Finding Hall blockers by matrix scaling

平井 広志
東京大学大学院 情報理工学系研究科
数理情報学専攻
JST さきがけ
hirai@mist.i.u-tokyo.ac.jp

Joint work with
林 興養（中央大学）

応用数理学会年会
離散システム研究部会
2022年9月9日

（２重確率）行列スケーリング（Sinkhorn 1964）

A: $n \times n$ 非負行列

正対角行列 R, C によって RAC を２重確率行列にできるか？

$RAC\,\mathbf{1} \approx \mathbf{1}, (RAC)^\top \mathbf{1} \approx \mathbf{1}$ としたい

応用：

- マルコフ連鎖の推定
- 線形計算の前処理
- 最適輸送, Entropic regularization (Cuturi 2013)
- and more

Sinkhorn のアルゴリズム

- 行正規化：$A \leftarrow RA \quad \text{s.t.} \ \ (RA)\mathbf{1} = \mathbf{1};\ R = \mathrm{diag}(\cdots \frac{1}{\sum_k A_{ik}} \cdots)$
- 列正規化：$A \leftarrow \ \ AC \ \ \text{s.t.} \ \ (AC)^\top\mathbf{1} = \mathbf{1};\ C = \ \mathrm{diag}(\cdots \frac{1}{\sum_k A_{kj}} \cdots)$
- これをくりかえす

$$\begin{array}{cc} & \begin{matrix}4 & 10 & 1 & 1\end{matrix} \\ \begin{matrix}6\\3\\4\\3\end{matrix} & \begin{pmatrix} & 6 & & \\ 1 & & 1 & 1 \\ & 4 & & \\ 3 & & & \end{pmatrix}\end{array} \xrightarrow{\text{行正規化}} \begin{array}{cc} & \begin{matrix}4/3 & 2 & 1/3 & 1/3\end{matrix} \\ \begin{matrix}1\\1\\1\\1\end{matrix} & \begin{pmatrix} & 1 & & \\ 1/3 & & 1/3 & 1/3 \\ & 1 & & \\ 1 & & & \end{pmatrix}\end{array}$$

$$\xrightarrow{\text{列正規化}} \begin{array}{cc} & \begin{matrix}1 & 1 & 1 & 1\end{matrix} \\ \begin{matrix}1/2\\9/4\\1/2\\3/4\end{matrix} & \begin{pmatrix} & 1/2 & & \\ 1/4 & & 1 & 1 \\ & 1/2 & & \\ 3/4 & & & \end{pmatrix}\end{array} \xrightarrow{\text{行正規化}} \cdot\ \cdot\ \cdot$$

２重確率スケーリング可能性の特徴付け

Sinkhorn, Knopp 1967, Rothblum, Schneider 1989

Thm: $A: n \times n$ 非負行列，以下は同値.

- Aは（近似的）２重確率スケーリング可能：

$$\forall \epsilon > 0, \exists R, C \text{ s.t. } \|RAC\mathbf{1} - \mathbf{1}\| < \epsilon, \|(RAC)^\top \mathbf{1} - \mathbf{1}\| < \epsilon$$

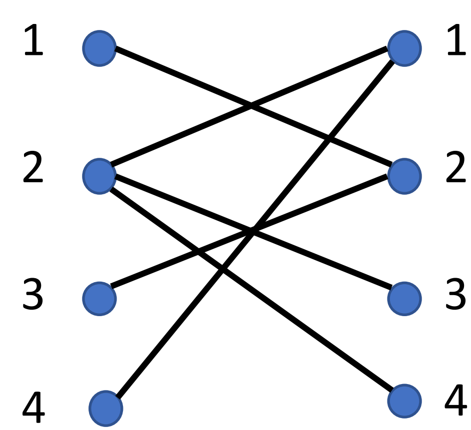
- Sinkhornアルゴリズムによって，Aは２重確率行列に収束

- ２部グラフG_Aに完全マッチングが存在

$$V(G_A) \coloneqq [n] \sqcup [n],\ E(G_A) \coloneqq \{\, ij \mid A_{ij} \neq 0 \,\}$$

$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} & 6 & & \\ 1 & & 1 & 1 \\ & 4 & & \\ 3 & & & \end{pmatrix} \end{array}$$

$G_A =$



Sinkhorn 反復による完全マッチングの存在判定

Linial, Samorodnitsky, Wigderson 2000

$G = (U \sqcup V, E)$: 2部グラフ, $|U| = |V| = n$

$$A(G)_{ij} := \begin{cases} 1 & \text{if } \; ij \in E \\ 0 & \text{otherwise} \end{cases}$$

初期化: $A \leftarrow A(G)$

1. 行正規化: $A \leftarrow \mathrm{diag}(1/\sum_k A_{ik})A$
2. 列正規化: $A \leftarrow A\,\mathrm{diag}\left(1/\sum_k A_{kj}\right)$
3. $\|A\mathbf{1} - \mathbf{1}\|_2 < 1/\sqrt{n}$ なら終了.
4. Go to 1.

Thm [LSW2000]

G に完全マッチング有り→ $n^2 \log n$ 反復以内に終了.

G に完全マッチング無し→ 終了しない.

- LSWアルゴリズムは，増加道アルゴリズムより遅いが，
根本的に異なる原理に基づいている点や
そのアルゴリズムの（異常な）単純さにおいて興味深い．

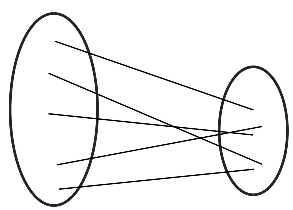
- 作用素スケーリングの最近の発展につながる重要な結果

- LSWアルゴリズムは，完全マッチングも非存在の証拠（Hall blocker）
も出力しない．

Hallの結婚定理：完全マッチングが存在 ⇔ Hall blockerが存在しない

Hall blocker

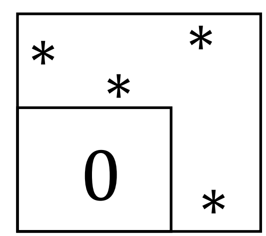
X



$\Gamma(X)$

$|X| > |\Gamma(X)|$

X



Y $\Gamma(X)$

$|X| + |Y| > n$

本研究の問題意識：

Sinkhorn反復によってHall blockerは特定できるか？

ねらい：作用素スケーリングへの展開（後述）

本研究の成果

多項式回の Sinkhorn 反復で Hall blocker は特定可能

初期化: $A \leftarrow A(G)$

1. 行正規化: $A \leftarrow \mathrm{diag}(1/\sum_k A_{ik})A$
2. 列正規化: $A \leftarrow A\,\mathrm{diag}\left(1/\sum_k A_{kj}\right)$
3. $p \coloneqq A\mathbf{1}$ を $p_{i_1} \leq p_{i_2} \leq \cdots \leq p_{i_n}$ とソートし,

   $\mathcal{X} \coloneqq \{\{i_1, i_2, \ldots, i_k\} \mid k = 1,2,\ldots,n\}$.
4. Go to 1.

LSW: $\|A\mathbf{1} - \mathbf{1}\|_2 < 1/\sqrt{n}$ なら終了

Thm (本研究)

G に完全マッチング無し→

$O(n^7 \log n)$回の反復のうちに $\mathcal{X}$ にはHall blockerが含まれる.

証明のアイデア 1: Sinkhorn 反復の交互最適化解釈

LSWの解析 ～ capacity 交互最適化解釈

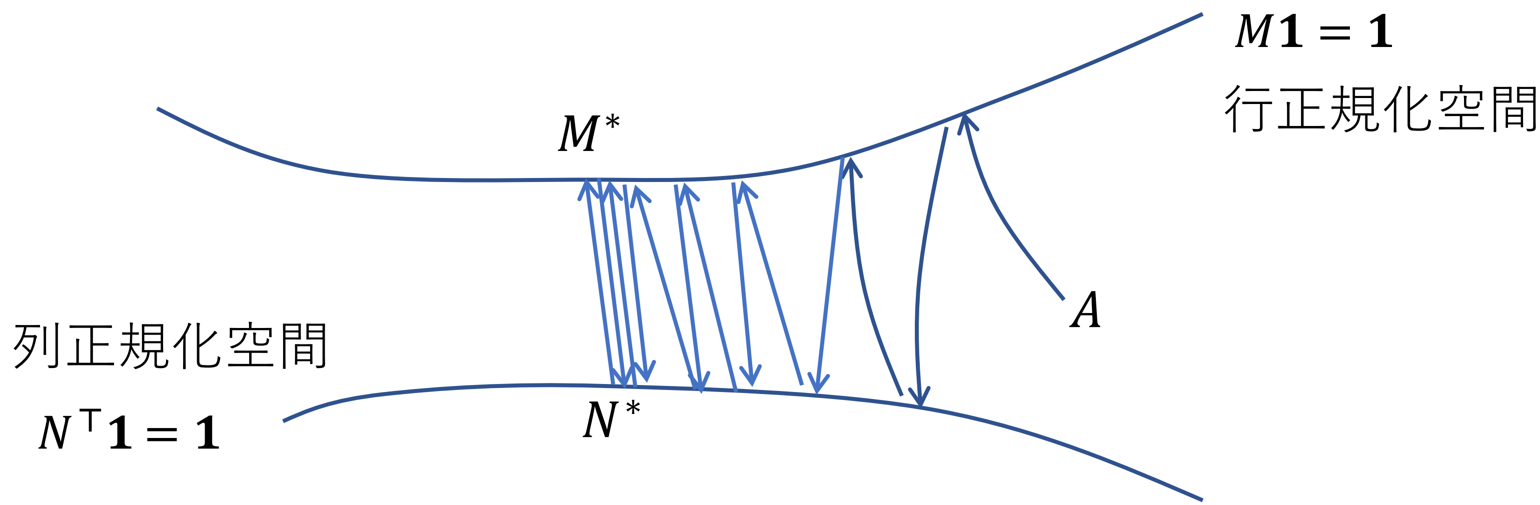
$$\text{inf.} \quad \log \frac{x^\top A y}{\prod_i x_i \prod_j y_j} \quad \text{s.t.} \quad x > 0, y > 0.$$

- $\text{opt} > -\infty \Leftrightarrow$ スケーリング可能 (RS89)
- 収束のスピード & 定数 κ: $\text{opt} > -\infty \Rightarrow \text{opt} > \kappa$ の評価

本研究の解析 ～ KLダイバージェンス交互最適化解釈（情報幾何解釈）

$$\text{inf.} \sum_{ij} M_{ij} \log \frac{M_{ij}}{N_{ij}} \quad \text{s.t.} \ M\mathbf{1} = \mathbf{1}, N^\top \mathbf{1} = \mathbf{1},$$
$$M, N \geq 0, \ \ \operatorname{supp} M, \operatorname{supp} N \subseteq \operatorname{supp} A$$

Sinkhorn反復の情報幾何学的解釈



Thm (Csiszar, Tusnady 1984, Gietl, Reffel 2013)

最小 KL ダイバージェンス点対 (M^*, N^*) に収束

- Sinkhorn反復が収束しないとき，　極限で M^*, N^*を振動
- 最適値 $D_{\mathrm{KL}}(M^* \,||\, N^*)$への収束レート～ $O(1/k)$

本研究：$p^* = N^*\mathbf{1}$ への収束評価

← Pinskerの不等式 & 収束先 (M^*, N^*) の特徴付け（次ページ）

証明のアイデア 2：収束先の特徴付け

Aas 2014:

- M^*, N^*はブロック対角化される.
- 各ブロックは，極限で

$$\begin{array}{c|ccccc|} & 1 & 1 & 1 & \cdots & 1 \\ \hline \alpha & & & & & \\ \alpha & & & & & \\ \vdots & & & & & \\ \alpha & & & & & \\ \hline \end{array} \quad \Leftrightarrow \quad \begin{array}{c|ccccc|} & \frac{1}{\alpha} & \frac{1}{\alpha} & \frac{1}{\alpha} & \cdots & \frac{1}{\alpha} \\ \hline 1 & & & & & \\ 1 & & & & & \\ \vdots & & & & & \\ 1 & & & & & \\ \hline \end{array}$$

と振動

$\alpha \coloneqq$ 列数 / 行数

本研究: **一般化 DM 分解**の対角ブロックたちに一致

行列の並べ替えの標準形

→DEMO

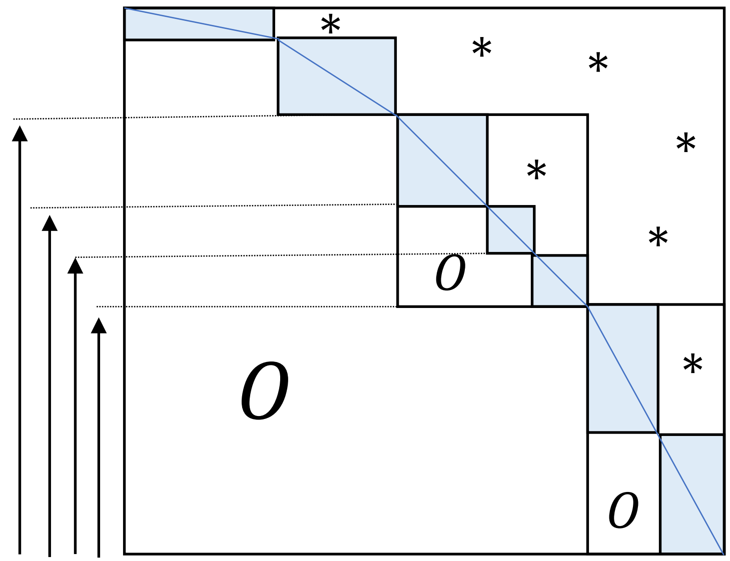
残りの部分

Dulmage-Mendelson 1958

DM分解＝

「最大」Hall blockerにそって
行と列を並べ替え

$A \quad \rightarrow$



ゼロブロックのサイズ
:= 行数 + 列数

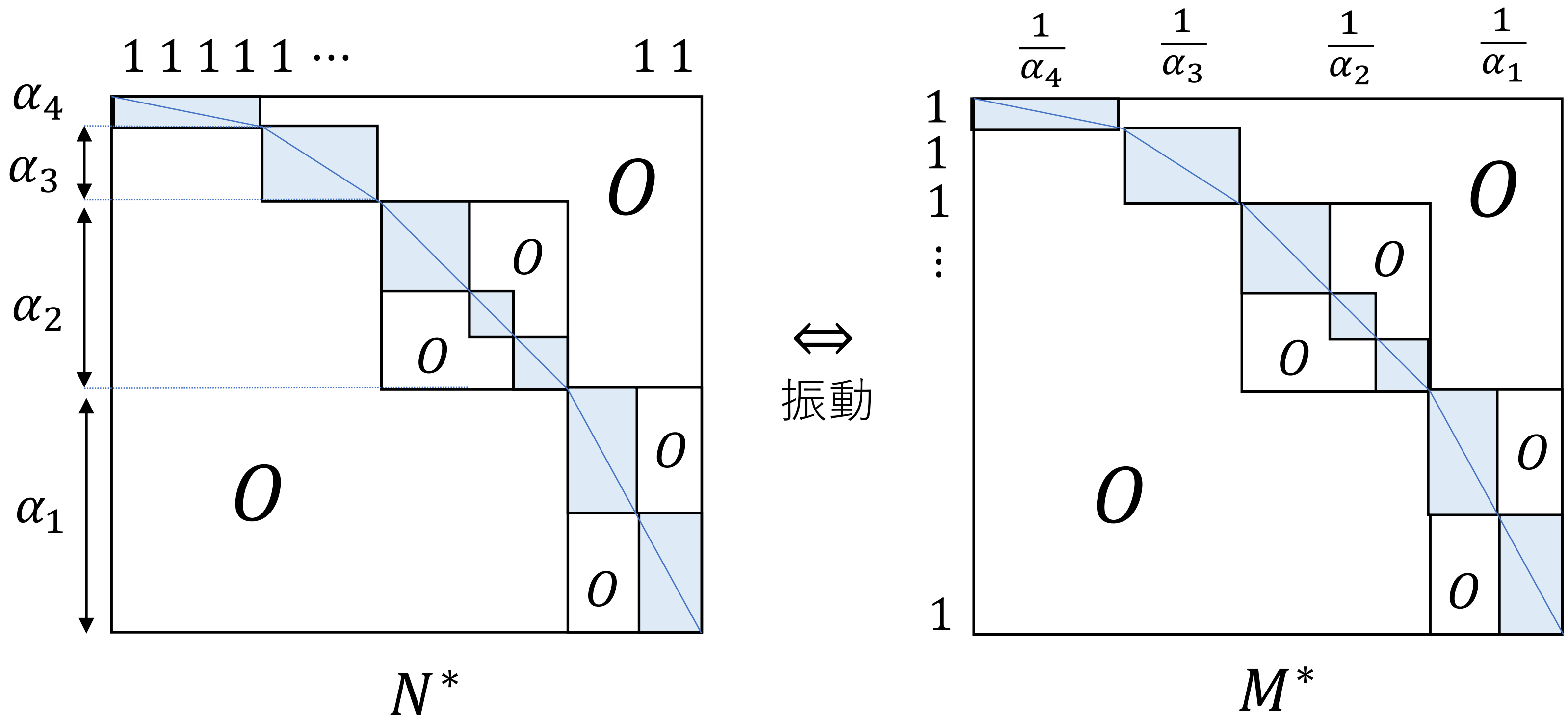
冨澤 1977 (unpublished)

一般化 DM 分解 ＝ 基本分割の理論に基づいて「残りの部分」も分解

ゼロブロックの重み付きサイズ:= α × 行数 + 列数

$$\alpha \in [0, \infty)$$

Sinkhorn 反復の極限



- $\alpha_1 < \alpha_2 < \alpha_3 < \cdots$ より，行和ベクトル $p^* = N^*\mathbf{1}$ を並べかえると DM分解の構造がわかる → Hall blockerもわかる
- 反復がすすむと $p = A\mathbf{1}$は，十分 $p^* = N^*\mathbf{1}$に近くなるので, $p = A\mathbf{1}$を並べかえてもわかる（収束評価を使う）．

まとめ

- Sinkhorn反復で Hall blocker も求まるようにした.
- 情報幾何 & DM分解との関係
- $O(n^7 \log n)$ を改良できるか？

作用素スケーリング（Gurvits 2004）への展開

$A_1, A_2, \ldots, A_m \in \mathbb{C}^{n\times n}$ がつくる完全正値作用素が

2 重確率 $\underset{\text{def}}{\Leftrightarrow} \sum_k A_k A_k^\dagger = I, \sum_k A_k^\dagger A_k = I$

正則行列 g, hによって，$g(A_1, A_2, \ldots, A_m)h^\dagger$ を2 重確率 にできるか？

作用素 Sinkhornアルゴリズム（Gurvitsのアルゴリズム）

1. 行正規化: $A_k \leftarrow gA_k$: $g(\sum_k A_k A_k^\dagger)g^\dagger = I$
2. 列正規化: $A_k \leftarrow A_k h^\dagger$: $h(\sum_k A_k^\dagger A_k)h^\dagger = I$
3. Go to 1.

Thm (Gurvitz 2004): 以下は同値.

- $(A_1, A_2, \dots, A_m)$ は（近似的）2重確率スケーリング可能.
- 作用素 Sinkhorn アルゴリズムが収束.
- $U^\dagger A_k V = \{0\}\ (\forall k),\ \dim U + \dim V > n$ となる
  ベクトル部分空間 $U, V \subseteq \mathbb{C}^n$ 無し.

Shrunk subspace: スケーリング不可能性の証拠

作用素 Sinkhorn 反復で Shrunk subspace は見つかるか？

Franks, Soma, Goemans 2022: （修正された Sinkhorn 反復で）見つかる

- K. Hayashi and H.Hirai: Finding Hall blockers by matrix scaling, arXiv, 2022

- 平井広志「行列スケーリングの数理」（読み物，演習資料）
  平井のページからダウンロード可

- 9/12: JST 数学関連３領域連携 WS「情報科学と拓く新しい数理科学」
  平井広志「行列スケーリングから非正曲率空間上の測地的凸最適化へ」

おわり