社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－AL－92 (9)
2003／11／7

MIN 3-SET COVER の近似困難性

大月 英明 †　　**平田 富夫** ‡

概要: 台集合 U と U の部分集合の族 $\mathcal{F}$ が与えられたとき, $\mathcal{F}$ からできるだけ少ない数の集合を選び, これらの和集合で U のすべての要素をカバーする問題を MIN SET COVER と呼ぶ. また, 部分集合それぞれに含まれる要素数が高々 $k(\geq 3)$ 個に制限されている問題は MIN k-SET COVER と呼ばれる. この論文では MAX 3-SAT(l) からのリダクションを用い, MIN 3-SET COVER は任意の $\delta > 0$ に関して $1+\epsilon_l/12l-\delta$ 近似が NP 困難であることを示す. ここで, ϵ_l はSAT から MAX 3-SAT(l) へのギャップ導入パラメータである.

Inapproximability of MIN 3-SET COVER

Hideaki Otsuki †　　**Tomio Hirata** ‡

Abstract: Given a finite set U and a collection of it's subsets $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$, MIN SET COVER is the problem of finding the minimum number of subsets covering all elements of U. MIN k-SET COVER is the restricted version of MIN SET COVER, in which the size of each F_i is not more than k. We show that it is not possible to approximate MIN 3-SET COVER within a ratio $1+\epsilon_l/12l-\delta$ (for any $\delta > 0$) unless $P = NP$, where ϵ_l is the gap introducing parameter for MAX 3-SAT(l).

1 はじめに

台集合 U とその部分集合の族 $\mathcal{F} = \{F_1, F_2, \ldots F_n\}$ が与えられたとき, MIN SET COVER は $\mathcal{F}$ から選んだ最小個数の F_i の和集合で U の要素をすべて被覆する問題である.

MIN SET COVER は代表的な NP 困難問題で, 研究も盛んに行われている. この問題はグリーディ法による近似アルゴリズムが知られており, その近似比率は $H_n = 1 + 1/2 + \cdots + 1/n \simeq \ln n$ である [1]. 近似困難性に関しては $c \log n (c > 0)$ の近似比率が証明されている [2]. 更に, 1998 年に, $NP \subset DTIME(n^{\log \log n})$ でない限り $(1-\epsilon)\ln n$ 近似が NP 困難であることが証明され [3], これにより, MIN SET COVER に関してはグリーディアルゴリズムが本質的に最適であるということが示された.

MIN k-SET COVER は各 F_i に含まれる要素数が高々 k 個の問題であり, $k \geq 3$ の場合はやはり NP 困難である. MIN SET COVER とは異なり, この問題は APX 完全であり, 実際に定数近似のアルゴリズムが与えられている [4]. また, 一般の k に関して, $\ln k - O(\ln \ln k)$ 近似が NP 困難であることが示されているが [5], 特に k が小さい場合, 近似比率の下界は具体的には求められていない. §

変数の出現回数がちょうど l 個であるような MAX 3-SAT の制限された問題は MAX 3-SAT(l) と呼ばれるが, この論文では MAX 3-SAT(l) から MIN 3-SET COVER へのギャップ保存リダクションを与え, MIN 3-SET COVER は任意の $\delta > 0$ に関して $1+\epsilon_l/12l-\delta$ 近似が NP 困難であることを示す. ここで, ϵ_l はSAT から MAX 3-SAT(l) へのギャップ導入パラメータである.

†名古屋大学大学院工学研究科電子情報学専攻
Department of Information Electronics, Nagoya University
‡名古屋大学大学院情報科学研究科
Department of Computer Science and Mathematical Informatics, Nagoya University

§発表申し込み後に文献 [11] の存在が分かった.

2 MIN 3-SET COVER の問題例の構成

3CNF 式 ϕ の充足可能な節の最大数を $OPT(\phi)$ で表す. Håstad [7] は PCP 定理 [8] を用い, 任意の 3CNF 式 ϕ が, 以下の性質を持つ 3CNF 式 ϕ' へ, 入力サイズの多項式時間で変換できることを証明した. ただし, $M(M')$ は $\phi(\phi')$ の節の数である.

$$\begin{aligned} OPT(\phi) = M &\rightarrow OPT(\phi') = M' \\ OPT(\phi) < M &\rightarrow OPT(\phi') < (1-\epsilon)M' \\ &\quad (\epsilon = 1/8) \end{aligned}$$

すなわち, P≠NP の仮定のもとでは, MAX 3-SAT に対し近似比率が $1-\epsilon$ よりも良い多項式時間アルゴリズムは存在しないことになる.

MAX 3-SAT(l) に関しては, 次のようなギャップ保存リダクションが知られている [9], [10]. 3CNF 式 ϕ_l は ϕ' から構成された MAX 3-SAT(l) の問題例で, を M_l 個の節を持つとする. このとき, $l = 29$ と $l = 5$ に対し, 以下の ϵ_l が得られている.

$$\begin{aligned} OPT(\phi') = M' &\rightarrow OPT(\phi_l) = M_l \\ OPT(\phi') < (1-\epsilon)M' &\rightarrow OPT(\phi_l) < (1-\epsilon_l)M_l \\ &\quad (\epsilon_{29} = \epsilon/43, \epsilon_5 = \epsilon_{29}/29). \end{aligned}$$

MIN 3-SET COVER の最適解のサイズを $OPT(U, \mathcal{F})$ と表すとき, 以下のような性質を持つ MIN 3-SET COVER の問題例 $(U, \mathcal{F})$ が構成できるとする.

$$\begin{aligned} OPT(\phi_l) = M_l &\rightarrow OPT(U, \mathcal{F}) = c \\ OPT(\phi_l) < (1-\epsilon_l)M_l &\rightarrow OPT(U, \mathcal{F}) \\ &\quad > (1+\sigma)c \\ &\quad (\sigma > 0) \end{aligned}$$

ここで, c は問題例 $(U, \mathcal{F})$ のパラメータである. この場合, MIN 3-SET COVER の問題例に対する最適解のサイズが $(1+\sigma)c$ よりも小さいかどうかを判定することにより, もとの SAT の問題例 ϕ の充足性が判定できる. したがって, $P \neq NP$ の仮定のもとでは近似比率 $1+\sigma$ 未満であるような MIN 3-SET COVER の近似アルゴリズムは存在しないことになる.

以下では Garey-Johnson [6] の手法を応用し, MAX 3-SAT(l) の問題例 ϕ から MIN 3-SET COVER の問題例 $(U, \mathcal{F})$ を構成する. [6] では, 変数への真偽値割り当てを表す集合族, 節の充足性を表す集合族, そして前の 2 つの集合でカバーできなかった要素をカバーするための集合族, の 3 種類の集合族を考え, 3-SAT から EXACT 3-SET COVER へのリダクションを与えている¶. この手法をそのまま MAX 3-SAT に利用しても, 前述の様なギャップを保存するリダクションは得られない. ここでは彼らの手法に修正を加えた集合族を定義し, 先に述べたような性質をもつ MIN 3-SET COVER の問題例を具体的に構成する.

まず, 3CNF 式 ϕ は変数 $x_1, \ldots, x_N$ と節 $C_1, \ldots, C_M$ をもつとする. さらに, 各変数 x_i の出現回数を l とし, $L = Nl$ とする. このとき, $L = Nl = 3M$ である.

各変数 $x_i (1 \le i \le N)$ に対し, $2l$ 個の要素

$$U_x^i = \{x_i[j], \bar{x}_i[j] | 1 \le j \le l\}$$

を対応させ

$$U_x = \bigcup_{i=1}^{N} U_x^i$$

とする. さらに各 $i (1 \le i \le N)$ に対して $2l$ 個の要素

$$U_T^i = \{a_i[j], b_i[j] | 1 \le j \le l\}$$

を用意し,

$$U_T = \bigcup_{i=1}^{N} U_T^i$$

¶[6] で実際に与えているのは, 3-SAT から 3-DIMENSIONAL MATCHING へのリダクションである.

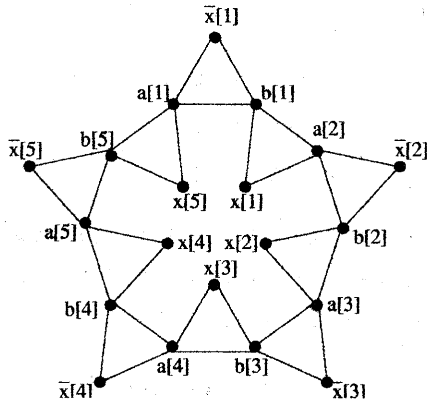

図 1: U_T^i のグラフ表現

とする．次に，各 $i(1 \le i \le N)$ について $2l$ 個の部分集合の族を次のように構成する．

$$
\begin{aligned}
\mathcal{T}_i^t &= \{\{\bar{x}_i[j], a_i[j], b_i[j]\} | 1 \le j \le l\} \\
\mathcal{T}_i^f &= \{\{x_i[j], a_i[j+1], b_i[j]\} \\
&\quad | 1 \le j \le l-1\} \cup \{x_i[l], a_i[1], b_i[l]\}.
\end{aligned}
$$

$\mathcal{T}_i \equiv \mathcal{T}_i^t \cup \mathcal{T}_i^f$ とし，$\mathcal{T} \equiv \bigcup_{1 \le i \le N} \mathcal{T}_i$ とすると $|\mathcal{T}| = 2L$ である．この集合族は各変数に与える真偽値と対応させるために用いる．例えば，変数 x_i に **true** を与える場合は $\mathcal{T}_i^t$, **false** を与える場合は $\mathcal{T}_i^f$ を対応させる．どの i についても，l 個未満の集合で U_T^i の要素をすべてカバーすることはできない．一方，$\mathcal{T}_i^t$ または $\mathcal{T}_i^f$ のどちらか一方で U_T^i をすべてカバーすることができる．したがって，ある真偽値割り当てを与えると，各 i について真偽割り当てに対応する集合族 $\mathcal{T}_i^t$ または $\mathcal{T}_i^f$ をカバーに含めることによって，U_T のすべての要素を $L = Nl$ 個の集合でカバーできる．以下では，ある真偽値割り当てに対応して $\mathcal{T}$ から部分集合を選ぶとき，各 i について"一貫した"選び方をする，ということにする．

節集合に対応して，M 個の要素

$$U_S = \{s[y] | 1 \le y \le M\}$$

を用意する．節 $C_y = (X_\alpha \vee X_\beta \vee X_\gamma)$ （ただし $X_\mu = x_\mu$ または $\bar{x}_\mu$）を考える．ここでは，ϕ における変数 x_μ の出現が何番目であるかに対応して部分集合を構成していく．C_y のリテラル X_μ が，3-CNF 式 ϕ において j_μ 番目に現れた変数であった場合，集合 $\mathcal{S}_y^k = \{X_\mu[j_\mu], s[y]\}$ を作る．3 つの集合ができるのでそれらの族を $\mathcal{S}_y = \{\mathcal{S}_y^k | k = 1,2,3\}$ とする．さらに $\mathcal{S} \equiv \bigcup_{1 \le y \le M} \mathcal{S}_y$ とする

この集合族は節 C_y の充足性を表すものである．ある真偽値割り当てのもとで節がすべて充足可能であれば，それぞれの C_y から **true** のリテラルを選び，それを含む $\mathcal{S}_y$ をカバーに含めれば，M 個の部分集合で U_S の要素をすべてカバーすることができる．

次に，

$$U_G = \{g(p) | 1 \le p \le L - M\}$$

を用意し，U_x の要素と U_G の要素を対にした集合族 $\mathcal{G}$ を構成する．

$$
\begin{aligned}
\mathcal{G} = \; & \{\{x_i[j], g[p]\}, \{\bar{x}_i[j], g[p]\} \\
& | 1 \le i \le N, 1 \le j \le l, \\
& 1 \le p \le L - M\}
\end{aligned}
$$

この集合族は，$\mathcal{T}, \mathcal{S}$ でカバーされていない U_x の要素をカバーするために利用される．

ここで，

$$
\begin{aligned}
U &= U_x \cup U_T \cup U_S \cup U_G \\
\mathcal{F} &= \mathcal{T} \cup \mathcal{S} \cup \mathcal{G}
\end{aligned}
$$

とし，$\mathcal{F}$ から最小個数の部分集合で U の要素をすべてカバーする SET COVER を考える．$\mathcal{F}$ の要素はすべて，高々3 個の要素しか含んでいない U の部分集合であるから，これは MIN 3-SET COVER の問題例である．これらの集合族の構成は，明らかに N の多項式時間で可能である．

3 MIN 3-SET COVER の近似困難性

上で構成した MIN 3-SET COVER の問題例が以下のような性質を持つことを示す.

$$
\begin{aligned}
OPT(\phi_l) = M & \rightarrow & OPT(U, \mathcal{F}) = 2L \\
OPT(\phi_l) \leq (1-\epsilon_l)M & \rightarrow & OPT(U, \mathcal{F}) \\
& & > (1+\frac{\epsilon_l}{12l})2L.
\end{aligned}
$$

ここで, $(U, \mathcal{F})$ の構成法より, U_T, U_S, U_G をすべてカバーするためには $L+|U_S|+|U_G| = L+M+(L-M) = 2L$ 個以上の部分集合が必要であることに注意されたい.

ϕ が充足可能な場合を考える. まず ϕ の最適割り当て π によって x_i に **true** が与えられる場合は $\mathcal{T}_i^t$ を, **false** が与えられる場合は $\mathcal{T}_i^f$ をカバーに含める. すなわち, π に対して一貫した選び方でカバーを構成する. このカバーを $\mathcal{C}_T$ とする. ここで $|\mathcal{C}_T| = L$ である. $\mathcal{C}_T$ は, π のもとで **false** になるリテラルに対応する U_x の要素をすべてカバーしている. 各節 C_y について, π のもとで C_y を充足させているリテラルを一つ選び, そのリテラルを含む S_y の部分集合をカバーに含める. このようにして構成したカバーを $\mathcal{C}_T \cup \mathcal{C}_S$ とする. ここで $|\mathcal{C}_T \cup \mathcal{C}_S| = L+M$ となる. π のもとで **true** になるリテラルに対応する U_x の要素 (L 個ある) のうち, $\mathcal{C}_S$ はちょうど M 個だけカバーしている. 残りの $L-M$ 個を $\mathcal{G}$ に含まれる集合で適切にカバーすることにより, U_G の要素もすべてカバーすることができる. 以上の構成から, このカバーは U をすべてカバーしており, そのサイズは $L+M+(L-M) = 2L$ となる. また U をカバーするために必要な集合のサイズは $2L$ であるから, これが最小カバーのサイズである.

次に ϕ の最適割り当てのもとで, ϵM 個の節が充足不可能な場合を考える.

U の任意のカバーを $\mathcal{C}_U$ とし

$$
\begin{aligned}
\mathcal{C}_U \cap \mathcal{T} & \equiv & \mathcal{C}_T \\
\mathcal{C}_U \cap \mathcal{S} & \equiv & \mathcal{C}_S \\
\mathcal{C}_U \cap \mathcal{G} & \equiv & \mathcal{C}_G
\end{aligned}
$$

と定義する. $\mathcal{C}_T$ によってカバーされている U_x の要素の集合を χ_T とする. さらに $\mathcal{C}_S$ によってカバーされている U_x の要素の集合を χ_S とし,

$$
\begin{aligned}
\chi_i & = & \{x_i[j] | x_i[j] \in \chi_S, 1 \leq j \leq l\} \\
\bar{\chi}_i & = & \{\bar{x}_i[j] | \bar{x}_i[j] \in \chi_S, 1 \leq j \leq l\}
\end{aligned}
$$

と定義する. 各 i に関して次のどれかが成り立っている.

1. $\chi_i = \emptyset$ かつ $\bar{\chi}_i = \emptyset$
2. $\chi_i \neq \emptyset$ かつ $\bar{\chi}_i = \emptyset$
3. $\chi_i = \emptyset$ かつ $\bar{\chi}_i \neq \emptyset$
4. $\chi_i \neq \emptyset$ かつ $\bar{\chi}_i \neq \emptyset$

上の条件を満たしている i の集合をそれぞれ I_1, I_2, I_3, I_4 とする.

補題 1 $|\mathcal{C}_U| \geq 2L + |I_4|/2$ である.

(証明) I_4 を次のように I_4' と $I_4'' = I_4 - I_4'$ に分割し, $d' = |I_4'|$, $d'' = |I_4''|$ とする.

$$I_4' = \{i \in I_4 | U_x^i \cap \chi_T \cap \chi_S = \emptyset\}$$

つまり, $i \in I_4'$ ならば U_x^i の要素は $\mathcal{C}_T$ と $\mathcal{C}_S$ とで重なってカバーされることはない. また, $i \in I_4''$ のときは U_x^i の要素で $\mathcal{C}_T$ と $\mathcal{C}_S$ とで重なってカバーされるものが少なくとも 1 個ある.

まず, $|\mathcal{C}_U| \geq 2L + d'$ であることを示す. U_T^i をカバーするためには, 各 $i (1 \leq i \leq N)$ について $\mathcal{C}_T$ の要素は l 個以上必要である. つまり, $|\mathcal{C}_T \cap \mathcal{T}_i| \geq l$ である. 特に, $i \in I_4'$ のときは $\mathcal{C}_T$ は一貫して選べな

いので U_T^i をカバーするのに T_i の要素が少なくとも $l+1$ 個以上必要となり, $|C_T \cap T_i| \geq l+1$ である. よって, $|C_T| \geq lN + d' = L + d'$ である. $(U, \mathcal{F})$ の構成法より, $|\mathcal{C}_S| \geq M$, $|\mathcal{C}_G| \geq L - M$ なので,

$$\begin{aligned} |\mathcal{C}_U| &= |\mathcal{C}_T| + |\mathcal{C}_S| + |\mathcal{C}_G| \\ &\geq L + d' + M + L - M = 2L + d' \end{aligned}$$

である.

次に, $|\mathcal{C}_U| \geq 2L + d''$ であることを示す. I_4'' の定義より, $\mathcal{C}_T$ と $\mathcal{C}_S$ とで重なってカバーされている U_x の要素が少なくとも d'' 個ある. したがって, $\mathcal{C}_T$ と $\mathcal{C}_S$ とでカバーされない U_x の要素は $|U_x| - |C_T| - |C_S| + d''$ 個以上あり, それらは $\mathcal{C}_G$ でカバーされていなければならない. よって, $|\mathcal{C}_G| \geq |U_x| - |C_T| - |C_S| + d''$ である. よって, カバー $\mathcal{C}_U$ の大きさは

$$\begin{aligned} |\mathcal{C}_U| &= |\mathcal{C}_T| + |\mathcal{C}_S| + |\mathcal{C}_G| \\ &\geq |C_T| + |C_S| + |U_x| - |C_T| - |C_S| + d'' \\ &= 2L + d'' \end{aligned}$$

である. $d' + d'' = |I_4|$ より, $C_U \geq 2L + \max(d', d'') \geq 2L + |I_4|/2$ となり, 補題が証明された.

補題 2 $|I_4| \geq \frac{\epsilon_l M}{l}$ である.

(証明) $\mathcal{C}_U = \mathcal{C}_T \cup \mathcal{C}_S \cup \mathcal{C}_G$ から次のような割り当て π を考える.

- $\{x_i | i \in I_1\} \to$ 任意
- $\{x_i | i \in I_2\} \to$ **true**
- $\{x_i | i \in I_3\} \to$ **false**
- $\{x_i | i \in I_4\} \to$ 任意

$\mathcal{C}_S$ は, それぞれの y に関して S_y から一つ以上の集合を含めている. $U_x^i (i \in I_2 \cup I_3)$ の要素をカバーしている $\mathcal{C}_S$ の要素 S_y^k があるとき, 節 C_y は π のもとで **true** である. したがって, π のもとで **false** になる節に対応する $s[y]$ をカバーする $\mathcal{C}_S$ の要素 S_y^k はすべて $U_x^i (i \in I_4)$ の要素をカバーしている. これらの S_y^k でカバーされる U_x^i の要素はたかだか $l|I_4|$ 個であり, U_x^i の各要素はたかだか一個の S_y^k によってカバーされるので, **false** になる節の個数はたかだか $l|I_4|$ 個しかない. π のもとで **false** になる節は少なくとも ϵM 個以上あるので $l|I_4| \geq \epsilon M$ である. よって補題が示された.

補題 1 と補題 2 より $\mathcal{C}_U$ のサイズは少なくとも

$$\begin{aligned} 2L + |I_4|/2 = 2L + \epsilon_l M/(2l) &= 2L(1 + \epsilon_l \frac{M}{4Ll}) \\ &= 2L(1 + \frac{\epsilon_l}{12l}) \end{aligned}$$

である.

以上より, 先に述べた性質を満たす MIN 3-SET COVER の問題例を与えることができた. MAX 3-SAT(l) の任意の $\delta > 0$ における $1 - \epsilon + \delta$ 近似が NP 困難であることから, 任意の $\delta > 0$ に関して MIN 3-SET COVER の $1 + \epsilon_l/12l - \delta$ 近似は NP 困難である.

4 考察

MIN k-SET COVER に関する $\sum_i^k 1/i - 1/2$ 近似比率のアルゴリズムが Duh と Fürer [4] によって与えられている. また [5] の結果からも, 実際の近似困難な閾値は $\ln k$ に近いものであると思われる ($k = 3$ のとき, $\sum_i^k 1/i - 1/2 = 1.333333333$, $\ln k = 1.098612289$, $1 + \epsilon_{29}/(12 \cdot 29) = 119713/119712 \simeq 1.000008353$, $1 + \epsilon_5/(12 \cdot 5) = 598561/598560 \simeq 1.000001671$,) .

謝辞 この結果を導いた後, Chelbik and Chelebikova [11] の結果 (MIN 3-SET COVER に関して $\frac{100}{99}$ という近似困難比率を得ている) の存在を知った。文献をご教示いただいた東京電気大学の陳致中先生に感謝いたします.

参考文献

[1] Johnson, D. S., "Approximation algorithms for combinatorial problems", *Journal of Comput. System Sci.*, **9**, pp.256-278, 1974.

[2] Raz, R. and Safra, S., "A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP ", *Proc. 29th Ann. ACM Symp. on Theory of Comp.*, pp.475-484, 1997.

[3] Feige, U, "A threshold of $\ln n$ for approximating set cover ", *Journal of ACM.*, **45**, pp.634-652, 1998.

[4] Duh, R. and Fürer, M. , "Approximation of k-set cover by semi-local optimization ", *Proc. 29th Ann. ACM Symp. on Theory of Comp.*, pp.256-265,1997.

[5] Trevisan, L. , "Non-approximability results for optimization problems of bounded degree instances, " *Proc. 33th Ann. ACM Symp. on Theory of Comp.*, pp.453-461, 2001.

[6] Garey, M. and Johnson, D. , Computers and Intractability, W. H. Freeman and Company, New York,1979.

[7] Håstad, J. , "Some optimal inapproximability results", *Journal of ACM.*, **48**, pp.798-859, 2001.

[8] Arora, S. and Safra, S., "Probablistic checking of proofs: a new characterization of NP" , *Proc of the 33rd Symposium on Foundations of Computer Science,* IEEE,pp.2-13,1992.

[9] Papadimitriou, C. and Yannakakis M. , "Optimization, Approximation, and Complexity Classes", *Journal of Computer and System Sciences* **43(3)**, pp.425-440,1991.

[10] Arora, S. and Lund, C. , "Hardness of Approximations." In Approximation Algorithms for NP-hard Problems, Dorit Hochbaum, Ed. PWS Publishing , 1996.

[11] Chelbik, M. and Chelebikova, J. " Inapproximability results for bounded variants of optimization problems", Proceedings of FCT2003, LNCS 2751, pp. 27-38(2003).