# New Evaluation Criteria for the Convergence of Continuous Evolutionary Algorithms

Ying Lin, *Student Member IEEE*, Jian Huang and Jun Zhang (Corresponding Author), *MIEEE*

***Abstract*—The first hitting time (FHT) plays an important role in convergence evaluation for evolutionary algorithms. However, the current criteria of the FHT are mostly under a hypothesis that never has been testified: the FHT subjects to the normal distribution. Aiming at more convincible evaluations, this paper investigates the distribution of the FHT through a goodness-of-fit test and discovers an unexpected result. Based on this result, this paper proposes a new set of criteria, which utilizes two types of relative frequency histograms. This paper validates the proposed criteria on the optimization problem of benchmark functions by the standard genetic algorithm (SGA) and the particle swarm optimization (PSO). The experiments show that the proposed criteria are effective to evaluate the convergent speed and the convergent stability of the evolutionary algorithms.**

## I. INTRODUCTION

EVONUTIONARY algorithms (EAs) imitate the evolution process in nature, where better-fit individuals have greater chances to survive and reproduce [1]. There is no doubt that EAs are algorithms of probability, whose outcomes are different in every trial even under the same algorithm structures and control parameters. As one of the outcomes from EAs, the first hitting time (FHT) of the target solution is not an exception. Analyzing a bunch of observations of the FHT and revealing the rule behind them have become a critical step in the evaluation of the convergent performance of EAs.

Researchers have been working on finding the theoretical complexity of the FHT. S. Droste [2][3] *et al.* analyzed the FHT of the (1+1) evolutionary algorithm with one individual and without crossover. Hierarchy results are reported on the linear functions of polynomial degree 2 and unimodal functions. J. He and X. Yao [4] extended the complexity analysis to population-based evolutionary algorithms. They also proposed a framework for the complexity of the FHT in both (1+1) evolutionary algorithms and population-based evolutionary algorithms with crossover [5]. Recently, Y. Zhou and J. He [6] first analyzed the FHT in constrained optimization problems, which resulted in important inferences about the penalty coefficient in the fitness function. However, due to the complex and dynamic structures of the evolutionary algorithms, it is still difficult to summarize a general complexity expression for the FHT. Up till now, researchers have been accustomed to analyze the observations of the FHT by statistical methods.

This work was supported by NSF of China Project No.60573066 and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, P.R. China.

Authors are with Department of Computer Science, SUN Yat-sen University, Guangzhou, P.R.China,. (Jun Zhang is the corresponding author, email:junzhang@ieee.org)

Most of the statistical methods are on the basis of the mean first hitting time (MFHT). The most common used statistical methods include the direct comparison and the t-test analysis [7] of the MHST, the curve of the cumulative frequency of the FHT [8], and the chart of the best fitness or the mean fitness versus generations or function evaluations [9][10]. Another note-worthy method is based on the statistical distribution of the FHT, which was systematically implemented into the long-path problem by J. Garnier and L. Kallel in [11]. All these methods are numbered in Table I.

TABLE I
THE MOST COMMON USED STATISTICAL METHOD ON THE FHT

| No. | Statistical Method |
|---|---|
| 1 | Direct comparison of the mean first hitting time [7] |
| 2 | t-test on the mean first hitting time [7] |
| 3 | The curve of the cumulative frequency of the FHT [8] |
| 4 | The chart of best fitness/mean fitness vs. generations/evaluations [9][10] |
| 5 | Analysis based on the statistical distribution discovered from the observations of the first hitting time [11] |

In the first and the second methods, if the FHT subjects to the normal distribution, using a mean value of the observations can be reasonable, because the mean value is at the same time the expectation and the mode of the FHT. However, if the hypothesis fails, the MFHT is no longer typical. Consequently, the results based on the MFHT can be rather deceptive, especially when the size of the observation set is not large enough. Both the third and the forth methods use charts, thus they can not afford the quantitative analysis. Besides, the forth method has to take the risk that some accidental situations may disturb the charts. The fifth method sounds to be very attractive, but finding the theoretical distribution for every algorithm in different problems is hard. Moreover, such a theoretical distribution may not exist in some situations. Except for the inaccurate judgment to the convergent speed, all the above methods ignore the convergent stability reflected by the FHT.

This paper investigates the distribution of the FHT by the case that the standard genetic algorithm (SGA) and the particle swarm optimization (PSO) optimize continuous functions. The chi-square goodness-of-fit test finds that the distribution of the FHT varies according to different functions and dimensions. This unexpected result leads to the



978-1-4244-1823-7/08/$25.00©2008 IEEE

Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

suspicion of the reliability of the MFHT-based criteria. Thus, this paper proposes new criteria based on the observed frequencies of the FHT. Two types of relative frequency histograms are utilized in the evaluation. The type I relative frequency histogram limits the proportion of its largest category to a predefined range by the dynamic category number and the tunable category boundaries. The range and the 'center' of the largest category are recorded for the evaluation of the convergent speed. The comparison of different EAs is fair because the ranges and the 'centers' are derived from the largest categories with similar proportion. Moreover, the 'center' is typical no matter which distribution the FHT is. The type II relative frequency histogram has a lower bound for the proportion of the largest category. In the comparison among different EAs, the category number is the maximum value that makes all the histograms satisfy the lower bound. The span of the histogram and the standard deviation of the proportion of each category are utilized in the evaluation of the convergent stability. This paper testifies the proposed criteria on the optimization problem of benchmark functions by the SGA and the PSO from [12]. The experiments show that the proposed criteria are effective.

The rest of this paper is organized as follows. Section II defines the optimization problem, the target solution and introduces the histogram. Section III displays the chi-square goodness-of-fit test on the distribution of the FHT. Section IV illustrates the proposed evaluation criteria. Section V provides the experiment results and discussions. Finally, Section VI draws a conclusion. For the sake of convenient descriptions, Table II lists the symbols used in this paper.

## II. INTRODUCTION TO RELATIVE DEFINITIONS

### A. The optimization problem and the target solution

An optimization problem must have an objective function set $\Theta=\{\theta_1,\theta_2,\cdots,\theta_n\},\ n\ge 1$, and a constrain set $\Omega=\{\omega_1,\omega_2,,\omega_m\},\ m\ge 0$. When $n>1$, the problem is a multi-objective one. And when $m=0$, the problem is constrain-free.

A target solution $s^*$ of an optimization problem is a solution which satisfies the following conditions:

1) For any $\theta_i$ belonging to $\Theta$,

$$\theta_i(s^*)\in[\theta_i^*-\varepsilon,\theta_i^*+\varepsilon], \quad (1)$$

where $\theta_i^*$ is the optimal value of the $i^{\text{th}}$ objective function and $\varepsilon$ is the error bound.

2) For any $\omega_i$ belonging to $\Omega$,

$$\omega_i(s^*)=1, \quad (2)$$

where

$$\omega_i(s)=\begin{cases}1, & s \text{ satisfies } \omega_i \\ 0, & \text{otherwise}\end{cases}. \quad .(3)$$

The FHT is the time for an algorithm to reach the target solution for the first time. It can be denoted by three terms: the absolute computer time [13], the generations [8]-[10] and the evaluations [9]. The evaluations is an indirect measurement, which calculates the FHT by counting the evaluation times of the most time-consuming component in an algorithm. These three terms adapt to different situations, but all can be used for statistical analysis.

TABLE II
NOTATION USED IN THIS PAPER

| Symbol | Quantity |
|---|---|
| $\Theta=(\theta_1,\theta_2,\cdots,\theta_n),\ n\ge 1$ | The object function set |
| $\Omega=(\omega_1,\omega_2,\cdots,\omega_m),\ m\ge 0$ | The constrain set |
| $s^*$ | The target solution of an optimization problem |
| $\varepsilon$ | The error bounds of the target solution |
| $N$ | The size of the observation set |
| $n_b$ | The number of categories in the histogram |
| $e_i,\ i=0,1,\cdots,n_b$ | The boundaries of the categories |
| $L, U$ | The lower and the upper bounds of the histogram |
| $p_i,\ i=1,2,\cdots,n_b$ | The relative frequencies/proportions of the categories in the histogram |
| $O_i,\ i=1,2,\cdots,n_b$ | The observed frequencies of the categories in the histogram |
| $f_{obj}(\cdot)$ | The PDF of the given object *obj* |
| $F$ | A theoretical distribution |
| $E_i,\ i=1,2,\cdots,n_b$ | The absolute frequencies of the categories in the expected distribution |
| $\Phi$ | The range of the largest category |
| $P^\alpha$ | The peak at level $\alpha$ |
| $\Psi$ | The span of the histogram |
| $\sigma_p$ | The standard deviation of the proportions of the categories in the histogram |

### B. The histogram

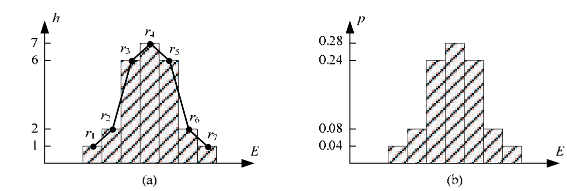



Fig.1. The histograms of absolute frequencies and relative frequencies when $N=25$ and $n_b=7$.

A multinomial experiment [13] samples $N$ observations $t_1,t_2,\cdots,t_N$ of the FHT from $N_0$ independent trails, which are performed under the same configurations. Then the $N$ observations will be classified into $n_b$ categories. Each category spans an closed-open interval $[e_i,e_{i+1})$ $(i=0,1,\cdots,n_b-2)$, except that the last category covers a closed interval $[e_{n_b-1},e_{n_b}]$. The width of each interval is equal, besides, the boundaries satisfy $e_0\le\min_{1\le i\le N}(t_i)$ and $e_{n_b}\ge\max_{1\le i\le N}(t_i)$. The absolute frequency of a category is the number of the observations that belong to this category. The relative frequency of a category is the proportion of the



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

observations in this category. A histogram is a graph that marks the categories on the horizontal axis and the absolute frequencies or the relative frequencies on the vertical axis. Fig.1(a) and Fig.1(b) display the histograms of absolute frequencies and relative frequencies by an example of 25 observations and 7 categories. As it can be seen, in the histogram, each bar corresponds to a category and there is no gap between bars.

As it is shown in Fig.1(a), a polygon connects the midpoints at the top of each bar in the absolute frequency histogram. When the size of the observation set and the number of the categories increase, the number of the midpoints rises. Thus, the edge between every two midpoints becomes shorter, which makes the polygon approaches the frequency curve of the FHT [13]. Suppose the absolute frequencies are transformed into the relative frequencies and the width of each bin is normalized to 1, the polygon actually simulates the curve of the probability density function (PDF) of the FHT. The PDF of the FHT satisfies

$$\int_{e_i}^{e_{i+1}} f_{\mathrm{FHT}}(t)\mathrm{d}t = p_{i+1}(e_{i+1}-e_i) = p_i, \quad i = 0,1,\cdots,n_b\,. \tag{4}$$

and

$$\int_{e_0}^{e_{n_b}} f_{\mathrm{FHT}}(t)\mathrm{d}t = \sum\nolimits_{i=0}^{n_b-1} p_{i+1}(e_{i+1}-e_i) = \sum\nolimits_{i=1}^{n_b} p_i = 1 \tag{5}$$

Assume a null hypothesis $H_0$: the FHT follows a certain theoretical distribution $F_0$. If the null hypothesis is not rejected, the scaled curve of the PDF from $F_0$ can fit the midpoints at the top of each bar in the histogram of the absolute frequencies.

## III. Goodness-of-fit test of the FHT on specified data

TABLE III
Test Functions

| Test functions | SD | $f_{\min}$ |
|---|---|---|
| $f_1(x)=\sum_{i=1}^{d} x_i^2$ | $[-100,100]^d$ | 0 |
| $f_2(x)=\sum_{i=1}^{d} \lvert x_i\rvert+\prod_{i=1}^{d}\lvert x_i\rvert$ | $[-10,10]^d$ | 0 |
| $f_3(x)=\sum_{i=1}^{d}\lfloor x_i+0.5\rfloor^2$ | $[-100,100]^d$ | 0 |
| $f_4(x)=\sum_{i=1}^{d} -ix^4+random[0,1)$ | $[-100,100]^d$ | 0 |
| $f_5(x)=\pi/30\{10\sin^2(\pi y_1) + \sum_{i=1}^{d-1}(y_i-1)^2[1+10\sin^2(\pi y_{i+1})] + (y_n-1)^2\}+\sum_{i=1}^{d}u(x_i,10,100,4)$ * | $[-50,50]^d$ | 0 |
| $f_6(x)=1/10\{10\sin^2(3\pi x_1) + \sum_{i=1}^{d-1}(x_i-1)^2[1+\sin^2(3\pi x_{i+1})] + (x_n-1)^2[1+\sin^2(2\pi x_n)]\} + \sum_{i=1}^{d}u(x_i,5,100,4)$ * | $[-50,50]^d$ | 0 |

* $y_i = 1+(x_i+1)/4$, $u(x_i,a,k,m)=\begin{cases}k(x_i-a)^m & x_i>a\\ 0 & -a\le x_i\le a\\ k(-x_i-a)^m & x_i<-a\end{cases}$

Most continuous problems can be summarized as a function optimization model. Thus, this paper takes the continuous function optimization into investigation. In that case, calculation of the optimized function is the most time-consuming component in EAs. For the sake of fair comparison, the evaluations of the optimized functions are selected to represent the FHT. The observation sets originate from 1,000 trials of the function optimization by the SGA and the PSO [12]. The dimensions of the optimized functions are 5, 15 and 30, with the error bounds of the target solution at 0.01, 0.1 and 0.1, respectively. The SGA uses 20 chromosomes with the probability of crossover at 0.7 and the probability of mutation at 0.07. The PSO uses 20 particles with $c_1 = c_2 = 2$ and a linearly declining $w$. The test functions are listed in Table III. The first four are unimodal functions while the last two are multimodal functions. $d$ in Table III indicates the dimensions of the function, and SD is the range of the independent variables. The investigation is performed on the basis of the above data.

### A. A goodness-of-fit test

The normal distribution is generally considered to be appropriate for the FHT. This section is going to testify this hypothesis through a goodness-of-fit test. Besides, another three theoretical distributions are also tested for comparison. They are the 2-parameter Weibull distribution, the log-normal distribution and the Gamma distribution. Except the normal distribution, the shapes of the curves of the PDF from the other three are dissymmetric. The PDFs of the four distributions are listed in Table IV.

TABLE IV
The PDF of the Testified Distribution

| Distribution | PDF |
|---|---|
| Normal distribution | $f_{\mathrm{N}}=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ |
| Weibull distribution (2-parameter) | $f_{\mathrm{WBL}}(x;m,\eta)=\dfrac{m(x)^{m-1}}{\eta^m}e^{-\left(\frac{x}{\eta}\right)^m}$ |
| Log-normal distribution (2-parameter) | $f_{\mathrm{LN}}(x;\mu,\sigma)=\dfrac{1}{\sqrt{2\pi}x\sigma}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$ |
| Gamma distribution | $f_{\Gamma}(x;\sigma,\alpha)=\dfrac{1}{\sigma\Gamma(\alpha)}\left(\dfrac{x}{\sigma}\right)^{\alpha-1}e^{-\frac{x}{\sigma}}$ * |

* $\Gamma(\cdot)$ is the gamma function.

Assume a null hypothesis $H_0$ that the FHT subjects to the distribution $F_0$, a goodness-of-fit test is used to test how well the observed frequencies fit $F_0$ [15]. The chi-square goodness-of-fit test ($\chi^2$ test) is one of the most common used methods. Because the $\chi^2$ test is especially suitable for large sets of observations and it allows for estimated parameters in $F_0$ [16], this paper chooses the $\chi^2$ test in the investigation.

To the opposite of the observed frequencies $O_i$ $(i=1,2,\cdots,n_b)$, the expected frequencies from $F_0$ is denoted by $E_i$ $(i=1,2,\cdots,n_b)$, which is calculated by



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

$$E_i = N \cdot p_i,\ \ i = 1, 2, \cdots, n_b\,. \tag{6}$$

$p_i$ is the probability for the observation to fall into the $i$th category when $H_0$ is not rejected. The $\chi^2$ test makes the decision to reject or not reject $H_0$ based on how large $O_i$ differs from $E_i$, which is measured by $\chi^2$ in (7).

$$\chi^2 = \sum_{i=1}^{n_b} (O_i - E_i)^2 / O_i \tag{7}$$

$\chi^2$ subjects to the chi-square distribution with the degree of freedom at $v = n_b - k - 1$, where $k$ is the number of the estimated parameters in $F_0$. $\chi^2_\alpha(v)$ is the value derived from the chi-square distribution table. If $\chi^2$ is not smaller than $\chi^2_\alpha(v)$, the small-probability-event occurs and $H_0$ is rejected at $\alpha$ level of significance. A smaller $\alpha$ produces a stronger rejection [15].

TABLE V

RESULTS OF $\chi^2$ TEST ON THE TEST FUNCTIONS OF DIFFERENT DIMENSIONS FROM SGA AND PSO

| $d$ | $\varepsilon$ | $f$ | SGA | | | | PSO | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $N$ | $F_{\text{wbl}}$ | $F_{\ln}$ | $F_\Gamma$ | $N$ | $F_{\text{wbl}}$ | $F_{\ln}$ | $F_\Gamma$ |
| 5 | $10^{-2}$ | 1 | 1 | 1 | **0** | **0** | 1 | **0** | 1 | 1 |
| | | 2 | 1 | **0** | 1 | **0** | 1 | **0** | 1 | 1 |
| | | 3 | 1 | 1 | **0** | **0** | 1 | **0** | 1 | 1 |
| | | 4 | 1 | 1 | **0** | **0** | **0** | 1 | 1 | **0** |
| | | 5 | 1 | 1 | 1 | 1 | 1 | **0** | 1 | 1 |
| | | 6 | 1 | 1 | 1 | 1 | **0** | **0** | 1 | 1 |
| 15 | $10^{-1}$ | 1 | 1 | 1 | **0** | **0** | **0** | 1 | **0** | **0** |
| | | 2 | 1 | 1 | **0** | **0** | **0** | 1 | **0** | **0** |
| | | 3 | 1 | 1 | **0** | **0** | **0** | 1 | **0** | **0** |
| | | 4 | 1 | 1 | **0** | 1 | **0** | 1 | **0** | **0** |
| | | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | 6 | 1 | 1 | **0** | **0** | **0** | 1 | 1 | 1 |
| 30 | $10^{-1}$ | 1 | **0** | **0** | 1 | 1 | 1 | 1 | 1 | 1 |
| | | 2 | 1 | 1 | 1 | 1 | **0** | 1 | **0** | **0** |
| | | 3 | 1 | 1 | **0** | 1 | 1 | 1 | 1 | 1 |
| | | 4 | 1 | 1 | **0** | **0** | **0** | 1 | **0** | **0** |
| | | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | 6 | **0** | 1 | **0** | **0** | **0** | 1 | **0** | **0** |
| $\rho$ (%) | | | 91 | 91 | 33 | 43 | 44 | 72 | 61 | 56 |

In the $\chi^2$ test of this paper, $n_b$ follows the rule in reference [17], that is,

$$n_b = \lfloor N / 2\log_2 N \rfloor. \tag{8}$$

$\alpha$ equals 0.05 and the maximum likelihood estimation is used to estimate the unknown parameters in the testified distributions. The results of $\chi^2$ test are listed in Table V. $N$, $F_{\text{WBL}}$, $F_{\text{LN}}$ and $F_\Gamma$ indicate the normal distribution, the Weibull distribution, the log-normal distribution and the Gamma distribution. $\varepsilon$ is the error bound of the target solution. '1' indicates the rejection to the null hypothesis that the FHT subjects to the specified distribution, while '0' indicates the opposite meaning. $\rho$ in the last row is the rejection rate for the distribution on the observation sets in the same column, which can be calculated by

$$\rho = \frac{\sum_{i=1}^{\kappa} H_i}{\kappa}. \tag{9}$$

$\kappa$ is the number of the observation sets. In Table V, $\kappa$ equals to 18.

As it's shown on the left part of Table V, the FHT from the SGA does not follow the normal distribution in most cases. The log-normal distribution and the Gamma distribution have better chances to be accepted. For the sake of further examination, Fig.2 displays three histograms of the absolute frequencies from the optimization of $f_1$ and $f_7$ by the SGA. The graphs in Fig.2 also depict the frequency curves of the four distributions. The graphs show that the distribution of the FHT is not close to a symmetric bell-shaped curve but a dissymmetric one with a tail on one side. It is also noticed that the higher the dimension is, the more symmetric the curve will be. From the right part of Table V, it can be seen that in the observation set from the PSO, the rejection rate of the normal distribution decreases. However, there is still no certain rule for the distribution of the FHT.

Concluded from Table V and Fig.3, it could be said that the distribution of the FHT does not always subject to the normal distribution; moreover, since the performance of an EA is influenced by many factors (e.g. the control parameters), it's quite probable that none of the theoretical distributions can fit all the cases. Thus, there is a great need to set up new criteria without statistical distribution models.

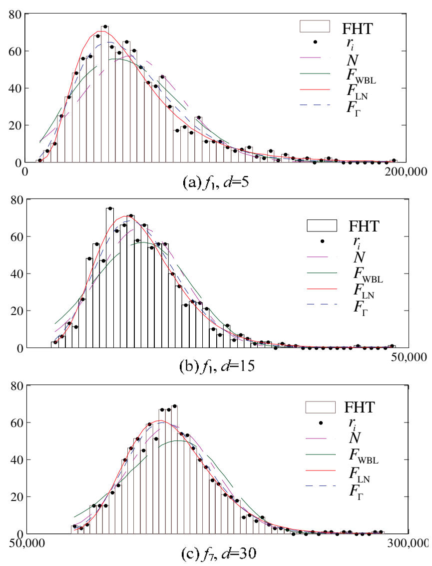



Fig.2. The graphs from the optimization of f1 and f7 with 5, 15 and 30 dimensions by the SGA

## IV. OBSERVED-FREQUENCY-BASED EVALUATION CRITERIA

Even though the histogram can not comply with a certain distribution, it does reflect the distribution of the FHT. The evaluation criteria proposed by this paper are on the basis of the relative frequency histogram. Applying different limitations to the constructing process can produce different types of histograms, which are used for the evaluation of the convergent speed and the convergent stability, respectively.



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

### A. *The type* I *relative frequency histogram*

Suppose $N$ observations $t_1, t_2, \cdots, t_N$ and an initial $n_b$, the boundaries of the categories are set to be

$$e_i = L + i \times \frac{U-L}{n_b},\ i = 0,1,\cdots,n_b \tag{10}$$

where $L$ and $U$ satisfy $\min_{1\le i\le n_b}(t_i) \ge L$ and $\max_{1\le i\le n_b}(t_i) \le U$. According to these boundaries and the relative frequencies in each category, the initial relative frequency histogram is born. The height of the bars in the histogram are denoted by $p_i (i = 1,2,\cdots,n_b)$. If $\max_{1\le i\le n_b}(p_i)$ is out of the range of $\alpha \pm \delta$, $n_b$ must be adjusted by the rule

$$n_b = \begin{cases} n_b - 1, & \max(p_i) < \alpha - \delta \\ n_b + 1, & \max(p_i) > \alpha + \delta \end{cases}. \tag{11}$$

Then the new boundaries are set according to (10) and another relative frequency histogram is produced. The above process carries on until the histogram satisfies the limitation $\alpha - \delta \le \max_{1\le i\le n_b}(p_i) \le \alpha + \delta$. But there is one situation that the adjustment of $n_b$ is too rough to fulfill the limitation, at that time a more precise adjustment is performed by tuning the boundaries of the largest category as follows:

$$e_i = \begin{cases} e_i + \lambda \cdot (U-L)/n_b, & p_i > \alpha + \delta,\ i \ne 1 \\ \max(L, e_i - \lambda \cdot w_b), & p_i < \alpha - \delta \end{cases} \tag{12}$$

$$e_{i+1} = \begin{cases} e_{i+1} - \lambda \cdot (L-U)/n_b, & p_i > \alpha + \delta,\ i \ne n_b + 1 \\ \min(U, e_{i+1} + \lambda \cdot w_b), & p_i < \alpha - \delta \end{cases} \tag{13}$$

where $i$ is the index of the largest category and $\lambda$ is a predefined coefficient for the step of the precise adjustment. When the limitation is finally met, the mean of the observations in the largest category is named the peak at level $\alpha$, denoted by $P^\alpha$. The range of the largest category is recorded as $\Phi$. The flowchart of the above process is shown in Fig.3.

Derived from the largest category, $P^\alpha$ implies that an algorithm is most probable to achieve a first hitting time near it. When the FHT does not subject to the normal distribution, the mean value of $t_i (i = 1,2,\cdots,n_b)$ is not any more the maximum likelihood estimation of the expectation. In that case, $P^\alpha$ is more typical than the mean value in describing the general convergent speed. Comparing $P^\alpha$ from different algorithms can reveal the relationship of their convergent speed in most cases. Fixing the parameter $\alpha$ forces a fair comparison. A large $\alpha$ improves the confidence of the judgment, but at the same time increases the error bounds of $P^\alpha$ by enlarging $\Phi$. The following rules are summarized in the comparison between two algorithms:

1) Suppose the size of $\Phi_A$ and $\Phi_B$ are at the same level:

   *Rule a*-1: If $\Phi_A \cap \Phi_B$ is nearly empty, the algorithm with a smaller $P^\alpha$ converges faster than the other.

   *Rule a*-2: If the area of $\Phi_A \cap \Phi_B$ is broad, the two algorithms have similar convergent speed.

2) Suppose $\Phi_A$ is much narrower than $\Phi_B$:

   *Rule b*-1: If $\Phi_A \cap \Phi_B = \phi$, the convergent speed of an algorithm is determined by $P^\alpha$. Smaller $P^\alpha$ indicates better convergent speed.

   *Rule b*-2: If $\Phi_A \cap \Phi_B \ne \phi$, the convergent speed relies on the relationship between $P_A^\alpha$ and $P_B^\alpha$. If $P_A^\alpha$ is smaller than or close to $P_B^\alpha$, algorithm $A$ is considered to converge faster than $B$; otherwise, algorithm $A$ has a slower convergent speed than $B$.

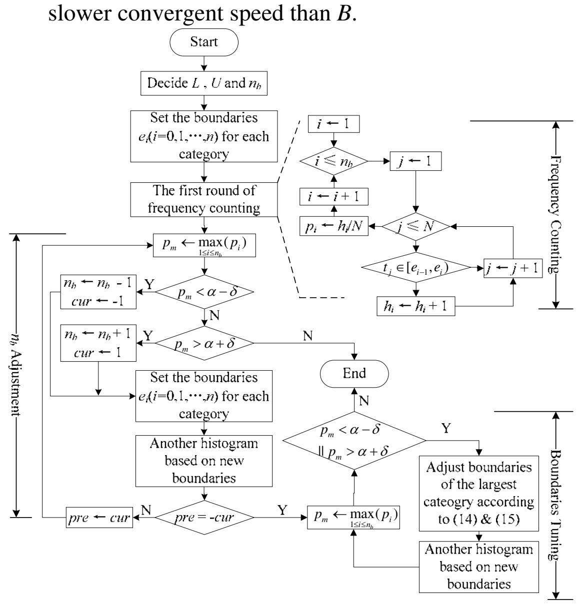



Fig.3. The Flowchart of the constructing process for the relative frequency histogram type I

The levels of the four rules above are depicted in Fig.4, which are also the steps to follow in practical cases.

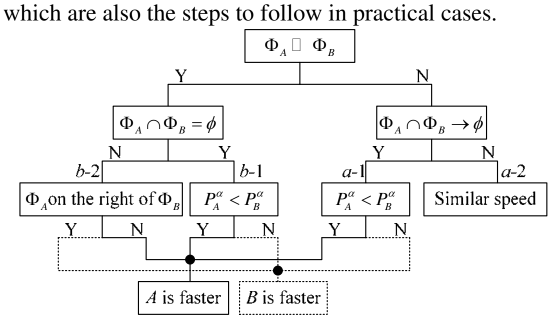



Fig.4. The levels of the rules in the comparison of convergent speed

### B. *The type* II *relative histogram*

Another relative frequency histogram is utilized to compare the convergent stability. This histogram uses the $n_b$ adjustment in Fig.4 to insure that the proportion of the largest category exceeds a predefined lower bound $\beta$. The histograms from different algorithms must use equal $n_b$ in the comparison. Thus, $n_b$ is determined as the maximum value that makes all the histograms satisfy the predefined



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

lower bound.

$\sigma_p$ records the standard deviation of $p_i\,(i=1,2,\cdots,n_b)$ in

TABLE VI
THE COMPARISON OF CONVERGENT SPEED BY $\mu_{\text{ALG}}$ AND THE PROPOSED CRITERIA

| $D$ | $f$ | ALG | $\Phi_{\text{ALG}}$ | $P^{\alpha}_{\text{ALG}}$ | $\max\limits_{1\le i\le n_b}(p_i)$ | $\text{MFHT}_{\text{ALG}}$ | $f$ | ALG | $\Phi_{\text{ALG}}$ | $P^{\alpha}_{\text{ALG}}$ | $\max\limits_{1\le i\le n_b}(p_i)$ | $\text{MFHT}_{\text{ALG}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 1 | SGA | [27913, 57535] | 42877 | 0.504 | 55662 | 4 | SGA †* | [2180, 17525] | **12099** | 0.492 | **19520** |
| | | PSO †* | [39188, 45762] | **41793** | 0.49 | **39220** | | PSO | [51176, 64221] | 57040 | 0.494 | 54015 |
| | 2 | SGA | [48682, 100344] | 75742 | 0.492 | 94395 | 5 | SGA †* | [1340, 14027] | **8225** | 0.494 | **19962** |
| | | PSO †* | [37361, 42382] | **40306** | 0.49 | **41075** | | PSO | [25926, 33093] | 29451 | 0.492 | 26557 |
| | 3 | SGA †* | [620, 4042] | **2810** | 0.496 | **4531** | 6 | SGA †* | [2940, 23209] | **14058** | 0.506 | **30922** |
| | | PSO | [19033, 28764] | 24882 | 0.494 | 27423 | | PSO | [28029, 35802] | 32442 | 0.502 | 33742 |
| 15 | 1 | SGA † | [48270, 76640] | **63584** | 0.494 | 74843 | 4 | SGA †* | [32740, 50620] | **41834** | 0.502 | **49084** |
| | | PSO | [73169, 76670] | 74771 | 0.494 | **74389** | | PSO | [83323, 89296] | 86098 | 0.506 | 86009 |
| | 2 | SGA †* | [44427, 68075] | **56730** | 0.498 | **64353** | 5 | SGA †* | [2520, 8754] | **6904** | 0.496 | **10665** |
| | | PSO | [65761, 70258] | 68356 | 0.498 | 69814 | | PSO | [56320, 62634] | 59540 | 0.508 | 59760 |
| | 3 | SGA †* | [10940, 17600] | **14004** | 0.502 | **15031** | 6 | SGA †* | [11518, 18231] | **15246** | 0.496 | **18291** |
| | | PSO | [74111, 81403] | 77242 | 0.49 | 75162 | | PSO | [66007, 70384] | 68236 | 0.51 | 68242 |
| 30 | 1 | SGA | [525202, 674498] | 596107 | 0.49 | 569638 | 4 | SGA | [386338, 546202] | 456114 | 0.502 | 447209 |
| | | PSO †* | [101101, 105368] | **102882** | 0.494 | **101385** | | PSO †* | [119470, 130421] | **125666** | 0.49 | **129647** |
| | 2 | SGA | [663950, 799940] | 726343 | 0.492 | 658107 | 5 | SGA †* | [38938, 61937] | **47570** | 0.502 | **48139** |
| | | PSO †* | [93470, 96595] | **94945** | 0.496 | **94800** | | PSO | [89354, 99156] | 93461 | 0.496 | 92740 |
| | 3 | SGA †* | [72545, 108818] | **90132** | 0.506 | **101837** | 6 | SGA | [135239, 175898] | 152697 | 0.49 | 27974 |
| | | PSO | [103929, 112325] | 107808 | 0.504 | 109247 | | PSO †* | [99312, 104226] | **102180** | 0.494 | **103102** |

† and * indicate the algorithm with faster convergent speed concluded by the proposed criteria and $\mu_{\text{ALG}}$ in-order.

TABLE VII
THE COMPARISON OF CONERGENT STABILITY BY $\sigma_{\text{ALG}}$ , $cv_{\text{ALG}}$ AND THE PROPOSED CRITERIA

| $D$ | $f$ | ALG | $\Psi_{\text{ALG}}$ | $\sigma_p^{\text{ALG}}$ | $\sigma_{\text{ALG}}$ | $cv_{\text{ALG}}$ | $f$ | ALG | $\Psi_{\text{ALG}}$ | $\sigma_p^{\text{ALG}}$ | $\sigma_{\text{ALG}}$ | $cv_{\text{ALG}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 1 | SGA | [6120, 159400] | 0.152 | 25824 | 0.464 | 4 | SGA | [3500, 63040] | **0.174** | 9357 | 0.477 |
| | | PSO †* | **[20960, 49079]** | **0.154** | **4152** | **0.106** | | PSO *# | **[25039, 81258]** | 0.141 | **8530** | **0.158** |
| | 2 | SGA | [13720, 196740] | 0.121 | 38768 | 0.412 | 5 | SGA | [1640, 135300] | **0.255** | 16717 | 0.837 |
| | | PSO †* | **[27180, 49520]** | **0.234** | **3468** | **0.084** | | PSO †*# | **[4618, 37199]** | 0.149 | **5382** | **0.203** |
| | 3 | SGA †* | **[840, 17400]** | **0.158** | **4345** | 0.488 | 6 | SGA | [2940, 170400] | **0.191** | 24507 | 0.793 |
| | | PSO | [7059, 41658] | 0.143 | 5793 | **0.208** | | PSO †*# | **[13480, 51500]** | 0.146 | **5335** | **0.158** |
| 15 | 1 | SGA | [30740, 159520] | **0.175** | 20375 | 0.272 | 4 | SGA | [16920, 113460] | **0.164** | 14775 | 0.301 |
| | | PSO †*# | **[65417, 80877]** | 0.165 | **2598** | **0.035** | | PSO †*# | **[71378, 101054]** | 0.148 | **4526** | **0.053** |
| | 2 | SGA | [24580, 172380] | **0.166** | 17458 | 0.273 | 5 | SGA | [3860, 61740] | **0.283** | 6208 | 0.560 |
| | | PSO †*# | **[61656, 78615]** | 0.143 | **2651** | **0.038** | | PSO *# | **[39372, 79179]** | 0.182 | **5405** | **0.090** |
| | 3 | SGA | **[6560, 34520]** | **0.203** | 5045 | 0.328 | 6 | SGA | [8380, 36300] | **0.181** | 4817 | 0.263 |
| | | PSO *# | [57818, 88018] | 0.152 | **4661** | **0.062** | | PSO *# | **[56156, 79138]** | 0.151 | **3296** | **0.048** |
| 30 | 1 | SGA | [321180, 796880] | 0.170 | 96645 | 0.161 | 4 | SGA | [189120, 799260] | 0.140 | 108472 | 0.241 |
| | | PSO †*# | **[92565, 108119]** | **0.244** | **2525** | **0.025** | | PSO †*# | **[111118, 155397]** | **0.180** | **7381** | **0.057** |
| | 2 | SGA | [344280, 799300] | 0.177 | 91223 | 0.140 | 5 | SGA | [21420, 142980] | **0.189** | 17527 | 0.364 |
| | | PSO †*# | **[88600, 102284]** | **0.226** | **2364** | **0.025** | | PSO †*# | **[76670, 118094]** | 0.153 | **6819** | **0.074** |
| | 3 | SGA | [37640, 289540] | 0.156 | 33354 | 0.328 | 6 | SGA | [87900, 284020] | **0.174** | 27974 | 0.192 |
| | | PSO †*# | **[93551, 168944]** | **0.237** | **8124** | **0.074** | | PSO †*# | **[94633, 114232]** | 0.139 | **3181** | **0.031** |

†, *, and # indicate the algorithm with better convergent stability concluded by the proposed criteria, $\sigma_{\text{ALG}}$ and $cv_{\text{ALG}}$ in-order.

a histogram. Since the sum of $p_i$ always equals 1, an especially large category reduces the proportions of the others. Thus, large $\sigma_p$ indicates the concentration on some categories, which implies a better stability on the algorithm's possible range. Besides $\sigma_p$, in the comparison of different algorithms, the size and the position of their possible ranges $\Psi=[\min\limits_{1\le i\le n_b}(t_i),\max\limits_{1\le i\le n_b}(t_i)]$ must be taken into consideration, too.

$\sigma_p$ describes the frequency of the vibration while $\Psi$ is for the amplitude of the vibration. The proposed criteria compare the convergent stability of the two algorithms as follows:

*Rule c*: Suppose the size of $\Psi_A$ and $\Psi_B$ is almost equal, the decision depends on $\sigma_p$. If $\sigma_p^A > \sigma_p^B$, algorithm $A$ performs better in term of convergent stability. If $\sigma_p^A \approx \sigma_p^B$, the stability of these two algorithms is alike.



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

*Rule d*: Suppose $\Psi_A$ is significantly narrower than $\Psi_B$, algorithm *A* is advantageous in the convergent stability because the FHT concentrates on small range.

Suppose $\Psi_A$ and $\Psi_B$ are on the same level:

*Rule e*-1: If both $\Psi_A$ and $\sigma_P^A$ are larger than those of *B*, *A* not only has a large category but also some small categories spreading over a wide range. In that case, the conclusion on convergent stability can be described by two terms: algorithm *A* does not vibrate very often, but once it does, the amplitude is great.

*Rule e*-2: If algorithm *A* has a narrower $\Psi$ and greater $\sigma_p$, *A* is definitely more stable than *B*.

## V. Experiments and Discussions

Based on the data from Section III, this section will implement the proposed criteria to compare the convergent performance of the SGA and the PSO. The differences between the conclusions from the proposed and the other criteria will be discussed.

The parameters in the proposed criteria are set as: $\alpha = 0.5$, $\delta = 0.01$, $\lambda = 0.01$, and $\beta = 0.35$. The initial value of $n_b$ is 8. Every comparison between the SGA and the PSO is performed on the observation sets from 500 trials. For the sake of steady conclusions, the size of the observation set is larger than normal.

### *A. The comparison in term of convergent speed*

The traditional criterion based on the mean FHT (MFHT) is chosen for comparison. Table VI lists all of the related outcomes, where 'ALG' indicates the name of the algorithm. The bold letters in the table are the values that are advantageous when comparing to the other one in the same unit. The superscripts † and * indicate the conclusions from the proposed and the traditional criteria, respectively.

When *d*=5, as it is shown by the first part of Table VI, the SGA outperforms the PSO from $f_3$ to $f_6$ while the PSO wins in $f_1$ and $f_2$. No matter judging from the traditional or the proposed criteria, these conclusions are the same. However, the two types of criteria differ on how much one algorithm outperforms the other. In $f_1$, the difference between $P_{\mathrm{SGA}}^{\alpha}$ and $P_{\mathrm{PSO}}^{\alpha}$ is much smaller than that between the two MFHT. Nevertheless, a much narrower $\Phi_{\mathrm{PSO}}$ confirms the advantage of the PSO according to *Rule b*-2. As for $f_5$ and $f_6$, the proposed criteria conclude much larger advantages on SGA. As it can be seen, $\Phi_{\mathrm{SGA}}$ and $\Phi_{\mathrm{PSO}}$ are not overlapped and their sizes are not significantly different. Thus, according to *Rule a*-1, the SGA converges much faster than the PSO because of the great difference between $P_{\mathrm{PSO}}^{\alpha}$ and $P_{\mathrm{SGA}}^{\alpha}$. We suppose the MFHT of the SGA is biased because the SGA sometimes achieves large evaluations. The type I histogram of $f_6$ is drawn to validate this conclusion. As it's shown in Fig.5, a large proportion of the SGA's histogram is on the left side of the PSO's. However, the SGA's histograms has a thin long tail on the right, which corresponds to our inference. The same thing happens in $f_5$. As it is seen in this case, the proposed criteria can reveal the fact hidden by the MFHT.

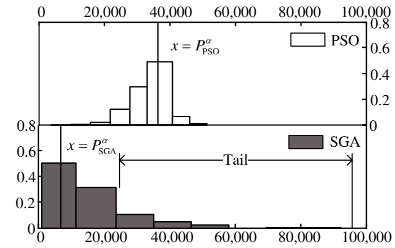



Fig.5. The comparison of the convergent speed between the PSO and the SGA on $f_6$ of 5 dimensions

The second part of Table VI displays the situations of the functions of 15 dimensions. As it can be seen, from $f_2$ to $f_6$, the size of $\Phi_{\mathrm{SGA}}$ and $\Phi_{\mathrm{PSO}}$ are on the same level and they are not overlapped, moreover, $P_{\mathrm{SGA}}^{\alpha}$ is smaller than $P_{\mathrm{PSO}}^{\alpha}$. Thus according to *Rule a*-1, the SGA converges faster than the PSO from $f_2$ to $f_6$, which is the same as the conclusions from the MFHT. The situation in $f_1$ is more ambiguous. Judging from the MFHT, the convergent speed of the PSO and the SGA are almost equal. However, Table VI shows that $\Phi_{\mathrm{PSO}}$ and $\Phi_{\mathrm{SGA}}$ only overlap each other by a small range near their right boundaries, besides, $\Phi_{\mathrm{PSO}}$ is much narrower than $\Phi_{\mathrm{SGA}}$. According to *Rule b*-2, the proposed criteria consider that the SGA converges faster than the PSO. In another 10,000 trials, the SGA converges faster than the PSO in 5,497 cases, which validates the conclusion from the proposed criteria.

As for the functions of 30 dimensions, the third part of Table VI shows that the proposed and the traditional criteria achieve the same conclusions.

### *B. The comparison in term of convergent stability*

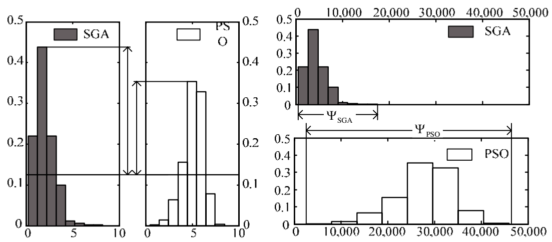
Table VII lists all the related outcomes, where $\sigma_{\mathrm{ALG}}$ and $cv_{\mathrm{ALG}}$ are the standard deviation and the coefficient of variance of the observations, respectively. $cv_{\mathrm{ALG}}$ is calculated by dividing $\sigma_{\mathrm{ALG}}$ with the mean of the observations. The proposed criteria are based on $\sigma_p^{\mathrm{ALG}}$ and $\Psi_{\mathrm{ALG}}$. The bold value has the same meaning as those in Table VII. The superscript †, * and # indicate the winner of the convergent stability according to the proposed criteria, $\sigma_{\mathrm{ALG}}$ and $cv_{\mathrm{ALG}}$, respectively.

The first part of Table VII displays the results of the functions of 5 dimensions. According to *Rule e*-2, the proposed criteria conclude that the PSO is more stable in $f_1$ and $f_2$ because of narrower $\Psi$ and greater $\sigma_p$. And following *Rule d*, PSO is also more stable in $f_5$ and $f_6$ because a much narrower $\Psi_{\mathrm{PSO}}$ is observed. However, the proposed criteria reverse the conclusion from $cv_{\mathrm{ALG}}$ on $f_3$. Fig.6(a) and (b) show the type II histogram of $f_3$ in two ways. The width of



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.

the categories is normalized to 1 in Fig.6(a), where the SGA displays a higher concentration. When the width of the categories is restored in Fig.6(b), the span of the SGA is smaller than that of the PSO. According to the above two points, the new criteria judge according to *Rule e*-2 and declaim that the SGA has better stability. Because the SGA converges much faster than the PSO on $f_3$ (refer to Table VI), a small denominator biases the value of $cv_{\mathrm{SGA}}$ . In this case, the coefficient of variance is deceptive and the proposed criteria are more convincible. Another difficult situation occurs in $f_4$, where both $\Psi_{\mathrm{SGA}}$ and $\sigma_p^{\mathrm{SGA}}$ are greater than those of the PSO. The proposed criteria conclude from both terms of the frequency and the amplitude of the vibration as *Rule e*-1.




(a) $f_3$, normalize the width of bins (b) $f_3$, restore the width of bins

Fig.6. The comparison of the convergent stability between the PSO and the SGA in $f_3$ and $f_4$ of 5 dimensions

As for functions of 15 dimensions, since $\Psi_{\mathrm{PSO}}$ is much narrower than $\Psi_{\mathrm{SGA}}$ in $f_1$, $f_2$ and $f_4$, the proposed criteria cite *Rule d* and conclude that the PSO converges with better stability. These conclusions are the same with those from $\sigma_{\mathrm{ALG}}$ and $cv_{\mathrm{ALG}}$ . In $f_3$, the size of $\Psi_{\mathrm{PSO}}$ and $\Psi_{\mathrm{SGA}}$ are very close and the difference between $\sigma_p^{\mathrm{SGA}}$ and $\sigma_p^{\mathrm{PSO}}$ is not significant. Thus, the proposed criteria use *Rule c* to judge that the SGA and the PSO have similar convergent stability. The proposed criteria apply *Rule e*-1 to $f_5$ and $f_6$, which analyze their situations in two ways.

Based on *Rule e*-2 and *Rule d*, the new criteria conclude that the PSO performs more stably than the SGA on all the functions of 30 dimensions, which is the same as the conclusions from $\sigma_{\mathrm{ALG}}$ and $cv_{\mathrm{ALG}}$ .

After all these experiments on the test functions, we summarize that the proposed criteria are effective and sometimes can solve the deceptive conclusion and reveal the hidden facts from the other criteria.

## VI. Conclusions

This paper investigates the distribution of the first hitting time in the function optimization problem by the SGA and the PSO. It discovers that the distribution of the FHT does not always comply with the normal distribution. This unexpected result leads to the idea of the new criteria without the usage of theoretical distributions. Therefore, this paper proposes a new set of criteria based on the observed frequencies of the FHT. The proposed criteria evaluate the convergent performance of an algorithm from the speed and the stability. Experiments show that the proposed criteria can work effectively in the function optimization problem.

The study of this paper is performed on the SGA and the PSO. However, since the histograms have no additional requests on algorithms, the application of the proposed criteria can be extended to other probability algorithms.

## References

[1] T. Back, U. Hammel, H.P. Schwefel: “Evolutionary computation: comments on history and current states,” *IEEE Trans. on Evolutionary Computation.* vol.1, No.1, pp. 3-17, Apr. 1997.

[2] S. Droste, T. Jansen, and I. Wegener, “On the analysis of (1+1) evolutionary algorithms,” in *Collection of SF1 531*, Feb. 1998.

[3] ---, “A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs,” in *Proc. CEC. 1998*, Anchorage, USA, pp.499-504, May. 1998.

[4] J.He, and X. Yao, “From an individual to a population: an analysis of the first hitting time of population-based evolutionary algorithms,” *IEEE Trans. on Evolutionary Computation*, vol.6, no.5, pp.495-501, Oct. 2002.

[5] ---, “Towards an analytic framework for analyzing the computation time of evolutionary algorithms,” *Artificial Intelligence*, vol.145, pp.59-97, 2003.

[6] Y. Zhou, and J. He, “A runtime analysis of evolutionary algorithms for constrained optimization problems,” *IEEE Trans. on Evolutionary Computation*, vol.11, no.5, pp.608-619, Oct. 2007.

[7] R. Caponetto, L. Fortuna, S. Fazzino, and M. G. Xibilia, “Chaotic sequence to improve the performance of evolutionary algorithms,” *IEEE Trans. on Evolutionary Computation*, vol.7, no.3, pp.290-305, June 2003.

[8] J. Zhang, H. Chung, and W.L. Lo, “Clustering-based adaptive crossover and mutation,” *IEEE Trans. on Evolutionary Computation*, vol.11, no.3, pp.326-335, June 2007.

[9] R. Salomon, “Evolutionary algorithms and gradient search: similarities and differences,” *IEEE Trans. on Evolutionary Computation*, vol.2, no.2, pp.45-55, July 1998.

[10] S. Tsutsui, A. Ghosh, “Genetic algorithms with a robust solution searching scheme,” *IEEE Trans. on Evolutionary Computation*, vol.1, no.3, pp.201-208, Sep. 1997.

[11] J. Garnier and L. Kallel, “Statistical distribution of the convergence time of evolutionary algorithms for long-path problems,” *IEEE Trans. on Evolutionary Computation*, vol.4, no.1, Apr. 2000.

[12] Y. Shi and R. C. Eberhart, “A modified particle swarm optimizer,” in *Proc. IEEE World Congr. Comput. Intell.*, *1998*, pp.69-73.

[13] K.C. Tan and T.H. Lee, “Evolutionary algorithms with dynamic population size and local exploration for multiobjective optimization,” *IEEE Trans. on Evolutionary Computation*, vol.5, no.6, pp.565-588, Dec. 2001.

[14] George E.P. Box, J. Stuart Hunter, and William G. Hunter, *Statistics for experiment: design, innovation, and discovery*, 2nd ed. New Jersey: Wiley. 2005.

[15] Prem S. Mann, *Introductory statistics*. 5th ed. New Jersey: Wiley. 2003.

[16] J.R. Green and D. Margerison, *Statistical treatment of experimental data*. 2nd ed. Netherlands: Elsevier Scientific Publishing Company. 1978.

[17] Z.H. Yang, *The goodness-of-fit test in simulation (Chinese)*, Hefei: Anhui Education Press, 1994.



Authorized licensed use limited to: Hanyang University. Downloaded on November 29,2023 at 07:48:56 UTC from IEEE Xplore. Restrictions apply.