

## 修士論文の和文要旨

研究科・専攻	大学院 情報理工学研究科 情報・ネットワーク工学専攻 博士前期課程		
氏名	豊田 健太	学籍番号	1631101
論文題目	ボルツマンマシンにおけるクラスター型モデルへの近似に関する情報幾何学的考察		
要旨	<p>ボルツマンマシンとは相互結合型ニューラルネットワークの一種であり、確率的な状態変化規則に従って状態を変化させていくネットワークである。また、状態変化を十分に繰り返すことで、ボルツマンマシンの状態が従う分布はある極限分布（平衡確率分布）へと一意に収束することが知られている。</p> <p>ボルツマンマシンの学習とは、ボルツマンマシンの平衡確率分布を観測データが従う未知の分布へと近づけていくことである。1985年、Hintonらはこれを実現させるための学習アルゴリズムを最初に提案した。この学習アルゴリズムでは、平衡確率分布に関する期待値計算を行っており、この計算にかかる時間はボルツマンマシンの素子数に関して指数関数的に増加してしまう。従って、素子数が多い場合には現実的な時間で学習を終えるのが困難となってしまう。</p> <p>ボルツマンマシンの学習の計算困難性を回避するための手法として、平均場近似がよく用いられる。平均場近似とは、各素子間の相互作用の強さをある種の期待値で置き換える手法である。この手法により、ボルツマンマシンの平衡確率分布に対する近似分布が導出され、近似分布に関する期待値計算を用いることで学習の計算困難性を回避することができる。しかし、平均場近似は少々強引な近似の仕方をしているため、導出される近似分布の精度を向上させる余地があると考えられる。</p> <p>本研究では、田中(1996)により示された平均場近似の情報幾何学的な解釈をベースとして、平均場近似の精度の向上を目指した。具体的には、クラスター型モデルと呼ばれるモデルを用いて、従来の平均場近似よりも近似精度が高い分布が満たす条件式を導出した。さらに、この条件式を満たす分布を導出するためのアルゴリズムを提案した。実際に、提案アルゴリズムに関するシミュレーションを行った結果、従来の平均場近似と比較して、提案アルゴリズムにより導出される近似分布の方が近似の精度が高いことを確認した。</p>		

平成29年度(2017年度)修士論文

# ボルツマンマシンにおけるクラスター型モデル への近似に関する情報幾何学的考察

電気通信大学大学院 情報理工学研究科  
情報・ネットワーク工学専攻

1631101 豊田 健太

指導教員

小川 朋宏 准教授

長岡 浩司 教授

図書館提出日 平成30年(2018年)3月13日

# 目次

第1章	序論	1
1.1	研究背景	1
1.2	研究概要	2
1.3	関連研究	2
1.4	本論文の構成	2
第2章	ボルツマンマシン	4
2.1	ホップフィールドネットワーク	4
2.1.1	モデル	4
2.1.2	状態変化規則	5
2.2	ボルツマンマシンへの拡張	5
2.2.1	状態変化規則	5
2.2.2	ボルツマンマシンの平衡確率分布	6
2.3	ボルツマンマシンの学習	7
2.3.1	学習の目的	7
2.3.2	学習アルゴリズム	7
2.3.3	学習アルゴリズムの問題点	8
第3章	ボルツマンマシンの情報幾何	9
3.1	統計多様体	9
3.2	Fisher 情報行列	10
3.3	指数型分布族	10
3.4	$e$ -射影, $m$ -射影	13
3.5	ボルツマンマシンの学習と $m$ -射影について	14
第4章	平均場近似の情報幾何	16
4.1	ボルツマンマシンにおける平均場近似	16
4.2	平均場近似の情報幾何学的解釈	17
4.3	平均場近似を用いた学習アルゴリズム	19
4.4	アルゴリズムの幾何学的イメージと収束先	20
第5章	クラスター型モデルへの近似	22
5.1	グラフ概念を用いた記法の導入	22
5.2	クラスター型モデルへの近似問題	24
5.3	$e$ -射影点が満たす条件式の導出	26
第6章	ビリーフプロパゲーションを用いた $e$ -射影点導出アルゴリズムについて	31
6.1	クラスターから次元鎖への変形	31
6.2	ビリーフプロパゲーション	32
6.3	提案アルゴリズム	34

---

第7章 数値実験	37
7.1 実行環境と各種パラメータの設定について	37
7.2 提案アルゴリズムの挙動に関するシミュレーション	37
7.2.1 シミュレーション内容	37
7.2.2 結果と考察	39
7.3 平均場近似と提案アルゴリズムの精度比較に関するシミュレーション	42
7.3.1 シミュレーション内容	42
7.3.2 結果と考察	43
7.4 提案アルゴリズムの計算時間に関するシミュレーション	43
7.4.1 シミュレーション内容	43
7.4.2 結果と考察	44
第8章 結論	48
8.1 本論文の成果	48
8.2 今後の課題	48
付録．A	49
A.1 準備	49
A.2 一般化平均場方程式	52
A.3 本論文との関係性	52
付録．B	57
B.1 提案アルゴリズム	57
B.2 提案アルゴリズムの情報幾何学的解釈	58
B.3 提案アルゴリズムの収束性	62
参考文献	65
謝辞	67

# 第1章 序論

## 1.1 研究背景

今日に至るまでに、人工知能を実現させるための技術として様々なモデル・手法が研究されてきた。これを実現させるための技術分野の一つとして、ニューラルネットワークが存在する。ニューラルネットワークとは、脳の神経回路の仕組みを模した数理モデルであり、これまでに多種多様なモデルが提案されてきた [3], [4], [5]。ニューラルネットワークを大別すると、階層型ニューラルネットワークと相互結合型ニューラルネットワークに分けられる。本研究で述べるボルツマンマシン [1] は相互結合型ニューラルネットワークの一種であり、確率的な状態変化規則に従って状態を離散時間的に変化させていくネットワークである。さらに、ボルツマンマシンの状態変化の過程はマルコフ連鎖として見做すことができるため、確率的な状態変化を繰り返すことで、ボルツマンマシンの状態が従う分布はある極限分布（平衡確率分布）へと一意に収束する。

ボルツマンマシンの用途としては、次の二点が挙げられる。一つ目は、組合せ最適化問題を解くことである。ボルツマンマシンは、シミュレーティッド・アニーリングと呼ばれる手法を用いることで、理論的には、組合せ最適化問題の大域的最適解を導出することができる。二つ目は、観測データが従う未知の分布をボルツマンマシンの平衡確率分布で表現することである。ボルツマンマシンの平衡確率分布を未知の分布へと近づけていくための学習アルゴリズムとして、最初に提案されたのが Hinton ら [1] によるものである。一方で、[1] で提案されているボルツマンマシンの学習アルゴリズムでは、平衡確率分布に関する期待値計算を行っており、この計算部分はボルツマンマシンの素子数に関して計算時間が指数関数的に増加してしまう。従って、素子数が多いボルツマンマシンを学習させる場合、現実的な時間で学習を終えるのが困難となってしまう。

ボルツマンマシンの学習の計算困難性を回避するための手法として、Weiss の分子場近似 [6]（以降、平均場近似と呼ぶ）がよく用いられる。平均場近似とは、元々は統計力学分野で生まれた手法であり、相互作用し合う素子から構成される巨視的な物理系に対し、各素子間の相互作用の強さをある種の期待値で置き換える手法である。平均場近似を用いることで、元の平衡確率分布に対する近似分布を導出することができる。しかし、平均場近似は少々強引な近似の仕方をしているため、導出される近似分布の精度を向上させる余地があると考えられる。

ボルツマンマシンや平均場近似を通常とは異なる視点から解析するための手段として、情報幾何学 [10] が用いられる。情報幾何学とは、確率分布を点とする空間（多様体）を考察の対象とした時の微分幾何であり、情報理論・統計学・学習理論などの様々な分野と密接に関わっている学問である。ボルツマンマシンの平衡確率分布を点とする空間を考えたとき、平均場近似により導出される近似分布は、元の平衡確率分布からある部分空間（各頂点間に結合が存在しないボルツマンマシンの平衡確率分布を点とする空間）への射影であることが知られている [12], [13]。この部分空間の範囲をさらに拡大させた時の射影を考えることによって、平均場近似の精度を向上させることが期待できる。

## 1.2 研究概要

本研究では、格子模型型のボルツマンマシンを対象としたときの平均場近似の精度を向上させるために、クラスター型模型とよばれるモデルを用いた。クラスター型模型とは、元の格子模型型のボルツマンマシンを短冊形にいくつかの部分グラフ（以降、クラスターと呼ぶ）へと分割させたモデルである。さらに、このクラスター型模型は各クラスターをループを含まないグラフへと変形することで、ピリーフプロパゲーション [7] を用いることができる。この手法により、各クラスター上で定義される平衡確率分布に関する効率的な期待値計算を行うことができる。また、格子模型型のボルツマンマシンとクラスター型模型の形状を固定し、格子模型型のボルツマンマシンの平衡確率分布から部分空間（クラスター型模型上で定められる平衡確率分布を要素とする空間）への射影が満たす条件式を導出した。

さらに本研究では、この条件式を満たす近似分布を数値計算で求めるためのアルゴリズムを提案し、このアルゴリズムにより導出される近似分布と平均場近似により導出される近似分布との間の精度比較、アルゴリズムの計算時間に関するシミュレーションを行った。シミュレーションの結果、提案したアルゴリズムにより導出される近似分布は、平均場近似により導出される近似分布よりも近似精度が高いことを確認した。さらに、格子模型型のボルツマンマシンのサイズを  $N$  行  $M$  列、クラスター型模型に含まれるクラスターの個数を  $Z$ 、各クラスターの列数を最大  $\hat{M}$  とした場合、考案したアルゴリズムの計算量オーダーは  $O(ZN2^{\hat{M}})$  であることを確認した。従って、格子模型型のボルツマンマシンのサイズが非常に大きくなる場合でも、クラスターの個数を増やして、最大の列数  $\hat{M}$  が小さいクラスター型模型を考え、提案アルゴリズムを実行することにより、平均場近似の場合よりも精度の良い近似分布を現実的な計算時間で導出することが可能となる。

## 1.3 関連研究

本研究の関連研究として、Xing らの研究 [2] が挙げられる。Xing らは、Weiss の分子場近似手法で用いられる条件式（平均場方程式）を拡張させたものとして、一般化平均場方程式 (generalized mean field equation) を導出した。本研究で導出した（部分空間への射影が満たす）条件式は、一般化平均場方程式と等価であることが示せる（付録 A 参照）。ただし、本論文では情報幾何学的なアプローチから条件式の導出を試みており、Xing らの導出方法とはまったく異なる方向から導き出している。さらに、本研究ではボルツマンマシンの平衡確率分布がなす分布族を多様体と見做し、この多様体が満たす幾何学的性質から条件式の導出を行っている。従って、この多様体と同じ幾何学的性質を備える他のモデルに対しても、本論文での議論はそのまま成立する。等価な幾何学的性質を持つモデルの例として、ボルツマンマシンを量子系へと拡張させたモデルなどが挙げられる [18]。

## 1.4 本論文の構成

この節では、本論文の構成について述べる。まず、第 2 章ではボルツマンマシンに関する基礎的な事項について説明を行う。第 3 章では、ボルツマンマシンの平衡確率分布を要素とする分布族を考えたときに、どのような性質が成り立つのかについて説明を行い、ボルツマンマシンの学習の収束先を直感的にわかりやすい形で与える。第 4 章では、文献 [12], [13], [16] の内容を参考に、平均場近似の情報幾何学的解釈について解説を行う。第 5 章では、格子模型型のボルツマンマシンの平衡確率分布から部分空間（クラスター型模型上で定められる平衡確率分布を要素とする空間）への射影が満たす条件式の導出を行う。第 6 章では、ピリーフプロパゲーションの手法を解説した後に、本論文で提案したアルゴリズムについて述べる。

第7章では、提案アルゴリズムに関するシミュレーションの結果と考察について述べる。最後に、第8章では結論と今後の課題について述べる。

## 第2章 ボルツマンマシン

ボルツマンマシン [1] とは，相互結合型ニューラルネットとして知られるホップフィールドネットワークの決定論的な状態変化規則を，温度パラメータと呼ばれる正の実数値を用いて確率論的な状態変化規則へと拡張させたモデルである．ボルツマンマシンの状態変化は一つ前の時刻の状態にしか依存しないため，マルコフ連鎖をなすことが分かる．さらに，このマルコフ連鎖は既約かつ非周期的という性質を満たすため，十分に状態変化を繰り返すことで，ボルツマンマシンの状態が従う確率分布はある一定の確率分布（平衡確率分布）へと収束する [9] ．

本章では，ボルツマンマシンの前身であるホップフィールドネットワークについて述べた後に，ホップフィールドネットワークからボルツマンマシンへの拡張，ボルツマンマシンの学習とその問題点について順に説明していく．

### 2.1 ホップフィールドネットワーク

#### 2.1.1 モデル

ホップフィールドネットワークとは，二種の状態を取りうる複数の素子から構成されるネットワークであり，後述する決定論的な状態変化規則に従って状態を離散時間的に変化させていくモデルである．以降では， $n$  個の素子から構成されるホップフィールドネットワークについて考える．まず，各素子を一意に指定するために，各素子に 1 から  $n$  の番号を割り当てる．また，素子  $i$  ( $1 \leq i \leq n$ ) の状態を表す変数を  $x_i \in \{0, 1\}$ ，しきい値を表す変数を  $h_i \in \mathbb{R}$ ，素子  $i$  から  $j$  への結合の度合いを表す変数（結合定数）を  $w_{i,j} \in \mathbb{R}$  で表すものとする．さらに，ホップフィールドネットワークの結合定数は対称でかつ自己結合がないという性質，すなわち

$$\forall i, j (i \neq j), \quad w_{i,j} = w_{j,i}, \quad \forall i, \quad w_{i,i} = 0 \quad (2.1)$$

が成り立つ．以下の図 2.1 は，素子数  $n = 3$  の場合のホップフィールドネットワークである．

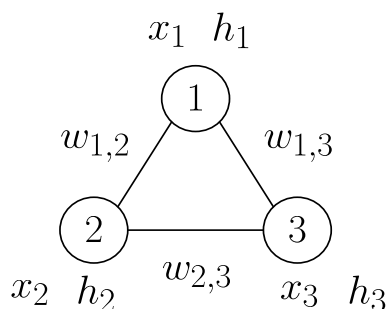


図 2.1: 3 素子のホップフィールドネットワーク



## 2.1.2 状態変化規則

各素子の状態変数を成分として持つ組を

$$x = (x_1, \dots, x_n) \quad (2.2)$$

と定める．以降の議論では， $x$  のことをネットワークの状態と呼ぶ．ネットワークの状態が  $x = (x_1, \dots, x_n)$  であるとき，素子  $i$  を除いた他の素子  $j (\neq i)$  から素子  $i$  への入力関数を以下のように定める．

$$I_i(x) := h_i + \sum_{j=1}^n w_{j,i} x_j \quad (h_i \text{は素子 } i \text{ のしきい値}) \quad (2.3)$$

このとき，ホップフィールドネットワークは以下の状態変化規則に従って，ネットワークの状態を離散時間的に変化させていく．

1. 素子  $i$  をランダムに選ぶ．
2. 以下の場合分けに従って，素子  $i$  の状態  $x_i$  を  $x'_i$  へと更新する．

$$x'_i = \begin{cases} 1 & \text{if } I_i(x) > 0 \\ 0 & \text{if } I_i(x) < 0 \\ x_i & \text{if } I_i(x) = 0 \end{cases} \quad (2.4)$$

ただし，他の素子  $j (\neq i)$  の状態は変化させない．

上記規則に従って状態変化を十分に繰り返すと，ネットワークの状態はエネルギー関数

$$U(x) := - \left( \sum_{i=1}^n h_i x_i + \sum_{i,j} w_{i,j} x_i x_j \right) \quad (2.5)$$

の極小値点に対応する状態へと収束する．この性質から，ホップフィールドネットワークは巡回セールスマン問題などの組み合わせ最適化問題を解くのに利用可能である．ただし，状態変化を十分に繰り返した時の収束先は，初期のネットワークの状態に依存した局所最小値点 (local minimum) である．

## 2.2 ボルツマンマシンへの拡張

## 2.2.1 状態変化規則

ボルツマンマシンとは，ホップフィールドネットワークの決定論的な状態変化規則を確率と温度の概念を用いて，以下のように拡張させたモデルである．

1. 素子  $i$  をランダムに選ぶ．
2. 確率

$$\Pr\{x'_i := 1\} = \frac{1}{1 + \exp\left[-\frac{I_i(x)}{T}\right]} \quad (2.6)$$

に従って，素子  $i$  の次の時刻の状態  $x'_i$  を 1 にする．

ただし，他の素子  $j (\neq i)$  の状態は変化させない．

(2.6) 式中の  $T$  は温度パラメータと呼ばれる正の実数値であり，ボルツマンマシンの状態変化を制御する役割を担っている．実際，温度パラメータを  $T \rightarrow \infty$  としたときの確率分布  $\Pr\{x'_i := 1\}$ ,  $\Pr\{x'_i := 0\}$  はそれぞれ

$$\Pr\{x'_i := 1\} = \Pr\{x'_i := 0\} = \frac{1}{2} \quad (2.7)$$

となる．このとき，素子  $i$  の状態は入力関数  $I_i(x)$  の値に依らずにランダムに変化する．また， $T \rightarrow 0$  としたとき，

$$\begin{cases} \Pr\{x'_i := 1\} = 1, & \Pr\{x'_i := 0\} = 0 & \text{if } I_i(x) > 0 \\ \Pr\{x'_i := 1\} = 0, & \Pr\{x'_i := 0\} = 1 & \text{if } I_i(x) < 0 \end{cases} \quad (2.8)$$

となり，素子  $i$  の状態は入力関数  $I_i(x)$  の正負に依存して決定論的に変化する．(2.8) 式より， $T \rightarrow 0$  としたときの状態変化規則はホップフィールドネットワークのものと同値となることから，ボルツマンマシンの状態変化規則はホップフィールドネットワークの状態変化規則の拡張となっていることが分かる．

組み合わせ最適化問題を解く上で，ホップフィールドネットワークはネットワークの初期状態に依存した局所最適解 (local optimum) しか得られなかったのに対し，ボルツマンマシンは，温度パラメータを高温から徐々に低温へと下げつつネットワークの状態を変化させていくアルゴリズム（焼きなまし法，もしくはシミュレーテッド・アニーリングともいう）を用いることで，理論的にはネットワークの初期状態に依らない大域的最適解を得ることが可能である [9]．しかし，理論通り大域的最適解を得るためには，温度パラメータを高温から低温へと徐々に下げる過程において，1 ステップごとに下げる温度を微小な値に設定する必要がある．従って，所望の解を得るまでに膨大な時間が必要となり，必ずしも実用的とはいえない．ゆえに，理論より早く温度を下げて近似解を得ることもある．

### 2.2.2 ボルツマンマシンの平衡確率分布

ボルツマンマシンの状態変化規則 (2.6) より，各時刻のネットワークの状態は一つ前の時刻の状態にのみ依存している．従って，ボルツマンマシンのネットワークの状態を変化させていく過程はマルコフ連鎖をなしていると解釈される．さらにこのマルコフ連鎖に関して，いくつかの性質（既約性，非周期性）が成立するため，十分に状態変化を繰り返すことでネットワークの状態が従う分布は以下の平衡確率分布  $P_\theta(x)$  へと収束する．

$$P_\theta(x) = \exp \left[ \sum_{i=1}^n h_i x_i + \sum_{1 \leq j < k \leq n} w_{j,k} x_j x_k - \Psi(\theta) \right] \quad (2.9)$$

ただし， $\Psi(\theta)$  は

$$\Psi(\theta) = \log \left( \sum_{x \in \{0,1\}^n} \exp \left[ \sum_{i=1}^n h_i x_i + \sum_{1 \leq j < k \leq n} w_{j,k} x_j x_k \right] \right) \quad (2.10)$$

である．また，結合定数に関する性質 (2.1) より，以下の等式が成立する．

$$\sum_{1 \leq j < k \leq n} w_{j,k} x_j x_k = \frac{1}{2} \sum_{j,k=1}^n w_{j,k} x_j x_k \quad (2.11)$$

従って，(2.9) 式は以下のようにも表現される．

$$P_\theta(x) = \exp \left[ \sum_{i=1}^n h_i x_i + \frac{1}{2} \sum_{j,k=1}^n w_{j,k} x_j x_k - \Psi(\theta) \right] \quad (2.12)$$

## 2.3 ボルツマンマシンの学習

### 2.3.1 学習の目的

$\{0, 1\}^n$  上の未知の確率分布に従って生成される  $N$  個の観測データ

$$x^{(1)} = (x_1^{(1)}, \dots, x_n^{(1)}), \dots, x^{(N)} = (x_1^{(N)}, \dots, x_n^{(N)}) \quad (2.13)$$

が与えられているとする, このとき, 観測データに関する経験分布  $q(x)$  は以下のように記述される.

$$q(x) = \frac{1}{N} \sum_{k=1}^N \delta(x, x^{(k)}), \quad \delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \quad (2.14)$$

ボルツマンマシンの学習の目的は, この経験分布  $q(x)$  を最もよく近似する平衡確率分布  $P_\theta(x)$  を導出することで, 未知の分布の推定を行うことである.

本論文では, 確率分布間の近似の指標として, Kullback-Leibler ダイバージェンス

$$D(q||P_\theta) := \sum_{x \in \{0, 1\}^n} q(x) \log \frac{q(x)}{P_\theta(x)} \quad (2.15)$$

を用いる. Kullback-Leibler ダイバージェンス  $D(\cdot||\cdot)$  は距離関数が満たすべき性質のうち, 非負性を満たす. ただし, 対称性や三角不等式は満たさないため, Kullback-Leibler ダイバージェンスは確率分布間の擬似的な距離関数としてみなされる.  $D(q||P_\theta)$  の値が最小となるようなパラメータ  $\theta$  を導出することが, 次の節で述べるボルツマンマシンの学習アルゴリズムのモチベーションとなっている.

### 2.3.2 学習アルゴリズム

Hinton ら [1] によって提案されたボルツマンマシンの学習アルゴリズムを以下に示す.

[ボルツマンマシンの学習アルゴリズム]

Initialize:  $t = 0, \theta^{(0)} = (h_i^{(0)}, w_{i,j}^{(0)})_{1 \leq i, j \leq n}$

step.1 以下の式に従って  $\frac{\partial D(q||P_{\theta^{(t)}})}{\partial h_i}, \frac{\partial D(q||P_{\theta^{(t)}})}{\partial w_{i,j}}$  を計算する.

$$\frac{\partial D(q||P_{\theta^{(t)}})}{\partial h_i} = E_{P_{\theta^{(t)}}}[x_i] - E_q[x_i] \quad (2.16)$$

$$\frac{\partial D(q||P_{\theta^{(t)}})}{\partial w_{i,j}} = E_{P_{\theta^{(t)}}}[x_i x_j] - E_q[x_i x_j] \quad (2.17)$$

step.2 パラメータ  $\theta^{(t)} = (h_i^{(t)}, w_{i,j}^{(t)})_{1 \leq i, j \leq n}$  を以下の式に従って更新する.

$$h_i^{(t+1)} = h_i^{(t)} - \alpha \frac{\partial D(q||P_{\theta^{(t)}})}{\partial h_i} \quad (2.18)$$

$$w_{i,j}^{(t+1)} = w_{i,j}^{(t)} - \alpha \frac{\partial D(q||P_{\theta^{(t)}})}{\partial w_{i,j}} \quad (2.19)$$

step.3  $\theta^{(t+1)} \simeq \theta^{(t)}$  ならば, アルゴリズムを終了する. でなければ,  $t = t + 1$  として,

step.1 へ戻る

アルゴリズム内のパラメータ  $\alpha$  は学習率と呼ばれる十分小さな正数であり,  $P_{\theta^{(t)}}, h_i^{(t)}, w_{i,j}^{(t)}$  はそれぞれ  $t$  回パラメータ更新した後の平衡確率分布, 素子  $i$  のしきい値, 素子  $i, j$  の間の結合定数である. また,  $E_{P_{\theta^{(t)}}}[\cdot], E_q[\cdot]$  はそれぞれ平衡確率分布  $P_{\theta^{(t)}}(x)$  と経験分布  $q(x)$  に関する期待値を表している.

### 2.3.3 学習アルゴリズムの問題点

前節のボルツマンマシンの学習アルゴリズムより、各しきい値、結合定数を 1 回更新するためには (2.16) 式と (2.17) 式中の平衡確率分布  $P_{\theta^{(t)}}(x)$  に関する期待値計算を行わなくてはならない。ボルツマンマシンの素子数が  $n$  個の場合、この期待値計算は  $2^n$  個の項の足し算となる。従って、この学習アルゴリズムの計算量オーダーは  $O(2^n)$  であり、素子数  $n$  が大きい場合には計算量的に困難となる。

## 第3章 ボルツマンマシンの情報幾何

ボルツマンマシンの平衡確率分布がなす集合を考えたとき，この集合は指数数型分布族となる．情報幾何学では，指数型分布族に関していくつかの重要な性質が成り立つことが知られており [10]，ボルツマンマシンを通常とは異なった視点から考察するのに役立つ．この章では，指数型分布族よりもさらに広い概念である統計多様体から議論を始める．次に，指数型分布族の定義と一般的な性質を述べた後に，2章で取り上げたボルツマンマシンの学習アルゴリズムの収束先を直感的に理解しやすい形で与える．

### 3.1 統計多様体

有限集合  $\mathcal{X} = \{1, \dots, m\}$  上の正の確率分布を要素とする集合を以下のように記述する．

$$\mathcal{P}(\mathcal{X}) = \left\{ p : \mathcal{X} \rightarrow \mathbb{R} \mid \forall x \in \mathcal{X}, p(x) > 0, \sum_{x \in \mathcal{X}} p(x) = 1 \right\} \quad (3.1)$$

次に， $\mathcal{P}(\mathcal{X})$  の部分集合であり， $n (< m)$  次元パラメータ  $\xi = (\xi^1, \dots, \xi^n)$  によって要素が一意に指定される分布族

$$\mathcal{S} = \{ p_\xi \in \mathcal{P}(\mathcal{X}) \mid \xi = (\xi^1, \dots, \xi^n) \in \Xi \} \subset \mathcal{P}(\mathcal{X}), \quad \Xi : \mathbb{R}^n \text{ の開集合} \quad (3.2)$$

を考える．任意の  $x \in \mathcal{X}$  に対して，関数  $\xi \mapsto p_\xi$  が一対一対応かつ  $C^\infty$  級であるとき， $\mathcal{S}$  を  $n$  次元統計多様体と呼ぶ．ここで取り上げた  $\xi = (\xi^1, \dots, \xi^n)$  のように，与えられた分布族の中から要素を一意に指定するパラメータのことを以降では座標系と呼ぶ．さらに，統計多様体の任意の要素  $p_\xi \in \mathcal{S}$  に関して，以下の条件式が成り立つ．

$$\forall i (1 \leq i \leq n), \quad \frac{\partial p_\xi(x)}{\partial \xi^i} = p_\xi(x) \left( \frac{\partial \log p_\xi(x)}{\partial \xi^i} \right) \quad (3.3)$$

$$\forall i (1 \leq i \leq n), \quad E_{p_\xi} \left[ \frac{\partial \log p_\xi(x)}{\partial \xi^i} \right] = 0 \quad (3.4)$$

なお，(3.4) 式は以下のようにして確かめられる．まず， $\forall p_\xi \in \mathcal{P}(\mathcal{X})$  に関して以下の等式が成立する．

$$\sum_{x \in \mathcal{X}} p_\xi(x) = 1 \quad (3.5)$$

(3.5) 式の両辺を変数  $\xi^i (1 \leq i \leq n)$  で偏微分すると，

$$\frac{\partial}{\partial \xi^i} \sum_{x \in \mathcal{X}} p_\xi(x) = 0 \quad (3.6)$$

となる．さらに (3.6) 式を同値変形すると

$$\begin{aligned}
 \frac{\partial}{\partial \xi^i} \sum_{x \in \mathcal{X}} p_\xi(x) &= 0 \\
 \iff \sum_{x \in \mathcal{X}} \frac{\partial p_\xi(x)}{\partial \xi^i} &= 0 \\
 \iff \sum_{x \in \mathcal{X}} p_\xi(x) \frac{\partial \log p_\xi(x)}{\partial \xi^i} &= 0 \\
 \iff E_{p_\xi} \left[ \frac{\partial \log p_\xi(x)}{\partial \xi^i} \right] &= 0
 \end{aligned}$$

となり，条件式 (3.4) が導出される．

統計多様体の例としては  $\mathcal{P}(\mathcal{X})$  が挙げられる． $\mathcal{P}(\mathcal{X})$  に含まれる分布は， $\mathcal{X}$  に含まれる  $(m-1)$  個の各要素の確率を与えることで一意に指定される．従って，以下のように座標系を取ることで， $\mathcal{P}(\mathcal{X})$  は  $(m-1)$  次元統計多様体となることが分かる．

$$\begin{aligned}
 \mathcal{P}(\mathcal{X}) &= \{p : \mathcal{X} \rightarrow \mathbb{R} \mid \forall x \in \mathcal{X}, p(x) > 0, \sum_{x \in \mathcal{X}} p(x) = 1\} \\
 &= \{p_\xi : \mathcal{X} \rightarrow \mathbb{R} \mid \xi = (\xi^1, \dots, \xi^{m-1}) \in \Xi\}
 \end{aligned} \tag{3.7}$$

ただし，パラメータ空間  $\Xi$  と座標系  $\xi = (\xi^1, \dots, \xi^{m-1})$  は以下のようにして定まっているとする．

$$\forall i \in \mathcal{X}, \xi^i = p(i) \tag{3.8}$$

$$\Xi = \left\{ (\xi^1, \dots, \xi^{m-1}) \in \mathbb{R}^{m-1} \mid \forall i, \xi^i > 0, \sum_{i=1}^{m-1} \xi^i < 1 \right\} \tag{3.9}$$

## 3.2 Fisher 情報行列

$n$  次元統計多様体  $\mathcal{S}$  に含まれる分布  $P_\theta \in \mathcal{S}$  が与えられているとする．このとき，第  $i$  行  $j$  列成分が以下の式で定まっている行列  $G(\theta) \in \mathbb{R}^{n \times n}$  を分布  $P_\theta \in \mathcal{S}$  に関する Fisher 情報行列と呼ぶ．

$$g_{i,j}(\theta) := E_{P_\theta} \left[ \left( \frac{\partial \log P_\theta(x)}{\partial \theta^i} \right) \left( \frac{\partial \log P_\theta(x)}{\partial \theta^j} \right) \right] \tag{3.10}$$

一般的に，Fisher 情報行列  $G(\theta)$  は正定値対称行列となる．

## 3.3 指数型分布族

指数型分布族の定義は以下の通りである．

**定義 1** (指数型分布族)

$n$  次元統計多様

$$\mathcal{S} = \{p_\theta : \mathcal{X} \rightarrow \mathbb{R} \mid \theta = (\theta^1, \dots, \theta^n) \in \Theta\}$$

が与えられているとする．このとき， $\mathcal{X}$  上の関数  $C(x), F_1(x), \dots, F_n(x)$  と  $\Theta$  上の関数  $\Psi(\theta)$  を用いて，任意の分布  $P_\theta \in \mathcal{S}$  が以下の形式

$$P_\theta(x) = \exp \left[ C(x) + \sum_{\alpha=1}^n \theta^\alpha F_\alpha(x) - \Psi(\theta) \right] \quad (3.11)$$

で与えられるとき， $\mathcal{S}$  を  $n$  次元指数型分布族という．ただし，(3.11) 式中の  $\Psi(\theta)$  は以下のように記述されるものとする．

$$\Psi(\theta) = \log \left( \sum_{x \in \mathcal{X}} \exp \left[ C(x) + \sum_{\alpha=1}^n \theta^\alpha F_\alpha(x) \right] \right) \quad (3.12)$$

また，このときの座標系  $\theta = (\theta^1, \dots, \theta^n)$  のことを自然座標系（または自然パラメータ）と呼ぶ．

次に指数型分布族の具体例をいくつか挙げていく．

#### 例 1 (ボルツマンマシン)

ボルツマンマシンの平衡確率分布 (2.9) がなす集合を  $\mathcal{M}$  とする．さらに，集合  $I$  ( $|I| = \frac{n(n+1)}{2}$ )，自然座標系  $\theta = (\theta^\alpha)_{\alpha \in I}$ ， $\mathcal{X} = \{0, 1\}^n$  上の関数  $C(x), F_\alpha(x)$  ( $\alpha \in I$ ) をそれぞれ以下のように定める．

$$\forall x \in \mathcal{X}, C(x) = 0, \quad I := \{i \mid 1 \leq i \leq n\} \cup \{(j, k) \mid 1 \leq j < k \leq n\} \quad (3.13)$$

$$\forall \alpha \in I, \quad \theta^\alpha = \begin{cases} h_i & \text{if } \alpha = i \\ w_{j,k} & \text{if } \alpha = (j, k) \end{cases} \quad F_\alpha = \begin{cases} x_i & \text{if } \alpha = i \\ x_j x_k & \text{if } \alpha = (j, k) \end{cases} \quad (3.14)$$

このとき，ボルツマンマシンの平衡確率分布 (2.9) は次のように式変形される．

$$\begin{aligned} P_\theta(x) &= \exp \left[ \sum_{i=1}^n h_i x_i + \sum_{1 \leq j < k \leq n} w_{j,k} x_j x_k - \Psi(\theta) \right] \\ &= \exp \left[ C(x) + \sum_{\alpha \in I} \theta^\alpha F_\alpha(x) - \Psi(\theta) \right] \end{aligned}$$

従って， $\mathcal{M}$  は  $\frac{n(n+1)}{2}$  次元指数型分布族となる．

#### 例 2 (有限集合上の正の確率分布全体)

有限集合  $\mathcal{X} = \{1, \dots, m\}$  上の正の確率分布を要素とする集合  $\mathcal{P}(\mathcal{X})$  についても指数型分布族となる．実際，自然座標系  $\theta = (\theta^\alpha)_{\alpha \in \mathcal{X}}$  と  $\mathcal{X}$  上の関数  $C(x), F_\alpha(x)$  ( $\alpha \in \mathcal{X}$ ) をそれぞれ以下のように定める．

$$\forall x \in \mathcal{X}, C(x) = 0 \quad (3.15)$$

$$\forall \alpha (1 \leq \alpha \leq m-1), \quad \theta^\alpha = \log \frac{p(\alpha)}{1 - \sum_{i=1}^{m-1} p(i)} = \log \frac{p(\alpha)}{p(m)} \quad (3.16)$$

$$F_\alpha(x) = \begin{cases} 1 & \text{if } \alpha = x \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

このとき，関数  $\Psi(\theta)$  は以下のように記述される．

$$\Psi(\theta) = \log \left( \sum_{x \in \mathcal{X}} \exp \left[ C(x) + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x) \right] \right) \quad (3.18)$$

(3.15) ~ (3.18) 式を用いると, 任意の  $p \in \mathcal{P}(\mathcal{X})$  は次のように式変形される .

$$\begin{aligned}
 p(x) &= \frac{\frac{p(x)}{p(m)}}{\frac{1}{p(m)}} = \frac{\frac{p(x)}{p(m)}}{\sum_{x' \in \mathcal{X}} \frac{p(x')}{p(m)}} = \frac{\exp \left[ \log \frac{p(x)}{p(m)} \right]}{\sum_{x' \in \mathcal{X}} \exp \left[ \log \frac{p(x')}{p(m)} \right]} \\
 &= \frac{\exp \left[ C(x) + \log \frac{p(x)}{p(m)} \right]}{\sum_{x' \in \mathcal{X}} \exp \left[ C(x') + \log \frac{p(x')}{p(m)} \right]} \\
 &= \frac{\exp \left[ C(x) + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x) \right]}{\sum_{x' \in \mathcal{X}} \exp \left[ C(x') + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x') \right]} \\
 &= \frac{\exp \left[ C(x) + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x) \right]}{\exp \left[ \log \left( \sum_{x' \in \mathcal{X}} \exp \left[ C(x') + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x') \right] \right) \right]} \\
 &= \frac{\exp \left[ C(x) + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x) \right]}{\exp [\Psi(\theta)]} \\
 &= \exp \left[ C(x) + \sum_{\alpha \in \mathcal{X}} \theta^\alpha F_\alpha(x) - \Psi(\theta) \right] \tag{3.19}
 \end{aligned}$$

従って,  $\mathcal{P}(\mathcal{X})$  は  $(m-1)$  次元指数型分布族となる .

$n$  次元指数型分布族  $\mathcal{S}$  に含まれる分布は, 自然パラメータ  $\theta = (\theta^i)_{1 \leq i \leq n}$  の値を決めることで一意に指定することができる . さらに, 次のようにして定まるパラメータ  $\eta(\theta) = (\eta_i(\theta))_{1 \leq i \leq n}$  でも分布を一意に指定することができる .

$$\eta_i(\theta) := E_{P_\theta}[F_i(x)] \tag{3.20}$$

このパラメータ  $\eta(\theta) = (\eta_i(\theta))_{1 \leq i \leq n}$  のことを期待値座標系 (もしくは期待値パラメータ) と呼ぶ . さらに, 期待値座標系  $\eta(\theta) = (\eta_i(\theta))_{1 \leq i \leq n}$  に関する  $P_\theta \in \mathcal{S}$  の Fisher 情報行列  $G(\eta) \in \mathbb{R}^{n \times n}$  は次のようになる .

$$G(\eta) := [g^{i,j}(\eta)] \tag{3.21}$$

$$g^{i,j}(\eta) := E_{P_\theta} \left[ \left( \frac{\partial \log P_\theta(x)}{\partial \eta_i} \right) \left( \frac{\partial \log P_\theta(x)}{\partial \eta_j} \right) \right] \tag{3.22}$$



以下に  $n$  次元指数型分布族  $S$  が満たす重要な性質をまとめておく [11] .

- (1)  $\eta_i(\theta) = \frac{\partial \Psi(\theta)}{\partial \theta^i}$
- (2)  $g_{i,j}(\theta) = \frac{\partial \eta_j(\theta)}{\partial \theta^i} = \frac{\partial \eta_i(\theta)}{\partial \theta^j} = \frac{\partial^2 \Psi(\theta)}{\partial \theta^i \partial \theta^j}$
- (3)  $G(\theta) = (G(\eta))^{-1}$
- (4)  $g^{i,j}(\eta) = \frac{\partial \theta^j(\eta)}{\partial \eta_i} = \frac{\partial \theta^i(\eta)}{\partial \eta_j}$
- (5)  $\forall P_\theta \in S$  に対し, ルジャンドル変換により  $\Phi(\eta)$  を

$$\Phi(\eta) := \max_{\theta} \left\{ \sum_{1 \leq i \leq n} \theta^i \eta_i - \Psi(\theta) \right\}$$

と定義すれば  $\theta^i = \frac{\partial \Phi(\eta)}{\partial \eta^i}$ ,  $g^{i,j}(\eta) = \frac{\partial^2 \Phi(\eta)}{\partial \eta_i \partial \eta_j}$  が成立する .

### 3.4 $e$ -射影, $m$ -射影

以下のような統計多様体  $S, \mathcal{M}$  が与えられているとする .

$$S = \{P_\theta : \mathcal{X} \rightarrow \mathbb{R} \mid \theta = (\theta^1, \dots, \theta^n)\} : n \text{ 次元統計多様体} \quad (3.23)$$

$$\mathcal{M} = \{q_{\bar{\theta}} : \mathcal{X} \rightarrow \mathbb{R} \mid \bar{\theta} = (\bar{\theta}^1, \dots, \bar{\theta}^m)\} \subset S : m \text{ 次元統計多様体 } (m \leq n) \quad (3.24)$$

このとき,  $P_\theta \in S$  から  $\mathcal{M}$  への  $e$ -射影  $q_{\bar{\theta}} \in \mathcal{M}$  は以下の条件式を満たす .

$$\forall_i (1 \leq i \leq m), \frac{\partial D(q_{\bar{\theta}} \| P_\theta)}{\partial \bar{\theta}^i} = 0 \quad (3.25)$$

また,  $P_\theta \in S$  から  $\mathcal{M}$  への  $m$ -射影  $q_{\bar{\theta}} \in \mathcal{M}$  は以下の条件式を満たす .

$$\forall_i (1 \leq i \leq m), \frac{\partial D(P_\theta \| q_{\bar{\theta}})}{\partial \bar{\theta}^i} = 0 \quad (3.26)$$

一般的に,  $e$ -射影,  $m$ -射影の条件を満たす分布は複数存在するため, それぞれの射影点は一意的に定まらない . ただし,  $S, \mathcal{M}$  が指数型分布族であるとき,  $m$ -射影点  $q_{\bar{\theta}} \in \mathcal{M}$  は一意に定まり, 次の等式を満たすことが知られている [10] .

$$q_{\bar{\theta}} = \min_{r \in \mathcal{M}} D(P_\theta \| r) \quad (3.27)$$

一方で,  $e$ -射影の一意性は保障されないままである .

ここで,  $S, \mathcal{M}$  が指数型分布族であるとし,  $\forall P_\theta \in S, \forall q_{\bar{\theta}} \in \mathcal{M}$  は以下のように記述されるものとする .

$$P_\theta(x) = \exp \left[ \sum_{\alpha=1}^n \theta^\alpha F_\alpha(x) - \Psi(\theta) \right] \quad (3.28)$$

$$q_{\bar{\theta}}(x) = \exp \left[ \sum_{\beta=1}^m \bar{\theta}^\beta F_\beta(x) - \Phi(\bar{\theta}) \right] \quad (3.29)$$

このとき， $m$ -射影の条件式 (3.26) は以下のように同値変形される．

$$\begin{aligned}
 \frac{\partial D(P_\theta || q_{\bar{\theta}})}{\partial \bar{\theta}^i} &= 0 \\
 \iff \frac{\partial}{\partial \bar{\theta}^i} \sum_{x \in \mathcal{X}} P_\theta(x) \{ \log P_\theta(x) - \log q_{\bar{\theta}}(x) \} &= 0 \\
 \iff - \sum_{x \in \mathcal{X}} P_\theta(x) \frac{\partial \log q_{\bar{\theta}}}{\partial \bar{\theta}^i} &= 0 \\
 \iff - \sum_{x \in \mathcal{X}} P_\theta(x) (F_i(x) - \bar{\eta}_i(\bar{\theta})) &= 0 \\
 \iff \bar{\eta}_i(\bar{\theta}) = \eta_i(\theta) \quad (1 \leq i \leq m) & \quad (3.30)
 \end{aligned}$$

条件式 (3.30) の両辺の期待値パラメータ  $\bar{\eta}_i(\bar{\theta})$ ,  $\eta_i(\theta)$  はそれぞれ分布  $q_{\bar{\theta}} \in \mathcal{M}$ ,  $P_\theta \in \mathcal{S}$  に関する期待値パラメータを意味している．従って  $m$ -射影とは，射影させる前の分布  $P_\theta \in \mathcal{S}$  が持ついくつかの期待値パラメータの値と一致するような  $\mathcal{M}$  上の点へと移す射影であると解釈することができる．

### 3.5 ボルツマンマシンの学習と $m$ -射影について

2章で挙げたボルツマンマシンの学習アルゴリズムの終了条件は，平衡確率分布  $P_{\theta^{(t)}} \in \mathcal{M}$  に関する期待値  $E_{P_{\theta^{(t)}}}[x_i]$ ,  $E_{P_{\theta^{(t)}}}[x_j x_k]$  と経験分布  $q \in \mathcal{P}(\mathcal{X})$  に関する期待値  $E_q[x_i]$ ,  $E_q[x_j x_k]$  に関して，以下が成立するとき満たされる．

$$\forall i (1 \leq i \leq n), E_{P_{\theta^{(t)}}}[x_i] \simeq E_q[x_i] \quad (3.31)$$

$$\forall j, k (1 \leq j < k \leq n), E_{P_{\theta^{(t)}}}[x_j x_k] \simeq E_q[x_j x_k] \quad (3.32)$$

$P_{\theta^{(t)}} \in \mathcal{M}$ ,  $q \in \mathcal{P}(\mathcal{X})$  に関する期待値パラメータをそれぞれ  $\eta(\theta^{(t)})$ ,  $\eta(q)$  としたとき，条件式 (3.31), (3.32) はそれぞれ以下のように記述し直せる．

$$\forall i (1 \leq i \leq n), \eta_i(\theta^{(t)}) \simeq \eta_i(q) \quad (3.33)$$

$$\forall j, k (1 \leq j < k \leq n), \eta_{j,k}(\theta^{(t)}) \simeq \eta_{j,k}(q) \quad (3.34)$$

さらに，経験分布  $q(x) \in \mathcal{P}(\mathcal{X})$  から指数型分布族  $\mathcal{M}$  への  $m$ -射影を  $P_{\hat{\theta}}(x) \in \mathcal{M}$ ，その期待値パラメータを  $\eta(\hat{\theta})$  としたとき，(3.30) 式より以下の同値関係が成り立つ．

$$\begin{aligned}
 \forall i (1 \leq i \leq n), \frac{\partial D(q || P_{\hat{\theta}})}{\partial h_i} = 0, \quad \forall j, k (1 \leq j < k \leq n), \frac{\partial D(q || P_{\hat{\theta}})}{\partial w_{j,k}} = 0 \\
 \iff \forall i (1 \leq i \leq n), \eta_i(\hat{\theta}) = \eta_i(q), \quad \forall j, k (1 \leq j < k \leq n), \eta_{j,k}(\hat{\theta}) = \eta_{j,k}(q) \quad (3.35)
 \end{aligned}$$

従って，2章で挙げたボルツマンマシンの学習アルゴリズムの収束先は，経験分布  $q \in \mathcal{P}(\mathcal{X})$  からボルツマンマシンの平衡確率分布がなす分布族  $\mathcal{M}$  への  $m$ -射影点  $P_{\hat{\theta}} \in \mathcal{M}$  である．さらに， $\mathcal{P}(\mathcal{X})$ ,  $\mathcal{M}$  は指数型分布族であるため， $m$ -射影点  $P_{\hat{\theta}} \in \mathcal{M}$  は一意に定まり，条件式

$$P_{\hat{\theta}} = \min_{r \in \mathcal{M}} D(q || r) \quad (3.36)$$

を満たす．ゆえに，ボルツマンマシンの学習アルゴリズムは大域的最適解を得るアルゴリズムとなっている．以下の図は，ボルツマンマシンの学習アルゴリズムにより，平衡確率分布  $P_{\theta^{(t)}} \in \mathcal{M}$  が  $m$ -射影点  $P_{\hat{\theta}} \in \mathcal{M}$  へと収束していく様子を表した図である．

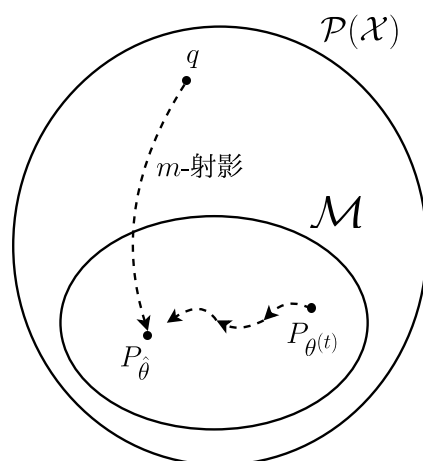


図 3.1: ボルツマンマシンの学習と  $m$ -射影

## 第4章 平均場近似の情報幾何

平均場近似 [6] とは、相互作用し合う素子から構成される巨視的な物理系に対し、各素子間の相互作用の強さをある種の期待値で置き換えることで、様々な物理量の導出を容易にする近似手法である。また、情報理論的に解釈すると、平均場近似とは Kullback-Leibler ダイバージェンスを近似の指標として、元の巨視的な物理系が従う確率分布を各素子の状態変数が統計的に独立となる場合の物理系が従う確率分布で近似する手法であると解釈することができる。

この章では、情報理論的な立場から平均場近似の仕組みについて述べた後に、平均場近似を情報幾何学の観点から考察する [11], [12], [13], [14]。さらに、平均場近似を用いたときのボルツマンマシンの学習アルゴリズムを具体的に記述した後に、学習アルゴリズムの挙動と収束先について幾何学的なイメージを用いながら解説する。

### 4.1 ボルツマンマシンにおける平均場近似

ボルツマンマシンの学習では、分布  $P_{\theta^{(t)}} \in \mathcal{M}$  に関する期待値パラメータ  $\eta_i(\theta^{(t)})$ ,  $\eta_{i,j}(\theta^{(t)})$  の導出部分が計算困難性の原因となっていた。ここで、期待値パラメータの導出が容易に行える分布からなる集合として、以下の分布族を考える。

$$\mathcal{M}_C = \{P_{\bar{h}} \mid \bar{h} = (\bar{h}_1, \dots, \bar{h}_n)\} \quad (4.1)$$

ただし、分布族  $\mathcal{M}_C$  に含まれる任意の要素  $P_{\bar{h}}$  は以下の形式で表現されるものとする。

$$P_{\bar{h}}(x) = \exp \left[ \sum_{i=1}^n \bar{h}_i x_i - \Phi(\bar{h}) \right] \quad (4.2)$$

$$\Phi(\bar{h}) = \log \sum_{x \in \mathcal{X}} \exp \left[ \sum_{i=1}^n \bar{h}_i x_i \right] \quad (4.3)$$

このとき、分布族  $\mathcal{M}_C$  は結合定数がすべて 0 となる（すなわち、素子間の結合が存在しない）ボルツマンマシンの平衡確率分布がなす分布族を表している。また、素子間の結合が存在しない場合、各素子の状態変数は明らかに統計的に独立となる。実際、素子  $i$  だけから構成されるボルツマンマシンの平衡確率分布を

$$P_{\bar{h}_i}(x_i) = \frac{\exp[\bar{h}_i x_i]}{\sum_{x_i \in \{0, 1\}} \exp[\bar{h}_i x_i]} \quad (4.4)$$

としたとき，任意の  $P_{\bar{h}}(x) \in \mathcal{M}_C$  は以下のように計算される．

$$\begin{aligned}
 P_{\bar{h}}(x) &= \exp \left[ \sum_{i=1}^n \bar{h}_i x_i - \Phi(\bar{h}) \right] \\
 &= \frac{\exp \left[ \sum_{i=1}^n \bar{h}_i x_i \right]}{\sum_{x \in \mathcal{X}} \exp \left[ \sum_{i=1}^n \bar{h}_i x_i \right]} \\
 &= \frac{\prod_{i=1}^n \exp[\bar{h}_i x_i]}{\sum_{x_1 \in \{0,1\}} \cdots \sum_{x_n \in \{0,1\}} \prod_{i=1}^n \exp[\bar{h}_i x_i]} \\
 &= \prod_{i=1}^n \frac{\exp[\bar{h}_i x_i]}{\sum_{x_i \in \{0,1\}} \exp[\bar{h}_i x_i]} \\
 &= \prod_{i=1}^n P_{\bar{h}_i}(x_i) \tag{4.5}
 \end{aligned}$$

上記計算結果より，素子間の結合が存在しない場合，各素子の状態変数は統計的に独立となる．ゆえに，任意の  $P_{\bar{h}} \in \mathcal{M}_C$  に関する期待値パラメータ  $\bar{\eta}(\bar{h}) = (\bar{\eta}_1(\bar{h}), \dots, \bar{\eta}_n(\bar{h}))$  は以下の式に従って容易に導出される．

$$\bar{\eta}_i(\bar{h}) = \sum_{x_i \in \{0,1\}} x_i P_{\bar{h}_i}(x_i) \tag{4.6}$$

さらに，任意の素子  $i, j$  の状態に関する期待値  $E_{P_{\bar{h}}}[x_i x_j]$  は以下の等式を満たす．

$$E_{P_{\bar{h}}}[x_i x_j] = \bar{\eta}_i(\bar{h}) \bar{\eta}_j(\bar{h}) \tag{4.7}$$

ボルツマンマシンにおける平均場近似とは，元の平衡確率分布  $P_\theta \in \mathcal{M}$  に対して，分布族  $\mathcal{M}_C$  の要素の中で Kullback-Leibler ダイバージェンス  $D(\cdot \| P_\theta)$  の値が最小となる分布  $P_{\bar{h}} \in \mathcal{M}_C$  で近似する手法である．また，平均場近似により導出される近似分布  $P_{\bar{h}} \in \mathcal{M}_C$  の期待値パラメータ  $\bar{\eta}(\bar{h}) = (\bar{\eta}_1(\bar{h}), \dots, \bar{\eta}_n(\bar{h}))$  は以下の非線形連立方程式の解であることが知られている [15]．

$$\begin{cases} \bar{\eta}_1(\bar{h}) = \frac{1}{1 + \exp \left[ -h_1 - \sum_{j=1}^n w_{1,j} \bar{\eta}_j(\bar{h}) \right]} \\ \vdots \\ \bar{\eta}_n(\bar{h}) = \frac{1}{1 + \exp \left[ -h_n - \sum_{j=1}^n w_{n,j} \bar{\eta}_j(\bar{h}) \right]} \end{cases} \tag{4.8}$$

一般的に，連立方程式 (4.8) 式は自己無撞着方程式と呼ばれている．

## 4.2 平均場近似の情報幾何学的解釈

$P_\theta \in \mathcal{M}$  から  $\mathcal{M}_C$  への  $e$ -射影  $q_{\bar{h}} \in \mathcal{M}_C$  を考える．このとき， $e$ -射影  $q_{\bar{h}} \in \mathcal{M}_C$  は以下の条件式を満たす．

$$\forall_i (1 \leq i \leq n), \frac{\partial}{\partial \bar{h}_i} D(q_{\bar{h}} \| P_\theta) = 0 \tag{4.9}$$

(4.9) 式の左辺について式変形を施すと、以下のように計算される。

$$\begin{aligned}
 \frac{\partial}{\partial \bar{h}_i} D(q_{\bar{h}} \| P_{\theta}) &= \frac{\partial}{\partial \bar{h}_i} \sum_{x \in \mathcal{X}} q_{\bar{h}}(x) \{ \log q_{\bar{h}}(x) - \log P_{\theta}(x) \} \\
 &= \frac{\partial}{\partial \bar{h}_i} \sum_{x \in \mathcal{X}} q_{\bar{h}}(x) \left\{ \sum_{k=1}^n (\bar{h}_k - h_k) x_k - \frac{1}{2} \sum_{\ell, m=1}^n w_{\ell, m} x_{\ell} x_m - \Phi(\bar{h}) + \Psi(\theta) \right\} \\
 &= \sum_{k=1}^n \left\{ \frac{\partial}{\partial \bar{h}_i} (\bar{h}_k - h_k) \right\} \bar{\eta}_k(\bar{h}) + \sum_{k=1}^n (\bar{h}_k - h_k) \left( \frac{\partial \bar{\eta}_k(\bar{h})}{\partial \bar{h}_i} \right) \\
 &\quad - \frac{1}{2} \sum_{\ell, m=1}^n w_{\ell, m} \left( \frac{\partial \bar{\eta}_{\ell}(\bar{h})}{\partial \bar{h}_i} \right) \bar{\eta}_m(\bar{h}) - \frac{1}{2} \sum_{\ell, m=1}^n w_{\ell, m} \bar{\eta}_{\ell}(\bar{h}) \left( \frac{\partial \bar{\eta}_m(\bar{h})}{\partial \bar{h}_i} \right) - \bar{\eta}_i(\bar{h}) \\
 &= \bar{\eta}_i(\bar{h}) + (\bar{h}_i - h_i) \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) - \frac{1}{2} \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) \sum_{m=1}^n w_{i, m} \bar{\eta}_m(\bar{h}) \\
 &\quad - \frac{1}{2} \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) \sum_{\ell=1}^n w_{\ell, i} \bar{\eta}_{\ell}(\bar{h}) - \bar{\eta}_i(\bar{h}) \\
 &= (\bar{h}_i - h_i) \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) - \frac{1}{2} \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) \sum_{j=1}^n w_{i, j} \bar{\eta}_j(\bar{h}) \\
 &\quad - \frac{1}{2} \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) \sum_{j=1}^n w_{i, j} \bar{\eta}_j(\bar{h}) \\
 &= \left( \frac{\partial \bar{\eta}_i(\bar{h})}{\partial \bar{h}_i} \right) \left( \bar{h}_i - h_i - \sum_{j=1}^n w_{i, j} \bar{\eta}_j(\bar{h}) \right) \tag{4.10}
 \end{aligned}$$

ただし、5 番目の等号は、ボルツマンマシンの結合定数に関する性質 (2.1) を用いている。また、4 番目の等号は、以下のように場合分けされる変数  $\left( \frac{\partial \bar{\eta}_k(\bar{h})}{\partial \bar{h}_i} \right)$  を用いた変形である。

$$\left( \frac{\partial \bar{\eta}_k(\bar{h})}{\partial \bar{h}_i} \right) = \begin{cases} g_{i, i}(\bar{h}) (= E_{q_{\bar{h}}}[x_i^2] - \{\bar{\eta}_i(\bar{h})\}^2) & \text{if } k = i \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

さらに、 $\left( \frac{\partial \bar{\eta}_k(\bar{h})}{\partial \bar{h}_i} \right)$  は  $\mathcal{M}_C$  の自然座標系  $\bar{h}$  に関する Fisher 情報行列  $G(\bar{h}) \in \mathbb{R}^{n \times n}$  の第  $i$  行  $k$  列成分である。一般的に、Fisher 情報行列は正定値対称行列となるため、行列  $G(\bar{h})$  のランクは  $n$  となる。(4.11) 式より Fisher 情報行列  $G(\bar{h})$  は対角行列となっているため、 $\forall i (1 \leq i \leq n)$  に対して、 $g_{i, i}(\bar{h}) \neq 0$  が成立する。ゆえに、 $e$ -射影の条件式 (4.9) は以下と同値である。

$$\forall i (1 \leq i \leq n), \bar{h}_i = h_i + \sum_{j=1}^n w_{i, j} \bar{\eta}_j(\bar{h}) \tag{4.12}$$

条件式 (4.12) 満たすとき,  $q_{\bar{h}} \in \mathcal{M}_C$  の期待値パラメータ  $\bar{\eta}_i(\bar{h})$  ( $1 \leq i \leq n$ ) は以下のように記述される .

$$\begin{aligned}
 \bar{\eta}_i(\bar{h}) &= \sum_{x_i \in \{0, 1\}} x_i q_{\bar{h}_i}(x_i) \\
 &= \frac{\exp[\bar{h}_i]}{1 + \exp[\bar{h}_i]} \\
 &= \frac{1}{1 + \exp[-\bar{h}_i]} \\
 &= \frac{1}{1 + \exp\left[-h_i - \sum_{j=1}^n w_{i,j} \bar{\eta}_j(\bar{h})\right]} \tag{4.13}
 \end{aligned}$$

(4.13) 式は, 平均場近似により導出される近似分布が満たす条件式 (4.8) そのものである . 従って, 平均場近似により導出される分布は  $P_\theta \in \mathcal{M}$  から  $\mathcal{M}_C$  への  $e$ -射影である [12] .

### 4.3 平均場近似を用いた学習アルゴリズム

平均場近似を用いたボルツマンマシンの学習アルゴリズムは以下のとおりである .

[平均場近似を用いたボルツマンマシンの学習アルゴリズム]

Intialize:  $t = 0, \theta^{(0)} = (h_i^{(0)}, w_{j,k}^{(0)})_{1 \leq i \leq n, 1 \leq j < k \leq n}$

step.1  $s = 0$  とし, 期待値パラメータ  $\bar{\eta}_i^{(0)}$  ( $1 \leq i \leq n$ ) を  $[0, 1]$  の間でランダムに初期化

For Loop1: (逐次代入法による平均場近似の解の導出)

step.2 期待値パラメータ  $\bar{\eta}_i^{(s)}$  ( $1 \leq i \leq n$ ) を以下の式で更新する .

$$\begin{cases}
 \bar{\eta}_1^{(s+1)} = \frac{1}{1 + \exp\left[-h_1^{(t)} - \sum_{j=1}^n w_{1,j}^{(t)} \bar{\eta}_j^{(s)}\right]} \\
 \vdots \\
 \bar{\eta}_n^{(s+1)} = \frac{1}{1 + \exp\left[-h_n^{(t)} - \sum_{j=1}^n w_{n,j}^{(t)} \bar{\eta}_j^{(s)}\right]}
 \end{cases} \tag{4.14}$$

step.3 以下の条件を満たすならば Loop1 を終了 .

$$\forall i, \bar{\eta}_i^{(s+1)} \simeq \bar{\eta}_i^{(s)} \tag{4.15}$$

でなければ,  $s = s + 1$  として, step.2 へ戻る .

End Loop1:

step.4 以下の式に従って,  $\Delta h_i$  ( $1 \leq i \leq n$ ),  $\Delta w_{j,k}$  ( $1 \leq j < k \leq n$ ) を導出する .

$$\Delta h_i = \bar{\eta}_i^{(s)} - E_q[x_i] \tag{4.16}$$

$$\Delta w_{j,k} = \bar{\eta}_j^{(s)} \bar{\eta}_k^{(s)} - E_q[x_j] E_q[x_k] \tag{4.17}$$

step.5 自然パラメータ  $\theta^{(t)} = (h_i^{(t)}, w_{j,k}^{(t)})_{1 \leq i \leq n, 1 \leq j < k \leq n}$  を以下の式に従って更新する .

$$h_i^{(t+1)} = h_i^{(t)} - \alpha \Delta h_i \tag{4.18}$$

$$w_{j,k}^{(t+1)} = w_{j,k}^{(t)} - \alpha \Delta w_{j,k} \tag{4.19}$$

step.6  $\theta^{(t+1)} \simeq \theta^{(t)}$  ならばアルゴリズムを終了する . でなければ,  $t = t + 1$  として step.1 へ戻る .

#### 4.4 アルゴリズムの幾何学的イメージと収束先

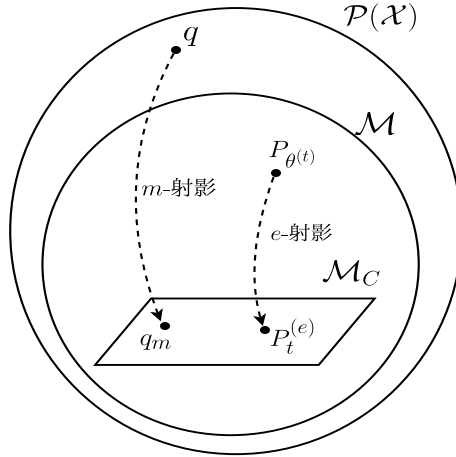


図 4.1: 平均場近似を用いた学習アルゴリズムの幾何学的イメージ

図 4.1 は、学習アルゴリズムの挙動と収束先を示している。まず、図 4.1 に記載されている各記号はそれぞれ以下のとおりである。

- $\mathcal{P}(\mathcal{X})$ : 状態空間  $\mathcal{X} = \{0, 1\}^n$  上の正の確率分布全体
- $\mathcal{M}$ : 素子数  $n$  のボルツマンマシンの平衡確率分布を要素とする分布族
- $\mathcal{M}_C$ : 素子間の結合が存在しない素子数  $n$  のボルツマンマシンの平衡確率分布を要素とする分布族
- $q \in \mathcal{P}(\mathcal{X})$ : 観測データに関する経験分布
- $P_{\theta^{(t)}} \in \mathcal{M}$ : 時刻  $t$  におけるボルツマンマシンの平衡確率分布
- $q_m \in \mathcal{M}_C$ : 経験分布  $q \in \mathcal{P}(\mathcal{X})$  から  $\mathcal{M}_C$  への  $m$ -射影
- $P_t^{(e)} \in \mathcal{M}_C$ : 時刻  $t$  における平衡確率分布  $P_{\theta^{(t)}} \in \mathcal{M}$  から  $\mathcal{M}_C$  への  $e$ -射影

このとき、学習アルゴリズムの挙動は以下のように解釈される。

step.1 ~ step.3:

$P_{\theta^{(t)}} \in \mathcal{M}$  から  $\mathcal{M}_C$  への  $e$ -射影  $P_t^{(e)} \in \mathcal{M}_C$  の導出。

step.4 ~ step.5:

$e$ -射影点  $P_t^{(e)} \in \mathcal{M}_C$  の期待値パラメータを用いて、次の時刻の平衡確率分布  $P_{\theta^{(t+1)}} \in \mathcal{M}$  の自然パラメータを導出。

step.6:

自然パラメータの値がほぼ更新されていない場合、アルゴリズムを終了。でなければ  $t = t+1$  として step.1 へ

次に、学習アルゴリズムの収束先について述べる。step.6 に記載されている学習アルゴリズムの終了条件は、step.4 の (4.16), (4.17) 式の値が任意の  $i, j, k$  に関して微小となるときに満たされる。ここで、 $q_m \in \mathcal{M}_C$  に関する期待値パラメータを  $\bar{\eta}(q_m) = (\bar{\eta}_1(q_m), \dots, \bar{\eta}_n(q_m))$ ,  $P_t^{(e)} \in \mathcal{M}_C$  に関する期待値パラメータを  $\bar{\eta}(P_t^{(e)}) = (\bar{\eta}_1(P_t^{(e)}), \dots, \bar{\eta}_n(P_t^{(e)}))$  とする。この



とき,  $m$ -射影の性質から (4.16) 式中の期待値  $E_q[x_i]$  は  $\bar{\eta}_i(q_m)$  と等価である. 従って, 学習アルゴリズムの終了条件は以下と同値となる.

$$\forall i(1 \leq i \leq n), \quad \bar{\eta}_i(q_m) \simeq \bar{\eta}_i(P_t^{(e)}) \quad (4.20)$$

期待値パラメータ  $\bar{\eta}(\bar{h})$  は  $\mathcal{M}_C$  に含まれる分布を一意に指定するものである. 従って, 条件式 (4.20) を満たすような平衡確率分布  $P_{\theta^{(t)}} \in \mathcal{M}$  とは,  $\mathcal{M}$  から  $\mathcal{M}_C$  への  $e$ -射影によって  $q_m \in \mathcal{M}_C$  に移される分布である. ただし,  $\mathcal{M}$  から  $\mathcal{M}_C$  への  $e$ -射影の一意性は保証されないため, この学習アルゴリズムの収束先は複数存在する場合がある.

## 第5章 クラスタ型モデルへの近似

この章では、形状が格子模型となるボルツマンマシンを対象として、ボルツマンマシンの平衡確率分布をクラスタ型モデルと呼ばれるグラフ上で定義される平衡確率分布で近似する問題を考える。さらに、この問題設定の下で元の平衡確率分布を最もよく近似するクラスタ型モデル上の平衡確率分布が満たす条件式の導出を行う。

### 5.1 グラフ概念を用いた記法の導入

この節では、以降の議論をしやすくするために、グラフ理論の記法に則ったいくつかの記号を導入する。

まず、頂点を要素とする集合  $V$  (以降では、頂点集合と呼ぶ) が与えられているとし、集合  $A$  を以下のように定義する。

$$A := \{(u, v) \mid u, v \in V\} \quad (5.1)$$

グラフ理論では、頂点  $u$  から頂点  $v$  への有向辺を数学的に表すために、順序対  $(u, v)$  が用いられる。従って、集合  $A$  は頂点集合  $V$  上の全ての有向辺からなる集合である。このとき、頂点集合  $V$  と部分集合  $E \subseteq A$  の組  $G = (V, E)$  のことを有向グラフと呼ぶ。

一方で、集合  $A$  が (5.2) 式のように定義されている場合、頂点集合  $V$  と部分集合  $E \subseteq A$  の組  $G = (V, E)$  のことを無向グラフと呼ぶ。

$$A := \{\{u, v\} \mid u, v \in V\} \quad (5.2)$$

以降では、グラフ  $G = (V, E)$  が与えられているとき、グラフ  $G$  の頂点集合と辺集合をそれぞれ  $V(G) (= V)$ ,  $E(G) (= E)$  として表す。また、 $V(G)$  の元はグラフ  $G$  の頂点、 $E(G)$  の元はグラフ  $G$  の辺とそれぞれ呼ぶ。

ボルツマンマシンは無向グラフで表されるため、各辺において頂点を一意に指定することができない。この問題を解決するために、次の図 5.1 のような状況を考える。

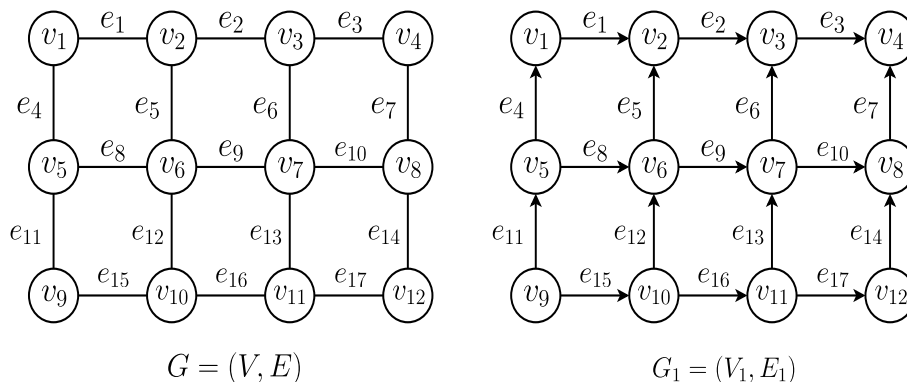


図 5.1: 格子模型型のボルツマンマシンと有向グラフ

図 5.1 の左側の無向グラフ  $G$  は格子模型型のボルツマンマシン、右側のグラフ  $G_1$  は無向グラフ  $G$  の各辺に向きを定めた有向グラフである。 $G_1$  は任意の頂点  $v_i$  に対して、それを始

点と終点とするような経路を持たない．このような条件を満たすとき，有向グラフは無閉路 (acyclic) であるという．なお，格子模型型のグラフの場合，acyclic な有向グラフは必ず存在する．グラフ理論では，有向辺  $e \in G_1$  の始点と終点をそれぞれ  $\text{In}(e), \text{Out}(e)$  と表現することが多い．本論文では，これらを  $\ell(e) := \text{In}(e), r(e) := \text{Out}(e)$  として記述する．ここで，格子模型型のボルツマンマシン  $G$  の無向辺  $e_i$  に関する頂点  $\ell(e_i), r(e_i)$  をそれぞれ，無閉路な有向グラフ  $G_1$  上の対応する有向辺  $e_i$  に関する頂点  $\ell(e_i), r(e_i)$  と同じものとする．このように定めることで，ボルツマンマシンの各辺の端点の中から一つの頂点を一意に指定することが可能となる．さらに，頂点  $v_i$  の状態変数を  $x_{v_i} \in \{0, 1\}$ ，頂点  $v_i$  のしきい値を  $h_{v_i} \in \mathbb{R}$ ，辺  $e$  により接続されている頂点間の結合定数を  $w_e \in \mathbb{R}$ ，ボルツマンマシン  $G$  の平衡確率分布がなす分布族を  $\mathcal{M}(G)$  として表すことにする．

次に，グラフ理論のカットの問題について説明する．

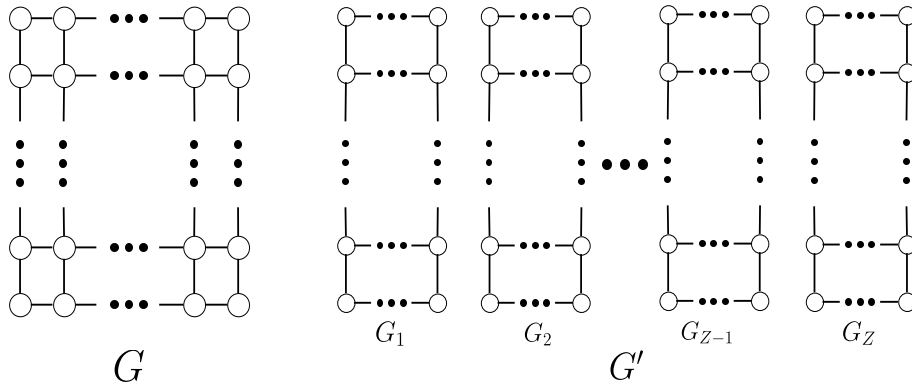


図 5.2: グラフのカット

グラフ  $G$  は格子模型型のボルツマンマシン，グラフ  $G'$  はボルツマンマシン  $G$  を短冊形に部分グラフ  $G_1 \sim G_Z$  へと分割させたグラフである．このとき，部分グラフ  $G_i$  と  $G_{i+1}$  に関するカット  $C_i$  とは，以下のようにして定義される辺集合のことである．

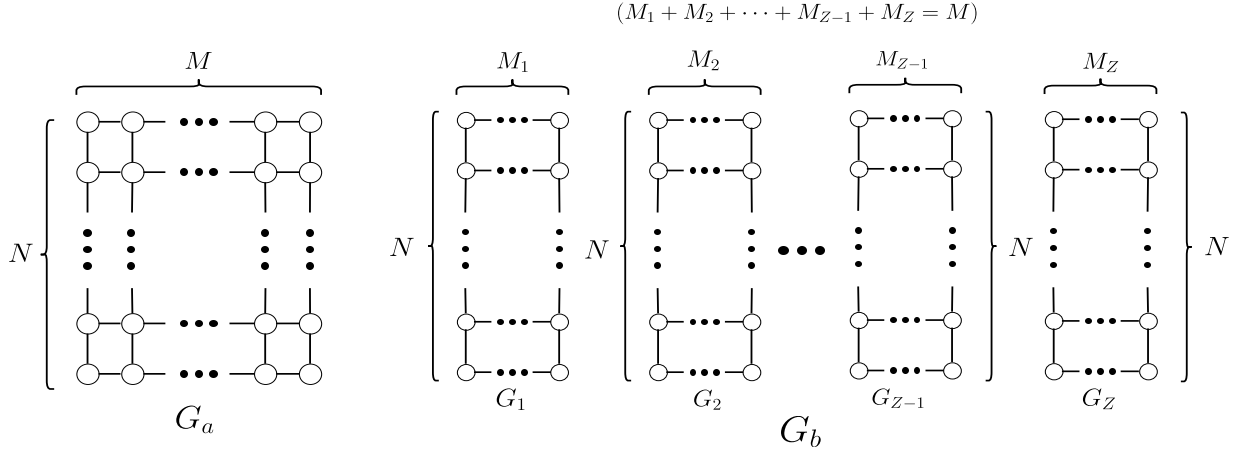
$$C_i := \{\{u, v\} \in E(G) \mid u \in V(G_i), v \in V(G_{i+1})\} \quad (5.3)$$

すなわち，カット  $C_i$  とは  $G$  に含まれる辺の中で， $G_i$  に含まれる頂点と  $G_{i+1}$  に含まれる頂点を接続している辺を要素とする集合である．さらに，カット  $C_1 \sim C_{Z-1}$  の和集合を

$$C := \bigcup_{i=1}^{Z-1} C_i \quad (5.4)$$

と定義する．

## 5.2 クラスタ型模型への近似問題


 図 5.3:  $N$  行  $M$  列格子模型とクラスタ型模型

まずはじめに，図 5.3 のようなグラフ  $G_a, G_b, G_1, \dots, G_Z$  が与えられているとする．グラフ  $G_a$  は  $N$  行  $M$  列格子模型型のボルツマンマシン，グラフ  $G_b$  はグラフ  $G_a$  を短冊形に  $Z$  個のクラスタ  $G_1 \sim G_Z$  へと分割させたグラフであり，本論文ではクラスタ型模型と呼ぶことにする．さらに，格子模型型のボルツマンマシンの列数  $M$ ，各クラスタの列数  $M_1 \sim M_Z$  に関して，以下の等式が成り立つものとする．

$$\sum_{i=1}^Z M_i = M \quad (5.5)$$

図 5.3 中の各グラフをボルツマンマシンと見做したとき，各グラフ上の平衡確率分布は以下のように与えられる．

最初に，グラフ  $G_a$  上で定められる平衡確率分布  $P_\theta^{G_a}(x)$  ( $x = (x_v)_{v \in V(G_a)}$ ) は，自然パラメータ  $\theta = (\theta^\alpha)_{\alpha \in V(G_a) \cup E(G_a)}$  と関数  $F_\alpha(x)$  ( $\alpha \in V(G_a) \cup E(G_a)$ ) をそれぞれ

$$\theta^\alpha = \begin{cases} h_v & \text{if } \alpha = v \text{ (}\alpha\text{が頂点 } v \in V(G_a) \text{ である場合)} \\ w_e & \text{if } \alpha = e \text{ (}\alpha\text{が辺 } e \in E(G_a) \text{ である場合)} \end{cases} \quad (5.6)$$

$$F_\alpha(x) = \begin{cases} x_v & \text{if } \alpha = v \text{ (}\alpha\text{が頂点 } v \in V(G_a) \text{ である場合)} \\ x_{\ell(e)} x_{r(e)} & \text{if } \alpha = e \text{ (}\alpha\text{が辺 } e \in E(G_a) \text{ である場合)} \end{cases} \quad (5.7)$$

としたとき，以下のように表現される．

$$P_\theta^{G_a}(x) = \exp \left[ \sum_{\alpha \in V(G_a) \cup E(G_a)} \theta^\alpha F_\alpha(x) - \Psi(\theta) \right] \quad (5.8)$$

$$\Psi(\theta) = \log \sum_{x \in \{0,1\}^{MN}} \exp \left[ \sum_{\alpha \in V(G_a) \cup E(G_a)} \theta^\alpha F_\alpha(x) \right] \quad (5.9)$$

次に，グラフ  $G_i$  ( $1 \leq i \leq Z$ ) 上で定められる平衡確率分布  $P_{\theta_i}^{G_i}(x_i)$  ( $x_i = (x_v)_{v \in V(G_i)}$ ) は，

自然パラメータ  $\bar{\theta}_i = (\bar{\theta}_i^\beta)_{\beta \in V(G_i) \cup E(G_i)}$  と関数  $G_\beta(x_i)$  ( $\beta \in V(G_i) \cup E(G_i)$ ) をそれぞれ

$$\bar{\theta}_i^\beta = \begin{cases} \bar{h}_v & \text{if } \beta = v \text{ (}\beta\text{が頂点 } v \in V(G_i) \text{ である場合)} \\ \bar{w}_e & \text{if } \beta = e \text{ (}\beta\text{が辺 } e \in E(G_i) \text{ である場合)} \end{cases} \quad (5.10)$$

$$G_\beta(x_i) = \begin{cases} x_v & \text{if } \beta = v \text{ (}\beta\text{が頂点 } v \in V(G_i) \text{ である場合)} \\ x_{\ell(e)}x_{r(e)} & \text{if } \beta = e \text{ (}\beta\text{が辺 } e \in E(G_i) \text{ である場合)} \end{cases} \quad (5.11)$$

としたとき、以下のように表現される。

$$P_{\bar{\theta}_i}^{G_i}(x_i) = \exp \left[ \sum_{\beta \in V(G_i) \cup E(G_i)} \bar{\theta}_i^\beta G_\beta(x_i) - \Phi_i(\bar{\theta}_i) \right] \quad (5.12)$$

$$\Phi_i(\bar{\theta}_i) = \log \sum_{x_i \in \{0,1\}^{M_i N}} \exp \left[ \sum_{\beta \in V(G_i) \cup E(G_i)} \bar{\theta}_i^\beta G_\beta(x_i) \right] \quad (5.13)$$

最後に、グラフ  $G_b$  上で定義される平衡確率分布  $P_{\bar{\theta}}^{G_b}(x)$  ( $x = (x_v)_{v \in V(G_b)}$ ) は、自然パラメータ  $\bar{\theta} = (\bar{\theta}^\mu)_{\mu \in V(G_b) \cup E(G_b)}$  を

$$\bar{\theta}^\mu = \bar{\theta}_i^\beta \quad \text{if } \exists i \in \{1, \dots, Z\}, \exists \beta \in V(G_b) \cup E(G_b), \mu = \beta \quad (5.14)$$

としたとき、以下のように表現される。

$$P_{\bar{\theta}}^{G_b}(x) = \prod_{i=1}^Z P_{\bar{\theta}_i}^{G_i}(x_i) \quad (5.15)$$

今、グラフ  $G_a$  上で定義される平衡確率分布  $P_{\bar{\theta}}^{G_a} \in \mathcal{M}(G_a)$  が一つ与えられているとする。このとき、 $P_{\bar{\theta}}^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_b)$  への  $e$ -射影  $q_{\bar{\theta}}^{G_b} \in \mathcal{M}(G_b)$  が持つ自然パラメータ  $\bar{\theta} = (\bar{\theta}^\mu)_{\mu \in V(G_b) \cup E(G_b)}$  はどのような条件式を満たすのか？

以上の内容までが問題設定となっている。以下では、この問題設定を与える意義を述べる。元のグラフ  $G_a$  から全ての辺を取り除いたグラフを  $G_c = (V(G_a), \emptyset)$  とした時、分布族  $\mathcal{M}(G_c), \mathcal{M}(G_b), \mathcal{M}(G_a)$  に関して以下の包含関係が成立する。

$$\mathcal{M}(G_c) \subset \mathcal{M}(G_b) \subset \mathcal{M}(G_a) \quad (5.16)$$

上記の包含関係より、以下の不等式が成立する。

$$\min_{P_{\bar{\theta}}^{G_b} \in \mathcal{M}(G_b)} D(P_{\bar{\theta}}^{G_b} \| P_{\bar{\theta}}^{G_a}) \leq \min_{P_{\bar{h}}^{G_c} \in \mathcal{M}(G_c)} D(P_{\bar{h}}^{G_c} \| P_{\bar{\theta}}^{G_a}) \quad (5.17)$$

ここで、分布  $q_{\bar{\theta}}^{G_b}, q_{\bar{h}}^{G_c}$  をそれぞれ

$$q_{\bar{\theta}}^{G_b} = \arg \min_{P_{\bar{\theta}}^{G_b} \in \mathcal{M}(G_b)} D(P_{\bar{\theta}}^{G_b} \| P_{\bar{\theta}}^{G_a}) \quad (5.18)$$

$$q_{\bar{h}}^{G_c} = \arg \min_{P_{\bar{h}}^{G_c} \in \mathcal{M}(G_c)} D(P_{\bar{h}}^{G_c} \| P_{\bar{\theta}}^{G_a}) \quad (5.19)$$

とする。このとき、ボルツマンマシンにおける平均場近似とは、元の平衡確率分布  $P_{\bar{\theta}}^{G_a} \in \mathcal{M}(G_a)$  に対して、分布  $q_{\bar{h}}^{G_c} \in \mathcal{M}(G_c)$  で近似する手法であった。さらに4.2節では、平均場近似により導出される近似分布  $q_{\bar{h}}^{G_c} \in \mathcal{M}(G_c)$  は、 $P_{\bar{\theta}}^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_c)$  への  $e$ -射影であることを示した。4.2節とほぼ同様の議論により、(5.18)式を満たす分布  $q_{\bar{\theta}}^{G_b} \in \mathcal{M}(G_b)$  は  $P_{\bar{\theta}}^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_b)$  への  $e$ -射影となることも示せる。従って、上記の問題設定の下で導出される  $e$ -射影は、平均場近似により導出される近似分布よりも近似精度が高い分布であることが期待される。

### 5.3 $e$ -射影点を満たす条件式の導出

この節では、5.2 節の問題設定を考慮した上で、 $e$ -射影点を満たす条件式と同値変形を行っていく。まず、集合  $U$  を

$$U := V(G_b) \cup E(G_b) \quad (5.20)$$

と定義する。このとき、 $P_\theta^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_b)$  への  $e$ -射影  $q_\theta^{G_b} \in \mathcal{M}(G_b)$  は以下の条件 (5.21) を満たす。

$$\forall \mu \in U, \quad \frac{\partial D(q_\theta^{G_b} \| P_\theta^{G_a})}{\partial \bar{\theta}^\mu} = 0 \quad (5.21)$$

但し、 $D(\cdot \| \cdot)$  は Kullback-Leibler ダイバージェンスである。また、3 章の条件式 (3.3), (3.4) より、以下の等式が成り立つ。

$$\forall \mu \in U, \quad \frac{\partial q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} = q_\theta^{G_b}(x) \left( \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right), \quad E_{q_\theta^{G_b}} \left[ \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right] = 0 \quad (5.22)$$

上記等式を用いて、(5.21) 式の左辺について計算すると

$$\begin{aligned} \frac{\partial D(q_\theta^{G_b} \| P_\theta^{G_a})}{\partial \bar{\theta}^\mu} &= \frac{\partial}{\partial \bar{\theta}^\mu} \sum_{x \in \{0,1\}^{MN}} q_\theta^{G_b}(x) \left( \log q_\theta^{G_b}(x) - \log P_\theta^{G_a}(x) \right) \\ &= \sum_{x \in \{0,1\}^{MN}} \left( \frac{\partial q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \left( \log q_\theta^{G_b}(x) - \log P_\theta^{G_a}(x) \right) \\ &\quad + \sum_{x \in \{0,1\}^{MN}} q_\theta^{G_b}(x) \left( \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \\ &= \sum_{x \in \{0,1\}^{MN}} q_\theta^{G_b}(x) \left( \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \left( \log q_\theta^{G_b}(x) - \log P_\theta^{G_a}(x) \right) \\ &\quad + E_{q_\theta^{G_b}} \left[ \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right] \\ &= E_{q_\theta^{G_b}} \left[ \left( \log q_\theta^{G_b}(x) - \log P_\theta^{G_a}(x) \right) \left( \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] \end{aligned} \quad (5.23)$$

となる。この計算結果を用いると、 $e$ -射影の条件式 (5.21) は以下と同値となる。

$$\forall \mu \in U, \quad E_{q_\theta^{G_b}} \left[ \left( \log q_\theta^{G_b}(x) - \log P_\theta^{G_a}(x) \right) \left( \frac{\partial \log q_\theta^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] = 0 \quad (5.24)$$

次に, 5.1 節で導入したカット  $C_i$  ( $1 \leq i \leq Z-1$ ) を用いると,  $\log q_{\bar{\theta}}^{G_b}(x) - \log P_{\theta}^{G_a}(x)$  は

$$\begin{aligned}
 \log q_{\bar{\theta}}^{G_b}(x) - \log P_{\theta}^{G_a}(x) &= \sum_{i=1}^Z \log q_{\bar{\theta}_i}^{G_i}(x_i) - \log P_{\theta}^{G_a}(x) \\
 &= \sum_{i=1}^Z \sum_{\beta \in V(G_i) \cup E(G_i)} (\bar{\theta}_i^{\beta} - \theta^{\beta}) G_{\beta}(x_i) \\
 &\quad - \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e F_e(x) - \sum_{i=1}^Z \Phi_i(\bar{\theta}_i) + \Psi(\theta) \\
 &= \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) F_{\beta}(x) - \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e F_e(x) \\
 &\quad - \sum_{i=1}^Z \Phi_i(\bar{\theta}_i) + \Psi(\theta) \tag{5.25}
 \end{aligned}$$

となる. この計算結果を (5.24) 式へ代入すると, (5.24) 式は以下の (5.26) 式と同値となる.

$$\begin{aligned}
 \forall \mu \in U, \quad E_{q_{\bar{\theta}}^{G_b}} \left[ \left\{ \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) F_{\beta}(x) \right\} \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \\
 = E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e F_e(x) \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \tag{5.26}
 \end{aligned}$$

ただし, (5.26) 式を導出する際に, 以下の条件式 (5.27) が成立することを用いた.

$$\begin{aligned}
 \forall i \in \{1, \dots, Z\} \\
 E_{q_{\bar{\theta}}^{G_b}} \left[ \Phi_i(\bar{\theta}_i) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] = 0, \quad E_{q_{\bar{\theta}}^{G_b}} \left[ \Psi(\theta) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] = 0 \tag{5.27}
 \end{aligned}$$

次に, (5.26) 式の右辺と左辺の式変形を行う. 分布  $q_{\bar{\theta}}^{G_b}$  の Fisher 計量  $g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right)$  ( $\mu, \xi \in U$ ) と期待値パラメータ  $\bar{\eta}_{\mu}(\bar{\theta})$  ( $\mu \in U$ ) はそれぞれ

$$g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right) := E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\xi}} \right) \right] \tag{5.28}$$

$$\bar{\eta}_{\mu}(\bar{\theta}) := E_{q_{\bar{\theta}}^{G_b}} [F_{\mu}(x)] \tag{5.29}$$

で定義される. さらに,  $\forall q_{\bar{\theta}}^{G_b} \in \mathcal{M}(G_b)$  に対して, 以下の性質が成り立つ.

$$\forall \mu \in U, \quad \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} = F_{\mu}(x) - \bar{\eta}_{\mu}(\bar{\theta}) \tag{5.30}$$

(5.26) 式の右辺を計算すると

$$\begin{aligned}
 & E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e F_e(x) \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] \\
 &= \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e E_{q_{\bar{\theta}}^{G_b}} \left[ F_e(x) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] \\
 &= \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e E_{q_{\bar{\theta}}^{G_b}} \left[ x_{\ell(e)} x_{r(e)} \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] \\
 &= \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} + \bar{\eta}_{\ell(e)}(\bar{\theta}) \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} + \bar{\eta}_{r(e)}(\bar{\theta}) \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] \\
 & \left[ \because \forall \mu \in U, \quad \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} = F_\mu(x) - \bar{\eta}_\mu(\bar{\theta}) \right] \\
 &= \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e \left\{ \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right]}_{0} \right\}_{*1} \\
 & \quad + \bar{\eta}_{r(e)}(\bar{\theta}) E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] \\
 & \quad + \bar{\eta}_{\ell(e)}(\bar{\theta}) E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right) \right] + \underbrace{\bar{\eta}_{\ell(e)}(\bar{\theta}) \bar{\eta}_{r(e)}(\bar{\theta}) E_{q_{\bar{\theta}}^{G_b}} \left[ \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right]}_{0} \right\}_{*2} \\
 &= \sum_{i=1}^{Z-1} \sum_{e \in C_i} \theta^e \left\{ \bar{\eta}_{r(e)}(\bar{\theta}) g_{\mu, \ell(e)} \left( q_{\bar{\theta}}^{G_b} \right) + \bar{\eta}_{\ell(e)}(\bar{\theta}) g_{\mu, r(e)} \left( q_{\bar{\theta}}^{G_b} \right) \right\} \\
 &= g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right) y^\xi \tag{5.31}
 \end{aligned}$$

となる．ただし，最後の等式中の  $y^\xi$  ( $\xi \in U$ ) は補助的な変数であり，以下のように場合分けされる．

$$y^\xi = \begin{cases} \theta^e \bar{\eta}_{r(e)}(\bar{\theta}) & \text{if } \exists e \in C, \forall f \in C, \xi = \ell(e) \wedge \xi \neq r(f) \\ \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}) & \text{if } \exists e \in C, \forall f \in C, \xi = r(e) \wedge \xi \neq \ell(f) \\ \theta^e \bar{\eta}_{r(e)}(\bar{\theta}) + \theta^f \bar{\eta}_{\ell(f)}(\bar{\theta}) & \text{if } \exists e, f \in C, \xi = \ell(e) = r(f) \\ 0 & \text{otherwise} \end{cases} \tag{5.32}$$

また， $*1, *2$  の部分が 0 となる理由は，それぞれ以下の通りである．まず，条件式 (5.22) により， $*2$  部分は 0 に等しくなる．次に， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right)$ ， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right)$ ， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right)$  をそれぞれ確率変数として見做す．このとき， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right)$  はカットされた辺  $e$  の左側に位置するクラスターに関する確率変数である．一方で， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right)$  はカットされた辺  $e$  の右側に位置するクラスターに関する確率変数である．従って， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right)$ ， $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right)$  は必ず統計的に独立となる．また，確率変数  $\left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^\mu} \right)$  は，頂点  $\mu$  (もしくは辺  $\mu$ ) を含む



クラスターに関する確率変数である．この節では， $\mu$  は任意に取っているため， $*^1$  は以下のように場合分けされる．

$$E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right]$$

$$\left\{ \begin{array}{l} \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right) \right]}_0 E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] = 0 \quad (5.33) \\ E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right) \right]}_0 = 0 \quad (5.34) \\ \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\ell(e)}} \right) \right]}_0 \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{r(e)}} \right) \right]}_0 \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right]}_0 = 0 \quad (5.35) \end{array} \right.$$

条件式 (5.22) より，(5.33) ~ (5.35) 式中の全ての波線部分は 0 と等しくなる．従って， $*^1$  は 0 と等価である．

次に，(5.26) 式の左辺を計算すると

$$\begin{aligned} & E_{q_{\bar{\theta}}^{G_b}} \left[ \left\{ \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) F_{\beta}(x) \right\} \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \\ &= \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) E_{q_{\bar{\theta}}^{G_b}} \left[ F_{\beta}(x) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \\ &= \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\beta}} + \bar{\eta}_{\beta}(\bar{\theta}) \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \\ &= \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) \left\{ E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\beta}} \right) \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] + \bar{\eta}_{\beta}(\bar{\theta}) \underbrace{E_{q_{\bar{\theta}}^{G_b}} \left[ \left( \frac{\partial \log q_{\bar{\theta}}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right]}_0 \right\} \\ &= \sum_{\beta \in U} (\bar{\theta}^{\beta} - \theta^{\beta}) g_{\mu, \beta} \left( q_{\bar{\theta}}^{G_b} \right) \\ &= g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right) z^{\xi} \quad (5.36) \end{aligned}$$

となる．ただし，最後の等式中の  $z^{\xi}$  ( $\xi \in U$ ) は以下のように定義されているものとする．

$$z^{\xi} := \bar{\theta}^{\xi} - \theta^{\xi} \quad (\forall \xi \in U) \quad (5.37)$$

(5.26) 式の右辺と左辺の式変形により，(5.26) 式は以下の (5.38) 式と同値となる．

$$\forall \mu \in U, \quad g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right) z^{\xi} = g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right) y^{\xi} \quad (5.38)$$

Fisher 情報行列  $g \left( q_{\bar{\theta}}^{G_b} \right) = \left[ g_{\mu, \xi} \left( q_{\bar{\theta}}^{G_b} \right) \right]_{\mu, \xi \in U}$  は正定値行列であるため，逆行列が存在する．従って，両辺に左から逆行列をかけることで，(5.38) 式はさらに

$$\forall \xi \in U, \quad z^{\xi} = y^{\xi} \quad (5.39)$$

と同値であることがわかる．

以上の議論により， $P_\theta^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_b)$  への  $e$ -射影  $q_\theta^{G_b} \in \mathcal{M}(G_b)$  が持つ自然パラメータ  $\bar{\theta} = (\bar{\theta}^\mu)_{\mu \in V(G_b) \cup E(G_b)}$  は以下の条件式 (5.40) ~ (5.43) を満たす．

$$\bar{\theta}^\mu = \begin{cases} \theta^\mu + \theta^e \bar{\eta}_{r(e)}(\bar{\theta}) & \text{if } \exists e \in C, \forall f \in C, \mu = \ell(e) \wedge \mu \neq r(f) & (5.40) \\ \theta^\mu + \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}) & \text{if } \exists e \in C, \forall f \in C, \mu = r(e) \wedge \mu \neq \ell(f) & (5.41) \\ \theta^\mu + \theta^e \bar{\eta}_{r(e)}(\bar{\theta}) + \theta^f \bar{\eta}_{\ell(f)}(\bar{\theta}) & \text{if } \exists e, f \in C, \mu = \ell(e) = r(f) & (5.42) \\ \theta^\mu & \text{otherwise} & (5.43) \end{cases}$$

条件式 (5.40) ~ (5.43) が意味するところは以下の通りである．まず (5.40) 式は，クラスタ  $G_1$  と列数が 2 以上のクラスタ  $G_i$  ( $2 \leq i \leq Z-1$ ) に含まれる右端の素子のしきい値が満たす条件式である．(5.41) 式は，クラスタ  $G_Z$  と列数が 2 以上のクラスタ  $G_i$  ( $2 \leq i \leq Z-1$ ) に含まれる左端の素子のしきい値が満たす条件式である．条件式 (5.42) は，列数が 1 のクラスタ  $G_i$  ( $2 \leq i \leq Z-1$ ) に含まれる素子のしきい値が満たす条件式である．そして，その他のしきい値と結合定数は  $P_\theta^{G_a} \in \mathcal{M}(G_a)$  が持つ自然パラメータ  $\theta = (\theta^\alpha)_{\alpha \in V(G_a) \cup E(G_a)}$  の値をそのまま代入するという条件式が (5.43) 式である．これらの条件式 (5.40) ~ (5.43) は，Xing ら [2] が示した一般化平均場方程式 (generalized mean field equation) と等価となる (付録.A 参照) ．

## 第6章 ビリーフプロパゲーションを用いた $e$ -射影点導出アルゴリズムについて

ビリーフプロパゲーションとは，ループのないグラフ上で定義される同時確率分布に関する周辺分布を高速で導出するための手法である．4章で紹介したクラスター型模型に含まれる各クラスターは，それぞれループのないグラフへと変形することが可能である．従って，ビリーフプロパゲーションを用いることで，各クラスター上で定義される平衡確率分布に関する周辺分布を高速で導出することができる．この章では，クラスター型模型をループのないグラフへと変形できることを示した後に，ビリーフプロパゲーションにより周辺分布がどのように導出されるのかについて解説する．さらに，5章で導出した条件式 (5.40) ~ (5.43) を満たす近似分布 ( $e$ -射影点) を求めるための提案アルゴリズムについて述べる．

### 6.1 クラスターから一次元鎖への変形

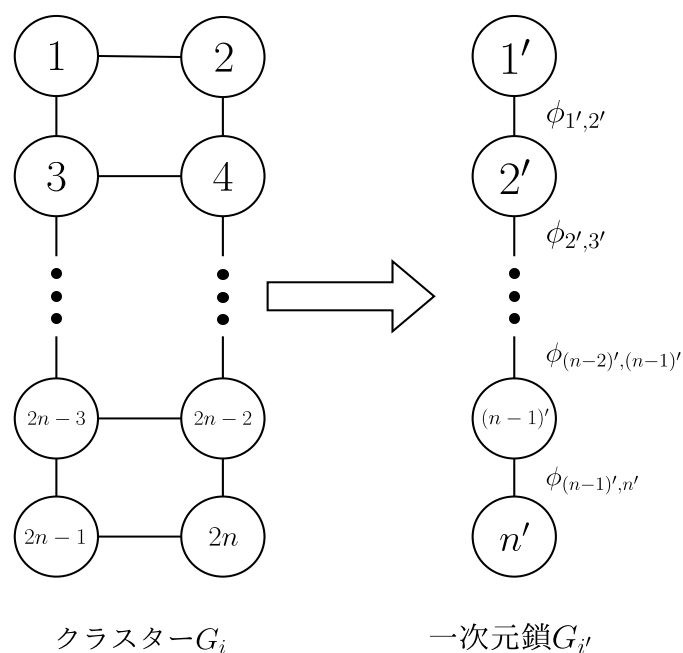


図 6.1: クラスターの変形

この節では，図 6.1 のクラスター  $G_i$  をクラスター型模型に含まれるクラスターのひとつとして捉え，どのようにして図 6.1 のような一次元鎖  $G_{i'}$  へと変形されるのかについて述べる．

まず，クラスター  $G_i$  の各頂点には，左上から順に 1 から  $2n$  までの番号が割り振られているとする． $x_i \in \{0, 1\}$  を頂点  $i$  ( $1 \leq i \leq 2n$ ) の状態変数， $h_i$  を頂点  $i$  のしきい値， $w_{i,j}$  ( $i < j$ ) を頂点  $i, j$  の結合定数， $x = (x_1, \dots, x_n)$  をグラフ  $(a)$  のネットワークの状態とする．このとき，クラスター  $G_i$  上で定義されるボルツマンマシンの平衡確率分布  $P(x) \in \mathcal{M}(G_i)$  は以

下のように記述される .

$$P(x) = \frac{\exp \left[ \sum_{i=1}^{2n} h_i x_i + \sum_{i=1}^n w_{2i-1,2i} x_{2i-1} x_{2i} + \sum_{i=1}^{2n-2} w_{i,i+2} x_i x_{i+2} \right]}{\sum_{x \in \{0,1\}^{2n}} \exp \left[ \sum_{i=1}^{2n} h_i x_i + \sum_{i=1}^n w_{2i-1,2i} x_{2i-1} x_{2i} + \sum_{i=1}^{2n-2} w_{i,i+2} x_i x_{i+2} \right]} \quad (6.1)$$

この平衡確率分布 (6.1) は一次元鎖  $G_{i'}$  上でも定義することができる . まず ,  $G_{i'}$  に含まれる各頂点に対し , 上から順に  $1'$  から  $n'$  が割り振られているとする . また ,  $G_{i'}$  の頂点  $i'$  ( $1 \leq i \leq n$ ) に関する状態変数  $x_{i'}$  は以下のように表現されるものとする .

$$x_{i'} = (x_{2i-1}, x_{2i}) \in \{0, 1\}^2 \quad (6.2)$$

すなわち , 頂点  $1', 2', \dots, n'$  はそれぞれクラスター  $G_i$  に含まれる頂点  $1$  と  $2, 3$  と  $4, \dots, 2n-1$  と  $2n$  を一つの頂点へと代替させたものとして捉えることができる . さらに , 一次元鎖  $G_{i'}$  内に記載されている  $\phi_{i',(i+1)'}$  ( $1 \leq i \leq n-1$ ) は隣接する二頂点の状態変数を変数として持つ関数であり ,  $\phi_{1',2'}(x_{1'}, x_{2'}), \dots, \phi_{(n-1)',n'}(x_{(n-1)'}, x_{n'})$  はそれぞれ以下のように定義されているとする .

$$\phi_{i',(i+1)'}(x_{i'}, x_{(i+1)'}) := \begin{cases} \sum_{j=2i-1}^{2i} h_j x_j + \sum_{j=2i-1}^{2i} w_{j,j+2} x_j x_{j+2} + w_{2i-1,2i} x_{2i-1} x_{2i} & \text{if } i \neq n-1 \\ \sum_{j=2n-3}^{2n} h_j x_j + \sum_{j=2n-3}^{2n-2} w_{j,j+2} x_j x_{j+2} + \sum_{j=n-1}^n w_{2j-1,2j} x_{2j-1} x_{2j} & \text{if } i = n-1 \end{cases} \quad (6.3)$$

一次元鎖  $G_{i'}$  のネットワークの状態を

$$x' = (x_{1'}, \dots, x_{n'}) \quad (6.4)$$

としたとき , 平衡確率分布 (6.1) は次のようにも表現される .

$$P(x') = \frac{\exp \left[ \sum_{i=1}^{n-1} \phi_{i',(i+1)'}(x_{i'}, x_{(i+1)'}) \right]}{\sum_{x_{1'}, \dots, x_{n'}} \exp \left[ \sum_{i=1}^{n-1} \phi_{i',(i+1)'}(x_{i'}, x_{(i+1)'}) \right]} \quad (6.5)$$

従って , クラスター  $G_i$  , 一次元鎖  $G_{i'}$  のどちらを用いても同一の平衡確率分布を定義することができるため , クラスター  $G_i$  は一次元鎖  $G_{i'}$  へと変形することが可能である . ここではクラスターの列数を 2 とした場合を例として取り上げたが , 列数を任意に指定したとしても , 関数  $\phi_{i',(i+1)'}$  ( $1 \leq i \leq n-1$ ) を適切に設定することで一次元鎖へと変形することが可能である . 従って , クラスター型模型はループを含まないグラフへと変形することができる .

## 6.2 ビリーフプロパゲーション

この節では , 前節の図 6.1 の一次元鎖  $G_{i'}$  上で定義される平衡確率分布 (6.5) を例として , ビリーフプロパゲーションにより周辺分布がどのように導出されるのかについて解説する .

平衡確率分布 (6.5) の変数  $x_{i'}$  に関する周辺分布  $P(x_{i'})$  は以下のように計算される .

$$\begin{aligned}
 P(x_{i'}) &= \frac{\sum_{x' \setminus x_{i'}} \exp \left[ \sum_{j=1}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right]}{\sum_{x_1', \dots, x_{n'}} \exp \left[ \sum_{j=1}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right]} \\
 &= \frac{\sum_{x' \setminus x_{i'}} \exp \left[ \sum_{j=1}^{i-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] \cdot \exp \left[ \sum_{j=i}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right]}{\sum_{x_{i'}' \setminus x_{i'}} \exp \left[ \sum_{j=1}^{i-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] \cdot \exp \left[ \sum_{j=i}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right]} \\
 &= \frac{\left( \sum_{x_1', \dots, x_{(i-1)'}} e^{\sum_{j=1}^{i-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'})} \right) \left( \sum_{x_{(i+1)'}, \dots, x_{n'}} e^{\sum_{j=i}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'})} \right)}{\sum_{x_{i'}' \setminus x_{i'}} \left( \sum_{x_1', \dots, x_{(i-1)'}} e^{\sum_{j=1}^{i-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'})} \right) \left( \sum_{x_{(i+1)'}, \dots, x_{n'}} e^{\sum_{j=i}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'})} \right)} \\
 &= \frac{\hat{P}_{T_{i'}}(x_{i'}) \times \hat{P}_{U_{i'}}(x_{i'})}{\sum_{x_{i'}'} \hat{P}_{T_{i'}}(x_{i'}) \times \hat{P}_{U_{i'}}(x_{i'})} \tag{6.6}
 \end{aligned}$$

ただし , 1, 2 番目の等式中の  $\sum_{x' \setminus x_{i'}}$  は  $x_{i'}$  以外の  $x'$  の成分に関する総和を意味している .

また ,  $\hat{P}_{T_{i'}}(x_{i'})$ ,  $\hat{P}_{U_{i'}}(x_{i'})$  はそれぞれ以下のように定義されているものとする .

$$\hat{P}_{T_{i'}}(x_{i'}) := \begin{cases} \sum_{x_1', \dots, x_{(i-1)'}} \exp \left[ \sum_{j=1}^{i-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] & \text{if } i \neq 1 \\ 1 & \text{if } i = 1 \end{cases} \tag{6.7}$$

$$\hat{P}_{U_{i'}}(x_{i'}) := \begin{cases} \sum_{x_{(i+1)'}, \dots, x_{n'}} \exp \left[ \sum_{j=i}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] & \text{if } i \neq n \\ 1 & \text{if } i = n \end{cases} \tag{6.8}$$

さらに ,  $\hat{P}_{T_{(i+1)'}}(x_{(i+1)'})$ ,  $\hat{P}_{U_{(i-1)'}}(x_{(i-1)'})$  に関して , 次の漸化式 (6.9), (6.10) が成立する .

$$\begin{aligned}
 \hat{P}_{T_{(i+1)'}}(x_{(i+1)'}) &= \sum_{x_1', \dots, x_{i'}} \exp \left[ \sum_{j=1}^i \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] \\
 &= \sum_{x_{i'}} \left( \sum_{x_1', \dots, x_{(i-1)'}} \exp \left[ \sum_{j=1}^{i-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] \right) \cdot \exp [\phi_{i',(i+1)'}(x_{i'}, x_{(i+1)'})] \\
 &= \sum_{x_{i'}} \hat{P}_{T_{i'}}(x_{i'}) \cdot \exp [\phi_{i',(i+1)'}(x_{i'}, x_{(i+1)'})] \tag{6.9}
 \end{aligned}$$

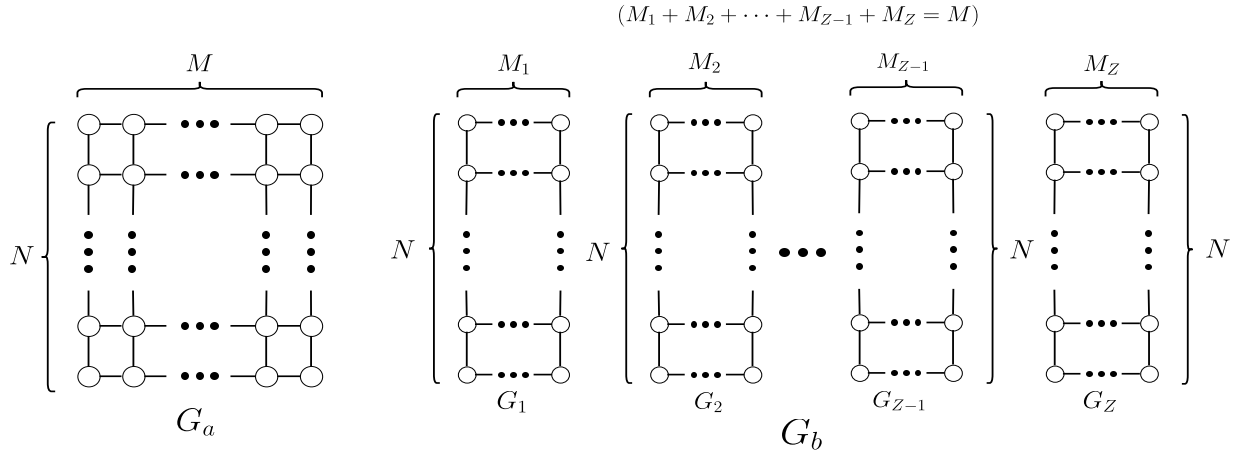
$$\begin{aligned}
 \hat{P}_{U_{(i-1)'}}(x_{(i-1)'}) &= \sum_{x_{i'}, \dots, x_{n'}} \exp \left[ \sum_{j=i-1}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] \\
 &= \sum_{x_{i'}} \left( \sum_{x_{(i+1)'}, \dots, x_{n'}} \exp \left[ \sum_{j=i}^{n-1} \phi_{j',(j+1)'}(x_{j'}, x_{(j+1)'}) \right] \right) \exp [\phi_{(i-1)',i'}(x_{(i-1)'}, x_{i'})] \\
 &= \sum_{x_{i'}} \hat{P}_{U_{i'}}(x_{i'}) \cdot \exp [\phi_{(i-1)',i'}(x_{(i-1)'}, x_{i'})] \tag{6.10}
 \end{aligned}$$

従って、一つ前のステップで導出した  $\hat{P}_{U_{i'}}(x_{i'}), \hat{P}_{T_{i'}}(x_{i'})$  を用いることで、 $\hat{P}_{U_{(i-1)'}}(x_{(i-1)'})$ ,  $\hat{P}_{T_{(i+1)'}}(x_{(i+1)'})$  を容易に導出することができる。漸化式 (6.9), (6.10) に従って  $\hat{P}_{T_{1'}}(x_{1'}), \dots, \hat{P}_{T_{n'}}(x_{n'}), \hat{P}_{U_{n'}}(x_{n'}), \dots, \hat{P}_{U_{1'}}(x_{1'})$  を順々に導出し、得られた計算結果を (6.6) 式に代入することで周辺分布が求まる。このような計算手法に基づいて、周辺分布の導出を行う手法をビリーフプロパゲーションと呼ぶ。

### 6.3 提案アルゴリズム

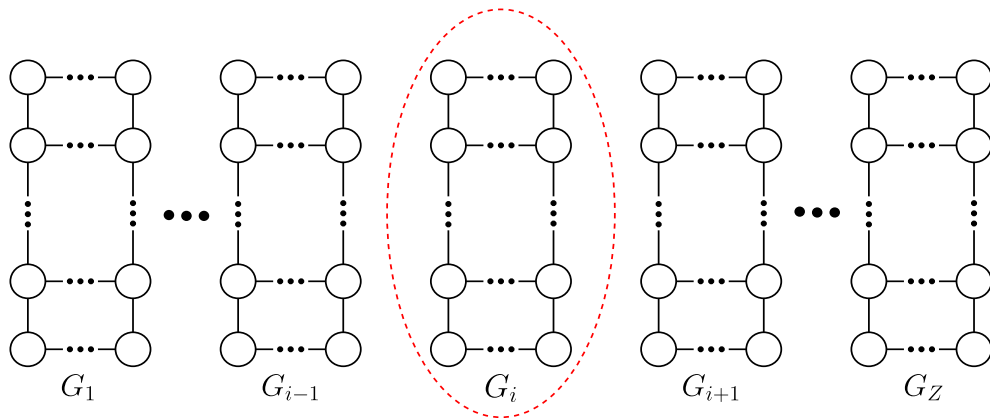
6.2 節で述べたビリーフプロパゲーションを用いて、条件式 (5.40) ~ (5.43) を満たすような近似分布 ( $e$ -射影点) が求まるようなアルゴリズムとして、以下を提案した。

[ビリーフプロパゲーションを用いた  $e$ -射影点導出アルゴリズム]



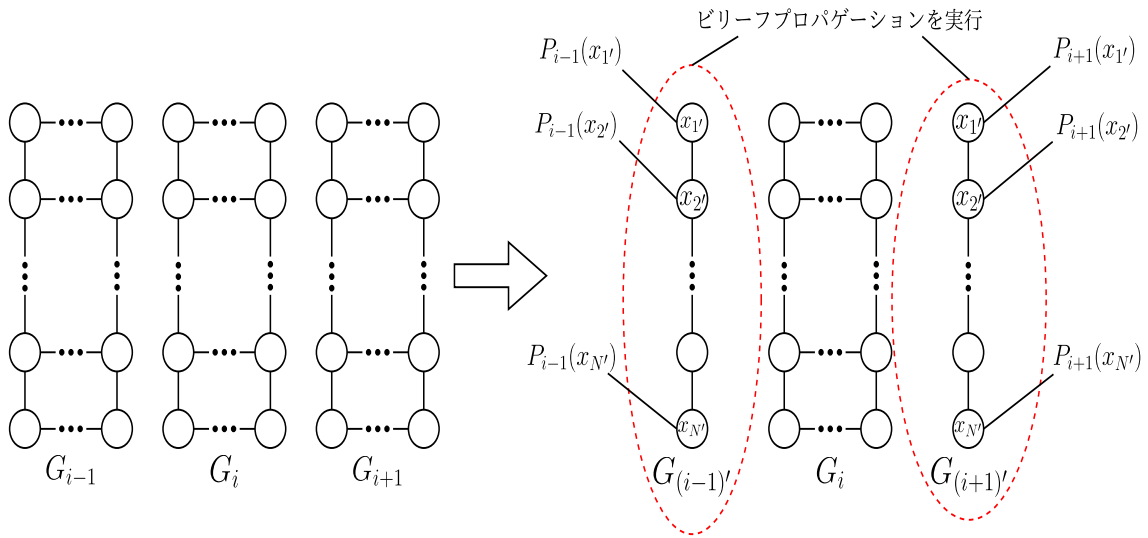
[Initialize]:  $t = 0, \bar{h}_v^{(0)} = h_v \quad (v \in V(G_b)), \bar{w}_e^{(0)} = w_e \quad (e \in E(G_b))$

step.1  $Z$  個のクラスターの中からクラスター  $G_i$  を任意に選ぶ。

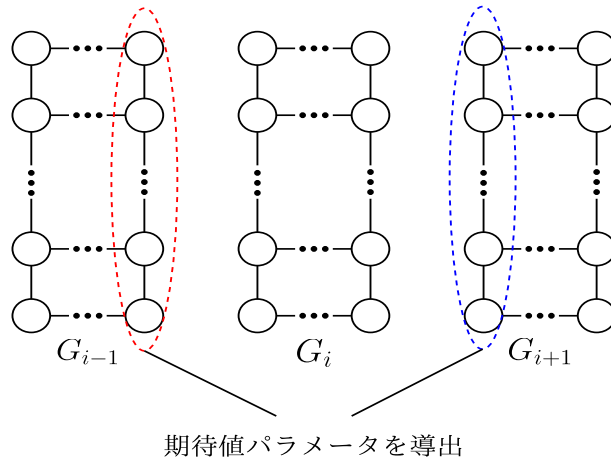


クラスター  $G_i$  を選択

step.2 クラスター  $G_i$  と隣接するクラスターを一次元鎖へと変形し、ビリーフプロパゲーションを用いて、一次元鎖の各頂点に関する周辺分布を導出する。

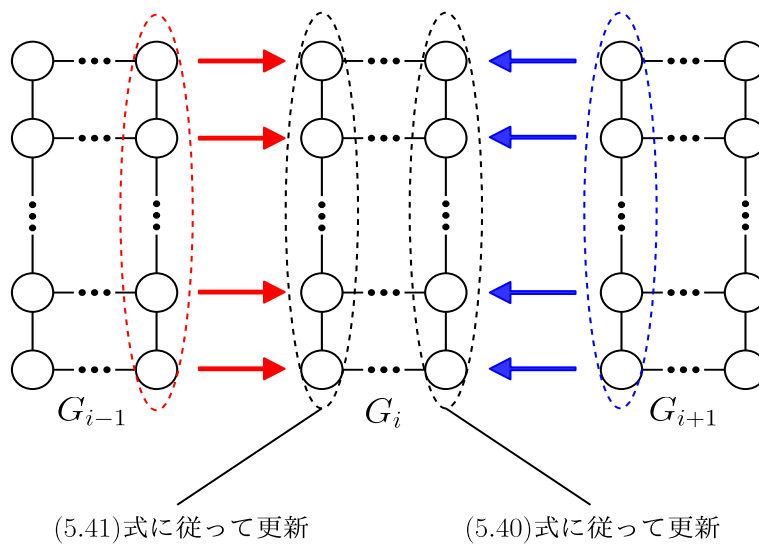


step.3 step.2 で導出した周辺分布を用いて，結合が切断された素子に関する期待値パラメータを導出する．

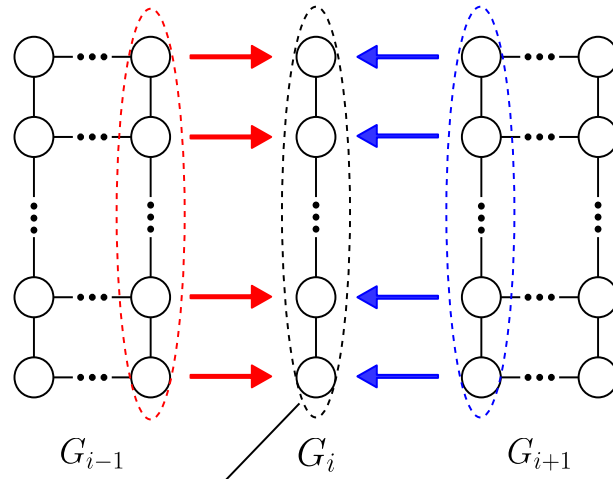


step.4 step.3 で導出した期待値パラメータを用いて，クラスター  $G_i$  の結合が切断された素子のしきい値を更新する．更新後の分布を  $q_{\theta^{(t+1)}}^{G_i}$  とする．

(i) クラスタ  $G_i$  の列数が2以上のとき



(ii) クラスタ  $G_i$  の列数が 1 の場合



(5.42)式に従って更新

step.5 分布  $q_{\theta}^{G_b}$  に関するノルム (7.1) 式を計算し, 以下の条件式

$$\left\| q_{\theta}^{G_b} \right\| < \epsilon \quad (\epsilon \text{ は十分小さな正数}) \quad (6.11)$$

を満たすならばアルゴリズムを終了. でなければ,  $t = t + 1$  として step.1 へ



## 第7章 数値実験

この章では、6章で提案したアルゴリズムの挙動に関するシミュレーションを行う。また、平均場近似により導出される近似分布よりも提案アルゴリズムにより導出される近似分布の方が近似精度が高いこと。さらには、提案アルゴリズムの計算時間などについても計算機上のシミュレーションで確認を行う。

### 7.1 実行環境と各種パラメータの設定について

本章におけるシミュレーションは、表 7.1 に示す計算機環境で行った。

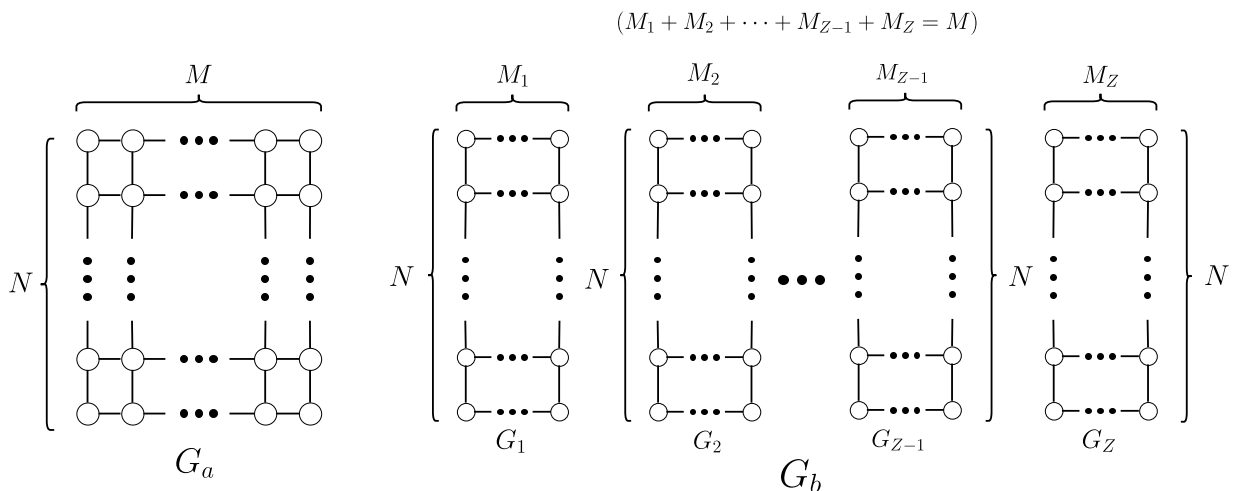
表 7.1: 計算機環境

OS	Ubuntu Mate 16.04.3 LTS
CPU	Intel(R) Core(TM) i7-5600U 2.60GHz
メモリ	8GB
使用言語	C 言語

また、本章で行うシミュレーションでは、しきい値と結合定数は一様分布に従って生成される  $[0, 1]$  の値をランダムに設定している。さらに、提案アルゴリズムの step.5 で用いているパラメータ  $\epsilon$  (十分小さな正数) は  $10^{-6}$  としている。

### 7.2 提案アルゴリズムの挙動に関するシミュレーション

#### 7.2.1 シミュレーション内容



この節では、6章で提案したアルゴリズムを実行することで所望の近似分布が導出されることを計算機上のシミュレーションで確認する。提案アルゴリズムの目的は、格子模型  $G_a$  上で定義される平衡確率分布  $P_\theta^{G_a} \in \mathcal{M}(G_a)$  が一つ与えられている時に、この分布  $P_\theta^{G_a} \in \mathcal{M}(G_a)$

から分布族  $\mathcal{M}(G_b)$  への  $e$ -射影  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  を導出することであった。5.3 節で  $e$ -射影点  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  が満たす条件式として (5.40) ~ (5.43) 式を導出しており、さらに、条件式 (5.43) に関しては提案アルゴリズムの初期化部分 ([Initialize]) を実行することで必ず満たされることが明らかである。従って、提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  が  $e$ -射影の条件式 (5.40) ~ (5.42) を満たしていることを確認すれば、提案アルゴリズムは意図したとおりに動作していると見做せる。本シミュレーションでは、提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  が条件式 (5.40) ~ (5.42) をどの程度満たしているのかを確認するための指標として、以下のノルムを導入する。

$$\begin{aligned}
 & \left\| q_{\hat{\theta}}^{G_b} \right\| \\
 &= \sum_{e \in C_1} \left| \hat{\theta}^{\ell(e)} - \left( \theta^{\ell(e)} + \theta^e \hat{\eta}_{r(e)}(\hat{\theta}) \right) \right| + \sum_{f \in C_{Z-1}} \left| \hat{\theta}^{r(f)} - \left( \theta^{r(f)} + \theta^f \hat{\eta}_{\ell(f)}(\hat{\theta}) \right) \right| \\
 &+ \sum_{i=2: M_i \geq 2}^{Z-1} \left( \sum_{e \in C_{i-1}} \left| \hat{\theta}^{r(e)} - \left( \theta^{r(e)} + \theta^e \hat{\eta}_{\ell(e)}(\hat{\theta}) \right) \right| + \sum_{f \in C_i} \left| \hat{\theta}^{\ell(f)} - \left( \theta^{\ell(f)} + \theta^f \hat{\eta}_{r(f)}(\hat{\theta}) \right) \right| \right) \\
 &+ \sum_{i=2: M_i=1}^{Z-1} \sum_{\substack{(e,f) \in C_{i-1} \times C_i \\ : r(e)=\ell(f)}} \left| \hat{\theta}^{r(e)} - \left( \theta^{r(e)} + \theta^e \hat{\eta}_{\ell(e)}(\hat{\theta}) + \theta^f \hat{\eta}_{r(f)}(\hat{\theta}) \right) \right| \quad (7.1)
 \end{aligned}$$

記号の濫用により、(7.1) 式の意味するところが理解しにくいと思われるため、各記号の意味を以下にまとめておく。

- $C_i, C_{i-1}$  : グラフ  $G_a$  に含まれる辺の中で、グラフ  $G_i$  ( $G_{i-1}$ ) に含まれる頂点とグラフ  $G_{i+1}$  ( $G_i$ ) に含まれる頂点を結ぶ辺の集合 (カット)
- $\ell(e), r(e)$  : 辺  $e$  の左側, 右側の端点
- $\theta^e$  : グラフ  $G_a$  上で定義されるボルツマンマシンの平衡確率分布  $P_{\theta}^{G_a} \in \mathcal{M}(G_a)$  に関する自然パラメータであり、辺  $e$  の端点の間の結合定数
- $\theta^{\ell(e)}, \theta^{\ell(f)}$  : グラフ  $G_a$  上で定義されるボルツマンマシンの平衡確率分布  $P_{\theta}^{G_a} \in \mathcal{M}(G_a)$  に関する自然パラメータであり、グラフ  $G_a$  をグラフ  $G_b$  のように分割する際に、結合が切断される辺  $e, f$  の左側の頂点に関するしきい値
- $\theta^{r(e)}, \theta^{r(f)}$  : グラフ  $G_a$  上で定義されるボルツマンマシンの平衡確率分布  $P_{\theta}^{G_a} \in \mathcal{M}(G_a)$  に関する自然パラメータであり、グラフ  $G_a$  をグラフ  $G_b$  のように分割する際に、結合が切断される辺  $e, f$  の右側の頂点に関するしきい値
- $\hat{\theta}^{\ell(e)}, \hat{\theta}^{\ell(f)}$  : 提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  に関する自然パラメータであり、グラフ  $G_a$  をグラフ  $G_b$  のように分割する際に、結合が切断される辺  $e, f$  の左側の頂点に関するしきい値
- $\hat{\theta}^{r(e)}, \hat{\theta}^{r(f)}$  : 提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  に関する自然パラメータであり、グラフ  $G_a$  をグラフ  $G_b$  のように分割する際に、結合が切断される辺  $e, f$  の右側の頂点に関するしきい値
- $\hat{\eta}_{\ell(e)}(\hat{\theta}), \hat{\eta}_{\ell(f)}(\hat{\theta})$  : 提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  に関する期待値パラメータであり、グラフ  $G_a$  をグラフ  $G_b$  のように分割する際に、結合が切断される辺  $e, f$  の左側の頂点に関する期待値パラメータ

- $\hat{\eta}_{r(e)}(\hat{\theta}), \hat{\eta}_{r(f)}(\hat{\theta})$  : 提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  に関する期待値パラメータであり, グラフ  $G_a$  をグラフ  $G_b$  のように分割する際に, 結合が切断される辺  $e, f$  の右側の頂点に関する期待値パラメータ

(7.1) 式の第一項目の絶対値の中身は, クラスタ  $G_1$  の右端の素子のしきい値に関する  $e$ -射影の条件式 (5.40) の(左辺)-(右辺)を表している. 第二項目の絶対値の中身は, クラスタ  $G_Z$  の左端の素子のしきい値に関する  $e$ -射影の条件式 (5.41) の(左辺)-(右辺)を表している. 第三項目の絶対値の中身は, 列数が 2 以上のクラスタ  $G_i (2 \leq i \leq Z-1)$  の左端の素子のしきい値に関する  $e$ -射影の条件式 (5.41) の(左辺)-(右辺)を表している. 第四項目の絶対値の中身は, 列数が 2 以上のクラスタ  $G_i (2 \leq i \leq Z-1)$  の右端の素子のしきい値に関する  $e$ -射影の条件式 (5.40) の(左辺)-(右辺)を表している. 第五項目の絶対値の中身は, 列数が 1 のクラスタ  $G_i (2 \leq i \leq Z-1)$  に含まれる素子のしきい値に関する  $e$ -射影の条件式 (5.42) の(左辺)-(右辺)を表している. 従って, 導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  が  $e$ -射影である場合, すべての項は 0 となる. (7.1) 式のノルムは定義の仕方より, 以下の条件式を満たす.

$$\forall q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b), \quad \left\| q_{\hat{\theta}}^{G_b} \right\| \geq 0 \quad (7.2)$$

$$\left\| q_{\hat{\theta}}^{G_b} \right\| = 0 \iff q_{\hat{\theta}}^{G_b} \text{は } P_{\hat{\theta}}^{G_a} \in \mathcal{M}(G_a) \text{ から } \mathcal{M}(G_b) \text{ への } e\text{-射影}$$

従って, (7.1) 式のノルムの値が小さいほど近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  は  $e$ -射影点  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  に近いものとして解釈される.

### 7.2.2 結果と考察

表 7.2: 提案アルゴリズムの挙動に関するシミュレーション

	グラフ $G_a$ のサイズ	クラスタの個数 $Z$	各クラスタの列数
パターン 1	$3 \times 12$	4	$M_i = 3 (i = 1, 2, 3, 4)$
パターン 2	$10 \times 60$	15	$M_i = 4 (1 \leq i \leq 15)$
パターン 3	$40 \times 4$	2	$M_i = 2 (i = 1, 2)$
パターン 4	$100 \times 4$	2	$M_i = 2 (i = 1, 2)$

提案アルゴリズムの挙動を確かめるために, 表 7.2 に記載されている 4 パターンに関してそれぞれシミュレーションを行った. パターン 1 からパターン 4 に対する結果のグラフ・表はそれぞれ以下の通りである.

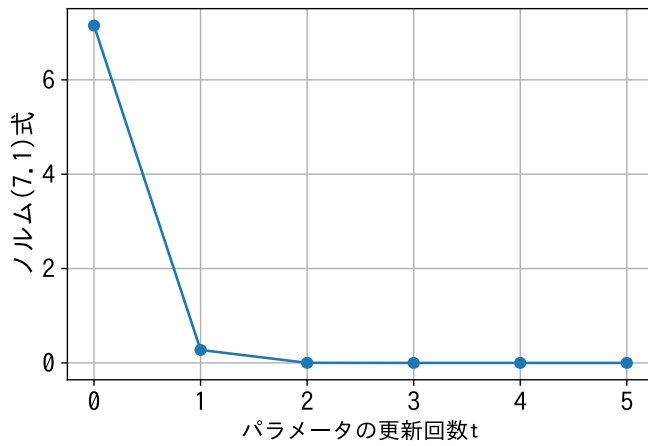


図 7.1: パターン 1 に関する結果

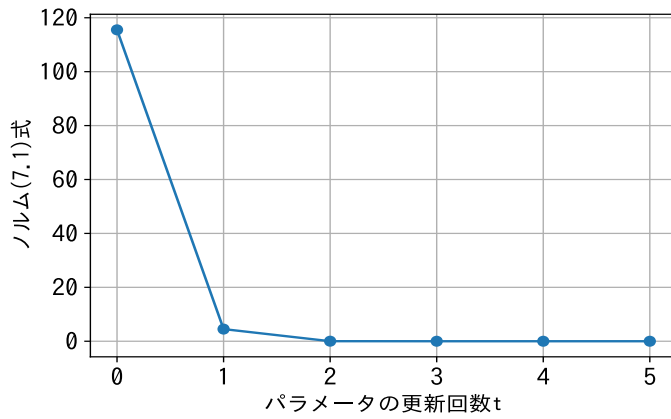
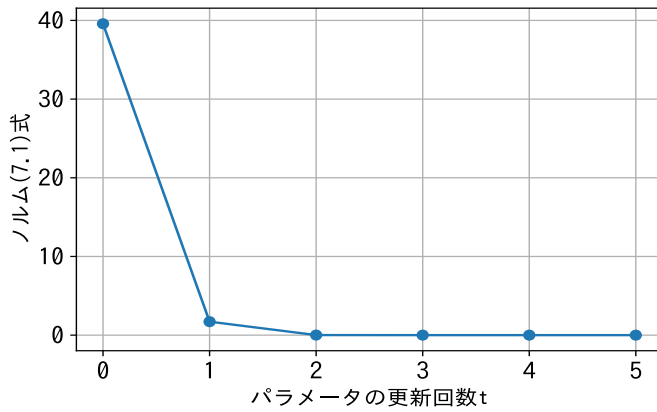
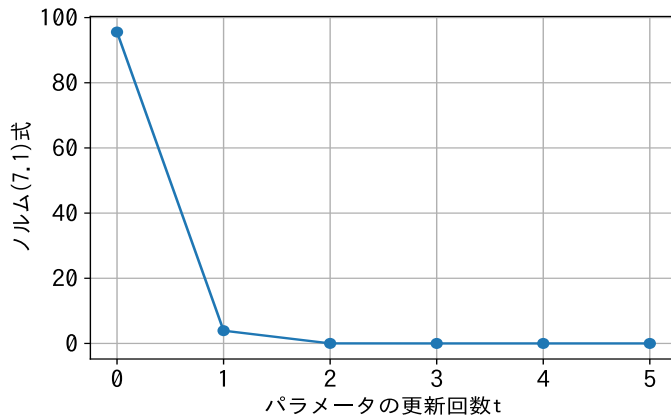


図 7.2: パターン 2 に関する結果



更新回数 $t$	ノルム (7.1) 式の値
0	39.5
1	1.7
2	$9.6 \times 10^{-3}$
3	$7.8 \times 10^{-5}$
4	$7.8 \times 10^{-7}$
5	$8.8 \times 10^{-9}$

図 7.3: パターン 3 に関する結果



更新回数 $t$	ノルム (7.1) 式の値
0	95.5
1	3.9
2	$1.5 \times 10^{-2}$
3	$7.5 \times 10^{-5}$
4	$4.4 \times 10^{-7}$
5	$3.1 \times 10^{-9}$

図 7.4: パターン 4 に関する結果

全てのパターンに対して、提案アルゴリズムを進めていくことでノルム (7.1) 式の値が指数関数的に 0 に近づいていることが分かる。従って、提案アルゴリズムは  $P_{\theta}^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_b)$  への  $e$ -射影  $q_{\theta}^{G_b} \in \mathcal{M}(G_b)$  を導出するアルゴリズムとして、上手く機能していると解釈される。

### 7.3 平均場近似と提案アルゴリズムの精度比較に関するシミュレーション

#### 7.3.1 シミュレーション内容

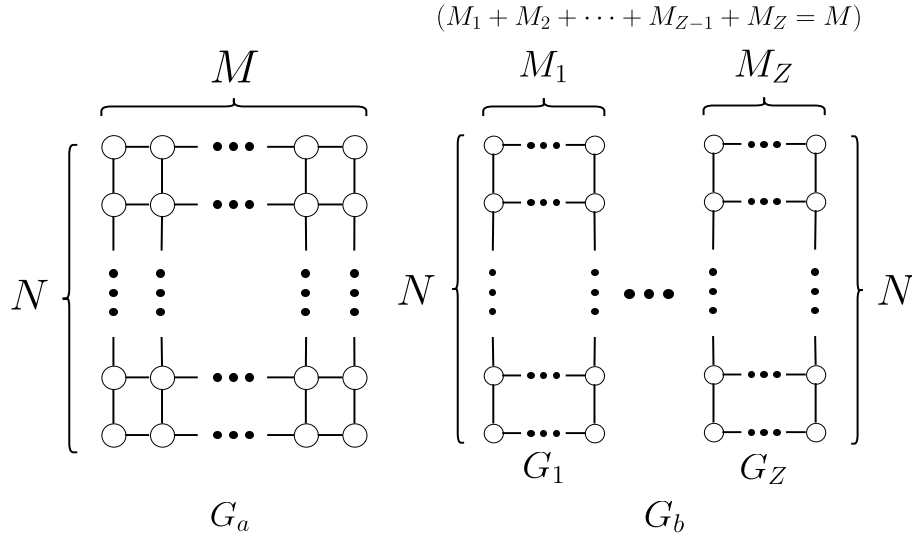


図 7.5: 格子模型  $G_a$ , クラスタ型模型  $G_b$

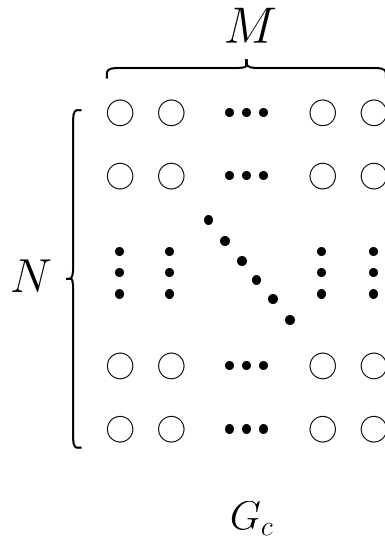


図 7.6: 結合が存在しないグラフ  $G_c$

この節では、格子模型  $G_a$  のサイズ（行数  $\times$  列数）、クラスタ型模型  $G_b$  に含まれるクラスタの個数  $Z$ 、各クラスタの列数をそれぞれ表 7.3 のように設定し、各設定の下で、提案アルゴリズムにより導出される近似分布  $q_{\theta}^{G_b} \in \mathcal{M}(G_b)$  と平均場近似により導出される近似分布  $q_{\bar{h}}^{G_c} \in \mathcal{M}(G_c)$  の精度の比較を行った。また、本シミュレーションでは、提案アルゴリズムと平均場近似により導出される近似分布  $q_{\theta}^{G_b} \in \mathcal{M}(G_b)$ 、 $q_{\bar{h}}^{G_c} \in \mathcal{M}(G_c)$  の精度を比較するための指標として、Kullback-Leibler ダイバージェンス  $D(\cdot \| P_{\theta}^{G_a})$  を用いた（但し、 $P_{\theta}^{G_a}$  は格子模型  $G_a$  上で定義されるボルツマンマシンの平衡確率分布である）。この Kullback-Leibler ダイバージェンスの値が小さいほど、元の平衡確率分布  $P_{\theta}^{G_a} \in \mathcal{M}(G_a)$  の近似分布として精度が高いと見做される。

## 7.3.2 結果と考察

表 7.3: 平均場近似と提案アルゴリズムの精度比較

グラフ $G_a$ のサイズ	クラスターの個数 $Z$	各クラスターの列数	$D(q_{\hat{\theta}}^{G_b} \  P_{\theta}^{G_a})$	$D(q_{\hat{h}}^{G_c} \  P_{\theta}^{G_a})$
$3 \times 6$	2	$M_i = 3 (i = 1, 2)$	$1.6 \times 10^{-3}$	$5.9 \times 10^{-2}$
	3	$M_i = 2 (i = 1, 2, 3)$	$1.4 \times 10^{-2}$	
	6	$M_i = 1 (1 \leq i \leq 6)$	$2.3 \times 10^{-2}$	
$3 \times 8$	2	$M_i = 4 (i = 1, 2)$	$6.0 \times 10^{-3}$	$1.0 \times 10^{-1}$
	4	$M_i = 2 (i = 1, 2, 3, 4)$	$1.7 \times 10^{-2}$	
	8	$M_i = 1 (1 \leq i \leq 6)$	$6.9 \times 10^{-2}$	
$3 \times 10$	2	$M_i = 5 (i = 1, 2)$	$1.0 \times 10^{-2}$	$9.2 \times 10^{-2}$
	5	$M_i = 2 (1 \leq i \leq 5)$	$2.2 \times 10^{-2}$	
	10	$M_i = 1 (1 \leq i \leq 10)$	$5.4 \times 10^{-2}$	

表 7.3 は、格子模型  $G_a$  のサイズ、クラスター型模型  $G_b$  に含まれるクラスターの個数、各クラスターの列数をそれぞれ変化させた時に、提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  と平均場近似により導出される近似分布  $q_{\hat{h}}^{G_c} \in \mathcal{M}(G_c)$  の精度をまとめた表である。すべてのパターンにおいて、提案アルゴリズムにより導出される近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  は、従来の平均場近似により導出される近似分布  $q_{\hat{h}}^{G_c} \in \mathcal{M}(G_c)$  よりも近似精度が上であるという結果となった。また、格子模型  $G_a$  のサイズを固定し、クラスターの個数  $Z$  を増加させた場合、近似分布  $q_{\hat{\theta}}^{G_b} \in \mathcal{M}(G_b)$  の精度は徐々に落ちていくことが確認できる。これは、クラスターの個数が増加していくにつれて、分布族  $\mathcal{M}(G_b)$  の範囲が狭まっていくことに起因している。

## 7.4 提案アルゴリズムの計算時間に関するシミュレーション

## 7.4.1 シミュレーション内容

各クラスターの列数と行数、クラスター型模型に含まれるクラスターの個数をそれぞれ増やしていく場合に、提案アルゴリズムの計算時間がどのように増加していくのかについて、計算機によるシミュレーションで確認をした。

## 7.4.2 結果と考察

表 7.4: 各クラスターの列数の増加に伴う計算時間の変化

クラスターの個数 $Z$	各クラスターの行数	各クラスターの列数	計算時間 [ミリ秒]
2	3	$M_i = 4 (i = 1, 2)$	7
2	3	$M_i = 5 (i = 1, 2)$	29
2	3	$M_i = 6 (i = 1, 2)$	132
2	3	$M_i = 7 (i = 1, 2)$	402
2	3	$M_i = 8 (i = 1, 2)$	$1.8 \times 10^3$
2	3	$M_i = 9 (i = 1, 2)$	$7.6 \times 10^3$
2	3	$M_i = 10 (i = 1, 2)$	$3.2 \times 10^4$
2	3	$M_i = 11 (i = 1, 2)$	$1.1 \times 10^5$
2	3	$M_i = 12 (i = 1, 2)$	$5.8 \times 10^5$
2	3	$M_i = 13 (i = 1, 2)$	$1.9 \times 10^6$
2	3	$M_i = 14 (i = 1, 2)$	$1.0 \times 10^7$

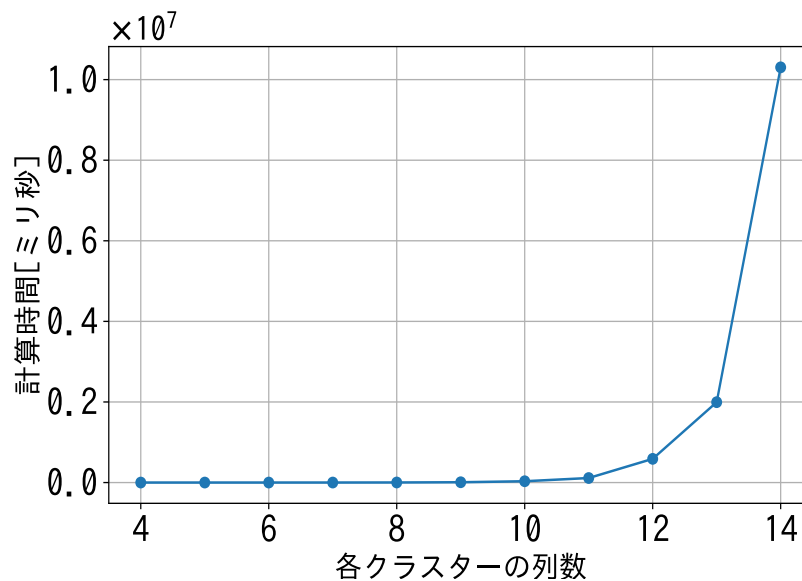


図 7.7: 列数の増加に伴う計算時間の変化

表 7.4 は、クラスターの個数、各クラスターの行数に関しては固定し、各クラスターの列数を 1 ずつ増やした場合の提案アルゴリズムの計算時間の増加具合を表したものである。図 7.7 は横軸を各クラスターの列数  $M_i$ 、縦軸を提案アルゴリズムの計算時間 [ミリ秒] とした時のグラフを表している。図 7.7 を見ると、アルゴリズムの計算時間は、各クラスターの列数の増加に関して指数関数的に増加していることが分かる。このように計算時間が一気に増加してしまう原因としては、提案アルゴリズムの step.2 におけるビリーフプロパゲーションの計算部分が挙げられる。



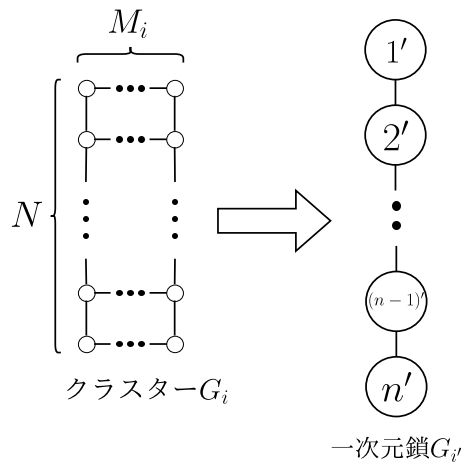


図 7.8: クラスタから一次元鎖への変形

図 7.8 は、クラスター型模型に含まれるクラスター  $G_i$  から一次元鎖  $G_{i'}$  への変形を表している。ピリーフプロパゲーションでは、一次元鎖  $G_{i'}$  の各頂点に関する周辺分布 (6.6) を導出している。また、この周辺分布 (6.6) を計算するために必要な項は、(6.9), (6.10) 式により導出される。クラスター  $G_i$  の列数が増加するにつれて、一次元鎖  $G_{i'}$  に含まれる各頂点の取りうる状態数は指数的に増加する。このとき、(6.9), (6.10) 式の計算時間も指数的に増大してしまう。このことから、図 7.7 のような結果が得られたと考えられる。

表 7.5: 各クラスタの行数の増加に伴う計算時間の変化

クラスタの個数 $Z$	各クラスタの行数	各クラスタの列数	計算時間 [ミリ秒]
2	10	$M_i = 4 (i = 1, 2)$	27
2	20	$M_i = 4 (i = 1, 2)$	69
2	30	$M_i = 4 (i = 1, 2)$	90
2	40	$M_i = 4 (i = 1, 2)$	112
2	50	$M_i = 4 (i = 1, 2)$	141
2	60	$M_i = 4 (i = 1, 2)$	168
2	70	$M_i = 4 (i = 1, 2)$	197
2	80	$M_i = 4 (i = 1, 2)$	236
2	90	$M_i = 4 (i = 1, 2)$	273
2	100	$M_i = 4 (i = 1, 2)$	287

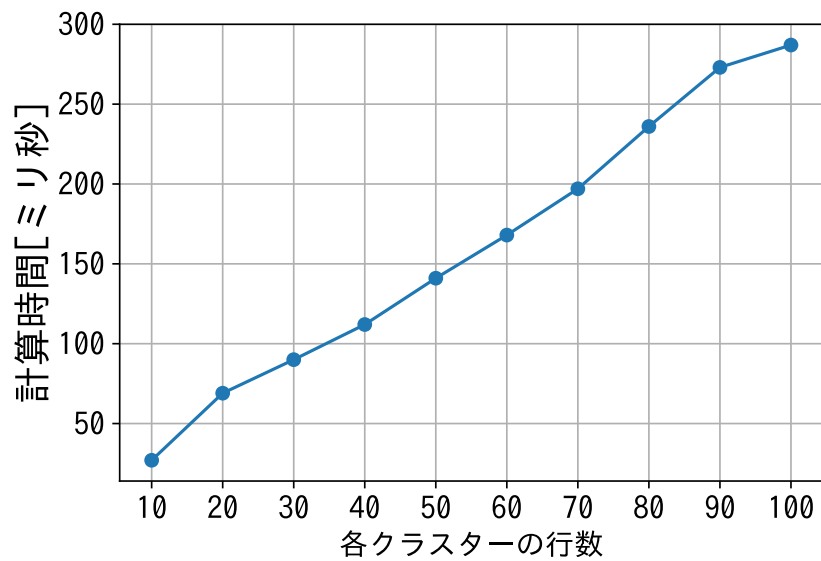


図 7.9: 行数の増加に伴う計算時間の変化

表 7.5 はクラスターの個数，各クラスターの列数に関しては固定し，各クラスターの行数を 10 ずつ増やした時の提案アルゴリズムの計算時間の増加具合を表したものである．図 7.9 は横軸を各クラスターの行数，縦軸を提案アルゴリズムの計算時間 [ミリ秒] とした場合のグラフである．図 7.9 を見ると，アルゴリズムの計算時間は，各クラスターの行数に関して線形関数的に増加していることが分かる．表 7.4 と比較して，表 7.5 ではクラスター型モデルのサイズ（素因数）がかなり大きいものをシミュレーションの対象としている．しかし，アルゴリズムの計算時間は表 7.4 と比べて，かなり少ない時間で済んでおり，計算機側での演算の限界（オーバーフロー）などが生じない限り，行数をさらに大きくとっても現実的な時間でアルゴリズムが終了すると考えられる．

表 7.6: クラスター数の増加に伴う計算時間の変化

クラスターの個数 $Z$	各クラスターの行数	各クラスターの列数	計算時間 [ミリ秒]
10	10	$M_i = 2 (1 \leq i \leq 10)$	7
20	10	$M_i = 2 (1 \leq i \leq 20)$	16
30	10	$M_i = 2 (1 \leq i \leq 30)$	24
40	10	$M_i = 2 (1 \leq i \leq 40)$	35
50	10	$M_i = 2 (1 \leq i \leq 50)$	51
60	10	$M_i = 2 (1 \leq i \leq 60)$	64
70	10	$M_i = 2 (1 \leq i \leq 70)$	68
80	10	$M_i = 2 (1 \leq i \leq 80)$	84
90	10	$M_i = 2 (1 \leq i \leq 90)$	89
100	10	$M_i = 2 (1 \leq i \leq 100)$	94

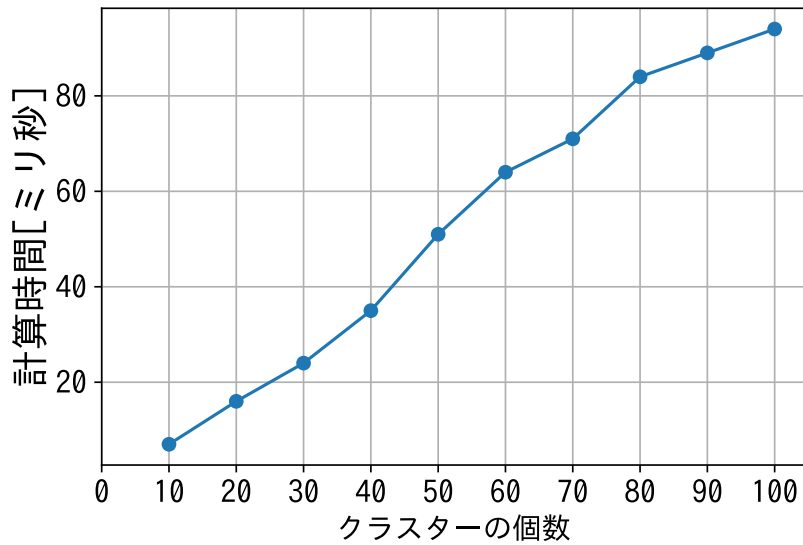


図 7.10: クラスタ数の増加に伴う計算時間を表したグラフ

表 7.6 は各クラスタの行数・列数に関しては固定し、クラスタの個数を 10 ずつ増やした時の提案アルゴリズムの計算時間の増加具合を表したものである。図 7.10 は横軸をクラスタの個数、縦軸を提案アルゴリズムの計算時間 [ミリ秒] とした場合のグラフである。図 7.10 を見ると、提案アルゴリズムの計算時間は、クラスタの個数に関して線形関数的に増加していることがわかる。従って、元のボルツマンマシンのサイズが非常に大きい場合に、クラスタの個数が非常に多いクラスタ型模型を用いて提案アルゴリズムを実行することで、近似分布を現実的な計算時間で導出することが可能となる。

## 第8章 結論

### 8.1 本論文の成果

本論文では、平均場近似の情報幾何学的な仕組みをベースにしつつ、平均場近似の精度を向上させるために、クラスター型模型というモデルを導入した。そして、格子模型型のボルツマンマシンの平衡確率分布から部分空間（クラスター型模型上で定義される平衡確率分布を要素とする分布族）への  $e$ -射影が満たす条件式を導出し、この条件式を満たす近似分布を数値計算で導出するためのアルゴリズムを提案した。その後、提案したアルゴリズムにより導出される近似分布が  $e$ -射影の条件式を満たすこと、提案アルゴリズムにより導出される近似分布は平均場近似により導出される近似分布よりも近似精度が高いことを計算機によるシミュレーションで確認した。

さらに、提案アルゴリズムの計算量オーダーを確かめるために、クラスター型模型に含まれるクラスターの個数  $Z$ 、各クラスターの列数（縦のサイズ） $M_1 \sim M_Z$ 、各クラスターの行数（横のサイズ） $N$  をそれぞれ増加させたときのアルゴリズムの処理時間を計測した。シミュレーションの結果、各クラスターの列数  $M_1 \sim M_Z$  の中の最大値を  $\hat{M}$  としたとき、提案アルゴリズムの計算量オーダーは  $\mathcal{O}(ZN2^{\hat{M}})$  であることを確認できた。このことから、元の格子模型型のボルツマンマシンのサイズが非常に大きい場合でも、クラスターの個数が非常に多いクラスター型模型を与え、提案アルゴリズムを実行することで平均場近似よりも精度の高い近似分布を現実的な計算時間で導出できると考えられる。

### 8.2 今後の課題

今後の課題は、ボルツマンマシンを量子系へと拡張させることである。本論文で扱っているボルツマンマシンは、各素子が  $\{0, 1\}$  の2状態を取りうるネットワークであった。それに対し、量子系へと拡張させたボルツマンマシンでは、各素子が取りうる状態は密度行列で表される。このとき、格子模型型のボルツマンマシンとクラスター型模型を量子系へと拡張させた時の  $e$ -射影の条件式は、5章で導出したものと等価となる。従って、量子系に拡張させたボルツマンマシンに対しても、提案アルゴリズムのフレームワークをほとんど変えることなく、適用することができると考えられる。

# 付録 A Xing らの研究 [2] について

本付録では，Xing らが導出した一般化平均場方程式 (generalized mean field equation) と本論文で導出した  $e$ -射影の条件式が結果的に等価となることを示す．

## A.1 準備

この節では，一般化平均場方程式について簡潔に述べるために，いくつかの概念を導入する．

まず，頂点の集合を  $V$  とし，集合  $A$  を以下のように定義する．

$$A := \{\{u, v\} \mid u, v \in V\} \quad (\text{A.1})$$

このとき，頂点集合  $V$  と部分集合  $E \subseteq A$  の組  $G = (V, E)$  を無向グラフと呼ぶ．以降の議論で扱うグラフは全て無向グラフであり，グラフ  $G$  に関する頂点集合を  $V(G)$ ，辺集合を  $E(G)$  として記述する．無向グラフ  $G$  上で確率分布を定めるとき，頂点集合  $V(G)$  に含まれる各頂点は確率変数として捉えられる．本付録では，無向グラフ上の頂点と確率変数との間の対応関係を次のように定める．まず，頂点  $v \in V(G)$  に対応する確率変数を  $x_v$ ，部分集合  $S \subseteq V(G)$  に含まれる頂点に対応する確率変数の集合を  $x_S$ ，無向グラフ上のすべて頂点に関する確率変数の集合を  $x = x_V$  とする．

グラフ  $G$  の頂点集合  $V(G)$  に対する分割 (クラスタリング)

$$\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_I\} \quad (\text{A.2})$$

が与えられているとする．このとき，クラスタリング  $\mathcal{C}$  に含まれる要素  $\mathcal{C}_i$  ( $1 \leq i \leq I$ ) のことをクラスターと呼ぶ．部分集合  $S \subseteq V(G)$  に含まれる任意の 2 頂点に対して，それらを接続する辺が  $E(G)$  に含まれるとき， $S$  をクリークと呼ぶ．グラフ  $G$  上のすべてのクリークの集合を

$$\mathcal{D} = \{D_1, \dots, D_K\} \quad (\text{A.3})$$

とする．さらに，クラスター  $\mathcal{C}_i$  の境界に位置するクリークの集合を

$$\mathcal{B}_i := \{D_\beta \in \mathcal{D} \mid D_\beta \cap \mathcal{C}_i \neq \emptyset, D_\beta \not\subseteq \mathcal{C}_i\} \quad (\text{A.4})$$

と定義する．クラスターの境界に位置するクリークとは，複数のクラスターと共通部分を持つクリークのことである．以下の図 A.1 は，クラスター  $\mathcal{C}_i, \mathcal{C}_\ell, \mathcal{C}_k$  の境界に位置するクリーク  $D_\beta$  を表している．

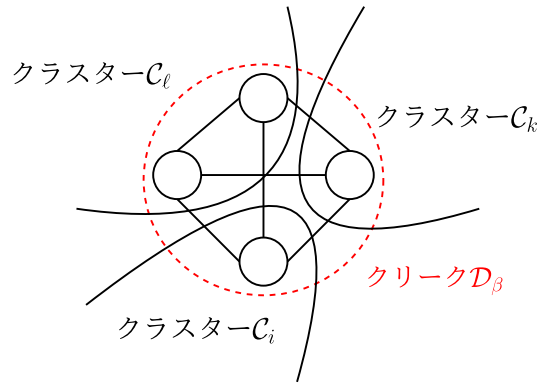


図 A.1: クラスターの境界に位置するクリーク

次に，クラスターのマルコフブランケット (Markov blanket) を以下で定義する．

定義 1 (クラスターのマルコフブランケット)

無向グラフ  $G$  と頂点集合  $V(G)$  の分割

$$\mathcal{C} = \{C_1, \dots, C_I\} \quad (\text{A.5})$$

が与えられているとする．このとき，クラスター  $C_i$  のマルコフブランケット  $MB_i$  とは，他のクラスター  $C_j$  ( $j \neq i$ ) に含まれる頂点の中で，クラスター  $C_i$  に含まれる頂点と隣接している頂点の集合である．例えば，以下の図 A.2 のような無向グラフ  $G$  とクラスタリング  $\mathcal{C}$  が与えられている場合，クラスター  $C_i$  のマルコフブランケット  $MB_i$  は

$$MB_i = \{3, 4, 5, 6\} \quad (\text{A.6})$$

となる．

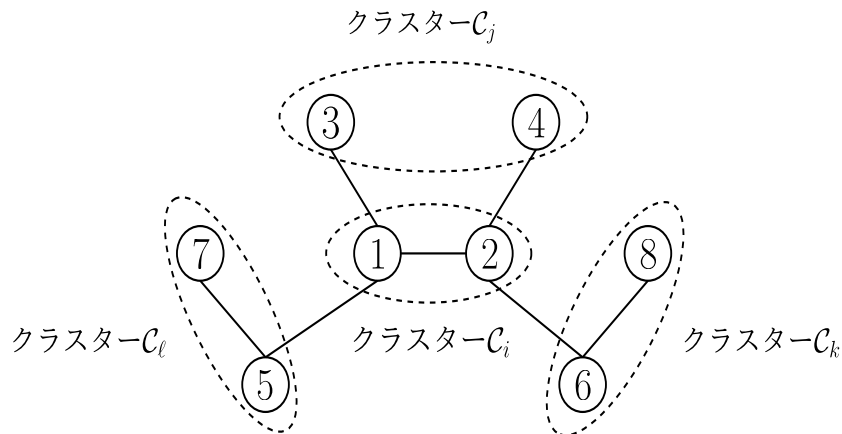


図 A.2: 4 個のクラスターに分割されている無向グラフ

無向グラフ上で定められる確率分布は，指数型分布族に含まれる分布の形式（ここでは，指数型表現と呼ぶ）で記述することができる．ここで，無向グラフ上の確率分布の指数型表現を以下のように定義する．

定義 2 (無向グラフ上の確率分布の指数型表現)

無向グラフ上のすべてのクリークの集合を  $\mathcal{D}$ ，各クリーク上で定義されるポテンシャル関数の集合を

$$\phi = \{\phi_\alpha \mid \alpha \in \mathcal{D}\} \quad (\text{A.7})$$

とする．さらに，ポテンシャル関数と同じ添字で表されるパラメータ集合を

$$\theta = \{\theta^\alpha \mid \alpha \in \mathcal{D}\} \quad (\text{A.8})$$

とおく．このとき，無向グラフ上で定められる確率分布  $p_\theta(x)$  は，以下の形式で表現される．

$$p_\theta(x) = \exp \left[ \sum_{\alpha \in \mathcal{D}} \theta^\alpha \phi_\alpha(x_{D_\alpha}) - A(\theta) \right] \quad (\text{A.9})$$

(A.9) 式を無向グラフ上の確率分布の指数型表現と呼ぶ．ただし， $A(\theta)$  は

$$A(\theta) = \log \left( \sum_x \exp \left[ \sum_{\alpha \in \mathcal{D}} \theta^\alpha \phi_\alpha(x_{D_\alpha}) \right] \right) \quad (\text{A.10})$$

としている．

次に，一般化平均場方程式を理解する上で重要な定義 3~5 を以下で導入する．

**定義 3 (クラスター因子分解可能なポテンシャル関数)**

クリーク  $D_\beta$  上のポテンシャル関数  $\phi_\beta(x_{D_\beta})$  が以下の形式で表現されるとき，ポテンシャル関数  $\phi_\beta(x_{D_\beta})$  はクラスター因子分解可能 (*cluster-factorizable*) であるという．

$$\phi_\beta(x_{D_\beta}) = F_\beta(\phi_{\beta_i}(x_{D_\beta \cap C_i}), \dots, \phi_{\beta_j}(x_{D_\beta \cap C_j})) \quad (\text{A.11})$$

ただし，(A.11) 式中の  $\phi_{\beta_i}(x_{D_\beta \cap C_i})$  は，クリーク  $D_\beta$  とクラスター  $C_i$  の共通部分に関するポテンシャル関数を表している．また，関数  $F_\beta(\cdot)$  は，与えられた引数の積もしくは和の形式で表現される関数である．

**定義 4 (平均場因子 (mean field factor))**

クラスター因子分解可能なポテンシャル関数  $\phi_\beta(x_{D_\beta})$  に対して，クリーク  $D_\beta$  と共通部分を持つクラスターの添字集合を  $I_\beta$  とおく．さらに，ポテンシャル関数  $\phi_\beta(x_{D_\beta})$  は以下のポテンシャル関数

$$\phi_{\beta_i}(x_{C_i \cap D_\beta}) \quad (i \in I_\beta) \quad (\text{A.12})$$

を因子として持つとする．このとき，平均場因子  $f_{i\beta}$  は，以下のように定義される．

$$f_{i\beta} := f_{i\beta}(x_{C_i \cap D_\beta}) := \langle \phi_{\beta_i}(x_{C_i \cap D_\beta}) \rangle_{q_i} \quad (i \in I_\beta) \quad (\text{A.13})$$

ここで， $\langle \cdot \rangle_{q_i}$  は，クラスター  $C_i$  に関する周辺分布  $q_i$  による期待値である．

**定義 5 (一般化平均場 (generalized mean field))**

無向グラフ  $G$  の頂点集合  $V(G)$  に関する分割 (クラスタリング)

$$\mathcal{C} = \{C_1, \dots, C_I\} \quad (\text{A.14})$$

が与えられているとする．このとき，クラスター  $C_j$  の一般化平均場  $\mathcal{F}_j$  は，以下で定義される．

$$\mathcal{F}_j := \{f_{i\beta} : D_\beta \in \mathcal{B}_j, i \in I_\beta, i \neq j\} \quad (\text{A.15})$$

## A.2 一般化平均場方程式

Xing らが導出した一般化平均場方程式は以下のとおりである。

定理 1 (一般化平均場方程式に関する定理)

無向グラフ  $G$  上で定められる確率分布  $p(x)$  と、頂点集合  $V(G)$  に関する分割

$$\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_I\} \quad (\text{A.16})$$

が与えられているとする。もし、クラスターの境界に位置するすべてのポテンシャル関数がクラスター因子分解可能であるとき、確率分布  $p(x)$  への一般化平均場近似により導出される近似分布  $q^{GMF}(x)$  は、以下の形式で表現される。

$$q^{GMF}(x) = \prod_{\mathcal{C}_i \in \mathcal{C}} q_i^{GMF}(x_{\mathcal{C}_i}) \quad (\text{A.17})$$

ただし、 $q_i^{GMF}(x_{\mathcal{C}_i})$  はクラスター  $\mathcal{C}_i$  に関する周辺分布であり、以下の条件式を満たす。

$$\forall_i (1 \leq i \leq I), \quad q_i^{GMF}(x_{\mathcal{C}_i}) = p(x_{\mathcal{C}_i} | \mathcal{F}_i) \quad (\text{A.18})$$

条件式 (A.18) のことを一般化平均場方程式と呼ぶ。

## A.3 本論文との関係性

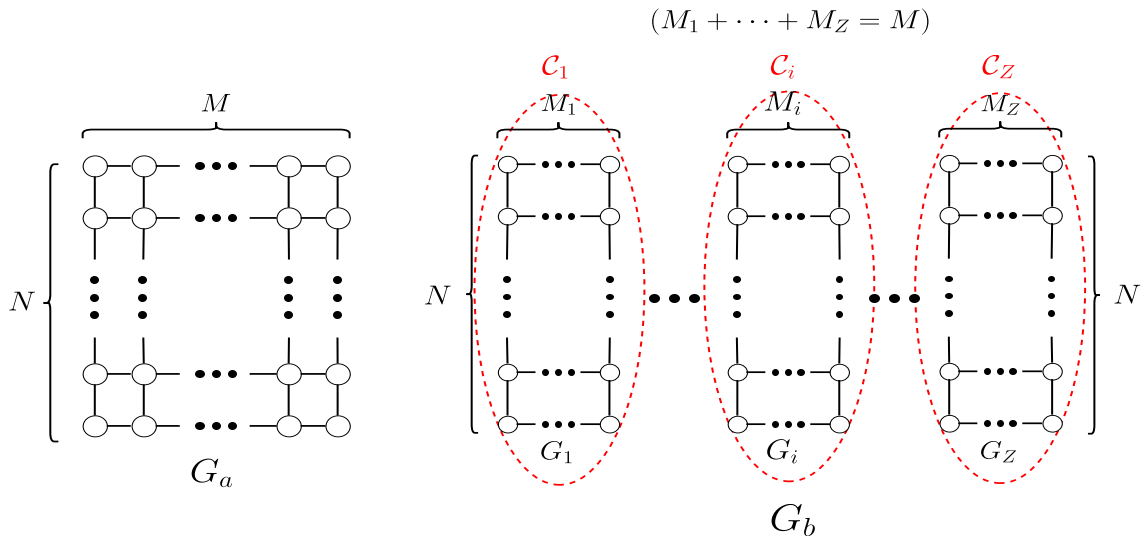


図 A.3:  $N$  行  $M$  列格子模型とクラスター型模型

図 A.3 は、5 章で導入した  $N$  行  $M$  列格子模型  $G_a$  とクラスター型模型  $G_b$  を表している。Xing らの論文中の記号を用いると、クラスター型模型  $G_b$  は、グラフ  $G_a$  にクラスタリング

$$\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_Z\} \quad (\forall_i, \mathcal{C}_i = V(G_i)) \quad (\text{A.19})$$

が与えられた時の状況と捉えることができる。さらに、5.2 節でのポテンシャル関数の定義より、クラスターの境界に位置するすべてのポテンシャル関数はクラスター因子分解可能である。従って、5.2 節で導入した問題設定において、定理 1 を適用することが可能である。このとき、一般化平均場方程式 (A.18) の右辺は、ボルツマンマシン  $G_a$  の平衡確率分布に関する条件付き周辺分布  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  となる。本研究では、ボルツマンマシン  $G_a$  の平衡確



率分布  $P_\theta^{G_a} \in \mathcal{M}(G_a)$  に対する近似分布として,  $P_\theta^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}(G_b)$  への  $e$ -射影  $q_\theta^{G_b} \in \mathcal{M}(G_b)$  を用いている. さらに 5.3 節では, この  $e$ -射影が満たす条件式として, 以下の (5.40) ~ (5.43) 式を導出した.

$$\bar{\theta}^\mu = \begin{cases} \theta^\mu + \theta^e \bar{\eta}_{r(e)}(\bar{\theta}) & \text{if } \exists e \in C, \forall f \in C, \mu = \ell(e) \wedge \mu \neq r(f) & (5.40) \\ \theta^\mu + \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}) & \text{if } \exists e \in C, \forall f \in C, \mu = r(e) \wedge \mu \neq \ell(f) & (5.41) \\ \theta^\mu + \theta^e \bar{\eta}_{r(e)}(\bar{\theta}) + \theta^f \bar{\eta}_{\ell(f)}(\bar{\theta}) & \text{if } \exists e, f \in C, \mu = \ell(e) = r(f) & (5.42) \\ \theta^\mu & \text{otherwise} & (5.43) \end{cases}$$

一般化平均場方程式 (A.18) と  $e$ -射影の条件式 (5.40) ~ (5.43) が等価であることを示すためには,  $e$ -射影  $q_\theta^{G_b} \in \mathcal{M}(G_b)$  に関する周辺分布  $q_{\theta_i}^{G_i} \in \mathcal{M}(G_i)$  と条件付き周辺分布  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  が等しいことを確認すれば良い. ゆえに, 以降の議論では周辺分布  $q_{\theta_i}^{G_i} \in \mathcal{M}(G_i)$  と条件付き周辺分布  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  が等しいことを示してゆく.

まず, 以降の計算式を簡潔にするために, 集合  $G(i), L_i, R_i$  を以下のように定義する.

$$G(i) := V(G_i) \cup E(G_i) \quad (A.20)$$

$$L_i := \{v \in V(G_i) \mid \exists e \in C_i, v = \ell(e)\} \quad (A.21)$$

$$R_i := \{v \in V(G_{i+1}) \mid \exists e \in C_i, v = r(e)\} \quad (A.22)$$

$G(i)$  はグラフ  $G_i$  に含まれる頂点と辺を要素とする集合,  $L_i$  はカット  $C_i$  に含まれる辺  $e$  の左側の頂点  $\ell(e)$  を要素とする集合,  $R_i$  はカット  $C_i$  に含まれる辺  $e$  の右側の頂点  $r(e)$  を要素とする集合としている. このとき,  $e$ -射影  $q_\theta^{G_b} \in \mathcal{M}(G_b)$  に関する周辺分布  $q_{\theta_i}^{G_i} \in \mathcal{M}(G_i)$  は以下のように式変形される (ただし, ここではクラスター  $G_i$  の列数は 2 以上の場合を想

定している) .

$$\begin{aligned}
 & q_{\bar{\theta}_i}^{G_i}(x_i) \\
 &= \exp \left[ \sum_{\beta \in G(i)} \bar{\theta}_i^\beta G_\beta(x_i) - \Phi_i(\bar{\theta}_i) \right] \\
 &= \frac{1}{\exp [\Phi_i(\bar{\theta}_i)]} \times \exp \left[ \sum_{\beta \in G(i)} \bar{\theta}_i^\beta G_\beta(x_i) \right] \\
 &= \frac{1}{\exp [\Phi_i(\bar{\theta}_i)]} \times \exp \left[ \sum_{\beta \in G(i) \setminus (L_i \cup R_{i-1})} \bar{\theta}_i^\beta G_\beta(x_i) + \sum_{\beta \in L_i} \bar{\theta}_i^\beta G_\beta(x_i) + \sum_{\beta \in R_{i-1}} \bar{\theta}_i^\beta G_\beta(x_i) \right] \\
 &= \frac{1}{\exp [\Phi_i(\bar{\theta}_i)]} \times \exp \left[ \sum_{\alpha \in G(i) \setminus (L_i \cup R_{i-1})} \theta^\alpha G_\alpha(x_i) \right] \times \exp \left[ \sum_{e \in C_i} \bar{\theta}_i^{\ell(e)} G_{\ell(e)}(x_i) \right] \\
 &\quad \times \exp \left[ \sum_{e \in C_{i-1}} \bar{\theta}_i^{r(e)} G_{r(e)}(x_i) \right] \\
 &= \frac{1}{\exp [\Phi_i(\bar{\theta}_i)]} \times \exp \left[ \sum_{\alpha \in G(i) \setminus (L_i \cup R_{i-1})} \theta^\alpha G_\alpha(x_i) \right] \\
 &\quad \times \exp \left[ \sum_{e \in C_i} \left( \theta^{\ell(e)} + \theta^e \bar{\eta}_{r(e)}(\bar{\theta}_{i+1}) \right) G_{\ell(e)}(x_i) \right] \\
 &\quad \times \exp \left[ \sum_{e \in C_{i-1}} \left( \theta^{r(e)} + \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}_{i-1}) \right) G_{r(e)}(x_i) \right] \\
 &= \frac{1}{\exp [\Phi_i(\bar{\theta}_i)]} \times \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) \right] \times \exp \left[ \sum_{e \in C_i} \theta^e \bar{\eta}_{r(e)}(\bar{\theta}_{i+1}) G_{\ell(e)}(x_i) \right] \\
 &\quad \times \exp \left[ \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}_{i-1}) G_{r(e)}(x_i) \right] \\
 &= \frac{\exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e \bar{\eta}_{r(e)}(\bar{\theta}_{i+1}) G_{\ell(e)}(x_i) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}_{i-1}) G_{r(e)}(x_i) \right]}{\sum_{x_i} \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e \bar{\eta}_{r(e)}(\bar{\theta}_{i+1}) G_{\ell(e)}(x_i) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}_{i-1}) G_{r(e)}(x_i) \right]}
 \end{aligned} \tag{A.23}$$

上記の式変形において, 4 番目の等号では条件式 (5.43), 5 番目の等号では条件式 (5.40), (5.41) を用いている .

次に, 条件付き周辺分布  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  の導出を行う . まず, 格子模型型のボルツマンマシ

ン  $G_a$  の平衡確率分布  $P_\theta^{G_a}(x)$  を以下のように式変形する .

$$\begin{aligned}
 P_\theta^{G_a}(x) &= P_\theta^{G_a}(x_1, \dots, x_Z) \\
 &= \exp \left[ \sum_{\alpha \in G(a)} \theta^\alpha F_\alpha(x) - \Psi(\theta) \right] \\
 &= \frac{1}{\exp[\Psi(\theta)]} \times \exp \left[ \sum_{\alpha \in G(a)} \theta^\alpha F_\alpha(x) \right] \\
 &= \frac{1}{\exp[\Psi(\theta)]} \times \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) \right] \times \exp \left[ \sum_{e \in C_i} \theta^e F_e(x) \right] \times \exp \left[ \sum_{e \in C_{i-1}} \theta^e F_e(x) \right] \\
 &\quad \times \exp \left[ \sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x) \right] \\
 &= \frac{1}{\exp[\Psi(\theta)]} \times \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) \right] \times \exp \left[ \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) G_{r(e)}(x_{i+1}) \right] \\
 &\quad \times \exp \left[ \sum_{e \in C_{i-1}} \theta^e G_{\ell(e)}(x_{i-1}) G_{r(e)}(x_i) \right] \times \exp \left[ \sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x) \right] \quad (\text{A.24})
 \end{aligned}$$

(A.24) 式中のポテンシャル関数  $G_{r(e)}(x_{i+1}), G_{\ell(e)}(x_{i-1})$  を平均場因子 (期待値パラメータ) へと置き換えたときの条件付き分布  $P_\theta^{G_a}(x_1, \dots, x_Z | \mathcal{F}_i)$  は

$$\begin{aligned}
 P_\theta^{G_a}(x_1, \dots, x_Z | \mathcal{F}_i) &= \frac{1}{\exp[\Psi(\theta)]} \times \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) \right] \times \exp \left[ \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\bar{\theta}_{i+1}) \right] \\
 &\quad \times \exp \left[ \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}_{i-1}) G_{r(e)}(x_i) \right] \times \exp \left[ \underbrace{\sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x)}_{*1} \right] \quad (\text{A.25})
 \end{aligned}$$

となる . このとき , (A.25) 式中の  $\exp[\Psi(\theta)]$  は以下のように記述される .

$$\begin{aligned}
 \exp[\Psi(\theta)] &= \sum_{x_1, \dots, x_Z} \left( \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) \right] \times \exp \left[ \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\bar{\theta}_{i+1}) \right] \right. \\
 &\quad \left. \times \exp \left[ \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\bar{\theta}_{i-1}) G_{r(e)}(x_i) \right] \times \exp \left[ \underbrace{\sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x)}_{*1} \right] \right) \quad (\text{A.26})
 \end{aligned}$$

(A.25), (A.26) 式中の  $*1$  は変数  $x_i$  を含まない項である . 従って , (A.25) 式は以下のように

記述し直せる .

$$\begin{aligned}
 & P_\theta^{G_a}(x_1, \dots, x_z | \mathcal{F}_i) \\
 &= \frac{\exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\overline{\theta_{i+1}}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\overline{\theta_{i-1}}) G_{r(e)}(x_i) \right]}{\sum_{x_i} \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\overline{\theta_{i+1}}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\overline{\theta_{i-1}}) G_{r(e)}(x_i) \right]} \\
 & \times \frac{\exp \left[ \sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x) \right]}{\sum_{x \setminus x_i} \exp \left[ \sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x) \right]} \tag{A.27}
 \end{aligned}$$

ただし, (A.27) 式中の  $\sum_{x \setminus x_i}$  は, 状態変数  $x_i$  以外の全ての状態変数  $x_j (j \neq i)$  に関する総和を意味している . また, 条件付き周辺分布  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  は以下の等式を満たす .

$$P_\theta^{G_a}(x_i | \mathcal{F}_i) = \sum_{x \setminus x_i} P_\theta^{G_a}(x_1, \dots, x_z | \mathcal{F}_i) \tag{A.28}$$

(A.28) 式を用いることで,  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  は以下のように式変形される .

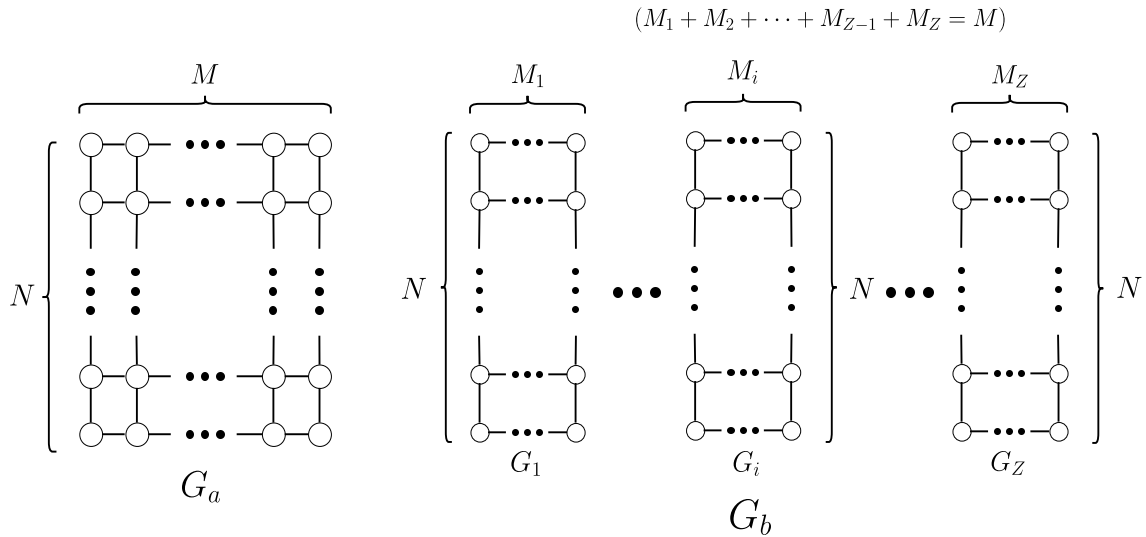
$$\begin{aligned}
 & P_\theta^{G_a}(x_i | \mathcal{F}_i) = \sum_{x \setminus x_i} P_\theta^{G_a}(x_1, \dots, x_z | \mathcal{F}_i) \\
 &= \frac{\exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\overline{\theta_{i+1}}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\overline{\theta_{i-1}}) G_{r(e)}(x_i) \right]}{\sum_{x_i} \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\overline{\theta_{i+1}}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\overline{\theta_{i-1}}) G_{r(e)}(x_i) \right]} \\
 & \times \frac{\sum_{x \setminus x_i} \exp \left[ \sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x) \right]}{\sum_{x \setminus x_i} \exp \left[ \sum_{\beta \in G(a) \setminus (C_i \cup C_{i-1} \cup G(i))} \theta^\beta F_\beta(x) \right]} \\
 &= \frac{\exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\overline{\theta_{i+1}}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\overline{\theta_{i-1}}) G_{r(e)}(x_i) \right]}{\sum_{x_i} \exp \left[ \sum_{\alpha \in G(i)} \theta^\alpha G_\alpha(x_i) + \sum_{e \in C_i} \theta^e G_{\ell(e)}(x_i) \bar{\eta}_{r(e)}(\overline{\theta_{i+1}}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}(\overline{\theta_{i-1}}) G_{r(e)}(x_i) \right]} \tag{A.29}
 \end{aligned}$$

(A.23), (A.29) 式より, 周辺分布  $q_{\theta_i}^{G_i} \in \mathcal{M}(G_i)$  と条件付き周辺分布  $P_\theta^{G_a}(x_i | \mathcal{F}_i)$  は等しい . また, クラスタ  $G_i$  の列数が 1 の場合も同様の手順で示せる . ゆえに, 本論文で導出した  $e$ -射影の条件式 (5.40) ~ (5.43) は, Xing らが導出した一般化平均場方程式 (A.18) と等価である .

# 付録B 提案アルゴリズムの情報幾何学的解釈と収束性について

本付録では、提案アルゴリズムの情報幾何学的解釈と収束性について解説する。

## B.1 提案アルゴリズム



まず、カット  $C_i (1 \leq i \leq Z-1)$  と和集合  $C$  は以下のように定義されていた。

$$C_i := \{\{u, v\} \in E(G_a) \mid u \in V(G_i), v \in V(G_{i+1})\} \quad (\text{B.1})$$

$$C := \bigcup_{i=1}^Z C_i \quad (\text{B.2})$$

次に、集合  $V_B$  と  $V_i (1 \leq i \leq Z)$  を以下のように定義する。

$$V_i := \{\ell(e), r(f) \mid e \in C_i, f \in C_{i-1}\} \quad (\text{B.3})$$

$$V_B := \{\ell(e), r(e) \mid e \in C\} \quad \left( = \bigcup_{i=1}^Z V_i \right) \quad (\text{B.4})$$

$V_i$  はクラスター  $G_i$  に含まれる頂点の中で、結合が切断されている頂点の集合を表している。ただし、

$$C_0 = C_Z = \emptyset \quad (\text{B.5})$$

としている。

本研究では、 $e$ -射影を導出するアルゴリズムとして以下を提案した。

[ビリーフプロパゲーションを用いた  $e$ -射影点導出アルゴリズム]

[Initialize]:  $t = 0$ ,  $\bar{h}_v^{(0)} = h_v$  ( $v \in V(G_b)$ ),  $\bar{w}_e^{(0)} = w_e$  ( $e \in E(G_b)$ )

step.1  $Z$  個のクラスターの中からクラスター  $G_i$  を任意に選ぶ .

step.2 クラスター  $G_i$  と隣接するクラスターを一次元鎖へと変形し, ビリーフプロパゲーションを用いて, 一次元鎖の各頂点に関する周辺分布を導出する .

step.3 step.2 で導出した周辺分布を用いて, 結合が切断された素子に関する期待値パラメータを導出する .

step.4 step.3 で導出した期待値パラメータを用いて, クラスター  $G_i$  の結合が切断された素子のしきい値  $(\bar{h}_v^{(t)})_{v \in V_i}$  を更新する . 更新後の分布を  $q_{\bar{\theta}^{(t+1)}}^{G_b}$  とする .

step.5 分布  $q_{\bar{\theta}^{(t+1)}}^{G_b}$  に関するノルム (7.1) 式を計算し, 以下の条件式

$$\left\| q_{\bar{\theta}^{(t+1)}}^{G_b} \right\| < \epsilon \quad (\epsilon \text{ は十分小さな正数}) \quad (\text{B.6})$$

を満たすならばアルゴリズムを終了 . でなければ,  $t = t + 1$  として step.1 へ

## B.2 提案アルゴリズムの情報幾何学的解釈

提案アルゴリズムの [Initialize] 部分より, 提案アルゴリズムは以下の分布族  $\mathcal{M}_{V_B} \subset \mathcal{M}(G_b)$  の中から最適な分布を導出するアルゴリズムであると考えられる .

$$\mathcal{M}_{V_B} := \left\{ q_{\bar{\theta}}^{G_b} \mid \forall \beta \in U \setminus V_B, \bar{\theta}^\beta = \theta^\beta \right\} \quad (U = V(G_b) \cup E(G_b)) \quad (\text{B.7})$$

提案アルゴリズムの時刻  $t$  における  $\mathcal{M}_{V_B}$  の要素  $q_{\bar{\theta}^{(t)}}^{G_b}$  は以下のように記述される .

$$\begin{aligned} \bar{\theta}^{(t)} &= (\bar{\theta}_1^{(t)}, \dots, \bar{\theta}_Z^{(t)}), \quad q_{\bar{\theta}^{(t)}}^{G_b}(x) = \prod_{j=1}^Z q_{\bar{\theta}_j^{(t)}}^{G_j}(x_j) \\ &\left( x = (x_v)_{v \in V(G_a)}, x_j = (x_v)_{v \in V(G_j)} \right) \end{aligned} \quad (\text{B.8})$$

ここで,  $\mathcal{M}_{i,t+1} \subset \mathcal{M}_{V_B}$  ( $1 \leq i \leq Z$ ) を以下で定義する .

$$\mathcal{M}_{i,t+1} := \left\{ q_{\bar{\theta}^{(t+1)}}^{G_b}(x) = \prod_{j=1}^Z q_{\bar{\theta}_j^{(t+1)}}^{G_j}(x_j) \mid \forall j \neq i, q_{\bar{\theta}_j^{(t+1)}}^{G_j}(x_j) = q_{\bar{\theta}_j^{(t)}}^{G_j}(x_j) \right\} \quad (\text{B.9})$$

$$= \left\{ q_{\bar{\theta}^{(t+1)}}^{G_b}(x) = \prod_{j=1}^Z q_{\bar{\theta}_j^{(t+1)}}^{G_j}(x_j) \mid \forall j \neq i, \bar{\theta}_j^{(t+1)} = \bar{\theta}_j^{(t)} \right\} \quad (\text{B.10})$$

$$= \left\{ q_{\bar{\theta}^{(t+1)}}^{G_b}(x) = \prod_{j=1}^Z q_{\bar{\theta}_j^{(t+1)}}^{G_j}(x_j) \mid \forall \gamma \in V_B \setminus V_i, \bar{\theta}_{(t+1)}^\gamma = \bar{\theta}_t^\gamma \right\} \quad (\text{B.11})$$

このとき, 以下の補題が成立する .

補題 1 (クラスター  $G_i$  の境界部分のしきい値更新と  $e$ -射影の関係)

クラスター  $G_i$  のしきい値  $\bar{h}_v^{(t)}$  ( $v \in V_i$ ) を  $e$ -射影の条件式 (5.40) ~ (5.42) に従って更新させた時の分布  $q_{\bar{\theta}^{(t+1)}}^{G_b} \in \mathcal{M}_{i,t+1}$  は,  $P_\theta^{G_a} \in \mathcal{M}(G_a)$  から  $\mathcal{M}_{i,t+1}$  への  $e$ -射影である . すなわち,

$$\forall \mu \in V_i, \quad \frac{\partial D \left( q_{\bar{\theta}^{(t+1)}}^{G_b} \parallel P_\theta^{G_a} \right)}{\partial \bar{\theta}^\mu} = 0 \quad (\text{B.12})$$

が成立する .

## 証明 1

任意の時刻  $t \geq 0$  において，以下が成立する．

$$\begin{aligned}
 & \log q_{\bar{\theta}(t)}^{G_b}(x) - \log P_{\theta}^{G_a}(x) \\
 &= \sum_{\beta \in U} \left( \bar{\theta}_{(t)}^{\beta} - \theta^{\beta} \right) F_{\beta}(x) - \sum_{e \in C} \theta^e F_e(x) - \sum_{j=1}^Z \Phi_j \left( \bar{\theta}_{j(t)} \right) + \Psi(\theta) \\
 &= \sum_{\beta \in U \setminus V_B} \left( \bar{\theta}_{(t)}^{\beta} - \theta^{\beta} \right) F_{\beta}(x) + \sum_{\gamma \in V_B} \left( \bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma} \right) F_{\gamma}(x) - \sum_{e \in C} \theta^e F_e(x) - \sum_{j=1}^Z \Phi_j \left( \bar{\theta}_{j(t)} \right) + \Psi(\theta) \\
 &= \sum_{\gamma \in V_B} \left( \bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma} \right) F_{\gamma}(x) - \sum_{e \in C} \theta^e F_e(x) - \sum_{j=1}^Z \Phi_j \left( \bar{\theta}_{j(t)} \right) + \Psi(\theta) \\
 &= \sum_{\gamma \in V_B \setminus V_i} \left( \bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma} \right) F_{\gamma}(x) + \sum_{\xi \in V_i} \left( \bar{\theta}_{(t)}^{\xi} - \theta^{\xi} \right) F_{\xi}(x) - \sum_{e \in C} \theta^e F_e(x) \\
 & \quad - \sum_{j=1}^Z \Phi_j \left( \bar{\theta}_{j(t)} \right) + \Psi(\theta) \tag{B.13}
 \end{aligned}$$

1 番目の等式は，(5.25) 式そのものである．また， $q_{\bar{\theta}(t)}^{G_b}$  は分布族  $\mathcal{M}_{V_B}$  の要素であるため，2 番目の等式中の第 1 項目は 0 となる．次に，(B.12) 式の左辺を計算すると，

$$\begin{aligned}
 & \frac{\partial D(q_{\bar{\theta}(t+1)}^{G_b} \| P_{\theta}^{G_a})}{\partial \bar{\theta}^{\mu}} \\
 &= E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left( \log q_{\bar{\theta}(t+1)}^{G_b}(x) - \log P_{\theta}^{G_a}(x) \right) \left( \frac{\partial \log q_{\bar{\theta}(t+1)}^{G_b}(x)}{\partial \bar{\theta}^{\mu}} \right) \right] \\
 &= \sum_{\gamma \in V_B \setminus V_i} \left( \bar{\theta}_{(t+1)}^{\gamma} - \theta^{\gamma} \right) E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_{\gamma}(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \\
 & \quad + E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left\{ \sum_{\xi \in V_i} \left( \bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi} \right) F_{\xi}(x) \right\} \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \\
 & \quad - \sum_{e \in C} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_e(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \\
 &= \sum_{\gamma \in V_B \setminus V_i} \left( \bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma} \right) \left\{ \bar{\eta}_{\gamma}^{(t)}(\bar{\theta}_{(t)}) \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) - \bar{\eta}_{\gamma}^{(t)}(\bar{\theta}_{(t)}) \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \\
 & \quad + E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left\{ \sum_{\xi \in V_i} \left( \bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi} \right) F_{\xi}(x) \right\} \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \\
 & \quad - \sum_{e \in C} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_e(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \\
 &= E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left\{ \sum_{\xi \in V_i} \left( \bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi} \right) F_{\xi}(x) \right\} \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \\
 & \quad - \sum_{e \in C} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_e(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}_{(t+1)}) \right\} \right] \tag{B.14}
 \end{aligned}$$

となる．1 番目の等式は (5.23) 式そのものである．2 番目の等式は (B.13), (5.27), (5.30) 式を用いている．以降では，(i)  $i = 1$ , (ii)  $i = Z$ , (iii)  $1 < i < Z$  の場合に分けて (B.14) 式の右辺を式変形していく．

(i)  $i = 1$  の場合 ( クラスタ  $G_1$  に関する更新 )

(B.14) 式の右辺

$$\begin{aligned}
 &= E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left( \sum_{e \in C_1} \theta^e \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) F_{\ell(e)}(x) \right) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \right] \\
 &\quad - \sum_{e \in C_1} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_e(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \right] \\
 &\quad - \sum_{e \in C \setminus C_1} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_e(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \right] \\
 &= \sum_{e \in C_1} \theta^e \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(e)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(e)}(x)] \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \\
 &\quad - \sum_{e \in C_1} \theta^e \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(e)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(e)}(x)] \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \\
 &\quad - \sum_{e \in C \setminus C_1} \theta^e \left\{ \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) - \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \\
 &= 0 \tag{B.15}
 \end{aligned}$$

1 番目の等式は  $e$ -射影の条件式 (5.40) を用いている .

(ii)  $i = Z$  の場合 ( クラスタ  $G_Z$  に関する更新 )

(B.14) 式の右辺

$$\begin{aligned}
 &= E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left\{ \sum_{e \in C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) F_{r(e)}(x) \right\} \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \right] \\
 &\quad - \sum_{e \in C_{Z-1}} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_{\ell(e)}(x) F_{r(e)}(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \right] \\
 &\quad - \sum_{e \in C \setminus C_{Z-1}} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_{\ell(e)}(x) F_{r(e)}(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \right] \\
 &= \sum_{e \in C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x)] \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \\
 &\quad - \sum_{e \in C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x)] \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \\
 &\quad - \sum_{e \in C \setminus C_{Z-1}} \theta^e \left\{ \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) - \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}(t)) \bar{\eta}_{\mu}^{(t+1)} (\bar{\theta}(t+1)) \right\} \\
 &= 0 \tag{B.16}
 \end{aligned}$$

1 番目の等式は ,  $e$ -射影の条件式 (5.41) を用いている .



(iii)  $1 < i < Z$  の場合

(B.14) 式の右辺

$$\begin{aligned}
 &= E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left\{ \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)}(\bar{\theta}(t)) F_{r(e)}(x) \right\} \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \right] \\
 &+ E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ \left\{ \sum_{f \in C_i} \theta^f \bar{\eta}_{r(f)}^{(t)}(\bar{\theta}(t)) F_{\ell(f)}(x) \right\} \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \right] \\
 &- \sum_{e \in C_{i-1}} \theta^e E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_{\ell(e)}(x) F_{r(e)}(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \right] \\
 &- \sum_{f \in C_i} \theta^f E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_{\ell(f)}(x) F_{r(f)}(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \right] \\
 &- \sum_{g \in C \setminus (C_{i-1} \cup C_i)} \theta^g E_{q_{\bar{\theta}(t+1)}^{G_b}} \left[ F_{\ell(g)}(x) F_{r(g)}(x) \left\{ F_{\mu}(x) - \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \right] \\
 &= \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)}(\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x)] \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \\
 &+ \sum_{f \in C_i} \theta^f \bar{\eta}_{r(f)}^{(t)}(\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(f)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(f)}(x)] \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \\
 &- \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)}(\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{r(e)}(x)] \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \\
 &- \sum_{f \in C_i} \theta^f \bar{\eta}_{r(f)}^{(t)}(\bar{\theta}(t)) \left\{ E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(f)}(x) F_{\mu}(x)] - E_{q_{\bar{\theta}(t+1)}^{G_b}} [F_{\ell(f)}(x)] \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \\
 &- \sum_{g \in C \setminus (C_{i-1} \cup C_i)} \theta^g \left\{ \bar{\eta}_{\ell(g)}^{(t)}(\bar{\theta}(t)) \bar{\eta}_{r(g)}^{(t)}(\bar{\theta}(t)) \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) - \bar{\eta}_{\ell(g)}^{(t)}(\bar{\theta}(t)) \bar{\eta}_{r(g)}^{(t)}(\bar{\theta}(t)) \bar{\eta}_{\mu}^{(t+1)}(\bar{\theta}(t+1)) \right\} \\
 &= 0 \tag{B.17}
 \end{aligned}$$

1 番目の等式は、 $e$ -射影の条件式 (5.40), (5.41) を用いている。また、クラスター  $G_i$  ( $1 < i < Z$ ) の列数が 1 の場合でも、(B.14) 式の右辺は 0 となる。

(i) ~ (iii) より、 $\forall V_i \subset V_B, \forall \mu \in V_i$  に対して (B.12) が成立する (証明終)

例えば、 $t$  時点の分布  $q_{\bar{\theta}(t)}^{G_b} \in \mathcal{M}_{V_B}$  からクラスター  $G_1$ , クラスタ  $G_2$ , クラスタ  $G_3$  の順に結合が切断されたしきい値を更新していく場合、提案アルゴリズムの幾何学的イメージは以下の図 B.4 のように表現される。

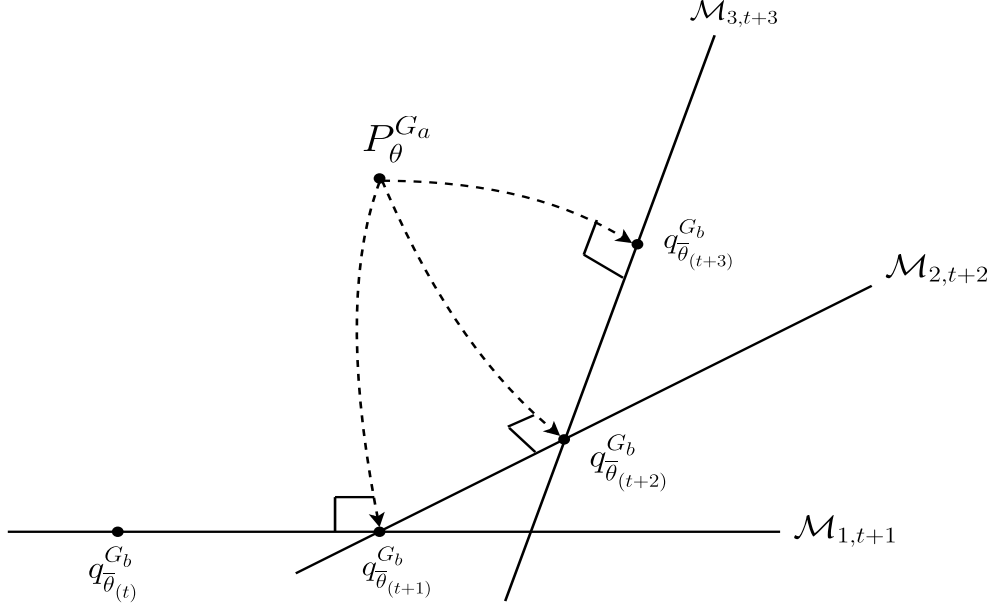


図 B.4: 提案アルゴリズムの幾何学的イメージ

### B.3 提案アルゴリズムの収束性

提案アルゴリズムについて，以下の補題が成立する．

**補題 2 (Kullback-Leibler ダイバージェンスに関するピタゴラス型の等式)**

$\forall t, \forall i (1 \leq i \leq Z)$  に対して，以下の等式が成立する．

$$\begin{aligned} q_{\bar{\theta}(t)}^{G_b}, q_{\bar{\theta}(t+1)}^{G_b} &\in \mathcal{M}_{i,t+1} \\ D\left(q_{\bar{\theta}(t)}^{G_b} \parallel P_{\theta}^{G_a}\right) &= D\left(q_{\bar{\theta}(t+1)}^{G_b} \parallel P_{\theta}^{G_a}\right) + D\left(q_{\bar{\theta}(t)}^{G_b} \parallel q_{\bar{\theta}(t+1)}^{G_b}\right) \end{aligned} \quad (\text{B.18})$$

ただし， $D(\cdot \parallel \cdot)$  は Kullback-Leibler ダイバージェンスとしている．

**証明 2**

まず (B.18) 式の左辺を計算すると，

$$\begin{aligned} &(\text{B.18) 式の左辺} \\ &= \sum_x q_{\bar{\theta}(t)}^{G_b}(x) \left( \log q_{\bar{\theta}(t)}^{G_b}(x) - \log P_{\theta}^{G_a}(x) \right) \\ &= \sum_x q_{\bar{\theta}(t)}^{G_b}(x) \left\{ \sum_{\gamma \in V_B \setminus V_i} (\bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma}) F_{\gamma}(x) + \sum_{\xi \in V_i} (\bar{\theta}_{(t)}^{\xi} - \theta^{\xi}) F_{\xi}(x) \right\} \\ &\quad - \sum_x q_{\bar{\theta}(t)}^{G_b}(x) \left\{ \sum_{e \in C} \theta^e F_e(x) + \sum_{j=1}^Z \Phi_j(\bar{\theta}_{j(t)}) - \Psi(\theta) \right\} \\ &= \sum_{\gamma \in V_B \setminus V_i} (\bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma}) \bar{\eta}_{\gamma}^{(t)}(\bar{\theta}_{(t)}) + \sum_{\xi \in V_i} (\bar{\theta}_{(t)}^{\xi} - \theta^{\xi}) \bar{\eta}_{\xi}^{(t)}(\bar{\theta}_{(t)}) \\ &\quad - \sum_{e \in C} \theta^e \bar{\eta}_{\ell(e)}^{(t)}(\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)}(\bar{\theta}_{(t)}) - \sum_{j=1}^Z \Phi_j(\bar{\theta}_{j(t)}) + \Psi(\theta) \end{aligned} \quad (\text{B.19})$$

となる . また , 2 番目の等式は (B.13) 式を用いている . 次に ,  $D \left( q_{\bar{\theta}(t+1)}^{G_b} \parallel P_{\theta}^{G_a} \right)$  と  $D \left( q_{\bar{\theta}(t)}^{G_b} \parallel q_{\bar{\theta}(t+1)}^{G_b} \right)$  をそれぞれ計算すると ,

$$\begin{aligned}
 & D \left( q_{\bar{\theta}(t+1)}^{G_b} \parallel P_{\theta}^{G_a} \right) \\
 &= \sum_x q_{\bar{\theta}(t+1)}^{G_b}(x) \left( \log q_{\bar{\theta}(t+1)}^{G_b}(x) - \log P_{\theta}^{G_a}(x) \right) \\
 &= \sum_x q_{\bar{\theta}(t+1)}^{G_b}(x) \left\{ \sum_{\gamma \in V_B \setminus V_i} (\bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma}) F_{\gamma}(x) + \sum_{\xi \in V_i} (\bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi}) F_{\xi}(x) \right\} \\
 &\quad - \sum_x q_{\bar{\theta}(t+1)}^{G_b}(x) \left\{ \sum_{e \in C} \theta^e F_e(x) + \sum_{j \neq i} \Phi_j(\bar{\theta}_{j(t)}) + \Phi_i(\bar{\theta}_{i(t+1)}) - \Psi(\theta) \right\} \\
 &= \sum_{\gamma \in V_B \setminus V_i} (\bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma}) \bar{\eta}_{\gamma}^{(t)}(\bar{\theta}_{(t)}) + \sum_{\xi \in V_i} (\bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi}) \bar{\eta}_{\xi}^{(t+1)}(\bar{\theta}_{(t+1)}) - \sum_{e \in C} \theta^e \bar{\eta}_e^{(t+1)}(\bar{\theta}_{(t+1)}) \\
 &\quad - \sum_{j \neq i} \Phi_j(\bar{\theta}_{j(t)}) - \Phi_i(\bar{\theta}_{i(t+1)}) + \Psi(\theta) \tag{B.20}
 \end{aligned}$$

$$\begin{aligned}
 & D \left( q_{\bar{\theta}(t)}^{G_b} \parallel q_{\bar{\theta}(t+1)}^{G_b} \right) \\
 &= \sum_x q_{\bar{\theta}(t)}^{G_b}(x) \left( \log q_{\bar{\theta}(t)}^{G_b}(x) - \log q_{\bar{\theta}(t+1)}^{G_b}(x) \right) \\
 &= \sum_{\xi \in V_i} (\bar{\theta}_{(t)}^{\xi} - \bar{\theta}_{(t+1)}^{\xi}) \bar{\eta}_{\xi}^{(t)}(\bar{\theta}_{(t)}) - \Phi_i(\bar{\theta}_{i(t)}) + \Phi_i(\bar{\theta}_{i(t+1)}) \tag{B.21}
 \end{aligned}$$

となる . (B.20), (B.21) 式より (B.18) 式の右辺は

(B.18) 式の右辺

$$\begin{aligned}
 &= \sum_{\gamma \in V_B \setminus V_i} (\bar{\theta}_{(t)}^{\gamma} - \theta^{\gamma}) \bar{\eta}_{\gamma}^{(t)}(\bar{\theta}_{(t)}) + \sum_{\xi \in V_i} (\bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi}) \bar{\eta}_{\xi}^{(t+1)}(\bar{\theta}_{(t+1)}) \\
 &\quad + \sum_{\xi \in V_i} (\bar{\theta}_{(t)}^{\xi} - \bar{\theta}_{(t+1)}^{\xi}) \bar{\eta}_{\xi}^{(t)}(\bar{\theta}_{(t)}) - \sum_{e \in C} \theta^e \bar{\eta}_e^{(t+1)}(\bar{\theta}_{(t+1)}) - \sum_{j=1}^Z \Phi_j(\bar{\theta}_{j(t)}) + \Psi(\theta) \tag{B.22}
 \end{aligned}$$

となる . (B.19), (B.22) 式より , (B.18) 式の左辺 - (B.18) 式の右辺 は

(B.18) 式の左辺 - (B.18) 式の右辺

$$\begin{aligned}
 &= \sum_{\xi \in V_i} (\bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi}) \bar{\eta}_{\xi}^{(t)}(\bar{\theta}_{(t)}) - \sum_{e \in C} \theta^e \bar{\eta}_{\ell(e)}^{(t)}(\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)}(\bar{\theta}_{(t)}) \\
 &\quad - \sum_{\xi \in V_i} (\bar{\theta}_{(t+1)}^{\xi} - \theta^{\xi}) \bar{\eta}_{\xi}^{(t+1)}(\bar{\theta}_{(t+1)}) + \sum_{e \in C} \theta^e \bar{\eta}_e^{(t+1)}(\bar{\theta}_{(t+1)}) \tag{B.23}
 \end{aligned}$$

となる . ここで , 補題 1 の証明のときと同様に , (i)  $i = 1$ , (ii)  $i = Z$ , (iii)  $1 < i < Z$  の場合分けを考えたときの (B.23) 式の右辺の値を計算していく .

(i)  $i = 1$  の場合

$$\begin{aligned}
 & \text{(B.23) 式の右辺} \\
 &= \sum_{e \in C_1} \theta^e \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) - \sum_{e \in C} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \\
 &\quad - \sum_{e \in C_1} \theta^e \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{\ell(e)}^{(t+1)} (\bar{\theta}_{(t+1)}) + \sum_{e \in C_1} \theta^e \bar{\eta}_{\ell(e)}^{(t+1)} (\bar{\theta}_{(t+1)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \\
 &\quad + \sum_{e \in C \setminus C_1} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \\
 &= 0
 \end{aligned} \tag{B.24}$$

(ii)  $i = Z$  の場合

$$\begin{aligned}
 & \text{(B.23) 式の右辺} \\
 &= \sum_{e \in C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) - \sum_{e \in C} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \\
 &\quad - \sum_{e \in C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t+1)} (\bar{\theta}_{(t+1)}) + \sum_{e \in C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t+1)} (\bar{\theta}_{(t+1)}) \\
 &\quad + \sum_{e \in C \setminus C_{Z-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \\
 &= 0
 \end{aligned} \tag{B.25}$$

(iii)  $1 < i < Z$  の場合

$$\begin{aligned}
 & \text{(B.23) 式の右辺} \\
 &= \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) + \sum_{f \in C_i} \theta^f \bar{\eta}_{r(f)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{\ell(f)}^{(t)} (\bar{\theta}_{(t)}) \\
 &\quad - \sum_{e \in C} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) - \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t+1)} (\bar{\theta}_{(t+1)}) \\
 &\quad - \sum_{f \in C_i} \theta^f \bar{\eta}_{r(f)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{\ell(f)}^{(t+1)} (\bar{\theta}_{(t+1)}) + \sum_{e \in C_{i-1}} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t+1)} (\bar{\theta}_{(t+1)}) \\
 &\quad + \sum_{f \in C_i} \theta^f \bar{\eta}_{\ell(f)}^{(t+1)} (\bar{\theta}_{(t+1)}) \bar{\eta}_{r(f)}^{(t)} (\bar{\theta}_{(t)}) + \sum_{e \in C \setminus (C_{i-1} \cup C_i)} \theta^e \bar{\eta}_{\ell(e)}^{(t)} (\bar{\theta}_{(t)}) \bar{\eta}_{r(e)}^{(t)} (\bar{\theta}_{(t)}) \\
 &= 0
 \end{aligned} \tag{B.26}$$

(i) ~ (iii) より, (B.18) 式は成立する (証明終)

また, Kullback-Leibler ダイバージェンスの非負性より,

$$D \left( q_{\bar{\theta}_{(t)}}^{G_b} \parallel q_{\bar{\theta}_{(t+1)}}^{G_b} \right) \geq 0 \tag{B.27}$$

となるため, 以下の不等式が成立する.

$$\forall t, \quad D \left( q_{\bar{\theta}_{(t)}}^{G_b} \parallel P_{\theta}^{G_a} \right) \geq D \left( q_{\bar{\theta}_{(t+1)}}^{G_b} \parallel P_{\theta}^{G_a} \right) \tag{B.28}$$

従って, 提案アルゴリズムは Kullback-Leibler ダイバージェンス  $D \left( \cdot \parallel P_{\theta}^{G_a} \right) \Big|_{\mathcal{M}_{V_B}}$  を単調減少させるアルゴリズムであるため収束する.

## 参考文献

- [1] D.H. Ackley, G.E. Hinton. and T.J Sejnowski, “A leaning algorithm for Boltzmann machines,” *Cognitive Science*, vol.9, no.1 pp.147-169, 1985
- [2] Eric P. Xing, Michael I. Jordan. and Stuart Russell, “A generalized mean field algorithm for variational inference in exponential families,” In *UAI*, 2003
- [3] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howrad, and L.D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural Computation*, Vol. 1, No. 4, pp.541-551, 1989
- [4] G.E. Hinton, S. Osindero, and Y.W. Teh, “A fast learning algorithm for deep belief net,” *Neural Computation*, Vol. 18, No. 7, pp. 1527-1554, 2006
- [5] R. Salakhutdinov and G.E. Hinton, “Deep Boltzmann Machines,” In *Proceedings of the 12th International Conference fo Artificial Intelligence and Statistics (AISTATS 2009)*, pp. 448-455, 2009
- [6] P. Weiss, “L’hypothèse du champ moléculaire et la propriété ferromagnétique,” *J. de Phys. Rad.* 6, pp.661-690, 1907
- [7] J. Pearl, “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,” Morgan Kaufmann, 1988
- [8] 長岡 浩司, 小嶋 徹也, 統計的モデルとしてのボルツマンマシン, 計算機統計学, 第 8 巻・第 1 号, pp.61-81, 1995
- [9] 上坂 吉則, ニューロコンピューティングの数学的基礎, 近代科学社, 1993
- [10] 甘利 俊一, 長岡浩司, 情報幾何の方法, 岩波講座 応用数学 6, 岩波書店, 1993
- [11] 金 聖培, 平均場近似の情報幾何学的考察, 電気通信大学大学院情報システム学研究科修士論文, 2008
- [12] T.Tanaka, “Information Geometry of Mean Field Theory,” *IEICE Trans. Fundamentals*, vol. E79-A, no. 5, pp. 709-715, May. 1996.
- [13] T.Tanaka, “Information Geometry of Mean-Field Approximation,” *Neural Computation*, 12 pp. 1951-1968, 2000.
- [14] 新井 利幸, 田中 利幸, 藤森 偉恭, 情報幾何学に基づく平均場ボルツマンマシン学習則の理論的検討, 電子情報通信学会論文誌, A vol. J81-A, no. 3 pp. 336-345, 1998.
- [15] 神嵐 敏弘 編, 深層学習 –Deep Learning–, 近代科学社, 2015.
- [16] 藤原 彰夫, 情報幾何学の基礎, 牧野書店, 2015.

- [17] 汪 金芳, 田栗 正章, 手塚 集, 樺島 祥介, 上田 修功, 統計科学のフロンティア 11 計算統計 I 確率計算の新しい手法, 岩波書店, 2003.
- [18] S.Amari and H.Nagaoka, “Methods of Information Geometry,” AMS&Oxford Univ. Press, 2000.

## 謝辞

長岡・小川研究室での2年間の研究生生活において、お世話になった全ての人に対して、この場をお借りして感謝の意を示します。

まず、本研究を進める上で終始丁寧な指導をして頂いた小川朋宏准教授に感謝いたします。小川先生からは、私が少しでも興味を示した学問分野に関して、本当に幅広く何でも教えていただきました。そして、レクチャーを受ける度に、小川先生のカバーしている学問領域の広さと物事の本質を見抜く力に感心しておりました。また、学問や研究に関するだけでなく、プログラミングや $\text{T}_\text{E}\text{X}$ の作法、さらには就活事情や社会常識など挙げればきりがなく、いろいろな事を学ばせていただきました。その中でも特に、「人前での発表は、聞き手のことを最大限配慮した内容にすること」や「複雑な概念をシンプルなものへと抽象化することで効率化を図ること」はとても重要であり、なおかつ実現するのが大変難しいということをも身をもって体感しました。今後社会に出て行った後もこれらについては実践し続けていこうと思います。またそれ以外にも、この2年間で学んだことを活かせる機会があれば積極的に活かし、今度は自分が人に教える側に回れればと思います。

次に、定例ゼミや確率論ゼミ、情報幾何学の講義などでお世話になった長岡浩司教授に感謝いたします。長岡先生が受け持つ授業やゼミでは、常に頭をフル回転させながら受講していた印象が強いです。そのおかげもあってか、学部時代の自分と比べると、物事を深く考え抜く力が格段に向上したと思います。また、毎月の定例ゼミでは、研究に関する鋭い指摘をしていただき、自身の理解の助けとなりました。また、甘利先生との共同著書「情報幾何の方法」は、研究を行う上で大変参考にさせていただきました。

次に、定例ゼミや記号論理学・集合・写像ゼミ、情報幾何学の講義などでお世話になった鈴木淳准教授に感謝いたします。鈴木先生には、大学院に入学した際の頃で、研究室のことについてまだまだあまり把握していない時期に、大変丁寧に接していただいた記憶があります。また、M1の後期に行われた情報幾何学の講義は、その頃ちょうど藤原先生の教科書を使って勉強していた自分にとって、自身の理解度をさらに深めるのに大変役立ちました。

次に、毎月の定例ゼミで熱心に発表を聞いていただき、さらに定例ゼミ中（もしくは定例ゼミ終了後）に手厚くレクチャーして頂いた朝倉大樹研究員に感謝いたします。この2年間、朝倉さんの数学にかける情熱を間近で見る度に、自分にはとても真似できないという意味で、とても感心しておりました。また、常に謙虚な姿勢で研究に取り組んでいる姿は、私にとって良い見本となりました。

次に、2年間の大学院生活において、共に励まし合い、共に学び続けてきた森君、福君にも感謝しております。森君はLinuxや $\text{T}_\text{E}\text{X}$ の扱いに慣れていない私に対して、嬉々として色々なことを教えてくれました。福君は研究室内での仕事を率先して引き受けていた印象が強く、また、自分が気づかないような細かなことへの気配りにも随分と助けられました。

最後に、大学院への進学を受け入れ、私の大学院生活を経済面・精神面で支えてくれた両親に深く感謝します。