On the complexity of Chamberlin-Courant on Nearly Structured Profiles

Neeldhara Misra[1], Chinmay Sonar[1], and P. R. Vaidyanathan[1]

Indian Institute of Technology, Gandhinagar, Gandhinagar 382355, India
`neeldhara.m|sonar.chinmay|pr.vaidyanathan@iitgn.ac.in`

Abstract. The Chamberlin-Courant voting rule is an important multi-winner voting rule. Although NP-hard to compute on general profiles, it is known to be polynomially solvable on single-crossing and single-peaked electorates by exploiting the structures of these domains. We consider the problem of generalizing the domain on which the voting rule admits efficient algorithms.

On the one hand, we show efficient algorithms on profiles that are k candidates or k voters away from the single-peaked and single-crossing domains. In particular, for profiles that are k candidates away from being single-peaked or single-crossing, we show algorithms whose running time is FPT in k. For profiles that are k voters away from being single-peaked or single-crossing, our algorithms are XP in k. These algorithms are obtained by a careful extension of known algorithms on structured profiles [10,2]. This provides a natural application for the work by Elkind and Lackner in [8], who study the problem of finding deletion sets to single-peaked and single-crossing profiles.

In contrast to these results, for a different, but equally natural way of generalizing these domain, we show severe intractability results. In particular, we show that the problem is NP-hard on profiles that can be "decomposed" into a constant number of single-peaked profiles. Also, if the number of crossings per pair of candidates in a profile is permitted to be at most three (instead of one), the problem continues be NP-hard. This stands in contrast with other attempts at generalizing these domains (such as single-peaked or single-crossing width), as it rules out the possibility of fixed-parameter (or even XP) algorithms when parameterized by the number of peaks, or the maximum number of crossings per candidate pair.

Keywords: NP-hardness, Chamberlin Courant, Single-Crossing Profiles, Single-Peaked Profiles, fixed-parameter algorithms, voting rules

1 Introduction

A traditional election setting consists of voters expressing their preferences over alternatives, where preferences can be modelled in several ways (approval ballots, ternary ballots, top-truncated lists, total orders, and so forth). Usually, given such a scenario, we would like to identify a winning alternative. In many applications, however, we need to identify not one, but a fixed *set of alternatives* that best represent the interests of the voters. Such a problem arises in a variety of scenarios like committee selection, parliamentary elections, movie recommendation systems, and so forth.

There are several ways of measuring how well a committee fares against a set of votes. When votes are approval ballots, for instance, the maximum or the sum of Hamming distances is often used as a measure of quality. We consider the setting of votes given as complete rankings, and focus on the well-studied Chamberlin-Courant rule [6], which achieves proportional representation. The way this voting rule works is the following. We begin by fixing a notion of a "dissatisfaction function" $\alpha : \mathbb{N} \to \mathbb{N}$, which simply specifies, by $\alpha(i)$, how unhappy a voter is when she is represented by a candidate who is ranked at the i^{th} position on her list. Given a committee with k candidates, a voter is represented by the candidate that she ranks the highest among candidates from X. If $\phi(v)$ denotes the candidate that is representing voter v, the optimal committee under the Chamberlin-Courant voting rule seeks to minimize either the sum or the maximum value of $\alpha(\text{pos}_v(\phi(v)))$, taken over all voters v (where $\text{pos}_v(c)$ denotes the ranking of the candidate c in the vote v).

The Chamberlin-Courant rule (and the closely related Monroe voting rule which we do not consider in the present work) has several desirable properties. It has been argued [10] that rules that achieve proportional representation are particularly well-suited for electing committees that need to make unanimous decisions, and in particular, that takes minority candidates into account. However, it turns out that finding an optimal committee under this rule is NP-hard, and it is therefore unlikely to admit an efficient algorithm.

On the other hand, there have been promising developments showing that the optimal Chamberlin-Courant committees can be computed efficiently on structured profiles which are commonly encountered in practical scenarios. Two such restrictions that have been particularly successful are the single-peaked and single-crossing domains. In a parallel development, [8] showed various efficient algorithms for detecting profiles that are close to being structured (that is, they admit the structure on all but a small number of candidates or voters). We combine these scenarios to address the following question: how well do the efficient algorithms on the restricted domains extend to profiles that are of the latter type, that is, they exhibit the properties of the domain on all but a small number of candidates or voters? We now turn to our findings in the context of this question and closely related issues.

Our Contributions and Methodology. A natural framework for addressing the problem of how well algorithms on structured domains scale up to nearly-structured ones is parameterized complexity. Readers are referred to [7] for a comprehensive introduction to this approach. To begin with, we show efficient algorithms on profiles that k candidates or k voters away from the single-peaked and single-crossing domains. In particular, for profiles that are k candidates away from being single-peaked or single-crossing, we show algorithms whose running time is FPT in k. For profiles that are k voters away from being single-peaked or single-crossing, our algorithms are XP in k. These algorithms are obtained by a careful extension of the the known algorithms [2,10] on the structured profiles. This provides a natural application for the work by Elkind and Lackner in [8], who study the problem of finding deletion sets to single-peaked and single-crossing profiles.

In contrast to these results, for a different, but equally natural way of generalizing these domain, we show severe intractability results. In particular, we show that the problem is NP-hard on profiles that can be "decomposed" into a constant number of single-peaked profiles. Also, if the number of crossings per pair of candidates in a profile is permitted to be at most three (instead of one), the problem continues be NP-hard. This stands in contrast with other attempts at generalizing these domains (such as single-peaked or single-crossing width [10]), as it rules out the possibility of fixed-parameter (or even XP) algorithms when parameterized by the number of peaks, or the maximum number of crossings per candidate pair.

Related Work. Our work builds primarily on two lines of work from before. We appeal to the known algorithms that determine the optimal Chamberlin-Courant committees on single-peaked profiles [2] and single-crossing profiles [10]. These results have been be extended to other multiwinner voting rules, which we do not consider in the present work. Also, efficient algorithms have been shown on more general preference restrictions such as single-peakedness on trees, or single-crossing width.

2 Technical Preliminaries

In this section, we introduce some of the notation and definitions that we will use. For a more detailed introduction to notions relating to restricted domains and voting rules, we refer the reader to the appropriate chapters in [4], and for a comprehensive introduction to parameterized algorithms, we refer the reader to [7].

For a positive integer ℓ, we denote the set $\{1, \ldots, \ell\}$ by $[\ell]$. We first define some general notions relating to voting rules. Let $V = \{v_i : i \in [n]\}$ be a set of n *voters* and $C = \{c_j : j \in [m]\}$ be a set of m *candidates*. If not mentioned otherwise, we denote the set of candidates, the set of voters, the number of candidates, and the number of voters by C, V, m, and n respectively.

Every voter v_i has a *preference* $\succ_i$ which is a complete order over the set C of candidates. We say voter v_i prefers a candidate $x \in C$ over another candidate $y \in C$ if $x \succ_i y$. We denote the set of all preferences over C by $\mathcal{L}(C)$. The n-tuple $(\succ_i)_{i\in[n]} \in \mathcal{L}(C)^n$ of the preferences of all the voters is called a *profile*. Note that a profile, in general, is a multiset of linear orders. For a subset $M \subseteq [n]$, we call $(\succ_i)_{i\in M}$ a sub-profile of $(\succ_i)_{i\in[n]}$. For a subset of candidates $D \subseteq C$, we use $\mathcal{P}|_D$ to denote the projection of the profile on the candidates in D alone. A *domain* is a set of profiles.

The rest of this section is organized as follows. We first define the Chamberlin-Courant voting rule. We then introduce the domain restrictions that are of interest to us, and the notion of closeness to a restricted domain. We finally define the problems that we will study subsequently.

Chamberlin-Courant. The Chamberlin–Courant voting rule is based on the notion of a *dissatisfaction function* or a *misrepresentation function*. This function specifies, for each $i \in [m]$, a voter's dissatisfaction from being represented by candidate she ranks in position i.

Definition 1. *For an* m*-candidate election, a* dissatisfaction function *is given by a non-decreasing function* $\alpha\colon [m] \to \mathbb{N}$ *with* $\alpha(1) = 0$.

A popular dissatisfaction function is Borda, given by $\alpha_B^m(i) = \alpha_B(i) = i - 1$. We now turn to the notion of an assignment function. Let k be a positive integer. A k-*CC-assignment function* for an election $E = (C, V)$ is a mapping $\Phi\colon V \to C$ such that $\|\Phi(V)\| \leqslant k$. For a given assignment function Φ, we say that voter $v \in V$ is *represented* by candidate $\Phi(v)$ in the chosen committee. There are several ways to measure the quality of an assignment function Φ with respect to a dissatisfaction function α; we use the following two:

1. $\ell_1(\Phi) = \sum_{i=1,\ldots,n} \alpha(\mathrm{pos}_{v_i}(\Phi(v_i)))$, and
2. $\ell_\infty(\Phi) = \max_{i=1,\ldots,n} \alpha(\mathrm{pos}_{v_i}(\Phi(v_i)))$.

We are now ready to define the Chamberlin-Courant voting rule, which is the primary focus of this paper.

Definition 2. *For every family of dissatisfaction functions* $\alpha = (\alpha^m)_{m=1}^\infty$*, and every* $\ell \in \{\ell_1, \ell_\infty\}$*, the* α-ℓ-CC voting rule *is a mapping that takes an election* $E = (C, V)$ *and a positive integer* k *with* $k \leqslant \|C\|$ *as its input, and returns a* k*-CC-assignment function* Φ *for* E *that minimizes* $\ell(\Phi)$ *(if there are several optimal assignments, the rule is free to return any of them).*

Chamberlin and Courant [6] originally proposed the utilitarian variants of their rules with a focus on the Borda dissatisfaction function. The egalitarian variant was considered by, for instance, Betzler et al. [2].

single-peaked Profiles. A preference profile is said be single-peaked if there exists an ordering σ over the candidates C such that the preference of every voter v has the following structure: v has a favorite candidate c (sometimes called the "peak" for v), and the further away a candidate $d \neq c$ is from c in σ, the less it is preferred by the voter v. The notion of single-peaked preferences was introduced by Black [3] and a formal definition is as follows.

Definition 3 (single-peaked Domain). *A preference* $\succ \in \mathcal{L}(C)$ *over a set of candidates* C *is called* single-peaked *with respect to an order* $\succ' \in \mathcal{L}(C)$ *if, for every pair of candidates* $x, y \in C$*, we have* $x \succ y$ *whenever we have either* $c \succ' x \succ' y$ *or* $y \succ' x \succ' c$*, where* $c \in C$ *is the candidate at the first position of* $\succ$*. A profile* $\mathcal{P} = (\succ_i)_{i \in [n]}$ *is called single-peaked with respect to an order* $\succ' \in \mathcal{L}(C)$ *if* $\succ_i$ *is single-peaked with respect to* $\succ'$ *for every* $i \in [n]$.

We now turn to the definition of a k-composite single-peaked profile, which is a natural generalization of the single-peaked notion above. We say that a profile is k-composite single-peaked if there is an ordering of the candidates σ and a partition of the candidate set into at most k parts such that each part induces a single-peaked profile on σ restricted to that part. We note, importantly, that this is different from the more well-studied notion of multipeaked profiles, where we have the additional constraint that the k parts have to additionally form intervals on a fixed global ordering. A similar notion called k-additional axis where the votes(rather than the candidates) are divided into k buckets and each bucket is single-peaked, has been studied in [9].

single-crossing Profiles. A preference profile is said to belong to the single-crossing domain if it admits a permutation of the voters such that for any pair of candidates a and b, there is an index $j\langle a, b\rangle$ such that either all voters v_j with $j < j\langle a, b\rangle$ prefer a over b and all voters v_j with $j > j\langle a, b\rangle$ prefer b over a, or vice versa. The formal definition is as follows.

Definition 4 (single-crossing Domain). *A profile $\mathcal{P} = (\succ_i)_{i\in[n]}$ of n preferences over a set C of candidates is called a* single-crossing profile *if there exists a permutation σ of $[n]$ such that, for every pair of distinct candidates $x, y \in C$, whenever we have $x \succ_{\sigma(i)} y$ and $x \succ_{\sigma(j)} y$ for two integers i and j with $1 \leqslant \sigma(i) < \sigma(j) \leqslant n$, we have $x \succ_{\sigma(k)} y$ for every $\sigma(i) \leqslant k \leqslant \sigma(j)$.*

As we did with single-peaked profiles, we generalize the notion of single-crossing domains to r-single-crossing domains in the following natural way: for every pair of candidates (a, b), instead of demanding one index where the preferences "switch" from one way to the other, we allow for r such switches. More formally, a profile is r-single-crossing if for every pair of candidates a and b, there exist r indices $j_0\langle a, b\rangle, j_1\langle a, b\rangle, \ldots, j_r\langle a, b\rangle, j_{r+1}\langle a, b\rangle$ with $j_0\langle a, b\rangle = 1$ and $j_{r+1}\langle a, b\rangle = n + 1$, such that for all $1 \leqslant i \leqslant r + 1$, all voters v_j with $j_i\langle a, b\rangle \leqslant j < j_{i+1}\langle a, b\rangle$ are unanimous in their preferences over a and b.

Nearly Structured Domains. Let $\mathcal{D} = \{\text{SP, SC}\}$ be a fixed domain, where SP refers to single-peaked domains, and SP denotes single-crossing domains. We say that a profile $\mathcal{P}$ over candidates C has a candidate (voter) modulator of size k to $\mathcal{D}$ if there exists a subset of at most k candidates (voters) such that the restriction of the profile to all but the chosen candidates (voters) belongs to the domain $\mathcal{D}$. Whenever a profile admits a k-sized candidate modulator to $\mathcal{D}$, we say that it is k-*close to $\mathcal{D}$ via candidates.* The notion of being k-*close to $\mathcal{D}$ via voters* is analogously defined.

The work of [9], ([5]) shows that it is polynomial-time to find the smallest candidate (voter) modulator to the domain of single-peaked (single-crossing) profiles respectively. While [8] showed 2-approximation and 6-approximation algorithms for finding the smallest voter and candidate modulator to the domains of single-peaked and single-crossing profiles, respectively. Therefore, in all our problem formulations, we assume that we are given an instance of an election with a modulator to either domain as a part of the input — since it is tractable to find such modulators in all cases.

Parameterized Complexity. A parameterized problem is denoted by a pair $(Q, k) \subseteq \Sigma^* \times \mathbb{N}$. The first component Q is a classical language, and the number k is called the parameter. Such a problem is *fixed–parameter tractable* (FPT) if there exists an algorithm that decides it in time $O(f(k)n^{O(1)})$ on instances of size n. On the other hand, a problem is said to belong to the class XP if there exists an algorithm that decides it in time $n^{O(f(k))}$ on instances of size n. We refer the reader to [7] for a more detailed introduction to the parameterized paradigm.

Problem Definition. We now define the main problem that we address in this work, which we denote by $\ell, \mathcal{D}$-CC Via χ, where ℓ is an aggregation function, $\mathcal{D}$ is a domain and χ is either candidates or voters, referring to the type of the modulator we are given as a part of the input.

> $\ell, \mathcal{D}$-CC Via χ **Parameter:** k
> **Input:** An election $E = (C, V)$, a committee size b, a target misrepresentation score R, a misrepresentation function α, and a k-sized χ modulator X to the domain $\mathcal{D}$.
> **Question:** Is there a committee of size b whose ℓ-misrepresentation score under the function α is at most R?

3 Tractability on Nearly Structured Preference Profiles

The goal of this section is to establish the following theorem.

Theorem 1. *For all $\ell \in \{\ell_1, \ell_\infty\}$ and for all $\mathcal{D} \in \{SP, SC\}$, the $(\ell, \mathcal{D})$-CC Via Candidates problem is in* FPT *and the $(\ell, \mathcal{D})$-CC Via Voters problem is in* XP.

We describe now informally our overall approach for solving the $(\ell, \mathcal{D})$-CC Via χ problem. First, we brute force through all possible "behaviors" of the solution on the modulator. Next, instead of solving the "vanilla" Chamberlin-Courant optimization problem on the part of the profile that is structured (according to the domain $\mathcal{D}$), we adapt our solution to account for the guessed behavior on the modulator. For ease of presentation, we define an intermediate auxiliary problem, which is an extension version of the original problem, described below.

In the extension problem corresponding to $(\ell, \mathcal{D})$, we are given, as usual, an election $E = (C, V)$, a committee size k, a target misrepresentation score R and a misrepresentation function α. In addition, we are also given a subset of candidates X and a partition of X into G and B. The promise is that the election induced by the votes V when restricted to the candidates $C \setminus X$ is structured according to the domain $\mathcal{D}$. The goal is to find an optimal Chamberlin-Courant committee among the ones that contain all candidates in G and contain none of the candidates in B. The formal definition is as follows. In the following, we say that a committee respects a partition $(D \uplus G \uplus B)$ of the candidate set C if it contains all of G and none of B.

> $(\ell, \mathcal{D})$-CC Extension
> **Input:** An election $E = (C, V)$, a partition of the candidates into $(D \uplus G \uplus B)$, a committee size b, a target misrepresentation score R, a misrepresentation function α; such that the election induced by (D, V) belongs to the domain $\mathcal{D}$.
> **Question:** Is there a committee of size b that respects $(D \uplus G \uplus B)$ and whose ℓ-misrepresentation score under the function α is at most R?

Before describing how to solve the $(\ell, \mathcal{D})$-CC Extension problem, we first establish that it is indeed useful for solving the $(\ell, \mathcal{D})$-CC Via χ problem. Let $\mathcal{D}$ be a fixed domain from {Single-Peaked, Single-Crossing}. First, consider the $(\ell, \mathcal{D})$-CC Via χ problem where we are given a k-sized candidate modulator as input, or that χ is fixed to be candidates. Let $(E = (C, V), b, R, \alpha, X)$, denoted by $\mathcal{I}$, be an instance of $(\ell, \mathcal{D})$-CC Via χ. Recall that X is a candidate modulator to the domain $\mathcal{D}$, in other words, the election induced by $(C \setminus X, V)$ has the structure of $\mathcal{D}$. Our algorithm proceeds as follows. For a subset of candidates $Y \subseteq X$, let:

$$\mathcal{J}_Y := (E = (C, V); (C \setminus X, Y, X \setminus Y), b, R, \alpha).$$

If $\mathcal{J}_Y$ is a Yes-instance of $(\ell, \mathcal{D})$-CC Extension for some $Y \subseteq X$, then our algorithm returns Yes and aborts. If, on the other hand, for every subset $Y \subseteq X$ of candidates it turns out that $\mathcal{J}_Y$ is a No-instance of $(\ell, \mathcal{D})$-CC Extension, then we return No. It is easy to see that whenever the algorithm returns Yes, assuming the correctness of the $(\ell, \mathcal{D})$-CC Extension procedure used, there exists a committee that has the desired misrepresentation score.

To argue the correctness of the algorithm, we show that if $\mathcal{I}$ is a Yes-instance then the algorithm does indeed produce a committee that can achieve the desired misrepresentation score. To this end, let $C^\star$ be a committee whose ℓ-misrepresentation score under the function α is at most R. Let $Y^\star$ denote $C^\star \cap X$. Then note that $C^\star$ is a committee that respects the partition $D := C \setminus X$, $G := Y^\star$, and $B := X \setminus Y^\star$. Further, note that since X is a candidate modulator to $\mathcal{D}$, the election induced by (D, V) belongs to the domain $\mathcal{D}$. Clearly, the instance $(E = (C, V); (D, G, B), b, R, \alpha)$ is a well-formed input to the $(\ell, \mathcal{D})$-CC Extension problem, and $C^\star$ is a valid solution to it. Assuming again the correctness of the $(\ell, \mathcal{D})$-CC Extension procedure used, we are done. Observe that the running time of our algorithm here is $2^k q(n, m)$, where $q(n, m)$ is the time required by the $(\ell, \mathcal{D})$-CC Extension procedure on an instance of size $n + m$.

We now turn to the $(\ell, \mathcal{D})$-CC Via χ problem where we are given a k-sized voter modulator as input, or that χ is fixed to be voters. Here a direct brute-force approach as in the previous case does not suggest itself, because of which we suffer a greater overhead in our running time. For simplicity, we first describe our algorithm for the egalitarian variant, that is, we fix $\ell = \ell_\infty$. We later describe the changes we need to make when we deal with the utilitarian variant.

Let $(E = (C, V), b, R, \alpha, X)$, denoted by $\mathcal{I}$, be an instance of $(\ell, \mathcal{D})$-CC VIA χ. Recall that X is a voter modulator to the domain $\mathcal{D}$, in other words, the election induced by $(C, V \setminus X)$ has the structure of $\mathcal{D}$. For every voter, we guess the candidate who represents that voter in an arbitrary but fixed, and valid, Chamberlin-Courant committee. For such a guess μ, let Y_μ denote the set of at most k candidates who have been chosen to represent the voters in the modulator. More specifically, a voter $v \in X$, let $\mu(v)$ denote the candidate that we have guessed as the representative for the voter v, and let $d(v)$ denote the set of candidates ranked higher than $\mu(v)$ by the voter v. Note that Y_μ is simply $\cup_{v \in X} \mu(v)$.

We first run the following easy sanity check: if, for $u, v \in X$, $u \neq v$, we have that $\mu(v) \in d(u)$, then we reject the guess Y. Otherwise, define $B_\mu := \cup_{v \in X} d(v)$ and $G_\mu := Y_\mu$, and let $D_\mu := C \setminus (G \cup B)$. Observe that B_μ and G_μ are disjoint because of the sanity check. Further, let:

$$\mathcal{J}_\mu := (E = (C, V \setminus X); (D_\mu, G_\mu, B_\mu), b, R, \alpha).$$

It is easily checked that $\mathcal{J}_\mu$ is a well-formed instance for $(\ell, \mathcal{D})$-CC EXTENSION. As before, we return YES if and only if there exists a guess μ for which $\mathcal{J}_\mu$ is a YES instance of $(\ell, \mathcal{D})$-CC EXTENSION. To see the correctness of this approach, let $C^\star$ be a committee whose ℓ-misrepresentation score under the function α is at most R. For each voter $v \in X$, let $\mu^\star(v)$ denote the top-ranking candidate from $C^\star$ in the vote of v. Let $Y^\star$ be given by $\cup_{v \in X} \mu^\star(v)$, and let $B^\star$ be the set of all candidates ranked higher than $\mu^\star(v)$ in the votes v from X. Observe that $C^\star$ does not contain any candidates from $B^\star$ by the definition of $\mu^\star$.

Now, as before, define: $G := Y^\star$, $B := B^\star$, and $D := C \setminus (G \cup B)$. Clearly, the instance $(E = (C, V \setminus X); (D, G, B), k, R, \alpha)$ is a well-formed input to the $(\ell, \mathcal{D})$-CC EXTENSION problem, and $C^\star$ is a valid solution to it. Assuming again the correctness of the $(\ell, \mathcal{D})$-CC EXTENSION procedure used, we are done. Observe that the running time of our algorithm here is $n^k q(n, m)$, where $q(n, m)$ is the time required by the $(\ell, \mathcal{D})$-CC EXTENSION procedure on an instance of size $n + m$. For the utilitarian version of the problem (where $\ell = \ell_1$), the procedure is identical, except that we use R' instead of R in the definition $\mathcal{J}_\mu$, where R' is $R - R_{X,\mu}$, and $R_{X,\mu}$ is the sum of the misrepresentation score of the candidate $\mu(v)$ with respect to the voter v, and the sum is over $v \in X$. It is easily verified that the other details work out in the same fashion.

The rest of this section is section is devoted to showing that the $(\ell, \mathcal{D})$-CC EXTENSION problem can be solved in polynomial time by adapting suitably the known algorithms for the Chamberlin-Courant problem on the relevant domain $\mathcal{D}$. These adaptations are sometimes subtle and in particular for the single-peaked case, we have to treat the utilitarian and the egalitarian variants separately (corresponding to $\ell = \ell_1$ and $\ell = \ell_\infty$ respectively).

3.1 $(\ell, \mathcal{D})$-CC Extension for the Single-Crossing Domain

In this section we demonstrate a polynomial time algorithm for the $(\ell, \mathcal{D})$-CC Extension problem for the case when $\mathcal{D} = \mathrm{SC}$. This builds closely on the algorithm shown by [10]. First, we show a structural property which is an easy adaptation of Lemma 5 in [10]. The statement corresponding to single-crossing profiles states that there is an optimal committee for which an optimal assignment assigns candidates in contiguous blocks over the single-crossing order. For the $(\ell, \mathcal{D})$-CC Extension problem, this continues to be the case for candidates c from $\mathcal{D}$ except that some candidates in the contiguous block may be assigned to candidates in G instead of being assigned to c. We now state this formally. In the statement below, an optimal b-CC assignment is considered only among committees that respect the annotation (D, G, B) in the given instance $\mathcal{I}$ of $(\ell, \mathcal{D})$-CC Extension.

Lemma 1 **($\star$).** *Let $\mathcal{I} = (E = (C, V); (D, G, B), b, R, \alpha)$ be an instance of (ℓ,SC)-CC* Extension. *Suppose $V = (v_1, \dots, v_n)$ is the single-crossing order of the votes and $C = (c_1, \dots, c_m)$ is an ordering of the candidates according to v_i. Then for every $b \in [m]$, every dissatisfaction function α for m candidates, and for every $\ell \in \{\ell_1, \ell_\infty\}$, there is an optimal b-CC assignment Φ for E under $\alpha - \ell - \mathrm{CC}$ such that for each candidate $c_i \in D$, if $\phi^{-1}(c_i) \neq \emptyset$, then there are two integers e_i and f_i, with $e_i < f_i$, such that for every vote v in the set of voters $V' = \{v_{e_i}, v_{e_i+1}, \dots, v_{f_i}\}$, $\phi(v) \in \{c_i\} \cup G$. Moreover, for each $i < j$ such that $\Phi^{-1}(c_i) \neq \emptyset$ and $\Phi^{-1}(c_j) \neq \emptyset$, it holds that $e_i < f_i$.*

Due to space considerations we omit the proof of the technical claim above, however, we note that it is along the lines of the proof in [10]. In particular, observe that if there are voters u, v, w appearing in that order in the single-crossing ordering, and for two candidates $c_1, c_2 \in D$, if u and w were to be assigned to c_1 and v were to be assigned to c_2, then this would imply that $c_1 \succ_u c_2$ and $c_1 \succ_w c_2$, while $c_2 \succ_v c_1$, violating the single-crossing structure of the election restricted to D. Since the only other assignments allowed are to candidates in D, the claim follows. We now have the following natural consequence.

Lemma 2. *(ℓ,SC)-CC* Extension *admits a polynomial time algorithm, both for when $\ell = \ell_1$ and when $\ell = \ell_\infty$.*

*Proof. (Sketch)*For Single-Crossing profiles we propose a modified version of the dynamic programming routine which was originally developed in [10]. Here, for $i \in \{0\} \cup [n]$, $j \in [m - |G| - |B|]$ and $t \in b - |G|$, we define $A[i, j, t]$ as the best possible misrepresentation score that can be achieved by a committee of size $t + |G|$ that respects the annotation (G, B, D) formed using a subset of first j candidates considering first i votes, where the candidates of D are ordered according to the ranking of the first voter in the single-crossing ordering and the voters are ordered according to the single-crossing ordering. The recurrence for single-crossing orders works by "guessing" the first voter v to be represented by the candidate c_j, and the optimal representation of the preceding voters is found recursively. In our setting, this approach continues to work, except that instead of simply adding up the misrepresentation score of c_j for all voters in the interval starting from v and ending at v_i, we check (for every vote in this interval) if there is a candidate from G who is ranked above c_j, and appropriately adjust the calculation of the misrepresentation score for such voters. The time complexity of above algorithm turns out to be $\mathcal{O}(mn^2k)$ (as calculating the misrepresentation score for each voter can take $\mathcal{O}(n)$ time). □

3.2 $(\ell, \mathcal{D})$-CC Extension for the Single-Peaked Domain

For the single-peaked domain, as alluded to earlier, we need to consider the utilitarian and egalitarian variants separately. We first consider $\ell = \ell_1$. In the following discussion the terms *first* and *last* are with respect to the societal order, which we denote by $\sqsupset$. A candidate c_i is said to be *smaller* than another candidate c_j if the candidate c_i appears before c_j in the societal order $\sqsupset$, and a candidate is said to be *larger* if it appears after the other candidate. Betzler et al [2] proposed separate algorithms for the utilitarian and egalitarian variants. To solve $(\ell, \mathcal{D})$-CC Extension in this setting, we extend the dynamic programming algorithm proposed by Betzler et al for the utilitarian setting.

Lemma 3. *$(\ell_1$,SP)-CC* Extension *admits a polynomial time algorithm.*

Proof. Recall that we are given an instance $(E = (C, V); G, B, D, b, r, R, \ell)$ of $(\ell_1$,SP)-CC Extension. If $b = |G|$, then there is nothing to do. If $b > |G|$, we assume without loss of generality that there is at least one voter whose top candidate does not belong to G, otherwise we may simply return Yes since the committee G is already good enough for any reasonable R[1]. The main semantics of the DP table employed previously is the following. For $i \in [m]$ and $j \in 1, \ldots, \min(i, k)$, we define $z(i, j)$ to be the total misrepresentation for a set of j winners from $\{c_1, \ldots, c_i\}$ including c_i. The final answer is given by $\min_{i \in \{k,\ldots,m\}} z(i, k)$.

We let d denote $|D|$ and let $c_1 \succ c_2 \succ \cdots c_d$ be the single-peaked order. As before, for $i \in [m]$ and $j \in 1, \ldots, \min(i, k)$, we define a modified DP table as follows: let $z(i, j)$ be the total misrepresentation for a set of j winners from $\{c_1, \ldots, c_i\}$ including $\{c_i\} \cup G$. Now, note that the final answer is given by $\min_{i \in \{b',\ldots,m\}} z(i, b')$, where $b' = |G| - b$. Observe that our solution respects the partition (G, B, D), since the semantics of z are such that the candidates G are always incorporated and no candidate from B is ever chosen. Towards describing the recurrence, we establish some notation. First, let $g^*(v)$ denote the highest-ranked candidate from G in the ordering of the voter v. Also, define:

$$g(p, i) := \sum_{v \in V} \max\{0, \min\{r(v, c_p) - r(v, c_i), r(v, g^*(v)) - r(v, c_i)\}\}$$

[1] If $R < \alpha(1) * n$, for instance, then it is already impossible to achieve for any committee.

Intuitively, $g(p, i)$ gives the potential gain of assigning candidate i to the voter v, assuming that the voter v was previously assigned to either the candidate c_p or $g^*(v)$. Both $d(p, i)$ and $g(i)$ can be precomputed in time $\mathcal{O}(nm^2)$ by performing one pass over the votes and two passes over the candidates.We are now ready to describe the main recurrence:

$$z[i, j] = \min_{j-1 \leqslant p \leqslant i-1} \left(z\,[p, j-1] - g(p, i) \right),$$

with the base case:

$$z[i, 1] = \min(r(v, c_i), r(v, g^*(v)).$$

Due to space constraints, our argument for correctness only focuses on the part that needs to be adapted appropriately from the proof of [2]. Let C^* be a committee that witnesses the value of $z[i, j]$. Let p be the largest index smaller than i (in the societal ordering) which is such that $c_p \in C^*$ and let $g^*(v)$ be c_q. If for a voter v it holds that $r(v, c_i) < r(v, c_p)$ and $r(v, c_i) < r(v, c_q)$, then note that $r(v, c_i) < r(v, c_t)$ for all $t < p$. Then the contribution of such a voter v to the misrepresentation of $z[p, i-1]$ is $\min(r(v, c_p), r(v, c_q))$. This implies that the improvement in the misrepresntation score of this voter obtained by reassigning the voter to the candidate c_i is precisely given by $g(p, i)$. For all other voters, an assignment to c_i does not improve their misrepresentation, so the algorithm does nothing in these situations. The correctness follows from the fact that the algorithm tries all possible values of p, and the inductively assumed correctness of $z[p, j-1]$. The time complexity of the core algorithm is $\mathcal{O}(m^2)$, as both i and j can take at most m values, coupled with the time to precompute $d(p, i)$ and $g(i)$, the total time complexity is $\mathcal{O}(nm^2)$. $\square$

We now turn to the egalitarian version of the rule, that is, $\ell = \ell_\infty$. Here again, the solution involves a straightforward adaptation of the approach of [2] to account for the constraints imposed by the (G, B, D) annotations in the extension problem.

Lemma 4. *(ℓ_∞,SP)-CC* Extension *admits a polynomial time algorithm.*

Proof. (Sketch.) Let q be the largest integer for which $\alpha(q) \leqslant R$. We first remove voters who have a candidate from G in their top q positions. Let V' denote the remaining set of voters. For a voter $v \in V'$, let $T_q(v)$ denote the top q candidates in v's ranking. Consider the set $M(v) := T_q(v) \setminus B$. Note that any valid committee must contain a candidate from $M(v)$ for all $v \in V'$. However, observe that the set $M(v) \subseteq D$, and therefore forms a continuous interval on the societal ordering of candidates in D. Therefore our problem reduces to finding a clique cover of size at most $b - |G|$ on the interval graph that is naturally defined by the votes in V', which can be found in time $O(nm)$. $\square$

4 Hardness for Generalized Restrictions on the Domain

4.1 3-composite single-peaked domains.

To show the hardness of computing an optimal ℓ_∞-CC committee on double-peaked domains, we reduce from the following variant of SAT, which is called LSAT. In an LSAT instance, each clause has at most three literals, and further the literals of the formula can be sorted such that every clause corresponds to at most three consecutive literals in the sorted list, and each clause shares at most one of its literals with another clause, in which case this literal is extreme in both clauses. The hardness of LSAT was shown in [1]. For ease of description, we will assume in the following reduction that every clause has exactly three literals, although it is easy to see that the reduction can be extended to account for smaller clauses as well.

Theorem 2. *Computing an optimal ℓ_∞-CC committee with respect to the Borda misrepresentation score is* NP*-hard even when the domain is a three-composite single-peaked domain.*

Proof. **(Sketch.)** Let ϕ be an instance of LSAT with variables $x_1, \ldots, x_n$ and clauses $C_1, \ldots C_m$. Towards constructing the election instance, we introduce one candidate for every literal in ϕ. Let p_1 and q_i denote the candidates corresponding to the variable x_i. We also introduce $(n+1)$ dummy candidates for each variable (which is a total of $n(n+1)$ dummy candidates). Let $d[i,j]$ denote the j^{th} dummy candidate corresponding to the variable x_i. We use C to denote the $2n$ candidates corresponding to the literals, and D to denote the set of dummy candidates. P and Q denote the candidates corresponding to the positive and the negated literals respectively.

Let us fix the ordering σ on the candidates as follows. The first $2n$ candidates are from C arranged according to the LSAT ordering. The last $n(n+1)$ candidates are from D and are arranged in an arbitrary but fixed order. Let σ' be the reverse of σ. For a subset of candidates X, the notation $\overline{X}$ refers to an ordering of X according to σ. For a subset of candidates $X \subset C$, who occupy adjacent positions in the LSAT ordering projected over C, the notation $\overrightarrow{C \setminus X}$ refers to an ordering according to σ of the candidates from $C \setminus X$ who appear after X in the LSAT ordering and similarly $\overleftarrow{C \setminus X}$ refers to an ordering according to σ' of the candidates from $C \setminus X$ who appear before X in the LSAT ordering. This notation easily yields an ordering which is single-peaked — $\overline{X} \succ \overrightarrow{C \setminus X} \succ \overleftarrow{C \setminus X}$. For a subset of candidates $X \subset C$, who occupy adjacent positions in the LSAT ordering projected over C, the notation $\overleftrightarrow{C \setminus X}$ refers to an ordering according to σ of the candidates from $C \setminus X$ who appear after X in the LSAT ordering followed by an ordering according to σ' of the candidates from $C \setminus X$ who appear before X in the LSAT ordering. This notation allows us to easily express an ordering which is single-peaked — $\overline{X} \succ \overleftrightarrow{C \setminus X}$.

We would now like to setup the votes in such a way that a winning committee corresponds to a valid satisfying assignment. We introduce one vote for every clause as follows. Suppose the clause c consists of the literals (ℓ_1, ℓ_2, ℓ_3), and let the candidates corresponding to these literals be t_1, t_2, t_3 respectively. If $\ell_1 < \ell_2 < \ell_3$ in the LSAT ordering, then we introduce the following vote:

$$\nu(c) := t_2 \succ t_1 \succ t_3 \succ \overleftrightarrow{(C \setminus \{t_1, t_2, t_3\})} \succ \overline{D}$$

For every variable x_i, we also introduce the following $(n+1)$ votes, with $1 \leqslant j \leqslant (n+1)$:

$$\nu(x_i, j) := d[i, j] \succ p_i \succ q_i \succ \overleftrightarrow{(P \setminus \{p_i\})} \succ \overleftrightarrow{(Q \setminus \{q_i\})} \succ \overleftrightarrow{D \setminus \{d[i, j]\}}$$

This completes a description of the profile. We fix the Borda misrepresentation target score at two and the committee size is set to n. It is easily checked that this profile is three-composite single-peaked with respect to the partition (P, Q, D). First we look at $\nu(c)$ – the votes based on a clause. $\nu(c)$ when projected on D is trivially single-peaked. $\nu(c)$ when projected on C is single-peaked, and hence when projected on $P, Q \subset C$ will remain single-peaked . Now we look at $\nu(x_i, j)$ – the votes based on variables, which are clearly single-peaked when projected over P, Q and D individually. We now prove the equivalence of these two instances.

In the forward direction, we simply pick the literals corresponding to a satisfying assignment. If a satisfying assignment does not set a variable, then we pick either p_i or q_i. This clearly satisfies every vote based on a clause $\nu(c)$, if a vote is not satisfied, then the corresponding clause will also not be satisfied. This trivially satisfies the votes based on variables $\nu(x_i, j)$, as we pick at least one from p_i and q_i satisfying $\nu(x_i, j)$ for all $1 \leqslant j \leqslant n+1$.

In the reverse direction, let W be a committee whose score is at most two. Observe that W must choose at least one of p_i or q_i, for all $1 \leqslant i \leqslant n$. Indeed, if not, then such a committee is forced to pick every $d[i, j]$, $1 \leqslant j \leqslant n+1$, which is a violation of the committee size. Since the committee has at most n candidates, it follows by a standard pigeon-hole argument that $|W \cap \{p_i, q_i\}| \leqslant 1$ for all $1 \leqslant i \leqslant n$, which implies that we pick exactly one of p_i or q_i. Therefore, the committee corresponds naturally to an unambiguous assignment of the variables. It is easily checked that this satisfies every clause, because an unsatisfied clause c would correspond to a voter $\nu(c)$ whose Borda misrepresentation score would exceed two. This completes the proof.

4.2 3-Crossing Profiles

In this section, we show the hardness of computing an optimal ℓ_∞-CC committee with respect to the Borda misrepresentation score with respect to three-crossing domains. The reduction is again from LSAT, and the construction is similar to the one used in the proof of Theorem 2 in that we again have candidates corresponding to literals and votes representing clauses. A committee corresponds to a satisfying assignment precisely when its misrepresentation score is at most two. The main difference from before is in how the candidates are ordered in the preferences of the voters.

Theorem 3. *Computing an optimal ℓ_∞-CC committee with respect to the Borda misrepresentation score is* NP*-hard even when the domain is three-crossing domain.*

Proof. **(Sketch.)** Let ϕ be an instance of LSAT with variables $x_1, \ldots, x_n$ and clauses $C_1, \ldots C_m$. Without loss of generality, let us assume that the ordering of the clauses in the LSAT instance is also given by $C_1, \ldots, C_m$. Towards constructing the election instance, we introduce one candidate for every literal in ϕ. Let p_i and q_i denote the candidates corresponding to the variable x_i. We also introduce $(n+1)$ dummy candidates for each variable (which is a total of $n(n+1)$ dummy candidates). Let $d[i, j]$ denote the j^{th} dummy candidate corresponding to the variable x_i. We use C to denote the $2n$ candidates corresponding to the literals, and D to denote the set of dummy candidates.

Towards describing the votes, let us fix an ordering σ on the candidates as follows. The first $2n$ candidates are from C arranged according to the LSAT ordering. The last $n(n+1)$ candidates are from D and are arranged in an arbitrary but fixed order. For a subset of candidates X, the notation $\overline{X}$ refers to an ordering of X according to σ. We would now like to setup the votes in such a way that a winning committee corresponds to a valid satisfying assignment. For $1 \leqslant i \leqslant m-1$, let G_i denote literals in the set $C_i \setminus C_{i+1}$, while we let G_m denote the literals in C_m. We are now ready to describe the votes. For every $1 \leqslant i \leqslant m$, we introduce the vote v_i, which has the literals of the clause C_i in the top three positions, and the remaining candidates are ranked as follows:

$$v_i := \overline{G_i} \succ \overline{G_{i+1}} \succ \cdots \succ \overline{G_m} \succ \overline{G_{i-1}} \succ \cdots \succ \overline{G_1} \succ \overline{D}$$

It is useful to note that the vote v_{i+1} can be thought of as a ranking obtained from the vote v_i by "pushing back" the tuple $\overline{G_i}$ to just behind $\overline{G_m}$. Therefore, the ordering among the G_i's in v_m is reverse of their ordering in v_1. Observe that if a literal occurs in $C_i \cap C_{i+1}$, then it appears among the top three positions of both v_i and v_{i+1}.

We now turn to the second part of our profile, which consists of votes corresponding to the variables. Here, for a subset of candidates X, we will use $\overline{\overline{X}}$ to refer to an ordering of X according to v_m. Now, for every variable x_i, we introduce the following $(n+1)$ votes, with $1 \leqslant j \leqslant (n+1)$.

$$v_{i,j} := d[i, j] \succ p_i \succ q_i \succ \overline{\overline{(C \setminus \{p_i, q_i\})}} \succ \overline{D \setminus \{d[i, j]\}}$$

This completes a description of the profile. We fix the Borda misrepresentation target score at two and the committee size is set to n. The argument for the equivalence of the instances is similar to that in the proof of Theorem 2, and we will revisit it shortly. It can be shown, by a careful case analysis [2], that this profile is three-crossing with respect to the following ordering of the votes:

$$v_1, v_2, \ldots, v_m, v_{1,1}, \ldots, v_{1,n+1}, \ldots, v_{i,1}, \ldots, v_{i,n+1}, \ldots, v_{n,1}, \ldots v_{n,n+1}$$

[2] We omit the details due to lack of space.

We now turn to a proof of equivalence. In the forward direction, we simply pick the literals corresponding to a satisfying assignment. If a satisfying assignment does not set a variable, then we pick either p_i or q_i. This clearly satisfies every vote based on a clause (or the assignment would not be a satisfying one), and trivially satisfies the votes based on variables.

In the reverse direction, let W be a committee whose score is at most two. Observe that W must choose at least one of p_i or q_i, for all $1 \leqslant i \leqslant n$. Indeed, if not, then such a committee is forced to pick every $d[i, j]$, $1 \leqslant j \leqslant n+1$, which is a violation of the committee size. Since the committee has at most n candidates, it follows by a standard pigeon-hole argument that $|W \cap \{p_i, q_i\}| \leqslant 1$ for all $1 \leqslant i \leqslant n$. Therefore, the committee corresponds naturally to an unambiguous assignment of the variables. It is easily checked that this satisfies every clause, because an unsatisfied clause c would correspond to a voter $v(c)$ whose Borda misrepresentation score would exceed two. This completes the proof. □

5 Concluding Remarks

We have made some progress in demonstrating that the Chamberlin-Courant voting rule can be computed efficiently on nearly-structured domains, and there are some notions of being "almost structured" for which the rule remains hard. Several specific problems remain open. The most pertinent issue is whether the problem admits a FPT algorithm when parameterized by the size of a voter modulator to either single-peaked or single-crossing profiles. The complexity of the utilitarian version of the voting rule on composite profiles or k-crossing profiles is also open.

References

1. Esther M. Arkin, Aritra Banik, Paz Carmi, Gui Citovsky, Matthew J. Katz, Joseph S. B. Mitchell, and Marina Simakov. *Choice Is Hard*, pages 318–328. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
2. Nadja Betzler, Arkadii Slinko, and Johannes Uhlmann. On the computation of fully proportional representation. *J. Artif. Intell. Res. (JAIR)*, 47:475–519, 2013.
3. Duncan Black. On the rationale of group decision-making. *The Journal of Political Economy*, pages 23–34, 1948.
4. Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel Procaccia. Handbook of computational social choice, 2016.
5. Robert Bredereck, Jiehua Chen, and Gerhard J Woeginger. Are there any nicely structured preference profiles nearby? *Mathematical Social Sciences*, 79:61–73, 2016.
6. John R Chamberlin and Paul N Courant. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, 77(03):718–733, 1983.
7. Marek Cygan, Fedor V Fomin, ukasz Kowalik, Daniel Lokshtanov, D 'aniel Marx, Marcin Pilipczuk, Micha Pilipczuk, and Saket Saurabh. *Parameterized algorithms.* Springer, Cham, Cham, 2015.

8. Edith Elkind and Martin Lackner. On detecting nearly structured preference profiles. In Carla E. Brodley and Peter Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 661–667. AAAI Press, 2014.
9. Gábor Erdélyi, Martin Lackner, and Andreas Pfandler. Computational aspects of nearly single-peaked electorates. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.
10. Piotr Skowron, Lan Yu, Piotr Faliszewski, and Edith Elkind. The complexity of fully proportional representation for single-crossing electorates. *Theor. Comput. Sci*, 569:43–57, 2015.