Eighth International Conference on Intelligent Systems Design and Applications

Notes on the Distinction of Gaussian and Cauchy Mutations

Kuo-Torng Lan[1] and Chun-Hsiung Lan[2]

[1]*Department of Information Technology*
Takming University of Science and Technolgy, Neihu, Taipei, Taiwan 114, R.O.C.

[2]*Graduate Institute of Management Sciences*
Nanhua University, Dalin, Chaiyi, Taiwan 622, R.O.C.

Abstract

In evolutionary computation, Gaussian and Cauchy mutations are two popular mutation techniques and completely discussed in this study. It is known that Cauchy mutation has the better ability of escaping local optima, and Gaussian mutation is excellent in local convergence. However, there still exist some vague characteristics in the abilities of local escape/convergence for these two mutations. Therefore, four closed-form equations of probability to clarify those vague behaviors of Gaussian and Cauchy mutations are derived in this paper, and successfully apply to explain the simulated results of benchmark functions. Finally, this paper verifies that Cauchy mutation can prevent the dilemma problem of choosing a proper mutation step size and achieve the acceptable performance except for some specific conditions for evolutionary computation.

Keywords: Gaussian mutation, Cauchy mutation, local escape, local convergence, mutation step size.

1. Introduction

Mutation is an important operator to derive offspring from parents and enhance the diversity of population in discussing the evolutionary computation. While the diversity of the population is larger, the population possesses a better searching capability on the searching space. In former researches, Gaussian mutation is the traditionally predominant choice of evolutionary algorithms, but Cauchy mutation captures many researchers' attention because of infinite variance for its probability density function (pdf). Rudolph [1] investigated the local convergence of EAs by adopting Cauchy mutation, and then concluded that Cauchy mutation has better capability of escaping local optima because of the more diversity of population. He also suggested that Guassian mutation is preferable to Cauchy mutation if a faster local convergence is desired [1].

The simulations of Yao [2] revealed that the fast evolutionary programming (FEP) (*i.e.* EAs with Chauchy mutation) will increase (decrease) the probability of finding a near-optimum while the distance between the current search point and the optimum is large (small). In addition, Yao concluded that Cauchy mutation possesses a higher probability to escape local optimum, but spends more time in exploiting the local neighborhood and thus has a weaker fine-tuning ability than Gaussian mutation in a small or mid-range region [2].

Due to the trade-off decision between the capability of escaping local optima and the capability of local convergence, many researchers try to solve such a dilemma in their heuristic algorithms. Chellapilla [3] linearly combined Gaussian and Cauchy mutations based on their merits to formulate an adaptive mean mutation operator (AMMO). Birru et al. [4] proposed a mixed technique of gradient method and EP with Cauchy mutation to raise the convergent rate and the solution quality. Kim et al. [5] also proposed a mixed mutation operator of Gaussian and Cauchy mutations to obtain the rough convergent characteristics. For lasting researches, Y. Liu and X. Yao [6] presented three techniques to control the mutation step size in FEP.

It is difficult to analyze the convergence of a mutation operator because it is hard to derive the meaningful mathematical theory. Rudolph's theorem [7] proved that the probability of achieving the global optimum will be "1" after infinite-generation evolutions under the situation with a relatively slow decrease of mutation step size. But, in practical applications, escaping local optima and converging to the global optimum should be finished during finite-generation evolutions. Based on above simulation results and theses, this paper derives four closed-form equations of probability to clarify the behaviors of Gaussian and Cauchy mutations, and proposes the suggestions of mutation applications. Section II derives mathematical expressions for Gaussian and Cauchy mutations and two inventive benchmark functions are applied to discriminate the capability of escaping local optima and the capability of local convergence for these two mutations shown in section III, and the simulated results are also demonstrated in the next section. Finally, section IV gives the concluding remarks of this study.

2. Analyses of Two Mutations

As a matter of convenience to describe the proposed theorem, searching the maximum value of objective function is the goal, and the dimension of

978-0-7695-3382-7/08 $25.00 © 2008 IEEE
DOI 10.1109/ISDA.2008.237

individuals $\bar{x}_i$ is set as one. That is $\bar{x}_i = x_i$, $\bar{x}_i' = x_i' = x_i + \Delta x = x_i + \Delta r \cdot X$; where Δr means the mutation step size and X means a random variable. While the Gaussian mutation operator (GMO) is adopted, X is a random variable with the Gaussian distribution. In the meantime, X is denoted as $X = N(0,1)$ and its pdf is $f_N(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. On the other hand, if the Cauchy mutation operator (CMO) is applied, the random variable X is Cauchy distribution and denoted as $X = C$; where the pdf of C is presented as $f_C(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right)$.

Different mutation operators and different mutation step sizes result in different variation of offspring. By inspecting the pdfs of Cauchy and Gaussian distributions and conducting complete simulations, researchers have concluded that the level of variance affects the ability of locally escaping/converging. In order to derive a closed-form expression of probability to explain the relationship among mutation operators and mutation step sizes, two major conditions, local escape and local convergence, of evolutionary algorithms have been discussed to explore the capability of searching solution space. These two major discussions are listed below.

Condition 1: Local Escape

In this case, the landscape with a valley (named valley landscape) hinders individuals from moving toward another hill where a better solution located. Fig. 1 is the valley landscape diagram. In addition, a higher probability of jumping over the valley means a better capability of escaping local optima.

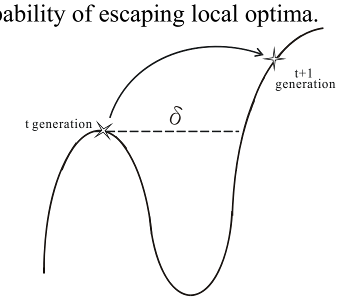


Fig. 1. The valley landscape of the objective function.

For GMO, the successful probability of jumping over the valley is

$$P(\Delta x > \delta) = \int_{\delta/\Delta r}^{\infty} f_N(v)dv = \frac{1}{2} erfc\left(\frac{\delta}{\sqrt{2}\Delta r}\right) \quad (1)$$

Where $erfc(\cdot)$ presents the complementary error function, δ means the width of barrier (i.e. valley width), and Δr is the mutation step size.

For CMO, the successful probability of jumping over the valley is

$$P(\Delta x > \delta) = \int_{\delta/\Delta r}^{\infty} f_C(v)dv = \frac{1}{2} - \frac{1}{\pi} Tan^{-1}\left(\frac{\delta}{\Delta r}\right) \quad (2)$$

Together with Eq. (1) and Eq. (2), it reveals that the successful probability of escaping local optima for CMO is better than GMO's. Then that CMO has a better capability of escaping local optima for a valley landscape than GMO is concluded.

Condition 2: Local Convergence

In this condition, consider the adaptive landscape containing a single convex hill where the global optimum (or a local optimum) is located (called hill landscape), as shown in Fig. 2. An individual has to climb up the top of hill to achieve the optimum. Because the convergent rate is used to be an indicator for measuring the performance in a lot of EA researches, the GMO and CMO are adopted to measure the convergent performances of EAs and completely discussed below:

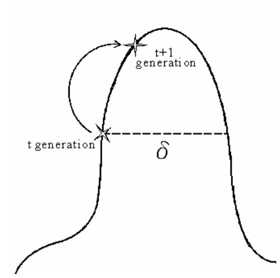


Fig. 2. The diagram of the hill landscape.

Firstly, the discussion goes to the successful probability of climbing up the hill for GMO:

$$P(0 < \Delta x < \delta) = \int_{0}^{\delta/\Delta r} f_N(v)dv = \frac{1}{2} - \frac{1}{2} erfc\left(\frac{\delta}{\sqrt{2}\Delta r}\right) \quad (3)$$

Where δ presents the width of barrier (called hillside width) and Δr is the mutation step size.

Secondly, the successful probability of climbing up the hill for CMO, is

$$P(0 < \Delta x < \delta) = \int_{0}^{\delta/\Delta r} f_C(v)dv = \frac{1}{\pi} Tan^{-1}\left(\frac{\delta}{\Delta r}\right) \quad (4)$$

Eq. (3) and Eq. (4) reveal that the probability of converging to local optimum for CMO in the hill landscape is less than GMO. That is, GMO has a better converging capability than CMO in the hill landscape. Fig. 3 is the diagram showing the probabilities of escaping/converging abilities with respect to $\delta/\Delta r$ for Eqs. (1) – (4).

As we know, Fig. 1 shows that a larger valley width δ or a smaller mutation step size Δr makes more difficult in jumping over the valley, and the quantitative analyses of escaping local optimum can be obtained from Fig. 3. With the same valley width

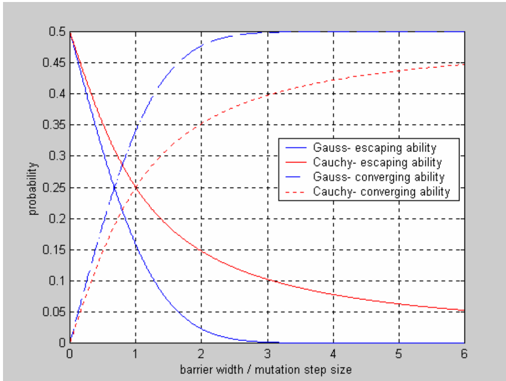


Fig. 3. The diagram showing the probabilities of escaping/converging abilities with respect to $\delta/_{\Delta r}$ for four equations.

(δ) and mutation step size (Δr), CMO has a better escaping ability than GMO (shown in Fig. 3). On the other hand, Fig. 2 shows that a smaller hillside width (δ) or a larger mutation step size (Δr) leads to more difficult in performing the fine-tuning to climb up the hilltop in the hill landscape. The quantitative analyses of local convergence can be gained from Fig. 3. It reveals that GMO is good in converging to local optima for the hill landscape than CMO with the same valley width (δ)) and mutation step size (Δr).

3. Simulation Results

Although different types of mutation make different probabilities to escape or search an adaptive landscape, the operation of an evolutionary algorithm does not depend on the mutation only. The evolution is an emergent behavior and consists of reproduction, mutation and selection. Especially, the selection dominates the convergence of the population and the goodness of an evolutionary algorithm. Lots of researchers proposed several excellent techniques to overcome the phenomenon of premature. Without losing generality, this research adopts a simple but powerful technique "ranking selection" to discriminate the merits between Gaussian and Cauchy mutations.

Two benchmark functions are selected in this study for representing two typical searching landscapes. They are:

$$\begin{aligned} f_1: f(x,y) &= 20 - 2.0 \times \exp(2.0) \\ &+ 20.0 \times \exp\left(-0.5 \cdot \sqrt{0.5\left(x^2+y^2\right)}\right) \\ &+ 2.0 \times \exp(2.0 \times 0.5(\cos(\pi \cdot x) + \cos(\pi \cdot y))) \qquad (5) \end{aligned}$$

f_1 is called Ackey function [8, p.43], and the ranges of x and y are both set from -5 to 5 (i.e. $x, y \in [-5, 5]$). Fig. 4 shows the diagram of Ackey function f_1.

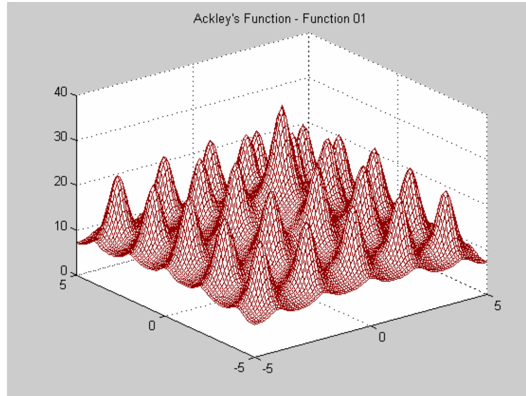


Fig. 4. The diagram of Ackey function

The second selected nonlinear multimodal function is f_2.

$$\begin{aligned} f_2: f(\bar{x}) &= 0.5 - \frac{\sin^2\sqrt{\sum x_i^2} - 0.5}{(1+0.001\cdot(\sum x_i^2))^2} \\ &+ 5\cdot\exp\left(-\frac{\sum x_i^2}{0.002}\right) \qquad (6) \end{aligned}$$

The original version was proposed by Schaffer [9], then Lan [10] modified it by adding an error term "$\left(5\cdot\exp\left[-\left[\frac{\sum x_i^2}{0.002}\right]\right]\right)$" to form a new landscape with a narrow -sharp- high beam. This new function is called the modified Schaffer function (Fig. 5). The ranges of x_i are set from -5 to 5.

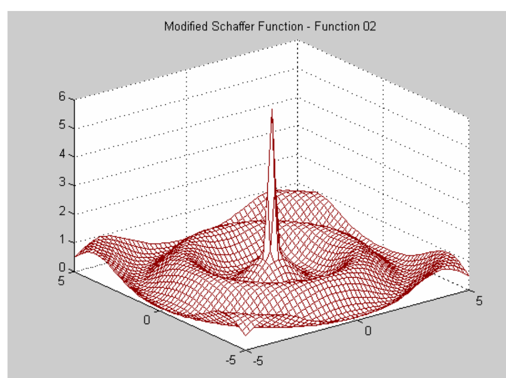
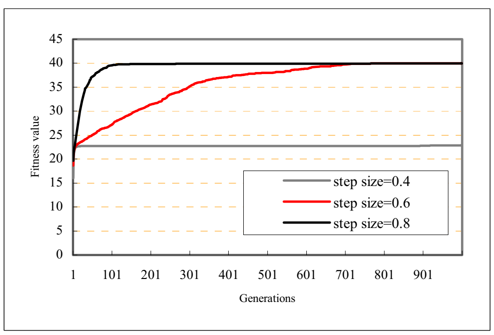


Fig. 5. The diagram of modified Schaffer function

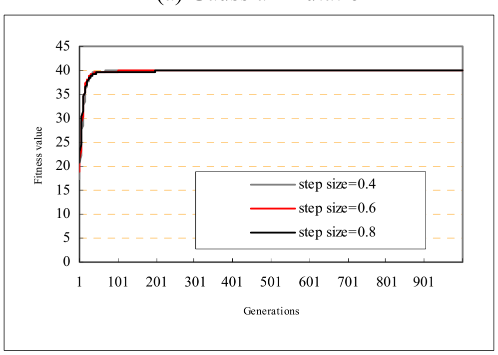
According to the concerns of section II, the performances of EAs adopting Gaussian or Cauchy mutations are investigated. Firstly, the local-distribution technique [10] is adopted to limit the initial population into a small region which is far from the global optimum, and thus the ability for driving the population to the global optimum can be explored. Secondly, the "ranking selection" is adopted for all EAs in order to compare which type of mutation possesses a better searching ability. Thirdly, the number of individuals among population is 20 and the evolutionary generations are set 1000. In addition, each evolutionary technique should run 100 epochs to compute its average to be the experimental result. The simulations of different mutation types and mutation step sizes are conducted in the following experiments.

Experiment 1:

For the benchmark function f_1, the initial distribution of population is set as $[-5,-4.95]^2$, and experimental results of two different mutations are shown in Fig. 6. In experiment 1, while the mutation step size is relatively small ($\Delta r = 0.4$), adopting Gaussian mutation will lead population to converge prematurely because Gaussian distribution has a smaller deviation. In addition, the increase of mutation step size leads the population to escape local optima and then tends to the global optimum (shown in Fig. 6(a)). Due to the larger variance of Cauchy distribution, even a small mutation step size ($\Delta r = 0.4$) is adopted, EAs still force the individuals to escape local optima and tend to the global optimum. Based on such phenomenon, adopting Cauchy mutation is an excellent choice because it has better capability of escaping local optimum.



(a) Gaussian mutation

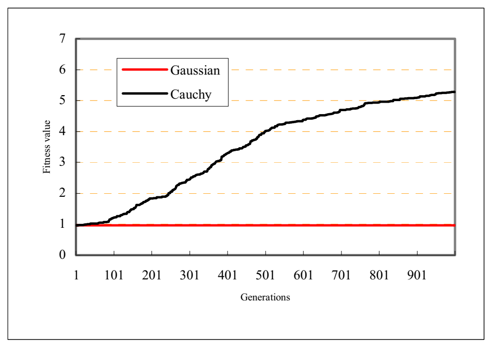


(b) Cauchy Mutation

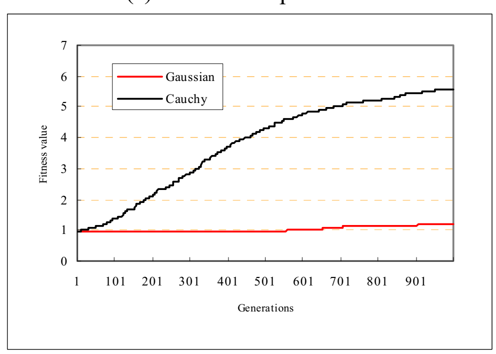
Fig. 6. The Gaussian/Cauchy mutation techniques for Ackey bechmark function

Experiment 2:

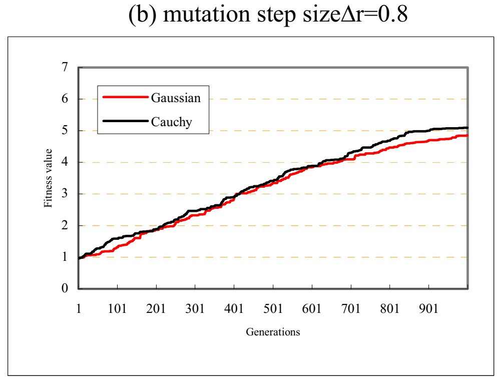
Benchmark function f_2 (Fig. 5) is applied to test the escaping/converging capabilities of the Gaussian and Cauchy mutations. The global optimum of f_2 is 6 and locates at the top of central beam in the modified Schaffer function. Like experiment 1, the initial distribution of the population is set as $[-5,-4.95]^2$. The experimental results (Figs. 7(a) & (b)) reveal that the Gaussian mutation cannot lead population to escape local optima because of a small mutation step size. The valley width is around 3.33 (obtained from Fig. 5), the proportions ($\delta/\Delta r$) of valley width to different mutation step sizes of Fig. 7(a) and (b) are 3.33/0.4=8.33 and 3.33/0.8=4.16 respectively, then the probabilities of escaping local optima for those mutation step sizes approximate zero (obtained from Fig. 3). On the other hand, Cauchy mutation has more chances to drive the population to escape local optima comparing to Gaussian mutation because the probabilities of escaping local optima are approximately 0.04 and 0.075 (obtained from Fig. 3).



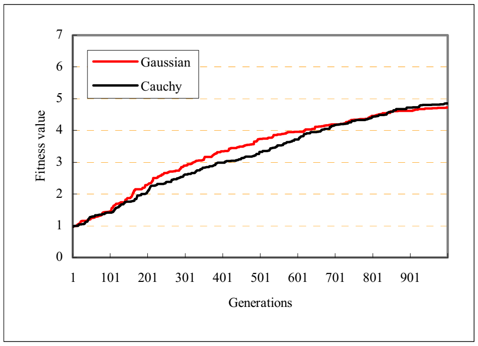
(a) mutation step size Δr=0.4



(b) mutation step sizeΔr=0.8



(c) mutation step size Δr=1.4



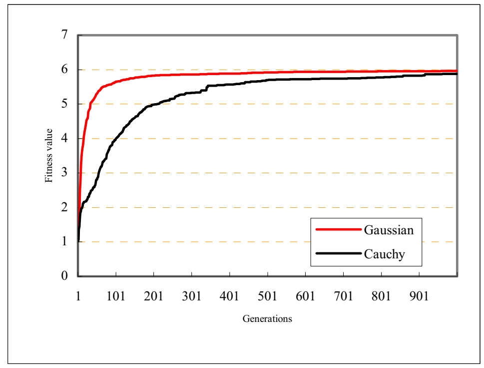
(d) mutation step size Δr=1.9

Fig. 7. The Gaussian/Cauchy mutation techniques for modified Schaffer bechmark function with initial distribution (-5.0 ~ -4.5)

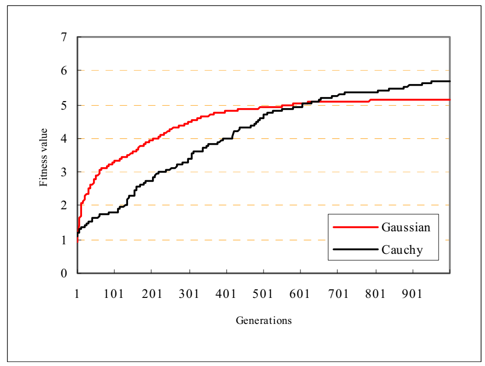
By a smaller mutation step size, EAs with Gaussian mutation are always trapped in the local optimum. Therefore, the increase of mutation step size will help the improvement of such premature situation. In Figs. 7(c) & (d), the proportions of the valley width to different mutation step sizes ($\delta/\Delta r$) are 3.33/1.4=2.37 and 3.33/1.9=1.75 respectively, thus the better escaping probabilities, 0.009 and 0.04, are obtained. These probabilities (0.009 and 0.04) are sufficient to escape the attraction of local optima after multi-generation evolutions. In fact, this experiment reveals while the mutation step size is over 1.4 (Gaussian mutation) or 0.4 (Cauchy mutation), EAs have excellent performance in leading the population to escape local optima for the modified Schaffer function.

Regardless of adopting Gaussian or Cauchy mutation, EAs also cannot precisely reach the global optimum with a larger mutation step size, and they only arrive at the hillside of the central beam where the global optimum is located. Since the mutation step size becomes larger, EAs raise their probabilities to escape local optima and jump up the hillside of the central beam. However, the larger mutation step size Δr will reduce the fine-tuning capability to achieve the peak of the hill. Obviously, Gaussian and Cauchy mutations do not have a significant contribution on searching the neighborhood of the hillside of a narrow beam if the larger mutation step size is adopted (shown in Figs. 7(c) & (d)).

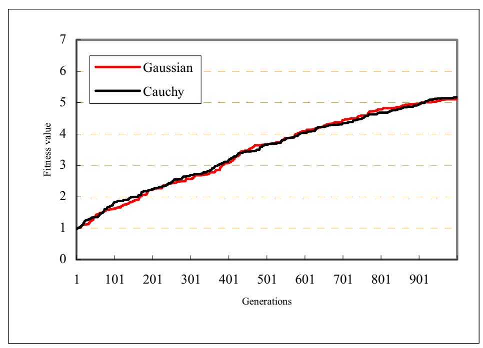
In order to explore the capability of local convergence, the initial distributed region of population is rearranged as $[-0.5,-0.4]^2$ (this region is located at the hillside of central beam). The simulation results (shown in Figs. 8(a) & (b)) reveal that Gaussian mutation possesses a faster capability of local convergence on searching the neighborhood of a narrow-beam landscape while the mutation step size is small. .



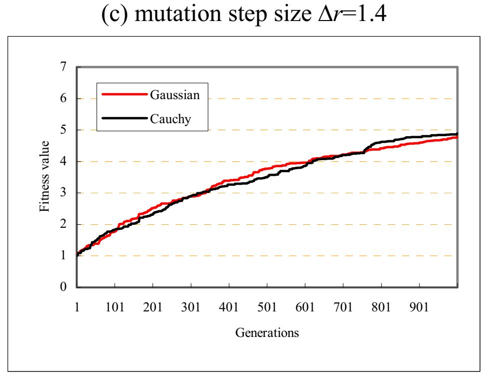
(a) mutation step size Δr=0.4



(b) mutation step size Δr=0.8



(c) mutation step size Δr=1.4



(d) mutation step size Δr=1.9

Fig. 8. The Gaussian/Cauchy mutation techniques for modified Schaffer bechmark function with initial distribution (-0.5 ~ -0.4)

The measurement of hillside width from Fig. 5 is around 0.33, and the proportions of the valley width to different mutation step sizes ($\delta/\Delta r$) are 0.826 (0.33/0.4=0.826) and 0.413 (0.33/0.8=0.413) respectively in Figs. 8(a) & (b). The probabilities of local convergence for Gaussian mutation are 0.29 and 0.16 approximately (calculated by Fig. 3). On the other hand, the probabilities of local convergence for Cauchy mutation are approximately 0.21 and 0.12. For the larger mutation step sizes (Figs. 8(c) & (d)), the proportions of the hillside width to different mutation step sizes ($\delta/\Delta r$) are 0.235 (0.33/1.4=0.235) and 0.174 (0.33/1.9=0.174) respectively.

For EAs with Gaussian or Cauchy mutation, there are no significant differences in local convergence if a larger mutation step size is adopted. For a small mutation step size, the EAs with Gaussian mutation converge little more faster than that with Cauchy. But through multi-generation evolutions, the predominance will become unobvious.

4. Conclusions

Based on the experimental results, the following findings are revealed. Firstly, a larger mutation step size can lead population to escape local optima and tend towards the global optimum. In addition, a smaller mutation step size can finely adjust population for suitable to achieve the global optimum. Secondly, while the proportion of the valley width to mutation step size ($\delta/\Delta r$) is large enough (for example, larger than 3 in one dimensional case), the premature situation will happen by adopting the Gaussian mutation under the target for jumping over the valley; thus, the Cauchy mutation is preferable for jumping over the valley.

In simplification, this paper mainly explores two basic abilities, local escape and local convergence, of mutations in evolutionary computation. Moreover, on the comparison of two popular mutation techniques (Gaussian and Cauchy), Cauchy mutation possesses more power in escaping local optima and converging to the global optimum. For local convergence, the Cauchy technique is nearly equal to the Gaussian after evolving more generations. Therefore, Cauchy mutation is suggested to avoid the dilemma problem of choosing a proper mutation step size and to achieve the acceptable performance for evolutionary computation.

In sum, to develop an ingenious selection technique for enhancing the diversity of population or to enlarge the variance of mutation by a fixed mutation step size instead of the efforts in determining mutation step size is the focal issue for future researches.

References

[1] G. Rudolph, "Local convergence rates of simple evolutionary algorithms with Cauchy mutations," *IEEE Trans. Evol. Compt.*, vol.1, no.1, pp.249-256, 1997.

[2] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," *IEEE Trans. Evol. Compt.*, vol.3, no.2, pp.82-102, 1999.

[3] K. Chellapilla, "Combininb mutation operators in evolutionary programming," *IEEE Trans. Evol. Compt.*, vol.2, no.3, pp.91-96, 1998.

[4] H. K. Birru, K. Chellapilla, and S. S. Rao, "Local search operators in fast evolutionary programming," in *Proc. Of IEEE Int. Conf. on Evol. Compt*., 1999, pp.1506-1513.

[5] N. G. Kim, J. M. Won, J. S. Lee and S. W. Kim, "Local convergence rate of evolutionary algorithm with combined mutation operator," in *Proc. Of IEEE Int. Conf. on Evol. Compt*., 2002, pp.261-266.

[6] Y. Liu and X. Yao, "How to control search step size in fast evolutionary programming," in *Proc. Of IEEE Int. Conf. on Evol. Compt*., 2002, pp.652-656.

[7] G. Rudolph, "Self-adaptive mutations may lead to premature convergence," *IEEE Trans. Evol. Compt.*, vol.5, no.4, pp.410-414, 2001.

[8] M. Gen and R. Cheng, *Genetic Algorithms & Engineering* Design, John Wiley &Sons. Inc., 1997.

[9] J. C. Potts, J. D. Giddens, and S. B. Yadav, "The development and evaluation of an improved genetic algorithm based on migration and artificial selection," *IEEE Trans. Syst. Man, and Cybern.*, vol.24, no.1, pp.73-86, 1994.

[10] K. T. Lan, *Analysis and Application of Competing Evolutionary Algorithms to Nonlinear Dynamic Systems*, Ph. D. thesis, Electrical Engineering Department, National Taiwan University of Science and Technology, Taipei, Taiwan, 2000.