

Arimoto アルゴリズムにおける遅い収束についての解析

Analysis for the Slow Convergence in Arimoto Algorithm

中川 健治*
Kenji Nakagawa

武井 由智†
Yoshinori Takei

渡部 康平*
Kohei Watabe

Abstract— In this paper, we investigate the convergence speed of the Arimoto algorithm. By analyzing the Taylor expansion of the defining function of the Arimoto algorithm, we will clarify the conditions for the exponential or $1/N$ order convergence and calculate the convergence speed. We show that the convergence speed of the $1/N$ order is evaluated by the derivatives of the Kullback-Leibler divergence with respect to the input probabilities. The analysis for the convergence of the $1/N$ order is new in this paper. Based on the analysis, we will compare the convergence speed of the Arimoto algorithm with the theoretical values obtained in our theorems for several channel matrices.

Keywords— channel capacity, discrete memoryless channel, Arimoto algorithm, convergence speed, Hessian matrix

1 はじめに

Arimoto [1] は、離散的無記憶通信路の通信路容量 C を計算するための逐次アルゴリズムを提案した。ベイズ確率に基づいて、アルゴリズムは入力確率と逆向き通信路行列との間の交互最小化によって与えられる。任意の通信路行列 Φ に対して、Arimoto アルゴリズムの収束が証明され、収束速度が評価されている。最悪のケースで、収束速度は $1/N$ オーダであり、通信路容量 C を達成する入力分布 λ^* が入力分布全体の集合 $\Delta(\mathcal{X})$ の内点である場合、収束は指数関数的であることが示されている。

本論文ではまず指数関数的な収束について検討し、収束速度を評価する。さらに、従来の研究では扱っていなかった $1/N$ オーダの収束も考察する。特に、入力アルファベットサイズ $m = 3$ の場合において、 $1/N$ オーダの収束を詳細に解析する。収束速度は、Kullback-Leibler 情報量の入力確率に関する微分によって評価される。

収束速度を評価するための基本的な考え方として、Arimoto アルゴリズムを定義する関数 $F(\lambda)$ が $\Delta(\mathcal{X})$ から $\Delta(\mathcal{X})$ への微分可能な写像であり、通信路容量 C を達成する入力分布 λ^* が $F(\lambda)$ の不動点であることに注目する。そして、不動点 $\lambda = \lambda^*$ のまわりでの $F(\lambda)$ の Taylor 展開を解析することにより、収束速度を評価する。本論文に関する詳細は [7] に記載する。

2 関連研究

Arimoto アルゴリズムに関連する多くの研究がある。例えば、異なるタイプの通信路への拡張 [5], [8], [9], Ari-

moto アルゴリズムの加速 [4], [10], ダイバージェンス幾何による Arimoto アルゴリズムの特徴付け [3], [4], [6], などである。Arimoto アルゴリズムの収束速度の解析に焦点を当てると、[1], [4], [10] において、 λ^* が $\Delta(\mathcal{X})$ の内部にある場合、Jacobi 行列によって指数関数的収束が調べられている。

3 通信路行列と通信路容量

離散的無記憶通信路 $X \rightarrow Y$ を考える。入力アルファベットを $\mathcal{X} = \{x_1, \dots, x_m\}$ とし、出力アルファベットを $\mathcal{Y} = \{y_1, \dots, y_n\}$ とする。

入力シンボル x_i が送信されたとき、出力シンボル y_j が受信される条件付き確率を $P_j^i = P(Y = y_j | X = x_i)$, $i = 1, \dots, m, j = 1, \dots, n$ とし、行ベクトル P^i を $P^i = (P_1^i, \dots, P_n^i)$, $i = 1, \dots, m$ と定義する。通信路行列 Φ が次のように定義される。

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P_1^1 & \dots & P_n^1 \\ \vdots & & \vdots \\ P_1^m & \dots & P_n^m \end{pmatrix}. \quad (1)$$

入力分布全体の集合を $\Delta(\mathcal{X}) \equiv \{\lambda = (\lambda_1, \dots, \lambda_m) | \lambda_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \lambda_i = 1\}$ とし、 $\Delta(\mathcal{X})$ の内部を $\Delta(\mathcal{X})^\circ \equiv \{\lambda = (\lambda_1, \dots, \lambda_m) \in \Delta(\mathcal{X}) | \lambda_i > 0, i = 1, \dots, m\}$ とする。さらに出力分布全体の集合を $\Delta(\mathcal{Y}) \equiv \{Q = (Q_1, \dots, Q_n) | Q_j \geq 0, j = 1, \dots, n, \sum_{j=1}^n Q_j = 1\}$ とする。

入力分布 $\lambda \in \Delta(\mathcal{X})$ に対する出力分布を $Q = \lambda\Phi$ とし、これを成分で表して $Q_j = \sum_{i=1}^m \lambda_i P_j^i$, $j = 1, \dots, n$ とする。このとき、相互情報量が $I(\lambda, \Phi) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i P_j^i \log P_j^i / Q_j$ と定義され、通信路容量 C が

$$C = \max_{\lambda \in \Delta(\mathcal{X})} I(\lambda, \Phi) \quad (2)$$

と定義される。

2つの出力分布 $Q = (Q_1, \dots, Q_n)$, $Q' = (Q'_1, \dots, Q'_n) \in \Delta(\mathcal{Y})$ に対して、Kullback-Leibler 情報量 $D(Q||Q')$ が次のように定義される [2]。

$$D(Q||Q') = \sum_{j=1}^n Q_j \log \frac{Q_j}{Q'_j}. \quad (3)$$

Arimoto アルゴリズムの収束速度を調べるための重要な命題は、入力分布 $\lambda = \lambda^*$ に関する次の Kuhn-Tucker 条件である。

定理 (Kuhn-Tucker 条件) 最大化問題 (2) において、入力分布 $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \Delta(\mathcal{X})$ が最大値を達成する

* 長岡技術科学大学 電気電子情報工学専攻 〒940-2188 新潟県長岡市上富岡町 1603-1 Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan, E-mail: nakagawa@nagaokaut.ac.jp, k.watabe@vos.nagaokaut.ac.jp

† 秋田工業高等専門学校 電気・電子・情報系 〒011-8511 秋田市飯島文京町 1 番 1 号 National Institute of Technology, Akita College, Akita 011-8511, Japan, E-mail: ytakei@akita-nct.ac.jp

ための必要十分条件は、次の条件を満たす \tilde{C} が存在することである；

$$D(P^i \| \lambda^* \Phi) \begin{cases} = \tilde{C}, & \text{for } i \text{ with } \lambda_i^* > 0, \\ \leq \tilde{C}, & \text{for } i \text{ with } \lambda_i^* = 0. \end{cases} \quad (4)$$

(4)において \tilde{C} は通信路容量 C に等しい。

この Kuhn-Tucker 条件は必要十分条件であるので、通信路容量を達成する入力分布 λ^* に関するすべての情報はこの条件から導出できる。

4 Arimoto アルゴリズム [1]

入力分布の系列

$$\{\lambda^N = (\lambda_1^N, \dots, \lambda_m^N)\}_{N=0,1,\dots} \subset \Delta(\mathcal{X}) \quad (5)$$

が Arimoto アルゴリズムによって以下のように定義される。まず、 $\lambda^0 = (\lambda_1^0, \dots, \lambda_m^0) \in \Delta(\mathcal{X})^\circ$, すなわち、 $\lambda_i^0 > 0, i = 1, \dots, m$ となる初期分布を λ^0 とする。このとき Arimoto アルゴリズムは次の漸化式で与えられる；

$$\lambda_i^{N+1} = \frac{\lambda_i^N \exp D(P^i \| \lambda^N \Phi)}{\sum_{k=1}^m \lambda_k^N \exp D(P^k \| \lambda^N \Phi)}, \quad (6)$$

$$i = 1, \dots, m, N = 0, 1, \dots$$

4.1 $\Delta(\mathcal{X})$ から $\Delta(\mathcal{X})$ 自身への写像

$F_i(\lambda)$ を Arimoto アルゴリズム (6) の定義関数とする。すなわち、

$$F_i(\lambda) = \frac{\lambda_i \exp D(P^i \| \lambda \Phi)}{\sum_{k=1}^m \lambda_k \exp D(P^k \| \lambda \Phi)}, \quad i = 1, \dots, m \quad (7)$$

である。

$F(\lambda) = (F_1(\lambda), \dots, F_m(\lambda))$ と定義すると、 $F(\lambda)$ は $\Delta(\mathcal{X})$ から $\Delta(\mathcal{X})$ への微分可能写像と考えることができる。そして、(6) は次のように表せる。

$$\lambda^{N+1} = F(\lambda^N), \quad N = 0, 1, \dots \quad (8)$$

本論文では、収束速度を解析するために

$$\text{rank } \Phi = m \quad (9)$$

を仮定する。

補題 1 通信路容量を達成する入力分布 λ^* は写像 $F(\lambda)$ の不動点である。すなわち、 $\lambda^* = F(\lambda^*)$ である。

Arimoto アルゴリズムの系列 λ^N は不動点 λ^* に収束する。すなわち、

$$\lambda^N \rightarrow \lambda^*, \quad N \rightarrow \infty \quad (10)$$

となる。我々は、関数 $F(\lambda)$ の $\lambda = \lambda^*$ のまわりでの Taylor 展開を用いて (10) の収束速度を調べる。

4.2 インデックスのタイプ

ここで、Kuhn-Tucker 条件 (4) におけるインデックス $i (i = 1, \dots, m)$ を、さらに詳しく、以下の 3 つのタイプに分類する。

$$D(P^i \| \lambda^* \Phi) \begin{cases} = C, & \text{for } i \text{ with } \lambda_i^* > 0 \text{ (タイプ I)}, \\ = C, & \text{for } i \text{ with } \lambda_i^* = 0 \text{ (タイプ II)}, \\ < C, & \text{for } i \text{ with } \lambda_i^* = 0 \text{ (タイプ III)}. \end{cases} \quad (11)$$

インデックスの集合を次のように定義する；

$$\text{インデックス全体} : \mathcal{I} \equiv \{1, \dots, m\},$$

$$\text{タイプ I のインデックス} : \mathcal{I}_I \equiv \{1, \dots, m_1\},$$

$$\text{タイプ II のインデックス} : \mathcal{I}_{II} \equiv \{m_1 + 1, \dots, m_1 + m_2\},$$

$$\text{タイプ III のインデックス} : \mathcal{I}_{III} \equiv \{m_1 + m_2 + 1, \dots, m\}.$$

$|\mathcal{I}| = m, |\mathcal{I}_I| = m_1, |\mathcal{I}_{II}| = m_2, |\mathcal{I}_{III}| = m - m_1 - m_2 \equiv m_3$ であり、 $\mathcal{I} = \mathcal{I}_I \cup \mathcal{I}_{II} \cup \mathcal{I}_{III}$ および $m = m_1 + m_2 + m_3$ が成り立つ。

任意の通信路において、 \mathcal{I}_I は空集合ではなく、 $|\mathcal{I}_I| = m_1 \geq 2$ となる。しかし、 \mathcal{I}_{II} および \mathcal{I}_{III} はある通信路に対して空集合となる場合がある。

4.3 収束速度の例

Arimoto アルゴリズムの収束速度の通信路行列による違いを考察する。

多くの通信路行列 Φ では、収束は指数関数的だが、特別な Φ では収束は非常に遅い。入力アルファベットサイズ $m = 3$ 、出力アルファベットサイズ $n = 3$ の場合に、インデックスのタイプ I, II, III を考慮して次の例を考える。

例 1 [タイプ I のインデックスのみが存在する場合]

この場合、 $\lambda_i^* > 0, i = 1, 2, 3$ となる。したがって、 $Q^* \equiv \lambda^* \Phi$ は $\Delta P^1 P^2 P^3$ の内部にある。この例の具体的な通信路行列として、

$$\Phi^{(1)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{pmatrix} \quad (12)$$

を考える。この $\Phi^{(1)}$ に対して、 $\lambda^* = (0.431, 0.431, 0.138)$ となる。

例 2 [タイプ I とタイプ II のインデックスがある場合]

この場合、一般性を失うことなく、 $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0$ としてよい。したがって、 Q^* は辺 $P^1 P^2$ 上にあり、 $D(P^i \| Q^*) = C, i = 1, 2, 3$ である。この例の具体的な通信路行列として、

$$\Phi^{(2)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{pmatrix} \quad (13)$$

を考える。この $\Phi^{(2)}$ に対して、 $\lambda^* = (0.500, 0.500, 0.000)$ となる。

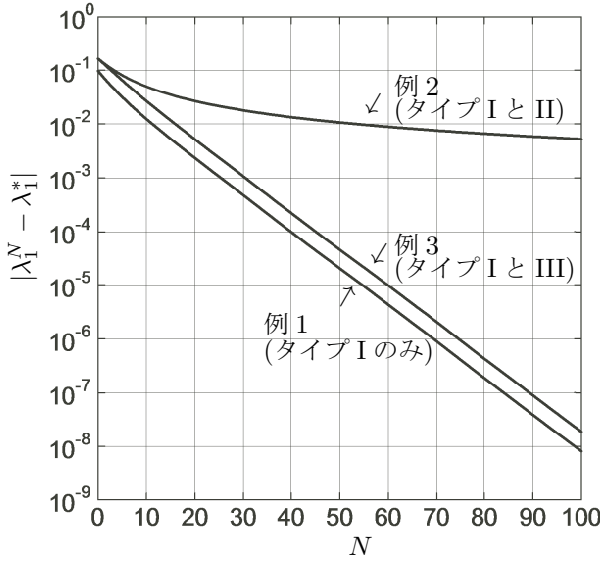


図 1: 例 1,2,3 における $|\lambda_1^N - \lambda_1^*|$ の収束速度の比較

例 3 [タイプ I とタイプ III のインデックスがある場合]
この場合、一般性を失うことなく $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0$ としてよい。したがって、 Q^* は辺 $P^1 P^2$ 上にあり、 $C = D(P^1 \| Q^*) = D(P^2 \| Q^*) > D(P^3 \| Q^*)$ である。この例の具体的な通信路行列として、

$$\Phi^{(3)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{pmatrix} \quad (14)$$

を考える。この $\Phi^{(3)}$ に対して、 $\lambda^* = (0.500, 0.500, 0.000)$ となる。

上記の $\Phi^{(1)}, \Phi^{(2)}, \Phi^{(3)}$ に対して、 $|\lambda_1^N - \lambda_1^*| \rightarrow 0$ の収束を図 1 に示す。この図から、例 1 および例 3 では収束が指数関数的であるのに対し、例 2 では収束が指数関数的よりも遅いことがわかる。

上記 3 つの例から、Arimoto アルゴリズムは、タイプ II のインデックスが存在する場合には非常に遅い収束となり、タイプ II のインデックスが存在しない場合には指数関数的に収束すると推測される。この現象を以下で解析する。

5 $F(\lambda)$ の $\lambda = \lambda^*$ のまわりでの Taylor 展開

Arimoto アルゴリズムの収束速度を調べるために、下記に示すとおり、 $F(\lambda)$ の不動点 $\lambda = \lambda^*$ のまわりでの Taylor 展開を利用する。

$$F(\lambda) = F(\lambda^*) + (\lambda - \lambda^*)J(\lambda^*) + \frac{1}{2!}(\lambda - \lambda^*)H(\lambda^*)^t(\lambda - \lambda^*) + o(\|\lambda - \lambda^*\|^2), \quad (15)$$

ただし、 ${}^t\lambda$ は λ の転置を表し、 $\|\lambda\|$ はユークリッドノルム $\|\lambda\| = (\lambda_1^2 + \dots + \lambda_m^2)^{1/2}$ を表す。

(15) において、 $J(\lambda^*)$ は $\lambda = \lambda^*$ における Jacobi 行

列である。すなわち、

$$J(\lambda^*) = \left(\frac{\partial F_i}{\partial \lambda_{i'}} \bigg|_{\lambda=\lambda^*} \right)_{i', i \in \mathcal{I}}. \quad (16)$$

また、 $H(\lambda^*) \equiv (H_1(\lambda^*), \dots, H_m(\lambda^*))$ であり、 $H_i(\lambda^*)$ は F_i の $\lambda = \lambda^*$ における Hesse 行列である。すなわち、

$$H_i(\lambda^*) = \left(\frac{\partial^2 F_i}{\partial \lambda_{i'} \partial \lambda_{i''}} \bigg|_{\lambda=\lambda^*} \right)_{i', i \in \mathcal{I}} \quad (17)$$

であり、 $(\lambda - \lambda^*)H(\lambda^*)^t(\lambda - \lambda^*)$ は m 次元ベクトル $((\lambda - \lambda^*)H_1(\lambda^*)^t(\lambda - \lambda^*), \dots, (\lambda - \lambda^*)H_m(\lambda^*)^t(\lambda - \lambda^*))$ を簡略に表現したものである。

さて、 $\lambda = \lambda^N$ を (15) に代入して、 $F(\lambda^*) = \lambda^*$ と $F(\lambda^N) = \lambda^{N+1}$ より

$$\begin{aligned} \lambda^{N+1} &= \lambda^* + (\lambda^N - \lambda^*)J(\lambda^*) \\ &+ \frac{1}{2!}(\lambda^N - \lambda^*)H(\lambda^*)^t(\lambda^N - \lambda^*) + o(\|\lambda^N - \lambda^*\|^2) \end{aligned} \quad (18)$$

となる。 $\mu^N \equiv \lambda^N - \lambda^*$ とおくと、(18) は

$$\mu^{N+1} = \mu^N J(\lambda^*) + \frac{1}{2!} \mu^N H(\lambda^*)^t \mu^N + o(\|\mu^N\|^2) \quad (19)$$

となる。(10) より、 $\mu^N \rightarrow 0, N \rightarrow \infty$ の収束を Taylor 展開 (19) によって調べる。

$\mu^N = \lambda^N - \lambda^*$ を成分で表して、 $\mu^N = (\mu_1^N, \dots, \mu_m^N)$ 、 $\mu_i^N \equiv \lambda_i^N - \lambda_i^*, i = 1, \dots, m$ とする。 $\sum_{i=1}^m \mu_i^N = 0, N = 0, 1, \dots$ となる。

5.1 Jacobi 行列 $J(\lambda^*)$ について

任意の m, n に対して、Jacobi 行列 $J(\lambda^*)$ を考える。 $J(\lambda^*)$ の成分 (16) を計算する。そのために次を定義する。

$$D_i \equiv D(P^i \| \lambda \Phi), i \in \mathcal{I}, \quad (20)$$

$$D_i^* \equiv D(P^i \| \lambda^* \Phi), i \in \mathcal{I}, \quad (21)$$

$$D_{i', i}^* \equiv \frac{\partial D_i}{\partial \lambda_{i'}} \bigg|_{\lambda=\lambda^*}, i', i \in \mathcal{I}. \quad (22)$$

定理 1

$$\frac{\partial F_i}{\partial \lambda_{i'}} \bigg|_{\lambda=\lambda^*} = e^{D_i^* - C} (\delta_{i', i} + \lambda_i^* D_{i', i}^*) + \lambda_i^* (1 - e^{D_{i', i}^* - C}), \quad i', i \in \mathcal{I},$$

$$= \begin{cases} \delta_{i', i} + \lambda_i^* (D_{i', i}^* + 1 - e^{D_{i', i}^* - C}), & i' \in \mathcal{I}, i \in \mathcal{I}_I, \\ \delta_{i', i}, & i' \in \mathcal{I}, i \in \mathcal{I}_{II}, \\ e^{D_i^* - C} \delta_{i', i}, & i' \in \mathcal{I}, i \in \mathcal{I}_{III}, \end{cases} \quad (23)$$

ここで、 $\delta_{i', i}$ は Kronecker デルタであり、インデックスの集合 $\mathcal{I}, \mathcal{I}_I, \mathcal{I}_{II}, \mathcal{I}_{III}$ は 4.2 節で定義されている。 $i \in \mathcal{I}_I \cup \mathcal{I}_{II}$ に対して $D_i^* = C$ であり、 $i \in \mathcal{I}_{II} \cup \mathcal{I}_{III}$ に対して $\lambda_i^* = 0$ であることを用いている。

5.2 Jacob 行列 $J(\lambda^*)$ の固有値

(23) より Jacobi 行列 $J(\lambda^*)$ は次の形になる。

$$J(\lambda^*) \equiv \begin{pmatrix} J^I & O & O \\ * & J^{II} & O \\ * & O & J^{III} \end{pmatrix}, \quad (24)$$

$$J^I \in \mathbb{R}^{m_1 \times m_1}, \quad (25)$$

$$J^{II} = I \text{ (単位行列)} \in \mathbb{R}^{m_2 \times m_2}, \quad (26)$$

$$J^{III} = \text{diag} \left(e^{D_{m_1+m_2+1}^*-C}, \dots, e^{D_m^*-C} \right) \in \mathbb{R}^{m_3 \times m_3}, \quad (27)$$

ここで, (11) におけるタイプ III より $D_{m_1+m_2+1}^* < C, \dots, D_m^* < C$,

O は適当なサイズのゼロ行列。

$\{\theta_1, \dots, \theta_m\} \equiv \{\theta_i | i \in \mathcal{I}\}$ を $J(\lambda^*)$ の固有値とする。(24) より, $J(\lambda^*)$ の固有値は J^I, J^{II}, J^{III} の固有値であるから, 次のようにおくことができる。

$\{\theta_i | i \in \mathcal{I}_I\}$: J^I の固有値,

$\{\theta_i | i \in \mathcal{I}_{II}\}$: J^{II} の固有値,

$\{\theta_i | i \in \mathcal{I}_{III}\}$: J^{III} の固有値。

J^I, J^{II}, J^{III} の固有値を以下において評価する。

定理 2 J^I の固有値は次を満たす。

$$0 \leq \theta_i < 1, i \in \mathcal{I}_I. \quad (28)$$

定理 3 (24), (26) より, J^{II} の固有値は次を満たす。

$$\theta_i = 1, i \in \mathcal{I}_{II}. \quad (29)$$

定理 4 (24), (27) より, J^{III} の固有値は $\theta_i = e^{D_i^*-C}, D_i^* < C, i \in \mathcal{I}_{III}$ であり, したがって次を満たす。

$$0 < \theta_i < 1, i \in \mathcal{I}_{III}. \quad (30)$$

6 収束速度について

定理 2, 3, 4 で $J(\lambda^*)$ の固有値に対する評価を得た。 $\theta_{\max} \equiv \max_{i \in \mathcal{I}} \theta_i$ を $J(\lambda^*)$ の最大固有値とすると, 定理 2, 3, 4 より, \mathcal{I}_{II} が空集合ならば $0 \leq \theta_{\max} < 1$ であり, \mathcal{I}_{II} が空集合でないならば $\theta_{\max} = 1$ となる。以下において, $\lambda^N \rightarrow \lambda^*$ あるいは $\mu^N \rightarrow \mathbf{0}$ の収束速度が, $0 \leq \theta_{\max} < 1$ ならば指数関数的であり, $\theta_{\max} = 1$ ならば $1/N$ オーダであることを示す。

6.1 $0 \leq \theta_{\max} < 1$ の場合の収束速度

定理 5 Jacobi 行列 $J(\lambda^*)$ の最大固有値 θ_{\max} が $0 \leq \theta_{\max} < 1$ を満たすとする。 $\theta_{\max} < \theta < 1$ となる任意の θ に対して $\delta > 0$ と $K > 0$ が存在して, $\|\lambda^0 - \lambda^*\| < \delta$ を満たす任意の初期ベクトル λ^0 に対して次が成り立つ;

$$\|\mu^N\| = \|\lambda^N - \lambda^*\| < K\theta^N, N = 0, 1, \dots, \quad (31)$$

すなわち, 収束は指数関数的である。ここで, θ^N は θ の N 乗を表す。

6.2 $\theta_{\max} = 1$ の場合の収束速度

$\theta_{\max} = 1$ となる場合には定理 5 は適用できない。すなわち, $\mu^N \rightarrow \mathbf{0}$ の収束速度は Jacobi 行列だけでは決まらず, Taylor 展開における 2 次項の Hesse 行列を調べる必要がある。

6.3 Hesse 行列について

従来の研究, 例えば [1],[4],[10] では, Jacobi 行列は考えられているが, Hesse 行列は扱われていない。関数 $F_i, i \in \mathcal{I}$ の $\lambda = \lambda^*$ における Hesse 行列の成分 (17) を計算する。 $D_{i,i',i''}^* \equiv \partial^2 D_i / \partial \lambda_{i'} \partial \lambda_{i''} |_{\lambda=\lambda^*}$ と定義する。

定理 6

$$\begin{aligned} \frac{\partial^2 F_i}{\partial \lambda_{i'} \partial \lambda_{i''}} \Big|_{\lambda=\lambda^*} &= e^{D_i^*-C} \left[(1 - e^{D_{i'}^*-C} + D_{i,i'}^*)(\delta_{ii''} + \lambda_i^*(1 - e^{D_{i''}^*-C})) \right. \\ &\quad + (1 - e^{D_{i''}^*-C} + D_{i,i''}^*)(\delta_{ii'} + \lambda_i^*(1 - e^{D_{i'}^*-C})) \\ &\quad \left. + \lambda_i^*(D_{i,i'}^* D_{i,i''}^* + D_{i,i',i''}^* + D_{i',i''}^* - e^{D_{i'}^*-C} D_{i,i''}^* - e^{D_{i''}^*-C} D_{i,i'}^* - \sum_{k=1}^{m_1} \lambda_k^* D_{k,i'}^* D_{k,i''}^*) \right], \quad (32) \end{aligned}$$

$i, i', i'' \in \mathcal{I}.$

特に, もし \mathcal{I}_{III} が空集合ならば, $i \in \mathcal{I}_{II}$ に対して

$$\frac{\partial^2 F_i}{\partial \lambda_{i'} \partial \lambda_{i''}} \Big|_{\lambda=\lambda^*} = \delta_{ii'} D_{i,i''}^* + \delta_{ii''} D_{i,i'}^*, i', i'' \in \mathcal{I}. \quad (33)$$

7 $m = 3, n$ が任意の場合の収束速度

定理 6 で, Hesse 行列は非常に複雑なので, 任意の通信路行列を調べることは困難である。したがって, 本節では, 特別な場合, すなわち $m = 3$ で n が任意の場合を考える。一般性を失うことなく $m = 3$ に対して, 次の網羅的な分類を得る。

(i) $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* > 0$,

(ii) $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0, D_3^* = C$,

(iii) $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0, D_3^* < C$ 。

(i) は例 1 の場合であり, $\mathcal{I}_I = \mathcal{I}, \mathcal{I}_{II} = \mathcal{I}_{III} = \emptyset$ となる。(24), (25) より

$$J(\lambda^*) = J^I \quad (34)$$

となる。定理 2 より, $0 \leq \theta_{\max} < 1$ となるので, 定理 5 より収束 $\mu^N \rightarrow \mathbf{0}$ は指数関数的である。

(ii) は後にして, (iii) を先に考える。(iii) は例 3 の場合であり, $\mathcal{I}_I = \{1, 2\}, \mathcal{I}_{II} = \emptyset, \mathcal{I}_{III} = \{3\}$ となる。(24), (27) より

$$J(\lambda^*) = \begin{pmatrix} J^I & O \\ * & J^{III} \end{pmatrix}, \quad (35)$$

$$J^I \in \mathbb{R}^{2 \times 2}, \quad (36)$$

$$J^{III} = e^{D_3^*-C}, 0 < J^{III} < 1 \quad (37)$$

となる。定理 2, 4 より, $0 < \theta_{\max} < 1$ となるので, 定理 5 より, 収束 $\mu^N \rightarrow \mathbf{0}$ は指数関数的である。

残るは (ii) である。(ii) は例 2 の場合であり, $\mathcal{I}_I = \{1, 2\}$, $\mathcal{I}_{II} = \{3\}$, $\mathcal{I}_{III} = \emptyset$ となる。(24), (26) より

$$J(\lambda^*) = \begin{pmatrix} J^I & O \\ * & J^{II} \end{pmatrix}, \quad (38)$$

$$J^I \in \mathbb{R}^{2 \times 2}, \quad (39)$$

$$J^{II} = 1 \quad (40)$$

となる。定理 2, 3 より $\theta_{\max} = 1$ となるので, 定理 5 を適用することはできない。収束速度の解析のために Hesse 行列を調べる。

7.1 $1/N$ オーダの収束

上述の (ii) の場合における $\mu^N \rightarrow \mathbf{0}$ の収束速度を調べ, それが $1/N$ オーダの収束であることを示す。

定理 7 $m = 3$ で n は任意とする。通信路容量を達成する $\lambda^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*)$ は $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0$ を満たし, $D_3^* = D(P^3 \| \lambda^* \Phi) = C$ であると仮定する。さらに, 下記の ρ について, $\rho > 0$ を仮定する。

このとき, $\mu^N = \lambda^N - \lambda^*$, $\mu^N = (\mu_1^N, \mu_2^N, \mu_3^N)$ に関して, 収束 $\mu^N \rightarrow \mathbf{0}$ は $1/N$ オーダであり, 次が成り立つ。

$$\lim_{N \rightarrow \infty} N\mu_1^N = -b_1/\rho, \quad (41)$$

$$\lim_{N \rightarrow \infty} N\mu_2^N = -b_2/\rho, \quad (42)$$

$$\lim_{N \rightarrow \infty} N\mu_3^N = 1/\rho, \quad (43)$$

ただし, $b_1 = \frac{\lambda_1^*(D_{1,2}^* - D_{1,3}^*)}{1 + D_{1,2}^*}$, $b_2 = \frac{\lambda_2^*(D_{1,2}^* - D_{2,3}^*)}{1 + D_{1,2}^*}$,

$$\rho = \lambda_1^* \frac{D_{1,3}^*(D_{1,2}^* - D_{1,3}^*)}{1 + D_{1,2}^*} + \lambda_2^* \frac{D_{2,3}^*(D_{1,2}^* - D_{2,3}^*)}{1 + D_{1,2}^*} - D_{3,3}^*,$$

$D_{i',i}^*$ は (20)-(22) で定義されている。

8 数値例

前節までの解析に基づいて, $m = n = 3$ を満たすいくつかの通信路行列に対して Arimoto アルゴリズムの収束速度を数値的に評価する。

指数関数的収束の場合には関数

$$L(N) \equiv -\frac{1}{N} \log \|\mu^N\| \quad (44)$$

を評価し, 定理 5 に基づいて, 十分大きな N に対して $L(N)$ と $-\log \theta_{\max}$ を比較する。一方, $1/N$ オーダの収束の場合には

$$N\mu^N = (N\mu_1^N, N\mu_2^N, N\mu_3^N) \quad (45)$$

の値を評価し, 定理 7 の極限值と比較する。

8.1 (i) の場合: $\lambda^* \in \Delta(\mathcal{X})^\circ$ のときの指数関数的収束

例 4 (12) の通信路行列 $\Phi^{(1)}$ を考える。すなわち,

$$\Phi^{(1)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}. \quad (46)$$

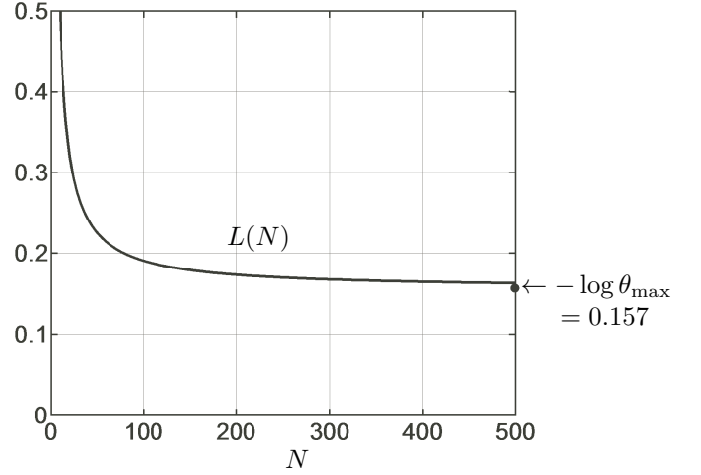


図 2: 例 4 における $L(N)$ の収束

このとき,

$$\lambda^* = (0.431, 0.431, 0.138), Q^* = (0.422, 0.422, 0.156),$$

$$J(\lambda^*) = \begin{pmatrix} 0.308 & -0.191 & -0.117 \\ -0.191 & 0.308 & -0.117 \\ -0.369 & -0.369 & 0.738 \end{pmatrix}$$

となり, $J(\lambda^*)$ の固有値は $(\theta_1, \theta_2, \theta_3) = (0.000, 0.500, 0.855)$ であり, $\theta_{\max} = \theta_3 = 0.855$ である。

$\lambda^0 = (1/3, 1/3, 1/3)$ を初期分布とすると, $N = 500$ に対して,

$$L(500) = 0.161 \div -\log \theta_{\max} = 0.157. \quad (47)$$

となり, $L(N)$ が定理 5 の理論値に収束していることがわかる。図 2 参照。

8.2 (ii) の場合: $1/N$ オーダの収束

例 5 (13) の通信路行列 $\Phi^{(2)}$ を考える。すなわち,

$$\Phi^{(2)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{pmatrix}. \quad (48)$$

このとき,

$$\lambda^* = (0.500, 0.500, 0.000), Q^* = (0.450, 0.450, 0.100),$$

$$J(\lambda^*) = \begin{pmatrix} 0.228 & -0.228 & 0.000 \\ -0.228 & 0.228 & 0.000 \\ -0.500 & -0.500 & 1.000 \end{pmatrix},$$

$$H_3(\lambda^*) = \begin{pmatrix} 0.000 & 0.000 & -1.000 \\ 0.000 & 0.000 & -1.000 \\ -1.000 & -1.000 & -3.990 \end{pmatrix}$$

となる。 $J(\lambda^*)$ の固有値は $(\theta_1, \theta_2, \theta_3) = (0.000, 0.456, 1.000)$ である。 $N = 500$ に対して,

$$N\mu^N = (-0.510, -0.510, 1.019) \quad (49)$$

$$\div \lim_{N \rightarrow \infty} N\mu^N = (-0.503, -0.503, 1.005). \quad (50)$$

となる。(50) は定理 7 の結果である。大きい N に対して $N\mu^N$ が定理 7 の極限值に収束していることが確認できる。図 3 参照。

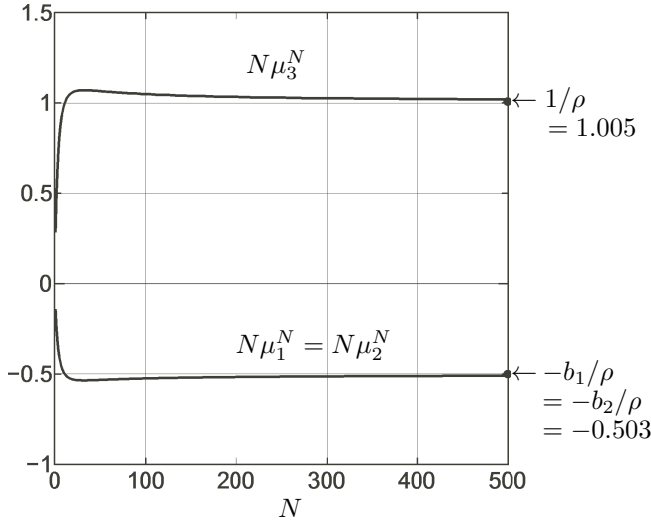


図 3: 例 5 における $N\mu_i^N$ の収束

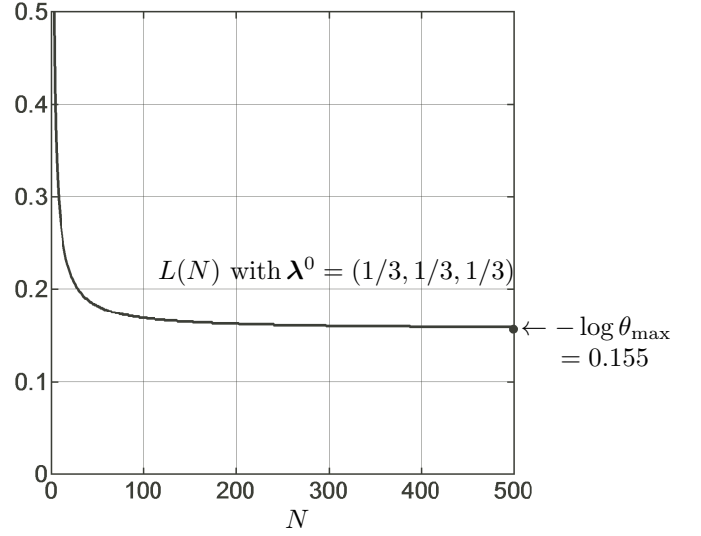


図 4: 例 6 における $L(N)$ の収束

8.3 (iii) の場合: $\lambda^* \in \partial\Delta(\mathcal{X})$ のときの指数関数的収束

例 6 (14) の通信路行列 $\Phi^{(3)}$ を考える。すなわち、

$$\Phi^{(3)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{pmatrix}. \quad (51)$$

このとき、

$$\lambda^* = (0.500, 0.500, 0.000), Q^* = (0.450, 0.450, 0.100),$$

$$J(\lambda^*) = \begin{pmatrix} 0.228 & -0.228 & 0.000 \\ -0.228 & 0.228 & 0.000 \\ -0.428 & -0.428 & 0.856 \end{pmatrix}$$

であり、 $J(\lambda^*)$ の固有値は $(\theta_1, \theta_2, \theta_3) = (0.000, 0.456, 0.856)$ となり、 $\theta_{\max} = \theta_3 = 0.856$ である。

$\lambda^0 = (1/3, 1/3, 1/3)$ を初期分布とすると、 $N = 500$ に対して

$$L(500) = 0.159 \div -\log \theta_{\max} = 0.155 \quad (52)$$

となり、 $L(N)$ が定理 5 の理論値に収束していることがわかる。図 4 参照。

9 まとめ

本研究において、Arimoto アルゴリズムの定義関数の Taylor 展開に基づいてその収束速度を調べた。具体的に Jacobi 行列 J と Hesse 行列 H を計算した。その結果、次の 3 つの条件が同値であることがわかった。すなわち、タイプ II のインデックスが存在すること、Jacobi 行列の最大固有値 $\theta_{\max} = 1$ であること、収束が $1/N$ オーダであること、の 3 つの条件である。さらに、 $1/N$ オーダの収束のときその収束速度を Kullback-Leibler 情報量の入力確率による微分によって表した。Hesse 行列による $1/N$ オーダの収束の解析は本研究で初めて行われた。

これらの解析に基づいて、いくつかの通信路行列に対する Arimoto アルゴリズムの収束速度を数値的に評価し、我々の定理で得られた理論値と精度良く一致することが確認できた。

10 謝辞

本研究は JSPS 科研費 JP17K00008 の助成を受けたものである。

参考文献

- [1] S. Arimoto, “An algorithm for computing the capacity of arbitrary discrete memoryless channels”, IEEE Trans. Inf. Theory, vol. 18, pp.14-20, Jan. 1972.
- [2] I. Csiszàr and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Orlando, 1982.
- [3] I. Csiszàr and G. Tusnády, “Information Geometry and Alternating Minimization Procedures”, Statistics and Decisions, Supplement Issue No.1, 205-237, 1984.
- [4] G. Matz and P. Duhamel, “Information Geometric Formulation and Interpretation of Accelerated Blahut-Arimoto-Type Algorithms”, in Proceedings of ITW2004, 2004.
- [5] I. Naiss and H. Permuter, “Extension of the Blahut-Arimoto Algorithm for Maximizing Directed Information”, IEEE Trans. Inf. Theory, vol. 59, no. 1, pp.204-222, Jan. 2013.
- [6] Z. Naja, F. Alberge, and P. Duhamel, “Geometrical interpretation and improvements of the Blahut-Arimoto’s algorithm”, in proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2009.
- [7] K. Nakagawa, Y. Takei and K. Watabe, “Analysis for the Slow Convergence in Arimoto Algorithm”, <http://arxiv.org/abs/1809.00752>
- [8] M. Rezaeian and A. Grant, “A Generalization of Arimoto-Blahut Algorithm”, in proceedings of IEEE ISIT 2004.
- [9] P. O. Vontobel, A. Kačičić, D. M. Arnold, H.-A. Loeliger, “A Generalization of the Blahut-Arimoto Algorithm to Finite-State Channels”, IEEE Trans. Inf. Theory, vol. 54, no. 5, pp.1887-1918, 2008.
- [10] Y. Yu, “Squeezing the Arimoto-Blahut Algorithm for Faster Convergence”, IEEE Trans. Inf. Theory, vol. 56, no. 7, pp.3149-3157, Jul. 2010.