Improved Approximations for Max Set Splitting and Max NAE SAT *

Jiawei Zhang and Yinyu Ye †

Department of Management Sciences

Henry B. Tippie College of Business

The University of Iowa

Iowa City, Iowa 52242, USA

Qiaoming Han ‡

Graduate School of Management Science and Engineering

Nanjing University

Nanjing, 210093, P.R.China

August 16, 2000, Revised January 22, 2001.

*This research was supported in part by NSF grants DMI-9908077 and DMS-9703490.

†This work was done while the author was visiting Fudan University, Shanghai, R.R.China.

‡This work was done while the author was visiting Computational Optimization Laboratory, Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242, USA.

Abstract

We present a 0.7499-approximation algorithm for Max-Set-Splitting in this paper. The previously best known result for this problem is a 0.7240-approximation by Andersson and Engebretsen [3], which is based on a semidefinite programming (SDP) relaxation. Our improvement is resulted from a strengthened SDP relaxation, an improved rounding method, and a tighter analysis compared with that in [3].

Key words. Max Set Splitting, Max NAE SAT, approximation algorithm, semidefinite programming relaxation.

1 Introduction

In the *Max Set Splitting* problem, we are given a finite set $U = \{1, 2, \cdots, m\}$, a collection $C = \{S_1, S_2, \cdots, S_n\}$ of subsets of U, and the nonnegative weight w_j for each subset $S_j \in C$. The goal is to find a partition of U into two subsets, $U = U_1 \cup U_2$, that maximizes the total weight of subsets in C which are split. A subset $S_j \in C$ is said to be split by the partition $U = U_1 \cup U_2$ if $S_j \cap U_1 \neq \emptyset$ and $S_j \cap U_2 \neq \emptyset$. This problem is also called *Max Hypergraph Cut*. *Max E_k-Set Splitting* is a special case of *Max Set Splitting* in which all subsets in C have cardinality exactly k.

For any fixed $k \geq 2$, *Max E_k-Set Splitting* was shown to be NP-hard by Lovász [12]. Petrank [15] proved that there exists some constant $\epsilon > 0$ such that the existence of a polynomial time $(1 - \epsilon)$-approximation algorithm would imply P=NP. *Max E_2-Set Splitting* is exactly the extensively studied MAX CUT problem, which is known to be NP-hard to approximate within $\frac{16}{17} + \epsilon$ for any constant $\epsilon > 0$ [16]. Goemans and Williamson [6], in a major breakthrough, have used semidefinite programming (SDP) and obtained an approximation algorithm for MAX CUT with a performance guarantee 0.87856. For *Max E_3-Set Splitting*, it has been shown by Kann et.al. [8] to be approximatable within the same performance guarantee as MAX CUT; and this bound has been improved to 0.87867 [19], and 0.90871 [18] by Zwick. Very recently, Guruswami [9] have showed that *Max E_3-Set Splitting* is not approximatable within a factor of $\frac{19}{20} + \epsilon$. For $k \geq 4$, *Max E_k-Set Splitting* is approximatable within $1 - 2^{1-k}$ [2, 8], and this is the best possible since Hastad [10] showed that it is hard to approximate within a factor of $1 - 2^{1-k} + \epsilon$ for any constant $\epsilon > 0$.

On approximating k-mixed *Max Set Splitting* problems, Arora et.al. [4] designed a PTAS for dense instances of the *Max Set Splitting* problem. For the general *Max Set Splitting* problem, the previously best known result was a 0.72405-approximation algorithm due to Andersson and Engebretsen [3]. Ageev and Sviridenko [1] gave a 1/2-approximation algorithm for this problem with the constraint $|U_1| = l$ for any given integer $l (1 \leq l \leq m)$.

Note that Zwick [19] also constructed an approximation algorithm for the general *Max Set Splitting* problem. Based on the strong numerical evidence, he conjectured that the algorithm has performance guarantee 0.7977. His conjecture has not been proved yet.

A closely related problem of *Max Set Splitting* is *Max NAE SAT*, which is a variant of the well-known *Max SAT* problem. In the *Max NAE SAT* problem, we are given a set of variables $U = \{x_1, x_2, \cdots, x_m\}$, a collection C of disjunctive boolean clauses of literals, where a literal is either a variable $x_i \in U$ or its negation $\bar{x}_i$. Each clause $S_j \in C$ is associated with a nonnegative weight w_j. The objective is to maximize the total weight of the satisfied clauses, i.e., contain at lease one true literal and at least one false literal, over all assignments of truth values to the variables of U. *Max (E_k-)Set Splitting* is a special case of *Max (E_k)-NAE SAT* in which all literals appear unnegated. All of the results for *Max (E_k-)Set Splitting* mentioned above hold for *Max (E_k)-NAE SAT*.

In this paper, building on the algorithm of Andersson and Engebretsen [3], we present an improved 0.74996−approximation algorithm for *Max Set Splitting*. Recall that the algorithm of Andersson and Engebretsen [3] has three steps: (1) solve a semidefinite programming (SDP) relaxation of *Max Set Splitting*, (2) find a partition by using the rounding method of Goemans and Williamson [6], (3) add a probabilistic postprocessing step, where the partition of the second step is perturbed.

We improve the algorithm of Andersson and Engebretsen [3] in the following ways. First, we present a strengthened SDP relaxation for *Max Set Splitting* by adding some valid inequalities. Second, we round the SDP solution by an improved rounding method, which was introduced by Nesterov [13], Zwick [19] and Ye [17]. Our improvement also relies on a tighter analysis compared with that in [3].

We also consider the satisfiable instances of the *Max NAE SAT* problem, i.e., all the clauses in C can be satisfied in an optimal assignment. In this case, the performance ratio of our algorithm can be improved to 0.8097. The only known result for satisfiable instances is a conjectured 0.8638-approximation algorithm [19]. For satisfiable *Max* E_3*-NAE SAT*, there is a 0.912-approximation algorithm [19].

This paper is organized as follows. In the next section, we present an strengthened SDP relaxation based approximation algorithm for *Max Set Splitting*. In Section 3, we analyze the quality of the partition resulted from rounding an optimal SDP solution. We analyze the performance ratio of the perturbation step and the final partition in Section 4. In Section 5, we extend our algorithm to *Max NAE SAT*. In Section 6, we show that the satisfiable *Max NAE SAT* can be approximated with a factor of (at least) 0.8097.

2 Semidefinite programming relaxation

The *Max Set Splitting* can be formulated as

$$
\text{(MSS)} \quad
\begin{array}{lll}
w^* := \text{Maximize} & \displaystyle\sum_{j=1}^{n} w_j z_j & \\
\text{subject to} & \displaystyle\frac{1}{|S_j|-1} \sum_{i,k \in S_j | i<k} \frac{1-x_i x_k}{2} \geq z_j \quad \text{for all} \quad S_j \in C \\
& x_i \in \{-1,1\}, \quad \text{for} \quad i=1,\ldots,m, \\
& z_j \in \{0,1\}, \quad \text{for} \quad i=1,\ldots,n.
\end{array}
$$

In the above formulation $z_j = 1$ if S_j is split, $z_j = 0$ otherwise. $x_i = 1$ if $i \in U_1$ and $x_i = -1$ if $i \in U_2$. The constraint

$$\frac{1}{|S_j| - 1} \sum_{i,k \in S_j | i<k} \frac{1 - x_i x_k}{2} \geq z_j \quad \text{for all} \quad S_j \in C$$

is valid since

$$\sum_{i,k \in S_j | i<k} \frac{1 - x_i x_k}{2} \begin{cases} = 0, & \text{if} \quad x_i = x_k \quad \text{for all} \quad i, k \in S_j; \\ \geq |S_j| - 1, & \text{otherwise.} \end{cases}$$

The following SDP relaxation of (MSS) is due to Andersson and Engebrestsen [3].

$$\text{(AE-SDP)} \qquad \begin{array}{ll} \text{Maximize} & \displaystyle\sum_{j=1}^{n} w_j z_j \\ \text{subject to} & \displaystyle\frac{1}{|S_j| - 1} \sum_{i,k \in S_j | i<k} \frac{1 - X_{ik}}{2} \geq z_j \quad \text{for all} \quad S_j \in C \\ & X_{ii} = 1, \quad \text{for} \quad i = 1, \ldots, m, \\ & X \succeq 0, \\ & 0 \leq z_j \leq 1, \quad \text{for} \quad i = 1, \ldots, n. \end{array}$$

We observed that the SDP relaxation (AE-SDP) can be strengthened by the following lemma:

Lemma 1 *If $t_i \in \{-1, 1\}$ for $i = 1, 2, \cdots, L$, then*

$$\sum_{i,k | 1 \leq i < k \leq L} t_i t_k \geq -\lfloor \frac{L}{2} \rfloor$$

where $\lfloor x \rfloor$ denotes the maximum integer which is less than or equal to x.

Proof. If $t_i \in \{-1, 1\}$ for $i = 1, 2, \cdots, L$, then

$$(\sum_{i=1}^{L} t_i)^2 \geq \begin{cases} 0, & \text{if } L \text{ is even:} \\ 1, & \text{otherwise.} \end{cases}$$

Therefore, the lemma is followed. ■

Corollary 1 *If $t_i \in \{-1, 1\}$ for $i = 1, 2, \cdots, L$, then*

$$\sum_{i,k|1\le i<k\le L} \frac{1 - t_i t_k}{2} \le \begin{cases} \frac{L^2}{4}, & \text{if } L \text{ is even;} \\ \frac{(L+1)(L-1)}{4}, & \text{if } L \text{ is odd.} \end{cases}$$

Now, we are ready to give a strengthened SDP relaxation of (MSS) by adding those valid inequalities provided in Lemma 1.

$$\begin{array}{llll} w^{SDP} := & \text{Maximize} & \sum_{j=1}^{n} w_j z_j & \\ \text{(MSS-SDP)} & & & \\ & \text{subject to} & \frac{1}{|S_j| - 1} \sum_{i,k\in S_j|i<k} \frac{1 - X_{ik}}{2} \ge z_j & \text{for all } S_j \in C \\ & & \sum_{i,k\in S_j|i<k} X_{ik} \ge -\lfloor \frac{|S_j|}{2} \rfloor & \text{for all } S_j \in C \\ & & X_{ii} = 1, \quad \text{for} \quad i = 1, \ldots, m, & \\ & & X \succeq 0, & \\ & & 0 \le z_j \le 1, \quad \text{for} \quad i = 1, \ldots, n. & \end{array}$$

Let (X^*, z^*) be an optimal solution of (MSS-SDP). Then, the approximation algorithm for *Max Set Splitting* is as follows.

Algorithm MSS

1. SDP Solving: Solve (MSS-SDP) to obtain the semidefinite matrix X^*.

Repeat the next two steps for enough times and output the best partition.

2. Randomized Rounding: Generate a vector u from a multivariate normal distribution with 0 mean and covariance matrix $\theta X^* + (1-\theta)I$, where I is the identity matrix and $0 < \theta \le 1$; that is, generate $u \in N(0, \theta X^* + (1-\theta)I)$, then assign $\hat{x} = \text{sign}(u)$, i.e.,

$$\hat{x}_i = \begin{cases} 1 & \text{if } u_i \ge 0, \\ -1 & \text{if } u_i < 0. \end{cases}.$$

The value of θ will be specified later in Section 4.

3. Perturbation: For each $i \in U$, we let

$$\tilde{x}_i = \begin{cases} -\hat{x}_i, & \text{with probability } p; \\ \hat{x}_i, & \text{with probability } 1-p. \end{cases}$$

Again, the value of $p \leq \frac{1}{2}$ will be specified later in Section 4. ∎

Note that the rounding method has been used by Zwick [19] and Ye [17], and generalized recently by Han et. al. [7]. When $\theta = 1$, it is exactly the method of Goemans and Williamson [6].

3 Analysis of rounding

Let $u(S_j) = 1$ if the subset S_j is split by the partition given by $\hat{x}$, $u(S_j) = 0$ otherwise. Then, the total weighted splitting value from the partition is $\sum_{j=1}^n w_j u(S_j)$. Since $u(S_j)$ is a random number, we are interested in the expected value of $\sum_{j=1}^n w_j u(S_j)$, i.e., $\sum_{j=1}^n w_j \mathrm{E}[u(S_j)]$.

Now, we consider any subset $S_j \in C$. We want to prove that, for some $\alpha_{|S_j|} > 0$, $\mathrm{E}[u(S_j)] \geq \alpha_{|S_j|} \cdot z_j^*$. Recall that (X^*, z^*) is an optimal solution of (MSS-SDP). For simplicity, we drop subscript j in the rest of this section.

We start by establishing a lower bound for $u(S)$. Let

$$\tau_{|S|} = \begin{cases} \frac{4}{|S|^2}, & \text{if } |S| \text{ is even;} \\ \frac{4}{(|S|+1)(|S|-1)}, & \text{if } |S| \text{ is odd.} \end{cases}$$

Lemma 2 *With probability* 1*, the following two inequalities hold:*

(i) $\quad u(S) \geq \max_{i,k \in S} \{\frac{1 - \hat{x}_i \hat{x}_k}{2}\}$,

(ii) $u(S) \geq \tau_{|S|} \sum_{i,k \in S | i<k} \frac{1-\hat{x}_i\hat{x}_k}{2}$.

Proof. We first observe that S is split if any pair of nodes in S is split by $\hat{x}$, i.e.,

$$u(S) \geq \max_{i,k \in S}\{\frac{1-\hat{x}_i\hat{x}_k}{2}\},$$

with probability 1. This proves (i).

To prove (ii), we note that $u(S) \in \{0,1\}$. By Corollary 1, we have

$$\tau_{|S|} \sum_{i,k \in S | i<k} \frac{1-\hat{x}_i\hat{x}_k}{2} \leq 1.$$

Therefore, if $u(S) = 1$, we have

$$\tau_{|S|} \sum_{i,k \in S | i<k} \frac{1-\hat{x}_i\hat{x}_k}{2} \leq 1 = u(S).$$

If $u(S) = 0$, then $\hat{x}_i\hat{x}_k = 1$ for all $i, k \in S$ which implies that

$$\tau_{|S|} \sum_{i,k \in S | i<k} \frac{1-\hat{x}_i\hat{x}_k}{2} = 0 \leq u(S).$$

■

In order to prove a lower bound on $\mathrm{E}[u(S)]$, we need the following theorem which is due to Goemans and Williamson (Theorem 3.1, [6]); see also Bertsimas and Ye (Proposition 1, [5]).

Theorem 1 *[6] Let X be a positive semidefinite matrix with $X_{ii} = 1$ for all i, and choose u from the multivariate normal distribution with mean 0 and covariance matrix X. Let $\hat{x} = sign(u)$. Then $\mathrm{E}[\hat{x}_i\hat{x}_k] = \frac{2}{\pi}\arcsin(X_{ik})$ for all $i \neq k$.*

Then, we have

Corollary 2 *Let X be a positive semidefinite matrix with $X_{ii} = 1$ for all i, $\theta \in (0, 1]$, and choose u from the multivariate normal distribution with mean 0 and covariance matrix $\theta X + (1-\theta)I$, where I is the identity matrix. Let $\hat{x} = sign(u)$. Then* $\mathrm{E}[\hat{x}_i \hat{x}_k] = \frac{2}{\pi}\arcsin(\theta X_{ik})$ *for $i \neq k$.*

Now we are ready to analyze the (expected) quality of the partition given by $\hat{x}$. For a given $\theta \in (0, 1]$, let X_S be an $|S| \times |S|$ real symmetric matrix and

$$\begin{array}{rll} \alpha_{|S|} := & \text{Minimize}_{\{X_S, y, z\}} & y \\ & \text{Subject to} & y \geq \max_{i,k|1 \leq i < k \leq |S|} \{\frac{1 - \frac{2}{\pi}\arcsin(\theta X_{ik})}{2z}\} \\ & & y \geq \frac{\tau_{|S|} \sum_{i,k|1 \leq i < k \leq |S|} (1 - \frac{2}{\pi}\arcsin(\theta X_{ik}))}{2z} \\ & & z = \min\{1, \frac{1}{|S|-1} \sum_{i,k|1 \leq i < k \leq |S|} \frac{1 - X_{ik}}{2}\} > 0 \\ & & \sum_{i,k|1 \leq i < k \leq |S|} X_{ik} \geq -\lfloor \frac{|S|}{2} \rfloor \\ & & X_{ii} = 1, \quad \text{for} \quad i = 1, \ldots, |S| \\ & & X_S \succeq 0. \end{array} \tag{1}$$

Then, we have

Lemma 3 $\mathrm{E}[u(S)] \geq \alpha_{|S|} \cdot z^*$.

Proof. Note that $\mathrm{E}[u(S)] \geq 0$. Therefore, the lemma is true if $z^* = 0$.

Now we consider the case $z^* > 0$. Let X^* be the optimal matrix solution of (MSS-SDP), and let X_S^* be the principle submatrix (of X^*) consisting of all components X_{ik}^* such that $i, k \in S$. Note that X_S^* is $|S| \times |S|$, symmetric and positive semidefinite, and its diagonal components equal 1.

By (i) of Lemma 2 and Corollary 2, we have

$$\begin{aligned}
\mathrm{E}[u(S)] &\geq \mathrm{E}\left[\max_{i,k\in S}\{\frac{1-\hat{x}_i\hat{x}_k}{2}\}\right] \\
&\geq \max_{i,k\in S}\{\frac{1-\mathrm{E}[\hat{x}_i\hat{x}_k]}{2}\} \\
&= \max_{i,k\in S}\{\frac{1-\frac{2}{\pi}\arcsin(\theta X^*_{ik})}{2}\} \qquad (2) \\
&= \max_{i,k\in S}\{\frac{1-\frac{2}{\pi}\arcsin(\theta X^*_{ik})}{2z^*}\}\cdot z^*.
\end{aligned}$$

On the other hand, by (ii) of Lemma 2 and Corollary 2, we have,

$$\begin{aligned}
\mathrm{E}[u(S)] &\geq \tau_{|S|}\mathrm{E}[\sum_{i,k\in S|i<k}\frac{1-\hat{x}_i\hat{x}_k}{2}] \\
&= \tau_{|S|}\sum_{i,k\in S|i<k}\frac{1-\mathrm{E}[\hat{x}_i\hat{x}_k]}{2} \\
&= \tau_{|S|}\sum_{i,k\in S|i<k}\frac{1-\frac{2}{\pi}\arcsin(\theta X^*_{ik})}{2} \qquad (3) \\
&= \frac{\tau_{|S|}\sum_{i,k\in S|i<k}(1-\frac{2}{\pi}\arcsin(\theta X^*_{ik}))}{2z^*}\cdot z^*.
\end{aligned}$$

Since (X^*, z^*) is a maximal solution of (MSS-SDP), we must have

$$z^* = \min\{1, \frac{1}{|S|-1}\sum_{i,k\in S|i<k}\frac{1-X^*_{ik}}{2}\}.$$

Then, one can verify that $(X, y, z) = (X^*_S, \frac{\mathrm{E}[u(S)]}{z^*}, z^*)$ is a feasible solution of (1). Thus

$$y = \frac{\mathrm{E}[u(S)]}{z^*} \geq \alpha_{|S|}$$

which gives us the desired result. ■

However, it is not easy to compute $\alpha_{|S|}$ directly from (1). What we are able to do is to provide a good lower bound for $\alpha_{|S|}$. The following table is the list of lower bounds for $\alpha_{|S|}$, $|S| = 2, ..., 15$, when $\theta = 0.9795$. The detail of how to compute these lower bounds will be presented in the Appendix.

Table 1. The lower bounds of $\alpha_{|S|}$ when $\theta = 0.9795$:

$\|S\|$	2	3	4	5	6	7	8
$\alpha_{\|S\|} \geq$	0.8711	0.9085	0.6952	0.6489	0.5778	0.5382	0.4940
$\|S\|$	9	10	11	12	13	14	15
$\alpha_{\|S\|} \geq$	0.4646	0.4297	0.4087	0.3845	0.3687	0.3510	0.3388

4 Perturbation

First, we strengthen the approximation quality for size $|S_j| = 2$ by the following analysis. Let $\phi(t) = \arccos(\theta(1-2t))/\pi$, $\gamma_\theta(\geq \frac{1}{2})$ be the minimizer of $\phi(t)/t$ in the interval $(0,1]$. It is easy to see that

$$\frac{\phi(\gamma_\theta)}{\gamma_\theta} = \min_{-1 \leq x < 1} \frac{1 - \frac{2}{\pi}\arcsin(\theta x)}{1-x}.$$

Let

$$W_2 = \sum_{j:\ |S_j|=2} w_j \quad \text{and} \quad B_2 = (\sum_{j:\ |S_j|=2} w_j z_j^*)/W_2.$$

The following lemma is an extension of the Theorem 3.1.1 of [6].

Lemma 4 *If $B_2 \geq \gamma_\theta$, then*

$$\sum_{j:\ |S_j|=2} w_j \mathrm{E}[u(S_j)] \geq \frac{\phi(B_2)}{B_2} \sum_{j:\ |S_j|=2} w_j z_j^*.$$

Proof. Let $\mu_j = w_j/W_2$, then $B_2 = \sum_{j:\ |S_j|=2} \mu_j z_j^*$. Note that $z_j^* = \frac{1-X^*_{i_j k_j}}{2}$ for $S_j = \{i_j, k_j\}$. Then we get

$$\sum_{j:\ |S_j|=2} w_j \mathrm{E}[u(S_j)] \quad = \quad \sum_{j:\ |S_j|=2} w_j \frac{1 - \frac{2}{\pi}\arcsin(\theta X^*_{i_j k_j})}{2}$$

$$
\begin{aligned}
&= \sum_{j:\ |S_j|=2} w_j \frac{\arccos(\theta X^*_{i_j k_j})}{\pi} \\
&= W_2 \cdot \sum_{j:\ |S_j|=2} \mu_j \frac{\arccos(\theta(1-2z^*_j))}{\pi} \\
&= W_2 \cdot \sum_{j:\ |S_j|=2} \mu_j \phi(z^*_j).
\end{aligned}
$$

Let

$$
\tilde{\phi}(t) = \begin{cases} \phi(t), & \text{if } t \geq \gamma_\theta; \\ \frac{\phi(\gamma_\theta)}{\gamma_\theta} t, & \text{if } 0 \leq t \leq \gamma_\theta. \end{cases}
$$

By an analysis similar to [6], one can show that $\tilde{\phi}(t)$ is a convex function in $[0,1]$ and $\tilde{\phi}(t) \leq \phi(t)$. Note that $\sum_{j:\ |S_j|=2} \mu_j = 1$, then

$$
\sum_{j:\ |S_j|=2} \mu_j \phi(z^*_j) \geq \sum_{j:\ |S_j|=2} \mu_j \tilde{\phi}(z^*_j) \geq \tilde{\phi}(\sum_{j:\ |S_j|=2} \mu_j z^*_j) = \tilde{\phi}(B_2) = \phi(B_2).
$$

The last equality holds since $B_2 \geq \gamma_\theta$. It follows that

$$
\sum_{j:\ |S_j|=2} w_j \mathrm{E}[u(S_j)] \geq W_2 \cdot \phi(B_2) = \frac{\phi(B_2)}{B_2} \sum_{j:\ |S_j|=2} w_j z^*_j.
$$

■

By the analysis of last two sections, after the first two steps of the algorithm the partition of $u(S_j)$ given by $\hat{x}$ has the following property:

$$
\mathrm{E}[u(S_j)] \geq \alpha_{|S_j|} z^*_j,
$$

where α_k, $k \geq 2$, is defined by (1). In particular, for subset S_j with size $|S_j| = 2$,

$$
\alpha_2 \geq \min_{-1 \leq x < 1} \frac{1 - \frac{2}{\pi}\arcsin(\theta x)}{1-x},
$$

and when $B_2 \geq \gamma_\theta$,

$$
\sum_{j:\ |S_j|=2} w_j \mathrm{E}[u(S_j)] \geq \frac{\phi(B_2)}{B_2} \sum_{j:\ |S_j|=2} w_j z^*_j.
$$

Now, let $\tilde{u}(S_j) = 1$ if the subset S_j has been split by the partition given by $\tilde{x}$ and $\tilde{u}(S_j) = 0$ otherwise (i.e., after the perturbation step). Recall that

$$\tilde{x}_i = \begin{cases} -\hat{x}_i, & \text{with probability } p; \\ \hat{x}_i, & \text{with probability } 1-p. \end{cases}$$

The following lemma is due to [3].

Lemma 5

$$\mathrm{E}[\tilde{u}(S_j)] \geq$$

$$\mathrm{E}[u(S_j)]\cdot\Big(1 - p(1-p)^{|S_j|-1} - p^{|S_j|-1}(1-p)\Big) + (1-\mathrm{E}[u(S_j)])\cdot\Big(1 - p^{|S_j|} - (1-p)^{|S_j|}\Big).$$

Then we analyze the ratio between Algorithm MSS and the SDP relaxation for subsets with different sizes.

Corollary 3 *For $k \geq 2$ and $p \leq 1/2$,*

$$\sum_{j:|S_j|=k} w_j \mathrm{E}[\tilde{u}(S_j)] \geq$$

$$\sum_{j:|S_j|=k} w_j z_j^* \cdot \{\alpha_k \cdot \Big(1 - p(1-p)^{k-1} - p^{k-1}(1-p)\Big) + (1-\alpha_k)\cdot\Big(1 - p^k - (1-p)^k\Big)\}.$$

For the subset with size $|S_j| = 2$, the ratio can be further improved.

Corollary 4

$$\begin{aligned} & \sum_{j:|S_j|=2} w_j \mathrm{E}[\tilde{u}(S_j)] \\ \geq \quad & \min_{\gamma_\theta \leq B_2 \leq 1} \frac{\phi(B_2)(1-2p)^2 + (2p-2p^2)}{B_2} \cdot \sum_{j:|S_j|=2} w_j z_j^*. \end{aligned}$$

Proof. If $B_2 \geq \gamma_\theta$, by Lemma 5 and recall the definition of B_2, we get

$$\begin{aligned}
& \sum_{j:|S_j|=2} w_j \mathrm{E}[\tilde{u}(S_j)] \\
\geq\ & \sum_{j:|S_j|=2} w_j \cdot \{\mathrm{E}[u(S_j)] \cdot (1-2p(1-p)) + (1-\mathrm{E}[u(S_j)]) \cdot \left(1-p^2-(1-p)^2\right)\} \\
=\ & \sum_{j:|S_j|=2} w_j \mathrm{E}[u(S_j)] \cdot (1-2p)^2 + \sum_{j:|S_j|=2} w_j (2p-2p^2) \\
\geq\ & \frac{\phi(B_2)}{B_2} \sum_{j:\ |S_j|=2} w_j z_j^* \cdot (1-2p)^2 + \frac{1}{B_2} \sum_{j:\ |S_j|=2} w_j z_j^* \cdot (2p-2p^2) \\
=\ & \frac{\phi(B_2)\,(1-2p)^2 + (2p-2p^2)}{B_2} \cdot \sum_{j:|S_j|=2} w_j z_j^*.
\end{aligned}$$

If $B_2 < \gamma_\theta$, we have

$$\begin{aligned}
& \sum_{j:|S_j|=2} w_j \mathrm{E}[\tilde{u}(S_j)] \\
\geq\ & \sum_{j:|S_j|=2} w_j \mathrm{E}[u(S_j)] \cdot (1-2p)^2 + \sum_{j:|S_j|=2} w_j (2p-2p^2) \\
\geq\ & \alpha_2 \sum_{j:\ |S_j|=2} w_j z_j^* \cdot (1-2p)^2 + \frac{1}{B_2} \sum_{j:\ |S_j|=2} w_j z_j^* \cdot (2p-2p^2) \\
=\ & (\alpha_2 \cdot (1-2p)^2 + \frac{(2p-2p^2)}{B_2}) \cdot \sum_{j:|S_j|=2} w_j z_j^*.
\end{aligned}$$

But

$$\begin{aligned}
& \alpha_2 \cdot (1-2p)^2 + \frac{(2p-2p)^2}{B_2} \\
=\ & \min_{-1\leq x<1} \frac{1-\frac{2}{\pi}\arcsin(\theta x)}{1-x} \cdot (1-2p)^2 + \frac{(2p-2p^2)}{B_2} \\
=\ & \frac{\phi(\gamma_\theta)}{\gamma_\theta} \cdot (1-2p)^2 + \frac{(2p-2p^2)}{B_2} \\
>\ & \frac{\phi(\gamma_\theta)}{\gamma_\theta} \cdot (1-2p)^2 + \frac{(2p-2p^2)}{\gamma_\theta} \\
\geq\ & \min_{\gamma_\theta \leq B_2 \leq 1} \frac{\phi(B_2)(1-2p)^2 + (2p-2p^2)}{B_2}.
\end{aligned}$$

■

For a given θ, we can run the Algorithm MSS for

$$p \in \{0, 0.005, 0.01, 0.015, 0.02, \cdots, 0.495, 0.5\},$$

and then choose the best output. Let $p_i = 0.005 \cdot i$ for $i = 0, 1, 2, \cdots 100$. For each p_i, we can get a ratio $\alpha_k^{(i)}$ for the subset with size k, where for $k \geq 3$,

$$\alpha_k^{(i)} = \alpha_k \cdot \left(1 - p_i(1-p_i)^{k-1} - p_i^{k-1}(1-p_i)\right) + (1-\alpha_k) \cdot \left(1 - p_i^k - (1-p_i)^k\right),$$

and

$$\alpha_2^{(i)} = \min_{\gamma_\theta \leq B_2 \leq 1} \frac{\phi(B_2)}{B_2} \cdot (1-2p_i)^2 + \frac{(2p_i - 2p_i^2)}{B_2}.$$

It is easy to see the performance guarantee of Algorithm MSS for given $0 \leq \theta \leq 1$ is at least $\max_{i=0,1,\ldots,100}\{\min_{k=2,\ldots,T} \alpha_k^{(i)}\}$.

We can see that when p_i increases, $\alpha_k^{(i)}$ increases for a larger k, and decreases for a smaller k. This motivates us to analyze the performance guarantee of Algorithm MSS by considering the contribution of different size (see [3]). We let

$$r_k = \frac{\sum_{j:|S_j|=k} w_j z_j^*}{\sum_{1 \leq j \leq n} w_j z_j^*},$$

then the following minimization problem gives us a lower bound of the overall performance guarantee:

$$\begin{array}{ll} \text{Minimize} & R \\ \text{subject to} & \sum_{k=1}^{T} \alpha_k^{(i)} r_k \leq R \quad \text{for} \quad i = 0, 1, \cdots, 100, \\ & \sum_{k=1}^{T} r_k = 1, \\ & 0 \leq r_k \leq 1 \quad \text{for} \quad k = 2, 3, \cdots, T. \end{array} \tag{4}$$

If we only consider the subsets with size $k \leq 15$, that is, $T = 15$, we have that $R \geq 0.74996$ when we set $\theta = 0.9795$. The worst case is $r_2 = 0.543$, $r_6 = 0.457$ and

$r_k = 0$ for other $k(\leq 15)$. Note that for $k \geq 15$, $\alpha_k^{(i)} \geq 0.79 > 0.7499$ even if we set $\alpha_k = 0$. Therefore it is sufficient to consider $k \leq 15$ for analyzing the performance guarantee of our algorithm. To summarize, we have

Theorem 2 *The worst-case performance ratio of Algorithm MSS is at least* 0.7499*.*

5 Max NAE SAT

As we mentioned in the introduction, *Max Set Splitting* is a special case of *Max NAE SAT* in which all literals appear unnegated. A small modification of Algorithm MSS described in Section 2 will also give a 0.7499-approximation algorithm for *Max NAE SAT*.

This can be seen as follows. If a variable x_i occurs negated in a clause, we define a new variable x_{m+i} which equals to $-x_i$ ([11]). This means that we will have $2m$ variables and then the corresponding SDP relaxation has an unknown variable which is a $2m \times 2m$ matrix. The SDP relaxation for *Max NAE SAT* is:

$$
\begin{aligned}
w^{SDP} := \qquad & \text{Maximize} \quad \sum_{j=1}^{n} w_j z_j \\
\text{(MNS-SDP)} \qquad & \\
& \text{subject to} \quad \frac{1}{|S_j|-1} \sum_{i,k \in S_j | i<k} \frac{1-X_{ik}}{2} \geq z_j \quad \text{for any} \quad S_j \in C \\
& \sum_{i,k \in S_j | i<k} X_{ik} \geq -\lfloor \frac{|S_j|}{2} \rfloor \quad \text{for any} \quad S_j \in C \\
& X_{ii} = 1, \quad \text{for} \quad i = 1, \ldots, 2m, \\
& X_{i,m+i} = -1, \quad \text{for} \quad i = 1, \ldots, m, \\
& X \succeq 0, \\
& 0 \leq z_j \leq 1, \quad \text{for} \quad i = 1, \ldots, n.
\end{aligned}
$$

The algorithm for *Max NAE SAT* is the same as Algorithm MSS except the rounding procedure. When we round the optimal solution X^* of (MNS-SDP) to $\{-1, 1\}$ solution $\hat{x}$, to make sure $\hat{x}_i = -\hat{x}_{m+i}$, we have to use the covariance matrix $\theta X^* + (1-\theta) NI$ instead of $\theta X^* + (1-\theta)I$, where

$$NI = \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}.$$

Note that NI is also semidefinite, and

$$\mathrm{E}[\hat{x}_i \hat{x}_{m+i}] = \frac{2}{\pi} \arcsin(\theta X^*_{i,m+i} + (1-\theta) \cdot (-1)) = \frac{2}{\pi} \arcsin(-1) = -1.$$

Therefore, with probability 1, $\hat{x}_i \hat{x}_{m+i} = -1$.

Using the same analysis as that for *Max Set Splitting*, we get a 0.7499-approximation algorithm for *Max NAE SAT*.

6 Satisfiable instances

In this section, we consider the satisfiable instances of *Max NAE SAT* problem, i.e., all the clauses in C can be satisfied in an optimal assignment. Our main result is

Theorem 3 *The satisfiable instances of Max NAE SAT can be approximated with a factor of (at least)* 0.8097 *in polynomial time.*

We first consider the case in which each clause contains at least 3 literals. Recall the definition of $\alpha_{|S|}$ and r_k. For solving the satisfiable instances, we have every $z_j^* = 1$ in the optimal solution of (MNS-SDP), and thus both of them can be strengthened as:

$$\alpha_{|S|} := \quad \text{Minimize}_{X,y,z} \qquad y$$

$$\begin{aligned}
\text{Subject to}\qquad & y \geq \max_{i,k|1\leq i,k\leq |S|}\{\frac{1-\frac{2}{\pi}\arcsin(\theta X_{ik})}{2}\} \\
& y \geq \frac{\tau_{|S|}\sum_{i,k|1\leq i<k\leq |S|}(1-\frac{2}{\pi}\arcsin(\theta X_{ik}))}{2} \\
& \frac{(|S|-4)(|S|-1)}{2} \geq \sum_{i,k|1\leq i<k\leq |S|} X_{ik} \geq -\lfloor\frac{|S|}{2}\rfloor \qquad (5) \\
& X_{ii} = 1, \quad \text{for} \quad i = 1,\ldots,|S| \\
& X \succeq 0.
\end{aligned}$$

and

$$r_k = \frac{\sum_{j:|S_j|=k} w_j z_j^*}{\sum_{1\leq j\leq n} w_j z_j^*} = \frac{\sum_{j:|S_j|=k} w_j}{\sum_{1\leq j\leq n} w_j}.$$

In particular, the lower bound of α_3 can be improved by the following lemma.

Lemma 6 *For the satisfiable instances, α_3 is bounded below by the optimal value of the following minimization problem*

$$\begin{aligned}
\textit{Minimize}\quad & \frac{1}{4}(3-\frac{2}{\pi}\arcsin(\theta X_{ij})-\frac{2}{\pi}\arcsin(\theta X_{il})-\frac{2}{\pi}\arcsin(\theta X_{jl})) \\
\textit{Subject to}\quad & X_{ij}+X_{il}+X_{jl} = -1 \qquad (6) \\
& -1 \leq X_{ij}, X_{il}, X_{jl} \leq 1
\end{aligned}$$

Proof. Note that, for the satisfiable instances, $z^* = 1$, which implies that $X_{ij}^* + X_{il}^* + X_{jl}^* = -1$. By the second constraint of (1) and the fact that $\tau_3 = 1/2$, we conclude the proof. ■

The optimal solution of (6) can be easily computed. See Lemma 12 in the Appendix and note that minimization problem (6) is a special case of minimization problem (8) in which $\lambda = -1$ and $\beta = -1$. we have verified that, if $\theta = 0.9795$, then

$\alpha_3 \geq 0.9088$ which is slightly better than the lower bound provided by (1) (see Table 1).

However, for $|S| \geq 4$, we will simply use the lower bound of $\alpha_{|S|}$ as defined in (1), since we have found that the lower bound is always achieved at $z^* = 1$. Table 2 is the list of lower bounds for $\alpha_{|S|}$, $|S| = 3, \ldots, 25$, when $\theta = 0.90$ for satisfiable instances.

Table 2. The lower bounds of $\alpha_{|S|}$ for satisfiable instances when $\theta = 0.90$:

$\|S\|$	3	4	5	6	7	8	9	10
$\alpha_{\|S\|} \geq$	0.8954	0.7203	0.6876	0.6203	0.5962	0.5545	0.5400	0.5150
$\|S\|$	11	12	13	14	15	16	17	18
$\alpha_{\|S\|} \geq$	0.5008	0.4848	0.4752	0.4617	0.4547	0.4447	0.4381	0.4299
$\|S\|$	19	20	21	22	23	24	25	
$\alpha_{\|S\|} \geq$	0.4250	0.4182	0.4139	0.4086	0.4048	0.4001	0.3969	

In order to give a lower bound for the performance guarantee of our algorithm, we consider LP (4) with one more constraint $r_2 = 0$ (since we are dealing with the case in which all clauses contain at least 3 literals). We choose $\theta = 0.90$ and need to consider $k \leq 25$. Then, we get that the optimal R is at least 0.8097 and the worst case is $r_3 = 0.5417$, $r_4 = 0.4381$, $r_{14} = 0.0085$, $r_{15} = 0.0019$, $r_{16} = 0.0060$, $r_{17} = 0.0012$, $r_{18} = 0.0026$ and $r_k = 0$ for other $k (\leq 25)$.

Lemma 7 *For the satisfiable instances of Max NAE SAT with every clause containing at least* 3 *literals, the algorithm MSS has a performance guarantee at least* 0.8097.

The following lemma, combing with Lemma 7, enables us to prove Theorem 3.

Lemma 8 *For the satisfiable instances of Max NAE SAT, if there exists a R-approximation for the case that every clause contains at least* 3 *literals, then there exists a R-approximation for the case that every clause contains at least* 2 *literals.*

Proof. Suppose that there exists at least one clause that contains 2 literals. If the clause is $y_i \vee y_j$, then we must have $y_i = \bar{y}_j$; if the clause is $y_i \vee \bar{y}_j$, then we have $y_i = y_j$. Then we can substitute y_i by either y_j or $\bar{y}_j$ such that all the clauses are still satisfiable. By this substitution, we reduced one literal and one clause that contains 2 literals. Note that after the substitution, y_j may appear twice in a single clause, or both y_j and $\bar{y}_j$ appear in a single clause. In the former case, we can simply delete one y_j; in the latter case, we can simply remove the clause.

By keeping doing this substitution, we get a reduced satisfiable instance where each clause contains at least 3 literals, and each literal appear at most once in a single clause. Since there are at most m literals, and n clauses, the substitution can be done by at most mn times.

Now we apply the R-approximation algorithm for the reduced satisfiable instance, which give us a true and false assignment for the reduced literals. We can easily extend this assignment to the original literals such that all the clauses that contain two literals are still satisfiable. Then we get a R-approximation for the original satisfiable instance. ■

Since *Max Set Splitting* is a special case of *Max NAE SAT*, we also obtained a 0.8097-approximation algorithm for the satisfiable *Max Set Splitting*.

Acknowledgment We would like to thank the referees for their helpful comments that improved the presentation of the paper. In particular, Lemma 8 was suggested by one of the referees, which led to the 0.8097-approximation for the satisfiable instances

of *Max NAE SAT*.

References

[1] A. A. Ageev and M. I. Sviridenko, "An Approximation Algorithm for Hypergraph Max k-Cut with Given Sizes of Parts," *Proceedings of ESA'2000*, Lecture Notes in Computer Science 1879, (2000) 32-41.

[2] P. Alimonti, "Non-oblivious Local Search for Graph and Hypergraph Coloring Problems," In *Proc. of 21st International Workshop on Graph-Theoretic Concepts in Comp. Sci.*, Lecture Notes in Computer Sciences 1017, Springer-Verlag, Berlin, (1995) 167-180.

[3] G. Andersson and L. Engebrestsen, "Better Approximation Algorithms for Set Splitting and Not-All-Equal SAT," *Information Processing Letters* 65, (1998) 305-311.

[4] S. Arora, D. Karger and M. Karpinski, "Polynomial Time Approximation Schemes for Dense Instances of NP-hard Problems," *Journal of Computer and System Science*, 58, (1999) 193-210 .

[5] D. Bertsimas and Y. Ye, "Semidefinite Relaxations, Multivariate Normal Distributions, and Order Statistics," *Handbook of Combinatorial Optimization*, Vol (3), D.-Z. Du and P.M. Pardalos (Eds.), Kluwer Academic Publishers, (1998) 1-19.

[6] M. X. Goemans and D. P. Williamson, "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming," *Journal of the ACM*, 42, (1995) 1115-1145.

[7] Q. Han, Y. Ye and J. Zhang, "An Improved Rounding Method and Semidefinite Programming Relaxation for Graph Partition," *Mathematical Programming*, to appear.

[8] V. Kann, J. Lagergren and A. Panconesi, "Approximability of Maximum Splitting of k-Sets and Some Other APX-complete Problems," *Information Processing Letters*, 58, (1996) 105-110.

[9] V. Guruswami, "The Approximability of Set Splitting Problems and Satisfiability Problems with no Mixed Clauses," *Proceedings of APPROX'2000,* Lecture Notes in Computer Science 1913 (2000) 155-166.

[10] J. Hastad, "Some Optimal Inapproximability Results," *29th STOC*, (1997) 1-10.

[11] H. Karloff and U. Zwick, "A 7/8-approximation algorithm for MAX 3SAT?" *38th FOCS*, (1997), 406-415.

[12] L. Lovász, "Coverings and Colorings of Hypergraphs," In *Proc. 4th Southeastern Conf. on Combinatorics, Graph Theory, and Computing*, Utilitas Mathematica Publishing, Winnipeg, (1973) 3-12.

[13] Yu. E. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optimization Methods and Software*, 9, (1998) 141–160.

[14] C. H. Papadimitriou and M. Yannakakis, "Optimization, Approximation, and Complexity Classes," *J. Comput. System. Sci.*, 43, (1991) 425-440.

[15] E. Petrank, "The Hardness of Approximation: Gap Location," *Computational Complexity*, 4, (1994) 133-157.

[16] L. Trevisan, G. B. Sorkin, M. Sudan and D. P. Williamson. "Gadgets, approximation and linear programming, " *SIAM Journal on Computing*, 29(6), (2000) 2074-2097.

[17] Y. Ye, "A .699-Approximation Algorithm for Max-Bisection," *Mathematical Programming*, 90, (2001) 101-111.

[18] U. Zwick, "Approximation Algorithms for Constraint Satisfaction Problems Involving at Most Three Variables Per Constraint," *9th SODA*, (1998) 201-210.

[19] U. Zwick, "Outward Rotations: A Tool for Rounding Solutions of Semidefinite Programming Relaxations, with Application to MAX CUT and Other Problems," *31th STOC*, (1999) 679-687.

7 Appendix: A lower bound for $\alpha_{|S|}$

In this appendix, we show how to numerically compute the lower bound for $\alpha_{|S|}$ that is defined by the minimum value of the optimization problem (1). We first parameterize the problem by introducing a parameter of X in (1):

$$\lambda := \sum_{i,k \in S | i<k} X_{ik}.$$

Noting the total number of variables

$$N_{|S|} := |\{X_{ik} : \ 1 \le i < k \le |S|\}| = \frac{|S|(|S|-1)}{2},$$

every variable $-1 \le X_{ik} \le 1$ for $i < k$, and by the third and fourth constraints of (1), we have

$$-\lfloor \frac{|S|}{2} \rfloor \le \lambda < N_{|S|}. \tag{7}$$

Secondly, based on the range of λ we can construct a close form for z in (1):

Lemma 9

$$z = \begin{cases} 1, & \textit{if and only if } \lambda \in [-\lfloor\frac{|S|}{2}\rfloor, \frac{(|S|-4)(|S|-1)}{2}]; \\ \frac{N_{|S|}-\lambda}{2(|S|-1)} > 0, & \textit{if and only if } \lambda \in [\frac{(|S|-4)(|S|-1)}{2}, N_{|S|}). \end{cases}$$

Proof. From the third constraint of (1),

$$z = \min\{1, \frac{1}{|S|-1}\sum_{i<k}\frac{1-X_{ik}}{2}\} > 0.$$

If $\lambda \in [-\lfloor\frac{|S|}{2}\rfloor, \frac{(|S|-4)(|S|-1)}{2}]$, then $\frac{1}{|S|-1}\sum_{i<k}\frac{1-X_{ik}}{2} \geq 1$, and thus $z = 1$. If $\lambda \in [\frac{(|S|-4)(|S|-1)}{2}, N_{|S|})$, then $\frac{1}{|S|-1}\sum_{i<k}\frac{1-X^*_{ik}}{2} \leq 1$, and thus $z = \frac{N_{|S|}-\lambda}{2(|S|-1)} > 0$.

One can verify that the reverse is also true. ■

Now we can construct a lower bound for $\alpha_{|S|}$ by relaxing the minimization problem (1) to a relatively "easier" parameterized problem for any given constants $\theta \in (0,1]$ and $\beta \in [-1,1)$, and for any given parameter $\lambda \in [-\lfloor\frac{|S|}{2}\rfloor, N_{|S|})$. Let

$$\begin{aligned} f(\lambda,\theta,\beta) := \quad & \text{Minimize} \quad \sum_{i<k}(1-\frac{2}{\pi}\arcsin(\theta X_{ik})) \\ & \text{Subject to} \quad \sum_{i<k} X_{ik} = \lambda \\ & \qquad\qquad\quad \beta \leq X_{ik} \leq 1 \quad \text{for all} \quad 1 \leq i < k \leq |S|. \end{aligned} \tag{8}$$

Note that the constraints of (8) are linear. Furthermore, via Lemma 9, let

$$\gamma_{\theta,\beta}(\lambda) := \begin{cases} \frac{\tau_{|S|}}{2} f(\lambda,\theta,\beta), & \text{if } \lambda \in [-\lfloor\frac{|S|}{2}\rfloor, \frac{(|S|-4)(|S|-1)}{2}]; \\ \tau_{|S|}(|S|-1)\frac{f(\lambda,\theta,\beta)}{N_{|S|}-\lambda}, & \text{if } \lambda \in [\frac{(|S|-4)(|S|-1)}{2}, N_{|S|}); \end{cases} \tag{9}$$

and, for given constants $\theta \in (0,1]$ and $\beta \in [-1,1)$, consider the one-dimensional minimization problem:

$$\gamma^*_{\theta,\beta} := \min_{-\lfloor\frac{|S|}{2}\rfloor \leq \lambda < N_{|S|}} \gamma_{\theta,\beta}(\lambda). \tag{10}$$

Then, we have

Lemma 10 *For any given $\theta \in (0,1]$ and $\beta \in [-1,1]$,*

$$\alpha_{|S|} \geq \min\{\frac{1-\frac{2}{\pi}\arcsin(\theta\beta)}{2}, \gamma^*_{\theta,\beta}\}.$$

Proof. Consider any minimal solution (y^*, X^*, z^*) of (1). If there exist $i < k$ such that $X^*_{ik} < \beta$, by the first inequality constraint of (1), we then have

$$\alpha_{|S|} \geq y^* \geq \frac{1-\frac{2}{\pi}\arcsin(\theta\beta)}{2}.$$

Otherwise $X^*_{ik} \geq \beta$ for all $i < k$, and then X^* is a feasible solution of (8) for some $\lambda(X^*) \in [-\lfloor\frac{|S|}{2}\rfloor, N_{|S|})$. From the second inequality constraint of (1), and then from Lemma 9, (9) and (10), we must have

$$\alpha_{|S|} \geq y^* \geq \frac{\tau_{|S|}\sum_{i<k}(1-\dfrac{2}{\pi}\arcsin(\theta X^*_{ik}))}{2z^*} \geq \frac{\tau_{|S|}}{2z^*}f(\lambda(X^*),\theta,\beta) = \gamma_{\theta,\beta}(\lambda(X^*)) \geq \gamma^*_{\theta,\beta}.$$

■

In order to provide a good lower bound of $\alpha_{|S|}$, for given $\theta \in (0,1]$ we can choose a suitable $\beta \in [-1,1)$, such that $\frac{1-\frac{2}{\pi}\arcsin(\theta\beta)}{2}$ and $\gamma^*_{\theta,\beta}$ are numerically close. For example, for $|S| = 6$ and $\theta = 0.9795$, we choose $\beta = -0.2471$. It is easy to verify that $\frac{1-\frac{2}{\pi}\arcsin(\theta\beta)}{2} > 0.5778$. If we can show that $\gamma^*_{\theta,\beta} > 0.5778$, then we can bound $\alpha_{|S|} > 0.5778$ for $|S| = 6$ and $\theta = 0.9795$. (We choose θ to maximize the overall approximation bound R).

Since $\frac{1-\frac{2}{\pi}\arcsin(\theta\beta)}{2}$ is easy to compute for any given θ and β, we focus on how to numerically calculate $\gamma^*_{\theta,\beta}$. Note that $\gamma^*_{\theta,\beta}$ is computed by one-dimensional search of λ on $\gamma_{\theta,\beta}(\lambda)$ of (9), and (8). Thus, the key is to construct an analytical form of $f(\lambda,\theta,\beta)$ of (8) for any given parameter λ and constants θ and β.

We present a minimal solution structure for linearly constrained minimization problem (8).

Lemma 11 *There is a minimal solution X^* of (8) such that every variable X^*_{ik} takes one of the three values, $\{\beta, \alpha, 1\}$, for some $\alpha \in (\beta, 1)$.*

Proof. First, in any minimal solution X^* of (8), each variable X^*_{ik} could take value β (the lower bound), 1 (the upper bound), or some value in $(\beta, 1)$ (the interior of $[\beta, 1]$). And every x^*_{ik} in $(\beta, 1)$ must satisfy the first order necessary, called the Karush-Kuhn-Tucker (KKT), condition. That is, there exists a Lagrange multiplier u for the only equality linear constraint in (8) such that every interior X^*_{ik} meets

$$-\frac{2\theta}{\pi\sqrt{1-(\theta X^*_{ik})^2}} = u. \tag{11}$$

Therefore, every interior X^*_{ik} (if exists) satisfies

$$(X^*_{ik})^2 = \frac{1-(\frac{2\theta}{\pi \cdot u})^2}{\theta^2}.$$

Let $\alpha' = \frac{\sqrt{1-(\frac{2\theta}{\pi \cdot u})^2}}{\theta}$. Then $\alpha' \geq 0$ and by the constraint $\beta \leq X_{ik} \leq 1$, we must have $\beta < \alpha' < 1$.

Now we complete the proof by considering the following cases:

Case 1. If all the interiors equal to α', then we must have $\beta < \alpha' < 1$. In this case, the lemma is true by letting $\alpha = \alpha'$.

Case 2. If all the interiors equal to $-\alpha'$, then we must have $\beta < -\alpha' < 1$. In this case, the lemma is also true by letting $\alpha = -\alpha'$.

Case 3. There exist at least two interiors, say X^*_{12} and X^*_{13}, such that $1 > X^*_{12} = \alpha'$ and $X^*_{13} = -\alpha' > \beta$. Suppose that $\alpha' > 0$. Since

$$2 - \frac{2}{\pi}\arcsin(\theta \cdot \alpha') - \frac{2}{\pi}\arcsin(\theta \cdot (-\alpha')) = 2 - 2 \cdot \frac{2}{\pi}\arcsin(\theta \cdot 0),$$

we can construct a new minimal solution by switching both $X^*_{12} = X^*_{13} = 0$ while keeping all other variables unchanged (note that all the constraints of (8) are still satisfied). In the new minimal solution, there should be no more interior variable equal α' or $-\alpha'$, because, otherwise, we have a minimal solution whose interior variables take both 0, and $\alpha'(> 0)$ or $-\alpha'(< 0)$, which contradicts (11) that all interior variables' square, $(X^*_{ik})^2$, are equal. Therefore, $\alpha' = 0$ and then Case 3 can be reduced to either Case 1 or Case 2. ■

Let the mixed integer and continuous function

$$f_{\theta,\beta}(\lambda, \overline{N}, \underline{N}) = N_{|S|} - \overline{N} \cdot \frac{2}{\pi} \arcsin(\theta) - \underline{N} \cdot \frac{2}{\pi} \arcsin(\theta\beta) - \\ (N_{|S|} - \overline{N} - \underline{N}) \cdot \frac{2}{\pi} \arcsin\left(\theta \frac{\lambda - \overline{N} - \underline{N}\beta}{N_{|S|} - \overline{N} - \underline{N}}\right). \tag{12}$$

Lemma 12 *For any $\lambda \in [-\lfloor \frac{|S|}{2} \rfloor, N_{|S|})$, there exist two nonnegative integers $\overline{N}(\lambda)$ and $\underline{N}(\lambda)$, such that*

(i). $\overline{N}(\lambda) + \underline{N}(\lambda) \leq N_{|S|} - 1$;

(ii). $N_{|S|}\beta + \overline{N}(\lambda)(1 - \beta) \leq \lambda \leq N_{|S|} - \underline{N}(\lambda)(1 - \beta)$;

(iii). The optimal value of (8) $f(\lambda, \theta, \beta) = f_{\theta,\beta}(\lambda, \overline{N}(\lambda), \underline{N}(\lambda))$.

Proof. Suppose that, in the optimal solution of (8), the number of X^*_{ik}s that equal to 1 is $\overline{N}'$, and the number of X^*_{ik}s that equal to β is $\underline{N}'$. It is easy to see that $\overline{N}'$ and $\underline{N}'$ are nonnegative integers and $\overline{N}' + \underline{N}' \leq N_{|S|}$.

If $\overline{N}' + \underline{N}' = N_{|S|}$, we must have $\overline{N}' < N_{|S|}$ and $\underline{N}' - 1 \geq 0$ since $\lambda < N_{|S|}$. In this case, let $\overline{N}(\lambda) = \overline{N}' \geq 0$ and $\underline{N}(\lambda) = \underline{N}' - 1 \geq 0$ such that $\overline{N}(\lambda) + \underline{N}(\lambda) \leq N_{|S|} - 1$. Then $N_{|S|} - \overline{N}(\lambda) - \underline{N}(\lambda) = 1$, and by the linear constraint of (8)

$$\frac{\lambda - \overline{N}(\lambda) - \underline{N}(\lambda)\beta}{N_{|S|} - \overline{N}(\lambda) - \underline{N}(\lambda)} = \beta.$$

Therefore, the optimal value of (8) is

$$f(\lambda,\theta,\beta)=N_{|S|}-\overline{N}'\cdot\frac{2}{\pi}\arcsin(\theta)-\underline{N}'\cdot\frac{2}{\pi}\arcsin(\theta\beta)=f_{\theta,\beta}(\lambda,\overline{N}(\lambda),\underline{N}(\lambda)).$$

If $\overline{N}'+\underline{N}'\le N_{|S|}-1$, then by Lemma 11, all other X_{ik}^*s must equal to

$$\alpha=\frac{\lambda-\overline{N}'-\underline{N}'\beta}{N_{|S|}-\overline{N}'-\underline{N}'}\in(\beta,1).$$

In this case, let $\overline{N}(\lambda)=\overline{N}'\ge 0$ and $\underline{N}(\lambda)=\underline{N}'\ge 0$ such that $\overline{N}(\lambda)+\underline{N}(\lambda)\le N_{|S|}-1$.. Therefore, the minimal objective value function of (8) also has the form

$$f(\lambda,\theta,\beta)=f_{\theta,\beta}(\lambda,\overline{N}(\lambda),\underline{N}(\lambda)).$$

We have proved (i) and (iii). In order to prove (ii), note that in either case we have

$$\beta\le\frac{\lambda-\overline{N}(\lambda)-\underline{N}(\lambda)\beta}{N_{|S|}-\overline{N}(\lambda)-\underline{N}(\lambda)}\le 1$$

and thus $N_{|S|}\beta+\overline{N}(\lambda)(1-\beta)\le\lambda\le N_{|S|}-\underline{N}(\lambda)(1-\beta)$. ■

Let

$$\gamma_{\theta,\beta}(\lambda,\overline{N},\underline{N}):=\begin{cases}\frac{\tau_{|S|}}{2}f_{\theta,\beta}(\lambda,\overline{N},\underline{N}), & \text{if } \lambda\in[-\lfloor\frac{|S|}{2}\rfloor,\frac{(|S|-4)(|S|-1)}{2}];\\ \tau_{|S|}(|S|-1)\frac{f_{\theta,\beta}(\lambda,\overline{N},\underline{N})}{N_{|S|}-\lambda}, & \text{if } \lambda\in[\frac{(|S|-4)(|S|-1)}{2},N_{|S|}).\end{cases}\tag{13}$$

Furthermore, for any given nonnegative integers $\overline{N}$ and $\underline{N}$, define a one-dimensional minimization problem

$$\gamma_{\theta,\beta}(\overline{N},\underline{N}):=\min_{\lambda:\ \in[-\lfloor\frac{|S|}{2}\rfloor,N_{|S|})\ \cap\ [N_{|S|}\beta+\overline{N}(1-\beta),N_{|S|}-\underline{N}(1-\beta)]}\gamma_{\theta,\beta}(\lambda,\overline{N},\underline{N}).\tag{14}$$

Lemma 12 leads to the following:

Corollary 5 *There exist two nonnegative integers $\overline{N}^*$ and $\underline{N}^*$, such that $\overline{N}^*+\underline{N}^*\le N_{|S|}-1$ and $\gamma^*_{\theta,\beta}\ge\gamma_{\theta,\beta}(\overline{N}^*,\underline{N}^*)$.*

Proof. Let λ^* be the minimizer of the minimization problem (10), i.e., $\gamma^*_{\theta,\beta} = \gamma_{\theta,\beta}(\lambda^*)$. By Lemma 12, there exist two nonnegative integers $\overline{N}^* := \overline{N}(\lambda^*)$ and $\underline{N}^* := \underline{N}(\lambda^*)$ such that

$$f(\lambda^*, \theta, \beta) = f_{\theta,\beta}(\lambda^*, \overline{N}^*, \underline{N}^*).$$

Then by the definitions of $\gamma_{\theta,\beta}(\lambda, \overline{N}, \underline{N})$ and $\gamma_{\theta,\beta}(\overline{N}, \underline{N})$, we get $\gamma^*_{\theta,\beta} \geq \gamma_{\theta,\beta}(\overline{N}^*, \underline{N}^*)$. ■

Remarks. For our purpose, we need to prove only that $\gamma_{\theta,\beta}(\overline{N}^*, \underline{N}^*)$ is a lower bound of $\gamma^*_{\theta,\beta}$. However, one can easily verify that $\gamma^*_{\theta,\beta} = \gamma_{\theta,\beta}(\overline{N}^*, \underline{N}^*)$.

Now, suppose we have known $\overline{N}^*$ and $\underline{N}^*$. Then the minimization problem (10) is reduced to a easier problem (14). Of course, we do not know $\overline{N}^*$ and $\underline{N}^*$ in advance. However, we can enumerate all pairs of nonnegative integers $(\overline{N}^*, \underline{N}^*)$ with $\overline{N}^* + \underline{N}^* \leq N_{|S|} - 1$, which should be no more than $O(|S|^4)$ many. For each pair, we numerically solve (14) and obtain $\gamma_{\theta,\beta}(\overline{N}^*, \underline{N}^*)$. Then, we select $\gamma^*_{\theta,\beta}$ among all candidate pairs as

$$\gamma^*_{\theta,\beta} = \min_{(\overline{N}^*, \underline{N}^*)|\ \overline{N}^* \geq 0,\ \underline{N}^* \geq 0,\ \overline{N}^* + \underline{N}^* \leq N_{|S|}-1} \gamma_{\theta,\beta}(\overline{N}^*, \underline{N}^*).$$

For any fixed pair $(\overline{N}^*, \underline{N}^*)$, solving problem of (14) is straightforward. If $\lambda^* \in [-\lfloor \frac{|S|}{2} \rfloor, \frac{(|S|-4)(|S|-1)}{2}]$, then from (13)

$$\gamma_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*) = \frac{\tau_{|S|}}{2} f_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*).$$

Since $f_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*)$ and thus $\gamma_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*)$ are monotonely decreasing in λ from (12), the minimum of $\gamma_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*)$ is obtained when λ^* reaches its upper limit, i.e.,

$$\lambda^* = \min\left\{\frac{(|S|-4)(|S|-1)}{2}, N_{|S|} - \underline{N}^*(1-\beta)\right\}.$$

In the case $\lambda^* \in [\frac{(|S|-4)(|S|-1)}{2}, N_{|S|})$, from (13)

$$\gamma_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*) = \tau_{|S|}(|S|-1) \cdot \frac{f_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*)}{N_{|S|}-\lambda} = \tau_{|S|}(|S|-1) \cdot$$

$$\frac{N_{|S|} - \overline{N}^* \cdot \frac{2}{\pi}\arcsin(\theta) - \underline{N}^* \cdot \frac{2}{\pi}\arcsin(\theta\beta) - (N_{|S|} - \overline{N}^* - \underline{N}^*) \cdot \frac{2}{\pi}\arcsin\left(\theta\frac{\lambda - \overline{N}^* - \underline{N}^*\beta}{N_{|S|} - \overline{N}^* - \underline{N}^*}\right)}{N_{|S|}-\lambda},$$

where again variable λ satisfies

$$N_{|S|}\beta + \overline{N}^*(1-\beta) \le \lambda \le N_{|S|} - \underline{N}^*(1-\beta).$$

Define a new variable

$$x = \frac{\lambda - \overline{N}^* - \underline{N}^*\beta}{N_{|S|} - \overline{N}^* - \underline{N}^*} \in [\beta, 1].$$

Then

$$\gamma_{\theta,\beta}(\lambda, \overline{N}^*, \underline{N}^*) = h(x) := \tau_{|S|}(|S|-1) \cdot \frac{A - \frac{2}{\pi}\arcsin(\theta x)}{B - x},$$

where

$$A = \frac{N_{|S|} - \overline{N}^* \cdot \frac{2}{\pi}\arcsin(\theta) - \underline{N}^* \cdot \frac{2}{\pi}\arcsin(\theta\beta)}{N_{|S|} - \overline{N}^* - \underline{N}^*} \text{ and } B = \frac{N_{|S|} - \overline{N}^* - \underline{N}^*\beta}{N_{|S|} - \overline{N}^* - \underline{N}^*}.$$

Note that $A, B \ge 1$. Therefore, minimizing $\gamma_{\theta,\beta}(\lambda)$ over $\lambda \in [N_{|S|}\beta + \overline{N}^*(1-\beta), N_{|S|} - \underline{N}^*(1-\beta)] \cap [\frac{(|S|-4)(|S|-1)}{2}, N_{|S|})$ is equivalent to minimizing $h(x)$ over

$$x \in [\beta, 1] \cap [\frac{\frac{(|S|-4)(|S|-1)}{2} - \overline{N}^* - \underline{N}^*\beta}{N_{|S|} - \overline{N}^* - \underline{N}^*}, \frac{N_{|S|} - \overline{N}^* - \underline{N}^*\beta}{N_{|S|} - \overline{N}^* - \underline{N}^*}).$$

But minimizing continuous $h(x)$ over the above interval can be done by computing its 2 stationary points and 2 boundary points, and then by choosing the optimal one among these points. In fact, for $A = B = 1$ and $|S| = 2$, $h(x)$ has been numerically calculated by Goemans and Williamson [6] for proving their 0.878 Max-Cut approximation result.

To summarize, for any given $|S|$, we can construct a lower bound for $\alpha_{|S|}$ by the following procedure:

Step 1. Choose suitable $\theta \in (0, 1]$, and $\beta \in [-1, 1)$.

Step 2. For every nonnegative integer pair $(\overline{N}^*, \underline{N}^*)$ such that $\overline{N}^* + \underline{N}^* \leq N_{|S|} - 1$, numerically solve the one-dimensional continuous minimization problem (14) to obtain its minimal value. Among all the minimal values $\gamma_{\theta,\beta}(\overline{N}^*, \underline{N}^*)$, choose the minimal one as $\gamma^*_{\theta,\beta}$.

Step 3. Choose $\min\{\frac{1-\frac{2}{\pi}\arcsin(\theta\beta)}{2}, \gamma^*_{\theta,\beta}\}$ as the lower bound for $\alpha_{|S|}$.

The lower bounds presented in Table 1 and Table 2 are generated by this procedure.