Genetic Algorithms with a Robust Solution Searching Scheme

Shigeyoshi Tsutsui, *Member, IEEE*, and Ashish Ghosh

Abstract---A large fraction of studies on GAs emphasize finding a globally optimal solution. Some other investigations have also been made for detecting multiple solutions. If a global optimal solution is very sensitive to noise or perturbations in the environment then there may be cases where it is not good to use this solution. In this paper, we propose a new scheme which extends the application of GAs to domains that require the discovery of robust solutions. Perturbations are given to the phenotypic features while evaluating the functional value of individuals, thereby reducing the chance of selecting sharp peaks (i.e., brittle solutions). A mathematical model for this scheme is also developed. Guidelines to determine the amount of perturbation to be added is given. We also suggest a scheme for detecting multiple robust solutions. The effectiveness of the scheme is demonstrated by solving different one- and two-dimensional functions having broad and sharp peaks.

I. MOTIVATION

Over the years, genetic algorithms (GAs) have proven useful in varieties of search and optimization problems [1, 2]. There are many theoretical and empirical studies that investigate or present ways to improve the performance of conventional GAs for difficult function optimization problems such as those posed by multimodal and deceptive functions. These include the CHC of Eshelman [3], the messy GAs of Goldberg *et al*. [4], the delta coding of Mathias and Whitley [5], the forking GAs of Tsutsui et al. [6], and the GAs incorporating ancestors'

influence of De et al. [7]. The main emphases of these efforts was placed on finding a global optimal solution. Some other investigations emphasize finding multiple solutions (peaks) including local optima. These include the crowding method of De Jong [8], the sharing scheme of Goldberg and Richardson [9], Deb and Goldberg [10], Horn and Nafpliotis [11], the deterministic crowding method of Mahfoud [12], and the sequential niche technique of Beasley et al. [13].

If a solution obtained by some search technique is very sensitive to small perturbations of its parameter values, it may not be good to use this solution in certain situations. For example, consider the problem of designing the optimal parameter values of a process control plant. Suppose, by some technique, we determine the parameter values which yield very high performance from the plant. Now if the activity of the plant changes heavily due to small variation of the parameter values, then it is very risky to use such a parameter set for the plant, because in practice some noise will always be involved with the parameter values. A similar situation is also observed in robotics where motion planning must be done using uncertain and noisy sensory feedback [14]. Let us consider another case. If the performance of a product is highly sensitive to the precision of its parts, then the product will be very difficult or costly to produce by machines because each and every machine has limited capability to handle precision. Thus in many optimization tasks, there is a need to determine solutions whose value will not change much due to small variation of the parameter values.

We describe this type of solution as *robust*. This paper provides the basic concept of a new technique which extends the application of GAs to domains that require identification of robust solutions. We call this new technique *GAs/RS³: GAs with a robust solution searching scheme*. A precursor of this

S. Tsutsui is with the Department of Management and Information Science, Hannan University, 5-4-33 Amamihigashi, Matsubara, Osaka 580, Japan
Ashish Ghosh is with the Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta 700035, India.
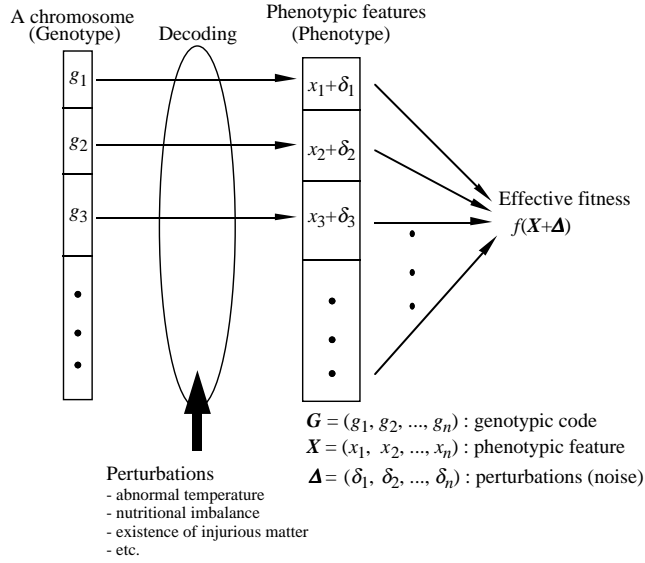
Fig. 1. A fitness evaluation model in a natural individual. The genotype undergoes a noisy decoding and interaction with the environment to generate a phenotype. Note this model is greatly simplified. The real world is much more complex due to the effects of pleiotropy and polygeny. Pleiotropy is the effect where a single gene affects multiple phenotypic characters. Polygeny is the effect where a single phenotypic character is affected by multiple genes [16].

line of research is found in Sebald and Fogel [15], where fault tolerant neural networks for pattern classification were designed by explicitly causing random faults in the networks during training using evolutionary programming. In the GAs/RS³ perturbations (noise) are added to the phenotypic parameter values while evaluating the functional value or fitness of individuals. Simulation results on various functions having broad and sharp peaks show the potential of the proposed concept.

The remainder of this article is organized as follows. First the basic model of the GAs/RS³ is outlined in Section II. Section III describes a simple mathematical model of the GAs/RS³ for one-dimensional problems. Empirical results are analyzed in Section IV. An attempt to combine the GAs/RS³ with a sharing scheme is made in Section V. The model of the GAs/RS³ is extended to multi-dimension functions in Section VI. Finally, concluding remarks and future scope of work are given in Section VII.

II. DESCRIPTION OF THE GAs/RS³

In nature, the phenotypic expression of an organism is determined in part by decoding the genotypic code of genes

```
gen = 0;
Pop(gen) = randomly initialized population {G¹, G², ..., Gᴺ};
Decode each genotype Gⁱ to produce the corresponding phenotype Xⁱ;
Add noise and set Yⁱ = Xⁱ + Δ;
Evaluate f(Yⁱ) for all i in Pop(gen);
while( termination condition == false ){
    gen += 1;
    Select Pop(gen) from Pop(gen - 1) based on f(Yⁱ);
    Apply genetic operators to {G¹, G², ..., Gᴺ} in Pop(gen);
    Decode each Gⁱ to produce Xⁱ;
    Set Yⁱ = Xⁱ + Δ;
    Evaluate f(Yⁱ) for all i in Pop(gen);
}
Evaluate f(Xⁱ) for all i in Pop(gen);
```

Fig. 2. Schematic model of the GAs/RS³, N is the population size.

in the chromosomes. During this decoding process there may be some perturbations, for example, caused by an abnormal temperature, a nutritional imbalance, existence of injurious matter, etc. (see Fig. 1). Stated loosely, if the individual has low fitness due to these perturbed phenotypic features, then the individual will not survive to produce offspring. Thus individuals and reproductive populations having "good" genotypic material would become extinct if they were highly sensitive to perturbations of phenotypic features. On the other hand, in noisy environments, reproductive units which are robust to these perturbations would have a better chance of surviving. We develop the GAs/RS³ with an aim to locating robust solutions by using this sort of natural genetic metaphor. GAs/RS³ uses the effect of perturbation of the phenotypic parameters while evaluating the functional values of individuals.

Approaches which give consideration to the existence of noise in calculating the fitness values are discussed in [17 - 19]. These efforts are mainly directed towards studying *noisy fitness functions*, i.e., noise is added to the fitness function. If $X = (x_1, x_2, ..., x_m)$ is a phenotypic parameter vector, $f(X)$ the evaluation function and δ a scalar noise, then the fitness of the individual will be $f(X) + \delta$ [17 - 19]. On the other hand, when we aim to detect robust solutions, it can be understood from the natural phenomena that noise is added during the process of decoding the genotypic codes to the phenotypic parameters (Fig. 1). Hence, to add noise to the phenotypic parameter X, i.e., to use an evaluation function of the form $f(X+\Delta)$ appears reasonable, where $\Delta = (\delta_1, \delta_2,..., \delta_m)$ is a random

vector. The solutions thus determined are expected to be more robust against perturbations or noise having the appropriate tested distribution.

Let **G** be a genotypic string (or chromosome) which generates the phenotypic parameter **X**. Then the model of the GAs/RS³ becomes as shown in Fig. 2. Here, it should be noted that adding noise in the form $f(X+\Delta)$ may appear like a mutation operation on a real-valued coding, but actually it is operationally different from mutation, since it does not have any direct effect on individual strings. The perturbations are used only for judging the quality of a solution and for selection.

III. MATHEMATICAL MODEL

This section describes a simple mathematical model of the GAs/RS³. Although there are differences between high- and low-dimensional problems, we consider **X** to be one-dimensional to demonstrate the effect of adding noise more clearly and denote **X** and **Δ** by x and δ, respectively. Extension to the multidimensional case is discussed in Section VI.

A. Formulation of the Effective Evaluation Function for the GAs/RS3

Consider the schema theorem of the simple GA using a proportional payoff selection scheme and a single-point crossover [1]:

$$M(H, t+1) \geq M(H,t) \cdot \frac{f(H,t)}{\overline{f(t)}} \cdot$$
$$\cdot \left[1 - p_c \frac{d(H)}{L-1} - o(H)p_m \right] \qquad (1)$$

where H represents a schema in population $P(t)$, $f(H,t)$ is the average fitness of representatives of H in $P(t)$, $\overline{f(t)}$ denotes the average fitness of the individuals in $P(t)$, p_c is the crossover probability, p_m is the mutation probability, $d(H)$ is the defining length of H, $o(H)$ is the order of H, L is the string length, and $M(H,t)$ is the expected number of representatives of schema H at generation t. Here the population size N is assumed to be large. When $N \rightarrow \infty$, $\overline{f(t)}$, the average fitness of the individuals in $P(t)$, can be described as

$$\overline{f(t)} = \sum_{i=1}^{N} f(x^i)/N$$
$$= \int_x f(x) \cdot p(x,t)dx, \qquad (2)$$

where $p(x,t)$ provides the distribution of parameter x in the population. In a similar manner, $f(H,t)$ can be described as

$$f(H,t) = \int_x f(x) \cdot p(x,H,t)dx, \qquad (3)$$

where $p(x,H,t)$ provides the distribution of parameter x in schema H in the population.

Now, consider the schema theorem corresponding to the GAs/RS³ where fitness is evaluated in the form $f(x+\delta)$.

$$M(H, t+1) \geq M(H,t) \cdot \frac{f'(H,t)}{\overline{f'(t)}} \cdot$$
$$\cdot \left[1 - p_c \frac{d(H)}{L-1} - o(H)p_m \right]. \qquad (4)$$

Here, $\overline{f'(t)}$ can also be described as in (2),

$$\overline{f'(t)} = \sum_{i=1}^{N} f(x^i + \delta^i)/N$$
$$= \int_x \int_{-\infty}^{\infty} f(x+\delta) \cdot p(x,t) \cdot q(\delta)d\delta dx$$
$$= \int_x \left[\int_{-\infty}^{\infty} f(x+\delta) \cdot q(\delta)d\delta \right] \cdot p(x,t)dx \qquad (5)$$
$$= \int_x F(x) \cdot p(x,t)dx,$$

where it is assumed that x^i and δ^i are mutually independent, $q(\delta)$ is the continuous density function of δ^i having defined mean value, and $F(x)$ is set as

$$F(x) \equiv \int_{-\infty}^{\infty} f(x+\delta) \cdot q(\delta)d\delta. \qquad (6)$$

Similarly, $f'(H,t)$ can be represented as

$$f'(H,t) = \int_x F(x) \cdot p(x,H,t)dx. \qquad (7)$$

Comparing (2) (the average fitness of the whole population for simple GAs) with (5) (the effective average fitness of the whole population for GAs/RS³), and (3) (the average fitness of representatives of H for simple GAs) with (7) (the average fitness of representatives of H for GAs/RS³), we can confirm that $F(x)$ corresponds to $f(x)$.

From the above derivation, we may conclude, for $N \rightarrow \infty$, that the average number of instances of each schema in the GAs/RS³ increases or decreases depending on $F(x)$ instead of $f(x)$, and thus the solution under GAs/RS³ evolve

so as to maximize $F(x)$ instead of maximizing the actual objective function $f(x)$. Hereafter we call $F(x)$ the *effective evaluation function* of $f(x)$ in the GAs/RS³. As is easily understood, $F(x)$ is equivalent to the expected value of $f(x)$ over $x+\delta$. If we assume $q(\delta)$ to be symmetric, i.e., $q(\delta) = q(-\delta)$, then $F(x)$ can be rewritten as

$$F(x) = \int_{-\infty}^{\infty} q(x - y) \cdot f(y)dy . \tag{8}$$

Thus the effective evaluation function for the GAs/RS³ can be formulated. In practice, the population size must be finite. If the population is sufficiently large, then this may yield approximate characteristics as indicated in (8). In general, the sufficiency of the population size will depend on the distribution of the noise. Hereafter, (8) is used for further discussion.

Here we must note that in practice $f(x)$ is unknown, and thus $F(x)$ cannot be calculated using mathematical formulas. So as to analyze the effect of noise mathematically, we have assumed that $f(x)$ is known completely. For computer implementation this information is not required.

B. Relationship between noise parameter and peak reduction

In this subsection, the appropriate size (σ) of a Gaussian noise to be added is estimated depending on the actual function when we assume the width of the sharp peak to be known. For the sake of simplicity, the peaks of functions are represented by rectangles, and for illustration define one rectangle having height h ($h > 0$) and width $2w$ ($w > 0$) as follows:

$$f(x) = \begin{cases} h: & -w \leq x \leq w \\ 0: & \text{otherwise} . \end{cases} \tag{9}$$

The peak of this function is spread from -w to w. The effective evaluation function corresponding to this is obtained from (8) as

$$F(x) = h\int_{-w}^{w} q(x - y)dy$$
$$= h\left[\Phi\left(\frac{x + w}{\sigma}\right) - \Phi\left(\frac{x - w}{\sigma}\right)\right], \tag{10}$$

where $\Phi(x)$ is the distribution function of the standard normal distribution defined by

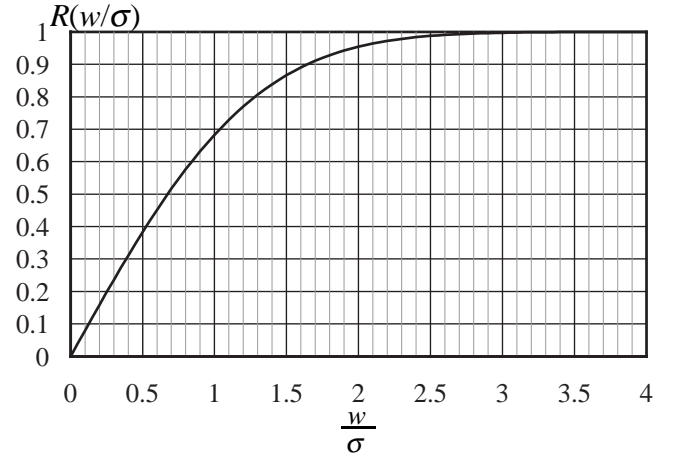$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy . \tag{11}$$

Fig. 3. Relationship between the reduction factor $R(w/\sigma)$ and w/σ, $2w$ is the width of a peak, σ is the standard deviation of the zero mean Gaussian noise. As the ratio w/σ increases, the reduction factor converge asymptotically to 1.0 .

By setting the derivative of the function $F(x)$ to zero, the peak point is obtained at $x = 0$ and the peak value, max $F(x)$, is obtained as

$$\max F(x) = F(0)$$
$$= h\left[2\Phi\left(\frac{w}{\sigma}\right) - 1\right]$$
$$= h \times R\left(\frac{w}{\sigma}\right). \tag{12}$$

Here, $R(w/\sigma)$ is the *reduction factor*. Fig. 3 plots the reduction factor versus w/σ. Fig. 4 shows the relationship between function $f(x)$ and function $F(x)$ for $w/\sigma = 4.0$, 2.0, 1.0, 0.5 and 0.25. Figs. 3 and 4 confirm that addition of Gaussian noise to phenotypic parameters reduces the effective height of the peaks and the effect is greater as the value of σ become larger. This can also be supported by a simple Fourier analysis given in Appendix A.

Amount of noise to be added (σ) can be estimated given the width ($2w$) and the reduction factor ($R(w/\sigma)$) of the peak. Let $2w$ be the width of a sharp peak. If σ takes values in the range [$2w$, $4w$], then w/σ has values in the range [0.5, 0.25] and the reduction factor is between 0.197 and 0.383 (see Fig. 3). Thus, σ can be roughly estimated when the allowable width of the sharp peaks and their reduction factors are given.
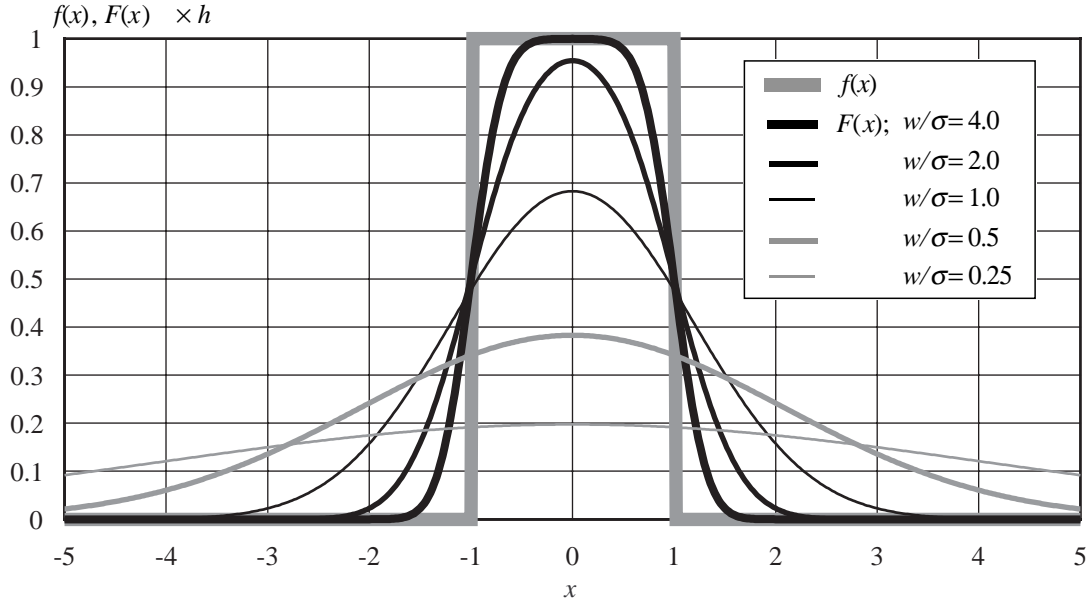
Fig. 4. Relationship between the actual function $f(x)$ and the effective evaluation function $F(x)$. The addition of more variable noise reduces the effective height of the peak.

IV. EMPIRICAL STUDY

To introduce the utility of the GAs/RS[3], we studied two one-dimensional functions. A simple GA (hereafter we refer to it as SGA) was used with stochastic universal sampling [20]. GA parameters were kept constant for all the simulations with mutation probability $p_m = 0.006$, crossover probability $p_c = 0.6$, population size $N = 100$, maximum number of function evaluations = 5,000, and phenotypic parameter x is encoded by a 30-bit string (Gray coded). These values follow many standard implementations in the literature. We performed 30 simulations for each experiment with randomly initializing the population.

1) Function f_a :   Consider a function f_a (Fig. 5a), which has one broad peak and one sharp peak, defined as

$$f_a = \begin{cases} 1: & -1 \le x \\ 2: & 1.5 \le x \le 1. \\ 0: & \text{otherwise} . \end{cases} \tag{13}$$

The parameter range is $-3 \le x \le 3$. We took $\sigma = 0.4$ ($w/\sigma = 0.25$). Fig. 5b shows the convergence process of the mean value of parameter x in the population with function evaluations. The SGA converged at $x = 1.6$, the center of the

highest peak. In contrast, the GAs/RS[3] converged to the center ($x = 0$) of the broad peak. It can be observed from Fig. 5b that at the initial stage of searching, the population moved towards the highest peak, but then quickly turned attention to the broad peak. This phenomenon may be explained as follows. In the early stage of evolution, the effect of adding noise is low since the diversity of the population is greater. As the search process proceeds, noise gradually affects and moves the population. Fig. 6 shows a typical distribution of the individuals in the population after 5,000 function evaluations for both the SGA and the GAs/RS[3]. In each case, the population converged to reside in a single peak.

2) Function f_b : Function f_b (Fig. 7a) has five unequal peaks in the range

$0 \le x \le 1$, and is a variant of the function used in [10]. It is defined as

$$f_b = \begin{cases} e^{-2\ln 2 \left(\frac{x-0.1}{0.8}\right)^2} |\sin(5\pi x)|^{0.5} & : 0.4 < x \le 0.6 \\ e^{-2\ln 2 \left(\frac{x-0.1}{0.8}\right)^2} \sin^6(5\pi x) & : \text{otherwise}. \end{cases} \tag{14}$$

As shown in Fig. 7a, the global optimum is located at $x = 0.1$ with the functional value 1.0. There are four sharp peaks. The third peak is broad compared to others and is located at $x =$
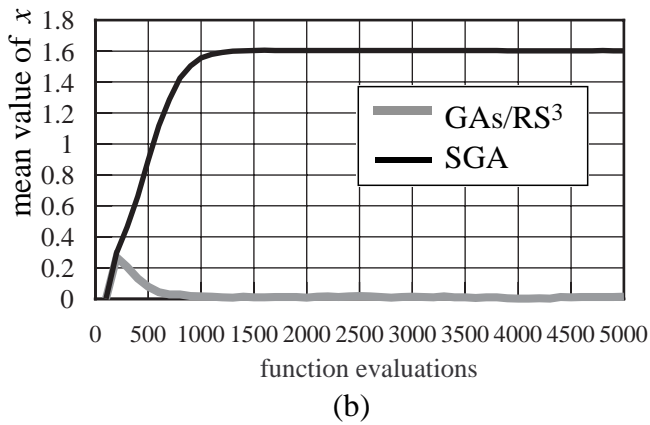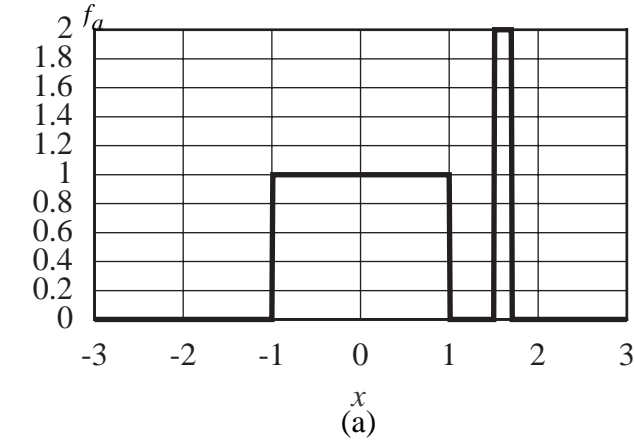
Fig. 5. Convergence process for function f_a. (a) shows the original function f_a and   (b) shows the variation of mean (over the population) value of x with function evaluations.

Fig. 6.   A typical distribution of the individuals in function f_a after 5,000 function evaluations for SGA (a) and for the GAs/RS³ (b)

0.486 with functional value 0.715. The effective width of the four sharp peaks can be estimated as follows. Although these four peaks have unequal heights defined by the exponential coefficient in (14), its main shape can be approximate by $h \times \sin^6(5\pi x)$, where h is the peak height. Express the effective peak by a rectangle with width $2w$ and height h. Then w may be estimated by letting the area of one bell shape (covered by $h \times \sin^6(5\pi x)$ and x axis) be equal to the area of the rectangle. Thus for the first bell shape the width is obtained as

$$h \times 2w \approx h \times \int_0^{0.2} \sin^6(5\pi x)\, dx = h \times \frac{1}{16}$$

$$\Rightarrow \quad w \approx 1/32 \,. \tag{15}$$

Since the functional form is the same for all other bell shapes, the effective width of all of these are the same and is 1/32. Referring to Fig. 3, w/σ can be chosen in (0, 0.65] so as to
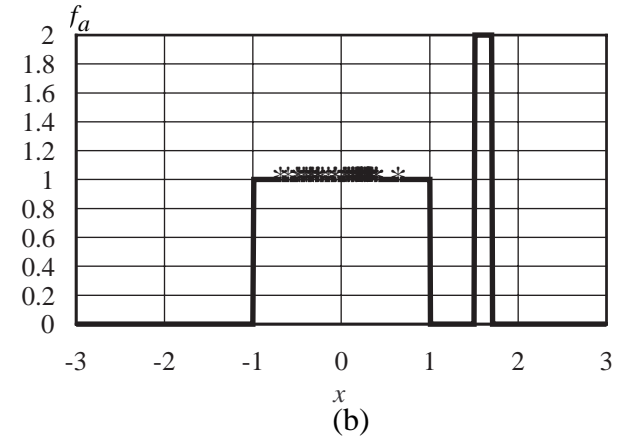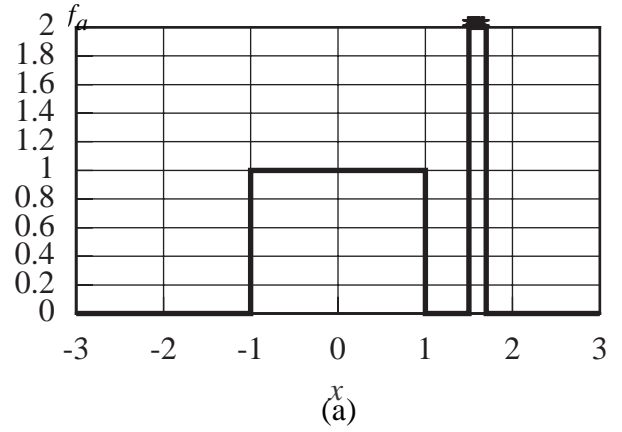
reduce the effective functional value by more than 50%. We chose $w/\sigma = 0.5$. Thus, $\sigma = 0.0625$ ($\sigma = w/0.5 = 2 \times 1/32$) was used (from Fig. 3, we can see the value of reduction factor $R(w/\sigma)$ for the sharp peaks is about 0.4). In this context it may be mentioned that we estimated the width of the effective peaks just for illustration. In practical problems, the maximum allowable width of a peak to be considered as a sharp peak must be given.

Fig. 7b shows a convergence process of the mean value of parameter x in the population with trials. The SGA converged at $x = 0.1$, the center of the highest peak. The GAs/RS³ converged to the robust peak ($x = 0.486$) zone. As before, we can observe from Fig. 7b that the population moved initially towards the highest peak but then approached the broad peak. Fig. 8 shows a typical distribution of the
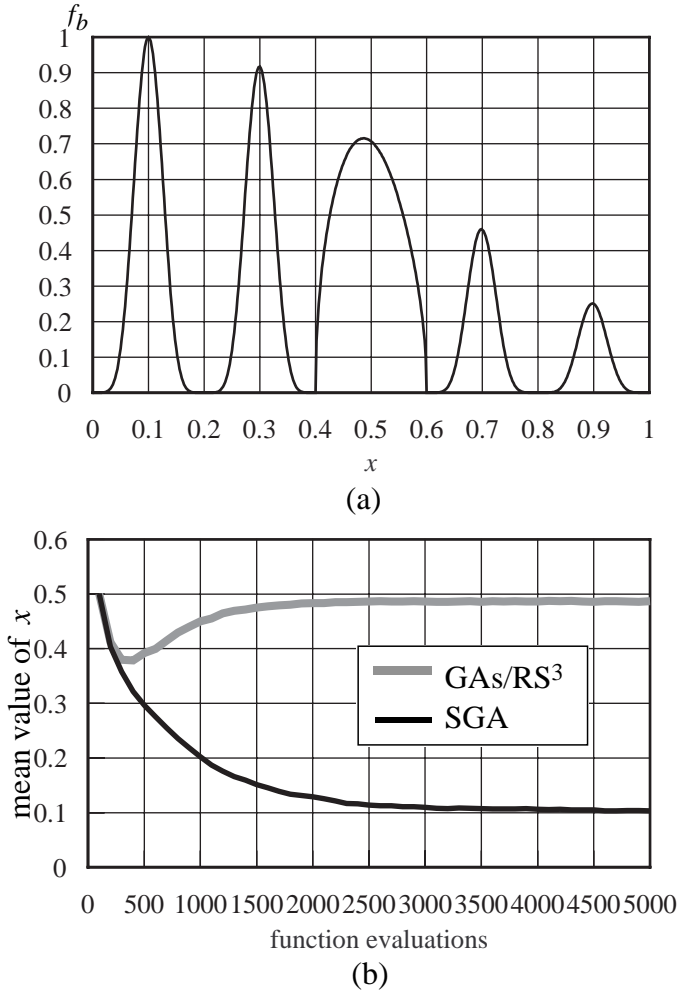
(a)

(b)

Fig. 7. A typical convergence process for function f_b. (a) shows the original function f_b and (b) shows the variation of mean (over the population) value of x with function evaluations.

(a)

(b)

Fig. 8. A typical distribution of the individuals in function f_b after 5,000 function evaluations for the SGA (a) and for the GAs/RS³ (b).

individuals in the population after 5,000 function evaluations for both the SGA and the GAs/RS³. From the figure, we can say that for the GAs/RS³ the population practically converged to the broad peak, as all but one individual resided within its bounding width.

V. DETECTION OF MULTIPLE ROBUST SOLUTIONS

Even if there is more than one robust solution, the proposed GAs/RS³ would normally find only one. This sort of problem 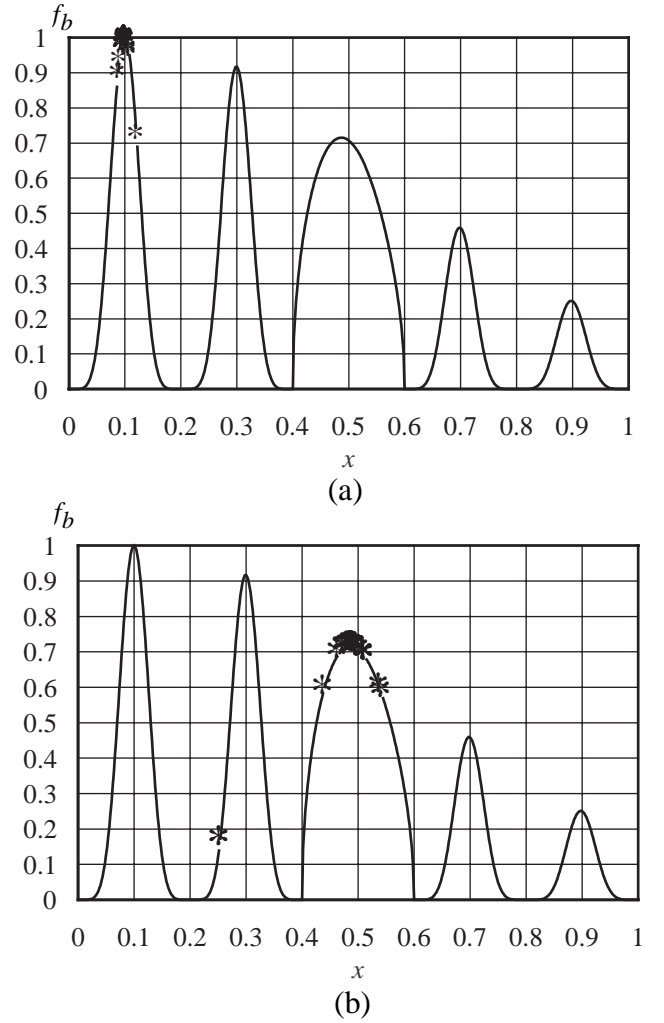with finite populations is known in evolutionary computation as genetic drift [8, 10]. One approach to find more than one robust solution is to combine the GAs/RS³ with the sharing scheme of Goldberg and Richardson [9] using a sharing parameter σ_{share} (> 0) to control the extent of sharing and a distance metric involving an index

α > 0. When the proximity of the individuals (distance) is defined in the decoded parameter space, it is called phenotypic sharing and when it is defined in the genotypic space (Hamming distance), it is termed genotypic sharing.

Here, we employed phenotypic sharing. Consider a function f_c (Fig. 9) which has two broad peaks and one sharp peak, defined as

$$f_c = \begin{cases} 1: -2.5 \le x \le -1.0 \text{ or } 0.0 \le x \le 1.5 \\ 2: 2.0 \le x \le 2.2 \\ 0: \text{otherwise}. \end{cases} \quad (16)$$

We used $\sigma_{share} = 1.0$ and $\alpha = 1.0$. With regard to noise size (i.e., standard deviation), we took $\sigma = 0.4$, the same value as used for function f_a in Section IV. Other parameters for the GAs/RS³ remained unchanged.

Fig. 9 shows a typical distribution of the individuals in the population after 5,000 trials by the SGA, the GAs/RS³, and the GAs/RS³ with the sharing scheme. The GAs/RS³ always converged to either of the two robust peaks. Fig. 9 (b) is an example when the population converged to the left robust peak. On the other hand, the GAs/RS³ with the sharing scheme converged the population to both broad peaks stably as shown in Fig. 9c.

VI. EXTENDING TO MULTIPLE DIMENSIONS

A. Calculation of reduction factor for multi-dimensions

It is of interest to determine how the calculation of the reduction factor scales up for higher-dimensional search spaces. First consider a two-dimensional evaluation function and represent it by $f(x_1, x_2)$. Then the effective evaluation function $F(x_1, x_2)$ of $f(x_1, x_2)$ for $N \rightarrow \infty$ can easily be formulated in a similar manner as (8) and is obtained as follows:

$$F(x_1, x_2) = \iint_{y_1 y_2} q_1(x_1 - y_1) q_2(x_2 - y_2) f(y_1, y_2) d y_1 d y_2 , \quad (17)$$

where, q_1 and q_2 are the density functions of noise added to the phenotypic parameters x_1 and x_2, respectively. Now we emulate a peak in two-dimensional search space by the following two-dimensional rectangular or box function having height h ($h > 0$), $2w_1$ ($w_1 > 0$) in the x_1 axis and width $2w_2$ ($w_2 > 0$) in the x_2 axis, which is a direct extension of the rectangular function represented by (9) in Section III.

$$f(x_1, x_2) = \begin{cases} h: -w_1 \le x_1 \le w_1 \wedge -w_2 \le x_2 \le w_2 \\ 0: \text{ otherwise}. \end{cases} \quad (18)$$
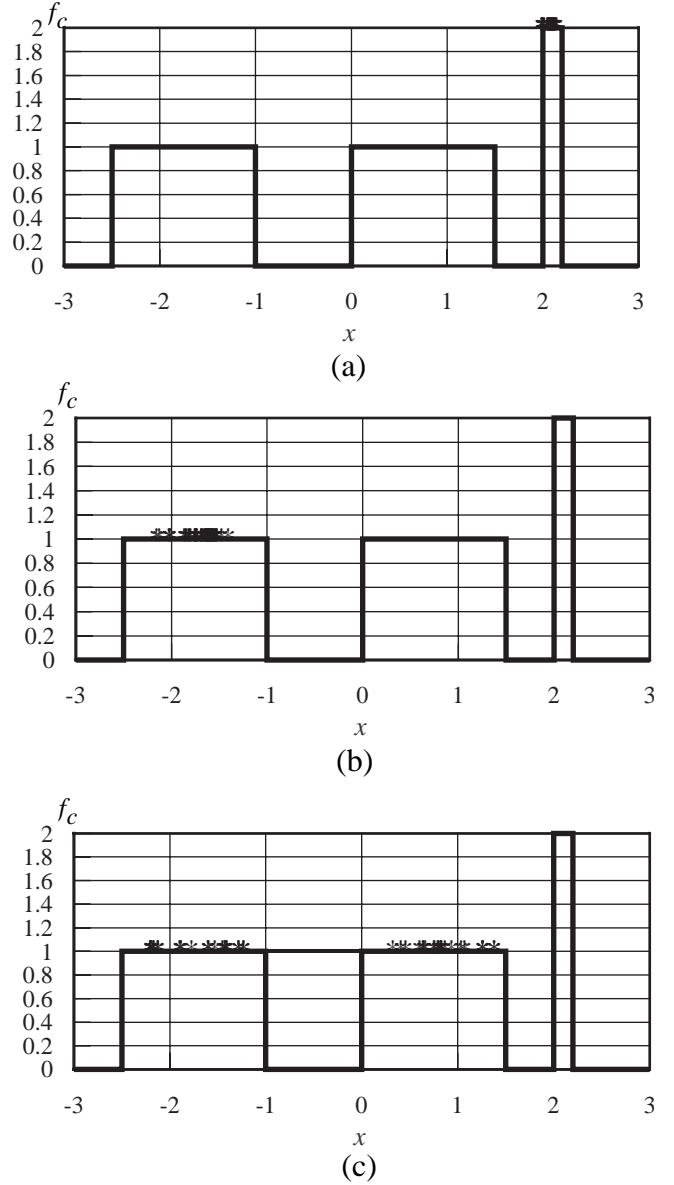

(a)

(b)

(c)

Fig. 9. A typical distribution of the individuals in function f_c after 5,000 function evaluations for the SGA (a), for the GAs/RS³ (c) and for the GAs/RS³ with sharing (c) Sharing allows the population to converge to multiple broad peaks.

As in Section III, again assume zero mean Gaussian noise for q_1 and q_2 with standard deviation σ_1 and σ_2, respectively. Then the effective evaluation function $F(x_1, x_2)$ of (18) is obtained
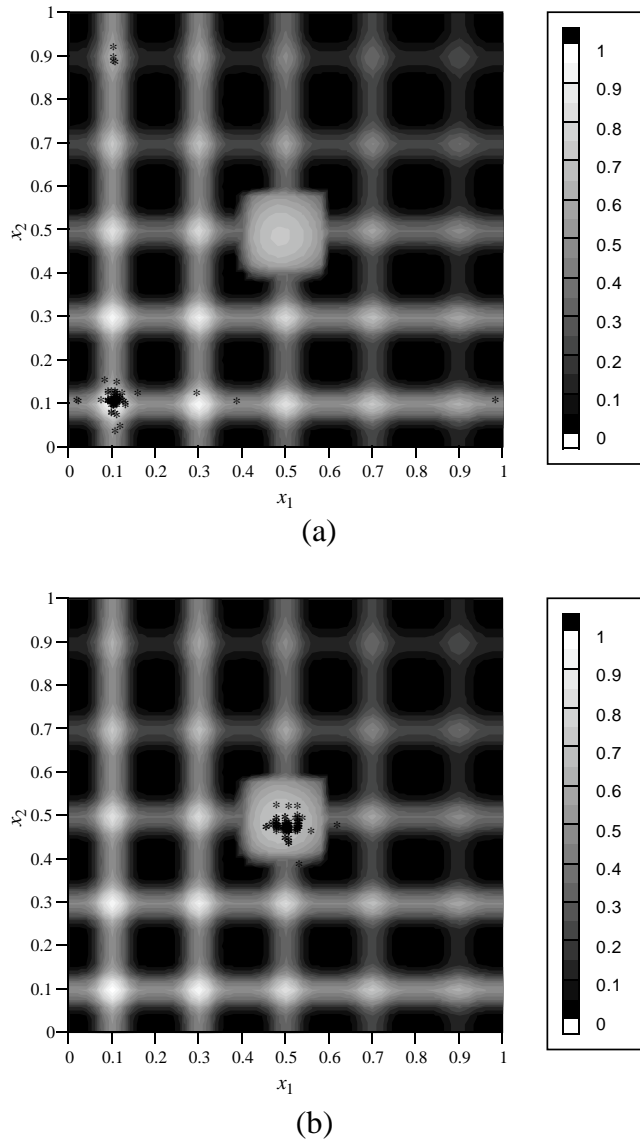
(a)

(b)

Fig. 10. A typical distribution of the individuals in function f_{b2} after 5,000 function evaluations for the SGA (a) and for the GAs/RS[3] (b).

in the same manner as in (10) as

$$F(x_1, x_2) = h \left[\Phi\left(\frac{x_1 + w_1}{\sigma_1} \right) - \Phi\left(\frac{x_1 - w_1}{\sigma_1} \right) \right]$$
$$\cdot \left[\Phi\left(\frac{x_2 + w_2}{\sigma_2} \right) - \Phi\left(\frac{x_2 - w_2}{\sigma_2} \right) \right]. \quad (19)$$

In a similar manner, when we approximate a peak of any function in an n- dimensional search space by an n-dimensional hyperbox function having height h ($h > 0$), width

$2w_i$ ($w_i > 0$) in the x_i axis, its effective function $F(X)$ is as follows:

$$F(X) = h \prod_{i=1}^{n} \left[\Phi\left(\frac{x_i + w_i}{\sigma_i} \right) - \Phi\left(\frac{x_i - w_i}{\sigma_i} \right) \right]. \quad (20)$$

where, σ_i is the standard deviation of the Gaussian noise added to the phenotypic parameter x_i. From (20) we can calculate R_n, the reduction factor for an n-dimensional search space, as follows:

$$R_n = \max F(X) / h$$
$$= \prod_{i=1}^{n} R\left(\frac{w_i}{\sigma_i} \right), \quad (21)$$

where

$$R\left(\frac{w_i}{\sigma_i} \right) = \left[2\Phi\left(\frac{w_i}{\sigma_i} \right) - 1 \right]. \quad (22)$$

The reduction factor R_n is the product of the reduction factor in each dimension i (see (20), (21), and (12)).

B. Experimental results on a multidimensional function

We used the following two-dimensional evaluation function f_{b2}, which is a direct extension of the function f_b (14), to observe the effect of GAs/RS[3] for a multidimensional function. This function has 25 peaks as shown in Fig. 10. The central peak is broad compared to the other 24 peaks and is located at $x_1 = x_2 = 0.486$ with functional value 0.715. The highest sharp peak is located at $x_1 = x_2 = 0.1$ with functional value 1.0. Here the noise size was $\sigma_1 = \sigma_2 = 0.0625$, which is the same as used earlier in the one-dimensional function f_b . The value of the reduction factor R_2 for the sharp peaks can be calculated from (21) as 0.16 (= 0.4 × 0.4).

Fig. 10 shows a typical distribution of the individuals in the population after 5,000 function evaluations by both the SGA and GAs/RS[3]. The SGA always converged to the highest peak as shown in Fig. 10a. In contrast, the GAs/RS[3] consistently moved the population to the broad peak as shown in Fig. 10b.

VII. SUMMARY

The basic concept of the GAs/RS[3] (genetic algorithms with a robust solution searching scheme), which extends the application of GAs to domains that require detection of robust solutions, was proposed in the present article. Perturbations

are given on the phenotypic features while evaluating the functional value of individuals. A mathematical model of GAs/RS[3] has been developed and described using simple one- and two-dimensional functions. A guideline to determine the amount of noise to be added has also been given. The effectiveness of the GAs/RS[3] was demonstrated by maximizing functions having broad and sharp peaks. By taking suitable noise sizes it has been shown that the GAs/RS[3] can converge the population to the robust solution against perturbation of the parameters. A consideration on combining the GAs/RS[3] with the sharing scheme was also made and it was found that this approach can be effective when we want to detect more than one robust solutions on different peaks.

Future work will focus on analyzing the behavior of GAs/RS[3] on more complex problems where many peaks interact, evaluating GAs/RS[3] on real-world problems, and extending the GAs/RS[3] for ordered representation based problems such as required in a scheduling system. So as to have a rough empirical estimate of N, various types of fitness evaluation techniques will also be adopted.

APPENDIX

A. Effect of Noise on the Height of Peaks

Here we give a simple Fourier analysis of the effective evaluation function $F(x)$ described by (8) to study its relationship with the original evaluation function $f(x)$ and noise density function $q(\delta)$. We recapitulate that high frequencies in the Fourier domain correspond to sharp peaks in the functional domain. Let $\hat{F}(\omega)$, $\hat{f}(\omega)$ and, $\hat{q}(\omega)$ be the Fourier transform of functions $F(x), f(x)$ and $q(x)$, respectively. Then, since (8) has the form of a convolution integral we get

$$\hat{F}(\omega) = \hat{q}(\omega) \cdot \hat{f}(\omega). \tag{23}$$

When we assume $q(\delta)$ to be a Gaussian noise N(0, σ), then,

$$\hat{q}(\omega) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\sigma^2 \omega^2}{2}}. \tag{24}$$

We see from above that $\hat{q}(\omega)$ decreases as ω increases. Thus, the magnitude of high frequency is reduced, i.e., addition of Gaussian noise to phenotypic parameters reduces the height of sharp peaks more. Also from (24) this effect is strengthened

as the value of σ increases. This results also support the analysis in Section III. B.

REFERENCES

[1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, pp. 28-33, 1989.

[2] Z. Michalewicz, *Genetic Algorithms + Data Structure = Evolution Programs*, 3rd, revised and extended edition, Springer-Verlag, Berlin Heidelberg, 1994.

[3] L. J. Eshelman, "The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination," *Foundations of Genetic Algorithms*, edited by G. J. E. Rawlins, Morgan Kaufmann, San Mateo, CA, pp. 265-283, 1991.

[4] D. E. Goldberg, K. Deb and B. Korb, " Messy Genetic Algorithms Revisited: Studies in Mixed Size and Scale," *Complex Systems,* no. 4, pp. 415-444 1990.

[5] K. Mathias and L. D. Whitley, "Changing Representations During Search: A Comparative Study of Delta Coding," *Evolutionary Computation*, vol. 2, no. 3, pp. 249-278, 1994.

[6] S. Tsutsui, Y. Fujimoto and A. Ghosh, "Forking GAs: GAs with Search Space Division Schemes," *Evolutionary Computation*, vol. 5, no. 1, pp. 61-80, 1997.

[7] S. De, A. Ghosh and S. K. Pal, "Fitness Evaluation in Genetic Algorithms with Ancestors' influence," *Genetic Algorithm for Pattern Recognition*, S. K. Pal and P. P. Wang (Eds), CRC Press, Boca Raton, pp. 1-23, 1996.

[8] K. A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*, A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Computer and Communication Sciences) in the University of Michigan, 1975.

[9] D. E. Goldberg and J. Richardson, "Genetic Algorithms with Sharing for Multimodal Function Optimization,"

Proc. of the 2nd Int. Conf. on Genetic Algorithms, J. J. Grefenstette (Ed.), Lawrence Erlbaum, Hillsdale, NJ, pp. 41-49, 1987.

[10] K. Deb and D. E. Goldberg, "An Investigation of Niche and Species Formation in Genetic Function Optimization," *Proc. of the 3rd Int. Conf. on Genetic Algorithms*, J. D. Schaffer (Ed.), Morgan Kaufmann, San Mateo, CA, pp. 42-50, 1989.

[11] J. Horn and N. Nafpliotis, "Multiobjective Optimization Using the Niched Pareto," *University of Illinois Urbana-Champaign IlliGAL Report* no. 93005, 1993.

[12] S. W. Mahfoud, "Crowding and preselection revisited," *Proc. of the 2nd Conf. on Parallel Problem Solving from Nature*, R. Männer and B. Manderick (Eds.), North-Holland, Amsterdam, pp. 27-34, 1992.

[13] D. Beasley, D. R. Bull, and R. R. Martin, "A Sequential Niche Technique for Multimodal Function Optimization," *Evolutionary Computation*, vol. 1, no. 2, pp. 101-125, 1993.

[14] S. J. Louis and G. Li, "Combining Robot Control Strategies Using Genetic Algorithms with Memory," *Evolutionary Programming VI,* P. J. Angeline, R. G. Reynolds, J. R. McDonnell and R. Eberhart (Eds.), Springer, Berlin, pp. 431-441, 1997.

[15] A. V. Sebald and D. B. Fogel, "Design of Fault Tolerant Neural Networks for Pattern Classification," *Proc. of the First Arnual Conf. on Evolutionary Programming*, D. B. Fogel and W. Atmar (Eds.) Evolutionary Programming Society, La Jolla, CA, pp. 90-99, 1992.

[16] D. B. Fogel, *Evolutionary Computation*, IEEE Press, Piscataway, NJ, pp. 38-43, 1995.

[17] J. M. Fitzpatrick and J. J. Grefenstette, "Genetic algorithms in noisy environments," *Machine Learning*, 3, pp. 101-120, 1988.

[18] U. Hammel and T. Bäck, "Evolution Strategies on Noisy Functions: How to Improve Convergence Properties," *Parallel Problem Solving from Nature - PPSN III*, Y. Davidor H.-P. Schwefel and R. Männer (Eds.), Springer, Berlin, pp. 159-168, 1994.

[19] B. L. Miller and D. E. Goldberg, "Genetic Algorithms, Selection Scheme, and the Varying Effect of Noise, *Evolutionary Computation,* vol. 4, no. 2 , pp. 113-131, 1996.

[20] J. E. .Baker, "Reducing Bias and Inefficiency in the Selection Algorithm," *Proc. of the 2nd Int. Conf. on Genetic Algorithms*, J. J. Grefenstette (Ed.), Lawrence Erlbaum, Hillsdale, NJ, pp. 14-21, 1987.