

# Arimoto-Blahut アルゴリズムにおける速い収束と遅い収束

## Fast and Slow Convergence in the Arimoto-Blahut Algorithm

中川健治

長岡技術科学大学

2023年3月10日 電子情報通信学会総合  
大会情報理論研究専門委員会企画 チュートリアルセッション

# 研究の目的

離散的無記憶通信路の通信路容量を計算するための Arimoto-Blahut アルゴリズムの速い収束と遅い収束について調べる

- 速い収束：指数的収束
- 遅い収束：オーダー  $O(1/N)$  の収束

指数的収束については多くの研究がある。

# 数列の収束速度について

一般に、0 に収束する実数列  $\{\mu^N\}_{N=0,1,\dots}$  に対して、

- 指数的収束

$$|\mu^N| < K \cdot (\theta)^N, \quad K > 0, \quad 0 < \theta < 1$$

- オーダー  $O(1/N)$  の収束

$$\mu^N \simeq \frac{K}{N}, \quad K \neq 0, \quad N \rightarrow \infty$$

$$\text{(i.e., } N\mu^N \rightarrow K)$$

# 定義

離散的無記憶通信路 :  $X \longrightarrow Y$

入力アルファベット :  $\mathcal{X} = \{x_1, \dots, x_m\}$

出力アルファベット :  $\mathcal{Y} = \{y_1, \dots, y_n\}$

入力分布全体の集合 :

$$\Delta(\mathcal{X}) = \{\lambda = (\lambda_1, \dots, \lambda_m) \mid \lambda_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \lambda_i = 1\}$$

出力分布全体の集合 :

$$\Delta(\mathcal{Y}) = \{Q = (Q_1, \dots, Q_n) \mid Q_j \geq 0, j = 1, \dots, n, \sum_{j=1}^n Q_j = 1\}$$

# 定義

通信路の条件付き確率：

$$P_j^i = P(Y = y_j | X = x_i), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

行ベクトル：

$$P^i = (P_1^i, \dots, P_n^i), \quad i = 1, \dots, m$$

通信路行列：

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

# 定義

入力分布  $\lambda \in \Delta(\mathcal{X})$  に対する出力分布  $Q \in \Delta(\mathcal{Y})$  :

$$Q = \lambda\Phi, \quad Q_j = \sum_{i=1}^m \lambda_i P_j^i, \quad j = 1, \dots, n$$

相互情報量 :

$$I(\lambda, \Phi) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i P_j^i \log \frac{P_j^i}{Q_j}$$

通信路容量 :

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

# 定義

Kullback-Leibler ダイバージェンス :

$$D(Q \| Q') = \sum_{j=1}^n Q_j \log \frac{Q_j}{Q'_j} \text{ for } Q, Q' \in \Delta(\mathcal{Y})$$

# Kuhn-Tucker 条件

入力分布  $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \Delta(\mathcal{X})$  が通信路容量  $C$  を達成するための必要十分条件は,

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



## Arimoto-Blahut アルゴリズム

入力分布の系列  $\{\lambda^N = (\lambda_1^N, \dots, \lambda_m^N)\}_{N=0,1,\dots}$  が次の漸化式で与えられる：

$$\lambda_i^{N+1} = \frac{\lambda_i^N \exp D(P^i \| \lambda^N \Phi)}{\sum_{k=1}^m \lambda_k^N \exp D(P^k \| \lambda^N \Phi)}, \quad N = 0, 1, \dots, i = 1, \dots, m.$$

初期分布： $\lambda^0 = (\lambda_1^0, \dots, \lambda_m^0)$ ,  $\lambda_i^0 > 0$ ,  $i = 1, \dots, m$ .

通信路容量を達成する分布  $\lambda^*$  が唯一とする。 $N \rightarrow \infty$  のとき  $\lambda^N$  は  $\lambda^*$  に収束する. i.e.,  $\lambda^N \rightarrow \lambda^*$ ,  $N \rightarrow \infty$ . Arimoto[1972]

## Arimoto-Blahut アルゴリズムの定義関数

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

$$F(\boldsymbol{\lambda}) \equiv (F_1(\boldsymbol{\lambda}), \dots, F_m(\boldsymbol{\lambda})) : \Delta(\mathcal{X}) \rightarrow \Delta(\mathcal{X})$$

Arimoto-Blahut アルゴリズムは  $\boldsymbol{\lambda}^{N+1} = F(\boldsymbol{\lambda}^N)$  と表せる。

$F(\boldsymbol{\lambda})$  の  $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$  における Taylor 展開に基づいて収束速度を調べる。

rank  $\Phi = m$  を仮定する

## $F(\lambda)$ の $\lambda = \lambda^*$ における Taylor 展開

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

$J(\lambda^*)$  : Jacobi 行列

$H(\lambda^*) = (H_1(\lambda^*), \dots, H_m(\lambda^*))$ , ここで  $H_i(\lambda^*)$  は  $F_i$  の  $\lambda^*$  における Hesse 行列

$\lambda = \lambda^N$ ,  $\mu^N \equiv \lambda^N - \lambda^*$  とおき, 2 次項までを考える :

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

$\mu^N$  の成分を  $\mu^N = (\mu_1^N, \dots, \mu_m^N)$  と書く。

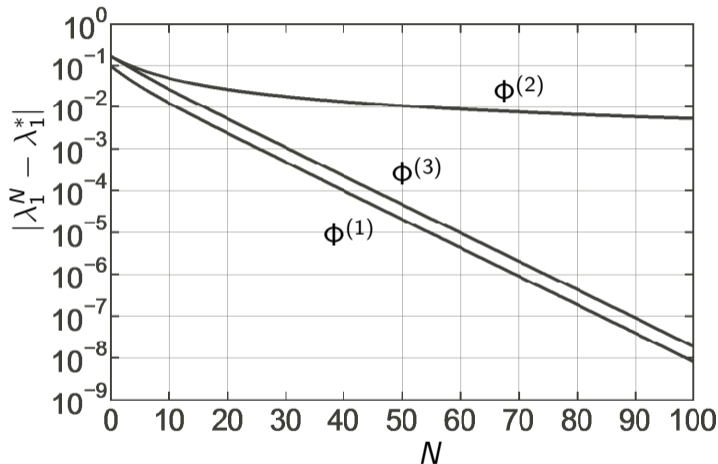
# 例題

$$\Phi^{(1)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}, \lambda^* = (0.431, 0.431, 0.138)$$

$$\Phi^{(2)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{pmatrix}, \lambda^* = (0.500, 0.500, 0.000)$$

$$\Phi^{(3)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{pmatrix}, \lambda^* = (0.500, 0.500, 0.000)$$

# $\phi^{(1)}$ , $\phi^{(2)}$ , $\phi^{(3)}$ に対する $|\lambda_1^N - \lambda_1^*|$ の収束速度の比較



多くの通信路行列  $\phi$  で収束は指数的. ある特別な  $\phi$  では収束は遅い.

## 添字 $i$ のタイプ

Kuhn-Tucker 条件をさらに分類して，添字  $i$  を 3 つのタイプに分ける．

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

任意の通信路行列  $\Phi$  に対してタイプ I の添字は存在する． $\Phi$  によってタイプ II, III の添字は存在する場合もしない場合もある．

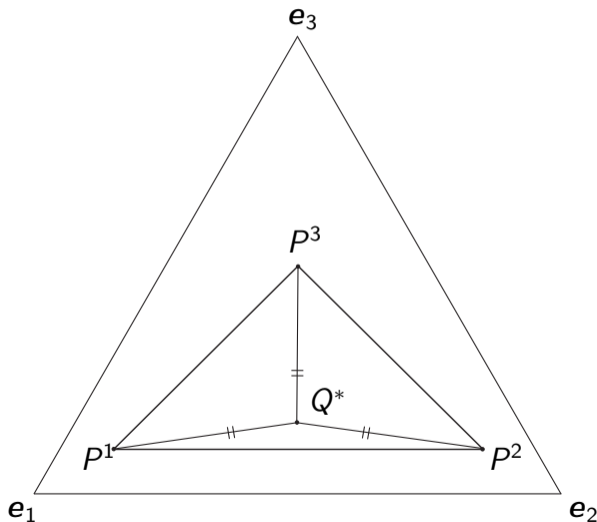
$$\mathcal{I}_I \equiv \{ \text{タイプ I の添字} \}, \mathcal{I}_{II} \equiv \{ \text{タイプ II の添字} \}, \\ \mathcal{I}_{III} \equiv \{ \text{タイプ III の添字} \}$$

## 例題における添字のタイプについて

通信路 行列	タイプ I の 添字 $\mathcal{I}_I$	タイプ II の 添字 $\mathcal{I}_{II}$	タイプ III の 添字 $\mathcal{I}_{III}$
$\Phi^{(1)}$	1,2,3	-	-
$\Phi^{(2)}$	1,2	3	-
$\Phi^{(3)}$	1,2	-	3

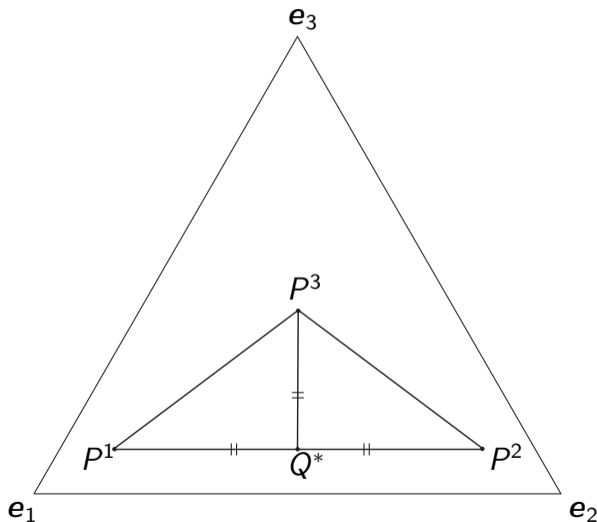
$\Phi^{(2)}$  においてタイプ II の添字が存在する。タイプ II の添字が存在するとき収束が遅い。

# $\Phi^{(1)}$ における $P^1, P^2, P^3$ と $Q^* = \lambda^* \Phi$ の位置関係

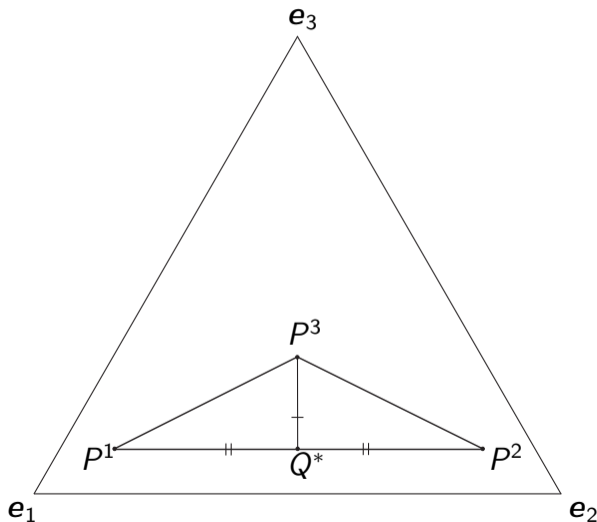




# $\Phi^{(2)}$ における $P^1, P^2, P^3$ と $Q^* = \lambda^* \Phi$ の位置関係



# $\Phi^{(3)}$ における $P^1, P^2, P^3$ と $Q^* = \lambda^* \Phi$ の位置関係



## Jacobi 行列の構造

$$J(\lambda^*) \equiv \left( \frac{\partial F_i}{\partial \lambda_{i'}} \bigg|_{\lambda=\lambda^*} \right) = \begin{pmatrix} J^I & 0 & 0 \\ * & J^{II} & 0 \\ * & 0 & J^{III} \end{pmatrix}$$

	対応する 添字	固有値 $\theta_i$	行列の 特徴
$J^I$	タイプ I	$0 \leq \theta_i < 1$	$I - J^I$ は正定値 対称行列に相似
$J^{II}$	タイプ II	$\theta_i = 1$	単位行列
$J^{III}$	タイプ III	$0 < \theta_i < 1$	対角行列

## タイプIIの添字の存在と収束速度の関係

$$\text{2次漸化式: } \mu^{N+1} = \mu^N J(\lambda^*) + \frac{1}{2!} \mu^N H(\lambda^*)^t \mu^N$$

$J(\lambda^*)$  の最大固有値を  $\theta_{\max}$  とする.

タイプIIの添字が存在しないとき

$0 < \theta_{\max} < 1$  となるので  $\mu^N \rightarrow \mathbf{0}$  は指数的収束である.

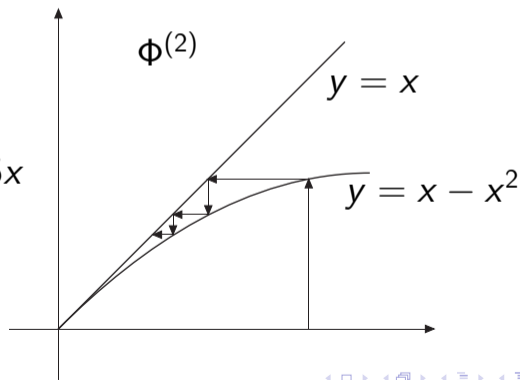
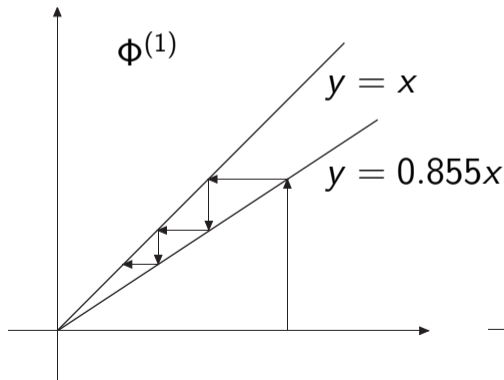
タイプIIの添字が存在するとき

$\theta_{\max} = 1$  となるので  $\mu^N \rightarrow \mathbf{0}$  の収束は  $J(\lambda^*)$  だけでは決まらず、Hesse 行列  $H(\lambda^*)$  も調べる必要がある.

## 例題 $\phi^{(1)}$ , $\phi^{(2)}$ の収束速度について

$\phi^{(1)}$  :  $\theta_{\max} = 0.855$ ,  $\|\mu^{N+1}\| = 0.855\|\mu^N\| + o(\|\mu^N\|)$ , 指数的収束

$\phi^{(2)}$  :  $\theta_{\max} = 1$ ,  $\mu_3^{N+1} = \mu_3^N - (\mu_3^N)^2$ , オーダー  $O(1/N)$  の収束



## タイプIIの添字が存在するときの Hesse 行列

タイプIIの添字が存在するとき，Hesse 行列は特別な形になる

$$H_i(\lambda^*) = \begin{pmatrix} 0 & \cdots & 0 & D_{i,m'+1}^* & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & D_{i,i-1}^* & 0 & \cdots & 0 \\ D_{i,m'+1}^* & \cdots & D_{i,i-1}^* & 2D_{i,i}^* & D_{i,i+1}^* & \cdots & D_{i,m}^* \\ 0 & \cdots & 0 & D_{i,i+1}^* & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & D_{i,m}^* & 0 & \cdots & 0 \end{pmatrix} \quad (1)$$

## 2次漸化式の形

定理 (Nakagawa [6])

タイプIIの添字が存在するとき (すなわち  $\mathcal{I}_{\text{II}} \neq \emptyset$  のとき),  
2次漸化式は

$$\mu_i^{N+1} = \mu_i^N + \mu_i^N \sum_{i' \in \mathcal{I}_{\text{II}}} r_{ii'} \mu_{i'}^N, \quad i \in \mathcal{I}_{\text{II}}$$

の形となる. ここで,  $r_{ii'}$  は  $J(\lambda^*)$ ,  $H(\lambda^*)$  で決まる実数である.

また,  $\{\mu_i^N\}, i \in \mathcal{I}_{\text{I}} \cup \mathcal{I}_{\text{III}}$  は  $\{\mu_i^N\}, i \in \mathcal{I}_{\text{II}}$  の線形結合で表される.

## 漸化式に関する仮定

$R \equiv (r_{ii'}), i, i' \in \mathcal{I}_{\text{II}}, \mathbf{1} \equiv (1, \dots, 1) \in \mathbb{R}^{m_2}, (m_2 = |\mathcal{I}_{\text{II}}|)$  として,

- (i)  $r_{ii'} \leq 0, i, i' \in \mathcal{I}_{\text{II}}$
- (ii)  $R$  は正則行列
- (iii)  $\boldsymbol{\sigma} \equiv -\mathbf{1} ({}^t R)^{-1}, \boldsymbol{\sigma} = (\sigma_i), i \in \mathcal{I}_{\text{II}}$  とおくととき,  $\sigma_i > 0, i \in \mathcal{I}_{\text{II}}$
- (iv)  $-r_{ii}\sigma_i > 1/2, i \in \mathcal{I}_{\text{II}}$

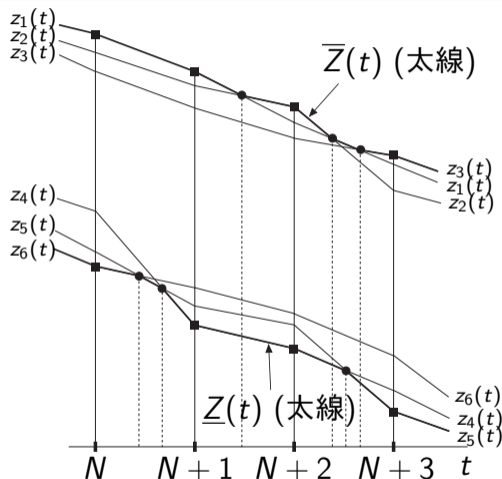
上記の仮定 (i)-(iv) のもとで,

$$\lim_{N \rightarrow \infty} N \mu_i^N = \sigma_i, i \in \mathcal{I}_{\text{II}}. \quad (2)$$

証明のポイントは、漸化式の値を線分で結んで補間し、連続的な折れ線として解析する。



# 漸化式で定義される系列を線分で結んで折れ線とする



連立微分方程式の解の存在を証明する Cauchy の折れ線を参考にした。

# 数値的評価

- 指数的収束の場合の評価尺度

$$L(N) = -\frac{1}{N} \log \|\mu^N\|$$

- $1/N$  オーダーの収束の場合の評価尺度

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

## $\Phi^{(1)}$ における指数的収束 (タイプ I の添字のみ)

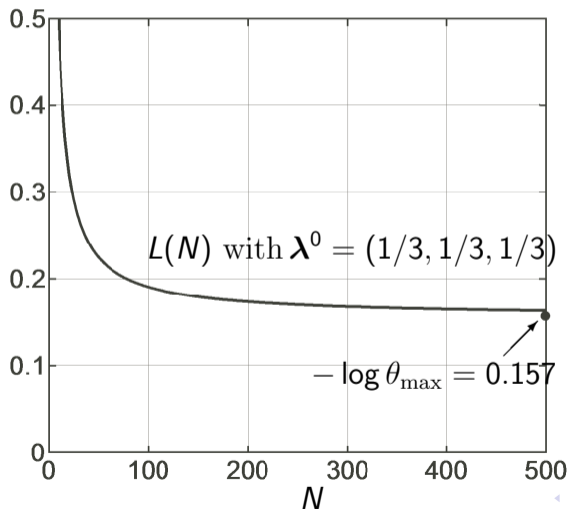
$$\Phi^{(1)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}$$

$$\lambda^* = (0.431, 0.431, 0.138)$$

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

$$\theta_{\max} = 0.855$$

# $\Phi^{(1)}$ における $L(N)$ の収束



## $\Phi^{(2)}$ におけるオーダー $O(1/N)$ の収束 (タイプ I と II の添字が存在)

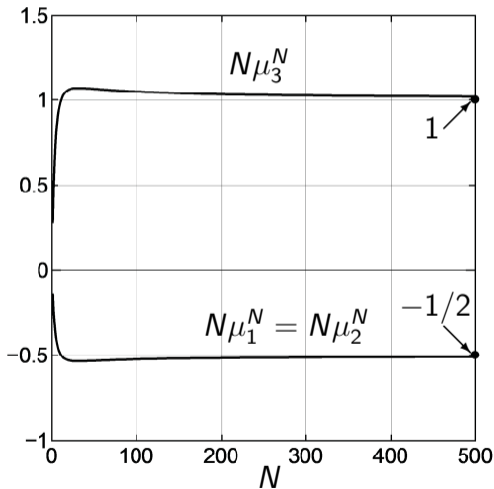
$$\Phi^{(2)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{pmatrix}, \quad \lambda^* = (0.500, 0.500, 0.000)$$

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

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

$$(\sigma_1, \sigma_2, \sigma_3) = (-1/2, -1/2, 1)$$

## $\Phi^{(2)}$ における $N\mu_i^N$ の収束



## $\Phi^{(3)}$ における指数的収束 (タイプ I と III の添字が存在)

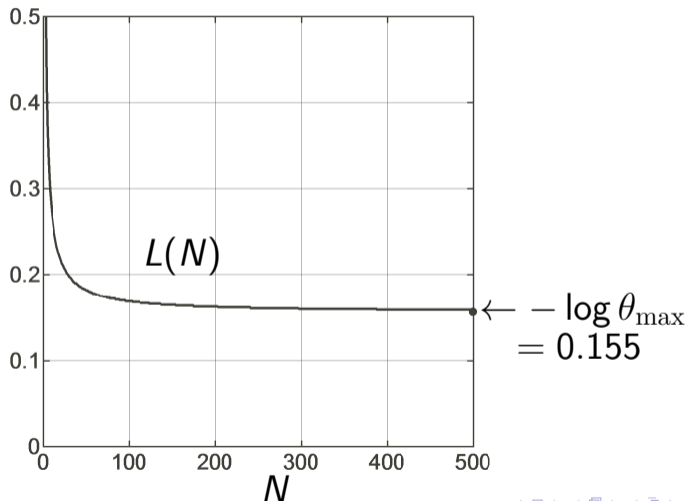
$$\Phi^{(3)} = \begin{pmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{pmatrix}$$

$$\lambda^* = (0.500, 0.500, 0.000)$$

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

$$\theta_{\max} = 0.856$$

## $\Phi^{(3)}$ における $L(N)$ の収束





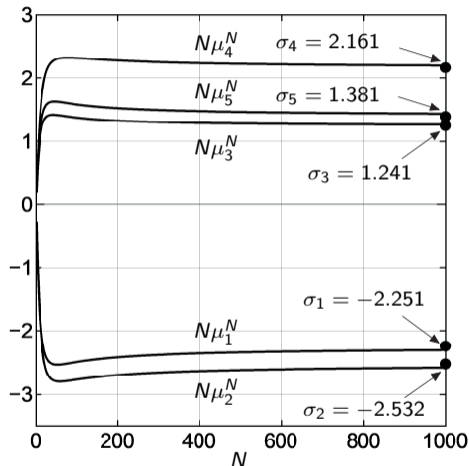
## オーダー $O(1/N)$ の収束の例

$$\Phi = \begin{pmatrix} 0.600 & 0.100 & 0.100 & 0.100 & 0.100 \\ 0.100 & 0.600 & 0.100 & 0.100 & 0.100 \\ 0.231 & 0.231 & 0.066 & 0.179 & 0.292 \\ 0.161 & 0.341 & 0.226 & 0.226 & 0.046 \\ 0.341 & 0.161 & 0.226 & 0.046 & 0.226 \end{pmatrix}$$

$$R = \begin{pmatrix} -0.522 & -0.020 & -0.223 \\ -0.020 & -0.401 & -0.078 \\ -0.223 & -0.078 & -0.401 \end{pmatrix}$$

$$\sigma = (-2.251, -2.532, 1.241, 2.161, 1.381)$$

# $N\mu_i^N$ と $\sigma_i$ の比較



$N = 1000$  で  $N\mu_i^N$  は  $\sigma_i^{(1)}$  に十分近い値に収束している。

# 相互情報量の $C$ への収束速度

定理 (Nakagawa[2021])

(1) タイプ III の添字が存在しないとき

(1-1) タイプ II の添字が存在しないとき

$$0 < C - I(\lambda^*, \Phi) < K \cdot (\theta)^{2N}$$

(1-2) タイプ II の添字が存在するとき

$$\lim_{N \rightarrow \infty} N^2 (C - I(\lambda^N, \Phi)) = \frac{1}{2} \sum_{j=1}^n \frac{1}{Q_j^*} \left( \sum_{i=1}^m \sigma_i P_j^i \right)^2 \cdot \dots \cdot (*)$$

(2) タイプ III の添字が存在するとき

(2-1) タイプ II の添字が存在しないとき

$$0 < C - I(\lambda^*, \Phi) < K \cdot (\theta)^N$$

(2-2) タイプ II の添字が存在するとき

$$0 < C - I(\lambda^*, \Phi) < K/N$$

Convergence speed of  $I(\lambda^N, \Phi^{(1)}) \rightarrow C$

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

Convergence speed of  $I(\lambda^N, \Phi^{(2)}) \rightarrow C$

$N^2 (C - I(\lambda^N, \Phi^{(2)})) \big _{N=500}$	0.516
Eq. (*)	0.500

Convergence speed of  $I(\lambda^N, \Phi^{(3)}) \rightarrow C$

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

## まとめと今後の課題

### まとめ

Arimoto-Blahut アルゴリズムの速い収束である指数的収束と遅い収束であるオーダー  $O(1/N)$  の収束について，定義関数の Taylor 展開を使って調べた。

### 今後の課題

Arimoto-Blahut アルゴリズムは EM アルゴリズムとしても定式化できる．一般の EM アルゴリズムでも遅い収束はあるのか？