Multiwinner Elections under Minimax Chamberlin-Courant Rule in Euclidean Space

Chinmay Sonar[1], **Subhash Suri**[1] and **Jie Xue**[2]
[1]University of California, Santa Barbara
[2]New York University, Shanghai
{csonar, suri}@cs.ucsb.edu, jiexue@nyu.edu

Abstract

We consider multiwinner elections in Euclidean space using the minimax Chamberlin-Courant rule. In this setting, voters and candidates are embedded in a d-dimensional Euclidean space, and the goal is to choose a committee of k candidates so that the *rank* of any voter's most preferred candidate in the committee is minimized. (The problem is also equivalent to the *ordinal* version of the classical k-center problem.) We show that the problem is NP-hard in any dimension $d \geq 2$, and also provably hard to approximate. Our main results are three polynomial-time approximation schemes, each of which finds a committee with provably good minimax score. In all cases, we show that our approximation bounds are tight or close to tight. We mainly focus on the 1-Borda rule but some of our results also hold for the more general r-Borda.

1 Introduction

Multiwinner elections are a classical problem in social science where the goal is to choose a fixed number of candidates, also called a *winning committee*, based on voters' preferences. The problem encapsulates a number of applications that range from choosing representatives in democracies to staff hiring and procurement decisions [Lu and Boutilier, 2011; Goel *et al.*, 2019; Skowron *et al.*, 2016]. Many facility location problems in operations research and public goods planning are equivalent to multiwinner elections: facilities are candidates and users are voters [Betzler *et al.*, 2013].

One of the central computational problems in this area is to design algorithms for computing a winning committee with *provable* guarantees of representation quality. In particular, let V be a set of n voters, $\mathcal{C}$ a set of m candidates, and k the size of the winning committee, for a positive integer k. We focus on *ordinal* preferences where each voter v ranks all the candidates in $\mathcal{C}$, from the most preferred (rank 1) to the least preferred (rank m). Elections under ordinal preferences are widely studied because in many settings, the rank ordering of candidates is both natural and easier to determine than a precise numerical value (cardinal preference) [Brandt *et al.*, 2016; Endriss, 2017; Munagala *et al.*, 2021]. We let $\sigma_v(c)$ denote the rank of candidate c in v's list, and use Chamberlin-Courant voting rule [Chamberlin and Courant, 1983] for evaluating the score of a committee. In the simplest version, also called 1-Borda, a voter v's score for a committee T is the rank of its most preferred committee member: $\sigma_v(T) = \min_{c \in T} \sigma_v(c)$. (In a generalized form called r-Borda, which we also consider, the score is the *sum* of the ranks of the r most preferred candidates in T.) This scoring function is often called the *misrepresentation* score to emphasize that we want to *minimize* it—more preferred candidates have smaller ranks.

The goal of a multiwinner election is to find a size-k committee $T \subseteq \mathcal{C}$ that optimizes some function $g(\sigma_{v_1}(T), \ldots, \sigma_{v_n}(T))$ of all the voter's scores. Two classical choices for g are the *sum* and the *max*. The former is the *utilitarian* objective and seeks to minimize the *sum* of the scores over all the voters. The latter is the *egalitarian* objective and minimizes the *maximum* (worst) of the scores over all the voters. Both versions of the multiwinner elections are NP-hard under *general preferences* [Lu and Boutilier, 2011; Betzler *et al.*, 2013], and as a result, an important line of research has been to examine natural settings with *structured* preference spaces [Betzler *et al.*, 2013; Yu *et al.*, 2013; Skowron *et al.*, 2015b; Elkind *et al.*, 2017b].

Our paper studies one such setting—and arguably one of the most natural—namely, the Euclidean space of preferences. The geometry of Euclidean space gives an intuitive and interpretable *positioning* of voters and candidates in many natural settings such as spatial voting and facility locations, but it also has important computational advantages: when candidates and voters are embedded in d-space, only a tiny fraction of all (exponentially many) $m!$ candidate orderings are *realizable*. In particular, the maximum number of realizable rankings is only (polynomially bounded) $O(m^{d+1})$. This important combinatorial property enables us derive much better bounds than what is possible in completely unstructured preferences spaces. Specifically, we explore algorithmic and hardness questions for *minimax* (egalitarian) multiwinner elections using the 1-Borda and r-Borda rules under ordinal *Euclidean* preferences. In this setting, voters and candidates are embedded in a d-dimensional Euclidean space, implicitly specifying each voter's ranking (closest to farthest) of the candidates. The goal is to choose a committee of k candidates minimizing the rank of the worst voter's most preferred candidate.

Remark: There are good reasons for using *ordinal* preferences even when cardinal distances are implied by an Euclidean embedding. The first is robustness: consider a voter v and two candidates c, c'. If their distances satisfy $d(v, c) < d(v, c')$, then clearly v prefers c to c', but it seems harder to argue that v's preference varies linearly (or even smoothly) with distance—for instance, would doubling the distance really halve the value to a voter? Another reason is that k-center solutions based on cardinal preferences are highly susceptible to the outlier effect—a few outlying voters may control the minimax value (i.e., radius) of the optimal solution even though all other voters have significantly better solution quality. By contrast, under the ordinal measure the (rank-based) solution seems more equitable because outliers are matched with a highly ranked candidate (irrespective of the distances).

1.1 Our Results

Our work shows that a number of interesting and encouraging approximation results are possible for Euclidean preferences, thus, offering new directions for research. Indeed, quoting [Elkind *et al.*, 2017b], "multidimensional domain restrictions offer many challenging research questions, but fast algorithms for these classes are very desirable." A brief summary of our main results is the following:

1. We show that the Euclidean minimax committee problem is NP-hard in every dimension $d \geq 2$; in one dimension, the problem is easy to solve optimally with dynamic programming. The complexity of this problem, also called the ordinal Euclidean k-center problem, was not known and had been an important folklore problem. Our proof shows that the problem also hard to approximate in the worst-case (see Theorem 2), which stands in sharp contrast with the 2-approximability of the *cardinal* k-center problem [Kleinberg and Tardos, 2006].
2. We then show a number of efficient approximation results for the problem, starting with a polynomial time algorithm to compute a size-k committee with a minimax score of $O(m/k)$ for any instance in dimension $d = 2$, and score of $O((m/k) \log k)$ for any instance in dimension $d \geq 3$. These scores are also shown to be essentially the best possible in worst-case.
3. Our next approximation uses the bicriterion framework to design a polynomial time algorithm that achieves the *optimal* minimax score $\sigma^\star$ possible for a size-k committee by constructing slightly larger committee, namely, of size $(1 + \epsilon)k$ for $d = 2$ and size $O(k \log m)$ for $d \geq 3$. (We also show that increasing the committee size by an *additive* constant is not sufficient.)
4. Our final approximation combines ordinal and cardinal features of the problem in a novel way, as follows. Suppose the optimal score of the k committee is $\sigma^\star$, and $d_v^\star$ is the distance of v to its rank $\sigma^\star$ candidate. We define a committee T to be δ*-optimal* if each voter has a representative in T within distance $\delta d_v^\star$. (That is, for each voter the committee contains a candidate whose distance to the voter is almost as good as distance to its σ^* rank candidate.) We show that a δ-optimal committee can be computed in polynomial time for $\delta = 3$, but unless P = NP, there is no polynomial-time algorithm to compute δ-optimal committees for any $\delta < 2$.

Due to limited space, we defer some of the proofs to the full version of this work [Sonar *et al.*, 2022].

2 Related Work

The literature on multi-winner elections is too large to summarize in this limited space; hence, we mostly survey the computational results closely related to our work. For a general introduction to multi-winner elections, we refer the reader to the works of [Elkind *et al.*, 2017a; Faliszewski *et al.*, 2017; Faliszewski *et al.*, 2019]. The work of [Aziz *et al.*, 2018] studies the computational complexity and axiomatic properties of various egalitarian committee scoring rules under the general preferences. For computing a winning committee under the Chamberlin-Courant rule, polynomial-time algorithms are known only for restricted preferences such as single-peaked, single-crossing, 1D Euclidean, etc. [Betzler *et al.*, 2013; Skowron *et al.*, 2015b; Elkind *et al.*, 2017b]. Very little is known about the more general d-dimensional Euclidean setting considered in our paper with the exception of a work of [Godziszewski *et al.*, 2021], which shows NP-hardness for the *approval set* voting rule for the utilitarian objective in 2-dimensional Euclidean elections.

Constant factor approximations are often easier to achieve under the *utilitarian* objective. For instance, [Munagala *et al.*, 2021] present several nearly-optimal approximation bounds; [Skowron *et al.*, 2015a] presents a constant factor approximation for minimizing the weighted sum of ranks of the winning candidates; and [Byrka *et al.*, 2018] presents a constant factor approximation for minimizing the sum when $\sigma_v(c)$ is an arbitrary cardinal value. In contrast, for the minimax objective, mostly inapproximability results are known, and only under general preferences [Skowron *et al.*, 2015a].

3 Preliminaries

Throughout the paper, an *election* is a pair $E = (\mathcal{C}, V)$, where $\mathcal{C} = \{c_1, \dots, c_m\}$ is the *candidate* set and $V = \{v_1, \dots, v_n\}$ is the *voter* set. The *preference list* of each voter is a total ordering (ranking) of $\mathcal{C}$, in which the most preferred candidate has rank 1 and the least preferred candidate has rank m.

We call an election $E = (\mathcal{C}, V)$ a d*-Euclidean election* if there exists a function $f : \mathcal{C} \cup V \to \mathbb{R}^d$, called a *Euclidean realization* of E, such that for any pair $c_i, c_j \in \mathcal{C}$, a voter v prefers c_i to c_j if and only if $\text{dist}(f(v), f(c_i)) < \text{dist}(f(v), f(c_j))$ where $\text{dist}(\cdot, \cdot)$ denotes the Euclidean distance (in case the candidates are equidistant, we break the ties arbitrarily). We assume that a Euclidean realization is part of the input; the decision problem of whether an election admits an Euclidean realization is computationally hard [Peters, 2017].

We use $\sigma_v(c)$ to denote the rank of candidate c in v's preference list, and use the Chamberlin-Courant voting rule [Chamberlin and Courant, 1983] for evaluating the score of a committee. In particular, a voter v's score for a committee T is the rank of its most preferred committee member

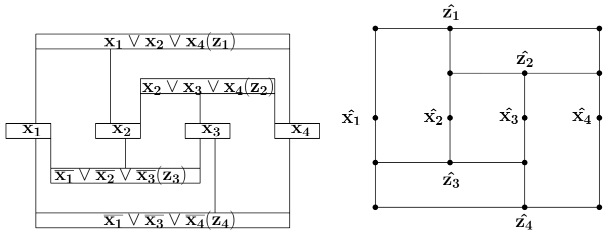

Figure 1: The rectangular (left) and the orthogonal (right) embedding of the PM-3SAT instance, respectively.

$\sigma_v(T) = \min_{c \in T} \sigma_v(c)$, and the *Euclidean minimax committee* problem is to choose a committee of size k that minimizes the maximum score (misrepresentation) of any voter. That is, minimize the following:

$$\sigma(T) = \max_{v \in V} \left(\min_{c \in T} \sigma_v(c) \right),$$

So the optimal committee score is always between 1 and m. (We mainly focus on this simpler 1-Borda scoring function, but some of our results also hold for the generalized r-Borda, as mentioned later.)

4 Hardness Results

We first show that the Euclidean minimax committee problem is NP-hard in any dimension $d \geq 2$. After that, we extend our proof to show that the problem is even hard to approximate within any sublinear factor of m, where m is the number of candidates in the election.

Our hardness reduction uses the NP-complete problem PLANAR MONOTONE 3-SAT (PM-3SAT) [de Berg and Khosravi, 2010]. An instance of PM-3SAT consists of a *monotone* 3-CNF formula ϕ where each clause contains either three positive literals or three negative literals, and a special "planar embedding" of the variable-clause incidence graph of ϕ described below. In the embedding, each variable/clause is drawn as a (axis-parallel) rectangle in the plane. The rectangles for the variables are drawn along the x-axis, while the rectangles for positive (resp., negative) clauses lie above (resp., below) the x-axis. All the rectangles are pairwise disjoint. If a clause contains a variable, then there is a vertical segment connecting the clause rectangle and the variable rectangle. Each such vertical segment is disjoint from all the rectangles except the two it connects. We call such an embedding a *rectangular embedding* of ϕ. See Figure 1 for an illustration. Given ϕ with a rectangular embedding, the goal of PM-3SAT is to decide if there exists a satisfying assignment for ϕ.

Suppose we are given a PM-3SAT instance consisting of the monotone 3-CNF formula ϕ with a rectangular embedding. Let $X = \{x_1, \ldots, x_n\}$ be the set of variables and $Z = \{z_1, \ldots, z_m\}$ be the set of clauses of ϕ (each of which consists of three literals). We construct (in polynomial time) a Euclidean minimax committee instance (E, k) in $\mathbb{R}^2$ such that ϕ has a satisfying assignment if and only if for the election E there exists a committee of size k with score at most 4. We begin by modifying the rectangular embedding of ϕ to another form that we call an *orthogonal embedding*. The details of the modifications are explained in the full version, but the resulting embedding satisfies the following three conditions: *(i)* no vertical segment crosses a horizontal segment; *(ii)* each horizontal segment s intersects exactly three vertical segments which correspond to the three variables contained in the clause corresponding to s; and *(iii)* the endpoints of all segments are *connection points*.

By properties *(ii)* and *(iii)*, there are three connection points on each clause segment, two of which are the left and right endpoints of the segment, and we call the middle one the *reference point* of the clause. We denote by $\hat{x}_1, \ldots, \hat{x}_n$ the reference points of the variables $x_1, \ldots, x_n$ and denote by $\hat{z}_1, \ldots, \hat{z}_m$ the reference points of the clauses $z_1, \ldots, z_m$. By shifting/scaling the segments properly without changing the topological structure of the orthogonal embedding, we can further guarantee that the x-coordinates (resp., y-coordinates) of the vertical (resp., horizontal) segments are distinct *even* integers in the range $\{1, \ldots, 2n\}$ (resp., $\{-2m, \ldots, 2m\}$). Therefore, all the connection points now have integral coordinates (which are even numbers) and the entire embedding is contained in the rectangle $[1, 2n] \times [-2m, 2m]$. Points on the segments of the orthogonal embedding that have integral coordinates partition each segment s into $\ell(s)$ unit-length segments, where $\ell(s)$ is the length of s. We call these unit-length segments the *pieces* of the orthogonal embedding. Let N be the total number of pieces. Clearly, $N = O(nm)$.

Our Euclidean minimax committee instance (E, k) consists of the following set of voters and candidates in the two-dimensional plane. (In fact, in our construction, each point is both a candidate and a voter, namely, $\mathcal{C} = V$. It is easy to modify the construction so that the set of voters is much larger by simply making multiple copies of each voter.)

Variable gadgets. For each variable x_i, we choose four points near the reference point $\hat{x}_i$ as follows. There are two (vertical) pieces incident to $\hat{x}_i$ in the orthogonal embedding, one above $\hat{x}_i$, the other below $\hat{x}_i$. On each of the two pieces, we choose two points with distances 0.01 and 0.02 from $\hat{x}_i$, respectively. We put a candidate and a voter at each of the four chosen points, and call these candidates/voters the x_i-*gadget*. We construct gadgets for all $x_1, \ldots, x_n$. The total number of candidates/voters in the variable gadgets is $4n$.

Clause gadgets. The second set of candidates/voters are constructed for the clauses $z_1, \ldots, z_m$. For each clause z_i, we put a candidate and a voter at the reference point $\hat{z}_i$, and call this candidate/voter the z_i-*gadget*. The total number of candidates/voters in the clause gadgets is m.

Piece gadgets. The last set of candidates/voters are constructed for connecting the variable gadgets and the clause gadgets. Consider a piece s of the orthogonal embedding, which is a unit-length segment. We distinguish the two endpoints of s as s^- and s^+ as follows. If s is a vertical piece above (resp., below) the x-axis, let s^- be the bottom (resp., top) endpoint of s and s^+ be the top (resp., bottom) endpoint of s. If s is a horizontal piece, then it must belong to the horizontal segment of some clause z_i. If s is to the left (resp., right) of the reference point $\hat{z}_i$, let s^- be the left (resp., right) endpoint of s and s^+ be the right (resp., left) endpoint of s.

For every piece s that is *not* adjacent to any clause reference point, we choose four points on s with distances 0.49, 0.8, 0.9, 1 from s^- (i.e., with distances 0.51, 0.2, 0.1, 0 from s^+), respectively. We put a candidate and a voter at each of the four chosen points, and call these the candidates/voters of the *s-gadget*. Note that we do not construct gadgets for the pieces that are adjacent to some clause reference point. Thus, the total number of candidates/voters in the piece gadgets is $4(N-3m)$, as each clause reference point is adjacent to three pieces.

By combining these three constructed gadgets, we obtain our election $E = (\mathcal{C}, V)$ instance with $4N + 4n - 11m$ candidates and voters. The size of the committee is $k = N + n - 3m$. We now prove that E has a committee of size k with score ≤ 4 iff ϕ is satisfiable.

The "if" part. Suppose ϕ is satisfiable and let $\pi : X \to \{\mathsf{true}, \mathsf{false}\}$ be an assignment which makes ϕ true. We construct a committee $T \subseteq \mathcal{C}$ of size k as follows. Our committee T contains one candidate in each variable gadget and each piece gadget (this guarantees $|T| = k$ as the total number of variable and piece gadgets is k). Consider a variable x_i. By our construction, the x_i-gadget contains four candidates which have the same x-coordinates as $\hat{x}_i$. If $\pi(x_i) = \mathsf{true}$ (resp., $\pi(x_i) = \mathsf{false}$), we include in T the topmost (resp., bottommost) candidate in the x_i-gadget. Now consider a piece s that is not adjacent to any clause reference point. We first determine a variable as the *associated* variable of s as follows. If s is vertical, then the associated variable of s is just defined as the variable whose vertical segment contains s. If s is horizontal, then s must belong to the horizontal segment of some clause z_j. In this case, we define the associated variable of s as the variable whose vertical segment intersects the left (resp., right) endpoint of the horizontal segment of z_j if s is to the left (resp., right) of the reference point $\hat{z}_j$. Let x_i be the associated variable of s. If $\pi(x_i) = \mathsf{true}$, then we include in T the candidate in the s-gadget that has distance 1 (resp., 0.9) from s^- if s is above (resp., below) the x-axis. Symmetrically, if $\pi(x_i) = \mathsf{false}$, then we include in T the candidate in the s-gadget that has distance 1 (resp., 0.9) from s^- if s is below (resp., above) the x-axis. This finishes the construction of T. The following lemma completes the "if" part of our proof.

Lemma 1. *The score of T in the election E is at most 4.*

The "only if" part. Suppose there exists a size-k committee $T \subseteq \mathcal{C}$ with score at most 4. We use that committee to construct a satisfying assignment $\pi : X \to \{\mathsf{true}, \mathsf{false}\}$. We first note the following property of the committee T.

Lemma 2. *T contains exactly one candidate in each variable gadget and exactly one candidate in each piece gadget.*

We note that the total number of variable and piece gadgets is k. Since $|T| = k$, using Lemma 2, we conclude that T has no budget to contain any candidate in the clause gadgets.

Corollary 1. *T contains no candidate in the clause gadgets.*

Recall that each variable gadget contains four candidates, two of which are above the x-axis while the other two are below the x-axis. Consider a variable x_i. By Lemma 2, T contains exactly one candidate in the x_i-gadget. If that candidate is above (resp., below) the x-axis, we set $\pi(x_i) = \mathsf{true}$ (resp., $\pi(x_i) = \mathsf{false}$). We show that π is a satisfying assignment of ϕ. It suffices to show that every positive (resp., negative) clause of ϕ contains at least one variable which is mapped to true (resp., false) by π. We only consider positive clauses, as the proof for negative clauses is similar. We need the following property of T.

Lemma 3. *Let s be a piece above the x-axis that is not adjacent to any clause reference point, and suppose x_i is the associated variable of s. If T contains the candidate in the s-gadget with distance 1 from s^-, then $\pi(x_i) = \mathsf{true}$.*

The rest of the argument to complete the "only if" part of our proof includes showing that for any positive clause z_i, π maps at least one variable of z_i to true. Due to limited space, we skip the details in the full version of the paper.

Finally, the reduction can clearly be done in polynomial time, and so we have established the following result.

Theorem 1. *Euclidean minimax committee is* NP-hard *in all dimensions $d \geq 2$. This claim holds even if the voter and candidate sets are identical.*

In fact, our construction also rules out the possibility of a PTAS as it is hard to decide in polynomial time whether the minimum score is ≤ 4 or ≥ 5. By slightly modifying the proof, we can also show that even computing an approximation for Euclidean minimax committee within any sublinear factor of $|\mathcal{C}|$ is hard.

Theorem 2. *For any constant $\epsilon > 0$, it is* NP-hard *to achieve a $|\mathcal{C}|^{1-\epsilon}$-approximation for Euclidean minimax committee in $\mathbb{R}^d$ for any $d \geq 2$.*

5 Approximately-Optimal Committees

We now complement the hardness results of the previous sections with nearly-optimal approximation algorithms.

5.1 Approximation Using ϵ-nets

Our first algorithm computes in polynomial-time a size-k committee of minimax score $O(m/k)$ for $d = 2$ and $O((m/k) \log k)$ for $d \geq 3$. Our algorithm uses the notion of *ϵ-nets*, which are commonly used in computational geometry [Toth *et al.*, 2017] for solving set cover and hitting set problems. Let us first briefly describe this notion.

Let X be a finite set of points in $\mathbb{R}^d$ and let $\mathcal{R}$ be a set of ranges (subsets of X) in $\mathbb{R}^d$. A subset $A \subseteq X$ is called an *ϵ-net* of $(X, \mathcal{R})$ if A intersects all those ranges in $\mathcal{R}$ that are ϵ-heavy, i.e., they contain at least an ϵ-fraction of the points in X. In other words, A is an ϵ-net for $(X, \mathcal{R})$ if $A \cap R \neq \emptyset$ for any $R \in \mathcal{R}$ with $|R \cap X| \geq \epsilon |X|$. There exists an ϵ-net of size $O(\frac{1}{\epsilon})$ for ranges defined by disks in $\mathbb{R}^2$, and of size $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$ for ranges defined by balls in $\mathbb{R}^d$, for any constant dimension $d \geq 3$ [Toth *et al.*, 2017]. In both cases, ϵ-nets can be computed in polynomial time.

Building on this result, we now present our algorithm.

Theorem 3. *Given a d-Euclidean election $E = (\mathcal{C}, V)$, we can compute in polynomial time a size k committee with minimax score $O(m/k)$ for $d = 2$ and score $O((m/k) \log k)$ for $d \geq 3$, where $m = |\mathcal{C}|$.*

Proof. In order to convey the intuition more clearly, let us first show how to find an $O(k)$-size committee with score at most $\lceil (m/k)\log k \rceil$. Given a d-Euclidean election, let $\mathcal{C}$ be the set of the m candidates with their embedding in $\mathbb{R}^d$. For each voter v, we consider a d-dimensional ball R_v centered at v containing the $\lceil (m/k)\log k \rceil$ closest points of $\mathcal{C}$ to v. Let $\mathcal{R}$ be the set of all these balls. Each ball of $\mathcal{R}$ is ϵ-heavy for $\epsilon = \log k/k$ because it contains an ϵ-fraction of the m candidates. Therefore, in polynomial time we can find an ϵ-net $T \subseteq \mathcal{C}$ for $(\mathcal{C}, \mathcal{R})$ of size $O(\frac{1}{\epsilon}\log\frac{1}{\epsilon}) = O(k)$. By the definition of ϵ-net, T contains at least one point from each R_v, and thus points of T form a committee of size $O(k)$ with minimax score $\lceil (m/k)\log k \rceil$.

To reduce the committee size to k while increasing the score by only a constant factor, we enlarge each ball R_v to include the $\alpha(m/k)\log k$ closest candidates of v, for an appropriate constant α. Each ball is now ϵ'-heavy, for $\epsilon' = \alpha \log k/k$, which guarantees an ϵ'-net $T \subseteq \mathcal{C}$ for $(\mathcal{C}, \mathcal{R})$ of size $O(\frac{1}{\epsilon'}\log\frac{1}{\epsilon'}) = O(k/\alpha)$. With an appropriate choice of α, we can ensure $|T| \leq k$ and achieve the score of $\alpha(m/k)\log k = O((m/k)\log k)$.

When $d \leq 2$, the ϵ-nets of this set system have size $O(\frac{1}{\epsilon})$, and therefore we can construct a committee of size k with score $O(m/k)$. □

The $O((m/k)\log k)$ and $O(m/k)$ bounds of Theorem 3 are essentially the best possible. In particular, we can construct instances of Euclidean elections in which no size-k committee can achieve the minimax score better than $\Omega(m/k)$.

Theorem 4. *For any $d \geq 1$, there exist Euclidean elections in $\mathbb{R}^d$ such that any committee $T \subseteq \mathcal{C}$ of size k has score $\Omega(m/k)$, where $m = |\mathcal{C}|$.*

On *all* the instances where $C \subseteq V$ (note that most representative elections do satisfy this condition because each candidate is also a voter), we can show that *any* committee has score $\Omega(m/k)$. In this case, the algorithm in Theorem 3 gives a *constant factor* approximation.

5.2 Approximation by Relaxing the Committee Size

The hardness result of Theorem 2 rules out any efficient algorithm for Euclidean election with a good approximation guarantee *but only under the rigid constraint that the committee size is at most k*. In this section, we study the problem in the setting where we are allowed to relax the committee size. Specifically, we ask the following natural question: *Can we efficiently compute a committee of size slightly larger than k whose score is close to the optimal score of a size-k committee?*

First, we show that if we are allowed to increase the committee size by a (small) *multiplicative* factor, then one can achieve (or improve) the optimal score of a size-k committee.

Theorem 5. *Given a d-Euclidean election, we can compute in polynomial time a committee of size $(1+\epsilon)k$, given any fixed $\epsilon > 0$, when $d = 2$ and of size $O(k \log m)$, where m is the number of candidates, when $d \geq 3$, whose score is smaller than or equal to the score of any size-k committee.*

Proof. We prove the result for a Euclidean election $E = (\mathcal{C}, V)$ in $d = 2$; the proof for higher dimensions is similar. Suppose we know that the optimal size-k committee of E has score $\sigma^\star$. We show how to compute a committee of size $(1+\epsilon)k$ whose score is at most $\sigma^\star$. For each voter $v \in V$, we consider the smallest disk R_v centered at v containing its closest $\sigma^\star$ candidates in $\mathcal{C}$. Let $\mathcal{R} = \{R_v : v \in V\}$. A *hitting set* of the set system $(\mathcal{C}, \mathcal{R})$ is a subset $H \subseteq \mathcal{C}$ such that $H \cap R \neq \emptyset$ for all $R \in \mathcal{R}$. If a committee has score at most $\sigma^\star$, then it must be a hitting set of $(\mathcal{C}, \mathcal{R})$, and since there is a size-k committee with score $\sigma^\star$, the minimum hitting set of $(\mathcal{C}, \mathcal{R})$ has size at most k. By using the PTAS for disk hitting set [Mustafa and Ray, 2010], we can compute a hitting set for $(\mathcal{C}, \mathcal{R})$ of size $(1+\epsilon)k$, which is the desired committee. Since we do not know the value of $\sigma^\star$, we simply try all values from 1 to m, and pick the smallest one for which we have a hitting set of size $(1+\epsilon)k$.

In higher dimensions, we can apply the same approach. The only difference is that we do not have a PTAS for ball hitting set in $\mathbb{R}^d$ for $d \geq 3$. But we can apply the greedy hitting set algorithm to compute a hitting set of size $O(k \log m)$ if a size-k hitting set exists. Therefore, the above algorithm computes a committee of size $O(k \log m)$ whose score is at most the score of any size-k committee. □

On the other hand, we prove that if we can only increase the committee size by an *additive* constant, we are not able to achieve any good approximation for the minimax score. (The proof of this theorem is a minor modification of Theorem 2 and is presented in the full version of the paper.)

Theorem 6. *Let $\alpha, \epsilon > 0$ be constants. Given a Euclidean election $E = (\mathcal{C}, V)$ in $\mathbb{R}^d$ for $d \geq 2$ and a number $k \geq 1$, it is* NP-hard *to compute a committee of size $k + \alpha$ whose score is at most $|\mathcal{C}|^{1-\epsilon} \cdot \sigma^\star$, where $\sigma^\star$ is the minimum score of a size-k committee.*

5.3 Minimax Committees for the r-Borda Rule

A natural generalization of the Chamberlin Courant rule is the so-called *r-Borda rule* where the score of each voter is determined by its nearest r candidates in the committee for a given $r \leq k$. More specifically, the score of a voter v with respect to a committee is the sum of the ranks of its nearest r candidates in the committee in the preference list of v. The minimax score of a committee T is the maximum over all voter scores. Our goal is to find a committee T of size k that minimizes $\sigma(T)$, where

$$\sigma(T) = \max_{v \in V}\left(\min_{Q \subseteq T, |Q| = r}\left(\sum_{c \in Q}\sigma_v(c)\right)\right).$$

We show that for any election $E = (\mathcal{C}, V)$ in $\mathbb{R}^d$, we can compute a size-k committee with an r-Borda score of $O((r^2 m/k)\log k)$ using a modification of the algorithm in Theorem 3. In $\mathbb{R}^2$, the score of the committee can be further improved to $O(r^2 m/k)$ using the same approach with ϵ-nets of size $O(1/\epsilon)$ for disks. Thus, we have the following result.

Theorem 7. *Given an election in any fixed dimension d, we can find in polynomial time a size-k committee with minimax*

r-Borda score $O((r^2m/k)\log k)$. Furthermore, if $d \leq 2$, the score can be further improved to $O(r^2m/k)$.

We observe that the bound $O(r^2m/k)$ in the above theorem is tight. In particular, there are instances for which an optimal committee's r-Borda score is $\Omega(r^2m/k)$—this can be verified using the instance described in the proof of Theorem 4—and so this serves as the benchmark score for r-Borda.

5.4 δ-Optimal Committees

In the previous section, we showed that for any instance we can find a minimax committee of optimal score if we increase the committee size by a small (multiplicative) factor. In this section, we suggest an alternative way to assess the approximation quality *while keeping the committee size k.*

To introduce this criterion, let us consider an election $E = (\mathcal{C}, V)$ and suppose the optimal score of a size-k committee is $\sigma^\star$. In our approximation, we are looking for a size-k committee in which the candidate closest to each voter $v \in V$ has rank not much larger than $\sigma^\star$ in the preference list of v. Our hardness proof shows that in general this is not possible because there may be many candidates at roughly the same distance from v, *but with a large difference in ranks*, and any polynomial time algorithm is bound to end up with a bad minimax score for some v. A natural way to get rid of this pathological situation is to treat two candidates with roughly the same distance from v as if they have similar ranks.

With this motivation, we introduce the following *δ-optimality criterion*. We say a committee $T \subseteq \mathcal{C}$ of size k is *δ-optimal*, for $\delta \geq 1$, if for each voter $v \in V$ the distance from v to its closest candidate in T is at most δ times the distance from v to its rank-$\sigma^\star$ candidate.

We now show how to compute a 3-optimal committee in polynomial time for any d-Euclidean election. Let $E = (\mathcal{C}, V)$ be a Euclidean election and $k \geq 1$ be the desired committee size. For convenience, let us first assume that the optimal score $\sigma^\star$ of a size-k committee of E is known. For each voter $v \in V$, define $d_v^\star$ as the distance from v to the rank-$\sigma^\star$ candidate in the preference list of v. We say a voter v is *satisfied* with a subset $T \subseteq \mathcal{C}$ if there exists a candidate $c \in T$ such that $\text{dist}(c, v) \leq 3d_v^\star$. We denote by $S[T] \subseteq V$ the subset of voters satisfied with T. Then a committee $T \subseteq \mathcal{C}$ (of size k) is 3-optimal if every $v \in V$ is satisfied with T. Our algorithm begins with an empty committee $T = \emptyset$ and iteratively adds new candidates to T using the following three steps until $S[T] = V$:

1. $\hat{v} \leftarrow \arg\min_{v \in V \setminus S[T]} d_v^\star$.
2. $\hat{c} \leftarrow$ a candidate within distance $d_{\hat{v}}^\star$ from $\hat{v}$.
3. $T \leftarrow T \cup \{\hat{c}\}$.

In words, in each iteration, we find the *unsatisfied* voter $\hat{v}$ with the minimum $d_{\hat{v}}^\star$, and then add to T a (arbitrarily chosen) candidate $\hat{c} \in \mathcal{C}$ within distance $d_{\hat{v}}^\star$ from $\hat{v}$. The algorithm terminates when $S[T] = V$, and so all voters are satisfied with T at the end.

We only need to show that $|T| \leq k$; if $|T| < k$, we can always add extra candidates while keeping all voters satisfied. Consider an optimal size-k committee $T_{\text{opt}} = \{c_1, \ldots, c_k\}$, with score $\sigma^\star$, and let $V_i \subseteq V$ be the subset of voters whose closest candidate in T_{opt} is c_i. Thus, $V_1, \ldots, V_k$ form a partition of V. We say two voters $v, v' \in V$ are *separated* if they belong to different V_i's.

Suppose our algorithm terminates in r iterations. We will show that $r \leq k$. Let $\hat{v}_j$ (resp., $\hat{c}_j$) be the voter $\hat{v}$ (resp., the candidate $\hat{c}$) chosen in the j-th iteration, and let T_j be the committee T at the beginning of the j-th iteration. We claim the following property of our greedy algorithm.

Lemma 4. *The voters $\hat{v}_1, \ldots, \hat{v}_r$ are pairwise separated.*

Thus, the voters $\hat{v}_1, \ldots, \hat{v}_r$ belong to different V_i's, which implies that $r \leq k$ and $|T| = r \leq k$, proving the correctness of our algorithm.

Finally, notice that in our algorithm we assumed $\sigma^\star$ is known, but this assumption is easy to get rid of. We can try each possible value from 1 to $m = |\mathcal{C}|$, and choose the smallest number $\sigma^\star \in [m]$ for which the algorithm returns a committee of size at most k. Thus, we proved the following.

Theorem 8. *Given a Euclidean election $E = (\mathcal{C}, V)$ in any dimension and $k \geq 1$, one can compute a 3-optimal committee of size k in polynomial time.*

Complementary to the above algorithmic result, we can also show the following hardness result.

Theorem 9. *For any $\delta < 2$, unless P=NP, there is no polynomial time algorithm to compute a δ-optimal committee of size k for a given Euclidean election in $\mathbb{R}^d$ for $d \geq 2$.*

6 Closing Remarks

We studied the multiwinner elections in Euclidean space under minimax Chamberlin-Courant voting rules. First, we settle the complexity of the problem by showing it is NP-hard for dimensions $d \geq 2$ (in contrast, the problem is known to be polynomial time solvable in 1d), but our main contribution is presenting several attractive (& nearly-optimal) approximation bounds which are elusive in the general setting. We believe that our algorithms are robust and will generalize to many other interesting questions, for instance, most of our algorithmic results (except for Thm. 5) extend to the recently studied egalitarian k-median rules [Gupta *et al.*, 2021] when considered for the Euclidean elections.

Our work suggests many natural directions, including resolving the complexity and approximation bounds for other important voting rules, such as the utilitarian variant of the CC rule and the general class of OWA rules [Skowron *et al.*, 2015a; Skowron *et al.*, 2016] for the Euclidean elections. Investigating the parameterized complexity of the problems we considered is also an interesting question since natural parameters such as the committee size or the number of candidates are small in many real-world elections. Another interesting avenue is to explore Euclidean elections when the positions of voters and candidates are only approximately known, for instance, each placed at some unknown point in a disk. Such data uncertainty naturally exists in many real-world applications.

References

[Aziz *et al.*, 2018] Haris Aziz, Piotr Faliszewski, Bernard Grofman, Arkadii Slinko, and Nimrod Talmon. Egalitarian committee scoring rules. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 56–62. International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[Betzler *et al.*, 2013] Nadja Betzler, Arkadii Slinko, and Johannes Uhlmann. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research*, 47:475–519, 2013.

[Brandt *et al.*, 2016] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

[Byrka *et al.*, 2018] Jaroslaw Byrka, Piotr Skowron, and Krzysztof Sornat. Proportional approval voting, harmonic k-median, and negative association. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107, page 26. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.

[Chamberlin and Courant, 1983] John R Chamberlin and Paul N Courant. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, 77(3):718–733, 1983.

[de Berg and Khosravi, 2010] Mark de Berg and Amirali Khosravi. Optimal binary space partitions in the plane. In *International Computing and Combinatorics Conference*, pages 216–225. Springer, 2010.

[Elkind *et al.*, 2017a] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3):599–632, 2017.

[Elkind *et al.*, 2017b] Edith Elkind, Martin Lackner, and Dominik Peters. Structured preferences. *Trends in computational social choice*, pages 187–207, 2017.

[Endriss, 2017] Ulle Endriss. Trends in computational social choice. https://Lulu.com, 2017. Accessed: 2021-06-14.

[Faliszewski *et al.*, 2017] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice*, 74(2017):27–47, 2017.

[Faliszewski *et al.*, 2019] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Committee scoring rules: Axiomatic characterization and hierarchy. *ACM Transactions on Economics and Computation (TEAC)*, 7(1):1–39, 2019.

[Godziszewski *et al.*, 2021] Michał T. Godziszewski, Paweł Batko, Piotr Skowron, and Piotr Faliszewski. An analysis of approval-based committee rules for 2d-euclidean elections. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5448–5455, 2021.

[Goel *et al.*, 2019] Ashish Goel, Anilesh K Krishnaswamy, Sukolsak Sakshuwong, and Tanja Aitamurto. Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation (TEAC)*, 7:1–27, 2019.

[Gupta *et al.*, 2021] Sushmita Gupta, Pallavi Jain, Saket Saurabh, and Nimrod Talmon. Even more effort towards improved bounds and fixed-parameter tractability for multiwinner rules. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, pages 217–223, 2021.

[Kleinberg and Tardos, 2006] Jon Kleinberg and Eva Tardos. *Algorithm design*. Pearson Education India, 2006.

[Lu and Boutilier, 2011] Tyler Lu and Craig Boutilier. Budgeted social choice: From consensus to personalized decision making. In *Twenty-Second International Joint Conference on Artificial Intelligence*, pages 280–286, 2011.

[Munagala *et al.*, 2021] Kamesh Munagala, Zeyu Shen, and Kangning Wang. Optimal algorithms for multiwinner elections and the chamberlin-courant rule. *EC '21: Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 697–717, 2021.

[Mustafa and Ray, 2010] Nabil H Mustafa and Saurabh Ray. Improved results on geometric hitting set problems. *Discrete & Computational Geometry*, 44:883–895, 2010.

[Peters, 2017] Dominik Peters. Recognising multidimensional euclidean preferences. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 642–648, 2017.

[Skowron *et al.*, 2015a] Piotr Skowron, Piotr Faliszewski, and Arkadii Slinko. Achieving fully proportional representation: Approximability results. *Artificial Intelligence*, 222:67–103, 2015.

[Skowron *et al.*, 2015b] Piotr Skowron, Lan Yu, Piotr Faliszewski, and Edith Elkind. The complexity of fully proportional representation for single-crossing electorates. *Theoretical Computer Science*, 569:43–57, 2015.

[Skowron *et al.*, 2016] Piotr Skowron, Piotr Faliszewski, and Jérôme Lang. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence*, 241:191–216, 2016.

[Sonar *et al.*, 2022] Chinmay Sonar, Subhash Suri, and Jie Xue. Multiwinner elections under minimax chamberlin-courant rule in euclidean space. *CoRR*, abs/2205.13598, 2022.

[Toth *et al.*, 2017] Csaba D Toth, Joseph O'Rourke, and Jacob E Goodman. *Handbook of discrete and computational geometry*. CRC press, 2017.

[Yu *et al.*, 2013] Lan Yu, Hau Chan, and Edith Elkind. Multiwinner elections under preferences that are single-peaked on a tree. In *Twenty-Third International Joint Conference on Artificial Intelligence*, pages 425–431, 2013.