

作用素スケーリング入門

相馬 輔
(統計数理研究所)

組合せ最適化セミナー 2024

概要

- **スケーリング問題**は様々な分野に共通に現れる線形代数の問題
- **行列スケーリング**: 統計学, 機械学習, 組合せ最適化...
- **作用素スケーリング**: 組合せ最適化, 計算量理論, 非可換代数, 関数解析...
- **交互最適化** (Sinkhorn アルゴリズム) で解ける
- **Brascamp-Lieb 多面体**など, 組合せ最適化で面白い多面体が現れる

講演概要

Part I: 行列スケールリング

- 行列スケールリングとは？
- Sinkhorn アルゴリズム，凸解析・組合せ最適化との関連，応用

Part II: 作用素スケールリング

- 作用素スケールリングとは？
- 作用素 Sinkhorn アルゴリズム，解析，応用

Part III: Brascamp-Lieb 不等式と作用素スケールリング

- Brascamp-Lieb 不等式と多面体
- Brascamp-Lieb 多面体上の最適化

講演概要

Part I: 行列スケーリング

- 行列スケーリングとは？
- Sinkhorn アルゴリズム，凸解析・組合せ最適化との関連，応用

Part II: 作用素スケーリング

- 作用素スケーリングとは？
- 作用素 Sinkhorn アルゴリズム，解析，応用

Part III: Brascamp-Lieb 不等式と作用素スケーリング

- Brascamp-Lieb 不等式と多面体
- Brascamp-Lieb 多面体上の最適化

行列スケールリング

入力: 非負行列 $A \in \mathbb{R}_+^{n \times n}$

出力: 次を満たす正対角行列 L, R :

$$(LAR)\mathbf{1} = \mathbf{1} \quad \text{かつ} \quad (LAR)^\top \mathbf{1} = \mathbf{1}$$

すなわち LAR が二重確率行列になる L, R

応用

- マルコフ連鎖の推定 [Sinkhorn 1964]
- 産業連関表の推定 (RAS 法) [Morioka, Tsuda 2011]
- 最適輸送 (Wasserstein 距離) の高速計算 [Wilson 1969; Peyré, Cuturi 2019]

Sinkhorn の定理

定理 (Sinkhorn (1964))

正行列 A (すなわち任意の i, j で $a_{ij} > 0$) に対し, 行列スケーリングの解 (L, R) が常に存在する.

効率的なアルゴリズムで解を見つけられるか?

Sinkhorn アルゴリズム [Sinkhorn 1964]

$A^\top \mathbf{1} = \mathbf{1}$ と仮定してよい。

$$A^{(0)} = A$$

$$A^{(2t+1)} = \text{Diag}(A^{(2t)} \mathbf{1})^{-1} A^{(2t)}, \quad (\text{行正規化})$$

$$A^{(2t+2)} = A^{(2t+1)} \text{Diag}((A^{(2t+1)})^\top \mathbf{1})^{-1}. \quad (\text{列正規化})$$

定理 (Sinkhorn (1964))

A が正行列ならば、 $A^{(t)}$ は二重確率行列に収束。

数值例

$$A^{(0)} = \begin{matrix} & & \mathbf{1.0000} & \mathbf{1.0000} & \mathbf{1.0000} & \mathbf{1.0000} \\ \mathbf{1.2000} & \left[\begin{array}{cccc} \mathbf{0.3062} & \mathbf{0.4189} & \mathbf{0.0214} & \mathbf{0.4535} \\ \mathbf{0.9089} & \mathbf{0.1533} & \mathbf{0.1564} & \mathbf{0.4889} & \mathbf{0.1104} \\ \mathbf{0.7675} & \mathbf{0.3142} & \mathbf{0.0410} & \mathbf{0.2224} & \mathbf{0.1899} \\ \mathbf{1.1235} & \mathbf{0.2263} & \mathbf{0.3838} & \mathbf{0.2672} & \mathbf{0.2462} \end{array} \right. & & & & \end{matrix}$$

周辺分布つき行列スケーリング

入力: 非負行列 $A \in \mathbb{R}_+^{m \times n}$, **周辺分布** $r \in \mathbb{R}_+^m, c \in \mathbb{R}_+^n$

出力: 正対角行列 L, R s.t.

$$(LAR)\mathbf{1} = r \quad \text{かつ} \quad (LAR)^\top \mathbf{1} = c.$$

- Sinkhorn の定理やアルゴリズムは容易に周辺分布つき設定に拡張できる

行列スケーリングの歴史

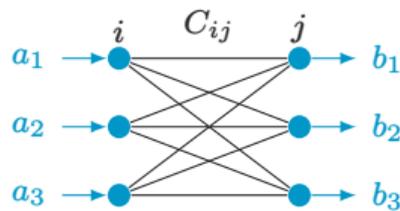
1930s	Kruithof	電話網の問題
1940s	Deming, Stephen	統計
1960s	Stone Sinkhorn Knopp	経済学 (RAS method) “行列スケーリング”の定式化 Sinkhorn アルゴリズム
1970s	Csiszár	情報理論
1990s	Wigderson et al.	理論計算機科学
2010s	Cuturi	機械学習

参考文献 [Idel 2016]

応用: 輸送問題

$C \in \mathbb{R}_+^{m \times n}$: 費用行列, $a \in \mathbb{R}_+^m$, $b \in \mathbb{R}_+^n$: 供給/需要

$$\min_{P \geq 0} C \bullet P \quad \text{s.t.} \quad P\mathbf{1} = a, P^\top \mathbf{1} = b$$



エントロピー正則化 [Wilson 1969]

$$\min_{P \geq 0} C \bullet P + \lambda \sum_{i,j} p_{ij} \log(p_{ij}) \quad \text{s.t.} \quad P\mathbf{1} = a, P^\top \mathbf{1} = b$$

→ 最適解は $P^* = LAR$ (L, R : 正対角行列, $a_{ij} = \exp(-c_{ij}/\lambda)$) の形

行列スケールリング!

- 近年機械学習で広く活用されている (Wasserstein 距離) [Peyré, Cuturi 2019]

応用: 産業連関表の推定

産業連関表... ある地域における産業間の取引量をまとめた表。

昨年度の表 A と今年度の行和 r , 列和 c があるとき, 今年度の表 P を推定したい。

	需要側				
供給側	産業 1	産業 2	...	産業 n	行和
産業 1	a_{11}	a_{12}	...	a_{1n}	r_1
産業 2	a_{21}	a_{22}	...	a_{2n}	r_2
⋮	⋮	⋮	⋮	⋮	⋮
産業 m	a_{m1}	a_{m2}	...	a_{mn}	r_m
列和	c_1	c_2	...	c_n	

KL ダイバージェンス最小化

$$\min_{P \geq 0} D_{KL}(P : A) \quad \text{subject to} \quad P\mathbf{1} = r, P^T\mathbf{1} = c$$

ここで

$$D_{KL}(P : A) = \sum_{i,j} \left(p_{ij} \log \frac{p_{ij}}{a_{ij}} - p_{ij} + a_{ij} \right)$$

は **Kullback–Leibler (KL) ダイバージェンス**。

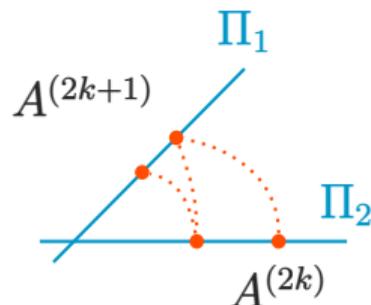
凸最適化としての解釈

以下, $r = c = \mathbf{1}_n$ の場合 (二重確率行列へのスケーリング) を考える.

KL ダイバージェンス最小化問題の双対問題は, 次の関数 f の最小化と等価.

$$f(x, y) = \sum_{i,j} a_{ij} e^{x_i + y_j} - \sum_i x_i - \sum_j y_j$$

- $\nabla_x f(x, y) = A\mathbf{1} - \mathbf{1}$, $\nabla_y f(x, y) = A^\top \mathbf{1} - \mathbf{1}$
- $\nabla f(x, y) = 0 \iff (L, R) = (\text{Diag}(e^x), \text{Diag}(e^y))$ が解
- f は凸関数. 停留点は (本質的に) 一意
- Sinkhorn アルゴリズムは交互最適化 $\implies O(1/\sqrt{t})$ 収束
- 各反復は Kullback–Leibler(KL) 射影 [Csiszár 1975]



交互最適化との等価性

$x, y \in \mathbb{R}^n$ に対応する A のスケーリングを $\tilde{A} = \text{Diag}(e^x)A \text{Diag}(e^y)$ と定める。

交互最適化

$$x^{(2t+1)} = \underset{x \in \mathbb{R}^n}{\text{argmin}} f(x, y^{(2t)})$$

$$y^{(2t+1)} = y^{(2t)}$$

$$x^{(2t+2)} = x^{(2t+1)}$$

$$y^{(2t+2)} = \underset{y \in \mathbb{R}^n}{\text{argmin}} f(x^{(2t+1)}, y)$$

Sinkhorn 反復

$$A^{(2t+1)} = \text{Diag}(A^{(2t)} \mathbf{1})^{-1} A^{(2t)}$$

$$A^{(2t+2)} = A^{(2t+1)} \text{Diag}((A^{(2t+1)})^\top \mathbf{1})^{-1}.$$

どちらも同じ行列 $A^{(t)}$ を定める。

近似行列スケーリング

入力: 非負行列 $A \in \mathbb{R}_+^{n \times n}$, $\varepsilon > 0$

出力: 正対角行列 $L_\varepsilon, R_\varepsilon$ s.t.

$$\|(L_\varepsilon A R_\varepsilon) \mathbf{1} - \mathbf{1}\|_1 < \varepsilon \quad \text{かつ} \quad \|(L_\varepsilon A R_\varepsilon)^\top \mathbf{1} - \mathbf{1}\|_1 < \varepsilon$$

- A が **近似スケーリング可能**

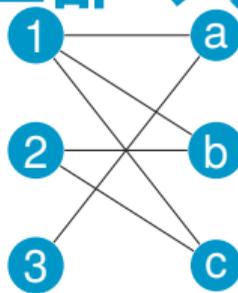
$\stackrel{\text{def}}{\iff} \forall \varepsilon > 0$ に対して近似スケーリング解が存在

例

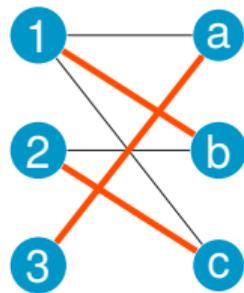
$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ は近似スケーリング可能だが，厳密スケーリング不可能．

近似スケーリングと二部マッチング

$$A = \begin{bmatrix} .3 & .2 & .5 \\ 0 & .8 & .5 \\ .7 & 0 & 0 \end{bmatrix}$$



台グラフ



完全マッチング

定理 (Sinkhorn, Knopp (1967))

$n \times n$ 非負行列 A (w.l.o.g. $A^T \mathbf{1} = \mathbf{1}$) に対して、以下は同値:

- 1 $\inf f > -\infty$
- 2 A は近似スケーリング可能
- 3 A の台グラフは完全マッチングをもつ

これらは $O(n^2 \log^2(1/\mu))$ 回の Sinkhorn 反復で判定可能.

μ : A の非零成分の最小値

証明の準備

Kullback–Leibler (KL) ダイバージェンス

$p, a \in \mathbb{R}_+^n$ に対し

$$D_{KL}(p : a) = \sum_i \left(p_i \log \frac{p_i}{a_i} - p_i + a_i \right)$$

ただし、 $a_i = 0$ かつ $p_i \neq 0$ を満たす i が存在するときは、
 $D_{KL}(p : a) = +\infty$ とする。

- $D_{KL}(p : a) \geq 0$
- $p = a \iff D_{KL}(p : a) = 0$
- **Pinsker の不等式:** $\sum_i p_i = \sum_i a_i =: N$ ならば

$$D_{KL}(p : a) \geq \frac{1}{2N} \|p - a\|_1^2$$

定理の証明 (1/3)

$$f(x, y) = \sum_{i,j} a_{ij} e^{x_i + y_j} - \sum_i x_i - \sum_j y_j$$

① \implies ②: t が偶数 (行正規化) の場合のみ考える.

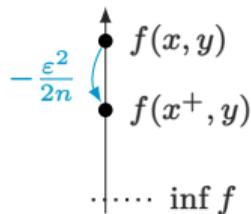
- $t = 0$ では $(x, y) = (\mathbf{0}, \mathbf{0})$ で $f(x, y) = n$.
- $x_i^+ = x_i - \log(A^{(t)}\mathbf{1})_i$ より

$$f(x^+, y) - f(x, y) = \sum_i (1 + \log(A^{(t)}\mathbf{1})_i - (A^{(t)}\mathbf{1})_i)$$

$$= -D_{KL}(\mathbf{1} : A^{(t)}\mathbf{1})$$

$$\leq -\frac{1}{2n} \|A^{(t)}\mathbf{1} - \mathbf{1}\|_1^2 \quad (\text{Pinsker})$$

よって, $\|A^{(t)}\mathbf{1} - \mathbf{1}\|_1 > \varepsilon$ である限り, f は $-\frac{\varepsilon^2}{2n}$ は減少する. 仮定より, $\inf f > -\infty$ なので, 有限回反復で行制約が満たされる.



※より強く $O(1/\sqrt{t})$ 収束が言えている

定理の証明 (2/3)

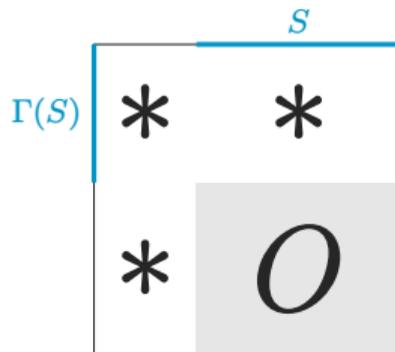
② \implies ③: $\|A\mathbf{1} - \mathbf{1}\|_1 \leq \varepsilon$, $A^\top \mathbf{1} = \mathbf{1}$ として,

Hall 条件 $|S| \leq |\Gamma(S)|$ (S : 列部分集合) を示す. ($\Gamma(S) = S$ の隣接点)

$$\begin{aligned} |S| &= \sum_{j \in S} \sum_{i \in \Gamma(S)} a_{ij} \leq \sum_{i \in \Gamma(S)} \sum_{j \in [n]} a_{ij} \\ &\leq \sum_{i \in \Gamma(S)} (1 + |\sum_{j \in [n]} a_{ij} - 1|) \\ &\leq |\Gamma(S)| + \|A\mathbf{1} - \mathbf{1}\|_1 \\ &\leq |\Gamma(S)| + \varepsilon \end{aligned}$$

$\varepsilon \rightarrow 0$ とすれば $|S| \leq |\Gamma(S)|$.

※上式より, $\varepsilon = 1/2$ に取れば Hall 条件を厳密に check できる



定理の証明 (3/3)

③ \implies ①:

演習問題

KL ダイバージェンス最小化問題 $\inf_P \{ D_{KL}(P : A) : P \geq 0, P\mathbf{1} = \mathbf{1}, P^\top \mathbf{1} = \mathbf{1} \}$ の Lagrange 双対問題は $\sup_{x,y} -f(x,y)$.

弱双対性から, $\sup -f(x,y) \leq \inf \{ D_{KL}(P : A) : P \geq 0, P\mathbf{1} = \mathbf{1}, P^\top \mathbf{1} = \mathbf{1} \}$.
いま, A の台グラフが完全マッチング M を持つので, P として

$$P_{ij} = \begin{cases} 1 & ij \in M \\ 0 & \text{otherwise} \end{cases}$$

を取ると, $P\mathbf{1} = \mathbf{1}, P^\top \mathbf{1} = \mathbf{1}$ かつ

$$D_{KL}(P : A) = \sum_{ij \in M} \log \frac{1}{a_{ij}} < +\infty$$

※ここから $\inf f = \Omega(n \log \mu)$ も従う

周辺分布つき行列スケーリングへの拡張

一般の行和 r , 列和 c へのスケーリングにも容易に拡張できる.

$$f(x, y) = \sum_{i,j} a_{ij} e^{x_i + y_j} - \sum_i r_i x_i - \sum_j c_j y_j$$

定理 (一般周辺分布の Sinkhorn–Knopp の定理)

非負 $m \times n$ 行列 A に対して, 以下は同値:

- 1 $\inf f > -\infty$
- 2 A が行和 r , 列和 c へ近似スケーリング可能
- 3 A の台グラフが (r, c) 輸送を持つ

これらは $O(n^2 \log^2(1/\mu))$ 回の Sinkhorn 反復で判定可能.

μ : A の非零成分の最小値

Part I のまとめ

入力 A : 非負行列,

スケーリング $\tilde{A} := LAR$ (L, R : 正対角行列)

スケーリング	関数 $f(x, y)$	A の台グラフの条件
二重確率 $\tilde{A}\mathbf{1} = \mathbf{1}, \tilde{A}^\top\mathbf{1} = \mathbf{1}$	$(e^x)^\top A(e^y) - \mathbf{1}^\top x - \mathbf{1}^\top y$	完全マッチングが存在 $\iff S \leq \Gamma(S) $
周辺分布 $\tilde{A}\mathbf{1} = r, \tilde{A}^\top\mathbf{1} = c$	$(e^x)^\top A(e^y) - r^\top x - c^\top y$	(r, c) 輸送が存在 $\iff c(S) \leq r(\Gamma(S))$

$(e^x) := \text{Diag}(e^x)$, S : 列部分集合, $\Gamma(S)$: S の隣接点

講演概要

Part I: 行列スケーリング

- 行列スケーリングとは？
- Sinkhorn アルゴリズム，凸解析・組合せ最適化との関連，応用

Part II: 作用素スケーリング

- 作用素スケーリングとは？
- 作用素 Sinkhorn アルゴリズム，解析，応用

Part III: Brascamp-Lieb 不等式と作用素スケーリング

- Brascamp-Lieb 不等式と多面体
- Brascamp-Lieb 多面体上の最適化

作用素スケーリング

行列スケーリングの一般化 (Gurvits (2004) が導入)

応用

- Edmonds 問題 [Gurvits 2004; Ivanyos, Qiao, Subrahmanyam 2017; Ivanyos, Qiao, Subrahmanyam 2018; Garg, Gurvits, Oliveira, Wigderson 2020] (cf. 平井先生の COSS 2019)
- Brascamp–Lieb 不等式 [Garg, Gurvits, Oliveira, Wigderson 2018]
- Quantum Schrödinger bridge [Georgiou, Pavon 2015]
- Paulsen 問題 [Kwok, Lau, Lee, Ramachandran 2018; Kwok, Lau, Ramachandran 2019; Hamilton, Moitra 2021]
- Tyler の M 推定量 [Franks, Moitra 2020], 行列正規分布の共分散推定 [Dutilleul 1999; Drton, Kuriki, Hoff 2021]
- 計算的不変式論 [Allen-Zhu, Garg, Li, Oliveira, Wigderson 2018].

作用素スケーリング

入力: 行列 $A_1, \dots, A_m \in \mathbb{C}^{n \times n}$

出力: 正則行列 $L, R \in \text{GL}(n)$ s.t. $\tilde{A}_i := LA_iR^\dagger$ が次を満たす:

$$\sum_{i=1}^m \tilde{A}_i \tilde{A}_i^\dagger = I \quad \text{かつ} \quad \sum_{i=1}^m \tilde{A}_i^\dagger \tilde{A}_i = I$$

†: 複素共役転置

例

$$n = 2, m = 3$$

$$A_1 = \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 \\ \gamma & 0 \end{bmatrix}$$

とする。このとき、

$$\sum_i A_i A_i^\dagger = \begin{bmatrix} |\alpha|^2 + |\beta|^2 & 0 \\ 0 & |\gamma|^2 \end{bmatrix}, \quad \sum_i A_i^\dagger A_i = \begin{bmatrix} |\alpha|^2 + |\gamma|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}$$

である。

演習問題

上の例を参考にして、行列スケーリングを作用素スケーリングで表現せよ。

近似版作用素スケーリング

入力: 行列 $A_1, \dots, A_m \in \mathbb{C}^{n \times n}$, $\varepsilon > 0$

出力: 正則行列 $L_\varepsilon, R_\varepsilon \in \text{GL}(n)$ s.t.

$$\left\| \sum_{i=1}^m \tilde{A}_i \tilde{A}_i^\dagger - I \right\|_1 < \varepsilon \quad \text{かつ} \quad \left\| \sum_{i=1}^m \tilde{A}_i^\dagger \tilde{A}_i - I \right\|_1 < \varepsilon.$$

(ここで $\tilde{A}_i = L_\varepsilon A_i R_\varepsilon^\dagger$)

核ノルム (トレースノルム) : $\|M\|_1 := \sum_{i=1}^n |\lambda_i(M)|$

- **近似スケーリング可能** $\stackrel{\text{def}}{\iff} \forall \varepsilon > 0$ に対して近似スケーリング解が存在

完全正值写像

- 行列間の線形写像 $\Phi : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ と $\Phi^* : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ をそれぞれ

$$\Phi(X) = \sum_{i=1}^m A_i X A_i^\dagger, \quad \Phi^*(X) = \sum_{i=1}^m A_i^\dagger X A_i$$

と定める。(Φ は**完全正值 (CP) 写像**, Φ^* は Φ の双対写像と呼ばれる)

- 正則行列 L, R に対し, “スケーリングされた写像” $\Phi_{L,R}$ と $\Phi_{L,R}^*$ を

$$\Phi_{L,R}(X) = \sum_{i=1}^m \tilde{A}_i X \tilde{A}_i^\dagger \quad \Phi_{L,R}^*(X) = \sum_{i=1}^m \tilde{A}_i^\dagger X \tilde{A}_i$$

で定める。ここで $\tilde{A}_i = L A_i R^\dagger$ 。

- 作用素スケーリングは与えられた Φ に対し, $\Phi_{L,R}(I) = I, \Phi_{L,R}^*(I) = I$ となる正則行列 L, R を求める問題。

行列スケーリングとの対比

	行列スケーリング	作用素スケーリング
入力	非負行列 $A : \mathbb{R}_+^{n \times n}$	$A_i \in \mathbb{C}^{n \times n} (i \in [m])$ or CP 写像 $\Phi : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$
対応する錐	$\mathbb{R}_+^n : A\mathbb{R}_+^n \subseteq \mathbb{R}_+^n$	$\text{PSD}_n : \Phi(\text{PSD}_n) \subseteq \text{PSD}_n$
出力	L, R : 正対角行列	L, R : 正則行列
スケーリング	$\tilde{A} = LAR$	$\tilde{A}_i = LA_iR^\dagger$ or $\tilde{\Phi} = \Phi_{L,R}$
制約	$A\mathbf{1} = \mathbf{1}, A^\top \mathbf{1} = \mathbf{1}$	$\Phi(I) = I, \Phi^*(I) = I$

※ PSD_n : $n \times n$ 半正定値エルミート行列の集合

作用素 Sinkhorn アルゴリズム [Gurvits 2004]

$\Phi^*(I) = I$ と仮定して良い。次の作用素の反復 $\Phi^{(0)}, \Phi^{(1)}, \dots$ を考える。

$$\Phi^{(0)} = \Phi$$

$$\Phi^{(2t+1)} = \Phi_{L,I}^{(2t)}$$

ここで L は $L\Phi^{(2t)}(I)L^\dagger = I$ を満たす行列

$$\Phi^{(2t+2)} = \Phi_{I,R}^{(2t+1)}$$

ここで R は $R\Phi^{(2t+1)*}(I)R^\dagger = I$ を満たす行列

※ Cholesky 分解や行列平方根を使えば L, R は簡単に求められる

定理 (Gurvits (2004))

Φ が近似スケーリング可能ならば、 $\Phi^{(t)}$ は解に収束する。

連続最適化としての解釈

$X, Y: n \times n$ 正定値エルミート行列

$$f(X, Y) = X \bullet \Phi(Y) - \log \det X - \log \det Y$$

補題

$$\nabla_X f(X, Y) = \Phi(Y) - X^{-1}, \quad \nabla_Y f(X, Y) = \Phi^*(X) - Y^{-1}$$

(証明) $\nabla_X \log \det X = X^{-1}$ と $X \bullet \Phi(Y) = \Phi^*(X) \bullet Y$ から従う。 □

補題

$\nabla f(X, Y) = O \iff L^\dagger L = X, R^\dagger R = Y$ を満たす (L, R) が解

(証明) $\nabla_X f(X, Y) = O \iff \Phi(Y) = X^{-1} \iff \Phi_{I,R}(I) = X^{-1}$.
 $X = L^\dagger L$ より $X^{-1} = L^{-1} L^{-\dagger}$ であるから、両辺に左右から L, L^\dagger を掛けると
 $L \Phi_{I,R}(I) L^\dagger = \Phi_{L,R}(I) = I$. □

交互最適化との等価性

X, Y に対応するスケーリング: $L^\dagger L = X, R^\dagger R = Y$ として $\Phi_{L,R}$

交互最適化

$$X \leftarrow \operatorname{argmin}_{X \succ 0} f(X, Y)$$

$$Y \leftarrow \operatorname{argmin}_{Y \succ 0} f(X, Y)$$

作用素 Sinkhorn

$$\Phi \leftarrow \Phi_{L,I} \quad (L\Phi(I)L^\dagger = I)$$

$$\Phi \leftarrow \Phi_{I,R} \quad (R\Phi^*(I)R^\dagger = I)$$

どちらも同じ $\Phi^{(t)}$ を定める.

作用素スケーリングの基本定理

行列スケーリングにおける Sinkhorn-Knopp の定理の拡張として，以下が成り立つ．

定理 (Gurvits (2004))

以下は同値:

- ① $\inf f > -\infty$
- ② Φ は近似スケーリング可能
- ③ 任意の部分空間 U に対し $\dim U \leq \dim(\sum_i A_i U)$

以下，線形代数を用いた Gurvits の証明を説明する．

cf. Hadamard 空間上の凸解析を用いた証明 [\[Hirai 2023\]](#)

定理の証明(1/3)

① \implies ②: L の更新 (X の更新) を考える. (X, Y) に対応するスケーリングを $\tilde{\Phi} = \Phi_{L,R}$ とする.

$$f(L^\dagger X' L, Y) = X' \bullet \tilde{\Phi}(I) - \log \det X' - \log \det X - \log \det Y$$

$$f(L^\dagger X' L, Y) - f(X, Y) = X' \bullet \tilde{\Phi}(I) - \log \det X' - I \bullet \tilde{\Phi}(I)$$

X' について最小化すると, $X' = \tilde{\Phi}(I)^{-1}$ より

$$f(X^+, Y) - f(X, Y) = n + \log \det \tilde{\Phi}(I) - \text{tr} \tilde{\Phi}(I)$$

$$= \sum_{i=1}^n (1 + \log \lambda_i(\tilde{\Phi}(I)) - \lambda_i(\tilde{\Phi}(I)))$$

$$= -D_{KL}(\mathbf{1} : \lambda(\tilde{\Phi}(I)))$$

$\lambda(M)$: M の固有値を並べたベクトル

$$\geq -\frac{1}{2n} \|\mathbf{1} - \lambda(\tilde{\Phi}(I))\|_1^2$$

よって, $\|I - \tilde{\Phi}(I)\|_1 > \varepsilon$ である限り f が $\frac{\varepsilon^2}{2n}$ 以上改善.

定理の証明 (2/3)

② \implies ③: W.l.o.g. $\sum_i A_i^\dagger A_i = I$ としてよい. 任意の部分空間 U を取る. $W = \sum_i A_i U$ とする. U への直交射影行列を Π_U とすると,

$$\begin{aligned} \dim U &= \Pi_U \bullet I = \Pi_U^2 \bullet \sum_i A_i^\dagger A_i \\ &= \sum_i \|A_i \Pi_U\|_F^2 \leq \sum_i \|\Pi_W A_i\|_F^2 \\ &= \Pi_W \bullet \sum_i A_i A_i^\dagger = \Pi_W \bullet I + \Pi_W \bullet \left(I - \sum_i A_i A_i^\dagger \right) \\ &\leq \dim W + \varepsilon \end{aligned}$$

よって, $\varepsilon \rightarrow 0$ とすると $\dim U \leq \dim(\sum_i A_i U)$.

$$A_i = \begin{array}{c|c} & U \\ \hline W & \begin{array}{cc} * & * \\ * & O \end{array} \end{array}$$

定理の証明 (3/3)

③ \implies ①: 行列スケーリングに帰着. $X, Y \succ O$ の固有分解を

$$X = P \operatorname{Diag}(e^x) P^\dagger = \sum_i e^{x_i} p_i p_i^\dagger, \quad Y = Q \operatorname{Diag}(e^y) Q^\dagger = \sum_j e^{y_j} q_j q_j^\dagger$$

(P, Q : ユニタリ行列) とおき, $P, Q \in U(n)$, $x, y \in \mathbb{R}^n$ の最適化問題にする.

$$\begin{aligned} f(X, Y) &= X \bullet \Phi(Y) - \log \det X - \log \det Y \\ &= \sum_{i,j} p_i p_i^\dagger \bullet \Phi(q_j q_j^\dagger) e^{x_i + y_j} - \sum_i x_i - \sum_j y_j \\ &\quad := (a_{P,Q})_{ij} \end{aligned}$$

非負行列 $(a_{P,Q})_{ij}$ に対する行列スケーリングの目的関数!

仮定より, 任意の P, Q に対して $A_{P,Q}$ は Hall 条件を満たすので, 近似スケーリング可能. $\therefore \inf_{x,y \in \mathbb{R}^n} f_{A_{P,Q}}(x, y) > -\infty$.

ユニタリ行列全体 $U(n)$ はコンパクトなので $\inf_{P,Q \in U(n)} \inf_{x,y \in \mathbb{R}^n} f_{A_{P,Q}}(x, y) > -\infty$.

Sinkhorn アルゴリズムの反復回数解析

現在の解が制約を ε 以上破る \implies 次の反復で f が $\frac{\varepsilon^2}{2n}$ 以上改善する.

難しい部分: $\inf_{X, Y \succ 0} f(X, Y)$ の下界は? ([Gurvits 2004] は特殊な A_i のみ解析)

Breakthrough [Garg, Gurvits, Oliveira, Wigderson 2020] 代数的不変式論を用いた下界の証明

定理 (degree bound [Derksen 2001])

$\dim U \leq \dim(\sum_i A_i U) \iff \det(\sum_{i=1}^m A_i \otimes X_i) \neq 0$. ここで, X_i は各成分が可換変数の $n \times n$ 行列

背後にある数学: 自由斜体, 非可換ランク, 幾何学的不変式論 (cf. 平井@COSS2019)

定理 ([Garg, Gurvits, Oliveira, Wigderson 2020])

A_i が整数行列の場合, $\inf f = -O(n \log(Mmn))$. ここで, M は A_i の成分の絶対値の最大値.

長方作用素への拡張

入力: 行列 $A_1, \dots, A_m \in \mathbb{C}^{n_1 \times n_2}$

出力: 正則行列 $L \in GL(n_1), R \in GL(n_2)$ s.t.

$$\sum_{i=1}^m \tilde{A}_i \tilde{A}_i^\dagger = \frac{1}{n_1} I_{n_1} \quad \text{かつ} \quad \sum_{i=1}^m \tilde{A}_i^\dagger \tilde{A}_i = \frac{1}{n_2} I_{n_2}$$

ここで, $\tilde{A}_i = LA_iR^\dagger$.

†: 複素共役転置

Gurvits の定理や作用素 Sinkhorn アルゴリズムは長方作用素へ容易に拡張できる.

長方作用素への拡張

$X: n_1 \times n_1$ 正定値行列, $Y: n_2 \times n_2$ 正定値行列, $\Phi: \mathbb{C}^{n_2 \times n_2} \rightarrow \mathbb{C}^{n_1 \times n_1}$: CP 写像

$$f(X, Y) = X \bullet \Phi(Y) - \frac{1}{n_1} \log \det X - \frac{1}{n_2} \log \det Y$$

- $\nabla f(X, Y) = 0 \iff L^\dagger L = X, R^\dagger R = Y$ を満たす (L, R) が解
- 交互最適化として自然に Sinkhorn アルゴリズムが得られる

$$\Phi^{(2t+1)} \leftarrow \Phi_{L, I}^{(2t)} \quad (L\Phi^{(2t)}(I_{n_2})L^\dagger = \frac{1}{n_1}I_{n_1})$$

$$\Phi^{(2t+2)} \leftarrow \Phi_{I, R}^{(2t+1)} \quad (R\Phi^{(2t+1)*}(I_{n_1})R^\dagger = \frac{1}{n_2}I_{n_2})$$

応用: 行列正規分布の共分散推定

[Drton, Kuriki, Hoff 2021]

G : 各成分が標準正規分布の $d \times d$ 行列

L : $d \times d_1$ 行列, R : $d \times d_2$ 行列

に対して, $L^\top GR$ の分布を **行列正規分布** と呼び,

$\Sigma_1 := R^\top R$, $\Sigma_2 := L^\top L$ を共分散行列と呼ぶ.

$$\boxed{L^\top} \quad \boxed{G} \quad \boxed{R}$$

$\sim N(0, 1)$

対数尤度関数

データ A_1, \dots, A_m に対して, 対数尤度関数は

$$\ell(\Sigma_1, \Sigma_2) = -\frac{1}{2} \operatorname{tr} \left(\sum_{i=1}^m \Sigma_2^{-1} A_i^\top \Sigma_1^{-1} A_i \right) - \frac{d_2}{2} \log \det \Sigma_1 - \frac{d_1}{2} \log \det \Sigma_2 + \text{const.}$$

- $(X, Y) \leftarrow (\Sigma_1^{-1}, \Sigma_2^{-1})$ と変数変換すると長方作用素の $f(X, Y)$ と一致.
- 作用素 Sinkhorn は **flipflop 法** [Duttilleul 1999] として知られていた

一般の周辺分布への拡張 [Franks 2018]

入力: 行列 $A_1, \dots, A_m \in \mathbb{C}^{n_1 \times n_2}$, **周辺分布** $p \in \mathbb{R}_+^{n_1}$, $q \in \mathbb{R}_+^{n_2}$
(ただし $p_1 \geq \dots \geq p_{n_1} \geq 0$, $q_1 \geq \dots \geq q_{n_2} \geq 0$)

出力: 正則**上三角**行列 $L \in \text{GL}(n_1)$, $R \in \text{GL}(n_2)$ s.t.

$$\sum_{i=1}^m \tilde{A}_i \tilde{A}_i^\dagger = \text{Diag}(p) \quad \text{かつ} \quad \sum_{i=1}^m \tilde{A}_i^\dagger \tilde{A}_i = \text{Diag}(q)$$

ここで, $\tilde{A}_i = LA_iR^\dagger$.

†: 複素共役転置

※近似スケーリングの定義も同様

一般の周辺分布への拡張 [Franks 2018]

$$f(X, Y) = X \bullet \Phi(Y) - \sum_{i=1}^{n_1} (p_i - p_{i+1}) \log \Delta_i(X) - \sum_{j=1}^{n_2} (q_j - q_{j+1}) \log \Delta_j(Y)$$

ここで $\Delta_i(X)$ は X の i 番目の**首座小行列式**. また $p_{n_1+1} := 0, q_{n_2+1} := 0$.

演習問題

$\nabla f(X, Y) = 0 \iff L^\dagger L = X, R^\dagger R = Y$ を満たす上三角 (L, R) が解

Sinkhorn 反復 [Franks 2018]

$$\Phi^{(2t+1)} \leftarrow \Phi_{L,I}^{(2t)}$$

$$(L\Phi^{(2t)}(I)L^\dagger = \text{Diag}(p))$$

$$\Phi^{(2t+2)} \leftarrow \Phi_{I,R}^{(2t+1)}$$

$$(R\Phi^{(2t+1)*}(I)R^\dagger = \text{Diag}(q))$$

一般周辺分布に対する定理

- 部分空間 $U \leq \mathbb{C}^{n_1}$, $V \leq \mathbb{C}^{n_2}$ が Φ 安定 $\stackrel{\text{def}}{\iff} \Pi_U \bullet \Phi(\Pi_V) = 0$
- 標準旗 $F_i := \langle \mathbf{e}_1, \dots, \mathbf{e}_i \rangle$ ($i = 1, \dots, n_1$), $G_j := \langle \mathbf{e}_1, \dots, \mathbf{e}_j \rangle$ ($j = 1, \dots, n_2$)

定理 ([Franks 2018])

$p \in \mathbb{R}_+^{n_1}$, $q \in \mathbb{R}_+^{n_2}$ ($p_1 \geq \dots \geq p_{n_1} \geq 0$, $q_1 \geq \dots \geq q_{n_2} \geq 0$, $\sum_i p_i = \sum_j q_j =: N$) とする。以下は同値:

- 1 $\inf f > -\infty$
- 2 Φ は (p, q) へ近似スケールリング可能
- 3 任意の Φ 安定部分空間 (U, V) に対し,

$$\sum_{i=1}^{n_1} (p_i - p_{i+1}) \dim(U \cap F_i) + \sum_{j=1}^{n_2} (q_j - q_{j+1}) \dim(V \cap G_j) \leq N.$$

Part II のまとめ

入力 A_1, \dots, A_m : 行列

\longleftrightarrow CP 写像 $\Phi(X) = \sum_{i=1}^m A_i X A_i^\dagger$, 双対 CP 写像 $\Phi^*(X) = \sum_{i=1}^m A_i^\dagger X A_i$,

スケーリング $\tilde{A}_i := L A_i R^\dagger$ (L, R : 正則上三角行列) $\longleftrightarrow \tilde{\Phi}(X) := \sum_{i=1}^m \tilde{A}_i X \tilde{A}_i^\dagger$

スケーリング	関数 $f(X, Y)$	A の代数的条件
二重確率 $\tilde{\Phi}(I) = I, \tilde{\Phi}^*(I) = I$	$X \bullet \Phi(Y) - \log \det X - \log \det Y$	$\dim U \leq \dim(\sum_i A_i U)$
長方作用素 $\tilde{\Phi}(I) = \frac{I}{n_1}, \tilde{\Phi}^*(I) = \frac{I}{n_2}$	$X \bullet \Phi(Y) - \frac{\log \det X}{n_1} - \frac{\log \det Y}{n_2}$	$\frac{\dim U}{n_1} \leq \frac{\dim(\sum_i A_i U)}{n_2}$
周辺分布 $\tilde{\Phi}(I) = p, \tilde{\Phi}^*(I) = q$	(略)	(略)

U : 部分空間

講演概要

Part I: 行列スケーリング

- 行列スケーリングとは？
- Sinkhorn アルゴリズム，凸解析・組合せ最適化との関連，応用

Part II: 作用素スケーリング

- 作用素スケーリングとは？
- 作用素 Sinkhorn アルゴリズム，解析，応用

Part III: Brascamp-Lieb 不等式と作用素スケーリング

- Brascamp-Lieb 不等式と多面体
- Brascamp-Lieb 多面体上の最適化

ブロック構造のある作用素スケーリング

応用では、 Φ, Φ^* はブロック対角行列である場合が多い。

$$L = \begin{bmatrix} L_1 & & \\ & L_2 & \\ & & L_3 \end{bmatrix}, \Phi(X) = \begin{bmatrix} \Phi_1(X) & & \\ & \Phi_2(X) & \\ & & \Phi_3(X) \end{bmatrix}, \Phi^*(X) = \begin{bmatrix} \Phi_1^*(X) & & \\ & \Phi_2^*(X) & \\ & & \Phi_3^*(X) \end{bmatrix}, R = \begin{bmatrix} R_1 & \\ & R_2 \end{bmatrix}$$

補題

周辺分布つき作用素スケーリングにおいて、 Φ, Φ^* が上のようなブロック構造を持つとする。このとき、スケーリング行列 L, R も同じブロック構造を持つと仮定しても、近似スケーリング可能性は（一般の上三角 L, R と）変わらない。

(証明) Sinkhorn アルゴリズムは Φ, Φ^* と同じブロック構造を持つ L, R を生成するので。 □

Φ 安定部分空間 (U, V) も同様の直和構造を持つ。

Brascamp–Lieb 不等式

$\mathbf{B} = (B_1, \dots, B_m)$: $B_\ell \in \mathbb{R}^{n_\ell \times n}$... 行フルランク行列, $p = (p_1, \dots, p_m) \in \mathbb{R}_+^m$

定理 (Brascamp, Lieb (1976))

次を満たす定数 $\text{BL}(\mathbf{B}, p) \in (0, +\infty]$ が存在する:

任意の非負可測関数 $f_\ell : \mathbb{R}^{n_\ell} \rightarrow \mathbb{R}_+$ ($\ell = 1, \dots, m$) に対して,

$$\int_{\mathbb{R}^n} \prod_{\ell=1}^m (f_\ell(B_\ell x))^{p_\ell} dx \leq \text{BL}(\mathbf{B}, p) \cdot \prod_{\ell=1}^m \left[\int_{\mathbb{R}^{n_\ell}} f_\ell(x_\ell) dx_\ell \right]^{p_\ell}.$$

演習問題

BL 不等式は Hölder 不等式, Loomis–Whitney 不等式, etc. の一般化であることを示せ

Brascamp–Lieb 不等式

BL 不等式の成り立つ最良の $BL(\mathbf{B}, p)$ を **Brascamp–Lieb 定数** という。以下、 $BL(\mathbf{B}, p)$ は最良定数であるとする。

定理 (Lieb (2002))

$$BL(\mathbf{B}, p) = \sup_{X_1, \dots, X_m \succ O} \left[\frac{\prod_{\ell=1}^m \det(X_\ell)^{p_\ell}}{\det(\sum_{\ell=1}^m p_\ell B_\ell^\top X_\ell B_\ell)} \right]^{1/2}$$

よって

$$-2 \log BL(\mathbf{B}, p) = \inf_{X_1, \dots, X_m \succ O} \left[\log \det \left(\sum_{\ell=1}^m p_\ell B_\ell^\top X_\ell B_\ell \right) - \sum_{\ell=1}^m p_\ell \log \det(X_\ell) \right]$$

(後で使う)

Brascamp–Lieb 多面体

Brascamp–Lieb 多面体 [Bennett, Carbery, Christ, Tao 2008]

B_ℓ : $n_\ell \times n$ 行列, 行フルランク ($\ell = 1, \dots, m$)

部分空間

$$B = \left\{ x \in \mathbb{R}^m : \begin{array}{l} \sum_{\ell=1}^m x_\ell \dim(B_\ell V) \geq \dim(V) \quad (V \leq \mathbb{R}^n) \\ \sum_{\ell=1}^m x_\ell n_\ell = n \\ x \geq \mathbf{0} \end{array} \right\}$$

ランク k の BL 多面体 $\stackrel{\text{def}}{\iff} n_\ell = k \ (\ell = 1, \dots, m)$

- $p \in B \iff \text{BL}(\mathbf{B}, p) < +\infty$ (後述)
- 線形マトロイド基多面体, 線形共通基多面体, 分数線形マトロイドマッチング多面体など, **組合せ最適化の重要な多面体**を含む [Barthe 1998; Garg, Gurvits, Oliveira, Wigderson 2018; Franks, Soma, Goemans 2023]

BL不等式と作用素スケーリング

定理 ([Bennett, Carbery, Christ, Tao 2008; Garg, Gurvits, Oliveira, Wigderson 2018])

以下は同値:

- 1 $BL(\mathbf{B}, p) < +\infty$
- 2 任意の $\varepsilon > 0$ に対し, 正則行列 $g_\ell \in GL(n_\ell)$, $h \in GL(n)$ が存在して $\tilde{B}_\ell := g_\ell B_\ell h$ が次を満たす:

$$\|\tilde{B}_\ell \tilde{B}_\ell^\top - p_\ell I_{n_\ell}\|_1 < \varepsilon \quad (\ell = 1, \dots, m), \quad \left\| \sum_{\ell=1}^m \tilde{B}_\ell^\top \tilde{B}_\ell - I_n \right\|_1 < \varepsilon$$

- 3 $p \in B$, i.e., $\sum_{\ell=1}^m p_\ell n_\ell = n$ かつ $\sum_{\ell=1}^m p_\ell \dim(B_\ell V) \geq \dim(V)$ ($V \leq \mathbb{R}^n$)

目標: 周辺分布つき作用素スケーリングを用いて証明する

作用素スケーリングとの対応 (1/2)

CP 写像 $\Phi : \mathbb{C}^{n \times n} \rightarrow \bigoplus_{\ell=1}^m \mathbb{C}^{n_\ell \times n_\ell}$ を

$$\Phi(Y) := \begin{bmatrix} B_1 Y B_1^\top & & \\ & \ddots & \\ & & B_m Y B_m^\top \end{bmatrix}, \quad \Phi^* : \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_m \end{bmatrix} \mapsto \sum_{\ell=1}^m B_\ell^\top X_\ell B_\ell$$

と定める.

$$\bar{p} := (\underbrace{p_1 \dots p_1}_{n_1 \text{ 個}} \dots \underbrace{p_m \dots p_m}_{n_m \text{ 個}}) \text{ とすると, } \text{Diag}(\bar{p}) = \begin{bmatrix} p_1 I_{n_1} & & \\ & \ddots & \\ & & p_m I_{n_m} \end{bmatrix}.$$

よって, 条件② $\iff \Phi$ が周辺分布 $(r, q) = (\bar{p}, \mathbf{1})$ に近似スケーリング可能.

作用素スケーリングとの対応 (2/2)

$$f(X, Y) = X \bullet \Phi(Y) - \sum_{i=1}^{n_1} (r_i - r_{i+1}) \log \Delta_i(X) - \sum_{j=1}^{n_2} (q_j - q_{j+1}) \log \Delta_j(Y)$$

これを先の Φ と $(r, q) = (\bar{p}, \mathbf{1})$ に適用する. ほとんどの i, j で $r_i - r_{i+1} = 0$, $q_j - q_{j+1} = 0$ となることに注意すると,

$$f(X_1, \dots, X_m; Y) = \sum_{\ell=1}^m X_\ell \bullet B_\ell Y B_\ell^\top - \sum_{\ell=1}^m p_\ell \log \det X_\ell - \log \det Y.$$

さらに X_1, \dots, X_m を固定すると $Y = (\sum_{\ell=1}^m B_\ell^\top X_\ell B_\ell)^{-1}$ が最適なので,

$$\begin{aligned} \inf f(X, Y) &= \inf_X \left(\sum_{\ell=1}^m n_\ell - \sum_{\ell=1}^m p_\ell \log \det X_\ell + \log \det \left(\sum_{\ell=1}^m B_\ell^\top X_\ell B_\ell \right) \right) \\ &= \text{const.} - 2 \log \text{BL}(\mathbf{B}, p). \end{aligned} \quad (\text{Lieb の結果})$$

よって, $\inf f > -\infty \iff \text{BL}(\mathbf{B}, p) < +\infty$.

Brascamp–Lieb 多面体の導出 (1/2)

定理 (再掲)

Φ が (r, q) へ近似スケールリング可能 \iff

(1) $\sum_i r_i = \sum_j q_j =: N$

(2) 任意の Φ 安定部分空間 (U, V) に対し

$$\sum_{i=1}^{n_1} (r_i - r_{i+1}) \dim(U \cap F_i) + \sum_{j=1}^{n_2} (q_j - q_{j+1}) \dim(V \cap G_j) \leq N.$$

これを先の Φ と $(r, q) = (\bar{p}, \mathbf{1})$ に使う.

(1): $\sum_{\ell=1}^m p_\ell n_\ell = n.$

Brascamp–Lieb 多面体の導出 (2/2)

(2): 任意の Φ 安定部分空間 $U = \bigoplus_{\ell=1}^m U_\ell$, V に対し,

$$\sum_{\ell=1}^m p_\ell \dim U_\ell + \dim V \leq n.$$

Φ 安定 $\iff U_\ell \leq (B_\ell V)^\perp$ なので, V を固定すると, $U_\ell = (B_\ell V)^\perp$ が $\dim U_\ell$ を最大にする. よって, U_ℓ を消去して

$$\sum_{\ell=1}^m p_\ell (n_\ell - \dim(B_\ell V)) + \dim V \leq n. \quad (V \leq \mathbb{R}^n)$$

$\sum_{\ell} p_\ell n_\ell = n$ より, $\sum_{\ell} p_\ell \dim(B_\ell V) \geq \dim V$ となる. □

BL 多面体の階層

一般ランク



ランク 2 BL = 分数線形マトロイドマッチング多面体

[Franks, Soma, Goemans 2023]

BL 多面体の劣モジュラ性

Bennett, Carbery, Christ, Tao (2008)

$$B = \left\{ x \in \mathbb{R}^m : \begin{array}{l} \sum_{\ell=1}^m x_{\ell} \dim(B_{\ell}V) \geq \dim(V) \quad (V \leq \mathbb{R}^n) \\ \sum_{\ell=1}^m x_{\ell} n_{\ell} = n \\ x \geq \mathbf{0} \end{array} \right\}$$

↓ $V := V^{\perp}$ と書き換え

$$B = \left\{ x \in \mathbb{R}^m : \begin{array}{l} \sum_{\ell=1}^m x_{\ell} \dim(\text{row}(B_{\ell}) \cap V) \leq \dim(V) \quad (V \leq \mathbb{R}^n) \\ \sum_{\ell=1}^m x_{\ell} n_{\ell} = n \\ x \geq \mathbf{0} \end{array} \right\}$$

BL 多面体の劣モジュラ性

$$B = \left\{ x \in \mathbb{R}^m : \begin{array}{l} \sum_{\ell=1}^m x_{\ell} \dim(\text{row}(B_{\ell}) \cap V) \leq \dim(V) \quad (V \leq \mathbb{R}^n) \\ \sum_{\ell=1}^m x_{\ell} n_{\ell} = n \\ x \geq \mathbf{0} \end{array} \right\}$$

$V \leq \mathbb{R}^n$ に対し

$$a_{\ell}(V) = \dim(\text{row}(B_{\ell}) \cap V), \quad r(V) = \dim(V)$$

とおくと

- a_{ℓ} : 優モジュラ関数 $a_{\ell}(U) + a_{\ell}(V) \leq a_{\ell}(U + V) + a_{\ell}(U \cap V)$
- r : モジュラ関数 $r(U) + r(V) = r(U + V) + r(U \cap V)$

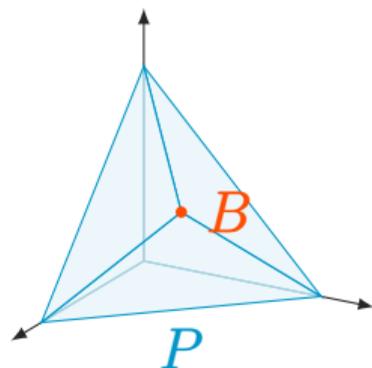
BL 多面体の劣モジュラ性

次の多面体 P を考える

$$P := \left\{ x \in \mathbb{R}^m : \sum_{\ell=1}^m a_{\ell}(V) x_{\ell} \leq r(V) \quad (V \leq \mathbb{R}^n) \right. \\ \left. x \geq \mathbf{0} \right\}$$

- $B = P \cap \{x : \sum_{\ell=1}^m a_{\ell}(\mathbb{R}^n) x_{\ell} = r(\mathbb{R}^n)\}$
- P : 二部マッチング多面体,
 B : 二部**完全**マッチング多面体に対応
- cf. ダイヤモンド直積上の劣モジュラ多面体

[Fujishige, Király, Makino, Takazawa, Tanigawa 2022]



Brascamp–Lieb 多面体上の線形最適化

重みベクトル $w \in \mathbb{R}^m$

$$\text{maximize } w^\top x \quad \text{subject to } x \in B \quad [\text{or } x \in P]$$

- 線形マトロイドの最大重み基，最大重み線形共通基，最大重み分数線形マトロイドマッチングなどの一般化
- **一般ランク:** 計算量は未解決
 - 作用素スケールリング+shrunk subspace による擬多項式時間分離オラクル [Garg, Gurvits, Oliveira, Wigderson 2018]
 - 多項式時間弱メンバーシップオラクル [Burgisser, Franks, Garg, Oliveira, Walter, Wigderson 2018]
 - 楕円体法を使わない直接的な擬多項式時間アルゴリズム [Franks, Soma, Goemans 2023]
- **ランク 2:** 重み付き非可換 Edmonds 問題による強多項式時間アルゴリズム [Hirai, Iwamasa, Oki, Soma 2023]

Part III のまとめ

- Brascamp–Lieb 不等式に現れる最良定数は作用素スケーリングにより計算可能
- $BL(\mathbf{B}, p) < +\infty$ となる p の全体 = Brascamp–Lieb 多面体.
- BL 多面体は，線形マトロイド基多面体，線形共通基多面体，分数線形マトロイドマッチング多面体など，組合せ最適化の重要な多面体を含む.
- 一般の BL 多面体上の線形最適化は未解決だが，ランク 2 以下の場合には強多項式時間アルゴリズムが存在する.

概要

- **スケーリング問題**は様々な分野に共通に現れる線形代数の問題
- **行列スケーリング**: 統計学, 機械学習, 組合せ最適化...
- **作用素スケーリング**: 組合せ最適化, 計算量理論, 非可換代数, 関数解析...
- **交互最適化** (Sinkhorn アルゴリズム) で解ける
- **Brascamp-Lieb 多面体**など, 組合せ最適化で面白い多面体が現れる

参考文献 I

- Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira, A. Wigderson (2018). “Operator Scaling via Geodesically Convex Optimization, Invariant Theory and Polynomial Identity Testing”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pp. 172–181.
- F. Barthe (1998). “On a reverse form of the Brascamp-Lieb inequality”. *Inventiones Mathematicae* 134.2, pp. 335–361.
- J. Bennett, A. Carbery, M. Christ, T. Tao (2008). “The Brascamp-Lieb inequalities: Finiteness, structure and extremals”. *Geometric and Functional Analysis* 17.5, pp. 1343–1415.
- H. J. Brascamp, E. H. Lieb (1976). “Best constants in Young’s inequality, its converse, and its generalization to more than three functions”. *Advances in Mathematics* 20.2, pp. 151–173.
- P. Burgisser, C. Franks, A. Garg, R. Oliveira, M. Walter, A. Wigderson (2018). “Efficient Algorithms for Tensor Scaling, Quantum Marginals, and Moment Polytopes”. In: *Proceedings of the IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 883–897.
- I. Csiszár (1975). “ I -divergence geometry of probability distributions and minimization problems”. *The Annals of Probability* 3, pp. 146–158.
- H. Derksen (2001). “Polynomial bounds for rings of invariants”. *Proceedings of the American Mathematical Society* 129.4, pp. 955–963.
- M. Drton, S. Kuriki, P. Hoff (2021). “Existence and Uniqueness of the Kronecker Covariance MLE”. *The Annals of Statistics* 49.5.
- P. Dutilleul (1999). “The MLE algorithm for the matrix normal distribution”. *Journal of Statistical Computation and Simulation* 64.2, pp. 105–123.
- C. Franks (2018). “Operator Scaling with Specified Marginals”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pp. 190–203.

参考文献 II

- C. Franks, A. Moitra (2020). “Rigorous Guarantees for Tyler’s M-estimator via Quantum Expansion”. In: *Proceedings of 33rd Conference on Learning Theory (COLT)*, pp. 1601–1632.
- C. Franks, T. Soma, M. X. Goemans (2023). “Shrunk subspaces via operator Sinkhorn iteration”. In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1655–1668.
- S. Fujishige, T. Király, K. Makino, K. Takazawa, S. Tanigawa (2022). “Minimizing Submodular Functions on Diamonds via Generalized Fractional Matroid Matchings”. *Journal of Combinatorial Theory, Series B* 157, pp. 294–345.
- A. Garg, L. Gurvits, R. Oliveira, A. Wigderson (2018). “Algorithmic and optimization aspects of Brascamp–Lieb inequalities, via Operator Scaling”. *Geometric and Functional Analysis* 28.1, pp. 100–145.
- (2020). “Operator Scaling: Theory and Applications”. *Foundations of Computational Mathematics*, pp. 223–290.
- T. T. Georgiou, M. Pavon (2015). “Positive contraction mappings for classical and quantum Schrödinger systems”. *Journal of Mathematical Physics* 56.3.
- L. Gurvits (2004). “Classical complexity and quantum entanglement”. *Journal of Computer and System Sciences* 69.3, pp. 448–484.
- L. Hamilton, A. Moitra (2021). “The Paulsen problem made simple”. *Israel Journal of Mathematics* 246.1, pp. 299–313.
- H. Hirai (2023). “Convex Analysis on Hadamard Spaces and Scaling Problems”. *Foundations of Computational Mathematics*.
- H. Hirai, Y. Iwamasa, T. Oki, T. Soma (2023). “Algebraic combinatorial optimization on the degree of determinants of noncommutative symbolic matrices”. *arXiv*.
- M. Idel (2016). “A review of matrix scaling and Sinkhorn’s normal form for matrices and positive maps”. *arXiv preprint arXiv:1609.06349*.

参考文献 III

- G. Ivanyos, Y. Qiao, K. V. Subrahmanyam (2017). “Non-commutative Edmonds’ problem and matrix semi-invariants”. *Computational Complexity* 26.3, pp. 717–763.
- (2018). “Constructive non-commutative rank computation is in deterministic polynomial time”. *Computational Complexity* 27.4, pp. 561–593.
- T. C. Kwok, L. C. Lau, A. Ramachandran (2019). “Spectral Analysis of Matrix Scaling and Operator Scaling”. In: *Proceedings of the 60th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1184–1204.
- T. C. Kwok, L. C. Lau, Y. T. Lee, A. Ramachandran (2018). “The Paulsen Problem, Continuous Operator Scaling, and Smoothed Analysis”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 182–189.
- E. H. Lieb (2002). “Gaussian kernels have only Gaussian maximizers”. In: *Inequalities*. Springer, pp. 595–624.
- R. Morioka, K. Tsuda (2011). *Information geometry of input-output table (in Japanese)*. Tech. rep. Technical Report IEICE, pp. 161–168.
- G. Peyré, M. Cuturi (2019). “Computational optimal transport”. *Foundations and Trends® in Machine Learning* 11.5-6, pp. 355–607.
- R. Sinkhorn, P. Knopp (1967). “Concerning nonnegative matrices and doubly stochastic matrices”. *Pacific Journal of Mathematics* 21, pp. 343–348.
- R. Sinkhorn (1964). “A relationship between arbitrary positive matrices and doubly stochastic matrices”. *The Annals of Mathematical Statistics* 35.2, pp. 876–879.
- A. G. Wilson (1969). “The use of entropy maximising models, in the theory of trip distribution, mode split and route split”. *Journal of transport economics and policy*, pp. 108–126.