DISTANCE RATIONALIZATION OF ANONYMOUS AND HOMOGENEOUS VOTING RULES

BENJAMIN HADJIBEYLI AND MARK C. WILSON

Abstract. The concept of distance rationalizability of voting rules has been explored in recent years by several authors. Most previous work has dealt with a definition in terms of preference profiles. However, most voting rules in common use are anonymous and homogeneous. In this case there is a much more succinct representation (using the voting simplex) of the inputs to the rule. This representation has been widely used in the voting literature, but rarely in the context of distance rationalizability.

Recently, the present authors showed, as a special case of general results on quotient spaces, exactly how to connect distance rationalizability on profiles for anonymous and homogeneous rules to geometry in the simplex. In this article we develop the connection for the important special case of votewise distances, recently introduced and studied by Elkind, Faliszewski and Slinko in several papers. This yields a direct interpretation in terms of well-developed mathematical topics not seen before in the voting literature, namely Kantorovich (also called Wasserstein) distances and the geometry of Minkowski spaces.

As an application of this approach, we prove some positive and some negative results about the decisiveness of distance rationalizable anonymous and homogeneous rules. The positive results connect with the recent theory of hyperplane rules, while the negative ones deal with distances that are not metrics, controversial notions of consensus, and the fact that the ℓ^1-norm is not strictly convex.

We expect that the above novel geometric interpretation will aid the analysis of rules defined by votewise distances, and the discovery of new rules with desirable properties.

1. Introduction

We are interested in the relation between two ways of describing voting rules (interpreted broadly), each of which has a geometric flavour.

The class of anonymous and homogeneous voting rules includes all rules used in practice, and most rules appearing in the research literature (Dodgson's rule is a notable exception). For such rules there is an obvious concise way to describe an input profile of preferences, using the *vote simplex*. This approach goes back at least as far as Young [22] and was extensively developed and popularized by Saari [17]. By allowing us to use geometric intuition, it aids in the analysis of many properties of anonymous and homogeneous voting rules.

The framework of *distance rationalizability* is a useful way to organize the huge number of voting rules that have been introduced. By decomposing a rule into a *consensus* and a notion of *distance* to that consensus, the framework allows systematic derivation of axiomatic properties of the rule from those of its components. This kind of analysis has been systematically carried out recently by Elkind, Faliszewski and Slinko [6, 5] and the present authors [7], following preliminary work by Campbell, Lerer and Nitzan [15, 10, 1] and Meskanen and Nurmi [13].

Until now, these two approaches have not been systematically connected. Specific distance-based rules have indeed been studied in the simplex or permutahedron, notably by Zwicker and coauthors [24, 23, 3]. However, a more general approach is lacking. As shown in [7], the theory can be developed simultaneously for social choice rules and social welfare rules, and for very general distances and consensus notions, in a way that clarifies the relationship between the profile-based and simplex-based representations.

Key words and phrases. social choice theory, collective decision-making, rankings, homogeneity, anonymity, simplex, Wasserstein metric, Kantorovich distance, Earth Mover distance.

MW thanks Elchanan Mossel and Miklos Racz for research hospitality and useful conversations.

1.1. **Our contribution.** After developing in Section 2 the basic notation and terminology, in Section 3 we show how the usual distance rationalizability approach on profiles connects with the geometric approach on the vote simplex. We focus in Section 3.3 on the special case of ℓ^p-votewise distances, which have been shown by Elkind, Faliszweski and Slinko to have many desirable properties. We show that the distance induced on the simplex is a *Wasserstein* distance. In particular when $p = 1$ this is induced by a norm and we can interpret the theory in terms of the geometry of Minkowski spaces. This gives a completely new perspective not yet seen in the voting literature, and suggests not only new voting rules but also a new tool for analysis of existing rules.

In particular, in Section 4 we use the simplex representation to explore, for the first time in the literature, the decisiveness of distance rationalizable rules. On the positive side, we give a sufficient condition in Corollary 6.4 for rule to be a *hyperplane rule* and therefore be rather decisive. For example, any rule defined using an ℓ^p votewise norm and the strong or weak unanimity consensus satisfies this condition. On the negative side, we show in Proposition 5.1 that large tied sets can occur easily with the commonly used ℓ^1 votewise metrics, unless the notion of consensus is very well chosen. This sheds light on some common consensus notions such as that of Condorcet. In Section 7 we make some recommendations for desirable properties of consensus sets and distances.

The approach adopted here and in [7] allows for systematic exploration of the space of aggregation rules and the construction of rules with guaranteed axiomatic properties.

2. Basic definitions

We use standard concepts of social choice theory. Not all of these concepts have completely standardized names.

Definition 2.1. *We fix a finite set $C = \{c_1, c_2, \dots, c_m\}$ of* **candidates** *and an infinite set $V^* = \{v_1, v_2, \dots, \}$ of potential* **voters**. *For each s with $1 \leq s \leq m$, an s-***ranking** *is a strict linear order of s elements chosen from C. The set of all s-rankings is denoted $L_s(C)$. When $s = m$ we write simply $L(C)$. When $s = 1$ we identify $L_1(C)$ with C in the natural way.*

Definition 2.2. *A* **profile** *is a function $\pi : V \to L(C)$ where $V \subset V^*$ is finite. We denote the set of all profiles by $\mathcal{P}$. An* **election** *is a triple (C, V, π) with $\pi \in \mathcal{P}$ and $\pi : V \to L(C)$. We denote the set of all elections with fixed C and V by $\mathcal{E}(C, V)$, and the class of all elections by $\mathcal{E}$.*

Definition 2.3. *A* **social rule of size** *s is a function R that takes each election $E = (C, V, \pi)$ to a nonempty subset of $L_s(C)$. When there is a unique s-ranking chosen, the word "rule" becomes "function". When $s = 1$, we have the usual* **social choice function**, *and when $s = |V|$ the usual* **social welfare function**.

For each subset D of $\mathcal{E}$ we can consider a **partial social rule with domain** *D to be defined as above, but with domain restricted to D.*

2.1. **Consensus.** Intuitively, a consensus is simply a socially agreed unique outcome on some set of elections. We now define it formally.

Definition 2.4. *An s-***consensus** *is a partial social function $\mathcal{K}$ of size s. The domain $D(\mathcal{K})$ of $\mathcal{K}$ is called an s-***consensus set** *and is partitioned into the inverse images $\mathcal{K}_r := \mathcal{K}^{-1}(\{r\})$.*

Several specific consensuses have been described in the literature. We list a few important ones. Some have been discussed by previous authors only in the case $s = 1$ but the definitions extend naturally.

Definition 2.5. *We use the following consensuses in this article.*

- *We denote by $\mathbf{S}^s$ the consensus $\mathcal{K}$ for which $\mathcal{K}_r$ is the election in which all voters agree that r is the ranking of the top s candidates. When $s = |V|$, we simply write* **S** *(called the* **strong unanimity consensus***), whereas when $s = 1$, for consistency with previous authors we denote it* **W**, *the* **weak unanimity consensus**.

- *The* 1-**Condorcet** *consensus* **C** *has domain consisting of all elections for which a Condorcet winner exists. That is, there is a candidate c such that for all other candidates b a fraction strictly greater than $1/2$ of voters rank c above b.*

2.2. **Distances.** We require a notion of distance on elections. We aim to be as general as possible.

Definition 2.6. *(distance) A* **distance** *(or* **hemimetric***) on $\mathcal{E}$ is a function $d : \mathcal{E} \times \mathcal{E} \to \mathbb{R}_+ \cup \{\infty\}$ that satisfies the identities*

- $d(x, x) = 0$,
- $d(x, z) \leq d(x, y) + d(y, z)$.

A **pseudometric** *is a distance that also satisfies $d(x, y) = d(y, x)$. A* **quasimetric** *is a distance that also satisfies $d(x, y) = 0 \Rightarrow x = y$. A* **metric** *is a distance that is both a quasimetric and a pseudometric. We call a distance* **standard** *if $d(E, E') = \infty$ whenever E and E' have different sets of voters or candidates (this term has not been used in previous literature).*

One commonly used class of distances consists of the **votewise** distances [5] (Definition 2.11 below). First we require some preliminary definitions.

Example 2.7. *(commonly used distances on $L(C)$) We discuss the following distances on $L(C)$ in this article.*

- *The* **discrete metric** d_H*, defined by*

$$d_H(\rho, \rho') = \begin{cases} 1 & \text{if } \rho = \rho' \\ 0 & \text{otherwise.} \end{cases}$$

- *The* **inversion metric** d_K *(also called the swap, bubblesort or Kendall-τ metric), where $d_K(\rho, \sigma)$ is the minimum number of swaps of adjacent elements needed to change ρ into σ.*

Definition 2.8. *A* **seminorm** *on a real vector space X is a real-valued function N satisfying the identities*

- $N(x + y) \leq N(x) + N(y)$
- $N(\lambda x) = |\lambda| N(x)$

for all $x, y \in X$ and all $\lambda \in \mathbb{R}$. Note that this implies that $N(0) = 0$ and $N(x) \geq 0$ for all $x \in X$.

A **norm** *is a seminorm that also satisfies*

- $N(x) = 0 \Rightarrow x = 0$.

Remark 2.9. *Every seminorm induces a pseudometric via $d(x, y) = ||x - y||$. This is a metric if and only if the seminorm is a norm.*

Example 2.10. *Consider an n-dimensional space X with fixed basis $e_1, \dots, e_n$ and corresponding coefficients x_i for each element $x \in X$. Fix p with $1 \leq p < \infty$ and define the ℓ^p-norm on X by*

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

When $p = \infty$ we define the ℓ^∞ norm by

$$||x||_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Definition 2.11. *(votewise distances)*

Fix candidate set C and voter set V, and a distance d on $L(C)$. Choose a family $\{N_n\}_{n \geq 1}$ of seminorms, where N_n is defined on $\mathbb{R}^n$. Extend d to a function on $\mathcal{P}(C, V)$ by taking $n = |V|$ and defining for $\sigma, \pi \in \mathcal{P}(C, V)$

$$d^{N_n}(\pi, \sigma) := N_n(d(\pi_1, \sigma_1), \dots, d(\pi_n, \sigma_n)).$$

This yields a distance on elections having the same set of voters and candidates. We complete the definition of the extended distance (which we denote by d^N) on $\mathcal{E}$ by declaring it to be standard.

We use the abbreviation d^p for d^{ℓ^p}, and sometimes we even use just d for d^N if the meaning is clear.

Remark 2.12. *Note that if d is a metric and N is a norm, then d^N is a metric.*

Example 2.13. *(famous votewise distances) The distances d^1_H and d^1_K are called respectively the* **Hamming metric** *and* **Kemeny metric**. *The Hamming metric measures the minimum number of voters whose preferences must be changed in order to convert one profile to another, and as such has an interpretation in terms of bribery. The Kemeny metric measures the minimum number of swaps of adjacent candidates required to convert one profile to another, and is related to models of voter error.*

Example 2.14. *(tournament distances) Given an election $E = (C, V, \pi)$, we form the* **pairwise tournament digraph** $\Gamma(E)$ *with nodes indexed by the candidates, where the arc from a to b has weight equal to the* net support *for a over b in a pairwise contest. Formally, there is an arc from a to b whose weight equals $n_{ab} - n_{ba}$, where n_{ab} denotes the number of rankings in π in which a is above b.*

*Let $M(E)$ be the weighted adjacency matrix of $\Gamma(E)$ (with respect to an arbitrarily chosen fixed ordering of C). Given a seminorm N on the space of all $|C| \times |C|$ real matrices, we define the N-***tournament distance** *by*

$$d^N(E, E') = N(M(E) - M(E')).$$

*A closely related distance is defined in the analogous way, but where each element of the adjacency matrix is replaced by its sign (1, 0, or −1). We call this the N-***reduced tournament distance**. *We denote the special cases where N is the ℓ^1 norm on matrices by d^T and d^{RT} respectively. Every (reduced) tournament distance is a pseudometric.*

2.3. **Combining consensus and distance.** In order for a rule to be definable via the DR construction, it is necessary that the following property holds, and we shall assume this from now on.

Definition 2.15. *Let d be a distance and $\mathcal{K}$ a consensus. Say that $(\mathcal{K}, d)$* **distinguishes consensus choices** *if whenever $x \in \mathcal{K}_r, y \in \mathcal{K}_{r'}$ and $r \neq r'$, then $d(x, y) > 0$.*

We use a distance to extend a consensus to a social rule in the natural way. The choice at a given election E consists of all s-rankings r whose consensus set $\mathcal{K}_r$ minimizes the distance to E. We introduce the idea of a score in order to use our intuition about positional scoring rules.

Definition 2.16. *(DR scores and rules)*

Suppose that $\mathcal{K}$ is an s-consensus and d a distance on $\mathcal{E}$. Fix an election $E \in \mathcal{E}$. The $(\mathcal{K}, d, E)$**-score** *of $r \in L_s(C)$ is defined by*

$$|r| := d(E, \mathcal{K}_r) := \inf_{E' \in \mathcal{K}_r} d(E, E').$$

The rule $R := \mathcal{R}(\mathcal{K}, d)$ is defined by

$$R(E) = \arg \min_{r \in L_s(C)} |r|. \tag{1}$$

We say that R is **distance rationalizable** *(DR) with respect to $(\mathcal{K}, d)$.*

Example 2.17. *(scoring rules) The* **positional scoring rule** *defined by a vector w of* **weights** *with $w_1 \geq \cdots \geq w_m$ and $w_1 > w_m$ elects all candidates with maximum score, where the score of a in the profile π is defined as $\sum_{v \in V} w_{r(\pi(v), a)}$. Every positional scoring rule has the form $\mathcal{R}(\mathbf{W}, d^1_w)$ where d_w is the distance on rankings defined by*

$$d_w(\rho, \rho') = \sum_{c \in C} |w_{r(\rho, c)} - w_{r(\rho', c)}|.$$

Remark 2.18. *Note that d_w is a metric on $L_s(C)$ if and only if $w_1, \ldots, w_s$ are all distinct. The* **plurality** *($w = (1, 0, 0, \ldots, 0)$) and* **Borda** *($w = (m-1, m-2, \ldots, 0)$) rules are special cases, where d_w simplifies to d_H^1 and d_S^1 respectively. The score of r under the rule defined by w is the difference $nw_1 - |r|$. For example, for Borda with m candidates the maximum possible score of a candidate c is $(m-1)n$, achieved only for those elections in $\mathbf{W}_c$. The score of c under Borda is exactly $n(m-1) - K$ where K is the total number of swaps of adjacent candidates needed to move c to the top of all preference orders in $\pi(E)$.*

Example 2.19. *(Copeland's rule)* **Copeland's rule** *can be represented as $\mathcal{R}(\mathbf{C}, d_{RT})$. Indeed, in an election E, the Copeland score of a candidate c (the number of points it scores in pairwise contests with other candidates) equals $n - 1 - s$, where s is the minimum number of pairwise results that must be changed for E to change to an election that belongs to $\mathbf{C}_c$.*

3. Simplex rules

Although it is far from the general case, most rules used in practice are in fact anonymous and homogeneous. In this case there is an appealing geometric interpretation. The probability distribution of votes is sufficient information to determine the output of the rule, and so profile space can be substantially compressed.

We have previously explored in detail the connection with distance rationalization [7]. We first recall the construction for anonymous rules. Recall that a **multiset** of weight n on an underlying set S of size M is "a set of n elements of S with repetitions". Formally, there is a function $f : S \to \mathbb{N}$ where $f(s)$ gives the multiplicity of s in the multiset.

Definition 3.1. *Let $E = (C, V, \pi) \in \mathcal{E}$. The* **vote number map** *$\mathcal{N}$ is the quotient map that associates E with the multiset $\mathcal{N}(E)$ on $L(C)$ of weight $|V|$, in which the multiplicity of $\pi \in L(C)$ is the number of voters in V having that preference order.*

A rule R is **anonymous** *if $R(E) = R(E')$ whenever $\mathcal{N}(E) = \mathcal{N}(E')$. We denote the quotient space by $\mathcal{V}$ and call it the set of* **anonymous profiles***.*

Remark 3.2. *$\mathcal{N}(E)$ simply keeps track of the numbers of votes of each type in π. A rule is anonymous if this information is enough to determine the output.*

Definition 3.3. *The* **vote distribution** *associated to E is the probability distribution on $L(C)$ induced by the multiset $\mathcal{N}(E)$, which we denote $\mathcal{D}(E)$. The vote distribution map defines an equivalence relation $\sim$ on $\mathcal{E}$ in the usual way: $E \sim E'$ if and only if $\mathcal{D}(E) = \mathcal{D}(E')$. We denote the quotient space by $\mathcal{P}$, and call it the set of* **anonymous and homogeneous profiles***.*

Remark 3.4. *Note that if $E = (C, V, \pi), E' = (C, V', \pi') \in \mathcal{E}$ and $\mathcal{N}(E) = \mathcal{N}(E')$, then $|V| = |V'|$. Thus there is a well-defined map $f : \mathcal{V} \to \mathcal{P}$ ("divide by the number of voters"), and $\mathcal{D} = f \circ \mathcal{N}$.*

Remark 3.5. *Let g be a permutation of V^*. For each $E = (C, V, \pi)$, define $g(E) = (C, V, g(\pi))$ where $g(\pi)$ is the profile formed by permuting the linear orders to match the permutation of the voters. A rule R is anonymous if and only if the identity $R(g(E)) = R(E)$ is satisfied.*

For $k \geq 1$, define kE to be an election $(C, kV, k\pi)$ where kV consists of k copies of each voter in V and $k\pi$ the corresponding copies of their preference orders (the exact order of the voters is irrelevant since we deal only with anonymous rules). An anonymous rule R is homogeneous if and only if the identity $R(kE) = R(E)$ is satisfied.

We call anonymous and homogeneous rules **simplex rules** for short, and now explain why. So far the discussion has been coordinate-free, but it is often useful to introduce coordinates. Given any linear ordering $\rho_1, \rho_2, \ldots, \rho_M$ on $L(C)$, we can introduce coordinates x_i such that x_i denotes the probability mass associated to ρ_i. The set of probability distributions $\Delta(L(C)$ is then coordinatized by the rational points of the standard simplex.

Definition 3.6. *The* **standard simplex** *in* $\mathbb{R}^M$ *is the set*

$$\Delta_M := \left\{ x \in \mathbb{R}^n \mid \sum_i x_i = 1, x_i \geq 0 \text{ for } 1 \leq i \leq M \right\}.$$

We let $\Delta_M^{\mathbb{Q}} := \mathbb{Q}^M \cap \Delta_M$ *denote the rational points of* Δ_M.

Remark 3.7. *For simplicity we sometimes write* x_t *for the component of* $x \in \Delta_M$ *corresponding to* $t \in L(C)$ *(instead of* x_i *where* i *is the number of* t *in some linear ordering on* $L(C)$*).*

Example 3.8. *An election on candidates* $C = \{a, b, c\}$ *having* 7 *voters of whom* 3 *have preference abc,* 2 *have preference bac and* 2 *have preference cba corresponds (under the lexicographic order on* $L(C)$*) to the anonymous profile* $(3, 0, 2, 0, 0, 2)$ *and hence to the point* $(3/7, 0, 2/7, 0, 0, 2/7) \in \Delta_6$.

Remark 3.9. *We always consider* Δ_M *as embedded in* $\mathbb{R}_M$. *It is contained in a unique hyperplane* H_M.

We can interpret each anonymous and homogeneous rule as being defined on $\Delta_M^{\mathbb{Q}}$. We can then interpret consensus sets as subsets of Δ_M, which allows us to use our geometric intuition. For example, the domain of $\mathbf{S}$ consists of the corners of the simplex, while the domain of $\mathbf{W}$ also lies on the boundary of the simplex (for example, $\mathbf{W}_a$ does not contain any points with nonzero coordinates corresponding to rankings with a not at the top).

The general approach of the last paragraph has been adopted by many previous authors. For example, Saari [16] simply uses the terminology "profiles" to refer to vote distributions, and all rules he considers are simplex rules by definition. Furthermore, he assumes continuity, in order to define rules directly on the entire simplex, rather than just on rational points.

3.1. **Distance rationalization in the simplex.**

We shall see in Section 3.2 how anonymous and homogeneous distance rationalizable rules defined using profiles can be interpreted using the simplex. The converse idea is to define distance rationalizable rules directly on the simplex rather than on profile space. We make the obvious definitions by analogy with those for profiles.

Definition 3.10. *Given fixed candidate set* C *of size* m *and a distance on* Δ_M *where* $M = m!$, *a* **partial social rule** *on* Δ_M *of size* s *and domain* $D \subseteq \Delta_M^{\mathbb{Q}}$ *is a mapping taking each element of* D *to a nonempty subset of* $L_s(C)$. *A* **consensus** *on* Δ_M *is a partial social rule that is single-valued at every point (a partial social function). Given a consensus* K *and distance* δ *on* Δ_M*, the rule* $\mathcal{R}(K, \delta)$ *is defined by*

$$R(x) = \arg \min_{r \in L_s(C)} \delta(x, K_r). \tag{2}$$

The most obvious distances mathematically are surely the ℓ^p metrics. The interpretation in terms of social choice is less compelling for $p > 1$, since we are measuring the amount of effort needed to change one election into another by transferring vote mass under a nonlinear penalty. The case $p = 1$ is by far the most commonly studied, and also arises directly from votewise distances, unlike the case $p > 1$,

To our knowledge, several fairly obvious rules of this type have not yet been studied in detail. Here is an example using the unanimity consensus.

Example 3.11. *Fix* p *with* $1 \leq p \leq \infty$ *and consider the social choice rule* $\mathcal{R}(\mathbf{S}^s, \ell^p)$ *defined on* Δ_M. *We claim that, this chooses the s-ranking(s) which occur most often as the initial s-rankings of the votes. For* $s = 1$ *this is simply plurality rule, while for* $s = m$ *it is the* **modal ranking rule** *[2] (the term* **plurality ranking rule** *may seem more logical, but it is important not to be confused with the ranking induced by plurality scores of candidates in the case where these scores all differ).*

The proof is straightforward: given a distribution $(x_1, \dots, x_M)$*, by summing terms corresponding to the same initial s-ranking we reduce the argument to the case* $s = m$*. Let* e_i *be the ith basis vector in* $\mathbb{R}^M$*, a corner of the simplex. Then* $\delta^p(x, e_i) - \delta^p(x, e_j) = (1 - x_i)^p - (1 - x_j)^p + x_j^p - x_i^p$*. Choosing* x_i *to be the maximal value among the entries of* x *shows this to be nonpositive for all* $j \neq i$*, so that this value of* i *is the minimizer (the same argument works for* $p = \infty$ *with a different computation). Thus the rule returns precisely the most frequent initial s-ranking(s) from the input profile.*

Thus in order to find more interesting rules, we need to use less common distances on Δ_M. For example, so-called **statistical distances** such as the Kullback-Leibler and Hellinger distances are heavily used in many application areas. We do not present a detailed study here, deferring it to future work. Instead, we now move on to explore distances on Δ_M induced by distances on $\mathcal{E}$.

3.2. **Quotient distances.** We can define a simplex rule by first starting with profiles and passing to the quotient space, provided the rule in question is anonymous and homogeneous (the case of Dodgson's rule $\mathcal{R}(\mathbf{C}, d^1_K)$ which is anonymous but not homogeneous shows that care must be taken).

Since votewise distances are very natural and the ℓ^1 norm is the most obvious choice for a votewise distance (because it just adds the distance from each voter), we obtain several rules with an ℓ^1 flavour in this way. For example, the Hamming metric yields a constant multiple of ℓ^1 via the Wasserstein construction as described below, while the Kemeny metric and ℓ^1 lead to rules such as Borda and Kemeny's rule. We discuss ℓ^1-votewise rules in more detail in Section 5.

Definition 3.12. *A distance is* **anonymous** *if it satisfies the identity* $d(g(E), g(E')) = d(E, E')$*.*

An anonymous distance is **homogeneous** *if* $d(kE, kE') = d(E, E')$ *for all* $k \geq 1$*.*

Every anonymous and homogeneous rule R corresponds to a simplex rule $\overline{R}$. When $R = \mathcal{R}(\mathcal{K}, d)$ where both $\mathcal{K}$ and d are anonymous and homogeneous, we can express $\overline{R}$ as $\mathcal{R}(K, \delta)$ in a nice way. The mapping from profiles to the simplex yields the obvious consensus $\overline{\mathcal{K}}$. The distance is a little more involved. The obvious idea is to use a quotient distance [4]. This concept is relatively little-known.

Definition 3.13. *We define* $\overline{d} : \Delta_M \times \Delta_M \to \mathbb{R}_+$ *to be the* **quotient distance** *induced by* $\sim$*.*

Remark 3.14. *The standard construction of quotient distance* $\overline{d}$ *is as follows [4].*

$$\overline{d}(x, y) = \inf \sum_{i=1}^{k} d(E_i, E'_i) \tag{3}$$

where the infimum is taken over all admissible paths*, namely paths such that* $E'_i \sim E_{i+1}$ *for* $1 \leq i \leq k - 1$*,* $E = E_1, E' = E'_k$*,* E *projects to* x *and* E' *to* y*.*

In general, quotient distances are tricky to work with, owing to the complicated definition. In our setup, it turns out that they are reasonably tractable.

Proposition 3.15. *Let* d *be an anonymous and homogeneous distance on* $\mathcal{E}$*. Then* $\overline{d}$ *is given by*

$$\overline{d}(x, y) = \inf_{E, E'} d(E, E')$$

where E, E' *range over all elections having an equal number of voters, such that* $\mathcal{D}(E) = x, \mathcal{D}(E') = y$*.*

Proof. Let $x, y \in \mathcal{P}$ and consider an admissible path and its corresponding sum

$$\sum_{i=1}^{k} d(E_i, E'_i)$$

where $k > 1$. We show that we can reduce the value of k. By homogeneity of d, we can change the size of the voter sets so that E_k and E'_{k-1} have the same sized voter set. Then we can choose a permutation g of V^* taking E_k to E'_{k-1}. Since d is anonymous, using the triangle inequality we obtain

$$\begin{aligned} d(E_{k-1}, E'_{k-1}) + d(E_k, E'_k) &= d(E_{k-1}, E'_{k-1}) + d(g(E_k), g(E'_k)) \\ &= d(E_{k-1}, E'_{k-1}) + d(E'_{k-1}, g(E'_k)) \\ &\geq d(E_{k-1}, g(E'_k)). \end{aligned}$$

This gives an admissible path with k replaced by $k-1$.

Thus without loss of generality, in computing $\overline{d}$ we need only deal with the case $k = 1$. However, as above by homogeneity we can always assume in this case that E and E' have the same number of voters. □

Remark 3.16. *Proposition 3.15 shows that we can replace* inf *by* min*, since by anonymity, once we fix the number of voters there are only a finite number of possible distances* $d(E, E')$ *to consider.*

We know [7] that given an anonymous and homogeneous distance d, if $\mathcal{R}(\mathcal{K}, d)$ is homogeneous then $\mathcal{R}(\mathcal{K}, d)$ corresponds exactly to the simplex rule $\mathcal{R}(\overline{\mathcal{K}}, \overline{d})$. Note that Dodgson's rule which is not homogeneous, shows that we cannot omit the qualification "if $\mathcal{R}(\mathcal{K}, d)$ is homogeneous".

3.3. **The ℓ^p-votewise case - Wasserstein distance.** In the special case of ℓ^p-votewise distances, we can describe $\overline{d}$ in more detail to a well-known construction from probability theory, the Wasserstein distance, which we now recall.

Let S be a finite set of size M, and let $\Delta(S)$ denote the set of probability distributions on S. For a distance d defined on S, the function $d_W^p : \Delta(S) \times \Delta(S) \to \mathbb{R}$ is defined by

$$d_W^p(x, y)^p = \inf_A \sum_{r, r' \in S} A_{r,r'} d(r, r')^p,$$

where the infimum is taken over all *couplings* of x and y, defined as nonnegative square matrices of size M whose marginals are x and y respectively (i.e. $\forall r, \sum_{r'} A_{rr'} = x_r$ and $\forall r', \sum_r A_{rr'} = y_{r'}$). Basically, it represents the minimum cost to move from one configuration to another, where the underlying distance d defines the cost of each movement. Indeed, this construction leads to a new distance.

Proposition 3.17. *If d is a distance on S, then d_W^p is a distance on $\Delta(S)$. If d is a metric, then so is d_W^p.*

Proof. See [19, Ch. 6]. □

Remark 3.18. *The function d_W^p goes by several names, some common ones being the* **l^p-transportation** *distance, the* **Kantorovich p-distance***, the* **p-Wasserstein distance***. When $p = 1$, it is also called the* **Earth Mover's** *distance or* **first Mallows metric***, and is used heavily in several areas of computer science, particularly image retrieval and pattern recognition.*

Now we are able to make the link with the votewise metrics, by applying the construction of d_W^p in the case $S = L(C)$. Scaling a votewise distance based on ℓ^p-norm gives a homogeneous distance with a special formula, and this turns out to be exactly the p-Wasserstein distance.

Remark 3.19. *A votewise distance based on ℓ^p is not homogeneous, as its value depends on the number of voters. However, we may make an equivalent homogeneous version by scaling. We define a new norm*

$$||x||_p^* = \frac{1}{n} \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

for $1 \leq p < \infty$, and analogously for $p = \infty$. Letting the votewise distance corresponding to this norm be denoted d_*^p, we see that $\mathcal{R}(\mathcal{K}, d^p) = \mathcal{R}(\mathcal{K}, d_*^p)$.

Proposition 3.20. *Let d be a distance on $L(C)$. Then $\overline{d_*^p} = d_W^p$.*

Proof. Since d_*^p is homogeneous, $\overline{d_*^p}(x,y) = \min_{E'} d_*^p(E,E')$, where $\overline{E} = x$, $\overline{E'} = y$ and $|V(E)| = |V(E')| = n$, say. Let $E = (C,V,\pi)$ and $E' = (C,V,\pi')$. Then

$$\overline{d_*^p}(x,y)^p = \min_{\pi'} \frac{1}{n} \sum_i d(\pi_i, \pi'_i)^p = \min_{\pi'} \frac{1}{n} \sum_{\substack{i \\ \pi_i = r, \pi'_i = r'}} d(r,r')^p = \min_a \frac{1}{n} \sum_{r,r' \in S} a_{r,r'} d(r,r')^p,$$

where the $a_{r,r'}$'s are integers such that for all $r \in S, \sum_r \frac{a_{r,r'}}{n} = x_r$ and for all $r' \in S, \sum_{r'} \frac{a_{r,r'}}{n} = y_{r'}$, which corresponds to the Wasserstein distance restricted to matrices A respecting the conditions and with coefficients of the form $\frac{k}{n}$ with $0 \leq k \leq n$. So clearly, $d_W^p(x,y) \leq \overline{d_*^p}(x,y)$.

Let assume that this inequality is strict: let $\epsilon > 0$ and A be such that $\sum_{r,r' \in S} A_{r,r'} d(r,r')^p \leq d_W^p(x,y) + \epsilon/2$. Since $\max_{r,r'} d(r,r') < \infty$ and we can choose n as big as we want, A can be approximated arbitrarily close by a matrix A' of the previous form: $\sum_{r,r' \in S} A'_{r,r'} d(r,r')^p \leq d_W^p(x,y) + \epsilon$. We can get arbitrarily close to $d_W^p(x,y)$ in this way, which proves the equality. □

Example 3.21. *Let d_1 be the ℓ^1 distance between probability measures on $L(C)$ and let $d = d_H$. Then $\overline{d_*^1} = \frac{1}{2} d_1$ (also called the* **total variation distance***). This was observed (without the current notation) in [14, Lemma 3.2]: if E, E' are elections on (C,V) with $|V| = n$, then*

$$d_1(\overline{E}, \overline{E'}) \leq \frac{2}{n} d_H(E, E')$$

and given $\overline{E}, \overline{E'} \in \Delta_M^{\mathbb{Q}}$, we can choose n and the preimages E, E' so that equality holds.

Example 3.22. *If $x \in \mathcal{V}$ has 2 abc voters and 3 bac voters, while y has 2 bac voters and 3 cba voters, the quotient distance corresponding to the normalized version of d_H^1 is $3/5$. Note that for the Kemeny metric the analogous quantity is $8/5$.*

4. Tied sets and decisiveness

All social rules used in practice encounter the problem of breaking ties. However, many commonly used social rules have the property that the tied subset is “small” (for example asymptotically negligible as $n \to \infty$, for fixed m). This is important, because if the tied region is small, then ties can be ignored for many purposes, whereas if the tied region is asymptotically large, our rule may suffer extreme lack of decisiveness.

In the worst case, the rule may do nothing, and simply return all possible s-rankings at every profile, making it useless. In the DR framework, this extreme indecision cannot occur, because some consensus set must be nonempty. However, it is certainly possible to have “large” subsets of profile space on which a DR rule is not single-valued. We investigate this question for simplex rules, giving both positive (few ties) and negative (many ties) results.

4.1. **Boundaries.**

Definition 4.1. *The* **boundary** *of the social rule $\mathcal{R}(\mathcal{K}, d)$ of size s is the set of all elections at which the minimum in (1) is attained for at least two distinct $r \in L_s(C)$. The boundary of the simplex rule $\mathcal{R}(K, \delta)$ on Δ_M is the set of all elements of Δ_M for which the minimum in (2) is attained for at least two distinct $r \in L_s(C)$.*

Example 4.2. *Suppose that $m = 2$, with alternatives a and b, $\mathcal{K} = \mathbf{C}$, and d is an anonymous neutral standard distance. The concepts of majority winner and Condorcet winner coincide when $m = 2$. When n is odd, the boundary $\mathcal{R}(\mathcal{K}, d)$ is empty, whereas when n is even, the elections having an equal number of ab and ba voters are in the boundary.*

On the other hand, for the simplex rule $\mathcal{R}(\overline{\mathbf{C}}, \overline{d})$, the boundary is the point $(1/2, 1/2)$.

Our intuition is that the boundary of a well-behaved simplex rule should be "small" in Δ_M, and be geometrically "nice". The relevant geometric theory is that of *Voronoi diagrams*. We now digress to review some known results, which will help develop a more refined intuition.

4.2. **Geometric background.** Voronoi theory is usually defined for a metric space, and most commonly for a **Minkowski space** (a finite-dimensional real vector space equipped with a norm). All definitions below work for an arbitrary metric space. The theory can no doubt be generalized to hemimetrics, but we do not deal with maximum generality here. Our main interest is in explaining that there are several interesting reasons why DR rules defined by ℓ^p-votewise distances may fail to be decisive.

For a fixed set of of **sites** (subsets of the entire space), the open **Voronoi cell** of each site X is defined as the set of points closer to X than to any other site. The boundaries of these cells are contained in the union of bisectors, where a **bisector** of the sites X, Y, denoted $\beta(X, Y)$, is defined to be the set of points equidistant from those two sites.

Interpreting the sites as the consensus sets $\mathcal{K}_r$, we see that the open Voronoi cell corresponding to $\mathcal{K}_r$ is precisely the set on which $\mathcal{R}(\mathcal{K}, d)$ is single-valued with value r. Also, the boundary as defined above is just the union of boundaries of all Voronoi cells.

We first discuss bisectors, because if these are well-behaved, so will the boundary of the rule be. We first restrict to the nicest situation (where our intuition is greatest), namely where sites are single distinct points X and Y in $\mathbb{R}^n$ under the Euclidean ℓ^2-norm. In this case $\beta(X, Y)$ is a hyperplane normal to the line joining the points, and the Voronoi cells are therefore convex polyhedra that tile the entire space.

However, this situation is rather special. In fact $\beta(X, Y)$ is a hyperplane for all X and Y if and only if the space is ℓ^2 [12]. Thus we should expect to see bisectors that are not hyperplanes. Of course, such bisectors may still be well-behaved, for example smooth hypersurfaces. Note that $\beta(X, Y)$ is known to be homeomorphic to a hyperplane provided the norm is strictly convex (recall that a norm is **strictly convex** if its unit ball contains no line segment) and even sometimes when it is not [8]. This does not preclude nasty behaviour such as two bisectors intersecting infinitely often, but for norms defined algebraically, such as ℓ^p, this does not happen.

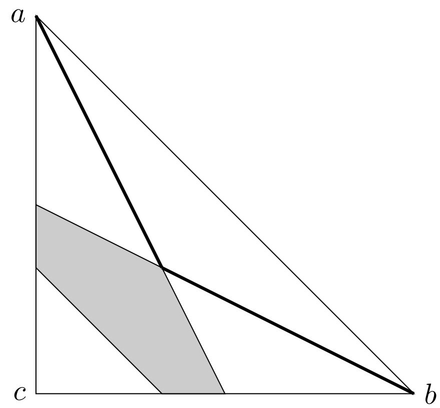
When the sites are not single points (in particular when they are not separated), bisectors may poorly behaved even in ℓ^2 (see Example 4.4). We conclude that well-behaved bisectors should not be expected in general, and we explore this in the next section.

4.3. **Large boundary.** The most obvious way for a rule to be rather indecisive is if the underlying distance does not distinguish points well. We say that a subset of a Minkowski space (possibly defined with a seminorm) is **large** if it contains an open ball, and **small** otherwise.

Example 4.3. *(pseudometric) Consider the Copeland rule whose boundary contains all points with no unique Copeland winner. This contains in particular the set where all candidates have the same Copeland score, for example because the majority tournament is a single cycle. This is a large subset of Δ_M. To see this, note that in terms of coordinates, the single cycle is described by $\binom{m}{2}$ equations of the form $\sum_{i \in S_{ab}} x_i > \sum_{i \in S_{ba}} x_i$, where S_{ab} denotes the set of rankings for which a is above b. The dimension of the simplex is $(m!-1)$, and so for $m \geq 3$ this subset is large (* a priori *it could even be empty, but we know it is not). The tied set contains a sufficently small neighborhood of every point for which all inequalities are strict, because small changes to the proportions of voter types will not change the majority tournament.*

In view of this example, we should require our distances to be quasimetrics. However, there are other more subtle problems that can occur. Note that in the next example, the distance is ℓ^2 and is hence as nice as could be expected: a metric induced by a norm that is strictly convex, symmetric, and algebraically defined. The problem is that the consensus notion is wrong – intuitively, consensus sets should be separated, because otherwise how could the consensus choice be uncontroversial?

FIGURE 4.1. Illustration for Example 4.4.



Example 4.4. *(non-separated consensus) Let $m = 3$ and consider Δ_6 with the usual ℓ^2 metric (induced from $\mathbb{R}^5$ which coordinatizes H_6 in the usual way).*

Consider the line segments L_1, L_2 that join the centre P of Δ_6 to the points $x_{abc} = 1$ and $x_{bca} = 1$ respectively (Figure 4.1 gives some intuition in lower dimension). Define a 1-consensus by letting $\mathcal{K}_a = L_1, \mathcal{K}_b = L_2$ and $\mathcal{K}_c$ be the single point $x_{cab} = 1$. Let H_1, H_2 be the hyperplanes (in Δ_6, i.e. having dimension 4) normal to L_1, L_2 at P. Let S be the set of points in Δ_6 that lie on the other side of H_1 from L_1 and on the other side of H_2 from L_2. Then each point of S lies on the bisector of L_1, L_2 with respect to the usual ℓ^2-metric d, because the closest point of L_1 is P and this is the closest point of L_2. Every point of S that is closer to the centre of Δ_6 than to $\mathcal{K}_c$ is in the boundary of the social choice rule $\mathcal{R}(\mathcal{K}, d)$, which is therefore large. Note that $\mathcal{K}$ satisfies anonymity and homogeneity, but not neutrality. Furthermore, every $\mathcal{K}_r$ is a convex polyhedral subset of the simplex.

If instead we define $\mathcal{K}$ symmetrically by letting $\mathcal{K}_c$ be the line segment L_3 joining the point $x_{cab} = 1$ to the centre of the simplex, then although the bisectors are large on Δ_6, the boundary of the rule is small. This is because points in S are now closer to L_3 than either L_1 or L_2. Of course, if we defined $\mathcal{K}$ in this way as an s-consensus, with abc, bca, cab as the possible outputs, then points of S will still yield a large boundary.

Also note that if the line segments did not meet at any point, the bisectors would all be small.

Thus we should require that consensus sets be separated. However, there is another common way in which bisectors can fail to behave well, which is when the underlying norm in a Minkowski space is not strictly convex. We now analyse a special case of this in some detail.

5. Analysis of ℓ^1-votewise metrics

Votewise distances based on the ℓ^1 norm are very commonly used. They are computationally easy and have a clear interpretation in terms of adding distances corresponding to each voter. In fact we are not aware of a named ℓ^p-votewise rule that has been defined for any $p \neq 1$. However, when we consider decisiveness, there are some potential negative consequences to using the ℓ^1 norm.

We first show in Proposition 5.1 that all ℓ^1-votewise metrics correspond to a (Wasserstein) distance on the simplex that is induced by a norm that is not strictly convex.

Let $c = (1, 1, \dots, 1)/M \in \mathbb{R}^M$ be the center of the simplex. Now, we translate the center c to the origin, and we denote by Δ' the image of the simplex under this translation. We denote by $\mathcal{H}$ the hyperplane containing Δ'. Our study of the geometry under the Wasserstein distance will be facilitated by the following observation.

Proposition 5.1. *Let d be an ℓ^1-votewise distance. Then*

(i) d induces a norm N on $\mathcal{H}$. Explicitly, for each $x \in \Delta'$, we have $N(x) = d^1_W(x + c, c)$;

(ii) the unit ball of the norm N is not strictly convex.

Proof. Part (i) is a well known property of the Wasserstein 1-distance (the norm is called the Kantorovich-Rubinstein norm [19, Ch 6]). For (ii), fix a ranking $r \in L(C)$ and consider the subset S_r of all points x where only the component corresponding to r is negative. In S_r, we have

$$N(x) = \sum_{r'} x_{r'} d_W^1(r', r).$$

Thus the intersection of S_r with the unit sphere is contained in a hyperplane and hence not strictly convex. □

Remark 5.2. *This result is not true for the other Wasserstein metrics, because they do not satisfy the homogeneity property of norms. For example, if $x \in \Delta'$ is such that $x_r \geq 0$ and $\forall r \neq r', x_{r'} \leq 0$, then $d_W^p(x+c,c)^p = \sum_{r'} |x_{r'}| d(r,r')^p$, and $d_W^p(\lambda x + c, c) = |\lambda|^{\frac{1}{p}} \sum_{r'} |x_{r'}| d(r,r')^p = |\lambda|^{\frac{1}{p}} d_W^p(x+c,c)$.*

We can now show that for any distance d, there are points whose bisector under the Wasserstein distance d_W^1 is large.

Proposition 5.3. *Consider a norm N induced over $\mathcal{H}$ by an ℓ^1-votewise metric. Let r, r_1, r_2 be rankings. We denote by d_1 and d_2 the distances $d(r, r_1)$ and $d(r, r_2)$. Let $\epsilon > 0$. We define x and y as the two points of $\mathcal{H}$ such that $x_r = -x_{r_1} = \frac{\epsilon}{d_1}$, $y_r = -y_{r_2} = \frac{\epsilon}{d_2}$ and all other components are equal to zero. Then, any point $z \in \mathcal{H}$ such that $z_r \geq 1 - \frac{\epsilon}{\min(d_1,d_2)}$ is equidistant from x and y according to N.*

Proof. Let z be such a point. Then, $x - z$ and $y - z$ have only one negative component: the one corresponding to the ranking r. So $N(x-z) = \sum_{r'} (x_{r'} - z_{r'}) d(r, r')$ and $N(y - z) = \sum_{r'} (y_{r'} - z_{r'}) d(r, r')$. Since the only components where x and y differ are r_1 and r_2, they are equidistant of z if and only if $(x_{r_1} - z_{r_1}) d_1 + (x_{r_2} - z_{r_2}) d_2 = (y_{r_1} - z_{r_1}) d_1 + (y_{r_2} - z_{r_2}) d_2$, which is equivalent to $x_{r_1} d_1 + x_{r_2} d_2 = y_{r_1} d_1 + y_{r_2} d_2$, which is in turn equivalent to $x_{r_1} d_1 = y_{r_2} d_2$, which is always true by definition. □

It follows that the behaviour of ℓ^1-votewise distances is rather counterintuitive.

Corollary 5.4. *Let d be an ℓ^1-votewise metric. Then there is a consensus $\mathcal{K}$ consisting of isolated points, such that the boundary of $\mathcal{R}(\mathcal{K}, d)$ is large.*

Proof. Write x_ϵ and y_ϵ for points x and y of the form defined in Proposition 5.3. That proposition implies that, by setting $\mathcal{K}_a = \{x_\epsilon\}$ and $\mathcal{K}_b = \{y_\epsilon\}$ and choosing a sufficiently small ϵ, then $\beta(\mathcal{K}_a, \mathcal{K}_b)$ will be large. Also, for any other candidate c, if we set $\mathcal{K}_c = \{x_{\epsilon'}\}$ with $\epsilon < \epsilon'$, then for any z such that $z_r \geq 1 - \frac{\epsilon}{\min(d_1,d_2)}$, $N(z, \mathcal{K}_a) = N(z, \mathcal{K}_b) < N(z, \mathcal{K}_c)$. □

Remark 5.5. *Note that the consensus in the proof of Corollary 5.4 is somewhat unnatural. For example, it is not neutral and does not intersect the boundary of Δ_M.*

The next question is how often this kind of situation happens. For simplicity we focus on the case $d = d_H$, when the induced norm is exactly ℓ^1. We can give an exact characterization of when two points have a large bisector. This is directly connected with the well-known integer partition problem.

Proposition 5.6. *Let $M \geq 1$ and let $x, y \in \mathbb{R}^M$. We denote by S the set of values $(x_i - y_i)$. Then x and y have a large bisector under ℓ^1 if and only if there exists a subset $S' \subset S$ such that $\sum_{e \in S'} e = \sum_{e \notin S'} e$.*

Proof. By definition $\beta(x,y) = \{z | \sum_i |x_i - z_i| = \sum_i |y_i - z_i|\}$. We divide $\mathbb{R}^M$ into 4^M subspaces corresponding to the possible signs of the the values $(x_i - z_i)$ and $(y_i - z_i)$. Let V be one of these subspaces: in V, the equality $\sum_i |x_i - z_i| = \sum_i |y_i - z_i|$ is equivalent to $\sum_i \epsilon_i (x_i - z_i) = \sum_i \epsilon'_i (y_i - z_i)$, where $\forall i, \epsilon_i, \epsilon'_i = \pm 1$. This is equivalent to $\sum_i (\epsilon_i - \epsilon'_i) z_i = \sum_i (\epsilon_i x_i - \epsilon'_i y_i)$.

There are two cases. First, if for some i,$\epsilon_i \neq \epsilon'_i$, then the linear equation in z is nontrivial and z lies in a hyperplane, so that $V \cap \beta(x, y)$ is small. The other case is when $\epsilon_i = \epsilon'_i$ for all i, in which case the left side of the equation is 0. If the right side is nonzero there is no solution, and $V \cap \beta(x, y) = \emptyset$. If the right side is zero, then $V \cap \beta(x, y)$ is large (for each i, it contains all points for which z_i is sufficiently large, for example). The right side is zero if and only if $\sum_i \epsilon_i(x_i - y_i) = 0$, which is equivalent to the fact that there exists $S' \subset S, \sum_{e \in S'} e = \sum_{e \notin S'} e$. □

The argument in the proof gives insight into the shape of any large bisector of two points: any ball included in the bisector is contained in cells where $(x_i - z_i)$ and $(y_i - z_i)$ are of the same sign, and thus in a subset defined by a set of equations $z_i \leq \min(x_i, y_i)$ or $z_i \geq \max(x_i, y_i)$ for all i. It implies, for example, that if the points are corners of the simplex, the large bisector in $\mathbb{R}^M$ intersects Δ_M in a small set. Thus, for example, large boundaries cannot occur with **S** (which also follows from Corollary 6.4 below).

Definition 5.7. *The standard decision problem PARTITION is defined as follows. Input is a vector $(x_1, \ldots, x_M)$ of nonegative integers. We must decide whether there is a subset $S \subseteq \{1, \ldots, M\}$ for which $\sum_{i \in S} x_i = \sum_{i \notin S} x_i$.*

Define the decision problem LARGE-BISECTOR as follows. Input is a pair (x, y) of points of $\mathbb{Q}_+^M$ (with coordinates expressed in lowest terms) and we must decide whether $\beta(x, y)$ contains an open ball under the ℓ^1 metric.

Remark 5.8. *Note that M is part of the input in each case. When M is bounded, PARTITION can be solved trivially by exhaustive enumeration of subsets. Note that if we let $K := \sum_i x_i$, then a standard dynamic programming algorithm solves PARTITION in $O(KM)$ time.*

Proposition 5.9. *LARGE-BISECTOR is NP-complete, and so the analogous decision problem for rational points of Δ_M is NP-hard.*

Proof. We use Proposition 5.6. Given an instance $(z_1, \ldots, z_M)$ of PARTITION, let $x_i = z_i, y_i = 0$. This gives an instance of LARGE-BISECTOR, which is a yes instance if and only if the original instance is a yes instance of PARTITION. Given an instance (x, y, M) of LARGE-BISECTOR, let $z_i = x_i - y_i$, and clear denominators, giving an instance of the known NP-complete problem PARTITION. Thus LARGE-BISECTOR is NP-complete since there is a polynomial-reduction each way between it and PARTITION. □

We suspect the analogue of LARGE-BISECTOR to be NP-hard for every l^1-votewise metric.

The situation is quite subtle, because large bisectors do not occur when the consensus sets are hyperplanes instead of points.

Proposition 5.10. *The bisector of two distinct hyperplanes under any norm on $\mathbb{R}^n$ is contained in a union of at most two hyperplanes.*

Proof. The distance from a point x to a hyperplane H defined by $a^T x = b$ is equal to $d(x, H) = \frac{|a^T x - b|}{||a||^*}$ where * denotes the dual norm (see for example [11]; the exact definition is not necessary here). Now, let H' be another hyperplane defined by the equation $a' x = b'$. We assume that $||a||^* = ||a'||^*$ (without loss of generality since multiplying by a scalar still defines the same hyperplane). The bisector of H and H' can be defined as the set of points x verifying $|a^T x - b| = |a'^T x - b'|$. So, we have two cases, depending on the sign of these absolute values: either $\sum_i (a_i - a'_i) x_i = b - b'$ or $\sum_i (a_i + a'_i) x_i = b + b'$. Since $H \neq H'$, each of these is the equation of a hyperplane. □

6. Small bisectors and hyperplane rules

All our results in this section show that the bisectors in question are contained in a finite union of hyperplanes. Rules which have a well-defined winner on each component of the complement in Δ_M of a finite set of hyperplanes have been studied recently. Mossel, Procaccia and Racz [14] call such simplex rules **hyperplane rules** and show their equivalence with the **generalized scoring rules** of Xia and Conitzer [21]. These rules can be defined axiomatically using **finite local consistency** [20]. Although originally introduced for social choice rules only, the definition extends to social welfare rules [2] and it is clear that it also extends to general s.

Most rules that have ever been studied by social choice theorists are hyperplane rules. A notable exception is Copeland's rule. In order to interpret Copeland's rule as a hyperplane rule, Mossel, Procaccia and Racz [14] require that the winner be (arbitrarily) specified on the tied region. This seems to us to be stretching the definition too far – we could do the same thing for any indecisive rule.

We now give a sufficient condition for a DR rule to be a hyperplane rule. We first recall some facts proved in [7].

Definition 6.1. *Let $\mathcal{K}$ be a homogeneous consensus and d a homogeneous distance. Say that $(\mathcal{K}, d)$ has the* **homogeneous minimizer property** *(HMP) if for each $E, E' \in \mathcal{E}$ with $E \sim E'$ and each r, $d(E, \mathcal{K}_r) = d(E', \mathcal{K}_r)$.*

Say that $(\mathcal{K}, d)$ satisfies the **votewise minimizer property** *(VMP) if the following condition is satisfied.*

> *For each $r \in L_s(C)$ and each election $E = (C, V, \pi) \in \mathcal{E}$, there exists a minimizer $(C, V, \pi^*) \in \mathcal{K}_r$ of the distance from E to $\mathcal{K}_r$, so that for all $i, d(\pi_i, \pi_i^*)$ depends only on π_i and r.*

We collect some basic results [7]. The proofs are immediate consequences of the definitions.

Proposition 6.2. *The following results hold.*

(i) If $(\mathcal{K}, d)$ satisfies the HMP then $d(E, \mathcal{K}_r)$ has the form $N(S)$ where S is the multiset of values of $d(\pi_i, \pi_i^*)$ occurring.
(ii) If $(\mathcal{K}, d)$ satisfies the VMP then it satisfies the HMP.
(iii) If $\mathcal{K}$ is an s-consensus, and for each $r \in L_s(C)$, there is a nonempty subset S_r of $L(C)$ such that $\mathcal{K}_r$ consists precisely of the elections for which no voter has a ranking in S_r, then $(\mathcal{K}, d)$ satisfies the VMP for every votewise d.
(iv) $\mathbf{S}^s$ satisfies the condition in (iii).

□

Proposition 6.3. *Let d be ℓ^p-votewise for some $1 \leq p < \infty$ and suppose that $(\mathcal{K}, d)$ satisfies the VMP. Then on Δ_M, $\beta(\mathcal{K}_r, \mathcal{K}_{r'})$ is defined by*

$$\sum_{t \in L(C)} x_t \delta(t, r)^p = \sum_{t \in L(C)} x_t \delta(t, r')^p.$$

Proof. We know that the distance between $E = (C, V, \pi)$ and the minimizer $m(E, r) = (C, V, \pi^*)$ equals $N(S)$ where S is the multiset with entries $\delta(t, r)$ occurring according to their multiplicities nx_t, for all $t \in L(C)$. The specific form of N then shows that $d(E, m(E, r))^p = n\left(\sum_t x_t \delta(t, r)^p\right)$. Applying the same argument for r' yields the result. □

Corollary 6.4. *Suppose that d is ℓ^p-votewise with $1 \leq p < \infty$, d is finite and not identically zero, and $\mathcal{K}$ satisfies condition (iii) of Proposition 6.2. Then $\mathcal{R}(\mathcal{K}, d)$ is a hyperplane rule.*

Proof. Since $d < \infty$ we may rearrange the formula in Proposition 6.3 to get $\sum_t (\delta(t, r) - \delta(t, r')) = 0$. It suffices to show that the linear function on the left side is not identically zero. That could only happen if $\delta(t, r) = \delta(t, r')$ for all t. However, letting the distance from x to $\mathcal{K}_r$ is attained at a point $m(x, r)$ where $m(x, r)_t = x_t$ for all $t \notin S$, and $d(x, \mathcal{K}_r) = \sum_{t \in S} x_t \delta(t, r)^p$. If $r \neq r'$ then by definition $S_r \neq S_{r'}$. Thus taking $t \notin S \cap S'$, without loss of generality $\delta(t, r) = 0$ and $\delta(t, r') \neq 0$. □

Corollary 6.5. *Every rule of the form $\mathcal{R}(\mathbf{S}^s, d^p)$, where $1 \leq p < \infty$ and d is a distance on $L(C)$ that is neither infinite nor identically zero, is a hyperplane rule.*

Remark 6.6. *This result does not extend to general distances. For example, Copeland's rule as we have defined it is not a hyperplane rule, yet it can be defined as* $\mathcal{R}(\mathbf{W}, d_{RT})$*. Also note that when* $p = \infty$*, we do not obtain a hyperplane rule. For example, every point* $x \in \Delta_M^{\mathbb{Q}}$ *for which every coordinate is nonzero is equidistant from all* $\mathbf{S}_r$*, so* $\mathcal{R}(\mathbf{S}, d^\infty)$ *is almost maximally indecisive.*

Remark 6.7. *Rules of the type described in Proposition 6.3 are rather special. Since the distance to* $\mathcal{K}_r$ *is of the form* $\sum_t x_t \delta(t, r)^p$*, each can be thought of as a differently weighted version of the rule with* $p = 1$*.*

7. Discussion and future work

We now summarize what we have learned about the boundary of a DR simplex rule.

- Using a pseudometric that is not a metric can easily lead to a large boundary.
- Large bisectors can occur even with ℓ^2, if consensus sets are not separated.
- Large bisectors can occur with ℓ^1-votewise rules, even for consensus sets that are isolated points, and it can be difficult to determine whether they do occur.
- Even when bisectors are large in the ambient space, using consensus sets on the boundary of the simplex often yields small bisectors on the simplex.
- Even when bisectors are large on the simplex, neutrality often makes the boundary of the rule small.

We have seen some desirable properties of consensus sets, such as homogeneity and neutrality. We argue that convexity (defined in the usual way via restriction from $\mathbb{R}^M$) of each $\mathcal{K}_r$ is another essential condition. In the following example, it seems ridiculous that a should win at the extra point.

Example 7.1. *Consider the case* $m = 3$*, and the consensus formed by extending* $\mathbf{W}$ *so that* a *is the winner whenever* $x_{bca} = x_{cba} = 1/2$ *(and similarly for* b, c*). This consensus is anonymous and homogeneous, but* $\mathcal{K}_a, \mathcal{K}_b, \mathcal{K}_c$ *are not convex.*

Remark 7.2. *In the simplex model, convexity (over* $\mathbb{Q}$*) is equivalent to the notion of* **consistency** *: if we split the voter set into two parts each of which elects* r*, the original voter set should elect* r*. It rules out the above example. Note that* $\mathbf{C}$ *and* $\mathbf{S}^s$ *are convex. In fact we do not know of a consensus that has been used in the literature that is not convex.*

Based on the above results, we suggest that the following criteria be required of consensus classes in the simplex (anonymity and homogeneity come for free)

- neutrality
- convexity
- separation
- intersecting the boundary of the simplex

while distances should be required to be metrics.

Note that the separation requirement rules out $\mathbf{C}$ as a consensus notion. This may of course be controversial. It may turn out that neutral rules based on $\mathbf{C}$ and using metrics do have small boundary (we do not know of a counterexample, but have no proof yet). However, it seems strange to consider a situation arbitrarily close to a complete tie among all rankings (the centre of the simplex) to be an election on which a "consensus" can be formed.

We saw above that ℓ^1 votewise distances can lead to major problems with decisiveness. However there are many natural examples of ℓ^1 votewise distances, as we have seen. We do not know of any "natural" simplex rule satisfying the above requirements for which the boundary is large. However, not all obvious rules have been thoroughly explored.

Systematic exploration of rules $\mathcal{R}(\mathcal{K}, d)$, where $\mathcal{K}$ and d satisfy the recommendations above, may prove fruitful in finding new rules with desirable properties. For example, by the results in this paper and [7], the rules $\mathcal{R}(\mathbf{S}^s, d^p)$ where d is a neutral metric on permutations, are anonymous, homogeneous, neutral, continuous, hyperplane rules. There are many neutral (also called right-invariant) metrics on permutations we have not discussed here, such as the ℓ^q-metrics (the cases $q = 1, 2, \infty$ being called Spearman's footrule, Spearman's rank correlation and the maximum displacement distance), and the Lee distance [4]. Even the rules $\mathcal{R}(\mathbf{W}, d^p)$ and $\mathcal{R}(\mathbf{S}, d^p)$ have not been fully explored, to our knowledge.

Even less understood are rules of the form $\mathcal{R}(\mathbf{C}, \ell^p)$. For example, when $p = 1$, we obtain a homogeneous version of the recently described *Voter Replacement Rule* [6]. Little is known about the Voter Replacement Rule other than that it is not homogeneous [7].

Beyond the realm of votewise and ℓ^p distances, we have already mentioned more general statistical distances. Finally, rules involving various matrix norms on the tournament matrices have not been well studied.

Distance-based aggregation of preferences is a more general procedure than we have studied here: it could be applied with many different input and output spaces [25]. If the input consists of the tournament matrix rather than the profile, there is a natural hypercube representation of the input in $\binom{m}{2}$ dimensions. Saari & Merlin [18] showed that the Kemeny rule can be described in this way using distance rationalization with respect to the ℓ^1 norm and $\mathbf{S}$. This is the same as using an elementwise norm on the weighted tournament matrix, in our framework. When using profiles as input, the simplex geometry is hard enough to visualize that some authors have used a fixed projection to the *permutahedron* and essentially used $\mathbf{S}$ as a consensus. The cases $p = 2$ (*mean proximity rules*) [24, 9] and $p = 1$ (*mediancenter rules*) [3] have received attention. These can be interpreted in our framework by changing the distance — detailed formulae might be interesting.

A question which partially motivated the present work remains unanswered. Does (a homogenization of) Dodgson's rule have a "small and nice" boundary? What about other Condorcet rules $\mathcal{R}(\mathbf{C}, d)$ where d is a votewise metric, or even rules based on d_T, such as the maximin rule?

References

[1] D.E. Campbell and S.I. Nitzan. "Social compromise and social metrics". English. *Social Choice and Welfare* 3. (1986), 1–16.

[2] Ioannis Caragiannis, Ariel D Procaccia, and Nisarg Shah. "Modal ranking: A uniquely robust voting rule". In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI).* 2014, pp. 616–622.

[3] Davide P Cervone et al. "Voting with rubber bands, weights, and strings". *Mathematical Social Sciences* 64. (2012), 11–27.

[4] Michel Marie Deza and Elena Deza. *Encyclopedia of Distances.* 00842. Springer, 2009, pp. 1–583.

[5] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. "Distance rationalization of voting rules". *Social Choice and Welfare* 45. (2015), 345–377.

[6] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. "Rationalizations of Condorcet-consistent rules via distances of hamming type". *Social Choice and Welfare* 39. (2012), 891–905.

[7] Benjamin Hadjibeyli and Mark C. Wilson. "Distance rationalization of social rules". *submitted* (2016).

[8] A.G. Horvath. "On bisectors in Minkowski normed spaces". *Acta Mathematica Hungarica* 89 (2000), 233–246.

[9] Sébastien Lahaie and Nisarg Shah. "Neutrality and geometry of mean voting". In: *Proceedings of the fifteenth ACM conference on Economics and computation.* ACM. 2014, pp. 333–350.

[10] Ehud Lerer and Shmuel Nitzan. "Some general results on the metric rationalization for social decision rules". *Journal of Economic Theory* 37. (1985). 00023, 191–201.

[11] O. L. Mangasarian. "Arbitrary-Norm Separating Plane". *Operations Research Letters* 24 (1997), 15–23.

[12] Heinrich Mann. “Untersuchungen über Wabenzellen bei allgemeiner Minkowskischer Metrik”. *Monatshefte für Mathematik* 42. (1935), 417–424.

[13] Tommi Meskanen and Hannu Nurmi. “Closeness Counts in Social Choice”. English. In: *Power, Freedom, and Voting*. Ed. by Matthew Braham and Frank Steffen. Springer Berlin Heidelberg, 2008, 289–306.

[14] Elchanan Mossel, Ariel D. Procaccia, and Miklós Z. Rácz. “A smooth transition from powerlessness to absolute power”. *Journal of Artificial Intelligence Research* (2013), 923–951.

[15] Shmuel Nitzan. “Some measures of closeness to unanimity and their implications”. English. *Theory and Decision* 13. (1981), 129–138.

[16] Donald G. Saari. *Basic geometry of voting*. Springer-Verlag, Berlin, 1995, pp. xii+300.

[17] Donald G. Saari. *Geometry of voting*. Vol. 3. Studies in Economic Theory. Springer-Verlag, Berlin, 1994, pp. xvi+372.

[18] Donald G. Saari and Vincent R. Merlin. “Changes that cause changes”. *Soc. Choice Welf.* 17. (2000), 691–705.

[19] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media, 2008.

[20] Lirong Xia and Vincent Conitzer. “Finite Local Consistency Characterizes Generalized Scoring Rules”. In: *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*. Ed. by Craig Boutilier. 2009, pp. 336–341.

[21] Lirong Xia and Vincent Conitzer. “Generalized scoring rules and the frequency of coalitional manipulability”. In: *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008), Chicago, IL, USA, June 8-12, 2008*. Ed. by Lance Fortnow, John Riedl, and Tuomas Sandholm. ACM, 2008, pp. 109–118.

[22] H. Young. “Social Choice Scoring Functions”. *SIAM Journal on Applied Mathematics* 28. (1975), 824–838. eprint: `http://epubs.siam.org/doi/pdf/10.1137/0128067`.

[23] William S Zwicker. “A characterization of the rational mean neat voting rules”. *Mathematical and Computer Modelling* 48. (2008), 1374–1384.

[24] William S. Zwicker. “Consistency without neutrality in voting rules: When is a vote an average?” *Mathematical and Computer Modelling* 48. (2008). 00022, 1357–1373.

[25] William S. Zwicker. “Universal and Symmetric Scoring Rules for Binary Relations”. *Preprint* (2014).

ENS Lyon
E-mail address: benjamin.hadjibeyli@ens-lyon.fr

University of Auckland
E-mail address: mcw@cs.auckland.ac.nz