Proof for Transformations of the Special Euclidean Group

We take a two-dimensional space as an example and the conclusion for higher dimensional space can be analogized accordingly.

Proof of reflection transformation. Denote θ as the angle of reflection. The reflection matrix can represent the reflection transformation, which is defined as follows:

$$\boldsymbol{M}_{Ref} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}.$$

Then we have $\boldsymbol{M}_{Ref}^T \boldsymbol{M}_{Ref} = \mathbf{I}$ and $\det \boldsymbol{M}_{Ref} = 1$. It implies $\boldsymbol{M}_{Ref} \in \mathbf{SO}(n)$, so $\boldsymbol{M}_{Ref} \in \mathbf{SE}(n)$ hlods. Then the proof completes.

Proof of inversion transformation. According to Eq.(5), if a matrix $\boldsymbol{A} \in \mathbf{SO}(n) \in \mathbf{SE}(n)$, we have $\boldsymbol{A}^T \boldsymbol{A} = \mathbf{I}$, then we can deduce $\boldsymbol{A}$ is an inversion matrix. Therefore it can represent the inversion transformation and the proof completes.

Proof of translation transformation. According to Eq.(6), if a matrix $\mathbf{A} = \begin{bmatrix} \mathbf{R} & \mathbf{v} \\ 0 & 1 \end{bmatrix} \in \mathbf{SE}(n)$, we just make $\mathbf{R}$ be $\mathbf{0}$, then $\boldsymbol{A}$ is a translation matrix. The proof completes.

Proof of rotation transformation. Denote θ as the angle of rotation. The reflection matrix can represent the rotation transformation, which is defined as follows:

$$\boldsymbol{M}_{Rot} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Then we have $\boldsymbol{M}_{Rot}^T \boldsymbol{M}_{Rot} = \mathbf{I}$ and $\det \boldsymbol{M}_{Rot} = 1$. It implies $\boldsymbol{M}_{Rot} \in \mathbf{SO}(n)$, so $\boldsymbol{M}_{Rot} \in \mathbf{SE}(n)$ hlods. Then the proof completes.

Proof of homothety transformation. According to Eq.(6), if a matrix $\boldsymbol{A} \in \mathbf{SE}(n)$, we can turn it into a diagonal matrix, then it is a homothety matrix about itself. The proof completes.

Proof of Key Patterns

As shown in Eq.(5), to simplify the notations, we take 2-dimensional space as an example, in which all involved embeddings are real. The conclusion in n-dimensional space can be analogized accordingly. For a given matrix $\boldsymbol{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we set $\overline{\boldsymbol{M}} = \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \\ \frac{c}{\sqrt{c^2+d^2}} & \frac{d}{\sqrt{c^2+d^2}} \end{bmatrix}$. If $\overline{\boldsymbol{M}} \in \mathbf{GL}(2, \mathbb{R})$, we have $ab + cd = 0$. We might as well let $\overline{\boldsymbol{M}}$ satisfy the conditions of $\mathbf{GL}(2, \mathbb{R})$ and denote it as $\boldsymbol{R}$. Then according to Eq.(6), we have the $\mathbf{A} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$, where $\boldsymbol{t} \in \mathbb{R}^2$. The inverse of $\boldsymbol{A}$ is $\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{R}^{-1} & -\boldsymbol{R}^{-1}\boldsymbol{t} \\ 0 & 1 \end{bmatrix}$. Note that since $\boldsymbol{R} \in \mathbf{GL}(2, \mathbb{R})$, we have $\boldsymbol{R}^{-1} = \boldsymbol{R}^\top$, which means that the the reverse of $\boldsymbol{R}$ is always existing so the the reverse of $\boldsymbol{A}$ is always existing.

We denote $\otimes$ as the the multiplication in tangent space. Next we will prove the seven patterns hold in tangent space as follows.

Proof of symmetry pattern. To prove the symmetry pattern, if $\boldsymbol{r}(\boldsymbol{h}, \boldsymbol{t})$ and $\boldsymbol{r}(\boldsymbol{t}, \boldsymbol{h})$ hold, we have the equality as follows:

$$(\boldsymbol{t} = \boldsymbol{r} \otimes \boldsymbol{h}) \wedge (\boldsymbol{h} = \boldsymbol{r} \otimes \boldsymbol{t}) \Rightarrow \boldsymbol{r} \otimes \boldsymbol{r} = \mathbf{1}.$$

Proof of antisymmetry pattern. In order to prove the antisymmetry pattern, we need to prove the following inequality when $\boldsymbol{r}(\boldsymbol{h}, \boldsymbol{t})$ and $\neg \boldsymbol{r}(\boldsymbol{t}, \boldsymbol{h})$ hold:

$$(\boldsymbol{t} = \boldsymbol{r} \otimes \boldsymbol{h}) \wedge (\boldsymbol{h} \neq \boldsymbol{r} \otimes \boldsymbol{t}) \Rightarrow \boldsymbol{r} \otimes \boldsymbol{r} \neq \mathbf{1}.$$

Proof of inversion pattern. To prove the inversion pattern, if $\boldsymbol{r}_1(\boldsymbol{h}, \boldsymbol{t})$ and $\boldsymbol{r}_2(\boldsymbol{t}, \boldsymbol{h})$ hold, we have the equality as follows:

$$(\boldsymbol{t} = \boldsymbol{r}_1 \otimes \boldsymbol{h}) \wedge (\boldsymbol{h} = \boldsymbol{r}_2 \otimes \boldsymbol{t}) \Rightarrow \boldsymbol{r}_1 = \boldsymbol{r}_2^{-1}.$$

Proof of composition pattern. In order to prove the antisymmetry pattern, we need to prove the following equality holds when $\boldsymbol{r}_1(\boldsymbol{h}, \boldsymbol{p})$, $\boldsymbol{r}_2(\boldsymbol{h}, \boldsymbol{t})$ and $\boldsymbol{r}_3(\boldsymbol{t}, \boldsymbol{p})$ hold:

$$(\boldsymbol{p} = \boldsymbol{r}_1 \otimes \boldsymbol{h}) \wedge (\boldsymbol{t} = \boldsymbol{r}_2 \otimes \boldsymbol{h}) \wedge (\boldsymbol{p} = \boldsymbol{r}_3 \otimes \boldsymbol{t}) \Rightarrow \boldsymbol{r}_1 = \boldsymbol{r}_2 \otimes \boldsymbol{r}_3.$$

Proof of hierarchy pattern. First of all, the hyperbolic space used by our model can capture the hierarchical structure of the data. Next we prove that the model can also logically derive the hierarchy pattern. To prove the hierarchy pattern, if $\boldsymbol{r}_1(\boldsymbol{h}, \boldsymbol{t})$ and $\boldsymbol{r}_1(\boldsymbol{h}, \boldsymbol{t})$ hold, we need to prove the following equality holds:

$$\boldsymbol{r}_1(\boldsymbol{h}, \boldsymbol{t}) \Rightarrow \boldsymbol{r}_2(\boldsymbol{h}, \boldsymbol{t}).$$

Then we need to prove

$$\boldsymbol{t} = \boldsymbol{r}_1 \otimes \boldsymbol{h} \Rightarrow \boldsymbol{t} = \boldsymbol{r}_2 \otimes \boldsymbol{h}. \tag{12}$$

We assume that the Eq.(12) holds, then we have

$$(\boldsymbol{t} = \boldsymbol{r}_1 \otimes \boldsymbol{h}) \wedge (\boldsymbol{t} = \boldsymbol{r}_2 \otimes \boldsymbol{h}) \Rightarrow \boldsymbol{r}_1 \otimes \boldsymbol{h} = \boldsymbol{r}_2 \otimes \boldsymbol{h}. \tag{13}$$

we denote $\boldsymbol{r}_1$ as $\begin{bmatrix} \mathbf{R}_1 & \mathbf{t}_1 \\ 0 & 1 \end{bmatrix}$ and $\boldsymbol{r}_2$ as $\begin{bmatrix} \mathbf{R}_2 & \mathbf{t}_2 \\ 0 & 1 \end{bmatrix}$, where $\boldsymbol{R}_1 = \begin{bmatrix} \frac{a_1}{\sqrt{a_1^2+b_1^2}} & \frac{b_1}{\sqrt{a_1^2+b_1^2}} \\ \frac{c_1}{\sqrt{c_1^2+d_1^2}} & \frac{d_1}{\sqrt{c_1^2+d_1^2}} \end{bmatrix}$ with $a_1 b_1 + c_1 d_1 = 0$, $\boldsymbol{R}_2 = \begin{bmatrix} \frac{a_2}{\sqrt{a_2^2+b_2^2}} & \frac{b_2}{\sqrt{a_2^2+b_2^2}} \\ \frac{c_2}{\sqrt{c_2^2+d_2^2}} & \frac{d_2}{\sqrt{c_2^2+d_2^2}} \end{bmatrix}$ with $a_2 b_2 + c_2 d_2 = 0$, $\boldsymbol{t}_1 \in \mathbb{R}^2$ and $\boldsymbol{t}_2 \in \mathbb{R}^2$. Then we bring $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ into the Eq.(13). Such we can get the solutions of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ after solving the equation, which can satisfy the Eq.(12). Then we have $\boldsymbol{r}_1(\boldsymbol{x}, \boldsymbol{y}) \Rightarrow \boldsymbol{r}_2(\boldsymbol{x}, \boldsymbol{y})$.

Proof of intersection pattern. In order to prove the intersection pattern, if $\boldsymbol{r}_1(\boldsymbol{h}, \boldsymbol{t})$, $\boldsymbol{r}_2(\boldsymbol{h}, \boldsymbol{t})$ and $\boldsymbol{r}_3(\boldsymbol{h}, \boldsymbol{t})$ hold, we have the following equality:

$$(\boldsymbol{t} = \boldsymbol{r}_1 \otimes \boldsymbol{h}) \wedge (\boldsymbol{t} = \boldsymbol{r}_2 \otimes \boldsymbol{h}) \Rightarrow \left(\begin{bmatrix} \mathbf{R}_1 & \mathbf{v}_1 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{h} \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{t} \\ 1 \end{bmatrix} \right) \wedge \left(\begin{bmatrix} \mathbf{R}_2 & \mathbf{v}_2 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{h} \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{t} \\ 1 \end{bmatrix} \right).$$

Then we have

$$\boldsymbol{R}_1 \otimes \boldsymbol{h} + \boldsymbol{v}_1 = \boldsymbol{t},\ \boldsymbol{R}_2 \otimes \boldsymbol{h} + \boldsymbol{v}_2 = \boldsymbol{t}.$$

We multiply the left and right parts of the first equation by $\boldsymbol{v}_2$ respectively, and we multiply the left and right parts of the second equation by $\boldsymbol{v}_1$ respectively. Then we have:

$$\boldsymbol{v}_2 \otimes \boldsymbol{R}_1 \otimes \boldsymbol{h} + \boldsymbol{v}_2 \otimes \boldsymbol{v}_1 = \boldsymbol{v}_2 \otimes \boldsymbol{t},\ \boldsymbol{v}_1 \otimes \boldsymbol{R}_2 \otimes \boldsymbol{h} + \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 = \boldsymbol{v}_1 \otimes \boldsymbol{t}.$$

Make the first formula subtract the second formula above, and we can get

$$(\boldsymbol{v}_2 \otimes \boldsymbol{R}_1 - \boldsymbol{v}_1 \otimes \boldsymbol{R}_2) \otimes \boldsymbol{h} = (\boldsymbol{v}_2 - \boldsymbol{v}_1) \otimes \boldsymbol{t}.$$

Since $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$, $v_1 \neq v_2$. Then we have

$$(\boldsymbol{v}_2 - \boldsymbol{v}_1)^{-1} \otimes (\boldsymbol{v}_2 \otimes \boldsymbol{R}_1 - \boldsymbol{v}_1 \otimes \boldsymbol{R}_2) \otimes \boldsymbol{h} = \boldsymbol{t}.$$

Next let $\boldsymbol{r}_3 = \begin{bmatrix} \mathbf{R}_3 & \boldsymbol{v}_3 \\ 0 & 1 \end{bmatrix}$, where $\mathbf{R}_3 = (\boldsymbol{v}_2 - \boldsymbol{v}_1)^{-1} \otimes (\boldsymbol{v}_2 \otimes \boldsymbol{R}_1 - \boldsymbol{v}_1 \otimes \boldsymbol{R}_2)$ and $\boldsymbol{v}_3 = \mathbf{0}$. Then we have $\boldsymbol{r}_1(\boldsymbol{x},\boldsymbol{y}) \wedge \boldsymbol{r}_2(\boldsymbol{x},\boldsymbol{y}) \Rightarrow \boldsymbol{r}_3(\boldsymbol{x},\boldsymbol{y})$.

Proof of mutual exclusion pattern. To prove the mutual exclusion pattern, if $\boldsymbol{r}_1(\boldsymbol{h},\boldsymbol{t})$ and $\boldsymbol{r}_2(\boldsymbol{h},\boldsymbol{t})$ hold, where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are orthogonal ($\boldsymbol{r}_1\boldsymbol{r}_2 = \mathbf{0}$), we have the formula as follows:

$$\begin{aligned} &(\boldsymbol{t} = \boldsymbol{r}_1 \otimes \boldsymbol{h}) \wedge (\boldsymbol{t} = \boldsymbol{r}_2 \otimes \boldsymbol{h}) \\ \Rightarrow &\boldsymbol{r}_1 \otimes \boldsymbol{h} = \boldsymbol{r}_2 \otimes \boldsymbol{h} \\ \Rightarrow &\boldsymbol{r}_1 \otimes \boldsymbol{r}_1 \otimes \boldsymbol{h} = \boldsymbol{r}_1 \otimes \boldsymbol{r}_2 \otimes \boldsymbol{h} \\ \Rightarrow &\boldsymbol{r}_1 \otimes \boldsymbol{r}_1 \otimes \boldsymbol{h} = \mathbf{0} \\ \Rightarrow &\perp . \end{aligned}$$

Therefore, we have $\boldsymbol{r}_1(\boldsymbol{x},\boldsymbol{y}) \wedge \boldsymbol{r}_2(\boldsymbol{x},\boldsymbol{y}) \Rightarrow \perp$.

Hierarchy Estimates

There are two main metrics to estimate how hierarchical a relation is, which is called the curvature estimate ξ_G (it means how much the graph is tree-like) and the Krackhardt hierarchy score Khs_G (it means how many small loops in the graph). Each of these two measures has its own merits: the global hierarchical behaviours can be captured by the curvature estimate, and the local behaviour can be captured by the Krackhardt score.

Curvature estimate. In order to estimate the curvature of a relation r, we restrict to the undirected graph G_r spanned by the edges labeled as r. As shown in (Gu et al. 2019a), for the given vertices $\{a,b,c\}$, we denote $\xi_{G_r}(a,b,c)$ as the curvature estimate of a triangle in G_r, which can be given as follows:

$$\begin{aligned} \xi_{G_r}(a,b,c) = &\frac{1}{2d_{G_r}(a,m)} \left(d_{G_r}(a,m)^2 + d_{G_r}(b,c)^2/4 \right. \\ &\left. - \left(d_{G_r}(a,b)^2 + d_{G_r}(a,c)^2\right)/2\right), \end{aligned}$$

where m denotes the midpoint of the shortest path from b to c. For the triangle, its curvature estimate is positive if it is in circles; its curvature estimate is zero if it is in lines; its curvature estimate is negative if it is in trees. Moreover, following (Gu et al. 2019a), suppose there is a triangle in a Riemannian manifold M lying on a plane, and we can use $\xi_M(a,b,c)$ to estimate the sectional curvature of the plane. Denote the total number of connected components in G_r as m_r, denote the number of nodes in the component $c_{i,r}$ as $N_{i,r}$ and denote the mean of the estimated curvatures of the sampled triangles as ξ_{G_r}. We then sample 1000 $w_{i,r}$ triangles from each connected component $c_{i,r}$ of G_r where $w_{i,r} = \frac{N_{i,r}^3}{\sum_{i=1}^{m_r} N_{i,r}^3}$. Next, we take the weighted average of the relation curvatures ξ_{G_r} with respect to the weights $\frac{\sum_{i=1}^{m_r} N_{i,r}^3}{\sum_r \sum_{i=1}^{m_r} N_{i,r}^3}$ for all the graphs.

Krackhardt hierarchy score. For the directed graph G_r which is spanned by the relation r, we denote the adjacency matrix (If there is an edge connecting node i to node j, we have $R_{i,j} = 1$. If there is not an edge connecting node i to node j, we have $R_{i,j} \neq 1$.) as R. Then we have:

$$\mathrm{Khs}_{G_r} = \frac{\sum_{i,j=1}^{n} R_{i,j}\left(1 - R_{j,i}\right)}{\sum_{i,j=1}^{n} R_{i,j}} \tag{14}$$

More details can refer to (Everett and Krackhardt 2012). For fully observed symmetric relations, each edge of which is in a two-edge loop, it can be seen that $\mathrm{Khs}_{G_r} = 0$. And for the anti-symmetric relations, we have $\mathrm{Khs}_{G_r} = 1$.

Experimental Details

We conduct the hyperparameter search for the negative sample size, learning rate, batch size, dimension and the number of epoch. We train each model for 150 epochs and use early stopping after 100 epochs if the validation MRR stops increasing. Moreover, we count the running time of each epoch of GIE on different datasets: 20s for WN18RR, 34s for FB15K237, 65s for YAGO3-10. We need to run 150 epochs in total. Therefore, the overall time for each datasets is between 50 minutes and 162 minutes.

In the test dataset, we measure and compare the performance of models listed in the paper for each test triplet (h,r,t). For all $h' \in \mathcal{E}$ (where $\mathcal{E}$ is the set of entities), we calculate the scores of triplets (h',r,t) and then denote the ranking of the triplet (h,r,t) among these triplets as $rank_h$. Next, for all $t' \in \mathcal{E}$, we compute the scores of triplets (h,r,t') and denote the ranking of the triplet (h,r,t) among these triplets as $rank_t$. According to the calculation above, MRR is calculated as follows, which is the mean of the reciprocal rank for the test dataset:

$$MRR = \frac{1}{2 * |\Omega|} \sum_{(h,r,t)\in\Omega} \frac{1}{rank_h(h,r,t)} + \frac{1}{rank_t(h,r,t)}.$$

where Ω denotes the set of observed triplets in the test dataset.

relation name	ξ_G	Khs	MuRE	MuRP	MuRS	MuRMP	MuRMP-a	GIE
also_see	-2.09	.24	.634	.705	.483	.692	.725	**.759**
hypernym	-2.46	.99	.161	.228	.126	.222	.232	**.262**
has_part	-1.43	1	.215	.282	.301	.134	.316	**.334**
member_meronym	-2.90	1	.272	.346	.138	.343	.350	**.360**
synset_domain_topic_of	-0.69	.99	.316	.430	.163	.421	**.445**	.435
instance_hypernum	-0.82	1	.488	.471	.258	.345	.491	**.501**
member_of_domain_region	-0.78	1	.308	.347	.201	.344	.349	**.404**
member_of_domain_usage	-0.74	1	.396	.417	.228	.416	.420	**.438**
derivationally_related_form	-3.84	0.4	.954	.967	.965	.967	**.970**	.968
similar_to	-1.00	0	**1**	**1**	**1**	**1**	**1**	**1**
verb_group	-0.5	0	.974	.974	.976	.976	.981	**.984**

Table 5: Comparison of hits@10 for WN18RR. MuRMP-a represents MuRMP-autoKT in (Wang et al. 2021).

	WN18RR				FB15k-237				YAGO3-10			
Analysis	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
GIE_1	.472	.438	.474	.534	.336	.231	.373	.526	.534	.477	.584	.680
GIE_2	.477	.440	.483	.541	.338	.236	.371	.526	.538	.479	.588	.674
GIE_3	.471	.436	.470	.531	.332	.228	.375	.523	.531	.473	.581	.671

Table 6: Analysis on different variants of scoring function.