

## Gaussian-Bernoulli RBMにおける最尤学習を効率化する直交制約

An efficient algorithm for maximum likelihood estimation under orthogonality constraints in Gaussian-Bernoulli RBM

唐木田亮<sup>\*1</sup> 岡田真人<sup>\*1\*2</sup> 甘利俊一<sup>\*2</sup>  
Ryo Karakida Masato Okada Shun-ichi Amari<sup>\*1</sup>東京大学大学院 新領域創成科学研究科  
Graduate School of Frontier Sciences, The University of Tokyo<sup>\*2</sup>理化学研究所 脳科学総合研究センター  
RIKEN Brain Science Institute

The restricted Boltzmann machine (RBM) is a bipartite graphical model widely used as a building block of deep neural networks. However, it is hard to train RBMs by using maximum likelihood estimation because many iterations of Gibbs sampling take too much computational time. In this study, we consider Gaussian-Bernoulli RBMs and reveal that if the weight matrices satisfy orthogonality constraints, we can easily compute analytical values of the likelihood function and its gradients without using any sampling method. We propose a novel algorithm based on geodesic flow under the orthogonality constraints and demonstrate its effectiveness in numerical experiments with natural image dataset.

## 1. はじめに

Restricted Boltzmann Machine (RBM) は完全 2 部グラフ型の生成モデルであり、深層学習の構成要素として開発が進んだ [Hinton 06, Lee 07]. 生成モデルの典型的な学習方法は尤度を評価関数とした最尤学習であるが、通常の RBM では尤度が解析的に計算できない。そこで、MCMC に基づくサンプリング、あるいは平均場近似のような近似推論が必要とされる。特に、サンプリングを有限回で止める Contrastive Divergence (CD) 学習が最尤学習の粗い近似として開発され、経験的には短時間で最尤推定値に近いパラメータに収束することから広く用いられている [Perpignan 05]. しかし、一般には CD 学習を含む近似アルゴリズムは学習の収束や尤度の最大化が保証されておらず、理論的な解の特徴付けが難しい点が短所である。

本研究では、連続値入力とバイナリ隠れ変数を持つ Gaussian-Bernoulli RBM において、結合パラメータ行列に直交制約を要請すると、尤度と尤度勾配の厳密値が解析的に求まることを明らかにする。直交制約下でのパラメータ更新を実現するために測地線勾配を使ったアルゴリズムを提案し、自然画像を使った数値実験で制約無しの既存の学習アルゴリズムと同等な性能を達成できることを示す。

## 2. モデル

Gaussian-Bernoulli(G-B) RBM はバイナリ隠れ変数  $h_i = \{0, 1\}$  ( $i = 1, \dots, M$ ) と連続入力変数  $v_i$  ( $i = 1, \dots, N$ ) を持つ RBM の一種で、モデル分布が以下で定義される:

$$p(\mathbf{h}, \mathbf{v}) = \exp\left(-\frac{1}{2\sigma^2}|\mathbf{v} - \mathbf{b}|^2 + \frac{1}{\sigma}\mathbf{h}^T W \mathbf{v} + \mathbf{c}^T \mathbf{h}\right) / Z. \quad (1)$$

ここで、モデル分散パラメータを  $\sigma^2$ 、規格化因子を  $Z$  とした。G-B RBM は深層ネットワークにおいて画像や音声といった連続値入力を受ける末端の 2 層としての利用 [Hinton 06, Lee 07] やモデルの拡張 [Ranzato 13] が進んでいる。本研究では、結合パラメータ行列  $W \in \mathbb{R}^{M \times N}$  とバイアス項パラメータ  $\mathbf{b}, \mathbf{c}$  の学習を考える。

最尤学習は、負の対数尤度  $L = -\int q(\mathbf{v}) \ln p(\mathbf{v}) d\mathbf{v}$  を最小化するパラメータを求める方法である。ここで、 $q(\mathbf{v})$  は入力

連絡先: 唐木田亮, karakida@mns.k.u-tokyo.ac.jp

分布,  $p(\mathbf{v})$  はモデル分布を表す。RBM において尤度の最急勾配は  $\Delta W = \langle \mathbf{h} \mathbf{v}^T \rangle_q - \langle \mathbf{h} \mathbf{v}^T \rangle_p$  となり、更新則は微小な学習係数  $\epsilon$  を用いて  $W_{t+1} = W_t + \epsilon \Delta W$  と定式化される [Hinton 06]. 記号  $\langle \cdot \rangle_q$  は入力分布から生成された訓練データに対する平均、 $\langle \mathbf{h} \mathbf{v}^T \rangle_p$  はモデル分布に対する平均で以下のように式変形できる:

$$\langle \mathbf{h} \mathbf{v}^T \rangle_p = \sigma \langle \mathbf{h} \mathbf{h}^T \rangle_{p(\mathbf{h})} W + \langle \mathbf{h} \rangle_{p(\mathbf{h})} \mathbf{b}^T. \quad (2)$$

すなわち、モデル平均は以下の周辺化されたモデル分布  $p(\mathbf{h})$  に従う  $2^M$  個の隠れ状態  $\mathbf{h}$  で和をとる必要があり、 $M$  が大きければ実用上は計算困難である:

$$p(\mathbf{h}) = \exp\left(-|\mathbf{W}^T \mathbf{h}|^2 / 2 + (\mathbf{W} \mathbf{b} / \sigma + \mathbf{c})^T \mathbf{h}\right) / Z. \quad (3)$$

## 3. 提案手法

## 3.1 直交制約と隠れ変数の独立性

結合行列  $W$  の探索空間を制限すれば、モデル平均 (2) が解析的に計算できることを示す。具体的には、 $A$  を各行が直交した  $M \leq N$  を満たす行列 ( $AA^T = I_M$ )、 $D$  を対角行列  $D = \text{diag}(d_1, d_2, \dots, d_M)$  とし  $W = DA$  を要請する。本研究では、 $W = DA$  を直交制約と呼ぶ。直交制約を式 (3) に代入すると、

$$p(\mathbf{h}) = \prod_{i=1}^M g(y_i)^{h_i} (1 - g(y_i))^{1-h_i}. \quad (4)$$

ただし、 $y_i = d_i^2 / 2 + d_i \mathbf{a}_i \mathbf{b} / \sigma + c_i$ 、 $A$  の  $i$  行ベクトルを  $\mathbf{a}_i$  とおいた。分布 (4) からわかるように、直交制約によってモデル分布  $p(\mathbf{h})$  は隠れ変数間が独立となる。直交制約  $W = DA$  は行列  $W$  の自由なパラメータ数を実効的に下げる正則化 [Fiori 05] となるだけでなく、周辺分布 (4) において隠れ素子間が独立になる事前知識を与えることに等しい。

直交制約下の独立な分布 (4) を使ってモデル平均 (2) は解析的に計算できて、結合行列  $A$  の勾配が以下の式で求まる:

$$\Delta A = D[\langle g(\mathbf{W} \mathbf{v} / \sigma + \mathbf{c}) \mathbf{v}^T \rangle_q - (\sigma K W + g(\mathbf{y}) \mathbf{b}^T)]. \quad (5)$$

シグモイド関数  $g(\cdot)$  を使って、 $K_{ii} = g(y_i)$  ( $i = 1, \dots, M$ )、 $K_{ij} = g(y_i)g(y_j)$  ( $i \neq j$ ) とおいた。

### 3.2 更新則

直交制約を満たす行列  $A$  について, 通常の最急勾配法  $A_{t+1} = A_t - \epsilon \Delta A$  は更新後のパラメータ  $A_{t+1}$  が直交制約を破ることに注意が必要である. この問題に対し, Stiefel 多様体上における測地線に沿った勾配を考えることで, 以下のように直交制約を保つ更新則が構成できることが知られている [Nishimori 05]:

$$A_{t+1} \leftarrow A_t \exp \left( \epsilon (A_t^T \Delta A - \Delta A^T A_t) / 2 \right). \quad (6)$$

この更新では  $A_t$  が直交制約を満たせば, 自動的に  $A_{t+1}$  も直交制約を満たす. 独立成分分析に代表される信号処理では, この測地線勾配が通常の最急勾配を上回る性能を發揮している [Nishimori 05, Fiori 05]. 本研究では, G-B RBM における  $A$  の更新を (6) に (5) を代入して得た. また, 制約の無い他のパラメータの更新則は通常の最急勾配で与えた,

$$\begin{aligned} d_i &\leftarrow d_i + \epsilon [\langle g(d_i \mathbf{a}_i \mathbf{v} / \sigma + c_i) \mathbf{a}_i \mathbf{v} \rangle_q - g(y_i) (\mathbf{a}_i \mathbf{b} + \sigma d_i)], \\ \mathbf{b} &\leftarrow \mathbf{b} + \epsilon [\langle \mathbf{v} \rangle_q - (\mathbf{b} + \sigma W^T g(\mathbf{y}))], \\ \mathbf{c} &\leftarrow \mathbf{c} + \epsilon [\langle g(W \mathbf{v} / \sigma + \mathbf{c}) \rangle_q - g(\mathbf{y})]. \end{aligned}$$

なお, G-B RBM では直交制約のもとで勾配だけでなく尤度も解析的に計算可能である. 負の対数尤度は以下で与えられる:

$$\begin{aligned} L = &\langle \frac{1}{2\sigma^2} \|\mathbf{v} - \mathbf{b}\|^2 - \sum_i \ln(1 + e^{d_i \mathbf{a}_i \mathbf{v} / \sigma + c_i}) \rangle_q \\ &+ \sum_i \ln(1 + e^{y_i}) + N \ln(\sqrt{2\pi}\sigma). \end{aligned} \quad (7)$$

制約無しの最尤学習では G-B RBM の尤度の計算が困難であるため, 学習途中の尤度を確認することが難しい. これに対し, 直交制約下では尤度を計算量  $O(MN)$  で計算できるため, 素子数が大きい場合も学習の収束状況を容易に計算して確認できる利点がある.

### 4. 数値実験

自然画像入力に対して提案手法の数値実験を行った. van Hateren 自然画像データセットから  $14 \times 14$  サイズのバッチを 50,000 枚取り出し, ZCA 白色化を行なった [Lee 07]. 訓練データを 40,000 枚, テストデータを 10,000 枚とした.  $\sigma^2$  は訓練データにおける分散値を与えて固定した.

図 1 に  $M = 16, N = 196$  の G-B RBM を提案手法 (赤) と直交制約無しの場合の既存手法 (黒) で学習した結果を示す. 図中の実線は学習の初期値が異なる 10 試行に対する平均と偏差を表し, 縦軸はテストデータにおける負の対数尤度 (7) である. ここでは既存手法として persistent CD 学習 [Tieleman 09] を使用し, CD 学習時の尤度を計算するために隠れ素子数  $M$  を小さく設定した. 提案手法はパラメータが  $W = DA$  に制限されているため, 制約無しの CD 学習よりパラメータの自由度が低いにもかかわらず, 既存手法と同様な尤度を達成した.

次に, 図 2 に  $M = N = 196$  の G-B RBM を提案手法で学習した結果を示す. 各パッチは, 学習の平衡点において  $d_i \neq 0$  を満たす  $W$  の  $i$  行ベクトルをランダムに取り出し  $14 \times 14$  に変形したものである. 自然画像に特有なガボールフィルタが抽出されており, 直交制約下でも制約無しと同様な特徴抽出がされた. なお, G-B RBM では入力データが独立情報源の線形混合で与えられた場合, 最尤推定値のひとつが独立成分分析の復元行列に対応することが理論解析 [Karakida 14] によって指摘されている. G-B RBM のモデル分布は最尤学習によって入力を独立な成分に分解する傾向があるため, 隠れ変数の独立性を要請することに等しい直交制約は G-B RBM において妥当な仮定と示唆される.

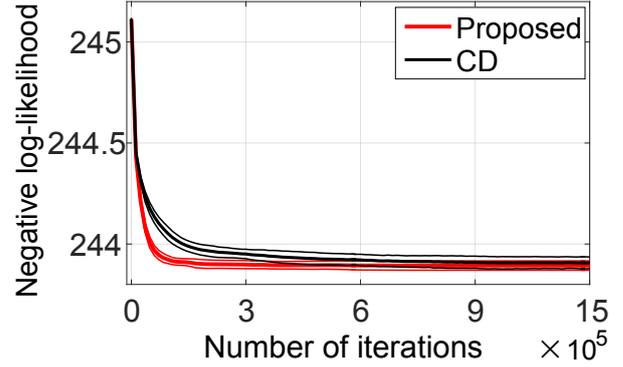


図 1: 自然画像入力における提案手法と既存手法の比較

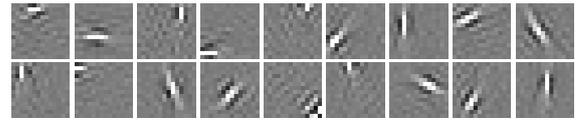


図 2: 自然画像入力において直交制約下で抽出された特徴

### 5. 結論

本研究では, Gaussian-Bernoulli RBM の結合行列を直交に制約すると尤度とその勾配が解析的に計算できることに注目し, 直交制約付きのアルゴリズムを提案した. また, 自然画像入力において提案手法の有効性を示した. 一般に, CD 学習は学習の発散・振動を避けるようにサンプリング方法やパラメータ初期値を選択する必要がある上に, 収束先が最尤推定値とは限らない. これに対し, 本手法は  $W = DA$  上の最尤推定値へ最急勾配に従って収束することが保証されている点, 学習中の尤度が容易に計算できるため収束や最大化を確認しやすい点で, CD 学習のような近似アルゴリズムよりも効率のよい最尤学習であるといえる.

同様の直交制約をバイナリ入力や教師ラベル付きデータに対応した RBM, あるいは RBM の一般系である Exponential Family Harmoniums モデルに適用することは今後の課題である. また, 深層モデルでは層間の結合が事前学習で直交行列に制約されると fine tuning の収束が加速することが示唆されており, 本研究の適用が期待できる [Saxe 13].

### 参考文献

- [Hinton 06] G. E. Hinton and R. Salakhutdinov, *Science*, 313(5786):504-507, 2006.
- [Lee 07] H. Lee, C. Ekanadham, and A. Ng, In *Proceedings of NIPS*, 2007.
- [Perpignan 05] M. A. Carreira-Perpignan, and G. E. Hinton, In *Proceedings of the 10<sup>th</sup> AISTATS*, 2005.
- [Ranzato 13] M. A. Ranzato, V. Mnih, J. M. Susskind, and G. E. Hinton, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2206-2222, 2013.
- [Nishimori 05] Y. Nishimori and S. Akaho, *Neurocomputing*, 67:106-135, 2005.
- [Fiori 05] S. Fiori, *The Journal of Machine Learning Research*, 6:743-781, 2005.
- [Tieleman 09] T. Tieleman and G. E. Hinton, In *Proceedings of the 26<sup>th</sup> ICML*, 2009.
- [Karakida 14] R. Karakida, M. Okada and S. Amari, *Deep Learning and Representation Learning Workshop: NIPS 2014*.
- [Saxe 13] A. Saxe, J. L. McClelland and S. Ganguli, *ICLR*, 2013.