Restricted Parameter Range Promise Set Cover Problems Are Easy

Hao Chen
Software Engineering Institute
East China Normal University
Shanghai 200062, P.R. China
E-mail: haochen@sei.ecnu.edu.cn

September, 2010

Abstract

Let \((U, S, d)\) be an instance of Set Cover Problem, where \(U = \{u_1, ..., u_n\}\) is a \(n\) element ground set, \(S = \{S_1, ..., S_m\}\) is a set of \(m\) subsets of \(U\) satisfying \(\bigcup_{i=1}^m S_i = U\) and \(d\) is a positive integer. In STOC 1993 M. Bellare, S. Goldwasser, C. Lund and A. Russell proved the NP-hardness to distinguish the following two cases of \textbf{GapSetCover}_\(\eta\) for any constant \(\eta > 1\). The Yes case is the instance for which there is an exact cover of size \(d\) and the No case is the instance for which any cover of \(U\) from \(S\) has size at least \(\eta d\). This was improved by R. Raz and S. Safra in STOC 1997 about the NP-hardness for \textbf{GapSetCover}_{c\log m} for some constant \(c\). In this paper we prove that restricted parameter range subproblem is easy. For any given function of \(n\) satisfying \(\eta(n) \geq 1\), we give a polynomial time algorithm not depending on \(\eta(n)\) to distinguish between

\[\textbf{YES}: \text{The instance } (U, S, d) \text{ where } d > \frac{2|S|}{\eta(n)-1}, \text{ for which there exists an exact cover of size at most } d;\]

\[\textbf{NO}: \text{The instance } (U, S, d) \text{ where } d > \frac{2|S|}{\eta(n)-1}, \text{ for which any cover from } S \text{ has size larger than } \eta(n)d.\]

Thus the large part subproblem of the NP-hard promise set cover problem is actually easy. The polynomial reduction of this restricted parameter range set cover problem is constructed by using the lattice.

\textbf{Index Terms}—Set Cover Problem, Lattice, Closest vector Problem(CVP), Shortest vector problem(SVP)
1 Introduction

**Set Cover Problem** is a classical combinatorial optimization problem. The instance of the problem is \((U, S)\), where where \(U = \{u_1, \ldots, u_n\}\) is a \(n\) element ground set, \(S = \{S_1, \ldots, S_m\}\) is a set of \(m\) subsets. The goal is to find the minimal size \(S' \subseteq S\) such that \(\bigcup_{S_i \in S'} S_i = U\). The classical result says that there is the polynomial time algorithm approximating the optimal solution to a factor \(t\), where \(t = \max_i \{|\{j : u_i \in S_j\}\}|\) is the largest number of the subsets in \(S\) for which one element in \(U\) may belong (see [10, 12, 15]). The greedy algorithm with polynomial time can also be used for approximating the optimal solution to a \(H_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k} \leq 1 + \ln k\) factor (see [5]), where \(k = \max_{i=1}^m |S_i|\) is the size of the largest set.

On the other hand, from the NP-completeness of 3-dimensional matching problem it is well-known that the following exact covering problem by 3-sets is NP-complete: given an instance \((U, S)\), where \(U\) is a ground set of \(3n\) elements and \(S\) is a collection of subsets with 3 elements, the goal is to determine if there is a sub-collection \(S' \subseteq S\) of pairwise disjoint subsets (an exact cover) such that \(\bigcup_{S \in S'} S = U\). In STOC 1993, Bellare, Goldwasser, Lund and Russell proved that the promise problem of approximating set cover to any constant factor is NP hard in [4] (see also [16]). Explicitly they proved that for any positive constant \(\eta > 1\) it is NP-hard to distinguish between the following YES and NO instances. The YES instance is the instance for which there is an exact cover of size \(d\), that is, there exist pairwise disjoint subsets \(S_{i_1}, \ldots, S_{i_d}\) in \(S\) satisfying \(\bigcup_{j=1}^d S_{i_j} = U\), and the No instance is the instance for which any cover of \(U\) from \(S\) has size at least \(\eta d\). The result of R. Raz and S. Safra in STOC 1997 ([19]) implies the NP-hardness of the promise problem \(\text{GapSetCover}_{c \ln m}\) for some constant \(c\). Feige [9] proved that there cannot be a \((1 - \varepsilon) \ln m\) approximate algorithm for the original set cover problem, for any \(\varepsilon > 0\), unless \(\text{NP} \subseteq \text{QP}\). Trevisan [20] indicated that Feige’s proof also implies that there is a constant \(c\) such that the Set Cover problem with sets of size \(k\) (where \(k\) is constant) has no \((\ln k - c \ln \ln k)\)-approximate algorithm for the original set cover problem unless \(\text{NP} = \text{P}\).

A variant of **Set Cover Problem** is the following vertex cover problem for \(k\)-uniform hypergraphs. An edge in a hypergraph is a subset of the vertices. A \(k\)-uniform hypergraph is \(G = (V, E)\), where \(V\) is the set of \(n\) vertices and the \(E\) is set of edges and each edge in \(E\) is a \(k\) element subset of \(V\). The **Vertex Cover Problem** for \(k\)-uniform hypergraphs \((k\) is a con-
stant) is defined as follows. For any given $k$-uniform hypergraph, to find the minimum size subset of vertices $V' \subseteq V$ such that $V' \cap e \neq \emptyset$ for each edge $e$. When $k = 2$ it is the classical vertex cover problem. There is a polynomial time greedy algorithm approximating the **Vertex Cover Problem** for $k$-uniform hypergraphs to a factor $k$. Approximating the vertex cover problem within a factor $k - 1 - \varepsilon$ for any $\varepsilon > 0$ and $k \geq 3$ was proved NP-hard in [7]. Khot and Regev proved that approximating the vertex cover problem for $k$-uniform hypergraphs to the factor $k - \varepsilon$ for any $\varepsilon > 0$ is NP-hard under the assumption that the **Unique Game Conjecture** is true (see [13]).

The main results of this paper are the following theorems.

**Theorem 1.** For any function of $n = |U|$ satisfying $\eta(n) > 1$ there is a polynomial time algorithm not depending on $\eta(n)$ to distinguish between

**YES:** The set cover problem instance $(U, S, d)$ where $d > \frac{2|S|}{3\eta(n)-1}$, for which there exists an exact cover of size at most $d$;

**NO:** The set cover problem instance $(U, S, d)$ where $d > \frac{2|S|}{3\eta(n)-1}$, for which any cover from $S$ has size larger than $\eta(n)d$.

**Corollary 1.** For any function of $m = |E|$ satisfying $\eta(m) \geq 1$ there is a polynomial time algorithm not depending on $\eta(m)$ to distinguish between

**YES:** The vertex cover problem for $k$-uniform hypergraphs instance $(G = (V, E), d)$ satisfying $d > \frac{|V|}{2\eta(m)}$, for which there exists a subset $V'$ (exact vertex cover) of size at most $d$ such that $|V' \cap e| = 1$ for each edge $e$ in $E$;

**NO:** The vertex cover problem for $k$-uniform hypergraphs instance $(G = (V, E), d)$ satisfying $d > \frac{|V|}{2\eta(m)}$, for which any vertex cover has size larger than $\eta(m)d$.

The reduction in Theorem 1 is based on polynomial time solvable computational problems for lattices.

## 2 Proofs of the Main Results

**Proof of Corollary 1.** Set $|V| = n$ and $|E| = m$. Let $B$ be the $m \times n$
matrix whose rows and columns correspond to the elements in $E$ and $G$ respectively. The entry $b_{ev}$ of $B$ at the position corresponding to the edge $e$ and vertex $v$ is 1 if the vertex $v$ is in the edge $e$ and otherwise $b_{ev} = 0$. Consider the lattice in $\mathbb{Z}^n$ defined by $L(B) = \{x \in \mathbb{Z}^n : B \cdot x = 0\}$.

We prove the following two lemmas.

**Lemma 1.** For YES instance of the vertex cover problem for $k$-uniform hypergraphs in Theorem 1, there exists a vector in $L(B)$ with $n$ non-zero coordinates.

2) For NO instance of the vertex cover problem for $k$-uniform hypergraphs in Theorem 1, any vector in $L(B)$ has at most $2(n - \eta(m)d)$ non-zero coordinates.

**Proof.** Let $1$ be the vector in $\mathbb{Z}^n$ (or $\mathbb{Z}^m$) with all $n$ (or $m$) coordinates 1. For any instance of vertex cover problem for $k$-uniform hypergraphs, it is obvious $B \cdot 1 = k \cdot 1$. For YES case, there exists an integral vector $x \in \mathbb{Z}^n$ with at most $d$ non-zero coordinates which equals to 1 such that $B \cdot x = 1$, since there exists an exact vertex cover of size at most $d$. Thus $B \cdot (kx - 1) = 0$. It is obvious that $(kx - 1)$ has $n$ non-zero coordinates. The conclusion in 1) is proved.

For NO case, let $x = x^+ - x^- \in \mathbb{Z}^n$ be any vector in the lattice $L(B)$, where $x^+$ and $x^-$ be two vectors in $\mathbb{Z}^n$ with all their coordinates non-negative integers. Set $x' \in \mathbb{Z}^n$ an integral vector which equals to $x^+$ at the non-zero positions of $x^+$ and takes any positive integer at the zero positions of $x^+$. It is obvious every coordinate of the vector $B \cdot x'$ is a positive vector, since this vector is a linear combination of all columns of the matrix $B$ with positive coefficients. Then we have every coordinates of the integral vector $B \cdot (x' - x^+)$ is positive integer. Note that $x' - x^+$ has all coordinates non-negative integers. Thus the vertex corresponding to the non-zero positions of the vector $x' - x^+$ is a vertex cover. Thus we have $n - |supp(x^+)| \geq \eta(m)d$, where $supp(v) \subseteq [n]$ is the set of non-zero positions of the vector $v \in \mathbb{Z}^n$. Similarly we can prove that $|supp(x^-)| \leq n - \eta(m)d$. The conclusion is proved.

**Lemma 2.** For NO instance of the vertex cover problem for $k$-uniform hypergraphs in Theorem 1, there exists a subspace $\mathbb{R}^{2(n-\eta(m)d)} \times \{0\}^{2\eta(m)d-n}$ in $\mathbb{Z}^n \otimes \mathbb{R}$ such that $L(B) \subseteq \mathbb{R}^{2(n-\eta(m)d)} \times \{0\}^{2\eta(m)d-n}$.
Proof. Let $x = (x_1, ..., x_n)$ be a vector in the lattice $L(B)$ with the largest number of non-zero coordinates $|\text{supp}(x)|$. From 2) of Lemma 1, we have $|\text{supp}(x)| \leq 2(n - \eta(m)d)$. Suppose there exists a vector $y = (y_1, ..., y_n) \in L(B)$ which has one non-zero coordinate position outside $\text{supp}(x)$. It is clear there exists an integer $t$ such that $x_i + ty_i \neq 0$ for those indices $i$ satisfying $x_i \neq 0$ or $y_i \neq 0$. Then the integral vector $x + ty \in L(B)$ has larger support than $x$, which has at least $|\text{supp}(x)| + 1$ non-zero positions. This is a contradiction.

Fixed any function $\eta(m) \geq 1$ we have the following algorithm.

Algorithm

Input: The vertex cover problem for $k$ uniform hypergraphs instance $(G = (V, E), d)$ satisfying $d > \frac{|E|}{2\eta(m)}$, for which the instance is YES or NO of Theorem 1.

Output: YES or NO.

Step 1. Find a base of the lattice $L(B)$;

Step 2. Check if there are at least $2\eta(m)d - n$ coordinate positions such that all these base vectors in $\mathbb{Z}^n \otimes \mathbb{R}$ are zero at the $2\eta(m)d - n$ positions. If yes, answer NO, if for every coordinate position, there exists a base vector which is non-zero at this position, answer YES.

Lemma 3. The above algorithm answers the promise problem in Theorem 1 correctly in polynomial time.

Proof. Since $2(n - \eta(m)d) < n$, it is clear the algorithm solve the promise problem correctly from Lemma 1 and 2. On the other hand the step 1 and 2 are all in polynomial time in $mn$ bits (the inputs size of the matrix $B$, see [18]). Thus the conclusion is proved.

Proof of Theorem 2. Set $|U| = n$ and $|S| = m$. Let $B$ be the $n \times m$ matrix whose rows and columns correspond to the the elements in $U$ and subsets in $S$ respectively. The entry $b_{us}$ of $B$ at the position corresponding to the point $u \in U$ and subset $S \in S$ is 1 if the element $u$ is in the subset $S$ and otherwise $b_{us} = 0$. Consider the lattice in $\mathbb{Z}^m$ defined by $L(B) = \{x \in \mathbb{Z}^m : B \cdot x = 0\}$. Let $B' = (B, 1)$ be the $n \times (m + 1)$ matrix
with the all 1 vector \(1 \in \mathbb{Z}^n\) appending to the matrix \(B\). We define lattice
\[
L(B') = \{x \in \mathbb{Z}^{m+1} : B' \cdot x = 0\} \subseteq \mathbb{Z}^{m+1}.
\]

**Lemma 4.** For NO instance of the set cover problem in Theorem 2, any vector in \(L(B)\) has at most \(2(m - \eta(n)d)\) non-zero coordinates.

**Proof.** Let \((U, S, d)\) be a NO instance. Let \(x = x^+ - x^- \in \mathbb{Z}^m\) be any vector in the lattice \(L(B)\), where \(x^+\) and \(x^-\) be two vectors in \(\mathbb{Z}^m\) with all their coordinates non-negative integers. Set \(x' \in \mathbb{Z}^m\) an integral vector which equals to \(x^+\) at the non-zero positions of \(x^+\) and takes any positive integer at the zero positions of \(x^+\). It is obvious every coordinate of the vector \(B \cdot x'\) is a positive vector, since this vector is a linear combination of all columns of the matrix \(B\) with positive coefficients. Then we have every coordinates of the integral vector \(B \cdot (x' - x^+)\) is positive integer. Note that \(x' - x^+\) has all coordinates non-negative integers. Thus the subsets in \(S\) corresponding to the non-zero positions of the vector \(x' - x^+\) is a cover of the set \(U\). Thus we have \(m - |\text{supp}(x^+)| \geq \eta(n)d\), where \(\text{supp}(v) \subseteq [m]\) is the set of non-zero positions of the vector \(v \in \mathbb{Z}^m\). Similarly we can prove that \(|\text{supp}(x^-)| \leq m - \eta(n)d\). The conclusion is proved.

**Proof of Corollary 1.** From the proof of Lemma 4 we note that for the NO instances of the promise set cover problem in Corollary 1, the lattice \(L(B)\) is the zero lattice. Thus it is easy to distinguish the YES and NO instances in Corollary 1. We can check if the lattice \(L(B)\) is zero and then find a rational solution \(y\) of the system of linear equations \(B \cdot y = 1\). When \(L(B)\) is not zero, it is YES instance. When \(L(B) = 0\), it is YES instance if the Hamming weight of \(y\) is smaller than \(m\) and it is NO instance if the Hamming weight of \(y\) is \(m\).

Similarly as Lemma 2 we can prove the following lemma.

**Lemma 5.** For NO instance of the set cover problem in Theorem 2, there exists a subspace \(\mathbb{R}^{2(m-\eta(n)d)} \times \{0\}^{2\eta(n)d-m}\) in \(\mathbb{Z}^m \otimes \mathbb{R}\) such that \(L(B) \subseteq \mathbb{R}^{2(m-\eta(n)d)} \times \{0\}^{2\eta(n)d-m}\).

**Lemma 6.** For the YES instance of the set cover problem in Theorem 2, if \(L(B) \subseteq \mathbb{R}^{2(m-\eta(n)d)} \times \{0\}^{2\eta(n)d-m}\) for some subspace \(\mathbb{R}^{2(m-\eta(n))} \times \{0\}^{2\eta(n)d-m} \subseteq \mathbb{Z}^m \otimes \mathbb{R}\). Set \(P : \mathbb{Z}^m \otimes \mathbb{R} \to \{0\}^{2(m-\eta(n)d)} \times \mathbb{R}^{2\eta(n)d-m}\) be the projection to that fixed orthogonal complimentary space \(\mathbb{R}^{m-2\eta(n)d} \times \{0\}^{2\eta(n)d-m}\). We have that \(L(B') - L(B)\) is not empty and any lattice vector
\[ \mathbf{x} \text{ in } \mathbf{L}(\mathbf{B}') - \mathbf{L}(\mathbf{B}) \text{ satisfies } ||P(\mathbf{x})|| \leq \sqrt{d}. \]

**Proof.** For the **YES** instance of Theorem 2, there exists a \( \mathbf{x}_0 \in \{0, 1\}^m \) with at most \( d \) coordinates satisfying \( \mathbf{B} \cdot \mathbf{x}_0 = 1 \) from the condition there exists an exact cover of size at most \( d \). Thus \( \mathbf{L}(\mathbf{B}') - \mathbf{L}(\mathbf{B}) \) is not empty and any vector in \( \mathbf{L}(\mathbf{B}') - \mathbf{L}(\mathbf{B}) \) is of the form \( \mathbf{x}_0 + \mathbf{x} \) where \( \mathbf{x} \in \mathbf{L}(\mathbf{B}) \subseteq \mathbb{R}^{2(m - \eta(n)d)}. \) The projection \( P \) 's image is in the orthogonal complementary of \( \mathbb{R}^{2(m - \eta(n)d)}. \) From the property \( \mathbf{x} \in \mathbb{R}^{2(m - \eta(n)d)}, \) we know that there are at most \( d \) nonzero coordinates (which are 1) in \( P(\mathbf{x}_0 + \mathbf{x}) \). The conclusion is proved.

For any function satisfying \( \eta(n) > 1 \) we have the following algorithm.

**Algorithm**

**Input:** The set cover problem instance \((\mathbf{U}, \mathbf{S}, d)\) satisfying \( d > \frac{2|\mathbf{S}|}{3\eta(n)-1}, \) for which the instance is **YES** or **NO** of Theorem 2.

**Output:** **YES** or **NO**.

**Step 1.** Find a base of the lattice \( \mathbf{L}(\mathbf{B}) \);

**Step 2.** Check if there are at least \( 2\eta(n)d - m \) coordinate positions such that all these base vectors in \( \mathbb{Z}^m \otimes \mathbb{R} \) are zero at the \( 2\eta(n)d - m \) positions. If there are less than \( 2\eta(n)d - m \) zero positions for all these base vectors, answer **YES**.

**Step 3.** If the answer of the previous step is yes, check if the lattice \( \mathbf{L}(\mathbf{B}') \) equals to \( \mathbf{L}(\mathbf{B}) \) by the natural inclusion \( \mathbf{L}(\mathbf{B}) \rightarrow \mathbf{L}(\mathbf{B}') \). If they are the same, answer **NO**.

**Step 4.** If the answer of the previous step is not. The \( \mathbf{L}(\mathbf{B}') \) and \( \mathbf{L}(\mathbf{B}) \) are not the same. Set \( P : \mathbb{Z}^m \otimes \mathbb{R} \rightarrow \mathbb{R}^{m - 2\eta(n)d} \) be the projection to that fixed \( \mathbb{R}^{m - 2\eta(n)d}. \) Take an arbitrary lattice vector \( \mathbf{x} \) in \( \mathbf{L}(\mathbf{B}') \) not in \( \mathbf{L}(\mathbf{B}) \), check if \( ||P(\mathbf{x})|| > \sqrt{d}, \) if yes answer **NO**, if not, answer **YES**.

**Lemma 7.** The above algorithm answers the promise problem in Theorem 2 correctly in polynomial time.

**Proof.** From the condition \( d > \frac{2|\mathbf{S}|}{3\eta(n)-1}, \) we have \( \eta(n)d - 2(m - \eta(n)d) > \)
If there exists a vector \( x_0 \in L(B') \setminus L(B) \) for NO instance, then \( P(x_0 + x) > \sqrt{\eta(n)d - 2(m - \eta(n)d)} > \sqrt{d} \). On the other hand, it is clear that all computation complexity in Step 1-4 is polynomial time of \( mn \). The conclusion is proved.

3 Conclusion

The implications of Theorem 1 are as follows. Firstly the \( 3\eta - 3 \) part sub-problem of the NP-hard promise set cover problem is easy when the function \( \eta(n) \) is a constant. Secondly if a problem can be reduced to the promise set cover problem within the restricted parameter range as described in Theorem 1 within polynomial time, it is an easy problem. Thirdly, from the results in Theorem 1, we can see that the instances constructed from SAT problem reduction to the set cover problem(or vertex cover problem for \( k \)-uniform hypergraph problem) in the previous works [4, 16, 19, 7] are not in the restricted parameter range in this paper.

Acknowledgment. The work was supported by Grant 10871068 of NSFC and DNRF-NSFC joint grant 11061130539.

REFERENCES

[1] M. Ajtai, The shortest vector problem in \( L_2 \) is NP-hard for randomized reductions(extended abstract), Proc. STOC 1998, pages 10-19.

[2] S. Arora, L. Babai, J. Stern and Z. Sweedyk, The hardness of approximating optima in lattices, codes and systems of linear equations, Journal of Computer and System Science, 54(2), pages 317-331, 1997, Preliminary version, FOCS 1993.

[3] N. Bansal and S. Khot, Inapproximability of hypergraph vertex cover and applications to scheduling problems, preprint 2009, see http://cs.nyu.edu/~khot/publications.html.

[4] M. Bellare, S. Goldwasser, C. Lund and A. Russell, Efficient multiprover interactive proofs with applications to approximation problems, Proc. STOC 1993, pages 113-131, 1993.
[5] V. Chvatal, A greedy heuristic for the set-covering problem. Mathematics of Operations Research, Vol. 4, pages 233-235, 1979.

[6] I. Dinur, G. Kindler, R. Raz, and S. Safra, An improved lower bound for approximating CVP. Combinatorica, Vol.23(2), pages 205-243, 2003, Preliminary version, STOC 1999.

[7] I. Dinur, V. Guruswami, S. Khot and O. Regev, A new multilayered PCP and the hardness of hypergraph vertex cover, SIAM Journal on Computing, Vol.34, no.5, 2005, pages 1129-1146, Preliminary version, STOC 2003.

[8] I. Dumer, D. Micciancio and M. Sudan, Hardness of approximating the minimum distance of a linear code, Proc FOCS 1999, Journal version, IEEE Transactions on Information Theory, 49(1), pages 22-37, 2003.

[9] U. Feige, A threshold of $lnn$ for approximating set cover, Journal of ACM, vol.45, no.4, pages 634-652, 1998.

[10] D. Hochbaum, Approximation algorithms for set covering and vertex cover problems. SIAM Journal on Computing, vol. 11, pages 555-556, 1982.

[11] I. Haviv and O. Regev, Tensor-based hardness of the shortest vector problem to within almost polynomial factors, Proc. STOC 2007, pages 469-477.

[12] D. S. Johnson, Approximation algorithms for combinatorial problems. Journal of Computer and System Sciences, vol.9, pages 256-278, 1974.

[13] S. Khot and O. Regev, Vertex cover might be hard to approximate to within 2 , Proc. of the 18th IEEE Conference on Computational Complexity, 2003.

[14] S. Khot, Hardness of approximating the shortest vector problem in lattices, Journal ACM, 52(5), pages 789-808, 2005, Preliminary version, FOCS 2004.

[15] L. Lovasz, On the ratio of optimal integral and fractional covers. Discrete Mathematics, Vol.13, pages 383-390, 1975.
[16] C. Lund and M. Yannakakis, Hardness of approximating minimization problem, J. ACM, 41(5), pages 960-981, 1994, Preliminary version, STOC 1993.

[17] D. Micciancio, The shortest vector problem is NP-hard to approximate within some constant, SIAM Journal of Computing, 30(6), pages 2008-2035, 2001, Preliminary version, FOCS 1998.

[18] D. Micciancio and S. Goldwasser, Complexity of lattice problems, A Cryptographic perspective, Kluwer Academic Publishers, 2002.

[19] R. Raz and S. Safra, A sub-constant error-probability low degree test and a sub-constant error-probability PCP characteristic of NP, Proc. STOC 1997, pages 475-484.

[20] L. Trevisan, Non-approximability results for optimization problems on bounded degree instances, Proc. STOC 2001, pages 453-461, 2001.