Threshold Arrangements and the Knapsack Problem

Günter Rote* André Schulz*

February 3, 2005

Abstract

We show that a combinatorial question which has been studied in connection with lower bounds for the knapsack problem by Brimkov and Dantchev (2002) is related to *threshold graphs*, *threshold arrangements*, and other well-studied combinatorial objects, and we correct an error in the analysis given in that paper.

Keywords: Knapsack problem, Threshold graphs, Real number model

1 Introduction

The complexity of the knapsack problem has been deeply investigated under various computational models [5, 7]. In particular, this problem plays an important role in algebraic complexity theory. Here, the considered model is a real number computation model, which was established by Blum *et al.* [2]. In this model, an arithmetic operation with infinite precision costs only constant time. One of the key theorems in algebraic complexity theory is the *Ben-Or lower bound theorem* [1], which states that solving a decision problem costs $\Omega(\log(\#c.c.))$ operations. Here $\#c.c.$ stands for the number of connected no-instances in the space of all input vectors of the decision problem. The knapsack problem is one of the problems where the Ben-Or theorem can be applied to get an $\Omega(n^2)$ complexity bound.

In [3] the authors aimed to prove an alternative statement for the $\Omega(n^2)$ bound of the knapsack problem. We will show that their main lemma, which claims an exact formula for certain combinatorial structure, is wrong. This will be done by presenting a counterexample in section 2.

In this note, we will also show that the combinatorial objects studied by [3] have been already studied in the literature, in various incarnations, and we mention a few of the most important results about them.

Moreover, we indicate in Section 4 that, for a completely independent reason, the main result of the paper [3] is also in error.

2 The Knapsack Problem

The *knapsack problem* is a decision problem. It asks if, for a given $a \in \mathbb{R}^n$, there exists some $x \in \{0,1\}^n$ such that $a^T x = 1$. In other words, we are looking for a subset of the a_i's whose sum equals 1. This subset is denoted by the characteristic vector x.

*Institut für Informatik, Freie Universität Berlin,Takustraße 9, D-14195 Berlin, Germany, email:{rote,schulza}@inf.fu-berlin.de

Supported by the Deutsche Forschungsgesellschaft (DFG) under grant RO 2338/2-1

For the application of the Ben-Or Theorem, we need to count the number of connected components of the set

$$C_{no} := \{a \in \mathbb{R}^n \mid \forall x \in \{0,1\}^n \ a^T x \neq 1\}.$$

This means that the unit hypercube $[0,1]^n$ of all possible input vectors $(a_1, \ldots, a_n)$ is dissected into cells by the $2^n - 1$ *knapsack hyperplanes* h_I which are given by the equations $\sum_{i \in I} a_i = 1$, for all nonempty subsets $I \subseteq \{1, 2, \ldots, n\}$. It is known that the knapsack arrangement has at least $4 \cdot 2^{\binom{n}{2}}$ cells [4]. Hence the algebraic decision tree complexity is $\Omega(n^2)$ [1].

For technical reasons, the authors of [3] restrict their attention to arguments with $1/3 < a_i < 2/3$. Let A_n denote the number of cells into which the hypercube $(1/2, 2/3)^n$ is dissected by the knapsack hyperplanes. Clearly the number of cells in $(1/2, 2/3)^n$ gives a lower bound for the number of cells in the unit hypercube.

The only hyperplanes h_I which intersect $(1/3, 2/3)^n$ are hyperlanes $h_{\{i,j\}} : a_i + a_j = 1$ where I contains exactly two elements. Thus a cell is characterized by specifying, for each pair $\{i, j\}$, on which side of the hyperplane $h_{\{i,j\}}$ it lies. So, every cell corresponds to a set $S_a = \{\, \{i, j\} \mid i \neq j,\ a_i + a_j < 1 \,\}$. The reader should notice that not all combinations of pairs $\{i, j\}$ will lead to proper defined sets. For example, the set $S_a = \{\{1, 2\}, \{3, 4\}\}$ will not define a proper cell. It induces the inequalities

$$\begin{array}{rcl} a_1 + a_2 & < & 1 \\ a_1 + a_3 & \geq & 1 \\ a_3 + a_4 & < & 1 \\ a_2 + a_4 & \geq & 1 \end{array}$$

From the first two inequalities follows that $a_2 < a_3$. On the other hand the last two inequalities lead to $a_2 > a_3$. Hence S_a does not correspond to a properly defined cell.

3 Threshold Arrangements and Threshold Graphs

The translation of the arrangement of the hyperplanes $h_{\{i,j\}}$ by the vector $(-1/2, \ldots, -1/2)$ will lead to a more natural arrangement. This new arrangement which consists of the hyperplanes $h'_{\{i,j\}} : a_i + a_j = 0$ is known as the *threshold arrangement* [8, exercise 5.4]. Clearly the number of cells is invariant under translation of the whole arrangement.

A *threshold graph* is a graph (V, E), for which a weight assignment $w : V \to \mathbb{R}$ and some $t \in \mathbb{R}$ exist, such that for any distinct vertices i, j

$$(i, j) \in E \Leftrightarrow w_i + w_j < t$$

There exist many other characterization of threshold graphs, for a survey see [6]. If we view the unordered pairs in S_a as the edges of a graph, the class of graphs that we obtain in this way is exactly the class of *threshold graphs.* This follows directly from the given characterization of threshold graphs. Thus A_n denotes not only the number of cells in the dissected hypercube, but also the number of threshold graphs and distinct sets S_a.

In [3, Lemma 1] it is claimed that $A_n = n!$. This statement is wrong. The first few values in this sequence are $1, 2, 8, 46, \ldots$. The first term which is not equal to $n!$ is $A_3 = 8$. Indeed, the examples of Table 1 prove that all 2^3 subsets S_a of

$\{\{1,2\},\{1,3\},\{2,3\}\}$ can occur as the edge set of some threshold graph. The reader can check that Lemma 1 in [1] actually proves a *lower bound* $A_n \geq n!$. We leave it to the interested reader to find the error in the proof.

a_1	a_2	a_3	S_a
0.6	0.5	0.55	$\{\}$
0.4	0.5	0.65	$\{\{1,2\}\}$
0.5	0.6	0.45	$\{\{1,3\}\}$
0.6	0.5	0.45	$\{\{2,3\}\}$
0.4	0.5	0.55	$\{\{1,2\},\{1,3\}\}$
0.5	0.4	0.55	$\{\{1,2\},\{2,3\}\}$
0.6	0.5	0.35	$\{\{1,3\},\{2,3\}\}$
0.4	0.5	0.45	$\{\{1,2\},\{1,3\},\{2,3\}\}$

Table 1: All graphs on three vertices are threshold graphs. We give one representative a for every set.

One possible alternative characterization of threshold graphs can be done by giving a construction scheme. Every threshold graph can be generated in the following way: We start with a single vertex and add the other vertices one after another in some order. A new vertex v can be either isolated (no edge between v and a previous vertex) or dominating (all previous vertices share an edge with v).

We call the a sequence of consecutive dominating or isolated vertices a *block*. If we change the order inside a block, we will still construct the same graph. Hence it suffices to analyze the ordered partitions of all permutation of $[n]$ to retrieve the numbers of possible threshold graphs. Together with the observation that the first block must consist of at least two elements, we obtain an expression for A_n in terms of Stirling numbers $s(n,k)$ of the second kind.

$$A_n = 2 \cdot \left(\sum_{k=0}^{n-1} k! s(n,k) - n(k-1)!\, s(n-1,k-1) \right)$$

Thus we are able to calculate A_n (and at the same time the whole sequence $A_1, A_2, \ldots, A_n$) in $O(n^2)$ arithmetic steps.

An exponential generating function for the numbers of labeled threshold graphs can be found in [6, chapter 17.2] and [8, page 106]:

$$G(x) = \sum_{n=1}^{\infty} A_n \frac{x^n}{n!} = \frac{e^x(1-x)}{(2-e^x)}$$

The generating function leads to the asymptotic bound

$$\frac{A_n}{n!} \approx \left(\frac{1}{\log 2} - 1\right)\left(\frac{1}{\log 2}\right)^n$$

4 Lower bounds for the knapsack problem

Brimkov and Dantchev [3] apply their Lemma 1 to prove a lower bound on the knapsack problem. The wrong analysis of the number of threshold graphs does not invalidate their application of the lemma to this problem, since they need only a lower bound on A_n. Moreover the important quantity in the proof is the asymptotic behavior of the logarithm of A_n, which is the same as for $n!$, apart from a constant factor: $\log A_n = \Theta(n \log n) = \Theta(\log n!)$.

However, the main result of their work is flawed for a different reason. Theorem 1 of the paper reads as follows:

> No algorithm solving the knapsack problem can achieve a time complexity $o(n \log n) \cdot f(a_1, \ldots, a_n)$ where f is an arbitrary continuous function of n variables.

The theorem tries to address algorithms whose running time is sensitive to the data $a_1, a_2, \ldots, a_n$ and does not just depend on n. For example, the well-known dynamic programming algorithm for the knapsack problem takes $O(n/\delta)$ time, where $\delta(a_1, a_2, \ldots, a_n)$ is the smallest difference between two *distinct* elements of the set of all sums that can be formed from subsets of the input numbers $a_1, a_2, \ldots, a_n$. The function δ, however, is discontinuous.

An obvious counterexample for the Theorem 1 of [3] is given by the function $f(a_1, a_2, \ldots, a_n) = 2^n + a_1 + a_2 + \cdots + a_n$, which is clearly a continuous function of $a_1, a_2, \ldots, a_n$. (Here, the additive term $a_1 + \cdots + a_n$ serves only to make the function more interesting.) The trivial algorithm which simply checks all 2^n subsets takes $O(n2^n) = O(n \cdot f(a_1, a_2, \ldots, a_n))$ time.

When one reads the proof of Theorem 1 in [3] one can get a glimpse of the authors' intentions. However, we could not think of a meaningful variation or modification of their statement which would be interesting. The trap into which the argument fell is apparently a confusion about the proper quantification of the variable n.

References

[1] Ben-Or, M. Lower bounds for algebraic computation trees. In *Proceedings of the 15th ACM Symposium on Theory of Computing* (1983), ACM Press, pp. 80–86.

[2] Blum, L., Cucker, F., Shub, M., and Smale, S. *Complexity and real computation.* Springer-Verlag New York, 1998.

[3] Brimkov, V., and Dantchev, S. An alternative to Ben-Or's lower bound for the knapsack problem complexity. *Applied Mathematics Letters* **15**, 2 (February 2002), 187–191.

[4] Bürgisser, P., Clausen, M., and Amin Shokrollahi, M. *Algebraic Complexity Theory*, Grundlehren der mathematischen Wissenschaften, Vol. 315. Springer, 1997.

[5] Garey, M., and Johnson, D. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman and Co., San Fransico, 1979.

[6] Mahadev, N., and Peled, U. *Threshold Graphs and Related Topics*, Ann. of Discrete Mathematics, Vol. 56. North-Holland, 1995.

[7] Martello, S., and Toth, P. *Knapsack Problems Algorithms and Computer Implementations.* Series in Discrete Mathematics and Optimization. Wiley-Interscience, 1990.

[8] Stanley, R. P. *Enumerative Combinatorics. Vol. 2.* Cambridge University Press, Cambridge, 1999.