A total-value greedy heuristic for the integer knapsack problem

Rajeev Kohli

Graduate School of Business, Columbia University, New York, NY 10027, USA

Ramesh Krishnamurti

School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

This paper examines a new greedy heuristic for the integer knapsack problem. The proposed heuristic selects items in non-increasing order of their maximum possible contribution to the solution value given the available knapsack capacity at each step. The lower bound on the performance ratio for this "total-value" greedy heuristic is shown to dominate the corresponding lower bound for the density-ordered greedy heuristic.

knapsack problem; heuristic; approximation algorithms; worst-case performance

1. Introduction

Approximate solutions to the integer knapsack problem can be identified using a number of procedures. The simplest of these is the greedy heuristic. Given an ordering of the items, the greedy heuristic places as many integral units of each new item into the knapsack as will fit in the remaining capacity. The heuristic can select arbitrarily-bad solutions if the items are placed in non-increasing order of their values or in non-decreasing order of their weights, but has a better performance if the items are placed in non-increasing order of their densities. If k units of the heaviest available item can fit into the knapsack, the density-ordered greedy heuristic has a worst-case solution value that is $k/(k+1)$ of the optimal solution value (Fisher [1]). Besides providing an approximate solution, the density-ordered greedy heuristic is used for designing approximation schemes for the integer knapsack problem. These schemes use the heuristic solution value to bound the value of the optimal, the time complexity of the algorithm depending upon the worst-case bound provided by the heuristic.

This paper examines the worst-case performance of the greedy heuristic if it selects items in non-increasing order of their maximum possible contribution to the solution value given the available knapsack capacity at each step. The lower bound on the performance ratio for this total-value greedy heuristic is shown to dominate the corresponding bound for the density-ordered greedy heuristic for all values of k. [1] In particular, the worst-case bound for the total-value greedy heuristic is given by $1/\sum_{i=1}^{\infty} 1/h(i)$, where $h(i)$ is an integer value given by the recursion

[1] The value of k can be obtained in $O(n)$ time for any problem. The computation of the worst-case performance ratio for both the density-ordered and total-value greedy heuristics considers the choice of only the first greedy item. Hence the worst-case performance bound for both heuristics is guaranteed in $O(n)$ time.

Correspondence to: Rajeev Kohli, Graduate School of Business, Columbia University, New York, NY 10027, USA.

Published in: *Operations Research Letters*

$h(1) = 1$, $h(2) = k + 1$, $h(i) = [h(i-1)] \cdot [h(i-1) + 1]$ for $i \geq 3$. For $k = 1$, the above bound is identical to the bound conjectured by Johnson [3] for the Any-Fit-Increasing heuristic for the bin-packing problem and has a value of 0.5913, as compared with the value of 0.5 for the density-ordered greedy heuristic. Also, since a total-value ordering of the items improves the bound on the optimal provided by the greedy heuristic, using this ordering can improve the complexity of approximation schemes that currently use bounds based on a density ordering of the items. Thus, for example, it can be verified that using a total-value ordering of the items can reduce the constant (but not the order) of the time complexity of Lawler's [4] extension of Ibbara and Kim's [2] approximation scheme for the integer knapsack problem.

2. The total-value greedy heuristic

Without loss of generality, let $C = 1$ denote the knapsack capacity. Let Problem P denote the integer knapsack problem defined over n items, where item i has unit weight w_i and unit value v_i. The optimal solution to Problem P consists of $x_i \geq 0$ integral units of item i that maximize $\sum_i v_i x_i$ without violating the capacity constraint $\sum_i w_i x_i \leq C$. Let Z denote the value of the optimal solution to Problem P.

At step j, the total-value greedy heuristic selects an item for which the value $v_i \lfloor C_j / w_i \rfloor$ across all items i is maximum, where C_j is the available knapsack capacity at step j. Let $n_j \geq 1$ units of item j be selected at step j of the heuristic (if $n_j = 0$ for some step j, ignore the item and the step and re-index the remaining items and steps). Let $V_j = n_j v_j$ denote the total value contributed by item j to the greedy solution. Without loss of generality, let $V_1 = n_1 v_1 = 1$ be the contribution of the first item selected by the total-value greedy heuristic. Then $Z_t \geq 1$, where Z_t denotes the solution value for the total-value heuristic.

Let $W_j = n_j w_j$ denote the capacity occupied by item j. Then $W_j > \frac{1}{2} C_j > C_{j+1}$, else at least one more unit of item j would have been selected by the heuristic at step j. Hence, $Z \leq 2$ because the maximum total value in capacity C_2 can be no greater than the total value in capacity $C_1 = 1$. Thus, the total-value heuristic performs at least as well as the density-ordered heuristic when $k = 1$, for which the optimal solution value is no greater than twice the heuristic solution value. Below, we show that the worst-case bound for the total-value heuristic is significantly better than that for the density-ordered heuristic.

Let $k_i = \lfloor 1/w_i \rfloor$ denote the maximum number of units of item i that fit in the unit capacity of the knapsack, $i = 1, 2, \ldots, n$. Thus, $k = \min_i k_i$. Let Problem P′ denote an integer knapsack problem with capacity $C' = 1$. Let item l of Problem P′ have unit weight $w_l = 1/(k_i + 1) + \varepsilon$ and unit value $v_l = 1/k_i$, where $\varepsilon > 0$ is an arbitrarily small number and w_i and v_i are the unit weight and unit value, respectively, of item i in Problem P, i, $l = 1, 2, \ldots, n$. Let Z' denote the optimal solution value for Problem P′. Lemma 1 shows that $Z \leq Z'$. Lemma 2 then constructs a relaxation P* of P′ by permitting additional items. The optimal solution value for Problem P* will be shown to be $Z^* < \sum_{i=1}^{\infty} 1/h(i)$, where $h(1) = 1$, $h(2) = k + 1$, $h(i) = [h(i-1)] \cdot [h(i-1) + 1]$ for $i \geq 3$ and integer. As $Z \leq Z' \leq Z^*$, a lower bound on the worst-case performance ratio for the total-value heuristic is

$$r(0) \geq \frac{Z_t}{Z^*} > \frac{1}{\sum_{i=1}^{\infty} \frac{1}{h(i)}}.$$

The bound will be shown to be asymptotically tight with a a value strictly greater than 0.5913.

Lemma 1. $Z \leq Z'$.

Proof. Consider item l of Problem P' with unit weight $w_l = 1/(k_i + 1) + \varepsilon$ and unit value $v_l = 1/k_i$. As $k_i = \lfloor 1/w_i \rfloor$ is integer, $k_i + 1 > 1/w_i$. Thus, $w_i \geq 1/(k_i + 1) + \varepsilon = w_l$, where ε is an arbitrarily small positive number. Also, $k_i v_i \leq 1$ implies $v_i \leq 1/k_i = v_l$. It follows that if item i appears in the optimal solution to Problem P, it can be replaced by item l without decreasing the optimal solution value or violating the capacity constraint of the knapsack. Thus $Z \leq Z'$. □

Let i^* be the smallest integer such that $h(i^*) = N \geq \max\{n, \max_i k_i\}$. Consider the following

Problem P*:

$$Z^* = \max \sum_{i=k}^{N} \frac{1}{i} y_i$$

$$\sum_{i=k}^{N} \left(\frac{1}{i+1} + \varepsilon \right) y_i \leq 1$$

$y_i \geq 0$ and integer for all i, $k \leq i \leq N$.

Note that $Z \leq Z^*$ because Problem P* contains at least all n items of Problem P′. Note also that (for expository convenience) the items are indexed k through N in Problem P*. Item $i = k$ is the heaviest available item and item $i = N$ is the lightest available item.

Theorem 2. $r > 1/\sum_{i=1}^{\infty} h(i)$.

Proof. We prove the theorem by showing that $y_k = k$ and $y_{h(i)} = 1$, for $2 \leq i \leq i^*$ in the optimal solution to Z^*. Consequently, the optimal solution value for P* is

$$Z^* = k \cdot \frac{1}{k} + \sum_{i=2}^{i^*} 1 \cdot \frac{1}{h(i)} < \sum_{i=1}^{\infty} \frac{1}{h(i)},$$

and hence

$$r = \frac{Z_t}{Z} \geq \frac{Z_t}{Z^*} > \frac{1}{\sum_{i=1}^{\infty} \frac{1}{h(i)}}.$$

We show first that $y_k = k$ in the optimal solution to Z^*, then use an inductive argument to show that $y_{h(i)} = 1$ for $2 \leq i \leq i^*$.

(i) $y_k = k$.

As item k has unit weight $w_k = 1/(k+1) + \varepsilon$, the number of units of item k that can fit the knapsack is $y_k \leq k$. Consider the integer solution

$y_k = k, \quad y_{k+1} = 1, \quad y_{(k+1)(k+2)} = 1$ and

$y_i = 0$ for all other i.

To show that the solution is feasible, observe that the total occupied weight corresponding to the solution is

$$k \cdot \left(\frac{1}{k+1} + \varepsilon \right) + 1 \cdot \left(\frac{1}{k+2} + \varepsilon \right) + 1 \cdot \left(\frac{1}{(k+1)(k+2)+1} + \varepsilon \right)$$
$$= 1 - \left(\frac{1}{(k+1)(k+2)} - \frac{1}{(k+1)(k+2)+1} - (k+2)\varepsilon \right),$$

which for sufficiently small ε is less than 1, the knapsack capacity. The corresponding lower bound on the value of Z^* for $y_k = k$ is

$$k \cdot \left(\frac{1}{k} \right) + 1 \cdot \left(\frac{1}{k+1} \right) + 1 \cdot \left(\frac{1}{(k+1)(k+2)} \right)$$
$$= \frac{k+2}{k+1} + \frac{1}{(k+1)(k+2)}.$$

Consider $y_k = j$, $j < k$, so that the corresponding j units, each with unit value $1/k$, contribute a total value of j/k. Since each such item has unit weight $1/(k+1) + \varepsilon$, the remaining knapsack capacity is $1 - j/(k+1) - j\varepsilon$. The densest feasible item is $k+1$ with unit weight $1/(k+2) + \varepsilon$ and unit value $1/(k+1)$. Thus, an upper bound on the value of the optimal solution if $y_k = j$ is

$$\frac{j}{k} + \frac{1}{k+1} \frac{1 - \frac{j}{k+1} - j\varepsilon}{\frac{1}{k+2} + \varepsilon}$$
$$< \frac{j}{k} + \frac{1}{k+1} \frac{(k+1-j)}{k+1} (k+2)$$
$$= \frac{k+2}{k+1} + \frac{j}{k(k+1)^2}$$
$$< \frac{k+2}{k+1} + \frac{1}{(k+1)(k+2)}.$$

As the lower bound on Z^* when $y_k = k$ exceeds its upper bound when $y_k < k$, $y_k = k$ in the optimal solution to P*.

Observe that after k units of item k are chosen by the optimal, the remaining knapsack capacity in Problem P* is $1/(k+1) - k\varepsilon$. Also, after k units of item k have been chosen, the heaviest item that can fit in the remaining capacity is item $k+1$ with weight $1/(k+2) + \varepsilon$ and unit value $1/(k+1)$.

(ii) $y_{h(i)} = 1$, $2 \leq i \leq i^*$.

We prove the result by induction on the value of i, $2 \leq i \leq i^*$.

Base case: For $i = 2$, to show that $y_{h(2)} = 1$.

As the weight of item $h(2)$ is $1/(k+2) + \varepsilon$, $y_{h(2)}$ cannot exceed 1. Consider the feasible integer solution

$y_k = k, \quad y_{h(2)} = y_{k+1} = 1,$

$y_{h(3)} = y_{(k+1)(k+2)} = 1,$

$y_i = 0$, otherwise.

As shown in (i) above, this solution is feasible. Thus, if $y_{h(2)} = 1$, a lower bound on the value of Z^* is

$$1 + \left(\frac{1}{k+1}\right) + 1 \cdot \left(\frac{1}{(k+1)(k+2)}\right)$$
$$= 1 + \frac{k+3}{(k+1)\cdot(k+2)}.$$

If $y_{h(2)} = 0$, then item $h(2)+1$ is the densest among the remaining feasible items. As item $h(2)+1$ has unit weight

$$\frac{1}{h(2)+2} + \varepsilon = \frac{1}{k+3} + \varepsilon$$

and unit value

$$\frac{1}{h(2)+1} = \frac{1}{k+2},$$

its density is strictly less than $(k+3)/(k+2)$. Thus, the value of Z^* is strictly less than the solution value obtained by assuming that the remaining capacity (after k units of item 1 are included) is filled at a density rate of $(k+3)/(k+2)$; i.e., if $y_{h(2)} = 0$, an upper bound on the value of Z^* is

$$1 + \left(\frac{1}{k+1} - k\varepsilon\right) \cdot \frac{k+3}{k+2},$$

which is strictly less than

$$1 + \frac{k+3}{(k+1)\cdot(k+2)},$$

the lower bound on Z^* when $y_{h(2)} = 1$. Hence $y_{h(2)} = 1$ in the optimal solution to Problem P*.

Induction hypothesis. Assume $y_{h(i)} = 1$, $2 \le i \le l$, for some $l \le i^* - 1$, in the optimal solution to Problem P*.

Induction step. To show that $y_{h(i)} = 1$, $2 \le i \le l+1$, in the optimal solution to Problem P*.

Case 1. $1 \le l \le i^* - 2$.

By the induction hypothesis, $y_{h(i)} = 1$, $2 \le i \le l$, in the optimal solution to Problem P*. To show that $y_{h(l+1)} = 1$, consider the integer solution

$$y_k = k,\ y_{h(i)} = 1 \quad \text{for } 2 \le i \le l+2, \text{ and}$$

$$y_i = 0 \text{ otherwise},$$

which provides a lower bound of $\sum_{i=1}^{l+2} 1/h(i)$ on the optimal solution to Problem P* when $y_{h(l+1)} = 1$. To show that this solution is feasible, observe that the occupied capacity of the items in this solution is

$$k \cdot \left(\frac{1}{k+1} + \varepsilon\right) + \sum_{i=2}^{l+2} \left(\frac{1}{h(i)+1} + \varepsilon\right)$$
$$= 1 - 1 + k \cdot \left(\frac{1}{k+1} + \varepsilon\right)$$
$$+ \sum_{i=2}^{l+2} \left(\frac{1}{h(i)+1} + \varepsilon\right)$$
$$= 1 - \frac{1}{k+1} + k\varepsilon + \sum_{i=2}^{l+2} \left(\frac{1}{h(i)+1} + \varepsilon\right).$$

Since

$$\frac{1}{h(2)} = \frac{1}{k+1} \quad \text{and}$$

$$-\frac{1}{h(i)} + \frac{1}{h(i)+1} = -\frac{1}{h(i+1)},$$

the above expression for the occupied capacity can be written as

$$1 - \frac{1}{h(l+3)} + (k+l+1)\varepsilon,$$

which for sufficiently small ε is less then 1, the knapsack capacity. Using a similar argument, it can be verified that $y_{h(l+1)} = 2$ is not feasible, if, as assumed in the induction hypothesis, $y_{h(i)} = 1$ for all i, $2 \le i \le l$.

Now consider $y_{h(l+1)} = 0$. The remaining knapsack capacity is $1/(h(l+1)) - (k+l-1)\varepsilon$ after k units of item k and one unit each of items $h(i)$, $2 \le i \le l$, are placed in the knapsack. Also, items k and $h(i)$, $2 \le i \le l$, contribute a value of $\sum_{i=1}^{l} 1/h(i)$ to Z^*. As $y_{h(l+1)} = 0$, and as items $h(l)+1, h(l)+2, \ldots, h(l+1)-1$, are not feasible, the densest feasible item that can fit in the remaining capacity is item $h(l+1)+1$, which has unit weight $1/(h(l+1)+2)+\varepsilon$, and unit value $1/(h(l+1)+1)$. Thus, an upper bound on the value of the optimal solution when $y_{h(l+1)} = 0$ is

$$Z < \sum_{i=1}^{l} \frac{1}{h(i)} + \left(\frac{1}{h(l+1)} - (k+l-1)\varepsilon \right) \cdot \left(\frac{\frac{1}{h(l+1)+1}}{\frac{1}{h(l+1)+2} + \varepsilon} \right)$$

$$< \sum_{i=1}^{l} \frac{1}{h(i)} + \frac{1}{h(l+1)} \frac{h(l+1)+1+1}{h(l+1)+1}$$

$$= \sum_{i=1}^{l} \frac{1}{h(i)} + \frac{1}{h(l+1)} + \frac{1}{h(l+2)}$$

$$= \sum_{i=1}^{l+2} \frac{1}{h(i)}.$$

The right hand side above is the lower bound on Z^* when $y_{h(l+1)} = 1$. Hence $y_{h(l+1)} = 1$ in the optimal solution to P*.

Case 2. $l = i^* - 1$.

Since

$$\frac{1}{h(i)} - \frac{1}{h(i)+1} = \frac{1}{h(i+1)},$$

the available capacity if $y_k = k$ and $y_{h(i)} = 1$ for $2 \le i \le l$ is

$$\frac{1}{k+1} - k\varepsilon - \sum_{i=1}^{l} \left(\frac{1}{h(i)+1} + \varepsilon \right)$$

$$= \frac{1}{h(l+1)} - (k+l-1)\varepsilon$$

$$= \frac{1}{N} - (k+l-1)\varepsilon.$$

Thus, only one unit of the smallest available item (i.e, item N) with unit weight $1/(N+1) + \varepsilon$ can fit in the available knapsack capacity. Hence $y_{h(l+1)} = y_N = 1$ in the optimal solution to P*. □

The example in Table 1 shows that the bound derived in Theorem 2 is achieved asymptotically as n tends to infinity. At the first step of the total-value heuristic, each item can contribute a total value of 1 to the unit knapsack capacity. Let the greedy select $h(n)$ units of item n. No other item is feasible and the greedy heuristic terminates after one step. The optimal solution consists of k units of item 1 and one unit each of items 2 through n. Item 1 occupies just over $k/(k+1)$ of the knapsack capacity, which is the minimum that must be occupied by an item that contributes the same solution value $(=1)$ as the

Table 1
Worst-case example for total-value greedy heuristic

Item	Weight	Value	Total Value (Step 1)
1	$\frac{1}{k+1} + \varepsilon$	$\frac{1}{k}$	1
2	$\frac{1}{h(2)+1} + \varepsilon = \frac{1}{k+2} + \varepsilon$	$\frac{1}{h(2)} = \frac{1}{k+1}$	1
⋮			
i	$\frac{1}{h(i)+1} + \varepsilon$	$\frac{1}{h(i)}$	1
⋮			
n−2	$\frac{1}{h(n-2)+1} + \varepsilon$	$\frac{1}{h(n-2)}$	1
n−1	$\frac{1}{h(n-1)+1} + \varepsilon$	$\frac{1}{h(n-1)}$	1
n	$\frac{1}{h(n)+1} + \varepsilon$	$\frac{1}{h(n)}$	1

Table 2
Bounds for density and total-value greedy heuristic

k	r (total-value)	r (density-ordered)
1	0.5913555	0.5000000
2	0.7026825	0.6666667
3	0.7678212	0.7500000
4	0.8101038	0.8000000
5	0.8396093	0.8333333

item selected by the total-value greedy heuristic. Each subsequent item in the optimal solution is selected to have the largest possible unit value and the smallest possible size that it can without violating the constraint that its total contribution *at the first greedy step* is no larger than 1. For example, given that the knapsack contains k units of item 1, less than $1/(k+1)$ of the total capacity is available for other items. Thus, every other optimal item must be small enough to have at least $k+1$ of its units fit into the knapsack at the first step of the greedy heuristic, and therefore have a unit value of at most $1/(k+1)$. Item 2 obtains this unit value and has the smallest possible weight (just in excess of $1/(k+2)$ of the knapsack capacity) for exactly $k+1$ units of the item to be feasible at the first step of the greedy heuristic. The subsequent items are generated in a similar fashion.

Table 2 presents the lower bound for the total-value greedy heuristic (accurate up to the first six decimal places) for $1 \leq k \leq 5$ and compares it to the lower bound $k/(k+1)$ for the density-ordered greedy heuristic. [2]

The density-ordered greedy heuristic selects items in non-increasing order of their densities v_i/w_i, $1 \leq i \leq n$. If we relax the integrality constraint (i.e., allow fractional amounts of items to be included in the knapsack), then the ordering obtained by the total-value greedy heuristic is the same as the ordering obtained by the density-ordered greedy heuristic, since $\lfloor 1/w_i \rfloor v_i = v_i/w_i$. Thus, the density-ordered greedy heuristic behaves like the total-value greedy heuristic with the integrality constraint removed. The effect of increasing k is to reduce the difference between $\lfloor 1/w_i \rfloor$ and $1/w_i$ for $1 \leq i \leq n$. Thus, as k increases, the ordering due to the total-value greedy heuristic tends to the ordering due to the density-ordered heuristic, and the difference between the two worst-case bounds decreases. Observe from Table 2 that the difference between the worst-case bound for the total-value and the density-ordered greedy heuristics is the largest for $k=1$. This difference decreases as k increases. As k approaches infinity, both heuristics obtain the optimal solution.

3. Conclusion

The principal benefit of the proposed total-value greedy heuristic appears to be that it considers all three problem parameters – unit weight, unit value, and knapsack capacity – in ordering items. A consequence of considering all three is that the total-value greedy heuristic dominates the worst-case performance of the density-ordered greedy heuristic for the integer knapsack problem.

There are certain conditions under which the density-ordered greedy heuristic identifies the optimal solution to the integer knapsack problem (see, e.g., Magazine, Nemhauser, and Trotter [5]). Similar conditions for the total-value greedy heuristic may be worth investigating. It may also be useful to examine the joint performance of the density-ordered and total-value greedy heuristics, and to consider the conditions under which the combination of the two heuristics identifies the optimal solution.

Acknowledgements

The authors thank Pavol Hell, Art Liestman, Sanjeev Mahajan, Prakash Mirchandani and Joe Peters for numerous helpful suggestions that have helped improve the presentation. Support by the Natural Sciences and Engineering Research Council of Canada under grant OGP0036809 is gratefully acknowledged by the second author.

[2] The bound for the total-value greedy heuristic is computed by taking the first six values of the recursive integer function $h(i)$. If $k=1$, these values are $h(1)=1$, $h(2)=2$, $h(3)=6$, $h(4)=42$, $h(5)=1806$ and $h(6)=3\,263\,442$. The sixth term in the summation corresponds to $1/h(6)=\frac{1}{3263442}=0.000\,000\,306\,424\,934$. Similarly, if $k=2$, $h(1)=1$, $h(2)=3$, $h(3)=12$, $h(4)=156$, $h(5)=24492$, and $h(6)=599\,882\,556$. In both cases, including terms for $n>6$ influences the value of the expression $\Sigma_{i=1}^{n} 1/h(i)$ only beyond the seventh decimal place.

References

[1] M.L. Fisher, Worst-case analysis of heuristic algorithms", *Management Sci.* **26**, 1–17 (1980).

[2] O.H. Ibarra and C.E. Kim, "Fast approximation àlgorithms for the knapsack and sum of subsets problems", *J. Assoc. Comput. Mach.* **22**, 463–468 (1975).

[3] D.S. Johnson, "Near-optimal bin packing algorithms", *Project MAC TR-109*, Ph.D. Dissertation Thesis, Massachusetts Institute of Technology (1973).

[4] E.L. Lawler, "Fast approximation algorithms for knapsack problems", *18th Annual Symposium on Foundations of Computer Science*, 1977, 206–213.

[5] M.J. Magazine, G.L. Nemhauser, and L.E. Trotter, Jr., "When the greedy solution solves a class of knapsack problems", *Oper. Res.* **23**, 207–217 (1975).