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, [2]National Institute of Informatics, [3]University of Edinburgh

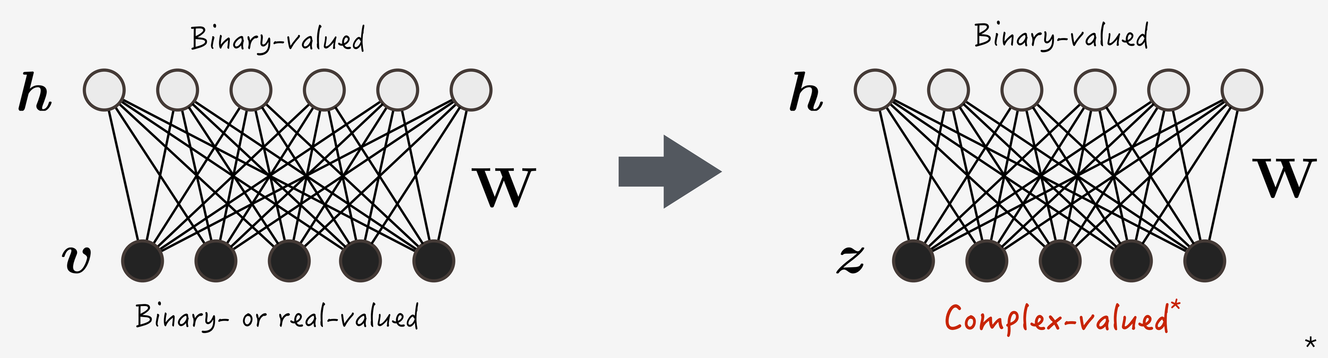
1. Introduction

Background

- The **RBM** (restricted Boltzmann machine), which is a probabilistic model that consists of visible and hidden units, has **often been used** in the pre-training scheme of deep neural networks, and as a feature extractor, a generator, etc.
- Although the RBM has been used in so many tasks, the conventional RBM **assumed visible units to be either binary-valued or real-valued**.
- In signal processing, **there are many cases where we have to deal with complex-valued actual data** such as complex spectra of speech, fMRI images, wireless signals, acoustic intensity, etc.

What we want to do

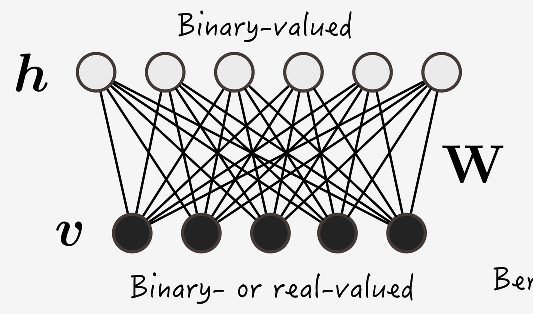
is **to define an extension of the RBM that deals with complex-valued data**, and evaluate its effectiveness through experiments using artificial data and speech spectra.



* Note that our interested complex is in rectangular form, not in polar form.

2. Conventional models

BB-RBM and GB-RBM



BB-RBM

Definition

$$p(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{N}e^{-E(\boldsymbol{v},\boldsymbol{h})}$$

$$E(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{v}^\top \mathbf{W}\boldsymbol{h}$$

GB-RBM

Definition

$$p(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{N}e^{-E(\boldsymbol{v},\boldsymbol{h})}$$

$$E(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{2}\boldsymbol{v}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{v} - \boldsymbol{v}^\top \boldsymbol{\Sigma}^{-1}\mathbf{W}\boldsymbol{h}$$

Bernoulli distribution

$$p(\boldsymbol{v}|\boldsymbol{h}) = \mathcal{B}(\boldsymbol{v};\boldsymbol{\sigma}(\mathbf{W}\boldsymbol{h}))$$

Sigmoid function

Gaussian distribution

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

An RBM was originally introduced as an undirected graphical model that defines the distribution of binary visible variables $\boldsymbol{v}$ with binary hidden (latent) variables $\boldsymbol{h}$ (often referred to as BB-RBM). The conditional probability $p(\boldsymbol{v}|\boldsymbol{h})$ forms Bernoulli distribution according to the definition. Therefore, it feeds **binary-valued**.

The RBM was later extended to deal with real valued-data known as a Gaussian-Bernoulli RBM (GB-RBM). The conditional probability $p(\boldsymbol{v}|\boldsymbol{h})$ turns to be Gaussian-distributed, which feeds **real-valued**.

4. Parameter estimation

In conventional RBMs, the parameters are estimated using gradient ascend. In CRBM, we estimate the parameters using **complex-valued gradient ascend** so as to maximize the log-likelihood $L = \log p(\boldsymbol{z})$ as:

$$\boldsymbol{\theta}^{(\text{new})} \leftarrow \boldsymbol{\theta}^{(\text{old})} + \alpha \cdot 2\frac{\partial L}{\partial \bar{\boldsymbol{\theta}}}, \quad \alpha \in \mathbb{C}$$

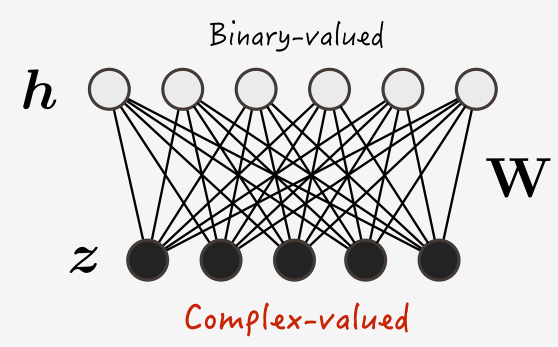
where

$$\frac{\partial L}{\partial \boldsymbol{\theta}} \triangleq \frac{1}{2}\left(\frac{\partial L}{\partial \Re(\boldsymbol{\theta})} - i\frac{\partial L}{\partial \Im(\boldsymbol{\theta})}\right)$$

(Wirtinger derivative)

3. Proposed model

Complex-valued RBM



Definition

$$p(\boldsymbol{z},\boldsymbol{h}) = \frac{1}{N}e^{-E(\boldsymbol{z},\boldsymbol{h})} \qquad \boldsymbol{\Phi} \triangleq \begin{bmatrix} \Delta(\boldsymbol{\gamma}) & \Delta(\boldsymbol{\delta}) \\ \Delta(\bar{\boldsymbol{\delta}}) & \Delta(\bar{\boldsymbol{\gamma}}) \end{bmatrix}$$

$$E(\boldsymbol{z},\boldsymbol{h}) = \frac{1}{2}\begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix}^H \boldsymbol{\Phi}^{-1}\begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{z} \\ \bar{\boldsymbol{z}} \end{bmatrix}^H \boldsymbol{\Phi}^{-1}\begin{bmatrix} \mathbf{W} \\ \bar{\mathbf{W}} \end{bmatrix}\boldsymbol{h}$$

$$p(\boldsymbol{z}|\boldsymbol{h}) = \mathcal{CN}(\boldsymbol{z};\mathbf{W}\boldsymbol{h},\Delta(\boldsymbol{\gamma}),\Delta(\boldsymbol{\delta}))$$

Complex Gaussian distribution

$$p(\boldsymbol{z}) = \mathcal{CN}(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Gamma},\mathbf{C})$$

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

where

Mean $\boldsymbol{\mu} \in \mathbb{C}^D \quad = \mathbb{E}[\boldsymbol{z}]$

Covariance $\boldsymbol{\Gamma} \in \mathbb{C}^{D\times D} \quad = \mathbb{E}[(\boldsymbol{z}-\boldsymbol{\mu})(\boldsymbol{z}-\boldsymbol{\mu})^H]$

Pseudo-covariance $\mathbf{C} \in \mathbb{C}^{D\times D} \quad = \mathbb{E}[(\boldsymbol{z}-\boldsymbol{\mu})(\bar{\boldsymbol{z}}-\bar{\boldsymbol{\mu}})^H]$

We define the extension of RBM, namely CRBM, so that it satisfies:

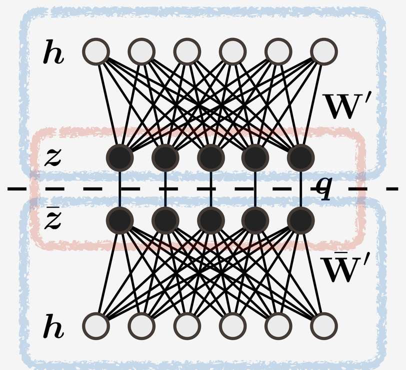
1. The conditional probability $p(\boldsymbol{z}|\boldsymbol{h})$ **forms complex Gaussian distribution**
2. The conditional probability $p(\boldsymbol{h}|\boldsymbol{z})$ forms Bernoulli distribution
3. No connections across dimensions like the standard RBM
4. **Having connections between real and imaginary** of each dimension

$\Leftrightarrow$ Covariance $\boldsymbol{\Gamma}$ and pseudo-covariance $\mathbf{C}$ are diagonal

$$\boldsymbol{\Gamma} = \Delta(\boldsymbol{\sigma}) \qquad \boldsymbol{\sigma} \in \mathbb{R}^D$$

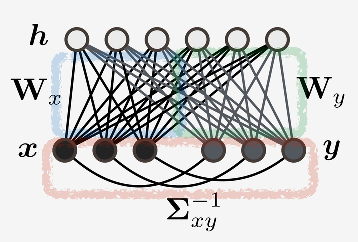
$$\mathbf{C} = \Delta(\boldsymbol{\delta}) \qquad \boldsymbol{\delta} \in \mathbb{C}^D$$

Another perceptive



The model that has connections with conjugate of visible units

$$E(\boldsymbol{z},\boldsymbol{h}) = \frac{1}{2}\boldsymbol{z}^H\Delta(\boldsymbol{p})\boldsymbol{z} + \frac{1}{2}\bar{\boldsymbol{z}}^H\Delta(\boldsymbol{p})\bar{\boldsymbol{z}} - \boldsymbol{z}^H\Delta(\boldsymbol{q})\bar{\boldsymbol{z}} - \bar{\boldsymbol{z}}^H\Delta(\bar{\boldsymbol{q}})\boldsymbol{z} - \boldsymbol{z}^H\mathbf{W}'\boldsymbol{h} - \bar{\boldsymbol{z}}^H\bar{\mathbf{W}}'\boldsymbol{h}$$



$$\tilde{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \in \mathbb{R}^{2D}$$

Ext. of RBM having connections between real and imaginary

$$E(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}) = \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{\Sigma}_x^{-1}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}^\top\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y} - \boldsymbol{x}^\top\boldsymbol{\Sigma}_{xy}^{-1}\boldsymbol{y} - \boldsymbol{x}^\top\boldsymbol{\Sigma}_x^{-1}\mathbf{W}_x\boldsymbol{h} - \boldsymbol{y}^\top\boldsymbol{\Sigma}_y^{-1}\mathbf{W}_y\boldsymbol{h}$$

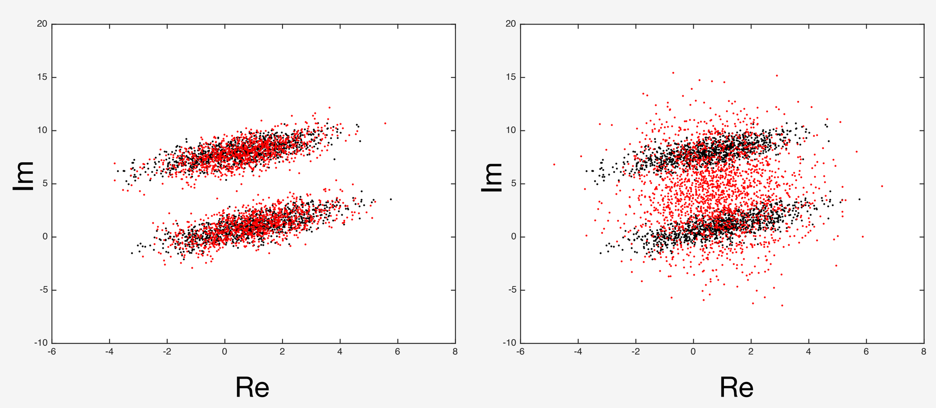
5. Experiments

Evaluation using artificial data

We first conducted an experiment using one-dimensional complex-valued artificial data (*N = 2000*), illustrated in the figure below (as black dots).

The red dots shows random examples generated from the trained CRBM and GB-RBM (*H=2*).

The **CRBM approximates the data distribution** more than the GB-RBM, because the **CRBM can capture the relationships between the real and imaginary parts**.



CRBM GB-RBM

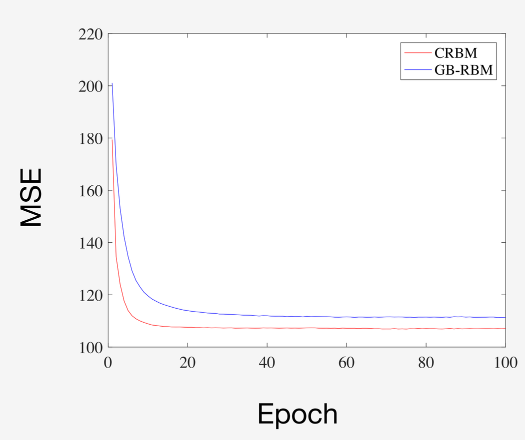
Evaluation using speech data

Second, we conducted an encoding-and-decoding experiment using speech data from the Repeated Harvard Sentence Prompts (REHASP) corpus (30 repeats of 30 sentences).

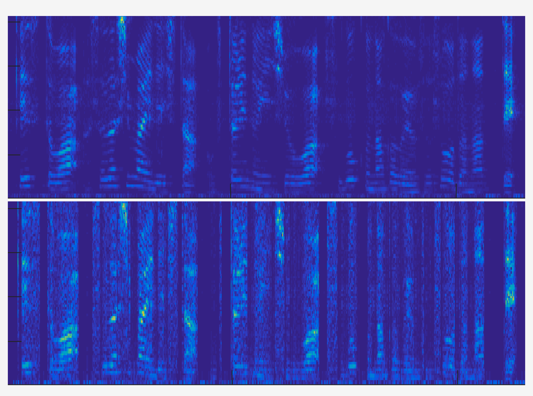
MSE curve: the CRBM converged more quickly than the GB-RBM, and the MSE in convergence of the CRBM is much smaller than that of the GB-RBM.

Reconstruction: the reconstructed spectra (below) was fairly closed to the original spectra (above).

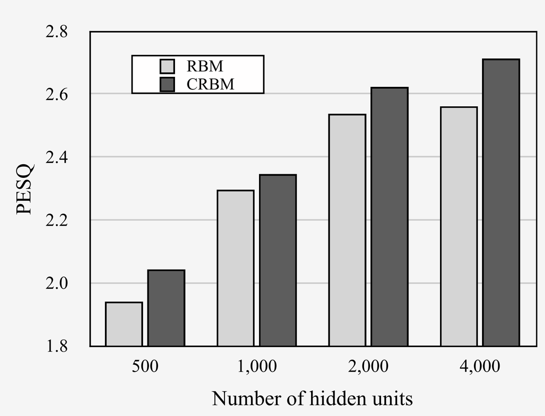
PESQ: the CRBM outperformed the GB-RBM in terms speech quality (PESQ).



MSE curve



Reconstruction



PESQ

Copyright © 2017 UEC&NII. All rights reserved.