Cone Semantics for Logics with Negation

Özgür Lütfü Özçep[1] , **Mena Leemhuis**[1] and **Diedrich Wolter**[2]

[1]University of Lübeck, Germany
[2]University of Bamberg, Germany

oezcep@ifis.uni-luebeck.de, mena.leemhuis@student.uni-luebeck.de, diedrich.wolter@uni-bamberg.de

Abstract

This paper presents an embedding of ontologies expressed in the $\mathcal{ALC}$ description logic into a real-valued vector space, comprising restricted existential and universal quantifiers, as well as concept negation and concept disjunction. Our main result states that an $\mathcal{ALC}$ ontology is satisfiable in the classical sense iff it is satisfiable by a partial faithful geometric model based on cones. The line of work to which we contribute aims to integrate knowledge representation techniques and machine learning. The new cone-model of $\mathcal{ALC}$ proposed in this work gives rise to conic optimization techniques for machine learning, extending previous approaches by its ability to model full $\mathcal{ALC}$.

1 Introduction

The idea of embedding words into low-dimensional continuous vector spaces has been implemented successfully in various algorithms with various applications in the realm of information retrieval [Goldberg and Levy, 2014; Pennington *et al.*, 2014; Levy and Goldberg, 2014]. However, these approaches are insensitive to the relational—or more generally: to the predicate-logical—structure of documents. The embedding idea was pushed further (see, e.g., Nickel *et al.*; Bordes *et al.* [2011; 2013] and, for an overview, Wang *et al.* [2017]) in order to design embeddings of knowledge graphs, i.e., of sets of triples of the form $(a\ R\ b)$ for objects a and b and relations R, or even embeddings of knowledge graphs with additional background knowledge, say an ontology consisting of axioms in some (expressive) logic [Mehran Kazemi and Poole, 2018; Kulmanov *et al.*, 2019].

Many classical approaches of knowledge graph embeddings such as TransE [Bordes *et al.*, 2013] suffer from a lack of full expressivity in the sense laid down by Mehran Kazemi and Poole [2018]: given a knowledge graph and a set of triples known to be true (positive set) as well as triples that are known to be false (negative set), the embedding is said to be *fully expressive* if it maps the relations and constants into a space such that $(a\ R\ b)$ holds in the embedding iff it is in the positive set and $(a\ R\ b)$ does not hold iff it is in the negative set. The lack of (full) expressivity of classical approaches to knowledge graph embeddings is due to a decision on how to

model relations R: rather than following the classical logic approach, relations are modelled with computationally feasible data structures (e.g., in case of TransE by vector translations) that fit into the general mathematical framework of continuous embeddings. In fact, following the translation approach of TransE one can see that the induced logic is not that of arbitrary relations but that of functional relations.

These observations motivate the research on embeddings that give a better compromise between the geometrical models that can be constructed by means of learning and the demands of ontologies. This paper proposes a new embedding of ontologies expressed in the $\mathcal{ALC}$ description logic into a real-valued vector space, providing restricted forms of existential and universal quantifiers as well as concept negation and concept disjunction. Our main result states that an $\mathcal{ALC}$ ontology is satisfiable in a classical sense iff it is satisfiable by a geometric model that interprets all concept descriptions as cones (concretely: axis-aligned cones). We derive this result by first considering the Boolean part of $\mathcal{ALC}$ ontologies and then generalizing this to $\mathcal{ALC}$ with concepts up to a fixed quantifier rank. The geometric models we use are partial and thus allow some uncertainty to be retained, i.e., if x is only known to be a member of the union of two atomic concepts, then our partial model will not commit to saying to which atomic concept x belongs. Put differently, we need partial models in order to represent the knowledge contained in the ontology *faithfully*: exactly those axioms derivable from the ontology should be represented in a partial model.

This paper continues by reviewing related approaches. Section 3 summarizes $\mathcal{ALC}$, Section 4 introduces al-cones and discusses their motivation. In Section 5 we present an embedding for propositional $\mathcal{ALC}$ and show its completeness. Section 6 then considers the non-propositional case. The paper concludes with a brief discussion of the results.

2 Related Work

When combining knowledge representation (KR) and machine learning (ML) techniques, ML algorithms are either used to learn an ontology or they exploit the ontologies as constraint specifications in order to optimize statistical models. One recent approach along this line that is based on the geometric interpretation of knowledge graph embeddings is presented in Kulmanov *et al.* [2019]. Here, the lightweight logic $\mathcal{EL}$++ is considered. The method is based on a geo-

metric interpretation of TransE [Bordes *et al.*, 2013], which interprets objects as vectors and relations as translations of these vectors. A concept is in this case an open n-ball with a fixed radius and relations are translations of the n-balls. The method is limited due to the restricted constructions available in $\mathcal{EL}++$, which in particular has no full concept negation.

In Gutiérrez-Basulto and Schockaert [2018], geometric interpretations are defined as classical interpretations of an ontology with two specific constraints: the domain is a Euclidean space of some dimension m and the interpretations of all (n-ary) relations are constrained to convex sets over n-wise cartesian products $\mathbb{R}^{nm}$ of $\mathbb{R}^m$. The main difference to classical knowledge graph approaches is the idea of interpreting n-ary relations (classically) by n-wise cartesian products and (non-classically) requiring them to be convex sets. Their approach achieves ontologies expressed as rules in datalog$^\pm$ [Calì *et al.*, 2009], which admits rules with existentials in the head of the rule (alias tuple-generating dependencies) and integrity constraints, i.e., rules of the form $\forall \vec{x} \psi(\vec{x}) \to \bot$ with a conjunction of atoms $\psi(\vec{x})$. Their main result is that an ontology under some specific constraint (quasi-chainedness) is satisfiable classically iff it is satisfiable by a geometric interpretation where all relations are interpreted by convex sets. The use of convex sets for the interpretation of relations can be justified by their importance as, on the one hand, computationally feasible data structures as used in convex optimization [Boyd and Vandenberghe, 2004] and, on the other hand, as a linguistically and cognitively justified structure for representing concepts [Gärdenfors, 2000].

Convexity is preserved under many operations, in particular it is preserved under intersection and projection, which are the main operations expressible in the allowed fragment of Gutiérrez-Basulto and Schockaert [2018]. Interestingly, due to the use of integrity constraints, these ontologies implicitly use some aspect of negation, namely that of disjointness. But this is not full negation: implicitly, negation is allowed to occur on the righthand side as atomic negation, but not as negation on the lefthand side of a rule. In the latter case negation allows us to express coverage aspects which amounts to the use of disjunction. And this restriction in the language is not surprising as convexity is neither preserved under set complement nor under set union.

In our approach, we try to keep the cake (stick to convexity) and eat it (allow for negation). Our approach follows that of Gutiérrez-Basulto and Schockaert [2018] in interpreting relations as tuples over a domain. Our approach also presumes finite satisfiability of the ontology but works dually by constructing concepts on the axes and then placing individuals on these. The reason is that we incorporate an additional structure, namely a scalar product, which in turn induces the negation operator. And this one constrains the potential places in which the negations of concepts can be placed. This, in particular, prevents adapting the quasi-chainedness property which was defined for datalog$^\pm$ and not $\mathcal{ALC}$.

3 The Description Logic $\mathcal{ALC}$

We are going to work with the description logic $\mathcal{ALC}$ [Baader, 2003]. We assume that there is a DL vocabulary

Name	Syntax	Semantics
top	$\top$	$\Delta^\mathcal{I}$
bottom	$\bot$	$\emptyset$
conjunction	$C \sqcap D$	$C^\mathcal{I} \cap D^\mathcal{I}$
disjunction	$C \sqcup D$	$C^\mathcal{I} \cup D^\mathcal{I}$
negation	$\neg C$	$\Delta^\mathcal{I} \setminus C^\mathcal{I}$
universal quantifier	$\forall R.C$	$\{x \in \Delta^\mathcal{I} \mid$ For all $y \in \Delta^\mathcal{I}:$ If $(x,y) \in R^\mathcal{I}$ then $y \in C^\mathcal{I}\}$
existential quantifier	$\exists R.C$	$\{x \in \Delta^\mathcal{I} \mid$ There is $y \in \Delta^\mathcal{I}$ $(x,y) \in R^\mathcal{I}$ and $y \in C^\mathcal{I}\}$

Table 1: Syntax and semantics for the DL $\mathcal{ALC}$

given by a set of constants N_c, a set of role names N_R and concept names N_C. The $\mathcal{ALC}$ concepts (concept descriptions) over $N_C \cup N_R$ are described by the grammar

$$C \longrightarrow A \mid \bot \mid \top \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \exists R.C \mid \forall R.C$$

where $A \in N_C$ is an atomic concept, $R \in N_R$ is a role symbol, and C stands for arbitrary concepts. A classical $\mathcal{ALC}$ *interpretation* is a pair $(\Delta, (\cdot)^\mathcal{I})$ consisting of a set Δ, called the *domain*, and an *interpretation function* $(\cdot)^\mathcal{I}$ which maps constants to elements in Δ, concept names to subsets of Δ, and role names to subsets of $\Delta \times \Delta$. The semantics of arbitrary concept descriptions for a given interpretation $\mathcal{I}$ is given in Table 1. An *ontology* $\mathcal{O}$ is defined as a pair $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ of a *terminological box (tbox)* $\mathcal{T}$ and an *assertional box (abox)* $\mathcal{A}$. A tbox consists of *general inclusion axioms (GCIs)* $C \sqsubseteq D$ ("C is subsumed by D") with concept descriptions C, D. An abox consists of a finite set of *assertions*, i.e., facts of the form $C(a)$ or of the form $R(a, b)$ for $a, b \in N_c$. An interpretation $\mathcal{I}$ *models* a GCI $C \sqsubseteq D$, for short $\mathcal{I} \models C \sqsubseteq D$, iff $C^\mathcal{I} \subseteq D^\mathcal{I}$. An interpretation $\mathcal{I}$ *models* an abox axiom $C(a)$, for short $\mathcal{I} \models C(a)$, iff $a^\mathcal{I} \in C^\mathcal{I}$ and it models an abox axiom of the form $R(a, b)$ iff $(a^\mathcal{I}, b^\mathcal{I}) \in R^\mathcal{I}$. An interpretation is a *model of an ontology* $(\mathcal{T}, \mathcal{A})$ iff it models all axioms appearing in $\mathcal{T} \cup \mathcal{A}$. An ontology $\mathcal{O}$ *entails* a (tbox or abox) axiom ax, for short $\mathcal{O} \models ax$, iff all models of $\mathcal{O}$ are also models of ax. The *quantifier rank* for arbitrary concepts C is defined as the maximal nesting of quantifiers in it. Formally, for an $\mathcal{ALC}$ concept define the quantifier rank by recursion as $qr(A) = 0$, $qr(\neg C) = qr(C)$, $qr(C_1 \wedge C_2) = max\{qr(C_1), qr(C_2)\}$, and $qr(\exists R.C) = qr(\forall R.C) = qr(C) + 1$.

Each tbox $\mathcal{T}$ generates a Boolean algebra, the so-called *Lindenbaum-Tarski algebra*, as follows: For concepts C, D let $\sim$ be the relation defined by $C \sim D$ iff $\mathcal{T} \models C \sqsubseteq D$ and $\mathcal{T} \models D \sqsubseteq C$. Relation $\sim$ is an equivalence relation inducing for each concept C an equivalence class $[C]$. Define operations $\sqcap, \sqcup, \neg$ on the equivalence classes by setting $[C] \sqcap [D] = [C \sqcap D]$, $[C] \sqcup [D] = [C \sqcup D]$ and $\neg[C] = [\neg C]$ which can be shown to fulfil the axioms of a Boolean algebra.

4 Al-Cone Models

Our geometric interpretations are based on axis-aligned cones (al-cones) which form a subclass of the class of convex cones. A *convex cone* X is a set that fulfils the following property: If $v, w \in X$, then also $\lambda v + \mu w \in X$ for all $\lambda, \mu \geq 0$. The *polar cone* X° for X is defined for Euclidean spaces with a scalar (dot) product $\langle \cdot, \cdot \rangle$ as follows:

$$X^\circ = \{v \in \mathbb{R}^n \mid \forall w \in X : \langle v, w \rangle \leq 0\}$$

In the sequel, we use the term "cone" to refer to convex cones. Important cones are the so-called n-orthants or n-hyperoctants. For dimension $n = 2$ these are the 4 quadrants. Using the abbreviations $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$ and $\mathbb{R}_- = \{x \in \mathbb{R} \mid x \leq 0\}$ we can define the 2^n n-dimensional hyperoctants $H_{(b_1,\dots,b_n)}$ with $b_i \in \{+, -\}$ by $\mathbb{R}_{b_1} \times \cdots \times \mathbb{R}_{b_n}$.

Cones can be constructed by unions of neighboring hyperoctants. Consider next to $+$, $-$, also the value u (for union) and define $\mathbb{R}_u = \mathbb{R}$. Then the axis aligned unions of hyperoctants are defined as sets $\mathbb{R}_{b_1} \times \cdots \times \mathbb{R}_{b_n}$ with $b_i \in \{+, -, u\}$. By further allowing for projections of all the sets mentioned above to the axes of the space we get the *al-cones*.

X is al-cone $:\Leftrightarrow X = X_1 \times \cdots \times X_n, X_i \in \{\mathbb{R}, \mathbb{R}_+, \mathbb{R}_-, \{0\}\}$

So, in n dimensions we have 4^n possible al-cones.

Let us now see how Boolean operations can be modeled. Aside set-based intersection to represent conjunction, special attention is required. All al-cones are convex cones. Convex cones are preserved under intersection, polarity, and other operations (such as projection), but not under the set-union operator. This applies to al-cones as well, but by interpreting $\sqcup$ using de Morgan we can circumvent that problem.

Proposition 1. *Al-cones X, Y in $\mathbb{R}^n$ are closed as follows:*

1. *The intersection $X \cap Y$ of X and Y is an al-cone.*

2. *The polar X° is an al-cone.*

3. $(X^\circ)^\circ = X$

4. $(X^\circ \cap Y^\circ)^\circ$, *the convex union of X, Y, is an al-cone.*

Proof sketch. (1) holds as X_i in the al-cone definition is closed under intersection, (2) can be shown by contradiction: Consider one component X_i, suppose it is $\mathbb{R}_+$ (other cases follow similarly). Let $v \in X$ be of the form $(0, 0, \dots, x_i, 0, \dots)$. Since X° should contain exactly those vectors u s.t. $\langle u, v \rangle \leq 0$ for any $v \in X$, clearly $(X^\circ)_i$ the i-th component of the polar of X, for short $(X^\circ)_i$, cannot contain $\mathbb{R}^+$, which disqualifies it from being either $\mathbb{R}$ or $\mathbb{R}^+$. On the other hand, $(X^\circ)_i$ must contain the entire $\mathbb{R}_-$ because for any non-positive u_i the defining condition holds. (3) is known to hold for arbitrary closed convex cones [Rockafellar, 1997, pp. 121–122], (4) immediately follows from (1), (2). $\square$

Before we further embark on technicalities let us motivate the specifics of our approach by discussing two questions: First, why should one use the polarity operator as the interpretation of concept negation? Second, why should we restrict ourselves to axis-aligned cones?

4.1 Motivation of Polarity-Based Negation

The use of the polarity operation for concept negation is motivated by the idea of providing an operator that always maps a concept to a disjoint concept such that the disjoint concept is maximally so w.r.t. the underlying similarity structure. Here, the similarity structure is given by the usual scalar product in the Euclidean space. Two reasons substantiate our choice.

As for the first reason, one can observe that the idea of considering polarity as a form of negation, is related to a general approach of defining negation on the basis of some binary complementary-relation com, as explained, e.g., in

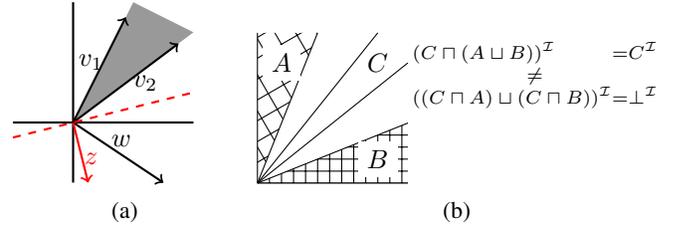Figure 1: Illustration of Farkas' Lemma (a) and counterexample distributivity for arbitrary cones (b)

Dunn [1996] for propositional logics: relation com holds between two propositions iff they are complementary in the sense that there is no situation where both propositions are true (but still both may be false). Then, negation of a proposition p can be defined as the disjunction of all those propositions q that are complementary to p. Of course we are interested here in reading subsets C of the embedding space $\mathbb{R}^n$ as concepts and points in $\mathbb{R}^n$ as objects falling into the extension of C. But, drawing a closer connection to the approach in Dunn [1996], one could equally think of the points of $\mathbb{R}^n$ as possible worlds (or states). Then set C becomes a proposition which is exactly true for points contained in it.

As for the second reason, one can consider the following fact, known as Farkas' Lemma (see Figure 1a).

Lemma 1 (Farkas' Lemma). *Let C be the convex cone generated by vectors $v_1, \dots, v_n \in \mathbb{R}^n$ (i.e., C = smallest convex cone containing all v_i) and let $w \in \mathbb{R}^n$. Then either $w \in C$ or there is $z \in \mathbb{R}^n$ such that $\langle z, v_i \rangle \leq 0$ for all i, and $\langle z, w \rangle > 0$.*

The lemma says that if one considers a vector w that is not in the convex cone, i.e., it is in the complement $\mathbb{R}^n \setminus C$ (the "usual" negation of C), then there is at least a verifier z that is similar to w (namely $\langle z, w \rangle > 0$) and is contained in the polar cone of C. Note that from $\langle z, v_i \rangle \leq 0$ for all i it follows that $\langle z, v \rangle \leq 0$ for all $v \in C$. In the other direction, if one can find a z of the polar cone being similar to w, then w is in the complement of C. In conclusion, polarity can be considered to provide a more intuitive model of negation for scalar-product based similarity structures than set complement.

4.2 Motivation for Considering Al-Cones

Firstly, the usefulness of al-cone models hinges on whether they are able to represent an interesting class of ontologies. In the remainder of this paper we show that al-cone models are indeed complete in the sense that an ontology is satisfiable classically iff it can be embedded into a geometric model based on al-cones.

Secondly, al-cones are motivated by their ability to link to ML. So far we have motivated the use of convex cones as proper structures to handle negation, yet we emphasize that convex cones also are used as computationally feasible data structures in the area of conic optimization (see, e.g., Boyd and Vandenberghe [2004], Section 4.3 on linear optimization problems and Section 4.4.2 on second-order cone programming) and therefore attractive for ML application. Here we only regard the case where one aims to learn a statistical model for data that can be characterized by some ontology

that has been specified in a logic beforehand, not the case of investigating logics induced by the intersection and polarity operators for arbitrary cones. In fact, the resulting logic cannot be guaranteed to lead to Boolean (and so not to full) $\mathcal{ALC}$ as it would not fulfil the distribution property for $\sqcap$ and $\sqcup$. A simple example is given in Figure 1b.

Let us sketch a specific incremental learning scenario. We are given an $\mathcal{ALC}$ tbox. The training data are points from $\mathbb{R}^n$ with a concept label or pairs of points with a role label. The dimension n is fixed in advance and represents features. For ease of exposition we assume that the number n equals the number of atoms in the Boolean algebra over the tbox. In this scenario the hypothesis set is chosen as al-cones in $\mathbb{R}^n$ and all other set systems of cones that result from rotating the al-cones w.r.t. the same (arbitrarily chosen) angle.

On receiving some training datum, the algorithm is required to choose some skeleton of al-cones and possibly rotate it with some angle α in order to fit the datum. We do not specify here how to choose the α, as there may be various possible rotations that can be chosen following an optimization principle, say. In the training phase, the algorithm may not be able to embed the data to any rotated skeleton. In this case, the algorithm would stop due to non-resolvable inconsistencies. These can be inconsistencies due to a too small feature dimension n chosen in the beginning or w.r.t. to our assumption to consider only rotations of al-cones and not other al-cone preserving transformations such as warping or due to inconsistencies w.r.t. the ontology. As the following propositions show, the inconsistency cannot be due to the fact that concepts are represented as al-cones.

5   Embedding for Propositional $\mathcal{ALC}$

Let us start by considering $\mathcal{ALC}$ ontologies where the tbox language amounts to propositional logic. That is, we first restrict definitions from Section 3 to the concept constructors $\sqcap, \sqcup, \neg$ corresponding to the Boolean operations $\wedge, \vee, \neg$. A Boolean $\mathcal{ALC}$ tbox (abox) consists of GCIs (abox axioms) using Boolean concepts only.

Let $(\mathcal{T}, \mathcal{A})$ be a Boolean $\mathcal{ALC}$ ontology. We are going to define a geometric interpretation of a special kind that can be a model or an anti-model of a Boolean ontology. It is an ordinary model in the sense that it is a classical predicate logical structure with a domain and interpretations for the atomic concept symbols. It is special in the sense that the domain is of a specific structure, namely a Euclidean space and concept names are interpreted by al-cones and their projections to subspaces. It is non-classical in the sense that some logical constructors are not interpreted by the corresponding set operations, but by operations on al-cones. The interesting aspect is now that logical aspects of the Boolean algebra provide global constraints on the model. These can be satisfied by choosing appropriate positions of the interpretations for all atomic concepts. As a consequence, an ontology is satisfiable classically iff it is satisfied by the geometric interpretation.

Definition 1. *A* Boolean al-cone interpretation $\mathcal{I}$ *is a structure* $(\Delta, (\cdot)^{\mathcal{I}})$ *where* Δ *is* $\mathbb{R}^n$ *for some* $n \in \mathbb{N}$, *and where* $(\cdot)^{\mathcal{I}}$ *maps each concept symbol A to some al-cone and each constant a to some element in* $\Delta \setminus \{\vec{0}\}$. *An* al-cone interpre-
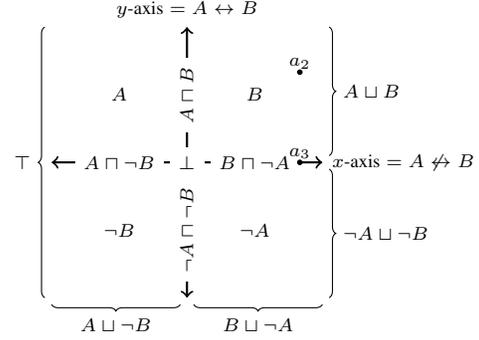

Figure 2: Al-cone model for a simple abox with empty tbox

tation for arbitrary Boolean $\mathcal{ALC}$ concepts *is defined recursively as* $(\top)^{\mathcal{I}} = \Delta$, $(\bot)^{\mathcal{I}} = \{\vec{0}\}$, $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$, $(\neg C)^{\mathcal{I}} = C^{\circ}$, *and* $(C \sqcup D)^{\mathcal{I}} = (\neg(\neg C \sqcap \neg D))^{\mathcal{I}}$. *The notions of an al-cone being a model and that of entailment are defined as in the classical case (but using al-cone interpretations).*

Let us consider a simple example of a Boolean ontology, consisting of an empty tbox and an abox constructed as follows: Consider the Lindenbaum-Tarski algebra of all Boolean concepts defined on two Boolean names, say A, B. Then the generated Boolean algebra of classes of equivalent concepts has $2^{2^2} = 16$ elements. Choose for each a "smallest" representative C_i, $i \in \{1, \ldots, 16\}$ w.r.t. some ordering. C_1 is the bottom concept and let us assume that C_2 is the concept B and C_3 is the concept $B \sqcap \neg A$. Now, for each i with $n \geq i \geq 2$ take a constant a_i and add abox axioms $C_i(a_i)$.

Thus we represent each of the 15 non-bottom concepts uniquely by some constant. For example, in Figure 2 the constant a_2 represents the concept $C_2 = B$. This ontology is satisfiable classically. There is an al-cone interpretation that fulfills the ontology in $\mathbb{R}^2$ that is constructed as follows: interpret A by the left upper quadrant and B by the right upper quadrant. This induces uniquely the positions of all other hyperoctants corresponding to the other concepts C_i. The interesting point is the localization of constants a_i. For each constant a_i one localizes the hyperoctant corresponding to C_i in the area and positions on it such that it does not coincide with one of the points already associated with a constant.

Figure 2 illustrates the positions of the concepts: $\bot$ is by definition the singleton $\{\vec{0}\}$, A is the left upper quadrant, $\top$ is the whole area, etc. One can check that the concepts are associated with appropriate al-cones. For example, the negation $\neg A$ of A is indeed the polar cone of the quadrant of A. Similarly, consider $B \sqcap \neg A$, which is interpreted as the positive x-axis $\mathbb{R}_+ \times \{0\}$. Its polar cone is the whole left area.

As another example consider the Boolean algebra induced by the atomic concepts A, B under the tbox axiom $A \sqsubseteq B$. Then, e.g., $A \sqcap \neg B \sqsubseteq \bot$ and so forth. This gives an embedding with axis-aligned cones as illustrated in Figure 3.

Using the construction idea of the examples one can show that $\mathcal{ALC}$-ontologies are classically satisfiable iff they are by a geometric model based on al-cones.

Proposition 2. *Boolean $\mathcal{ALC}$-ontologies are classically satisfiable iff they are by a geometric model on some finite* $\mathbb{R}^n$
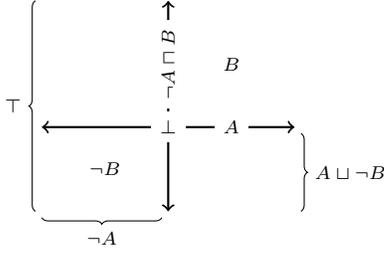
Figure 3: Al-cone model for tbox $\{A \sqsubseteq B\}$

based on al-cones of the form $b_1 \times \cdots \times b_n$ with $b_i \in \{\{0\}, \mathbb{R}_+, \mathbb{R}_-, \mathbb{R}\}$ for $i \in \{1, \ldots, n\}$.

Proof sketch. For the general construction we use the encoding from the examples. The set of encodings $\{u, +, -, 0\}$ used in the example forms a diamond shaped Boolean algebra B_4 with $\bot \leq +, - \leq \top$. As the $\mathcal{ALC}$-ontology $\mathcal{O}$ the induced Boolean algebra is finite, say 2^k, with a finite number of atoms m. We have to chose a dimension n such that $4^n \geq 2^k$ and $n \geq m/2$, the idea being that we have to represent all of the 2^k Boolean concepts by embedding the atoms of the Boolean algebra directly onto the half-axes of $\mathbb{R}$. Now one can see that B_4^n is the n-wise Cartesian product of B_4 where $\sqcap, \sqcup, \neg$ are defined component-wise based on the $\sqcap$, $\sqcup, \neg$ defined on B_4. But in this way $\sqcap$ over B_4 is nothing else than set intersection, $\neg$ is polarity, and $\sqcup$ is defined by de Morgan. So, if $\mathcal{O}$ is satisfiable, then we can find a geometric model of it.

"$\leftarrow$": Given a geometric model, one can build a classical model by "forgetting" any named object that is not on an axis by moving it outside to one of the axes that cover the al-cone. $\square$

Geometric models are more specific than classical models in the sense that they impose an underlying structure on the domain given by, firstly, the dimensions representing some latent features that usually are not even mentioned in the ontology and, secondly, a scalar product $\langle \cdot, \cdot \rangle$ over the space. Geometric interpretations are more general than classical interpretations in the sense that they are partial interpretations. Actually, when embedding an ontology into some space one expects the embedding to represent not a single, but all interpretations that make the ontology true. And this is (up to some exceptions, see below) the case in the construction of Proposition 2. This allows partial information to be encoded. Consider, e.g., the difference between a_2 and a_3 in the geometric model of Fig. 2. The individual a_3 is completely identified w.r.t. the given concepts A, B: it lies in the extension of B and in the extension of $\neg A$. For a_2 we "only" know that it must be an B, but we do not know whether it is also an A. Hence this geometric model correctly reflects this specific knowledge provable in the ontology: one can prove that a_2 is a B but one cannot prove that it is an A nor that it is a $\neg A$.

The geometric model illustrated in Fig. 2 does not reflect the whole knowledge (not) contained in an ontology in the correct way. For example, it is not possible to represent an object b which is known to be in $A \sqcup B$, but neither known

to be an A nor known to be a B. The reason is that by our construction no place is left to reflect this partial knowledge.

The discussion above motivates the definition of faithful geometric models of a given ontology. As within ontologies one has to deal with many models there are actually two different adaptations, a weak one and a strong one. Orthogonally to this distinction one may also consider the faithfulness w.r.t. the abox only (to the data only) or to the tbox. An even finer distinction can be given by considering special subclasses of the abox (which we will do here by considering concept assertions vs. role assertions) and the tbox (which we will do here by considering concepts with a bounded quantifier rank).

Definition 2. *Let $\mathcal{O}$ be a classically consistent (DL) ontology (or any other representation allowing the distinction between abox and tbox). For a geometric interpretation $\mathcal{I}$ we have the following notions of being a faithful model for $\mathcal{O}$: $\mathcal{I}$ is a strongly concept-faithful model of $\mathcal{O}$ iff for each concept C and each constant b: if $b^{\mathcal{I}} \in C^{\mathcal{I}}$, then $\mathcal{O} \models C(b)$; $\mathcal{I}$ is a weakly concept-faithful model of $\mathcal{O}$ iff for each concept C and each constant b: if $b^{\mathcal{I}} \in C^{\mathcal{I}}$, then $\mathcal{O} \cup \{C(b)\}$ is satisfiable classically; $\mathcal{I}$ is a strongly (weakly) abox-faithful model of $\mathcal{O}$ iff it is strongly (weakly) concept-faithful and for each role R and constants a, b: if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$, then $\mathcal{O} \models R(a, b)$ (resp. $\mathcal{O} \cup \{R(a, b)\}$ is satisfiable classically); $\mathcal{I}$ is a strongly (weakly) tbox-faithful model iff for all tbox axioms τ: if $\mathcal{I} \models \tau$, then $\mathcal{O} \models \tau$ (resp. $\mathcal{O} \cup \{\tau\}$) is satisfiable.*

When we consider classical geometric models such as those defined by Gutiérrez-Basulto and Schockaert [2018], it will in general not be possible to ensure strong faithfulness. So, the notion of strong faithfulness makes sense only for models which allow for representing partial information.

Now let us come back to our al-cones. We present two ways to gain faithfulness, the first using al-cones, the second a subclass of al-cones.

Proposition 3. *For classically satisfiable Boolean $\mathcal{ALC}$-ontologies there is a strongly concept-faithful and tbox-faithful geometric model on some finite $\mathbb{R}^{2n}$ based on al-cones of the form $b_1 \times \cdots \times b_{2n}$ with $b_{2i} \in \{0, +, -, u\}$ and $b_{2i+1} = b_{2i}$. Here n is the number of atomic elements in the Boolean algebra generated by the tbox of the ontology.*

Proof sketch. First a geometric model of the given tbox is generated in size n (using the method described in Prop. 2). In the resulting model every dimension is doubled, $A \sqcup B = \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}$ for instance becomes $A \sqcup B = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_- \times \mathbb{R} \times \mathbb{R}$. To ensure strong concept-faithfulness it must not be possible for any object in the abox to get embedded into a concept properly subsumed by its most specific concept (msc). In each of the subconcepts of the object's msc, one can find an entry $\mathbb{R}$ of the msc a $\mathbb{R}_+$ or a $\mathbb{R}_-$. With doubled dimension there is a $\mathbb{R}_+ \times \mathbb{R}_+$ or $\mathbb{R}_- \times \mathbb{R}_-$. Thus if the corresponding $\mathbb{R}$ in the concept is replaced by, e.g., $\mathbb{R}_+ \times \mathbb{R}_-$ the object is in the concept, but in none of its subconcepts. $\square$

When disregarding $\mathbb{R}_+, \mathbb{R}_-$ in al-cones we are left with halfspaces, which are sufficient to define faithful models.

Proposition 4. *For classically satisfiable Boolean $\mathcal{ALC}$-ontologies there is a strongly concept-faithful and tbox-faithful geometric model on some finite $\mathbb{R}^n$ using sets of the form $b_1 \times \cdots \times b_n$ with $b_i = \{0\}$ or $b_i = \mathbb{R}$.*

Proof sketch. Assume that an $\mathcal{ALC}$ tbox with n atomic elements is given. Arbitrarily assign them to one of the following n spaces $\{0\} \times \{0\} \times \cdots \times \mathbb{R} \times \cdots \times \{0\} \times \{0\}$. Then assign other concepts applying disjunction component-wise. □

This result follows also from the well-known fact that orthogonal subspaces make up a Boolean algebra.

6 The Non-Boolean Case

Our main task is to define semantics for roles and quantifiers. We choose to define crisp extensions for roles. As $\mathcal{ALC}$ does not have role negation, we do not get areas of uncertainty for roles. More importantly, we do not consider special structures for roles (in particular no cones). Following the characterization of $\mathcal{ALC}$ according to Schild [1991] our definition of the existential quantifier has to satisfy $\exists R.\bot = \bot$ and $\exists R.(C \sqcup D) = \exists R.C \sqcup \exists R.D$. As we will define the existential quantifier via the universal quantifier we have to ensure that $\forall R.\top = \top$ and $\forall R.(C \sqcap D) = \forall R.C \sqcap \forall R.D$.

Definition 3. *An al-cone interpretation $\mathcal{I}$ is a Boolean al-cone interpretation $(\Delta, (\cdot)^{\mathcal{I}})$ mapping additionally roles R to subsets over $\Delta \times \Delta$ such that $\vec{0}$ does not appear in the first or second argument of $R^{\mathcal{I}}$. An al-cone interpretation of $\mathcal{ALC}$ concepts is defined recursively as for Boolean concepts and defining the concepts of the form $\forall R.C$ as minimal al-cones containing the set $\{x \in \Delta^{\mathcal{I}} \mid \text{For all } y \in \Delta^{\mathcal{I}}: \text{if } (x, y) \in R^{\mathcal{I}} \text{ then } y \in C^{\mathcal{I}}\}$. The interpretation of the existential quantifier is given by de Morgan, i.e., $(\exists R.C)^{\mathcal{I}} = (\neg \forall R. \neg C)^{\mathcal{I}}$.*

Constructions of the Boolean case do not work for non-Boolean $\mathcal{ALC}$ ontologies due to the fact that the algebra of $\mathcal{ALC}$ concepts is not atomic anymore. Consider the simple example $\mathcal{O}_n = \{loves(narcis, narcis), Vain(narcis)\}$ from Baader and Küsters [2006]. Then $narcis$ fulfils all concepts of the form $\exists R^n.Vain$, which, connected conjunctively, give a chain of concepts C_i that become infinitely narrower and narrower, defined by $C_0 := \exists R.Vain$ and $C_i := C_{i-1} \sqcap \exists R^{i+1}.Vain$. We take a midway between the approach of Gutiérrez-Basulto and Schockaert [2018] and our aim of representing knowledge abox- and tbox-faithfully. The rationale for this is to support ML applications such as prediction. Given an ML model, prediction is meant to categorize (new) objects w.r.t. the concepts represented in the model. If a geometric model has m dimensions, it can be asked queries of the form $A(a)$ for concept label A and an m-vector a. More generally, we consider not only atomic concepts A but arbitrary concepts in $\mathcal{ALC}$. Of course, in a classical model the extension of all concepts are determined by those of the atomic ones. But the point of our approach is to work with partial models. So let us assume that the queries are of the form $C(a)$ (assertion check) or of the form $C(x)$ with variable x (concept retrieval).

Now, the class of queries that are answered in a prediction scenario are constrained in size, where the size is formally specified by some parameter. We will assume in the following that the parameter is the quantifier rank (see Sect. 3). In order to find all potential inconsistencies in an ontology we have to incorporate also the length of role paths in the abox. Let m be the minimum of the length of paths and the expected maximal quantifier rank of the queries in a prediction scenario. The set of $\mathcal{ALC}$ concepts of maximal quantifier rank m still form a Boolean algebra and we have still $\forall R.\top = \top$ as well as $\forall R.(C \sqcap D) = \forall R.C \sqcap \forall R.D$. So, within this class we will have atomic elements, say n, which will be put on the axes of the finite dimensional space $\mathbb{R}^n$.

Proposition 5. *$\mathcal{ALC}$-ontologies are classically satisfiable iff they are satisfiable by a (strongly abox faithful, m-rank-concepts faithful) geometric model on some finite $\mathbb{R}^n$ using sets of the form $b_1 \times \cdots \times b_n$ with $b_i \in \{\{0\}, \mathbb{R}_+, \mathbb{R}_-, \mathbb{R}\}$.*

Proof sketch. "→": Assume that the ontology $\mathcal{O}$ is satisfiable classically. Consider the Lindenbaum-Tarski algebra of maximal m-rank $\mathcal{ALC}$ concepts induced by the given tbox. Assume the number of atomic concepts is n. As in the Boolean case put these concepts onto the axes of a Euclidean space $\mathbb{R}^n$ respecting the Boolean operators. We obtain an al-cone interpretation as follows: Each atomic concept A is interpreted by the al-cone where it is placed according to the Lindenbaum-Tarski algebra embedding. We have to specify the interpretation of the constants and the roles. For a constant a find the most specific concept $msc_{Rank \leq m}(a)$ of rank smaller than m provable in the ontology and place a at some area not occupied already by some individual. In particular, if $C(a)$ follows from the ontology, a is going to be true in the geometric model. For each role assertion $R(a, b)$ entailed by the ontology put $(a, b) \in R^I$.

"←": Assume that $\mathcal{O}$ is not satisfiable classically. There is a complete closed tableau. But this derivation can be mimicked in the geometric model as all clashes are on concept level, i.e., considering only concept assertions of the form $C(a)$ and $\neg C(a)$ but not role assertions. □

7 Conclusion and Outlook

Starting from an interpretation of negation as a polarity operator we presented embeddings of $\mathcal{ALC}$ ontologies that interpret all concepts as axis-aligned cones. This result adds an interesting alternative to embeddings considered so far.

The resulting embeddings can be used in prediction scenarios where one is interested in predicting whether new samples belong to a concept. An open problem is whether roles can be interpreted by cones (or other feasible structures).

As a side product of defining negations by polarity we obtain partial models. So there are individuals for which one does not *know* whether they belong to a concept or not. This is different from the approaches considered in classical embedding scenarios. A fine-grained treatment of this uncertainty could proceed by considering *partial models* as in Hartonas [2016] or by providing an epistemic operator □ [Donini *et al.*, 1998, for example]. Such an operator would allow us to distinguish between instances known to be in a specific concept C, i.e., those in □C and instances which could be in the concept or its negation.

References

[Baader and Küsters, 2006] Franz Baader and Ralf Küsters. Nonstandard inferences in description logics: The story so far. In D.M. Gabbay, S.S. Goncharov, and M. Zakharyaschev, editors, *Mathematical Problems from Applied Logic I*, volume 4 of *International Mathematical Series*, pages 1–75. Springer-Verlag, 2006.

[Baader, 2003] Franz Baader. Description logic terminology. In F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider, editors, *The Description Logic Handbook*, pages 485–495. Cambridge University Press, 2003.

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2787–2795, 2013.

[Boyd and Vandenberghe, 2004] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[Calì *et al.*, 2009] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. Datalog+-: A unified approach to ontologies and integrity constraints. In *Proceedings of the 12th International Conference on Database Theory*, pages 14–30. ACM Press, 2009.

[Donini *et al.*, 1998] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, Werner Nutt, and Anderea Schaerf. An epistemic operator for description logics. *Artificial Intelligence*, 100(1):225–274, 1998.

[Dunn, 1996] Jon Michael Dunn. Generalized ortho-negation. In Heinrich Wansing, editor, *Negation—A Notion in Focus*, Perspektiven der Analytischen Philosophie / Perspectives in Analytical Philosophy 7, pages 3–26. De Gruyter, 1996.

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

[Goldberg and Levy, 2014] Yoav Goldberg and Omer Levy. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *ArXiv e-prints*, February 2014.

[Gutiérrez-Basulto and Schockaert, 2018] Víctor Gutiérrez-Basulto and Steven Schockaert. From knowledge graph embedding to ontology embedding? An analysis of the compatibility between vector space representations and rules. In Michael Thielscher, Francesca Toni, and Frank Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October – 2 November 2018.*, pages 379–388. AAAI Press, 2018.

[Hartonas, 2016] Chrysafis Hartonas. Reasoning with incomplete information in generalized galois logics without distribution: The case of negation and modal operators. In Katalin Bimbó, editor, *J. Michael Dunn on Information Based Logics*, pages 279–312. Springer International Publishing, Cham, 2016.

[Kulmanov *et al.*, 2019] Maxat Kulmanov, Wang Liu-Wei, Yuan Yan, and Robert Hoehndorf. El embeddings: Geometric construction of models for the description logic EL++. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, 2019.

[Levy and Goldberg, 2014] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2177–2185, 2014.

[Mehran Kazemi and Poole, 2018] Seyed Mehran Kazemi and David Poole. SimplE Embedding for Link Prediction in Knowledge Graphs. *arXiv e-prints*, page arXiv:1802.04868, Feb 2018.

[Nickel *et al.*, 2011] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 809–816, USA, 2011. Omnipress.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

[Rockafellar, 1997] Ralph Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1997.

[Schild, 1991] Klaus Schild. A correspondence theory for terminological logics: Preliminary report. In *Proceedings of IJCAI-91*, pages 466–471, 1991.

[Wang *et al.*, 2017] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, Dec 2017.