Average performance of greedy heuristics for the integer knapsack problem

Rajeev Kohli [a], Ramesh Krishnamurti [b,*], Prakash Mirchandani [c]

[a] Graduate School of Business, Columbia University, New York, NY 10027, USA
[b] School of Computing Science, Simon Fraser University, Burnaby, Canada, BC V5A 1S6
[c] Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, USA

Abstract

This paper derives a lower bound on the average performance of a total-value greedy heuristic for the integer knapsack problem. This heuristic selects items in order of their maximum possible contribution to the solution value at each stage. We show that, as for the worst-case bound, the average performance bound for the total-value heuristic dominates the corresponding bound for the density-ordered greedy heuristic.

Keywords: Heuristics; Combinatorial optimization; Integer programming; Knapsack; Probabilistic algorithm

1. Introduction

Given a set of items with corresponding unit values (i.e., unit profits) and unit weights, and a knapsack capacity limit, the integer knapsack problem selects integer units of those items that maximize the total value and for which the total weight does not exceed the knapsack capacity limit. Though simple to state, the integer knapsack problem is NP-hard; with real unit weights and unit values (as considered in this paper), this problem is strongly NP-hard. Moreover, the problem is of significant economic importance because it often arises as a sub-problem in practice while solving many large-scale integer programming problems. Consequently, this problem has attracted operations researchers and computer scientists who have extensively studied it and its variants from both theoretical and computational viewpoints. Solution procedures proposed for the problem include exact algorithms (e.g., dynamic programming, branch-and-bound) and heuristic procedures. (See, for example, [18,20] for comprehensive treatments of the knapsack problem, and [1] for an exact algorithm using dynamic programming for the integer knapsack problem.) The heuristic procedures for approximately solving the knapsack problem include the intuitively appealing density ordered heuristic (which picks the item with the highest unit value to unit weight ratio at each stage) and the total-value heuristic (which picks the item that contributes the highest total value given the remaining knapsack capacity at each stage). These heuristic procedures are

* Corresponding author.
E-mail address: ramesh@cs.sfu.ca (R. Krishnamurti).

Published in: ***European Journal of Operational Research***

particularly relevant for the knapsack problem with real unit weights and unit values. In this paper, we analyze and compare the average-case behavior of these two greedy heuristics for the integer knapsack problem.

Despite its intricate nature, probabilistic analysis has been used to study the performance of heuristics for a variety of other problems, including the k-median problem [10], the traveling salesman problem [5,13], bin packing and its variants [2,4,6,8,14,21], satisfiability [11,15], the quadratic-assignment problem [12], and multiprocessor list scheduling [3]. Although a variety of approaches can be used (see, e.g., [7]), most researchers assume independent, identically distributed observations from a specific density function. The analyses generally involve the calculation of rather complex conditional probabilities, which makes it difficult to obtain exact results. Furthermore, changing the assumptions for the problem data often significantly alters the analysis as well as the results.

The model used to analyse the average performance of heuristics in this paper was introduced in [15] for the maximum satisfiability problem, and further developed in [17] for the minimum satisfiability problem. The analysis does not require independence or specific statistical distributions for the problem data (i.e., the unit weights and unit values of the available items). To illustrate the proposed approach, consider using the density-ordered greedy heuristic [9] to solve a problem in which the knapsack has unit capacity and $n = 2$ items are available. The first item has weight 1/2 and value 1/2, and the second item has weight $(1/2) + \varepsilon$ and value $(1/2) + \delta$, where $\varepsilon, \delta > 0$ are arbitrarily small. Suppose δ is randomly generated, and is arbitrarily smaller than ε (or arbitrarily greater than ε) with probability p (or $(1 - p)$). Then the density of the first item equals 1, and the density of the second item is less than 1 when $\delta < \varepsilon$ and greater than 1 when $\delta > \varepsilon$. Thus, when $\delta < \varepsilon$, the density-ordered greedy heuristic selects two units of the first item and has solution value equal to 1. However, if $\delta > \varepsilon$, the density-ordered greedy heuristic selects one unit of the second item and has solution value equal to $(1/2) + \delta$. Regardless of the value of δ, the optimal solution consists of two units of the first item and has objective function value equal to 1. Thus, the performance ratio for the density-ordered greedy heuristic equals 1 with probability p and is arbitrarily close to 1/2 with probability $1 - p$. Hence, for this data-generating mechanism, the expected performance ratio for the density-ordered greedy heuristic can be made arbitrarily close to $p \cdot 1 + (1 - p) \cdot (1/2) = (1 + p)/2$, a function that increases linearly from 1/2 (the worst-case performance bound for the density-ordered heuristic) to 1 as p increases from 0 to 1.

In the above example, the greedy heuristic terminates in one step, and the only possible solutions are the worst-case and the optimal. More generally, one does not know the mechanism by which the problem data are generated, the number of steps before the greedy heuristic terminates, or the probability with which the greedy heuristic selects an optimal item at any step. However, for any data-generating mechanism (and hence for any set of problem instances), we define p to be the minimum probability with which the greedy heuristic selects an optimal item for a knapsack subproblem at any step of the greedy heuristic. We derive tight lower bounds on the expected performance ratios for the total-value [16] and density-ordered [9] greedy heuristics as a function of this probability value, and show that the lower bound on the expected performance ratio for the total-value greedy heuristic strictly dominates the lower bound on the expected performance ratio for the density-ordered greedy heuristic.

Although we leave open the question of theoretically deriving the value of this probability for a specific distribution, we demonstrate using illustrative examples that there exist distributions for which this probability can obtain every possible value. Consequently, the bounds obtained are tight in the sense that there exists at least one data-generating distribution that achieves the bound for any given value of this probability. We also conduct a computational study to develop empirical estimates of this probability value.

This paper is organized as follows. In Section 2, we introduce the notation and formally describe the density-ordered and total-value greedy heuristics. In Section 3, we derive the lower bound on the expected performance ratio for the total-value greedy heuristic. Section 4 analyzes the expected performance of the density-ordered greedy heuristic. Section 5 concludes the paper with some future research directions.

2. Notation and heuristic description

Let $0 < w_i \leqslant 1$ denote the unit weight and $v_i > 0$ the unit value of item i, $1 \leqslant i \leqslant n$. Let C denote the knapsack capacity; without loss of generality, we assume $C = 1$. The integer knapsack problem, denoted by P, is to select an integer number of units of item i for each i, $i = 1, 2, \ldots, n$, so that the total value of the selected items is maximized subject to the knapsack capacity constraint. Let Z denote the optimal solution value of Problem P.

Procedure GREEDY, described below, is a common description for both the total-value and the density-ordered greedy heuristics. Let C_j denote the knapsack capacity available at step j of the greedy heuristic, where $C_1 = C$. At step j (Line 6 of Procedure GREEDY), the total-value greedy heuristic selects as many units as are feasible of an item j that contributes the largest possible value in the available capacity. The density-ordered greedy heuristic, on the other hand, selects at step j as many units as possible of the densest remaining item. Without loss of generality, by reindexing items if necessary, we assume that $n_j \geqslant 1$ units of item j are selected at step j of the greedy heuristic.

Procedure GREEDY
1. **begin**
2. $\quad j := 1, C_1 := 1;$
3. $\quad$ **while** $j \leqslant n$ and $C_j \geqslant \min_{j \leqslant i \leqslant n} w_i$
4. $\quad$ **do**
5. $\quad\quad$ **begin**
6. $\quad\quad\quad$ select item with highest total value (density); reindex it item j;
7. $\quad\quad\quad n_j := \lfloor \frac{C_j}{w_j} \rfloor, C_{j+1} := C_j - n_j w_j;$
8. $\quad\quad\quad j := j + 1;$
9. $\quad\quad$ **end;**
10. **end.**

We detail below the notation used in the rest of the paper. $W_j = n_j w_j$ denotes the capacity occupied by item j. We note that $W_j > C_j/2$, otherwise the greedy heuristic would select at least one more unit of item j. $V_j = n_j v_j$ denotes the total value contributed by item j to the greedy solution. The number of steps in which the greedy heuristic terminates is denoted by $q \leqslant n$. P_j denotes the knapsack sub-problem arising at step j of the greedy heuristic, $1 \leqslant j \leqslant q$. Z_j denotes the value of the optimal solution to Problem P_j ($Z_1 = Z$ because Problem P is identical to Problem P_1). Also, p_j denotes the probability that the n_j units of item j selected by the greedy heuristic also appear in some optimal solution to Problem P_j; we call item j a greedy-optimal item. We let $p = \min_{1 \leqslant j \leqslant q} p_j$.

3. Total-value greedy heuristic

Let $r_{tv}(s)$ denote the ratio of the total-value heuristic solution value to the optimal solution value, given that the first $s \geqslant 0$ items selected by the heuristic are greedy-optimal. Note that $r_{tv}(q) = 1$ because there is no feasible item after the last step of the greedy heuristic. We derive the worst-case bound for $r_{tv}(s)$ for all s, $0 \leqslant s < q$. Using this bound, we derive a lower bound on $E[r_{tv}]$, the expected performance ratio for the total-value heuristic, as a function of p. We begin by stating the following theorem, which provides an upper bound on the optimal solution value in terms of the heuristic solution value [16].

Theorem 1. *If the remaining knapsack capacity at step j is at most C/k for some integer k, then an upper bound on the optimal solution value to Problem P_j is given by $\sum_{i=k}^{\infty} (1/h(i))V_1$, where $h(1) = 1$, $h(i) = [h(i-1)][h(i-1)+1]$ for $i \geqslant 2$ and integer.*

A consequence of Theorem 1 is that the worst-case performance ratio for the total-value greedy heuristic is $1/\sum_{i=1}^{\infty} 1/h(i) = 0.5913\ldots$ Table 1 presents a worst-case example (with $C = 1$) that achieves this bound.

Lemmas 2–5 are technical lemmas that are used to prove Theorem 6. Essentially, the first three of these lemmas develop bounds on the contributed

Table 1
Problem $T(0)$: Worst-case example for total-value greedy heuristic

Item number i	Unit weight w_i	Unit value v_i
1	$(1/2)+\varepsilon$	1
2	$(1/3)+\varepsilon$	1/2
3	$(1/7)+\varepsilon$	1/6
$\vdots$	$\vdots$	$\vdots$
n	$(1/(h(n)+1))+\varepsilon$	$1/h(n)$

The greedy heuristic solution consists of $h(n)$ units of item n. The optimal solution consists of one unit each of items 1 through n. The function $h(i)$ is recursively defined as $h(1)=1$, $h(i+1)=h(i)(h(i)+1)$.

values of the items chosen at each stage of the greedy heuristic, and the last one places an upper bound on the density of a feasible item at any step of the heuristic. Together, these lemmas allow us to derive a lower bound on the total-value heuristic solution value.

The following lemma can be easily shown.

Lemma 2. *If $W_i > 2C_{i+1}$, then $V_i \geqslant 3V_{i+1}$ for any i, $i = 1,2,\ldots,q-1$.*

Lemma 2 says that if the capacity occupied by the ith greedy item is at least twice the remaining capacity at the next step, then the contributed value of the ith greedy item is at least three times the contributed value of the $i+1$th greedy item. The following lemma develops a bound on the contributed value of the greedy-optimal items with respect to the contributed value of the first item selected by the greedy heuristic that is not in the optimal solution.

Lemma 3. *Suppose the first $s \geqslant 1$ items selected by the total-value heuristic are greedy-optimal. Then, $V_i \geqslant 2\cdot 3^{s-i}V_{s+1}$ for all i, $1 \leqslant i \leqslant s < n$.*

Proof. The lemma is proved by induction on the integer $t = s-i+1$ for $1 \leqslant i \leqslant s$.

Base case: To show that $V_i \geqslant 2\cdot 3^{s-i}V_{s+1}$ if $t = s-i+1 = 1$ (i.e., if $i = s$), or equivalently, $V_s \geqslant 2V_{s+1}$.

As item $s+1$ occupies no more than the available capacity at step $s+1$ of the total-value greedy heuristic, $C_{s+1} \geqslant W_{s+1}$. Also, $W_s > C_{s+1}$ since the total-value greedy heuristic would otherwise have selected at least one more unit of item s. Thus, $W_s > C_{s+1} \geqslant W_{s+1}$, which implies that the capacity available at step s of the total-value greedy heuristic is $C_s = W_s + C_{s+1} > 2C_{s+1} \geqslant 2W_{s+1}$. Consequently, at least $2n_{s+1}$ units of item $s+1$ can fit into the knapsack at step s, contributing a value no less than $2V_{s+1}$ to the greedy solution value. Since the greedy heuristic selects item s at step s, $V_s \geqslant 2V_{s+1}$.

Induction hypothesis: Assume $V_i \geqslant 2\cdot 3^{s-i}V_{s+1}$ for all t, $1 \leqslant t \leqslant l$ (i.e., for all $s \geqslant i \geqslant s-l+1$), where $l < s$ is integer.

Induction step. To prove that $V_i \geqslant 2\cdot 3^{s-i}V_{s+1}$ for $t = l+1$ (i.e., for $i = s-l$).

Since $W_{s-l} > C_{s-l+1}$, it follows that

$$C_{s-l} = W_{s-l} + C_{s-l+1} > 2C_{s-l+1} = 2W_{s-l+1} + 2C_{s-l+2}.$$

If $2C_{s-l+2} \geqslant W_{s-l+1}$, then $C_{s-l} > 3W_{s-l+1}$. Thus, $V_{s-l} \geqslant 3V_{s-l+1}$ and the inductive step follows from the inductive hypothesis. Else, $W_{s-l+1} > 2C_{s-l+2}$, which implies

$$C_{s-l} > 6C_{s-l+2} = 2\cdot 3W_{s-l+2} + 2\cdot 3C_{s-l+3}.$$

If $2C_{s-l+3} \geqslant W_{s-l+2}$, then $C_{s-l} > 3^2W_{s-l+2}$. Thus, $V_{s-l} \geqslant 3^2V_{s-l+2}$ and the inductive step follows. Else, $W_{s-l+2} > 2C_{s-l+3}$, which implies

$$C_{s-l} > 2\cdot 3^2W_{s-l+3} + 2\cdot 3^2C_{s-l+4}.$$

Proceeding in a similar fashion it can be shown that either the inductive step follows at an intermediate stage, or

$$C_{s-l} > 2\cdot 3^{l-1}W_s + 2\cdot 3^{l-1}C_{s+1}.$$

Consequently, if $2C_{s+1} \geqslant W_s$, then the inductive step follows. Alternatively,

$$C_{s-l} > 2\cdot 3^lC_{s+1} \geqslant 2\cdot 3^lW_{s+1}$$

and again, the inductive step follows since at least $2\cdot 3^ln_{s+1}$ units of item $s+1$ can fit into C_{s-l}, the available capacity when $t = l+1$. □

Lemma 3 developed a bound on the contributed value of the greedy-optimal items with respect to the contributed value of the first non-optimal item chosen by the total-value heuristic. In contrast, the following lemma develops a bound on the contributed value of the greedy-optimal items with respect to the contributed value of any feasible item at step l of the heuristic.

Lemma 4. *Let t denote a feasible item at step l of the greedy heuristic, $1 \leqslant l \leqslant q$. Let $W_{l,t} = \lfloor C_l/w_t \rfloor w_t$ and $V_{l,t} = \lfloor C_l/w_t \rfloor v_t$ respectively denote the occupied weight and contributed value of item t at step l. If $3W_{l,t} \leqslant 2C_l$, then item i selected by the greedy heuristic at step $i, 1 \leqslant i \leqslant l$, contributes a value $V_i \geqslant 3^{l-i}V_{l,t}$ to the greedy solution value.*

Proof. As $W_{l-1} > C_l$ and $C_{l-1} = W_{l-1} + C_l, C_{l-1} > 2C_l \geqslant 3W_{l,t}$. Thus, $V_{l-1} \geqslant 3V_{l,t}$. As in the proof of Lemma 3, an inductive argument can be used to complete the proof. □

The following lemma imposes an upper bound on the density of an item given the remaining capacity at any stage of the algorithm.

Lemma 5. *Let t denote a feasible item at step i of the greedy heuristic, and let item i be chosen by the heuristic at step $i, 1 \leqslant i \leqslant n$. If item t is such that at least k units of the item fit in capacity C_i (i.e., $\lfloor C_i/w_t \rfloor \geqslant k$), then the maximum density of item t is $((k+1)/k)(V_i/C_i)$.*

Proof. Let $\lfloor C_i/w_t \rfloor = k + j$, where $j \geqslant 0$ is an integer. Then $(C_i/(k+j)) \geqslant w_t > (C_i/(k+j+1))$. As no item at step i has total value greater than V_i,

$$v_t \left\lfloor \frac{C_i}{w_t} \right\rfloor = v_t(k+j) \leqslant V_i.$$

Thus, $v_t \leqslant (V_i/k+j)$ and the density of item t is given by

$$\frac{v_t}{w_t} < \frac{\frac{V_i}{k+j}}{\frac{C_i}{k+j+1}} = \frac{k+j+1}{k+j}\frac{V_i}{C_i} \leqslant \frac{k+1}{k}\frac{V_i}{C_i}. \qquad \square$$

Theorem 6 uses the bounds derived in Lemmas 2–5 to derive a lower bound on $r_{tv}(s)$ for $s \leqslant n-2$ and $q \geqslant s+1$. It can be verified that $r_{tv}(s) = 1$ if $s = n-1$ or $s = n$, or if the heuristic terminates in $q = s$ steps.

Theorem 6

$$r_{tv}(s) > \frac{\sum_{i=1}^{s} 3^{s-i+1} + 1}{\sum_{i=1}^{s} 3^{s-i+1} + \sum_{j=1}^{\infty} \frac{1}{h(j)}}$$

for $s \leqslant n-2$ and $q \geqslant s+1$,

where $h(1) = 1,\ h(i) = [h(i-1)] \cdot [h(i-1)+1]$ for $i \geqslant 2$ and integer.

Proof

Case 1: $W_i > 2C_{i+1}$ for all i, $i = 1, 2, \ldots, s$.

As $W_i > 2C_{i+1}$, $V_i \geqslant 3V_{i+1}$ by Lemma 2 for all $i = 1, 2, \ldots, s$. Thus, $V_i \geqslant 3^{s-i+1}V_{s+1}$ for all i, $i = 1, 2, \ldots, s$. It follows that

$$Z_t(s) = \sum_{i=1}^{q} V_i \geqslant \sum_{i=1}^{s+1} V_i = \sum_{i=1}^{s} 3^{s-i+1} m_i V_{s+1} + V_{s+1},$$

where $m_i \geqslant 1$ is a multiplier such that $V_i = 3^{s-i+1} m_i V_{s+1}$, and $m_i \geqslant m_{i+1}$ because $V_i \geqslant 3V_{i+1}$, $1 \leqslant i \leqslant s$.

Consider Problem P_{s+1} defined over capacity C_{s+1}. Since item $s+1$ selected by the greedy heuristic contributes a total value of V_{s+1} to Z_{s+1}, it follows from Theorem 1 that $Z_{s+1} \leqslant \sum_{j=1}^{\infty} (1/h(j))V_{s+1}$. Thus, the optimal solution to Problem P is bounded by

$$Z \leqslant \sum_{i=1}^{s} 3^{s-i+1} m_i V_{s+1} + \left(\sum_{j=1}^{\infty} \frac{1}{h(j)} \right) V_{s+1}.$$

It follows that

$$\begin{aligned} r_{tv}(s) &= \frac{Z_t(s)}{Z} \\ &\geqslant \frac{\sum_{i=1}^{s} 3^{s-i+1} m_i V_{s+1} + V_{s+1}}{\sum_{i=1}^{s} 3^{s-i+1} m_i V_{s+1} + \left(\sum_{j=1}^{\infty} \frac{1}{h(j)} \right) V_{s+1}} \\ &\geqslant \frac{\sum_{i=1}^{s} 3^{s-i+1} + 1}{\sum_{i=1}^{s} 3^{s-i+1} + \left(\sum_{j=1}^{\infty} \frac{1}{h(j)} \right)}, \end{aligned}$$

where the second inequality follows since $m_i \geqslant 1$, for $1 \leqslant i \leqslant s$.

Case 2: $W_i \leqslant 2C_{i+1}$ for some i, $i = 1, 2, \ldots, s$.

Consider the largest index l for which (i) $W_l \leqslant 2C_{l+1}$ and (ii) $W_i > 2C_{i+1}$ if $l+1 \leqslant i \leqslant s$. (If $W_s \leqslant 2C_{s+1}$, we set $l = s$.) Lemma 2 implies $V_i = 3^{s-i+1} m_i V_{s+1}$ for all $l+1 \leqslant i \leqslant s$, where

$m_i \geqslant m_{i+1} \geqslant 1$. Similarly, Lemma 4 implies $V_i \geqslant 3^{l-i} V_l$ for $1 \leqslant i \leqslant l-1$ and Lemma 3 implies $V_l \geqslant 2 \cdot 3^{s-l} V_{s+1}$. Thus, $V_i = 2 \cdot 3^{s-i} m_i V_{s+1}$ for $1 \leqslant i \leqslant l-1$, where $m_i \geqslant m_l \geqslant 1$ for $1 \leqslant i \leqslant l-1$. It follows that

$$Z_t(s) = \sum_{i=1}^{q} V_i \geqslant \sum_{i=1}^{s+1} V_i = \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + V_{s+1}.$$

Let i^* denote the largest item in the optimal solution for Problem P_{s+1}. Note that V_{i^*}, the total value of item i^* in capacity C_{s+1}, equals $\lfloor C_{s+1}/w_{i^*} \rfloor v_{i^*}$. To bound the value of the optimal solution, and hence obtain the lower bound of $r_{tv}(s)$, we consider the following sub-cases.

Case 2(a): $w_{i^*} > (2/3)C_{s+1}$.

As the weight of item i^* exceeds 2/3 of C_{s+1}, it follows from Lemma 5 that no other optimal item for Problem P_{s+1} can have density exceeding $4V_{s+1}/(3C_{s+1})$. Thus, the optimal solution value for Problem P_{s+1} is no greater than $V_{i^*} + (1/3)(4V_{s+1}/3)$. It follows that

$$Z \leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + V_{i^*} + \frac{4}{9} V_{s+1}.$$

As $V_{i^*} \leqslant V_{s+1}$,

$$Z \leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + \frac{13}{9} V_{s+1}.$$

Thus,

$$r_{tv}(s) \geqslant \frac{\sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i + \sum_{i=l+1}^{s} 3^{s-i+1} m_i + 1}{\sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i + \sum_{i=l+1}^{s} 3^{s-i+1} m_i + \frac{13}{9}}.$$

Multiplying the numerator and denominator by 3/2 and subtracting $(1/2) + (1/2)\sum_{i=l+1}^{s} 3^{s-i+1}$ from the numerator and denominator gives

$$r_{tv}(s) > \frac{\sum_{i=1}^{s} 3^{s-i+1} + 1}{\sum_{i=1}^{s} 3^{s-i+1} + \frac{5}{3}}.$$

As $\sum_{j=1}^{\infty} 1/h(j) > 1 + (1/2) + (1/6) = 5/3$,

$$r_{tv}(s) > \frac{\sum_{i=1}^{s} 3^{s-i+1} + 1}{\sum_{i=1}^{s} 3^{s-i+1} + \left(\sum_{j=1}^{\infty} \frac{1}{h(j)} \right)}.$$

Thus, we obtain the desired result.

Case 2(b): $(1/2)C_{s+1} < w_{i^*} \leqslant (2/3)C_{s+1}$.

Since $w_{i^*} \leqslant (2/3)C_{s+1}$, we use Lemma 4 and obtain $V_i = 3^{s-i+1} m_i V_{i^*}$ for all i, $1 \leqslant i \leqslant s$, where $m_i \geqslant m_{i+1} \geqslant 1$ are suitable multipliers. Thus, the heuristic solution has value

$$Z_t(s) \geqslant \sum_{i=1}^{s} 3^{s-i+1} m_i V_{i^*} + V_{s+1}.$$

As $w_{i^*} > (1/2)C_{s+1}$, Theorem 1 implies that $Z_{s+1} \leqslant V_{i^*} + \sum_{j=2}^{\infty} (1/h(j)) V_{s+1}$. Thus,

$$\begin{aligned} r_{tv}(s) &\geqslant \frac{\sum_{i=1}^{s} 3^{s-i+1} m_i V_{i^*} + V_{s+1}}{\sum_{i=1}^{s} 3^{s-i+1} m_i V_{i^*} + V_{i^*} + \sum_{j=2}^{\infty} \frac{1}{h(j)} V_{s+1}} \\ &\geqslant \frac{\sum_{i=1}^{s} 3^{s-i+1} m_i V_{i^*} + V_{i^*}}{\sum_{i=1}^{s} 3^{s-i+1} m_i V_{i^*} + V_{i^*} + \sum_{j=2}^{\infty} \frac{1}{h(j)} V_{i^*}} \\ &\geqslant \frac{\sum_{i=1}^{s} 3^{s-i+1} + 1}{\sum_{i=1}^{s} 3^{s-i+1} + \sum_{j=1}^{\infty} \frac{1}{h(j)}}. \end{aligned}$$

Case 2(c): $(1/3)C_{s+1} < w_{i^*} \leqslant (1/2)C_{s+1}$.

If the optimal solution takes two units of item i^*, the remaining capacity is less than $C_{s+1}/3$ and arguments similar to those used in Case 2(a) may be used to show that

$$Z \leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + \frac{13}{9} V_{s+1}.$$

If the optimal takes one unit of item i^*, then $v_{i^*} \leqslant V_{s+1}/2$. Let j^* be the next largest item in the optimal solution to Problem P_{s+1}. Then $w_{j^*} \leqslant w_{i^*} \leqslant (1/2)C_{s+1}$ and $v_{j^*} \leqslant V_{s+1}/2$. If $w_{j^*} > (1/3)C_{s+1}$ the capacity remaining after items i^* and j^* are included is less than $(1/3)C_{s+1}$, and Lemma 5 implies that the densest item that can be fitted in this capacity can contribute no more than $(4/9)V_{s+1}$ to the optimal solution. Thus,

$$\begin{aligned} Z &\leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + v_{i^*} + v_{j^*} + \frac{4}{9} V_{s+1} \\ &\leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + \frac{13}{9} V_{s+1}. \end{aligned}$$

If $w_{j^*} \leqslant (1/3)C_{s+1}$, then by Lemma 5 the maximum density among the optimal items (not including item i^*) is given by $(4/3)(V_{s+1}/C_{s+1})$. Since the remaining capacity after including item i^* is at most $(2/3)C_{s+1}$, an upper bound on the optimal solution value is

$$Z \leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + v_{i^*} + \frac{2}{3} \cdot \frac{4}{3} V_{s+1} < \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + \frac{13}{9} V_{s+1}.$$

Using this bound on Z, the desired lower bound on $r_{tv}(s)$ follows using the steps in Case 2(a).

Case 2(d): $w_{i^*} < (1/3)C_{s+1}$.

From Lemma 5, the maximum density of any item in the optimal is at most $(4/3)(V_{s+1}/C_{s+1})$. Thus, an upper bound on the optimal solution value is

$$Z \leqslant \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + \frac{4}{3} V_{s+1} < \sum_{i=1}^{l} 2 \cdot 3^{s-i} m_i V_{s+1} + \sum_{i=l+1}^{s} 3^{s-i+1} m_i V_{s+1} + \frac{13}{9} V_{s+1}.$$

Once again, the desired lower bound on $r_{tv}(s)$ follows using the steps in Case 2(a). □

Problem $T(s)$ in Table 2 shows that the bound obtained in Theorem 6 is tight if $q = s+1$ and $n \to \infty$. The total-value heuristic selects one unit each of items $1, 2, \ldots, s$ and $h(n-s)$ units of item n. The optimal solution consists of 1 unit each of items 1 through n.

Theorem 7 uses the above bound on $r_{tv}(s)$ to obtain a lower bound on $E[r_{tv}]$.

Theorem 7

$$E[r_t] \begin{cases} = 1 & \text{if } n = 1; \\ \geqslant (1-p)\left(\sum_{s=0}^{n-2} \frac{\sum_{i=1}^{s} 3^{s-i+1} + 1}{\sum_{i=1}^{s} 3^{s-i+1} + \sum_{j=1}^{\infty} \frac{1}{h(j)}} \cdot p^s \right) + p^{n-1} & \text{if } n \geqslant 2. \end{cases}$$

Proof. If $n = 1$, both the greedy and optimal solutions comprise as many units as possible of the available item and $E[r_{tv}] = 1$. Consider $n \geqslant 2$. If $s = 0$, the total-value greedy heuristic obtains at least its worst-case performance ratio of $r_{tv}(0)$ with probability $1 - p_1$. If $q > s \geqslant 1$, the total-value greedy heuristic obtains at least its worst-case performance ratio of $r_{tv}(s)$ with probability

Table 2
Problem $T(s)$: Worst-case example when total-value greedy heuristic selects s greedy-optimal items

Item number i	Unit weight w_i	Unit value v_i	Total value at step 1 $v_i \lfloor C_1/w_i \rfloor$	Available capacity C_i	Contribution to greedy solution $v_i \lfloor C_i/w_i \rfloor$	Type of item
1	$2^{s-1}(1+\varepsilon)$	3^s	3^s	$2^s(1+\varepsilon) - \varepsilon$	3^s	Greedy-optimal
2	$2^{s-2}(1+\varepsilon)$	3^{s-1}	$3^{s-1}(2^2-1)$	$2^{s-1}(1+\varepsilon) - \varepsilon$	3^{s-1}	Greedy-optimal
⋮	⋮	⋮	⋮	⋮	⋮	Greedy-optimal
$s-1$	$2(1+\varepsilon)$	3^2	$3^2(2^{s-1}-1)$	$2^2(1+\varepsilon) - \varepsilon$	3^2	Greedy-optimal
s	$1+\varepsilon$	3	$3(2^{s-1})$	$2(1+\varepsilon) - \varepsilon$	3	Greedy-optimal
$s+1$	$(1/2)+\varepsilon$	1	$2^{s+1}-1$	Not applicable	Not applicable	Optimal but not chosen by greedy
$s+2$	$(1/3)+\varepsilon$	1/2	$2^{s+1}-1$[a]	Not applicable	Not applicable	Optimal but not chosen by greedy
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$n-1$	$1/(h(n-s-1)+1)+\varepsilon$	$1/h(n-s-1)$	$2^{s+1}-1$[a]	Not applicable	Not applicable	Optimal but not chosen by greedy
n	$1/(h(n-s)+1)+\varepsilon$	$1/h(n-s)$	$2^{s+1}-1$[a]	Not applicable	1	Greedy but not optimal

The greedy heuristic solution consists of one unit each of items 1 through s and $h(n-s)$ units of item n. The optimal solution consists of one unit each of items 1 through n.

[a] Upper bound.

$p_1p_2\cdots p_s(1-p_{s+1})$. If $s=q$, none of the remaining $n-q$ items are feasible at step $q+1$, and hence $r_{tv}(q)=1$ with probability $p_1p_2\ldots p_q$. Thus, a lower bound on the expected performance ratio of the total-value greedy heuristic is

$$\sum_{s=0}^{q-1}(p_0p_1\cdots p_s(1-p_{s+1})r_{tv}(s))+p_0p_1\cdots p_q\cdot 1,$$

where $p_0=1$. In this expression, the term corresponding to $s=q-1$ and the last term can be written as

$$\begin{aligned}&p_1p_2\cdots p_{q-1}(1-p_q)r_{tv}(q-1)+p_1p_2\cdots p_q\cdot 1\\&\quad=p_1p_2\cdots p_{q-1}\left[(1-p_q)r_{tv}(q-1)+p_q\cdot 1\right]\\&\quad\geqslant p_1p_2\cdots p_{q-1}[(1-p)r_{tv}(q-1)+p\cdot 1],\end{aligned}$$

where the inequality follows since $r_{tv}(q-1)\leqslant 1$ and $p\leqslant p_q$. By a similar reasoning, the conditions $r_{tv}(i-1)<r_{tv}(i)$ and $p\leqslant p_i$, $1\leqslant i\leqslant q-1$, can be shown to imply

$$E[r_{tv}]\geqslant\sum_{s=0}^{q-1}(p^s(1-p)r_{tv}(s))+p^q\cdot 1.$$

The above expression is non-increasing in q. As $q\leqslant n$ and $r_{tv}(s)=1$ if $s=n-1$ or n (i.e., the total-value greedy heuristic finds the optimal solution if $s=n-1$ or $s=n$), the above expression is minimized if $q=n-1$. Thus,

$$E[r_{tv}]\geqslant(1-p)\sum_{s=0}^{n-2}(p^sr_{tv}(s))+p^{n-1}\cdot 1.$$

Substituting the lower bound for the value of $r_{tv}(s)$ from Theorem 6 yields the desired result. □

The bound on the average performance of the total-value greedy heuristic cannot be written as a closed form expression for large n. However, the bound on $E[r_{tv}(n)]$ achieves a lower asymptote as n tends to infinity, and the value of $E[r_{tv}(n)]$ increases from the worst-case bound (>0.5913) to 1 as p increases from 0 to 1.

To show that the bound derived in Theorem 7 is tight, consider the following example. Problem $T(s)$ is generated with probability $p^s(1-p)$, $0\leqslant s\leqslant n-1$, where $T(s)$ is shown in Table 2. At each step, the probability of the greedy heuristic selecting a greedy-optimal item is p. If a non-optimal item is selected by the greedy heuristic at step $s+1$, it terminates at step $q=s+1$, achieving the worst-case performance ratio given that it selects s greedy-optimal items. For n items, the value of q is permitted to range from 1 to $n-1$, the greedy heuristic obtaining the optimal solution if $q=n-1$ or n. The expected performance ratio for the data-generating mechanism in this example achieves the lower bound described in Theorem 7.

4. Density-ordered greedy heuristic

We now analyse the average case performance of the density-ordered greedy heuristic. Recollect that at any step j, the density-ordered greedy heuristic selects as many units as are feasible of an item denoted by j. Let $r_d(s)$ denote the performance ratio of the density-ordered greedy heuristic given that it selects greedy-optimal items at each of its first $s\geqslant 0$ steps. We begin by stating the worst-case bound for $r_d(s)$, the performance ratio for the greedy heuristic if it selects greedy-optimal items at each of its first s steps. The proof for Lemma 8 is straightforward and is based on the observation that at each stage, the heuristic uses the densest feasible item to fill up at least half the remaining knapsack capacity. As before, $r_d(s)=1$ if $s\geqslant n-1$ or if $s=q$.

Lemma 8. $r_d(s)\geqslant 1-(1/2^{s+1})$ *if* $s\leqslant n-2$ *and* $q\geqslant s+1$.

Theorem 9 characterizes the lower bound on the expected performance ratio $E[r_d]$ for the density-ordered greedy heuristic (as for the total-value greedy heurisitic, p denotes the minimum probability with which the heuristic selects a greedy-optimal item at any step).

Theorem 9

$$E[r_d]\begin{cases}=1\\ \quad\textit{if}\ n=1;\\ \geqslant(1-p)\left(\frac{1-p^{n-1}}{1-p}-\frac{1-(p/2)^{n-1}}{2-p}\right)+p^{n-1}\\ \quad\textit{if}\ n\geqslant 2.\end{cases}$$

Proof. Using arguments similar to those used in Theorem 7, we can show that

$$E[r_d] \geqslant (1-p)\sum_{s=0}^{n-2}\left(p^s\left(1-\frac{1}{2^{s+1}}\right)\right)+p^{n-1}.$$

Simplifying the right hand side of this expression, we obtain

$$E[r_d] \geqslant (1-p)\left(\frac{1-p^{n-1}}{1-p}-\frac{1-(p/2)^{n-1}}{2-p}\right)+p^{n-1},$$

which is the desired result. □

To show that the bound derived in Theorem 9 is tight, consider the following example. Problem $D(s)$ is generated with probability $p^s(1-p)$, $0 \leqslant s \leqslant n-1$, where $D(s)$ is shown in Table 3. At each step, the probability of the greedy heuristic selecting a greedy-optimal item is p. The expected performance ratio for this data-generating mechanism can be verified to be the lower bound on $E[r_d]$ described in Theorem 9. For large n, this bound asymptotically approaches the limit

$$\lim_{n\to\infty} E[r_d] \geqslant \lim_{n\to\infty}\left\{(1-p)\left(\frac{1-p^{n-1}}{1-p} - \frac{1-(p/2)^{n-1}}{2-p}\right)+p^{n-1}\right\} = \frac{1}{2-p}.$$

As p increases from 0 to 1, the lower bound on the expected value of the performance ratio increases from the worst case bound of 1/2 to 1. Note that the above data-generating mechanism can be parameterized with any value of p, for each value of which the above bound is strictly obtained.

How do the total-value and the density heuristics compare? Unlike the density-ordered greedy heuristic, the bound on the average performance of the total-value greedy heuristic cannot be written as a closed form expression for large n. For any value of n, however, the bound on the expected performance ratio is larger for the total-value greedy heuristic than for the density-ordered greedy heuristic. Thus, the total-value greedy heuristic dominates the density-ordered greedy heuristic with regard to both the worst-case performance, and the average-case performance for the same value of p.

5. Conclusion

This research identifies several open questions. First, it would be useful to derive the value of p for different data generating distributions. While we do not do so theoretically, we have conducted a computational investigation to estimate p. To do so, we generate six of the seven classes of knapsack problems described in [19]. (Because it generates equal densities for all items, we do not test the class of subset sum instances.) We use a *data range* value of 10,000 (the range over which the unit value and weight of an item vary), and generate 100 problem instances in each class. To estimate a lower bound on p for the total-value (density) heuristic, we record the proportion of problem instances for which the optimal solution and the total-value (density) heuristic solution match exactly. (We obtain the optimal solution using dynamic programming.) This analysis results in an

Table 3
Problem $D(s)$: Worst-case example when density-ordered greedy heuristic selects s greedy-optimal items

Item number i	Unit weight w_i	Unit value v_i
1	$(1/2)+\varepsilon$	$(1/2)+\varepsilon$
2	$(1/4)+\varepsilon$	$(1/4)+\varepsilon$
$\vdots$	$\vdots$	$\vdots$
s	$(1/2^{s+1})+\varepsilon$	$(1/2^{s+1})+\varepsilon$
$s+1$	$(1/2)\left(1-\sum_{i=1}^{s} w_i\right)+s\varepsilon$	$(1/2)\left(1-\sum_{i=1}^{s} w_i\right)+s\varepsilon$
$\vdots$	$\vdots$	$\vdots$
n	$1-\sum_{i=1}^{s} w_i$	$1-\sum_{i=1}^{s} w_i$

The greedy heuristic solution consists of one unit each of items 1 through $s+1$. The optimal solution consists of one unit each of items 1 through s and item n.

estimated mean for the value of p to be 0.79 for the total-value heuristic and 0.76 for the density heuristic.

Second, the preceding bounds are valid regardless of the sizes of the available items. The worst-case bound for the density-ordered greedy heuristic improves to $k/(k+1)$ if at least $k = \max_i k_i$ units of each item i can fit into the knapsack. Similarly, for the total-value greedy heuristic, the worst-case bound improves to $1/\sum_{i=1}^{\infty} 1/h(i)$, where $h(i)$ is an integer value given by the recursion $h(1) = 1$, $h(2) = k + 1$, $h(i) = [h(i-1)]\cdot[h(i-1)+1]$ for $i \geqslant 3$ (see [16] for details). Using these bounds, it may be useful to examine the bounds on the expected performance ratio for both greedy heuristics as a function of k.

Finally, it may also be useful to examine the average performance of the better of the solutions obtained by running the density-ordered and total-value greedy heuristics for every problem instance. The joint worst-case bound for the two heuristics is 2/3 [16], which is larger than the worst-case bound for either heuristic alone. It is possible that the average performance of the joint heuristic also dominates the independent average performances of the two greedy heuristics.

References

[1] R. Andonov, V. Poirriez, S. Rajopadhye, Unbounded knapsack problem: Dynamic programming revisited, European Journal of Operational Research 123 (2) (2000) 394–407.

[2] J. Bentley, D.S. Johnson, F.T. Leighton, C.C. McGeoch, Some unexpected expected behavior results for bin packing, Proceedings of the 16th ACM Symposium on the Theory of Computing, 1984, pp. 279–288.

[3] O.J. Boxma, A probabilistic analysis of multiprocessing list scheduling: The Erlang case, Stochastic Models 1 (1985) 209–220.

[4] J.L. Bruno, P.J. Downey, Probabilistic bounds for dual bin packing, Acta Informatica 22 (1985) 333–345.

[5] R.E. Burkard, Travelling salesman and assignment problems: A survey, Annals of Discrete Mathematics 4 (1979) 193–215.

[6] E.G. Coffman, K. So, M. Hofri, A.C. Yao, A stochastic model of bin-packing, Information and Control 44 (1980) 105–115.

[7] E.G. Coffman, G.S. Lueker, A.H.G. Rinnooy Kan, Asymptotic methods in the probabilistic analysis of sequencing and packing heuristics, Management Science 34 (3) (1988) 266–290.

[8] J. Csirik, J.B.G. Frenk, A. Frieze, G. Galambos, A.H.G. Rinnooy Kan, A probabilistic analysis of the next fit decreasing bin packing heuristic, Operations Research Letters 5 (1986) 233–236.

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

[10] M.L. Fisher, D.S. Hochbaum, Probabilistic analysis of the planar K-median problem, Mathematics of Operations Research 5 (1) (1980) 27–34.

[11] J. Franco, M. Paull, Probabilistic analysis of the Davis Putnam procedure for solving the satisfiability problem, Discrete Applied Mathematics 5 (1983) 77–87.

[12] J.B.G. Frenk, M. Van Houweninge, A.H.G. Rinnooy Kan, Asymptotic properties of the quadratic assignment problem, Mathematics of Operations Research 10 (1) (1985) 100–116.

[13] R. Karp, Probabilistic analysis of partitioning algorithms for the travelling-salesman problem in the plane, Mathematics of Operations Research 2 (3) (1977) 209–224.

[14] R.M. Karp, M. Luby, A. Marchetti-Spaccamela, A probabilistic analysis of multidimensional bin packing problems, Proceedings of the 16th ACM Symposium of the Theory of Computing 1984, pp. 289–298.

[15] R. Kohli, R. Krishnamurti, Average performance of heuristics for satisfiability, SIAM Journal on Discrete Mathematics 2 (4) (1989) 508–523.

[16] R. Kohli, R. Krishnamurti, A total-value greedy heuristic for the integer knapsack problem, Operations Research Letters 12 (2) (1992) 65–72.

[17] R. Kohli, R. Krishnamurti, P. Mirchandani, The minimum satisfiability problem, SIAM Journal on Discrete Mathematics 7 (2) (1994) 275–283.

[18] S. Martello, P. Toth, Knapsack Problems: Algorithms and Computer Implementations, John Wiley and Sons, Chichester, UK, 1990.

[19] S. Martello, D. Pisinger, P. Toth, New trends in exact algorithms for the 0–1 knapsack problem, European Journal of Operational Research 123 (2) (2000) 325–332.

[20] D. Pisinger, P. Toth, Knapsack problems, in: D.-Z. Du, P. Pardalos, (Eds.), Handbook of Combinatorial Optimization, vol. 1, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 299–428.

[21] P.W. Shor, The average-case analysis of some on-line algorithms for bin packing, Combinatorica 6 (2) (1986) 179–200.