Adaptive Isolation Model using Data Clustering for Multimodal Function Optimization

Shin Ando, Jun Sakuma, Shigenobu Kobayashi

Interdisciplinary School of Science and Engineering, Tokyo Institute of Technology
4259 Nagatsutacho, Midori-ku
Yokohama, Kanagawa, Japan

Abstract

In this paper, we propose a GA model called Adaptive Isolation Model(AIM), for multimodal optimization. It uses a data clustering algorithm to detect clusters in GA population, which identifies the attractors in the fitness landscape. Then, subpopulations which makes-up the clusters are isolated and optimized independently. Meanwhile, the region of the isolated subpopulations in the original landscape are suppressed. The isolation increases comprehensiveness, i.e., the probability of finding weaker attractors, and the overall efficiency of multimodal search. The advantage of the AIM is that it does not require distance between the optima as a presumed parameter, as it is estimated from the variance/covariance matrix of the subpopulation.

Further, AIM's behavior and efficiency is equivalent to basic GA in unimodal landscape, in terms of number of evaluation. Therefore, it is applied recursively to all subpopulations until they converge to a suboptima. This makes AIM suitable for locally-multimodal landscapes, which have closely located attractors that are difficult to distinguish in the initial run.

The performance of AIM is evaluated in several benchmark problems and compared to iterated hill-climbing methods.

1 Introduction

The multimodal optimization algorithm have many practical needs in real-world problems. In many instances, preparing multiple solutions, preferably with varying characteristics, is a better general strategy. The objective of multimodal optimization is to search multiple optima/suboptima comprehensively, or to certain extent. In nature, the process of evolution has produced great diversity in population and their functions. In the field of evolutionary computation, many researchers have attempted to reproduce and utilize such mechanism in the framework of evolutionary algorithms, which were successfully applied to many benchmark problems.

In this paper, we describe a framework of GA for multimodal optimization called Adaptive Isolation Model(AIM), which use clustering algorithm to identify different region of attractors, then search each attractors independently.

With proper setting, it is reasonable to assume that GA population will form a 'cluster', which has higher density near the center and sparse at the marginal region, on the attracting region of the landscape. Such clusters can be detected using proper clustering algorithm.

In AIM, the subpopulation in the detected clusters are isolated and optimized in later process. On the other hand, the fitness function in original GA is modified to suppress the region of isolated subpopulation, so that the effect of the isolated attractor is also removed from remaining population.

In multimodal optimization, searching each of the attractors independently from the rest is beneficial, in terms of efficiency and comprehensiveness. The efficiency is induced since the crossovers will only include parents from same attractors, thereby reducing the number of offspring sampled outside of the attractors. With regard to comprehensiveness, the probability of finding suboptima in weaker attractors increases by isolating strong attractors. This is due to combination of excluding parents in strong attractors from crossover, and prohibiting children in strong attractor to replace parents in weak attractors.

Many multimodal optimization algorithms have attempted to avoid repeatedly finding or converging to same optima by applying penalty function to a region within a predefined range of known optima. However, this approach introduces the range of the penalty function as a critical problem dependent parameter, which made the algorithm impractical for unknown problems. One of the advantages of AIM is that it does not require such distance/range parameter, since it is estimated from the range of the subpopulation.

In AIM, isolation occurs when more than one clusters are detected within the population. This implementation allows AIM to behave as an unimodal GA in unimodal landscape. On the other hand, many multimodal optimization algorithms propose two-phase algorithm, with fast convergence algorithm ensuing the multimodal search algorithm. This is due to the inefficiency of the multimodal GA in unimodal landscape compared to the standard GA.

Meanwhile, many researches have reported the difficulty in a locally multimodal landscape, where 'clusters' or groups of suboptima are located closely together in a relatively small region. In such landscape, it is difficult to distinguish suboptima within that cluster. [8] proposed to reapply multimodal-optimizer to the subpopulation to distinguish such suboptima. Therefore, in practice, users cannot switch to 'convergence-mode' without somehow confirming local landscape's unimodality.

With regards to above, another advantage of AIM comes from its efficiency in unimodal landscape. AIM can be recursively applied to its isolated subpopulation, without analysis of local landscape's unimodality/multimodality.

In Sec. 2, we give an overview of GAs for multimodal optimization. Then in Sec. 3, we describe the implementation of AIM. In Sec. 4, we evaluates AIM's performance on several benchmarks problems. Sec. 5, 6 give discussion and our conclusion.

2 Related Works

In multimodal-landscape, there are multiple optima/suboptima and their attractors, i.e., convex surrounding each. When GA is applied to such landscape, after certain generations, its population will be distributed among different attractors. In Fig.1, GA population is distributed among four attractors in the fitness landscape.

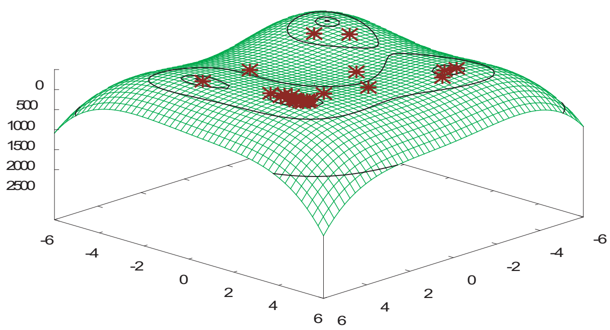

Figure 1: A Snapshot of GA Population on Multimodal Landscape

The attractors can have different characteristics, which can largely affect the GA's behavior[4, 5]. Ikeda et al. [4] showed a type of deception caused by presence of heterogeneous attractors. Such landscape are known to exist in Job Shop scheduling[4] and Fletcher-Powell's Function[14]. They identified three properties of strong attractors, which can attract larger number of population. They were: a) size, b) better fitness in initial sampling, and c)search efficiency. Ikeda et al. termed 'UV-structure' as a landscape when suboptima have significantly stronger attractors than that of the optima and claimed such landscape caused deception.

They proposed Innately Split Model(ISM)[4], where multiple GA population are initialized on small subdivisions of the landscape. The division helped to remove the effect of strong attractors from others and ‘localize’ the search process.

An implication from [4] is that strong attractors are detrimental to search in the rest of landscape. Assuming random selection of the parents, strong attractors have better possibility of having parents in the crossover since they have larger number of individuals, therefore slowing search speed in other attractors. Further, with slower search speed, the individual in weaker attractors are more likely to be replaced by offspring in stronger attractors.

To that end, it is beneficial to 'localize‘ a) selection and b) reproduction procedures, to prevent a strong attractor from taking away individuals from other attractors. In addition, localizing the reproduction, i.e., crossing-over parents from same attractors will reduce the number of offspring generated outside the attractors, which will improve efficiency of the search. [14] referred to a) as survival selection and b) as reproductive selection.

We can also classify conventional approaches to multimodal optimization into two categories, with regards to type of localization is used. First group of methods, including Crowding[2] and its variations Deterministic[9] and Probabilistic Crowding[10], and Species Conserving GA[8], modifies the selection procedure. These method try to apply selection to the parent-child pair that are neighbors or same ‘species’ in terms of Euclidean distance, so that parents are replaced by individuals from same attractors, thereby localize the selection process.

The second group of methods use multi-populational implementation, e.g. Multi-National GA[15], ISM[4], CBN[13]. These approaches, intend to assign one population to an attractor. In effect, parents from same attractor will be crossed-over, thus localizing the reproductive selection. In addition, ANS[14], which uses nearest-neighbor approach to select the parents of genetic operator, can also be categorized in the latter group.

AIM also belongs to the latter, since it isolates subpopulation on each attractors. However, unlike ISM which divides the landscape at the predefined threshold, AIM will detect the attractors heuristically using GA and clustering algorithm.

3 Adaptive Isolation Model

In unimodal landscape, GA population can converge to a single optima after iteration of proper genetic operators and selection. The population forms a cluster over the attractor of the optima, densely populated near the center and sparsely at the margin. In multimodal landscape, populations are distributed among several attractors and forming multiple clusters on each attractors. In both cases, we can assume that the clusters represent the local convex around the optima. The main procedure of AIM is to detect such clusters and isolate them into independent subpopulations.

The detection of clusters is done by repeatedly applying clustering algorithm to GA population. The isolation consist of a) recording the subpopulation that compose the cluster and removing them from original population, and b) modifying the fitness function to suppress the region of the attractors. The isolated subpopulation is optimized independently. The range of suppression is calculated from variance/covariance of the cluster.

CBN[13] also uses density-based clustering algorithm to split and merge multiple population, but does not isolate subpopulation.

Following are the detailed description of the AIM procedure.

1. Apply GA step, i.e., crossover operator(UNDX) and selection(MGG), k times.

2. Run clustering algorithm on GA population.

3. If the clustering algorithm detects $nc(> 2)$ components, isolate subpopulations which belong to components $C_1, ...C_{nc-1}$. Further, supplement individual to maintain the size of original population.

4. Adjust the fitness function to suppress the region of the isolated components C_i. The clusters are identified as Gaussian components as described in Section 3.1, and the region within 2σ of its

mean are suppressed. Individuals within that region will have the fitness value of f_{worst}, which is the value of the worst individual in the run.

5. Repeat the steps 1 through 4, until the termination criteria is met. The AIM is terminated after: a) fitness value reaches target value, b) performing certain number of evaluation, or c) when certain proportion of the population are trapped in suppressed region.

6. After termination, if isolated subpopulations exist, AIM is applied to each. In other word, AIM is applied recursively to subpopulations, until the subpopulation is distributed over only one attractor.

Fig.3 and Fig.4 shows the pseudo code and the flow chart of the AIM's algorithm.

In the the unimodal landscape, clustering algorithm detect only one cluster as AIM population converge to a single optima. Therefore AIM behaves as UNDX+MGG in unimodal landscape. We will show in Section 4 that AIM has convergence speed equivalent to UNDX+MGG in unimodal landscape, thus computational cost of applying AIM to subpopulation is not infeasible.

Fig.2 shows the snapshots of population when AIM is applied to the Rastrigin function. In each snapshot, members of population detected as clusters are represented by 'Δ', while others are represented by '+'. The region of Gaussian components, denoted by circles, are suppressed in the original fitness landscape. In this example, the periodically placed optima induce clustering algorithm to cluster individuals on several attractor together.

3.1 Clustering Algorithm

There are several requirement for clustering algorithms used in AIM.

1. Does not require number of clusters as parameters.

2. Able to identify clusters from noisy data (noise from data).

aimflow.eps

Figure 3: Flow chart of AIM

```
pop : {(x_11,...,x_1n),(x_21,...,x_2n),...,(x_m0,...,x_mn)}

function CLS(pop){return k, subpop_i (i = 0 : U, i = 1 to k : G)}

function restrict(x_1,...,x_n,G){
    d = maharanobis(G, x_1,...,x_n)
    if (d < threshold) return feasible
    else return infeasible
}
function AIM(pop, maxcomponent){
gen = 0;
while[gen < maxgen AND pop ≠ ∅]
do
    for[count = 0 to 100]
    do
        pop ← mggstep(pop)
    done
    k, subpop_{nc+1},...,subpop_{nc+k} ← CLS(pop)
    if (k > 1)
    pop = subpop_0 + subpop_{nc+k}
    obj = obj × restrict(G_{nc+1}) × ... × restrict(G_{nc+k-1})
    nc+ = k − 1
done
if (overlap) merge(subpops)
AIM(subpop_i)
return
}
```

Figure 4: Pseudo code of AIM

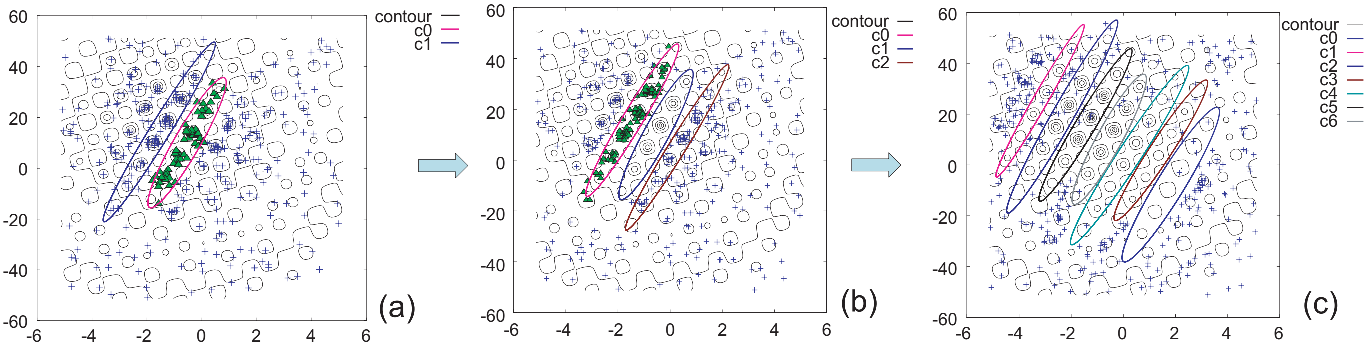

Figure 2: Detection and Isolation in AIM

The example of AIM's behavior in two-dimensional, rotated, and scaled Rastrigin function is shown. In the left figure(a), two components c_0 and c_1 are found and c_0 is isolated. The individuals indicated by Δ belong to c_0, and $+$ indicate the rest of the population. The circle indicate the 2σ threshold line. The fitness value of the region within the circle will be suppressed. Following several GA steps (b), another component c_2 is found and c_1 is isolated. After more iteration (c), components $c_0 - c_6$, indicated by circles, are isolated.

3. Computationally stable.

4. Scale-invariant (preferably).

Scale-invariance ensures that the result of the clustering will not affected by linear transformation of the data/domain, which is an important feature for algorithm applied to real-world problems. In many instance of real-world problems, variables may have different units of measurement and different in the order of thousands. This is referred to as an ill-scaled domain[5]. Additionally, strong correlation among variables also induce ill-scale property into attractors. Ill-scaled domain can be problematic for real-coded GA[5] and optimization algorithms in general.

We used a clustering algorithm called Localized MAP-EM(LMAP-EM)[6], which is an EM algorithm based on MAP learning. While LMAP-EM is model-free, we used Gaussian distribution to model components for mathematical convenience.

The data is modeled as a mixture of normal distributions as follow.

$$P_j(\vec{x}\,|\vec{\mu}_j) = \frac{1}{m}\exp\left(-\frac{(\vec{x}-\vec{\mu}_j)^T(\vec{x}-\vec{\mu}_j)}{2\sigma^2}\right) \quad (1)$$

$$P(\vec{x}\,|\vec{\mu}) = \sum_{j=1}^{m} P_j(\vec{x}\,|\vec{\mu}_j) \quad (2)$$

m is the number of components and $\Theta_j = (\vec{\mu}_j, \sigma)$ represents the parameters of each component.

Assuming the latent variables as

$$Z = \{z_{ij}\,|i = 1, \ldots, N, j = 1, \ldots, m\} \quad (3)$$

where N is the data size. $z_{ij} = 1$ if data $\vec{x}_i$ is generated by jth component $P_j(\vec{x}\,|\Theta_j)$ and $z_{ij} = 0$ otherwise.

In LMAP-EM, hard-assignment is assumed, i.e., each datapoint is assigned to only one component. More formally, the following equality for $p(z_{ij}\,|x_i)$, the probability that data x_i was generated by jth component is assumed.

$$p(z_{ij}\,|x_i) = \{0, 1\}, \sum_{j=1}^{m} p(z_{ij}\,|x_i) = 1 \quad (4)$$

Under this assumption, the log likelihood of the model is expressed as the sum of likelihood of each component as follows.

$$L(\Theta) = \sum_{j=1}^{m}\left\{\log P_j(\Theta_j\,|pa(\Theta_j)) + \sum_{i=1}^{N} z_{ij}\log P_j(\vec{x}_j\,|\Theta_j)\right\} \quad (5)$$

where $pa(\Theta_j)$ is the prior distribution of the parameter.

The E-step of LMAP-EM, new latent variables Z', which increase the localized likelihood (6) is generated.

$$\sum_{i=1}^{N} z'_{ij} \log P_j\left(X, Z \left| \Theta_j \right.\right) \geq \sum_{i=1}^{N} z_{ij} \log P_j\left(X, Z \left| \Theta_j \right.\right) \tag{6}$$

In the M-step, the component parameter is updated with equation (7). This is a simplified version of the M-step, localized likelihood is increased by steepest decent method in [6].

$$\vec{\mu} = \sum_{i=1}^{N} z_{ij}\vec{x} \Big/ \sum_{i=1}^{N} z_{ij} \tag{7}$$

After the localized likelihood has converged, we remove the subset and apply the same procedure on the remaining data. This is repeated until all data are removed.

Since each component is estimated locally and sequentially, the number of components is not required in the algorithm. Furthermore, use of Gaussian component with full parameters ensures that the result of clustering is consistent with the linear transformation of the domain. This is computationally stable as it does not require inversion of the matrix, therefore meeting 1,3,4 of the requirement mentioned.

3.2 Customizing L-MAP EM

Following are the implementation of the LMAP-EM that is customized for use with AIM.

With LMAP-EM, we obtain a mixture model of one uniform distribution and Gaussian distributions by labeling each data cluster as a Gaussian distribution or part of uniform distribution. We determine type of distribution for each cluster by comparing Akaike's Information Criteria(AIC) of two presumed mixture models. AIC is a statistical index for selecting probabilistic model with better prediction power, calculated as follows

$$AIC = -2\times\left[\begin{array}{l} log\,likelihood\,of\,the \\ estimated\,model \end{array}\right] + 2\times\left[\begin{array}{l} \#\,estimated \\ parameters \end{array}\right]$$

We prepare one model which include the data as a new Gaussian component, and another model which include data in a uniform distribution. We will select a mixture model with smaller AIC.

The initial component size L_z is a required parameter in this algorithm. This parameter is statistically determined from the number of data needed to sufficiently distinguish Uniform distribution from Normal distribution.

3.3 UNDX and MGG

In this section, we describe the genetic operators and replacement scheme used in AIM. Unimodal Normal Distribution Crossover (UNDX)[11, 3] produce offspring by a normally distributed probability function defined by the parents. A variety of real-coded crossover operators were derived from UNDX and have been successfully applied to many benchmark problems [11, 7, 3].

UNDX produces offspring based on following Gaussian distribution (8).

$$\mathbf{y} = \mathbf{g} + \sum_{i=1}^{\mu-1} w_i \mathbf{d}^{(i)} + \sum_{i=\mu}^{n} v_i D \mathbf{e}^{(i)} \tag{8}$$

where,

$$\mathbf{g} = \frac{1}{\mu-1}\sum_{i=1}^{\mu-1} \mathbf{x}^{(i)}$$

is the center of the mass, and

$$\mathbf{d}^{(i)} = \mathbf{x}^{(i)} - \mathbf{g} : i = 1 - \mu$$

$$\mathbf{e}^{(i)} = \mathbf{d}^{(i)} \Big/ \left|\mathbf{d}^{(i)}\right|$$

$$D = \left| \mathbf{x}^{(\mu)} - \mathbf{g} - \sum_{i=1}^{\mu-1} \kappa \mathbf{e}^{(i)} \right|$$

The normally distributed variables

$$w_i = N\left(0, \sigma_\zeta^2\right), v_i = N\left(0, \sigma_\eta^2\right)$$

whose parameters σ_ζ and σ_η are suggested in [11] as

$$\sigma_\eta = 0.35 \Big/ \sqrt{n-\mu-2}, \sigma_\zeta = 1 \Big/ \sqrt{\mu-2}$$

The Minimal Generation Gap (MGG) [12] is a replacement scheme used to maintain diversity in GA population. In MGG, large number of offspring is generated from one operation of crossover. Then two individuals, one by elitist and one roulette-wheel, are selected out of all offspring and parents and returned to the population. Following is the procedure for one cycle of MGG.

1. From Population P, select m parents at random.
2. Generate λ offspring by iterating the crossover of the same parents.
3. selects one elitist and one roulette-selected individual parents and offspring and return to the population.

4 Experiment

In this section, the performance of proposed method is tested on unimodal and multimodal benchmarks and artificial deceptive function.

4.1 Standard Benchmark Functions

First, we studied the proposed methods behavior in three standard benchmark functions. Sphere Function F_1(9), Rosenbrock's Function F_2(10), which has a nonlinear dependency, and Rastrigin's Function F_3(11). F_3 has a big-valley structure and 4^n local optima within the given range.

$$F_1(\vec{x}) = \sum_{i=1}^{n} x_i^2 \ \ [-5.12 \leq x_i \leq 5.12] \tag{9}$$

$$F_2(\vec{x}) = \sum_{i=2}^{n} \left\{ 100\left(x_1 - x_i^2\right)^2 + \left(x_i - 1\right)^2 \right\} \ [-10 \leq x_i \leq 10] \tag{10}$$

$$F_3(\vec{x}) = 10n + \sum_{i=1}^{n} x_i \sin\sqrt{|x_i|} \ [-1.5 \leq x_i \leq 2.5] \tag{11}$$

Following parameters were used in the experiment. Dimension $D = 5$, population size $M = 400$, # of parents: $m = 6$, offspring size: $\lambda = 20$, number of iteration: 20. For AIM, the minimum component size: $L_Z = 35$, and AIM interval $k = L_Z$. Each run was terminated when best fitness value reached 1.0^{-10}.

Result

For F_1 and F_2, AIM detected one component throughout all of the runs, therefore its behavior and efficiency were equivalent to that of basic UNDX+MGG in terms of evaluation. The convergence curve of AIM and UNDX+MGG are shown in Figures 5 and 6. The vertical and the horizontal axis denote the best obtained fitness and the number of evaluation.

curve_sphere_aim_vs_mgg.eps

Figure 5: Optimization of Sphere Function by UNDX+MGG and AIM. Best Fitness vs. # of Evaluation.

In optimization of unimodal functions F_1 and F_2, the AIM population quickly converge to compose a very large cluster, consisting of more than half of the population. Figure 7 shows the size of the largest cluster detected in the primary population while optimizing F_1, F_2, and F_3. In multimodal landscape F_3, the primary (unisolated) population quickly forms a large cluster after all but one attractor has been isolated.

To this end, we suggest to use the size of cluster to signal AIM to switch from multimodal search to unimodal search, for the sake of improving convergence

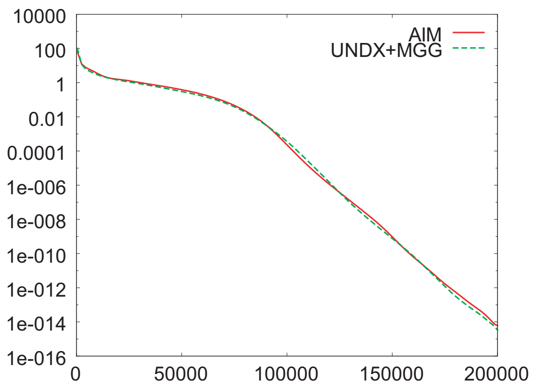

Figure 6: Optimization of Rosenbrock Function by UNDX+MGG and AIM Best Obtained Fitness vs # of Evaluation.

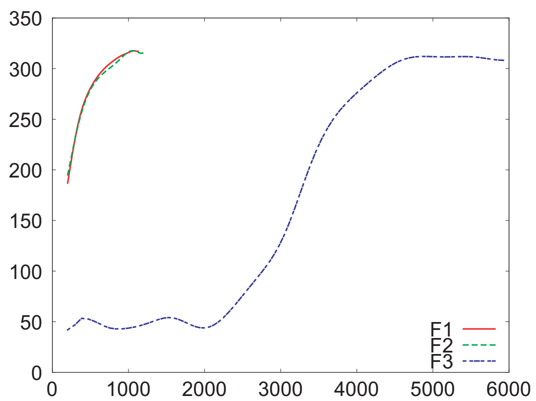

Figure 7: Size of the largest cluster in Optimization of F_1, F_2, and F_3

AIM	IHC(σ=0.01)	IHC(σ=0.05)	IHC(σ=0.1)
1002	1022	212.2	24.2

Table 1: # of Suboptima found with AIM and IHC within 8.69×10^{12} evaluations.

speed. This is a practical modification, since it is difficult to perform multimodal search at the presence of very large cluster.

In F_3, AIM found 1024 suboptima after applying AIM to 2123 subpopulations and evaluating $N_e = 8.69 \times 10^{12}$ individuals in average . When use (UNDX+MGG) with small population ($M = 100$) in unimodal search. The unimodal search is terminated when the best individual is within $1/10^8$ of suboptimal value.

We compared the result of Rastrigin function with iterated hill-climbing(IHC). IHC is a primitive method but can be more effective than GA in some problems. An HC process is started from random point. After one HC process converge, next HC is started from random initial point if the total number of evaluations is less than N_e. Here we simulate HC with (1+1)+ES with a fixed mutation rate σ_{mut}.

The comparison of average number of suboptima found in 5 runs by IHC and AIM is shown in Table 1. It shows that step size is the critical parameter to the performance of hill-climbing.

4.2 Locally multimodal function

The Shubert's function 12 is often used as a benchmark for multimodal optimization[8].

$$F_4(x) = \prod_{i=0}^{n} \sum_{j=1}^{5} j \cos((j+1)x_i + j) \qquad (12)$$

We applied AIM to two-dimensional Shubert Function. It has over 1000 symmetrically distributed suboptima within the range ($-10 \leq x_i \leq 10$). There are 3^n pairs of optima, each pair within a close proximity of 4^n suboptima. The contour of F_4 is drawn with dark line in Fig.4.2.

We applied AIM parameters equivalent to the previous experiment. The AIM was terminated when

75% of the population are trapped in suppressed region, or one cluster include 75% of the population. in the latter case, $M = 40$ best individuals of the largest clusters were used as initial population of UNDX+MGG.

Following describes a typical behavior of AIM applied to F_4. In a typical primary run of AIM, 30 clusters were detected before 75% of the population. shown in Figure 4.2.

The secondary runs, i.e. application of AIM to isolated subpopulation, can detect clusters with higher resolution. This is due to the increase in overall density of the GA population. Clusters detected in a typical secondary run of AIM is shown in Figure 4.2

In the typical run AIM found 1231 suboptima after 3.39×10^{11} evaluation. The subpopulations of unimodal search, are shown in Figure 4.2. In five runs, AIM found 1220 suboptima in 5.58×10^{12} evaluations on average.

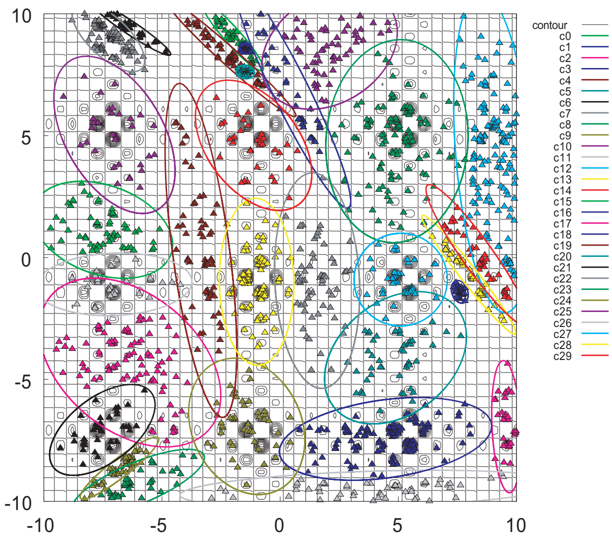

Figure 8: Components found in a typical primary run of AIM

5 Discussion

The behavior of IHC shows the critical effect of distance parameter in multimodal optimization. Con-

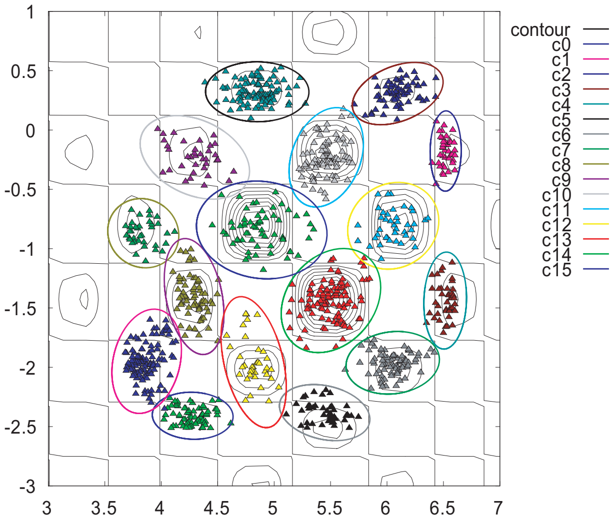

Figure 9: Components found in a typical secondary run of AIM

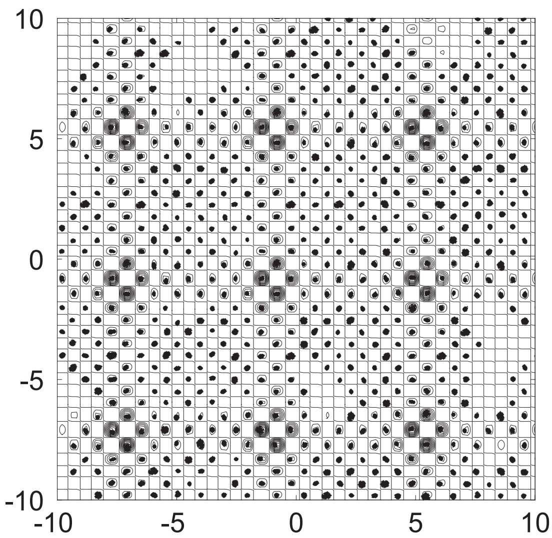

Figure 10: Subpopulation of AIM which were applied Unimodal Search

ventionally, Sequential Niching[1] also used penalty function to suppress regions around the known optima. However, it used a predefined the radii for penalty function, which was a problem dependent parameter. Since no study, to our knowledge, has proposed general method for estimating such parameters, the primary advantage of the proposed method comes from excluding such parameter from the algorithm. In this case, we use variance/covariance of subpopulation to estimate the range of attractors. In doing so, we are exploiting the GA's property as a populational search, by extracting information of the fitness landscape from the subpopulation.

Further, since many conventional works proposed to use fast converging algorithm after multimodal search, the timing of switch to unimodal optimization were another crucial parameter. AIM exclude this parameter as well, since it has same convergence speed as basic GA in unimodal landscape, and can be applied recursively to subpopulations. However, the efficiency can be improved by using largest cluster size to signal AIM to switch to unimodal search mode.

Another advantage of AIM is its reduced space complexity. While conventional multimodal optimization algorithms use large population size for highly multimodal landscape [8], this algorithm use relatively small population on the memory to find large number of optima. This contributes to improved convergence speed. This advantage comes from isolating multiple optima as a cluster, and recursively applying multimodal optimization algorithm.

The parameters used in AIM are standard GA parameters, except for the population size. While the analysis of AIM's behavior with different population size remain as our future work, we suggest an empirically determined value of $10 \times L_z$. The suggested values for minimal components size L_z are shown in Table 2

	$D \leq 5$	$D \leq 8$	$D \leq 1$
L_z	35	40	50

Table 2: Suggested values for $D \leq 10$ Dimensional Domain

6 Conclusion

We proposed a framework for combining clustering algorithm with GA for multimodal optimization. The method is aimed to isolate the attractors in the landscape to benefit multimodal search in terms of efficiency and robustness against deception.

AIM is capable of comprehensive multimodal search by isolating and suppressing strong attractors. It can also search unimodal landscape at the equivalent evaluation cost as the basic GA.

The main contribution of AIM are a) elimination of distance parameter, b) recursive application, and c) reduced time complexity for multimodal search.

We also note that AIM is scale-invariant, i.e., the result is equivalent to the linear transformation of the domain. This is an important property for real-world problems, where the variable have different dimension and significantly different scale.

In our future work, we plan to apply this method to problems with deception and higher dimensionality. We are also implementing the clustering algorithm to run online to reduce the computational cost of clustering procedure.

References

[1] D. Beasley, D. R. Bull, and R. R. Martin. A sequential niche technique for multimodal function optimization. *Evolutionary Computation*, 1(2):101–125, 1993.

[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, A. Anand, and D. Joshi. A computationally efficient evolutionary algorithm for real-parameter optimization. *Evolutionary Computation*, 10, 2002.

[4] 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.

[5] S. K. J. Sakuma. Edx. In *Proceedings of GECCO2002, Late Breaking Paper*, 2002.

[6] S. K. J. Sakuma. Localized map-em on inter-dependent prior distribution (in japanese). In *Proceedings of Workshop on Information-Based Induction Sciences*, 2003.

[7] S. Kimura, I. Ono, H. Kita, and S. Kobayashi. 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] 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 Concep-tion, Design, and Application of Intelligent Systems*, pages 494–497, Singapore, 1996. World Scientific.

[13] 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.

[14] K. Takahashi. A real-coded genetic algorithm using distance dependent alternation model for complex function optimization. pages 219–227, March 2000.

[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.