On Detecting Nearly Structured Preference Profiles

Edith Elkind
University of Oxford, UK
elkind@cs.ox.ac.uk

Martin Lackner
Vienna University of Technology, Austria
lackner@dbai.tuwien.ac.at

Abstract

Structured preference domains, such as, for example, the domains of single-peaked and single-crossing preferences, are known to admit efficient algorithms for many problems in computational social choice. Some of these algorithms extend to preferences that are close to having the respective structural property, i.e., can be made to enjoy this property by performing minor changes to voters' preferences, such as deleting a small number of voters or candidates. However, it has recently been shown that finding the optimal number of voters or candidates to delete in order to achieve the desired structural property is NP-hard for many such domains. In this paper, we show that these problems admit efficient approximation algorithms. Our results apply to all domains that can be characterized in terms of forbidden configurations; this includes, in particular, single-peaked and single-crossing elections. For a large range of scenarios, our approximation results are optimal under a plausible complexity-theoretic assumption. We also provide parameterized complexity results for this class of problems.

1 Introduction

Collective decision-making plays an important role in the functioning of multi-agent systems. Typically, it is assumed that agents are given a set of alternatives (sometimes also called the *candidates*), and need to select a non-empty subset of this set; each agent's preferences over the candidates are usually represented by a total order over the candidate set. Making a collectively optimal choice in this setting is often a difficult problem, as evidenced by Arrow's classic impossibility result (1951). Therefore, collective decision-making is often studied under the assumption that agents' preferences satisfy additional constraints.

Perhaps the most famous example of a restricted preference domain is the domain of *single-peaked preferences* (Black 1958); other examples include *single-caved preferences* (Inada 1964), *single-crossing preferences* (Mirrlees 1971), *value-restricted preferences* (Sen 1966), and *group-separable preferences* (Inada 1964; 1969). Many of these domains enjoy desirable social choice-theoretic properties, such as transitivity of the majority relation and existence of a strategyproof social choice rule (Barberà and Moreno 2011). Moreover, it has recently been shown that many hard algorithmic problems pertaining to voting and elections become easier if the voters' preferences can be assumed to be single-peaked or single-crossing (Faliszewski et al. 2011; Brandt et al. 2010; Cornaz, Galand, and Spanjaard 2012; 2013; Skowron et al. 2013); it seems plausible that some of these results could be extended to other restricted domains. Further, some of these efficient algorithms can be modified to work for preference profiles that are close to being single-peaked or single-crossing, for an appropriate notion of closeness (Faliszewski, Hemaspaandra, and Hemaspaandra 2014; Cornaz, Galand, and Spanjaard 2012; 2013; Yang and Guo 2014a; 2014b).

Now, suppose that we have a polynomial-time algorithm for some voting-related problem on a restricted domain $\mathcal{D}$. To use this algorithm, we may have to be able to detect whether a given election belongs to $\mathcal{D}$. For commonly studied domains, such as single-peaked and single-crossing preferences this can be done efficiently (Bartholdi and Trick 1986; Escoffier, Lang, and Öztürk 2008; Bredereck, Chen, and Woeginger 2013b; Elkind, Faliszewski, and Slinko 2012). However, it has recently been shown that determining if an election is close to being in a restricted domain is computationally difficult, for many such domains and many notions of closeness. More specifically, Erdelyi et al. (2013) focus on the single-peaked domain, and investigate the complexity of computing the "distance" between a given election and this domain. They consider a variety of distance measures, such as the number of voters or candidates that need to be deleted or the number of candidate swaps that need to be performed to make the election single-peaked, as well as distances that are based on splitting voters or candidates into several groups. In particular, Erdelyi et al. show that finding a minimum-size set of voters to delete in order to make an election single-peaked is NP-hard; in contrast, for candidate deletion this problem is in P. A related paper by Bredereck et al. (2013a) only considers two distance measures, namely, the candidate deletion distance and the voter deletion distance, but explores several restricted preference domains, including single-caved and single-crossing preferences, best-/worst-/medium-/value-restricted preferences and group-separable preferences. It shows that many of the associated computational problems are NP-hard; an important exception is the problem of finding a minimum-size set of voters whose deletion results in a single-crossing election, which admits a polynomial-time algorithm. These NP-hardness results present a difficulty if one wants to use efficient algorithms for nearly structured domains, as some of these algorithms rely on knowing the "distance" to the respective domain.

It is then natural to ask if these hardness results can be circumvented using approximation algorithms and/or parameterized algorithms. The main contribution of our paper is answering this question in the affirmative for the voter deletion distance and the candidate deletion distance, and for a large family of restricted domains. Specifically, our results apply to any restricted domain that can be characterized in terms of *forbidden configurations* (see Section 2); this includes all domains discussed by Bredereck et al. (2013a). We demonstrate that for any such domain $\mathcal{D}$ the problem of finding the smallest number of voters/candidates to delete in order to obtain an election in $\mathcal{D}$ admits an efficient approximation algorithm. To do so, we reduce our problem to the classic HITTING SET problem. The approximation ratio on our algorithm is determined by the size of the largest forbidden configuration used to characterize $\mathcal{D}$, which is typically a small integer. For the voter deletion distance and several restricted domains (including, notably, the single-peaked domain), we can improve the approximation ratio of our algorithm by using a more elaborate reduction to HITTING SET; this approach results in a 2-approximation algorithm. We show that this result is optimal subject to the Unique Games Conjecture (Khot and Regev 2008), which is a well-known complexity-theoretic assumption.

Our reduction to HITTING SET also allows us to use parameterized algorithms for this problem, resulting in FPT algorithms for our problem. For a summary of approximation and FPT results, we refer to Table 1.

For voter deletion, we also consider the setting where we need to delete more than half of the voters. In this case, it is more natural to focus on computing the number of surviving voters. We show that this problem is W[1]-complete, and cannot be approximated within $n^{1-\epsilon}$ unless $\mathsf{P} \neq \mathsf{NP}$.

We omit some proofs due to space constraints.

2 Preliminaries

Given a positive integer s, we write $[s]$ to denote the set $\{1, \dots, s\}$. When discussing fixed-parameter algorithms, we use the standard notation of parameterized complexity, and write $\mathcal{O}^*(f(k))$ as a shorthand for $\mathcal{O}(f(k) \cdot n^{\mathcal{O}(1)})$, i.e., the $\mathcal{O}^*$ notation ignores polynomial factors.

Elections and restricted preference domains. An *election* is described by a set of *candidates* $C = \{c_1, \dots, c_m\}$ and a list of *votes* $V = (v_1, \dots, v_n)$, where each v_i, $i \in [n]$, is a complete order over C; we refer to v_i as the vote, or *preferences*, of voter i, and write $E = (C, V)$. The list of votes V is sometimes called the *preference profile*. If v_i ranks candidate a above candidate b, we write $a >_i b$ or $v_i : ab$. Given a list of votes V', we write $V' \subseteq V$ if V' can be obtained from V by deleting some of the votes; further, given $V' \subseteq V$, we write $V \setminus V'$ to denote the list of votes that can be obtained from V by removing the votes in V'.

In what follows, we discuss restricted preference domains, i.e., sets of elections that satisfy certain properties. The most prominent examples of such domains are single-peaked preferences and single-crossing preferences.

Definition 1. A vote v_i over a candidate set C is said to be *single-peaked with respect to a complete order* $\succ$ *over* C if for every triple of candidates $a, b, c \in C$ such that $a \succ b \succ c$ or $c \succ b \succ a$ it holds that $a >_i b$ implies $b >_i c$. An election $E = (C, V)$ is said to be *single-peaked* if there exists a complete order $\succ$ over C such that every vote in V is single-peaked with respect to $\succ$.

Definition 2. An election $E = (C, V)$, where $V = (v_1, \dots, v_n)$, is said to be *single-crossing with respect to* V if for every pair of candidates $a, b \in C$ such that $a >_1 b$ all voters in V that rank a above b precede all voters in V that rank b above a. Further, $E = (C, V)$ is said to be *single-crossing* if the votes in V can be permuted so that E is single-crossing with respect to the resulting ordering V'.

Our results apply to several other restricted preference domains, including worst-/best-/medium-/value-restricted, single-caved and group-separable preferences. We define the first four of these domains in Section 3. The remaining definitions are omitted, as they are not essential for our presentation; see, e.g., (Bredereck, Chen, and Woeginger 2013a).

3 Configurations

A *condition* on a set of variables $X = \{x_1, \dots, x_t\}$ is a Boolean formula with pairwise comparisons of $x_1, \dots, x_t$ as atoms. For instance, $\phi : x_1 > x_2 \wedge x_3 > x_4$ (or short: $\phi : x_1x_2 \wedge x_3x_4$) is a condition on $\{x_1, x_2, x_3, x_4\}$. Let $||\phi||$ denote the description size of a condition. Since we only consider conditions over domains of small constant size, the representation details do not affect the complexity of our algorithms. A *configuration* is a set of conditions $\Phi = \{\phi_1, \dots, \phi_s\}$, where all ϕ_i, $i \in [s]$, are conditions over the same set of variables. We denote by $s(\Phi)$ the number of conditions in Φ and by $X(\Phi)$ the set of variables that occur in Φ; also, we write $t(\Phi) = |X(\Phi)|$. We refer to a configuration Φ with $s(\Phi) = s$, $t(\Phi) = t$ as an (s, t)-*configuration*.

The following definition plays a central role in this paper.

Definition 3. Given a mapping $\xi : X \to C$ and a condition ϕ over X, let $\xi(\phi)$ denote the Boolean formula obtained by replacing all variables in ϕ according to ξ. We say that a vote v over C *fulfills* ϕ *with respect to* ξ (and write $v \models_\xi \phi$) if v is a model for $\xi(\phi)$. An election $E = (C, V)$ is said to *contain a configuration* $\Phi = \{\phi_1, \dots, \phi_s\}$ with $X(\Phi) = X$ if there exists a mapping $\xi : X \to C$ and s distinct votes $v_{i_1}, \dots, v_{i_s} \in V$ such that $v_{i_j} \models_\xi \phi_j$ for all $j \in [s]$.

Example 1. Consider an election $E = (C, V)$, where $C = \{c_1, c_2, c_3, c_4\}$, $V = (v_1, v_2)$, $v_1 : c_1c_2c_3c_4$, $v_2 : c_4c_1c_2c_3$, and a configuration $\Phi = \{\phi_1, \phi_2\}$, where $\phi_1 : abc$, $\phi_2 : bca$. Then E contains Φ. Indeed, if we set $\xi(a) = c_4$, $\xi(b) = c_1$, $\xi(c) = c_2$, $v_{i_1} = v_2$, $v_{i_2} = v_1$, we get $v_{i_1} \models_\xi \phi_1$, $v_{i_2} \models_\xi \phi_2$.

We will now introduce five configurations that will play an important role in this paper.

Γ	Approximation		FPT runtime for VDEL		FPT runtime for CDEL
	VDEL	CDEL	$k < n/2$	$k \geq n/2$	
Single-peaked / Single-caved	2	P	$\mathcal{O}^*(1.28^k)$	$\mathcal{O}^*(2.08^k)$	P
Single-crossing	P	6	P	P	$\mathcal{O}^*(5.07^k)$
Best-/Medium-/Worst-restricted	2	3	$\mathcal{O}^*(1.28^k)$	$\mathcal{O}^*(2.08^k)$	$\mathcal{O}^*(2.08^k)$
Value-restricted	3	3	$\mathcal{O}^*(2.08^k)$	$\mathcal{O}^*(2.08^k)$	$\mathcal{O}^*(2.08^k)$
Group-separable	2	4	$\mathcal{O}^*(1.28^k)$	$\mathcal{O}^*(2.08^k)$	$\mathcal{O}^*(3.15^k)$

Table 1: Approximation and FPT algorithms for Γ-VDEL and Γ-CDEL.

Definition 4. The *α-configuration* is a $(2, 4)$-configuration Φ_α with conditions

$$\phi_1 : abc \wedge db, \quad \phi_2 : cba \wedge db.$$

The *worst-diverse configuration* is a $(3, 3)$-configuration Φ_W with conditions

$$\phi_1 : ac \wedge bc, \quad \phi_2 : ab \wedge cb, \quad \phi_3 : ba \wedge ca.$$

The *best-diverse configuration* is a $(3, 3)$-configuration Φ_B with conditions

$$\phi_1 : ab \wedge ac, \quad \phi_2 : ba \wedge bc, \quad \phi_3 : ca \wedge cb.$$

The *medium-diverse configuration* is a $(3, 3)$-configuration Φ_M with conditions

$$\phi_1 : bac \vee cab, \quad \phi_2 : abc \vee cba, \quad \phi_3 : acb \vee bca.$$

The *value-diverse configuration* is a $(3, 3)$-configuration Φ_C with conditions

$$\phi_1 : abc, \quad \phi_2 : bca, \quad \phi_3 : cab.$$

An election is said to be *worst-restricted* if it contains no occurrences of Φ_W; *best-restricted*, *medium-restricted*, and *value-restricted* elections are defined similarly.

We will now formulate two simple conditions on configurations.

Definition 5. A configuration Φ is *exact* if every preference order over $X(\Phi)$ fulfills at most one condition in Φ. Further, Φ is *partitioning* if every preference order over $X(\Phi)$ fulfills exactly one condition in Φ.

Observe that Φ_α, Φ_W, Φ_B, Φ_M, and Φ_C are exact configurations; further, Φ_W, Φ_B, and Φ_M are partitioning, but Φ_α and Φ_C are not.

The notion of partitioning configuration will play an important role in Section 5. We will now describe an efficient algorithm for checking whether an election E contains an exact configuration Φ.

Proposition 6. *Given an exact configuration Φ with $s(\Phi) = s$, $t(\Phi) = t$ and an election $E = (C, V)$ with $|C| = m$, $|V| = n$, we can detect whether E contains Φ in time $\mathcal{O}(||\Phi||\, nm^t)$.*

Proof. We can go over all ordered t-tuples of elements of C. Each such tuple can be interpreted as a mapping ξ from $X = X(\Phi)$ to C. For each such mapping, we set $\Phi' = \Phi$ and go over the votes in V one by one. For each vote $v \in V$, we check whether $v \models_\xi \phi_i$ for some $\phi_i \in \Phi'$; this can be done in time $\mathcal{O}(||\Phi||)$. Note that, since Φ is exact, there can be at most one such condition. If $v \models_\xi \phi_i$, we remove ϕ_i from Φ', and repeat this process with the next vote in V. If Φ' becomes empty, we return "yes" and stop. If all votes in V have been processed, but Φ' remains non-empty, we move on to the next mapping $\xi : X \to C$ (and reset $\Phi' = \Phi$). If we have enumerated all mappings $\xi : X \to C$, we stop and output "no". The correctness of this algorithm and the bound on its running time are immediate. □

If Φ is not exact, the algorithm described in the proof of Proposition 6 may fail to work correctly. However, by considering all mappings $\xi : X \to C$ and all ordered s-tuples of voters in V, we can check whether E contains Φ in time $\mathcal{O}(||\Phi||\, n^s m^t)$.

We say that a preference domain $\mathcal{D}$ is *characterized by a set of forbidden configurations* $\Gamma = \{\Phi_1, \ldots, \Phi_\gamma\}$ if for every election E we have $E \in \mathcal{D}$ if and only if E does not contain any of the configurations in Γ.

By definition, the domains of worst-restricted, best-restricted, medium-restricted, and value-restricted elections can be characterized by sets of forbidden configurations that consist of a single $(3, 3)$-configuration each. Moreover, the following results are known.

- The domain of single-peaked preferences is characterized by the set of forbidden configurations $\{\Phi_\alpha, \Phi_W\}$ (Ballester and Haeringer 2011).
- The domain of single-crossing preferences is characterized by a set of forbidden configurations $\{\Phi_\gamma, \Phi_\delta\}$, where Φ_γ is a $(3, 6)$-configuration and Φ_δ is a $(4, 4)$-configuration (Bredereck, Chen, and Woeginger 2013b).
- The domain of single-caved preferences is characterized by a set of forbidden configurations $\{\Phi_{\bar\alpha}, \Phi_B\}$, where $\Phi_{\bar\alpha}$ is a $(2, 4)$-configuration (Ballester and Haeringer 2011).
- The domain of group-separable preferences is characterized by a set of forbidden configurations $\{\Phi_\beta, \Phi_M\}$, where Φ_β is a $(2, 4)$-configuration (Ballester and Haeringer 2011). Each of the configurations $\Phi_{\bar\alpha}$, Φ_β, Φ_γ, Φ_δ is exact.

We set $\Gamma_W = \{\Phi_W\}$, $\Gamma_B = \{\Phi_B\}$, $\Gamma_M = \{\Phi_M\}$, $\Gamma_C = \{\Phi_C\}$, $\Gamma_{sp} = \{\Phi_\alpha, \Phi_W\}$, $\Gamma_{scv} = \{\Phi_{\bar\alpha}, \Phi_B\}$, $\Gamma_{sc} = \{\Phi_\gamma, \Phi_\delta\}$, $\Gamma_{gs} = \{\Phi_\beta, \Phi_M\}$.

We will now define the two families of computational problems that will be the focus of this paper. Both families are parameterized by a set of configurations Γ.

Γ-VDEL	
Instance:	An election $E = (C, V)$
Question:	Find the smallest k such that for some $V' \subseteq V$ with $\|V'\| = k$ the election $(C, V \setminus V')$ contains no configurations from Γ

Γ-CDEL	
Instance:	An election $E = (C, V)$
Question:	Find the smallest k such that for some $C' \subseteq C$ with $\|C'\| = k$ the restriction of E to $C \setminus C'$ contains no configurations from Γ

4 A Simple Conversion to Hitting Set

In this section, we describe a straightforward transformation from Γ-VDEL and Γ-CDEL to the classic HITTING SET problem. We start by defining this problem formally.

d-HITTING SET	
Instance:	A finite set A and a collection $\mathcal{T}$ of subsets of A, where $\|S\| \leq d$ for all $S \in \mathcal{T}$
Question:	Find the smallest k such that there is a set $A' \subseteq A$ with $\|A'\| = k$ satisfying $A' \cap S \neq \emptyset$ for each $S \in \mathcal{T}$

Theorem 7. *Let Γ be a set of exact configurations, let $||\Gamma|| = \sum_{\Phi\in\Gamma} ||\Phi||$, and let $s = \max_{\Phi\in\Gamma} s(\Phi)$, $t = \max_{\Phi\in\Gamma} t(\Phi)$. Then an instance $E = (C, V)$ of* Γ-VDEL *(respectively,* Γ-CDEL*) with $|C| = m$, $|V| = n$ can be reduced to an instance $(A, \mathcal{T})$ of d-*HITTING SET *with $d = s$ (respectively, $d = t$) in time $\mathcal{O}(||\Gamma|| nm^t)$ so that the optimal number of voters (respectively, candidates) to delete in E equals the optimal size of the hitting set for $(A, \mathcal{T})$.*

Proof. We first consider Γ-VDEL. Given an election $E = (C, V)$, we set $A = V$. Further, for each occurrence of a forbidden configuration from Γ in E we add the corresponding set of voters to $\mathcal{T}$. We obtain an instance of d-HITTING SET with $d = \max_{\Phi\in\Gamma} s(\Phi)$. For Γ-CDEL, the reduction is similar: we set $A = C$, and the sets in $\mathcal{T}$ correspond to sets of candidates in occurrences of configurations from Γ in E.

Let $(A, \mathcal{T})$ be the instance of d-HITTING SET produced by our reduction. Suppose that we can eliminate all occurrences of the configurations in Γ from E by deleting a set of voters $V' \subseteq V$. Then V' intersects every set in $\mathcal{T}$, so $(A, \mathcal{T})$ admits a hitting set of size $|V'|$. Conversely, if A' is a hitting set for $(A, \mathcal{T})$, then by deleting the corresponding voters from V we ensure that our election contains no configurations in Γ. A similar argument works for Γ-CDEL.

To implement this reduction, we go over all configurations in Γ, and, for each configuration Φ, detect all occurrences of Φ in E using a modification of the algorithm described in Proposition 6. This establishes the bound on the running time of our reduction. □

This simple conversion enables us to use the techniques developed for d-HITTING SET in order to solve Γ-VDEL and Γ-CDEL whenever all configurations in Γ are exact and $t = \max_{\Phi\in\Gamma} t(\Phi)$ is bounded by a small constant; this is the case for all sets of forbidden configurations considered in this paper. These techniques include, in particular, approximation algorithms and FPT algorithms for d-HITTING SET. However, the running time and/or solution quality of these algorithms often depends on the value of d. Thus, it would be desirable to have a reduction that produces an instance of d-HITTING SET with a smaller value of d. We will now see that this is indeed possible for Γ-VDEL, for several important sets of forbidden configurations Γ, including the one that characterizes single-peaked preferences.

5 An Improved Conversion to Hitting Set

Our improved conversion from Γ-VDEL to d-HITTING SET relies on the notion of a partitioning configuration (Definition 5). Given an election $E = (C, V)$ and a mapping $\xi : X \to C$, a partitioning (s, t)-configuration $\Phi = \{\phi_1, \ldots, \phi_s\}$ with $X(\Phi) = X$ induces a partition of V into s sets $V_1^\xi, \ldots, V_s^\xi$, where $V_i^\xi = \{v \in V \mid v \models_\xi \phi_i\}$ for each $i \in [s]$. Using this observation, for Γ-VDEL we can strengthen Theorem 7 as follows.

Theorem 8. *Let Γ be a set of exact configurations, let $||\Gamma|| = \sum_{\Phi\in\Gamma} ||\Phi||$, and let $s = \max_{\Phi\in\Gamma} s(\Phi)$, $t = \max_{\Phi\in\Gamma} t(\Phi)$, where $s \geq 3$. Suppose also that Γ contains exactly one configuration Φ^+ with $s(\Phi^+) = s$, and this configuration is partitioning. Then, given an instance $E = (C, V)$ of* Γ-VDEL *with $|C| = m$, $|V| = n$, where the optimal solution size is less than $\frac{n}{s-1}$, we can construct s instances of $(s-1)$-*HITTING SET *in time $\mathcal{O}(||\Gamma|| nm^t)$ so that the optimal number of voters to delete in E equals $\min_{i\in[s]} |A_i|$, where A_i is an optimal hitting set for the i-th instance.*

Proof. We construct s instances of $(s-1)$-HITTING SET, denoted by $(A, \mathcal{T}_1), (A, \mathcal{T}_2), \ldots, (A, \mathcal{T}_s)$. We set $A = V$. The sets $\mathcal{T}_1, \ldots, \mathcal{T}_s$ are constructed in three steps.

Step 1. Let $\Phi^+ = \{\phi_1, \ldots, \phi_s\}$ be the unique configuration in Γ with $s(\Phi^+) = s$; let $X = X(\Phi^+)$. As explained above, for every mapping $\xi : X \to C$, Φ^+ defines a partition of V into sets of votes $V_1^\xi, \ldots, V_s^\xi$. Pick a mapping ξ that maximizes the size of the smallest set in $\{V_1^\xi, \ldots, V_s^\xi\}$. For each $i \in [s]$, initialize $\mathcal{T}_i$ by setting $\mathcal{T}_i = \{\{v\} \mid v \in V_i^\xi\}$.

Step 2. Let $V_{-i} = V \setminus V_i^\xi$ for all $i \in [s]$. We will now iterate over all mappings $\xi' : X \to C$ and for every such mapping we consider its induced partition of V_{-i}. We denote the sets in this partition by $V_1^{\xi',i}, \ldots, V_s^{\xi',i}$ and assume without loss of generality that $|V_1^{\xi',i}| \leq \ldots \leq |V_s^{\xi',i}|$. For each tuple $(v_{i_1}, \ldots, v_{i_{s-1}}) \in V_1^{\xi',i} \times \cdots \times V_{s-1}^{\xi',i}$ we add the set $\{v_{i_1}, \ldots, v_{i_{s-1}}\}$ to $\mathcal{T}_i$.

Step 3. It remains to deal with configurations in $\Gamma \setminus \{\Phi^+\}$; by our assumption, we have $s(\Phi) \leq s - 1$ for every $\Phi \in \Gamma \setminus \{\Phi^+\}$. We handle them in the same way as in Theorem 7, i.e., for each $i \in [s]$ and each $\Phi' \in \Gamma \setminus \{\Phi\}$ we add to $\mathcal{T}_i$ all sets of voters that correspond to occurrences of Φ' in E.

This completes the description of our reduction. The bound on its running time is immediate. Also, each $\mathcal{T}_i$, $i \in [s]$, only contains sets of size $s - 1$ or less, i.e., we have constructed s instances of $(s-1)$-HITTING SET. Let

A_i be an optimal solution for $(A, \mathcal{T}_i)$. To complete the proof, we will show that for each $i \in [s]$ (1) removing the voters in A_i from E results in an election that contains no configurations from Γ, and (2) if one can ensure that E contains no configurations from Γ by deleting a set of votes $V' = \{v_{i_1}, \dots, v_{i_k}\}$, $k < \frac{n}{s-1}$, then V' is a hitting set for at least one of the instances $(A, \mathcal{T}_1), \dots, (A, \mathcal{T}_s)$.

To prove the first claim, fix $i \in [s]$ and consider a configuration $\Phi \in \Gamma$. If $\Phi \neq \Phi^+$, the claim is immediate: A_i intersects each set of voters corresponding to an occurrence of Φ in E, so by removing A_i we eliminate all occurrences of Φ. Now, suppose that $\Phi = \Phi^+ = \{\phi_1, \dots, \phi_s\}$. Consider some occurrence of Φ^+ in E; it corresponds to a mapping $\xi' : X \to C$ and a set of votes $\{v_{i_1}, \dots, v_{i_s}\}$, where $v_{i_j} \models_{\xi'} \phi_j$ for $j \in [s]$. If $\xi' = \xi$, then we have $V_i^\xi \subseteq A_i$, and hence no vote in $(C, V \setminus A_i)$ fulfills ϕ_i with respect to ξ'. Now, suppose that $\xi' \neq \xi$. If $v_{i_j} \in V_i^\xi$ for some $j \in [s]$, then $v_{i_j} \in A_i$, and we are done. Otherwise we have $v_{i_j} \in V_j^{\xi',i}$ for all $j \in [s]$. But then the set $\{v_{i_j} \mid 1 \leq j \leq s-1\}$ belongs to $\mathcal{T}_i$, and therefore A_i intersects it. This completes the proof of our first claim.

To prove the second claim, consider a set of votes $V' = \{v_{i_1}, \dots, v_{i_k}\}$ such that $E' = (C, V \setminus V')$ contains no configurations from Γ. Note first that V' has to contain at least one of $V_1^\xi, \dots, V_s^\xi$; indeed, if $V \setminus V'$ intersects each of $V_1^\xi, \dots, V_s^\xi$, the votes in the intersection would correspond to an occurrence of Φ^+. Thus, suppose that $V_i^\xi \subseteq V'$ for some $i \in [s]$. We will now argue that V' is a hitting set for $(A, \mathcal{T}_i)$. Consider a set $S \in \mathcal{T}_i$. If S is a singleton that has been added to $\mathcal{T}_i$ during the first step, then we are done, since $S = \{v_j\}$ for some $v_j \in V_i^\xi$, and $V_i^\xi \subseteq V'$. If S corresponds to an occurrence of a configuration in $\Gamma \setminus \{\Phi^+\}$, we are done, too, since V' hits all occurrences of this configuration.

Finally, suppose that $S = \{v_{j_1}, \dots, v_{j_{s-1}}\}$ where $v_{j_\ell} \in V_\ell^{\xi',i}$ for all $\ell \in [s-1]$. We will now argue that if $V' \cap S = \emptyset$, then $|V'| \geq \frac{n}{s-1}$. To see this, suppose that $V' \cap S = \emptyset$. Then $V_s^{\xi',i} \subseteq V'$. Indeed, if this is not the case, consider a vote $v_j \in V_s^{\xi',i}$. All of the votes in $S \cup \{v_j\}$ are present in $E' = (C, V \setminus V')$, and hence E' contains an occurrence of Φ^+, a contradiction. As we also have $V_i^\xi \subseteq V'$ for some $i \in [s]$, and $V_s^{\xi',i} \cap V_i^\xi = \emptyset$, it follows that $|V'| \geq |V_s^{\xi',i}| + |V_i^\xi|$. It remains to prove that $|V_s^{\xi',i}| + |V_i^\xi| \geq \frac{n}{s-1}$.

To establish this, let $y = |V_i^\xi|$ and $z_j = |V_j^{\xi',i}|$ for $j \in [s]$; we need to show that $y + z_s \geq \frac{n}{s-1}$. We have $z_1 \leq z_2 \leq \cdots \leq z_s$, and hence $z_s \geq \frac{1}{s-1}(z_2 + \cdots + z_s)$ Further, by our choice of ξ we have $y \geq z_1$ and therefore $z_2 + \cdots + z_s = n - y - z_1 \geq n - 2y$. Thus, we obtain

$$y + z_s \geq y + \frac{n-2y}{s-1} \geq \frac{n}{s-1} + y\frac{s-3}{s-1} \geq \frac{n}{s-1},$$

where the last inequality follows since we assume $s \geq 3$.

We have argued that if V' fails to intersect some set in $\mathcal{T}_i$, then $|V'| \geq \frac{n}{s-1}$. This completes the proof. □

The constraint on the true size of the optimal solution for Γ-VDEL in Theorem 8 may appear to be significant (and difficult to check). However, in Section 6 we will see that it does not affect our ability to design efficient approximation algorithms for Γ-VDEL.

Further, we remark that Theorem 8 only provides an improved reduction for Γ-VDEL, and not for Γ-CDEL. This is because there is no direct analogue to the notion of a partitioning configuration for the latter problem.

Optimality of the Conversion

We will now show that the improved conversion is optimal, by providing a reduction from d-HITTING SET to Γ-VDEL and thus establishing equivalence between these two problems. To make the theorem as widely applicable as possible, we introduce the notion of a *solid subconfiguration*.

Definition 9. Consider a configuration Φ with $X = X(\Phi)$, and a subset X' of X with $|X'| \geq 2$. Let $\Phi[X']$ be the restriction of Φ to X'. We say that $\Phi[X']$ is a *solid subconfiguration* if for every $x, y \in X'$ there exists a condition $\phi \in \Phi$ such that ϕ implies $x > y$.

Example 2. Consider the configuration Φ_α. The subconfiguration $\Phi_\alpha[\{a, b, c\}]$ is solid, whereas $\alpha[\{a, b, d\}]$ is not solid since neither ϕ_1 nor ϕ_2 implies $b > d$.

Theorem 10. *Consider a set of configurations Γ and a configuration $\Phi \in \Gamma$. Suppose that there exists an election $E = (C, V)$ with $V = (v_1, \dots, v_r)$ such that $r \geq 3$ and*

1. *E contains Φ;*
2. *for every $\Phi' \in \Gamma$ there exists a solid subconfiguration of Φ' such that for every $i \in [r-1]$ the election $(C, V \setminus \{v_i\})$ does not contain this solid subconfiguration.*

Then there exist a polynomial-time reduction from $(r-1)$-HITTING SET to Γ-VDEL that is approximation-preserving.

The conditions of Theorem 10 are satisfied by $\Gamma \in \{\Gamma_W, \Gamma_B, \Gamma_M, \Gamma_{sp}, \Gamma_{scv}, \Gamma_{gs}\}$ with $r = 3$ and by Γ_C with $r = 4$. However, they are not satisfied by Γ_{sc} (for any r). This is not surprising since Γ_{sc}-VDEL is in P, and thus a reduction from HITTING SET would imply P = NP.

We will now illustrate how these conditions are satisfied for $r = 3$ and Γ_{sp}. Consider the configuration Φ_α and the election $E = (C, V)$, where $C = \{a, b, c, d\}$, $V = (v_1, v_2, v_3)$, $v_1 : dabc$, $v_2 : dcba$, $v_3 : dacb$. This election satisfies the conditions of Theorem 10. Indeed, it contains Φ_α in the first two votes. If one of these two votes is deleted, the resulting election no longer contains the solid subconfiguration $\Phi_\alpha[\{a, b, c\}]$ (see Example 2). Further, Φ_W is a solid subconfiguration by itself, and it can be eliminated by removing any of the three votes. Thus, 2-HITTING SET admits an approximation-preserving reduction to Γ_{sp}-VDEL.

Finally, let us remark that, while Theorem 10 works for $r = 4$ and Γ_C, it does not work for $r = 4$ and Γ_W, Γ_B, or Γ_M. The reason is that Γ_W, Γ_B, and Γ_M are partitioning, and this can be shown to imply that the second condition of Theorem 10 does not hold for $r = 4$.

6 Approximation Algorithms

We now present the first application of our reductions. Since d-HITTING SET allows for a factor-d approximation (Even

d	2	3	4	5	6
c_d	1.28	2.08	3.15	4.11	5.07

Table 2: Currently best algorithms for d-HITTING SET with a runtime of $\mathcal{O}^*({c_d}^k)$, where k is the optimal solution size

and Bar-Yehuda 1981), we are able to approximate Γ-VDEL and Γ-CDEL up to a constant factor for all sets of forbidden configurations Γ considered in Section 3.

Theorem 11. *Let Γ be a set of configurations and let $s = \max_{\Phi\in\Gamma} s(\Phi)$, $t = \max_{\Phi\in\Gamma} t(\Phi)$. Then Γ-VDEL admits a polynomial-time s-approximation algorithm, and Γ-CDEL admits a polynomial-time t-approximation algorithm. Moreover, if Γ contains a unique configuration Φ with $s(\Phi) = s$, $s \geq 3$, and this configuration is partitioning, then Γ-VDEL admits an $(s-1)$-approximation algorithm.*

Proof. The first claim follows immediately from Theorem 7 and the fact that d-HITTING SET admits a polynomial-time d-approximation algorithm. Now, suppose that Γ contains a unique configuration Φ with $s(\Phi) = s$, and Φ is partitioning. We then use the reduction described in the proof of Theorem 8, and obtain s instances of $(s-1)$-HITTING SET. We run the $(s-1)$-approximation algorithm for $(s-1)$-HITTING SET, and obtain s sets $A_1, \ldots, A_s$. We return the set of voters that corresponds to the smallest of these sets.

To see why this approach is correct, observe first that by Theorem 8 each of the sets $A_1, \ldots, A_s$ corresponds to a feasible solution to our instance of Γ-VDEL. Now, let k be the size of the optimal solution for our instance of Γ-VDEL. If $k < \frac{n}{s-1}$, then by Theorem 8 one of our instances of $(s-1)$-HITTING SET has a hitting set of size k, so $\min_{i\in[s]} |A_i| \leq (s-1)k$. Otherwise we have $(s-1)k \geq n$, so even the solution that deletes all voters (and hence any of the sets A_i) is within a factor of $(s-1)$ from optimal. □

Corollary 12. *For $\Gamma \in \{\Gamma_W, \Gamma_B, \Gamma_M, \Gamma_{sp}, \Gamma_{scv}, \Gamma_{gs}\}$, the problem Γ-VDEL can be approximated within a factor of 2, and Γ_C-VDEL can be approximated within a factor of 3. Moreover, the problem Γ-CDEL can be approximated within a factor of 3 for $\Gamma \in \{\Gamma_W, \Gamma_B, \Gamma_M, \Gamma_C\}$, within a factor of 4 for $\Gamma = \Gamma_{gs}$, and within a factor of 6 for $\Gamma = \Gamma_{sc}$.*

The reduction in the proof of Theorem 10 is approximation preserving, and it is known that a d-approximation of d-HITTING SET is optimal under the assumption that the Unique Games Conjecture holds (Khot and Regev 2008). Thus, we immediately obtain the following result.

Corollary 13. *Assuming the Unique Games Conjecture, the approximation results for Γ-VDEL in Table 1 are optimal.*

7 Fixed-Parameter Algorithms

Fixed-parameter algorithms for Γ-VDEL can be obtained by utilizing FPT algorithms for d-HITTING SET (Chen, Kanj, and Xia 2010; Wahlström 2007; Fernau 2010). The currently best runtimes for d-HITTING SET are displayed in Table 2.

Theorem 14. *Let Γ be a set of configurations, and let $s = \max_{\Phi\in\Gamma} s(\Phi)$, $t = \max_{\Phi\in\Gamma} t(\Phi)$. Then Γ-VDEL can be solved in time $\mathcal{O}^*({c_s}^k)$, and Γ-CDEL can be solved in time $\mathcal{O}^*({c_t}^k)$, where k is the size of the optimal solution and c_s is taken from Table 2. Moreover, if $k < n/2$, Γ contains a unique configuration Φ with $s(\Phi) = s$, and Φ is partitioning, then Γ-VDEL can be solved in time $\mathcal{O}^*(c_{s-1}^k)$, where k is the size of the optimal solution.*

8 Deleting Almost All Votes

The approximation algorithm described in Section 6 is useful when the size of the optimal solution for Γ-VDEL does not exceed $n/2$. However, it may also be the case that, to eliminate configurations in Γ, we need to delete almost all voters. In this case, it is trivial to find a 2-approximate solution to Γ-VDEL: simply deleting all voters provides a 2-approximation. Thus, a more fine-grained approach is to try to approximate the number of *surviving* voters; we refer to this variant of our problem as Γ-VDEL$^-$. It turns out that Γ-VDEL$^-$ is hard to approximate for many sets Γ.

Theorem 15. *Consider a set of configurations Γ and a configuration $\Phi \in \Gamma$. Suppose that there exists an election $E = (C, V)$ with $V = (v_1, \ldots, v_n)$ such that $n \geq 3$ and*

1. *E contains Φ;*
2. *for every $\Phi' \in \Gamma$ there exists a solid subconfiguration of Φ' such that for every $i \in [n-1]$ the election $(C, V \setminus \{v_i\})$ does not contain this solid subconfiguration.*

Then there exists a polynomial-time reduction from INDEPENDENT SET to Γ VDEL$^-$ that is approximation-preserving.

Since INDEPENDENT SET cannot be approximated within $n^{1-\epsilon}$ unless P = NP (Håstad 1999; Zuckerman 2006), we obtain the following corollary.

Corollary 16. *For $\Gamma \in \{\Gamma_W, \Gamma_B, \Gamma_M, \Gamma_C, \Gamma_{sp}, \Gamma_{scv}, \Gamma_{gs}\}$, Γ-VDEL$^-$ cannot be approximated within $n^{1-\epsilon}$ unless P = NP.*

Finally, we characterize the parameterized complexity of Γ-VDEL$^-$.

Theorem 17. *Γ-VDEL$^-$ parameterized by the size of the optimal solution is W[1]-complete.*

9 Conclusions

We have investigated the complexity of approximating the distance between a given election and a restricted preference domain, for two natural distance measures and many well-known restricted preference domains. Our results are broadly positive: they include polynomial-time approximation algorithms whose approximation ratio is bounded by a small constant and reasonably fast FPT algorithms. The reader may wonder if improving the approximation ratio of our algorithms, e.g., for Γ_{sp}-VDEL from 3 to 2, by going through a more complicated reduction was worth the effort. Observe, however, that the running time of the algorithms for nearly single-peaked elections typically scales exponentially (or faster!) with the distance from the single-peaked domain (Faliszewski, Hemaspaandra, and Hemaspaandra 2014); thus, a constant-factor improvement in approximation ratios translates into significant improvement in the running time of these algorithms.

Acknowledgements

Part of this work was done while the first author was employed by Nanyang Technological University, Singapore, and supported by National Research Foundation (Singapore) grant NRF2009-08. The second author was supported by the Austrian Science Fund (FWF): P25518-N23 and by the Vienna Science and Technology Fund (WWTF) project ICT12-15.

References

Arrow, K. 1951. *Social Choice and Individual Values*. John Wiley and Sons.

Ballester, M., and Haeringer, G. 2011. A characterization of the single-peaked domain. *Social Choice and Welfare* 36(2):305–322.

Barberà, S., and Moreno, B. 2011. Top monotonicity: A common root for single peakedness, single crossing and the median voter result. *Games and Economic Behavior* 73(2):345–359.

Bartholdi, III, J., and Trick, M. 1986. Stable matching with preferences derived from a psychological model. *Operations Research Letters* 5(4):165–169.

Black, D. 1958. *The Theory of Committees and Elections*. Cambridge University Press.

Brandt, F.; Brill, M.; Hemaspaandra, E.; and Hemaspaandra, L. 2010. Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 715–722.

Bredereck, R.; Chen, J.; and Woeginger, G. 2013a. Are there any nicely structured preference profiles nearby? In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 62–68.

Bredereck, R.; Chen, J.; and Woeginger, G. 2013b. A characterization of the single-crossing domain. *Social Choice and Welfare* 41(4):989–998.

Chen, J.; Kanj, I. A.; and Xia, G. 2010. Improved upper bounds for vertex cover. *Theor. Comput. Sci.* 411(40-42):3736–3756.

Cornaz, D.; Galand, L.; and Spanjaard, O. 2012. Bounded single-peaked width and proportional representation. In *Proceedings of the 20th European Conference on Artificial Intelligence*, 270–275.

Cornaz, D.; Galand, L.; and Spanjaard, O. 2013. Kemeny elections with bounded single-peaked or single-crossing width. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 76–82.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2012. Clone structures in voters' preferences. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 496–513.

Erdélyi, G.; Lackner, M.; and Pfandler, A. 2013. Computational aspects of nearly single-peaked electorates. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*.

Escoffier, B.; Lang, J.; and Öztürk, M. 2008. Single-peaked consistency and its complexity. In *Proceedings of the 18th European Conference on Artificial Intelligence*, 366–370.

Even, S., and Bar-Yehuda, R. 1981. A linear-time approximation algorithm for the weighted vertex cover problem. *Journal of Algorithms* 2(2):198–203.

Faliszewski, P.; Hemaspaandra, E.; Hemaspaandra, L.; and Rothe, J. 2011. The shield that never was: Societies with single-peaked preferences are more open to manipulation and control. *Information and Computation* 209(2):89–107.

Faliszewski, P.; Hemaspaandra, E.; and Hemaspaandra, L. 2014. The complexity of manipulative attacks in nearly single-peaked electorates. *Artificial Intelligence* 207:69–99.

Fernau, H. 2010. Parameterized algorithms for d-hitting set: The weighted case. *Theoretical Computer Science* 411(1618):1698 – 1713.

Håstad, J. 1999. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica* 182(1):105–142.

Inada, K. 1964. A note on the simple majority decision rule. *Econometrica* 32(4):525–531.

Inada, K. 1969. The simple majority decision rule. *Econometrica* 37(3):490–506.

Khot, S., and Regev, O. 2008. Vertex cover might be hard to approximate to within $2 - \epsilon$. *Journal of Computer and System Sciences* 74(3):335 – 349.

Mirrlees, J. 1971. An exploration in the theory of optimal income taxation. *Review of Economic Studies* 38:175–208.

Sen, A. K. 1966. A possibility theorem on majority decisions. *Econometrica* 34:491–499.

Skowron, P.; Yu, L.; Faliszewski, P.; and Elkind, E. 2013. The complexity of fully proportional representation for single-crossing electorates. In *Proceedings of the 6th International Symposium on Algorithmic Game Theory*, 1–12.

Wahlström, M. 2007. *Algorithms, measures, and upper bounds for satisfiability and related problems*. PhD Thesis, Linköping University.

Yang, Y., and Guo, J. 2014a. The control complexity of r-approval: from the single-peaked case to the general case. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*. To appear.

Yang, Y., and Guo, J. 2014b. Election controls with small single-peaked width. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*. To appear.

Zuckerman, D. 2006. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 681–690.