



Tokyo Tech

Arimotoの指数計算アルゴリズム

2023.3.10(Fri) 午後

東京工業大学 工学院

實松 豊

● 流れ

- Arimotoの指数計算アルゴリズムを紹介する
- 誤り指数と正復号指数
- アルゴリズムの導出, Arimoto-Blahutとの対比
- アルゴリズムによって導出される指数の別表現
- Oohama-Jitsumatsu (isit2015) の正復号指数を求めるアルゴリズムを紹介
- (時間があれば) CVXPY

本発表の目的:

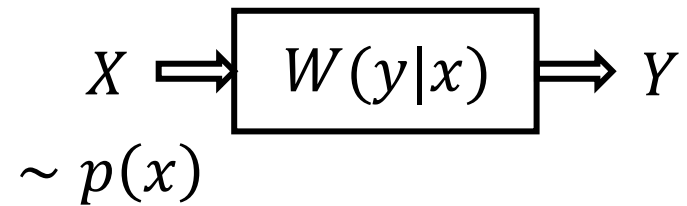
Arimotoの指数計算アルゴリズムの見事さを伝えること.

Arimotoのアルゴリズムの紹介

- 強逆定理と指数計算アルゴリズム

- Arimoto-Blahutアルゴリズム 1972年
- Arimotoの強逆定理（正復号指数）1973年
- Arimotoの指数計算アルゴリズム 1976年

- 離散無記憶通信路を対象とする.



- p に関する最大化・最小化は X のアルファベット上の確率分布全体に対して行う.

- 原稿で記述したコスト制約付き通信路符号化への拡張は、省略する.
- Arimoto1976は、有歪み情報源符号化に対する計算アルゴリズムも提案されているが、時間の都合により省略する.

$p(x)$ 入力分布,
 $W(y|x)$ 通信路の遷移確率行列

● Gallagerの誤り指数(1965)

$$P_e \leq \exp[-n\{-\rho R + \max_p E_0(\rho, p)\}], \quad 0 \leq \rho \leq 1$$

$$E_0(\rho, p) = -\log \left[\sum_y \left\{ \sum_x p(x) W(y|x)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right] \quad \dots \text{Gallager関数}$$

● Arimotoの強逆定理 (1973)

$$P_e \geq 1 - \exp[-n\{-\rho R + \min_p E_0(\rho, p)\}], \quad -1 \leq \rho \leq 0$$



$$E_r(R|W) = \max_{\rho \in [0,1]} \left\{ -\rho R + \max_p E_0(\rho, p) \right\}$$

$$G_{AR}(R|W) = \max_{\rho \in [-1,0]} \left\{ -\rho R + \min_p E_0(\rho, p) \right\}$$

注意) $E_0(\rho, p)$ の凸性

- Csiszár-Körnerの誤り指数(C.-K.の教科書1981, Csiszár-Körner-Martón, 1977)

$$E_{\text{CK}}(R|W) = \max_{q_X \in \mathcal{P}(\mathcal{X})} \min_{q_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(q_{Y|X}||W|q_X) + |I(q_X, q_{Y|X}) - R|^+\}$$

- Dueck-Körnerの正復号指数(1979)

$$G_{\text{DK}}(R|W) = \min_{q_X \in \mathcal{P}(\mathcal{X})} \min_{q_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(q_{Y|X}||W|q_X) + |R - I(q_X, q_{Y|X})|^+\}$$

- CsiszárとKörnerの誤り指数及びDueck-Körnerの正復号指数は情報理論的な量で表現される
- Gallagerの誤り指数及びArimotoの正復号指数は、パラメトリックな表現（Arimotoアルゴリズムは、こちらに基づく）

(解説)

$$E_0(\rho, p) = -\log \left[\sum_y \left\{ \sum_x p(x) W(y|x)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right]$$

- 通信路容量に対するArimto-Blahutアルゴリズム

$$C(W) = \max_{p(x)} I(p, W)$$



$$C(W) = \max_p \max_{q_{X|Y}} \sum_{x,y} p(x) W(y|x) \log \frac{q_{X|Y}(x|y)}{p(x)}$$

元の最大化問題と答えが一致する2重最大化問題を見つける。

KLダイバージェンスの性質を利用

(補足) 1972年当時, Divergenceの用語は使用されていない

- エラー指数・正復号指数に対するアルゴリズム

$$C_\rho(W) = \max_p \frac{1}{\rho} E_0(\rho, p)$$



?

高い難易度: パラメータ ρ のため, Rényi エントロピーのような形式になっている. 左の話の類推から, すんなり出てくるものではない. 深い洞察力を要する.

$$\frac{1}{\rho} E_0(\rho, p) \rightarrow I(p, W) \quad (\rho \rightarrow 0)$$

- Arimoto(1976)は, 次の関数を導入した

$$F^{(\rho)}(p, q_{X|Y}) = -\frac{1}{\rho} \log \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p^{1+\rho}(x) q_{X|Y}^{-\rho}(x|y) W(y|x)$$

定理1(Arimoto 1976)

固定された $\rho \in (-1, \infty)$, $p(x)$ に対し,
 $F^{(\rho)}(p, q_{X|Y})$ を最大にする $q_{X|Y}$ は

$$q_{X|Y}(x|y) = \frac{p(x)W(y|x)^{\frac{1}{1+\rho}}}{\sum_{x'} p(x')W(y|x')^{\frac{1}{1+\rho}}}$$

で与えられ, 次が成り立つ.

$$\max_{q_{X|Y}} F^{(\rho)}(p, q_{X|Y}) = \frac{1}{\rho} E_0(\rho, p)$$

一方, 固定された $\rho \in (-1, \infty)$ と $q_{X|Y}$ に対し,
 $F^{(\rho)}(p, q_{X|Y})$ を最大にする p は

$$p(x) = \frac{\{\sum_y q_{X|Y}^{-\rho}(x|y)W(y|x)\}^{-1/\rho}}{\sum_{x'} \{\sum_y q_{X|Y}^{-\rho}(x'|y)W(y|x')\}^{-1/\rho}}$$

で与えられ, 次が成り立つ.

$$\max_p F^{(\rho)}(p, q_{X|Y}) = \log \sum_x \left[\sum_y q_{X|Y}^{-\rho}(x|y)W(y|x) \right]^{-1/\rho}$$

$$E_0(\rho, p) = -\log \left[\sum_y \left\{ \sum_x p(x) W(y|x)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right]$$

- 次の関数を定義する

$$A(\rho, q_{X|Y}) = \rho \log \sum_x \left[\sum_y q_{X|Y}^{-\rho}(x|y) W(y|x) \right]^{-1/\rho}$$

- Arimotoアルゴリズムが収束することより以下が成立する.

$$\max_p \frac{1}{\rho} E_0(\rho, p) = \max_p \max_{q_{X|Y}} F^{(\rho)}(p, q_{X|Y}) = \max_{q_{X|Y}} \frac{1}{\rho} A(\rho, q_{X|Y})$$

これより, 誤り指数と正復号指数の別表現が得られる.

$$E_r(R|W) := \max_{\rho \in [0,1]} \left\{ -\rho R + \max_p E_0(\rho, p) \right\} = \max_{\rho \in [0,1]} \left\{ -\rho R + \max_{q_{X|Y}} A(\rho, q_{X|Y}) \right\} =: \hat{E}_r(R|W)$$

$$G_{AR}(R|W) := \max_{\rho \in [-1,0]} \left\{ -\rho R + \min_p E_0(\rho, p) \right\} = \max_{\rho \in [-1,0]} \left\{ -\rho R + \min_{q_{X|Y}} A(\rho, q_{X|Y}) \right\} =: \hat{G}_{AR}(R|W)$$

別表現 → あとで使う.

Algorithm 1 Arimoto's algorithm for error and strong converse exponents in channel coding

Require: The conditional probability of the channel W , $\rho \in (-1, 0) \cup (0, \infty)$ and $\nu \geq 0$. Choose initial $p^{[0]}(x)$ such that all components are nonzero.

for $i = 0, 1, 2, \dots$ **do**

$$q_{X|Y}^{[i]}(x|y) = \frac{p^{[i]}(x)W(y|x)^{\frac{1}{1+\rho}}}{\sum_{x'} p^{[i]}(x')W(y|x')^{\frac{1}{1+\rho}}},$$
$$p^{[i+1]}(x) = \frac{\left\{ \sum_y q_{X|Y}^{[i]}(x|y)^{-\rho} W(y|x) \right\}^{-1/\rho}}{\sum_{x'} \left\{ \sum_y q_{X|Y}^{[i]}(x'|y)^{-\rho} W(y|x') \right\}^{-1/\rho}}.$$

end for

- $\rho \rightarrow 0$ の極限でABアルゴリズムに一致する。
- ρ が正でも負でも同じ更新式が通用する（驚き）。

Oohama-Jitsumatsuのアルゴリズム (概要)

- エラー指数と正復号指数には, タイプの理論により導出されるパラメトリックでない表現がある.
- Oohama-J.のアルゴリズムは, Dueck – Körnerの正復号指数 (パラメトリックでない正復号指数の表現) に基づくアルゴリズムである.

$$G_{\text{DK}}(R|W) = \min_{q_X \in \mathcal{P}(\mathcal{X})} \min_{q_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(q_{Y|X} || W|q_X) + |R - I(q_X, q_{Y|X})|^+\}$$

- コスト制約なし(Oohama-J., isit2015),
- コスト制約ありへの拡張(J.-Oohama, IT Trans.2020),
- 第3のアルゴリズム(J.-Oohama, arXiv:2205.06027, submitted to IT Trans.) を提案.
- 長所 : Dueck – Körnerの正復号指数を達成する同時分布が求まる.
- 正復号指数にしか適用できない. (誤り指数はminimax)
- ➔ Arimotoの指数計算アルゴリズムの際立つ秀逸さ.

補題1: 次式が成立する.

$$\begin{aligned} G_{\text{DK}}(R|W) &= \min_{q_{XY}} \{ D(q_{Y|X} \| W | q_X) + |R - I(q_X, q_{Y|X})|^+ \} \\ &= \max_{0 \leq \lambda \leq 1} \min_{q_{XY}} \{ D(q_{Y|X} \| W | q_X) + \lambda [R - I(q_X, q_{Y|X})] \} \end{aligned}$$

補題1より, 次の目的関数の同時分布 q_{XY} に関する最小化を考える.

$$\begin{aligned} \Theta^{(\lambda)}(q_{XY}|W) &:= D(q_{Y|X} \| W | q_X) - \lambda I(q_X, q_{Y|X}) \\ &= \mathbb{E}_{q_{XY}} \left[\log \frac{q_{Y|X}^{1-\lambda}(Y|X) q_Y^\lambda(Y)}{W(Y|X)} \right] \end{aligned}$$

次の性質が成り立つ.

性質1: 固定された $\lambda \in [0,1]$ に対し $\Theta^{(\lambda)}(q_{XY}|W)$ は q_{XY} に関して下に凸である.

証明は省略(J.-Oohama, IT Trans, 2020 参照.)

Oohama-Jitsumatsuのアルゴリズム (準備 2)

Oohama-J. (isit2015)は二重最小化のための次の関数を導入した.

定義 1

$$J^{(\lambda)}(q, p) = \mathbb{E}_{q_{XY}} \left[\log \frac{p_{Y|X}^{1-\lambda}(Y|X)p_Y^\lambda(Y)}{W(Y|X)} \right] + D(q_{XY} \| p_{XY})$$

補題 2 固定された q に対し, $J^{(\lambda)}(q, p)$ は, $p = q$ のとき最小値

$$J^{(\lambda)}(q, q) = \Theta^{(\lambda)}(q|W)$$

を取る.

補題 3 固定された p に対し, $J^{(\lambda)}(q, p)$ は,

$$q(x, y) = \frac{1}{K} p_{X|Y}^\lambda(x|y) p_X^{1-\lambda}(x) W(y|x) = q^*(p)$$

のとき最小値

$$J^{(\lambda)}(q^*(p), p) = -\log K$$

と取る. ここで,

$$K = \sum_{x,y} p_{X|Y}^\lambda(x|y) p_X^{1-\lambda}(x) W(y|x).$$

Algorithm 1 Oohama and Jitsumatsu's algorithm (isit2015)

Require: The conditional probability of the channel W and $\lambda \in (0, 1)$. Choose the initial joint probability distribution $q_{XY}^{[0]}$ such that $q_{XY}^{[0]}(x, y) = 0$ if $W(y|x) = 0$ and $q_{XY}^{[0]}(x, y) > 0$ if $W(y|x) > 0$.

for $i = 0, 1, 2, \dots$, **do**

$$\begin{aligned} & q_{XY}^{[i+1]}(x, y) \\ &= \frac{q_X^{[i]}(x)^{1-\lambda} q_{X|Y}^{[i]}(x|y)^\lambda W(y|x)}{\sum_{x'} \sum_{y'} q_X^{[i]}(x')^{1-\lambda} q_{X|Y}^{[i]}(x'|y')^\lambda W(y'|x')} \end{aligned}$$

end for

- 本アルゴリズムは p と q の交互最適化に基づくが, $p^{[i+1]} = q^{[i+1]}$ のため, $q^{[i]}$ のみ更新すればよい.

- 次の関数を定義する.

$$F_{\text{OJ}}^{(\lambda)}(p_{XY}) = -\log \sum_{x,y} p_{X|Y}^\lambda(x|y) p_X^{1-\lambda}(x) W(y|x)$$

- アルゴリズムが収束することから次の等式が成立する.

$$\min_{q_{XY}} \Theta^{(\lambda)}(q_{XY}|W) = \min_{q_{XY}} \min_{p_{XY}} J^{(\lambda)}(p_{XY}, q_{XY}) = \min_{p_{XY}} F_{\text{OJ}}^{(\lambda)}(p_{XY})$$

- これより Dueck-Körner 指数の別表現が得られる.

$$\begin{aligned} G_{\text{DK}}(R|W) &:= \min_{q_{XY}} \{D(q_{Y|X}||W|q_X) + |R - I(q_X, q_{Y|X})|^+\} \\ &= \max_{0 \leq \lambda \leq 1} \min_{q_{XY}} \{\lambda R + \Theta^{(\lambda)}(q_{XY}|W)\} \\ &= \max_{0 \leq \lambda \leq 1} \min_{p_{XY}} \{\lambda R + F_{\text{OJ}}^{(\lambda)}(p_{XY}|W)\} =: \hat{G}_{\text{DK}}(R|W) \end{aligned}$$

正復号指数の表現間の関係

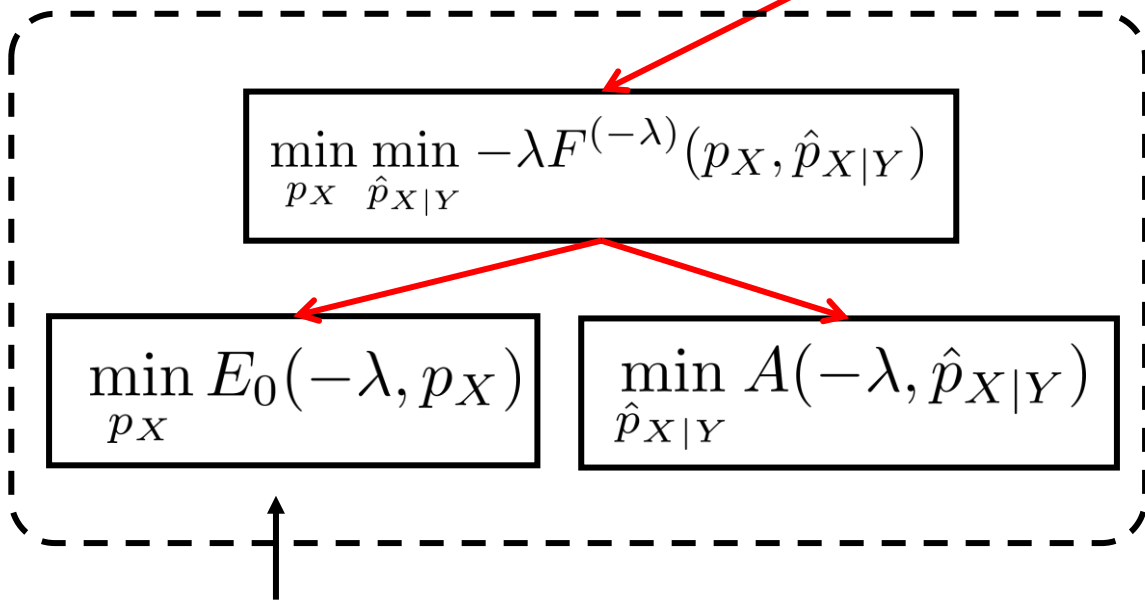
2つのアルゴリズムの関係を示すために、関数 \tilde{J} を導入する。

赤線は、最適化により等号が示されることを意味する。

$$\min_{p_X} \min_{\hat{p}_{X|Y}} \min_{q_{XY}} \min_{\hat{q}_{XY}} \tilde{J}^{(\lambda)}(p_X, \hat{p}_{X|Y}, q_{XY}, \hat{q}_{XY})$$

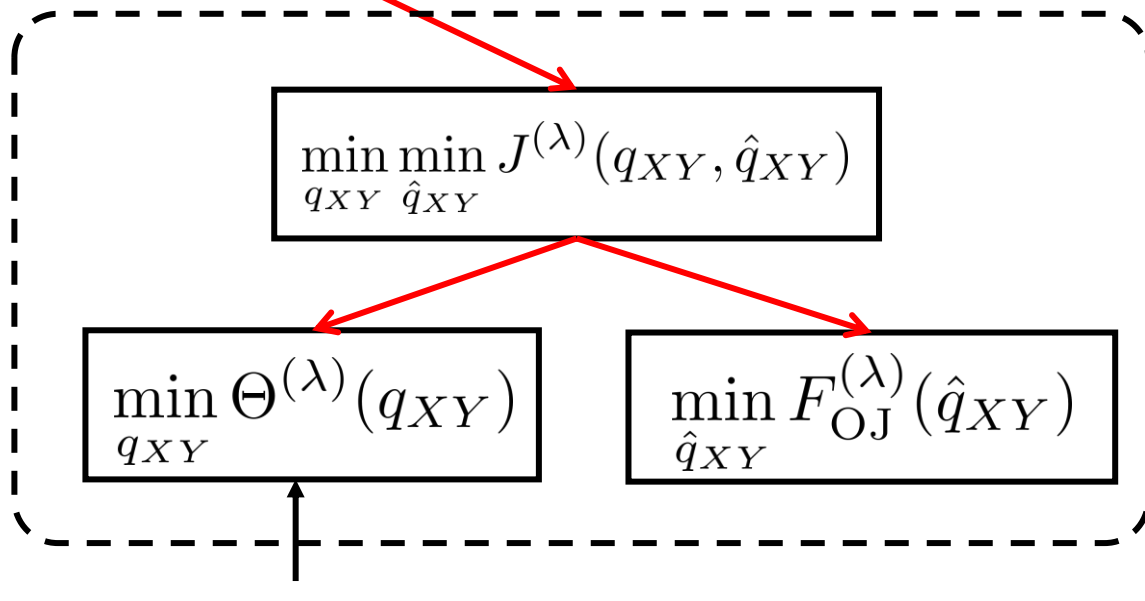
$$\begin{aligned} \hat{J}^{(\lambda)}(p_X, \hat{p}_{X|Y}, q_{XY}, \hat{q}_{XY}) \\ := J^{(\lambda)}(q_{XY}, \hat{q}_{XY}) \\ + (1 - \lambda)D(q_X || p_X) \\ + \lambda D(q_{X|Y} || p_{X|Y} | q_X) \end{aligned}$$

Arimotoの指数計算アルゴリズム



Arimotoの正復号指数の最適化問題

Oohama-J.のアルゴリズム



D&K指数の最適化問題

- Arimotoアルゴリズム
 - 誤り指数と正復号指数のパラメトリックな表現に基づく
 - **誤り指数と正復号指数の両方に適用できる.**
 - $p(x)$ と $q(x|y)$ を交互に最適化する
 - 有歪み情報源符号化でも同様のアルゴリズムを提案
- Oohama-J.のアルゴリズム
 - Dueck-Körnerの正復号指数の表現からの導出を示した.
 - **誤り指数に適用できない**
 - 同時分布 $q(x, y)$ を更新する
 - 有歪み情報源符号化用の同様のアルゴリズムを提案(isit2016)
- 収束スピード：
 - 理論的解析は未だされていない。例題ではArimotoの方が速い。

J.-Oohama(arXiv:2205.06027)の紹介

- 誤り指数の2つの異なる表現は関数としては一致する。

$$E_r(R|W) = \max_{\rho \in [0,1]} \left\{ -\rho R + \max_{p_X} E_0(\rho, p_X) \right\}$$

$$E_{CK}(R|W) = \max_{q_X} \min_{q_{Y|X}} \{ D(q_{Y|X} || W|q_X) + |I(q_X, q_{Y|X}) - R|^+ \}$$

- 従来知られている証明 → Csiszár-Körnerの教科書の練習問題 Problem 23 (10.24 in 2nd Edition). (誘導付き)

- minimax の点の上界と下界を示す。

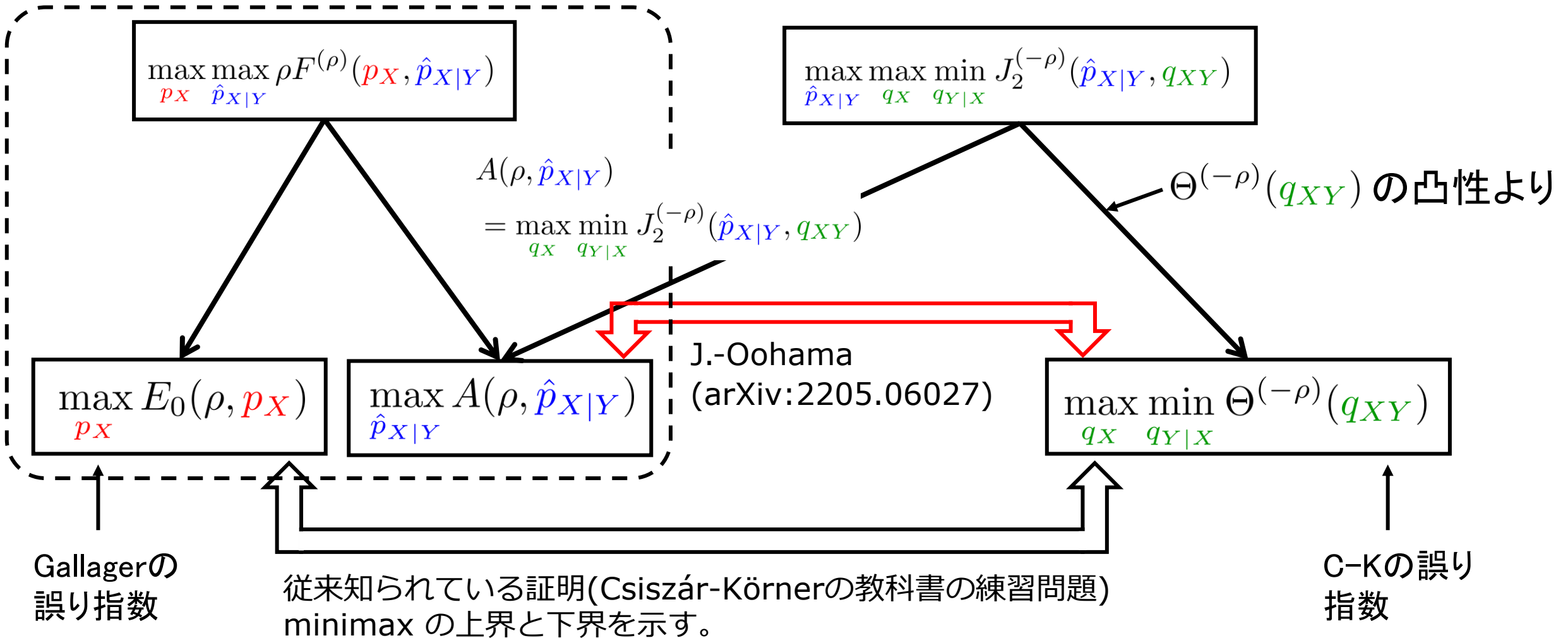
- 我々の証明 $E_r(R|W)$ の別表現と $E_{CK}(R|W)$ が一致することを示した。

$$\hat{E}_r(R|W) = \max_{\rho \in [0,1]} \left\{ -\rho R + \max_{\hat{p}_{X|Y}} \rho \log \sum_x \left[\sum_y \hat{p}_{X|Y}^{-\rho}(x|y) W(y|x) \right]^{-1/\rho} \right\}$$

- 各ステップの最適分布が陽に分かる。
- 正復号指数のアルゴリズム導出の副産物

誤り指数の異なる表現の一致

Arimotoの指数計算アルゴリズム



- Arimotoの指数計算アルゴリズム
 - 誤り指数と正復号指数の両方に対応できる.
- OohamaとJitsumatsuのアルゴリズム
 - 正復号指数を計算できる. 誤り指数には対応していない
- 交互最適化 → 別表現を与える.
 - アルゴリズムによる最適分布への収束 → 2つの表現の一致
- 今回の発表で省略したもの
 - 第3のアルゴリズム
 - 有歪み情報源符号化における正復号指数

Arimoto-Blahutアルゴリズムに関する話題(1)

- 汎用ソルバ

- CVXPY:

- https://www.cvxpy.org/examples/applications/Channel_capacity_BV4.57.html

- CVXは凸最適化のためのMatlabライブラリ。
CVXPYは、そのpython移植版

プログラムの詳細は、上のリンクを見てください。

ポイント: 目的関数を相互情報量として定義し、制約条件を書く。
あとは最大化せよと指示するだけ。

■ 通信路容量の計算：CVXPYはデフォルト設定。

(遷移確率行列は、中川先生の論文の例題から選びました)

通信路容量	W_1	W_2	W_3	W_4
CVXPY	0.46579975878 22469	0.44706882491 30339	0.44706749870 19183	0.285830638846 80815
Arimoto-Blahut ($\epsilon = 1.0 \times 10^{-8}$)	0.46579975878 211427	0.44706882487 37066	0.44706748513 425076	0.285868465006 82567
Arimoto-Blahut ($\epsilon = 1.0 \times 10^{-10}$)	0.46579975878 21180	0.44706882491 25958	0.44706749855 570566	0.285868465080 14924
Arimoto-Blahut ($\epsilon = 1.0 \times 10^{-20}$)	0.46579975878 21180	0.44706882491 25997	0.44706749870 19189	0.285868465080 15680

$$W_1 = \begin{bmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.250 & 0.250 & 0.500 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.300 & 0.300 & 0.400 \end{bmatrix}$$

$$W_4 = \begin{bmatrix} 0.600 & 0.100 & 0.100 & 0.100 & 0.100 \\ 0.100 & 0.600 & 0.100 & 0.100 & 0.100 \\ 0.231 & 0.231 & 0.066 & 0.179 & 0.292 \\ 0.161 & 0.341 & 0.226 & 0.226 & 0.046 \\ 0.522 & 0.160 & 0.046 & 0.227 & 0.046 \end{bmatrix}$$

$$W_3 = \begin{bmatrix} 0.800 & 0.100 & 0.100 \\ 0.100 & 0.800 & 0.100 \\ 0.350 & 0.350 & 0.300 \end{bmatrix}$$

$$\max_x |p^{[i+1]}(x) - p^{[i]}(x)| < \epsilon \text{ で停止}$$

● AB algorithmとCVXPY(最適化手法はデフォルト)の実行時間

単位(ms) [反復回数]	W_1	W_2	W_3	W_4
CVXPY	2.99	1.99	1.99	2.99
Arimoto-Blahut ($\epsilon = 1.0 \times 10^{-8}$)	1.99 [92]	47.3 [3873]	0.99 [98]	89.2 [7208]
($\epsilon = 1.0 \times 10^{-10}$)	1.99 [122]	86.7 [7272]	0.99 [127]	187.4 [14587]
($\epsilon = 1.0 \times 10^{-20}$)	4.98 [216]	529.1 [21836]	6.97 [274]	1061.1 [44401]

- デフォルト設定では実行時間は短い。精度は不明。
- CVXPYではパラメタを設定できるが、意味を正しく理解する必要がある。
→cvxのソルバに精通しないと怖い。
- デフォルト設定で使用されるのは、embedded conic solver (ECOS), an interior-point solver for second-order cone programming (SOCP)
- 物は使いよう……

Arimoto-Blahutアルゴリズムに関する話題(2)

- 有本先生による日本語による記事
 - J-STAGE(<https://www.jstage.jst.go.jp/>)で多数見つかる.
- 応用数理、2002年12巻1号～4号、応用数理の遊歩道 (28) ～ (31)
 - ➔ **重要**
 - 応用数理の遊歩道(30)に有本先生がアルゴリズムを発見した経緯の記述がある。
 - マルコフ連鎖がエルゴード的ならば、 $p^{[t]} = p^{[t-1]} \mathbf{P}$ が定常分布 p^* に収束する。KLダイバージェンスの非不性を用いると、このことが証明できるが、“逆に、このような単調性を保証するものは、KLダイバージェンスしかない”と直感し、その証明を試みていたときであった。”
- 計測と制御、昭和50年1月号 (特集、制御理論の新しい展開) 「情報理論とそのシステム理論との関連分野における発展」
- 計測と制御、昭和45年4月号 「情報理論から見たシステムの同定と推定」