Optimization, Theoretical Computer Science and Algebraic Geometry: Convexity and Beyond

Gradient Descent for Unbounded Convex Functions on Hadamard Manifolds and Its Applications to Scaling Problems

Keiya Sakabe: The University of Tokyo (→ Ruhr-Universität Bochum)

RIMS & MFO
18th, Feb. 2025

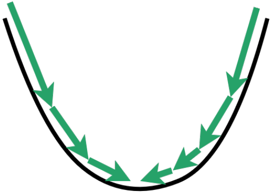
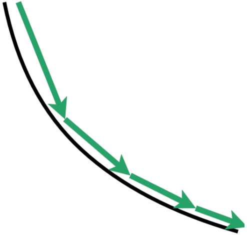
Paper for This Talk

“Gradient Descent for Unbounded Convex Functions on Hadamard Manifolds and Its Applications to Scaling Problems.”

Joint work with Hiroshi Hirai (Nagoya University)

- FOCS 2024 (Proc. pp. 2387–2402)
- arXiv: 2404.09746

Unbounded Convex Optimization

Usual Convex Optimization	**Unbounded Convex Optimization**
	
Convergence Meaningful!	**Divergence** **Meaningless?**

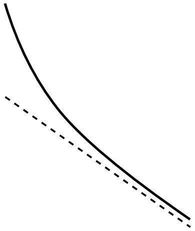
Goal: To understand and make use of such diverging sequences

Background: The objective function can be unbounded in:

- Geometric programming
- Matrix scaling
- Operator scaling
- Non-commutative optimization [Bürgisser et al., FOCS'19], (cf. Geometric Invariant Theory)

Unbounded Convex Optimization: Precisely

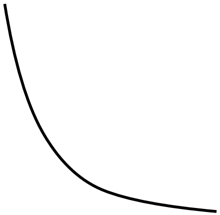
$\inf_x \|\nabla f(x)\| > 0$ case:



e.g., $e^{-x} - x$

$\log\left(e^{-x} + e^{-2x}\right)$

$\inf_x \|\nabla f(x)\| = 0$ case:



e.g., $-\log x$

$-\sqrt{x}$

We treat $\inf_x \|\nabla f(x)\| > 0$ case

- In non-commutative optimization, $\inf_x \|\nabla f(x)\| > 0 \iff \inf_x f(x) = -\infty$
- In general, $\inf_x \|\nabla f(x)\| > 0 \implies \inf_x f(x) = -\infty$

Main Theorem

Gradient Descent (GD) for Unbounded Convex Function
$\to$ Converges to a Specific Point at Infinity

Main Theorem

M: Hadamard manifold

$f : M \to \mathbb{R}$, convex, L-smooth, $\inf_{x \in M} \|\nabla f(x)\| > 0$

in cone topology

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

$$x_{i+1} \coloneqq \exp_{x_i}(-\nabla f(x_i)/L)$$

- New even in $M = \mathbb{R}^d$ case
- Application: Geometric Invariant Theory
  - $x_i \to \xi^*$: destabilizing 1-PSG
  - Left-right action: GD limit of operator scaling

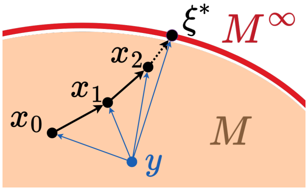


Table of Contents

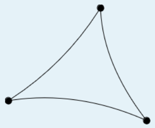
Hadamard Manifold

Hadamard manifold M

Riemannian manifold s.t.

- complete
- simply-connected
- sectional curvature ≤ 0

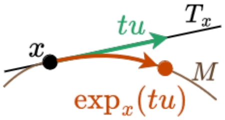
$\left.\right\}$ CAT(0)



Examples: $\mathbb{R}^n$ (Euclidean sp.), $\mathbb{H}^n$ (hyperbolic sp.), P_n (next slide)

Exponential map

$\exp_x(tu) :=$ "point at distance t along geodesic toward u"
$(t \in \mathbb{R}_+, u \in S_x := \{v \in T_x \mid \|v\| = 1\})$



⋆ Unique geodesic between $\forall x, y \in M$

⋆ $\forall y \neq x$ is uniquely represented as $y = \exp_x(tu)$ $(t \in \mathbb{R}_{>0}, u \in S_x)$

Set of Positive Definite Hermitian Matrices

P_n: The set of $n \times n$ positive definite Hermitian matrices

Usual distance (i.e. Frobenius norm, PD cone) → not Hadamard (not complete)

P_n is Hadamard with the following metric:

$$d(X, X + \delta H) \coloneqq \left\| X^{-1/2} \delta H X^{-1/2} \right\|_{\mathrm{F}}$$

- Any unit-speed geodesic: $\gamma(t) = g e^{tH} g^{\dagger}$ ($g \in GL_n$, H: Hermitian, $\|H\|_{\mathrm{F}} = 1$)
- $d(X, Y) = \|\log(X^{-1/2} Y X^{-1/2})\|_{\mathrm{F}}$
- Natural metric in the viewpoint of symmetric space (e.g., [Bridson and Haefliger, 1999])

Geodesically Convex Optimization

Geodesic Convexity

M : Hadamard manifold

$f : M \to \mathbb{R}$ is **convex** / L-**smooth**

$\iff$ $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ is convex / L-smooth (γ: any geodesic)

$\iff$ $\nabla^2 f(x) \geq O$ / $\|\nabla^2 f\|_{\mathrm{op}} \leq L$ $(\forall x \in M)$

Gradient descent

$$x_{i+1} \coloneqq \exp_{x_i}(-h \nabla f(x_i)) \qquad (h > 0 : \text{step-size})$$

- L-smooth $\implies$ $h \coloneqq 1/L$ (or less)

cf. $\exists$ optimum $\implies$ $x_i \to$ optimum (e.g. [Zhang and Sra, COLT'16])

Boundary—Points at Infinity (See e.g. [Sakai, 1996])

Fix $x_0 \in M$

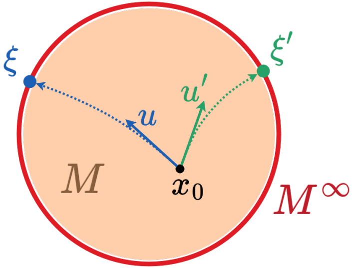
Boundary

Cone topology: visual compactification $M \cup M^\infty$

- $\forall x \in M \ : \ x = \exp_{x_0}(tu) \qquad (\exists! u \in S_{x_0}, t \in \mathbb{R}_+)$
- $\forall \xi \in M^\infty\!: \ \xi = \exp_{x_0}(\infty u) \qquad (\exists! u \in S_{x_0})$

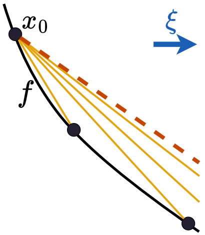
  $\rightarrow M^\infty \simeq S_{x_0} (:= \{v \in T_{x_0} \mid \|v\| = 1\})$

⋆ Independent of x_0



Recession Function: A Tool for Unboundedness

$f : M \to \mathbb{R}$: convex



Recession function [Kapovich, Leeb, and Millson, 2009; Hirai, 2024]

$$f^\infty(\xi) \coloneqq \lim_{t\to\infty} \frac{f(\exp_{x_0}(tu_\xi)) - f(x_0)}{t} \qquad \left(\begin{array}{ccc} S_{x_0} & \simeq & M^\infty \\ \Psi & & \Psi \\ u_\xi & \leftrightarrow & \xi \end{array}\right)$$

Independent of $x_0 \implies f^\infty : M^\infty \to \mathbb{R} \cup \{\infty\}$

Well-known in Euclidean convex analysis (e.g. [Rockafellar, 1970])

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

Prop:
- $\inf_{\xi\in M^\infty} f^\infty(\xi) < 0 \iff \inf_{x\in M} \|\nabla f(x)\| > 0$
- $\inf_{\xi\in M^\infty} f^\infty(\xi) < 0 \implies \exists!\ \xi^* \coloneqq \operatorname{argmin}_{\xi\in M^\infty} f^\infty$ [Kapovich, Leeb, and Millson, 2009]

ξ^* is "The direction in which f decreases most rapidly"

Weak Duality

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

Lemma (Weak duality)

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

Proof: $\forall x \in M, \xi \in M^\infty$,

$$\begin{aligned} f^\infty(\xi) &= \lim_{t \to \infty} \frac{f(\exp_x(tu_\xi)) - f(x)}{t} \\ &\geq \lim_{t \to 0} \frac{f(\exp_x(tu_\xi)) - f(x)}{t} = \langle \nabla f(x), u_\xi \rangle \geq -\|\nabla f(x)\| \end{aligned}$$

$$\begin{pmatrix} S_x & \simeq & M^\infty \\ \Uparrow & & \Uparrow \\ u_\xi & \leftrightarrow & \xi \end{pmatrix}$$

- ⋆ **Strong duality** (equality)? → Achieved at the limit of GD!

Main Theorem

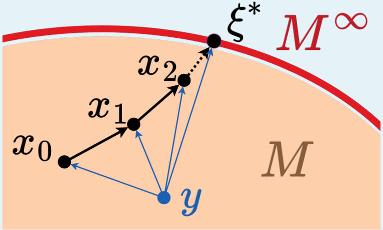
M: Hadamard manifold, $f : M \to \mathbb{R}$: convex, L-smooth
(x_i): gradient descent, step-size $\le 1/L$

Theorem [Hirai and Sakabe, FOCS'24]

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

- $\inf_{x \in M} \|\nabla f(x)\| = \sup_{\xi \in M^\infty} -f^\infty(\xi)$ (strong duality)
- $x_i \to \xi^*$ in cone top. (ξ^*: minimizer of f^∞)
  equivalent to: convergence of direction
- $\|\nabla f(x_i)\| \to \inf_{x \in M} \|\nabla f(x)\|$

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



- Proof → appendix
- Similar properties hold for gradient flow (continuous limit)
  → Generalization to *symmetric spaces with Finsler metric* [Hirai, 2025]

Main Theorem: Euclidean Case

Our result is new even in $M = \mathbb{R}^d$ case:

Theorem (Euclidean case)

$f : \mathbb{R}^d \to \mathbb{R}$: convex, L-smooth, $\inf_{x \in \mathbb{R}^d} \|\nabla f(x)\| > 0$

$x_{i+1} \coloneqq x_i - \nabla f(x_i)/L$

Gradient vector converges to the minimum norm point in $\overline{\nabla f(\mathbb{R}^d)}$:

$$\lim_{i \to \infty} \nabla f(x_i) = \operatorname*{argmin}_{p \in \overline{\nabla f(\mathbb{R}^d)}} \|p\|$$

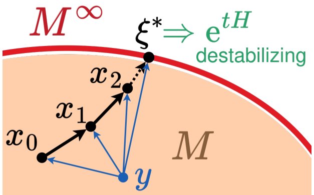
Remark:

Strong duality is obtained by $f^\infty(u) = \sup_{p \in \nabla f(\mathbb{R}^d)} \langle u, p \rangle$ (e.g. [Rockafellar, 1970])

Table of Contents

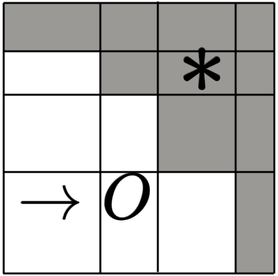
Application Overview

- GD limit $\Rightarrow$ **Destabilizing 1-PSG** in Geometric Invariant Theory



- GD Limit of Unscalable Operator Scaling: **Simultaneous Block-Triangular**

$\sigma_i A_k \tau_i^\dagger =$
(for all k)



Introduction: Matrix Scaling

Matrix scaling problem [Sinkhorn, 1964]

Input: $A \in \mathbb{R}^{n \times n}_{\geq 0}$
Find: Positive diagonal R, C s.t. RAC is doubly stochastic:

$$\|(RAC)\mathbf{1} - \mathbf{1}\| < \varepsilon, \quad \|(RAC)^\top \mathbf{1} - \mathbf{1}\| < \varepsilon$$

Applications:
Markov chain, computational linear algebra, optimal transport + entropy regularization, ...

Algorithm: Sinkhorn iteration

Repeat:

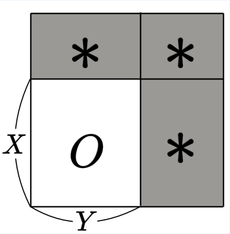
1. Update R s.t. $(RAC)\mathbf{1} = \mathbf{1}$ (row scaling)
2. Update C s.t. $(RAC)^\top \mathbf{1} = \mathbf{1}$ (column scaling)

Matrix Scalability

Thm. Scalability [Knopp and Sinkhorn, 1967]

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

$\iff$ A has no **Hall blocker:**

permutation P, Q s.t. $PAQ =$ 

$|X| + |Y| \geq n + 1$

Unscalable case:

Sinkhorn iteration can find a Hall blocker [Hayashi, Hirai, and Sakabe, 2022+]

Opeartor Scaling

Operator scaling problem [Gurvits, 2004]

Input: $\mathcal{A} \coloneqq (A_1, \ldots, A_N) \in \mathbb{C}^{N(n\times n)}$
Find: $g, h \in GL_n$ s.t. $g\mathcal{A}h^\dagger$ is doubly stochastic:

$$\left\| \sum_{k=1}^{N} gA_k h^\dagger h A_k^\dagger g^\dagger - I \right\|_{\mathrm{F}} < \varepsilon, \quad \left\| \sum_{k=1}^{N} hA_k^\dagger g^\dagger g A_k h^\dagger - I \right\|_{\mathrm{F}} < \varepsilon \qquad (I : \text{identity})$$

Related to:

- noncommutative Edmonds' problem [Garg, Gurvits, Oliveira, and Wigderson, FOCS'16; Ivanyos, Qiao, and Subrahmanyam, 2017]
- Brascamp–Lieb inequality [Garg, Gurvits, Oliveira, and Wigderson, STOC'17]

Scalability problem: Special case of **null cone membership problem** in GIT

Null Cone Membership Problem

$G \subseteq GL_n(\mathbb{C})$: reductive algebraic group
$\pi : G \to GL(V)$: rational representation (V: finite dim.)
$v \in V$: given

Null cone membership problem

Determine whether $0 \in \overline{\pi(G)v}$

- $0 \notin \overline{\pi(G)v} \iff$ **stable**
- $0 \in \overline{\pi(G)v} \iff p(v) = 0$ ($\forall p$: homogeneous invariant polynomial)

(Hilbert–Mumford, e.g. [Derksen and Kemper, 2015])

Appears in computational algebra, quantum information, statistics, functional analysis, etc... [Bürgisser et al., FOCS'19]

Unsolved: polynomial time?

Previous Researches: Threshold Approach

Noncommutative Optimization [Bürgisser et al., FOCS'19]

minimize $\quad F_v(x) \coloneqq \log\langle v, \pi(x)v\rangle = 2\log\|\pi(g)v\| \qquad (x = g^\dagger g \in P_n \cap G)$

- F_v: Convex on Hadamard manifold $P_n \cap G$ (**Kempf–Ness function**)
- $\inf_x F_v(x) > -\infty \iff 0 \notin \overline{\pi(G)v}$

In some actions, $\inf_x F_v(x) > -\infty \implies \inf_x F_v(x) \geq C$ (threshold)

Optimization algorithms:

- Gradient descent [Bürgisser et al., FOCS'19]
- Box constrained Newton's method [Bürgisser et al., FOCS'19]
- Interior-point method [Hirai, Nieuwboer, and Walter, FOCS'23]

$K \subseteq G$: maximum compact subgroup of G
$K\backslash G \simeq P_n \cap G$: quotient set
P_n: $n \times n$ positive definite Hermitian matrices
$\langle \bullet, \bullet\rangle, \|\bullet\|$: invariant under K

Our Target: Unstable Case

Thm. Unstability (Kempf–Ness, Hilbert–Mumford) (e.g. [Georgoulas, Robbin, and Salamon, 2021])

The followings are equivalent:

(a) $\inf_x F_v(x) = -\infty \quad (\iff 0 \in \overline{\pi(G)v})$

(b) $\inf_x \|\nabla F_v(x)\| > 0$

(c) $\exists$ 1-PSG $t \mapsto \mathrm{e}^{tH}$ s.t. $\lim_{t\to\infty} \pi\left(\mathrm{e}^{tH}\right) v = 0$ (**destabilizing** 1-PSG)

Q: Given an unstable input, can we find a destabilizing 1-PSG (in polynomial time)?

Known: Left-right action case (operator scaling case) → Yes (later)

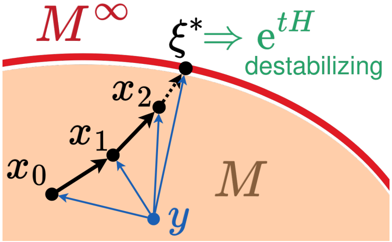
Implication of Main Theorem

Q: Given an unstable input, can we find a destabilizing 1-PSG (in polynomial time)?

Fact: e^{tH} with $\xi^\star = \lim_{t\to\infty} \mathrm{e}^{tH}$: destabilizing 1-PSG

Our result: GD limit $\Rightarrow$ destabilizing 1-PSG

- ☺ Conceptually new approach
- ☹ Finite iterations of GD $\neq$ destabilizing 1-PSG
  $\Longrightarrow$ We need GD + **rounding**



Unknwon: rounding in general?
complexity?

Known: rounding in left-right action case [Franks, Soma, and Goemans, SODA'23]

Left-Right Action

Left-right action [Garg, Gurvits, Oliveira, and Wigderson, FOCS'16]

$G = SL_n \times SL_n$
$V = \mathbb{C}^{N(n\times n)}$
$\mathcal{A} := (A_1, \dots, A_N) \in V$
$\pi(g,h)\mathcal{A} = g\mathcal{A}h^\dagger := (gA_1h^\dagger, \dots, gA_Nh^\dagger)$: **left-right action**

(un)**scalable** in operator scaling $\iff$ (un)**stable** in left-right action

c.f. Left-right torus action

$\mathbb{T}_n := \{g \in SL_n \mid g : \text{diagonal}\}$
$G = \mathbb{T}_n \times \mathbb{T}_n$
$V = \mathbb{C}^{n\times n}$
$\pi(g,h)A := gAh^\dagger$

(un)scalable in matrix scaling $\iff$ (un)stable in left-right torus action

Operator Scaling as Convex Optimization

Kempf–Ness function: $F_{\mathcal{A}}(x,y) \coloneqq \log \sum_{k=1}^{N} \operatorname{tr} x A_k y A_k^{\dagger} \qquad (x, y \in P_n \cap SL_n)$

Stationary condition of $F_{\mathcal{A}} \iff g\mathcal{A}h^{\dagger}$ is doubly-stochastic ($g \coloneqq \sqrt{x}, h \coloneqq \sqrt{y}$):

$$\nabla F_{\mathcal{A}}(x,y) = \begin{pmatrix} \sum_{k=1}^{N} g A_k h^{\dagger} h A_k^{\dagger} g^{\dagger} - I \\ \sum_{k=1}^{N} h A_k^{\dagger} g^{\dagger} g A_k h^{\dagger} - I \end{pmatrix} \qquad \text{(under appropriate normalization)}$$

Algorithms (stable case):

- operator Sinkhorn [Gurvits, 2004]
- gradient descent [Kwok, Lau, and Ramachandran, FOCS'19]

Open problem: Operator Sinkhorn limit in unscalable case [Garg and Oliveira, 2018]

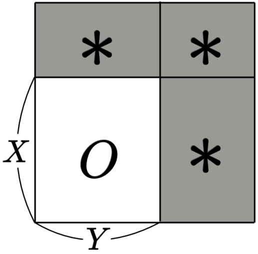
Our result: GD limit ⇒ simultaneous block-triangular form

Unstability Certificate: Vanishing Subspace

Destabilizing 1-PSG $\approx$ **Vanishing subspace**:

$UA_kV =$

(for all k)



U, V: unitary

$\dim X + \dim Y \geq n + 1$

Algorithms to find vanishing subspaces:

- Wong sequence [Ivanyos, Qiao, and Subrahmanyam, 2018]
- modified Sinkhorn iteration [Franks, Soma, and Goemans, SODA'23]
- submodular optimization [Hamada and Hirai, 2021]

Gradient Descent Limit for Operator Scaling

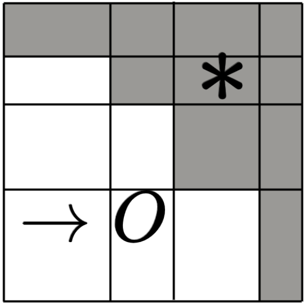
Theorem [Hirai and Sakabe, FOCS'24]

$\{(x_i, y_i)\}_i$: GD, step-size $\leq 1/2$

$$\begin{cases} x_i = \sigma_i^\dagger \operatorname{diag} p^i \sigma_i \ (p^i : \text{nondecreasing}) \\ y_i = \tau_i^\dagger \operatorname{diag} q^i \tau_i \ (q^i : \text{nonincreasing}) \end{cases}$$

$\sigma_i \mathcal{A} \tau_i^\dagger \to$ **simultaneous block-triangular**

$\sigma_i A_k \tau_i^\dagger =$
(for all k)



- maximal zero-block $=$ maximal vanishing subspace
- vector-space generalization of **Dulmage–Mendelsohn decomposition** [Dulmage and Mendelsohn, 1958]
- divided subspaces $=$ special case of **Harder–Narasimhan filtration** of quiver representation [Iwamasa, Oki, and Soma, 2024]

Table of Contents

Summary

- Gradient descent for unbounded convex function on Hadamard manifold
  - **Strong duality** is achieved at the limit of GD
  - Convergence to the minimizer of recession function (at infinity)
- Application: Geometric invariant theory
  - Gradient descent limit $\Rightarrow$ destabilizing 1-PSG
  - Operator scaling (left-right action): simultaneous block-triangular form
- Future works
  - Convergence rates
  - Rounding in general

Message: Unbounded convex optimization is also a meaningful problem!

References I

Bridson, M. R. and A. Haefliger (1999). *Metric spaces of non-positive curvature*. Vol. 319. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, pp. xxii+643. DOI: 10.1007/978-3-662-12494-9.

Bürgisser, P. et al. (2019). “Towards a theory of non-commutative optimization: geodesic 1st and 2nd order methods for moment maps and polytopes”. *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*. IEEE Comput. Soc. Press, Los Alamitos, CA, pp. 845–861. DOI: 10.1109/FOCS.2019.00055.

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.

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.

Garg, A., L. Gurvits, R. Oliveira, and A. Wigderson (2018). “Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via Operator Scaling”. *Geom. Funct. Anal.*, 28.1, pp. 100–145.

Garg, A., L. Gurvits, R. Oliveira, and A. Wigderson (2020). “Operator scaling: theory and applications”. *Found. Comput. Math.*, 20.2, pp. 223–290. DOI: 10.1007/s10208-019-09417-z.

Garg, A. and R. Oliveira (2018). “Recent progress on scaling algorithms and applications”. *CoRR*, abs/1808.09669. arXiv: 1808.09669.

Georgoulas, V., J. W. Robbin, and D. A. Salamon (2021). *The moment-weight inequality and the Hilbert-Mumford criterion—GIT from the differential geometric viewpoint*. Vol. 2297. Lecture Notes in Mathematics. Springer, Cham, pp. vii+190. DOI: 10.1007/978-3-030-89300-2.

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.

References II

Hamada, M. and H. Hirai (2021). "Computing the nc-rank via discrete convex optimization on $\mathrm{CAT}(0)$ spaces". *SIAM J. Appl. Algebra Geom.*, 5.3, pp. 455–478. DOI: 10.1137/20M138836X.

Hayashi, K., H. Hirai, and K. Sakabe (to appear). "Finding Hall blockers by matrix scaling". *Mathematics of Operations Research*. to appear.

Hirai, H. (2024). "Convex analysis on Hadamard spaces and scaling problems". *Foundations of Computational Mathematics.*, 24.6, pp. 1979–2016.

Hirai, H. (2025). *A scaling characterization of nc-rank via unbounded gradient flow*. arXiv:2502.05412, 2025.

Hirai, H., H. Nieuwboer, and M. Walter (2023). "Interior-point methods on manifolds: theory and applications". *Proceedings of the 64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023*. IEEE, pp. 2021–2030. DOI: 10.1109/FOCS57990.2023.00123.

Hirai, H. and K. Sakabe (2024). "Gradient Descent for Unbounded Convex Functions on Hadamard Manifolds and Its Applications to Scaling Problems". *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024*. IEEE, pp. 2387–2402.

Ivanyos, G., Y. Qiao, and K. V. Subrahmanyam (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 (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.

Iwamasa, Y., T. Oki, and T. Soma (2024). *Algorithmic aspects of semistability of quiver representations*. arXiv:2407.06493, 2024.

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.

References III

Knopp, P. and R. Sinkhorn (1967). “Concerning nonnegative matrices and doubly stochastic matrices.”. *Pacific Journal of Mathematics*, 21.2, pp. 343–348.

Kwok, T. C., L. C. Lau, and A. Ramachandran (2021). “Spectral Analysis of Matrix Scaling and Operator Scaling”. *SIAM Journal on Computing*, 50.3, pp. 1034–1102. DOI: 10.1137/20M1315981.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

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.

Sinkhorn, R. (1964). “A relationship between arbitrary positive matrices and doubly stochastic matrices”. *Ann. Math. Statist.*, 35, pp. 876–879. DOI: 10.1214/aoms/1177703591.

Zhang, H. and S. Sra (2016). “First-order Methods for Geodesically Convex Optimization”. *Proceedings of the 29th Annual Conference on Learning Theory, PMLR*. Vol. 49, pp. 1617–1638.