Generalizing Knowledge Graph Embedding with Universal Orthogonal Parameterization

Rui Li [1] **Chaozhuo Li** [2] **Yanming Shen** [3] **Zeyu Zhang** [1] **Xu Chen** [1]

Abstract

Recent advances in knowledge graph embedding (KGE) rely on Euclidean/hyperbolic orthogonal relation transformations to model intrinsic logical patterns and topological structures. However, existing approaches are confined to rigid relational orthogonalization with *restricted dimension* and *homogeneous geometry*, leading to deficient modeling capability. In this work, we move beyond these approaches in terms of both dimension and geometry by introducing a powerful framework named GoldE, which features a *universal* orthogonal parameterization based on a generalized form of Householder reflection. Such parameterization can naturally achieve *dimensional extension* and *geometric unification* with theoretical guarantees, enabling our framework to simultaneously capture crucial logical patterns and inherent topological heterogeneity of knowledge graphs. Empirically, GoldE achieves state-of-the-art performance on three standard benchmarks. Codes are available at `https://github.com/xxrep/GoldE`.

1. Introduction

Knowledge graphs (KGs) structurally organize vast human knowledge into the factual triples, where each triple (h, r, t) signifies the existence of a relation r between head entity h and tail entity t. Imbued with a wealth of factual knowledge, KGs have facilitated a myriad of downstream applications (Saxena et al., 2020; Yasunaga et al., 2022; Sun et al., 2023). However, real-world KGs such as Freebase (Bollacker et al., 2008) are plagued by incompleteness (Bordes et al., 2013). This motivates substantial research on Knowledge Graph Embedding (KGE), which learns expressive representations of entities and relations for predicting the missing links.

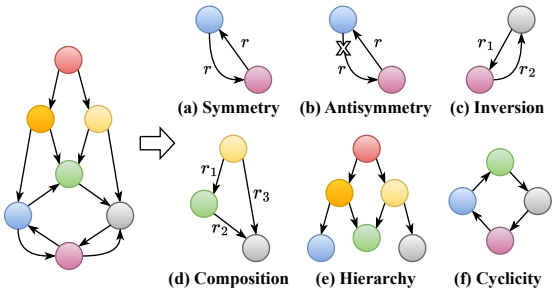


Figure 1. Illustrations of logical patterns (a-d) (Sun et al., 2019) and topological structures (e-f) (Chami et al., 2020).

The key to KGE lies in capturing crucial *logical patterns* (e.g., symmetry, antisymmetry, inversion and composition) (Sun et al., 2019) and inherent *topological structures* (e.g., cyclicity and hierarchy) (Chami et al., 2020) as illustrated in Figure 1. Recent advances highlight the effectiveness of *orthogonal relation transformations* for this principle. As a pioneering work, RotatE (Sun et al., 2019) represents relations as 2-dimensional Euclidean rotations, i.e., isometries on 1-spheres (Sommer, 2013), which is able to model the four logical patterns and cyclical structures. The subsequent works can be categorized into two branches: (1) "*Dimension Branch*" (Zhang et al., 2019; Gao et al., 2020; Cao et al., 2021; Li et al., 2022) extends the (Euclidean) orthogonal relation transformations into higher-dimensional spaces for better modeling capacity; (2) "*Geometry Branch*" (Balazevic et al., 2019; Chami et al., 2020) leverages hyperbolic isometries to preserve the hierarchical structures.

However, *it takes two to tango*—none of existing approaches can simultaneously break the restrictions of dimension and geometry for orthogonal relation transformations. As shown in Table 1, the dimensional extensions are entrenched in Euclidean geometry, while the hyperbolic relational orthogonalization remains constrained in low-dimensional spaces due to the significant computational demand and complexity. Moreover, all these approaches are specifically designed on homogeneous geometry, which is inadequate to preserve the topological heterogeneity of KGs—in some regions cyclical, in others hierarchical (Gu et al., 2019; Skopek et al., 2020; Wang et al., 2021). In light of such limitations, a challenging

[1]Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China [2]Key Laboratory of Trustworthy Distributed Computing and Service (MOE), Beijing University of Posts and Telecommunications, China [3]School of Computer Science and Technology, Dalian University of Technology, China. Correspondence to: Xu Chen <xu.chen@ruc.edu.cn>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Table 1. Recent models' capability of modeling logical patterns and topological structures of KGs. $\mathbb{E}$, $\mathbb{P}$ and $\mathbb{Q}$ denote Euclidean, elliptic and hyperbolic geometries for orthogonal relation transformations, respectively. Note that Euclidean models essentially act by isometries on hyperspheres (Sommer, 2013), which can be viewed as special cases of our elliptic parameterization.

Model	Geometry of Orth.	Dimension of Orth.	Logical Patterns				Topological Structures	
			Symmetry	Antisymmetry	Inversion	Composition	Cyclicity	Hierarchy
RotatE (Sun et al., 2019)	$\mathbb{E}$	2	✔	✔	✔	✔	✔	✗
Rotate3D (Gao et al., 2020)	$\mathbb{E}$	3	✔	✔	✔	✔	✔	✗
DualE (Cao et al., 2021)	$\mathbb{E}$	3	✔	✔	✔	✗	✔	✗
QuatE (Zhang et al., 2019)	$\mathbb{E}$	4	✔	✔	✔	✗	✔	✗
HousE (Li et al., 2022)	$\mathbb{E}$	k	✔	✔	✔	✔	✔	✗
RefH (Chami et al., 2020)	$\mathbb{Q}$	2	✔	✔	✔	✔	✗	✔
RotH (Chami et al., 2020)	$\mathbb{Q}$	2	✔	✔	✔	✔	✗	✔
AttH (Chami et al., 2020)	$\mathbb{Q}$	2	✔	✔	✔	✔	✗	✔
GoldE	$\mathbb{P}(\mathbb{E}), \mathbb{Q}$	k	✔	✔	✔	✔	✔	✔

question arises: *can we generalize these approaches in both dimension and geometry to achieve the best of all worlds?*

In this paper, we give an affirmative answer by presenting a general framework named GoldE, which represents relations with a *universal orthogonal parameterization*. To establish the universality, we first derive a generalized form of Householder reflection (Householder, 1958) based on a quadratic inner product. By treating such reflection as the elementary operator, we can then theoretically design elegant mappings to parameterize orthogonal relation transformations of typical geometric types (i.e., Euclidean, elliptic and hyperbolic). Each of them breaks through the dimensional rigidity without losing degrees of freedom, enabling superior capacity for modeling logical patterns and corresponding topologies. Considering the inherent topological heterogeneity of KGs, we further integrate the designed mappings within a product manifold (Ficken, 1939) to unify multiple geometric types of orthogonal transformations. In this way, such parameterization can simultaneously achieve dimensional extension and geometric unification for relational orthogonalization, empowering our framework to better match heterogeneous topologies and thus provide higher-quality representations.

Contributions: To the best of our knowledge, GoldE is the first framework that generalizes existing KGE approaches *in both dimension and geometry* of the orthogonal relation transformations. With theoretical guarantees, we establish a *universal orthogonal parameterization*, based on which our GoldE can *simultaneously capture crucial logical patterns and inherent topological heterogeneity* as shown in Table 1. Empirically, we conduct extensive experiments over three standard benchmarks, and GoldE consistently outperforms current state-of-the-art baselines across all the datasets.

2. Preliminaries

2.1. Problem Setup

Given the entity set $\mathcal{V}$ and relation set $\mathcal{R}$, a knowledge graph can be formally defined as a collection of factual triples $\mathcal{F} = \{(h, r, t)\} \subseteq \mathcal{V} \times \mathcal{R} \times \mathcal{V}$, where each triple represents that there is a relation r between head entity h and tail entity t. As an effective technique for automatically inferring missing links, KGE encodes entities and relations into expressive representations, and measures the plausibility of each triple with a pre-defined score function.

2.2. Geometric Background

We briefly introduce some necessary background on elliptic and hyperbolic geometry. More relevant details are available in the standard texts (Do Carmo, 1992; Eisenhart, 1997).

Elliptic Manifold: Let $\mathbf{p} \in \mathbb{R}^k_+$ be a weighting vector where all elements are positive. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, the elliptic inner product (Özdemir, 2016) is defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{p}} = \mathbf{x}^\top \mathrm{diag}(\mathbf{p}) \mathbf{y} = p_1 x_1 y_1 + \cdots + p_k x_k y_k.$$

For any $\beta > 0$, the elliptic manifold (ellipsoid) $\mathbb{P}^k_{\mathbf{p},\beta}$ is the following submanifold in $\mathbb{R}^k$:

$$\mathbb{P}^k_{\mathbf{p},\beta} = \{\mathbf{x} \in \mathbb{R}^k : \langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{p}} = \beta\}. \tag{1}$$

If one set $\mathbf{p}$ to an all-ones vector $\mathbf{1}$, $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{p}} = \sum_{i=1}^k x_i y_i$ is the Euclidean inner product, based on which the ellipsoid $\mathbb{P}^k_{\mathbf{p},\beta}$ is specialized into the hypersphere $\mathbb{S}^k_\beta$.

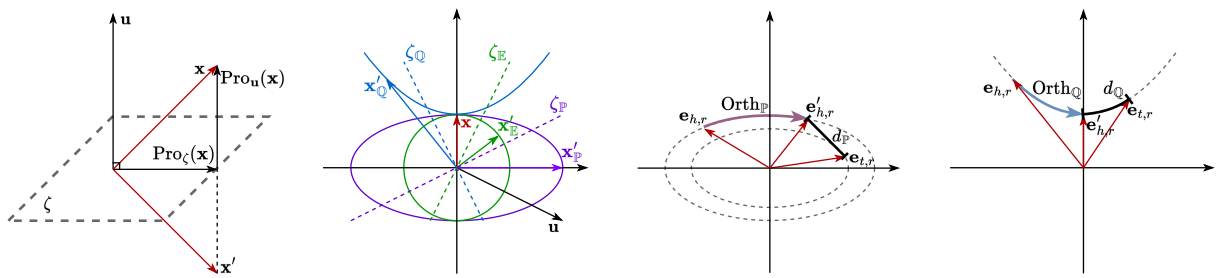
Hyperbolic Manifold: This paper chooses the hyperboloid model of hyperbolic space for its simplicity and numerical stability (Nickel & Kiela, 2018). Let $\mathbf{q} \in \mathbb{R}^k$ be a weighting vector with the first element equal to -1 and the remaining elements equal to $+1$. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, the hyperbolic (Lorentz) inner product is defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{q}} = \mathbf{x}^\top \mathrm{diag}(\mathbf{q}) \mathbf{y} = -x_1 y_1 + x_2 y_2 + \cdots + x_k y_k.$$

For any $\beta > 0$, the hyperboloid $\mathbb{Q}^k_\beta$ is denoted as:

$$\mathbb{Q}^k_\beta = \{\mathbf{x} \in \mathbb{R}^k : \langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{q}} = -\beta, x_1 > 0\}. \tag{2}$$

For the sake of clarification, all notations used in this paper are listed in Appendix A.



(a) (b) (c) (d)

Figure 2. (a) Illustration of generalized Householder reflection; (b) Three types of reflections of $\mathbf{x}$ about the hyperplanes ζ orthogonal to $\mathbf{u}$ under different weighting vectors, i.e., Euclidean (green lines), elliptic (purple lines) and hyperbolic (blue lines). Note that all quantities are displayed in the uniformly weighted Euclidean space, thus the non-Euclidean hyperplanes $\zeta_{\mathbb{P}}$ and $\zeta_{\mathbb{Q}}$ do not appear perpendicular to $\mathbf{u}$; (c) Elliptic orthogonal parameterization in 2-dimensional space; (d) Hyperbolic orthogonal parameterization in 2-dimensional space.

3. Methodology

3.1. Generalized Orthogonal Matrix and Mapping

In the first step, we seek an elegant parameterization method for modeling orthogonal matrices in Euclidean, elliptic and hyperbolic spaces of arbitrary dimension k. To establish this universality, we take a broader perspective by introducing a generalized space equipped with a quadratic inner product.

Quadratic form: For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, the quadratic inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}}$ with respect to $\mathbf{w} \in \mathbb{R}^k$ is defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} = \mathbf{x}^\top \mathrm{diag}(\mathbf{w}) \mathbf{y} = \sum_{i=1}^{k} w_i x_i y_i, \tag{3}$$

where $\mathbf{w}$ is a weighting vector with no elements equal to 0, ensuring that $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}}$ is non-degenerate.

Generalized orthogonal matrix: In the generalized space endowed with $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}}$, a real square matrix $\mathbf{G} \in \mathbb{R}^{k \times k}$ is orthogonal if it satisfies:

$$\langle \mathbf{G}\mathbf{x}, \mathbf{G}\mathbf{y} \rangle_{\mathbf{w}} = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} \Leftrightarrow \mathbf{G}^\top \mathrm{diag}(\mathbf{w}) \mathbf{G} = \mathrm{diag}(\mathbf{w}). \tag{4}$$

We call the set of all $\mathbf{G}$ as the generalized orthogonal group in dimension k, denoted by $\mathbf{O}_{\mathbf{w}}(k)$. According to Equation (1) and (2), the generalized orthogonal matrix $\mathbf{G} \in \mathbf{O}_{\mathbf{w}}(k)$ can be naturally specialized into Euclidean, elliptic and hyperbolic spaces by setting $\mathbf{w}$ to $\mathbf{1}$, $\mathbf{p}$ and $\mathbf{q}$, respectively. Based on this observation, we turn to modeling such generalized orthogonal matrices for establishing the universality.

Generalized Householder matrix: Before fully covering all $\mathbf{G} \in \mathbf{O}_{\mathbf{w}}(k)$, we start with the simplest case: *elementary reflection matrix*, which is also known as the Householder matrix (Householder, 1958) in Euclidean space. Here we derive the generalized form of Householder matrix in the quadratic inner product space.

Geometrically, the generalized Householder matrix $\mathbf{H}$ describes an elementary reflection from $\mathbf{x}$ to $\mathbf{x}'$ about a a hyperplane ζ as shown in Figure 2(a). Given a normal vector $\mathbf{u} \in \mathbb{R}^k$ of ζ, the projection of $\mathbf{x}$ onto $\mathbf{u}$ is denoted as:

$$\mathrm{Pro}_{\mathbf{u}}(\mathbf{x}) = \frac{\langle \mathbf{u}, \mathbf{x} \rangle_{\mathbf{w}}}{\langle \mathbf{u}, \mathbf{u} \rangle_{\mathbf{w}}} \mathbf{u}. \tag{5}$$

By the law of vector addition, the projection of $\mathbf{x}$ onto the hyperplane ζ, i.e., $\mathrm{Pro}_{\zeta}(\mathbf{x})$, is calculated as:

$$\mathrm{Pro}_{\zeta}(\mathbf{x}) = \mathbf{x} - \mathrm{Pro}_{\mathbf{u}}(\mathbf{x}). \tag{6}$$

Based on the concept of projections, we move to the elementary reflection of $\mathbf{x}$ with respect to ζ in the quadratic inner product space. The resulting vector $\mathbf{x}'$ is given by:

$$\mathbf{x}' = \mathrm{Pro}_{\zeta}(\mathbf{x}) - \mathrm{Pro}_{\mathbf{u}}(\mathbf{x}) = \underbrace{(\mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^\top \mathrm{diag}(\mathbf{w})}{\mathbf{u}^\top \mathrm{diag}(\mathbf{w}) \mathbf{u}})}_{\mathbf{H}(\mathbf{u},\mathbf{w})} \mathbf{x}, \tag{7}$$

where $\mathbf{I} \in \mathbb{R}^{k \times k}$ is the identity matrix. Notably, the induced reflection matrix $\mathbf{H}(\mathbf{u}, \mathbf{w})$, with normal vector $\mathbf{u}$ and weighting vector $\mathbf{w}$ as variables, can be viewed as a generalization of Euclidean Householder matrix ($\mathbf{w} = \mathbf{1}$) in the quadratic inner product space. It is straightforward to demonstrate that the generalized Householder matrix $\mathbf{H}(\mathbf{u}, \mathbf{w})$ is orthogonal, which satisfies the inner product invariance in Equation (4). Figure 2(b) illustrates the Euclidean, elliptic and hyperbolic reflections in 2-dimensional space, described by $\mathbf{H}$ with $\mathbf{w}$ equal to $\mathbf{1}$, $\mathbf{p}$ and $\mathbf{q}$, respectively.

Orthogonal mapping: By taking the elementary reflections as basic operators, we can design a mapping to represent the generalized orthogonal transformations. Formally, given a series of vectors $\mathbf{U} = \{\mathbf{u}_c\}_{c=1}^{n}$ where $\mathbf{u}_c \in \mathbb{R}^k$, we define the mapping with respect to $\mathbf{w} \in \mathbb{R}^k$ as follows:

$$\mathrm{Orth}(\mathbf{U}, \mathbf{w}) = \prod_{c=1}^{n} \mathbf{H}(\mathbf{u}_c, \mathbf{w}). \tag{8}$$

Since each $\mathbf{H}(\mathbf{u}_c, \mathbf{w})$ preserves the quadratic inner product, one can easily verify that the output of $\mathrm{Orth}(\mathbf{U}, \mathbf{w})$ is also a generalized orthogonal matrix according to Equation (4). Moreover, we have the following theorem:

Theorem 3.1. *When $n = k$, the image of* Orth *is the set of all $k \times k$ generalized orthogonal matrices, i.e.,* $\mathrm{Image}(\mathrm{Orth}) = \mathbf{O}_{\mathbf{w}}(k)$. *(See proof in Appendix B)*

This theorem guarantees that the output of Orth completely covers the generalized orthogonal group $\mathbf{O}_{\mathbf{w}}(k)$. Therefore, such a mapping can naturally represent all the orthogonal matrices of Euclidean, elliptic and hyperbolic types with the weighting vector $\mathbf{w}$ equal to $\mathbf{1}$, $\mathbf{p}$ and $\mathbf{q}$, respectively.

3.2. GoldE: Universal Orthogonal Parameterization

In the following, we leverage the derived mapping Orth as the cornerstone to develop the *elliptic* (Section 3.2.1) and *hyperbolic* (Section 3.2.2) orthogonal parameterizations for modeling relations, which are further integrated in a product manifold to establish our GoldE framework (Section 3.2.3).

3.2.1. Elliptic Orthogonal Parameterization

To better capture logical patterns and cyclical structures of KGs, we propose to model each relation as a k-dimensional elliptic orthogonal transformation between head and tail entities with the designed mapping Orth in Equation (8).

Given an elliptic weighting vector $\mathbf{p} \in \mathbb{R}^k_+$, the initial embedding $\mathbf{e}_v \in \mathbb{R}^k$ for entity $v \in \mathcal{V}$ can be regarded as lying on the ellipsoid $\mathbb{P}^k_{\mathbf{p},\beta_{\mathbf{e}_v,\mathbf{p}}}$, where $\beta_{\mathbf{e}_v,\mathbf{p}} = \langle \mathbf{e}_v, \mathbf{e}_v \rangle_{\mathbf{p}}$. Considering that each entity typically exhibits different characteristics when involving different relations (Wang et al., 2014), we define a learnable weighting vector $\mathbf{p}_r$ for each relation $r \in \mathcal{R}$ to associate entities with the r-specific ellipsoid. For each triple (h, r, t), the r-associated embeddings of head entity h and tail entity t are denoted as $\mathbf{e}_{h,r} \in \mathbb{P}^k_{\mathbf{p}_r,\beta_{\mathbf{e}_h,\mathbf{p}_r}}$ and $\mathbf{e}_{t,r} \in \mathbb{P}^k_{\mathbf{p}_r,\beta_{\mathbf{e}_t,\mathbf{p}_r}}$. We then represent relation r as an elliptic orthogonal transformation between $\mathbf{e}_{h,r}$ and $\mathbf{e}_{t,r}$. Formally, the embedding of relation r is defined as $\mathbf{U}_r \in \mathbb{R}^{k\times k}$, in which each row $\mathbf{U}_r[i] \in \mathbb{R}^k$ is a normal vector for elementary reflection. Taking $\mathbf{U}_r$ and $\mathbf{p}_r$ as input to the mapping Orth, we can apply the r-specific elliptic orthogonal transformation to the head embedding $\mathbf{e}_{h,r}$:

$$\begin{aligned} \mathbf{e}'_{h,r} &= \mathrm{Orth}_{\mathbb{P}}(\mathbf{U}_r, \mathbf{p}_r)\mathbf{e}_{h,r} \\ &= \prod_{i=1}^{k} \mathbf{H}(\mathbf{U}_r[i], \mathbf{p}_r)\mathbf{e}_{h,r}. \end{aligned} \tag{9}$$

Based on Theorem 3.1, any k-dimensional elliptic orthogonal relation transformation can be naturally represented by Equation (9) without any loss of degrees of freedom.

Objective function: If triple (h, r, t) holds, the transformed head embedding $\mathbf{e}'_{h,r}$ is expected to be close to the tail embedding $\mathbf{e}_{t,r}$. Figure 2(c) illustrates the elliptic orthogonal parameterization in 2-dimensional space. Note that $\mathbf{e}'_{h,r}$ and $\mathbf{e}_{t,r}$ lie on the ellipsoids that have the same $\mathbf{p}_r$ but different β. To measure such proximity, we introduce the extrinsic Mahalanobis distance (Mitchell & Krzanowski, 1985) as the score function for each (h, r, t):

$$\begin{aligned} s_r(h,t) &= -d_{\mathbb{P}}(\mathbf{e}'_{h,r}, \mathbf{e}_{t,r}) \\ &= -\sqrt{(\mathbf{e}'_{h,r} - \mathbf{e}_{t,r})^\top \mathrm{diag}(\mathbf{p}_r)(\mathbf{e}'_{h,r} - \mathbf{e}_{t,r})}. \end{aligned} \tag{10}$$

Note that it is also possible to normalize all entities onto one ellipsoid and measure similarity using the elliptic geodesic distance. However, the extrinsic measurement usually works better in practice (Sun et al., 2019; Wang et al., 2021).

Connections to Euclidean models: As shown in Table 1, a series of existing models such as RotatE, QuatE and HousE represent relations as Euclidean orthogonal transformations with respective dimensions. All these models can be viewed as the special cases of the proposed elliptic parameterization by setting $\mathbf{p}_r$ to $\mathbf{1}$ for all $r \in \mathcal{R}$. With the designed orthogonal mapping Orth, our parameterization is applicable to the space of any dimension k. Moreover, we highlight that such elliptic generalization moves beyond Euclidean models by introducing an additional relation-specific scaling operation. Formally, we can prove the following claim:

Claim 3.2. *The proposed elliptic parameterization can be reformulated as Euclidean parameterization equipped with element-wise scaling transformations determined by $\sqrt{\mathbf{p}_r}$. (See proof in Appendix C)*

Such scaling transformations enable an adaptive adjustment of the margin in loss function to fit for complex relations (Chao et al., 2021). However, elliptic geometry is inherently not the optimal choice for preserving the hierarchical structures due to its mismatched geometric characteristics, which then motivates us to investigate hyperbolic parameterization.

3.2.2. Hyperbolic Orthogonal Parameterization

As a complement for preserving the hierarchical structures of KGs with low distortion, we further represent relations as k-dimensional hyperbolic orthogonal transformations.

In order to rigorously measure similarity with the hyperbolic geodesic distance[1], we begin by projecting the initial entity embeddings $\mathbf{e}_h$ and $\mathbf{e}_t$ to the hyperboloid via the hyperbolic exponential map $g_\beta : \mathbb{R}^{k-1} \to \mathbb{Q}^k_\beta$. Given any $\mathbf{x} \in \mathbb{R}^{k-1}$, g_β is defined as (Mishne et al., 2023):

$$g_\beta(\mathbf{x}) = (\sqrt{\beta}\cosh(\frac{\|\mathbf{x}\|}{\sqrt{\beta}}), \sqrt{\beta}\sinh(\frac{\|\mathbf{x}\|}{\sqrt{\beta}})\frac{\mathbf{x}}{\|\mathbf{x}\|}). \tag{11}$$

Following (Chami et al., 2020), we define a learnable radius of curvature β_r for each relation r to encode a variety of hierarchies. For each triple (h, r, t), $\mathbf{e}_h$ and $\mathbf{e}_t$ are mapped

[1]Another alternative is the extrinsic Lorentzian distance (Law et al., 2019). However, such distance does not satisfy the triangle inequality and may underestimate the plausibility of triples.

to the r-specific hyperboloid $\mathbb{Q}^k_{\beta_r}$ via g_{β_r}, resulting in the r-associated embeddings $\mathbf{e}_{h,r} = g_{\beta_r}(\mathbf{e}_h)$ and $\mathbf{e}_{t,r} = g_{\beta_r}(\mathbf{e}_t)$.

For parameterizing hyperbolic orthogonal transformation of any dimension k between $\mathbf{e}_{h,r}$ and $\mathbf{e}_{t,r}$, a straightforward strategy is to leverage the designed orthogonal mapping Orth in the case of $\mathbf{w} := \mathbf{q} = (-1, 1, \ldots, 1)^\top$. Given the relation embedding $\mathbf{U}_r \in \mathbb{R}^{k\times k}$, the mapping $\mathrm{Orth}(\mathbf{U}_r, \mathbf{q})$ is able to represent every hyperbolic orthogonal relation matrix $\mathbf{G}_r \in \mathbf{O}_{\mathbf{q}}(k)$ according to Theorem 3.1.

Nevertheless, although $\mathbf{G}_r$ satisfies the inner product invariance in Equation (4), it is not guaranteed to preserve the sign of the first element of $\mathbf{e}_{h,r}$. In other words, $\mathbf{e}'_{h,r}$ might be transformed to the lower sheet of the hyperboloid with the first element less than 0, leading to the inability to calculate the geodesic distance between $\mathbf{e}'_{h,r} \notin \mathbb{Q}^k_{\beta_r}$ and $\mathbf{e}_{t,r} \in \mathbb{Q}^k_{\beta_r}$. To remedy this deficiency, we restrict the orthogonal relation matrix $\mathbf{G}_r$ into the *positive* subgroup of $\mathbf{O}_{\mathbf{q}}(k)$ such that for any $\mathbf{x} \in \mathbb{Q}^k_{\beta_r}$, the first element of $\mathbf{G}_r\mathbf{x}$ is positive, thereby ensuring that $\mathbf{G}_r\mathbf{x} \in \mathbb{Q}^k_{\beta_r}$. Essentially, the full hyperbolic orthogonal group $\mathbf{O}_{\mathbf{q}}(k)$ can be classified into two distinct connected components based on the following proposition:

Proposition 3.3. *For any $\mathbf{x} \in \mathbb{Q}^k_\beta$, every hyperbolic orthogonal matrix $\mathbf{G} \in \mathbf{O}_{\mathbf{q}}(k)$ falls into two subsets according to the sign of its first element G_{11} for which $|G_{11}| \geq 1$:*

- *The matrix in the positive subset $\mathbf{O}^+_{\mathbf{q}}(k) = \{\mathbf{G} \in \mathbf{O}_{\mathbf{q}}(k) : G_{11} \geq +1\}$ preserves the sign of the first element of $\mathbf{x}$;*
- *The matrix in the negative subset $\mathbf{O}^-_{\mathbf{q}}(k) = \{\mathbf{G} \in \mathbf{O}_{\mathbf{q}}(k) : G_{11} \leq -1\}$ reverses the sign of the first element of $\mathbf{x}$;*
- *$\mathbf{O}^+_{\mathbf{q}}(k)$ is a multiplicative subgroup of $\mathbf{O}_{\mathbf{q}}(k)$.*

(See proof in Appendix D)

For fully covering the positive subgroup $\mathbf{O}^+_{\mathbf{q}}(k)$, we employ the polar decomposition (Gallier, 2005) to express any $\mathbf{G} \in \mathbf{O}^+_{\mathbf{q}}(k)$ as the product of a Euclidean orthogonal matrix and a hyperbolic boost matrix. Formally, we have the following proposition, based on which a positive hyperbolic mapping $\mathrm{Orth}_{\mathbb{Q}}$ is defined for relational parameterization:

Proposition 3.4. *Every (positive) hyperbolic orthogonal matrix $\mathbf{G} \in \mathbf{O}^+_{\mathbf{q}}(k)$ can be expressed by a mapping $\mathrm{Orth}_{\mathbb{Q}}$ such that for $\mathbf{U} \in \mathbb{R}^{(k-1)\times(k-1)}$ and $\mathbf{b} \in \mathbb{R}^{k-1}$:*

$$\begin{aligned} G &= \mathrm{Orth}_{\mathbb{Q}}(\mathbf{U}, \mathbf{b}) \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{Orth}(\mathbf{U}, \mathbf{1}) \end{bmatrix} \begin{bmatrix} \sqrt{\|\mathbf{b}\|^2 + 1} & \mathbf{b}^\top \\ \mathbf{b} & \sqrt{\mathbf{I} + \mathbf{b}\mathbf{b}^\top} \end{bmatrix}. \end{aligned}$$

(See proof in Appendix E)

Based on Proposition 3.4, we can represent each relation r as a hyperbolic orthogonal transformation whose output space is restricted to $\mathbb{Q}^k_{\beta_r}$. Specifically, two types of parameters are defined for each r: $\mathbf{U}_r \in \mathbb{R}^{(k-1)\times(k-1)}$ and $\mathbf{b}_r \in \mathbb{R}^{k-1}$. Taking $\mathbf{U}_r$ and $\mathbf{b}_r$ as input to the mapping $\mathrm{Orth}_{\mathbb{Q}}$, we apply such restricted orthogonal transformation to $\mathbf{e}_{h,r} \in \mathbb{Q}^k_{\beta_r}$:

$$\begin{aligned} \mathbf{e}'_{h,r} &= \mathrm{Orth}_{\mathbb{Q}}(\mathbf{U}_r, \mathbf{b}_r)\mathbf{e}_{h,r} \qquad (12) \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \prod_{i=1}^{k-1} \mathbf{H}(\mathbf{U}_r[i], \mathbf{1}) \end{bmatrix} \begin{bmatrix} \sqrt{\|\mathbf{b}_r\|^2 + 1} & \mathbf{b}_r^\top \\ \mathbf{b}_r & \sqrt{\mathbf{I} + \mathbf{b}_r\mathbf{b}_r^\top} \end{bmatrix} \mathbf{e}_{h,r}. \end{aligned}$$

Objective function: Proposition 3.4 theoretically ensures that the transformed head embedding $\mathbf{e}'_{h,r}$ lies on the same hyperboloid as the tail embedding $\mathbf{e}_{t,r}$, enabling us to measure their similarity using the hyperbolic geodesic distance. Formally, the score function for each (h, r, t) is defined as:

$$\begin{aligned} s_r(h, t) &= -d_{\mathbb{Q}}(\mathbf{e}'_{h,r}, \mathbf{e}_{t,r}) \\ &= -\sqrt{\beta}\cosh^{-1}\left(-\frac{\langle \mathbf{e}'_{h,r}, \mathbf{e}_{t,r}\rangle_{\mathbf{q}}}{\beta}\right). \end{aligned} \qquad (13)$$

Connections to hyperbolic models: Existing hyperbolic models such as AttH, RotH and RefH (Chami et al., 2020) represent each relation as the combination of a rotation (or reflection) and a Möbius addition in the Poincaré ball. Since the Poincaré ball is isometric to the hyperboloid (Nickel & Kiela, 2017), and the Möbius addition is equivalent to the hyperbolic boost (Tabaghi & Dokmanic, 2021), all these models can be viewed as the special cases of our proposed hyperbolic parameterization. Note that these models parameterize rotations/reflections using the 2-dimensional Givens transformations, while our approach breaks the dimensional restriction without loss of degrees of freedom, thereby generalizing them and achieving superior modeling capacity.

3.2.3. Mixed Orthogonal Parameterization

Although our elliptic and hyperbolic parameterizations are advantageous for embedding the corresponding topologies (cyclicity or hierarchy), real-world KGs typically exhibit non-uniform structures (Gu et al., 2019; Wang et al., 2021). Considering such inherent topological heterogeneity, we further introduce a mixed orthogonal parameterization in the product manifold (Gu et al., 2019) to unify multiple geometric spaces, culminating in the universal GoldE framework.

Product manifold: Beyond individual elliptic or hyperbolic spaces, we leverage the Cartesian product to describe a heterogeneous embedding space for GoldE: $\mathbb{D}^k = \bigtimes_{i=1}^m \mathbb{X}_i^{k_i}$, where $\mathbb{X}_i^{k_i} \in \{\mathbb{E}, \mathbb{P}, \mathbb{Q}\}$[2] denotes the i-th component space of dimension k_i[3] such that the total dimension $k = \sum_{i=1}^m k_i$. The resulting product space is a "non-constantly" curved manifold that is more flexible than a single constant curvature manifold (Skopek et al., 2020).

[2]Note that the notation $\mathbb{E}$ is redundant since Euclidean orthogonal parameterization is the special case of our proposed elliptic orthogonal parameterization according to Claim 3.2.

[3]One can assign identical dimensions to component spaces of the same geometric type (i.e., $k_i \in \{k_{\mathbb{P}}, k_{\mathbb{Q}}\}$) for reducing the number of hyperparameters (Gu et al., 2019).

For each triple (h, r, t), GoldE represents relation r as a component-wise orthogonal transformation between head entity h and tail entity t in the product manifold. Formally, let $\mathbf{e}_{h,r}^{(i)}$ and $\mathbf{e}_{t,r}^{(i)}$ denote the i-th sub-embeddings of h and t, which are associated to the r-specific (elliptic or hyperbolic) space as described in Section 3.2.1 and 3.2.2. GoldE applies the generalized orthogonal transformation to $\mathbf{e}_{h,r}^{(i)}$ according to the geometric type of the i-th component space $\mathbb{X}_i$:

$$\mathbf{e}_{h,r}^{\prime(i)} = \begin{cases} \mathrm{Orth}_{\mathbb{P}}(\mathbf{U}_r^{(i)}, \mathbf{p}_r^{(i)})\mathbf{e}_{h,r}^{(i)}, & \text{for } \mathbb{X}_i = \mathbb{P} \\ \mathrm{Orth}_{\mathbb{Q}}(\mathbf{U}_r^{(i)}, \mathbf{b}_r^{(i)})\mathbf{e}_{h,r}^{(i)}, & \text{for } \mathbb{X}_i = \mathbb{Q} \end{cases}, \quad (14)$$

where $\mathbf{U}_r^{(i)}$, $\mathbf{p}_r^{(i)}$ and $\mathbf{b}_r^{(i)}$ are the parameters of relation transformations in the i-th component space.

Objective function: Since the orthogonal transformations guarantee the inner product invariance in each component space, the transformed head embedding $\mathbf{e}'_{h,r}$ remains in the product space $\mathbb{D}^k$. The distance between $\mathbf{e}'_{h,r}$ and $\mathbf{e}_{t,r}$ can be calculated by performing decomposition (Ficken, 1939; Turaga & Srivastava, 2016). Formally, the score function for measuring the plausibility of each (h, r, t) is defined as:

$$s_r(h, t) = -d_{\mathbb{D}}(\mathbf{e}'_{h,r}, \mathbf{e}_{t,r}) = -\sum_{i=1}^{m} d^l_{\mathbb{X}_i}(\mathbf{e}_{h,r}^{\prime(i)}, \mathbf{e}_{t,r}^{(i)}), \quad (15)$$

where l is the norm indicator. We highlight that such mixed orthogonal parameterization in the product space naturally achieves the geometric unification, enabling improved representations by better matching the geometry of the embedding space to the heterogeneous structures of KGs.

Complexity analysis: The learnable parameters of GoldE include $\{\mathbf{e}_v\}_{v\in\mathcal{V}}$ and $\{\mathbf{U}_r, \mathbf{p}_r, \mathbf{b}_r, \beta_r\}_{r\in\mathcal{R}}$. Compared to the previous models (Sun et al., 2019; Chami et al., 2020; Li et al., 2022), the extra space cost is proportional to the number of relation types, which is usually much smaller than the number of entities. Therefore, GoldE has the same space complexity as the existing KGE models, i.e., $O(k|\mathcal{V}|)$. In the aspect of time cost for processing a single triple, one may observe that the time complexity of Equation (14) is $O(k_i^3)$, in which k_i matrix-vector multiplications incur high computational costs. However, it is worth noting that these multiplications can be entirely replaced by vector operations via efficient computation (refer to Appendix G for details), thereby reducing the time complexity to $O(k_i^2)$.

Modeling capability: Benefiting from the dimensional extension and geometric unification achieved by the designed parameterization, GoldE is able to simultaneously model the logical patterns and heterogeneous structures as shown in Table 1. Formally, we have the following claim:

Claim 3.5. *GoldE is capable of capturing the symmetry, antisymmetry, inversion and composition patterns, as well as the inherent cyclical and hierarchical structures in KGs. (See proof in Appendix F)*

3.3. Optimization

Following (Sun et al., 2019; Gao et al., 2020; Li et al., 2022), we use the self-adversarial negative sampling loss (Sun et al., 2019) to optimize our GoldE framework. Formally, given a positive triple (h, r, t), the loss function is defined as:

$$\begin{aligned} L = &-\log\sigma(\gamma + s_r(h, t)) \\ &- \sum_{i=1}^{g} p(h'_i, r, t'_i)\log\sigma(-s_r(h'_i, t'_i) - \gamma), \end{aligned} \quad (16)$$

where γ is a pre-defined margin, σ is the sigmoid function, g is the number of negative samples, (h'_i, r, t'_i) is the i-th negative sample against (h, r, t), and $p(h'_i, r, t'_i)$ determines the proportion of (h'_i, r, t'_i) in the current optimization as defined in (Sun et al., 2019). Note that we initially define the entity embeddings of GoldE in Euclidean space, and all the relation transformations (including g_β) are differentiable and bijective (Chami et al., 2019). This enables us to learn the parameters using the standard Euclidean optimization techniques for the numerical stability (Mishne et al., 2023).

4. Experiments

To comprehensively validate the effectiveness of GoldE, we conduct extensive experiments on link prediction task, and attempt to answer the following questions:

(1) How does GoldE perform on link prediction compared to existing KGE approaches? (Section 4.2)

(2) What is the effect of configuring different geometries and dimensions for orthogonal relation transformations on the performance of GoldE? (Section 4.3)

(3) From a fine-grained perspective, does GoldE effectively model different relation types? (Section 4.4)

(4) Can GoldE still learn high-quality representations even when the embedding size k is restricted? (Section 4.5)

4.1. Experimental Setup

Datasets: We evaluate GoldE framework on three standard benchmarks: WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova & Chen, 2015) and YAGO3-10 (Mahdisoltani et al., 2015). Refer to Appendix H for statistical details.

Baselines: We compare GoldE to a series of advanced KGE approaches, including (1) non-orthogonal models: TransE (Bordes et al., 2013), DistMult (Yang et al., 2015), ComplEx (Trouillon et al., 2016), ConvE (Dettmers et al., 2018) and MuRP (Balazevic et al., 2019); (2) Euclidean orthogonal models: RotatE (Sun et al., 2019), Rotate3D (Gao et al., 2020), QuatE (Zhang et al., 2019), DualE (Cao et al., 2021), ReflectE (Zhang et al., 2022), HousE (Li et al., 2022) and CompoundE (Ge et al., 2023); (3) hyperbolic orthogonal models: RefH, RotH and AttH (Chami et al., 2020).

Table 2. Link prediction results on WN18RR, FB15k-237 and YAGO3-10. Best results are in **bold** and second best results are underlined. [†]: Results are taken from (Nguyen et al., 2018); [◇]: Results are taken from (Dettmers et al., 2018). [∗]: HousE-pro is a variant of HousE with additional projection operations (Li et al., 2022). Other results are taken from the corresponding original papers.

Model	WN18RR					FB15k-237					YAGO3-10				
	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE†	3384	.226	-	-	.501	357	.294	-	-	.465	-	-	-	-	-
DistMult◇	5110	.430	.390	.440	.490	254	.241	.155	.263	.419	5926	.340	.240	.380	.540
ComplEx◇	5261	.440	.410	.460	.510	339	.247	.158	.275	.428	6351	.360	.260	.400	.550
ConvE◇	4187	.430	.400	.440	.520	224	.325	.237	.356	.501	1671	.440	.350	.490	.620
MuRP	-	.481	.440	.495	.566	-	.335	.243	.367	.518	-	.354	.249	.400	.567
RotatE	3340	.476	.428	.492	.571	177	.338	.241	.375	.533	1767	.495	.402	.550	.670
Rotate3D	3328	.489	.442	.505	.579	165	.347	.250	.385	.543	-	-	-	-	-
QuatE	3472	.481	.436	.500	.564	176	.311	.221	.342	.495	-	-	-	-	-
DualE	-	.482	.440	.500	.561	-	.330	.237	.363	.518	-	-	-	-	-
ReflectE	1730	.488	.450	.501	.559	148	.358	.263	.396	.546	-	-	-	-	-
HousE	1885	.496	.452	.511	.585	165	.348	.254	.384	.534	1449	.565	.487	.616	.703
HousE-pro∗	1303	.511	.465	.528	.602	153	.361	.266	.399	.551	1415	.571	.491	.620	.714
CompoundE	-	.491	.450	.508	.576	-	.357	.264	.393	.545	-	-	-	-	-
RefH	-	.461	.404	.485	.568	-	.346	.252	.383	.536	-	.576	.502	.619	.711
RotH	-	.496	.449	.514	.586	-	.344	.246	.380	.535	-	.570	.495	.612	.706
AttH	-	.486	.443	.499	.573	-	.348	.252	.384	.540	-	.568	.493	.612	.702
GoldE	**790**	**.525**	**.476**	**.542**	**.615**	**142**	**.370**	**.277**	**.405**	**.558**	**536**	**.588**	**.510**	**.634**	**.723**

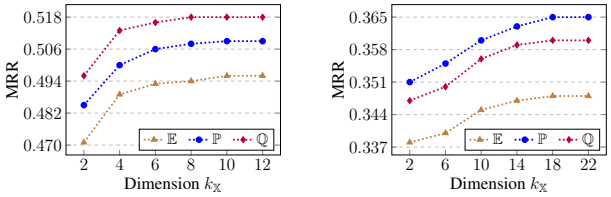
Table 3. Link prediction results for our GoldE framework under different geometric combinations.

Geometry of Orth.	WN18RR		FB15k-237		YAGO3-10	
	MRR	H@10	MRR	H@10	MRR	H@10
$\mathbb{E}^k$	.496	.585	.348	.534	.565	.703
$\mathbb{P}^k$	.505	.593	.358	.547	.572	.708
$\mathbb{Q}^k$	.513	.606	.354	.542	.580	.714
$(\mathbb{P}_i^{k_\mathbb{P}})^{m_\mathbb{P}}$	.509	.602	.365	.553	.578	.712
$(\mathbb{Q}_j^{k_\mathbb{Q}})^{m_\mathbb{Q}}$	.518	.610	.360	.549	.585	.718
$(\mathbb{P}_i^{k_\mathbb{P}})^{m_\mathbb{P}} \times (\mathbb{Q}_j^{k_\mathbb{Q}})^{m_\mathbb{Q}}$	**.525**	**.615**	**.370**	**.558**	**.588**	**.723**

Implementation details: For fair comparisons, we follow (Li et al., 2022) to fix the embedding dimension k, ensuring that the number of parameters is comparable to the baselines. Moreover, we follow the most natural way (Gu et al., 2019) to construct the product manifold $\mathbb{D}$ as the combination of both $\mathbb{P}$ and $\mathbb{Q}$ for GoldE, i.e., $\mathbb{D}^k = \bigtimes_{i=1}^{m_\mathbb{P}} \mathbb{P}_i^{k_\mathbb{P}} \times \bigtimes_{j=1}^{m_\mathbb{Q}} \mathbb{Q}_j^{k_\mathbb{Q}}$. The hyperparameters $m_\mathbb{P}$ and $k_\mathbb{P}$ ($m_\mathbb{Q}$ and $k_\mathbb{Q}$) determine the number and dimension of elliptic (hyperbolic) components, respectively. More details can be found in Appendix I.

4.2. Performance Comparison

Table 2 summarizes the main results on link prediction task. One can observe that GoldE framework consistently outperforms the baselines on all metrics across the three datasets. In particular, compared to HousE, the most relevant baseline that represents relations as k-dimensional Euclidean Householder rotations, GoldE achieves an average absolute gain of 2.5% in MRR, and even surpasses its projection-enhanced variant (i.e., HousE-pro) with 1.3% absolute improvement in MRR. Moreover, GoldE also outperforms exisiting hyperbolic orthogonal models (i.e., RefH, RotH and AttH) by a clear margin. Such new state-of-the-art results demonstrate the superiority of our universal orthogonal parameterization.



(a) MRR vs. $k_\mathbb{X}$ on WN18RR (b) MRR vs. $k_\mathbb{X}$ on FB15k-237

Figure 3. MRR results of GoldE under the product of $\mathbb{X}$ with varying dimension $k_\mathbb{X}$ ($\mathbb{X} \in \{\mathbb{E}, \mathbb{P}, \mathbb{Q}\}$) on WN18RR and FB15k-237.

4.3. Effect of Universal Orthogonal Parameterization

Effect of geometric unification: To investigate the effect of geometries for orthogonal relation transformations, we explore the performance of GoldE under different geometric combinations[4]. From the results summarized in Table 3, one can observe the following: (1) Both elliptic and hyperbolic parameterizations are more effective than the Euclidean counterpart, indicating the importance of additional scaling operations and modeling inherent hierarchies, respectively; (2) As we expected, hyperbolic parameterization performs better on the two datasets (i.e., WN18RR and YAGO3-10) dominated by hierarchies, while the elliptic counterpart excels on the dataset (i.e., FB15k-237) dominated by cycles. Note that the product of multiple $\mathbb{P}$ (or $\mathbb{Q}$) outperforms its single case, since such homogeneous combination is able to embed a variety of cycles (or hierarchies) (Gu et al., 2019); (3) By combining multiple $\mathbb{P}$ and $\mathbb{Q}$, our GoldE can naturally preserve both cyclical and hierarchical structures, thereby achieving the optimal performance on all the three datasets.

[4]To simplify, we use a shorthand notation for repeated components: $(\mathbb{X}_i^{k_\mathbb{X}})^{m_\mathbb{X}} = \bigtimes_{i=1}^{m_\mathbb{X}} \mathbb{X}_i^{k_\mathbb{X}}$. Note that the Cartesian product of $\mathbb{E}$ is redundant since it satisfies $\mathbb{E}^k = \bigtimes_{i=1}^{m_\mathbb{E}} \mathbb{E}_i^{k_\mathbb{E}}$, while this equality does not hold for non-Euclidean manifolds (Skopek et al., 2020).

Table 4. MRR for the models tested on each relation of WN18RR.

Relation Name	#Triple	RotH	HousE	GoldE
hypernym	1,251	0.190	0.182	**0.226**
instance_hypernym	122	0.362	0.395	**0.453**
member_meronym	253	0.266	0.275	**0.296**
synset_domain_topic_of	114	0.406	0.396	**0.430**
has_part	172	0.213	0.217	**0.255**
member_of_domain_usage	24	0.350	0.415	**0.455**
member_of_domain_region	26	0.356	0.281	**0.403**
derivationally_related_form	1,074	0.947	0.958	**0.961**
also_see	56	0.583	0.638	**0.665**
verb_group	39	0.820	0.968	**0.975**
similar_to	3	**1.000**	**1.000**	**1.000**

Table 5. Link prediction results for 32-dimensional embeddings.

	WN18RR		FB15k-237		YAGO3-10	
Model	MRR	H@10	MRR	H@10	MRR	H@10
RotatE	.387	.491	.290	.458	-	-
QuatE	.421	.467	.293	.460	-	-
MuRE	.458	.525	.313	.489	.283	.478
MuRP	.465	.544	.323	.501	.230	.392
RefH	.447	.518	.312	.489	.381	.530
RotH	.472	.553	.314	.497	.393	.559
AttH	.466	.551	.324	.501	.397	.566
GoldE	**.482**	**.558**	**.335**	**.506**	**.420**	**.586**

Such improvements confirm the effectiveness of the geometric unification in our universal orthogonal parameterization.

Effect of dimensional extension: To verify the expressiveness of the dimensional extension in our parameterization, we further conduct experiments for GoldE under the product of $\mathbb{X}$ with varying dimension $k_{\mathbb{X}}$. Figure 3 exhibits the MRR results on WN18RR and FB15k-237. As expected, whether in Euclidean, elliptic, or hyperbolic parameterizations, the higher-dimensional orthogonal transformations enhance modeling capability, leading to superior performance compared to their lower-dimensional counterparts.

4.4. Fine-grained Performance Analysis

To verify the modeling capability of our GoldE from a fine-grained perspective, we report its performance per relation on WN18RR in Table 4. Compared to the hyperbolic model RotH and the Euclidean model HousE, our GoldE achieves the best performance on all 11 relation types. Specifically, GoldE not only obtains significant improvements on hierarchical relations (e.g., *hypernym* and *instance_hypernym*), but also performs well on symmetric relations (e.g., *derivationally_related_form* and *verb_group*). These advanced fine-grained results demonstrate the superior and comprehensive modeling capability of our GoldE framework.

4.5. Results with Restricted Embedding Size

We also follow (Chami et al., 2020) to evaluate our GoldE under the restriction of fixing the embedding size k to 32. It is challenging for KGE models to offer qualified representations in such a restricted setting. Table 5 reports the experimental results on the three datasets. We observe that GoldE consistently surpasses the baselines by a clear margin across all datasets, confirming the higher-quality representations learned by our universal orthogonal parameterization. See Appendix J for more results of GoldE with varying k.

5. Related Work

Early KGE approaches explore various geometric operations such as translation (Bordes et al., 2013; Wang et al., 2014; Lin et al., 2015; Ji et al., 2015) and scaling (Yang et al., 2015; Trouillon et al., 2016) to model relations. As a representative work, RotatE (Sun et al., 2019) goes beyond the previous methods by modeling relations as 2-dimensional rotations between entities to handle the crucial logical patterns in KGs. Inspired by the effectiveness of orthogonal transformations, subsequent research mainly advances in two directions.

Dimension direction: A series of works (Zhang et al., 2019; Gao et al., 2020; Cao et al., 2021; Li et al., 2022) seek higher-dimensional orthogonal transformations for better modeling capacity. Recently, HousE (Li et al., 2022) introduces the Euclidean Householder reflections to represent relations as k-dimensional Euclidean rotations, which can be viewed as the Euclidean specialization of our proposal.

Geometry direction: Another branch of KGE models (Balazevic et al., 2019; Chami et al., 2020) exploits hyperbolic geometry for preserving the hierarchical structures in KGs. AttH (Chami et al., 2020) represents relations as hyperbolic orthogonal transformations to accommodate both hierarchical structures and logical patterns, yet remains constrained by low-dimensional space and homogeneous geometry.

Notably, our GoldE framework generalizes existing KGE models in terms of both dimension and geometry. Moreover, the designed parameterization is orthogonal to advanced techniques such as regularizers (Lacroix et al., 2018) and graph encoders (Vashishth et al., 2020; Zhu et al., 2021), allowing for their integration to achieve better performance.

6. Conclusion

In this paper, we introduce GoldE, a powerful and general framework that generalizes existing KGE approaches based on a universal orthogonal parameterization. Empowered by such parameterization, our framework can simultaneously achieve dimensional extension and geometric unification for relational orthogonalization, thereby possessing the superior capability of capturing logical patterns and heterogeneous structures in KGs. Experimental results over three datasets comprehensively validate the effectiveness of our proposal. In future work, we will explore the potential of GoldE in modeling hyper-relational and temporal knowledge graphs.

Acknowledgements

This work is supported in part by National Natural Science Foundation of China (No. 62102420), Beijing Outstanding Young Scientist Program NO. BJJWZYJH 012019100020098, Intelligent Social Governance Platform, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative, Renmin University of China, Public Computing Cloud, Renmin University of China, fund for building world-class universities (disciplines) of Renmin University of China, Intelligent Social Governance Platform.

Impact Statement

This work presents GoldE based on a universal orthogonal parameterization, aiming to advance the field of knowledge graph embedding in terms of both dimension and geometry. We highlight that the designed orthogonal parameterization is a general technique, even with the potential to enhance various foundational models such as RNN (Vorontsov et al., 2017), GNN (Guo et al., 2022) and Transformer (Zhang et al., 2021), thus possessing the capability of facilitating a wide range of downstream tasks. As for the negative impact, GoldE may encode the biases present in the training graph, which can lead to stereotyped predictions for missing links between certain nodes (e.g., users on a social or e-commerce platform). We expect future studies to mitigate this issue.

References

Balazevic, I., Allen, C., and Hospedales, T. M. Multi-relational poincaré graph embeddings. In *Advances in Neural Information Processing Systems*, pp. 4465–4475, 2019.

Barrett, J. F. The hyperbolic theory of special relativity. *arXiv preprint arXiv:1102.0462*, 2011.

Bergstra, J. and Bengio, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305, 2012.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 1247–1250, 2008.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, pp. 2787–2795, 2013.

Cao, Z., Xu, Q., Yang, Z., Cao, X., and Huang, Q. Dual quaternion knowledge graph embeddings. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, pp. 6894–6902, 2021.

Cartan, É. *The theory of spinors*. M.I.T. Press, 1966.

Chami, I., Ying, Z., Ré, C., and Leskovec, J. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 4869–4880, 2019.

Chami, I., Wolf, A., Juan, D., Sala, F., Ravi, S., and Ré, C. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6901–6914, 2020.

Chao, L., He, J., Wang, T., and Chu, W. Pairre: Knowledge graph embeddings via paired relation vectors. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 4360–4369, 2021.

Clark, P. L. Quadratic forms chapter i: Witt's theory. *University of Georgia*, 2013.

Dettmers, T., Minervini, P., Stenetorp, P., and Riedel, S. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 1811–1818, 2018.

Dieudonné, J. A. *Sur les groupes classiques*. Hermann, 1967.

Do Carmo, M. P. *Riemannian geometry*. Birkhäuser, 1992.

Eisenhart, L. P. *Riemannian geometry*. Princeton University Press, 1997.

Feng, B.-J., Cheng, X., Xu, H.-N., and Xue, W.-F. Corporate credit ratings based on hierarchical heterogeneous graph neural networks. *Machine Intelligence Research*, 21:257–271, 2024.

Ficken, F. A. The riemannian and affine differential geometry of product-spaces. *Annals of Mathematics*, pp. 892–913, 1939.

Gallier, J. Notes on group actions, manifolds, lie groups, and lie algebras. *University of Pennsylvania*, 27, 2005.

Gallier, J. *Geometric methods and applications: for computer science and engineering*, volume 38. Springer Science & Business Media, 2011.

Gao, C., Sun, C., Shan, L., Lin, L., and Wang, M. Rotate3d: Representing relations as rotations in three-dimensional space for knowledge graph embedding. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 385–394, 2020.

Ge, X., Wang, Y., Wang, B., and Kuo, C. J. Compounding geometric operations for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6947–6965, 2023.

Gu, A., Sala, F., Gunel, B., and Ré, C. Learning mixed-curvature representations in product spaces. In *7th International Conference on Learning Representations*, 2019.

Guo, K., Zhou, K., Hu, X., Li, Y., Chang, Y., and Wang, X. Orthogonal graph neural networks. In *Thirty-Sixth AAAI Conference on Artificial Intelligence*, pp. 3996–4004, 2022.

Householder, A. S. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM (JACM)*, pp. 339–342, 1958.

Hu, J., Wu, L., Chen, Y., Hu, P., and Zaki, M. J. Graphflow+: Exploiting conversation flow in conversational machine comprehension with graph neural networks. *Machine Intelligence Research*, 21:272–282, 2024.

Ji, G., He, S., Xu, L., Liu, K., and Zhao, J. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 687–696, 2015.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, 2015.

Lacroix, T., Usunier, N., and Obozinski, G. Canonical tensor decomposition for knowledge base completion. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 2869–2878, 2018.

Law, M. T., Liao, R., Snell, J., and Zemel, R. S. Lorentzian distance learning for hyperbolic representations. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 3672–3681, 2019.

Li, C., Wang, S., Yang, D., Li, Z., Yang, Y., Zhang, X., and Zhou, J. PPNE: property preserving network embedding. In *Database Systems for Advanced Applications*, volume 10177, pp. 163–179. Springer, 2017.

Li, C., Cao, Y., Zhu, Y., Cheng, D., Li, C., and Morimoto, Y. Ripple knowledge graph convolutional networks for recommendation systems. *Machine Intelligence Research*, 21:481–494, 2024.

Li, R., Zhao, J., Li, C., He, D., Wang, Y., Liu, Y., Sun, H., Wang, S., Deng, W., Shen, Y., Xie, X., and Zhang, Q. House: Knowledge graph embedding with householder parameterization. In *International Conference on Machine Learning*, pp. 13209–13224, 2022.

Li, R., Chen, X., Li, C., Shen, Y., Zhao, J., Wang, Y., Han, W., Sun, H., Deng, W., Zhang, Q., and Xie, X. To copy rather than memorize: A vertical learning paradigm for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6335–6347, 2023.

Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 2181–2187, 2015.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 212–220, 2017.

Liu, Y., Wang, Z., Zhou, X., and Zheng, L. A study of using synthetic data for effective association knowledge learning. *Machine Intelligence Research*, 20:194–206, 2023.

Mahdisoltani, F., Biega, J., and Suchanek, F. Yago3: A knowledge base from multilingual wikipedias. In *7th Biennial Conference on Innovative Data Systems Research*, 2015.

Miller, G. A. Wordnet: a lexical database for english. *Communications of the ACM*, pp. 39–41, 1995.

Mishne, G., Wan, Z., Wang, Y., and Yang, S. The numerical stability of hyperbolic representation learning. In *International Conference on Machine Learning*, pp. 24925–24949, 2023.

Mitchell, A. F. and Krzanowski, W. J. The mahalanobis distance and elliptic distributions. *Biometrika*, 72(2): 464–467, 1985.

Nguyen, D. Q., Nguyen, T. D., Nguyen, D. Q., and Phung, D. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 327–333, 2018.

Nickel, M. and Kiela, D. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, pp. 6338–6347, 2017.

Nickel, M. and Kiela, D. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 3776–3785, 2018.

Özdemir, M. An alternative approach to elliptical motion. *Advances in Applied Clifford Algebras*, 26:279–304, 2016.

O'Meara, O. T. *Introduction to quadratic forms*, volume 117. Springer, 2013.

Rui, Y., Carmona, V. I. S., Pourvali, M., Xing, Y., Yi, W.-W., Ruan, H.-B., and Zhang, Y. Knowledge mining: A cross-disciplinary survey. *Machine Intelligence Research*, 19:89–114, 2022.

Saxena, A., Tripathi, A., and Talukdar, P. P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4498–4507, 2020.

Skopek, O., Ganea, O., and Bécigneul, G. Mixed-curvature variational autoencoders. In *8th International Conference on Learning Representations*, 2020.

Sommer, G. *Geometric computing with Clifford algebras: theoretical foundations and applications in computer vision and robotics*. Springer Science & Business Media, 2013.

Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., Shum, H.-Y., and Guo, J. Think-on-graph: Deep and responsible reasoning of large language model with knowledge graph. *arXiv preprint arXiv:2307.07697*, 2023.

Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations*, 2019.

Tabaghi, P. and Dokmanic, I. On procrustes analysis in hyperbolic space. *IEEE Signal Process. Lett.*, 28:1120–1124, 2021.

Toutanova, K. and Chen, D. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality*, pp. 57–66, 2015.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. Complex embeddings for simple link prediction. In *Proceedings of the 33nd International Conference on Machine Learning*, pp. 2071–2080, 2016.

Turaga, P. K. and Srivastava, A. *Riemannian computing in computer vision*, volume 1. Springer, 2016.

Vashishth, S., Sanyal, S., Nitin, V., and Talukdar, P. P. Composition-based multi-relational graph convolutional networks. In *8th International Conference on Learning Representations*, 2020.

Vorontsov, E., Trabelsi, C., Kadoury, S., and Pal, C. On orthogonality and learning recurrent networks with long term dependencies. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 3570–3578, 2017.

Wang, S., Wei, X., dos Santos, C. N., Wang, Z., Nallapati, R., Arnold, A. O., Xiang, B., Yu, P. S., and Cruz, I. F. Mixed-curvature multi-relational graph neural network for knowledge graph completion. In *Proceedings of the Web Conference*, pp. 1761–1771, 2021.

Wang, Z., Zhang, J., Feng, J., and Chen, Z. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1112–1119, 2014.

Wilson, R. C., Hancock, E. R., Pekalska, E., and Duin, R. P. W. Spherical and hyperbolic embeddings of data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2255–2269, 2014.

Yan, H., Li, C., Long, R., Yan, C., Zhao, J., Zhuang, W., Yin, J., Zhang, P., Han, W., Sun, H., Deng, W., Zhang, Q., Sun, L., Xie, X., and Wang, S. A comprehensive study on text-attributed graphs: Benchmarking and rethinking. In *Advances in Neural Information Processing Systems*, pp. 17238–17264, 2023.

Yang, B., Yih, W., He, X., Gao, J., and Deng, L. Embedding entities and relations for learning and inference in knowledge bases. In *3rd International Conference on Learning Representations*, 2015.

Yang, H.-R. and Ni, W. Continuous-time distributed heavy-ball algorithm for distributed convex optimization over undirected and directed graphs. *Machine Intelligence Research*, 19:75–88, 2022.

Yasunaga, M., Bosselut, A., Ren, H., Zhang, X., Manning, C. D., Liang, P. S., and Leskovec, J. Deep bidirectional language-knowledge graph pretraining. In *Advances in Neural Information Processing Systems*, pp. 37309–37323, 2022.

Zhang, A., Chan, A., Tay, Y., Fu, J., Wang, S., Zhang, S., Shao, H., Yao, S., and Lee, R. K. On orthogonality constraints for transformers. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 375–382, 2021.

Zhang, P., Guo, J., Li, C., Xie, Y., Kim, J., Zhang, Y., Xie, X., Wang, H., and Kim, S. Efficiently leveraging multi-level user intent for session-based recommendation via atten-mixer network. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pp. 168–176, 2023.

Zhang, Q., Wang, R., Yang, J., and Xue, L. Knowledge graph embedding by reflection transformation. *Knowl. Based Syst.*, 238:107861, 2022.

Zhang, S., Tay, Y., Yao, L., and Liu, Q. Quaternion knowledge graph embeddings. In *Advances in Neural Information Processing Systems*, pp. 2731–2741, 2019.

Zhao, J., Qu, M., Li, C., Yan, H., Liu, Q., Li, R., Xie, X., and Tang, J. Learning on large-scale text-attributed graphs via variational inference. In *The Eleventh International Conference on Learning Representations*, 2023.

Zhu, Z., Zhang, Z., Xhonneux, L. A. C., and Tang, J. Neural bellman-ford networks: A general graph neural network framework for link prediction. In *Advances in Neural Information Processing Systems*, pp. 29476–29490, 2021.

A. Glossary of Symbols

Table 6. Glossary of variables and symbols used in this paper.

Symbol	Shape	Description
$\mathcal{V}$	-	Set of entities
$\mathcal{R}$	-	Set of relations
$\mathcal{F}$	-	Set of factual triples
h, t	-	Head entity and tail entity
r	-	Relation type
$\mathbb{E}^k$	-	k-dimensional Euclidean space, equivalent to $\mathbb{R}^k$
$\mathbb{P}^k$	-	Elliptic subspace embedded in $\mathbb{R}^k$
$\mathbb{Q}^k$	-	Hyperbolic subspace embedded in $\mathbb{R}^k$
$\mathbb{D}^k$	-	Product manifold with total dimension k
$\mathbb{X}_i^{k_i}$	-	The i-th component space of product manifold $\mathbb{D}$
m	$\mathbb{R}_+$	Number of component spaces
β_r	$\mathbb{R}_+$	Param. of $r \in \mathcal{R}$ that determines the curvature of manifold
$\mathbf{p}_r$	$\mathbb{R}_+^k$	r-specific elliptic weighting vector
$\mathbf{q}$	$\mathbb{R}^k$	Hyperbolic weighting vector $(-1, 1, \ldots, 1)$
$\mathbf{e}_v$	$\mathbb{R}^k$	Embedding of entity $v \in \mathcal{V}$
$\mathbf{e}_{v,r}$	$\mathbb{R}^k$	r-associated representation of entity $v \in \mathcal{V}$
$\mathbf{b}_r$	$\mathbb{R}^k$	Param. of $r \in \mathcal{R}$ for hyperbolic boost
$\mathbf{U}_r$	$\mathbb{R}^{k\times k}$	Param. of $r \in \mathcal{R}$ for generalized orthogonal transformation

B. Proof of Theorem 3.1

This theorem is a special case of the "Cartan-Dieudonné Theorem" (Cartan, 1966; Dieudonné, 1967), which is one of the fundamental results of geometric algebra (Gallier, 2011). Since this result is not used in the remainder of this paper and the proof is somewhat intricate, we encourage the beginning reader to read the statement and then skip to the next part.

B.1. Proof of Theorem B.1 (Cartan-Dieudonné Theorem)

Theorem B.1. (Cartan-Dieudonné Theorem). *In the non-degenerate quadratic space V endowed with $\langle\cdot,\cdot\rangle_{\mathbf{w}}$, every element $\mathbf{G}$ of the orthogonal group $\mathbf{O}_{\mathbf{w}}(k)$ can be expressed as a product of at most k reflections.*

Proof. (1) The first step is to prove the following lemma: *Suppose that $\mathbf{G} \in \mathbf{O}_{\mathbf{w}}(k)$ satisfies such a condition: for every anisotropic vector $\mathbf{x}$, the vector $\mathbf{Gx} - \mathbf{x}$ is nonzero and isotropic*[5]*. Then, $k \geq 4$, k is even, and $\mathbf{G}$ is a rotation.*

It is evident that we cannot have $k = 1$. If $k = 2$, let $\mathbf{x}$ be an anisotropic vector. Since $\mathbf{Gx} - \mathbf{x}$ is isotropic and nonzero, $\mathbf{Gx}$ must be linearly independent from $\mathbf{x}$. This implies that the determinant of the quadratic form with respect to the basis $\mathbf{Gx} - \mathbf{x}$ and $\mathbf{x}$ is equal to 0, contradicting the non-degeneracy of the space. We therefore must have $k \geq 3$.

Since $\mathbf{Gx} - \mathbf{x}$ is isotropic for all anisotropic $\mathbf{x}$, we have $q(\mathbf{Gx} - \mathbf{x}) = \langle \mathbf{Gx} - \mathbf{x}, \mathbf{Gx} - \mathbf{x}\rangle_{\mathbf{w}} = 0$. In fact, this holds for all $\mathbf{x}$. To see this, let $\mathbf{y}$ be a nonzero isotropic vector. There is a plane containing $\mathbf{y}$ and splitting the space V (O'Meara, 2013), hence there is a vector $\mathbf{z}$ with $q(\mathbf{z}) \neq 0$ and $\langle \mathbf{y}, \mathbf{z}\rangle_{\mathbf{w}} = 0$. Then, we have $q(\mathbf{y} + \epsilon\mathbf{z}) \neq 0$ for any $\epsilon \in \mathbb{R}\backslash\{0\}$, hence

$$q(\mathbf{G}(\mathbf{y} + \epsilon\mathbf{z}) - (\mathbf{y} + \epsilon\mathbf{z})) = 0, \quad q(\mathbf{Gz} - \mathbf{z}) = 0.$$

It follows that

$$q(\mathbf{Gy} - \mathbf{y}) + 2\epsilon \langle \mathbf{Gy} - \mathbf{y}, \mathbf{Gz} - \mathbf{z}\rangle_{\mathbf{w}} = 0.$$

[5]A vector $\mathbf{x} \in V$ is said to be **isotropic** if $\langle \mathbf{x}, \mathbf{x}\rangle_{\mathbf{w}} = 0$ and **anisotropic** otherwise.

If we substitute $\epsilon = 1$, and then $\epsilon = -1$, and then add, we obtain $q(\mathbf{Gy} - \mathbf{y}) = 0$ as we asserted. Therefore, if we define the space $W := (\mathbf{G} - 1)V$, we have $q(\mathbf{y}) = 0$ for any $y \in W$, i.e., $q(W) = 0$. Then, the orthogonal complement of W is $W^{\perp} = \{\mathbf{x} \in V \mid \forall \mathbf{y} \in W, \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} = 0\}$. Now for any $\mathbf{x} \in V$ and $\mathbf{y} \in W^{\perp}$, we have

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{Gy} - \mathbf{y} \rangle_{\mathbf{w}} &= \langle \mathbf{Gx}, \mathbf{Gy} - \mathbf{y} \rangle_{\mathbf{w}} - \langle \mathbf{Gx} - \mathbf{x}, \mathbf{Gy} - \mathbf{y} \rangle_{\mathbf{w}} \\
&= \langle \mathbf{Gx}, \mathbf{Gy} - \mathbf{y} \rangle_{\mathbf{w}} \\
&= \langle \mathbf{Gx}, \mathbf{Gy} \rangle_{\mathbf{w}} - \langle \mathbf{Gx}, \mathbf{y} \rangle_{\mathbf{w}} \\
&= \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} - \langle \mathbf{Gx}, \mathbf{y} \rangle_{\mathbf{w}} \\
&= - \langle \mathbf{Gx} - \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} \\
&= 0.
\end{aligned}$$

Hence $\mathbf{Gy} - \mathbf{y}$ is perpendicular to all of V. Since V is non-degenerate, we conclude that $\mathbf{Gy} = \mathbf{y}$ for all $\mathbf{y} \in W^{\perp}$. Then, $q(W^{\perp}) = 0$ by the given condition. Thus we get $W = W^{\perp}$. Therefore, $k = \dim(V) = \dim(W) + \dim(W^{\perp}) = 2\dim(W)$ is even, hence at least 4. Moreover, since $\mathbf{G}$ acts as the identity on a maximal totally isotropic subspace of V (Clark, 2013), $\mathbf{G}$ is a rotation with the determinant equal to $+1$, completing Step 1.

(2) Now we can prove the theorem. The proof is by induction on k. For $k = 1$, the result is trivial, so let $k > 1$.

Case 1: Suppose there exists an anisotropic vector $\mathbf{x} \in V$ such that $\mathbf{Gx} = \mathbf{x}$. Then the restriction of $\mathbf{G}$ to the hyperplane ζ orthogonal to $\mathbf{x}$ is an element of $\mathbf{O_w}(k-1)$. By the inductive assumption, this restriction is a product of at most $k-1$ reflections taken with respect to lines in ζ. One can naturally view each of them as the reflection on all of V, and the same product of reflections agrees with $\mathbf{G}$ on ζ. Moreover, it also agrees with $\mathbf{G}$ on $\mathbf{x}$, since $\mathbf{G}$ and all of the reflections are equal to the identity on $\mathbf{x}$. Thus $\mathbf{G}$ is itself equal to the product of the at most $k-1$ reflections.

Case 2: Next suppose that there is an anisotropic vector $\mathbf{x}$ such that $\mathbf{Gx} - \mathbf{x}$ is anisotropic, i.e., $q(\mathbf{Gx} - \mathbf{x}) \neq 0$. To facilitate presentation, we denote the elementary reflection in Equation (7) as $\tau_{\mathbf{u}}$. In this way, we have

$$\begin{aligned}
\tau_{\mathbf{Gx}-\mathbf{x}}(\mathbf{Gx}) &= \mathbf{Gx} - 2\frac{\langle \mathbf{Gx}, \mathbf{Gx} - \mathbf{x} \rangle_{\mathbf{w}}}{\langle \mathbf{Gx} - \mathbf{x}, \mathbf{Gx} - \mathbf{x} \rangle_{\mathbf{w}}}(\mathbf{Gx} - \mathbf{x}) \\
&= \mathbf{Gx} - 2\frac{\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{w}} - \langle \mathbf{Gx}, \mathbf{x} \rangle_{\mathbf{w}}}{\langle \mathbf{Gx} - \mathbf{x}, \mathbf{Gx} - \mathbf{x} \rangle_{\mathbf{w}}}(\mathbf{Gx} - \mathbf{x}) \\
&= \mathbf{Gx} - \frac{\langle \mathbf{Gx} - \mathbf{x}, \mathbf{Gx} - \mathbf{x} \rangle_{\mathbf{w}}}{\langle \mathbf{Gx} - \mathbf{x}, \mathbf{Gx} - \mathbf{x} \rangle_{\mathbf{w}}}(\mathbf{Gx} - \mathbf{x}) \\
&= \mathbf{Gx} - (\mathbf{Gx} - \mathbf{x}) \\
&= \mathbf{x},
\end{aligned}$$

which implies that $\tau_{\mathbf{Gx}-\mathbf{x}}\mathbf{G}$ leaves $\mathbf{x}$ fixed. Based on the first case, the orthogonal matrix $\tau_{\mathbf{Gx}-\mathbf{x}}\mathbf{G}$ is a product of at most $k-1$ reflections, hence $\mathbf{G}$ is a product of at most k reflections.

Case 3: According to Case 1 and 2, any $\mathbf{G}$ that does not satisfy the condition in Step 1 is a product of at most k reflections. Now, let's consider the $\mathbf{G}$ that for every anisotropic vector $\mathbf{x} \in V$, $\mathbf{Gx} - \mathbf{x}$ is isotropic and nonzero. Based on Step 1, we conclude that k is even and $\mathbf{G}$ is a rotation matrix. Then, $\mathbf{G}' = \tau\mathbf{G}$ is a reflection that cannot satisfy the condition in Step 1. Hence $\mathbf{G}'$ can be expressed as a product of at most k reflections based on Case 1 and 2. Therefore, $\mathbf{G}$ is itself a product of at most $k+1$ reflections. However, $\mathbf{G}$ cannot be a product of $k+1$ reflections, since $k+1$ is odd and $\mathbf{G}$ is a rotation matrix. Hence $\mathbf{G}$ is a product of at most k reflections, qed. □

B.2. Proof of Theorem 3.1

Proof. First, the designed mapping Orth is a product of n elementary reflections, and its output is a generalized orthogonal matrix that satisfies the quadratic inner product invariance, i.e., $(\mathbf{H}_n \cdots \mathbf{H}_1)^{\top}\mathrm{diag}(\mathbf{w})(\mathbf{H}_n \cdots \mathbf{H}_1) = \mathrm{diag}(\mathbf{w})$. Therefore, we have $\mathrm{Image}(\mathrm{Orth}) \subseteq \mathbf{O_w}(k)$. Then, according to the Cartan-Dieudonné Theorem, every generalized orthogonal matrix $\mathbf{G} \in \mathbf{O_w}(k)$ can be expressed as $\prod_{c=1}^{z} \mathbf{H}(\mathbf{u}_c, \mathbf{w})$ for a certain $z \leq k$. Since the $k \times k$ identity matrix $\mathbf{I} \in \mathrm{Image}(\mathrm{Orth})$ for any value of n, we can get $\mathbf{G} = \mathbf{GI} = \prod_{c=1}^{k} \mathbf{H}(\mathbf{u}_c, \mathbf{w})$. Therefore, we also have $\mathbf{O_w}(k) \subseteq \mathrm{Image}(\mathrm{Orth})$ when $n = k$. On the whole, we conclude that $\mathrm{Image}(\mathrm{Orth}) = \mathbf{O_w}(k)$. □

C. Proof of Claim 3.2

Proof. We show that the elliptic parameterization is equivalent to Euclidean parameterization equipped with element-wise scaling transformations determined by $\sqrt{\mathbf{p}_r}$. Specifically, the distance function of elliptic parameterization in Equation (10) can be reformulated as follows:

$$\begin{aligned}
d_{\mathbb{P}}(\mathbf{e}'_{h,r}, \mathbf{e}_{t,r}) &= \sqrt{(\mathbf{e}'_{h,r} - \mathbf{e}_{t,r})^\top \mathrm{diag}(\mathbf{p}_r)(\mathbf{e}'_{h,r} - \mathbf{e}_{t,r})} \\
&= \sqrt{(\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}'_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r})^\top (\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}'_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r})} \\
&= \|\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}'_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r}\| \\
&= \|\mathrm{diag}(\sqrt{\mathbf{p}_r}) \prod_{i=1}^{k} \mathbf{H}(\mathbf{U}_r[i], \mathbf{p}_r)\mathbf{e}_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r}\| \\
&= \|\mathrm{diag}(\sqrt{\mathbf{p}_r})(\mathbf{I} - 2\frac{\mathbf{U}[k]\mathbf{U}[k]^\top \mathrm{diag}(\mathbf{p}_r)}{\mathbf{u}^\top \mathrm{diag}(\mathbf{p}_r)\mathbf{u}}) \prod_{i=1}^{k-1} \mathbf{H}(\mathbf{U}_r[i], \mathbf{p}_r)\mathbf{e}_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r}\| \\
&= \|(\mathbf{I} - 2\frac{\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{U}[k]\mathbf{U}[k]^\top \mathrm{diag}(\sqrt{\mathbf{p}_r})^\top}{\mathbf{U}[k]^\top \mathrm{diag}(\sqrt{\mathbf{p}_r})^\top \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{U}[k]})\mathrm{diag}(\sqrt{\mathbf{p}_r}) \prod_{i=1}^{k-1} \mathbf{H}(\mathbf{U}_r[i], \mathbf{p}_r)\mathbf{e}_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r}\| \\
&= \|\mathbf{H}(\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{U}_r[k], \mathbf{1})\mathrm{diag}(\sqrt{\mathbf{p}_r}) \prod_{i=1}^{k-1} \mathbf{H}(\mathbf{U}_r[i], \mathbf{p}_r)\mathbf{e}_{h,r} - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_{t,r}\| \\
&= \| \prod_{i=1}^{k} \mathbf{H}(\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{U}_r[i], \mathbf{1})\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_h - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_t\| \\
&= \|\mathrm{Orth}(\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{U}_r, \mathbf{1})\mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_h - \mathrm{diag}(\sqrt{\mathbf{p}_r})\mathbf{e}_t\|.
\end{aligned}$$

This expression implies that the head embedding $\mathbf{e}_h \in \mathbb{R}^k$ is first scaled by $\sqrt{\mathbf{p}_r}$, and then performs a Euclidean orthogonal transformation, with the expectation of being similar to the tail embedding $\mathbf{e}_t \in \mathbb{R}^k$, which is also scaled by $\sqrt{\mathbf{p}_r}$. □

D. Proof of Proposition 3.3

Proof. For any $\mathbf{G} \in \mathbf{O}_{\mathbf{q}}(k)$ and $\mathbf{x} \in \mathbb{Q}_\beta^k$, the hyperbolic orthogonal transformation $\mathbf{Gx}$ can be expressed in the following block form:

$$\mathbf{Gx} = \begin{bmatrix} a & \mathbf{b}^\top \\ \mathbf{c} & \mathbf{A} \end{bmatrix} \begin{bmatrix} x_1 \\ \mathbf{x}_{2:k} \end{bmatrix} = \begin{bmatrix} ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k} \\ x_1\mathbf{c} + \mathbf{A}\mathbf{x}_{2:k} \end{bmatrix},$$

where $a \in \mathbb{R}$, $\mathbf{b}, \mathbf{c} \in \mathbb{R}^{k-1}$ and $\mathbf{A} \in \mathbb{R}^{(k-1)\times(k-1)}$. We prove that (1) the absolute value of the first element of $\mathbf{G}$ is greater than or equal to 1, i.e., $|G_{11}| = |a| \geq 1$; (2) when $a \geq 1$ (or $a \leq -1$), the transformation $\mathbf{Gx}$ preserves (or reverses) the sign of the first element of $\mathbf{x}$, i.e., $\mathrm{sign}(x_1) = \mathrm{sign}(ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k})$ for $a \geq 1$ and $\mathrm{sign}(x_1) \neq \mathrm{sign}(ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k})$ for $a \leq -1$.

According to the definition of generalized orthogonal matrix in Equation (4), one can verify that the hyperbolic orthogonal matrix $\mathbf{G} \in \mathbf{O}_{\mathbf{q}}(k)$ satisfies the inner product invariance:

$$\mathbf{G}^\top \mathrm{diag}(\mathbf{q})\mathbf{G} = \mathrm{diag}(\mathbf{q}) \Rightarrow \mathbf{G}\mathrm{diag}(\mathbf{q})\mathbf{G}^\top = \mathrm{diag}(\mathbf{q}).$$

Since $\mathbf{q} = (-1, 1, \ldots, 1) \in \mathbb{R}^k$, the first element of $\mathbf{G}$ (i.e., $G_{11} = a$) requires that

$$a^2 - \|\mathbf{b}\|^2 = 1 \Rightarrow |a| = \sqrt{1 + \|\mathbf{b}\|^2} \geq 1.$$

We then consider the following square equation:

$$\begin{aligned}
(\frac{\mathbf{b}}{a} + \frac{\mathbf{x}_{2:k}}{x_1})^2 &= (\frac{\mathbf{b}}{a})^2 + 2\frac{\mathbf{b}^\top \mathbf{x}_{2:k}}{ax_1} + (\frac{\mathbf{x}_{2:k}}{x_1})^2 \\
&= (\frac{\|\mathbf{b}\|^2}{a^2}) + 2\frac{\mathbf{b}^\top \mathbf{x}_{2:k}}{ax_1} + (\frac{\|\mathbf{x}_{2:k}\|^2}{x_1^2}) \\
&= (\frac{a^2 - 1}{a^2}) + 2\frac{\mathbf{b}^\top \mathbf{x}_{2:k}}{ax_1} + (\frac{\|\mathbf{x}_{2:k}\|^2}{x_1^2}) \geq 0.
\end{aligned}$$

Since $\mathbf{x} \in \mathbb{Q}_\beta^k$, one can easily derive that $x_1^2 > \|\mathbf{x}_{2:k}\|^2$ according to the definition of hyperboloid in Equation (2). Based on this, we further have the following inequality:

$$-2\frac{\mathbf{b}^\top \mathbf{x}_{2:k}}{ax_1} \leq (\frac{a^2-1}{a^2}) + (\frac{\|\mathbf{x}_{2:k}\|^2}{x_1^2}) < (\frac{a^2-1}{a^2}) + 1 < 2.$$

When $a \geq +1$ and $x_1 > 0$, we have

$$2ax_1 > -2\mathbf{b}^\top \mathbf{x}_{2:k} \Rightarrow ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k} > 0.$$

When $a \geq +1$ and $x_1 < 0$, we have

$$2ax_1 < -2\mathbf{b}^\top \mathbf{x}_{2:k} \Rightarrow ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k} < 0.$$

When $a \leq -1$ and $x_1 > 0$, we have

$$2ax_1 < -2\mathbf{b}^\top \mathbf{x}_{2:k} \Rightarrow ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k} < 0.$$

When $a \leq -1$ and $x_1 < 0$, we have

$$2ax_1 > -2\mathbf{b}^\top \mathbf{x}_{2:k} \Rightarrow ax_1 + \mathbf{b}^\top \mathbf{x}_{2:k} > 0.$$

Therefore, we can observe that the hyperbolic orthogonal matrix $\mathbf{G}$ with its first element $G_{11} = a \geq +1$ preserves the sign of x_1, while $\mathbf{G}$ with $G_{11} \leq -1$ reverses the sign of x_1. Moreover, the set of all $\mathbf{G}$ with $G_{11} \geq +1$, namely the positive subset $\mathbf{O}_\mathbf{q}^+(k) = \{\mathbf{G} \in \mathbf{O}_\mathbf{q}(k) : G_{11} \geq +1\}$, forms a multiplicative subgroup of $\mathbf{O}_\mathbf{q}(k)$, since (1) for any $\mathbf{G}_1, \mathbf{G}_2 \in \mathbf{O}_\mathbf{q}^+(k)$, their multiplication also preserves the quadratic inner product of $\mathbf{x}$ and the sign of x_1, i.e., $\mathbf{G}_1\mathbf{G}_2 \in \mathbf{O}_\mathbf{q}^+(k)$; (2) $\mathbf{O}_\mathbf{q}^+(k)$ contains the identity matrix $\mathbf{I}$; (3) every matrix $\mathbf{G} \in \mathbf{O}_\mathbf{q}^+(k)$ has an inverse $\mathbf{G}^{-1} \in \mathbf{O}_\mathbf{q}^+(k)$ such that $\mathbf{G}\mathbf{G}^{-1} = \mathbf{I}$. □

E. Proof of Proposition 3.4

Proof. The positive hyperbolic orthogonal matrix $\mathbf{G} \in \mathbf{O}_\mathbf{q}^+(k)$ can be written in block form as follows:

$$\mathbf{G} = \begin{bmatrix} a & \mathbf{b}^\top \\ \mathbf{c} & \mathbf{A} \end{bmatrix},$$

where $a \geq +1$, $\mathbf{b}, \mathbf{c} \in \mathbb{R}^{k-1}$ and $\mathbf{A} \in \mathbb{R}^{(k-1)\times(k-1)}$. Since $\mathbf{G}^\top \mathrm{diag}(\mathbf{q})\mathbf{G} = \mathbf{G}\mathrm{diag}(\mathbf{q})\mathbf{G}^\top = \mathrm{diag}(\mathbf{q})$, we get

$$\begin{bmatrix} a & \mathbf{c}^\top \\ \mathbf{b} & \mathbf{A}^\top \end{bmatrix} \begin{bmatrix} -a & -\mathbf{b}^\top \\ \mathbf{c} & \mathbf{A} \end{bmatrix} = \begin{bmatrix} a & \mathbf{b}^\top \\ \mathbf{c} & \mathbf{A} \end{bmatrix} \begin{bmatrix} -a & -\mathbf{c}^\top \\ \mathbf{b} & \mathbf{A}^\top \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{I} \end{bmatrix},$$

then

$$\begin{aligned} \mathbf{A}^\top \mathbf{A} &= \mathbf{I} + \mathbf{b}\mathbf{b}^\top, \\ \mathbf{c}^\top \mathbf{c} &= a^2 - 1, \\ \mathbf{A}^\top \mathbf{c} &= a\mathbf{b}, \\ \mathbf{b}^\top \mathbf{b} &= a^2 - 1, \\ \mathbf{A}\mathbf{b} &= a\mathbf{c}. \end{aligned}$$

From $\mathbf{A}^\top \mathbf{A} = \mathbf{I} + \mathbf{b}\mathbf{b}^\top$, we get that $\mathbf{A}^\top \mathbf{A}$ is symmetric and positive definite. Geometrically, it is well known that $\frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top \mathbf{b}}$ is the orthogonal projection onto the line determined by $\mathbf{b}$. Consequently, $\mathbf{b}\mathbf{b}^\top$ has the eigenvalue 0 with multiplicity $k-2$ and the eigenvalue $\mathbf{b}^\top \mathbf{b} = a^2 - 1$ with multiplicity 1. The eigenvectors associated with 0 are orthogonal to $\mathbf{b}$ and the eigenvectors associated with $a^2 - 1$ are proportional with $\mathbf{b}$. It follows that $\mathbf{I} + \mathbf{b}\mathbf{b}^\top$ has the eigenvalue 1 with multiplicity $k-2$ and the eigenvalue a^2 with multiplicity 1, the eigenvectors being as before. Now, $\mathbf{A}$ has the polar form $\mathbf{A} = \mathbf{Q}\mathbf{S}$, where $\mathbf{Q}$ is a Euclidean orthogonal matrix, $\mathbf{S}$ is a positive semi-definite symmetric matrix and $\mathbf{S}^2 = \mathbf{S}^\top \mathbf{S} = \mathbf{A}^\top \mathbf{A} = \mathbf{I} + \mathbf{b}\mathbf{b}^\top$. Since $a \geq +1$, $\mathbf{S} = \sqrt{\mathbf{I} + \mathbf{b}\mathbf{b}^\top}$ has the eigenvalue 1 with multiplicity $k-2$ and the eigenvalue a with multiplicity 1, the eigenvectors being as before. Then, since $\mathbf{b}$ is an eigenvector of $\mathbf{S}$ for the eigenvalue a, we have

$$\mathbf{A}\mathbf{b} = \mathbf{Q}\mathbf{S}\mathbf{b} = \mathbf{Q}(\mathbf{S}\mathbf{b}) = \mathbf{Q}(a\mathbf{b}) = a\mathbf{Q}\mathbf{b}.$$

Since we also have $\mathbf{A}\mathbf{b} = a\mathbf{c}$, this implies

$$a\mathbf{Q}\mathbf{b} = a\mathbf{c} \Rightarrow \mathbf{Q}\mathbf{b} = \mathbf{c}.$$

It follows that

$$\mathbf{G} = \begin{bmatrix} a & \mathbf{b}^\top \\ \mathbf{c} & \mathbf{A} \end{bmatrix} = \begin{bmatrix} a & \mathbf{b}^\top \\ \mathbf{Q}\mathbf{b} & \mathbf{Q}\mathbf{S} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \sqrt{\|\mathbf{b}\|^2 + 1} & \mathbf{b}^\top \\ \mathbf{b} & \sqrt{\mathbf{I} + \mathbf{b}\mathbf{b}^\top} \end{bmatrix}.$$

Based on Theorem 3.1, we can further parameterize the Euclidean orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{(k-1)\times(k-1)}$ as the composition of $k-1$ Euclidean Householder reflections:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{Orth}(\mathbf{U}, \mathbf{1}) \end{bmatrix} \begin{bmatrix} \sqrt{\|\mathbf{b}\|^2 + 1} & \mathbf{b}^\top \\ \mathbf{b} & \sqrt{\mathbf{I} + \mathbf{b}\mathbf{b}^\top} \end{bmatrix}.$$

□

F. Proof of Claim 3.5

Proof. Our proposed GoldE framework models each relation r as a generalized orthogonal transformation between entities. To facilitate presentation, we denote the r-specific generalized orthogonal matrix derived from Equation (14) as $\mathbf{G}_r$:

$$\mathbf{G}_r^{(i)} = \begin{cases} \mathrm{Orth}_{\mathbb{P}}(\mathbf{U}_r^{(i)}, \mathbf{p}_r^{(i)}), & \text{for } \mathbb{X}_i = \mathbb{P} \\ \mathrm{Orth}_{\mathbb{Q}}(\mathbf{U}_r^{(i)}, \mathbf{b}_r^{(i)}), & \text{for } \mathbb{X}_i = \mathbb{Q} \end{cases}.$$

Due to the orthogonality of the relation matrix and the geometric unification of the embedding space, GoldE is capable of simultaneously capturing the crucial logical patterns and the inherent topological structures in KGs. For simplicity, we omit the indices (i) of component spaces in the following proofs.

Symmetry/antisymmetry pattern: A relation r is symmetric (antisymmetric) if $\forall x, y$

$$r(x, y) \Rightarrow r(y, x) \quad (r(x, y) \Rightarrow \neg r(y, x)).$$

If $r(x, y)$ and $r(y, x)$ hold, we have

$$\mathbf{e}_y = \mathbf{G}_r\mathbf{e}_x \wedge \mathbf{e}_x = \mathbf{G}_r\mathbf{e}_y \Rightarrow \mathbf{G}_r\mathbf{G}_r = \mathbf{I}.$$

Otherwise, if $r(x, y)$ and $\neg r(y, x)$ hold, we have

$$\mathbf{e}_y = \mathbf{G}_r\mathbf{e}_x \wedge \mathbf{e}_x \neq \mathbf{G}_r\mathbf{e}_y \Rightarrow \mathbf{G}_r\mathbf{G}_r \neq \mathbf{I}.$$

Inversion pattern: A relation r_1 is inverse to relation r_2 if $\forall x, y$

$$r_2(x, y) \Rightarrow r_1(y, x).$$

If $r_1(x, y)$ and $r_2(y, x)$ hold, we have

$$\mathbf{e}_y = \mathbf{G}_{r_1}\mathbf{e}_x \wedge \mathbf{e}_x = \mathbf{G}_{r_2}\mathbf{e}_y \Rightarrow \mathbf{G}_{r_1} = \mathbf{G}_{r_2}^\top.$$

Composition pattern: A relation r_1 is composed of relation r_2 and relation r_3 if $\forall x, y, z$

$$r_2(x, y) \wedge r_3(y, z) \Rightarrow r_1(x, z).$$

If $r_1(x, z), r_2(x, y)$ and $r_3(y, z)$ hold, we have

$$\mathbf{e}_z = \mathbf{G}_{r_1}\mathbf{e}_x \wedge \mathbf{e}_y = \mathbf{G}_{r_2}\mathbf{e}_x \wedge \mathbf{e}_z = \mathbf{G}_{r_3}\mathbf{e}_y \Rightarrow \mathbf{G}_{r_1} = \mathbf{G}_{r_3}\mathbf{G}_{r_2}.$$

Cyclical and hierarchical structures: The fidelity of representations arises from the correspondence between the structure of the data and the geometry of the space (Gu et al., 2019). Our GoldE is endowed with a product manifold composed of multiple model spaces, enabling improved representations by better matching the geometry of the embedding space to the heterogeneous structures of KGs. On the one hand, the orthogonal relation transformations in the elliptic component spaces are naturally suitable for capturing the cyclical structures (Wilson et al., 2014; Liu et al., 2017). On the other hand, the hyperbolic orthogonal transformation in Equation (12) inherently encodes the hierarchical structures—the Euclidean orthogonal matrix models the transformation between entities at the same level of hierarchies, and the hyperbolic boost matrix models the transformation between entities at different levels of hierarchies (Tabaghi & Dokmanic, 2021). □

G. Efficient Computation

The training time cost of GoldE primarily stems from the elliptic and hyperbolic orthogonal relation transformations in Equation (9) and (12), where k matrix-vector multiplications incur the time complexity of $O(k^3)$. We show that these matrix-vector multiplications can be entirely replaced by vector-vector operations to achieve efficient computation.

For the elliptic orthogonal relation transformations in Equation (9), the entity embedding $\mathbf{e}_{h,r}$ iteratively performs k elliptic Householder reflections. Each iteration can be achieved by vector-vector operations as follows:

$$\mathbf{e}_{h,r}^{i+1} = \mathbf{H}(\mathbf{U}_r[i], \mathbf{p}_r)\mathbf{e}_{h,r}^{i} = (\mathbf{I} - 2\frac{\mathbf{U}_r[i]\mathbf{U}_r[i]^\top \mathrm{diag}(\mathbf{p}_r)}{\mathbf{U}_r[i]^\top \mathrm{diag}(\mathbf{p}_r)\mathbf{U}_r[i]})\mathbf{e}_{h,r}^{i} = \mathbf{e}_{h,r}^{i} - 2\frac{\langle \mathbf{U}_r[i], \mathbf{e}_{h,r}^{i}\rangle_{\mathbf{p}_r}}{\langle \mathbf{U}_r[i], \mathbf{U}_r[i]\rangle_{\mathbf{p}_r}}\mathbf{U}_r[i], \tag{17}$$

where $\mathbf{e}_{h,r}^{1}$ is the initial entity embedding $\mathbf{e}_{h,r}$. In this way, the time complexity of Equation (9) is reduced to $O(k^2)$.

For the hyperbolic orthogonal relation transformations in Equation (12), the entity embedding $\mathbf{e}_{h,r}$ is transformed by a hyperbolic boost matrix and a Euclidean orthogonal matrix. We first replace the hyperbolic boost matrix with its equivalent form (Barrett, 2011) to avoid calculating the matrix square root, and then accelerate the transformation via block matrix multiplication. Formally, the hyperbolic boost transformation can be achieved by vector-vector operations as follows:

$$\begin{bmatrix} \sqrt{\|\mathbf{b}_r\|^2+1} & \mathbf{b}_r^\top \\ \mathbf{b}_r & \sqrt{\mathbf{I}+\mathbf{b}_r\mathbf{b}_r^\top} \end{bmatrix}\mathbf{e}_{h,r} = \begin{bmatrix} \gamma & -\gamma\mathbf{v}_r^\top \\ -\gamma\mathbf{v}_r & \mathbf{I}+\frac{\gamma^2}{1+\gamma}\mathbf{v}_r\mathbf{v}_r^\top \end{bmatrix}\begin{bmatrix} e_{h,r}^{1} \\ \mathbf{e}_{h,r}^{2:k} \end{bmatrix} = \begin{bmatrix} \gamma e_{h,r}^{1} - \gamma\langle \mathbf{v}_r, \mathbf{e}_{h,r}^{2:k}\rangle_{\mathbf{1}} \\ -\gamma e_{h,r}^{1}\mathbf{v}_r + \mathbf{e}_{h,r}^{2:k} + \frac{\gamma^2\langle \mathbf{v}_r, \mathbf{e}_{h,r}^{2:k}\rangle_{\mathbf{1}}}{1+\gamma}\mathbf{v}_r \end{bmatrix}, \tag{18}$$

where $\mathbf{v}_r \in \mathbb{R}^{k-1}$, $\|\mathbf{v}_r\| < 1$, $\gamma = \frac{1}{\sqrt{1-\|\mathbf{v}_r\|^2}}$, $e_{h,r}^{1}$ and $\mathbf{e}_{h,r}^{2:k}$ are the first and the remaining $k-1$ entries of $e_{h,r}$, respectively. The subsequent Euclidean orthogonal transformation can also be replaced by vector-vector operations, which is analogous to Equation (17). Therefore, the time complexity of Equation (12) is also reduced to $O(k^2)$.

Table 7 shows the convergence time required for training the models on three standard benchmarks. We select RotatE (for its simplicity) and HousE (for its advanced performance) as the baselines. By using the efficient computation, our GoldE framework costs comparable training time to these two models. Combined with the link prediction results in Table 2, one can see that the GoldE is capable of achieving superior effectiveness without sacrificing the efficiency.

Table 7. Training time of RotatE, HousE and our GoldE on three datasets.

Model	WN18RR	FB15k-237	YAGO3-10
RotatE	4h	6h	10h
HousE	2h	3h	11h
GoldE	2h	5h	13h

H. Datasets

Table 8 summarizes the detailed statistics of three benchmark datasets.

WN18RR (Dettmers et al., 2018) and FB15k-237 (Toutanova & Chen, 2015) datasets are subsets of WN18 (Bordes et al., 2013) and FB15k (Bordes et al., 2013) respectively with inverse relations removed to avoid test leakage. WN18 is extracted from WordNet (Miller, 1995), a database featuring lexical relations between words. FB15k is extracted from Freebase (Bollacker et al., 2008), a large-scale KG containing general knowledge facts.

YAGO3-10 is a subset of YAGO3 (Mahdisoltani et al., 2015), containing 123,182 entities and 37 relations. Most of the triples in YAGO3-10 are descriptive attributes of people, such as citizenship, gender, profession and marital status.

Table 8. Statistics of five standard benchmarks.

Dataset	#Entity	#Relation	#Training	#Validation	#Test
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466
YAGO3-10	123,182	37	1,079,040	5,000	5,000

I. Implementation Details

Evaluation: We follow the filtered ranking protocol (Bordes et al., 2013) for evaluation. For a test triple (h, r, t), we rank it against all negative triplets (h, r, t') or (h', r, t) that do not appear in the knowledge graph. We report mean rank (MR), mean reciprocal rank (MRR) and HITS at N (H@N) as performance metrics.

Fair comparisons: To ensure fair comparisons, we control the total number of parameters of GoldE to be similar to the baselines as shown in Table 9. Specifically, we follow (Li et al., 2022) to fix the embedding size k of each entity as 800, 600, 1000 on WN18RR, FB15k-237 and YAGO3-10 datasets, respectively.

Table 9. Comparison of the number of parameters. The results of baselines are taken from the original papers.

Dataset	RotatE	Rotate3D	DualE	HousE	GoldE
WN18RR	40.95M	61.44M	32.76M	32.57M	31.23M
FB15k-237	29.32M	44.57M	11.64M	12.13M	15.93M
YAGO3-10	123.18M	-	-	122.91M	118.52M

Selection of Product Manifold: For reducing the number of hyperparameters, we set identical dimensions for component spaces of the same geometric type to describe the product space as $\mathbb{D}^k = \bigtimes_{i=1}^{m_{\mathbb{P}}} \mathbb{P}_i^{k_{\mathbb{P}}} \times \bigtimes_{j=1}^{m_{\mathbb{Q}}} \mathbb{Q}_j^{k_{\mathbb{Q}}}$, where $m_{\mathbb{P}}$ and $m_{\mathbb{Q}}$ denote the number of elliptic and hyperbolic component spaces, $k_{\mathbb{P}}$ and $k_{\mathbb{Q}}$ denote the dimension of elliptic and hyperbolic component spaces. We omit $\mathbb{E}$ since the Euclidean orthogonal parameterization is the special case of our proposed elliptic orthogonal parameterization according to Claim 3.2. Considering that the embedding size k is fixed to ensure fair comparisons, we can define the product manifold $\mathbb{D}^k$ with only three hyperparameters: the embedding size of one partition $k_* = m_{\mathbb{P}} k_{\mathbb{P}}$, the number of elliptic component spaces $m_{\mathbb{P}}$, and the number of hyperbolic component spaces $m_{\mathbb{Q}}$. Note that the selected product manifold may not be the optimal case due to the simplified dimension configuration. One potential direction is to design a signature[6] estimator (Gu et al., 2019) for KGs, and we leave this as the future work.

Hyperparameters: We use Adam (Kingma & Ba, 2015) as the optimizer and fine-tune the hyperparameters on the validation dataset. The hyperparameters are tuned by the random search (Bergstra & Bengio, 2012), including batch size b, self-adversarial sampling temperature α, fixed margin γ, learning rate lr, dimension k_*, number of elliptic component spaces $m_{\mathbb{P}}$, and number of hyperbolic component spaces $m_{\mathbb{Q}}$. The hyperparameter search space is shown in Table 10.

Table 10. Hyperparameter search space.

Hyperparameter	Search Space	Type
b	$\{500, 800, 1000, 1500, 2000\}$	Choice
α	$[0.5, 2.0]$	Range
γ	$\{6, 8, 10, 12, 16, 20, 24, 28\}$	Choice
lr	$\{0.0001, 0.0005, 0.001, 0.003\}$	Choice
k_*	$\{0, 200, 400, 600, 800\}$	Choice
$m_{\mathbb{P}}, m_{\mathbb{Q}}$	$\{2, 4, 8, 10, 15, 20, 25, 30\}$	Choice

J. More Results of GoldE with Varying Embedding Size

Table 11 shows the full results of GoldE with embedding size k equal to 32. Such low-dimensional setting follows (Chami et al., 2020). One can observe that our GoldE framework consistently outperforms the baselines on all metrics across the three datasets, demonstrating the superior modeling capability of our designed universal orthogonal parameterization.

Table 11. Full results of GoldE with embedding size k equal to 32.

	WN18RR				FB15k-237				YAGO3-10			
Model	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
RotatE	.387	.330	.417	.491	.290	.208	.316	.458	-	-	-	-
QuatE	.421	.396	.430	.467	.293	.212	.320	.460	-	-	-	-
MuRE	.458	.421	.471	.525	.313	.226	.340	.489	.283	.187	.317	.478
MuRP	.465	.420	.484	.544	.323	.235	.353	.501	.230	.150	.247	.392
RefH	.447	.408	.464	.518	.312	.224	.342	.489	.381	.302	.415	.530
RotH	.472	.428	.490	.553	.314	.223	.346	.497	.393	.307	.435	.559
AttH	.466	.419	.484	.551	.324	.236	.354	.501	.397	.310	.437	.566
GoldE	**.482**	**.447**	**.493**	**.558**	**.335**	**.246**	**.360**	**.506**	**.420**	**.330**	**.465**	**.586**

[6]The signature of product manifold refers to the number of component spaces of each geometric type and their dimensions.

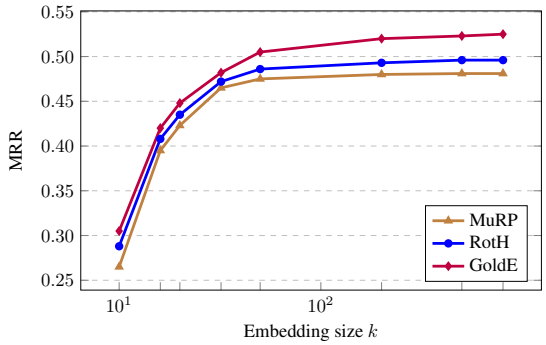


Figure 4. MRR results of the models with embedding size $k \in \{10, 16, 20, 32, 50, 200, 500, 800\}$ on WN18RR.

To delve deeper into the impact of different embedding sizes, we further conduct experiments for GoldE with varying values of k. The parameters in all cases are tuned, and the results are averaged over 5 runs with different random initializations. Figure 4 exhibits the MRR performance curves on WN18RR. It reveals that GoldE consistently surpasses the baselines across a broad range of embedding sizes, comprehensively confirming the effectiveness of our proposal.

K. Limitations and Future Work

The proposed GoldE framework follows the traditional learning paradigm of KGE (Li et al., 2023). Specifically, GoldE (as well as existing KGE approaches) models each training triple independently to implicitly capture the multi-hop logical patterns. However, the single-triple constraints inevitably have biases during optimization. These biases will accumulate with the increase of relation hops and make the perceived logical patterns unreliable, thereby hindering the generalization ability. A possible solution is to combine our framework with a path/graph encoder (Li et al., 2023; Zhang et al., 2023) to simultaneously process multiple structurally-dependent training triples. Furthermore, GoldE purely accesses the graph structures for symbolic inference. We plan to incorporate the textual attributes (Li et al., 2017; Yan et al., 2023; Zhao et al., 2023) of entities and relations in our future work to enhance the quality of learned representations. In addition, we would also like to investigate other theories and methodologies (Yang & Ni, 2022; Liu et al., 2023; Li et al., 2024; Hu et al., 2024; Feng et al., 2024) to unveil their potential significance in the realm of knowledge mining (Rui et al., 2022).