European Journal of Operational Research 255 (2016) 357–363

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

Asymptotic behavior of the quadratic knapsack problem

Joachim Schauer*

Department of Statistics and Operations Research, University of Graz, Universitaetsstr. 15, Graz A-8010, Austria

ARTICLE INFO

Article history:
Received 25 August 2015
Accepted 6 June 2016
Available online 11 June 2016

Keywords:
Quadratic knapsack problem
Asymptotic analysis
Random instances

ABSTRACT

We study the quadratic knapsack problem on instances where the profits are independent random variables defined on the interval [0, 1] and the knapsack capacity is proportional to the number of items (we assume that the weights are arbitrary numbers from the interval [0, 1]). We show asymptotically that the objective value of a very easy heuristic is not far away from the optimal solution. More specifically we show that the ratio of the optimal solution and the objective value of this heuristic almost surely tends to 1 as the size of the knapsack instance tends to infinity. As a consequence using randomly generated instances following this scheme seems to be inappropriate for studying the performance of heuristics and (to some extend) exact methods. However such instances are frequently used in the literature for this purpose. Additionally we introduce a class of test instances based on hidden cliques for which finding a good solution is much harder. We support this by theoretical observations as well as by performing computational experiments.

1. Introduction

In the quadratic knapsack problem (QKP) we are given a set of n items each with an integer profit p_j and weight w_j. Moreover for each pair of items an additional profit p_{ij} is given which accounts for interdependencies between the involved items. We look for a subset of items whose total weight does not exceed a given capacity bound c and whose total profit is maximized. The problem can be modeled by the well known quadratic integer programing formulation:

$$(QKP) \qquad \max \sum_{i=1}^{n} p_i x_i + \sum_{1 \le i < j \le n} p_{ij} x_i x_j \tag{1}$$

$$\text{s.t.} \sum_{i=1}^{n} w_i x_i \le c \tag{2}$$

$$x_i \in \{0, 1\}, \quad i = 1, \ldots, n \tag{3}$$

Here, $x_j = 1$ means that item I_j is included into the knapsack. Note that some profits p_{ij} might also be 0, meaning that the involved items do not influence each other at all.

The quadratic knapsack problem is a classical combinatorial optimization problem which has many real-world applications (see e.g. Pisinger, 2007, Kellerer, Pferschy, & Pisinger, 2004, Section 12). The *NP*-hardness of QKP can be easily seen by a reduction from the maximum clique problem. When negative profits are allowed in a QKP instance (Rader Jr. & Woeginger, 2002) proved that the problem becomes inapproximable, however for instances containing only non-negative profits the approximability behavior remains open due to the following connection to the *densest-k subgraph* problem: this problem asks for a k vertex induced subgraph of a given graph $G = (V, E)$ containing the maximum number of edges. This problem can be modeled as a *QKP*: the vertices v_i of V correspond to items I_i and all items have profits $p_i = 0$ and weights $w_i = 1$. The quadratic profit p_{ij} is set to 1 whenever $(v_i, v_j) \in E$ and otherwise to 0. The knapsack constraint is set to k. Note that the densest-k subgraph problem is infamous for its open approximability status. The existence of a *PTAS* for this problem is still open (under the standard assumption $\mathcal{P} \neq \mathcal{NP}$) and the best known approximation algorithm of Bhaskara, Charikar, Chlamtac, Feige, and Vijayaraghavan (2010) has a performance ratio of $O(n^{\frac{1}{4}})$. Inapproximability results are only known under weaker complexity assumptions: Feige (2002) and Khot (2006) ruled out the existence of a PTAS under complexity assumptions dealing with average case hardness. Alon, Arora, Manokaran, Moshkovitz, and Weinstein (2011) showed that an hardness assumption on random k-AND formulas rules out the existence of any constant factor approximation algorithm. Finally (Alon et al., 2011) even showed superconstant inapproximation results based on the hidden clique assumption. Clearly all these results also hold for QKP.

Due to the need for solutions of QKP instances originating in real world problems there are many papers focusing on exact algorithms and (meta-)heuristics in the literature. In fact many of

* Tel.: +43 3163807241.
E-mail address: joachim.schauer@uni-graz.at

http://dx.doi.org/10.1016/j.ejor.2016.06.013
0377-2217/

them computationally show a very good performance, which in some sense contrasts the hardness of approximation results from above. Caprara, Pisinger, and Toth (1999) provided a fast and successful branch and bound algorithm where the calculation of the upper bounds is based on Lagrangian relaxation. Billionnet and Soutif (2004) presented an algorithm based on Lagrangian decomposition. Note that the algorithm of Caprara et al. (1999) works better for dense instances, whereas the algorithm of Billionnet and Soutif (2004) has its strength on spare instances. Both were able to solve instances of about 400 items. Helmberg, Rendl, and Weismantel (2000) presented bounds based on semi-definite programing and cutting planes. Pisinger, Rasmussen, and Sandvik (2007) introduced an aggressive reduction strategy in order to fix some variables to its optimal solution values and demonstrated that instances of size up to 1500 variables can often be reduced significantly to instances containing about 100 items only. After the reduction strategy they used an adaption of the branch and bound algorithm of Caprara et al. (1999). Recently (Fomeni, Kaparis, & Letchford, 2014) presented a cut and branch algorithm which is especially successful on sparse instances. Létocart, Nagih, and Plateau (2012) gave a method for improving the computation time for getting upper bounds which can be incorporated in the above approaches. Note that Pisinger (2007) surveys some of these methods (and others not mentioned here) in detail.

Notable meta-heuristic approaches are the genetic algorithms of Julstrom (2005) and Julstrom (2012). The algorithm based on Tabu search and GRASP from Yang, Wang, and Chu (2013) seems to be the currently best performing meta heuristic for QKP: even for instances containing 2000 items the gap between the best found solution and an upper bound from Caprara et al. (1999) was small (< 1.5 percent). A heuristic by Fomeni and Letchford (2013) which modified the classical dynamic program for the $0-1$ knapsack problem together with a simple local search was able to find the optimum for almost all standard instances of size up to 380 items.

1.1. Main contribution

The main result of this paper indicates that the mismatch between the hardness results and the good performance of the heuristic approaches can be explained by the test instances used. The standard benchmark instances have its origin in Gallo, Hammer, and Simeone (1980) and were subsequently used in most computational papers. First a density δ is defined which corresponds to the probability that a profit p_i or p_{ij} is non-zero. Whenever a profit p_{ij} $(i < j)$ or p_i is non-zero it is drawn uniformly at random from the interval [1, 100]. The weights w_i are drawn uniformly at random from the interval [1, 50]. Finally the capacity is drawn uniformly at random from the interval $[0, \sum w_i]$ (in more recent papers this interval was changed to $[50, \sum w_i]$).

We will show that the following very easy heuristic will produce a solution for instances of type (Gallo et al., 1980) whose objective value is (asymptotically) very close to the optimal solution value: sort the items in non-decreasing order of their weights and include the items greedily as long as they fit.

More specifically we will show that QKP restricted to instances where the profits are independent and (to some extent) identically distributed random variables, the weights are arbitrary numbers from the interval [0, 1] and the knapsack constraint is linear in the number of items has the following property: the ratio of the solution value of the above algorithm and the optimal solution value almost surely tends to 1. Note that the result even holds whenever the profits are defined on an interval $[0, M]$ for some constant M. Hence, the instances of Gallo et al. (1980) fit into this scheme whenever the drawn capacity is linear in the number of items.

Pisinger et al. (2007) introduced other types of instances which we will discuss in Section 3. Moreover we will introduce a new class of instances for which the main result of this paper is not applicable. This class is based on the hidden clique assumption, which has been studies a lot in the past, but not in the context of QKP. In Section 4 we will computationally show that they are indeed hard to solve for selected state of the art heuristic approaches from the literature.

1.2. Asymptotic results for the quadratic assignment problem

For the quadratic assignment problem and a related family of optimization problems called *generic optimization problems* similar results are known due to Burkard and Fincke (1985). In a generic optimization problem a set E is given and the set F of feasible solutions S is a subset of the power set of E (i.e. $F \subseteq \mathcal{P}(E)$). Moreover every element $e \in E$ gets assigned a cost $c(e)$. The sum objective function is defined in the classical way for every set S as:

$$C(S) = \sum_{e \in S} c(e)$$

Burkard and Fincke (1985) showed that whenever the costs $c(e)$ are i.i.d. random variables defined on the interval [0, 1] and some additional assumptions are fulfilled then for generic optimization problems with sum objective function the ratio of the worst and best feasible solution asymptotically tends to 1 in probability. The most important additional assumptions are the following:

1. The size of every feasible solution S is the same.
2. $\exists \lambda > 0 : \lim_{n\to\infty}(\lambda|S| - \log|F|) = \infty$

Assumption 2 is crucial and this property (though not stated explicitly in this form) is also central for our problem. It states that the number of feasible solutions is relatively small compared to the number of elements in a feasible solution S. Assumption 1 however does not hold for QKP. Later this result was extended to almost sure convergence by Szpankowski (1995). Similar results also hold for bottleneck objective functions.

Note however that there exists a result by Dyer, Frieze, and McDiarmid (1986) which states that certain branch and bound algorithms will take more than exponential time for quadratic assignment instances with random costs. Hence, the results mentioned in this section as well as our result do not imply that finding the optimal solution of such a problem is easy. They just imply that finding good solutions is easy.

2. Asymptotic analysis

For deriving our main result we need the following Chernoff–Hoeffding bounds that are based on Angluin and Valiant (1979) and can be found in McDiarmid (1998, Theorem 2.3 p. 200):

Theorem 1. *Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $0 \le X_k \le 1$ for each k. Let $S_n = \sum X_k$ and let $\mu = E(S_n)$. Then for any $0 \le \varepsilon \le 1$:*

$$P[S_n \ge (1+\varepsilon)\mu] \le e^{-\frac{1}{3}\varepsilon^2\mu}$$

$$P[S_n \le (1-\varepsilon)\mu] \le e^{-\frac{1}{2}\varepsilon^2\mu}$$

In this section we assume that the linear profits P_i of a quadratic knapsack instance are i.i.d. random variables and the quadratic profits P_{ij} are i.i.d. as well. Both distributions are defined on the interval [0, 1]. Note that the two distributions may differ, but the P_i and the P_{ij} are assumed to be mutually independent. Additionally the weights are arbitrary numbers from [0, 1]. This means that the weights might be random or part of the input. Moreover we assume that the knapsack capacity c is linear in n (i.e. $c = \lambda n$ for some constant $\lambda \le 1$). We moreover assume that