Neurosymbolic Artificial Intelligence
XX(X):1–43

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Trustworthy Knowledge Base Embeddings: A Foundational Study of Box Semantics

Mena Leemhuis[1] and Oliver Kutz[1]

Abstract

Knowledge base embeddings are a widely applied technique, used for instance to improve link prediction tasks on knowledge graphs by using the geometric regularities occurring during learning. Techniques where ontological concepts are interpreted as boxes have shown to be particularly useful in this context, as they are both suitably expressive and of computational cost allowing practical implementations. However, to use those regularities for learning, it is necessary to determine and understand the possible biases in the approach: how do we distinguish what is learned due to regularities in the data from what is simply based on the representational limitations of the embedding? In this paper, we establish that there are some severe limitations in expressivity when modeling description logic ontologies with box embeddings in intended target languages such as $\mathcal{ELHO}(\circ)^{\bot}$. We illustrate that, under some weak assumptions, box semantics always satisfy Helly's Property, and is thus too weak to semantically capture $\mathcal{ELHO}(\circ)^{\bot}$ in an adequate way. We then characterize how so-called Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$ ontologies can be determined and discuss other restrictions of representability arising from Helly's property, namely the restricted faithfulness of the embeddings.

Keywords

Knowledge Base Embeddings, Ontologies, Neurosymbolic AI

[1] Research Centre for Knowledge-Based Artificial Intelligence (KRDB), Faculty of Engineering, Free University of Bozen-Bolzano, Italy

Corresponding author:
Mena Leemhuis, Free University of Bozen-Bolzano, Italy.
Email: mena.leemhuis@unibz.it

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

Introduction

Knowledge Graphs (KGs) (Hogan et al. 2021) are a widely used representation of diverse knowledge in form of $(subject, predicate, object)$-triples, e.g., $(alice, loves, bob)$. As KGs tend to be highly incomplete, it is necessary to predict missing triples. For this task, various techniques for *Knowledge Graph Embedding (KGE)* have turned out to be useful as they allow for using geometric regularities for learning and thus connect an abstract graph-based view with a vector-based representation. Though these approaches show a promising result quality, they do not incorporate background knowledge. Several techniques have been proposed to include background knowledge in the form of an ontology. Approaches are, e.g., based on sequence modeling, graph propagation and *Knowledge Base Embeddings (KBEs)* (see (Chen et al. 2025) for a survey). The basic idea of KBE is to model individuals as points in a geometric space, concepts as convex sets and relations and logical operations as geometric operations between the individuals or concepts. Subconcept relations are modeled as subset relations and an individual belongs to a concept if its representation is a member of the respective convex set, mimicking the set-based Tarskian semantics. This ensures that newly inferred triples adhere to the background knowledge. There are many different KBE approaches, varying in the choice of the representations of concepts and relations. For instance, they can be based on representing concepts as spheres (Kulmanov et al. 2019), closed convex cones (Özçep et al. 2020) or boxes (Xiong et al. (2022) and others). These approaches propose to use the ontology for enhancing the result quality and interpretability. To have an interpretable result, it is, however, necessary that the embedding actually acts in a predictable way, e.g., with Tarskian style semantics and compositional behavior. Additionally, it needs to be determined whether a specific ontology can be modeled at all. Even when having a consistent and interpretable embedding of the ontology, it is necessary to ensure that the embedding represents the geometric regularities of the training data, and not a bias imposed by possible restrictions of the embedding approach. This leads us to two questions that need to be answered for every KBE approach:

(1) Is the training procedure of the approach able to find an embedding where geometric regularities precisely reflect the information of the knowledge base?

(2) Does such an embedding always exist? If not, under what conditions does it exist?

We will focus here on the more general question (2), which is a basis for particular improvements in the training procedure considered in (1). These questions have been discussed for some specific KBE approaches, namely by Lacerda et al. (2024) in the context of the description logic $\mathcal{ELH}$ and convex sets, and by Özçep et al. (2020) in the context of closed convex cones. Abboud et al. (2020) and Boratko et al. (2021) considered the expressivity of box embeddings, however, boxes were used to model relations and not concepts. Here, we are following the lines of Lacerda et al. (2024) but focus on embeddings based on boxes. Such box embeddings are widely used as they exhibit low computational cost and are able to represents various fragments of the description logic $\mathcal{ELHO}(\circ)^{\bot}$ (the exact expressivity varies for different approaches). Thus, they are of higher expressivity than the approach of Lacerda et al. (2024) and of lower complexity

than the one of Özçep et al. (2020). Though the box embedding approaches are widely used, their expressivity has not been thoroughly examined.

Bourgaux et al. (2024) pointed out that box embedding approaches exhibit problems. For instance, some approaches are not able to model consequences of axioms. This means that whilst an axiom might hold in a geometric representation, its consequences might not necessarily be satisfied (thus it is not a 'full model' of the knowledge base). These problems are, however, problems exhibited by *specific* box embedding approaches. We want to dig deeper into this problem and here extend the work of Bourgaux et al. (2024) in order to understand the general pattern. First, we are considering an abstract box embedding method to determine which properties need to be fulfilled by such an embedding to simulate classical semantics to various degrees. Based on these results, we define the concept of an *optimal* box embedding approach, based on some basic assumptions about box semantics. For such an optimal box embedding we assume that, first, the existence/findability of embeddings is a well-defined notion (without considering their practical learnability), and second, it only imposes those restrictions that all box embedding approaches exhibit. Is it then possible to embed each $\mathcal{ELHO}(\circ)^{\bot}$-ontology such that the ontology is satisfiable if and only if there is a box model of it? In other words, are the limitations of current box embedding approaches based only on the specific (learning) approach used, or are these limitations based on general properties of box semantics? We show in the following that, in addition to restrictions imposed by specific box embedding techniques, also the latter is the case. Thus it is in fact not possible to find a correct box embedding for each $\mathcal{ELHO}(\circ)^{\bot}$-ontology under some weak and widely accepted assumptions on box semantics.* This result is based on an analysis of *Helly's Property* (going back to Helly (1923)), a well-known fact about intersections of convex sets that can be applied to box semantics. Based on this property, we define the notion of *Helly-satisfiable ontologies*. Therefore, although we show that box embeddings have general limitations we also show that it is possible to determine whether an ontology is problematic. Thus, our result does not argue against using box embeddings but opens up a way to determine for which ontologies a box embedding could lead potentially to problems. We extend this notion to *Helly-faithfulness* to determine not only whether an ontology is representable in general but also whether it is possible to do so bias-free. Both these notions can not only be used to determine problematic ontologies but also to increase the trustworthiness of given embeddings. If an embedding is learned, then it is possible to identify problematic parts as the ones that are incorrectly modeled due to limitations of the approach. These parts of the result can then be handled with special care. Additionally, we analyze how the implemented box embedding approaches can be considered as special cases of the generalized box interpretation and thus are also affected by these results.

The paper is structured as follows: First, we discuss the preliminaries on description logics, ontology embeddings and box embeddings in the section "Preliminaries and

*These general standard assumptions on box embeddings, outlined in more detail below, include that conjunction is modeled as set-intersection and that the bottom concept is modeled as the empty set.

Name	Syntax	Semantics
top	$\top$	Δ
bottom	$\bot$	$\varnothing$
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{x \in \Delta \mid \exists y \in \Delta : (x,y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
role concatenation	$(R_1 \circ R_2)^{\mathcal{I}}$	$\{(a,c) \mid \exists b \in \Delta : (a,b) \in R_1^{\mathcal{I}}, (b,c) \in R_2^{\mathcal{I}}\}$

Table 1. Syntax and semantics of $\mathcal{ELHO}(\circ)^{\bot}$ (Baader et al. 2005)

Foundations". In the section "Towards Trustworthy and Interpretable Box Embeddings" we introduce and motivate the problems of the lack of classical semantics, completeness and faithfulness in detail and give an informal overview of the results of the paper. Then, in section "A Generalized Box Interpretation", a general box interpretation is given and set in context of the existing box embedding methods. Next, in section "Expressivity of Boxes", we discuss the expressivity of box embeddings formally, both regarding satisfiability and faithfulness. We end with conclusions and a discussion of open problems and future work.[†] [‡]

Preliminaries and Foundations

In the following, an overview on description logics is given. After that, knowledge graph embeddings and ontology embeddings are introduced with a focus on geometric ontology embeddings based on boxes.

Description Logics

Ontologies are widely used to represent structured information of the world. One way of representing ontologies is with the help of *Description Logics (DL)* (Baader et al. 2007). We are focusing here on the $\mathcal{ELHO}(\circ)^{\bot}$-fragment of the well-known description logic $\mathcal{EL}^{++}$(Baader et al. 2005) due to its computational advantages, as subsumption is polynomial. Prominent examples for ontologies in $\mathcal{ELHO}(\circ)^{\bot}$ are, e.g., SNOMED (Donnelly 2006) for clinical documentation and the Gene Ontology (Ashburner et al. 2000) for modeling genes and their interactions.

A DL vocabulary is given by a set of individual names $\mathbf{I}$, a set of role names $\mathbf{R}$ and concept names $\mathbf{C}$. The $\mathcal{ELHO}(\circ)^{\bot}$ concepts over $\mathbf{C} \cup \mathbf{R}$ are described by the grammar

$$C \longrightarrow A \mid \{a\} \mid \bot \mid \top \mid C \sqcap C \mid \exists R.C$$

where $A \in \mathbf{C}$ is an atomic concept, $a \in \mathbf{I}$ is an individual name, $R \in \mathbf{R}$ is a role symbol, and C stands for arbitrary concepts. $\{a\}$ denotes a nominal concept. An *ontology* $\mathcal{O}$

[†]The detailed versions of all sketched proofs can be found in the appendix.
[‡]This paper is an extended version of the paper (Leemhuis and Kutz 2025) presented at the 19th Conference on Neurosymbolic Learning and Reasoning 2025.

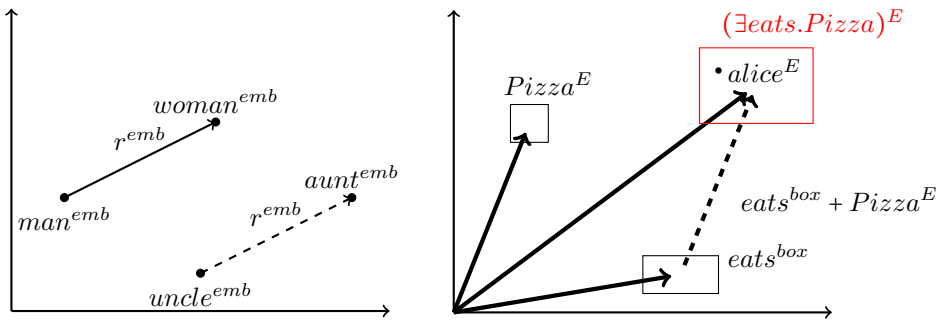


Figure 1. (a) Example for an embedding with TransE; (b) Example for an embedding with TransBox;

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, *role inclusions* $R_1 \circ \cdots \circ R_k \sqsubseteq R$ and *role hierarchies* $R_1 \sqsubseteq R$. $C \equiv D$ is used as an abbreviation of $C \sqsubseteq D$ and $D \sqsubseteq C$. In the following, sets of arbitrary GCIs, not necessarily part of the Tbox, are denoted as $\mathfrak{T}$, sets of arbitrary individual assertions, not necessarily part of the Abox, are denoted as $\mathfrak{A}$. Each ontology can be translated adhering to the following normal forms: all general concepts inclusions can be represented as follows (for $C, D \in \mathbf{C}, E \in \mathbf{C} \cup \{\bot\}$)

$$C \sqsubseteq E \qquad C \sqsubseteq \exists R.D \qquad C \sqcap D \sqsubseteq E \qquad \exists R.C \sqsubseteq E$$

and all role inclusions can be represented as $R_1 \sqsubseteq R$ or $R_1 \circ R_2 \sqsubseteq R$ for $R, R_1, R_2 \in \mathbf{R}$.

An Abox consists of a finite set of *individual assertions*, i.e., facts of the form $a : C$ or of the form $(a, b) : R$ for $a, b \in \mathbf{I}, C \in \mathbf{C}$ and $R \in \mathbf{R}$. An *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 individual names 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. A concept inclusion $C \sqsubseteq D$ is represented as $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, a role inclusion $R_1 \circ \cdots \circ R_k \sqsubseteq R$ as $R_1^{\mathcal{I}} \circ \cdots \circ R_k^{\mathcal{I}} \subseteq R^{\mathcal{I}}$. An interpretation $\mathcal{I}$ *models* an Abox axiom $a : C$, for short $\mathcal{I} \models a : C$, iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and it models an Abox axiom of the form $(a, b) : R$ 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 set of definable concepts in an interpretation $\mathcal{I}$ of $\mathcal{O}$ is defined as

$$DC_{\mathcal{O}}^{\mathcal{I}} = \{A \subseteq \Delta \mid \mathcal{I} \models \mathcal{O}, A = \varphi^{\mathcal{I}} \text{ for some } \varphi \in \mathcal{ELHO}(\circ)^{\bot}\}$$

In this paper, the focus lies on finite ontologies.

From Knowledge Graph Embeddings to Ontology Embeddings

The raise of interest in neurosymbolic AI has led to many approaches enhancing embedding techniques with background knowledge information. Also in the area of knowledge graph embeddings (KGE) this gained increased attention. Knowledge graphs (KGs) are set of $(subject, predicate, object)$-triples. They are especially considered in the context of representing structured large scale data. Examples are DBpedia (Lehmann et al. 2015) and Wikidata (Vrandecic and Krötzsch 2014). KGs are often incomplete and error-prone. A way to tackle this issue is to use KGE to predict missing links with the help of geometric regularities. The basic idea is to learn an embedding where subject and object are modeled as points in a vector space and the relation is modeled as some geometric operation, in case of TransE (Bordes et al. 2013), e.g., as a translation. If such an embedding is trained for given data, it is then possible to predict missing links by applying the geometric operation representing the predicate to a subject and determine whether the resulting point is close to an object representation. In Figure 1 (a), an example can be seen, modeling the relation r representing "is male form of" as a translation vector. When training the embedding with several triples including this relation, e.g., *(man,is_male_form_of,woman)*, then it is possible to predict a missing link, e.g., between *uncle* and *aunt* if $uncle^{emb} + r^{emb} \approx aunt^{emb}$. This is a heavily studied area, see, e.g., Hogan et al. (2021) for an overview. Though KGE allows for a vector representation of graph data, it does not incorporate background knowledge information. Therefore, the inferred links possibly interfere with the common knowledge. To overcome this issue, ontology embeddings (Chen et al. 2025) can be considered.

Ontology embeddings try to incorporate background knowledge information in form of ontologies into the learning process. There, different techniques are possible, especially based on sequence modeling, graph propagation or geometric modeling. The former two suffer from issues with interpretability, as they are mostly not able to represent the ontological structure (Chen et al. 2025). In contrast, geometric ontology embeddings are based on a tight connection between ontology and embedding, as concepts are represented as regions in the embedding space and logical operations can be represented as geometric operations between these regions. These geometric ontology embeddings can be roughly discriminated into two types, the ones for simple and for complex ontologies (Chen et al. 2025). Whereas the former are only able to model concept hierarchies, the latter are able to model more complex ontologies such as fragments of $\mathcal{EL}^{++}$ or $\mathcal{ALC}$. We are interested here in models able to represent more complex ontologies.

The basic idea of such approaches is to learn an embedding of the data and the ontology in a low dimensional vector space such that instances are modeled as points, concepts are modeled as convex regions and relations and logical operations are modeled as geometric operations. Examples for such approaches are ELEmbeddings (Kulmanov et al. 2019) and EmEL++ (Mondal et al. 2021) representing concepts as spheres, Özçep et al. (2020) represents concepts as closed convex cones and several works considering concepts represented as axis-aligned boxes. Concept conjunction can be modeled as set intersection of the respective convex regions and relations can be modeled, e.g.,

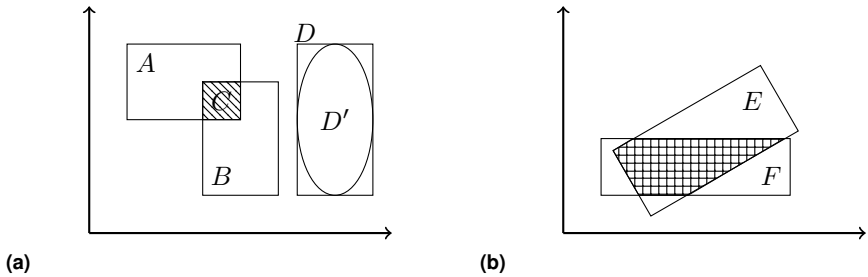


Figure 2. (a) A, B, C and D are examples for axis-aligned boxes. The intersection of two axis-aligned boxes (e.g., A and B) results always in an axis-aligned box (in this case C). $D = BoxHull(D')$, thus D is the box hull of D'; (b) Intersection with a box that is not axis-aligned (in this case E) does not necessarily lead to a box.

via translation in style of TransE. These approaches have a high interpretability, as they model ontological axioms directly in a geometric fashion. We will focus in the following on embeddings based on boxes, as they are, in contrast to spheres, closed under intersection. Thus, it is possible to model axioms of the form $A \sqcap B \equiv C$. In contrast to cone embeddings, the complexity of box embeddings is sufficiently lower, making them more applicable in real-world scenarios. This is also witnessed by the number of box embedding approaches proposed in the last years. They will be discussed in detail in the following.

Boxes

Boxes are chosen as a basis for many embeddings due to their good computational properties and simple representation. A box in some $\mathbb{R}^n$, for $n \in \mathbb{N}$, is defined as an axis-aligned hyperrectangle. It can be represented by its lower corner $l_c \in \mathbb{R}^n$ and upper corner $u_c \in \mathbb{R}^n$, with $l_c \leq u_c$, where $\leq$ is applied element wise. Then, $Box(C) = \{x \in \mathbb{R}^n \mid l_c \leq x \leq u_c\}$. Let $BoxHull(A)$ be the smallest box containing all elements of a set A. This can be defined as

$$BoxHull(A) = \{(x_1, \ldots, x_n)^T \mid x_i \in ConvHull(\{a_i \mid a \in A\}) \text{ for } 1 \leq i \leq n\}$$

where $ConvHull(X)$ is the convex hull of X and a_i is the value at the i-th dimension of vector a. The set $\mathcal{B}^n = \{BoxHull(X) \mid X \subseteq \mathbb{R}^n\}$, thus the set of all boxes in $\mathbb{R}^n$ including the whole space $\mathbb{R}^n$ and the empty set, is closed under set intersection. Properties of boxes are widely researched, e.g., in the context of *intersection graphs* and *boxicity* (Roberts 1969). One main advantage of axis-aligned boxes is their closure under intersection. The intersection of two axis-aligned boxes A and B is defined as follows:

$$C := A \cap B = \{x \in \mathbb{R}^n \mid l_c \leq x \leq u_c \text{ where } l_c^i = \max(l_a^i, l_b^i) \text{ and } u_c^i = \min(u_a^i, u_b^i)\}$$

Thus, the intersection of two axis-aligned boxes always results in an axis-aligned box. In Figure 2 (a), an example for such an intersection of two boxes can be seen, resulting in

an axis-aligned box. Box D is an example for the box hull of an arbitrary set. Figure 2 (b) exemplifies the need of axis-aligned boxes: the intersection of a box F with a non-axis-aligned box E leads not to a box, therefore, it is not closed under intersection. This would be problematic, as it complicates the learning: the intersection of E and F can not be represented by defining center and offset but needs more parameters. Therefore, arbitrary complex representation would be necessary, interfering with the aim of computational simplicity and potentially leading to overfitting. Therefore axis-aligned boxes are considered and the term "box" refers in the following solely to axis-aligned boxes. Note here that arbitrary dimensional boxes are considered, and that the examples are restricted to the two-dimensional case for representation reasons only.

Box Embeddings

The basic principle of box embeddings is to find a mapping between an ontology and a box representation. Concepts of the ontology are represented as boxes, whereas logical operations in the ontology language, such as conjunctions, are represented with the help of some geometric operation in the embedding space. Thus, we give a basic definition of such box embeddings below, inspired by existing KBE approaches whilst abstracting from the specifics of those approaches. The level of abstraction of our definition is chosen primarily to support our general study of box embeddings and their theoretical properties. A thorough examination of a broader framework for abstract embedding models and their theoretical properties, also covering for instance other basic geometries, is left for future work.
In the following definition, the core commonality is that concepts will be interpreted as boxes and individuals as subsets of $\mathbb{R}^n$. Note here that we can recover the simpler version of an individual as a point in space by modeling them as singleton sets.

Definition 1. Box Embedding Method / Box Embedding. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^\bot$ ontology.*[§] *Assume $\mathbf{C}, \mathbf{R}, \mathbf{I}$ are the finite sets of concepts, roles, and individuals, respectively, that appear in $\mathcal{O}$. A* box embedding method *S_M for $\mathcal{ELHO}(\circ)^\bot$ in $\mathbb{R}^n$ for some fixed $n \in \mathbb{N}$, provides functions for logical constants*

- $f_\top \in \mathcal{B}^n$ *(interpreting $\top$ as a box)*
- $f_\bot \in \mathcal{B}^n$ *(interpreting $\bot$ as a box)*
- $f_\sqcap : \mathcal{B}^n \times \mathcal{B}^n \to \mathcal{B}^n$ *(interpreting conjunction),*
- $f_\exists : (\wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n)) \times \mathcal{B}^n \to \mathcal{B}^n$ *(interpreting existential restriction, where $\wp(\cdot)$ denotes the powerset operation), and*
- $f_\circ : \big(\wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n), \wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n)\big) \to \wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n)$ *(interpreting role composition)*

[§]Note that here and in the following, we are focusing on $\mathcal{ELHO}(\circ)^\bot$ due to the fact that actual box-based embedding approaches are considering $\mathcal{ELHO}(\circ)^\bot$. It is straightforwardly possible to extend these results and definitions to more expressive ontologies.

It furthermore provides Boolean-valued functions for sentences

- $f_{\sqsubseteq} : \mathcal{B}^n \times \mathcal{B}^n \to \{0,1\}$ *(interpreting subsumption),*
- $f_R : (\wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n)) \times (\wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n)) \to \{0,1\}$ *(interpreting role instantiation)*
- $f_: : \wp(\mathbb{R}^n) \times \mathcal{B}^n \to \{0,1\}$ *(interpreting instantiation)*

Given a method S_M, a box embedding *E for $\mathcal{O}$ is an interpretation function that maps*

- $\top^E = f_\top$ *and* $\bot^E = f_\bot$ *(top and bottom concept set by method)*
- *each concept $C \in \mathbf{C}$ to a box $B \in \mathcal{B}^n$ in $\mathbb{R}^n$*
- *each individual name to a subset of $\mathbb{R}^n$*
- *each role $R \in \mathbf{R}$ to a subset of $\wp(\mathbb{R}^n) \times \wp(\mathbb{R}^n)$*
- *each nominal concept $\{c\}$ to a box $B \in \mathcal{B}^n$ in $\mathbb{R}^n$*

E can now be extended recursively to arbitrary $\mathcal{ELHO}(\circ)^\bot$-concepts as follows:

- $(C \sqcap D)^E = f_\sqcap(C^E, D^E)$
- $(\exists R.C)^E = f_\exists(R^E, C^E)$
- $(R \circ S)^E = f_\circ(R^E, S^E)$

Given a method S_M and a box embedding E, we can define satisfaction. We use the notation $E \Vdash \phi$ to denote that a box interpretation E models, or satisfies, a certain statement ϕ. Specifically, given a box interpretation E we define:

- *For an Abox assertions $a : C$ set: $E \Vdash a : C \iff f_:(a^E, C^E) = 1$;*
- *For an Abox assertions $(a,b) : R$ set: $E \Vdash (a,b) : R \iff f_R(a^E, b^E, R^E) = 1$;*
- *For a GCI $C \sqsubseteq D$ set: $E \Vdash C \sqsubseteq D \iff f_\sqsubseteq(C^E, D^E) = 1$.*

The main idea of this embedding method is to be as general as possible. Concepts (including the top and bottom concept) are modeled as boxes. Note that $\mathcal{B}^n$ includes also $\mathbb{R}^n$ and the empty set, thus $\top$ can be modeled classically as $\mathbb{R}^n$ and $\bot$ classically as $\varnothing$. The representations of conjunction of concepts and existential restriction of concepts have a box as an outcome. To exemplify the generality of this basic box embedding definition, different readings of $E \Vdash a : C$ thus of $f_:(\cdot,\cdot)$ can be considered. It could be, e.g., the case that a^E is interpreted as a small set in $\mathbb{R}^n$. Then, $E \Vdash a : C$ could be the case if a^E and C^E are overlapping, thus a is 'somewhat' a member of C. We can also consider the case where $a^E \subseteq C^E$, which would be an interpretation closer to the classical interpretation. When interpreting individuals as points, thus $a^E \in \mathbb{R}^n$ and $E \Vdash a : C$ if $a^E \in C^E$, then we arrive at the classical, intuitive definition of $a : C$. Note that this box embedding deviates

highly from classical interpretations and also from an interpretation that would have been expected to be an interpretation of an ontology based on boxes. A more classical interpretation based on boxes will be presented in Definition 22. These embedding definitions presented here should make it clear that the term "box embedding" alone does not imply anything about the actual semantics of such an embedding. The tendency is to assume that a box embedding fulfills basic expected properties such as that the subconcept relationship remains transitive, i.e. that $E \models A \sqsubseteq B$ and $E \models B \sqsubseteq C$ should lead to $E \models A \sqsubseteq C$. This is, however, dependent on the definition of the embedding and not trivially fulfilled. Therefore, we define here a box embedding as an abstract method with the basic property of mapping concepts to boxes, and concept forming operations such as conjunction and existential restrictions also mapping to boxes, as explained in detail above. This is inspired by *abstract description systems* (Baader et al. 2002; Kutz et al. 2002) which provided a similar syntactic and semantic abstraction of a number of systems of modal, hybrid, and description logics, and thus enabled a systematic study of combination methods and transfer results. Note here that in our context it is important that $\top^E$ and $\bot^E$ are not modeled individually for every embedding E but are fixed by the embedding method S_M. Otherwise, this would lead to an even more non-standard behavior, e.g., that $E \models \mathcal{O}$, $\mathcal{O}' \subseteq \mathcal{O}$ but $E \not\models \mathcal{O}'$. In general, it would also be possible to train a function such as $f_\sqcap(\cdot,\cdot)$ based on the given input data. This is not considered here, as it is not done for existing box-based KBE approaches. This abstract definition of a box embedding is considered in the context of actual KBE approaches in the next section.

Example 2. A non-intuitive box embedding. *Let S_M be a box embedding method such that each individual is mapped to a point in $\mathbb{R}^n$. Let $E \models C \sqsubseteq D$ iff $C^E \cap D^E = \emptyset$, $E \models a : C$ iff $a^E \notin C^E$. Conjunction is defined as box hull, thus $f_\sqcap(C^E, D^E) = BoxHull(C^E \cup D^E)$. The other definitions are omitted here for simplicity. This embedding method shows some classical behavior, e.g., $(A \sqcap A)^E = A^E$ but also non-classical behavior, e.g., the subconcept relation is symmetric, thus, $E \models C \sqsubseteq D$ implies $E \models D \sqsubseteq C$. The definition of the individuals and their membership in combination with the definition of conjunction leads to the fact that if $E \not\models a : C$, it could be the case that $E \models a : C \sqcap D$.*

Box Embeddings in the Context of Knowledge Base Embeddings

We are focusing here on representing these convex sets as boxes, as they show a good tradeoff between expressivity and computational properties. Especially, they are closed under intersection (in contrast to spheres) and easy to be handled computationally (in contrast to cones). Box embedding approaches in the context of KBE are BoxEL (Xiong et al. 2022), ELBE (Peng et al. 2022), Box^{2}EL (Jackermeier et al. 2024) and TransBox (Yang et al. 2025). All these approaches model a subconcept classically as subset relation, thus $E \models A \sqsubseteq B$ iff $A^E \subseteq B^E$ and individual assertions classically as $E \models a : C$ iff $a^E \subseteq C^E$ where a^E is either defined as a point or a box in $\mathbb{R}^n$. $\top^E = \mathbb{R}^n$ and

$\bot^E = \varnothing$ in all these approaches.[¶] Conjunction is mostly defined as set-intersection, thus $(A \sqcap B)^E = A^E \cap B^E$ with exception of TransBox that offers as an additional variant to define the conjunction as an approximation of the intersection. These approaches mainly differ in the representation of relations. An overview on these approaches can be found in Bourgaux et al. (2024). These approaches are able to model $\mathcal{ELHO}(\circ)^\bot$ ontologies or fragments of it. Note here that, though these approaches seem to have a classical semantics on first sight (as their definition is not as abstract as the one of Definition 1), they still do not have a classical semantics as will discussed in section "Towards Trustworthy and Interpretable Box Embeddings" and pointed out in Example 4.

Example 3. Embedding with TransBox. *An example for a box embedding based on TransBox can be seen in Figure 1 (b). It is based on the idea of modeling relations also as boxes and states that a triple (a, r, b) holds if $a^E \in r^{box} + b^E$. In the example, the concept "Pizza" and the role "eats" are represented as boxes, the individual "alice" as a point in the space. Then it is the case that "Alice eats pizza" if the point representing Alice is part of the box representing $\exists eats.Pizza$. This box is determined by adding up the centers resp. the offsets of the boxes of "Pizza" and "eats".*

These approaches try to find a tradeoff between expressivity and complexity. Thus, e.g., representing individuals as boxes eases the training but leads to unexpected behavior as demonstrated in the following example.

Example 4. *Assume an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ is given and let $\mathcal{T} = \{A \sqcap B \sqsubseteq \bot\}$ and let $\mathcal{A} = \{a : A, a : B\}$. This ontology is clearly not satisfiable. Let an embedding method S_M be defined by interpreting conjunction as set intersection, $\bot$ as empty set and individuals as boxes. The rest can be defined classically, as it is not considered here. Now, a possible embedding E such that $E \not\models A \sqcap B \sqsubseteq \bot$ could be trivially modeled by using two non-intersecting boxes in $\mathbb{R}^n$ to represent the two concepts. Let a^E be now the empty box. This is a box, thus the definition is in line with the definition of the embedding method. Then, clearly, $E \not\models a : A$ and $E \not\models a : B$. This example is an adapted version of (Bourgaux et al. 2024, Example 1) and ELBE (Peng et al. 2022) suffers from a similar problem. This shows that such problems are not only easily to be made up with artificial examples but also based on real-world KBE approaches.*

An embedding is not constructed by hand but learned based on the given definition of the embedding method and the training data. For given input data, thus a KG with an underlying ontology, some neural network based learning method is used to learn an embedding based on a loss function. The ontology is first transformed to normal form to ease the training process. The dimensionality n of the embedding is set as a hyperparameter. The choice is based on a trade off between representability and learnability. Then, all concepts are initialized as some (random) boxes in the n-dimensional vector space and individuals and relations are instantiated depending on the

[¶]Note here that representing $\bot$ as empty set, though copying the classical interpretation, is not self-evident. In the work of Özçep et al. (2020), e.g., $\bot$ is interpreted as the point of origin in the context of interpreting concepts as closed convex cones.

exact embedding approach used. After this initialization, the representation of concepts, individuals and relations is stepwise optimized via the loss function. An axiom of type $A \sqsubseteq B$ is, e.g., modeled via a loss function that rewards short distance between A and B and is zero if the box representing A is part of the box representing B. The overall loss function is a weighted sum of the losses for each of the normal forms. A global optimum is found if the loss equals zero. Otherwise some axioms have not been modeled correctly. Our main question is whether for a general box embedding approach, it is always possible to find for a given ontology such an embedding where the loss is zero. Additionally, the question is whether such a zero-loss-embedding then has an interpretable semantics and allows for statements about the satisfiability of the ontology. Note here that the goal is not to find the one perfect embedding approach but to analyze the pitfalls and restrictions of existing approaches, especially as the assumption of classical behavior could lead to overlooking problematic unexpected behavior of the embedding.

Towards Trustworthy and Interpretable Box Embeddings

Knowledge base embeddings evolved out of the need of including background knowledge information in form of an ontology into a link prediction process. In contrast to other ontology embedding approaches, they rely on a tight connection between ontological information and data and are especially not based on a post-processing step. But how tight is this connection? In the last section we exemplified in Example 4 that a small adaptation of the embedding approach could lead to a loss of soundness and to unexpected behavior. A trustworthy embedding approach needs to have an understandable and intuitive semantics. The ontology is based on Tarskian semantics, thus, e.g., when the axioms of an ontology are satisfied in an interpretation, then also their entailments are satisfied. This leads especially to the question whether there is some kind of correspondence between classical models based on interpretations (of an description logic ontology) and geometric models based on some embedding method. Thus, is there are Tarskian semantic for the box embedding (and also for the other KBE methods)? This discussion was started in Bourgaux et al. (2024).

They define several steps to determine to what extend an embedding resp. an embedding method follows a classical semantics. Here, they are slightly adapted to focus on box embeddings. In the following, we will give the definition of (Bourgaux et al. 2024), discuss them in detail and show what needs to be fulfilled to have an understandable Tarskian style semantics or what needs to be tested to understand regarding which aspects the embedding method might deviate from a classical semantics.

Definition 5. *(Bourgaux et al. 2024, Def. 3) Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$. The embedding E interpreted under the embedding method S_M as defined in Definition 1 is a* shallow model *of*

- *$\mathcal{A}$ if for every individual assertion α of $\mathcal{A}$, $E \models \alpha$,*
- *$\mathcal{T}$ if for every axiom τ of $\mathcal{T}$, $E \models \tau$,*

- *$\mathcal{O}$ if it is a shallow model of $\mathcal{A}$ and $\mathcal{T}$.*

Although, this resembles the classical interpretation of a model in the DL-sense, in contrast to classical models, fulfillment of Tbox and Abox does neither imply that the entailments of the ontology are also modeled nor that the ontology is satisfiable classically. Therefore, they are called "shallow" here, as they are only enforced to fulfill the surface of the ontology, namely only the axioms and individual assertions. Entailments of the ontology are not necessarily represented in a shallow model. Models of this type are highly dependent on the representation of Abox and Tbox and their existence does not even state whether an ontology is satisfiable. Examples have been given in Example 2 showing that a shallow model could have a non-intuitive semantics and in Example 4 showing that the existence of a shallow model does not imply that the ontology is satisfiable, thus the presented embedding method is not sound. Therefore, in the following, *soundness* and *completeness* of shallow models are defined.

Definition 6. *(Bourgaux et al. 2024, Prop. 1,2)*

- *We say that S_M is* sound *if the existence of a shallow model (under S_M) for an ontology $\mathcal{O}$ implies that $\mathcal{O}$ is satisfiable.*
- *We say that S_M is* complete *if for every satisfiable ontology $\mathcal{O}$, there is a shallow model (under S_M) for $\mathcal{O}$.*

The existence of a sound and/or complete shallow model still does not guarantee that the shallow model resembles a classical interpretation. It is possible that, e.g., entailments are not modeled correctly or that the embedding is inconsistent. It only depicts that there is a correspondence between existence of some shallow model and satisfiability of an ontology. A trivial example for such a sound model would be based on an embedding method including a satisfiability checker where an embedding is only produced if the ontology is proven to be satisfiable.

Therefore, as a next step, entailment closure should be enforced, thus an embedding that satisfies axioms and assertions of the ontology should also satisfy their entailments. This reduces the dependence of the embedding from the exact definition of the ontology. Missing entailment closure is a problem of many box-based KBE approaches: though they are able to model axioms and assertions, the entailments are, especially when they are more complex, often not modeled properly.

Example 7. A sound but not entailment closed embedding. *Let S_M be an arbitrary embedding method such that $f_{\sqsubseteq}$ is not transitive. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^{\perp}$ ontology with $\mathcal{T} = \{A \sqsubseteq B, B \sqsubseteq C\}$ and $\mathcal{A} = \{\}$. Let E be a shallow model of $\mathcal{O}$ under S_M such that $E \mathrel{\sim\!\!\!\vdash} A \sqsubseteq B$ and $E \mathrel{\sim\!\!\!\vdash} B \sqsubseteq C$. As $f_{\sqsubseteq}$ is not transitive, $E \mathrel{\sim\!\!\!\vdash} A \sqsubseteq C$ would not necessarily be the case in E, though it follows trivially from the ontology. Another example would be to interpret $f_{\sqcap}$ of S_M as $f_{\sqcap}(A^E, B^E) = BoxHull(A^E \backslash B^E)$. Then $E \not\mathrel{\sim\!\!\!\vdash} A \sqcap A \equiv A$, a trivial tautology in S_M, namely idempotence. This, though, does not need to have any impact on soundness and completeness of S_M.*

Therefore, *entailment closure* is introduced.

Definition 8. *(Bourgaux et al. 2024, Def. 4) Let $\mathcal{O}$ be a classically consistent (DL) ontology. We say that a shallow model E is*

- *Tbox-entailment closed if for every GCI τ if $\mathcal{O} \models \tau$, then $E \models \tau$;*
- *Abox-entailment closed if for every assertion α, if $\mathcal{O} \models \alpha$, then $E \models \alpha$;*
- *KB-entailment closed if it is Tbox-entailment closed and Abox-entailment closed.*

Instead of demanding closure over all entailments, it would also be possible to examine the depth up to which the entailments are modeled correctly. The depth could be, e.g., approximated by considering the minimal depth of a tableau used to prove an entailment. Such strategies are not considered further here but could be interesting to consider for real-world use cases. An embedding modeling only GCIs directly entailed by the Tbox would be of lower quality than one able to model entailments of depth n for some $n > 1$. This could ease the training procedure by still modeling a sufficient correctness.

As box embeddings methods, especially KBE methods are not only used to model the information entailed by the ontology correctly but are also used to infer new facts, e.g., by doing link prediction, it is not sufficient to model the ontology and its entailments correctly, it is also necessary that all GCIs or assertions modeled in the embedding are consistent with the ontology. Thus, an assertion learned based on the data should be in line with the ontology. This is called *weak faithfulness*.

Example 9. An embedding not weakly faithful. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^{\bot}$ ontology with $\mathcal{T} = \{\}$ and $\mathcal{A} = \{a : \exists R.C\}$. Let the embedding method S_M such that individuals be represented as points in $\mathbb{R}^n$, let $E \models a : C$ iff $a^E \in C^E$, let R be represented as translation, thus $E \models (a, b) : R$ iff $a^E + R^E = b^E$ and assume that $\exists R.C$ is defined classically. Let additionally $\bot^E = \varnothing$. Now consider as domain the one dimensional space $\mathbb{R}$ and a shallow model E under S_M. Let $a^E = 1$ and $C^E = 0$, thus $E \models C \sqsubseteq \bot$. Assume that $R^E = -1$. Then $E \models a : \exists R.C$, as $a^E + R^E = 0$ and thus, $E \models a : \exists R.\bot$ which is clearly unsatisfiable in a classical interpretation of an ontology. Note that for cone embeddings exactly this problem occurred and needed to be actively circumvented (see (Leemhuis et al. 2022) for an in-depth discussion).*

Thus, we define *weak faithfulness*.

Definition 10. (Özçep et al. 2020). *Let $\mathcal{O}$ be a classically consistent (DL) ontology. We say that a shallow model E of $\mathcal{O}$ is*

- weakly Tbox-faithful[‖] *if for every GCI τ: if $E \models \tau$, then $\mathcal{O} \cup \{\tau\}$ is satisfiable.*
- weakly Abox-faithful *if for every assertion α: if $E \models \alpha$, then $\mathcal{O} \cup \{\alpha\}$ is satisfiable.*

[‖] In line with the literature, we stick here to the names of "Tbox-" and "Abox-"faithful. Note, though, that not only Tbox axioms and Abox assertions are checked but GCIs and assertions in general.

- weakly faithful *if it is weakly Abox and weakly Tbox faithful.*

Note that in the definitions of weak faithfulness, both by Özçep et al. (2020) and by Bourgaux et al. (2024), it is stated that $\mathcal{O} \cup \{\tau\}$ (resp. $\mathcal{O} \cup \{\alpha\}$) needs to be satisfiable for every α, τ. In the following, it is argued that this restriction is not sufficient to gain a meaningful semantics. For weak faithfulness, the GCIs are tested separately, however, they could interfere with each other. Thus, the ontology should be satisfiable by considering all newly inferred GCIs and assertions at the same time. Therefore, we define *global faithfulness*:

Definition 11. *Let $\mathcal{O}$ be a classically consistent (DL) ontology. We say that a shallow model E of $\mathcal{O}$ is*

- globally Tbox-faithful *if for every set of GCIs $\mathfrak{T}$: if $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} \mathfrak{T}$, then $\mathcal{O} \cup \mathfrak{T}$ is satisfiable.*
- globally Abox-faithful *if for every set of assertions $\mathfrak{A}$: if $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} \mathfrak{A}$, then $\mathcal{O} \cup \mathfrak{A}$ is satisfiable.*
- globally faithful *if for every set of GCIs and assertions $\mathfrak{T} \cup \mathfrak{A}$: if $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} \mathfrak{T} \cup \mathfrak{A}$, then $\mathcal{O} \cup \mathfrak{T} \cup \mathfrak{A}$ is satisfiable.*

Note that this means that for a globally faithful shallow model E, there is a classical interpretation $\mathcal{I}$ such that if $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} \mathfrak{T} \cup \mathfrak{A}$, then $\mathcal{I} \models \mathfrak{T} \cup \mathfrak{A}$.

Proposition 12. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an ontology and S_M an embedding method. There exists a shallow model E of $\mathcal{O}$ under S_M such that E is weakly but not globally faithful.*

Proof. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an ontology with $\mathcal{T} = \{\}$ and $\mathcal{A} = \{a_1 : A, a_2 : B\}$. The ontology is clearly satisfiable. Let E be a shallow model such that $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} A \sqcap B \sqsubseteq \bot$ and $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} a_1 : B$. E is entailment closed (as there aren't any non-trivial entailments). $\mathcal{O} \cup \{A \sqcap B \sqsubseteq \bot\}$ is satisfiable, same as $\mathcal{O} \cup \{a_1 : B\}$ is satisfiable. Thus, E is weakly faithful. However, $\mathcal{O} \cup \{A \sqcap B \sqsubseteq \bot, a_1 : B\}$ is clearly not satisfiable. Thus, E is not globally faithful. An embedding method can be constructed as done in Example 4 by interpreting individuals as boxes and therefore allowing an individual to be represented by the empty box.

When now considering an KB-entailment closed and globally faithful shallow model E, we can guarantee that E is consistent in a classical semantics. However, it still does not resembles classical semantics in the sense exemplified in the following example.

Example 13. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^{\bot}$-ontology with $\mathcal{T} = \{\}$ and $\mathcal{A} = \{\}$ and $\mathbf{C} = \{A, B, C\}$, $\mathbf{I} = \{a_1, a_2\}$ and $\mathbf{R} = \{\}$. Let S_M be an embedding method and E a shallow model such that $E \mathrel{\ooalign{$\models$\cr\hidewidth$\sim$\hidewidth}} \mathfrak{T} \cup \mathfrak{A}$ for $\mathfrak{T} = \{A \sqsubseteq B, B \sqsubseteq C\}$ and $\mathfrak{A} = \{a_1 : A, a_1 : B, a_2 : B, a_2 : C\}$. Note that this embedding is trivially entailment closed (as there is nothing entailed by the ontology except tautologies such as $A \equiv A$ which we assume to hold in the embedding). It is also weakly (and globally) faithful, as $\mathcal{O} \cup \mathfrak{T} \cup \mathfrak{A}$ is satisfiable. Now assume that $\mathfrak{T} \cup \mathfrak{A}$ are the only (non-trivial) statements satisfied in E.*

Let the subconcept relation $f_{\sqsubseteq}$ be defined as being non-transitive, thus $E \not\models A \sqsubseteq C$. In each classical model $\mathcal{I}$ of $\mathcal{O} \cup \{A \sqsubseteq B, B \sqsubseteq C\}$, it is the case that $\mathcal{I} \models A \sqsubseteq C$. Thus, it is necessary to consider entailment closure not only based on the entailments of the ontology but also based on the expected entailments of the embedding.

Therefore, we introduce embedding-entailment closure as a notion of entailment closure on embedding level.

Definition 14. *Let $\mathcal{O}$ be a classically consistent (DL) ontology. We say that a shallow model E is* embedding-entailment closed *if for every set of GCIs $\mathfrak{T}$ and assertions $\mathfrak{A}$ if $E \models \mathfrak{T} \cup \mathfrak{A}$ and $\mathcal{O} \cup \mathfrak{T} \cup \mathfrak{A} \models \mathfrak{T}$, then $E \models \mathfrak{T}$.*

Note that every shallow model E that is embedding-entailment closed is also KB-entailment closed (that follows immediately when choosing $\mathfrak{T} = \mathfrak{A} = \emptyset$). Up to now we have defined faithfulness and entailment closure based on one shallow model E of an ontology. Thus, it was sufficient that the embedding method S_M had the *ability* to model an entailment closed/faithful embedding of an ontology. A stronger constraint is to enforce the embedding method S_M to produce only KB-entailment closed/weakly faithful shallow models. Thus, a *guarantee* needs to be given.

Definition 15. *(Bourgaux et al. 2024, Property 5) We say that S_M is guaranteed to be weakly faithful resp. KB-entailment closed if, for every satisfiable ontology $\mathcal{O}$, S_M always produces a shallow model E of $\mathcal{O}$ such that E is weakly faithful resp. entailment closed.*

Under a guaranteed condition, weak faithfulness implies global faithfulness and entailment closure implies embedding-entailment closure. These two stronger conditions are therefore only necessary when considering specific embeddings.

Corollary 16. *Let $\mathcal{O}$ be a classically consistent (DL) ontology.*

1. *Let S_M be guaranteed to be weakly faithful. Then S_M is also guaranteed to be globally faithful.*
2. *Let S_M be guaranteed to be KB-entailment closed. Then S_M is also embedding-entailment closed.*

Proof. Let $\mathcal{O}$ be a classically consistent (DL) ontology.

1. Let S_M guaranteed to be weakly faithful. Assume for contradiction that there is a shallow model E of $\mathcal{O}$ under S_M such that E is not globally faithful but weakly faithful. Let $E \models \mathfrak{T} \cup \mathfrak{A}$ and assume $\mathfrak{T} \cup \mathfrak{A}$ has some enumeration. Due to weak faithfulness, for $\gamma_1 \in \mathfrak{T} \cup \mathfrak{A}$, $\mathcal{O} \cup \gamma_1$ is satisfiable. Let $\mathcal{O}_1 = \mathcal{O} \cup \gamma_1$. Then E is also a shallow model of $\mathcal{O}_1$ and due to guaranteed faithfulness, it is also weakly faithful. This can now be repeated inductively: As E is a shallow model of $\mathcal{O}_n$ and $\mathcal{O}_n$ is satisfiable, due to weak faithfulness, with $E \models \gamma_{n+1}$ for $\gamma_{n+1} \in \mathfrak{T}$ it follows that $\mathcal{O}_n \cup \gamma_{n+1}$ is satisfiable. Then $\mathcal{O}_{n+1} = \mathcal{O}_n \cup \gamma_{n+1}$. And therefore $\mathcal{O} \cup \mathfrak{T} \cup \mathfrak{A}$ is satisfiable, thus global faithfulness follows, a contradiction to the assumption.

2. The proof follows analogously.

As proven in the following Theorem 17, such a guarantee enables a close connection to a classical semantics. When additionally assuming that S_M is sound and complete, then it is not only the case that $\mathcal{O}$ is satisfiable by a classical interpretation if and only if it is satisfiable by a shallow model E under S_M. Additionally, it can be stated that there is a model $\mathcal{I}$ of $\mathcal{O}$ such that for all γ it is the case that $\mathcal{I} \models \gamma$ if and only if $E \mathrel{\mathfrak{E}\!\!-} \gamma$. Thus, there is an interpretation exactly resembling what E represents. Note that this is not equivalent to E being a classical interpretation.

Theorem 17. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^\bot$ ontology.*
Assume the following condition:

Pseudo-Tarski *S_M is guaranteed to be weakly faithful and guaranteed to be KB-entailment closed.*

Then for any shallow model E of $\mathcal{O}$ under S_M, there exists a classical interpretation $\mathcal{I}$ with $\mathcal{I} \models \mathcal{O}$ such that for all GCIs and assertions γ in language $\mathcal{ELHO}(\circ)^\bot$:

$$\mathcal{I} \models \gamma \text{ if and only if } E \mathrel{\mathfrak{E}\!\!-} \gamma$$

Proof. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^\bot$ ontology. Let S_M be a guaranteed weakly faithful and guaranteed entailment closed. Due to the guarantee of faithfulness and entailment closure and with Corollary 16, each shallow model E is KB- and embedding-entailment closed and globally faithful. Let $E \mathrel{\mathfrak{E}\!\!-} \mathfrak{T} \cup \mathfrak{A}$ and let $\mathfrak{T} \cup \mathfrak{A}$ be maximal, thus for all γ with $E \mathrel{\mathfrak{E}\!\!-} \gamma$, $\gamma \in \mathfrak{T} \cup \mathfrak{A}$. Due to global faithfulness, with $E \mathrel{\mathfrak{E}\!\!-} \mathfrak{T} \cup \mathfrak{A}$ it follows that $\mathcal{O}' := \mathcal{O} \cup \mathfrak{T} \cup \mathfrak{A}$ is satisfiable. Due to KB-entailment closure, $E \mathrel{\mathfrak{E}\!\!-} \mathcal{O}$ and due to embedding entailment closure $E \mathrel{\mathfrak{E}\!\!-} \gamma$ if $\mathcal{O}' \models \gamma$. Let $\mathcal{I}$ be the minimal model of $\mathcal{O}'$, minimal in the sense that $\mathcal{I} \models \gamma$ implies $\mathcal{O}' \models \gamma$ for all γ. Now let γ be an arbitrary GCI or assertion. If $E \mathrel{\mathfrak{E}\!\!-} \gamma$, then by definition, $\gamma \in \mathcal{O}'$ and thus $\mathcal{I} \models \gamma$, as $\mathcal{I}$ is a model of $\mathcal{O}'$. If $\mathcal{I} \models \gamma$, then by definition of $\mathcal{I}$, it is the case that $\mathcal{O}' \models \gamma$ and thus, due to entailment closure, $E \mathrel{\mathfrak{E}\!\!-} \gamma$.

As will be discussed in section "Expressivity of Boxes", box embedding methods (based on some weak assumptions) can never be complete. Additionally, KBE approaches should be of a sufficiently low computational cost and especially not every ontology is needed to be represented correctly. It would be sufficient to focus on specific, widely used ontologies and to enforce that there is at least one adequate, thus entailment closed and weakly faithful embedding. Enforcing guaranteed properties would be therefore a too severe restriction. A second variant is to consider again a specific shallow model E based on some S_M for a specific ontology $\mathcal{O}$. Independent of whether S_M has some guaranteed properties, it is necessary to be able to determine whether this shallow model E resembles the properties of a classical interpretation in a sufficient way (thus, that there is a model $\mathcal{I}$ of $\mathcal{O}$ such that for all γ it is the case that $\mathcal{I} \models \gamma$ if and only if $E \mathrel{\mathfrak{E}\!\!-} \gamma$). In the following, we show that if E is globally faithful and embedding-entailment closed, then this is actually the case. This result allows for defining procedures

to test for specific KBE approaches and specific embeddings of these approaches whether these embeddings resemble a classical semantics. Such a procedure would allow to increase the trustworthiness of the KBE approaches, as it allows for clearly stating the expressivity not only of the approach in general but also based on the exact ontology and embedding modeled. That this can be defined follows directly out of Theorem 17.

Corollary 18. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^\bot$ ontology.*
Assume the following condition:

Weak Pseudo-Tarski *E is a shallow model of $\mathcal{O}$ that is embedding entailment closed and globally faithful.*

Then there exists a classical interpretation $\mathcal{I}$ with $\mathcal{I} \vDash \mathcal{O}$ such that for all GCIs and assertions γ in language $\mathcal{ELHO}(\circ)^\bot$:

$$\mathcal{I} \vDash \gamma \text{ if and only if } E \mathrel{\mathfrak{E}\!\!\sim} \gamma$$

Proof. The statement follows directly with Theorem 17 as the pseudo-Tarski property implies the weak pseudo-Tarski property.

To summarize: a KBE approach can only then be used in a trustworthy fashion if it has a trustworthy (and known) semantics. By Bourgaux et al. (2024), the existing KBE approaches have been examined based on their ability for soundness, completeness, guaranteed and possible weak faithfulness and entailment closure. We extend this consideration in the following: we are not only interested in the fact that most of the embedding approaches are neither sound, complete, faithful nor entailment closed. We want to determine whether it is possible to define such an embedding method as defined in Theorem 17 that simulates a classical semantics. In section "Expressivity of Boxes" we show that this is impossible when assuming some straightforward semantics for a box embedding method. Having this negative result, we want to determine the positive: which ontologies are representable by a box embedding method in a pseudo-Tarski fashion? Is it possible to restrict the ontologies in such a way that box embedding methods can handle them?

This is discussed in the following from different angles: (i) as a starting point, a general box interpretation is defined to present a general, expressive embedding as a starting point that is closer to a classical interpretation than the general embedding method presented in Definition 1. (ii) Based on this interpretation, we discuss which ontologies are actually representable (in a pseudo-Tarski fashion), thus not influenced by the missing completeness. (iii) The learnability is considered: does the geometric construction introduce a bias into the learning approach?

Point (ii) and (iii) are introduced in the following and discussed in detail in section "$\mathcal{ELHO}(\circ)^\bot$ under HP-Semantics".

Completeness

When looking at Table 4 in (Bourgaux et al. 2024), it turns out that there are sound box embedding approaches (especially BoxEL (Xiong et al. 2022)) but none of the

box embeddings is complete. The proofs of this fact in Bourgaux et al. (2024) are based on individual properties of the respective box embedding approaches. This opens up one of the main questions of this paper: is it possible to find a box embedding approach that is complete? As this question is especially interesting for embedding methods of a sufficiently classical semantics, the question can be changed to: Is there a guaranteed entailment closed, weakly faithful embedding method S_M that is sound and complete? We are focusing here on one specific embedding method based on some basic assumptions. In section "Expressivity of Boxes" it turns out that this question has to be answered partly negatively. Though, that specific box embedding method is sound, it is not complete, meaning there are always $\mathcal{ELHO}(\circ)^\bot$ ontologies (and even $\mathcal{EL}_\bot$ ontologies) that can't be modeled by such a box embedding. As the specific interpretation considered for this result shares many properties of classical KBE approaches, this result is also directly applicable to standard KBE approaches and shows that they can not be complete, even if their specific problems of completeness would not have been the case.

This result is vital on the way towards a trustworthy (and high quality) KBE approach. Assume data and ontology are given and an embedding has been trained. If the loss of the trained model is greater than zero, then not all axioms and assertions of the ontology have been modeled correctly. This could happen out of several reasons and is not necessarily problematic. It could be, e.g., the case that the learning scheme is unable to find a perfect embedding and is stuck in a local minima. It could also be the case that the dimension of the embedding has been chosen too small and therefore the ontology is not representable.

Another point, where an advantage of the embedding comes into play is when handling erroneous assertions. Due to the effort of the embedding to model a regular, low dimensional embedding, such outliers are implicitly detected and corrected, leading to a non-problematic non-zero loss.

This is in principle nice to have, however, has the downside that it is not obvious whether such a corrected outlier was actually erroneous or not. At this point, it is necessary to determine whether the ontology was embeddable consistently at all. Due to the non-completeness of the approaches, it could be the case that even based on a perfect learning scheme and an appropriate dimensionality, an embedding being a shallow model of the ontology is not found. In this case, assertions (and also GCIs) are modeled incorrectly but not due to data regularities but solely due to a lack of representational ability. It is still possible that the embedding has the weak-pseudo-Tarski property based on some ontology. The embedded ontology is, however, not the ontology that should have been modeled but an ontology modeling some other facts. In that case improvements in data quality or the learning approach or a higher dimension do not solve the issue. Problems due to non-completeness need to be determined before considering other problems of the embedding approach.

Non-complete embedding approaches are still useful: first, they can be used for the ontologies that can be modeled. We determine these in section "Helly-Satisfiability". Additionally, this information can be used as a post-processing step: if a result, e.g., a predicted link, seems implausible, it can be tested whether this link has any connection to a non-representable part of the ontology.

Faithfulness

Assume now that an embedding approach exists that constructs for each ontology representable by a method S_M a weakly pseudo-Tarski shallow model. Due to interpreting the domain as real-valued vector space, it is possible to argue about similarity of individuals or concepts by considering distances between them. This is the so-called *geometric regularity principle*, a basic principle of models such as word2vec (Mikolov et al. 2013) and used ever since. In KGE-approaches, this principle is only valid to a limited extent, as modeling relations geometrically influences the representation (Paulheim et al. 2025). By introducing conceptual information in form of geometric constraints, the tight connection between data regularities and geometric regularities is even more relaxed, as the modeling is restricted by box-shaped constraints. This leads to a trade off: on the one hand, data could incorporate a bias that hints towards a specific assertion that seems to be plausible but still is incorrect, e.g., if there is only correlation and not causality. Then, the embedding should ignore this similarity and inject the ontological information to overcome this issue. On the other hand, the embedding approach should be bias-free in the sense that it is able to build a geometric model without being influenced by geometric restrictions.

Example 19. *Consider an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{\}$ and $\mathcal{A} = \{a : A, b : B\}$. Now, assume a primitive embedding approach S_M is given, that is only able to model exactly one box. Let E under S_M be a shallow model of $\mathcal{O}$. Then, $A^E = B^E$, thus the box representing A equals the box representing B, and it can be inferred that $a^E \in B^E$ and $b^E \in A^E$. This is in line with the ontology, thus the embedding is consistent. However, it seems not to be reasonable as there is no hint in the data pointing towards such an equivalence. In this case, the restrictions of the embedding approach prevent from finding a good solution.*

Therefore, we want to determine whether it is in theory possible to find a bias-free embedding, thus an embedding that is not restricted by geometric regularities but solely models the data. If such an embedding exists, then the trustworthiness of the embedding approach increases: The risk of inferring knowledge solely based on geometric regularities and not based on data regularities decreases. Note, however, that this does not guarantee for a totally geometric-bias free approach: it is still possible that the learning approach prefers the biased version over the unbiased one due to lower training complexity. It could also be the case that the chosen dimension introduces such a geometric bias. The level of bias of the approach can be described via the notion of *strong faithfulness*:

Definition 20. (Özçep et al. 2020). *Let $\mathcal{O}$ be a classically consistent (DL) ontology. We say that a shallow model E of $\mathcal{O}$ under S_M is*

- strongly Tbox-faithful *if for every GCI τ: if $E \models\!\!\!\sim \tau$, then $\mathcal{O} \models \tau$.*
- strongly Abox-faithful *if for every assertion α: if $E \models\!\!\!\sim \alpha$, then $\mathcal{O} \models \alpha$.*
- strongly faithful *if it is strongly Abox-faithful and strongly Tbox-faithful.*

Corollary 21. *Let $\mathcal{O}$ be a classically consistent DL ontology. If a shallow model E of $\mathcal{O}$ is strongly faithful, then it is also globally faithful (and thus also weakly faithful).*

The existence of a strongly faithful shallow model E of $\mathcal{O}$ means that the underlying embedding method S_M has the ability to model exclusively the axioms and assertions that are entailed by the ontology. Such an embedding is extremely large and can be considered as the most overfitted embedding possible: nothing new can be inferred, as exactly the existing information is modeled. Therefore, such an embedding is not relevant for usage in practice. It is, however, necessary for determining whether it is in theory possible, as then the actually learned embedding does not suffer from this type of bias.

Whereas Bourgaux et al. (2024) states that all box-based KBE approaches approaches considered are not strongly faithful, we are again interested in the question on regaining strong faithfulness: what is the influence of the restriction of box embeddings on modeling strongly faithful embeddings? Are there any ontologies that can be modeled strongly faithful despite the restricted expressivity of box embedding methods? This will is discussed in detail in section "Helly-Faithfulness".

After this informal introduction to the topic, in the following we discuss the three mentioned aspects in detail: in section "A Generalized Box Interpretation", a generalized box interpretation is presented, after that, in section "Expressivity of Boxes" general restrictions of box embeddings are discussed. Their influence to completeness and faithfulness is then discussed in the sections "Helly-Satisfiability" and "Helly-Faithfulness", resp.

A Generalized Box Interpretation

First, the question is tackled whether there is a specific embedding method S_M that is entailment closed, weakly faithful and possibly sound and complete and thus resembles a classical semantics as discussed in Theorem 17. This embedding should be also general enough to be able to act as a generalization of existing box-based KBE approaches or at least to allow for adapting the insights that are based on this embedding method to the KBE methods. This is particularly relevant to discriminate between issues with specific embedding approaches and problems inherent to box-based embeddings in general. Our first aim is to gain insights into the commonalities of all these approaches to find a generalized box embedding that can be (i) used as a basis for further considerations on the general ability of box embeddings and (ii) clarifies which fragments of the ontologies can be represented with the help of a Tarskian-style semantics. Starting with the general embedding method as defined in Definition 1, it is now considered how this can be specified.

Classically, the domain is $\mathbb{R}^n$ for some predefined n. All concepts are interpreted as boxes: the top concept as the whole space $\mathbb{R}^n$, the bottom concept as the empty set, and other concepts as specific boxes. Furthermore, conjunction of concepts is typically defined as set-intersection, with the exception of TransBox (Yang et al. 2025) where an approximated intersection is considered. We use here the classical intersection, as it is more often used and closer to an intuitive semantics. The main difference between

the various approaches lies in the definition of relations used. To define a well-behaved semantics for relations, we interpret them as sets over $\mathbb{R}^n \times \mathbb{R}^n$, however with the added condition that existential role restrictions always transform boxes into boxes. As the main feature of representations of relations is to model exactly this (or a subset of this) relation, it can be considered as general. This is exemplified in Example 25. We arrive at the following definition.

Definition 22. *A* box interpretation ξ *is a structure* $(\Xi, \cdot^\xi)$*, where* $\Xi = \mathbb{R}^n$ *for some* $n \in \mathbb{N}$*, and where* $\cdot^\xi$ *maps each concept name* $A \in \mathbf{C}$ *to some box in* $\mathcal{B}^n$*, each individual name* $c \in \mathbf{I}$ *to a point* $c^\xi \in \Xi$*, each nominal concept* $\{c\}$ *to the box* $\{c^\xi\}$*, and each role* $R \in \mathbf{R}$ *to a subset* $R^\xi \subseteq \Xi \times \Xi$ *such that for every* $B \in \mathcal{B}^n$*:* $R^{-1}(B) \in \mathcal{B}^n$*. A* box interpretation for arbitrary $\mathcal{ELHO}(\circ)^\bot$-concepts *is defined recursively as*

$$(\top)^\xi = \Xi \qquad (\bot)^\xi = \varnothing \qquad (C \sqcap D)^\xi = C^\xi \cap D^\xi$$
$$(\exists R.C)^\xi = \{x \in \Xi \mid \textit{there is } y \in \Xi \textit{ with } (x,y) \in R^\xi \textit{ and } y \in C^\xi\}$$
$$(R \circ S)^\xi = \left\{(a,c) \mid \exists b \in \Xi : (a,b) \in R^\xi, (b,c) \in S^\xi\right\}$$

A box interpretation ξ *models an Abox axiom* $a : C$ *for short* $\xi \Vdash a : C$ *iff* $a^\xi \in C^\xi$ *and it models an Abox axiom of the form* $(a,b) : R$ *iff* $(a^\xi, b^\xi) \in R^\xi$*.*

Clearly the structure $(\Xi, \cdot^\xi)$ can be considered as an instance of the embedding method S_M defined in Definition 1. When talking about box interpretations in the following, we are referring to the one defined here. Note, that this can be directly adapted to weaker ontologies such as in $\mathcal{EL}_\bot$ or even in ontologies not considering roles at all. Thus, this interpretation and its semantic can also be used as a building block for understanding other approaches based on box embeddings, e.g., the ones considering solely hierarchical structures. This interpretation is inspired by classical interpretations as defined in Table 1. It differs only in the assertion that classical interpretations consider concepts as arbitrary sets whereas in a box interpretation each concept is represented as a box. The definition of the roles ensures that each $(\exists R.C)^\xi$ results in a box, independent of the choice of C. In contrast to classical interpretations, the box interpretation allows for the notion of convexity and dimensionality.

The box interpretations of Definition 22 can be interpreted as a special type of classical interpretation in the following sense:

Proposition 23. *Let* ξ *be a box interpretation of an* $\mathcal{ELHO}(\circ)^\bot$*-ontology* $\mathcal{O}$ *such that* $\xi \Vdash \mathcal{O}$*. Then*

1. ξ *is entailment closed,*
2. ξ *is weakly faithful, and*
3. $\mathcal{O}$ *is satisfiable in standard DL semantics*

Proof. Let $\mathcal{O}$ be an $\mathcal{ELHO}(\circ)^\bot$-ontology. Let $\xi \vDash \mathcal{T} \cup \mathcal{A}$. First, it is shown that each concept in ξ is a box. Each concept symbol is interpreted as a box. $(\exists R.C)^\xi$ is defined

as $R^{-1}(C^\xi)$ and as C^ξ is a box, by definition also $(\exists R.C)^\xi$ is a box. Boxes are closed under intersection. Therefore, also $(C \sqcap D)^\xi$ is a box for arbitrary concepts C, D. Note that $\varnothing \in \mathcal{B}^n$, thus by definition $(C \sqcap D)^\xi = \varnothing$ also results in a box. $\bot^\xi, \top^\xi$ and $\{c\}^\xi$ for nominals $\{c\}$ are boxes by definition.

In the following, it is shown that the box interpretation is a special case of a classical interpretation. As classical interpretations are entailment-closed and weakly faithful, the proposition follows.

Let $\mathcal{I}$ be a classical interpretation and let ξ be a box interpretation that models all Tbox and Abox axioms. Let $\Delta = \Xi$, $c^\mathcal{I} = c^\xi$ for $c \in \mathbf{I}$. Let $C^\mathcal{I} = \{a \mid a \in C^\xi\}$ for $C \in \mathbf{C}$ and $R^\mathcal{I} = \{(a,b) \mid (a,b) \in R^\xi\}$ for $R \in \mathbf{R}$. Thus, $\mathcal{I} \vDash ax$ iff $\xi \Vdash ax$ for all assertions ax. Therefore, ξ can be interpreted as classical interpretation and thus is entailment closed and weakly faithful. As $\mathcal{I}$ is a classical interpretation, (3) follows trivially.

Corollary 24. *The embedding method $(\Xi, \cdot^\xi)$ is guaranteed to be entailment closed and weakly faithful and sound and thus fulfills the pseudo-Tarski property.*

This box interpretation is not only interesting on its own but also strongly connected to the existing embedding approaches, as exemplified in detail for TransBox in the following example.

Example 25. *This box interpretation is general in the sense that it allows for interpreting some of the existing box embedding methods as special cases. Consider, e.g., TransBox and especially the example mentioned in Figure 1 (b). There, individuals are defined as points in $\mathbb{R}^n$, concepts as boxes, and $\bot$ and $\top$ can be interpreted as the empty space and $\mathbb{R}^n$, resp. A direct translation of R^{box} to R^ξ is the following: $R^\xi = \{(a^\xi, b^\xi) \mid a^\xi \in R^{box} + b^\xi$ and $b^\xi \in \Xi\}$. As R^{box} is a box, also $R^{box} + b^\xi$ is a box. As translation with R^{box} is linear, also each translation of an arbitrary box results in a box. With $\mathbf{C} = \{Pizza\}, \mathbf{R} = \{eats\}$ and $\mathbf{I} = \{alice\}$, this definition leads to the example mentioned in Figure 1 (b).*

Now, after showing that this general box interpretation is useful as a generalization of existing approaches, the expressivity of this interpretation is discussed.

Expressivity of Boxes

Now, assume that a perfect learning approach and an arbitrary high dimensional embedding space are given. In the following, it is shown that even under these ideal circumstances, there exists not always an embedding as defined in Definition 22 that models the ontology. Therefore, these box representations ease the training but come to the prize of a restricted expressivity, as not every satisfiable $\mathcal{ELHO}(\circ)^\bot$-ontology has a box interpretation that satisfies $\mathcal{O}$: there is a satisfiable $\mathcal{ELHO}(\circ)^\bot$-ontology not having any box interpretation. Thus, box embedding methods, introduced also to improve trustworthiness of the approaches, by themselves add a source of lack of trustworthiness due to limits in expressivity. The problem can be visualized with the following example:

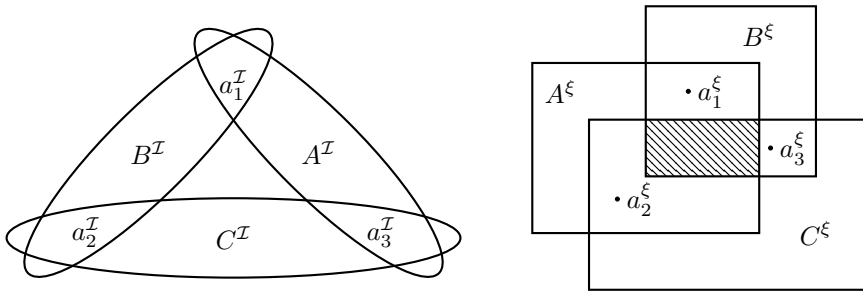


Figure 3. (a) a classical interpretation not fulfilling Helly's property; (b) a box interpretation with $\xi \nVdash A \sqcap B \sqcap C = \bot$ (as the shaded region represents $A^{\xi} \cap B^{\xi} \cap C^{\xi}$)

Example 26. *Given the ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{A \sqcap B \sqcap C = \bot\}$ and $\mathcal{A} = \{a_1 : A \sqcap B, a_2 : B \sqcap C, a_3 : A \sqcap C\}$. A possible DL-interpretation satisfying $\mathcal{O}$ can be seen in Figure 3 (a). However, the attempt to find a box interpretation leads to interpretations such as the one shown in Figure 3 (b). With boxes, it is necessary to dismiss either the axiom $A \sqcap B \sqcap C = \bot$ or to model one of the individuals incorrectly.*

This property is well-known and is called *Helly's Property (HP)*.

Definition 27. Helly's Property. *(adapted from (Eckhoff 1988)) A family B fulfills* Helly's Property *if it is the case that: $\bigcap_{b \in B} b \neq \varnothing$ if and only if for all $b_1, b_2 \in B$: $b_1 \cap b_2 \neq \varnothing$.*

It is in fact a long-standing result that boxes need to fulfill Helly's property.

Proposition 28. *(adapted from (Eckhoff 1988)) Each finite family $B \subset \mathcal{B}^n$ of axis-parallel boxes in $\mathbb{R}^n$ fulfills Helly's property, for any $n \in \mathbb{N}$.*

This property is not to be confused with the commonly known *Helly's Theorem* (Helly 1923) about the intersection of convex sets in $\mathbb{R}^n$ for fixed $n \in \mathbb{N}$. Helly's property is independent of the dimensionality of the boxes, thus using a higher dimension does not solve the issue. Examples that interfere with Helly's property can be found in many real world problems, e.g., in project management.

Example 29. Project Management Triangle (Van Wyngaard et al. 2012). *The project management triangle depicts a basic economic principle. The main aim of a production system is to optimize production cost, production time and scope (thus, the number of features of the product) at the same time to have a cheap, quickly produced but complex product. This is, however, not possible. Each two of the features can be optimized at the same time, coming to the cost of neglecting the third. Thus, it is possible to produce a cheap product fast but then without complex features. Producing a complex product cheaply is time consuming and producing a complex product fast is costly (see Figure 4 for a visualization). Note that this triangle resembles the shape of the counterexample for Helly's property in Figure 3.*

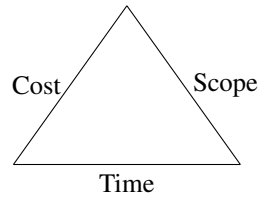


Figure 4. Project management triangle (Van Wyngaard et al. 2012)

Next to this well-known example, where the main aim is to optimize three features against each other, this property is also of importance in other use cases, where the focus is on pairwise intersection directly.

Example 30. Real-world example. *Other real-world examples can be found, e.g., in the animal domain when considering attributes of animals, e.g., "flying", "aquatic" and "mammal". There are aquatic mammals such as dolphins, there are flying mammals such as bats and (at least somehow) flying aquatic animals, namely flying fish. There are, however, no flying aquatic mammals, thus Helly's property is not fulfilled.*

Also in some of the standard real-world ontologies those problems can be found at least implicitly.

Example 31. Real-world ontologies. *When considering Helly's property on ontology level, interference with this property is only possible if disjointness axioms are modeled. An ontology of the animals of the last example would, e.g., consists of three concepts* aquatic animal, mammal *and* flying animal *where each of the two are pairwise intersecting. Such an ontology is not in line with Helly's property, as the Tbox contains an axioms stating that* $aquatic\_animal \sqcap mammal \sqcap flying\_animal \sqsubseteq \bot$. *Many real-world ontologies such as GALEN (Rector et al. 1996) and SNOMED (Donnelly 2006) are not modeling disjointness axioms at all and thus always fulfilling Helly's property. Therefore, in context of these ontologies it seems that at first sight, adherence to Helly's property is not problematic and therefore is not of any relevance. These axioms are, however, often not missing due to the fact that the concepts are allowed to intersect but due to the fact that the modeling of disjointness axioms has been neglected by the ontology developers, as hierarchies are mostly considered more important than disjointness axioms. Even if such a disjointness axiom is not modeled, it could exist both based on implicit knowledge and based on data regularities. One especially fatal problem occurring due to Helly's property is the consideration of contraindications of drugs. Thus, there could be three drugs being all pairwise compatible having severe side effects given all together.*

These three examples exemplify real-world use cases interfering with Helly's property. This necessitates further considerations on this property: what implications does it have that an embedding approach is not able to model something interfering with Helly's property?

In the following, the intuition of the restrictions of box interpretations given in Example 26 is proven formally with the help of box interpretations.

Proposition 32. *There exists a classically satisfiable $\mathcal{ELHO}(\circ)^\bot$ ontology $\mathcal{O}$ such that no box interpretation in $\mathbb{R}^n$, for arbitrary $n \in \mathbb{N}$, satisfies $\mathcal{O}$.*

Proof. The proof is based on Example 26. Given the ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{A \sqcap B \sqcap C = \bot\}$ and $\mathcal{A} = \{a_1 : A, a_1 : B, a_2 : B, a_2 : C, a_3 : C, a_3 : A\}$. $\mathcal{O}$ is satisfiable, see Figure 3 (a) for a possible model. The box interpretation as defined in Definition 22 models concepts as axis-parallel boxes. Therefore, it is due to Proposition 28 necessary that each interpretation fulfills Helly's property. As due to $\mathcal{A}$, the boxes representing A, B and C need to intersect pairwise (and are especially non-empty) with Proposition 28 it follows that $A^\xi \cap B^\xi \cap C^\xi \neq \varnothing$. Therefore, $\xi \nVdash A \sqcap B \sqcap C \sqsubseteq \bot$ and therefore $\xi \nVdash \mathcal{O}$.

This proof directly leads to the following corollary stating that not only $\mathcal{ELHO}(\circ)^\bot$ ontologies are suffering from this issue but all structures that are at least able to model concept conjunction and disjointness. The structure defined in Proposition 32 is a *Helly antipattern*, i.e. a collection of sentences that can not be satisfied in structures satisfying Helly's property. Obviously, this does not require the full expressivity of $\mathcal{ELHO}(\circ)^\bot$ as recorded in the following corollary:

Corollary 33. *Let $L \subseteq \mathcal{ELHO}(\circ)^\bot$ be any sublanguage of $\mathcal{ELHO}(\circ)^\bot$ that contains conjunctions, disjointness, and allows instantiations (expressing non-emptiness of concepts). Then L admits Helly antipatterns.*

Not every ontology suffer from restrictions due to Helly's property. There are many ontologies that can be modeled correctly. Therefore, it is necessary to be able to determine whether an ontology suffer from problems with Helly's property or not. Thus: (i) Is it possible to effectively check, for a given ontology $\mathcal{O}$, whether it has one interpretation fulfilling Helly's property, thus whether it has a box interpretation satisfying $\mathcal{O}$? The theoretical existence of such an interpretation does not state that it is constructable, thus of practical relevance for an embedding approach. Thus: (ii) when it exists, can we effectively construct a Helly-satisfying interpretation?

Due to Corollary 33, all the following considerations are not only relevant for box-based KBE-approaches but for all approaches modeling conceptual information, concept conjunction and disjointness. It is not even necessary to state this information explicitly, also implicit disjointness can not be modeled due to the restriction of box embedding methods to fulfill Helly's property. Thus, these problems are also occurring in query embedding with boxes (see, e.g., (Ren et al. 2020)) or when modeling relations as boxes (Abboud et al. 2020) and in many other related areas.

$\mathcal{ELHO}(\circ)^\bot$ under HP-Semantics

First, in the following, *Helly-satisfiability* is considered, thus the question of how to determine for a given ontology whether there is at least one box interpretation fulfilling HP, thus whether (in theory) there is a box embedding being a model of the ontology.

After that, *Helly-faithfulness* is considered, thus the question of what influence HP has on the faithfulness of the embedding.

Helly-Satisfiability

As shown in Proposition 32, not every classically satisfiable ontology is representable by a box embedding based on some general assumptions as defined in Definition 22. Therefore, we now want to consider Helly's property in ontologies further: What does it mean that an ontology "does not fulfill HP"? Can we determine the fragment of ontologies efficiently that fulfill HP?

First note that HP is tested on interpretation level.

Definition 34. *An interpretation $\mathcal{I}$ of an $\mathcal{ELHO}(\circ)^{\bot}$-ontology $\mathcal{O}$ is* Helly-closed *if for all $A^{\mathcal{I}}, B^{\mathcal{I}}, C^{\mathcal{I}} \in DC^{\mathcal{I}}_{\mathcal{O}}$, the set of definable concepts, Helly's property (Definition 27) is fulfilled. Thus if $A^{\mathcal{I}}, B^{\mathcal{I}}, C^{\mathcal{I}}$ are pairwise intersecting, then also $A^{\mathcal{I}} \cap B^{\mathcal{I}} \cap C^{\mathcal{I}} \neq \emptyset$.*

One challenge here is that it is not sufficient to test HP for all concept symbols or elements of the Abox but it needs to be tested for every definable concept in the interpretation.

An ontology thus "fulfills HP" if it has at least one Helly-closed model. This leads to the notion of *Helly-satisfiability*.

Definition 35. *An $\mathcal{ELHO}(\circ)^{\bot}$-ontology $\mathcal{O}$ is* Helly-satisfiable *if it has a model $\mathcal{I}$ that is Helly-closed.*

In the following, properties of Helly-satisfiable ontologies are discussed. Note that Helly's property relies on the Abox-level: Consider an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{A \sqcap B \sqcap C = \bot\}$ and $\mathcal{A} = \{\}$. A Helly-closed interpretation of $\mathcal{O}$ can be seen, e.g., in Figure 5 (a) (Page 32). This ontology has, however, the same Tbox as the ontology defined in Example 26 that does not have a Helly-closed interpretation. Therefore, it is not possible to define a rule on Tbox-level to capture adherence to HP.

Thus, the Abox needs to be considered. As a first step towards a procedure for checking whether an ontology is Helly-satisfiable, an Abox closure rule is defined as a closure procedure. It is based on the straightforward idea to enforce a Helly-closed interpretation by adding a new individual witnessing the intersection of three concepts every time when the premise of HP is fulfilled by the Abox, thus these three concepts are pairwise intersecting. Such an individual can be added, as it needs to exist in every Helly-closed model of the ontology.

Definition 36. Helly-Abox closure rule. *For all concept descriptions A, B, C: if $\mathcal{A}$ contains $\{a : A, a : B, b : B, b : C, c : A, c : C\}$ but there is no individual d with $\{d : A, d : B, d : C\} \in \mathcal{A}$. Then, add a new individual e to $\mathbf{I}$ and $\mathcal{A}' = \mathcal{A} \cup \{e : A, e : B, e : C\}$.*

Now, let a satisfiable ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be given. The Helly-Abox closure rule is applied until no new element is added. This rule checks whether there is a case where three concepts are pairwise intersecting. Then a new individual is added at the intersection of all three to circumvent contradictions to HP. If the resulting ontology $\mathcal{O}' = (\mathcal{T}, \mathcal{A}')$ is not satisfiable anymore, then the ontology is not Helly-satisfiable.

Corollary 37. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable ontology. Let $\mathcal{O}' = (\mathcal{T}, \mathcal{A}')$ be the ontology after applying the Abox-closure rule until termination. If $\mathcal{O}'$ is not satisfiable then the ontology is not Helly-satisfiable.*

The other direction of this statement, however, does not follow: Even if the extended ontology $\mathcal{O}'$ is satisfiable, it is not necessarily Helly-satisfiable. This is due to the fact that, as mentioned in Definition 34, HP needs to be tested on each definable concept in an interpretation. Especially due to existential restrictions, it is possible that individuals interfering with HP are implicitly enforced to exist in every possible model of the ontology but are not explicitly stated in the set of individuals $\mathbf{I}$.

Proposition 38. *There is a satisfiable ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{O}' = (\mathcal{T}, \mathcal{A}')$ being the Helly-Abox closed version of $\mathcal{O}$ such that $\mathcal{O}'$ is satisfiable but not Helly-satisfiable.*

Proof. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{A \sqcap B \sqcap C \sqsubseteq \bot, D \sqsubseteq \exists R.(A \sqcap C)\}$ and $\mathcal{A} = \{a : A, a : B, b : B, b : C, c : D\}$. This ontology is Abox-closed, thus $\mathcal{O}' = \mathcal{O}$, and satisfiable. As $\mathcal{O} \models D \sqsubseteq \exists R.(A \sqcap C)$ and $\mathcal{O} \models c : D$, it follows that $\mathcal{O} \cup \{A \sqcap C \sqsubseteq \bot\}$ is not satisfiable. In each model $\mathcal{I}$ of $\mathcal{O}$, there must be an element $d^{\mathcal{I}}$ with $(c^{\mathcal{I}}, d^{\mathcal{I}}) \in R^{\mathcal{I}}$ and $d^{\mathcal{I}} \in (A \sqcap C)^{\mathcal{I}}$. Therefore, the concepts A, B and C are pairwise intersecting in each model of $\mathcal{O}$ but not intersecting all three due to the Tbox axiom and thus interference with HP.

Therefore, a more in-depth strategy is needed to test for Helly-satisfiability. One idea is to create an interpretation and check for this interpretation whether it is Helly-closed. If it is Helly-closed, then the ontology is Helly-satisfiable. However, if not, then there could be a different Helly-closed interpretation. Therefore, we need to test the Helly-closure for an interpretation that is generic in the sense that this interpretation is Helly-closed if and only if the ontology is Helly-satisfiable. The idea is to find a *Helly-companion* of this ontology. This Helly-companion extends the ontology in such a way that all possibly Helly-incompatible aspects are incorporated in the Abox and thus, the Abox-closure rule is sufficient to test for Helly-satisfiability. This companion then trivially leads to a Helly-closed interpretation if the ontology is Helly-satisfiable.

A Helly-companion $\mathcal{O}'$ is an extension of the ontology $\mathcal{O}$, where the set of concept and role symbols remain unchanged, only individuals are potentially added to include the implicit instances that could lead to problems with Helly-satisfiability (see the proof of Proposition 38 for an example). These potentially added individuals are enforced by condition 2. of the definition. This condition adds for each necessary non-empty concept a witness. Then it is sufficient to apply the Helly-Abox closure. Note here that "concept C" is condition 2 again considers arbitrary concept descriptions and not only concepts in $\mathbf{C}$.

Definition 39. *An ontology $(\mathcal{T}', \mathcal{A}') = \mathcal{O}' \supseteq \mathcal{O} = (\mathcal{T}, \mathcal{A})$, with the sets $\mathcal{T}'$ and $\mathcal{A}'$ finite, is a* Helly-companion *of $\mathcal{O}$ if*

1. $\mathbf{I}(\mathcal{O}') \supseteq \mathbf{I}(\mathcal{O}), \mathbf{C}(\mathcal{O}') = \mathbf{C}(\mathcal{O}), \mathbf{R}(\mathcal{O}') = \mathbf{R}(\mathcal{O})$*; (Signature extends only ind. names)*

2. *If for some concept C we have $\mathcal{O}' \cup \{C \sqsubseteq \bot\}$ is inconsistent, then there exists a $d \in \mathbf{I}(\mathcal{O}')$ such that $\mathcal{O}' \models d : C$; (Every necessarily non-empty concept is witnessed.)*

3. *$\mathcal{A}'$ is Helly-Abox closed for $\mathbf{I}(\mathcal{O}')$. (All Helly scenarios are witnessed.)*

This is exemplified in the following example.

Example 40. Helly companion. *Consider the ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{\}$ and $\mathcal{A} = \{a_1 : A, a_1 : B, a_2 : B, a_2 : C, a_3 : \exists R.(A \sqcap C)\}$. This is an ontology similar to the one in the proof of Proposition 32 but with an empty Tbox. The ontology is Helly-Abox closed, however $\mathcal{O}' = \mathcal{O}$ is not a Helly-companion, as $\mathcal{O}' \cup \{A \sqcap C \sqsubseteq \bot\}$ is inconsistent. To define $\mathcal{O}'$ as a Helly-companion of $\mathcal{O}$ a new individual a_4 needs to be added, thus $\mathbf{I}' = \mathbf{I} \cup \{a_4\}$. As $\mathcal{O} \cup \{A \sqcap C \sqsubseteq \bot\}$ is inconsistent, the Abox $\mathcal{A}$ is extended to $\mathcal{A}' = \mathcal{A} \cup \{a_4 : A, a_4 : C, (a_3, a_4) : R\}$. As a next step, the Helly-Abox-closure needs to be applied to $\mathcal{O}'$. Thus, an individual a_5 is added to $\mathbf{I}'$, thus $\mathbf{I}'' = \mathbf{I}' \cup \{a_5\}$ and $\mathcal{A}'' = \mathcal{A}' \cup \{a_5 : A, a_5 : B, a_5 : C\}$. As every necessarily non-empty concept is witnessed in $\mathcal{O}''$ and $\mathcal{A}''$ is Helly-Abox closed, $\mathcal{O}''$ is a Helly-companion of $\mathcal{O}$. With the help of this Helly-companion, the definition of a Helly-closed interpretation of the ontology is trivially possible. As an interpretation satisfying $\mathcal{O}''$ also satisfies $\mathcal{O}$, there is a Helly-closed interpretation of $\mathcal{O}$. Therefore, $\mathcal{O}$ is Helly-satisfiable.*

The notion of an Helly-companion now allows for identifying the Helly-satisfiable ontologies as exactly the ones that have a consistent Helly-companion.

Proposition 41. *A satisfiable $\mathcal{ELHO}(\circ)^\bot$-ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ is Helly-satisfiable if and only if there exists a consistent Helly-companion $\mathcal{O}'$ of $\mathcal{O}$.*

Proof sketch. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^\bot$ ontology.

→ Let $\mathcal{O}$ be Helly-satisfiable. A Helly-companion can be constructed by applying a tableau-like algorithm to the Abox. If, e.g., $\{a : A\} \subseteq \mathcal{A}$ and $\{A \sqsubseteq B\} \subseteq \mathcal{T}$, then $\mathcal{A}' = \mathcal{A} \cup \{a : B\}$. Then, the Abox-closure rule is applied. This makes it necessary to apply the tableau-like algorithm again to witness every necessarily non-empty concept. These two steps are repeated until nothing new is added. Next, it is discussed whether $\mathcal{O}'$ is satisfiable. Assume for the sake of contradiction that $\mathcal{O}'$ is not satisfiable. $\mathcal{O}$ is satisfiable by definition. Fulfilling condition 2 does not change the satisfiability of an ontology $\mathcal{O}'$ compared to $\mathcal{O}$, as it only makes implicitly existing individuals explicit. Thus, the Abox-closure needs to cause the loss of satisfiability of $\mathcal{O}'$. This, however, could only be the case if $\mathcal{O}$ would not be Helly-satisfiable. This is a contradiction to the assumption and therefore, $\mathcal{O}'$ is a satisfiable Helly-companion.

← Let $\mathcal{O}'$ be a consistent Helly-companion of $\mathcal{O}$. Thus, there is an interpretation $\mathcal{I}$ such that $\mathcal{O}' \models \mathcal{I}$. As $\mathcal{O} \subseteq \mathcal{O}'$, it follows that $\mathcal{I}$ is also a model of $\mathcal{O}$. It remains to show that there is a model $\mathcal{I}$ of $\mathcal{O}'$ that is Helly-closed. As $\mathcal{A}'$ is Helly-Abox closed and every necessarily non-empty concept is witnessed, a trivial interpretation of $\mathcal{O}'$ can be used (thus the interpretation resembling the Abox). This is by definition Helly-closed.

The detailed proof and especially the proof of termination of the construction of a Helly-companion can be found in the appendix.

This shows that it is (i) possible to determine whether an ontology is not Helly-satisfiable, thus whether there can't be a box embedding approach relying on the standard assumptions correctly modeling this ontology. In the detailed proof, it turns out that for a Helly-satisfiable ontology $\mathcal{O}$, not only a Helly-companion exist but also always a finite one. Thus, it also shows (ii) that if an ontology is Helly-satisfiable then a finite Helly-satisfiable model exists, thus there is (at least a theoretical) possibility to construct such a model.

Proposition 42. *Given a Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology $\mathcal{O}$, a finite model fulfilling HP can be found in finite time.*

Proof. This follows directly out of the proof of Proposition 41, as such a model can be directly defined with the help of the Helly-companion.

The previous results lead directly to the following theorem:

Theorem 43. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^{\bot}$ ontology. It is in finite time possible to test whether $\mathcal{O}$ is Helly-satisfiable and, if it is the case, a finite Helly-closed model of $\mathcal{O}$ can be found in finite time.*

In this section it was shown that Helly's property is a relevant restriction of box embedding approaches. These approaches can't be complete and independent of the exact modeling strategy, there will be always $\mathcal{ELHO}(\circ)^{\bot}$ ontologies that can't be modeled. It turned out that Helly's property acts especially on Abox-level. In the following example, the relevance of this result is discussed for those real-world ontologies that have an empty Abox and thus seem, on first sight, not influenced by Helly's property.

Example 44. HP for an empty Abox. *Many real world ontologies such as the Gene Ontology have an empty Abox. Are these ontologies not influenced by problems emerging from Helly's property, as ontologies with an empty Abox are always Helly-satisfiable? Also for these ontologies, HP is relevant, due to two reasons: first, it is plausible to assume that axioms of the form $A \sqcap B \sqsubseteq C$ are implying that $A \sqcap B \not\sqsubseteq \bot$, as otherwise the axiom $A \sqcap B \sqsubseteq \bot$ could have been stated directly (see, e.g., (Jackermeier et al. 2024)). This goes in line with the* coherence principle *for ontologies (see, e.g., Osman et al. (2021)) stating that each concept defined in an ontology should also be satisfiable, thus should have non-empty interpretations. Therefore, it is possible to make this coherence explicitly by extending $\mathbf{I}$ by individuals and the Abox by instantiations of the respective concepts. Then, the Abox is populated and therefore, Helly's property need to be considered. One even more relevant problem occurs in the context of faithfulness: The adherence to HP enforces the embedding to model specific GCIs and assertions that are neither following from the ontology nor from the data but solely from the restrictions of the embedding. This is discussed in detail in the next section in Example 49.*

These considerations are a first step towards a characterization of the exact representability of box embeddings: Is Helly's property not only a necessary but also a sufficient condition? Thus, is every Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology representable via box embeddings? This is, up to our knowledge, still an open question and topic of future work. We continue in the following to focus on Helly's property and discuss how it not only influences the completeness of the box embeddings but also its faithfulness.

Helly-Faithfulness

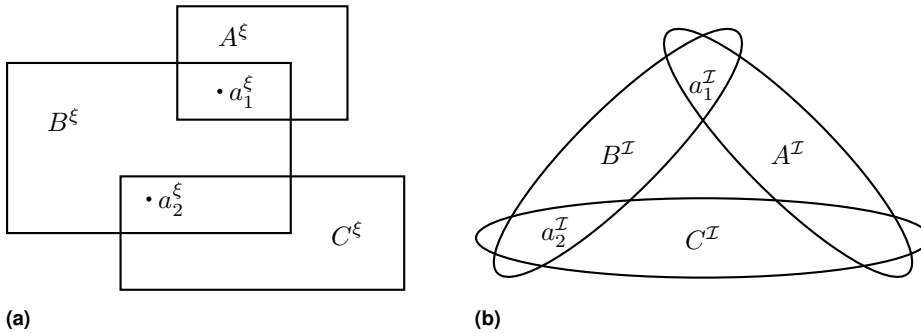
As argued in section "Towards Trustworthy and Interpretable Box Embeddings", not only satisfiability but also faithfulness** is relevant for discussing the expressivity of box embedding approaches. The basic idea of faithfulness is to allow the embedding approach to be solely based on data regularities and not on restrictions of the geometric embedding. Therefore, e.g., a new link should be predicted based on the data regularities and not because the embedding is not able to represent the case where this link does not exists. A faithful embedding thus is an embedding where exactly and only the information is represented that is stated in the ontology. The definition can be found in Definition 20. It has been shown for the DL $\mathcal{EL}$ based on modeling concepts as convex regions (Lacerda et al. 2024) and also for $\mathcal{ALC}$ based on modeling concepts as closed convex cones (Özçep et al. 2023) that each $\mathcal{EL}$ resp. $\mathcal{ALC}$ ontology is satisfiable if and only if it has a faithful model based on the resp. representation. First note, that each ontology trivially has a faithful model (when not considering HP). Now the question is whether and how HP influences the existence of such a model. Obviously, an ontology that is not Helly-satisfiable can't have a faithful Helly-closed model, as it has no Helly-closed model at all. Therefore, we relax the question: can every Helly-satisfiable ontology be satisfied by a faithful box interpretation? This question has to be answered negatively: Helly's property does not only influence the completeness, it also influences the faithfulness even for Helly-satisfiable ontologies. Consider the following example:

Example 45. *Consider an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$. Assume that $\mathbf{I} = \{a_1, a_2\}$ and let $\mathcal{T} = \{A \sqcap B \sqcap C = \bot\}$ and $\mathcal{A} = \{a_1 : A \sqcap B, a_2 : B \sqcap C\}$. This is representable by a box interpretation ξ by assuming that $\xi \Vdash A \sqcap C = \bot$. An example for ξ can be seen in Figure 5 (a). Thus, the ontology is Helly-satisfiable. The interpretation is, however, not a Tbox-faithful one, as $\xi \Vdash A \sqcap C \sqsubseteq \bot$ but $\mathcal{O} \not\models A \sqcap C \sqsubseteq \bot$. A Tbox-faithful interpretation (but without considering boxes) can be seen in Figure 5 (b). There, $A^{\mathcal{I}}$ and $C^{\mathcal{I}}$ are intersecting, which shows the missing knowledge on whether the conjunction of A and C is empty or not.*

This can also be exemplified by a real-world example:

Example 46. Problems with faithfulness in a real-world setting. *Consider the ontology of Example 45 in the context of the animal example of Example 30. Let A be the concept* aquatic animal*, B be* mammal *and C be* flying animal*. Then, every embedding ξ of*

** In the following, the term "faithfulness" refers to strong faithfulness.



(a) **(b)**

Figure 5. (a) a Helly-closed box interpretation, enforcing $A^\xi \cap C^\xi \subseteq \varnothing$; (b) a non Helly-closed but faithful classical interpretation of the ontology defined in Example 19.

this ontology would model that there aren't any aquatic animal that can fly. Though, this is plausible based on the ontology, it is also plausible that there are flying aquatic animals. Box embeddings would therefore decrease the expressive capabilities even when the ontology is in principle representable.

Thus, if a Helly-satisfiable ontology also has models not fulfilling HP, then faithfulness is lost: to be in line with Helly's property, implicit assumptions on the geometry need to be taken. In the example, it is assumed that $A \sqcap C \sqsubseteq \bot$. This is, however, not due to data regularities but solely due to the fact that it could not be modeled otherwise. In reality, it could be possible that the data hints towards the fact that A and C should be combineable. The following proposition states in which cases Helly's property leads to problems with gaining a strongly faithful interpretation. An ontology $\mathcal{O}$ could have a model $\mathcal{I}$ that is not Helly-closed, therefore, some three concepts are pairwise intersecting but not all three. This could be due to two reasons: either the non-intersection is an axiom of the ontology or this non-intersection is only entailed by this specific interpretation. The second case is not a problem, as such an interpretation would not be a strongly faithful interpretation anyway, also not in a non-restricted setting: there is a disjointness entailed by $\mathcal{I}$ that is not entailed by $\mathcal{O}$. The first case, however, is problematic: it states that it is not possible to model that all of the three concepts are pairwise intersecting and therefore, there can't be any strongly faithful model (see also Example 45).

Proposition 47. *Given an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$. Let $\mathcal{I}$ be a model of $\mathcal{O}$ such that*

1. *$\mathcal{I}$ is not Helly-closed and*
2. *$\mathcal{O} \cup \mathfrak{A}$ is not Helly-satisfiable for some set $\mathfrak{A}$ of individual assertions with $\mathcal{I} \models \mathfrak{A}$*

If at least one such model of $\mathcal{O}$ exists, then $\mathcal{O}$ has no strongly faithful Helly-closed interpretation.

Proof. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an ontology. Let $\mathcal{I}$ be a model of $\mathcal{O}$ such that $\mathcal{I}$ is not Helly-closed and $\mathcal{O} \cup \mathfrak{A}$ is not Helly-satisfiable for some set $\mathfrak{A}$ of individual assertions with

$\mathcal{I} \models \mathfrak{A}$. As $\mathcal{I} \models \mathfrak{A}$ and $\mathcal{I}$ is a model of $\mathcal{O}$, $\mathcal{O} \cup \mathfrak{A}$ is also satisfiable. Let $\mathcal{O}' = \mathcal{O} \cup \mathfrak{A}$. $\mathcal{O}'$ has a model but not a Helly-closed model. Thus, for all models $\mathcal{I}$ of $\mathcal{O}'$ by definition of Helly-closure, it follows that there are $A^{\mathcal{I}}, B^{\mathcal{I}}, C^{\mathcal{I}} \in DC_{\mathcal{O}}^{\mathcal{I}}$ that are pairwise intersecting but $A^{\mathcal{I}} \cap B^{\mathcal{I}} \cap C^{\mathcal{I}} = \varnothing$. Thus $\mathcal{I} \models A \sqcap B \sqcap C \sqsubseteq \bot$ for all $\mathcal{I}$. As $\mathcal{T} = \mathcal{T}'$, thus the Tbox of $\mathcal{O}'$ remained unchanged from the Tbox of $\mathcal{O}$, that means that $\mathcal{O} \models A \sqcap B \sqcap C \sqsubseteq \bot$. Therefore, there can't be any Helly-closed model $\mathcal{I}$ of $\mathcal{O}$ such that $\mathcal{I} \not\models A \sqcap B \sqsubseteq \bot$ and $\mathcal{I} \models B \sqcap C \sqsubseteq \bot$ and $\mathcal{I} \not\models A \sqcap C \sqsubseteq \bot$. Therefore, there can't be a Helly-closed interpretation of $\mathcal{O}$ that is strongly faithful.

It is not the case that all Helly-satisfiable models have a faithful model. Therefore, we are defining *Helly-faithfulness* in the style of Helly-satisfiabilty to determine the type of ontologies that are faithfully satisfiable with box interpretations.

Definition 48. *An $\mathcal{ELHO}(\circ)^{\bot}$-ontology $\mathcal{O}$ is* Helly-faithful *if there is no model of $\mathcal{O}$ that fulfills the properties as defined in Proposition 47.*

Helly-faithfulness in this context does not mean that the specific model is actually faithful. It means that Helly's property does not influence the considerations of the faithfulness. Thus, it is a necessary but not sufficient condition.

Example 49. Example 44 continued. *As discussed in Example 44, real-world ontologies often have an empty Abox. This, however, does not influence the considerations on faithfulness. Also for an empty Abox, the same problem as discussed in Example 19 can occur. Let ξ be a box interpretation of an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{A \sqcap B \sqcap C \sqsubseteq \bot\}$, $\mathcal{A} = \{\}$. Then all three intersections $(A \sqcap B)^{\xi}, (B \sqcap C)^{\xi}$ and $(A \sqcap C)^{\xi}$ can be either empty or not. As the Tbox includes $A \sqcap B \sqcap C \sqsubseteq \bot$, at least one of these three intersections needs to be empty. Then, e.g., $\xi \models A \sqcap B \sqsubseteq \bot$, however, $\mathcal{O} \not\models A \sqcap B \sqsubseteq \bot$ and no faithful Helly-closed model of $\mathcal{O}$ exists, even for an empty Abox.*

In this section, the influence of Helly's property on strong faithfulness of box embeddings has been discussed. This is a first step towards defining for which ontologies a faithful model based on box interpretations can be found. Faithfulness is a hard to accomplish goal, as has been shown, e.g., by Özçep et al. (2023) where strong faithfulness for cone-based interpretation in $\mathcal{ALC}$ has been discussed. There, an infinite dimensional space was needed for such a definition. Whereas $\mathcal{ALC}$ is of higher expressivity than $\mathcal{ELHO}(\circ)^{\bot}$, one of the main problems remains also in simpler ontologies: when modeling existential restrictions faithfully, it is necessary to model for each role R with $\exists R.\top \not\equiv \bot$ that $(\exists R.\exists R.\top)^{\mathcal{I}} \subset (\exists R.\top)^{\mathcal{I}}$, that $(\exists R.\exists R.\exists R.\top)^{\mathcal{I}} \subset (\exists R.\exists R\top)^{\mathcal{I}}$ and so on to gain a faithful interpretation $\mathcal{I}$ of the ontology $\mathcal{O}$. The classical example for this problem is Narcissus who loves himself but also loves someone who loves themselves and so on (Baader and Küsters 2006). These and other questions need to be tackled to consider strong faithfulness for box interpretation. This is the topic of future work. Here, we have been able to show some necessary restrictions of the ontology, as without a Helly-faithful ontology, a faithful model definitely does not exists.

Conclusion and Outlook

KBE approaches are useful tools for inference tasks. However, a KBE approach is only trustworthy and transparent if the bias of the approach is known, thus if we are able to detect and distinguish whether an inference is based on regularities in the data or restrictions of the model. We showed that KBE approaches based on boxes introduce a structural bias in the form of Helly's property imposed on the learned embedding. This bias has influence both on whether the interpretation is consistent and whether the inferences are based on geometric regularities. These results are not only relevant for the case of KBE but need to be considered in all approaches considering boxes for representing structured information based on conjunctions, disjointness, and instantiation. In future work, it is necessary to consider the dimensionality of interpretations in order to determine whether an ontology can not only be modeled in theory but also in a restricted low-dimensionality environment. Additionally, efficient tools and implementations needs to be developed that allow to be able to detect for an ontology in practice whether it suffers from problems with Helly's property. This would provide the basis to develop strategies to actively circumvent certain problems before starting the training process. Another interesting question is whether there are axioms that are preferably learned: for disjointness of concepts, it is, e.g., enough if the two respective boxes are disjoint in one dimension. In contrast, for non-disjointness, it is necessary that two boxes intersect in every dimension. There are use cases where it is appropriate to search for interpretations that only partially model the given ontology. Examples can include cases where the ontology is inconsistent, contains 'non-essential' axioms, or idiosyncratic individuals that could be omitted. Thus it will be essential to understand the deeper interplay between constraints imposed by the embedding semantics, restrictions imposed by the learning approach, and requirements imposed by the ontology languages. Therefore, the capabilities for representation of existing KBE-approaches could be considered further, thus, e.g., for which ontologies they allow for weakly/strongly faithful or entailment closed embeddings. Additionally, the abstract box embedding method could be used to find constraints as necessary and sufficient conditions for faithful, entailment closed, sound or complete models.

References

Abboud R, Ceylan I, Lukasiewicz T and Salvatori T (2020) BoxE: A Box Embedding Model for Knowledge Base Completion. In: *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*. pp. 9649–9661.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* 25(1): 25–29. DOI:10.1038/75556.

Baader F, Brandt S and Lutz C (2005) Pushing the $\mathcal{EL}$ Envelope. In: *IJCAI'05: Proceedings of the 19th International Joint Conference on Artificial Intelligence*. pp. 364–369. DOI: 10.25368/2022.144.

Baader F, Calvanese D, McGuinness DL, Nardi D and Patel-Schneider PF (eds.) (2007) *The Description Logic Handbook: Theory, Implementation and Applications*. Second edition. Cambridge University Press. DOI:10.1017/CBO9780511711787.

Baader F and Küsters R (2006) Nonstandard Inferences in Description Logics: The Story So Far. In: Gabbay D, Goncharov S and Zakharyaschev M (eds.) *Mathematical Problems from Applied Logic I*, *International Mathematical Series*, volume 4. Springer, pp. 1–75.

Baader F, Lutz C, Sturm H and Wolter F (2002) Fusions of description logics and abstract description systems. *J. Artif. Intell. Res.* 16: 1–58. DOI:10.1613/JAIR.919.

Baader F and Sattler U (2001) An Overview of Tableau Algorithms for Description Logics. *Studia Logica* 69(1): 5–40. DOI:10.1023/a:1013882326814.

Boratko M, Zhang D, Monath N, Vilnis L, Clarkson KL and McCallum A (2021) Capacity and Bias of Learned Geometric Embeddings for Directed Graphs. In: Ranzato M, Beygelzimer A, Dauphin YN, Liang P and Vaughan JW (eds.) *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. pp. 16423–16436.

Bordes A, Usunier N, Garcia-Duran A, Weston J and Yakhnenko O (2013) Translating Embeddings for Modeling Multi-Relational Data. In: *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*. p. 2787–2795.

Bourgaux C, Guimarães R, Koudijs R, Lacerda V and Ozaki A (2024) Knowledge Base Embeddings: Semantics and Theoretical Properties. In: *Proceedings of the TwentyFirst International Conference on Principles of Knowledge Representation and Reasoning*. International Joint Conferences on Artificial Intelligence Organization, pp. 823–833. DOI: 10.24963/kr.2024/77.

Chen J, Mashkova O, Zhapa-Camacho F, Hoehndorf R, He Y and Horrocks I (2025) Ontology Embedding: A Survey of Methods, Applications and Resources. *IEEE Transactions on Knowledge and Data Engineering* : 1–20DOI:10.1109/tkde.2025.3559023.

Donnelly K (2006) SNOMED-CT: The Advanced Terminology and Coding System for EHealth. *Studies in Health Technology and Informatics* 121: 279–290.

Eckhoff J (1988) Intersection properties of boxes. Part I: An upper-bound theorem. *Israel Journal of Mathematics* 62(3): 283–301. DOI:10.1007/bf02783298.

Helly E (1923) Über Mengen konvexer Körper mit gemeinschaftlichen Punkten. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 32: 175–176.

Hogan A, Blomqvist E, Cochez M, D'Amato C, Melo GD, Gutierrez C, Kirrane S, Gayo JEL, Navigli R, Neumaier S, Ngomo ACN, Polleres A, Rashid SM, Rula A, Schmelzeisen L, Sequeda J, Staab S and Zimmermann A (2021) Knowledge Graphs. *ACM Computing Surveys (CSUR)* 54(4): 1–37. DOI:10.1145/3447772.

Jackermeier M, Chen J and Horrocks I (2024) Dual Box Embeddings for the Description Logic $\mathcal{EL}^{++}$. In: *Proceedings of the ACM Web Conference 2024*, WWW '24. ACM, pp. 2250–2258. DOI:10.1145/3589334.3645648.

Kulmanov M, Liu-Wei W, Yan Y and Hoehndorf R (2019) EL Embeddings: Geometric Construction of Models for the Description Logic $\mathcal{EL}^{++}$. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint

Conferences on Artificial Intelligence Organization, pp. 6103–6109. DOI:10.24963/ijcai.2019/845.

Kutz O, Wolter F and Zakharyaschev M (2002) Connecting abstract description systems. In: Fensel D, Giunchiglia F, McGuinness DL and Williams M (eds.) *Proceedings of the Eighth International Conference on Principles and Knowledge Representation and Reasoning (KR-02), Toulouse, France, April 22-25, 2002*. Morgan Kaufmann, pp. 215–226.

Lacerda V, Ozaki A and Guimarães R (2024) FaithEL: Strongly TBox Faithful Knowledge Base Embeddings for $\mathcal{EL}$. In: Kirrane S, Simkus M, Soylu A and Roman D (eds.) *Rules and Reasoning - 8th International Joint Conference, RuleML+RR 2024, Bucharest, Romania, September 16-18, 2024, Proceedings*, *Lecture Notes in Computer Science*, volume 15183. Springer, pp. 191–199. DOI:10.1007/978-3-031-72407-7_14.

Leemhuis M and Kutz O (2025) Understanding the expressive capabilities of knowledge base embeddings under box semantics. In: H Gilpin L, Giunchiglia E, Hitzler P and van Krieken E (eds.) *Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning*, *Proceedings of Machine Learning Research*, volume 284. PMLR, pp. 303–321.

Leemhuis M, Özçep Ö and Wolter D (2022) Knowledge Graph Embeddings with Ontologies: Reification for Representing Arbitrary Relations. In: Bergmann R, Malburg L, Rodermund S and Timm I (eds.) *German Conference on Artificial Intelligence (Künstliche Intelligenz)*, number 13404 in Lecture Notes in Computer Science. Springer, Springer International Publishing, pp. 146–159. DOI:10.1007/978-3-031-15791-2_13.

Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, van Kleef P, Auer S and Bizer C (2015) DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6(2): 167–195. DOI:10.3233/SW-140134.

Mikolov T, Sutskever I, Chen K, Corrado G and Dean J (2013) Distributed Representations of Words and Phrases and Their Compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems* 2: 3111–3119.

Mondal S, Bhatia SK and Mutharaju R (2021) EmEL++: Embeddings for $\mathcal{EL}^{++}$ Description Logic. In: Martin A, Hinkelmann K, Fill HG, Gerber A, Lenat D, Stolle R and van Harmelen F (eds.) *Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021)*, *CEUR Workshop Procedings*, volume 2846.

Osman I, Yahia SB and Diallo G (2021) Ontology Integration: Approaches and Challenging Issues. *Inf. Fusion* 71: 38–63. DOI:10.1016/J.INFFUS.2021.01.007.

Özçep Ö, Leemhuis M and Wolter D (2020) Cone Semantics for Logics with Negation. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. International Joint Conferences on Artificial Intelligence Organization, pp. 1820–1826. DOI:10.24963/ijcai.2020/252.

Özçep Ö, Leemhuis M and Wolter D (2023) Embedding Ontologies in the Description Logic $\mathcal{ALC}$ by Axis-Aligned Cones. *Journal of Artificial Intelligence Research* 78: 217–267. DOI:10.1613/jair.1.13939.

Paulheim H, Hubert N and Portisch J (2025) What Do Knowledge Graph Embeddings Learn to Represent? In: *Handbook on Neurosymbolic AI and Knowledge Graphs*. IOS Press, pp. 166–195. DOI:10.3233/faia250206.

Peng X, Tang Z, Kulmanov M, Niu K and Hoehndorf R (2022) Description Logic $\mathcal{EL}^{++}$ Embeddings with Intersectional Closure. *ArXiv* DOI:10.48550/arXiv.2202.14018.

Rector AL, Rogers JE and Pole P (1996) The GALEN high level ontology. In: *Medical Informatics Europe'96*. IOS Press, pp. 174–178.

Ren H, Hu W and Leskovec J (2020) Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. *arXiv* DOI:10.48550/arXiv.2002.05969.

Roberts FS (1969) On the Boxicity and Cubicity of a Graph. *Recent progress in combinatorics* 1(1): 301–310.

Van Wyngaard CJ, Pretorius JHC and Pretorius L (2012) Theory of the triple constraint — A conceptual review. In: *2012 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, pp. 1991–1997. DOI:10.1109/ieem.2012.6838095.

Vrandecic D and Krötzsch M (2014) Wikidata: a Free Collaborative Knowledgebase. *Commun. ACM* 57(10): 78–85. DOI:10.1145/2629489.

Xiong B, Potyka N, Tran TK, Nayyeri M and Staab S (2022) Faithiful Embeddings for $\mathcal{EL}^{++}$ Knowledge Bases. In: *The Semantic Web – ISWC 2022*. Springer International Publishing, pp. 22–38. DOI:10.1007/978-3-031-19433-7_2.

Yang H, Chen J and Sattler U (2025) TransBox: $\mathcal{EL}^{++}$-closed Ontology Embedding. In: *Proceedings of the ACM on Web Conference 2025*, WWW '25. ACM, p. 22–34. DOI: 10.1145/3696410.3714672.

Proofs of Section "Helly-Satisfiability"

The basic idea is to define an algorithm that constructs a Helly-companion by iteratively extending the Abox of the ontology with the concepts that need to be populated in every model of the ontology. This is done based on a construction principle similar to a tableau-algorithm. Therefore, first, the transformation rules are given. They are adapted from the $\mathcal{ALC}$-tableau, see, e.g., (Baader and Sattler 2001). Assume in the following for simplicity that the ontology is given in normal form.

The $\rightarrow_{\sqcap}$-rule
Condition: $\mathcal{A}$ contains $x : (C_1 \sqcap C_2)$, but not both $x : C_1$ and $x : C_2$.
Action: $\mathcal{A}' := \mathcal{A} \cup \{x : C_1, x : C_2\}$.

The $\rightarrow_{\sqsubseteq}$-rule
Condition: $\mathcal{T}$ contains $C \sqsubseteq D$, $\mathcal{A}$ contains $x : C$ (resp. $x : C_1$ and $x : C_2$ if $C = C_1 \sqcap C_2$), but not $x : D$.
Action: $\mathcal{A}' = \mathcal{A} \cup \{x : D\}$.

The $\rightarrow_{\exists\sqsubseteq}$-rule
Condition: $\mathcal{T}$ contains $\exists R.C \sqsubseteq D$, $\mathcal{A}$ contains $(x, y) : R, y : C$, but not $x : D$.
Action: $\mathcal{A}' = \mathcal{A} \cup \{x : D\}$.

The $\rightarrow_{R\sqsubseteq S}$-rule
Condition: $\mathcal{T}$ contains $R \sqsubseteq S$, $\mathcal{A}$ contains $(x, y) : R$, but not $(x, y) : S$.
Action: $\mathcal{A}' = \mathcal{A} \cup \{(x, y) : S\}$.

The $\to_\circ$-rule

Condition: $\mathcal{T}$ contains $R_1 \circ R_2 \sqsubseteq S$, $\mathcal{A}$ contains $(x,y) : R_1, (y,z) : R_2$, but not $(x,z) : S$.
Action: $\mathcal{A}' = \mathcal{A} \cup \{(x,z) : S\}$.

The $\to_\exists$-rule

Condition: $\mathcal{A}$ contains $x : \exists R.C$, but there is no individual name z such that $z : C$ and $(x,z) : R$ are in $\mathcal{A}$.
Action: $\mathcal{A}' := \mathcal{A} \cup \{y : C, (x,y) : R\}$ where y is an individual name not occurring in $\mathcal{A}$.

With these transformation rules, a tableau-inspired algorithm can be defined.

Definition 50. Adapted tableau algorithm for $\mathcal{ELHO}(\circ)^\bot$. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELHO}(\circ)^\bot$-ontology. Apply the transformation rules iteratively to $\mathcal{A}$ by preferring all other rules over the $\to_\exists$-rule. The blocking condition is defined as usual (see (Baader and Sattler 2001)): the application of the rule $\to_\exists$ to an individual x is blocked by an individual y in an Abox $\mathcal{A}$ iff $\{D \mid x : D \in \mathcal{A}\} \sqsubseteq \{D' \mid y : D' \in \mathcal{A}\}$.*

Based on classical tableau algorithms and due to the simplicity of $\mathcal{ELHO}(\circ)^\bot$, it can be proven that the algorithm terminates and that $\mathcal{O}' = (\mathcal{T}, \mathcal{A}')$ is satisfiable if $\mathcal{O}$ is satisfiable. The tableau algorithm allows for constructing a canonical interpretation.

Definition 51. Baader and Sattler (2001). *The* canonical interpretation $\mathcal{I}_\mathcal{A}$ *of $\mathcal{A}$ is defined as follows:*

- *the domain $\mathfrak{A}^{\mathcal{I}_\mathcal{A}}$ consists of the individual names occurring in $\mathcal{A}$*
- *for all concept names P we define $P^{\mathcal{I}_\mathcal{A}} := \{x \mid x : P \in \mathcal{A}\}$*
- *for all role names R we define $R^{\mathcal{I}_\mathcal{A}} := \{(x,y) \mid (x,y) : R \in \mathcal{A}\}$*

This algorithm is now extended by including the Abox closure rule of Definition 36. Therefore, first the application of the Abox closure rule is considered independently of the tableau. Observe that for a given interpretation $\mathcal{I}$ all concepts in $\mathcal{I}$ (thus $DC^\mathcal{I}_\mathcal{O}$) need to fulfill HP. Therefore, it is not sufficient to test only for concepts occurring in $\mathcal{A}$. Especially, it is necessary to construct a Helly-companion for $\mathcal{O}$ that fulfills HP. Special problems arise due to the consideration of roles: If $(a,a) : R \in \mathcal{A}$, then it is necessary to test HP for each $\exists R.\top, \exists R.\exists R.\top$ etc. To circumvent this problem, a graph-based view is applied to the concepts.

Definition 52. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a $\mathcal{ELHO}(\circ)^\bot$-ontology. Let a be an individual in $\mathcal{A}$. For a, a directed graph $G_a = (V_a, E_a)$ representing its relations is modeled as follows:*

$$V_a^0 = \{a_0\}$$
$$E_a^0 = \{(a_0, x) \mid (a,x) : R \in \mathcal{A} \text{ for } R \in \mathbf{R}\}$$

If V_a^{i-1}, E_a^{i-1} are given, construct V_a^i, E_a^i as follows:

$$V_a^i = \bigcup\{y \mid \exists x : (x,y) \in E_a^{i-1} \text{ and } y \notin V_a^j \text{ for some } 0 \leq j < i \text{ and } R \in \mathbf{R}\}$$
$$E_a^i = \bigcup\{(x,y) \mid (x,y) : R \in \mathcal{A} \text{ and } x \in V_a^i\}$$

The procedure stops when $V_a^i = \varnothing$ thus no new nodes are added. Then $V_a = \bigcup_i V_a^i$, $E_a = \bigcup_i E_a^i$ This construction terminates, as $\mathbf{I}$ and $\mathcal{A}$ are finite and it is checked for duplicates.

Note that such a graph can be directly translated into an assertion by considering the paths in the graph. For example let there be a $v \in V_a$ with $(a,v) \in E_a$ and $C \in \mathcal{C}(v)$ for $\mathcal{C}(v) = \{A \mid v : A \in \mathcal{A}\} \cup \top$ (thus $\mathcal{C}(v)$ represents the concepts asserted to v in the Abox). Then, $\mathcal{O} \models a : \exists R.C$.

With the help of this graph, the Abox closure rule can be applied.

Definition 53. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^\perp$-ontology. The Abox closure rule as defined in Definition 36 is applied to ontology $\mathcal{O}$ for all individual names a, b, c occurring in $\mathcal{A}$ as follows:*

First, create for each of a, b, c the relation graph G_a, G_b, G_c as defined in Definition 52. For each $v, w \in \{a, b, c\}$ now the combined relation graph $G_{v \cap w}$ is defined, thus the graph representing only concept representations asserted to both v and w. Let $G_{v\cap w} = (V_{v\cap w}, E_{v\cap w})$.

$$V^0_{v\cap w} = \{\{v_0, w_0\}\}$$
$$E^0_{v\cap w} = \{(\{v_0, w_0\}, \{x, y\}) \mid (v, x) : R, (w, y) : R \in \mathcal{A} \text{ for } R \in \mathbf{R}\}$$

The rest is defined analogously to Definition 52. Based on the same argument as above, $G_{v\cap w}$ is finite.

If for some v, w $E_{v\cap w} = \varnothing$ and $\mathcal{C}(v) \cap \mathcal{C}(w) = \top$, then there is no non-trivial concept description A with $v : A, w : A \in \mathcal{A}$ and no $R \in \mathbf{R}$ with $(v, x) : R, (w, y) : R \in \mathcal{A}$ for some individuals x, y. Thus, the closure rule is trivially fulfilled in this case.

Therefore, assume that for each of $v, w \in \{a, b, c\}$, $E_{v\cap w} \neq \varnothing$ or $\mathcal{C}(v) \cap \mathcal{C}(w) \supset \{\top\}$. Then, the premise of the Abox closure rule is non-trivially fulfilled and the conclusion needs to be tested and possibly a new individual needs to be added.

Thus, test whether there is an individual d such that:

- *$(\mathcal{C}(a) \cap \mathcal{C}(b)) \cup (\mathcal{C}(b) \cap \mathcal{C}(c)) \cup (\mathcal{C}(a) \cap \mathcal{C}(c)) \subseteq \mathcal{C}(d)$ (thus d shares all concepts that a, b and b, c and a, c respectively share) and*
- *for each $v, w \in \{a, b, c\}$: Iteratively test for each edge and each node in $G_{v\cap w}$ whether an individual mimicking the modeled relation can be found. Therefore, construct the graph G_d and find for each edge in G_d matching edges in $G_{v\cap w}$. Start with edges $(d, z) \in E_d$. (d, z) can be matched with $(\{v, w\}, \{x, y\}) \in E_{v\cap w}$, if $\mathcal{R}(v, x) \cap \mathcal{R}(w, y) \subseteq \mathcal{R}(d, z)$ where $\mathcal{R}(a, b) = \{R \mid R \in \mathbf{R} \text{ and } R(a, b) \in \mathcal{A}\}$ and if $\mathcal{C}(x) \cap \mathcal{C}(y) \subseteq \mathcal{C}(z)$. Thus, (d, z) needs to have all relations that (v, x) and (w, y) share and z needs to have all concepts that x and y share. This*

procedure is continued stepwise (thus, e.g., for edges $(z,u) \in E_d$ as match with all $(\{x,y\},\{s,t\}) \in E_{v\cap w}$ for which in the last step a match has been found). If for all edges in $E_{v\cap w}$ a match is found, then d has also a witness for each concept that v and w share.

If this is the case, then the Abox closure rule does not need to be applied. Otherwise, new individuals need to be defined Add a new individual d_i for $0 < i \leq |V_{a\cap b}| + |V_{b\cap c}| + |V_{a\cap c}| - 3$ for each node in $G_{a\cap b}, G_{b\cap c}$ and $G_{a\cap c}$ except for the root nodes. For the root nodes add one individual d_0. For each d_i add the corresponding concepts and roles to the Abox. For a d_i corresponding to node $\{x,y\}$ of the graph, let $\mathcal{C}(d_i) = \mathcal{C}(x) \cap \mathcal{C}(y)$ and add $C(d_i)$ to $\mathcal{A}$ for all $C \in \mathcal{C}(d_i)$. For d_0 add $\bigcup_{A\in(\mathcal{C}(a)\cap\mathcal{C}(b))\cup(\mathcal{C}(b)\cap\mathcal{C}(b))\cup(\mathcal{C}(a)\cap\mathcal{C}(c))} d_0 : A$ to $\mathcal{A}$. For $i,j \geq 0$, add $(d_i,d_j) : R$ to $\mathcal{A}$ if the corresponding $\{x,y\},\{t,u\}$ have $(x,t) : R$ and $(y,u) : R$ in $\mathcal{A}$. Now, as new individuals have been added, the process is started again. This is repeated until nothing is added.

Note that this construction is highly inefficient, as many individuals are added unnecessarily. These are, however, only finitely many and therefore not problematic as will be proven later on. This construction does not only apply the Abox closure rule but additionally considers the relational part. The relational part is, however, necessary to consider to gain an interpretation fulfilling HP. The process of Definition 53 is applied to all individuals in $\mathcal{A}$, thus also to the individuals newly added during the process. Therefore, it needs to be proven that this process terminates.

Corollary 54. *The application of the (extended) Abox closure rule as stated in Definition 53 to an Abox $\mathcal{A}$ of an $\mathcal{ELHO}(\circ)^{\perp}$-ontology $\mathcal{O}$ terminates.*

Proof. 1. First, consider the case where $E_{a\cap b} = \emptyset, E_{b\cap c} = \emptyset, E_{a\cap c} = \emptyset$ for all individuals a,b,c considered during the process. In this case only concepts are considered. Assume n is the number of different concepts occurring in $\mathcal{A}$, as $\mathcal{A}$ is finite by definition. Then, there are worst-case 2^n individuals to be added, as worst-case only 2^n different concept combinations are possible. If worst-case 2^n individuals have been added, then for each case when the premise of the Abox closure rule is fulfilled, there is an individual fulfilling the conclusion.

2. Now, consider the case where at least one of $E_{a\cap b} \neq \emptyset, E_{b\cap c} \neq \emptyset, E_{a\cap c} \neq \emptyset$ for some individuals a,b,c.

   Observe the following facts: for each group of newly added individuals d_i, all of these elements except d_0 are not able to introduce new situations on which the Abox closure needs to be applied. This is the case, as each d_i mimics the concepts represented in some subgraph of $G_{a\cap b}, G_{b\cap c}$ or $G_{a\cap c}$.

   Thus, only d_0 needs to be considered. Note that d_0 does not have an incoming edge. Therefore, the same argument as for the case solely based on concepts can be used: There are finitely many graphs G_v, one for each individual in $\mathcal{A}$. There are only finitely many variants to combine these graphs. Therefore, only finitely many individuals can be added.

Until now, the process applies the Abox closure only to a given Abox.

As HP needs to be valid not only for the Abox but for an interpretation, it is additionally necessary to consider HP for concepts not present in the Abox but present in each possible interpretation satisfying an ontology. Thus, the Helly companion of an ontology needs to be defined (see Definition 39).

Therefore, in the following, the tableau algorithm is combined with the Abox closure rule to define a Helly-companion. Note that this is not the only possible Helly-companion. After that, it is shown that this approach terminates and leads to a model that fulfills HP.

Definition 55. *In the following, a tableau algorithm incorporating the Abox closure rule is defined. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology. Repeat the following two steps until the application of the Abox closure rule does not introduce any new individuals.*

1. *Apply the tableau algorithm as defined in Definition 50 on $\mathcal{O}$ until a blocking condition is reached. Then, materialize the blocking, thus add for each a blocked by b, for a the successors of b.*
2. *Apply the Abox closure as defined in Definition 53.*

Now, it is shown that this adapted tableau algorithm terminates.

Corollary 56. *The algorithm as defined in Definition 55 terminates.*

Proof. Each application of the tableau algorithm terminates and each application of the Abox closure terminates. Therefore, it remains to show that the combination of both terminates. It is again sufficient to consider the newly added d_0. All other added individuals represent concept descriptions that already exist in $\mathcal{A}$. If these would not be complete regarding the application of the transformation rules, then also the concepts used to construct the individuals would not have been complete.

However, for a new individual d_0 it can be the case that for $d_0 : A \sqcap B \sqcap C$ there is a Tbox axiom $A \sqcap B \sqcap C \sqsubseteq D$ (in its corresponding normal form). However, the Tbox is by definition finite. Therefore, there are only finitely many axioms of this type. Therefore, the algorithm terminates.

Now, it can be shown that this construction leads to a Helly-companion.

Corollary 57. *Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology. Let $\mathcal{O}' = (\mathcal{T}', \mathcal{A}')$ be the result of the application of the modified tableau as defined in Definition 55. Then, $\mathcal{O}'$ is a Helly-companion.*

Proof. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology. Let $\mathcal{O}' = (\mathcal{T}', \mathcal{A}')$ be the result of the application of the modified tableau as defined in Definition 55. It is shown that $\mathcal{O}'$ is a Helly-companion.

1. $\mathcal{O} \subseteq \mathcal{O}'$, $\mathbf{I}(\mathcal{O}') \supseteq \mathbf{I}(\mathcal{O}), \mathbf{C}(\mathcal{O}') = \mathbf{C}(\mathcal{O}), \mathbf{R}(\mathcal{O}') = \mathbf{R}(\mathcal{O})$ follows trivially based on the definition.

2. Every necessarily non-empty concept is witnessed due to the transformation rules of the standard tableau and due to the fact that the tableau is sound and complete.

3. The process of Definition 55 terminates and directly models the Abox closure rule. Additionally, it due to the graph-based view, all complex concept descriptions including relations are considered and checked for HP. This means that $\mathcal{A}'$ is Helly-closed.

Corollary 58. *Let $\mathcal{O}$ be a Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology and let $\mathcal{O}'$ be the result of the application of Definition 55. The interpretation as defined in Definition 51 is a model of $\mathcal{O}$.*

Proof. Let $\mathcal{O}$ be a Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology and let $\mathcal{O}'$ be the result of the application of Definition 55. Let $\mathcal{I}_{\mathcal{A}}$ be the canonic interpretation of $\mathcal{O}'$. Assume by contradiction that $\mathcal{I}_{\mathcal{A}}$ is not a model of $\mathcal{O}$. Therefore, there must be a $d^{\mathcal{I}_{\mathcal{A}}} \in (A \sqcap B \sqcap C)^{\mathcal{I}_{\mathcal{A}}}$ with $\mathcal{O} \models A \sqcap B \sqcap C = \bot$. When applying the standard tableau algorithm to a satisfiable ontology, a concept description A is only added to $\mathcal{A}'$ if $\mathcal{O} \cup \{A \sqsubseteq \bot\}$ is inconsistent. Therefore, for all of these added concepts and same for all added roles, the Abox closure rule needs to be applied and thus HP needs to be tested. By definition, for fulfilling HP, always the least specific concept is added. Therefore such a $d^{\mathcal{I}_{\mathcal{A}}}$ would directly interfere with HP, thus, the ontology can't be Helly-satisfiable, a contradiction.

Corollary 59. *Let $\mathcal{O}$ be a Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology and let $\mathcal{O}'$ be the result of the application of Definition 55. The canonic interpretation $\mathcal{I}_{\mathcal{A}}$ of $\mathcal{O}'$ as defined in Definition 51 fulfills HP.*

Proof. Let $\mathcal{O}$ be a Helly-satisfiable $\mathcal{ELHO}(\circ)^{\bot}$-ontology and let $\mathcal{O}'$ be the result of the application of Definition 55. $\mathcal{O}'$ is a Helly-companion of $\mathcal{O}$ (see Corollary 57). Therefore, it is Helly-closed. The canonic model of $\mathcal{O}'$ can be defined without the need to infer new conceptual information and without adding new relations or individuals. Therefore, the canonic model of $\mathcal{O}'$ is also Helly-closed and thus fulfills HP.

Now, Proposition 42 can be proven. Thus, if an ontology is Helly-satisfiable, then there is a construction procedure for a finite model that satisfies HP.

Proof of Proposition 42. Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a Helly-satisfiable ontology in normal form.

Let $\mathcal{I}_{\mathcal{A}}$ be the canonic model of $\mathcal{O}'$ constructed based on $\mathcal{O}$ with the adapted tableau as defined in Definition 55.

$\mathcal{I}_{\mathcal{A}}$ is constructable in finite time, as proven in Corollary 56. It is a model of $\mathcal{O}$ due to Corollary 58 and it fulfills HP due to Corollary 59. Thus, a finite model of $\mathcal{O}$ fulfilling HP can be found in finite time.

With this result, the proof of Proposition 41 follows.

Proof of Proposition 41. → Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a Helly-satisfiable $\mathcal{ELHO}(\circ)^{\perp}$-ontology. Then apply the adapted tableau algorithm as defined in Definition 55 to get $\mathcal{O}'$ that is a Helly-companion as proven in Corollary 57. As proven in Corollary 58, a model can be defined, therefore, $\mathcal{O}'$ is satisfiable.

← Let $\mathcal{O}$ not be Helly-satisfiable. Consider an arbitrary Helly-companion $\mathcal{O}'$ of $\mathcal{O}$. It is shown that the resulting ontology $\mathcal{O}'$ is not satisfiable. As $\mathcal{O}$ is not Helly-satisfiable, there is in each model $\mathcal{I}$ of $\mathcal{O}$, $a^{\mathcal{I}}, b^{\mathcal{I}}, c^{\mathcal{I}} \in \Delta$ with $a^{\mathcal{I}} \in (A \sqcap B)^{\mathcal{I}}$, $b^{\mathcal{I}} \in (B \sqcap C)^{\mathcal{I}}$, $c^{\mathcal{I}} \in (A \sqcap C)^{\mathcal{I}}$ but $A^{\mathcal{I}} \cap B^{\mathcal{I}} \cap C^{\mathcal{I}} = \varnothing$ for some concept descriptions A, B, C.

Assume for the sake of contradiction that the Helly-companion is satisfiable. Each Helly-companion contains witnesses for each concept description that is known to be non-empty. Therefore, the canonic interpretation $\mathcal{I}_{\mathcal{A}}$ of $\mathcal{O}'$ would be a model of $\mathcal{O}'$ (as $\mathcal{O}'$ is assumed to be satisfiable). But this model fulfills by definition of the Helly-companion HP and therefore, $\mathcal{O}$ would be Helly-satisfiable. A contradiction to the assumption.