

複素RBM：制限ボルツマンマシンの複素数拡張と音声信号への応用と評価

中鹿 亘^{1,a)} 高木 信二² 山岸 順一²

概要：ディープニューラルネットワークの事前学習や特徴抽出器として，制限ボルツマンマシン (restricted Boltzmann machine; RBM) がこれまで様々な分野で用いられてきたが，いずれのアプローチでも入力特徴量はバイナリまたは実数値が仮定されてきた．しかし，複素スペクトル，fMRI 画像，音響インテンシティなど，実データに基づいた画像認識や音声信号処理では複素数データを取り扱う場合が多い．本稿では，音声信号の複素スペクトルから直接潜在的な特徴量を抽出する，RBM の拡張モデル (complex-valued RBM; CRBM; 複素 RBM) を用いた音声信号モデリングの有効性を検証する．また，勾配法の一つである Adam を複素拡張した複素 Adam についても新たに提案し，その有効性を評価する．

Evaluation of modeling speech signals using complex extension of restricted Boltzmann machine

TORU NAKASHIKA^{1,a)} SHINJI TAKAKI² JUNICHI YAMAGISHI²

1. はじめに

近年，ディープラーニングを用いた手法が飛躍的に高い精度を上げ，画像認識や音声認識など，幅広い分野において盛んに研究されている [7]．これまでに数多くのディープラーニング手法が提案されているが，最も代表的なモデルとして，制限ボルツマンマシン (restricted Boltzmann machine; RBM) を多層に積み重ねた Deep Belief Net (DBN) が挙げられる [3]．また，RBM 単独でも特徴量抽出手法としてしばしば用いられ，様々な RBM の拡張モデルも提案されている [6], [9], [11], [12]．以上のように，RBM はこれまで様々な分野で用いられてきたにもかかわらず，いずれのアプローチでも入力特徴量はバイナリまたは実数値が仮定されてきた [2], [3], [8]．しかし，音声の複素スペクトル，MRI 画像，音響インテンシティなど，実データに基づいた画像認識や音声信号処理では複素数データを取り扱う場合

が多い．

音声認識や音声合成では音響特徴量として MFCC やメルケプストラム特徴量，STRAIGHT スペクトルなど，(位相スペクトルと比較して，振幅スペクトルの方が認識や合成に効果的であることが知られているため) 振幅スペクトルに基づいた特徴量が利用される．RBM を用いた特徴抽出でも入力には実数値のメルケプストラム特徴量が利用される場合が多い．しかしこれらは位相情報が欠落しており，元の音声データに対して少なからず情報の損失が存在するため，位相情報を含めた複素数データをそのまま表現することが重要となる．ニューラルネットワークや多変量解析手法である非負値行列因子分解 (Non-negative matrix factorization; NMF) では，複素数データを表現する拡張表現が提案されている [4], [10]．本研究では，複素数データ (複素スペクトル) から直接潜在的な特徴量を抽出する，ディープラーニングや特徴抽出手法の根幹をなす RBM の拡張モデル (complex-valued RBM; CRBM; 複素 RBM) を新たに定義し，その有効性を人工データを使った実験及び音声符号化・復号化実験を通じて評価する．

複素数の入力に対し，実部と虚部を連結した実数ベクト

¹ 電気通信大学
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

² 国立情報学研究所
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

a) nakashika@uec.ac.jp

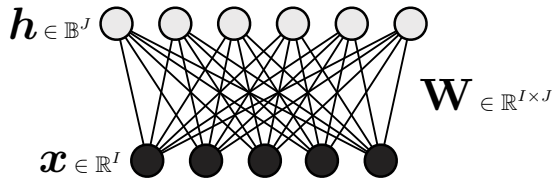


図 1 従来モデル：RBM のグラフ表現（GB-RBM の場合）。

ルを可視素子とすることで，従来の RBM でも複素入力に対応することは可能であるが，複素 RBM は従来の RBM による表現と異なり，複素入力の各素子の実部と虚部には接続（または複素入力とその共役との接続）が存在するため，実部と虚部に相関のある複素数データをより正確に表現することができる．また，従来の RBM と同様，可視素子間及び隠れ素子間には接続がないと仮定しているため，Gibbs サンプリングや CD (contrastive divergence) 法 [3] を用いてパラメータを効率よく推定することができる．一方，ボルツマンマシンの複素入力拡張として DUBM (directional-unit Boltzmann machine) [13] が提案されているが，各素子が全結合されているためパラメータ推定は困難であった．また，複素 RBM と DUBM には，前者では複素可視素子を実部と虚部による直交座標で表現し，後者では絶対値と偏角による極座標で表現するという違いがある．そのため複素 RBM では，隠れ素子が与えられた時の可視素子が複素正規分布に従うので，分布から直接複素乱数を生成させることができる．

2. 従来モデル：RBM

Restricted Boltzmann machine (RBM) は，入力データを表現する可視素子と，潜在的な情報を表現する隠れ素子の間に双方向な接続重みが存在する（ただし可視素子間または隠れ素子間には接続はない）と仮定した確率モデルである．元々 RBM はバイナリ値の入力が想定されていた [2] が，実数値データを扱うために Gaussian-Bernoulli RBM が後に提案された [8]．さらに学習を安定させるために Improved Gaussian-Bernoulli も提案されている [1]^{*1}． I 次元の実数値データ（可視素子） $x \in \mathbb{R}^I$ と， J 個のバイナリ値の隠れ素子 $h \in \mathbb{B}^J$ を表現する RBM は以下のように定義される．

$$p(x; \theta) = \sum_h p(x, h; \theta) \quad (1)$$

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

$$E(x, h; \theta) = \frac{1}{2} x^\top \Sigma^{-1} x - b^\top \Sigma^{-1} x - c^\top h - x^\top \Sigma^{-1} W h \quad (3)$$

$$U(\theta) = \int \sum_h e^{-E(x, h; \theta)} dx \quad (4)$$

^{*1} 本稿ではこの改善モデルを単に RBM と呼ぶ．

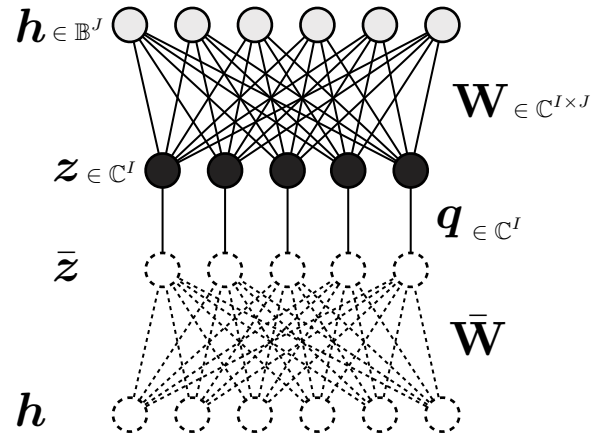


図 2 提案モデル：複素 RBM のグラフ表現．

ただし $\theta = \{b, c, W, \sigma\}$ は RBM のパラメータ集合であり， $b \in \mathbb{R}^I$ ， $c \in \mathbb{R}^J$ はそれぞれ可視素子と隠れ素子のバイアス， $W \in \mathbb{R}^{I \times J}$ は可視素子-隠れ素子間の双方向結合重み， $\Sigma \triangleq \Delta(\sigma^2)$ ， $\sigma \in \mathbb{R}^I$ ($\cdot^2$ は要素ごとの 2 乗を表す) は可視素子の分散を表すパラメータとする (Δ はベクトルを対角成分とする対角行列を表す)．パラメータ θ は， N 個の学習データ集合 $\{x_n\}_{n=1}^N$ の対数尤度 $L(\theta) = \log \prod_{n=1}^N p(x_n; \theta)$ を最大化するように，確率的勾配法を用いて更新される．このとき，計算困難なモデル分布の期待値が勾配に含まれるが，CD (contrastive divergence) 法 [3] を用いて効率よく近似することができる．

式 (1) ~ (4) の定義により，隠れ素子が与えられたときの可視素子の条件付き確率 $p(x|h)$ ，および可視素子が与えられたときの隠れ素子の条件付き確率 $p(h|x)$ は，それぞれ以下の単純な式で計算することができる．

$$p(x|h) = \mathcal{N}(x; b + W h, \Sigma) \quad (5)$$

$$p(h|x) = \mathcal{B}(h; f(c + W^\top \Sigma^{-1} x)) \quad (6)$$

ここで， $\mathcal{N}(\cdot; \mu, \Sigma)$ は平均 μ ，分散共分散行列 Σ の多変量正規分布， $\mathcal{B}(\cdot; \pi)$ は成功確率 π の多次元ベルヌーイ分布， $f(\cdot)$ は要素毎のシグモイド関数を表す．

なお，式 (3) で定義されるエネルギー関数を，以下の式：

$$E(x, h; \theta) = -b^\top x - c^\top h - x^\top W h \quad (7)$$

に置き換えることで，条件付き確率 $p(x|h)$ は，

$$p(x|h) = \mathcal{B}(x; f(b + W h)) \quad (8)$$

となり， x は I 次元のバイナリ値データと見做することができる．

3. 提案モデル：複素 RBM

3.1 定義

確率密度関数（パラメータ推定時におけるコスト関数）は従来同様実数であるが，入力の実数ではなく複素数となる RBM の拡張モデル（Complex-valued RBM；複素 RBM と

呼ぶ)を考える．本稿では， I 次元の複素数データ $z \in \mathbb{C}^I$ を可視素子とする複素 RBM を以下のように定義する．

$$p(z; \theta) = \sum_h p(z, h; \theta) \quad (9)$$

$$p(z, h; \theta) = \frac{1}{U(\theta)} e^{-E(z, h; \theta)} \quad (10)$$

$$E(z, h; \theta) = \frac{1}{2} \begin{bmatrix} z \\ \bar{z} \end{bmatrix}^H \Phi^{-1} \begin{bmatrix} z \\ \bar{z} \end{bmatrix} - \begin{bmatrix} b \\ \bar{b} \end{bmatrix}^H \Phi^{-1} \begin{bmatrix} z \\ \bar{z} \end{bmatrix} - 2c^\top h \quad (11)$$

$$- \begin{bmatrix} z \\ \bar{z} \end{bmatrix}^H \Phi^{-1} \begin{bmatrix} \mathbf{W} \\ \bar{\mathbf{W}} \end{bmatrix} h$$

$$U(\theta) = \int \sum_h e^{-E(z, h; \theta)} dz \quad (12)$$

ただし， $b \in \mathbb{C}^I$ ， $c \in \mathbb{R}^J$ はそれぞれ可視素子と隠れ素子のバイアス， $\mathbf{W} \in \mathbb{C}^{I \times J}$ は可視素子-隠れ素子間の複素結合重みを表し，オーバーラインは複素共役， H はエルミート転置を表す．また，

$$\Phi \triangleq \begin{bmatrix} \Gamma & \mathbf{C} \\ \mathbf{C}^H & \Gamma^H \end{bmatrix} \quad (13)$$

であり，

$$\Gamma \triangleq \Delta(\gamma), \quad \gamma \in \mathbb{R}^+ \quad (14)$$

$$\mathbf{C} \triangleq \Delta(\delta), \quad \delta \in \mathbb{C}^I$$

はそれぞれ複素数 z の分散と擬分散（共役複素数との共分散）を表すパラメータとする．結局複素 RBM のパラメータは $\theta = \{b, c, \mathbf{W}, \gamma, \delta\}$ となる．ここで，

$$p \triangleq \frac{\gamma}{\gamma^2 - |\delta|^2} \in \mathbb{R}^I \quad (15)$$

$$q \triangleq -\frac{\delta}{\gamma^2 - |\delta|^2} \in \mathbb{C}^I \quad (16)$$

を導入する（ただし分数線は要素除算を表す）と，

$$\Phi^{-1} = \begin{bmatrix} \Delta(p) & \Delta(q) \\ \Delta(\bar{q}) & \Delta(p) \end{bmatrix} \quad (17)$$

となることから，式 (11) で定義されるエネルギー関数は

$$E(z, h; \theta) = z^H \Delta(p) z + \Re(z^H \Delta(q) \bar{z}) - 2\Re(z^H \Delta(p) b) - 2\Re(z^H \Delta(q) \bar{b}) - 2c^\top h - 2\Re(z^H \Delta(p) \mathbf{W}) h - 2\Re(z^H \Delta(q) \bar{\mathbf{W}}) h \quad (18)$$

と書き直すことができ，エネルギー関数は実数値となること，複素可視素子 z の各次元は共役複素数との結合が存在するが，通常の RBM のように次元間の結合は存在しないことが確認できる．さらに，

$$b' \triangleq \Delta(p)b + \Delta(q)\bar{b} \quad (19)$$

$$\mathbf{W}' \triangleq \Delta(p)\mathbf{W} + \Delta(q)\bar{\mathbf{W}} \quad (20)$$

を用いることで，式 (11) は

$$E(z, h; \theta) = \frac{1}{2} z^H \Delta(p) z + \frac{1}{2} \bar{z}^H \Delta(p) \bar{z} + z^H \Delta(q) \bar{z} + \bar{z}^H \Delta(\bar{q}) z - z^H b' - \bar{z}^H \bar{b}' - 2c^\top h - z^H \mathbf{W}' h - \bar{z}^H \bar{\mathbf{W}}' h \quad (21)$$

となり，図 2 で示す（厳密には，図では $\mathbf{W}'$ を $\mathbf{W}$ と表記している）ように z と h ， $\bar{z}$ と h の関係性は互いに共役空間を挟んで鏡像の関係にあることが分かる．さらに， z とその共役 $\bar{z}$ は， q の重みで接続されている．

以上の定義により，隠れ素子が与えられたときの可視素子の条件付き確率，および可視素子が与えられたときの隠れ素子の条件付き確率をそれぞれ以下のように表すことができる．

$$p(z|h) = \mathcal{CN}(z; b + \mathbf{W}h, \Gamma, \mathbf{C}) \quad (22)$$

$$p(h|z) = \mathcal{B}(h; f(2c + 2\Re(\mathbf{W}'^H z))) \quad (23)$$

ただし， $\mathcal{CN}(\cdot; \mu, \Gamma, \mathbf{C})$ は平均 μ ，分散共分散行列 Γ ，擬分散共分散行列 $\mathbf{C}$ の多変量複素正規分布：

$$p(z) = \frac{1}{\pi^D \sqrt{\det(\Gamma) \det(\mathbf{C})}} \cdot \exp \left\{ -\frac{1}{2} \begin{bmatrix} z - \mu \\ \bar{z} - \bar{\mu} \end{bmatrix}^H \begin{bmatrix} \Gamma & \mathbf{C} \\ \mathbf{C}^H & \Gamma^H \end{bmatrix}^{-1} \begin{bmatrix} z - \mu \\ \bar{z} - \bar{\mu} \end{bmatrix} \right\} \quad (24)$$

$$\mathbf{Q} = \bar{\Gamma} - \mathbf{C}^H \Gamma^{-1} \mathbf{C} \quad (25)$$

を表す．

3.2 パラメータ推定

本稿では，複素 RBM のパラメータを， N 個の独立な複素数データ集合 $\{z_n\}_{n=1}^N$ の対数尤度：

$$L(\theta) = \log \prod_{n=1}^N p(z_n; \theta) \quad (26)$$

$$= \sum_{n=1}^N \log \sum_{h_n} p(z_n, h_n; \theta) \quad (27)$$

を最大化するように複素勾配降下法によって更新する（または，次節で述べる複素 Adam で更新する）．複素勾配降下法は学習率 $\alpha > 0$ を用いて

$$\theta^{(\text{new})} \leftarrow \theta^{(\text{old})} + 2\alpha \frac{\partial L}{\partial \theta} \quad (28)$$

を繰り返し計算することでパラメータを更新する．ただし，式 (28) における複素数の偏微分はウェルティンガーの

微分：

$$\frac{\partial L}{\partial \theta} = \frac{1}{2} \left(\frac{\partial L}{\partial \Re(\theta)} - i \frac{\partial L}{\partial \Im(\theta)} \right) \quad (29)$$

である (i は虚数単位)。

一方, $\frac{\partial L}{\partial \theta}$ は,

$$\frac{\partial L}{\partial \theta} = \mathbb{E}_{p(\mathbf{h}|\mathbf{z})} \left[-\frac{\partial E(\mathbf{z}, \mathbf{h})}{\partial \theta} \right] - \mathbb{E}_{p(\mathbf{z}', \mathbf{h}')} \left[-\frac{\partial E(\mathbf{z}', \mathbf{h}')}{\partial \theta} \right] \quad (30)$$

と導けるので, 各パラメータのエネルギー関数に対する偏微分を用いて計算される。ただし, 式 (30) 右辺の第 1 項は観測データが与えられた時の隠れ素子の期待値, 第 2 項は式 (10) で定義されるモデルの期待値を表す。モデルの期待値は一般に計算困難であるが, $p(\mathbf{z}|\mathbf{h})$ と $p(\mathbf{h}|\mathbf{z})$ の計算及びそれらに従う乱数サンプリングが容易であるため, 従来の RBM と同様, CD 法 (contrastive divergence) を用いて近似することができる。具体的には, モデルの期待値を, あるデータ $\mathbf{z}_n$ が与えられたときの隠れ素子の条件付き確率 $p(\mathbf{h}_n|\mathbf{z}_n)$ に従う乱数 $\tilde{\mathbf{h}}_n$ を生成し, その隠れ素子が与えられたときの可視素子の条件付き確率 $p(\mathbf{z}_n|\tilde{\mathbf{h}}_n)$ に従う乱数 (再構築データ) $\tilde{\mathbf{z}}_n$ が与えられた時の隠れ素子の期待値として計算する。

エネルギー関数に対するパラメータの偏微分は解析的に求めることができ, それぞれ以下の通りになる。

$$-\frac{\partial E}{\partial \mathbf{b}} = \Delta(\mathbf{p})\bar{\mathbf{z}} + \Delta(\bar{\mathbf{q}})\mathbf{z} \quad (31)$$

$$-\frac{\partial E}{\partial \mathbf{c}} = \mathbf{h} \quad (32)$$

$$-\frac{\partial E}{\partial \mathbf{W}} = (\Delta(\mathbf{p})\bar{\mathbf{z}} + \Delta(\bar{\mathbf{q}})\mathbf{z})\mathbf{h}^\top \quad (33)$$

$$-\frac{\partial E}{\partial \gamma} = (\mathbf{p}^2 + |\mathbf{q}|^2) \circ \frac{\partial E}{\partial \mathbf{p}} + 2\Re(\mathbf{p} \circ \mathbf{q} \circ \frac{\partial E}{\partial \mathbf{q}}) \quad (34)$$

$$-\frac{\partial E}{\partial \delta} = \mathbf{p}^2 \circ \frac{\partial E}{\partial \mathbf{q}} + \bar{\mathbf{q}}^2 \circ \frac{\partial E}{\partial \bar{\mathbf{q}}} + 2\mathbf{p} \circ \bar{\mathbf{q}} \circ \frac{\partial E}{\partial \mathbf{p}} \quad (35)$$

ただし, $\circ, |\cdot|$ はそれぞれ要素ごとの積, 絶対値を表し,

$$\frac{\partial E}{\partial \mathbf{p}} = \frac{1}{2}|\mathbf{z}|^2 - \Re(\mathbf{z} \circ (\bar{\mathbf{b}} + \bar{\mathbf{W}}\mathbf{h})) \quad (36)$$

$$\frac{\partial E}{\partial \bar{\mathbf{q}}} = \frac{1}{2}\bar{\mathbf{z}}^2 - \bar{\mathbf{z}} \circ (\bar{\mathbf{b}} + \bar{\mathbf{W}}\mathbf{h}) \quad (37)$$

である。分散および擬似分散の更新は, 他のパラメータと比較してスケールが異なるため, 安定して学習させるために実際には $\gamma \triangleq e^r$, $\delta \triangleq e^s$ とおき, r および s でパラメータ更新を行う。

3.3 複素 Adam による最適化

ニューラルネットワークのパラメータ最適化手法として Adam (adaptive moment estimation)[5] が提案され, 勾配降下法よりも収束が早く, 性能が高いことが報告されている。本研究では, 実数値で定義されていた Adam を複素数に拡張した複素 Adam (CAdam) を提案する。

複素 Adam では, 式 (28) の代わりに, ハイパーパラメータである減衰率 $\beta_1 \in [0, 1]$, $\beta_2 \in [0, 1]$ を用いて,

$$\beta_1^{(\text{new})} \leftarrow \beta_1 \cdot \beta_1^{(\text{old})} \quad (38)$$

$$\beta_2^{(\text{new})} \leftarrow \beta_2 \cdot \beta_2^{(\text{old})} \quad (39)$$

$$\begin{aligned} \mathbf{m}^{(\text{new})} &\leftarrow \beta_1^{(\text{new})} \mathbf{m}^{(\text{old})} \\ &+ (1 - \beta_1^{(\text{new})}) \left(\frac{\partial L}{\partial \Re(\theta)} + i \frac{\partial L}{\partial \Im(\theta)} \right) \end{aligned} \quad (40)$$

$$\begin{aligned} \mathbf{v}^{(\text{new})} &\leftarrow \beta_2^{(\text{new})} \mathbf{v}^{(\text{old})} \\ &+ (1 - \beta_2^{(\text{new})}) \left(\left(\frac{\partial L}{\partial \Re(\theta)} \right)^2 + \left(\frac{\partial L}{\partial \Im(\theta)} \right)^2 \right) \end{aligned} \quad (41)$$

$$\theta^{(\text{new})} \leftarrow \theta^{(\text{old})} + \alpha \cdot \frac{\sqrt{1 - \beta_2^{(\text{new})}}}{1 - \beta_1^{(\text{new})}} \cdot \frac{\mathbf{m}^{(\text{new})}}{\sqrt{\mathbf{v}^{(\text{new})}}} \quad (42)$$

と各パラメータを更新する。ただし, 式 (41) における $\cdot^2$ は要素ごとの二乗, 式 (42) における $\sqrt{\cdot}$ は要素ごとの二乗根, 分数線は要素除算を表す。

3.4 従来の RBM を用いた複素数表現との違い

複素数 $z = x + iy$ は, 実部と虚部の連結ベクトル $\tilde{z} = [\mathbf{x}^\top \mathbf{y}^\top]^\top \in \mathbb{R}^{2I}$ を用いることで, 以下のように従来の RBM で表現することもできる。

$$p(\mathbf{x}, \mathbf{y}; \theta) = \sum_{\mathbf{h}} p(\mathbf{x}, \mathbf{y}, \mathbf{h}; \theta) \quad (43)$$

$$p(\mathbf{x}, \mathbf{y}, \mathbf{h}; \theta) = \frac{1}{U(\theta)} e^{-E(\mathbf{x}, \mathbf{y}, \mathbf{h}; \theta)} \quad (44)$$

$$\begin{aligned} E(\mathbf{x}, \mathbf{y}, \mathbf{h}; \theta) &= \frac{1}{2} \mathbf{x}^\top \Sigma_x^{-1} \mathbf{x} + \frac{1}{2} \mathbf{y}^\top \Sigma_y^{-1} \mathbf{y} \\ &\quad - \mathbf{b}_x^\top \Sigma_x^{-1} \mathbf{x} - \mathbf{b}_y^\top \Sigma_y^{-1} \mathbf{y} - \mathbf{c}^\top \mathbf{h} \\ &\quad - \mathbf{x}^\top \Sigma_x^{-1} \mathbf{W}_x \mathbf{h} - \mathbf{y}^\top \Sigma_y^{-1} \mathbf{W}_y \mathbf{h} \end{aligned} \quad (45)$$

$$U(\theta) = \iint \sum_{\mathbf{h}} e^{-E(\mathbf{x}, \mathbf{y}, \mathbf{h}; \theta)} d\mathbf{x} d\mathbf{y} \quad (46)$$

ただし, $\Sigma_x = \Delta(\sigma_x^2)$, $\Sigma_y = \Delta(\sigma_y^2)$ である。この場合, 例えばエネルギー関数に対する実部, 虚部のバイアスパラメータの偏微分はそれぞれ

$$-\frac{\partial E}{\partial \mathbf{b}_x} = \frac{\mathbf{x}}{\sigma_x^2} \quad (47)$$

$$-\frac{\partial E}{\partial \mathbf{b}_y} = \frac{\mathbf{y}}{\sigma_y^2} \quad (48)$$

となる。

一方で, $\mathbf{z} = \mathbf{x} + i\mathbf{y}$, $\mathbf{b} = \mathbf{b}^R + i\mathbf{b}^I$, $\mathbf{W} = \mathbf{W}^R + i\mathbf{W}^I$, $\mathbf{q} = \mathbf{q}^R + i\mathbf{q}^I$ とし, 複素 RBM のエネルギー関数 (式 (11) 右辺) を書き換えると

$$\begin{aligned} &\frac{1}{2} \mathbf{x}^\top \Sigma_x^{-1} \mathbf{x} + \mathbf{x}^\top \Sigma_{xy}^{-1} \mathbf{y} + \frac{1}{2} \mathbf{y}^\top \Sigma_y^{-1} \mathbf{y} \\ &\quad - \mathbf{b}_x^\top \Sigma_x^{-1} \mathbf{x} - \mathbf{b}_y^\top \Sigma_y^{-1} \mathbf{y} - 2\mathbf{c}^\top \mathbf{h} \\ &\quad - \mathbf{x}^\top \Sigma_x^{-1} \mathbf{W}_x \mathbf{h} - \mathbf{y}^\top \Sigma_y^{-1} \mathbf{W}_y \mathbf{h} \end{aligned} \quad (49)$$

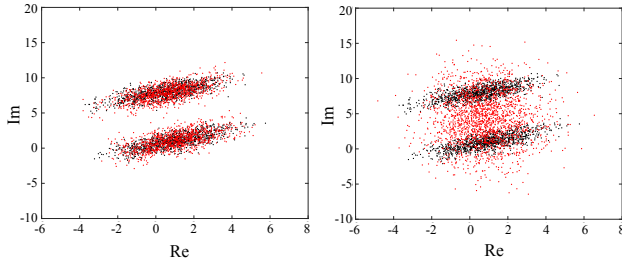


図 3 人工的に生成した 1 次元複素数データ (黒点) と、学習された複素 RBM (上図) 及び RBM (下図) からランダムに生成されたデータ (赤点) を示す。

となる。ただし

$$\Sigma_x = \Delta \left(\frac{1}{2(p + q^R)} \right) \quad (50)$$

$$\Sigma_y = \Delta \left(\frac{1}{2(p - q^R)} \right) \quad (51)$$

$$\Sigma_{xy} = \Delta \left(\frac{1}{2q^I} \right) \quad (52)$$

$$b_x = b^R + \frac{q^I}{p + q^R} \circ b^I \quad (53)$$

$$b_y = b^I + \frac{q^I}{p - q^R} \circ b^R \quad (54)$$

$$W_x = W^R + \Delta \left(\frac{q^I}{p + q^R} \right) W^I \quad (55)$$

$$W_y = W^I + \Delta \left(\frac{q^I}{p - q^R} \right) W^R \quad (56)$$

といた。式 (45) と式 (49) を比較すると、複素 RBM によるモデル化では x と y のクロス項 ($x^T \Sigma_{xy}^{-1} y$) が含まれていることが分かる。すなわち複素 RBM では、RBM による複素数表現に加えて、特徴量次元ごとに実部と虚部との関係性を考慮した拡張表現であると言える。

さらに RBM による複素数表現では、式 (47)(48) で示されるように、観測データの実部と虚部のバイアスがそれぞれ独立して (例えば実部バイアスの更新では実部のみの情報が用いられる) 計算されるのに対し、式 (31) による複素 RBM のバイアスパラメータの更新式では、実部と虚部の両方が用いられて更新される。そのため複素 RBM では複素数のデータ構造を保ったまま学習を行うことができると期待される。

4. 評価実験

4.1 人工データを用いた評価実験

複素 RBM の有効性を確認するため、まず初めに、人工的に生成した $N = 2000$ 個の 1 次元複素数データを用いた実験を行った。この人工データは、図 3 (黒点) に示すように、実部と虚部に相関を持つように生成された。本実験では、複素 RBM による人工データのモデリングと、実部・虚部を表す 2 次元の可視素子を持つ RBM によるモデリングの結果を比較した。両モデルとも隠れ素子数 2 とし、学

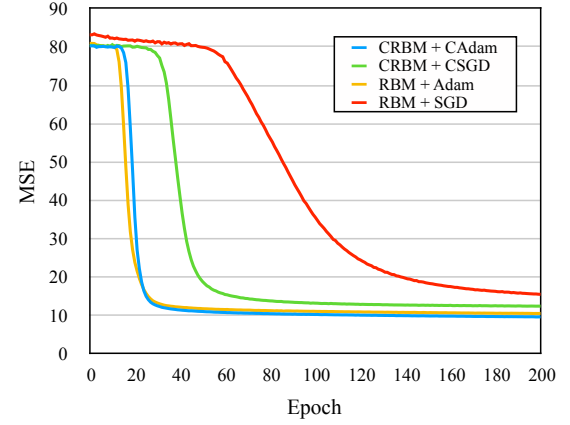


図 4 各繰り返しステップにおける複素 RBM (複素確率的勾配法を用いた場合と複素 Adam を用いた場合) と RBM (確率的勾配降下法を用いた場合と Adam を用いた場合) の再構築エラー。

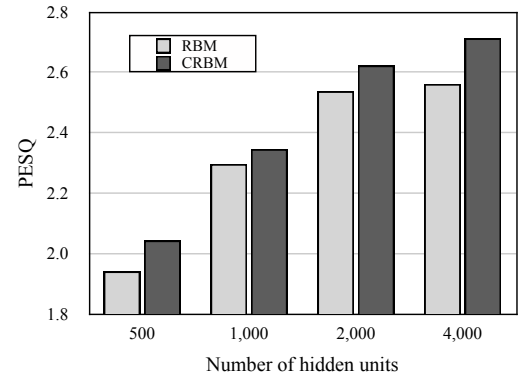


図 5 隠れ素子数を変えた場合の、複素 RBM 及び RBM によって復元された音声の PESQ による品質評価。

習率 0.01、モーメント係数 0.1、バッチサイズ 20、繰り返し回数 200 の確率的勾配降下法によって学習を行った。その後、両方のモデルについて、学習データを入力し、隠れ素子の期待値を計算 (エンコード)、この隠れ素子から逆に可視素子の確率を計算 (デコード) し、乱数を生成させた。複素 RBM 及び RBM によって生成された乱数をそれぞれ図 3 上図 (赤点) 及び下図 (赤点) に示す。図 3 より、複素 RBM の方が RBM よりも、元の複素人工データ分布をよく表現できていることが分かる。これは複素 RBM では実部と虚部の相関を表現する構造を持っているのに対し、RBM では相関を表現することができないからだと考えられる。

4.2 音声データを用いた評価実験

次に、音声データを用いて提案法である複素 RBM の評価実験を行なった。本実験では ATR 音声データベースセット A 女性アナウンサー (FTK) 50 文音声 (約 4.2 分、サンプリングレート 20kHz を 16kHz にダウンサンプリング) をモデルの学習に、同じ話者の別の 53 文を評価に用いた。それぞれの音声データについて、窓幅 256、64 サンプルオー

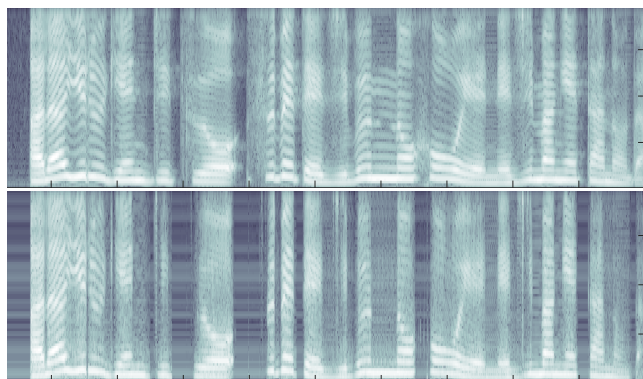


図 6 オリジナルの振幅スペクトル（上）と複素 RBM によって復元されたスペクトル（下）．縦軸は周波数，横軸は時間を表す．

オーバーラップの短時間フーリエ変換を施した複素スペクトル (129 次元) を，複素 PCA で 40 次元に圧縮した特徴量を複素 RBM の可視素子とした．なお，複素 PCA で 40 次元に圧縮したデータから複素スペクトルを復元し，得られた音声の PESQ (Perceptual evaluation of speech quality) 値が 4.46 であり，複素 PCA による品質劣化が少ないことから，本実験では次元削減のため，可視素子として複素スペクトルではなく複素 PCA を用いた．まず，学習率 0.01，モーメント係数 0.1，バッチサイズ 100，繰り返し回数 200 の複素確率的勾配降下法と，学習率 0.001，減衰率 $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ，バッチサイズ 100，繰り返し回数 200 の複素 Adam の 2 つのモデル（隠れ素子数は 1,000）を学習させた．また比較手法として，同じ複素数データの実部と虚部を連結したベクトルを可視素子とした RBM（隠れ素子数は同様に 1000）を，同様の条件の確率的勾配降下法と Adam を用いて学習させた．それぞれの手法について，学習時の各更新ステップ（Epoch）毎の再構築エラー（MSE）をプロットしたものを図 4 に示す．まず，複素確率的勾配降下法を用いた複素 RBM（CARBM+CSGD）と確率的勾配降下法を用いた RBM（RBM+SGD）を比較すると，前者の方が収束が早く，収束時におけるエラーが低いことが分かる．また，複素 Adam を用いた複素 RBM（CARBM+CAAdam）と Adam を用いた RBM（RBM+Adam）を比較すると，わずかながら複素 RBM の方が収束時の再構築エラーが低くなっている．複素 Adam 及び Adam を用いた方が精度が高いことから，以降の実験ではこれらを用いる．

最後に，隠れ素子数を 500，1,000，2,000，4,000 と変化させて複素 RBM 及び RBM を学習し，それぞれのモデルで評価用音声を復元した音声について，オリジナル音声を参照した PESQ による客観的比較を行った．図 5 に示すように，どの隠れ素子数においても，複素 RBM の方が PESQ 値が高いことが分かる．また，複素 RBM と RBM のいずれにおいても隠れ素子数を増やすと PESQ 値が向上しているが，RBM では隠れ素子数が 2,000 の時に飽和している．一方，複素 RBM では 4,000 まで隠れ素子数を増

加させても PESQ 値の向上が見られる．

また，オリジナルの振幅スペクトルと，隠れ素子数が 1,000 の複素 RBM によって復元されたスペクトルを図 6 に示す．図 6 より，復元されたスペクトルはオリジナルのスペクトルに近く，フォルマントや基本周波数など，音声の特徴をよく捉えられていることが分かる．

5. おわりに

本研究では，複素数データを可視素子とする RBM の拡張モデル（複素 RBM）を提案し，複素数の人工データ及び音声データを用いて性能評価を行なった．後者の実験では，従来の RBM によって復元された音声よりも，複素 RBM によって復元された音声の方が PESQ 値が高く，品質向上が見られた．また，複素 RBM の学習では，複素 Adam によってパラメータを推定する方が，勾配降下法よりも効果的であることが確認できた．今後は複素 RBM を音声合成や声質変換へ応用したい．

参考文献

- [1] Cho, K., Ilin, A. and Raiko, T.: Improved learning of Gaussian-Bernoulli restricted Boltzmann machines, *Artificial Neural Networks and Machine Learning-ICANN 2011*, Springer, pp. 10–17 (2011).
- [2] Freund, Y. and Haussler, D.: Unsupervised learning of distributions of binary vectors using two layer networks (1994).
- [3] Hinton, G. E., Osindero, S. and Teh, Y.-W.: A fast learning algorithm for deep belief nets, *Neural computation*, Vol. 18, No. 7, pp. 1527–1554 (2006).
- [4] Kameoka, H., Ono, N. and Kashino, K.: Complex NMF: A new sparse representation for acoustic signals, *ICASSP*, pp. 3437–3440 (2009).
- [5] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *ICLR* (2015).
- [6] Krizhevsky, A. and Hinton, G. E.: Factored 3-way restricted Boltzmann machines for modeling natural images, *Journal of Machine Learning Research* (2010).
- [7] LeCun, Y., Bengio, Y. and Hinton, G.: Deep learning, *Nature*, Vol. 521, No. 7553, pp. 436–444 (2015).
- [8] Lee, H., Ekanadham, C. and Ng, A. Y.: Sparse deep belief net model for visual area V2, pp. 873–880 (2008).
- [9] Nakashika, T., Takiguchi, T. and Minami, Y.: 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).
- [10] Nemoto, I. and Kono, T.: Complex neural networks, *Systems and computers in Japan*, Vol. 23, No. 8, pp. 75–84 (1992).
- [11] Salakhutdinov, R. and Hinton, G. E.: Deep Boltzmann Machines, *AISTATS* (2009).
- [12] Sohn, K., Zhou, G., Lee, C. and Lee, H.: Learning and Selecting Features Jointly with Point-wise Gated Boltzmann Machines, *ICML (2)* (2013).
- [13] Zemel, R. S., Williams, C. and Mozer, M. C.: Lending direction to neural networks, *Neural Networks*, Vol. 8, No. 4, pp. 503–512 (1995).