Sample-based Crowding Method for Multimodal Optimization in Continuous Domain

Shin Ando
Yokohama National University
79-5 Tokiwadai, Hodogaya
Yokohama, Kanagawa, Japan
ando@slab.dnj.ynu.ac.jp

Einoshin Suzuki
Yokohama National University
79-5 Tokiwadai, Hodogaya
Yokohama, Kanagawa, Japan
suzuki@slab.dnj.ynu.ac.jp

Shigenobu Kobayashi
Tokyo Institute of Technology
4259 Nagatsutacho, Midori-ku
Yokohama, Kanagawa, Japan
kobayashi@fe.dis.titech.ac.jp

Abstract- We propose a selection scheme called Sample-based Crowding, which is aimed to improve the performance of Genetic Algorithms for multimodal optimization in ill-scaled and locally multimodal domains. These domains can be problematic for conventional approaches, but are commonly found in real-world optimization problems. The principle of Crowding is to apply a tournament selection to a parent-child pair with a high similarity. In the Sample-based Crowding, we determine such pairs based on a statistical comparison of the fitness values, which are sampled from the region between the pairs. Further, we take into account the ranks of the parents among the sampled values in the selection process, to determine their indispensability. These measurements are scale-invariant, which enables the proposed method to search a domain without presuming the distance between the optima or the scaling and the correlation of the variables.

The proposed approach is evaluated in two benchmark problems with an ill-scaled and a locally multimodal landscape. The proposed method has a substantial advantage in terms of comprehensiveness compared to the conventional approaches, despite the additional cost of evaluations.

1 Introduction

The purpose of the Multimodal Optimization (MO) is to obtain a set of optima/suboptima of a fitness landscape comprehensively or at an acceptable level of efficiency. It can provide solutions for many instances of real-world problems and design tasks, where preparing multiple solutions, preferably with varying features, can be a better strategy.

In nature, the process of evolution has produced a great diversity in its population and their functions. The researchers of Evolutionary Computation have attempted to reproduce and utilize such a mechanism in the framework of evolutionary algorithms. Their works have been successfully applied to many multimodal benchmark tasks [3, 8, 9, 5].

However, these conventional methods commonly report difficulties in following domains: a) locally multimodal domains, where optima and suboptima are closely located, and b) ill-scaled domains, where the variables have significantly different scales or strong correlations. In such domains, it is difficult to estimate the neighbors of an individual based on the Euclidean distance. Particularly, determining the critical parameters, e.g. sharing distance[5] or minimum species distance[8], which presume a distribution of the optima, can be troublesome. The above types of domains are commonly found in real-world problems and present difficulty for many optimization algorithms.

In this paper, we propose a Sample-based Crowding (SC) method, in which we calculate the similarity of a parent and a child based on the statistical comparison of the sampled fitness values instead of using the Euclidean distance. Thereafter, a tournament selection is applied to the most similar pair. Further, we introduce a 'indispensability' index for the parent individuals, which is based on the ranks of a parent's fitness among the sampled values, to determine whether the parents can be replaced.

The above measurements are invariant to the affine transformation of the domain, which enables the proposed methods to search the problem domain without the user's estimate of a) the distance between the optima, and b) the scaling and the correlation of the variables.

This paper is organized as follows: In Section 2, we refer to the related works of the multimodal optimization by evolutionary algorithms and the important concepts introduced by them. In Section 3, we describe the characteristics of the problem domains, which can be problematic for the conventional approaches. In Section 5, we describe the implementation of our proposed method. In Section 6, we evaluate the proposed method in two artificial benchmark problems, and discuss its performance. Section 7 gives our conclusion and the summary of this work.

2 Related Works

In this paper, we refer to the convex around a suboptimum as its attractor. (e.g., Figure 1 shows P and Q on one attractor and R on another). When studying the behavior of GAs in a continuous domain, it is important to consider the fitness landscape's attractors as well as its optima. The attractors can have different characteristics which can largely affect the GA's behavior[6, 12].

To perform a multimodal search in a landscape with multiple optima/suboptima, GA must distribute its population among different attractors. This is referred to as the Niching of the population.

When multiple attractors exist, a large number of individuals are drawn to relatively stronger attractors, in turn, abandoning weaker attractors. The characteristics which make the attractors weak or strong are discussed in [6].

0-7803-9363-5/05/$20.00 ©2005 IEEE.

Here, we will refer to two groups of multimodal evolutionary algorithms, each of which employ important concepts in multimodal optimization: 1)a modified selection approach and 2) a multi-populational approach.

The first group of methods, including Crowding[2] and its variations, i.e., Deterministic[9] and Probabilistic Crowding[10], modifies the selection procedure to induce the Niching behavior of GA. The general concept of Crowding is to have the individuals compete for survival against a similar offspring. This, in effect, prevents individuals in the weaker attractors to be replaced by one in the stronger attractors. Fitness Sharing[5] and Species Conserving GA[8] also induce the Niching behavior by modifying the process of the survival selection, therefore we categorize them into this group. It is important to note that the above approaches use a standard reproduction procedure, thus maintain the global search of the standard GA. Consequently, their efficiency decreases significantly in a highly multimodal domain.

The second group of methods, which includes Multi-National GA[15], Innately Split Model[6], Clustering Based Niching[14], employ a multi-populational implementation. These methods attempt to assign one population per attractor, thus attempt to localize the search within each attractor. The Multi-populational approach can increase the speed of convergence once the subpopulations are properly distributed among the attractors in the landscape. However, this type of approach lacks a reliable methodology to distribute the subpopulations over multiple attractors comprehensively and efficiently.

Our proposed method belongs to the first group, and is aimed to improve the comprehensiveness of the Niching behavior of Crowding, by using a robust estimation of the similarity among the individuals.

3 Problematic Domains for Multimodal Optimization

In this section, we will describe two types of domains that can be problematic for the multimodal optimization GA.

3.1 Ill-scaled Domain

In real-world problems, some variables may have significant sensitivity to the fitness function than others or strong correlations among them. More formally, we will give the following terms. Assume a fitness function $f(\mathbf{x}|\mathbf{x} \in \mathbf{R}^D)$, whose approximation within the attractor of optima $\mathbf{x}_0$ can be given by a hyper-parabolic function (1)

$$\begin{aligned} f(\mathbf{x}-\mathbf{x_0}) &= (\mathbf{A}(\mathbf{x}-\mathbf{x_0}))^T\mathbf{A}(\mathbf{x}-\mathbf{x_0}) \\ &= ((\mathbf{x}-\mathbf{x_0}))^T\mathbf{H}(\mathbf{x}-\mathbf{x_0}) \end{aligned} \tag{1}$$

where $\mathbf{A}$ is a full rank matrix.

Let $\mathrm{A} = \{\alpha_1, \ldots, \alpha_n\}$ be the Eigen values of $\mathbf{H}$. If the attractor is ill-scaled, there are significant differences among the Eigen values, i.e.,

$$\exists\, \alpha_i, \alpha_j \in \mathrm{A}\,|\alpha_i \ll \alpha_j$$

This type of attractor is called a k-tablet structure in [12].

We will consider a domain with one or more ill-scaled attractors as an ill-scaled domain. It is important to note that the Eigen values and vectors are generally different for each attractor. Which means, the attractors in different parts of the domain may have different variables that are ill-scaled or highly correlated, thus it is impractical to correct the ill-scaled landscape by re-scaling the variables.

An ill-scaled domain can be problematic for multimodal optimization methods that employ the Euclidean distance to determine the similarities among the individuals or to presume the distance between the optima. For example, in Figure1, it is difficult, even with *a priori* knowledge of the domain, to choose an optimal value of sharing distance for Fitness Sharing or minimum species distance for Species Conserving GA based on guidelines given in [5] and [8, 3].

The above parameters represent an estimate of the distance between the optima, therefore determine the 'resolution' of the search. While a search with a high resolution, i.e., using small values for the above parameters, is suitable for distinguishing the two optima in Figure 1, it will induce unnatural Niching within the convex of the attractor. A search with a low resolution, i.e., assuming a large distance between optima, will result in obtaining only one of the two optima.

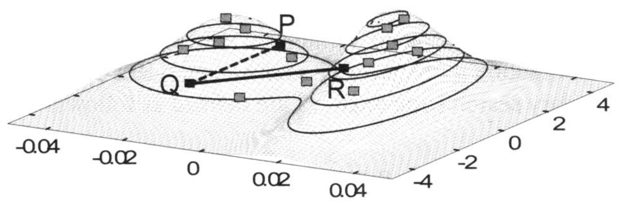

Figure 1: Pop. Snapshot on Ill-scaled Attractors

For Crowding methods, an individual can often be further apart from the individuals on the same attractor, e.g. P and Q in Figure 1, than from many individuals on another attractor, e.g. R, in terms of the Euclidean distance. As a result, the Niching behavior of Crowding is nullified.

On the other hand, use of measurements such as Mahalanobis' generalized distance in multimodal domain, require a mixture of models that can represent a GA population distributed over the multiple attractors. Preparation of such models can be difficult and time consuming, and may be unsuitable for an implementation with GA.

3.2 Local multimodality

A locally multimodal domain contains 'clusters' of optima and suboptima. Distinguishing each suboptimum within such a 'cluster' is commonly difficult for multimodal GAs. For example, Shubert Function has several suboptima closely located near the optima, which are difficult to find by multimodal GAs [8].

The difficulty comes from having to estimate the distance between the optima using a single distance parameter. A locally multimodal landscape presents a dilemma for the level of resolution similar to the one described in Section

3.1. A search with a high resolution is capable of distinguishing the suboptima within a 'cluster', but can severely detriment the efficiency of the search. On the other hand, a search with a low resolution will compromise the comprehensiveness for the efficiency, and may result in finding only one suboptimum within a cluster.

4 Basic Implementations

Our proposed method employs the Real-coded GA implementation proposed by [7] and [13] as basic GA operators. This section gives a brief overview of their works.

4.1 Genetic Operator

We implement ENDX[7], which is an extension of UNDX[11], as the crossover operator of real number vectors. It has been successful in many real-number benchmark problems. ENDX generates each offspring $\mathbf{y} \in \mathbf{R}^D$ probabilistically from parents $\{\mathbf{x_1}, \ldots, \mathbf{x_m}\}$.

$$\mathbf{y} = \mathbf{g} + \xi \mathbf{d} + \frac{1}{m-2} \sum_{i=3}^{m} \eta_i \mathbf{x}'_i \tag{2}$$

where, $\mathbf{x}_i$ and $\mathbf{y}$ are D-dimensional vectors, n is the dimension of the domain, $m = n + 2$ is the number of parents used in a crossover, and $\xi = N\left(0, \alpha^2\right)$, $\eta = N\left(0, \beta^2\right)$ are normally distributed variables. Further, the center of the distribution $\mathbf{g}$, the primary search component $\mathbf{d}$, and the secondary search component $\mathbf{x}'_i$ are defined as follows.

$$\mathbf{g} = \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}$$

$$\mathbf{d} = \mathbf{x}_1 - \mathbf{x}_2$$

$$\mathbf{x}'_i = \mathbf{x}_i - \frac{1}{m-2} \sum_{j-3}^{m} \mathbf{x}_j$$

Moreover, the values for the parameters of the crossover α and β, are suggested as follows.

$$4\alpha^2 + 2\left(m-3\right)\beta^2 = 1,\ \beta = \frac{0.35}{\sqrt{m-3}} \tag{3}$$

4.2 Replacement Scheme

Our replacement scheme is based on Minimal Generation Gap (MGG) [13]. Following is the procedure for one cycle of the replacement scheme. Step 3 is described in more detail in Section 5.1.

1. Randomly select m parents from the population P.
2. Generate λ offspring by iterating the crossover operation.
3. Apply the selection to the best child and the neighboring parent(s).

5 Proposed Methods

This section describes our proposal of Sample-based Crowding scheme, in which we attempt to select a neighboring parent-child pair based on a statistical comparison of the sampled fitness values, instead of the Euclidean distance. Further, the ranks of parents' fitness among the samples are used to evaluate the parent's indispensability. The above measurements are consistent to the linear transformation, therefore induce SC's robustness in an ill-scaled domain.

5.1 Sample-based Crowding

Assume a maximization problem whose fitness landscape are as shown in Figure 1. We can see intuitively that the individuals P and Q are neighbors, while R is not, as the latter is on a different attractor. More formally, we consider the mean fitness M_s over path s

$$M_s = \frac{\int_s f\left(\mathbf{x}\right) d\mathbf{x}}{\int_s d\mathbf{x}}$$

where $f\left(\mathbf{x}\right)$ is the fitness function.

If M_{PQ} is significantly larger (or smaller in minimization) than M_{QR}, the pair is more likely to be on the same attractor than the other.

Since the integral of the fitness function is usually incalculable, it is more practical to compare the means statistically by sampled fitness values. Consider $\mathbf{f}_{PQ}$ as the fitness values sampled from a normal distribution around the path PQ. Such $\mathbf{f}_{PQ}$ can be obtained as the fitness values of the offspring, obtained by the ENDX crossover of P and Q.

After obtaining $\mathbf{f}_{PQ}$ and $\mathbf{f}_{QR}$, we can compare the mean fitness $\bar{\mathbf{f}}_{PQ}$ and $\bar{\mathbf{f}}_{QR}$ using the Wilcoxon's Rank-sum Test[4]. We denote the results of the Rank-Sum Tests as follows:

- $\mathbf{f}_{PQ} \underset{\gamma}{\ll} \mathbf{f}_{QR}$ denotes that $\bar{\mathbf{f}}_{PQ}$ is better than $\bar{\mathbf{f}}_{QR}$ with the significance level of γ, and
- $\mathbf{f}_{PQ} \underset{\gamma}{\approx} \mathbf{f}_{QR}$ indicates neither of the two is better than the other with the significance level of γ.

The pair with the higher mean fitness is considered as neighbors, thus competes for survival in the tournament selection.

In the conventional Crowding, the similarities of several pairs are evaluated before a selection is applied to the most similar pair. However, in SC, due to the large number of evaluations required, we only evaluate the similarity of parent p_0 and the best child c and that of p_1 and c.

After each crossover, the similarity among the best child c and the two parents p_0 and p_1 are evaluated as follows. First, we crossover p_0 and c to obtain $\mathbf{K}_0 = \{k_{0i} | i = 1, \ldots, \eta\}$, which are fitness values of their offspring. Next, we obtain $\mathbf{K}_1 = \{k_{1i} | i = 1, \ldots, \eta\}$ by the crossover of p_1 and c. Note that the variance parameters are reduced in order to sample from a region close to the path between c and p_0. Tournament selection is applied to c and p_i if $K_i \gg K_j$.

While $\mathbf{K}_0 = \{k_{0i} | i = 1, \ldots, \eta\}$ may include better fitness values than that of $\{p_0, p_1, c\}$, we do not consider the samples for replacement, as we currently do not have the means to assess the attractors they belong to. However, it is part of our future work to devise such a methodology, in order to improve the efficiency of the proposed method.

Further, we evaluate the 'indispensability' of the parents p_i by their ranks. R_i denotes the rank of p_i's fitness among $\mathbf{K}_i$. p_i is not subject to replacement if R_i is above the threshold R_{th}. This index is introduced to preserve the parents when all of the parents and the child are on different attractors.

The summary of Sample-based Crowding procedure is as follows:

1. Crossover $p_0, p_1, \ldots, p_m$ and select the best offspring c.
2. Crossover $p_0, c, \ldots, p_m$ to samples the fitness values $\mathbf{K}_0$.
3. Crossover $c, p_1, \ldots, p_m$ to sample the fitness values $\mathbf{K}_1$.
4. If $\mathbf{K}_0 \underset{\gamma}{\ll} \mathbf{K}_1$, apply the tournament selection to p_1 and c.
5. If $\mathbf{K}_0 \underset{\gamma}{\gg} \mathbf{K}_1$, apply the tournament selection to p_0 and c.
6. Otherwise, consider the ranks R_0, R_1.

In step 6, p_i is dispensable when $R_i > R_{th}$, and the elitist selection is applied to the replaceable parent(s) and c. Note that the replacement does not occur if neither of the parents are replaceable.

Figure 2 shows the pseudo code of Sample-based Crowding. $ENDX(\{p_i\})$ denotes the crossover of parents $\{p_i\}$ based on (2).

$parents\{p_0, p_1, \ldots, p_{m-1}\} \subset Generation_j$

$for(i = 1\ to\ \lambda)$

do

    $c_i = ENDX(p_0, p_1, \ldots, p_{m-1})$

$done$

$c = \arg\max_{c_i} f(c_i)$

$for(i = 1\ to\ \eta)$

do

    $k_{0i} = f(ENDX(p_0, p_c, \ldots, p_{m-1}))$

    $k_{1i} = f(ENDX(p_0, p_c, \ldots, p_{m-1}))$

$done$

$\mathbf{K}_0 = \{k_{0i}\}, \mathbf{K}_1 = \{k_{1i}\}$

$R_0 = rank(f(p_0), \mathbf{K}_0), R_1 = rank(f(p_1), \mathbf{K}_1)$

$if\left(\mathbf{K}_a \underset{\gamma}{<<} \mathbf{K}_b\right)$

$then$

    $s_0 = tournament(p_b, c),\ s_1 = p_a$

$else$

    $if(R_0, R_1 > R_{th})$

    $then$

        $s_0 = tournament(p_0, p_1, c)$

        $s_1 = tournament(\{p_0, p_1, c\} - s_0)$

    $else$

        $if(R_i > R_{th}, R_j < R_{th})$

        $then$

            $s_0 = tournament(p_i, c),\ s_1 = p_j$

        $else$

            $s_0 = p_0,\ s_1 = p_1$

        fi

    fi

fi

$Generation_{j+1} = Generation_j - \{p_0, p_1\} + \{s_0, s_1\}$

Figure 2: Pseudo Code of Sample based Crowding

6 Experiments

We compare Deterministic Crowding (DC) [9], Probabilistic Crowding (PC) [9] and Sample-based Crowding (SC) in two benchmarks with properties described in Section3. We note that, since the proposed method exploit the scale-invariant measurements, its results are the same as described below in the benign-scaled domain problems as well. The values suggested in [7] are used for the parameters of the ENDX crossover, i.e., the number of parents $\mu = D + 2$, the number of offspring $\lambda = 100$. D is the dimension of the problem domain. Further, no mutation is used, therefore the reproduction is done solely by the crossover.

6.1 Locally Multimodal Domain

Shubert Function (4) is often used as a benchmark for the locally multimodal problem [8].

$$F_1(x) = \prod_{i=1}^{D} \sum_{j=1}^{5} j \cos((j+1) x_i + j) \tag{4}$$

where D is the dimension of the domain.

Within the range $(-4 \le x_i \le 8)$, it has $66 \times 2^D + 1$ suboptima/optima distributed symmetrically. There are 2^D pairs of global optima, each pair within a cluster of 4^D suboptima. Figures 3 and 4 show the contour of F_1 in light lines, along with a snapshot of GA population indicated by dark '+'.

Sample-based Crowding and DC are applied to 2-dimensional F_1. The GA parameters are: the population size $M = 500$ and the termination step $T = 30,000$. The performance of the two methods are robust to the small changes in these parameters. The parameters of SC are empirically set to the significance level $\gamma = 0.1$, the rank threshold $R_{th} = 0.6$, the sample size $\eta = 100$. Generally, more stringent values, i.e., a smaller γ and a larger R_{th}, cause the search to be slower but more comprehensive.

Figures 3 and 4 show the populations of Sample-based Crowding and DC after 30,000 selections from a typical run (the individuals are shown as dark '+', and the contours of the fitness functions are shown by the light lines). Note that the number of evaluations is larger for SC than DC, as SC evaluates the sampled individuals in order to determine the neighbors as described in section 5.1. In DC, the parents on the suboptimal attractors are repeatedly replaced when the nearest offspring is created on a nearby optimal attractor. As a result, the final population merely occupies the optimal attractors and few strong suboptimal attractors.

On the other hand, the result of our Sample-based Crowding shows a larger number of the suboptimal attractors occupied by a set of individuals, thus is more comprehensive compared to that of DC. The unoccupied suboptimal attractors in Figure 3 were never occupied throughout the run. To obtain a comprehensive set of suboptima, we should increase the size of the population to an extent where high probability of sampling is assured for each of the attractors.

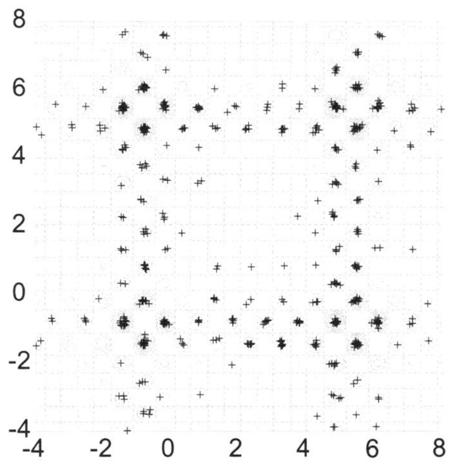

Figure 3: Snapshot of Sample-based Crowding population in F_1. The dark '+' indicate the individuals, and the light lines indicate the contours of the fitness functions.

Further, we applied three Crowding methods to 2, 5 and 10-dimensional F_1. Here, we use the population size $M = 250 \times D$, and the termination step $150,000 \times D$. Other parameters use the same values from the previous experiment. After the termination of each run, we ran a hill-climbing algorithm ($(1+1)$ES[1] with the mutation rate of 0.01) starting from each individual. After all the iterations converged, we count the number of suboptima in F_1, each of which has more than one individual within the distance of 0.01. The coordinates of the suboptima of F_1 were identi-

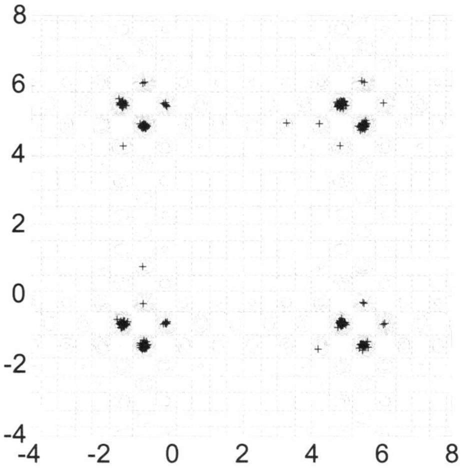

Figure 4: Snapshot of DC population in F_1. The dark '+' indicate the individuals, and the light lines indicate the contours of the fitness functions.

fied prior to the experiment. In each run, the number of evaluations for DC and PC is $D \times 15 \times 10^6$, while $D \times 45 \times 10^6$ for SC. The number of evaluations for SC is larger, since it includes the evaluation of the samples for identifying the neighbors and indispensability of individuals.

Table 1 shows the number of optima or suboptima found in the experiment using Deterministic, Probabilistic, and Sample-based Crowding in 10 runs. It shows that the proposed approach can maintain a significantly larger number of suboptima than the conventional methods. This is significant when the number of global optima is much smaller than the number of individuals, in 2-dimensional and 5-dimensional domain. The results of DC and PC for 10-dimensional Shubert function are rather misleading, since DC and PC are incapable of finding the suboptima in higher dimension as well. However, by finding the global optima, there exists 2^{11} of which, they can close in on SC's performance on the number of optima or suboptima found.

	D=2	D=5	D=10
SC	123 ± 12.0	320 ± 39.8	$146 \times 10^1 \pm 206$
DC	21.1 ± 2.85	151 ± 24.1	$121 \times 10^1 \pm 196$
PC	20.0 ± 2.39	132 ± 4.90	$119 \times 10^1 \pm 272$

Table 1: The number of optima/suboptima found by Crowding Methods with Hill-climbing in 5-Dimensional Shubert Function. The Sample-based Crowding can maintain a significantly larger number of suboptima than Deterministic or Probabilistic Crowding. The results of 10-dimensional Shubert function are rather misleading, since the conventional methods close in on proposed methods' number of optima or suboptima found, due to the large number of the global optima, there exists 2^{11} of which, even though they are incapable of finding the suboptima.

6.2 Ill-scaled Domain

First, we propose an artificial benchmark problem with the property of an ill-scaledness, a nonlinear dependence, and a multimodality. We defined a two-peak Rosenbrock function F_2 (6), derived as the product of Rosenbrock's function (5), as such benchmark. This is a minimization problem where the proper estimation of the distance between the optima is difficult. Note that its multimodality is relatively low, thus the multimodal GAs should be able to converge without the aid of greedy algorithms.

F_2 (6) has one dominant parabolic attractor and one weaker parabolic attractor, both of which are placed parallel to each other. The optimal value of $F_2(\vec{x}) = 0$ is found along the dominant attractor and at a point $\mathbf{x}$ ($x_i = 1$) in the weaker attractor.

$$F_R(\vec{x}) = \sum_{i=2}^{n} \left\{100\left(x_1 - x_i^2\right)^2 + (x_i - 1)^2\right\} \quad (5)$$

$$F_2(\vec{x}) = F_R(\vec{x}) \times \sum_{i=2}^{n} \left\{100\left(x_1 - (x_i - 4)^2\right)^2\right\} \quad (6)$$

We assess the performance of Sample-based Crowding and DC for searching the weak attractor by recording $F_R(y)$ for each generation's best individual y. $F_R(y)$ is roughly the best fitness found in the weaker attractor.

We ran Sample-based Crowding and DC 10 times respectively. Each run was terminated after 1.0×10^6 evaluations. We used the population size $M = 40$ and the offspring size λ=100. The parameters for Sample-based Crowding are the same from the previous experiment.

Figure 5 and 6 show the convergence of $F_R(y)$ for 10 runs of DC and Sample-based Crowding. The results of DC runs show an increase in fitness value, which indicates that the best individual in the weaker attractor was replaced by one in the stronger dominant attractor. In 8 out of 10 runs, the weaker attractor was completely abandoned. Meanwhile, Sample-based Crowding maintained a steady improvement at various convergence speed depending on the number of individuals on the weaker attractor. It reached the exact optima (1,1) within the observed time frame in two runs .

6.3 Parameter Setting

The sample size η is an important parameter in the proposed approach. In this section, we performed a grid search for the parameter η. We applied Sample-based Crowding to $D = 5, 10, 15, 20, 25$ dimensional F_2 function. Here we tried $\eta = 100, 50, 20$ for the sample size parameter of SC. Other parameters are maintained from the previous experiment. 10 runs are conducted for each D and η. We set an objective fitness value $f_o = 1.0 \times 10^{-5}$. Each run is terminated when the best individual reaches f_o or the number of evaluated solutions reaches 1.0×10^{10}. For SC, it includes the evaluation of the samples for identifying the neighbors and the indispensability of the individuals. We ran SC 10 times, and counted the number of runs where

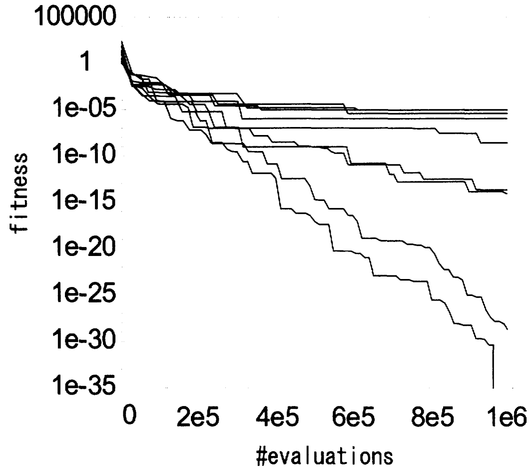

Figure 5: Best fitness value on weaker attractor in Sample-based Crowding

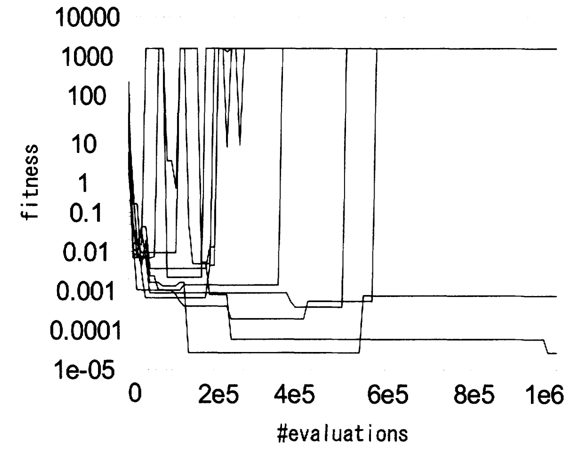

Figure 6: Best fitness value on weaker attractor in DC

$F_R(y)$ reached f_o within the limit of the number of evaluations. We denote this number by $\#(f_o)$. We will note that neither Deterministic Crowding nor Probabilistic Crowding were able to obtain f_o.

Figure 7 shows the $\#(f_o)$ when using the different values of η for dimensions $D = 5$ to 25. It shows that the increase in dimension does not necessary cause a significant increase in the number of sample size required for Sample-based Crowding. The number of evaluations required for SC is $(1 + \frac{\eta}{\lambda})$ times that of DC. In this experiment, $\eta = 20$ is sufficient for optimizing the 15-dimensional F_2, in which case the ratio of the number of evaluations for SC and DC is 1.2. It is our conjecture, that η is affected more by the complexity of the domain than its dimensionality. Additionally, this experiment shows that Sample-based Crowding does not affect the capabilities of ENDX to search a high-dimensional, nonlinear problems, and also that its parameters can be chosen independently from other GA parameters.

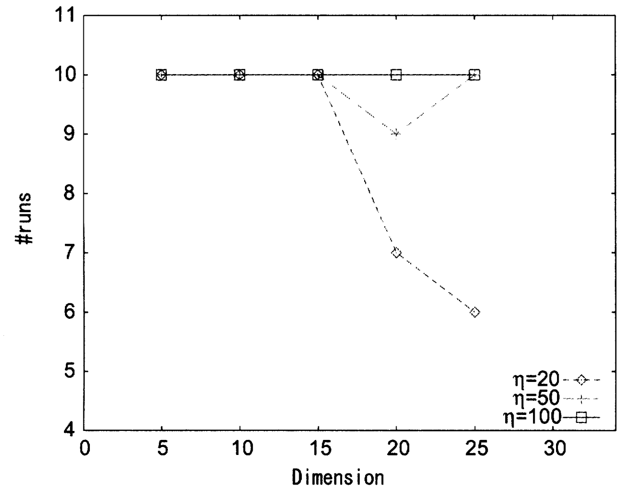

Figure 7: Performance of Sample-based Crowding with Different Sample-sizes η in D-dimensional Two-peak Rosenbrock function

7 Conclusion

In this paper, we proposed a Sample-based Crowding method, in which we introduced a scale-invariant measurement to determine the neighboring individuals. This allowed the proposed method to run without a presumption on the distance between the optima or the scaling and the correlations of the variables. In two artificial benchmark problems with an ill-scale and a local multimodality characteristics, our proposed approach showed the advantage to the conventional approaches in terms of comprehensiveness, i.e., obtaining more suboptima in one run.

It is probable in real-world problems, that a set of desirable suboptima exists near a global optimum. The results of the experiments in Sections 6.1 and 6.2 indicate that the conventional methods converge their populations to few global optima in such domains. Therefore, to maintain useful suboptima in their populations, the user must carefully observe and terminate the runs before the suboptima are eliminated. It is more desirable for a multimodal optimization algorithm to maintain a set of useful suboptima along with the global optima, as demonstrated by the proposed method.

The increase in the number of evaluations is an intrinsic characteristic of our approach. However, as our experiments in Section 6.2 showed, the disadvantage is not significant and can also be offset by the increased comprehensiveness. By obtaining a larger number of suboptima in a single run, the proposed method requires a smaller number of runs to obtain a more comprehensive set of optima and suboptima than the conventional methods. As our future work, we intend to apply the proposed method to more complex and practical problems.

Bibliography

[1] T. Back. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms.* Oxford University Press, 1996.

[2] K. A. De Jong. An analysis of behavior of a class of genetic adaptive systems. *Ph.D. thesis, University ofMichigan*, 1975.

[3] K. Deb and D. E. Goldberg. An investigation of niche and species formation in genetic function optimization. In J. D. Schaffer, editor, *Proceedings of the Third InternationalConference on Genetic Algorithms*, pages 42–50, SanMateo, California, 1989. Morgan Kaufmann.

[4] J. L. Devore. *Probability and Statistics for Engineering and Sciences.* Duxbury Press, 1995.

[5] D. E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimization. In *Proceedings of the Second International Conference on Genetic Algorithms*, pages 41–49, Hillsdale, New Jersey, 1987.

[6] K. Ikeda and S. Kobayashi. GA based on the UV-structure hypothesis and its application to JSP. *Lecture Notes in Computer Science*, 1281:273–282, March 2000.

[7] S. Kimura, I. Ono, H. Kita, and S. Kobayashi. An extension of UNDX based on guidelines for designing crossover operators: Proposition and evalution of ENDX (in Japanese). *Transactions of the Society of Instrument and Control Engineers*, 36:1162–1171, 2000.

[8] J. Li, M. Balazs, G. Parks, and P. Clarkson. A genetic algorithm using species conservation for multimodal function optimization. *Evolutionary Computation*, 10(3):207–234, 2002.

[9] S. W. Mahfoud. Crowding and preselection revisited. In R. Männer and B. Manderick, editors, *Parallel*

problem solving from nature 2, pages 27–36, Amsterdam, North-Holland, 1992.

[10] O. Mengshoel and D. Goldberg. Probabilistic crowding: Deterministic crowding with probabilistic replacement. pages 409–416. Morgan KauRmann, 1999.

[11] I. Ono and S. Kobayashi. A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover. In *Proceedings of 7th International Conference on Genetic Algorithms*, pages 246–253, 1997.

[12] J. Sakuma and S. Kobayashi. k-tablet structure and crossover on latent variables for real-coded GA. In *Proceedings of GECCO2002, Late Breaking Paper*, 2002.

[13] H. Satoh, M. Yamamura, and S. Kobayashi. Minimal generation gap model for GAs considering both, exploration and exploitation. In *Proceedings of IIZUKA: Methodologies for the Conception, Design, and Application of Intelligent Systems*, pages 494–497, Singapore, 1996. World Scientific.

[14] F. Streichert, G. Stein, H. Ulmer, and A. Zell. A clustering based niching ea for multimodal search spaces. In *Proceedings of Evolution Artificielle (LNCS 2935)*, pages 293–304. Springer-Verlag, 2003.

[15] R. K. Ursem. Multinational gas: Multimodal optimization techniques in dynamic environments. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 19–26. Morgan Kaufmann, 2000.