

PTAS for Minimax Approval Voting

Jarosław Byrka*, Krzysztof Sornat**

Institute of Computer Science, University of Wrocław
Joliot-Curie 15
50-383 Wrocław, Poland

Abstract. We consider Approval Voting systems where each voter decides on a subset of candidates he/she approves. We focus on the optimization problem of finding the committee of fixed size k , minimizing the maximal Hamming distance from a vote. In this paper we give a PTAS for this problem and hence resolve the open question raised by Carragianis et al. [AAAI'10]. The result is obtained by adapting the techniques developed by Li et al. [JACM'02] originally used for the less constrained Closest String problem. The technique relies on extracting information and structural properties of constant size subsets of votes.

1 Introduction

Approval Voting systems are widely considered [2] as an alternative to traditional elections, where each voter may select and support at most some small number of candidates. In Approval Voting each voter decides about every single candidate if he approves the candidate or does not approve him/her. A result is obtained by applying a predefined election rule to the set of collected votes.

In this paper we study the problem of implementing an appropriate election rule and focus on the Minimax objective [3]: we minimize the biggest dissatisfaction over voters. The resulting optimization problem is denoted *MAV*, and it is to select a committee composed of exactly k candidates, and minimizing the maximal symmetric difference between the committee and the set of approved candidates by a single voter.

Using the string terminology, votes are encoded as strings, and the goal is to find a string encoding a committee minimizing the maximal Hamming distance to an input string. Unlike in the related Closest String problem, in *MAV* there is also a constraint: the selected committee must be of fixed size k , and hence in the string terminology there must be exactly k ones in the string.

1.1 Related work and our results

Many different objective functions have been proposed and studied in the context of selecting the committee based on the set of votes collected in an Approval Voting system [1,2]. Clearly, optimizing the sum of Hamming distances to all votes

* jby@cs.uni.wroc.pl

** krzysztof.sornat@cs.uni.wroc.pl

is an easy task and can be done by simply selecting the k candidates approved by the largest number of voters. By contrast, Minimax Approval Voting was shown by LeGrand [6] to be NP-hard. LeGrand et al. [7] obtained 3-approximation by a very simple k -completion algorithm. Next, Carragianis et al. [5] gave the currently best 2-approximation algorithm. The algorithm was obtained by rounding a fractional solution to the natural LP relaxation of the problem, and obtained approximation ratio essentially matches the integrality gap of the LP.

In this paper we give a PTAS for the Minimax Approval Voting problem. Our work is based on the PTAS for Closest String [8], which is a similar problem to *MAV* but there we do not have the restriction on the number of 1's in the result. Technically, our contribution is the method of handling the number of 1's in the output. We also believe that our presentation is somewhat more intuitive.

Approval Voting systems are also analyzed in respect of manipulability, see e.g., [1] or [5]. In particular, [5] proved that each strategy-proof algorithm for *MAV* must have approximation ratio at least $2 - \frac{2}{k+1}$, which implies that our PTAS cannot be strategy-proof.

1.2 Definitions

We will use the following notation:

n – number of voters,

m – number of candidates,

$s_i \in \{0, 1\}^m$ – a vote of voter i ,

$s_i[j] = 1$ if voter i approves candidate j ,

$s_i[j] = 0$ if voter i does not approve candidate j ,

$S = \{s_1, s_2, \dots, s_n\}$ – the set of collected votes,

$s^{(1)} = |\{j : s[j] = 1\}|$ – the number of 1's in s .

For $x, y \in [0, 1]^m$ we define a distance $d(x, y) = \sum_{j=1}^m |x[j] - y[j]| = \|x - y\|_1$.

For $x, y \in \{0, 1\}^m$, $d(x, y)$ is called the Hamming distance.

Definition 1.

$$OPT = \min_{\substack{x \in \{0,1\}^m \\ x^{(1)}=k}} \max_{i \in \{1,2,\dots,n\}} d(x, s_i)$$

Let s_{OPT} be an optimal solution, i.e., $\max_{i \in \{1,2,\dots,n\}} d(s_{OPT}, s_i) = OPT$.

WLOG we assume that $n > k$. If not, we copy the first string $k - n + 1$ times.

1.3 The main idea behind our algorithm

The general idea behind our PTAS is to find a small enough subset X of votes that is a “good representation” of the whole set of votes S . Then the candidates are partitioned into those for which voters in X agree and the rest of candidates. For the “consensus candidates” we fix our decision to the decision induced by votes in X (additionally correcting the number of selected candidates in the “consensus” set). Next, we consider the optimization problem of finding a proper

subset of the remaining candidates to join the committee. The key insight is that there exists a small enough subset X such that the induced decision for the “consensus candidates” will not be a big mistake.

1.4 Organization of the paper

First, in Section 2 we formalize the information we may extract from subset of votes, and introduce a measure of inaccuracy of such a subset. Next, in Section 3 we prove the existence of a small subset of votes with stable inaccuracy. In Section 4 we show that the optimization problem of deciding the part of the committee not induced by the subset of votes can be approximated with only a small additional loss in the objective function. Finally, in Section 5 we give an algorithm considering all subsets of a fixed size and show that, in the iteration when the algorithm happens to consider a subset with stable inaccuracy, it will produce a $(1 + \epsilon)$ -approximate solution to MAV .

2 Extracting information from subsets

We consider subsets of votes and analyze the information they carry. We measure the inaccuracy of this information with respect to the set of all votes. We show that there exists a small subset with stable inaccuracy, i.e., the drop of inaccuracy after including one more vote is small.

Let us define an inaccuracy function $ina : 2^S \mapsto \mathbb{R}_{\geq 0}$ that measures the inaccuracy if we will consider subset $Y \subseteq S$ instead of S . The smaller the $ina(Y)$ is the better the common parts of strings in Y represent s_{OPT} .

Definition 2. For all $Y \subseteq S, Y \neq \emptyset$ we define functions $t(Y) \in \{0, 1\}^m$ and $ina(Y) \in \mathbb{R}_{\geq 0}$ as follows:

$$(t(Y))[j] = \begin{cases} 0 & \text{if } \forall_{y \in Y} y[j] = 0 \\ 1 & \text{if } \forall_{y \in Y} y[j] = 1 \\ s_{OPT}[j] & \text{otherwise,} \end{cases}$$

$$ina(Y) = d(t(Y), s_{OPT}).$$

Intuitively $t(Y)$ is the optimal solution s_{OPT} changed at positions where all strings from Y agree. Also we define the pattern of a subset of votes.

Definition 3. For all $Y \subseteq S, Y \neq \emptyset$ we define pattern $p(Y) \in \{0, 1, *\}^m$ as:

$$(p(Y))[j] = \begin{cases} 0 & \text{if } \forall_{y \in Y} y[j] = 0 \\ 1 & \text{if } \forall_{y \in Y} y[j] = 1 \\ * & \text{otherwise.} \end{cases}$$

It represents positions that all strings in Y agree. “*” encodes a mismatch.

Note that (from Definitions 2 and 3) $t(Y)$ is an optimal solution s_{OPT} over-written by a pattern p_r on no-star positions:

$$(t(Y))[j] = \begin{cases} s_{OPT}[j] & \text{if } (p(Y))[j] = * \\ (p(Y))[j] & \text{otherwise.} \end{cases}$$

The inaccuracy function has the following properties:

Lemma 1. $\forall_{s_{i_1} \in S}$, for all sequences $\{s_{i_1}\} = Y_1 \subseteq Y_2 \subseteq \dots \subseteq Y_n = S$ we have

$$OPT \geq ina(Y_1) \geq ina(Y_2) \geq \dots \geq ina(Y_n) = 0$$

Proof. It is easy to see that

$$\begin{aligned} ina(Y_1) &\stackrel{\text{def.}}{=} d(t(Y_1), s_{OPT}) = d(t(\{s_{i_1}\}), s_{OPT}) = d(s_{i_1}, s_{OPT}) \leq OPT, \\ ina(Y_n) &= ina(S) = d(s_{OPT}, s_{OPT}) = 0. \end{aligned}$$

Still we need to prove $ina(Y_i) \geq ina(Y_{i+1})$. Pattern $p(Y_{i+1})$ is built on strings from $Y_i \subseteq Y_{i+1}$ and strings from $Y_{i+1} \setminus Y_i$. So $p(Y_{i+1})$ has at least as many * as $p(Y_i)$ has. Therefore $t(Y_{i+1})$ has at least as many positions as $t(Y_i)$ has that agree with optimal solution s_{OPT} , so $d(t(Y_i), s_{OPT}) \geq d(t(Y_{i+1}), s_{OPT})$. Using definition of the inaccuracy function (Definition 2) we prove the lemma. ■

Intuitively $ina(Y) - ina(Y \cup \{y\})$ is the decrease of the inaccuracy from adding element y to set Y . We will show that, when adding one more element y to sets Y, Z such that $Y \subseteq Z$, the inaccuracy decrease more in a case of adding y to the smaller set Y than adding y to the bigger set Z .

Lemma 2. If we artificially extend the $ina(\cdot)$ function for the empty set: $ina(\emptyset) = 2 \cdot OPT$, then function $ina(\cdot)$ is supermodular¹, i.e.,

$$\forall_{Y \subseteq Z \subseteq S} \quad \forall_{s \in S} \quad ina(Z) - ina(Z \cup \{s\}) \leq ina(Y) - ina(Y \cup \{s\}) \quad (1)$$

Proof. Let fix Y, Z and s such that $Y \subseteq Z \subseteq S$ and $s \in S$.

Case 1: $Z = \emptyset$:

Then also $Y = \emptyset$, and inequality (1) holds obviously.

Case 2: $Z \neq \emptyset, Y = \emptyset$:

We have:

$$\begin{aligned} ina(Z) - ina(Z \cup \{s\}) &\leq OPT = \\ &= 2 \cdot OPT - OPT \leq ina(\emptyset) - ina(\{s\}) = ina(Y) - ina(Y \cup \{s\}), \end{aligned} \quad (2)$$

because we use respectively: Lemma 1 and the fact that Z has at least one element; definition of $ina(\cdot)$ for empty set and upperbound for $ina(\cdot)$ function; assumption that $Y = \emptyset$.

¹ according to [11], $f : 2^S \mapsto \mathbb{R}$ is supermodular iff

$$\begin{aligned} \forall_{Y, Z \subseteq S} \quad f(Y) + f(Z) &\leq f(Y \cup Z) + f(Y \cap Z) \text{ which is equivalent with} \\ \forall_{Y \subseteq Z \subseteq S} \quad \forall_{s \in S} \quad f(Z) - f(Z \cup \{s\}) &\leq f(Y) - f(Y \cup \{s\}). \end{aligned}$$

Case 3: $Z \neq \emptyset, Y \neq \emptyset$:

From definition of $ina(\cdot)$ we have:

$$ina(Z) - ina(Z \cup \{s\}) = d(t(Z), s_{OPT}) - d(t(Z \cup \{s\}), s_{OPT}) =$$

counting a difference by considering two cases for value of s_{OPT} we obtain

$$= \left| \{j : s_{OPT}[j] = 1 \wedge t(Z \cup \{s\})[j] = 1 \wedge t(Z)[j] = 0\} \right| + \\ + \left| \{j : s_{OPT}[j] = 0 \wedge t(Z \cup \{s\})[j] = 0 \wedge t(Z)[j] = 1\} \right| =$$

using definition of function $t(\cdot)$:

$$= \left| \{j : s_{OPT}[j] = 1 \wedge s[j] = 1 \wedge \forall_{z \in Z} z[j] = 0\} \right| + \\ + \left| \{j : s_{OPT}[j] = 0 \wedge s[j] = 0 \wedge \forall_{z \in Z} z[j] = 1\} \right| \leq$$

taking an universal quantifier over a smaller subset we obtain:

$$\leq \left| \{j : s_{OPT}[j] = 1 \wedge s[j] = 1 \wedge \forall_{y \in Y} y[j] = 0\} \right| + \\ + \left| \{j : s_{OPT}[j] = 0 \wedge s[j] = 0 \wedge \forall_{y \in Y} y[j] = 1\} \right| =$$

reversing all previous transformations finally we obtain:

$$= ina(Y) - ina(Y \cup \{s\}).$$

■

3 Existence of a stable subset

Lemma 3. For any fixed $R \in \mathbb{N}_{\geq 1}$ there exists a subset $X \subseteq S, |X| = R$ such that

$$\forall_{s \in S \setminus X} \quad ina(X) - ina(X \cup \{s\}) \leq \frac{OPT}{R}. \quad (3)$$

We say such X is $\frac{OPT}{R}$ -stable.

It means that there exists such a subset of votes X that adding one more vote into X the inaccuracy decreases by at most $\frac{OPT}{R}$.

Proof. First, we construct S_r satisfying (3) with at most R elements.

Let us construct a sequence of subsets $S_1 \subset S_2 \subset \dots \subset S_n = S, |S_i| = i$. We take $S_1 = \{s_{i_1}\}$, where s_{i_1} is any element of S and for $r \in \{2, 3, \dots, n\}$ we take $S_r = S_{r-1} \cup \{s_{i_r}\}$ where s_{i_r} is such a vote that after adding it the inaccuracy function decreases the most, i.e.,

$$s_{i_r} = \arg \max_{s \in S \setminus S_{r-1}} (ina(S_{r-1}) - ina(S_{r-1} \cup \{s\})). \quad (4)$$

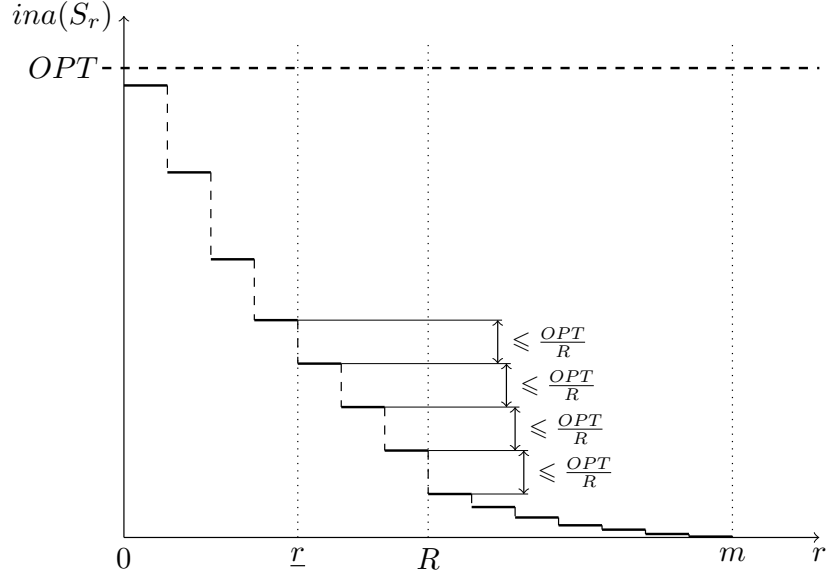


Fig. 1. The $ina(\cdot)$ function for the sequence of subsets $S_1 \subset S_2 \subset \dots \subset S_n = S$.

We have

$$\begin{aligned} \min_{r \in \{1, 2, \dots, R\}} ina(S_r) - ina(S_{r+1}) &\leq \frac{1}{R} \left(\sum_{r=1}^R ina(S_r) - ina(S_{r+1}) \right) = \\ &= \frac{1}{R} (ina(S_1) - ina(S_{R+1})) \leq \frac{OPT}{R}, \end{aligned} \quad (5)$$

because (from Lemma 1) we know that $ina(S_1) \leq OPT$ and $ina(S_{R+1}) \geq 0$. Let \underline{r} be a minimizer for the left-hand side of (5), then (by the choice of $s_{i_{\underline{r}}}$ in (4)) we have:

$$\max_{s \in S \setminus S_{\underline{r}}} (ina(S_{\underline{r}}) - ina(S_{\underline{r}} \cup \{s\})) \leq \frac{OPT}{R}, \quad (6)$$

thus $S_{\underline{r}}$ satisfies (3), see Figure 1. If $S_{\underline{r}}$ has less elements than R we can extend $S_{\underline{r}}$ to an R -elements subset X by adding any elements of S . It follows from the supermodularity of $ina(\cdot)$. From Lemma 2 we have:

$$\forall_{s \in S \setminus S_{\underline{r}}} ina(X) - ina(X \cup \{s\}) \leq ina(S_{\underline{r}}) - ina(S_{\underline{r}} \cup \{s\}),$$

and hence also:

$$\max_{s \in S \setminus S_{\underline{r}}} (ina(X) - ina(X \cup \{s\})) \leq \max_{s \in S \setminus S_{\underline{r}}} (ina(S_{\underline{r}}) - ina(S_{\underline{r}} \cup \{s\})). \quad (7)$$

Finally, taking (6) and (7) we obtain:

$$\max_{s \in S \setminus X} (ina(X) - ina(X \cup \{s\})) \leq \frac{OPT}{R}.$$

■

Of course we cannot construct such a subset efficiently if we do not know s_{OPT} . How to find a proper subset X ? For constructing our PTAS we will fix $R \in \mathbb{N}_{\geq 1}$ and consider all subsets $Y \subseteq S$ with cardinality R . There is less than $n^R \in Poly(n)$ such subsets. For clarity, we will use $Y \subseteq S$ in arguments valid for all subsets considered by the algorithm, and $X \subseteq S$ for a $\frac{OPT}{R}$ -stable subset of votes.

For a fixed $Y \subseteq S, Y \neq \emptyset$, WLOG we reorder candidates in such a way that $p(Y)$ is a lexicographically smallest permutation:

$$p(Y) = ** \dots * 00 \dots 011 \dots 1.$$

The first part (from the left) is called “star positions” or “star part”. The remaining part is called “no-star part”. We define $p^{(*)}(Y)$ as the number of $*$ in $p(Y)$ and we denote it β :

$$\beta = p^{(*)}(Y) = \left| \{j : (p(Y))[j] = *\} \right|.$$

In our PTAS we essentially fix the “no-star part” of the answer to the pattern $p(Y)$ and optimize over the choices for the “star part” of the outcome. If the number of stars or number of 1’s on star positions of s_{OPT} is small enough, then there is only $Poly(m, n)$ possible solutions and we can consider all of them. Let us analyze the size of the “star part”.

Lemma 4. *For all $Y \subseteq S$ we have*

$$\beta = p^{(*)}(Y) \leq |Y| \cdot OPT$$

Proof. Consider an arbitrary $Y = \{y_1, y_2, \dots, y_{|Y|}\}$. We can construct Y in the following 3 phases:

1. $Y := \{s_{OPT}\}$
2. for $i = 1$ to $|Y|$ do
 $Y := Y \cup \{y_i\}$
3. $Y := Y \setminus \{s_{OPT}\}$

After that we obtain set Y . Let us calculate how many stars $p(Y)$ has. In Phase 1 there are no stars. In each step in Phase 2 we add at most OPT stars, because $\forall_{i \in \{1, 2, \dots, |Y|\}} d(y_i, s_{OPT}) \leq OPT$. In Phase 3 we can at most decrease the number of stars. So $\beta \leq |Y| \cdot OPT$. ■

Note that for X from Lemma 3 we have

$$p^{(*)}(X) \leq |X| \cdot OPT = R \cdot OPT. \quad (8)$$

Let us now introduce some more notation. Assuming $Y \subseteq S$ and hence also $\beta = p^{(*)}(Y)$ are fixed, we will use the following notation to denote the “star part” and the “no-star part” of a string $x \in \{0, 1\}^m$:

$$x' = x[1] \cdot x[2] \cdot \dots \cdot x[\beta],$$

$$x'' = x[\beta + 1] \cdot x[\beta + 2] \cdot \dots \cdot x[m],$$

where “ \cdot ” is a concatenation of strings (letters). So we divide x into two parts: $x = x' \cdot x''$.

Let us now define a k -completion of $x \in \{0, 1\}^m$ (definition from [7]) to be a $y \in \{0, 1\}^m$ such that $y^{(1)} = k$ and $d(y, x)$ is the minimum possible Hamming distance between x and any vector with k of 1’s. To obtain a k -completion we only add or only delete a proper number of 1’s. To be more specific in this paper we assume the k -completion is always obtained by changing bits at positions with the smallest possible index².

In the following lemma we will show that for the pattern from a stable subset X we can change the number of 1’s in the “no-star part” to the properly guessed number of 1’s losing only twice the stability constant.

Lemma 5. *If $X \subseteq S$ is $(\epsilon_1 \cdot OPT)$ -stable, z'' is a k'' -completion of $(p(X))''$, where $k'' = (s''_{OPT})^{(1)}$, then*

$$\forall_{i \in \{1, 2, \dots, n\}} d(s'_{OPT} \cdot z'', s_i) \leq (1 + 2\epsilon_1) \cdot OPT \quad (9)$$

Proof. WLOG there is insufficient number of 1’s in no-star part of pattern $p(X)$, i.e., $k'' \geq ((p(X))'')^{(1)}$. The other case is symmetric.

Let us fix $s_i \in S$ and consider all combinations of values in strings $(p(X))''$, z'' , s''_i , s''_{OPT} at the same position j . $\alpha_a \in \mathbb{N}$, for $a \in \{1, 2, \dots, 12\}$, counts the number of positions j with combination a , see Table 1.

We have:

$$d(z'', s''_i) = |\{j : z''[j] \neq s''_i[j]\}| =$$

we consider two cases for value of s_{OPT} at position j :

$$= |\{j : z''[j] \neq s''_i[j] \wedge (z''[j] = s_{OPT} \vee z''[j] \neq s_{OPT})\}| =$$

we divide it into two components:

$$\begin{aligned} &= |\{j : s_{OPT} = z''[j] \neq s''_i[j] \quad \quad \quad \}| + \\ &+ |\{j : \quad \quad \quad z''[j] \neq s''_i[j] = s_{OPT}\}| = \end{aligned}$$

² Any other deterministic rule would work for us just as well.

| index of a combination | combinations | | | | | | | | | | | |
|-----------------------------|--------------|------------|------------|------------|------------|------------|------------|------------|------------|---------------|---------------|---------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $(p(X))''[j]$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| $z''[j]$ | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $s_i''[j]$ | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| $s_{OPT}''[j]$ | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| number of occurrences | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | α_7 | α_8 | α_9 | α_{10} | α_{11} | α_{12} |
| $d(z''[j], s_i''[j])$ | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| $d(s_{OPT}''[j], s_i''[j])$ | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |

Table 1. Combinations of values in strings $(p(X))''$, z'' , s_i'' , s_{OPT}'' . There is only 12 combinations (no $2^4 = 16$), because by the assumption $k'' \geq ((p(X))'')^{(1)}$ we never change from 1 in $((p(X))'')^{(1)}$ to 0 in z'' .

we use case counts from Table 1 to count positions in both components:

$$= \underbrace{(\alpha_2 + \alpha_7 + \alpha_{11})}_{\text{first component}} + \underbrace{(\alpha_3 + \alpha_6 + \alpha_{10} - \alpha_3 - \alpha_6 - \alpha_{10})}_{=0} + \underbrace{(\alpha_4 + \alpha_5 + \alpha_9)}_{\text{second component}} =$$

and we use the definition of the Hamming distance:

$$= (d(s_{OPT}'', s_i'') - \alpha_3 - \alpha_6 - \alpha_{10}) + (\alpha_4 + \alpha_5 + \alpha_9). \quad (10)$$

Since $(z'')^{(1)} = k'' = (s_{OPT}'')^{(1)}$,

$$\sum_{k=5}^{12} \alpha_k = \alpha_3 + \alpha_4 + \alpha_7 + \alpha_8 + \alpha_{11} + \alpha_{12}$$

$$\alpha_5 = \alpha_3 + \alpha_4 - \alpha_6 - \alpha_9 - \alpha_{10}. \quad (11)$$

Also

$$\alpha_4 + \alpha_8 + \alpha_9 \leq \epsilon_1 \cdot OPT, \quad (12)$$

because X is $\epsilon_1 \cdot OPT$ -stable. Now we are ready to prove equation (9).

$$\begin{aligned} & d(s'_{OPT} \cdot z'', s_i) \stackrel{\text{def.}}{=} d(s'_{OPT}, s'_i) + d(z'', s'_i) \stackrel{(10)}{=} \\ & \stackrel{(10)}{=} d(s'_{OPT}, s'_i) + d(s''_{OPT}, s''_i) - \alpha_3 - \alpha_6 - \alpha_{10} + \alpha_4 + \alpha_5 + \alpha_9 \stackrel{(11)}{=} \\ & \stackrel{(11)}{=} \underbrace{d(s_{OPT}, s_i)}_{\leq OPT} + 2 \underbrace{(\alpha_4 - \alpha_6 - \alpha_{10})}_{\stackrel{(12)}{\leq \epsilon_1 \cdot OPT}} \stackrel{(12)}{\leq} (1 + 2\epsilon_1) \cdot OPT. \end{aligned}$$

■

4 An auxiliary optimization problem

In this section we will consider the optimization problem obtained after guessing the number of 1's in the two parts and fixing the “no-star part” of the outcome. It has variables for all the positions of the “star part” and constraints for all the original votes $s_i \in S$.

Let us define the optimization problem in terms of the integer program $IP_{(13)-(17)}(Y, k')$ by (13)-(17):

$$\min q \quad (13)$$

$$(s')^{(1)} = k' \quad (14)$$

$$\forall_{i \in \{1, 2, \dots, n\}} \quad d(s', s'_i) \leq q - d(s''_{ALG}, s''_i) \quad (15)$$

$$q \geq 0 \quad (16)$$

$$\forall_{j \in \{1, 2, \dots, \beta\}} \quad s'[j] \in \{0, 1\} \quad (17)$$

where $Y \subseteq S$, $k = k' + k''$, and s''_{ALG} is the k'' -completion of $(p(Y))''$. Recall that $\beta = p^{(*)}(Y)$ and $(p(Y))''$ is the “no-star part” of the pattern $p(Y)$.

In the LP relaxation (17) is replaced with:

$$\forall_{j \in \{1, 2, \dots, \beta\}} \quad s'[j] \in [0, 1] \quad (18)$$

Constraints (13)-(16),(18) are linear because

$$(s')^{(1)} = \sum_{j=1}^{\beta} s'[j],$$

$$d(s', s'_i) = \sum_{j=1}^{\beta} \left(\chi(s'_i[j] = 0) \cdot s'[j] + \chi(s'_i[j] = 1) \cdot (1 - s'[j]) \right)$$

are linear functions of $s'[j]$, where $j \in \{1, 2, \dots, \beta\}$.

Lemma 6. $\forall_{R \in \mathbb{N}_{\geq 1}, Y \subseteq S, |Y| \leq R, k' \in \mathbb{N}, \epsilon_2 > 0}$ we can find $(1 + 2\epsilon_2)$ -approximation solution for $IP_{(13)-(17)}(Y, k')$ by solving the LP and considering at most

$$(3n)^{\frac{3R \ln(2)}{(\epsilon_2)^2}} + m^{\frac{3R^2 \ln(6)}{(\epsilon_2)^2}} \quad \text{cases.}$$

Proof. Let us fix constants $\epsilon_2 \in (0, \frac{1}{2})$ (for $\epsilon_2 \geq \frac{1}{2}$ we could use 2-approximation from [5]). Consider three cases:

Case 1: $\beta \leq \frac{3R \ln(3n)}{(\epsilon_2)^2}$

There is 2^β possibilities for s' .

$$2^\beta \leq 2^{\frac{3R \ln(3n)}{(\epsilon_2)^2}} = e^{\ln(3n) \frac{3R \ln(2)}{(\epsilon_2)^2}} = (3n)^{\frac{3R \ln(2)}{(\epsilon_2)^2}} \in \text{Poly}(n),$$

because ϵ_2 and R are fixed constants. So we will check (in polynomial time) all possibilities for s' and we will find optimal solution for the integer program.

Case 2: $k' \leq \frac{3R^2 \ln(6)}{(\epsilon_2)^2}$

There is $Poly(m)$ possibilities for s' because we can upperbound the number of possibilities of setting 1's into β positions by:

$$\binom{\beta}{k'} \leq \beta^{k'} \leq \beta^{\frac{3R^2 \ln(6)}{(\epsilon_2)^2}} \leq m^{\frac{3R^2 \ln(6)}{(\epsilon_2)^2}} \in Poly(m),$$

because ϵ_2 and R are fixed constants.

Case 3: $\beta > \frac{3R \ln(3n)}{(\epsilon_2)^2} \wedge k' > \frac{3R^2 \ln(6)}{(\epsilon_2)^2}$

We denote an optimal solution of the $IP_{(13)-(17)}(Y, k')$ by $((s')^{IP}, q^{IP})$. Let us use LP relaxation and denote an optimal solution of the LP by $((s')^{LP}, q^{LP})$. Obviously we have $q^{LP} \leq q^{IP}$. We can solve the LP in polynomial time but we may obtain a fractional solution. We want to round it independently. We will use a randomized rounding defined by distributions on each position $j \in \{1, 2, \dots, \beta\}$:

$$P(s'[j] = 1) = (s')^{LP}[j], \quad P(s'[j] = 0) = 1 - (s')^{LP}[j]. \quad (19)$$

We can estimate the expected value of a distance to such a random solution s' :

$$\begin{aligned} \forall_{i \in \{1, 2, \dots, n\}} \quad \mathbb{E}[d(s', s'_i)] &\stackrel{\text{def.}}{=} \mathbb{E} \left[\sum_{j=1}^{\beta} |s'[j] - s'_i[j]| \right] = \\ &= \mathbb{E} \left[\sum_{j=1}^{\beta} \left(\chi(s'_i[j] = 0) \cdot s'[j] + \chi(s'_i[j] = 1) \cdot (1 - s'[j]) \right) \right] \stackrel{\text{lin. of } \mathbb{E}}{=} \\ &\stackrel{\text{lin. of } \mathbb{E}}{=} \mathbb{E} \sum_{j=1}^{\beta} \left(\chi(s'_i[j] = 0) \cdot \mathbb{E}[s'[j]] + \chi(s'_i[j] = 1) \cdot \mathbb{E}[1 - s'[j]] \right) \stackrel{(19)}{=} \\ &\stackrel{(19)}{=} \sum_{j=1}^{\beta} \left(\chi(s'_i[j] = 0) \cdot (s')^{LP}[j] + \chi(s'_i[j] = 1) \cdot (1 - (s')^{LP}[j]) \right) \stackrel{\text{def.}}{=} \\ &\stackrel{\text{def.}}{=} d((s')^{LP}, s'_i) \stackrel{(15)}{\leq} q^{LP} - d(s''_{ALG}, s'_i). \end{aligned} \quad (20)$$

$d(s', s'_i)$ is a sum of β independent 0-1 variables. For $\epsilon' \in (0, 1)$ using Chernoff's bound [9] we have:

$$P\left(d(s', s'_i) \geq (1 + \epsilon') \cdot \mathbb{E}[d(s', s'_i)]\right) \leq \exp\left(-\frac{1}{3}(\epsilon')^2 \cdot \mathbb{E}[d(s', s'_i)]\right).$$

If we take $\epsilon' = \frac{\epsilon_2 \cdot q^{IP}}{\mathbb{E}[d(s', s'_i)]}$ then we obtain:

$$\exp\left(-\frac{1}{3} \cdot \frac{(\epsilon_2)^2 \cdot (q^{IP})^2}{\mathbb{E}[d(s', s'_i)]}\right) \geq P\left(d(s', s'_i) \geq \mathbb{E}[d(s', s'_i)] + \epsilon_2 \cdot q^{IP}\right) \stackrel{(20)}{\geq}$$

$$\stackrel{(20)}{\geq} P\left(d(s', s'_i) \geq q^{LP} - d(s''_{ALG}, s''_i) + \epsilon_2 \cdot q^{IP}\right). \quad (21)$$

We want to know an upperbound for the probability that we make an error greater than $\epsilon_2 \cdot q^{IP}$ for at least one vote:

$$\begin{aligned} & P\left(\exists_{i \in \{1, 2, \dots, n\}} : d(s', s'_i) \geq q^{LP} - d(s''_{ALG}, s''_i) + \epsilon_2 \cdot q^{IP}\right) \stackrel{(21)}{\leq} \\ & \stackrel{(21)}{\leq} n \cdot \exp\left(-\frac{1}{3} \cdot \frac{(\epsilon_2)^2 \cdot (q^{IP})^2}{\mathbb{E}[d(s', s'_i)]}\right) \leq n \cdot \exp\left(-\frac{1}{3}(\epsilon_2)^2 \cdot q^{IP}\right), \end{aligned} \quad (22)$$

where the last inequality is because of:

$$\mathbb{E}[d(s', s'_i)] \stackrel{(20)}{\leq} q^{LP} - d(s''_{ALG}, s''_i) \leq q^{IP}.$$

We want to further upperbound the probability in (22). From the assumption about β and from Lemma 4 we have:

$$\begin{aligned} \frac{3R \ln(3n)}{(\epsilon_2)^2} < \beta \stackrel{\text{Lem.4}}{\leq} |Y| \cdot OPT \leq R \cdot OPT \leq R \cdot q^{IP}, \quad \text{equivalently} \\ \frac{1}{3} > n \cdot \exp\left(-\frac{1}{3}(\epsilon_2)^2 \cdot q^{IP}\right). \end{aligned} \quad (23)$$

So, finally we have:

$$P\left(\exists_{i \in \{1, 2, \dots, n\}} : d(s', s'_i) \geq q^{LP} - d(s''_{ALG}, s''_i) + \epsilon_2 \cdot q^{IP}\right) \stackrel{(22), (23)}{<} \frac{1}{3}. \quad (24)$$

So with probability at least $\frac{2}{3}$ we obtain:

$$\begin{aligned} \forall_{i \in \{1, 2, \dots, n\}} \quad d(s' \cdot s''_{ALG}, s_i) &= d(s', s'_i) + d(s''_{ALG}, s''_i) \stackrel{(24)}{<} \\ &\stackrel{(24)}{<} q^{LP} - d(s''_{ALG}, s''_i) + \epsilon_2 \cdot q^{IP} + d(s''_{ALG}, s''_i) \leq (1 + \epsilon_2) \cdot q^{IP}. \end{aligned} \quad (25)$$

We can also obtain a wrong number of 1's. The solution s'_{ALG} for that is to take the k' -completion of s' . We will show that the additional error for such operation is not so big. Expected number of 1's in s' is equal k' :

$$\mathbb{E}[(s')^{(1)}] \stackrel{\text{def.}}{=} \mathbb{E}\left[\sum_{j=1}^{\beta} s'[j]\right] \stackrel{\text{lin. of } \mathbb{E}}{=} \mathbb{E}\sum_{j=1}^{\beta} (s')^{LP}[j] \stackrel{\text{def.}}{=} ((s')^{LP})^{(1)} \stackrel{(14)}{=} k'.$$

We want to know how much we lose taking the k' -completion. Similar as before, $(s')^{(1)} = \sum_{j=1}^{\beta} s'[j]$ is a sum of β independent 0-1 variables. For $\epsilon'' \in (0, 1)$ using Chernoff's bound [9] we have:

$$P\left((s')^{(1)} \geq (1 + \epsilon'') \cdot k'\right) \leq \exp\left(-\frac{1}{3}(\epsilon'')^2 \cdot k'\right),$$

$$P\left((s')^{(1)} \leq (1 - \epsilon'') \cdot k'\right) \leq \exp\left(-\frac{1}{2}(\epsilon'')^2 \cdot k'\right).$$

Taking both inequalities together, $\epsilon'' = \frac{\epsilon_2}{R}$ and using assumption $k' > \frac{3R^2 \ln(6)}{(\epsilon_2)^2}$ we have:

$$\begin{aligned} P\left(\left|(s')^{(1)} - k'\right| \geq \epsilon'' \cdot k'\right) &\leq 2 \cdot \exp\left(-\frac{1}{3}(\epsilon'')^2 \cdot k'\right) \leq \\ &\leq 2 \cdot \exp\left(-\frac{1}{3} \frac{(\epsilon_2)^2}{R^2} \cdot k'\right) < \frac{1}{3}. \end{aligned}$$

So with probability at least $\frac{2}{3}$ the error from taking the k' -completion is not greater than $\epsilon'' \cdot k' = \frac{\epsilon_2}{R} \cdot k' \leq \frac{\epsilon_2}{R} \cdot \beta \stackrel{\text{Lem.4}}{\leq} \frac{\epsilon_2}{R} \cdot |Y| \cdot OPT \leq \epsilon_2 \cdot OPT \leq \epsilon_2 \cdot q^{IP}$.

Combining the above with (25) we obtain a $(1 + 2\epsilon_2)$ -approximate solution with probability at least $\frac{1}{3}$. We may derandomize the algorithm analogously to how it was done in the PTAS for the Closest String problem [8]. For more on derandomization techniques see [10]. \blacksquare

5 Algorithm and its complexity analysis

Now we are ready to combine the ideas into a single algorithm.

Algorithm ALG(R)

Input: $S = \{s_1, s_2, \dots, s_n\} \in (\{0, 1\}^m)^n, 0 \leq k \leq m, R \in \mathbb{N}_{\geq 1}$

Output: $s_{ALG} \in \{0, 1\}^m$

- 1: **for** each R -element subset $Y = \{s_{i_1}, s_{i_2}, \dots, s_{i_R}\} \subseteq S$ **do**
 - 2: **for** each division k into two parts $k = k' + k''$ **do**
 - 3: $s'_{ALG} \leftarrow k''$ -completion of $(p(Y))'$
 (if not possible, then skip this inner iteration)
 - 4: $s'_{ALG} \leftarrow$ approximation solution of $IP_{(13)-(17)}(Y, k')$ using Lemma 6
 (if $LP_{(13)-(16),(18)}(Y, k')$ infeasible, then skip this inner iteration)
 - 5: evaluate $s'_{ALG} \cdot s''_{ALG}$ by computing $\max_{i \in \{1, 2, \dots, n\}} d(s_i, s'_{ALG} \cdot s''_{ALG})$
 - 6: **end for**
 - 7: **end for**
 - 8: $s_{ALG} \leftarrow$ the best solution from a loop in lines 1-7
-

It remains to argue that for a large enough parameter R the above algorithm will at some point consider a subset of votes X that leads to an accurate enough approximation of the Minimax objective function of our problem.

Theorem 1. $\forall \epsilon \in (0, 1)$ we may compute a $(1 + \epsilon)$ -approximate solution to Minimax Approval Voting in $O(\text{Poly}(n, m))$ time.

Proof. Let $\epsilon_0 = \frac{\epsilon}{3} < \frac{1}{3}$.

By Lemma 3, there exists an $\frac{\epsilon_0 \cdot OPT}{2}$ -stable set of votes $X \subseteq S$ of cardinality $|X| = R = \lceil \frac{2}{\epsilon_0} \rceil$.

Consider algorithm ALG(R). In one iteration it will consider X and k', k'' such that $(s'_{OPT})^{(1)} = k'$. Recall that s''_{ALG} is the specific k'' -completion of $(p(X))''$. By Lemma 5 we have:

$$d(s'_{OPT} \cdot s''_{ALG}, s_i) \leq (1 + \epsilon_0) \cdot OPT,$$

hence $(s' = s'_{OPT}, q = (1 + \epsilon_0) \cdot OPT)$ is a feasible solution to $IP_{(13-17)}(X, k')$ and the optimal value of $IP_{(13-17)}(X, k')$ is at most $(1 + \epsilon_0) \cdot OPT$.

By Lemma 6 with $\epsilon_2 = \frac{\epsilon_0}{2}$ we find a $(1 + \epsilon_0)$ -approximate solution (s'_{ALG}, q_{ALG}) to $IP_{(13-17)}(X, k')$. So we have:

$$q_{ALG} \leq (1 + \epsilon_0) \cdot (1 + \epsilon_0) \cdot OPT \stackrel{\epsilon_0 < 1}{\leq} (1 + 3\epsilon_0) \cdot OPT = (1 + \epsilon) \cdot OPT.$$

It remains to observe, that $s_{ALG} = s'_{ALG} \cdot s''_{ALG}$ is a solution to MAV of cost $q_{ALG} \leq (1 + \epsilon) \cdot OPT$.

The algorithm examined $O(n^R) = O(n^{\lceil \frac{6}{\epsilon} \rceil}) \in O(\text{Poly}(n))$ subsets Y , $O(m)$ choices of k' and each time considered

$$O\left((3n)^{\frac{108 \cdot \lceil \frac{6}{\epsilon} \rceil \cdot \ln(2)}{\epsilon^2}} + m^{\frac{108 \cdot \lceil \frac{6}{\epsilon} \rceil^2 \cdot \ln(6)}{\epsilon^2}}\right) \in O(\text{Poly}(n, m)) \text{ cases.}$$

■

6 Concluding remarks

We showed the existence of a PTAS for Minimax Approval Voting by considering all subsets of a fixed size R . If not the discovered supermodularity for the inaccuracy function $ina(\cdot)$, we would simply consider all subsets of size at most R . Although the supermodularity was not essential for our result, it shows that larger subsets of votes are generally more stable (in the sense of definition in Lemma 3). It seems to suggest that an algorithm considering a smaller number of larger subsets of votes would potentially be more efficient in practice. Perhaps the most interesting open question is whether by randomly sampling a number of subsets of votes to examine, one could obtain a more practical FPRAS for the problem.

Another interesting direction is the optimization of the Minimax objective function subject to a restriction that the voting system must be incentive compatible. According to [5] the best possible approximation ratio in this setting is between $2 - \frac{2}{k+1}$ and $3 - \frac{2}{k+1}$, and a natural challenge is to narrow this gap.

Finally, we know the complexity of the two extreme objectives, i.e., Minimax and Minisum. The latter is easily optimized by selecting the k most often approved candidates. The optimization problem for intermediate objectives such as optimizing the sum of squares of the Hamming distances remains unexplored, and it would be interesting to learn which objective functions are more difficult to approximate than Minimax in the context of Approval Voting systems.

Acknowledgments

We want to thank Katarzyna Staniewicz for many helpful proofreading comments. Also we want to thank reviewers for their valuable suggestions. Krzysztof Sornat was supported by local grant 2139/M/II/14.

References

1. H. Aziz, S. Gaspers, J. Gudmundsson, S. Mackenzie, N. Mattei, T. Walsh. Computational Aspects of Multi-Winner Approval Voting. arXiv preprint arXiv:1407.3247v1, 2014.
2. S.J. Brams, P.C. Fishburn. Approval Voting. 2nd edition, Springer, 2007.
3. S.J. Brams, D.M. Kilgour, M.R. Sanver. A minimax procedure for electing committees. *Public Choice*, 132(3-4): pages 401-420, 2007.
4. J. Byrka, K. Sornat. PTAS for Minimax Approval Voting. arXiv preprint arXiv:1407.7216v2, 2014.
5. I. Caragiannis, D. Kalaitzis, E. Markakis. Approximation algorithms and mechanism design for minimax approval voting. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 737–742, 2010.
6. R. LeGrand. Analysis of the minimax procedure. Technical Report WUCSE-2004-67. Department of Computer Science and Engineering, Washington University, St. Louis, Missouri, 2004.
7. R. LeGrand, E. Markakis, A. Mehta. Some results on approximating the minimax solution in approval voting. *Proceedings of 6th AAMAS*, pages 1193–1195, 2007.
8. M. Li, B. Ma, L. Wang. On the closest string and substring problems. *Journal of the ACM*, vol. 49: pages 157–171, 2002.
9. R. Motvani, P. Raghavan. *Randomized Algorithms*. Cambridge University Press, Chapter 4.1, 1995.
10. P. Raghavan. Probabilistic construction of deterministic algorithms: Approximate packing integer programs. *Journal of Computer and System Sciences*, 37(2), pages 130-143, 1988.
11. A. Schrijver. *Combinatorial Optimization: polyhedra and efficiency*. Springer, page 766, 2003.