

Recognising Multidimensional Euclidean Preferences

Dominik Peters

Department of Computer Science
University of Oxford, UK
dominik.peters@cs.ox.ac.uk

Abstract

Euclidean preferences are a widely studied preference model, in which decision makers and alternatives are embedded in d -dimensional Euclidean space. Decision makers prefer those alternatives closer to them. This model, also known as multidimensional unfolding, has applications in economics, psychometrics, marketing, and many other fields. We study the problem of deciding whether a given preference profile is d -Euclidean. For the one-dimensional case, polynomial-time algorithms are known. We show that, in contrast, for every other fixed dimension $d > 1$, the recognition problem is equivalent to the existential theory of the reals (ETR), and so in particular NP-hard. We further show that some Euclidean preference profiles require exponentially many bits in order to specify any Euclidean embedding, and prove that the domain of d -Euclidean preferences does not admit a finite forbidden minor characterisation for any $d > 1$. We also study dichotomous preferences and the behaviour of other metrics, and survey a variety of related work.

1 Introduction

The study of preferences spans a multitude of fields: economics (and particularly game theory and social choice), political science, psychology, multi-agent systems, marketing, and others. An important element of working with preferences is *understanding* them by constructing models for them and identifying underlying structure. For example, in a psychological model of preferences, we aim to discover which underlying psychological process has generated the preferences we now observe (Coombs 1964). In political science, one might wonder about the underlying structure of ‘political space’ by analysing voter preferences (Merrill and Grofman 1999). In economics, working with well-structured preferences often allows a model to become analytically tractable.

This paper will use the lens and language of *computational social choice*. In recent years, much work in computational social choice has focussed on identifying structure in a given preference profile (Elkind, Lackner, and Peters 2016). The reason for this is simple: many of the problems social

choice aims to solve (such as preference aggregation, committee selection, or fair and efficient allocation) are computationally hard. However, if we manage to identify underlying, hopefully low-dimensional structure in the input profile, we can exploit this structure to guide algorithms. This approach has been successful, particularly for the domain of *single-peaked preferences*: once we have identified an *axis* on which an input profile is single-peaked, we can efficiently find Kemeny, Young, and Dodgson winners (Brandt et al. 2015), identify an optimal committee according to the Chamberlin–Courant rule (Betzler, Slinko, and Uhlmann 2013), control elections (Faliszewski et al. 2011), and solve the stable roommates problem (Bartholdi and Trick 1986).

This leaves the question of whether we can, in fact, efficiently find a certificate for single-peakedness (or another desired domain restriction) that such algorithms can use. The answer is often positive: there are efficient and certificate-producing recognition algorithms for preferences that are *single-peaked* (Escoffier, Lang, and Öztürk 2008), *single-crossing* (Elkind, Faliszewski, and Slinko 2012), *one-dimensional Euclidean* (Doignon and Falmagne 1994), or *single-peaked on a tree* (Trick 1989). Indeed, each of these preference domains is quite well-understood, for example in terms of forbidden-substructure characterisations (Ballester and Haeringer 2011), concise representations of all certificates (Peters and Elkind 2016; Bartholdi and Trick 1986), containment relations between the different domains (Elkind, Faliszewski, and Skowron 2014), and the probability with which a random preference profile falls within a given domain (Bruner and Lackner 2015).

Intriguingly, there is another preference domain for which we do not have a comparable amount of understanding, yet it is an extremely popular modelling choice across disciplines. Known as *spatial preferences*, or as the *d -Euclidean domain*, or as *multidimensional unfolding*, this preference domain contains profiles that can be ‘embedded’ into d -dimensional Euclidean space. Precisely, a preference profile is d -Euclidean if we can assign every voter and every alternative a point in \mathbb{R}^d such that voters prefer those alternatives that are closer to them (according to the usual Euclidean metric) to those that are further away. This characterisation of preferences has intuitive appeal: considering \mathbb{R}^d as a continuous ‘policy space’, within which alternatives can vary along multiple dimensions, each voter is identified

with an *ideal point* (Bennett and Hays 1960).

In practice, the embedding of voters and alternatives into \mathbb{R}^d is hidden, and we only have access to the ordinal ranking data in form of a preference profile (Anshelevich, Bhardwaj, and Postl 2015). Given this data, can we recover an appropriate embedding into \mathbb{R}^d that explains the preferences? This problem is known as multidimensional unfolding (Bennett and Hays 1960), and a large variety of methods that attempt to estimate embeddings have been proposed in the statistics and psychometrics literature (for a modern exposition see, e.g., Borg and Groenen 2005). None of these methods is guaranteed to return a suitable embedding whenever it exists and terminate in polynomial time.

We analyse this problem from a formal computational perspective. In particular, we prove that the decision problem of identifying d -Euclidean preference profiles is NP-hard for each fixed $d \geq 2$. More precisely, we prove that the problem is equivalent to the existential theory of the reals (ETR), and thus is in fact $\exists\mathbb{R}$ -complete. The recognition problem is thus unlikely to be contained in NP (though it is decidable in PSPACE). Using recent results about hyperplane arrangements, we deduce that, for each fixed $d \geq 2$, there exist d -Euclidean preference profiles such that the coordinates of any d -Euclidean embedding require exponentially many bits to specify. Thus, there is provably no polynomial-time algorithm that, on input a preference profile, outputs a d -Euclidean embedding if one exists.

Some restricted preference domains admit characterisations by a finite list of *forbidden subprofiles*, such as preferences that are single-peaked (Ballester and Haeringer 2011), single-crossing (Bredereck, Chen, and Woeginger 2013), or single-peaked on a circle (Peters and Lackner 2017). In contrast, Chen, Pruhs, and Woeginger (2015) show that the 1-Euclidean domain does not admit such a characterisation. They conjectured that the same is true for the domain of d -Euclidean preferences, for each fixed $d \geq 2$. We use a connection with the theory of oriented matroids to prove their conjecture.

The results in this paper cast some doubt on the fruitfulness of exploiting structure in d -Euclidean preferences in computational social choice, and they also limit the extent to which the spatial preference model can be used to reliably explain observed preference data. Future work could explore ways of mitigating this situation. At the end of this paper, we briefly sketch one way that this could be done, namely by replacing the Euclidean ℓ_2 -metric by the ℓ_1 - or ℓ_∞ -metric. We show that the recognition problems corresponding to these metrics are contained in NP.

2 Preliminaries

Euclidean preferences Let A be a finite set of *alternatives* or *candidates*. A *preference relation* \succsim over A (usually referred to as a *vote*) is a complete and transitive binary relation over A . We denote by \succ the strict part of \succsim , that is, $a \succ b$ if and only if $a \succsim b$ but $b \not\succeq a$; and we denote by \sim the indifference part of \succsim , with $a \sim b$ if and only if $a \succsim b$ and $b \succsim a$. A *profile* V over A is a set of votes over A . For notational convenience, we give voters names like v or i , and refer to their preference relation by \succsim_v and \succsim_i .

Let V be a profile over A . We say that V is *d -Euclidean* (where $d \geq 1$) if there exists a map $x : V \cup A \rightarrow \mathbb{R}^d$ with

$$a \succ_v b \implies \|x(v) - x(a)\| < \|x(v) - x(b)\| \quad (1)$$

for all $v \in V$ and $a, b \in A$. Thus, voter v prefers those alternatives which are closer to v according to the embedding x . Here, $\|\cdot\|$ refers to the usual Euclidean ℓ_2 -norm on \mathbb{R}^d ,

$$\|(x_1, \dots, x_d)\| = \|(x_1, \dots, x_d)\|_2 = \sqrt{x_1^2 + \dots + x_d^2}.$$

The (*open*) *ball* of radius r centred at c is $B(c, r) = \{x \in \mathbb{R}^d : \|x - c\| < r\}$.

Typically, we will consider profiles of *strict orders* in which every vote \succsim_v is antisymmetric (that is, we do not allow ties). Notice that in any d -Euclidean embedding of a profile of strict orders, no two alternatives can reside at the same point of \mathbb{R}^d .

In some applications, it makes sense to consider preferences that include indifferences (ties). An extreme case, which nevertheless finds many applications, is that of *dichotomous preferences*. A vote \succsim is *dichotomous* if there are no three alternatives $a, b, c \in A$ with $a \succ b \succ c$. Equivalently, \succsim is dichotomous if A splits into approved and non-approved alternatives. That is, we can write $A = A_1 \cup A_2$ with $A_1 \cap A_2 = \emptyset$ satisfying $a \succ b$ iff $a \in A_1$ and $b \in A_2$. We then say that the voter *approves* of the alternatives in A_1 , while the voter does not approve the alternatives in A_2 . A dichotomous vote can (and will) be specified by just giving the set A_1 of approved alternatives.

Our definition of Euclidean preferences applies to dichotomous preferences as well. Following the terminology of Elkind and Lackner (2015), we call a profile of dichotomous preferences *d -DE* (Dichotomous Euclidean) if it is d -Euclidean. In this context, the definition requires that there is an embedding $x : V \cup A \rightarrow \mathbb{R}^d$ so that for each voter $v \in V$, the set of approved alternatives of v coincides with the set of alternatives contained in some ball $B(x(v), r_v)$ centred at $x(v)$.

Some authors define Euclidean preferences in a subtly different way from us. One popular definition involves reversing the direction of the implication arrow in (1). Notice that under this definition, whenever a voter is equidistant between two alternatives, the voter is free to break the tie in either way. In particular, when $d \geq 2$, then *every* preference profile is “ d -Euclidean” under this definition – place all voters at the origin, and position alternatives on the unit sphere around the origin. In the area of multidimensional unfolding, embeddings of this type are said to include *degeneracies* (see, e.g., Busing, Groenen, and Heiser 2005). Our definition circumvents this issue by just outright disallowing degeneracies.

Bogomolnaia and Laslier (2007) use another different definition: they replace the implication in (1) by an if-and-only-if. This definition is equivalent to ours for strict preferences, but is much more restrictive for preferences including ties: Bogomolnaia and Laslier’s definition requires that whenever a voter v is indifferent between a and b , then a and b are equidistant to v . Our definition does not impose any relation on the relative distances in cases of ties, which makes it more applicable to the case of dichotomous orders.

Existential Theory of the Reals (ETR) The *language of the first-order theory of the reals* consists of formulas using as symbols (i) a countable collection of variable symbols x_i , (ii) constant symbols 0 and 1, (iii) addition, subtraction, multiplication symbols, (iv) the equality (=) and inequality (<) symbols, (v) Boolean connectives (\vee, \wedge, \neg), (vi) universal and existential quantifiers (\forall, \exists). The *theory of the reals* consists of all true sentences in this language, interpreted using the obvious semantics (where quantifiers quantify over the real numbers \mathbb{R}).

The *existential theory of the reals (ETR)* consists of the true sentences of the form

$$\exists x_1 \in \mathbb{R} \exists x_2 \in \mathbb{R} \dots \exists x_n \in \mathbb{R} \quad F(x_1, x_2, \dots, x_n)$$

with $F(x_1, x_2, \dots, x_n)$ a quantifier-free formula in the language just defined. In other words, F is a Boolean combination of equalities and inequalities of real polynomials.

The *decision problem* of ETR is the problem of deciding whether a given sentence of the above form is true, that is whether it is a member of ETR. Schaefer (2010) introduced the complexity class $\exists\mathbb{R}$ as the class of decision problems that admit a polynomial-time many-one reduction to the decision problem of ETR. Thus, $\exists\mathbb{R}$ captures the computational complexity of the existential theory of the reals. We say that a problem A is $\exists\mathbb{R}$ -hard if all problems in $\exists\mathbb{R}$ reduce to A in polynomial time. We say that A is $\exists\mathbb{R}$ -complete if it is contained in $\exists\mathbb{R}$ and is $\exists\mathbb{R}$ -hard.

From the definition of ETR it is not even clear that the decision problem of ETR is decidable. By introducing a quantifier-elimination procedure, Tarski (1948) showed that ETR is in fact decidable. Since then, a variety of algorithmic improvement have been made over Tarski's procedure, and there exist algorithms with a singly-exponential time dependence in the number of variables (Grigor'ev 1988; Renegar 1992). In addition, Canny (1988) obtained the astonishing result that ETR can be solved in polynomial space. Thus $\exists\mathbb{R} \subseteq \text{PSPACE}$. ETR can be used to solve the propositional satisfiability problem (3SAT); thus, we have the containments $\text{NP} \subseteq \exists\mathbb{R} \subseteq \text{PSPACE}$.

Several $\exists\mathbb{R}$ -complete problems are known, and many of them are questions of the form "can a given combinatorial object be geometrically represented?". Particular examples include recognising intersection graphs of line segments in the plane (Schaefer 2010), of unit disk graphs (Kang and Müller 2012), or of unit distance graphs (Schaefer 2013).

Arrangements of hyperplanes Our exposition and terminology follows Kang and Müller (2012).

An (*affine*) d -hyperplane is a set of form $h = \{x \in \mathbb{R}^d : c^T x = b\} \subseteq \mathbb{R}^d$ for some $c \in \mathbb{R}^d$ and $b \in \mathbb{R}$. A particular example of a hyperplane, for given $p, q \in \mathbb{R}^d$, is the set $\{x \in \mathbb{R}^d : \|x - p\| = \|x - q\|\}$ of points that are equidistant to p and q ; in two dimensions, this is the *perpendicular bisector*. Any hyperplane h divides $\mathbb{R}^d \setminus h$ into two connected components, namely the half-planes $c^T x > b$ and $c^T x < b$. We can give h an *orientation* by (arbitrarily) designating one of these components as h 's *positive side* h^+ , and the other as h 's *negative side* h^- . We call a hyperplane with a chosen orientation an *oriented hyperplane*. An *oriented hyperplane*

arrangement (h_1, \dots, h_n) is a finite ordered collection of oriented hyperplanes in \mathbb{R}^d .

Given a fixed oriented hyperplane arrangement $\mathcal{H} = (h_1, \dots, h_n)$, we can assign to each point $x \in \mathbb{R}^d$ its *sign vector* $\sigma(x) \in \{-, 0, +\}^n$ by setting $\sigma(x)_i = +$ if $x \in h_i^+$, $\sigma(x)_i = 0$ if $x \in h_i$, and $\sigma(x)_i = -$ if $x \in h_i^-$. Thus, the sign vector of x records, for each oriented hyperplane in the arrangement, on which side of the hyperplane x lies. The *combinatorial description* $\mathcal{D}(\mathcal{H}) = \{\sigma(x) : x \in \mathbb{R}^d\}$ of \mathcal{H} is the set of all sign vectors induced by the arrangement \mathcal{H} . If $\mathcal{D}(\mathcal{H}) = \mathcal{D}(\mathcal{H}')$, we say that \mathcal{H} and \mathcal{H}' are *isomorphic*.

A connected component of $\mathbb{R}^d \setminus \mathcal{H} = \mathbb{R}^d \setminus (h_1 \cup \dots \cup h_n)$ is called a *cell*. All points in the same cell have the same sign vector.

3 The Recognition Problem

Here is the formal definition of the recognition problem whose complexity we will investigate in this section.

d -EUCLIDEAN

Instance: a profile V of strict orders over A

Question: is V d -Euclidean?

Proposition 1. d -EUCLIDEAN is contained in $\exists\mathbb{R}$ for every $d \geq 1$. In particular it is contained in PSPACE.

Proof. This is almost immediate from the definition of d -Euclidean preferences. Namely, a profile is d -Euclidean if and only if there *exist* reals $x_{r,i} \in \mathbb{R}$ for each $r \in A \cup V$ and $i = 1, \dots, d$ such that whenever $a \succ_r b$, we have

$$\begin{aligned} \|x_v - x_a\| &< \|x_v - x_b\| \\ \iff \sum_{i=1}^d (x_{v,i} - x_{a,i})^2 &< \sum_{i=1}^d (x_{v,i} - x_{b,i})^2. \end{aligned}$$

Thus, the problem is equivalent to asking whether a system of polynomial inequalities has a solution. This system can be constructed in polynomial time, given the profile. \square

Our hardness proofs will use the following problem about combinatorial descriptions of hyperplane arrangements.

d -REALISABILITY

Instance: a set $S \subseteq \{-, +\}^n$ of sign vectors with $(-, \dots, -), (+, \dots, +) \in S$

Question: is there an oriented d -hyperplane arrangement \mathcal{H} with $S \subseteq \mathcal{D}(\mathcal{H})$?

For example, the set $S = \{(+, +, +, +), (-, +, +, -), (-, +, -, +), (-, +, -, -), (-, -, -, +), (-, -, -, -)\}$ is 2-realised by the four red lines in Figure 1, where the red label of the line is placed on the positive side of the line.

Theorem 2 (Kang and Müller 2012). d -REALISABILITY is $\exists\mathbb{R}$ -complete for $d \geq 2$.

Kang and Müller establish this by a reduction from SIMPLE STRETCHABILITY, the problem of deciding whether an arrangement of pseudolines can be stretched into an isomorphic arrangement of lines. That problem is $\exists\mathbb{R}$ -complete

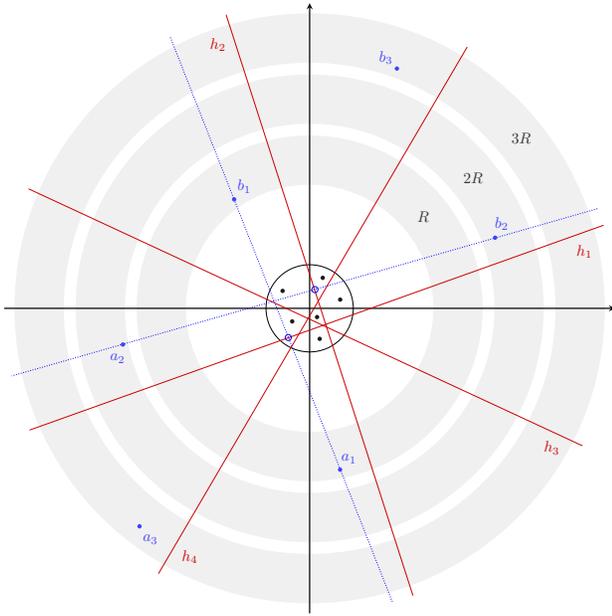


Figure 1: 2-Euclidean embedding of a profile obtained from a 2-realizable sign vector set S through the reduction of Theorem 3. Black dots represent voters v_σ , positioned within the unit ball and within the cell with sign vector σ induced by the red hyperplanes (lines). The red labels are on the positive side of each line. Blue circles denote the points p_i . Blue dots correspond to alternatives; note that a_i and b_i are at radius $R \cdot i \pm 2$ from the origin.

by Mněv's (1985) universality theorem, a deep topological result about representing semialgebraic varieties. Shor (1991) gives a direct proof of NP-hardness by a reduction from SAT.

We are now ready to prove our main result.

Theorem 3. d -EUCLIDEAN is $\exists\mathbb{R}$ -complete for each $d \geq 2$.

Proof. We have already seen that d -EUCLIDEAN is contained in $\exists\mathbb{R}$ (Proposition 1). We now show $\exists\mathbb{R}$ -hardness by a reduction from d -REALISABILITY.

Let $S \subseteq \{-, +\}^n$ be a given set of sign vectors with $(-, \dots, -), (+, \dots, +) \in S$. We construct a profile of $|S|$ votes over a total of $2n$ alternatives. Precisely, we take as alternatives the set $A = \{a_1, b_1, \dots, a_n, b_n\}$. For each $\sigma \in S$, we introduce a voter v_σ with strict order \succ_σ such that

$$\{a_1, b_1\} \succ_\sigma \{a_2, b_2\} \succ_\sigma \dots \succ_\sigma \{a_n, b_n\}$$

and specified by

$$\begin{cases} a_i \succ_\sigma b_i & \iff \sigma_i = +, \\ b_i \succ_\sigma a_i & \iff \sigma_i = -. \end{cases}$$

This completes the description of the reduction. We now show its correctness.

Suppose the profile constructed is d -Euclidean, and let $x : V \cup A \rightarrow \mathbb{R}^d$ be a Euclidean embedding. Take the oriented

hyperplane arrangement $\mathcal{H} = (h_1, \dots, h_n)$ defined by

$$h_i = \{x \in \mathbb{R}^d : \|x - x(a_i)\| = \|x - x(b_i)\|\},$$

$$h_i^+ = \{x \in \mathbb{R}^d : \|x - x(a_i)\| < \|x - x(b_i)\|\},$$

$$h_i^- = \{x \in \mathbb{R}^d : \|x - x(a_i)\| > \|x - x(b_i)\|\}.$$

Then, clearly, $S \subseteq \mathcal{D}(\mathcal{H})$: Let $\sigma \in S$ and let $i \in \{1, \dots, d\}$. If $\sigma_i = +$, we have $a_i \succ_\sigma b_i$, and thus by definition of Euclidean preferences, we must have $\|x(v_\sigma) - x(a_i)\| < \|x(v_\sigma) - x(b_i)\|$ and hence $x(v_\sigma) \in h_i^+$ so that $\sigma(x(v_\sigma))_i = + = \sigma_i$. Similarly if $\sigma_i = -$. It follows that $\sigma \in \mathcal{D}(\mathcal{H})$. Hence $S \subseteq \mathcal{D}(\mathcal{H})$.

Conversely, suppose that $S \subseteq \mathcal{D}(\mathcal{H})$ for some oriented d -hyperplane arrangement \mathcal{H} . By applying an appropriate scaling map $x \mapsto \lambda x$ if needed, we may assume that every cell of \mathcal{H} intersects the unit ball $B(0, 1) \subseteq \mathbb{R}^d$. Write $\mathcal{H} = (h_1, \dots, h_n)$ with $h_i = \{x \in \mathbb{R}^d : u_i^T x = b_i\}$, where without loss of generality $\|u_i\| = 1$, so that u_i is a unit vector. Further, we will say that $h_i^+ = \{x : u_i^T x > b_i\}$ and $h_i^- = \{x : u_i^T x < b_i\}$.

We now construct a Euclidean embedding $x : A \cup V \rightarrow \mathbb{R}^d$. We start by placing the voter v_σ corresponding to $\sigma \in S$ at an arbitrary point $x(v_\sigma) \in B(0, 1)$ of the cell of \mathcal{H} with sign vector σ . This exists by our assumption that $S \subseteq \mathcal{D}(\mathcal{H})$.

Next, for each $i = 1, \dots, d$, pick some point $p_i \in h_i \cap B(0, 1)$ (this is possible because $B(0, 1)$ meets both h_i^- and h_i^+ since $(-, \dots, -), (+, \dots, +) \in S$). Following an argument by Kang and Müller (2012), we set for $r > 0$

$$w_{i,r}^+ := p_i + r u_i \quad \text{and} \quad w_{i,r}^- := p_i - r u_i,$$

$$B_{i,r}^+ := B(w_{i,r}^+, r) \quad \text{and} \quad B_{i,r}^- := B(w_{i,r}^-, r).$$

Note that $B_{i,r}^+ \subseteq h_i^+$ and $B_{i,r}^- \subseteq h_i^-$. In fact (see picture), $\bigcup_{r>0} B_{i,r}^+ = h_i^+$ and $\bigcup_{r>0} B_{i,r}^- = h_i^-$.

Hence, for all r sufficiently large, we have

$$x(v_\sigma) \in B_{i,r}^+ \quad \text{for } \sigma \text{ with } \sigma_i = +, \text{ and}$$

$$x(v_\sigma) \in B_{i,r}^- \quad \text{for } \sigma \text{ with } \sigma_i = -.$$

Fix a value $R > 4$ of r for which this holds. We now pick the positions of the alternatives in the Euclidean embedding: Set

$$x(a_i) = w_{i,R-i}^+ \quad \text{and} \quad x(b_i) = w_{i,R-i}^-.$$

We are left to verify that the map $x : A \cup V \rightarrow \mathbb{R}^d$ thus constructed actually corresponds to voters' preferences. First let us show that, according to the embedding x , every voter's preference has the form

$$\{a_1, b_1\} \succ_\sigma \{a_2, b_2\} \succ_\sigma \dots \succ_\sigma \{a_n, b_n\}.$$

So let v_σ be a voter, let $1 \leq i < j \leq n$, and let $c_i \in \{a_i, b_i\}$ and $c_j \in \{a_j, b_j\}$. Then

$$\begin{aligned} & \|x(v_\sigma) - x(c_i)\| \\ & \leq \|x(v_\sigma) - p_i\| + \|p_i - x(c_i)\| \quad (\text{triangle inequality}) \\ & \leq 2 + R \cdot i \quad (u_i \text{ is a unit vector}) \\ & < R \cdot j - 2 \quad (j > i \text{ and } R > 4) \\ & \leq \|x(c_j) - p_j\| - \|p_j - x(v_\sigma)\| \quad (\text{as before}) \\ & \leq \|x(v_\sigma) - x(c_j)\|. \quad (\text{reverse triangle inequality}) \end{aligned}$$

Hence $c_i \succ_{\sigma} c_j$, as desired. Finally, we need to confirm that

$$\|x(v_{\sigma}) - x(a_i)\| < \|x(v_{\sigma}) - x(b_i)\| \iff \sigma_i = +.$$

So suppose $\sigma_i = +$. By choice of R , we have $x(v_{\sigma}) \in B_{i,R \cdot i}^+$, so that $\|x(v_{\sigma}) - x(a_i)\| < R \cdot i$. On the other hand, we have $\|x(v_{\sigma}) - x(b_i)\| \geq R \cdot i$: for suppose not. Then $x(v_{\sigma}) \in B_{i,R \cdot i}^-$, and thus $x(v_{\sigma}) \in B_{i,R \cdot i}^+ \cap B_{i,R \cdot i}^- = \emptyset$, a contradiction. \square

Certainly, this hardness result implies that it is also hard to recognise d -Euclidean profiles of *weak orders* (since strict orders form a special case). For *dichotomous* orders, hardness does not follow immediately, but a similar reduction can be used. The argument employed is almost identical to the hardness result for recognising unit disk graphs (Kang and Müller 2012); see the full version (Peters 2016).

Theorem 4. *For each $d \geq 2$, it is $\exists\mathbb{R}$ -complete to decide whether a profile of dichotomous votes is d -Euclidean.*

4 Precision

In this section, we consider the question of how many bits are needed to specify a Euclidean embedding $x : A \cup V \rightarrow \mathbb{R}^d$. We only consider the ‘natural encoding’ where the coordinates of each point are given as rational numbers. Note that if every d -Euclidean profile were to admit an embedding that can be specified in polynomially many bits, then this would put the problem d -EUCLIDEAN in NP. Yet in this section we show that there is a family of profiles which need exponentially many bits in order to specify any Euclidean embedding. This result, by itself, does not rule out that the *decision* problem d -EUCLIDEAN is in NP: there could be a ‘clever’ way to certify that an embedding exists, without explicitly giving the embedding. On the other hand, our result shows that the *function* problem associated with the problem d -EUCLIDEAN is *provably* not in P.

Let us now make precise the notion of the *size* of an embedding $x : A \cup V \rightarrow \mathbb{R}^d$. Here, we follow the definitions of McDiarmid and Müller (2013). The number of bits needed to store an integer $n \in \mathbb{Z}$ is the number of digits in its binary representation plus its sign: $\text{size}(n) := 1 + \lceil \log_2(n + 1) \rceil$. We represent a rational number $q \in \mathbb{Q}$ as a pair of integers representing a fraction: if $q = m/n$ in lowest terms, we set $\text{size}(q) := \text{size}(m) + \text{size}(n)$. The size of a rational vector $x \in \mathbb{Q}^d$ is $\text{size}(x) := \sum_{i=1}^d \text{size}(x_i)$. Finally, the size of a *rational* Euclidean embedding $x : A \cup V \rightarrow \mathbb{Q}^d$ is

$$\text{size}(x) := \sum_{r \in A \cup V} \text{size}(x(r)).$$

Before we establish the promised lower bound, let us first give a corresponding upper bound. Namely, while some d -Euclidean profiles require exponentially many bits to specify, (single-)exponentially many bits are always enough. To see this, we will first need a guarantee that every d -Euclidean profile admits a rational embedding, because we have only assigned sizes to rational embeddings.

Theorem 5. *Every d -Euclidean profile admits a rational embedding. Further, for each $d \geq 2$, there is a constant*

$c = c(d)$ such that any d -Euclidean profile with n voters and m alternatives admits a rational embedding $x : A \cup V \rightarrow \mathbb{Q}^d$ with $\text{size}(x) \leq 2^{c(n+m)}$.

This theorem is an essentially immediate corollary of a general result due to Basu, Pollack, and Roy (1996) about the bit sizes of solutions to polynomial inequalities.

For the lower bound, we use techniques developed by McDiarmid and Müller (2013) and Kang and Müller (2012) and apply them to the reduction of Theorem 3. The proof appears in the full version of this paper (Peters 2016). For a profile V over alternative set A , we define $\|V\| := |V| + |A|$.

Theorem 6. *Fix $d \geq 2$. For a d -Euclidean preference profile V , let $e(V)$ denote the minimum size of a rational Euclidean embedding of V . For each $m \geq 1$, let $e(n + m)$ be the maximum $e(V)$ among d -Euclidean preference profiles V with $\|V\| = n + m$. Then $e(n + m) \geq 2^{\Omega(n+m)}$.*

5 Forbidden Subprofile Characterisations

We say that the profile V *contains* the profile W if we can obtain W by first deleting some alternatives and voters from V , and then relabelling the remaining alternatives and reordering the remaining voters. A preference domain \mathcal{R} (that is, a set of profiles) may then be *characterised* by *forbidden subprofiles* by giving a set \mathcal{S} of profiles such that $\mathcal{R} = \{\text{profile } V : V \text{ does not contain any } W \in \mathcal{S}\}$.

We call a preference domain \mathcal{R} *hereditary* if it is closed under containment. That is, if $V \in \mathcal{R}$ and V contains W , then $W \in \mathcal{R}$. The d -Euclidean domain is hereditary for any $d \geq 1$. Any hereditary domain \mathcal{R} is characterised by its complement $\overline{\mathcal{R}}$ (its set of counterexamples) in this way.

Consider a possible characterisation of the d -Euclidean domain by a set \mathcal{S} of forbidden subprofiles. We will call this characterisation *good* if the set \mathcal{S} is polynomial-time recognisable: that is, given a profile, there should be a polynomial-time algorithm deciding whether the given profile is contained in \mathcal{S} . Certainly, if \mathcal{S} were *finite*, then \mathcal{S} provides a good characterisation. However, there exist infinite characterisations that are still good in this sense, for example for interval graphs (Lekkerkerker and Boland 1962) and matrices with the consecutive ones property (Tucker 1972).

However, given the complexity result of Section 3, it is a straightforward observation that for each $d \geq 2$, no good characterisation by forbidden substructures will exist for the d -Euclidean domain, subject to a reasonable complexity-theoretic assumption.

Proposition 7. *For each $d \geq 2$, the set of d -Euclidean preference profiles does not admit a good characterisation by forbidden substructures unless $\exists\mathbb{R} \subseteq \text{coNP}$.*

Proof. Suppose a good characterisation by \mathcal{S} exists. We give a coNP-algorithm that recognises d -Euclidean preferences: Given an input profile, guess some subprofile, guess a relabeling of voters and alternatives in this subprofile, and check whether the result is contained in \mathcal{S} . \square

By a similar argument, no *finite* characterisation can exist unless $\text{P} = \exists\mathbb{R} = \text{NP}$. In the remainder of this section, we prove this weaker result *without* appealing to any

complexity-theoretic assumptions. To do this, we use a connection between the theory of arrangements of hyperplanes and the theory of *ordered matroids*.

Theorem 8 (Bokowski and Sturmfels 1989). *There exist infinitely many nonrealisable uniform oriented matroids of rank 3 such that every proper minor of them is realisable.*

To prove our result about d -Euclidean preferences, we will use the examples from Theorem 8 and apply to them the chain of many-one reductions that yielded the hardness result of Theorem 3. We present this argument in several lemmas, whose proofs appear in the full version (Peters 2016).

Lemma 9. *Fix $d \geq 2$. For every $n_0 \in \mathbb{N}$, there exists $n > n_0$ and a set $S \subseteq \{-, +\}^n$ with $(+, \dots, +), (-, \dots, -) \in S$ such that S is not d -realisable, but for each $i = 1, \dots, n$, the set $S_{-i} = \{s_{-i} : s \in S\}$ obtained by deleting coordinate i is d -realisable.*

Lemma 10. *Fix $d \geq 2$. For every $m_0 \in \mathbb{N}$, there is $m > m_0$ such that there exists a preference profile over m alternatives which is not d -Euclidean, yet removing any alternative yields a d -Euclidean profile.*

It is worth noting that all these infinitely many minimal counterexamples have the shape $\{a_1, b_1\} \succ \{a_2, b_2\} \succ \dots \succ \{a_n, b_n\}$, and in particular they are single-peaked.

Theorem 11. *The domain of d -Euclidean preferences does not admit a finite characterisation by forbidden configurations, for any fixed $d \geq 2$.*

Proof. Suppose for a contradiction that such a characterisation exists, and let M be the maximum number of alternatives in any of the forbidden configurations. By Lemma 10, there exists a profile V over at least $M+1$ alternatives which is not d -Euclidean. Since the forbidden configurations characterise the d -Euclidean domain, one of the configurations must be contained in V . Considering the size of the configuration, it must be contained in V even with one alternative deleted. However this profile is d -Euclidean, contradicting the fact that the d -Euclidean domain is hereditary. \square

6 Other metrics

We have defined Euclidean preferences using the usual Euclidean ℓ_2 -metric, measuring distances by shortest paths in the plane. Other choices of metric may be preferred in certain contexts. We now look at the ℓ_1 -metric and the ℓ_∞ -metric. The ℓ_1 -metric is also known as the *cityblock* or *Manhattan* distance, because it measures distances by shortest paths on a grid like the street network of Manhattan. Formally, the ℓ_1 -norm is defined by

$$\|(x_1, \dots, x_d)\|_1 := |x_1| + \dots + |x_d|.$$

Thus, the ℓ_1 -distance $\|x - y\|_1$ of two points x and y is the sum of the absolute distances along each coordinate axis. The ℓ_∞ -metric, on the other hand, measures the *maximum* distance along a coordinate axis:

$$\|(x_1, \dots, x_d)\|_\infty := \max\{|x_1|, \dots, |x_d|\}.$$

For each of these metrics (or indeed, any metric space), we can obtain a notion of d -Euclidean preferences. For certain

settings, the ℓ_∞ metric has a nice interpretation as corresponding to ‘pessimistic’ voters who judge candidate according to their (subjectively) worst feature. The ℓ_1 metric also has intuitive appeal – see Eguia (2011) and the references therein for arguments in favour of using Euclidean preferences with respect to this metric.

Comparing the ℓ_1 - and ℓ_∞ -metrics to the ℓ_2 -metric we have used so far, one gets the sense that ℓ_1 and ℓ_∞ are more ‘discrete’ or ‘combinatorial’ than the more geometric ℓ_2 . Supporting this intuition, we find that the complexity of the recognition problem changes (unless $\text{NP} = \exists\mathbb{R}$):

Theorem 12. *The problems of recognising preference profiles that are d -Euclidean with respect to the ℓ_1 -metric or the ℓ_∞ -metric are contained in NP for every $d \geq 1$.*

Proof. We start with the ℓ_1 -metric and show containment in NP by giving a nondeterministic reduction to linear programming. For each of the d coordinate axes of \mathbb{R}^d , nondeterministically guess in which order the points corresponding to voters and alternatives appear along that axis. Once we have decided these orderings, we can rewrite the definition of ℓ_1 -Euclidean preferences without using absolute values. Then, we can replace strict inequalities with weak inequalities by introducing additive ‘slack’ constants (Elkind and Faliszewski 2014, Prop. 3). The result is a linear program producing an ℓ_1 -Euclidean embedding if one exists.

$$d_{v,c,i} = \pm(x_{v,i} - x_{c,i}) \quad (\text{distance from } v \text{ to } c \text{ along } i)$$

$$d_{v,c} = \sum_{i=1}^d d_{v,c,i} \quad (\text{distance from } v \text{ to } c)$$

$$d_{v,a} \leq d_{v,b} - 1 \quad \text{when } a \succ_v b$$

$$x_{y,i} \leq x_{y',i} - 1 \quad \text{when } y \text{ occurs left of } y' \text{ on axis } i$$

In constraints of the first form, the \pm can be replaced by plus or minus at ‘compile’-time so that the quantity reflects $|x_{v,i} - x_{c,i}|$. The argument for the ℓ_∞ -metric is similar: here we additionally guess for each pair (v, c) in which direction the maximum distance is achieved. \square

7 Conclusions

The results of this paper are bad news for the d -Euclidean domain: because producing a Euclidean embedding will in general be infeasible, we are stuck with heuristic algorithms that may or may not produce a correct output in their allotted time. In some sense, our hardness results show that the estimation algorithms developed for the multidimensional unfolding problem over the past several decades are best possible, in the sense that we cannot hope for exact efficient algorithms. Still, future developments in ETR-solver technology might allow solving practical instances in reasonable time, and perhaps some of the ideas in the area of multidimensional unfolding can be formalised and yield exact algorithms. We have run some preliminary experiments on the PrefLib dataset (Mattei and Walsh 2013) using the `nlsat` solver (Jovanović and De Moura 2012) which appears to be the strongest ETR-solver available. However, `nlsat` was unable to decide whether *any* of the non-trivial PrefLib profiles that we tried was 2- or 3-Euclidean within a time bound of one hour.

Acknowledgements I thank Edith Elkind and Martin Lackner for helpful discussions, several anonymous reviewers for their feedback, and Jürgen Bokowski for correspondence about the theory of oriented matroids. This work was supported by EPSRC and by ERC under grant number 639945 (ACCORD).

References

- Anshelevich, E.; Bhardwaj, O.; and Postl, J. 2015. Approximating optimal social choice under metric preferences. In *AAAI '15*, 777–783.
- Ballester, M. A., and Haeringer, G. 2011. A characterization of the single-peaked domain. *Social Choice and Welfare* 36(2):305–322.
- Bartholdi, J., and Trick, M. A. 1986. Stable matching with preferences derived from a psychological model. *Operations Research Letters* 5(4):165–169.
- Basu, S.; Pollack, R.; and Roy, M.-F. 1996. On the combinatorial and algebraic complexity of quantifier elimination. *Journal of the ACM* 43(6):1002–1045.
- Bennett, J. F., and Hays, W. L. 1960. Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika* 25(1):27–43.
- Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research* 47(1):475–519.
- Bogomolnaia, A., and Laslier, J.-F. 2007. Euclidean preferences. *Journal of Mathematical Economics* 43(2):87–98.
- Bokowski, J., and Sturmfels, B. 1989. An infinite family of minor-minimal nonrealizable 3-chirotopes. *Mathematische Zeitschrift* 200(4):583–589.
- Borg, I., and Groenen, P. J. 2005. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.
- Brandt, F.; Brill, M.; Hemaspaandra, E.; and Hemaspaandra, L. A. 2015. Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates. *Journal of Artificial Intelligence Research* 53:439–496.
- Bredereck, R.; Chen, J.; and Woeginger, G. J. 2013. A characterization of the single-crossing domain. *Social Choice and Welfare* 41(4):989–998.
- Bruner, M.-L., and Lackner, M. 2015. On the likelihood of single-peaked preferences. Technical Report arXiv:1505.05852 [cs.GT].
- Busing, F. M.; Groenen, P. J.; and Heiser, W. J. 2005. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika* 70(1):71–98.
- Canny, J. 1988. Some algebraic and geometric computations in PSPACE. In *STOC '88*, 460–467.
- Chen, J.; Pruhs, K.; and Woeginger, G. J. 2015. The one-dimensional Euclidean domain: Finitely many obstructions are not enough. Technical Report arXiv:1506.03838 [cs.GT].
- Coombs, C. H. 1964. *A Theory of Data*. John Wiley & Sons.
- Doignon, J.-P., and Falmagne, J.-C. 1994. A polynomial time algorithm for unidimensional unfolding representations. *Journal of Algorithms* 16(2):218–233.
- Eguia, J. X. 2011. Foundations of spatial preferences. *Journal of Mathematical Economics* 47(2):200–205.
- Elkind, E., and Faliszewski, P. 2014. Recognizing 1-Euclidean preferences: An alternative approach. In *Algorithmic Game Theory*. Springer. 146–157.
- Elkind, E., and Lackner, M. 2015. Structure in dichotomous preferences. In *IJCAI '15*, 2019–2025.
- Elkind, E.; Faliszewski, P.; and Skowron, P. 2014. A characterization of the single-peaked single-crossing domain. In *AAAI '14*, 654–660.
- Elkind, E.; Faliszewski, P.; and Slinko, A. 2012. Clone structures in voters' preferences. In *EC '12*, 496–513.
- Elkind, E.; Lackner, M.; and Peters, D. 2016. Preference restrictions in computational social choice: Recent progress. In *IJCAI '16*, 4062–4065.
- Escoffier, B.; Lang, J.; and Öztürk, M. 2008. Single-peaked consistency and its complexity. In *ECAI '08*, 366–370.
- Faliszewski, P.; Hemaspaandra, E.; Hemaspaandra, L. A.; and Rothe, J. 2011. The shield that never was: Societies with single-peaked preferences are more open to manipulation and control. *Information and Computation* 209(2):89–107.
- Grigor'ev, D. Y. 1988. Complexity of deciding Tarski algebra. *Journal of Symbolic Computation* 5(1):65–108.
- Jovanović, D., and De Moura, L. 2012. Solving non-linear arithmetic. In *Automated Reasoning*. Springer. 339–354.
- Kang, R. J., and Müller, T. 2012. Sphere and dot product representations of graphs. *Discrete & Computational Geometry* 47(3):548–568.
- Lekkerkerker, C. G., and Boland, J. 1962. Representation of a finite graph by a set of intervals on the real line. *Fundamenta Mathematicae* 51(1):45–64.
- Mattei, N., and Walsh, T. 2013. Preflib: A library for preferences <http://www.preflib.org>. In *Algorithmic Decision Theory*. Springer. 259–270.
- McDiarmid, C., and Müller, T. 2013. Integer realizations of disk and segment graphs. *Journal of Combinatorial Theory, Series B* 103(1):114–143.
- Merrill, S., and Grofman, B. 1999. *A unified theory of voting: Directional and proximity spatial models*. Cambridge University Press.
- Mněv, N. E. 1985. Realizability of combinatorial types of convex polyhedra over fields. *Journal of Soviet Mathematics* 28(4):606–609.
- Peters, D., and Elkind, E. 2016. Preferences single-peaked on nice trees. In *AAAI '16*, 594–600.
- Peters, D., and Lackner, M. 2017. Preferences single-peaked on a circle. In *AAAI '17*.
- Peters, D. 2016. Recognising multidimensional Euclidean preferences. *arXiv:1602.08109 [cs.GT]*.
- Renegar, J. 1992. On the computational complexity and geometry of the first-order theory of the reals. *Journal of Symbolic Computation* 13(3):255–299.
- Schaefer, M. 2010. Complexity of some geometric and topological problems. In *Graph Drawing: 17th International Symposium*, volume 5849, 334–344. Springer.
- Schaefer, M. 2013. Realizability of graphs and linkages. In *Thirty Essays on Geometric Graph Theory*. Springer. 461–482.
- Shor, P. 1991. Stretchability of pseudolines is NP-hard. *Applied Geometry and Discrete Mathematics, Amer. Math. Soc., Providence, RI* 4:531–554.
- Tarski, A. 1948. *A decision method for elementary algebra and geometry*. Rand report. Rand Corporation.
- Trick, M. A. 1989. Recognizing single-peaked preferences on a tree. *Mathematical Social Sciences* 17(3):329–334.
- Tucker, A. 1972. A structure theorem for the consecutive 1's property. *Journal of Combinatorial Theory, Series B* 12(2):153–162.