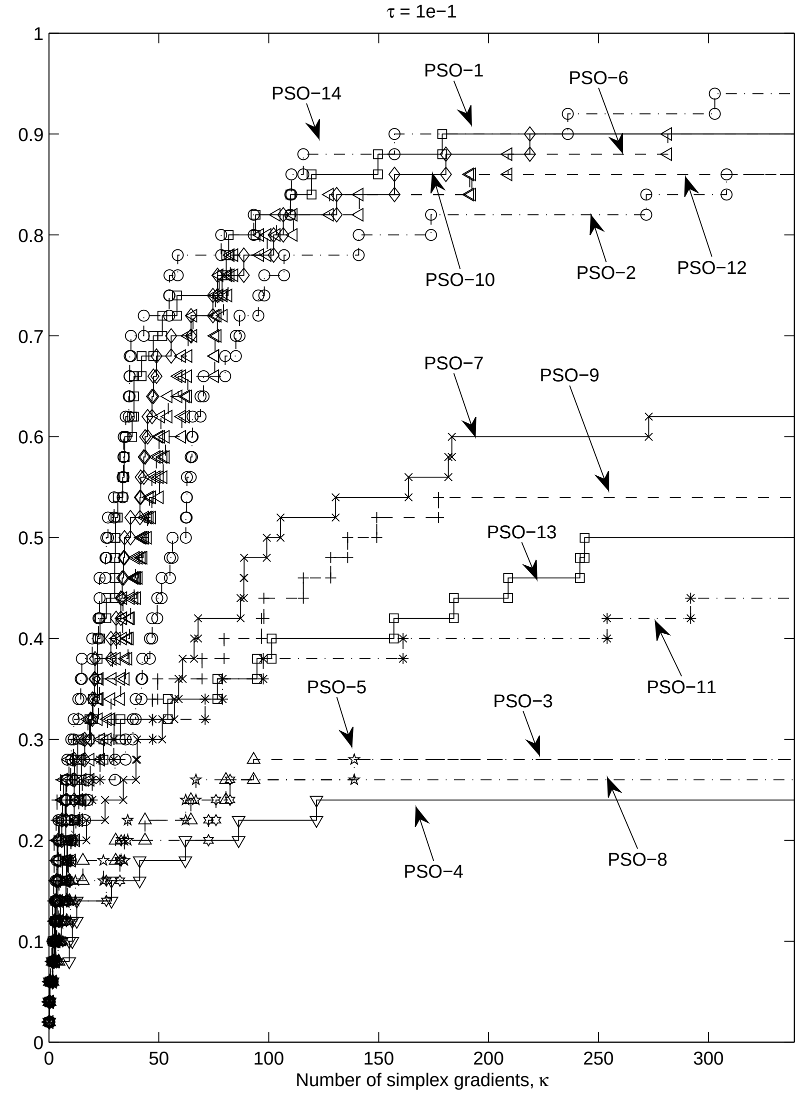

Figure 8: Data profiles for 50 CEC2005 test functions ($\tau = 10^{-1}$).

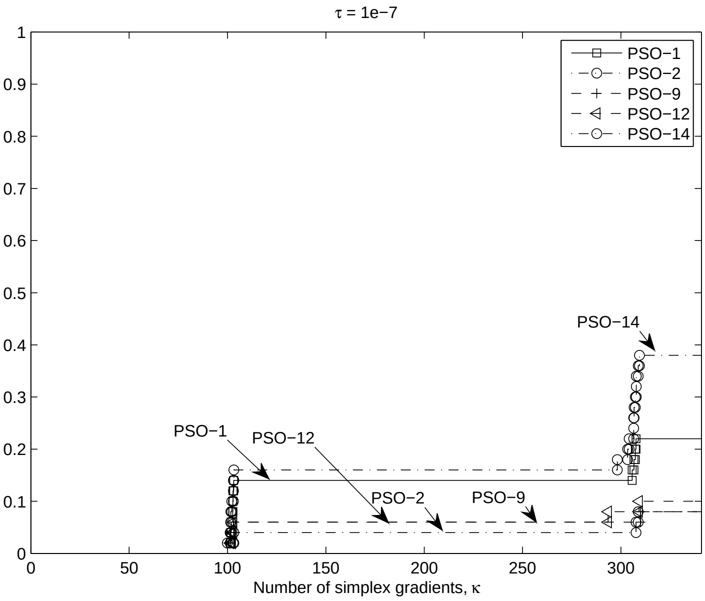

Figure 9: Data profiles for 50 CEC2005 test functions ($\tau = 10^{-7}$). Only the best five PSOs' profiles are presented.

$24\% = 40\% - 16\%$, and the performance difference between PSO-14 and PSO-2 is more than 35%. Therefore, the answer to the second question proposed in the beginning of this section is "yes", $(\omega = 0.42, \phi = 1.55)$ is an example.

In the following, we provide two additional experiments whose results are interesting and helpful in selecting parameter combination.

We test the 14 PSO algorithms on the 50 CEC2005 problems (see Table 2) and the other 40 non-CEC2005 (see Table 3) problems, respectively. Because the CEC2005 problems are often regarded as difficult global optimization problems, and the 40 non-CEC2005 problems are much easier partly due to their lower dimensions, our experiments can be used to show how different PSOs perform on difficult or easy problems.

Figure 8 and Figure 9 show the data profiles on the 50 CEC2005 problems for $\tau = 10^{-1}$ and $\tau = 10^{-7}$, respectively. Similarly, Figure 10 and Figure 11 show the data profiles on the 40 non-CEC2005 problems for $\tau = 10^{-1}$ and $\tau = 10^{-7}$, respectively. Because some PSOs solve very few functions or even none function when $\tau = 10^{-7}$, we present only the best five PSO algorithms' profiles in Figure 9 and Figure 11.

From Figure 8 and Figure 10 we can see that, when the accuracy is low, whatever the test functions is hard or easy, the unstable PSOs perform much worse than any stable PSO. This result supports that the stable region (36) is meaningful.

For the CEC2005 functions, PSO-14 is the best algorithm for almost all budget of computational cost. Finally, it can solve about 95% problems when $\tau = 10^{-1}$ while only about 38% problems when $\tau = 10^{-7}$. The performance difference between low accuracy and high accuracy is about $57\% (= 95\% - 38\%)$.

For the non-CEC2005 function, PSO-9 is the best PSO algorithm when the budget of computational cost is low, while PSO-14 performs the best when the budget is

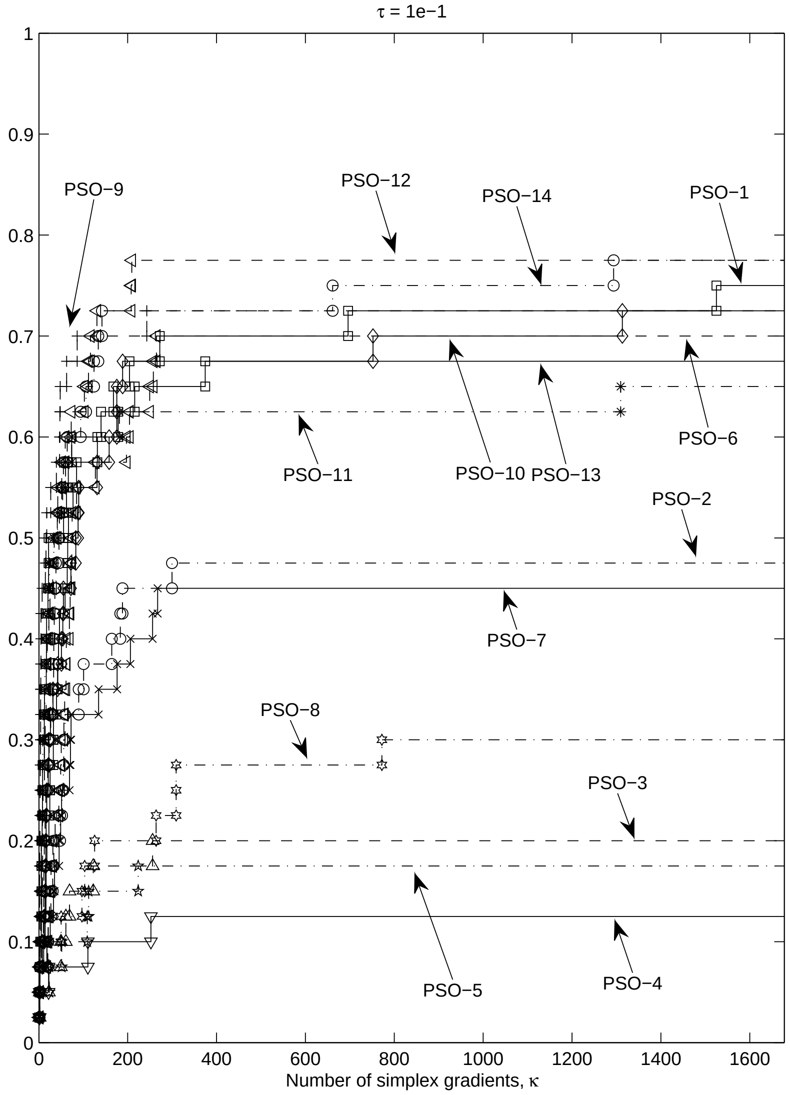

Figure 10: Data profiles for 40 non-CEC2005 test functions ($\tau = 10^{-1}$).

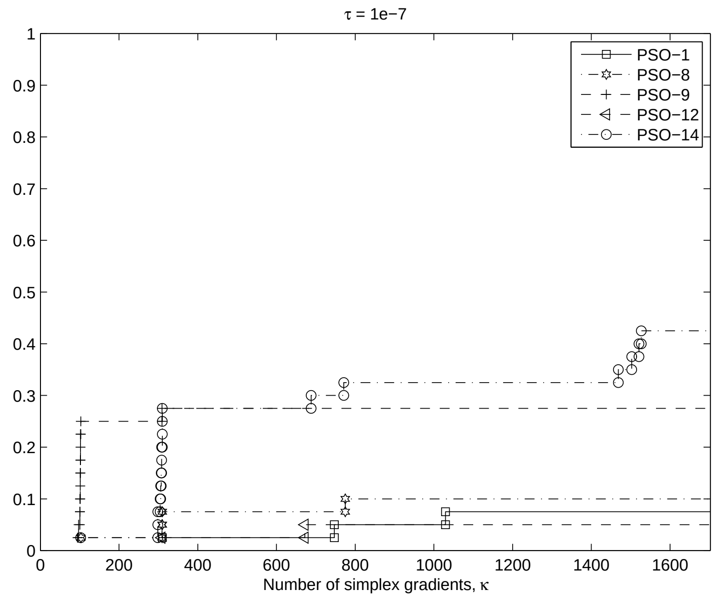

Figure 11: Data profiles for 40 non-CEC2005 test functions ($\tau = 10^{-7}$). Only the best five PSOs' profiles are presented.

high. Finally, PSO-14 can solve about 78% problems when $\tau = 10^{-1}$ while about 43% problems when $\tau = 10^{-7}$. The performance difference between low accuracy and high accuracy is about $35\%(= 78\% - 43\%)$.

Therefore, for the CEC2005 functions, it is not hard to find solutions with low accuracy but hard to find solution with high accuracy. In another word, the hardness of the CEC2005 functions lie in finding solutions with high accuracy.

We note that, PSO-2 (which adopt the known best parameter combination $\omega = 0.7298, \phi = 1.49618$) performs well on the CEC2005 functions, especially when $\tau = 10^{-7}$, it is one of the best five PSOs. However, PSO-2 does not perform well on the non-CEC2005 functions, it can solve only 48% problems when $\tau = 10^{-1}$ while no more than 5% problems when $\tau = 10^{-7}$. It is the reason why PSO-2 lie outside the top 5 PSO algorithms for the whole 90 test problems (see Figure 6).

Similar to PSO-2, PSO-1 performs better on CEC2005 functions than on non-CEC2005 functions. For example, when $\tau = 10^{-7}$, PSO-1 can solve about 22% CEC2005 problems while only about 8% non-CEC2005 functions. However, PSO-1 performs better than PSO-2, it is always one of the best five PSOs, whatever the test functions are the CEC2005 functions or not.

On the other hand, PSO-8 and PSO-9 perform better on non-CEC2005 functions than on CEC2005 functions. For example, when $\tau = 10^{-7}$, PSO-9 can solve about 8% CEC2005 problems while about 27% non-CEC2005 problems. PSO-8 does not perform well on CEC2005 functions, however, it is one of the best five PSOs for non-CEC2005 functions. It is the reason why PSO-8 lies inside the top 5 PSO algorithms for the whole 90 test problems (see Figure 6).

PSO-14 performs much better than any other PSO algorithm, whatever the test functions are CEC2005 functions or not. It is the reason why PSO-14 performs very well on the whole 90 test functions.

At the end of this section, we need to note that, our conclusions in this section are made based only on 3100 function evaluations in each run. Such a budget of computational cost is enough to compute at least 100 simplex gradients for any problems tested in this section. When the budget increases greatly, it is not guaranteed that the conclusions still hold.

7 Conclusion and future work

In this paper, we analyze the stability of PSO. Our main innovation is that, based on a weak assumption of stagnation and a weak order-2 definition of stability, we derived an order-2 stable region for canonical PSO. The same stable region has been proposed in some PSO literatures, however, under stricter assumption of stagnation and often stronger definition of stability.

Our work shows that the classical stagnation assumption is too strict and not necessary for stability analysis of the canonical PSO. Our work also shows that in the order-2 definition of stability, the covariances keep constant over time is too strict and not necessary, too.

We hope that the weak stagnation assumption will open a road to comprehensive understanding about the theoretical properties of PSO. For example, the knowledge diffusion in the swarm and the following behaviors of non-dominant particles. Those are part of our future work.

In this paper, we also provided some experimental results. Our experimental results have shown that, although a stable PSO is not sufficient to be an effective PSO, a stable PSO often performs better than an unstable PSO when the required accuracy is low. In this sense, the obtained stable region is meaningful. Our experimental results also show that, canonical PSO with parameter $(\omega = 0.42, \phi = 1.55)$ performs very well, even better than some known best parameter combinations of PSO on a large set of benchmark problems.

Finally, it is interesting and valuable to analyze what kind of parameter combinations inside the stable region proposed in this paper can result in an effective PSO. That is also our future work.

Acknowledgements

I would like to thank three anonymous reviewers for their helpful comments, which improve this paper greatly. I would also like to thank Professor Maurice Clerc for sending me the codes of the CEC2005 test functions and test functions for the Standard PSO 2011.

Appendix: Proof of Theorem 4

Proof. If $\omega = 0$, then there are three real roots $0, 0, \mu^2 + \sigma^2$ for equation (33). $\lim\limits_{t\to\infty} E[R^2(t)] = 0$ requires

$$\mu^2 + \sigma^2 < 1.$$

If $\omega \in (-1, 0)$, then because

$$\Phi(\omega) = -2\omega^2\mu^2 < 0,\ \Phi(0) = -\omega^3 > 0, \Phi(-\omega) = -2\omega^2\sigma^2 < 0,\ \Phi(+\infty) = +\infty,$$

$\lim_{t\to\infty} E[R^2(t)] = 0$ only requires $\Phi(1) > 0$, i.e.,

$$(1-\omega)\mu^2 + (1+\omega)\sigma^2 < (1+\omega)^2(1-\omega). \tag{41}$$

If $\omega \in (0,1)$, then because

$$\Phi(\omega) = -2\omega^2\sigma^2 < 0,\ \Phi(+\infty) = +\infty.$$

$\lim_{t\to\infty} E[R^2(t)] = 0$ also requires $\Phi(1) > 0$, i.e., (41) must hold. Moreover, we need to assure the norm of another two solutions of (33) less than 1.

Based on the condition (41), we can ensure that the characteristic equation (33) has a root $r_3 \in (\omega, 1)$. What is more, we can rewrite (33) into the following style

$$\Phi(r) = (r - r_3)\left[r^2 + (\omega - \mu^2 - \sigma^2 + r_3)r + \frac{\omega^3}{r_3}\right] = 0. \tag{42}$$

So we obtain the other two roots of (33)

$$r_{4,5} = \frac{(\mu^2 + \sigma^2 - \omega - r_3) \pm \sqrt{\Delta_1}}{2},$$

where $\Delta_1 = (\mu^2 + \sigma^2 - \omega - r_3)^2 - 4\omega^3/r_3$.

$\lim_{t\to\infty} E[R^2(t)] = 0$ requires $\max\{|r_4|, |r_5|\} < 1$, which results in the following condition

$$\omega + r_3 - 1 - \frac{\omega^3}{r_3} < \mu^2 + \sigma^2 < \omega + r_3 + 1 + \frac{\omega^3}{r_3}. \tag{43}$$

Because (41) requires

$$(\mu^2 + \sigma^2) < (1+\omega)^2 - 2\omega\sigma^2/(1-\omega).$$

From (43), combing with $r_3 \in (\omega, 1) \subset (0,1)$, we have

$$\omega + r_3 + 1 + \frac{\omega^3}{r_3} \geq (1+\omega)^2.$$

This implies that the right side inequality of (43) is not necessary.

Furthermore, from (33) we obtain

$$\begin{aligned}
\Phi(-1) &= -1 + (\omega - \mu^2 - \sigma^2) - (\omega\mu^2 - \omega\sigma^2 - \omega^2) - \omega^3 \\
&= -(1-\omega)(1-\omega^2) - (1+\omega)\mu^2 - (1-\omega)\sigma^2 \\
&< 0.
\end{aligned}$$

On the other hand, from (42) we have

$$\Phi(-1) = (-1 - r_3)\left[1 - (\omega - \mu^2 - \sigma^2 + r_3) + \frac{\omega^3}{r_3}\right].$$

This implies that

$$\mu^2 + \sigma^2 > \omega + r_3 - 1 - \frac{\omega^3}{r_3}$$

always holds. Thus the left side inequality of (43) is not necessary, too.

The above analysis shows that (41) is the sufficient and necessary condition for $\lim_{t\to\infty} E[R^2(t)] = 0$. □

References

Blackwell, T. M. (2012). A study of collapse in bare bones particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 16:354–372.

Blackwell, T. M. and Bratton, D. (2008). Examination of particle tails. *Journal of Artificial Evolution and Application*, 2008.

Clerc, M. (2011). Standard particle swarm optimization: from 2006 to 2011. *http://clerc.maurice.free.fr /pso/*.

Clerc, M. and Kennedy, J. (2002). The particle swarm - explosion, stability and convergence in a multidimensional complex space. *IEEE Transanctions on Evolutionary Computation*, 6:58–73.

Eberhart, R. C. and Kennedy, J. (1995). A new optimizer using particle swarm theory. In *The sixth international symposium on micro machine and human science, Nagoya, Japan, Piscataway: IEEE*, pages 39–43.

Eberhart, R. C. and Shi, Y. (2000). Comparing inertia weights and constriction factors in particle swarm optimization. In *IEEE congress on evolutionary computation (CEC),San Diego, CA. Piscataway: IEEE*, pages 84–88.

Evers, G. I. (2009). An automatic regrouping mechanism to deal with stagnation in particle swarm optimization. Master's thesis, University of Texas-Pan American.

Gazi, V. (2012). Stochastic stability analysis of the particle dynamics in the pso algorithm. In *IEEE International Symposium on Intelligent Control (ISIC), part of 2012 IEEE Multi-Conference on Symstems and Control, Dubrovnik, Croatia*, pages 708–713.

Jiang, M., Luo, Y. P., and Yang, S. Y. (2007). Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm. *Information Processing Letters*, 102:8–16.

Kadirkamanathan, V., Selvarajah, K., and Fleming, P. J. (2006). Stability analysis of the particle dynamics in particle swarm optimizer. *IEEE Transactions on Evolutionary Computation*, 10:245–255.

Kennedy, J. (2003). Bare bones particle swarms. In *Proceedings of the IEEE Swarm Intelligence Symposium, Indianapolis, IN*, pages 80–87.

Kennedy, J. and Eberhart, R. C. (1995). Particle swarm optimization. In *In proceeding of the IEEE iternational conference on neural networks IV, Piscataway: IEEE*, pages 1942–1948.

Kennedy, J. and Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. In *Proceedings of the conference on systems, man, and cybernetics, Piscataway: IEEE*, pages 4104–4109.

Kennedy, J. and Mendes, R. (2002). Population structure and particle swarm performance. In *Proceedings of the IEEE congress on evolutionary computation (CEC) Honolulu, HI. Piscataway: IEEE*, pages 1671–1676.

Martínez, J. L. F. and Gonzalo, E. G. (2008). The generalized pso: a new door to pso evolution. *Journal of Artificial Evolution and Application*, 2008.

Mendes, R. (2004). *Population topologies and their influence in particle swarm performance*. PhD thesis, Departamento de Informatica, Escola de Engenharia, Universidade do Minho.

Mendes, R., Cortes, P., Rocha, M., and Neves, J. (2002). Particle swarms for feedforward neural net training. In *Proceedings of the international joint conference on neural networks Honolulu, HI. Piscataway: IEEE*, pages 1895–1899.

Mendes, R., Kennedy, J., and Neves, J. (2003). Watch thy neighbor or how the swarm can learn from its environment. In *Proceedings of the IEEE swarm intelligence symposium (SIS), Piscataway: IEEE*, pages 88–94.

Moré, J. and Wild, S. (2009). Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20:172–191.

Ozcan, E. and Mohan, C. (1999). Particle swarm optimization: surfing the waves. In *Proceeding of Congress on Evolutionary Computation, Piscataway, NJ: IEEE Service Center*, pages 1939–1944.

Poli, R. (2008). Analysis of the publications on the applications of particle swarm optimisation. *Journal of Artificial Evolution and Applications*, pages 1–10.

Poli, R. (2009). Mean and variance of the sampling distribution of particle swarm optimizers during stagnation. *IEEE Transactios on Evolutionary Computation*, 13:712–721.

Poli, R., Kennedy, J., and Blackwell, T. (2007). Particle swarm optimization: an overview. *Swarm Intelligence*, 1:33–57.

Shi, Y. and Eberhart, R. C. (1998). A modified particle swarm optimizer. In *Proceedings of the IEEE international conference on evolutionary computation, Piscataway: IEEE*, pages 69–73.

Taboga, M. (2012). *Lectures on probability theory and mathematical statistics (Second edition)*. Amazon CreateSpace.

Trelea, I. C. (2003). The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters*, 85:317–325.

van den Bergh, F. and Engelbrecht, A. P. (2006). A study of particle swarm optimization particle trajectories. *Information Science*, 176:937 – 971.