Qualitative Reasoning about Directions in Semantic Spaces

Steven Schockaert
Cardiff University
Cardiff, UK
s.schockaert@cs.cardiff.ac.uk

Jae Hee Lee
Australian National University
Canberra, Australia
jae-hee.lee@anu.edu.au

Abstract

We introduce a framework for qualitative reasoning about directions in high-dimensional spaces, called $\mathcal{EER}$, where our main motivation is to develop a form of commonsense reasoning about semantic spaces. The proposed framework is, however, more general; we show how qualitative spatial reasoning about points with several existing calculi can be reduced to the realisability problem for $\mathcal{EER}$ (or REER for short), including $\mathcal{LR}$ and calculi for reasoning about betweenness, collinearity and parallelism. Finally, we propose an efficient but incomplete inference method, and show its effectiveness for reasoning with $\mathcal{EER}$ as well as reasoning with some of the aforementioned calculi.

1 Introduction

High-dimensional Euclidean spaces play a prominent role in areas such as computational linguistics, information retrieval and cognition [Turney and Pantel, 2010; Deerwester *et al.*, 1990; Gärdenfors, 2000]. These Euclidean spaces, often called semantic or conceptual spaces, are typically learned from large text corpora, using dimensionality reduction methods such as singular value decomposition or multidimensional scaling [Turney and Pantel, 2010]. In the resulting representations, points or vectors are used to encode the meaning of words, documents, concepts or entities.

The coordinates of the points or vectors in such a semantic space are essentially arbitrary; what matters is their spatial relationship. For example, the relative distance between points in a semantic space is used to discover synonymous words [Turney and Pantel, 2010] or to find the documents that best satisfy a query [Deerwester *et al.*, 1990]. The extent to which four points approximately form a parallelogram can be used to discover analogical proportions [Miclet *et al.*, 2008; Schockaert and Prade, 2011; Mikolov *et al.*, 2013]; e.g. the approach from [Mikolov *et al.*, 2013] was able to learn that "king is to queen what man is to woman". Geometric betweenness has been used for implementing a form of commonsense reasoning called interpolation. For example, the approach from [Derrac and Schockaert, 2014b] discovered that "candy store is conceptually between grocery store and toy store", meaning that natural properties that are true for both grocery stores and toy stores tend to be true for candy stores as well. Finally, it is possible to identify directions in a semantic space that correspond to interpretable semantic relations, such as e.g. "more violent than" in a space of films [Vig *et al.*, 2012; Kovashka *et al.*, 2012; Derrac and Schockaert, 2014a].

In this paper, we focus on qualitative reasoning about the latter type of direction relations. To date, semantic spaces have only received limited attention from the spatial reasoning community. Some authors have looked at logics of comparative similarity [Williamson, 1988; Sheremet *et al.*, 2010] in semantic spaces (including non-Euclidean spaces). Furthermore, [Davis *et al.*, 1999] discusses a language for expressing betweenness, collinearity and parallelism. Although the results in [Davis *et al.*, 1999] are proven for points in $\mathbb{R}^2$, they are relevant for higher-dimensional spaces (see Section 3). To the best of our knowledge, qualitative reasoning about directions has not yet been considered for high-dimensional spaces.

Although we only focus on reasoning aspects in this paper, our overall aim is to improve methods for extracting semantic relations from large text corpora. Most relation extraction approaches in the literature are based on lexical or syntactic patterns [Hearst, 1992; Carlson *et al.*, 2010; Nakashole *et al.*, 2012], which typically leads to high-recall, low-precision results. To increase precision while keeping recall high, implausible assertions are usually removed based on some form of constraint reasoning [Carlson *et al.*, 2010; Nakashole *et al.*, 2011], e.g. based on learned constraint rules such as "city X can only be the capital of country Y if X is located in Y". In this paper, we are specifically interested in relations that encode relative properties between entities of the same type, e.g. relations such as "funnier than" for films, for which suitable constraint rules are difficult to find. As an alternative to constraint rules, we propose to measure the plausibility of such relational facts in terms of the consistency of an associated set of spatial constraints. In particular, we consider relational facts plausible if they can be interpreted as direction relations in a semantic space. An important advantage of our approach is that it allows us to use existing semantic space representations of entities (as points) and relative properties (as directions) to help assess the plausibility of relational facts about entities or properties for which no such representation can be obtained (e.g. because they are too

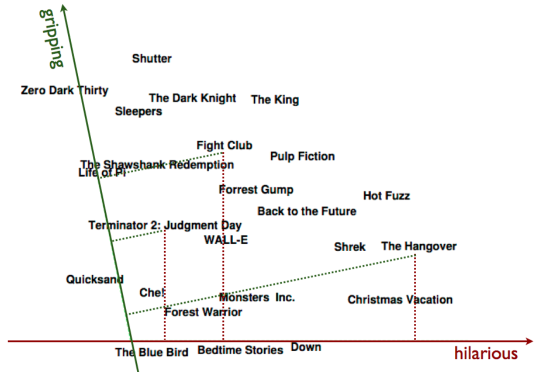

Figure 1: Directions corresponding to the relative properties 'more hilarious than' and 'more gripping than'.

infrequent in the text corpus).

The remainder of this paper is structured as follows. In the next section, we introduce $\mathcal{EER}$, a formalism for reasoning about directions in high-dimensional spaces. Section 3 subsequently shows how reasoning in several existing calculi can be reduced to the consistency problem for $\mathcal{EER}$ and characterises the computational complexity of the main decision problem. Then, in Section 4, we propose an efficient, but incomplete method for finding models of consistent sets of $\mathcal{EER}$ formulas, after which we provide experimental results in Section 5. In addition to evaluating the feasibility of reasoning about $\mathcal{EER}$ instances, we show how our implementation can be used to find solutions to problem instances of existing calculi such as $\mathcal{LR}$, based on the reductions from Section 3. To the best of our knowledge, this constitutes the first practical method that can find solutions for these calculi.

2 Euclidean Embedding of Rankings

Several authors have proposed to model relative properties as directions in a suitable semantic space. For example, [Kovashka *et al.*, 2012] learns directions modelling relations such as "shinier than" and "more formal than" in a semantic space of shoes (based on visual features), while [Derrac and Schockaert, 2014a] learns directions modelling relations such as "more gripping than" in a semantic space of films (based on textual features). The latter approach is illustrated in Figure 1, which shows directions corresponding to the relations "more gripping than" and "more hilarious than". Given these directions, we can rank films according to how much they have these properties by comparing the orthogonal projections of the corresponding points. For example, in the case of Figure 1, the Hangover is considered to be more hilarious than Fight Club, which is in turn considered more hilarious than Terminator 2. Note that while Figure 1 illustrates the principle for $\mathbb{R}^2$, semantic spaces will typically be of higher dimension.

Formally, each relative property, modelled as a vector $v \in \mathbb{R}^d$, induces a total preorder $\preceq_v$ on $\mathbb{R}^d$, ranking points based on the relative position of their orthogonal projections on the directed line $\lambda v, \lambda \in \mathbb{R}$. We define:

$$p \preceq_v q \quad \text{iff} \quad v^T(q-p) \geq 0 \tag{1}$$

We also use $p \sim_v q$ as an abbreviation for $(p \preceq_v q) \wedge (q \preceq_v p)$. Similarly, a strict partial order $\prec_v$ is defined as follows:

$$p \prec_v q \quad \text{iff} \quad v^T(q-p) > 0 \tag{2}$$

Definition 1. Let a set of variables $\{x_1, ..., x_n\}$ and a positive integer m be given. An $\mathcal{EER}$-formula[1] is a statement of the form $x_i \preceq_\ell x_j$ or a statement of the form $x_i \prec_\ell x_j$, with $1 \leq i, j \leq n$ and $1 \leq \ell \leq m$. We call the pair (π, ρ) a d-dimensional solution of a set of $\mathcal{EER}$-formulas Θ, with $\pi : \{x_1, ..., x_n\} \to \mathbb{R}^d$ and $\rho : \{1, ..., m\} \to \mathbb{R}^d$, if for every formula in Θ of the form

$$x_i \preceq_\ell x_j \qquad (\text{or} \quad x_i \prec_\ell x_j)$$

it holds that

$$\pi(x_i) \preceq_{\rho(\ell)} \pi(x_j) \qquad (\text{or} \quad \pi(x_i) \prec_{\rho(\ell)} \pi(x_j)),$$

which means that $\pi(x_i)$ is (strictly) before $\pi(x_j)$ when projected orthogonally on the directed line $\lambda\rho(\ell), \lambda \in \mathbb{R}$ (cf. (1) and (2)). If such a d-dimensional solution exists, we say that Θ is realisable in $\mathbb{R}^d$. The problem of deciding whether an $\mathcal{EER}$-formula is realisable is called REER.

In the following, we will use abbreviations such as $p \prec_\ell q \sim_\ell r$, which corresponds to the set $\{p \prec_\ell q, q \preceq_\ell r, r \preceq_\ell q\}$.

Definition 2. We say that a set of $\mathcal{EER}$ formulas Θ contains a cycle if there are formulas $\alpha_1, ..., \alpha_r$ in Θ, with each α_i of the form $y_i \prec_\ell y_{i+1}$ or $y_i \preceq_\ell y_{i+1}$, such that $y_{r+1} = y_1$ and at least one of the formulas α_i is of the form $y_i \prec_\ell y_{i+1}$.

The following proposition is easy to show.

Proposition 3. *Let Θ be defined as in Definition 1. For $d \geq m$, it holds that Θ is realisable in $\mathbb{R}^d$ iff Θ does not contain any cycles.*

Example 4. The $\mathcal{EER}$-formulas

Terminator 2 $\prec_{\textit{hilarious}}$ *Fight Club* $\prec_{\textit{hilarious}}$ *Hangover*
Hangover $\prec_{\textit{gripping}}$ *Terminator 2* $\prec_{\textit{gripping}}$ *Fight Club*

are realisable for $d = 2$. A solution is depicted in Figure 1.

Given a set of $\mathcal{EER}$-formulas Θ, we write $\dim(\Theta)$ for the lowest $d \in \mathbb{N}$ such that Θ is realisable in $\mathbb{R}^d$, where we define $\dim(\Theta) = +\infty$ if Θ is not realisable for any $d \in \mathbb{N}$ (i.e. if Θ contains a cycle). Intuitively, $\dim(\Theta)$ reflects the degree to which the set of assertions encoded by Θ is regular. To identify implausible assertions, we can impose a level of regularity by treating a set Θ as inconsistent if $\dim(\Theta) > d^*$, for some $d^* \in \mathbb{N}$. Such inconsistencies can be restored in the usual way, e.g. by choosing a maximal subset Θ^* for which $\dim(\Theta^*) \leq d^*$ based on confidence scores for each of the assertions in Θ. The view that $\dim(\Theta)$ reflects the regularity of Θ could also be used to identify plausible conclusions which are missing from Θ: we say that an $\mathcal{EER}$-formula $x \prec_i y$ is a plausible consequence of Θ if $\dim(\Theta \cup \{x \prec_i y\}) <$

[1] $\mathcal{EER}$ stands for *Euclidean Embedding of Rankings*.

$\dim(\Theta \cup \{y \preceq_i x\})$, and similarly for formulas of the form $x \preceq_i y$.

In this way, our approach supports a form of commonsense reasoning about relational facts based on dimensionality reduction, similar in spirit to how singular value decomposition was used for automatically extending ConceptNet in [Speer *et al.*, 2008] or how tensor decomposition was used for automatically extending YAGO in [Nickel *et al.*, 2012]. However, due to its geometric nature, our approach also allows us to combine qualitative knowledge with existing semantic space representations: given a set of $\mathcal{EER}$-formulas, one merely needs to find a solution (π, ρ) in which the maps π and ρ are fixed for objects and rankings with a known semantic space representation.

3 Reductions from Existing Calculi

In this section, we first show how the realisability problem for atomic $\mathcal{LR}$-formulas can be reduced to REER in $\mathbb{R}^2$, which will enable us to apply the implementation from Section 4 for checking the consistency of sets of atomic $\mathcal{LR}$-formulas. Subsequently, in Section 3.2 we will show how reasoning about betweenness, collinearity and parallelism can be reduced to REER in $\mathbb{R}^d$ for any d. This will allow us to characterise the computational complexity of REER, and could again be useful for using $\mathcal{EER}$ reasoners to check the consistency of constraints about betweenness, collinearity and parallelism.

3.1 Reduction from $\mathcal{LR}$

The $\mathcal{LR}$ calculus [Ligozat, 1993; Scivos and Nebel, 2005] is based on a set of 9 jointly exhaustive and pairwise disjoint relations between triples of points: l (left), r (right), b (behind), c (closer), f (further), e_{12}, e_{13}, e_{23} and eq. The main relations are l and r, where intuitively $l(p_1, p_2, p_3)$ holds if p_3 is located left of the oriented line defined by $\overrightarrow{p_1p_2}$ and $r(p_1, p_2, p_3)$ holds if p_3 is right of this line. Formally:

$$\begin{aligned} l(p_1,p_2,p_3) &\quad \text{iff} \quad \sin(\mathit{ang}(\overrightarrow{p_1p_2}, \overrightarrow{p_1p_3})) > 0 \\ r(p_1,p_2,p_3) &\quad \text{iff} \quad \sin(\mathit{ang}(\overrightarrow{p_1p_2}, \overrightarrow{p_1p_3})) < 0 \end{aligned}$$

where $\mathit{ang}(\overrightarrow{pq}, \overrightarrow{pr})$ is the oriented angle between $\overrightarrow{pq}$ and $\overrightarrow{pr}$. The remaining relations cover the cases where the three points are collinear or coincide:

$$\begin{aligned} b(p_1,p_2,p_3) &\quad \text{iff} \quad \exists \lambda > 1\,.\,\overrightarrow{p_3p_2} = \lambda \cdot \overrightarrow{p_1p_2} \\ c(p_1,p_2,p_3) &\quad \text{iff} \quad \exists 0 < \lambda < 1\,.\,\overrightarrow{p_3p_2} = \lambda \cdot \overrightarrow{p_1p_2} \\ f(p_1,p_2,p_3) &\quad \text{iff} \quad \exists \lambda < 0\,.\,\overrightarrow{p_3p_2} = \lambda \cdot \overrightarrow{p_1p_2} \\ e_{13}(p_1,p_2,p_3) &\quad \text{iff} \quad p_1 = p_3 \neq p_2 \\ e_{23}(p_1,p_2,p_3) &\quad \text{iff} \quad p_2 = p_3 \neq p_1 \\ e_{12}(p_1,p_2,p_3) &\quad \text{iff} \quad p_1 = p_2 \neq p_3 \\ eq(p_1,p_2,p_3) &\quad \text{iff} \quad p_1 = p_2 = p_3 \end{aligned}$$

Definition 5. An $\mathcal{LR}$-formula is an expression of the form $\bigvee_{i=1}^{k} \xi_i(x_1, x_2, x_3)$ with $\xi_i \in \{l, r, b, c, f, e_{12}, e_{13}, e_{23}, eq\}$ and x_1, x_2, x_3 are taken from a set of variables X. An $\mathcal{LR}$-formula is called *atomic*, if $k = 1$. We call an $X \to \mathbb{R}^2$ mapping π a solution of a set of $\mathcal{LR}$-formulas Ψ if the constraint $\xi(\pi(x_1), \pi(x_2), \pi(x_3))$ is satisfied for every formula of the form $\xi(x_1, x_2, x_3)$ in Ψ. If such a solution exists, we say that Ψ is satisfiable.

The satisfiability problem for $\mathcal{LR}$ is $\exists\mathbb{R}$-complete and the satisfiability problem for atomic $\mathcal{LR}$-formulas NP-hard [Lee, 2013], where $\exists\mathbb{R}$ is the class of problems that can be reduced in polynomial time to the problem of checking the consistency of a set of algebraic equations over the real numbers [Schaefer, 2010]. The class $\exists\mathbb{R}$ is known to be in PSPACE and to contain NP, but its exact relationship with the polynomial hierarchy is still unknown.

We now show that for every set of atomic $\mathcal{LR}$-formulas Ψ, in polynomial time we can construct a set of $\mathcal{EER}$ formulas Θ such that Ψ is satisfiable iff Θ is realisable in $\mathbb{R}^2$. For a formula in Ψ of the form $eq(x_1, x_2, x_3)$, we add the following formulas to Θ:

$$\begin{aligned} x_1 \sim_i x_2 \sim_i x_3 \prec_i y \prec_i z \\ x_1 \sim_j x_2 \sim_j x_3 \prec_j z \prec_j y \end{aligned}$$

where fresh rankings i and j and fresh variables y and z are used for every formula. Note that $x_3 \prec_i y \prec_i z$ and $x_3 \prec_j z \prec_j y$ imply that rankings i and j will correspond to different lines in any two-dimensional solution, in which case $x_1 \sim_i x_2 \sim_i x_3$ and $x_1 \sim_j x_2 \sim_j x_3$ implies that x_1, x_2 and x_3 coincide. For a formula in Ψ of the form $e_{12}(x_1, x_2, x_3)$, we add the following formulas to Θ:

$$x_1 \sim_i x_2 \sim_i x_3 \qquad\qquad x_1 \sim_j x_2 \prec_j x_3$$

where i and j are again fresh rankings for every formula. The formulas $x_2 \sim_i x_3$ and $x_2 \prec_j x_3$ are sufficient to guarantee that the lines corresponding to rankings i and j cannot coincide and that x_3 cannot coincide with x_2. From $x_1 \sim_i x_2$ and $x_1 \sim_j x_2$ we then find that x_1 and x_2 will coincide in any two-dimensional solution. Formulas of the form $e_{23}(x_1, x_2, x_3)$ and $e_{13}(x_1, x_2, x_3)$ are treated in an entirely analogous way.

For a formula in Ψ of the form $b(x_1, x_2, x_3)$ we add the following $\mathcal{EER}$ formulas to Θ:

$$x_1 \sim_i x_2 \sim_i x_3 \qquad\qquad x_3 \prec_j x_1 \prec_j x_2$$

with i and j fresh rankings. Indeed, $x_1 \sim_i x_2 \sim_i x_3$ guarantees that x_1, x_2 and x_3 are collinear in any two-dimensional solution, while $x_3 \prec_j x_1 \prec_j x_2$ guarantees that the points appear in the correct order. Formulas of the form $c(x_1, x_2, x_3)$ and $f(x_1, x_2, x_3)$ are treated entirely analogously.

For a formula in Ψ of the form $l(x_1, x_2, x_3)$ we add the following $\mathcal{EER}$ formulas to Θ:

$$x_3 \prec_i x_2 \prec_i x_1 \qquad\qquad x_3 \prec_j x_1 \prec_j x_2 \tag{3}$$

Similarly, for a formula of the form $r(x_1, x_2, x_3)$ we add:

$$x_2 \prec_i x_1 \prec_i x_3 \qquad\qquad x_1 \prec_j x_2 \prec_j x_3 \tag{4}$$

where each time, i and j are fresh rankings. These formulas alone are not sufficient, as the following lemma reveals.

Lemma 6. *Suppose Θ contains the $\mathcal{EER}$ formulas in* (3)*. In any solution (π, ρ) of Θ in $\mathbb{R}^2$ it holds that:*

$$\begin{aligned} l(\pi(x_1), \pi(x_2), \pi(x_3)) &\quad \textit{iff} \quad \sin(\mathit{ang}(\rho(i), \rho(j))) > 0 \\ r(\pi(x_1), \pi(x_2), \pi(x_3)) &\quad \textit{iff} \quad \sin(\mathit{ang}(\rho(i), \rho(j))) < 0 \end{aligned}$$

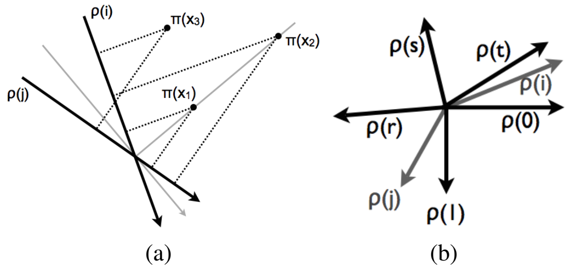

Figure 2: (a) A solution of the constraints in (3); (b) Illustration of Lemma 8.

Similarly, if Θ *contains the formulas in* (4) *then for any solution* (π, ρ) *it holds that:*

$$l(\pi(x_1), \pi(x_2), \pi(x_3)) \quad \textit{iff} \quad \sin(\textit{ang}(\rho(i), \rho(j))) < 0$$
$$r(\pi(x_1), \pi(x_2), \pi(x_3)) \quad \textit{iff} \quad \sin(\textit{ang}(\rho(i), \rho(j))) > 0$$

This lemma is illustrated in Figure 2(a), showing a solution of (3) where $\sin(\textit{ang}(\rho(i), \rho(j))) > 0$, which implies $l(\pi(x_1), \pi(x_2), \pi(x_3))$. Unfortunately, there is no obvious way to enforce $\sin(\textit{ang}(\rho(i), \rho(j))) > 0$ using additional $\mathcal{EER}$ formulas. We can overcome this issue by fixing the rankings 0 and 1 and adding $\mathcal{EER}$ constraints to Θ that impose that $\sin(\textit{ang}(\rho(0), \rho(1))) > 0$ iff $\sin(\textit{ang}(\rho(i), \rho(j))) > 0$. We say that a vector v is between the vectors u and w if $\mu v = \lambda u + (1 - \lambda)v$ for some $\mu > 0$ and $0 < \lambda < 1$.

Lemma 7. *Suppose* Θ *contains the following constraints:*

$$a \prec_i b \prec_i c \qquad b \prec_j a \prec_j c \qquad b \prec_k c \prec_k a \qquad (5)$$

Then it holds that $\rho(j)$ *is between* $\rho(i)$ *and* $\rho(k)$ *in any two-dimensional solution* (π, ρ) *of* Θ.

Using this lemma, we can show the following result.

Lemma 8. *Suppose* Θ *contains the following constraints:*

$$a_1 \prec_0 a_2 \prec_0 a_3 \quad a_2 \prec_1 a_1 \prec_1 a_3 \quad a_2 \prec_r a_3 \prec_r a_1 \qquad (6)$$
$$b_1 \prec_1 b_2 \prec_1 b_3 \quad b_2 \prec_r b_1 \prec_r b_3 \quad b_2 \prec_s b_3 \prec_s b_1 \qquad (7)$$
$$c_1 \prec_r c_2 \prec_r c_3 \quad c_2 \prec_s c_1 \prec_s c_3 \quad c_2 \prec_t c_3 \prec_t c_1 \qquad (8)$$
$$d_1 \prec_s d_2 \prec_s d_3 \quad d_2 \prec_t d_1 \prec_t d_3 \quad d_2 \prec_i d_3 \prec_i d_1 \qquad (9)$$
$$e_1 \prec_t e_2 \prec_t e_3 \quad e_2 \prec_i e_1 \prec_i e_3 \quad e_2 \prec_j e_3 \prec_j e_1 \qquad (10)$$

In any solution (π, ρ) *of* Θ *it holds that:*

$$\sin(\textit{ang}(\rho(0), \rho(1))) > 0 \;\; \textit{iff} \;\; \sin(\textit{ang}(\rho(i), \rho(j))) > 0$$

As illustrated in Figure 2(b), the reason we need five instances of (5) in (6)-(10) is to ensure that the constraints have a solution for every choice of $\rho(i)$ and $\rho(j)$ such that $\sin(\textit{ang}(\rho(i), \rho(j))) > 0$ iff $\sin(\textit{ang}(\rho(0), \rho(1))) > 0$, provided that $\rho(i)$ is not between $\rho(0)$ and $\rho(1)$. As we will see below, we can always ensure that the latter requirement is satisfied in our construction.

Hence, for every $\mathcal{LR}$-formula of the form $l(x_1, x_2, x_3)$ in Ψ, we add the formulas in (3) to Θ, in addition to the formulas in (6)–(10), where $a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3, d_1, d_2, d_3, e_1, e_2, e_3$ are fresh variables and r, s, t, i, j are fresh rankings for every $\mathcal{LR}$-formula; for every $\mathcal{LR}$-formula of the form $r(x_1, x_2, x_3)$ in Ψ we similarly add the formulas in (4) and (6)–(10).

Let (π, ρ) be a solution of Θ. It is clear that π satisfies all formulas of the form $\xi(x_1, x_2, x_3)$ from Ψ with $\xi \in \{b, c, f, e_{12}, e_{13}, e_{23}, \textit{eq}\}$. From Lemma 6 and 8 it follows that:

1. either $r(\pi(x_1), \pi(x_2), \pi(x_3))$ holds for every formula $r(x_1, x_2, x_3) \in \Psi$ and $l(\pi(x_1), \pi(x_2), \pi(x_3))$ holds for every formula $l(x_1, x_2, x_3) \in \Psi$; or
2. $l(\pi(x_1), \pi(x_2), \pi(x_3))$ holds for every formula $r(x_1, x_2, x_3) \in \Psi$ and $r(\pi(x_1), \pi(x_2), \pi(x_3))$ holds for every formula $l(x_1, x_2, x_3) \in \Psi$.

In the former case, we find that the restriction of π to X is a solution of Ψ. In the latter case, we define a new solution (π', ρ') by reflecting all points and vectors over the Y-axis. It is easy to see that (π', ρ') is a solution of Θ and that the restriction of π' to X is a solution of Ψ. In each case, we have that if Θ is realisable in $\mathbb{R}^2$ then Ψ is satisfiable.

Conversely, if π is a solution of Ψ, we need to construct a solution (π, ρ) of Θ. For a formula of the form $l(x_1, x_2, x_3)$, it is easy to verify that we can always find rankings i and j such that (3) is satisfied and $\sin(\textit{ang}(\rho(i), \rho(j))) > 0$. For formulas of the form $r(x_1, x_2, x_3)$ we can similarly find rankings i and j such that (4) is satisfied and $\sin(\textit{ang}(\rho(i), \rho(j))) > 0$. Once these vectors have been fixed, we choose $\rho(0)$ and $\rho(1)$ such that none of the other vectors $\rho(i)$ is between $\rho(0)$ and $\rho(1)$. Then π and ρ can always be extended such that the formulas (6)–(10) are satisfied. For formulas of the form $\xi(x_1, x_2, x_3)$ with $\xi \in \{b, c, f, e_{12}, e_{13}, e_{23}, \textit{eq}\}$ it is straightforward to see how ρ needs to be extended to ensure that (π, ρ) is a solution of Θ. Hence we obtain the following proposition.

Proposition 9. *Let* Ψ *be a set of atomic* $\mathcal{LR}$*-formulas and let* Θ *be the corresponding set of* $\mathcal{EER}$ *formulas. It holds that* Ψ *is satisfiable iff* Θ *is realisable in* $\mathbb{R}^2$.

3.2 Betweenness, Collinearity and Parallelism

In this section, we will consider the following spatial relations between points in $\mathbb{R}^d$:

- *Col*(x, y, z) iff x, y and z are distinct and collinear.
- *Bet*(x, y, z) iff *Col*(x, y, z) and y is between x and z.
- *Gen*(x, y, z) iff x, y and z are in general linear position (i.e. distinct and not collinear).
- *Par*(w, x, y, z) iff the line wx is distinct from and parallel to yz, with $w \neq x$ and $y \neq z$.

These relations are studied in [Davis *et al.*, 1999], with the exception of *Gen*. They are of particular interest in the context of semantic spaces, as they form the basis for commonsense reasoning methods such as interpolation and analogical reasoning [Miclet *et al.*, 2008; Schockaert and Prade, 2013; Derrac and Schockaert, 2014b].

Definition 10. An $\mathcal{L}$-formula is an expression of the form $\xi(x, y, z)$, with $\xi \in \{\textit{Col}, \textit{Bet}, \textit{Gen}\}$, or *Par*$(w, x, y, z)$. Let Ψ be a set of $\mathcal{L}$-formulas over the set of variables X. An $X \to \mathbb{R}^d$ mapping π which satisfies each of the $\mathcal{L}$-formulas in Ψ is called a solution of Ψ. If such a solution exists, we say that Ψ is realisable in $\mathbb{R}^d$.

It follows from Lemma 1 in [Davis *et al.*, 1999] that checking the realisability in $\mathbb{R}^2$ of a set of $\mathcal{L}$-formulas is $\exists\mathbb{R}$-hard (even when only the predicates *Col* and *Par* are considered). We can extend this result to $\mathbb{R}^d$ for any $d > 2$.

Proposition 11. *The problem of checking whether a set of $\mathcal{L}$-formulas is realisable in $\mathbb{R}^d$ is $\exists\mathbb{R}$-hard for any $d \geq 2$.*

Proof. Let Ψ_1 be a set of $\mathcal{L}$-formulas over X and let $d \geq 2$. We define a new set of $\mathcal{L}$-formulas, imposing that every x in X is in the convex hull of three fresh variables a, b, c:

$$\Psi_2 = \Psi_1 \cup \{Bet(a, d_x, b), Bet(d_x, x, c) \mid x \in X\}$$

where d_x is a fresh variables for every $x \in X$. It is easy to see that Ψ_1 is realisable in $\mathbb{R}^2$ iff Ψ_2 is realisable in $\mathbb{R}^d$. □

We now show how the realisability problem for a set of $\mathcal{L}$-formulas Ψ can be reduced to the realisability problem for a corresponding set of $\mathcal{EER}$-formulas Θ. To express that x, y and z are collinear, we will use $\mathcal{EER}$ formulas that encode that they are located on the intersection of $d-1$ linearly independent hyperplanes. In particular, for every formula in Ψ of the form $Col(x, y, z)$ or $Bet(x, y, z)$, we add the following $\mathcal{EER}$ formula to Θ for every $i \in \{1, ..., d-1\}$

$$x \sim_{r_i} y \sim_{r_i} z \tag{11}$$

where r_i is a fresh ranking. To impose that the directions corresponding to rankings $r_1, ..., r_{d-1}$ are linearly independent, we also add the following $\mathcal{EER}$-formulas to Θ:

$$a_2 \sim_{r_1} a_3 \sim_{r_1} a_4 \sim_{r_1} ... \sim_{r_1} a_{d-1} \sim_{r_1} a_d \prec_{r_1} a_1 \tag{12}$$
$$a_1 \sim_{r_2} a_3 \sim_{r_2} a_4 \sim_{r_2} ... \sim_{r_2} a_{d-1} \sim_{r_2} a_d \prec_{r_2} a_2 \tag{13}$$
$$a_1 \sim_{r_3} a_2 \sim_{r_3} a_4 \sim_{r_3} ... \sim_{r_3} a_{d-1} \sim_{r_3} a_d \prec_{r_3} a_3 \tag{14}$$
$$... \tag{15}$$
$$a_1 \sim_{r_{d-1}} ... \sim_{r_{d-1}} a_{d-2} \sim_{r_{d-1}} a_d \prec_{r_{d-1}} a_{d-1} \tag{16}$$

where $a_1, ..., a_d$ are fresh variables. For a formula of the form $Col(x, y, z)$ we finally add three additional $\mathcal{EER}$-formulas to Θ, to impose that x, y, z need to be distinct:

$$x \prec_{r_d} y \qquad y \prec_{r_{d+1}} z \qquad x \prec_{r_{d+2}} z \tag{17}$$

with r_d, r_{d+1}, and r_{d+2} fresh rankings. For a formula of the form $Bet(x, y, z)$, instead of (17), we add $x \prec_{r_d} y \prec_{r_d} z$. For a formula of the form $Gen(x, y, z)$, we only need to add the following three formulas to Θ:

$$x \prec_{r_1} y \prec_{r_1} z \qquad y \prec_{r_2} x \prec_{r_2} z \qquad x \prec_{r_3} z \prec_{r_3} y$$

with r_1, r_2 and r_3 fresh rankings. Finally, for a formula of the form $Par(w, x, y, z)$, we add the following formulas to Θ ($i \in \{1, ..., d-1\}$):

$$w \sim_{r_i} x \prec_{r_i} y \sim_{r_i} z$$

as well as $w \prec_{r_d} x$ and $y \prec_{r_{d+1}} z$. We can show the following result.

Proposition 12. *Let Ψ be a set of $\mathcal{L}$-formulas and let Θ be the corresponding set of $\mathcal{EER}$-formulas. For any $d \geq 2$ it holds that Ψ is realisable in $\mathbb{R}^d$ iff Θ is realisable in $\mathbb{R}^d$.*

From (1) and (2) it immediately follows that REER can be formulated as checking the consistency of a set of polynomial inequalities, and thus belongs to $\exists\mathbb{R}$. From Propositions 11 and 12, also it follows that REER in $\mathbb{R}^d$ is $\exists\mathbb{R}$-hard for $d \geq 2$. Hence we have the following result.

Corollary 13. *For any $d \geq 2$, REER in $\mathbb{R}^d$ is $\exists\mathbb{R}$-complete.*

4 An Iterative Method for Finding Solutions

Let Θ be a set of $\mathcal{EER}$ constraints without cycles. Finding a solution of Θ (or deciding that no solution exists) in $\mathbb{R}^d$ boils down to solving a set of inequalities of the form $v^T(q-p) \geq 0$ or $v^T(q-p) > 0$. While this can be accomplished, in principle, using generic solvers for polynomial inequalities over the reals, initial experiments revealed that solvers such as QEPCAD[2] and Redlog[3] do not scale beyond trivial instances.

Instead we propose a heuristic method, based on the observation that the inequalities can be efficiently solved using linear programming if either all direction vectors or all points are known. This gives rise to a procedure, inspired by the well-known Expectation-Maximisation method, which alternatingly (1) fixes the points and estimates optimal coordinates for the direction vectors, and then (2) fixes the direction vectors and estimates optimal coordinates for the points. Note that this procedure can only be used to find solutions of consistent problem instances (or approximate solutions of inconsistent instances), which is what is usually needed in applications. We now discuss the method in more detail.

Initialisation: A simple method for initialising the coordinates $(c_1, ..., c_d)$ of a point $\pi(x_i)$ is to sample each c_j from a uniform distribution. As a more informed alternative, we also consider a method which uses multi-dimensional scaling to assign the coordinates such that two points $\pi(x_i)$ and $\pi(x_j)$ are close to each other iff they appear in similar positions in the rankings. Specifically, we use topological sort to obtain a complete ranking of the variables $x_1, ..., x_n$ for each direction (assuming there are no cycles). This allows us to define a distance on the set of variables as $dist(x_i, x_j) = \sqrt{\sum_\ell (rank(x_i, \ell) - rank(x_j, \ell))^2}$, where $rank(x_i, \ell)$ is the position of variable x_i in the ℓ^{th} ranking. Multi-dimensional scaling (MDS) is then used to find coordinates $\pi(x_1), ..., \pi(x_n)$ in $\mathbb{R}^d$ such that the Euclidean distance between $\pi(x_i)$ and $\pi(x_j)$ best approximates $dist(x_i, x_j)$ for each x_i and x_j. Finally, we also consider a hybrid method, where we generate an initialisation using MDS as well as 49 random initialisations. For each of these 50 initialisations, we estimate the corresponding directions and we select the initialisation which satisfies most of the given $\mathcal{EER}$ formulas.

Estimating directions/coordinates: Assume the coordinates of $\pi(x_1), ..., \pi(x_n)$ are fixed, and let us write $\pi(x_i) = (c^i_1, ..., c^i_d)$. Then we can use linear programming (LP) to find suitable values for the coordinates of $\rho(1), ..., \rho(m)$, by minimizing $\sum_{i,j,k} e^{i,j}_k$ s.t.:

$$\sum_{l=1}^{d} (c^i_l - c^j_l) \cdot \lambda^k_l \leq -1 + e^{i,j}_k \qquad \forall (x_i \prec_k x_j) \in \Theta$$
$$\sum_{l=1}^{d} (c^i_l - c^j_l) \cdot \lambda^k_l \leq e^{i,j}_k \qquad \forall (x_i \preceq_k x_j) \in \Theta$$
$$e^{i,j}_k \geq 0 \qquad \forall 1 \leq i, j \leq n, 1 \leq k \leq m$$

[2] http://www.usna.edu/CS/qepcadweb

[3] http://redlog.eu

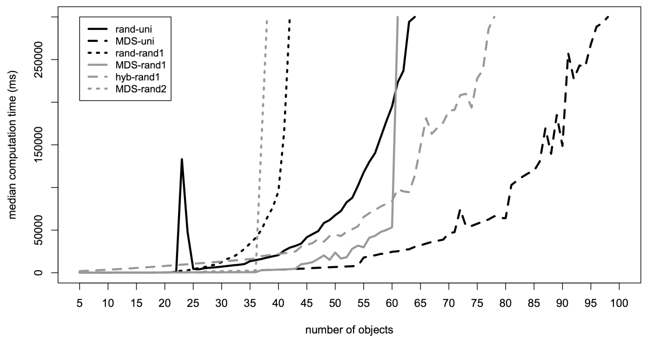

Figure 3: $\mathcal{EER}$ constraints in a semantic space of films.

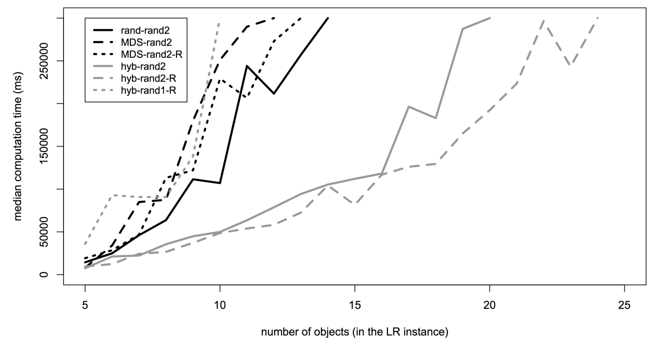

Figure 4: Results for the LR-LEFT-ALL benchmark.

It is easy to show that the objective function reaches its optimal value of 0 iff (π, ρ) is a d-dimensional solution of Θ, where $\pi(x_i)$ is defined as before and $\rho(i) = (\lambda_1^i, ..., \lambda_d^i)$. In the same way, we can estimate the coordinates of $\pi(x_1), ..., \pi(x_n)$ by taking the values λ_l^k as constants and the values c_l^i as variables.

We have experimented with a few variations of this approach. One possibility is to require that $e_k^{i,j} \in \{0, 1\}$ and use a Mixed Integer Programming (MIP) solver instead. This has the advantage that the solution of the linear program would explicitly maximise the number of satisfied $\mathcal{EER}$ formulas. However, MIP has a higher complexity than LP, and accordingly this approach was found not to be competitive in initial experiments. We will, however, consider a variant in which the objective function is replaced by

$$\textit{minimise} \sum \{ (\mu_k^{i,j})^N \cdot e_k^{i,j} \mid (x_i \prec_k x_j) \in \Theta \} \tag{18}$$

where each constant $\mu_k^{i,j}$ is sampled from a uniform distribution over $[0, 1]$ and $N \geq 1$ is a parameter. Because in each step new values $\mu_k^{i,j}$ are used, the method will prioritise different $\mathcal{EER}$ formulas each time and thus avoid getting stuck in local optima. The parameter N controls the exploitation-exploration trade-off by determining to what extent the method prioritises a small number of $\mathcal{EER}$ formulas.

Finally, in some variants we reset the values of $\pi(x_i)$ to the best values encountered so far if there has not been any improvement for T consecutive iterations (we used $T = 250$).

5 Experimental Results

We have used CBC[4] for solving the linear programs, with a timeout of 300 seconds; all reported values are the median of 50 executions. In Figures 3 and 4, we write *rand*, *MDS* and *hyb* to denote configurations where the initialisation used random values, MDS, or the hybrid method. We write *uni*, *rand1* and *rand2* to indicate whether the uniform objective or the randomised objective (18) was used, and in the latter case, whether we set $N = 1$ or $N = 2$. Finally, names ending with *-R* denote configurations where we reset the values of $\pi(x_i)$ after 250 iterations without improvement.

For the results in Figure 3, we have considered a 20-dimensional space of films, for which 40 directions have been obtained using the method from [Derrac and Schockaert, 2014a]. From this space, we generate $\mathcal{EER}$ instances with 40 rankings and a variable number of objects. In Figure 3 we observe that (i) MDS substantially outperforms random initialisations and performs better than the hybrid method for sufficiently small problem instances, (ii) $N = 1$ is better than $N = 2$, and (iii) the uniform objective outperforms the randomised objective. We found that resetting the values of $\pi(x_i)$ after 250 iterations makes no difference (results not shown).

Figure 4 illustrates the effectiveness of our reduction from the $\mathcal{LR}$ calculus. The problem we consider is the LR-LEFT-ALL benchmark problem that was proposed in [Lee and Wolter, 2011]. To the best of our knowledge, our method is the first that can find solutions for problem sizes larger than 5 (e.g. Mathematica, QEPCAD and Redlog could not find a solution for size 6, even when given several hours). Interestingly, [van Delden and Mossakowski, 2013] presents a heuristic method that can derive inconsistencies in sets of $\mathcal{LR}$-formulas. This means that a general purpose $\mathcal{LR}$ solver could be obtained by combining our method with the method from [van Delden and Mossakowski, 2013] in a portfolio solver.

In Fig. 4, we observe that $N = 2$ leads to better results than $N = 1$, and that using the hybrid initialisation and resetting the values of $\pi(x_i)$ after 250 iterations makes a substantial difference. The different conclusions we can draw from Figures 3 and 4 reflect the different nature of the problem instances: the problems in Figure 3 involve a larger number of variables[5] and dimensions, while the problems from Figure 4 appear computationally harder.

6 Conclusions

We have studied the problem of reasoning about direction relations in high-dimensional spaces. Our primary motivation was to support a form of commonsense reasoning about relative properties, based on the view that relative properties such as 'more romantic than' correspond to direction relations in a suitable semantic space. Even though the proposed framework is quite simple, we have shown that several types

[4]https://projects.coin-or.org/Cbc

[5]Note however that Fig. 4 only refers to the number of $\mathcal{LR}$ objects, not including the variables introduced in the translation.

of qualitative spatial relations can be reduced to our framework, including $\mathcal{LR}$ relations and relations that express betweenness, collinearity and parallelism. As a result we obtained that the consistency problem in our framework is $\exists\mathbb{R}$-complete. Finally, we have proposed a heuristic method for finding solutions, based on linear programming, and experimentally showed that it scales sufficiently well to be useful in practice.

7 Acknowledgment

Steven Schockaert has been supported by an ERC Starting Grant (637277).

References

[Carlson *et al.*, 2010] Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka, Jr., and Tom M. Mitchell. Coupled semi-supervised learning for information extraction. In *Proc. of the 3rd ACM International Conf. on Web Search and Data Mining*, pages 101–110, 2010.

[Davis *et al.*, 1999] E. Davis, N.M. Gotts, and A.G. Cohn. Constraint networks of topological relations and convexity. *Constraints*, 4:241–280, 1999.

[Deerwester *et al.*, 1990] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[Derrac and Schockaert, 2014a] J. Derrac and S. Schockaert. Characterising semantic relatedness using interpretable directions in conceptual spaces. In *Proc. of the 21st European Conference on Artificial Intelligence*, pages 243–248. 2014.

[Derrac and Schockaert, 2014b] J. Derrac and S. Schockaert. Enriching taxonomies of place types using Flickr. In *Proc. of the 8th International Symposium on Foundations of Information and Knowledge Systems*, pages 174–192. 2014.

[Gärdenfors, 2000] P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000.

[Hearst, 1992] Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics*, pages 539–545, 1992.

[Kovashka *et al.*, 2012] A. Kovashka, D. Parikh, and K. Grauman. Whittlesearch: Image search with relative attribute feedback. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2973–2980, 2012.

[Lee and Wolter, 2011] Jae Hee Lee and Diedrich Wolter. A new perspective on reasoning with qualitative spatial knowledge. In *Proceedings of IJCAI-2011 Workshop on Benchmarks and Applications of Spatial Reasoning*, 2011.

[Lee, 2013] Jae Hee Lee. *Qualitative Reasoning about Relative Directions: Computational Complexity and Practical Algorithm*. PhD thesis, Universität Bremen, 2013.

[Ligozat, 1993] Gérard F. Ligozat. Qualitative triangulation for spatial reasoning. In *Proceedings of the European Conference on Spatial Information Theory*, pages 54–68. 1993.

[Miclet *et al.*, 2008] Laurent Miclet, Sabri Bayoudh, and Arnaud Delhay. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *Journal of Artificial Intelligence Research*, 32:793–824, 2008.

[Mikolov *et al.*, 2013] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of NAACL-HLT*, pages 746–751, 2013.

[Nakashole *et al.*, 2011] Ndapandula Nakashole, Martin Theobald, and Gerhard Weikum. Scalable knowledge harvesting with high precision and high recall. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pages 227–236, 2011.

[Nakashole *et al.*, 2012] N. Nakashole, G. Weikum, and F. Suchanek. PATTY: A taxonomy of relational patterns with semantic types. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1135–1145, 2012.

[Nickel *et al.*, 2012] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. Factorizing YAGO: Scalable machine learning for linked data. In *Proceedings of the 21st International Conference on World Wide Web*, pages 271–280, 2012.

[Schaefer, 2010] Marcus Schaefer. Complexity of some geometric and topological problems. volume 5849 of *Lecture Notes in Computer Science*, pages 334–344. 2010.

[Schockaert and Prade, 2011] Steven Schockaert and Henri Prade. Interpolation and extrapolation in conceptual spaces: A case study in the music domain. In *Proceedings of the 5th International Conference on Web Reasoning and Rule Systems*, pages 217–231, 2011.

[Schockaert and Prade, 2013] Steven Schockaert and Henri Prade. Interpolative and extrapolative reasoning in propositional theories using qualitative knowledge about conceptual spaces. *Artificial Intelligence*, 202:86–131, 2013.

[Scivos and Nebel, 2005] Alexander Scivos and Bernhard Nebel. The finest of its class: The natural point-based ternary calculus $\mathcal{LR}$ for qualitative spatial reasoning. In *Spatial Cognition IV. Reasoning, Action, Interaction*, pages 283–303. Springer Berlin Heidelberg, 2005.

[Sheremet *et al.*, 2010] Mikhail Sheremet, Frank Wolter, and Michael Zakharyaschev. A modal logic framework for reasoning about comparative distances and topology. *Annals of Pure and Applied Logic*, 161(4):534 – 559, 2010.

[Speer *et al.*, 2008] Robert Speer, Catherine Havasi, and Henry Lieberman. Analogyspace: reducing the dimensionality of common sense knowledge. In *Proceedings of the 23rd AAAI Conference on Artificial intelligence*, pages 548–553, 2008.

[Turney and Pantel, 2010] P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188, 2010.

[van Delden and Mossakowski, 2013] André van Delden and Till Mossakowski. – mastering left and right – different approaches to a problem that is not straight forward. In *Proceedings of the 36th Annual German Conference on AI*, pages 248–259. 2013.

[Vig *et al.*, 2012] Jesse Vig, Shilad Sen, and John Riedl. The tag genome: Encoding community knowledge to support novel interaction. *ACM Transactions on Interactive Intelligent Systems*, 2(3):13:1–13:44, 2012.

[Williamson, 1988] Timothy Williamson. First-order logics for comparative similarity. *Notre Dame Journal of Formal Logic*, 29(4):457–481, 1988.