Multiple Referenda and Multiwinner Elections Using Hamming Distances: Complexity and Manipulability

Georgios Amanatidis
Athens University of Economics and Business
gamana@aueb.gr

Nathanaël Barrot
LAMSADE
Université Paris Dauphine
nathanael.barrot@lamsade.dauphine.fr

Jérôme Lang
LAMSADE
Université Paris Dauphine
lang@lamsade.dauphine.fr

Evangelos Markakis
Athens University of Economics and Business
markakis@gmail.com

Bernard Ries
LAMSADE
Université Paris Dauphine
bernard.ries@dauphine.fr

ABSTRACT

We study multiple referenda and committee elections, when the ballot of each voter is simply a set of approved binary issues (or candidates). Two well-known rules under this model are the commonly used candidate-wise majority, also called the minisum rule, as well as the minimax rule. In the former, the elected committee consists of the candidates approved by a majority of voters, whereas the latter picks a committee minimizing the maximum Hamming distance to all ballots.

As these rules are in some ways extreme points in the whole spectrum of solutions, we consider a general family of rules, using the Ordered Weighted Averaging (OWA) operators. Each rule is parameterized by a weight vector, showing the importance of the i-th highest Hamming distance of the outcome to the voters. The objective then is to minimize the weighted sum of the (ordered) distances. We study mostly computational, but also manipulability properties for this family. We first exhibit that for many rules, it is NP-hard to find a winning committee. We then proceed to identify cases where the problem is either efficiently solvable, or approximable with a small approximation factor. Finally, we investigate the issue of manipulating such rules and provide conditions that make this possible.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Non-numerical Algorithms and Problems; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Algorithms; Theory

Keywords

Approval voting; Minisum; Minimax; Ordered weighted averaging operators

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

1. INTRODUCTION

Multiple referenda (making a collective decision over several binary propositions) and committee elections (electing several winners out of a set of candidates) are two similar problems, though they are generally studied apart. In both cases, the voters have to decide about a common truth value for each proposition (accept or reject), or for each candidate (elect or do not elect). Both cases may also present some constraints on the set of feasible combined decisions. For example, the number of persons elected in a committee may be subject to cardinality constraints (e.g., the committee should be of size exactly 6 or between 6 and 10).

Let us start with a detailed example. Five coauthors of a paper have to decide about whether they should use Dropbox (d or $\overline{d}$), whether they should have a physical meeting (m or $\overline{m}$), whether they should work more and prove more results before submission (w or $\overline{w}$) and whether they should tick the "possibly accept as short paper" box (a or $\bar{a}$). Here are their votes (where, for instance, the 1 in the cell (A_1, d) means that A_1 votes in favor of using Dropbox).

	d	m	w	a
A_1	1	1	0	0
A_2	0	1	0	0
A_3	1	0	0	1
A_4	1	1	1	0
A_5	0	0	1	1

A proposition-wise majority vote would give 1100 as the collective decision. Is this a fair decision? Arguably not: A_5 will complain that everything was decided against her will, which can be considered unfair for such a small set of voters. A more fair solution, proposed in the context of committee elections by [5], is *minimax approval voting*: the winning committee is the one minimizing the maximum Hamming distance to all votes (after possibly applying a tie-breaking rule). Applying this to our example would rule out 1100 (with a distance of 4 to A_5), and the winning decision would be one among the committees 1111, 1110, 1101, 0111, 1010, 1001, 0101, 1000, 0100 and 0010.

Minimax approval voting does make decisions that are more fair in the Rawlsian sense, assuming that the utility an agent draws from a decision decreases linearly with the Hamming distance from her optimal decision. However, there are some objections to this rule, which we review below.

The first objection is computational complexity: finding an optimal subset is NP-hard [10]. These bad news are tempered by the fact that there are efficient approximation algorithms (a PTAS in [21], and more recently, for the cardinality constrained case, a factor of 3 in [20], 2 in [7] and a PTAS in [6]).

The second and most important objection has to do with the extreme nature of the rule. The agent with the worst utility (the largest distance) may have a huge influence, even if everyone else agrees. Assume that instead of 5 voters we now have 21 voters: twenty unanimously vote 1100, and the twenty-first is A_5, who again votes 0011. Minimax approval voting will output one of 1010, 0110, 1001, 1010, 0000 or 1111, with everyone at Hamming distance 2 from the outcome; A_5, preventing the otherwise unanimous decision 1100 to be taken, can be seen as having too much influence. In a large population of voters, it is to be expected that there will be voters with opposite preferences. This perhaps means that minimax approval voting is tailored only for very small electorates, where this phenomenon is less likely to occur (at least not in such a drastic way). Related to this objection, it is well-known that minimax approval voting is not strategyproof [20]; we do not view this as an objection, since unless we impose strong domain restrictions, finding strategyproof ways of making collective decisions is hopeless. But what is specific to minimax approval voting is that the tremendous influence of the least happy voter may give to his potential manipulative votes a huge impact.

Our last objection has to do with the decisive power of the rule. As we have seen in our two examples, these decisions may be fair, but the set of winning committees (before applying a tie-breaking rule) can be very large, and indeed, a drawback of minimax approval voting applied to a small number of voters and propositions is its indecisiveness. We would like to have rules where it is more likely to have a unique outcome without the need for tie-breaking.

Contribution: The aim of this paper is to show that we can remedy the second objection without making things worse in terms of computational complexity. We believe the last objection is also less likely to occur under the proposed framework. To achieve this, we consider a family of voting rules which generalizes the minimax rule, as well as the standard commonly used rule, referred to as *minisum*, that outputs all candidates approved by a majority of voters. Our general setting makes use of *Ordered Weighted Averaging* operators (in short, OWA) [30]. Each such operator weighs appropriately the distances of a decision to the votes according to the *rank* they have if we order them from largest to smallest. Here we take the Hamming distance as our distance function. As an example, minimax approval voting would correspond to the weight vector $(1, 0, \ldots, 0)$, since we only care for the maximum Hamming distance and all other distances have weight 0. In minisum approval voting, all distances have equal weights, and this would correspond to $(1/n, \ldots, 1/n)$ for a population of n voters. Between minimax and minisum, we have a continuum of rules which are parameterized by the vector of weights, and that can be fine-tuned according to the application domain.

Given this framework, our study mainly focuses on the complexity of computing a winning committee, given a weight vector of an OWA operator. We first establish various NP-completeness results, showing that even for vectors that may slightly differ from minisum (for which a winning committee can be computed efficiently) the problem is NP-complete. We next identify some families of vectors where we can have polynomial time exact algorithms and then moving on, we design approximation algorithms based on Linear Programming for some of the NP-complete families. We also investigate the performance of the minisum algorithm with regard to its approximation quality for an OWA operator. We show that this achieves a constant approximation factor in most interesting cases, degrading smoothly for vectors that move away from minisum. Finally, we consider the issue of manipulating such elections. It is already known that the minimax rule is manipulable and we extend this result to other OWA operators as well.

Related Work: Two recent works use OWA operators in a voting context, namely [12] and [28]. The work of [12] generalizes positional scoring rules by weighing scores according to their rank in the ordered list of scores obtained by the candidate from the votes. The work of [28] generalizes Chamberlin and Courant's proportional representation rule, with the score of a committee being the sum of the individual scores it obtains from different voters. The individual score obtained by a committee from a vote is computed by weighing the scores of the members of the committee by their rank in the list of scores given to them by the voter. The latter work is the most closely related work to ours, since they also deal with multiwinner elections; where our models strongly depart is that in their model, the weights bear on the scores obtained by various candidates (or items) and not (like in our model and also in [12]) on the scores obtained for different voters.

Another related series of works is [15, 8], who study the properties of propositionwise majority (or *minisum*) under the assumption that agents have Hamming-induced preferences: for instance, under this assumption, they show that the outcome of propositionwise majority cannot be Pareto-dominated and that it belongs to the top cycle (*a fortiori*, propositionwise majority voting is Condorcet-consistent). Finally, the computational aspects of minimax approval voting have been studied in [20, 7, 6], and the computational aspects of other multiwinner voting rules based on approval ballots in [25, 24, 2].

Minimax approval voting has also been proposed independently (and earlier) in [14], under the name "egalitarian merging". Furthermore, it has been used (much more recently) in judgment aggregation [16]. We believe that although minimax approval voting was defined in the context of committee elections, it also makes sense (and perhaps even more so) in the context of multiple referenda. It can be further applicable in "budgeted social choice" where the goal is to choose a collective set of items subject to budget constraints [22], as it offers a more egalitarian way of making choices than existing methods.

Outline: In the rest of the paper we consider two versions of approval voting elections: one where there is a constraint on the size of the committee (or equivalently, on the number of propositions accepted), and one without any constraints. Cardinality constraints clearly make sense in committee elections, in budgeted social choice, and also in multiple referenda when accepted propositions imply a cost (such as the decision of building or not each of a set of common facilities); The outline of the paper is as follows. In Section 2 we formally define our general framework. Then we address computational issues for various choices of OWA operators. We show in Section 3 that NP-hardness holds in most cases. In Section 4 we give positive results; specifically, we identify cases where the problem is either efficiently solvable or approximable with a small approximation factor. Finally, in Section 5, we show that, unsurprisingly, our rules are manipulable in most interesting cases; in fact, we do not know of any strategyproof rule in our family, other than minisum. Finally, we give several research directions in Section 6.

2. DEFINITIONS AND NOTATION

In the setting we consider, each voter casts a ballot consisting of a subset of candidates or issues. The two main applications we have in mind are *approval-based committee elections* and *multiple referenda*. In approval-based committee elections, a voter's ballot expresses his approval for a subset of the candidates under consideration. In multiple referenda, it expresses the binary issues which the voter wishes to see adopted. Our setting applies to yet other domains too, such as the selection of a common set of objects. To avoid repetitions, we define all the relevant notions using the terminology of approval-based committee elections, but everything transfers to multiple referenda by replacing "candidates" by issues, and "committees" by "bundles of issues". A uniform view of multiple referenda and multiwinner elections (and a review of existing approaches), under the general umbrella of voting over combinatorial domains, can be found in [17].

We use n to denote the number of voters and m to denote the number of candidates. We denote the set of voters by N, and the set of candidates by A. An approval ballot simply specifies a subset of A, *i.e.*, the subset of candidates that a voter approves. A voting profile P is a tuple $P = (P_1, ..., P_n)$, where $P_i \subseteq A$ denotes the preference of voter i. Note that we can also represent the preferences of voters as binary vectors in $\{0, 1\}^m$, where the 1s indicate approvals. We will mostly stick to the set representation, but when convenient, we will also switch to the binary vector representation (as in Section 5). Under this notation, an election is specified by a tuple (N, A, P).

There are several ways of using approval voting for committee elections; see [13] for a review. Arguably the most commonly used method (referred to as *minisum*) consists of electing the k candidates approved most often or, if we have no cardinality constraint, the candidates that are approved by a majority of voters. Another interesting method that has attracted some attention is *minimax approval voting*. To describe this, we first define the Hamming distance between two ballots Q and T, as their symmetric difference, i.e., the total number of candidates in which they differ: $d_{\mathcal{H}}(Q, T) = |Q \setminus T| + |T \setminus Q| = |Q| + |T| - 2|Q \cap T|$. Minimax approval voting [5] selects a committee S that minimizes $\max_{i \in N} d_{\mathcal{H}}(S, P_i)$. Note that replacing max by sum in this quantity leads back to the standard multi-winner approval rule (hence the name "minisum" – see [5]).

We now introduce a family of voting rules that generalize the minisum and the minimax rules. For this, we use Ordered Weighted Averaging Operators (OWA) [30]. Each rule in this family is specified by a weight vector and selects the outcome that minimizes the weighted sum of Hamming distances, after ordering them in nonascending order. To be more precise, given a preference profile for n voters, $P = (P_1, ..., P_n)$, and a subset of candidates S, we let $\mathcal{H}(P, S)$ be the n-dimensional vector that contains the Hamming distances of the P_is from S in nonascending order. Let now $\mathbf{w} = \{\mathbf{w}_n\}_{n \in \mathbb{N}}$, be a collection of weight vectors such that for every number of voters $n \in \mathbb{N}$, $\mathbf{w}_n = (\mathbf{w}_n(1), \mathbf{w}_n(2), \ldots, \mathbf{w}_n(n)) \in [0, 1]^n$ and the coordinates of $\mathbf{w}_n$ sum up to 1. Here $\mathbf{w}_n(i)$ is the weight attached to the i-th largest Hamming distance of the selected outcome to the n voters. The voting rule then, which we refer to as $\mathbf{w}$-*Approval Voting*, or in short $\mathbf{w}$-AV, is as follows:

$\mathbf{w}$-AV: Given an election (N, A, P) with n voters, select a subset of candidates S, so as to minimize the dot product $\mathbf{w}_n \cdot \mathcal{H}(P, S)$.

In some settings, such as committee elections, the subset of candidates to be elected has to be of a certain size. In that case, we are interested in the following variant:

$(k, \mathbf{w})$-AV: Given an election (N, A, P) with n voters, and an integer k, select a subset of candidates S, with $|S| = k$, so as to minimize the dot product $\mathbf{w}_n \cdot \mathcal{H}(P, S)$.

For the decision version of the above problems, we are also given a parameter $\alpha \in \mathbb{Q}$, and we ask whether there exists a set of candidates S, such that $\mathbf{w}_n \cdot \mathcal{H}(P, S) \leq \alpha$.

Notice that $\mathbf{w}$-AV is a generalization of both the minisum and the minimax rules in approval voting (resp. k-minisum and k-minimax when the committee size is restricted). Indeed, let MINISUM (resp. k-MINISUM) denote the problem of finding the committee that minimizes the sum of the Hamming distances. Then, this is equivalent to $\mathbf{w}$-AV (resp. $(k, \mathbf{w})$-AV) for $\mathbf{w}_n = (1/n, 1/n, \ldots, 1/n)$. Similarly, let MINIMAX and k-MINIMAX denote the problems of minimizing the maximum Hamming distance. For $\mathbf{w}_n = (1, 0, \ldots, 0)$, these are equivalent to $\mathbf{w}$-AV and $(k, \mathbf{w})$-AV.

EXAMPLE 1. *Consider again our introductory example. Recall that minisum selects* 1100 *and that minimax selects a committee such that* $\max_{i=1 \ldots 5} d_{\mathcal{H}}(S, P_i) = 3$. *Consider the vector* $\mathbf{w} = (\frac{6}{16}, \frac{4}{16}, \frac{3}{16}, \frac{2}{16}, \frac{1}{16})$, *which can be seen as a hybrid of minisum and minimax.* $\mathbf{w}$*-AV winning committees are* 1000 *and* 1101, *with* $\mathcal{H}(P, 1000) = \mathcal{H}(P, 1101) = (3, 2, 2, 2, 1)$ *and* $\mathbf{w} \cdot \mathcal{H}(P, 1000) = \mathbf{w}_5 \cdot \mathcal{H}(P, 1101) = (\frac{6}{16} \cdot 3) + (\frac{4}{16} \cdot 2) + (\frac{3}{16} \cdot 2) + (\frac{2}{16} \cdot 2) + (\frac{1}{16} \cdot 1) = \frac{35}{16}$. *Now, if we add the constraint that the winning committee(s) should have cardinality 2, then the* $(2, \mathbf{w})$*-AV winning committee is* 1001, *with* $\mathbf{w} \cdot \mathcal{H}(P, 1001) = \frac{36}{16}$.

In most of the paper we focus on OWA vectors that are *nonincreasing*. This nonincreasingness condition is classical in the study of OWAs, because it corresponds to a fairness criterion: the higher a voter in the ordered list (across all voters) of Hamming distances between her preferred committee P_i and a possible output committee S, the more she counts for evaluating the quality of S. Among the family of nonincreasing OWA vectors, we find, at two extremities, $(1, 0, \ldots, 0)$ corresponding to MINIMAX , and $(1, 1, \ldots, 1)$ corresponding to MINISUM. Arguably, MINIMAX is the fairest, and MINISUM the less fair, in this family of rules corresponding to nonincreasing vectors, with a continuum inbetween.

3. HARDNESS RESULTS

We already know that the decision versions of $\mathbf{w}$-AV and $(k, \mathbf{w})$-AV are NP-complete, since they are generalizations of MINIMAX and k-MINIMAX, which correspond to $\mathbf{w}_n = (1, 0, \cdots, 0)$. The interesting and intriguing question is to understand for what choices of $\mathbf{w}$ we would still have NP-hardness. We know for example that for $\mathbf{w}_n = (1/n, 1/n, \cdots, 1/n)$, the problem is efficiently solvable. We investigate this question further in this section.

3.1 Hardness of w-AV

An interesting family of vectors is the family $\mathbf{f}^i$ defined by $\mathbf{f}^i_n = (\frac{1}{n-i}, \ldots, \frac{1}{n-i}, 0, \ldots, 0)$, where i is the number of 0's, for $i = 0, \ldots, n-1$; this family ranges from minisum approval voting (corresponding to $\mathbf{f}^0$) to minimax approval voting (corresponding to $\mathbf{f}^{n-1}$). It is a subfamily of nonincreasing (and thus, fair) vectors. In the sequel, i can sometimes be a function of the number of voters in the instance at hand.

We will prove that finding a winning committee for $\mathbf{f}^i$-AV is NP-hard, for a large range of values for i. This shows that NP-hardness remains as we slowly move away from MINIMAX and is still present even if we come relatively close to MINISUM. To prove it, we will study the decision version of $\mathbf{f}^i$-AV and show that it is NP-complete. Our hardness result holds even for the special case of balanced elections, where all the candidates are approved by exactly half of the voters.

We first start with MINIMAX, i.e., with $\mathbf{f}^{n-1}$-AV, which we know already that it is NP-hard [10]. We prove that MINIMAX remains NP-hard, even when restricted to balanced elections. The proof (which we omit due to space restrictions) is based on a reduction from a balanced variant of 3-SAT, which is known to be NP-complete [4].

THEOREM 1. *The decision version of* MINIMAX *is NP-complete, even when restricted to balanced elections.*

Before we continue, we remark that in general the parameter i in $\mathbf{f}^i$-AV can be a function of the number of voters, n, i.e., $i := i(n) : \mathbb{N} \to \mathbb{N}$. For the sake of readability we will write simply i instead of $i(n)$ in the statements of results. We can now show the following hardness result, using a reduction from MINIMAX on balanced elections.

THEOREM 2. *For any constant c, and any $1 \leq i \leq n^{1-1/c}$, the decision version for $\mathbf{f}^{n-i}$-AV is NP-complete, even when restricted to balanced elections.*

Finally, we also show that even when we come close to MINISUM, $\mathbf{f}^i$-AV is intractable.

THEOREM 3. *For any constant c, and any i, with $1 \leq i \leq n^{1-1/c}$, the decision version for $\mathbf{f}^i$-AV is NP-complete, even when restricted to balanced elections.*

PROOF. First, the problem is clearly in NP. The hardness proof is based on a reduction from the decision version of $\mathbf{f}^{n-i}$-AV to the decision version of $\mathbf{f}^i$-AV, for any $1 \leq i \leq n-1$. Consider a balanced election E for $\mathbf{f}^{n-i}$-AV, $E = (N, A, P)$, and a constant α. We construct a balanced election E' for $\mathbf{f}^i$-AV with $N' = N$, $A' = A$, $P' = \bar{P} = (\bar{P}_1, \ldots, \bar{P}_n)$ and $\alpha' = \frac{m(n-2i)}{2(n-i)} + \frac{i}{n-i}\,\alpha$; here $\bar{P}_i$ denotes $A \setminus P_i$. Let $n = |N| = |N'|$ and $m = |A| = |A'|$. It is easy to see that E' remains balanced since P' is the complementary profile of P, i.e. the profile composed of the complementary ballots of P. We claim that the election E is a yes instance for $\mathbf{f}^{n-i}$-AV if and only if E' is a yes instance for $\mathbf{f}^i$-AV. To prove it, we will study the $\mathbf{f}^{n-i}$-AV score of a committee C for E, which is $\mathbf{f}_n^{n-i} \cdot \mathcal{H}(P, C)$, in comparison to its $\mathbf{f}^i$-AV score for E', which is $\mathbf{f}_n^i \cdot \mathcal{H}(\bar{P}, C)$. Consider a committee $C \subseteq A$. Since E is a balanced election, each candidate (belonging to C or not) increases the minisum score of C by $\frac{n}{2}$. Thus, the minisum score of any committee C has the same value and is equal to $\frac{mn}{2}$. The minisum score can be expressed as follows:

$$\text{minisum}(C, P) = \sum_{k=1}^{n} \mathcal{H}(P, C)_k = \sum_{k=1}^{i} \mathcal{H}(P, C)_k + \sum_{k=i+1}^{n} \mathcal{H}(P, C)_k$$

where $\mathcal{H}(P, C)_k$ is the k-th largest Hamming distance of a voter from C. This is equivalent to:

$$\frac{mn}{2} = i \cdot \mathbf{f}_n^{n-i} \cdot \mathcal{H}(P, C) + \sum_{k=i+1}^{n} \mathcal{H}(P, C)_k. \qquad (1)$$

In addition, given an approval ballot P_k of a voter $k \in N$, we know that

$$d_{\mathcal{H}}(C, P_k) = m - d_{\mathcal{H}}(C, \bar{P}_k),$$

since $\bar{P}_k$ is the complement of preference P_k. This implies:

$$\sum_{k=i+1}^{n} \mathcal{H}(P, C)_k = \sum_{k=1}^{n-i} (m - \mathcal{H}(\bar{P}, C)_k),$$

because voters corresponding to the first coordinates of $\mathcal{H}(P, C)$ will correspond to the last coordinates of $\mathcal{H}(\bar{P}, C)$. Thus, we have:

$$\sum_{k=i+1}^{n} \mathcal{H}(P, C)_k = (n-i)m - (n-i) \cdot \mathbf{f}_n^i \cdot \mathcal{H}(\bar{P}, C), \qquad (2)$$

Then, from Equations (1) and (2), we obtain:

$$\frac{mn}{2} = i \cdot \mathbf{f}_n^{n-i} \cdot \mathcal{H}(P, C) + (n-i)m - (n-i) \cdot \mathbf{f}_n^i \cdot \mathcal{H}(\bar{P}, C)\ ,$$

which is equivalent to:

$$\mathbf{f}_n^i \cdot \mathcal{H}(\bar{P}, C) = \frac{m(n-2i)}{2(n-i)} + \frac{i}{n-i}\, \mathbf{f}_n^{n-i} \cdot \mathcal{H}(P, C)\,.$$

Hence, the $\mathbf{f}^{n-i}$-AV instance has a solution with score at most α if and only if the $\mathbf{f}^i$-AV has a solution with score at most α'. □

A related question is the complexity of outputting *all* winning committees for $\mathbf{f}^i$-AV. We remark that clearly, for minisum with an even number of voters, the number of such committees can be exponential in the number of candidates (consider an election where all the candidates are approved by half of the voters; then all the committees are winning committees). On the other hand, with an odd number of voters, there are no candidates approved by exactly half of the voters, hence there is exactly one winning committee. Interestingly, this observation does not extend to $\mathbf{f}^i$-AV for $i \geq 1$.

PROPOSITION 4. *For any $i \geq 1$, and under $\mathbf{f}^i$-AV, there exists a collection of elections such that the number of winning committees is exponential in the number of candidates, even when n is odd.*

We conclude this subsection with another interesting family of vectors. Namely, we denote by $\mathbf{m}^i$-AV the rule defined by $\mathbf{m}_n^i = (0, \ldots, 0, 1, 0 \ldots, 0)$, where the 1 is at position i, for $i = 1, \ldots, n$; this family corresponds to the median operators, and ranges from the minimax solution (corresponding to $\mathbf{m}^1$) to the minimin solution (corresponding to $\mathbf{m}^n$). The problem of finding a winning committee for $\mathbf{m}^n$-AV is polynomial, since it is sufficient to choose any one of the P_is to obtain a winning committee. But the decision version of $\mathbf{m}^i$-AV is NP-complete for a wide range of values:

THEOREM 5. *For any constant c, and any i, with $1 \leq i \leq n^{1-1/c}$, the decision version for $\mathbf{m}^i$-AV is NP-complete.*

Furthermore, we have a hardness result when i is equal to $n/2$.

THEOREM 6. *The decision version of $\mathbf{m}^{n/2}$-AV is NP-complete.*

3.2 Hardness of $(k, \mathbf{w})$-AV

We now move to the version where the size of the committee is restricted to be of a certain size. Note that an NP-hardness result for $\mathbf{w}$-AV only implies a Turing reduction for $(k, \mathbf{w})$-AV. We are not aware of any way in which hardness results for $\mathbf{w}$-AV can yield immediately NP-hardness (i.e., a Karp reduction) for $(k, \mathbf{w})$-AV. In fact, the hard constraint on the size of the committee requires a different approach for creating a Karp reduction. As a result, the reductions we present in this subsection are based on completely different ideas than in Section 3.1.

As we will show, the decision version of $(k, \mathbf{w})$-AV is NP-complete, for a wide range of families of vectors. We first introduce the following notion, for which our hardness results apply.

DEFINITION 1. *Let $\beta \geq 2$ be an integer. We say that the family of vectors $\mathbf{w}$ is β-restricted if for any $n \in \mathbb{N}$:*

i. $\mathbf{w}_n = \big(a_1(n), a_2(n), \ldots, a_n(n)\big) \in [0,1]^n$ *with* $\sum_{i=1}^n a_i(n) = 1$

ii. $a_1(n) > \sum_{i=\ell+1}^n a_i(n)$ *where* $\ell = \lceil n + 4 - 2n^{\frac{1}{\beta}} \rceil$

This is a family of vectors where the first coordinate, i.e., the weight on the maximum distance is relatively large; in particular, larger compared to the sum of weights given to the lower distances. For instance, for $\beta = 2$, the weight on the last $2\sqrt{n}$ coordinates should be small. β-restrictedness corresponds to a fairness criterion (even if it does not imply nonincreasingness, nor is implied by it). Also, it obviously generalizes the minimax objective, since $(1, 0, ..., 0)$ belongs to the family. By $(k, \mathbf{w})$-AV$_\beta$ we denote the decision version of $(k, \mathbf{w})$-AV on β-restricted families of vectors.

THEOREM 7. *For any integer $\beta \geq 2$ and any β-restricted family of vectors $\mathbf{w}$, $(k, \mathbf{w})$-AV$_\beta$ is NP-complete.*

PROOF. For the sake of readability we are going to drop the argument n in the vectors, e.g., we will write $\mathbf{w}_n = (a_1, a_2, \ldots, a_n)$ instead of $\mathbf{w}_n = \big(a_1(n), a_2(n), \ldots, a_n(n)\big)$.

Clearly, $(k, \mathbf{w})$-AV$_\beta$ is in NP. To prove hardness, we will provide a reduction from a certain variant of the Vertex Cover problem. In particular, let VC$_{\triangle\text{-free}}(k)$ be the problem of deciding whether there exists a vertex cover of size k in a simple planar triangle-free graph. This problem is known to be NP-complete [23].

Here we consider a variant with restricted values for k that we call VC$^{\beta}_{\triangle\text{-free}}(k)$, and we define as follows:

VC$^{\beta}_{\triangle\text{-free}}(k)$: Fix a positive integer $\beta \geq 2$. Given a simple planar triangle-free graph $G = (V, E)$, and a positive integer k, with $k < |E|^{\frac{1}{\beta}}$, is there a vertex cover of G of size k?

LEMMA 8. *For any constant integer $\beta \geq 2$, VC$^{\beta}_{\triangle\text{-free}}(k)$ is NP-complete.*

We are now going to give a reduction from VC$^{\beta}_{\triangle\text{-free}}(k)$. The reduction exploits ideas from the reduction obtained for k-minimax approval voting in [19]. Let $G = (V, E)$ be a simple planar triangle-free graph, and k be an integer such that $k < |E|^{\frac{1}{\beta}}$. We will construct a voting profile $P = (P_1, \ldots, P_n)$ with n voters and m candidates, as well as integers k' and α, so that:

$$G \text{ has a vertex cover of size } k \iff$$
$$\text{there is a set } S \text{ of size } k', \text{ such that } \mathbf{w}_n \boldsymbol{\cdot} \mathcal{H}(P, S) \leq \alpha\,.$$

For each $v_i \in V$, we consider a candidate c_i and for each $e_j \in E$ a voter j. The profile $P = (P_1, ..., P_{|E|})$ is then defined by $P_j = e_j$ for all j. Hence, each voter j only approves 2 candidates, the two endpoints of edge e_j. Let $k' = k$ and $\alpha = k$. The construction is polynomial, and notice that $m = |V|, n = |E|$.

Suppose that G has a vertex cover of size k. The corresponding set S of candidates will have a Hamming distance of k or $k - 2$ from any voter (depending on whether S contains one or both preferences of the voter; it contains at least one). Therefore, any element of $\mathcal{H}(P, S)$ is at most k, which yields:

$$\mathbf{w}_n \boldsymbol{\cdot} \mathcal{H}(P, S) \leq \sum_{i=1}^{n} a_i k = k = \alpha\,.$$

For the other direction, suppose that G has no vertex cover of size k. Then, any k-subset S of candidates will have a Hamming distance of $k + 2$ from at least one voter. Next, we are going to bound the number of voters that have distance $k-2$ from S. Notice that since G is planar and triangle-free, it is well known that k vertices can induce at most $2k - 4$ edges. It follows that at most $2k - 4$ edges can have both their endpoints covered by any set of k vertices. Equivalently, any k-subset S of candidates will have Hamming distance $k - 2$ from at most $2k - 4$ voters. Notice that

$$2k - 4 < 2|E|^{\frac{1}{\beta}} - 4 = 2n^{\frac{1}{\beta}} - 4\,.$$

Therefore, at least the first $\ell = \lceil n+4-2n^{\frac{1}{\beta}} \rceil$ distances in $\mathcal{H}(P, S)$ are at least k each, and the very first element is definitely $k+2$. And the last $\lfloor 2n^{\frac{1}{\beta}} - 4 \rfloor$ elements, may be $k-2$ or higher by the previous arguments. This gives:

$$\begin{aligned}
\mathbf{w}_n \boldsymbol{\cdot} \mathcal{H}(P, S) &\geq a_1(k + 2) + \sum_{i=2}^{\ell} a_i k + \sum_{i=\ell+1}^{n} a_i(k - 2) \\
&= 2a_1 + \sum_{i=1}^{n} a_i k - \sum_{i=\ell+1}^{n} 2a_i = \\
&= k + 2\left(a_1 - \sum_{i=\ell+1}^{n} a_i\right) > k\,,
\end{aligned}$$

which concludes the proof. □

Note that the above proof implies that the NP-completeness holds when k is a polynomially small fraction of the number of candidates. Actually, this is not necessary; we could have chosen $k' = k + \rho$, where ρ is polynomially bounded by n, m and add ρ candidates approved by all the voters. The proof would still go through.

Moreover, now it is relatively easy to generalize the NP-completeness to families of vectors $\mathbf{w}$ that have anything in their first coordinates. Actually, in the next theorem we generalize to families of vectors where a restriction applies only on the last coordinates of a subfamily of vectors indexed by some polynomially bounded function. The restriction in that case is very similar to the one in Theorem 7. Before we state the theorem, we give one more definition, which again corresponds to a fairness criterion.

DEFINITION 2. *Let $\beta \geq 2$ be an integer and $h(n) : \mathbb{N} \to \mathbb{N}$ be a polynomial-time computable function, such that $\forall n \in \mathbb{N},\ n \leq h(n) \leq p(n)$, where $p(n)$ is a polynomial. We say that a family of vectors $\mathbf{w}$ is (h, β)-restricted if for any $n \in \mathbb{N}$:*

i. $\mathbf{w}_n = \big(a_1(n), a_2(n), \ldots, a_n(n)\big) \in [0, 1]^n$ *with* $\sum_{i=1}^{n} a_i(n) = 1$

ii. If $n = h(n'), n' \in \mathbb{N}$, then $a_{n-n'+1} > \sum_{i=\ell+1}^{n} a_i$, where $\ell = \lceil n + 4 - 2(n')^{\frac{1}{\beta}} \rceil$

By $(k, \mathbf{w})$-AV$_{h,\beta}$ we denote the decision version of $(k, \mathbf{w})$-AV on (h, β)-restricted families of vectors.

THEOREM 9. *For any $\beta, h(n)$ as in the above definition, and for any (h, β)-restricted family of vectors $\mathbf{w}$, $(k, \mathbf{w})$-AV$_{h,\beta}$ is NP-complete.*

Finally, we prove NP-completeness for two families of vectors that are not entirely covered by the above results. Recall from Subsection 3.1 that by $\mathbf{f}^i$ we denote the family of vectors defined by $\mathbf{f}^i_n = \left(\frac{1}{n-i}, \ldots, \frac{1}{n-i}, 0, \ldots, 0\right)$.

THEOREM 10. *Let $c : \mathbb{N} \to \mathbb{N}$ with $1 \leq c(n) \leq n - 1$. Then, the decision version of $(k, \mathbf{f}^{c(n)})$-AV is NP-complete, even when restricted to balanced elections.*

The next family highlights that even slight deviations from MINISUM make the problem hard. We will revisit this family at the end of the next section too.

THEOREM 11. *Fix some integer $\beta \geq 2$. Let $\mathbf{w}_n = \frac{1}{n-\sum \varepsilon_i} \big(1, \ldots, 1, 1 - \varepsilon_1, 1 - \varepsilon_2, \ldots, 1 - \varepsilon_{c(n)}\big)$, where every $\varepsilon_i \in (0, 1]$, and $c : \mathbb{N} \to \mathbb{N}$ with $c(n) \leq \frac{n}{2} - 2\left(\frac{n}{2}\right)^{\frac{1}{\beta}}$. Then, the decision version of $(k, \mathbf{w})$-AV is NP-complete, even when restricted to balanced elections.*

4. POSITIVE RESULTS: EXACT AND APPROXIMATION ALGORITHMS

In this section we present families of vectors where we can compute an optimal solution, either exactly or approximately, in polynomial time.

4.1 Exact Algorithms

Recall that for $\mathbf{w}_n = (1/n, 1/n, \ldots, 1/n)$, $\mathbf{w}$-AV becomes MINISUM, and $(k, \mathbf{w})$-AV becomes k-MINISUM. It is known that computing an optimal minisum (resp. k-minisum) solution can be done in polynomial time [5]. Given the negative results of the previous section, we cannot hope to have polynomial time algorithms for rules that assign a relatively high score on the maximum Hamming distance. It is meaningful however to consider families of rules, where the maximum distance does not play a significant role. For example, in a large population of voters, it is anticipated that there will always be some voters whose preferences will be completely disjoint from the selected committee. In such cases it would make sense to disregard the first few (constant) largest distances and assign positive weight to the remaining coordinates of the weight vector. This is true for case *(i)* in Theorem 12 below. On top of case *(i)*, we also present an additional family of vectors for which we can still have polynomial time algorithms.

THEOREM 12. *Let $c, a_1, a_2, \ldots, a_c \in \mathbb{N}$ be fixed constants. Then, for the following families of vectors an optimal solution for both $\mathbf{w}$-AV and $(k, \mathbf{w})$-AV can be computed in polynomial time:*

i. $\mathbf{w}_n = \left(0, 0, \ldots, 0, \frac{1}{n-c}, \frac{1}{n-c}, \ldots, \frac{1}{n-c}\right)$,

ii. $\mathbf{w}_n = \frac{1}{\alpha}\left(0, 0, \ldots, 0, a_1, a_2, \ldots, a_c\right)$, *where* $\alpha = \sum_{i=1}^{n} a_i$.

PROOF. Due to limited space, we only prove (*i.*). The proof of (*ii.*) differs in the way we choose the committee S. Instead of utilizing the minisum solution on subsets of voters, one would need to solve several Integer Linear Programs, each with a constant number of constraints. We state the proof in terms of $(k, \mathbf{w})$-AV, but it is essentially identical for the case of $\mathbf{w}$-AV.

Assume that we have an instance of $(k, \mathbf{w})$-AV, with $\mathbf{w}_n = \left(0, \ldots, 0, \frac{1}{n-c}, \ldots, \frac{1}{n-c}\right)$. Let $P = (P_1, P_2, \ldots, P_n)$ be the voters' profile. There are $\binom{n}{c}$ subsets of voters of size $n - c$. For each one of them we find a corresponding k-MINISUM optimal solution. Among these solutions, let S be the one of minimum cost (for the corresponding k-MINISUM instance). Clearly, this can be done in polynomial time. We claim that S is an optimal solution for $(k, \mathbf{w})$-AV.

Suppose that the voters $x_1, \ldots, x_{n-c}$ defined the instance that produced S, and let $P_x = (P_{x_1}, \ldots, P_{x_{n-c}})$ be their preferences. Suppose, also, that S is not optimal for the initial problem, and let S' be a k-subset of the candidates, such that $\mathbf{w}_n \cdot \mathcal{H}(P, S')$ is minimum. Assume that voters $v_1, \ldots, v_{n-c}$ have the $n - c$ smallest distances from S' among all the voters. We then have:

$$\mathbf{w}_n \cdot \mathcal{H}(P, S') < \mathbf{w}_n \cdot \mathcal{H}(P, S) \leq \mathbf{f}^0_{n-c} \cdot \mathcal{H}(P_x, S) \,. \qquad (3)$$

The first inequality follows from S's suboptimality. For the second inequality, note that if we add new voters, while keeping S fixed, we can only reduce the weighted sum of the $n - c$ smallest Hamming distances (recall that $\mathbf{f}^0_{n-c} = \frac{1}{n-c}(1, ..., 1)$).

Moreover, if S'' is any optimal solution for the profile $P_v = (P_{v_1}, \ldots, P_{v_{n-c}})$ under $\mathbf{f}^0_{n-c}$, we have

$$\mathbf{f}^0_{n-c} \cdot \mathcal{H}(P_v, S'') \leq \mathbf{f}^0_{n-c} \cdot \mathcal{H}(P_v, S') = \mathbf{w}_n \cdot \mathcal{H}(P, S') \,, \qquad (4)$$

where the inequality follows from S'''s optimality and the equality follows from the choice of $v_1, \ldots, v_{n-c}$.

Combining (3) and (4) leads to

$$\mathbf{f}^0_{n-c} \cdot \mathcal{H}(P_v, S'') < \mathbf{f}^0_{n-c} \cdot \mathcal{H}(P_x, S) \,.$$

However, this contradicts the choice of S. Both S and S'' are among the $\binom{n}{c}$ k-MINISUM optimal solutions defined above. Moreover, S was the one with minimum cost (attained for the profile P_x), which yields a contradiction with the inequality above. We conclude that S is indeed optimal. □

4.2 LP-Based Approximation Algorithms

Next, we present two algorithms based on linear programming, for approximating $\mathbf{f}^{n-c(n)}$-AV and $(k, \mathbf{f}^{n-c(n)})$-AV, with $c : \mathbb{N} \to \mathbb{N}$ and $1 \leq c(n) \leq n-1$. This is the special case where we care for the $c(n)$ largest distances, since $\mathbf{f}^{n-c(n)}_n = \frac{1}{c(n)}(1, \ldots, 1, 0, \ldots, 0)$. By Theorems 7 and 10, we know that the decision versions of these problems are NP-complete.

First we present the LP approximation for $(k, \mathbf{f}^{n-c(n)})$-AV. Our approach follows closely the work of [7]. Given a profile P and a function $c(n)$, the algorithm uses the following Integer Linear Program (ILP).

$$\begin{aligned}
&\text{minimize: } q \\
&\text{subject to: } q + \frac{2}{c(n)} \sum_{\ell=1}^{c(n)} \sum_{a \in P_{i_\ell}} x_a \geq k + \frac{1}{c(n)} \sum_{\ell=1}^{c(n)} |P_{i_\ell}| \,, \\
&\qquad\qquad\qquad\qquad \forall\, i_1 < i_2 < \ldots < i_{c(n)} \in N \\
&\qquad \sum_{a \in A} x_a = k \\
&\qquad x_a \in \{0, 1\} \,, \quad \forall a \in A
\end{aligned}$$

The variable x_a denotes whether candidate a is included in the solution ($x_a = 1$) or not ($x_a = 0$). The first constraint essentially lower-bounds the value of q by the average of the $c(n)$ largest distances of the voters from the k candidates included in the solution. This is easier to see if we write the constraint as

$$q \geq \frac{1}{c(n)} \sum_{\ell=1}^{c(n)} \left(k + |P_{i_\ell}| - 2 \sum_{a \in P_{i_\ell}} x_a \right) .$$

The LP-based algorithm solves the LP relaxation in which the integrality constraint has been relaxed to $0 \leq x_a \leq 1$. In this way, a fractional solution is obtained with the x-variables having values in $[0, 1]$. Then, the algorithm includes the candidates with the k largest x-variables in the final solution (by breaking ties arbitrarily).

Here, we should note that the relaxed LP may have a superpolynomial number of constraints. So, in order to solve it in polynomial time, a separation oracle is needed [27]. It is not hard, though, to identify fast a violated constraint given an infeasible solution $(q, \vec{x})$. First, we can easily check whether $\sum_{a \in A} x_a \neq k$. If $\sum_{a \in A} x_a = k$, then for each voter i we compute $d_i = k + |P_i| - 2\sum_{a \in P_i} x_a$ in linear time, and then we sort the d_is. Let $i_1, i_2, \ldots, i_{c(n)}$ be the voters with the largest d_is. If $q \geq \frac{1}{c(n)} \sum_{\ell=1}^{c(n)} \left(k + |P_{i_\ell}| - 2\sum_{a \in P_{i_\ell}} x_a \right)$ then the solution $(q, \vec{x})$ is feasible, otherwise, we can see that $q < \frac{1}{c(n)} \sum_{\ell=1}^{c(n)} \Big(k + |P_{i_\ell}|$ $- 2\sum_{a \in P_{i_\ell}} x_a \Big)$ is a separating hyperplane.

THEOREM 13. *The LP-based algorithm above, has approximation ratio at most 2 for $(k, \mathbf{f}^{n-c(n)})$-AV, for any $c : \mathbb{N} \to \mathbb{N}$ with $1 \leq c(n) \leq n$.*

It is not hard to see that the particular LP relaxation has an integrality gap of almost 2 and thus we cannot hope to do any better using this LP formulation.

FACT 14. *The LP relaxation used in Theorem 13 has integrality gap at least $2 - \frac{2}{k}$.*

We continue with the LP approximation for $\mathbf{f}^{n-c(n)}$-AV. The notation as well as the actual algorithm are exactly the same as before, but we use a slightly different ILP:

minimize: q

subject to: $q + \frac{2}{c(n)} \sum_{\ell=1}^{c(n)} \sum_{a \in P_{i_\ell}} x_a - \sum_{a \in A} x_a \geq \frac{1}{c(n)} \sum_{\ell=1}^{c(n)} |P_{i_\ell}|$,

$\forall\, i_1 < i_2 < \ldots < i_{c(n)} \in N$

$x_a \in \{0,1\}$, $\forall a \in A$

As before, the first constraint lower-bounds the value of q by the average of the $c(n)$ largest distances of the voters from the set of candidates included in the solution. Again, the algorithm proceeds by solving the LP relaxation to get a fractional solution. Then, a candidate a is included in our solution if and only if $x_a \geq 1/2$. The relaxed LP may have a superpolynomial number of constraints, but we can construct a separation oracle using the same ideas as before.

THEOREM 15. *The LP-based algorithm above, has approximation ratio at most* 2 *for* $\mathbf{f}^{n-c(n)}$*-AV, for any* $c : \mathbb{N} \rightarrow \mathbb{N}$ *with* $1 \leq c(n) \leq n$.

Again, the LP relaxation has an almost matching integrality gap.

FACT 16. *The LP relaxation used in Theorem 15 has integrality gap at least* $2 - \frac{2}{m}$.

4.3 Performance of Minisum as an Approximation Algorithm

An interesting question is whether we can use other existing algorithms in approval voting to produce approximate solutions for our family. In this subsection, we focus on the questions of whether $\mathbf{w}$-AV (resp. $(k,\mathbf{w})$-AV) can be well approximated by the optimal solution to MINISUM (resp. k-MINISUM). The algorithm for solving MINISUM is quite simple to implement, it is also strategyproof, and hence a guarantee that it achieves a good approximation for any voting rule in our family would be a desirable property. It is known that in the case of MINIMAX, the minisum solution achieves an approximation factor of 3 [20]. Since the minimax solution is an extreme situation among the OWA operators, we expect that for most other vectors, minisum should not have an approximation ratio worse than 3. Here, we first extend the factor 3 approximation for all the families of non-increasing vectors and then also give improved approximation guarantees for vectors that are close to MINISUM in the sense defined below. We begin with a lemma that extends the 3-approximation for non-increasing vectors.

LEMMA 17. *Let* $\mathbf{w}$ *be a family of vectors as in the definition of* $\mathbf{w}$*-AV. Moreover, let* $\mathbf{w}_n$ *be non-increasing, for all* $n \in \mathbb{N}$. *Then, an optimal Minisum solution (resp. k-Minisum solution) achieves an approximation ratio of at most* 3 *for* $\mathbf{w}$*-AV (resp.* $(k,\mathbf{w})$*-AV).*

Notice that the above lemma holds in the special case where $\mathbf{w} = \mathbf{f}^i$, where i is allowed to be a function of n. We use Lemma 17 to prove the next theorem for vectors of the form $\mathbf{w}_n = \frac{1}{n-E(n)}\big(1, \ldots, 1, 1-\varepsilon_1, 1-\varepsilon_2, \ldots, 1-\varepsilon_{n-c(n)}\big)$, where $c : \mathbb{N} \rightarrow \mathbb{N}$ with $1 \leq c(n) \leq n$, $E(n) = \sum_{i=1}^{n-c(n)} \varepsilon_i$, and $\varepsilon_i : \mathbb{N} \rightarrow [0,1]$ for all $i \in \{1, \ldots, n-c(n)\}$. We call such a family of vectors a (c,E)-*reduced* family. Note that when the ε_is are small or even close to zero, then the voting rule is close to the minisum rule. Note also that under this definition, vectors $\mathbf{f}_n^{n-c(n)}$ are $(c(n), n-c(n))$-reduced.

THEOREM 18. *Let* $\mathbf{w}_n = \frac{1}{n-E(n)}\big(1, \ldots, 1, 1-\varepsilon_1, \ldots, 1-\varepsilon_{n-c(n)}\big)$ *be a* (c,E)*-reduced family of vectors. An optimal Minisum (resp. k-Minisum) solution achieves an approximation ratio of at most* $\min\left\{3\frac{n-E(n)}{c(n)}, \frac{n}{n-E(n)}\right\}$ *for* $\mathbf{w}$*-AV (resp.* $(k,\mathbf{w})$*-AV). If, moreover,* $\varepsilon_1(n) \leq \varepsilon_2(n) \leq \ldots \leq \varepsilon_{n-c(n)}(n)$ *for all* $n \in \mathbb{N}$, *then the above ratio is at most* $\min\left\{3, \frac{n}{n-E(n)}\right\}$.

PROOF. We give the proof in terms of $\mathbf{w}$-AV and Minisum, but it is the same for $(k,\mathbf{w})$-AV and k-Minisum. Let P be a profile and let S and O be optimal Minisum and $\mathbf{w}$-AV solutions respectively on input P. Let $OPT = \mathbf{w}_n \cdot \mathcal{H}(P,O)$. First, we prove that $\mathbf{w}_n \cdot \mathcal{H}(P,S) \leq \frac{n}{n-E(n)} OPT$. Indeed,

$$\begin{aligned}
\mathbf{w}_n \cdot \mathcal{H}(P,S) &= \tfrac{1}{n-E(n)}\big[(1,1,\ldots,1)\cdot\mathcal{H}(P,S) \\
&\quad - (0,\ldots,0,\varepsilon_1,\ldots,\varepsilon_{n-c(n)})\cdot\mathcal{H}(P,S)\big] \\
&\leq \tfrac{n}{n-E(n)}\left[\tfrac{1}{n}(1,1,\ldots,1)\cdot\mathcal{H}(P,S)\right] \\
&\leq \tfrac{n}{n-E(n)}\left[\tfrac{1}{n}(1,1,\ldots,1)\cdot\mathcal{H}(P,O)\right] \\
&\leq \tfrac{n}{n-E(n)}\left[\mathbf{w}_n\cdot\mathcal{H}(P,O)\right] = \tfrac{n}{n-E(n)}\, OPT\,.
\end{aligned}$$

Note that $\frac{1}{n}(1,\ldots,1)\cdot\mathcal{H}(P,S) \leq \frac{1}{n}(1,\ldots,1)\cdot\mathcal{H}(P,O)$ follows by the Minisum optimality of S. We also used the fact that, with the same set of distances, the largest dot product is given by the vector that is more biased towards the larger distances, to get $\frac{1}{n}(1,\ldots,1)\cdot\mathcal{H}(P,O) \leq \mathbf{w}_n\cdot\mathcal{H}(P,O)$.

Next we prove that $\mathbf{w}_n\cdot\mathcal{H}(P,S) \leq 3\frac{n-E(n)}{c(n)} OPT$. Notice that this ratio is better when $c(n)$ is a small fraction of n and the ε_is are close to 1. In the proof below, we denote by O' an optimal solution to $\mathbf{f}^{n-c(n)}$-AV. Recall that, by Lemma 17, we have $\mathbf{f}_n^{n-c(n)}\cdot\mathcal{H}(P,S) \leq 3\,\mathbf{f}_n^{n-c(n)}\cdot\mathcal{H}(P,O')$. We now have

$$\begin{aligned}
\mathbf{w}_n\cdot\mathcal{H}(P,S) &\leq \mathbf{f}_n^{n-c(n)}\cdot\mathcal{H}(P,S) \\
&\leq 3\,\mathbf{f}_n^{n-c(n)}\cdot\mathcal{H}(P,O') \\
&\leq 3\,\mathbf{f}_n^{n-c(n)}\cdot\mathcal{H}(P,O) \\
&= 3\,\tfrac{1}{c(n)}\,(1,\ldots,1,0,\ldots,0)\cdot\mathcal{H}(P,O) \\
&= 3\,\tfrac{n-E(n)}{c(n)}\left[\tfrac{1}{n-E(n)}\,(1,\ldots,1,0,\ldots,0)\cdot\mathcal{H}(P,O)\right] \\
&\leq 3\,\tfrac{n-E(n)}{c(n)}\left[\mathbf{w}_n\cdot\mathcal{H}(P,O)\right] \leq 3\,\tfrac{n-E(n)}{c(n)}\, OPT\,.
\end{aligned}$$

If, moreover, $\varepsilon_1(n) \leq \ldots \leq \varepsilon_{n-c(n)}(n)$, for all $n \in \mathbb{N}$, then $\mathbf{w}_n$ is non-increasing for all $n \in \mathbb{N}$, and by Lemma 17 we have $\mathbf{w}_n\cdot\mathcal{H}(P,S) \leq 3\,OPT$. ☐

Now, if we restrict Theorem 18 to $(c(n), n-c(n))$-reduced vectors we get an improvement on Lemma 17.

COROLLARY 19. *Let* $c : \mathbb{N} \rightarrow \mathbb{N}$ *with* $1 \leq c(n) \leq n$. *An optimal Minisum solution (resp. k-Minisum solution) achieves an approximation ratio of at most* $\min\left\{3, \frac{n}{c(n)}\right\}$ *for* $\mathbf{f}^{n-c(n)}$*-AV (resp.* $(k,\mathbf{f}^{n-c(n)})$*-AV).*

5. MANIPULABILITY

In this section, we briefly report on the manipulability of the $\mathbf{w}$-AV rules. We note first that the Gibbard-Satterthwaite theorem [26, 11] (or other known impossibility results) do not apply in our setting since we are in a domain of Hamming-induced preferences. Not surprisingly however, we will show that most of the rules in our family are manipulable. Let us, first, formally define manipulability in our setting. Given a profile P and an algorithm R, we denote by $R(P)$ the outcome of the algorithm on profile P. We also denote by P_{-i} the preferences of all voters besides i. Hence, we can also write P as (P_i, P_{-i}). Manipulability means that some voter i has an incentive to unilaterally change her preference so as to reduce the distance of P_i from the outcome of the algorithm.

DEFINITION 3. *An algorithm R is manipulable if for some profile P, there exists a voter i and a preference $P_i' \subseteq A$ such that*

$$d_{\mathcal{H}}(P_i, R(P_i', P_{-i})) < d_{\mathcal{H}}(P_i, R(P_i, P_{-i}))\,.$$

Note that so far we did not deal with how the ties are resolved in an election. It was implied that ties break arbitrarily, and this was sufficient. Here, we introduce a simple deterministic tie-breaking rule which assigns distinct IDs to the candidates (e.g., their index); ties are resolved by selecting the candidates with the smallest ID.

It is known that minimax approval voting is manipulable and that minisum approval voting is not (see [5, 20]). How does this extend to $\mathbf{w}$-AV (or $(k, \mathbf{w})$-AV) rules in general? Here again, we will focus mostly on non-increasing vectors, for the reason explained in Section 2. Due to space constraints, we only give an overview of our results and we omit the proofs. Note that for special cases, e.g., for a small, fixed number of candidates, one can obtain stronger results. We start with a theorem that holds for both $\mathbf{w}$-AV and $(k, \mathbf{w})$-AV.

THEOREM 20. *Let $\mathbf{w}_n = (w_1, w_2, \ldots, w_n)$ be any non-increasing family of vectors, such that for any $n \in \mathbb{N}$ it holds that $w_1 > w_{\lfloor \frac{2n}{3} \rfloor}$. Then, any exact algorithm for either $\mathbf{w}$-AV or $(k, \mathbf{w})$-AV, that breaks ties with the smallest-ID-first tie-breaking rule is manipulable.*

To illustrate the main idea of the proof, we give next a small example that extends to prove Theorem 20 in the case of $\mathbf{w}$-AV.

Consider an election E, with 4 voters $\{1, 2, 3, 4\}$, and 4 candidates $\{x_1, x_2, x_3, x_4\}$, and the following profile P:

$$P_1 : (0100); P_2 : (0101); P_3 : (0110); P_4 : (0111)$$

Figure 1 shows the $\mathbf{w}$-AV scores for some committees. One can easily verify that committees that are not mentioned have even larger $\mathbf{w}$-AV scores. The $\mathbf{w}$-AV scores of these 4 committees are

C	$\mathbf{w}_4 \cdot \mathcal{H}(P, C)$
0100	$2 \cdot w_1 + 1 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$
0101	$2 \cdot w_1 + 1 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$
0110	$2 \cdot w_1 + 1 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$
0111	$2 \cdot w_1 + 1 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$

Table 1: $\mathbf{w}$-AV scores with true preferences

equal and minimal, so there are 4 co-winning committees. Then, according to the tie-breaking rule, the committee (0111) is elected. The manipulation comes from voter 1 by voting (1000) instead of her true preferences. With this new vote, we have new $\mathbf{w}$-AV scores summarized in Figure 2. Then, with a non-increasing $\mathbf{w}$ such that

C	$\mathbf{w}_4 \cdot \mathcal{H}(P, C)$
0100	$2 \cdot w_1 + 2 \cdot w_2 + 1 \cdot w_3 + 1 \cdot w_4$
0101	$3 \cdot w_1 + 2 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$
0110	$3 \cdot w_1 + 2 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$
0111	$4 \cdot w_1 + 1 \cdot w_2 + 1 \cdot w_3 + 0 \cdot w_4$

Table 2: $\mathbf{w}$-AV scores with manipulation of voter 1

$w_1 > w_4$, we obtain a unique winning committee (0100).

This proof can be extended by adding an equal number of copies of voters 1, 2, 3 and 4, and by adding dummy candidates approved by all the voters. Actually, this gives a slightly better result than stated in Theorem 20, since it suffices to have $w_1 > w_{\lfloor \frac{3n}{4} \rfloor + 1}$.

For $\mathbf{w}$-AV and for strictly decreasing vectors we can state a stronger result than Theorem 20, for any tie-breaking rule.

PROPOSITION 21. *Let $\mathbf{w}$ be any family of strictly decreasing vectors. Then, any exact algorithm for $\mathbf{w}$-AV is manipulable independently of the tie-breaking rule used.*

Moreover, for $(k, \mathbf{w})$-AV we can state a result that holds for vectors that are not necessarily non-increasing.

PROPOSITION 22. *Let $\mathbf{w}_n = (w_1, w_2, \ldots, w_n)$ be a family of vectors, such that for any $n \in \mathbb{N}$, there exists some j with $\lfloor \frac{n+3}{2} \rfloor \leq j \leq \lfloor \frac{2n+1}{3} \rfloor$ and $w_1 > w_j$. Then, any exact algorithm for $(k, \mathbf{w})$-AV that breaks ties with the smallest-ID-first tie-breaking rule is manipulable.*

6. CONCLUSIONS

We have introduced a family of voting rules that generalize the standard ('minisum') rule and the minimax approval rule in committee elections and multiple referenda. By making use of Ordered Weighted Average operators, we are able to remedy the extreme behaviour of the minimax approval rule while retaining the idea of fairness to voters: our parameterized family of rules allows for fine-tuning the trade-off between fairness and utilitarian efficiency.

We have shown that although winner determination for rules belonging to this family is typically NP-hard, still there are cases where a winning committee can be computed efficiently. We also designed approximation algorithms for some of the cases where NP-hardness holds. Finally, we have addressed manipulability issues, and shown that, unsurprisingly, most of these rules are manipulable. Most of our results (both for computation and manipulability) hold for families of nonincreasing vectors, which correspond to "fair" rules that lay between pure egalitarianism (MINIMAX) and pure utilitarianism (MINISUM).

An interesting question is whether one can fully characterize the members of the family that admit polynomial time algorithms. As far as manipulability is concerned, it would be interesting to address the computational resistance to manipulability (as studied for other multiwinner voting rules in [24]), as well as to obtain a full characterization of strategyproof rules of our family (depending on n, m and $\mathbf{w}$). This would complete existing results by [3] for committee elections and by [18, 29] for multi-issue elections.

We have focused here on computation and manipulability; obviously, going further with an axiomatic study of our rules, along the general line for multiwinner rules exposed in [9], is the next major step in the study of our rules.

Finally, there is no reason to stick only to the Hamming distance for defining the satisfaction of a voter. Other possibilities, suggested in [13] and recently further studied in [1], are to measure the satisfaction of a voter by the fraction of her approved candidates that are elected (called *satisfaction approval voting*) or a concave function of the number of her approved candidates, such as in *proportional approval voting*. For all these other means of defining the satisfaction of a voter, one can define an OWA-based family of rules in the same way as we generalized minimax approval voting.

7. ACKNOWLEDGEMENTS

Research co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: "THALES - Investing in knowledge society through the European Social Fund", and by the ANR project 14-CE24-0007-01 – CoCoRICo-CoDec. We also thank the COST Action IC1205 on Computational Social Choice.

8. REFERENCES

[1] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. Justified representation in approval-based committee voting. In *AAAI-15*, 2015.

[2] H. Aziz, S. Gaspers, J. Gudmundson, S. Mackenzie, N. Mattei, and T. Walsh. Computational aspects of multi-winner approval voting. In *AAMAS-15*, 2015.

[3] S. Barbera, H. Sonnenschein, and L. Zhou. Voting by committees. *Econometrica*, 59(3):595–609, 1991.

[4] P. Berman, M. Karpinski, and A. D. Scott. Approximation hardness of short symmetric instances of MAX-3SAT. Technical Report TR03-049, Electronic Colloquium on Computational Complexity, 2003.

[5] S. J. Brams, D. M. Kilgour, and M. R. Sanver. A minimax procedure for electing committees. *Public Choice*, 3-4(132):401–420, 2007.

[6] J. Byrka and K. Sornat. PTAS for minimax approval voting. In *10th Conference on Web Interactions and Network Economics (WINE 2014)*, pages 203–217, 2014.

[7] I. Caragiannis, D. Kalaitzis, and E. Markakis. Approximation algorithms and mechanism design for minimax approval voting. In *AAAI*, pages 737–742, 2010.

[8] T. Cuhadaroğlu and J. Lainé. Pareto efficiency in multiple referendum. *Theory and Decision*, 72(4):525–536, 2012.

[9] E. Elkind, P. Faliszewski, P. Skowron, and A. Slinko. Properties of multiwinner voting rules. In *AAMAS-14*, 2014.

[10] M. Frances and A. Litman. On covering problems of codes. *Theory of Computing Systems*, 30:113–119, 1997.

[11] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41(4):587–602, 1973.

[12] J. Goldsmith, J. Lang, N. Mattei, and P. Perny. Voting with rank-dependent scoring rules. In *Proceedings of AAAI-14*, 2014.

[13] D. M. Kilgour. Approval balloting for multi-winner elections. In *Handbook on Approval Voting*, chapter 6. Springer, 2010.

[14] S. Konieczny and R. Pino-Perez. On the logic of merging. In *Proceedings of KR-1998*, pages 488–498, 1998.

[15] G. Laffond and J. Lainé. Condorcet choice and the Ostrogorski paradox. *Social Choice and Welfare*, 32(2):317–333, 2009.

[16] J. Lang, G. Pigozzi, M. Slavkovik, and L. van der Torre. Judgment aggregation rules based on minimization. In *TARK*, pages 238–246, 2011.

[17] J. Lang and L. Xia. Voting in combinatorial domains. In *Handbook of Computational Social Choice*. Cambridge University Press, 2014.

[18] M. Le Breton and A. Sen. Separable preferences, strategyproofness, and decomposability. *Econometrica*, 67(3):605–628, 1999.

[19] R. LeGrand. Analysis of the minimax procedure. Technical Report WUCSE-2004-67, Washington University, St Louis, 2004.

[20] R. LeGrand, E. Markakis, and A. Mehta. Some results on approximating the minimax solution in approval voting. In *AAMAS*, pages 198:1–198:3, 2007.

[21] M. Li, B. Ma, and L. Wang. On the closest string and substring problems. *Journal of the ACM*, 49:157–171, 2002.

[22] T. Lu and C. Boutilier. Budgeted social choice: From consensus to personalized decision making. In *IJCAI*, pages 280–286, 2011.

[23] C. E. V. Madhavan. Approximation algorithm for maximum independent set in planar traingle-free graphs. In *Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 381–392, 1984.

[24] R. Meir, A. D. Procaccia, J. S. Rosenschein, and A. Zohar. Complexity of strategic behavior in multi-winner elections. *Journal of Artificial Intelligence Research*, 33:149–178, 2008.

[25] A. D. Procaccia, J. S. Rosenschein, and A. Zohar. On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3):353–362, 2008.

[26] M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[27] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.

[28] P. Skowron, P. Faliszewski, and J. Lang. Finding a collective set of items: From proportional multirepresentation to group recommendation. In *COMSOC*, 2014.

[29] L. Xia and V. Conitzer. Strategy-proof voting rules over multi-issue domains with restricted preferences. In *WINE-10*, pages 402–414, 2010.

[30] R. R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 18:183–190, 1988.