Complex-valued restricted Boltzmann machine for direct learning of frequency spectra

Citation for published version:
Nakashika, T, Takaki, S & Yamagishi, J 2017, Complex-valued restricted Boltzmann machine for direct learning of frequency spectra. in Proceedings Interspeech 2017. pp. 4021-4025, Interspeech 2017, Stockholm, Sweden, 20-24 August. DOI: 10.21437/Interspeech.2017-584

Digital Object Identifier (DOI):
10.21437/Interspeech.2017-584

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings Interspeech 2017

Complex-valued restricted Boltzmann machine for direct learning of frequency spectra

Toru Nakashika[1], *Shinji Takaki*[2], *Junichi Yamagishi*[2,3]

[1]University of Electro-Communications, Tokyo, Japan
[2]National Institute of Informatics, Tokyo, Japan
[3]University of Edinburgh, Edinburgh, UK

nakashika@uec.ac.jp,takaki@nii.ac.jp,jyamagis@nii.ac.jp

Abstract

In this paper, we propose a new energy-based probabilistic model where a restricted Boltzmann machine (RBM) is extended to deal with complex-valued visible units. The RBM that automatically learns the relationships between visible units and hidden units (but without connections in the visible or the hidden units) has been widely used as a feature extractor, a generator, a classifier, pre-training of deep neural networks, etc. However, all the conventional RBMs have assumed the visible units to be either binary-valued or real-valued, and therefore complex-valued data cannot be fed to the RBM.

In various applications, however, complex-valued data is frequently used such examples include complex spectra of speech, fMRI images, wireless signals, and acoustic intensity. For the direct learning of such the complex-valued data, we define the new model called "complex-valued RBM (CRBM)" where the conditional probability of the complex-valued visible units given the hidden units forms a complex-Gaussian distribution. Another important characteristic of the CRBM is to have connections between real and imaginary parts of each of the visible units unlike the conventional real-valued RBM. Our experiments demonstrated that the proposed CRBM can directly encode complex spectra of speech signals without decoupling imaginary number or phase from the complex-value data.

Index Terms: restricted Boltzmann machine, deep learning, complex-valued data, speech encoding

1. Introduction

Deep learning is one of the recent hottest topics in wide research fields such as artificial intelligence, machine learning, and signal processing including image classification, speech recognition, etc[1]. Many models have been proposed so far as a tool of deep learning; one of the most widely-used and famous models is a deep belief-net (DBN) [2] that stacks multiple restricted Boltzmann machines (RBMs) layer-by-layer. The RBM, which is a probabilistic model that consists of visible and hidden units, has often been used alone as a feature extractor, a generator, and a classifier as well as a pre-training scheme of deep neural networks, and many extensions of the RBM have been also proposed [3, 4, 5, 6]. Although the RBM has been used in so many tasks, the RBM traditionally assumed visible units to be either binary-valued or real-valued [2, 7, 8].

Concerning speech signal processing, representations based on amplitude spectra of speech such as MFCC and mel-cepstra are normally used as input features of speech recognition or output features of speech synthesis, because it is known that the amplitude spectra are more effective and relevant to our auditory field than phase spectra in such tasks. However, these features theoretically lack phase information. More specifically, the use of the amplitude-based features only cannot represent the original complex values correctly. In other signal processing as well, there are many cases where we have to deal with complex-valued actual data such as fMRI images, wireless signals, acoustic intensity, etc. Other machine learning models, that is neural networks, Boltzmann machines, and non-negative matrix factorization (NMF) [9], have their extensions proposed to represent complex-valued data [10, 11, 12].

In this paper, we newly propose an extension of the RBM that deals with complex-valued data, and evaluate its effectiveness through experiments using artificial data and speech spectra. The proposed model called "complex-valued RBM (CRBM)" consists of complex-normal visible units and Bernoulli hidden units.

The CRBM has important characteristics of having no connections across dimensions in the same layers but having connections between real and imaginary parts of each of the visible units unlike the conventional RBM. Therefore, it is easy to estimate the parameters using Gibbs sampling or contrastive divergence [2], and it is expected to aggregate the information propagated from the complex-valued visible units into the hidden units. Such characteristics can not be seen in an extension of Boltzmann machine (directional-unit Boltzmann machine (DUBM) [11]) that feeds complex-valued data having connections across dimensions, which makes the parameter estimation difficult. Another difference between the proposed CRBM and the DUBM is the form of complex-valued visible units; the visible units in the CRBM are in rectangular form having real and imaginary components, while those in the DUBM are in polar form having phase components with amplitude components. Since the conditional probability of visible units given hidden units in the CRBM form a complex-normal distribution, which makes the real and imaginary components Gaussian-distributed, respectively, we can generate samples from the distribution straightfowardly.

This paper is organized as follows: In Section 2 we overview the conventional real-valued RBM. In Section 3 we define the proposed Complex-valued RBM and show its parameter estimation algorithm. In Section 4 we show our experimental results and conclude our findings in Section 5.

2. Preliminary

A restricted Boltzmann machine (RBM [2, 13]), one of the most-widely used energy-based models, is convenient for representing latent features that are cannot be observed but surely exist in the background. An RBM was originally introduced as an undirected graphical model that defines the distribution

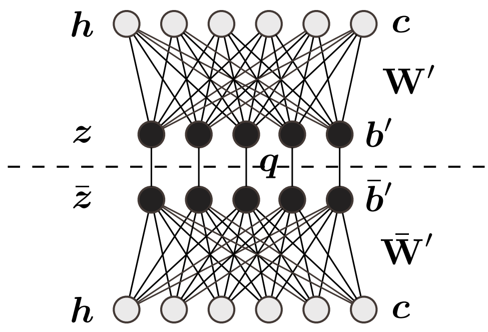

Figure 1: *Graphical representation of a complex-valued RBM.*

of binary visible variables with binary hidden (latent) variables, and was later extended to deal with real valued-data known as a Gaussian-Bernoulli RBM (GB-RBM) [2]. It has been, however, reported that the original GB-RBM had some difficulties because the training of the parameters was unstable. Later, an improved learning method for a GB-RBM has been proposed by Cho *et al.* [13] to overcome the difficulties[1]. In the modeling using an RBM, the joint probability $p(\boldsymbol{v}, \boldsymbol{h})$ of real-valued visible units $\boldsymbol{v} \in \mathbb{R}^I$ and binary-valued hidden units $\boldsymbol{h} \in \{0,1\}^J$ (I and J indicate the numbers of dimensions in the visible and hidden units, respectively) is defined as follows:

$$p(\boldsymbol{v};\boldsymbol{\theta}) = \sum_{\boldsymbol{h}} p(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\theta}) \tag{1}$$

$$p(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\theta}) = \frac{1}{U(\boldsymbol{\theta})} e^{-E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\theta})} \tag{2}$$

$$E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{v}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{v} - \boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{v} - \boldsymbol{c}^\top \boldsymbol{h} - \boldsymbol{v}^\top \boldsymbol{\Sigma}^{-1}\mathbf{W}\boldsymbol{h} \tag{3}$$

$$U(\boldsymbol{\theta}) = \int \sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\theta})} d\boldsymbol{v} \tag{4}$$

where $\boldsymbol{\theta} = \{\boldsymbol{b}, \boldsymbol{c}, \mathbf{W}, \boldsymbol{\sigma}\}$ indicates a set of parameters, which contains bias parameters of the visible units $\boldsymbol{b} \in \mathbb{R}^I$, bias parameters of the hidden units $\boldsymbol{c} \in \mathbb{R}^J$, the connection weight parameters between visible-hidden units $\mathbf{W} \in \mathbb{R}^{I\times J}$, and the standard deviation parameters associated with the dimension independent Gaussian visible units $\boldsymbol{\sigma} \in \mathbb{R}^I$ that defines $\boldsymbol{\Sigma} \triangleq \Delta(\boldsymbol{\sigma}^2)$ (the function $\Delta(\cdot)$ returns a diagonal matrix whose diagonal vector is the argument). From the above definition, the conditional probabilities $p(\boldsymbol{v}|\boldsymbol{h})$ and $p(\boldsymbol{h}|\boldsymbol{v})$ form simple distributions as:

$$p(\boldsymbol{v}|\boldsymbol{h}) = \mathcal{N}(\boldsymbol{v}; \boldsymbol{b} + \mathbf{W}\boldsymbol{h}, \boldsymbol{\Sigma}) \tag{5}$$

$$p(\boldsymbol{h}|\boldsymbol{v}) = \mathcal{B}(\boldsymbol{h}; \boldsymbol{f}(\boldsymbol{c} + \mathbf{W}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{v})) \tag{6}$$

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathcal{B}(\cdot; \boldsymbol{\pi})$, and $\boldsymbol{f}(\cdot)$ indicate the multivariate Gaussian distribution with the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, the multi-dimensional Bernoulli distribution with the success probabilities $\boldsymbol{\pi}$, and an element-wise sigmoid function, respectively.

[1] In the remaining of this paper, we refer to the improved GB-RBM just as an RBM.

3. Complex-valued RBM

3.1. Definition

In this section, we will define an extension of the RBM (complex-valued RBM; CRBM) that can feed complex-valued data but has a real-valued probability distribution (cost function used in parameter estimation) like the conventional RBM. In the CRBM, we will give a "restriction" where there are no connections between visible units or hidden units, which enables easy estimation of parameters just as an RBM does; however, in order to capture the relationships between the real and imaginary parts of each complex-valued visible unit, we will allow the model to have connections between the real and imaginary parts.

Based on the above extension, we formulate the CRBM that has I-dimensional complex-valued visible units $\boldsymbol{z} \in \mathbb{C}^I$ and J-dimensional binary-valued hidden units $\boldsymbol{h} \in \{0,1\}^J$ as follows in this paper:

$$p(\boldsymbol{z};\boldsymbol{\theta}) = \sum_{\boldsymbol{h}} p(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta}) \tag{7}$$

$$p(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta}) = \frac{1}{U(\boldsymbol{\theta})} e^{-E(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta})} \tag{8}$$

$$\begin{aligned} E(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta}) =& \frac{1}{2}\begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix}^H \boldsymbol{\Phi}^{-1} \begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix} \\ &- \begin{bmatrix} \boldsymbol{b} \\ \bar{\boldsymbol{b}} \end{bmatrix}^H \boldsymbol{\Phi}^{-1} \begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix} - 2\boldsymbol{c}^\top \boldsymbol{h} \\ &- \begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix}^H \boldsymbol{\Phi}^{-1} \begin{bmatrix} \mathbf{W} \\ \bar{\mathbf{W}} \end{bmatrix} \boldsymbol{h} \end{aligned} \tag{9}$$

$$U(\boldsymbol{\theta}) = \int \sum_{\boldsymbol{h}} e^{-E(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta})} d\boldsymbol{z}, \tag{10}$$

where $\bar{\cdot}$ and $\cdot^H$ denote complex-conjugate and Hermitian-transpose, respectively. $\boldsymbol{b} \in \mathbb{C}^I$, $\boldsymbol{c} \in \mathbb{R}^J$, and $\mathbf{W} \in \mathbb{C}^{I\times J}$ are bias parameters of the visible units and the hidden units, and the *biased* connection weights between visible and hidden units, respectively. In order to make the restrictions, the extended covariance matrix $\boldsymbol{\Phi}$ consists of a covariance matrix $\boldsymbol{\Gamma}$ and a pseudo-covariance matrix $\mathbf{C}$, both of which are diagonal matrices, as

$$\boldsymbol{\Phi} \triangleq \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{C} \\ \mathbf{C}^H & \boldsymbol{\Gamma}^H \end{bmatrix} \tag{11}$$

and

$$\begin{aligned} \boldsymbol{\Gamma} &\triangleq \Delta(\gamma), \quad \gamma \in \mathbb{R}+^I \\ \mathbf{C} &\triangleq \Delta(\boldsymbol{\delta}), \quad \boldsymbol{\delta} \in \mathbb{C}^I \end{aligned} \tag{12}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are variance and pseudo-variance parameters of the complex-valued visible units, respectively. To summarize, the set of parameters of the CRBM is $\boldsymbol{\theta} = \{\boldsymbol{b}, \boldsymbol{c}, \mathbf{W}, \gamma, \boldsymbol{\delta}\}$.

Introducing two vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ defined as

$$\boldsymbol{p} \triangleq \frac{\boldsymbol{\gamma}}{\boldsymbol{\gamma}^2 - |\boldsymbol{\delta}|^2} \in \mathbb{R}^I \tag{13}$$

$$\boldsymbol{q} \triangleq -\frac{\boldsymbol{\delta}}{\boldsymbol{\gamma}^2 - |\boldsymbol{\delta}|^2} \in \mathbb{C}^I \tag{14}$$

where the fraction bar denotes element-wise division, we can rewrite the energy function in Eq. (9) as follows:

$$\begin{aligned}&E(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta})=\\&\boldsymbol{z}^H\Delta(\boldsymbol{p})\boldsymbol{z}+\Re(\boldsymbol{z}^H\Delta(\boldsymbol{q})\bar{\boldsymbol{z}})-2\Re(\boldsymbol{z}^H\Delta(\boldsymbol{p})\boldsymbol{b})\\&-2\Re(\boldsymbol{z}^H\Delta(\boldsymbol{q})\bar{\boldsymbol{b}})-2\boldsymbol{c}^\top\boldsymbol{h}-2\Re(\boldsymbol{z}^H\Delta(\boldsymbol{p})\mathbf{W})\boldsymbol{h}\\&-2\Re(\boldsymbol{z}^H\Delta(\boldsymbol{q})\bar{\mathbf{W}})\boldsymbol{h},\end{aligned}\tag{15}$$

which confirms that 1) the above energy function E and the probability distribution are real-valued, and that 2) there are connections between the complex-valued visible units and their conjugates for each dimension but no connections between different dimensions.

Furthermore, when we use unbiased parameters:

$$\boldsymbol{b}'\triangleq\Delta(\boldsymbol{p})\boldsymbol{b}+\Delta(\boldsymbol{q})\bar{\boldsymbol{b}}\tag{16}$$

$$\mathbf{W}'\triangleq\Delta(\boldsymbol{p})\mathbf{W}+\Delta(\boldsymbol{q})\bar{\mathbf{W}},\tag{17}$$

the energy function E becomes

$$\begin{aligned}&E(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta})=\\&\frac{1}{2}\boldsymbol{z}^H\Delta(\boldsymbol{p})\boldsymbol{z}+\frac{1}{2}\bar{\boldsymbol{z}}^H\Delta(\boldsymbol{p})\bar{\boldsymbol{z}}+\boldsymbol{z}^H\Delta(\boldsymbol{q})\bar{\boldsymbol{z}}\\&+\bar{\boldsymbol{z}}^H\Delta(\bar{\boldsymbol{q}})\boldsymbol{z}-\boldsymbol{z}^H\boldsymbol{b}'-\bar{\boldsymbol{z}}^H\bar{\boldsymbol{b}}'-2\boldsymbol{c}^\top\boldsymbol{h}\\&-\boldsymbol{z}^H\mathbf{W}'\boldsymbol{h}-\bar{\boldsymbol{z}}^H\bar{\mathbf{W}}'\boldsymbol{h},\end{aligned}\tag{18}$$

which indicates that $\boldsymbol{z}$ and $\bar{\boldsymbol{z}}$ are symmetric to each other as shown in Figure 1.

From the above definition, the conditional probabilities $p(\boldsymbol{z}|\boldsymbol{h})$ and $p(\boldsymbol{h}|\boldsymbol{z})$ can be derived as follows:

$$p(\boldsymbol{z}|\boldsymbol{h})=\mathcal{CN}(\boldsymbol{z};\boldsymbol{b}+\mathbf{W}\boldsymbol{h},\boldsymbol{\Gamma},\mathbf{C})\tag{19}$$

$$p(\boldsymbol{h}|\boldsymbol{z})=\mathcal{B}(\boldsymbol{h};\boldsymbol{f}(2\boldsymbol{c}+2\Re(\mathbf{W}'^H\boldsymbol{z})))\tag{20}$$

where $\mathcal{CN}(\cdot;\boldsymbol{\mu},\boldsymbol{\Gamma},\mathbf{C})$ is a multivariate complex normal distribution a mean vector $\boldsymbol{\mu}$, a covariance matrix $\boldsymbol{\Gamma}$, and a pseudo-covariance matrix $\mathbf{C}$:

$$\begin{aligned}p(\boldsymbol{z})=&\frac{1}{\pi^D\sqrt{\det(\boldsymbol{\Gamma})\det(\mathbf{Q})}}\\&\cdot\exp\left\{-\frac{1}{2}\begin{bmatrix}\boldsymbol{z}-\boldsymbol{\mu}\\\bar{\boldsymbol{z}}-\bar{\boldsymbol{\mu}}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{\Gamma}&\mathbf{C}\\\mathbf{C}^H&\boldsymbol{\Gamma}^H\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{z}-\boldsymbol{\mu}\\\bar{\boldsymbol{z}}-\bar{\boldsymbol{\mu}}\end{bmatrix}\right\}\end{aligned}\tag{21}$$

$$\mathbf{Q}=\bar{\boldsymbol{\Gamma}}-\mathbf{C}^H\boldsymbol{\Gamma}^{-1}\mathbf{C}.\tag{22}$$

3.2. Parameter estimation

In this paper, we estimate the parameters of the CRBM $\boldsymbol{\theta}$ using complex-valued gradient ascend so as to maximize the log-likelihood of the complex-valued training data $\boldsymbol{z}$:

$$\begin{aligned}L(\boldsymbol{\theta})&=\log p(\boldsymbol{z};\boldsymbol{\theta})&(23)\\&=\log\sum_{\boldsymbol{h}}p(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta})&(24)\\&=\log\sum_{\boldsymbol{h}}e^{-E(\boldsymbol{z},\boldsymbol{h};\boldsymbol{\theta})}-\log\int\sum_{\tilde{\boldsymbol{h}}}e^{-E(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{h}};\boldsymbol{\theta})}d\tilde{\boldsymbol{z}}.&(25)\end{aligned}$$

The complex-valued gradient ascend iteratively updates each parameter with a learning rate $\alpha>0$ as:

$$\boldsymbol{\theta}^{(\mathrm{new})}\leftarrow\boldsymbol{\theta}^{(\mathrm{old})}+\alpha\cdot2\frac{\partial L}{\partial\bar{\boldsymbol{\theta}}},\tag{26}$$

where the partial gradients in Eq. (26) are the Wirtinger derivative:

$$\frac{\partial L}{\partial\boldsymbol{\theta}}=\frac{1}{2}\left(\frac{\partial L}{\partial\Re(\boldsymbol{\theta})}-i\frac{\partial L}{\partial\Im(\boldsymbol{\theta})}\right).\tag{27}$$

The partial gradients of each parameter include the expectations of the partial gradients to the energy function along the training data and the model, which can be approximated and easily calculated with contrastive divergence (CD) [2] in a similar way to the conventional RBM because of the restriction of the CRBM.

The partial gradients of each parameter to the energy function can be derived as:

$$-\frac{\partial E}{\partial\boldsymbol{b}}=\Delta(\boldsymbol{p})\bar{\boldsymbol{z}}+\Delta(\bar{\boldsymbol{q}})\boldsymbol{z}\tag{28}$$

$$-\frac{\partial E}{\partial\boldsymbol{c}}=\boldsymbol{h}\tag{29}$$

$$-\frac{\partial E}{\partial\mathbf{W}}=(\Delta(\boldsymbol{p})\bar{\boldsymbol{z}}+\Delta(\bar{\boldsymbol{q}})\boldsymbol{z})\boldsymbol{h}^\top\tag{30}$$

$$-\frac{\partial E}{\partial\boldsymbol{\gamma}}=(\boldsymbol{p}^2+|\boldsymbol{q}|^2)\circ\frac{\partial E}{\partial\boldsymbol{p}}+2\Re(\boldsymbol{p}\circ\boldsymbol{q}\circ\frac{\partial E}{\partial\boldsymbol{q}})\tag{31}$$

$$-\frac{\partial E}{\partial\boldsymbol{\delta}}=\boldsymbol{p}^2\circ\frac{\partial E}{\partial\boldsymbol{q}}+\bar{\boldsymbol{q}}^2\circ\frac{\partial E}{\partial\bar{\boldsymbol{q}}}+2\boldsymbol{p}\circ\bar{\boldsymbol{q}}\circ\frac{\partial E}{\partial\boldsymbol{p}},\tag{32}$$

where $\circ$ and $|\cdot|$ denote element-wise product and absolute, respectively, and

$$\frac{\partial E}{\partial\boldsymbol{p}}=\frac{1}{2}|\boldsymbol{z}|^2-\Re(\boldsymbol{z}\circ(\bar{\boldsymbol{b}}+\bar{\mathbf{W}}\boldsymbol{h}))\tag{33}$$

$$\frac{\partial E}{\partial\boldsymbol{q}}=\frac{1}{2}\bar{\boldsymbol{z}}^2-\bar{\boldsymbol{z}}\circ(\bar{\boldsymbol{b}}+\bar{\mathbf{W}}\boldsymbol{h}).\tag{34}$$

The gradients of variance and pseudo variance tend to be larger than those of the other parameters. For stable training, we replace the parameters as $\boldsymbol{\gamma}\triangleq e^{\boldsymbol{r}}$ and $\boldsymbol{\delta}\triangleq e^{\boldsymbol{s}}$, and update using the gradients of $\boldsymbol{r}$ and $\boldsymbol{s}$, respectively.

4. Experiments

4.1. Evaluation using artificial data

In order to evaluate the effectiveness of the proposed CRBM, we first conducted an experiment using one-dimensional complex-valued artificial data ($N=2000$). The artificially created data is illustrated in Figure 2 as black dots, which has correlations between the real and imaginary parts. In this experiment, we compared the CRBM with a GB-RBM having two visible units; one is for the real part, another is for the imaginary part. We trained both models with two hidden units using stochastic gradient ascend (SGD) with a learning rate of 0.01, a momentum of 0.1, a batch size of 20, and a number of epochs as 200. After the training, we randomly generated samples from the models; the samples from the CRBM and the GB-RBM are shown as red dots on the above and on the below of Figure 2, respectively. As shown in Figure 2, we can see that the proposed CRBM could represent the distribution of the complex-valued artificial data more accurately than the GB-RBM. This is because the CRBM can capture the relationships between the real and imaginary parts while GB-RBM does not capture the correlations between them.

4.2. Evaluation using speech data

Second, we conducted an encoding-and-decoding experiment using speech data from the Repeated Harvard Sentence Prompts

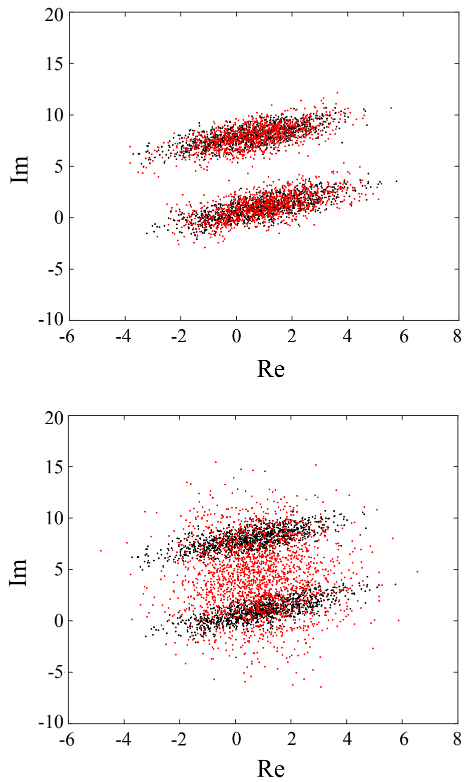

Figure 2: *Artificially created 1D complex-valued data (black dots) and random samples (red dots) generated from the trained models: the proposed CRBM (above) and the conventional GB-RBM (below).*

(REHASP) corpus[2]. From the corpus, we randomly selected 30 repeats of 30 sentences, processed the short-time Fourier transform (STFT) with a windows length of 256 with overlapping 64 samples, and trained the CRBM with 200 hidden units (i.e., $I = 129$ and $J = 200$). For evaluation, we first estimated hidden units from the test data that was different from the training data (encoding), and then reconstructed the visible units from the hidden units (decoding). The performance of the encoding-and-decoding was evaluated using an objective criteria of mean-squared error (MSE) between the original data and the reconstructed one. In this experiment, we trained the model with a learning rate of 0.01, a momentum of 0.1, a batch size of 100, and a number of epochs as 500. For comparison, we also trained a GB-RBM with doubled visible units, that is a super vector having the real and imaginary parts of the complex-valued spectra, with the same configuration.

Figure 3 shows an example of the reconstruction using the CRBM. As shown in Figure 3, the reconstructed spectra was fairly closed to the original spectra. Figure 4 shows MSE calculated during the training, comparing the proposed CRBM with the GB-RBM. We notice from Figure 4 that the CRBM converged more quickly than the GB-RBM, and the MSE in convergence (at around 100 epochs) of the CRBM is much smaller than that of the GB-RBM. For the test data, we also obtained the MSE of 51.9 from the CRBM, which outperformed the MSE of 54.1 from the GB-RBM.

[2] http://datashare.is.ed.ac.uk/handle/10283/561

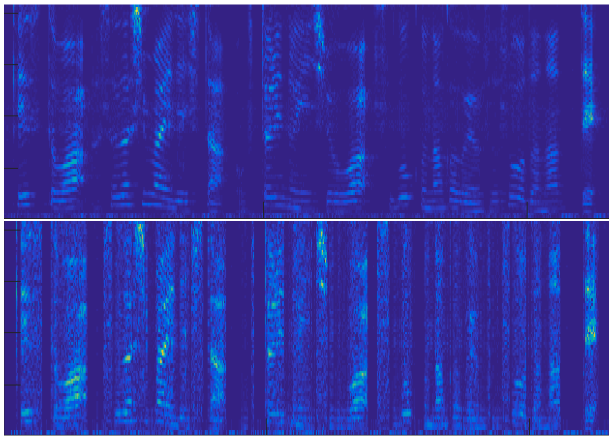

Figure 3: *Amplitude spectra of the original speech (above) and of the reconstructed one from the CRBM (below).*

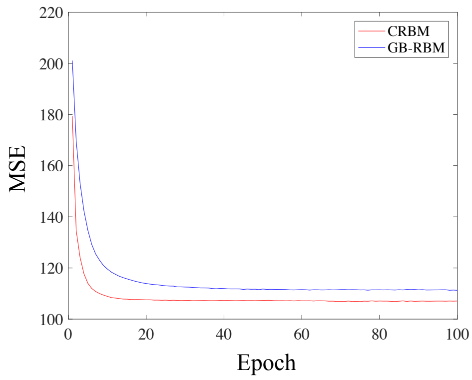

Figure 4: *MSE during the training of the CRBM (red line) and the GB-RBM (blue line).*

5. Conclusion

In this paper, we proposed an extension of restricted Boltzmann machine (RBM) that can feed complex-valued visible units, called complex-valued RBM (CRBM). We formulated the model and showed that the conditional probability of visible units given hidden units formed a complex-valued normal distribution. To evaluated the performance of the proposed CRBM, we conducted experiments using artificial complex-valued data and complex spectra of speech. Through the experiments, we conclude that the model would be effective to represent complex-valued data especially when the real and imaginary parts are correlated with each other. In the future, we would like to fully investigate the performance of the CRBM when using speech data in terms of subjective criteria, and to apply the model to represent other complex-valued actual data in the real world. Future work also includes extensions of the CRBM; e.g., the model stacking multiple hidden layers layer-by-layer like deep Boltzmann machine (DBM) [3], and the model having complex-normal hidden units instead of Bernoulli hidden units so as to extract complex-valued latent features.

Acknowledgements: This work was partially supported by ACT-I from the Japan Science and Technology Agency (JST), by MEXT KAKENHI Grant Numbers (26280066, 15H01686, 16K16096, 16H06302), and by The Telecommunications Advancement Foundation Grants.

6. References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[3] R. Salakhutdinov and G. E. Hinton, "Deep Boltzmann machines," *AISTATS*, pp. 448–455, 2009.

[4] A. Krizhevsky and G. E. Hinton, "Factored 3-way restricted boltzmann machines for modeling natural images," *Journal of Machine Learning Research*, 2010.

[5] K. Sohn, G. Zhou, C. Lee, and H. Lee, "Learning and Selecting Features Jointly with Point-wise Gated Boltzmann Machines," *ICML (2)*, 2013.

[6] T. Nakashika, T. Takiguchi, and Y. Minami, "Non-Parallel Training in Voice Conversion Using an Adaptive Restricted Boltzmann Machine," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2032–2045, 2016.

[7] Y. Freund and D. Haussler, "Unsupervised learning of distributions of binary vectors using two layer networks," *Computer Research Laboratory*, pp. 912–919, 1994.

[8] H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area V2," in *Advances in neural information processing systems*, 2008, pp. 873–880.

[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[10] I. Nemoto and T. Kono, "Complex neural networks," *Systems and computers in Japan*, vol. 23, no. 8, pp. 75–84, 1992.

[11] R. S. Zemel, C. Williams, and M. C. Mozer, "Lending direction to neural networks," *Neural Networks*, vol. 8, no. 4, pp. 503–512, 1995.

[12] H. Kameoka, N. Ono, and K. Kashino, "Complex NMF: A new sparse representation for acoustic signals," *ICASSP*, pp. 3437–3440, 2009.

[13] K. Cho, A. Ilin, and T. Raiko, "Improved learning of Gaussian-Bernoulli restricted Boltzmann machines," in *Proc. ICANN*. Springer, 2011, pp. 10–17.