COMBINATORICA 2(4) (1982) 385—393

AN ANALYSIS OF THE GREEDY ALGORITHM FOR THE SUBMODULAR SET COVERING PROBLEM

Laurence A. WOLSEY

Received 19 October 1981

We consider the problem: $\min \{\sum_{j\in S} f_j\colon z(S)=z(N),\ S\subseteq N\}$ where z is a nondecreasing submodular set function on a finite set N. When z is integer-valued and $z(\emptyset)=0$, it is shown that the value of a greedy heuristic solution never exceeds the optimal value by more than a factor $H(\max_j z(\{j\}))$, where $H(d)=\sum_{i=1}^{d}\frac{1}{i}$.

This generalises earlier results of Dobson and others on the applications of the greedy algorithm to the integer covering problem: $\min\{fy\colon Ay\geqq b,\, y\in\{0,1\}\}$ where $a_{ij}, b_i\geqq 0$ are integer, and also includes the problem of finding a minimum weight basis in a matroid.

1. Introduction

Several authors have very recently studied the behaviour of the greedy heuristic for various versions of the integer covering problem

$$(C)\qquad \min\left\{\sum_{j=1}^{n} f_j y_j\colon\ Ay \geqq b,\quad y_j\in\{0,1\}\quad j=1,\dots,n\right\}$$

where $\qquad a_{ij}\geqq 0,\ b_i, f_j > 0\quad$ for all i and j.

On the other hand the optimality of the greedy algorithm for finding a minimum weight basis in a matroid is by now a classic result. Here we consider a generalisation of both problems, namely the submodular set covering problem:

$$(Q)\qquad Z=\min_{S\subset N}\left\{\sum_{j\in S} f_j\colon\ z(S)=z(N)\right\}$$

where $z\colon \mathcal{P}(N)\to R$ is a nondecreasing, submodular set function on $N=\{1,\dots,n\}$. A function is *submodular* if $z(A)+z(B)\geqq z(A\cup B)+z(A\cap B)$.

To see that the integer covering problem (C) is a special case of (Q), it suffices to take $z(S)=\sum_{i=1}^{m}\min\left\{\sum_{j\in S} a_{ij},\ b_i\right\}$, while we obtain the minimum weight spanning set of a matroid by taking z to be the rank function of the matroid. Another

---

AMS subject classification (1980): 68 C 05, 68 C 25; 90 C 10, 05 B 35 case of (Q) of practical interest is the set covering problem with capacity restrictions:

$$\min\left\{\sum_{j=1}^{n} f_j y_j:\ \sum_{j=1}^{n} a_{ij}x_{ij} \geqq 1 \quad i=1,\dots m, \quad \sum_{i=1}^{m} x_{ij} \leqq d_j y_j \quad j=1,\dots,n,\right.$$
$$\left. x_{ij} \geqq 0, \quad y_j \in \{0,1\} \quad \text{with} \quad a_{ij}\in\{0,1\}.\right\}$$

Note that z' defined by $z'(S)=\min\{z_0, z(S)\}$ is submodular and nondecreasing whenever z is, so the apparently more general constraint $z(S)\geqq z_0$ also fits the model.

The main result of the paper is to show that if a greedy heuristic is applied to problem (Q), the value Z^G of a greedy heuristic solution always satisfies $Z^G \leqq (1+\log_e \gamma)Z$ where γ is one of several possible problem parameters. In the special case that z is integer-valued, the analysis gives $Z^G/Z \leqq H(\max_j z(\{j\})-z(\emptyset))$. This leads to an error factor of $H\left(\max_j \sum_{i=1}^{m} a_{ij}\right)$ for problem (C) with integer data, which is a result of Dobson [3], generalising earlier results of Johnson [5], Lovász [6] and Chvátal [1] for the set covering problem. If z is the rank function of a matroid, $\max_j (z(\{j\})-z(\emptyset))=1$, $H(1)=1$, and greedy is optimal, see Rado [9] and many others.

The problem (Q) and its analysis is also closely related to the problem $\max_{S\subseteq N}\{z(S):\ \sum_{j\in S} f_j \leqq f_0\}$ which has been studied extensively in [7, 8, 10].

The outline of the paper is as follows. In the following section we give an integer programming reformulation of (Q), describe the greedy heuristic and prove the main result. In Section 3 we indicate how a similar analysis can be carried out for a continuous version of (Q), the problem: $Z_R=\min\{w(y):\ w(y)=w(h),\ y\geqq 0\}$ where w is a concave submodular nondecreasing function on R^n_+. In addition we deduce a highly negative result for the family of set covering problems with unit costs and duplicate rows: $\min\left\{\sum_{j=1}^{n} y_j:\ \sum_{j=1}^{n} a_{ij}y_j \geqq 1\ i=1,\dots,m,\ y_j\in\{0,1\}\ j\in N\right\}$ where $a_{ij}\in\{0,1\}$ for all i and j. Among all "black box" algorithms looking only at values of the subroutine $z(S)=\sum_{j=1}^{m}\min\{\sum_{j\in S} a_{ij}, 1\}=\sum_{i=1}^{m}\max_{j\in S} a_{ij}$ (i.e., $z(S)$ is the number of rows covered by the set S of columns), there is no approximation algorithm making a polynomial number of calls of the subroutine that guarantees less than τ times the optimal value for all problem instances for any fixed value of τ.

2. Problem reformulation and the greedy heuristic

First we present a reformulation of (Q) as a linear integer program. For this it is useful to view two alternative properties of submodular functions.

Let $\varrho_j(S)=z(S\cup\{j\})-z(S)$.

Proposition 1 [7]. *A set function $z\colon \mathscr{P}(N)\to R$ is submodular and nondecreasing if and only if either*

a) $\varrho_j(S)\geqq\varrho_j(T)\geqq 0\quad \forall S\subseteq T\subseteq N$ *or*

b) $z(T)\leqq z(S)+\sum_{j\in T\setminus S}\varrho_j(S)\quad \forall S,T\subseteq N.$ ∎

Consider now the linear integer program:

$$(Q^I)\qquad Z_I = \min \sum_{j\in N} f_j y_j \quad \text{s.t.}$$
$$\sum_{j\in N} \varrho_j(S) y_j \geqq z(N)-z(S) \quad \forall S \subseteqq N \quad y_j \in \{0, 1\} \quad j \in N.$$

The following result shows us that problems (Q) and (Q^I) are equivalent, and that $Z=Z_I$.

Proposition 2. *$T \subseteqq N$ is feasible in (Q) if and only if its characteristic vector y^T is feasible in (Q^I).*

Proof. Suppose T is feasible in (Q). Then $\sum_{j\in N} \varrho_j(S) y_j^T = \sum_{j\in T\setminus S} \varrho_j(S) \geqq z(T) - z(S) = z(N)-z(S) \forall S \subset N$. The first equality holds as $\varrho_j(S)=0$ if $j\in S$, the inequality follows from Proposition 1, and $z(T)=z(N)$ as T is feasible in (Q).

Conversely if y^T is feasible in (Q^I), and we consider the constraint indexed by T, $0=\sum_{j\in N} \varrho_j(T) y_j^T \geqq z(N)-z(T)$, and hence $z(T)=z(N)$. ∎

Now we present the greedy algorithm.

A Greedy Heuristic for (Q)

Set $t=1$. $S^0=\emptyset$. Stop if $z(\emptyset)=z(N)$.

Iteration t. Let $\theta^t = \min_{j\in N\setminus S^{t-1}} \left\{\frac{f_j}{\varrho_j(S^{t-1})}\right\}$.

Let $\arg\min \left\{\frac{f_j}{\varrho_j(S^{t-1})}\right\} = j_t$.

Let $\varrho_t = \varrho_{j_t}(S^{t-1})$.

Set $S^t = S^{t-1} \cup \{j_t\}$, and $\sigma_t = z(S^t) - z(S^{t-1})$.

Stop if $z(S^t)=z(N)$, and set $T=t$.
Otherwise set $t=t+1$.

We say that S^T is a *greedy heuristic* solution with value $Z^G = \sum_{j\in S^T} f_j$. Evidently Z^G provides an upper bound for Z. Note also that because of submodularity $0<\theta^1 \leqq \theta^2 \leqq \ldots \leqq \theta^T$.

Theorem 1. *If the greedy algorithm is applied to (Q)*

i) $$Z^G/Z \leqq 1+\log_e \max_{j,r} \left\{\frac{\varrho_j(S^0)}{\varrho_j(S^r)} \colon \varrho_j(S^r) > 0\right\},$$

ii) $$Z^G/Z \leqq 1+\log_e \theta^T/\theta^1,$$

iii) $$Z^G/Z \leqq 1+\log_e \left\{\frac{z(N)-z(\emptyset)}{z(N)-z(S^{T-1})}\right\}.$$

If z is integer-valued,

iv) $$Z^G/Z \leqq H(\max_j z(\{j\}) - z(\emptyset)),$$

where $H(d) = \sum_{i=1}^{d} \frac{1}{i}$ *for d a positive integer.*

Before proving the theorem we need one preliminary result.

Proposition 3. *Let* $0 < u_1 \leqq u_2 \leqq \ldots \leqq u_n$, *and* $x_1 \geqq x_2 \geqq \ldots \geqq x_n > 0$. *If* $S = \sum_{i=1}^{n-1} u_i(x_i - x_{i+1}) + u_n x_n = u_1 x_1 + \sum_{i=1}^{n-1} (u_{i+1} - u_i) x_{i+1}$, *then*

$$S \leqq (\max_i u_i x_i) \left[1 + \log_e \min\left(\frac{x_1}{x_n}, \frac{u_n}{u_1}\right)\right].$$

If $(x_i)_{i=1}^n$ *are integer,* $S \leqq (\max_i u_i x_i) H(x_1)$.
If $\{u_i\}_{i=1}^n$ *are integer,* $S \leqq (\max_i u_i x_i) H(u_n)$.

Proof. Taking $u_i \leqq (\max_i u_i x_i)/x_i$, $S \leqq (\max_i u_i x_i)\left[\sum_{i=1}^{n-1}\left(1 - \frac{x_{i+1}}{x_i}\right) + 1\right] \leqq (\max_i u_i x_i) \times \left[1 + \sum_{i=1}^{n-1}\left(\log_e \frac{x_i}{x_{i+1}}\right)\right] = (\max_i u_i x_i)\left[1 + \log_e \frac{x_1}{x_n}\right]$, where the second inequality uses the fact that $1 - \frac{1}{x} \leqq \log_e x \, \forall x \geqq 1$. If $\{x_i\}_{i=1}^n$ are integer, $1 - \frac{x_{i+1}}{x_i} \leqq \frac{1}{x_i} + \frac{1}{x_{i+1}} + \ldots + \frac{1}{x_{i+1}+1}$ if $x_i > x_{i+1} \geqq 1$, so that $S \leqq (\max_i u_i x_i)\left[\sum_{i=1}^{n-1}\left(1 - \frac{x_{i+1}}{x_i}\right) + 1\right] \leqq (\max_i u_i x_i)\left[\frac{1}{x_1} + \ldots + \frac{1}{x_n + 1} + 1\right] \leqq (\max_i u_i x_i) H(x_1)$.

Taking $x_i \leqq (\max u_i x_i)/u_i$, und using an identical argument completes the proof. ∎

$$\text{Let} \quad k_1 = \max_{j,r}\left\{\frac{\varrho_j(S^0)}{\varrho_j(S^r)} : \varrho_j(S^r) > 0\right\}; \quad k_2 = \frac{\theta^T}{\theta^1}; \quad k_3 = \frac{z(N) - z(\emptyset)}{z(N) - z(S^{T-1})}.$$

Proof of Theorem 1. To analyse the heuristic it is necessary to obtain lower bounds on Z. For this we consider the following linear programming relaxation of (Q^I):

$$(Q^L) \qquad Z^L = \min \sum_{j \in N} f_j y_j : \sum_{j \in N} \varrho_j(S^t) y_j \geqq z(N) - z(S^t) \quad t = 0, \ldots, T-1$$
$$y_j \geqq 0, \quad j \in N.$$

Our aim will be to find appropriate dual feasible solutions for (Q^L) whose value will provide a lower bound on Z^L and hence on Z.

i) and ii). Let $\theta^* = (\theta^1, \theta^2 - \theta^1, \ldots, \theta^T - \theta^{T-1})$. For a given j, there exists $r \leqq T$ such that $\varrho_j(S^{r-1}) > 0$ and $\varrho_j(S^r) = 0$. Apply Proposition 3 with $0 < \theta^1 \leqq \ldots \leqq \theta^r$

and $\varrho_j(S^0)\geqq\ldots\geqq\varrho_j(S^{r-1})>0$. We obtain that

$$\theta^1\varrho_j(S^0)+(\theta^2-\theta^1)\varrho_j(S^1)+\ldots+(\theta^r-\theta^{r-1})\varrho_j(S^{r-1})$$

$$\leqq \{\max_{t=1,\ldots,r}\theta^t\varrho_j(S^{t-1})\}\left[1+\log_e\min\left\{\frac{\varrho_j(S^0)}{\varrho_j(S^{r-1})}\cdot\frac{\theta^r}{\theta^1}\right\}\right] \leqq f_j[1+\log_e\min\{k_1, k_2\}],$$

where $\theta^t\varrho_j(S^{t-1})\leqq f_j$ is a consequence of the greedy heuristic.

Hence $(1+\log_e\min\{k_1, k_2\})^{-1}\theta^*$ is dual feasible for (Q^L), and therefore

$$(1+\log_e\min\{k_1, k_2\})^{-1}[\theta^1(z(N)-z(S^0))+$$

$$+(\theta^2-\theta^1)(z(N)-z(S^2))+\ldots+(\theta^T-\theta^{T-1})(z(N)-z(S^{T-1}))]\leqq Z^L\leqq Z.$$

But

$$Z^G=\sum_{t=1}^{T}\theta^t(z(S^t)-z(S^{t-1}))=\theta^1(z(N)-z(S^0))+\sum_{t=2}^{T}(\theta^t-\theta^{t-1})(z(N)-z(S^{t-1}))$$

and hence

$$Z^G\leqq Z(1+\log_e\min\{k_1, k_2\}).$$

iii) Define $u^t\in R^T$ by $u_i^t=\theta^t$ if $i=t$, $u_i^t=0$ otherwise.

$$u^t(\varrho_j(S^0), \ldots, \varrho_j(S^{T-1}))=\theta^t\varrho_j(S^{t-1})\leqq f_j,$$

and h.nce u^t is dual feasible for $t=1, \ldots, T$. It follows that

$$\max_{t=1,\ldots,T}u^t(z(N)-z(S^0), \ldots, z(N)-z(S^{T-1}))=\max_{t=1,\ldots,T}\theta^t(z(N)-z(S^{t-1}))\leqq Z^L\leqq Z.$$

Now applying Proposition 3 with $0<\theta^1\leqq\ldots\leqq\theta^T$, and $z(N)-z(S^0)\geqq z(N)-$ $-z(S^1)\geqq\ldots\geqq z(N)-z(S^{T-1})$ gives

$$Z^G=\sum_{t=1}^{T-1}\theta^t(z(S^t)-z(S^{t-1}))+\theta^T(z(N)-z(S^{T-1}))$$

$$\leqq\max_t\{\theta^t(z(N)-z(S^{t-1}))\}\left[1+\log_e\frac{z(N)-z(S^0)}{z(N)-z(S^{T-1})}\right]\leqq Z(1+\log_e k_3).$$

iv). If z is integer-valued, $\varrho_j(S^t)$ is integer for all j and t, and from Proposition 3, we obtain

$$\theta^1\varrho_j(S^0)+\ldots+(\theta^r-\theta^{r-1})\varrho_j(S^{r-1})\leqq f_j\,H(\max_j\varrho_j(S^0)).$$

The rest of the proof follows that of i) and ii) above. ▮

Corollary. *For the problem of finding a minimum weight set that is a spanning set in each of p matroids, there exists a greedy heuristic for which* $Z^G/Z\leqq H(p)$.

Proof. Let r_i be the rank function of matroid i. Take $z=\sum_{i=1}^{p}r_i$ and apply the greedy heuristic to the resulting problem (Q). As $z(S)=z(N)$ only if $r_i(S)=r_i(N)$ for all i, the result follows from Theorem 1. ▮

It is perhaps of interest to note that when $z(S)$ is the rank function of a matroid i.e. $p=1$ above, the proof of Theorem 1 shows not only that the greedy algorithm is optimal but also the polyhedron

$$\{y: \sum_{j\in N} [r(S\cup\{j\})-r(S)]y_j \geqq z(N)-z(S) \forall S\subseteqq N, \quad y_j\geqq 0, \quad j\in N\}$$

has integer vertices, and hence is the blocker of the bases of the matroid.

3. Further results and extensions

First we consider a continuous version of the earlier model, namely the problem:

$$(R) \qquad Z_R = \min\left\{\sum_{j=1}^{n} f_j y_j \colon\ w(y) = w(h),\ y_j \geqq 0, \quad j = 1, \ldots, n\right\},$$

where $w\colon R_+^n \to R$ is nondecreasing, submodular $\big(w(x)+w(y)\geqq w(x\vee y)+w(x\wedge y)\big)$, piecewise linear and concave. Again there is no gain in generality with the constraint set $\{y\colon w(y)\geqq w_0,\ 0\leqq y\leqq h\}$ as both $\bar{w}(y)=w(y\wedge h)$ and $w'(y)=\min\big(w(y), w_0\big)$ are submodular whenever w is submodular.

Using the properties of w, it is easily shown, see [9], that:

$$w(y) \leqq w(x) + \sum_{\{s\colon y_s > x_s\}} t_s\left[w\left(x+\frac{y_s-x_s}{t_s}e_s\right)-w(x)\right] \forall x,\ y\in R_+^n,\ t_s \geqq 1.$$

Paralleling the earlier development, we now describe the continuous greedy heuristic for (R):

Let $\varrho_j(x)=\lim\limits_{\varepsilon\downarrow 0^+} \dfrac{w(x+\varepsilon e_j)-w(x)}{\varepsilon}$ where e_j is the unit vector in direction j.

A Continuous Greedy Algorithm for (R)

Set $t=1$. $y^0=(0, \ldots, 0)$. Stop if $w(0)=w(h)$.
Iteration t. Let $\theta^t=\min\limits_{j\in N} \{f_j/\varrho_j(y^{t-1})\}$
Let $\arg\min \{f_j/\varrho_j(y^{t-1})\} = j_t$.
Let $\varrho_t=\varrho_{j_t}(y^{t-1})$.
Let $\sigma_t=\varepsilon_t\varrho_t$, where $\varepsilon_t=\max\{\varepsilon\colon w(y^{t-1}+\varepsilon e_{j_t})-w(y^{t-1})=\varepsilon\varrho_t\}$.
Set $y^t=y^{t-1}+\varepsilon_t e_{j_t}$, so that $w(y^t)=w(y^{t-1})+\sigma_t$.
Stop if $w(y^t)=w(h)$, and set $T=t$.
Otherwise set $t=t+1$.

We call y^T a *continuous greedy solution* with value $Z^{CG}=\sum_{j=1}^{n} f_j y_j^T$. A lower bound on Z_R is now obtained from the linear program:

$$Z_R^L = \min \sum_{j=1}^{n} f_j y_j$$

$$(R^L) \qquad \sum_{j=1}^{n} \varrho_j(y^{t-1}) y_j \geqq w(h) - w(y^{t-1}), \quad t = 1, \ldots, T$$

$$y_j \geqq 0, \quad j = 1, \ldots, n.$$

Defining all other terms in identical fashion, we obtain:

Theorem 2. *If the continuous greedy algorithm is applied to* (R) *and terminates with a feasible solution* y^{CG},

$$Z^{CG}/Z_R \leqq 1 + \log_e \min \left\{ \left[\max_{j,r} \frac{\varrho_j(y^0)}{\varrho_j(y^r)} : \varrho_j(y^r) > 0 \right], \frac{\theta^T}{\theta^1}, \frac{w(h)-w(0)}{w(h)-w(y^{T-1})} \right\}.$$

Proof. The proof is identical to that of Theorem 1, once we have shown that (R^L) is indeed a relaxation of (R). This will follow as in Proposition 2 if we can show that

$$w(y) \leqq w(x) + \sum_{j=1}^{n} \varrho_j(x) y_j \forall x, y \in R_n^+.$$

But $w(y) \leqq w(x) + \sum_{\{s: y_s > x_s\}} t_s \left[w\left(x + \frac{y_s - x_s}{t_s} e_s \right) - w(x) \right] \forall t_s \geqq 1,$ and setting

$$t_s = (y_s - x_s)/\varepsilon$$

$$w(y) \leqq w(x) + \sum_{\{s: y_s > x_s\}} (y_s - x_s) \frac{w(x + \varepsilon e_s) - w(x)}{\varepsilon} \forall \varepsilon \to 0^+.$$

Hence $w(y) \leqq w(x) + \sum_{\{s: y_s > x_s\}} (y_s - x_s) \varrho_s(x) \leqq w(x) + \sum_{j=1}^{n} \varrho_j(x) y_j$ as $\varrho_j(x) \geqq 0$ and $x, y \geqq 0$. ∎

The linear programming covering problem: min $\{fy: Ay \geqq b, y \geqq 0\}$ treated in [4] is one special case of problem (R). The results here and in [4] suggest that initial problem scaling is of importance for the worstcase results. Dobson [3] has taken this further, and shown how rescaling in the course of the greedy algorithm (i.e., changing the submodular function) can significantly improve certain worst case performance.

To conclude the paper, we now consider a somewhat different question. Let $\mathscr{Q}$, $\mathscr{C}$ denote the families of all problems of the form (Q), (C) respectively. Given an algorithm for $\mathscr{Q}$, or some subclass of $\mathscr{Q}$ such as the family of integer covering problems $\mathscr{C}$, that works only by looking at function values $z(S)$, can we say anything about the performance guarantees we can obtain from looking at a given number of function values?

For a class $\mathscr{D}$ of problems, we say that an approximation algorithm H has *performance measure* α^H if $Z^H \leqq \alpha^H Z$ for all problems $D \in \mathscr{D}$, where Z^H is any solution value from the algorithm H, and Z is the optimal value.

We say that algorithm H is an $O(n^q)$ *black box algorithm* for $\mathscr{D}$ if there exists a constant M such that for every problem $D \in \mathscr{D}$ with $|N|=n$, algorithm H always terminates before looking at more than Mn^q function values.

It is immediate that the greedy algorithm is an $O(n^2)$ black box algorithm for $\mathscr{Q}$, and if $\mathscr{Q}^p \subseteq \mathscr{Q}$ is the subclass of problems with integer data and $z(N) \leqq p$, we know from Theorem 1 that the greedy algorithm has a performance measure of $\{1+\log_e p\}$. However, for the whole class $\mathscr{Q}$ our results do not given an obvious performance measure.

Let $\mathscr{C}^*$ be the subclass of $\mathscr{Q}$ consisting of set covering problems with unit costs and duplicate rows, i.e., $\min\left\{\sum_{j=1}^{n} y_j \colon \sum_{j=1}^{n} a_{ij}y_j \geqq 1,\ i=1,\ldots,m,\ y_j \in \{0,1\}\ j \in N\right\}$ with $a_{ij} \in \{0,1\}$ for all i and j. Perhaps not surprisingly in view of the above remark, we have:

Theorem 3. *There is no polynomial* $(O(n^q)$ *for any* $q)$ *black box algorithm for* $\mathscr{C}^*$ *having performance measure less than* τ *for any finite value of* τ.

Proof. Fix τ and q. We consider two families of functions $v_n^{r,\tau}$ and $u_n^{r,\tau}$.

Given the set $N=\{1,\ldots,n\}$, we take a set $R \subseteq N$ with $|R|=r$ as a special set. We define $v_n^{r,\tau}$ and $u_n^{r,\tau}$ as follows:

$$v_n^{r,\tau}(S) = |S| \quad \text{if} \quad |S| < \tau r, \quad \text{and} \quad R \not\subset S$$

$$v_n^{r,\tau}(S) = \tau r \quad \text{otherwise,}$$

and

$$u_n^{r,\tau}(S) = |S| \quad \text{if} \quad |S| < \tau r$$

$$u_n^{r,\tau}(S) = \tau r \quad \text{otherwise.}$$

We observe immediately that the optimal value of problem (Q) equals r when $z \equiv v_n^{r,\tau}$, and equals τr when $z \equiv u_n^{r,\tau}$.

What is more, we claim that any black box algorithm requires at least $\binom{n}{r}\Big/\binom{\tau r}{r} = O(n^r)$ calls of the function z to distinguish between $v_n^{r,\tau}$ and $u_n^{r,\tau}$. Note first that any set S that gives information (i.e., $v_n^{r,\tau}(S) \neq u_n^{r,\tau}(S)$) has $|S| < \tau r$, and hence contains less than $\binom{\tau r}{r}$ r-tuples. However the total number of r-tuples is $\binom{n}{r}$, and hence at least $\binom{n}{r}\Big/\binom{\tau r}{r}$ sets must be examined so as to be sure that the special set R has not been missed (if it exists).

Now we invoke a result from [2] that any nondecreasing set function satisfying the following condition:

$$\text{'If} \quad z(S \cup \{j\}) = z(S) \quad \text{for some} \quad S \subset N \quad \text{and} \quad j \in N-S, \quad \text{then}$$

$$Z(T \cup \{j\}) = Z(T) \quad \text{for all} \quad T \supset S\text{'}$$

is 'order equivalent' $\big(z(A) \geqq z(B)$ if and only if $z'(A) \geqq z'(B)\big)$ to a location function $z'(S) = \sum_{i=1}^{m} \max_{j \in S} a_{ij}$, where $a_{ij} \in \{0, 1\}$.

Both $v_n^{r,\tau}$ and $u_n^{r,\tau}$ satisfy the above condition, so let $v_n'^{r,\tau}$ and $u_n'^{r,\tau}$ be two order equivalent location functions. It follows that any black box algorithm H requires at least $O(n^r)$ function calls to distinguish between $v_n'^{r,\tau}$ and $u_n'^{r,\tau}$, and hence to guarantee that $\alpha^H < \tau$.

Letting $r = q+1$, we have shown that for any q and τ, there is no $O(n^q)$ black box algorithm with performance measure less than τ. ▮

Acknowledgement. I am most grateful to G. de Ghellinck for his formulation of Proposition 3 permitting a further shortening of the proof of Theorem 1.

References

[1] V. Chvatal, A Greedy Heuristics for the Set-Covering Problem, *Math. of Oper. Res.* **4** (3), (1979), 233—235.

[2] G. Cornuéjols, G. L. Nemhauser and L. A. Wolsey, A Canonical Representation of Simple Plant Location Problems and Its Applications, *SIAM J. Alg. Disc. Math.* **1**, (1980), 261—272.

[3] G. Dobson, Worst Case Analysis of Greedy Heuristics for Integer Programming with Nonnegative Data, *Technical Report SOL 80—25, Stanford, October 1980.*

[4] M. L. Fisher and L. A. Wolsey, On the Greedy Heuristic for Covering and Packing Problems, *CORE, DP 8124, Louvain-la-Neuve, May 1981.*

[5] D. S. Johnson, Approximation Algorithms for Combinatorial problems, *J. Comput. System Sci.* **9**, (1974), 256—298.

[6] L. Lovász. On the Ratio of Optimal Integral and Fractional Covers, *Discrete Math.*, **13**, (1975), 383—390.

[7] G. L. Nemhauser, L. A. Wolsey and M. L. Fisher, An Analysis of Approximations for Maximizing Submodular Set Functions — I. *Math. Prog.*, **14** (1978), 265—294.

[8] G. L. Nemhauser and L. A. Wolsey, Maximising Submodular Set Functions: Formulations and Analysis of Algorithms, *CORE, DP 7832, Louvain La-Neuve, August 1978, to appear in Ann. of. Disc. Math.*

[9] R. Rado, Note on Independence Functions, *Proc. London Math. Soc.*, **7**, (1957), 300—320.

[10] L. A. Wolsey, Maximising Real-Valued Submodular Functions: Primal and Dual Heuristics for Location Problems, *Math. of Oper. Res.* **7**, (1982), 410—425,

L. A. Wolsey

Center for Operations Research and Econometrics
Université Catholique de Louvain
1348 Louvain-la-Neuve
Belgium