UNIVERSITY OF MARYLAND
COLLEGE OF COMPUTER, MATHEMATICAL, AND NATURAL SCIENCES
Department of Computer Science

Department of Computer Science
8125 Paint Branch Drive
College Park, Maryland 20742
301-405-2662

MS Scholarly Paper Approval Form

Gholami, Iman	Jan 2023
Student's Full Name (Last, First, Middle)	Month & Year Started (Entry Year)
MS Along the Way	Spring 2025
Name of Graduate Program	Graduation Term (Term/Year)

Title of MS Scholarly Paper:

Bi-Criteria Metric Distortion

The purpose of the Scholarly Paper to demonstrate a student's research abilities, mirroring the standards of a conference or journal paper in their respective field. The paper could be based on:

- An ongoing research project by the student
- A paper submitted by the student to a conference (co-authorship is permitted)
- A class project undertaken by the student
- A comprehensive research survey within area of research

I have thoroughly reviewed the paper and deem it satisfactory.

Mohammad Hajiaghayi	*Mohammad Hajiaghayi*	4/12/2025
Faculty Name	Signature	Date

Bi-Criteria Metric Distortion

Kiarash Banihashem* Diptarka Chakraborty† Shayan Chashm Jahan‡ Iman Gholami§
MohammadTaghi Hajiaghayi¶ Mohammad Mahdavi‖ Max Springer**

Abstract

Selecting representatives based on voters' preferences is a fundamental problem in social choice theory. While cardinal utility functions offer a detailed representation of preferences, ordinal rankings are often the only available information due to their simplicity and practical constraints. The metric distortion framework addresses this issue by modeling voters and candidates as points in a metric space, with distortion quantifying the efficiency loss from relying solely on ordinal rankings. Existing works define the cost of a voter with respect to a candidate as their distance and set the overall cost as either the sum (utilitarian) or maximum (egalitarian) of these costs across all voters. They show that deterministic algorithms achieve a best-possible distortion of 3 for any metric when considering a single candidate.

This paper explores whether one can obtain a better approximation compared to an optimal candidate by relying on a committee of k candidates ($k \geq 1$), where the cost of a voter is defined as its distance to the closest candidate in the committee. We answer this affirmatively in the case of line metrics, demonstrating that with $O(1)$ candidates, it is possible to achieve optimal cost. Our results extend to both utilitarian and egalitarian objectives, providing new upper bounds for the problem. We complement our results with lower bounds for both the line and 2-D Euclidean metrics.

1 Introduction

One of the fundamental challenges in the social choice theory is to elect representatives based on voters' preferences, ideally represented by cardinal utility functions that assign numerical values to each outcome. However, in most real-world scenarios, voters only provide ordinal information, such as preference orders among outcomes/candidates. This raises a natural question of how worse, if at all, a voting mechanism performs given ordinal information than cardinal information. Procaccia and Rosenschein [29] introduced the notion of *distortion* to measure such an efficiency loss – how different voting rules respond to the lack of cardinal information. Many practical voting scenarios can be formulated by considering both voters and candidates lying on a metric space (see [15]). The distance to candidate locations determines voters' cardinal preferences for candidates – voters rank candidates based on ascending distance, with the closest candidate being the most preferable and the farthest candidate being the least preferable. The worst-case behavior of any ordinal preference order-based voting rule/mechanism is captured by the notion of *metric distortion*, introduced by Anshelevich *et al.* [3]. A voting mechanism, without access to the actual distances among the set of voters and candidates, seeks to minimize a specific cost function, which depends on the distances.

*Computer Science Department, University of Maryland, College Park. kiarash@umd.edu.
†School of Computing, National University of Singapore. diptarka@comp.nus.edu.sg.
‡Computer Science Department, University of Maryland, College Park. schjahan@umd.edu.
§Computer Science Department, University of Maryland, College Park. igholami@umd.edu.
¶Computer Science Department, University of Maryland, College Park. hajiagha@umd.edu.
‖Computer Science Department, University of Maryland, College Park. mahdavi@umd.edu.
**Mathematics Department, University of Maryland, College Park. mss423@umd.edu.

Distortion is defined in relation to this cost: For any voting mechanism f, its distortion is the worst-case ratio (across all instances) of the cost of the solution produced by f compared to the optimal cost.

Given a single fixed candidate, the cost for a voter is defined as its distance from the given candidate. Then, the overall cost is set to be an objective that combines these values across all the voters. Depending on the specific context in the literature, the following two objectives have widely been considered: A utilitarian objective, which aims to minimize the total individual costs for all voters, and an egalitarian cost, which minimizes the maximum cost experienced by any voter. Different variants of the metric distortion problem under the above objective functions have received significant attention, e.g. [22, 23, 20, 2, 24].

For the classical metric distortion problem, it is known that a distortion of 3 can be achieved for any metric. Further, it is widely known that the candidate chosen by any deterministic method cannot achieve a distortion factor of less than 3, even in a line metric. In other words, without knowing the exact distances, it is not possible to obtain a better than 3-approximation to the optimal cost. This leads to an intriguing question: Can we obtain a better approximation (distortion) by selecting more than one candidate? Specifically, assume that the algorithm is allowed to choose $k > 1$ candidates, and we set the cost for each voter to be its distance to the closest chosen candidate. Can we design an algorithm for which the overall cost (either utilitarian or egalitarian) is at most α times the cost of an optimal candidate for some $\alpha < 3$?

In this paper, we answer the above question affirmatively for the line metric. We not only attain better distortion by allowing the selection of more than one candidate but, in fact, attain optimal cost with $O(1)$ candidates. We complement our upper-bound results with various lower-bound constructions showing impossibility results in line and 2-D Euclidean metric.

Our work provides a *bicriteria* perspective to metric distortion, which, to our knowledge, has not been considered before. The pursuit of improved *bi-criteria* approximation results for various classical optimization problems has already been well-investigated in the literature. Indeed, when the metric is known, and the goal is to pick k candidates that minimize the overall cost across voters, the election problem becomes an instance of either the k-median or the k-center clustering (for the utilitarian and egalitarian objective, respectively). For these problems, numerous (constant-factor) approximation algorithms are known which select up to $O(k)$ centers (instead of k centers), e.g. [18, 31, 1]. Therefore, it is quite natural to explore a similar question in the metric distortion problem, where the underlying metric is not directly given.

Our work additionally extends the existing line of research on k-committee elections [17, 13, 7], which also select a committee of k candidates and aim to minimize some loss function across all voters. The key distinction is that we consider a single candidate as our baseline for calculating distortion, while these works consider a baseline of k candidates (see Section 1.2). Specifically, for the utilitarian cost we consider (i.e., the sum of distances of voters to their closest candidate in the committee), Caragiannis, Shah, and Voudouris [7] show that the ratio of the cost for any deterministic method compared to the cost of an optimal committee can be unbounded in the worst case. This result fails to suggest a choice of candidates, as all possible choices are the same in the worst case. In contrast, we show that by choosing the right baseline, one can make a meaningful distinction between different choices even though the underlying metric is unknown.

1.1 Our contribution

Our first result shows that when the underlying metric is a line, a committee of two candidates can achieve a 1-distortion of 1 under the *sum-cost* objective. Here, *1-distortion* refers to the worst-case ratio between the cost of the selected committee and the cost of an optimal single-winner candidate. In fact, we will prove a stronger result by showing that it is possible to output a list of 2 candidates that always contains an optimal one. Formally, we prove the following theorem.

Theorem 1. There exists an algorithm for the 2-committee utilitarian election on the line metric that is guaranteed to choose an optimum candidate in the elected committee. Consequently, the 1-distortion of the algorithm is 1.

We further show that when the metric is not a line, one can obtain a distortion of $1 + \frac{2}{m-1}$.

Theorem 2. There exists an algorithm for the $(m-1)$-committee utilitarian election on the general metric that obtains a 1-distortion of, at most $1 + 2/(m-1)$. Additionally, no algorithm can obtain a 1-distortion better than $1 + 2/(m-1)$ when choosing $m-1$ candidates, even if the metric space is 2-D Euclidean.

Note that the above theorem immediately implies that no algorithm can find a set of size $m-1$ guaranteed to contain an optimum, even if the underlying metric is 2-D. We further study the *max-cost* objective, showing that one can obtain 1-distortions of 1, 1.5, and 2 using sets of size four, three, and two, respectively.

Theorem 3. For any $k \in \{2, 3, 4\}$, there exists an algorithm for the k-committee egalitarian election on the line metric, which obtains a 1-distortion of at most $3 - k/2$. Furthermore, there is no algorithm that can obtain a 1-distortion better than $3 - k/2$.

For the *max-cost* objective, even though there exists an algorithm that selects one candidate with a distortion of 3, we show that no algorithm can achieve a 1-distortion better than 3, even when choosing a committee of $m-1$ candidates.

Theorem 4. There is no algorithm for the $(m-1)$-committee egalitarian election on 2-D Euclidean metric such that can obtain a 1-distortion better than 3.

It is important to note that the obtained lower bounds apply to deterministic algorithms.

Objective	Metric Space	Committee Size (Out of m)	Lower-Bound	Upper-Bound
Sum	1D	≥ 2	1	1
		1	3[3]	3[20]
	2D	$m-1$	$1 + \frac{2}{m-1}$	$1 + \frac{2}{m-1}$
Max	1D	≥ 4	1	1
		3	1.5	1.5
		2	2	2
		1	3	3
	2D	$m-1$	3	3

Table 1: State-of-the-Art; Blue colored bounds are results of this paper

Our approach to the problem involves many novel techniques that we believe are of independent interest. Most notably, for the line metric, we propose an algorithm that essentially finds the order of candidates and voters. Specifically, we identify the order for all the candidates in a set of *core candidates*, which has the following property: If c_i is a core candidate and c_j is not, then all voters prefer c_i to c_j. We additionally find the order of all voters with the minor caveat that we may potentially have ties for voters who have the same exact preference for all core candidates. We refer to Section 3 for more details.

1.2 Related works

Since its introduction, the metric distortion framework has remained central in investigating the performance of single-winner voting. Gkatzelis, Halpern, and Shah [20] showed the existence of a deterministic mechanism with a 3-distortion factor, settling a long-standing open question, a simpler proof of which was later given by Kizilkaya and Kempe [25]. Although a distortion factor of 3 is unavoidable for any deterministic mechanism, an intriguing question arises about whether randomization could surpass this 3-factor barrier. In fact, it was conjectured that a randomized mechanism could achieve a distortion of 2. However, Charikar and Ramakrishnan [9], and Pulyassary and Swamy [30] independently established the non-existence of such a randomized mechanism, refuting the conjecture. The quest of breaking below 3-factor using randomization remained. Charikar *et al.* [10] recently answered this positively by developing a randomized mechanism with a 2.753-distortion factor. On a different line of work, Anshelevich *et al.* [4] demonstrate that when the threshold approval set of each voter is known – containing all candidates whose cost is within an appropriately chosen factor of the voter's cost for their most preferred candidate – a distortion of $1 + \sqrt{2}$ can be achieved.

In the k-committee election problem (the single-winner election being a special case with $k = 1$), the aim is to select k candidates from a pool of m candidates based on ordinal preferences provided by n voters. For any mechanism f, its distortion is defined as the worst-case ratio (across all instances) of the cost of the solution produced by f compared to the optimal cost. When the cost for a voter is considered as the sum of distances to all committee members, Goel, Hulett, and Krishnaswamy [21] showed that the problem reduces to the single-winner election.

On the other hand, Caragiannis, Shah, and Voudouris [7] considered a general cost function – each voter's cost is the distance to the q-th (for some integer $q \geq 1$) nearest committee member. They identified a trichotomy: For $q \leq k/3$, the distortion is unbounded; for $q \in (k/3, k/2]$, it is $\Theta(n)$; and for $q > k/2$, the problem reduces to the single-winner election. As an immediate corollary, for the 2-committee election, with each voter's cost being its distance to the nearest committee member (i.e., $q = 1$), we get a distortion of $\Theta(n)$. Further, with the same cost function, for the k-committee election when $k \geq 3$, the distortion is unbounded (even for line metric). However, when the positions of candidates are known, for $k = m - 1$, Chen, Li, and Wang [11] demonstrated that single-vote rules achieve a distortion of 3 and provided a matching lower bound.

One of the most basic versions – where both voters and candidates are positioned on a real line – has already garnered significant attention in computational social choice theory. If we know the locations of the voters and candidates on this real line, a straightforward dynamic programming approach can solve the k-committee election problem optimally in $O(nk \log n)$ time. Strict preference profiles, with voters and candidates on a real line (also known as 1-D Euclidean), exhibit many intriguing properties, including being single-peaked and single-crossing [5, 28, 16]. Given a preference profile and the order of voters, deciding whether it is 1-D Euclidean can be done in polynomial time [12]. Furthermore, if the input preference order is consistent with 1-D Euclidean, [12] provides an efficient construction of a mapping realizing that.

Another closely related question to the problem of optimal candidate selection is the facility location problem [27], where the goal is to place facilities at locations in a metric space to minimize the cost of serving agents. Unlike the candidate selection problem, where candidates are restricted to a fixed set, facilities in the facility location problem can be placed anywhere in the space [19].

A related solution concept to the matroid we propose is the Condorcet winning set: a set of candidates such that no other candidate is preferred by at least half the voters over every member of the set [14]. The Condorcet dimension, defined as the minimum cardinality of a Condorcet winning set, is known to be at most logarithmic in the number of candidates. Caragiannis *et al.* [6] partially reaffirm this logarithmic bound when considering committee-selected alternatives. In the metric ranking framework, Lassota *et al.* [26] recently demonstrated that the Condorcet dimension is at most three under the Manhattan or ℓ_∞ norms. Independently and concurrently, Charikar *et al.* [8] showed that Condorcet sets of size six always exist.

2 Preliminaries

Metric space. Let us consider a domain $\mathcal{X}$ and a distance function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. We call $(\mathcal{X}, d)$ a *metric space* if the distance function d satisfies the following properties:

- **Positive definite**: For all $x, y \in \mathcal{X}$, $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$.
- **Symmetry**: For all $x, y \in \mathcal{X}$, $d(x, y) = d(y, x)$.
- **Triangle inequality**: For all $x, y, z \in \mathcal{X}$, $d(x, y) \leq d(x, z) + d(z, y)$.

In this paper, we consider

- **Line metric (1-D Euclidean metric)**: The domain is $\mathcal{X} = \mathbb{R}$, and for any two points $p, q \in \mathbb{R}$, their distance is $d(p, q) = |p - q|$.
- **2-D Euclidean metric**: The domain is $\mathcal{X} = \mathbb{R}^2$, and for any two points $p = (p_x, p_y), q = (q_x, q_y) \in \mathbb{R}^2$, their distance is $d(p, q) = \|p - q\|_2 := \left((p_x - q_x)^2 + \left(p_y - q_y\right)^2\right)^{1/2}$.

Election. An *election instance* $\mathcal{E} = (V, C, \succ)$ consists of a set $V = \{v_1, \dots, v_n\}$ of n voters and a set $C = \{c_1, \dots, c_m\}$ of m candidates. Each voter $v_i \in V$ has a linear order $\succ_i$ over the candidates, where $c_j \succ_i c_k$ indicates that voter v_i prefers c_j over c_k. We refer to $\succ_i$ as the *ordinal preference* of voter v_i. Furthermore, $\succ = \{\succ_1, \dots, \succ_n\}$ is called the *preference profile* of the voters. Additionally, the j-th candidate in the ordinal preference of voter v_i is denoted by $\succ_{i,j}$.

We consider the voters and candidates to lie in the same metric space $(\mathcal{X}, d)$. For ease of exposition, we extend the notion of distance d to be defined directly on the set of voters and candidates instead of the points they occupy in the underlying space $\mathcal{X}$. We say a (distance) metric d is *consistent* with an election instance $\mathcal{E} = (V, C, \succ)$, denoted as $d \triangleright \mathcal{E}$, when for any voter v_i, $c_j \succ_i c_k$ if $d(v_i, c_j) \leq d(v_i, c_k)$.

Social cost. Let us consider an election instance $\mathcal{E} = (V, C, \succ)$, and a distance metric $d \triangleright \mathcal{E}$. Let $I = (\mathcal{E}, d)$ denote the instance $\mathcal{E}$ with d being its underlying distance metric. For any subset of candidates $S \subseteq C$, and a voter $v \in V$, we use $d(v, S)$ to denote the distance between the voter v to its nearest neighbor in S, i.e., $d(v, S) := \min_{c \in S} d(v, c)$. In this paper, we focus on the following two *social costs*:

- **Sum-cost (Utilitarian objective)**: For any subset of candidates $S \subseteq C$, its *sum-cost*, denoted by $\texttt{cost}_s(S, I)$ is defined as $\texttt{cost}_s(S, I) := \sum_{v \in V} d(v, S)$.
- **Max-cost (Egalitarian objective)**: For any subset of candidates $S \subseteq C$, its *max-cost*, denoted by $\texttt{cost}_m(S, I)$ is defined as $\texttt{cost}_m(S, I) := \max_{v \in V} d(v, S)$.

When it is clear from the context, we drop I and simply use $\texttt{cost}_s(S)$ and $\texttt{cost}_m(S)$.

Voting rule and distortion. A (deterministic) *voting rule* (also referred to as *mechanism*) f is a function that maps an election instance $\mathcal{E}$ to a subset of candidates S. We use algorithms and mechanisms interchangeably throughout this paper. In this paper, we are interested in comparing the cost of voting rules that selects k-committees with the cost of an optimal single candidate. We call a single candidate c_{opt} *optimal* if

$$\texttt{cost}(c_{\text{opt}}) = \min_{c \in C} \texttt{cost}(c).$$

Throughout the paper we use $\texttt{OPT} = \texttt{cost}(c_{\text{opt}})$ to refer to the cost of an optimal single candidate.

To capture how good a voting rule is in the worst case, the notion of distortion is used. For any voting rule f, its *distortion*, or more specifically, *1-distortion* is defined as

$$\texttt{1-distortion}(f) := \sup_{\mathcal{E}} \sup_{d \rhd \mathcal{E}} \frac{\texttt{cost}(f(\mathcal{E}))}{\texttt{OPT}}$$

where the cost function $\texttt{cost}$ in the above definition could be either $\texttt{cost}_s$ or $\texttt{cost}_m$ depending on the context. In other words, the 1-distortion compares the cost of the mechanism to the cost of an optimal candidate in the worst case.

3 Line metric election: the order of candidates and voters

In this section, we present an algorithm designed to determine the order of candidates and voters for any line metric election instance $\mathcal{E} = (V, C, \succ)$ based on the voters' preference profile. This step is essential for establishing the upper bounds discussed in Section 4. Moreover, this approach may prove useful for future research, as it offers a general method for obtaining the total order of candidates and voters as explained in the following.

This algorithm focuses on a specific subset of candidates with properties useful for the purposes of this paper and potential future work on the line metric distortion.

Definition 5 (Core)**.** In an election $\mathcal{E} = (V, C, \succ)$, a subset $A \subseteq C$ is called *a core* if for any voter v_i, any candidate $c_j \in A$, and any candidate $c_k \in C \setminus A$, we have $c_j \succ_i c_k$.

The algorithm introduced in this section determines the order only for a subset C^*, referred to as the *determined candidates*, rather than for all candidates. It expands C^* iteratively and establishes the order of candidates within C^*. By the end of the algorithm, C^* forms a core subset of candidates.

Additionally, for some voters with similar preferences, it may be impossible to distinguish whether they are on the right or left; thus, they are considered to be at the same point. These properties of the voter and candidate order retrieved by the algorithm are accounted for in the analysis and other sections of this paper. Overall, obtaining the order of voters and candidates with this level of accuracy remains significant for the intended purposes.

The algorithm consists of three parts. The first part, called SplitLine, involves dividing the line into two halves and determining whether each candidate occurs on the left or right side (some of them remain undetermined). The second part, denoted as SortCandidates, finds the order of candidates in each half and merges the sorted lists to obtain a total ordering. The final part, referred to as SortVoters, determines the voters' ordering based on the retrieved order of candidates. Algorithm 1 demonstrates how these three components contribute to sorting the candidates and voters.

SplitLine. In this part, we first find a pivot to split the line at that point. To achieve this, we introduce the following definitions. We arbitrarily choose the first voter as the pivot voter, associated with two pivot candidates as follows.

Definition 6 (Pivot Voter and Candidates)**.** The voter v_1 is called the *pivot voter*. Additionally, the two nearest candidates to the pivot voter are called the *pivot candidates*. Without loss of generality, assume that c_1 and c_2 are the two nearest candidates to the pivot voter v_1, with c_1 positioned to the left of c_2. Note that c_1 and c_2 may both be on the same side of v_1.

Algorithm 1 Sort candidates and voters

Input: Election instance $\mathcal{E}$.

Output: Sequence S_C where is the order of determined candidates and sequence S_V the order of voters.

function SortCandiatesAndVoters($\mathcal{E} = (V, C, \succ)$)
  $(L, R) \leftarrow$ SplitLine($\mathcal{E}$)
  $S_C \leftarrow$ SortCandidates($\succ_1, L, R$)
  $S_V \leftarrow$ SortVoters($\mathcal{E}, S_C$)
  return (S_C, S_V)

Finally, the point where the line is split is defined as follows:

Definition 7 (Pivot Point)**.** The midpoint of the line segment between the two pivot candidates is called the *pivot point*, denoted by p. Let L and R be the subsets of candidates on the left and right sides of p, respectively.

Finally, We aim to determine whether a candidate belongs to L or R. Therefore, we formally define determined and undetermined candidates as follows:

Definition 8. A candidate is *determined* if it is known whether the candidate belongs to L or R based on Definition 7. Otherwise, the candidate is *undetermined*. Let $C^* = L \cup R$ be the set of determined candidates

Initially, based on Definition 6, we know that $c_1 \in L$ and $c_2 \in R$. Thus, $C^* = \{c_1, c_2\}$. Through several iterations, we expand L and R by adding as many candidates as possible. Therefore, at the end of each iteration we update C^* such that $C^* = L \cup R$ again. We also ensure that C^* always forms a consecutive subset of candidates on the line. The process for determining new candidates in each iteration is as follows:

If there exists a voter v_i such that two candidates c_k and c_j satisfy $c_k \succ_i c_j$, where $c_j \in C^*$ but $c_k \notin C^*$, then we can determine c_k's membership as follows:

- If both c_1 and c_2 are closer to v_i than c_j, then c_k belongs to the opposite side of c_j.
- Otherwise, c_k belongs to the same side as the candidate in $\{c_1, c_2\}$ that is closer to v_i.

Algorithm 2 is pseudocode for the function Determine, which determines a candidate if the above conditions hold and adds it to the corresponding set. Algorithm 3 determines as many candidates as possible in each iteration while maintaining the succession of the determined candidates (see 3.1 for proofs). We call these candidates C_{new}. At the end of the iteration, it merges C_{new} into C^*. It is important to note that we do not add each point immediately to C^*; instead, we merge them all at the end. This approach ensures that C^* remains a consecutive list of candidates, which is crucial for our analysis.

The output of this function is L and R, which represent the determined candidates on the left and right sides of the pivot point p, respectively.

SortCandidates. The goal of this part is to sort the candidates in L and R. We know that L lies to the left of p, and R lies to the right of p, where p is the midpoint of the segment connecting the two nearest candidates of v_1 (see Definitions 6 and 7). Consequently, in the ordinal preference of v_1, candidates in L with higher preferences are positioned to the right of those with lower preferences. Similarly, candidates in R with higher preferences are positioned to the left of those with lower preferences. By combining these two observations, we can sort the candidates based on their positions along the line. Algorithm 4 presents the pseudocode for this approach.

Algorithm 2 Determine a candidate with a determined candidate in a voter's ordinal preference

Input: Ordinal preference of voter v_i, denoted $\succ_i$; candidates c_j and c_k where $c_k \succ_i c_j$, c_j is determined but c_k is not; two sets L and R containing currently determined candidates on the left and right sides of the p, respectively; and pivot candidates, c_1 and c_2.

Output: Updated sets L and R including candidate c_k.

function DETERMINE($\succ_i, c_j, c_k, L, R, c_1, c_2$)
    if $c_1 \succ_i c_k \wedge c_2 \succ_i c_k$ **then**
        Add c_k to L if $c_j \in R$, otherwise add c_k to R
    else
        Add c_k to L if $c_1 \succ_i c_2$, otherwise add c_k to R
    return (L, R)

Algorithm 3 Determine Candidates

Input: Election instance $\mathcal{E}$.

Output: Two subsets L and R of candidates, where candidates are on the left or right of the pivot point p.

function SPLITLINE($\mathcal{E} = (V, C, \succ)$)
    $L \leftarrow \{c_1\}$
    $R \leftarrow \{c_2\}$
    $C^* \leftarrow \{c_1, c_2\}$
    repeat
        $C_{new} \leftarrow \emptyset$
        while $\exists(v_i, c_j, c_k) : c_k \succ_i c_j \wedge c_k \notin C^* \wedge c_j \in Det$ **do**
            $(L, R) \leftarrow$ DETERMINE($\succ_i, c_k, c_j, L, R, c_1, c_2$)
            $C_{new} \leftarrow C_{new} \cup \{c_k\}$
        $C^* \leftarrow C^* \cup C_{new}$
    until $C_{new} = \emptyset$
    return (L, R)

SORTVOTERS. This part focuses on sorting voters given the sorted determined candidates. The key observation is that a voter who prefers one candidate over another tends to be closer to the candidate they prefer more. Consequently, we have a method to compare two voters. For a pair of voters v_i and v_j, assume k is the smallest index where the ordinal preferences of v_i and v_j differ. v_i is on the left side of v_j if $\succ_{i,k}$ is on the left side of $\succ_{j,k}$. If $\succ_{i,k}$ and $\succ_{j,k}$ are not determined, assuming v_i and v_j are at the same position does not affect the further analysis (see Sections 3.1).

It is important to note that the voter sorting process, denoted as SORTVOTERS($\mathcal{E} = (V, C, \succ), S_C$), takes the election and the sorted determined candidates as input and outputs S_V, a sequence of all voters in the order determined by the above approach.

In the next subsection, we provide a detailed proof demonstrating why the mentioned algorithm functions correctly.

3.1 Analysis

In this part, we focus on showing that we correctly calculated the order of candidates and voters. As mentioned earlier, this approach does not determine the exact order of all candidates or accurately rank all voters. First, recall that the algorithm returns the sequence S_C as the order of candidates (Line 3). This sequence contains

Algorithm 4 Sort determined candidates

Input: Preference order of v_1, denoted as $\succ_i$, L and R, candidates on the left and the right side of pivot p.
Output: Sequence S_C, sorted candidates in L and R by their position left to right.

```
function SortCandidates(≻_1, L, R)
    S_C is an empty sequence of candidates.
    for c_i in ≻_1 do
        if c_i ∈ L then
            Add c_i to the left of S_C.
        else if c_i ∈ R then
            Add c_i to the right of S_C.
    return S_C
```

the set of determined candidates, denoted as $C^* = L \cup R$. Additionally, in Line 4 of Algorithm 1, we calculate the order of voters based on S_C. As previously noted, this sequence is not entirely accurate due to limited available information. The following definition formalizes this order.

Definition 9. For a subset of candidates $A \subseteq C$, the order of voters *with respect to A* is a sequence S of voters from left to right, such that there is a tie between voters v_i and v_j if and only if their ordinal preferences over the candidates in A are identical.

Now, we can formally introduce the main theorem implied by the mentioned algorithm:

Theorem 10. Let $(\mathbb{R}, d)$ be the line metric. For an election $\mathcal{E} = (V, C, \succ)$ such that $d \triangleright \mathcal{E}$, SortCandiatesAndVoters correctly identifies:

1. the subset $C^* = L \cup R$ which is a core subset of candidates,
2. the order of C^* candidates, denoted as S_C,
3. and the order of voters with respect to C^*, denoted as S_V.

For the remainder of this section, all lemmas consider a specific line metric election instance.

The following lemma demonstrates that a prefix of a voter's ordinal preference forms a consecutive subsequence of candidates.

Lemma 11. For any voter $v_i \in V$ and $1 \leq k \leq m$, the k most preferred candidates of voter v_i form a consecutive subsequence of candidates.

Proof. Assume the condition does not hold for a voter v_i and some k. Since any single candidate forms a consecutive subsequence, we must have $k \geq 2$. Let A denote the k most preferred candidates of v_i. By assumption, there exist two candidates $c_l, c_r \in A$ such that a candidate $c_m \in C \setminus A$ lies between them.

Assume without loss of generality that c_l is to the left of c_m and c_r is to the right of c_m. Further, suppose v_i is located to the left of c_m. Since $c_m \notin A$, we must have $c_r \succ_{v_i} c_m$. However, since the candidates are arranged in a line, it follows that $c_m \succ_{v_i} c_r$, which is a contradiction (illustrated in Figure 1). □

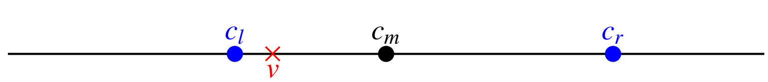


Figure 1: This figure is the illustration of the succession of nearest of any voter.

According to Lemma 11, we have:

Corollary 12. c_1 and c_2 are consecutive.

Next, recalling the definitions of determined and undetermined candidates (8), the following lemma formally proves that the function DETERMINE in Algorithm 2 correctly determines an undetermined candidate.

Lemma 13. *Assume that C^* is a consecutive subset of determined candidates, including c_1 and c_2. If there exists a voter v_i and candidates $c_j \in C^*$ and $c_k \notin C^*$ such that $c_k \succ_i c_j$, then the function DETERMINE correctly determines c_k.*

Proof. Let us consider two cases regarding the positioning of c_1, c_2, and c_k in the ordinal preference of v_i.

Case 1: Both c_1 and c_2 are positioned before c_k in the ordinal preference of v_i.

Without loss of generality, we can assume that c_j is in R. We use proof by contradiction to show that c_k is in L. Assume that c_i is in R. Consider the prefix of the ordinal preference of v_i ending with c_k. By Lemma 11, they must form a consecutive set of candidates. Therefore, c_j must be on the right side of c_k. However, since C^* consists of consecutive candidates, c_k would necessarily be positioned on the right side of c_j. Because of the contradiction we can conclude c_k is in L (Illustrated in Figure 2)

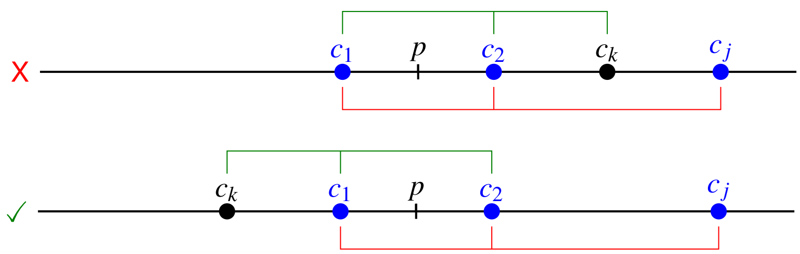


Figure 2: For a voter v_i, if we have $c_1 \succ_i c_k$, $c_2 \succ_i c_k$, and $c_k \succ_i c_k$, then c_k and c_j cannot both be in the same side of p. The figure above illustrates this contradiction, while the one below shows that they can be on opposite sides.

Case 2: At least one of c_1 and c_2 is positioned after c_k in the ordinal preference of v_i.

Without loss of generality, assume that c_1 precedes c_2 in the ordinal preference of v_i. Consider the prefix of the ordinal preference of v_i that contains c_1 and c_k but not c_2. By Lemma 11, this prefix must form a consecutive sequence of candidates, meaning c_2 cannot lie between c_1 and c_k. If c_k were in R, c_2 would lie between c_1 and c_k, as c_1 and c_2 are consecutive. This contradiction implies that c_k is in L (Illustrated in Figure 3).

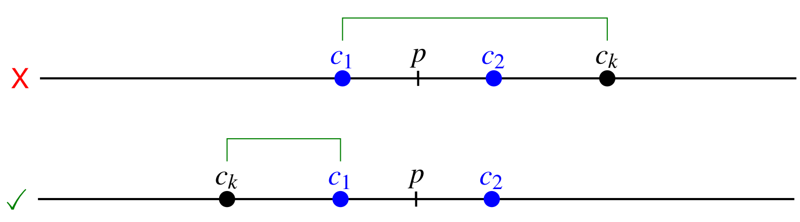


Figure 3: For a voter v_i, if we have $c_1 \succ_i c_2$ and $c_k \succ_i c_2$, then c_k is in L. The figure above illustrates that if c_k were in R, then c_2 would be in the consecutive subsequence of c_1 and c_k which is a contradiction. On the other hand, the one below shows that c_k must be in L.

□

As explained in Lemma 13, determining a new candidate works if the current set of C^* is consecutive. Therefore, in the following Lemma we show that in the begininegof each iteration, C^* is a consecutive subsequence of candidates.

Lemma 14. In Algorithm 3, in the beginning of each iteration (Line 5), all determined candidates, denoted as C^*, form a consecutive subsequence of all candidates.

Proof. Let C_i^* be the set of determined candidates in the beginning of i-th iteration. We prove that for any i, C_i^* is a consecutive subsequence. $C_1^* = \{c_1, c_2\}$ satisfies the condition. (Corollary 12).

In the $(i-1)$-th iteration $(i \geq 2)$, For any voter v_j, let $last_j$ be the least preferred candidate of v_j in C_{i-1}^*. Then by Lemma 13, by the end of iteration $i-1$ any candidate c_k such that $c_k \succ_v last_j$ is determined. We call this set of candidates P_j; note that $C_{i-1}^* \subseteq P_j$. As P_j is a prefix of the ordinal preference of v_j, by Lemma 11, P_j is a consecutive subsequence of candidates. Also P_j contains c_1 and c_2 because $\{c_1, c_2\} \subseteq C_{i-1}^* \subseteq P_j$.

Since $C_i^* = C_{i-1}^* \cup C_{new}$ (Line 10) and $C_{new} = \bigcup_j (P_j \setminus C_{i-1}^*)$, we have $C_i^* = \bigcup_j P_j$. Observe however that a union of intersecting intervals is always an interval; this can be seen from an easy inductive argument. Since P_j all contain $\{c_1, c_2\}$, it follows that C_i^* is a consecutive subsequence as well. □

Next, we show that C^* is a core subset of candidates.

Lemma 15. By the end of the algorithm, C^* is core. i.e. For any voter v_i, candidate $c_j \in C^*$, $c_k \notin C^*$, we have $c_j \succ_i c_k$.

Proof. Assume otherwise that $c_k \succ_i c_i$. Then, we can determine c_k using c_j and voter v_i according to Lemma 13, and consequently, Algorithm 3 cannot have terminated at this point. Therefore, $c_j \succ_i c_i$ for any voter v_i. □

We now demonstrate that the function SortCandidates in Algorithm 4 returns the determined candidates in sorted order from left to right.

Lemma 16. Function SortCandidates correctly sorts determined candidates, denoted as $C^* = L \cup R$ from left to right.

Proof. Recall that sets L and R contain determined candidates (Definition 8). Now, consider the ordinal preference of v_1. For any pair of candidates $c_i, c_j \in L$, we show that $d(c_i, c_1) < d(c_j, c_1)$ if $c_i \succ_1 c_j$.

There are two cases:

- c_1 is on the right side of v_1; we have:

$$\begin{aligned} d(c_i, c_1) - d(c_j, c_1) &= d(c_i, v_1) - d(c_1, v_1) - d(c_j, v_1) + d(c_1, v_1) \\ &= d(c_i, v_1) - d(c_j, v_1) \end{aligned}$$

- c_1 is on the left side of v_1; we have:

$$\begin{aligned} d(c_i, c_1) - d(c_j, c_1) &= d(c_i, v_1) + d(c_1, v_1) - d(c_j, v_1) - d(c_1, v_1) \\ &= d(c_i, v_1) - d(c_j, v_1) \end{aligned}$$

Therefore, we have:

$$d(c_i, c_1) < d(c_j, c_1) \Leftrightarrow d(c_i, v_1) < d(c_j, v_1)$$

Based on the definition of ordinal preference, we have $d(c_i, v_1) < d(c_j, v_1)$ if $c_i \succ_1 c_j$. Consequently $d(c_i, c_1) < d(c_j, c_1)$ if $c_i \succ_1 c_j$.

Similarly, for any pair of candidates $c_i, c_j \in R$, we can show that $d(c_i, c_2) < d(c_j, c_2)$ if $c_i \succ_1 c_j$.

The function SortCandidates processes candidates according to $\succ_1$, adding them to the left of S_C if they are in L and to the right otherwise. Therefore, S_C contains candidates in the correct order from left to right. □

Next, we show the correctness of function SortVoters.

Lemma 17. function SortVoters finds the order of voters with respect to determined candidates.

Proof. By Lemma 16, we have the order of a determined candidates. Assume that for candidate c_i, $order_{c_i}$ is the index of c_i from left to right. As SplitLine determines a prefix for each candidate, for a voter v_i, we let sequence $sorted_j = order_{\succ_{i,1}}, order_{\succ_{i,2}}, \ldots order_{\succ_{i,k}}$ where k is the number of determined candidates. Now, we can compare voters' *sorted* sequences lexicographically. A voter with smaller *sorted* is on the left side of a voter with larger *sorted*. It is important to highlight that this approach may result in ties among some voters in the ordering. □

By Lemmas 15, 16, and 17, we have Theorem 10 proved.

4 Line metric upper bounds

4.1 Sum-cost objective

In this section, we study the sum objective, where the cost of a set of candidates is defined as the sum of the distance of voters to the candidate set. We consider the line metric, where voters and candidates are located on the real line, and utilize the orderings obtained in Section 3 to present upper bounds on distortion. We start by showing that it is possible to choose a set of three candidates, such that an optimal candidate is always chosen. We then improve this by showing that we can always omit one of the three selected candidates, resulting in a shortlist of size two that includes an optimal candidate. We note that this not only shows a distortion of two based on our distortion metric, but it also ensures that the actual optimal single candidate is included in the selected list.

First, we identify the optimal single candidate when candidates and voters are positioned on the real line. In the following proofs, we often refer to the median voter. When the number of voters is even, we can consider either of the middle two candidates as the median.

Lemma 18. When considering the sum of distances objective and the candidates and voters are located on the real line, one of the candidates directly to the right or left of the median voter will be an optimal candidate.

Proof. Assume, without loss of generality, that there exists a candidate c_l who is the first candidate to the left of the median voter v. Consider any candidate c to the left of c_l. As we move from c to c_l, all voters to the right of v (including v) will experience a *reduction* in their distance to the candidates by the difference between c and c_l, while the distance for other voters can increase by at most the same amount. Since v is the median voter, there are at least as many voters to the right of v (including v) as there are to its left (excluding v). Therefore, the total cost cannot increase when moving from c to c_l, implying that the cost of c_l is *at most* the cost of c. Similarly, if a candidate c_r directly to the right of v exists, any candidate farther to the right will have a distance that is at least as great. Hence, either c_l or c_r will be an optimal candidate that minimizes the total cost. □

Next, we show that we can select three candidates to ensure an optimal candidate, which is immediately to the left or right of the median voter is selected.

Lemma 19. There exists a voting rule for 3-committee election on the line metric such that the resulting committee includes an optimal candidate.

Proof. First, we can determine the order of a subset of candidates C^* and all voters based on the algorithms in Section 3. Additionally, by Lemma 15, any candidate $c \notin C^*$ cannot be an optimal candidate, since c will be farther from all voters than any candidate in C^*. Thus, we can ignore the candidates outside C^* and proceed with the remaining candidates.

Now, we consider the median voter v in the ordering of voters. We note that there might be a tie between multiple voters for this position, but we will only utilize the voter's closest candidate, which will be the same for all tied voters.

After identifying the median voter, we consider v's closest candidate c. This candidate will be either the one immediately to the left or the right of the median voter v. Next, we use the ordering of the candidates to find the candidates c_l to the left of c and c_r to the right of c. We claim that the set $\{c, c_l, c_r\}$ is guaranteed to include an optimal candidate: if c is to the left of the median voter, then c_r will be the first candidate to the right of the median, and one of c or c_r will be an optimal candidate by Lemma 18. Similarly, if c is to the right of the median, one of c or c_l will be an optimal candidate. Therefore, the set $\{c, c_l, c_r\}$ will contain an optimal candidate.

Finally, we note that candidates c_l and c_r might not exist. In such cases, the candidates to the left and right of v are still selected, if they exist.

□

Next, we show that we can omit either c_l or c_r, choosing only two candidates, while still including an optimal candidate.

Theorem 20. There exists a voting rule for 2-committee election on the line metric such that the resulting committee includes an optimal candidate.

Proof. By Lemma 19, we can select three candidates that include an optimal candidate. Let c_1, c_2, and c_3 denote these candidates from left to right. We then define the sets V_1, V_2, and V_3 as the sets of voters who

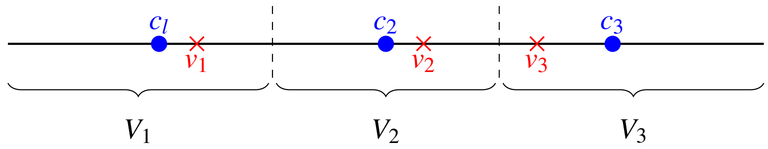


Figure 4: A figure illustrating three candidates c_1, c_2, and c_3 along with the possible locations of voters closest to each candidate, V_1, V_2, and V_3. Voters v_1, v_2 and v_3 show examples of voters in each set.

prefer the corresponding candidate to the other two, as illustrated in Figure 4. Thus, we can state that

$$\begin{aligned}
\texttt{cost}_s(c_2) &= \sum_{v \in V} d(v, c_2) \\
&= \sum_{v \in V_1} d(v, c_2) + \sum_{v \in V_2} d(v, c_2) + \sum_{v \in V_3} d(v, c_2) \\
&\leq \sum_{v \in V_1} d(v, c_2) + \sum_{v \in V_2} d(v, c_2) + \sum_{v \in V_3} (d(v, c_3) + d(c_2, c_3)) && \text{(Triangle Inequality)} \\
&\leq \sum_{v \in V_1} d(v, c_2) + \sum_{v \in V_2} d(v, c_3) + \sum_{v \in V_3} (d(v, c_3) + d(c_2, c_3)) && (\forall_{v \in V_2}\ d(v, c_2) \leq d(v, c_3)) \\
&= \sum_{v \in V_1} (d(v, c_3) - d(c_3, c_2)) + \sum_{v \in V_2} d(v, c_3) + \sum_{v \in V_3} (d(v, c_3) + d(c_2, c_3)) \\
& && (\forall_{v \in V_1}\ d(v, c_2) = d(v, c_3) - d(c_2, c_3)) \\
&= \texttt{cost}_s(c_3) - |V_1| \cdot d(c_2, c_3) + |V_3| \cdot d(c_2, c_3) \\
&= \texttt{cost}_s(c_3) + (|V_3| - |V_1|) \cdot d(c_2, c_3).
\end{aligned}$$

Similarly, we can show that $\texttt{cost}_s(c_2) \leq \texttt{cost}_s(c_1) + (|V_1| - |V_3|) \cdot d(c_2, c_1)$. Now, depending on whether $|V_1| < |V_3|$ or not, we can see that either $\texttt{cost}_s(c_2) \leq \texttt{cost}_s(c_1)$ or $\texttt{cost}_s(c_2) \leq \texttt{cost}_s(c_3)$. So we can disregard one of c_1 and c_3 based on this and still find the optimum, as c_2 is guaranteed to be a better candidate than the discarded one. □

Corollary 21. There exists a voting rule for 2-committee election on the line metric with 1-distortion of 1.

Finally, we show that for a general metric, it is possible to achieve a 1-distortion of $1 + \frac{2}{m-1}$ when choosing $k = m - 1$ candidates.

Theorem 22. There exists a voting rule choosing $m - 1$ out of m candidates achieving a 1-distortion of $1 + \frac{2}{m-1}$ for the sum-cost objective.

Proof. For each voter v, let first_v be their closest candidate. Then, we claim that the voting rule that chooses the $m - 1$ candidates appearing most frequently in first achieves the desired distortion. For a given instance, let C be the set of candidates selected by this algorithm and c_{opt} be the optimal single candidate. If $c_{\text{opt}} \in C$, then we get a 1-distortion of 1 and we are done. Otherwise, C includes every candidate except for c_{opt}. Now, we can bound the optimal cost $\texttt{OPT}$ in this instance as

$$\begin{aligned}\texttt{OPT} &= \sum_{i\in[n]} d(c_{\text{opt}}, v)\\ &= \sum_{\substack{v\in V\\ \mathit{first}_v = c_{\text{opt}}}} d(c_{\text{opt}}, v) + \sum_{\substack{v\in V\\ \mathit{first}_v \neq c_{\text{opt}}}} d(c_{\text{opt}}, v)\\ &\geq \sum_{\substack{v\in V\\ \mathit{first}_v = c_{\text{opt}}}} d(c_{\text{opt}}, v) + \sum_{\substack{v\in V\\ \mathit{first}_v \neq c_{\text{opt}}}} d(C, v). && (\mathit{first}_v \in C \text{ if } \mathit{first}_v \neq o)\end{aligned}$$

Let v' be the voter closest to c_{opt} such that $\mathit{first}_v \neq o$. Then, we can use this to state that

$$\begin{aligned}\texttt{OPT} &\geq \sum_{\substack{v\in V\\ \mathit{first}_v \neq c_{\text{opt}}}} d(c_{\text{opt}}, v)\\ &\geq \sum_{\substack{v\in V\\ \mathit{first}_v \neq c_{\text{opt}}}} d(c_{\text{opt}}, v')\\ &\geq \frac{m-1}{m} \cdot n \cdot d(c_{\text{opt}}, v') && (\mathit{first}_v = c_{\text{opt}} \text{ for at most } \tfrac{n}{m} \text{ voters based on choice of } C)\end{aligned}$$

and therefore

$$\frac{n}{m} d(c_{\text{opt}}, v') \leq \frac{1}{m-1}\texttt{OPT}. \tag{1}$$

Now, we can bound the cost of C as follows:

$$\begin{aligned}\texttt{cost}_m(C) &= \sum_{v\in V} d(C, v)\\ &= \sum_{\substack{v\in V\\ \mathit{first}_v = c_{\text{opt}}}} d(C, v) + \sum_{\substack{v\in V\\ \mathit{first}_v \neq c_{\text{opt}}}} d(C, v)\\ &\leq \sum_{\substack{v\in V\\ \mathit{first}_v = c_{\text{opt}}}} (d(c_{\text{opt}}, v) + d(c_{\text{opt}}, v') + d(C, v')) + \sum_{\substack{v\in V\\ \mathit{first}_v \neq c_{\text{opt}}}} d(C, v) && \text{(Triangle inequality)}\\ &\leq \sum_{\substack{v\in V\\ \mathit{first}_v = c_{\text{opt}}}} (d(c_{\text{opt}}, v) + d(c_{\text{opt}}, v') + d(c_{\text{opt}}, v')) + \sum_{\substack{v\in V\\ \mathit{first}_v \neq o}} d(C, v) && (\mathit{first}_{v'} \neq c_{\text{opt}} \text{ and } \mathit{first}_{v'} \in C)\\ &\leq \texttt{OPT} + \sum_{\substack{v\in V\\ \mathit{first}_v = c_{\text{opt}}}} 2d(c_{\text{opt}}, v') && \text{(Previous upper bound for OPT)}\\ &= \texttt{OPT} + 2 \cdot \frac{n}{m} d(c_{\text{opt}}, v') && (\mathit{first}_v = c_{\text{opt}} \text{ for at most } \tfrac{n}{m} \text{ voters})\\ &\leq (1 + \frac{2}{m-1})\texttt{OPT} && \text{(By Equation 1)}\end{aligned}$$

□

4.2 Max-cost objective

Next, we focus on the max objective, where the cost of a set of candidates is the maximum distance of any voter from the set. We show that selecting four candidates is sufficient for achieving a distortion of 1 compared to the single optimal candidate. In addition, we show that we can achieve distortion factors of $3/2$

and 2 by selecting three or two candidates respectively. We note that in different parts of this section, we refer to the leftmost and rightmost voters v_l and v_r which we identify using the algorithms in Section 3. While the orderings in that section may have ties, we are only concerned with the most preferred candidate of each voter, which will be the same for tied voters.

We begin by determining a lower bound for the optimal cost $\texttt{OPT}$ when only one candidate is selected.

Lemma 23. Let v_l and v_r be the leftmost and rightmost candidates respectively. If the distance between v_l and v_r is d (i.e., $d = d(v_l, v_r)$), then the cost for the optimal single candidate $\texttt{OPT}$ satisfies $\texttt{OPT} \geq \frac{d}{2}$.

Proof. Let c_{opt} be an optimal candidate. The distance between v_l and c_{opt} is $d(v_l, c_{\text{opt}})$, and the distance between v_r and c_{opt} is $d(v_r, c_{\text{opt}})$. By the triangle inequality:

$$d(v_l, c_{\text{opt}}) + d(v_r, c_{\text{opt}}) \geq d.$$

Thus,

$$\max(d(v_l, c_{\text{opt}}), d(v_r, c_{\text{opt}})) \geq \frac{d}{2}.$$

Therefore, $\texttt{OPT} \geq \frac{d}{2}$.

□

Based on the algorithms in Section 3, we can determine the order of candidates in C^* and the voters. We denote the leftmost voter as v_l and the rightmost voter as v_r. Similarly, we denote the closest candidate to v_l as c_l and the closest candidate to v_r as c_r.

Now, we show that if there are no candidates between v_l and v_r, a distortion of 1 can be achieved with 2 candidates.

Lemma 24. If there are no candidates placed between v_l and v_r, a distortion of 1 can be achieved by selecting c_l and c_r.

Proof. In this case, the voters are next to each other in a block, with c_l being the first candidate immediately to the left of all voters and c_r being the first candidate immediately to the right of all voters. Therefore, by selecting c_l and c_r, we ensure that the closest candidate to each voter is included. Therefore, the cost of this set is at most that of the single optimal candidate, and a distortion factor of 1 is achieved compared to this benchmark. □

From now on, we assume at least one candidate is located between v_l and v_r. This implies that the optimal single candidate c_{opt} is also between v_l and v_r, as any candidate outside the interval has a cost of at least $d(v_l, v_r)$.

Next, we show that selecting two candidates is sufficient to achieve a distortion of 2 compared to the optimal single candidate.

We first prove the following lemma for the left-most and right-most voters, denoted v_l and v_r.

Lemma 25. If the distance between v_l and v_r is d, then for every voter v, either $d(v, v_l) \leq \texttt{OPT}$ or $d(v, v_r) \leq \texttt{OPT}$.

Figure 5: Illustration of the candidates c_l and c_r, and the voters v_l and v_r, demonstrating why choosing these candidates does not achieve a distortion of 1

Proof. Since v is a voter between v_l and v_r (recall v_l is the left-most and v_r is the right-most voter), we have:

$$d(v_l, v) + d(v, v_r) = d.$$

Thus, either $d(v_l, v) \leq \frac{d}{2}$ or $d(v, v_r) \leq \frac{d}{2}$.

By Lemma 23, $\frac{d}{2} \leq \texttt{OPT}$. Therefore, we conclude:

$$d(v_l, v) \leq \frac{d}{2} \leq \texttt{OPT}, \quad \text{or} \quad d(v, v_r) \leq \frac{d}{2} \leq \texttt{OPT}.$$

□

Now, we show that it's possible to select two candidates and achieve a distortion of 2.

Theorem 26. Let v_l and v_r be the leftmost and rightmost voters, and c_l and c_r the closest candidates to v_l and v_r. Then, set $\{c_l, c_r\}$ has a cost at most twice the optimal single candidate.

Proof. Since c_l is closest to v_l and we know $d(v_l, c_{\text{opt}}) \leq \texttt{OPT}$, it follows that:

$$d(v_l, c_l) \leq d(v_l, c_{\text{opt}}) \leq \texttt{OPT}.$$

Similarly, $d(v_r, c_r) \leq \texttt{OPT}$.

For any voter v, by Lemma 25, either $d(v, v_l) \leq \texttt{OPT}$ or $d(v, v_r) \leq \texttt{OPT}$. Without loss of generality, assume that $d(v, v_l) \leq \texttt{OPT}$. Then:

$$d(v, c_l) \leq d(v, v_l) + d(v_l, c_l) \leq \texttt{OPT} + \texttt{OPT} = 2 \cdot \texttt{OPT}.$$

Similarly, in the other case:

$$d(v, c_r) \leq 2 \cdot \texttt{OPT}.$$

Thus, for every voter, the distance to the closest candidate in $\{c_l, c_r\}$ is at most $2 \cdot \texttt{OPT}$. Hence, this selection achieves a distortion of 2.

□

Figure 5 illustrates how the closest candidates to v_l and v_r lying outside the interval between them leads to a distortion factor larger than 1. Specifically, if c_l is to the left of v_l or c_r is to the right of v_r, some voters between v_l and v_r might be further than $\texttt{OPT}$ from both c_l and c_r.

On the other hand, if the candidates c_l and c_r are placed between v_l and v_r, as shown in Figure 6, this selection would achieve a distortion of 1. In this case, the following property holds for any voter v:

$$d(v, c_l) \leq \max(d(v, v_l), \texttt{OPT}) \quad \text{and} \quad d(v, c_r) \leq \max(d(v, v_r), \texttt{OPT})$$

as c_l lies between v_l and c_{opt} and c_r lies between v_r and c_{opt}.

Figure 6: Illustration of the candidates c_l and c_r positioned between the voters v_l and v_r. In this case, we achieve the optimal answer instead of a two approximation.

By Lemma 25, we know that for every voter v, either $d(v, v_l) \leq \texttt{OPT}$ or $d(v, v_r) \leq \texttt{OPT}$. Combining these, it follows that:

$$d(v, c_l) \leq OPT \quad \text{or} \quad d(v, c_r) \leq OPT.$$

To improve the distortion factor, we attempt to include the leftmost and rightmost candidates that lie between the first and last voters by selecting more candidates. We first demonstrate that introducing an additional candidate allows us to achieve a distortion of $3/2$ and then show that selecting two additional candidates guarantees a distortion of 1.

Theorem 27. There exists a voting rule to select three candidates that achieves a distortion of $3/2$ compared to the optimal single candidate.

Proof. Let v_l and v_r be the leftmost and rightmost voters, and c_l and c_r their closest candidates. Using the ordering of the candidates, let c'_r be the candidate immediately to the left of c_r. We claim that the set $\{c_l, c_r, c'_r\}$ achieves a distortion of $3/2$. First, we note that any voter to the left of c_l or the right of c'_r has their closest candidate included in this set. Now, one of c'_r or c_r must be the rightmost candidate between v_l and v_r. Let this be candidate c^*_r. We already know that any voter outside the interval between c_l and c^*_r has their closest candidate included. In addition, we have

$$d(c_l, c^*_r) \leq d(c_l, v_l) + d(v_l, c_{\text{opt}}) + d(c_{\text{opt}}, c^*_r) \leq 3\texttt{OPT}.$$

Therefore, the minimum distance of any voter in this interval to endpoints c_l, c^*_r is at most $3\texttt{OPT}/2$. So, the maximum distance of any voter to the candidates is at most $3/2\texttt{OPT}$ and we achieve a distortion of $3/2$. □

Finally, we show that selecting four candidates allows us to achieve a distortion of 1.

Theorem 28. There exists a voting rule to select four candidates that achieves a distortion of 1 compared to the optimal single candidate.

Proof. We again consider the leftmost and rightmost voters v_l and v_r and their closest candidates c_l and c_r. In addition, we choose candidate c'_l to the right of c_l and c'_r to the left of c_r based on the ordering of candidates. Now, one of c_l and c'_l will be the first candidate to the right of v_l: if c_l is to the right of v_l, it must be the first such candidate. Otherwise, it is the first candidate to the left of v_l, so c'_l is to the right of v_l. Let this candidate be c^*_l and define c^*_r similarly. Now, any voter v to the left of c'_l has their closest candidate included in the set, as v is either between c_l and c'_l, so one of these two is v's closest candidate, or v is to the left of c_l, in which case c_l must be v's closest candidate. Similarly, the closest candidate to each voter to the right of c'_r is selected.

Next, since both c^*_l and c^*_r are between v_l and v_r and $d(v_l, v_r) \leq 2\texttt{OPT}$, we have $d(c^*_l, c^*_r) \leq 2\texttt{OPT}$. Therefore, any voter between c^*_l and c^*_r has a distance of at most $\texttt{OPT}$ to the closer candidate in $\{c^*_l, c^*_r\}$. As every voter is either between c^*_l and c^*_r or outside the interval c'_l, c'_r, the distance of each voter to the closest candidate in our selected set is at most $\texttt{OPT}$ and we get a distortion of 1. □

5 Lower Bound on Distortion Factor

5.1 Lower bound for sum-cost

In this section, we provide lower bounds for distortion in k-committee election when considering the sum-cost objective. First, we extend the lower bound of [7] to show that achieving bounded distortion is not possible for general k even if we allow to select $\omega(k)$ candidates. Furthermore, the lower bound holds for the line metric. Next, we show that for general metrics, it is not possible to achieve a distortion of 1 when compared to the single optimal candidate if at least one candidate is not selected. We argue about this lower bound using instances in the two-dimensional plane (2-D Euclidean metric), showing that using the line metric space is crucial to our positive results.

Theorem 29. For $k \geq 3$, there exists an instance for the k-committee election (with respect to the sum-cost) where any algorithm choosing at most $\lceil \frac{k+1}{2} \rceil \lfloor \frac{k+1}{2} \rfloor - 1$ candidates has unbounded distortion when compared to an optimal choice of k candidates.

Proof. Let $a = \lceil \frac{k+1}{2} \rceil$ and $b = \lfloor \frac{k+1}{2} \rfloor$. We consider an instance of the problem with ab candidates located on a line. We partition these candidates into a groups $C_1, C_2, \ldots, C_a$ of size b each. Let the j-th candidate in set C_i be $c_{i,j}$. We consider a voter $v_{i,j}$ for each such candidate and set the preferences of voter $v_{i,j}$ as follows:

- For any two candidates $c_{i',j'}$ and $c_{i'',j''}$ with $i' < i''$, $c_{i',j'} \succ c_{i'',j''}$ if $i'' > i$ and $c_{i'',j''} \succ c_{i',j'}$ otherwise. This means that in voter $v_{i,j}$'s preference order, each set C_i occupies a consecutive segment, with $C_i \succ C_{i-1} \succ \ldots \succ C_1 \succ C_{i+1} \succ \ldots \succ C_a$. (By the notation, $C_r \succ C_s$, we mean every candidate in the group C_s is preferred over any candidate in the group C_r.)
- With C_i, the voter $v_{i,j}$'s preference order is

$$c_{i,j} \succ c_{i,j-1} \succ \ldots \succ c_{i,1} \succ c_{i,j+1} \succ \ldots \succ c_{i,b}.$$

- The preference of voter $v_{i,j}$ for candidates in $C_{i'}$ is in increasing order of j if $i' > i$ and decreasing order of j if $' < i$.

Now, we show that any deterministic algorithm ALG choosing at most $ab - 1 = \lceil \frac{k+1}{2} \rceil \lfloor \frac{k+1}{2} \rfloor - 1$ candidates will have unbounded distortion when compared to an optimal choice of $a + b - 1 = k$ candidates. As $ab - 1 < ab$, at least one candidate c_{i^*,j^*} is not chosen by ALG in this instance. Then, consider arbitrary constants $\ell > 2$ and $\varepsilon > 0$. Now, for any $i \neq i^* \in [a]$ and $j \in [b]$, we place candidate $c_{i,j}$ and voter $v_{i,j}$ in point $\ell 3^i + \varepsilon 3^{j-b}$ if $i \neq i^*$ and point $\ell 3^i + 3^{j-b}$ if $i = i^*$.

Next, we first show that this placing respects the voters' preference orders. We can see that for each $i \in [a]$, $j \in [b]$, the distance of voter $v_{i,j}$ to any candidate in C_{i+1} is at least

$$\ell 3^{i+1} - \ell 3^i - 1 \geq 2\ell 3^i - 1 \geq \ell 3^i + \ell - 1 > \ell 3^i$$

while the voter's distance to any candidate in C_1 is at most

$$\ell 3^i + 1 - \ell 3^1 \leq l3^i.$$

Therefore, any candidate in C_1 is closer to $v_{i,j}$ than any candidate in C_{i+1}. In addition, the distance of this voter to any candidate in C_i is at most 1, while its distance to the closest candidate in C_{i-1} is at least

$$\ell 3^i - \ell 3^{i-1} - 1 \geq 2\ell 3^{i-1} - 1 \geq \ell > 1.$$

Now, since candidate groups C'_i are ordered left to right on the line based on increasing i', and voter $v_{i,j}$ is on the same point as $c_{i,j}$, we can say

$$C_i \succ C_{i-1} \succ \ldots \succ C_1 \succ C_{i+1} \succ \ldots \succ C_a$$

holds for voter $v_{i,j}$. The preference for candidates within each $C_{i'}$ where $i' \neq i$ will also be consistent with the ordering, as they are all to the left or right of $v_{i,j}$. Finally, within C_i, the candidates $c_{i,j'}$ are ordered left to right in increasing order of j', with voter $v_{i,j}$ coinciding with candidate $c_{i,j}$. Since the distance of $v_{i,j}$ to $v_{i,1}$ is less than its distance to $v_{i,j+1}$, the preference order for $v_{i,j}$ is as desired.

Now, we compare the cost of our algorithm and an optimal solution, choosing at most $a+b-1 = k$ candidates. Since the algorithm does not choose c_{i^*,j^*}, v_{i^*,j^*} has a cost of at least $3^{j^*-b-1} \geq 3^{-b}$ from its closest candidate. So the cost of $\texttt{ALG}$ is at least 3^{-b}. On the other hand, we can consider choosing every candidate in C_{i^*}, along with $c_{i,1}$ for every $i \neq i^* \in [a]$. For each voter $v_{i,j}$, if $i = i^*$, this choice has a cost of 0 as $c_{i,j}$ is chosen, and if $i \neq i^*$, its distance to $c_{i,1}$ is at most $\varepsilon 3^{j-b} \leq \varepsilon$. So, the total cost is at most $(a-1)b\varepsilon$, and only $a+b-1 = k$ candidates are selected. Now, the distortion will be at least

$$\frac{3^{-b}}{(a-1)b\varepsilon}$$

which can be arbitrarily increased to take any value by proper choice of ε.

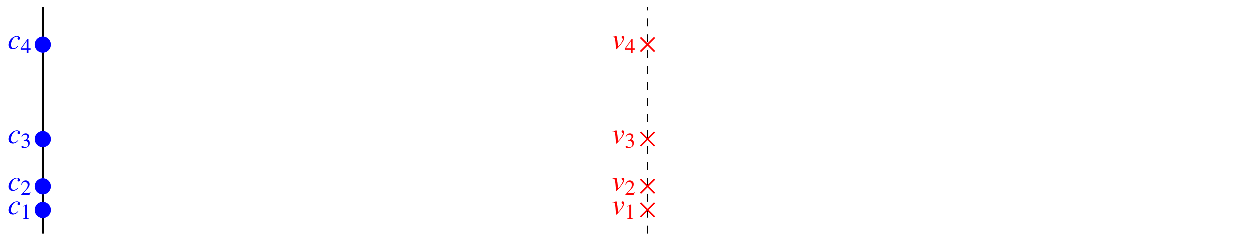
□

Remark 30. We note that this instance has a recursive structure with two levels that can be generalized to more levels. In a generalization with ℓ levels, we can define each voter and candidate using a sequence of ℓ indices with values in $[a]$. Then, a voter $v_{i_1,\ldots,i_\ell}$'s preference order for candidates $c_{j_1,\ldots,j_\ell}$ is determined by the first index where sequences $i_1, \ldots, i_\ell$ and $j_1, \ldots, j_\ell$ defer. If this index is different for two candidates, the one with more matching indices is preferred, and if both differ with the voter at the same index i, the same pattern of preference $i \succ i-1 \succ \ldots \succ 1 \succ i+1 \succ \ldots \succ a$ is used. Then, suppose $c_{i^*_1,\ldots,i^*_\ell}$ is a candidate not chosen by an algorithm selecting at most $a^\ell - 1$ candidates. In that case, we can place the candidates and voters with a similar recursive structure such that the set of candidates $\{c_{i^*_1,\ldots,i^*_{r-1},i_r,1,\ldots,1} \mid r \in [\ell], i_r \in [a] \setminus \{i^*_r\}\} \cup \{c_{i^*_1,\ldots,i^*_\ell}\}$ has as a multiple of ε as its total cost. In contrast, the closest candidate except $c_{i^*_1,\ldots,i^*_\ell}$ to $v_{i^*_1,\ldots,i^*_\ell}$ has a constant distance to it not dependent on ε. This leads to an unbounded distortion when comparing any algorithm selecting at most $a^\ell - 1$ candidates to the optimal selection of $\ell(a-1)+1$ candidates. In particular, choosing $a = 2$ and $\ell = \lceil \log_2 k + 1 \rceil$, we can show that any algorithm for selecting k candidates cannot achieve an unbounded distortion when compared to the optimal selection of $\lceil \log_2(k+1) \rceil + 1$ candidates.

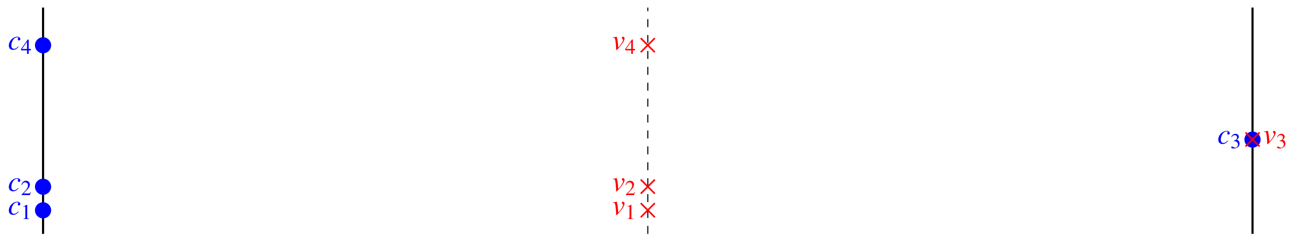
Next, We construct an example in a two-dimensional plane, where picking even $m-1$ candidates out of m candidates won't guarantee that an optimal candidate is chosen. Furthermore, we provide construction of such an instance so that if we consider the distance to the closest candidate among the $m-1$ selected candidates for each voter, we still get a distortion of at least $1 + \frac{2}{m-1} - \varepsilon$, for any given $\varepsilon > 0$, compared to the optimal single candidate.

Theorem 31. Any deterministic algorithm for the $(m-1)$-committee election problem (with respect to the sum-cost) that, given voters' preference orderings on m candidates, selects at most $m-1$ candidates must have a 1-distortion factor of at least $1 + \frac{2}{m-1} - \varepsilon$, for any $\varepsilon > 0$.

Proof. We consider a family of $m+1$ instances $I_0, I_1, \cdots, I_m$ with m voters $\{v_1, v_2, \cdots, v_m\}$ and m candidates $\{c_1, c_2, \cdots, c_m\}$ on a two-dimensional plane. In terms of voters' and candidates' locations on the plane, each



(a) The instance I_0



(b) I_3: Instance with voter v_3 and candidate c_3 moved

Figure 7: Figures of the lower bound instances. In (a), all candidates are located on the line $x = -\ell$, with voters with matching y-coordinates on the line $x = 0$. In (b), voter v_3 and candidate c_3 are moved to the line $x = \ell$, while keeping the same y-coordinate.

of these instances is distinct. However, in terms of voters' preference orders on candidates (i.e., ordinal information), $I_0, I_1, \cdots, I_m$ are the same. To be more specific, let us describe these instances.

Let $\ell = 3/\varepsilon - 1$. We first describe the instance I_0. For each $i \in [m]$, the candidate c_i is located at the point $\left(-\ell, 2^{i-1}/2^m\right)$. For each $i \in [m]$, the voter v_i is located at the point $\left(0, 2^{i-1}/2^m\right)$. (See Figure 7.) It is straightforward to observe that, for each $i \in [m]$, the preference order of the voter v_i is

$$c_i \succ c_{i-1} \succ \cdots \succ c_1 \succ c_{i+1} \succ c_{i+2} \succ \cdots \succ c_m.$$

Next, for each $j \in [m]$, we define the instance I_j as follows: For each $i \neq j \in [m]$, the locations of the candidate c_i and the voter v_i are the same as that in the instance I_0. Both the candidate c_j and the voter v_j are located at the point $\left(\ell, 2^{i-1}/2^m\right)$. (See Figure 7.) It is easy to observe that for any voter v_i (where $i \neq j$), the distance to any candidate c_j remains the same (as in I_0). Since the locations of other candidates remain the same, the preference order for any voter v_i (where $i \neq j$) remains the same as in I_0. Let us now consider the voter v_j. Note that its closest candidate is c_j, and the distance to any other candidate c_i remains the same as in I_0. Thus, the preference order of the voter v_j is

$$c_j \succ c_{j-1} \succ \cdots \succ c_1 \succ c_{j+1} \succ c_{j+2} \succ \cdots \succ c_m$$

which is the same as that in I_0.

Let us now consider any arbitrary deterministic algorithm ALG that selects at most $m - 1$ candidates. Now, suppose given the preference orders of voters as in I_0 (the same as in $I_1, \cdots, I_m$), ALG selects a set C of candidates where $|C| \leq m - 1$. Suppose $c_k \notin C$, for some $k \in [m]$. Now, consider the instance I_k. Note, since voters' preference orders, i.e., ordinal information, are the same as in I_0, ALG also selects the set C for the instance I_k. Then

$$\begin{aligned}\texttt{cost}_s(C, I_k) &= \sum_{i\in[m]} d(v_i, C) \\ &\geq \sum_{i\neq k} \ell + 2\ell = (m+1)\ell.\end{aligned}$$

On the other hand, selecting only the candidate c_k for the instance I_k would lead to the cost of

$$\texttt{cost}_s(c_k, I_k) = \sum_{i\in[m]} \|v_i - c_k\|_2 \leq \sum_{i\neq k}(\ell+1) = (m-1)(\ell+1)$$

and thus the optimal cost $\texttt{OPT}(I_k) \leq (m-1)(\ell+1)$.

Hence, the distortion factor of $\texttt{ALG}$ on the instance I_k is

$$\begin{aligned}\texttt{1-distortion(ALG)} &= \frac{\texttt{cost}_s(C, I_k)}{\texttt{OPT}(I_k)} \\ &\geq \frac{(m+1)\ell}{(m-1)(\ell+1)} \\ &= \left(1 + \frac{2}{m-1}\right)\left(1 - \frac{1}{\ell+1}\right) \\ &= 1 + \frac{2}{m-1} - \frac{1}{\ell+1}\left(1 + \frac{2}{m-1}\right) \\ &\geq 1 + \frac{2}{m-1} - \varepsilon \qquad \text{(since } \ell = 3/\varepsilon - 1).\end{aligned}$$

□

Next, we generalize our lower bound to any algorithm that picks at most r candidates.

5.2 Lower bounds for the max-cost

In this section, we provide lower bounds on the distortion for the committee election problem when considering the max-objective. In the 2-D Euclidean space, we show that even when choosing $m-1$ candidates out of m, we cannot guarantee a distortion of less than 3, matching the distortion achieved by simple algorithms when selecting only one candidate. In addition, we provide lower bounds when considering the line metric, showing that our algorithms that select at most $k = 2$ and $k = 3$ candidates, respectively, achieve tight bounds in terms of distortion.

We first show that any deterministic algorithm choosing at most $k < m$ candidates (out of m candidates) cannot achieve a distortion strictly lower than 3 when compared to a single optimal candidate, even when the candidates and voters are located in the two-dimensional plane.

Theorem 32. Any deterministic algorithm for the k-committee election (with respect to the max-cost) that, selects at most $k < m$ candidates out of m candidates, must have a 1-distortion of at least $3 - \varepsilon$ for any $\varepsilon > 0$.

Proof. We proceed similarly to Theorem 31, using a family of instances $I_0, I_1, \ldots, I_m$ on the two-dimensional plane, such that the voters' ordinal preferences are the same in all of these instances, but the locations and distances of candidates and voters vary. In each instance, we have m voters $\{v_1, v_2, \ldots, v_m\}$ and m candidates $\{c_1, c_2, \ldots, c_m\}$.

Let $\ell = 3/\varepsilon - 1$. We use the same instance I_0 as in Theorem 31, where for each $i \in [m]$, candidate c_i is located at $(-\ell, 2^{i-1}/2^m)$ and voter v_i is located at $(0, 2^{i-1}/2^m)$. Then, for each $i \in [m]$, voter v_i will have the ordinal preference

$$c_i \succ c_{i-1} \succ \cdots \succ c_1 \succ c_{i+1} \succ c_{i+2} \succ \cdots \succ c_m.$$

in instance I_0.

Next, we define instance I_j for each $j \in [m]$. For each $i \in [m] \setminus \{j\}$, v_i and c_i will remain in the same location as in I_0. Meanwhile, candidate c_j is moved to $(\ell, 2^{j-1}/2^m)$ and voter v_j to $(2\ell, 2^{j-1}/2^m)$. It can be seen that for any $i \neq j$, the distance of voter v_i to candidate c_j will be unchanged compared to I_0, and therefore the voter's ordinal preference will remain the same. For voter j, its closest candidate will remain c_j. In addition, its preference for the other candidates will remain unchanged. Therefore, voter v_j's ordinal preference will remain the same as in I_0 too and all voters' ordinal preferences are identical in I_0 and I_j. An example of these instances is illustrated in Figure 8.

Now, consider an arbitrary deterministic algorithm ALG that selects $k < m$ candidates. Since $k < m$, when running ALG on I_0, there exists a candidate c_j that is not in the set of selected candidates C. Now, consider the algorithm's performance on I_j. Since the voter's ordinal preferences are the same in I_0 and I_j, and ALG operates using only this information, its output on I_j must be the same as in I_0 and therefore it does not select c_j. Now, we can lower bound the cost of ALG with respect to the max objective as

$$\begin{aligned} \texttt{cost}_m(C, I_0) &\geq d(v_j, C) \\ &\geq 3\ell. && \text{(Since } c_j \notin C\text{)} \end{aligned}$$

On the other hand, if we only select c_j, we get

$$\begin{aligned} \texttt{cost}_m(c_j, I_0) &= \max_{i \in [m]} \left\| v_i - c_j \right\|_2 \\ &\leq \ell + 1. \end{aligned}$$

This shows that the optimal cost $\texttt{OPT}_m(I_j) \leq \ell + 1$. Finally, combining these two, we get that the distortion of ALG on instance I_j is at least

$$\begin{aligned} \texttt{1-distortion(ALG)} &\geq \frac{\texttt{cost}_m(C, I_k)}{\texttt{OPT}_m(I_k)} \\ &\geq \frac{3\ell}{\ell + 1} \\ &= 3 - \frac{3}{\ell + 1} \\ &= 3 - \varepsilon. && (\ell = 3/\varepsilon - 1) \end{aligned}$$

□

Next, we show that our upper bounds for the line metric are tight by proving matching lower bounds. We begin with the case $k = 2$.

Theorem 33. Any deterministic algorithm for the 2-committee election (with respect to the max-cost) when voters and candidates are located on a line must have a 1-distortion of at least $2 - \varepsilon$ for any $\varepsilon > 0$.

(a) The instance I_0

(b) I_3: Instance with voter v_3 and candidate c_3 moved

Figure 8: Figures of the lower bound instances for the max objective. In (a), all candidates are located on the line $x = -\ell$, with voters with matching y-coordinates on the line $x = 0$. In (b), candidate c_3 is moved to the line $x = \ell$, and voter v_3 is moved to the line $x = 2\ell$ while keeping the same y-coordinates.

Proof. We consider the following instance with three voters $\{v_1, v_2, v_3\}$ and three candidates $\{c_1, c_2, c_3\}$. Let the preferences of voters be as follows:

$$v_1 : c_1 \succ c_2 \succ c_3$$
$$v_2 : c_2 \succ c_1 \succ c_3$$
$$v_3 : c_3 \succ c_2 \succ c_1.$$

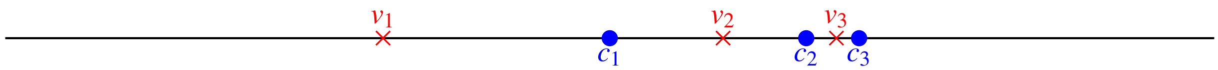
Next, we consider three possible placements for the voters and candidates that respect the above (desired) preference order of voters, as shown in Figure 9. In all instances, v_1 is located at point -1, and v_3 at point 1. In the first instance, v_2 is located at 0.5. Candidates c_1, c_2, c_3 are located at points $0, 1-\varepsilon$ and $1+\varepsilon/2$ respectively. It is easy to see that the voters' preferences in this instance match our desired orders. Now, we can see that in this instance, c_1 has a distance of at most 1 to every voter, while both c_2 and c_3 have a distance of at least $2-\varepsilon$ from v_1. Therefore, not choosing c_1 will lead to a distortion of at least $2-\varepsilon$.

In the second instance, we place v_2 and c_2 at point 0, c_1 at point $-2+\varepsilon$, and c_3 at point $2-\varepsilon/2$. Again, we can see that the voters' preferences will follow the desired orders. In this instance, c_2 has a distance of at most 1 from all voters, while v_2 is at a distance of at least $2-\varepsilon$ from the other candidates. Therefore, not choosing c_2 will lead to a distortion of at least $2-\varepsilon$.

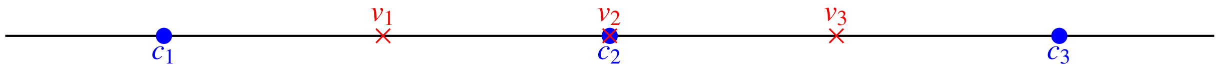
Finally, in the third instance, we place c_1 at point -1, c_2 and v_2 at point $-1+\varepsilon$ and c_3 at point 0. This placement will also respect the desired preference orders for each voter. Additionally, since c_3 has a distance of at most 1 from all voters and v_3 has a distance of at least $2-\varepsilon$ from the other candidates, not choosing c_3 leads to a distortion of at least $2-\varepsilon$.

Now, since these instances cannot be distinguished based on the voters' ordinal preferences, and not choosing any of the candidates leads to a distortion of at least $2-\varepsilon$, any deterministic algorithm selecting two candidates cannot achieve a distortion better than $2-\varepsilon$.

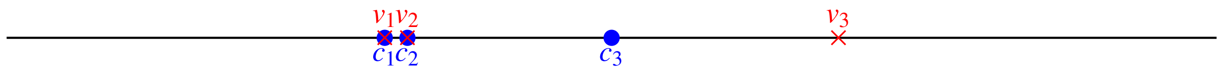
□



(a) First instance: Choosing c_1 achieves a cost of 1, while the other two candidates have a distances of at least $2 - \varepsilon$ to v_1.



(b) Second instance: Choosing c_2 achieves a cost of 1, while the other two candidates have a distance of at least $2 - \varepsilon$ to v_2.



(c) Third instance: Choosing c_3 achieves a cost of 1, while the other two candidates have a distances of at least $2 - \varepsilon$ to v_3.

Figure 9: Possible locations of the voters and candidates in the lower bound instance for $k = 2$.

Lastly, we provide a tight lower bound for the case $k = 3$.

Theorem 34. Any deterministic algorithm for the 3-committee election (with respect to the max-cost) when voters and candidates are located on a line must have a 1-distortion of at least $3/2 - \varepsilon$ for any $\varepsilon > 0$.

Proof. We consider instances $\{I_1, I_2, I_3, I_4\}$ each with four voters $\{v_1, v_2, v_3, v_4\}$ and four candidates $\{c_1, c_2, c_3, c_4\}$ such that the preferences of voters are as follows in every instance:

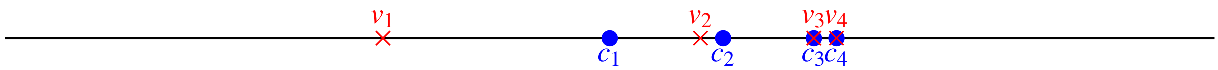
$$\begin{aligned} v_1 &: c_1 \succ c_2 \succ c_3 \succ c_4 \\ v_2 &: c_2 \succ c_1 \succ c_3 \succ c_4 \\ v_3 &: c_3 \succ c_4 \succ c_2 \succ c_1 \\ v_4 &: c_4 \succ c_3 \succ c_2 \succ c_1. \end{aligned}$$

For instance I_i, we choose the location of voters and candidates so that candidate c_i has a distance of at most 2 to each voter, while voter v_i has a distance of at least $3 - 2\varepsilon$ to every candidate except c_i. This leads to a distortion of at least $3/2 - \varepsilon$ for any deterministic algorithm ALG, as these instances cannot be distinguished based on ordinal preferences, and at least one candidate c_i is not chosen in instance I_i by ALG. These instances are shown in Figure 10.

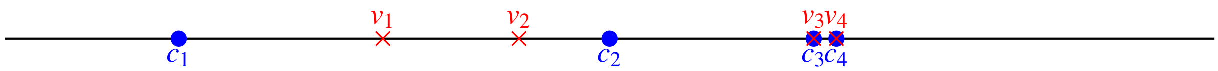
In instance I_1, we have voters v_1, v_2, v_3, v_4 located at points $-2, 1 - \varepsilon, 2 - \varepsilon, 2$ and candidates c_1, c_2, c_3, c_4 located at points $0, 1, 2 - \varepsilon, 2$ respectively. It can be seen that the ordinal preference of each voter in this instance matches the desired ordering. Now, c_1 has a distance of at most 2 to every voter in this instance, while v_1 is at a distance of at least 3 from every candidate except c_1. So, if c_1 is not chosen, we get a distortion of at least $3/2$.

In instance I_2, we have voters v_1, v_2, v_3, v_4 located at points $-2, -1 - \varepsilon, 2 - \varepsilon, 2$ and candidates c_1, c_2, c_3, c_4 located at points $-4 + \varepsilon, 0, 2 - \varepsilon, 2$ respectively. Once again, we can see that the ordering for voters' preferences is respected:

$$\begin{aligned} &d(v_1, c_1) = 2 - \varepsilon < d(v_1, c_2) = 2 < d(v_1, c_3) = 4 - \varepsilon < d(v_1, c_4) = 4 \\ &d(v_2, c_2) = 1 + \varepsilon < d(v_2, c_1) = 3 - 2\varepsilon < d(v_2, c_3) = 3 < d(v_1, c_4) = 3 + \varepsilon \\ &d(v_3, c_3) = 0 < d(v_3, c_4) = \varepsilon < d(v_3, c_2) = 2 - \varepsilon < d(v_3, c_1) = 6 - 2\varepsilon \\ &d(v_4, c_4) = 0 < d(v_4, c_3) = \varepsilon < d(v_4, c_2) = 2 < d(v_4, c_1) = 6 - \varepsilon. \end{aligned}$$



(a) I_1: Choosing c_1 achieves a cost of 2, while the other candidates have a distances of at least 3 to v_1.



(b) I_2: Choosing c_2 achieves a cost of 2, while the other candidates have a distances of at least 3 to v_2.

Figure 10: Possible locations of the voters and candidates in the lower bound instance for $k = 3$.

In addition, since c_2 is at a distance of at most 2 from every voter and every candidate except c_2 has a distance of at least $3 - 2\varepsilon$ from v_2, not choosing c_2 leads to a distortion of at least $3/2 - \varepsilon$.

Based on the symmetry in voters' preferences between c_1, c_2 and c_4, c_3, we create mirror versions of I_1 and I_2 as I_4 and I_3 respectively, so that not choosing c_4 or c_3 would lead to a distortion of at least $3/2 - \varepsilon$. Therefore, since any deterministic algorithm must omit one of the candidates, we cannot achieve a distortion of better than $3/2 - \varepsilon$ in this case. □

References

[1] S. Alamdari and D. Shmoys. A bicriteria approximation algorithm for the k-center and k-median problems. In *Approximation and Online Algorithms: 15th International Workshop, WAOA 2017, Vienna, Austria, September 7–8, 2017, Revised Selected Papers 15*, pages 66–75. Springer, 2018.

[2] I. Anagnostides, D. Fotakis, and P. Patsilinakos. Dimensionality and coordination in voting: The distortion of stv. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4776–4784, 2022.

[3] E. Anshelevich, O. Bhardwaj, E. Elkind, J. Postl, and P. Skowron. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264:27–51, 2018.

[4] E. Anshelevich, A. Filos-Ratsikas, C. Jerrett, and A. A. Voudouris. Improved metric distortion via threshold approvals. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9460–9468, 2024.

[5] D. Black. On the rationale of group decision-making. *Journal of political economy*, 56(1):23–34, 1948.

[6] I. Caragiannis, E. Micha, and J. Peters. Can a few decide for many? the metric distortion of sortition. In *Forty-first International Conference on Machine Learning*.

[7] I. Caragiannis, N. Shah, and A. A. Voudouris. The metric distortion of multiwinner voting. *Artificial Intelligence*, 313:103802, 2022.

[8] M. Charikar, A. Lassota, P. Ramakrishnan, A. Vetta, and K. Wang. Six candidates suffice to win a voter majority. *arXiv preprint arXiv:2411.03390*, 2024.

[9] M. Charikar and P. Ramakrishnan. Metric distortion bounds for randomized social choice. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2986–3004. SIAM, 2022.

[10] M. Charikar, P. Ramakrishnan, K. Wang, and H. Wu. Breaking the metric voting distortion barrier. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1621–1640. SIAM, 2024.

[11] X. Chen, M. Li, and C. Wang. Favorite-candidate voting for eliminating the least popular candidate in a metric space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1894–1901, 2020.

[12] E. Elkind and P. Faliszewski. Recognizing 1-euclidean preferences: An alternative approach. In *International Symposium on Algorithmic Game Theory*, pages 146–157. Springer, 2014.

[13] E. Elkind, P. Faliszewski, P. Skowron, and A. Slinko. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48:599–632, 2017.

[14] E. Elkind, J. Lang, and A. Saffidine. Condorcet winning sets. *Social Choice and Welfare*, 44(3):493–517, 2015.

[15] J. M. Enelow and M. J. Hinich. *The spatial theory of voting: An introduction*. CUP Archive, 1984.

[16] B. Escoffier, J. Lang, and M. Öztürk. Single-peaked consistency and its complexity. In *ECAI 2008*, pages 366–370. IOS Press, 2008.

[17] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice*, 74(2017):27–47, 2017.

[18] D. Feldman, A. Fiat, M. Sharir, and D. Segev. Bi-criteria linear-time approximations for generalized k-mean/median/center. In *Proceedings of the twenty-third annual symposium on Computational geometry*, pages 19–26, 2007.

[19] M. Feldman, A. Fiat, and I. Golomb. On voting and facility location. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 269–286, 2016.

[20] V. Gkatzelis, D. Halpern, and N. Shah. Resolving the optimal metric distortion conjecture. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1427–1438. IEEE, 2020.

[21] A. Goel, R. Hulett, and A. K. Krishnaswamy. Relating metric distortion and fairness of social choice rules. In *Proceedings of the 13th Workshop on Economics of Networks, Systems and Computation*, pages 1–1, 2018.

[22] A. Goel, A. K. Krishnaswamy, and K. Munagala. Metric distortion of social choice rules: Lower bounds and fairness properties. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 287–304, 2017.

[23] D. Kempe. An analysis framework for metric voting based on lp duality. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2079–2086, 2020.

[24] F. E. Kizilkaya and D. Kempe. Generalized veto core and a practical voting rule with optimal metric distortion. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pages 913–936, 2023.

[25] F. E. Kizilkaya and D. Kempe. Plurality veto: A simple voting rule achieving optimal metric distortion, 2023.

[26] A. Lassota, A. Vetta, and B. von Stengel. The condorcet dimension of metric spaces. *arXiv preprint arXiv:2410.09201*, 2024.

[27] M. Mahdian, Y. Ye, and J. Zhang. Approximation algorithms for metric facility location problems. *SIAM Journal on Computing*, 36(2):411–432, 2006.

[28] J. A. Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.

[29] A. D. Procaccia and J. S. Rosenschein. The distortion of cardinal preferences in voting. In *International Workshop on Cooperative Information Agents*, pages 317–331. Springer, 2006.

[30] H. Pulyassary and C. Swamy. On the randomized metric distortion conjecture. *arXiv preprint arXiv:2111.08698*, 2021.

[31] D. Wei. A constant-factor bi-criteria approximation guarantee for k-means++. *Advances in neural information processing systems*, 29, 2016.