Order-2 Stability Analysis of Particle Swarm Optimization

Qunfeng Liu
liuqf@dgut.edu.cn
College of Computing, Dongguan University of Technology,
Dongguan, 523808, China
doi:10.1162/EVCO_a_00129

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

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

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:

$$v_{ij}(k + 1) = \omega v_{ij}(k) + C_{1,ij}(p_{ij}(k) - x_{ij}(k)) + C_{2,ij}(g_{ij}(k) - x_{ij}(k)), \quad (1a)$$

$$x_{ij}(k + 1) = x_{ij}(k) + v_{ij}(k + 1). \quad (1b)$$

Here $x_{ij}(k)$, $v_{ij}(k)$ are the position and the velocity of particle $i$ on dimension $j$ at iteration $k$, respectively, $i = 1, 2, \ldots, N$, $j = 1, \ldots, n$, $k = 1, 2, \ldots$, 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 $\omega, \phi_1, \phi_2$ are three important parameters in PSO. The variable $\omega$ is the inertial weight, which determines the level of inertia; $\phi_1$ and $\phi_2$ are often called the learning factors, which determine the amplitudes of personal best experiments and social best experiments, respectively.

© 2015 by the Massachusetts Institute of Technology Evolutionary Computation 23(2): 187–216
\( p_i(k) \) and \( g_i(k) \) are the personal best and the neighborhood best of particle \( i \), respectively, that is to say,

\[
p_i(k) = \arg \min_{0 \leq t \leq k} f(x_i(t)), \quad g_i(k) = \arg \min_{l \in N_i} f(p_l(k)),
\]

where \( N_i \) is the neighborhood of particle \( i \) which contains all the particles that 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 was first proposed 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, the most work has been done on stability analysis or convergence analysis. PSO stability analysis explores how different essential factors of PSO, especially the parameter combinations and network topology, affect particle swarm dynamics, and moreover, under what kind of conditions does the particle swarm converge to some constant position. Most existing stability analyses focus on the dynamic behaviors or trajectories of particles, the stable regions or convergence regions, and the sample distributions (see, e.g., 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).

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 (e.g., Clerc and Kennedy, 2002; van den Bergh and Engelbrecht, 2006). Some of them adopted conservative theories, for example, the Lyapunov theory (Gazi, 2012; Kadirkamanathan et al., 2006), which is very conservative (Poli et al., 2007), and therefore only sufficient condition can be obtained (Kadirkamanathan et al., 2006). Some of them are strict, for example, in Poli (2009), the (mixed) order-2 original moments are required to be constant over time, and in Blackwell (2012), all covariances are required to be constant over time. See Section 2 for more details.

This paper provides a stability analysis of PSO based on the assumption of weak stagnation. Weak stagnation only requires that the position of the swarm best particle is constant, while all other particles’ best positions are allowed to improve.

This paper points out that there are several advantages to adopting the weak stagnation assumption. First of all, this paper will show that the obtained stable region is not worse than the existing stable regions. This implies that the classical stagnation assumption is too strict and not necessary for PSO stability analysis. Second, the weak stagnation assumption allows the particles to be classified into four different types; this is very important for a more comprehensive understanding about the particle behavior. For example, this work analyzes the conditions needed for knowledge diffusion among
different type of particles, and analyzes the leading behavior of the best particle and the following behavior of other particles.

Based on the weak stagnation assumption and a new definition of stability, this work then focuses on deriving 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, this paper will show that the conditions of stability used in this work are the weakest conditions among the existing stability analyses for canonical PSO and additional conditions bring no benefit.

Although the present 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 $\omega$ and $C_1 \sim U(0, \phi_1), C_2 \sim U(0, \phi_2)$. In this sense, both the PSO with inertia weight and the PSO with constriction are canonical PSOs.

Although a stable PSO whose parameter combination lies inside the stable region is not guaranteed to be an efficient PSO, the present experimental results show that, when the required accuracy is low, a stable PSO often performs better than an unstable PSO. In this sense, the obtained stable region is meaningful. The present 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.

The rest of this paper is organized as follows. In Section 2, there is a review of existing stability analyses of PSO, especially their definitions of stability and the resulting stable regions. In Section 3, the assumption of weak stagnation is proposed which provides some advantages of weak stagnation in the stability analysis of PSO. In Section 4, the stable region of PSO is derived based on a new definition of stability. In Section 5, the present stability analysis is compared with existing stability analyses. In Section 6, some experimental results are reported to show that the present stable region is meaningful. The final section provides conclusions and directions for future work.

2 Existing Stability Analyses

This section covers several existing PSO stability analyses. For the present purpose, the focus is mainly on the definitions of stability and the resulting stable regions.

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

$$\lim_{t \to \infty} x(t) = y,$$  \hspace{1cm} (2)

where $y$ 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 regions

$$\omega \in (0, 1), \quad \phi_1 + \phi_2 < 2(1 + \omega)$$  \hspace{1cm} (3)

or

$$\omega \in (-1, 1), \quad \phi_1 + \phi_2 < 2(1 + \omega)$$  \hspace{1cm} (4)

were obtained. In the rest of this paper, the region defined in Equation (3) will not be adopted, partly because most works adopt $\omega \in (-1, 1)$, and partly for simplicity.

After the stable region in Equation (4) is obtained, there are two approaches to deal with the randomness of $C_1, C_2$. The first approach is simply to replace $C_1, C_2$ with their maximal values $\phi_1$ and $\phi_2$, respectively, which results in the same stable region as
in Equation (4). The second approach is to 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

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

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

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

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

**Definition 1:** If the condition in Equation (6) holds, then PSO is order-1 stable and the corresponding parameter region is its order-1 stable region.

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

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

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

is derived using Lyapunov stability analysis.

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

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

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

Because both the region defined in Equation (7) and the region defined in Equation (8) adopt the Lyapunov condition, this work will refer to them as Lyapunov stable regions for convenience.

It has been pointed out (Jiang et al., 2007; Poli, 2009) that order-1 stability is not enough to ensure convergence; 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

$$\lim_{t \to \infty} E[|x(t) - y|^2] = 0 \quad (9)$$
is adopted to ensure order-2 stability, where \( y = \lim_{t \to \infty} E[x(t)] \). This condition requires that \( x(t) \) mean square converges to \( y \) (Taboga, 2012). Based on the conditions of mean square convergence, the following region

\[
\frac{5\phi - \sqrt{25\phi^2 - 336\phi + 576}}{24} < \omega < \frac{5\phi + \sqrt{25\phi^2 - 336\phi + 576}}{24}
\]

(10)
is derived (\( \phi_1 = \phi_2 = \phi \)).

Having looked at the region from Jiang et al. (2007), the focus now shifts to Poli (2009). In Poli (2009), the following conditions

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

(11)

are adopted to ensure order-2 stability, where \( \beta_0 \) and \( \beta_1 \) are constant vectors that are not presented explicitly in Poli (2009). Then under the classical stagnation assumption, and with \( \phi_1 = \phi_2 = \phi \), the following line

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

was obtained. On this line, the magnitude of the largest eigenvalue of an important matrix (\( M \) in Poli, 2009) is 1. In other words, Poli (2009) implicitly presented an order-2 stable region

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

(12)

By reformulating Equation (12) as follows

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

then the same inequalities for \( \omega \) are obtained as in Equation (10). Therefore, both Equations (10) and (12) define the same stable region. For convenience, we call both the region in Equation (10) and the region in Equation (12) order-2 stable regions.

Martínez and Gonzalo (2008) proposed a generalized continuous PSO, which extended the definitions of stability and the resulting 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

\[
E[x(t)] = y
\]

(13)

and

\[
E[x(t) - y][x(t - k) - y] = \gamma_k, \quad k = 0, 1, 2, \ldots
\]

(14)

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

\[
\omega \in (-1, 1), \quad \phi_1 + \phi_2 < \frac{24(1 - \omega^2)}{7 - 5\omega}.
\]

(15)
can be derived from the weak stationary condition. For convenience, we call the region in Equation (15) the weak stationary stable region. It is clear that when \( \phi_1 = \phi_2 \), this region is the same as the order-2 stable region in Equation (12).
3 Weak Stagnation and Its Advantages

Existing stability analyses reviewed in the above section often adopted the assumption of stagnation, which requires that all particles’ best position must remain constant. This is not a realistic assumption. This paper proposes to adopt the weak stagnation assumption, whose definition is as follows.

**Definition 2:** If the whole swarm’s best position \( x_i(K) \) remains constant since the \( K \)th iteration until the \((K + M)\)th \((M \geq 3)\) iteration, then PSO is in the weak stagnation state during iteration \((K, K + M)\), and \( x_i(K) \) is 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, classical stagnation may never happen in reality. However, under 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 (e.g., Evers, 2009). Moreover, if the global optimum is found, then the whole particle swarm will stay in the weak stagnation state.

Moreover, the weak stagnation assumption allows the particles to be classified into one dominant particle and several types of nondominant particles. Such a classification allows the conditions for knowledge sharing among these different type of particles to be analyzed, as well as the behaviors of different type of particles.

The remainder of this section provides the classification of particles and a simple condition for knowledge sharing among the dominant particle and the nondominant particles. The section then provides the leading behavior of the dominant particle, which is important for the later analysis.

3.1 Classification of Particles and Knowledge Diffusion

In the weak stagnation state, all particles can be classified into the following four types:

- Type I: the dominant particle;
- Type II: particles that are informed by the dominant particle, that is, particles whose neighborhood best particle is the dominant particle;
- Type III: neighborhood best particles not including the dominant particle;
- Type IV: particles which are informed by the type III particles.

Here, the notation \( T_i(i = 1, 2, 3, 4) \) is used to denote the sets of particles that belong to these four types.

Note that the classification of particles depends on the topology (social network). For example, for the gbest topology, there are no type III or type IV particles, so \( T_3 = T_4 = \emptyset \). For other topologies except gbest, four types of particles are often extant.

Given any topology, the type I particle possesses the best experience or knowledge, the type II particles possess worse experience, and the type IV particles (if any exist) possess the worst experience. In what follows, the conditions are discussed under which 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 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

\[ T_2 \cap T_3 \neq \emptyset. \]  

\[(16)\]

**Proof:** If \( T_3 = \emptyset \), then there is only the type I particle to be the informer, that is, 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 \( T_3 \), that is, \( T_2 \cap T_3 \neq \emptyset \). In other words, after the 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 that becomes a type III particle, then knowledge diffusion is impossible. \( \square \)

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

When 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 the particle swarm can benefit from the shutting off of knowledge diffusion.

### 3.2 Leading Behavior of the Dominant Particle

From dynamic Equations (1a) and (1b) is obtained the following dynamic equation for particle \( i \) on dimension \( j \)

\[ x_{ij}(k + 1) = \alpha_1 x_{ij}(k) + \alpha_2 x_{ij}(k - 1) + \alpha_3 p_{ij}(k) + \alpha_4 g_{ij}(k). \]

\[(17)\]

where \( i = 1, \ldots, N \), \( j = 1, \ldots, n \) and

\[ \alpha_1 = 1 + \omega - C_{1,ij} - C_{2,ij}, \]

\[ \alpha_2 = -\omega, \]

\[ \alpha_3 = C_{1,ij}, \]

\[ \alpha_4 = C_{2,ij}. \]

\[(18)\]

Note that, here and after, \( \alpha_1, \alpha_2, \alpha_3, \) and \( \alpha_4 \) 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 is applied, \( \alpha_1(k) \) is used, denoting the value of \( \alpha_1 \) at iteration \( k \).

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

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

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

\[ \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1. \]  

\[(19)\]
This implies that the value of $x_{ij}$ at iteration $k + 1$ is always a random convex combination of its current value $x_{ij}(k)$, its past value $x_{ij}(k - 1)$, its current personal best value $p_{ij}(k)$ and its current neighborhood best value $g_{ij}(k)$.

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

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

Suppose that the PSO is in the 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 also equals the swarm’s best. Therefore, $p_d(k) = g_d(k) = x_d(K), k \in [K, K + M)$, where $x_d(K)$ is the stagnation point. Thus for the dominant particle $d$, the following simpler SDE obtains

$$x_d(k + 1) = \alpha_1 x_d(k) + \alpha_2 x_d(k - 1) + \tilde{\alpha}_3 x_d(K), \quad (20)$$

for $k \in [K, K + M)$, where $\tilde{\alpha}_3 = \alpha_3 + \alpha_4$. 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 cannot solve Equation (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 $(K, K + M)$, $x_d(K)$ is the stagnation point, and particle $d$ is the dominant particle. Then the solutions of the SDE in Equation (20) satisfy the following formula

$$x_d(K + t) = x_d(K) + R(t)(x_d(K + 1) - x_d(K)), \quad t \in [0, M), \quad (21)$$

where $R(t)$ satisfies the following recursion

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

**PROOF:** Theorem 2 is proved by reduction. When $t = 0$ and $t = 1$, Equation (21) implies that

$$x_d(K) = x_d(K),$$

$$x_d(K + 1) = x_d(K) + (x_d(K + 1) - x_d(K)).$$

It is clearly true.

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

$$x_d(K + k - 1) = x_d(K) + R(k - 1)(x_d(K + 1) - x_d(K)),$$

$$x_d(K + k) = x_d(K) + R(k)(x_d(K + 1) - x_d(K)).$$

Next it must be proven that Equation (21) holds for $t = k + 1$.

From Equation (20),

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

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

$$x_d(K + k + 1) = x_d(K) + R(k + 1)(x_d(K + 1) - x_d(K)),$$

where $R(k + 1)$ satisfies $R(k + 1) = \alpha_1(K + k + 1)R(k) - \omega R(k - 1).$ \(\square\)
Setting \( k = K \) in Equation (20), and combining with \( \alpha_2 = -\omega \) yields

\[
x_d(K + 1) - x_d(K) = \omega(x_d(K) - x_d(K - 1)).
\]

Combining this result with Equation (21), any point sampled by the dominant particle during weak stagnation depends not only on the stagnation point \( x_d(K) \), but also \( x_d(K + 1) \) or \( x_d(K - 1) \).

From Equation (21), the following conclusions can be drawn 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 \( \alpha_1 \).
- 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 \( x_d(K + 1) \) and \( x_d(K) \).
- In the general case, \( R(t) \) is different for different dimensions; hence the points sampled by the dominant particle distribute randomly around \( x_d(K) \). The exact distribution depends on \( \alpha_1 \).

A further analyses of \( R(t) \) is undertaken in the next section.

4 Stability Analysis of PSO

In this section, a new definition of order-2 stability is first proposed, and then an order-2 stable region is derived.

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 the weak stagnation state forever. Then the PSO is said to be stable if

\[
\lim_{t \to \infty} D[x_d(K + t)] = 0,
\]

where \( x_d(K) \) is the stagnation point, and particle \( d \) is the dominant particle.

**Remark 1:** Definition 3 implies implicitly that the stability of the PSO is determined only by the dominant particle. Although for some iterations \( T_2 \cap T_3 \) may equal empty set, there are some particles whose evolution is independent of the dominant particle. However, as long as the 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 Definition 3 requires the variance to converge to zero, it satisfies order-2 stability. What is more, the following theorem shows that it also satisfies order-1 stability.

**Theorem 3:** Definition 3 satisfies order-1 stability. Especially, if \( P \{ C_1 + C_2 \neq 0 \} = 1 \), that is, there is almost no probability that \( C_1 + C_2 \) equals 0 then

\[
\lim_{t \to \infty} E[x_d(K + t)] = x_d(K).
\]
PROOF: Since
\[ D[x_d(K + t)] = E[x_d(K + t) - E(x_d(K + t))]^2, \]
\( \lim_{t \to \infty} D[x_d(K + t)] = 0 \) implies that \( E[x_d(K + t)] \) exists almost surely for all \( t > 0 \), and moreover, there exists a constant vector \( y \) such that \( x_d(K + t) \) converges to the constant vector \( y \) almost surely (Taboga, 2012). In other words, we have
\[
P\{ \lim_{t \to \infty} x_d(K + t) = y \} = 1,
\]
therefore
\[
\lim_{t \to \infty} E[x_d(K + t)] = y,
\]
that is, Definition 3 satisfies order-1 stability.

In the following, it will be proven that \( y \) equals \( x_d(K) \). The proof is by contradiction.

On one hand, if \( x_d(K + t) \) converges to a constant vector, this implies that, taking the limitation in Equation (1a) (where the \( i, j \) have been dropped for convenience), yields
\[
\lim_{t \to \infty} v_d(K + t) = \left( C_1 + C_2 \right) \left( x_d(K) - y \right) \left( 1 - \omega \right).
\]
On the other hand, taking the limit in Equation (1b), yields
\[
\lim_{t \to \infty} v_d(K + t) = 0.
\]
Therefore, if \( P\{C_1 + C_2 \neq 0\} = 1 \), then \( \lim_{t \to \infty} E[x_d(K + t)] = y = x_d(K) \). \( \square \)

Hence, the present definition of stability satisfies both order-1 stability and order-2 stability. Moreover, the following corollary holds, which can be derived straightforwardly from Theorem 2 and Equation (24).

**Corollary 1:** Suppose that the swarm is in the 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 the PSO is stable or not.

**Lemma 1:** Suppose that the swarm is in the weak stagnation state forever. Then the 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
\[
x_d(K + t) = x_d(K) + R(t)(x_d(K + 1) - x_d(K)).
\]
Because the stagnation point \( x_d(K) \) and \( x_d(K + 1) \) are constant,
\[
D[x_d(K + t)] = D[R(t)][x_d(K + 1) - x_d(K)]^2.
\]
Thus \( \lim_{t \to \infty} D[x_d(K + t)] = 0 \) is equivalent to \( \lim_{t \to \infty} D[R(t)] = 0 \). \( \square \)

Since \( E[R^2(t)] = D[R(t)] + [E(R(t))]^2 \), the following corollary can be derived straightforwardly from Lemma 1 and Corollary 1.
COROLLARY 2: Suppose that the swarm is in the weak stagnation state forever. If \( P \{ C_1 + C_2 \neq 0 \} = 1 \), then the PSO is stable if and only if
\[
\lim_{t \to \infty} E[R^2(t)] = 0.
\] (27)

For canonical PSO and many of its variants, \( C_1 + C_2 \) is always greater than 0, so the condition \( P \{ C_1 + C_2 \neq 0 \} = 1 \) always holds. Therefore, in the rest of this paper, 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 sections, \( E[R^2(t)] \) is computed and then the stable region is derived. First, denote
\[
\mu = E(\alpha_1), \quad \sigma^2 = D(\alpha_1).
\] (28)

Especially for canonical PSO,
\[
\mu = 1 + \omega - \frac{\phi_1 + \phi_2}{2}, \quad \sigma^2 = \frac{\phi_1^2 + \phi_2^2}{12}.
\]

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

From Equation (22)
\[
R^2(t + 1) = \alpha_1^2(K + t)R^2(t) + \omega^2 R^2(t - 1) - 2\omega\alpha_1(K + t)R(t)R(t - 1),
\]
and
\[
R(t + 1)R(t) = \alpha_1(K + t)R^2(t) - \omega R(t)R(t - 1).
\]

Thus,
\[
E[R^2(t + 1)] = (\mu^2 + \sigma^2)E[R^2(t)] + \omega^2 E[R^2(t - 1)] - 2\mu\omega E[R(t)R(t - 1)].
\] (29)

and
\[
E[R(t + 1)R(t)] = \mu E[R^2(t)] - \omega E[R(t)R(t - 1)].
\] (30)

From Equation (29), \( E[R(t)R(t - 1)] \) and \( E[R(t + 1)R(t)] \), which can be substituted into Equation (30). Finally, the following difference equation can be obtained
\[
E[R^2(t + 2)] + (\omega - \mu^2 - \sigma^2)E[R^2(t + 1)] + (\omega\mu^2 - \omega\sigma^2 - \omega^2)E[R^2(t)] - \omega^3 E[R^2(t - 1)] = 0,
\] (31)

which satisfies
\[
E[R^2(0)] = 0, \quad E[R^2(1)] = 1, \quad E[R^2(2)] = \mu^2 + \sigma^2.
\] (32)

The characteristic equation of Equation (31) is
\[
\Phi(r) = r^3 + (\omega - \mu^2 - \sigma^2)r^2 + (\omega\mu^2 - \omega\sigma^2 - \omega^2)r - \omega^3 = 0,
\] (33)
which is the same as in Jiang et al. (2007). Through solving the characteristic equation, three solutions of this characteristic equation are obtained. Setting \( \lim_{t \to \infty} E[R^2(t)] = 0 \), and letting the norm of all these three solutions be less than 1, then the following theorem is obtained, where the proof can be found in the Appendix.

THEOREM 4: Suppose that \( E[R^2(t)] \) is the solution of Equation (31), then the sufficient and necessary condition for \( \lim_{t \to \infty} E[R^2(t)] = 0 \) is
\[
(1 - \omega)\mu^2 + (1 + \omega)\sigma^2 < (1 + \omega)^2(1 - \omega).
\] (34)
4.3 Stable Region

From Corollary 2 and Theorem 4, the stable region can be obtained.

**Theorem 5:** The PSO is stable if and only if the parameters $\omega$, $\phi_1$, and $\phi_2$ satisfy the condition

$$
\begin{align*}
&\omega \in (-1, 1) \\
&(1 - \omega)\mu^2 + (1 + \omega)\sigma^2 < (1 + \omega)(1 - \omega),
\end{align*}
$$

where $\mu$, $\sigma^2$ are defined in Equation (28).

**Remark 2:** We note that the stable region in Equation (35) is suitable for any PSO whose SDE is Equation (17), where $C_1$, $C_2$ are allowed 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 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

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

Then it is straightforward to obtain the conditions. □

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

$$
\begin{align*}
&\omega \in (-1, 1) \\
&\phi \in \left(0, \frac{12(1 - \omega^2)}{7 - 5\omega}\right).
\end{align*}
$$

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

In closing this section, the following remark shows what kind of topology can be applied by order-2 stability analysis.

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

The next section compares the present stability analysis with some existing stability analyses mentioned in Section 2.
Stability Analysis of PSO

5 Comparison and Discussion

This section compares the different definitions of stability covered in Section 2 and the stable regions that result 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 the general case. This section focuses on the canonical PSO with $\phi_1 = \phi_2 = \phi$.

First, all stable regions discussed in this paper are graphed in Figure 2. In Figure 2, the triangular regions ABC and ABF are the order-1 stable regions in Equations (4) and (5), respectively; the concave triangle ABD (horizontal lines) and ABE (vertical lines) are the Lyapunov stable regions from Equations (7) and (8), respectively; the convex ABGE region is the order-2 stable region from Equation (36). Not surprisingly, the order-2 stable region is larger than the Lyapunov stable regions but smaller than the order-1 stable region ABF.

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

First of all, 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 in Equation (36). This fact results straightforwardly in the following proposition.

**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 words, the classical stagnation assumption is too strict and not necessary.

In the previous works under consideration (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 applied to any N-dimensional arbitrary particle. However, the assumptions underlying classical stagnational are unlikely to even actually happen, so the analyses based on the classical stagnation assumption have diminished meaning.

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

The focus now turns to a comparison of the definitions of stability.
In Poli (2009), the conditions of stability include Equations (6) and (11), which require limits not only on expectation and variance but also on the covariance $E[x(t) - y][x(t - 1) - y]$. In Blackwell (2012), all the covariances are required to keep constant over time (see Equation (14)), which is a stricter condition. However, in both Jiang et al. (2007) and this paper, there is no bound on the covariances. Moreover, from our definition of stability and Theorem 3, having the variance converge to zero can ensure
not only order-2 stability but also order-1 stability. Therefore, the following proposition follows.

**Proposition 2:** Definition 3 is enough to derive an order-2 stable region for canonical PSO with \( \phi_1 = \phi_2 = \phi \). In other words, letting the variance of the dominant particle’s position converge to zero is sufficient for deriving an order-2 stable region, and other conditions are not necessary.

The focus of this section now turns to comparing the definition of stability (see also Equation (9)) proposed in Jiang et al. (2007) with that proposed in this paper.

First, let \( y = \lim_{t \to \infty} E[x(t)] \), then the following relationship is obtained:

\[
E[x(t) - y]^2 = E[x(t) - E[x(t)] + E[x(t)] - y]^2 = D[x(t)] + (y - E[x(t)])^2.
\]

This implies that \( E[x(t) - y]^2 > D[x(t)] \) often holds. However, taking the limit yields

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

Therefore, the definition of stability proposed in Jiang et al. (2007) is equivalent to that proposed in this paper. Not surprisingly, they yield the same stable region.

However, since \( E[x(t) - y]^2 > D[x(t)] \) almost always holds, the present definition can be regarded as an extension of the definition proposed in Jiang et al. (2007). In order to compute \( E[x(t) - y]^2 \), it is necessary to first compute \( \lim_{t \to \infty} E[x(t)] \) and \( D[x(t)] \). In this sense, the present definition is easier to compute than that proposed in Jiang et al. (2007). Moreover, the present definition need not touch upon order-1 stability, which is guaranteed automatically (see Theorem 3).

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

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

### 6 Experimental Results

This section presents experimental results. The main purpose is to test:

- Does a stable PSO perform better than an unstable PSO?
- Is there any parameter combination that is better than some known best parameter combinations?

#### 6.1 Algorithm Configurations and Test Problems

Specifically, 14 parameter combinations of the canonical PSO are tested which adopt the gbest topology with 20 particles in the swarm. The only difference among them is that they adopt different parameter combinations \( (\omega, \phi_1 = \phi_2 = \phi) \). Table 1 shows the 14 parameter combination, and their distribution is illustrated in Figure 3.
Table 1: Fourteen parameter combinations of PSO tested in this paper.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>PSO-1</th>
<th>PSO-2</th>
<th>PSO-3</th>
<th>PSO-4</th>
<th>PSO-5</th>
<th>PSO-6</th>
<th>PSO-7</th>
<th>PSO-8</th>
<th>PSO-9</th>
<th>PSO-10</th>
<th>PSO-11</th>
<th>PSO-12</th>
<th>PSO-13</th>
<th>PSO-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\omega)</td>
<td>(1/(2 \ln 2))</td>
<td>0.7298</td>
<td>0.7298</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.42</td>
<td>0.42</td>
<td>0.2</td>
<td>(-0.2)</td>
<td>(-0.2)</td>
<td>(-0.42)</td>
<td>(-0.7)</td>
<td>0.42</td>
</tr>
<tr>
<td>(\phi)</td>
<td>0.5 + \ln 2</td>
<td>1.49618</td>
<td>2.1</td>
<td>2</td>
<td>1.8</td>
<td>0.5</td>
<td>2</td>
<td>2.6</td>
<td>0.8</td>
<td>1.4</td>
<td>0.5</td>
<td>1</td>
<td>0.4</td>
<td>1.55</td>
</tr>
</tbody>
</table>
Among these 14 parameter combinations, there are four unstable combinations (PSO-3, PSO-4, PSO-5, and PSO-8) and 10 stable combinations. Among the 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 the PSO literature and can be regarded as two of the known best parameter combinations. Among the other eight stable combinations, half of them (PSO-6, PSO-7, PSO-9, and PSO-14) lie inside the region $\omega \in (0, 1)$, while the other half (PSO-10, PSO-11, PSO-12, and PSO-13) lie inside the region $\omega \in (-1, 0)$.

Our test set includes 25 CEC2005 test functions and 26 other test functions. These 51 test functions are used to test the standard PSO 2011, for which the code was 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. In total, test 90 functions are used.

6.2 Data Profile Technique

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

Specifically, each PSO algorithm $s \in S$ is used to solve each test problem $p \in P$ for 50 runs, where $S$ denotes the set of 14 PSO algorithms and $P$ denotes the set of 90 test
Table 2: Properties of the 50 CEC2005 test functions.

<table>
<thead>
<tr>
<th>Function name</th>
<th>$d$</th>
<th>Minimal value</th>
<th>Search space</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEC2005 F1</td>
<td>10, 30</td>
<td>$-450$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F2</td>
<td>10, 30</td>
<td>$-450$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F3</td>
<td>10, 30</td>
<td>$-450$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F4</td>
<td>10, 30</td>
<td>$-310$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F5</td>
<td>10, 30</td>
<td>$390$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F6</td>
<td>10, 30</td>
<td>$-180$</td>
<td>$[0, 600]$</td>
</tr>
<tr>
<td>CEC2005 F7</td>
<td>10, 30</td>
<td>$-140$</td>
<td>$[-32, 32]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F8</td>
<td>10, 30</td>
<td>$-330$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F9</td>
<td>10, 30</td>
<td>$90$</td>
<td>$[-0.5, 0.5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F10</td>
<td>10, 30</td>
<td>$-460$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F11</td>
<td>10, 30</td>
<td>$-130$</td>
<td>$[-3, 1]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F12</td>
<td>10, 30</td>
<td>$-300$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F13</td>
<td>10, 30</td>
<td>$120$</td>
<td>$[-100, 100]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F14</td>
<td>10, 30</td>
<td>$120$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F15</td>
<td>10, 30</td>
<td>$120$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F16</td>
<td>10, 30</td>
<td>$10$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F17</td>
<td>10, 30</td>
<td>$10$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F18</td>
<td>10, 30</td>
<td>$10$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F19</td>
<td>10, 30</td>
<td>$10$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F20</td>
<td>10, 30</td>
<td>$10$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F21</td>
<td>10, 30</td>
<td>$360$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F22</td>
<td>10, 30</td>
<td>$360$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F23</td>
<td>10, 30</td>
<td>$360$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F24</td>
<td>10, 30</td>
<td>$260$</td>
<td>$[-5, 5]^{d}$</td>
</tr>
<tr>
<td>CEC2005 F25</td>
<td>10, 30</td>
<td>$260$</td>
<td>$[-2, 5]^{d}$</td>
</tr>
</tbody>
</table>

problems. In each run, the algorithm does not stop until 3,100 function evaluations have been reached. The history of the found minimal function values is preserved. That is to say, each run results in a vector whose length is 3,100, for which the $k$th element is the minimal function value found during $k$ function evaluations. After 50 runs are finished, the component wise average of the 50 vectors is computed. 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, the result is a $3,100 \times 90 \times 14$ 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:

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

(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:

$$f(x_0) - f(x) \geq (1 - \tau)(f(x_0) - f_L).$$

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

<table>
<thead>
<tr>
<th>Function name (abbreviation)</th>
<th>$d$</th>
<th>Minimal value</th>
<th>Search space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere (f26)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-100, 100]^d$</td>
</tr>
<tr>
<td>Rastrigin (f27)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-5.12, 5.12]^d$</td>
</tr>
<tr>
<td>Six hump camel back (f28)</td>
<td>2</td>
<td>$-1.031628$</td>
<td>$[-5.5]^d$</td>
</tr>
<tr>
<td>Step (f29)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-100, 100]^d$</td>
</tr>
<tr>
<td>Rosenbrock (f30)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-2, 2]^d$</td>
</tr>
<tr>
<td>Ackley (f31)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-32, 32]^d$</td>
</tr>
<tr>
<td>Griewank (f32)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-600, 600]^d$</td>
</tr>
<tr>
<td>Salomon (f33)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-100, 100]^d$</td>
</tr>
<tr>
<td>Normalized Schwefel (f34)</td>
<td>10, 30</td>
<td>$-418.98288727274338$</td>
<td>$[-512, 512]^d$</td>
</tr>
<tr>
<td>Quartic (f35)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-1.28, 1.28]^d$</td>
</tr>
<tr>
<td>Rotated hyper-ellipsoid (f36)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-100, 100]^d$</td>
</tr>
<tr>
<td>Norwegian (f37)</td>
<td>10, 30</td>
<td>1</td>
<td>$[-1.1, 1.1]^d$</td>
</tr>
<tr>
<td>Alpine (f38)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-10, 10]^d$</td>
</tr>
<tr>
<td>Branin (f39)</td>
<td>2</td>
<td>0.397887</td>
<td>$[-5, 15]^d$</td>
</tr>
<tr>
<td>Easom (f40)</td>
<td>2</td>
<td>$-1$</td>
<td>$[-100, 100]^d$</td>
</tr>
<tr>
<td>Goldstein Price (f41)</td>
<td>2</td>
<td>3</td>
<td>$[-2, 2]^d$</td>
</tr>
<tr>
<td>Shubert (f42)</td>
<td>2</td>
<td>$-186.7309$</td>
<td>$[-10, 10]^d$</td>
</tr>
<tr>
<td>Hartmann (f43)</td>
<td>3</td>
<td>$-3.86278$</td>
<td>$[0, 1]^d$</td>
</tr>
<tr>
<td>Shekel (f44)</td>
<td>4</td>
<td>$-10.5364$</td>
<td>$[0, 10]^d$</td>
</tr>
<tr>
<td>Levy (f45)</td>
<td>10, 30</td>
<td>0</td>
<td>$[-10, 10]^d$</td>
</tr>
<tr>
<td>Michalewicz (f46)</td>
<td>10</td>
<td>$-9.66015$</td>
<td>$[0, \pi]^d$</td>
</tr>
<tr>
<td>Shifted Griewank (f47)</td>
<td>10, 30</td>
<td>$-180$</td>
<td>$[-600, 600]^d$</td>
</tr>
<tr>
<td>Design of a gear train (f48)</td>
<td>4</td>
<td>$2.7e - 12$</td>
<td>$[12, 60]^d$</td>
</tr>
<tr>
<td>Pressure vessel (f49)</td>
<td>4</td>
<td>$7197.72893$</td>
<td>$[1.125, 0.625, 0, 0]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\times [12.5, 12.5, 240, 240]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\times [70, 3, 0.5]$</td>
</tr>
<tr>
<td>Tripod (f50)</td>
<td>2</td>
<td>0</td>
<td>$[-100, 100]^d$</td>
</tr>
<tr>
<td>Compression spring (f51)</td>
<td>3</td>
<td>$2.6254214578$</td>
<td>$[1, 0.207]^d$</td>
</tr>
</tbody>
</table>

by any algorithm within a given number $\mu_f = 3,100$ of function evaluations. $t_{p,s} = \infty$ if the condition in Equation (38) is not satisfied after $\mu_f$ evaluations.

It can be seen from the condition in Equation (38) that tolerance $\tau$ determines the accuracy of solution $x \in f_L$. Low tolerance $\tau$ means that a solution with high accuracy to $f_L$ is required. In practice, $\tau = 10^{-k}$, and $k = 1, 2, 3, 4$ are often used. Therefore, this work will state that the required accuracy is high when $\tau \leq 10^{-7}$, and 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 $\kappa$ 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 that $\mu_f$ is set to equal 3,100 is that it is enough to compute at least 100 simplex gradients for the largest dimension problem ($d = 30$). 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

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

(39)
where the performance ratio \( r_{p,s} \) is defined by

\[
    r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}, s \in S\}}.
\]

(40)

We can see that \( \rho_s(1) \) is the fraction of problems for which algorithm \( s \) performs the best, while for sufficiently large \( \alpha \), \( \rho_s(\alpha) \) 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

At this point, the focus now falls on the first question proposed at 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, it is clear that PSO-14, PSO-1, PSO-12, PSO-10, and PSO-6 are the best PSO algorithms, they can solve at least 80% of the problems, while PSO-3, PSO-4, PSO-5, and PSO-8 are the worst PSO algorithms, only solving under 30% of the problems. It is interesting to note 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 solved about 27% of the problems. In contrast, even the worst stable algorithms (PSO-11) can solve about 54% of the problems. The performance difference between the worst stable PSO algorithm (PSO-11) and the best unstable PSO algorithm (PSO-8) is about 27% (\( = 54\% - 27\% \)).

Next, the focus moves to increased accuracy, that is, decreasing \( \tau \) (see the convergence condition in Equation (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% of the problems. Therefore, we only present the best five PSOs profiles.

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

Therefore, based on the present experimental results, the answer to the first question proposed at the beginning of this section is “no.” That is to say, there is no guarantee that a stable PSO always performs better than an unstable PSO. Actually, this is not surprising. The stable analysis of PSO only requires that, when the swarm cannot find a 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 in Equation (36) is meaningful.

### 6.4 The “Best” Parameter Settings

At this point, the focus turns to the second question proposed at the beginning of this section. Note that, in general, there is probably no best parameter setting for canonical PSO. The present 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) that adopt two of the known best parameter combinations perform well, especially PSO-1. However,
there exist parameter combinations that may be better than PSO-1 ($\omega = 0.7298, \phi = 1.49618$) and ($\omega = 1/(2 \ln 2), \phi = 0.5 + \ln 2$). For example, the present 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 $40\% - 16\% = 24\%$, 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,” (e.g., $\omega = 0.42, \phi = 1.55$).

In the following, two additional experiments are described whose results are interesting and helpful in selecting a parameter combination.
The 14 PSO algorithms are tested on the 50 CEC2005 problems (see Table 2) and the 40 non-CEC2005 (see Table 3) problems. 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, the present 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
Figure 6: Data profiles ($\tau = 10^{-7}$). Only the best five PSO profiles are presented.

Figure 7: Performance profiles ($\tau = 10^{-7}$). Only the best five PSO profiles are presented.
PSOs solve very few functions or even no functions when $\tau = 10^{-7}$ only the best five PSO algorithms’ profiles are presented in Figure 9 and Figure 11.

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

For the CEC2005 functions, PSO-14 is the best algorithm for almost all budgets 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 high. Finally, PSO-14 can solve about 78\% problems when $\tau = 10^{-1}$, and 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 solutions with high accuracy. In other words, the difficulty of the CEC2005 functions lies in finding solutions with high accuracy.

Note that PSO-2 (which adopts the known best parameter combination $\omega = 0.7298$, $\phi = 1.49618$) performs well on the CEC2005 functions, and 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, as it can solve only 48\% of the problems when $\tau = 10^{-1}$ and no more than 5\% problems when $\tau = 10^{-7}$. This is the reason why PSO-2 lies outside the top five PSO algorithms for the entire 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\% of the CEC2005 problems and only about 8\% of the non-CEC2005 functions. However, PSO-1 performs better than PSO-2, and it is always one of the best five PSOs, whether 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\% of the CEC2005 problems and about 27\% of the non-CEC2005 problems. PSO-8 does not
Figure 10: Data profiles for 40 non-CEC2005 test functions ($\tau = 10^{-1}$).

perform well on CEC2005 functions; however, it is one of the best five PSOs for non-CEC2005 functions. This is the reason why PSO-8 lies inside the top five PSO algorithms for the entire 90 test problems (see Figure 6).

PSO-14 performs much better than any other PSO algorithm, whether the test functions are CEC2005 functions or not. This is the reason why PSO-14 performs very well over the entire 90 test functions.

At this section draws to a close, note that the conclusions are made based only on 3,100 function evaluations in each run. Such a budget of computational cost is enough
Figure 11: Data profiles for 40 non-CEC2005 test functions ($\tau = 10^{-7}$). Only the best five PSO profiles are presented.

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

This paper gives an analysis of the stability of PSO. The current work’s main innovation
is that, based on a weak assumption of stagnation and a weak order-2 definition of
stability, an order-2 stable region was derived for canonical PSO. The same stable region
has been proposed in some PSO literatures; however, under a stricter assumption of
stagnation and often a 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, keeping the covariances constant over time, is too strict
and also not necessary.

The results of this work suggest that the weak stagnation assumption will open
a road to comprehensive understanding about the theoretical properties of PSO. For
example, important properties to explore are knowledge diffusion in the swarm and
the following behaviors of non-dominant particles. Future work will address these
properties.

This paper also provided some experimental results. The 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
Q. Liu

this sense, the obtained stable region is meaningful. Present experimental results also show that canonical PSO with parameters $\omega = 0.42$ and $\phi = 1.55$ 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 those parameter combinations inside the stable region proposed in this paper that can result in an effective PSO. That will also be addressed in future work.

Acknowledgments

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). 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 code of the CEC2005 test functions and test functions for the Standard PSO 2011.

References


Appendix: Proof of Theorem 4

**Proof:** If \( \omega = 0 \), then there are three real roots 0, 0, \( \mu^2 + \sigma^2 \) for Equation (33). Then \( \lim_{t \to \infty} E[R^2(t)] = 0 \) requires

\[
\mu^2 + \sigma^2 < 1.
\]

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

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

\( \lim_{t \to \infty} E[R^2(t)] = 0 \) only requires \( \Phi(1) > 0 \), that is,

\[
(1 - \omega)\mu^2 + (1 + \omega)\sigma^2 < (1 + \omega)^2(1 - \omega).
\]

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

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

\( \lim_{t \to \infty} E[R^2(t)] = 0 \) also requires \( \Phi(1) > 0 \), that is, Equation (41) must hold. Moreover, the norm of two other solutions of Equation (33) must be less than 1.
Based on the condition in Equation (41), it can be ensured that the characteristic Equation (33) has a root \( r_3 \in (\omega, 1) \). What is more, Equation (33) can be rewritten in the following style

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

The other two roots of Equation (33) are then obtained

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

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

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

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

Note that Equation (41) requires

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

From Equation (43), combining with \( r_3 \in (\omega, 1) \subset (0, 1) \)

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

This implies that the right-hand side inequality of Equation (43) is not necessary. Furthermore, from Equation (33)

\[
\Phi(-1) = -1 + (\omega - \mu^2 - \sigma^2) - (\omega\mu^2 - \omega\sigma^2 - \omega^3) - \omega^3
\]

\[
= -(1 - \omega)(1 - \omega^2) - (1 + \omega)\mu^2 - (1 - \omega)\sigma^2 < 0.
\]

On the other hand, from Equation (42)

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

This implies that

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

always holds. Thus, the left-hand side inequality of Equation (43) is also not necessary.

The above analysis shows that Equation (42) is the sufficient and necessary condition for \( \lim_{t \to \infty} E[R^2(t)] = 0 \). \( \square \)