Self-Adaptive Genetic Algorithms with Simulated Binary Crossover

Kalyanmoy Deb
Kanpur Genetic Algorithms Laboratory (KanGAL)
Department of Mechanical Engineering
Indian Institute of Technology Kanpur
PIN 2 08 016, India
deb@iitk.ac.in

Hans-Georg Beyer
Systems Analysis Group
Department of Computer Science/XI
University of Dortmund
44221 Dortmund, Germany
beyer@ls11.informatik.uni-dortmund.de

Technical Report No. CI-61/99
March 1999
Department of Computer Science/XI
University of Dortmund
44221 Dortmund, Germany

Abstract

Self-adaptation is an essential feature of natural evolution. However, in the context of function optimization, self-adaptation features of evolutionary search algorithms have been explored only with evolution strategy (ES) and evolutionary programming (EP). In this paper, we demonstrate the self-adaptive feature of real-parameter genetic algorithms (GAs) using simulated binary crossover (SBX) operator and without any mutation operator. The connection between the working of self-adaptive ESs and real-parameter GAs with SBX operator is also discussed. Thereafter, the self-adaptive behavior of real-parameter GAs is demonstrated on a number of test problems commonly-used in the ES literature. The remarkable similarity in the working principle of real-parameter GAs and self-adaptive ESs shown in this study suggests the need of emphasizing further studies on self-adaptive GAs.

1 Introduction

Self-adaptation is a phenomenon which makes evolutionary algorithms flexible and closer to natural evolution. Among the evolutionary methods, self-adaptation properties have been explored with evolution strategies (ESs) (Bäck, 1997; Beyer, 1996; Hansen and Ostermier, 1996; Rechenberg, 1973; Saravanan, Fogel, and Nelson, 1995; Schwefel, 1977, 1987) and evolutionary programming (EP) (Fogel, Angeline, and Fogel 1995), although there exist some studies of self-adaptation in genetic algorithms (GAs) with mutation operator (Bäck, 1992). Despite such studies, there exists no formal definition of self-adaptation or description of properties an algorithm should have in order for it to qualify to be a self-adaptive algorithm. In this paper, we do not try to answer this question, rather recognize the importance of such a study in the near future. When applied to function optimization, there are a number of reasons why evolutionary algorithmists should pay attention to self-adaptation:

1. Knowledge of lower and upper bounds for the optimal solution may not be known a priori,
2. It may be desired to know the optimal solution with arbitrary precision,
3. The objective function and the optimal solution may change with time.

In many problems, the lower and upper bounds for the optimal solution may not be known a priori. Some evolutionary search algorithms such as binary-coded genetic algorithms (GAs) require information about lower and upper bounds on each problem variable, so that a linear mapping of the decoded values of a binary string-coding can be used. This forces the search to concentrate only within the chosen lower and upper bounds. If the true optimal solution does not lie within this range, fixed coding scheme of binary-coded GAs will not be able to find the true optimal solution. In this respect, ES and EP are alternatives where precise knowledge of such range is not required. Real-parameter GAs, where problem variables are not coded in any string-structure, rather are used directly, can eliminate the rigidity of fixed coding used in binary-coded GAs.

Some search and optimization problems require the optimal solution to be found with arbitrary precision. In such cases, the fixed-length coding scheme has another disadvantage. Since a fixed-length coding is used, the algorithm offers a lower bound on the precision that can be achieved in a GA. For example, if ℓ_i bits are used to code a problem variable in the range $[x_i^l, x_i^u]$ in a binary-coded GA, the maximum attainable precision is $(x_i^u - x_i^l)/(2^{\ell_i} - 1)$. Although the precision in the optimal solution can be increased by increasing the string length ℓ_i, it has been shown elsewhere (Goldberg, Deb, and Clark, 1992) that even for simple problems the required population size is of the order of the string length. One other approach to achieve more precision is to use a variable-length coding or a coarse-to-fine grained coding, both of which makes the algorithm more complex and subjective to the way decisions are made in switching from coarse to fine grained coding (Schaefer, 1987, Schraudolph and Belew, 1990). Once again, real-parameter GAs with direct use of problem variables can practically achieve any precision in the problem variables, simply because the real numbers are used directly.

One of the challenging optimization problems and more commonly-found problems in real-world search and optimization is a problem with an objective function which changes with time. In such problems, function landscapes and consequently the optimal solution change with time. When such problems are to be solved for optimality, the search procedure needs to be flexible enough to adapt to the new function landscape as quickly as it changes. A difficulty of the population-based optimizers is that once the search has narrowed near the previous optimal solution, the diversity in the population may not be enough for the search to get out of there and proceed towards the new optimal solution. Often, in these cases, diversity preserving mechanisms (large mutation rate, inclusion of a niching operator, and others) must be used. However, the maintenance of diversity does not come free; a portion of the total function evaluations is always spent in areas of non-interest in current iteration, as an investment for diversity needed at a later iteration when the function landscape changes. ES or EP without self-adaptation cannot tackle such problems and are not flexible enough to respond to change landscapes. However, the invent of self-adaptation with both ES and EP allowed such problems to be solved with an addition of extra strategy parameters which control the degree of search power in their major mutation-based search operators. Although not obvious, such self-adaptive behavior is also possible to achieve with real-parameter GAs with specialized crossover operators.

In this paper, we show the self-adaptive behavior of real-parameter GAs with one such crossover operator on a number of different fitness landscapes commonly used in self-adaptive ES studies. The simulated binary crossover operator (SBX) operator (Deb and Agrawal, 1995) uses a probability distribution around two parents to create two children solutions. Unlike other real-parameter crossover operators, SBX uses a probability distribution which is similar in principle to the probability of creating children solution in crossover operators used in binary-coded GAs. There are two aspects which give real-parameter GAs with SBX their self-adaptive power: (i) children solutions closer to parent solutions are more likely to be created, and (ii) the span of children solutions is proportional to the span of parent solutions.

In the remainder of the paper, we briefly discuss the working principle of the simulated binary crossover (SBX) operator. Thereafter, we mention different variants of self-adaptive ESs and show that there is a similarity in working of real-parameter GAs with SBX and at least one variant of self-adaptive ESs. The self-adaptive behavior of real-parameter GAs with SBX operator is then demonstrated by applying them on a number of test problems. Finally, based on the current study, a number of plausible extensions are suggested.

2 Genetic Algorithms with Simulated Binary Crossover (SBX)

There exists a number of real-parameter GA implementations, where crossover and mutation operators are applied directly on real parameter values. One of the early implementations was by Wright (1990), where a linear crossover operator created three solutions $(x_i^{(1,t)} + x_i^{(2,t)})$, $(1.5x_i^{(1,t)} - 0.5x_i^{(2,t)})$, and $(-0.5x_i^{(1,t)} + 1.5x_i^{(2,t)})$ from two parent solutions $x_i^{(1,t)}$ and $x_i^{(2,t)}$ at generation t and choose the best two solutions as children solutions. Goldberg introduced the concept of virtual alphabets in the context of real-coded GAs (Goldberg, 1991). Eshelman and Schaffer (1993) have introduced the notion of *interval* schemata for real-coded genetic algorithms and suggested a blend crossover (BLX-α) operator. For two parent solutions $x_i^{(1,t)}$ and $x_i^{(2,t)}$ (assuming $x_i^{(1,t)} < x_i^{(2,t)}$), the BLX-$\alpha$ randomly picks a solution in the range $[x_i^{(1,t)} - \alpha(x_i^{(2,t)} - x_i^{(1,t)}), x_i^{(2,t)} + \alpha(x_i^{(2,t)} - x_i^{(1,t)})]$. Thus, if u is a random number between 0 and 1, the following is a child solution:

$$x_i^{(1,t+1)} = (1-\gamma)x_i^{(1,t)} + \gamma x_i^{(2,t)}, \tag{1}$$

where $\gamma = (1+2\alpha)u - \alpha$. If α is zero, this crossover creates a random solution in the range ($x_i^{(1,t)}$, $x_i^{(2,t)}$). In a number of test problems, they have reported that BLX-0.5 (with $\alpha = 0.5$) performs better than BLX operators with any other α value. However, it is important to note that the factor γ is uniformly distributed for a fixed value of α. However, BLX-α has an interesting property: the location of the child solution depends on the difference in parent solutions. This will be clear if we rewrite equation 1 as follows:

$$\left(x_i^{(1,t+1)} - x_i^{(1,t)}\right) = \gamma\left(x_i^{(2,t)} - x_i^{(1,t)}\right).$$

If the difference in the parent solution is small, the difference between the child and parent solutions is also small and vice versa. We believe that this is an essential property for any search algorithm to exhibit self-adaptation. This is because the spread of current population dictates the spread of solutions in the resulting population. In problems where self-adaptation should cause the population to either converge or diverge for better regions in the search space, assignment of children population proportional to parent population may help achieve the task. However, as we shall see later, there is more than such proportional assignment which is needed for the self-adaptation to work.

Ono and Kobayashi (1997) suggested a unimodal normally distributed crossover (UNDX) operator, where three parent solutions are used to create two or more children solutions. Children solutions are created from an ellipsoidal probability distribution with one axis is formed along the line joining two of the three parent solutions and the extent of the orthogonal direction is decided by the perpendicular distance of the third parent from the axis. Unlike the BLX operator, this operator assigns more probability for creating solutions near the center of the first two parents than near the parents. Recently, a multi-parental UNDX operator with more than three parents is also suggested (Kita, Ono, and Kobayashi, 1998). Like the BLX operator, this operator also assigns children solutions proportional to the spread of parent solutions, thereby making a GA with this operator potential to exhibit self-adaptation.

In the year 1995, the first author and his students have developed the simulated binary crossover (SBX), which works from two parent solutions to create two children solutions (Deb and Agrawal, 1995; Deb and Kumar, 1995). As the name suggests, the SBX operator simulates the working principle of the single-point crossover operator on binary strings. In those studies, authors showed that this crossover operator *respects* the *interval* schemata processing, in the sense that common interval schemata between parents are preserved in children. The procedure of computing the children solutions $x_i^{(1,t+1)}$ and $x_i^{(2,t+1)}$ from parent solutions $x_i^{(1,t)}$ and $x_i^{(2,t)}$ is described as follows. A spread factor β is defined as the ratio of the absolute difference in children values to that of the parent values:

$$\beta = \left| \frac{x_i^{2,t+1} - x_i^{1,t+1}}{x_i^{2,t} - x_i^{1,t}} \right|. \tag{2}$$

First, a random number u between 0 and 1 is created. Thereafter, from a specified probability distribution function, the ordinate β_q is found so that the area under the probability curve from 0 to β_q is equal to the chosen random number u. The probability distribution used to create a child solution is derived to have a similar *search power* as that in a single-point crossover in binary-coded GAs and is given as follows (Deb and Agrawal, 1995):

$$\mathcal{P}(\beta) = \begin{cases} 0.5(\eta+1)\beta^\eta, & \text{if } \beta \le 1; \\ 0.5(\eta+1)\frac{1}{\beta^{\eta+2}}, & \text{otherwise.} \end{cases} \tag{3}$$

Figure 1 shows the above probability distribution with $\eta = 2$ and 5 for creating children solutions from two parent solutions ($x_i^{(1,t)} = 2.0$ and $x_i^{(2,t)} = 5.0$) in the real space. In the above expressions, the distribution index η is any nonnegative real number. A large value of η gives a higher probability for creating near parent solutions and a small value of η allows distant solutions to be selected as children solutions. Using equation 3, we calculate β_q by equating the area under the probability curve equal to u, as follows:

$$\beta_q = \begin{cases} (2u)^{\frac{1}{\eta+1}}, & \text{if } u \le 0.5; \\ \left(\frac{1}{2(1-u)}\right)^{\frac{1}{\eta+1}}, & \text{otherwise.} \end{cases} \tag{4}$$

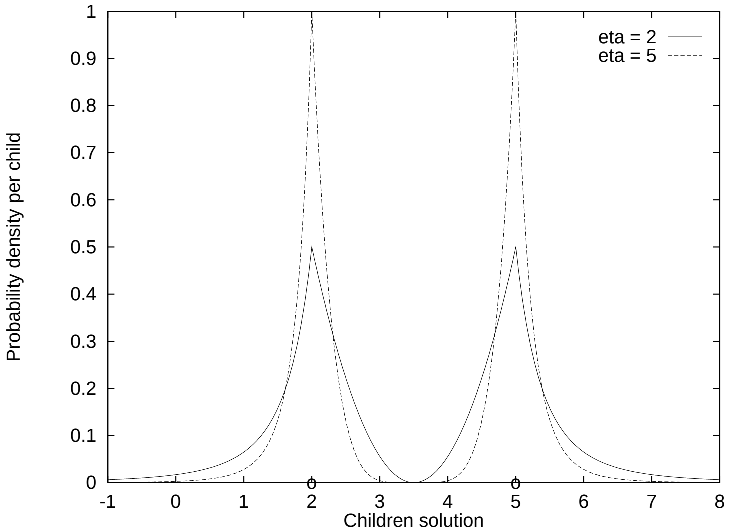


Figure 1: Probability distribution for creating children solutions of continuous variables. Parents are marked with an 'o'.

After obtaining β_q from the above probability distribution, the children solutions are calculated as follows:

$$x_i^{(1,t+1)} = 0.5\left[(1+\beta_q)x_i^{(1,t)} + (1-\beta_q)x_i^{(2,t)}\right], \tag{5}$$

$$x_i^{(2,t+1)} = 0.5\left[(1-\beta_q)x_i^{(1,t)} + (1+\beta_q)x_i^{(2,t)}\right]. \tag{6}$$

Thus, the following step-by-step procedure is followed to create two children solutions ($x_i^{(1,t+1)}$ and $x_i^{(2,t+1)}$) from two parent solutions ($x_i^{(1,t)}$ and $x_i^{(2,t)}$):

Step 1: Choose a random number $u \in [0, 1)$.

Step 2: Calculate β_q using equation 4.

Step 3: Compute children solutions using equations 5 and 6.

Note that two children solutions are symmetric about the parent solutions. This is deliberately used to avoid any bias towards any particular parent solution in a single crossover operation. Another interesting aspect of this crossover operator is that for a fixed η the children solutions has a spread which is proportional to that of the parent solutions. With simple algebraic manipulations, it can be shown that 99% of crossovers lie within $\beta_q \in [(0.01)^{\frac{1}{1+\eta}}, 1/(0.01)^{\frac{1}{1+\eta}}]$. For $\eta = 2$, this range is $\beta_q \in [0.215, 4.64]$. However, for a fixed β_q, the difference in children solutions is proportional to that of parent solutions:

$$\left(x_i^{(2,t+1)} - x_i^{(1,t+1)}\right) = \beta_q \left(x_i^{(2,t)} - x_i^{(1,t)}\right). \tag{7}$$

This has an important implication. Let us consider two scenarios: (i) Two parents are far away from each other, and (ii) two parents are closer to each other. For illustration, both these cases (with parent solutions $x_i^{(1,t)} = 2.0$ and $x_i^{(2,t)} = 5.0$ in the first case and with parent solutions $x_i^{(1,t)} = 2.0$ and $x_i^{(2,t)} = 2.5$ in the second case) and the corresponding probability distributions with $\eta = 2$ are shown in Figures 2 and 3, respectively. For an identical random number, β_q, as calculated by using equation 4, are the same for both

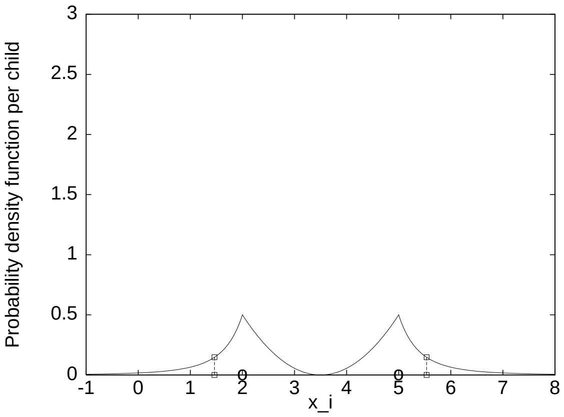


Figure 2: Probability distribution of children solutions with distant parents.

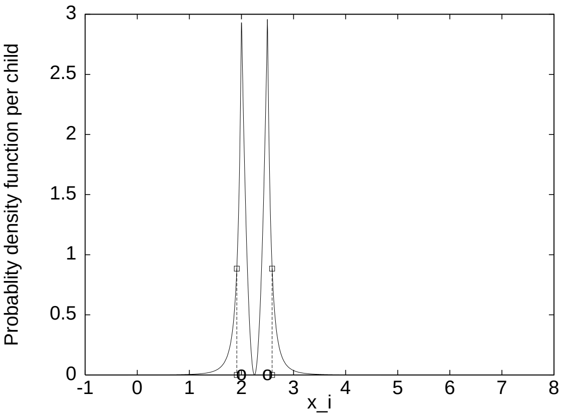


Figure 3: Probability distribution of children solutions with closely spaced parents.

cases. From equation 7, it is clear that in the first case the children are likely to be more widely spread than in the second case. Figures2 and 3 also show the corresponding children solutions (marked with a box) for $u = 0.8$ or $\beta_q = 2.5^{1/3}$. Figures clearly show that if the parent values are far from each other (the first case), solutions away from parents are possible to be created. Compare the child solution $x_i^{(1,t+1)} = 1.464$ with parent $x_i^{(1,t)} = 2.0$. But if the parent values are close by (the second case), distant children solutions are not likely. Compare the child solution $x_i^{(1,t+1)} = 1.911$ with parent $x_i^{(1,t)} = 2.0$ creating using the same random number as in the first case. In initial populations, where the solutions are randomly placed (like the first case), this allows almost any values to be created as a child solution. But when the solutions tend to converge due to the action of genetic operators (like the second case), distant solutions are not allowed, thereby focusing the search to a narrow region.

It is interesting to note that both equations 5 and 6 can be written in the form of equation 1 with following relationships: $\gamma = 0.5(1 \mp \beta_q)$. However, it is important that, unlike in BLX-α operator, the equivalent γ term in SBX operator is not uniformly distributed (refer to equation 4). The SBX operator biases solutions near each parent more favorably than solutions away from parents. Essentially, SBX operator has two properties:

1. The extent of children solutions is in proportion to the parent solutions, and

2. Near parent solutions are monotonically more likely to be chosen as children solutions than solutions distant from parents.

BLX-α operator has the first property, but it does not have the second property in its true sense. Although in BLX operator, any solution within a certain distance from the parents are favored, there is no bias assigned to any of the solutions. We argue that assigning a fixed probablity to all solutions near parents is a too generic property for a GA to show adequate self-adaptation. Later, we show that when a crossover operator has both the above properties, the resulting GA works similar to that of self-adaptive ESs. But before we discuss that similarity, let us describe briefly the existing self-adaptive ESs.

3 Self-Adaptive Evolution Strategies

A simple adaptive evolution strategy was suggested for the (1+1)-ES by Rechenberg (1973). Depending on the success or failure of mutations in past few generations, the mutation strength is increased or decreased by a simple rule. There are three different ways self-adaptation is used in ES—(i) a hierarchically organized population-based meta-ES (Herdy, 1992), (ii) adaptation of covariance matrix (CMA) determining the probability distribution for mutation (Hansen and Ostermeier, 1995), and (iii) explicit use of self-adaptive control parameters (Rechenberg, 1973; Schwefel, 1974). The meta-ES method of self-adaptation uses two levels of ESs—the top level optimizes the strategy parameters (such as mutation strengths), a solution of which is used to optimize the true objective function in the lower level ES. Although the idea is simple, it involves a number of additional strategy parameters to be fixed and the needed number of overall function evaluations may not make it suitable for real-world applications. The third method (CMA) records the population history for some number of iterations before executing expensive numerical computation of finding covariance and variance information among object variables. Although application to a number of test problems shows promising results, the algorithm is difficult to implement and clearly there is lack of any motivation of whether such complicated computations resembles any event of natural evolution. Nevertheless, the CMA approach seems interesting and readers may refer to the literature for more details (Hansen and Ostermeier, 1996; 1997). We now discuss the third type of self-adaptive ES where the strategy parameters are explicitly coded and updated in each generation. Although there exists other ways to update (Rechenberg, 1994), we discuss here the lognormal update rules. A recent study by the second author reveals that there is a relationship between the lognormal update rule with other learning rules (Beyer, 1996). There are basically three different implementations which are in use.

3.1 Isotropic self-adaptation

In this self-adaptive ES, a single mutation strength σ is used for all variables. In addition to N object variables, the strategy parameter σ is also used in a population member. Here are the update rules:

$$\sigma^{(t+1)} = \sigma^{(t)} \exp(\tau_0 N(0,1)), \quad (8)$$

$$x_i^{(t+1)} = x_i^{(t)} + \sigma^{(t+1)} N_i(0,1), \quad (9)$$

where $N(0,1)$ and $N_i(0,1)$ are realizations of a one-dimensional normally distributed random variable with mean zero and standard deviation one. The parameter τ_0 is the learning parameter which is $\tau_0 \propto N^{-1/2}$, where N is the dimension of the variable vector (Schwefel, 1974). Beyer (1996) has shown that, for the sphere model, the optimal learning parameter for (1,λ)-ES is $\tau_0 = c_{1,\lambda}/\sqrt{N}$, where $c_{1,\lambda}$ is the progress coefficient. We use $c_{\mu,\lambda}$ or $c_{\mu/\mu,\lambda}$ as constant of proportionality in corresponding ES, although they may not be optimal. The above update rule for σ requires an initial value. In all simulations here, we choose a $\sigma^{(0)} = (x_i^u - x_i^l)/\sqrt{12}$ which assumes a uniform distribution of solutions within the specified range of x_i values.

3.2 Non-isotropic self-adaptation

A different mutation strength σ_i is used for each variable. This is capable of learning to self-adapt to problems where variables are unequally scaled in the objective function. In addition to N object variables, there are N other strategy parameters. The update rules are as follows:

$$\sigma_i^{(t+1)} = \sigma_i^{(t)} \exp\left(\tau' N(0,1) + \tau N_i(0,1)\right), \tag{10}$$

$$x_i^{(t+1)} = x_i^{(t)} + \sigma_i^{(t+1)} N_i(0,1), \tag{11}$$

where $\tau' \propto (2n)^{-1/2}$ and $\tau \propto (2n^{1/2})^{-1/2}$. Due to lack of any theoretical results on this self-adaptive ES, we use the progress coefficient of the (μ, λ)-ES or $(\mu/\mu, \lambda)$-ESs as the constant of proportionality of both τ' and τ. Similar initial values for $\sigma_i^{(0)}$ as discussed for isotropic self-adaptive ESs are used here.

3.3 Correlated self-adaptation

Here, different mutation strengths and rotation angles are used to represent the covariances for pair-wise interactions among variables. Thus, in addition to N object variables there are a total of N mutation strengths and $N(N-1)/2$ rotation angles used explicitly in each population member. The update rules are as follows:

$$\sigma_i^{(t+1)} = \sigma_i^{(t)} \exp\left(\tau' N(0,1) + \tau N_i(0,1)\right), \tag{12}$$

$$\alpha_j^{(t+1)} = \alpha_j^{(t)} + \beta N_j(0,1), \tag{13}$$

$$\vec{x}^{(t+1)} = \vec{x}^{(t)} + \vec{N}\left(\vec{0}, C(\vec{\sigma}^{(t+1)}, \vec{\alpha}^{(t+1)})\right), \tag{14}$$

where $\vec{N}\left(\vec{0}, C(\vec{\sigma}^{(t+1)}, \vec{\alpha}^{(t+1)})\right)$ is a realization of correlated mutation vector with a zero mean vector and covariance matrix C. The parameter β is fixed as 0.0873 (or 5^o) (Schwefel, 1974). The parameters τ' and τ are used the same as before. We initialize the rotation angles within zero and 180 degrees at random.

4 Connection Between GAs with SBX and Self-Adaptive ESs

Without loss of generality, we try to argue the similarity in the working of GAs with SBX and self-adaptive ESs by considering only isotropic self-adaptive ESs. Under isotropic self-adaptive ES, the difference (say Δ) between the child ($x_i^{(t+1)}$) and its parent ($x_i^{(t)}$) can be written from equations 8 and 9:

$$\Delta = \left(\sigma^{(t)} \exp(\tau_0 N(0,1))\right) N(0,1). \tag{15}$$

Thus, an instantiation of Δ is a normal distribution with zero mean and a variance which depends on $\sigma^{(t)}$, τ_0, and the instantiation of the lognormal distribution. For our argument, there are two aspects of this procedure:

1. For a particular realization of lognormal distribution, the difference Δ is normally distributed with zero mean. That is, children solutions closer to parents are monotonically more likely to be created than children solutions away from parents.

2. The standard deviation of Δ is proportional to the mutation strength $\sigma^{(t)}$, which signifies, in some sense, the population diversity.

Under the SBX operator, we write the term Δ using equations 5 and 6 as follows:

$$\Delta = \frac{\delta_p}{2}(\beta_q - 1), \tag{16}$$

where δ_p is the absolute difference in two parent solutions. There is a similarity between equations 15 and 16. The above equation suggests that an instantiation of Δ depends on the distribution of $(\beta_q - 1)$ for a particular pair of parents. The distribution of β_q has its mode at $\beta_q = 1$, thus, the distribution of $(\beta_q - 1)$ will have its mode at zero. Although, we have not used a normal distribution for $(\beta_q - 1)$ here, Figure 2 or 3 suggests that a small Δ has a higher probability to be created than a large Δ and that this distribution is montonic to the distance from a parent. The variance of this distribution depends on δ_p, which signifies the population diversity. Thus, there is a remarkable similarity in the way children solutions are assigned in both isotropic self-adaptive ES and in GAs with SBX. In both cases, the children solutions closer to parent solutions are assigned more probability to be created than solutions away from parents and the variance of this probability distribution depends on the current population diversity.

In a self-adaptive ES, the mutation strength gets continuously updated depending on the fitness landscape. For example, if the fitness landscape is such that the population needs to concentrate in a narrow region in the search space for improvement in the fitness, a self-adaptive ES usually evolves mutation strengths to become smaller and smaller, so that search concentrates near the parents rather than away from parents. This is precisely how a self-adaptive ES works on sphere model to achieve continuously improving performance. The outcome of continuously reducing mutation strength is that most population members come closer and closer. When population members come closer in a real-parameter GA, the effective variance of probability distribution under SBX operator also reduces. This, in turn, creates children solutions which are also not far away from each other. This helps to produce continuously closer population members, thereby producing the effect of increased precision like that in the self-adaptive ES. A similar phenomenon occurs when a fitness function demands the population to diverge to get to the optimal region or demands other kind of variations in the search process.

5 Simulation Results

In this section. we present simulation results of real-parameter GAs with SBX operator on a number of different test problems borrowed from the ES literature. For handling multi-variable problems, SBX is used variable-by-variable with a variable-wise probability of 0.5. This means that, on an average, 50% of variables get crossed using the SBX operator and rest of the variables get passed on to the children solutions unchanged. This is in agreement with single-point, two-point, or uniform crossovers used in binary-coded GAs, where, on an average, 50% bits get changed in one crossover operation. However, we would like to mention here that since variable-wise crossover is used, GAs with the current implementation of SBX may face difficulty in problems where variable interactions are important. This so called *linkage issue* is an important matter in search and optimization problems and has been recognized by many researchers (Goldberg, Korb, and Deb, 1989; Harik, 1997; Kargupta, 1996; Schwefel, 1977). However, more research must be carried out to find linkages among variables adaptively using evolutionary algorithms. Wherever recombinative ES is used, the dominant crossover is used on object variables and intermediate recombination is used on strategy parameters, as suggested by Schwefel (1974). In all methods, no special effort is spent to find the best parameter settings, instead a reasonable set of parameter values are used. In all test problems, we use $N = 30$ variables.

5.1 Sphere model

This is the most commonly-used test function chosen for studying (both experimentally and theoretically) self-adaptation properties of ES. We consider several variants of this function in the following subsections.

5.1.1 Function F1-1: Quadratic function

First, we consider the sphere model, where the objective is to minimize the following N-variable function:

$$\text{F1-1:} \quad \text{Minimize} \sum_{i=1}^{N} (x_i - x_i^*)^2, \tag{17}$$

where x_i^* is the optimal value of the i-th variable. In the simulations presented first, we have chosen $x_i^* = 0.0$. Populations are initialized in $x_i \in [-1.0, 1.0]$. Real-parameter GAs with SBX are used to find the optimal solution for this function. Tournament size $s = 2$ and distribution index $\eta = 1$ for SBX operator are used. A population size of 100 is used. To investigate the effect of SBX operator alone, we have not used any mutation operator. The Euclidean distance $R = \sqrt{\sum_{i=1}^{N} x_i^2}$ of the best solution in a population from the minimum is plotted with generation number in Figure 4. The ordinate axis is drawn in

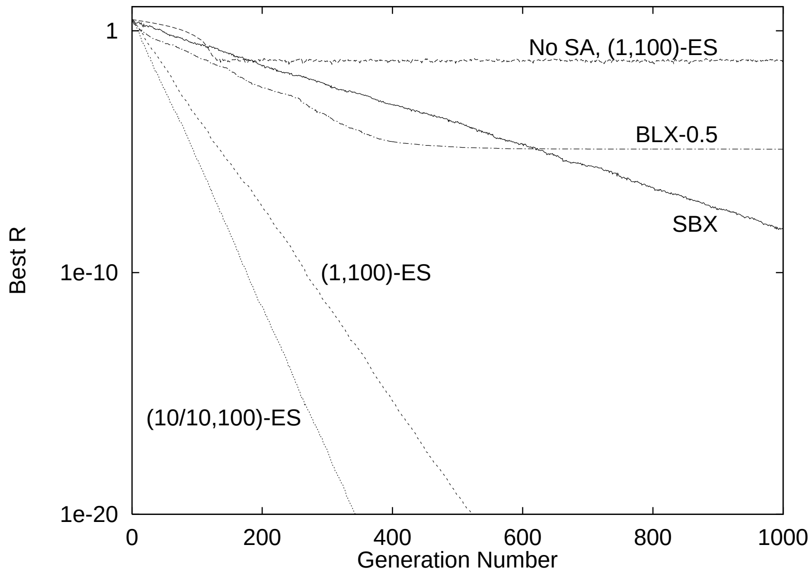


Figure 4: Population-best distance R from the minimum obtained with several evolutionary algorithms on test function F1-1.

logarithmic scale and the figure shows that real-parameter GAs with SBX (solid line) are able to maintain increased precision with generation number. Here, it is worth mentioning that the straight lines refer to linear convergence, that is, an exponential decrease of the residual distance to the optimum. The figure also shows the performance of several other evolutionary algorithms. ES with no self-adaptation (marked as 'No SA') and with $\mu = 1$ and $\lambda = 100$ are shown next. We have used a fixed mutation strength of 0.01 here. The children population size $\lambda = 100$ is used to keep the number of function evaluation same as that in the real-parameter GAs. The figure re-confirms an already established (Beyer, 1996) fact that a comma ES without self-adaptation cannot find continuously increasing precision. From a theoretical analysis on the sphere model it was found that a non-self-adaptive ES gets stuck at a distance R_∞ given by (Beyer, 1996):

$$R_\infty = \frac{\sigma N}{2c_{\mu,\lambda}}.$$

For (1,100)-ES, $c_{1,100} = 2.51$ and the above equation yields $R_\infty = 0.0598$. An average of population-best R values from generation 800 till 1000 generations is calculated from the simulation data and it is found to be 0.0585, which is in good agreement (within 2%) with the theory.

Next, we apply isotropic self-adaptive ESs ((1,100)-ES and (10/10,100)-ES). In all self-adaptive ES runs, we have used $\tau = c_{\mu,\lambda}/\sqrt{N}$ to update the mutation strength. For example, $c_{1,100} = 2.51$ and $c_{10/10,100} = 1.94$ are used (Beyer, 1995). The figure also shows an already established fact that with proper learning parameter update, self-adaptive ESs can find continuously improving precision. The theory suggests that for the sphere model with $(1, \lambda)$-ES, the relative distance (R) change is $c_{1,\lambda}^2/(2N)$, or 0.1048 with $N = 30$ (Beyer, 1996). When the slope of the (1,100)-ES plot is calculated from the above figure, it is found to be 0.1011, which is about 3% from the theoretical estimate. Both these calculations and experimental results give us confidence in our self-adaptive ES implementations.

There are two aspects to notice in Figure 4. First, the introduction of crossover enhances the performance of self-adaptive ES in this problem (Beyer, 1995). Second, the performance of the self-adaptive ES is much better than that of real-parameter GAs with SBX operator. This test function is unimodal and isotropic ES uses this problem knowledge by using only one mutation strength parameter for all variables. On the other hand, real-parameter GAs with SBX does not use any such information and hence the progress rate comparison between the two algorithms is not proper. Moreover, there is a mismatch of effective selection pressure in both algorithms, with more selection pressure for the self-adaptive ESs. But, in the elliptic test function, independent mutation strengths are needed to achieve self-adaptation and both algorithms may then be compared. However, what is important here to note that real-parameter GAs with SBX operator is able to maintain increased precision with generation number.

Before we leave this test problem, we would like to mention that when real-parameter GAs with BLX-0.5 is used on this problem, adequate self-adaptive behavior is not observed. GAs get stuck at solutions away (at a distance $R = 1.276(10^{-5})$) from the optimum. Although it has some capabilities of self-adaptation compared to (1,100)-ES without self-adaptation, clearly the self-adaptive power is not adequate.

5.1.2 Function F1-2: Biased population

It has been shown elsewhere (Fogel and Beyer, 1996) that an initial population symmetrically placed around the true optimum may induce a bias for better performance of a recombinative self-adaptive ES. We take the clue from that study and use the same sphere model as that used in the previous subsection, but here we initialize the population far away from the optimum and in a narrow range $x_i \in [10 - \epsilon, 10 + \epsilon]$, where ϵ is a small positive number. However the minimum solution is at $x_i^* = 0.0$. The average distance from initial population from the minimum solution is thus $R_0 = 10\sqrt{N}$ (or, 54.772 with $N = 30$). We choose three different values of $\epsilon = 10^{-5}$, 10^{-10}, and 10^{-15}. Figures 5 and 6 show the population-best distance from optimum (R) and the population standard deviation[1] in the variable vectors (averaged over all $N = 30$ variables). Identical GA parameter settings as before are used here. With real-parameter GAs with SBX operator, the initial population begins with a standard deviation of the order of ϵ and grows to a large value. Thereafter, the population standard deviation reduces as in test function F1-1 and GAs converge to the true optimum. Notice, how GAs require larger number of generations to bring the population standard deviation to a reasonable limit with smaller ϵ values. This behavior of GAs is very similar to that observed with self-adaptive ESs (Bäck, 1997). Once the population has the correct population variance and it is near the optimum, the rate of convergence to the optimum with increasing precision is independent of how the population was initialized. These results show that although 100 members in the initial population was confined to a tiny region, GAs with SBX operator can come out of there mainly with function value information and converge to the correct optimum.

[1]This quantity is calculated by first finding the mean $\bar{x}$ of all population members. Thereafter, the standard deviation is computed as $\sqrt{\sum_{j=1}^{n}(x^{(j)} - \bar{x})^T(x^{(j)} - \bar{x})/(n-1)}$, where n is the population size and $x^{(j)}$ is the x-vector of the j-th population member.

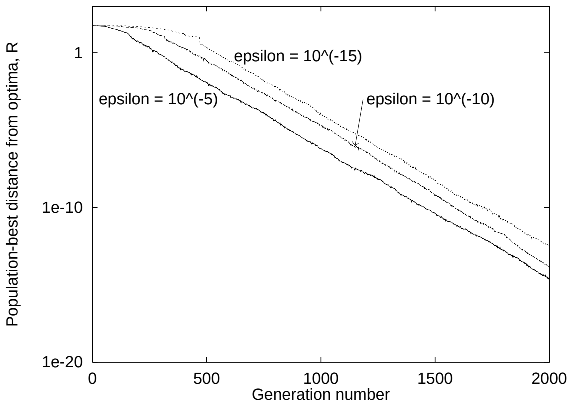


Figure 5: Population-best distance from optimum (R) for populations initialized at different ranges $[10-\epsilon, 10+\epsilon]$ for the function F1-2.

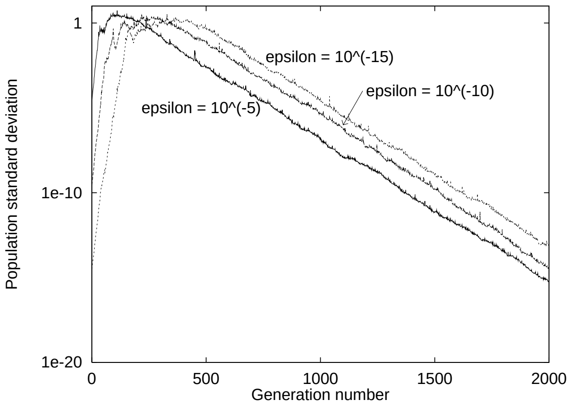


Figure 6: Population standard deviation with generation number for populations initialized at different ranges $[10-\epsilon, 10+\epsilon]$ for the function F1-2.

5.1.3 Function F1-3: Time-varying function

In order to investigate the performance of real-parameter GAs with SBX on time-varying functions, we now choose the same function as F1-1, but x_i^* now varies with generation number in the range $[-1.0, 1.0]$ at random. The optimum is changed after every 1,000 generations so that at the time of a change in the optimum the population diversity has reduced substantially. The best function value and average population standard deviation (as defined earlier) in all variables are plotted versus generation number in Figure 7. GA parameter settings same as that in F1-1 are used here. The figure shows that even though all

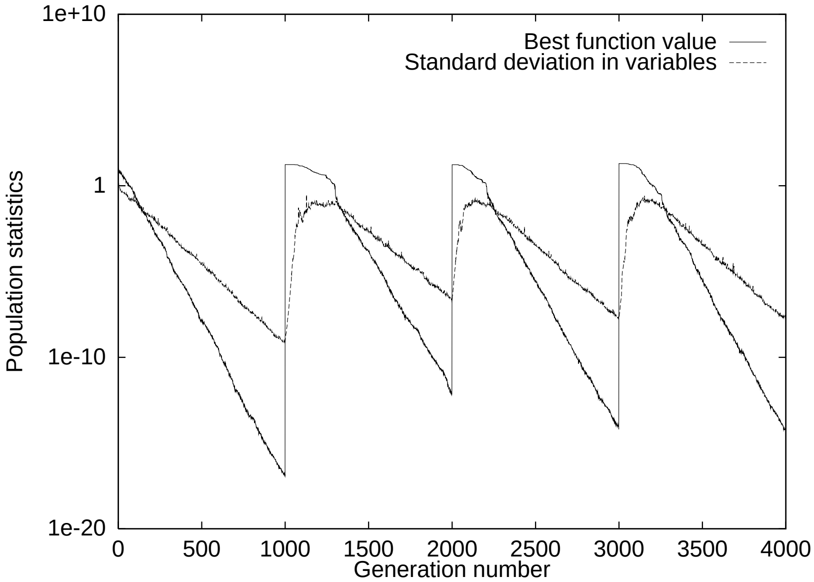


Figure 7: Population-best function value and average of population standard deviation in variables are shown with generation number, as test function F1-3 changes its optimum after every 1,000 generations.

population members are all within a small range (in the order of 10^{-10}) at the end of 999 generations, the population with SBX operator can diverge and can get adapted to a changed optimum. This happens not only once, but as many times as there is a change in the function.

The sudden jump in the best objective function value at generation 1,000 suggests that the new function value at the old optimum (at 999-th generation) is not good. The population standard deviation in all variables at generation is of the order of 10^{-8}, suggesting that the population has lost diversity with respect to the new function. Even then, GAs with SBX operator can quickly increase its population diversity and converge to the new optimum. First, the population gets more diverse to come near the current optimum and then decrease diversity to converge closer and closer to the current optimum. This is exactly the self-adaptive behavior which self-adaptive ESs are expected to exhibit (Bäck, 1997), and we find here similar self-adaptive behavior of real-parameter GAs with the SBX operator.

5.1.4 Function F1-4: Multi-modal function

We now choose one test problem which is not quadratic in the search space. Moreover, in the range $[-1.0, 1.0]$, where the population is initialized, the function is unimodal, but just outside this range the function has other local attractors. We simply add a non-linear term to the sphere model:

$$\text{F1-4:} \quad \text{Minimize} \sum_{i=1}^{N} \left(x_i^2 + 10(1 - \cos(\pi x_i)) \right). \tag{18}$$

The interesting aspect is that the function has a global optimum (all $x_i = 0$) in the range where the population is initialized. But beyond this range, there exist a number of local optima—the nearest one for each variable is at $x_i = -2$ and $x_i = 2$. Figure 8 shows an one-dimensional version of this function. The range, where the population is initialized, is shown by drawing two vertical dashed lines. Figure 9 shows

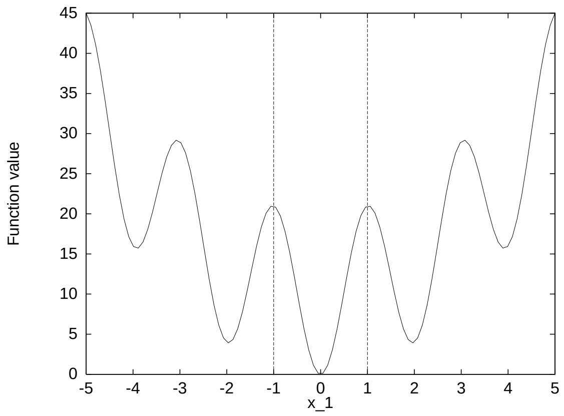


Figure 8: An one-dimensional version of F1-4.

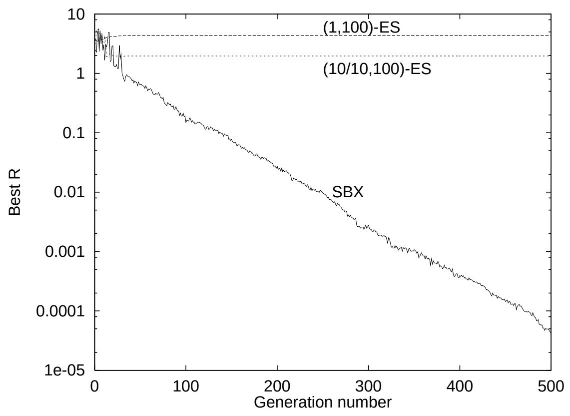


Figure 9: Population-best distance R from optimum for F1-4.

the performance of real-parameter GAs with SBX with $\eta = 1$. The figure also shows the performance of isotropic self-adaptive (1,100)-ES and (10/10,100)-ES with identical parameter settings as used in Function F1-1. Now, self-adaptive ES seems not able to converge to the global attractor (all $x_i = 0$ having function value equal to zero), although the initial population was placed in the global basin. In self-adaptive ESs, the mutation strength for each parameter needs an adaptation time within which update of mutation strengths and corresponding fitness landscape should make a suitable agreement. If either due to improper use of learning rate or other ES parameters or due to a complex fitness landscape this agreement does not happen, the mutation strength does not get adapted properly. Since in this function, the landscape just outside $[-1, 1]$ has a non-agreeing landscape compared to that inside the region $[-1, 1]$ for each variable, self-adaptive ES gets confused whether to increase or decrease mutation strengths for variables.

However, as suggested in Beyer (1996, page 335), if lower and upper bounds on variables are known with confidence, self-adaptive ES may be used with a small initial mutation strength $\sigma^{(0)}$. It is intuitive that when a small initial mutation strength $\sigma^{(0)}$ is used, the mutated solutions are not likely to be outside $[-1, 1]$, and thus self-adaptive ES will be confined in the global basin. When an initial mutation strength $\sigma^{(0)}$ one-tenth of what used in the above runs is used, the ES converges to the correct optimum. However, a small initial mutation strength may not have the desired properties in other functions where divergence from the initial population is necessary to get to the true optimum (such as test function F1-2, F1-3, or ridge functions). Nevertheless, the study of this multi-modal test function suggests the importance of the initial mutation strength in successful working of self-adaptive ES, particularly in non-linear problems, a matter which is not adequately addressed in the ES literature.

In real-parameter GAs with SBX, there is an equal probablity of creating solutions inside or outside the region bounded by two parent solutions. Thus, many children solutions will be created outside $[-1, 1]$. But since the generation of children solutions are mainly governed by the distance between parent solutions and above function has an overall quadratic structure with its attractor at the global optimum, real-parameter GAs do not face much difficulty in converging to the true optimum. Notice how similar the performance of GAs with SBX in this figure is with respect to that in Figure 4, where the simple sphere model was considered.

5.2 Elliptic model

In this function, every variable has an unequal contribution to the objective function. We consider a few variants of the elliptic function.

5.2.1 Function F2-1: Elliptic function

It is similar to the sphere model, but not every variable has equal contribution to the objective function:

$$\text{F2-1:} \quad \text{Minimize} \sum_{i=1}^{N} 1.5^{i-1} x_i^2. \tag{19}$$

Since each variable has unequal contribution to the objective function, self-adaptive ESs with isotropic mutation strength is not adequate to solve this problem. First, we use real-parameter GAs with SBX and then show that its self-adaptive behavior is similar to that in a non-isotropic self-adaptive ES, where a separate mutation strength is used for each variable.

Figure 10 shows the objective function value of the best solution in the population. The population is initialized in $x_i \in [-1.0, 1.0]$. Once again, we use the same GA parameters as before, but use tournament size 3 to compare the performance with self-adaptive ES. The performance of non-isotropic (10/10, 100)-ES with lognormal update of learning parameter is also shown. The performance of BLX-0.5 on this function shows that BLX-0.5 does not have adequate self-adaptive power.

Figure 11 plots the population standard deviation in x_1, x_{15} and x_{30} variables in the population for the real-parameter GAs with SBX operator. Since they are scaled as 1.0, $1.5^{14} \approx 292$, and $1.5^{29} \approx 127,834$, respectively, the 30-th variable is likely to have smaller variance than the 1st variable. The figure shows this fact clearly. Since, ideal mutation strengths for these variables are also likely to be inversely proportionate as 1.5^{i-1}, we find similar ordering with non-isotropic self-adaptive ES as well (Figure 12). Thus, there is a remarkable similarity by which the both real-parameter GAs with SBX and self-adaptive ES work. In the former case, the population diversity gets adapted based on the need of the fitness landscape, which in turn helps the SBX operator to create x_i values of children solutions proportionately. In the latter case, the population diversity gets controlled by independent mutation strengths, which get adapted based on the fitness landscape.

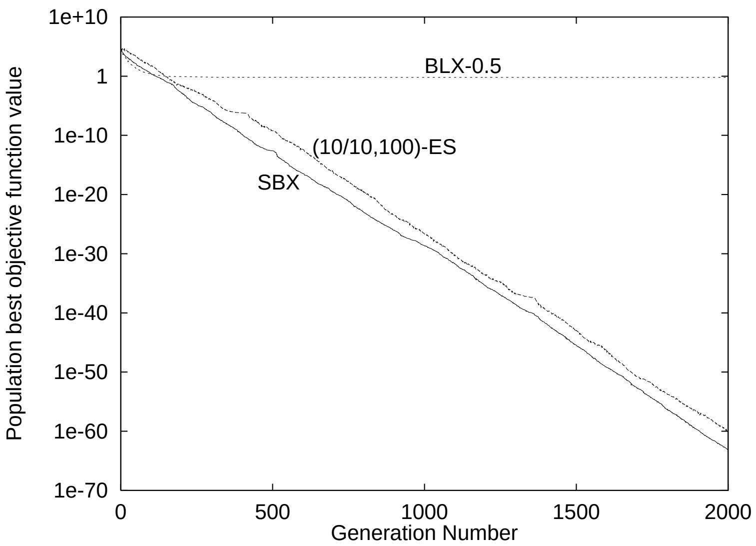


Figure 10: The best objective function value in the population is shown for three evolutionary algorithms—real-parameter GAs with SBX, self-adaptive (10/10,100)-ES, and real-parameter GAs with BLX-0.5 operator on function F2-1.

5.2.2 Function F2-2: Time varying elliptic function

Like in the sphere model, we construct a test problem where the elliptic function changes its optimum solution occasionally with generation. We use the following function:

$$\text{F2-2:} \quad \text{Minimize} \sum_{i=1}^{N} r_i (x_i - x_i^*)^2. \tag{20}$$

where r_i is an randomly shuffled array of integers between 1 to N. After every 1,000 generations, this array is changed to another permutation of integers from 1 to N. In addition, the optimum (x_i^* values) of function is also changed to a random value in the range $[-1.0, 1.0]$. The parameter setting of tournament size of 3 for real-parameter GAs and $\mu = \rho = 10$ for self-adaptive ES (which are used in the previous experiment) make the corresponding algorithm too sluggish to adapt to the changes made at every 1,000 generations. Thus, in this experiments, we use tournament size of 2 for GAs and $\mu = \rho = 15$ with $c_{15/15,100} = 1.558$ for ESs. Figure 13 shows the population-best objective function value and the population standard deviation in the x_{15} variable. It is clear that although the population deviations are quite small at the end of 999 generations, the population can adapt to the change in the function landscape. A similar performance plots are observed with (15/15, 100)-ES with non-isotropic self-adaptation in Figure 14. In this figure, the population-best objective function value and the mutation strength for x_{15} variable are shown. The remarkable similarity in both figures suggests that for chosen parameter settings both real-parameter GAs with SBX operator and self-adaptive ES have very similar working principles.

5.2.3 Functions F2-3 and F2-4: Multi-modal function

Like before, we choose a multi-modal elliptic test function to investigate the performance of self-adaptive ESs and GAs with SBX operator. Before we discuss the function, we first show simulation results of non-isotropic self-adaptive ES and real-parameter GAs with SBX on the following elliptic function:

$$\text{F2-3:} \quad \text{Minimize} \sum_{i=1}^{N} i x_i^2. \tag{21}$$

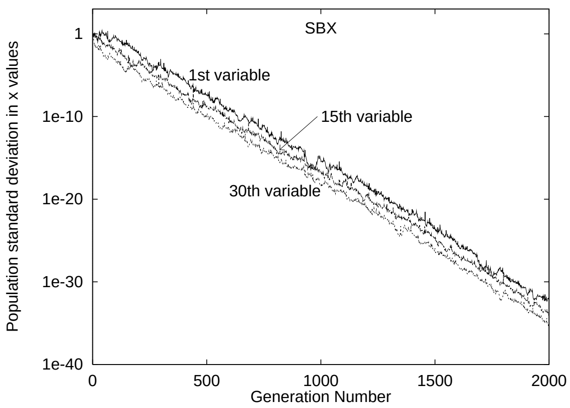


Figure 11: Population standard deviation in variables x_1, x_{15}, and x_{30} are shown with real-parameter GAs with the SBX operator for the function F2-1.

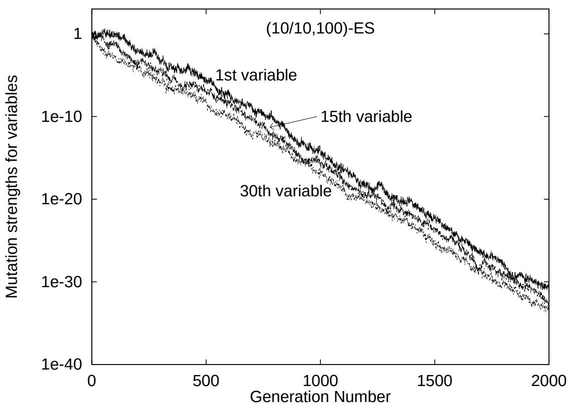


Figure 12: Mutation strengths for variables x_1, x_{15}, and x_{30} are shown with self-adaptive (10/10,100)-ES for the function F2-1.

Figure 15 shows that the (10/10,100)-ES with self-adaptation is able to converge near the optimum. The same is true with real-parameter GAs with SBX, but the rate of convergence is smaller compared to that of the self-adaptive ES.

We now construct a multi-modal test function by adding a cosine term as follows:

$$\text{F2-4:} \quad \text{Minimize} \sum_{i=1}^{N} i \left(x_i^2 + 10(1 - \cos(\pi x_i)\right). \qquad (22)$$

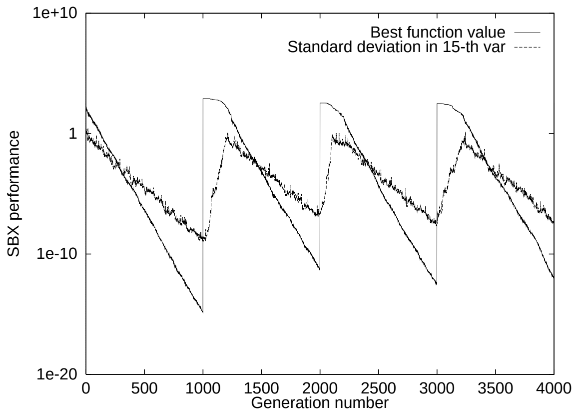
Figure 16 shows the performance of real-parameter GAs and self-adaptive ESs with identical parameter setting as on the elliptic function F2-3. Now, self-adaptive ES seems *not* able to converge to the global attractor (all $x_i = 0$ having function value equal to zero), for the same reason as that described in section 5.1.4—the initial mutation strength being too large to keep the population within the global basin. When all initial mutation strengths reduced to one-tenth of their original values, non-isotropic ESs had no trouble in finding the true optimum with increased precision. We observe no such difficulty in GAs with SBX operator and a performance similar to that in Function F2-3 is observed here.

5.3 Correlated function

Next, we consider a function where pair-wise interactions of variables exist. The following Schwefel's function is chosen:

$$\text{F3-1:} \quad \text{Minimize} \sum_{i=1}^{N} \left(\sum_{j=1}^{i} x_j \right)^2. \qquad (23)$$

The population is initialized at $x_i \in [-1.0, 1.0]$. Figure 17 shows the performance of real-parameter GAs with SBX ($\eta = 1$) and tournament size 3, non-isotropic (4,100)-ES, and correlated (4,100)-ES. Although all methods have been able to find increased precision in obtained solutions, the rate of progress for the real-parameter GAs with SBX is slower compared to that of the correlated self-adaptive ESs. However, GAs with SBX makes a steady progress towards the optimum and performs better than the non-isotropic (4,100)-ES. The reason for the slow progress of real-parameter GAs is as follows. In the SBX operator, variable-by-variable crossover is used with a probability of 0.5. Correlations (or linkage) among the variables are not explicitly considered in this version of SBX. Although some such information comes via the population diversity in variables, it is not enough to progress faster towards the optimum in this problem compared to the correlated ES, where pairwise interactions among variables are explicitly considered.



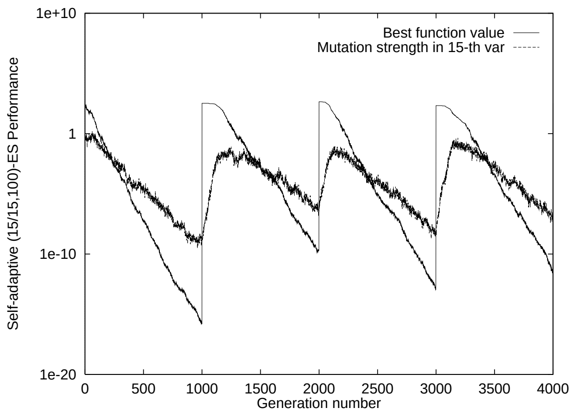
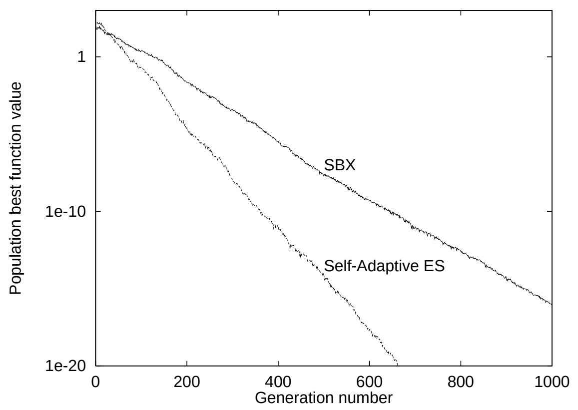


Figure 13: Population-best objective function value and the standard deviation of x_{15} variable are shown for real-parameter GAs with SBX operator on function F2-2.

Figure 14: Population-best objective function value and the mutation strength for x_{15} variable are shown for self-adaptive (15/15,100)-ES on function F2-2.



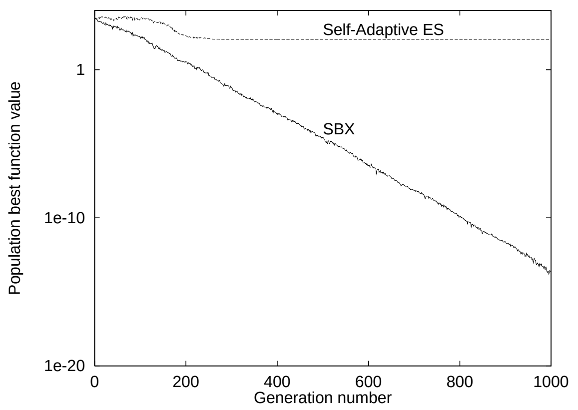


Figure 15: Population-best objective function value for real-parameter GAs and self-adaptive ESs are shown for the elliptic function F2-3.

Figure 16: Population-best objective function value with real-parameter GAs and self-adaptive ESs are shown for the multi-modal function F2-4.

Clearly, a better crossover operator handling the linkage issue but with the concept of probability distribution to create children solutions is in order to solve such problems faster. One such implementation is suggested in Section 6.

Besides the linkage issue discussed above, there is another mismatch between the GAs with SBX and the correlated self-adaptive ESs used above. In GAs, a selection pressure of 3 (best solution in a population gets a maximum of three copies after the tournament selection operation), whereas in the correlated self-adaptive (4,100)-ESs, only 4 best solutions are picked from 100 children solutions. In order to alleviate this mismatch in selection pressures, we use a different algorithm (we call ES-SBX) where a (μ, λ)-ES is used, but λ children solutions are created from μ parent solutions only by the action of SBX operator alone. Each parent $(x^{(1,t)})$ mates with other parents in exactly λ/μ crossovers, everytime creating one child solution using equation 5. Like the SBX operator used in GAs, every variable is crossed with a probability 0.5. If a variable is not to be crossed, its value $(x_i^{(}1,t))$ in the first parent is directly passed on to the child. No explicit mutation operator is used. The rest of the algorithm is exactly the same as that in a (μ, λ)-ES. As shown in Figure 17, this new algorithm with (4,100)-ES-SBX is able to achieve

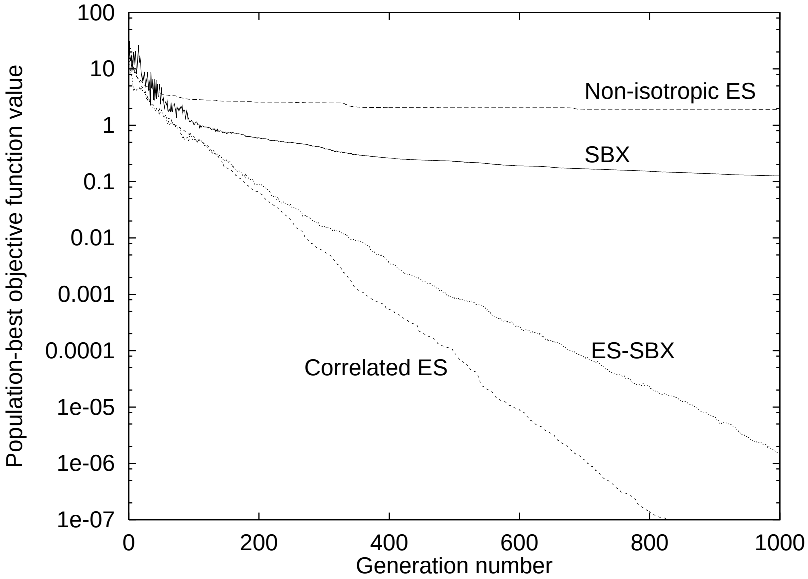


Figure 17: The best objective function value in the population is shown for four evolutionary algorithms—real-parameter GAs with SBX, non-isotropic (4,100)-ES, and correlated self-adaptive (4,100)-ES, and (4,100)-ES-SBX—on function F3-1.

much better performance than GAs with SBX operator, although no explicit pair-wise correlations among variables are used in any operator. However, we recognize that the linkage issue may still be a problem in general, but this simple change in the algorithm seems to fair well in a problem where finding pair-wise linkage information is important.

5.4 Ridge model

Above problems have tested the ability of algorithms to converge near the true optimum with increasing precision. We now test algorithms for an opposing characteristic. We consider the following function, which is largely known as the ridge functions in ES literature:

$$\text{Maximize} \quad v^T x - d\left(\|(v^T x)v - x\|\right)^{\alpha}, \tag{24}$$

where x is the N-dimensional variable vector and v is the ridge axis (or the direction vector specifying the ridge axis). Thus, the first term is the projection of x vector along the ridge axis. The term inside $\|\ \|$ is the orthogonal component specifying the distance to the ridge axis. Since the objective is to maximize the overall function, the subgoals are to maximize the distance along the ridge axis and minimize the distance to the ridge axis. The parameters d and α govern the shape of ridge functions. Higher values of d makes the second term more prominent compared to the first term and therefore makes an algorithm difficult to progress along the ridge axis. In all simulations here, we use a comparatively large $d = 1$. The parameter α has a direct effect on the fitness landscape. Usually, two values of α are commonly used—$\alpha = 2$ is known as the parabolic ridge function and $\alpha = 1$ is known as the sharp ridge function. Here, we shall consider the parabolic ridge function only.

The ridge functions have their theoretical maximum solution at infinity along the ridge axis in the search space. Since the maximum solution lies on the ridge axis, this function tests two aspects: converging ability on the axis, and diverging ability in the direction of the ridge axis. We test the performance of real-parameter GAs with SBX and the non-isotropic self-adaptive ES.

5.4.1 Function F4-1: Parabolic ridge

Here we consider the ridge function with $\alpha = 2$. At first, we choose $v_1 = 1$ and all other $v_i = 0$ for $i = 2, \ldots, N$. This makes the first coordinate axis as the ridge axis. In all simulations, we use $N = 30$ and the population is initialized in $x_i \in [-2, 2]$. Real-parameter GAs with SBX ($\eta = 1$) and with binary tournament selection are used. A population of size 100 is used. Figure 18 plots the distance along the ridge axis (the term $v^T x$) with generation number. The figure shows that real-parameter GAs are able to

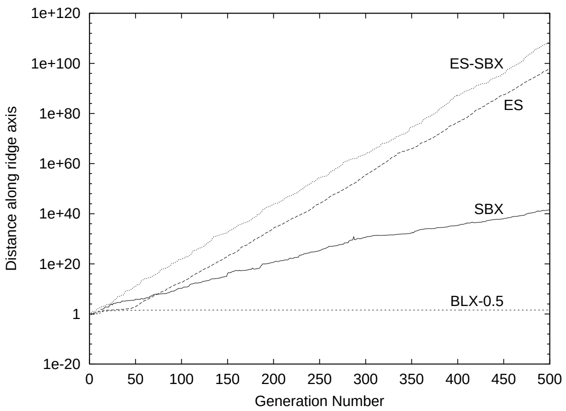


Figure 18: Performance of real-parameter GAs with SBX and BLX-0.5, non-isotropic self-adaptive (10/10,100)-ESs, and (10,100)-ES-SBX are shown on the ridge function F4-1.

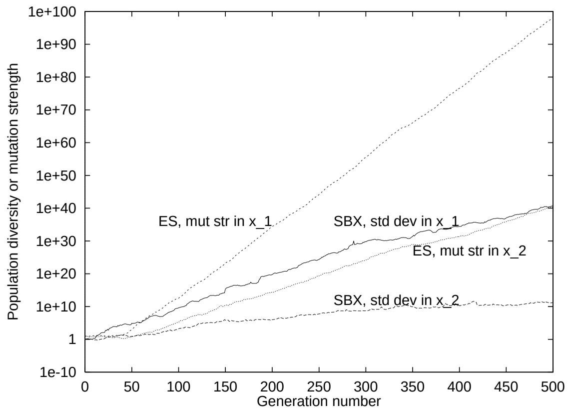


Figure 19: Population standard deviation in x_1 and x_2 for GAs with SBX and population average mutation strengths in x_1 and x_2 are shown for the ridge function F4-1.

progress towards infinity along the ridge axis exponentially with generation number (the ordinate axis is plotted in logarithmic scale). A run with non-isotropic self-adaptive (10/10,100)-ES with one independent mutation strength parameter for each variable shows a similar behavior, but with much better progress rate. Since the optimum is at infinity along the ridge axis, an algorithm with faster divergence characteristics is desired. Since BLX-0.5 allows creation of more diverse children solutions than parents, we tried using real-paramater GAs with BLX-0.5 operator. The figure shows poor self-adaptive power of the BLX-0.5 operator.

Figure 19 shows the population standard deviation in x_1 and x_2 for real-parameter GAs with SBX operator. The figure also shows the population average mutation strengths in x_1 and x_2 as they evolve for the parabolic ridge function. It is clear that in both algorithms the population diversity or mutation strength for x_1 grows much faster than that in x_2. This is what we have expected, because for the parabolic ridge function there are two subgoals. By increasing the population diversity or mutation strength at faster rate will allow the corresponding algorithm to move quicker along the ridge axis. On the other hand, since the secondary goal is to come closer to the ridge axis, an ideal situation would be reduce the population diversity or mutation strengths of other variables (x_2 to x_N) as small as possible. However, both algorithms resorted to increase these quantities with generation, but at a rate much smaller than the growth in x_1 variable.

Next, we apply (10,100)-ES-SBX (where SBX is used as the only search operator in an ES, which is described in the previous subsection). SBX operator with $\eta = 1$ is used. Figure 18 shows that a better progress compared to all other algorithms is obtained with ES-SBX algorithm.

5.4.2 Function F4-2: Parabolic ridge with rotated ridge axis

In this case, we use a random $\vec{v}$, so that the ridge axis is now not along a coordinate direction. Parameters as that used in F4-1 for real-parameter GAs with SBX and self-adaptive ES are chosen here and an identical $\vec{v}$ is used for all algorithms. Figure 20 shows the progress towards the ridge axis with real-parameter GAs, with non-isotropic self-adaptive ESs, and with ES-SBX algorithm. The self-adaptive ES has a faster progress rate. However, it is worth mentioning that a non-recombinative self-adaptive (10,100)-ES performs poorly in this function. However, notice that due to the parameter interactions, the progress along the

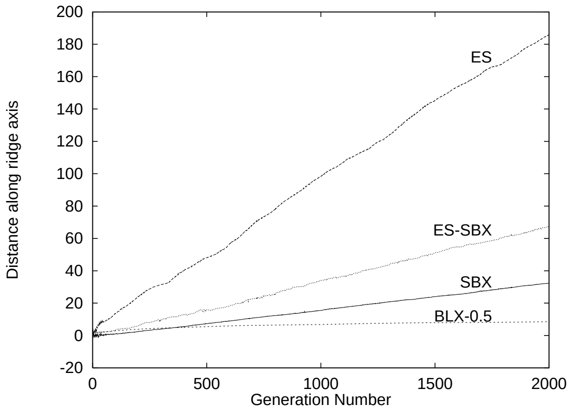


Figure 20: Performance of real-parameter GAs with SBX and BLX-0.5, non-siotropic self-adaptive (10/10,100)-ESs, and (10,100)-ES-SBX are shown on the rotated ridge function F4-2.

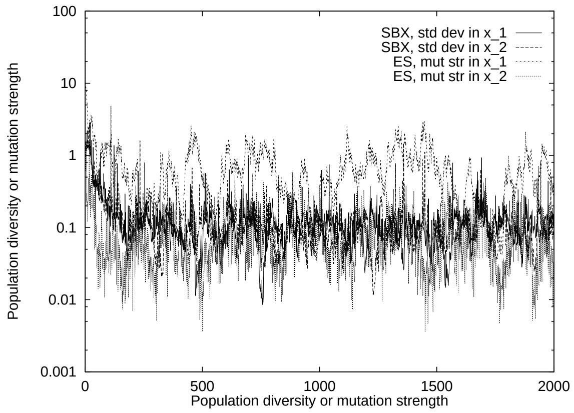


Figure 21: Population standard deviation in x_1 and x_2 for GAs with SBX and population average mutation strengths in x_1 and x_2 are shown for the rotated ridge function F4-2.

ridge axis is now not exponential to the generation number, rather the progress is linear. Since to improve along the ridge, all N variables need to be changed in a particular way, the progress slows down. However, both algorithms have been able to maintain a steady progress. Real-parameter GAs with BLX-0.5 also has progress towards optimum, but the progress is slow. The (10,100)-ES-SBX algorithm (with $\eta = 1$) is able to find better performance than GAs with the SBX operator.

Figure 21 shows the population average standard deviation in x_1 and x_2 for real-parameter GAs with SBX operator and the population average mutation strength in x_1 and x_2. In contrary to their evolution in the parabolic ridge F4-1 (Figure 19), here, these quantities in both algorithms reduce and fluctuate in a certain range. More importantly, these quantities for variables x_0 and x_1 now varies in the same range. This can also be explained as follows. In the rotated ridge function, the ridge axis $\vec{v}$ is a random direction, other than any coordinate direction. For the same reason, any orthogonal direction to the ridge axis is also a non-coordinate direction. In order to simultaneously satisfy both subgoals of maximizing progress towards the ridge axis and minimizing the distance to the ridge axis, the best an algorithm can do is to have a compromise in the rate of growth in each variable direction. In this case, both algorithms make careful changes to variables by keeping the population diversity or the mutation strength small so that a compromise in both subgoals is achieved. Naturally, this reduces the overall progress rate along the ridge axis. However, it may be mentioned here that such a linear behavior of rotated ridge functions is inevitable for any self-adaptive strategies which work with adaptations in variable directions independently. For a strategy which has the capability to independently adapt variations in arbitrary directions (not necessarily the coordinate directions), an exponential progress towards optimum with generation number may be possible (Hansen and Ostermeier, 1998).

6 Future Studies

This study suggests a number of extensions, which are outlined in the following:

1. Real-parameter GAs with SBX operator can be compared with other self-adaptive ES implementations.

2. Other probability distributions, such as lognormal probability distribution, can also be investigated for self-adaptive behavior in real-parameter GAs.
3. Other existing real-parameter crossover operators can be investigated for their self-adaptive behavior.
4. Properties for an efficient crossover operator can be developed for real-parameter GAs.
5. A generic real-parameter crossover operator with more than two parents can be investigated for a faster progress rate.
6. A real-parameter crossover efficient for problems having correlated interactions among object variables can be investigated.
7. Real-parameter GAs can be compared with self-adaptive ESs in real-world complex search and optimization problems.
8. Properties for an algorithm to exhibit self-adaptation and test suites can be developed for testing self-adaptive feature of an algorithm.
9. Discrete programming with real-parameter GAs with a modified SBX operator can be investigated for self-adaptation.

We discuss the above extensions in somewhat more details.

In this study, we have shown that the real-parameter GAs with SBX operator has the self-adaptive behavior, similar to that in a self-adaptive ES with lognormal update of self-adaptive parameters. Other self-adaptive ES implementations also exist and it will be interesting to compare the performance of real-parameter GAs with SBX operator with them. Of them, the ES implementations by Hansen and Ostermier (1995) can be investigated.

It is intuitive that there is nothing special about the polynomial probability distribution used in the SBX operator. However, it is important that the properties of the SBX operator described in section 2 must be preserved in a probability distribution for crossover. Thus, other more standard probability distributions used in ES, such as lognormal distribution can be used. In this regard, the following probability distribution as a function of β can be investigated, instead of equation 3:

$$\mathcal{P}(\beta) = \frac{1}{\sqrt{2\pi}\tau}\frac{1}{\beta}\exp\left(-\frac{1}{2}\frac{(\ln\beta)^2}{\tau^2}\right), \quad \beta > 0. \tag{25}$$

This probability distribution has its mode (maximum) at $\beta = \exp(-\tau^2)$, mean at $\bar{\beta} = \exp(\tau^2/2)$, and exactly 50% probability of finding a $0 < \beta \leq 1$ and the rest 50% probability of finding $\beta > 1$. The above distribution has a variance $\sigma_\beta^2 = \exp(\tau^2)(\exp(\tau^2) - 1)$. Since all statistical properties for this distribution are known, it may be easier to compare the performance of such crossover operators with self-adaptive ES which also uses lognormal update rule.

Besides the SBX operator, there exist other real-parameter crossover operators such as BLX-α (Eshelman and Schaffer, 1992) and UNDX operators (Ono and Kobayashi, 1997) which can be investigated for self-adaptive behavior. We have investigated BLX-0.5 operator in some test problems in this study and our findings are not positive. However, BLX-α with other α values can be tried for their self-adaptive behavior. The UNDX operator creates children solutions proportional to the parent solutions, but gives preference to solutions that are near the mean of the parent solutions. This property is contrary to what SBX operator does, and it will be an interesting study to investigate whether real-parameter GAs with the UNDX operator has the adequate self-adaptive behavior.

Studies of self-adaptation with various real-parameter crossover operators will reveal and allow us to find properties which are needed in an efficient real-parameter crossover operator. Such properties will help us to create problem specific crossover operators, if needed. The evidence of self-adaptive behavior of real-parameter GAs with SBX operator in this study suggests that, besides the properties mentioned in Section 2, the following are also important properties that a self-adaptive real-parameter GA should have in their search operator:

- The crossover operator must produce children population which has the same mean as that in the parent population.
- The variance of the resulting children population may be larger than that of the parent population.

The first property distinguishes the crossover operator from the selection operator. The primary task of a crossover operator is to search the region represented by the parent solutions. There is no reason why a crossover operator should have any bias towards any particular region in the search space. It is precisely the task of the selection operator to guide the search towards better regions in the search space. The second property helps to maintain a balance between the spread of solutions under selection and crossover operators. Since selection emphasizes good solutions by eliminating bad solutions in a population, it may, in general, reduce the variance of the population. If the crossover operator also has a tendency to reduce the variance of the population, the overall search algorithm may not have adequate power to adapt to any function landscape. Thus, it may be desirable to have a crossover operator which, in general, increases the variance of the parent population.

In self-adaptive ES studies, it has been observed that using multi-parent crossovers provides better local convergence properties. In this study, we have only confined crossovers having two parents. However, a suitable extension to the SBX operator with more than two parents may lead to an algorithm with a faster progress rate. In the following, we outline one such possible algorithm. It has been discussed earlier that in the SBX operator the variance of children distribution depends on the distance between the two parents. It is also important to use a distribution which assigns more probability for creating near-parent solutions than solutions away from the parents. It may not be of significant importance what distribution is actually used—whether the polynomial distribution used in this study or any other distribution with the above property. In order to be closer with mutation operators used in ES studies, a normal distribution with zero-mean and a standard deviation proportional to the distance between two parents can be used, instead of the polynomial distribution. In such a scheme, a parent is first identified. A second parent is randomly chosen to compute the distance (either variable-by-variable or vector-wise) between both parents. Thereafter, a child is created with a normal distribution having its mean at the first parent and standard deviation proportional to the distance. In principle, such a crossover operator (it will still be a crossover operator, because more than one parents will be used to create a child solution) should also have the self-adaptive power. Once such a principle is established, a multi-parent (more than two parents) crossover can be designed. The distance measure used to define the standard deviation for the normal distribution can be computed as a weighted sum of the distances multiple parents have from the parent being mutated.

In solving the correlated functions and generalized ridge functions, it is observed that progress towards the optimum is linear, instead of exponential, to the generation number. One way to speed up the progress would be to use a correlated SBX operator which exploits the pair-wise interactions among variables. One such procedure would be to first use variable-by-variable SBX crossover. Thereafter, the resulting children solutions can be modified further by performing a line SBX operator on pairs of variables. In a line SBX operator, children solutions are found along the line joining the two parents. Such a pair-wise crossover for variables can allow progress of solutions in directions where correlated interactions exist.

The primary reason for emphasizing the research in real-parameter optimizations using evolutionary optimization techniques is their use in real-world search and optimization problems of science and engineering. With the indication of self-adaptive behavior of real-parameter GAs in this study, such GAs can be tried to solve real-world search and optimization problems, where self-adaptation is an essential feature needed in an algorithm to solve a problem.

In this connection, it is important to note that there exists no study identifying properties that a search and optimization algorithm should have in order for it to be qualified as a self-adaptive algorithm. Research efforts in this direction is needed so as to develop better algorithms. Such a study may also help develop a suite of test problems for identifying self-adaptation properties in an algorithm. So far, self-adaptive evolutionary methods are applied only to a few simple functions. We have observed in this study that existing self-adaptive ES algorithms are vulnerable to changes in these test problems. A recent study (Grünz and Beyer, in press) has shown that the theoretically optimal parameter settings for a non-recombinative self-adaptive ES does not work well for recombinative ESs. Thus, there is a need of studies developing problems which will test various aspects of self-adaptation—convergence with arbitrary precision, adaptation to non-stationary functions, finding true optimum in functions with inadequate knowledge of the location of optimum, and others.

Since a discrete version of the SBX operator can be used to solve discrete programming problems (Deb and Goyal, 1997, 1998) efficiently, GAs can be investigated for their self-adaptiveness in discrete search space problems. The extension of SBX to discrete search space is simple and straightforward. Instead of using a continuous probability density function, a discrete probability distribution (allowing non-zero probabilities only to acceptable solutions, and keeping the shape of the distribution similar to that in SBX operator used in continuous search space) can be used to create children solutions. Rudolph (1994) used a self-adaptive ES on discrete search space problems. GAs with discrete-version of the SBX operator can be tried to check if GAs can also exhibit self-adaptation in those problems.

7 Conclusions

In this paper, we have demonstrated that real-parameter genetic algorithms (GAs) with simulated binary crossover (SBX) exhibit self-adaptive behavior on a number of test problems. In this respect, a connection between the self-adaptive evolution strategies (ESs) with lognormal update methods and real-parameter GAs with SBX operator has been discussed. A non-rigorous analysis has shown that both methods use similar probability distributions in creating children solutions, although a self-adaptive ES uses a single parent and a GA with SBX uses two parents. Applications of both methods in a number of test problems borrowed from the ES literature, including sphere models and ridge models, reveal the remarkable similarity in their performances. Based on this study, a number of extensions to this study have also been suggested.

Acknowledgments

The first author acknowledges the support provided by the Alexander von Humboldt Foundation, Germany during the course of this study. The second author acknowledges support as Heisenberg Fellow of the Deutsche Forschungsgemeinschaft under grant Be1578/4-1.

References

Bäck, T. (1997). Self-adaptation. In T. Bäck, D. Fogel, and Z. Michalewicz (Eds.) *Handbook of Evolutionary Computation*. New York: Oxford University Press.

Bäck, T. (1992). The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm. In R. Männer and B. Manderick (Eds.) *Parallel Problem Solving from Nature, II*, 85–94.

Beyer, H.-G. (1996). Toward a theory of evolution strategies: Self-adaptation. *Evolutionary Computation, 3*(3), 311–347.

Beyer, H.-G. (1995). Toward a theory of evolution strategies: On the benefit of sex—the $(\mu/\mu, \lambda)$-theory. *Evolutionary Computation, 3*(1), 81–111.

Deb, K. and Agrawal, R. B. (1995) Simulated binary crossover for continuous search space. *Complex Systems, 9* 115–148.

Deb, K. and Goyal, M. (1998). A robust optimization procedure for mechanical component design based on genetic adaptive search. *Transactions of the ASME: Journal of Mechanical Design, 120*(2), 162–164.

Deb, K., and Goyal. M. (1996). A combined genetic adaptive search (GeneAS) for engineering design. *Computer Science and Informatics, 26*(4), 30–45.

Deb, K. and Kumar, A. (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multi-modal and multi-objective problems. *Complex Systems, 9*(6), 431–454.

Eshelman, L. J. and Schaffer, J. D. (1993). Real-coded genetic algorithms and interval schemata. In D. Whitley (Ed.), *Foundations of Genetic Algorithms, II* (pp. 187–202).

Fogel, L. J., Angeline, P. J., and Fogel, D. B. (1995). An evolutionary programming approach to self-adaptation on finite state machines. In J. R. McDonnell, R. G. Reynolds, and D. B. Fogel (Eds.) *Proceedings of the Fourth International Conference on Evolutionary Programming*, 355–365.

Goldberg, D. E. (1991). Real-coded genetic algorithms, virtual alphabets, and blocking. *Complex Systems, 5*(2), 139–168. (Also IlliGAL Report 90001)

Goldberg, D. E., Deb, K., and Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems, 6*, 333–362.

Goldberg, D. E., Korb, B., and Deb, K. (1989). Messy genetic algorithms: Motivation, analysis, and first results, *Complex Systems, 3*, 93–530.

Gordon, V. S. and Whitley, D. (1993). Serial and parallel genetic algorithms as function optimizers. In S. Forrest (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 177–183).

Grünz, L. and Beyer, H.-G. (in press). Some observations on the interaction of recombination and self-adaptation in evolution strategies. Proceedings of the *Congress on Evolutionary Computation*.

Hansen, N. and Ostermeier, A. (October, 1998). Personal communication.

Hansen, N. and Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strageties: The covariance matrix adaptation. *Proceedings of the IEEE International Conference on Evolutionary Computation*, 312–317.

Hansen, N. and Ostermeier, A. (1997). Convergence properties of evolution strategies with the derandomized covariance matrix adaptation: The $(\mu/\mu_I, \lambda)$-CMA-ES. *European Congress on Intelligent Techniques and Soft Computing*, 650–654.

Harik, G. (1997). Learning gene linkage to efficiently solve problems of bounded difficulty using genetic algorithms (IlliGAL Report No. 97005). Urbana: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.

Herdy, M. (1992). Reproductive isolation as strategy parameter in hierarchically organized evolution stretegies. In R. Männer and B. Manderick (Eds.) *Parallel Problem Solving from Nature, II*, 207–217.

Johnson, N. L. and Kotz, S. (1970). *Continuous univariate distributions—1*, Boston: Houghton Mifflin.

Kargupta, H. (1996). The gene expression messy genetic algorithm. *Proceedings of the IEEE International Conference on Evolutionary Computation*. 814–819.

Kita, H., Ono, I., and Kobayashi, S. (1998). The multi-parent unimodal normal distribution crossover for real-coded genetic algorithms. Unpublished document.

Michalewicz, Z. and Janikow, C. Z. (1991). Handling constraints in genetic algorithms. In R. K. Belew and L. B. Booker (Eds.), *Proceedings of the Fourth International Conference on Genetic Algorithms*, 151–157.

Ono, I. and Kobayashi, S. (1997). A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover. *Proceedings of the Seventh International Conference on Genetic Algorithms*, 246–253.

Oyman, A. I., Beyer, H.-G., and Schwefel, H.-P. (1998). Where elitists start limping: Evolution strategies at ridge functions. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (Eds.) *Parallel Problem Solving from Nature, V*, 34–43.

Rechenberg, I. (1994). *Evolutionsstrategie'94*. Stuttgart: Frommann-Holzboog Verlag.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Stuttgart: Frommann-Holzboog Verlag.

Rudolph, G. (1994). An evolutionary algorithm for integer programming. In Y. Davidor, H.-P. Schwefel, and R. Männer (Eds.) *Parallel Problem Solving from Nature, III*, 139–148.

Saravanan, N., Fogel, D. B., and Nelson, K. M. (1995). A comparison of methods for self-adaptation in evolutionary algorithms. *BioSystems, 36*, 157–166.

Schraudolph, N. N. and Belew, R. K. (1990). Dynamic parameter encoding for genetic algorithms. Technical Report No. LAUR90-2795. Los Alamos: Los Alamos National Laboratory.

Schwefel, H.-P. (1974). Collective phenomena in evolutionary systems. In P. Checkland and I. Kiss (Eds.), *Problems of Constancy and Change—the Complementarity of Systems Approaches to Complexity*. 1025–1033.

Schwefel, H.-P. (1987). Collective intelligence in evolving systems. In W. Wolff, C-J Soeder, and F. R. Drepper (Eds.) *Ecodynamics, Contributions to Theoretical Ecology*, 95–100.

Shaefer, C. G. (1987). The ARGOT strategy: Adaptive representation genetic optimizer technique. In J. J. Grefenstette (Ed.), *Proceedings of the Second International Conference on Genetic Algorithms*, 50–58.

Weinberg, R. (1970). Computer simulation of a living cell. (Doctoral dissertation, University of Michigan). *Dissertations Abstracts International, 31*(9), 5312B. (University Microfilms No. 71-4766)

Wright, A. (1991). Genetic algorithms for real parameter optimization. In G. J. E. Rawlins (Ed.), *Foundations of Genetic Algorithms*. (pp.205–220).