Approximation of Quadratic Set Covering Problem

Bruno Escoffier, Peter L. Hammer

Approximation of the Quadratic Set Covering Problem[§]

Bruno Escoffier[*], Peter L. Hammer[†]

Résumé

Nous étudions dans cet article l'approximation polynomiale du problème de la couverture d'ensemble quadratique. Ce problème, qui apparaît dans de nombreuses applications, est une généralisation naturelle du problème usuel de la couverture d'ensemble. Nous montrons que ce problème est très difficile à approcher dans le cas général, et même dans des cas particuliers classiques (quand la taille des ensembles ou quand la fréquence des éléments est bornée par une constante). Nous nous focalisons ensuite sur le cas convexe et nous donnons des résultats d'approximation à la fois positifs et négatifs. Dans un dernier temps, nous étudions la version non pondérée de ce problème.

Mots-clefs : NP-complétude, approximation polynomiale, couverture d'ensemble

Abstract

We study in this article polynomial approximation of the Quadratic Set Covering problem. This problem, which arises in many applications, is a natural generalization of the usual Set Covering problem. We show that this problem is very hard to approximate in the general case, and even in classical subcases (when the size of each set or when the frequency of each element is bounded by a constant). Then we focus on the convex case and give both positive and negative approximation results. Finally, we tackle the unweighted version of this problem.

Key words : NP-completeness, polynomial approximation, set covering

[*]LAMSADE, Université Paris-Dauphine, 75775 Paris cedex 16, France. `{escoffier}@lamsade.dauphine.fr`

[†]RUTCOR, Rutgers, the State University of New Jersey, 640 Bartholomew Road Piscataway, NJ 08854-8003. `hammer@rutcor.rutgers.edu`

[§]This work has been accomplished while Bruno Escoffier was a visitor at DIMACS, as part of a partnership between DIMACS and LAMSADE, supported by NSF and CNRS.

1 Introduction

The Set Covering problem is one of the most famous problems in complexity and approximation theory. Given a set $\mathcal{C} = \{c_1, \cdots, c_n\}$ of elements and a collection $\mathcal{S} = \{S_1, \cdots, S_m\}$ of subsets of $\mathcal{C}$, the goal is to find a subset $\mathcal{S}' \subseteq \mathcal{S}$ of minimum cardinality such that $\cup_{S_j \in \mathcal{S}'} S_j = \mathcal{C}$. In the Weighted Set Covering problem, a weight w_i being given for every set S_i, we want to minimize $\sum_{S_i \in \mathcal{S}'} w_i$ or, equivalently, $\sum_{i=1}^{m} w_i s_i$ where $s_i = 1$ if S_i is chosen in the solution and $s_i = 0$ otherwise.

We study in this paper a generalization of this problem where we change the objective function. We want to find a cover which minimizes, instead of a linear function $\sum_{i=1}^{m} w_i s_i$, a quadratic function $\sum_i w_i s_i + \sum_{i<j} w_{ij} s_i s_j$ (see section 2.2 for a formal definition).

The main motivation of this study comes from Logical Analysis of Data [5, 8], which is a methodology, based on a logical analysis, to detect structural information about datasets. A typical situation arising in this research area ([13]), coming from medicine, is the following: we want to determine if an illness can be related to some other medical parameters of patients, such as finding a correlation between heart attack and cholesterol for instance. To analyze this, we collect data on these parameters for both ill and healthy people. More formally, each person gives data on several criteria (weight, cholesterol,...) and is represented as a point in $\mathbb{R}^p$ (p is the number of criteria). Thus we have a set Ω_+ of positive points (ill people) and a set Ω_- of negative points (healthy people). A first step in the analysis of this data produces a collection of positive and negative patterns. A positive (resp. negative) pattern is a hypercube in $\mathbb{R}^p$ which contains no negative points (resp. no positive points). This collection is such that every point is covered. From a medical point of view, we would like to find a sub-collection of patterns such that:

- every point is covered,
- the volume of intersections between positive and negative patterns is as small as possible (for criteria relevance).

If we define w_{ij} as the volume of the intersection between the positive pattern $S_i^+ \in \mathcal{S}^+$ and the negative pattern $S_j^- \in \mathcal{S}^-$, then the problem is to find a cover of all points such that $\sum_{i,j} w_{ij} s_i^+ s_j^-$ is minimized. This is an instance of Quadratic Set Covering.

This problem also arises in other applications, such as the location of access points in a wireless network ([1, 2]), the facility layout problem ([4]), or line planning in public transports ([6, 16]).

Several heuristic techniques have been used to solve this problem but, to our knowledge, it has not been studied from the point of view of polynomial approximation so far.

This paper is organized as follows.

We recall in section 2 some basic definitions of approximation and some of the most important results dealing with the approximability properties of the standard Set Covering problem, and we define precisely the Quadratic Set Covering problem.

In section 3, we show that, without adding assumptions on weights, the problem is almost not approximable at all (not $2^{q(|\mathcal{C}|)}$-approximable for any polynomial q), even in the case of bounded degree instances (bounds on frequency of elements and on size of sets).

We focus in section 4 on the convex case. In this case, we can use classical continuous relaxation techniques to get good approximation ratios. More precisely, in the convex case, Quadratic Set Covering is approximable within ratio $O(\log^2(|\mathcal{C}|))$ by a randomized algorithm, and f^2 approximable (by a deterministic algorithm) if the frequency of each element is bounded by f. We also provide lower bounds which match somehow these upper bounds: Convex Quadratic Covering is not $c\log^2(|\mathcal{C}|)$-approximable in the general case, for some constant $c > 0$, unless $P = NP$, and not $(f-1)^2 - \varepsilon$-approximable if the frequency of each element is bounded by f, unless $P = NP$.

We conclude the paper in the last section, where we tackle the unweighted version of Quadratic Set Covering (the weights are either 0 or 1).

2 Preliminaries

2.1 Polynomial approximation and Set Covering

Let's now recall briefly standard definitions on polynomial approximation of NP-hard optimization problems.

Definition 1 *An algorithm A is a ρ-approximation for a minimization problem Π if and only if for any instance I of problem Π, A computes (in polynomial time) a feasible solution $A(I)$ such that*

$$m_\Pi(A(I)) \leq \rho \times opt_\Pi(I)$$

where $m_\Pi(.)$ is the objective function and $opt_\Pi(.)$ is the value of an optimal solution.

Definition 2 *A problem Π is ρ-approximable if and only if there exists an algorithm which is a ρ-approximation.*

The Set Covering problem has been widely studied from the polynomial approximation point of view, since the greedy algorithm given by Johnson [17]. Let's mention the following results:

- The Set Covering problem is approximable within ratio $1+\ln(|\mathcal{C}|)$ [17], even in the weighted case ([7]).
- Raz and Safra [20] showed that the Set Covering problem is inapproximable within ratio $c\ln(|\mathcal{C}|)$ for some $c>0$, unless $P=NP$. Feige [11] proved that the problem is not $(1-\varepsilon)\ln(|\mathcal{C}|)$-approximable, unless $NP\subseteq QP$.
- If the cardinality of each set S_i is bounded above by a constant $k\geq 3$, then the problem is $1+\ln(k)$-approximable [19] (this result has been improved by Duh and Fürer [10] who gave a $\sum_{i=1}^{k} 1/i-1/2$-approximation), but not $(\ln(k)-c\ln(\ln(k)))$ approximable, for some constant c, unless $P=NP$ ([23]).
- If each element appears in at most f sets ($f\geq 2$), then the problem is f-approximable, even in the weighted case ([3]), and is not $f-1-\varepsilon$-approximable, for any $\varepsilon>0$, unless $P=NP$ ([9]).

2.2 Quadratic Set Covering

As mentioned in the introduction, we are interested in a generalization of Set Covering called Quadratic Set Covering. In this problem, we are given:

- a set $\mathcal{C}=(c_1,\cdots,c_n)$ of elements;
- a set $\mathcal{S}=(S_1,\cdots,S_m)$ of subsets of $\mathcal{C}$ such that $\bigcup_{i=1}^{m} S_i=\mathcal{C}$;
- A weight $w_i>0$, $i=1,\cdots,m$ and a weight $w_{ij}\geq 0$ for all $i<j$, $i,j=1,\cdots,m$ (we chose to consider in all this paper, without loss of generality, $w_{ij}=0$ for $i>j$).

The goal is to find a cover of $\mathcal{C}$, i.e. a subset $\mathcal{S}'$ of $\mathcal{S}$ with $\bigcup_{i|S_i\in\mathcal{S}'} S_i=\mathcal{C}$, such that $m_{QSC}(\mathcal{S}')$ is minimized, where:

$$m_{QSC}(\mathcal{S}')=\sum_{i|S_i\in\mathcal{S}'} w_i+\sum_{i<j,(S_i,S_j)\in\mathcal{S}'} w_{ij}$$

Note that the restriction to positive weights is natural and classical in polynomial approximation theory, since it ensures that the value of any solution is strictly positive. We allow quadratic weights w_{ij} to be 0 in order to capture the usual Set Covering problem as a particular case (set $w_i=1$ for all i, and $w_{ij}=0$ for all (i,j)). However, all the results would be the same under the restriction $w_{ij}>0$.

We will also use another natural formulation of Quadratic Set Covering, as a mathematical program. Let A be the $n\times m$ matrix where $A_{ij}=1$ if $c_i\in S_j$, $A_{ij}=0$ otherwise.

Then Quadratic Set Covering can be expressed by the following program (which will be denoted also by QSC):

$$(QSC)\begin{cases}\min\ \sum_{i=1}^{m} w_i s_i + \sum_{i<j, i,j=1,\cdots,m} w_{ij} s_i s_j \\ s.t. \left| \begin{array}{l} As \geq \mathbf{1}_n \\ s \in \{0,1\}^m \end{array}\right.\end{cases}$$

where $\mathbf{1}_n$ denotes the vector $[1, 1, \cdots, 1]$ of $\mathbb{R}^n$. Since $s_i^2 = s_i$ ($s_i \in \{0,1\}$), we can write the objective function with the matrix W of weights (with $w_{ii} = w_i$, $i = 1, \cdots, m$ and $w_{ij} = 0$ for $i > j$):

$$(QSC)\begin{cases}\min\ s^t W s \\ s.t. \left| \begin{array}{l} As \geq \mathbf{1}_n \\ s \in \{0,1\}^m \end{array}\right.\end{cases}$$

We will also study the so-called convex case, classical when dealing with the minimization of a mathematical program, i.e. when the matrix W is positive semi-definite (when for any vector $v \in \mathbb{R}^m$, $v^t W v \geq 0$).

3 Approximation of Quadratic Set Covering: general case

We investigate in this section the approximation properties of Quadratic Set Covering. We show that this problem is very hard to approximate, even in the case where the size of each set is bounded, and in the case where the frequency of each element is bounded. More precisely, the following property holds:

Proposition 3 *Quadratic Set Covering is not approximable within approximation ratio $2^{q(|\mathcal{C}|)}$, for any polynomial q, unless $P = NP$. This result remains true even if each element appears at most 3 times and each set contains one element.*

This result follows from a reduction from the problem of 3-colorability. In this problem, we are given a graph G and we want to determine whether we can color the vertices of G with 3 colors such that two any adjacent vertices don't have the same color. This problem is NP-complete ([12]).

Consider a graph G with n vertices. We construct an instance of Quadratic Set Covering with n elements c_i and $3n$ sets $S_{ij}, i = 1 \cdots n, j = 1 \cdots 3$. Element c_i belongs to set $S_{ij}, j = 1 \cdots 3$.

The weights are the following:

- The weight $w_{(ij)}$ relative to set S_{ij} is 1.
- For any edge (v_i, v_j) in G and any $k = 1, \cdots, 3$, the weight $w_{(ik)(jk)}$ corresponding to sets S_{ik} and S_{jk} is nM (where M is a large positive integer to be specified).
- All other weights are 0.

In other words, the objective function to minimize (expressed as a quadratic function of variables s_i) is:

$$\sum_{i=1}^{n} \sum_{j=1}^{3} s_{ij} + nM \sum_{(i,j)\in E} \sum_{k=1}^{3} s_{ik}s_{jk}$$

We claim that if G is 3-colorable, then $opt(\mathcal{C}, \mathcal{S}, W) \leq n$, else $opt(\mathcal{C}, \mathcal{S}, W) \geq nM$.

First, consider a 3-coloring of G. Take S_{ik} in the cover if and only if v_i is colored with color k. Then we can easily see that each element is covered. Moreover, $s_{ik}s_{jk} = 0$ for all $(i, j) \in E$ (since it's a coloring). Hence, the value of this cover is $\sum_{i=1}^{n} \sum_{j=1}^{3} s_{ij} = n$, and thus $opt(\mathcal{C}, \mathcal{S}) \leq n$.

Conversely, consider a solution of Quadratic Set Covering with value at most $nM - 1$. For any i, at least one s_{ik}, $k = 1, 2, 3$, is equal to 1; we color v_i with color k for one k such that $s_{ik} = 1$. Since $s_{ik}s_{jk} = 0$ for $(i, j) \in E$, this is a proper coloring and the graph is 3-colorable.

This transformation is polynomial for $M = 2^{q(n)}$, so we get an inapproximability result of $2^{q(n)} = 2^{q(|\mathcal{C}|)}$ for any polynomial q.

Note that we have, since $|\mathcal{S}| = 3|\mathcal{C}|$, an analogous result for approximation ratios which depend on $|\mathcal{S}|$.

Proposition 4 *Quadratic Set Covering is not approximable within approximation ratio $2^{q(|\mathcal{S}|)}$, for any polynomial q, unless $P = NP$.*

4 Approximation in the convex case

Since Quadratic Set Covering is in some way too hard to be approximated, we now focus on the case where the objective function is convex (which is a classical particular case when dealing with minimization of functions). As mentioned in section 2.2, the function is convex when $v^tWv \geq 0$ for any vector v.

4.1 Lower bounds

Since the problem is a generalization of the standard Set Covering, all the lower bounds for the latter problem hold, obviously, for the former. So, we already have a trivial inapproximability result of $(1-\varepsilon)\ln(|\mathcal{C}|)$ ([11]), but we can improve this result, using the following proposition:

Proposition 5 *If Set Covering is hard to approximate within ratio ρ on a set of instances, then Convex Quadratic Set Covering is hard to approximate within ratio ρ^2 on the same set of instances.*

This result easily follows from the following reduction. Suppose that Set Covering is hard to approximate within ratio ρ on a set of instances, and consider an instance $(\mathcal{S},\mathcal{C})$ in this set. We consider the instance of Quadratic Set Covering with the same sets $\mathcal{S}$ and $\mathcal{C}$, and with weights $w_{ii}=1$ $(i=1,\cdots,m)$ and $w_{ij}=2$ $(i,j=1,\cdots,m,\ i<j)$.

Then, for a vector $v\in\mathbb{R}^m$, $v^tWv=\sum_{i=1}^m v_i^2+\sum_{i<j}2v_iv_j=(\sum_{i=1}^m v_i)^2$. Hence:

- we are in the convex case;
- a cover of size k has value k^2 in this instance of Quadratic Set Covering.

Thus, a ρ approximation algorithm for the usual Set Covering problem is a ρ^2 approximation algorithm for Quadratic Set Covering, and vice versa. The result follows.

With the inapproximability results of [20] and [11], we get:

Corollary 6 *Convex Quadratic Set Covering is not $(1-\varepsilon)\ln^2(|\mathcal{C}|)$-approximable unless $NP\subseteq QP$. It is not $c\ln^2(|\mathcal{C}|)$-approximable, for some constant c, unless $P=NP$.*

With the results of [9] and [23] on bounded degree instances, we obtain:

Corollary 7 *When restricted to instances where each element appears in at most f sets ($f\geq 2$), Convex Quadratic Set Covering is not $(f-1)^2-\varepsilon$-approximable, for any $\varepsilon>0$, unless $P=NP$.*

Corollary 8 *When restricted to instances where the size of each set is at most k, Convex Quadratic Set Covering is not $(\ln(k)-c\ln(\ln(k)))^2$-approximable, for some constant c, unless $P=NP$.*

4.2 Positive results

We present here some positive approximation results. Algorithms are based on a resolution of a continuous relaxation of the problem, which is a classical technique used to solve the standard Set Covering problem (see [15] for instance).

As mentioned in section 2, Convex Quadratic Set Covering can be expressed by the following mathematical program:

$$(QSC)\ \begin{cases} \min\ s^t W s \\ s.t. \left| \begin{array}{l} As \geq \mathbf{1}_n \\ s \in \{0,1\}^m \end{array} \right. \end{cases}$$

We are interested in a continuous relaxation of this integer program, noted $(LQSC)$:

$$(LCQSC)\ \begin{cases} \min\ s^t W s \\ s.t. \left| \begin{array}{l} As \geq \mathbf{1}_n \\ s \in [0,1]^m \end{array} \right. \end{cases}$$

The general problem of Min Quadratic Programming (minimization of a quadratic function under linear constraints with real variables) is known to be in NP ([24]), but is NP-hard in the general case ([21]). However, Kozlov et al. [18] proved that it's polynomially solvable in the convex case.

We use this property to obtain, with usual rounding techniques, positive approximation results: an f^2-approximation algorithm when each element appears in at most f sets, and, as a consequence, an m^2-approximation algorithm in the general case. We improve the latter result by showing that a classical randomized algorithm for Set Cover gives, when applied to Quadratic Set Cover, an $O(\log^2(n))$-approximate solution with probability at least 1/2. Note that these upper bounds, although found by simple techniques, are relatively close to the corresponding lower bounds ($(f-1)^2 - \varepsilon$ unless $P = NP$ and $(1-\varepsilon)\ln^2(|\mathcal{C}|)$ unless $NP \subseteq QP$ respectively).

Proposition 9 *If each element appears in at most f sets, then Convex Quadratic Set Covering is approximable within ratio f^2.*

Let s^* be an optimal solution of the continuous relaxation $(LCQSC)$ of an instance $(CQSC)$ of Convex Quadratic Set Covering. We use the following standard rounding technique ([14]). Let f be the maximal number of sets in which an element appears. Set $s_i = 1$ if $s_i^* \geq \frac{1}{f}$, $s_i = 0$ otherwise. This solution is a valid cover since, for each element

c_i, the constraint $\sum_j A_{ij} s_j \geq 1$ implies that for at least one S_j containing c_i, $s_j^* \geq 1/f$. Moreover, since, for all i, $0 \leq s_i \leq f s_i^*$:

$$m(s) = s^t W s \leq f^2 s^{*t} W s^* = f^2 m(s^*)$$

Since $opt_{(LQSC)} \leq opt_{(QSC)}$, the result follows.

We immediately get the following corollary.

Corollary 10 *Quadratic Set Covering is approximable within ratio* m^2.

Let's recall briefly the following well known randomized algorithm for Set Covering (see [15]):

- Find an optimal solution s^* to the $[0,1]$ relaxation of Set Covering;
- For any i, flip between 0 and 1 with probability s_i^* to get 1, $c\log(n)$ times independently (for a suitable constant c). Take set S_i in the solution when you get 1 at least one time among the $c\log(n)$ flips.

The probability that the output solution $\mathcal{S}'$ is a cover is at least 3/4, and the probability that this solution is a $O(\log(n))$-approximation is at least 3/4.

We would like to use this algorithm for Convex Quadratic Set Covering, but we cannot use directly the continuous relaxation $(LCQSC)$ to get a similar result. Actually, we have to modify the objective function. Let's consider the following mathematical program $(LCQSC')$:

$$(LCQSC') \left\{ \begin{array}{l} \min \ \sum_{i=1}^m w_{ii} s_i + s^t W s \\ s.t. \left| \begin{array}{l} As \geq \mathbf{1}_n \\ s \in [0,1]^m \end{array} \right. \end{array} \right.$$

In other words, we add to the objective function the linear term $\sum_{i=1}^m w_{ii} s_i$. W hasn't changed, hence we can solve optimally $(LCQSC')$ (the function is still convex). Let s^* be an optimal solution of $(LCQSC')$, and $\hat{s}$ be an optimal solution to the initial integer Quadratic Set Covering (QSC). We have:

$$m_{(LCQSC')}(s^*) \leq m_{(LCQSC')}(\hat{s}) = 2\sum_{i=1}^m w_{ii}\hat{s}_i + \sum_{i<j} w_{ij}\hat{s}_i\hat{s}_j \leq 2m_{(QSC)}(\hat{s})$$

Consider now the solution given by the randomized algorithm, when choosing s_i^* for the flips probabilities.

Since constraints are the same as in the Set Covering problem, we get here also a solution $\mathcal{S}'$ which is a cover with probability at least $3/4$. The probability $P(S_i)$ that S_i is chosen in the solution $\mathcal{S}'$ is at most $c\log(n)s_i^*$, hence:

$$\begin{aligned}
E_{(QSC)}(\mathcal{S}') &= \sum_{i=1}^{m} w_{ii}P(S_i) + \sum_{i<j} w_{ij}P(S_i)P(S_j) \\
&\leq c\log(n)\sum_{i=1}^{m} w_{ii}s_i^* + c^2\log^2(n)\sum_{i<j} w_{ij}s_i^*s_j^* \\
&\leq c^2\log^2(n)m_{(LCQSC')}(s^*) \leq 2c^2\log^2(n)m_{(QSC)}(\hat{s})
\end{aligned}$$

Then, using Markov's inequality, we get that $\mathcal{S}'$ is a $O(\log^2(n))$-approximation, with probability at least 3/4.

This leads to the following result:

Proposition 11 *There exists a (polynomial) randomized algorithm which outputs, with probability at least 1/2, a cover which is an* $O(\log^2(|\mathcal{C}|))$*-approximation for Convex Quadratic Set Covering.*

5 Discussion

We gave in this article the first results, in terms of polynomial approximation, for the Quadratic Set Covering problem. Following our analysis, the first question that comes to mind may be to determine whether the problem is constant approximable or not in the convex case when the size of each set is bounded by a constant k. This fact is trivial for the usual Set Covering problem since any solution is a k-approximation.

Another interesting question concerns the approximation of 0-1 Quadratic Set Covering, that is the subcase of Quadratic Set Covering where all the weights are either 0 or 1. We can think of several natural versions of 0-1 Quadratic Set Covering.

First, if all the w_i and all the w_{ij} are equal to 1 (which can be considered as the most restricted quadratic generalization of Set Covering), the problem is trivially related to Set Covering, since each solution of size k has value $k(k+1)/2$. Hence, positive and negative results are transferred (from Set Covering to this version of Quadratic Set Covering) from

ρ to, roughly speaking, ρ^2.

The question of 0-1 Quadratic Set Covering seems more interesting when $w_i = 1$ (as in the usual Set Covering problem) but $w_{ij} \in \{0, 1\}$. We have obviously a trivial lower bound of $O(\ln^2(n))$.

For positive results, we first note that this version of 0-1 Quadratic Set Covering is not necessarily convex. To see this, consider the minimization of $\sum_{i=1}^m s_i^2 + \sum_{i=2}^m s_1 s_i$: this function is negative, as soon as $m \geq 6$, for $s_1 = 1$ and $s_i = -1/2$, $i = 1, \cdots, m$. Hence, we cannot use the approximation algorithms given in section 4.

We can however show the following:

Proposition 12 *When $w_i = 1$ (for all i) and $w_{ij} \in \{0, 1\}$ (for all $i < j$), Quadratic Set Covering is approximable within ratio $O(|\mathcal{C}|)$.*

This follows from a result of [22] who showed that Set Covering can be approximated within ratio $O(\ln(n/opt))$ ($n = |\mathcal{C}|$). If we apply this algorithm on an instance $(\mathcal{S}, \mathcal{C}, W)$ of 0-1 Quadratic Set Covering, we get a solution $\mathcal{S}'$ such that:

$$m_{0-1QSC}(\mathcal{S}') \leq m_{SC}^2(\mathcal{S}') \leq O\left(\ln^2(n/opt_{SC}(\mathcal{S}, \mathcal{C}))opt_{SC}^2(\mathcal{S}, \mathcal{C})\right)$$

Since $w_i = 1$ for all i, $opt_{SC}(\mathcal{S}, \mathcal{C}) \leq opt_{0-1QSC}(\mathcal{S}, \mathcal{C}, W)$, thus:

$$\frac{m_{0-1QSC}(\mathcal{S}')}{opt_{0-1QSC}(\mathcal{S}, \mathcal{C}, W)} \leq O\left(\ln^2(n/opt_{SC}(\mathcal{S}, \mathcal{C}))opt_{SC}(\mathcal{S}, \mathcal{C})\right) \leq O(n)$$

Bridging the gap between the $O(\ln^2(n))$ lower bound and the $O(n)$ approximation algorithm seems to be an interesting question.

Finally, we can also consider the most general version of 0-1 Quadratic Set Covering where weights belong to $\{0, 1\}$, without any other restriction (note that this doesn't fulfil the requirements of our definition of Quadratic Set Covering since w_i may be 0). In this case, we cannot provide any approximation guarantee since we have the following result:

Proposition 13 *When all weights belong to $\{0, 1\}$ (with no other restriction), then it is NP-hard to decide whether the optimal value of Quadratic Set Covering is 0 or not.*

We prove this last result with a reduction from the very well known problem Sat. In this problem, we are given a set of m binary variables $x_1, \cdots, x_m$ and a set of n clauses $C_1, \cdots, C_n$, and we want to determine whether there exists a truth assignment which satisfies all the clauses. Starting from such an instance of Sat, we construct the following instance of 0-1 Quadratic Set Covering:

- there are $m+n$ elements $c_i, \cdots, c_m, v_1, \cdots, v_n$ and $2n$ sets $s_1^+, \cdots, s_n^+, s_1^-, \cdots, s_n^-$;
- if x_j appears positively (resp. negatively) in C_i, then element c_i is in s_j^+ (resp. in s_j^-);
- element v_i is in s_i^+ and s_i^-;
- 'quadratic' weight $w_{i^+i^-}$ corresponding to sets (s_i^+, s_i^-) is 1, all other weights are 0.

We claim that the optimum of this instance of 0-1 Quadratic Set Covering is 0 if and only if the Sat formula is satisfiable.

If we have a truth assignment which satisfies all the clauses, then consider in the solution the set s_i^+ if x_i is true, s_i^- otherwise. The value of this solution is obviously 0; moreover, element c_i is covered since C_i is satisfied, and v_i is covered since s_i^+ or s_i^- is in the solution.

Conversely, if there is a cover with value 0, then exactly one of the sets s_i^+ and s_i^- is in the solution (with similar arguments). We set x_i to true if s_i^+ is in the solution, to false otherwise. Covering element c_i means that clause C_i is satisfied, hence the truth assignment satisfies the formula.

Note that, starting from restricted versions of Sat shows that proposition 13 still holds when the frequency of each element or when the size of each set is bounded by a constant.

References

[1] Amaldi E., Capone A., Cesana M., Malucelli F. and Palazzo F.: WLAN Coverage Planning: Optimization Models and Algorithms. Proceedings of the IEEE Vehicular Technology Conference (VTC Spring), 2004.

[2] Amaldi E., Capone A., Cesana M., Malucelli F. and Palazzo F.: Optimizing WLAN Radio Coverage. Proceedings of the IEEE International Conference on Communications, 2004.

[3] Bar-Yehuda R. and Even S.: A linear-time approximation algorithm for the weighted vertex cover problem. J. Algorithms, **2** (1981):198-203.

[4] Bazaraa M. S.: Computerized layout design: A branch and bound approach. AIIE Transactions, **7**(4) (1975):432-438.

[5] Boros E., Hammer P. L., Ibaraki T., Kogan A., Mayoraz E. and Muchnik I. B.: An Implementation of Logical Analysis of Data. IEEE Trans. Knowl. Data Eng. **12**(2) (2000):292-306.

[6] Ceder A. and Israeli Y.: Scheduling considerations in designing transit routes at the network level. Proceedings of the Fifth International Workshop on Computer-Aided Scheduling of Public Transport, 1990. Lecture Notes in Economics and Mathematical Systems, **386** (1992):113-136.

[7] Chvátal V.: A greedy heuristic for the set covering problem. Math. Oper. Res. **4** (1979):233–235.

[8] Crama Y., Hammer P. L. and Ibaraki T.: Cause-effect relationships and partially defined boolean functions. Annals of Operations Research, **16** (1988):299–326.

[9] Dinur I., Guruswami V., Khot S. and Regev O.: A new multilayered PCP and the hardness of hypergraph vertex cover. Proceedings of STOC'03 (2003):595–601.

[10] Duh R. and Fürer M.: Approximation of k-set cover by semi-local optimization. Proceedings of STOC'97 (1997):256-265.

[11] Feige U.: A threshold of $\ln(n)$ for approximating set cover. J. ACM **45** (1998):634–652.

[12] Garey M. R., Johnson D.S. and Stockmeyer L.: Some simplified NP-complete graph problems. Theoretical Computer Science **1** (1976):237–267.

[13] Hammer P. L.: Non linear set covering problems in Logical Analysis of Data. SIAM conference on optimization, 2005.

[14] Hochbaum D. S.: Approximation Algorithms for the Set Covering and Vertex Cover Problems. SIAM Journal on Computing. **11**(3) (1982):555-556.

[15] Hochbaum D. S. (editor): Approximation Algorithms for NP-Hard Problems. PWS Publishing Company, Boston, MA. (1997).

[16] Israeli Y. and Ceder A.: Transit route design using scheduling and multiobjective programming techniques. Proceedings of the Sixth International Workshop on Computer-Aided Scheduling of Public Transport. Lecture Notes in Economics and Mathematical Systems. **430** (1995):56-75.

[17] Johnson D. S.: Approximation algorithms for combinatorial problems. J. Comput. System Sci. **9** (1974):256–278.

[18] Kozlov M. K., Tarasov S. P. and Khachiyan L. G.: Polynomial solvability of convex quadratic programming. Dokl. AKad. Nauk SSSR **248** (1979):1049–1051. Translated in Soviet Math Dokl. **20** (1979):1108–1111.

[19] Lovász L.: On the ratio of optimal integral and fractional covers. Discrete Mathematics **13** (1975):383–390.

[20] Raz R. and Safra S.: A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP. Proceedings of STOC'97, (1997):475–484.

[21] Sahni S.: Computationnaly related problems. SIAM J. of Computing **3**(4) (1974):262–279.

[22] Srinivasan A.: Improved approximations of packing and covering problems. Proceedings of STOC'95, (1995):268–276.

[23] Trevisan L.: Error-Correcting Codes and Pseudorandom Projections. Proceedings of RANDOM-APPROX (2001):7–9.

[24] Vavasis S.: Quadratic Programming is in NP. Information Processing Letters **36** (1990):73–77.