Available online at www.sciencedirect.com

ScienceDirect

Information Processing Letters 102 (2007) 8–16

Information Processing Letters

www.elsevier.com/locate/ipl

Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm

M. Jiang *, Y.P. Luo, S.Y. Yang

Department of Automation, Tsinghua University, Beijing 100084, China

Received 18 January 2006; received in revised form 20 October 2006

Available online 17 November 2006

Communicated by F. Meyer auf der Heide

Abstract

This letter presents a formal stochastic convergence analysis of the standard particle swarm optimization (PSO) algorithm, which involves with randomness. By regarding each particle's position on each evolutionary step as a stochastic vector, the standard PSO algorithm determined by non-negative real parameter tuple $\{\omega, c_1, c_2\}$ is analyzed using stochastic process theory. The stochastic convergent condition of the particle swarm system and corresponding parameter selection guidelines are derived.

Keywords: Particle swarm optimization; Analysis of algorithms; Stochastic convergence analysis; Stochastic optimization; Parameter selection

1. Introduction

The particle swarm optimization (PSO) is an algorithm for finding optimal regions of complex search spaces through the interaction of individuals in a population of particles [2]. It was developed by Kennedy and Eberhart [3] based on the social behavior metaphor. The algorithm searches a solution space by adjusting the trajectories of individual vectors, called "particles" as they are conceptualized as moving points in multidimensional space. Each particle is assigned a randomized velocity. The individual particles are attracted stochastically toward the positions of their own best fitness achieved so far and the best fitness achieved so far by any of their neighbors.

Since the first formal analysis of a simple particle swarm system presented by Ozcan and Mohan [4,5], the PSO algorithm has been theoretically analyzed by van den Bergh [8], Clerc and Kennedy [2], Yasuda et al. [9], and Trelea [7]. Although those results provide insights into how particle swarm system works, all those analysis discard the randomness in the standard PSO algorithm, and are all based on a simplified deterministic algorithm. Obviously, those analytical results more or less deviate from the real particle swarm system due to the loss of randomness.

By regarding each particle's position on each evolutionary step as a stochastic vector, the standard PSO algorithm can be analyzed using stochastic process theory. The expectation and variance of the particle's position in a simplified one-particle one-dimensional particle swarm system is calculated, and corresponding convergence property is analyzed. After that, the present work gives convergence analysis to the standard par-

* Corresponding author.
E-mail address: jming99@mails.tsinghua.edu.cn (M. Jiang).

0020-0190/$ – see front matter
doi:10.1016/j.ipl.2006.10.005

ticle swarm system, considering the random influence thoroughly. As far to the authors' knowledge, this is the first contribution to analyze the stochastic standard PSO algorithm, instead of a simplified deterministic one.

The term *Convergence* is widely used in this letter, with different meanings. If it is used to describe a deterministic sequence $\{A(t)\}$, $t = 0, 1, \ldots$, *Convergence* refers to the property that the limit

$$\lim_{t\to\infty} A(t) = A$$

exists, where A is the convergent value (a constant), and both $A(t)$ and A are scalars or vectors.

When dealing with a particle swarm system, because particles' positions involve randomness, thus no deterministic convergent result can be obtained. In this letter, *Convergence* of a particle swarm is defined as the property that for $\forall i \in \{1, 2, \ldots, M\}$, $\vec{X}_i(t)$ converges in mean square to $\vec{P}$ (i.e. $\lim_{t\to\infty} E|\vec{X}_i(t) - \vec{P}|^2 = 0$), where M is the population size of the swarm, $\vec{P}$ is a position in the search space, and $\vec{X}_i(t)$ is the position of the ith particle in the swarm at time t. Under this situation, it is said that the particle swarm system converges to $\vec{P}$. Given $E\vec{X}_i(t)$ as the expectation of $\vec{X}_i(t)$ and $D\vec{X}_i(t)$ the variance, obviously the condition that $\vec{X}_i(t)$ converges in mean square to $\vec{P}$ equals to that $E\vec{X}_i(t)$ converges to $\vec{P}$ and $D\vec{X}_i(t)$ converges to 0 simultaneously.

Generally speaking, this work can be seen as an extension of van den Bergh's work [8], though the perspectives are different. The result derived in this letter extends all above mentioned theoretical analysis results. A stochastic convergent condition of the standard particle swarm system is derived. Guidelines for parameter selection are directly given by the analysis result, both in formular and graphical form.

2. Particle swarm optimization algorithm

2.1. Standard algorithm

The standard PSO algorithm maintains a population of M particles, the PSO formulae define each particle as a potential solution to a problem in D-dimensional space, with particle i represented $\vec{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iD})$, where $i = 1, 2, \ldots, M$. Each particle also maintains a memory of its previous best position, $\vec{P}_i = (P_{i1}, P_{i2}, \ldots, P_{iD})$, and a velocity along each dimension, represented as $\vec{V}_i = (V_{i1}, V_{i2}, \ldots, V_{iD})$. The $\vec{P}$ vector of the particle with the best fitness in the neighborhood is designated $\vec{P}_g$. At each iteration, $\vec{P}_g$ and the $\vec{P}$ vector of the current particle are combined to adjust the velocity of the particle along each dimension, and that velocity is then used to compute a new position for the particle. The portion of the adjustment to the velocity influenced by the individual's previous best position is considered the cognition component, and the portion influenced by the best in the neighborhood is the social component [3].

Without loss of generality, consider a minimization task and use symbol f to denote the objective function that is being minimized. The update equation for the dth dimension of the personal best position $\vec{P}_i$ is presented in Eq. (1), with the dependence on the time step t made explicit.

$$P_i^d(t+1) = \begin{cases} P_i^d(t) & \text{if } f(\vec{X}_i(t+1)) \geqslant f(\vec{P}_i(t)), \\ X_i^d(t+1) & \text{if } f(\vec{X}_i(t+1)) < f(\vec{P}_i(t)). \end{cases} \tag{1}$$

In standard PSO algorithm [6], at iteration t, the dth dimension of particle i's velocity and position are updated using Eqs. (2) and (3) separately, where ω, c_1 and c_2 are non-negative constant real parameters, $r_{1,i}^d(t)$ and $r_{2,i}^d(t)$ are two independent uniform random numbers distributed in the range [0, 1].

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)\big(P_i^d(t) - X_i^d(t)\big) + c_2 r_{2,i}^d(t)\big(P_g^d(t) - X_i^d(t)\big), \tag{2}$$

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1). \tag{3}$$

The velocity update equation can also be described using Eq. (4), where χ, φ_1 and φ_2 are non-negative constant real parameters [2]. Obviously, by choosing appropriate parameters, Eqs. (2) and (4) are identical. In this letter, Eqs. (1)–(3) are used as standard PSO update equations.

$$V_i^d(t+1) = \chi\big(V_i^d(t) + \varphi_1 r_{1,i}^d(t)\big(P_i^d(t) - X_i^d(t)\big) + \varphi_2 r_{2,i}^d(t)\big(P_g^d(t) - X_i^d(t)\big)\big). \tag{4}$$

There exist many factors that would influence the convergence property and performance of PSO algorithm, including selection of ω, c_1 and c_2, velocity clamping, position clamping, topology of neighborhood, etc. This letter focuses on analyzing the relationship between convergence property of the standard PSO algorithm and the parameter range of ω, c_1 and c_2. Factors such as velocity clamping, position clamping, topology of neighborhood do may influence the convergence property and performance of the standard PSO algorithm, but the discussion of those factors is beyond the scope of this letter. At the same time, the situation with variable parameter values during evolution is also not discussed here. That means, the standard PSO algorithm studied here is only determined by fixed parameter tuple $\{\omega, c_1, c_2\}$. Velocity and position clamping are not considered, and star topology is investigated, i.e., the neighborhood of any particle is the whole swarm.

2.2. One-dimensional algorithm

When the particle swarm operates on an optimization problem, the values of $\vec{P}_i$ and $\vec{P}_g$ are constantly updated, as the system evolves toward an optimum. For analysis purpose, consider the situation that $\vec{P}_i$ and $\vec{P}_g$ keep constant during a period of time, then all particles evolve independently. Thus, only particle i needs to be studied. For i is chosen arbitrarily, the result can be applied to all other particles. At the same time, it appears from Eqs. (2) and (3) that each dimension is updated independently from the others. Thus, without loss of generality, the algorithm description can be reduced to the one-dimensional case. By omitting particle and dimension notations, and considering discrete time situation, update equations become:

$$V_{t+1} = \omega V_t + c_1 r_{1,t}(P_i - X_t) + c_2 r_{2,t}(P_g - X_t), \quad (5)$$

$$X_{t+1} = X_t + V_{t+1}. \quad (6)$$

It should be noticed that the above simplification is only for analysis purpose, the original standard algorithm will be recalled after the analysis is finished.

According to [8], by substituting Eq. (5) into Eq. (6), the following non-homogeneous recurrence relation is obtained:

$$X_{t+1} = \big(1 + \omega - (c_1 r_{1,t} + c_2 r_{2,t})\big) X_t - \omega X_{t-1} + c_1 r_{1,t} P_i + c_2 r_{2,t} P_g. \quad (7)$$

Notice that there exist random numbers in Eq. (7), and that X_0, X_1 are also random numbers, thus each X_t should be regarded as a random variable, and the iterative process $\{X_t\}$ should be regarded as a stochastic process. The expectation and variance of each random variable X_t can then be calculated, and the convergence property of the iterative process can be analyzed.

3. Convergence analysis

As stated in last section, considering the one-particle one-dimensional PSO algorithm with fixed P_i and P_g, the particle's position at iteration t, i.e., X_t is a random variable, thus particle's trajectory can be regarded as a stochastic process. In this section, particle position's expectation and variance will be calculated, which are deterministic processes rather than stochastic processes, thus the corresponding guaranteed convergence properties can be directly analyzed. The analysis is based on Eq. (7) instead of Eqs. (5) and (6). After that, the assumption of fixed P_i is removed, and the convergence property of particle's cognition is analyzed. Remember that those analysis are based on the simplified one-particle one-dimensional PSO system. In Section 3.4, the original standard M-particle D-dimensional PSO system will be recalled, and the result obtained from one-particle one-dimensional PSO system can be applied to analyze the convergent condition of the standard PSO system.

3.1. Convergence analysis of the expectation of particle's position

In this subsection, the iteration equation of EX_t is obtained, where EX_t is the expectation of random variable X_t. Based on the iteration equation, the convergent condition of sequence $\{EX_t\}$ is analyzed.

According to Eq. (7), iteration equation of sequence $\{EX_t\}$ can be obtained.

$$EX_{t+1} = \left(1 + \omega - \frac{c_1 + c_2}{2}\right) EX_t - \omega EX_{t-1} + \frac{c_1 P_i + c_2 P_g}{2}. \quad (8)$$

The characteristic equation of the iterative process shown in Eq. (8) is

$$\lambda^2 - \left(1 + \omega - \frac{c_1 + c_2}{2}\right)\lambda + \omega = 0. \quad (9)$$

Theorem 1. *Given $\omega, c_1, c_2 \geqslant 0$, if and only if $0 \leqslant \omega < 1$ and $0 < c_1 + c_2 < 4(1+\omega)$, iterative process $\{EX_t\}$ is guaranteed to converge to $(c_1 P_i + c_2 P_g)/(c_1 + c_2)$.*

Proof. The convergent condition of iterative process $\{EX_t\}$ is that the absolute values (or complex modulus) of both eigenvalues λ_1, λ_2 are less than 1. That is,

$$\frac{1}{2}\left|\left(1 + \omega - \frac{c_1 + c_2}{2}\right) \pm \sqrt{\left(1 + \omega - \frac{c_1 + c_2}{2}\right)^2 - 4\omega}\right| < 1.$$

Consider two cases.

(1) $(1 + \omega - \frac{c_1+c_2}{2})^2 < 4\omega$.

Here, both eigenvalues are complex numbers. $|\lambda_1|^2 = |\lambda_2|^2 = \frac{1}{4}(1+\omega-\frac{c_1+c_2}{2})^2 + \frac{1}{4}[4\omega - (1+\omega-\frac{c_1+c_2}{2})^2] = \omega$, so $\max\{|\lambda_1|, |\lambda_2|\} < 1$ requires only $\omega < 1$. Condition (1) itself requires $\omega > 0$ and $2(1+\omega-2\sqrt{\omega}) < c_1 + c_2 < 2(1+\omega+2\sqrt{\omega})$.

(2) $(1+\omega-\frac{c_1+c_2}{2})^2 \geqslant 4\omega$.

Here, both eigenvalues are real numbers. Remember that ω, c_1, c_2 are all non-negative real numbers, condition (2) is equal to $\omega \geqslant 0$, and $c_1+c_2 \leqslant 2(1+\omega-2\sqrt{\omega})$ or $c_1+c_2 \geqslant 2(1+\omega+2\sqrt{\omega})$.

If $c_1+c_2 \leqslant 2(1+\omega-2\sqrt{\omega})$, $\max\{|\lambda_1|, |\lambda_2|\} < 1$ requires only

$$\frac{1}{2}\left[\left(1+\omega-\frac{c_1+c_2}{2}\right)+\sqrt{\left(1+\omega-\frac{c_1+c_2}{2}\right)^2-4\omega}\right] < 1.$$

Thus it leads to $\omega < 1$ and $0 < c_1+c_2 \leqslant 2(1+\omega-2\sqrt{\omega})$.

If $c_1+c_2 \geqslant 2(1+\omega+2\sqrt{\omega})$, $\max\{|\lambda_1|, |\lambda_2|\} < 1$ requires only

$$\frac{1}{2}\left[\left(1+\omega-\frac{c_1+c_2}{2}\right)-\sqrt{\left(1+\omega-\frac{c_1+c_2}{2}\right)^2-4\omega}\right] > -1.$$

Thus it leads to $\omega < 1$ and $2(1+\omega+2\sqrt{\omega}) \leqslant c_1+c_2 < 4(1+\omega)$.

Synthesize cases (1) and (2), the guaranteed convergent condition of iterative process $\{EX_t\}$ is

$$0 \leqslant \omega < 1 \quad \text{and} \quad 0 < c_1+c_2 < 4(1+\omega). \tag{10}$$

When iterative process $\{EX_t\}$ is convergent, the convergent value EX can be calculated using $EX = (1+\omega-\frac{c_1+c_2}{2})EX - \omega EX + \frac{c_1P_i+c_2P_g}{2}$. That gets $EX = (c_1P_i+c_2P_g)/(c_1+c_2)$. □

As a matter of fact, if $0 \leqslant \omega < 1$ and $c_1 = c_2 = 0$, then $X_t = X_0 + \frac{\omega(1-\omega^t)}{1-\omega}V_0$. That is not an interesting case, and is not considered convergent in this letter. Similar results can be found in [7–9], but none of those results explicitly takes particle's position as a stochastic variable, so those results are somewhat vague in concept, and a reasonable explanation is hard to be given. From above analysis, it is now clear what those results actually mean.

3.2. Convergence analysis of the variance of particle's position

As we know, a sequence of stochastic variables may not converge even if corresponding expectation sequence converges. To further study the convergence property, the variance sequence should be studied. In this subsection, the iteration equation of DX_t is obtained, where DX_t is the variance of random variable X_t. Based on the iteration equation, the convergent condition of sequence $\{DX_t\}$ is analyzed.

In order to make the procedure of calculating DX_t clear, some symbols should be introduced firstly. Let $\nu = (c_1+c_2)/2$, $\mu = (c_1P_i+c_2P_g)/(c_1+c_2)$, $\psi = 1+\omega-\nu$, $R_t = c_1r_{1,t}+c_2r_{2,t}-\nu$, $Q_t = ((c_1c_2)/(c_1+c_2))(r_{2,t}-r_{1,t})(P_g-P_i)$, and $Y_t = X_t-\mu$, then from Eq. (7), it gets

$$Y_{t+1} = (\psi-R_t)Y_t - \omega Y_{t-1} + Q_t. \tag{11}$$

Obviously, Y_t is also a random variable, and $DY_t = DX_t$, $EY_t = EX_t-\mu$. Since $r_{1,t}, r_{2,t}$ are two independent uniform random numbers ranged in [0, 1], it is obvious that $ER_t = EQ_t = 0$, $DR_t = ER_t^2 = \frac{1}{12}(c_1^2+c_2^2)$, $DQ_t = EQ_t^2 = \frac{1}{6}(c_1c_2/(c_1+c_2))^2(P_g-P_i)^2$, and $E(R_tQ_t) = (c_1c_2(c_2-c_1)/12(c_1+c_2))(P_g-P_i)$. Notice that DR_t, DQ_t and $E(R_tQ_t)$ are constants, let $R = DR_t$, $Q = DQ_t$, and $T = E(R_tQ_t)$.

Notice that Y_t, Y_{t-1} are both independent on R_t, Q_t, but Y_t and Y_{t-1} are dependent. Thus EY_{t+1}^2, EY_{t+2}^2, and $E(Y_{t+1}Y_t)$ can be calculated as follows:

$$\begin{aligned} EY_{t+1}^2 &= (\psi^2+R)EY_t^2 + \omega^2EY_{t-1}^2 + Q \\ &\quad - 2\omega\psi E(Y_tY_{t-1}) - 2TEY_t, \end{aligned} \tag{12}$$

$$\begin{aligned} EY_{t+2}^2 &= (\psi^2+R)EY_{t+1}^2 + \omega^2EY_t^2 + Q \\ &\quad - 2\omega\psi E(Y_{t+1}Y_t) - 2TEY_{t+1}, \end{aligned} \tag{13}$$

$$E(Y_{t+1}Y_t) = \psi EY_t^2 - \omega E(Y_tY_{t-1}). \tag{14}$$

Then, $(12)*\omega+(13)$ is calculated to eliminate items $E(Y_{t+1}Y_t)$ and $E(Y_tY_{t-1})$, get

$$\begin{aligned} &EY_{t+2}^2 + \omega EY_{t+1}^2 \\ &\quad = (\psi^2+R)\big(EY_{t+1}^2+\omega EY_t^2\big) \\ &\qquad + \omega^2\big(EY_t^2+\omega EY_{t-1}^2\big) \\ &\qquad + Q(1+\omega) - 2\omega\psi^2EY_t^2 - 2T(EY_{t+1}+\omega EY_t). \end{aligned} \tag{15}$$

Substitute $DY_t = EY_t^2-(EY_t)^2$, $EY_{t+2} = \psi EY_{t+1} - \omega EY_t$, and $\omega EY_{t-1} = \psi EY_t - EY_{t+1}$ into Eq. (15), the iteration equation of DY_t is obtained.

$$\begin{aligned} DY_{t+2} &= (\psi^2+R-\omega)DY_{t+1} - \omega(\psi^2-R-\omega)DY_t \\ &\quad + \omega^3DY_{t-1} + R\big[(EY_{t+1})^2+\omega(EY_t)^2\big] \\ &\quad - 2T(EY_{t+1}+\omega EY_t) + Q(1+\omega). \end{aligned} \tag{16}$$

Remember that $DY_t = DX_t$, $EY_t = EX_t-\mu$, the iteration equation of DX_t can be obtained.

$$\begin{aligned} DX_{t+2} &= (\psi^2+R-\omega)DX_{t+1} \\ &\quad - \omega(\psi^2-R-\omega)DX_t + \omega^3DX_{t-1} \\ &\quad + R\big[(EX_{t+1}-\mu)^2+\omega(EX_t-\mu)^2\big] \\ &\quad - 2T\big[EX_{t+1}-\mu+\omega(EX_t-\mu)\big] \\ &\quad + Q(1+\omega). \end{aligned} \tag{17}$$

The characteristic equation of the iterative process shown in Eq. (17) is

$$\lambda^3 - (\psi^2 + R - \omega)\lambda^2 + \omega(\psi^2 - R - \omega)\lambda - \omega^3 = 0. \tag{18}$$

The iteration equation and characteristic equation of iterative process $\{DX_t\}$ are both quite complex, and it is hard to analyze these two equations directly. Fortunately, the convergent condition of the iterative process $\{DX_t\}$ defined in Eq. (17) is comparatively simple. Before discussing convergence property of iterative process $\{DX_t\}$, an auxiliary theorem is introduced. Let $f(\lambda) = \lambda^3 - (\psi^2 + R - \omega)\lambda^2 + \omega(\psi^2 - R - \omega)\lambda - \omega^3$, and $\lambda_1, \lambda_2, \lambda_3$ are three roots of the characteristic equation (18).

Theorem 2. *Given* $0 \leqslant \omega < 1$ *and* $c_1 + c_2 > 0$, *then* $f(1) > 0$ *is the necessary and sufficient condition of* $\max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} < 1$.

Proof. If $\omega = 0$, then two among three eigenvalues are zeros. Without loss of generality, let $\lambda_2 = \lambda_3 = 0$, then $\lambda_1 = \psi^2 + R > 0$. It also can be easily obtained that $f(1) = 1 - \lambda_1$. Thus $|\lambda_1| < 1$ is equal to $f(1) > 0$.

Consider another special case. If $\psi = 0$, i.e., $c_1 + c_2 = 2(1 + \omega)$, then $\lambda_1 = -\omega$ and λ_2, λ_3 are roots of equation $\lambda^2 - R\lambda - \omega^2 = 0$. Obviously, $|\lambda_1| = \omega < 1$. Since $R > 0$ and $\omega^2 \geqslant 0$, it gets $\max\{|\lambda_2|, |\lambda_3|\} = \frac{1}{2}(R + \sqrt{R^2 + 4\omega^2})$. Thus, the convergent condition is $\frac{1}{2}(R + \sqrt{R^2 + 4\omega^2}) < 1$. This leads to $1 - R - \omega^2 > 0$. At this point, $f(1) = (1 - R - \omega^2)(1 + \omega)$. Obviously, $1 - R - \omega^2 > 0$ and $f(1) > 0$ are equal, so $\max\{|\lambda_2|, |\lambda_3|\} < 1$ is equal to $f(1) > 0$.

Now consider general situations when $0 < \omega < 1$, $c_1 + c_2 > 0$ and $c_1 + c_2 \neq 2(1 + \omega)$. In order to go on with the proof, some function values should be evaluated firstly.

$$f(0) = -\omega^3 < 0; \qquad f(\omega) = -2\omega^2 R < 0;$$
$$f(-\omega) = -2\psi^2\omega^2 < 0.$$

Based on property of cubic equation with one unknown, it is known that $\lambda_1\lambda_2\lambda_3 = \omega^3 > 0$. This means that the characteristic equation (18) has three positive roots, or one positive root and two negative roots.

To prove the necessity, reduction to absurdity is adopted. $f(1)$ should not equal to zero, otherwise 1 is the eigenvalue, violating condition $\max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} < 1$. Then assume $f(1) < 0$. According to conclusions in elementary mathematics, because $f(-\omega)$, $f(0)$, $f(\omega)$ and $f(1)$ have the same sign, the number of roots in the interval $(-\omega, 0)$, $(0, \omega)$, and $(\omega, 1)$ must all be even. Thus there must be at least one root located in interval $(1, \infty)$ to satisfy $\lambda_1\lambda_2\lambda_3 = \omega^3 > 0$, violating condition $\max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} < 1$.

Now the sufficiency. Since $f(1) > 0$ and $f(\omega) < 0$, so the number of the characteristic equation's roots in the interval $(\omega, 1)$ should be odd. The number of roots in the interval $(\omega, 1)$ could not be 3, for this will cause $\lambda_1\lambda_2\lambda_3 > \omega^3$, so the number could only be 1. Without loss of generality, let the root be λ_1. Because $f(-\omega)$, $f(0)$ and $f(\omega)$ have the same sign, the number of roots in the interval $(-\omega, 0)$, $(0, \omega)$ must all be even. Thus, the other two roots λ_2, λ_3 can both be located in the interval $(-\omega, 0)$ or $(0, \omega)$; or be located in interval $(-\infty, -\omega)$ or $(1, \infty)$. Obviously, λ_2, λ_3 cannot be located in interval $(-\infty, -\omega)$ or $(1, \infty)$, for that cause $|\lambda_1\lambda_2\lambda_3| > \omega^3$. Thus, λ_2, λ_3 can only be located in the interval $(-\omega, 0)$ or $(0, \omega)$. Apparently, $\max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} < 1$ is satisfied. □

Now the parameter range to guarantee the convergence of sequence $\{DX_t\}$ can be calculated using $f(1) > 0$, where

$$f(1) = -(c_1 + c_2)\omega^2 + \left(\frac{1}{6}c_1^2 + \frac{1}{6}c_2^2 + \frac{1}{2}c_1c_2\right)\omega + c_1 + c_2 - \frac{1}{3}c_1^2 - \frac{1}{3}c_2^2 - \frac{1}{2}c_1c_2.$$

Let $A = c_1 + c_2$, $B = -(\frac{1}{6}c_1^2 + \frac{1}{6}c_2^2 + \frac{1}{2}c_1c_2)$, $C = \frac{1}{3}c_1^2 + \frac{1}{3}c_2^2 + \frac{1}{2}c_1c_2 - c_1 - c_2$. Knowing that $f(1) > 0$ and ω is a real number, it is easy to get that both $A\omega^2 + B\omega + C < 0$ and $B^2 - 4AC \geqslant 0$ should be satisfied. That is,

$$\begin{aligned} g(c_1, c_2) = {} & (c_1 + c_2)^4 - 48(c_1 + c_2)^3 \\ & + 2(c_1c_2 + 72)(c_1 + c_2)^2 \\ & + 24c_1c_2(c_1 + c_2) + c_1^2c_2^2 \\ \geqslant {} & 0, \end{aligned} \tag{19}$$

$$\frac{c_1^2 + c_2^2 + 3c_1c_2 - \sqrt{g(c_1, c_2)}}{12(c_1 + c_2)} < \omega < \frac{c_1^2 + c_2^2 + 3c_1c_2 + \sqrt{g(c_1, c_2)}}{12(c_1 + c_2)}. \tag{20}$$

As a matter of fact, Eqs. (19) and (20) implies that $c_1 + c_2 < 4(1 + \omega)$; or, in other words, $f(1) > 0$ implies that $c_1 + c_2 < 4(1 + \omega)$. This is because

$$\sqrt{g(c_1, c_2)} < \sqrt{g(c_1, c_2) + 3(c_1 + c_2)^2(c_1^2 + c_2^2)} = \left|2(c_1 + c_2)^2 - 12(c_1 + c_2) - c_1c_2\right|$$

and $2(c_1 + c_2)^2 - 12(c_1 + c_2) - c_1c_2 < 0$ holds in the parameter range determined by Eq. (19). Thus,

$$\frac{c_1^2+c_2^2+3c_1c_2-\sqrt{g(c_1,c_2)}}{12(c_1+c_2)}$$
$$> \frac{c_1^2+c_2^2+3c_1c_2+2(c_1+c_2)^2-12(c_1+c_2)-c_1c_2}{12(c_1+c_2)}$$
$$= \frac{1}{4}(c_1+c_2)-1$$

and $(c_1^2+c_2^2+3c_1c_2-\sqrt{g(c_1,c_2)})/12(c_1+c_2) < \omega$ implies that $c_1+c_2 < 4(1+\omega)$.

Theorem 3. *Given $\omega, c_1, c_2 \geqslant 0$, if and only if $0 \leqslant \omega < 1$, $c_1+c_2 > 0$ and $f(1) > 0$ are all satisfied together, iterative process $\{DX_t\}$ is guaranteed to converge to $\frac{1}{6}(c_1c_2/(c_1+c_2))^2(P_g-P_i)^2(1+\omega)/f(1)$, where $f(1) = -(c_1+c_2)\omega^2 + (\frac{1}{6}c_1^2+\frac{1}{6}c_2^2+\frac{1}{2}c_1c_2)\omega + c_1+c_2-\frac{1}{3}c_1^2-\frac{1}{3}c_2^2-\frac{1}{2}c_1c_2$.*

Proof. The iteration equation of DX_t, Eq. (17), contains items related to EX_t, thus the condition shown in Theorem 1 should be satisfied firstly to make DX_t convergent. As stated above, $f(1) > 0$ implies that $c_1 + c_2 < 4(1+\omega)$. Thus conditions $0 \leqslant \omega < 1$, $c_1+c_2 > 0$, and $f(1) > 0$ together make sure that the conditions stated in Theorem 1 are satisfied.

After EX_t is convergent, the convergent condition of iterative process $\{DX_t\}$ is that the absolute values(or complex modulus) of the three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ are all less than 1. Theorem 2 proves that, given $0 \leqslant \omega < 1$ and $c_1+c_2 > 0$, $f(1) > 0$ is the necessary and sufficient condition of $\max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} < 1$.

Thus, $0 \leqslant \omega < 1$, $c_1+c_2 > 0$, and $f(1) > 0$ together give the necessary and sufficient condition to guarantee iterative process $\{DX_t\}$ convergent. If iterative process $\{DX_t\}$ is convergent, the convergent value can be calculated using

$$\begin{aligned} DX = {} & (\psi^2+R-\omega)DX - \omega(\psi^2-R-\omega)DX \\ & + \omega^3 DX + Q(1+\omega) \\ & + R\big[(EX-\mu)^2 + \omega(EX-\mu)^2\big] \\ & - 2T\big[EX-\mu+\omega(EX-\mu)\big]. \end{aligned}$$

That gets $DX = \frac{1}{6}(c_1c_2/(c_1+c_2))^2(P_g-P_i)^2(1+\omega)/f(1)$. □

3.3. Convergence analysis of particle's cognition

In above analysis, it is supposed that the values of $\vec{P}_i$ and $\vec{P}_g$ keep constant, which is not the case in real problem solving. Here, the assumption that $\vec{P}_i$ keeps constant is removed. That is, during the evolution process, $\vec{P}_i$ is constantly updated according to Eq. (1). But the value of $\vec{P}_g$ is still supposed to keep constant and be the best position found so far. As a matter of fact, this assumption is reasonable, because the value of $\vec{P}_g$ only influences the final convergent position, and it does not influence the convergence property at all.

Here we still focus on the one-dimensional one-particle simplified PSO system. The relationship between P_i and P_g is given in the following theorem.

Theorem 4. *Given $\omega, c_1, c_2 \geqslant 0$, if iterative process $\{DX_t\}$ is guaranteed to converge and $f(1) < \frac{c_2^2(1+\omega)}{6}$, then iterative process $\{P_i(t)\}$ will converge to P_g with probability 1.*

Proof. If the iterative process $\{DX_t\}$ is guaranteed to converge, then iterative process $\{EX_t\}$ is also guaranteed to converge. So X_t will converge to a random distribution with expectation $EX = (c_1P_i + c_2P_g)/(c_1+c_2)$ and variance $DX = \frac{1}{6}(c_1c_2/(c_1+c_2))^2(P_g-P_i)^2(1+\omega)/f(1)$. No matter what the value of P_i is, if $f(1) < \frac{c_2^2(1+\omega)}{6}$, then $(P_g-EX)^2 = (c_1/(c_1+c_2))^2(P_g-P_i)^2 < DX$. Hence it is obvious that $\mathrm{Prob}(X_t = P_g) > 0$, which directly leads to $\mathrm{Prob}(\lim_{t\to\infty} P_i(t) = P_g) = 1$ due to the update equation of P_i. And if $P_i(t) = P_g$, it will be stable there. Thus it is evident that iterative process $\{P_i(t)\}$ will converge to P_g with probability 1. □

It should be noticed that the condition $f(1) < \frac{c_2^2(1+\omega)}{6}$ is only a sufficient condition to ensure the convergence, but not a necessary one. As a matter of fact, this condition is to be relaxed in future.

3.4. Convergence analysis of standard particle swarm system

The above analysis in this section are all based on the one-particle one-dimensional simplified PSO system. Here, the original standard M-particle D-dimensional stochastic PSO system is recalled.

When the particle swarm operates on an optimization problem, the value of $\vec{P}_i$ and $\vec{P}_g$ are constantly updated, as the system evolves toward an optimum. As we can see from above analysis, when $\vec{P}_g$ is fixed, under certain conditions, $\vec{P}_i$ will evolve toward $\vec{P}_g$. And if $\vec{P}_g$ changes, $\vec{P}_i$ will evolve to the new $\vec{P}_g$. Regarding the convergence property of the standard PSO system, Theorem 5 can be obtained.

Theorem 5. *Given $\omega, c_1, c_2 \geqslant 0$, if $0 \leqslant \omega < 1$, $c_1 + c_2 > 0$, and $0 < f(1) < \frac{c_2^2(1+\omega)}{6}$ are all satisfied to-*

gether, the standard particle swarm system determined by parameter tuple $\{\omega, c_1, c_2\}$ *will converge in mean square to* $\vec{P}_g$.

Proof. From the results of Theorems 1 and 3, given $\omega, c_1, c_2 \geqslant 0$, if $\vec{P}_i$ and $\vec{P}_g$ keep constant during a period of time, then if $0 \leqslant \omega < 1$, $c_1 + c_2 > 0$, and $f(1) > 0$ are all satisfied together, for each dimension d of particle i, the conclusion is that $\{EX_i^d(t)\}$ converges to $(c_1 P_i^d + c_2 P_g^d)/(c_1 + c_2)$, and $\{DX_i^d(t)\}$ converges to $\frac{1}{6}(c_1c_2/(c_1+c_2))^2(P_g^d - P_i^d)^2(1+\omega)/f(1)$. It appears from Eqs. (2) and (3) that each dimension of particle i is updated independently from the others. Thus, it can be concluded that $\{E\vec{X}_i(t)\}$ converges to $(c_1\vec{P}_i + c_2\vec{P}_g)/(c_1+c_2)$, and $\{D\vec{X}_i(t)\}$ converges to $\frac{1}{6}(c_1c_2/(c_1+c_2))^2(\vec{P}_g - \vec{P}_i)^2(1+\omega)/f(1)$. And from the result of Theorem 4, if $f(1) < \frac{c_2^2(1+\omega)}{6}$, each $\vec{P}_i$ converges to $\vec{P}_g$ with probability 1, thus it can be immediately obtained that $\{E\vec{X}_i(t)\}$ will finally converge to $\vec{P}_g$, and $\{D\vec{X}_i(t)\}$ will finally converge to 0. That means, each sequence $\{\vec{X}_i(t)\}$ will stochastically evolve toward $\vec{P}_g$ until it converges in mean square to $\vec{P}_g$. This conclusion applies to each particle, thus the whole particle swarm system will converge to $\vec{P}_g$. □

It should be noticed that Theorem 5 only declares that each particle in the standard particle swarm would converge in mean square to the best position found so far by the swarm, i.e., $\vec{P}_g$. It does not mean that the convergent position is an optimal one, or even a local optimal one. If a local optimal position is desired, the idea from Guaranteed Locally Convergence Particle Swarm Optimiser (GCPSO) proposed in van den Bergh [8] can be adopted, which would not influence the analysis results derived in this letter.

4. Parameter selection guidelines

The convergence analysis of the expectation and variance of particle's position, and convergence analysis of particle's cognition in Section 3 lead to some limitation on the relationship among the parameter tuple used in standard PSO algorithm, that is, $0 \leqslant \omega < 1$, $c_1 + c_2 > 0$, and $0 < f(1) < \frac{c_2^2(1+\omega)}{6}$. These conditions can be used to effectively guide the parameter selection of PSO algorithm. Although this parameter area is quite restricted, the suggested and widely used parameters in literatures all fall into this area, such as $\omega = 0.729$, $c_1 = 2.8 * \omega$, $c_2 = 1.3 * \omega$ [1], $\omega = 0.729$, $c_1 = c_2 = 1.49$ [2], and $\omega = 0.6$, $c_1 = c_2 = 1.7$ [7].

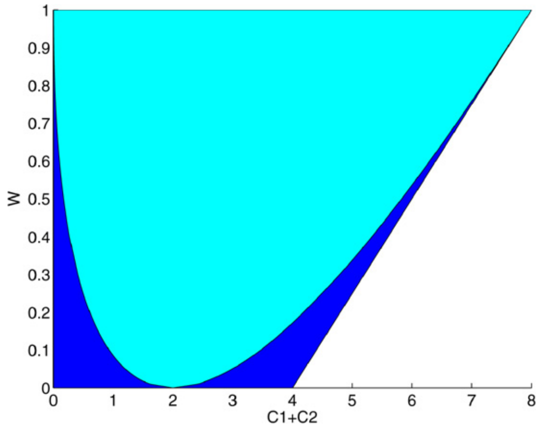

Fig. 1. Parameter range to guarantee the convergence of iterative process $\{EX_t\}$. The cyan (light) area corresponds to case with complex eigenvalues, the blue (dark) area corresponds to case with real eigenvalues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

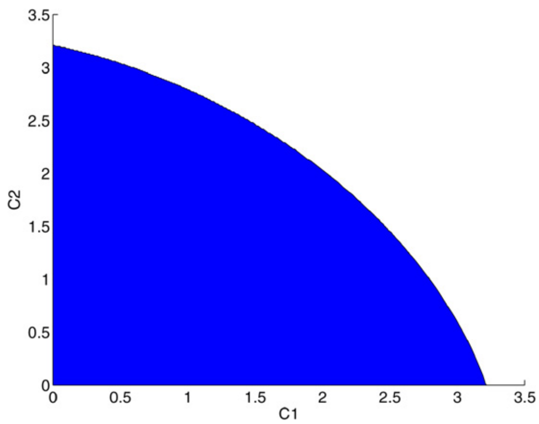

Fig. 2. Relationship between c_1 and c_2 to guarantee the convergence of iterative process $\{DX_t\}$.

Maybe the analysis result presented in this work can help to explain why those parameters work well.

The corresponding graphical illustrations of parameter ranges are given as follows in this section.

The parameter range to guarantee the convergence of iterative process $\{EX_t\}$ is illustrated in Fig. 1. The cyan (light) area in Fig. 1 corresponds to case (1) discussed in Theorem 1, and the blue (dark) area in Fig. 1 corresponds to case (2) discussed in Theorem 1.

The parameter ranges to guarantee the convergence of iterative process $\{DX_t\}$ are illustrated in Figs. 2–4. In Fig. 2, the relationship between c_1 and c_2, which is determined by Eq. (19), is illustrated. The relationship between lower and higher range of ω and c_1, c_2, which is determined by Eq. (20) and $\omega \geqslant 0$, are illustrated in Figs. 3 and 4, separately.

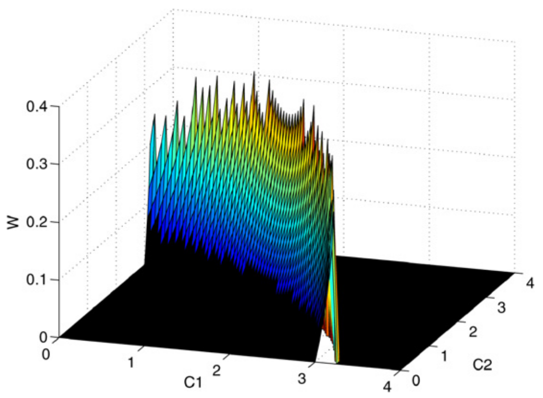

Fig. 3. Relationship between lower range of ω and c_1, c_2 to guarantee the convergence of iterative process $\{DX_t\}$.

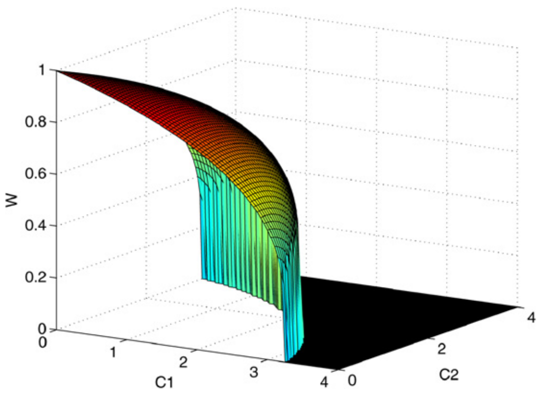

Fig. 4. Relationship between higher range of ω and c_1, c_2 to guarantee the convergence of iterative process $\{DX_t\}$.

Because the condition $f(1) < \frac{c_2^2(1+\omega)}{6}$ is only a sufficient condition, the relationship among parameter tuple $\{\omega, c_1, c_2\}$ determined by this condition is not illustrated.

If we use PSE to denote the parameter set to guarantee the convergence of iterative process $\{EX_t\}$, use PSD to denote the parameter set to guarantee the convergence of iterative process $\{DX_t\}$, and use PSC to denote the parameter set to satisfy $0 < f(1) < \frac{c_2^2(1+\omega)}{6}$, then it is obvious that the relationship should be PSE $\supset$ PSD $\supset$ PSC. If we use PSS to denote the parameter set to ensure the convergence of iterative process $\{P_i(t)\}$, it is obvious that the relationship should be PSD $\supseteq$ PSS and PSS $\supseteq$ PSC.

The parameter selection of PSO algorithm in literatures favors $c_1 = c_2 = c$, so more detailed discussion on this condition is given. At this time, Eq. (10) becomes

$$0 \leqslant \omega < 1 \quad \text{and} \quad 0 < c < 2(1+\omega). \tag{21}$$

And $f(1) > 0$ becomes

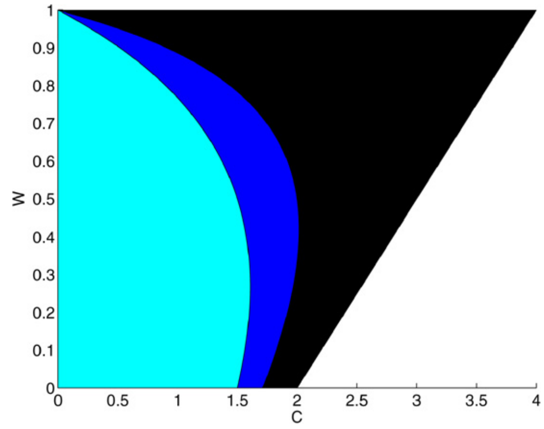

Fig. 5. Relationship between ω and c when $c_1 = c_2 = c$ to simultaneously guarantee the convergence of iterative processes $\{EX_t\}$, $\{DX_t\}$, and $\{P_i(t)\}$. The black (dark) area shows the convergence condition of process $\{EX_t\}$. The cyan (light) area shows the convergence condition of process $\{DX_t\}$. And the blue (grey) area shows a convergence condition of process $\{P_i(t)\}$ and the total particle swarm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\frac{5c - \sqrt{25c^2 - 336c + 576}}{24} < \omega < \frac{5c + \sqrt{25c^2 - 336c + 576}}{24}. \tag{22}$$

The synthetic convergence conditions are simultaneously illustrated in Fig. 5. The relationship between ω and c, which is determined by Eq. (21), is illustrated in the black (dark) area in Fig. 5. The relationship between ω and c, which is determined by Eq. (22) and $\omega \geqslant 0$, is illustrated in the cyan (light) area in Fig. 5. The relationship between ω and c, which is determined by $0 < f(1) < \frac{c_2^2(1+\omega)}{6}$, is illustrated in the blue (grey) area in Fig. 5. Obviously, the convergence condition of iterative process $\{DX_t\}$ is much stronger than the convergence condition of iterative process $\{EX_t\}$, and the convergence condition of particle's cognition is the strongest.

5. Conclusions

The stochastic process theory is applied to analyze the standard particle swarm optimization algorithm determined by parameter tuple $\{\omega, c_1, c_2\}$, considering the randomness thoroughly. The analysis results lead to a convergent condition for the standard particle swarm system, and the corresponding parameter ranges, both in formular and graphical form, are given. This result is helpful to understand the mechanism of standard PSO algorithm and select appropriate parameters to make PSO algorithm more powerful.

The result derived in this letter declares that the standard particle swarm system can converge in a sense of probability, but only a sufficient condition ensuring the convergence is given, and the detailed convergent procedure is still not quite clear now.

Some more factors influencing the convergence of particle swarm system are needed to be considered in future. Further research is also needed to clarify the relationship between PSO performance and parameter selection to guarantee convergence.

References

[1] A. Carlisle, G. Dozier, An off-the-shelf PSO, in: Proceedings of the Workshop on Particle Swarm Optimization, Indianapolis, USA, 2001.

[2] M. Clerc, J. Kennedy, The particle swarm-explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation 6 (1) (2002) 58–73.

[3] J. Kennedy, R.C. Eberhart, Particle swarm optimization, in: Proc. IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp. 1942–1948.

[4] E. Ozcan, C.K. Mohan, Analysis of a simple particle swarm optimization system, in: Intelligent Engineering Systems through Artificial Neural Networks, 1998, pp. 253–258.

[5] E. Ozcan, C.K. Mohan, Particle swarm optimization: surfing the waves, in: Proc. IEEE Congress on Evolutionary Computation (CEC 1999), Washington, DC, USA, 1999, pp. 1939–1944.

[6] Y. Shi, R.C. Eberhart, A modified particle swarm optimizer, in: Proceedings of the IEEE Congress on Evolutionary Computation (CEC 1998), Piscataway, NJ, 1998, pp. 69–73.

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

[8] F. van den Bergh, An analysis of particle swarm optimizers, PhD Dissertation, University of Pretoria, Nov. 2001.

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