A Tight Analysis of the Greedy Algorithm for Set Cover

Petr Slavík *

November 19, 1995

Abstract

We establish significantly improved bounds on the performance of the greedy algorithm for approximating *set cover*. In particular, we provide the first substantial improvement of the 20 year old classical harmonic upper bound, $H(m)$, of Johnson, Lovasz, and Chvátal, by showing that the performance ratio of the greedy algorithm is, in fact, *exactly* $\ln m - \ln\ln m + \Theta(1)$, where m is the size of the ground set. The difference between the upper and lower bounds turns out to be less than 1.1. This provides the first tight analysis of the greedy algorithm, as well as the first upper bound that lies below $H(m)$ by a function going to infinity with m.

We also show that the approximation guarantee for the greedy algorithm is better than the guarantee recently established by Srinivasan for the randomized rounding technique, thus improving the bounds on the *integrality gap*.

Our improvements result from a new approach which might be generally useful for attacking other similar problems.

Keywords: Approximation Algorithms, Combinatorial Optimization, Fractional Set Cover, Greedy Algorithm, Partial Set Cover, Set Cover.

*Department of Mathematics, State University of New York at Buffalo, Buffalo, NY 14214, USA.
E-mail: `slavik@math.buffalo.edu`

1 Introduction

Set cover is one of the oldest and most studied NP-hard problems ([8], [4], [7], [9], [1], etc.). Given a ground set U of m elements, the goal is to cover U with the smallest possible number of sets.

One of the best polynomial time algorithms for approximating *set cover* is the greedy algorithm: at each step choose the unused set which covers the largest number of remaining elements. Johnson and Lovasz ([7],[9]) showed that the performance ratio of the greedy method is no worse than $H(m)$, where $H(m) = 1 + \cdots + 1/m$ is the m^{th} harmonic number, a value which is clearly between $\ln m$ and $\ln m + 1$. Chvátal in [1] extended their results to the weighted version of the problem.

Other, more complex approximation algorithms, have also been studied. For example, Halldórsson's "local improvements" modification of the greedy algorithm ([5], [6]) improved the upper bound to about $H(m) - 0.43$ and suggested that for large ground sets this improvement can be made even stronger. Srinivasan's analysis of randomized rounding ([12]) showed some further improvements on the performance ratio in special cases, making it appear at that point that the randomized rounding algorithm was better than the greedy method.

The original classical analysis of the greedy algorithm has remained essentially unchanged for the last 20 years, despite the fact that $H(m)$ is not known to be a lower bound on the performance ratio of the greedy algorithm. In fact, Johnson in his well-known paper [7] provides a lower bound of only about $0.48 \ln m$. A straightforward modification of his approach gives a lower bound of about $0.72 \ln m$.

This state of affairs was put more sharply into focus by recent hardness results (see e.g. [10] and [3]). In particular, Feige proved a very strong result showing that for any $\epsilon > 0$, no polynomial time algorithm can approximate *set cover* within $(1-\epsilon)\ln m$ unless $NP \subset TIME[n^{O(\log\log n)}]$. Hence, under a plausible structural complexity assumption, the performance ratio of any polynomial time algorithm can improve on the harmonic bound by at most a function $f(m) = o(\ln m)$.

In this paper, we provide the first substantial improvement of this kind. Using a new approach, we show that the performance ratio of the greedy algorithm is exactly $\ln m - \ln\ln m + \Theta(1)$ (the lower and upper bounds differ by less than 1.1). This provides the first tight analysis of the greedy algorithm, as well as the first upper bound that lies below $H(m)$ by a function going to infinity with m. Clearly, Feige's recent hardness result adds to the relevance of our paper, since lower-order improvements in the bound for the greedy algorithm are now the best we can hope for; moreover, this improved analysis holds out the prospect of further identifying the exact approximation threshold at which the set cover problem becomes intractable.

The second part of our paper extends our results to fractional covers. A fractional cover consists of fractions of covering sets with the condition that for each point the fractions add up to at least 1. This enables us to compare our results with Srinivasan's bounds, and show that our bound on the integrality gap for the greedy algorithm is significantly better than that for randomized rounding.

2 Overview

Let U be a finite set, $|U| = m$, and let $\mathbf{S} = \{S_1, \ldots, S_n\}$ be a cover of U, that is $\bigcup S_i = U$. Let $\mathbf{S}^* = \{S_{i_1}, \ldots, S_{i_k}\}$ be a subcover of $\mathbf{S}$ (that is $\mathbf{S}^* \subset \mathbf{S}$ and $\mathbf{S}^*$ itself is a cover of U). Denote by c_{min} the number of sets in a minimum subcover.

The *greedy algorithm* for approximating a minimum subcover, at each step, simply chooses the covering set with the maximum number of elements left, deletes these elements from the remaining covering sets, and repeats this process until the ground set U is covered. We can assume that in case of a tie, the set with smaller subscript is chosen. We denote by c_{greedy} the size of the subcover output by the greedy algorithm.

Previous approximation guarantees were generally based on assuming some knowledge of m and c_{min}, and then using various techniques to obtain bounds on c_{greedy}. Our approach is different. We start with numbers c_{min} and c_{greedy}, and obtain bounds on m. The key to our approach is the introduction of "greedy numbers" $N(k,l)$, which turn out to satisfy a simple recursion relation and which we prove have the following property: $N(k,l)$ is the size of the smallest set U for which it is possible to have a cover of U with $c_{min} = l$ and $c_{greedy} = k$. This allows us to abstract from the *set cover* problem to an analysis of the function $N(k,l)$ leading to the establishment of both upper and lower bounds on the performance ratio. In fact we show that for any set U with $|U| = m \geq 2$ and for any cover $\mathbf{S}$ of U

$$\frac{c_{greedy}}{c_{min}} < \ln m - \ln\ln m + 0.78, \tag{1}$$

and that for all $m \geq 2$ there is a cover $\mathbf{S}$ of some set U with $|U| = m$ such that

$$\frac{c_{greedy}}{c_{min}} > \ln m - \ln\ln m - 0.31. \tag{2}$$

Notice that our approach leads to lower bounds in a stronger form than is customary. Namely, (2) holds for *all* values of m.

Generalization of our analysis to fractional covers is almost straightforward. The arguments are little more subtle, but lead, essentially, to the same upper and lower bounds. On our way to proving these bounds, we show that

$$c_{greedy} \leq \left(c^*_{min} - \frac{1}{2}\right)(\ln m - \ln c^*_{min}) + c^*_{min} \tag{3}$$

which significantly improves on Srinivasan's bound for the randomized rounding algorithm - see [12] and Section 4. Here c^*_{min} is the "cost" of the minimum fractional cover.

3 Performance Bounds

Let us first define the "greedy numbers" $N(k,l)$. For given $l \geq 2$, set

$$a_1 = 1$$

and

$$a_i = \lceil \frac{a_1 + \cdots + a_{i-i}}{l-1} \rceil \tag{4}$$

for $i = 2, 3, \ldots$. We then define the (k,l) greedy number $N(k,l)$ as

$$N(k,l) = \sum_{i=1}^{k} a_i \quad \text{for k=1,2,\ldots} \tag{5}$$

Obviously, for any $2 \leq l \leq k$, $N(l,l) = l$ and

$$N(k+1,l) = N(k,l) + \lceil \frac{N(k,l)}{l-1} \rceil = \lceil \frac{l}{l-1} N(k,l) \rceil. \tag{6}$$

Hence we can recursively generate $N(k,l)$ for any $k \geq l \geq 2$.

Consider now the *set cover* problem. The case $c_{min} = 1$, that is the case where U can be covered by one set, is not interesting, since the greedy algorithm will also output a single set, hence $c_{greedy} = c_{min}$. Therefore in what follows, we will consider only covers for which $c_{min} \geq 2$.

Set $m = |U|$. At each step, i, of the greedy algorithm, we delete q_i elements. We have $k = c_{greedy}$

steps, hence

$$q_1 + q_2 + \cdots + q_k = m. \tag{7}$$

We know that U can be covered by $l = c_{min}$ sets. By the pigeon hole principle, at least one of the sets in the minimum cover contains at least $\lceil \frac{m}{l} \rceil$ elements. Hence

$$q_1 \geq \lceil \frac{m}{l} \rceil.$$

Similarly,

$$q_2 \geq \lceil \frac{m - q_1}{l} \rceil$$

and, in general,

$$q_i \geq \lceil \frac{m - (q_1 + \cdots + q_{i-1})}{l} \rceil \tag{8}$$

for $i = 2, \ldots, k$. Using (7), we can rewrite (8) in the form

$$q_i \geq \lceil \frac{q_i + \cdots + q_k}{l} \rceil. \tag{9}$$

Solving for q_i gives

$$q_i \geq \lceil \frac{q_{i+1} + \cdots + q_k}{l - 1} \rceil \quad \text{for } i = 1, \ldots, k. \tag{10}$$

Clearly, $a_1 \leq q_k$, $a_2 \leq q_{k-1}$, ..., $a_k \leq q_1$, hence

$$N(k, l) = \sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} q_i = m = |U|. \tag{11}$$

Therefore, if $m < N(k, l)$ and $c_{min} = l$, it must be that $c_{greedy} < k$. The following example shows that the opposite is also true, namely, for any $k \geq l \geq 2$ and any $m \geq N(k, l)$ one can easily construct a cover $\mathbf{S}$ of some set U such that $|U| = m$, $c_{min} = l$, and $c_{greedy} = k$.

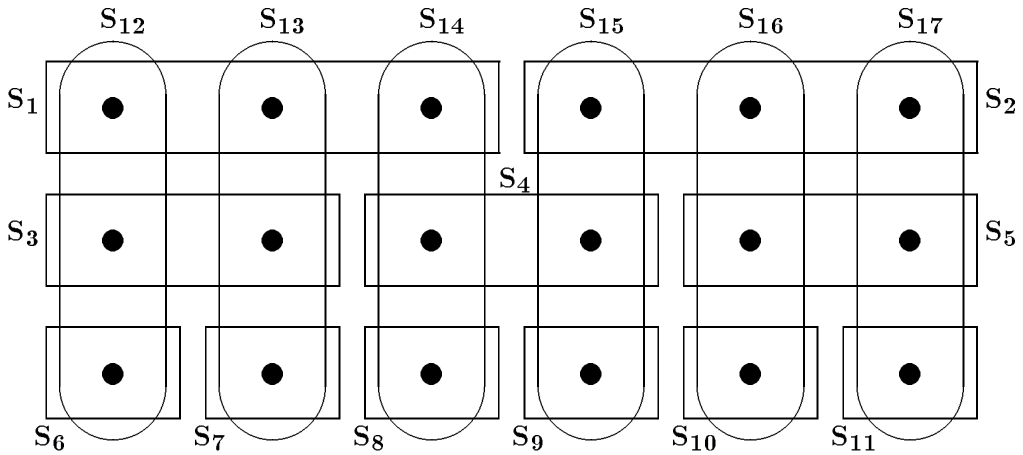

Figure 1: Example of a cover for which $c_{greedy} = k = 11$, $c_{min} = l = 6$, and $m = N(k, l) = 18$.

Example 1 Let $k \geq l \geq 2$, and $m \geq N(k, l)$ be given. Define $U = \{1, 2, \ldots, m\}$. Since $m \geq N(k, l)$ there are positive integers $q_1 \geq q_2 \geq \cdots \geq q_k$ satisfying (10) and (7). Define a cover $\mathbf{S}$ of U in the following way.

(i) For $i = 1, \ldots, k$ set

$$S_i = \{q_1 + \cdots + q_{i-1} + 1, q_1 + \cdots + q_{i-1} + 2, \ldots, q_1 + \cdots + q_i\},$$

Thus each S_i contains exactly q_i elements, the sets are disjoint, and $\bigcup_{i=1}^k S_i = U$.

(ii) Set $d = \lceil m/l \rceil$. Then one can write $m = l_1 d + l_2(d-1)$ for some l_1, l_2 such that $l_1 + l_2 = l$. Define

$$S_{k+i} = \{i, i+l, \ldots, i + l(d-1)\} \quad \text{for } i = 1, \ldots, l_1$$

and

$$S_{k+i} = \{i, i+l, \ldots, i + l(d-2)\} \quad \text{for } i = l_1 + 1, \ldots, l.$$

Then, clearly, the first l_1 sets contain d elements each, the next l_2 sets contain $d-1$ elements each, the sets are disjoint, and $\bigcup_{i=1}^l S_{k+i} = U$. Thus $c_{min} = l$.

Figure 1 shows this construction for $m = 18$, $d = 3$, $l = 6$, $k = 11$.

We claim that the greedy algorithm outputs a cover $\mathbf{S}^* = \{S_1, \ldots, S_k\}$. Indeed, $q_1 \geq d$, hence in the first iteration S_1 is chosen. Assume that after the i-th step, the sets $S_1, \ldots, S_i$ have been chosen, leaving $q_{i+1} + \cdots + q_k$ elements to cover. Set $r = \max\{|S_{k+j}^{(i)}| \ : \ j = 1, \ldots, l\}$, where $S_j^{(i)}$ denotes the set S_j after the i-th iteration of the greedy algorithm, that is after deleting the elements belonging to already chosen sets. Because of the construction of the sets in the cover, we have $r - 1 \leq |S_{k+j}^{(i)}| \leq r$ for all $j = 1, \ldots, l$. Hence $q_{i+1} + \cdots + q_k = \sum_j |S_{k+j}^{(i)}| > (r-1)l$ and thus $q_{i+1} \geq r$ by inequality (9). As a result, the greedy algorithm will choose the set S_{i+1} in its $(i+1)$-st step. Thus $c_{greedy} = k$. □

It is now clear that, for fixed $c_{min} = l$, $m < N(k,l)$ implies $c_{greedy} < k$. On the other hand, if $m \geq N(k,l)$, there are covers for which $c_{min} = l$ and $c_{greedy} = k$. This proves the following Lemma:

Lemma 1 *For any set U with $|U| \geq 2$ and for any cover $\mathbf{S}$ of U,*

$$\frac{c_{greedy}}{c_{min}} \leq \max\{\frac{k}{l} \mid N(k,l) \leq |U|\}. \tag{12}$$

Moreover, there are covers for which the equality is attained.

Lemma 1 establishes a tight bound on the quotient c_{greedy}/c_{min}. Unfortunately, by itself, it is of little practical use since we know almost nothing about the numbers $N(k,l)$. Using (6), one can evaluate the bound for the quotient c_{greedy}/c_{min} for small m, but it does not say much about asymptotic behavior. Let us now establish some lower and upper bounds on $N(k,l)$. This will enable us to find upper and lower bounds on c_{greedy}/c_{min}.

Using (6), one can easily show that

$$N(k,l) \geq \left(\frac{l}{l-1}\right)^{k-l} N(l,l) = \left(\frac{l}{l-1}\right)^{k-l} l \geq e^{\frac{k-l}{l}} l \tag{13}$$

which gives $k \leq l(\ln N(k,l) - \ln l + 1)$, hence, by Lemma 1,

$$c_{greedy} \leq c_{min}(\ln m - \ln c_{min} + 1). \tag{14}$$

But we can actually obtain a much better bound. The proof is in the Appendix.

Lemma 2 *For any set U with $m = |U| \geq 2$ and any cover $\mathbf{S}$ of U,*

$$c_{greedy} \leq (c_{min} - 1/2)(\ln m - \ln c_{min}) + l. \tag{15}$$

The fact that we can multiply by $c_{min} - 1/2$ instead of by c_{min} makes our analysis of the greedy algorithm stronger than previous results. This improvement together with Lemma 1 is crucial for establishing the following upper bound on c_{greedy}/c_{min} proved in the Appendix.

Theorem 1 *For any set U with $|U| = m \geq 2$ and for any cover $\mathbf{S}$ of U, the greedy algorithm outputs a cover of size c_{greedy} satisfying*

$$\frac{c_{greedy}}{c_{min}} \leq \ln m - \ln\ln m + 3 + \ln\ln 32 - \ln 32 \approx \ln m - \ln\ln m + 0.78. \tag{16}$$

The following lower bound on the performance of the greedy algorithm nicely complements the above result.

Theorem 2 *For every $m \geq 2$, there exist a set U with $|U| = m$ and a cover $\mathbf{S}$ of U, such that the greedy algorithm outputs a cover $\mathbf{S}^*$ with cost c_{greedy} satisfying*

$$\frac{c_{greedy}}{c_{min}} > \ln m - \ln\ln m - 1 + \ln 2 \approx \ln m - \ln\ln m - 0.31. \tag{17}$$

This Theorem shows that the upper bound (16) is tight (up to a constant). That essentially completes the analysis of the performance of the greedy algorithm for approximating the *set cover*.

Note: For any real number $u \geq 2$ set $M(u) = \max\{k/l \mid N(k,l) \leq u\}$, that is for $u = m$, $M(u) =$"the worst case for c_{greedy}/c_{min}". Figure 2 shows the graphs of $M(u)$ (in the middle), and the functions $\ln u - \ln\ln u - 1 + \ln 2$ (lower bound) and $-\ln\ln u + 3 + \ln\ln 32 - \ln 32$ (upper bound), for small and large values of u. This shows that some improvements of the constants in (16) and particularly in (17) may still be possible.

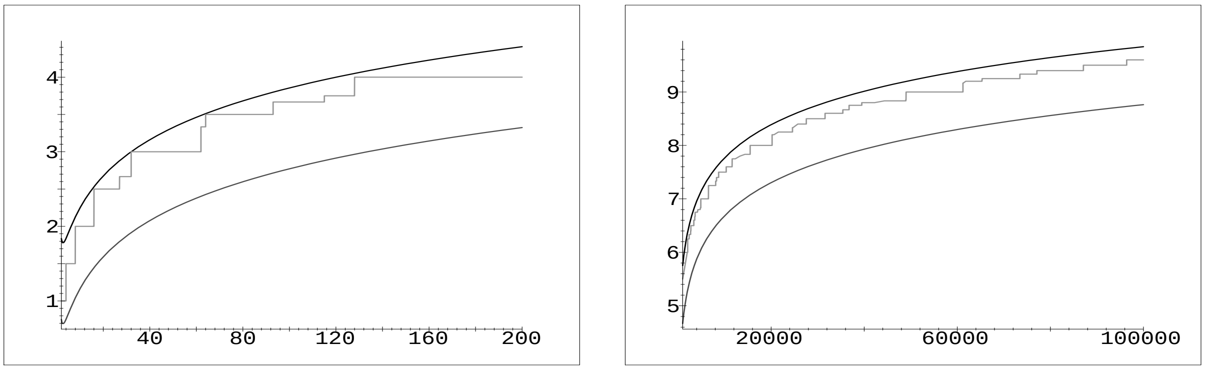

Figure 2: Graphs of M(u) and the lower and upper bounds

4 Fractional Covers

Srinivasan in [12] showed that the approximation guarantee for the randomized rounding algorithm is

$$c_{rra} \leq c^*_{min}\left(\ln(\frac{m}{c^*_{min}}) + O(\ln\ln(\frac{m}{c^*_{min}})) + O(1)\right), \tag{18}$$

making it appear at that point that, at least in some cases, the performance ratio for the randomized rounding technique was better than the performance ratio for the greedy algorithm. (Here $c^*_{min} \leq c_{min}$ is the optimum value of the LP relaxation of the *set cover* problem.)

Our bounds in the previous section are incomparable with those of [12]. In this section we further improve our estimates and show that the performance guarantee of the greedy algorithm is better than that of the randomized rounding technique.

Following [9], we define a fractional cover $\mathbf{T}$ of U to be a system of weights $\mathbf{T} = \{t_1, \ldots, t_n\}$ such that for all points $x \in U$ we have

$$\sum_{\{j \mid x \in S_j\}} t_j \geq 1.$$

Denote by $c^*(\mathbf{T})$ the "cost" of the fractional cover $\mathbf{T}$, i.e.

$$c^*(\mathbf{T}) = \sum_{j=1}^{n} t_j$$

and let

$$c^*_{min} = \min_{\mathbf{T}} c^*(\mathbf{T}).$$

This formulation is equivalent to the LP relaxation of the set cover problem considered by Srinivasan in [12]. Obviously, $c^*_{min} \leq c_{min}$.

Let us follow the steps in Section 3. Set $l^* = c^*_{min}$. A simple argument shows that $c^*_{min} = 1$ implies $c_{min} = 1$, hence by considering only those covers for which $c_{min} = l \geq 2$, we actually consider covers for which $c^*_{min} = l^* > 1$. We define generalized greedy numbers as follows. Set $a^*_1 = 1$ and

$$a^*_i = \lceil \frac{a^*_1 + \cdots + a^*_{i-i}}{l^* - 1} \rceil.$$

Define

$$N^*(k, l^*) = \sum_{i=1}^{k} a^*_i \quad \text{for k=1,2,\ldots.} \tag{19}$$

We have again that

$$N^*(k+1, l^*) = N^*(k, l^*) + \lceil \frac{N^*(k, l^*)}{l^* - 1} \rceil = \lceil \frac{l^*}{l^* - 1} N^*(k, l^*) \rceil, \tag{20}$$

and, with a small adjustment, $N^*(\lfloor l^* \rfloor, l^*) = \lfloor l^* \rfloor$ for any $l^* > 1$.

Now,

$$q_1 l^* = q_1 \sum_{j=1}^{n} t_j \geq \sum_{j=1}^{n} |S_j| t_j = \sum_{j=1}^{n} (\sum_{x \in S_j} t_j) = \sum_{x \in U} (\sum_{\{j \mid x \in S_j\}} t_j) \geq m,$$

hence $q_1 \geq \lceil \frac{m}{l^*} \rceil$. Similarly, $q_i \geq \lceil \frac{m - (q_1 + \cdots + q_{i-1})}{l^*} \rceil$ and thus, as before, $\sum_{i=1}^{k} a^*_i \leq \sum_{i=1}^{k} q_i$. From the discussion above, it is clear that $m < N^*(k, c^*_{min})$ implies $c_{greedy} < k$ hence the following counterpart of Lemma 1 holds.

Lemma 3 *For any set U with $|U| = m \geq 2$ and for any cover $\mathbf{S}$ of U,*

$$c_{greedy} \leq \max\{k \mid N(k, c^*_{min}) \leq m\}.$$

Careful analysis shows that

$$N^*(k, l^*) \geq \left(\frac{l^*}{l^* - 1}\right)^{k - \lfloor l^* \rfloor} N^*(\lfloor l^* \rfloor, l^*) \geq \left(\frac{l^*}{l^* - 1}\right)^{k - \lfloor l^* \rfloor} \lfloor l^* \rfloor \geq e^{\frac{k - l^*}{l^*}} l^* \tag{21}$$

for $k \geq l^*$, hence

$$\ln N^*(k, l^*) \geq k/l^* + \ln l^* - 1. \tag{22}$$

Thus Lemma 3 gives the following.

Proposition 1 *For any set U with $|U| = m \geq 1$ and for any cover* $\mathbf{S}$ *of U, the cover output by the greedy algorithm satisfies*

$$c_{greedy} \leq c^*_{min}(\ln m - \ln c^*_{min} + 1). \tag{23}$$

Proposition 1 already improves on Srinivasan's bound (18). Proceeding as in Section 3, we can further improve (23) and obtain the following analogy of Lemma 2. The proof is in the Appendix.

Theorem 3 *For any set U with $|U| = m \geq 1$ and for any cover* $\mathbf{S}$ *of U, the cover output by the greedy algorithm satisfies*

$$c_{greedy} \leq \left(c^*_{min} - \frac{1}{2}\right)(\ln u - \ln c^*_{min}) + c^*_{min}. \tag{24}$$

Theorem 3 shows that the performance guarantee for the greedy algorithm is substantially better than the performance guarantee (18) for the randomized rounding algorithm. This immediately gives an improved bound on the integrality gap

$$\left(1 - \frac{1}{2c^*_{min}}\right)(\ln u - \ln c^*_{min}) + 1.$$

Moreover, inequality (24) is of the same form as the bound in Lemma 2, only c_{min} is replaced by c^*_{min}. Thus a simple repetition of the steps in the proof of Theorem 1 proves that

$$\frac{c_{greedy}}{c^*_{min}} \leq \ln m - \ln\ln m + \ln 2 + \epsilon$$

for all m large enough.

It is obvious that for any cover, $c^*_{min} \leq c_{min}$, hence for every $m \geq 2$, there are covers for which

$$\frac{c_{greedy}}{c^*_{min}} > \ln m - \ln\ln m + \ln 2 - 1.$$

5 Minimum Partial Cover

Let us now briefly mention an extension of our results to the *partial cover* problem which further generalizes *set cover.* The goal here is to find a minimum subcollection of sets covering at least a p-fraction, $0 < p \leq 1$, of the ground set. Greedy algorithm works exactly as in the "complete" cover case, hence only a slight modification of the approach of Section 2 would prove the following:

Theorem 4 *Let U be finite set of size $|U| = m$ and* $\mathbf{S}$ *a cover of U, and let $0 < p \leq 1$ be such that $\lceil pm \rceil = u \geq 2$. Denote by c_{min} the size of a minimum p-partial cover of U. The greedy algorithm outputs a p-partial cover of size c_{greedy} satisfying*

$$\frac{c_{greedy}}{c_{min}} \leq \ln u - \ln\ln u + 3 + \ln\ln 32 - \ln 32 \approx \ln u - \ln\ln u + 0.78. \tag{25}$$

Moreover, for any $u \geq 2$ there is a cover $\mathbf{S}$ *of U, such that the greedy algorithm outputs a p-partial cover* $\mathbf{S}^*$ *with cost c_{greedy} satisfying*

$$\frac{c_{greedy}}{c_{min}} > \ln u - \ln\ln u - 1 + \ln 2 \approx \ln u - \ln\ln u - 0.31. \tag{26}$$

One can similarly generalize the results of Section 4. The only difference would be the definition of p-partial fractional cover, where we want the condition $\sum_{\{j \mid x \in S_j\}} t_j \geq 1$ to hold for at least $u = \lceil pm \rceil$ points.

For a more detailed treatment of partial covers see [11].

6 Conclusion

Simple modification of our analysis to Halldórsson's "local improvements" algorithm shows that the performance ratio of this algorithm is also exactly $\ln m - \ln\ln m + \Theta(m)$, with slightly better constants than in (16) and (17). Since Halldórsson's algorithm has the best performance guarantee among the known polynomial time approximation algorithms, our analysis suggests a direction in possible further improvement of Feige's hardness result.

Acknowledgements

I would like to thank Eugene Kleinberg, my advisor, for many useful discussions. I am also very grateful to Jon Kleinberg for his comments, references, and suggestions. In particular, he pointed me towards establishing the result in Theorem 1 and helped me with the Introduction.

References

[1] V. Chvátal. A Greedy Heuristic for the Set-covering Problem. *Mathematics of Operations Research* 4(1979), pp. 233-235.

[2] P. Crescenzi and V. Kann. A Compendium of NP Optimization Problems. Technical Report SI/RR-95/02, Department of Computer Science, University of Rome "La Sapienza", 1995.

[3] U. Feige. A Threshold of $\ln n$ for Approximating Set Cover. Manuscript, 1995.

[4] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness* W.H. Freeman, New York, NY (1979).

[5] M.M. Halldórsson. Approximating Discrete Collections via Local Improvements. *Proc. 6th Annual ACM-SIAM Symposium on Discrete Algorithms* (1995).

[6] M.M. Halldórsson. Approximating Set Cover via Local Improvements. JAIST Technical Report IS-RR-95-0002F, 1995.

[7] D.S. Johnson. Approximation Algorithms for Combinatorial Problems. *Journal of Computer and System Sciences* 9(1974), pp. 256-278.

[8] R.M. Karp. Reducibility among Combinatorial Problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, 85-103, Plenum Press, New York, NY (1972).

[9] L. Lovasz. On the Ratio of Optimal Integral and Fractional Covers. *Discrete Mathematics* 13(1975), pp. 383-390.

[10] C. Lund and M. Yannakakis. On the Hardness of Approximating Minimization Problems. *Journal of the ACM* 41(1994), pp. 960-982.

[11] P. Slavík. Improved Performance of the Greedy Algorithm for Minimum Set Cover and Minimum Partial Cover. Technical Report 95-45, Department of Computer Science, SUNY at Buffalo, 1995.

[12] A. Srinivasan. Improved Approximation of Packing and Covering Problems. In *Proc. ACM Symposium on Theory of Computing* (1995), pp. 268-276.

Appendix

Proof[Lemma 2]: Since the function $y = 1/x$ is convex (=concave up), we have

$$\ln(a+b) - \ln a > \frac{b}{a+\frac{b}{2}} = \frac{2b}{2a+b}$$

for any $a, b > 0$. Hence, by the first part of inequality (13),

$$\ln N(k,l) \geq \ln l + (k-l)(\ln l - \ln(l-1)) > \ln l + (k-l)\frac{2}{2l-1}. \tag{27}$$

Rearranging (27) immediately gives

$$k \leq (l - 1/2)\,(\ln N(k,l) - \ln l) + l \tag{28}$$

for any $k \geq l \geq 2$. The discussion preceding Lemma 1 easily concludes the proof. □

Proof[Theorem 1]: We can rewrite (28) in the form

$$\frac{k}{l} \leq (1 - \frac{1}{2l})(\ln N(k,l) - \ln l) + 1,$$

hence for any real $u \geq N(k,l)$

$$\ln u - \frac{k}{l} \geq \frac{\ln u - \ln l}{2l} + \ln l - 1. \tag{29}$$

That motivates the rest of the proof. To simplify the reasoning, allow $l \geq 2$ to be a real number. Define $g(l,u) = \dfrac{\ln u - \ln l}{2l} + \ln l - 1$ and $f(u) = \min_{2 \leq l \leq u} g(l,u)$. One can easily show, that for fixed $1 \leq u \leq 2e^3$, $g(l,u)$ is an increasing function of l, and for fixed $u > 2e^3$, $g(l,u)$ is a unimodal function with both relative and absolute minimum at $l = \hat{l}$, where $\hat{l}$ satisfies $\ln u = h(\hat{l})$ with

$$\ln u = h(\hat{l}) = \ln \hat{l} + 2\hat{l} - 1. \tag{30}$$

Therefore

$$f(u) = \begin{cases} g(2,u) = \frac{\ln u}{4} + \frac{3\ln 2}{4} - 1 & \text{for } 1 \leq u \leq 2e^3, \\ g(\hat{l},u) = \ln \hat{l} - \frac{1}{2\hat{l}} = \ln h^{-1}(\ln u) - \frac{1}{2h^{-1}(\ln u)} & \text{for } u > 2e^3. \end{cases}$$

Here h^{-1} is the inverse of h. Since h is increasing, so is h^{-1}. Let us establish a lower bound on $f(u)$ for large u. Clearly, for any small $\omega > 0$, $\ln u = 2\hat{l} + \ln \hat{l} - 1 \leq (2+\omega)\hat{l}$, for $\hat{l}$ big enough. Hence $\hat{l} \geq \frac{\ln u}{2+\omega}$, and $\ln \hat{l} \geq \ln\ln u - \ln(2+\omega) \geq \ln\ln u - \ln 2 - \omega/2$. Also, $1/(2\hat{l})$ can be made arbitrarily small, hence for any $\epsilon > 0$ there exists u_o such that $f(u) \geq \ln\ln u - \ln 2 - \epsilon$ for all $u \geq u_o$. Therefore for any u large enough, and for any k and l such that $N(k,l) \leq u$, the inequality (29) implies

$$\ln u - \frac{k}{l} \geq g(l,u) \geq f(u) \geq \ln\ln u - \ln 2 - \epsilon. \tag{31}$$

Hence, by Lemma 1,

$$\frac{c_{greedy}}{c_{min}} \leq \ln m - \ln\ln m + \ln 2 + \epsilon < \ln m - \ln\ln m + 0.69 + \epsilon$$

for m large enough. By actually checking "small" values of m and $\hat{l}$ ($\hat{l} \leq 17$, $m \leq 4 \cdot 10^{15}$), one can show that

$$\frac{c_{greedy}}{c_{min}} \leq \ln m - \ln\ln m + 3 + \ln\ln 32 - \ln 32 < \ln m - \ln\ln m + 0.78$$

for all $m \geq 2$. □

Proof[Theorem 2]: In order to prove the Theorem, we will need the following Lemma.

Lemma 4 *For any $k \geq l \geq 2$*

$$\frac{k}{l} \geq (1 - \frac{1}{2l-1})(\ln N(k,l) - \ln l) + \frac{2}{l} + \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k-l}. \tag{32}$$

Proof: Obviously,

$$N(i+1,l) \leq N(i,l)\frac{l}{l-1} + \frac{l-2}{l-1} = N(i,l)\left(1 + \frac{1}{l-1} + \frac{l-2}{(l-1)N(i,l)}\right).$$

Using the fact that $\ln(1+a) < \frac{a}{2}(1 + \frac{1}{1+a})$, for any $a > 0$ and inequality (13), we have

$$\begin{aligned}
\ln N(i+1,l) - \ln N(i,l) &\leq \ln\left(1 + \frac{1}{l-1} + \frac{l-2}{(l-1)N(i,l)}\right) \\
&< \frac{1}{2}\left(\frac{1}{l-1} + \frac{l-2}{(l-1)N(i,l)}\right)\left(1 + \frac{1}{\frac{l}{l-1} + \frac{l-2}{(l-1)N(i,l)}}\right) \\
&< \frac{1}{2}\left(\frac{1}{l-1} + \frac{l-2}{(l-1)N(i,l)}\right)\left(1 + \frac{1}{\frac{l}{l-1}}\right) \\
&\leq \frac{2l-1}{2l}\left[\frac{1}{l-1} + \frac{l-2}{l(l-1)}\left(\frac{l-1}{l}\right)^{i-l}\right].
\end{aligned}$$

Adding the above inequalities for $i = l, \ldots, k-1$, and using (13) again, we get

$$\begin{aligned}
\ln N(k,l) - \ln l &\leq \frac{2l-1}{2l}\cdot\frac{k-l}{l-1} + \frac{2l-1}{2l}\cdot\frac{l-2}{l(l-1)}\cdot\frac{1-\left(\frac{l-1}{l}\right)^{k-l}}{1-\frac{l-1}{l}} \\
&\leq \frac{2l-1}{2l(l-1)}\left[(k-l) + (l-2)\left(1 - \left(\frac{l-1}{l}\right)^{k-l}\right)\right].
\end{aligned}$$

Now we can multiply the whole inequality by $2(l-1)/(2l-1)$, rearrange the terms and obtain (32). □

Proof[Theorem 2 - continue]: Let m be arbitrary. Taking advantage of the condition (30) define l to be the largest integer such that $2l - 1 + \ln l < \ln m$ and let u satisfy $\ln u = 2l - 1 + \ln l$. Thus we have

$$\ln u = 2l - 1 + \ln l < \ln m < 2l + 1 + \ln(l+1), \tag{33}$$

therefore

$$\ln m - \ln u < 2 + \ln(l+1) - \ln l < 2 + \frac{1}{l} \tag{34}$$

and, for $l \geq 3$ (that is for $m \geq 446$),

$$\ln m \leq 2l + 1 + \ln(l+1) < 2(2l-1). \tag{35}$$

Using (33) we have

$$l = \frac{\ln u - \ln l + 1}{2} \leq \frac{\ln u}{2} = \frac{\ln m}{2} - \frac{\ln m - \ln u}{2}.$$

Since $\ln(b-a) \le \ln b - a/b$ for any $0 < a < b$, we obtain

$$\ln l \le \ln\left(\frac{\ln m}{2}\right) - \frac{\ln m - \ln u}{\ln m},$$

which by (35) can be reduced to

$$\ln l \le \ln\ln m - \ln 2 - \frac{\ln m - \ln u}{2(2l-1)}. \tag{36}$$

Let k be such that $N(k,l) \le m < N(k+1,l)$. Using inequality (32) we have

$$\ln N(k+1,l) - \frac{k+1}{l} \le \ln l + \frac{\ln N(k+1,l) - \ln l}{2l-1} - \frac{2}{l} - \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k+1-l}.$$

Now we can add (and subtract) some terms and use the above inequalities to obtain

$$\begin{aligned}
\ln N(k+1,l) - \frac{k}{l} &\le \frac{1}{l} + \ln l + \frac{\ln u - \ln l}{2l-1} + \frac{\ln m - \ln u}{2l-1} \\
&\quad + \frac{\ln N(k+1,l) - \ln m}{2l-1} - \frac{2}{l} - \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k+1-l} \\
&< \ln\ln m - \ln 2 - \frac{\ln m - \ln u}{2(2l-1)} + \frac{2l-1}{2l-1} + \frac{\ln m - \ln u}{2l-1} \\
&\quad + \frac{\ln N(k+1,l) - \ln N(k,l)}{2l-1} - \frac{1}{l} - \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k+1-l} \\
&< \ln\ln m - \ln 2 + 1 + \frac{\ln m - \ln u}{2(2l-1)} \\
&\quad + \frac{\ln N(k+1,l) - \ln N(k,l)}{2l-1} - \frac{1}{l} - \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k+1-l} \\
&< \ln\ln m - \ln 2 + 1 + \frac{2+\frac{1}{l}}{2(2l-1)} + \frac{\frac{1}{l-1} + \frac{l-2}{l(l-1)}\left(\frac{l-1}{l}\right)^{k-l}}{2l} \\
&\quad - \frac{1}{l} - \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k+1-l} \\
&= \ln\ln m - \ln 2 + 1 + \left[\frac{2+\frac{1}{l}}{2(2l-1)} + \frac{1}{2l(l-1)} - \frac{1}{l}\right] \\
&\quad + \left[\frac{l-2}{2l(l-1)^2}\left(\frac{l-1}{l}\right)^{k+1-l} - \frac{l-2}{l}\left(\frac{l-1}{l}\right)^{k+1-l}\right] \\
&< \ln\ln m - \ln 2 + 1.
\end{aligned}$$

Thus for each $m \ge 446$ there are l and k, such that $N(k,l) < m$ and

$$\frac{k}{l} > \ln N(k+1,l) - \ln\ln m + \ln 2 - 1 > \ln m - \ln\ln m + \ln 2 - 1. \tag{37}$$

Checking small values of m, one can show that (37) is true for all $m \ge 2$. $\square$

Proof[Theorem 3]: In order to simplify the notation, let us omit the "*" when referring to N^* and l^*. Hence $l > 1$ is now a real number. Repeating the proof of Lemma 2, we get

$$\ln N(k,l) \geq \ln\lfloor l \rfloor + \frac{2(k - \lfloor l \rfloor)}{2l-1},$$

hence

$$\ln N(k,l) \geq \ln l + \frac{2(k-l)}{2l-1} + \omega,$$

where

$$\omega = \frac{2(l - \lfloor l \rfloor)}{2l-1} + \ln\lfloor l \rfloor - \ln l.$$

Set $\alpha = l - \lfloor l \rfloor$, that is $0 \leq \alpha < 1$. Then

$$\omega = \frac{2\alpha}{2l-1} + \ln(l-\alpha) - logl \geq \frac{2\alpha}{2l-1} - \frac{\alpha}{l-\alpha} = \frac{\alpha(1-2\alpha)}{(2l-1)(l-\alpha)}.$$

Thus for $0 \leq \alpha \leq 1/2$ and any $k \geq l > 1$, we have $\omega \geq 0$, hence

$$\ln N(k,l) \geq \ln l + \frac{2(k-l)}{2l-1}. \tag{38}$$

If $\alpha = l - \lfloor l \rfloor \geq 1/2$, then $l \geq 3/2$, hence $N(\lceil l \rceil, l) = \lfloor l \rfloor + 2$. Therefore

$$\ln N(k,l) \geq \ln(\lfloor l \rfloor + 2) + \frac{2(k - \lfloor l \rfloor - 1)}{2l-1} = \ln l + \frac{2(k-l)}{2l-1} + \epsilon,$$

where

$$\epsilon = \frac{2(l - \lfloor l \rfloor - 1)}{2l-1} + \ln(\lfloor l \rfloor + 2) - \ln l.$$

Similarly as above,

$$\epsilon \geq \frac{2\alpha - 2}{2l-1} + \frac{2-\alpha}{l+1} = \frac{2l + 3\alpha - 4}{(2l-1)(l+1)} \geq 0,$$

hence (38) is valid for $1/2 \leq \alpha < 1$ as well. Rearranging (38) completes the proof. □