

最適化：モデリングとアルゴリズム @政策研究大学院大学

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

○坂部圭哉 (東京大学大学院情報理工学系研究科)
平井広志 (名古屋大学多元数理科学研究科)

2024 年 3 月 21 日

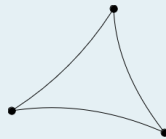
- ① Hadamard 多様体上の測地的凸関数
- ② 作用素スケーリング問題への応用
- ③ まとめ

- ① Hadamard 多様体上の測地的凸関数
- ② 作用素スケーリング問題への応用
- ③ まとめ

Definition (Hadamard 多様体)

Hadamard 多様体：以下を満たす Riemann 多様体 M

- 完備
- 単連結
- 断面曲率が非正 (三角形の内角の和が 180° 以下)



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

測地線に関する著しい性質：

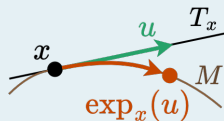
- 任意の 2 点間に唯一の測地線 (長さ最小の曲線) が存在
- 測地線は、前後に無限に延伸可能

指数写像

T_x : 点 $x \in M$ での接空間

指数写像 $\exp_x : T_x \rightarrow M \stackrel{\text{def.}}{\iff}$

$\exp_x(u) :=$ 「 u の方向に測地線に沿って距離 $\|u\|$ だけ進んだ点」



Definition (測地的凸・ L -平滑)

M : Hadamard 多様体

- $f : M \rightarrow \mathbb{R}$ が (測地的) 凸関数
 - $\iff$ 任意の測地線上に制限した 1 変数関数が凸
 - $\iff$ Hessian $\nabla^2 f$ が半正定値
- 凸関数 $f : M \rightarrow \mathbb{R}$ が L -平滑 ($L > 0$)
 - $\iff$ 任意の測地線上に制限した 1 変数関数が L -平滑
 - $\iff$ Hessian $\nabla^2 f$ の最大固有値が L 以下

Definition (最急降下列: 最急降下法による点列)

点列 $\{x_i\}$ が最急降下列 $\iff x_{i+1} := \exp_{x_i}(-h\nabla f(x_i))$ ($h > 0$: ステップ幅)

- L -平滑 $\implies$ ステップ幅 $h := 1/L$ (以下)
- $\exists$ 最適解 $\implies x_i \rightarrow$ 最適解 (e.g. [Zhang and Sra 2016])

研究目的: f が下に非有界なとき (i.e. $\inf f = -\infty$), 最急降下列の振る舞いは?

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

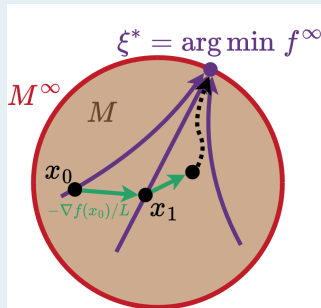
f : Hadamard 多様体上の L -平滑凸関数

$\inf_{x \in M} \|\nabla f(x)\| > 0$ ($\inf f = -\infty$ より強い条件)

$\{x_i\}$: 最急降下列 (ステップ幅 $h = 1/L$)

主結果

- 勾配ノルムの収束：
 $\|\nabla f(x_i)\| \rightarrow \inf_{x \in M} \|\nabla f(x)\|$
- $\{x_i\}$ は無限遠に飛んでいく.
その飛んでいく方向の特徴付け：
 $x_i \rightarrow \xi^*$ (錐位相での収束)
(ξ^* は後退関数 f^∞ の最小化元)



錐位相・後退関数 f^∞ を次頁以降で説明する：

Cartan–Hadamard の定理 : $\exp_x : T_x \rightarrow M$ 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 $M \cup M^\infty \cong \text{閉球}$

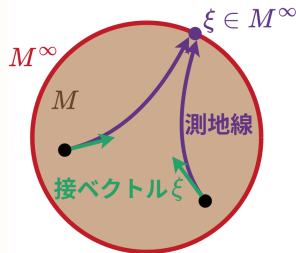
具体的には…

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

- 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]

$$\inf_{\xi \in M^\infty} f^\infty(\xi) < 0$$

Lemma (弱双対性・本研究)

$$\sup_{\xi \in M^\infty} -f^\infty(\xi) \leq \inf_{x \in M} \|\nabla f(x)\|.$$

証明： $\forall x \in M, \xi \in M^\infty$ について

$$f^\infty(\xi) = \lim_{t \rightarrow \infty} \frac{f(\exp_x(t\xi)) - f(x)}{t} \geq \lim_{t \rightarrow 0} \frac{f(\exp_x(t\xi)) - f(x)}{t} = \langle \nabla f(x), \xi \rangle \geq -\|\nabla f(x)\|$$

- 強双対性 (等号成立) は？ → 最急降下法の極限で達成されることを示す。

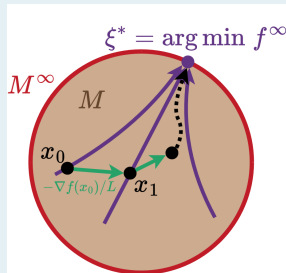
f : Hadamard 多様体上の L -平滑凸関数

$$\inf_{x \in M} \|\nabla f(x)\| > 0$$

$\{x_i\}$: f の最急降下列 (ステップ幅 $h = 1/L$)

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)



強双対性：

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

以下の不等式が重要：

Lemma (本研究)

$$f(x_{i+1}) - f(x_i) \leq -\frac{\|\nabla f(x_{i+1})\|^2}{L}$$

Remarks:

- 勾配流 $\dot{x}(t) = -\nabla f(x(t))$ における $\dot{f}(x(t)) = -\|\nabla f(x(t))\|^2$ に相当
- $f(x_{i+1}) - f(x_i) \leq -\frac{\|\nabla f(x_i)\|^2}{2L}$ は well-known だが, $\|\nabla f(x_{i+1})\|$ を用いた評価が重要
- Euclid 空間 $\mathbb{R}^d$ の場合は, L -平滑凸と同値な不等式 (e.g. [Beck 2017])

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + \frac{\|\nabla f(x) - \nabla f(y)\|^2}{2L} \quad (\forall x, y \in \mathbb{R}^d)$$

に $y = x_i, x = x_{i+1}$ を代入することで, より強い評価が得られる：

$$f(x_{i+1}) - f(x_i) \leq -\frac{\|\nabla f(x_i)\|^2 + \|\nabla f(x_{i+1})\|^2}{2L}.$$

ベクトルの差分は，Hessian ベクトル積の積分：

$$\begin{aligned}
 \tau_{x_{i+1} \rightarrow x_i} \nabla f(x_{i+1}) &= \nabla f(x_i) + \int_0^1 \tau_{x_i^{(\lambda)} \rightarrow x_i} \nabla^2 f(x_i^{(\lambda)}) \tau_{x_i \rightarrow x_i^{(\lambda)}} (-\nabla f(x_i)/L) d\lambda \\
 &= \left(I - \int_0^1 \frac{1}{L} \tau_{x_i^{(\lambda)} \rightarrow x_i} \nabla^2 f(x_i^{(\lambda)}) \tau_{x_i \rightarrow x_i^{(\lambda)}} d\lambda \right) \nabla f(x_i) \\
 &=: \textcolor{brown}{B} \nabla f(x_i) \quad \boxed{\textcolor{brown}{B} \text{ は半正定値対称, } \|\textcolor{brown}{B}\|_{\text{op}} \leq 1}
 \end{aligned}$$

$x_i^{(\lambda)} :=$ 測地線 $x_i \rightarrow x_{i+1}$ 上の $\lambda : (1 - \lambda)$ 内分点 $(x_i^{(0)} = x_i, x_i^{(1)} = x_{i+1})$

$$\begin{aligned}
 f(x_{i+1}) - f(x_i) &= \int_0^1 \left\langle -\nabla f(x_i)/L, \tau_{x_i^{(\lambda)} \rightarrow x_i} \nabla f(x_i^{(\lambda)}) \right\rangle d\lambda && \text{(直線積分)} \\
 &\leq \left\langle -\nabla f(x_i)/L, \tau_{x_{i+1} \rightarrow x_i} \nabla f(x_{i+1}) \right\rangle && \text{(凸性)} \\
 &= -\langle \nabla f(x_i), \textcolor{brown}{B} \nabla f(x_i) \rangle / L && \text{(}\textcolor{brown}{B}\text{の定義)} \\
 &\leq -\langle \nabla f(x_i), \textcolor{brown}{B} \textcolor{brown}{B} \nabla f(x_i) \rangle / L = -\|\nabla f(x_{i+1})\|^2 / L && \text{(}\textcolor{brown}{B}\text{の性質)}
 \end{aligned}$$

- 勾配ノルムの非増大性 $\|\nabla f(x_{i+1})\| \leq \|\textcolor{brown}{B}\|_{\text{op}} \|\nabla f(x_i)\| \leq \|\nabla f(x_i)\|$

$\kappa := \lim_{i \rightarrow \infty} \|\nabla f(x_i)\|$ ($\because \{\|\nabla f(x_i)\|\}$ は減少列)

$$\frac{\kappa^2}{L} = \lim_{i \rightarrow \infty} \frac{\|\nabla f(x_i)\|^2}{L} \leq \lim_{i \rightarrow \infty} -\frac{f(x_i) - f(x_0)}{i} \quad (\text{先ほどの Lemma})$$

$$\leq \left(\lim_{i \rightarrow \infty} \frac{d(x_0, x_i)}{i} \right) \sup_{\xi \in M^\infty} -f^\infty(\xi) \quad (f^\infty \text{ の定義})$$

$$\leq \left(\lim_{i \rightarrow \infty} d(x_i, x_{i+1}) \right) \sup_{\xi \in M^\infty} -f^\infty(\xi) \quad (\text{三角不等式})$$

$$\leq \frac{\kappa}{L} \inf_{x \in M} \|\nabla f(x)\| \quad (\text{最急降下列} \cdot \text{弱双対性})$$

$$\leq \frac{\kappa^2}{L} \quad (\kappa \text{ の定義})$$

より等号成立．特に

- $\bullet \quad \kappa = \lim_{i \rightarrow \infty} \|\nabla f(x_i)\| = \inf_{x \in M} \|\nabla f(x)\| = \sup_{\xi \in M^\infty} -f^\infty(\xi)$ (強双対性・勾配ノルムの収束)

- $\bullet \quad 2 \text{ 行目の等号成立} \implies \lim_{i \rightarrow \infty} \frac{f(x_i) - f(x_0)}{d(x_0, x_i)} = f^\infty(\xi^*) \text{ (i.e. 最小値)} \implies x_i \rightarrow \xi^*$

- ① Hadamard 多様体上の測地的凸関数
- ② 作用素スケーリング問題への応用
- ③ まとめ

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 問題と関連 (後述)
- 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 $g \mapsto Ug, h \mapsto Vh$ と変換しても不変
→ 正定値行列 $X := g^\dagger g, Y := h^\dagger h$ 上の問題とみなせる

測地的凸関数 $F_A(X, Y)$ ($X, Y \in P_n$) の最適化

- $\nabla_X F_A(X, Y) = \frac{1}{n} \left(\sum_{k=1}^N \sqrt{X} A_k Y A_k^\dagger \sqrt{X} - I \right)$ ($\nabla_Y F_A$ 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挙動は未解明 [Garg and Oliveira 2018]

→ **本研究: 最急降下法を用いた場合の挙動を解明**

→ 一般化 Dulmage–Mendelsohn 分解による特徴づけを得た

☆行列スケーリング問題では, 両アルゴリズムの漸近挙動は本質的に同じ
[Hayashi, Hirai, and Sakabe, to appear]

$A = \{A_k\}$: 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」

消滅空間族 $\mathcal{S}_A := (A \text{ の消滅空間全体の集合})$

Lemma[Hamada and Hirai 2017]

$\mathcal{S}_A$ はモジュラ束

$$(X, Y) \wedge (X', Y') := (X \cap X', Y + Y'), \quad (X, Y) \vee (X', Y') := (X + X', Y \cap Y')$$

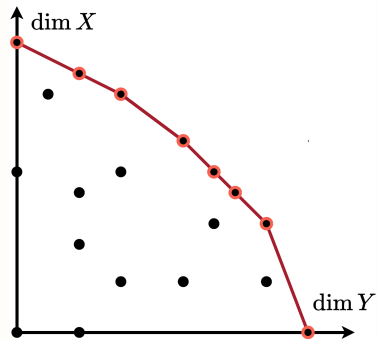
(証明：演算結果が $\mathcal{S}_A$ に入る・線型空間が $(\cap, +)$ によりモジュラ束をなす)

一般化 Dulmage–Mendelsohn 分解 [Hamada and Hirai 2017] 17/24

$(\dim X, \dim Y)$ が極大になる $(X, Y) \in \mathcal{S}_A$ たちは部分モジュラ束

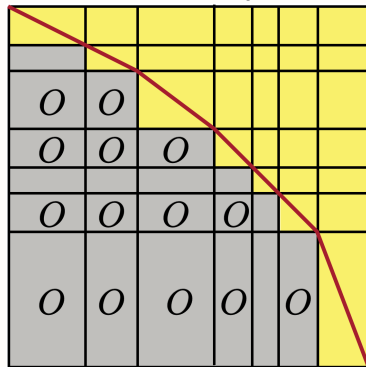
極大鎖 $(\mathbb{C}^n, \{0\}) = (X_0, Y_0) > \cdots > (X_\theta, Y_\theta) = (\{0\}, \mathbb{C}^n)$ をとって基底変換する

$\rightarrow A_1, \dots, A_N$ 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.

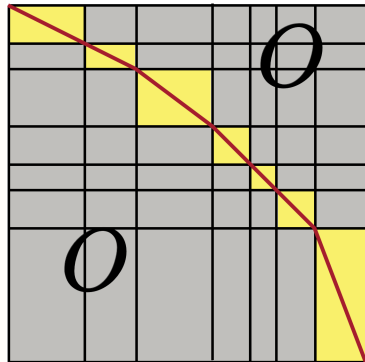
$\{(x_i, y_i)\}_i$: $F_A(X, Y)$ に関する最急降下列

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

- $\|\nabla f(x_i)\| \rightarrow \inf \|\nabla f\|$ 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

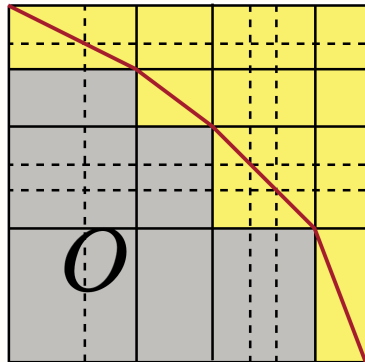


$\{(x_i, y_i)\}_i: F_A(X, Y)$ に関する最急降下列

Theorem 2 (本研究)

固有値分解 $x_i = \sigma_i^\dagger \exp(\text{diag } p^i) \sigma_k$ (降順), $y_i = \tau_i^\dagger \exp(\text{diag } q^i) \tau_k$ (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] で, 対応する消滅空間が求まる
→ 非可換 Edmonds 問題に関する certificate (次頁)



非可換 Edmonds 問題 [Ivanyos, Qiao, and Subrahmanyam 2017]

変数付き行列 $A = x_1 A_1 + \cdots + x_N A_N$ のランク (非可換ランク) は？
 $\{x_k\}$ は非可換変数 (i.e. $x_k x_l \neq x_l x_k$)

(非可換ランク) $= 2n - (\max_{(X,Y) \in \mathcal{S}_A} \dim X + \dim Y)$ [Fortin and Reutenauer 2004]
そこで以下を考える：

Problem

$\max_{(X,Y) \in \mathcal{S}_A} \dim X + \dim Y$ を達成する (X, Y) を求めよ

- 多項式アルゴリズム：存在するが複雑 ([Ivanyos, Qiao, and Subrahmanyam 2018; Franks, Soma, and Goemans 2023]) ([Hamada and Hirai 2021] ← ビットサイズ大)

本研究の帰結：単純な最急降下法により解ける

- TODO: 計算量解析—多項式か？
- 群軌道上ノルムに関する unstable 1-parameter subgroup 問題に一般化

- ① Hadamard 多様体上の測地的凸関数
- ② 作用素スケーリング問題への応用
- ③ まとめ

- Hadamard 多様体上の下に非有界な凸関数への最急降下法
 - 非常に一般的な設定
 - 勾配ノルムの収束
 - 点は無限遠へ飛んでいくが、その方向は収束
 - 強双対性 $\sup -f^\infty = \inf \|\nabla f\|$
- 作用素スケーリング：スケール不可能な行列の場合
 - ブロック対角行列に収束 (ユニタリ同値を除く)
 - 粗い DM 分解に収束 (線形収束)
 - 一般化 DM 分解の構造により特徴づけ
- 今後の課題
 - ユニタリ同値性の解消 → 数値実験ではうまくいく
 - 非可換 Edmonds 問題に対応する部分空間の計算量解析：多項式？
 - 作用素 Sinkhorn アルゴリズムで同様の収束性があるか
 - 群軌道上ノルム最小化問題への一般化
 - Kronecker 標準形の数値計算への応用

- 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>.
- 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.
- Zhang, H. and S. Sra (2016). *First-order Methods for Geodesically Convex Optimization*. arXiv: 1602.06053 [math.OC].