Analysis of Particle Swarm Optimization Convergence Time

Chao-Hong Chen

NCLab Report No. NCL-TR-2009007
July 2009

Natural Computing Laboratory (NCLab)
Department of Computer Science
National Chiao Tung University
329 Engineering Building C
1001 Ta Hsueh Road
HsinChu City 300, TAIWAN
http://nclab.tw/

國立交通大學

資訊科學與工程研究所

碩 士 論 文

粒 子 群 最 佳 化 之 收 斂 時 間 分 析

Analysis of Particle Swarm Optimization Convergence Time

研 究 生：陳昭宏

指導教授：陳穎平　教授

中 華 民 國 九 十 八 年 六 月

粒 子 群 最 佳 化 之 收 斂 時 間 分 析

Analysis of Particle Swarm Optimization Convergence Time

研 究 生：陳昭宏　　Student：Chao-Hong Chen

指導教授：陳穎平　　Advisor：Ying-Ping Chen

國 立 交 通 大 學

資 訊 科 學 與 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

粒子群最佳化之收斂時間分析

學生：陳昭宏　　　　　　　　　　　　　　　　　　指導教授：陳穎平

國立交通大學資訊科學與工程研究所碩士班

摘　　　要

在本論文中，我們分析了粒子群最佳化中粒子交互作用的收斂時間。我們提出了能夠描述粒子交互作用的統計模型，利用這個模型我們得到了關於收斂時間的理論結果。在進行了理論分析之後，經由執行粒子群最佳化在某些測試函式上，我們使用實驗來驗證我們推導出的結果。

Analysis of Particle Swarm Optimization Convergence Time

Student：Chao-Hong Chen Advisor：Dr. Ying-Ping Chen

Institutes of Computer Science and Engineering
National Chiao Tung University

ABSTRACT

In this thesis we analyze the convergence time of particle swarm optimization (PSO) on the facet of particle interaction. We propose a statistical model of PSO which captures the behavior of PSO particle interaction, and we use it to obtain results about convergence time. After the theoretical analysis we use experiments to verify our results by running real PSO on benchmark functions.

誌　　謝

感謝所有在這兩年中幫助過我的人，因為有你們這篇論文才得以完成。感謝所有看這篇論文的人，因為有你們這篇論文才有存在的意義。

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Particle swarm optimizer (PSO), introduced by James Kennedy and Russell C. Eberhart in 1995[1, 2], is a stochastic, population-based algorithm for solving optimization problem. It is kind of a swarm intelligence[3, 4, 5] based on social-psychological principles. In PSO, each potential solution is considered as a particle flying through the n-dimensional search space searching for the optimal solution, the velocity of each particle is determined by the inertia, personal experience and particle interaction.

As [6] shows, PSO is an efficient optimization framework. Although PSO has been widely applied in many fields[7, 8, 9, 10], the understanding of PSO in theoretical aspect is still limited. Most of previous theoretical results [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21] are derived under the system that assumes fixed attractor or single particle, which do not take particle interaction into consideration. Due to the lack of theoretical analysis of PSO particle interaction which is one of the most mechanism of PSO, we will propose a simplified model of PSO with non-fixed attractors. And the behavior of PSO will be analyzed under this model.

1.2 Thesis objectives

Due to the lack of theoretical analysis of PSO particle interaction, this thesis primarily aims to provide a theoretical analysis of the behavior particle interaction. There are three objectives in this thesis:

- To analyze the behavior of particle interaction, we will propose a model of PSO which captures the behavior of particle interaction.

- Using the proposed model, the convergence of PSO will be analyzed.
- The experimental results will be obtained to examine our model and results.

1.3 Road map

The thesis is composed by five chapters, the organization is given as follows:

- Chapter 1 consists of motivation, objective and organization of this thesis.
- Chapter 2 gives a detail description of standard PSO algorithm and survey of previous theoretical works of PSO.
- Chapter 3 describes the proposed statistical model of PSO and the analysis of PSO convergence time.
- Chapter 4 provides experimental results obtained by running real PSO, and use these results to verify the results obtained in chapter 3.
- Chapter 5 summaries and concludes this thesis, and some future objectives are provided.

Chapter 2

Background

In this chapter, we first give a detail description of PSO algorithm, and then we survey previous theoretical works.

2.1 Particle swarm optimization

In this section, we will describe the framework of PSO in detail, the algorithm of standard PSO is described in figure 2.1.

procedure PSO(Objective function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$)
    Initialize m particles
    while the stopping criterion is not satisfied **do**
        for $i = 1, 2, \ldots, m$ **do**
            if $\mathcal{F}(\mathbf{X_i}) < \mathcal{F}(\mathbf{Pb_i})$ **then**
                $\mathbf{Pb_i} \leftarrow \mathbf{X_i}$
                if $\mathcal{F}(\mathbf{Pb_i}) < \mathcal{F}(\mathbf{Nb})$ **then**
                    $\mathbf{Nb} \leftarrow \mathbf{Pb_i}$
                end if
            end if
        end for
        for $i = 1, 2, \ldots, m$ **do**
            $\mathbf{V_i} \leftarrow w\mathbf{V_i} + C_p(\mathbf{Pb_i} - \mathbf{X_i}) + C_n(\mathbf{Nb} - \mathbf{X_i})$
            $\mathbf{X_i} \leftarrow \mathbf{X_i} + \mathbf{V_i}$
        end for
    end while
end procedure

Figure 2.1: Standard PSO

Note, in this thesis we will use boldface to represent vectors, ex. $\mathbf{X_i}$, $\mathbf{V_i}$. And without loss of generality we assume that the objective of PSO is to minimize the objective function.

According to figure 2.1, at first PSO will initialize the m particles where m is swarm size which is a parameter of PSO algorithm, each particle contains three informations: location($\mathbf{X_i}$), velocity($\mathbf{V_i}$) and personal best position($\mathbf{Pb_i}$). In each generation, each particle will update the personal best position($\mathbf{Pb_i}$) and neighborhood best position($\mathbf{Nb}$) according to the objective function's return value. $\mathbf{Nb}$ can be updated by all members within the neighborhood, but $\mathbf{Pb_i}$ can only be updated by particle itself. After the update of personal and neighborhood best position, each particle will update their velocity according to $\mathbf{Pb_i}$ and $\mathbf{Nb}$, there are some parameters in the velocity update formula: w is the weight of inertia which is a constant, C_p and C_n are random variable sampled from uniform distribution $U(0, c_p)$ and $U(0, c_n)$ where c_p and c_n are acceleration coefficients of PSO. Finally, each particle will update it's position according to the velocity and then go to next generation.

2.2 Literature review

In this section, we will review previous theoretical works of PSO. The model used in these theoretical works all assume global structure which means all particles are within the same neighborhood, every particle update the same $\mathbf{Nb}$, in here we denote this global attractor by $\mathbf{Gb}$.

To the best we know, the first analysis is [11] proposed by Kennedy, in this work some simplification are used:

$$\mathbf{Gb} = \mathbf{Pb_i} = p \text{ where } p \text{ is a constant}, m = 1, w = 1, n = 1$$

the update rules thus become

$$V \leftarrow V + \varphi(p - x)$$
$$x \leftarrow x + V$$

where $\varphi = C_p + C_n$, in this work Kennedy further assumed that $\varphi = c_p + c_n$ is a constant, the particle trajectories under different φ are plotted. The particle trajectories are sinusoidal waves, for $\varphi = 4.0$ the trajectory explodes linearly and for $\varphi > 4.0$ the explosion becomes exponential, this suggests user to set c_p and c_n smaller than 2.0. Another behavior mentions in this study is "drunkard's walk", since the initial velocity, location and the fixed attractor p scale the amplitude of the system, when a trajectory is headed away from p and approaching the extreme of its cycle, the effect of decreasing V as a function of increasing $(p - x)$ eventually causes the particle to slow down and turn around. If $(p-x)$ is weighted by a very small(or, in some case, very large) random number, then the movement toward the extreme will continue farther than it would have which results in a larger value of x, as the particle is hurled outside its previous bound and v is a function of of previous velocity and $(p-x)$, the further x travels from p the greater will v be. This will cause the cycles approach the new limit and then inevitably exceed it again, a fix proposed in this study is set the limit of velocity V_{max}.

Shi and Eberhart[22] analyzed the impact of the inertia weight(w) and maximum velocity(V_{max}), they conducted a number of experiments to find good settings of w and V_{max}. When we lack of knowledge regarding the selection of V_{max}, this study suggests that set $V_{max} = X_{max}$ and inertia weight $w = 0.8$ is a good starting point. Furthermore if a time varying inertia weight is employed , even better performance can be expected.

Using the same simplified system in [11], Ozcan and Mohan[12] analyzed the particle trajectories by solving the equations:

$$V(t+1) = V(t) + \varphi(p - x(t-1))$$

$$x(t+1) = x(t-1) + v(t)$$

the conclusion of this study is there are four search types according to the value of φ. Using type 1($\varphi = 0$ or $\varphi \geq 4$), search is conducted by increasing step sizes in the space. For other types, a particle follows a path on a sinusoidal wave searching a space bounded by the amplitude of the sine wave. In type 3($2-\sqrt{3} < \varphi \leq 2+\sqrt{3}$) the amplitude of the sine wave is approximately the initial velocity, and in type 2($0 < \varphi \leq 2-\sqrt{3}$) and type 4($2+\sqrt{3} < \varphi < 4$) the amplitude decrease and increase respectively as φ increase. The simulations in [11] correspond exactly to the analytical results in this work. Later Ozcan and Mohan[13] generalized their results to a multiple multi-dimensional particles($m \geq 1, n \geq 1$) with $\mathbf{Gb} = p_n, \mathbf{Pb_i} = p_i$ where p_n and p_i are not necessary the same, similar results were obtained. And due to the increasing step sizes, the setting of velocity limit(V_{max}) is suggested in both [12, 13].

With the same simplified system in [13], Clerc and Kennedy[14] defined a simplified dynamic system:

$$v(t+1) = v(t) + \varphi y(t)$$
$$y(t+1) = -v(t) + (1-\varphi)y(t)$$

where $y(t) = p - x(t)$ and $p = \frac{c_p\mathbf{Pb_i}+c_n\mathbf{Gb_i}}{c_p+c_n}$ the weight average of the two best which is assumed to be a constant. Let

$$P_t = \begin{bmatrix} v(t) \\ y(t) \end{bmatrix}$$

be the current point in $\mathbb{R}^2$ and

$$M = \begin{bmatrix} 1 & \varphi \\ -1 & 1-\varphi \end{bmatrix}$$

the matrix of the system, we have $P_t = M^t P_0$. The eigenvalues of M are

$$e_1 = 1 - \frac{\varphi}{2} + \frac{\sqrt{\varphi^2 - 4\varphi}}{2}$$
$$e_2 = 1 - \frac{\varphi}{2} - \frac{\sqrt{\varphi^2 - 4\varphi}}{2}.$$

The behavior of this system is governed by the two eigenvalue e_1,e_2 of this system. The system converges when $e_1 < 1$ and $e_2 < 1$. Another contribution of this work is the introduction of the constriction coefficient and different classes of constriction models. The objective of this theoretically derived constriction coefficient is to prevent the velocity to grow out of bounds, with the advantage that, theoretically, velocity clamping(V_{max}) is no longer required. According to this study, the velocity equation changes to

$$V(t+1) = \chi\left[V(t) + c_p(\mathbf{Pb}_i - x(t)) + c_n(\mathbf{Gb}_i - x(t))\right]$$

where χ is the constriction coefficient calculated as

$$\chi = \frac{2\kappa}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}$$

with $\varphi \geq 4$ and $\kappa \in [0, 1]$. The constant κ controls the speed of convergence, for $\kappa \approx 0$ fast convergence to a stable point is obtained, while $\kappa \approx 1$ results in slow convergence.

Trelea[18] analyzed the stability under the similar dynamic system in [14], the update rules in this work are

$$V(t+1) = aV(t) + C_p(\mathbf{Pb}_i - x(t)) + C_n(\mathbf{Gb}_i - x(t))$$
$$x(t+1) = cx(t) + dV(t+1).$$

To simplify the analysis, the acceleration coefficients are set to their expected value. Thus, the update rules can be expressed as

$$V(t+1) = aV(t) + b(p - x(t))$$
$$x(t+1) = cx(t) + dV(t+1),$$

where

$$b = \frac{c_p + c_n}{2}$$
$$p = \frac{c_p}{c_p + c_n}\mathbf{Pb}_i + \frac{c_n}{c_p + c_n}\mathbf{Gb}_i.$$

The dynamic system is defined as follows:

$$y(t+1) = Ay(k) + Bp$$

with

$$y(t) = \begin{bmatrix} x(t) \\ V(t) \end{bmatrix}, A = \begin{bmatrix} 1-b & a \\ -b & a \end{bmatrix}, B = \begin{bmatrix} b \\ b \end{bmatrix}.$$

The eigenvalues of the system λ_1 and λ_2 are the solution of the equation:

$$\lambda^2 - (a - b + 1)\lambda + a = 0.$$

The system will converge if the parameters satisfy following condition:

$$a < 1, b > 0, 2a - b + 2 > 0.$$

which is a triangle in the (a, b) plane. Trelea also conducts some experiments to examine the effect of the parameters a and b.

Keiichiro, et al.[16] also analyzed the particle trajectories using similar dynamic system in [14] but take inertia weight into consideration. The update equations in this work are

$$V(t+1) = wV(t) + \varphi y(t)$$
$$y(t+1) = -wV(t) + (1-\varphi)y(t)$$

where $y(t) = p - x(t)$. The system is defined as

$$\begin{bmatrix} V(t+1) \\ y(t+1) \end{bmatrix} = \begin{bmatrix} w & \varphi \\ -w & 1-\varphi \end{bmatrix} \begin{bmatrix} V(t) \\ y(t) \end{bmatrix} = M \begin{bmatrix} V(t) \\ y(t) \end{bmatrix}$$

where M is the matrix of the system. The eigenvalues of M are

$$\lambda_1 = \frac{w+1-\varphi+\sqrt{(w+1-\varphi)^2-4w}}{2}$$
$$\lambda_2 = \frac{w+1-\varphi-\sqrt{(w+1-\varphi)^2-4w}}{2}.$$

According to the stability theory, the behavior of the particle is stable if and only if $\lambda_1 < 1$ and $\lambda_2 < 1$. Since the value of eigenvalues are function of w and φ, there are four cases:

- $w = 0$: The system is stable when $0 < \varphi < 2$, and each particle converges to p.
- $\varphi < w + 1 - 2\sqrt{w}$: The system is stable when $0 \leq w < 1$, and each particle converges to p.
- $w + 1 - 2\sqrt{w} < \varphi < w + 1 + 2\sqrt{w}$: The system is stable when $0 \leq w < 1$, and each particle converges to p.
- $w + 1 + 2\sqrt{w} < \varphi$: Under the condition $w + 1 + 2\sqrt{w} < \varphi < 2w + 2$, the system is stable when $0 \leq w < 1$, and each particle converges to p.

In the third case, φ can be expressed as $\varphi = w + 1 - 2\sqrt{w} + 4\kappa_1\sqrt{w}$, and let $\kappa_2 = \frac{c_p}{c_p+c_n}$. The velocity update rule become

$$V(t+1) = wV(t) + 2(1-\kappa_2)\varphi(\mathbf{Pb_i} - x) + 2\kappa_2\varphi(\mathbf{Gb_i} - x).$$

The relation between the particle trajectories of the system and the parameters is summarized as follows:

- w : If $0 \leq w < 1$, the convergence tendency of the system strengthens as w become small. If $1 < w$, the divergence tendency strengthens as w becomes large.

- κ_1 : $0 < \kappa_1 < 1$, the system becomes vibrational as κ_1 becomes large.

- κ_2 : $0 < \kappa_2 < 1$, p is put in the neighborhood of $\mathbf{Pb_i}$ if $\kappa_2 \approx 0$, and p is put in the neighborhood of $\mathbf{Gb}$ if $\kappa_2 \approx 1$.

Keiichiro also examined the relationship between particle trajectories of the system and the parameters w, κ_1 and κ_2 through experiments.

Zheng[17] analyzed the same dynamic system in [16], let $\Delta = (\varphi - 1)^2 - 2(\varphi - 1) + w^2$, the conclusions of the analysis are summarized as follows:

- $w = 1$: The particle trajectories are sinusoidal waves which are consist with [12, 13].

- $\Delta < 0$: The particle trajectories converge.

- $\Delta > 0$: The particle trajectories diverge.

Since a large inertia weight facilitates a global search while a small inertia weight facilitates a local search, a novel PSO is proposed in this work:

$$\mathbf{V_i} \leftarrow wV + \varphi_1(\mathbf{Pb_i} - \mathbf{X_i}) + \varphi_2(\mathbf{Gb} - \mathbf{X_i})$$
$$\mathbf{X_i} \leftarrow \mathbf{X_i} + \mathbf{V_i}$$

where $\varphi_i = b_i r_i + d_i$, $b_i = 1.5$, $r_i \sim U(0, 1)$ and $d_i = 0.5$ for $i = 1, 2$. w is an inertia weight linearly increasing from 0.4 to 0.9. In the experiment section of this study, the novel PSO outperform the standard PSO.

All works above assume the acceleration coefficients as constants, Kadirkamanathan, et al.[20] analyzed the system similar in [14] which take stochastic acceleration coefficients and inertia weight in to consideration. The update rules in this work are:

$$V(t+1) = wV(t) + \alpha_t^{(l)}(\mathbf{Pb_i} - x(t)) + \alpha_t^{(g)}(\mathbf{Gb} - x(t))$$
$$x(t+1) = x(t) + V(t+1)$$

where $\alpha_t^{(l)} \sim U(0, c_p)$ and $\alpha_t^{(g)} \sim U(0, c_n)$. Let $\alpha_t = \alpha_t^{(l)} + \alpha_t^{(g)}$ and

$$p = \frac{\alpha_t^{(l)}\mathbf{Pb_i} + \alpha_t^{(g)}\mathbf{Gb}}{\alpha_t},$$

we have

$$V(t+1) = wV(t) + \alpha_t(p - x(t))$$
$$x(t+1) = x(t) + V(t+1).$$

The state-space representation of the PSO system is given by

$$\xi(t+1) = A\xi(t) + Bu_t$$
$$y_t = C\xi_t$$

where

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ V(t) \end{bmatrix}, u_t = -\alpha_t(y_t - p), \xi(t) = \begin{bmatrix} x(t) - p \\ V(t) \end{bmatrix},$$
$$A = \begin{bmatrix} 1 & w \\ 0 & w \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

The main result of this work is when $|w| < 1$ and $w \neq 0$, if

$$c_p + c_n < \left(\frac{2(1 - 2|w| + w^2)}{1 + w} \right)$$

the system is asymptotically stable.

Jiang, et al. [21] also analyzed the stochastic acceleration coefficients with similar system in [20], by calculating the expected value, $E[X(t)]$ they showed that given $w, c_p, c_n \geq 0$, the iterative process $E[X(t)]$ is guaranteed to converge to $(c_p\mathbf{Pb_i} + c_n\mathbf{Gb})/(c_p + c_n)$ if and only if $0 \leq w < 1$ and $0 < c_p + c_n < 4(1 + w)$. And by more advanced analysis of the iterative process $E[X(t)]$ they obtained some convergence criteria for parameters w, c_p and c_n.

A good survey about PSO particle trajectories analysis is provided in [15, 19]. As we can see, all the results mentioned above assumes fixed attractor, although in [20, 21] the attractor p is time variant, but the local best($\mathbf{Pb_i}$) and global best($\mathbf{Gb}$) are still constants. Since the particles interaction of PSO progresses through update of global best or neighborhood best, this assumption eliminate all effect of the particle interaction. Thus, the behavior of particle interaction is shown in none of these studies.

Chapter 3

Analysis of PSO convergence time

In this chapter, we first propose our statistical model of PSO, and then we use it to analyze the convergence time of PSO.

3.1 Statistical model of PSO

From the above section, we can see that particle interaction is an important part of PSO, but there are not too much literature discussed about this. Although there are some theoretical studies have discussed about particle interaction and the behavior of PSO, most of these studies are based on the assumption that the attractor is fixed. It seems to be inevitable to assume that the attractor is fixed, since each particle not only keeps its own experience(inertia and $\mathbf{Pb_i}$) but shares collective knowledge($\mathbf{Nb}$), any slight change in these quantities will result in a new state. The analysis of the overall behavior thus become intractable due to the complication of state transition, and fixed attractor thus becomes necessary simplification.

In order to take particle interaction into consideration, we use an alternative view of PSO that regards the swarm as a unity. Instead of consider the detail configuration of each particle in the swarm, we consider the overall behavior of the swarm. To do this, we convert the state of the entire swarm into a statistical abstraction. The exact locations of particles are not traced but modeled with a distribution θ over $\mathbb{R}^n$, and velocities are viewed as random vector $\mathcal{V} \in \mathbb{R}^n$. To concentrate on the particle interaction, we use the social-only model of PSO[23] which works without personal experience. The swarm size m is considered as the number of samples from distribution θ, since the geographic knowledge is embodied in the distribution, the neighborhood attractor can be viewed as the best of the m samples.

More specifically, each particle $\mathbf{P_i}$ is a random vector sampled from θ, and its velocity $\mathbf{V_i}$ is sampled from $\mathcal{V}$. The neighborhood attractor is defined as

$$\mathbf{P_a} := \min_{\mathbf{P_i}}\{\mathcal{F}(P_1), \mathcal{F}(P_2), \ldots, \mathcal{F}(P_m)\}.$$

At each generation particle $\mathbf{P_i}$ its position to $\mathbf{P_i}+w\mathbf{V_i}+C(\mathbf{P_a}-\mathbf{P_i})$, the distributions of next time step are thus the statistical characterization which can be denoted by functions of the observed value:

$$\theta \leftarrow \mathcal{T}_p(\mathbf{P_1}, \ldots, \mathbf{P_m}),$$
$$\mathcal{V} \leftarrow \mathcal{T}_v(\mathbf{P_1}, \ldots, \mathbf{P_m}, \mathbf{V_1}, \ldots, \mathbf{V_m}).$$

Since w is a constant the distribution $\mathcal{V}$ can be removed because given two random vectors $\mathbf{X} \sim \theta$ and $\mathbf{V} \sim \mathcal{V}$, we can simply let θ' be the distribution of $\mathbf{X}' := \mathbf{X} + w\mathbf{V}$.

Now the only question remain is which distribution is suitable for description of the swarm and how to update the distribution. For simplicity, we use product distribution in this thesis i.e. the position of each dimension is independently sampled from the distribution θ_i. Now consider the random variable $X \sim \theta_i$ and let $E[X] = \mu$. If we divide the support of θ_i into s disjoint regions $R_1, \ldots, R_s$ such that $Prob[X \in R_i] = 1/s$ for $i = 1, 2, \ldots, s$, and each region is associated with a random variable of velocity $V_i \sim \mathcal{V}_i$. Pick $x_i \in R_i$ for each region, when s is sufficiently large the swarm can be characterized by

$$\begin{aligned}\sum_{i=1}^{s} \frac{1}{s}(x_i + V_i) &= \sum_{i=1}^{s} \frac{x_i}{s} + \sum_{i=1}^{s} \frac{V_i}{s} \\ &\approx \mu + \sum_{i=1}^{s} \frac{V_i}{s},\end{aligned}$$

and each component of $\sum_{i=1}^{s} V_i/s$ can be approximated with a normal distribution by central limit theorem. Therefore normal distribution seems to be a reasonable choice.

We let the distribution of i-th dimension, θ_i be $N(\mu_i, \sigma_i^2)$ where $N(\mu_i, \sigma_i^2)$ is the normal distribution with mean μ_i and variance σ_i^2. Now the update of distribution becomes simply calculate the mean and the variance. The mean can be calculated intuitively by take the average of updated positions, and the variance is calculated by maximum likelihood estimation(MLE) which will be describe in next section. The statistical model of PSO is summarized in figure 3.1, the distribution θ is represented by $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, and the acceleration coefficient $C \sim U(0, c)$.

procedure STATISTICAL MODEL OF PSO(Objective function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$)
  Initialize:$\sigma \leftarrow \sigma_{\mathbf{0}}$,$\mu \leftarrow \mu_{\mathbf{0}}$
  while the stopping criterion is not satisfied **do**
    for $i = 1, 2, \ldots, m$ **do**
      for $j = 1, 2, \ldots, n$ **do**
        $\mathbf{P_i}_j \sim N(\mu_j, \sigma_j^2)$
      end for
    end for
    $\mathbf{P_a} = \min_{\mathbf{P_i}}\{\mathcal{F}(\mathbf{P_i})\}$
    for $i = 1, 2, \ldots, m$ **do**
      $\mathbf{P'_i} \leftarrow \mathbf{P_i} + C(\mathbf{P_a} - \mathbf{P_i})$
    end for
    $\mu \leftarrow (\sum_{i=1}^{m} \mathbf{P'_i})/m$
    $\sigma^2 \leftarrow \mathrm{MLE}(\mathbf{P'_1}, \mathbf{P'_2}, \ldots, \mathbf{P'_m})$
  end while
end procedure

Figure 3.1: Statistical model of PSO

3.2 Convergence time analysis

In this section we first describe the maximum likelihood estimation which used in the update of distribution, and then we obtain the results about PSO convergence time.

Let $\sigma_{\mathbf{t}}{}_i^2$ and $\mu_{\mathbf{t}}{}_i^2$ be the variance and mean of i-th dimension in t-th generation, let $y_j = \mathbf{P_j}_i$ and $y'_j = \mathbf{P'_j}_i$ for $j = 1, 2, \ldots, m$ and let $y_a = \mathbf{P_a}_i$, $\overline{y} = (1/m)\sum_{j=1}^{m} y'_j$. To estimate the variance of i-th dimension for $t+1$-th generation, we use maximum likelihood estimation (MLE), the likelihood function $L(\sigma_{\mathbf{t}}{}_i^2)$ is defined as the joint probability:

$$L(\sigma_{\mathbf{t}}{}_i^2) := \prod_{j=1}^{m}\left(\frac{1}{\sqrt{2\pi\sigma_{\mathbf{t+1}}{}_i^2}}\right)\exp\left(\frac{-(y'_j - \overline{y})^2}{2\sigma_{\mathbf{t+1}}{}_i^2}\right) = \left(\frac{1}{\sqrt{2\pi\sigma_{\mathbf{t+1}}{}_i^2}}\right)^m \exp\left(\frac{-\sum_{j=1}^{m}(y'_j - \overline{y})^2}{2\sigma_{\mathbf{t+1}}{}_i^2}\right).$$

To find ${\sigma_{\mathbf{t+1}}}_i^2$ that maximizes $L({\sigma_{\mathbf{t}}}_i^2)$, we differentiate $L({\sigma_{\mathbf{t}}}_i^2)$ with respect to ${\sigma_{\mathbf{t+1}}}_i^2$:

$$L'({\sigma_{\mathbf{t}}}_i^2) = -(\frac{m}{2})(\frac{1}{\sqrt{2\pi}})^m {\sigma_{\mathbf{t+1}}}_i^{-m-2} \cdot \exp(\frac{-\sum_{j=1}^m (y'_j - \overline{y})^2}{2{\sigma_{\mathbf{t+1}}}_i^2})$$
$$+(\frac{1}{\sqrt{2\pi}})^m \cdot \frac{\sum_{j=1}^m (y'_j - \overline{y})^2}{2} {\sigma_{\mathbf{t+1}}}_i^{-m-4} \cdot \exp(\frac{-\sum_{j=1}^m (y'_j - \overline{y})^2}{2{\sigma_{\mathbf{t+1}}}_i^2}),$$

the value of ${\sigma_{\mathbf{t+1}}}_i^2$ that maximizes $L({\sigma_{\mathbf{t}}}_i^2)$ is $\sum_{j=1}^m (y'_j - \overline{y})^2/m$, so in our model of PSO the results of MLE is ${\sigma_{\mathbf{t+1}}}_i^2 = \sum_{j=1}^m (y'_j - \overline{y})^2/m$ for $i = 1, 2, \ldots, n$.

After introduce MLE, we now discuss about the convergence time of PSO. There are many different definitions of convergence, in here we define the state of swarm is converge if all particles crowd into some specific location. Since in our viewpoint, the swarm is seen as a distribution, so the state of particles crowd into some specific location is refer to the variance of the distribution becomes very small. Here comes another problem, what's the definition of "very small"? Since under difference circumstances the definition of "very small" is different, so in here we do not give exact value, instead we define the variance is "very small" if the variance is less than some constant $\epsilon > 0$. So the state of convergence refer to the variance for every dimension is less than ϵ. Use this definition, now we start our analysis of convergence time.

To estimate the variance after distribution update, we need following lemma from [24]:

Lemma 1. *Let $X_1, X_2, \ldots, X_m \sim N(\mu, \sigma^2)$, define $S = \sum_{i=1}^m (X_i - \overline{X})^2/(m-1)$ where $\overline{X} = \sum_{i=1}^m X_i/m$. We have $(m-1)S \sim \sigma^2 \chi^2_{m-1}$ where χ^2_{m-1} is the chi-square distribution with $m-1$ degrees of freedom.*

Using this lemma we can obtain the following result:

Lemma 2. *Given the swarm size m, acceleration coefficient c and variance of i-th dimension at t-th generation ${\sigma_{\mathbf{t}}}_i^2$ we have $E[{\sigma_{\mathbf{t+1}}}_i^2] = [\frac{1}{3}c^2 - c + 1][(m-1)/m]{\sigma_{\mathbf{t}}}_i^2$.*

Proof. From above we know ${\sigma_{\mathbf{t+1}}}_i^2 = \sum_{j=1}^{m}(y'_j - \overline{y})^2/m$, the expected value is

$$\begin{aligned}
E[\frac{1}{m}\sum_{j=1}^{m}(y'_j - \frac{\sum_{k=1}^{m} y'_k}{m})^2] &= \frac{1}{m}E[\sum_{j=1}^{m}(y_j + C(y_a - y_j) - \frac{\sum_{k=1}^{m} y_k + C(y_a - y_k)}{m})^2] \\
&= \frac{1}{m}E[\sum_{j=1}^{m}(\frac{m(1-C)y_j - (1-C)\sum_{k=1}^{m} y_k}{m})^2] \\
&= \frac{1}{m^3}E[(1-C)^2\sum_{j=1}^{m}(my_j - \sum_{k=1}^{m} y_k)^2] \\
&= \frac{1}{m}E[(1-C)^2]E[\sum_{j=1}^{m}(y_j - \frac{1}{m}\sum_{k=1}^{m} y_k)^2],
\end{aligned}$$

Let $S = \sum_{j=1}^{m}(y_j - \frac{1}{m}\sum_{k=1}^{m} y_k)^2$, since $y_j \sim N({\mu_{\mathbf{t+1}}}_i, {\sigma_{\mathbf{t+1}}}_i^2)$ for $j = 1, 2, \ldots, m$ and $y_1, y_2, \ldots, y_m$ are i.i.d., by lemma 1 $S \sim {\sigma_{\mathbf{t}}}_i^2\chi^2_{m-1}$, thus $E[S] = (m-1){\sigma_{\mathbf{t}}}_i^2$. Using this, we can obtain

$$\begin{aligned}
E[{\sigma_{\mathbf{t+1}}}_i^2] &= \frac{1}{m}E[(1-C)^2]E[\sum_{j=1}^{m}(y_j - \frac{1}{m}\sum_{k=1}^{m} y_k)^2] \\
&= \frac{1}{m}E[1 - 2C + C^2]E[S] \\
&= \frac{1}{m}(\frac{1}{3}c^2 - c + 1)(m-1){\sigma_{\mathbf{t}}}_i^2 \\
&= (\frac{1}{3}c^2 - c + 1)\frac{m-1}{m}{\sigma_{\mathbf{t}}}_i^2
\end{aligned}$$

□

Lemma 2 is derived under the situation that ${\sigma_{\mathbf{t}}}_i^2$ is given, the following lemma will derive the relationship between $E[{\sigma_{\mathbf{t}}}_i^2]$ and $E[{\sigma_{\mathbf{t+1}}}_i^2]$.

Lemma 3. $E[{\sigma_{\mathbf{t+1}}}_i^2] = (\frac{1}{3}c^2 - c + 1)\frac{m-1}{m}E[{\sigma_{\mathbf{t}}}_i^2]$.

Proof.

$$\begin{aligned}
E[{\sigma_{\mathbf{t}}}_i^2] &= \int_{\sigma^2 \in \mathbb{R}^+} E[{\sigma_{\mathbf{t+1}}}_i^2 | {\sigma_{\mathbf{t}}}_i^2 = \sigma^2] Prob\{{\sigma_{\mathbf{t}}}_i^2 = \sigma^2\} d\sigma^2 \\
&= \int_{\sigma^2 \in \mathbb{R}^+} (\frac{1}{3}c^2 - c + 1)\frac{m-1}{m}\sigma^2 Prob\{{\sigma_{\mathbf{t}}}_i^2 = \sigma^2\} d\sigma^2 \qquad \text{(by lemma 2)} \\
&= (\frac{1}{3}c^2 - c + 1)\frac{m-1}{m}\int_{\sigma^2 \in \mathbb{R}^+} \sigma^2 Prob\{{\sigma_{\mathbf{t}}}_i^2 = \sigma^2\} d\sigma^2 \\
&= (\frac{1}{3}c^2 - c + 1)\frac{m-1}{m}E[{\sigma_{\mathbf{t}}}_i^2].
\end{aligned}$$

□

Using all the previous results, we can obtain the relationship of convergence time and parameters of PSO:

Theorem 4. *Given swarm size m, acceleration coefficient c, ϵ and $\sigma_{\mathbf{0}}$, let $h = \max_i\{{\sigma_{\mathbf{0}}}_i^2\}$ we have $E[{\sigma_{\mathbf{t}}}_i^2] < \epsilon$ for $i = 1, 2, \ldots, n$ when $[\frac{1}{3}c^2 - c + 1][(m-1)/m] < 1$ and $t > \log(\epsilon/{\sigma_0}_h^2)/\log([\frac{1}{3}c^2 - c + 1][(m-1)/m])$.*

Proof. From lemma 3 we know

$$\begin{aligned} E[{\sigma_{\mathbf{t}}}_i^2] &= (\frac{1}{3}c^2 - c + 1)^t(\frac{m-1}{m})^t E[{\sigma_{\mathbf{0}}}_i^2] \\ &= (\frac{1}{3}c^2 - c + 1)^t(\frac{m-1}{m})^t {\sigma_{\mathbf{0}}}_i^2 \end{aligned}$$

Since $[\frac{1}{3}c^2 - c + 1][(m-1)/m] < 1$ we have

$$\begin{aligned} E[{\sigma_{\mathbf{t}}}_i^2] &= (\frac{1}{3}c^2 - c + 1)^t(\frac{m-1}{m})^t {\sigma_{\mathbf{0}}}_i^2 \\ &< (\frac{1}{3}c^2 - c + 1)^t(\frac{m-1}{m})^t {\sigma_{\mathbf{0}}}_h^2 \\ &< \epsilon \end{aligned}$$

The last inequality is hold because

$$\begin{aligned} &\log\left((\frac{1}{3}c^2 - c + 1)^t(\frac{m-1}{m})^t {\sigma_{\mathbf{0}}}_h^2\right) \\ =&t\log\left((\frac{1}{3}c^2 - c + 1)(\frac{m-1}{m})\right) + \log {\sigma_{\mathbf{0}}}_h^2 \\ <&\log(\epsilon/{\sigma_0}_h^2) + \log {\sigma_{\mathbf{0}}}_h^2 \\ =&\log\epsilon \end{aligned}$$

□

We have two corollaries immediately from theorem 4:

Corollary 5. *Given swarm size m, acceleration coefficient c, and level of convergence ϵ such that $[\frac{1}{3}c^2 - c + 1][(m-1)/m] < 1$ and $\epsilon < 1$, we have $E[{\sigma_{\mathbf{t}}}_i^2] < \epsilon$ for $i = 1, 2, \ldots, n$ for $t = O(-\log\epsilon)$.*

Corollary 6. *Given swarm size m, c, and ϵ such that $[\frac{1}{3}c^2 - c + 1][(m-1)/m] < 1$ and $\epsilon < 1$, there exist a constant $c' < 1$ such that for $t = O(-1/\log c'(1-1/m))$ we have $E[\sigma_{\mathbf{t}_i}^2] < \epsilon$ for $i = 1, 2, \ldots, n$.*

Corollary 5 reveals the linear relationship between the level of convergence and convergence time, and the interpretation of corollary 6 is that when the swarm size is large enough, enlarge swarm size can only influences convergence time slightly. In next chapter, We will examine the above two corollaries.

Chapter 4

Experiments

In tis section, we examine corollary 5 and 6 by running standard PSO, we use two objective functions in our experiments:

- Sphere function [25]:

$$f_1(\mathbf{x}) = \sum_{i=1}^{D} x_i^2 \text{ where } \mathbf{x} \in [-100, 100]^D.$$

- Schwefel's problem 1.2 [25]:

$$f_2(\mathbf{x}) = \sum_{i=1}^{D} \left(\sum_{j=1}^{i} x_j^2 \right) \text{ where } \mathbf{x} \in [-100, 100]^D.$$

We set $D = 10$ for both $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ in the following experiments.

First, we examine the corollary 5, the parameters we used in PSO are given in the following: $c_p = 1$, $c_n = 1$, $w = 1/(2\ln 2)$ and swarm size = 50. The value of ϵ is from 10^{-1} to 10^{-10}, for each value of ϵ we perform 100 runs, for each run we count the number of generations from initialization to the state that all dimension's variance is smaller than ϵ, and then we calculate the mean number of generations from the 100 runs. The result of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are shown in figure 4.1 and 4.2 respectively, the x-axis represents the value of ϵ and the y-axis represent the mean number of generation.

The comparison of these experimental results and our theoretical results is shown in figure 4.3 and 4.4. From figure 4.3, we can see that the experimental results of $f_1(\mathbf{x})$ is very close to $-4.6 \log \epsilon + 43 = O(-\log \epsilon)$, and from figure 4.4, the experimental results of $f_2(\mathbf{x})$ are very close to $-4.7 \log \epsilon + 43.5 = O(-\log \epsilon)$. The experimental results are very

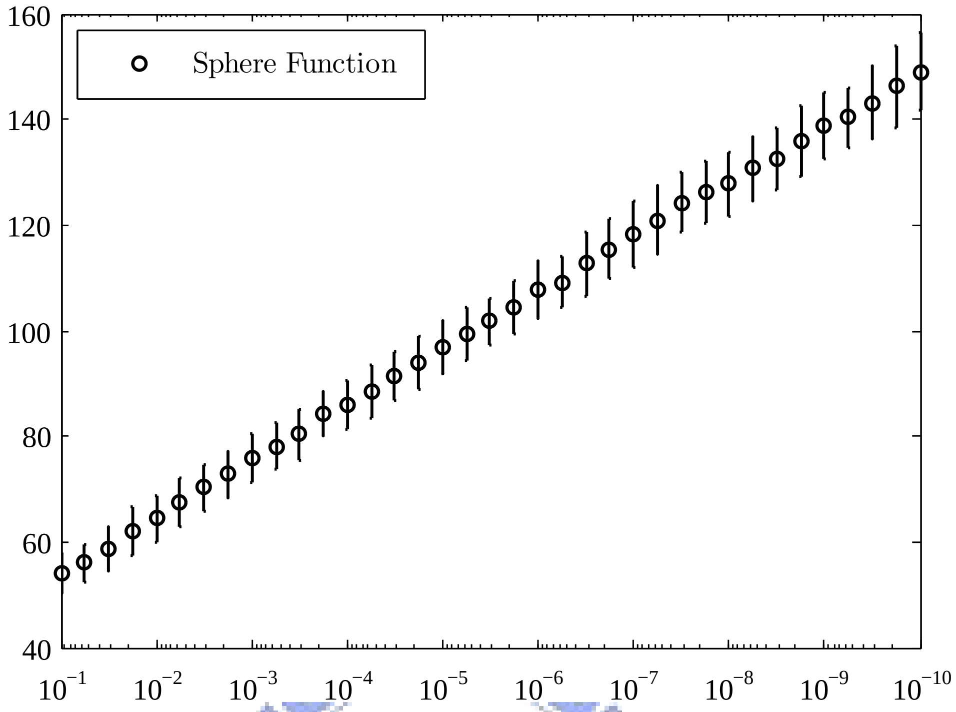

Figure 4.1: The experimental results of $f_1(\mathbf{x})$, the x-axis represents the value of ϵ and y-axis represents the mean number of generation.

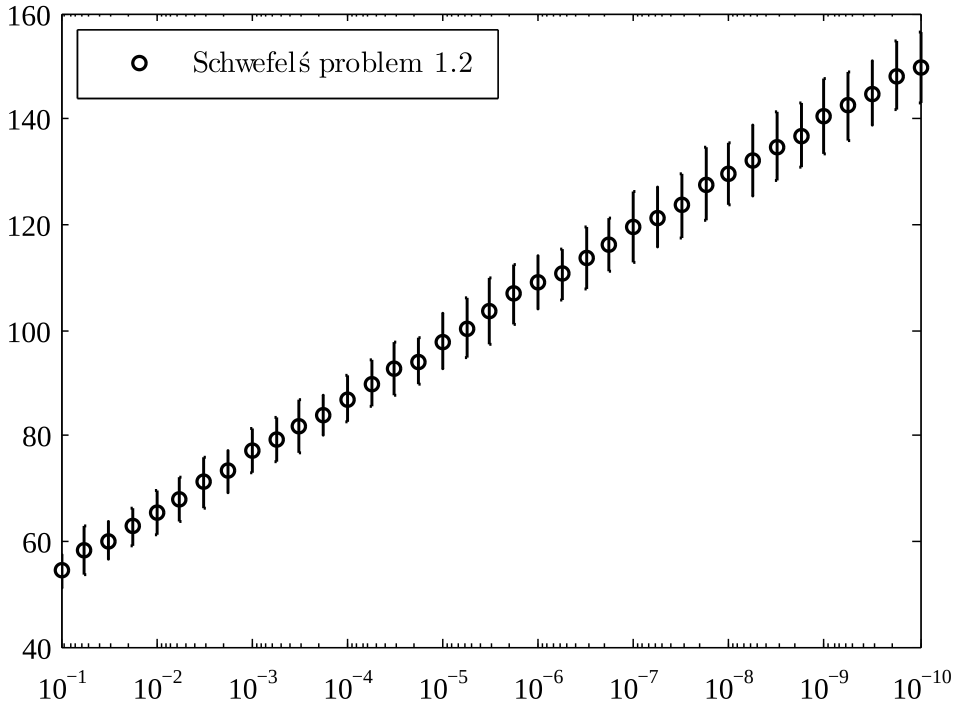

Figure 4.2: The experimental results of $f_2(\mathbf{x})$, the x-axis represents the value of ϵ and y-axis represents the mean number of generation.

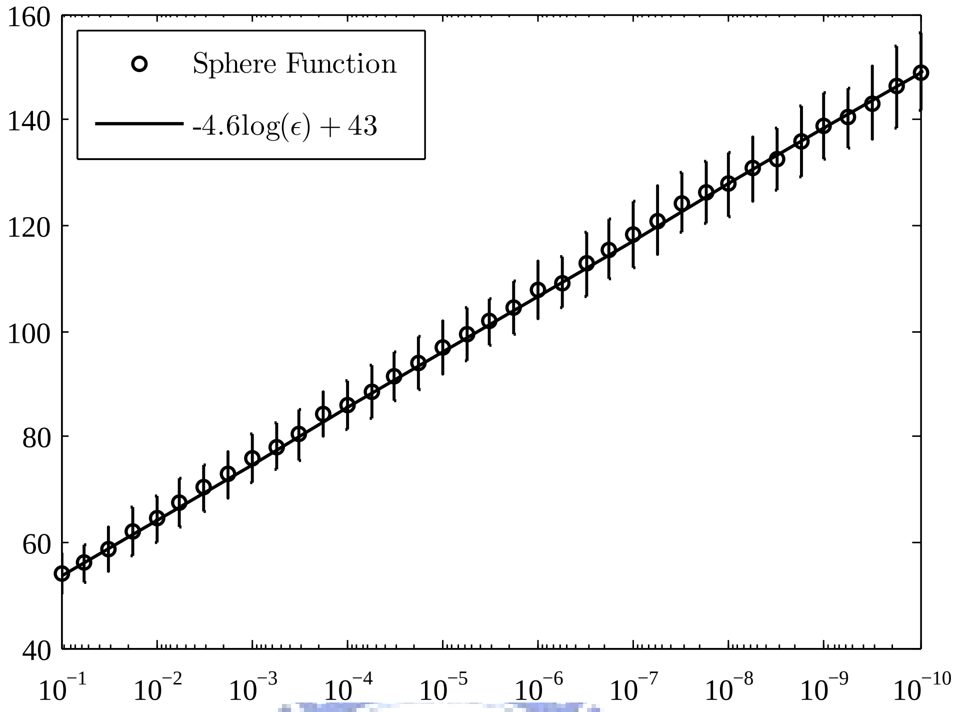

Figure 4.3: Comparison of experimental results and theoretical results from corollary 5 of $f_1(\mathbf{x})$, the x-axis represents the value of ϵ and y-axis represents the mean number of generation, the experimental results are very close to $O(-\log \epsilon)$.

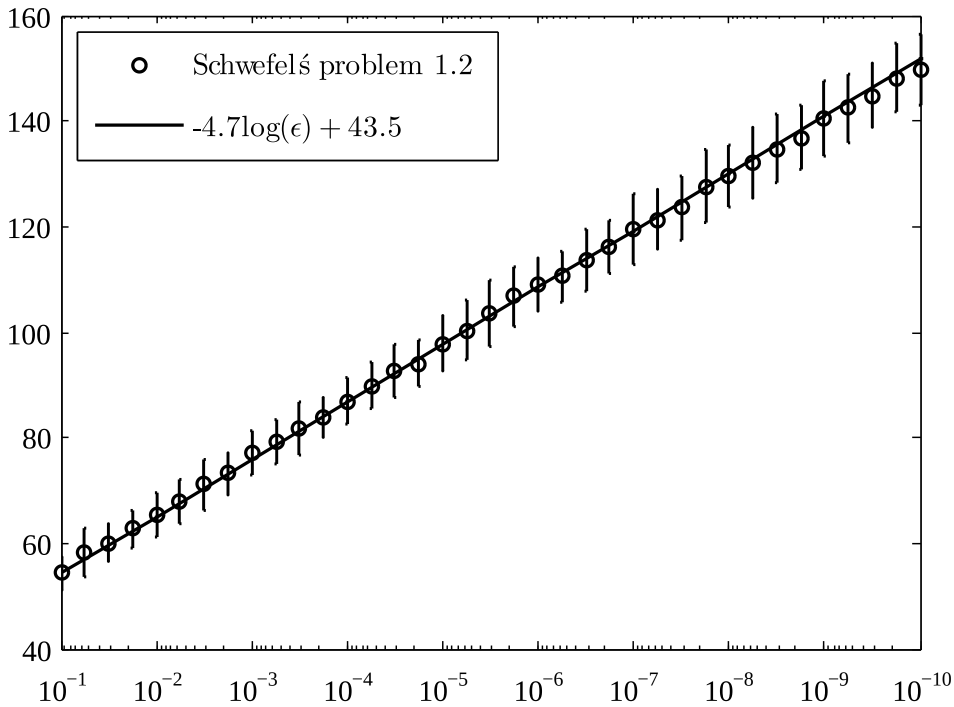

Figure 4.4: Comparison of experimental results and theoretical results from corollary 5 of $f_2(\mathbf{x})$, the x-axis represents the value of ϵ and y-axis represents the mean number of generation, the experimental results are very close to $O(-\log \epsilon)$.

close to our estimation in corollary 5, the value of $-\ln \epsilon$ and the convergence time are linearly related.

After examine corollary 5, now we examine corollary 6. The parameters we used in PSO are given in the following: $c_p = 1$, $c_n = 1$, $w = 1/(2\ln 2)$ and $\epsilon = 10^{-6}$, the swarm size if from 50 to 1000 with step 5, for each swarm size we perform 100 runs, and records the mean as before. The results are shown in figure 4.5, 4.6, 4.7 and 4.8, the x-axis represents the swarm size and y-axis represents the mean number of generation.

The comparison of experimental and theoretical results is shown in figure 4.9, 4.10,4.11 and 4.12. From figure 4.9 and 4.10, we can see that the convergence time is close to $-64.9/\log 0.555(1-1/m) = O(-1/\log c'(1-1/m))$ where $c' = 0.555$, and in figure 4.11 and 4.12 the convergence time is close to $-99.75/\log 0.405(1-1/m) = O(-1/\log c'(1-1/m))$ where $c' = 0.405$. As we can see from these figures, when the swarm size become large, the increase of convergence time is very small which conform with our estimation in corollary 6.

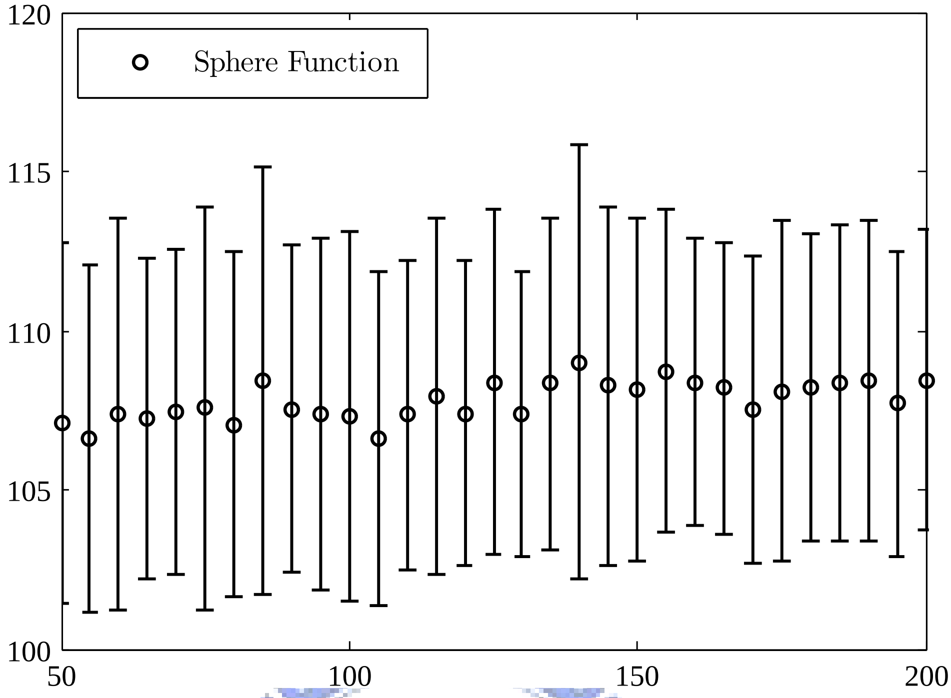

Figure 4.5: The experimental results of $f_1(\mathbf{x})$, x-axis represents the swarm size from 50 to 200 and y-axis represents the mean number of generation.

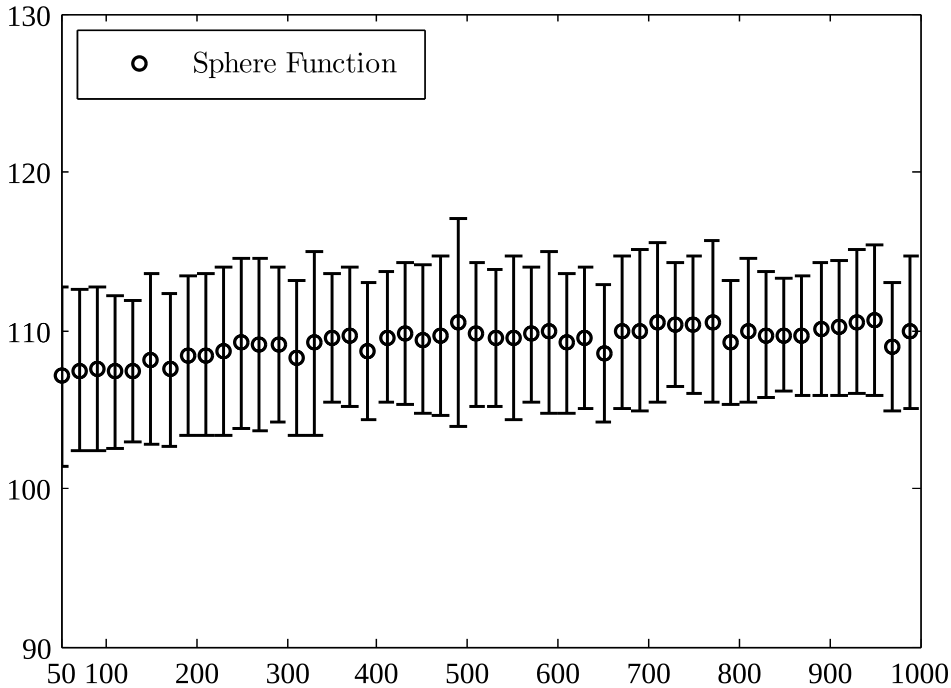

Figure 4.6: The experimental results of $f_1(\mathbf{x})$, x-axis represents the swarm size from 50 to 1000 and y-axis represents the mean number of generation.

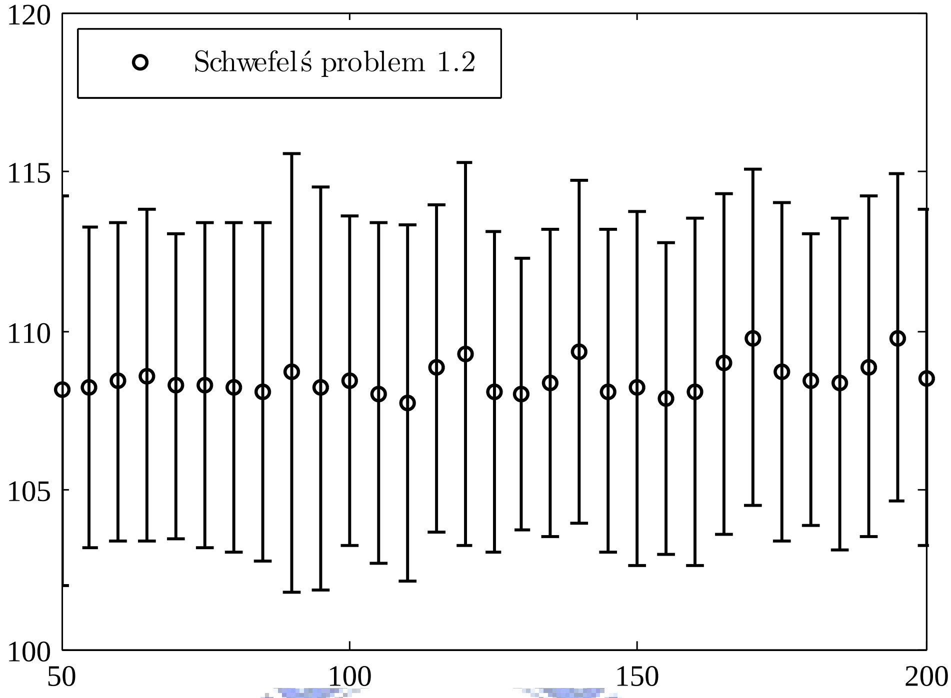

Figure 4.7: The experimental results of $f_2(\mathbf{x})$, x-axis represents the swarm size from 50 to 200 and y-axis represents the mean number of generation.

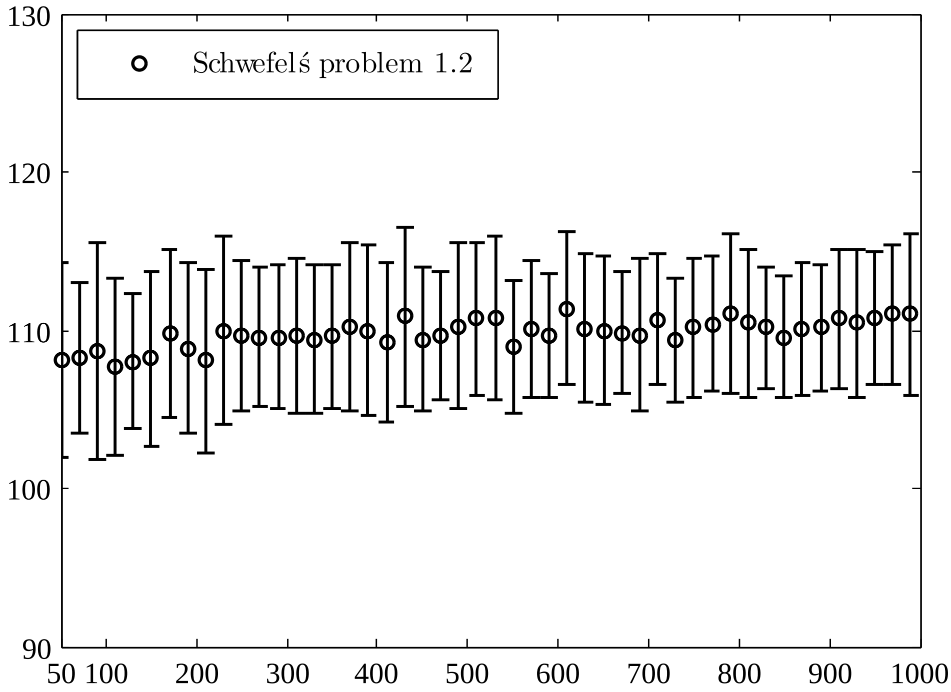

Figure 4.8: The experimental results of $f_2(\mathbf{x})$, x-axis represents the swarm size from 50 to 1000 and y-axis represents the mean number of generation.

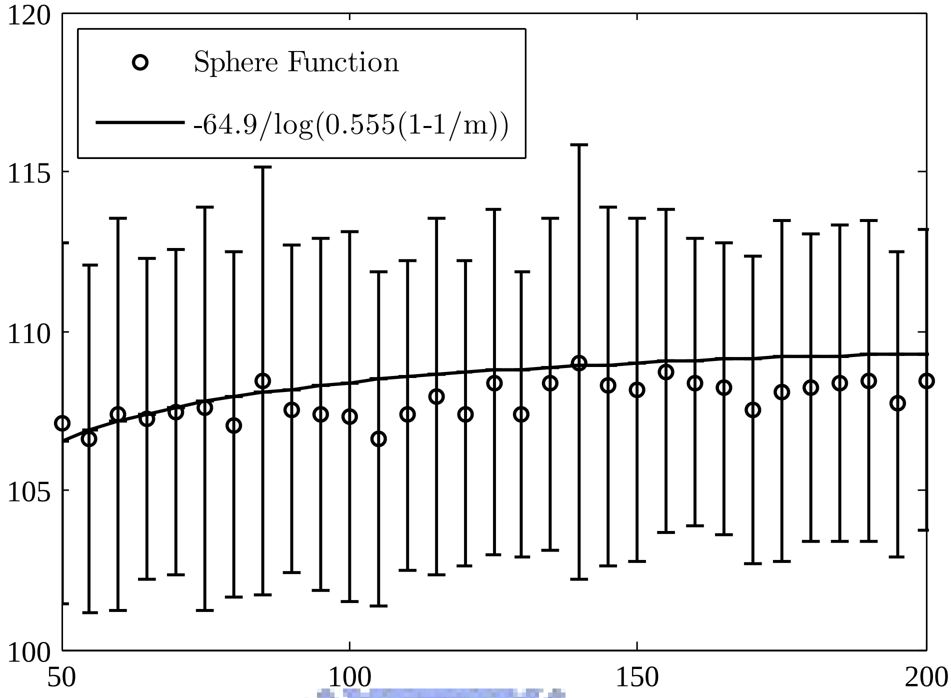

Figure 4.9: Comparison of experimental results and theoretical results from corollary 6 of $f_1(\mathbf{x})$, x-axis represents the swarm size from 50 to 200 and y-axis represents the mean number of generation, the experimental results are very close to $O(-1/\log c'(1-1/m))$ with $c' < 1$.

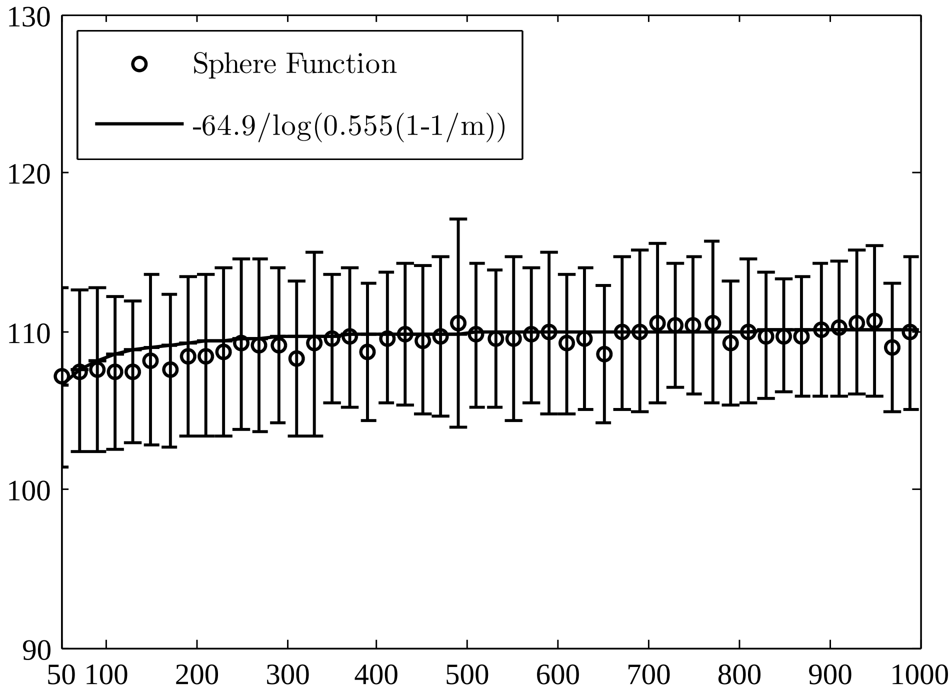

Figure 4.10: Comparison of experimental results and theoretical results from corollary 6 of $f_1(\mathbf{x})$, x-axis represents the swarm size from 50 to 1000 and y-axis represents the mean number of generation, the experimental results are very close to $O(-1/\log c'(1-1/m))$ with $c' < 1$.

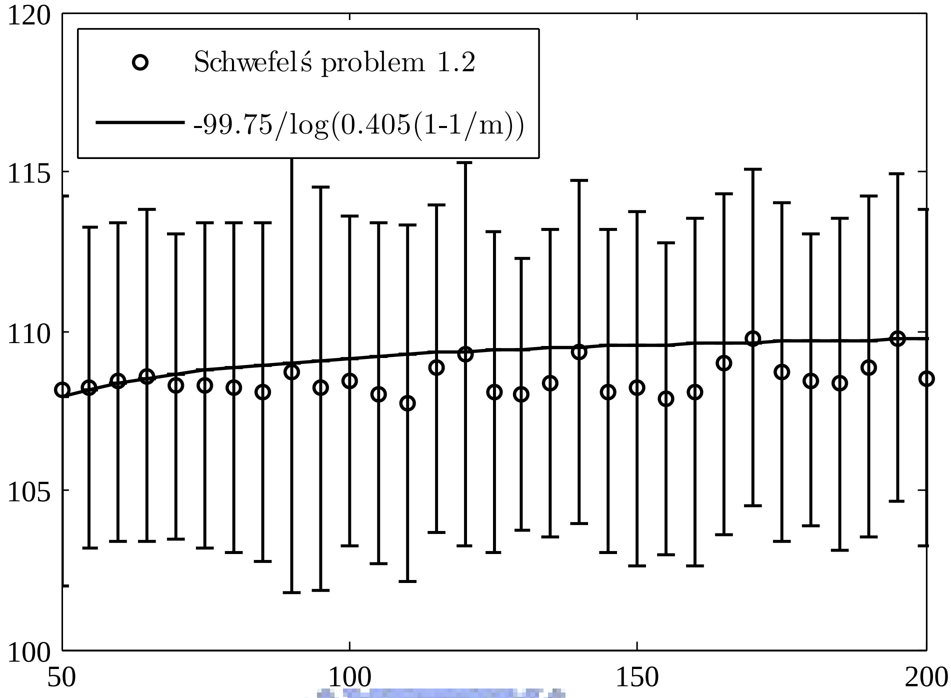

Figure 4.11: Comparison of experimental results and theoretical results from corollary 6 of $f_2(\mathbf{x})$, the x-axis represents the swarm size from 50 to 200 and y-axis represents the mean number of generation, the experimental results are very close to $O(-1/\log c'(1-1/m))$ with $c' < 1$.

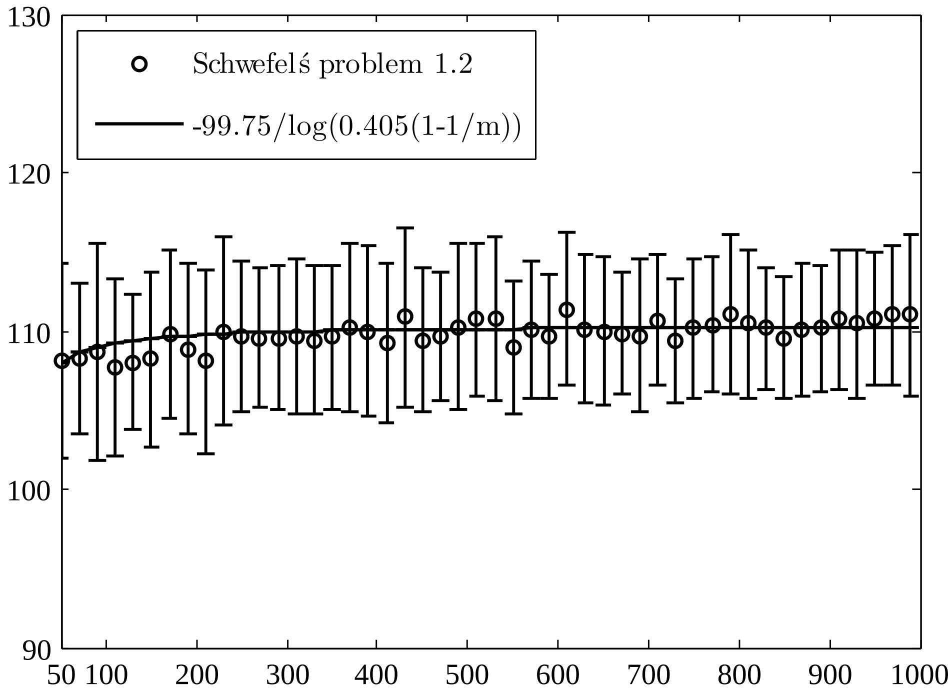

Figure 4.12: Comparison of experimental results and theoretical results from corollary 6 of $f_2(\mathbf{x})$, the x-axis represents the swarm size from 50 to 1000 and y-axis represents the mean number of generation, the experimental results are very close to $O(-1/\log c'(1-1/m))$ with $c' < 1$.

Chapter 5

Conclusions

5.1 Summary

In this thesis, a simplified statistical model of PSO which captures the behavior of particle interaction is proposed. And convergence time of PSO is analyzed, our results reveals the relationship between convergence time and the level of convergence, and relationship between convergence time and swarm size. The experimental results are obtained to verify our results, which shows that our estimation is very close to real PSO.

5.2 Main conclusions

To capture the behavior of particle interaction, the proposed model in this thesis does not assumes fixed attractor, so the effect of particle interaction can remain on our model. Although the dimensions are still treated independently like the other previous studies, unlike the previous work, we do not analyze the particle trajectories, since the whole swarm is seen as a distribution, we concentrate on the state of the swarm which is represented by a distribution. The criteria of convergence in our model is showed in this thesis, and the relationships between parameters and convergence time are analyzed.

5.3 Future work

There are still some future works remain. First, the relation between PSO and number of dimension, i.e. the relation between t and $E[{\sigma_{\mathbf{t}(n)}}^2]$ where ${\sigma_{\mathbf{t}(n)}}^2 = \max\{\sigma_{\mathbf{t}1}^2, \sigma_{\mathbf{t}2}^2, \ldots, \sigma_{\mathbf{t}n}^2\}$. Second, the results obtained in this thesis are independent of objective function, in chapter 4 we only use two different objective functions, more functions should be used to examine our estimation. Third, the distribution we used in this thesis is normal distribution, but there might be some objective functions that make the swarm becomes different probability distribution. If so, then our estimation will fail on these objective functions, more sophisticated model is needed to provide good estimation. Finally, the model we proposed in this thesis does not take personal experience in to consideration, again we need a more sophisticated model to analyze the behavior influenced by personal experience.

Bibliography

[1] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks*, vol. 1942-1948, 1995.

[2] J. Kennedy and R. C. Eberhart, "A new optimizer using paritcle swarm theory," in *Proceeding of the Sixth Int. Symposium on Micromachine and Human Science*, Nagoya, Japan, 1995, pp. 39–43.

[3] G. Beni and J. Wang, "Swarm intelligence in cellular robotic systems," in *Proceedings of NATO Advanced Workshop on Robots and Biological Systems*, 1989, pp. 26–30.

[4] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press, 1999, ISBN: 0-19-513159-2.

[5] R. Eberhart, Y. Shi, and J. Kennedy, *Swarm Intelligence.* Morgan Kaufmann Publishers, 2001, ISBN: 1-55860-595-9.

[6] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*, vol. 3, 1999, pp. 1945–1950.

[7] F. Pan, G. Wang, and Y. Liu, "A multi-objective-based non-stationary UAV assignment model for constraints handling using PSO," in *GEC '09: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, 2009, pp. 459–466.

[8] P. Yan and L. Tang, "PSO algorithm for a scheduling parallel unit batch process with batching," in *GEC '09: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, 2009, pp. 703–708.

[9] T. M. Alkhamis and M. A. Ahmed, "Simulation-based optimization for repairable systems using particle swarm algorithm," in *WSC '05: Proceedings of the 37th conference on Winter simulation*, 2005, pp. 857–861.

[10] R. Wrobel and P. Mellor, "Particle swarm optimisation for the design of brushless permanent magnet machines," in *Industry Applications Conference, 2006. 41st IAS Annual Meeting. Conference Record of the 2006 IEEE*, vol. 4, Oct. 2006, pp. 1891–1897.

[11] J. Kennedy, "The behavior of particles," in *Proceedings of the 7th International Conference on Evolutionary Programming*, 1998, pp. 581–589.

[12] E. Ozcan and C. K. Mohan, "Analysis of a simple particle swarm optimization system," *Intelligent Engineering Systems Through Artificial Neural Networks*, vol. 8, pp. 253–258, 1998.

[13] ——, "Particle swarm optimization: surfing the waves," in *Proceedings of 1999 IEEE Congress on Evolutionary Computation (CEC 99)*, 1999, pp. 1939–1944.

[14] M. Clerc and J. Kennedy, "The particle swarm—explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.

[15] F. van den Bergh, "An analysis of particle swarm optimizers," Ph.D. dissertation, University of Pretoria, 2002.

[16] K. Yasuda, A. Ide, and N. Iwasaki, "Adaptive particle swarm optimization," in *Proceedings of 1999 IEEE International Conference on Systems, Man and Cybernetics*, 2003, pp. 1554–1559.

[17] Y.-L. Zheng, L.-H. Ma, L.-Y. Zhang, and J.-X. Qian, "On the convergence analysis and parameter selection in particle swarm optimization," in *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, 2003, pp. 1802–1807.

[18] I. C. Trelea, “The particle swarm optimization algorithm: convergence analysis and parameter selection,” *Information Processing Letters*, vol. 85, pp. 317–325, 2003.

[19] F. van den Bergh and A. P. Engelbrecht, “A study of particle swarm optimization particle trajectories,” *Information Sciences*, vol. 176, no. 8, pp. 937–971, April 2006.

[20] V. Kadirkamanathan, K. Selvarajah, and P. J. Fleming, “Stability analysis of the particle dynamics in particle swarm optimizer,” *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 245–255, 2006.

[21] M. Jiang, Y. Luo, and S. Yang, “Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm,” *Information Processing Letters*, vol. 102, pp. 8–16, 2007.

[22] Y. Shi and R. C. Eberhart, “Parameter selection in particle swarm optimization,” in *Proceedings of the 7th International Conference on Evolutionary Programming*, 1998, pp. 591–600.

[23] J. Kennedy, “The particle swarm: social adaptation of knowledge,” in *Proceedings of the 1997 IEEE International Conference on Evolutionary Computation*, 1997, pp. 303–308.

[24] M. R. Spiegel, *Schaum's Outline of Theory and Problems of Probability and Statistics.* Mcgraw-Hill, 1975, ISBN: 0-07060-220-4.

[25] X. Yao, Y. Liu, and G. Lin, “Evolutionary programming made faster,” *Evolutionary Computation, IEEE Transactions on*, vol. 3, no. 2, pp. 82–102, Jul 1999.