Deliberation via Matching[†]

Kamesh Munagala[‡] Qilin Ye[‡] Ian Zhang[‡]

November 2, 2025

Abstract

We study deliberative social choice, where voters refine their preferences through small-group discussions before collective aggregation. We introduce a simple and easily implementable *deliberation via matching* protocol: for each pair of candidates, we form an arbitrary maximum matching among voters who disagree on that pair, and each matched pair deliberates. The resulting preferences (individual and deliberative) are then appropriately weighted and aggregated using the weighted uncovered set tournament rule.

We show that our protocol has a tight distortion bound of 3 within the metric distortion framework. This breaks the previous lower bound of 3.11 for tournament rules without deliberation and matches the lower bound for deterministic social choice rules without deliberation. Our result conceptually shows that tournament rules are just as powerful as general social choice rules, when the former are given the minimal added power of pairwise deliberations. We prove our bounds via a novel *bilinear* relaxation of the non-linear program capturing optimal distortion, whose vertices we can explicitly enumerate, leading to an analytic proof. Loosely speaking, our key technical insight is that the distortion objective, as a function of metric distances to any three alternatives, is *both* supermodular and convex. We believe this characterization provides a general analytical framework for studying the distortion of other deliberative protocols, and may be of independent interest.

[†]This work is supported by NSF grant IIS-2402823.

[‡]Department of Computer Science, Duke University, Durham, NC 27708-0129, USA. Email: `{munagala,qilin.ye,ian.zhang}@duke.edu`.

1 Introduction

Collective decision-making lies at the core of both democratic governance and algorithmic social choice. Classical models assume that voters possess fixed, exogenous preferences over a set of alternatives, which are then aggregated through a social choice rule. Yet in practice, preferences are rarely static: individuals deliberate, exchange arguments, and frequently revise their views in response to others' reasoning. A large body of research in deliberative democracy, most notably "deliberative polling" and "citizens' assemblies", shows that when individuals are given balanced information and structured opportunities for deliberation, their policy preferences can shift substantially and become more informed [16, 22]. This underscores the normative intuition that collective decisions should emerge from public reasoning rather than isolated votes.

At the same time, empirical work indicates that deliberation is most effective in small groups. Large assemblies or unstructured online forums often suffer from coordination challenges, conformity pressures, and polarization effects, where participants reinforce existing biases [7, 14, 29]. In contrast, small, balanced groups promote reasoned exchange and opinion updating [22, 17], while maintaining manageability and diversity of perspectives. Beyond these empirical considerations, small-group deliberation is also more practical in large-scale settings: it can be implemented in parallel, either through many simultaneous discussions among small groups of participants, or via automated or AI-assisted mediators [6, 24, 28, 9, 8]. These advantages motivate theoretical models that capture the benefits of structured, small-group deliberation rather than full-group discussion.

Recent theoretical work has begun to formalize this intuition [15, 19]. In these models, voters engage in local deliberations that modify their ordinal preferences, and the resulting rankings are then aggregated using a social choice rule. Such frameworks allow us to study the fundamental algorithmic question of whether structured, small-group deliberation provably improves the efficiency of collective decisions.

We study this question through the lens of the *metric distortion* framework [2], a quantitative model for evaluating the efficiency of social choice rules. In this framework, both voters and alternatives are embedded in an unknown metric space that captures their underlying preferences: Voters prefer alternatives that are closer to them in this latent metric. A social choice rule then selects a single winner, based only on the voters' ordinal rankings over alternatives. The distortion of a rule measures how far the chosen winner can be from the voters in terms of total distance in the worst case, compared to the welfare-optimal alternative that minimizes the total distance to all voters had the latent metric been revealed. Thus, a smaller distortion indicates a decision rule that better preserves social welfare despite having only ordinal information.

Within this setting, it is known that any deterministic rule must incur a distortion of at least 3 [2, 27, 18, 25]. A prominent and widely studied subclass of such rules are *tournament rules*, which base their decision on the outcomes of pairwise contests between alternatives. These rules are quite classic, dating back to Ramon Llull in the 1300s [33], and further, are the simplest type of rules with bounded metric distortion. Tournament rules are appealing not only for their analytical simplicity but also for their low cognitive complexity: Voters need only compare two alternatives at a time rather than rank all options simultaneously. However, any tournament rule (that only uses pairwise ranking information about candidates) has a lower bound of 3.11 on distortion [12], which is worse than the deterministic optimum of 3.[1] This motivates the following question:

> *Can small-group deliberation, where voters refine their pairwise comparisons through deliberation, improve the distortion of tournament rules while preserving their simplicity?*

The recent work of [19] provided the first affirmative answer for *three-person deliberation*. In their model, every group of three voters deliberates between every pair of alternatives, and each group collectively chooses between any two alternatives by favoring the one with the smaller *average distance* in the latent metric, that is, the alternative closer to the group's barycenter. When the resulting tournament graph

[1] The best known upper bound on the distortion of tournament rules is much larger, around 3.93 [12].

is aggregated through the well-known Copeland tournament rule [33, 26], the authors showed that the distortion of such a three-person deliberation protocol is strictly better than 3, thereby surpassing the lower bounds for both tournament and general social choice rules without deliberation. This result established that structured local deliberation can provably improve social welfare. However, their analysis relied on solving a high-dimensional non-convex program numerically, and importantly, left open both the analytical tractability and the effectiveness of *two-person deliberation*, the smallest and most practical form of deliberation.

1.1 Our Contribution: Deliberation via Matching

In this paper, we propose a novel and natural protocol for deliberation based on *pairwise discussions* (groups of size 2). Our protocol, called *deliberation via matching*, proceeds as follows. For every pair of candidates (X, Y), we form an arbitrary maximum matching among voters who disagree on their relative ranking, and each matched pair deliberates. The result of each deliberation updates their pairwise preferences according to the sum of the latent distances in the underlying metric. These refined pairwise preferences are then aggregated using the *λ-weighted uncovered set* tournament rule [27, 23], where $\lambda \in [0.5, 1]$ is a parameter controlling the strength of dominance required in pairwise comparisons. A scalar parameter $w \geqslant 0$ controls the influence of deliberation: each matched pair contributes weight w to its joint outcome, while unmatched individual votes retain unit weight.

This protocol differs from prior work [19] that required all groups of voters of a fixed size to deliberate, in that (i) the protocol is more natural to state and is simpler to analyze, (ii) it is parsimonious in requiring only one deliberation per voter for each pair of candidates, and (iii) it allows precise control over how individual votes and pairwise deliberations between voters with opposing preferences are weighted when constructing the tournament graph.

Within the metric distortion framework, we prove the following main theorem:

Theorem 1.1 (Main Theorem, proved in Section 5)**.** *The deliberation-via-matching protocol with pairwise (two-person) deliberation achieves a metric distortion of* 3 *for an appropriate choice of* (λ, w).

This breaks the 3.11 lower bound for tournament rules without deliberation and matches the deterministic optimum of 3 for any social choice rule without deliberation. Conceptually, this shows that in the metric distortion framework, *tournament rules are just as powerful as general social choice rules, provided the former rules are given the minimal added power of pairwise deliberations*.

We complement this positive result with several lower bounds. We first show in Section 4.3 that *any deterministic social choice rule* that uses the outcomes of individual votes and pairwise deliberations has distortion at least 2, even for $m = 2$ candidates. In Section 4, we also show that for $m = 2$ candidates, this bound is tight, and the deliberation-via-matching protocol yields the optimal distortion of 2.

We finally show in Section 6 that the bound of 3 is optimal for the deliberation-via-matching protocol, in the sense that for any (λ, w), there is an instance with distortion at least 3. This shows that Theorem 1.1 cannot be improved for this protocol[2], though we leave open the question of matching the bound of 2 via a different protocol that uses pairwise deliberations (but possibly not a tournament social choice rule).

Technical Contribution: Bilinear Forms, Supermodularity, and Convexity. Beyond the quantitative bounds, our main technical contribution is to develop a novel analytical framework for studying deliberations. As pointed out in [19], the key difficulty in analyzing deliberative protocols is that the distortion objective is the solution to a non-linear, non-convex program over the distribution of voter–candidate distances, often with unbounded support. This contrasts with classical social choice, where distortion typically arises as the

[2]The lower bound also applies when the maximum matching is chosen in a benign fashion as opposed to adversarially.

solution to a linear program [23]. The resulting non-linearity severely limits both the classes of deliberative protocols that can be analyzed and the intuition one can draw from such analyses.

Our main contribution in Section 5 is to show that, for the deliberation-via-matching protocol, this non-linear program can be reformulated as a *bilinear optimization problem*, where the two linear components correspond to voter masses and metric distances, respectively, each with its own constraint set. This reformulation, which relies crucially on the "matching" structure of the protocol, enables both an analytic proof of the distortion bound and a clear characterization of the structure of worst-case instances.

In more detail, our first key observation in the reduction is that this objective has a *supermodular* structure in the metric distances. This structure allows us to show that the worst-case instance has voter preferences in a *monotonic* order, where the relative strengths of preferences to three given candidates are monotonic. We next show that our specific way of writing the objective function is also *convex* in the metric variables, which allows us to use Jensen's inequality to create a small collection of groups of voters based on how they are matched in deliberations, and what the outcomes of these matchings are. We collapse these groups into singleton weighted voters, yielding a bilinear objective with a small number of probability masses, and separate linear constraints on these masses and the metric distances. We then enumerate all vertices of the former polytope (at most six) and solve the resulting linear programs to show our distortion bound. We explicitly produce the corresponding dual certificates, yielding a fully analytic proof of the distortion bound.

As a warm-up, we analyze the special case with only two alternatives in Section 4. In this setting, the deliberation-via-matching rule admits a direct and elegant analysis: by pairing voters who disagree and letting each pair support the alternative with the smaller total distance, we show that any winner must be backed, in effect, by at least two-thirds of the electorate. This immediately yields an optimal distortion bound of 2, improving upon the classic bound of 3 for deterministic rules without deliberation in this case.[3] The two-candidate analysis captures the essential geometry of deliberation and serves as the basis for our general multi-candidate distortion bound in Section 5.

Taken together, our results suggest that small-group deliberation can be both *powerful* and *tractable*: Even minimal pairwise interactions suffice to make the well-studied tournament rules match the distortion bounds of general social choice rules. More broadly, our bilinear form characterization provides a new methodological foundation for analyzing deliberative extensions of social choice mechanisms.

1.2 Related Work

Our work lies at the intersection of metric distortion, deliberative social choice, and sampling-based decision mechanisms. We briefly touch on the most relevant lines of research.

Metric Distortion and Tournament Rules. The metric distortion framework was introduced by Anshelevich et al. [2], building on earlier work by Procaccia and Rosenschein [30], to study how well deterministic voting rules can approximate the social optimum when only ordinal information is available. They showed that the Copeland rule has distortion at most 5, and that no deterministic rule can achieve distortion below 3. Later work tightened the upper bound to 3 via novel social choice rules such as the *matching uncovered set* [18, 27] and *plurality veto* [25]. For randomized voting rules, the work of [5] showed a lower bound of 2. This lower bound was subsequently improved to 2.11 by [11]. An upper bound of 3 follows from random dictatorship [5], and this was improved to 2.74 in [13]. We refer the reader to [3] for a survey.

A particularly important subclass of deterministic voting rules are *tournament rules*, which make decisions based solely on the outcomes of pairwise majority contests between candidates. Tournament rules are attractive because they rely only on pairwise comparisons, requiring voters to reason about two alternatives

[3] This also yields a bound of 4 for the multiple candidate case via standard arguments [2].

at a time. On the positive side, Munagala and Wang [27] and Kempe [23] defined a class of *weighted tournament rules*, in which pairwise majority margins are aggregated with varying strengths. Their specific rule, the weighted uncovered set, achieves distortion at most $2 + \sqrt{5} \approx 4.236$. Subsequently, Charikar et al. [12] gave a rule with an improved upper bound of 3.93, and proved a lower bound of approximately 3.11 on the distortion of any deterministic tournament rule, improving a lower bound of 3 in [20]. This result shows that in the absence of deliberation, tournament rules cannot match the optimal distortion of 3 achievable by general deterministic mechanisms, so that these rules are fundamentally limited despite their simplicity and centuries-long history [33].

Our work revisits this barrier through the lens of deliberation. We show that by allowing pairs of voters to refine their comparisons before aggregation, a tournament-based rule can in fact achieve the optimal distortion bound of 3 without deliberation, thereby escaping the 3.11 bound for non-deliberative tournaments.

Deliberative Social Choice: From Sortition to Dyads. The idea that deliberation can improve collective decisions has a long pedigree in political philosophy and deliberative democracy, for example through deliberative polling and citizens' assemblies [16, 22]. Many deliberative systems in practice use *sortition*, which is the random sampling of participants into discussion bodies, to reduce biases and improve legitimacy.

Several theoretical models of voter interaction within the metric distortion framework have been recently proposed [1, 4, 21]. Caragiannis et al. [10] examine models of sortition where a large random sample of voters deliberates to compute a consensus or median point, achieving a logarithmic (in the number of alternatives) bound on the sample size required to attain distortion arbitrarily close to one. However, this assumes a single large deliberative body, which raises issues of coordination and bias in practice.

In contrast, our focus is on *small-group deliberation* rather than sortition. Here, Fain et al. [15] studied a two-person bargaining model under metric preferences, while Goel, Goyal, and Munagala [19] proposed a general model in which all groups of up to k voters deliberate between pairs of alternatives and the resulting tournament aggregated via the Copeland rule. They showed that groups of size $k = 3$ suffice to beat the deterministic distortion bound, with an analysis that relied on numerical optimization. For groups of size $k = 2$ (the setting we consider), they showed a distortion bound of 4.414, which we vastly improve to 3. We note that the protocol in [19] required all pairs of voters to deliberate between a pair of alternatives, regardless of their preference between them, while our protocol requires only one deliberation per voter for a pair of alternatives.

Finally, focusing on two-person deliberation is natural, since most discussions unfold through back-and-forth exchanges between pairs of participants. Communication research models such dyadic interactions as the basic unit of conversational dynamics [32, 31], and many deliberative settings, such as in-person debates and online replies, can be viewed as networks of such exchanges [34]. Modeling this atomic form of deliberation allows us to capture the core mechanism by which deliberation influences preferences and to develop analytic tools that extend naturally to larger deliberative settings.

In summary, our approach departs from prior work in both the protocol and analysis technique: It relies solely on two-person deliberations, with one per voter per candidate pair, and precisely balances individual and deliberative inputs. Moreover, unlike previous analyses that required numerical optimization, our results follow from an explicit bilinear formulation that yields an analytic proof of optimality.

2 Preliminaries

We begin by reviewing the metric distortion framework and the class of tournament rules used in our analysis, following [2, 27, 19].

Metric Distortion Framework. Let $\mathcal{C} = \{c_1, \ldots, c_m\}$ denote a finite set of m candidates (alternatives), and let V denote a finite set of n voters. Each voter $v \in V$ has a ranking over the candidates that is *consistent* with an underlying latent metric space $(\mathcal{M}, d)$ that contains both voters and candidates as points. If v ranks candidate X higher than Y, then $d(v, X) \leqslant d(v, Y)$. The metric d is not known to the social planner, who only observes the ordinal rankings induced by it. For any two candidates $X, Y \in \mathcal{C}$, let XY denote the set of voters who prefer X to Y, with cardinality $|XY|$. Should ties exist, i.e., $d(v, X) = d(v, Y)$, we handle them in any consistent way that counts each tied voter toward exactly one of XY, YX. We let σ be the set of preference orderings over candidates for each voter.

For any candidate $X \in \mathcal{C}$, we define its *social cost* with respect to a metric d to be

$$SC(X, d) = \sum_{v \in V} d(v, X).$$

When the metric d is clear from context, we simply write $SC(X)$. Let $X^* = \arg\min_{X \in \mathcal{C}} SC(X)$ denote the socially optimal (1-median) alternative. Given a social choice rule $\mathcal{S}$ that maps the profile of rankings to a winning candidate $\mathcal{S}(\sigma)$, the *distortion* of $\mathcal{S}$ is defined as

$$\text{Distortion}(\mathcal{S}) = \sup_{\sigma} \sup_{d \text{ consistent with } \sigma} \frac{SC(\mathcal{S}(\sigma), d)}{SC(X^*, d)}.$$

A smaller distortion indicates that $\mathcal{S}$ achieves better welfare despite only knowing ordinal information.

Tournament Rules. A *tournament graph* on the candidates is a complete directed graph, with weights $f(XY) \in [0, 1]$ for each directed edge $X \to Y$, so that for every pair of candidates (X, Y), we have $f(XY) + f(YX) = 1$. In the setting without deliberation, $f(XY)$ represents the fraction of voters that prefer X over Y; however, the weights we construct later will also reflect the outcome of deliberation. A tournament rule takes such a weighted graph as input and outputs the winning candidate.

Among many tournament-based social choice rules, we focus on the *λ-weighted uncovered set (WUS)* rule of [27, 23], which builds on the classic uncovered set rules [26]. Given a tournament with edge weights $f(XY) \in [0, 1]$, a candidate X is in the *λ-weighted uncovered set* WUS_λ if for every other candidate Y, either

1. $f(XY) \geqslant 1 - \lambda$, or
2. there exists a third candidate Z such that $f(XZ) \geqslant 1 - \lambda$ and $f(ZY) \geqslant \lambda$.

It is known that for $\lambda \in [1/2, 1]$, WUS_λ is nonempty [27]. Furthermore, for $\lambda = (\sqrt{5} - 1)/2 \approx 0.618$, the rule selecting any candidate from WUS_λ achieves distortion at most $2 + \sqrt{5} \approx 4.236$ [27, 23]. The special case where $\lambda = 1/2$ is the standard notion of uncovered set [26]; the classic Copeland rule due to Llull [33] that chooses any candidate that beats the most number of others in simple majority voting between them chooses an outcome that belongs to this set.

Small-Group Deliberation. We next recall the 2-person deliberation model with *averaging* introduced in [19]. A deliberation involves two voters u, v and a pair of candidates (X, Y). Under the *averaging model*, the pair collectively supports the alternative with smaller total distance, or equivalently,

$$X \text{ wins against } Y \quad \text{iff} \quad d(u, X) + d(v, X) \leqslant d(u, Y) + d(v, Y).$$

3 Deliberation via Matching Protocol

We now describe our main protocol, *Deliberation via Matching*, which implements two-person deliberation between voters who disagree on a pair of candidates. The protocol defines a weighted tournament over candidates, parameterized by a deliberation weight $w \geqslant 0$ and the λ-weighted uncovered set parameter $\lambda \in [1/2, 1]$. These parameters will be optimized later.

Matching Step. Fix two distinct candidates $X, Y \in C$. Let XY denote the set of voters who prefer X to Y, and YX denote those who prefer Y to X.

Form an arbitrary maximum cardinality matching M_{XY} between voters in XY and voters in YX; that is, select $|M_{XY}| = \min\{|XY|, |YX|\}$ disjoint pairs (u_i, v_i) with $u_i \in XY$ and $v_i \in YX$ for $i = 1, \cdots, |M_{XY}|$. Each pair (u_i, v_i) represents a two-person deliberation between voters with opposing preferences on (X, Y). Any remaining voters (those not matched) are said to be *unmatched*. Note all unmatched voters must have the same preference: either they all prefer X (if $|XY| \geqslant |YX|$) or all prefer Y (if $|XY| < |YX|$).

In the averaging model of deliberation, let W_{XY} denote the number of matched pairs that favor X, and $W_{YX} = |M_{XY}| - W_{XY}$ the number that favor Y.

Aggregation Step. We define the *weighted pairwise score* of X against Y as

$$\text{score}(XY; w) = \frac{|XY| + w \cdot W_{XY}}{n},$$

and symmetrically $\text{score}(YX; w) = (|YX| + w \cdot W_{YX})/n$. We divide by n so that the score() function is independent of n, the number of voters. The total score for the pair (X, Y) is therefore $\text{score}(XY; w) + \text{score}(YX; w) = 1 + w \cdot |M_{XY}|/n$. We define the normalized score to be

$$f(XY; w) = \frac{\text{score}(XY; w)}{\text{score}(XY; w) + \text{score}(YX; w)} \tag{1}$$

and define $f(YX; w)$ likewise so that $f(XY; w) + f(YX; w) = 1$. When the context is clear (e.g. w is a prescribed constant), we may simply write $f(XY; w)$ and $\text{score}(XY; w)$ as $f(XY)$ and $\text{score}(XY)$.

Applying the above procedure to every ordered pair of candidates (X, Y) defines a weighted tournament graph on C where the weight on edge (X, Y) is $f(XY; w)$. The final collective decision is obtained by applying the λ-weighted uncovered set rule WUS_λ (as defined in Section 2) to this tournament.

Parameters. The protocol is governed by two parameters:

- the *deliberation weight* $w \geqslant 0$, controlling the relative influence of two-person deliberation outcomes versus individual preferences, and
- the *uncovering parameter* $\lambda \in [1/2, 1]$, which determines the strength of the dominance condition used in the λ-weighted uncovered set rule.

When $w = 0$, the protocol reduces to a standard tournament rule without deliberation. As w increases, the outcomes of matched deliberations receive greater emphasis, interpolating smoothly between non-deliberative aggregation and fully deliberative pairwise refinement.

4 Warm-up: A Simple Distortion Bound for the Copeland Rule

We first consider the setting in deliberation-via-matching where we set $\lambda = 0.5$ and $w = 1$. This means the deliberation outcomes are given the same importance as individual votes, and we run the Copeland rule to aggregate the tournament into a winner. In the Copeland rule, candidate A beats B if $f(AB; 1) \geqslant 0.5$. The rule outputs any candidate that beats the greatest number of other candidates. We show that this protocol has distortion exactly 4.

Towards this end, we analyze the setting with only $m = 2$ candidates and show a distortion of 2. Since the Copeland winner lies in the uncovered set [2], a standard argument shows that the distortion for any $m \geqslant 2$ candidates will be at most the square of the distortion for two candidates, showing an upper bound of

4 for general number of candidates. Despite the simplicity of this analysis, we show that the bound of 4 is tight for this setting of (λ, w).

For $m = 2$ candidates, we note that in the absence of deliberation, any deterministic social choice rule has a worst-case distortion of 3 [2], while we show that the deliberation-via-matching protocol achieves a distortion of 2. We further show that this bound is tight for $m = 2$ candidates *regardless* of the deterministic social choice rule used, and the way pairwise deliberations are constructed. This provides an unconditional lower bound for metric distortion with pairwise deliberations. Similarly, we show a lower bound of 1.5 for randomized social choice rules.

The setting of (λ, w) in this section isolates the geometric effect of pairwise deliberation without the additional complexity of tournament aggregation. It therefore acts as a warm-up for the more general analysis in the following section, where we extend the same reasoning to find the optimal (λ, w) that yields distortion 3.

4.1 Preliminaries

Since this section mainly focuses on the $m = 2$ candidate case, we specialize the notation to this setting. Let the candidates be A and B, separated by distance $d(A, B)$ in the latent metric. Let AB (respectively BA) denote the set of voters who prefer A (respectively B), so that $|AB| + |BA| = n$ is the total number of voters. Let M denote the arbitrary matching formed between voters in AB and those in BA according to the deliberation-via-matching protocol (Section 3). Each matched pair $(u, v) \in M \subseteq AB \times BA$ deliberates between A and B and supports the alternative with the smaller total distance to the pair. Define

$$M_A = \{(u, v) \in M : A \text{ wins}\} = \{(u, v) \in M : d(u, A) + d(v, A) \leqslant d(u, B) + d(v, B)\},$$
$$M_B = \{(u, v) \in M : B \text{ wins}\} = \{(u, v) \in M : d(u, A) + d(v, A) > d(u, B) + d(v, B)\}.$$

Observe M_A, M_B partition M, and recall that the number of A-wins pairs (resp. B-win pairs) are $W_A = |M_A|$ (resp. $W_B = |M_B|$) by definition. The electorate now splits into three types of voters: (i) Those that contribute to A-wins, grouped as pairs from $AB \times BA$; (ii) Those that contribute to B-wins, also grouped as pairs; and (iii) Unmatched voters, all of whom belong to AB if $|AB| \geqslant |BA|$ and BA otherwise. Ties can be apportioned in any way as long as every tie pair is counted once.

In the protocol in Section 3, we will set $\lambda = 1/2$ and $w = 1$. This means we set

$$\text{score}(AB) = \frac{|AB| + W_A}{n},$$

and apply the Copeland rule with $f(AB) = \text{score}(AB)/(\text{score}(AB) + \text{score}(BA))$, so that A is the winner if $\text{score}(AB) \geqslant \text{score}(BA)$, and B is the winner otherwise.

We note that the classic Copeland rule declares A as the winner if and only if $|AB| \geqslant |BA|$; it is well known that this rule, as well as any other deterministic rule relying solely on ordinal information, has distortion $\geqslant 3$ even on two candidates [2]. With deliberation, we instead declare A as the winner if and only if $|AB| + W_A \geqslant |BA| + W_B$, and we show this simple change leads to an improved distortion of 2.

4.2 Analysis of the Copeland Rule for Two Candidates

Assume A is the winner. To bound the distortion, we aim to maximize $SC(A)/SC(B)$, where SC denotes the social cost.

Upper-bounding $SC(A)$. For every voter v, we have by triangle inequality

$$d(v, A) \leqslant d(v, B) + \mathbf{1}[v \in BA] \cdot d(A, B) = \begin{cases} d(v, B) & \text{if } v \in AB \\ d(v, B) + d(B, A) & \text{if } v \in BA. \end{cases} \tag{2}$$

Based on the outcomes of the matching, we split $SC(A)$ into three sums and analyze them separately:

$$SC(A) = \sum_{(u,v) \in M_A} [d(u, A) + d(v, A)] + \sum_{(u,v) \in M_B} [d(u, A) + d(v, A)] + \sum_{v \text{ unmatched}} d(v, A).$$

- For $(u, v) \in M_A$: as A wins the deliberation, we have $d(u, A) + d(v, A) \leqslant d(u, B) + d(v, B)$.
- For $(u, v) \in M_B$: assume $u \in AB$ and $v \in BA$, so that the corresponding applications of Equation (2) give $d(u, A) + d(v, A) \leqslant d(u, B) + d(v, B) + d(A, B)$.
- Equation (2) is also directly applicable on the sum over unmatched voters.

Observe that the total additional copies of $d(A, B)$ that appear in $SC(A)$ equals W_B plus number of unmatched BA voters; this is equivalent to $|BA| - W_A$. Hence,

$$SC(A) \leqslant SC(B) + (|BA| - W_A) \cdot d(A, B). \tag{3}$$

Lower-bounding $SC(B)$. For any pair $(u, v) \in M_A$, the deliberation constraint and triangle inequality imply

$$\begin{cases} d(u, B) + d(v, B) \geqslant d(u, A) + d(v, A) \\ d(u, A) + d(u, B) \geqslant d(A, B) \\ d(v, A) + d(v, B) \geqslant d(A, B) \end{cases} \implies d(u, B) + d(v, B) \geqslant d(A, B).$$

We now lower bound $SC(B)$ as follows:

- Each $(u, v) \in M_A$ contributes $d(A, B)$ to $SC(B)$, and there are W_A such pairs.
- The remaining $|AB| - W_A$ voters in AB each contribute at least $d(A, B)/2$ to $SC(B)$, since $d(v, A) \leqslant d(v, B)$ and $d(v, A) + d(v, B) \geqslant d(A, B)$, which imply $d(v, B) \geqslant d(A, B)/2$.

Therefore,

$$SC(B) \geqslant W_A \cdot d(A, B) + (|AB| - W_A) \cdot d(A, B)/2 = (|AB| + W_A)/2 \cdot d(A, B). \tag{4}$$

Combining Equation (3) and Equation (4), we see that

$$\begin{aligned} \frac{SC(A)}{SC(B)} &\leqslant \frac{SC(B) + (|BA| - W_A) \cdot d(A, B)}{SC(B)} \leqslant 1 + \frac{(|BA| - W_A) \cdot d(A, B)}{(|AB| + W_A)/2 \cdot d(A, B)} \\ &= 1 + \frac{2(|BA| - W_A)}{|AB| + W_A} = \frac{2n}{|AB| + W_A} - 1 = \frac{2}{\text{score}(AB; w = 1)} - 1. \end{aligned} \tag{5}$$

We now bound the distortion of the protocol.

Theorem 4.1. *The metric distortion of the deliberation via matching protocol with the Copeland Rule for any 2-candidate instance is bounded by* 2*.*

Proof. By Equation (5), it suffices to show that if A wins, then $\text{score}(AB) \geqslant 2/3$. To prove this claim, we first assume $|AB| \leqslant |BA|$, so that $|AB| = W_A + W_B$. Since A is the winner,

$$|AB| + W_A \geqslant |BA| + W_B = |BA| + (|AB| - W_A) = n - W_A \qquad \Rightarrow \qquad 2W_A \geqslant n - |AB|.$$

But we also have $W_A \leqslant |AB|$, so $|AB| \geqslant n/3$. Hence

$$n \cdot \text{score}(AB) = |AB| + W_A \geqslant |AB| + \frac{n - |AB|}{2} = \frac{n + |AB|}{2} \geqslant \frac{n + (n/3)}{2} = \frac{2n}{3}.$$

If instead $|AB| \geqslant |BA|$ so that $|BA| = W_A + W_B$ and $|AB| \geqslant n/2$, then since A is the winner,

$$|AB| + W_A \geqslant |BA| + W_B = |BA| + (|BA| - W_A) = 2|BA| - W_A \qquad \Rightarrow \qquad 2W_A \geqslant 2|BA| - |AB|.$$

If $|AB| \geqslant 2n/3$ there is nothing to show, so we assume $n/2 \leqslant |AB| \leqslant 2n/3$. In this case, the above inequality becomes $2W_A \geqslant 2(n - |AB|) - |AB| = 2n - 3|AB|$. Then,

$$n \cdot \text{score}(AB) = |AB| + W_A \geqslant |AB| + \frac{2n - 3|AB|}{2} = \frac{2n - |AB|}{2} \geqslant \frac{2n - 2n/3}{2} = \frac{2n}{3}.$$ □

By the uncovered set property of the Copeland rule, the distortion for any number of candidates is upper bounded by the square of the distortion on two candidates [2]. This directly implies the following corollary.

Corollary 4.2. *For any number m of candidates, the deliberation-via-matching protocol with $\lambda = 0.5$ and $w = 1$ has distortion at most 4.*

4.3 Lower Bounds

We first show the following lower bound on the distortion of *any* social choice rule that only uses voter preferences and the outcomes of pairwise deliberations. In particular, this shows that the bound in Theorem 4.1 is tight for $m = 2$ candidates, and cannot be improved by either running the deliberations differently or using a different social choice rule.

Theorem 4.3. *Any deterministic social choice rule that uses individual preferences and the outcomes of pairwise deliberations has distortion at least 2, even with $m = 2$ candidates.*

Proof. We construct two instances X and Y with two candidates A and B, which have the same voter preferences, but $SC(B)/SC(A) = 2$ in X and $SC(A)/SC(B) = 2$ in Y. In both instances, the metric is on a line where A is at -1 and B is at 1. For X, we place two voters at $A = -1$ and one voter at $B = 1$, and set the deliberation between a voter at -1 and a voter at 1 to prefer B. For Y, we place two voters at 0 (which prefer A) and one voter at B.

The preference profile of the voters and the deliberation profiles are identical for these two instances. Thus no deterministic social choice rule can give distortion better than 2, regardless of the protocol used for constructing deliberating pairs. □

The same pair of instances shows the following corollary. The proof follows by observing that the best any social choice rule can do on the above instance is randomize equally between A and B.

Corollary 4.4. *Any randomized social choice rule that uses individual preferences and the outcomes of pairwise deliberations has distortion at least 1.5, even with $m = 2$ candidates.*

We finally show that for the setting of $\lambda = 0.5, w = 1$, the deliberation-via-matching rule has distortion exactly 4 for any $m \geqslant 2$ candidates, showing the analysis in Corollary 4.2 is tight.

Theorem 4.5. *The deliberation-via-matching protocol with $\lambda = 0.5, w = 1$ has distortion at least* 4.

Proof. We construct an instance with 3 candidates A, B, and C, and 3 voters. The metric is a line and we place A at 0, B at 1, and C at 2. We place two voters at $B = 1$ who prefer A over C and one voter at $C = 2$. We also set the deliberations between a voter at $B = 1$ and a voter at $C = 2$ to prefer C. When $w = 1$, we see that $f(AC) = f(CB) = 0.5$. Thus, candidate A is in the λ-uncovered set for $\lambda = 0.5$. Since $SC(A)/SC(B) = 4$, the distortion is at least 4. □

5 Optimal Distortion Bound: Proof of Theorem 1.1

The analysis in Section 4 focused on the 2-candidate setting and squared the resulting distortion bound. Though this yielded a tight bound for the specific setting of $\lambda = 0.5$ and $w = 1$, the same approach will not yield a better bound than 4 for other (λ, w). For analyzing the general setting of λ-WUS, we follow previous work [2, 27] and consider three candidates A, B, C, where the rule selects A, the social optimum is B, and there is a candidate C such that $f(AC; w) \geqslant 1 - \lambda$ and $f(CB; w) \geqslant \lambda$. Finding the resulting worst case distortion now becomes the solution to a non-linear program with unbounded support, and simplifying this program to show a tight distortion bound forms our main technical contribution.

Specifically, we show a specific setting of (λ, w) which yields distortion bound 3 for the deliberation via matching protocol, showing Theorem 1.1. We further show in Section 6 that *any* choice of (λ, w) yields distortion at least 3 for the deliberation via matching protocol, showing our analysis is tight.

5.1 The λ-Weighted Uncovered Set

In this section, we first assume a fixed (λ, w) and use them implicitly to ease notation. We later optimally choose these parameters in Section 5.5 and analyze the distortion with the chosen parameters. For now, recall that, for any ordered pair of candidates (X, Y), we defined

$$\text{score}(XY; w) = |XY| + w \cdot W_{XY}, \qquad f(XY; w) = \frac{\text{score}(XY; w)}{\text{score}(XY; w) + \text{score}(YX; w)} \tag{6}$$

as in Equation (1), where $|XY|$ is number of voters preferring X to Y, W_{XY} is the number of deliberation pairs that favor X, and $w \geqslant 0$ controls the weight placed on the deliberative outcomes. We then select a winner using the λ-weighted uncovered set rule on this tournament by selecting any candidate in the *λ-weighted uncovered set* WUS_λ as the winner. Throughout this section, we write $f(XY)$ and $\text{score}(XY)$, with the w-dependence implicit whenever the context is clear.

Using the analysis technique for uncovered set tournament rules in [2, 27], suppose B is the optimal candidate and A is the outcome of our protocol. Either $f(AB) \geqslant 1 - \lambda$ directly, or there exists another candidate C such that $f(AC) \geqslant 1 - \lambda$ and $f(CB) \geqslant \lambda$. It therefore suffices to consider three such candidates A, B, C and the worst-case instance over these as:

$$\begin{aligned} \text{Distortion} = \quad & \sup \ \frac{SC(A)}{SC(B)} \\ \text{Subject to} \quad & \text{either } \ (f(AB) \geqslant 1 - \lambda), \\ & \text{or} \quad \big(f(AC) \geqslant 1 - \lambda \text{ and } f(CB) \geqslant \lambda\big). \end{aligned} \tag{7}$$

Since the first case $f(AB) \geqslant 1 - \lambda$ is a restriction of the second case with $C = B$, it further suffices to upper bound the distortion in the second case only, with $f(AC) \geqslant 1 - \lambda$ and $f(CB) \geqslant \lambda$.

As two maximum matchings ((A, C) on $AC \times CA$ and (C, B) on $CB \times BC$) will occur, we introduce the following notation to avoid ambiguity by highlighting both alternatives involved. We say a voter pair (u, v)

is an $\boldsymbol{A \succ C}$ **pair** if the pair deliberates in the (A, C) matching and favors A. Similarly, a $C \succ A$ pair is one that deliberates in the (A, C) matching and favors C. The $C \succ B$ and $B \succ C$ pairs are defined analogously.

5.2 A Mathematical Program for Distortion

We now show a change of variables under which the above program can be written with a bilinear objective, and separate constraints capturing the metric and the voter distribution on this metric.

Definition 5.1. Given an instance, define three variables X, Y, Z on the electorate V by

$$X(v) = d(v, C) - d(v, A), \qquad Y(v) = d(v, B) - d(v, C), \qquad Z(v) = d(v, C). \tag{8}$$

Then $X(v)$ quantifies voter v's relative preference between A and C, and $Y(v)$ between C and B.

We will show below that for the instances that achieve worst case distortion, it suffices to specify the values of (X, Y, Z) for each voter. This definition also directly implies that X encodes all information needed to determine outcomes of the (A, C) deliberation: $|AC|, |CA|$, as well as the distribution of voters in V. Consequently, we may compute $f(AC)$ given X. Likewise, Y encodes $f(CB)$. To see this, take the (A, C) matching for instance, in which the protocol is concerned with two properties: (i) whether a voter v prefers A over C, i.e., $d(v, A) \leqslant d(v, C)$, and (ii) whether a pair of voters $(u, v) \in AC \times CA$ collectively prefers A over C, i.e., $d(u, A) + d(v, A) \leqslant d(u, C) + d(v, C)$. Our X encodes both: $v \in AC$ if and only if $X(v) \geqslant 0$, and the deliberation of (u, v) outputs $(A \succ C)$ if and only if $X(u) + X(v) \geqslant 0$. Therefore, the values of X, Y on V completely encode all quantities of interest.

Continuum of Voters and the Objective Function. In the discussion below, we will consider a more general setting where the underlying metric space is finite and voters V form a distribution over this metric, and define a continuum. This essentially relaxes the discrete voters into a continuous space, and such a transformation cannot reduce distortion, since the discrete case is a special case. We therefore view the voters as forming a distribution over the metric space. We will write $\rho(v)$ as the density of a voter at v and normalize $\sum_{v \in V} \rho(v)$ into unit mass. The variables $X, Y \colon V \to \mathbb{R}$ are fixed by the instance and determine $f(AC), f(CB)$ through their one-dimensional distributions, $\mathcal{D}_X, \mathcal{D}_Y$.

It now follows from Definition 5.1 that $SC(A)/SC(B) = [\mathbb{E}Z - \mathbb{E}X]/[\mathbb{E}Z + \mathbb{E}Y]$, where the expectation is over the distribution of voters over the underlying metric space. We now transform the objective into a linear form, observing for $R > 0$ that if $SC(A)/SC(B) = [\mathbb{E}Z - \mathbb{E}X]/[\mathbb{E}Z + \mathbb{E}Y] > R + 1$, then

$$\mathbb{E}X + (R + 1) \cdot \mathbb{E}Y + R \cdot \mathbb{E}Z < 0. \tag{9}$$

We will choose R appropriately and show that the global minimum of the LHS of Equation (9) is at least zero, and this will imply a distortion of at most $R + 1$.

Note that this objective is bilinear, since both the values (X, Y, Z) and the voter distribution over these values are variables, and the support of (X, Y, Z) can be unbounded. Our main contribution below is to relax the problem so that this support becomes constant, and the constraints capturing λ-WUS become linear.

Remark. At several points, we will use an exchange argument over pairs of voters; these arguments can be extended to the continuum over voters by shifting the probability mass appropriately, and we omit the simple details. Further, since we assumed the metric space is finite, the optimization problem above will also have finite size, with the variables corresponding to metric distances and voter masses. We will transform this program in several steps below, noting that these steps will preserve the finite nature of the program.

Simplifying Z. We first show that the worst-case instances will use a specific setting of Z as a function of (X, Y). We subsequently analyze properties of this function. Note from Equation (9) that given fixed X and Y, we should point-wise minimize Z such that $\{(X(v), Y(v), Z(v))\}_{v \in V}$ is still metric feasible in the sense that Equation (8) can be realized in some latent metric space. This leads to the following key lemma. In the lemma below, by $\|X\|_\infty$, we mean $\max_v |X(v)|$.

Lemma 5.2. *Fix real-valued functions X, Y on the electorate V. For any real-valued function Z on V, in order for (X, Y, Z) to be realized by some metric d under Equation (8), it is necessary and sufficient that*

$$Z(v) \geqslant Z_{\min}(v) = \max\left\{\frac{\|X\|_\infty + X(v)}{2}, \frac{\|Y\|_\infty - Y(v)}{2}, \frac{\|X+Y\|_\infty + X(v) - Y(v)}{2}\right\} \qquad \textit{for all } v. \tag{10}$$

Proof. We first prove necessity. Because d is nonnegative, we must have $d(v, C) = Z(v) \geqslant 0, d(v, A) = Z(v) - X(v) \geqslant 0$, and $d(v, B) = Z(v) + Y(v) \geqslant 0$ from Equation (8). Triangle inequalities for (v, A, C) imply

$$|d(v, A) - d(v, C)| = |X(v)| \leqslant d(A, C) \leqslant d(v, A) + d(v, C) = 2Z(v) - X(v).$$

Taking supremum over the first $\leqslant$ gives $d(A, C) \geqslant \|X\|_\infty$; combining with the second $\leqslant$ gives

$$2Z(v) - X(v) \geqslant \|X\|_\infty \qquad \text{so} \qquad Z(v) \geqslant \frac{\|X\|_\infty + X(v)}{2}. \tag{11}$$

The remaining two terms can be obtained analogously by enforcing triangle inequalities on (v, B, C) and (v, A, B), respectively.

For sufficiency, assume Equation (10) and define $d(A, C) = \|X\|_\infty, d(B, C) = \|Y\|_\infty$, and $d(A, B) = \|X + Y\|_\infty$. Let $d(v, C) = Z(v)$, $d(v, A) = Z(v) - X(v)$, and $d(v, B) = Z(v) + Y(v)$. Then (A, B, C) satisfy triangle inequalities, and for each voter, the inequalities established in the necessity part show triangle inequality: For instance, for (v, A, C), we have

$$|d(v, A) - d(v, C)| = |X(v)| \leqslant \|X\|_\infty = d(A, C) \leqslant 2Z(v) - X(v) = d(v, A) + d(v, C),$$

and likewise for (v, B, C) and (v, A, B), so triangle inequalities also hold among these pairs. Finally, to complete the metric, it remains to specify voter-to-voter distances. Note that the current metric defines a graph on $V \cup \{A, B, C\}$ with edges between every pair of candidates, and between each voter and candidate. Thus for two voters $u \neq v$, we can define $d(u, v)$ to be the distance between u and v in this graph. □

Unless otherwise indicated, given a pair (X, Y) defined on V, we will from now on default to defining Z by $Z_{\min}(X, Y)$ as stated in Equation (10).

The Bilinear Objective. From Definition 5.1, X alone determines the (A, C) matching and thus $f(AC)$; similarly Y determines $f(CB)$. From Equation (9), given $R > 0$, the distortion is at most $R + 1$ if the following functional is non-negative:

$$\Phi_R(X, Y) = \mathbb{E}X + (R+1) \cdot \mathbb{E}Y + R \cdot \mathbb{E}[Z_{\min}(X, Y)].$$

Combining these observations, we obtain the following mathematical program with bilinear objective:

$$\begin{array}{rl} \text{Minimize} & \Phi_R(X, Y) = \mathbb{E}X + (R+1) \cdot \mathbb{E}Y + R \cdot \mathbb{E}Z \\ \text{over} & X, Y \text{ on } V, \quad Z = Z_{\min}(X, Y) \text{ from Equation (10)} \\ \text{Subject to} & \text{(i) } f(AC) \text{ is induced by } \textit{some} \text{ matching determined by } X; \\ & \text{(ii) } f(CB) \text{ is induced by } \textit{some} \text{ matching determined by } Y; \\ & \text{(iii) } f(AC) \geqslant 1 - \lambda, \quad f(CB) \geqslant \lambda. \end{array} \tag{12}$$

Note that the above optimization is both over (X, Y) and the choice of the matchings given (X, Y). We will subsequently show that the optimal matching has a specific form that enables writing the final constraints as a set of linear constraints where the variables capture the distribution of the voters. This will make the entire program bilinear, with separate linear constraints for (X, Y) and for the distribution. In summary, we seek to find the smallest R^* under which the infimum of feasible Φ's remains non-negative. Define

$$\mathrm{OPT}(R) = \inf\{\Phi_R(X, Y) : (X, Y) \text{ feasible under Equation (12)}\}, \qquad R^* = \inf\{R > 0 : \mathrm{OPT}(R) \geqslant 0\}.$$

Then the supremum of distortion equals $R^* + 1$.

5.3 Super-modularity and Counter-monotone Coupling

In this section, we show a key structural property of $Z_{\min}$ from Equation (10): it is *supermodular* as a function of X, Y. The objective (Equation (7)) is minimized when $Z = Z_{\min}$, and by supermodularity, this happens when X is paired counter-monotonically with Y, as defined below.

We define a **coupling** of X and Y to be any joint assignment $\{(X(v), Y(v)) : v \in V\}$ that preserves the multisets $\{X(v)\}, \{Y(v)\}$ (equivalently, their distributions $\mathcal{D}_X, \mathcal{D}_Y$ under ρ). By Definition 5.1, the $f(\cdot)$-constraints are oblivious to the choice of coupling. Because $\mathbb{E}X$ and $\mathbb{E}Y$ are coupling-invariant, in the function $\Phi_R(X, Y)$, only the term $Z = Z_{\min}(X, Y)$ may change as we vary the coupling. Below, we prove that whenever two voters v_1, v_2 satisfy $X(v_1) < X(v_2)$ and $Y(v_1) < Y(v_2)$, swapping their Y-values weakly decreases $\mathbb{E}Z$ and hence the objective. Consequently, the optimal coupling is counter-monotone: descending X values are paired with ascending Y values. This shows that it suffices to examine instances whose induced variables X and Y from Definition 5.1 are coupled in this manner, and we will do so once the following lemma is proven.

Lemma 5.3 (Counter-monotone Coupling of X, Y)**.** *Fix distributions $\mathcal{D}_X, \mathcal{D}_Y$ on V and $R > 0$. Then, over $X \sim \mathcal{D}_X$ and $Y \sim \mathcal{D}_Y$, the objective $\Phi_R(X, Y)$ is minimized by one where X and Y are coupled counter-monotonically: if $X(v_1) \leqslant X(v_2)$ then $Y(v_1) \geqslant Y(v_2)$.*

Given $X \in \mathcal{D}_X$ and $Y \in \mathcal{D}_Y$, their expectations are fixed. Then, as discussed earlier, for any R, to minimize $\Phi_R(X, Y)$ in Equation (12), it suffices to minimize $\mathbb{E}Z$. We prove this claim via an exchange argument: as long as the coupling involves pairs $(x_1, y_1) = (X(u_1), Y(v_1))$ and $(x_2, y_2) = (X(u_2), Y(v_2))$ with $x_1 < x_2$ and $y_1 < y_2$, swapping them (pairing x_1 with y_2 and x_2 with y_1) does not increase $\mathbb{E}Z$.

This exchange argument, and consequently the entirety of Lemma 5.3, follows directly from showing submodularity of the associated functions, which we establish now. We define the relevant notions first.

Definition 5.4 (Submodular and Supermodular Functions)**.** A function $f : \mathbb{R}^2 \to \mathbb{R}$ is **submodular** if for all $x_1 \leqslant x_2, y_1 \leqslant y_2$,

$$f(x_1, y_1) + f(x_2, y_2) \leqslant f(x_1, y_2) + f(x_2, y_1).$$

Equivalently, f has decreasing differences in (x, y): for every $x_1 < x_2$, the increment $\Delta_x f(y) = f(x_2, y) - f(x_1, y)$ is nonincreasing in y. Analogously, f is **supermodular** if the inequality holds with $\geqslant$.

Lemma 5.5. *Fix $A, B, C \in \mathbb{R}$. The function $H(x, y) = \max\{A + x, B + y, C + x + y\}$ is submodular.*

Proof. The graph of H is the upper envelope of three planes, $z = A + x, z = B + y$, and $z = C + y + x$. Partition the (x, y)-plane into regions where one of these planes is on top. Then, the boundaries are defined by a horizontal line $A + x = C + x + y$, a vertical line $B + y = C + x + y$, and a diagonal line $A + x = B + y$. Observe that along any vertical line with fixed x, $H(x, \cdot)$ as a function of y has the following shape:

- Below the boundary $y_0(x) = \min\{A - C, x + (A - B)\}$, the top plane is $z = A + x$, which has slope 0 in the y-direction; and

- Above $y_0(x)$, the top plane is either $z = B + y$ or $z = C + x + y$, both having slope 1 in y.

Consequently, on each vertical line, H is flat in y up to a threshold $y_0(x)$; from there, it increases in y with slope 1. Crucially, the threshold $y_0(x) = \min\{A - C, x + (A - B)\}$ is also nondecreasing in x.

Now fix $y_1 \leqslant y_2$. From the geometric observation above, for x,

$$\Delta_y H(x) = H(x, y_2) - H(x, y_1) = \begin{cases} 0 & y_2 \leqslant y_0(x) \\ y_2 - y_1 & y_1 \geqslant y_0(x) \\ y_2 - y_0(x) & y_1 < y_0(x) < y_2. \end{cases}$$

Because $y_0(x)$ is nondecreasing in x, the function $x \mapsto \Delta_y H(x)$ is nonincreasing: when we slide the vertical line to the right, the threshold $y_0(x)$ can only move up, shrinking the portion of $[y_1, y_2]$ above it. Now take $x_1 \leqslant x_2$. The preceding monotonicity gives

$$H(x_2, y_2) - H(x_2, y_1) = \Delta_y H(x_2) \leqslant \Delta_y H(x_1) = H(x_1, y_2) - H(x_1, y_1).$$

This completes the proof. □

Proof of Lemma 5.3. The proof consists of two steps. First, given a *frozen* baseline $c = \|X + Y\|_\infty$, along with $\|X\|_\infty$ and $\|Y\|_\infty$ which are fixed by $\mathcal{D}_X$ and $\mathcal{D}_Y$, a local counter-monotone swap never increases $\mathbb{E}Z$. Indeed, by applying Lemma 5.5, the mapping

$$(x, -y) \mapsto h_c(x, -y) = \max\{\|X\|_\infty + x, \|Y\|_\infty + (-y), c + x + (-y)\},$$

as a function of x and $-y$ is submodular. Flipping the sign of the second argument, we see that $(x, y) \mapsto h_c(x, -y)$ as a mapping of x and y is *super*modular. Therefore, when $x_1 < x_2$ and $y_1 < y_2$ (and hence $(-y_1) > (-y_2)$), we have

$$h_c(x_1, y_1) + h_c(x_2, y_2) \geqslant h_c(x_1, y_2) + h_c(x_2, y_1), \tag{13}$$

so a local counter-monotone swap weakly decreases $\mathbb{E}[h_c(X, Y)]$, conditioned on $\|X + Y\|_\infty$ being fixed.

Second, we claim that local counter-monotone swaps do not worsen (increase) $\|X + Y\|_\infty$. To see this, suppose $x_1 < x_2$ and $y_1 < y_2$. Then we have

$$x_1 + y_1 \leqslant \min(x_1 + y_2, x_2 + y_1) \leqslant \max(x_1 + y_2, x_2 + y_1) \leqslant x_2 + y_2.$$

The maximum absolute value of these four terms will therefore be attained at one of the extremes, which implies

$$\max\{|x_1 + y_1|, |x_2 + y_2|\} \geqslant \max\{|x_1 + y_2|, |x_2 + y_1|\}.$$

In particular, this means the swap does not increase $\|X + Y\|_\infty$. As $h_c(x, -y)$ is nondecreasing in c, this proves that local swaps indeed always help. We therefore iteratively perform local swaps until no swap is available, which happens precisely when the resulting coupling is counter-monotone. Since the objective did not increase in this process, the proof is complete. □

Remark 5.6 (Mapping V onto $[0, 1]$). A useful alternate perspective, which we will frequently use later, is to view X, Y as real-valued functions on $[0, 1]$.

Recall V is normalized into unit mass so it naturally maps to $[0, 1]$. Using Lemma 5.3, we consider a mapping $V \mapsto [0, 1]$ such that X (resp. Y) can be viewed as a decreasing (resp. increasing) function on $[0, 1]$: for instance, we map voters with largest $X(v)$ (and most negative $Y(v)$) to near 0 and map voters with most negative $X(v)$ (and largest $Y(v)$) to near 1. Figure 1 shows one hypothetical example of X and Y. As we assumed that the metric space is finite, X and Y will be piecewise constant step functions.

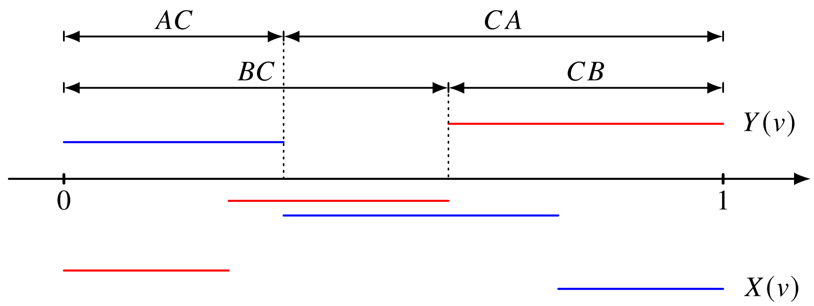


Figure 1: Counter-monotonic coupling of X (in blue) and Y (in red). Note the graphs partition $[0, 1]$ into AC/CA (by X) and BC/CB (by Y).

5.4 Tight f Constraints and the Optimal Matchings

A fixed instance I may admit various matchings for the candidate pair (A, C) and thus potentially different values for W_{AC}, the number of matchings where the outcome is $A \succ C$. Consequently, the values of $f(AC)$ need not be unique; neither for $f(CB)$. However, we note that for fixed (X, Y) and distribution ρ over voters, the constraints for $f(AC)$ and $f(CB)$ in our mathematical program are made most slack by choosing the matchings with the most number of AC wins for X (resp. CB wins for Y). Call them the **A-optimal (A, C) matching** and the **C-optimal (C, B) matching**, respectively.

We now show two properties of the optimal solution. We first prove that a specific type of optimal matchings pairs "prefixes" (most polar voters) of one side with the "suffixes" (most indifferent, i.e., least polar) of the other. Next, we prove a "continuity" result: that it suffices to tighten the inequalities $f(AC) \geqslant 1 - \lambda$ and $f(CB) \geqslant \lambda$ into equalities. These results pave way to a clean structural reduction (Equation (14)) that leads us to Section 5.5.

5.4.1 Property 1: Prefix Property of Matchings

Fix the instance I as well as some matching for candidates (A, C). Consider two $A \succ C$ pairs (u_1, v_1) and (u_2, v_2). Suppose $X(u_1) \geqslant X(u_2) \geqslant 0 \geqslant X(v_1) \geqslant X(v_2)$. Then we have $X(u_1) + X(v_1) \geqslant 0$ and $X(u_2) + X(v_2) \geqslant 0$. It is easy to check that $X(u_1) + X(v_2) \geqslant 0$ and $X(u_2) + X(v_1) \geqslant 0$. This means we can replace the matchings with (u_1, v_2) and (u_2, v_1). This means the matchings can be made *counter-monotone*. Further, suppose $X(u_1) \geqslant X(u_2) \geqslant 0$ and u_1 does not participate in an $A \succ C$ pair, while u_2 is matched to v_2 in an $A \succ C$ pair. Then we can replace (u_2, v_2) with (u_1, v_2).

Analogously, if $0 \geqslant X(v_1) \geqslant X(v_2)$ and v_1 does not participate in an $A \succ C$ pair while v_2 is matched to u_2 in an $A \succ C$ pair, we can replace (u_2, v_2) with (u_2, v_1). This is feasible for the f constraint since W_{CA} cannot increase in this process, and W_{AC} is preserved.

Iterating this process, we obtain a new (A, C) matching that also has W_{AC} pairs satisfying $A \succ C$. Additionally, in this new (A, C) matching, these pairs come from pairing the W_{AC} mass of highest $X(u) \geqslant 0$ (the **prefix** of AC) with the W_{AC} mass with highest $X(v) < 0$ (the **suffix** of CA) counter-monotonically, meaning that between these two blocks of mass W_{AC}, the highest positive $X(u)$ is matched to the lowest $X(v) < 0$, and so on. This is shown in Figure 2.

An identical result can be shown for (C, B), so that any $C \succ B$ matchings of mass W_{CB} can be assumed to satisfy the prefix property. In particular, by starting with an optimal matching and repeating the procedure above, we arrive at the following conclusion.

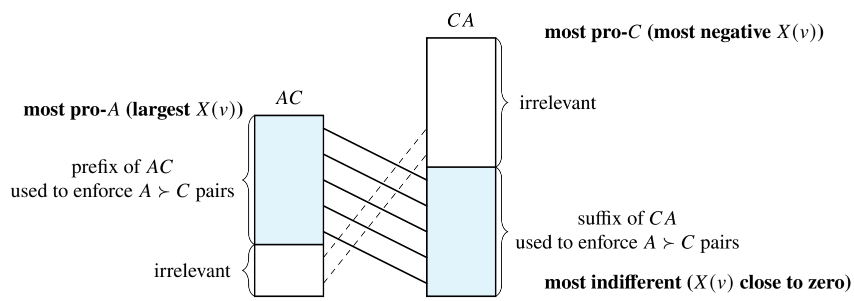


Figure 2: The prefix-suffix structure of $A \succ C$ pairs.

Lemma 5.7 (Prefix Property of Optimal Matchings)**.** *Fix an instance I and an A-optimal matching for candidates (A, C). Suppose this matching admits a mass W_{AC} of $A \succ C$ pairs. Then, there exists another A-optimal (A, C) matching (hence also with mass W_{AC} of $A \succ C$ pairs) such that:*

(i) It takes place between the W_{AC} masses of highest $X(u) \geqslant 0$ and highest $X(v) < 0$; and

(ii) It couples the two blocks counter-monotonically: highest $X(u)$ with lowest $X(v)$, and so on.

An equivalent version holds for the (C, B) matching, where if W_{CB} is the largest admissible mass of $C \succ B$ pairs, then they can be assumed to be coupled counter-monotonically between the W_{CB} mass with highest $Y(u) \geqslant 0$ and the W_{CB} mass with highest $Y(v) < 0$.

Recall from Remark 5.6 that we may view X, Y as monotonic functions on $[0, 1]$. The previous observations yield a four-interval decomposition: two blocks of size W_{AC}, one for each of AC, CA, and two complementary blocks. A similar decomposition follows for (C, B). We focus only on the cases where $|AC| \leqslant |CA|$ and $|BC| \leqslant |CB|$, but the other cases also admit a similar partition, the major difference being the location of the unmatched voters.

Lemma 5.8 (Four-interval partition of $[0, 1]$ by X and Y)**.** *When $|AC| \leqslant |CA|$, the range $[0, 1]$ can be partitioned into four consecutive, possibly empty intervals that describe the (AC, CA) matching, as shown in Table 1. By the structure of the A-optimal matching, the A-win blocks always lie on the leftmost of the $[0, |AC|]$ (AC block) and $[|AC|, 1]$ (CA block). Similarly, when $|BC| \leqslant |CB|$, we can perform the same partition based on Y which describes the (BC, CB) matching. Since we analyze the C-optimal matching and Y is increasing, the C-win blocks lie on the* rightmost *of the $[0, |BC|]$ and $[|BC|, 1]$ blocks.*

X	Interval	$[0, W_{AC}]$	$[W_{AC}, \lvert AC\rvert]$	$[\lvert AC\rvert, \lvert AC\rvert + W_{AC}]$	$[\lvert AC\rvert + W_{AC}, 1]$
	Role	AC A-win	AC A-loss	CA A-win	CA A-loss/unmatched
	Length	W_{AC}	W_{CA}	W_{AC}	$W_{CA} + (1 - 2\lvert AC\rvert)$
Y	Interval	$[0, W_{BC}]$	$[W_{BC}, \lvert BC\rvert]$	$[\lvert BC\rvert, 1 - \lvert BC\rvert + W_{BC}]$	$[1 - \lvert BC\rvert + W_{BC}, 1]$
	Role	BC C-loss	BC C-win	CB unmatched/C-loss	CB C-win
	Length	W_{BC}	W_{CB}	$(1 - 2\lvert BC\rvert) + W_{BC}$	W_{CB}

Table 1: Two different partitions of $[0, 1]$ induced by X and Y.

5.4.2 Property 2: Tightness of $f(AC) \geqslant 1-\lambda$ and $f(CB) \geqslant \lambda$

Since the constraints for $f(AC)$ and $f(CB)$ in our mathematical program are made most slack by choosing the A-optimal (A, C) matching and C-optimal (C, B) matchings, we will now restrict our f values to correspond to the optimal matchings. Furthermore, we define M_{AC} and M_{CB} to be the optimal (A, C) and (C, B) matchings which satisfy the prefix property in Lemma 5.7. Similarly W_{AC} and W_{CB} correspond to the number of wins in the optimal matching.

We now show that without loss of generality, we can assume that both the $f(AC) \geqslant 1-\lambda$ and $f(CB) \geqslant \lambda$ constraints are tight.

Lemma 5.9 (Tightness of $f(AC)$-constraint)**.** *Let I be an instance under which $f(AC) \geqslant 1-\lambda$ and $f(CB) \geqslant \lambda$. Then there exists an instance I' such that $f(AC') = 1-\lambda$, $f(CB') = f(CB)$, and $\Phi_R(X', Y') \leqslant \Phi_R(X, Y)$. Consequently, we may assume without loss of generality that $f(AC) = 1-\lambda$ in Program (12).*

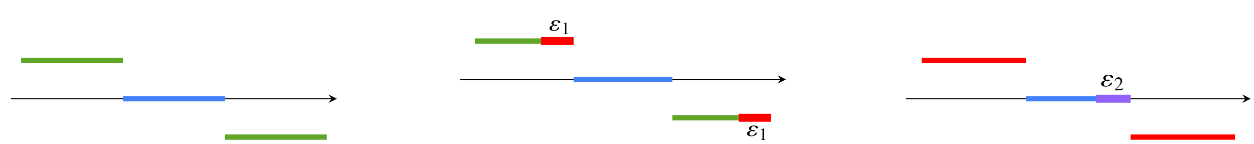
Proof. For $t \geqslant 0$, we define a parameterized instance I_t with variables $X_t(v) = X(v) - t$ and keep $Y_t = Y$ unchanged for all t. Since $f(CB)$ is determined by Y, $f(CB)$ also remains unchanged. Define $AC(t)$, $CA(t)$, $M_{AC}(t)$, and $W_{AC}(t)$ parametrically to be the values of AC, CA, M_{AC}, and W_{AC} induced by $X_t(v)$. As $X(v)$ decreases by t, $\|X\|_\infty$ and $\|X+Y\|_\infty$ increase by no more than t, so overall,

$$\|X_t\|_\infty + X_t(v) \leqslant \|X\|_\infty + t + X(v) - t = \|X\|_\infty + X(v),$$

and similarly $\|X_t+Y\|_\infty + X_t(v) - Y_t(v) \leqslant \|X+Y\|_\infty + X(v) - Y(v)$. Therefore, $Z_{\min}(X_t, Y_t) \leqslant Z_{\min}(X, Y)$. Consequently, $\Phi_R(X_t, Y_t) = \mathbb{E}X_t + (R+1) \cdot \mathbb{E}Y_t + R \cdot \mathbb{E}Z_{\min}(X_t, Y_t) \leqslant \Phi_R(X, Y)$ for all t.

It remains now to find a t^* that attains the equality $f(AC(t^*)) = 1-\lambda$. Observe that at $t = 0$, we have $f(AC(0)) \geqslant 1-\lambda$ by assumption; on the other hand, if $t > \|X\|_\infty$ then $X_t < 0$ everywhere, making every voter prefer C over A, at which point $f(AC(t)) = 0$. We argue that if ties are distributed appropriately as t increases, then $f(AC(t))$ is a continuous function. Since $f(AC(0)) \geqslant 1-\lambda$ and $f(AC(t)) = 0$ for $t > \|X\|_\infty$, this would imply that there is some value of t^* where $f(AC(t^*)) = 1-\lambda$.

Note by Equation (6) that a discontinuity of $f(AC(t))$ is either caused by a discontinuity in $|AC(t)|$ or $W_{AC}(t)$. A discontinuity in $|AC(t)|$ happens precisely when $S_t = \{v \mid X_t(v) = 0\}$ has nonzero mass. A discontinuity in $W_{AC}(t)$ can occur when either S_t has nonzero mass, or $P_t = \{(u,v) \in M_{AC}(t) \mid X_t(u) + X_t(v) = 0\}$ has nonzero mass. Fix an arbitrary t. Since X is a piecewise step function, let $t' > t$ be the earliest time step after t at which either $S_{t'}$ or $P_{t'}$ has nonzero mass. We first handle the case where $P_{t'}$ has nonzero mass. Since $t' > t$, the deliberation for every pair in $P_{t'}$ initially prefers A. Now arbitrarily select a subset of pairs $P' \subseteq P_{t'}$ with mass $\varepsilon_1 > 0$. We set the deliberation for every pair in P' to prefer C. Clearly $W_{AC}(t')$ decreases by at most ε_1, implying that $W_{AC}(t')$, and hence $f(AC(t'))$, changes continuously.



(a) Original: green segments ($P_{t'}$) are $A \succ C$ pairs; blue segment ($S_{t'}$) is AC.

(b) Handling $P_{t'}$. Carve out P' (red) of mass ε_1 (greens) and attribute them to $C \succ A$ pairs.

(c) Handling $S_{t'}$. Carve out S' of mass ε_2 (purple) and set it to CA. Note $W_{AC}(t')$ may change.

Figure 3: A visualization of continuous tie-handling on X. Left (a): At time step t', both $P_{t'}$ and $S_{t'}$ have positive mass. Middle (b): We handle $P_{t'}$ by continuously allocating increasing subsets $P' \subseteq P_{t'}$ to $C \succ A$ pairs. Right (c): We then handle $S_{t'}$ by continuously allocating increasing subsets $S' \subseteq S_{t'}$ to CA and argue that the change in $W_{AC}(t')$ is also continuous.

We now handle the case where $S_{t'}$ has nonzero mass. Since $t' > t$, every voter in $S_{t'}$ is initially in AC. We arbitrarily choose a subset $S' \subseteq S_{t'}$ with mass $\varepsilon_2 > 0$ and set the ties so that every voter in S' is in CA.

Clearly $|AC|$ decreases by exactly ε_2. To bound the change in $W_{AC}(t')$, first note that because $S_{t'}$ is initially the unique set on which $X_t = 0$, no $(A \succ C)$ pairs can take place on it. Consequently, our operation cannot decrease $W_{AC}(t')$. On the other hand, if $W_{AC}(t')$ increased by more than ε_2, then we could remove the set S' and achieve a matching for the original instance with a higher value of $W_{AC}(t')$, which contradicts the original optimality of $W_{AC}(t')$.

Thus $|AC(t')|$ and $W_{AC}(t')$ are both continuous in the change, so by Equation (6), $f(AC(t'))$ changes continuously for this step as well. This implies there is some value of t^* where $f(AC(t^*)) = 1 - \lambda$. $\square$

Lemma 5.10 (Tightness of $f(CB)$-constraint)**.** *If I is such that $f(AC) \geqslant 1 - \lambda$ and $f(CB) \geqslant \lambda$, then there exists an instance I' with $f(AC') = f(AC)$, $f(CB') = \lambda$, and $\Phi_R(X', Y') \leqslant \Phi_R(X, Y)$. Consequently, we may assume without loss of generality that $f(CB) = \lambda$ in Program (12).*

Proof. For $t \geqslant 0$, we keep $X_t = X$ unchanged for all t and let $Y_t(v) = Y(v) - t$. Then $\|Y\|_\infty$ and $\|X + Y\|_\infty$ increase by no more than t as $Y(v)$ decreases by t, so

$$\|Y_t\|_\infty - Y_t(v) \leqslant \|Y\|_\infty + t - (Y(v) - t) = \|Y\|_\infty - Y(v) + 2t,$$

and similarly $\|X_t + Y_t\|_\infty + X_t(v) - Y_t(v) \leqslant \|X + Y\|_\infty + X(v) - Y(v) + 2t$. By Equation (10) these imply $Z_{\min}(X_t, Y_t) \leqslant Z_{\min}(X, Y) + t$. Then $\Phi_R(X_t, Y_t) \leqslant \Phi_R(X, Y)$, as the coefficient of $\mathbb{E}Y$ is $R + 1$, greater than that of $\mathbb{E}Z$. The rest of the proof mirrors that of Lemma 5.9 by noting that when $t > \|Y\|_\infty$, we have $f(CB(t)) = 0$ because $Y_t < 0$ everywhere. $\square$

With the two properties of optimal matchings in place, we now restate Program (12) in its updated form.

$$\begin{array}{rl}
\text{Minimize} & \Phi_R(X,Y) = \mathbb{E}X + (R+1)\cdot\mathbb{E}Y + R\cdot\mathbb{E}Z \\
\text{over} & X, Y \text{ on } V, \quad Z = Z_{\min}(X,Y) \;\text{ from Equation (10)} \\
& \\
\text{Subject to} & f(AC) \text{ is induced by an } A\text{-optimal } (AC, CA) \text{ matching;} \\
& f(CB) \text{ is induced by a } C\text{-optimal } (CB, BC) \text{ matching;} \\
& f(AC) = 1 - \lambda, \quad f(CB) = \lambda.
\end{array} \tag{14}$$

We will next show that this program can be made to have a constant number of variables, and it is a bilinear program with two disjoint sets of variables and linear constraints over these.

5.5 Bilinear Program and the Distortion of 3

For the remainder of the section, we fix $\lambda^* = (3 - \sqrt{3})/2 \approx 0.634$ and $w^* = \sqrt{3} - 1 \approx 0.732$, and show that the deliberation-via-matching protocol under these parameters has distortion at most 3. This setting is optimal by the lower bound construction in Section 6. Further, assuming parameters reduces the number and complexity of the cases we need to consider below.

We first show that this setting of parameters, combined with the tightness of the f constraints implies the sizes of $|AC|$ and $|CB|$ satisfy some nice properties. This leads to a collection of instances that we then simplify using the convexity of the variable Z and the max norms into bilinear programs with a constant number of variables, which we can easily solve via vertex enumeration.

5.5.1 Bounding Sizes of $|AC|$ and $|CB|$

We begin by characterizing the range of possible sizes of $|AC|$ and $|CB|$ when $f(AC) = 1 - \lambda^*$ and $f(CB) = \lambda^*$.

Lemma 5.11. *When $f(AC) = 1 - \lambda^*$, we have $0.25 \leqslant |AC| \leqslant 0.50$. Similarly, when $f(CB) = \lambda^*$, we have $0.50 \leqslant |CB| \leqslant 0.75$. In particular, for instances where $f(AC) = 1 - \lambda^*$ and $f(CB) = \lambda^*$, we always have $|AC| \leqslant |CA|$ and $|BC| \leqslant |CB|$.*

Proof. Recall that since $W_{AC} + W_{CA} = \min(|AC|, 1 - |AC|)$, and $|AC| + |CA| = 1$, we have

$$\text{score}(AC) = \frac{|AC| + w^* \cdot W_{AC}}{n} \qquad \text{and} \qquad \text{score}(AC) + \text{score}(CA) = \frac{1 + w^* \cdot \min(|AC|, 1 - |AC|)}{n}.$$

Since $W_{AC} \leqslant \min(|AC|, 1 - |AC|)$, this means

$$1 - \lambda^* = f(AC) = \frac{\text{score}(AC)}{\text{score}(AC) + \text{score}(CA)} \leqslant \frac{|AC| + w^* \cdot \min(|AC|, 1 - |AC|)}{1 + w^* \cdot \min(|AC|, 1 - |AC|)}.$$

The RHS is strictly increasing as a function of $|AC|$ when $0 \leqslant |AC| \leqslant 1$. When setting $|AC| = 0.25$, we can verify that the RHS is $1 - \lambda^*$, implying that 0.25 is the smallest possible value of $|AC|$ to achieve $f(AC) = 1 - \lambda^*$.

Similarly, we have $W_{AC} \geqslant 0$, so that

$$1 - \lambda^* = f(AC) \geqslant \frac{|AC|}{1 + w^* \cdot \min(|AC|, 1 - |AC|)}.$$

Again, the RHS is strictly increasing as a function of $|AC|$ when $0 \leqslant |AC| \leqslant 1$. Setting $|AC| = 0.5$, we have the RHS is $1 - \lambda^*$, implying that 0.5 is the largest possible value of $|AC|$.

An identical argument for $|CB|$ shows that $0.5 \leqslant |CB| \leqslant 0.75$. □

Lemma 5.12. *When $f(AC) = 1 - \lambda^*$, we have $|AC| + W_{AC} = 0.5$. Similarly, when $f(CB) = \lambda^*$, we have $|BC| + W_{BC} = 0.5$.*

Proof. By Lemma 5.11, we have $|AC| \leqslant |CA|$, so

$$1 - \lambda^* = f(AC) = \frac{|AC| + w^* \cdot W_{AC}}{1 + w^* \cdot |AC|}.$$

Solving for $|AC|$ in terms of W_{AC}, we have $|AC| + W_{AC} = 0.5$ as desired. For $f(CB)$, we have by Lemma 5.11 that $|CB| \geqslant |BC|$, so

$$\lambda^* = f(CB) = \frac{|CB| + w^* \cdot W_{CB}}{1 + w^* \cdot (1 - |CB|)}.$$

Solving for W_{CB} in terms of $|CB|$, we have $W_{CB} = 1.5 - 2|CB|$. Substituting $|CB| = 1 - |BC|$ and $W_{CB} = |BC| - W_{BC}$ gives $|BC| + W_{BC} = 0.5$. □

We now consider the partitions induced on the number line as given in Table 1. Our goal will be to write a bilinear programming relaxation of Program (14), where we have variables for each interval in the partition. There are two cases based on the sizes of $|AC|$ and $|BC|$.

5.5.2 Case 1: $|AC| \leqslant |BC|$

We first consider the case where $|AC| \leqslant |BC|$. We know from Lemma 5.11 that $|AC| \leqslant |CA|$ and $|BC| \leqslant |CB|$, so the partitions induced by X and Y are shown in Table 1. As shown in Figure 4, we partition the range into 9 intervals labeled 1 through 9. The top row of the figure depicts how the 9 intervals relate to the partition induced by X, while the bottom row depicts how the same 9 intervals relate to the partition induced by Y. For the X partition, intervals 1 and 2 correspond to $[0, W_{AC}]$, interval 3 corresponds to $[W_{AC}, |AC|]$, intervals 4 and 5 correspond to $[|AC|, |AC| + W_{AC}]$, and intervals 6 through 9 correspond to

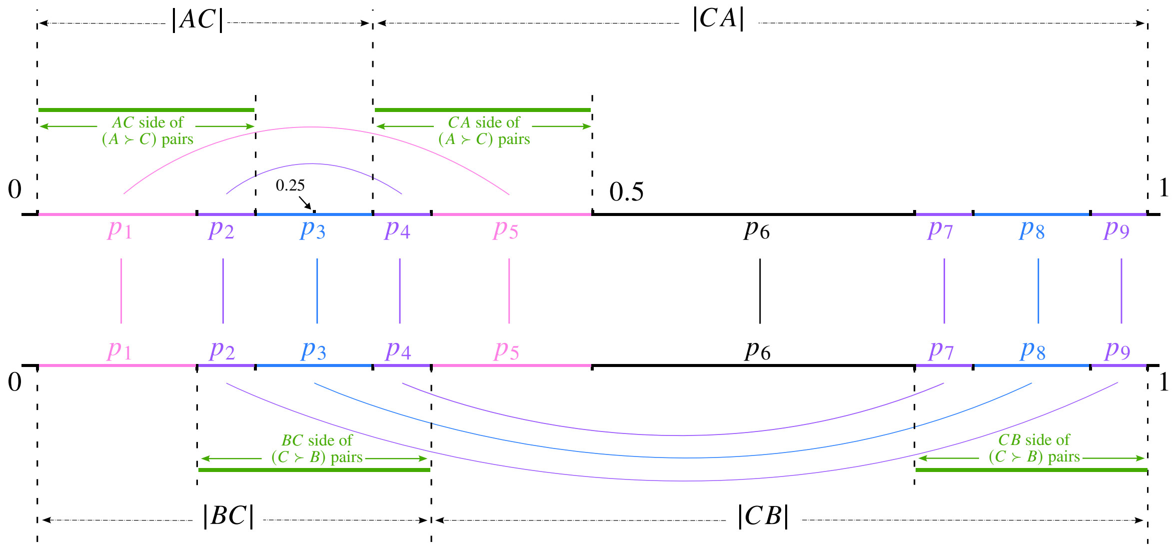


Figure 4: The lines go from 0 to 1, capturing cumulative voter mass. The top line represents X values in decreasing order and the bottom line represents Y values in increasing order. A voter appears at the same position in both lines. The pink masses p_1 and p_5 represent a set of $A \succ C$ matching pairs. This means $p_1 = p_5$. The same holds for the purple masses p_2 and p_4. The masses p_1, p_2, p_3 correspond to non-negative X values, hence together capture $|AC|$. The masses $p_1, \ldots, p_4$ have non-positive Y values and together capture $|BC|$. The $C \succ B$ pairs are captured by the pairs of masses (p_2, p_9), (p_3, p_8) and (p_4, p_7).

$[|AC| + W_{AC}, 1]$. For the Y partition, interval 1 corresponds to $[0, W_{BC}]$, intervals 2 through 4 correspond to $[W_{BC}, |BC|]$, intervals 5 and 6 correspond to $[|BC|, 1 - |BC| + W_{BC}]$, and intervals 7 through 9 correspond to $[1 - |BC| + W_{BC}, 1]$. The interpretations of each segment of the X and Y partitions are given in Table 1.

We now show that it suffices to consider instances where for each interval, the values of X and Y are constant across it. In particular, we can replace the voters in an interval with a weighted voter whose X (resp. Y) value is equal to the average of the X (resp. Y) values of the voters in that interval, and the weight being the probability mass of voters in that interval. This will follow from the lemma below.

Lemma 5.13 (Compacting an Interval). *For any two voters u, v with values $(X(u), Y(u))$ and $(X(v), Y(v))$, let $\mu = (X(u) + X(v))/2$ and $\nu = (Y(u) + Y(v))/2$. Then if we replace u and v with two voters with identical values (μ, ν), the objective in Equation (9) does not increase.*

Proof. First, note that

$$\begin{aligned}\max\left(X(u) + Y(u), X(v) + Y(v)\right) &\geqslant \frac{X(u) + Y(u) + X(v) + Y(v)}{2} \\ &= \mu + \nu \geqslant \min\left(X(u) + Y(u), X(v) + Y(v)\right),\end{aligned}$$

which means $\|X + Y\|_\infty$ cannot increase. A similar argument shows that $\|X\|_\infty, \|Y\|_\infty$ cannot increase. Next, fixing the values of these norms, $Z_{\min}$ in Equation (10) is the maximum of three linear functions, and is therefore a convex function of X, Y. By Jensen's inequality, this means $\mathbb{E}Z$ cannot increase when we replace values by their means. Finally, $\mathbb{E}X, \mathbb{E}Y$ are preserved by this transformation. □

Using this lemma in each interval, we can replace the voters in each interval with a weighted voter with X, Y values equal to the means of X, Y values of voters in the interval. This does not increase the objective in Equation (9). We now show that the f constraints are preserved. Consider for example the intervals $(1, 5)$

that define a set of $A \succ C$ matchings, each with non-negative sum of X values. Simple averaging over the pairs of matched voters shows that the sum of the mean values of X in the two intervals is non-negative, so that the new weighted voters also define an $A \succ C$ matching. Further, voters who initially preferred A to C map to a weighted voter with the same preference. This shows the f constraints are preserved in this process.

For each interval $i \in [1, 9]$ let X_i and Y_i denote the uniform value of X and Y respectively on that interval. Also let p_i denote the length of interval i. We first show that $p_2 = p_4$. By Lemma 5.12, we have $W_{AC} + |AC| = 0.5$, so the midpoint of interval 3 must be at 0.25. Similarly, we have $W_{BC} + |BC| = 0.5$ so the midpoint of intervals 2, 3, and 4 collectively must also be at 0.25. Together, this implies $p_2 = p_4$. Since intervals 2 through 4 collectively are centered around 0.25, and interval 5 ends at $W_{AC} + |AC| = 0.5$, we also have $p_1 = p_5$. Finally, we define intervals 7 through 9 to be the intervals that match with intervals 2 through 4 in the (B, C) matching. Thus we have $p_2 = p_9$, $p_3 = p_8$, and $p_4 = p_7$. In total we have the constraints $p_1 = p_5$, $p_2 = p_4 = p_7 = p_9$, and $p_3 = p_8$.

We now describe the constraints on X and Y induced by the matching constraints. Recall that intervals 1 and 2 correspond to the section of AC where A wins the deliberation, and intervals 4 and 5 correspond to the section of CA where A wins the deliberation. The A-optimal matching pairs interval 1 with interval 5 and interval 2 with interval 4. Thus we have the constraints $X_1 + X_5 \geqslant 0$ and $X_2 + X_4 \geqslant 0$. Similarly for Y, we have the constraints $Y_2 + Y_9 \geqslant 0$, $Y_3 + Y_8 \geqslant 0$, and $Y_4 + Y_7 \geqslant 0$. We note that our relaxation will not need to enforce the constraints that $X(u) + X(v) \leqslant 0$ for a deliberation where C wins against A (or the corresponding constraint for Y).

By the counter-monotonic coupling of X and Y, we have $X_i \geqslant X_{i+1}$ and $Y_i \leqslant Y_{i+1}$ for all $i \in [8]$. Finally, since the section AC corresponds to positive X values and the section CB corresponds to positive Y values, we have $X_3 \geqslant 0$ and $Y_5 \geqslant 0$. We have the following relaxation of Program (14):

$$
\begin{array}{lll}
\min & \mathbb{E}X + (R+1) \cdot \mathbb{E}Y + R \cdot \mathbb{E}Z\,, & (15) \\
\text{s.t.} & \mathbb{E}X = \sum_{i=1}^{9} p_i \cdot X_i, \qquad \mathbb{E}Y = \sum_{i=1}^{9} p_i \cdot Y_i \qquad \text{and} \qquad \mathbb{E}Z = \sum_{i=1}^{9} p_i \cdot Z_i & \\
& Z_i \geqslant Z_{\min}(X_i, Y_i) \qquad \forall i \in [9] & \text{(Set of linear constraints)} \\
& X_i \geqslant X_{i+1} \qquad \text{and} \qquad Y_i \leqslant Y_{i+1} \qquad \forall\, i \in [8] & \text{(Counter-monotonicity)} \\
& X_3 \geqslant 0 \qquad \text{and} \qquad Y_5 \geqslant 0 & \\
& X_1 + X_5 \geqslant 0 \qquad \text{and} \qquad X_2 + X_4 \geqslant 0 & (A \succ C \text{ matchings in } X) \\
& Y_2 + Y_9 \geqslant 0, \qquad Y_3 + Y_8 \geqslant 0 \qquad \text{and} \qquad Y_4 + Y_7 \geqslant 0 & (C \succ B \text{ matchings in } Y) \\
& \sum_{i=1}^{9} p_i = 1 \qquad \text{and} \qquad \sum_{i=1}^{5} p_i = 0.5 & (|AC| + W_{AC} = 0.5) \\
& p_1 = p_5, \qquad p_2 = p_4 = p_7 = p_9 \qquad \text{and} \qquad p_3 = p_8 & \text{(Coupling of masses)} \\
& Z_i, p_i \geqslant 0\,,\ \forall\, i \in [9]\,. &
\end{array}
$$

We note that the constraint $Z_i \geqslant Z_{\min}(X_i, Y_i)$ corresponds to a set of linear inequalities by Equation (10), which we write out explicitly in Appendix A. Since $\|X\|_\infty + X(v) \geqslant 0$, we must have $Z_i \geqslant 0$. We include this constraint explicitly in the relaxation to aid in our analysis.

Since the above program contains a multiplicative term when computing the expectation of each variable, it is a bilinear program, where the objective multiplies the p_i variables with the (X_i, Y_i, Z_i) variables, and there are separate linear constraints for the p_i and the (X_i, Y_i, Z_i). In order to solve this program efficiently, we separate the constraints into two parts, where the first one has variables for each p_i and the second one has the remaining variables. If we absorb the $\mathbb{E}X, \mathbb{E}Y, \mathbb{E}Z$ constraints into the objective, the two sets of constraints are disjoint. Since for any fixed (X_i, Y_i, Z_i) variables, the bilinear program is linear in the p_i

variables, its optimum is achieved at a vertex of the polytope of the p_i. This means the overall optimum is also achieved at such a point, and it therefore suffices to enumerate all extreme points of the first set of constraints (the ones capturing p_i) and solve the bilinear program at every such extreme point.

Isolating the p_i variables, and grouping the equal terms, we have a polytope defined by the following constraints:

$$
\begin{aligned}
&2p_1 + 4p_2 + 2p_3 + p_6 = 1, \\
&2p_1 + 2p_2 + p_3 = 0.5, \\
&p_1, p_2, p_3, p_6 \geqslant 0.
\end{aligned}
$$

Eliminating p_3, p_6, the above reduces to the interior of a triangle on (p_1, p_2) with vertices given by the point set $\{(0,0), (0,0.25), (0.25,0)\}$. Therefore, the 3 extreme points of the polytope are given by

$$(p_1, p_2, p_3, p_6) = \{(0,0,0.5,0), (0,0.25,0,0), (0.25,0,0,0.5)\}.$$

For each of the 3 extreme points, we substitute the p_i variables into Program (15) and solve the resulting LP. For $R = 2$, the optimal objective value at each such extreme point is exactly 0, implying that the maximum distortion is at most 3. We present the verifiable dual certificates in Appendix A.

5.5.3 Case 2: $|AC| > |BC|$

This case uses the same ideas as the previous one. We again have $|AC| \leqslant |CA|$ and $|BC| \leqslant |CB|$, so the partitions induced by X and Y are the same as before. We show the (A, C) and (C, B) matchings and the corresponding set of intervals in Figure 5. For the X partition, interval 1 corresponds to $[0, W_{AC}]$, intervals 2 through 4 correspond to $[W_{AC}, |AC|]$, interval 5 corresponds to $[|AC|, |AC| + W_{AC}]$, and intervals 6 and 7 correspond to $[|AC| + W_{AC}, 1]$. For the Y partition, intervals 1 and 2 correspond to the segment $[0, W_{BC}]$, interval 3 corresponds to $[W_{BC}, |BC|]$, intervals 4 through 6 correspond to $[|BC|, 1 - |BC| + W_{BC}]$, and interval 7 corresponds to $[1 - |BC| + W_{BC}, 1]$. The interpretations of each segment of the X and Y partitions are given in Table 1.

Define X_i, Y_i, and p_i as in the previous case. We first show that $p_2 = p_4$. By Lemma 5.12, we have $W_{BC} + |BC| = 0.5$, so the midpoint of interval 3 must be at 0.25. Similarly, we have $W_{AC} + |AC| = 0.5$ so the midpoint of intervals 2, 3, and 4 collectively must also be at 0.25. This implies $p_2 = p_4$. Since intervals 2 through 4 collectively are centered around 0.25, and interval 5 ends at $|AC| + W_{AC} = 0.5$, we also have $p_1 = p_5$. Finally, we define interval 7 to be the interval that matches with interval 3 in the (B, C) matching. Thus $p_3 = p_7$.

For the matching constraints, interval 1 is mapped with interval 5 in the (A, C) matching and interval 3 is mapped with interval 7 in the (B, C) matching, so we have the constraints $X_1 + X_5 \geqslant 0$ and $Y_3 + Y_7 \geqslant 0$. By the counter-monotonic coupling of X and Y, we have $X_i \geqslant X_{i+1}$ and $Y_i \leqslant Y_{i+1}$ for all $i \in [6]$. Finally, since the section AC corresponds to positive X values, and the section CB corresponds to positive Y values, we have $X_4 \geqslant 0$ and $Y_4 \geqslant 0$. We have the following bilinear programming relaxation:

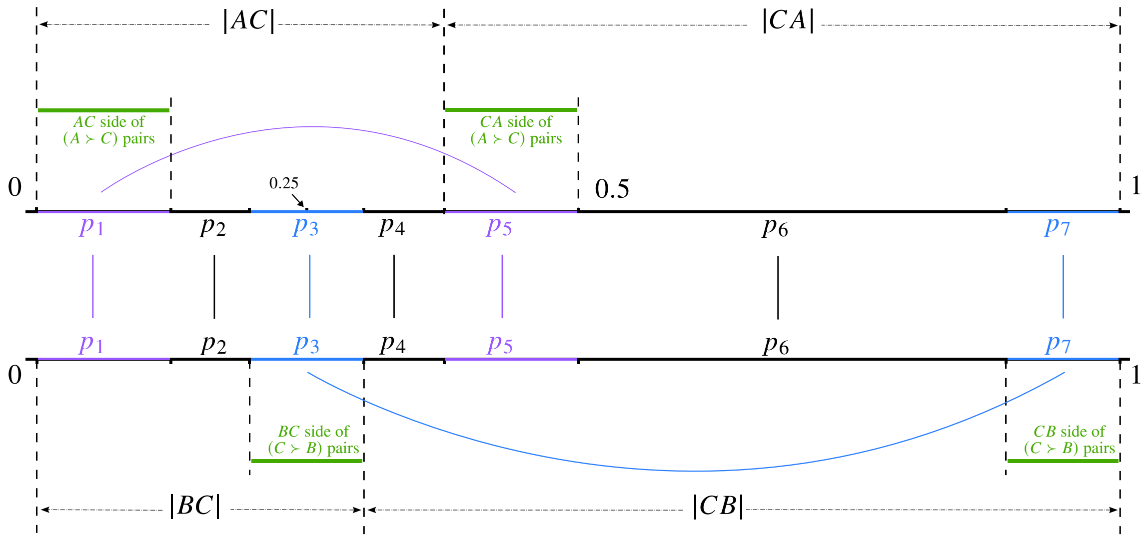


Figure 5: Interval split for Case 2. The interpretation of this figure is similar to Figure 4. Note that analogous to that case, we have $p_2 = p_4$.

$$
\begin{aligned}
\min \quad & \mathbb{E}X + (R+1)\cdot\mathbb{E}Y + R\cdot\mathbb{E}Z\,, && (16)\\
\text{s.t.} \quad & \text{First three constraints in Program (15)} \\
& X_4 \geqslant 0 \qquad \text{and} \qquad Y_4 \geqslant 0 \\
& X_1 + X_5 \geqslant 0 \qquad \text{and} \qquad Y_3 + Y_7 \geqslant 0 && (AC, CB \text{ matchings})\\
& \sum_{i=1}^{7} p_i = 1 \qquad \text{and} \qquad \sum_{i=1}^{5} p_i = 0.5 && (|AC| + W_{AC} = 0.5)\\
& p_1 = p_5, \qquad p_2 = p_4 \qquad \text{and} \qquad p_3 = p_7 && (\text{Coupling of masses})\\
& Z_i, p_i \geqslant 0\,,\ \forall i \in [7]\,.
\end{aligned}
$$

As before, the polytope over p is given by:

$$
\begin{aligned}
2p_1 + 2p_2 + 2p_3 + p_6 &= 1,\\
2p_1 + 2p_2 + p_3 &= 0.5,\\
p_1, p_2, p_3, p_6 &\geqslant 0.
\end{aligned}
$$

Eliminating p_3, p_6, the above again reduces to the interior of a triangle on (p_1, p_2) with vertices $\{(0,0), (0,0.25), (0.25,0)\}$. The polytope therefore has vertices given by

$$(p_1, p_2, p_3, p_6) = \{(0,0,0.5,0), (0,0.25,0,0.5), (0.25,0,0,0.5)\}.$$

Solving the resulting linear programs again shows that for $R = 2$, the objective is at least zero at each extreme point, hence showing the distortion is at most 3. Again, we present the dual certificates in Appendix A. This completes the proof of Theorem 1.1.

6 Lower Bound Instances for Any (λ, w)

To complete the full picture, we now show that the analysis in the previous section is tight by presenting a construction that yields a lower bound on distortion for any (λ, w). This lower bound is optimized at 3, hence informing our choice of parameters above.

Definition 6.1 (Permissible Ranges for $|AC|, |CB|$)**.** For $\lambda \in (1/2, 1)$ and $w > 0$, with $f(AC) = 1-\lambda$, one must have $AC_{\min} \leqslant |AC| \leqslant AC_{\max}$, and with $f(CB) = \lambda$, $CB_{\min} \leqslant |CB| \leqslant CB_{\max}$, where

$$AC_{\min}(\lambda, w) = \frac{1-\lambda}{1+\lambda w} \qquad CB_{\max}(\lambda, w) = \frac{\lambda(1+w)}{1+\lambda w}$$

$$CB_{\min}(\lambda, w) = \begin{cases} \dfrac{\lambda-(1-\lambda)w}{1-(1-\lambda)w} & \text{if } w \leqslant \dfrac{2\lambda-1}{1-\lambda} \\ \dfrac{\lambda}{1+(1-\lambda)w} & \text{if } w > \dfrac{2\lambda-1}{1-\lambda} \end{cases} \qquad AC_{\max}(\lambda, w) = \begin{cases} \dfrac{1-\lambda}{1-(1-\lambda)w} & \text{if } w \leqslant \dfrac{2\lambda-1}{1-\lambda} \\ \dfrac{(1-\lambda)(1+w)}{1+(1-\lambda)w} & \text{if } w > \dfrac{2\lambda-1}{1-\lambda}. \end{cases}$$

These are the quantities that allow the $f(\cdot)$ constraints to be satisfied by winning all deliberations (lower bounding the set sizes) or winning zero deliberation (upper bounding the set sizes). Observe that $AC_{\min} + CB_{\max} = AC_{\max} + CB_{\min} = 1$, regardless of λ, w. The four quantities are found by solving equations. For example, $AC_{\min}$ and $AC_{\max}$ are found by respectively solving:

$$\begin{aligned} \text{find} \quad & |AC| = AC_{\min} \\ \text{s.t.} \quad & m_{AC} = \min\{|AC|, |CA|\}, \\ & 0 \leqslant |AC| \leqslant 1, \qquad W_{AC} = m_{AC}, \\ & |AC| + w \cdot W_{AC} = (1-\lambda)(1 + w \cdot m_{AC}) \end{aligned} \qquad \begin{aligned} \text{find} \quad & |AC| = AC_{\max} \\ \text{s.t.} \quad & m_{AC} = \min\{|AC|, |CA|\}, \\ & 0 \leqslant |AC| \leqslant 1, \qquad W_{AC} = 0, \\ & |AC| + w \cdot W_{AC} = (1-\lambda)(1 + w \cdot m_{AC}) \end{aligned}$$

The quantities $CB_{\max}, CB_{\min}$ can be computed similarly. We now describe three types of instances. For all three examples, we assume V has unit mass.

Example 6.2 (Collinear Points $A - B - C$)**.** Embed $V \cup \{A, B, C\}$ on $\mathbb{R}$. Put $A = 0, B = 1$, and $C = 2$. Place voter v_B of mass $AC_{\max}$ at B and v_C with the remaining mass $CB_{\min} = 1 - AC_{\max}$ at C. Then:

- A vs. C. Arbitrate v_B in favor of A. Then $f(AC) = 1 - \lambda$ is satisfied by $|AC| = AC_{\max}$, with A winning zero deliberations.
- C vs. B. All (C, B) deliberations are ties, and we arbitrate all of them into $C \succ B$ pairs. Then $f(CB) = \lambda$ with $|CB| = CB_{\min}$ and C winning every deliberation matching.

This instance has distortion $SC(A)/SC(B) = (AC_{\max} + 2CB_{\min})/(CB_{\min})$. This generalizes Theorem 4.5.

Example 6.3 (Co-located B and C)**.** Embed $V \cup \{A, B, C\}$ on $\mathbb{R}$. Put $A = 0$ and $B = C = 1$. Place voter v_A of mass $AC_{\min}$ at A, and v_{BC} of remaining mass $CA_{\max} = 1 - AC_{\min}$ at B (equivalently C). Then:

- A vs. C. All (A, C) deliberations are ties and we arbitrate as $A \succ C$. Then $f(AC) = 1 - \lambda$ is satisfied by $|AC| = AC_{\min}$ along with A winning all deliberations.
- C vs. B. All (C, B) deliberations are also ties; we arbitrate in favor of $B \succ C$. Then $f(CB) = \lambda$ by $|CB| = CB_{\max}$, with C winning zero deliberation.

This instance has distortion $SC(A)/SC(B) = CB_{\max}/AC_{\min}$.

Example 6.4 (Triangular Instance)**.** Embed A, B, C on an equilateral triangle with side length 2, and partition voters into three point masses of ordinal preferences ACB, CBA, and BAC, respectively. Define their voter-candidate distances by the following table, where $\eta = 1 - CB_{\min} - AC_{\min} = AC_{\max} - AC_{\min} = CB_{\max} - CB_{\min}$.

We note that this instance can be embedded in $(\mathbb{R}^2, \ell_1)$ by placing $A = (0, 0)$, $B = (1, 1)$, $C = (2, 0)$, $ACB = (1, 0)$, $CBA = (2, 1)$, and $BAC = (1, 1)$.

Cluster	Mass	$d(v, A)$	$d(v, B)$	$d(v, C)$
ACB	η	1	1	1
CBA	$CB_{\min}$	3	1	1
BAC	$AC_{\min}$	2	0	2

- In this instance, $|AC| = AC_{\max}$ and $|CB| = CB_{\max}$.
- A vs. C. In the (A, C) deliberation, A is unable to win any: either ACB, BAC when paired with CBA results in $C \succ A$. However, because $|AC| = AC_{\max}$, this is exactly enough to ensure $f(AC) = 1 - \lambda$.
- C vs. B. By the same token, BAC beats both ACB, CBA in the (C, B) deliberation, so every pair outputs $(B \succ C)$. Still, as $|CB| = CB_{\max}$ we nevertheless reach $f(CB) = \lambda$.

This instance has distortion

$$\frac{SC(A)}{SC(B)} = \frac{(AC_{\max} - AC_{\min}) + 3 \cdot CB_{\min} + 2 \cdot AC_{\min}}{(AC_{\max} - AC_{\min}) + CB_{\min}}.$$

The Distortion Lower Bound Over (λ, w). Aggregating Examples 6.2 to 6.4, we obtain a (piecewise) lower bound of the distortion of our rule with parameters (λ, w). For each (λ, w), we compute the distortions $d_1(\lambda, w)$ from Example 6.2, $d_2(\lambda, w)$ from Example 6.3, and $d_3(\lambda, w)$ from Example 6.4. We then set $\mathcal{D}(\lambda, w) = \max_i d_i(\lambda, w)$ and plot it in Figure 6. This creates a 2D plane of lower bounds of the (λ, w) deliberation-via-matching protocol, with global minimizer (λ^*, w^*) attaining value $\mathcal{D}(\lambda^*, w^*) = 3$. By Theorem 1.1, we conclude that our parameter choice (λ^*, w^*) is tight and uniquely optimal. The exact algebraic expressions are supplemented in Appendix B.

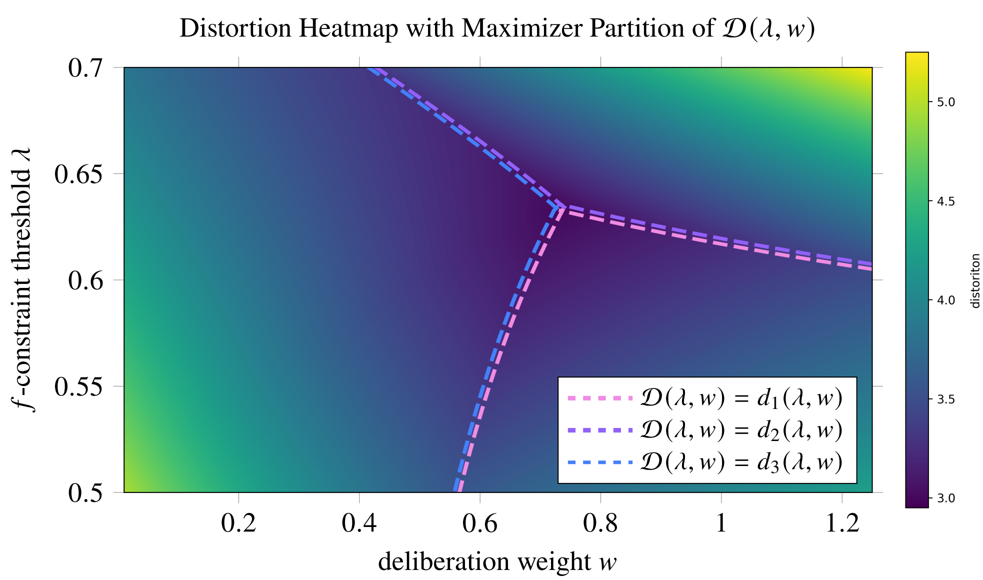


Figure 6: Distortion heatmap and the maximizer (arg max) partition of the (λ, w)-plane induced by d_1, d_2, d_3. Each color shows the decision region $\arg\max_i d_i(\lambda, w)$; dashed curves represent the decision boundaries $d_i = d_j$. As d_i quickly blows up, we only plot (λ, w) over $[0.5, 0.7] \times [0, 1.25]$. The unique global minimum of $\mathcal{D}(\lambda, w)$ is 3, attained by (λ^*, w^*) (this is also the unique intersection of all three decision boundaries).

Finally, observe that across all three examples, the lower bounds hold for every maximum cardinality matching, so long as we apportion preference and deliberation ties as described. Therefore, the tightness of Theorem 1.1 is robust to the *choice* of matchings, which justifies using an *arbitrary* maximum matching in our protocol.

7 Open Questions

The main open question from our work is to close the gap between the lower bound of 2 for deterministic social choice rules with pairwise deliberations, and the upper bound of 3. Further, can the bounds be improved via a randomized protocol? Our analysis crucially uses the λ-WUS tournament rule in order to restrict the analysis to three candidates. Can our bilinear relaxation extend to other types of tournament rules, for instance, those considered in [12]? Further, how can a protocol analogous to matching voters with opposing preferences be extended to deliberating groups of size more than two, and will such an extension also be amenable to bilinear relaxations? These questions make us believe that social choice with small-group deliberation is an exciting research direction with the potential for deep mathematical analysis, and the design of novel protocols with practical significance.

Acknowledgment: We used GPT-5 to assist with paraphrasing and clarifying portions of the text, generating code for the optimization routines, and drafting Appendix A based on the corresponding dual solution. All content was reviewed, verified, and finalized by the authors, who take full responsibility for the paper's accuracy and contributions.

References

[1] Ben Abramowitz, Elliot Anshelevich, and Wennan Zhu. Awareness of voter passion greatly improves the distortion of metric social choice. In Ioannis Caragiannis, Vahab Mirrokni, and Evdokia Nikolova, editors, *Web and Internet Economics*, pages 3–16, Cham, 2019. Springer International Publishing.

[2] Elliot Anshelevich, Onkar Bhardwaj, Edith Elkind, John Postl, and Piotr Skowron. Approximating optimal social choice under metric preferences. *Artif. Intell.*, 264:27–51, 2018.

[3] Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A. Voudouris. Distortion in social choice problems: An annotated reading list. *SIGecom Exch.*, 19(1):12–14, 2021.

[4] Elliot Anshelevich, Aris Filos-Ratsikas, and Alexandros A. Voudouris. The distortion of distributed metric social choice. *Artificial Intelligence*, 308:103713, 2022.

[5] Elliot Anshelevich and John Postl. Randomized social choice functions under metric preferences. *J. Artif. Intell. Res.*, 58:797–827, 2017.

[6] Lisa P. Argyle, Christopher A Bail, E. Busby, Joshua R Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences of the United States of America*, 120, 2023.

[7] Solomon E. Asch. Opinions and social pressure. *Scientific American*, 193(5):31–35, 1955.

[8] Joshua Ashkinaze, Emily Fry, Narendra Edara, Eric Gilbert, and Ceren Budak. Plurals: A system for guiding LLMs via simulated social ensembles. In Naomi Yamashita, Vanessa Evers, Koji Yatani,

Sharon Xianghua Ding, Bongshin Lee, Marshini Chetty, and Phoebe O. Toups Dugas, editors, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI 2025, YokohamaJapan, 26 April 2025- 1 May 2025*, pages 245:1–245:21. ACM, 2025.

[9] Michiel A. Bakker, Martin J. Chadwick, Hannah R. Sheahan, Michael Henry Tessler, Lucy Campbell-Gillingham, Jan Balaguer, Nat McAleese, Amelia Glaese, John Aslanides, Matthew M. Botvinick, and Christopher Summerfield. Fine-tuning language models to find agreement among humans with diverse preferences. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NeurIPS '22, 2022.

[10] Ioannis Caragiannis, Evi Micha, and Jannik Peters. Can a few decide for many? The metric distortion of sortition. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

[11] Moses Charikar and Prasanna Ramakrishnan. Metric distortion bounds for randomized social choice. In Joseph (Seffi) Naor and Niv Buchbinder, editors, *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 2986–3004. SIAM, 2022.

[12] Moses Charikar, Prasanna Ramakrishnan, Zihan Tan, and Kangning Wang. Metric distortion for tournament voting and beyond. In Itai Ashlagi and Aaron Roth, editors, *Proceedings of the 26th ACM Conference on Economics and Computation, EC 2025, Stanford University, Stanford, CA, USA, July 7-10, 2025*, pages 790–818. ACM, 2025.

[13] Moses Charikar, Kangning Wang, Prasanna Ramakrishnan, and Hongxun Wu. Breaking the metric voting distortion barrier. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 1621–1640. SIAM, 2024.

[14] Morton Deutsch and Harold Benjamin Gerard. A study of normative and informational social influences upon individual judgement. *Journal of abnormal psychology*, 51 3:629–36, 1955.

[15] Brandon Fain, Ashish Goel, Kamesh Munagala, and Sukolsak Sakshuwong. Sequential deliberation for social choice. In Nikhil R. Devanur and Pinyan Lu, editors, *Web and Internet Economics - 13th International Conference, WINE 2017, Bangalore, India, December 17-20, 2017, Proceedings*, volume 10660 of *Lecture Notes in Computer Science*, pages 177–190. Springer, 2017.

[16] J.S. Fishkin. *Democracy and Deliberation: New Directions for Democratic Reform*. Yale University Press, 1991.

[17] Bailey Flanigan, Ariel D. Procaccia, and Sven Wang. Distortion under public-spirited voting. In Kevin Leyton-Brown, Jason D. Hartline, and Larry Samuelson, editors, *Proceedings of the 24th ACM Conference on Economics and Computation, EC 2023, London, United Kingdom, July 9-12, 2023*, page 700. ACM, 2023.

[18] Vasilis Gkatzelis, Daniel Halpern, and Nisarg Shah. Resolving the optimal metric distortion conjecture. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1427–1438. IEEE, 2020.

[19] Ashish Goel, Mohak Goyal, and Kamesh Munagala. Metric distortion of small-group deliberation. In Michal Koucký and Nikhil Bansal, editors, *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025*, pages 1568–1579. ACM, 2025.

[20] Ashish Goel, Anilesh Kollagunta Krishnaswamy, and Kamesh Munagala. Metric distortion of social choice rules: Lower bounds and fairness properties. In Constantinos Daskalakis, Moshe Babaioff, and Hervé Moulin, editors, *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, Cambridge, MA, USA, June 26-30, 2017*, pages 287–304. ACM, 2017.

[21] Mohak Goyal and Sahasrajit Sarmasarkar. Metric distortion under probabilistic voting. In Itai Ashlagi and Aaron Roth, editors, *Proceedings of the 26th ACM Conference on Economics and Computation, EC 2025, Stanford University, Stanford, CA, USA, July 7-10, 2025*, page 840. ACM, 2025.

[22] Sean Ingham and Ines Levin. Can deliberative minipublics influence public opinion? Theory and experimental evidence. *Political Research Quarterly*, 71(3):654–667, 2018.

[23] David Kempe. An analysis framework for metric voting based on LP duality. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(02):2079–2086, Apr. 2020.

[24] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[25] Fatih Erdem Kizilkaya and David Kempe. Plurality veto: A simple voting rule achieving optimal metric distortion. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 349–355. ijcai.org, 2022.

[26] Nicholas R. Miller. Graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, 21(4):769–803, 1977.

[27] Kamesh Munagala and Kangning Wang. Improved metric distortion for deterministic social choice rules. In Anna R. Karlin, Nicole Immorlica, and Ramesh Johari, editors, *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019*, pages 245–262. ACM, 2019.

[28] Online deliberation platform. `https://stanforddeliberate.org/`.

[29] Vincent Price, Lilach Nir, and Joseph N. Cappella. Normative and informational influences in online political discussions. *Communication Theory*, 16(1):47–74, 03 2006.

[30] Ariel D. Procaccia and Jeffrey S. Rosenschein. The distortion of cardinal preferences in voting. In *Proceedings of the 10th International Conference on Cooperative Information Agents*, CIA'06, page 317–331, Berlin, Heidelberg, 2006. Springer-Verlag.

[31] Sally J. Scholz. Dyadic deliberation versus discursive democracy. *Political Theory*, 30(5):746–750, 2002.

[32] Denise Haunani Solomon, Miriam Brinberg, Graham D Bodie, Susanne Jones, and Nilam Ram. A dynamic dyadic systems approach to interpersonal communication. *Journal of Communication*, 71(6):1001–1026, 09 2021.

[33] George G. Szpiro. *Numbers Rule: The Vexing Mathematics of Democracy, from Plato to the Present*. Princeton University Press, 2010.

[34] Amy Zhang, Bryan Culbertson, and Praveen Paritosh. Characterizing online discussion using coarse discourse sequences. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):357–366, May 2017.

A Explicit Dual Construction

We now present an analytic proof based on the LP solutions to the bilinear program, via the dual certificate of optimality of the corresponding LP. We first note that by Equation (10), the term $Z_i \geqslant Z_{\min}(X_i, Y_i)$ can be captured by the following set of linear constraints:

$$\begin{aligned}
&M_X \geqslant X_i, \quad M_X \geqslant -X_i, \quad M_Y \geqslant Y_i, \quad M_Y \geqslant -Y_i && \forall i \in [9],\\
&M_{X+Y} \geqslant X_i + Y_i, \quad M_{X+Y} \geqslant -(X_i + Y_i) && \forall i \in [9],\\
&Z_i \geqslant \tfrac{1}{2}(M_X + X_i), \quad Z_i \geqslant \tfrac{1}{2}(M_Y - Y_i), \quad Z_i \geqslant \tfrac{1}{2}\big(M_{X+Y} + X_i - Y_i\big) && \forall i \in [9],
\end{aligned}$$

The variables M_X, M_Y, and M_{X+Y} represent $\|X\|_\infty$, $\|Y\|_\infty$, and $\|X+Y\|_\infty$ respectively.

A.1 Dual Certificates for Case 1

A.1.1 Vertex $(p_1, p_2, p_3, p_6) = (0, 0, 0.5, 0)$

We consider the LP obtained by substituting $p_3 = p_8 = 0.5$ into Program (15). The primal objective (for $R = 2$) is

$$\Phi_2 \;=\; \tfrac{1}{2}(X_3 + X_8) \;+\; \tfrac{3}{2}(Y_3 + Y_8) \;+\; (Z_3 + Z_8).$$

A valid dual certificate is given by the following nonnegative multipliers on the displayed constraints:

$$\begin{aligned}
0.5 &\quad \text{on} \quad M_{X+Y} + X_8 + Y_8 \geqslant 0,\\
1 &\quad \text{on} \quad Z_3 - \tfrac{1}{2}M_{X+Y} + \tfrac{1}{2}Y_3 - \tfrac{1}{2}X_3 \geqslant 0,\\
1 &\quad \text{on} \quad Z_8 \geqslant 0,\\
1 &\quad \text{on} \quad Y_3 + Y_8 \geqslant 0,\\
1 &\quad \text{on} \quad X_3 \geqslant 0.
\end{aligned}$$

Adding the weighted inequalities (left-hand sides minus right-hand sides) with these multipliers gives

$$\begin{aligned}
0.5\big(M_{X+Y} + X_8 + Y_8\big) + 1\big(Z_3 - \tfrac{1}{2}M_{X+Y} + \tfrac{1}{2}Y_3 - \tfrac{1}{2}X_3\big)&\\
+\, 1 \cdot Z_8 + 1\big(Y_3 + Y_8\big) + 1 \cdot X_3 \;\geqslant\; 0.&
\end{aligned}$$

Collecting terms on the left-hand side, all M-terms cancel and the sum simplifies exactly to

$$\tfrac{1}{2}X_3 + \tfrac{1}{2}X_8 + \tfrac{3}{2}Y_3 + \tfrac{3}{2}Y_8 + Z_3 + Z_8 \;=\; \Phi_2.$$

Hence $\Phi_2 \geqslant 0$ for every primal feasible point.

A.1.2 Vertex $(p_1, p_2, p_3, p_6) = (0, 0.25, 0, 0)$

For V_2 the mass pattern places $p = 0.25$ on indices $2, 4, 7, 9$. The objective (for $R = 2$) becomes

$$\Phi_2 \;=\; 0.25(X_2 + X_4 + X_7 + X_9) \;+\; 0.75(Y_2 + Y_4 + Y_7 + Y_9) \;+\; 0.5(Z_2 + Z_4 + Z_7 + Z_9).$$

A valid choice of multipliers (all nonnegative) on the primal constraints is:

$$\begin{array}{rll}
0.25 & \text{on} & M_Y + Y_4 \geqslant 0, \\
0.25 & \text{on} & M_{X+Y} + X_7 + Y_7 \geqslant 0, \\
0.25 & \text{on} & M_{X+Y} + X_9 + Y_9 \geqslant 0, \\
0.5 & \text{on} & Z_2 - \frac{1}{2}M_{X+Y} + \frac{1}{2}Y_2 - \frac{1}{2}X_2 \geqslant 0, \\
0.5 & \text{on} & Z_4 - \frac{1}{2}M_{X+Y} + \frac{1}{2}Y_4 - \frac{1}{2}X_4 \geqslant 0, \\
0.5 & \text{on} & Z_7 - \frac{1}{2}M_Y + \frac{1}{2}Y_7 \geqslant 0, \\
0.5 & \text{on} & Z_9 \geqslant 0, \\
0.5 & \text{on} & X_2 + X_4 \geqslant 0, \\
0.5 & \text{on} & Y_2 + Y_9 \geqslant 0, \\
0.25 & \text{on} & Y_4 + Y_7 \geqslant 0.
\end{array}$$

Summing these weighted inequalities yields on the left hand side

$$0.25X_2 + 0.25X_4 + 0.25X_7 + 0.25X_9 + 0.75Y_2 + 0.75Y_4 + 0.75Y_7 + 0.75Y_9 + 0.5Z_2 + 0.5Z_4 + 0.5Z_7 + 0.5Z_9,$$

which is precisely Φ_2. Thus $\Phi_2 \geqslant 0$.

A.1.3 Vertex $(p_1, p_2, p_3, p_6) = (0.25, 0, 0, 0.5)$

For V_3 we use the mass assignment $p_1 = 0.25,\ p_5 = 0.25,\ p_6 = 0.5$. The objective (for $R = 2$) is

$$\Phi_2 \ = \ 0.25(X_1 + X_5) + 0.5X_6 \ + \ 0.75(Y_1 + Y_5) + 1.5\,Y_6 \ + \ 0.5(Z_1 + Z_5) + 1.0\,Z_6.$$

A valid set of nonnegative multipliers is

$$\begin{array}{rll}
0.5 & \text{on} & M_Y + Y_1 \geqslant 0, \\
0.5 & \text{on} & M_{X+Y} + X_6 + Y_6 \geqslant 0, \\
0.5 & \text{on} & Z_1 - \frac{1}{2}M_{X+Y} + \frac{1}{2}Y_1 - \frac{1}{2}X_1 \geqslant 0, \\
0.5 & \text{on} & Z_5 - \frac{1}{2}M_{X+Y} + \frac{1}{2}Y_5 - \frac{1}{2}X_5 \geqslant 0, \\
1 & \text{on} & Z_6 - \frac{1}{2}M_Y + \frac{1}{2}Y_6 \geqslant 0, \\
0.5 & \text{on} & Y_6 - Y_5 \geqslant 0, \\
0.5 & \text{on} & X_1 + X_5 \geqslant 0, \\
1 & \text{on} & Y_5 \geqslant 0.
\end{array}$$

Summing these weighted inequalities yields on the left hand side

$$0.25X_1 + 0.25X_5 + 0.5X_6 + 0.75Y_1 + 0.75Y_5 + 1.5Y_6 + 0.5Z_1 + 0.5Z_5 + 1.0Z_6,$$

which equals Φ_2, and therefore $\Phi_2 \geqslant 0$.

A.2 Dual certificates for Case 2

A.2.1 Vertex $(p_1, p_2, p_3, p_6) = (0, 0, 0.5, 0)$

We use the mass assignment $p_3 = p_7 = 0.5$. The objective is

$$\Phi_2 = \tfrac{1}{2}X_3 + \tfrac{1}{2}X_7 + \tfrac{3}{2}Y_3 + \tfrac{3}{2}Y_7 + Z_3 + Z_7.$$

A valid dual certificate is given by the following nonnegative multipliers on the displayed constraints:

$$\begin{aligned}
0.5 &\quad \text{on} \quad M_X + X_7 \geqslant 0,\\
1 &\quad \text{on} \quad Z_3 - \tfrac{1}{2}M_X - \tfrac{1}{2}X_3 \geqslant 0,\\
1 &\quad \text{on} \quad X_3 - X_4 \geqslant 0,\\
1.5 &\quad \text{on} \quad Y_3 + Y_7 \geqslant 0,\\
1 &\quad \text{on} \quad X_4 \geqslant 0,\\
1 &\quad \text{on} \quad Z_7 \geqslant 0.
\end{aligned}$$

Multiply and sum these inequalities with the listed multipliers. On the left-hand side the M-terms cancel:

$$0.5(M_X + X_7) \; + \; 1\big(Z_3 - \tfrac{1}{2}M_X - \tfrac{1}{2}X_3\big) \; + \; 1(X_3 - X_4)$$
$$+ \; 1.5(Y_3 + Y_7) \; + \; 1 \cdot X_4 \; + \; 1 \cdot Z_7 \; \geqslant \; 0.$$

Grouping and simplifying the left hand side yields exactly

$$\tfrac{1}{2}X_3 + \tfrac{1}{2}X_7 + \tfrac{3}{2}Y_3 + \tfrac{3}{2}Y_7 + Z_3 + Z_7 \; = \; \Phi_2.$$

Thus $\Phi_2 \geqslant 0$.

A.2.2 Vertex $(p_1, p_2, p_3, p_6) = (0, 0.25, 0, 0.5)$

We use the mass assignment $p_2 = p_4 = 0.25, p_6 = 0.5$. A valid dual certificate is given by the following nonnegative multipliers on the displayed constraints:

$$\begin{aligned}
0.5 &\quad \text{on} \quad M_Y + Y_2 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad M_{X+Y} + X_6 + Y_6 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad Z_2 - \tfrac{1}{2}M_{X+Y} + \tfrac{1}{2}Y_2 - \tfrac{1}{2}X_2 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad Z_4 - \tfrac{1}{2}M_{X+Y} + \tfrac{1}{2}Y_4 - \tfrac{1}{2}X_4 \; \geqslant \; 0,\\
1 &\quad \text{on} \quad Z_6 - \tfrac{1}{2}M_Y + \tfrac{1}{2}Y_6 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad X_2 - X_3 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad X_3 - X_4 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad Y_5 - Y_4 \; \geqslant \; 0,\\
0.5 &\quad \text{on} \quad Y_6 - Y_5 \; \geqslant \; 0,\\
1 &\quad \text{on} \quad X_4 \; \geqslant \; 0,\\
1 &\quad \text{on} \quad Y_4 \; \geqslant \; 0.
\end{aligned}$$

Summing these weighted inequalities yields on the left hand side

$$0.25(X_2 + X_4) + 0.5X_6 + 0.75(Y_2 + Y_4) + 1.5Y_6 + 0.5(Z_2 + Z_4) + 1.0Z_6.$$

This is exactly

$$\Phi_2 = \mathbb{E}X + 3\mathbb{E}Y + 2\mathbb{E}Z$$

under the mass assignment $p_2 = p_4 = 0.25,\; p_6 = 0.5$. Therefore $\Phi_2 \geqslant 0$ for this vertex.

A.2.3 Vertex $(p_1, p_2, p_3, p_6) = (0.25, 0, 0, 0.5)$

We use the mass assignment $p_1 = p_5 = 0.25,\ p_6 = 0.5$. A valid dual certificate is given by the following nonnegative multipliers on the displayed constraints:

$$\begin{array}{rll} 0.5 & \text{on} & M_Y + Y_1 \geqslant 0, \\ 0.5 & \text{on} & M_{X+Y} + X_6 + Y_6 \geqslant 0, \\ 0.5 & \text{on} & Z_1 - \frac{1}{2}M_{X+Y} + \frac{1}{2}Y_1 - \frac{1}{2}X_1 \geqslant 0, \\ 0.5 & \text{on} & Z_5 - \frac{1}{2}M_{X+Y} + \frac{1}{2}Y_5 - \frac{1}{2}X_5 \geqslant 0, \\ 1 & \text{on} & Z_6 - \frac{1}{2}M_Y + \frac{1}{2}Y_6 \geqslant 0, \\ 0.5 & \text{on} & Y_6 - Y_5 \geqslant 0, \\ 0.5 & \text{on} & X_1 + X_5 \geqslant 0, \\ 1 & \text{on} & Y_5 \geqslant 0. \end{array}$$

Summing these weighted inequalities yields on the left hand side

$$0.25(X_1 + X_5) + 0.5X_6 + 0.75(Y_1 + Y_5) + 1.5Y_6 + 0.5(Z_1 + Z_5) + 1.0Z_6,$$

which equals the objective

$$\Phi_2 = \mathbb{E}X + 3\mathbb{E}Y + 2\mathbb{E}Z$$

for the mass assignment $p_1 = p_5 = 0.25,\ p_6 = 0.5$. Thus $\Phi_2 \geqslant 0$ for this vertex.

Therefore, for each of the six extreme points of the p-polytope in Case 1 and 2, the nonnegative dual multipliers above satisfy

$$\sum_j \lambda_j(\mathrm{LHS}_j - \mathrm{RHS}_j) = \Phi_2,$$

hence each yields an analytic dual certificate proving $\Phi_2 \geqslant 0$ and therefore a distortion bound of at most 3.

B Closed-form Distortion Lower Bound

We now derive closed-form lower bounds to the (λ, w) deliberation-via-matching protocol's distortion based on Examples 6.2 to 6.4. Let $\tau(\lambda) = \dfrac{2\lambda - 1}{1 - \lambda}$.

Distortion of Example 6.2. These instances have distortion $(AC_{\max} + 2CB_{\min})/CB_{\min} = 1 + 1/CB_{\min}$. Hence

$$d_1(\lambda, w) = \begin{cases} 1 + \dfrac{1 - (1-\lambda)w}{\lambda - (1-\lambda)w} = \dfrac{1 + \lambda - 2(1-\lambda)w}{\lambda - (1-\lambda)w} & \text{if } w \leqslant \tau(\lambda) \\[2ex] 1 + \dfrac{1 + (1-\lambda)w}{\lambda} = \dfrac{\lambda + 1 + (1-\lambda)w}{\lambda} & \text{if } w > \tau(\lambda). \end{cases}$$

Distortion of Example 6.3. In these instances, $d_2(\lambda, w) = CB_{\max}/AC_{\min} = \dfrac{\lambda(1+w)}{1-\lambda}$.

Distortion of Example 6.4. We first rewrite the symbolic distortion by replacing $CB_{\min}$ with the simpler $CB_{\max}$ by

$$d_3 = \frac{0.5 \cdot (AC_{\max} - AC_{\min}) + 1.5 \cdot CB_{\min} + AC_{\min}}{0.5 \cdot (AC_{\max} - AC_{\min}) + 0.5 \cdot CB_{\min}} = \frac{3 - 2AC_{\max} + AC_{\min}}{1 - AC_{\min}} = \frac{3 - 2AC_{\max} + AC_{\min}}{CB_{\max}},$$

where we multiplied both sides of the first fraction by 2 and used the identity $AC_{\max} + CB_{\min} = 1$. Then,

$$d_3(\lambda, w) = \begin{cases} \dfrac{2 + \lambda + (\lambda^2 + 6\lambda - 4)w - 3\lambda(1-\lambda)w^2}{\lambda(1+w)(1-(1-\lambda)w)} & \text{if } w \leqslant \tau(\lambda) \\[2ex] \dfrac{2 + \lambda + (3\lambda^2 - 2\lambda + 2)w + (\lambda - \lambda^2)w^2}{\lambda(1+w)(1+(1-\lambda)w)} & \text{if } w > \tau(\lambda). \end{cases}$$

Finally, the piecewise analytic lower bound is given by $\mathcal{D}(\lambda, w) = \max\{d_1(\lambda, w), d_2(\lambda, w), d_3(\lambda, w)\}$.