

Multimodal Function Optimization Using Local Ruggedness Information

Jian Zhang and Xiaohui Yuan and Bill P. Buckles

Department of Electrical Engineering and Computer Science

Tulane University

New Orleans, LA 70118

{zhangj, yuanx, buckles}@eecs.tulane.edu

Abstract

In multimodal function optimization, niching techniques create diversification within the population, thus encouraging heterogeneous convergence. The key to the effective diversification is to identify the similarity among individuals. Without knowledge of the fitness landscape, it is usually determined by uninformative assumptions. In this article, we propose a method to estimate the sharing distance for niching and the population size. Using the Probably Approximately Correct (PAC) learning theory and the ϵ -cover concept, we prove a PAC neighborhood of a local optimum exists for a given population size. The PAC neighbor distance is further derived. Within this neighborhood, we uniformly sample the fitness landscape and compute its subspace fitness distance correlation (FDC) coefficients. An algorithm for estimating the granularity feature is described. The sharing distance and the population size are determined when above procedure converges. Experiments demonstrate that by using the estimated population size and sharing distance an Evolutionary Algorithm (EA) can correctly identify multiple optima.

Introduction

Evolutionary Algorithms have been successful in solving single-optimum problems, such as pattern recognition (Dasgupta & Michalewicz 1997) and image processing (Yuan, Zhang, & Buckles 2002). When optimizing complex problem with many local optima, EAs suffer from premature or slow convergence. To overcome difficulties imposed by multiple local optima, hybrid EAs incorporating local search are developed. Such techniques include clustering (Törn 1978), stochastic approximation (Liang, Yao, & Newton 1999) and parallel local search (Guo & Yu 2003). Meanwhile, niching or speciation is a particular mechanism that allows and maintains several subpopulations so that each optimum can attract a number of them (Mahfoud 1995). Among niching strategies, sharing is an approach that divides the fitness of an individual by the number of “similar” individuals. Determining the similarity among individuals is nontrivial and is usually based on user assumptions.

In our previous work (Zhang, Yuan, & Buckles 2003), we found that population size is problem-dependent and can

be estimated by analyzing the ruggedness of a fitness landscape. With the information on population size, we are able to determine the sharing distance as well. This is, however, infeasible for a free-form function optimization problem without an impractical number of samples. Hence, we propose an approximating approach that employs the Fitness Distance Correlation coefficient (Jones & Forrest 1995) as a measure of local ruggedness. FDC has been developed as a measure of problem difficulty for Genetic Algorithms. Generally, one FDC coefficient is unable to uncover the variation in ruggedness of a fitness landscape. Moreover, computing FDC requires the knowledge of the global optimum, which is usually unknown. Based on PAC learning theory, we divide the search space into subspaces. A few samples are drawn from each subspace and the largest one is assumed to be the “local” optimum¹. Subspace FDC coefficients are computed and used as a guide for further subgrouping. The granularity of a fitness landscape is obtained by iterating this procedure until it converges.

The rest of this article is organized as follows. In the next section, we briefly discuss the concept of sharing distance in niching techniques and PAC learning theory. In Section “PAC Neighborhood Distance” we prove that based on the initial population a neighborhood distance exists. The analytical result is developed using PAC learning theory. In Section “Granularity of A Fitness Landscape”, we present the concept of subspace FDC as a measure for subspace granularity. An iterative algorithm for the estimation of the overall granularity is described next. The sharing distance and population size are estimated in the next section. In Section “Experiments and Discussions”, we demonstrate the results on 1-dimensional and 2-dimensional functions. This article is closed with conclusions.

Background

Sharing Method

Sharing (Goldberg & Richardson 1987) is a popular and successful niching method. It attempts to maintain a diverse population with members distributed among niches in a multimodal fitness landscape. To diversify its population, it reduces the fitness of an individual within a neighborhood de-

¹Without loss of generality, we assume the optimization is to find the maximum.

defined by the sharing function. This rewards individuals that uniquely exploit regions of the fitness landscape by discouraging redundant solutions. The shared fitness is defined as

$$f_{o_i}(d_{ij}) = \frac{f_i}{\sum_{j=1}^m sh(d_{ij})} \quad (1)$$

where f_i is the raw fitness of individual a_i , m is the number of individuals in the population, and d_{ij} is the distance between the i th and the j th individuals. The sharing function $sh(\cdot)$ reaches a maximum of 1 at zero, decreases monotonically with distance, and falls to zero for distances that are greater than σ_{sh} . For example, the triangular sharing function shown in Figure 1 is given by

$$sh(d_{ij}) = \begin{cases} 1 - \frac{d_{ij}}{\sigma_{sh}} & \text{if } d_{ij} < \sigma_{sh} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $sh(d_{ij})$ measures the amount of sharing or similarity between two individuals. The parameter σ_{sh} is vital to the

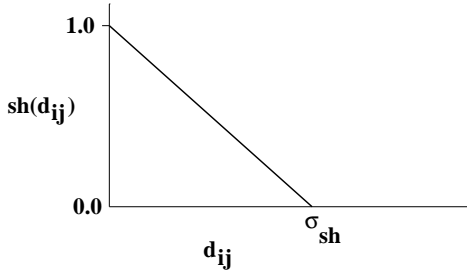


Figure 1: Triangular sharing function

performance of the sharing method and, ideally, it should approximate the peaks' widths. Unfortunately, in many real applications the number of peaks is unknown. We propose a local search algorithm to estimate the granularity, which is a ruggedness measure, of fitness landscape. Without knowing the number of peaks, we derive a sharing distance from the estimated granularity.

PAC Learning and Population Size

In this section, we briefly review PAC learning theory and the prior research on using PAC learning to study population size.

Let \mathcal{X} be the set of all possible instances over which the target function is defined and \mathcal{C} refer to some subset of target concepts that a learner might learn. The learner considers a *hypothesis space* \mathcal{H} when attempting to learn the target concept $c \in \mathcal{C}$. After observing a sequence of examples, a learner must output a hypothesis $h \in \mathcal{H}$ that is a good approximation of c with high probability. In order to measure the extent to which the output hypothesis h approximates the actual target concept c , the *error* of hypothesis h with respect to c and a distribution \mathcal{D} is defined. \mathcal{D} is the probability that h mismatches an instance drawn randomly according to \mathcal{D} ,

$$error(h) \equiv P_r(c(x) \neq h(x))$$

where the probability P_r is over the distribution \mathcal{D} . "High probability" is indicated by the confidence parameter δ . The

number of training examples that is required by the learner largely determines PAC-learnability. The minimum number of training samples that is needed to attain PAC-learnability is *sample complexity* (Mitchell 1997). The sample complexity m for a finite hypothesis space is

$$m \geq \frac{1}{\phi} (\ln(|\mathcal{H}|) + \ln \frac{1}{\delta}) \quad (3)$$

where ϕ is the bound on $error(h)$, $|\mathcal{H}|$ is the size of hypothesis space \mathcal{H} .

It is suggested that sample complexity and population size are similar concepts (Hernández-Aguirre, Buckles, & Martínez-Alcantara 2000). Based on the ruggedness of a fitness landscapes, we applied PAC learning theory to generate a problem-dependent population size for real-coded EAs (Zhang, Yuan, & Buckles 2003).

$$\tilde{m} = \lceil \frac{1}{\phi} (\ln \lceil \frac{1}{\phi} \rceil + \ln \frac{1}{\delta}) \rceil \quad (4)$$

where $\phi = g(\tau)/|\mathcal{S}|$, $g(\tau) = \tau$ when $n = 1$; $g(\tau) = \pi\tau^2/4$ when $n = 2$. $\lceil 1/\phi \rceil$ defines the size of hypothesis space, such that with confidence δ , $0 < \delta < 1$, the initial population forms an ϵ -cover of search space \mathcal{S} with probability greater than $1 - \delta$ and $\epsilon = \tau$.

To estimate the granularity of fitness landscape, a decomposition of the fitness function is performed. The granularity of the fitness landscape is defined as the period of the frequency component after which 10 percent of energy is ignored. The proposed PAC population sizes for real-coded EAs are shown to be effective and efficient. However, in situations where fitness function is piece-wise or discontinuous, the decomposition-based approach is no longer applicable. Therefore, a numerical method shall be developed to estimate the granularity. In the following sections, we describe a PAC learning based iterative granularity approximation method.

PAC Neighborhood Distance

Given the granularity of a function, an "ideal" initial population size can be determined analytically using PAC learning theory (Zhang, Yuan, & Buckles 2003). However, in most function optimization problems, the granularity is unknown. To infer such knowledge from a free-form function is non-trivial. Fortunately, PAC learning theory is a statistical method that connects the population size and granularity in an EA-based optimization problem.

The concept of ϵ -cover is defined in a pseudometric space. Let W denote a set and function $\xi : W \times W \rightarrow R_+$ be a pseudometric. Therefore, (W, ξ) forms a pseudometric space (metric space). A set $C = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$ is an ϵ -cover of S , $S \subseteq W$, when the following two constraints are satisfied:

1. $\epsilon > 0$,
2. $\forall x \in S, \exists \gamma_i \in C$ such that $\xi(x, \gamma_i) \leq \epsilon$.

Based on PAC population size bounded by Equation 4, the following theorem can be proven that infers a neighborhood distance from a given population size.

Theorem

Draw a population of individuals uniformly on a fitness landscape. Assume the population size is m , then it forms an ϵ -cover of the n -dimensional search space \mathcal{S} , where the PAC neighborhood distance posed by initial individuals is bounded by ϵ with probability greater than $1 - \delta$, $0 < \delta < 1$. The PAC Neighbor Distance α is given by

$$\alpha = \frac{2|\mathcal{S}|}{\sqrt{\ln(\frac{1}{\delta} - 1)^2 + 4m - (\ln \frac{1}{\delta} - 1)}} \quad (5)$$

where $|\mathcal{S}|$ denotes the size of space \mathcal{S} .

Proof. Given a population size m , using Equation 4 we have

$$m = \frac{1}{\phi} \left(\ln \frac{1}{\phi} + \ln \frac{1}{\delta} \right)$$

where $\phi = g(x)/|\mathcal{S}|$. By substituting $\frac{1}{\phi}$ with Y , we get

$$m = Y \left(\ln Y + \ln \frac{1}{\delta} \right). \quad (6)$$

Since $\ln(1 + x) \leq x$ holds for $x > -1$,

$$\ln Y \leq Y - 1, \text{ for } Y > 0. \quad (7)$$

Combining Equation 6 and Equation 7, we have the following inequality

$$m \leq Y \left[(Y - 1) + \ln \frac{1}{\delta} \right] \quad (8)$$

Two solutions exist for Equation 8. Considering the sign of PAC distance, we shall have the unique positive solution

$$\alpha = \frac{2|\mathcal{S}|}{\sqrt{\ln(\frac{1}{\delta} - 1)^2 + 4m - (\ln \frac{1}{\delta} - 1)}} \quad (9)$$

□

Let $\epsilon|_m$ denote the granularity given a population size m . $\epsilon|_m$ is approximated with a hypersphere defined by α , i.e. α is the radius of this hypersphere. For 1-dimensional and 2-dimensional cases, the granularity of a given population size m , denoted by $\epsilon|_m$, are the functions of α as $\epsilon|_m = \alpha$ and $\epsilon|_m = \sqrt{\frac{4\alpha}{\pi}}$ respectively.

We notice that Equation 7 provides a loose upper bound on $\ln(x)$ for $x > 0$. The neighborhood distance is decided more precisely if a tighter bound on $\ln(x)$ can be found. When using real numbers to represent individuals, the mapping process from genotype to phenotype is trivial. Thus the continuous fitness landscapes structure is not affected by applying genetic operators.

Granularity of A Fitness Landscape

To infer the granularity of a fitness landscape, we adopt the concept of fitness distance correlation (FDC).

Given m individuals $\vec{x}_i, i = \{1, 2, \dots, m\}$ sampled from an unknown fitness landscape, each individual maps to a fitness value $f(\vec{x}_i)$. Let $D = \{d_i, i = \{1, 2, \dots, m\}\}$ denote

the distance to the global optimum. FDC, denoted by R , is computed as

$$R = \frac{c_{FD}}{\sigma_F \sigma_D} \quad (10)$$

where

$$c_{FD} = \frac{1}{n} \sum_{i=1}^m (f(\vec{x}_i) - \bar{f})(d_i - \bar{d})$$

is the covariance of F , $F = \{f(\vec{x}_i), i = \{1, 2, \dots, m\}\}$, and D . $\sigma_F, \sigma_D, \bar{f}$ and \bar{d} are the standard deviations and means of F and D respectively.

Given the PAC neighborhood distance $\mathcal{D}_{PAC} = \epsilon|_m$ in a n -dimensional search space \mathcal{S} , an initial individual is represented as a vector, i.e. $\vec{x} = \{x^1, x^2, \dots, x^n\}$. Thus, we have a neighborhood about the k th dimension, i.e. $[x^k - \mathcal{D}_{PAC}/2, x^k + \mathcal{D}_{PAC}/2]$. A new individual, \vec{x}^k , is randomly drawn such that $\vec{x}^k = \{x^1, x^2, \dots, \hat{x}^k, \dots, x^n\}$ and $x^k - \mathcal{D}_{PAC}/2 \leq \hat{x}^k \leq x^k + \mathcal{D}_{PAC}/2$.

When maximizing an ideal fitness function, the fitness is expected to increase as distance to global optima decreases and the FDC coefficient is always negative, i.e. $R < 0$ (Jones & Forrest 1995). Therefore, we know that when $R > 0$ the search is led away from the optimum. In a multimodal situation, it implies a valley between two local optima. Figure 2 illustrates examples of an ideal fitness function and a multimodal function.

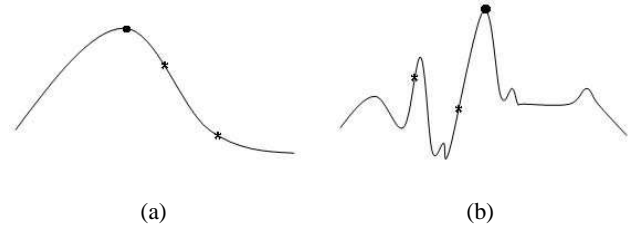


Figure 2: (a) is an ideal fitness landscape. The dot locates the optimum and stars are the search individuals. In a multimodal fitness landscape, the FDC is positive ($R > 0$) given the search individuals and optimum as shown in (b).

With the initial population of size m , we split the search space \mathcal{S} into m subspaces, i.e.

$$\mathcal{S} = \bigcup_{i=1}^m \mathcal{S}_i$$

In each subspace \mathcal{S}_i , l samples are randomly drawn on its k th dimension², i.e. $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_l\}$, which include the individual in the initial population. We assume the ‘‘local’’ optimum is within these samples and denote it as \vec{x}^* ,

²We consider one dimension at a time by holding the remaining dimensions unchanged. Therefore, drawing a sample on the k th dimension varies only on the k th component of a vector \vec{x} . Also notice that such random drawings are bounded by the PAC neighborhood, i.e. the new value \vec{x}^k is in the range of $[x^k - \mathcal{D}_{PAC}/2, x^k + \mathcal{D}_{PAC}/2]$.

$f(\vec{x}^*) = \max(f(\vec{x}_i))$. Hence a $l - 1$ subset is generated that excludes the “local” optimum. Denote this subset as X^{-*} , $X^{-*} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_l\} - \{\vec{x}^*\}$. The fitness of X^{-*} is F^{-*} and the sample distance to \vec{x}^* is D^{-*} . The FDC of this sample set is computed using Equation 10. We name it as *Subspace FDC*, denoted by r .

To reduce computational complexity and to capture the local landscape feature, the number of samples selected from the subspace shall be small. From our experimental study, a maximum of 4 samples are adequate to satisfy the computation requirements. Based on subspace FDC, an algorithm is described below that estimates subspace granularity, $\tilde{\tau}$.

Algorithm – 1

-
1. Randomly choose the initial population size as m and compute \mathcal{D}_{PAC} using Equation 5;
 2. Compute fitness $f(\vec{x}_i)$ for $\vec{x}_i, i = 1, 2, \dots, m$;
 3. On the k th dimension,
 - (a) Generate two neighbors of \vec{x}_i , \vec{x}_i^1 and \vec{x}_i^2 , and compute $r^3 = r(\vec{x}_i, \vec{x}_i^1, \vec{x}_i^2)$ using Equation 10;
 - (b) Randomly generate another neighbor of \vec{x}_i , \vec{x}_i^3 , and compute $r^4 = r(\vec{x}_i, \vec{x}_i^1, \vec{x}_i^2, \vec{x}_i^3)$;
 - (c) When $\text{sgn}(r^3) + \text{sgn}(r^4) \geq 0$, $\tilde{\tau}_i = 2|(\arg \max_A F - \arg \min_A F)|$, where

$$\text{sgn}(x) = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$

$$A = \{\vec{x}_i, \vec{x}_i^1, \vec{x}_i^2, \vec{x}_i^3\}, \text{ and } F = \{f(\vec{x}_i), f(\vec{x}_i^1), f(\vec{x}_i^2), f(\vec{x}_i^3)\},$$
 When $\text{sgn}(r^3) + \text{sgn}(r^4) < 0$, $\tilde{\tau}_i = \mathcal{D}_{PAC}$.
 4. Repeat from step 3 for all dimensions.
 5. Repeat from step 2 for all subspaces.
-

Using Algorithm 1, we compute the subspace granularity of a fitness landscape of an n -dimensional function. Let $\hat{\tau}_k$ and $\sigma_{\hat{\tau}_k}$ denote the mean and standard deviation of local granularity of the k th dimension respectively, the granularity of a fitness landscape is computed as

$$\tau = \hat{\tau}_k |_{\sigma_{\hat{\tau}_k} = \min_k(\sigma_{\hat{\tau}_k})} \quad (11)$$

Sharing Distance and Population Size

The sharing distance and the population size are closely related and can be derived from the granularity of a fitness landscape.

Combining granularity τ computed from Equation 11 with Equation 4, the population size can be computed as

$$\tilde{m} = \lceil \frac{1}{g(\tau)/|\mathcal{S}|} (\ln \lceil \frac{1}{g(\tau)/|\mathcal{S}|} \rceil + \ln \frac{1}{\delta}) \rceil$$

where in 1-dimensional function, $g(\tau) = \tau$ and in 2-dimensional function, $g(\tau) = \pi\tau^2/4$. This re-calculated

population size is more suitable for exploring the search space.

When approximating the sharing distance, we consider a local optimum as the mid-point of a peak. Therefore, the sharing distance σ_{sh} is calculated as half of the estimated granularity of a fitness landscape, i.e.

$$\sigma_{sh} = \frac{1}{2}\tau \quad (12)$$

Experiments and Discussions

Table 1 lists three 1-dimensional and one 2-dimensional multimodal functions used in our experiments. Function

Table 1: Fitness functions.

Function	Domain
$f_1(x) = \sin^6(5\pi x)$	$x \in [0, 1]$
$f_2(x) = e^{-2 \log 2 \cdot (\frac{x-0.1}{0.8})^2} \cdot \sin^6(5\pi x)$	$x \in [0, 1]$
$f_3(x) = \sum_{i=1}^{10} \frac{1}{(k_i(x-a_i))^2 + c_i}$	$x \in [0, 10]$
$f_4(x) = \begin{cases} h - \frac{2hx^2}{r^2} & \text{if } x < \frac{r}{2} \\ \frac{2h(x-r)^2}{r^2} & \text{if } \frac{r}{2} \leq x < r \\ 0 & \text{otherwise} \end{cases}$	$r \in [0.02, 0.1]$ $h \in [0.1, 1]$

$f_1(x)$ has 5 equally spaced maxima of equal height 1.0. Function $f_2(x)$ is similar to $f_1(x)$, but its maxima decrease in height exponentially from 1.0 to 0.25. The global optimum of Shekel’s Foxhole function, $f_3(x)$, is located at $x = 0.699$ with fitness value of 15.7206.

In the bell function $f_4(x)$, r is the radius of the cone, h is the height and x is the Euclidean distance from the center of the cone. With a different number of peaks and different values of r and h , this function provides tunable complexity. In our experiments, 30 peaks are randomly generated in a 2-dimensional space. The radii of the bells are generated with values in the range from 0.02 to 0.1 and the bell heights range from 0.1 to 1. The maximum locates at $(x_1, x_2) = (0.91, 0.67)$ and its function value of 1.4728. Due to the randomness, some cones are overlapped or too small to be shown. Therefore 24 maxima are considered, which are illustrated in Figure 4(a).

For each test function, 50 runs are performed. A PAC population size is generated based on Equation 4. Tests are terminated upon a pre-determined number of generations. In the experiments for 1-dimensional functions, the maximum generation is 30. In the experiments for 2-dimensional bell function, we set the maximum generation to 50. Figure 3 shows the typical population distribution for 1-dimensional functions. The output for the bell function is shown as a contour map in Figure 4(b).

Table 2 lists the results for our test functions. To demonstrate the effectiveness and efficiency of using estimated granularity in multimodal function optimization, we record the mean and the standard deviation of global maximum found in 50 runs, $\bar{\mathcal{M}}$ and $\sigma_{\mathcal{M}}$ respectively. We also average the number of generations at which the maximum values are found, $\bar{\mathcal{G}}$. It is shown that the average numbers of generations for 1-dimensional functions are below 20, and

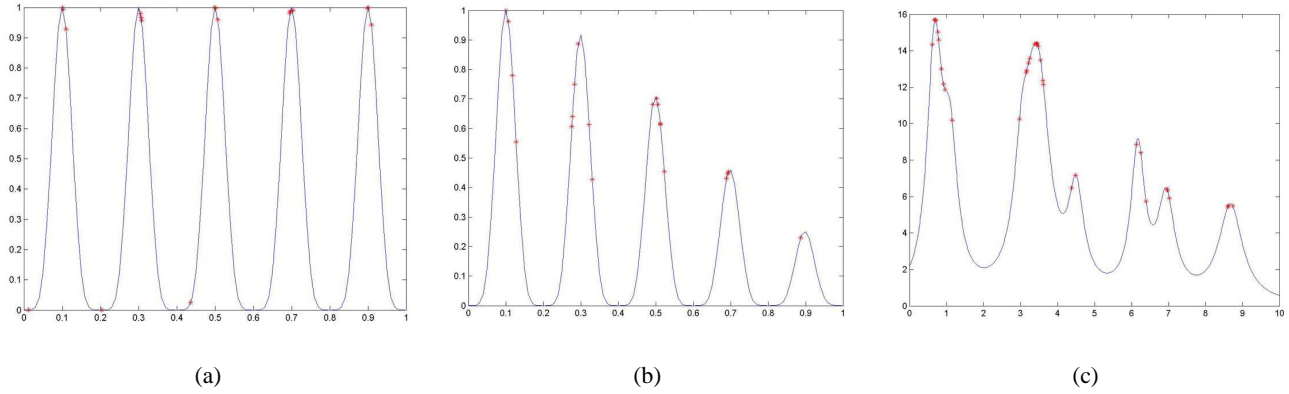


Figure 3: Population distribution on three experimental functions (a) $f_1(x)$ contains 5 equal maxima, (b) $f_2(x)$ has one global maximum and 4 local maxima, and (c) Foxhole function $f_3(x)$ also has one global maximum and 5 irregular local maxima. The dots represent positions of EA individuals at the termination point.

for 2-dimensional function it takes around 30 generations. By comparing the average number of peaks found, pf , with the actual number of peaks, No. Max., we show that EAs converge correctly with small variations in evaluations, σ_{pf} . Our results show that the global convergence is reliable for

Table 2: Convergence performance on $f_1 - f_4$ for 50 runs. NoG denotes the number of generations EA takes to converge. Pop denotes the population size.

Function	f_1	f_2	f_3	f_4
Pop	18	20	35	76
σ_{sh}	0.1071	0.1017	0.6335	0.145
NoG	30	30	30	50
\mathcal{M}	0.999	0.9895	15.697	1.446
$\sigma_{\mathcal{M}}$	0.001	0.015	0.062	0.057
\mathcal{G}	15.54	14.88	17.98	31.14
No. Max	5	5	6	23
pf	4.84	4.46	5.42	18.32
σ_{pf}	0.37	0.5	0.67	1.659

all tests performed. In the peer work (Guo & Yu 2003), it takes more than a hundred generations to converge. Our algorithm exhibits a tremendous reduction in terms of evaluation time. This shows that the initial populations are within vicinity of the optima. At the same time, a majority of local optima are found with reduced search. We observe that over 90% of local optima are located for 1-dimensional tests, and over 18 local optima are found for bell function. Obviously, the local search obtains a rough knowledge of the ruggedness feature of fitness landscapes. Therefore, the estimated population size and the sharing distance are more appropriate for multimodal function optimization than randomly generated parameters.

Conclusions

By analyzing fitness distance correlations in subspaces within the initial population, we describe a granularity approximation scheme for fitness landscapes. The knowledge gained is further used to estimate the sharing distance as well as the population size. The sharing method requires determination of the sharing distance, which is usually based upon assumptions about the number of peaks and their distribution. Population size, one important factor in multimodal function optimization, has influences on the formation of stable subpopulations. A small population is not able to fully explore the search space, thus miss local optima. On the other hand, a large population takes longer to converge and form stable groups around local optima. Knowledge of the ruggedness feature of certain problems is critical in deciding both sharing distance and population size for EAs.

Our method divides the search space into subspaces based on the PAC neighborhood distance. It is grounded in PAC learning theory and computes subspace FDC coefficients. In the computation, a few samples are drawn from each subspace and the largest one is assumed to be the “local” optimum. Using subspace FDC, granularity of a fitness landscape is estimated and is employed to set the sharing distance and population size for an EA. Our experiments verify that the population size and the sharing distance are appropriate for quick convergences on multiple local optima including the global optimum.

References

- Dasgupta, D., and Michalewicz, Z., eds. 1997. *Evolutionary Algorithms in Engineering Applications*. Berlin, Germany: Springer.
- Goldberg, D., and Richardson, J. 1987. Genetic algorithms with sharing for multimodal function optimization. In Grefenstette, J. J., ed., *Proc. 2nd Intern. Conf. on Genetic Algorithms*, 41–49. Hillsdale, NJ: Lawrence Erlbaum Associates.

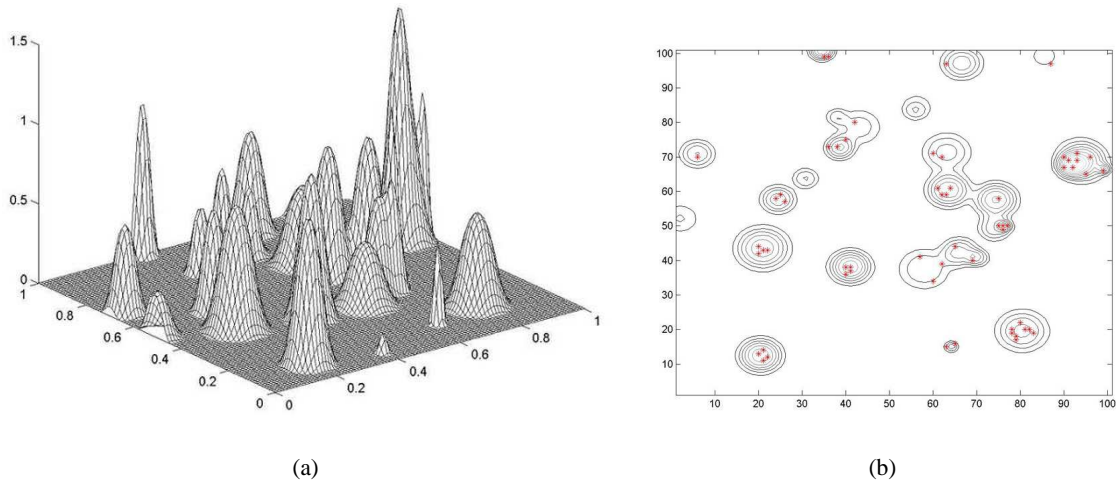


Figure 4: Optimizing 2-dimensional Bell function. (a) is the 3-dimensional illustration. The contour image in (b) shows the distribution of EA individuals (represented by dots) in the fitness landscape.

Guo, G., and Yu, S. 2003. Evolutionary parallel local search for function optimization. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*.

Hernández-Aguirre, A.; Buckles, B. P.; and Martínez-Alcantara, A. 2000. The PAC population size of a genetic algorithm. In *the 12th International Conference on Tools with Artificial Intelligence*, 199–202.

Jones, T., and Forrest, S. 1995. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In Eshelman, L., ed., *Proceedings of the Sixth International Conference on Genetic Algorithms*, 184–192. San Francisco, CA: Morgan Kaufmann.

Liang, K.-H.; Yao, X.; and Newton, C. 1999. Combining landscape approximation and local search in global optimization. In *Proceedings of the Congress on Evolutionary Computation*, volume 2, 1514–1520. IEEE Press.

Mahfoud, S. 1995. *Niching Methods for Genetic Algorithms*. doctoral dissertation, University of Illinois at Urbana-Champaign.

Mitchell, T. M. 1997. *Machine Learning*. McGraw-Hill.

Törn, A. 1978. a search-clustering approach to global optimization. In Dixon, L., and Szego, G., eds., *Toward Global Optimization 2*. Amsterdam: North-Holland. 49–62.

Yuan, X.; Zhang, J.; and Buckles, B. P. 2002. Feature based approach for image correspondence estimation. In *Proceeding of the 2nd Int'l Conf. on Image and Graphics*.

Zhang, J.; Yuan, X.; and Buckles, B. P. 2003. Sample complexity of real-coded evolutionary algorithms. In *Proceedings of the 16th International FLAIRS Conference*.