Local Optima and Weight Distribution in the Number Partitioning Problem

Khulood Alyahya and Jonathan E. Rowe

School of Computer Science
University of Birmingham, B15 2TT, UK
{kya020,J.E.Rowe}@cs.bham.ac.uk
http://www.cs.bham.ac.uk

Abstract. This paper investigates the relation between the distribution of the weights and the number of local optima in the Number Partitioning Problem (NPP). The number of local optima in the 1-bit flip landscape was found to be strongly and negatively correlated with the coefficient of variation (CV) of the weights. The average local search cost using the 1-bit flip operator was also found to be strongly and negatively correlated with the CV of the weights. A formula based on the CV of the weights for estimating the average number of local optima in the 1-bit flip landscape is proposed in the paper. The paper also shows that the CV of the weights has a potentially useful application in guiding the choice of heuristic search algorithm.

Keywords: Combinatorial optimisation, phase transition, partitioning problem, makespan scheduling, fitness landscape.

1 Introduction

The number partitioning problem (NPP) is a classical problem in theoretical computer science. It is one of Garey and Johnson's six basic NP-complete problems [4]. It has many practical applications such as multiprocessor scheduling. The optimisation version of the problem can be defined as: given a set $A = \{a_1, \ldots, a_n\}$ of positive integers (weights) drawn at random from the set $\{1, 2, .., M\}$, the goal is to partition A into two disjoint subsets S, S' such that the discrepancy between them $|\sum_{a_i \in S} a_i - \sum_{a_i \in S'} a_i|$ is minimised. A partition is called perfect, if the discrepancy between the two subsets is 0 when the sum of the original set is even, or 1 when the sum is odd. Equivalently, the problem can be viewed as minimising the maximum sum over the two subsets:

$$f(x) = \max \left\{ \sum_{a_i \in A} a_i x_i, \sum_{a_i \in A} a_i (1 - x_i) \right\}, x \in \{0, 1\}^n \tag{1}$$

The NPP has an easy-hard phase transition, with many perfect partitions with probability tending to 1 in the easy phase, then the number of perfect partitions drops to zero with probability tending to 1 in the hard phase. The transition is

T. Bartz-Beielstein et al. (Eds.): PPSN XIII 2014, LNCS 8672, pp. 862–871, 2014.

determined by the control parameter $k = \log_2 M/n$, which corresponds to the number of bits required to encode the numbers in the set divided by the size of the set. For $\log_2 M$ and n tending to infinity, the transition occurs at the critical value of $k_c = 1$ [2,6]. The pairs of weights that are placed in the same subset or in opposite subsets in all optimal solutions of an NPP instance, form the backbone of that instance. There is a very sharp increase in the backbone size of the optimal solutions in the NPP as one approaches the phase transition boundary, after which the backbone tends to be complete giving a unique optimal solution [8].

Theoretical analysis of randomized local search shows that it can be a good approximation algorithm for the NPP [12]. Generally, when using local search metaheuristics the average local search cost can vary across problem instances of the same size by many orders of magnitude. A number of models of local search cost for various NP-complete problems such as the Traveling Salesman Problem (TSP), the Boolean Satisfiability Problem (SAT), and the Job-Shop Scheduling Problem (JSP) [7,13,11], have been developed as functions of the problem fitness landscape features. These models attempt to provide explanations for some or ideally all of this variability in the search cost. They also aim to give insights into why one problem instance is more difficult than the other and how the local search is influenced by the properties of the landscape.

The phase transition in the NPP provides an explanation for the increase in the local search cost for problem instances in the hard phase [1], agreeing with the intuition that having fewer optimal solutions usually yield an increase in the local search cost. However, there is considerable variability in the average local search cost for instances in the hard phase that the phase transition fails to account for. The average local search cost was found to vary by many orders of magnitude for hard instances drawn from different distributions such as uniform, normal, negatively and positively skewed distributions [1]. Also, it has been shown in [9] that the number of local optima is not dependent on the easy-hard phase transition, which suggests that the control parameter k is not necessarily good for predicting which problem instances will be easy or hard for local search.

In this paper, we investigate the relation between the distributions of the weights and the fitness landscape features of the NPP. We examine how the number of local optima and the average cost of local search are influenced by the distributions of the weights.

2 Weights Coefficient of Variation and NPP Landscape

Most of the existing analyses of the NPP assume that weights are drawn uniformly at random from a given range. Paper [1] shows that when different distributions are used, there can be large changes in local search performance in hard instances, most noticeable in the 1-bit flip landscape. These changes are mostly due to the difference in the number of local optima between instances drawn from the different distributions. The results shown in [1] suggest that the variability of the weights is what causes the difference in the number of local optima between instances of the NPP. To measure the variability of the weights we use the coefficient of variation (CV) which provides a measure of relative variability or dispersion. CV is defined as the ratio of the standard deviation σ to the mean μ:

$$CV = \frac{\sigma}{\mu} \tag{2}$$

We conducted a series of experiments in order to test the assumption that variability of the weights results in different number of local optima. In the experiments, instances from small problem sizes were considered to allow exhaustive enumeration of the entire search space, the problem sizes considered are $n = 12, 14, 16, 18, 20$. For each problem size, 700 instances from the hard phase ($k = log_2 M/n > 1$) were randomly generated with different values of CV.

The rest of this section gives an analysis of the obtained experimental results, focusing on the number of local optima and the average local search cost and how they correlate with the CV. The following definitions will be used throughout this paper:

Search Space. The search space X is the finite set of all the candidate solutions. Since the fitness function of NPP is a pseudo-Boolean function the search space size is 2^n. The binary representation of NPP creates a symmetry in the search space, in the sense that a solution and its bitwise complement have the same fitness value. Thus, the number of unique solutions is 2^{n-1}.

Neighbourhood. A neighbourhood is a mapping $N : X \to P(X)$, that associates each solution with a set of candidate solutions, called neighbours which can be reached by applying the neighbourhood operator once. The set of neighbours of x is called $N(x)$, and $x \notin N(x)$.

We consider two different neighbourhood operators in this paper: The 1 hamming operator (H1), the neighbourhood using this operator is the set of points that are reached by 1-bit flip mutation of the current solution x, hence the neighbourhood size is $|N(x)| = n$. The second operator is the 1+2 hamming operator (H1+2), the neighbourhood here includes the hamming one neighbours plus the hamming two neighbours of the current solution x which can be reached by 2-bits flip mutation, the neighbourhood size for this operator is $|N(x)| = n + (n(n-1)/2)$.

Fitness Landscape. The fitness landscape of a combinatorial optimisation problem is a triple (X, N, f), where f is the objective function $f : X \to R$, X is the search space and N is the neighbourhood operator function [10].

Local Optima. A point $x \in X$ is a local optimum iff $\forall y \in N(x), f(y) \geq f(x)$. The number of local optima found in the fitness landscape will be referred to as m.

2.1 Number of Local Optima

We investigate here if the variability of the weights correlates with the number of local optima in the NPP landscapes induced by the H1 and H1+2 neighbourhood operators. Figure 1 shows that the local optima fraction of the search

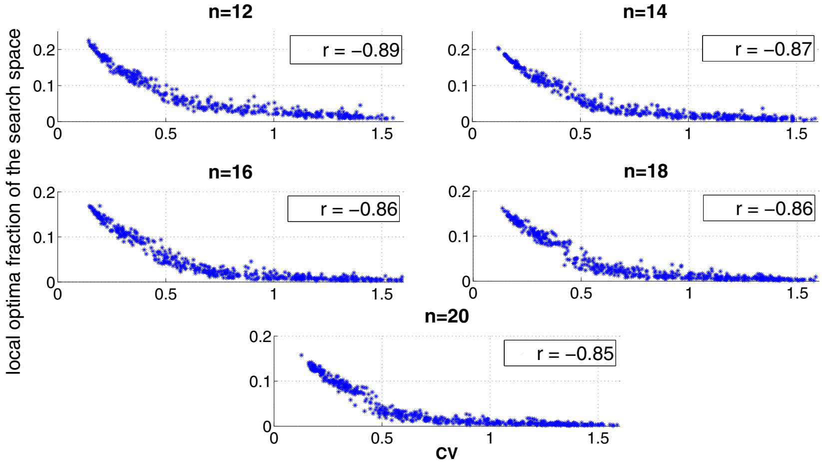

Fig. 1. Local optima as a fraction of the search space in the H1 landscape versus the coefficient of variation CV for problem sizes $n = 12$, 14, 16, 18 and 20 and for 700 instances for each problem size. The correlation coefficients (r) between CV and the the fraction of the local optima are shown for each plot.

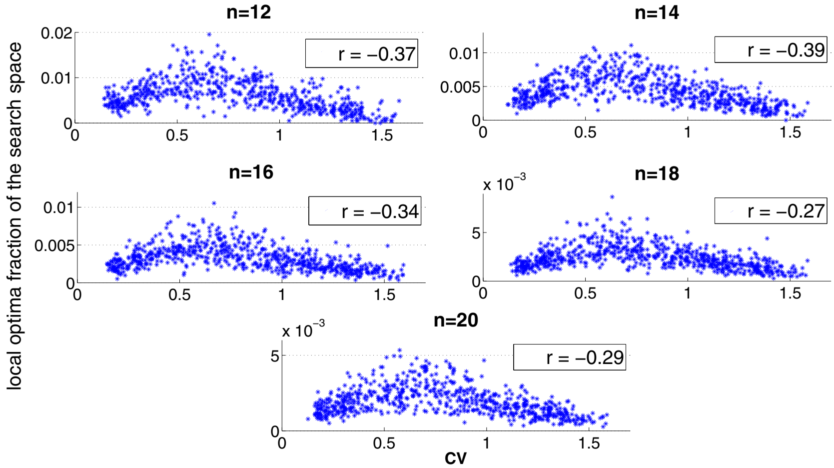

Fig. 2. Local optima as a fraction of the search space in the H1+2 landscape versus the coefficient of variation CV for problem sizes $n = 12$, 14, 16, 18 and 20 and for 700 instances for each problem size. The correlation coefficients (r) between CV and the the fraction of the local optima are shown for each plot.

space ($m/2^n$) in the H1 landscape is highly and negatively correlated with the coefficient of variation.

The intuition behind the strong correlation between the number of local optima in the H1 landscape and the CV is that, for smaller values of CV, the similarity of the weights provides many ways to arrange the weights such that moving one of the weights from one subset to the other does not lead to a better solution, resulting in a larger number of local optima in the H1 Landscape. Contrarily, in instances with larger values of CV, the discrepancy of the weights enables the same application of the 1-bit flip move operator to lead to a better solution most of the time, which result in fewer local optima.

Figure 2 shows that the fraction of local optima in the H1+2 landscape has a weak negative correlation with the CV and slightly higher fractions of local optima for instances with $(0.4 < CV < 0.8)$. In both landscapes, the fraction of local optima decreases as n gets larger. The correlation coefficients also get weaker as the problem size grows, with a faster decay in the correlation between the CV and the number of local optima in the H1+2 landscape. Higher orders of n would need to be studied to examine if and how the correlation between the CV and the number of local optima in both landscapes continue to exist in larger problem sizes.

2.2 Average Number of Local Optima

The number of local optima typically influences the performance of local search metaheuristics, and for the NPP it has been shown in [1] that the number of local optima does indeed influence the local search cost. Given that, it is interesting to be able to estimate the average number of local optima in the landscape of a given NPP instance. For instances of the NPP with weights drawn from a uniform distribution, Ferreira and Fontanari [3] derived the following formula, using statistical mechanics analysis, for the average fraction of local minima in the H1 landscape.

$$\frac{\langle m \rangle}{2^n} = \sqrt{\frac{24}{\pi}}\, n^{-3/2} \tag{3}$$

Here we propose a generalized formula for estimating the average fraction of local minima in the H1 landscape of the NPP. The formula does not require the knowledge of the distribution from which the weights are drawn and only depends on the CV of the weights and the size of the problem. The proposed formula is based on the data we observed in figure 1 and it is as follows:

$$\frac{\langle m \rangle}{2^n} = a \exp(-b\, CV) \tag{4}$$

Where the values of the coefficients a and b depend on the size of the problem. Figure 3 shows the estimation of the fraction of the local optima using this formula and with the values of a and b, determined by the least squares fit method, as shown in figure 4.

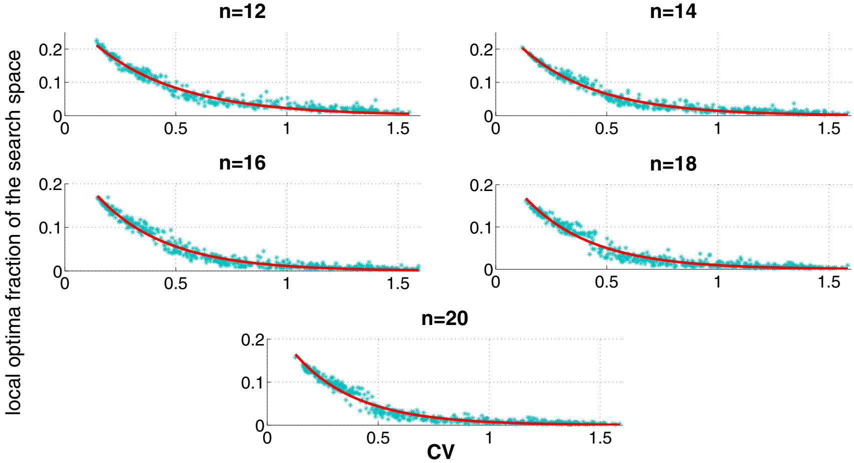

Fig. 3. The fraction of Local optima versus the coefficient of variation CV for problem sizes $n = 12$, 14, 16, 18 and 20 and for 700 instances for each problem size. The least-squares fit lines were obtained using Eq. (4) with values of a and b as shown in figure 4 and the r^2 values for the regression models ≈ 0.97.

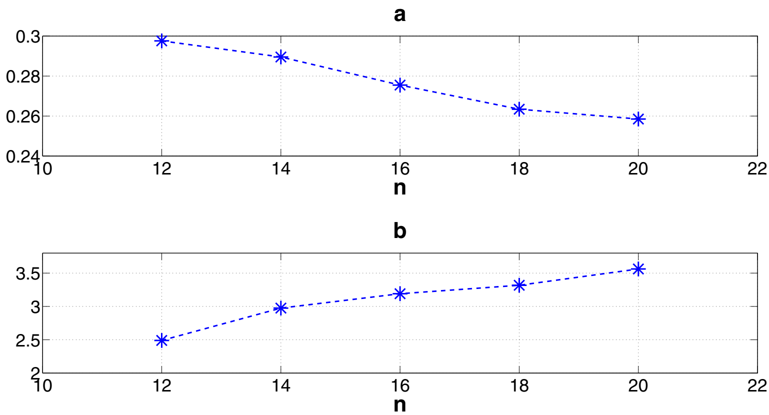

Fig. 4. The values of the a and b coefficients from Eq. (4) for problem sizes $n = 12$, 14, 16, 18 and 20, estimated using regression models. The r^2 values for the regression models are $0.96, 0.97, 0.97, 0.96, 0.97$ respectively.

2.3 Cost of Local Search

To examine how the cost of finding the optimal solution varies for different values of CV and to investigate if the coefficient of variation can be used to guide the choice of local search neighbourhood operator, steepest descent with random restart (Algorithm 1) was run with two neighbourhood operators, the H1 operator and the larger neighbourhood operator H1+2. The algorithm was run for 1000 times for each instance. The cost of finding the global optima is then calculated using the number of used fitness evaluations.

Algorithm 1. Steepest Descent with Random Restarts

repeat
  Chose $x \in \{0,1\}^n$, uniformly at random
  repeat
    choose $x' \in N(x)$, *such that* $f(x') = min_{y \in N(x)} f(y)$
    replace x *with* x' *if* $f(x') < f(x)$
  until $f(x) \leq f(x')$
until $f(x)$ is the optimal solution

Figures 5 and 6 show, respectively, the results of the steepest descent runs with H1 and H1+2 neighbourhoods operators. The figures show that the average cost of local search using H1 operator and the CV of the weights are strongly and negatively correlated, while for the H1+2 the correlation is weakly negative.

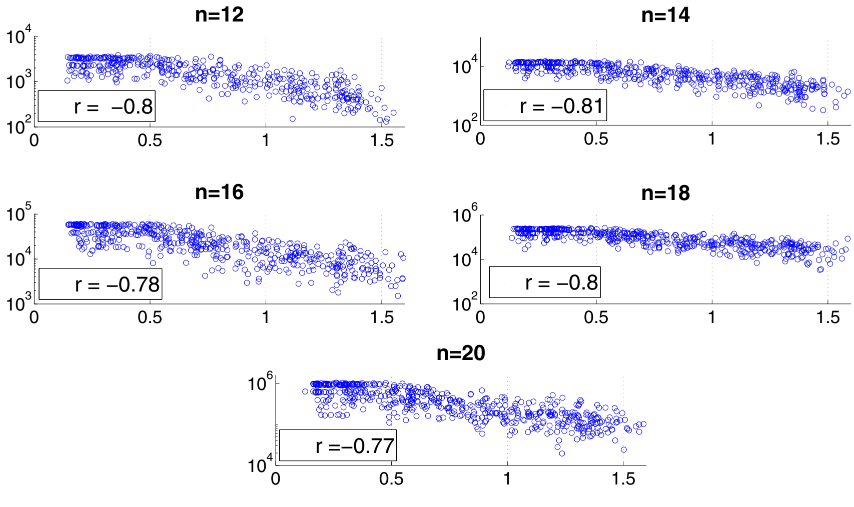

Fig. 5. The cost of steepest descent search with H1 neighbourhood operator for problem sizes n = 12, 14, 16, 18 and 20 and for the 700 instances for each problem size. The x-axes represent the coefficient of variation CV and the y-axes represent the average number of fitness evaluations used to find the global optimum in log scale. Each data point represents the average of 1000 runs of steepest descent. The correlation coefficients (r) between CV and the cost of local search are shown for each plot.

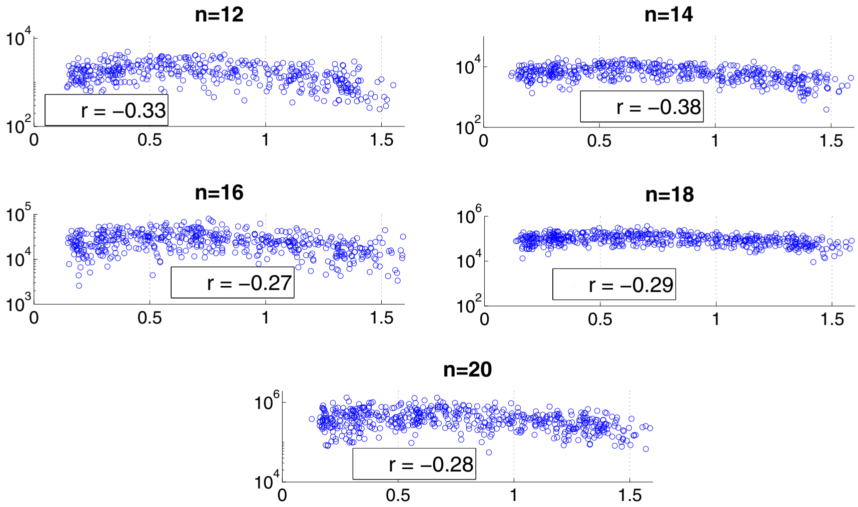

Fig. 6. The cost of steepest descent search with H1+2 neighbourhood operator for problem sizes $n = 12$, 14, 16, 18 and 20 and for the 700 instances for each problem size. The x-axes represent the coefficient of variation CV and the y-axes represent the average number of fitness evaluations used to find the global optimum in log scale. Each data point represents the average of 1000 runs of steepest descent. The correlation coefficients (r) between CV and the cost of local search are shown for each plot.

The landscape induced by the H1+2 operator has far less number of local optima than the landscape induced by the H1 operator and the difference between the number of local optima between the two landscapes is very large for smaller values of CV. Intuitively, a decrease in the number of local optima should yield a decrease in local search cost but if this decrease is not large (i.e. if the difference between the number of local optima between the two landscapes is not large) then it is possible that the advantage of having less local optima be offset by the number of fitness evaluations needed to explore the much larger neighbourhood of the H1+2 operator. To identify the values of CV, where such decrease in the number of local optima would make the use of H1+2 neighbourhood operator be more effective than the H1 neighbourhood operator, we compared the performance of the two operators. Figure 7 shows the number of instances where the performance of the algorithm with the H1 operator was significantly better than the H1+2 performance, the number of instances where the performance of the algorithm with the H1+2 operator was significantly better than the H1 performance, and the instances where the two performances were not statistically significantly different. The Wilcoxon rank-sum test was used to determine the significance of the differences between the performances of the algorithm ($p < 0.05$).

The figure shows that instances with small CV values ($CV < 0.5$), the performance of the H1+2 operator is better than the H1 operator, which is not surprising due to the low number of local optima in the H1+2 landscape and the very big difference between the number of local optima in the H1+2 landscape compared to the H1 landscape which suggest that the algorithm probably had to do far less restarts when using the H1+2 operator. For instances with large values of CV ($CV \geq 0.5$), the performance of the H1 operator is better than H1+2 operator, even though the number of local optima is less in the landscapes induced by the H1+2 operator. This could be explained by the large number of fitness evaluations needed to explore the much larger neighbourhood of the H1+2 operator. These results show that the CV of the weights has a potentially useful application in guiding the selection of the most suitable neighbourhood operator for a given NPP instance.

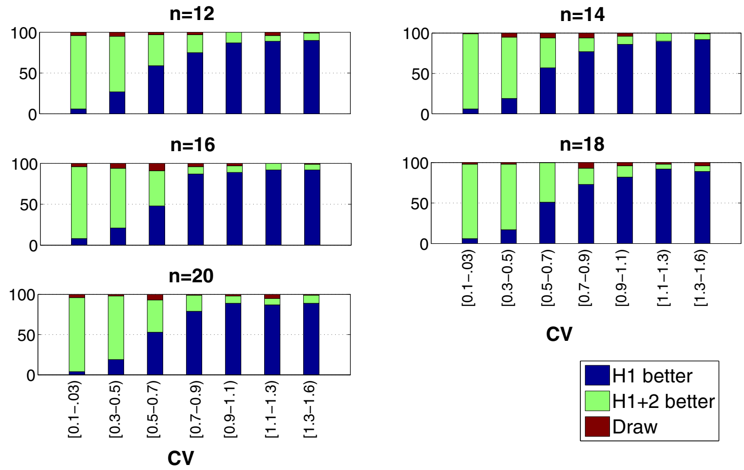

Fig. 7. Coefficient of variation CV against number of instances steepest descent with H1 neighbourhood preformed significantly better, number of instances steepest descent with H1+2 neighbourhood preformed significantly better, and instances where the performance of the two neighbourhood operators was not significantly different. For each problem instance steepest descent was run for 1000 times. The Wilcoxon rank-sum test was used to determine the significance of the differences.

3 Conclusions

In this paper, we examined how the variability of the weights influences the NPP landscape by looking at how the landscape features of the NPP change with the different values of the coefficient of variation (CV) of the weights. The CV of the weights can be easily calculated for a given instance of NPP, and does not require the knowledge of the distribution from which the weights are drawn. We found that the number of local optima and the average cost of local search in the H1 landscape are strongly and negatively correlated with the CV. For the landscapes induced by the H1+2 operator, we found that both the number of local optima and the average search cost have a weak negative correlation with the CV. We also showed what could be a practical use of the CV of the weights for guiding the choice of the move operators of local search heuristics.

We proposed a formula to estimate the average number of local optima in the H1 landscape that depends only on the problem size and the CV of the weights, exploiting the strong correlation between the CV and the number of local optima in the H1 landscape. We still need as future work to look at larger problem sizes to be able to predict the relation between the two coefficients (a and b) in the formula and the size of the problem. For that we are going to use some sampling techniques to estimate the number of local optima [5].

References

1. Alyahya, K., Rowe, J.E.: Phase transition and landscape properties of the number partitioning problem. In: Blum, C., Ochoa, G. (eds.) EvoCOP 2014. LNCS, vol. 8600, pp. 206–217. Springer, Heidelberg (2014)
2. Borgs, C., Chayes, J., Pittel, B.: Phase transition and finite-size scaling for the integer partitioning problem. Random Structures & Algorithms 19(3-4), 247–288 (2001)
3. Ferreira, F.F., Fontanari, J.F.: Probabilistic analysis of the number partitioning problem. Journal of Physics A: Mathematical and General 31(15), 3417 (1998)
4. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. Series of books in the mathematical sciences. W. H. Freeman (1979)
5. Garnier, J., Kallel, L.: How to detect all maxima of a function. In: Theoretical Aspects of Evolutionary Computing. Natural Computing Series, pp. 343–370. Springer, Heidelberg (2001)
6. Mertens, S.: A physicist's approach to number partitioning. Theoretical Computer Science 265(1-2), 79–108 (2001)
7. Prügel-Bennett, A., Tayarani-Najaran, M.: Maximum satisfiability: Anatomy of the fitness landscape for a hard combinatorial optimization problem. IEEE Transactions on Evolutionary Computation 16(3), 319 (2012)
8. Slaney, J., Walsh, T.: Backbones in optimization and approximation. In: IJCAI, pp. 254–259 (2001)
9. Stadler, P.F., Hordijk, W., Fontanari, J.F.: Phase transition and landscape statistics of the number partitioning problem. Physical Review E 67(5), 056701 (2003)
10. Stadler, P.F., Stephens, C.R.: Landscapes and effective fitness. Comments on Theoretical Biology 8(4-5), 389–431 (2002)
11. Watson, J.P., Whitley, L.D., Howe, A.E.: Linking search space structure, run-time dynamics, and problem difficulty: A step toward demystifying tabu search. J. Artif. Int. Res. 24(1), 221–261 (2005)
12. Witt, C.: Worst-case and average-case approximations by simple randomized search heuristics. In: Diekert, V., Durand, B. (eds.) STACS 2005. LNCS, vol. 3404, pp. 44–56. Springer, Heidelberg (2005)
13. Zhang, W.: Configuration landscape analysis and backbone guided local search.: Part i: Satisfiability and maximum satisfiability. Artificial Intelligence 158(1), 1–26 (2004)