Generic Hardness Estimation using Fitness and Parameter Landscapes applied to Robust Taboo Search and the Quadratic Assignment Problem

Erik Pitzer, Andreas Beham, and Michael Affenzeller
University of Applied Sciences Upper Austria
School of Informatics, Communication and Media
Softwarepark 11, 4232 Hagenberg, AUSTRIA
{erik.pitzer, andreas.beham, michael.affenzeller}@fh-hagenberg.at

ABSTRACT

Fitness landscape analysis methods have become an increasingly popular topic for research. The future application of these methods to metaheuristics can yield advanced self-adaptive metaheuristics and knowledge bases that can take the role of expert systems in the field of optimization. One important feature of such an expert system would be the prediction of algorithm effort on a certain instance. Estimating whether a certain algorithm is able to tackle the problem adequately or not is a valuable piece of information that currently only an experienced human expert can give. The ability to generate such an advice automatically is, therefore, an important milestone. While fitness landscape analysis methods have been developed for exactly this purpose, it has been shown in the past that single-value analyses have limited applicability. Here, a general method for extracting fitness landscape features will be shown in combination with regression models that indicate a strong correlation between the actual and the predicted effort. Significant potential to increase the prediction quality arises when combining several measures each derived from several different sampling trajectories.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

General Terms

Experimentation, Measurement

Keywords

parameter landscapes, fitness landscapes, quadratic assignment problem, HeuristicLab, sampling trajectories, neutrality, tabu search, Robust Taboo Search

1. INTRODUCTION

In the past many attempts have been made to describe and predict problem instance hardness. There are two basic possibilities, one is to simply try and repeatedly execute various optimization methods and measure the effort necessary to solve a particular problem instance. While this reveals relevant information about the problem instance, it provides little practical value as the process is time consuming. Therefore, methods under the umbrella of fitness landscape analysis have been developed to create a more practical approach. Again, there are several approaches that have been employed in the past. One approach is to exhaustively enumerate the solution space to identify local optima, basins of attraction, barriers and their connecting structures[25, 20]; While this gives good descriptions and insight into fitness landscapes and the inherent structure and provides the basis for fundamental analysis of whole problem classes, it is even more impractical than repeated optimization due to its enormous calculation efforts. Furthermore, problem-specific properties of problem instances can be derived to describe hardness, such as flow dominance[19] for the quadratic assignment problem or depot eccentricity for the vehicle routing problem[21]. On the other hand, stochastic fitness landscape analysis methods have been developed that enable the rapid measurement of certain characteristic features of problem instances. Most of these measures were intended to directly measure problem hardness.

We are skeptical that any single fitness landscape measure can effectively capture problem hardness on its own. In combination, however, using different "perspectives" of a fitness landscape, these measures can jointly provide useful insights. In this particular paper we have improved upon the modest correlation between single fitness landscape analysis values and were able to build good predictors of problem hardness using simple combinations of these different perspectives.

In the past, there have already been several attempts to create problem hardness measures. In essence, most papers about fitness landscape analysis have the ultimate goal of problem hardness prediction. One comprehensive prediction of problem hardness for the Quadratic Assignment Problem (QAP) has been created in [19], where, in addition to pure fitness landscape analysis methods, some problem-specific measures are used. Moreover, some of their analysis measures require expensive measurements such as knowledge of the global optimum for the Fitness Distance Correlation Coefficient[15] or a list of all local optima. Another interesting analysis is performed in [5], where a theoretical decomposition of QAP fitness landscapes is described that yields a method to directly and efficiently calculate the auto correlation coefficient without stochastic sampling. In [23] another summary and analysis of combinatorial optimization problems with the help of fitness landscape analysis is given. There, the focus lies on the algorithm selection problem which is linked to hardness prediction as different hardnesses of instances obtained with different algorithms can be used as a selection criterion.

In this work we concentrate on practicability and speed of the analysis methods that can initially be used before applying any algorithm to quickly get an estimate of problem hardness. With these measures—or more exactly a combination thereof—we are building a simple regression model of problem hardness for instances of the quadratic assignment problem library.

This article is structured as follows: In Section 2 we give an introduction of the quadratic assignment problem (QAP), followed by an overview of the public quadratic assignment problem library (QAPLIB) in Section 2.1. Section 3 describes a specialized variant of the tabu search algorithm, namely the Robust Taboo Search that has been conceived to efficiently solve QAPs. Section 4 describes how parameter landscapes were explored and visualized as the basis for effort estimation. In Section 5 we give a short introduction on fitness landscape analysis methods, sampling strategies and measures as well as an overview of the experimental setup to derive these results from the QAPLIB. Finally, Section 6 details how hardness can be measured and predicted in general and reports results on the application to the QAPLIB.

2. QUADRATIC ASSIGNMENT PROBLEM

The Quadratic Assignment Problem (QAP) was introduced in [17] and is a well-known problem in the field of operations research. It is the topic of many studies, treating the improvement of optimization methods as well as reporting successful application to practical problems in keyboard design, facility layout planning and re-planning as well as in circuit design[8, 11, 6]. The problem is NP hard in general and, thus, the best solution cannot easily be computed in polynomial time. Many different optimization methods have been tried, among them popular metaheuristics such as tabu search [26, 13] and genetic algorithms[7].

The problem can be described as finding the best assignment for a set of facilities to a set of locations so that each facility is assigned to exactly one location which in turn houses only this facility. An assignment is considered better than another when the flows between the assigned facilities have to be hauled over smaller distances.

The QAP is also a generalization of the traveling salesman problem (TSP). Conversion of a TSP can be achieved by using a special weight matrix[17, 19] where the flow between the "facilities" is modeled as a ring that involves all of them exactly once. The flow in this case can be interpreted as the salesman that travels from one city to another.

More formally the problem can be described by an $N \times N$ matrix W with elements w_{ik} denoting the weights between facilities i and k and an $N \times N$ matrix D with elements d_{xy} denoting the distances between locations x and y. The goal is to find a permutation π with $\pi(i)$ denoting the element at position i so that the following objective is achieved:

$$\min \sum_{i=1}^{N} \sum_{k=1}^{N} w_{ik} \cdot d_{\pi(i)\pi(k)} \tag{1}$$

A permutation is restricted to contain every number just once, hence, it satisfies the constraint of a one-to-one assignment between facilities and locations:

$$\forall_{i,k} i \neq k \Leftrightarrow \pi(i) \neq \pi(k) \tag{2}$$

The complexity of evaluating the quality of an assignment according to Eq. (1) is $O(N^2)$, however several optimization algorithms move from one solution to another through small changes, such as by swapping two elements in the permutation. These moves allow to reduce the evaluation complexity to $O(N)$ and even $O(1)$ if the previous qualities are memorized[26]. Despite changing the solution in small steps iteratively, these algorithms can, nevertheless, explore the solution space and interesting parts thereof quickly. The complete enumeration of such a "swap" neighborhood contains $N*(N-1)/2$ moves and, therefore, grows quickly with the problem size. This poses a challenge for solving larger instances of the QAP.

The QAP can also be used to model cases when there are more locations than facilities and also when there are more facilities than locations. In these cases dummy facilities with zero flows or dummy locations with a high distance can be defined.

2.1 QAPLIB

The quadratic assignment problem library (QAPLIB)[4] is a collection of benchmark instances from different contributors. According to their website[1], it originated at the Graz University of Technology and is now maintained by the University of Pennsylvania, School of Engineering and Applied Science. It includes the instance descriptions in a common format, as well as optimal and best-known solutions or lower bounds and consists of a total of 137 instances from 15 contributing sources which cover real-world as well as random instances. The sizes range from 10 to 256 although smaller instances are more frequent. All 103 instances between 12 and 50 have been selected for this study with the exception of `esc16f`, which does not have any flows.

3. ROBUST TABOO SEARCH

Tabu search (TS) is a general metaheuristic that was proposed by Glover in [10]. It behaves like a local search, except that it will always make a move in each iteration, even if the current fitness deteriorates. The name, however, results from the use of a memory that forbids to make certain moves and forces the search trajectory to explore new parts of the fitness landscape.

The Robust Taboo Search (RTS) algorithm was proposed by Taillard in [26] and the code has since been further developed[2]. The main difference between the tabu search described by Glover lies in the stochastic choice of tabu tenures. For every move, the time to keep it tabu is a random variable. In an earlier version of the algorithm, the tabu tenure varies between a minimum and a maximum value but in a later version the minimum parameter is abandoned and

[1] `http://www.seas.upenn.edu/qaplib/`
[2] `http://mistic.heig-vd.ch/taillard/`

the random variable is drawn from a left-skewed distribution instead. This random tabu tenure lowers the possibility of search cycles returning to the same solution over and over.

Another important aspect of the RTS is its aspiration tenure. When applying the tabu search to the QAP it was discovered that in some cases the search process got stuck in suboptimal regions of the fitness landscape. For this reason, a parameter was introduced that would diversify the search after a number of iterations. The diversification aims to perform moves which have not been seen in the past few iterations. There are, thus, two important parameters in the RTS that have to be configured for the search to perform well on a given problem instance: These are the *maximum tabu tenure* and the *aspiration tenure*.

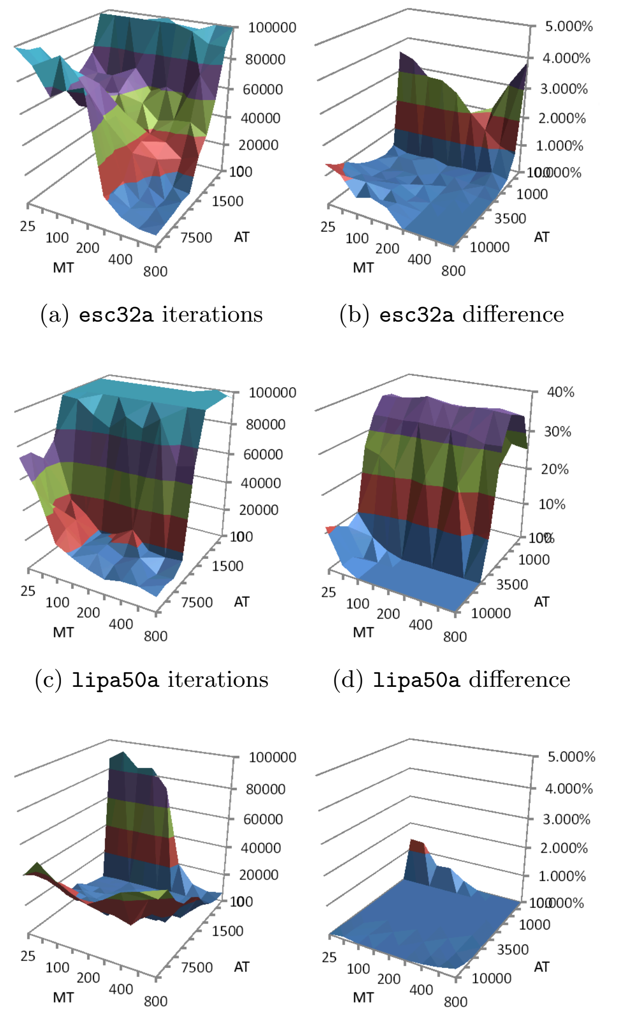
4. PARAMETER GRID AND EFFORT VISUALIZATION

The performance of metaheuristics usually depends to a large degree on the choice of their parameters. Good parameters result in fast convergence, high quality solutions, or ideally both. Bad parameters can significantly delay or prevent convergence and are usually undesirable. However, there is no best a-priori parameterization of metaheuristics and, thus, it takes time and computational resources to find the right settings. A task that has to be performed for every new instance. Additionally, some of the instances might be rather difficult to parameterize because of a small set of parameters that work well in comparison to the overall parameter space while in other instances the parameters do not seem to affect the performance and equal results are obtained for any of them.

In order to have an estimation on the performance of the RTS for several instances of the QAPLIB, tests were performed that sample the parameter space at certain intervals and visualize the performance for all of these samples. The results indicate, for each instance, the optimal setting with respect to the chosen grid and the expected performance. A total of nine different tabu tenures (25, 50, 100, 150, 200, 300, 400, 600, and 800) and ten different aspiration tenures (100, 500, 750, 1000, 1500, 2000, 3500, 5000, 7500, and 10000) have been chosen, resulting in 90 different configurations. Each of these have been evaluated 20 times on every instance to account for stochastic variability, resulting in a total of 183,600 runs that have been conducted. The tests have been made with HeuristicLab[29, 30] and the resulting experiment can be downloaded[3] and studied.

In these tests, the number of iterations until convergence to the optimal or best-known solution were recorded as well as the quality deviation if this solution could not be found within 100,000 iterations. The visualizations were created using pivot charts in Microsoft Excel and are shown in Figure 1. As can be seen, the quality difference and the iterations show a similar picture of the algorithm's performance. The iterations allow to judge convergence speed, while the quality difference allows to judge the achievable solution quality when the best-known solution could not be found.

[3] http://dev.heuristiclab.com/AdditionalMaterial



(a) esc32a iterations (b) esc32a difference

(c) lipa50a iterations (d) lipa50a difference

(e) els19 iterations (f) els19 difference

Figure 1: Parameter landscapes showing the performance of various configurations and instances: Figures (a), (c), and (e) show the average number of iterations till convergence to the best known solution, while Figures (b), (d), and (f) show the average scaled quality difference in percent. The axes are labeled MT for the setting of *maximum tabu tenure* and AT for different values of alternative aspiration tenure.

5. FITNESS LANDSCAPE ANALYSIS

Since the inception of fitness landscapes in Sewall Wright's description of populations' biological evolution in [32] it has also been applied to evolution-inspired algorithms and optimization methods in general. While apparently an intuitive concept, it is important to give a formal description as well. Mathematically, a fitness landscape $\mathcal{F}$ is defined by the solution space S, a fitness function $f : S \to \mathbb{R}$ and a notion of connectivity $\mathcal{X}$. In many cases, connectivity can be defined in terms of neighboring solution candidates through a neighborhood function $N : S \to \mathcal{P}(S)$ where $\mathcal{P}(S)$ is the power set of S. Alternatively, and more conveniently, we can also use a distance measure $d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ between two solution candidates as our connectivity. While the first form is more practical when thinking about mutation operators of evolutionary algorithms, the second form is more general and better comparable across different algorithm types and manipulation operators. This is also the reason, why connectivity is often omitted from the definition of a fitness landscape and a distance metric is implicitly assumed.

Based on the formal definition, several possibilities of fitness landscape analysis (FLA) of a problem instance have been devised. While several theoretical approaches have been proposed[24, 3, 5], we focus on automatic sampling-based fitness landscape analysis techniques. Therefore, the first step in analyzing a problem instance is deriving interesting samples. While a random sample would obviously give a good overview of the fitness distribution in general, it has little to say about the neighborhood structure. In fact, we use random samples for establishing a base-line or average fitness which can then be used to compare fitness improvement over different problem instances with different extents of fitness values. While relative difference between the obtained solution's fitness $f(s)$ and the best known solution f^*, given as $(f(s) - f^*)/f^*$, can be used to compare different algorithms on the same problem instance, a different range of fitness values makes this infeasible for comparing different problem instances. Therefore, we use the average fitness of a random sample $\hat{f}$ to equalize this ratio and obtain a *scaled difference* as $(f(s) - f^*)/(\hat{f} - f^*)$.

5.1 Trajectories

In addition to random samples, which have been used for calculating the scaled difference, we are using series of adjacent solution candidates, i.e. trajectories inside the fitness landscape, to explore the neighborhood structures in more detail. In the following sections, different exploration strategies are explained that greatly increase the insight into fitness landscapes due to their complementarity to random walks. These different walk types should be seem complementary instead of competing in later analysis, as different perspectives of the fitness landscape are important to obtain a complete view.

5.1.1 Random Walks

Probably the most popular way of trajectory generation in the field of fitness landscape analysis is the random walk. Here, based on a randomly-selected initial solution candidate a neighbor is chosen at random. In discrete spaces such as the permutation encoding used most often for the quadratic assignment problem any of the popular mutation operators can be used to obtain such a random neighbor. In our study, we have used only the SWAP-2 mutator as it is also the only operator used for the Robust Taboo Search. However, it can be insightful to include other mutation operators and, hence, other mutation landscapes when performing a comprehensive study of problem instances.

5.1.2 Adaptive and Up-Down Walks

Another interesting trajectory is similar to what a trajectory based optimization algorithm would see: A trajectory that seeks to improve solution quality. In an adaptive walk several neighboring solutions are examined first and the most promising is chosen in the next step. This simple optimization scheme, tends to quickly become trapped in local optima. Therefore, we have also implemented up-down walks[14] which consecutively seek positive and negative extrema instead of single-minded optimization only. Since we are not trying to find a good solution when doing fitness landscape analysis, but instead try to collect as much information about the landscape as possible this presents a good compromise between perceiving optimization-relevant properties of the landscape and the exploratory attitude required to gain maximum insight.

5.1.3 Neutral Walks

A fascinating property of fitness landscapes is neutrality[22, 16, 1]. Neutral areas are connected neighborhood structures that have the same fitness. While low-dimensional cases can be problematic for simple algorithms that get stuck because of a missing direction to continue, in higher-dimensional instance they can be beneficial. An intelligent population-based algorithm that can deal with neutrality can spread across a neutral area without fitness penalty and increase population diversity. Once the "edges" are reached more neighboring solutions are accessible than from a single local optimum. To explore this property, we perform special neutral walks that measure the extents of these areas by remembering the staring point and subsequently seeking new neighbors with equal fitness that increase the distance to the starting point. Then, the next neutral area is searched by moving away from the last point on the last neutral area. This trajectory repeatedly searches for and crosses neutral areas more or less diametrically.

5.2 Measures

So far, most FLA measures have been proposed only for random walks. Most of these can easily be extended to other trajectories as well. In the following, we will give a short overview of the employed fitness landscape analysis measures.

5.2.1 Ruggedness

One of the first general measures to describe fitness landscapes, introduced in [31], was auto correlation. It describes the average correlation between consecutive fitness values in e.g. a random walk. Based on these measures, the correlation length can be defined as the first step that is not significantly correlated with the initial step in the autocorrelation function[12]. This definition differs slightly from the more ubiquitous standard definition where the correlation length is typically assumed to be the reciprocal value of the auto correlation of the first step[14]. The definition of the auto correlation function $\rho(\varepsilon)$ and the variance-corrected auto correlation coefficient ρ are defined as follows:

$$\rho(\varepsilon) := \frac{\mathrm{E}(f_i \cdot f_{i+\varepsilon}) - \mathrm{E}(f_i) \cdot \mathrm{E}(f_{i+\varepsilon})}{\mathrm{Var}(f_i)} \tag{3}$$

$$\rho := \frac{\mathrm{Cov}(f_i, f_{i+\varepsilon})}{\sqrt{\mathrm{Var}(f_i) \cdot \mathrm{Var}(f_{i+\varepsilon})}} \tag{4}$$

where f_i is the random variable describing the fitness trajectories, $f_{i+\varepsilon}$ is the series of fitness values shifted by ε, $\mathrm{E}(x)$ is the expected value of x and $\mathrm{Var}(x)$ is the variance of x.

5.2.2 Information Analysis

Related to ruggedness, by analysis of consecutive steps of a trajectory, is the information analysis[28], which is based on the discretization of fitness changes and a subsequent slope analysis. Given a sequence of fitness values $\{f_i\}_{i=0}^{n}$, initially the sequence is transformed into a sequence of fitness differences $\{d_i\}_{i=1}^{n} = \{f_i - f_{i-1}\}_{i=1}^{n}$ which is then discretized by a relaxed sign function which gives -1 or 1 when x exceed $-\varepsilon$ or ε and 0 otherwise, yielding a sequence of slopes $\{s_i\}_{i=1}^{n}$.

Two simple measures can directly be derived from this sequence. One is the *partial information content* which gives the relative number of slope changes. This simply reduces the sequence of slopes s by removing all zeros and all consecutive equal slopes giving s'. The resulting quotient is denoted as $M(\varepsilon) = |s'|/n$, where n is the length of the trajectory. Another simple measure is the smallest value of ε for which no slopes remain after applying the relaxed sign function, i.e. the maximum fitness difference between consecutive steps in the trajectory. This value, the *information stability*, is denoted as ε^*. In addition, the following slightly more complex entropy measures have been defined for determining the *information content* $H(\varepsilon)$ and *density basin information* $h(\varepsilon)$:

$$H(\varepsilon) := -\sum_{p \neq q} P_{[pq]} \log_6 P_{[pq]} \tag{5}$$

$$h(\varepsilon) := -\sum_{p = q} P_{[pq]} \log_3 P_{[pq]} \tag{6}$$

These information measures are based on the idea of "an ensemble of objects, which are characterized by their size, form, and distribution"[28]. In other words, each solution candidate together with its surroundings is taken as an individual and forms either a flat area, a local optimum or neither. Finally, the information measures give an estimate of the difficulty to reconstruct the configuration of these individual objects. One addition in comparison to previous ruggedness analysis was the "zoom level" ε, which allows considering the landscape structures at various extents.

An additional, seemingly trivial, measure that carries a good deal of information is the so-called *regularity*. It is simply the number of distinct fitness values that were seen during the landscape exploration. Typically, this measure is similar for different but large trajectories or other samples of the landscape but can be an important indicator. A higher number gives finer grained comparability of solution candidates and, hence, can improve homing in towards the optimum.

5.2.3 Intermediate Walk Lengths and Distances

In addition to the measures described in the previous sections, the different trajectories provide some even more basic but informative characteristics. While the total number of steps for any of random, up-down, and neutral walks is usually predetermined, the intermediate sub-walk lengths are interesting. In the up-down walks, we can analyze the number of steps it takes to reach the next local maximum followed by the number of steps to the next local minimum. In addition we can analyze the upper and lower fitness levels that are usually achieved. Similarly, for the neutral walk it is revealing to examine the number of steps inside a neutral area. Moreover, we can also look at the difference in terms of solution representation distance[2] between entry and exit points in the neutral area. At a first glance this might appear to be the same as the number of steps, however, firstly, different mutation operators have different magnitudes of change in the solution representation, and secondly, the concrete steps in the neutral area might lurch about in several dimensions while actually staying close to the starting point. For these additional intermediate walks we include both average length and variance and for neutral areas also the average distance between start end end points of the neutral walks and the corresponding variance.

5.3 Experiments

One important aspect of our analysis was the execution time. Therefore, we have only used fitness landscape analysis methods that give good insights in relatively short time. For this reason, we exclude all methods that require exhaustive analysis, knowledge of global or local optima. Instead, we rely solely on the application of stochastic, trajectory-based fitness landscape measures. We have used three trajectories, a random walk, a neutral walk, and an up-down walk and collected several measures for each of the produced fitness trajectories. We have conducted the analysis using four different mutation operators and five different analysis algorithms. For this evaluation, however, only three different algorithms and one mutation operator were used. In Table 1, the number of steps per walk type as well as the considered neighbors and the resulting run-times are shown. In summary, the presented fitness landscape results have been obtained quickly, especially in comparison to the optimization results for different parameter configurations, described in Section 4. While the fitness landscape analysis took only seven minutes, the effort prediction took 11 hours when calculated in parallel on 14 Blade Systems, each with two Intel Xeon E5420 quad core CPUs with 2.50 GHz.

In addition to the measured results, we have added the auto correlation coefficients which are calculated as $c = 1/(1-x)$ and the normalized auto correlation coefficient and normalized correlation length values which are simply the actual values divided by the respective problem sizes.

Table 1: Trajectory configurations and average run-times per problem instance

trajectory	steps	neighbors	average run-time
random	100,000	1	59 secs
up-down	100,000	10	190 secs
neutral	10,000	100	140 secs

6. HARDNESS ESTIMATION

Classically, problem hardness is defined as the difficulty to find optimal or good results for a certain problem instance. This is typically measured as the necessary effort to reach a certain quality, or the quality achieved after making a certain effort[33, 23, 18]. However, the parameterization of metaheuristics, such as Robust Taboo Search, can have a significant impact on its performance and success. Therefore, to estimate a more general problem difficulty we are measuring the effort on a whole parameter grid as shown in Figures 1a through 1f. As described in Section 4 we have tested 90 configurations with 20 repetitions for each problem instance to get an insight into the algorithm's performance on different instances. Additionally, for each of these instances, we have calculated 44 different fitness landscape analysis values. We have limited the iterations in the RTS to 100,000 steps and stopped execution once the best-known or global best solution had been found.

6.1 Hardness Measurement

With the experiments of different parameter settings at hand, there are several possibilities for deriving a hardness value. There are two basic measures that we can use to derive hardness. One is the achieved quality after the maximum number of iterations and the other is the expended effort in case the global optimum or best known solution was found earlier. Each of these measures describes the performance of a certain algorithm run. In addition we can create a combined measure of difficulty. In summary, we can create a single performance measure that spans the whole range between early successful solution over late successful solution, non-successful solution close to the optimum to non-successful solution far from the optimum. Given the maximum number of iterations and an assumed worst case scaled difference of 1 (which is the average scaled difference in a random sample) the performance for an individual run can be expressed as $p = i/i_{\max} + s$ where i is the number of iterations and s the scaled difference.

We have performed 20 repetitions of each run to account for stochasticity. For these repetitions we can directly use the averages of iterations and scaled differences. Alternatively we can use the average combined performance as described above. In this case, however, we should emphasize that the scaled difference plays a more important role than the number of iterations which is reflected by a different scaling that puts more emphasis on the solution quality than convergence speed. Experiments with unscaled addition have shown that quality can otherwise easily be overshadowed by convergence speed when taking averages over the combined measure.

Many metaheuristic algorithms, in contrast to exact methods, have several parameters that can have a large impact on the performance of the algorithms. For this reason, we have tried 90 different parameter configurations of the Robust Taboo Search as described earlier. Now, these different experiments can be combined further to derive a notion of hardness pertaining to a particular problem. The most straightforward measurement of problem hardness we have used is the average performance over all studied parameter configurations. In addition, we have also included the performance of the best parameter configuration as the minimum hardness of a problem. The difference between the average performance over all configurations and the best possible configuration can be seen as another related measure of hardness, albeit not problem hardness, but specificity of parameter selection or the difficulty of finding the right parameter combination for a problem instance.

6.2 Simple Correlations

An initial analysis revealed interesting correlations between fitness landscape measures and average hardness, which are summarized in Table 2. Only few typical FLA results have a correlation with average problem hardness, despite their aspiration to do just that. Correlations with average iterations are comparable (not shown) while other hardness measures such as effort of the best configuration or average scaled difference have hardly any correlations with any of the FLA results.

In the scatter plot of average hardness and problem size in Figure 2 we can see that there seems to be a simple correlation. Interestingly, there also seem to be two classes of problems with different correlation coefficients.

Table 2: Simple Correlations with Average Hardness

Variable	r^2
U/D Auto Correlation Coefficient	0.59
Rnd Auto Correlation Coefficient	0.59
Problem Size	0.56
U/D Down Walk Len	0.44
U/D Correlation Len	0.41
U/D Up Walk Len	0.41
U/D Inf Content	0.36

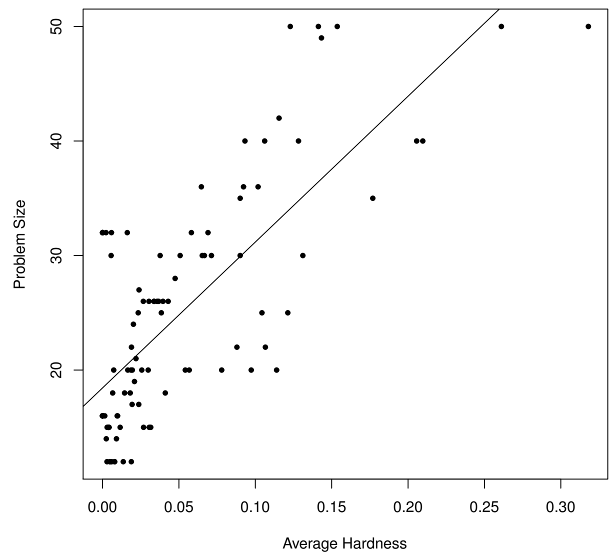


Figure 2: Average Hardness vs. Problem Size

6.3 Regression

To show the improved usability of multiple "perspectives" of problem instances' fitness landscapes we have built several regression models to describe the problem hardness with easily obtainable fitness landscape measures. Table 3 shows the results of linear regression of average hardness. We have used the Lasso method[27, 9] to efficiently select variables from an initial set of 44 measures and created eight different regression models using different trajectories only or the whole variable set. For each of these configurations we selected two models, one with the smallest error, and another model with the least number of variables denoted as either *best* or *minimal* model. The obtained variable selections were then input to HeuristicLab and a tenfold cross validation was performed where the reported values are predictions using the respective test sets. We have built similar models for all other hardness measures, such as minimal hardness (results of most successful parameter setting), individually for just average iterations and average scaled difference and specificity. However, the strongest correlation was observed for the, intuitively most interesting, combined effort with scaled influences, paying more attention to scaled difference than iterations.

Most interestingly, the typically ubiquitous random walk trajectories create the models with least correlation. We

Table 3: Cross validated regression models for average hardness: A "•" denotes inclusion of the value, while "○" denotes the inclusion of the value divided by the problem size, and "∼" denotes inclusion of the variable's variance along the trajectory. The symbols on the left of each column shown the inclusion in the best model while the symbols on the right denote the inclusion in the minimal model.

Measure	Rnd best	Rnd min	U/D best	U/D min	Neut best	Neut min	Full best	Full min
Problem Size	•	•	•	•	•	•	•	•
Rnd Auto Correlation	○						○	○
Rnd Correlation Len	○							
Rnd Density Basin Inf		•					•	•
Rnd Part Inf Content	•							
U/D Auto Correlation			•	•			○	
U/D Correlation Len			○				•	
U/D Density Basin Inf			•				•	
U/D Regularity			•				•	
U/D Up Walk Len			•				∼	
U/D Down Walk Len			∼	∼			∼	
U/D Lower Level							∼	
Neut Auto Correlation					•	○		○
Neut Correlation Len					•	○	•	•
Neut Density Basin Inf					•			
Neut Inf Stability					•			
Neut Regularity					•		•	
Neut Avg Walk Dist					•	•	•	•
Neut Avg Walk Len					∼	∼	∼	∼
Best Train r^2	0.68		0.76		0.77		0.84	
Best Test r^2	**0.60**		**0.67**		**0.70**		**0.74**	
Min Train r^2	0.65		0.72		0.75		0.77	
Min Test r^2	**0.59**		**0.67**		**0.70**		**0.71**	

can see in Table 3 that both the up-down walk as well as the neutral walk generate measures that are superior. One could argue that this is due to the larger number of variables that were used in these configurations, however, even using only the corresponding measures selected for the random walk in e.g. a neutral walk model creates slightly superior regressions. This emphasizes the importance of these alternative trajectories that are often ignored. Another interesting point is the near equal performance of the minimal models in the trajectory specific cases. Even though, they are slightly worse on the training sets, the performance on the test sets is almost on par with the bigger models.

Figure 3 shows a scatter plot of the predicted values in the full model using only the test sets of the respective folds plotted against the target values. We can see acceptable correlation between these two sets. Some interesting cases can be seen for easy instances, where the predicted values vary wildly. Even more interestingly, these are instances which are easy for the Robust Taboo Search while they are difficult for many other algorithms. We postulate that FLA methods are able to measure this difficulty, which coincidentally does not pose a problem for the Robust Taboo Search.

7. CONCLUSIONS

We have shown how popular fitness landscape analysis measures can be used effectively by combining them not only with each other but also by measuring them along different trajectories. These measures have been used to fulfill their original intention of predicting problem hardness. While no perfect prediction was achieved, given the

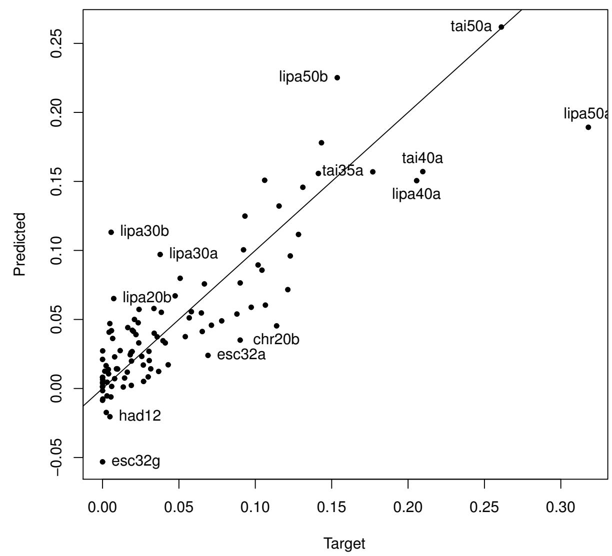


Figure 3: Scatter plot of average actual hardness vs. predicted hardness using several trajectories' measures.

simplicity of the method and the promising results we are confident that extensions of this method can perform even better. In this work we have emphasized the importance of different *perspectives* of a fitness landscape, obtained by performing several different analyses, combining different measures and different sampling strategies, in our case different types of walks. We have also observed that the most prevalent walk type, the random walk, might not be the best choice to obtain insight into a particular fitness landscape, but that more advanced sampling strategies such as up-down or neutral walks can be advantageous. In the future, we plan to extend this approach using more measures and more sampling strategies to obtain a more complete picture through even more "perspectives" of a problem instance's fitness landscape.

8. ACKNOWLEDGMENTS

This work was in part supported by the Austrian Research Promotion Agency (FFG) within the Josef Ressel Center "Heureka!".

9. REFERENCES

[1] L. Barnett. Ruggedness and neutrality—the NKp family of fitness landscapes. In *ALIFE: Proceedings of the sixth international conference on Artificial life*, pages 18–27, Cambridge, MA, USA, 1998. MIT Press.

[2] A. Beham, E. Pitzer, and M. Affenzeller. A new metric to measure distances between solutions to the quadratic assignment problem. In *Proceedings of the IEEE 3rd International Symposium on Logistics and Industrial Informatics (LINDI 2011)*, 2011.

[3] Y. Borenstein and R. Poli. Decomposition of fitness functions in random heuristic search. In *FOGA'07: Proceedings of the 9th International Conference on*

Foundations of Genetic Algorithms, pages 123–137, Berlin, Heidelberg, 2007. Springer-Verlag.

[4] R. E. Burkard, S. E. Karisch, and F. Rendl. QAPLIB - A quadratic assignment problem library. *Journal of Global Optimization*, 10(4):391–403, June 1997.

[5] F. Chicano, G. Luque, and E. Alba. Elementary landscape decomposition of the quadratic assignment problem. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, GECCO'10, pages 1425–1432, New York, NY, USA, 2010.

[6] S. A. de Carvalho Jr. and S. Rahmann. Microarray layout as quadratic assignment problem. In *Proceedings of the German Conference on Bioinformatics (GCB), volume P-83 of Lecture Notes in Informatics*, 2006.

[7] Z. Drezner. Extensive experiments with hybrid genetic algorithms for the solution of the quadratic assignment problem. *Computers & Operations Research*, 35(3):717–736, 2008. Part Special Issue: New Trends in Locational Analysis.

[8] A. N. Elshafei. Hospital layout as a quadratic assignment problem. *Operational Research Quarterly*, 28(1):167–179, 1977.

[9] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

[10] F. Glover. Tabu search – part I. *ORSA Journal on Computing*, 1(3):190–206, 1989.

[11] P. M. Hahn and J. Krarup. A hospital facility layout problem finally solved. *Journal of Intelligent Manufacturing*, 12:487–496, 2001.

[12] W. Hordijk. A measure of landscapes. *Evol. Comput.*, 4(4):335–360, 1996.

[13] T. James, C. Rego, and F. Glover. Multistart tabu search and diversification strategies for the quadratic assignment problem. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(3):579–596, may 2009.

[14] T. Jones. *Evolutionary Algorithms, Fitness Landscapes and Search*. PhD thesis, University of New Mexico, Albuquerque, New Mexico, 1995.

[15] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.

[16] Y. Katada. Estimating the degree of neutrality in fitness landscapes by the Nei's standard genetic distance – an application to evolutionary robotics. In *2006 IEEE Congress on Evolutionary Computation*, pages 483–490, 2006.

[17] T. C. Koopmans and M. Beckmann. Assignment problems and the location of economic activities. *Econometrica, Journal of the Econometric Society*, 25(1):53–76, January 1957.

[18] W. G. Macready and D. H. Wolpert. What makes an optimization problem hard? *Complexity*, 5:40–46, 1996.

[19] P. Merz and B. Freisleben. Fitness landscape analysis and memetic algorithms for the quadratic assignment problem. *Evolutionary Computation, IEEE Transactions on*, 4(4):337–352, nov 2000.

[20] E. Pitzer and M. Affenzeller. *Recent Advances in Intelligent Engineering Systems*, chapter A Comprehensive Survey on Fitness Landscape Analysis, pages 167–196. Springer, 2011.

[21] E. Pitzer, S. Vonolfen, , A. Beham, M. Affenzeller, V. Bolshakov, and G. Merkurieva. Structural analysis of vehicle routing problems using general fitness landscape analysis and problem specific measures. In *Proceedings of the 14th International Asia Pacific Conference on Computer Aided System Theory*, 2012. accepted.

[22] C. M. Reidys and P. F. Stadler. Neutrality in fitness landscapes. *Applied Mathematics and Computation*, 117(2-3):321–350, 1998.

[23] K. Smith-Miles and L. Lopes. Review: Measuring instance difficulty for combinatorial optimization problems. *Comput. Oper. Res.*, 39(5):875–889, May 2012.

[24] P. Stadler and G. Wagner. The algebraic theory of recombination spaces. *Evol.Comp.*, 5:241–275, 1998.

[25] P. F. Stadler. *Biological Evolution and Statistical Physics*, chapter Fitness Landscapes, pages 187–207. Springer, 2002.

[26] E. D. Taillard. Robust taboo search for the quadratic assignment problem. *Parallel Computing*, 17:443–455, 1991.

[27] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1):267–288, 1996.

[28] V. K. Vassilev, T. C. Fogarty, and J. F. Miller. Information characteristics and the structure of landscapes. *Evol. Comput.*, 8(1):31–60, 2000.

[29] S. Wagner. *Heuristic Optimization Software Systems - Modeling of Heuristic Optimization Algorithms in the HeuristicLab Software Environment*. PhD thesis, Johannes Kepler University, Linz, Austria, 2009.

[30] S. Wagner, A. Beham, G. K. Kronberger, M. Kommenda, E. Pitzer, M. Kofler, S. Vonolfen, S. M. Winkler, V. Dorfer, and M. Affenzeller. Heuristiclab 3.3: A unified approach to metaheuristic optimization. In *Actas del séptimo congreso español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB'2010)*, page 8, September 2010.

[31] E. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63(5):325–336, 1990.

[32] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the Sixth International Congress of genetics*, volume 1, pages 356–366, 1932.

[33] B. Xin, J. Chen, and F. Pan. Problem difficulty analysis for particle swarm optimization: deception and modality. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, GEC '09, pages 623–630, 2009.