Greedy differential approximations for

C. Bazgan, J. Monnot, V. Th. Paschos, and F. Serrière

LAMSADE, Université Paris-Dauphine
Place du Maréchal De Lattre de Tassigny, 75775 Paris Cedex 16, France,

We present in this paper differential approximation results for min set cover and min weighted set cover. We first show that the differential approximation ratio of the natural greedy algorithm for min set cover is bounded below by $1.365/\Delta$ and above by $4/(\Delta+1)$, where Δ is the maximum set-cardinality in the min set cover-instance. Next, we study an approximation algorithm for min weighted set cover and provide a tight lower bound of $1/\Delta$.

1 Introduction

Given a family $\mathcal{S} = \{S_1, \ldots, S_m\}$ of subsets of a ground set $C = \{c_1, \ldots, c_n\}$ (we assume that $\cup_{S_i \in \mathcal{S}} S_i = C$), a set-cover of C is a sub-family $\mathcal{S}' \subseteq \mathcal{S}$ such that $\cup_{S_i \in \mathcal{S}'} S_i = C$; MIN SET COVER is the problem of determining a minimum-size set-cover of C. MIN WEIGHTED SET COVER consists of considering that sets of $\mathcal{S}$ are weighted by positive weights; the objective becomes then to determine a minimum total-weight cover of C.

Given $I = (\mathcal{S}, C)$ and a cover $\hat{\mathcal{S}}$, the sub-instance $\hat{I}$ of I induced by $\hat{\mathcal{S}}$ is the instance $(\hat{\mathcal{S}}, C)$. For simplicity, we identify in what follows a feasible (resp., optimal) cover $\mathcal{S}'$ (resp., $\mathcal{S}^*$) by the set of indices N' (resp., N^*) of the sets of the cover, i.e., $S' = \{S_i : \ i \in N'\}$ (resp., $S^* = \{S_i : \ i \in N^*\}$).

For an instance $(\mathcal{S}, C)$ of MIN SET COVER, its *characteristic graph* $B = (L, R; E)$ is a bipartite graph B with color-classes $L = \{1, \ldots, m\}$, corresponding to the members of the family $\mathcal{S}$ and $R = \{c_1, \ldots, c_n\}$, corresponding to the elements of the ground set C; the edge-set E of B is defined as $E = \{(i, c_j) : c_j \in S_i\}$.

A cover $\mathcal{S}'$ of C is said to be *minimal* (or *minimal for the inclusion*) if removal of any set $S \in \mathcal{S}'$ results in a family that is not anymore a cover for C.

Consider an instance $(\mathcal{S}, C)$ of MIN SET COVER and a minimal set-cover $\mathcal{S}'$ for it. Then, for any $S_i \in \mathcal{S}'$, there exists $c_j \in C$ such that S_i is the only set in $\mathcal{S}'$ covering c_j. Such a c_j will be called *non-redundant with respect to* $S_i \in \mathcal{S}'$; furthermore, S_i itself will be called *non-redundant for* $\mathcal{S}'$. With respect to the characteristic bipartite graph B' corresponding to the sub-instance I' of I induced by $\mathcal{S}'$ (it is the subgraph B' of B induced by $L' \cup R$ where $L = N'$), for any $i \in L'$, there exists a $c \in R$ such that $d(c) = 1$, where, for a vertex v of a graph G, $d(v)$ denotes the degree of v. In particular, there exists at least $|N'|$ non-redundant elements, one for each set. For simplicity, we will consider only one non-redundant element with respect to $S_i \in \mathcal{S}'$. Moreover, we assume that this element is c_i for the set $i \in N'$. Thus, the set of non-redundant elements with respect to $\mathcal{S}'$ considered here is $C_1 = \{c_i,\ i \in N'\}$.

In this paper we study differential approximability for MIN SET COVER in both unweighted and weighted versions. Differential approximability is analyzed using the so-called differential approximation ratio defined, for an instance I of an **NPO** problem Π (an optimization problem is in **NPO** if its decision version is in **NP**) and an approximation algorithm computing a solution S for Π in I, as $\delta_{\mathtt{A}}(I) = |\omega(I) - m_{\mathtt{A}}(I, S)|/|\omega(I) - \mathrm{opt}(I)|$ where $\omega(I)$ is the value of the worst Π-solution for I, $m_{\mathtt{A}}(I, S)$ is the value of S and $\mathrm{opt}(I)$ is the value of an optimal Π-solution for I. For an instance $I = (\mathcal{S}, C)$ of MIN SET COVER, $\omega(I) = m$, the size of the family $\mathcal{S}$. Obviously, this is the maximum-size cover of I. Finally, standard approximability is analyzed using the standard approximation ratio defined as $m_{\mathtt{A}}(I, S)/\mathrm{opt}(I)$.

Surprisingly enough, differential approximation, although introduced in [1] since 1977, has not been systematically used until the 90's ([2, 3] are, to our knowledge, the most notable uses of it) when a formal framework for it and a more systematic use started to be drawn ([4]). In general, no apparent links exist between standard and differential approximations in the case of minimization problems, in the sense that there is no evident transfer of a positive, or negative, result from one paradigm to the other. Hence a "good" differential approximation result does not signify anything for the behavior of the approximation algorithm studied when dealing with the standard framework and vice-versa. As already mentioned, the differential approximation ratio measures the quality of the computed feasible solution according to both optimal value and the value of a worst feasible solution. The motivation for this measure is to look for the placement of the computed feasible solution in the interval between an optimal solution and a worst-case one. Even if differential approximation ratio is not as popular as the standard one, it is interesting enough to be investigated for some fundamental problems as MIN SET COVER, in order to observe how they behave under several approximation criteria. Such joint investigations can significantly contribute to a deeper apprehension of the approximation mechanisms for the problems dealt. A further motivation for the study of differential approximation is the stability of the differential approximation ratio under affine transformations of the objective function. This stability often serves in order to derive differential approximation results for minimization (resp., maximization) problems by analyzing approximability of their maximization (resp., minimization) equivalents under affine transformations. We will apply such transformation in Section 4.

We study in this paper the performance of two approximation algorithms. The first one is the classical greedy algorithm studied, for the unweighted case and for the standard approximation ratio, in [5, 6] and, more recently, in [7]. For this algorithm, we provide a differential approximation ratio bounded below by $1.365/\Delta$ when $\Delta = \max_{S_i \in \mathcal{S}}\{|S_i|\}$ is sufficiently large. We next deal with MIN

WEIGHTED SET COVER and analyze the differential approximation performance of a simple greedy algorithm that starts from the whole $\mathcal{S}$ considering it as solution for MIN WEIGHTED SET COVER and then it reduces it by removing the heaviest of the remaining sets of $\mathcal{S}$ per time until the cover becomes minimal. We show that this algorithm achieves differential approximation ratio $1/\Delta$.

Differential approximability for both MIN SET COVER and MIN WEIGHTED SET COVER have already been studied in [8] and discussed in [9]. The differential approximation ratios provided there are $1/\Delta$, for the former, and $1/(\Delta + 1)$, for the latter. Our current work improves (quite significantly for the unweighted case), these old results. Note also that an approximation algorithm for MIN SET COVER has been analyzed also in [4] under the assumption $m \geqslant n$, the size of the ground set C. It has been shown that, under this assumption, MIN SET COVER is approximable within differential approximation ratio $1/2$. More recently, in [9], under the same assumption, MIN SET COVER has been proved approximable within differential approximation ratio $289/360$.

It is proved in [4] that *if $\neq$, then inapproximability bounds for standard (and differential) approximation for* MAX INDEPENDENT SET *hold as differential inapproximability bounds for* MIN SET COVER. *Consequently, unless $=$,* MIN SET COVER *is not differentially approximable within $O(n^{\epsilon-(1/2)})$, for any $\epsilon > 0$.* This result implies that approximation ratios of the same type as in standard approximation (for example, $O(1/\ln \Delta)$, or $O(1/\log n)$) are extremely unlikely for MIN SET COVER in differential approximation. Consequently, differential approximation results for MIN SET COVER cannot be trivially achieved by simply transposing the existing standard approximation results to the differential framework. This is a further motivation of our work.

In what follows, we deal with non-trivial instances of (unweighted) MIN SET COVER. An instance I is non-trivial for unweighted MIN SET COVER if the two following conditions hold simultaneously: (i) no set $S_i \in \mathcal{S}$ is a proper subset of a set $S_j \in \mathcal{S}$, and (ii) no element in C is contained in I by only one subset of $\mathcal{S}$ (i.e., there is no non-redundant set for $\mathcal{S}$).

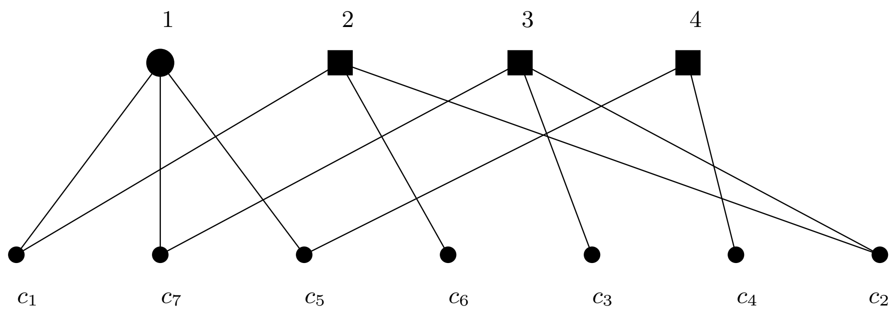
2 The natural greedy algorithm for MIN SET COVER

Let us first note that a lower bound of $1/\Delta$ can be easily proved for the differential ratio of any algorithm computing a minimal set cover. We analyze in this section the differential approximation performance of the following very classical greedy algorithm for MIN SET COVER, called SCGREEDY in the sequel:

1. compute $S_i \in \text{argmax}_{S \in \mathcal{S}}\{|S|\}$; set $N'' = N'' \cup \{i\}$ (N'' is initialized to $\emptyset$);
2. update I setting: $\mathcal{S} = \mathcal{S} \setminus \{S_i\}$, $C = C \setminus S_i$ and, for any $S_j \in \mathcal{S}$, $S_j = S_j \setminus S_i$;
3. repeat Steps 1 to 2 until $C = \emptyset$;
4. range N'' in the order sets have been chosen and assume $N'' = \{i_1, \ldots, i_k\}$;
5. Set $N' = N''$; for $j = k$ downto 1: if $N' \setminus \{i_j\}$ is a cover then $N' = N' \setminus \{i_j\}$;
6. output N' the minimal cover computed in Step 5.

Theorem 1. *For Δ sufficiently large, algorithm* SCGREEDY *achieves differential approximation ratio $1.365/\Delta$.*

Proof. Consider N'' and the sets $\mathcal{S}'' = \{S'_{i_1}, S'_{i_1} \ldots, S'_{i_k}\}$, computed in Step 4 with their residual cardinalities, i.e., as they have been chosen during Steps 1 and 2; remark that, so-considered, the set $\mathcal{S}''$ forms a partition on C. On the other hand, consider solution N' output by the algorithm SCGREEDY and remark that family $\{S'_i : \ i \in N'\}$ does not necessarily cover C.



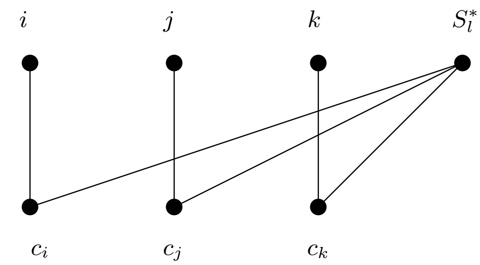
An example of application of Step 5 of .

As an example, assume some MIN SET COVER-instance $(\mathcal{S}, C)$ with $C = \{c_1, \ldots, c_7\}$ and suppose that execution of Steps 1 to 4 has produced a cover $N'' = \{1, 2, 3, 4\}$ (given by the sets $\{S_1, S_2, S_3, S_4\}$). Figure 1 illustrates characteristic graph B', i.e., the subgraph of $B = (L, R; E)$ induced by $L' \cup R$ where L' and R correspond to the sets N'' and C respectively. It is easy to see that N'' is not minimal and application of Step 5 of SCGREEDY drops the set S_1 out of N''; hence, $N' = \{2, 3, 4\}$. The residual parts of S_2, S_3 and S_4 are $S'_2 = \{c_2, c_6\}$, $S'_3 = \{c_3\}$ and $S'_4 = \{c_4\}$, respectively. Note that Step 5 of SCGREEDY is important for the solution returned in Step 6, since solution N'' computed in Step 4 may be a worst solution (see the previous example) and then, $\delta(I, \mathcal{S}') = 0$.

Consider an optimal solution N^* given by the sets S_i, $i \in N^*$ and denote by $\{S^*_i\}$, $S^*_i \subseteq S_i$, $i \in N^*$, an arbitrary partition of C (if an element c is covered by more than one sets S_i, $i \in N^*$, then c is randomly assigned to one of them). Let $C'_1 = \{c_i : \ i \in N' \setminus N^*\}$ be a set of non-redundant elements with respect to N'; obviously, by construction $|C'_1| = |N' \setminus N^*|$. Finally, set $N^*_1 = \{j \in N^* : \ \exists c \in C'_1, c \in S^*_j\}$. We deduce $N^*_1 \subseteq N^* \setminus N'$, since any element $c \in C'_1$ is non-redundant for N' (otherwise, there would exist at least a $c \in C'_1$ covered twice: one time by a set in $N' \setminus N^*$ and one time by a set in $N' \cap N^*$, absurd by the construction of C'_1). Finally, set $\bar{N} = \{1, \ldots, m\} \setminus (N' \cup N^*)$. Observe that, using the notations just introduced, we have:

$$\delta\left(I, \mathcal{S}'\right) = \frac{|N^*_1| + |N^* \setminus (N' \cup N^*_1)| + \left|\bar{N}\right|}{|N' \setminus N^*| + \left|\bar{N}\right|} \qquad (1)$$

Consider the bipartite graph $B'' = (L'', R''; E'')$ with $L'' = N_1^* \cup (N' \setminus N^*)$, $R'' = C_1'$ and $(i, c_j) \in E''$ iff $i \in S_j^*$ or $i = j$. This graph is a partial graph of the characteristic bipartite graph B' induced by $L' = N_1^* \cup (N' \setminus N^*)$ and $R' = C_1'$. By construction, B'' is not connected and, furthermore, any of its connected components is of the form of Figure 2.



A connected component of B''.

For $i = 1, \ldots, \Delta$, denote by x_i the number of connected components of B'' corresponding to sets S_l^* of cardinality i. Then, by construction of this sub-instance, we have:

$$|N_1^*| = \sum_{i=1}^{\Delta} x_i \tag{2}$$

$$|N' \setminus N^*| = |C_1'| = \sum_{i=1}^{\Delta} i \cdot x_i \tag{3}$$

Consider $z \in [1, \Delta]$ such that $|C_1'| = i_0 |N_1^*|$ where $i_0 = \Delta / z$. One can easily see that i_0 is the average cardinality of sets in N_1^* (when we consider the sets S_i^*, $i \in N_1^*$, that form, by construction, a partition on C_1'). Indeed,

$$i_0 = \frac{1}{|N_1^*|} \sum_{i \in N_1^*} |S_i^*| = \frac{\sum_{i=1}^{\Delta} i \cdot x_i}{\sum_{i=1}^{\Delta} x_i} \tag{4}$$

We have immediately from (1), (2) and (3):

$$\delta(I, \mathcal{S}') \geqslant \frac{|N_1^*|}{|N' \setminus N^*|} = \frac{|N_1^*|}{|C_1'|} = \frac{1}{i_0} = \frac{z}{\Delta} \tag{5}$$

Consider once more the component of Figure 2, suppose that set S_ℓ^* has cardinality i and denote it by $S_\ell^* = \{c_{\ell_1}, \ldots, c_{\ell_i}\}$ with $\ell_1 < \ldots < \ell_i$. By greedy rule of SCGREEDY, we deduce that the sets $S'_{\ell_1}, \ldots, S'_{\ell_i}$ (recall that we only consider the residual part of the set) have been chosen in this order (cf., Steps 4 and 5 of SCGREEDY) and verify $|S'_{\ell_p}| \geqslant i+1-p$ for $p = 1, \ldots, i$. Consequently, there exist $(i-1)+(i-2)+\ldots+1 = i(i-1)/2$ elements of C not included in C'_1. Iterating this observation for any connected component of B'' we can conclude that there exists a set $C_2 \subseteq C$, outside set C_1, of size at least $|C_2| \geqslant \sum_{i=1}^{\Delta} i(i-1)x_i/2$. Elements of C_2 are obviously covered, with respect to N^*, by sets either from N_1^*, or from $N^* \setminus N_1^*$. Suppose that sets of N_1^* of cardinality i (there exist x_i such sets), $i = 1, \ldots, \Delta$, cover a total of $k_i x_i$ elements of C_2. Therefore, there exists a subset $C'_2 \subseteq C_2$ of size at least: $|C'_2| \geqslant \sum_{i=1}^{\Delta}((i(i-1)/2) - k_i)x_i$. The elements of C'_2 are covered in N^* by sets in $N^* \setminus N_1^*$. In order that C'_2 is covered, a family $N_2^* \subseteq N^* \setminus N_1^*$ of size

$$|N_2^*| \geqslant \frac{1}{\Delta} \cdot \sum_{i=1}^{\Delta} \left(\frac{i(i-1)}{2} - k_i \right) x_i \tag{6}$$

is needed. Dealing with N_2^*, suppose that for a $y \in [0,1]$: (i) $(1-y)|N_2^*|$ sets of N_2^* belong to $N^* \setminus N'$ (indeed, they belong to $N^* \setminus (N' \cup N_1^*)$) and (ii) $y|N_2^*|$ sets of N_2^* belong to $N^* \cap N'$.

We study the two following cases: $y \leqslant (\Delta-1)/\Delta$ and $y \geqslant (\Delta-1)/\Delta$.

The first case is equivalent to $(1-y) \geqslant 1/\Delta$ and then, taking into account that $k_i \leqslant \Delta - i$, we obtain: $(1-y)|N_2^*| \geqslant |N_2^*|/\Delta \geqslant \sum_{i=1}^{\Delta}((i(i-1)/(2\Delta^2)) + (i/\Delta) - 1)x_i$. Using (1), (2), (3) and (6), we deduce:

$$\delta(I, \mathcal{S}') \geqslant \frac{|N_1^*| + |N_2^*|/\Delta}{|N' \setminus N^*|} \geqslant \frac{\sum_{i=1}^{\Delta} \left(\frac{i(i-1)}{2\Delta^2} + \frac{i}{\Delta} \right) x_i}{\sum_{i=1}^{\Delta} i \cdot x_i} = \frac{1}{\Delta} + \frac{\sum_{i=1}^{\Delta} f(i) x_i}{\sum_{i=1}^{\Delta} i \cdot x_i} \tag{7}$$

where $f(x) = x(x-1)/(2\Delta^2)$, with $1 \leqslant x \leqslant \Delta$. We will now show the following inequality (see also (4)) that $i_0 = (\sum_{i=1}^{\Delta} i x_i)/(\sum_{i=1}^{\Delta} x_i)$):

$$\frac{\sum_{i=1}^{\Delta} f(i) \cdot x_i}{\sum_{i=1}^{\Delta} i \cdot x_i} \geqslant \frac{f(i_0)}{i_0} \tag{8}$$

Remark that (8) is equivalent to $\sum_{i=1}^{\Delta} f(i) \cdot (x_i / \sum_{i=1}^{\Delta} x_i) \geqslant f(i_0)$. On the other hand, since f is convex, we have by Jensen's theorem $\sum_{i=1}^{\Delta} z_i f(i) \geqslant f(\sum_{i=1}^{\Delta} i z_i)$, where $z_i \in [0,1]$, $\sum_{i=1}^{\Delta} z_i = 1$. Setting $z_i = x_i / \sum_{i=1}^{\Delta} x_i$, (8) follows.

Thus, since $i_0 = \Delta/z$ and we study an asymptotic ratio in Δ, (7) becomes

$$\delta(I, \mathcal{S}') \geqslant \frac{1}{\Delta} + \frac{1}{2\Delta^2} \left(\frac{\Delta}{z} - 1 \right) \approx \frac{1}{\Delta} + \frac{1}{2\Delta z} \tag{9}$$

Expression (9) is decreasing with z, while (5) is increasing with z. Equality of both ratios is reached when $2z^2 - 2z - 1 = 0$, i.e., for $z = (2 + \sqrt{12})/4 \approx 1.365$.

We now deal with case $y \geqslant (\Delta-1)/\Delta$. Sub-family $N_2^* \cap N'$ (of size $y|N_2^*|$) is, by hypothesis, common to both N' (the cover computed by SCGREEDY) and N^*. Minimality of N' implies that, for any set $i \in N_2^* \cap N'$, there exists at least one element of C non-redundant with respect to S_i. So, there exist at least $|C_3| = |N_2^* \cap N'|$ elements of C outside C_1' and C_2.

Some elements of C_3 can be covered by sets in N_1^*. In any case, for the sets $\{j_1, \ldots, j_{x_i}\}$ of N_1^* of cardinality i with respect to the partition S_ℓ^*, there exist at most $(\Delta-(i+k_i))x_i$ elements of C_3 that can belong to them (so, these elements are covered by the residual set $S_{j_p} \setminus S_{j_p}^*$ for $p = 1, \ldots, x_i$). Thus, there exist at least

$$|C_3'| = |C_3| - \sum_{i=1}^{\Delta} (\Delta-(i+k_i))\, x_i = y\,|N_2^*| - \sum_{i=1}^{\Delta} (\Delta-(i+k_i))\, x_i \qquad (10)$$

elements of C_3 not covered by sets in N_1^*. Since initial instance $(\mathcal{S}, C)$ is non-trivial, elements of C_3' are also contained in sets N_3 either from $N^* \setminus N_1^*$, or from $\bar{N}$. So, the family N_3 has size at least $|C_3'|/\Delta$. Moreover, using (6), (10) and $y \leqslant 1$, we get:

$$|N_3| \geqslant \frac{y\,|N_2^*|}{\Delta} - \sum_{i=1}^{\Delta} \frac{(\Delta-(i+k_i))\, x_i}{\Delta} \geqslant y \sum_{i=1}^{\Delta} \frac{i(i-1)}{2\Delta^2} x_i + \sum_{i=1}^{\Delta} (\frac{i}{\Delta} - 1) x_i \qquad (11)$$

We so deduce: $\delta(I, \mathcal{S}') \geqslant (|N_1^*| + |N_3 \setminus \bar{N}| + |\bar{N}|)/(|N' \setminus N^*| + |\bar{N}|)$, which, taking into account that $|\bar{N}| \geqslant |\bar{N} \cap N_3|$, finally becomes:

$$\delta\,(I, \mathcal{S}') \geqslant \frac{|N_1^*| + |N_3|}{|N' \setminus N^*| + |N_3|} \qquad (12)$$

Note, furthermore, that function $(a+x)/(b+x)$ is increasing with x, for $a \leqslant b$ and $x > -b$. Therefore, using (2), (3), (11) and $y \geqslant (\Delta-1)/\Delta$, (12) becomes:

$$\delta\,(I, \mathcal{S}') \geqslant \frac{\sum_{i=1}^{\Delta} \left(\frac{(\Delta-1)\cdot i(i-1)}{2\Delta^3} + \frac{i}{\Delta} \right) x_i}{\sum_{i=1}^{\Delta} \left(i + \frac{(\Delta-1)\cdot i(i-1)}{2\Delta^3} + \frac{i}{\Delta} - 1 \right) x_i} \qquad (13)$$

Set now $f(x) = (\Delta-1)\cdot(x(x-1)/2\Delta^3) + (x/\Delta)$; (13) can now be expressed as:

$$\delta\,(I, \mathcal{S}') \geqslant \frac{\sum_{i=1}^{\Delta} f(i) x_i}{\sum_{i=1}^{\Delta} (f(i) + i - 1)\, x_i} \qquad (14)$$

With the same arguments, as for the convexity of f, we deduce from (14):

$$\delta\,(I, \mathcal{S}') \geqslant \frac{f\,(i_0)}{f\,(i_0) + i_0 - 1} = \frac{\frac{(\Delta-1)\cdot i_0(i_0-1)}{2\Delta^3} + \frac{i_0}{\Delta}}{i_0 + \frac{(\Delta-1)\cdot i_0(i_0-1)}{\Delta^3} + \frac{i_0}{\Delta} - 1} \qquad (15)$$

Recall that we have fixed $i_0 = \Delta/z$. If one assumes that Δ is arbitrarily large, one can simplify calculations by replacing $i_0 - 1$ by i_0. Then, (15) becomes:

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\frac{i_0^2}{2\Delta^2} + \frac{i_0}{\Delta}}{\frac{i_0^2}{2\Delta^2} + \frac{i_0}{\Delta} + i_0} \geqslant \frac{\frac{1}{2z^2} + \frac{1}{z}}{\frac{1}{2z^2} + \frac{1}{z} + \frac{\Delta}{z}} \approx \frac{1}{2z\Delta} + \frac{1}{\Delta} \tag{16}$$

Ratio given by (5) is increasing with z, while the one of (16) is decreasing with z. Equality of both ratios is reached when $2z^2 - 2z - 1 = 0$, i.e., for $z \approx 1.365$.

So, in any of the cases studied above, the differential approximation ratio achieved by SCGREEDY is greater than, or equal to, $1.365/\Delta$ and the proof of the theorem is now complete. □

Proposition 1. *There exist* MIN SET COVER-*instances where the differential approximation ratio of* SCGREEDY *is* $4/(\Delta + 2)$ *for any* $\Delta \geqslant 3$.

Proof. Assume a fixed $t > 1$, a ground set $C = \{c_{ij} : i = 1, \ldots, t-1, j = 2, \ldots, t, j > i\}$ and a system $\mathcal{S} = \{S_1, \ldots, S_t\}$, where $S_i = \{c_{ji} : j < i\} \cup \{c_{ij} : j > i\}$, for $i = 1, \ldots, t$. Denote by $I_t = (\mathcal{S}, C)$ the instance of MIN SET COVER defined on C and $\mathcal{S}$.

Remark that the smallest cover for C includes at least $t-1$ sets of $\mathcal{S}$. Indeed, consider a family $\mathcal{S}' \subseteq \mathcal{S}$ of size less than $t-1$. Then, there exists $i_0 < j_0$ such that neither S_{i_0}, nor S_{j_0} belong to $\mathcal{S}'$. In this case element $c_{i_0 j_0} \in C$ is not covered by $\mathcal{S}'$. Note finally that, for I_t, the maximum size of the subsets of $\mathcal{S}$ is $\Delta = t - 1$. Indeed, for any $i = 1, \ldots, t$, $|\{c_{ji} : j < i\}| = i - 1$ and $|\{c_{ij} : j < i\}| = t - i$; so, $|S_i| = t - 1$.

Fix an even Δ and build the following instance $(\mathcal{S}, C)$ for MIN SET COVER:

$$C = \left\{a_{ij}, a'_{ij} : i, j = 1, \ldots, \frac{\Delta}{2}, j > i\right\} \cup \left\{b_{ij} : i = 1, \ldots, \frac{\Delta + 2}{2}, j = 1, \ldots, \Delta\right\}$$

$$S_i^1 = \{a_{ji} : j < i\} \cup \{a_{ij} : j > i\} \cup \left\{b_{ji} : j = 1, \ldots, \frac{\Delta + 2}{2}\right\},\ i = 1, \ldots, \frac{\Delta}{2}$$

$$S_i^2 = \left\{a'_{ij} : j < i\right\} \cup \left\{a'_{ij} : j > i\right\} \cup \left\{b_{jk} : j = 1, \ldots, \frac{\Delta + 2}{2}, k = i + \frac{\Delta}{2}\right\},\ i = 1, \ldots, \frac{\Delta}{2}$$

$$\mathcal{S}^j = \left\{S_i^j : i = 1, \ldots, \frac{\Delta}{2}\right\},\ j = 1, 2$$

$$\mathcal{S}^3 = \left\{\left\{\{b_{ij} : j = 1, \ldots, \Delta\},\ i = 1, \ldots, \frac{\Delta + 2}{2}\right\} i = 1, \ldots, \frac{\Delta + 2}{2}\right\}$$

Set $\mathcal{S} = \mathcal{S}^1 \cup \mathcal{S}^2 \cup \mathcal{S}^3$. Notice that, $\forall S_i \in \mathcal{S}$, $|S_i| = \Delta$. Hence, during its first iteration, SCGREEDY can choose a set in $\mathcal{S}^3$. Such a choice does not reduce cardinalities of the remaining sets in this sub-family; so, during its first $(\Delta + 2)/2$ iterations, SCGREEDY can exclusively choose all sets in $\mathcal{S}^3$. Remark that such choices entail that the surviving instance is the union of two disjoint instances $I_{\Delta/2}$ (i.e., instances of type I_t, as the ones defined in the beginning of this section, with $t = \Delta/2$), induced by the sub-systems $(\mathcal{S}^1, \{a_{ij}\})$ and $(\mathcal{S}^2, \{a'_{ij}\})$. According to what has been discussed at the beginning of the section, any cover for such instances uses at least $(\Delta/2) - 1$ sets. So, for a set-cover $\mathcal{S}'$ computed by SCGREEDY (remark that $\mathcal{S}'$ is minimal), we finally have: $m(\hat{I}, \mathcal{S}') \geqslant (3\Delta/2) - 1$. Furthermore, since $\mathcal{S}^1 \cup \mathcal{S}^2$ is a cover for C: $\text{opt}(\hat{I}) = \Delta$; finally, $\omega(\hat{I}) = (3\Delta/2) + 1$. In all, $\delta(\hat{I}, \mathcal{S}') = 4/(\Delta + 2)$.

For the case of odd Δ (that is omitted here), it can be proved that there exist an instance of MIN SET COVER for which attains differential ratio at least $4/(\Delta + 1)$. □

3 Differential approximation for MIN WEIGHTED SET COVER

Consider an instance $I = (\mathcal{S}, C, \boldsymbol{w})$ of MIN WEIGHTED SET COVER, where $\boldsymbol{w}$ denotes the vector of the weights on the subsets of $\mathcal{S}$ and the following algorithm, denoted by in what follows:

- order sets in $\mathcal{S}$ in decreasing weight-order (i.e., $w_1 \geqslant \ldots \geqslant w_m$); let $N = \{1, \ldots, m\}$ be the set of indices in the (so-ordered) $\mathcal{S}$; set $N' = N$;
- for $i = 1$ to m: if $N' \setminus \{i\}$ covers C, then set $N' = N' \setminus \{i\}$;
- output N'.

Proposition 2. *achieves differential approximation ratio bounded below by $1/\Delta$. This bound is asymptotically tight.*

Proof. We use in what follows notations introduced in Section 2. Observe that $N \setminus N' = \bar{N} \cup (N^* \setminus N')$ and $N \setminus N^* = \bar{N} \cup (N' \setminus N^*)$ where we recall that $\bar{N} = N \setminus (N^* \cup N')$. Denoting, for any $i \in N$, by w_i the weight of S_i, and, for any subset $\mathcal{X} \subseteq N$, by $w_{\mathcal{X}}$, the total weight of the sets with indices in $\mathcal{X}$, i.e., the quantity $\sum_{i \in \mathcal{X}} w_i$, the differential approximation ratio of becomes

$$\delta(I, N') = \frac{w_{N \setminus N'}}{w_{N \setminus N^*}} \tag{17}$$

Let $C_c = \{c_j : \exists i \in N' \cap N^*, c_j \in S_i\}$ be the set of elements covered by $N' \cap N^*$ and let $\bar{C}_c = C \setminus C_c$ be the complement of C_c with respect to C. It is easy to see that both $N' \setminus N^*$ and $N^* \setminus N'$ cover $\bar{C}_c$. Obviously, $C'_1 \subseteq \bar{C}_c$ (recall that $C'_1 = \{c_i : i \in N' \setminus N^*\}$ is a set of non-redundant elements with respect to sets of $N' \setminus N^*$ and that any element of C'_1 is covered by sets in $N^* \setminus N'$).

Consider the sub-instance of I induced by $(N' \setminus N^* \cup N^* \setminus N', C'_1)$. Fix an index $i \in N^* \setminus N'$ and denote by $S^*_i = \{c_{i_1}, \ldots, c_{i_k}\}$ the restriction of S_i to C'_1, i.e., $S^*_i = S_i \cap C'_1$. Assume that $S^*_i \neq \emptyset$; as it will be understood just below if this is not the case, then the approximation ratio of will be even better. Obviously, since sets $i_1, \ldots, i_k$ have been chosen by (i.e., $\{i_1, \ldots, i_k\} \subseteq N'$), $w_{i_j} \leqslant w_i$ and, $k \leqslant \Delta$, we get: $\sum_{j=1}^{k} w_{i_j} \leqslant \Delta w_i$. Summing it for all $i \in N^* \setminus N'$, we obtain $w_{N' \setminus N^*} \leqslant \Delta w_{N^* \setminus N'}$ and then, $w_{N \setminus N^*} \leqslant \Delta w_{N^* \setminus N}$. Expression (17) suffices now to conclude the proof of the ratio. The proof of the tightness is omitted here. □

4 MAX HYPERGRAPH INDEPENDENT SET

An instance $(\mathcal{S}, C)$ of MIN SET COVER can also be seen as a hypergraph H where C is the set of its vertices and $\mathcal{S}$ is the set of its hyper-edges. Then MIN SET COVER consists of determining the smallest set of hyper-edges covering C. The "dual" of this problem is the well-known MIN HITTING SET problem, where, on $(\mathcal{S}, C)$ one wishes to determine the smallest subset of C hitting any set in $\mathcal{S}$. MIN HITTING SET and MIN SET COVER are approximate equivalent in both standard and differential paradigms (see, for example, [10]; the former is the same as the latter modulo the inter-change of the roles of $\mathcal{S}$ and C). On the other hand another well-known combinatorial problem is MAX HYPERGRAPH INDEPENDENT SET where given $(\mathcal{S}, C)$, one wishes to determine the largest subset C' of C such that no $S_i \in \mathcal{S}$ is a proper subset of C'. It is easy to see that for MAX HYPERGRAPH INDEPENDENT SET and MIN HITTING SET, the objective function of the one is an affine transformation of the objective function of the other, since a hitting set is the complement with respect to C of a hypergraph independent set. Consequently, the differential approximation ratios of these two problems coincide, and coincide also (as we have seen just above) with the differential approximation ratio of MIN SET COVER. Hence, the results of the previous sections identically apply for MAX HYPERGRAPH INDEPENDENT SET also.

References

1. Ausiello, G., D'Atri, A., Protasi, M.: On the structure of combinatorial problems and structure preserving reductions. In: Proc. ICALP'77. Lecture Notes in Computer Science, Springer-Verlag (1977)
2. Ausiello, G., D'Atri, A., Protasi, M.: Structure preserving reductions among convex optimization problems. J. Comput. System Sci. (1980) 136–153
3. Zemel, E.: Measuring the quality of approximate solutions to zero-one programming problems. Math. Oper. Res. (1981) 319–332
4. Demange, M., Paschos, V. Th.: On an approximation measure founded on the links between optimization and polynomial approximation theory. Theoret. Comput. Sci. (1996) 117–141
5. Johnson, D.S.: Approximation algorithms for combinatorial problems. J. Comput. System Sci. (1974) 256–278
6. Lovász, L.: On the ratio of optimal integral and fractional covers. Discrete Math. (1975) 383–390
7. Slavík, P.: A tight analysis of the greedy algorithm for set cover. In: Proc. STOC'96. (1996) 435–441
8. Demange, M., Grisoni, P., Paschos, V. Th.: Differential approximation algorithms for some combinatorial optimization problems. Theoret. Comput. Sci. (1998) 107–122
9. Hassin, R., Khuller, S.: z-approximations. J. Algorithms (2001) 429–442
10. Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., Protasi, M.: Complexity and approximation. Combinatorial optimization problems and their approximability properties. Springer, Berlin (1999)