Proceedings of the 23rd International Conference on Principles of Knowledge Representation and Reasoning
KR meets Machine Learning and Explanation

BoxLitE: A Faithful Knowledge Base Embedding Based on Convex Optimization

Bruno F. Lourenço[1] , **Hesham Morgan**[2] , **Ana Ozaki**[3] , **Aleksandar Pavlović**[4] , **Emanuel Sallinger**[2]
[1]Institute of Statistical Mathematics, Japan
[2]TU Wien, Austria
[3]University of Oslo, Norway
[4]University of Applied Sciences Campus Vienna, Austria
bruno@ism.ac.jp, hesham.morgan@tuwien.ac.at, anaoz@uio.no, aleksandar.pavlovic@hcw.ac.at, sallinger@dbai.tuwien.ac.at

Abstract

Knowledge base (KB) embeddings aim at combining the capability of classical knowledge graph embeddings to generalize the information present in facts, the ABox, with conceptual knowledge represented in an ontology language, the TBox. Several authors have recently explored the idea of mapping concepts to *convex regions* in a vector space. This is useful to represent hierarchies, typically present in TBoxes, since more general concepts can be mapped to larger regions, containing those regions associated with more specific concepts. However, the power of convexity is rarely leveraged during the actual learning tasks. Here, we introduce BoxLitE, a KB embedding model for DL-Lite$^{\mathcal{H}}$ that allows for convex optimization. We show that for any satisfiable DL-Lite$^{\mathcal{H}}$ KB, there is a BoxLitE embedding that is a weakly faithful model. As a proof of concept, we show how to formulate the KB embedding task as a convex optimization problem and how to obtain embeddings with such desirable faithfulness property.

1 Introduction

Knowledge base (KB) embeddings combine the capability of knowledge graph embeddings to perform inductive reasoning for link prediction with deductive reasoning, using logic expressions present in an ontology (Bourgaux et al., 2024). Several authors have recently explored the idea of mapping concepts in a KB to regions in a vector space (Gutiérrez-Basulto and Schockaert, 2018; Pavlović and Sallinger, 2023a; Pavlović, Sallinger, and Schockaert, 2025; Morgan et al., 2026). Region-based embeddings are important for KBs as they can naturally represent hierarchies: more general concepts can be mapped to larger regions, containing those regions associated with more specific concepts.

Although concepts are usually mapped to *convex regions* in KBEs, e.g., balls, boxes, and cones (Kulmanov et al., 2019; Abboud et al., 2020; Xiong et al., 2022; Pavlović and Sallinger, 2023b; Lütfü Özçep, Leemhuis, and Wolter, 2020), the power of convexity is rarely leveraged during the actual learning tasks. While being convex is not synonymous with "easy", convexity brings a number of theoretical and practical benefits, in particular: all local minima must be global. Under convexity, there are a number of efficient algorithms for many classes of convex problems such as linear programs and second-order cone programs (SOCPs) (Nesterov and Nemirovskii, 1994; Ben-Tal and Nemirovski, 2001).

For KBs expressed in description logic, most of the literature work on region-based embeddings focuses on the $\mathcal{EL}$ ontology language (Yang, Chen, and Sattler, 2025; Jackermeier, Chen, and Horrocks, 2024; Lacerda, Ozaki, and Guimarães, 2024a; Xiong et al., 2022; Kulmanov et al., 2019; Mondal, Bhatia, and Mutharaju, 2021; Peng et al., 2022; Lacerda, Ozaki, and Guimarães, 2024b), while some consider $\mathcal{ALC}$ (Lütfü Özçep, Leemhuis, and Wolter, 2020; Leemhuis, Özçep, and Wolter, 2022). However, in practice, ontologies present in large-scale KBs tend to use features in the DL-Lite$^{\mathcal{H}}$ ontology language (Artale et al., 2009), due to its simple but versatile expressivity, featuring low computational complexity (Artale et al., 2009). Despite its broad use, only a few works investigate logics in the DL-Lite family in a KB embedding context (Li et al., 2022; Imenes, Guimarães, and Ozaki, 2023; Bourgaux, Ozaki, and Pan, 2021) and none provide a region-based embedding implementation.

In this paper, we introduce BoxLitE, a KB embedding model for DL-Lite$^{\mathcal{H}}$ KBs that allows for convex optimization. In particular, we study the notion of *weakly faithful models* (Lütfü Özçep, Leemhuis, and Wolter, 2020; Bourgaux et al., 2024) in the optimization task. This notion states that axioms that hold in the embedding are *consistent* with the KB and *entailments of the KB are satisfied* in the embedding. In detail,

- we select DL-Lite$^{\mathcal{H}}$ an **ontology language** that allows to exploit the **advantages of convex optimization**;
- we **design, implement, and evaluate** BoxLitE, a KB embedding approach that allows for convex optimization;
- we prove that every satisfiable DL-Lite$^{\mathcal{H}}$ has a BoxLitE embedding that is a **weakly faithful model**;
- we introduce a novel and efficient form of **negative sampling based on existential concepts**;
- in contrast to commonly used unconstrained optimization approaches, our BoxLitE approach can enforce the **TBox axioms using convex constraints** while leaving the ABox terms in the objective function.

Main Contribution. Our main contribution lies on theoretical grounds, establishing the existence of a faithful model for DL-Lite$^{\mathcal{H}}$ and the proposal of a novel approach that computes

embeddings for DL-Lite$^{\mathcal{H}}$ KBs. BoxLitE's implementation and the empirical results serve as a proof of concept that our theoretical results translate into practical settings. One of the motivations for this work is to try to solve link prediction using theoretically sounds techniques from a convex optimization perspective. We would like to check how much performance can we get from a purely convex approach. In this way, our work stands in contrast to previous approaches that relied on nonconvexity, yielding more complicated optimization problems.

There are three primary sources of nonconvexity in contemporary KB embedding methods (see more details in Section 7): the way *negative sampling* is used by most works (Bordes et al., 2013; Sun et al., 2019; Yang et al., 2015; Trouillon et al., 2016; Kazemi and Poole, 2018; Balazevic, Allen, and Hospedales, 2019; Xiong et al., 2022; Abboud et al., 2020; Pavlović and Sallinger, 2024b); the *design of loss terms* involving operations that do not preserve convexity; and the choice of the *ontology language*, in particular, languages that allow conjunction on the left-hand side of concept inclusions, such as $\mathcal{EL}$ and $\mathcal{ALC}$, are likely to lead to nonconvexity. This motivated us to consider DL-Lite$^{\mathcal{H}}$ (Artale et al., 2009), which is a simple but practically useful ontology language without conjunction on the left-hand side.

On the algorithmic side, ADAM (Kingma and Ba, 2015), the method of choice of several works, is often used in theoretically unsound ways. For example, ADAM requires the objective function to be differentiable; however, it is not uncommon to see this requirement being ignored. For instance, in (Xiong et al., 2022, Sections 4.3, 4.5) and (Abboud et al., 2020, Section 4), the authors describe nondifferentiable objective functions that are optimized via ADAM[1]. The rationale for that is not explained in the papers, but a reasonable guess seems to be that these functions are typically differentiable almost everywhere (in a measure theory sense), so there may be an underlying belief that iterates are unlikely to reach a point of nondifferentiability and the algorithm may still work in practice. Besides, widely used tools such as PyTorch may attempt to differentiate nondifferentiable functions for the user by, e.g., selecting subgradients/supgradients when available (PyTorch, 2026). Unfortunately, there are well-known examples in the optimization literature showing that when a gradient method is applied to a nondifferentiable function, it may fail to find an optimal solution, even if the function is differentiable at the points generated by the method, e.g., see (Beck, 2017, Section 8.1.2). In the optimization community, some works explore the convergence properties of SGD methods when applied to nondifferentiable functions (Bolte and Pauwels, 2020; Davis et al., 2019), but, as far as we know, similar results have not been proven for ADAM.

When performance is good, one may be tempted to overlook these issues, but when it is not, it may be hard to know what is to blame. Is it the method, the parameter choice, the optimization algorithm, or a combination of those? Our proposed approach *does* include nondifferentiable terms as well, but, in contrast, our method of choice for solving the underlying optimization problem is suitable for handling nondifferentiability (Section 7). In this sense, our approach is conceptually sound from an optimization point of view.

Organization. Our paper is organized as follows. Section 2 provides basic definitions. Section 3 defines the semantics of our BoxLitE approach. Section 4 studies BoxLitE's faithfulness properties. Section 5 formulates BoxLitE's convex optimization problem and shows faithfulness properties that are ensured by the problem formulation. Section 6 discusses BoxLitE's proof of concept implementation and experiments. Section 8 provides a discussion on optimization in KB embeddings and concludes our paper. Omitted proofs and additional details are in the arxiv version (Lourenço et al., 2026).

2 Basic Definitions: DL-Lite Ontologies

Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be *finite*, non-empty, mutually disjoint sets of *concept*, *role*, and *individual* names, respectively. We denote by $\mathsf{N_R^-}$ the set $\mathsf{N_R} \cup \{R^- \mid R \in \mathsf{N_R}\}$, by $\mathsf{N_C^\exists}$ the set $\mathsf{N_C} \cup \{\exists R \mid R \in \mathsf{N_R^-}\}$, and by $\mathsf{N_{C\neg}^\exists}$ the set $\mathsf{N_C^\exists} \cup \{\neg E \mid E \in \mathsf{N_C^\exists}\}$. DL-Lite$^{\mathcal{H}}$ role and concept inclusions are of the form $S \sqsubseteq T$ and $B \sqsubseteq C$, resp., where $S, T \in \mathsf{N_R^-}$ are roles[2] and B, C are concepts built as follows: $S ::= R \mid R^-$, $B ::= A \mid \exists S$, $C ::= B \mid \neg B$, with $R \in \mathsf{N_R}$ and $A \in \mathsf{N_C}$. A DL-Lite$^{\mathcal{H}}$ *TBox* (we may also use the less technical term *ontology*) is a (finite) set of DL-Lite$^{\mathcal{H}}$ concept and role inclusions. *Assertions* are of the form $D(a)$ (called *concept assertions*) or $R(a, b)$ (called *role assertions*), where $D \in \mathsf{N_C^\exists}$, $R \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. A DL-Lite$^{\mathcal{H}}$ *knowledge base* (KB) is a pair $(\mathcal{T}, \mathcal{A})$ where $\mathcal{T}$ is a DL-Lite$^{\mathcal{H}}$ TBox and $\mathcal{A}$ is a (finite) set of assertions, called *ABox*. Following the KBE literature (e.g. (Xiong et al., 2022)) and to simplify our presentation, we assume that DL-Lite$^{\mathcal{H}}$ TBoxes are in *named form*, meaning that they contain only concept inclusions where one of the concepts is a concept name. The semantics is given as usual by interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. We call a DL-Lite$^{\mathcal{H}}$ *axiom* an expression that is a role inclusion (RI), a concept inclusion (CI), or an assertion. We write $\mathcal{I} \models \alpha$ if $\mathcal{I}$ satisfies an axiom α. An interpretation $\mathcal{I}$ satisfies a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, written $\mathcal{I} \models \mathcal{K}$, if it satisfies the axioms in $\mathcal{T}$ and $\mathcal{A}$. We say that $\mathcal{K}$ is *satisfiable* if such an interpretation $\mathcal{I}$ exists. Also, $\mathcal{K}$ *entails* an axiom α, written $\mathcal{K} \models \alpha$, iff, for all interpretations $\mathcal{I}$, if $\mathcal{I} \models \mathcal{K}$ then $\mathcal{I} \models \alpha$. We say that an axiom α is *consistent* with a KB $\mathcal{K}$ if there is an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{K} \cup \{\alpha\}$.

Canonical Model. Our construction is based on previous definitions of canonical models for DL-Lite$^{\mathcal{H}}$ (e.g., (Kontchakov et al., 2010)). Though here, we ensure that *inclusions* hold in the model *only if* they are entailed by the ontology. Before we provide the definition of the canonical model, we introduce the following notions. Assume $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is a satisfiable

[1] In (Xiong et al., 2022), for example, loss terms for concepts assertions are expressed using the 2-norm and the regularization term considered therein use a maximum of functions. Both correspond to terms that are not differentiable in general, e.g., the 2-norm function is not differentiable at the origin, which can be checked directly by the definition of Fréchet differentiability.

[2] A *role* is a role name or the inverse of a role name.

DL-Lite$^{\mathcal{H}}$ KB. We say that a concept C is *satisfiable w.r.t.* $\mathcal{K}$ if there is an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{K}$ and $C^{\mathcal{I}} \neq \emptyset$. We write concepts of the form $B \sqcap C$ (with the *conjunction* operator) only to falsify concept inclusions of the form $B \sqsubseteq \neg C$. In an interpretation $\mathcal{I}$, the meaning of $(B \sqcap C)^{\mathcal{I}}$ is $B^{\mathcal{I}} \cap C^{\mathcal{I}}$. Let $\mathsf{N}_{\mathsf{C}\sqcap}^{\exists}$ be the set $\mathsf{N}_{\mathsf{C}}^{\exists} \cup \{D \sqcap E \mid D, E \in \mathsf{N}_{\mathsf{C}}^{\exists}\}$. Assume $\Delta_{\mathcal{K}} := \{c_D \mid D \in \mathsf{N}_{\mathsf{C}\sqcap}^{\exists},\ D \text{ is satisfiable w.r.t. } \mathcal{K}\}$ is disjoint from N_{I}. Given a role S, we write $\overline{S}$ as the result of switching between a role name and its inverse: $\overline{S} = S^-$ if $S \in \mathsf{N}_{\mathsf{R}}$ and $\overline{S} = R$ if $S = R^-$ with $R \in \mathsf{N}_{\mathsf{R}}$.

Definition 1 (Canonical Model)**.** *The canonical model $\mathcal{I}_{\mathcal{K}}$ for a satisfiable DL-Lite$^{\mathcal{H}}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is:*

- $\Delta^{\mathcal{I}_{\mathcal{K}}} := \mathsf{N}_{\mathsf{I}} \cup \Delta_{\mathcal{K}}, \qquad a^{\mathcal{I}_{\mathcal{K}}} := a$, *for all* $a \in \mathsf{N}_{\mathsf{I}}$,
- $A^{\mathcal{I}_{\mathcal{K}}} := \{a \in \mathsf{N}_{\mathsf{I}} \mid \mathcal{K} \models A(a)\} \cup \{c_D \in \Delta_{\mathcal{K}} \mid \mathcal{K} \models D \sqsubseteq A\}$ *for all* $A \in \mathsf{N}_{\mathsf{C}}$,
- $R^{\mathcal{I}_{\mathcal{K}}} := \{(a,b) \in \mathsf{N}_{\mathsf{I}} \times \mathsf{N}_{\mathsf{I}} \mid \mathcal{K} \models R(a,b)\} \cup \{(a, c_{\exists S}) \in \mathsf{N}_{\mathsf{I}} \times \Delta_{\mathcal{K}} \mid \mathcal{K} \models \exists \overline{S}(a),\ \mathcal{K} \models \overline{S} \sqsubseteq R\} \cup \{(c_{\exists S}, a) \in \Delta_{\mathcal{K}} \times \mathsf{N}_{\mathsf{I}} \mid \mathcal{K} \models \exists \overline{S}(a),\ \mathcal{K} \models S \sqsubseteq R\} \cup \{(c_{\exists S}, c_{\exists \overline{S}}) \in \Delta_{\mathcal{K}} \times \Delta_{\mathcal{K}} \mid \mathcal{K} \models S \sqsubseteq R\} \cup \{(c_D, c_{\exists S}) \in \Delta_{\mathcal{K}} \times \Delta_{\mathcal{K}} | \mathcal{K} \models D \sqsubseteq \exists \overline{S},\ \mathcal{K} \models \overline{S} \sqsubseteq R\} \cup \{(c_{\exists S}, c_D) \in \Delta_{\mathcal{K}} \times \Delta_{\mathcal{K}} \mid \mathcal{K} \models D \sqsubseteq \exists \overline{S},\ \mathcal{K} \models S \sqsubseteq R\}$, *for all* $R \in \mathsf{N}_{\mathsf{R}}$.

Theorem 1. *Let $\mathcal{K}$ be a satisfiable DL-Lite$^{\mathcal{H}}$ KB and let $\mathcal{I}_{\mathcal{K}}$ be the canonical model of $\mathcal{K}$. Then, for all DL-Lite$^{\mathcal{H}}$ axioms α, we have that $\mathcal{I}_{\mathcal{K}} \models \alpha$ iff $\mathcal{K} \models \alpha$.*

3 BoxLitE Semantics

Here we introduce a semantics for DL-Lite$^{\mathcal{H}}$ inspired by box-based geometric models (Abboud et al., 2020) and suitable for defining a KB embedding model that can be optimized using convex optimization. We define a geometric model that uses axis-aligned hyper-rectangles, called *boxes*.

Let ϵ and s_Ω be fixed but arbitrary [3] positive constants with $\epsilon \leq s_\Omega$. Let $\boldsymbol{\epsilon}$ and $\mathbf{s_\Omega}$ denote the d-dimensional vectors whose value in each dimension is equal to ϵ and s_Ω respectively. Henceforth, we denote elementwise comparison operators with $\leq_d$ and $\geq_d$. Using $\boldsymbol{\epsilon}$ and $\mathbf{s_\Omega}$, we define an axis-aligned hyper-rectangle, called the *universe box*:

$$\Omega = \{\mathbf{x} \in \mathbb{R}^d \mid -\mathbf{s_\Omega} \leq_d \mathbf{x} \leq_d \mathbf{s_\Omega}\}$$

in our d-dimensional Euclidean space, where $d > 0$. The universe box is useful to establish basic properties of boxes (Theorem 2) that are needed in our faithfulness proofs.

Definition 2. *A* box *is an element of the set* Box*, defined as:*

$$\mathsf{Box} := \{\{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{L} + \boldsymbol{\epsilon} \leq_d \mathbf{x} \leq_d \mathbf{U} - \boldsymbol{\epsilon}\} \mid \mathbf{0} \leq_d (\mathbf{U} - \mathbf{L}) \leq_d 2\mathbf{s_\Omega},\ \mathbf{L}, \mathbf{U} \in \mathbb{R}^d\}.$$

[3] In our work and, in particular, in the implementation, we take $0 < \epsilon < \epsilon_{max}$, where $\epsilon_{max} = 0.5$ and $\epsilon_{max} \leq s_\Omega/8$. Making ϵ and s_Ω learnable parameters would not enhance the representation capabilities of our model but allow for infinitely many equivalent solutions of our learned embeddings which only differ in scale.

For a non-empty box, the lower *and* upper *bounds are denoted* $\mathbf{L}$ *and* $\mathbf{U}$ *(resp.) of* $\mathbf{X}$*. If* $\mathbf{X}$ *is the empty box then we set* $\mathbf{L} := \mathbf{U} := \mathbf{0}$*. We denote with* $\mathbf{x}[i]$ *the* i*-th dimension of a vector* $\mathbf{x}$*. If* $\mathbf{X}$ *is a box with lower and upper bounds* $\mathbf{L}$ *and* $\mathbf{U}$*, resp., then* $\mathbf{X}[i]$ *is the pair* $(\mathbf{L}[i], \mathbf{U}[i])$*.*

Box Definition Intuition. The definition of Box allows boxes to have a width of $\leq 2\mathbf{s_\Omega}$ and to span outside of the universe Ω, while we require vectors associated with individual names to be within Ω (as we will see in Definition 3). This design is motivated (in the spirit of Abboud et al. (2020)) by associating any individual name with a position and a bump vector, where the embeddings of a pair of individual names (a, b) is retrieved by translating ("*bumping*") the position of a with the bump of b and vice versa. Thus, if both the position and bump vectors are within the universe box, i.e., bounded by $\mathbf{s_\Omega}$, the embeddings of any individual pair is bounded by $2\mathbf{s_\Omega}$. As we associate any concept and role name with a set of boxes that shall contain the embeddings of individual pairs, it is sufficient that box widths are bounded by $2\mathbf{s_\Omega}$. As for $\boldsymbol{\epsilon}$'s intuition, most convex optimization tools do not allow strict inequalities or handle them less efficiently. Yet, we need them to define empty boxes for concepts/roles that are unsatisfiable. So we employ in Definition 2 *nonstrict inequalities* and add a small positive $\boldsymbol{\epsilon}$ to emulate strict ones.

Definition 3. *A* box interpretation η *is a function that maps:*

- *each individual name* $a \in \mathsf{N}_{\mathsf{I}}$ *to two vectors* $\eta(a) = (\mathsf{pos}(a), \mathsf{bump}(a))$*, namely, a position* $\mathsf{pos}(e) \in \Omega$ *and a bump* $\mathsf{bump}(e) \in \Omega$*;*
- *each concept name* $A \in \mathsf{N}_{\mathsf{C}}$ *to a box* $\eta(A) \in \mathsf{Box}$*;*
- *each role name* $R \in \mathsf{N}_{\mathsf{R}}$ *to three boxes* $\eta(R) = (\mathsf{Head}(R), \mathsf{Tail}(R), \mathsf{Bump}(R))$*, which we call* R*'s head* $\mathsf{Head}(R)$*, tail* $\mathsf{Tail}(R)$ *and bump box* $\mathsf{Bump}(R)$*.*

We extend the mapping function η *to arbitrary DL-Lite$^{\mathcal{H}}$ concept and role expressions as follows:*

$$\eta(\neg C) := \overline{\eta(C)}, \quad \eta(R^-) := (\mathsf{Tail}(R), \mathsf{Head}(R), \mathsf{Bump}(R)),$$
$$\eta(\exists R) := \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{L}_{\mathbf{R}}^{\mathbf{H}} - \mathbf{U}_{\mathbf{R}}^{\mathbf{B}} + \boldsymbol{\epsilon} \leq_d \mathbf{x} \leq_d \mathbf{U}_{\mathbf{R}}^{\mathbf{H}} - \mathbf{L}_{\mathbf{R}}^{\mathbf{B}} - \boldsymbol{\epsilon}\}$$

where $\mathbf{L}_{\mathbf{R}}^{X}$*,* $\mathbf{U}_{\mathbf{R}}^{X}$ *with* $X \in \{\mathbf{H}, \mathbf{T}, \mathbf{B}\}$ *are the lower and upper bounds of* $\mathsf{Head}(R)$*,* $\mathsf{Tail}(R)$*, and* $\mathsf{Bump}(R)$*; and where* $\overline{\eta(C)}$ *represents* $\eta(C)$*'s complement box. Furthermore, for any* $C \in \mathsf{N}_{\mathsf{C}}^{\exists}$*, the complement box* $\overline{\eta(C)}$ *is defined as follows:*

$$\overline{\eta(C)} := \{\mathbf{x} \in \mathbb{R}^d \mid (-\mathbf{s_\Omega} - \mathbf{L_C} + \boldsymbol{\epsilon}) \leq_d \mathbf{x} \leq_d (\mathbf{s_\Omega} - \mathbf{U_C} - \boldsymbol{\epsilon})\}$$

with $\mathbf{L_C}$ *and* $\mathbf{U_C}$ *being* $\eta(C)$*'s lower and upper bounds.*

Box Interpretation Intuition. At first sight, an intuitive definition for the complement of a box would be the set complement. However, we cannot use this notion as it leads to non-convexity. Thus, a different convex-preserving definition of box complements is required that satisfies important properties of the usual complement, as shown in Theorem 2. Next, the box interpretation of inverse roles $\eta(R^-)$ swaps the head and tail boxes of $\eta(R)$. As for the existential boxes, we define $\eta(\exists R)$ as $\mathsf{Head}(R)$ enlarged by the boundaries of

$\mathsf{Bump}(R)$, to ensure that it contains the embeddings of all subjects of R assertions, following $\exists R$'s semantics.

Theorem 2. *For any box interpretation η:*

i) *for all $C \in \mathsf{N}_\mathsf{C}^\exists$, $\overline{\eta(C)} \in \mathsf{Box}$;*

ii) *for all $C \in \mathsf{N}_\mathsf{C}^\exists$, $\overline{\overline{\eta(C)}} = \eta(C)$;*

iii) *for all $C, D \in \mathsf{N}_\mathsf{C}^\exists$, if $\eta(C) \subseteq \eta(D)$ then $\overline{\eta(D)} \subseteq \overline{\eta(C)}$.*

Box Consistency. A box interpretation is *box consistent* if for all $C \in \mathsf{N}_\mathsf{C}^\exists$ we have that $\eta(C) \cap \overline{\eta(C)} = \emptyset$.

Definition 4. *A box interpretation η satisfies*

- *$R(a, b)$ with $R \in \mathsf{N}_\mathsf{R}$ and $a, b \in \mathsf{N}_\mathsf{I}$ iff*

$$i)\ \mathsf{pos}(a) + \mathsf{bump}(b) \in \mathsf{Head}(R)$$
$$ii)\ \mathsf{pos}(b) + \mathsf{bump}(a) \in \mathsf{Tail}(R)$$
$$iii)\ \mathsf{bump}(a), \mathsf{bump}(b) \in \mathsf{Bump}(R)$$

- *$R \sqsubseteq S$ with $R, S \in \mathsf{N}_\mathsf{R}^-$ iff* $\mathsf{Bump}(R) \subseteq \mathsf{Bump}(S)$
$\mathsf{Head}(R) \subseteq \mathsf{Head}(S)$ $\quad$ $\mathsf{Tail}(R) \subseteq \mathsf{Tail}(S)$

- *$C \sqsubseteq D$, $C \in \mathsf{N}_\mathsf{C}^\exists$ and $D \in \mathsf{N}_{\mathsf{C}\neg}^\exists$ iff $\eta(C) \subseteq \eta(D)$.*

Regarding concept assertions with $C \in \mathsf{N}_\mathsf{C}^\exists$ and $a \in \mathsf{N}_\mathsf{I}$, η satisfies $C(a)$ if $\mathsf{pos}(a) \in \eta(C)$ and η falsifies $C(a)$ if $\mathsf{pos}(a) \in \overline{\eta(C)}$. Otherwise the status of $C(a)$ is unknown[4].

We write $\eta \models \alpha$ to indicate that η satisfies an axiom α, and write $\eta \not\models \alpha$ otherwise. Also, if $\eta \models \alpha$ for every axiom α in a KB $\mathcal{K}$ then we say that η satisfies $\mathcal{K}$ or, equivalently, that η is a model of $\mathcal{K}$, in symbols, $\eta \models \mathcal{K}$.

The most intricate part of Definition 4 corresponds to role assertions, so we explain this in more details. Items $i)$ and $ii)$ are as in Abboud et al. (2020). Item $iii)$ is helpful to embed existential concepts. Recall that $\eta(\exists R)$ is $\mathsf{Head}(R)$ enlarged by the boundaries of $\mathsf{Bump}(R)$. Thus, if $R(a, b)$ is satisfied in a box interpretation, $\mathsf{pos}(a)$ may lie at most the size of $\mathsf{Bump}(R)$ away from $\mathsf{Head}(R)$ (see Figure 1).

With this, we have defined BoxLitE's semantics in terms of box interpretations. In contrast to (Abboud et al., 2020), we (i) define *complement boxes*, which allow the formulation of convex constraints that guarantee the satisfaction of concept disjointness axioms in our embedding solutions and (ii) associate roles with additional *bump* boxes that constrain the maximal bump of an individual that is accepted by a role embedding. The latter is important to distinguish between axioms of the form $R \sqsubseteq S$ and $\{\exists R \sqsubseteq \exists S, \exists R^- \sqsubseteq \exists S^-\}$. In the next section, we define the notion of a faithful model and analyse which faithfulness properties for DL-Lite$^\mathcal{H}$ can be satisfied by BoxLitE's box interpretations.

[4]Geometric models for languages that have negation normally need to allow for the 'unknown' truth status. This happens in the Cone Semantics (Lütfü Özçep, Leemhuis, and Wolter, 2020), tailored for $\mathcal{ALC}$, which has negation. Since DL-Lite$^\mathcal{H}$ allows for $B \sqsubseteq \neg C$ we use complement boxes and the 'unknown' truth state.

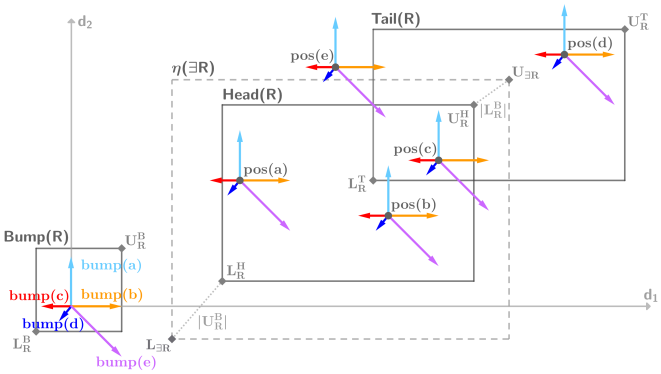

Figure 1: A two-dimensional box interpretation η is shown. It maps a role $R \in \mathsf{N}_\mathsf{R}$ to the $\mathsf{Head}(R)$, $\mathsf{Tail}(R)$, and $\mathsf{Bump}(R)$ boxes. Additionally $\eta(\exists R)$ is visualized. The depicted box interpretation satisfies the assertions $\{ R(a, b), R(b, d), R(b, c), R(c, d), R(c, c), \exists R(a), \exists R(b), \exists R(c) \}$ over $\{a, b, c, d, e\} \in \mathsf{N}_\mathsf{I}$.

4 Model Faithfulness

In this section, we study model faithfulness (Lütfü Özçep, Leemhuis, and Wolter, 2020), which is a property that is useful to show that geometric models correctly represent the knowledge present in KBs. For KB completion, one is particularly interested in preserving the conceptual knowledge in TBoxes while allowing new assertions to hold.

Definition 5. *(Adapted (Lütfü Özçep, Leemhuis, and Wolter, 2020; Bourgaux et al., 2024)) Let $\mathcal{L}$ be an ontology language and $\mathcal{K}$ a satisfiable KB in $\mathcal{L}$. A box interpretation η is*

- weakly KB faithful *for $\mathcal{L}$ and $\mathcal{K}$ if for every KB axiom α in $\mathcal{L}$, $\eta \models \alpha$ implies that α is consistent with $\mathcal{K}$;*
- KB entailed *for $\mathcal{L}$ and $\mathcal{K}$ if for every KB axiom α in $\mathcal{L}$ that is entailed by $\mathcal{K}$, $\eta \models \alpha$.*

*We may omit "weakly" and just say "faithful". We may also omit $\mathcal{L}$ and/or $\mathcal{K}$ if they are clear from the context. If a box η interpretation satisfies both of the properties we say that η is a KB **faithful model**. These notions can be adapted for the case in which $\mathcal{L}$ is a description logic with TBox and ABox axioms and for the case where we consider only TBox axioms (or only ABox axioms) for TBox (or ABox) faithful models.*

Proposition 1 and Proposition 2 give basic conditions for faithfulness in DL-Lite$^\mathcal{H}$.

Proposition 1. *Let $\mathcal{T}$ be a DL-Lite$^\mathcal{H}$ TBox. If $\eta \models \mathcal{T}$ then η is (weakly) TBox faithful.*

Proposition 1 follows from the fact that, in DL-Lite$^\mathcal{H}$, TBoxes are always consistent, that is, inconsistencies only happen when considering a TBox and an ABox.

Proposition 2. *Let $\mathcal{K}$ be a DL-Lite$^\mathcal{H}$ KB and η a box interpretation that is box consistent. If $\eta \models \mathcal{K}$ then η is (weakly) KB faithful.*

We are now ready to state Theorem 3, which is the main result of this section.

Theorem 3. *There exists a suitable s_Ω such that every satisfiable DL-Lite$^\mathcal{H}$ KB $\mathcal{K}$ has a box interpretation $\eta_{\mathcal{I}_\mathcal{K}}$ that is a KB faithful model and box consistent.*

Sketch. The proof strategy of Theorem 3 consists of first defining a mapping that translates classical finite interpretations into box interpretations. Then, given a DL-Lite$^\mathcal{H}$ KB $\mathcal{K}$, we construct the canonical model $\mathcal{I}_\mathcal{K}$ in Definition 1 and use this mapping to create a box interpretation $\eta_{\mathcal{I}_\mathcal{K}}$ (using $s_\Omega = 4$) that is a KB faithful model of $\mathcal{K}$. In fact, our proof provides a stronger guarantee: the embedding is KB entailed and strongly KB faithful (Lütfü Özçep, Leemhuis, and Wolter, 2020; Bourgaux et al., 2024). □

Informally, Theorem 3's proof proposes for every satisfiable DL-Lite$^\mathcal{H}$ KB $\mathcal{K}$ an algorithm for constructing a box interpretation (i.e., a BoxLitE embedding) that is KB faithful and box consistent. Using this algorithm, we can compute an upper bound for the minimal number of dimensions that a box interpretation requires to satisfy certain faithfulness properties, leading to Corollaries 1 and 2.

Corollary 1. *Let $\mathcal{K}$ be a DL-Lite$^\mathcal{H}$ KB with an empty ABox. Then for any $d \geq d_{\min}$ there is a box interpretation η with dimensionality d such that $d_{\min} = |\mathsf{N_C}| + 3|\mathsf{N_R}|$ and η is a TBox faithful model of $\mathcal{K}$.*

Corollary 2. *Let $\mathcal{K}$ be a satisfiable DL-Lite$^\mathcal{H}$ KB. Then for any $d \geq d_{\min}$ there is a box interpretation η with dimensionality d such that: $d_{\min} = |\mathsf{N_C}| + |\mathsf{N_R}|(2 + |\mathsf{N_I}| + 2|\mathsf{N_R}|)$ and η is a KB faithful model of $\mathcal{K}$.*

With this we have finished the theoretical analysis of BoxLitE's faithfulness properties. What now remains to show is how our BoxLitE approach can be translated into a convex optimization problem that *ensures* certain faithfulness properties. We investigate this in the next section.

5 BoxLitE's Convex Optimization Problem

Here, we formulate the representation of KBs with BoxLitE as a constrained convex optimization problem (Boyd and Vandenberghe, 2004; Rockafellar, 1997) and investigate how the formulation relates to faithfulness properties. In Section 5.1, we describe how TBox axioms are translated to constraints. In Section 5.2, we discuss the objective and scoring functions. In Section 5.3, we establish that our problem formulation ensures the KB faithful and TBox entailed properties for DL-Lite$^\mathcal{H}$. In summary, the TBox corresponds to a part of the constraints of the convex optimization problem and the ABox corresponds to a part of the objective function of the problem. In this way, a solver can handle both the ABox and TBox simultaneously by minimizing the objective function subject to the constraints.

5.1 From TBox Axioms to Constraints

Given a box interpretation η, we concatenate all of η's parameters into a single vector z of $n := (2|\mathsf{N_I}| + 2|\mathsf{N_C}| + 6|\mathsf{N_R}|)d$ dimensions[5]. We call this vector $z \in \mathbb{R}^n$ an *embedding solution*. Conversely, given such a $z \in \mathbb{R}^n$, we can reconstruct η. However, not all z leads to an η that satisfies Definition 3. Therefore, we devise constraints that z must satisfy, such that the corresponding η has desirable properties.

In particular, given an arbitrary DL-Lite$^\mathcal{H}$ TBox $\mathcal{T}$, we want to ensure that any feasible solution for a box interpretation (i) is box consistent, i.e., for all $C \in \mathsf{N}_\mathsf{C}^\exists$ it holds that $\eta(C) \cap \eta(\neg C) \neq \emptyset$, (ii) ensures that any box width is positive and bounded by $2\mathbf{s_\Omega}$, (iii) ensures that any individual embedding is within the universe box, and (iv) guarantees that any TBox axiom is satisfied in the embedding solution. In the following, we define convex constraints for Points (i)–(iv).

Box Consistency Constraints. Intuitively, Point (i) ensures that the set of feasible solutions only contains those solutions that do not predict any contradictions. However, enforcing box consistency directly is non-convex, due to the need of computing intersection boxes (see Section 7). Thus, we need to define a convex alternative that guarantees non-overlap of $\eta(C)$ and $\eta(\neg C)$. We ensure Point (i) by reserving one dimension i_C per $C \in \mathsf{N}_\mathsf{C}^\exists$ and defining one constraint per dimension i_C, formally described in (1). By reserving one dimension i_C per $C \in \mathsf{N}_\mathsf{C}^\exists$, the minimal number of dimensions of our method depends on the TBox. In particular, our method requires at least $|\mathsf{N_C}| + 2|\mathsf{N_R}|$ dimensions [6] for a TBox with $|\mathsf{N_C}|$ concept names and $|\mathsf{N_R}|$ role names.

$$\frac{\mathbf{L_C}[i_C] + \mathbf{U_C}[i_C]}{2} \leq -\frac{s_\Omega}{2} \tag{1}$$

The inequality in (1) ensures box consistency.

Box Width Constraints. As mentioned in Point (ii), our method assumes that the width of any box with lower and upper bounds $\mathbf{L}$ and $\mathbf{U}$ is positive and bounded by $2\mathbf{s_\Omega}$, i.e., $\mathbf{0} \leq_d \mathbf{U} - \mathbf{L} \leq_d 2\mathbf{s_\Omega}$. The following inequalities ensure this requirement:

$$\begin{aligned} \mathbf{0} \leq_d \mathbf{U_C} - \mathbf{L_C} \leq_d 2\mathbf{s_\Omega} \\ \mathbf{0} \leq_d \mathbf{U}^X_\mathbf{R} - \mathbf{L}^X_\mathbf{R} \leq_d 2\mathbf{s_\Omega}, \end{aligned} \tag{2}$$

where $C \in \mathsf{N_C}$, $R \in \mathsf{N_R}$, and $X \in \{\mathbf{H}, \mathbf{T}, \mathbf{B}\}$.

Universe Constraints. As mentioned in Point (iii), our method assumes that positions and bumps of individual embeddings are within the universe box Ω (see Definition 3). The following inequalities ensure this property:

$$\begin{aligned} -\mathbf{s_\Omega} + \boldsymbol{\epsilon} \leq_d \mathsf{pos}(a) \leq_d \mathbf{s_\Omega} - \boldsymbol{\epsilon} \\ -\mathbf{s_\Omega} + \boldsymbol{\epsilon} \leq_d \mathsf{bump}(a) \leq_d \mathbf{s_\Omega} - \boldsymbol{\epsilon}, \end{aligned} \tag{3}$$

where $a \in \mathsf{N_I}$.

TBox Axiom Constraints. Finally, to guarantee the satisfaction of TBox axioms in feasible solutions (Point (iv)), first

[5] For each individual in $\mathsf{N_I}$ we have two vectors, one for the position and one for the bump. For each concept in $\mathsf{N_C}$ we also have two vectors, one for the upper corner and one for the lower corner of the box. For each role in $\mathsf{N_R}$ we have 3 boxes: head, tail, and bump; each box needs the upper and lower corner vectors.

[6] We have $2|\mathsf{N_R}|$ because of $\exists R$ and $\exists R^-$ for each $R \in \mathsf{N_R}$.

we need to define the boundaries of concept and role embeddings. Based on the definitions in Section 3, the boundaries of $\eta(\neg C)$, $\eta(\exists R)$, and $\eta(R^-)$ are defined as follows:

$$\begin{aligned}
\mathbf{L}_{\neg\mathbf{C}} &\coloneqq -\mathbf{s}_{\mathbf{\Omega}} - \mathbf{L}_{\mathbf{C}},\ \mathbf{U}_{\neg\mathbf{C}} \coloneqq \mathbf{s}_{\mathbf{\Omega}} - \mathbf{U}_{\mathbf{C}},\\
\mathbf{L}_{\exists\mathbf{R}} &\coloneqq \mathbf{L}^{\mathbf{H}}_{\mathbf{R}} - \mathbf{U}^{\mathbf{B}}_{\mathbf{R}},\ \mathbf{U}_{\exists\mathbf{R}} \coloneqq \mathbf{U}^{\mathbf{H}}_{\mathbf{R}} - \mathbf{L}^{\mathbf{B}}_{\mathbf{R}},\\
\mathsf{Head}(R^-) &\coloneqq \mathsf{Tail}(R),\ \mathsf{Tail}(R^-) \coloneqq \mathsf{Head}(R),\\
\mathsf{Bump}(R^-) &\coloneqq \mathsf{Bump}(R).
\end{aligned}$$

Let $B \in \mathsf{N}^{\exists}_{\mathsf{C}}$, $C \in \mathsf{N}^{\exists}_{\mathsf{C}\neg}$, and $R, S \in \mathsf{N}^{-}_{\mathsf{R}}$. To satisfy concept and role inclusions within any embedding solution z (Definition 4), we employ the following linear inequalities:

- $B \sqsubseteq C$ corresponds to $\mathbf{L}_{\mathbf{C}} \leq_d \mathbf{L}_{\mathbf{B}}$ and $\mathbf{U}_{\mathbf{B}} \leq_d \mathbf{U}_{\mathbf{C}}$,
- $R \sqsubseteq S$ corresponds to $\mathbf{L}^X_{\mathbf{S}} \leq \mathbf{L}^X_{\mathbf{R}}, \mathbf{U}^X_{\mathbf{R}} \leq \mathbf{U}^X_{\mathbf{S}}$ with $X \in \{\mathbf{H}, \mathbf{T}, \mathbf{B}\}$.

Problem Size of the Translation. We point out that the size of BoxLitE's problem formulation increases linearly w.r.t. the size of the KB $\mathcal{K}$ and $|\mathsf{N}_{\mathsf{C}} \cup \mathsf{N}_{\mathsf{R}} \cup \mathsf{N}_{\mathsf{I}}|$. Let $\mathcal{T}$ be $\mathcal{K}$'s TBox, the number of inequalities, as described in this section, gives us $O((|\mathsf{N}_{\mathsf{I}}| + |\mathsf{N}_{\mathsf{C}}| + |\mathsf{N}_{\mathsf{R}}| + |\mathcal{T}|)d)$ constraints.

5.2 Optimization: Ranking Assertions

When ranking assertions, we would like the distance between the position of an individual w.r.t to a box associated with a concept name to affect the score of the corresponding assertion: the smaller the distance to the elements of the box, the higher the score. The same intuition also applies for roles, but needs to take into account the way role assertions are satisfied in the embedding model (Definition 4).

To express this intuition, we employ a *signed distance* function. The signed distance assigns nonpositive values to the assertions satisfied in the embedding, and positive values to those that are violated. Moreover, it reflects *how* an assertion is satisfied (or violated): for concept assertions, individuals with embeddings deeper inside the box have a smaller negative value than those closer to the border, while individuals with embeddings outside the box have positive values, and those values are higher the farther away the embeddings of the individuals are from the box.

Signed Distance. We move on to the formal definition of signed distance to a set. First, the *Euclidean distance to a set* $S \subseteq \mathbb{R}^n$ is defined as the function $\mathsf{dist}_e : \mathbb{R}^n \to \mathbb{R}$ such that $\mathsf{dist}_e(y, S) \coloneqq \inf\{\|x - y\|_2 \mid x \in S\},\ \forall y \in \mathbb{R}^n$. Let S^c be $\mathbb{R}^n \setminus S$. Then, the *signed distance* (also called *oriented distance*) to S is

$$\mathsf{sdist}(y, S) \coloneqq \begin{cases} \mathsf{dist}_e(y, S) & \text{if } y \notin S \\ -\mathsf{dist}_e(y, S^c) & \text{if } y \in S, \end{cases} \tag{4}$$

e.g., see Chapter 7 of Delfour and Zolésio (2011). Given $y \in \mathbb{R}^n$, we denote by $y^+ \in \mathbb{R}^n$ the vector that corresponds to replacing the negative components of y by 0. E.g., $(1, -2, 3, -4)^+ = (1, 0, 3, 0)$.

Proposition 3. *The signed distance to* $\mathbb{R}^n_-$ *satisfies*

$$\mathsf{sdist}(y, \mathbb{R}^n_-) = \begin{cases} \|y^+\|_2 & \text{if } y \notin \mathbb{R}^n_- \\ \max_{i \in \{1,\ldots,n\}} y_i & \text{if } y \in \mathbb{R}^n_-, \end{cases} \tag{5}$$

where $\mathbb{R}^n_- = \{y \in \mathbb{R}^n \mid y_i \leq 0, 1 \leq i \leq n\}$.

Let $\mathbf{B}$ be a box with bounds $\mathbf{L}$ and $\mathbf{U}$. We define $\mathsf{dist}(\mathbf{B}, \mathbf{x})$ as the signed distance of the concatenated vector $(\mathbf{L} + \boldsymbol{\epsilon} - \mathbf{x}) \oplus (\mathbf{x} - \mathbf{U} + \boldsymbol{\epsilon})$ to $\mathbb{R}^{2d}_-$. That is,

$$\mathsf{dist}(\mathbf{B}, \mathbf{x}) \coloneqq \mathsf{sdist}((\mathbf{L} + \boldsymbol{\epsilon} - \mathbf{x}) \oplus (\mathbf{x} - \mathbf{U} + \boldsymbol{\epsilon}), \mathbb{R}^{2d}_-).$$

We now define the loss terms using dist.

Concept Assertion Loss. We define the loss for concept assertions $D(a)$ as follows:

$$\mathcal{L}_{\mathsf{concept}}(D, a) \coloneqq \mathsf{dist}(\eta(D), \mathsf{pos}(a)).$$

Intuitively, minimizing the loss $\mathcal{L}_{\mathsf{concept}}(D, a)$ of a concept assertion $D(a)$ pushes an individual's position embedding $\mathsf{pos}(a)$ into D's concept embedding $\eta(D)$.

Role Assertion Loss. Similarly, the loss $\mathcal{L}_{\mathsf{role}}(S, a, b)$ of a role assertion $S(a, b)$ pushes the translated individual embedding $\mathsf{pos}(a) + \mathsf{bump}(b)$ and, respectively, $\mathsf{pos}(b) + \mathsf{bump}(a)$ into S's head box $\mathsf{Head}(S)$ and S's tail box $\mathsf{Tail}(S)$. Furthermore, it pushes the bumps of a and b into S's bump box $\mathsf{Bump}(S)$. Based on the distance function dist, we define the loss for role assertions $S(a, b)$ as follows:

$$\begin{aligned}
\mathcal{L}_{\mathsf{role}}(S, a, b) \coloneqq \max \big(&\mathsf{dist}(\mathsf{Head}(S), \mathsf{pos}(a) + \mathsf{bump}(b)),\\
&\mathsf{dist}(\mathsf{Tail}(S), \mathsf{pos}(b) + \mathsf{bump}(a)),\\
&\mathsf{dist}(\mathsf{Bump}(S), \mathsf{bump}(a)),\\
&\mathsf{dist}(\mathsf{Bump}(S), \mathsf{bump}(b))\big).
\end{aligned}$$

In practice, KBs typically contain only positive assertions and no explicit negative ones. Therefore, if we were to minimise only the loss terms of the assertions, BoxLitE would tend to assign high scores to all potential assertions, not distinguishing between true and false assertions. A common way to address this issue in KB embedding methods is *negative sampling* (Bordes et al., 2013; Sun et al., 2019; Abboud et al., 2020). In that approach, given a role assertion $R(h, t)$ contained in the KB, one constructs a *corrupted* assertion by exchanging either the head h or tail t of the role assertion by a randomly chosen individual from N_{I}. Since the number of assertions that hold is usually much smaller than the number of all possible assertions that can be made, most corrupted assertions tend to be false. KB embedding models are then trained to assign high scores to ABox assertions in the training data and low scores to corrupted ones. However, negative sampling leads to nonconvex loss terms, which are often incompatible with convex optimization (see Section 7); and it is computationally expensive, as competitive performance often requires sampling hundreds or even thousands of negative assertions per ABox assertion (Abboud et al., 2020; Lu and Hu, 2020; Pavlović and Sallinger, 2023b). Next, we adapt negative sampling to concepts in $\mathsf{N}^{\exists}_{\mathsf{C}}$, leading to convex regularization terms that are fast to compute and push individual embeddings into complement boxes as required.

Negative Concept Regularization. Our approach builds on three observations: (1) directly pushing individual embeddings outside a box (i.e., into its geometric complement)

leads to nonconvex optimization terms; (2) BoxLitE introduces convex boxes representing the complement of concept embeddings, which we can use to regularize the scores; and (3) BoxLitE does not offer convex representations for the complement of role embeddings, yet a role's domain and range can be expressed as existential concept embeddings. Based on these observations, we introduce a negative concept regularization term that for any potential concept assertion $D(a) \notin \mathcal{A}$ with $D \in \mathsf{N}_\mathsf{C}^\exists$ and $a \in \mathsf{N}_\mathsf{I}$ pushes $\mathsf{pos}(a)$ into $\overline{\eta(D)}$, reducing the score of $D(a)$. This regularization term keeps plausibility scores for arbitrary assertions low, while the assertion loss terms selectively increase the scores of ABox assertions explicitly contained in the KB:

$$\mathcal{L}_{\mathsf{negative}}(D, a) \coloneqq \mathcal{L}_{\mathsf{concept}}(\neg D, a). \tag{6}$$

Box Width Regularization. A second strategy to keep scores for arbitrary assertions within a reasonable range is to regularize the box size of concept and role embeddings. Specifically, for any concept name $D \in \mathsf{N}_\mathsf{C}$ we regularize the size of its concept box $\eta(D)$; and for any role name $S \in \mathsf{N}_\mathsf{R}$, we regularize the size of its head $\mathsf{Head}(S)$, tail $\mathsf{Tail}(S)$, and bump box $\mathsf{Bump}(S)$. Formally, we define the box regularization term $\mathcal{R}_{\mathsf{width}}(\mathbf{B})$ of a box $\mathbf{B}$ with bounds $\mathbf{U}$ and $\mathbf{L}$, as:

$$\mathcal{R}_{\mathsf{width}}(\mathbf{B}) \coloneqq \|\mathbf{U} - \mathbf{L}\|_2. \tag{7}$$

Objective Function. We now assemble these loss and regularization terms into our objective function:

$$\begin{aligned}
&\max\left(\max_{D(a)\in\mathcal{A}} \mathcal{L}_{\mathsf{concept}}(D,a), \max_{S(a,b)\in\mathcal{A}} \mathcal{L}_{\mathsf{role}}(S,a,b)\right) + \\
&\lambda_1 \max_{D(a)\in N_C^\exists \setminus \mathcal{A}} \mathcal{L}_{\mathsf{negative}}(D,a) + \\
&\lambda_2\Big(\sum_{D\in\mathsf{N_C}} \mathcal{R}_{\mathsf{width}}(\eta(D)) + \\
&\quad \sum_{S\in\mathsf{N_R}} \mathcal{R}_{\mathsf{width}}(\mathsf{Head}(S)) + \mathcal{R}_{\mathsf{width}}(\mathsf{Tail}(S))\Big) + \\
&\lambda_3 \sum_{S\in\mathsf{N_R}} \mathcal{R}_{\mathsf{width}}(\mathsf{Bump}(S)).
\end{aligned}$$

Scores. To rank the assertions, we define two scoring functions, one for concept and one for role assertions.

The score $s(D, a)$ of concept assertions $D(a)$ is:

$$s(D, a) \coloneqq -\mathcal{L}_{\mathsf{concept}}(D, a)$$

and the score $s(S, a, b)$ of role assertions $S(a, b)$ is:

$$s(S, a, b) \coloneqq -\mathcal{L}_{\mathsf{role}}(S, a, b).$$

The scoring functions of the concept and role assertions are the negative of the corresponding loss functions. The intuition for this is that solving the optimization problem, i.e., minimizing the concept and role assertion loss for ABox assertions, maximizes their score.

5.3 DL-Lite$^\mathcal{H}$ KB Faithfulness

Translating TBox axioms to linear inequalities that are used as convex constraints in the optimization problem (see Section 5.1), guarantees any BoxLitE embedding solution z to satisfy the concept inclusions in the TBox and to be KB faithful for DL-Lite$^\mathcal{H}$.

Let $\mathcal{C}_\mathcal{K} \subseteq \mathbb{R}^n$ be the set of z's such that the constraints of Section 5.1 are satisfied. That is, $z \in \mathcal{C}_\mathcal{K}$ if and only if: (a) for each TBox axiom the corresponding inequalities are satisfied; and (b) the box consistency and universe constraints are satisfied. Also, let $f_{\boldsymbol{\lambda}} : \mathbb{R}^n \to \mathbb{R}$ be the function that maps z to the objective value for a given choice of nonnegative hyperparameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3_+$.

Theorem 4. *Let $\mathcal{K}$ be a satisfiable DL-Lite$^\mathcal{H}$ KB. For nonnegative $\boldsymbol{\lambda}$ the following optimization problem is convex.*

$$\min_{z\in\mathbb{R}^n} f_{\boldsymbol{\lambda}}(z), \quad \text{subject to } z \in \mathcal{C}_\mathcal{K}. \tag{8}$$

In particular, $f_{\boldsymbol{\lambda}}$ is a convex function, $\mathcal{C}_\mathcal{K}$ is a polyhedral set and the following items hold.

i) For d as in Corollary 1, and s_Ω as in Theorem 3, $\mathcal{C}_\mathcal{K}$ is nonempty.

ii) Any embedding solution $z \in \mathcal{C}_\mathcal{K}$ corresponds to a box consistent interpretation that is TBox faithful.

iii) If $\lambda_1 = 0$ and there is an optimal solution z^* such that $f_{\boldsymbol{\lambda}}(z^*) \leq 0$ then the box interpretation corresponding to z^* is KB faithful.

iv) Suppose that $\lambda_1 = \lambda_2 = \lambda_3 = 0$. For d as in Corollary 2, and s_Ω as in Theorem 3, there is an optimal solution z^* s.t. $f_{\boldsymbol{\lambda}}(z^*) \leq 0$ holds.

Informally, Theorem 4 states that we can find a box interpretation for a satisfiable $\mathcal{K}$ via convex optimization over a polyhedral set. Any $z \in \mathcal{C}_\mathcal{K}$ (whether optimal or not) corresponds to a box interpretation η that is TBox faithful and box consistent. Item $i)$ gives a bound on the minimum required d to ensure that $\mathcal{C}_\mathcal{K}$ is nonempty, but this estimate seems to be conservative. Items $iii)$ and $iv)$ imply that if we wish to find a solution that is KB faithful, this can be done by setting the hyperparameters associated to regularization terms to 0, setting d to be sufficiently large and solving (8).

Finally, it turns out that (8) can be formulated as a *second-order cone program* (SOCP) (Lobo et al., 1998), (Ben-Tal and Nemirovski, 2001, Lecture 3).

Theorem 5. *For nonnegative $\boldsymbol{\lambda}$, the problem in* (8) *can be reformulated as an equivalent SOCP.*

Theorem 5 is important because it shows that (8) can be solved efficiently via high-quality open-source solvers such as SeDuMi (Sturm, 1999) and SDPT3 (Tütüncü, Toh, and Todd, 2003) or commercial solvers such as Gurobi (Gurobi Optimization, LLC, 2023) and MOSEK (ApS, 2026). The conversion of a problem as in (8) to a SOCP that can be handled by the aforementioned solvers, although tedious, can be automated by modelling tools for convex optimization such as CVXPY (Agrawal et al., 2018).

6 Proof of Concept

Here, we present empirical evidence for the theoretical foundations established so far. We evaluate BoxLitE's perfor-

mance on subsets of the Family dataset (Imenes, Guimarães, and Ozaki, 2023), providing first results for its scalability and reasoning capabilities. In our experiments, we only consider KBs with satisfiable concepts.

Reproducibility. We implemented BoxLitE's optimization problem in Python 3.12 using CVXPY (Diamond and Boyd, 2016; Agrawal et al., 2018) for modeling and MOSEK (ApS, 2026) for solving it. In our evaluation we include experiments with other KBE approaches using classical stochastic gradient descent (SGD). We use PyKEEN 1.11.1 (Ali et al., 2021) in these experiments. We ran each of our experiments on an Apple Mac Mini Desktop Computer with M4 Chip with 10 Core CPU and 10 Core GPU: 16GB (Shared Memory). The code is available at https://github.com/AleksVap/BoxLitE. We focus on the following questions.

(Q1) What is the reasoning performance and what is the effect of the regularization terms on the results?

(Q2) How does increasing the dataset size affect the prediction performance, compilation, and solving time?

(Q3) How well does our method compare with classical embedding methods based on stochastic gradient descent?

Experimental Setup. To answer each of these questions, we have created a set of datasets (F_v1-4) of varying sizes from the family dataset (Imenes, Guimarães, and Ozaki, 2023). We derived these datasets by sampling k assertions of the family dataset's ABox with forest fire sampling (Leskovec, Kleinberg, and Faloutsos, 2005), a popular sampling technique for large graphs. Furthermore, since the family dataset solely provides role assertions in its ABox, we selected all concept inclusions in the family dataset that only include roles and extended the TBox by the disjointness axiom $\exists \mathsf{hasFather}^- \sqsubseteq \neg\exists \mathsf{hasMother}^-$. We list the TBox of the created datasets in Figure 2.

Evaluation Setup. To evaluate BoxLitE's performance, we created a set of inferred role assertions by (i) adding any role assertion that logically follows from each dataset and (ii) removing any assertion that occurs in the ABox. We randomly split this set into a validation set (20%), used for model selection, and a test set (80%), used for evaluating the performance of the selected model. We use the standard evaluation setting for KB completion [7] (Abboud et al., 2020; Pavlović and Sallinger, 2024a; Xiong et al., 2022).

Dataset Properties. Table 1 lists the number of assertions and individuals of the train, validation, and test sets of F_v1-4. We sampled each of these datasets individually. Thus, although datasets F_v1-4 gradually increase in size, they are different from each other.

[7]The evaluation of a KB embedding model typically needs a set of true and corrupted role assertions. True role assertions $R(a, b)$ of the KB are corrupted by replacing a or b by any $c \in \mathsf{N_I}$ such that the corrupted assertion is not within the KB. The performance of KB embedding models is typically measured using the filtered versions (Bordes et al., 2013) of the *mean reciprocal rank* (MRR) and H@k, the proportion of true assertions within the predicted assertions whose rank is at maximum k.

$\mathsf{relative}^- \sqsubseteq \mathsf{relative}$ $\quad$ $\mathsf{hasSibling} \sqsubseteq \mathsf{relative}$

$\mathsf{hasChild} \sqsubseteq \mathsf{relative}$ $\quad$ $\mathsf{hasParent} \sqsubseteq \mathsf{relative}$

$\mathsf{hasFather} \sqsubseteq \mathsf{hasParent}$ $\quad$ $\mathsf{hasMother} \sqsubseteq \mathsf{hasParent}$

$\mathsf{spouse}^- \sqsubseteq \mathsf{spouse}$ $\quad$ $\mathsf{hasSibling}^- \sqsubseteq \mathsf{hasSibling}$

$\mathsf{spouse} \sqsubseteq \mathsf{relative}$ $\quad$ $\exists \mathsf{hasFather}^- \sqsubseteq \neg\exists \mathsf{hasMother}^-$

Figure 2: TBox of datasets F_v1-4.

Name	#Train	#Val	#Test	#Individuals
F_v1	300	110	440	155
F_v2	500	209	837	233
F_v3	1001	432	1731	368
F_v4	3014	1226	4908	895

Table 1: Dataset properties: Number of train, validation, and testing assertions, and individuals.

(Q1) Performance. In Table 2 we present the link prediction results on the test set for BoxLitE with the three regularization terms that appear in the objetive function (Section 5.2). To study the effect of each of these terms, we also performed experiments in which we remove them. We denote by BoxLitEi the version of BoxLitE without the regularization term multiplied by λ_i. The removal of each of the regularization terms decreases the overall performance of the model, with the removal of the term associated with λ_2 being the one that most negatively impacts the results.

(Q2) Scalability. The prediction performance on the test set, reported in Table 2, reduces slowly with increasing dataset sizes. Regarding the time required for each instance, we recall that a problem modelled through CVXPY is first compiled and then sent to a solver of the user's choice, which in our case is MOSEK. Given a specific choice of hyperparameters $\lambda_1, \lambda_2, \lambda_3$, Table 3 displays the compilation and solving time required for obtaining an optimal solution to the problem in Theorem 4, for F_v1-4. In our implementation, we tested 354 hyperparameter configurations for each dataset. While changing the hyperparameters requires solving the optimization problem again, it does not require a recompilation. Overall, compilation does not take more than a couple of seconds and all solution times were less than 30 seconds, which is quite reasonable considering that the final SOCP corresponding to F_v4 has around 3.5 million variables and 2.3 million constraints, which has size comparable to some of the instances that appear in Mittelmann's benchmark of large SOCPs (Mittelmann, 2026). Even more, the compilation time in Table 3 grows linearly with the number of training axioms, while the solving time increases only sublinearly.

(Q3) Comparison. Comparing BoxLitE with other KBEs in a direct way is tricky since KBEs in the literature consider other languages and are mostly optimized using SGD. BoxLitE is the first KBE for DL-Lite$^{\mathcal{H}}$ ontologies and the first that solves link prediction via convex optimization. In

Dataset	Model	MRR	H@1	H@3	H@10
F_v1	BoxLitE1	.632	.435	.800	.962
	BoxLitE2	.249	.134	.284	.475
	BoxLitE3	.698	.524	.842	.984
	BoxLitE	.720	.545	.870	.979
	BoxE	.826	.719	.918	.986
	RotatE	.474	.295	.584	.805
	ComplEx	.322	.206	.359	.535
F_v2	BoxLitE1	.428	.268	.519	.726
	BoxLitE2	.144	.072	.142	.273
	BoxLitE3	.541	.342	.668	.897
	BoxLitE	.549	.352	.685	.905
	BoxE	.432	.337	.461	.604
	RotatE	.209	.147	.218	.314
	ComplEx	.341	.265	.357	.488
F_v3	BoxLitE1	.364	.239	.396	.626
	BoxLitE2	.116	.055	.108	.224
	BoxLitE3	.414	.280	.472	.681
	BoxLitE	.433	.288	.509	.708
	BoxE	.894	.827	.946	.998
	RotatE	.177	.104	.199	.298
	ComplEx	.184	.119	.205	.284
F_v4	BoxLitE1	.339	.204	.415	.592
	BoxLitE2	.059	.023	.54	.110
	BoxLitE3	.409	.272	.483	.638
	BoxLitE	.444	.249	.571	.763
	BoxE	.626	.444	.772	.929
	RotatE	.245	.129	.300	.450
	ComplEx	.134	.087	.156	.203

Table 2: Test Results on F_v1-4. Average of 3 runs for SGD methods. Standard deviation nearly 0 in all cases.

the discussion, we include an argument for why it is challenging to design convex optimization approaches for languages with conjunctions, which appear in other papers. To illustrate how our approach roughly compares with classical embedding methods such as BoxE, RotatE, and ComplEx, based on SGD, we run those methods on the ABox part of F_v1-4. We see that the results of BoxLitE are better than RotatE and ComplEx, but still behind BoxE. One exception is F_v2 where our method performs better. None of the SGD methods had any rule injections to reflect the DL-Lite$^{\mathcal{H}}$ TBox axioms used in BoxLitE. This is a disadvantage for the SGD methods. On the other hand, BoxLitE has negative sampling applied only to existential assertions. A possible main reason for the performance gap is BoxLitE's hyperparameter optimization (HPO). It currently relies on grid search, which ensures full control over the explored parameter space, vital for ablations. Yet, grid search does not adaptively guide the hyperparameter search. By contrast, the SGD approaches, implemented in PyKEEN (Ali et al., 2021), make use of more efficient, model-based optimization techniques. Incorporating such adaptive HPO methods in future work could lead to more effective exploration of BoxLitE's hyperparameter landscape and performance gains.

Dataset	Compilation Time	Solving Time
F_v1	0.34	11.91
F_v2	0.54	13.14
F_v3	1.05	16.78
F_v4	3.12	26.24

Table 3: Time in seconds split by dataset for BoxLitE.

7 Discussion on Optimization in KBEs

In this section, we discuss some aspects related to differentiability and primary causes for nonconvexity in previous KB embedding works. We also recall how we address each challenge in our approach.

Nondifferentiability. In our approach nondifferentiability is not an issue because in view of Theorem 5 the underlying optimization problem can be cast as a second-order cone program. Informally, the nondifferentiable part of the problem gets embedded into the conic constraints and our solver of choice (MOSEK) (ApS, 2026) can handle this kind of problem without theoretical issues.

Negative sampling. Negative sampling as described in (Sun et al., 2019, Section 3.3) minimizes terms of the form:

$$-\log(\sigma(\gamma - d(x))) - \sum_{i=1}^{n} w_i(\log(\sigma(d(x_i) - \gamma))),$$

where σ is a sigmoid function (e.g., $1/(1+e^{-x})$), w_i are nonnegative weights, γ is a margin parameter, the x_i's are "negative samples" and d is a distance-like function which may include a p-norm term. Generally speaking, a function of the form $-\log(\sigma(d(z)-\gamma)$ is neither convex (nor concave) nor differentiable everywhere as a function of z. This can be seen by considering the 1-dimensional case, where d is the absolute value function and z is scalar, so that we obtain the function $-\ln(\sigma(|z|-\gamma)) = \ln(1+e^{\gamma-|z|})$, which, albeit continuous, is nonconvex and nondifferentiable at $z=0$. In contrast, BoxLitE adopts negative sampling in a convex way, using the negative concept regularization terms (Section 5.2) that push individuals into complement boxes.

Nonconvex loss terms. The loss terms of previous works often contain operations that do not preserve convexity in general. We briefly take a look at this issue in two closely related works. To be fair, none of these two works claim that their optimization problems are convex. For BoxE, it is not clear whether the distance function considered in (Abboud et al., 2020, Section 4) is convex as a *function of the parameters that need to be optimized during learning*, since it includes division and multiplication by a width term that represents box sizes. Division and multiplication are in general not operations that preserve convexity. In BoxEL, the authors consider loss terms that are quotients of volumes of boxes (or approximations thereof), e.g., see (Xiong et al., 2022, Section 4.4). Again, division does not preserve convexity in general. In contrast, all of BoxLitE's loss terms together with the distance function in our approach are convex.

The logic fragment. A final source of nonconvexity seems to be the choice of description logic fragments. Approaches based on minimizing loss terms together with logical languages that include, say, conjunction on the left-hand side typically lead to nonconvexity. Suppose that the conjunction of concepts C, D is interpreted as the set intersection of the corresponding boxes $\eta(C), \eta(D)$, as observed in the works for the $\mathcal{EL}$ ontology language. In this case, the axiom $C \sqcap D \sqsubseteq E$ translates to the constraints $\mathbf{L}_E \leq \max(\mathbf{L}_C, \mathbf{L}_D)$ and $\min(\mathbf{U}_C, \mathbf{U}_D) \leq \mathbf{U}_E$, where $\mathbf{L}_X, \mathbf{U}_X$ indicate the lower and upper bounds of the box associated to a concept X. Because the set $\mathcal{S} \coloneqq \{(a, b, c) \in \mathbb{R}^3 \mid \min(a, b) \leq c\}$ is not convex[8], these constraints are not convex in general. Since the optimal sets of convex functions are convex, it is not possible to devise a convex $f : \mathbb{R}^3 \to \mathbb{R}$ such that "$(a, b, c) \in \mathcal{S} \Leftrightarrow (a, b, c)$ is optimal for f". In particular, absent extenuating circumstances, if the bounds of C, D, E are parameters to be optimized during learning, it is impossible to construct a nonnegative *convex* loss function that is zero if and only if $C \sqcap D \sqsubseteq E$ holds. Regarding *cone semantics*, although convex optimization is mentioned as one motivation for using cones (Lütfü Özçep, Leemhuis, and Wolter, 2020), it is not explained how exactly convex optimization fits in the picture of their approach. In another work, the authors show how to use axis-aligned cones and pairs of unions of convex cones to solve certain multi-label classification problems (Leemhuis, Özçep, and Wolter, 2022) via SVMs. The approach described in (Leemhuis, Özçep, and Wolter, 2022) is significantly different from the loss function minimization approach described in other papers. Moreover, the Propositional $\mathcal{ALC}$ language used in (Leemhuis, Özçep, and Wolter, 2022) is different from DL-Lite$^{\mathcal{H}}$, which we consider in this work, since DL-Lite$^{\mathcal{H}}$ features roles and inverses.

8 Conclusion and Future Work

We propose BoxLitE, a KB embedding method that allows for convex optimization and ensures the satisfaction of TBox axioms. We prove that for any DL-Lite$^{\mathcal{H}}$ KB, there is a faithful embedding solution that is a KB model. We implement and evaluate BoxLitE's convex problem formulation for KB embeddings in CVXPY and MOSEK. The results reveal that MOSEK finds a solution that is KB faithful and that satisfies the concept inclusions in the TBox for a prototypical ontology. Within a few seconds of solving time, MOSEK finds embedding solutions on subsets (F_v1 to F_v4) of a real-world KB that achieve good link prediction results. In the future, we would like to consider more efficient methods for the HPO and evaluation. Also, we want to study how to ensure other theoretical properties in convex KBE methods. As languages that allow for conjunctions on the left of inclusions often lead to nonconvexity, another line lies in studying sound nonconvex KBE approaches that can ensure the construction of faithful models. Finally, we would like to investigate, using nonconvex tools, the effect of pushing negative samples outside the concept box into an 'unknown' truth state, and how this affects prediction results.

[8]It suffices to observe that $(1, 0, 0), (0, 1, 0) \in \mathcal{S}$, but $0.5(1, 0, 0) + 0.5(0, 1, 0) = (0.5, 0.5, 0) \notin \mathcal{S}$.

Acknowledgements

This work was supported by the "Strategic Research Projects" grant from ROIS (Research Organization of Information and Systems). Bruno F. Lourenço's work was partially supported by the JSPS Grant-in-Aid for Early-Career Scientists 23K16844. Ana Ozaki was supported by the Research Council of Norway, projects (316022, 322480) and Integreat - Norwegian Centre for knowledge-driven machine learning (332645). Emanuel Sallinger's and Hesham Morgan's work on this paper was funded by the Vienna Science and Technology Fund (WWTF) [Grant ID: 10.47379/VRG18013, 10.47379/ICT25032, 10.47379/NXT22018, 10.47379/ICT2201, 10.47379/DCDH001], and by the Austrian Science Fund (FWF) 10.55776/COE12.

AI Declaration

Gen AI tools were only used to help find grammatical mistakes and to aid in the process of debugging the code.

References

Abboud, R.; Ceylan, İ. İ.; Lukasiewicz, T.; and Salvatori, T. 2020. BoxE: A box embedding model for knowledge base completion. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *NeurIPS*.

Agrawal, A.; Verschueren, R.; Diamond, S.; and Boyd, S. 2018. A rewriting system for convex optimization problems. *Journal of Control and Decision* 5(1):42–60.

Ali, M.; Berrendorf, M.; Hoyt, C. T.; Vermue, L.; Sharifzadeh, S.; Tresp, V.; and Lehmann, J. 2021. PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *Journal of Machine Learning Research* 22(82):1–6.

ApS, M. 2026. *MOSEK Optimizer API for Python 11.1.3*.

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The DL-Lite family and relations. *J. Artif. Intell. Res.* 36:1–69.

Balazevic, I.; Allen, C.; and Hospedales, T. 2019. TuckER: Tensor factorization for knowledge graph completion. In *EMNLP-IJCNLP*, 5185–5194. Association for Computational Linguistics.

Beck, A. 2017. *First-Order Methods in Optimization*. SIAM.

Ben-Tal, A., and Nemirovski, A. 2001. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM.

Bolte, J., and Pauwels, E. 2020. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. *Mathematical Programming* 188(1):19–51.

Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In Burges, C. J. C.; Bottou, L.; Ghahramani, Z.; and Weinberger, K. Q., eds., *27th Annual*

Conference on Neural Information Processing Systems, 2787–2795.

Bourgaux, C.; Guimarães, R.; Koudijs, R.; Lacerda, V.; and Ozaki, A. 2024. Knowledge base embeddings: Semantics and theoretical properties. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *KR*.

Bourgaux, C.; Ozaki, A.; and Pan, J. Z. 2021. Geometric models for (temporally) attributed description logics. In Homola, M.; Ryzhikov, V.; and Schmidt, R. A., eds., *(DL*, volume 2954 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Boyd, S., and Vandenberghe, L. 2004. *Convex Optimization*. Cambridge, England: Cambridge University Press.

Davis, D.; Drusvyatskiy, D.; Kakade, S.; and Lee, J. D. 2019. Stochastic subgradient method converges on tame functions. *Foundations of Computational Mathematics* 20(1):119–154.

Delfour, M. C., and Zolésio, J. P. 2011. *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization, Second Edition*. Society for Industrial and Applied Mathematics.

Diamond, S., and Boyd, S. 2016. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* 17(83):1–5.

Gurobi Optimization, LLC. 2023. Gurobi Optimizer Reference Manual.

Gutiérrez-Basulto, V., and Schockaert, S. 2018. From knowledge graph embedding to ontology embedding? an analysis of the compatibility between vector space representations and rules. In Thielscher, M.; Toni, F.; and Wolter, F., eds., *KR*, 379–388. AAAI Press.

Imenes, A.; Guimarães, R.; and Ozaki, A. 2023. Marrying query rewriting and knowledge graph embeddings. In *RuleML+RR*, 126–140. Berlin, Heidelberg: Springer-Verlag.

Jackermeier, M.; Chen, J.; and Horrocks, I. 2024. Dual box embeddings for the description logic EL++. In *Proceedings of the ACM Web Conference*, WWW, 2250–2258. Association for Computing Machinery.

Kazemi, S. M., and Poole, D. 2018. Simple embedding for link prediction in knowledge graphs. In Bengio, S.; Wallach, H. M.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *NeurIPS*, 4289–4300.

Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. In Bengio, Y., and LeCun, Y., eds., *ICLR*.

Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyaschev, M. 2010. The combined approach to query answering in dl-lite. In Lin, F.; Sattler, U.; and Truszczynski, M., eds., *KR*. AAAI Press.

Kulmanov, M.; Liu-Wei, W.; Yan, Y.; and Hoehndorf, R. 2019. EL embeddings: Geometric construction of models for the description logic EL++. In Kraus, S., ed., *IJCAI*, 6103–6109. ijcai.org.

Lacerda, V.; Ozaki, A.; and Guimarães, R. 2024a. FaithEL: Strongly tbox faithful knowledge base embeddings for *E*. In Kirrane, S.; Simkus, M.; Soylu, A.; and Roman, D., eds., *RuleML+RR*, volume 15183 of *Lecture Notes in Computer Science*, 191–199. Springer.

Lacerda, V.; Ozaki, A.; and Guimarães, R. 2024b. Strong faithfulness for ELH ontology embeddings. *TGDK* 2(3):2:1–2:29.

Leemhuis, M.; Özçep, O. L.; and Wolter, D. 2022. Learning with cone-based geometric models and orthologics. *Annals of Mathematics and Artificial Intelligence* 90(11–12):1159–1195.

Leskovec, J.; Kleinberg, J. M.; and Faloutsos, C. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. In Grossman, R.; Bayardo, R. J.; and Bennett, K. P., eds., *ACM SIGKDD*, 177–187. ACM.

Li, W.; Zheng, X.; Gao, H.; Ji, Q.; and Qi, G. 2022. Cosine-based embedding for completing lightweight schematic knowledge in dl-litecore. *Applied Sciences* 12(20).

Lobo, M. S.; Vandenberghe, L.; Boyd, S.; and Lebret, H. 1998. Applications of second-order cone programming. *Linear Algebra and its Applications* 284(1–3):193–228.

Lourenço, B. F.; Morgan, H.; Ozaki, A.; Pavlović, A.; and Sallinger, E. 2026. BoxLitE: A faithful knowledge base embedding based on convex optimization. *CoRR* abs/2605.23937.

Lu, H., and Hu, H. 2020. Dense: An enhanced non-abelian group representation for knowledge graph embedding. *CoRR* abs/2008.04548.

Lütfü Özçep, O.; Leemhuis, M.; and Wolter, D. 2020. Cone semantics for logics with negation. In Bessiere, C., ed., *IJCAI*, 1820–1826. International Joint Conferences on Artificial Intelligence Organization. Main track.

Mittelmann, H. 2026. Large second order cone benchmark. https://plato.asu.edu/ftp/socp.html. [Online; accessed 17-February-2026].

Mondal, S.; Bhatia, S.; and Mutharaju, R. 2021. Emel++: Embeddings for EL++ description logic. In Martin, A.; Hinkelmann, K.; Fill, H.; Gerber, A.; Lenat, D.; Stolle, R.; and van Harmelen, F., eds., *AAAI-MAKE*, volume 2846 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Morgan, H.; Pavlović, A.; Ozaki, A.; Lourenço, B. F.; and Sallinger, E. 2026. Convexity for trustworthy knowledge graph embeddings: From interpretability to reliability. In *Proceedings of the 19th Forschungsforum der österreichischen Fachhochschulen (FHK)*.

Nesterov, Y., and Nemirovskii, A. 1994. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM.

Pavlović, A., and Sallinger, E. 2023a. Building bridges: Knowledge graph embeddings respecting logical rules. In

Kimelfeld, B.; Martinez, M. V.; and Angles, R., eds., *Proceedings of the 15th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW)*, volume 3409 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Pavlović, A., and Sallinger, E. 2023b. ExpressivE: A spatio-functional embedding for knowledge graph completion. In *the 11th International Conference on Learning Representations (ICLR 2023)*. OpenReview.net.

Pavlović, A., and Sallinger, E. 2024a. Raising the efficiency of knowledge graph embeddings while respecting logical rules. In Montoya, G.; Sallinger, E.; and Vargas-Solar, G., eds., *Proceedings of the 16th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW)*, CEUR Workshop Proceedings. CEUR-WS.org.

Pavlović, A., and Sallinger, E. 2024b. SpeedE: Euclidean geometric knowledge graph embedding strikes back. In Duh, K.; Gomez, H.; and Bethard, S., eds., *Findings of the Association for Computational Linguistics: NAACL 2024*, 69–92. Association for Computational Linguistics.

Pavlović, A.; Sallinger, E.; and Schockaert, S. 2025. Faithful Differentiable Reasoning with Reshuffled Region-based Embeddings. In *Proceedings of the 22nd International Conference on Principles of Knowledge Representation and Reasoning*, 489–499.

Peng, X.; Tang, Z.; Kulmanov, M.; Niu, K.; and Hoehndorf, R. 2022. Description logic EL++ embeddings with intersectional closure. *CoRR* abs/2202.14018.

PyTorch. 2026. Pytorch - autograd mechanics. https://docs.pytorch.org/docs/stable/notes/autograd.html#gradients-for-non-differentiable-functions. [Online; accessed 17-February-2026].

Rockafellar, R. T. 1997. *Convex Analysis* . Princeton University Press.

Sturm, J. F. 1999. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software* 11(1-4):625–653.

Sun, Z.; Deng, Z.; Nie, J.; and Tang, J. 2019. RotatE: Knowledge graph embedding by relational rotation in complex space. In *ICLR*. OpenReview.net.

Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; and Bouchard, G. 2016. Complex embeddings for simple link prediction. In Balcan, M., and Weinberger, K. Q., eds., *ICML*, volume 48 of *JMLR Workshop and Conference Proceedings*, 2071–2080. JMLR.org.

Tütüncü, R. H.; Toh, K. C.; and Todd, M. J. 2003. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming* 95(2):189–217.

Xiong, B.; Potyka, N.; Tran, T.; Nayyeri, M.; and Staab, S. 2022. Faithful embeddings for E^{++} knowledge bases. In Sattler, U.; Hogan, A.; Keet, C. M.; Presutti, V.; Almeida, J. P. A.; Takeda, H.; Monnin, P.; Pirrò, G.; and d'Amato, C., eds., *ISWC*, volume 13489 of *Lecture Notes in Computer Science*, 22–38. Springer.

Yang, B.; Yih, W.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. In Bengio, Y., and LeCun, Y., eds., *ICLR*.

Yang, H.; Chen, J.; and Sattler, U. 2025. Transbox: El++-closed ontology embedding. In *Proceedings of the ACM on Web Conference*, WWW, 22–34. Association for Computing Machinery.