Knapsack problems: Linear relaxations and greedy algorithms

A. A. Korbut*, I. Kh. Sigal †

*Institute for Economics and Mathematics RAS, korbut@emi.nw.ru
†Dorodnicyn Computing Center RAS, isigal@ccas.ru

Introduction. Discrete optimization plays an increasingly important role in various control problems. One of the most well-known discrete optimization problems is the knapsack problem [1] which models a broad range of practical problems (distribution of indivisible resources, choice of projects, cutting-stock problems, cryptography, financial decisions etc.). The exact solution of knapsack problems can be very laborious due to their NP-hardness [1]. Therefore the main attention nowadays is payed to the development of approximate methods (and this tendency is characteristic for the entire domain of discrete optimization). One of the possible approaches consists in solving the linear relaxation of the problem (its optimal value yields a good upper bound for the optimal value of the original problem). Another approach is connected with the use of greedy algorithms [1, 2, 4–7]. The first approach was investigated in [4] where the ratio Δ of the values of objective functions for the optimal solution of the linear relaxation and the optimal integer solution was considered. The paper [6] is dedicated to the second approach. Here the ratio δ of the objective function values for the optimal and the greedy solutions was analyzed. In both papers some estimates for these ratios were established and the connections between them were found. These results generalize some results from [5]. A series of numerical experiments (for one-dimensional and multidimensional knapsack problems) was also performed. All these experiments showed that the actual behavior of Δ and δ was much better than it could be expected from theoretical estimates. For example, for randomly generated one-dimensional problems with 3000 variables the average (over 10 instances) value of Δ was 1.000002, while for the third estimate from (9) it was 1.000780 (which is also more than satisfactory). For these problems the average value of δ was 1.000239. Consequently, both approaches mentioned above can be used for finding good approximate solutions (which can be subsequently used for finding optimal solutions).

1. One-dimensional Boolean knapsack. It consists in finding

$$f^* = \max \sum_{j=1}^{n} c_j x_j \tag{1}$$

subject to

$$\sum_{j=1}^{n} a_j x_j \leq b, \tag{2}$$

$$x_j \in \{0, 1\}, \ \ j = 1, 2, \ldots, n. \tag{3}$$

Here all a_j, c_j and b are positive. The coefficients in (2) are such that

$$a_j \leq b, \ \ j = 1, 2, \ldots, n, \tag{4}$$

$$\sum_{j=1}^{n} a_j > b. \tag{5}$$

The variables x_j are numbered in a non-increasing order of the ratios c_j/a_j, i.e.

$$\frac{c_1}{a_1} \geq \frac{c_2}{a_2} \geq \cdots \geq \frac{c_n}{a_n}. \tag{6}$$

The condition (6) is called the regularity condition. We denote $c_{\max} = \max c_j$, $c_{\min} = \min c_j$, $j = 1, 2, \cdots, n$. (1) and (4) imply that $f^* \geq c_{\max}$. The linear relaxation of the problem (1) – (3) is the linear programming problem defined by (1), (2) and

$$0 \leq x_j \leq 1, \ \ j = 1, 2, \ldots n. \tag{7}$$

Let the optimal solution of (1), (2), (7) be $x^{LR} = (x_1^{LR}, x_2^{LR}, \cdots, x_n^{LR})$ and its optimal value f^{LR}. If (6) is satisfied then x^{LR} can be found explicitly (the theorem of Dantzig, cf. [2]). We find the critical index s from the inequalities $\sum_{j=1}^{s-1} a_j \leq b < \sum_{j=1}^{s} a_j$ and let $x_k^{LR} = 1, k = 1, \ldots s-1, x_k^{LR} = 0, k = s+1, \ldots, n$ and $x_s = (b - \sum_{j=1}^{s} a_j)/a_s$.

It is evident that

$$f^* \le f^{LR}, \tag{8}$$

that is, f^{LR} yields an upper bound for the optimal value. A natural measure of the quality of this bound is the ratio $\Delta = f^{LR}/f^*$. It turns out that this ratio can be in general arbitrarily close to 2 (cf. [1]). In [4] the following estimates for Δ were obtained:

$$\Delta < 1+\frac{c_{\max}}{f^*}; \quad \Delta = 1+O\left(\frac{1}{n}\right); \quad \Delta < 1+\frac{c_{\max}}{sc_{\min}}. \tag{9}$$

Here s is the critical index.

Now we consider primal greedy methods for the problem (1) – (3). The greedy method starts with a feasible solution $x = (0, 0, \dots, 0)$ and consecutively replaces zeroes by ones in the order defined by (6) if each such replacement retains feasibility. Let $x^G = (x_1^G, x_2^G, \dots, x_n^G)$ be the last feasible solution, f^G – the corresponding objective function value. More formally, the greedy solution x^G is obtained in the following way. We set $x_1^G = 1$ (it is possible in view of (4)) and for $k = 2, \dots, n$

$$x_k^G = \begin{cases} 1, & \sum_{j=1}^{k-1} a_j x_j^G + a_k \le b, \\ 0, & \sum_{j=1}^{k-1} a_j x_j^G + a_k > b. \end{cases} \tag{10}$$

It is clear that $f^G \le f^* \le f^{LR}$. We suppose that the greedy value satisfies

$$f^G \ge c_{\max}. \tag{11}$$

Now we demonstrate some estimates for the ratio $\delta = f^*/f^G$. In [6] the following conditions were obtained which guarantee that $\delta \le 2$:

1) if $c_{\max}/c_{\min} \le 2$, $f^G \ge sc_{\min}$ and the regularity condition (6) holds then

$$\delta = \frac{f^*}{f^G} \le 2; \tag{12}$$

2) if the conditions (6) and (10) are satisfied then the inequality (12) holds;

3) $\delta = 1 + O(1/n)$.

In [6] the relations between Δ and δ were also considered. In particular, it was proved that for every problem (1) – (3) the validity of (6) and (11) implies the inequality $\Delta < 1 + \frac{1}{\delta}$.

2. One-dimensional integer knapsack. This problem is the problem (1) – (3) where (3) is replaced by

$$x_j \quad \text{are non-negative integers}, \quad j = 1, 2, \dots, n. \tag{13}$$

The validity of the regularity condition (6) is supposed for this formulation too. The linear relaxation of this problem is determined by (1), (2) and the condition

$$x_j \ge 0, \quad j = 1, 2, \dots, n. \tag{14}$$

The optimal solution x^{LR} of the linear relaxation can be found by a straightforward generalization of the theorem of Dantzig (cf. above).

It was shown in [6] that $1 \le \Delta < 1 + \frac{c_{\max}}{f^*} < 2$. In other words, the estimate (9) holds also for the one-dimensional integer knapsack.

3. Multidimensional Boolean knapsack. This problem consists in finding

$$f^* = \max \sum_{j=1}^{n} c_j x_j \tag{15}$$

subject to

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n, \tag{16}$$

$$x_j \in \{0, 1\}, \quad j = 1, 2, \dots, n. \tag{17}$$

Here all a_{ij}, c_j and b are positive, $a_{ij} \le b_j$. Besides, we suppose that $m < n$ and $\sum_{j=1}^{n} a_{ij} > b_i, i = 1, 2, \dots, m$.

In the linear relaxation of this problem the conditions (17) are replaced by

$$0 \le x_j \le 1, \quad j = 1, 2, \dots, n. \tag{18}$$

The optimal solution of the linear relaxation we denote again by x^{LR} and the corresponding optimal value by f^{LR}. We recall (cf. [1]) that x^{LR} contains at most m components with a non-zero fractional part.

We consider now the estimates $\Delta = f^{LR}/f^*$ and $\delta = f^*/f^G$ for the problem (15) – (17). In [6] it was shown that the inequalities $1 \le \Delta < m + 1$ hold. For $m = 1$ we get the corresponding inequalities for the one-dimensional problem. If we impose an additional assumption that every set of m items can be put in the knapsack (that is, each Boolean vector with m components equal to one is feasible) then $1 \le \Delta < 2$.

Let x be a feasible vector obtained by replacing all fractional components of x^{LR} by zeroes, f^x – the corresponding objective function value. In the greedy solution x^G the components are set to 1 according to the sequence defined by the inequalities $c_{j_1} \ge c_{j_2} \ge \dots \ge c_{j_n}$. It was shown in [6] that if $f^G \ge f^x$ then

$$1 \le \delta = \frac{f^*}{f^G} \le m + 1. \tag{19}$$

As we mentioned in the Introduction, for the one-dimensional and multidimensional knapsack problems a series of numerical experiments was performed. These experiments suggest (and in a certain sense confirm) the following hypothesis: $\Delta = \Delta_n = 1 + \alpha_n$, where α_n is positive and decreasing with the growth of n. It is reasonable to suppose that $\alpha_n \to 0$ when $n \to \infty$, and analogously for δ. In other words, this means that these algorithms are in a certain sense asymptotically optimal. Therefore the use of greedy methods can be recommended for practical computation, especially for large-scale problems.

4. The average behavior of greedy algorithms. Another justification for the recommended use of greedy algorithms is given by the results about their average behavior. This analysis concerns simultaneously primal (cf. (10)) and dual greedy algorithms. Informally, the dual greedy solution x^{DG} for the problem (1) – (3) is the optimal solution x^{LR} of the linear relaxation in which the fractional component is replaced by zero. Dual greedy solutions are in general not better than the primal ones. The analysis of the average behavior requires some probabilistic structure on the set of data. We shall suppose that (cf. [7])

1) the coefficients $c_j, a_j, j = 1, 2, \ldots, n$ are independent random variables uniformly distributed on [0, 1],

2) the right-hand side b is proportional to the number of variables: $b = \lambda n$ where $0 < \lambda < 1$.

Thus all objective function values become random variables too. Let A_n be an approximate algorithm for the problem with n variables, f^{A_n} – the objective function value for A_n. We say that A_n has an asymptotic tolerance $t > 0$ if $\mathbf{P}(f^* - f^{A_n} \leq t) \to 1$ when $n \to \infty$. It has been proved (cf. [7] and the references therein) that if $\lambda > \frac{1}{2} - \frac{t}{3}$ then both the primal and the dual greedy algorithms for (1) – (3) have asymptotic tolerance t. We call $1/2 - t/3$ the critical value and denote it by λ_0. Thus, for all $\lambda > \lambda_0$ both greedy methods are in a certain sense asymptotically good.

A series of numerical experiments was performed (their results are summarized in [7]). The goal was the comparison of the behavior of primal and dual greedy methods in dependence on the number of variables n and on the values of $\lambda < \lambda_0$. For this purpose a program was developed which generated and solved series of N instances. In this program the approximate objective function values were compared not with the optimal value f^* (which is difficult to find) but with its upper bound f^{LR} which can be computed with complexity $O(n \log n)$. The tolerance t varied from 0.01 to 0.03, the sample size N – from 100 to 500. Several dozens of instances were solved. The results were very similar. For all values of λ the values of f^{DG}, f^G, f^{LR} differed insignificantly. For relatively small λ the growth of the objective functions when λ increased was very rapid, and this growth delayed when λ was approaching its critical value. This empirical fact needs a theoretical explanation. We stress once more that we took $f^{LR}/f^G = \Delta\delta$ as the deviation measure. E.g., for an instance with $n = 3700, t = 0.01, N = 500$ the values of this measure were 1.0000098 for $\lambda = 0.30$ and 1.0000047 for $\lambda = 0.45$. This means that the actual behavior of greedy methods is still better. This is another confirmation of the recommendations we made above.

The generalization of this approach for the case of arbitrary distributions is presented in [3]. No experiments were performed.

Acknowledgements. This work was supported by RFBR grant 11-07-00360-a.

References

[1] H. Kellerer, U. Pferschy and D. Pisinger, *Knapsack problems.* Springer Verlag, 2004.

[2] I.Kh. Sigal and A.P. Ivanova, *Introduction to applied duscrete optimization: Models and computational algorithms, 2nd ed.* (In Russian) Fizmatlit, 2007.

[3] G. Diubin and A. Korbut, *The average behaviour of greedy algorithms for the knapsack problem: General distributions.* // Math. Meth. Oper. Res. 2003. V. 57. — P. 449 – 479.

[4] N.N. Galim'yanova, A.A. Korbut and I.Kh. Sigal. *Ratios of optimal values of objective functions of the knapsack problem and its linear relaxation.* // J. Comput. and Systems Sci. Int. 2009. V.48. — P. 906 – 913.

[5] I.Kh. Sigal, *Estimates of the deviation of approximate solutions from an optimal solution in certain problems of discrete optimization.*// J. Comput. and Systems Sci. Int. 2004. V. 43. — P. 94 – 103.

[6] A.A. Korbut and I.Kh. Sigal. *Exact and greedy solutions of the knapsack problem: the ratio of values of objective functions.* // J. Comput. and Systems Sci. Int. 2010. V. 49. — P. 757 – 764.

[7] G.N. Diubin, A.A. Korbut, I.Kh. Sigal. *The average behaviour of greedy algorithms for the knapsack problem: A computational experiment.* (In Russian) // In: Economico-matematicheskie issledovania. Matematicheskie modeli i informatsionnye tehnologii. Vol III. Nauka, 2003. — P. 46 – 54.