HousE: Knowledge Graph Embedding with Householder Parameterization

Rui Li [1] [*]   **Jianan Zhao** [2]   **Chaozhuo Li** [3]   **Di He** [4]   **Yiqi Wang** [5]   **Yuming Liu** [6]
Hao Sun [6]   **Senzhang Wang** [7]   **Weiwei Deng** [6]   **Yanming Shen** [1]   **Xing Xie** [3]   **Qi Zhang** [6]

Abstract

The effectiveness of knowledge graph embedding (KGE) largely depends on the ability to model intrinsic relation patterns and mapping properties. However, existing approaches can only capture some of them with insufficient modeling capacity. In this work, we propose a more powerful KGE framework named HousE, which involves a novel parameterization based on two kinds of Householder transformations: (1) *Householder rotations* to achieve superior capacity of modeling relation patterns; (2) *Householder projections* to handle sophisticated relation mapping properties. Theoretically, HousE is capable of modeling crucial relation patterns and mapping properties simultaneously. Besides, HousE is a generalization of existing rotation-based models while extending the rotations to high-dimensional spaces. Empirically, HousE achieves new state-of-the-art performance on five benchmark datasets. Our code is available at https://github.com/anrep/HousE.

1. Introduction

Knowledge graphs (KGs) store massive human knowledge as a collection of factual triples, where each triple (h, r, t) represents a relation r between head entity h and tail entity t. With a wealth of human knowledge, KGs have demonstrated their effectiveness in a myriad of downstream applications (Xiong et al., 2017). However, real-world KGs such as Freebase (Bollacker et al., 2008) and Yago (Suchanek et al., 2007)) usually suffer from incompleteness. Knowledge

---
[*]Work done during internship at MSRA. [1]Department of Computer Science and Technology, Dalian University of Technology, Dalian, China [2]University of Notre Dame, Indiana, USA [3]Microsoft Research Asia, Beijing, China [4]Peking University, Beijing, China [5]Michigan State University, Michigan, USA [6]Microsoft, Beijing, China [7]Central South University, Changsha, China. Correspondence to: Yanming Shen <shen@dlut.edu.cn>, Chaozhuo Li <cli@microsoft.com>.
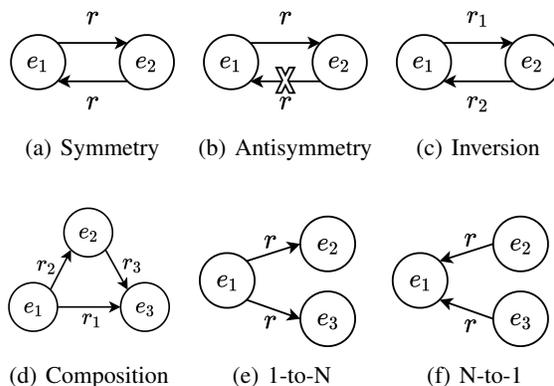
Figure 1. Illustrations of four relation patterns (a-d) (Sun et al., 2019) and two challenging RMPs (e-f) (Bordes et al., 2013).

Graph Embedding (KGE), which learns low-dimensional representations for entities and relations, excels as an effective tool for predicting missing links.

A crucial challenge of KGE lies in how to model relation patterns (e.g., symmetry, antisymmetry, inversion and composition) and relation mapping properties (RMPs, i.e., 1-to-1, 1-to-N, N-to-1 and N-to-N) (Bordes et al., 2013; Sun et al., 2019) as shown in Figure 1. Most works design specific vector spaces and operations to capture such patterns and RMPs. For example, TransE (Bordes et al., 2013) represents relations as translations, which fails in modeling symmetry and RMPs. Recently, RotatE (Sun et al., 2019) represents relations as rotations in the complex plane to model the four relation patterns, but it is incapable of handling RMPs due to the distance-preserving property of rotations. Rotate3D (Gao et al., 2020) and QuatE (Zhang et al., 2019) introduce quaternions to extend rotations to 3-dimensional and 4-dimensional spaces, and achieve better performance with larger model capacity.

However, as far as we know, none of the existing methods is capable of modeling all the relation patterns and RMPs as shown in Table 1, leading to the sub-optimal performance. Furthermore, some advanced approaches, such as (Sun et al., 2019; Gao et al., 2020; Zhang et al., 2019), are specifically designed on 2,3,4 dimensional spaces, which may be inadequate to capture the sophisticated structures of KGs (Zhang et al., 2019). Therefore, this brings us a question: *is there*

Table 1. Recent models' capability of modeling relation patterns and relation mapping properties (RMPs). TransX represents many TransE's variants, such as TransH (Wang et al., 2014), TransR (Lin et al., 2015) and TransD (Ji et al., 2015).

Model	Symmetry	Antisymmetry	Inversion	Composition	RMPs	Dim. of Rotation
TransE	✗	✔	✔	✔	✗	-
TransX	✔	✔	✗	✗	✔	-
DistMult	✔	✗	✗	✗	✔	-
ComplEx	✔	✔	✔	✗	✔	-
RotatE	✔	✔	✔	✔	✗	2-D
Rotate3D	✔	✔	✔	✔	✗	3-D
QuatE	✔	✔	✔	✗	✔	4-D
DualE	✔	✔	✔	✗	✔	3-D
HousE-r	✔	✔	✔	✔	✗	k-D
HousE	✔	✔	✔	✔	✔	k-D

a framework to handle all the above relation patterns and RMPs with more powerful modeling capacity?

In this paper, we give an affirmative answer by proposing a more powerful and general framework named HousE based on Householder parameterization. We prove that the composition of $2\lfloor\frac{k}{2}\rfloor$ Householder reflections (Householder, 1958) can represent any k-dimensional rotations. This unique property of Householder reflections provides us a natural way to model high-dimensional rotations with more degree of freedom. We call this kind of rotations as *Householder rotations*, based on which a simple model named HousE-r is proposed to achieve superior capacity of modeling relation patterns. Nevertheless, HousE-r is plagued by the sophisticated RMPs due to the distance-preserving nature of pure Householder rotations. To remedy this deficiency, we modify the vanilla Householder reflections to *Householder projections*, which can flexibly adjust the relative distances between points. The Householder projections are further integrated with HousE-r to establish the final HousE. By enjoying the merits of Householder rotations and Householder projections, HousE is theoretically capable of modeling all the relation patterns and RMPs shown in Table 1. Moreover, our proposal is a general framework and existing rotation-based models can be viewed as the special cases of HousE. Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to utilize Householder parameterization to build a more powerful and general KGE framework named HousE.

- We present a simple way to represent relations as high-dimensional Householder rotations, which endows HousE with better modeling capacity.

- We further modify the vanilla Householder reflections to Householder projections. By combining the Householder projections and rotations, HousE is able to model all the relation patterns and RMPs in Table 1.

- We conduct extensive experiments over five benchmarks and our proposal consistently outperforms SOTA baselines over all the datasets.

2. Problem Setup

Given the entity set $\mathcal{E}$ and relation set $\mathcal{R}$, a knowledge graph can be formally defined as a collection of factual triples $\mathcal{D} = \{(h, r, t)\}$, in which head/tail entities $h, t \in \mathcal{E}$ and relation $r \in \mathcal{R}$. To predict missing links, KGE maps entities and relations to distributed representations, and defines a score function to measure the plausibility of each triple.

Following a series of previous works (Bordes et al., 2013; Sun et al., 2019; Gao et al., 2020; Song et al., 2021), we define the score function as a distance function $d_r(h, t)$. The distance of the positive triple $(h, r, t) \in \mathcal{D}$ is expected to be smaller than the corrupted negative triples (h', r, t) or (h, r, t'), which can be generated by randomly replacing the entity h or t with other entities.

In the training process, the self-adversarial negative sampling (Sun et al., 2019) is used to optimize the models in a contrastive way. Given a positive triple and its negative samples, the loss function is defined as follows:

$$
\begin{aligned}
L = & -\log \sigma(\gamma - d_r(h, t)) \\
& - \sum_{i=1}^{l} p(h'_i, r, t'_i) \log \sigma(d_r(h'_i, t'_i) - \gamma) \\
& + \frac{\lambda}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \|e\|_2^2,
\end{aligned}
\tag{1}
$$

where γ is a pre-defined margin, σ is the sigmoid function, l denotes the number of negative samples, (h'_i, r, t'_i) is a negative sample against (h, r, t), $p(h'_i, r, t'_i)$ is the weight of negative sampling defined in (Sun et al., 2019), λ is the regularization coefficient. For the sake of clarification, notations used in this paper are listed in Appendix A.
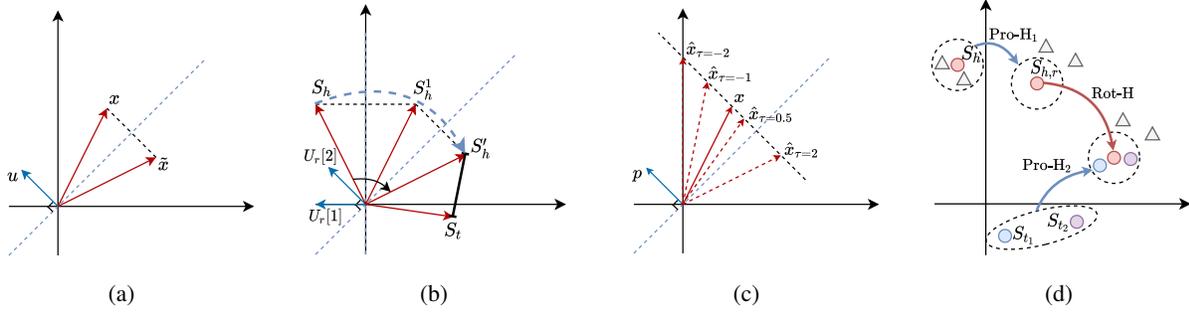
Figure 2. (a) Householder reflection in 2-dimensional space; (b) HousE-r models relation r as a 2-dimensional Householder rotation composed of two Householder reflections; (c) Modified Householder reflection in 2-dimensional space with different values of τ; (d) To model (h, r, t_1) and (h, r, t_2), HousE first utilizes relational Householder projections Pro-H$_1$ and Pro-H$_2$ (blue lines) to change the relative distance between entities, such as increasing the distance between S_h and the negative samples (Marked by triangles) and decreasing the distance between two positive tail entities S_{t_1} and S_{t_2}, then HousE performs a relational Householder rotation Rot-H (red lines) from projected head embedding $S_{h,r}$ to the projected tail embedding. Note that we omit the row (element) indices $[i]$ for simplicity.

3. Methodology

3.1. HousE-r: Relational Householder Rotations

In the first step, we seek to develop a general framework to model relations as rotations in the space of any dimension k, going beyond (Sun et al., 2019; Gao et al., 2020; Zhang et al., 2019), for better modeling capacity. In order to parameterize a k-dimensional rotation matrix, a straight-forward strategy is to randomly initialize a matrix and restrict it to a rotation matrix after every gradient descent update. However, such a naive approach may lead to the complicated optimization process and cannot fully cover the set of all $k \times k$ rotation matrices. In this paper, we theoretically prove that any k-dimensional rotations can be represented as $2\lfloor \frac{k}{2} \rfloor$ Householder reflections (Householder, 1958). Inspired by this theorem, we design an elegant parameterization based on Householder reflections to model k-dimensional rotations without any special optimizing procedure.

As the basic mathematical operator in this work, *Householder matrix* (Householder, 1958) represents the reflection (*Householder reflection*) about a hyperplane containing the origin. Given a unit vector $u \in \mathbb{R}^k$, the $k \times k$ Householder matrix H, taking u as variable, is defined as $H(u)$:

$$H(u) = I - 2uu^\top, \qquad (2)$$

where $\|u\|_2^2 = 1$ and I is the $k \times k$ identity matrix. Geometrically, as shown in Figure 2(a), the Householder matrix transforms x to $\widetilde{x}$ by a reflection about the hyperplane orthogonal to the normal vector u:

$$\widetilde{x} = H(u)x = x - 2\langle x, u \rangle u, \qquad (3)$$

where $\langle \cdot \rangle$ denotes the dot product.

Based on the Householder matrices, we can design a mapping to represent rotations. Specifically, given a series of

unit vectors $U = \{u_c\}_{c=1}^{2n}$ where $u_c \in \mathbb{R}^k$ and n is a positive integer, we define the mapping as follows:

$$\text{Rot-H}(U) = \prod_{c=1}^{2n} H(u_c). \qquad (4)$$

The output of Rot-H is a $k \times k$ orthogonal matrix with determinant 1, i.e., a rotation matrix (Artin, 2016), since each Householder matrix $H(u_c)$ is orthogonal and its determinant is -1. Moreover, we also prove that any rotation can be expressed as the composition of Householder reflections.

Formally, we have the following theorem:

Theorem 3.1. *When* $n = \lfloor \frac{k}{2} \rfloor$, *the image of* Rot-H *is the set of all* $k \times k$ *rotation matrices, i.e.,* $\text{Image}(\text{Rot-H}) = \mathbf{SO}(k)$, $\mathbf{SO}(k)$ *is the* k-*dimensional special orthogonal group. (See proof in Appendix B)*

Theorem 3.1 provides us a natural way to represent relations as high-dimensional rotations for better modeling capacity. We call such rotations composed of $2\lfloor \frac{k}{2} \rfloor$ Householder reflections as the *Householder rotations*.

Given a triple (h, r, t), we denote the embeddings of head entity h and tail entity t as $S_h \in \mathbb{R}^{d \times k}$ and $S_t \in \mathbb{R}^{d \times k}$, where d is the embedding size of entities and k is the dimension size of each row vector. Recall that in RotatE (Sun et al., 2019), Rotate3D (Gao et al., 2020) and QuatE (Zhang et al., 2019), each element (row) in the entity embeddings is represented as a 2-dimensional, 3-dimensional, 4-dimensional vector (i.e., $k = 2, 3, 4$). More generally, HousE-r represents each row of entity embeddings as a k-dimensional vector, i.e., $S_h[i], S_t[i] \in \mathbb{R}^k, i \in \{1, \ldots, d\}$. To model each relation as a row-wise k-dimensional rotation between head and tail entities, the embedding of relation r is denoted as $U_r \in \mathbb{R}^{d \times 2n \times k}$, where $n = \lfloor \frac{k}{2} \rfloor$. Each row $U_r[i] \in \mathbb{R}^{2n \times k}$ is composed of $2n$ k-dimensional unit vectors, i.e., $U_r[i][j] \in \mathbb{R}^k$ and $\|U_r[i][j]\|_2^2 = 1, j \in \{1, \ldots, 2n\}$.

We propose to parameterize the relational Householder rotations by using the mapping Rot-H in Equation (4). Formally, for each triple (h, r, t), HousE-r applies r-specific Householder rotations to the i-th row of head embedding h:

$$S'_h[i] = \text{Rot-H}(U_r[i])S_h[i]$$
$$= \prod_{j=1}^{2n} H(U_r[i][j])S_h[i]. \quad (5)$$

Based on Theorem 3.1, any k-dimensional relational rotations can be represented by Equation (5). As illustrated in Figure 2(b), a 2-dimensional rotation can be viewed as the composition of 2 Householder reflections.

Distance function of HousE-r. The distance function measures the distance between the rotated head entity embedding S'_h and the tail entity embedding S_t:

$$d_r(h, t) = \sum_{i=1}^{d} \|S'_h[i] - S_t[i]\|_2. \quad (6)$$

Modeling capability of HousE-r. Theoretically, HousE-r can model and infer symmetry, antisymmetry, inversion and composition patterns. The definitions of these relation patterns are listed in Appendix C for clarity.

Claim 3.2. *HousE-r can model the symmetry/antisymmetry pattern. (See proof in Appendix D.1)*

Claim 3.3. *HousE-r can model the inversion pattern. (See proof in Appendix D.2)*

Claim 3.4. *HousE-r can model the composition pattern. (See proof in Appendix D.3)*

Efficient computation. The time complexity of Equation (5) is $O(2nk^2)$, in which $2n$ matrix-vector multiplications incur high computational costs. However, it is worth noting that these matrix multiplications can be replaced by the vector operations. Formally, based on Equation (3), the j-th matrix-vector multiplication can be expressed as:

$$S_h^j[i] = H(U_r[i][j])S_h^{j-1}[i]$$
$$= S_h^{j-1}[i] - 2\langle S_h^{j-1}[i], U_r[i][j]\rangle U_r[i][j], \quad (7)$$

where $S_h^0[i] = S_h[i]$. Through such iterated vector operations, the time complexity can be reduced to $O(2nk)$.

Connections to RotatE, Rotate3D and QuatE. As shown in Table 1, the rotations of RotatE, Rotate3D and QuatE are modeled in 2-dimensional, 3-dimensional and 4-dimensional spaces, respectively. Geometrically, they can be viewed as the special cases of HousE-r by setting the rotation dimension k to 2, 3 and 4, respectively. For example, as shown in Figure 2(b), HousE-r in 2-dimensional space

is equivalent to RotatE since any rotation in a plane can be represented by two conjunctive Householder reflections. Moreover, unlike previous models restricting rotations to a fixed and low-dimensional space, HousE-r can easily model high-dimensional rotations by enlarging the value of k.

Limitation of Householder rotations. On the other side of the coin, HousE-r is not the panacea as the pure Householder rotations suffer from the challenge of *indistinguishable representations* in modeling RMPs. Considering the ideal case of no-error embedding, we have the following deductions:

- For a 1-to-N relation r, when (h, r, t_1) and (h, r, t_2) hold, $S_{t_1}[i] = S_{t_2}[i]$.

- For an N-to-1 relation r', when (h_1, r', t) and (h_2, r', t) hold, $S_{h_1}[i] = S_{h_2}[i]$.

One can see that the embeddings of different entities tend to be identical when facing the complex RMPs, leading to the uninformative representations. Thus, it is meaningful to tackle this challenge in our proposal.

3.2. HousE: Improved HousE-r with Relational Householder Projections

To handle the sophisticated RMPs, some projection operations have been proposed and shown their effectiveness (Wang et al., 2014; Lin et al., 2015; Ji et al., 2015). The relational projections enable the KGE models to generate relation-specific representations for each entity (Wang et al., 2014). However, existing projections are irreversible transformations, leading to the failure in modeling inversion and composition patterns (Sun et al., 2019). Differently, we propose the novel invertible projections named *Householder projections* by modifying the vanilla Householder matrices to tackle the limitation of HousE-r.

More concretely, given a unit vector $p \in \mathbb{R}^k$, i.e., $\|p\|_2^2 = 1$ and a real scalar τ, the $k \times k$ modified Householder matrix $M(p, \tau)$ is defined as:

$$M(p, \tau) = I - \tau pp^\top. \quad (8)$$

Note that the modified Householder matrix $M(p, \tau)$ has $k-1$ eigenvalues equal to 1 and one eigenvalue equal to $1-\tau$. Thus, $M(p, \tau)$ is invertible when $\tau \neq 1$. Geometrically, the modified Householder matrix transforms x to $\hat{x}$ by a projection along the axis p:

$$\hat{x} = M(p, \tau) = x - \tau\langle x, p\rangle p, \quad (9)$$

where τ determines the position of $\hat{x}$ on the axis p. Figure 2(c) illustrates several projected results with different values of τ in two-dimensional space.

Moreover, based on the modified Householder matrices, given a series of real scalars $T = \{\tau_c\}_{c=1}^m$ and unit vectors

$P = \{p_c\}_{c=1}^m$ where m is a positive integer and $p_c \in \mathbb{R}^k$, we define the mapping:

$$\text{Pro-H}(P, T) = \prod_{c=1}^m M(p_c, \tau_c). \tag{10}$$

The output of $\text{Pro-H}(P, T)$ is an invertible matrix since the product of invertible matrices is also an invertible matrix. We name such projections composed of m modified Householder reflections as *Householder projections*. Different from the rigidly distance-preserving Householder rotations, the Householder projections can reversibly change the relative distance between two points, and thus provide a suitable solution for modeling RMPs without sacrificing the capability of modeling relation patterns.

Specifically, we incorporate the relational Householder rotations and relational Householder projections under a unified framework named HousE to enjoy the merits from both sides. The relational Householder projections enable relation-specific representations for each entity and the relational Householder rotations enable high-dimensional rotations between projected entities. As shown in Figure 2(d), given the input triple (h, r, t), HousE first learns the relation(r)-specific representations $S_{h,r}$ and $S_{t,r}$ for head and tail entities via Householder projections, respectively. Then, $S_{h,r}$ is transformed by the high-dimensional Householder rotations to be close to $S_{t,r}$.

In the phase of relational Householder projections, we define two types of parameters for each relation r: the axes $P_r \in \mathbb{R}^{d \times m \times k}$ and the scalars $T_r \in \mathbb{R}^{d \times m}$, where m is a positive integer. Each row $P_r[i] \in \mathbb{R}^{m \times k}$ is composed of m k-dimensional unit vectors (projection axes), i.e., $P_r[i][j] \in \mathbb{R}^k$ and $\|P_r[i][j]\|_2^2 = 1$. Each row $T_r[i]$ is composed of m real values (projection scalars).

We propose to parameterize the relational Householder projections by using the mapping Pro-H in Equation (10). Considering the head and tail parts of a relation usually have different implicit types (Bordes et al., 2011), HousE utilizes two sets of independent projection parameters $\{P_{r,1}, T_{r,1}\}$ and $\{P_{r,2}, T_{r,2}\}$ for each relation r to project h and t, respectively. Formally, For each triple (h, r, t), HousE transforms each row of head entity h and tail entity t with r-specific Householder projections:

$$
\begin{aligned}
S_{h,r}[i] &= \text{Pro-H}(P_{r,1}[i], T_{r,1}[i]) S_h[i] \\
&= \prod_{j=1}^m M(P_{r,1}[i][j], T_{r,1}[i][j]) S_h[i], \\
S_{t,r}[i] &= \text{Pro-H}(P_{r,2}[i], T_{r,2}[i]) S_t[i] \\
&= \prod_{j=1}^m M(P_{r,2}[i][j], T_{r,2}[i][j]) S_t[i].
\end{aligned} \tag{11}
$$

Algorithm 1 Forward procedure of HousE
___
1: **Input:** An input triple (h, r, t), head (tail) entity embedding S_h (S_t), projection parameters $\{T_{r,1}, T_{r,2}\}$ and $\{P_{r,1}, P_{r,2}\}$, rotation parameters U_r, embedding size d, rotation dimension k, number of modified Householder reflections m.
2: **Output:** Distance δ
3: $\delta \leftarrow 0$
4: $n \leftarrow \lfloor \frac{k}{2} \rfloor$
5: **for** $i = 1$ **to** d **do**
6:    /* Relational Householder projections */
   $S_{h,r}[i] \leftarrow \prod_{j=1}^m M(P_{r,1}[i][j], T_{r,1}[i][j]) S_h[i]$
   $S_{t,r}[i] \leftarrow \prod_{j=1}^m M(P_{r,2}[i][j], T_{r,2}[i][j]) S_t[i]$
7:    /* Relational Householder rotations */
   $S'_{h,r}[i] \leftarrow \prod_{j=1}^{2n} H(U_r[i][j]) S_{h,r}[i]$
8:    $\delta \leftarrow \delta + \|S'_{h,r}[i] - S_{t,r}[i]\|_2$
9: **end for**
10: **Return:** δ
___

After that, HousE models the row-wise Householder rotations between the projected head point $S_{h,r}$ and projected tail point $S_{t,r}$, which is the same as the one in Equation (5). If (h, r, t) holds, we expect the rotated head point $S'_{h,r}[i] = \text{Rot-H}(U_r[i]) S_{h,r}[i] \approx S_{t,r}[i]$, where $U_r[i]$ is composed of $2 \lfloor \frac{k}{2} \rfloor$ r-specific Householder reflections.

As shown in Algorithm 1, for each triple (h, r, t), HousE first utilizes the relational Householder projection to generate r-specific representations $S_{h,r}$ and $S_{t,r}$ for h and t, as in line 6. Then, HousE applies the relational Householder rotation to the projected head embedding $S_{h,r}$, as in line 7. The rotated result $S'_{h,r}$ is expected to be close to the projected tail embedding $S_{t,r}$. Note that we replace the matrix-vector multiplications in line 6 and 7 with the vector operations in Equation (9) and (3) for efficient computation.

The learnable parameters of HousE include $\{S_e\}_{e \in \mathcal{E}}$ and $\{U_r, P_{r,1}, P_{r,2}, T_{r,1}, T_{r,2}\}_{r \in \mathcal{R}}$. Compared to previous models (Sun et al., 2019; Zhang et al., 2019), the extra cost is proportional to the number of relation types, which is usually much smaller than the number of entities. Therefore, the total number of parameters in HousE is about $O(dk|\mathcal{E}|)$.

Distance function of HousE. For each triple (h, r, t), the distance function of HousE is defined as:

$$d_r(h, t) = \sum_{i=1}^d \|S'_{h,r}[i] - S_{t,r}[i]\|_2. \tag{12}$$

Modeling capability of HousE. HousE can model and infer all the relation patterns and RMPs as shown in Table 1 (we also discuss other relation patterns in Appendix E). Formally, we can achieve the following claims:

Claim 3.5. *HousE can model the symmetry/antisymmetry*

Table 2. Link prediction results on WN18 and FB15k. Best results are in **bold** and second best results are underlined. [†]: Results are taken from (Nguyen et al., 2018); [◇]: Results are taken from (Kadlec et al., 2017). Other results are taken from the original papers.

	WN18					FB15k				
Model	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE†	-	.495	.113	.888	.943	-	.463	.297	.578	.749
DistMult◇	655	.797	-	-	.946	42.2	.798	-	-	.893
ComplEx	-	.941	.936	.945	.947	-	.692	.599	.759	.84
ConvE	374	.943	.935	.946	.956	51	.657	.558	.723	.831
RotatE	309	.949	.944	.952	.959	40	.797	.746	.830	.884
Rotate3D	214	.951	.945	.953	.961	<u>39</u>	.789	.728	.832	.887
QuatE	388	.949	.941	.954	.960	41	.770	.700	.821	.878
DualE	-	.951	.945	.956	.961	-	.790	.734	.829	.881
HousE-r	<u>155</u>	<u>.953</u>	<u>.947</u>	<u>.956</u>	<u>.964</u>	<u>39</u>	<u>.807</u>	<u>.758</u>	<u>.839</u>	<u>.893</u>
HousE	**137**	**.954**	**.948**	**.957**	**.964**	**38**	**.811**	**.759**	**.847**	**.898**

Table 3. 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). Other results are taken from the corresponding original papers.

	WN18RR					FB15k-237					YAGO3-10				
Model	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	.43	.39	.44	.49	254	.241	.155	.263	.419	5926	.34	.24	.38	.54
ComplEx◇	5261	.44	.41	.46	.51	339	.247	.158	.275	.428	6351	.36	.26	.4	.55
ConvE◇	4187	.43	.40	.44	.52	224	.325	.237	.356	.501	1671	.44	.35	.49	.62
RotatE	3340	.476	.428	.492	.571	177	.338	.241	.375	.533	1767	.495	.402	.55	.67
Rotate3D	3328	.489	.442	.505	.579	165	.347	.250	<u>.385</u>	<u>.543</u>	-	-	-	-	-
QuatE	3472	.481	.436	.500	.564	176	.311	.221	.342	.495	-	-	-	-	-
DualE	-	.482	.440	.500	.561	-	.330	.237	.363	.518	-	-	-	-	-
Rot-Pro	2815	.457	.397	.482	.577	201	.344	.246	.383	.540	1797	.542	.443	.596	.669
HousE-r	<u>1885</u>	<u>.496</u>	<u>.452</u>	<u>.511</u>	<u>.585</u>	<u>165</u>	<u>.348</u>	<u>.254</u>	.384	.534	<u>1449</u>	<u>.565</u>	<u>.487</u>	<u>.616</u>	<u>.703</u>
HousE	**1303**	**.511**	**.465**	**.528**	**.602**	**153**	**.361**	**.266**	**.399**	**.551**	**1415**	**.571**	**.491**	**.620**	**.714**

pattern. (See proof in Appendix D.4)

Claim 3.6. *HousE can model the inversion pattern. (See proof in Appendix D.5)*

Claim 3.7. *HousE can model the composition pattern. (See proof in Appendix D.6)*

Claim 3.8. *HousE can model the relation mapping properties. (See proof in Appendix D.7)*

Connections to TransH, TransR and TransD. Previous works such as TransH, TransR and TransD also focus on designing the projection operations to ensure that the same entity has different representations under different relations. However, as shown in Table 1, these methods will undermine the ability to infer inversion and composition patterns due to the irreversible projection operations. Note that, the projection operation of TransH is a special case of HousE if we set the scalar $\tau = 1$, which essentially is the irreversible transformation. Different from these works, HousE utilizes an invertible matrix derived by a series of modified Householder matrices to generate relation-specific entity representations. Such invertible projections enable our proposal to model relation mapping properties without sacrificing the capability in modeling relation patterns.

4. Experiment

4.1. Experimental Setup

Datasets. We evaluate our proposals on five widely-used benchmarks: WN18 (Bordes et al., 2013), FB15k (Bordes et al., 2013), WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova & Chen, 2015) and YAGO3-10 (Mahdis-oltani et al., 2015). Refer to Appendix F for more details.

Baselines. We compare our models with a number of baselines. For non-rotation models, we report TransE (Bordes et al., 2013), DistMult (Yang et al., 2015), ComplEx (Trouil-lon et al., 2016) and ConvE (Dettmers et al., 2018). For rotation-based models, we report RotatE (Sun et al., 2019), Rotate3D (Gao et al., 2020), QuatE (Zhang et al., 2019), DualE (Cao et al., 2021) and Rot-Pro (Song et al., 2021).

Implementation details. To ensure fair comparisons, we set a smaller embedding size d for HousE-r and HousE, so that the total numbers of parameters are comparable to baselines. More details can be found in Appendix G.

4.2. Main Results

The experimental results are summarized in Table 2 and Table 3. Compared to all the baselines, both HousE-r and

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

Relation Name	RotatE	QuatE	HousE-r	HousE
hypernym	0.154	0.172	<u>0.182</u>	**0.207**
instance_hypernym	0.324	0.362	<u>0.395</u>	**0.440**
member_meronym	0.255	0.236	<u>0.275</u>	**0.312**
synset_domain_topic_of	0.334	0.395	<u>0.396</u>	**0.428**
has_part	0.205	0.210	<u>0.217</u>	**0.232**
member_of_domain_usage	0.277	0.372	<u>0.415</u>	**0.453**
member_of_domain_region	0.243	0.140	<u>0.281</u>	**0.395**
derivationally_related_form	0.957	0.952	<u>0.958</u>	**0.958**
also_see	0.627	0.607	<u>0.638</u>	**0.640**
verb_group	0.968	0.930	<u>0.968</u>	**0.968**
similar_to	1.000	1.000	<u>1.000</u>	**1.000**

Table 5. MRR for the models tested on RMPs in FB15k-237.

Task	RMPs	RotatE	HousE
Predicting Head (MRR)	1-to-1	0.498	**0.514**
	1-to-N	0.475	**0.479**
	N-to-1	0.088	**0.114**
	N-to-N	0.260	**0.286**
Predicting Tail (MRR)	1-to-1	0.490	**0.502**
	1-to-N	0.071	**0.086**
	N-to-1	0.747	**0.778**
	N-to-N	0.367	**0.392**

HousE achieve SOTA performance, demonstrating the effectiveness of the Householder framework.

Table 2 shows the results on WN18 and FB15k, from which we observe that even with only Householder rotations, HousE-r already consistently outperforms the baselines over both datasets. Moreover, by combining Householder rotations and Householder projections together, HousE further achieves new state-of-the-art results on both WN18 and FB15k datasets. Considering that the main relation patterns in WN18 and FB15k are symmetry, antisymmetry and inversion, the superior performance of HousE-r and HousE reveals their effectiveness in modeling these patterns.

Table 3 summarizes the results on WN18RR, FB15k-237 and YAGO3-10. On these datasets, HousE-r surpasses most of the baselines. The only comparable exception is Rotate3D on FB15k-237 which models relations as 3-d rotations. However, HousE-r uses much less parameters than Rotate3D as shown in Appendix G and achieves similar performance, which also verifies the superior modeling capacity of Householder rotations. The improvements over existing rotations-based baselines (RotatE, Rotate3D, QuatE and DualE) demonstrate the superiority of high-dimensional rotations. Moreover, HousE consistently outperforms HousE-r along with all the baselines by a large margin on the three datasets across all metrics, benefiting from the ability to model relational mapping properties.

4.3. Fine-grained Performance Analysis

To further verify the modeling capacity of our proposal from a fine-grained perspective, we report the performance on each relation of WN18RR following (Zhang et al., 2019). As shown in Table 4, compared to two rotation-based baselines RotatE and QuatE, we observe that:

(1) HousE-r surpasses all the baselines on all 11 relation types, confirming the superior modeling capacity of the Householder rotations.

(2) By incorporating the Householder projections, HousE achieves more significant improvements on the challenging

1-to-N and N-to-1 relations. For example, HousE outperforms RotatE on 1-to-N relation *member_of_domain_region* and N-to-1 relation *instance_hypernym* with 62.55% and 35.80% relative gains, respectively.

4.4. Capability of Modeling RMPs

In order to further demonstrate the effectiveness of HousE in modeling RMPs, we report the detailed results of our proposal on different RMPs[2] in FB15k-237.

Table 5 exhibits the results on different types of RMPs. One can see that HousE outperforms RotatE across all RMP types. For example, on the challenging N-to-1 (predicting head) and 1-to-N (predicting tail) tasks, HousE achieves 29.55% and 21.13% relative improvements over RotatE. Such advanced performance of HousE owes to the powerful modeling capability of the Householder projections.

4.5. Hyperparameter Sensitivity Analysis

Dimension of rotations. To verify the expressiveness of high-dimensional rotations, we conduct experiments for our models under varying rotation dimension k. Figure 3(a) and 3(b) show the results on WN18RR and FB15k-237.

As expected, on both datasets, HousE-r and HousE rotated in higher-dimensional spaces achieve better performance than the ones rotated in lower-dimensional spaces, since the high-dimensional rotations bring the superior modeling capacity. Moreover, HousE consistently surpasses HousE-r by a large margin across all rotation dimensions, demonstrating the effectiveness of the integrated Householder projections. For example, on WN18RR, HousE with 4-dimensional rotations already outperforms HousE-r with 12-dimensional rotations.

Number of modified Householder matrices. As shown in Equation (10), a Householder projection is composed of

---

[2]Following (Sun et al., 2019), for each relation r, we compute the average number of heads per tail (hpt_r) and the average number of tails per head (tph_r). If $hpt_r < 1.5$ and $tph_r < 1.5$, r is treated as 1-to-1; if $hpt_r \geq 1.5$ and $tph_r \geq 1.5$, r is treated as N-to-N; if $hpt_r < 1.5$ and $tph_r \geq 1.5$, r is treated as 1-to-N; if $hpt_r \geq 1.5$ and $tph_r < 1.5$, r is treated as N-to-1.
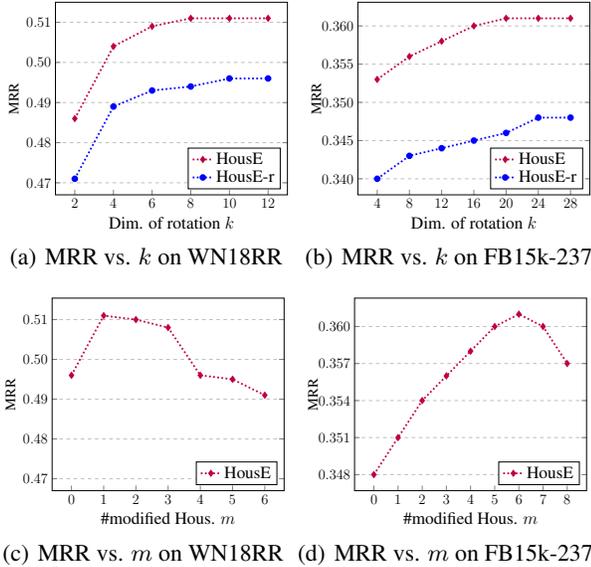
Figure 3. (a) and (b) show the MRR results of HousE and HousE-r with varying rotation dimensions on WN18RR and FB15k-237; (c) and (d) show the MRR results of HousE with different numbers of modified Householder matrices on WN18RR and FB15k-237.

m modified Householder matrices. Here we investigate the impact of m on the performance (MRR) of HousE. Figure 3(c) and 3(d) show the results on WN18RR and FB15k-237.

With the increase of m, the performance of HousE first improves and then drops on both datasets. This is because the larger m provides greater projection capability, but the over-complicated projections also aggravate the risk of overfitting. Moreover, the values of m for the best performance on the two datasets are different ($m = 1$ on WN18RR and $m = 6$ on FB15k-237) due to the distinct graph densities. Specifically, WN18RR is a sparse KG dataset with the average degree of 2.19, while FB15k-237 is a much denser KG with the average degree of 18.71. Thus, the larger m is needed for modeling the richer graph information in FB15k-237.

4.6. Superiority of Householder Projections

To verify the effectiveness of the proposed Householder projections, we design two variants of HousE by replacing the Householder projections with previous irreversible projections used in TransH (Wang et al., 2014) and TransR (Lin et al., 2015), dubbed HousH and HousR respectively. Table 6 shows the experimental results on WN18RR and FB15k-237. Compared to HousE-r without any projections, the performance of HousH and HousR is barely improved on FB15k-237, and even degraded on WN18RR. It reveals that the irreversible projections may hinder the modeling capability. Moreover, HousE significantly outperforms HousH and HousR on both datasets, demonstrating the superiority of the invertible Householder projections in HousE.

Table 6. Performance of different variants.

	WN18RR		FB15k-237	
Variants	MRR	H@10	MRR	H@10
HousH	.491	.584	.347	.537
HousR	.488	.580	.349	.538
HousE-r	.496	.585	.348	.534
HousE-r$^+$	.500	.591	.351	.538
HousE	.511	.602	.361	.551
HousE$^+$	.514	.606	.366	.552

4.7. Additional Translations

To explore the potential of our proposal, we also incorporate translations (Bordes et al., 2013) into HousE-r and HousE, dubbed HousE-r$^+$ and HousE$^+$ respectively. The translations are directly deployed after the Householder rotations. From Table 6, we see that these two variants both outperform their original versions. This is because the translations provide a natural way to represent the hierarchical property of KGs (Bordes et al., 2013), which also endows our proposal with more comprehensive modeling capacity.

5. Related Work

Translation-based models. TransE (Bordes et al., 2013) is the first model that represents each relation as a translation between entities. This simple model is effective in modeling antisymmetry, inversion and composition patterns, but fails in handling symmetry pattern and RMPs. To tackle TransE's limitations, a set of variants (Wang et al., 2014; Lin et al., 2015; Ji et al., 2015; Xiao et al., 2015) are subsequently proposed. TransH (Wang et al., 2014) projects entities to a relation-specific hyperplane and performs translation on this hyperplane. TransR (Lin et al., 2015) models entities and relations in distinct spaces and conducts relation-specific projections with normal linear transformations. However, these models lose the ability to model inversion and composition patterns since irreversible linear transformations are performed on head and tail entities (Sun et al., 2019).

Rotation-based models. Following ComplEx (Trouillon et al., 2016) which extends DistMult (Yang et al., 2015) to complex number systems, RotatE (Sun et al., 2019) represents each relation as a 2-dimensional rotation in complex plane to model symmetry, antisymmetry, inversion and composition patterns. Rotate3D (Gao et al., 2020) and QuatE (Zhang et al., 2019) extend the rotations to 3-dimensional and 4-dimensional spaces by introducing the quaternion number system. Recently, DualE (Cao et al., 2021) utilizes dual quaternions to combine translations and rotations in 3-d space for modeling multiple relations.

Neural-network-based models. There are also some models using neural networks for KGE. R-GCN (Schlichtkrull

et al., 2018) introduces graph neural networks as the graph encoders. ConvE (Dettmers et al., 2018) exploits convolution operations to facilitate the score calculation. However, such methods lack of explicit geometrical explanations on modeling relation patterns and RMPs.

6. Conclusion

In this paper, we propose HousE, a novel KGE framework based on Householder parameterization. HousE models relations as high-dimensional Householder rotations to capture crucial relation patterns. Moreover, with Householder projections, HousE generates relation-specific embeddings for each entity to model RMPs. Experimental results on five datasets demonstrate the superiority of our proposal.

Acknowledgements

References

Artin, E. *Geometric algebra*. Courier Dover Publications, 2016.

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., Weston, J., Collobert, R., and Bengio, Y. Learning structured embeddings of knowledge bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

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.

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.

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.

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

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.

Kadlec, R., Bajgar, O., and Kleindienst, J. Knowledge base completion: Baselines strike back. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pp. 69–74, 2017.

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

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.

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.

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.

Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., and Welling, M. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607, 2018.

Song, T., Luo, J., and Huang, L. Rot-pro: Modeling transitivity by projection in knowledge graph embedding. In *Advances in Neural Information Processing Systems*, 2021.

Suchanek, F. M., Kasneci, G., and Weikum, G. Yago: a core of semantic knowledge. In *Proceedings of the 16th*

International Conference on World Wide Web, pp. 697–706, 2007.

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.

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.

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.

Xiao, H., Huang, M., Hao, Y., and Zhu, X. Transa: An adaptive approach for knowledge graph embedding. *CoRR*, 2015.

Xiong, C., Power, R., and Callan, J. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 1271–1279, 2017.

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.

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

A. Notations

Table 7. Notations used in this paper.

Symbol	Shape	Description
$\mathcal{E}$	-	Set of entities
$\mathcal{R}$	-	Set of relations
$\mathcal{D}$	-	Set of factual triples
h, t	-	Head entity and tail entity
r	-	Relation type
d	$\mathbb{R}$	Embedding size
k	$\mathbb{R}$	Rotation dimension
m	$\mathbb{R}$	Number of modified Householder matrices
S_e	$\mathbb{R}^{d \times k}$	Representation of entity $e \in \mathcal{E}$
$S_{e,r}$	$\mathbb{R}^{d \times k}$	r-specific representation of entity $e \in \mathcal{E}$
U_r	$\mathbb{R}^{d \times 2 \lfloor \frac{k}{2} \rfloor \times k}$	Param. of $r \in \mathcal{R}$ for Householder rotation
P_r	$\mathbb{R}^{d \times m \times k}$	Param. of $r \in \mathcal{R}$ for Householder projection axes
T_r	$\mathbb{R}^{d \times m}$	Param. of $r \in \mathcal{R}$ for Householder projection scalars

B. Proofs of Theorem 3.1

B.1. Proof of Lemma B.1

In order to prove Theorem 3.1, we first prove an auxiliary Lemma B.1.

Lemma B.1. *Any $k \times k$ orthogonal matrix Q can be decomposed into the product of $k - 1$ or k Householder matrices.*

Proof. From the Householder QR decomposition (Householder, 1958), we can upper triangularize any full-rank matrix $W \in \mathbb{R}^{k \times k}$ by using $k - 1$ Householder reflections, i.e.,

$$H(u_{k-1})H(u_{k-2}) \cdots H(u_1)W = R,$$

where $R \in \mathbb{R}^{k \times k}$ is an upper triangular matrix and its first $n - 1$ diagonal elements are all positive.

When Household QR decomposition is performed on an orthogonal matrix Q, we can get:

$$H(u_{k-1})H(u_{k-2}) \cdots H(u_1)Q = R.$$

Note that R here is both upper triangular and orthogonal (i.e., $RR^T = I$) since it is a product of k orthogonal matrices. It establishes that R is a diagonal matrix, where the first $k - 1$ diagonal entries are equal to $+1$ and the last diagonal entry is either $+1$ or -1.

If the last diagonal entry of R is equal to $+1$, we have

$$H(u_{k-1})H(u_{k-2}) \cdots H(u_1)Q = I.$$

As each Householder matrix $H(u_i)$ is its own inverse, we obtain that

$$Q = H(u_1) \cdots H(u_{k-1}). \tag{13}$$

If the last diagonal entry of R is equal to -1, we can set $u_k = e_k = (0, \ldots, 0, 1)^\top \in \mathbb{R}^k$ and consequently get

$$H(u_k)R = H(u_k)H(u_{k-1}) \cdots H(u_1)Q = I.$$

Since $H(u_i)$ is its own inverse, we also obtain that

$$Q = H(u_1)\cdots H(u_{k-1})H(u_k). \tag{14}$$

From Equation (13) and (14), we can see that any $k \times k$ orthogonal matrix can be decomposed into the product of $k - 1$ or k Householder matrices. $\square$

B.2. Proof of Theorem 3.1

Proof. We first prove that when $n = \lfloor \frac{k}{2} \rfloor$, the image of Rot-H is a subset of $\mathbf{SO}(k)$, i.e., Rot-H$(U) \subset \mathbf{SO}(k)$. Note that each Householder matrix is symmetric and orthogonal and its determinant is -1. Therefore, the product of $2n$ Householder matrices is an orthogonal matrix with determinant $+1$, i.e., a rotation matrix (Artin, 2016), which means Rot-H$(U) \subset \mathbf{SO}(k)$.

Then we also prove that its converse is also valid, i.e., any $k \times k$ rotation matrix can be expressed as the product of $2\lfloor \frac{k}{2} \rfloor$ Householder matrices $H(u_i)$. Note that a rotation matrix $\widetilde{Q}$ is a special orthogonal matrix with determinant $+1$ (Artin, 2016), i.e. $\det(\widetilde{Q}) = +1$, and thus $\widetilde{Q}$ can be decomposed into the product of $k - 1$ or k Householder matrices based on Lemma B.1. Moreover, since $\det(H(u_i)) = -1$ and the determinant of a product of matrices is the product of their determinants, we can naturally derive that any $k \times k$ rotation matrix can be decomposed into the product of $2\lfloor \frac{k}{2} \rfloor$ Householder matrices, i.e., $\mathbf{SO}(k) \subset$ Rot-H(U). All in all, we have Rot-H$(U) = \mathbf{SO}(k)$. $\square$

C. Definitions

Definition C.1. A relation r is **symmetric** (**antisymmetric**) if $\forall x, y$

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

A clause with such form is a **symmetry** (**antisymmetry**) pattern.

Definition C.2. A relation r_1 is **inverse** to relation r_2 if $\forall x, y$

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

A clause with such form is an **inversion** pattern.

Definition C.3. A relation r_1 is **composed** of relation r_2 and relation r_3 if $\forall x, y, z$

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

A clause with such form is a **composition** pattern.

Following (Bordes et al., 2013), there are four relation mapping properties:

Definition C.4. A relation r is **1-to-1** if a *head* can appear with at most one *tail*.

Definition C.5. A relation r is **1-to-N** if a *head* can appear with many *tails*.

Definition C.6. A relation r is **N-to-1** if many *heads* can appear with the same *tail*.

Definition C.7. A relation r is **N-to-N** if many *heads* can appear with many *tails*.

D. Proofs of Claims

We denote the r-specific Householder rotation matrix and Householder projection matrices as $\widetilde{Q}_r$ and $\{W_{r,1}, W_{r,2}\}$ respectively:

$$\begin{aligned}
\widetilde{Q}_r &= \text{Rot-H}(U_r[i]), \\
W_{r,1} &= \text{Pro-H}(P_{r,1}[i], T_{r,1}[i]), \\
W_{r,2} &= \text{Pro-H}(P_{r,2}[i], T_{r,2}[i]).
\end{aligned}$$

For simplicity, we also omit the row indices $[i]$ of entity representations in the following proofs.

D.1. Proof of Claim 3.2

Proof. if $r(x, y)$ and $r(y, x)$ hold, we have

$$S_y = \widetilde{Q}_r S_x \wedge S_x = \widetilde{Q}_r S_y \Rightarrow \widetilde{Q}_r \widetilde{Q}_r = I$$

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

$$S_y = \widetilde{Q}_r S_x \wedge S_x \neq \widetilde{Q}_r S_y \Rightarrow \widetilde{Q}_r \widetilde{Q}_r \neq I$$

$\square$

D.2. Proof of Claim 3.3

Proof. if $r_1(x, y)$ and $r_2(y, x)$ hold, we have

$$S_y = \widetilde{Q}_{r_1} S_x \wedge S_x = \widetilde{Q}_{r_2} S_y \Rightarrow \widetilde{Q}_{r_1} = \widetilde{Q}_{r_2}^T$$

$\square$

D.3. Proof of Claim 3.4

Proof. if $r_1(x, z), r_2(x, y)$ and $r_3(y, z)$ hold, we have

$$S_z = \widetilde{Q}_{r_1} S_x \wedge S_y = \widetilde{Q}_{r_2} S_x \wedge S_z = \widetilde{Q}_{r_3} S_y \Rightarrow \widetilde{Q}_{r_1} = \widetilde{Q}_{r_3} \widetilde{Q}_{r_2}$$

$\square$

D.4. Proof of Claim 3.5

Proof. if $r(x, y)$ and $r(y, x)$ hold, we have

$$W_{r,2} S_y = \widetilde{Q}_r W_{r,1} S_x \wedge W_{r,2} S_x = \widetilde{Q}_r W_{r,1} S_y$$
$$\Rightarrow (W_{r,2}^{-1} \widetilde{Q}_r W_{r,1})(W_{r,2}^{-1} \widetilde{Q}_r W_{r,1}) = I$$

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

$$W_{r,2} S_y = \widetilde{Q}_r W_{r,1} S_x \wedge W_{r,2} S_x \neq \widetilde{Q}_r W_{r,1} S_y$$
$$\Rightarrow (W_{r,2}^{-1} \widetilde{Q}_r W_{r,1})(W_{r,2}^{-1} \widetilde{Q}_r W_{r,1}) \neq I$$

$\square$

D.5. Proof of Claim 3.6

Proof. if $r_1(x, y)$ and $r_2(y, x)$ hold, we have

$$W_{r_1,2} S_y = \widetilde{Q}_{r_1} W_{r_1,1} S_x \wedge W_{r_2,2} S_x = \widetilde{Q}_{r_2} W_{r_2,1} S_y$$
$$\Rightarrow W_{r_1,2}^{-1} \widetilde{Q}_{r_1} W_{r_1,1} = (W_{r_2,2}^{-1} \widetilde{Q}_{r_2} W_{r_2,1})^{-1}$$

$\square$

D.6. Proof of Claim 3.7

Proof. if $r_1(x, z), r_2(x, y)$ and $r_3(y, z)$ hold, we have

$$W_{r_1,2} S_z = \widetilde{Q}_{r_1} W_{r_1,1} S_x$$
$$\wedge W_{r_2,2} S_y = \widetilde{Q}_{r_2} W_{r_2,1} S_x$$
$$\wedge W_{r_3,2} S_z = \widetilde{Q}_{r_3} W_{r_3,1} S_y$$
$$\Rightarrow W_{r_1,2}^{-1} \widetilde{Q}_{r_1} W_{r_1,1} = (W_{r_3,2}^{-1} \widetilde{Q}_{r_3} W_{r_3,1})(W_{r_2,2}^{-1} \widetilde{Q}_{r_2} W_{r_2,1})$$

$\square$

D.7. Proof of Claim 3.8

In order to model sophisticated RMPs, we expect to tackle the challenge of indistinguishable representations with House-holder projections as mentioned in Section 3.1. For the N-to-1 relations, here we take a 2-to-1 relation r as an example with two triples (h_1, r, t) and (h_2, r, t). Householder projections can adjust the relative distance between entity h_1 and h_2 according to relation r. Formally, the original distance between h_1 and h_2 is defined as: $s = \|S_{h_1} - S_{h_2}\|_2$. After applying a modified Householder matrix, the relative distance between the projected representations is:

$$\hat{s}^2 = \|S_{h_1,r} - S_{h_2,r}\|_2^2 = s^2 + (\tau_r^2 - 2\tau_r)s^2 \cos^2 \theta_{s,p_r}.$$

It is clear that the learnable τ_r determines the increase or decrease of the relative distance: (1) when $0 < \tau_r < 2$, $\hat{s} \leq s$; (2) when $\tau_r = 0$ or 2, $\hat{s} = s$; (3) when $\tau_r < 0$ or $\tau > 2$, $\hat{s} \geq s$. Moreover, the term $\cos \theta_{s,p_r}$ is determined by the relative positions between the entities and the projection axis p_r. This reveals that the Householder projections can adaptively change the relative distance between entities based on their positions. With such projections, one can obtain similar r-specific representations $S_{h_1,r}$ and $S_{h_2,r}$ for h_1 and h_2, while the original representations S_{h_1} and S_{h_2} can be still distinguishable. The same is also true for 1-to-N relations.

E. Discussion on Other Relation Patterns

E.1. Multiplicity

The multiplicity pattern has been investigated in DualE (Cao et al., 2021). Formally, it has the following definition:

Definition E.1. Relation $r_1, r_2, \ldots, r_N$ are multiple if $\forall i \in \{1, \ldots, N\}$, (h, r_i, t) can hold in KGs simultaneously. A clause with such form is a multiplicity pattern.

DualE utilizes dual quaternions to represent each relation as a 3-dimensional rotation followed by a translation. It proves that the combination of rotations and translations can model multiple relations, since for any given rotation applied to the head entity h, there is always a corresponding translation to transform the rotated head entity to the tail entity t.

In our proposed HousE, the relational Householder projections can be regarded as a special translation along the projection axes. Thus, HousE is similar to DualE in terms of multiplicity modeling capacity. What's more, as shown in Section 4.7, our proposal can also easily integrate translations to achieve better performance. Geometrically, DualE can be viewed as a special case of HousE$^+$ with 3-dimensional rotations.

E.2. Transitivity.

Rot-Pro (Song et al., 2021) focuses on modeling the transitivity pattern, which is formally defined as:

Definition E.2. A relation r is transitive if for any instances (e_1, r, e_2) and (e_2, r, e_3) of relation r, (e_1, r, e_3) is also an instance of r. A clause with such form is a transitivity pattern.

Rot-Pro theoretically shows that the transitive relations can be modeled with a special orthogonal projections, which is designed to project the points onto the rotated axes. This kind of projections can be viewed as a 2-dimensional case of TransH's projections.

HousE can be reduced to Rot-Pro if we set the rotation dimension to 2 and the projection scalars to 1. However, in our opinion, such projections may not be the optimal way to handle transitivity. As shown in (Song et al., 2021), Rot-Pro tends to project the entities under the transitive relation to a same point and the phase of relational rotation tends to be $2n\pi(n = 0, 1, 2, \ldots)$. We can see that such solution is a subset of the solution of modeling symmetric relations, which means that the modeled transitive relations must be symmetric and the antisymmetric transitive relations are ignored. Therefore, how to comprehensively model the transitive relations is still a challenging problem, and we will take this as the future work.

F. Datasets

Table 8 summarizes the detailed statistics of five benchmark datasets:

WN18 (Bordes et al., 2013) is extracted from WordNet (Miller, 1995), a database featuring lexical relations between words.

FB15k (Bordes et al., 2013) contains relation triples from Freebase (Bollacker et al., 2008), a large-scale knowledge graph

Table 8. Statistics of five standard benchmarks.

Dataset	#entity	#relation	#training	#validation	#test
WN18	40,943	18	141,442	5,000	5,000
FB15k	14,951	1,345	483,142	50,000	59,071
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

containing general knowledge facts. The main relation patterns in WN18 and FB15k are symmetry, antisymmetry and inversion.

The WN18RR (Dettmers et al., 2018) and FB15k-237 (Toutanova & Chen, 2015) datasets are subsets of WN18 and FB15k respectively with inverse relations removed. The key of link prediction on WN18RR and FB15k-237 boils down to model and infer the symmetry, antisymmetry and composition patterns.

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.

G. Implementation details

Table 9 shows the amount of parameters used in our models and several recent competitive baselines: RotatE, Rotate3D, QuatE and DualE. To ensure fair comparisons, we set the smaller embedding size d to represent each entity and relation in HousE-r and HousE, so that the total number of parameters is similar to other baselines. Specifically, we fix the number of parameters $d \times k$ to represent a single entity as 1000, 1200, 800, 600, 1000 on WN18, FB15k, WN18RR, FB15k-237 and YAGO3-10, respectively. Hyperparameter d denotes the embedding size and k is the rotation dimension. The larger rotation dimension k leads to the smaller embedding size d.

From Table 9, one can see that our proposed models have similar numbers of parameters compared to the baselines. The only exception is QuatE on WN18RR and FB15k-237. We have tried to increase the number of parameters of QuatE by enlarging the embedding size d on these two datasets, while carefully tuning hyperparameters simultaneously. Unfortunately, the performance of QuatE drops with more free parameters. Thus, to ensure the fairness of performance comparison, we report the parameter numbers of QuatE with the best link prediction results.

Table 9. Number of free parameters comparison. The results of baselines are taken from the original papers.

Model	RotatE	Rotate3D	QuatE	DualE	HousE-r	HousE
WN18	40.95M	122.90M	49.15M	65.53M	40.88M	41.03M
FB15k	31.25M	50.23M	26.08M	26.08M	24.40M	27.63M
WN18RR	40.95M	61.44M	16.38M	32.76M	32.57M	32.84M
FB15k-237	29.32M	44.57M	5.82M	11.64M	12.13M	13.36M
YAGO3-10	123.18M	-	-	-	122.91M	122.99M

Table 10 shows the convergence time required for the model training on five datasets. RotatE is the simplest rotation-based model with the linear time complexity, which is selected as the baseline. Compared to RotatE, our proposed HousE-r and HousE cost comparable or even less training time on these datasets by using the efficient computation in Equation (7). Combined with the link prediction results in Table 2 and 3, one can see that our proposal is capable of improving model effectiveness without sacrificing the efficiency.

Table 10. Training time of RotatE and our proposal on five datasets.

Model	WN18RR	FB15k-237	WN18	FB15k	YAGO3-10
RotatE	4h	6h	4h	9h	10h
HousE-r	1.5h	3h	3h	8h	11h
HousE	1.5h	5h	3h	9h	13h

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, rotation dimension k, number of modified Householder reflections m for Householder projections, and regularization coefficient λ. The hyper-parameter search space is shown in Table 11.

Table 11. Hyperparameter search space.

Hyperparameter	Search Space	Type
b	$\{500, 800, 1000, 1500, 2000\}$	Choice
α	$[0.5, 2.0]$	Range
γ	$\{5, 7, 9, 10, 11, 16, 20, 24, 28\}$	Choice
lr	$[0.0001, 0.003]$	Range
k	$\{2, 4, 8, 12, 16, 20, 25, 30\}$	Choice
m	$\{1, 2, 3, 4, 6, 8\}$	Choice
λ	$[0, 0.3]$	Range