COMBINATORIAL OPTIMIZATION THROUGH ORDER STATISTICS

Extended Abstract

Wojciech Szpankowski*
Department of Computer Science
Purdue University
West Lafayette, IN 47907

Abstract

A mathematical model studied in this paper can be formulated as follows: find the optimal value of \( Z_{\text{max}} = \max_{\alpha \in B_n} \{ \sum_{i \in S_n(\alpha)} w_i(\alpha) \} \) (respectively \( Z_{\text{min}} \), where \( \Sigma \) is an operator, \( n \) is an integer, \( B_n \) is the set of all feasible solutions, \( S_n(\alpha) \) is the set of all objects belonging to the \( \alpha \)-th feasible solution, and \( w_i(\alpha) \) is the weight assigned to the \( i \)-th object. Our interest lies in finding an asymptotic solution to this optimization problem in a probabilistic framework. Using some novel results from order statistics, we investigate in a uniform manner a large class of combinatorial optimization problems such as: the assignment problem, the traveling salesman problem, the minimum spanning tree, the minimum weighted \( k \)-clique, the bottleneck assignment and traveling salesman problems, location problems on graphs, and so forth. For example, we provide some sufficient conditions that assure asymptotic optimality of a greedy algorithm.

1. INTRODUCTION

We consider in this paper a class of generalized optimization problems in a probabilistic framework. A general mathematical model is as follows: for some integer \( n \) define
\[
Z_{\text{max}} = \max_{\alpha \in B_n} \{ \sum_{i \in S_n(\alpha)} w_i(\alpha) \} \text{ (respectively } Z_{\text{min}} \text{),}
\]
where \( \Sigma \) is an operator (e.g., \( \Sigma = \sum \) or \( \Sigma = \min \), etc.), \( B_n \) is the set of all feasible solutions, \( S_n(\alpha) \) is the set of all objects belonging to the \( \alpha \)-th feasible solution, and \( w_i(\alpha) \) is the weight assigned to the \( i \)-th object. For example, in the traveling salesman problem [BOR62, KAR76, LLK85, WEI80] the operator \( \Sigma \) becomes a sum \( \Sigma \) operator, \( B_n \) represents the set of all Hamiltonian paths, \( S_n(\alpha) \) is the set of edges that fall into the \( \alpha \)-th Hamiltonian path, and \( w_i(\alpha) \) is the length of the \( i \)-th edge; for the bottleneck traveling salesman problem [GAG78, WEI80] the operator \( \Sigma \) becomes "min" operator. Some other examples include the assignment problem [BOR62, FHR87, WAL79, WEI80, LLK85], the minimum spanning tree [BOL85, KNU73], the minimum weighted \( k \)-clique problem [LUK81, BOL85] the bottleneck and capacity assignment problems [WEI80], geometric location problems [PAP81], and some others not directly related to optimization such as the height and depth of digital trees [KNU73, SZP88a, SZP91], the maximum queue length, hashing with lazy deletion, pattern matching, edit distance [AHU74, UKK90] and so forth. In our probabilistic framework, we assume that the weights \( w_i(\alpha) \) are random variables drawn from a common distribution function \( F(\cdot) \). Our interest lies in finding an asymptotic solution of \( Z_{\text{max}} \) (\( Z_{\text{min}} \)) and the \( k \)-th best solution \( Z_{(k)} \) in some probability sense for a large class of distribution functions \( F(\cdot) \), and apply these findings to design efficient heuristic algorithms that achieve asymptotically the optimal performance.

Our analysis does not assume any a priori information regarding the distribution of inputs. We, however, identify two classes of distributions that lead to precise asymptotic expansions for \( Z_{\text{max}} \) and \( Z_{\text{min}} \). We present several new results that are grouped into two categories, namely general results (Section 3.2) and specific solutions (Section 3.3). All of these results are obtained in a systematic way using a variety of tools from order statistics (Section 3.1). We have novel results in all three facets mentioned above. In particular, Lemma 1a of Section 3.1 establishes a simple characteristic of a probabilistic behavior of the \( k \)-th best solution \( Z_{(k)} \) of our optimization problem. Such a generalization of the problem is necessary in

---

*This research was supported by AFOSR grant 90-0107, in part by the NSF grant CCR-8900305, and in part by grant R01 LM05118 from the National Library of Medicine.
many fields of science, most notably in molecular biology (DNA and protein foldings), pattern recognition, and so forth. Moreover, Lemma 3 presents a systematic approach to analyze the limiting distribution of the optimal solutions \( Z_{\text{max}} \) and \( Z_{\text{min}} \). These probabilistic tools are next used to design practical heuristic algorithms for some optimization problems. This is simply achieved by comparing (in some probability sense) the performance of the optimal solution with the performance of a specific solution, and therefore controlling the performance quality of the heuristic. In particular, we present sufficient conditions under which a greedy algorithm achieves asymptotically the same performance as the optimal one (cf. Theorem 4). Our next result concentrates on a large class of bottleneck optimization problems and demonstrates a constructive probabilistic approach to design semi-optimal algorithms for arbitrary distribution functions (cf. Theorem 5). Finally, the list of our general results is concluded by a finding concerning the additive objective function (i.e., \( \Xi = \Sigma \)) that is virtually without probabilistic assumptions however it restricts the size of the input (cf. Theorem 6). We have also several specific results and suggest several approximate algorithms. In particular, we discuss the linear assignment problem (cf. Problem 1), maximal properties of digital trees (cf. Problem 2), the optimal weighted \( k \)-clique problem (cf. Problem 3), the optimal location problem (cf. Problem 4). In all categories we have obtained new results, and more importantly all of these problems - previously treated by disparate methods - are analyzed in this paper in a uniform manner.

2. PROBLEM STATEMENT

Let \( n \) be an integer (e.g., number of vertices in a graph, number of keys in a digital tree, etc.), and \( S \) a set of objects (e.g., set of vertices, keys, etc). We shall investigate the optimal values \( Z_{\text{max}} \) and \( Z_{\text{min}} \) defined as follows

\[
Z_{\text{max}} = \max_{\alpha \in B_n} \{ \sum_{i \in S_n(\alpha)} w_i(\alpha) \} \quad Z_{\text{min}} = \min_{\alpha \in B_n} \{ \sum_{i \in S_n(\alpha)} w_i(\alpha) \},
\]

where \( B_n \) is a set of all feasible solutions, \( S_n(\alpha) \) is a countable set of objects from \( S \) belonging to the \( \alpha \)-th feasible solution, and \( w_i(\alpha) \) is the weight assigned to the \( i \)-th object in the \( \alpha \)-th feasible solution (in addition, by \( w_{ij} \) we denote a weight assigned to a pair of objects \((i,j)\) in \( S \)). Throughout this paper, we adopt the following assumptions:

(A) The cardinality \( |B_n| \) of \( B_n \) is fixed and equal to \( m \). The cardinality \( |S_n(\alpha)| \) of the set \( S_n(\alpha) \) does not depend on \( \alpha \in B_n \) and for all \( \alpha \) it is equal to \( N \), i.e., \( |S_n(\alpha)| = N \).

(B) For all \( \alpha \in B_n \) and \( i \in S_n(\alpha) \) the weights \( w_i(\alpha) \) (i.e., the weights \( w_{ij} \)) are identically and independently distributed (i.i.d) random variables with common distribution function \( F(\cdot) \), and the mean value \( \mu \) and the variance \( \sigma^2 \).

The assumption (B) defines a probabilistic model of our problem (2.1). We shall explore the asymptotic behaviors of \( Z_{\text{max}} \) and \( Z_{\text{min}} \) as \( n \) becomes large (in probability and/or almost surely sense).

There are many combinatorial problems that fall into our formulation (2.1). For example, the linear assignment problem \([BOL85, WAL79, FHR87]\), the traveling salesman problem \([LLK85]\), the minimum spanning tree \([AHU74, BOL85]\), the minimum weighted \( k \)-clique problem \([BOL85, LUE81]\), and so forth. In particular, in the linear assignment problem \( |B_n| = n! \), \( |S_n(\alpha)| = n \) and the weights \( w_i(\alpha) \) are elements of a matrix; in the traveling salesman problem \( B_n \) is the set of all Hamiltonian paths and \( S_n(\alpha) \) is a set of \( n \) edges in the \( \alpha \)-th Hamiltonian path, that is, \( N = |S_n(\alpha)| = n \); in the \( k \)-clique problem \( B_n \) is defined as a set of all \( k \)-cliques in a graph and \( S_n(\alpha) \) the set of edges belonging to the \( \alpha \)-th \( k \)-clique, and so forth.

So far, we have restricted our attention to problems which can be represented as (2.1). In practice some other objective functions are important. For example, in a class of bottleneck and capacity problems \([GaG78, HoS86 \text{SZP90}]\) the operator ‘\( \Sigma \)’ in (2.1) is replaced by ‘\( \max \)’ and ‘\( \min \)’, respectively. Therefore, we extend (2.1) to the following

\[
Z_{\text{max}} = \max_{\alpha \in B_n} \{ \Xi_{i \in S_n(\alpha)} w_i(\alpha) \} \quad Z_{\text{min}} = \min_{\alpha \in B_n} \{ \Xi_{i \in S_n(\alpha)} w_i(\alpha) \}
\]

where \( \Xi \) is an operator applied to a set \( \{ w_i(\alpha), i \in S_n(\alpha) \} \), e.g., in (2.1) \( \Xi = \Sigma \). For example, in the bottleneck and capacity assignment problems, and bottleneck and capacity traveling salesmen problems \([GaG78, HoS86]\) the operator \( \Xi \) becomes either "\( \max \)" or "\( \min \)" respectively.