Adaptive Elitist-Population Based Genetic Algorithm for Multimodal Function Optimization

Kwong-Sak Leung and Yong Liang

Department of Computer Science & Engineering,
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
{ksleung, yliang}@cse.cuhk.edu.hk

Abstract. This paper introduces a new technique called adaptive elitist-population search method for allowing unimodal function optimization methods to be extended to efficiently locate all optima of multimodal problems. The technique is based on the concept of adaptively adjusting the population size according to the individuals' dissimilarity and the novel elitist genetic operators. Incorporation of the technique in any known evolutionary algorithm leads to a multimodal version of the algorithm. As a case study, genetic algorithms(GAs) have been endowed with the multimodal technique, yielding an adaptive elitist-population based genetic algorithm(AEGA). The AEGA has been shown to be very efficient and effective in finding multiple solutions of the benchmark multimodal optimization problems.

1 Introduction

Interest in the multimodal function optimization is expanding rapidly since real-world optimization problems often require the location of multiple optima in the search space. Then the designer can use other criteria and his experiences to select the best design among generated solutions. In this respect, evolutionary algorithms(EAs) demonstrate the best potential for finding more of the best solutions among the possible solutions because they are population-based search approach and have a strong optimization capability. However, in the classic EA search process, all individuals, which may locate on different peaks, eventually converge to one peak due to genetic drift. Thus, standard EAs generally only end up with one solution. The genetic drift phenomenon is even more serious in EAs with the elitist strategy, which is a widely adopted method to improve EAs' convergence to a global optimum of the problems.

Over the years, various population diversity mechanisms have been proposed that enable EAs to maintain a diverse population of individuals throughout its search, so as to avoid convergence of the population to a single peak and to allow EAs to identify multiple optima in a multimodal domain. However, various current population diversity mechanisms have not demonstrated themselves to be very efficient as expected. The efficiency problems, in essence, are related to

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2723, pp. 1160–1171, 2003.

some fundamental dilemmas in EAs implementation. We believe any attempt of improving the efficiency of EAs has to compromise these dilemmas, which include:

- *The elitist search versus diversity maintenance dilemma*: EAs are also expected to be global optimizers with unique global search capability to guarantee exploration of the global optimum of a problem. So the elitist strategy is widely adopted in the EAs search process. Unfortunately, the elitist strategy concentrates on some "super" individuals, reduces the diversity of the population, and in turn leads to the premature convergence.
- *The algorithm effectiveness versus population redundancy dilemma*: For many EAs, we can use a large population size to improve their effectiveness including a better chance to obtain the global optimum and the multiple optima for a multimodal problem. However, the large population size will notably increase the computational complexity of the algorithms and generate a lot of redundant individuals in the population, thereby decrease the efficiency of the EAs.

Our idea in this study is to strike a tactical balance between the two contradictory issues of the two dilemmas. We propose a new adaptive elitist-population search technique to identify and search multiple peaks efficiently in multimodal problems. We incorporate the technique in genetic algorithms(GAs) as a case study, yielding an adaptive elitist-population based genetic algorithm(AEGA).

The next section describes the related work relevant to our proposed technique. Section 3 introduces the adaptive elitist-population search technique and describes the implementation of the algorithm. Section 4 presents the comparison of our results with other multimodal evolutionary algorithms. Section 5 draws some conclusion and proposes further directions of research.

2 Related Work

In this section we briefly review the existing methods developed to address the related issues: elitism, niche formation method, and clonal selection principle of an artificial immune network.

2.1 Elitism

It is important to prevent promising individuals from being eliminated from the population during the application of genetic operators. To ensure that the best chromosome is preserved, elitist methods copy the best individual found so far into the new population [4]. Different EAs variants achieve this goal of preserving the best solution in different ways, e.g. GENITOR [8] and CHC [2]. However, "elitist strategies tend to make the search more exploitative rather than explorative and may not work for problems in which one is required to find multiple optimal solutions" [6].

2.2 Evolving Parallel Subpopulations by Niching

Niching methods extend EAs to domains that require the location and maintenance of multiple optima. Goldberg and Richardson [1] used Holland's sharing concept [3] to divide the population into different subpopulations according to similarity of the individuals. They introduced a sharing function that defines the degradation of the fitness of an individual due to the presence of neighboring individuals. The sharing function is used during selection. Its effect is such that when many individuals are in the same neighborhood they degrade each other's fitness values, thus limiting the uncontrolled growth of a particular species.

Another way of inducing niching behavior in a EAs is to use crowding methods. Mahfoud [7] improved standard crowding of De Jong [4], namely deterministic crowding, by introducing competition between children and parents of identical niche. Deterministic crowding works as follows. First it groups all population elements into $n/2$ pairs. Then it crosses all pairs and mutates the offspring. Each offspring competes against one of the parents that produced it. For each pair of offspring, two sets of parent-child tournaments are possible. Deterministic crowding holds the set of tournaments that forces the most similar elements to compete. Similarity can be measured using either genotypic or phenotypic distances. But deterministic crowding fails to maintain diversity when most of the current populations have occupied a certain subgroup of the peaks in the search process.

2.3 Clonal Selection Principle

The clonal selection principle is used to explain the basic features of an adaptive immune response to an antigenic stimulus. This strategy suggests that the algorithm performs a greedy search, where single members will be optimized locally (exploitation of the surrounding space) and the newcomers yield a broader exploration of the search space. The population of clonal selection includes two parts. First one is the clonal part. Each individual will generate some clonal points and select best one to replace its parent, and the second part is the newcomer part, the function of which is to find new peaks. Clonal selection algorithm also incurs expensive computational complexity to get better results of the problems [5].

All the techniques found in the literature try to give all local or global optimal solutions an equal opportunity to survive. Sometimes, however, survival of low fitness but very different individuals may be as, if not more, important than that of some highly fit ones. The purpose of this paper is to present a new technique that addresses this problem. We show that using this technique, a simple GA will converge to multiple solutions of a multimodal optimization problem.

3 Adaptive Elitist-Population Search Technique

Our technique for the multimodal function maximization presented in this paper achieves adaptive elitist-population searching by exploiting the notion of the

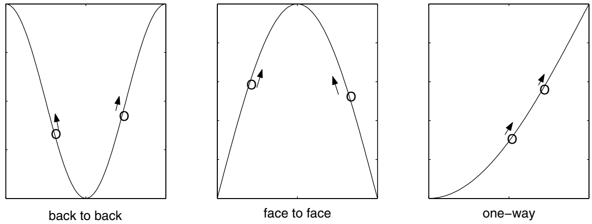

Fig. 1. The relative ascending direction of both individuals being considered: back to back, face to face and one-way.

relative ascending directions of both individuals (and for a minimization problem this direction is called relative descending direction).

For a high dimension maximization problem, every individual generally has many ascending directions. But along the line, which is uniquely defined by two individuals, each individual only has one ascending direction, called the relative ascending direction toward the other one. Moreover, the relative ascending directions of both individuals only have three probabilities: back to back, face to face and one-way (Fig.1). The individuals located in different peaks are called dissimilar individuals. We can measure the dissimilarity of the individuals according to the composition of their relative ascending directions and their distance. The distance between two individuals $x_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ and $x_j = (x_{j1}, x_{j2}, \cdots, x_{jn})$ is defined by:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \tag{1}$$

In this paper we use the above definition of distance, but the method we described will work for other distance definitions as well.

3.1 The Principle of the Individuals' Dissimilarity

Our definition of the principle of the individuals' dissimilarity, as well as the operation of the AEGA, depends on the relative ascending directions of both individuals and a parameter we call the distance threshold, which denoted by σ_s. The principle to measure the individuals' dissimilarity is demonstrated as follows:

- If the relative ascending directions of both individuals are back to back, these two individuals are dissimilar and located on different peaks;
- If the relative ascending directions of both individuals are face to face or one-way, and the distance between two individuals is smaller than σ_s, these two individuals are similar and located on the same peak.

In niching approach, the distance between two individuals is the only measurement to determine whether these two individuals are located on the same peak,

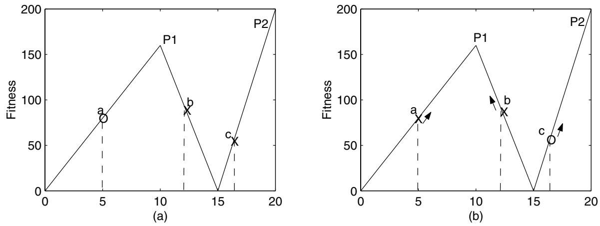

Fig. 2. Determining subpopulations by niching method and the relative ascending directions of the individuals.

but this is often not accurate. Suppose, for example, that our problem is to maximize the function shown in Fig.2. P_1 and P_2 are two maxima and assume that, in a particular generation, the population of the GA consists of the points shown. The individuals a and b are located on the same peak, and the individual c is on another peak. According to the distance between two individuals only, the individuals b and c will be put into the same subpopulation, and the individual a into another subpopulation (Fig.2-(a)). Since the fitness of c is smaller than that of b, the probability of c surviving to the next generation is low. This is true even for a GA using fitness sharing, unless a sharing function is specifically designed for this problem. However, the individual c is very important to the search, if the global optimum P_2 is to be found. Applying our principle, the relative ascending directions of both individuals b and c are back to back, and they will be considered to be located on different peaks (Fig.2-(b)). Identifying and preserving the "good quality" of individual c is the prerequisite for genetic operators to maintain the diversity of the population. We propose to solve the problems by using our new elitist genetic operators described below.

3.2 Adaptive Elitist-Population Search

The goal of the adaptive elitist-population search method is to adaptively adjust the population size according to the features of our technique to achieve:

- a single elitist individual searching for each peak; and
- all the individuals in the population searching for different peaks in parallel.

For satisfying multimodal optimization search, we define the elitist individuals in the population as the individuals with the best fitness on different peaks of the multiple domain. Then we design the elitist genetic operators that can maintain and even improve the diversity of the population through adaptively adjusting the population size. Eventually the population will exploit all optima of the mulitmodal problem in parallel based on elitism.

Elitist Crossover Operator: The elitist crossover operator is composed based on the individuals' dissimilarity and the classical crossover operator. Here we

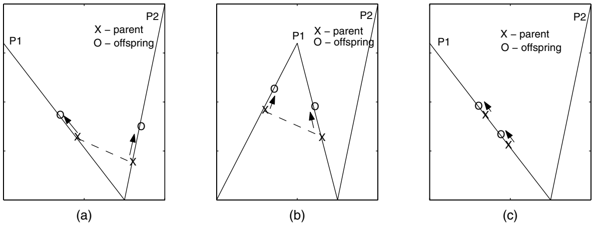

Fig. 3. A schematic illustration that the elitist crossover operation.

have chosen the random uniformly distributed variable to perform crossover (with probability p_c), so that the offspring c_i and c_j of randomly chosen parents p_i and p_j are:

$$\begin{aligned} c_i &= p_i \pm \mu_1 \times (p_i - p_j) \\ c_j &= p_j \pm \mu_2 \times (p_i - p_j) \end{aligned} \quad (2)$$

where μ_1, μ_2 are uniformly distributed random numbers over $[0, 1]$ and the signs of μ_1, μ_2 are determined by the relative directions of both p_i and p_j.

The algorithm of the elitist crossover operator is given as follows:

Input: g–number of generations to run, N–population size
Output: P_g–the population to next generation

for $t \longleftarrow 1$ **to** $N/2$ **do**
  $p_i \longleftarrow$ random select from population $P_g(N)$);
  $p_j \longleftarrow$ random select from population $P_g(N-1)$;
  determine the relative directions of both p_i and p_j;
    if back to back **then** $\mu_1 < 0$ and $\mu_2 < 0$;
    if face to face **then** $\mu_1 > 0$ and $\mu_2 > 0$;
    if one-way **then** $\mu_1 > 0, \mu_2 < 0$ or $\mu_1 < 0, \mu_2 > 0$;
  $c_i \longleftarrow p_i + \mu_1 \times (p_i - p_j)$;
  $c_j \longleftarrow p_j + \mu_2 \times (p_i - p_j)$;
  if $f(c_1) > f(p_1)$ **then** $p_1 \longleftarrow c_1$;
  if $f(c_2) > f(p_2)$ **then** $p_2 \longleftarrow c_2$;
  if the relative directions of p_1 and p_2 are face to face or one-way, and
   $d(p_1, p_2) < \sigma_s$, **then**
    if $f(p_1) > f(p_2)$ **then** $P_g \longleftarrow P_g/p_2$ and $N \longleftarrow N-1$;
    if $f(p_2) > f(p_1)$ **then** $P_g \longleftarrow P_g/p_1$ and $N \longleftarrow N-1$;
end for

As shown above, through determining the signs of the parameters μ_1 and μ_2 by the relative directions of both p_1 and p_2, the elitist crossover operator generates the offspring along the relative ascending direction of its parents (Fig.3), thus the search successful rate can be increased and the diversity of the population be maintained. Conversely, if the parents and their offspring are determined to be on the same peak, the elitist crossover operator could select the elitist to be retained by eliminating all the redundant individuals to increase the efficiency of the algorithm.

Elitist Mutation Operator: The main function of the mutation operator is finding a new peak to search. However, the classical mutation operator cannot satisfy this requirement well. As shown in Fig.4-(a), the offspring is located on a new peak, but since its fitness is not better than its parent, so it is difficult to be retained, and hence the new peak cannot be found by this mutation operation. We design our elitist mutation operator for solving this problem based on any mutation operator, but the important thing is to determine the relative directions of the parent and the child after the mutation operation. Here we use the uniform neighborhood mutation (with probability p_m):

$$c_i = p_i \pm \lambda \times r_m \tag{3}$$

where λ is a uniformly distributed random number over $[-1, 1]$, r_m defines the mutation range and it is normally set to $0.5 \times (b_i - a_i)$, and the $+$ and $-$ signs are chosen with a probability of 0.5 each.

The algorithm of the elitist mutation operator is given as follows:

Input: g–number of generations to run, N–population size
Output: P_g–the population to next generation

for $t \longleftarrow 1$ **to** N **do**
  $c_t \longleftarrow p_t + \lambda \times r_m$;
  determine the relative directions of both p_t and c_t;
  if face to face or one-way and $f(c_t) > f(p_t)$ **then**
    $p_t \longleftarrow c_t$ and **break** ;
  else if back to back, **then**
    for $s \longleftarrow 1$ **to** $N - 1$ **do**
      if $[d(p_s, c_t) < \sigma_s]$ and $[f(p_s) > f(c_t)]$ **then break;**
      else $P_g \longleftarrow P_g \cup \{c_t\}$ and $N \longleftarrow N + 1$;
    end for
  end if
end for

As shown above, if the direction identification between the parent and offspring demonstrates that these two points are located on different peaks, the parent is passed on to the next generation and its offspring is taken as a new individual candidate. If the candidate is on the same peak with another individual, the distance threshold σ_s will be checked to see if they are close enough for fitness competition for survival. Accordingly, in Fig.4-(a), the offspring will be conserved in the next generation, and in Fig.4-(b), the offspring will be deleted. Thus, the elitist mutation operator can improve the diversity of the population to find more multiple optima.

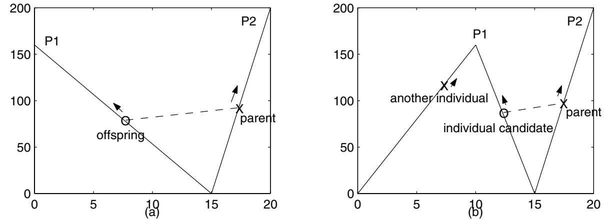

Fig. 4. A schematic illustration that the elitist mutation operation.

3.3 The AEGA

In this section, we will present the outline of the Adaptive Elitist-population based Genetic Algorithm(AEGA). Because our elitist crossover and mutation operators can adaptively adjust the population size, our technique completely simulates the "survival for the fittest" principle without any special selection operator. On the other hand, since the population of AEGA includes most of the elitist individuals, a classical selection operator could copy some individuals to the next generation and deleted others from the population, thus the selection operator will decrease the diversity of the population, increase the redundancy of the population, and reduce efficiency of the algorithm. Hence, we design the AEGA without any special selection operator. The pseudocode for AEGA is shown bellow. We can see that the AEGA is a single level parallel (individuals) search algorithm same as the classical GA, but the classical GA is to search for single optimum. The niching methods is a two-level parallel (individuals and subpopulations) search algorithms for multiple optima. So in terms of simplicity in the algorithm structure, the AEGA is better than the other EAs for multiple optima.

The structure of the AEGA:

begin
  $t \longleftarrow 0$;
  Initialize $P(t)$;
  Evaluate $P(t)$;
  while (not termination condition) **do**
    Elitist crossover operation $P(t+1)$;
    Elitist mutation operation $P(t+1)$;
    Evaluate $P(t+1)$;
    $t \longleftarrow t+1$;
  end while
end

4 Experimental Results

The test suite used in our experiments include those multimodal maximization problems listed in Table 1. These types of functions are normally regarded as difficult to be optimized, and they are particularly challenging to the applicability and efficiency of the multimodal evolution algorithms. Our experiments of multimodal problems were divided into two groups with different purpose. We report the results of each group below.

Table 1. The test suite of multimodal functions used in our experiments.

Deb's function(5 peaks):
$f_1(x) = \sin^6(5\pi x)$, $x \in [0,1]$;
Deb's decreasing function(5 peaks):
$f_2(x) = 2^{-2((x-0.1)/0.9)^2} \sin^6(5\pi x)$, $x \in [0,1]$;
Roots function(6 peaks):
$f_3(x) = \frac{1}{1+\|x^6-1\|}$, where $x \in C$, $x = x_1 + ix_2 \in [-2,2]$;
Multi function(64 peaks):
$f_4(x) = x_1 \sin(4\pi x_1) - x_2 \sin(4\pi x_2 + \pi) + 1$; $x_1, x_2 \in [-2,2]$.

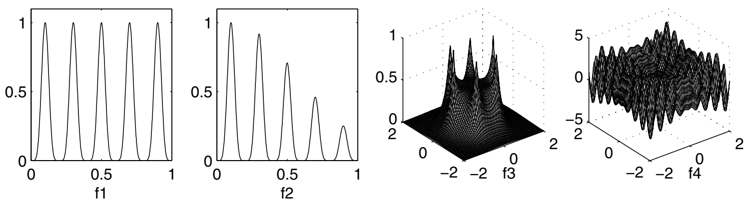

Fig. 5. The suite of the multimodal test function.

Explanatory Experiments: This group of experiments on $f_3(x)$ aims to exhibit the evolution details (particularly, the adaptive adjusting population size details) of the AEGA for 2-D case, and also to demonstrate the parameter control of the search process of the AEGA. In applying the AEGA to solve the 2-D problem $f_3(x)$, we set the initial population size $N = 2$ and the distance threshold $\sigma_s = 0.4$ or 2.

Fig.6 demonstrate clearly how the new individual are generated on a newly discovered peak and how the elitist individuals reach each optimum in the multimodal domain. Fig.6-(a) shows the 2 initial individuals, when $\sigma_s = 0.4$, the population size is increased to 8 at the $50th$ generation(Fig.6-(b)), and Fig.6-(c) show the 4 individuals in the population at the $50th$ generation when $\sigma_s = 2$. When σ_s is smaller, new individuals are generated more easily. At the $50th$ generation, the result of $\sigma_s = 0.4$ seems to be better; but finally, both settings of the AEGA can find all the 6 optima within 200 generations (Fig.6-(d)). This means the change of the distance threshold does not necessarily influence the efficiency of the AEGA.

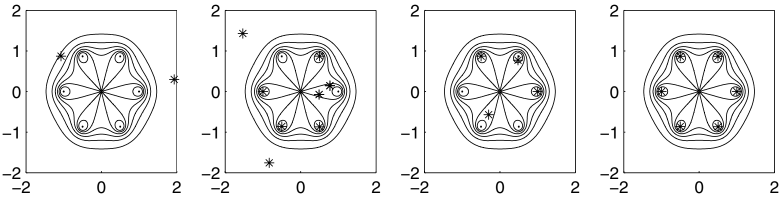

Fig. 6. A schematic illustration that the AEGA to search on the Roots function, (a) the initial population; (b) the population at $50th$ generation ($\sigma_s = 0.4$); (c) the population at $50th$ generation ($\sigma_s = 2$); (d) the final population at $200th$ generation ($\sigma_s = 0.4$ and 2).

Comparisons: To assess the effectiveness and efficiency of the AEGA, its performance is compared with the fitness sharing, determining crowding and clonal selection algorithms. The comparisons are made in terms of the solution quality and computational efficiency on the basis of applications of the algorithms to the functions $f_1(x) - f_4(x)$ in the test suite. As each algorithm has its associated overhead, a time measurement was taken as a fair indication of how effectively and efficiently each algorithm could solve the problems. The solution quality and computational efficiency are therefore respectively measured by the number of multiple optima maintained and the running time for attaining the best result by each algorithm. Unless mentioned otherwise, the time is measured in seconds as measured on the computer.

Tables 2 lists the solution quality comparison results in terms of the numbers of multiple optima maintained when the AEGA and other three multimodal algorithms are applied to the test functions, $f_1(x) - f_4(x)$. We have run each algorithms 10 times. We can see, each algorithm can find all optima of $f_1(x)$. In the AEGA, two initial individuals increase to 5 individuals and find the 5 multiple optima. For function $f_2(x)$, crowding algorithm cannot find all optima for each time. For function $f_3(x)$, crowding cannot get any better result. Sharing and clonal algorithms need to increase the population size for improving their performances. The AEGA still can use two initial individuals to find all multiple optima. For function $f_4(x)$, crowding, sharing and clonal algorithms cannot get any better results, but the successful rate of AEGA for finding all multiple optima is higher than 99%. Figs.7 and 8 show the comparison results of the AEGA and the other three multimodal algorithms for $f_1(x)$ and $f_2(x)$, respectively. The circles and starts represent the initial populations of AEGA and the final solutions respectively. In the AEGA process, we have only used 2 individuals in the initial population. In 200 generations, finally the 5 individuals in the population can find the 5 multiple optima. These clearly show why the AEGA is significantly more efficient than the other algorithms. On the other hand, the computational efficiency comparison results are also shown in Tables 2. It is clear from these results that the AEGA exhibits also a very significant outperformance of many orders compared to the three algorithms for all test functions. All these comparisons show the superior performance of the AEGA in efficacy and efficiency.

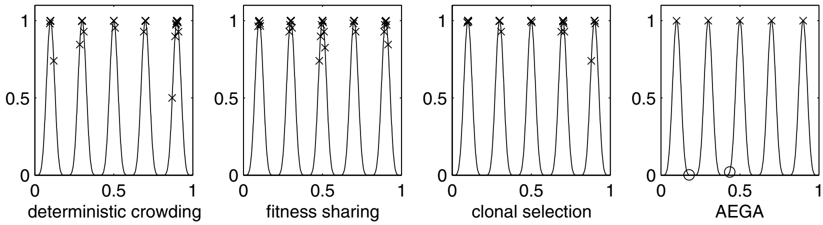

Fig. 7. A schematic illustration that the results of the algorithms for $f_1(x)$.

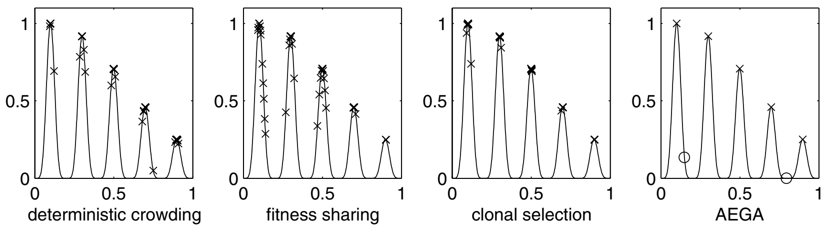

Fig. 8. A schematic illustration that the results of the algorithms for $f_2(x)$.

Table 2. Comparison of results of the algorithms for $f_1(x) - f_4(x)$.

Algorithms	Initial population				Generations				Nb of Peaks				Running time(s)			
	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
Deterministic Crowding	100	100	500	1000	20000	20000	20000	20000	5	4.8	3.2	13.2	6.09	6.14	51.5	121
Fitness Sharing	100	100	500	500	200	200	2000	5000	5	5	5.1	18.7	2.44	2.45	41.2	236
Clonal Selection	50	50	200	500	200	200	1000	5000	5	5	5.8	30.1	1.10	1.12	37.1	187
AEGA	2	2	2	50	200	200	200	5000	5	5	6	63.7	0.01	0.01	0.11	9.4

5 Conclusion and Future Work

In this paper we have presented the adaptive elitist-population search method, a new technique for evolving parallel elitist individuals for multimodal function optimization. The technique is based on the concept of adaptively adjusting the population size according to the individuals' dissimilarity and the elitist genetic operators.

The adaptive elitist-population search technique can be implemented with any combinations of standard genetic operators. To use it, we just need to introduce one additional control parameter, the distance threshold, and the population size is adaptively adjusted according to the number of multiple optima. As an example, we have endowed genetic algorithms with the new multiple technique, yielding an adaptive elitist-population based genetic algorithm(AEGA).

The AEGA then has been experimentally tested with a difficult test suite consisted of complex multimodal function optimization examples. The performance of the AEGA is compared against the fitness sharing, determining crowing and clonal selection algorithms. All experiments have demonstrated that the AEGA consistently and significantly outperforms the other three multimodal evolutionary algorithms in efficiency and solution quality, particularly with efficiency speed-up of many orders.

We plan to apply our technique to hard multimodal engineering design problems with the expectation of discovering novel solutions. We will also need to investigate the behavior of the AEGA on the more theoretical side.

Acknowledgment. This research was partially supported by RGC Earmarked Grant 4212/01E of Hong Kong SAR and RGC Research Grant Direct Allocation of the Chinese University of Hong Kong.

References

1. D.E. Goldberg and J.Richardson, Genetic algorithms with sharing for multimodal function optimization, *Proc.* 2^{nd} *ICGA*, pp.41-49, 1987
2. Eshelman.L.J. The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. *In Rawlins, G.J.E., editor, Foundation of Genetic Algorithms* pp.265-283, Morgan Kaufmann, San Mateo, California, 1991
3. Holland and J.H. Adaptation in Natural and ArtificialSystem. *University of Michigan Press*, Ann Arbor, Michigan, 1975.
4. K.A. De Jong, An analysis of the behavior of a class of genetic adaptive systems, *Doctoral dissertation, Univ. of Michigan*, 1975
5. L.N.Castro and F.J.Zuben, Learning and Optimization Using the Clonal Selection Principle, *IEEE Transactions on EC*, vol .6 pp.239-251, 2002
6. Sarma.J. and De Jong, Generation gap methods. *Handbook of Evolutionary Computation*, pp.C2.7:1-C2.7:5, 1997
7. S.W.Mahfoud. Niching Methods for Genetic algorithms, *Doctoral Dissertation, IlliGAL Report 95001, University of Illinois at Urbana Champaign, Illinois Genetic Algorithm Laboratory*, 1995
8. Whitley.D., The GENITOR algorithm and selection pressure: why rank-based allocation of reproductive trials is best. *Proc.* 3^{rd} *ICGA*, pp.116-121, 1989