

# From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules

Víctor Gutiérrez-Basulto, Steven Schockaert

School of Computer Science and Informatics

Cardiff University, Cardiff, UK

{gutierrezbasultov, schockaerts1}@cardiff.ac.uk

## Abstract

Recent years have witnessed the successful application of low-dimensional vector space representations of knowledge graphs to predict missing facts or find erroneous ones. However, it is not yet well-understood to what extent ontological knowledge, e.g. given as a set of (existential) rules, can be embedded in a principled way. To address this shortcoming, in this paper we introduce a general framework based on a view of relations as regions, which allows us to study the compatibility between ontological knowledge and different types of vector space embeddings. Our technical contribution is two-fold. First, we show that some of the most popular existing embedding methods are not capable of modelling even very simple types of rules, which in particular also means that they are not able to learn the type of dependencies captured by such rules. Second, we study a model in which relations are modelled as *convex* regions. We show particular that ontologies which are expressed using so-called quasi-chained existential rules can be exactly represented using convex regions, such that any set of facts which is induced using that vector space embedding is logically consistent and deductively closed with respect to the input ontology.

## 1 Introduction

Knowledge graphs (KGs), i.e. sets of (*subject,predicate,object*) triples, play an increasingly central role in fields such as information retrieval and natural language processing (Dong et al. 2014; Camacho-Collados, Pilehvar, and Navigli 2016). A wide variety of KGs are currently available, including carefully curated resources such as WordNet (Miller 1995), crowdsourced resources such as Freebase (Bollacker et al. 2008), ConceptNet (Speer, Chin, and Havasi 2017) and WikiData (Vrandečić and Krötzsch 2014), and resources that have been extracted from natural language such as NELL (Carlson et al. 2010). However, despite the large scale of some of these resources, they are, perhaps inevitably, far from complete. This has sparked a large amount of research on the topic of automated knowledge base completion, e.g. random-walk based machine learning models (Gardner and Mitchell 2015) and factorization and embedding approaches (Wang et al. 2017). The main premise underlying these approaches is that many plausible

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

triples can be found by exploiting the regularities that exist in a typical knowledge graph. For example, if we know (*Peter Jackson, Directed, The fellowship of the ring*) and (*The fellowship of the ring, Has-sequel, The two towers*), we may expect the triple (*Peter Jackson, Directed, The two towers*) to be somewhat plausible, if we can observe from the rest of the knowledge graph that sequels are often directed by the same person.

Due to their conceptual simplicity and high scalability, *knowledge graph embeddings* have become one of the most popular strategies for discovering and exploiting such regularities. These embeddings are  $n$ -dimensional vector space representations, in which each entity  $e$  (i.e. each node from the KG) is associated with a vector  $\mathbf{e} \in \mathbb{R}^n$  and each relation name  $R$  is associated with a scoring function  $s_R : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  that encodes information about the likelihood of triples. For the ease of presentation, we will formulate KG embedding models such that  $s_{R_1}(\mathbf{e}_1, \mathbf{f}_1) < s_{R_2}(\mathbf{e}_2, \mathbf{f}_2)$  iff the triple  $(e_1, R_1, f_1)$  is considered *more* likely than the triple  $(e_2, R_2, f_2)$ . Both the entity vectors  $\mathbf{e}$  and the scoring functions  $s_R$  are learned from the information in the given KG. The main assumption is that the resulting vector space representation of the KG is such that it captures the important regularities from the considered domain. In particular, there will be triples  $(e, R, f)$  which are not in the original KG, but for which  $s_R(\mathbf{e}, \mathbf{f})$  is nonetheless low. They thus correspond to facts which are plausible, given the regularities that are observed in the KG as a whole, but which are not contained in the original KG. The number of dimensions  $n$  of the embedding essentially controls the cautiousness of the knowledge graph completion process: the fewer dimensions, the more regularities can be discovered by the model, but the higher the risk of unwarranted inferences. On the other hand, if the number of dimensions is too high, the embedding may simply capture the given KG, without suggesting any additional plausible triples.

For example, in the seminal TransE model (Bordes et al. 2013), relations are modelled as vector translations. In particular, the TransE scoring function is given by  $s_R(\mathbf{e}, \mathbf{f}) = d(\mathbf{e} + \mathbf{r}, \mathbf{f})$ , where  $d$  is the Euclidean distance and  $\mathbf{r} \in \mathbb{R}^n$  is a vector encoding of the relation name  $R$ . Another popular model is DistMult (Yang et al. 2015), which corresponds to the choice  $s_R(\mathbf{e}, \mathbf{f}) = -\sum_{i=1}^n e_i r_i f_i$ , where we write  $e_i$  for the  $i^{th}$  coordinate of  $\mathbf{e}$ , and similar for  $\mathbf{f}$  and  $\mathbf{r}$ .

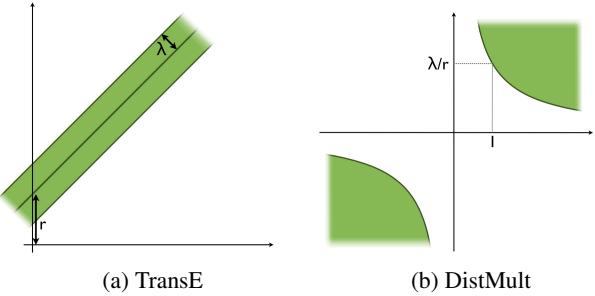


Figure 1: Region based view of knowledge graph embedding models.

To date, surprisingly little is understood about the types of regularities that existing embedding methods can capture. In this paper, we are particularly concerned with the types of (hard) rules that such models are capable of representing. To allow us to precisely characterize what regularities are captured by a given embedding, we will consider hard thresholds  $\lambda_R$  such that a triple  $(e, R, f)$  is considered valid iff  $s_R(e, f) \leq \lambda_R$ . In fact, KG embeddings are often learned using a max-margin loss function which directly encodes this assumption. The vector space representation of a given relation  $R$  can then be viewed as a region  $\eta(R)$  in  $\mathbb{R}^{2n}$ , defined as follows:

$$\eta(R) = \{e \oplus f \mid s_R(e, f) \leq \lambda_R\}$$

where we write  $\oplus$  for vector concatenation. In particular, note that  $(e, R, f)$  is considered a valid triple iff  $e \oplus f \in \eta_R$ . Figure 1 illustrates the types of regions that are obtained for the TransE and DistMult models.

This region-based view will allow us to study properties of knowledge graph embedding models in a general way, by linking the kind of regularities that a given embedding model can represent to the kind of regions that it considers. Furthermore, the region-based view of knowledge graph embedding also has a number of practical advantages. First, such regions can naturally be defined for relations of any arity, while the standard formulations of knowledge graph embedding models are typically restricted to binary relations. Second, and perhaps more fundamentally, it suggests a natural way to take into account prior knowledge about dependencies between different relations. In particular, for many knowledge graphs, some kind of ontology is available, which can be viewed as a set of rules describing such dependencies. These rules naturally translate to spatial constraints on the regions  $\eta_R$ . For instance, if we know that  $R(X, Y) \rightarrow S(X, Y)$  holds, it would be natural to require that  $\eta_R \subseteq \eta_S$ . If a knowledge graph embedding captures the rules of a given ontology in this sense, we will call it a *geometric model* of the ontology. By requiring that the embedding of a knowledge graph should be a geometric model of a given ontology, we can effectively exploit the knowledge contained in that ontology to obtain higher-quality representations. Indeed, there exists empirical support for the usefulness of (soft) rules for learning embeddings (Demeester, Rocktäschel, and Riedel 2016;

Niepert 2016; Wang and Cohen 2016; Minervini et al. 2017). A related advantage of geometric models over standard KG embeddings is that the set of triples which is considered valid based on the embedding is guaranteed to be logically consistent and deductively closed (relative to the given ontology). Finally, since geometric models are essentially “ontology embeddings”, they could be used for ontology completion, i.e. for finding plausible missing rules from the given ontology similar to how standard KG embedding models are used to find plausible missing triples from a KG.

**Objective and Contributions.** The main aim of this paper is to analyze the implications of choosing a particular type of geometric representation on the kinds of logical dependencies that can be faithfully embedded. To the best of our knowledge, this paper is the first to investigate the expressivity of embedding models in the latter sense.

Our technical contribution is two-fold. First, we show that the most popular approaches to KG embedding are actually not compatible with the notion of a geometric model. For instance, as we will see, the representations obtained by DistMult (and its variants) can only model a very restricted class of subsumption hierarchies. This is problematic, as it not only means that we cannot impose the rules from a given ontology for learning knowledge graph embeddings, but also that the types of regularities that are captured by such rules cannot be learned from data.

Second, to overcome the above shortcoming, we propose a novel framework in which relations are modelled as arbitrary convex regions in  $\mathbb{R}^k$ , with  $k$  the arity of the relation. We particularly show that convex geometric models can properly express the class of so-called *quasi-chained existential rules*. While convex geometric models are thus still not general enough to capture arbitrary existential rules, this particular class does subsume several key ontology languages based on description logics and important fragments of existential rules. Finally, we show that to capture arbitrary existential rules, a further generalization is needed, based on a non-linear transformation of the vector concatenations.

Missing proofs can be found in a extended version with an appendix under <https://tinyurl.com/yb696el8>

## 2 Background

In this section we provide some background on knowledge graph embedding and existential rules.

### 2.1 Knowledge Graph Embedding

A wide variety of KG embedding methods have already been proposed, varying mostly in the type of scoring function that is used. One popular class of methods was inspired by the TransE model. In particular, several authors have proposed generalizations of TransE to address the issue that TransE is only suitable for one-to-one relations (Wang et al. 2014; Lin et al. 2015): if  $(e, R, f)$  and  $(e, R, g)$  were both in the KG, then the TransE training objective would encourage  $f$  and  $g$  to be represented as identical vectors. The main idea behind these generalizations is to map the entities to a relation-specific subspace before applying the translation. For instance, the TransR scoring function is given by

$s_R(\mathbf{e}, \mathbf{f}) = d(M_r \mathbf{e} + \mathbf{r}, M_r \mathbf{f})$ , where  $M_r$  is an  $n \times n$  matrix (Lin et al. 2015). As a further generalization, in STransE a different matrix is used for the head entity  $e$  and for the tail entity  $f$ , leading to the scoring function  $s_R(\mathbf{e}, \mathbf{f}) = d(M_r^h \mathbf{e} + \mathbf{r}, M_r^t \mathbf{f})$  (Nguyen et al. 2016).

A key limitation of DistMult (cf. Section 1) is the fact that it can only model symmetric relations. A natural solution is to represent each entity  $e$  using two vectors  $\mathbf{e}_h$  and  $\mathbf{e}_t$ , which are respectively used when  $e$  appears in the head (i.e. as the first argument) or in the tail (i.e. as the second argument). In other words, the scoring function then becomes  $s_R = -\sum_i e_i^h r_i f_i^t$ , where we write  $\mathbf{e}_h = (e_1^h, \dots, e_n^h)$  and similar for  $\mathbf{f}_t$ . The problem with this approach is that there is no connection at all between  $\mathbf{e}_h$  and  $\mathbf{e}_t$ , which makes learning suitable representations more difficult. To address this, the ComplEx model (Trouillon et al. 2016) represents entities and relations as vectors of complex numbers, such that  $\mathbf{e}_t$  is the component-wise conjugate of  $\mathbf{e}_h$ . Let us write  $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle$ , with  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ , for the bilinear product  $\sum_{i=1}^n a_i b_i c_i$ . Furthermore, for a complex vector  $\mathbf{a} \in \mathbb{C}^n$ , we write  $re(\mathbf{a})$  and  $im(\mathbf{a})$  for the real and imaginary parts of  $\mathbf{a}$  respectively. It can be shown (Kazemi and Poole 2018) that the scoring function of ComplEx is equivalent to

$$s_R(e, f) = -\langle re(\mathbf{e}), re(\mathbf{r}), re(\mathbf{f}) \rangle - \langle re(\mathbf{e}), im(\mathbf{r}), im(\mathbf{f}) \rangle \\ - \langle im(\mathbf{e}), re(\mathbf{r}), im(\mathbf{f}) \rangle + \langle im(\mathbf{e}), im(\mathbf{r}), re(\mathbf{f}) \rangle$$

Recently, in (Kazemi and Poole 2018), a simpler approach was proposed to address the symmetry issue of DistMult. The proposed model, called SimplE, avoids the use of complex vectors. In this model, the DistMult scoring function is used with a separate representation for head and tail mentions of an entity, but for each triple  $(e, R, f)$  in the knowledge graph, the triple  $(f, R^{-1}, e)$  is additionally considered. This means that each such triple affects the representation of  $\mathbf{e}_h$ ,  $\mathbf{e}_t$ ,  $\mathbf{f}_h$  and  $\mathbf{f}_t$ , and in this way, the main drawback of using separate representations for head and tail mentions is avoided.

The RESCAL model (Nickel, Tresp, and Kriegel 2011) uses a bilinear scoring function  $s_R(e, f) = -\mathbf{e}^T M_r \mathbf{f}$ , where the relation  $R$  is modelled as an  $n \times n$  matrix  $M_r$ . Note that DistMult can be seen as a special case of RESCAL in which only diagonal matrices are considered. Similarly, it is easy to verify that ComplEx also corresponds to a bilinear model, with a slightly different restriction on the type of considered matrices. Without any restriction on the type of considered matrices, however, the RESCAL model is prone to overfitting. The neural tensor model (NTN), proposed in (Socher et al. 2013) further generalizes RESCAL by using a two-layer neural network formulation, but similarly tends to suffer from overfitting in practice.

**Expressivity.** Intuitively, the reason why KG embedding models are able to identify plausible triples is because they can only represent knowledge graphs that exhibit a certain type of regularity. They can be seen as a particular class of dimensionality reduction methods: the lower the number of dimensions  $n$ , the stronger the KG model enforces some notion of regularity (where the exact kind of regularity depends on the chosen KG embedding model).

However, when the number of dimensions is sufficiently high, it is desirable that any KG can be represented in an exact way, in the following sense: for any given set of triples  $P = \{(e_1, R_1, f_1), \dots, (e_m, R_m, f_m)\}$  which are known to be valid and any set of triples  $N = \{(e_{m+1}, R_{m+1}, f_{m+1}), \dots, (e_k, R_k, f_k)\}$  which are known to be false, given a sufficiently high number of dimensions  $n$ , there always exists an embedding and thresholds  $\lambda_R$  such that

$$\forall (e, R, f) \in P . s_R(\mathbf{e}, \mathbf{f}) \leq \lambda_R \quad (1)$$

$$\forall (e, R, f) \in N . s_R(\mathbf{e}, \mathbf{f}) > \lambda_R \quad (2)$$

A KG embedding model is called *fully expressive* (Kazemi and Poole 2018) if (1)–(2) can be guaranteed for any disjoint sets of triples  $P$  and  $N$ . If a KG embedding model is not fully expressive, it means that there are *a priori* constraints on the kind of knowledge graphs that can be represented, which can lead to unwarranted inferences when using this model for KG completion. In contrast, for fully expressive models, the types of KGs that can be represented is determined by the number of dimensions, which is typically seen as a hyperparameter, i.e. this number is tuned separately for each KG to avoid (too many) unwarranted inferences.

It turns out that translation based methods such as TransE, STransE and related generalizations are not fully expressive (Kazemi and Poole 2018), and in fact put rather severe restrictions on the types of relations that can be represented in the sense of (1)–(2). For instance, it was shown in (Kazemi and Poole 2018) that translation based methods can only fully represent a knowledge graph  $G$  if each of its relations  $R$  satisfies the following properties for every subset of entities  $S$ :

1. If  $R$  is reflexive over  $S$ , then  $R$  is also symmetric and transitive over  $S$ .
2. If  $\forall s \in S . (e, R, s) \in G$  and  $\exists s \in S . (f, R, s) \in G$  then we also have  $\forall s \in S . (f, R, s) \in G$ .

However, both ComplEx and SimplE have been shown to be fully expressive.

**Modelling Textual Descriptions.** Several methods have been proposed which aim to learn better knowledge graph embeddings by exploiting textual descriptions of entities (Zhong et al. 2015; Xie et al. 2016; Xiao et al. 2017) or by extracting information about the relationship between two entities from sentences mentioning both of them (Toutanova et al. 2015). Apart from improving the overall quality of the embeddings, a key advantage of such approaches is that they allow us to predict plausible triples involving entities which do not occur in the initial knowledge graph.

## 2.2 Existential Rules

Existential rules (a.k.a. Datalog $^\pm$ ) are a family of rule-based formalisms for modelling ontologies. An existential rule is a datalog-like rule with existentially quantified variables in the head, i.e. it extends traditional datalog with *value invention*. As a consequence, existential rules describe not only constraints on the currently available knowledge or data, but also *intentional* knowledge about the domain of discourse.

The appeal of existential rules comes from the fact that they are extensions of the prominent  $\mathcal{EL}$  and  $DL$ -Lite families of description logics (DLs) (Baader et al. 2017). For instance, existential rules can describe  $k$ -ary relations, while DLs are constrained to unary and binary relations.

**Syntax.** Let  $C, N$  and  $V$  be infinite disjoint sets of *constants*, (*labelled*) *nulls* and *variables*, respectively. A *term*  $t$  is an element in  $C \cup N \cup V$ ; an *atom*  $\alpha$  is an expression of the form  $R(t_1, \dots, t_n)$ , where  $R$  is a *relation name* (or *predicate*) with *arity*  $n$  and terms  $t_i$ . We denote with  $\text{terms}(\alpha)$  the set  $\{t_1, \dots, t_n\}$  and with  $\text{vars}(\alpha)$  the set  $\text{terms}(\alpha) \cap V$ . An *existential rule*  $\sigma$  is an expression of the form

$$B_1 \wedge \dots \wedge B_n \rightarrow \exists X_1, \dots, X_j. H_1 \wedge \dots \wedge H_k, \quad (3)$$

where  $B_1, \dots, B_n$  for  $n \geq 0$ ,  $H_1, \dots, H_k$  for  $k \geq 1$ , are atoms with terms in  $C \cup V$  and  $X_m \in V$  for  $1 \leq m \leq j$ . From here on, we assume w.l.o.g that  $k = 1$  (Calì, Gottlob, and Kifer 2013); we omit in this case the subindex. We use  $\text{body}(\sigma)$  and  $\text{head}(\sigma)$  to refer to  $\{B_1, \dots, B_n\}$  and  $\{H\}$ , respectively. We call  $\text{evars}(\sigma) = \{X_1, \dots, X_j\}$  the *existential variables* of  $\sigma$ ; if  $\text{evars}(\sigma) = \emptyset$ ,  $\sigma$  is called a *datalog rule*. We further allow *negative constraints* (or simply *constraints*) which are expressions of the form  $B_1 \wedge \dots \wedge B_n \rightarrow \perp$ , where the  $B_i$ s are as above and  $\perp$  denotes the truth constant *false*. A finite set  $\Sigma$  of existential rules and constraints is called an *ontology*; and a *datalog program* if  $\Sigma$  contains only datalog rules and constraints.

Let  $\mathfrak{R}$  be a set of relation names. A *database*  $D$  is a finite set of *facts* over  $\mathfrak{R}$ , i.e. atoms with terms in  $C$ . A *knowledge base (KB)*  $\mathcal{K}$  is a pair  $(\Sigma, D)$  with  $\Sigma$  an ontology (or a datalog program) and  $D$  a database.

**Semantics.** An *interpretation*  $\mathcal{I}$  over  $\mathfrak{R}$  is a (possibly infinite) set of atoms over  $\mathfrak{R}$  with terms in  $C \cup N$ . An interpretation  $\mathcal{I}$  is a *model* of  $\Sigma$  if it satisfies all rules and constraints:  $\{B_1, \dots, B_n\} \subseteq \mathcal{I}$  implies  $\{H\} \subseteq \mathcal{I}$  for every  $\sigma$  defined as above in  $\Sigma$ , where existential variables can be witnessed by constants or labelled nulls, and  $\{B_1, \dots, B_n\} \not\subseteq \mathcal{I}$  for all constraints defined as above in  $\Sigma$ ; it is a *model* of a database  $D$  if  $D \subseteq \mathcal{I}$ ; it is a model of a KB  $\mathcal{K} = (\Sigma, D)$ , written  $\mathcal{I} \models \mathcal{K}$ , if it is a model of  $\Sigma$  and  $D$ . We say that a KB  $\mathcal{K}$  is satisfiable if it has a model. We refer to elements in  $C \cup N$  simply as *objects*, call atoms  $\alpha$  containing only objects as terms *ground*, and denote with  $\mathfrak{O}(\mathcal{I})$  the set of all objects occurring in  $\mathcal{I}$ .

**Example 1.** Let  $D = \{\text{Wife}(anna), \text{Wife}(marie)\}$  be a database and  $\Sigma$  an ontology composed by the rules:

$$\text{Wife}(X) \wedge \text{Married}(X, Y) \rightarrow \text{Husband}(Y) \quad (4)$$

$$\text{Wife}(Y) \rightarrow \exists X. \text{Husband}(X) \wedge \text{Married}(X, Y) \quad (5)$$

$$\text{Husband}(X) \wedge \text{Wife}(X) \rightarrow \perp \quad (6)$$

Then, an example of a model of  $\mathcal{K} = (\Sigma, D)$  is the set of atoms  $D \cup \{\text{Husband}(o_1), \text{Husband}(o_2), \text{Married}(o_1, anna), \text{Married}(o_2, marie)\}$  where  $o_i$  are labelled nulls. Note that e.g.  $\{\text{Married}(anna, marie), \text{Husband}(marie)\}$  is not included in any model of  $\mathcal{K}$  due to (6).

**Notation.** We use  $a, b, c, a_1, \dots$  for constants and  $X, Y, Z, X_1, \dots$  for variables. We write  $\mathfrak{R}_k$  for the set

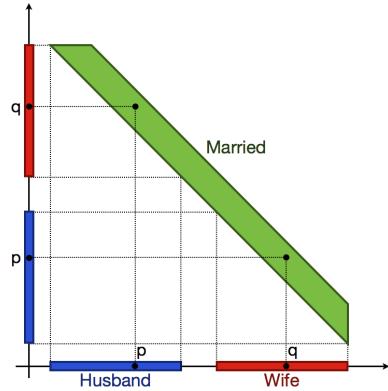


Figure 2: A geometric model of the KB from Example 1.

of relation names from  $\mathfrak{R}$  which have arity  $k$ . Given a KB  $\mathcal{K}$ , we use  $C(\mathcal{K}), \mathfrak{R}(\mathcal{K})$  and  $\mathfrak{R}_k(\mathcal{K})$  to denote, respectively, the set of constants, relation names and  $k$ -ary relation names occurring in  $\mathcal{K}$ . For vectors  $\mathbf{x} = (x_1, \dots, x_m)$  and  $\mathbf{y} = (y_1, \dots, y_k)$ , we denote their concatenation by  $\mathbf{x} \oplus \mathbf{y} = (x_1, \dots, x_m, y_1, \dots, y_k)$ .

### 3 Geometric Models

In this section, we formalize how regions can be used for representing relations, and what it means for such representations to satisfy a given knowledge base. The resulting formalization will allow us to study the expressivity of knowledge graph embedding models. It will also provide the foundations of a framework for knowledge base completion, based on embeddings that are jointly learned from a given database and ontology. We first define the geometric counterpart of an interpretation.

**Definition 1** (Geometric interpretation). Let  $\mathfrak{R}$  be a set of relation names and  $\mathbf{X} \subseteq C \cup N$  be a set of objects. An  $m$ -dimensional geometric interpretation  $\eta$  of  $(\mathfrak{R}, \mathbf{X})$  assigns to each  $k$ -ary relation name  $R$  from  $\mathfrak{R}$  a region  $\eta(R) \subseteq \mathbb{R}^{k \cdot m}$  and to each object  $o$  from  $\mathbf{X}$  a vector  $\eta(o) \in \mathbb{R}^m$ .

An example of a 1-dimensional geometric interpretation of  $(\{\text{Husband}, \text{Wife}, \text{Married}\}, \{p, q\})$  is depicted in Figure 2. Note that in this case, the unary predicates *Husband* and *Wife* are represented as intervals, whereas the binary predicate *Married* is represented as a convex polygon in  $\mathbb{R}^2$ . We now define what it means for a geometric interpretation to satisfy a ground atom.

**Definition 2** (Satisfaction of ground atoms). Let  $\eta$  be an  $m$ -dimensional geometric interpretation of  $(\mathfrak{R}, \mathbf{X})$ ,  $R \in \mathfrak{R}_k$  and  $o_1, \dots, o_k \in \mathbf{X}$ . We say that  $\eta$  satisfies a ground atom  $R(o_1, \dots, o_k)$ , written  $\eta \models R(o_1, \dots, o_k)$ , if  $\eta(o_1) \oplus \dots \oplus \eta(o_k) \in \eta(R)$ .

For  $\mathbf{Y} \subseteq \mathbf{X}$ , we will write  $\phi(\mathbf{Y}, \eta)$  for the set of ground atoms over  $\mathbf{Y}$  which are satisfied by  $\eta$ , i.e.:

$$\{R(o_1, \dots, o_k) \mid R \in \mathfrak{R}_k, o_1, \dots, o_k \in \mathbf{Y}, \eta \models R(o_1, \dots, o_k)\}$$

If  $\mathbf{Y} = \mathbf{X}$ , we also abbreviate  $\phi(\mathbf{Y}, \eta)$  as  $\phi(\eta)$ . For example, if  $\eta$  is the geometric interpretation from Figure 2, we find:

$$\phi(\eta) = \{\text{Husband}(p), \text{Wife}(p), \text{Married}(p, q), \text{Married}(q, p)\}$$

The notion of satisfaction in Definition 2 can be extended to propositional combinations of ground atoms in the usual way. Specifically,  $\eta$  satisfies a rule  $B_1 \wedge \dots \wedge B_n \rightarrow C$ , with  $B_1, \dots, B_n, C$  ground atoms, if  $\eta \models C$  or  $\{B_1, \dots, B_n\} \not\subseteq \phi(\eta)$ . Now consider the case of a non-ground rule, e.g.:

$$R(X, Y) \wedge S(Y, Z) \rightarrow T(X, Z) \quad (7)$$

Intuitively what we want to encode is whether  $\eta$  satisfies every possible grounding of this rule, i.e. whether for any objects  $o_x, o_y, o_z$  such that  $\eta(o_x) \oplus \eta(o_y) \in \eta(R)$  and  $\eta(o_x) \oplus \eta(o_z) \in \eta(S)$  it holds that  $\eta(o_x) \oplus \eta(o_z) \in \eta(T)$ . However, since an important aim of vector space representations is to enable inductive generalizations, this property of  $\eta$  should not only hold for the constants occurring in the given knowledge base, but also for any possible constants whose representation we might learn from external sources (Zhong et al. 2015; Xie et al. 2016; Xiao et al. 2017). As a result, we need to impose the following stronger requirement for  $\eta$  to satisfy (7): for every  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^m$  such that  $\mathbf{x} \oplus \mathbf{y} \in \eta(R)$  and  $\mathbf{y} \oplus \mathbf{z} \in \eta(S)$ , it has to hold that  $\mathbf{x} \oplus \mathbf{z} \in \eta(T)$ . Note that a rule like (7) thus naturally translates into a spatial constraint on the representation of the relation names. Finally, let us consider an existential rule:

$$R(X, Y) \rightarrow \exists Z . S(X, Y, Z) \quad (8)$$

For  $\eta$  to be a model of this rule, we require that for every  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$  such that  $\mathbf{x} \oplus \mathbf{y} \in \eta(R)$  there has to exist a  $\mathbf{z} \in \mathbb{R}^m$  such that  $\mathbf{x} \oplus \mathbf{y} \oplus \mathbf{z} \in \eta(S)$ . These intuitions are formalized in the following definition of a geometric model.

**Definition 3.** Let  $\mathcal{K} = (\Sigma, D)$  be a knowledge base and  $\mathfrak{O}$  a (possibly infinite) set of objects. A geometric interpretation  $\eta$  of  $(\mathfrak{R}(\mathcal{K}), \mathfrak{O})$  is called an  $m$ -dimensional geometric model of  $\mathcal{K}$  if

1.  $\phi(\eta) = \mathcal{M}$ , for some model  $\mathcal{M}$  of  $\mathcal{K}$ , and
2. for any set of points  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq \mathbb{R}^m$ ,  $\eta$  can be extended to a geometric interpretation  $\eta^*$  such that
  - (a) for each  $i \in \{1, \dots, n\}$  there is a fresh constant  $c_i \in \mathfrak{C} \setminus \mathfrak{O}(\mathcal{M})$  such that  $\eta^*(c_i) = \mathbf{v}_i$ ,
  - (b)  $\phi(\eta^*) = \mathcal{M}'$  for some model  $\mathcal{M}'$  of  $(\Sigma, D \cup \phi(\eta^*))$ .

The first point in Definition 3 ensures that we can view geometric models as geometric representations of classical models. The second point in Definition 3 ensures that we can use geometric models to introduce objects from external sources, without introducing any inconsistencies. It captures the fact that the logical dependencies between the relation names encoded in  $\Sigma$  should be properly captured by the spatial relationships between their geometric representations, as was illustrated in (8). Naturally,  $\mathcal{M}'$  might contain additional (in comparison to  $\mathcal{M}$ ) nulls to witness existential demands over the new constants. For datalog programs, however,  $\eta^*$  is completely determined by  $\eta$  and the fact that  $\eta^*(c_i) = \mathbf{v}_i$  for  $1 \leq i \leq n$ , that is, only Conditions 1 is necessary. For instance, the geometric interpretation depicted in Figure 2 is a geometric model of the rules from Example 1.

**Practical Significance of Geometric Models.** The framework presented in this section offers several key advantages over standard KG embedding methods. First, it allows us

to take into account a given ontology when learning the vector space representations, which should lead to higher-quality representations, and thus more faithful predictions, in cases where such an ontology is available. Also note that the region based framework can be applied to relations of any arity. Conversely, the framework also naturally allows us to obtain plausible rules from a learned geometric model, as this geometric model may (approximately) satisfy rules which are not entailed by the given ontology. Moreover, our framework allows for a tight integration of deductive and inductive modes of inference, as the facts and rules that are satisfied by a geometric model are deductively closed and logically consistent.

**Modelling Relations as Convex Regions.** While, in principle, arbitrary subsets of  $\mathbb{R}^{k \cdot m}$  can be used for representing  $k$ -ary relations, in practice the type of considered regions will need to be restricted in some way. This is needed to ensure that the regions can be efficiently learned from data and can be represented compactly. Moreover, the purpose of using vector space representations is to enable inductive inferences, but this is only possible if we impose sufficiently strong regularity conditions on the representations. For this reason, in this paper we will particularly focus on *convex* geometric interpretations, i.e. geometric interpretations in which each relation is represented using a convex region. While this may seem like a strong assumption, the vast majority of existing KG embedding models in fact learn representations that correspond to such convex geometric interpretations. Moreover, when learning regions in high-dimensional spaces, strong assumptions such as convexity are needed to avoid overfitting, especially if the amount of training data is limited. Finally, the use of convex regions is also in accordance with cognitive models such as conceptual spaces (Gärdenfors 2000), and more broadly with experimental findings in psychology, especially in cases where we are presented with few training examples (Rosseel 2002).

One may wonder whether it is possible to go further and restrict attention e.g. to convex models that are induced by vector translations. For instance, we could consider regions which are such that  $\mathbf{x} \oplus \mathbf{y} \in \eta(R)$  means that we also have  $\mathbf{u} \oplus \mathbf{v} \in \eta(R)$  whenever  $\mathbf{y} - \mathbf{x} = \mathbf{v} - \mathbf{u}$ , i.e. only the vector difference between  $\mathbf{x}$  and  $\mathbf{y}$  matters. Note that TransE and most of its generalizations aim to learn representations that correspond to such regions. Alas, as the next example illustrates, such translation-based regions do not have the desired generality, in the sense that they cannot properly capture even simple rules.

**Example 2.** For instance, consider rules (5)-(6) in Example 1. For the ease of presentation, let us write  $C_H$  for  $\eta(\text{Husband})$  and  $C_W$  for  $\eta(\text{Wife})$ , i.e. we assume that  $\text{Husband}(a)$  holds for a constant  $a$  iff  $a \in C_H$ . Let us furthermore assume that  $C_H$  and  $C_W$  are convex. We will also assume that a translation-based region is used to represent  $\text{Married}$ . Note that in such a case, the region  $\eta(\text{Married})$  in  $\mathbb{R}^{2n}$  can be characterized by a region  $C_M$  in  $\mathbb{R}^n$  such that  $\text{Married}(a, b)$  holds iff  $b - a \in C_M$ . To capture the logical dependencies encoded by the rules, the following spatial

relationships would then have to hold:

$$C_H \supseteq \{\mathbf{p} + \mathbf{r} \mid \mathbf{p} \in C_W, \mathbf{r} \in C_M\} \quad (9)$$

$$C_W \subseteq \{\mathbf{p} + \mathbf{r} \mid \mathbf{p} \in C_H, \mathbf{r} \in C_M\} \quad (10)$$

However, (9) and (10) entail<sup>1</sup> that  $C_W \subseteq C_H$ . Since, by rule (4), the concepts Wife and Husband are disjoint, we would have to choose  $C_W = C_H = \emptyset$  and would not be able to represent any instances of these concepts.

It is perhaps not surprising that translation based representations are not suitable for modelling rules, since they are already known not to be fully expressive in the sense of (1)–(2). As we discussed in Section 2.1, there are several bilinear models which are known to be fully expressive, and which may thus be thought of as more promising candidates for defining suitable types of regions. We address whether bilinear models are able to represent ontologies in the next section.

## 4 Limitations of Bilinear Models

As already mentioned, translation based approaches incur rather severe limitations on the kinds of databases and ontologies that can be modelled. In this section, we show that while bilinear models are fully expressive, and can thus model any database, they are not suitable for modelling ontologies. This motivates the need for novel embedding methods, which are better suited at modelling ontologies; this will be the focus of the next section.

Let us consider the following common type of rules:

$$R(X, Y) \rightarrow S(X, Y) \quad (11)$$

and a bilinear model in which each relation name  $R$  is associated with an  $n \times n$  matrix  $M_r$  and a threshold  $\lambda_r$ . We then say that (11) is satisfied if for each  $\mathbf{e}, \mathbf{f} \in \mathbb{R}^n$ , it holds that:

$$(\mathbf{e}^T M_r \mathbf{f} \geq \lambda_r) \Rightarrow (\mathbf{e}^T M_s \mathbf{f} \geq \lambda_s) \quad (12)$$

where  $\mathbf{e}^T$  denotes the transpose of  $\mathbf{e}$ . It turns out that bilinear models are severely limited in how they can model sets of rules of the form (11). This limitation stems from the following result.

**Proposition 1.** Suppose that (12) is satisfied for the matrices  $M_r, M_s$  and some thresholds  $\lambda_r, \lambda_s$ . Then there exists some  $\alpha \geq 0$  such that  $M_r = \alpha M_s$ .

If  $\alpha = 0$  then the rule (11) must be satisfied trivially, in the sense that the following rule is also satisfied for the matrix  $M_s$  and threshold  $\lambda_s$ :

$$\top \rightarrow S(X, Y)$$

Let us consider the case where  $\alpha > 0$ . Note that for the thresholds  $\lambda_r$  and  $\lambda_s$  we only need to consider the values -1 and 1 since other thresholds can always be simulated by

<sup>1</sup>Indeed, suppose that  $\mathbf{q} \in C_W$ , then by (10) there must exist some  $\mathbf{p} \in C_H$  and  $\mathbf{r} \in C_M$  such that  $\mathbf{q} = \mathbf{p} + \mathbf{r}$ . By (9) we furthermore have  $\mathbf{q} + \mathbf{r} \in C_H$ . Since  $\mathbf{q}$  is between  $\mathbf{p}$  and  $\mathbf{q} + \mathbf{r}$ , both of which belong to  $C_H$ , by the convexity of  $C_H$  it follows that  $\mathbf{q} \in C_H$ .

rescaling the matrices  $M_r$  and  $M_s$ . Now assume that the following rules are given:

$$R_1(X, Y) \rightarrow S(X, Y)$$

...

$$R_k(X, Y) \rightarrow S(X, Y)$$

By Proposition 1, we know that for  $i \in \{1, \dots, k\}$  there is some  $\alpha_i$  such that  $M_{r_i} = \alpha_i M_{s_i}$ . If  $\lambda_{r_i} = \lambda_{r_j}$  we thus have that either the rule  $R_i(X, Y) \rightarrow R_j(X, Y)$  or the rule  $R_j(X, Y) \rightarrow R_i(X, Y)$  is satisfied (depending on whether  $\alpha_i \geq 1$  and on whether  $\lambda_{r_i}$  is 1 or -1). This means in particular that we can always find two rankings  $R_{\tau_1}, \dots, R_{\tau_p}$  and  $R_{\sigma_1}, \dots, R_{\sigma_q}$  such that  $\{R_1, \dots, R_k\} = \{R_{\tau_1}, \dots, R_{\tau_p}, R_{\sigma_1}, \dots, R_{\sigma_q}\}$  and:

$$\forall 1 \leq i < p. R_{\tau_i}(X, Y) \rightarrow R_{\tau_{i+1}}(X, Y)$$

$$\forall 1 \leq i < q. R_{\sigma_i}(X, Y) \rightarrow R_{\sigma_{i+1}}(X, Y)$$

This clearly puts drastic restrictions on the type of subsumption hierarchies that can be modelled using bilinear models. Moreover, these limitations carry over to DistMult and ComplEx, as these are particular types of bilinear models. Due to the close links between DistMult and SimplE, it is also easy to see that the latter model has the same limitations.

In fact, the use of different vectors for head and tail mentions of entities in the SimplE model leads to even further limitations. To illustrate this, let us consider a rule of the following form:

$$R(X, Y) \wedge S(Y, Z) \rightarrow T(X, Z) \quad (13)$$

where we say that the SimplE representation defined by the vectors  $\mathbf{r}, \mathbf{ri}, \mathbf{s}, \mathbf{si}, \mathbf{t}, \mathbf{ti}$  and corresponding thresholds  $\lambda_r, \lambda_{ri}, \lambda_s, \lambda_{si}, \lambda_t, \lambda_{ti}$  satisfies (13) if for all entity vectors  $\mathbf{e}_h, \mathbf{e}_t, \mathbf{f}_h, \mathbf{f}_t, \mathbf{g}_h, \mathbf{g}_t$  it holds that:

$$\langle \mathbf{e}_h, \mathbf{r}, \mathbf{f}_t \rangle \geq \lambda_r \wedge \langle \mathbf{f}_h, \mathbf{ri}, \mathbf{e}_t \rangle \geq \lambda_{ri} \quad (14)$$

$$\wedge \langle \mathbf{f}_h, \mathbf{s}, \mathbf{g}_t \rangle \geq \lambda_s \wedge \langle \mathbf{g}_h, \mathbf{si}, \mathbf{f}_t \rangle \geq \lambda_{si}$$

$$\Rightarrow \langle \mathbf{e}_h, \mathbf{t}, \mathbf{g}_t \rangle \geq \lambda_t \wedge \langle \mathbf{g}_h, \mathbf{ti}, \mathbf{e}_t \rangle \geq \lambda_{ti}$$

Then we can show the following result.

**Proposition 2.** Suppose  $\mathbf{r}, \mathbf{ri}, \mathbf{s}, \mathbf{si}, \mathbf{t}, \mathbf{ti}$  and  $\lambda_r, \lambda_{ri}, \lambda_s, \lambda_{si}, \lambda_t, \lambda_{ti}$  define a SimplE representation satisfying (13). Then one of the following two rules is satisfied as well:

$$R(X, Y) \wedge S(Y, Z) \rightarrow \perp \quad (15)$$

$$\top \rightarrow T(X, Z) \quad (16)$$

## 5 Relations as Arbitrary Convex Regions

In this section we consider arbitrary convex geometric models, and show that they can correctly represent a large class of existential rules. We particularly show that KBs  $\mathcal{K}$  based on *quasi-chained rules* are properly captured by convex geometric models, in the sense that for each finite model  $\mathcal{I}$  of  $\mathcal{K}$ , there exists a convex geometric model  $\eta$  such that  $\mathcal{I} = \phi(\eta)$ .

**Quasi-chained Rules.** We say that an existential rule  $\sigma$ , defined as in (3) above, is *quasi-chained (QC)* if for all  $1 \leq i \leq n$

$$|(\text{vars}(B_1) \cup \dots \cup \text{vars}(B_{i-1})) \cap \text{vars}(B_i)| \leq 1$$

An ontology is quasi-chained if all its rules are either quasi-chained or quasi-chained negative constraints.

Note that quasi-chainedness is a natural and useful restriction. Quasi-chained rules are indeed closely related to the well-known chain-datalog fragment of datalog (Shmueli 1987; Ullman and Gelder 1988) in which important properties, e.g. reachability, are still expressible. Furthermore, prominent Horn description logics can be expressed using decidable fragments of quasi-chained existential rules. For example,  $\mathcal{ELH}\mathcal{I}$  ontologies<sup>2</sup> can be embedded into the *guarded fragment* (Calì, Gottlob, and Kifer 2013) of QC existential rules. Further, QC existential rules subsume *linear existential rules*, which only allow rule bodies that consist of a single atom and capture a  $k$ -ary extension of  $DL\text{-}Lite_{\mathcal{R}}$ .

We next show the announced result that geometric models properly capture quasi-chained ontologies.

**Proposition 3.** *Let  $\mathcal{K} = (\Sigma, D)$ , with  $\Sigma$  a quasi-chained ontology, and let  $\mathcal{M}$  be a finite model of  $\mathcal{K}$ . Then  $\mathcal{K}$  has a convex geometric model  $\eta$  such that  $\phi(\eta) = \mathcal{M}$ .*

To clarify the intuitions behind this proposition, we show how an  $m$ -dimensional geometric model  $\eta$  satisfying  $\phi(\eta) = \mathcal{M}$  can be constructed, where  $m = |\mathfrak{O}(\mathcal{M})|$ . Let  $x_1, \dots, x_m$  be an enumeration of the elements in  $\mathfrak{O}(\mathcal{M})$ , then for each  $x_i$ ,  $\eta(x_i)$  is defined as the vector in  $\mathbb{R}^m$  with value 1 in the  $i^{th}$  coordinate and 0 in all others. Further, for each  $R \in \mathfrak{R}_k(\mathcal{K})$ , we define  $\eta(R)$  as follows, where CH denotes the convex-hull:

$$\eta(R) = \text{CH}\{\eta(y_1) \oplus \dots \oplus \eta(y_k) \mid R(y_1, \dots, y_k) \in \mathcal{M}\} \quad (17)$$

A proof that  $\phi(\eta) = \mathcal{M}$ , and that  $\eta$  satisfies Conditions 1 and 2 from Definition 3, is provided in the appendix.

For the next corollary we assume that the quasi-chained ontology  $\Sigma$  belongs to fragments enjoying the *finite model property (FMP)*, i.e. if a KB  $\mathcal{K}$  is satisfiable, it has a finite model, e.g. where  $\Sigma$  is weakly-acyclic (Fagin et al. 2005), guarded, linear, or a quasi-chained datalog program. The following then is a direct consequence of Proposition 3.

**Corollary 1.** *Let  $\mathcal{K} = (\Sigma, D)$  with  $\Sigma$  as above. It holds that  $\mathcal{K}$  is satisfiable iff  $\mathcal{K}$  has a convex geometric model.*

Intuitively, we require logics enjoying the FMP since the construction in the proof of Proposition 3 uses one dimension for each object that appears in a given model of the knowledge base. For ontologies expressed in fragments without the FMP, we can thus not guarantee the existence of an Euclidean model using this argument.

A natural question is whether there is a way of defining a convex  $n$ -dimensional geometric model for an  $n$  considerably smaller than  $m = |\mathfrak{O}(\mathcal{M})|$  for some model  $\mathcal{M}$ . For the case of datalog rules, where  $|\mathfrak{O}(\mathcal{M})| = |\mathbf{C}(\mathcal{K})|$ , it turns out that this is in general not possible.

**Proposition 4.** *For each  $n \in \mathbb{N}$ , there exists a KB  $\mathcal{K} = (\Sigma, D)$  with  $\Sigma$  a datalog program, over a signature with  $n$*

<sup>2</sup>We assume they are in a suitable normal form (Baader et al. 2017)

constants and  $n$  unary predicates such that  $\mathcal{K}$  does not have a convex geometric model in  $\mathbb{R}^m$  for  $m < n - 1$ .

To see this, consider the knowledge base  $\mathcal{K} = (\Sigma, D)$  with  $D = \{A_i(a_j) \mid 1 \leq i \neq j \leq n\}$ , for some  $n \in \mathbb{N}$ , and  $\Sigma$  consisting of the following rule

$$A_1(X) \wedge \dots \wedge A_n(X) \rightarrow \perp \quad (18)$$

It is clear that  $\mathcal{K}$  is satisfiable. Now, let  $\eta$  be an  $n - 2$  dimensional convex geometric model of  $\mathcal{K}$ . Clearly, for each  $a_j \in \mathbf{C}(\mathcal{K})$ , it holds that  $\eta(a_j) \in \bigcap_{i \neq j} \eta(A_i)$  and thus  $\bigcap_{i \neq j} \eta(A_i) \neq \emptyset$ . Using Helly's Theorem<sup>3</sup>, it follows that  $\bigcap_{i=1}^n \eta(A_i)$  contains some point  $p$ . Further, let  $\eta^*$  be the extension of  $\eta$  to  $\mathbf{C}(\mathcal{K}) \cup \{d\}$  defined by  $\eta^*(d) = p$ . Then  $\mathcal{K} \cup \phi(\eta^*)$  contains  $D \cup \{A_i(d) \mid i \in [1, n]\}$  which together with (18) implies that  $\mathcal{K} \cup \phi(\eta^*)$  does not have a convex model. Thus,  $\eta$  cannot be an  $n - 2$  dimensional convex geometric model  $\mathcal{K}$ , and the dimensionality of any convex model of  $\mathcal{K}$  has to be at least  $n - 1$ .

Note that the model  $\eta$  that we constructed above is  $m$ -dimensional, but the lower bound from Proposition 4 only states that at least  $m - 1$  dimensions are needed in general. In fact, it is easy to see that such an  $m - 1$ -dimensional convex geometric model indeed exists for datalog programs. In particular, let  $H$  be the hyperplane defined by  $H = \{(p_1, \dots, p_m) \mid p_1 + \dots + p_m = 1\}$  then clearly  $\eta(x_i) \in H$  for every constant  $x_i$  and  $\eta(R) \subseteq H \oplus \dots \oplus H$ . In other words, each  $\eta(x_i)$  is located in an  $m - 1$  dimensional space, and  $\eta(R)$  is a subset of an  $k \cdot (m - 1)$  dimensional space.

**Beyond Quasi-chained Rules.** The main remaining question is whether the restriction to QC rules is necessary. The next example illustrates that if a KB contains rules that do not satisfy this restriction, it may not be possible to construct a convex geometric model.

**Example 3.** Consider  $\Sigma$  consisting of the following rule:

$$R_1(X, Y) \wedge R_2(X, Y) \rightarrow \perp$$

and let  $D = \{R_1(a_1, a_1), R_1(a_2, a_2), R_2(a_1, a_2), R_2(a_2, a_1)\}$ . Then clearly  $\mathcal{M} = D$  is a model of the knowledge base  $(\Sigma, D)$ . Now suppose this KB had a convex geometric model  $\eta$ . Let  $\eta^*$  be an extension of  $\eta$  to the fresh constant  $b$ , defined by  $\eta^*(b) = 0.5\eta(a_1) + 0.5\eta(a_2)$ . Note that we then have:

$$\begin{aligned} \eta^*(b) \oplus \eta^*(b) &= 0.5(\eta(a_1) \oplus \eta(a_1)) + 0.5(\eta(a_2) \oplus \eta(a_2)) \\ &= 0.5(\eta(a_1) \oplus \eta(a_2) + 0.5(\eta(a_2) \oplus \eta(a_1))) \end{aligned}$$

and thus, by the convexity of  $\eta(R_1)$  and  $\eta(R_2)$ , it follows that  $\eta^* \models R_1(b, b) \wedge R_2(b, b)$ . This means that  $(\Sigma, D \cup \phi(\eta^*))$  does not have a model, which contradicts the assumption that  $\eta$  was a geometric model.

## 6 Extended Geometric Models

As shown in Section 5, there are knowledge bases which have a finite model but which do not have a convex geometric model. To deal with arbitrary knowledge bases, one possible approach is to simply drop the convexity requirement.

<sup>3</sup>This theorem states that if  $C_1, \dots, C_k$  are convex regions in  $\mathbb{R}^n$ , with  $k > n$ , and each  $n + 1$  among these regions have a non-empty intersection, it holds that  $\bigcap_{i=1}^k C_i \neq \emptyset$ .

In this section, we briefly explore another solution, based on the idea that for each relation symbol  $R \in \mathfrak{R}_k(\mathcal{K})$ , we can consider a function  $f_R$  which embeds  $k$ -tuples into another vector space. This can be formalized as follows

**Definition 4** (Extended convex geometric interpretation). *Let  $\mathfrak{R}$  be a set of relation names and  $\mathbf{X} \subseteq \mathbf{C} \cup \mathbf{N}$  be a set of objects. An  $m$ -dimensional extended convex geometric interpretation of  $(\mathfrak{R}, \mathbf{X})$  is a pair  $((f_R)_{R \in \mathfrak{R}}, \eta)$ , where for each  $R \in \mathfrak{R}_k$ ,  $f_R$  is a  $\mathbb{R}^{k \cdot m} \rightarrow \mathbb{R}^{l_R}$  mapping, for some  $l_R \in \mathbb{N}$ , and  $\eta$  assigns to each  $R \in \mathfrak{R}_k$  a convex region  $\eta(R)$  in  $\mathbb{R}^{l_R}$  and to each constant  $c$  from  $\mathbf{X}$  a vector  $\eta(c) \in \mathbb{R}^m$ .*

We can now adapt the definition of satisfaction of a ground atom as follows.

**Definition 5** (Satisfaction of ground atoms). *Let  $((f_R)_{R \in \mathfrak{R}}, \eta)$  be an extended convex geometric interpretation of  $(\mathfrak{R}, \mathbf{X})$ ,  $R \in \mathfrak{R}_k$  and  $o_1, \dots, o_k \in \mathbf{X}$ . We say that  $\eta$  satisfies a ground atom  $R(o_1, \dots, o_k)$ , written  $\eta \models R(o_1, \dots, o_k)$ , if  $f_R(\eta(o_1) \oplus \dots \oplus \eta(o_k)) \in \eta(R)$ .*

The notion of extended convex geometric model is then defined as in Definition 3, by simply using extended convex geometric models instead of (standard) geometric models.

Note that we almost trivially have that every knowledge base  $\mathcal{K} = (\Sigma, D)$  which has a finite model  $\mathcal{M}$  also has an extended convex geometric model. Indeed, to construct such a model, we can choose  $\eta$  for constants from  $\mathbf{X}$  arbitrarily, as long as  $\eta(o_1) \neq \eta(o_2)$  if  $o_1 \neq o_2$ . We can then define  $f_R$  as follows:  $f_R(\mathbf{x}) = 1$  if  $\mathcal{M}$  contains a ground atom  $R(o_1, \dots, o_k)$  such that  $\mathbf{x} = \eta(o_1) \oplus \dots \oplus \eta(o_k)$ , and  $f_R(\mathbf{x}) = 0$  otherwise. Finally we can define  $\eta(R) = \{1\}$ . It can be readily checked that the extended convex geometric interpretation which is constructed in this way is indeed an extended convex geometric model of  $\mathcal{K}$ .

The extended convex geometric model which is constructed in this way is uninteresting, however, as it does not allow us to use the geometric representations of the constants to induce any knowledge which is not already given in  $\mathcal{K}$ . Specifically, suppose  $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^m$  and let  $\eta^*$  be the extension of  $\eta$  to  $\mathbf{X} \cup \{o_1, \dots, o_n\}$ , then for  $o'_1, \dots, o'_k \in \mathbf{X} \cup \{o_1, \dots, o_n\}$  and  $R \in \mathfrak{R}_k$ , we have  $\eta^* \models R(o'_1, \dots, o'_k)$  iff  $\mathcal{M}$  contains some atom  $R(p_1, \dots, p_k)$  such that  $\eta(p_1) = \eta^*(o'_1), \dots, \eta(p_k) = \eta^*(o'_k)$ . This means that in practice, we need to impose some restrictions on the functions  $f_R$ . Note, however, that we cannot restrict  $f_R$  to be linear, as that would lead to the same restrictions as we encountered for standard convex geometric models. For instance, it is easy to verify that the knowledge base from Example 3 cannot have an extended geometric model in which  $f_{R_1}$  and  $f_{R_2}$  are linear.

One possible alternative would be to encode each function  $f_R$  as a neural network, but there are still several important open questions related to this choice. First, it is far from clear how we would then be able to check whether an extended convex geometric interpretation is a model of a given ontology. In contrast, for standard convex geometric interpretations, we can use standard linear programming techniques to check whether a given existential rule is satisfied. It is furthermore unclear which types of neural networks would be

needed to guarantee that all types of existential rules can be captured.

## 7 Related Work

Various approaches to KG completion have been proposed that are based on neural network architectures (Socher et al. 2013; Niepert 2016; Minervini et al. 2017). Interestingly, some of these approaches can be seen as special cases of the extended convex geometric models considered in Section 6. For example, in the E-MLP model (Socher et al. 2013), to predict whether  $(e, R, f)$  is a valid triple, the concatenation of the vectors  $e$  and  $f$  is fed into a two-layer neural network.

Instead of constructing tuple representations from entity embeddings, some authors have also considered approaches that directly learn a vector space embedding of entity tuples (Turney 2005; Riedel et al. 2013). For each relation  $R$  a vector  $r$  can then be learned such that the dot product  $r \cdot t$  reflects the likelihood that a tuple represented by  $t$  is an instance of  $R$ . This model does not put any a priori restrictions on the kind of relations that can be modeled, although it is clearly not suitable for modelling rules (e.g. it is easy to see that this model carries over the limitations of bilinear models). Moreover, as enough information needs to be available about each tuple, this strategy has primarily been used for modelling knowledge extracted from text, where representations of word-tuples are learned from sentences that contain these words.

Note that KG embedding methods model relations in a soft way: their associated scoring function can be used to rank ground facts according to their likelihood of being correct, but no attempt is made at modelling the exact extension of relations. This means that logical dependencies among relations cannot be modeled, which makes such representations fundamentally different from the geometric representations that we have considered in this paper. Nonetheless, some authors have used logical rules to improve the predictions that are made in a KG completion setting. For example, in (Wang, Wang, and Guo 2015), a mixed integer programming formulation is used to combine the predictions made from a given KG embedding with a set of hard rules. Specifically, the aim of this approach is to determine the most plausible set of facts which is logically consistent with the given rules. Another strategy, used in (Demeester, Rocktäschel, and Riedel 2016), is to incorporate background knowledge in the loss function of the learning problem. Specifically, the authors propose to take advantage of relation inclusions, i.e. rules of the form  $R(X, Y) \rightarrow S(X, Y)$ , for learning better tuple embeddings. The main underlying idea is to translate such a rule to the soft constraint that  $r \cdot t \leq s \cdot t$  should hold for each tuple  $t$ . This is imposed in an efficient way by restricting tuple embeddings to vectors with non-negative coordinates and then requiring that  $r_i \leq s_i$  for each coordinate  $r_i$  of  $r$  and corresponding coordinate  $s_i$  of  $s$ . However, this strategy cannot straightforwardly be generalized to other types of rules.

To overcome this shortcoming, neural network architectures dealing with arbitrary Datalog-like rules have been recently proposed (Niepert 2016; Minervini et al. 2017). Other related approaches include (Wang and Cohen 2016;

Rocktäschel and Riedel 2017; Sourek et al. 2017). However, such methods essentially use neural network methods to simulate deductive inference, but do not explicitly model the extension of relations, and do not allow for the tight integration of induction and deduction that our framework supports. Moreover, these methods are aimed at learning (soft versions of) first-order rules from data, rather than constraining embeddings based on a given set of (hard) rules.

Within KR research, (Hohenecker and Lukasiewicz 2017) recently made first steps towards the integration of ontological reasoning and deep learning, obtaining encouraging results. Indeed, the developed system was considerably faster than the state of the art RDFox (Nenov et al. 2015), while retaining high-accuracy. Initial results have also been obtained in the use of ontological reasoning to derive human-interpretable explanations from the output of a neural network (Sarker et al. 2017).

## 8 Conclusions and Future Work

We have argued that knowledge base embedding models should be capable of representing sufficiently expressive classes of rules, a property which, to the best of our knowledge, has not yet been considered in the literature. We found that the commonly used translation-based and bilinear models are prohibitively restrictive in this respect. In light of this, we argue that more work is needed to better understand how different kinds of rules can be geometrically represented. In this paper, we have initiated this analysis, by studying knowledge base embeddings in which relations are represented as convex regions in a space of tuples. These tuples are simply represented as concatenations of the vector representations of the individual arguments, and can thus be obtained using standard approaches for learning entity embeddings.

Our main finding is that using this convex-regions approach, knowledge bases that are restricted to the important class of quasi-chained existential rules can be faithfully encoded, in the sense that any set of facts which is induced using that vector space embedding is logically consistent and deductively closed with respect to the input ontology. Note that this is an essential requirement if we want to exploit symbolic knowledge when learning embeddings. For example, one common strategy is to encode (soft versions of) the given rules in the loss function, but for such a strategy to be successful, we should ensure that the considered representation is actually capable of satisfying the corresponding (soft) constraints. We thus believe this paper provides an important step towards a comprehensive integration of neural embeddings and KR technologies, laying important foundations to develop methods that combine deductive and inductive reasoning in a tighter way than current approaches.

As future work, the most important next step is to develop practical region-based embedding models. Allowing arbitrary polytopes would likely lead to overfitting, but we believe that by appropriately restricting the types of regions that are allowed and regularizing the embedding model in an appropriate way, it will be possible to make more accurate predictions than existing knowledge graph embedding models. For example, note that translation based models, as

well as bilinear models when restricted to positive coordinates, are special cases of region based models, so a natural approach would be to learn region based models that are regularized to stay close to these standard approaches. From a theoretical point of view, an important open problem is to characterize particular classes of extended convex geometric models that are sufficiently expressive to model arbitrary existential rules (or interesting sub-classes). Indeed, the non-linear representation from Section 6 is too general to be practically useful, and we therefore need to characterize what types of knowledge bases can be captured by different kinds of simple neural network architectures. Finally, it would be interesting to extend our framework to model recently introduced ontology languages especially tailored for KGs (Krötzsch et al. 2017), which include means for representing annotations on data and relations.

## Acknowledgments

The authors were supported by EU’s Horizon 2020 programme under the Marie Skłodowska-Curie grant 663830 and ERC Starting Grant 637277 ‘FLEXILOG’, respectively.

## References

- Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.
- Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1247–1250.
- Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *Proc. NIPS*, 2787–2795.
- Calì, A.; Gottlob, G.; and Kifer, M. 2013. Taming the infinite chase: Query answering under expressive relational constraints. *J. Artif. Intell. Res.* 48:115–174.
- Camacho-Collados, J.; Pilehvar, M. T.; and Navigli, R. 2016. Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence* 240:36–64.
- Carlson, A.; Betteridge, J.; Kisiel, B.; Settles, B.; Hruschka Jr., E. R.; and Mitchell, T. M. 2010. Toward an architecture for never-ending language learning. In *Proc. AAAI*, 1306–1313.
- Demeester, T.; Rocktäschel, T.; and Riedel, S. 2016. Lifted rule injection for relation embeddings. In *Proc. EMNLP*, 1389–1399.
- Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmann, T.; Sun, S.; and Zhang, W. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *SIGKDD*, 601–610.
- Fagin, R.; Kolaitis, P. G.; Miller, R. J.; and Popa, L. 2005. Data exchange: semantics and query answering. *Theor. Comput. Sci.* 336(1):89–124.

- Gärdenfors, P. 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Gardner, M., and Mitchell, T. M. 2015. Efficient and expressive knowledge base completion using subgraph feature extraction. In *Proc. of EMNLP-15*, 1488–1498.
- Hohenecker, P., and Lukasiewicz, T. 2017. Deep learning for ontology reasoning. *arXiv preprint arxiv:1705.10342*.
- Kazemi, S. M., and Poole, D. 2018. SimplE embedding for link prediction in knowledge graphs. *arXiv preprint arXiv:1802.04868*.
- Krötzsche, M.; Marx, M.; Ozaki, A.; and Thost, V. 2017. Attributed description logics: Ontologies for knowledge graphs. In *Proc. of ISWC-17*.
- Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, 2181–2187.
- Miller, G. A. 1995. Wordnet: a lexical database for english. *Communications of the ACM* 38:39–41.
- Minervini, P.; Demeester, T.; Rocktäschel, T.; and Riedel, S. 2017. Adversarial sets for regularising neural link predictors. In *Proc. of UAI-17*.
- Nenov, Y.; Piro, R.; Motik, B.; Horrocks, I.; Wu, Z.; and Banerjee, J. 2015. RDFox: A highly-scalable RDF store. In *Proc. of ISWC-15*, 3–20.
- Nguyen, D. Q.; Sirts, K.; Qu, L.; and Johnson, M. 2016. STransE: a novel embedding model of entities and relationships in knowledge bases. In *Proc. of NAACL-HLT*, 460–466.
- Nickel, M.; Tresp, V.; and Kriegel, H.-P. 2011. A three-way model for collective learning on multi-relational data. In *Proc. ICML*, 809–816.
- Niepert, M. 2016. Discriminative gaifman models. In *Proc. of NIPS-16*, 3405–3413.
- Riedel, S.; Yao, L.; McCallum, A.; and Marlin, B. M. 2013. Relation extraction with matrix factorization and universal schemas. In *Proc. HLT-NAACL*, 74–84.
- Rocktäschel, T., and Riedel, S. 2017. End-to-end differentiable proving. In *Proc. NIPS*, 3791–3803.
- Rosseel, Y. 2002. Mixture models of categorization. *Journal of Mathematical Psychology* 46(2):178 – 210.
- Sarker, M. K.; Xie, N.; Doran, D.; Raymer, M.; and Hitzler, P. 2017. Explaining trained neural networks with semantic web technologies: First steps. In *Proc. of NeSy-17*.
- Shmueli, O. 1987. Decidability and expressiveness of logic queries. In *Proc. of PODS-87*, 237–249.
- Socher, R.; Chen, D.; Manning, C. D.; and Ng, A. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Proc. NIPS*, 926–934.
- Sourek, G.; Svatos, M.; Zelezný, F.; Schockaert, S.; and Kuzelka, O. 2017. Stacked structure learning for lifted relational neural networks. In *Proc. ILP*, 140–151.
- Speer, R.; Chin, J.; and Havasi, C. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proc. AAAI*, 4444–4451.
- Toutanova, K.; Chen, D.; Pantel, P.; Poon, H.; Choudhury, P.; and Gamon, M. 2015. Representing text for joint embedding of text and knowledge bases. In *Proc. of EMNLP-15*, 1499–1509.
- Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; and Bouchard, G. 2016. Complex embeddings for simple link prediction. In *Proc. ICML*, 2071–2080.
- Turney, P. D. 2005. Measuring semantic similarity by latent relational analysis. In *Proc. IJCAI*, 1136–1141.
- Ullman, J. D., and Gelder, A. V. 1988. Parallel complexity of logical query programs. *Algorithmica* 3:5–42.
- Vrandečić, D., and Krötzsch, M. 2014. Wikidata: a free collaborative knowledge base. *Communications of the ACM* 57:78–85.
- Wang, W. Y., and Cohen, W. W. 2016. Learning first-order logic embeddings via matrix factorization. In *Proc. of IJCAI-16*, 2132–2138.
- Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, 1112–1119.
- Wang, Q.; Mao, Z.; Wang, B.; and Guo, L. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* 29(12):2724–2743.
- Wang, Q.; Wang, B.; and Guo, L. 2015. Knowledge base completion using embeddings and rules. In *Proc. IJCAI*, 1859–1866.
- Xiao, H.; Huang, M.; Meng, L.; and Zhu, X. 2017. Ssp: Semantic space projection for knowledge graph embedding with text descriptions. In *Proc. AAAI*, volume 17, 3104–3110.
- Xie, R.; Liu, Z.; Jia, J.; Luan, H.; and Sun, M. 2016. Representation learning of knowledge graphs with entity descriptions. In *Proc. of AAAI*, 2659–2665.
- Yang, B.; Yih, W.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proc. of ICLR-15*.
- Zhong, H.; Zhang, J.; Wang, Z.; Wan, H.; and Chen, Z. 2015. Aligning knowledge and text embeddings by entity descriptions. In *EMNLP*, 267–272.