Analysis of the Convergence Speed of the Arimoto-Blahut Algorithm by the Second-Order Recurrence Formula

Kenji Nakagawa, *Member, IEEE,* Yoshinori Takei, *Member, IEEE,* Shin-ichiro Hara, *non-member,* and Kohei Watabe, *Member, IEEE,*

***Abstract*—In this paper, we investigate the convergence speed of the Arimoto-Blahut algorithm. For many channel matrices, the convergence speed is exponential, but for some channel matrices it is slower than exponential. By analyzing the Taylor expansion of the defining function of the Arimoto-Blahut algorithm, we will make the conditions clear for the exponential or slower convergence. The analysis of the slow convergence in this paper is new. Based on this analysis, we will compare the convergence speeds of the Arimoto-Blahut algorithm numerically with the values obtained in our theorems for several channel matrices. The purpose of this paper is to obtain a complete understanding of the convergence speed of the Arimoto-Blahut algorithm.**

***Index Terms*—channel capacity, discrete memoryless channel, Arimoto-Blahut algorithm, convergence speed, Hessian matrix, second-order recurrence formula.**

I. Introduction

We consider a discrete memoryless channel with the input and output alphabets of finite sizes. The channel capacity C of the channel is defined as the maximum value of the mutual information between the input and output, where the maximum is taken over the entire input probability distributions. By Shannon's channel coding theorem, C is equal to the maximum information rate that can be transmitted with arbitrarily small error probability.

Methods for calculating the value of the channel capacity C fall into two categories. One is the direct method of solving the mutual information maximization problem [14], [15]. This is a convex optimization problem regarding the input distribution and is generally considered to be solved by the method of Lagrange multipliers. However, the constraint that the probability must be between 0 and 1 should be carefully considered, hence the channels where the channel capacity can be calculated explicitly are limited.

The other is a sequential computation algorithm represented by the Arimoto-Blahut algorithm [2], [3], [4]. In the Arimoto-Blahut algorithm, a series of input probability distributions are computed by a recurrence formula, and they converge to the input probability distribution that achieves the channel capacity C. The Arimoto-Blahut algorithm can be applied to any channel, and its capacity is computed.

Based on the original Arimoto-Blahut algorithm, much work has been done. In [12], [17], [20], applicable channels were extended to include more than just discrete memoryless channels. In [11], [21], algorithms to accelerate the Arimoto-Blahut algorithm were proposed. Furthermore, in [8], [11], [13], [14], the Arimoto-Blahut algorithm is characterized by divergence geometry.

In this paper, we study the convergence speed of the Arimoto-Blahut algorithm. Although the convergence speed is exponential for many channels, we found that for some channels, the convergence speed is slower than exponential. Exponential convergence has been studied in [2], [3], [21], but no study of convergence speed other than exponential has been conducted.

A. Purpose of Study

The purpose of our study is to classify discrete memoryless channels and to determine when the Arimoto-Blahut algorithm converges quickly and when it converges slowly.

First, we show that only two kinds of convergence speeds occur. One is the exponential convergence, which is a fast convergence, and the other is the convergence of order $O(1/N)$, which is a slow convergence. Then, we evaluate the convergence speed theoretically in each case.

Exponential convergence is evaluated by the eigenvalues of the Jacobian matrix of the defining function F of the Arimoto-Blahut algorithm. Its analysis is not difficult. On the other hand, a convergence speed of order $O(1/N)$ is determined not only by the Jacobian matrix of F but also requires the analysis of the Hessian matrix, i.e., the second derivative of F. The convergence speed of the order $O(1/N)$ is analyzed for the first time in this paper. The Hessian matrix of F yields a second-order nonlinear recurrence formula, and then we will investigate the convergence speed of the recurrence formula. In general, a nonlinear recurrence formula can rarely be solved in an explicit form, and the recurrence formula in this paper also cannot be solved explicitly. Therefore, a method to investigate the speed of convergence must be devised. The key lemma is Lemma 6 on the convergence speed of a recurrence formula defined by a quadratic polynomial in one variable. In this

The material in this paper was presented in part at the 2018 International Symposium on Information Theory and its Applications (ISITA2018).
K. Nakagawa and K. Watabe are with the Department of Electrical and Electronics and Information Engineering, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan (e-mail: nakagawa@nagaokaut.ac.jp; k_watabe@vos.nagaokaut.ac.jp). Y. Takei is with the National Institute of Technology, Akita College, Akita 011-8511, Japan (e-mail: ytakei@akita-nct.ac.jp). S. Hara is with the Department of General Education, Nagaoka University of Technology (e-mail: sinara@blade.nagaokaut.ac.jp).

Manuscript received April 19, 2005; revised September 17, 2014.

paper, we also investigate the speed at which the mutual information converges to the channel capacity C.

Based on these analyses, we numerically calculate the convergence speeds of the Arimoto-Blahut algorithm for several channels and compare them with the theoretical values obtained in this paper.

II. Arimoto-Blahut algorithm

A. Channel matrix and channel capacity

Consider a discrete memoryless channel $X \to Y$ with the input source X and the output source Y. Let $\mathcal{X} = \{x_1, \dots, x_m\}$ be the input alphabet and $\mathcal{Y} = \{y_1, \dots, y_n\}$ be the output alphabet. The conditional probability that the output symbol y_j is received when the input symbol x_i is transmitted is denoted by $P^i_j = P(Y = y_j | X = x_i)$, $i = 1, \dots, m, j = 1, \dots, n$, and the row vector P^i is defined by $P^i = (P^i_1, \dots, P^i_n)$, $i = 1, \dots, m$. The channel matrix Φ is defined by

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P^1_1 & \dots & P^1_n \\ \vdots & & \vdots \\ P^m_1 & \dots & P^m_n \end{pmatrix}. \tag{1}$$

We assume that for any $j\,(j = 1, \dots, n)$, there exists at least one $i\,(i = 1, \dots, m)$ with $P^i_j > 0$. This assumption means that there are no useless output symbols.

The set of input probability distributions on the input alphabet $\mathcal{X}$ is denoted by $\Delta(\mathcal{X}) \equiv \{\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) | \lambda_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \lambda_i = 1\}$. The interior of $\Delta(\mathcal{X})$ is denoted by $\Delta(\mathcal{X})^\circ \equiv \{\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) \in \Delta(\mathcal{X}) \,|\, \lambda_i > 0,\ i = 1, \dots, m\}$. Similarly, the set of output probability distributions on the output alphabet $\mathcal{Y}$ is denoted by $\Delta(\mathcal{Y}) \equiv \{Q = (Q_1, \dots, Q_n) | Q_j \geq 0, j = 1, \dots, n, \sum_{j=1}^n Q_j = 1\}$, and its interior $\Delta(\mathcal{Y})^\circ$ is similarly defined as $\Delta(\mathcal{X})^\circ$.

Let $Q = \boldsymbol{\lambda}\Phi$ be the output distribution for an input distribution $\boldsymbol{\lambda} \in \Delta(\mathcal{X})$ and write its components as $Q_j = \sum_{i=1}^m \lambda_i P^i_j$, $j = 1, \dots, n$. Then, the mutual information is defined by $I(\boldsymbol{\lambda}, \Phi) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i P^i_j \log \left(P^i_j / Q_j\right)$, where $\log$ is the natural logarithm. The channel capacity C is defined by

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

The unit of C is nat/symbol.

The Kullback-Leibler divergence $D(Q \| Q')$ for two output distributions $Q = (Q_1, \dots, Q_n)$, $Q' = (Q'_1, \dots, Q'_n) \in \Delta(\mathcal{Y})$ is defined [7] by

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

An important proposition for investigating the convergence speed of the Arimoto-Blahut algorithm is the Kuhn-Tucker condition on the input distribution $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ that achieves the maximum of (2).

Theorem: (Kuhn-Tucker condition [6]) In the maximization problem (2), a necessary and sufficient condition for the input distribution $\boldsymbol{\lambda}^* = (\lambda^*_1, \dots, \lambda^*_m) \in \Delta(\mathcal{X})$ to achieve the maximum is that there is a certain constant $\tilde{C}$ with

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

In (4), $\tilde{C}$ is equal to the channel capacity C.

Since the Kuhn-Tucker condition is a necessary and sufficient condition, all the information about the capacity-achieving input distribution $\boldsymbol{\lambda}^*$ can be derived from this condition.

B. Definition of the Arimoto-Blahut algorithm

A sequence $\{\boldsymbol{\lambda}^N = (\lambda^N_1, \dots, \lambda^N_m)\}_{N=0,1,\dots} \subset \Delta(\mathcal{X})$ of input distributions is defined by the Arimoto-Blahut algorithm as follows [2], [4]. First, let $\boldsymbol{\lambda}^0 = (\lambda^0_1, \dots, \lambda^0_m)$ be an initial distribution taken in $\Delta(\mathcal{X})^\circ$, i.e., $\lambda^0_i > 0,\ i = 1, \dots, m$. Then, the Arimoto-Blahut algorithm is given by the recurrence formula

$$\lambda^{N+1}_i = \frac{\lambda^N_i \exp D(P^i \| \boldsymbol{\lambda}^N \Phi)}{\displaystyle\sum_{k=1}^m \lambda^N_k \exp D(P^k \| \boldsymbol{\lambda}^N \Phi)}, \tag{5}$$
$$i = 1, \dots, m, N = 0, 1, \dots.$$

On the convergence of this Arimoto-Blahut algorithm, the following results were obtained by Arimoto in [2], [3].

In [2], [3], defining

$$C(N+1, N) \equiv - \sum_{i=1}^m \lambda^{N+1}_i \log \lambda^{N+1}_i + \sum_{i=1}^m \sum_{j=1}^n \lambda^{N+1}_i P^i_j \log \frac{\lambda^N_i P^i_j}{\displaystyle\sum_{k=1}^m \lambda^N_k P^k_j} \tag{6}$$

$$= -D(\boldsymbol{\lambda}^{N+1} \| \boldsymbol{\lambda}^N) + \sum_{i=1}^m \lambda^{N+1}_i D(P^i \| Q^N), \tag{7}$$

Arimoto obtained the following results.

Theorem A1: [2] If the initial input distribution $\boldsymbol{\lambda}^0$ is in $\Delta(\mathcal{X})^\circ$, then

$$\lim_{N \to \infty} C(N+1, N) = C. \tag{8}$$

Theorem A2: [2] If $\boldsymbol{\lambda}^0$ is the uniform distribution, then

$$0 \leq C - C(N+1, N) \leq \frac{\log m - h(\boldsymbol{\lambda}^*)}{N}, \tag{9}$$

where $\boldsymbol{\lambda}^*$ is the capacity-achieving input distribution, and $h(\boldsymbol{\lambda}^*)$ is the entropy of $\boldsymbol{\lambda}^*$.

Theorem A3: [3] Assume that $\boldsymbol{\lambda}^*$ is unique and belongs to $\Delta(\mathcal{X})^\circ$. Then,

i) $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*,\ N \to \infty$, and
ii) for sufficiently small arbitrary $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that

$$0 \leq C - C(N+1, N) \leq \epsilon \cdot (\theta)^{N - N_0},\ N \geq N_0, \tag{10}$$

where θ is a constant with $0 \le \theta < 1$ and is unrelated to ϵ and N_0. Further, in the expression $\epsilon \cdot (\theta)^{N-N_0}$, $\cdot$ denotes multiplication, and $(\theta)^{N-N_0}$ means θ to the power of $N - N_0$.

We can see from Theorem A3, under the assumptions of Theorem A3, that the first term of (7) converges to 0, and the second term converges to C.

C. Function from $\Delta(\mathcal{X})$ to $\Delta(\mathcal{X})$

Let $F_i(\boldsymbol{\lambda})$ be the defining function of the Arimoto-Blahut algorithm (5), i.e.,

$$F_i(\boldsymbol{\lambda}) = \frac{\lambda_i \exp D(P^i \| \boldsymbol{\lambda}\Phi)}{\sum_{k=1}^{m} \lambda_k \exp D(P^k \| \boldsymbol{\lambda}\Phi)},\ i = 1, \ldots, m. \tag{11}$$

Define $F(\boldsymbol{\lambda}) \equiv (F_1(\boldsymbol{\lambda}), \ldots, F_m(\boldsymbol{\lambda}))$. Then, $F(\boldsymbol{\lambda})$ is a differentiable function from $\Delta(\mathcal{X})$ to $\Delta(\mathcal{X})$, and (5) is represented by $\boldsymbol{\lambda}^{N+1} = F(\boldsymbol{\lambda}^N)$.

In this paper, for the analysis of the convergence speed, we assume

$$\operatorname{rank} \Phi = m. \tag{12}$$

Concerning this assumption, we see that in [2], [3], for the analysis of the convergence speed, the uniqueness of the capacity-achieving $\boldsymbol{\lambda}^*$ is assumed, which is a necessary condition for (12). In fact, we have

Lemma 1: The capacity-achieving input distribution $\boldsymbol{\lambda}^*$ is unique under the assumption (12).

Proof: By Csiszár[7], p.137, Eq. (37), for arbitrary $Q \in \Delta(\mathcal{Y})$,

$$\sum_{i=1}^{m} \lambda_i D(P^i \| Q) = I(\boldsymbol{\lambda}, \Phi) + D(\boldsymbol{\lambda}\Phi \| Q). \tag{13}$$

By the assumption (12), we see that there exists $Q^0 \in \Delta(\mathcal{Y})$ [15] with

$$D(P^1 \| Q^0) = \ldots = D(P^m \| Q^0) \equiv C^0. \tag{14}$$

Substituting $Q = Q^0$ into (13), we have $C^0 = I(\boldsymbol{\lambda}, \Phi) + D(\boldsymbol{\lambda}\Phi \| Q^0)$. Because C^0 is a constant,

$$\max_{\boldsymbol{\lambda} \in \Delta(\mathcal{X})} I(\boldsymbol{\lambda}, \Phi) \Longleftrightarrow \min_{\boldsymbol{\lambda} \in \Delta(\mathcal{X})} D(\boldsymbol{\lambda}\Phi \| Q^0). \tag{15}$$

Define $W \equiv \{\boldsymbol{\lambda}\Phi \mid \boldsymbol{\lambda} \in \Delta(\mathcal{X})\}$. Then, W is a closed convex set, and thus by Cover [6], p.297, Theorem 12.6.1, $Q = Q^*$, which achieves $\min_{Q \in W} D(Q \| Q^0)$, exists and is unique. By the assumption (12), the mapping $\Delta \ni \boldsymbol{\lambda} \mapsto \boldsymbol{\lambda}\Phi \in W$ is one to one; therefore, $\boldsymbol{\lambda}^*$ with $Q^* = \boldsymbol{\lambda}^*\Phi$ is unique. ■

Remark 1: Because of the equivalence (15), the Arimoto-Blahut algorithm can be obtained by Csiszár [8], Chapter 4, "Minimizing information distance from a single measure", Theorem 5.

Lemma 2: The capacity-achieving input distribution $\boldsymbol{\lambda}^*$ is the fixed point of the function $F(\boldsymbol{\lambda})$. That is, $F(\boldsymbol{\lambda}^*) = \boldsymbol{\lambda}^*$.

Proof: In the Kuhn-Tucker condition (4), let us define m_1 as the number of indices i with $\lambda_i^* > 0$, i.e.,

$$\lambda_i^* \begin{cases} > 0, & i = 1, \ldots, m_1, \\ = 0, & i = m_1 + 1, \ldots, m, \end{cases} \tag{16}$$

by reordering the input symbols (if necessary), then

$$D(P^i \| \boldsymbol{\lambda}^*\Phi) \begin{cases} = C, & i = 1, \ldots, m_1, \\ \le C, & i = m_1 + 1, \ldots, m. \end{cases} \tag{17}$$

We have

$$\sum_{k=1}^{m} \lambda_k^* \exp D(P^k \| \boldsymbol{\lambda}^*\Phi) = \sum_{k=1}^{m_1} \lambda_k^* e^C = e^C, \tag{18}$$

hence, by (11), (16), and (18),

$$F_i(\boldsymbol{\lambda}^*) = \begin{cases} e^{-C} \lambda_i^* e^C, & i = 1, \ldots, m_1, \\ 0, & i = m_1 + 1, \ldots, m, \end{cases} \tag{19}$$

$$= \lambda_i^*,\ i = 1, \ldots, m, \tag{20}$$

which shows $F(\boldsymbol{\lambda}^*) = \boldsymbol{\lambda}^*$. ■

The sequence $\{\boldsymbol{\lambda}^N\}_{N=0,1,\ldots}$ of the Arimoto-Blahut algorithm converges to the fixed point $\boldsymbol{\lambda}^*$, i.e., $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$, $N \to \infty$. We will investigate the convergence speed by using the Taylor expansion of $F(\boldsymbol{\lambda})$ about $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$.

D. Convergence speed of $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$

Now, we define two kinds of convergence speeds for investigating $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$.

(i) Exponential convergence
$\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$ is the *exponential convergence* if

$$\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\| < K \cdot (\theta)^N, N = 0, 1, \ldots, \quad K > 0,\ 0 \le \theta < 1, \tag{21}$$

where $\|\boldsymbol{\lambda}\|$ denotes the Euclidean norm, i.e., $\|\boldsymbol{\lambda}\| = \left(\lambda_1^2 + \ldots + \lambda_m^2\right)^{1/2}$.

(ii) Convergence of order $O(1/N)$
$\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$ is the *convergence of order* $O(1/N)$ if

$$\lim_{N \to \infty} N\left(\lambda_i^N - \lambda_i^*\right) = K_i \ne 0,\ i = 1, \ldots, m. \tag{22}$$

E. Type of index

Now, we classify the indices $i\,(i = 1, \ldots, m)$ in the Kuhn-Tucker condition (4) in more detail into the following 3 types.

$$D(P^i \| \boldsymbol{\lambda}^*\Phi) \begin{cases} = C, & \text{for } i \text{ with } \lambda_i^* > 0 \text{ [type-I]}, \\ = C, & \text{for } i \text{ with } \lambda_i^* = 0 \text{ [type-II]}, \\ < C, & \text{for } i \text{ with } \lambda_i^* = 0 \text{ [type-III]}. \end{cases} \tag{23}$$

Let us define the sets of indices as follows.

$$\text{all the indices} : \mathcal{I} \equiv \{1, \ldots, m\}, \tag{24}$$

$$\text{type-I indices: } \mathcal{I}_{\mathrm{I}} \equiv \{1, \ldots, m_1\}, \tag{25}$$

$$\text{type-II indices: } \mathcal{I}_{\mathrm{II}} \equiv \{m_1 + 1, \ldots, m_1 + m_2\}, \tag{26}$$

$$\text{type-III indices: } \mathcal{I}_{\mathrm{III}} \equiv \{m_1 + m_2 + 1, \ldots, m\}. \tag{27}$$

We have $|\mathcal{I}| = m$, $|\mathcal{I}_{\mathrm{I}}| = m_1$, $|\mathcal{I}_{\mathrm{II}}| = m_2$, $|\mathcal{I}_{\mathrm{III}}| = m - m_1 - m_2 \equiv m_3$, $\mathcal{I} = \mathcal{I}_{\mathrm{I}} \cup \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}}$ and $m = m_1 + m_2 + m_3$. $\mathcal{I}_{\mathrm{I}}$ is not empty, and $|\mathcal{I}_{\mathrm{I}}| = m_1 \ge 2$ for any channel matrix, but $\mathcal{I}_{\mathrm{II}}$ and $\mathcal{I}_{\mathrm{III}}$ may be empty for some channel matrices.

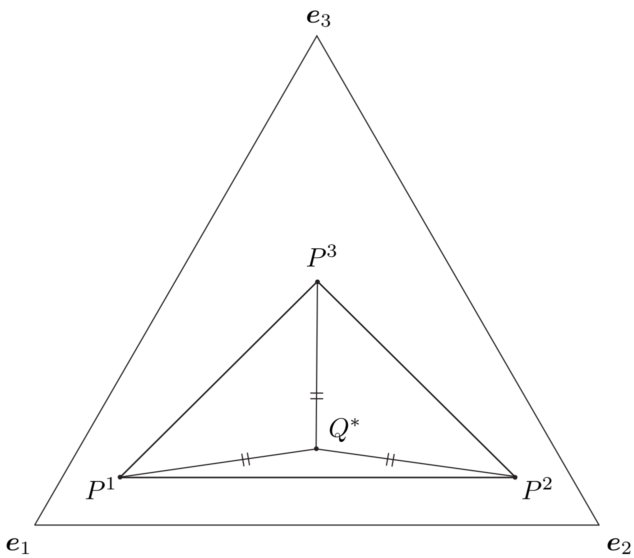

Fig. 1. Positional relation of row vectors P^1, P^2, P^3 of $\Phi^{(1)}$ and Q^* in Example 1.

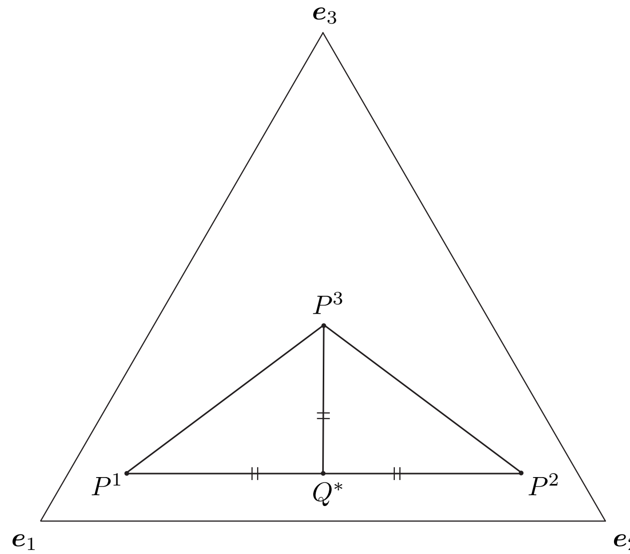

Fig. 2. Positional relation of row vectors P^1, P^2, P^3 of $\Phi^{(2)}$ and Q^* in Example 2.

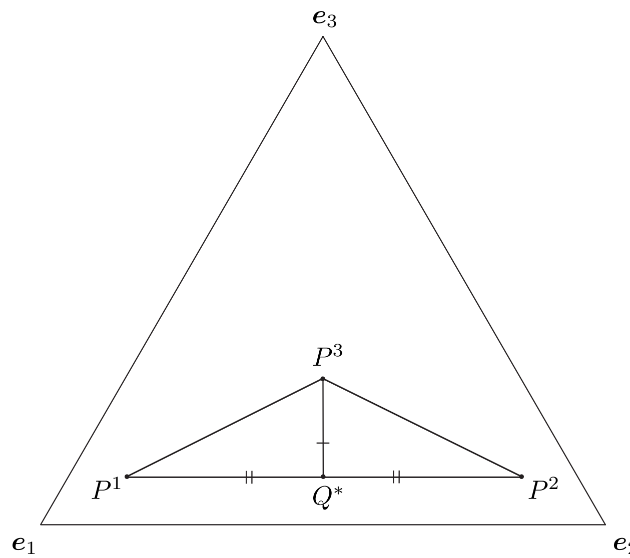

Fig. 3. Positional relation of row vectors P^1, P^2, P^3 of $\Phi^{(3)}$ and Q^* in Example 3.

F. Examples of convergence speed

Let us consider the dependence of the convergence speed of the Arimoto-Blahut algorithm on the channel matrices.

For many channel matrices Φ, the convergence is exponential, but for some special Φ, the convergence is very slow. Let us consider the following three examples with input alphabet size $m = 3$ and output alphabet size $n = 3$, taking types-I, -II and -III into account.

Example 1: (only type-I) If only type-I indices exist, then $\lambda_i^* > 0,\ i = 1, 2, 3$; hence, $Q^* \equiv \boldsymbol{\lambda}^*\Phi$ is in the interior of $\triangle P^1P^2P^3$. See Fig. 1. As a concrete channel matrix of this example, let us consider

$$\Phi^{(1)} \equiv \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}. \tag{28}$$

For this $\Phi^{(1)}$, we have $C = 0.323\,$[nat/symbol], $\boldsymbol{\lambda}^* = (0.431, 0.431, 0.138)$ and $Q^* = (0.422, 0.422, 0.156)$. The vertices of the large triangle in Fig. 1 are $\boldsymbol{e}_1 = (1, 0, 0)$, $\boldsymbol{e}_2 = (0, 1, 0)$, $\boldsymbol{e}_3 = (0, 0, 1)$. We have $D(P^i\|Q^*) = C,\ i = 1, 2, 3$, then considering the analogy to Euclidean geometry, $\triangle P^1P^2P^3$ can be regarded as an "acute triangle."

Example 2: (types-I and -II) If there are type-I and type-II indices, we can assume $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0$ without loss of generality; hence, Q^* is on the side P^1P^2 and $D(P^i\|Q^*) = C,\ i = 1, 2, 3$. See Fig. 2. As a concrete channel matrix of this example, let us consider

$$\Phi^{(2)} \equiv \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{pmatrix}. \tag{29}$$

For this $\Phi^{(2)}$, we have $C = 0.310\,$[nat/symbol], $\boldsymbol{\lambda}^* = (0.500, 0.500, 0.000)$ and $Q^* = (0.450, 0.450, 0.100)$. Considering the analogy to Euclidean geometry, $\triangle P^1P^2P^3$ can be regarded as a "right triangle."

Example 3: (types-I and -III) If there are type-I and type-III indices, we can assume $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0$ without loss of generality; hence, Q^* is on the side P^1P^2 and $C = D(P^1\|Q^*) = D(P^2\|Q^*) > D(P^3\|Q^*)$. See Fig. 3. As a concrete channel matrix of this example, let us consider

$$\Phi^{(3)} \equiv \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{pmatrix}. \tag{30}$$

For this $\Phi^{(3)}$, we have $C = 0.310\,$[nat/symbol], $\boldsymbol{\lambda}^* = (0.500, 0.500, 0.000)$ and $Q^* = (0.450, 0.450, 0.100)$. Considering the analogy to Euclidean geometry, $\triangle P^1P^2P^3$ can be regarded as an "obtuse triangle."

For the above $\Phi^{(1)}, \Phi^{(2)}, \Phi^{(3)}$, we show in Fig. 4 the speed of convergence of $|\lambda_1^N - \lambda_1^*| \to 0$. By Fig. 4, we see that in Examples 1 and 3, the convergence is exponential, while in Example 2, the convergence is slower than exponential.

From the above three examples, it is inferred that the Arimoto-Blahut algorithm converges exponentially when type-

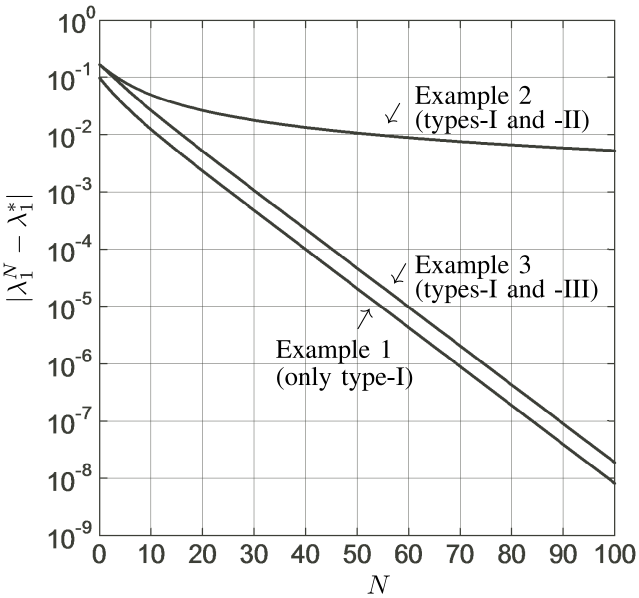

Fig. 4. Comparison of the convergence speed in Examples 1,2,3.

II indices do not exist and slowly when type-II indices exist. We will analyze this phenomenon in this paper.

Remark 2: Related to the above investigation of triangles, we have given a geometric investigation of the channel capacity in [15] based on divergence geometry.

G. Analysis of convergence speed

We will show in this paper for which channel the convergence speed of the Arimoto-Blahut algorithm is exponential and for which channel it is of order $O(1/N)$.

In previous studies [2], [21], the exponential convergence has been investigated. In [2], [21], the authors proved that the convergence speed is exponential if the capacity-achieving input distribution $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)$ satisfies $\lambda_i^* > 0,\ i = 1, \ldots, m$. In this case, all the indices are of type-I and there are no type-II indices. When there are no type-II indices, the convergence speed is determined by the Jacobian matrix of the defining function $F(\boldsymbol{\lambda})$ of the Arimoto-Blahut algorithm.

In this paper, we will investigate the case in which type-II indices exist. If there exist type-II indices, the convergence speed of the Arimoto-Blahut algorithm is determined not only by the Jacobian matrix, but also by the Hessian matrix. We will analyze the nonlinear recurrence formula defined by a quadratic polynomial, which is obtained from the Hessian matrix. Then, we will show, under some assumptions, that the convergence speed of the recurrence formula is of order $O(1/N)$.

III. Taylor expansion of $F(\boldsymbol{\lambda})$ about $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$

We will examine the convergence speed of the Arimoto-Blahut algorithm by the Taylor expansion of $F(\boldsymbol{\lambda})$ about the fixed point $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. The Taylor expansion of the function $F(\boldsymbol{\lambda}) = (F_1(\boldsymbol{\lambda}), \ldots, F_m(\boldsymbol{\lambda}))$ about $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ is

$$\begin{aligned} F(\boldsymbol{\lambda}) &= F(\boldsymbol{\lambda}^*) + (\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) J(\boldsymbol{\lambda}^*) \\ &+ \frac{1}{2!}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) H(\boldsymbol{\lambda}^*)\, {}^t(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) + o\left(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*\|^2\right), \end{aligned} \tag{31}$$

where ${}^t\boldsymbol{\lambda}$ denotes the transpose of $\boldsymbol{\lambda}$, and $o\left(\|\boldsymbol{\lambda}\|^2\right)$ means $\lim_{\|\boldsymbol{\lambda}\| \to 0} o\left(\|\boldsymbol{\lambda}\|^2\right) / \|\boldsymbol{\lambda}\|^2 = 0$.

In (31), $J(\boldsymbol{\lambda}^*)$ is the Jacobian matrix of $F(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, i.e.,

$$J(\boldsymbol{\lambda}^*) = \left(\left. \frac{\partial F_i}{\partial \lambda_{i'}} \right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} \right)_{i', i = 1, \ldots, m}. \tag{32}$$

In this paper, we assume that the input probability distribution $\boldsymbol{\lambda}$ is a row vector; thus, the Jacobian matrix $J(\boldsymbol{\lambda}^*)$ is

$$J(\boldsymbol{\lambda}^*) = \begin{array}{c} \\ \uparrow \\ i' \\ \downarrow \end{array} \overset{\leftarrow i \rightarrow}{\begin{pmatrix} \left. \dfrac{\partial F_1}{\partial \lambda_1} \right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} & \cdots & \left. \dfrac{\partial F_m}{\partial \lambda_1} \right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} \\ \vdots & & \vdots \\ \left. \dfrac{\partial F_1}{\partial \lambda_m} \right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} & \cdots & \left. \dfrac{\partial F_m}{\partial \lambda_m} \right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} \end{pmatrix}} \in \mathbb{R}^{m \times m}, \tag{33}$$

i.e., $\partial F_i / \partial \lambda_{i'} |_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*}$ is the (i', i) component.

Lemma 3: Every row sum of $J(\boldsymbol{\lambda}^*)$ is equal to 0.

Proof: By (11), we have $\sum_{i=1}^m F_i(\boldsymbol{\lambda}) = 1$; thus, by (32), the lemma holds. ■

In (31), $H(\boldsymbol{\lambda}^*) \equiv (H_1(\boldsymbol{\lambda}^*), \ldots, H_m(\boldsymbol{\lambda}^*))$, where $H_i(\boldsymbol{\lambda}^*)$ is the Hessian matrix of $F_i(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, i.e.,

$$H_i(\boldsymbol{\lambda}^*) = \left(\left. \frac{\partial^2 F_i}{\partial \lambda_{i'} \partial \lambda_{i''}} \right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} \right)_{i', i'' = 1, \ldots, m}, \tag{34}$$

and $(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) H(\boldsymbol{\lambda}^*)\, {}^t(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*)$ is an abbreviated expression of the m dimensional row vector
$\left((\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) H_1(\boldsymbol{\lambda}^*)\, {}^t(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*), \ldots, (\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) H_m(\boldsymbol{\lambda}^*)\, {}^t(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*)\right)$.

Remark 3: $\lambda_1, \ldots, \lambda_m$ satisfy the constraint $\sum_{i=1}^m \lambda_i = 1$, but in (31), (32), (34) we consider $\lambda_1, \ldots, \lambda_m$ as independent (or constraint free) variables to have the Taylor series approximation (31). This approximation is justified as follows. By the Kuhn-Tucker condition (4), $D(P^i \| Q^*) \le C < \infty,\ i = 1, \ldots, m$; hence, by the assumption put below (1), we have $Q_j^* > 0,\ j = 1, \ldots, n$. See [2]. For $\epsilon > 0$, define $\mathcal{Q}_\epsilon^* \equiv \{Q = (Q_1, \ldots, Q_n) \in \mathbb{R}^n \mid \|Q - Q^*\| < \epsilon\}$, i.e., $\mathcal{Q}_\epsilon^*$ is an open ball in $\mathbb{R}^n$ centered at Q^* with radius ϵ. Note that $Q \in \mathcal{Q}_\epsilon^*$ is free from the constraint $\sum_{j=1}^n Q_j = 1$. Taking $\epsilon > 0$ sufficiently small, we can have $Q_j > 0, j = 1, \ldots, n$, for any $Q \in \mathcal{Q}_\epsilon^*$. The function $F(\boldsymbol{\lambda})$ is defined for $\boldsymbol{\lambda}$ with $Q_j = (\boldsymbol{\lambda}\Phi)_j > 0,\ j = 1, \ldots, n$, even if some $\lambda_i < 0$. Therefore, the domain of definition of $F(\boldsymbol{\lambda})$ can be extended to $\Phi^{-1}(\mathcal{Q}_\epsilon^*) \subset \mathbb{R}^m$, where $\Phi^{-1}(\mathcal{Q}_\epsilon^*)$ is the inverse image of $\mathcal{Q}_\epsilon^*$ by the mapping $\mathbb{R}^m \ni \boldsymbol{\lambda} \mapsto \boldsymbol{\lambda}\Phi \in \mathbb{R}^n$. $\Phi^{-1}(\mathcal{Q}_\epsilon^*)$ is an open neighborhood of $\boldsymbol{\lambda}^*$ in $\mathbb{R}^m$. Then, $F(\boldsymbol{\lambda})$ is a function of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) \in \Phi^{-1}(\mathcal{Q}_\epsilon^*)$ as independent variables (free from the constraint $\sum_{i=1}^m \lambda_i = 1$). We can consider (31) to be the Taylor expansion by independent variables $\lambda_1, \ldots, \lambda_m$;

then, substituting $\boldsymbol{\lambda} \in \Delta(\mathcal{X}) \cap \Phi^{-1}(\mathcal{Q}_\epsilon^*)$ into (31), we obtain the approximation for $F(\boldsymbol{\lambda})$ about $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$.

Now, substituting $\boldsymbol{\lambda} = \boldsymbol{\lambda}^N$ into (31) and using $F(\boldsymbol{\lambda}^*) = \boldsymbol{\lambda}^*$ and $F(\boldsymbol{\lambda}^N) = \boldsymbol{\lambda}^{N+1}$, we have

$$\begin{aligned}\boldsymbol{\lambda}^{N+1} &= \boldsymbol{\lambda}^* + (\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*)J(\boldsymbol{\lambda}^*)\\ &+ \frac{1}{2!}(\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*)H(\boldsymbol{\lambda}^*)\,{}^t(\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*) + o\left(\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\|^2\right).\end{aligned} \tag{35}$$

Then, by substituting $\boldsymbol{\mu}^N \equiv \boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*$, (35) becomes

$$\boldsymbol{\mu}^{N+1} = \boldsymbol{\mu}^N J(\boldsymbol{\lambda}^*) + \frac{1}{2!}\boldsymbol{\mu}^N H(\boldsymbol{\lambda}^*)\,{}^t\boldsymbol{\mu}^N + o\left(\|\boldsymbol{\mu}^N\|^2\right). \tag{36}$$

Then, we will investigate the convergence $\boldsymbol{\mu}^N \to \mathbf{0}$, $N \to \infty$, based on the Taylor expansion (36). Let $\mu_i^N \equiv \lambda_i^N - \lambda_i^*$, $i = 1, \ldots, m$, denote the components of $\boldsymbol{\mu}^N = \boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*$; then, we can write $\boldsymbol{\mu}^N$ by components as $\boldsymbol{\mu}^N = (\mu_1^N, \ldots, \mu_m^N)$. We have $\sum_{i=1}^m \mu_i^N = 0$, $N = 0, 1, \ldots$, because $\sum_{i=1}^m \lambda_i^N = \sum_{i=1}^m \lambda_i^* = 1$.

A. The Jacobian matrix $J(\boldsymbol{\lambda}^*)$

Let us consider the Jacobian matrix $J(\boldsymbol{\lambda}^*)$. We are assuming $\operatorname{rank}\Phi = m$ in (12), hence $m \le n$. We will calculate the components (32) of $J(\boldsymbol{\lambda}^*)$.

Defining $D_i \equiv D(P^i \| \boldsymbol{\lambda}\Phi)$ and $F_i \equiv F_i(\boldsymbol{\lambda})$, $i = 1, \ldots, m$, we can write (11) as

$$F_i = \frac{\lambda_i e^{D_i}}{\sum_{k=1}^m \lambda_k e^{D_k}}, \; i = 1, \ldots, m. \tag{37}$$

From (37), it follows that

$$F_i \sum_{k=1}^m \lambda_k e^{D_k} = \lambda_i e^{D_i}. \tag{38}$$

Then, differentiating both sides of (38) with respect to $\lambda_{i'}$, we have

$$\frac{\partial F_i}{\partial \lambda_{i'}} \sum_{k=1}^m \lambda_k e^{D_k} + F_i \frac{\partial}{\partial \lambda_{i'}} \sum_{k=1}^m \lambda_k e^{D_k} = \delta_{i'i} e^{D_i} + \lambda_i e^{D_i} \frac{\partial D_i}{\partial \lambda_{i'}}, \tag{39}$$

where $\delta_{i'i}$ is the Kronecker delta.

Before substituting $\boldsymbol{\lambda} = \boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)$ into both sides of (39), we define the following symbols. Remember that the integer m_1 was defined in (16). See also (25).

Let us define

$$Q^* \equiv Q(\boldsymbol{\lambda}^*) = \boldsymbol{\lambda}^*\Phi, \tag{40}$$

$$Q_j^* \equiv Q(\boldsymbol{\lambda}^*)_j = \sum_{i=1}^m \lambda_i^* P_j^i = \sum_{i=1}^{m_1} \lambda_i^* P_j^i, \; j = 1, \ldots, n, \tag{41}$$

$$D_i^* \equiv D(P^i \| Q^*), \; i = 1, \ldots, m, \tag{42}$$

$$D_{i',i}^* \equiv \left.\frac{\partial D_i}{\partial \lambda_{i'}}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*}, \; i', i = 1, \ldots, m, \tag{43}$$

$$F_i^* \equiv F_i(\boldsymbol{\lambda}^*), \; i = 1, \ldots, m. \tag{44}$$

Lemma 4: We have

$$\left.\sum_{k=1}^m \lambda_k e^{D_k}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} = e^C, \tag{45}$$

$$\frac{\partial D_i}{\partial \lambda_{i'}} = -\sum_{j=1}^n \frac{P_j^{i'} P_j^i}{Q_j}, \; i', i = 1, \ldots, m, \tag{46}$$

$$\left.\frac{\partial}{\partial \lambda_{i'}} \sum_{k=1}^m \lambda_k e^{D_k}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} = e^{D_{i'}^*} - e^C, \; i' = 1, \ldots, m, \tag{47}$$

$$F_i^* = \lambda_i^*, \; i = 1, \ldots, m. \tag{48}$$

Proof: Eq. (45) was proved in (18). Eq. (46) is proved by a simple calculation. Eq. (47) is proved as follows.

$$\begin{aligned}\left.\frac{\partial}{\partial \lambda_{i'}} \sum_{k=1}^m \lambda_k e^{D_k}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} &= \left.\sum_{k=1}^m \left(\delta_{i'k} e^{D_k} + \lambda_k e^{D_k} \frac{\partial D_k}{\partial \lambda_{i'}}\right)\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*}\\ &= e^{D_{i'}^*} + \sum_{k=1}^{m_1} \lambda_k^* e^C \left(-\sum_{j=1}^n \frac{P_j^k P_j^{i'}}{Q_j^*}\right)\\ &= e^{D_{i'}^*} - e^C \sum_{j=1}^n P_j^{i'} \frac{1}{Q_j^*} \sum_{k=1}^{m_1} \lambda_k^* P_j^k\\ &= e^{D_{i'}^*} - e^C.\end{aligned} \tag{49}$$

Note that $Q_j^* > 0$, $j = 1, \ldots, n$, from Remark 3. Eq. (48) is the result of Lemma 2. ■

Substituting the results of Lemma 4 into (39), we have

$$\left.\frac{\partial F_i}{\partial \lambda_{i'}}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} e^C + \lambda_i^* \left(e^{D_{i'}^*} - e^C\right) = \delta_{i'i} e^{D_i^*} + \lambda_i^* e^{D_i^*} D_{i',i}^*. \tag{50}$$

Consequently, we have

Theorem 1: The components of the Jacobian matrix $J(\boldsymbol{\lambda}^*)$ are given as follows.

$$\begin{aligned}&\left.\frac{\partial F_i}{\partial \lambda_{i'}}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*}\\ &= e^{D_i^* - C}\left(\delta_{i'i} + \lambda_i^* D_{i',i}^*\right) + \lambda_i^* \left(1 - e^{D_{i'}^* - C}\right), \; i', i \in \mathcal{I},\end{aligned} \tag{51}$$

$$= \begin{cases} \delta_{i'i} + \lambda_i^* \left(D_{i',i}^* + 1 - e^{D_{i'}^* - C}\right), \; i' \in \mathcal{I}, \; i \in \mathcal{I}_{\mathrm{I}}, \\ \delta_{i'i}, \; i' \in \mathcal{I}, \; i \in \mathcal{I}_{\mathrm{II}}, \\ e^{D_i^* - C} \delta_{i'i}, \; i' \in \mathcal{I}, \; i \in \mathcal{I}_{\mathrm{III}}, \end{cases} \tag{52}$$

where the sets of indices $\mathcal{I}$, $\mathcal{I}_{\mathrm{I}}$, $\mathcal{I}_{\mathrm{II}}$, $\mathcal{I}_{\mathrm{III}}$ were defined in (24)-(27). Note that $D_i^* = C$ for $i \in \mathcal{I}_{\mathrm{I}} \cup \mathcal{I}_{\mathrm{II}}$ and $\lambda_i^* = 0$ for $i \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}}$.

B. Eigenvalues of the Jacobian matrix $J(\boldsymbol{\lambda}^*)$

From (52), we see that the Jacobian matrix $J(\boldsymbol{\lambda}^*)$ is of the form

$$J(\boldsymbol{\lambda}^*) \equiv \begin{pmatrix} J^{\mathrm{I}} & O & O \\ * & J^{\mathrm{II}} & O \\ * & O & J^{\mathrm{III}} \end{pmatrix}, \tag{53}$$

$$J^{\mathrm{I}} \equiv (\partial F_i/\partial \lambda_{i'}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*})_{i,i'\in\mathcal{I}_{\mathrm{I}}} \in \mathbb{R}^{m_1\times m_1}, \tag{54}$$

$$\begin{aligned} J^{\mathrm{II}} &\equiv (\partial F_i/\partial \lambda_{i'}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*})_{i,i'\in\mathcal{I}_{\mathrm{II}}} \\ &= I \text{ (the identity matrix)} \in \mathbb{R}^{m_2\times m_2}, \end{aligned} \tag{55}$$

$$\begin{aligned} J^{\mathrm{III}} &\equiv (\partial F_i/\partial \lambda_{i'}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*})_{i,i'\in\mathcal{I}_{\mathrm{III}}} \\ &= \mathrm{diag}\left(e^{D_i^*-C},\, i\in\mathcal{I}_{\mathrm{III}}\right) \in \mathbb{R}^{m_3\times m_3}, \end{aligned} \tag{56}$$

O denotes the all-zero matrix of appropriate size,

$*$ denotes an appropriate matrix.

In (56), $\mathrm{diag}\left(e^{D_i^*-C},\, i\in\mathcal{I}_{\mathrm{III}}\right)$ denotes the diagonal matrix with diagonal components $e^{D_i^*-C}$, $i\in\mathcal{I}_{\mathrm{III}}$. Here, $e^{D_i^*-C} < 1$, $i\in\mathcal{I}_{\mathrm{III}}$ holds from type-III in (23).

Let $\{\theta_1,\ldots,\theta_m\} \equiv \{\theta_i \,|\, i\in\mathcal{I}\}$ be the set of eigenvalues of $J(\boldsymbol{\lambda}^*)$. By (53), the eigenvalues of $J(\boldsymbol{\lambda}^*)$ are the eigenvalues of J^{I}, J^{II}, and J^{III}, hence we can put

$\{\theta_i \,|\, i\in\mathcal{I}_{\mathrm{I}}\}$: the set of eigenvalues of J^{I},
$\{\theta_i \,|\, i\in\mathcal{I}_{\mathrm{II}}\}$: the set of eigenvalues of J^{II},
$\{\theta_i \,|\, i\in\mathcal{I}_{\mathrm{III}}\}$: the set of eigenvalues of J^{III}.

We will evaluate the eigenvalues of J^{I}, J^{II}, and J^{III} as follows.

C. Eigenvalues of J^{I}

First, we consider the eigenvalues of J^{I}. Let $J^{\mathrm{I}}_{i'i}$ be the (i',i) component of J^{I}, then by (52),

$$J^{\mathrm{I}}_{i'i} = \delta_{i'i} + \lambda_i^* D^*_{i',i},\ \ i',i\in\mathcal{I}_{\mathrm{I}}. \tag{57}$$

Let $I\in\mathbb{R}^{m_1\times m_1}$ denote the identity matrix and define $B\equiv I-J^{\mathrm{I}}$. Let $B_{i'i}$ be the (i',i) component of B. Then, from (57),

$$\begin{aligned} B_{i'i} &= -\lambda_i^* D^*_{i',i} &(58)\\ &= \lambda_i^* \sum_{j=1}^{n} \frac{P_j^{i'}P_j^i}{Q_j^*},\ i',i\in\mathcal{I}_{\mathrm{I}}. &(59) \end{aligned}$$

Let $\{\beta_i \,|\, i\in\mathcal{I}_{\mathrm{I}}\}$ be the set of eigenvalues of B, then we have $\theta_i = 1-\beta_i$, $i\in\mathcal{I}_{\mathrm{I}}$. To calculate the eigenvalues of B, we will define the following matrices. Similar calculations are performed in [21].

Let us define

$$\Phi_1 \equiv \begin{pmatrix} P^1 \\ \vdots \\ P^{m_1} \end{pmatrix} \in \mathbb{R}^{m_1\times n}, \tag{60}$$

$$\Gamma \equiv \left(-D^*_{i',i}\right) = \left(\sum_{j=1}^{n} \frac{P_j^{i'}P_j^i}{Q_j^*}\right)_{i',i\in\mathcal{I}_{\mathrm{I}}} \in \mathbb{R}^{m_1\times m_1}, \tag{61}$$

$$\Lambda \equiv \mathrm{diag}\left(\lambda_1^*,\ldots,\lambda_{m_1}^*\right) \in \mathbb{R}^{m_1\times m_1}. \tag{62}$$

Furthermore, define

$$\sqrt{\Lambda} \equiv \mathrm{diag}\left(\sqrt{\lambda_1^*},\ldots,\sqrt{\lambda_{m_1}^*}\right) \in \mathbb{R}^{m_1\times m_1}, \tag{63}$$

$$\Omega \equiv \mathrm{diag}\left((Q_1^*)^{-1},\ldots,(Q_n^*)^{-1}\right) \in \mathbb{R}^{n\times n}, \tag{64}$$

$$\sqrt{\Omega} \equiv \mathrm{diag}\left((Q_1^*)^{-1/2},\ldots,(Q_n^*)^{-1/2}\right) \in \mathbb{R}^{n\times n}. \tag{65}$$

Then, we have, by calculation,

$$\begin{aligned} \sqrt{\Lambda}B\sqrt{\Lambda}^{-1} &= \sqrt{\Lambda}\Gamma\sqrt{\Lambda} &(66)\\ &= \sqrt{\Lambda}\Phi_1\Omega\ {}^t\Phi_1\ {}^t\sqrt{\Lambda} &(67)\\ &= \sqrt{\Lambda}\Phi_1\sqrt{\Omega}\ {}^t\sqrt{\Omega}\ {}^t\Phi_1\ {}^t\sqrt{\Lambda} &(68)\\ &= \sqrt{\Lambda}\Phi_1\sqrt{\Omega}\ {}^t\left(\sqrt{\Lambda}\Phi_1\sqrt{\Omega}\right). &(69) \end{aligned}$$

From (16), $\sqrt{\Lambda}$ is a regular matrix and from the assumption (12), $\mathrm{rank}\,\Phi_1 = m_1$. Therefore, by $m_1\le m\le n$, we have $\mathrm{rank}\,\sqrt{\Lambda}\Phi_1\sqrt{\Omega} = m_1$, and thus from (69), $\sqrt{\Lambda}B\sqrt{\Lambda}^{-1}$ is symmetric and positive definite. In particular, all the eigenvalues $\beta_1,\ldots,\beta_{m_1}$ of B are positive. Without loss of generality, let $\beta_1\ge\ldots\ge\beta_{m_1}>0$. By (59), every component of B is nonnegative and by Lemma 3, every row sum of B is equal to 1. Hence, by the Perron-Frobenius theorem [10]

$$1=\beta_1\ge\beta_2\ge\ldots\ge\beta_{m_1}>0. \tag{70}$$

Because $\theta_i = 1-\beta_i$, $i\in\mathcal{I}_{\mathrm{I}}$, we have

$$0=\theta_1\le\theta_2\le\ldots\le\theta_{m_1}<1, \tag{71}$$

therefore,

Theorem 2: The eigenvalues of J^{I} satisfy

$$0\le\theta_i<1,\ i\in\mathcal{I}_{\mathrm{I}}. \tag{72}$$

D. Eigenvalues of J^{II}

Second, we consider the eigenvalues of J^{II}. From (53), (55), we have

Theorem 3: The eigenvalues of J^{II} satisfy

$$\theta_i = 1,\ i\in\mathcal{I}_{\mathrm{II}}. \tag{73}$$

E. Eigenvalues of J^{III}

Third, we consider the eigenvalues of J^{III}. From (53), (56), we have

Theorem 4: The eigenvalues of J^{III} are $\theta_i = e^{D_i^*-C}$, $D_i^* < C$, $i\in\mathcal{I}_{\mathrm{III}}$, hence

$$0<\theta_i<1,\ i\in\mathcal{I}_{\mathrm{III}}. \tag{74}$$

Remark 4: From the above consideration, we know that all the eigenvalues of the Jacobian matrix $J(\boldsymbol{\lambda}^*)$ are real.

Lemma 5: If the eigenvalues of J^{I} and J^{III} are distinct, i.e.,

$$\theta_i \ne \theta_{i'},\ i\in\mathcal{I}_{\mathrm{I}},\ i'\in\mathcal{I}_{\mathrm{III}}, \tag{75}$$

then $J(\boldsymbol{\lambda}^*)$ is diagonalizable. In particular, if $\mathcal{I}_{\mathrm{III}}=\emptyset$, then (75) holds, and thus $J(\boldsymbol{\lambda}^*)$ is diagonalizable.

Proof: See Appendix A. ■

IV. On the exponential convergence

We obtained in Theorems 2, 3, and 4 the evaluation for the eigenvalues of $J(\boldsymbol{\lambda}^*)$. Let $\theta_{\max} \equiv \max_{i\in\mathcal{I}} \theta_i$ be the maximum eigenvalue of $J(\boldsymbol{\lambda}^*)$, then by Theorems 2, 3, and 4, we have $0 \le \theta_{\max} < 1$ if $\mathcal{I}_{\mathrm{II}}$ is empty, and $\theta_{\max} = 1$ if $\mathcal{I}_{\mathrm{II}}$ is not empty. First, we show that the convergence is exponential if $\mathcal{I}_{\mathrm{II}}$ is empty.

Theorem 5: Assume $\mathcal{I}_{\mathrm{II}} = \emptyset$, then for any θ with $\theta_{\max} < \theta < 1$, there exist $\delta > 0$ and $K > 0$, such that for arbitrary initial distribution $\boldsymbol{\lambda}^0$ with $\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\| < \delta$, we have

$$\|\boldsymbol{\mu}^N\| = \|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\| < K \cdot (\theta)^N,\ N = 0, 1, \ldots, \tag{76}$$

i.e., the convergence is exponential.

Proof: See Appendix B. ■

V. On the convergence of order $O(1/N)$

We will consider the second-order recurrence formula obtained by truncating the Taylor expansion of $F(\boldsymbol{\lambda})$ up to the second-order term and analyze the convergence of order $O(1/N)$ of the sequence defined by the second-order recurrence formula.

A. The Hessian matrix $H_i(\boldsymbol{\lambda}^*)$

If $0 \le \theta_{\max} < 1$, then the convergence speed of $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$ is determined by the Jacobian matrix $J(\boldsymbol{\lambda}^*)$ because of Theorem 5. However, if $\theta_{\max} = 1$, the convergence speed is not determined only by $J(\boldsymbol{\lambda}^*)$; hence, we must investigate the Hessian matrix. In [2], [21], the Jacobian matrix is considered, but the Hessian matrix is neglected in the literature.

Now, we will calculate the Hessian matrix

$$H_i(\boldsymbol{\lambda}^*) = \left(\left.\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} \right)_{i',i''\in\mathcal{I}},\ i \in \mathcal{I} \tag{77}$$

of $F_i(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. We have

Theorem 6: The components of the Hessian matrix $H_i(\boldsymbol{\lambda}^*)$, $i \in \mathcal{I}$, are given as follows.

$$\begin{aligned} &\left.\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} \\ &= e^{D_i^*-C}\Big\{\delta_{ii'}D^*_{i,i''} + \delta_{ii''}D^*_{i,i'} + \lambda_i^*\left(D^*_{i,i'}D^*_{i,i''} + D^*_{i,i',i''}\right) \\ &\quad + \left(\delta_{ii'} + \lambda_i^* D^*_{i,i'}\right)\left(1 - e^{D^*_{i''}-C}\right) \\ &\quad + \left(\delta_{ii''} + \lambda_i^* D^*_{i,i''}\right)\left(1 - e^{D^*_{i'}-C}\right)\Big\} \\ &\quad + 2\lambda_i^*\left(1 - e^{D^*_{i'}-C}\right)\left(1 - e^{D^*_{i''}-C}\right) \\ &\quad - \lambda_i^*\left(e^{D^*_{i'}-C}D^*_{i',i''} + e^{D^*_{i''}-C}D^*_{i',i''} + E_{i',i''} - D^*_{i',i''}\right), \\ &i, i', i'' \in \mathcal{I}, \end{aligned}$$

where $D^*_{i,i',i''} \equiv \partial^2 D_i/\partial\lambda_{i'}\partial\lambda_{i''}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}$ and $E_{i',i''} \equiv \sum_{k=1}^{m_1} \lambda_k^* D^*_{k,i'}D^*_{k,i''}$.

In particular, if $i \in \mathcal{I}_{\mathrm{II}}$, then $\lambda_i^* = 0$ by (23), and thus

$$\begin{aligned} &\left.\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} \\ &= \delta_{ii''}\left(1 - e^{D^*_{i'}-C} + D^*_{i,i'}\right) + \delta_{ii'}\left(1 - e^{D^*_{i''}-C} + D^*_{i,i''}\right). \end{aligned} \tag{78}$$

Further, if $i \in \mathcal{I}_{\mathrm{II}}$ and $D^*_{i'} = C$ holds for all $i' \in \mathcal{I}$, then by (78), we have

$$\left.\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} = \delta_{ii'}D^*_{i,i''} + \delta_{ii''}D^*_{i,i'},\ i', i'' \in \mathcal{I}. \tag{79}$$

Proof: See Appendix C. ■

B. Analysis of the convergence of order $O(1/N)$

We consider a recurrence formula obtained by truncating the Taylor expansion (36) up to the second-order term and write the variables as $\bar{\boldsymbol{\mu}}^N = (\bar{\mu}_1^N, \ldots, \bar{\mu}_m^N)$. That is, we have

$$\bar{\boldsymbol{\mu}}^{N+1} = \bar{\boldsymbol{\mu}}^N J(\boldsymbol{\lambda}^*) + \frac{1}{2!}\bar{\boldsymbol{\mu}}^N H(\boldsymbol{\lambda}^*)\,{}^t\bar{\boldsymbol{\mu}}^N. \tag{80}$$

The recurrence formula (80) is called the *second-order recurrence formula* of the Taylor expansion (36). We investigate the convergence speed of $\bar{\boldsymbol{\mu}}^N \to \mathbf{0}$. The convergence speed of $\bar{\boldsymbol{\mu}}^N \to \mathbf{0}$ seems to be the same as that of the original $\boldsymbol{\mu}^N \to \mathbf{0}$, but the proof has not been obtained. Numerical comparison will be done in Chapter VI. In this chapter, we will prove that, if $\mathcal{I}_{\mathrm{II}} \ne \emptyset$, there exists an initial vector $\bar{\boldsymbol{\mu}}^0$ such that $\bar{\boldsymbol{\mu}}^N \to \mathbf{0}$ is the convergence of order $O(1/N)$. Furthermore, we will consider the condition that $\bar{\boldsymbol{\mu}}^N \to \mathbf{0}$ is the convergence of order $O(1/N)$ for arbitrary initial vector $\bar{\boldsymbol{\mu}}^0$.

The convergence of order $O(1/N)$ will be proved by the following three steps.

- Step 1: Represent $\bar{\mu}_i^N$ with types-I and -III indices by $\bar{\mu}_i^N$ with type-II indices.
- Step 2: Obtain the recurrence formula satisfied by $\bar{\mu}_i^N$ with type-II indices.
- Step 3: Prove that the convergence of $\bar{\mu}_i^N$ with type-II indices is of order $O(1/N)$ for some initial vector $\bar{\boldsymbol{\mu}}^0$.

C. Step 1

Here, we consider the types-I and -III indices together. Then, substitute $\mathcal{I}_{\mathrm{I}} \cup \mathcal{I}_{\mathrm{III}} = \{1, \ldots, m'\}$ and $\mathcal{I}_{\mathrm{II}} = \{m' + 1, \ldots, m\}$. We have $m_2 = m - m'$ and $|\mathcal{I}_{\mathrm{II}}| = m_2$. The purpose of step 1 is to represent $\bar{\boldsymbol{\mu}}^N_{\mathrm{I,III}} \equiv (\bar{\mu}_1^N, \ldots, \bar{\mu}_{m'}^N)$ by $\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \equiv (\bar{\mu}^N_{m'+1}, \ldots, \bar{\mu}_m^N)$.

In the Jacobian matrix, by changing the order of J^{II} and J^{III}, we have

$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} J^{\mathrm{I}} & O & O \\ * & J^{\mathrm{III}} & O \\ * & O & J^{\mathrm{II}} \end{pmatrix}. \tag{81}$$

Then, by defining

$$J' \equiv \begin{pmatrix} J^{\mathrm{I}} & O \\ * & J^{\mathrm{III}} \end{pmatrix}, \tag{82}$$

we have

$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} J' & O \\ * & J^{\mathrm{II}} \end{pmatrix}. \tag{83}$$

The eigenvalues of $J(\boldsymbol{\lambda}^*)$ are θ_i, $i = 1, \ldots, m$, and then the eigenvalues of J' are θ_i, $i = 1, \ldots, m'$ with $0 \le \theta_i < 1$, and those of J^{II} are θ_i, $i = m'+1, \ldots, m$ with $\theta_i = 1$.

Now, let $\boldsymbol{a}_i$ be a right eigenvector of $J(\boldsymbol{\lambda}^*)$ for θ_i and define

$$A \equiv (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m) \in \mathbb{R}^{m \times m}. \tag{84}$$

Under the assumption (75), by choosing the eigenvectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ appropriately, we can make A a regular matrix. In fact, because $J(\boldsymbol{\lambda}^*)$ is diagonalizable by Lemma 5, the direct sum of all the eigenspaces spans the whole $\mathbb{R}^m$ (See [19], p.161, Example 4).

For $i = m'+1, \ldots, m$, define

$$\boldsymbol{e}_i = (0, \ldots, 0, \overset{i\,\mathrm{th}}{\overset{\vee}{1}}, 0, \ldots, 0) \in \mathbb{R}^m, \quad i = m'+1, \ldots, m. \tag{85}$$

Then, because $\theta_i = 1$, we can take

$$\boldsymbol{a}_i = {}^t\boldsymbol{e}_i, \; i = m'+1, \ldots, m. \tag{86}$$

Therefore, we have

$$A = \left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{m'}, {}^t\boldsymbol{e}_{m'+1}, \ldots, {}^t\boldsymbol{e}_m\right) \tag{87}$$

$$= \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{m'1} & & & \\ \vdots & & \vdots & & O & \\ a_{1m'} & \cdots & a_{m'm'} & & & \\ \hline a_{1,m'+1} & \cdots & a_{m',m'+1} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{m'm} & 0 & \cdots & 1 \end{array}\right) \tag{88}$$

$$\equiv \begin{pmatrix} A_1 & O \\ A_2 & I \end{pmatrix}, \tag{89}$$

where

$$A_1 \equiv \begin{pmatrix} a_{11} & \cdots & a_{m'1} \\ \vdots & & \vdots \\ a_{1m'} & \cdots & a_{m'm'} \end{pmatrix}, \tag{90}$$

$$A_2 \equiv \begin{pmatrix} a_{1,m'+1} & \cdots & a_{m',m'+1} \\ \vdots & & \vdots \\ a_{1m} & \cdots & a_{m'm} \end{pmatrix}. \tag{91}$$

Because A is regular, A_1 is also regular by (89). $J(\boldsymbol{\lambda}^*)$ is diagonalized by A, i.e., $A^{-1}J(\boldsymbol{\lambda}^*)A = \Theta$, where

$$\Theta = \begin{pmatrix} \Theta_1 & O \\ O & I \end{pmatrix}, \; \Theta_1 = \begin{pmatrix} \theta_1 & & O \\ & \ddots & \\ O & & \theta_{m'} \end{pmatrix}, \quad 0 \le \theta_i < 1, \; i = 1, \ldots, m'. \tag{92}$$

Using only the first-order term of the Taylor expansion (36), we have

$$\boldsymbol{\mu}^{N+1}A = \boldsymbol{\mu}^N J(\boldsymbol{\lambda}^*)A \tag{93}$$

$$= \boldsymbol{\mu}^N A\Theta, \tag{94}$$

thus, by $\boldsymbol{\mu}^N = (\boldsymbol{\mu}^N_{\mathrm{I,III}}, \boldsymbol{\mu}^N_{\mathrm{II}})$, (94), and (89),

$$\boldsymbol{\mu}^{N+1}_{\mathrm{I,III}}A_1 + \boldsymbol{\mu}^{N+1}_{\mathrm{II}}A_2 = \left(\boldsymbol{\mu}^N_{\mathrm{I,III}}A_1 + \boldsymbol{\mu}^N_{\mathrm{II}}A_2\right)\Theta_1. \tag{95}$$

Hence, if the second- and higher-order terms are negligible, we have

$$\boldsymbol{\mu}^N_{\mathrm{I,III}}A_1 + \boldsymbol{\mu}^N_{\mathrm{II}}A_2 \to \boldsymbol{0} \text{ (exponentially)}, \; N \to \infty. \tag{96}$$

To show that the second- and higher-order terms are negligible, let us consider $\boldsymbol{\mu}^{N+1}\boldsymbol{a}_i$ as a function of $\boldsymbol{\mu}^N\boldsymbol{a}_1, \ldots, \boldsymbol{\mu}^N\boldsymbol{a}_{m'}$. If the Taylor expansion (36) satisfies the condition that

$$\boldsymbol{\mu}^{N+1}\boldsymbol{a}_i \text{ is divisible by } \boldsymbol{\mu}^N\boldsymbol{a}_i, \tag{97}$$

i.e., if $\boldsymbol{\mu}^N\boldsymbol{a}_i = 0$ implies that $\boldsymbol{\mu}^{N+1}\boldsymbol{a}_i = 0$, then, we have

$$\boldsymbol{\mu}^{N+1}\boldsymbol{a}_i = \theta_i\boldsymbol{\mu}^N\boldsymbol{a}_i\left(1 + o(|\boldsymbol{\mu}^N\boldsymbol{a}_i|)\right), \; N \to \infty, \quad i = 1, \ldots, m', \tag{98}$$

and hence (96) holds. However, in general, (97) is difficult to prove. We will show later in Examples 6 and 7 that (97) holds.

In what follows, we assume (96) and regard it as $\boldsymbol{\mu}^N_{\mathrm{I,III}}A_1 + \boldsymbol{\mu}^N_{\mathrm{II}}A_2 = \boldsymbol{0}$. Then, we replace $\boldsymbol{\mu}^N$ by $\bar{\boldsymbol{\mu}}^N$ to have $\bar{\boldsymbol{\mu}}^N_{\mathrm{I,III}}A_1 + \bar{\boldsymbol{\mu}}^N_{\mathrm{II}}A_2 = \boldsymbol{0}$ and hence

$$\bar{\boldsymbol{\mu}}^N_{\mathrm{I,III}} = -\bar{\boldsymbol{\mu}}^N_{\mathrm{II}}A_2A_1^{-1}. \tag{99}$$

The validity of (99) will be checked by numerical examples.

Step 1 is achieved by (99).

D. Step 2

The purpose of step 2 is to obtain a recurrence formula satisfied by $\bar{\mu}^N_i$, $i = m'+1, \ldots, m$.

The i-th component of (80) for $i = m'+1, \ldots, m$ is

$$\bar{\mu}^{N+1}_i = \bar{\mu}^N_i + \frac{1}{2!}\bar{\boldsymbol{\mu}}^N H_i(\boldsymbol{\lambda}^*)\,{}^t\bar{\boldsymbol{\mu}}^N, \quad i = m'+1, \ldots, m. \tag{100}$$

We will represent the second term on the right hand side of (100) by $\bar{\boldsymbol{\mu}}^N_{\mathrm{II}}$.

Let $H_{i,i'i''}$ be the (i', i'') component of the Hessian matrix $H_i(\boldsymbol{\lambda}^*)$. Then, by (78) of Theorem 6, we have

$$H_{i,i'i''} = \left.\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}$$
$$= \delta_{ii''}\left(1 - e^{D^*_{i'}-C} + D^*_{i,i'}\right) + \delta_{ii'}\left(1 - e^{D^*_{i''}-C} + D^*_{i,i''}\right), \tag{101}$$
$$i = m'+1, \ldots, m, \; i', i'' = 1, \ldots, m.$$

Here, for the simplicity of symbols, define

$$S_{ii'} \equiv 1 - e^{D^*_{i'}-C} + D^*_{i,i'}, \quad i = m'+1, \ldots, m, i' = 1, \ldots, m. \tag{102}$$

Then, we can write (101) as

$$H_{i,i'i''} = \delta_{ii''}S_{ii'} + \delta_{ii'}S_{ii''}. \tag{103}$$

Further, writing $A_1^{-1} \equiv (\zeta_{i'i''})$, $i', i'' = 1, \ldots, m'$, $T_{ii'} \equiv -\sum_{k,k'=1}^{m'} a_{i'k}\zeta_{kk'}S_{ik'}$, $i, i' = m'+1, \ldots, m$, and $r_{ii'} \equiv T_{ii'} + D^*_{i,i'}$, $i, i' = m'+1, \ldots, m$, we have the following theorem.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3095406

Theorem 7: $\{\bar{\mu}_i^N\}$, $i=m'+1,\ldots,m$, satisfies the recurrence formula

$$\bar{\mu}_i^{N+1}=\bar{\mu}_i^N+\bar{\mu}_i^N\sum_{i'=m'+1}^{m} r_{ii'}\bar{\mu}_{i'}^N, \qquad (104)$$
$$i=m'+1,\ldots,m.$$

Proof: See Appendix D. ∎

Step 2 is achieved by (104).

E. Step 3

The purpose of step 3 is to prove that $\bar{\mu}_i^N\to 0$, $i=1,\ldots,m$, is the convergence of order $O(1/N)$. We will define the canonical form of the recurrence formula (104). Writing $R\equiv(r_{ii'})$, $i,i'=m'+1,\ldots,m$, and $\mathbf{1}\equiv(1,\ldots,1)\in\mathbb{R}^{m_2}$, we consider the equation $R\,{}^t\boldsymbol{\sigma}=-{}^t\mathbf{1}$ for the variables $\boldsymbol{\sigma}=(\sigma_{m'+1},\ldots,\sigma_m)$. Assuming that R is regular, we have

$${}^t\boldsymbol{\sigma}=-R^{-1}\,{}^t\mathbf{1}, \qquad (105)$$

and then we assume $\boldsymbol{\sigma}>0$, i.e., $\sigma_i>0, i=m'+1,\ldots,m$. Further, by substituting

$$\nu_i^N\equiv\bar{\mu}_i^N/\sigma_i,\ i=m'+1,\ldots,m, \qquad (106)$$
$$p_{ii'}\equiv -r_{ii'}\sigma_{i'},\ i,i'=m'+1,\ldots,m, \qquad (107)$$

the recurrence formula (104) becomes

$$\nu_i^{N+1}=\nu_i^N-\nu_i^N\sum_{i'=1}^{m} p_{ii'}\nu_{i'}^N, \qquad (108)$$
$$i=m'+1,\ldots,m,$$

where $\boldsymbol{p}_i\equiv(p_{i,m'+1},\ldots,p_{i,m})$ is a probability vector.

The recurrence formula (108) is called the *canonical form* of (104).

For the analysis of (108), we prepare the following lemma.

Lemma 6: Let us define a positive sequence $\{\nu^N\}_{N=0,1\ldots}$ by the recurrence formula

$$\nu^{N+1}=\nu^N-\left(\nu^N\right)^2,\ N=0,1,\ldots, \qquad (109)$$
$$0<\nu^0\le 1/2. \qquad (110)$$

Then, we have

$$\lim_{N\to\infty} N\nu^N=1. \qquad (111)$$

Proof: Since the function $g(\nu)\equiv\nu-\nu^2$ satisfies $0<g(\nu)<\nu$ for $0<\nu\le 1/2$, we see $0<\nu^{N+1}<\nu^N$, $N=0,1,\ldots$ by mathematical induction. Thus, $\nu^\infty\equiv\lim_{N\to\infty}\nu^N\ge 0$ exists, and by (109), $\nu^\infty=\nu^\infty-(\nu^\infty)^2$ holds. Then, we have $\nu^\infty=0$.

Next, by (109), we have

$$\frac{1}{N}\left(\frac{1}{\nu^N}-\frac{1}{\nu^0}\right)=\frac{1}{N}\sum_{l=0}^{N-1}\left(\frac{1}{\nu^{l+1}}-\frac{1}{\nu^l}\right) \qquad (112)$$
$$=\frac{1}{N}\sum_{l=0}^{N-1}\frac{1}{1-\nu^l}. \qquad (113)$$

Applying the proposition that "the arithmetic mean of the first n terms of a convergent sequence converges to the same limit as the original sequence" ([1], p.37, Exercise 2) to (113), we have

$$\lim_{N\to\infty}\frac{1}{N\nu^N}=\lim_{N\to\infty}\frac{1}{1-\nu^N} \qquad (114)$$
$$=1, \qquad (115)$$

which proves the lemma. ∎

Lemma 7: For the sequence $\{\nu_i^N\}$, $i=m'+1,\ldots,m$, defined by (108) with the initial values $\nu_i^0=1/2$, we have

$$\lim_{N\to\infty} N\nu_i^N=1,\ i=m'+1,\ldots,m. \qquad (116)$$

Proof: By mathematical induction, we see that $\nu_{m'+1}^N=\ldots=\nu_m^N$ holds for $N=0,1,\ldots$, and thus (108) becomes $\nu_i^{N+1}=\nu_i^N-\left(\nu_i^N\right)^2$, $i=m'+1,\ldots,m$. Therefore, (116) holds by Lemma 6. ∎

Theorem 8: For (104) with the initial values $\bar{\mu}_i^0=\sigma_i/2$, $i=m'+1,\ldots,m$, we have

$$\lim_{N\to\infty} N\bar{\mu}_i^N=\sigma_i,\ i=m'+1,\ldots,m. \qquad (117)$$

Further, under the assumption (99), we have

$$\lim_{N\to\infty} N\bar{\mu}_i^N=-\left(\boldsymbol{\sigma}A_2A_1^{-1}\right)_i,\ i=1,\ldots,m', \qquad (118)$$

where $\boldsymbol{\sigma}$, A_1, A_2 were defined by (105), (91).

Proof: From $\nu_i^N=\bar{\mu}_i^N/\sigma_i$, $i=m'+1,\ldots,m$, and Lemma 7, we obtain (117). Further, by (99), we obtain (118). ∎

Step 3 is achieved by Theorem 8. Summarizing the above, we have the following theorem.

Theorem 9: If type-II indices exist, then there exists an initial vector $\bar{\boldsymbol{\mu}}^0$ such that $\bar{\boldsymbol{\mu}}^N\to\mathbf{0}$ is the convergence of order $O(1/N)$.

Remark 5: In Theorem 9, we must choose a specific initial vector $\bar{\boldsymbol{\mu}}^0$, but we want to prove it for an arbitrary initial vector. If the existence of $\lim_{N\to\infty}N\mu_i^N$, $i=1,\ldots,m$, is proved for arbitrary $\bar{\boldsymbol{\mu}}^0$, then we have (117) and (118). However, the analysis for the convergence speed of the recurrence formula (104) for an arbitrary initial vector is very difficult, so it has not been achieved. We will give a proof under the assumption that a conjecture holds.

F. On the initial vector

In this section, we will investigate the convergence of (108) for an arbitrary initial vector.

Now, for the sake of simplicity, we will change the indices and symbols of the canonical form (108). Noting that $m-m'=m_2$, we change the indices from $m'+1,\ldots,m$ to $1,\ldots,m_2$, and define

$$\xi_i^N\equiv\nu_{m'+i},\ i=1,\ldots,m_2, \qquad (119)$$
$$q_{ii'}\equiv p_{m'+i,m'+i'},\ i,i'=1,\ldots,m_2, \qquad (120)$$

which are just shifting the indices. By the above change, the canonical form (108) becomes

$$\xi_i^{N+1}=\xi_i^N-\xi_i^N\sum_{i'=1}^{m_2} q_{ii'}\xi_{i'}^N,\ i=1,\ldots,m_2, \qquad (121)$$

where $\boldsymbol{q}_i\equiv(q_{i1},\ldots,q_{i,m_2})\in\mathbb{R}^{m_2}$ is a probability vector.

G. Diagonally dominant condition

For the probability vectors $\boldsymbol{q}_i,\ i=1,\cdots,m_2$, we assume

$$q_{ii} > \sum_{i'=1, i'\neq i}^{m_2} q_{ii'},\ i=1,\dots,m_2. \tag{122}$$

This assumption means that the matrix made by arranging the row vectors $\boldsymbol{q}_1,\dots,\boldsymbol{q}_{m_2}$ vertically is diagonally dominant. We call (122) the *diagonally dominant condition*. By numerical calculation, we confirmed that (122) holds for all our examples.

H. Conjecture

Our goal is to prove $\lim_{N\to\infty} N\xi_i^N = 1,\ i=1,\dots,m_2$, for any initial vector $\boldsymbol{\xi}^0$. We found that we can prove this if the following conjecture holds; however, the proof of this conjecture has not yet been obtained.

Conjecture: By reordering the indices if necessary, there exists an N_0 such that the following inequalities hold.

$$\xi_1^N \geq \xi_2^N \geq \ldots \geq \xi_{m_2}^N,\ N \geq N_0. \tag{123}$$

Many numerical examples we observed seem to support this conjecture.

We have

Theorem 10: We assume that the sequence defined by the recurrence formula (121) satisfies (122) and (123). Then, for arbitrary initial values with $0 < \xi_i^0 \leq 1/2,\ i=1,\dots,m_2$, we have

$$\lim_{N\to\infty} N\xi_i^N = 1,\ \ i=1,\dots,m_2. \tag{124}$$

Proof: See Appendix E. ■

I. Special cases where the conjecture holds

Now, we consider some special cases where the conjecture (123) holds.

If $m_2=1$, then the variable is only ξ_1^N, hence we do not need to consider the inequality condition.

If $m_2=2$, the canonical form (121) becomes

$$\xi_1^{N+1} = \xi_1^N - q_{11}\left(\xi_1^N\right)^2 - q_{12}\xi_1^N\xi_2^N, \tag{125}$$

$$\xi_2^{N+1} = \xi_2^N - q_{21}\xi_2^N\xi_1^N - q_{22}\left(\xi_2^N\right)^2, \tag{126}$$

$$q_{11}>0,\ q_{12}>0,\ q_{11}+q_{12}=1, \tag{127}$$

$$q_{21}>0,\ q_{22}>0,\ q_{21}+q_{22}=1. \tag{128}$$

By calculation, we have

$$\xi_1^{N+1} - \xi_2^{N+1} = \left(\xi_1^N - \xi_2^N\right)\left(1 - q_{11}\xi_1^N - q_{22}\xi_2^N\right). \tag{129}$$

By Lemma 12 in Appendix E, we have $1-q_{11}\xi_1^N - q_{22}\xi_2^N > 0$ for any $N=0,1,\ldots$. Hence, if $\xi_1^0 \geq \xi_2^0$, then $\xi_1^N \geq \xi_2^N$, $N=0,1,\ldots$. Thus, the conjecture (123) holds with $N_0=0$. Note that, when $m_2=2$, the diagonally dominant condition (122) is not necessary, because it is not used in the above proof.

In the case of $m_2 \geq 3$, this problem is very difficult. We have a sufficient condition for (123) in the following lemma.

Lemma 8: For i', consider $q_{ii'},\ i=1,\dots,m_2$. Assume that $q_{ii'}$ are equal for all i except i', i.e., $q_{1i'}=\ldots=q_{i'-1,i'}=q_{i'+1,i'}=\ldots=q_{m_2,i'}$. Then, the conjecture (123) holds.

Proof: By calculation, we have

$$\begin{aligned}&\xi_1^{N+1} - \xi_2^{N+1}\\ &= \left(\xi_1^N - \xi_2^N\right)\left(1 - q_{11}\xi_1^N - q_{22}\xi_2^N - \sum_{i'=3}^{m_2} q_{1i'}\xi_{i'}^N\right).\end{aligned} \tag{130}$$

The second factor on the right hand side of (130) is positive for all $N=0,1,\dots$ by Lemma 12. Hence, if $\xi_1^0 \geq \xi_2^0$, then $\xi_1^N \geq \xi_2^N$ holds for $N=0,1,\ldots$. Similarly, by considering any pair ξ_i^N and $\xi_{i'}^N$, we see that (123) holds. ■

Summarizing the above, we have

Theorem 11: Under the assumptions of Lemma 8 and the diagonally dominant condition (122), the convergence of the recurrence formula (121) is of order $O(1/N)$ for arbitrary initial vector $\boldsymbol{\xi}^0=(\xi_1^0,\dots,\xi_{m_2}^0)$ with $0<\xi_i^0\leq 1/2,\ i=1,\dots,m_2$, and

$$\lim_{N\to\infty} N\xi_i^N = 1,\ i=1,\dots,m_2. \tag{131}$$

VI. Numerical Evaluation

Based on the analysis in the previous sections, we will evaluate numerically the convergence speed of the Arimoto-Blahut algorithm for several channel matrices.

In Examples 4 and 5 below, we will investigate the exponential convergence, where all the indices are of type-I, in other words, the capacity-achieving $\boldsymbol{\lambda}^*$ is in $\Delta(\mathcal{X})^\circ$ (the interior of $\Delta(\mathcal{X})$). In Example 5, we will discuss how the convergence speed depends on the choice of the initial distribution $\boldsymbol{\lambda}^0$. Next, in Examples 6 and 7, we will consider the convergence of order $O(1/N)$. In these examples, there exist type-II indices. It will be confirmed that the convergence speed is accurately approximated by the values obtained in Theorem 8. In Example 8, we will investigate the exponential convergence, where all the indices are of type-I or -III, i.e., there exist no type-II indices.

Here, in the case of exponential convergence, we will evaluate the values of the function

$$L(N) \equiv -\frac{1}{N}\log\|\boldsymbol{\mu}^N\|. \tag{132}$$

Based on the results of Theorem 5, i.e., $\|\boldsymbol{\mu}^N\| = \|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\| < K\cdot(\theta)^N$, $\theta \doteqdot \theta_{\max}$, we will compare $L(N)$ for large N with $-\log\theta_{\max}$ (or other values).

On the other hand, in the case of the convergence of order $O(1/N)$, we will evaluate

$$N\boldsymbol{\mu}^N = (N\mu_1^N,\dots,N\mu_m^N). \tag{133}$$

We will compare $N\boldsymbol{\mu}^N$ for large N with the values obtained in Theorem 8.

A. Exponential convergence where all the indices are of type-I

Example 4: Consider the channel matrix $\Phi^{(1)}$ of (28), i.e.,

$$\Phi^{(1)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}, \tag{134}$$

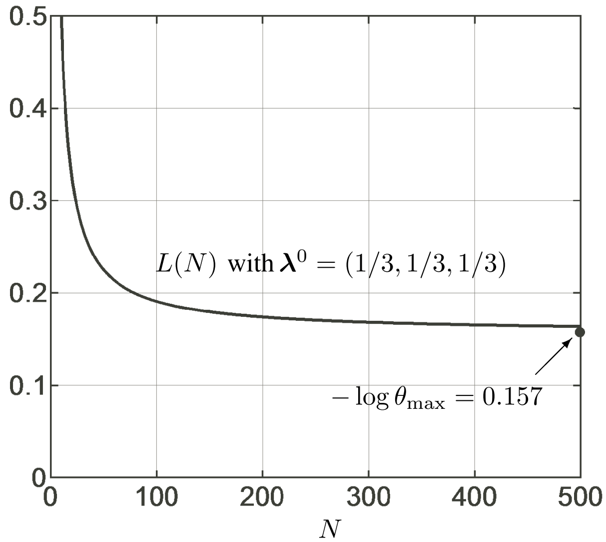

Fig. 5. Convergence of $L(N)$ in Example 4 with initial distribution $\boldsymbol{\lambda}^0 = (1/3, 1/3, 1/3)$.

and an initial distribution $\boldsymbol{\lambda}^0 = (1/3, 1/3, 1/3)$. We have

$$C = 0.323 \,[\text{nat/symbol}], \tag{135}$$

$$\boldsymbol{\lambda}^* = (0.431, 0.431, 0.138), \tag{136}$$

$$Q^* = (0.422, 0.422, 0.156), \tag{137}$$

$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} 0.308 & -0.191 & -0.117 \\ -0.191 & 0.308 & -0.117 \\ -0.369 & -0.369 & 0.738 \end{pmatrix}. \tag{138}$$

By (136), we see that $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* > 0$, thus all the indices are of type-I.

The eigenvalues of $J(\boldsymbol{\lambda}^*)$ are $(\theta_1, \theta_2, \theta_3) = (0.000, 0.500, 0.855)$. Then, $\theta_{\max} = \theta_3 = 0.855$. We have, for $N = 500$,

$$L(500) = 0.161 \doteqdot -\log\theta_{\max} = 0.157. \tag{139}$$

We can see from Fig. 5 that $L(N)$ for large N is accurately approximated by the value $-\log\theta_{\max}$.

Example 5: Let us consider another channel matrix. Define

$$\Phi^{(4)} \equiv \begin{pmatrix} 0.793 & 0.196 & 0.011 \\ 0.196 & 0.793 & 0.011 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}. \tag{140}$$

We have

$$C = 0.294 \,[\text{nat/symbol}], \tag{141}$$

$$\boldsymbol{\lambda}^* = (0.352, 0.352, 0.296), \tag{142}$$

$$Q^* = (0.422, 0.422, 0.156), \tag{143}$$

$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} 0.443 & -0.260 & -0.183 \\ -0.260 & 0.443 & -0.183 \\ -0.218 & -0.218 & 0.436 \end{pmatrix}. \tag{144}$$

By (142), we see that all the indices are of type-I.

The eigenvalues of $J(\boldsymbol{\lambda}^*)$ are $(\theta_1, \theta_2, \theta_3) = (0.000, 0.618, 0.702)$. Then, $\theta_{\max} = \theta_3 = 0.702$. Write the second largest eigenvalue as θ_{sec}, then $\theta_{\text{sec}} = \theta_2 = 0.618$.

We show in Fig. 6 the graph of $L(N)$ with initial distribution $\boldsymbol{\lambda}_1^0 \equiv (1/3, 1/3, 1/3)$ by the solid line, and the graph with initial distribution $\boldsymbol{\lambda}_2^0 \equiv (1/2, 1/3, 1/6)$ by the dotted line.

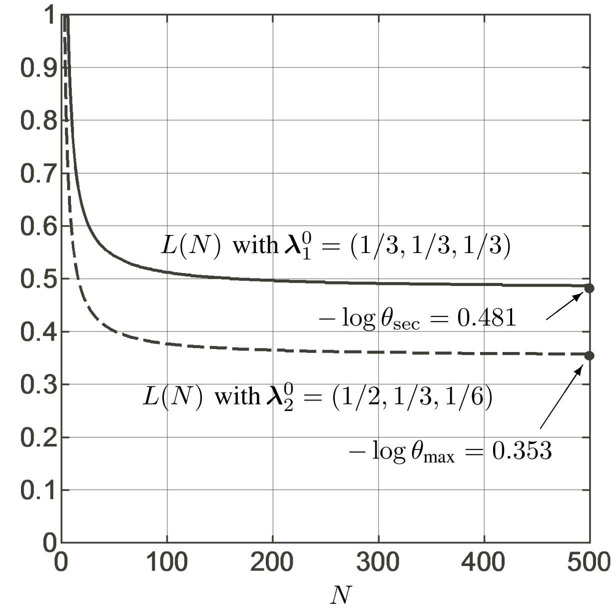

Fig. 6. Convergence of $L(N)$ in Example 5 with initial distribution $\boldsymbol{\lambda}_1^0 = (1/3, 1/3, 1/3)$ and $\boldsymbol{\lambda}_2^0 = (1/2, 1/3, 1/6)$.

The larger $L(N)$ is, the faster the convergence, hence the convergence with $\boldsymbol{\lambda}_1^0$ is faster than with $\boldsymbol{\lambda}_2^0$. The convergence speed varies depending on the choice of the initial distribution. What kind of initial distribution yields faster convergence? We will investigate this question below.

Let us define

$$\boldsymbol{\mu}_1^0 \equiv \boldsymbol{\lambda}_1^0 - \boldsymbol{\lambda}^* = (-0.019, -0.019, 0.038), \tag{145}$$

$$\boldsymbol{\mu}_2^0 \equiv \boldsymbol{\lambda}_2^0 - \boldsymbol{\lambda}^* = (0.148, -0.019, -0.129). \tag{146}$$

We will execute the following calculation by assuming $\boldsymbol{\mu}^{N+1} = \boldsymbol{\mu}^N J(\boldsymbol{\lambda}^*)$, $N = 0, 1, \ldots$ holds exactly.

Here, we will investigate for general m, n. We assume (75). Let $\boldsymbol{b}_{\max}$ be the left eigenvector of $J(\boldsymbol{\lambda}^*)$ for $\theta_{\max}$, and let $\boldsymbol{b}_{\max}^{\perp}$ be the orthogonal complement of $\boldsymbol{b}_{\max}$, i.e., $\boldsymbol{b}_{\max}^{\perp} \equiv \{\boldsymbol{\mu} \in \mathbb{R}^m \,|\, \boldsymbol{\mu}\,{}^t\boldsymbol{b}_{\max} = 0\}$.

Lemma 9: If

$$\boldsymbol{\mu}^N \in \boldsymbol{b}_{\max}^{\perp},\ N = 0, 1, \ldots, \tag{147}$$

then for any θ with $\theta_{\text{sec}} < \theta < 1$, we have $\|\boldsymbol{\mu}^N\| < K \cdot (\theta)^N$, $K > 0$, $N = 0, 1, \ldots$.

Proof: See Appendix F. ■

Because $\theta_{\text{sec}} < \theta_{\max}$, if (147) holds, then the convergence speed is faster than $\theta_{\max}$ by Lemma 9. The next lemma gives a necessary and sufficient condition that guarantees (147).

Lemma 10: $\boldsymbol{\mu} J(\boldsymbol{\lambda}^*) \in \boldsymbol{b}_{\max}^{\perp}$ holds for any $\boldsymbol{\mu} \in \boldsymbol{b}_{\max}^{\perp}$ if and only if ${}^t\boldsymbol{b}_{\max}$ is a right eigenvector for $\theta_{\max}$.

Proof: See Appendix G. ■

If ${}^t\boldsymbol{b}_{\max}$ is a right eigenvector, then by Lemma 10, any $\boldsymbol{\mu}^0 \in \boldsymbol{b}_{\max}^{\perp}$ yields (147), hence the convergence becomes faster.

Now, we will evaluate the convergence speed for the initial vectors (145) and (146). For $J(\boldsymbol{\lambda}^*)$ of (144), $\theta_{\max} = 0.702$ and $\theta_{\text{sec}} = 0.618$. The left eigenvector for $\theta_{\max}$ is $\boldsymbol{b}_{\max} = (-0.500, 0.500, 0.000)$. We can confirm that ${}^t\boldsymbol{b}_{\max}$ is a right eigenvector for $\theta_{\max}$ and $\boldsymbol{\mu}_1^0\,{}^t\boldsymbol{b}_{\max} = 0$, thus by Lemmas 9 and 10, we have $\lim_{N\to\infty} L(N) \doteqdot -\log\theta_{\mathrm{sec}}$. Then, by the solid line in Fig. 6, we have, for $N = 500$,

$$L(500) = 0.489 \doteqdot -\log\theta_{\mathrm{sec}} = 0.481. \tag{148}$$

On the other hand, we have $\boldsymbol{\mu}_2^{0\,t}\boldsymbol{b}_{\max} \neq 0$, thus by Lemma 10, we have $\lim_{N\to\infty} L(N) \doteqdot -\log\theta_{\max}$. Then, by the dotted line, we have, for $N = 500$,

$$L(500) = 0.360 \doteqdot -\log\theta_{\max} = 0.353. \tag{149}$$

Checking Example 4 in this way, we can see that $\boldsymbol{b}_{\max} = (-0.431, -0.431, 0.862)$ is a left eigenvector for $\theta_{\max} = 0.855$, but ${}^t\boldsymbol{b}_{\max}$ is not a right eigenvector. Thus, by Lemma 10, we have $\lim_{N\to\infty} L(N) \doteqdot -\log\theta_{\max}$ and (139).

B. Convergence of order $O(1/N)$

In the case of the convergence of order $O(1/N)$, there exist type-II indices.

Example 6: Consider the channel matrix $\Phi^{(2)}$ of (29), i.e.,

$$\Phi^{(2)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{pmatrix}, \tag{150}$$

and an initial distribution $\boldsymbol{\lambda}^0 = (1/3, 1/3, 1/3)$. We have

$$C = 0.310\,[\text{nat/symbol}], \tag{151}$$
$$\boldsymbol{\lambda}^* = (0.500, 0.500, 0.000), \tag{152}$$
$$Q^* = (0.450, 0.450, 0.100), \tag{153}$$
$$D^*_{1,1} = -1.544,\ D^*_{1,2} = -0.456,\ D^*_{1,3} = -1, \tag{154}$$
$$D^*_{2,2} = -1.544,\ D^*_{2,3} = -1,\ D^*_{3,3} = -2, \tag{155}$$
$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} 1+\lambda_1^* D^*_{1,1} & \lambda_2^* D^*_{1,2} & 0 \\ \lambda_1^* D^*_{2,1} & 1+\lambda_2^* D^*_{2,2} & 0 \\ \lambda_1^* D^*_{3,1} & \lambda_2^* D^*_{3,2} & 1 \end{pmatrix} \tag{156}$$
$$= \begin{pmatrix} 0.228 & -0.228 & 0.000 \\ -0.228 & 0.228 & 0.000 \\ -0.500 & -0.500 & 1.000 \end{pmatrix}. \tag{157}$$

By (152) and (153), we see that $\lambda_3^* = 0$ and $C = D(P^1\|Q^*) = D(P^2\|Q^*) = D(P^3\|Q^*)$, hence, $i = 3$ is a type-II index.

The eigenvalues of $J(\boldsymbol{\lambda}^*)$ are $(\theta_1, \theta_2, \theta_3) = (0.000, 0.456, 1.000)$.

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \tag{158}$$
$$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},\ A_2 = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{159}$$

By (159), the eigenvectors $\boldsymbol{a}_1$ for $\theta_1 = 0$ and $\boldsymbol{a}_2$ for $\theta_2 = 0.456$ are

$$\boldsymbol{a}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},\ \boldsymbol{a}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}. \tag{160}$$

We will prove (97). First, for $\boldsymbol{a}_1$, $\boldsymbol{\mu}^{N+1}\boldsymbol{a}_1 = \boldsymbol{\mu}^N\boldsymbol{a}_1 = \boldsymbol{0}$, thus (97) is trivial. Next, for $\boldsymbol{a}_2$, we will prove that $\boldsymbol{\mu}^{N+1}\boldsymbol{a}_2 = \mu_1^{N+1} - \mu_2^{N+1}$ is divisible by $\boldsymbol{\mu}^N\boldsymbol{a}_2 = \mu_1^N - \mu_2^N$. In fact, by $\lambda_1^* = \lambda_2^*$ we have

$$\mu_1^{N+1} - \mu_2^{N+1} = \lambda_1^{N+1} - \lambda_2^{N+1} \tag{161}$$
$$= F_1(\boldsymbol{\lambda}^N) - F_2(\boldsymbol{\lambda}^N) \tag{162}$$
$$= \frac{\lambda_1^N \exp D(P^1\|\boldsymbol{\lambda}^N\Phi) - \lambda_2^N \exp D(P^2\|\boldsymbol{\lambda}^N\Phi)}{\sum_{k=1}^{3} \lambda_k^N \exp D(P^k\|\boldsymbol{\lambda}^N\Phi)}. \tag{163}$$

If we substitute $\lambda_1^N = \lambda_2^N$ in (163), we have by calculation, $\mu_1^{N+1} - \mu_2^{N+1} = 0$. Because $\lambda_1^N = \lambda_2^N$ is equivalent to $\mu_1^N = \mu_2^N$, we see that $\mu_1^N - \mu_2^N = 0$ implies $\mu_1^{N+1} - \mu_2^{N+1} = 0$, thus (97) holds, and then we can consider (99).

For $\bar{\boldsymbol{\mu}}^N_{\mathrm{I,\,III}} = (\bar{\mu}_1^N, \bar{\mu}_2^N)$, $\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} = (\bar{\mu}_3^N)$, we have $\bar{\boldsymbol{\mu}}^N_{\mathrm{I,\,III}} = -\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} A_2 A_1^{-1}$ by (99), hence

$$\bar{\mu}_1^N = \bar{\mu}_2^N = -(1/2)\bar{\mu}_3^N. \tag{164}$$

Further, the Hessian matrix $H_3(\boldsymbol{\lambda}^*)$ is

$$H_3(\boldsymbol{\lambda}^*) = \begin{pmatrix} 0 & 0 & D^*_{3,1} \\ 0 & 0 & D^*_{3,2} \\ D^*_{3,1} & D^*_{3,2} & 2D^*_{3,3} \end{pmatrix}. \tag{165}$$
$$= \begin{pmatrix} 0.000 & 0.000 & -1.000 \\ 0.000 & 0.000 & -1.000 \\ -1.000 & -1.000 & -4.000 \end{pmatrix}, \tag{166}$$

then, we have by (164),

$$\frac{1}{2}\bar{\boldsymbol{\mu}}^N H_3\, {}^t\bar{\boldsymbol{\mu}}^N = -\left(\bar{\mu}_3^N\right)^2, \tag{167}$$

and the second-order recurrence formula

$$\bar{\mu}_3^{N+1} = \bar{\mu}_3^N - \left(\bar{\mu}_3^N\right)^2. \tag{168}$$

By Lemma 6 and (164), we have $\lim_{N\to\infty} N\bar{\mu}_3^N = 1$, $\lim_{N\to\infty} N\bar{\mu}_1^N = \lim_{N\to\infty} N\bar{\mu}_2^N = -1/2$.

By the numerical simulation, $N\boldsymbol{\mu}^N$ for $N = 500$ is

$$N\boldsymbol{\mu}^N = (-0.510, -0.510, 1.019) \tag{169}$$
$$\doteqdot \lim_{N\to\infty} N\bar{\boldsymbol{\mu}}^N = (-1/2, -1/2, 1). \tag{170}$$

See Fig. 7. We can confirm that $N\boldsymbol{\mu}^N$ for large N is near the values obtained in Theorem 8.

Example 7: Consider a channel matrix

$$\Phi^{(5)} \equiv \begin{pmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 & 0.1 & 0.1 \\ s & s & t & 0.1 & 0.1 \\ s & s & 0.1 & t & 0.1 \\ s & s & 0.1 & 0.1 & t \end{pmatrix}, \tag{171}$$

where $s \equiv 0.238$, $t \equiv 0.324$, $(2s + t + 0.2 = 1)$, (172)

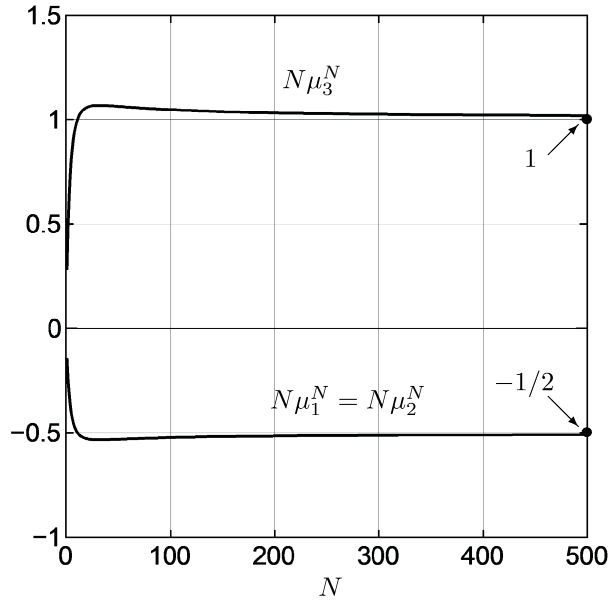

Fig. 7. Convergence of $N\mu_i^N$ in Example 6.

and an initial distribution $\boldsymbol{\lambda}^0 = (1/5, 1/5, 1/5, 1/5, 1/5)$. For this $\Phi^{(5)}$, we have

$$C = 0.198\,[\text{nat/symbol}], \tag{173}$$

$$\boldsymbol{\lambda}^* = (0.5,\, 0.5,\, 0,\, 0,\, 0), \tag{174}$$

$$Q^* = (0.35,\, 0.35,\, 0.1,\, 0.1,\, 0.1), \tag{175}$$

$$D^*_{1,1} = -19/14,\, D^*_{1,2} = -9/14,\, D^*_{1,3} = -1,\, D^*_{1,4} = -1,\, D^*_{1,5} = -1, \tag{176}$$

$$D^*_{2,2} = -19/14,\, D^*_{2,3} = -1,\, D^*_{2,4} = -1,\, D^*_{2,5} = -1, \tag{177}$$

$$D^*_{3,3} = -1.576 \equiv -\alpha,\, D^*_{3,4} = -1.072 \equiv -\beta,\, D^*_{3,5} = -\beta, \tag{178}$$

$$D^*_{4,4} = -\alpha,\, D^*_{4,5} = -\beta,\, D^*_{5,5} = -\alpha, \tag{179}$$

$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} 9/28 & -9/28 & 0 & 0 & 0 \\ -9/28 & 9/28 & 0 & 0 & 0 \\ -1/2 & -1/2 & 1 & 0 & 0 \\ -1/2 & -1/2 & 0 & 1 & 0 \\ -1/2 & -1/2 & 0 & 0 & 1 \end{pmatrix}. \tag{180}$$

By (174) and (175), we see that $\lambda_3^* = \lambda_4^* = \lambda_5^* = 0$ and $C = D(P^1 \| Q^*) = \ldots = D(P^5 \| Q^*)$, thus $i = 3, 4, 5$ are type-II indices.

The eigenvalues of $J(\boldsymbol{\lambda}^*)$ are $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (0, 9/14, 1, 1, 1)$ and

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{181}$$

$$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},\ A_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}. \tag{182}$$

We can prove (97) in a similar way as in Example 6.

For $\bar{\boldsymbol{\mu}}^N_{\mathrm{I,\,III}} = (\bar{\mu}_1^N,\, \bar{\mu}_2^N)$, $\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} = (\bar{\mu}_3^N,\, \bar{\mu}_4^N,\, \bar{\mu}_5^N)$, we have $\bar{\boldsymbol{\mu}}^N_{\mathrm{I,\,III}} = -\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} A_2 A_1^{-1}$ by (99), hence

$$\bar{\mu}_1^N = \bar{\mu}_2^N = -(\bar{\mu}_3^N + \bar{\mu}_4^N + \bar{\mu}_5^N)/2. \tag{183}$$

Further, the Hessian matrix $H_3(\boldsymbol{\lambda}^*)$ is

$$H_3(\boldsymbol{\lambda}^*) = \begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ -1 & -1 & -2\alpha & -\beta & -\beta \\ 0 & 0 & -\beta & 0 & 0 \\ 0 & 0 & -\beta & 0 & 0 \end{pmatrix}, \tag{184}$$

where α and β were defined in (178). Then, we have by (183)

$$\frac{1}{2}\bar{\boldsymbol{\mu}}^N H_3(\boldsymbol{\lambda}^*)\, {}^t\bar{\boldsymbol{\mu}}^N = -(\alpha-1)\left(\bar{\mu}_3^N\right)^2 - (\beta-1)\bar{\mu}_3^N\bar{\mu}_4^N - (\beta-1)\bar{\mu}_3^N\bar{\mu}_5^N. \tag{185}$$

Similarly,

$$\frac{1}{2}\bar{\boldsymbol{\mu}}^N H_4(\boldsymbol{\lambda}^*)\, {}^t\bar{\boldsymbol{\mu}}^N = -(\beta-1)\bar{\mu}_3^N\bar{\mu}_4^N - (\alpha-1)\left(\bar{\mu}_4^N\right)^2 - (\beta-1)\bar{\mu}_4^N\bar{\mu}_5^N, \tag{186}$$

$$\frac{1}{2}\bar{\boldsymbol{\mu}}^N H_5(\boldsymbol{\lambda}^*)\, {}^t\bar{\boldsymbol{\mu}}^N = -(\beta-1)\bar{\mu}_3^N\bar{\mu}_5^N - (\beta-1)\bar{\mu}_4^N\bar{\mu}_5^N - (\alpha-1)\left(\bar{\mu}_5^N\right)^2. \tag{187}$$

Therefore, by substituting $\alpha' \equiv \alpha - 1$, $\beta' \equiv \beta - 1$, we have

$$\bar{\mu}_3^{N+1} = \bar{\mu}_3^N - \alpha'\left(\bar{\mu}_3^N\right)^2 - \beta'\bar{\mu}_3^N\bar{\mu}_4^N - \beta'\bar{\mu}_3^N\bar{\mu}_5^N, \tag{188}$$

$$\bar{\mu}_4^{N+1} = \bar{\mu}_4^N - \beta'\bar{\mu}_3^N\bar{\mu}_4^N - \alpha'\left(\bar{\mu}_4^N\right)^2 - \beta'\bar{\mu}_4^N\bar{\mu}_5^N, \tag{189}$$

$$\bar{\mu}_5^{N+1} = \bar{\mu}_5^N - \beta'\bar{\mu}_3^N\bar{\mu}_5^N - \beta'\bar{\mu}_4^N\bar{\mu}_5^N - \alpha'\left(\bar{\mu}_5^N\right)^2. \tag{190}$$

From (105), we have $\boldsymbol{\sigma} = (\sigma_3, \sigma_4, \sigma_5) = (1.389, 1.389, 1.389)$, so the canonical form for (188), (189), (190) is

$$\nu_3^{N+1} = \nu_3^N - 0.8\left(\nu_3^N\right)^2 - 0.1\nu_3^N\mu_4^N - 0.1\nu_3^N\nu_5^N, \tag{191}$$

$$\nu_4^{N+1} = \nu_4^N - 0.1\nu_3^N\nu_4^N - 0.8\left(\nu_4^N\right)^2 - 0.1\nu_4^N\nu_5^N, \tag{192}$$

$$\nu_5^{N+1} = \nu_5^N - 0.1\nu_3^N\nu_5^N - 0.1\nu_4^N\nu_5^N - 0.8\left(\nu_5^N\right)^2. \tag{193}$$

Eqs. (191), (192), (193) satisfy the assumptions of Lemma 8 and the diagonally dominant condition (122). Then, by Theorem 11, they converge for arbitrary initial values and

$$\lim_{N\to\infty} N\nu_i^N = 1,\, i = 3, 4, 5. \tag{194}$$

Therefore, by (106)

$$\lim_{N\to\infty} N\bar{\mu}_i^N = \sigma_i = 1.389,\, i = 3, 4, 5, \tag{195}$$

and by (183),

$$\lim_{N\to\infty} N\bar{\mu}_i^N = -3\sigma_3/2 = -2.083,\, i = 1, 2. \tag{196}$$

We will show in Fig. 8 the comparison of the numerical results and the values of (195), (196).

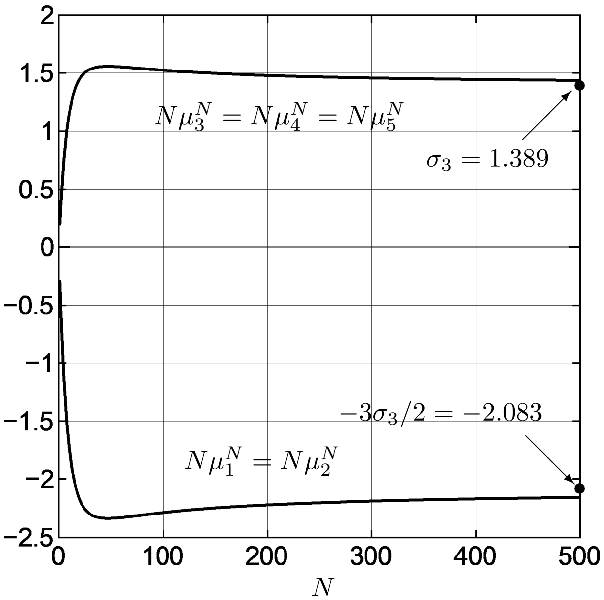

Fig. 8. Convergence of $N\mu_i^N$ in Example 7.

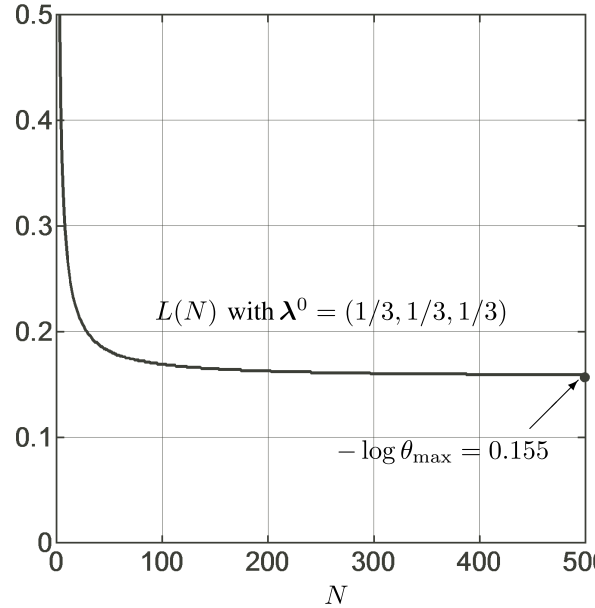

Fig. 9. Convergence of $L(N)$ in Example 8 with initial distribution $\boldsymbol{\lambda}^0 = (1/3, 1/3, 1/3)$.

C. Exponential convergence where indices of types-I and -III exist

Example 8: Consider the channel matrix $\Phi^{(3)}$ of (30)

$$\Phi^{(3)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{pmatrix}, \tag{197}$$

and an initial distribution $\boldsymbol{\lambda}^0 = (1/3, 1/3, 1/3)$. We have

$$C = 0.310 \,\text{[nat/symbol]}, \tag{198}$$
$$\boldsymbol{\lambda}^* = (0.500, 0.500, 0.000), \tag{199}$$
$$Q^* = (0.450, 0.450, 0.100), \tag{200}$$
$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} 0.228 & -0.228 & 0.000 \\ -0.228 & 0.228 & 0.000 \\ -0.428 & -0.428 & 0.856 \end{pmatrix}. \tag{201}$$

By (199) and (200), we see that $\lambda_1^* > 0, \lambda_2^* > 0, \lambda_3^* = 0$ and $C = D(P^1\|Q^*) = D(P^2\|Q^*) > D(P^3\|Q^*)$, thus $i = 1, 2$ are type-I indices, and $i = 3$ is a type-III index.

The eigenvalues of $J(\boldsymbol{\lambda}^*)$ are $(\theta_1, \theta_2, \theta_3) = (0.000, 0.456, 0.856)$. Then, $\theta_{\max} = \theta_3 = 0.856$. We have for $N = 500$

$$L(500) = 0.159 \doteqdot -\log\theta_{\max} = 0.155. \tag{202}$$

See Fig. 9.

Extending this result, we have the following lemma.

Lemma 11: Assume that type-II indices do not exist and $\theta_{\max} = \max_{i\in\mathcal{I}_{\text{III}}} \theta_i$, i.e., the maximum eigenvalue of $J(\boldsymbol{\lambda}^*)$ is achieved in J^{III}. Then, the convergence speed does not depend on the choice of initial distribution. In other words, $\lim_{N\to\infty} L(N) = -\log\theta_{\max}$ holds for arbitrary initial distribution, hence the convergence speed cannot be increased any more.

Proof: Let $\theta_{\max} = \theta_{i^*}$, $i^* \in \mathcal{I}_{\text{III}}$. We have $\theta_{i^*} > 0$ by Theorem 4. J^{III} is diagonal by (56), thus we can take ${}^t\boldsymbol{e}_{i^*} = {}^t(0, \ldots, 0, \overset{i^*\text{th}}{\overset{\vee}{1}}, 0, \ldots, 0)$ as a right eigenvector for θ_{i^*}. However, $\boldsymbol{e}_{i^*}$ is not a left eigenvector for θ_{i^*}. In fact, since every row sum of $J(\boldsymbol{\lambda}^*)$ is 0 by Lemma 3, putting $\mathbf{1} \equiv (1, \ldots, 1) \in \mathbb{R}^m$, we have $J(\boldsymbol{\lambda}^*)^t\mathbf{1} = \mathbf{0}$. If $\boldsymbol{e}_{i^*}$ were a left eigenvector for θ_{i^*}, then $0 = \boldsymbol{e}_{i^*} J(\boldsymbol{\lambda}^*)^t\mathbf{1} = \theta_{i^*}\boldsymbol{e}_{i^*}{}^t\mathbf{1} = \theta_{i^*} > 0$, which is a contradiction. Therefore, by Lemmas 9 and 10, the convergence speed does not depend on the choice of initial distribution and $\lim_{N\to\infty} L(N) = -\log\theta_{\max}$. ■

VII. Convergence speed of $I(\boldsymbol{\lambda}^N, \Phi) \to C$

Based on the results obtained so far, we will consider the convergence speed that the mutual information $I(\boldsymbol{\lambda}^N, \Phi)$ tends to C as $N \to \infty$. We will show that if $\mathcal{I}_{\text{III}} = \emptyset$ and $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$ is the convergence of order $O(1/N)$, then $I(\boldsymbol{\lambda}^N, \Phi) \to C$ is of order $O(1/N^2)$. Including this fact, we have the following theorem.

Theorem 12: Let $\mathcal{I}_{\text{III}} = \emptyset$.

If $\|\boldsymbol{\mu}^N\| < K_1 \cdot (\theta)^N$, $0 \le \theta_{\max} < \theta < 1$, $K_1 > 0$, $N = 0, 1, \ldots$, then

$$0 < C - I(\boldsymbol{\lambda}^N, \Phi) < K_2 \cdot (\theta)^{2N}, \quad K_2 > 0,\ N = 0, 1, \ldots. \tag{203}$$

Meanwhile, if $\lim_{N\to\infty} N\mu_i^N = \sigma_i \neq 0$, $i = 1, \ldots, m$, then

$$\lim_{N\to\infty} N^2 \left(C - I(\boldsymbol{\lambda}^N, \Phi)\right) = \frac{1}{2}\sum_{j=1}^{n} \frac{1}{Q_j^*}\left(\sum_{i=1}^{m} \sigma_i P_j^i\right)^2. \tag{204}$$

Next, let $\mathcal{I}_{\text{III}} \neq \emptyset$.

If $\|\boldsymbol{\mu}^N\| < K_1 \cdot (\theta)^N$, $0 \le \theta_{\max} < \theta < 1$, $K_1 > 0$, $N = 0, 1, \ldots$, then

$$0 < C - I(\boldsymbol{\lambda}^N, \Phi) < K_2 \cdot (\theta)^N, \quad K_2 > 0,\ N = 0, 1, \ldots. \tag{205}$$

Meanwhile, if $\lim_{N\to\infty} N\mu_i^N = \sigma_i \neq 0,\ i = 1, \ldots, m$, then

$$0 < C - I(\boldsymbol{\lambda}^N, \Phi) < K/N, \quad (206)$$
$$K > 0,\ N = 0, 1, \ldots.$$

Proof: See Appendix H. ■

We will show in the following tables the evaluation of the convergence speed of $I(\boldsymbol{\lambda}^N, \Phi^{(k)}) \to C$ for $k = 1, 2, 3$, where $\Phi^{(1)}$, $\Phi^{(2)}$, and $\Phi^{(3)}$ were defined in Examples 1, 2, and 3, and examined in Examples 4, 6, and 8, respectively.

TABLE I
CONVERGENCE SPEED OF $I(\boldsymbol{\lambda}^N, \Phi^{(1)}) \to C$.

$-(1/N)\log\left(C - I(\boldsymbol{\lambda}^N, \Phi^{(1)})\right)\|_{N=500}$	0.324
$-2\log\theta_{\max}$	0.313

TABLE II
CONVERGENCE SPEED OF $I(\boldsymbol{\lambda}^N, \Phi^{(2)}) \to C$.

$N^2\left(C - I(\boldsymbol{\lambda}^N, \Phi^{(2)})\right)\|_{N=500}$	0.516
Eq. (204)	0.500

TABLE III
CONVERGENCE SPEED OF $I(\boldsymbol{\lambda}^N, \Phi^{(3)}) \to C$.

$-(1/N)\log\left(C - I(\boldsymbol{\lambda}^N, \Phi^{(3)})\right)\|_{N=500}$	0.163
$-\log\theta_{\max}$	0.155

We can see from these tables that the convergence speed of $I\left(\boldsymbol{\lambda}^N, \Phi\right) \to C$ is accurately approximated by the results of Theorem 12.

VIII. CONCLUSION

In this paper, we investigated the convergence speed of the Arimoto-Blahut algorithm. We showed that the capacity-achieving input distribution $\boldsymbol{\lambda}^*$ is the fixed point of $F(\boldsymbol{\lambda})$ and analyzed the convergence speed by the Taylor expansion of $F(\boldsymbol{\lambda})$ about $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. We concretely calculated the Jacobian matrix J of the first-order term of the Taylor expansion and the Hessian matrix H of the second-order term. The analysis of the convergence speed by the Hessian matrix H was done for the first time in this paper.

We showed that if type-II indices do not exist, then the convergence of $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$ is exponential, and if type-II indices exist, then the convergence of the second-order recurrence formula obtained by truncating the Taylor expansion is of order $O(1/N)$ for some initial vector. Further, we considered the condition for the convergence of order $O(1/N)$ for an arbitrary initial vector. Next, we considered the convergence speed of $I(\boldsymbol{\lambda}^N, \Phi) \to C$ and showed that the type-III indices concern the convergence speed. In particular, if there exist no type-III indices and $\boldsymbol{\lambda}^N \to \boldsymbol{\lambda}^*$ is of order $O(1/N)$, then $I(\boldsymbol{\lambda}^N, \Phi) \to C$ is $O(1/N^2)$.

Based on these analyses, the convergence speeds for several channel matrices were numerically evaluated. As a result, it was confirmed that the convergence speed of the Arimoto-Blahut algorithm is accurately approximated by the values obtained by our theorems.

In this paper and in [2], [21], the capacity-achieving $\boldsymbol{\lambda}^*$ is used to analyze the convergence speed of the sequence $\{\boldsymbol{\lambda}^N\}$. Using the limit vector is a common method for the analysis of the convergence speed. For example, see [9], Proposition 4.4. However, $\boldsymbol{\lambda}^*$ is not known before it is calculated by the Arimoto-Blahut algorithm, etc. Determining whether the convergence speed is fast or slow by a simple calculation from the channel matrix Φ without using $\boldsymbol{\lambda}^*$ would be effective, so we would like to consider finding this condition as a future work.

REFERENCES

[1] L.V. Ahlfors, *Complex Analysis* (third edition), McGraw-Hill, 1979.
[2] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels", IEEE Trans. Inf. Theory, vol. 18, pp. 14-20, Jan. 1972.
[3] S. Arimoto, *Information Theory*, Kyoritsu Shuppan, Tokyo, 1976. (in Japanese)
[4] R. E. Blahut, "Computation of channel capacity and rate-distortion functions", IEEE Trans. Inf. Theory, vol. 18, no. 4, pp. 460-473, Jul. 1972.
[5] T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series*, third edition, Chelsea Publishing Company, New York, 1991.
[6] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, June 2006.
[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Orlando, 1982.
[8] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures", Statistics and Decisions, Supplement Issue No.1, 205-237, 1984.
[9] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, 1987,
[10] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Second Ed., 1985.
[11] G. Matz and P. Duhamel, "Information geometric formulation and interpretation of accelerated Blahut-Arimoto-Type algorithms", in proceedings of ITW2004, 2004.
[12] 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.
[13] 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.
[14] K. Nakagawa and F. Kanaya, "A new geometric capacity characterization of a discrete memoryless channel", IEEE Trans. Inf. Theory, vol. 34, no. 2, pp. 318-321, 1988.
[15] K. Nakagawa, K. Watabe and T. Sabu, "On the search algorithm for the output distribution that achieves the channel capacity", IEEE Trans. Inf. Theory, vol. 63, no. 2, pp. 1043-1062, Feb. 2017.
[16] K. Nakagawa, Y. Takei and K. Watabe, "Analysis for the slow convergence in Arimoto algorithm", in proceedings of ISITA2018, pp. 764-768, Oct. 2018.
[17] M. Rezaeian and A. Grant, "A generalization of Arimoto-Blahut algorithm", in proceedings of IEEE ISIT 2004.
[18] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos*, CRC Press, 1998.
[19] I. Satake, *Linear Algebra*, Marcel Dekker, 1975.
[20] P. O. Vontobel, A. Kavčić, 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.
[21] Y. Yu, "Squeezing the Arimoto-Blahut algorithm for faster convergence", IEEE Trans. Inf. Theory, vol. 56, no. 7, pp. 3149-3157, Jul. 2010.

Appendix A
Proof of Lemma 5

Since $\sqrt{\Lambda}B\sqrt{\Lambda}^{-1}$ is symmetric by (69), it is diagonalizable, hence J^{I} is diagonalizable because $J^{\mathrm{I}} = I - B$. Therefore, there exists a regular matrix $\Pi \in \mathbb{R}^{m_1 \times m_1}$ with

$$\Pi^{-1} J^{\mathrm{I}} \Pi = \Theta_1, \tag{207}$$

where Θ_1 is a diagonal matrix whose diagonal components are the eigenvalues of J^{I}, i.e., $\Theta_1 = \mathrm{diag}(\theta_i,\, i \in \mathcal{I}_{\mathrm{I}}) \in \mathbb{R}^{m_1 \times m_1}$. We can write it by components as $\Theta_1 = (\theta_i \delta_{i'i})\,,\, i', i \in \mathcal{I}_{\mathrm{I}}$.

Next, let Θ_2 be a diagonal matrix whose diagonal components are the eigenvalues of J^{II} and J^{III}, i.e., $\Theta_2 = \mathrm{diag}(\theta_i,\, i \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}}) \in \mathbb{R}^{(m_2+m_3)\times(m_2+m_3)}$. We have $\Theta_2 = (\theta_i \delta_{i'i})\,,\, i', i \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}}$. Then, by (53), we have

$$J(\boldsymbol{\lambda}^*) = \begin{pmatrix} J^{\mathrm{I}} & O \\ U & \Theta_2 \end{pmatrix}, \tag{208}$$

where $U \in \mathbb{R}^{(m_2+m_3)\times m_1}$ is an appropriate matrix.

Now, we will prove that there exists a unique matrix $V \in \mathbb{R}^{(m_2+m_3)\times m_1}$ that satisfies

$$V\Theta_1 - \Theta_2 V = U\Pi. \tag{209}$$

Define the components of V by $V \equiv (v_{i'i})\,,\, i' \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}},\, i \in \mathcal{I}_{\mathrm{I}}$. Then,

$$\begin{aligned} (V\Theta_1)_{i'i} &= \sum_{k \in \mathcal{I}_{\mathrm{I}}} v_{i'k} \theta_i \delta_{ki} = v_{i'i}\theta_i, \quad (210) \\ & \quad i' \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}},\, i \in \mathcal{I}_{\mathrm{I}}, \\ (\Theta_2 V)_{i'i} &= \sum_{k \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}}} \theta_{i'} \delta_{ki'} v_{ki} = v_{i'i}\theta_{i'}, \quad (211) \\ & \quad i' \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}},\, i \in \mathcal{I}_{\mathrm{I}}. \end{aligned}$$

Further, defining the components of $U\Pi$ by $U\Pi = \left(u^{\Pi}_{i'i}\right)\,,\, i' \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}},\, i \in \mathcal{I}_{\mathrm{I}}$, both sides of (209) are represented by components as

$$(\theta_i - \theta_{i'})\, v_{i'i} = u^{\Pi}_{i'i},\; i' \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}},\, i \in \mathcal{I}_{\mathrm{I}}. \tag{212}$$

By Theorem 2, the eigenvalues of J^{I} are less than 1, hence different from the eigenvalues 1 of J^{II}. Further, $\theta_i \neq \theta_{i'},\, i \in \mathcal{I}_{\mathrm{I}},\, i' \in \mathcal{I}_{\mathrm{III}}$ by the assumption (75), so we have

$$v_{i'i} = \frac{u^{\Pi}_{i'i}}{\theta_i - \theta_{i'}},\; i' \in \mathcal{I}_{\mathrm{II}} \cup \mathcal{I}_{\mathrm{III}},\, i \in \mathcal{I}_{\mathrm{I}}, \tag{213}$$

which shows the existence and uniqueness of $V \in \mathbb{R}^{(m_2+m_3)\times m_1}$ that satisfies (209).

Now, define

$$\tilde{\Pi} \equiv \begin{pmatrix} \Pi & O \\ V & I \end{pmatrix} \in \mathbb{R}^{m \times m}. \tag{214}$$

Then, by noting (207), (208), (209), we have

$$\tilde{\Pi}^{-1} J(\boldsymbol{\lambda}^*) \tilde{\Pi} = \begin{pmatrix} \Theta_1 & O \\ O & \Theta_2 \end{pmatrix}, \tag{215}$$

which proves the lemma.

Appendix B
Proof of Theorem 5

Consider the line segment with the starting point $\boldsymbol{\lambda}^*$ and the end point $\boldsymbol{\lambda}^N$, i.e.,

$$\boldsymbol{\lambda}(t) \equiv (1-t)\boldsymbol{\lambda}^* + t\boldsymbol{\lambda}^N,\; 0 \le t \le 1. \tag{216}$$

The components of (216) are $\lambda_i(t) = (1-t)\lambda_i^* + t\lambda_i^N,\; i = 1, \ldots, m$. Let us define

$$f(t) \equiv F(\boldsymbol{\lambda}(t)) \in \Delta(\mathcal{X}) \tag{217}$$

and write its components as $f(t) = (f_1(t), \ldots, f_m(t))$. We have

$$\begin{aligned} \frac{df_i(t)}{dt} &= \sum_{i'=1}^{m} \frac{d\lambda_{i'}(t)}{dt} \left.\frac{\partial F_i}{\partial \lambda_{i'}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}(t)} \quad (218) \\ &= \sum_{i'=1}^{m} (\lambda_{i'}^N - \lambda_{i'}^*) \left.\frac{\partial F_i}{\partial \lambda_{i'}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}(t)} \quad (219) \\ &= \left((\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*) J(\boldsymbol{\lambda}(t))\right)_i,\; i = 1, \ldots, m, \quad (220) \end{aligned}$$

thus

$$\frac{df(t)}{dt} = (\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*) J(\boldsymbol{\lambda}(t)). \tag{221}$$

Now, by the relation between the matrix norm and the maximum eigenvalue [10], p.347, Lemma 5.6.10, for $\epsilon \equiv \theta - \theta_{\max} > 0$, there exists a vector norm $\|\cdot\|'$ in $\mathbb{R}^m$ whose associated matrix norm $\|\cdot\|'$ satisfies

$$\theta_{\max} \le \|J(\boldsymbol{\lambda}^*)\|' < \theta_{\max} + \epsilon. \tag{222}$$

(Note that $'$ does not denote the derivative.) By the continuity of norm, for any ϵ_1 with $0 < \epsilon_1 < \theta_{\max} + \epsilon - \|J(\boldsymbol{\lambda}^*)\|'$, there exists $\delta' > 0$ such that if $\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*\|' < \delta'$, then $\left|\|J(\boldsymbol{\lambda})\|' - \|J(\boldsymbol{\lambda}^*)\|'\right| < \epsilon_1$, especially, $\|J(\boldsymbol{\lambda})\|' < \|J(\boldsymbol{\lambda}^*)\|' + \epsilon_1$. Thus,

$$\begin{aligned} \|J(\boldsymbol{\lambda})\|' &< \|J(\boldsymbol{\lambda}^*)\|' + \theta_{\max} + \epsilon - \|J(\boldsymbol{\lambda}^*)\|' \quad (223) \\ &= \theta < 1. \quad (224) \end{aligned}$$

By the mean value theorem, there exists $t^N \in [0,1]$, which satisfies

$$\begin{aligned} \|\boldsymbol{\lambda}^{N+1} - \boldsymbol{\lambda}^*\|' &= \|F(\boldsymbol{\lambda}^N) - F(\boldsymbol{\lambda}^*)\|' \quad (225) \\ &= \|f(1) - f(0)\|' \quad (226) \\ &\le \left\| \left.\frac{df(t)}{dt}\right|_{t=t^N} \right\|' (1-0) \quad (227) \\ &= \|(\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*) J(\boldsymbol{\lambda}(t^N))\|' \quad \text{(by (221))} \quad (228) \\ &\le \|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\|' \, \|J(\boldsymbol{\lambda}(t^N))\|'. \quad (229) \end{aligned}$$

Here, if $\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\|' < \delta'$, then we have $\|J(\boldsymbol{\lambda}^N)\|' < \theta < 1$ by (224), so $\|\boldsymbol{\lambda}^{N+1} - \boldsymbol{\lambda}^*\|' < \delta'$ by (229). Thus, by mathematical induction, if the initial distribution $\boldsymbol{\lambda}^0$ satisfies $\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|' < \delta'$, then $\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\|' < \delta'$ for all N, and so $\|J(\boldsymbol{\lambda}^N)\|' < \theta < 1$ by (224).

Therefore, by (224), (229), $\|\boldsymbol{\lambda}^{N+1} - \boldsymbol{\lambda}^*\|' < \theta\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\|' < \ldots < \theta^{N+1}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|'$, so we have

$$\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\|' < (\theta)^N \|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|',\; N = 0, 1, \ldots. \tag{230}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3095406

By the equivalence of norms in the finite dimensional vector space [10], [18], we can change the norm from $\|\cdot\|'$ to the Euclidean norm $\|\cdot\|$ to have

$$\|\boldsymbol{\lambda}^N - \boldsymbol{\lambda}^*\| \leq K \cdot (\theta)^N,\ K > 0,\ N = 0, 1, \ldots. \tag{231}$$

APPENDIX C
PROOF OF THEOREM 6 (CALCULATION OF HESSIAN MATRIX $H_i(\boldsymbol{\lambda}^*)$)

We will calculate the Hessian matrix $H_i(\boldsymbol{\lambda}^*)$ of $F_i(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, i.e., $H_i(\boldsymbol{\lambda}^*) = (\partial^2 F_i / \partial\lambda_{i'}\partial\lambda_{i''}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*})$.

Differentiating both sides of (39) with respect to $\lambda_{i''}$, we have

$$\begin{aligned}
&\underbrace{\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}} \sum_{k=1}^{m} \lambda_k e^{D_k}}_{\star 1} + \underbrace{\frac{\partial F_i}{\partial\lambda_{i'}}}_{\star 2} \underbrace{\frac{\partial}{\partial\lambda_{i''}} \sum_{k=1}^{m} \lambda_k e^{D_k}}_{\star 3} \\
&\quad + \frac{\partial F_i}{\partial\lambda_{i''}} \frac{\partial}{\partial\lambda_{i'}} \sum_{k=1}^{m} \lambda_k e^{D_k} + F_i \underbrace{\frac{\partial^2}{\partial\lambda_{i'}\partial\lambda_{i''}} \sum_{k=1}^{m} \lambda_k e^{D_k}}_{\star 4} \\
&= \delta_{ii'} e^{D_i} \frac{\partial D_i}{\partial\lambda_{i''}} + \delta_{ii''} e^{D_i} \frac{\partial D_i}{\partial\lambda_{i'}} + \lambda_i e^{D_i} \frac{\partial D_i}{\partial\lambda_{i''}} \frac{\partial D_i}{\partial\lambda_{i'}} \\
&\quad + \lambda_i e^{D_i} \frac{\partial^2 D_i}{\partial\lambda_{i'}\partial\lambda_{i''}}.
\end{aligned} \tag{232}$$

On the left hand side of (232), part $\star 1$ is evaluated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ by (45). Part $\star 2$ is the component of the Jacobian matrix which is evaluated by (51). Part $\star 3$ is evaluated by (47). Part $\star 4$ is evaluated as follows. We have

$$\begin{aligned}
&\frac{\partial^2}{\partial\lambda_{i'}\partial\lambda_{i''}} \sum_{k=1}^{m} \lambda_k e^{D_k} = \frac{\partial}{\partial\lambda_{i''}} \left(e^{D_{i'}} + \sum_{k=1}^{m} \lambda_k e^{D_k} \frac{\partial D_k}{\partial\lambda_{i'}} \right) \\
&= e^{D_{i'}} \frac{\partial D_{i'}}{\partial\lambda_{i''}} + \sum_{k=1}^{m} \left(\delta_{ki''} e^{D_k} \frac{\partial D_k}{\partial\lambda_{i'}} + \lambda_k e^{D_k} \frac{\partial D_k}{\partial\lambda_{i''}} \frac{\partial D_k}{\partial\lambda_{i'}} \right. \\
&\quad \left. + \lambda_k e^{D_k} \frac{\partial^2 D_k}{\partial\lambda_{i'}\partial\lambda_{i''}} \right) \\
&= e^{D_{i'}} \frac{\partial D_i'}{\partial\lambda_{i''}} + e^{D_{i''}} \frac{\partial D_{i''}}{\partial\lambda_{i'}} + \sum_{k=1}^{m} \lambda_k e^{D_k} \frac{\partial D_k}{\partial\lambda_{i'}} \frac{\partial D_k}{\partial\lambda_{i''}} \\
&\quad + \underbrace{\sum_{k=1}^{m} \lambda_k e^{D_k} \frac{\partial^2 D_k}{\partial\lambda_{i'}\partial\lambda_{i''}}}_{\star 5},
\end{aligned} \tag{233}$$

and part $\star 5$ becomes, at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$,

$$\begin{aligned}
\sum_{k=1}^{m} \lambda_k e^{D_k} \frac{\partial^2 D_k}{\partial\lambda_{i'}\partial\lambda_{i''}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} &= e^C \sum_{k=1}^{m_1} \lambda_k^* \sum_{j=1}^{n} \frac{P_j^k P_j^{i'} P_j^{i''}}{\left(Q_j^*\right)^2} \\
&= e^C \sum_{j=1}^{n} \frac{P_j^{i'} P_j^{i''}}{Q_j^*} \sum_{k=1}^{m_1} \frac{\lambda_k^* P_j^k}{Q_j^*} \\
&= -e^C D^*_{i',i''}.
\end{aligned} \tag{234}$$

Therefore, part $\star 4$ becomes, at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$,

$$\begin{aligned}
\frac{\partial^2}{\partial\lambda_{i'}\partial\lambda_{i''}} \sum_{k=1}^{m} \lambda_k e^{D_k} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} &= e^{D^*_{i'}} D^*_{i',i''} + e^{D^*_{i''}} D^*_{i',i''} \\
&\quad + e^C \sum_{k=1}^{m_1} \lambda_k^* D^*_{k,i'} D^*_{k,i''} - e^C D^*_{i',i''}.
\end{aligned} \tag{235}$$

Let us define $D^*_{i,i',i''} \equiv \partial^2 D_i / \partial\lambda_{i'}\partial\lambda_{i''}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}$ and $E_{i',i''} \equiv \sum_{k=1}^{m_1} \lambda_k^* D^*_{k,i'} D^*_{k,i''}$. Then, based on the above calculation, we have

$$\begin{aligned}
&\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} e^C \\
&\quad + \left\{ e^{D_i^* - C} \left(\delta_{ii'} + \lambda_i^* D^*_{i,i'} \right) + \lambda_i^* \left(1 - e^{D^*_{i'} - C} \right) \right\} \\
&\qquad \times \left(e^{D^*_{i''}} - e^C \right) \\
&\quad + \left\{ e^{D_i^* - C} \left(\delta_{ii''} + \lambda_i^* D^*_{i,i''} \right) + \lambda_i^* \left(1 - e^{D^*_{i''} - C} \right) \right\} \\
&\qquad \times \left(e^{D^*_{i'}} - e^C \right) \\
&\quad + F_i^* \left(e^{D^*_{i'}} D^*_{i',i''} + e^{D^*_{i''}} D^*_{i',i''} + e^C E_{i',i''} - e^C D^*_{i',i''} \right) \\
&= \delta_{ii'} e^{D_i^*} D^*_{i,i''} + \delta_{ii''} e^{D_i^*} D^*_{i,i'} + \lambda_i^* e^{D_i^*} D^*_{i,i'} D^*_{i,i''} \\
&\quad + \lambda_i^* e^{D_i^*} D^*_{i,i',i''}.
\end{aligned} \tag{236}$$

By rearranging this equation, we obtain, using (48) of Lemma 4,

$$\begin{aligned}
&\frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} \\
&= e^{D_i^* - C} \Big\{ \delta_{ii'} D^*_{i,i''} + \delta_{ii''} D^*_{i,i'} + \lambda_i^* \left(D^*_{i,i'} D^*_{i,i''} + D^*_{i,i',i''} \right) \\
&\quad + \left(\delta_{ii'} + \lambda_i^* D^*_{i,i'} \right) \left(1 - e^{D^*_{i''} - C} \right) \\
&\quad + \left(\delta_{ii''} + \lambda_i^* D^*_{i,i''} \right) \left(1 - e^{D^*_{i'} - C} \right) \Big\} \\
&\quad + 2\lambda_i^* \left(1 - e^{D^*_{i'} - C} \right) \left(1 - e^{D^*_{i''} - C} \right) \\
&\quad - \lambda_i^* \left(e^{D^*_{i'} - C} D^*_{i',i''} + e^{D^*_{i''} - C} D^*_{i',i''} + E_{i',i''} - D^*_{i',i''} \right).
\end{aligned} \tag{237}$$

APPENDIX D
PROOF OF STEP2

For $i = m' + 1, \ldots, m$, let $H_{i,i'i''}$ be the (i', i'') component of the Hessian matrix $H_i(\boldsymbol{\lambda}^*)$. Then, by Theorem 6,

$$\begin{aligned}
H_{i,i'i''} &= \frac{\partial^2 F_i}{\partial\lambda_{i'}\partial\lambda_{i''}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} \\
&= \delta_{ii''} \left(1 - e^{D^*_{i'} - C} + D^*_{i,i'} \right) + \delta_{ii'} \left(1 - e^{D^*_{i''} - C} + D^*_{i,i''} \right),
\end{aligned} \tag{238}$$

$$i = m' + 1, \ldots, m,\ i', i'' = 1, \ldots, m.$$

Here, for the simplicity of symbols, define

$$S_{ii'} \equiv 1 - e^{D^*_{i'} - C} + D^*_{i,i'},\ i' = 1, \ldots, m, \tag{239}$$

then (238) becomes

$$H_{i,i'i''} = \delta_{ii''} S_{ii'} + \delta_{ii'} S_{ii''}. \tag{240}$$

We have $S_{ii'} = D^*_{i,i'}$ for $i' = m' + 1, \ldots, m$, then by (240), the Hessian matrix $H_i(\boldsymbol{\lambda}^*)$ is given by (241)-(245).

Therefore, by (99),

$$\begin{aligned}
\frac{1}{2} \bar{\boldsymbol{\mu}}^N H_i(\boldsymbol{\lambda}^*)\, {}^t\bar{\boldsymbol{\mu}}^N &= \frac{1}{2} \left(\bar{\boldsymbol{\mu}}^N_{\mathrm{I,III}}, \bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \right) \begin{pmatrix} O & H_i^1 \\ {}^tH_i^1 & H_i^2 \end{pmatrix} \begin{pmatrix} {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{I,III}} \\ {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \end{pmatrix} \\
&= \frac{1}{2} \left(-\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} A_2 A_1^{-1}, \bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \right) \begin{pmatrix} O & H_i^1 \\ {}^tH_i^1 & H_i^2 \end{pmatrix} \begin{pmatrix} -{}^tA_1^{-1}\, {}^tA_2\, {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \\ {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \end{pmatrix} \\
&= \frac{1}{2} \bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \left(-A_2 A_1^{-1} H_i^1 - {}^tH_i^1\, {}^tA_1^{-1}\, {}^tA_2 + H_i^2 \right) {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{II}}.
\end{aligned} \tag{246}$$

$$H_i(\boldsymbol{\lambda}^*) = \left(\begin{array}{ccc:ccccccc}
 & & & 0 & \dots & 0 & S_{i1} & 0 & \dots & 0 \\
 & O & & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
 & & & 0 & \dots & 0 & S_{im'} & 0 & \dots & 0 \\
\hdashline
0 & \dots & 0 & 0 & \dots & 0 & D^*_{i,m'+1} & 0 & \dots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & D^*_{i,i-1} & 0 & \dots & 0 \\
S_{i1} & \dots & S_{im'} & D^*_{i,m'+1} & \dots & D^*_{i,i-1} & 2D^*_{i,i} & D^*_{i,i+1} & \dots & D^*_{i,m} \\
0 & \dots & 0 & 0 & \dots & 0 & D^*_{i,i+1} & 0 & \dots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & D^*_{i,m} & 0 & \dots & 0
\end{array}\right) \tag{241}$$

$$\equiv \begin{pmatrix} O & H_i^1 \\ {}^tH_i^1 & H_i^2 \end{pmatrix}, \tag{242}$$

where $O \in \mathbb{R}^{m' \times m'}$, $H_i^1 \in \mathbb{R}^{m' \times m_2}$, $H_i^2 \in \mathbb{R}^{m_2 \times m_2}$ and

$$H_i^1 \equiv \left(\delta_{ii''} S_{ii'}\right),\ i' = 1, \ldots, m', i'' = m'+1, \ldots, m, \tag{243}$$

$${}^tH_i^1 \equiv \left(\delta_{ii'} S_{ii''}\right),\ i' = m'+1, \ldots, m, i'' = 1, \ldots, m', \tag{244}$$

$$H_i^2 \equiv \left(\delta_{ii''} D^*_{ii'} + \delta_{ii'} D^*_{ii''}\right),\ i', i'' = m'+1, \ldots, m. \tag{245}$$

Now, define

$$G_i \equiv -A_2 A_1^{-1} H_i^1 \in \mathbb{R}^{m_2 \times m_2}, \tag{247}$$

and $A_1^{-1} \equiv (\zeta_{i'i''})$. Further, let $G_{i,i'i''}$ be the (i', i'') component of G_i. Then, by (247), we have

$$G_{i,i'i''} = -\sum_{k,k'=1}^{m'} a_{i'k} \zeta_{kk'} H^1_{i,k'i''} \tag{248}$$

$$= -\sum_{k,k'=1}^{m'} a_{i'k} \zeta_{kk'} \delta_{ii''} S_{ik'} \tag{249}$$

$$= -\delta_{ii''} \sum_{k,k'=1}^{m'} a_{i'k} \zeta_{kk'} S_{ik'}, \tag{250}$$

$$i', i'' = m'+1, \ldots, m.$$

Define

$$T_{ii'} \equiv -\sum_{k,k'=1}^{m'} a_{i'k} \zeta_{kk'} S_{ik'},\ i' = m'+1, \ldots, m, \tag{251}$$

then (250) becomes

$$G_{i,i'i''} = \delta_{ii''} T_{ii'}. \tag{252}$$

Thus, we have

$$G_i = \left(\delta_{ii''} T_{ii'}\right),\ {}^tG_i = \left(\delta_{ii'} T_{ii''}\right), \tag{253}$$

$$i', i'' = m'+1, \ldots, m.$$

Hence, by (246),

$$\frac{1}{2} \bar{\boldsymbol{\mu}}^N H_i(\boldsymbol{\lambda}^*)\, {}^t\bar{\boldsymbol{\mu}}^N = \frac{1}{2} \bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \left(G_i + {}^tG_i + H_i^2\right) {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{II}}. \tag{254}$$

Define

$$\hat{H}_i \equiv G_i + {}^tG_i + H_i^2, \tag{255}$$

and

$$r_{ii'} \equiv T_{ii'} + D^*_{i,i'}, \tag{256}$$

and let $\hat{H}_{i,i'i''}$ be the (i', i'') component of $\hat{H}_i$. Then, by (255), (253), (245), and (256), we have

$$\hat{H}_{i,i'i''} = \delta_{ii''} T_{ii'} + \delta_{ii'} T_{ii''} + \left(\delta_{ii''} D^*_{i,i'} + \delta_{ii'} D^*_{i,i''}\right) \tag{257}$$

$$= \delta_{ii''} \left(T_{ii'} + D^*_{i,i'}\right) + \delta_{ii'} \left(T_{ii''} + D^*_{i,i''}\right) \tag{258}$$

$$= \delta_{ii''} r_{ii'} + \delta_{ii'} r_{ii''}. \tag{259}$$

Therefore, (254) becomes

$$\frac{1}{2} \bar{\boldsymbol{\mu}}^N H_i(\boldsymbol{\lambda}^*)\, {}^t\bar{\boldsymbol{\mu}}^N = \frac{1}{2} \bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \hat{H}_i\, {}^t\bar{\boldsymbol{\mu}}^N_{\mathrm{II}} \tag{260}$$

$$= \frac{1}{2} \sum_{i',i''=m'+1}^{m} \bar{\mu}^N_{i'} \hat{H}_{i,i'i''} \bar{\mu}^N_{i''} \tag{261}$$

$$= \frac{1}{2} \sum_{i',i''=m'+1}^{m} \bar{\mu}^N_{i'} \left(\delta_{ii''} r_{ii'} + \delta_{ii'} r_{ii''}\right) \bar{\mu}^N_{i''} \tag{262}$$

$$= \frac{1}{2} \left(\sum_{i'=m'+1}^{m} \bar{\mu}^N_{i'} r_{ii'} \bar{\mu}^N_i + \sum_{i''=m'+1}^{m} \bar{\mu}^N_i r_{ii''} \bar{\mu}^N_{i''} \right) \tag{263}$$

$$= \bar{\mu}^N_i \sum_{i'=m'+1}^{m} r_{ii'} \bar{\mu}^N_{i'}. \tag{264}$$

Summarizing the above, the recurrence formula satisfied by $\bar{\mu}^N_i$, $i = m'+1, \ldots, m$, is

$$\bar{\mu}^{N+1}_i = \bar{\mu}^N_i + \bar{\mu}^N_i \sum_{i'=m'+1}^{m} r_{ii'} \bar{\mu}^N_{i'}, \tag{265}$$

$$i = m'+1, \ldots, m.$$

Appendix E
Proof of Theorem 10

We can prove Theorem 10 for any $m_2 \geq 3$, but because the symbols become complicated, we will give a proof for the case of $m_2 = 3$. The proof for $m_2 = 3$ does not lose generality, hence the extension to $m_2 \geq 3$ is easy. We should prove the following Theorem.

Theorem 10 (case of $m_2 = 3$): Let us consider a sequence $\{\xi_i^N\}$, $i = 1, 2, 3$, $N = 0, 1, \dots$ defined by

$$\xi_i^{N+1} = \xi_i^N - \xi_i^N \sum_{i'=1}^{3} q_{ii'}\xi_{i'}^N,\ i = 1, 2, 3, \tag{266}$$

$$0 < \xi_i^0 \leq 1/2,\ i = 1, 2, 3, \tag{267}$$

where $\boldsymbol{q}_i \equiv (q_{i1}, q_{i2}, q_{i3})$ is a probability vector.

Further, we assume the diagonally dominant condition (122), i.e.,

$$q_{ii} > \sum_{i'=1, i' \neq i}^{3} q_{ii'},\ i = 1, 2, 3, \tag{268}$$

and the conjecture (123), i.e.,

$$\xi_1^N \geq \xi_2^N \geq \xi_3^N,\ N \geq N_0. \tag{269}$$

Then, we have

$$\lim_{N\to\infty} N\xi_i^N = 1,\ i = 1, 2, 3. \tag{270}$$

Lemma 12: $0 < \xi_i^N \leq 1/2$, $i = 1, 2, 3$, holds for $N = 0, 1, \dots$.

Proof: We prove by mathematical induction. For $N = 0$, the assertion holds by (267). Assuming that the assertion holds for N, by noting $1/2 \leq 1 - \sum_{i'=1} q_{ii'}\xi_{i'}^N < 1$, we have $0 < \xi_i^{N+1} \leq 1/2$. ∎

Lemma 13: The sequence $\{\xi_i^N\}$, $N = 0, 1, \dots$ is strictly decreasing.

Proof: Because $\xi_i^N - \xi_i^{N+1} = \xi_i^N \sum_{i'=1}^{3} q_{ii'}\xi_{i'}^N > 0$ holds by Lemma 12. ∎

Lemma 14: $\lim_{N\to\infty} \xi_i^N = 0$, $i = 1, 2, 3$.

Proof: $\xi_i^\infty \equiv \lim_{N\to\infty} \xi_i^N \geq 0$ exists by Lemmas 12 and 13. Then, $\xi_i^\infty = \xi_i^\infty - \xi_i^\infty \sum_{i'=1}^{3} q_{ii'}\xi_{i'}^\infty$ holds by (266), hence we have $\xi_i^\infty = 0$. ∎

Lemma 15: $\liminf_{N\to\infty} N\xi_1^N \geq 1$.

Proof: By (269), for $N \geq N_0$,

$$\xi_1^{N+1} = \xi_1^N - \xi_1^N \sum_{i'=1}^{3} q_{1i'}\xi_{i'}^N \tag{271}$$

$$\geq \xi_1^N - \xi_1^N \sum_{i'=1}^{3} q_{1i'}\xi_1^N \tag{272}$$

$$= \xi_1^N - \left(\xi_1^N\right)^2. \tag{273}$$

Now, we define a sequence $\{\hat{\xi}_1^N\}$, $N = 0, 1, \dots$ by the recurrence formula

$$\hat{\xi}_1^0 = \xi_1^{N_0}, \tag{274}$$

$$\hat{\xi}_1^{N+1} = \hat{\xi}_1^N - \left(\hat{\xi}_1^N\right)^2,\ N = 0, 1, \dots. \tag{275}$$

Then, we will prove

$$\xi_1^{N+N_0} \geq \hat{\xi}_1^N,\ N = 0, 1, \dots \tag{276}$$

by mathematical induction. For $N = 0$, (276) holds by the assumption (274). Assume that (276) holds for N. Because the function $f(\xi) = \xi - \xi^2$ is monotonically increasing in $0 < \xi \leq 1/2$, by (273), we have $\xi_1^{N+N_0+1} \geq \xi_1^{N+N_0} - (\xi_1^{N+N_0})^2 \geq \hat{\xi}_1^N - (\hat{\xi}_1^N)^2 = \hat{\xi}_1^{N+1}$; thus, (276) also holds for $N+1$. Therefore, $\liminf_{N\to\infty} N\xi_1^N \geq \liminf_{N\to\infty} N\hat{\xi}_1^N = \lim_{N\to\infty} N\hat{\xi}_1^N = 1$, where the last equality is because of Lemma 6. ∎

Lemma 16: $\limsup_{N\to\infty} N\xi_3^N \leq 1$.

Proof: By (269), for $N \geq N_0$, we have $\xi_3^{N+1} = \xi_3^N - \xi_3^N \sum_{i'=1}^{3} q_{3i'}\xi_{i'}^N \leq \xi_3^N - \xi_3^N \sum_{i'=1}^{3} q_{3i'}\xi_3^n = \xi_3^N - \left(\xi_3^N\right)^2$. Now, we define a sequence $\{\hat{\xi}_3^N\}$ by

$$\hat{\xi}_3^0 = \xi_3^{N_0}, \tag{277}$$

$$\hat{\xi}_3^{N+1} = \hat{\xi}_3^N - \left(\hat{\xi}_3^N\right)^2,\ N = 0, 1, \dots. \tag{278}$$

Then, we can prove $\xi_3^{N+N_0} \leq \hat{\xi}_3^N$, $N = 0, 1, \dots$ in a similar way as the proof of Lemma 15. Therefore, we have $\limsup_{N\to\infty} N\xi_3^N \leq \limsup_{N\to\infty} N\hat{\xi}_3^N = \lim_{N\to\infty} N\hat{\xi}_3^N = 1$, where the last equality is because of Lemma 6. ∎

Lemma 17: Let $\tau_1^N \equiv \sum_{i'=1}^{3} q_{1i'}\xi_{i'}^N$, $\tau_3^N \equiv \sum_{i'=1}^{3} q_{3i'}\xi_{i'}^N$. Then, there exists a constant $K > 0$ with

$$\tau_1^N - \tau_3^N \geq K(\xi_1^N - \xi_3^N),\ N \geq N_0. \tag{279}$$

Proof: We have

$$\tau_1^N - \tau_3^N = \sum_{i'=1}^{3} q_{1i'}\xi_{i'}^N - \sum_{i'=1}^{3} q_{3i'}\xi_{i'}^N \tag{280}$$

$$\geq q_{11}\xi_1^N + q_{12}\xi_3^N + q_{13}\xi_3^N - q_{31}\xi_1^n - q_{32}\xi_1^N - q_{33}\xi_3^N \tag{281}$$

$$= (q_{11} - q_{31} - q_{32})\xi_1^N - (q_{33} - q_{12} - q_{13})\xi_3^N,\ N \geq N_0, \tag{282}$$

and $q_{11} - q_{31} - q_{32} = q_{33} - q_{12} - q_{13}$ holds by $q_{11} + q_{12} + q_{13} = q_{31} + q_{32} + q_{33} = 1$. Defining $K \equiv q_{11} - q_{31} - q_{32} = q_{33} - q_{12} - q_{13}$, we have $\tau_1^N - \tau_3^N \geq K(\xi_1^N - \xi_3^N)$. By the assumption (268), we have $q_{11} > 1/2, q_{12} + q_{13} < 1/2$, $q_{33} > 1/2, q_{31} + q_{32} < 1/2$, thus $K > 0$. ∎

Lemma 18: $\sum_{N=0}^{\infty}(\xi_1^N - \xi_3^N) < \infty$.

Proof: We have

$$\frac{\xi_1^N}{\xi_3^N} - \frac{\xi_1^{N+1}}{\xi_3^{N+1}} = \frac{\xi_1^N}{\xi_3^N} - \frac{\xi_1^N(1-\tau_1^N)}{\xi_3^N(1-\tau_3^N)} \tag{283}$$

$$= \frac{\xi_1^N}{\xi_3^N} \cdot \frac{\tau_1^N - \tau_3^N}{1 - \tau_3^N}. \tag{284}$$

By (269), we have $\xi_1^N/\xi_3^N \geq 1$, $N \geq N_0$, and by Lemma 17, $\tau_1^N - \tau_3^N \geq K(\xi_1^N - \xi_3^N)$, $K > 0$, $N \geq N_0$. Further, by Lemma 12, we have $0 < \tau_3^N \leq 1/2$, thus $1/(1-\tau_3^N) > 1$. Therefore, by (284),

$$\frac{\xi_1^{N_0}}{\xi_3^{N_0}} - \frac{\xi_1^{N+1}}{\xi_3^{N+1}} > K \sum_{l=N_0}^{N} (\xi_1^l - \xi_3^l),\ N \geq N_0, \tag{285}$$

hence $\sum_{l=N_0}^{N}(\xi_1^l - \xi_3^l)$ has an upper bound $K^{-1}(\xi_1^{N_0}/\xi_3^{N_0})$ which is unrelated to N. Then, the sum is convergent. ■

Here, we cite the following theorem.

Theorem B: ([5], p.31) Let $\{a_N\}_{N=0,1,\ldots,}$ be a decreasing positive sequence. If $\sum_{N=0}^{\infty} a_N$ is convergent, then $Na_N \to 0,\ N \to \infty$.

Lemma 19: The sequence $\{\xi_1^N - \xi_3^N\}$ is decreasing for $N \geq N_0$.

Proof: We have

$$\xi_1^{N+1} - \xi_3^{N+1} = \xi_1^N - \xi_1^N \sum_{i'=1}^{3} q_{1i'}\xi_{i'}^N - \xi_3^N + \xi_3^N \sum_{i'=1}^{3} q_{3i'}\xi_{i'}^N \tag{286}$$

$$\leq \xi_1^N - \xi_1^N \sum_{i'=1}^{3} q_{1i'}\xi_3^N - \xi_3^N + \xi_3^N \sum_{i'=1}^{3} q_{3i'}\xi_1^N \tag{287}$$

$$= \xi_1^N - \xi_1^N\xi_3^N - \xi_3^N + \xi_3^N\xi_1^N \tag{288}$$

$$= \xi_1^N - \xi_3^N,\ N \geq N_0, \tag{289}$$

hence the assertion holds. ■

Lemma 20: $\lim_{N\to\infty} N(\xi_1^N - \xi_3^N) = 0$.

Proof: The assertion holds by Lemmas 18, 19, and Theorem B. ■

Summarizing the above, we have

Theorem 10 (case of $m_2 = 3$):

$$\lim_{N\to\infty} N\xi_i^N = 1,\ i = 1, 2, 3.$$

Proof: By Lemmas 16 and 20, we have

$$\limsup_{N\to\infty} N\xi_1^N = \limsup_{N\to\infty} \left(N\xi_1^N - N\xi_3^N + N\xi_3^N\right) \tag{290}$$

$$\leq \limsup_{N\to\infty} N\left(\xi_1^N - \xi_3^N\right) + \limsup_{N\to\infty} N\xi_3^N \tag{291}$$

$$\leq 1, \tag{292}$$

thus, together with Lemma 15, we have

$$\lim_{N\to\infty} N\xi_1^N = 1. \tag{293}$$

Further, we have

$$\lim_{N\to\infty} N\xi_3^N = \lim_{N\to\infty} \left(N\xi_3^N - N\xi_1^N + N\xi_1^N\right) \tag{294}$$

$$= -\lim_{N\to\infty} N(\xi_1^N - \xi_3^N) + \lim_{N\to\infty} N\xi_1^N \tag{295}$$

$$= 1. \tag{296}$$

Finally, by (293), (296) and the squeeze theorem, we have

$$\lim_{N\to\infty} N\xi_2^N = 1. \tag{297}$$

■

Appendix F
Proof of Lemma 9

Let $0 = \theta_1 \leq \ldots \leq \theta_{m-1} \leq \theta_m < 1$ be the eigenvalues of $J(\boldsymbol{\lambda}^*)$. We have $\theta_{\max} = \theta_m$, $\theta_{\sec} = \theta_{m-1}$. Let $\boldsymbol{b}_i$, $i = 1, \ldots, m$, be the left eigenvector of $J(\boldsymbol{\lambda}^*)$ for θ_i, $i = 1, \ldots, m$. We have $\boldsymbol{b}_{\max} = \boldsymbol{b}_m$. Because of (75), we can assume that $\{\boldsymbol{b}_i\}_{i=1,\ldots,m}$ forms a basis of $\mathbb{R}^m$. Suppose $\boldsymbol{\mu}^N \in \boldsymbol{b}_{\max}^{\perp}$ for $N = 0, 1, \ldots$. Then, $\boldsymbol{\mu}^N$ is uniquely represented as

$$\boldsymbol{\mu}^N = \sum_{i=1}^{m-1} c_i^N \boldsymbol{b}_i,\ c_i^N \in \mathbb{R} \tag{298}$$

in the $m-1$ dimensional subspace $\boldsymbol{b}_{\max}^{\perp}$. By (298), we have

$$\boldsymbol{\mu}^{N+1} = \boldsymbol{\mu}^N J(\boldsymbol{\lambda}^*) \tag{299}$$

$$= \sum_{i=1}^{m-1} c_i^N \boldsymbol{b}_i J(\boldsymbol{\lambda}^*) \tag{300}$$

$$= \sum_{i=1}^{m-1} c_i^N \theta_i \boldsymbol{b}_i. \tag{301}$$

Comparing the coefficients of $\boldsymbol{\mu}^{N+1} = \sum_{i=1}^{m-1} c_i^{N+1}\boldsymbol{b}_i$ and (301), we have $c_i^{N+1} = \theta_i c_i^N = \ldots = (\theta_i)^{N+1} c_i^0$, $i = 1, \ldots, m-1$, thus $\boldsymbol{\mu}^N = \sum_{i=1}^{m-1} (\theta_i)^N c_i^0 \boldsymbol{b}_i$. Therefore,

$$\|\boldsymbol{\mu}^N\| \leq \sum_{i=1}^{m-1} (\theta_i)^N |c_i^0| \|\boldsymbol{b}_i\| \tag{302}$$

$$\leq K \cdot (\theta_{m-1})^N,\ K > 0 \tag{303}$$

$$= K \cdot (\theta_{\sec})^N. \tag{304}$$

Appendix G
Proof of Lemma 10

Suppose that ${}^t\boldsymbol{b}_{\max}$ is a right eigenvector for $\theta_{\max}$. For any $\boldsymbol{\mu} \in \boldsymbol{b}_{\max}^{\perp}$, $\boldsymbol{\mu}J(\boldsymbol{\lambda}^*){}^t\boldsymbol{b}_{\max} = \theta_{\max}\boldsymbol{\mu}{}^t\boldsymbol{b}_{\max} = 0$ holds. Thus, we obtain $\boldsymbol{\mu}J(\boldsymbol{\lambda}^*) \in \boldsymbol{b}_{\max}^{\perp}$.

Conversely, suppose $\boldsymbol{\mu}J(\boldsymbol{\lambda}^*) \in \boldsymbol{b}_{\max}^{\perp}$ for any $\boldsymbol{\mu} \in \boldsymbol{b}_{\max}^{\perp}$. Our goal is to show $J(\boldsymbol{\lambda}^*){}^t\boldsymbol{b}_{\max} = \theta_{\max}{}^t\boldsymbol{b}_{\max}$, which is equivalent to

$$\boldsymbol{\mu}J(\boldsymbol{\lambda}^*){}^t\boldsymbol{b}_{\max} = \theta_{\max}\boldsymbol{\mu}{}^t\boldsymbol{b}_{\max} \text{ holds for any } \boldsymbol{\mu}. \tag{305}$$

We will prove (305). Since we can write $\boldsymbol{\mu}$ uniquely as $\boldsymbol{\mu} = K\boldsymbol{b}_{\max} + \tilde{\boldsymbol{\mu}}$ with $K \in \mathbb{R}$ and $\tilde{\boldsymbol{\mu}} \in \boldsymbol{b}_{\max}^{\perp}$, we have

$$\boldsymbol{\mu}J(\boldsymbol{\lambda}^*){}^t\boldsymbol{b}_{\max} = K\boldsymbol{b}_{\max}J(\boldsymbol{\lambda}^*){}^t\boldsymbol{b}_{\max} + \tilde{\boldsymbol{\mu}}J(\boldsymbol{\lambda}^*){}^t\boldsymbol{b}_{\max} \tag{306}$$

$$= K\theta_{\max}\boldsymbol{b}_{\max}{}^t\boldsymbol{b}_{\max} + 0 \ \text{ (by the assumption)} \tag{307}$$

$$= \theta_{\max}K\boldsymbol{b}_{\max}{}^t\boldsymbol{b}_{\max} + \theta_{\max}\tilde{\boldsymbol{\mu}}{}^t\boldsymbol{b}_{\max} \ \text{ (by } \tilde{\boldsymbol{\mu}} \in \boldsymbol{b}_{\max}^{\perp}\text{)} \tag{308}$$

$$= \theta_{\max}\boldsymbol{\mu}{}^t\boldsymbol{b}_{\max}, \tag{309}$$

which proves (305).

Appendix H
Proof of Theorem 12

Define $Q^N \equiv \boldsymbol{\lambda}^N \Phi$. Noting that $\sum_{i=1}^m \lambda_i^* D(P^i \| Q^*) = C$ holds by the Kuhn-Tucker condition (4), we have

$$0 < C - I(\boldsymbol{\lambda}^N, \Phi) = C - \sum_{i=1}^m \lambda_i^N \sum_{j=1}^n P_j^i \log \left(\frac{P_j^i}{Q_j^*} \cdot \frac{Q_j^*}{Q_j^N} \right) \tag{310}$$

$$= C - \sum_{i=1}^m \left(\lambda_i^* + \mu_i^N \right) D(P^i \| Q^*) + D(Q^N \| Q^*) \tag{311}$$

$$= - \sum_{i=1}^m \mu_i^N D(P^i \| Q^*) + D(Q^N \| Q^*). \tag{312}$$

We will evaluate $D(Q^N \| Q^*)$ in (312). Defining $R^N \equiv Q^N - Q^*$ with $R_j^N = Q_j^N - Q_j^*$, $j = 1, \ldots, n$, we have

$$D(Q^N \| Q^*) = \sum_{j=1}^n \left(Q_j^* + R_j^N \right) \log \left(1 + \frac{R_j^N}{Q_j^*} \right) \tag{313}$$

$$= \sum_{j=1}^n \left(Q_j^* + R_j^N \right) \left\{ \frac{R_j^N}{Q_j^*} - \frac{1}{2} \left(\frac{R_j^N}{Q_j^*} \right)^2 + o\left(\| R^N \|^2 \right) \right\} \tag{314}$$

$$= \sum_{j=1}^n R_j^N + \sum_{j=1}^n \frac{\left(R_j^N \right)^2}{Q_j^*} - \frac{1}{2} \sum_{j=1}^n \frac{\left(R_j^N \right)^2}{Q_j^*} + o\left(\| R^N \|^2 \right) \tag{315}$$

$$= 0 + \frac{1}{2} \sum_{j=1}^n \frac{1}{Q_j^*} \left(\sum_{i=1}^m \mu_i^N P_j^i \right)^2 + o\left(\| \boldsymbol{\mu}^N \|^2 \right). \tag{316}$$

Let $\mathcal{I}_{\mathrm{III}} = \emptyset$. Then, we have $D(P^i \| Q^*) = C$, $i = 1, \ldots, m$. By $\sum_{i=1}^m \mu_i^N = 0$, the first term of (312) is 0. Therefore, if $\| \boldsymbol{\mu}^N \| < K_1 \cdot (\theta)^N$, then we obtain (203) by (316). Meanwhile, if $\lim_{N \to \infty} N \mu_i^N = \sigma_i$, then we obtain (204) by (316).

Next, let $\mathcal{I}_{\mathrm{III}} \neq \emptyset$. By (312), we have

$$C - I(\boldsymbol{\lambda}^N, \Phi) \leq \sum_{i=1}^m |\mu_i^N| D(P^i \| Q^*) + O\left(\| \boldsymbol{\mu}^N \|^2 \right). \tag{317}$$

Therefore, if $\| \boldsymbol{\mu}^N \| < K_1 \cdot (\theta)^N$, then we obtain (205) by (317). Meanwhile, if $\lim_{N \to \infty} N \mu_i^N = \sigma_i$, then we obtain (206) by (317).

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17K00008.

Kenji Nakagawa (M'85) received the B.S., M.S. and D.S. degrees from Tokyo Institute of Technology, Tokyo, Japan, 1980, 1982 and 1986, respectively. In 1985, he joined NTT (Nippon Telegraph and Telephone Corp.). Since 1992, he has been an associate professor of Nagaoka University of Technology, and from 2012 he is a professor. His research interests include information theory, performance evaluation of network, and queueing theory. He is a member of IEEE, IEICE, Operations Research Society of Japan, Mathematical Society of Japan and Japanese Society of Engineering Education.

Yoshinori Takei (A'01–M'02) was born in Tokyo, Japan, in 1965. He received the B.Sc. and M.Sc. degrees in mathematics in 1990 and 1992, respectively, from Tokyo Institute of Technology and the Dr. Eng. degree in information processing in 2000, from Tokyo Institute of Technology, Yokohama, Japan. From 1992 to 1995, he was with Kawasaki Steel System R&D Inc., Chiba, Japan. From 1999 to 2000, he was an Assistant Professor with the Department of Electrical and Electronic Engineering, Tokyo Institute of Technology. From 2000 to 2017, he was with the Department of Electrical Engineering, Nagaoka University of Technology, Niigata, Japan, where he was finally an Associate Professor. From 2017, he has been with the Department of Electrical and Information Engineering, National Institute of Technology, Akita College, where he is a Professor. His current research interests include computational complexity theory, public-key cryptography, randomized algorithms, and digital signal processing. Especially, he is now interested in the application of non-commutative harmonic analysis to data-mining. Dr. Takei is a member of IEEE, LA, SIAM, ACM, AMS, IPSJ and IEICE.

Shin-ichiro Hara received the B.S., M.S. and D.S. degrees from Kyoto University, Kyoto, Japan, 1983, 1985 and 1990, respectively. Since 1992, he has been an associate professor of Nagaoka University of Technology, and from 2013 he is a professor. His research interest is algebraic topology, especially simplicial objects of various categories and homotopy theories of them. He is a member of Mathematical Society of Japan.

Kohei Watabe (S'11–M'14) received his B.E. and M.E. degrees in Engineering from Tokyo Metropolitan University, Tokyo, Japan, in 2009 and 2011, respectively. He also received the Ph.D. degree from Osaka University, Japan, in 2014. He was a JSPS research fellow (DC2) from April 2012 to March 2014. He was an Assistant Professor of the Graduate School of Engineering, Nagaoka University of Technology, from April 2014 to October 2019. He has been an Associate Professor of the Graduate School of Engineering, Nagaoka University of Technology, since November 2019. He is a member of the IEEE and the IEICE.