Stagnation Analysis in Particle Swarm Optimization

Ming JIANG, Yupin LUO, Shiyuan YANG

Department of Automation, Tsinghua University, Beijing 100084, P.R. China

jming99@mails.tsinghua.edu.cn, {luo, ysy-dau}@tsinghua.edu.cn

Abstract – Particle swarm optimization (PSO) has shown to be an efficient, robust and simple optimization algorithm, and has been successfully applied to many different kinds of problems. But it is still an open problem that why PSO can be successful. Most of current PSO studies are empirical, with only a few theoretical analyses, and these theoretical studies concentrate mainly on simplified PSO systems, discarding randomness. In order to improve the understanding of real stochastic PSO algorithm, this paper presents a formal stochastic analysis of the stochastic PSO algorithm, which involves with randomness. The stochastic properties of particle trajectories in stagnation phase are studied in details.

I 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[1]. It was developed by Kennedy and Eberhart[2] 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.

PSO algorithm has shown to be an efficient, robust and simple optimization algorithm, and has been successfully applied to many different kinds of problems. Although many empirical studies have confirmed the success of PSO algorithm, it's still an open question that why PSO can be successful when applied to optimization problems. To gain a better and more general understanding of the behavior of particle swarm, in-depth theoretical analyses of particle trajectories are necessary. Since the first formal analysis of a simple particle swarm system presented by Ozcan and Mohan[3, 4], the PSO algorithm has been theoretically analyzed by van den Bergh[5], Clerc and Kennedy[1], Yasuda *et al*[6], and Trelea[7]. Although those results do provide insights into how particle swarm system works, all those analyses discard the randomness in the 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. Recently, researchers have begun to make progress in the analysis of randomness in PSO algorithm. Clerc[8] mathematically analyzed the stochastic behavior of the particles when the swarm is in stagnation, but he did not regard the velocity as stochastic variable, and thus he seemed unaware of the dependent relationship between velocity and the stochastic parameters. Jiang *et al*[9] studied the stochastic convergence property of the standard PSO algorithm, and gave a sufficient condition to ensure the stochastic convergence of the particle swarm system.

The particle swarm system is thought to be in stagnation, if arbitrary particle i's history best position $\vec{P}_i$ and the total swarm's history best position $\vec{P}_g$ keep constant over some time steps. When the particle swarm system is in stagnation, by regarding each particle's position on each evolutionary step as a stochastic vector, the PSO algorithm can be analyzed using stochastic process theory. Some stochastic characteristics (including expected value, variance, and auto-variance) of particle's position are obtained, both in explicit and implicit representations, and corresponding properties are analyzed.

The remainder of the paper is organized as follows: Section II provides a short introduction to PSO. Section III studies the properties of the expectation sequence of particle's position. Section IV studies the properties of the variance sequence of particle's position. Section V studies the properties of the auto-covariance sequence of particle's position. Section VI gives some discussion and the conclusion is given in Section VII.

II PARTICLE SWARM OPTIMIZATION

Standard algorithm

The 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 $\vec{X}_i = (X_{i1}, X_{i2}, ..., X_{iD})$, where i=1,2,…,M. Each particle also maintains a memory of its previous best position, $\vec{P}_i = (P_{i1}, P_{i2}, ..., P_{iD})$, and a velocity along each dimension, represented as $\vec{V}_i = (V_{i1}, V_{i2}, ..., V_{iD})$. The $\vec{P}$ vector of the particle with

1-4244-0708-7/07/$20.00 ©2007 IEEE

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

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1) \quad (2)$$

$$V_i^d(t+1) = \chi(V_i^d(t) + \phi_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + \phi_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))) \quad (3)$$

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[2].

In standard PSO algorithm[10], at iteration t, the dth dimension of particle i's velocity and position are updated using Eqs. (1) and (2) separately, where ω, c_1 and c_2 are constant real parameters, $r_{1,i}^d(t)$ and $r_{2,i}^d(t)$ are two independent uniform random numbers in the range [0,1].

The velocity update equation can also be described using Eq. (3), where χ, Φ_1, Φ_2 are constant real parameters[1]. Obviously, if appropriate parameters are chosen, Eqs. (1) and (3) are identical. In this paper, Eqs. (1) and (2) are used as standard PSO update equations.

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 paper focuses on analyzing how the selection of parameter tuple {ω, c_1, c_2} would influence the trajectories of particles in the standard PSO algorithm. Factors such as velocity clamping, position clamping, topology of neighborhood may influence the trajectories of particles, but the discussion of those factors is beyond the scope of this paper. 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 {ω, c_1, c_2}. Velocity and position clamping are not considered, and the neighborhood of any particle is the whole swarm.

Simplified one-dimensional one-particle algorithm

When the particle swarm system is in stagnation, arbitrary $\vec{P}_i$ and $\vec{P}_g$ would keep constant over some time steps, then it's easy to find out that all particles would 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. (1) and (2) that each dimension is updated independently from the other dimensions. Thus, without loss of generality, the algorithm description can be reduced to the one-dimensional case. Notice that vectors $\vec{X}_i$, $\vec{V}_i$, $\vec{P}_i$ and $\vec{P}_g$ degenerate to numeric values now, represented as X_i, V_i, P_i and P_g. By omitting particle and dimension notations, and considering discrete time situation, update equations become:

$$V_{t+1} = \omega V_t + c_1 r_{1,t}(P_i - X_t) + c_2 r_{2,t}(P_g - X_t) \quad (4)$$

$$X_{t+1} = X_t + V_{t+1} \quad (5)$$

It should be noticed that the above simplification is only for analysis purpose, and the analytical results can apply to the total particle swarm. All analyses in this paper are based on the simplified one-dimensional one-particle system, and the extension to the real D-dimensional M-particle system can be easily obtained.

According to [5], by substituting Eq. (1) into Eq. (2), the following non-homogeneous recurrence relation is obtained:

$$X_{t+1} = (1 + \omega - c_1 r_{1,t} - c_2 r_{2,t}) X_t - \omega X_{t-1} + c_1 r_{1,t} P_i + c_2 r_{2,t} P_g \quad (6)$$

Notice that there exist random numbers in Eq. (6), and that the values of 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. Some stochastic characteristics (such as expected value, variance, and auto-covariance) related to each random variable X_t can then be obtained, and the property of the stochastic particle trajectory can be analyzed.

III Expected Value of Particle's Position

As stated in last section, when considering the simplified 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, the iteration equation of EX_t is obtained, where EX_t is the expected value of particle's position X_t. Based on the iteration equation, the property of sequence $\{EX_t\}$ is analyzed.

According to Eq. (6), the iteration equation of sequence

$\{EX_t\}$ is

$$EX_{t+1} = (1+\omega-\frac{c_1+c_2}{2})EX_t - \omega EX_{t-1} + \frac{c_1P_i + c_2P_g}{2} \quad (7)$$

Obviously, if $c_1+c_2=0$, the sequence will increase linearly with time unless $c_1=c_2=0$. It's easy to see that neither of those two cases is interesting, thus we suppose $c_1+c_2\neq 0$ in the following discussion. Let $\psi = 1+\omega-\frac{c_1+c_2}{2}$, $\mu = \frac{c_1P_i + c_2P_g}{c_1+c_2}$, and $Y_t = X_t - \mu$, then

$$EY_{t+1} = \psi EY_t - \omega EY_{t-1} \quad (8)$$

The corresponding characteristic equation is

$$\lambda^2 - \psi\lambda + \omega = 0 \quad (9)$$

Given λ_{E1}, λ_{E2} as the two characteristic roots, EY_0 and EY_1 the initial conditions, then for any $t\geqslant 2$, we can get

$$EY_t = \begin{cases} [k_1(t-1)+k_2]\lambda_{E1}^{t-1}, & \text{if } \lambda_{E1} = \lambda_{E2} \\ k_3\lambda_{E1}^t + k_4\lambda_{E2}^t, & \text{otherwise} \end{cases} \quad (10)$$

and

$$EX_t = EY_t + \mu \quad (11)$$

where

$k_1 = EY_1 - \lambda_{E1}EY_0$, $k_2 = EY_1$, $k_3 = \frac{EY_1 - \lambda_{E2}EY_0}{\lambda_{E1}-\lambda_{E2}}$, and $k_4 = \frac{EY_1 - \lambda_{E1}EY_0}{\lambda_{E2}-\lambda_{E1}}$.

It's obviously that all properties of iteration process $\{EY_t\}$ are just the same as those of iteration process $\{EX_t\}$. At the same time, properties of iteration process $\{EY_t\}$ totally depend on $\lambda_{\max E} = \max\{|\lambda_{E1}|, |\lambda_{E2}|\}$, so it's necessary to study the relationship between the parameter tuple $\{\omega, c_1, c_2\}$ and $\lambda_{\max E}$, which is illustrated in the following theorem.

Theorem 1: Given a threshold $\lambda_{thE} > 0$, if and only if $2(1-\lambda_{thE})(1-\frac{\omega}{\lambda_{thE}}) < c_1+c_2 < 2(1+\lambda_{thE})(1+\frac{\omega}{\lambda_{thE}})$ and $|\omega| < \lambda_{thE}^2$ are satisfied together, then $\lambda_{\max E} < \lambda_{thE}$.

Proof: Consider two cases.

(1) $\psi^2 < 4\omega$

Here, both eigenvalues are complex numbers. And $|\lambda_{E1}|^2 = |\lambda_{E2}|^2 = \omega$, so $\lambda_{\max E} < \lambda_{thE}$ is identical to $\omega < \lambda_{thE}^2$. Condition (1) itself requires $\omega>0$ and $2(1+\omega-2\sqrt{\omega}) < c_1+c_2 < 2(1+\omega+2\sqrt{\omega})$.

(2) $\psi^2 \geq 4\omega$

Here, both eigenvalues are real numbers. Consider two more cases.

(2.1) If $\psi<0$, then $\lambda_{\max E} = -\frac{1}{2}(\psi - \sqrt{\psi^2-4\omega})$, and $\lambda_{\max E} < \lambda_{thE}$ is identical to $-\frac{1}{2}(\psi - \sqrt{\psi^2-4\omega}) < \lambda_{thE}$. One result is $0 < \omega < \lambda_{thE}^2$ and $2(1+\omega+2\sqrt{\omega}) \leq c_1+c_2 < 2(1+\lambda_{thE})(1+\frac{\omega}{\lambda_{thE}})$; and another result is $-\lambda_{thE}^2 < \omega \leq 0$ and $2(1+\omega) \leq c_1+c_2 < 2(1+\lambda_{thE})(1+\frac{\omega}{\lambda_{thE}})$

(2.2) If $\psi\geqslant 0$, then $\lambda_{\max E} = \frac{1}{2}(\psi + \sqrt{\psi^2-4\omega})$, and $\lambda_{\max E} < \lambda_{thE}$ is identical to $\frac{1}{2}(\psi + \sqrt{\psi^2-4\omega}) < \lambda_{thE}$. One result is $0 < \omega < \lambda_{thE}^2$ and $2(1-\lambda_{thE})(1-\frac{\omega}{\lambda_{thE}}) < c_1+c_2 \leq 2(1+\omega-2\sqrt{\omega})$; and another result is $-\lambda_{thE}^2 < \omega \leq 0$ and $2(1-\lambda_{thE})(1-\frac{\omega}{\lambda_{thE}}) < c_1+c_2 \leq 2(1+\omega)$

Synthesize case (1) and case (2), it is concluded that $\lambda_{\max E} < \lambda_{thE}$ is guaranteed if and only if $2(1-\lambda_{thE})(1-\frac{\omega}{\lambda_{thE}}) < c_1+c_2 < 2(1+\lambda_{thE})(1+\frac{\omega}{\lambda_{thE}})$ and $|\omega| < \lambda_{thE}^2$ are satisfied together. ■

Fig. 1 shows three examples of parameter ranges corresponding to different threshold λ_{thE}. The red(dark) area is the parameter range to guarantee $\lambda_{\max E}<1$, the cyan(grey) area is the parameter range to guarantee

λ_{maxE}<0.8, and the yellow(light) area is the parameter range to guarantee λ_{maxE}<0.5.

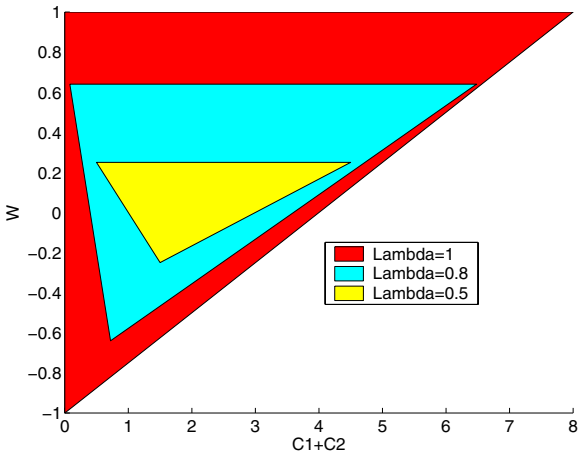

Figure 1. Parameter ranges to guarantee λ_{maxE}<λ_{thE}. The red(dark) area corresponds to λ_{thE}=1, the cyan(grey) area corresponds to λ_{thE}=0.8, and the yellow(light) area corresponds to λ_{thE}=0.5.

Now we can study the property of iteration sequence $\{EY_t\}$, which is the same as that of iteration sequence $\{EX_t\}$. Obviously, if $|\lambda_{\max E}|>1$, the sequence will diverge, which is not a desired property. If $|\lambda_{\max E}|=1$, the sequence will maintain constant amplitude vibration or increase linearly with time, neither of which is a desired property. For other cases, i.e. $|\lambda_{\max E}|<1$, sequence $\{EY_t\}$ will converge to zero and sequence $\{EX_t\}$ will converge to EX=μ. Based on above analysis, the parameter range to guarantee the convergence of iteration sequence $\{EX_t\}$ can thus be obtained, illustrated in the following corollary.

Corollary 1: If and only if $0<c_1+c_2<4(1+\omega)$ and $|\omega|<1$ are satisfied together, iterative process $\{EX_t\}$ is guaranteed to converge to $\dfrac{c_1P_i+c_2P_g}{c_1+c_2}$.

The parameter range to guarantee the convergence of sequence $\{EX_t\}$ is shown in the red(dark) area in Fig 1. When sequence $\{EX_t\}$ can converge, the convergence speed is determined by λ_{maxE}. As a matter of fact, if $|\omega|<1$ and c_1=c_2=0, then $X_t=X_0+\sum_{i=1}^{t}\omega^iV_0$. Although the sequence can converge, this is not an interesting case at all, and is not considered convergent in this paper.

Results similar to Corollary 1 can be found in [6, 7], but neither of those two 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. Trelea gave detailed analyses to properties of characteristic equation shown in Eq. (9), interested readers can refer to [7].

IV Variance of Particle's Position

In this section, the iteration equation of DX_t is obtained, where DX_t is the variance of particle's position X_t. Based on the iteration equation, the property of sequence $\{DX_t\}$ is analyzed. For the reason explained in last section, we still suppose $c_1+c_2\neq0$ in the following discussion.

In order to make the procedure of calculating DX_t clear, some symbols should be introduced firstly. Symbols ψ, μ, and Y_t are the same as that defined in Section III. Two new symbols are $R_t=c_1r_{1,t}+c_2r_{2,t}-\dfrac{c_1+c_2}{2}$ and $Q_t=\dfrac{c_1c_2}{c_1+c_2}(r_{2,t}-r_{1,t})(P_g-P_i)$, then from Eq. (6), get

$$Y_{t+1}=(\psi-R_t)Y_t-\omega Y_{t-1}+Q_t \tag{12}$$

Obviously, Y_t is also a random variable, and DY_t=DX_t, EY_t=EX_t-μ. Since $r_{1,t}$ and $r_{2,t}$ are two independent uniform random number ranged in [0, 1], it's obvious that

$$ER_t=EQ_t=0,\quad DR_t=ER_t^2=\frac{1}{12}(c_1^2+c_2^2),$$

$$DQ_t=EQ_t^2=\frac{1}{6}\left(\frac{c_1c_2}{c_1+c_2}\right)^2(P_g-P_i)^2,\quad\text{and}$$

$$E(R_tQ_t)=\frac{c_1c_2(c_2-c_1)}{12(c_1+c_2)}(P_g-P_i).$$

Because DR_t, DQ_t and $E(R_tQ_t)$ are all constants and are independent on time step t, we can 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, and $E(Y_{t+1}Y_t)$ can be calculated, as shown in Eqs (13) and (14).

We know that $EY_t^2=DY_t+(EY_t)^2$. Through simple calculation, the iteration equation of DY_t can be obtained, as shown in Eq. (15). And remember that DY_t=DX_t, EY_t=EX_t-μ, the iteration equation of DX_t can also be easily obtained, as shown in Eq. (16).

$$EY_{t+1}^2 = (\psi^2 + R)EY_t^2 + \omega^2 EY_{t-1}^2 + Q - 2\omega\psi E(Y_t Y_{t-1}) - 2TEY_t \quad (13)$$

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

$$DY_{t+2} = (\psi^2 + R - \omega)DY_{t+1} - \omega(\psi^2 - R - \omega)DY_t + \omega^3 DY_{t-1} + R\cdot[(EY_{t+1})^2 + \omega(EY_t)^2] - 2T(EY_{t+1} + \omega EY_t) + Q(1+\omega) \quad (15)$$

$$DX_{t+2} = (\psi^2 + R - \omega)DX_{t+1} - \omega(\psi^2 - R - \omega)DX_t + \omega^3 DX_{t-1} + g(t+1) \quad (16)$$

where $g(t+1) = R\cdot[(EX_{t+1} - \mu)^2 + \omega(EX_t - \mu)^2] - 2T[EX_{t+1} - \mu + \omega(EX_t - \mu)] + Q(1+\omega)$ and $g(0) = g(1) = 0$.

Both iterative processes $\{DX_t\}$ and $\{DY_t\}$ have the same characteristic equation, which is

$$\lambda^3 - (\psi^2 + R - \omega)\lambda^2 + \omega(\psi^2 - R - \omega)\lambda - \omega^3 = 0 \quad (17)$$

Before studying the property of iteration sequence $\{DX_t\}$, the property of the characteristic equation is studied firstly. Given λ_{D1}, λ_{D2}, λ_{D3} as the three characteristic roots, and let

$$f(\lambda) = \lambda^3 - (\psi^2 + R - \omega)\lambda^2 + \omega(\psi^2 - R - \omega)\lambda - \omega^3 \quad (18)$$

then we can determine the intervals in which the three roots are located. Remember that we suppose $c_1+c_2 \neq 0$, which leads to $R>0$.

First of all, consider two special cases.

If $\omega=0$, then two among three eigenvalues are zeros. Without loss of generality, let $\lambda_{D2}=\lambda_{D3}=0$, then $\lambda_{D1} = \psi^2 + R > |\omega|$.

If $\psi=0$, i.e. $c_1+c_2=2(1+\omega)$, then $\lambda_{D3}=-\omega$ and λ_{D1}, λ_{D2} are roots of equation $\lambda^2 - R\lambda - \omega^2 = 0$. We can get that $\lambda_{D1} = \frac{1}{2}(R + \sqrt{R^2 + 4\omega^2}) > |\omega|$. Since $\lambda_{D1}\lambda_{D2}\lambda_{D3} = \omega^3$, $|\lambda_{D2}| < |\omega|$ must be satisfied.

Now we will study two general cases.

If $\omega>0$, it is easily verified that

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

According to conclusions in elementary mathematics, because f(-ω), f(0), f(ω) have the same sign, the number of roots in the interval (-ω, 0), (0, ω) must be even. Thus there must be one root located in interval (ω, ∞) to satisfy $\lambda_{D1}\lambda_{D2}\lambda_{D3} = \omega^3 > 0$.

If $\omega<0$, it is easily verified that

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

Likely, according to conclusions in elementary mathematics, there must be one root located in the interval (ω, 0) and one root located in the interval (0, -ω). The third root must be located in the interval (-ω, ∞) to satisfy $\lambda_{D1}\lambda_{D2}\lambda_{D3} = \omega^3 < 0$.

Without loss of generality, suppose $|\lambda_{D1}| \geq |\lambda_{D2}| \geq |\lambda_{D3}|$, then it is clear that $\lambda_{D1} > |\omega| \geq |\lambda_{D2}| \geq |\lambda_{D3}|$, and λ_{D1} must be a positive real number.

Now let's get back to study the property of iteration sequence $\{DX_t\}$. Given DX_0, DX_1 and DX_2 the initial conditions, then for any $t\geqslant 3$, we can get the explicit representation of DX_t, which is shown in Eq. (19).

It's obviously that the property of iteration process $\{DX_t\}$ is totally dependent on the value $\lambda_{\max D} = \max\{|\lambda_{D1}|, |\lambda_{D2}|, |\lambda_{D3}|\}$, i.e. $\lambda_{\max D} = \lambda_{D1}$, so it's necessary to study the relationship between the parameter tuple $\{\omega, c_1, c_2\}$ and λ_{maxD}, which is illustrated in the following theorem.

Theorem 2: Given a threshold $\lambda_{thD}>0$, if and only if $|\omega| < \lambda_{thD}$ and $f(\lambda_{thD}) > 0$ are satisfied together, then $\lambda_{\max D} < \lambda_{thD}$.

Proof: As the above analytical result shows, if $|\omega| < \lambda_{thD}$ and $f(\lambda_{thD}) > 0$ are satisfied together, then λ_{maxD} must be located in the interval $(|\omega|, \lambda_{thD})$, so $\lambda_{\max D} < \lambda_{thD}$.

Similarly, if $\lambda_{\max D} < \lambda_{thD}$, because $\lambda_{\max D} > |\omega|$, so λ_{maxD} falls into the interval $(|\omega|, \lambda_{thD})$, which leads to $|\omega| < \lambda_{thD}$ and $f(\lambda_{thD}) > 0$. ■

Given a threshold $\lambda_{th}\geqslant 1$, then $\lambda_{\max D} < \lambda_{th}$ implies that $\lambda_{\max E} < \lambda_{th}$. The reason is as follows. As Theorem 2

$$DX_t = \begin{cases} m_1\lambda_{D1}^t + [m_2(t-3)+m_3]\lambda_{D2}^{t-1} + \sum_{j=2}^{t-1}[m_4\lambda_{D1}^j + (m_5(j-2)+m_6)\lambda_{D2}^{j-1}]g(t+1-j), & \text{if } \lambda_{D2}=\lambda_{D3} \\ \sum_{i=1}^{3} n_i\lambda_{Di}^t + \sum_{i=1}^{3}\sum_{j=2}^{t-1} n_{i+3}\lambda_{Di}^j g(t+1-j), & \text{otherwise} \end{cases} \quad (19)$$

where

$$m_1 = \frac{DX_2 - 2\lambda_{D2}DX_1 + \lambda_{D2}^2 DX_0}{(\lambda_{D2}-\lambda_{D1})^2},\quad m_2 = \frac{DX_2 - (\lambda_{D1}+\lambda_{D2})DX_1 + \lambda_{D1}\lambda_{D2}DX_0}{\lambda_{D2}-\lambda_{D1}},$$

$$m_3 = \frac{(2\lambda_{D2}-3\lambda_{D1})DX_2 - (\lambda_{D2}^2 - 3\lambda_{D1}^2)DX_1 + \lambda_{D1}\lambda_{D2}(\lambda_{D2}-2\lambda_{D1})DX_0}{(\lambda_{D2}-\lambda_{D1})^2},\quad m_4 = \frac{1}{(\lambda_{D2}-\lambda_{D1})^2},\quad m_5 = \frac{1}{\lambda_{D2}-\lambda_{D1}},$$

$$m_6 = \frac{\lambda_{D2}-2\lambda_{D1}}{(\lambda_{D2}-\lambda_{D1})^2}.\quad n_1 = \frac{DX_2 - (\lambda_{D2}+\lambda_{D3})DX_1 + \lambda_{D2}\lambda_{D3}DX_0}{(\lambda_{D1}-\lambda_{D2})(\lambda_{D1}-\lambda_{D3})},\quad n_2 = \frac{DX_2 - (\lambda_{D1}+\lambda_{D3})DX_1 + \lambda_{D1}\lambda_{D3}DX_0}{(\lambda_{D2}-\lambda_{D1})(\lambda_{D2}-\lambda_{D3})},$$

$$n_3 = \frac{DX_2 - (\lambda_{D1}+\lambda_{D2})DX_1 + \lambda_{D1}\lambda_{D2}DX_0}{(\lambda_{D3}-\lambda_{D1})(\lambda_{D3}-\lambda_{D2})},\quad n_4 = \frac{1}{(\lambda_{D1}-\lambda_{D2})(\lambda_{D1}-\lambda_{D3})},\quad n_5 = \frac{1}{(\lambda_{D2}-\lambda_{D1})(\lambda_{D2}-\lambda_{D3})}, \text{ and}$$

$$n_6 = \frac{1}{(\lambda_{D3}-\lambda_{D1})(\lambda_{D3}-\lambda_{D2})}.$$

declares, $\lambda_{\max D} < \lambda_{th}$ is identical to $|\omega| < \lambda_{th}$ and $f(\lambda_{th}) > 0$, these two conditions directly lead to $\psi^2 < \frac{(\lambda_{th}+\omega)^2}{\lambda_{th}} - \frac{\lambda_{th}+\omega}{\lambda_{th}-\omega}R < \frac{(\lambda_{th}+\omega)^2}{\lambda_{th}}$. Obviously, $\lambda_{th} + \frac{\omega^2}{\lambda_{th}} - (\lambda_{th}^2 + \frac{\omega^2}{\lambda_{th}^2}) = (1-\lambda_{th})(\lambda_{th} - \frac{\omega^2}{\lambda_{th}^2}) \le 0$ is satisfied due to $\lambda_{th} \geqslant 1$ and $|\omega| < \lambda_{th}$. Thus $\psi^2 < (\lambda_{th} + \frac{\omega}{\lambda_{th}})^2$, which directly leads to $2(1-\lambda_{th})(1-\frac{\omega}{\lambda_{th}}) < c_1 + c_2 < 2(1+\lambda_{th})(1+\frac{\omega}{\lambda_{th}})$. At the same time, $\lambda_{th} \geqslant 1$ and $|\omega| < \lambda_{th}$ together guarantee $|\omega| < \lambda_{th}^2$. Thus $\lambda_{\max E} < \lambda_{th}$ is guaranteed.

When considering the situation that $c_1=c_2=c$, Fig. 2 shows three examples of parameter ranges corresponding to different threshold λ_{thD}. The red(dark) area is the parameter range to guarantee $\lambda_{maxD}<1$, the cyan(grey) area is the parameter range to guarantee $\lambda_{maxD}<0.8$, and the yellow(light) area is the parameter range to guarantee $\lambda_{maxD}<0.5$.

Now we can study the property of iteration sequence $\{DX_t\}$. Obviously, if $\lambda_{maxD}>1$, the sequence will diverge, which is not a desired property. If $\lambda_{maxD}=1$, the sequence will also diverge unless $Q(1+\omega)=0$, which also is not desired. If $\lambda_{maxD}<1$, then $\lambda_{maxE}<1$, too. Thus, if $\lambda_{maxD}<1$, the expectation sequence will converge to $EX=\mu$ and the variance sequence will also converge. The convergent value can be easily calculated to be $DX = \frac{Q(1+\omega)}{f(1)}$. Based on above analysis, the parameter range to guarantee the convergence of iteration sequence $\{DX_t\}$ can thus be obtained, illustrated in the following corollary.

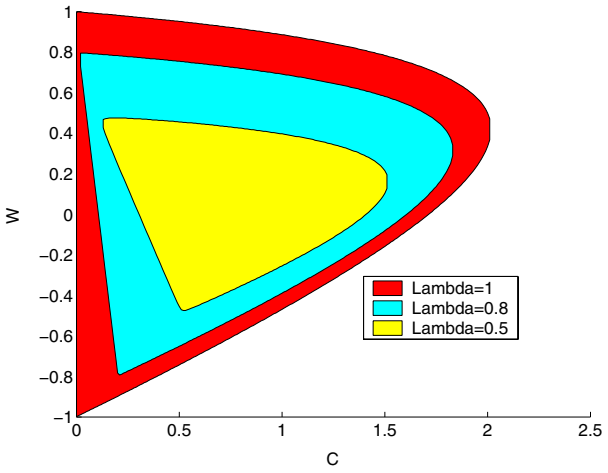

Figure 2. Parameter ranges to guarantee $\lambda_{maxD}<\lambda_{thD}$. The red(dark) area corresponds to $\lambda_{thD}=1$, the cyan(grey) area corresponds to $\lambda_{thD}=0.8$, and the yellow(light) area corresponds to $\lambda_{thD}=0.5$.

Corollary 2: If and only if $|\omega| < 1$ and $f(1) > 0$ are satisfied together, iterative process $\{DX_t\}$ is guaranteed to converge to $\frac{Q(1+\omega)}{f(1)}$.

$$E(X_{t+\tau}X_t) = \psi E(X_{t+\tau-1}X_t) - \omega E(X_{t+\tau-2}X_t) + \frac{c_1P_i + c_2P_g}{2} EX_t \quad (20)$$

The parameter range to guarantee the convergence of sequence $\{DX_t\}$ is shown in the red(dark) area in Fig 2. When sequence $\{DX_t\}$ can converge, the convergence speed is determined by λ_{maxD}.

V Auto-covariance of Particle's Position

In this section, the auto-covariance sequences of particle's position will be analyzed. Given a parameter $\tau>0$, according to Eq. (6), and notice that X_t, $X_{t+\tau-1}$ and $X_{t+\tau-2}$ are all independent on $r_{1,t+\tau-1}$ and $r_{2,t+\tau-1}$, we can get the iteration equation of $E(X_{t+\tau}X_t)$, as shown in Eq. (20).

Thus the iteration equation of $\gamma_t(\tau)$ can be obtained, which is

$$\gamma_t(\tau) = \psi\gamma_t(\tau-1) - \omega\gamma_t(\tau-2) \quad (21)$$

where $\gamma_t(\tau) = Cov(X_{t+\tau}, X_t)$ is the auto-covariance of random sequence $\{X_t\}$. The corresponding characteristic equation is

$$\lambda^2 - \psi\lambda + \omega = 0 \quad (22)$$

Obviously, $\gamma_t(-\tau) = \gamma_{t-\tau}(\tau)$ and $\gamma_t(0) = DX_t$ are satisfied. According to Eq. (21), we can easily obtain that $\gamma_t(1) = \psi\gamma_t(0) - \omega\gamma_t(-1) = \psi DX_t - \omega\gamma_{t-1}(1)$. This directly leads to $\gamma_t(1) = (-\omega)^t\gamma_0(1) + \psi\sum_{i=0}^{t-1}(-\omega)^i DX_{t-i}$. Given $\lambda_{\gamma1}$, $\lambda_{\gamma2}$ as the two characteristic roots, then when t is fixed, for any $\tau \geqslant 2$, we can get

$$\gamma_t(\tau) = \begin{cases} [(\tau-1)\cdot u_1(t) + u_2(t)]\lambda_{\gamma1}^{\tau-1}, & \text{if } \lambda_{\gamma1} = \lambda_{\gamma2} \\ u_3(t)\lambda_{\gamma1}^{\tau} + u_4(t)\lambda_{\gamma2}^{\tau}, & \text{otherwise} \end{cases} \quad (23)$$

where $u_1(t) = \gamma_t(1) - \lambda_{\gamma1}\gamma_t(0)$, $u_2(t) = \gamma_t(1)$, $u_3(t) = \frac{\gamma_t(1) - \lambda_{\gamma2}\gamma_t(0)}{\lambda_{\gamma1} - \lambda_{\gamma2}}$, and $u_4(t) = \frac{\gamma_t(1) - \lambda_{\gamma1}\gamma_t(0)}{\lambda_{\gamma2} - \lambda_{\gamma1}}$.

The characteristic equation of auto-covariance sequence is the same as that of expectation sequence, thus given $\lambda_{\max\gamma} = \max\{|\lambda_{\gamma1}|, |\lambda_{\gamma2}|\}$, $\lambda_{maxD}<1$ also implies that $\lambda_{max\gamma}<1$. At the same time, the value of each $\gamma_t(\tau)$ depends on all values of $DX_T(1\leq T\leq t)$. Thus, if iteration sequence $\{DX_t\}$ can converge, then the auto-covariance sequence can also converge. For any fixed t, when $\tau\to\infty$, $\gamma_t(\tau)$ is guaranteed to converge to 0. And when $t\to\infty$, the convergent values are

$$\gamma(0) = DX$$
$$\gamma(1) = \frac{\psi}{1+\omega} DX \quad (24)$$
$$\gamma(\tau) = \begin{cases} [(\tau-1)\cdot v_1 + v_2]\lambda_{\gamma1}^{\tau-1}, & \text{if } \lambda_{\gamma1} = \lambda_{\gamma2} \\ v_3\lambda_{\gamma1}^{\tau} + v_4\lambda_{\gamma2}^{\tau}, & \text{otherwise} \end{cases}$$

where $v_1 = \gamma(1) - \lambda_{\gamma1}\gamma(0)$, $v_2 = \gamma(1)$, $v_3 = \frac{\gamma(1) - \lambda_{\gamma2}\gamma(0)}{\lambda_{\gamma1} - \lambda_{\gamma2}}$, and $v_4 = \frac{\gamma(1) - \lambda_{\gamma1}\gamma(0)}{\lambda_{\gamma2} - \lambda_{\gamma1}}$.

VI Discussions

In preceding sections, we studied the simplified one-dimensional one-particle swarm system, which is in stagnation. It is easy to extend the result to the original D-dimensional M-particle swarm system because each particle evolves independently and different dimensions are independent. From the analytical results derived above, it is clear that when the particle swarm system is in stagnation, at any time step t, the trajectory of any arbitrary particle follows a stochastic distribution with expected value EX_t and variance DX_t, which are given in Eqs. (11) and (19) separately. The auto-covariance of the trajectory is also a time-variable function, given in Eq. (23). All these rules remain valid until a better position $\vec{X}_i$ (and thus $\vec{P}_i$ and $\vec{P}_g$) is discovered. When a better position is discovered, the new trajectory can be determined again after recalculating corresponding coefficients.

Holland discussed the balance between *exploration* and *exploitation* that an algorithm must maintain[11]. Exploration ability is related to the algorithm's tendency to explore new regions of the search space, while exploitation is the tendency to search a smaller region more thoroughly. Researchers in PSO community used to believe that inertia weight balances exploration and exploitation in PSO algorithm. But the theoretical results derived in this paper give a different new explanation. In our opinion, inertia weight can not balance exploration and exploitation by itself in PSO algorithm. The factor to balance exploration and exploitation should be the value of λ_{maxD}. The larger λ_{maxD} is, the stronger is the exploration ability of the PSO algorithm. An empirical evidence can be found in [12], which shows that the relationship between explore ability of PSO algorithm and ω is not monotone. When

$c_1=c_2=1.49$, the relationship between λ_{maxD} and ω is shown in Fig. 3, which is well fitted with the experimental result given in [12] if only non-negative ω is considered.

If given $c_1=c_2=2$, which is the case in original PSO algorithm, the relationship between λ_{maxD} and ω is also shown in Fig. 3. It can be clearly seen that the minimum λ_{maxD} appears near $\omega=0.4$, that is just why Shi and Eberhart[13] found that ω decrease linearly from 0.9 to 0.4 is the best choice, and not any other value smaller than 0.4.

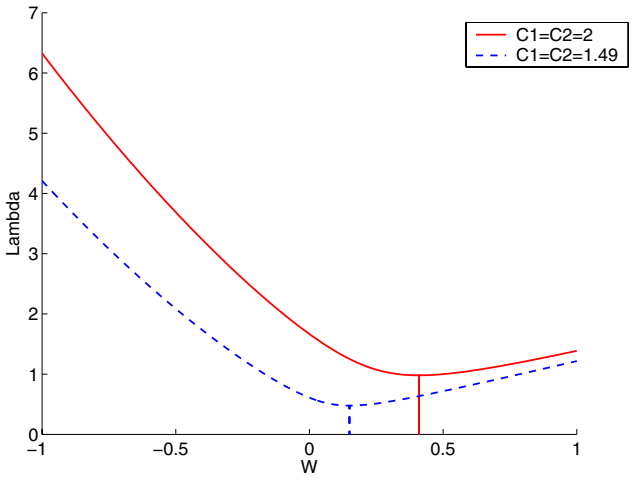

Figure 3. Relationship between λ_{maxD} and ω.

VII Conclusions

This paper studies a particle swarm system which is in stagnation. By regarding each particle's position as a stochastic vector, the stochastic process theory is applied to analyze the standard PSO algorithm determined by parameter tuple $\{\omega, c_1, c_2\}$, and the randomness in PSO is considered thoroughly. The analytical results give important characteristics (including expected value, variance, and auto-covariance) of the particle trajectory, both in explicit and implicit representations. This result is helpful to understand the mechanism of real stochastic PSO algorithm.

This paper also studies the convergent properties of expected value, variance, auto-covariance of the particle trajectories. The guaranteed convergent conditions are derived, and if the sequences can converge, corresponding convergent values are also given.

Some more factors may influence the particle trajectories of particle swarm system, but they are not studied in this paper. Those factors are needed to be considered in future, including velocity clamping, position clamping, different kinds of topology, etc. If those factors are incorporated, the derived results given in this paper may need to be modified. Furthermore, this study focuses on the particle swarm system in stagnation, which does not consider the interactions between particles. We know that the interactions among particles in PSO algorithm is quite an important factor that would influence the performance of PSO algorithm, thus this factor should be incorporated in future.

References

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

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

[3] E. Ozcan and C.K. Mohan, Analysis of a simple particle swarm optimization system, Intelligent Engineering Systems Through Artificial Neural Networks, 1998, pp. 253-258.

[4] E. Ozcan and 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.

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

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

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

[8] M. Clerc. Stagnation analysis in particle swarm optimisation or what happens when nothing happens. Technical Report CSM-460, Department of Computer Science, University of Essex. Aug. 2006. [Online]. Available: http://cswww.essex.ac.uk/technical-reports/2006/csm-460.pdf

[9] M. Jiang, Y.P. Luo, and S.Y. Yang, Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm, Information Processing Letters, in press.

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

[11] J. Holland. Adaption in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI, 1975.

[12] Xiao-Feng Xie, Wen-Jun Zhang, De-Chun Bi. Optimizing semiconductor devices by self-organizing particle swarm. Congress on Evolutionary Computation (CEC), Oregon, USA, 2004: 2017-2022.

[13] Y. Shi and R. C Eberhart. Parameter selection in particle swarm optimization. Evolutionary Programming VII: Proceedings of the Seventh Annual Conference on Evolutionary Programming, New York. 1998. 591-600