

Hadamard 多様体上の非有界凸関数に対する 最急降下法と 作用素スケーリング問題への応用

平井広志 (名古屋大学) ○坂部圭哉 (東京大学)

RAOTA @ つくば 2024
2024 年 5 月 18 日

H. Hirai and K. Sakabe,
Gradient descent for unbounded convex functions on Hadamard manifolds and its
applications to scaling problems. arXiv:2404.09746

Hadamard 多様体

「完備」かつ「CAT(0) (単連結・非正曲率)」な Riemann 多様体 M

例：Euclid 空間 $\mathbb{R}^n$ ・双曲空間 $\mathbb{H}^n$ ・ $n \times n$ 正定値対称行列全体の集合 P_n

関数の測地的凸 / L -平滑性

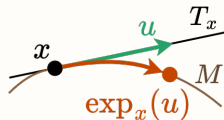
M : Hadamard 多様体

$f : M \rightarrow \mathbb{R}$ が凸 / L -平滑 $\iff$ 任意の測地線の上に制限した 1 変数関数が凸 / L -平滑

最急降下法

$x_{i+1} := \exp_{x_i}(-h\nabla f(x_i))$ ($\exp_x$: 指数写像, $h > 0$)

- L -平滑 $\implies$ ステップ幅 $h := 1/L$ (以下)
- **Fact:** $\exists$ 最適解 $\implies x_i \rightarrow$ 最適解 (e.g. [Boumal 2023])



研究目的

凸関数 f が下に非有界なとき (i.e. $\inf f = -\infty$), 最急降下法の振る舞いは？

結論：

- 勾配ノルムの収束
- x_i が飛んでいく無限遠の方向の特徴づけ (錐位相・後退関数)

応用：

- 行列スケーリング ([Hayashi, Hirai, and Sakabe, to appear])
- 作用素スケーリング (本発表)

Cartan–Hadamard の定理： $\exp_x : T_x(\text{接空間}) \rightarrow M$ は微分同相

Definition (錐位相 (e.g. [Sakai 1996]))

$M \cong T_x \cong \mathbb{R}^d \cong \text{開球}$ (同相)

M に境界 M^∞ を付け加えてコンパクト化 $M \cup M^\infty \cong \text{閉球}$

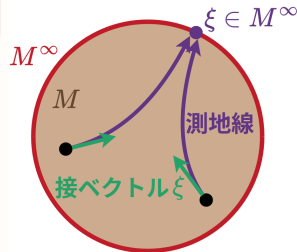
具体的には…

$S_x := \{u \in T_x \mid \|u\| = 1\}$

- M の点 $y = \exp_x(d(y, x)u)$ ($u \in S_x$): 極座標表示
- M^∞ の点 $\xi = \exp_x(\infty u)$ ($u \in S_x$): 方向 u ・ 距離 ∞ の点
 $\iff$ 測地的半直線 $t \mapsto \exp_x(tu)$ と同一視

$\implies \underline{S_x \cong M^\infty}$ を同一視

※ x の取り方に依らない



凸関数 $f : M \rightarrow \mathbb{R}$

Definition (後退関数 [Kapovich, Leeb, and Millson 2009])

後退関数 $f^\infty : M^\infty \rightarrow \mathbb{R} \cup \{\infty\} \stackrel{\text{def.}}{\iff}$

$$f^\infty(\xi) := \lim_{t \rightarrow \infty} \frac{f(\exp_x(t\xi)) - f(x)}{t}$$

- 「 ξ 方向に無限に進んだときの f の増加率」
- 始点 x の取り方に依存せず M^∞ 上で定義

Obs: $\inf_{\xi \in M^\infty} f^\infty(\xi) < 0 \implies \inf_{x \in M} f(x) = -\infty$

- $f^\infty(\xi^*) = \inf_{\xi \in M^\infty} f^\infty$ が唯一存在 [Kapovich, Leeb, and Millson 2009]

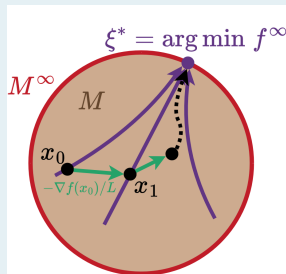
f : Hadamard 多様体上の L -平滑凸関数

$\inf_{\xi \in M^\infty} f^\infty(\xi) < 0$ ($\iff \inf_{x \in M} \|\nabla f(x)\| > 0$)

$\{x_i\}$: 最急降下法 (ステップ幅 $h = 1/L$) による点列

Theorem (本研究)

- $\|\nabla f(x_i)\| \rightarrow \inf_{x \in M} \|\nabla f(x)\|$
- $x_i \rightarrow \xi^*$ (錐位相での収束)



等式が成立：

$$\lim_{i \rightarrow \infty} \|\nabla f(x_i)\| = \inf_{x \in M} \|\nabla f(x)\| = \sup_{\xi \in M^\infty} -f^\infty(\xi) = -f^\infty\left(\lim_{i \rightarrow \infty} x_i\right)$$

作用素スケーリング問題 [Gurvits 2004]

入力: N 個の $n \times n$ 行列 $A = \{A_k\}_{k=1,2,\dots,N}$

出力: 正則行列 g, h で, 以下を満たすもの:

$$\left\| \sum_{k=1}^N g A_k h^\dagger h A_k^\dagger g^\dagger - I \right\|_F = 0 \quad \left\| \sum_{k=1}^N h A_k^\dagger g^\dagger g A_k h^\dagger - I \right\|_F = 0 \quad (I \text{ は単位行列})$$

- 非可換 **Edmonds 問題** ([Ivanyos, Qiao, and Subrahmanyam 2017]) と関連
- 群軌道上ノルム最小化の一例 (cf. Invariant theory (e.g. [Derksen and Kemper 2015]))
- g, h を対角行列に制限 $\implies$ **行列スケーリング問題**

Definition

$\{A_k\}$ が **スケール可能** $\iff$ 任意の $\varepsilon > 0$ に対し, 上式の $\|\cdot\|_F < \varepsilon$ を満たす解 g, h が存在

- ユニタリ行列によって $g \mapsto Ug, h \mapsto Vh$ と変換しても $\|\cdot\|_F$ は不変
→ 正定値行列 $X := g^\dagger g, Y := h^\dagger h$ 上の問題とみなせる

$$\text{minimize}_{X,Y \in P_n} \quad F_A(X,Y) := \log \left(\text{tr} \sum_{k=1}^N X A_k Y A_k^\dagger \right) - \frac{1}{n} \log \det X - \frac{1}{n} \log \det Y$$

- 勾配 $\nabla_X F_A(X,Y) = \frac{1}{n} \left(\sum_{k=1}^N \sqrt{X} A_k Y A_k^\dagger \sqrt{X} - I \right)$ ($\nabla_Y F_A$ も同様)
→ 作用素スケーリング問題は、 F_A の停留点を求める問題

○ F_A は、下に非有界となり得る：

Proposition ([Gurvits 2004])

A がスケール可能 $\iff \inf F_A(X,Y) > -\infty$

標準的なアルゴリズム：作用素 Sinkhorn アルゴリズム ([Gurvits 2004])：

「 X の最適化」と「 Y の最適化」を交互反復

ただし、スケール不可能なときの漸近挙動は未説明 [Garg and Oliveira 2018]

→ **本研究：最急降下法を用いた場合の漸近挙動を解明**

→ 一般化 Dulmage–Mendelsohn 分解による特徴づけを得た

☆行列スケーリングでは両手法の挙動は本質的に同じ [Hayashi, Hirai, and Sakabe, to appear]

準備：消滅空間

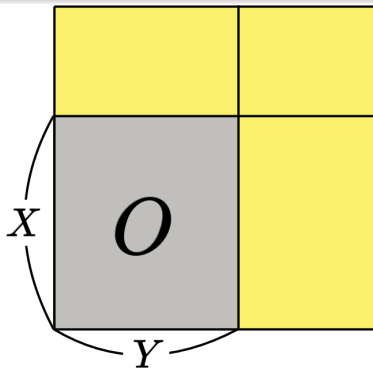
8/14

$A = \{A_k\}$: 作用素スケーリングの入力

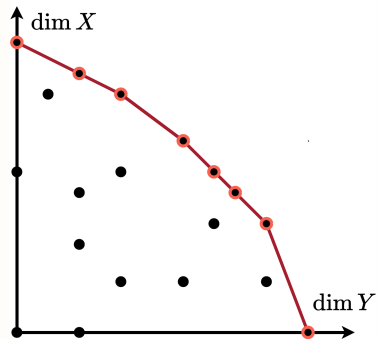
Definition (消滅空間)

$X \subseteq \mathbb{C}^n, Y \subseteq \mathbb{C}^n$ が **消滅空間** $\stackrel{\text{def}}{\iff} x^\top A_k y = 0 \ (\forall x \in X, \forall y \in Y, \forall k = 1, 2, \dots, N)$
 $\iff$ 「 X や Y を含む基底で A_k を表記したとき,
 X, Y に含まれる A_k の要素がすべて 0」

$A_k =$

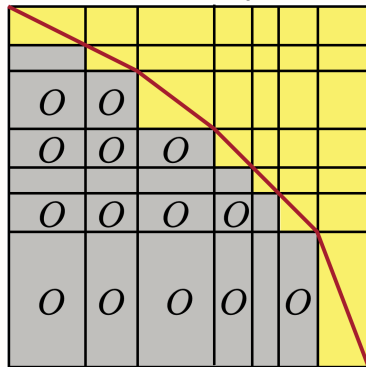


$(\dim X, \dim Y)$ が極大になる消滅空間 (X, Y) たちはモジュラ束 [Hamada and Hirai 2017]
 → 極大鎖 $(\mathbb{C}^n, \{0\}) = (X_0, Y_0) > \cdots > (X_\theta, Y_\theta) = (\{0\}, \mathbb{C}^n)$ をとって基底変換する
 → $A_1, \dots, A_N$ は同時ブロック三角化 (一般化 Dulmage–Mendelsohn 分解)



$(\dim X, \dim Y)$ の図

$$A_k =$$



対応する一般化 DM 分解

- 通常の DM 分解 [Dulmage and Mendelsohn 1958] は、斜め 45° 部分のみ分割
- $N = 2$ (行列 2 つ) のとき、Kronecker 標準形 [Gantmacher 1959] に対応 (本研究)

$\{(x_i, y_i)\}_i$: $F_A(X, Y)$ に対する最急降下法による点列

Theorem 1 (本研究)

$\{x_i^{1/2} A_k y_i^{1/2}\}_k$ は射影空間で, 一般化 DM 分解に対応した**ブロック対角行列**に

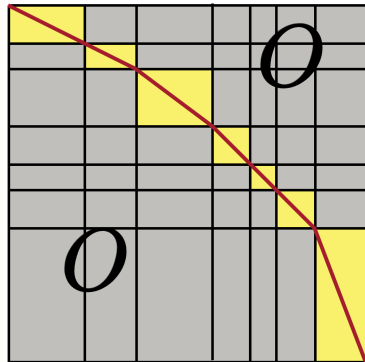
ユニタリ同値を除いて収束

$\iff$ 適切な U_i, V_i により $\{U_i x_i^{1/2} A_k y_i^{1/2} V_i^\dagger\}_k$ が収束

- ξ^* (F_A^∞ 最小化元) の DM 分解による特徴付けと,
 $\lim_{i \rightarrow \infty} \|\nabla f(x_i)\| = \inf_{x \in M} \|\nabla f(x)\| =$
 $\sup_{\xi \in M^\infty} -f^\infty(\xi) = -f^\infty(\lim_{i \rightarrow \infty} x_i)$ による.

△ 適切なユニタリ行列の発見が難しい

→ 数値実験によると, x_i や y_i を固有値分解すればよい

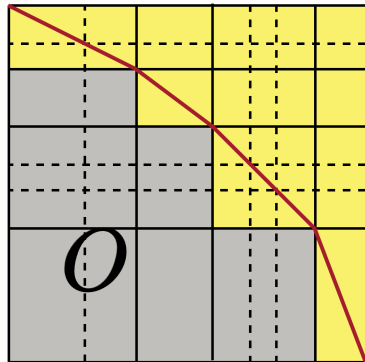


$\{(x_i, y_i)\}_i$: $F_A(X, Y)$ に対する最急降下法による点列

Theorem 2 (本研究)

固有値分解 $x_i = \sigma_i^\dagger \exp(\text{diag } p^i) \sigma_i$ (降順), $y_i = \tau_i^\dagger \exp(\text{diag } q^i) \tau_i$ (昇順)
 $\implies \left\{ \sigma_i A_k \tau_i^\dagger \right\}_k$ は, 粗い一般化 DM 分解に収束 (線形収束)

- 傾き (列数) / (行数) が等しい対角ブロックをまとめる
→ 粗い一般化 DM 分解
- $x_i \rightarrow \xi^*$ と F_A の減少量解析による
- 線形空間の rounding [Franks, Soma, and Goemans 2023] が利用可能だろう
→ 非可換 Edmonds 問題に関する certificate



- Hadamard 多様体上の下に非有界な凸関数に対する最急降下法
→勾配ノルムの収束・点列の錐位相での収束
- 作用素スケーリング：スケール不可能な行列の場合
→行列の収束・収束先の DM 分解による特徴づけ
- 今後の課題
 - 作用素スケーリング：2 つの収束性の統合
 - 作用素 Sinkhorn アルゴリズムでは？
 - Kronecker 標準形を計算するアルゴリズム
- 本日の内容の詳細：プレプリント
"Gradient descent for unbounded convex functions on Hadamard manifolds and its applications to scaling problems." (arXiv:2404.09746)

- Beck, A. (2017). *First-Order Methods in Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics. doi: 10.1137/1.9781611974997. eprint: <https://epubs.siam.org/doi/pdf/10.1137/1.9781611974997>.
- Boumal, N. (2023). *An introduction to optimization on smooth manifolds*. Cambridge University Press, Cambridge, pp. xviii+338.
- Derksen, H. and G. Kemper (2015). *Computational Invariant Theory*. Springer Berlin, Heidelberg.
- Dulmage, A. L. and N. S. Mendelsohn (1958). "Coverings of bipartite graphs". *Canadian J. Math.*, 10, pp. 517–534. doi: 10.4153/CJM-1958-052-0.
- Fortin, M. and C. Reutenauer (2004). "Commutative/noncommutative rank of linear matrices and subspaces of matrices of low rank.". eng. *Séminaire Lotharingien de Combinatoire [electronic only]*, 52, B52f, 12 p., electronic only–B52f, 12 p., electronic only.
- Franks, C., T. Soma, and M. X. Goemans (2023). "Shrunk subspaces via operator Sinkhorn iteration". *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, Philadelphia, PA, pp. 1655–1668. doi: 10.1137/1.9781611977554.ch62.
- Gantmacher, F. R. (1959). *The theory of matrices*. Vols. 1, 2. Translated by K. A. Hirsch. Chelsea Publishing Co., New York, Vol. 1, x+374 pp. Vol. 2, ix+276.
- Garg, A. and R. M. de Oliveira (2018). "Recent progress on scaling algorithms and applications". *CoRR*, abs/1808.09669. arXiv: 1808.09669.
- Gurvits, L. (2004). "Classical complexity and quantum entanglement". *J. Comput. System Sci.*, 69.3, pp. 448–484. doi: 10.1016/j.jcss.2004.06.003.
- Hamada, M. and H. Hirai (2017). *Maximum vanishing subspace problem, CAT(0)-space relaxation, and block-triangularization of partitioned matrix*. arXiv:1705.02060, 2017.
- Hamada, M. and H. Hirai (2021). "Computing the nc-rank via discrete convex optimization on CAT(0) spaces". *SIAM J. Appl. Algebra Geom.*, 5.3, pp. 455–478. doi: 10.1137/20M138836X.

- Ivanyos, G., Y. Qiao, and K. V. Subrahmanyam (Sept. 2017). “Non-commutative Edmonds’ problem and matrix semi-invariants”. *computational complexity*, 26.3, pp. 717–763. doi: 10.1007/s00037-016-0143-x.
- Ivanyos, G., Y. Qiao, and K. V. Subrahmanyam (Dec. 2018). “Constructive non-commutative rank computation is in deterministic polynomial time”. *computational complexity*, 27.4, pp. 561–593. doi: 10.1007/s00037-018-0165-7.
- Kapovich, M., B. Leeb, and J. Millson (2009). “Convex functions on symmetric spaces, side lengths of polygons and the stability inequalities for weighted configurations at infinity”. *J. Differential Geom.*, 81.2, pp. 297–354.
- Sakai, T. (1996). *Riemannian geometry*. Vol. 149. Translations of Mathematical Monographs. Translated from the 1992 Japanese original by the author. American Mathematical Society, Providence, RI, pp. xiv+358. doi: 10.1090/mmono/149.