

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Convergence and Spectral Radius Analysis and Parameter Selection for the Particle Swarm Optimization Algorithm Based on the Stochastic Process

¹Long-Hua Ma, ²Ming Xu, ³Meng Shao and ⁴Zhe-Ming Lu

¹School of Information Science and Engineering, Zhejiang University Ningbo Institute of Technology, Ningbo 315100, China

²Mechanical Engineering Building, Zhaohui Campus, Zhejiang University of Technology, 18 ChaoWang Road, Hangzhou 310014, People Republic of China

³Computer Centre, Hangzhou First People's Hospital, Hangzhou 310006, China

⁴School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China

Abstract: Randomness and parameter selection in the Particle Swarm Optimization (PSO) algorithm had great influence on its performance. This study presented a formal convergence and spectral radius analysis of the standard PSO algorithm model, where some of the parameters were stochastic. Based on the analysis of the relationship of $\{\omega, c_1, c_2\}$, a sufficient condition was given to guarantee that the PSO algorithm was mean-square convergent, using the stochastic process theory. Then, the mean spectral radius was constructed. According to the relationship between the spectral radius and the convergent speed, it was shown that, a small spectral radius lead to a faster convergent speed than a big one. By optimizing the mean spectral radius of the PSO algorithm in the mean-square convergent region, a minimum spectral radius and corresponding parameter selection guidelines were derived to guarantee that the PSO algorithm was mean-square convergent and had a fast convergent speed in the stochastic sense. Finally, one parameter selection $\{c_1 = c_2 = 2, \omega = 0.4222\}$ was proposed. with the parameter, the study gave examples whose performance on benchmark functions were superior to previously published results.

Key words: Particle swarm optimization, stochastic convergence, spectral radius, parameter selection

INTRODUCTION

As a famous optimization method based on stochastic population, Particle Swarm Optimization (PSO), was first proposed by Eberhart and Kennedy (1995) and Kennedy and Eberhart (1995). PSO has been successfully and widely used in a number of optimization problems (Ren and Zhong, 2011; Chun-Hong *et al.*, 2012; Zhuo *et al.*, 2011; Tu *et al.*, 2011). High speed of convergence and relative simplicity of iteration process enable it to be a highly suitable candidate for solving complex problems.

It is well known that the parameters of the standard PSO algorithm have a great effect on the performance. And some of the parameters have interactions which are stated as the exploration and exploitation tradeoff. Since 1995, many attempts have been made to improve the performance of the PSO algorithm and it has been theoretically analyzed since then. Some of those researches are representative. For instance, an empirical

parameter selection method with the inertia weight was proposed by Shi and Eberhart (1999). Eberhart and Shi (2000) compared the inertia weights with construction the factors in PSO and found empirically that an inertia weight of 0.7298 and acceleration coefficient of 1.49618 is a good parameter choice. Clerc and Kennedy (2002) and Trelea (2003), analyzed and proposed a generalized model of the PSO was theoretically the mechanism of coefficients set to control the convergent tendencies of the system. All of those efforts make the PSO algorithm more clearly, however, few of them considered the randomness in the PSO algorithm which is very important to the PSO algorithm. Thus, the research results are more or less different from the real PSO algorithm.

In the standard PSO algorithm, since each particle in the search space has a randomized velocity, it can be analyzed with the stochastic process theory. Jin *et al.* (2007), analyzed the convergent region of the PSO algorithm in a stochastic sense and a sufficient condition of mean-square convergence of the PSO algorithm was

given. However, the convergent speed is also important in addition to the convergent precision in the PSO algorithm. Using equivalent transform, the standard PSO algorithm model can be changed to a linear time-invariant system. According to the relationship between the spectral radius of the characteristic equation and the convergent speed, the algorithm with a small spectral radius converges faster than that with a big one. In this study, the stochastic process theory is used to construct the mean spectral radius. By optimizing the mean spectral radius of the standard PSO algorithm in the mean-square convergent region, a minimum mean spectral radius and corresponding parameter selection guidelines are derived to guarantee that the standard PSO algorithm is mean-square convergent and has a fast convergent speed in the stochastic sense.

PSO ALGORITHM

Standard PSO algorithm model: The standard PSO algorithm has a swarm containing M particles to solve a problem in D-dimension space. In the algorithm, $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ denotes the current position of the ith particle of the swarm and $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ represents the rate of the ith velocity for the particle. In every step, each particle maintains a memory of its previous best position $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ it has ever visited and global best position $P_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ is also preserved. In standard PSO model, at iteration tth, the dth dimension of particle ith is manipulated according to the Eq. 1 and 2:

$$v_{id}(t+1) = \omega \cdot v_{id}(t) + \varphi_{1d}(t) \cdot (p_{id}(t) - x_{id}(t)) + \varphi_{2d}(t) \cdot (p_{gd}(t) - x_{id}(t)) \tag{1}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{2}$$

where, $\varphi_{1d}(t) = c_1 r_{1d}(t)$, $\varphi_{2d}(t) = c_2 r_{2d}(t)$, $i = 1, 2, \dots, M$, $d = 1, 2, \dots, D$, $t = 1, 2, \dots, T$, T is the maximum iteration. ω is called inertia weight. c_1 and c_2 are called cognitive coefficient and social coefficient, respectively. $r_{1d}(t)$ and $r_{2d}(t)$ are two independent uniform random numbers distributed in the range of (0, 1).

At each iteration, X_i and P_i of the current particle combined to adjust the position and velocity of the particle along each dimension. If the position or velocity exceeds the initialization position or velocity bound, it can be manipulated according to the Eq. 3 or 4, respectively. Without loss of generality, considering a minimization problem, the update equation for the dth dimension of P_g is presented in Eq. 5:

$$V_{id}(t+1) = \begin{cases} V_{dmax} & V_{id}(t+1) \geq V_{dmax} \\ V_{dmin} & V_{id}(t+1) \leq V_{dmin} \\ V_{id}(t+1) & \text{otherwise} \end{cases} \tag{3}$$

$$X_{id}(t+1) = \begin{cases} X_{dmax} & X_{id}(t+1) \geq X_{dmax} \\ X_{dmin} & X_{id}(t+1) \leq X_{dmin} \\ X_{id}(t+1) & \text{otherwise} \end{cases} \tag{4}$$

$$P_{id}(t+1) = \begin{cases} P_{id}(T) & f(x_i(t+1)) \geq f(p_i) \\ X_{id}(t+1) & f(x_i(t+1)) < f(p_i) \end{cases} \tag{5}$$

where, v_{dmax} and v_{dmin} are velocity upper bound and lower bound of dth dimension; x_{dmax} and x_{dmin} are the position upper bound and lower bound of dth dimension. Eq. 3 guarantees that the particles do not oscillate around a position and Eq. 4 guarantees that the particles do not escape the search space. In Eq. 5, the symbol f denotes the objective function that is being minimized.

Dynamical PSO algorithm model: In order to analysis the dynamical model of PSO, here let $d = 1$; then Eq. 1 and 2 can be write as follows:

$$\begin{bmatrix} v_i(t+1) \\ x_i(t+1) \end{bmatrix} = \begin{bmatrix} \omega_i(t) & -\varphi_1(t) \\ \omega_i(t) & 1 - \varphi_1(t) \end{bmatrix} \begin{bmatrix} v_i(t) \\ x_i(t) \end{bmatrix} + \begin{bmatrix} \varphi_{1i}(t) & \varphi_{2i}(t) \\ \varphi_{1i}(t) & \varphi_{2i}(t) \end{bmatrix} \begin{bmatrix} P_i(t) \\ P_g(t) \end{bmatrix} \tag{6}$$

In order to express clearly, define $A_i(t)$, $B_i(t)$, $z_i(t)$ and $p_i(t)$ as follows:

$$A_i(t) = \begin{bmatrix} \omega_i(t) & -\varphi_1(t) \\ \omega_i(t) & 1 - \varphi_1(t) \end{bmatrix}, B_i(t) = \begin{bmatrix} \varphi_{1i}(t) & \varphi_{2i}(t) \\ \varphi_{1i}(t) & \varphi_{2i}(t) \end{bmatrix}, z_i(t) = \begin{bmatrix} v_i(t) \\ x_i(t) \end{bmatrix}, p_i(t) = \begin{bmatrix} P_i(t) \\ P_g(t) \end{bmatrix}$$

Then, Eq. 6 can be derived to be:

$$z_i(t+1) = A_i(t)z_i(t) + B_i(t)p_i(t) \tag{7}$$

Hypothesize that $\omega(t)$, $\varphi_{1i}(t)$, $\varphi_{2i}(t)$, $p_i(t)$ and $p_g(t)$ are all invariant. Then $\omega = \omega(t)$, $\varphi_1 = \varphi_{1i}(t)$, $\varphi_2 = \varphi_{2i}(t)$, $p_i = p_i(t)$, $p_g = p_g(t)$, so:

$$A = \begin{bmatrix} \omega & -\varphi \\ \omega & 1 - \varphi \end{bmatrix}, B = \begin{bmatrix} \varphi_1 & \varphi_2 \\ \varphi_1 & \varphi_2 \end{bmatrix}, z_i = \begin{bmatrix} v_i \\ x_i \end{bmatrix}, p = \begin{bmatrix} P_i \\ P_g \end{bmatrix}$$

In this way, Eq. 7 is simplified as:

$$z(t+1) = Az(t) + Bp \tag{8}$$

Let $y(t) = p - x(t)$ and $p = (\varphi_1 p_i + \varphi_2 p_g) / \varphi$, then Eq. 7 can be derived to be:

$$\begin{bmatrix} v(t+1) \\ y(t+1) \end{bmatrix} = \begin{bmatrix} \omega & \varphi \\ -\omega & 1-\varphi \end{bmatrix} \begin{bmatrix} v(t) \\ y(t) \end{bmatrix} \quad (9)$$

$$Z^*(t+1) = A^* Z^*(t) \quad (10)$$

Where:

$$z^*(t) = \begin{bmatrix} v(t) \\ y(t) \end{bmatrix}, A^* = \begin{bmatrix} \omega & \varphi \\ -\omega & 1-\varphi \end{bmatrix}$$

All appearance, the Eq. 10 is a linear constant system. Therefore, it can be analyzed with the method that is used in the linear constant system.

Otherwise, in order to analyze the system with stochastic process theory, hypotheses that $\omega(t)$, $p_i(t)$ and $p_g(t)$ are constant. If the particle has the best position of the swarm, then $p_i(t) = p_g(t)$ and the Eq. 6 can be changed to:

$$\begin{bmatrix} v(t+1) \\ x(t+1) \end{bmatrix} = \begin{bmatrix} \omega & \varphi(t) \\ \omega & 1-\varphi(t) \end{bmatrix} \begin{bmatrix} v(t) \\ x(t) \end{bmatrix} + \begin{bmatrix} \varphi(t)p \\ \varphi(t)p \end{bmatrix} \quad (11)$$

where, the initial condition can be express as follows:

$$\begin{cases} V(1) = \omega v(0) - \varphi(0)x(0) + \varphi(0)p \\ X(0) = \omega v(0) + (1 - \varphi(0))x(0) + \varphi(0)p \end{cases} \quad (12)$$

Here, $\varphi(t)$ is a stochastic variable with stochastic process theory. It's easy to get some direct conclusions:

- $v(t)$ and $x(t)$ are stochastic process corresponding to $\varphi(i)$, ($i = 0, 1, \dots, t-1$)
- Stochastic variable $\varphi(i)$ and $\varphi(j)$ are mutually independent if and only if $i \neq j$, ($0 = i, j = t$)
- Stochastic variable $v(t)$, $\varphi(t)$, $x(t)$ and $\varphi(i)$ are mutually independent if $i = t$
- Define the mathematical expectation of stochastic variable $\varphi(t)$ as $E(\varphi(t)) = (c_1 + c_2)/2 = \mu$ and the standard deviation of $\varphi(t)$ is $d(\varphi(t)) = (c_1^2 + c_2^2)/12 = \sigma^2$
- Define a stochastic variable $u(t) = \omega + 1 - \varphi(t)$, then $E(u(t)) = \omega + 1 - \mu = u$, $D(u(t)) = \sigma^2$

Parameters analysis: There are many parameters that would influence the performance of the PSO algorithm which are population size M , velocity bound (v_{\min} , v_{\max}), position bound (x_{\min} , x_{\max}), maximum iteration t , inertia weight ω , accelerative coefficients c_1 and c_2 .

Population size M represents the population of particles that can search the space synchronously and it determines the evaluate order at each iteration. Generally speaking, increasing the population size M , it is advantaged to search the space drastically which can decrease the maximum iteration but increase the evaluate

order at the same time. Therefore, the best tradeoff solution is that population size M is chosen between 20 and 30 (Carlisle and Dozier, 2001).

Position bound and velocity bound are used to clamp the positions and velocities. The former is determined by the problem and the later can be set artificially. A too small value of speed will cause the particles to trap in the local optima. On the contrary, a too large value will cause the particles to oscillate in a position. So, setting a suitable velocity bound (Ozcan and Mohan, 1998) is in favor of optimizing the PSO algorithm.

Inertia weight ω , cognitive parameter c_1 and social parameter c_2 are three most important parameters which influence the performance of the PSO algorithm greatly. A large ω facilitates a global search while a small one makes for a local search. Therefore, some kinds of mechanisms are proposed as Linearly Decreasing Weight (LDW) (Shi and Eberhart, 1998; Fu *et al.*, 2012), Constriction Factor Method (CFM) (Shi and Eberhart, 1999; Liu *et al.*, 2010), etc. Whereas, all of the mechanisms set that c_1 is equal to c_2 and there is no research basically under different c_1 and c_2 .

This study focuses on analyzing the relationship between the performance of the standard PSO algorithm and the parameter range of $\{\omega, c_1, c_2\}$ to guarantee that the algorithm is mean-square convergence and has a fast convergent speed in the stochastic sense. The other factors are beyond the discussing scope of this study. Additionally, the situation with variable $\{\omega, c_1, c_2\}$ during evolution is not discussed here either which means that the parameters of $\{\omega, c_1, c_2\}$ in the standard PSO algorithm are fixed, determined by the mean spectral radius.

SPECTRAL RADIUS ANALYSIS

Mean-square convergence analysis: The position sequence $\{x(0), x(1), \dots, x(t), \dots\}$ in the iteration of PSO is a sequence of the stochastic process and if the sequence $x(t)$ is mean-square convergence, the mathematical expectation $E(u(t))$ and the standard deviation $d(u(t))$ must be convergent first.

Theorem 1: The sufficient condition of convergence for the sequence $\{E(x(0)), E(x(1)), \dots, E(x(t)), \dots\}$ is given by Eq. 13 and the sequence $\{E(x(t))\}$ will converge to p :

$$\begin{cases} 0 < \mu < 2(\omega + 1) \\ |\alpha| < 1 \end{cases} \quad (13)$$

Proof: From Eq. 11, it can be derived that:

$$x(t+2) = \omega + 1 - \varphi(t)x(t+1) - \omega x(t) + \varphi(t+1)p \quad (14)$$

Then, according to Eq. 14, the iteration equation of sequence $\{E(x(t))\}$ can be obtained:

$$E(x(t+2))-E(u(t+1)x(t+1))+E(x(t)) = \mu p \quad (15)$$

Because $u(t+1)$ and $x(t+1)$ are mutually independent, Eq. 15 can be written as follows:

$$E(x(t+1))-uE(x(t+1))+E(x(t)) = \mu p \quad (16)$$

The characteristic equation of the iterative process shown in Eq. 16 is:

$$\lambda^2 - u\lambda + \omega = 0 \quad (17)$$

When the eigenvalues of the characteristic equation Eq. 17 are in the unit circle, according to Schurz criterion, the sufficient and necessary conditions for the convergence of the solution of Eq. 17 can be obtained:

$$\begin{cases} |\omega+1| > |u| \\ |\omega| < 1 \end{cases} \quad (18)$$

Thus, it leads to $0 < \mu < 2(\omega+1)$ and $|\omega| < 1$. As a result, $0 < \mu < 2(\omega+1)$ and $|\omega| < 1$ can be specified as the sufficient condition of convergence for the iterative process $\{E(x(t))\}$.

When iterative process $\{E(x(t))\}$ is convergent, the convergent value $E(x(t))$ can be calculated using $E(x(t))-uE(x(t))+\omega E(x(t)) = \mu p$ and thus $\lim_{t \rightarrow \infty} E(x(t)) = p$.

Theorem 2: The sufficient condition of convergence for the sequence $\{d(x(0)), d(x(1)), \dots, d(x(t)), \dots\}$ is given by Eq. 14 and the sequence $\{d(x(t))\}$ will converge to 0:

$$\begin{cases} 1 + a_2 + a_1 + a_0 > 0 \\ 1 - a_2 + a_1 - a_0 > 0 \\ 1 > |a_0| \\ 1 - a_0^2 > |a_1 - a_2 a_0| \\ 0 < \mu < 2(\omega+1) \\ |\omega| < 1 \end{cases} \quad (19)$$

Proof: From the definition of standard deviation, it follows that:

$$D(x(t+2)) = E(x^2(t+2)) - E^2(x(t+2)) \quad (20)$$

where, $E(x^2(t+2))$, the first part of the right side of Eq. 20, can be calculated as follows:

$$\begin{aligned} E(x^2(t+2)) &= E((u(t+1)x(t+1) - \omega x(t) + \varphi(t+1)p)^2) \\ &= E(u^2(t+1)x^2(t+1)) + \omega^2 x^2(t) + \varphi^2(t+1)p^2 \\ &\quad - 2u(t+1)\omega x(t+1)x(t) - 2\varphi(t+1)\omega p x \\ &\quad + 2u(t+1)\varphi(t+1)p x(t+1) \end{aligned} \quad (21)$$

Notice that $u^2(t+1)$ and $x^2(t+1)$, $u(t+1)$ and $x(t+1)x(t)$, $u(t+1)\varphi(t+1)$ and $x(t+1)$, $\varphi(t+1)$ and $x(t)$ are mutually independent and then $E(u^2(t+1)) = E((1+\omega-\varphi(t+1))^2) = u^2 + \sigma^2$, $E(\varphi^2(t+1)) = u^2 + \sigma^2$ and $E(u(t+1)\varphi(t+1)) = u\mu + \sigma^2$. Thus, from Eq. 21, it can be derived to be:

$$\begin{aligned} E(x^2(t+2)) &= (u^2 + \sigma^2)E(x^2(t+1)) + \omega^2 E(x^2(t)) \\ &\quad + (\mu^2 + \sigma^2)p^2 - 2u\omega E(x(t+1)x(t)) + \\ &\quad 2(u\mu + \sigma^2)pE(x(t+1)) - 2\mu\omega p E(t) \end{aligned} \quad (22)$$

The second part of Eq. 20 $E^2(x(t+2))$ can be changed to:

$$\begin{aligned} E^2(x(t+2)) &= (uE(x(t+1)) - \omega E(x(t)) + \mu p)^2 \\ &= u^2 E^2(x(t+1)) + \omega^2 E^2(x(t)) + 2u\mu p E(x(t+1)) \\ &\quad - 2u\omega E(x(t+1))E(x(t)) + \mu^2 p^2 - 2\mu\omega p E(x(t)) \end{aligned} \quad (23)$$

Equation 22 and 23 can be combined and simplified as follows:

$$\begin{aligned} D(x(t+2)) &= (u^2 + \sigma^2)D(x(t+1)) + 2u\mu p E(x(t+1)) \\ &\quad - 2u\omega E(x(t+1))E(x(t)) - 2\mu\omega p E(x(t)) + \omega 2D(x(t)) \end{aligned} \quad (24)$$

Then $E(x(t+1)x(t)) - E(x(t+1))E(x(t))$ should be computed, where:

$$\begin{aligned} E(x(t+1)x(t)) &= E(x(t)(u(t)x(t) - \omega x(t-1) + \varphi(t)p)) \\ &= E(u(t)E(x^2(t)) - \omega E(x(t)x(t-1))) + pE(\varphi(t))E(x(t)) \\ &= uE(x^2(t)) - \omega E(x(t)x(t-1)) + \mu p E(x(t)) \end{aligned}$$

Then it follows that:

$$\begin{aligned} E(x(t+1)x(t)) - E(x(t+1))E(x(t)) \\ = uD(x(t)) - \omega(E(x(t)x(t-1)) - E(x(t))E(x(t-1))) \end{aligned} \quad (25)$$

Equation 24 and 25 can be combined and simplified, then the iteration equation of $D(x(t+2))$ can be obtained:

$$\begin{aligned} D(x(t+2)) &+ (\omega - u^2 - \sigma^2)D(x(t+1)) + \\ & (u^2\omega - \sigma^2\omega - \omega^2)D(x(t)) - \omega^3 D(x(t-1)) \\ & = \sigma^2(E(x+1)) - p^2 \end{aligned} \quad (26)$$

The characteristic equation of the iterative process shown in Eq. 26 is:

$$\lambda^2 + (\omega - u^2 - \sigma^2)\lambda + (u^2\omega - \sigma^2\omega - \omega^2)\lambda - \omega^3 = 0 \quad (27)$$

When the eigenvalues of the characteristic equation of Eq. 27 are in the unit round, according to Jury criterion (Liu *et al.*, 2010), the sufficient and necessary conditions for the convergence of the solution of Eq. 27 can be obtained:

$$\begin{cases} 1+a_2+a_1+a_0 > 0 \\ 1-a_2+a_1+a_0 > 0 \\ 1 > |a_0| \\ 1-a_0^2 > |a_1-a_2a_0| \end{cases} \quad (28)$$

Besides, since Eq. 26 contains $E(x(t+1))$, the iterative process $\{E(x(t))\}$ must be convergent which has been proved in Theorem 1. As a result, Eq. 28 and 13 can be specified as the sufficient conditions of convergence for the iterative process $\{d(x(t))\}$.

When the iterative process $\{d(x(t))\}$ is convergent, the convergent value $\{d(x(t))\}$ can be calculated using $D(x(t))+(\omega-u^2-\sigma^2)D(x(t))+(\omega^2\omega-\sigma^2\omega^2D)x(t)-\omega^3D(x(t)) = \sigma^2(p-p)^2$ That gets $\lim_{t \rightarrow \infty} D(x(t)) = 0$.

Theorem 3: Stochastic variable $x(t)$ is mean-square convergent to p , when Theorem 1 and Theorem 2 are satisfied simultaneously.

Proof: According to Theorem 1 and 2, parameters $\{\omega, \mu, \sigma\}$ satisfy Eq. 13 and 14, then $\lim_{t \rightarrow \infty} E(x(t)) = p$ and $\lim_{t \rightarrow \infty} D(x(t)) = 0$. Thus:

$$\begin{aligned} \lim_{t \rightarrow \infty} E((x(t)-p)^2) &= \lim_{t \rightarrow \infty} E(x^2(t) - 2px(t) + p^2) \\ &= \lim_{t \rightarrow \infty} (D(x(t)) + E^2(x(t)) - 2pE(x(t)) + p^2) = 0 + p^2 - 2p^2 + p^2 = 0 \end{aligned}$$

From above analysis, it's clear that if Eq. 13 and 14 are satisfied, the stochastic variable $x(t)$ is mean-square convergent to p . Otherwise, Eq. 14 can be transformed to:

$$\begin{cases} \sqrt{\mu^2 - \sigma^2} < \omega < \frac{\mu^2 - \sigma^2 + \sqrt{\mu^2 - \sigma^2}}{4\mu} \\ \frac{\mu^2 - \sigma^2 - \sqrt{\mu^2 - \sigma^2}}{4\mu} < \omega < \frac{\mu^2 - \sigma^2 + \sqrt{\mu^2 - \sigma^2}}{4\mu} \end{cases} \quad (29)$$

In the model of PSO, stochastic variable $\varphi_1(t)$ and $\varphi_2(t)$ are usually considered as two uniformly distributed random variables by $[0, c_1]$ and $[0, c_2]$. In the practical uses, $c_1=c_2=c$ and $0 < \omega < 1$ are always set. So, it is obvious that $\mu = c$, $\sigma^2 = c^2/6$ and Eq. 29 can be transformed to:

$$\frac{5c - \sqrt{25c^2 - 336c + 576}}{24} < \omega < \frac{5c + \sqrt{25c^2 - 336c + 576}}{24} \quad (30)$$

Based on above analysis, the parameter range to guarantee the convergence of iterative process $\{E(x(t))\}$ and $\{d(x(t))\}$ in the condition of $c_1 = c_2 = c$ is illustrated in Fig. 1. The blue area together with the red area in Fig. 1 corresponds to the condition of Theorem 1 and the blue area in Fig. 1 corresponds to the condition of Theorem 2. Obviously, the convergence of iterative process $\{d(x(t))\}$ is stronger than the convergence of iterative process $\{E(x(t))\}$ and the red area in Fig. 1. is the mean-square convergence.

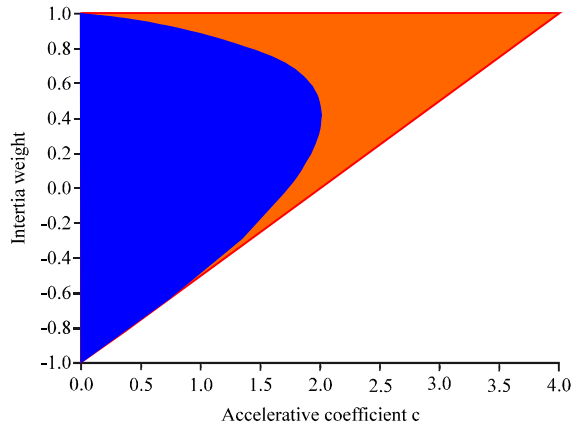


Fig. 1: Relationship between ω and c to guarantee the convergence of mathematical expectation $\{E(x(t))\}$ and the standard deviation $\{d(x(t))\}$

The blue area together with the red area corresponds to the convergence condition of process $\{E(x(t))\}$ and the blue area corresponds to the convergence condition of process $\{d(x(t))\}$.

Mean spectral radius optimization: In the standard PSO algorithm model, when given constant ω and φ , the model can be described as a linear constant system equation of Eq. 10. Here, in order to analyze the convergent ability of the model, spectral radius of the characteristic matrix A^* in Eq. 10 must be gotten firstly. Based on that, mean spectral radius can be constructed to analyze the model with constant ω and random variable φ by $[0, 4]$.

Here, the characteristic equation of Eq. 10 is:

$$\lambda^2 + (\varphi - 1 - \omega)\lambda + \omega = 0 \quad (31)$$

Thus, the eigenvalues of equation of Eq. 10 are $\lambda_1 = (a+b)/2$ and $\lambda_2 = (a-b)/2$. Where, $a = \omega + 1 - \varphi$, $b = \sqrt{\Delta}$, $\Delta = (\varphi - 1 - \omega)^2$. The spectral radius of matrix A^* is defined as $\rho(A^*) = \max\{|\lambda| : \lambda \in \lambda(A^*)\}$. When $0 < \varphi \leq 2(\omega + 1)$ and $0 < \omega \leq 1$, two cases must be considered:

- $\Delta < 0$. Here, both eigenvalues are complex numbers. Condition (1) equal to $(1 - \sqrt{\omega})^2 < \varphi < (1 + \sqrt{\omega})^2$, then $|\lambda_1|^2 = |\lambda_2|^2 = \frac{1}{4}(1 + \omega - \varphi)^2 + \frac{1}{4}[4\omega - (1 + \omega - \varphi)^2] = \omega$. So, $\max\{|\lambda| : \lambda \in \lambda(A^*)\} = \sqrt{\omega}$
- $\Delta \geq 0$. Here, both eigenvalues are real numbers, Condition (2) equal to $0 < \varphi \leq (1 - \sqrt{\omega})^2$ or $(1 + \sqrt{\omega})^2 \leq \varphi < 4$. When $0 < \varphi \leq (1 - \sqrt{\omega})^2$, $\max\{|\lambda| : \lambda \in \lambda(A^*)\} = (a+b)/2$. And when if $(1 + \sqrt{\omega})^2 \leq \varphi < 4$, $\max\{|\lambda| : \lambda \in \lambda(A^*)\} = (-a+b)/2$

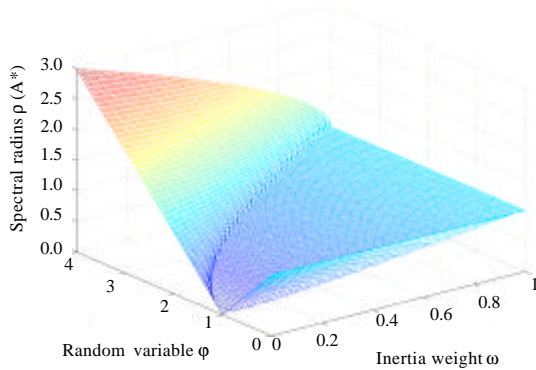


Fig. 2: Spectral radius $\rho(a^*)$ with different ω and φ

Thus, synthesize cases (1) and (2), the spectral radius can be described as Eq. 32 and showed in Fig. 2:

$$\rho(A^*) = \begin{cases} \frac{a+b}{2} & 0 < \varphi < (1-\sqrt{\omega})^2 \\ \sqrt{\omega} & (1-\sqrt{\omega})^2 < \varphi < (1+\sqrt{\omega})^2 \\ \frac{-a+b}{2} & (1+\sqrt{\omega})^2 < \varphi < 4 \end{cases} \quad (32)$$

In a linear constant system of Eq. 10, if the spectral radius $\rho(a^*)$ is less than 1, the model is convergence. And the smaller the spectral radius is, the faster the convergent speed will be. But in real PSO algorithm model, φ is random variable, by constructing and optimizing the mean spectral radius in mean-square convergence region, the PSO algorithm gets the best performance.

In the area $\rho(a^*) > 1$, PSO tend to diverge while in the area $\rho(a^*) < 1$, PSO is inclined to converge.

Lemma 1: Given constant ω and $c_1 = c_2 = 2$ (φ is random variable by $[0, 4]$), when $\omega = 0.4222$, the PSO algorithm has the best convergent precision and convergent speed.

Proof: From above analysis, it follows that $\varphi = \varphi_1 + \varphi_2$ and stochastic variable φ_1 and φ_2 are two uniformly distributed random variable by $[0, c_1]$ and $[0, c_2]$. Then the probability density function of φ can be described as:

$$f(\varphi) = \begin{cases} \varphi/(C_1 C_2) & 0 < \varphi \leq C_2 \\ 1/C_1 & C_2 < \varphi < C_1 \\ C_1 + C_2 - \varphi/(C_1 C_2) & C_1 \leq \varphi < C_1 + C_2 \\ 0 & \text{else} \end{cases} \quad (33)$$

where, $c_1 = \max\{c_1, c_2\}$, $C_2 = \min\{c_1, c_2\}$. Especially, in the condition of $c_1 = c_2$, φ obeys Simpson distribution $[0, c_1 + c_2]$. When $c_1 = c_2 = 2$, then:

$$f(\varphi) = \begin{cases} \frac{\varphi}{4} & 0 \leq \varphi \leq 2 \\ 1 - \frac{\varphi}{4} & 2 < \varphi \leq 4 \\ 0 & \text{else} \end{cases} \quad (34)$$

Thus, in the condition of constant ω and random variable φ , the mean spectral radius can be calculated by integrating $\rho(a^*)$ with the probability density function of ω to evaluate the speed of the PSO algorithm. Thus, this study defines:

$$\overline{\rho(A^*)} = \int_0^4 f(\varphi) \rho(A^*) d\varphi \quad (35)$$

where, $s = \{\varphi \in (0, 4) \times \omega \in (\underline{\omega}, \overline{\omega})\}$. $\underline{\omega}$ and $\overline{\omega}$ are the lower and the upper limits of ω in Eq. 30, where the model is mean-square convergence. When $c_1 = c_2 = 2$, by calculating, $\underline{\omega} = 1/3$ and $\overline{\omega} = 1/2$ can be obtained. Here, in order to proof the question integrality, let $s = \{\varphi \in (0, 4) \times \omega \in [0, 1]\}$. If Eq. 32 and 34 are put to Eq. 35, then:

$$\overline{\rho(A^*)} = \begin{cases} p_1 + p_2 + \frac{p_{31} - p_{32}}{8} & 3 - 2\sqrt{2} < \omega < 1 \\ p_1 + p_5 + \frac{p_{31} - p_{33} + p_{41} - p_{42}}{8} & 0 < \omega < 3 - 2\sqrt{2} \end{cases} \quad (36)$$

Where:

$$p_1 = \frac{1}{8} \left(\frac{8}{3} \omega^{1.5} - 3\omega^2 + \frac{1}{3} + \omega(\omega+1) \ln(\omega) \right)$$

$$p_2 = -\frac{\omega^{2.5}}{4} - \frac{\omega^{1.5}}{2} + 2\omega - \frac{\sqrt{\omega}}{4}$$

$$p_{31} = \frac{(3-\omega)^3}{6} - \frac{(\omega^2 - 10\omega + 9)^{1.5}}{3} + \frac{3-\omega}{2}$$

$$((3-\omega)\sqrt{\omega^2 - 10\omega + 9} - 4\omega \ln(3-\omega + \sqrt{\omega^2 - 10\omega + 9}))$$

$$p_{32} = \frac{-8\sqrt{\omega}^3}{3} + 2\omega(3-\omega) - 2(3-\omega)\omega \ln(2\sqrt{\omega})$$

$$p_{33} = -\frac{(1-\omega)^3}{3} + \frac{(1-\omega)^2(3-\omega)}{2} - \frac{(\omega^2 - 6\omega + 1)^{1.5}}{3} +$$

$$\frac{3-\omega}{2} \cdot ((1-\omega)\sqrt{\omega^2 - 6\omega + 1} - 4\omega \ln(1-\omega + \sqrt{\omega^2 - 6\omega + 1}))$$

$$p_{41} = \frac{(1-\omega)^3}{3} + \frac{(1-\omega)^2(\omega+1)}{2} + \frac{(\omega^2 - 6\omega + 1)^{1.5}}{3} +$$

$$\frac{\omega+1}{2} \cdot ((1-\omega)\sqrt{\omega^2 - 6\omega + 1} - 4\omega \ln(1-\omega + \sqrt{\omega^2 - 6\omega + 1}))$$

$$p_{42} = \frac{(2\sqrt{\omega})^3}{3} + \frac{(2\sqrt{\omega})^2(\omega+1)}{2} + \frac{\omega+1}{2} \cdot (-4\omega \ln(2\sqrt{\omega}))$$

$$P5 = \omega^2 + \omega$$

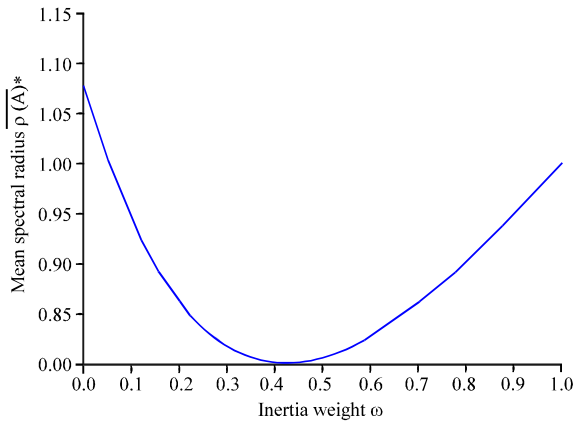


Fig. 3: Mean spectral radius $\overline{\rho(A^*)}$ with different ω , in the case of $c_1 = c_2 = 2$, $\omega = 0.4222$, $\min\{\overline{\rho(A^*)}\} = 0.8027$

Table 1: Different $c \in [1, 2.017]$ values for optimal ρ

$c_1 = c_2 = c$	$\min \overline{\rho(A^*)}$	Corresponding ω
2.017	0.8111	0.4333
2	0.8027	0.4222
1.9	0.7528	0.3590
1.8	0.7017	0.2997
1.7	0.6495	0.2447
1.6	0.5963	0.1946
1.5	0.5426	0.1501
1.4	0.4890	0.1118
1.3	0.4367	0.0802
1.2	0.3875	0.0555
1.1	0.3442	0.0375
1.0	0.3107	0.0255

From Eq. 36, it's easy to conclude that when $\omega = 0.4222$, $\overline{\rho(A^*)}$ reaches the minimum value and $\min_{\omega \in [0,1]} \overline{\rho(A^*)} = 0.8027$. The relationship between ω and $\overline{\rho(A^*)}$ is illustrated in Fig. 3. Obviously, when $0 = \omega < 0.4222$, $\overline{\rho(A^*)}$ decreases a lot with the increasing of ω . And when $0.4222 < \omega < 1$, $\overline{\rho(A^*)}$ increases lot with the increasing of ω . The relationship between ω and $\overline{\rho(A^*)}$ is illustrated in Fig. 3. Therefore, if ω is in the mean-square convergent region, it can be also concluded that when $\omega = 0.4222$, then $\min_{\omega \in [0,1]} \overline{\rho(A^*)} = 0.8027$ and the PSO algorithm model has the best convergent precision and convergent speed.

Lemma 1 is just one case in which $c_1 = c_2 = 2$. And choosing different values of c in $[1, 2.017]$, the optimal $\overline{\rho(A^*)}$ and corresponding ω changes, synchronously. With the same method, 12 representative cases are listed in Table 1. It's clear that the optimal $\overline{\rho(A^*)}$ and corresponding ω decrease a lot with the decreasing of c .

OPTIMIZATION EXPERIMENTS

Test conditions: The PSO algorithm was used to optimize several kinds of benchmark functions. In order to compare their performance four benchmark functions were tested which are Sphere, Rosenbrock, Rastrigin and Griewank,

also used in (Shi and Eberhart, 1999) and (Zheng *et al.*, 2003). All particular information of the four functions is summarized in Table 2. In all cases, the global minimum is situated at or near the center of the search space. The Sphere and the Rosenbrock functions have a single minimum while the others have multiple local minima. Here, three parameter sets are chosen to compare the performance of the PSO algorithm. According to above analysis, in the case of $c_1 = c_2 = 2$, when $\omega = 0.4222$, the PSO algorithm has the best performance, so the parameter set 1 $\{c_1 = c_2 = 2, \omega = 0.4222\}$ is selected. In order to compare the performance of the PSO algorithm with different ω , the other two parameter sets have the same c_1 and c_2 , $c_1 = c_2 = 2$ which are LDW mechanism and Sine Changing Weight (SINW) mechanism according to Eq. 37 and 38. So parameter set 2 $\{c_1 = c_2 = 2, \omega$ changes with LDW $\}$ and parameter set 3 $\{c_1 = c_2 = 2, \omega$ change with SINW $\}$ are selected.

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{\text{IterationMax}} \cdot \text{Iteration} \quad (37)$$

$$\omega = \frac{1}{4} \sin\left(\frac{\pi}{2} + \frac{\pi((\text{Iteration} \bmod 200) - 1)}{100}\right) + 0.5 \quad (38)$$

The population size M is set to be 20 and 40. The maximum iteration number T was fixed to 1000. Each optimization experiment runs 20 times with random initialvalue of x and v in the range of $[x_{\min}, x_{\max}]$ indicated in Table 2. During the optimization process, x and v are not allowed to “fly” outside the defined region. Equation 3 and 4 are used to guarantee variable bounds and it is assumed that $v_{\max} = (x_{\max} - x_{\min})/2$ and $v_{\min} = -v_{\max}$.

Optimization results and analysis: Figure 4 shows the experimental results of PSO algorithm with three parameter sets which are applied in the four benchmark functions in the case of population size $M = 20$, respectively. By looking at the shapes of the curves in all the figures in Fig. 4, it is easy to see that the curve of set 1 converges quickly in all cases, corresponding to the curve of other two sets. Otherwise, for the Sphere function, the convergence precise of set 1 is about 8 times higher than the curve of set 3 and for the Rosenbrock, Rastrigin and Griewank functions, the convergence precise of set 1 is a little higher than the other curves. On the other hand, four performance criterions of three parameter sets in the case of population size $M = 20$ are reported in Table 3 which is the mean, standard deviation, best and worst of the final fitness form 20 times experiments. From Table 3, it is clear that corresponding to all the 4 criterions for the four benchmark functions, the values of set 1 is smaller than the values of other two sets, or close to the values of other two sets.

Table 2: Some benchmark function used in the paper

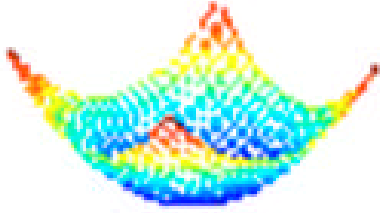
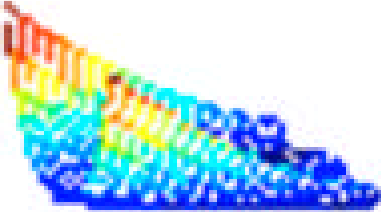
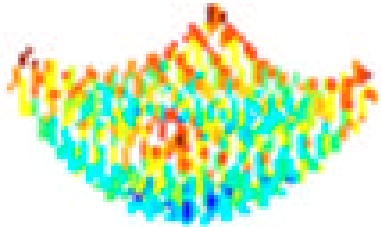
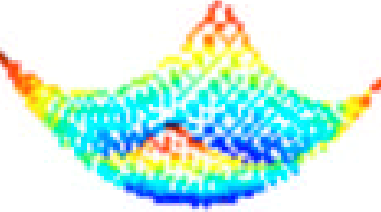
Function	Formula	Dim.	Range $[x_{min}, x_{max}]$	Optimal f	Optimal \bar{x}	Sketch in 2D
Sphere	$f_1(\bar{x}) = \sum_{i=1}^n x_i^2$	30	$[-100,100]^n$	0	$[0, \dots, 0]t$	
Rosenbrock	$f_2(\bar{x}) = \sum_{i=1}^n (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$	30	$[-2.048, 2.048]^n$	0	$[1, \dots, 1]t$	
Rastrigin	$f_3(\bar{x}) = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10)$	30	$[-5.12, 5.12]^n$	0	$[0, \dots, 0]t$	
Griewank	$f_4(\bar{x}) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(x_i / \sqrt{i}) + 1$	30	$[-600, 600]^n$	0	$[0, \dots, 0]t$	

Table 3: PSO algorithm performance with three parameter sets in the case of population size $M = 20$

Function	Sets	Mean	Std.	Best	Worst
Sphere	Set1	2.68e-10	9.95e-9	3.16e-13	4.47e-9
	Set2	500.02	2236.06	0.0008	10000
	Set3	0.0737	0.3296	6.50e-8	1.4739
Rosenbrock	Set1	30.7523	16.5883	23.825	83.0354
	Set2	101.809	202.405	23.087	925.675
	Set3	60.9237	95.5723	19.5344	444.949
Rastrigin	Set1	76.6472	25.8689	25.8689	138.369
	Set2	92.1423	32.2652	23.2652	140.392
	Set3	77.0504	31.8387	31.8387	137.522
Griewank	Set1	0.0224	2.59e-13	2.59e-13	0.0754
	Set2	0.0561	0.0031	0.0031	0.2262
	Set3	0.0546	7.63e-7	7.63e-7	0.1425

Figure 5 shows experimental results of PSO algorithm with three parameter sets which are applied in the four

benchmark functions in the case of population size $M = 40$, respectively. But four performance criterions of three parameter sets in the case of $M = 40$ is listed in Table 4. The experimental results of population size $M = 40$ is similar to that of population size $M = 20$.

From above experimental results, it can be seen that the PSO algorithm of set 1 converges fast and has a good convergent precision. The performance of the PSO algorithm of set 2 is little worse than the PSO algorithm of set 2, perhaps the inertia weight is sine changing among $[0.25, 0.75]$ and come into the mean-square convergent region $[1/3, 1/2]$, continuously. Otherwise, the mean inertia weight 0.5 is close to 0.4222. The performance of the PSO algorithm of set 3 is the worst, perhaps the inertia weight

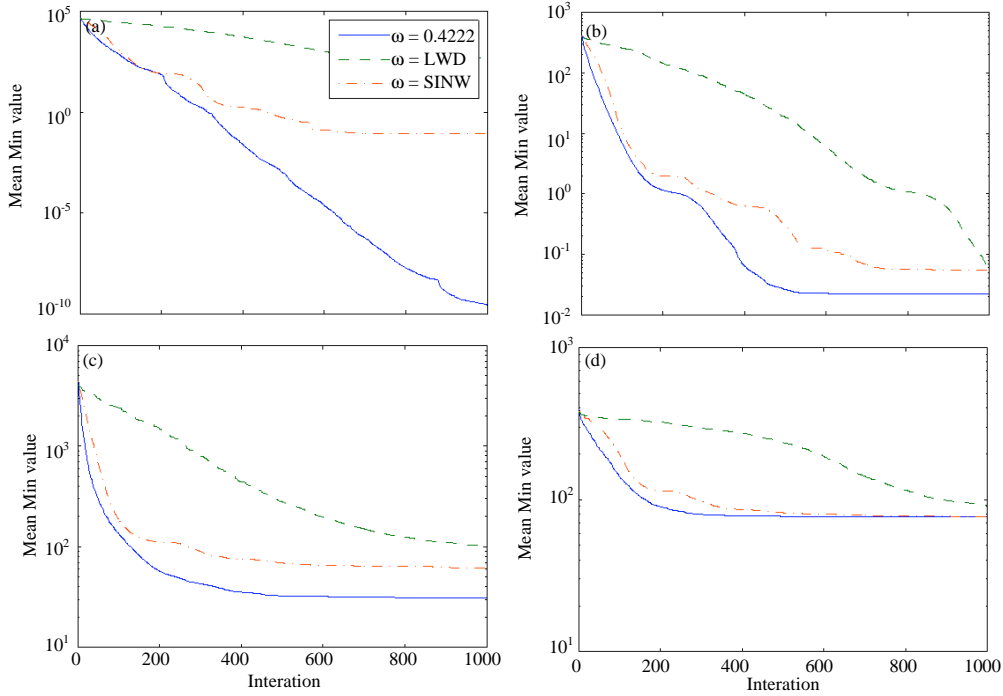


Fig. 4(a-d): Experimental results of PSO algorithm with three parameter sets applied in four benchmark functions in the case of population size $M = 20$, (a) Sphere function, (b) Griewank function, (c) Rosenbrock function and (d) Rastrigin function

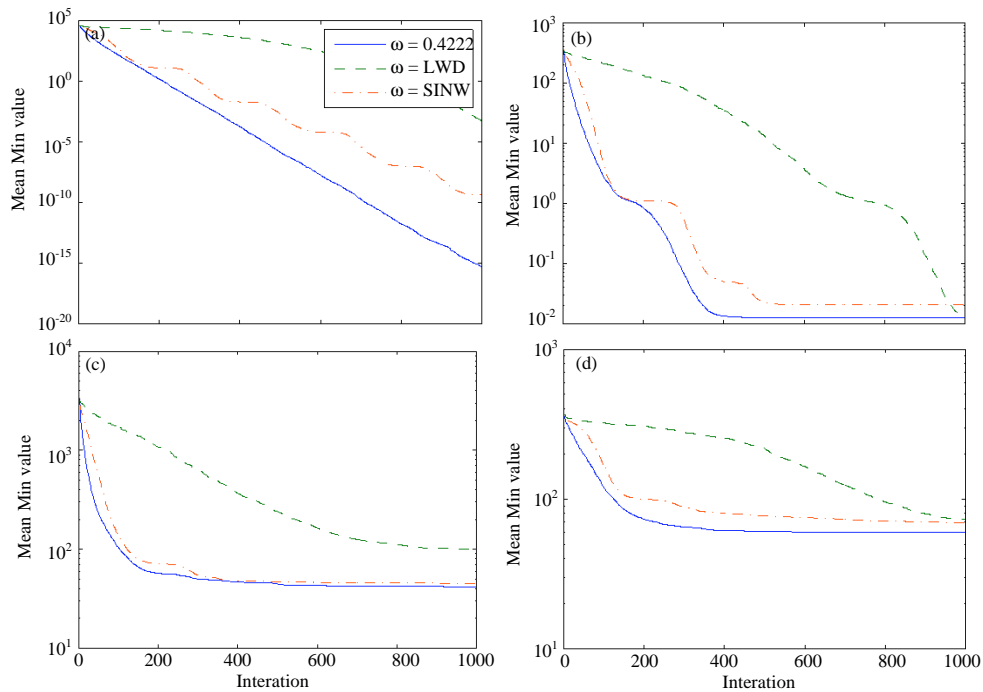


Fig. 5(a-d): Experimental results of PSO algorithm with three parameter sets applied in four benchmark functions in the case of population size $M = 40$, (a) Sphere function, (b) Griewank function, (c) Rosenbrock function and (d) Rastrigin function

Table 4: PSO algorithm performance with three parameter sets in the case of population size $M = 40$

Function	Sets	Mean	Std.	Best	Worst
Sphere	Set1	4.75e-16	1.36e-15	1.27e-18	6.15e-15
	Set2	0.0005	0.0006	0.0001	0.0026
	Set3	3.95e-10	7.84e-10	2.79e-12	3.57e-9
Rosenbrock	Set1	41.7185	47.9009	22.0274	208.082
	Set2	100.924	227.464	19.261	1056.41
	Set3	44.8479	44.3868	22.1448	208.478
Rastrigin	Set1	60.2233	21.9816	32.8336	101.556
	Set2	93.1944	31.0366	29.7631	146.486
	Set3	44.8479	26.0383	31.842	118.470
Griewank	Set1	0.0125	0.0123	0	0.0392
	Set2	0.0126	0.0136	0	0.0573
	Set3	0.0206	0.0331	5.56e-7	0.1469

is linearly decreasing from 0.9 to 0.4. A big inertia weight at the beginning is exceeded the mean-square convergent region $[1/3, 1/2]$, but come into the region in the end. So, it's obvious that the curves of the set 2 converge slowly at first but converge quickly in the end.

The characteristic of the test function influence the performance of the PSO algorithm greatly. Here, the Sphere and the Rosenbrock functions have a single minimum while the Rastrigin and Griewank functions have multiple local minima. According to the experimental results, for the Sphere and Griewank functions, in most cases, the PSO algorithm of three sets can converge to the optimal target basically while it's hard to converge to the optimal target for the Rosenbrock and Rastrigin functions. Perhaps the sketch in 2D of the Sphere and Griewank functions listed in Table 2 are simple, although the Griewank function has multiple local minima while the sketch in 2D of the Rosenbrock and Rastrigin functions is complex.

In most cases, the convergent precision improves as the number of particles increases. But in some cases, increasing the number of particles decreases the convergent precision. Some examples showed in Table 3 and 4 can illuminate this problem. Perhaps it has something to do with the characteristic of the test function itself. Considering the efficiency problem of the computer, 20 is a better choice for the population size. Obviously, the solid blue curve drops sharply and converges quickly correspondingly.

CONCLUSION

In this study, the stochastic processes theory is used to construct the mean spectral radius. By optimizing the mean spectral radius of the PSO algorithm under the mean-square convergent region, a minimum spectral radius and corresponding parameter selection guidelines is derived to guarantee that the PSO algorithm is mean-square convergent and has a fast convergent speed in the stochastic sense. One parameter selection

$\{c_1 = c_2 = 2, \omega = 0.4222\}$ is proposed and examples of performance on benchmark functions superior to previously published results are given.

As to the further research, to clarify the relationship between the cognitive parameter and the social parameter, the convergent performance in the condition of the social parameter is unequal to the cognitive parameter. Otherwise, proposing a robustness PSO algorithm is also important in the further study.

ACKNOWLEDGMENTS

This study is partially supported by the financial support from the National Nature Science Foundation of China under grant No. 61070003, 61071128 and 61272020 and the Zhejiang Provincial Natural Science Foundation of China under grant No. R1090052 and Y1101184. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers which have improved the presentation.

REFERENCES

- Carlisle, A. and G. Dozier, 2001. An off-the-shelf PSO. Proceeding of the Workshop on Particle Swarm Optimization, April 6-7, 2001, Purdue School of Engineering and Technology, Indianapolis, IN., USA., pp: 1-6.
- Chun-Hong, C., W. Li-Min, H. Chun-Yan, Z. Da-Zhe and Z. Bin, 2012. Geese PSO optimization in geometric constraint solving. Inform. Technol. J., 11: 504-507.
- Clerc, M. and J. Kennedy, 2002. The particle Swarm-explosion, stability and convergence in a multidimensional complex space. IEEE Trans. Evol. Computat., 6: 58-73.
- Eberhart, R. and J. Kennedy, 1995. A new optimizer using particle swarm theory. Proceeding 6th International Symposium on Micro Machine and Human Science, October 04-06, 1995, Nagoya, pp: 39-43.
- Eberhart, R.C. and Y. Shi, 2000. Comparing inertia weights and constriction factors in particle swarm optimization. Proceedings of the Congress on Evolutionary Computation, July 16-19, 2000, IEEE Service Center, San Diego, California, pp: 84-88.
- Fu, Y.G., M.Y. Ding and C.P. Zhou, 2012. Phase angle-encoded and quantum-behaved particle swarm optimization applied to three-dimensional route planning for UAV. IEEE Trans. Syst. Man Cybernet. Part A: Syst. Hum., 42: 511-526.
- Jin, X.L., L.H. Ma, T.J. Wu and J.X. Qian, 2007. Convergence analysis of the particle swarm optimization based on stochastic processes. Acta Automatica Sinica, 33: 1263-1268.

- Kennedy, J. and R. Eberhart, 1995. Particle swarm optimization. Proceedings of the International Conference on Neural Networks, November 27-December 1, 1995, Piscataway, New Jersey, pp: 1942-1948.
- Liu, L.L., S.X. Yang and D.W. Wang, 2010. Particle swarm optimization with composite particles in dynamic environments. *IEEE Trans. Syst. Man Cybernet. Part B: Cybernet.*, 40: 1634-1648.
- Ozcan, E. and C. Mohan, 1998. Analysis of a simple Particle swarm optimization system. Proceedings of the Artificial Neural Networks in Engineering Conference, November 1-4, 1998, StLouis, MI., USA., pp: 253-258.
- Ren, B. and W. Zhong, 2011. Multi objective optimization using chaos based PSO. *Inform. Technol. J.*, 10: 1908-1916.
- Shi, Y. and R.C. Eberhart, 1998. Parameter selection in particle swarm optimization. Proceedings of the 7th International Conference on Evolutionary Programming, March 25-27, 1998, Springer-Verlag, London, UK., pp: 591-600.
- Shi, Y. and R.C. Eberhart, 1999. Empirical study of particle swarm optimization. Proceeding of the IEEE International Conference on Evolutionary Computer, July 6-9, 1999, Piscataway, NJ., USA., pp: 1945-1950.
- Trelea, I.C., 2003. The particle swarm optimization algorithm: Convergence analysis and parameter selection. *Inform. Process Lett.*, 85: 317-325.
- Tu, J., Y. Zhan and F. Han, 2011. An improved PSO algorithm coupling with prior information for function approximation. *Inform. Technol. J.*, 10: 2226-2231.
- Zheng, Y.L., L.H. Ma, L.Y. Zhang and J.X. Qian, 2003. On the convergence analysis and parameter selection in particle swarm optimization. Proceeding of the 2nd International Conference on Machine Learning and Cybernetics, Volume 3, November 2-5, 2003, Xi'an, China, pp: 1802-1807.
- Zhuo, Y., G. Li-Li and W. Jian, 2011. Frequency domain subspace decomposition realization of UWB synchronization based on PSO. *Inform. Technol. J.*, 10: 1234-1239.