Characterizations of Trajectory Structure of Fitness Landscapes Based on Pairwise Transition Probabilities of Solutions*

Márk Jelasity
Research Group of Artificial Intelligence, JATE,
H-6720 Szeged, Aradi Vértanúk tere 1, Hungary
jelasity@inf.u-szeged.hu

Boglárka Tóth, Tamás Vinkó
Students of Department of JATE University,
H-6720 Szeged, Árpád tér 2, Hungary
h531733,h531661@stud.u-szeged.hu

Abstract- Characterization of trajectory structure of fitness landscapes is a major problem of evolutionary computation theory. In this paper a hardness measure of fitness landscapes will be introduced which is based on statistical properties of trajectories. These properties are approximated with the help of a heuristic based on the transition probabilities between the elements of the search space. This makes it possible to compute the measure for some well-known functions: a ridge function, a long path function, a fully deceptive function and a combinatorial problem: the subset sum problem. Using the same transition probabilities the expected number of evaluations needed to reach the global optimum from any point in the space are approximated and examined for the above problems.

1 Introduction

This paper is concerned with the characterization of fitness landscapes w.r.t. optimizers that use a stochastic hill-climbing heuristic. Members of the field of evolutionary computation belong to this class of optimizers. The definition of *hardness*, maybe the most important feature of a landscape is far from clear (see [NK98] for a summary of the available measures) so we need to clarify the problem we would like to tackle.

1.1 Theory or Practice?

The first question is whether one would like to give a *method* for characterizing fitness functions in practice i.e. a method which can be used to make predictions on interesting existing (maybe black box) problems or one would like to gain *theoretical insights* of the working of the optimizer.

In the first case the method needs to run fast; preferably faster than the optimizer itself otherwise it would be easier to run the optimizer and see what happens. There are attempts to give such practically useful measures like those based on the observed trajectories of several runs with some fixed set of parameters [Kal98]. These approaches are useful for e.g. studying the effects of certain parameter settings but from a theoretical point of view they have their limitations especially in the case of black box functions. The main problem is that these methods do not take the whole search space into account but instead only a very small fraction of it; one needs to know at least the global optimum to give a useful characterization. The two well-known measures: fitness distance correlation (FDC) [JF95] and epistasis variance [Dav91] do not belong to this first class; both need to know all the solutions in order to make predictions. If based only on a relatively small sample these measures can be highly misleading as described in [NK98] and as confirmed by our own experience.

Theoretically motivated measures need not be computed efficiently but they should be capable of providing a characterization of easy and hard functions and maybe useful in suggesting constructed easy and hard problems. While the characterization suggested here belongs to this class, techniques will be suggested that make it possible to examine relatively small problems empirically.

1.2 Multimodality

Another question is multimodality. The main criterion of success seems to be finding the global optimum. This practice has serious drawbacks however. The first is that there can be multiple global optima that may be arbitrarily far from each other which makes the original FDC meaningless. This problem is not serious since it is possible to choose the distance from the closest global optimum. The more serious problem is that in engineering practice where evolutionary computa-

*M. Jelasity, B. Tóth, and T. Vinkó. Characterizations of trajectory structure of fitness landscapes based on pairwise transition probabilities of solutions. In *Proceedings of the 1999 Congress on Evolutionary Computation (CEC99)*, pages 623–630. IEEE Press, 1999. This research was supported by the grant OTKA T25721 of the Hungarian Academy of Sciences.

tion has its main applications it is not always necessary to find a global optimum. Intuitively, it is enough to find solutions that are judged good enough by the engineers. This means that local optima are not simply obstacles in the way of success; their distribution and the structure of their attraction areas are essential from the point of view of problem characterization.

For example if a function has a unique global maximum that is a "needle in the haystack" but in the same time it has another local optimum which is almost as good as the global one but it has a large area of attraction then this function should be characterized as fairly easy. However, both FDC and epistasis variance would predict that it is a hard problem if the global and local optima are far from each other. Our approach emphasizes the role of the local optima.

1.3 One Number?

The practice of searching for a single number as a measure of problem hardness is is similar to the efforts of psychologists to characterize human intelligence with a single IQ. Both have the drawback of oversimplification. Our approach emphasizes the role of the interpretation of certain figures and our measure of difficulty is in fact a function of the stopping criterion of the search but is it possible to take other factors into account as well.

1.4 The Idea

The basic idea is to examine the trajectories of the space w.r.t. a given operator and stopping criterion. The ending points of these trajectories form a very interesting set: these are the points the search is expected to converge to. Statistical measures over this set w.r.t. some properties such as fitness or probability of being the result of the search are the best candidates for being a hardness measure. In this paper a measure of this kind will be suggested.

There are practical problems when such measures have to be computed since a huge amount of calculation is needed. As it was mentioned, it would be possible to obtain the trajectories of the search by running the algorithm and collecting the history of the process. Our approach is different. We define transition probabilities between the elements of the space. These values define the probability of every possible trajectory and also they make it possible to identify the points the algorithm is accepted to converge to. A heuristic for approximating the number of trajectories leading to a given point will be given and statistical measures based on this data will provide the hardness measure.

Based on these transition probabilities it is also possible to approximate the expected number of evaluations needed to get from a point to a given other point. This values can be used as distance measures and plots can be drawn that depict the convergence relations.

2 Basic Notions

This section introduces a model of stochastic hill-climbing search. This model will be used in two ways. The first application is an iteration formula used for approximating the expected number of steps of reaching the global optimum or any set of solutions from a point of the search space. The second and maybe more original application is to characterize fitness functions using the notion of *endpoints*. Let $S = \{s_1, s_2, \ldots, s_M\}$ be the search space. Let us fix S to be the binary space $\{0,1\}^l$.

2.1 A novel distance notion

Let us define an ordering of the solutions as was done in [Vos91].

Definition 1 *We say that $s_i \leq s_j$ iff $f(s_i) \leq f(s_j)$, where f is an arbitrary fitness function of type $S \to \mathbb{R}$. This means that s_M is a global optimum.*

For every mutation operator the probability of getting from a given solution to another one can be given. For example the operator of the stochastic hill-climber is that every bit in the solution is flipped with a given t probability. In this case the probability of getting from a given solution to another one is $t^d(1-t)^{l-d}$, where d is the number of different bits and l is the length of the solutions. This model is used in this paper. However, any mutation or other genetic operator can be chosen.

Let $Pr_{ik}^{(j)}$ denote the probability of getting from s_j to s_k via the application of the mutation operator j times. This notation is important for the main iteration formula. First of all we have to compute the values $Pr_{ik}^{(1)}$ for every index. Let s_i and s_k be solutions, let the probability of flipping a bit be t, let the Hamming distance between s_i and s_k be d and let the length of a bit string be l. Then

$$Pr_{ik}^{(1)} = \begin{cases} t^d(1-t)^{l-d} & \text{iff } s_i < s_k \\ \sum_{s_i \geq s_k} t^d(1-t)^{l-d} & \text{iff } i = k \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Note that the case $i = k$ is very important, because $Pr_{ii}^{(1)}$ denotes the probability of no motion, i.e. the sum of the probabilities of constructing a worse solution by the mutation operator. The only exception is the case $i = M$, because at this point we have reached the global optimum and that is the end of the search, so we choose $Pr_{MM}^{(1)} = 0$. This guarantees that the following iteration formula will measure the distance from the global optimum appropriately. $Pr_{iM}^{(j)}$ is computed by the following iteration formula:

$$Pr_{iM}^{(j)} = Pr_{ii}^{(1)} Pr_{iM}^{(j-1)} + \sum_{s_k > s_i} Pr_{ik}^{(1)} Pr_{kM}^{(j-1)}, \tag{2}$$

where $j = 2, 3, \ldots$. This formula means that $Pr_{iM}^{(j)}$ is the probability of getting from s_i to the global optimum in j steps. If $i = M$ i.e. the solution is the global optimum then

$Pr^{(j)}_{MM}$ need not be computed because it is 0. The expected value of the number of steps from a solution to the global maximum is given by the limes of the series

$$E^{(j)}_{iM} = E^{(j-1)}_{iM} + jPr^{(j)}_{iM} \qquad (j = 2, 3, \ldots). \tag{3}$$

This formula gives a new distance notion for the stochastic hill-climbing search which is in fact the expected number of function evaluations. This distance is more expensive to calculate than the Hamming distance, which is the basis of the FDC, but it provides more accurate and more informative results. In the numerical experiments the following formula was used for checking convergence:

$$jPr^{(j)}_{iM} < (j-1)Pr^{(j-1)}_{iM} \quad \text{and} \quad jPr^{(j)}_{iM} < \varepsilon. \tag{4}$$

Note that if the expected number of steps is zero then the solution at hand is a global optimum or the global optimum cannot be reached at all. The later is not possible with the operator we are using.

2.2 Endpoints

Our definition of endpoints is motivated by the very simple though quite common stopping criterion: if the best solution does not improve after a given number of evaluations then the program will stop. A solution will be called an endpoint if it is not expected to improve in a given number of steps.

Definition 2 *s_i is endpoint if $P^{(1)}_{ii}$ is greater than a given bound (near 1).*

Note that a given bound K corresponds to a stopping criterion of $1/(1-K)$ steps without improvement. Also note that the set of endpoints depends on this parameter; the parameter of the stopping criterion. The notion of endpoints depends only on the transition probabilities just like the expected number of evaluations.

Though the notion of endpoints is independent from the calculation of the expected number of evaluations, it is possible to generalize our above formulas to handle the set of endpoints instead of the global optima alone. Let Z denote the set of the endpoints. Then

$$Pr^{(j)}_{iZ} = \sum_{z \in Z} \left(Pr^{(1)}_{ii} Pr^{(j-1)}_{iz} + \sum_{s_k > s_i} Pr^{(1)}_{ik} Pr^{(j-1)}_{kz} \right), \tag{5}$$

where $j = 2, 3, \ldots$. Note that in this case $Pr^{(1)}_{zi}$ $(z \in Z)$ is set to 0 for all $s_i \in S$, similarly to the case of the global optimum in Eq. (1). Eq. (3) is changed to

$$E^{(j)}_{iZ} = E^{(j-1)}_{iZ} + jPr^{(j)}_{iZ} \qquad (j = 2, 3, \ldots). \tag{6}$$

Note that Eq. (5) gives the probability of reaching one of the endpoints from s_i in j steps and Eq. (6) gives the expected value of the number of steps from a solution to an endpoint.

Endpoints are essential since they are the possible results of the search. This motivates our approach that statistical measures of certain properties of these endpoints are the best candidates for being a good hardness measure. Clearly Eq. (6) says nothing about a given endpoint, only about the set of endpoints while the most important question is: what is the probability of reaching a given endpoint starting from a random element of the space? This probability can be approximated using the transition probabilities which define a weighted graph over the search space S. Let us determine a spanning forest in this graph in the following way: let the roots of the trees be the endpoints and for every other point let us select the outgoing edge (transition) with the highest probability. It is easy to see that this method provides us with a spanning tree with a maximal weight in $O(S)$ time. Note that the edges of these trees point *towards* their roots.

The endpoints of a given function describe the expected results of the stochastic hill-climber. The probability of reaching a given endpoint can now be approximated with the proportion of the points of the tree rooted from the given point in S. For instance if $|S| = 100$ and the tree of a given endpoint contains 10 points then we say that the probability of reaching that endpoint is 0.1. Using this probabilities it is possible to approximate the average fitness of the results of multiple optimization runs. This values can be compared to the actual observed values as will be done in Section 3.3.

It is also possible to characterize the deceptiveness of the problem with these probabilities. Let us introduce our *deceptiveness coefficient*, a number from $[0, 1]$, which is based on the notion of endpoints.

Definition 3 *Let B_K be a set of bounds (minimal transition probabilities as parameters of being an endpoint) from $[a, 1)$ where $0 < a$ for a given bound $K \in B_K$. Let F_{min} be the minimum, F_{max} be the maximum of the fitness of endpoints, and let E be the expected value of fitness. Let*

$$s_K = 1 - \frac{E - F_{min}}{F_{max} - F_{min}}.$$

If $F_{max} - F_{min} = 0$ then let s_K be 0. Then the deceptiveness coefficient is the mean value of s_K, where K takes the values of B_K.

This number characterizes the problems: 1 indicates that the problem is misleading, 0 means that it is very friendly. In section 3 this coefficient will be shown for some well-known problems. Note that this coefficient depends on a set of parameters. It is possible to give a coefficient for every stopping criterion by using the one element set containing the bound corresponding to that criterion.

3 Empirical results

In this section our method will be demonstrated via empirical results. A stochastic hill-climber was used with the same operator which has been mentioned in the section 2. First, a survey will be given of the studied functions and some explanation on how to read the iteration and the other figures. Finally our heuristic for determining the probabilities of the endpoints will be validated via some empirical results.

3.1 Studied functions

Some well-known functions have been examined with our method. These are the Ridge function, the Long Path problem, Liepins and Vose fully deceptive problem and the Subset Sum problem. The iteration figures (figures that show the expected number of evaluations for the space) will be shown for these functions and the deceptiveness coefficient will also be discussed. Note that in the iteration figures noise has been added to the distances and fitnesses so that identical points can be distinguished. For all the examined problems size was set to 10 bits and the probability of the mutation operator of the hill-climber was 0.1.

3.1.1 Ridge function

Ridge functions were introduced by Quick, Rayward-Smith and Smith in [QRSS98]. In our experiments a 10-bit version was used. FDC predicts that ridge functions are very misleading while the hill-climber and the GA easily solves the function. Our results show that this function is easy. The ridge function hasn't got local optima as it is shown in the iteration figure. The iteration stopped after 488 steps, so the iteration figure shows that from every point in the search space the global maximum can be reached in 250 steps.

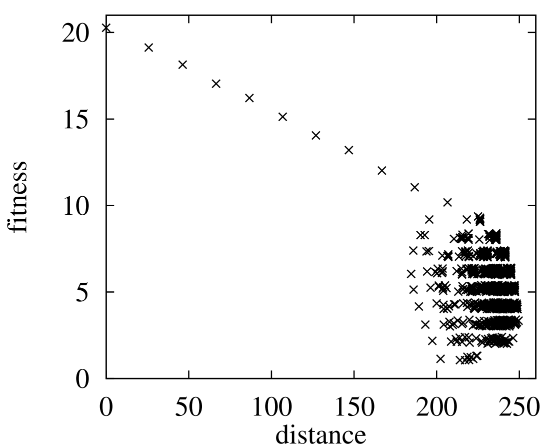

Figure 1: The iteration figure of the 10 bit Ridge function

The iteration figure shows that there is a path in the search space, as follows from the construction of the ridge function. On the other hand, the points which are not on the path were gathered at the beginning of it. It is also clear that from an arbitrary point the global maximum can be reached only via walking through the path.

Let us examine the deceptiveness figure with respect to the number of endpoints. Obviously, from bound 1 to 0.96 there is only one endpoint, the global maximum so it is predicted that the hill-climber always finds the global optimum. One can easily see that when the bound is 0.96 there are only two endpoints and every other point tends to the worst one. This is why the deceptiveness is 1. When the bound is under 0.96 the number of the endpoints suddenly starts to grow,

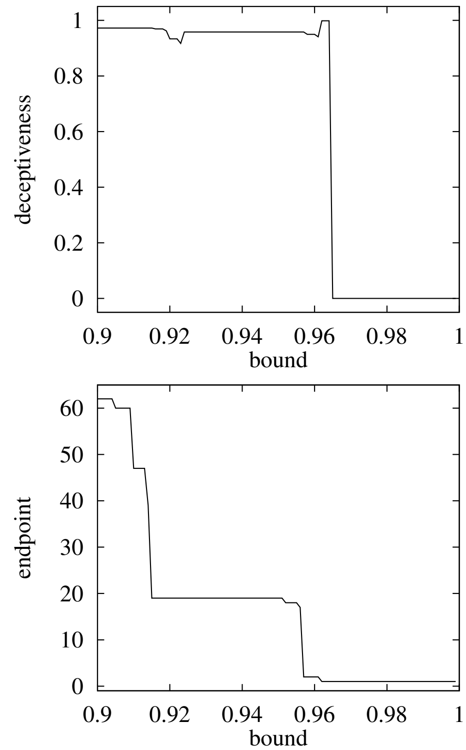

Figure 2: The deceptiveness figure and the number of endpoints of the 10 bit Ridge function

because all the points on the path become endpoints. This is the explanation of the jump of deceptiveness since as it has been mentioned the search process walks trough the path so most of the points tend to the beginning of it. Additional decreasing of the bound has essentially no effect on the value of deceptiveness because independently of the new endpoints all the remaining points tend to the beginning of the path.

Note that our coefficient which depends on the stopping criterion clearly shows that the friendliness of the ridge function as claimed by [QRSS98] heavily depends on the stopping criterion; the allowed probability of staying in place has to exceed 0.96 to make the problem friendly.

3.1.2 Long Path

Long Path problem is introduced by Horn, Goldberg and Deb in [HGD94]. This problem was constructed to be difficult for the hill-climber. First of all, note that the problem size was set to 11 bits, because this function is defined as odd bits problem. In this case formula (3) converged after 1257 steps for all the points.

The iteration figure clearly reflects the structure of this problem, surely there is a long path. In the case when some points with low fitness are closer to the global optimum than points which are at the beginning of the path show that these points do not have to walk through the path, they are able to jump into it. It is interesting that the shape of the path is not linear. It means that inside the path bigger steps can be

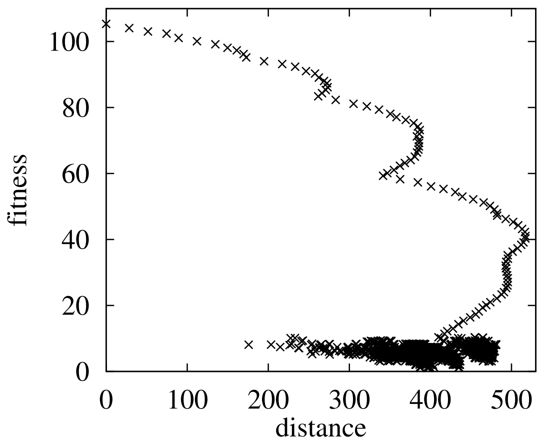

Figure 3: The iteration figure of the Long Path problem

taken i.e. there are shortcuts. Recall that the interpretation of our plots is essentially different from the interpretation of the FDC plots. In our case the distance is the expected number of evaluations for reaching the optimum from the given point and this distance is not a direct function of the encoding alone as in the case of Hamming distance. This is why the structure of the plot indicates shortcuts why a similar structure in an FDC plot would simply indicate deceptiveness.

In figure 4 from bound 1 to 0.96 the situation is the same as with the ridge function. Under 0.96 the points of the path gradually become endpoints. The deceptiveness of the function in the area where the points of the long path become endpoints as the bound decreases is lower than in the case of the ridge function. This is due to the above mentioned shortcuts. When all the points of the path become endpoints there is no significant change anymore. Note that since in this case the path is longer than the path of the ridge function deceptiveness does not increase as fast. The other interesting result to note is that our figures clearly show the structural similarity between the long path and the ridge function.

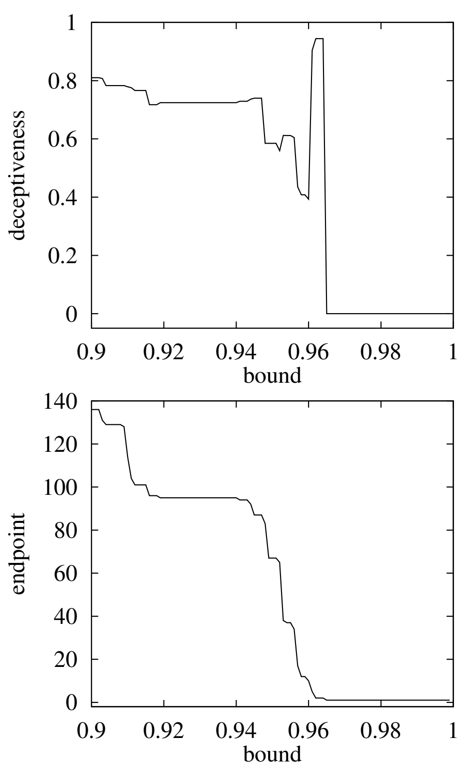

Figure 4: The deceptiveness figure and the number of endpoints of the Long Path problem

3.1.3 Liepins and Vose fully deceptive function

This problem was introduced in [LV91]. This function is fully deceptive. There is one global optimum and a local one. The global optimum is fairly independent of the whole structure of the function so no trajectories converge to it. The only possibility to find it is blind coin tossing.

If there is only one endpoint, the global maximum, then Eq. (3) converges for none of the points after 20000 iteration steps. It means that from all points of the search space the global optimum is unreachable under these conditions. That is why the iteration figure is meaningless though is clear that the points are very far from the optimum. For this reason Eqs. (5) and (6) were used.

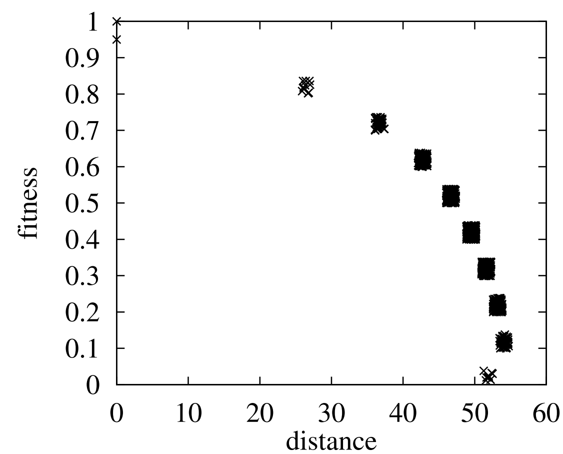

Figure 5: The generalized iteration figure of the Liepins and Vose function

The generalized iteration formula with bound 0.999 results in figure 5. There are two endpoints, the global and the local optimum. In this case all points convergent after 215 iteration steps, so it is clear that all points can easily reach the endpoints.

Examining the deceptiveness figure it is obvious that all points tend to the local optimum, even if the number of endpoints grows. It can be seen that this function is consistently deceptive independently of the stopping criterion (at least in the range we examined).

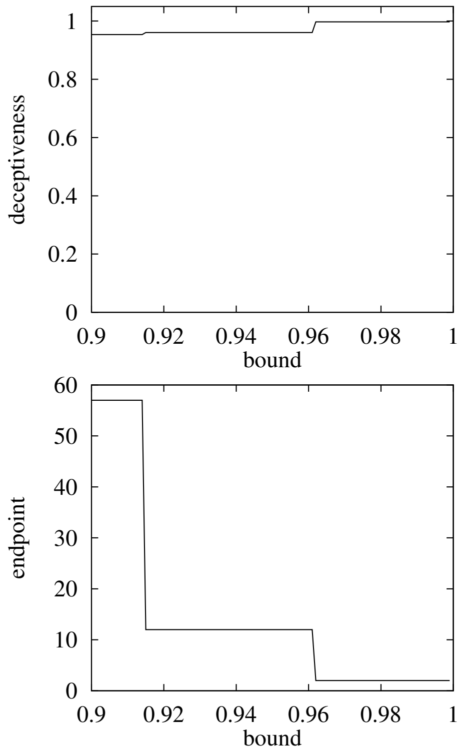

Figure 6: The deceptiveness figure and the number of endpoints of the Liepins and Vose function

3.1.4 Subset Sum problem

The subset sum problem is a combinatorial optimization problem. In this problem we are given a set $W = \{w_1, \ldots, w_n\}$ of n integers and a large integer N. We would like to find a $V \subseteq W$ such that the sum of the elements in V is closest to N. For more details see [KBH93]. We examined a 10 bit problem here.

Unlike the previous functions this is a highly multimodal function as it can be seen in the iteration figure. There

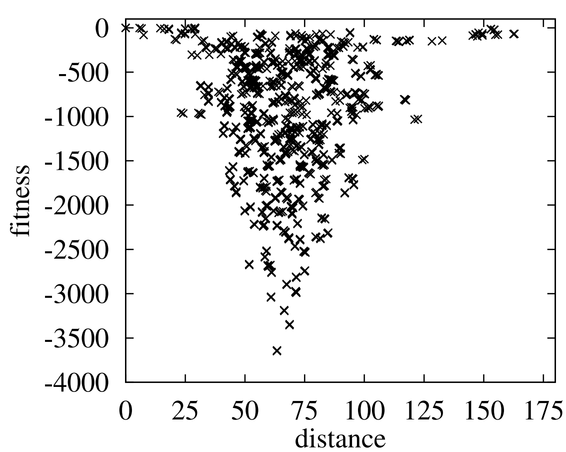

Figure 7: The iteration figure of the Subset Sum problem

are many points with high fitnesses which are far from the global optimum. Note that there are few points for which the formula had not converge, but the iteration figure correctly shows the structure of this function. Note that in this case the maximum iteration of formula (3) was 3000 and 1005 points converged. Let us note again that although the FDC plot of the subset sum problem is very similar in this case it does not mean that it means the same.

Let us examine the deceptiveness figure. Deceptiveness never goes above 0.5 so it shows that although this problem has many local optima most of the points tend to the endpoints with good fitness. Note that there are also large fallings and jumps because of the few endpoints, but the deviation of the deceptiveness is very little. So in this case the deceptiveness coefficient is very informative. An important observation about the figure of the number of endpoints is that in this case the number of endpoints grows smoothly. This is the reason why in the deceptiveness figure there is no plateau.

These results are in agreement with earlier investigations concerning the subset sum problem (e.g. [Jel97] and [KBH93]) where it was found that the problem is easy (with the suggested encoding and with the applied test problem generation method) and according to our earlier results the subset sum problem indeed has an enormous number of relatively good local optima and these local optima can be very far from each other.

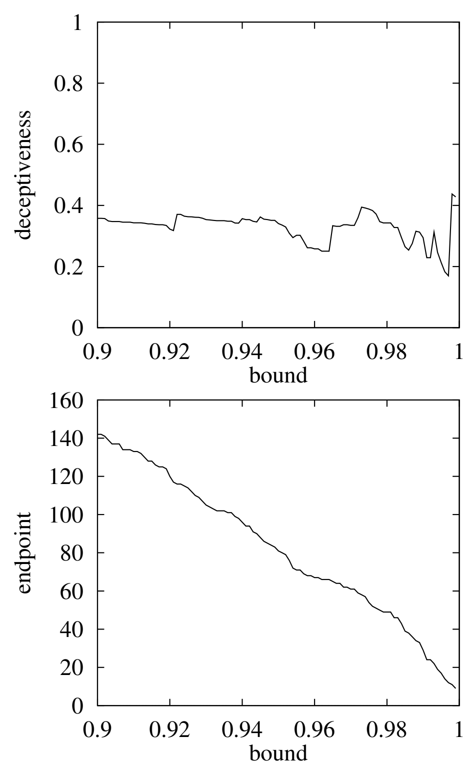

Figure 8: The deceptiveness figure and the number of endpoints of the Subset Sum problem

3.2 Discussion

This section gives a guide on how to read the figures in general. First notice that iteration figures are not correlation figures so the information shown by them is only indirectly related to our figures based in the approximated expected number of evaluation (referred to as iteration figures). If the iteration formulas converge for all points then the horizontal axis of the iteration figure shows the expected value of the steps from a given solution to the global optimum (or the set of endpoints). The structure of the plot is not necessarily important e.g. it is not necessarily good to have a linear form like in the case of FDC. However if there are many points with good fitness far from the global optimum then the given function has many good quality local optima.

The figure showing the deceptiveness coefficient as a function of the stopping criterion is very interesting. First of all it should be noted that the coefficient is most informative if the deviation is small. If the number of endpoints is small then the deviation may be large since as additional endpoints jump in with the decreasing bound the value of the coefficient may be altered significantly. However most of the real problems behave like the subset sum problem i.e. they have many good local optima. On the other hand, regions with small deviation carry much information about the structure of the problem; the sudden change of the deception coefficient in the case of the ridge and long path problems is related to their special structure.

function	dec.coeff.	deviation
Ridge	0.63	0.46
Long Path	0.46	0.36
Liepins and Vose	0.973	0.018
Subset Sum	0.33	0.046

Table 1: The deceptiveness and its standard deviation for the studied functions

Table 1 shows the deceptiveness coefficient and its standard deviation for the studied functions. It can be observed that in the case of the fully deceptive problem the result is correct and the same can be said about the Subset Sum problem. However, in the case of the other two functions these values are not so informative because of the large deviation but here the deceptiveness as the function of stopping criterion still accurately shows the structure of the space.

Finally note that according to this measure the ideal function is the constant function. This is quite evident since in that case we can reach the global optimum with zero evaluations with any kind of parameter setting. Strangely enough, there is a tendency among the known measures to regard the constant function as hard. FDC and the measure suggested in [NK98] are good examples. This is related to the observation that plateaus are hard for hill-climbers and if a partial sample of the space contains similar elements then the problem can be expected to be hard. However if the plateaus have good fitness then why would it be a problem to get stuck in them?

3.3 Validation of the Model

In this subsection we have a look at that the expected value of fitness as predicted by the same heuristic used for calculating the deceptiveness coefficient as was mentioned in section 2. For every bound the averages of 500 hill-climber runs are shown.

In figure 9 it can be seen that the sudden performance growth predicted by the spanning forest over the transition probability graph is not followed by the hill-climber. Note that at the sudden growth of the prediction the standard deviation of the average results of the hill-climber is very large.

On the other hand in the case of the Subset Sum problem the prediction coincides with the real result. This is related to the fact that in this case the number of the endpoints decreases smoothly with the increasing bound. In the case of the Liepins and Vose function the predictions follow the observed behavior. These plots show that the heuristic used for determining the weights of the endpoints is valid since using the weights the weighted sum of the fitnesses of the endpoints predict the result of the hill-climber quite well.

4 Summary

In this paper we suggested that the hardness measures of functions should take the local optima into account. The suggested measure is based on the fitnesses and weights of endpoints, the solutions to which the optimizer can be expected to converge. The weight is the probability that the optimizer will converge to the given point if started from a random solution. This measure depends on the stopping criterion of the optimizer since the set of endpoints depends on this parameter.

A heuristic was also suggested for calculating this coefficient. The time complexity of the method is $O(|S|^2)$ so only relatively small spaces can be considered. This method involves the calculation of the transition probabilities between solutions w.r.t. an operator. These transition probabilities were used also for approximating the expected number of evaluations needed to reach the global optimum and plots were presented showing these values for several spaces.

In general if the number of relatively good local optima is large then this measure seems to be reliable but if this is not true then the behavior of the coefficient as a function of the stopping criterion still carries a lot of information about the function at hand. A possible source of problems can be if the deviation of the fitnesses of the endpoints is too small since the coefficient is relative to the minimal and maximal fitness of the set of endpoints. This problem can be solved if the lowest acceptable fitness is given as a parameter.

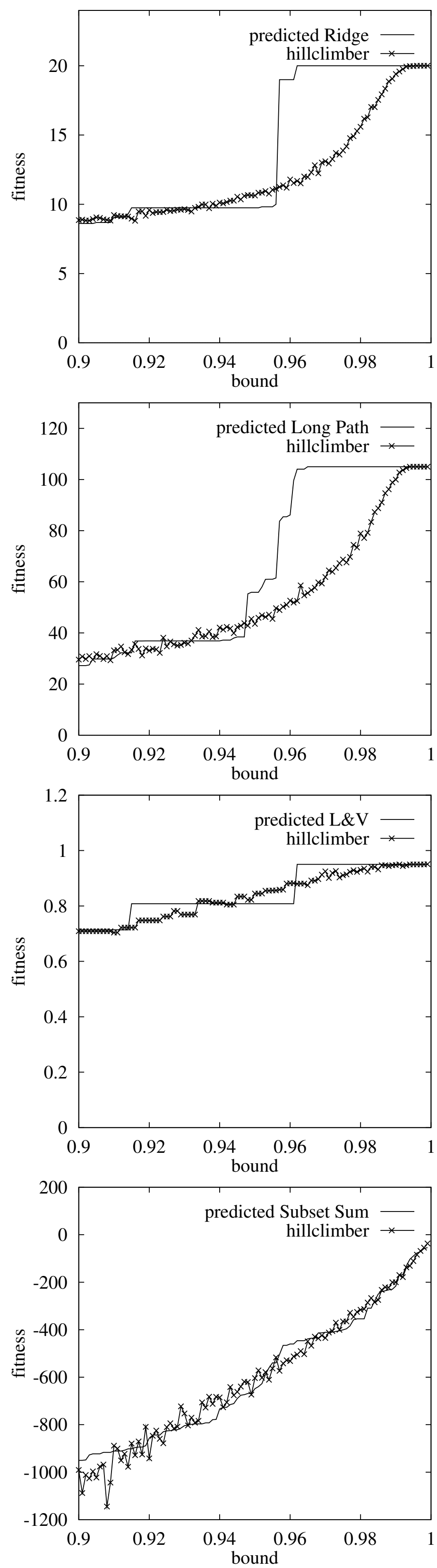

Figure 9: The prediction versus the results of the hill-climber

Bibliography

[Dav91] Yuval Davidor. Epistatsis variance: A viewpoint on GA-hardness. In Rawlins [Raw91], pages 23–35.

[EBSS98] Agoston E. Eiben, Thomas Bäck, Marc Schoenauer, and Hans-Paul Schwefel, editors. *Parallel Problem Solving from Nature - PPSN V*, volume 1498 of *Lecture Notes in Computational Science*. Springer-Verlag, 1998.

[HGD94] J. Horn, D. E. Goldberg, and K. Deb. Long path problems. In Yuval Davidor, Hans-Paul Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature 3*, volume 866 of *Lecture Notes in Computational Science*, pages 149–158. Springer-Verlag, 1994.

[Jel97] Márk Jelasity. A wave analysis of the subset sum problem. In Thomas Bäck, editor, *Proceedings of the Seventh International Conference on Genetic Algorithms*, pages 89–96, San Francisco, California, 1997. Morgan Kaufmann.

[JF95] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In L. J. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.

[Kal98] Leila Kallel. Inside GA dynamics: Ground basis for comparison. In Eiben et al. [EBSS98], pages 57–66.

[KBH93] S. Khuri, T. Bäck, and J. Heitkötter. An evolutionary approach to combinatorial optimization problems. In *The Proceedings of CSC'94*, 1993.

[LV91] G. Liepins and Michael D. Vose. Deceptiveness and genetic algorithm dynamics. In Rawlins [Raw91], pages 36–50.

[NK98] Bart Naudts and Leila Kallel. Comparison of summary statistics of fitness landscapes. Submitted for publication, 1998.

[QRSS98] R. J. Quick, V. J. Rayward-Schmith, and G. D. Smith. Fitness distance correlation and ridge functions. In Eiben et al. [EBSS98], pages 77–86.

[Raw91] G. J. E. Rawlins, editor. *Foundations of Genetic Algorithms*. Morgan Kaufmann, 1991.

[Vos91] Michael D. Vose. Generalizing the notion of schemata in genetic algorithms. *Artificial Intelligence*, 50(3):385–396, 1991.