

---

# Election Control by Manipulating Issue Significance

---

Andrew Estornell, Sanmay Das, Edith Elkind, Yevgeniy Vorobeychik  
Computer Science & Engineering, Washington University in St. Louis  
{aestornell,sanmay,yvorobeychik}@wustl.edu  
elkind@cs.ox.ac.uk

## Abstract

Integrity of elections is vital to democratic systems, but it is frequently threatened by malicious actors. The study of algorithmic complexity of the problem of manipulating election outcomes by changing its structural features is known as election control. One means of election control that has been proposed is to select a subset of issues that determine voter preferences over candidates. We study a variation of this model in which voters have judgments about relative importance of issues, and a malicious actor can manipulate these judgments. We show that computing effective manipulations in this model is NP-hard even with two candidates or binary issues. However, we demonstrate that the problem is tractable with a constant number of voters or issues. Additionally, while it remains intractable when voters can vote stochastically, we exhibit an important special case in which stochastic voting enables tractable manipulation.

## 1 INTRODUCTION

Fair elections are at the core of democratic systems. However, elections are increasingly subject to attack by malicious parties who aim to achieve personal goals at the expense of the social good [Caldwell et al., 2019]. The problem of election vulnerability to malicious attack has been studied in the broader literature on election control and bribery [Bartholdi III et al., 1992, Tomz and Houweling, 2008, Faliszewski and Rothe, 2016]. However, in much of this literature, control is exercised through a change in the election structure (e.g., adding and removing candidates), or directly the preferences of a subset of voters (bribery). A major means of election control that has often been overlooked in the research

literature is manipulation of *issues* that ultimately determine voter preferences over candidates.

A recent model of election control through issue selection attempts to bridge this gap [Lu et al., 2019]. The basis of this model is the spatial theory of voting [Downs, 1957, Enelow and Hinich, 1984], in which voters and candidates are represented as points in issue space, and a distance metric determines relative preferences, with voters preferring candidates who are similar to them on issues. In control through issue selection, a malicious party can select a subset of issues that then determines similarity and, consequently, voter preferences.

While capturing some of the intuition about the kinds of manipulations we commonly see (through, say, the spread of misinformation and fake news), control through issue selection nevertheless misses an essential factor: what is ultimately important, and what is at the core of manipulation is the *relative significance, or salience, of issues*, with issue selection being a rather extreme special case. A recent examples that illustrates this point is Brexit: Until 2016, the significance of the issue of U.K. membership in the EU was comparatively negligible [Khetani-Shah and Deutsch, 2019]. In 2016, it became one of the central issues, with considerable evidence pointing to Russian interference as a factor [Harper et al., 2019, Sabbagh et al., 2019]. In general, malicious parties can impact perceptions of relative issue importance in a variety of ways. For example, fake social media accounts can be used to coordinate widespread mentions of particular issues, increasing their salience compared to others. Similarly, influential individuals, such as celebrities or politicians, may be willing to accept payments to be more or less vocal about particular issues. To reflect the relative difficulty or cost of these actions we limit the attacker by one of two constraints.

Our model is a significant generalization of the work of Lu et al. [2019]. In our version, preferences of a voter over candidates are generated based on similarity in is-

sue space, *weighted by the relative importance of issues*. We study the complexity of this problem in the context of plurality elections for two common models of voter behavior in the spatial framework: 1) deterministic voting, in which voters always vote for their most preferred candidate, and 2) stochastic voting, where the probability of a voter voting for a candidate is a monotonic function of weighted issue similarity. We show that the control problem is in general NP-hard in either case, even with only 2 candidates. Indeed, for the deterministic case we demonstrate hardness even with only  $\Theta(\ell)$  voters, where  $\ell$  is the number of issues. Next, we exhibit several tractable special cases. In the deterministic case, if the number of voters is  $\Theta(\log(\ell))$ , or the number of issues is constant, election control is in P. In stochastic voting, in turn, control is tractable if the probability of voting for a candidate is linear in their weighted distance from the voter.

**Related Work** The complexity of controlling elections has seen extensive treatment, starting with the work of Bartholdi III et al. [1992]; see Hemaspaandra et al. [2007], Menton [2013], Erdélyi et al. [2015], Chen et al. [2017] for further examples and the survey by Faliszewski and Rothe [2016] for an overview. Variations of this problem consider attacks that add, remove, partition or clone candidates or voters, for a variety of voting rules. However, most of the prior election control literature considers election models in which voter preferences over candidates are given, rather than generated based on distance in issue space. The spatial model of elections, in turn, has received considerable attention in prior literature [Davis and Hinich, 1966, Enelow and Hinich, 1984, 1990, McKelvey and Ordeshook, 1990, Merrill and Groffman, 1999, Anshelevich et al., 2018, Anshelevich and Postl, 2017]. However, most of this research has focused on problems other than election control. For example, extensive literature exists on game-theoretic models in which candidates opportunistically select positions in issue space [Downs, 1957, Shen and Wang, 2017, Sabato et al., 2017].

A direct precursor to our model, combining election control with spatial theory of voting, is Lu et al. [2019], who study the model in which an adversary can select an arbitrary subset of issues in an election. These issues are then used to generate voter preferences over candidates, with voters preferring candidates who are closest to them on the adversarially selected issues. We significantly generalize this model by allowing the adversary to change relative importance of issues.

## 2 MODEL

Let  $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_m\}$  and  $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  be a set of candidates and voters, respectively. Each candidate and

voter is a vector over a set of issues  $[\ell]$  defined by  $\mathbf{c}_i = \langle c_{i,1}, \dots, c_{i,\ell} \rangle \in \mathbb{R}^\ell$  and  $\mathbf{v}_j = \langle v_{j,1}, \dots, v_{j,\ell} \rangle \in \mathbb{R}^\ell$ . We consider plurality elections in which each voter is asked to report the most preferred candidate, and the winner is the candidate who tallies the most votes.

Suppose that the relative importance of issues to voters is determined by a weight vector  $\mathbf{w} = \langle w_1, \dots, w_\ell \rangle$ , where  $0 \leq w_k \leq 1$  for  $k \in [\ell]$ , and  $\|\mathbf{w}\|_1 = 1$ . In this model, a voter’s preferences over candidates are determined by the weighted distance from each candidate in issue space. Formally, the weighted distance between a voter  $\mathbf{v}_j$  and candidate  $\mathbf{c}_i$  is

$$d_{(i,j)} = \left( \sum_{k=1}^{\ell} w_k |c_{i,k} - v_{j,k}|^p \right)^{1/p},$$

$p \geq 1$ , and the voter prefers a candidate who is closer according to this weighted distance measure.

Without loss of generality, suppose that  $\mathbf{c}_1$  is the attacker’s preferred candidate. In our model, an attacker aims to influence the election by modifying the relative importance of issues. Specifically, the attacker changes  $\mathbf{w}$  into a modified preference vector  $\mathbf{w}' = \mathbf{w} + \mathbf{x}$ . Restrictions on the attacker’s strength are given in one of the following forms.

1. **Normed Budget Constraint:** Given a budget  $B \in \mathbb{R}$  the attacker’s total perturbation must be less than the budget, i.e.  $\|\mathbf{x}\|_p \leq B$ .
2. **Interval Constraint:** Given a set of intervals  $I = I_1 \times \dots \times I_\ell \subset [0, 1]^\ell$ , the attacker’s new weight vector must fall within  $I$ , i.e.  $w_k \in I_k$  for all  $k \in [\ell]$ .

For both of these constraints, we consider two paradigms for voters selecting their desired candidate.

1. **Deterministic voting:** Voter  $\mathbf{v}_j$  votes for candidate  $\mathbf{c}_i^* \in \operatorname{argmin}_{\mathbf{c}_i} d_{i,j}$ , breaking ties according to candidate order  $\mathbf{c}_1 \prec \dots \prec \mathbf{c}_m$ .
2. **Stochastic voting:**  $\mathbb{P}(\mathbf{v}_j \text{ votes for } \mathbf{c}_i) = f(\mathbf{v}_j, \mathbf{c}_i)$ , where  $f : \mathcal{V} \times \mathcal{C} \rightarrow [0, 1]$  is a proper probability mapping and  $\sum_i f(\mathbf{v}_j, \mathbf{c}_i) = 1$  for all  $j \in [n]$ .

Both are common models translating voter-candidate distance into voting behavior Enelow and Hinich [1990].

When voters select candidates deterministically, we consider two objectives for the attacker. The first objective (*Max Support*) is to maximize the total number of votes for their preferred candidate,  $\mathbf{c}_1$ :

$$\operatorname{maximize}_{\mathbf{x}} |\{\mathbf{v}_j \in \mathcal{V} : \mathbf{c}_1 \in \operatorname{argmin}_{\mathbf{c}_i} d_{i,j}\}|$$

The second objective (*Majority Vote*) is to win the plurality vote:

$$\mathbf{c}_1 \in \operatorname{argmax}_{\mathbf{c}_i \in \mathcal{C}} |\{\mathbf{v}_j \in \mathcal{V} : \mathbf{c}_i \in \operatorname{argmin}_{\mathbf{c}_e} d_{\ell,j}\}|,$$

For stochastic voting, we consider the objective of maximizing the expected number of votes for  $\mathbf{c}_1$ :

$$\operatorname{maximize}_{\mathbf{x}} \sum_{j=1}^n f(\mathbf{v}_j, \mathbf{c}_1).$$

### 3 DETERMINISTIC VOTING

We begin by investigating our model of election control when voters always vote for their most preferred candidate. For compactness of notation we introduce, for every candidate and voter pair, a preference vector which gives  $\mathbf{v}_j$ 's unweighted preference for  $\mathbf{c}_1$  over  $\mathbf{c}_i$  on each issue. Let  $a_{j,k}^{(i)} = |c_{i,k} - v_{j,k}|^p - |c_{1,k} - v_{j,k}|^p$ . The preference vector for  $\mathbf{v}_j$  of  $\mathbf{c}_1$  over  $\mathbf{c}_i$  is  $\mathbf{a}_j^{(i)} = \langle a_{j,1}^{(i)}, \dots, a_{j,\ell}^{(i)} \rangle$ . The condition for  $\mathbf{v}_j$  voting for  $\mathbf{c}_1$  is then

$$\sum_{k=1}^{\ell} w_k a_{j,k}^{(i)} \geq 0 \quad \text{for all } i \in [m] \setminus \{1\}.$$

In the case of only two candidates, we omit the index  $i$  and denote the preference vector for  $\mathbf{c}_1$  over  $\mathbf{c}_2$  by  $\mathbf{a}_j$ .

In this section, we will show that both *Max Support* and *Majority Vote* are NP-hard, even when there are only two candidates, issues are binary, and the attacker has no constraints on their strength. If these assumptions are further restricted such that there are only  $\Theta(\ell)$  voters, where  $\mathbf{v}_k$  agrees with  $\mathbf{c}_1$  on exactly  $k$  issues, then *MaxSupport* is still NP-hard.

Although both objectives are hard, even with several strong restrictions, we present sufficient conditions for the problem to become tractable. The following positive results hold for the normed budget constraint when  $p \in \{1, 2, \infty\}$ , and for the interval constraint with any  $p \geq 1$ . Under the normed budget, with  $p \notin \{1, 2, \infty\}$ , results also hold, however, the computed  $\mathbf{x}$  may break the attacker's budget by some small  $\epsilon > 0$ , i.e. for both objectives we obtain, in polynomial time, a vector  $\mathbf{x}$  that is guaranteed to have  $\|\mathbf{x}\|_p \leq B + \epsilon$ . The tractable cases are as follows: *Max Support* has a polynomial time algorithm when either there are at most  $\Theta(\log(\ell))$  voters, or the number of issues and the number of values each issue can take on are both constant, and *Majority Votes* has a polynomial time algorithm when either the number of voters is constant, or the number of issues and the number of values each issue can take on are both constant.

### 3.1 HARDNESS OF CONTROL IN DETERMINISTIC SETTINGS

First we show that even for two candidates and binary issues *Majority Vote* (WTCP) is NP-complete, and that *Max Support* (TCWMS) is NP-hard. Both of these results are the product of hardness reductions from the problem of election control by issue selection Lu et al. [2019]. We first give a formal definition of each problem used in the reductions.

**Definition 1.** Let  $\mathcal{C}'$  be a set of  $m'$  candidates and  $\mathcal{V}'$  be a set of  $n'$  voters. Both candidates and voters are vectors in  $\{0, 1\}^\ell$ , indicating positions on  $[\ell']$  issues. An adversary selects a subset  $S \subset [\ell']$  with  $S \neq \emptyset$  for the objective of either determining if  $\mathbf{c}_1$  can win the plurality (TCIS), or maximizing the total number of votes  $\mathbf{c}_1$  received (TCMS).

TCIS was shown to be NP-complete and TCMS to be NP-hard by [Lu et al., 2019].

**Theorem 1.** For 2 candidates,  $n$  voters, and  $\ell$  binary issues, the problem of maximizing the number of votes for  $\mathbf{c}_1$ , TCWMS, is NP-hard, even when  $B = \infty$  in the normed budget constraint, or when  $I = [0, 1]^\ell$  in the interval constraint.

*Proof.* To prove this claim we will reduce from the problem of 0-1 issue selection on binary issues with two candidates (TCMS). An instance of TCMS is defined by a set of  $n'$  voters  $\mathcal{V}' = \{\mathbf{v}'_1, \dots, \mathbf{v}'_{n'}\}$  and two candidates  $\mathbf{c}'_1, \mathbf{c}'_2$  all of which select positions on  $\ell'$  binary issues. The objective of TCMS is to maximize the number of votes for  $\mathbf{c}'_1$  subject to the constraint that  $\mathbf{w}' \in \{0, 1\}^{\ell'}$  and  $\|\mathbf{w}'\|_1 > 0$ . To reduce from a given instance of TCMS we will add a set of voters that forces any optimal solution to have  $\mathbf{w} \in \{0, c\}^\ell$  for some constant  $c$  that can be associated with 1 in the instance of TCMS. Since  $B = \infty$ , we may assume the adversary is selecting the weight vector  $\mathbf{w}$  rather than a perturbation  $\mathbf{x}$ . Without loss of generality we may assume  $\mathbf{c}'_1 = \langle 1, 1, \dots, 1 \rangle$  and  $\mathbf{c}'_2 = \langle 0, 0, \dots, 0 \rangle$ .

First, let  $\ell = 2\ell' + 2$  and  $\mathbf{c}_1 = \langle 1, 1, \dots, 1 \rangle, \mathbf{c}_2 = \langle 0, 0, \dots, 0 \rangle$ . To encode  $\mathcal{V}'$ , let  $\mathcal{V}_1$  be a set of voters obtained by mapping each  $\mathbf{v}'_j \in \mathcal{V}'$  to a voter  $\mathbf{v}_j$  where  $v_{j,k} = v_{j,k+\ell'+1} = v'_{j,k}$  for  $k \in [\ell']$  and  $v_{j,\ell'+1} = 1, v_{j,2\ell'+2} = 0$ . Compactly, each voter can be represented as  $\mathbf{v}_j = \langle v'_{j,1}, \dots, v'_{j,\ell'}, 1, v_{j,1}, \dots, v_{j,\ell'}, 0 \rangle$ . Next we will introduce five more sets of voters which will force any optimal  $\mathbf{w}$  to be binary.

For each  $r \in [\ell' + 1]$  construct  $8n'\ell'$  identical copies of a voter who has  $v_{j,k} = 1$  when  $(k+r) \bmod \ell \leq \ell' + 1$ , and  $v_{j,k} = 0$  otherwise. Denote this set of voters as  $\mathcal{V}_2$ . Let  $\mathcal{V}_3$  be a set of voters obtained by taking each  $\mathbf{v}_j \in \mathcal{V}_2$

and flipping their opinion, i.e. for each  $\mathbf{v}_{j_1} \in \mathcal{V}_2$  add  $\mathbf{v}_{j_2}$  to  $\mathcal{V}_3$  where  $v_{j_2,k} = 1 - v_{j_1,k}$  for all  $k \in [\ell]$ . Note that  $|\mathcal{V}_2| = |\mathcal{V}_3| = 8n'\ell'(\ell' + 1)$ .

Now let  $\mathcal{V}_4$  be the set of  $4n'\ell'$  voters such that each  $\mathbf{v}_j \in \mathcal{V}_4$  has  $v_{j,\ell'+1} = v_{j,2\ell'+2} = 1$  and  $v_{j,k} = 0$  for all other  $k$ .

For each  $r \in [\ell']$  we create  $2n'$  voters of the form  $v_{j,r} = v_{j,r+\ell'+1} = 1$ , and for each  $k \neq r$ ,  $v_{j,k} = 1$ ,  $v_{j,k+\ell'+1} = 0$ . Call this set of  $2n'\ell'$  voter  $\mathcal{V}_5$ . Lastly, for each  $r \in [\ell']$  create  $2n'$  voters with  $v_{j,k} = 1$ ,  $v_{j,k+\ell'+1} = 0$  for  $k \neq r, \ell'+1$ , and  $v_{j,r} = v_{j,r+\ell'+1} = 1$ ,  $v_{j,\ell'+1} = v_{j,2\ell'+2} = 0$ . Call this set of  $2n'\ell'$  voters  $\mathcal{V}_6$ . Let  $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \mathcal{V}_3 \cup \mathcal{V}_4 \cup \mathcal{V}_5 \cup \mathcal{V}_6$ .

Note that all voters outside of  $\mathcal{V}_1$  have at least  $2n'$  copies of themselves. Therefore no optimal solution will have a voter  $\mathbf{v}_{j_1} \in \mathcal{V}_1$  vote for  $\mathbf{c}_1$  if doing so meant losing any voter  $\mathbf{v}_{j_2} \notin \mathcal{V}_1$ . As a result we will first examine criteria of optimal solutions over  $\mathcal{V}_1^c$ .

Note that the preference vector of  $\mathbf{v}_j$  has  $a_{j,k} = 1$  if  $v_{j,k} = c_{1,k}$  and  $a_{j,k} = -1$  if  $v_{j,k} = c_{2,k}$ .

Consider the voters in  $\mathcal{V}_2$ , each of which was created according to some  $r \in [\ell' + 1]$ . For each  $\mathbf{v}_j \in \mathcal{V}_2$ , we have that

$$\langle \mathbf{a}_j, \mathbf{w} \rangle \geq 0 \iff \sum_{k_1 \in I_r} w_{k_1} - \sum_{k_2 \in [\ell] \setminus I_r} w_{k_2} \geq 0$$

where  $I_r = \{k \in [\ell] : (k+r) \bmod \ell \leq \ell'+1\}$ . Similarly, for each voter  $\mathbf{v}_j \in \mathcal{V}_3$  we have

$$\langle \mathbf{a}_j, \mathbf{w} \rangle \geq 0 \iff \sum_{k_1 \in I_r} w_{k_1} - \sum_{k_2 \in [\ell] \setminus I_r} w_{k_2} \leq 0$$

Therefore, all  $16n'(\ell' + 1)$  voters can be made to vote for  $\mathbf{c}_1$  if

$$\sum_{k_1 \in I_r} w_{k_1} = \sum_{k_2 \in [\ell] \setminus I_r} w_{k_2} \quad \forall r \in [\ell' + 1]$$

The above system of linear equations has a unique solution, namely  $w_k = w_{k+\ell'+1}$  for all  $k$ . For any  $\mathbf{v}_j \in \mathcal{V}_2 \cup \mathcal{V}_3$ , there are strictly more copies of  $\mathbf{v}_j$  than there are total voters in all other voter sets combined. Therefore any optimal solution must have all voters in  $\mathcal{V}_2 \cup \mathcal{V}_3$  voting for  $\mathbf{c}_1$ . As a result we will work under the assumption that  $w_k = w_{k+\ell'+1}$  for all  $k$ .

All voters in  $\mathcal{V}_4$  are of the form  $v_{j,r} = v_{j,r+\ell'+1} = 0$  for some  $r \in [\ell']$ ,  $v_{j,\ell'+1} = v_{j,2\ell'+2} = 1$ , and  $v_{j,k} = 1$ ,  $v_{j,k+\ell'+1} = 0$  for all  $k \neq r, \ell' + 1$ . If  $\mathcal{V}_4$  is made to vote for  $\mathbf{c}_1$  then

$$w_{\ell'+1} + w_{2\ell'+2} \geq \max_k \{w_k + w_{k+\ell'+1} : k \in [\ell']\},$$

which would immediately imply that

$$0 < w_{\ell'+1} = w_{2\ell'+2} \geq \max_k \{w_k : k \in [\ell']\}. \quad (1)$$

Since there are more copies of each voter in  $\mathcal{V}_4$  than there are total remaining voters and since every voter in  $\mathcal{V}_2 \cup \mathcal{V}_3 \cup \mathcal{V}_4$  can be made to vote for  $\mathbf{c}_1$ , no optimal solution would have any of these voters vote for  $\mathbf{c}_2$ , and Equation 1 holds.

Finally, consider the voters in  $\mathcal{V}_5$  and  $\mathcal{V}_6$ . Each voter in  $\mathcal{V}_5$  is of the form  $v_{j,r} = v_{j,r+\ell'+1} = 0$  for some  $r \in [\ell']$ ,  $v_{j,k} = 1$ ,  $v_{j,k+\ell'+1} = 0$  for all  $k \neq r$ . Each voter in  $\mathcal{V}_6$  is of the form  $v_{j,r} = v_{j,r+\ell'+1} = 1$  for some  $r \in [\ell']$ ,  $v_{j,\ell'+1} = v_{j,2\ell'+2} = 0$ , and  $v_{j,k} = 1$ ,  $v_{j,k+\ell'+1} = 0$  for all  $k \neq r, \ell' + 1$ . Note that for each  $r$  there are  $2n'$  copies of the corresponding voter in  $\mathcal{V}_5$  and of the corresponding voter in  $\mathcal{V}_6$ , and that for each  $r$  either the set of voters in  $\mathcal{V}_5$  vote for  $\mathbf{c}_1$  or the voters in  $\mathcal{V}_6$  vote for  $\mathbf{c}_1$ . To see this, fix any  $r$  and consider the voters in either set. If the voter from  $\mathcal{V}_5$  votes for  $\mathbf{c}_1$  then it must be the case that

$$\begin{aligned} \sum_{k \neq \ell'+1, 2\ell'+2} w_k a_{j,k} &\geq w_{\ell'+1} + w_{2\ell'+2} \\ \iff w_r = w_{r+\ell'+1} = w_{\ell'+1} = w_{2\ell'+2}. \end{aligned}$$

Alternatively, if the voter is in  $\mathcal{V}_6$ , then

$$\begin{aligned} -w_r + w_{r+\ell'+1} + \sum_{k \neq r} w_k - w_{k+\ell'+1} &\geq 0 \\ \iff -w_r - w_{r+\ell'+1} &\geq 0. \end{aligned}$$

Both of the constraints cannot hold since  $w_{\ell'+1} = 0$  would imply that  $\max_k \{w_k, k \in [\ell]\} = 0$  and  $\sum_k w_k = 0$ . Again, there are  $2n'$  copies of each voter in  $\mathcal{V}_5, \mathcal{V}_6$  and there remain only  $n'$  voters left, so it must be the case that any optimal solution gains either the voter in  $\mathcal{V}_5$  or in  $\mathcal{V}_6$  for each  $r$ . Therefore, any optimal solution must have  $w_k \in \{0, w_{\ell'+1}\}$  with  $w_{\ell'+1} > 0$ .

Thus, as the only voters left to sway are those in  $\mathcal{V}_1$ , which corresponds to  $\mathcal{V}$ , it must be the case that there is a maximum of  $8\ell'^2 n' + 13\ell' n' + \alpha$  voters if and only if an optimal solution in the given instance of TCMS attains  $\alpha$  voters.  $\square$

When there are only two candidates, the problem of winning the plurality vote becomes a special case of maximizing the number of votes for  $\mathbf{c}_1$ . The proof of Theorem 1 can be easily extended to the problem of winning the plurality, by adding a set of voters who agree with  $\mathbf{c}_2$  on all issues, such that this set ‘‘cancels out’’ any votes for  $\mathbf{c}_1$  from the constructed voters. These new voters can clearly not be won over by any nonzero weight vector. This yields the following theorem.

**Theorem 2.** For 2 or more candidates,  $n$  voters, and  $\ell$  binary issues, the problem of determining if  $\mathbf{c}_1$  can win the plurality, TCWP, is NP-complete, even when  $B = \infty$  in the normed budget constraint or when  $I = [0, 1]^\ell$  in the interval constraint.

Next, we proceed to considerably strengthen the hardness result in Theorem 1. When there are only two candidates and issues are binary, a partial order can be induced on voters by the number of issues they agree with  $\mathbf{c}_1$  on. That is, the set of voters can be partitioned into  $\ell$  tiers  $S_r = \{\mathbf{v}_j \in \mathcal{V} : \sum_{k=1}^{\ell} a_{j,k} = 2r - \ell\}$ , where  $S_r$  is the set of voters who agree with  $\mathbf{c}_1$  on exactly  $r$  issues. We now show that even if for all  $r \in [\ell]$  there are only a constant number of voters who agree with  $\mathbf{c}_1$  on exactly  $r$  issues, maximizing the number of votes for  $\mathbf{c}_1$  is still NP-hard.

**Theorem 3.** Suppose there are only two candidates,  $\ell$  binary issues, and  $\Theta(\ell)$  voters. Suppose further that for each  $r \in [\ell]$   $|\{\mathbf{v}_j \in \mathcal{V} : \sum_{k=1}^{\ell} a_{j,k} = 2r - \ell\}| \in \Theta(1)$ . Then maximizing the number of voters for  $\mathbf{c}_1$  is NP-hard, even when  $B = \infty$  in the normed budget constraint, or when  $I = [0, 1]^\ell$  in the interval constraint.

*Proof.* To prove this claim, we will reduce from TCWMS, which was shown to be NP-hard in Theorem 1. An instance of TCWMS is defined by two candidates  $\mathbf{c}'_1, \mathbf{c}'_2$ , a voter set  $\mathcal{V}'$ , and a set of  $\ell'$  issues taking on values 0, 1.

In the constructed instance of our problem let  $\ell = n'^2 \ell'^2$  and assume w.l.o.g. that  $\mathbf{c}_1 = \langle 1, 1, \dots, 1 \rangle$ ,  $\mathbf{c}_2 = \langle 0, 0, \dots, 0 \rangle$ . We will construct a set of voters  $\mathcal{V}$  that encodes the voters in  $\mathcal{V}'$  such that the election only depends on issues  $[\ell']$ . To do this, first decompose the set of voters into disjoint sets  $\mathcal{V}' = S_1 \cup S_2 \cup \dots \cup S_{\ell'}$ , such that  $S_r = \{\mathbf{v}' \in \mathcal{V}' : \sum_{k'=1}^{\ell'} a'_{k'} = 2r - \ell'\}$ . Starting at  $r = 1$ , iterate through each  $\mathbf{v}'_j \in S_r$  and create a voter  $\mathbf{v}_j$  such that  $v_{j,k} = v'_{j,k}$  for all  $k \in [\ell']$ ,  $v_{j,k+\ell'} = 1$  for all  $k \in [\frac{j^2+j}{2} : \frac{j^2+3j}{2}]$  and  $v_{j,k+\ell'} = 0$  otherwise. Under this construction, for any index  $s \in [\ell' + 1 : n'^2 \ell'^2]$ , there is only a single  $j$  for which  $v_{j,s} = 1$ . That is, each voter has either  $\sum_s w_s$  or  $-\sum_s w_s$ , for  $s \in [\ell' + \frac{j^2+j}{2} : \ell' + \frac{j^2+3j}{2}]$ , as terms in  $\langle \mathbf{w}, \mathbf{a}_j \rangle$ . Therefore, each of these index sets  $[\ell' + \frac{j^2+j}{2} : \ell' + \frac{j^2+3j}{2}]$  can be associated with a single index, say  $j^*$ , where for all  $j \neq j^*$ ,  $\mathbf{a}_{j,j^*} = -1$  and  $\mathbf{a}_{j^*,j^*} = 1$ . Under this simplified version of indices, we see that for  $\mathbf{v}_j$ , and for  $k \in [n']$  that the contribution from all  $k^*$  issues, to the voters preference sum is,  $-w_{1^*} - \dots - w_{j^*-1} + w_{j^*} - w_{j^*-1} - \dots - w_{n'^*}$ . If we take any three of these sums as linear inequalities for voters  $\mathbf{v}_{j_1}, \mathbf{v}_{j_2}, \mathbf{v}_{j_3}$ , we get

$$-w_{1^*} - \dots - w_{j_1^*-1} + w_{j_1^*} - w_{j_1^*-1} - \dots - w_{n'^*} \geq 0$$

$$\begin{aligned} -w_{1^*} - \dots - w_{j_2^*-1} + w_{j_2^*} - w_{j_2^*-1} - \dots - w_{n'^*} &\geq 0 \\ -w_{1^*} - \dots - w_{j_3^*-1} + w_{j_3^*} - w_{j_3^*-1} - \dots - w_{n'^*} &\geq 0 \end{aligned}$$

Since each  $0 \leq w_{k^*} \leq 1$ , the only satisfying assignment to these three inequalities is  $w_{k^*} = 0$  for all  $k^*$ . Therefore the objectives of both problems align and this restricted version of WTCMS is NP-hard.  $\square$

### 3.2 TRACTABLE SPECIAL CASES

We now return to the setting when there are  $m$  candidates,  $n$  voters, and issues are real-valued. Although both *Max Support* and *Majority Vote* are NP-hard even with several strong restrictions, we now show sufficient conditions for either objective to be computed efficiently, as well as algorithms to do so.

Under the normed budget constraint when  $p \in \{1, 2, \infty\}$ , or under the interval constraint when  $p \geq 1$ , if the number of voters is  $\Theta(\log(\ell))$ , or the number of issues is constant and issue values are from a set of constant size, then *Max Support* can be computed in polynomial time. For the normed budget constraint with  $p \notin \{1, 2, \infty\}$ , if the maximum number of votes for  $\mathbf{c}_1$  is  $\alpha$  when  $\|\mathbf{x}\|_p \leq B$ , then for  $\epsilon > 0$  a perturbation  $\mathbf{x}'$  where  $\|\mathbf{x}'\|_p \leq B + \epsilon$  and  $\mathbf{c}_1$  obtains  $\alpha' \geq \alpha$  votes, can be found in polynomial time with respect to the input size and  $\log(\frac{1}{\epsilon})$ . Moreover, as  $\epsilon \rightarrow 0$ ,  $\alpha'$  asymptotically approaches  $\alpha$ .

Under similar assumptions on the number of voters or issues, the objective of *Majority Vote* can be computed in polynomial time for the normed budget constraint with  $p \in \{1, 2, \infty\}$ , or for the interval constraint with  $p \geq 1$ . In the case of  $p \notin \{1, 2, \infty\}$ , suppose that  $\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x}} \{\|\mathbf{x}\|_p : \mathbf{c}_1 \text{ wins the election}\}$ , assuming  $\mathbf{c}_1$  can be made to win the election. For  $\epsilon > 0$  a perturbation  $\mathbf{x}'$  with  $\mathbf{c}_1$  winning the election and  $\|\mathbf{x}'\|_p - \|\mathbf{x}^*\|_p \leq \epsilon$ , can be found in polynomial time with respect to the problem size and  $\log(\frac{1}{\epsilon})$ . This  $\mathbf{x}'$  might break the attacker's budget by at most  $\epsilon$ . If  $\|\mathbf{x}'\|_p \leq B$  then, simply taking  $\mathbf{x} = \mathbf{x}'$  wins the election within the budget constraint. However, in the case when  $B < \|\mathbf{x}'\|_p \leq B + \epsilon$ , it will be unknown whether there exists a  $\mathbf{x}$  with  $\|\mathbf{x}\|_p \leq B$  such that  $\mathbf{c}_1$  wins the election. If the attacker is allowed to break their budget by  $\epsilon$ , i.e.  $\|\mathbf{x}\|_p \leq B + \epsilon$  then taking  $\mathbf{x} = \mathbf{x}'$  wins  $\mathbf{c}_1$  the election. Further, as  $\epsilon \rightarrow 0$ ,  $\mathbf{x}'$  asymptotically approaches  $\mathbf{x}^*$ .

The existence of polynomial time algorithms for these two objectives is particularly interesting, given that the problem was NP-hard in the case of control by issue selection *even for a single voter* [Lu et al., 2019].

In both cases we use Algorithm 1, where *unanimity-program*, refers to an optimization program in which all

voters in the given demographic,  $D \subset \mathcal{V}$ , are made to unanimously vote for a given candidate. Recall that for

---

**Algorithm 1:** Maximizing votes for  $\mathbf{c}_1$

---

**Result:** weight vector achieving the most voters  
**for**  $D \in 2^{\mathcal{V}}$  **do**  
    Solve unanimity-program over  $D$ ;  
    **if** unanimity-program feasible and is within  
        budget restriction **then**  
        | Store  $|D|$  and  $\mathbf{w}_D$ ;  
    **end**  
**end**  
**return**  $\operatorname{argmax} \{|D| : \mathbf{w}_D\}$

---

a given candidate–voter pair  $\mathbf{c}_i, \mathbf{v}_j$ , the vector  $\mathbf{a}_j^{(i)}$  gives  $\mathbf{v}_j$ 's per-issue preference for  $\mathbf{c}_1$  over  $\mathbf{c}_i$ .

Under the normed budget constraint, the unanimity program for a demographic,  $D \subset \mathcal{V}$ , is given by

$$\begin{aligned} & \text{minimize}_{\mathbf{x}} \|\mathbf{x}\|_p \\ & \text{s.t. } \|\mathbf{x} + \mathbf{w}\|_1 = 1 \\ & \quad 0 \leq w_k + x_k \leq 1 \quad \forall k \in [\ell] \\ & \quad \langle \mathbf{w} + \mathbf{x}, \mathbf{a}_j^{(i)} \rangle \geq 0 \quad \forall i \in [m], \forall \mathbf{v}_j \in D \end{aligned} \quad (2)$$

and under the interval constraint, the unanimity program is given by the following linear feasibility problem:

$$\begin{aligned} & \|\mathbf{x} + \mathbf{w}\|_1 = 1 \\ & 0 \leq w_k + x_k \leq 1 \quad \forall k \in [\ell] \\ & w_k + x_k \in I_k \quad \forall k \in [\ell] \\ & \langle \mathbf{w} + \mathbf{x}, \mathbf{a}_j^{(i)} \rangle \geq 0 \quad \forall i \in [m], \forall \mathbf{v}_j \in D \end{aligned} \quad (3)$$

**Theorem 4.** *Suppose there are  $m$  candidates,  $\ell$  real-valued issues, and  $n \in \Theta(\log(\ell))$  voters and the attacker is restricted by  $\|\mathbf{x}\|_p \leq B$ , where  $p \in \{1, 2, \infty\}$ . Then Algorithm 1 computes Max Support for  $\mathbf{c}_1$ , in polynomial time.*

*Proof.* Since  $|\mathcal{V}| = n \in \Theta(\log(\ell))$ ,  $|2^{\mathcal{V}}| = |2^{\Theta(\log(\ell))}| \in \Theta(\ell)$ . Each subset of voters  $D \in 2^{\mathcal{V}}$  is referred to as a demographic. For any  $D \in 2^{\mathcal{V}}$ , determining if all voters in  $D$  can be made to unanimously vote for  $\mathbf{c}_1$  can be computed by solving Program 2. We minimize over  $\|\mathbf{x}\|_p$  in order to determine if the minimum change to  $\mathbf{w}$ , such that all of  $D$  votes for  $\mathbf{c}_1$ , is larger than  $B$ . That is, demographics that cannot be made to vote for  $\mathbf{c}_1$  come in two forms: those where the constraint set is infeasible, and those where the value of the optimal solution is greater than the budget  $B$ . By selecting the largest viable demographic, the maximum votes for  $\mathbf{c}_1$  can be found. For  $p \in \{1, 2, \infty\}$  Program 2 can be solved in polynomial time. When  $p = 1, \infty$  the program reduces to a

linear program, and when  $p = 2$  the program reduces to a positive definite quadratic program, all of which have polynomial time algorithms.  $\square$

**Theorem 5.** *Suppose there are  $m$  candidates,  $\ell$  real-valued issues, and  $n \in \Theta(\log(\ell))$  voters, and the attacker is restricted by  $\|\mathbf{x}\|_p \leq B$ , where  $p \notin \{1, 2, \infty\}$ . Let  $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \{\|\mathbf{x}\|_p : \mathbf{c}_1 \text{ has maximum votes and } \|\mathbf{x}\|_p \leq B\}$ . Then for any  $\epsilon > 0$  a perturbation,  $\mathbf{x}'$ , can be computed in polynomial time with respect to the problem size and  $\log(\frac{1}{\epsilon})$ , that obtains at least as many votes as  $\mathbf{x}^*$  and  $\|\mathbf{x}'\|_p - \|\mathbf{x}^*\|_p \leq \epsilon$ .*

*Proof.* Similarly to Theorem 4, Program 2 can be solved for each demographic. In contrast to Theorem 4, when  $p \notin \{1, 2, \infty\}$  we are solving a general convex program, and thus polynomial time solutions will be off by at most a factor of  $\epsilon$ . For a given demographic,  $D$ , suppose the optimal solution to Program 2 is  $\mathbf{x}_D^*$ . Then for  $\epsilon > 0$  we can obtain a solution  $\mathbf{x}'_D$  such that  $\|\mathbf{x}'_D\|_p \leq \|\mathbf{x}_D^*\|_p + \epsilon$ . As before, demographics that cannot be made to unanimously vote for  $\mathbf{c}_1$  come in two forms: demographics in which the constraints of the Program 2 are infeasible, and demographics for which the optimal  $\mathbf{x}_D^*$  has  $\|\mathbf{x}_D^*\|_p > B$ . We need not consider demographics of the first type, since the  $\epsilon$  approximation of the convex program will not return a vector if the constraint set is infeasible. Via the same strategy as Theorem 4, we solve each program and take the largest demographic that can be made to unanimously vote for  $\mathbf{c}_1$ . The key difference in this case, is that we may be selecting a demographic  $D'$  that has more voters than the optimal solution, and requires budget  $B + \epsilon$  to obtain. Thus, if  $\mathbf{x}^*$  is the smallest vector, with  $\|\mathbf{x}^*\|_p \leq B$ , that obtains the maximum votes for  $\mathbf{c}_1$ , then  $\|\mathbf{x}^*\|_p \leq \|\mathbf{x}'\|_p \leq B + \epsilon$ .  $\square$

**Theorem 6.** *Suppose there are  $m$  candidates,  $\ell$  real-valued issues, and  $n \in \Theta(\log(\ell))$  voters and the attacker has the interval constraint for some interval  $I \subset [0, 1]^\ell$ . Then Algorithm 1 computes Max Support for  $\mathbf{c}_1$ , in polynomial time.*

*Proof.* Each unanimity program is now given by Program 3, which is simply a feasibility LP. Therefore determining if a particular demographic can be made to vote for  $\mathbf{c}_1$  can be done in polynomial time. As stated in 4, there are  $\theta(\ell)$  demographics that need to be checked and thus the maximum number of votes for  $\mathbf{c}_1$  can be computed in polynomial time.  $\square$

**Theorem 7.** *Suppose there are  $m$  candidates,  $n$  voters, and  $\ell \in \Theta(1)$  issues, each of which take on values from a set of constant size. Then Algorithm 1 computes Max*

Support for  $\mathbf{c}_1$  in polynomial time, for the normed budget restriction when  $p \in \{1, 2, \infty\}$ .

*Proof.* Since  $\ell \in \Theta(1)$  and positions are selected from a set of constant size, say  $r$ , then only  $r^\ell$  distinct voters can exist. So, there may be  $n$  voters, but at most  $\Theta(1)$  of them that need to be investigated. Let  $\mathcal{V}' \subset \mathcal{V}$  be the set of all unique voters in  $\mathcal{V}$ . For each  $\mathbf{v} \in \mathcal{V}'$  we also keep track of the number of times  $\mathbf{v}$  appears in  $\mathcal{V}$ . So,  $|\mathcal{V}'| \in \Theta(1)$  and there are only a constant number of programs to solve, the only difference being that we now choose the feasible demographic representing the maximum number of voters in  $\mathcal{V}$ , rather than  $\mathcal{V}'$ . As stated before, each program can be efficiently solved when  $p \in \{1, 2, \infty\}$ .  $\square$

**Theorem 8.** *Suppose there are  $m$  candidates,  $n$  voters,  $\ell \in \Theta(1)$  issues, each of which take on values from a set of constant size, and the attacker is restricted by  $\|\mathbf{x}\|_p \leq B$ , where  $p \notin \{1, 2, \infty\}$ . Let  $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \{\|\mathbf{x}\|_p : \mathbf{c}_1 \text{ has maximum votes and } \|\mathbf{x}\|_p \leq B\}$ . Then for any  $\epsilon > 0$  a perturbation,  $\mathbf{x}'$ , can be computed in polynomial time with respect to the problem size and  $\log(\frac{1}{\epsilon})$ , that obtains at least as many votes as  $\mathbf{x}^*$  and  $\|\mathbf{x}'\|_p - \|\mathbf{x}^*\|_p \leq \epsilon$ .*

*Proof.* We again use the idea in the proof of Theorem 7 by keeping track of the unique voters. Once we have the set of unique voters, the proof is identical to that of Theorem 4.  $\square$

**Theorem 9.** *Suppose there are  $m$  candidates,  $n$  voters, and  $\ell \in \Theta(1)$  issues, each of which take on values from a set of constant size. Then Algorithm 1 computes Max Support for  $\mathbf{c}_1$  in polynomial time, under the interval constraint for given intervals  $I \subset [0, 1]^\ell$ .*

*Proof.* After constructing the set of unique voters, we solve a constant number of linear programs and take the vector yielding the largest number of votes for  $\mathbf{c}_1$ .  $\square$

**Theorem 10.** *Suppose that there are  $m$  candidates and either  $n \in \Theta(1)$ , or  $\ell \in \Theta(1)$  where each issue takes on values from a set of constant size. Then under the budgeted constraint for  $p \in \{1, 2, \infty\}$ , Majority Vote can be computed in polynomial time.*

*Proof.* In this setting the number of unique partitions of  $\mathcal{V}$  is constant. Thus, if there are  $m$  candidates, there are  $m^{\Theta(1)}$  unique ways in which each partition can be assigned to a candidate. This assignment of disjoint demographics to candidates is equivalent to that particular demographic being made to vote for that candidate. Each pairing, for a given partition  $P$ , can be given by a set  $A = \{(\mathbf{v}_{j_p}, \mathbf{c}_{i_p}), p \in P\}$ . To check if there exists

a weight vector such that the given pairing is attainable, one need only solve Program 2, with the additional set of linear constraints that  $\langle \mathbf{w} + \mathbf{x}, \mathbf{a}_{j_p}^{(i_p, i)} \rangle \geq 0$  for all  $j \in [n]$ ,  $i \in [m]$ , and for all  $(\mathbf{v}_{j_p}, \mathbf{c}_{i_p}) \in A$ , where  $\mathbf{a}_{j_p}^{(i_p, i)}$  is the preference vector of  $\mathbf{v}_j$  for  $\mathbf{c}_{i_p}$  over  $\mathbf{c}_i$ . There are only  $m^{\Theta(1)}$  programs to solve, each of which takes polynomial time.  $\square$

**Theorem 11.** *Suppose that there are  $m$  candidates and either  $n \in \Theta(1)$ , or  $\ell \in \Theta(1)$  where each issue takes on values from a set of constant size. Suppose further that the attacker is restricted by  $\|\mathbf{x}\|_p \leq B$ , where  $p \notin \{1, 2, \infty\}$ . Let  $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \{\|\mathbf{x}\|_p : \mathbf{c}_1 \text{ wins and } \|\mathbf{x}\|_p \leq B\}$ . Then for any  $\epsilon > 0$  a perturbation,  $\mathbf{x}'$ , can be computed in polynomial time with respect to the problem size and  $\log(\frac{1}{\epsilon})$ , that  $\mathbf{c}_1$  wins the election and  $\|\mathbf{x}'\|_p - \|\mathbf{x}^*\|_p \leq \epsilon$ .*

*Proof.* As in the proof of Theorem 10, the feasibility of each assignment of voters to candidates can be formulated a convex program. An assignment of voters to candidates is valid if the program is feasible and if the optimal solution has value at most  $B$ . For  $p \notin \{1, 2, \infty\}$  these programs cannot be solved exactly in polynomial time, but for any  $\epsilon > 0$  where  $\frac{1}{\epsilon}$  is polynomial with respect to the problem size, a solution  $\mathbf{x}'$ , with  $\|\mathbf{x}'\|_p - \|\mathbf{x}^*\|_p \leq \epsilon$ , can be computed efficiently. Thus, we obtain solutions for  $\mathbf{x}'$  for each program and take the one with the smallest one under the  $l_p$  norm such that  $\mathbf{c}_1$  wins the election.  $\square$

**Theorem 12.** *Suppose that there are  $m$  candidates and either  $n \in \Theta(1)$ , or  $\ell \in \Theta(1)$  where each issue takes on values from a set of constant size. Then under the interval constraint, for  $I \subset [0, 1]^\ell$ , Majority Vote can be computed in polynomial time.*

*Proof.* Under the interval constraint each possible assignment of voters to candidates can be formulated as a linear program. As shown in the proof of Theorem 10 there are only a constant number of such assignments and thus we need only solve a constant number of linear programs and then choose the assignment of voters to candidates such that  $\mathbf{c}_1$  wins the election.  $\square$

## 4 STOCHASTIC VOTING

Another common model for candidate selection is that of stochastic voting, where votes are cast via a distribution over candidates Schofield et al. [1998]. More precisely, let  $f$  be a function that maps weighted distance between a voter  $\mathbf{v}_j$  and a candidate  $\mathbf{c}_i$  to a probability of the voter voting for this candidate. Next, we show that election control in this setting for general  $f$  is NP-hard even when

we only have 2 candidates. For this, suppose that  $f$  belongs to a general class of *sigmoidal* functions Udell and Boyd [2013], of which the logistic function is a well-known member.

**Definition 2.** A function  $f : [l, u] \rightarrow \mathbb{R}$  is said to be sigmoidal if it is Lipschitz continuous and one of the following is true:  $f$  is convex,  $f$  is concave, or there exists  $z \in [l, u]$  such that  $f$  is concave on  $[l, z]$  and convex on  $[z, u]$ .

We now show the hardness of maximizing the expected number of votes for  $\mathbf{c}_1$  even in the two-candidate case.

**Theorem 13.** Suppose there are 2 or more candidates,  $n$  voters and  $\ell$  issues, where votes are cast via a sigmoidal function. Then maximizing the expected number of votes for  $\mathbf{c}_1$  is NP-hard, even when  $B = \infty$ , or  $I = [0, 1]^\ell$ .

*Proof.* We reduce from the known NP-hard problem Max-2SAT. An instance of Max-2SAT can be defined by a set of  $\ell$  Boolean variables  $B = \{b_1, \dots, b_\ell\}$  and a set of  $n$  clauses  $\Phi = \{(x_{1,1} \vee x_{1,2}), \dots, (x_{n,1} \vee x_{n,2})\}$  where each  $x_k \in \{-b_k, b_k\}$ . Let the number of issues be  $\ell + 1$ , let  $\mathbf{c}_1 = \langle 1, 1, \dots, 1 \rangle$  and let  $\mathbf{c}_2 = \langle 0, 0, \dots, 0 \rangle$ . Define  $\beta_1, \beta_2 \in \Theta(n)$ . Create  $4\ell^2 n^2 (\beta_1 + \beta_2)$  voters of the form  $v_{j,k} = 0$  for all  $k \in [\ell]$  and  $v_{j,\ell+1} = 1$ . For each Boolean variable,  $b_r \in B$  create  $n^2 \beta_1$  voters of the form  $v_{j,k} = 0.5$  if  $k \notin \{r, \ell + 1\}$ ,  $v_{j,r} = 1$ , and  $v_{j,\ell+1} = 0$ . Additionally, for each  $b_r$ , create  $n^2 \beta_2$  voters of the form  $v_{j,k} = 0.5$  for all  $k \neq r$  and  $v_{j,r} = 0$ . Finally, we encode each clause as a voter. Clauses can take on one of the three forms and we map each form to a voter in the following way:

1.  $(b_{r_1} \vee b_{r_2})$  yields  $v_{j,r_1} = v_{j,r_2} = 1 - v_{j,\ell+1} = 0$  and  $v_{j,k} = 0.5$  for all  $k \neq r_1, r_2, \ell + 1$ .
2.  $(\neg b_{r_1} \vee b_{r_2})$  yields  $1 - v_{j,r_1} = v_{j,r_2} = 1$  and  $v_{j,k} = 0.5$  for all  $k \neq r_1, r_2$ .
3.  $(\neg b_{r_1} \vee \neg b_{r_2})$  yields  $v_{j,r_1} = v_{j,r_2} = 1$ ,  $v_{j,\ell+1} = 0$ , and  $v_{j,k} = 0.5$  for all  $k \neq r_1, r_2, \ell + 1$ .

Under this construction, the attacker's objective function can be formulated as

$$4n^2 \ell^2 (\beta_1 + \beta_2) \theta\left(-\sum_{k=1}^{\ell} w_k + w_{\ell+1}\right) + \sum_{j=1}^n \theta(\langle \mathbf{w}, \mathbf{a}_j \rangle) + \sum_{k=1}^{\ell+1} \left( n^2 \beta_1 \theta(-w_k - w_{\ell+1}) + n^2 \beta_2 \theta(w_k - w_{\ell+1}) \right).$$

To complete the proof, we show that this objective is maximized by values that can be mapped back to binary values that satisfy the maximum number of clauses.

Since  $\theta$  defines a probability,  $0 \leq \theta(x) \leq 1$  for any value of  $x$ . Suppose  $\theta$  is a sigmoid function defined by some sharpness factor  $\alpha$ . The objective function can be examined by each of its terms, starting with,  $\theta(-\sum_{k=1}^{\ell} w_k + w_{\ell+1})$ . Since more than  $\frac{3}{4}$  of voters contribute to this term, it must be the case that  $\theta(-\sum_{k=1}^{\ell} w_k + w_{\ell+1}) \geq \frac{3}{4}$ , which implies that  $-\sum_{k=1}^{\ell} w_k + w_{\ell+1} \geq 0$  as desired. Further, since  $0 \leq w_k \leq 1$ , it must be the case that  $w_{\ell+1} \geq \frac{3}{4}$ .

Next consider the term  $\beta_1 \theta(-w_k - w_{\ell+1}) + \beta_2 \theta(w_k - w_{\ell+1}) \leq \beta_1 \theta(-w_k - 3/4) + \beta_2 \theta(w_k - 3/4)$ . Each of these terms is maximized at one of the extremes,  $w_k = 0$  or  $w_k = 1$ . We can always choose the coefficients  $\beta_1, \beta_2$  so that the value of the sum agrees at  $w_k = 0$  and  $w_k = 1$ . Lastly, we look at the terms that encode the clauses, namely  $\theta(\langle \mathbf{w}, \mathbf{a}_j \rangle)$ . Clearly any satisfying assignment of the clause yields a greater value for this term than any non-satisfying assignment. However, it remains to be shown that oversatisfying any number of clauses does not lead to a greater value of the sum than exactly satisfying any single clause. We will look at clauses of the form  $(b_{r_1} \vee b_{r_2})$ , but symmetric analysis holds for the other two cases. Each of these clauses produces a term  $\theta(w_{r_1} + w_{r_2} - w_{\ell+1})$ . So the marginal gain for over satisfying this clause is at most  $\theta(1 + 1 - 3/4) - \theta(1 - 3/4) = (1 + e^{-\frac{5\alpha}{4}})^{-1} - (1 + e^{-\frac{\alpha}{4}})^{-1}$ , where  $\alpha$  is the sharpness of the sigmoid function. For the sake of analysis, assume  $\alpha \geq n$ , although similar techniques work for smaller  $\alpha$ . Then, the gain from exactly satisfying a single clause is  $\theta(1 - 3/4) - \theta(-3/4) = (1 + e^{-\frac{\alpha}{4}})^{-1} - (1 + e^{\frac{3\alpha}{4}})^{-1}$ .

Thus, the gain for exactly satisfying a single clause, compared to over satisfying all  $n$  clauses is  $(1 + e^{-\frac{\alpha}{4}})^{-1} - (1 + e^{\frac{3\alpha}{4}})^{-1} - n((1 + e^{-\frac{5\alpha}{4}})^{-1} - (1 + e^{-\frac{\alpha}{4}})^{-1})$ . This term is positive for  $n \geq 10$ . Therefore no optimal solution will over satisfy clauses unless all satisfiable clauses have been satisfied. Therefore maximizing the expected number of voters and maximizing the number of satisfied clauses are equivalent.  $\square$

#### 4.1 A TRACTABLE SPECIAL CASE

While in general election control is hard even in the stochastic model of voting, we now exhibit a special case which is tractable. Specifically, we show that for  $m$  candidates  $n$  voters and  $\ell$  real-valued issues, if  $f$  is linear, the problem of maximizing the expected number of votes for  $\mathbf{c}_1$  is in P, under either strength constraint.

**Theorem 14.** Suppose that  $f$  is linear, and there are  $m$  candidates,  $n$  voters,  $\ell$  real-valued issues, and the attacker is restricted by either the normed budget constraint, for any  $p \geq 1$ , or the interval constraint. Then



maximizing the expected number of votes for  $\mathbf{c}_1$  is computable in polynomial time.

*Proof.* We may more explicitly represent  $f(\mathbf{v}_j, \mathbf{c}_i)$  as  $f(x_1, \dots, x_{m-1})$ , where each variable  $x_i$  represents  $\langle \mathbf{w}, \mathbf{a}_j^{(i)} \rangle$ . Suppose that  $f$  is linear. Since  $\mathbf{a}_j^{(i)}$  is a constant value, and the dot product is also linear,  $f$  depends linearly on each  $w_k$ . Therefore maximizing the expected number of voters can be formulated as a convex program

$$\begin{aligned} & \text{maximize}_{\mathbf{x}} \sum_j f(\langle \mathbf{w} + \mathbf{x}, \mathbf{a}_j^{(2)} \rangle, \dots, \langle \mathbf{w} + \mathbf{x}, \mathbf{a}_j^{(m)} \rangle) \\ & \text{s.t. } \|\mathbf{x}\|_p \leq B \\ & \quad 0 \leq w_k + x_k \leq 1 \quad \forall k \in [\ell] \\ & \quad \|\mathbf{w} + \mathbf{x}\|_1 = 1 \end{aligned}$$

In the case of the interval constraint,  $\|\mathbf{x}\|_p \leq B$  is replaced by  $\ell$  linear constraints of the form  $w_k + x_k \in I_k$  for intervals  $I_k \subset [0, 1]$ , making the above program a linear program. Therefore, under the interval constraint, the optimal  $\mathbf{x}$  is computable in polynomial time. However, under the budget constraint the program is not linear, but is still solvable in polynomial time. This can be seen by the fact that the objective function is linear with respect to  $\mathbf{x}$  and the only non-linear constraint is  $\|\mathbf{x}\|_p \leq B$ . As a result, we need only figure out the ratio of increase that each  $x_k$  gives to the objective function relative to increase given to  $\|\mathbf{x}\|_p$ .

Since  $f$  linearly depends on inner products (linear functions) we may decompose the objective sum into

$$\left( \sum_{k=1}^{\ell} b_k x_k \right) + C$$

for some constant values  $b_1, \dots, b_\ell, C$ . The constant  $C$  can be ignored for the purposes of maximization. First assume that we do not have the constraints  $0 \leq w_k + x_k \leq 1$  or  $\|\mathbf{w} + \mathbf{x}\|_1 = 1$ . Then the program reduces to

$$\text{maximize}_{\mathbf{x}} \sum_{k=1}^{\ell} b_k x_k \quad \text{s.t. } \|\mathbf{x}\|_p \leq B \quad (4)$$

The optimal solution to this program must have the property that,  $|b_{k_1} x_{k_1}^{-(p-1)}| = |b_{k_2} x_{k_2}^{-(p-1)}| \quad \forall k_1, k_2 \in [\ell]$  Further, any optimal solution must have  $\|\mathbf{x}\|_p = B$ , and  $\text{sign}(x_k) = \text{sign}(b_k)$ . Therefore the vector maximizing Program (4) can be computed analytically as

$$\mathbf{x}^* = \frac{1}{p-1} \left\langle x_1, \text{sign}(b_2) |x_1|^{p-1} \sqrt[p-1]{b_2}, \dots, \text{sign}(b_\ell) |x_1|^{p-1} \sqrt[p-1]{b_\ell} \right\rangle$$

Since  $\|\mathbf{x}\|_p = B$ , the value of  $x_1$  is unique.

However, this may not be a feasible solution to the original program since both the constraints that  $0 \leq w_k +$

$x_k \leq 1$  and  $\sum_{k=1}^{\ell} x_k = 0$  have been ignored. The first constraint,  $-w_k \leq x_k \leq 1 - w_k$ , may be violated if  $x_k$  was made too large or too small in  $\mathbf{x}^*$ . This is fixed by iteratively computing solutions to Program 4 in the following way. If any  $x_k$  does not satisfy the  $-w_k \leq x_k \leq 1 - w_k$ , truncate  $x_k$  to either value, depending on the sign( $b_k$ ). Then remove the truncated variables from the program and re-compute the solution. This must terminate in  $\ell$  or fewer steps.

Next we need only deal with the constraint that  $\sum_{k=1}^{\ell} x_k = 0$ . This constraint can be satisfied by partitioning the  $x_k$  values depending on  $\text{sign}(x_k)$ . Sort each partition by value of  $|\frac{b_k}{x_k^{p-1}}|$ . If  $\sum_{k=1}^{\ell} x_k \neq 0$  then the weight on the  $x_k$ 's with the smallest ratio of  $|\frac{b_k}{x_k^{p-1}}|$  can be shifted onto  $x_k$ 's onto elements in the opposite partition with the largest ratio.

Thus both constraints are satisfied without decreasing the objective value, and we have the maximum number expected votes that  $\mathbf{c}_1$  can obtain.  $\square$

## 5 CONCLUSION

We consider the problem of election control in the spatial model of voting, where a preference of a voter for a candidate is determined by the distance between them in issue space weighted by the relative importance of issues to voters. We suppose that a malicious actor aims to skew election results in favor of their preferred candidate by changing the relative perceived importance of issues. We show that this problem is NP-hard for the adversary even when there are only 2 candidates, and whether voters cast their votes deterministically or stochastically. On the other hand, we exhibit several special cases which are tractable, including settings with a constant number of voters or issues. Our model of spatial voting opens a novel direction in election control, but still makes a number of limiting assumptions, including an assumption that all voters have the same relative preferences over issues. Relaxing these is a natural subject for future work.

## Acknowledgments

This research was partially supported by the NSF (IIS-1903207, IIS-1910392), and ARO (W911NF1910241).

## References

E. Anshelevich and J. Postl. Randomized social choice functions under metric preferences. *Journal of Artificial Intelligence Research*, 58:797–827, 2017.

- E. Anshelevich, O. Bhardwaj, E. Elkind, J. Postl, and P. Skowron. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264:27–51, 2018.
- J. J. Bartholdi III, C. A. Tovey, and M. A. Trick. How hard is it to control an election? *Mathematical and Computer Modelling*, 16(8-9):27–40, 1992.
- L. A. Caldwell, H. Przybyla, and K. Stewart. Senate intelligence report finds ‘extensive’ Russian election interference. In *NBC News*. 2019.
- J. Chen, P. Faliszewski, R. Niedermeier, and N. Talmon. Elections with few voters: Candidate control can be easy. *Journal of Artificial Intelligence Research*, 60: 937–1002, 2017.
- O. A. Davis and M. J. Hinich. A mathematical model of preference formation in a democratic society. In *Mathematical Applications in Political Science II*, pages 175–208. Southern Methodist University Press, 1966.
- A. Downs. *An Economic Theory of Democracy*. Columbia University Press, 1957.
- J. M. Enelow and M. J. Hinich. *The spatial theory of voting: An introduction*. Cambridge University Press, Cambridge, MA, 1984.
- J. M. Enelow and M. J. Hinich. *Advances in the spatial theory of voting*. Cambridge University Press, 1990.
- G. Erdélyi, M. R. Fellows, J. Rothe, and L. Schend. Control complexity in Bucklin and fallback voting: A theoretical analysis. *Journal of Computer and System Sciences*, 81(4):632–660, 2015.
- P. Faliszewski and J. Rothe. Control and bribery in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, pages 146–168. Cambridge University Press, 2016.
- T. Harper, C. Wheeler, and R. Kerbaj. Revealed: the Russia report. In *The Sunday Times*. 2019.
- E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Anyone but him: The complexity of precluding an alternative. *Artificial Intelligence*, 171(5-6):255–285, 2007.
- S. Khetani-Shah and J. Deutsch. Brexit timeline: From referendum to EU exit. In *Politico Pro*. 2019.
- J. Lu, D. Zhang, Z. Rabinovich, S. Obraztsova, and Y. Vorobeychik. Manipulating elections by selecting issues. In *Proceedings of AAMAS’19*, pages 529–537, 2019.
- R. D. McKelvey and P. C. Ordeshook. A decade of experimental research on spatial models of elections and committees. In *Advances in the spatial theory of voting*, pages 99–144. 1990.
- C. Menton. Normalized range voting broadly resists control. *Theory of Computing Systems*, 53(4):507–531, 2013.
- S. Merrill and B. Groffman. *A Unified Theory of Voting: Directional and Proximity Spatial Models*. Cambridge University Press, 1999.
- I. Sabato, S. Obraztsova, Z. Rabinovich, and J. S. Rosenschein. Real candidacy games: a new model for strategic candidacy. In *Proceedings of AAMAS’17*, pages 867–875. 2017.
- D. Sabbagh, P. Walker, and L. Harding. No 10 blocks Russia EU referendum report until after election. In *The Guardian*. 2019.
- N. Schofield, I. Sened, and D. Nixon. Nash equilibrium in multiparty competition with ‘stochastic’ voters. *Annals of Operations Research*, 84:3–27, 1998.
- W. Shen and Z. Wang. Hotelling-Downs model with limited attraction. In *Proceedings of AAMAS’17*, pages 660–668, 2017.
- M. Tomz and R. Van Houweling. Candidate positioning and voter choice. *American Political Science Review*, 2008.
- M. Udell and S. Boyd. Maximizing a sum of sigmoids. *Optimization and Engineering*, pages 1–25, 2013.