

# On the Equivalence of Holographic and Complex Embeddings for Link Prediction

Katsuhiko Hayashi

NTT Communication Science Laboratories, Seika-cho, Kyoto 619 0237, Japan  
hayashi.katsuhiko@lab.ntt.co.jp

Masashi Shimbo

Nara Institute of Science and Technology, Ikoma, Nara 630 0192, Japan  
shimbo@is.naist.jp

## Abstract

We show the equivalence of two state-of-the-art models for link prediction/knowledge graph completion: Nickel et al.'s holographic embeddings and Trouillon et al.'s complex embeddings. We first consider a spectral version of the holographic embeddings, exploiting the frequency domain in the Fourier transform for efficient computation. The analysis of the resulting model reveals that it can be viewed as an instance of the complex embeddings with a certain constraint imposed on the initial vectors upon training. Conversely, any set of complex embeddings can be converted to a set of equivalent holographic embeddings.

## 1 Introduction

Recently, there have been efforts to build and maintain large-scale knowledge bases represented in the form of a graph (*knowledge graph*) (Auer et al., 2007; Bollacker et al., 2008; Suchanek et al., 2007). Although these knowledge graphs contain billions of relational facts, they are known to be incomplete. *Knowledge graph completion* (KGC) (Nickel et al., 2015) aims at augmenting missing knowledge in an incomplete knowledge graph automatically. It can be viewed as a task of *link prediction* (Liben-Nowell and Kleinberg, 2003; Hasan and Zaki, 2011) studied in the field of statistical relational learning (Getoor and Taskar, 2007). In recent years, methods based on vector embeddings of graphs have been actively pursued as a scalable approach to KGC (Bordes et al., 2011; Socher et al., 2013; Guu et al., 2015; Yang et al., 2015; Nickel et al., 2016; Trouillon et al., 2016b).

In this paper, we investigate the connection

between two models of graph embeddings that have emerged along this line of research: The *holographic embeddings* (Nickel et al., 2016) and the *complex embeddings* (Trouillon et al., 2016b). These models are simple yet achieve the current state-of-the-art performance in KGC.

We begin by showing that holographic embeddings can be trained entirely in the frequency domain induced by the Fourier transform, thereby reducing the time needed to compute the scoring function from  $O(n \log n)$  to  $O(n)$ , where  $n$  is the dimension of the embeddings.

The analysis of the resulting training method reveals that the Fourier transform of holographic embeddings can be regarded as an instance of the complex embeddings, with a specific constraint (viz. conjugate symmetry property) cast on the initial values.

Conversely, we also show that every set of complex embeddings has a set of holographic embeddings (with real vectors) that is equivalent, in the sense that their scoring functions are equal up to scaling.

## 2 Preliminaries

Let  $i$  denote the imaginary unit,  $\mathbb{R}$  be the set of real values, and  $\mathbb{C}$  the set of complex values. We write  $[\mathbf{v}]_j$  to denote the  $j$ th component of vector  $\mathbf{v}$ . A superscript T (e.g.,  $\mathbf{v}^T$ ) represents vector/matrix transpose. For a complex scalar  $z$ , vector  $\mathbf{z}$ , and matrix  $\mathbf{Z}$ ,  $\bar{z}$ ,  $\bar{\mathbf{z}}$ , and  $\bar{\mathbf{Z}}$  represent their complex conjugate, and  $\text{Re}(z)$ ,  $\text{Re}(\mathbf{z})$ , and  $\text{Re}(\mathbf{Z})$  denote their real parts, respectively.

Let  $\mathbf{x} = [x_0 \cdots x_{n-1}]^T \in \mathbb{R}^n$  and  $\mathbf{y} = [y_0 \cdots y_{n-1}]^T \in \mathbb{R}^n$ . Note that the vector indices start from 0 for notational convenience. The *circular convolution* of  $\mathbf{x}$  and  $\mathbf{y}$ , denoted by  $\mathbf{x} * \mathbf{y}$ , is

defined by

$$[\mathbf{x} * \mathbf{y}]_j = \sum_{k=0}^{n-1} x_{[(j-k) \bmod n]} y_k, \quad (1)$$

where mod denotes modulus. Likewise, *circular correlation*  $\mathbf{x} \star \mathbf{y}$  is defined by

$$[\mathbf{x} \star \mathbf{y}]_j = \sum_{k=0}^{n-1} x_{[(k-j) \bmod n]} y_k. \quad (2)$$

While circular convolution is commutative, circular correlation is not; i.e.,  $\mathbf{x} * \mathbf{y} = \mathbf{y} * \mathbf{x}$ , but  $\mathbf{x} \star \mathbf{y} \neq \mathbf{y} \star \mathbf{x}$  in general. As it can be verified with Eqs. (1) and (2),  $\mathbf{x} \star \mathbf{y} = \text{flip}(\mathbf{x}) * \mathbf{y}$ , where  $\text{flip}(\mathbf{x}) = [x_{n-1} \cdots x_0]^T$  is a vector obtained by arranging the components of  $\mathbf{x}$  in reverse.

For  $n$ -dimensional vectors, naively computing circular convolution/correlation by Eqs. (1) and (2) requires  $O(n^2)$  multiplications. However, we can take advantage of the Fast Fourier Transform (FFT) algorithm to accelerate the computation: For circular convolution, first compute the discrete Fourier transform (DFT) of  $\mathbf{x}$  and  $\mathbf{y}$ , and then compute the inverse DFT of their elementwise vector product, i.e.,

$$\mathbf{x} * \mathbf{y} = \mathfrak{F}^{-1}(\mathfrak{F}(\mathbf{x}) \odot \mathfrak{F}(\mathbf{y})),$$

where  $\mathfrak{F} : \mathbb{R}^n \rightarrow \mathbb{C}^n$  and  $\mathfrak{F}^{-1} : \mathbb{C}^n \rightarrow \mathbb{R}^n$  respectively denote the DFT and inverse DFT, and  $\odot$  denotes the elementwise product. DFT and inverse DFT can be computed in  $O(n \log n)$  time with the FFT algorithm, and thus the computation time for circular convolution is also  $O(n \log n)$ . The same can be said of circular correlation. Since  $\mathfrak{F}(\text{flip}(\mathbf{x})) = \overline{\mathfrak{F}(\mathbf{x})}$ , we have

$$\mathbf{x} \star \mathbf{y} = \mathfrak{F}^{-1}(\overline{\mathfrak{F}(\mathbf{x})} \odot \mathfrak{F}(\mathbf{y})). \quad (3)$$

By analogy to how the Fourier transform is used in signal processing, the original real space  $\mathbb{R}^n$  is called the “time” domain, and the complex space  $\mathbb{C}^n$  where DFT vectors reside is called the “frequency” domain.

### 3 Holographic embeddings for knowledge graph completion

#### 3.1 Knowledge graph completion

Let  $\mathcal{E}$  and  $\mathcal{R}$  be the finite sets of *entities* and (*binary*) *relations* over entities, respectively. For each relation  $r \in \mathcal{R}$  and each pair  $s, o \in \mathcal{E}$  of entities,

Table 1: Correspondence between operations in time and frequency domains.  $\mathbf{r} \leftrightarrow \boldsymbol{\rho}$  indicates  $\boldsymbol{\rho} = \mathfrak{F}(\mathbf{r})$  (and also  $\mathbf{r} = \mathfrak{F}^{-1}(\boldsymbol{\rho})$ ).

operation	time	frequency
scalar mult.	$\alpha \mathbf{x}$	$\alpha \mathfrak{F}(\mathbf{x})$
summation	$\mathbf{x} + \mathbf{y}$	$\mathfrak{F}(\mathbf{x}) + \mathfrak{F}(\mathbf{y})$
flip	$\text{flip}(\mathbf{x})$	$\overline{\mathfrak{F}(\mathbf{x})}$
convolution	$\mathbf{x} * \mathbf{y}$	$\mathfrak{F}(\mathbf{x}) \odot \mathfrak{F}(\mathbf{y})$
correlation	$\mathbf{x} \star \mathbf{y}$	$\overline{\mathfrak{F}(\mathbf{x})} \odot \mathfrak{F}(\mathbf{y})$
dot product	$\mathbf{x} \cdot \mathbf{y}$	$\frac{1}{n} \mathfrak{F}(\mathbf{x}) \cdot \mathfrak{F}(\mathbf{y})$

we are interested in whether  $r(s, o)$  holds<sup>1</sup> or not; we write  $r(s, o) = +1$  if it holds, and  $r(s, o) = -1$  if not. To be precise, given a *training set*  $\mathcal{D} = \mathcal{R} \times \mathcal{E} \times \mathcal{E} \times \{-1, +1\}$  such that  $(r, s, o, y) \in \mathcal{D}$  indicates  $y = r(s, o)$ , our task is to design a *scoring function*  $f : \mathcal{R} \times \mathcal{E} \times \mathcal{E} \rightarrow \mathbb{R}$  such that for each of the triples  $(r, s, o)$  not observed in  $\mathcal{D}$ , function  $f$  should give a higher value if  $r(s, o) = +1$  is more likely, and a smaller value for those that are less likely. If necessary,  $f(r, s, o)$  can be converted to probability by  $P[r(s, o) = +1] = \sigma(f(r, s, o))$ , where  $\sigma : \mathbb{R} \rightarrow (0, 1)$  is a sigmoid function.

Dataset  $\mathcal{D}$  can be regarded as a directed graph in which nodes represent entities  $\mathcal{E}$  and edges are labeled by relations  $\mathcal{R}$ . Thus, the task is essentially that of *link prediction* (Liben-Nowell and Kleinberg, 2003; Hasan and Zaki, 2011). Often, it is also called *knowledge graph completion*.

#### 3.2 Holographic embeddings (HolE)

Nickel et al. (2016) proposed *holographic embeddings* (HolE) for knowledge graph completion. Using training data  $\mathcal{D}$ , this method learns the vector embeddings  $\mathbf{e}_k \in \mathbb{R}^n$  of entities  $k \in \mathcal{E}$  and the embeddings  $\mathbf{w}_r \in \mathbb{R}^n$  of relations  $r \in \mathcal{R}$ . The score for triple  $(r, s, o)$  is then given by

$$f_{\text{HolE}}(r, s, o) = \mathbf{w}_r \cdot (\mathbf{e}_s \star \mathbf{e}_o). \quad (4)$$

Eq. (4) can be evaluated in time  $O(n \log n)$  if  $\mathbf{e}_s \star \mathbf{e}_o$  is computed by Eq. (3).

### 4 Spectral training of HolE

To compute the circular correlation in the scoring function of Eq. (4) efficiently, Nickel et al. (2016) used Eq. (3) in Section 2 and FFT. In this section, we extend this technique further, and con-

<sup>1</sup>Depending on the context, letter  $r$  is used either as an index to an element in  $\mathcal{R}$  or the binary relation it signifies.

sider training HoIE solely in the frequency domain. That is, real-valued embeddings  $\mathbf{e}_k, \mathbf{w}_r \in \mathbb{R}^n$  in the original “time” domain are abolished, and instead we train their DFT counterparts  $\boldsymbol{\epsilon}_k = \mathfrak{F}(\mathbf{e}_k) \in \mathbb{C}^n$  and  $\boldsymbol{\omega}_k = \mathfrak{F}(\mathbf{w}_r) \in \mathbb{C}^n$  in the frequency domain. This formulation eliminates the need of FFT and inverse FFT, which are the major computational bottleneck in HoIE. As a result, Eq. (4) can be computed in time  $O(n)$  directly from  $\boldsymbol{\epsilon}_k$  and  $\boldsymbol{\omega}_k$ .

Indeed, equivalent counterparts in the frequency domain exist for not only convolution/correlation but all other computations needed for HoIE: scalar multiplication, summation (needed when vectors are updated by stochastic gradient descent), and dot product (used in Eq. (4)). The frequency-domain equivalents for these operations are summarized in Table 1. All of these can be performed efficiently (in linear time) in the frequency domain.

In particular, the following relation holds for the dot product between any “time” vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ .

$$\mathbf{x} \cdot \mathbf{y} = \frac{1}{n} \tilde{\mathfrak{F}}(\mathbf{x}) \cdot \tilde{\mathfrak{F}}(\mathbf{y}), \quad (5)$$

where the dot product on the right-hand side is the complex inner product defined by  $\mathbf{a} \cdot \mathbf{b} = \bar{\mathbf{a}}^T \mathbf{b}$ . Eq. (5) is known as Parseval’s theorem (also called the *power theorem* in (Smith, 2007)), and it states that dot products in two domains are equal up to scaling.

After embeddings  $\boldsymbol{\epsilon}_k, \boldsymbol{\omega}_r \in \mathbb{C}^n$  are learned in the frequency domain, their time-domain counterparts  $\mathbf{e}_k = \mathfrak{F}^{-1}(\boldsymbol{\epsilon}_k)$  and  $\mathbf{w}_r = \mathfrak{F}^{-1}(\boldsymbol{\omega}_r)$  can be recovered if needed, but this is not required as far as computation of the scoring function is concerned. Thanks to Parseval’s theorem, Eq. (4) can be directly computed from the frequency vectors  $\boldsymbol{\epsilon}_k, \boldsymbol{\omega}_r \in \mathbb{C}^n$  by

$$f_{\text{HoIE}}(r, s, o) = \frac{1}{n} \boldsymbol{\omega}_r \cdot (\bar{\boldsymbol{\epsilon}}_s \odot \boldsymbol{\epsilon}_o). \quad (6)$$

#### 4.1 Conjugate symmetry of spectral components

A complex vector  $\boldsymbol{\xi} = [\xi_0 \dots \xi_{n-1}]^T \in \mathbb{C}^n$  is said to be *conjugate symmetric* (or *Hermitian*) if  $\xi_j = \bar{\xi}_{[(n-j) \bmod n]}$  for  $j = 0, \dots, n-1$ , or, in other words, if it can be written in the form

$$\boldsymbol{\xi} = \begin{cases} \left[ \begin{array}{ccc} \xi_0 & \boldsymbol{\gamma} & \text{flip}(\bar{\boldsymbol{\gamma}}) \end{array} \right]^T, & \text{if } n \text{ is odd,} \\ \left[ \begin{array}{ccc} \xi_0 & \boldsymbol{\gamma} & \xi_{n/2} \quad \text{flip}(\bar{\boldsymbol{\gamma}}) \end{array} \right]^T, & \text{if } n \text{ is even,} \end{cases}$$

for some  $\boldsymbol{\gamma} \in \mathbb{C}^{\lceil n/2 \rceil - 1}$  and  $\xi_0, \xi_{n/2} \in \mathbb{R}$ .

The DFT  $\mathfrak{F}(\mathbf{x})$  is conjugate symmetric if and only if  $\mathbf{x}$  is a real vector. Thus, maintaining conjugate symmetry of “frequency” vectors is the key to ensure their “time” counterparts remain in real space. Below, we verify that this property is indeed preserved with stochastic gradient descent. Moreover, conjugate symmetry provides a sufficient condition under which dot product takes a real value. It also has implications on space requirement. These topics are covered in the rest of this section.

#### 4.2 Vector initialization and update in frequency domain

Typically, at the beginning of training HoIE, each individual embedding is initialized by a random vector. When we train HoIE in the frequency domain, we could first generate a random real vector, regard them as a HoIE vector in the time domain, and compute its DFT as the initial value in the frequency domain. An alternative, easier approach is to directly generate a random complex vector that is conjugate symmetric, and use it as the initial frequency vector. This guarantees the inverse DFT to be a real vector, i.e., there exists a valid corresponding image in the time domain.

Given a training set  $\mathcal{D}$  (see Section 3.1), HoIE is trained by minimizing the following objective function over parameter matrix  $\Theta = [\mathbf{e}_1 \dots \mathbf{e}_{|\mathcal{E}|} \quad \mathbf{w}_1 \dots \mathbf{w}_{|\mathcal{R}|}] \in \mathbb{R}^{n \times (|\mathcal{E}| + |\mathcal{R}|)}$ :

$$\sum_{(r,s,o,y) \in \mathcal{D}} \log\{1 + \exp(-y f_{\text{HoIE}}(r, s, o))\} + \lambda \|\Theta\|_F^2 \quad (7)$$

where  $\lambda \in \mathbb{R}$  is the hyperparameter controlling the degree of regularization, and  $\|\cdot\|_F$  denotes the Frobenius norm.

In our version of spectral training of HoIE, the parameters matrix consists of frequency vectors  $\boldsymbol{\epsilon}_k$  and  $\boldsymbol{\omega}_r$  instead of  $\mathbf{e}_k$  and  $\mathbf{w}_r$ , i.e.,  $\Theta = [\boldsymbol{\epsilon}_1 \dots \boldsymbol{\epsilon}_{|\mathcal{E}|} \quad \boldsymbol{\omega}_1 \dots \boldsymbol{\omega}_{|\mathcal{R}|}] \in \mathbb{C}^{n \times (|\mathcal{E}| + |\mathcal{R}|)}$ . Let us discuss the stochastic gradient descent (SGD) update with respect to these frequency vectors. In particular, we are interested in whether conjugate symmetry of vectors is preserved by the update.

Suppose vectors  $\boldsymbol{\omega}_r, \boldsymbol{\epsilon}_s, \boldsymbol{\epsilon}_o$  are conjugate symmetric. Neglecting the contribution from the regularization term<sup>2</sup> in Eq. (7), we see that in

<sup>2</sup>It can be easily verified that the contribution from the regularization term to SGD update do not violate conjugate symmetry.

an SGD update step,  $\alpha \partial f_{\text{HoLE}} / \partial \mathbf{w}_r$ ,  $\alpha \partial f_{\text{HoLE}} / \partial \mathbf{e}_s$ , and  $\alpha \partial f_{\text{HoLE}} / \partial \mathbf{e}_o$  are respectively subtracted from  $\mathbf{w}_r$ ,  $\mathbf{e}_s$ ,  $\mathbf{e}_o$ , where  $\alpha \in \mathbb{R}$  is a factor not depending on these parameters. Noting the equalities

$$\mathbf{w}_r \cdot (\mathbf{e}_s \star \mathbf{e}_o) = \mathbf{e}_s \cdot (\mathbf{w}_r \star \mathbf{e}_o) = \mathbf{e}_o \cdot (\mathbf{w}_r \star \mathbf{e}_s)$$

(see (Nickel et al., 2016, Eq. (12), p. 1958)) and their frequency counterparts

$$\mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o) = \mathbf{e}_s \cdot (\overline{\mathbf{w}_r} \odot \mathbf{e}_o) = \mathbf{e}_o \cdot (\mathbf{w}_r \odot \mathbf{e}_s),$$

obtained through the translation of Table 1, we can derive

$$\begin{aligned} \frac{\partial f_{\text{HoLE}}}{\partial \mathbf{w}_r} &= \overline{\mathbf{e}_s} \odot \mathbf{e}_o, \\ \frac{\partial f_{\text{HoLE}}}{\partial \mathbf{e}_s} &= \overline{\mathbf{w}_r} \odot \mathbf{e}_o, \\ \frac{\partial f_{\text{HoLE}}}{\partial \mathbf{e}_o} &= \mathbf{w}_r \odot \mathbf{e}_s. \end{aligned}$$

As seen from above, conjugation, scalar multiplication, summation, and elementwise product are used in the SGD update. And it is straightforward to verify that all these operations preserve conjugate symmetry. It follows that if  $\mathbf{w}_r$ ,  $\mathbf{e}_s$ ,  $\mathbf{e}_o$  are initially conjugate symmetric, they will remain so during the course of training, which assures that the inverse DFTs of the learned embeddings are real vectors.

### 4.3 Real-valued dot product

In the scoring function of HoLE (Eq. (4)), dot product is used for generating a real-valued ‘‘score’’ out of two vectors,  $\mathbf{w}_r$  and  $\mathbf{e}_s \odot \mathbf{e}_o$ . Likewise, in Eq. (6), the dot product is applied to  $\mathbf{w}_r$  and  $\mathbf{e}_s \odot \mathbf{e}_o$ , which are complex-valued. However, provided that the conjugate symmetry of these vectors is maintained, their dot product is always real. This follows from Parseval’s theorem; the inverse DFTs of these frequency vectors are real, and thus their dot product is also real. Therefore, the dot product of the corresponding frequency vectors is real as well, according to Eq. (5).

### 4.4 Space requirement

A general complex vector  $\boldsymbol{\xi} \in \mathbb{C}^n$  can be stored in memory as  $2n$  floating-point numbers, i.e., one each for the real and imaginary part of a component. In our spectral representation of HoLE, however, the first  $\lfloor n/2 \rfloor$  components suffice to specify the frequency vector  $\boldsymbol{\xi}$ , since the vector is conjugate symmetric. Moreover,  $\xi_0$  (and  $\xi_{n/2}$  if  $n$

is even) are real values. Thus, a spectral representation of HoLE can be specified with exactly  $n$  floating-point numbers, which can be stored in the same amount of memory as needed by the original HoLE.

## 5 Relation to Trouillon et al.’s complex embeddings

### 5.1 Complex embeddings (Complex)

Trouillon et al. (2016b) proposed a model of embedding-based knowledge graph completion, called *complex embeddings* (Complex). The objective is similar to Nickel et al.’s; the embeddings  $\mathbf{e}_k$  of entities and  $\mathbf{w}_r$  of relations are to be learned. In their model, however, these vectors are complex-valued, and are based on the eigen-decomposition of complex matrix  $\mathbf{X}_r = \mathbf{E}\mathbf{W}_r\overline{\mathbf{E}}^T$  that encodes relation  $r \in \mathcal{R}$  over pairs of entities, where  $\mathbf{X}_r \in \mathbb{C}^{|\mathcal{E}| \times |\mathcal{E}|}$ ,  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_{|\mathcal{E}|}]^T \in \mathbb{C}^{|\mathcal{E}| \times n}$ , and  $\mathbf{W}_r = \text{diag}(\mathbf{w}_r) \in \mathbb{C}^{n \times n}$  is a diagonal matrix (with diagonal elements  $\mathbf{w}_r \in \mathbb{C}^n$ ). In practice,  $\mathbf{X}_r$  needs to be a real matrix, because its  $(r, s)$ -component must define the score for  $r(s, o)$ . To this end, Trouillon et al. simply extracted the real part; i.e.,  $\mathbf{X}_r = \text{Re}(\mathbf{E}\mathbf{W}_r\overline{\mathbf{E}}^T)$ . Trouillon et al. (2016a) advocated this approach, by showing that any real matrix  $\mathbf{X}_r$  can be expressed in this form.

With this formulation, the score for triple  $(r, s, o)$  is given by

$$f_{\text{Complex}}(r, s, o) = \text{Re} \left( \sum_{j=0}^{n-1} [\mathbf{w}_r]_j [\mathbf{e}_s]_j \overline{[\mathbf{e}_o]_j} \right). \quad (8)$$

### 5.2 Equivalence of holographic and complex embeddings

Now let us rewrite Eq. (8). Noting the definition of complex dot product, i.e.,  $\mathbf{a} \cdot \mathbf{b} = \overline{\mathbf{a}}^T \mathbf{b}$ , we have

$$\begin{aligned} \sum_{j=0}^{n-1} [\mathbf{w}_r]_j [\mathbf{e}_s]_j \overline{[\mathbf{e}_o]_j} &= (\mathbf{e}_s \odot \overline{\mathbf{e}_o})^T \mathbf{w}_r \\ &= \overline{(\mathbf{e}_s \odot \overline{\mathbf{e}_o})} \cdot \mathbf{w}_r \quad (\because \mathbf{a} \cdot \mathbf{b} = \overline{\mathbf{a}}^T \mathbf{b}) \\ &= \overline{(\mathbf{e}_s \odot \mathbf{e}_o)} \cdot \mathbf{w}_r \\ &= \mathbf{w}_r \cdot \overline{(\mathbf{e}_s \odot \mathbf{e}_o)} \quad (\because \mathbf{a} \cdot \mathbf{b} = \overline{\mathbf{b} \cdot \mathbf{a}}) \end{aligned}$$

and since  $\text{Re}(\mathbf{z}) = \text{Re}(\overline{\mathbf{z}})$ ,

$$\text{Re}(\mathbf{w}_r \cdot \overline{(\mathbf{e}_s \odot \mathbf{e}_o)}) = \text{Re}(\mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o)).$$

Thus, Eq. (8) can be written as

$$f_{\text{Complex}}(r, s, o) = \text{Re}(\mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o)). \quad (9)$$

Here, a marked similarity is noticeable between Eq. (9) and Eq. (6), the scoring function of our spectral version of HoIE (spectral HoIE); ComplEx extracts the real part of complex dot product, whereas in the spectral HoIE, dot product is guaranteed to be real because all embeddings satisfy conjugate symmetry. Indeed, Eq. (6) can be equally written as

$$f_{\text{HoIE}}(r, s, o) = \frac{1}{n} \text{Re}(\mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o)). \quad (10)$$

although the operator  $\text{Re}(\cdot)$  in this formula is redundant, since the inner product is guaranteed to be real-valued. Nevertheless, Eq. (10) elucidates the fact that the spectral HoIE can be viewed as an instance of ComplEx, with the embeddings constrained to be conjugate symmetric to make the inner product in Eq. (10) real-valued.

Conversely, given a set of complex embeddings for entities and relations, we can construct their equivalent holographic embeddings, in the sense that  $f_{\text{ComplEx}}(r, s, o) = c f_{\text{HoIE}}(r, s, o)$  for every  $r, s, o$ , where  $c > 0$  is a constant. For each  $n$ -dimensional complex embeddings  $\mathbf{x} \in \{\mathbf{e}_k\}_{k \in \mathcal{E}} \cup \{\mathbf{w}_r\}_{r \in \mathcal{R}} \subset \mathbb{C}^n$  computed by ComplEx, we make a corresponding HoIE  $\mathfrak{h}(\mathbf{x}) \in \mathbb{R}^{2n+1}$  as follows: For a given complex embedding  $\mathbf{x} = [x_0 \cdots x_{n-1}] \in \mathbb{C}^n$ , first compute  $\mathfrak{s}(\mathbf{x}) \in \mathbb{C}^{2n+1}$  by

$$\begin{aligned} \mathfrak{s}(\mathbf{x}) &= \left[ 0 \quad x_0 \quad \cdots \quad x_{n-1} \quad \overline{x_{n-1}} \quad \cdots \quad \overline{x_0} \right]^T \\ &= \left[ 0 \quad \mathbf{x} \quad \text{flip}(\overline{\mathbf{x}}) \right]^T \end{aligned} \quad (11)$$

and then define  $\mathfrak{h}(\mathbf{x}) = \mathfrak{F}^{-1}(\mathfrak{s}(\mathbf{x}))$ . Since  $\mathfrak{s}(\mathbf{x})$  is conjugate symmetric,  $\mathfrak{h}(\mathbf{x})$  is a real vector.

To verify if this conversion defines an equivalent scoring function for any triple  $(r, s, o)$ , let us now suppose complex embeddings  $\mathbf{w}_r \in \mathbb{C}^n$  and  $\mathbf{e}_s, \mathbf{e}_o \in \mathbb{C}^n$  are given. Since we regard real vectors  $\mathfrak{h}(\mathbf{w}_r), \mathfrak{h}(\mathbf{e}_s), \mathfrak{h}(\mathbf{e}_o) \in \mathbb{R}^{2n+1}$  as the holographic embeddings of  $r, s$  and  $o$ , respectively, the HoIE score for the triple  $(r, s, o)$  is given as

$$\begin{aligned} f_{\text{HoIE}}(r, s, o) &= \mathfrak{h}(\mathbf{w}_r) \cdot (\mathfrak{h}(\mathbf{e}_s) \star \mathfrak{h}(\mathbf{e}_o)) \\ &= \frac{1}{n} \mathfrak{s}(\mathbf{w}_r) \cdot (\overline{\mathfrak{s}(\mathbf{e}_s)} \odot \mathfrak{s}(\mathbf{e}_o)) \quad (\because \text{Eq. (6)}) \\ &= \frac{1}{n} \mathfrak{s}(\mathbf{w}_r) \cdot \left[ 0 \quad \overline{\mathbf{e}_s} \odot \mathbf{e}_o \quad \text{flip}(\overline{\overline{\mathbf{e}_s} \odot \mathbf{e}_o}) \right]^T \quad (\because \text{Eq. (11)}) \\ &= \frac{1}{n} \left[ 0 \quad \mathbf{w}_r \quad \text{flip}(\overline{\mathbf{w}_r}) \right]^T \cdot \left[ 0 \quad \overline{\mathbf{e}_s} \odot \mathbf{e}_o \quad \text{flip}(\overline{\mathbf{e}_s} \odot \mathbf{e}_o) \right]^T \\ &= \frac{1}{n} \left( \mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o) + \text{flip}(\overline{\mathbf{w}_r}) \cdot \text{flip}(\overline{\overline{\mathbf{e}_s} \odot \mathbf{e}_o}) \right) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{n} \left( \mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o) + \overline{\mathbf{w}_r} \cdot \overline{(\overline{\mathbf{e}_s} \odot \mathbf{e}_o)} \right) \\ &= \frac{1}{n} \left( \mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o) + \overline{\mathbf{w}_r} \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o) \right) \\ &= \frac{2}{n} \text{Re}(\mathbf{w}_r \cdot (\overline{\mathbf{e}_s} \odot \mathbf{e}_o)) \\ &= \frac{2}{n} f_{\text{ComplEx}}(r, s, o), \end{aligned}$$

which shows that  $\mathfrak{h}(\cdot)$  (or  $\mathfrak{s}(\cdot)$ ) gives the desired conversion from ComplEx to HoIE.

## 6 Conclusion

In this paper, we have shown that the holographic embeddings (HoIE) can be trained entirely in the frequency domain. If stochastic gradient descent is used for training, the conjugate symmetry of frequency vectors is preserved, which ensures the existence of the corresponding holographic embedding in the original real space (time domain). Also, this training method eliminates the need of FFT and inverse FFT, thereby reducing the computation time of the scoring function from  $O(n \log n)$  to  $O(n)$ .

Moreover, we have established the equivalence of HoIE and the complex embeddings (ComplEx): The spectral version of HoIE is subsumed by ComplEx as a special case in which conjugate symmetry is imposed on the embeddings. Conversely, every set of complex embeddings can be converted to equivalent holographic embeddings.

Many systems for natural language processing, such as those for semantic parsing and question answering, benefit from access to information stored in knowledge graphs. We plan to further investigate the property of spectral HoIE and ComplEx in these applications.

## Acknowledgments

We thank the anonymous reviewers for helpful comments. This work was partially supported by JSPS Kakenhi Grants 26730126 and 15H02749.

## References

- Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A nucleus for a web of open data. In *The Semantic Web: Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference (ISWC '07/ASWC '07)*. Springer, Lecture Notes in Computer Science 4825, pages 722–735.

- Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD '08)*. pages 1247–1250.
- Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI '11)*. pages 301–306.
- Lise Getoor and Ben Taskar. 2007. *Introduction to Statistical Relational Learning*. Adaptive Computation and Machine Learning. The MIT Press.
- Kelvin Guu, John Miller, and Percy Liang. 2015. Traversing knowledge graphs in vector space. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP '15)*. pages 318–327.
- Mohammad Al Hasan and Mohammed J. Zaki. 2011. A survey of link prediction in social networks. In Charu C. Aggarwal, editor, *Social Network Data Analytics*, Springer, chapter 9, pages 243–275.
- David Liben-Nowell and Jon Kleinberg. 2003. The link prediction problem for social networks. In *Proceedings of the 12nd Annual ACM International Conference on Information and Knowledge Management (CIKM '03)*. pages 556–559.
- Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2015. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE* pages 1–18.
- Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI '16)*. pages 1955–1961.
- III Julius O. Smith. 2007. *Mathematics of the Discrete Fourier Transform (DFT): with Audio Applications*. W3K Publishing, 2nd edition.
- Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems 26 (NIPS '13)*. pages 926–934.
- Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: A core of semantic knowledge unifying Wordnet and Wikipedia. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. pages 697–706.
- Théo Trouillon, Christopher R. Dance, Éric Gaussier, and Guillaume Bouchard. 2016a. Decomposing real square matrices via unitary diagonalization. arXiv.math eprint 1605.07103, arXiv.
- Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016b. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on Machine Learning (ICML '16)*. pages 2071–2080.
- Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR '15)*.