Joint Seminar of AFSA and Suri CS

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

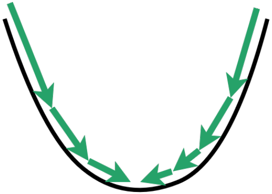
to appear in FOCS 2024, arXiv: 2404.09746

Keiya Sakabe (The University of Tokyo)
Joint work with: Hiroshi Hirai (Nagoya University)

Oct. 10, 2024

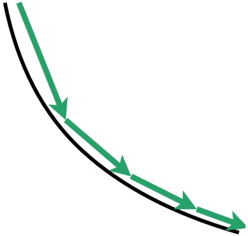
Unbounded Convex Optimization

Usual Convex Optimization



Convergence
Meaningful!

Unbounded Convex Optimization



Divergence
Meaningless?

Research Overview

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

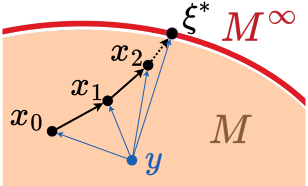


Table of Contents

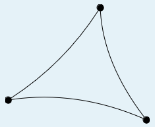
Hadamard Manifold

Hadamard manifold M

Riemannian manifold s.t.

- complete
- simply-connected
- sectional curvature ≤ 0

$\Big\}$ $\mathrm{CAT}(0)$

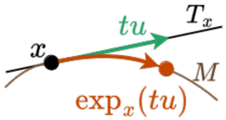


Examples: $\mathbb{R}^n$ (Euclidean sp.), $\mathbb{H}^n$ (hyperbolic sp.), P_n ($n \times n$ p.d. Hermitian matrices)

- ⋆ Unique geodesic between $\forall x, y \in M$
- ⋆ Any geodesic can be extended infinitely $\to$ Exponential map

Exponential map

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



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)

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

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)
- $\exists$ optimum $\implies$ $x_i \to$ optimum (e.g. [Zhang and Sra, 2016])

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

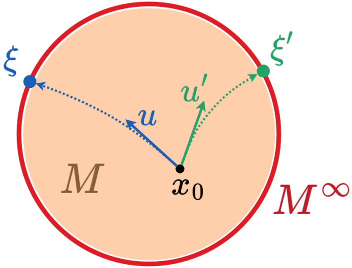
Fix $x_0 \in M$

Cone topology: compactification $\underbrace{M \cup M^\infty}$ Boundary

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

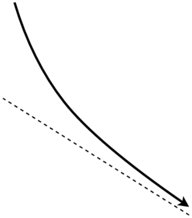
  $\rightarrow M^\infty \simeq S_{x_0} (:= \{v \in T_{x_0} \mid \|v\| = 1\})$

⋆ Independent of x_0



Recession Function—A Tool for Unboundedness

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



Recession function [Kapovich, Leeb, and Millson, 2009; Hirai, 2022+]

$$f^\infty(\xi) := \lim_{t\to\infty} \frac{f(\exp_{x_0}(tu))}{t} \qquad \begin{pmatrix} S_{x_0} & \simeq & M^\infty \\ \uplus & & \uplus \\ u & \leftrightarrow & \xi \end{pmatrix}$$

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

Generalization of Euclidean case (e.g. [Rockafellar, 1970])

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

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

Weak Duality

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

Lemma (Weak duality, this work)

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

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

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

$$\begin{pmatrix} S_x & \simeq & M^\infty \\ \Psi & & \Psi \\ u & \leftrightarrow & \xi \end{pmatrix}$$

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

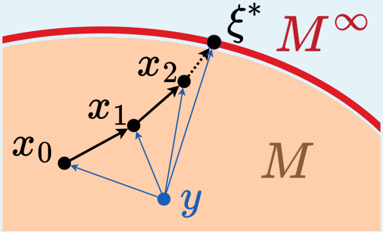
Main Theorem

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

Theorem (this work)

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^∞)
- $\|\nabla f(x_i)\| \to \inf_{x \in M} \|\nabla f(x)\|$



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

☹ No theoretical convergence rates

Main Theorem: Euclidean Case

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])

Descent Lemma

step-size $h \coloneqq 1/L$

Lemma

- $\|\nabla f(x_{i+1})\| \leq \|\nabla f(x_i)\|$ (well-known)
- $f(x_{i+1}) - f(x_i) \leq -\dfrac{\|\nabla f(x_{i+1})\|^2}{L}$ **(descent lemma, this work)**

Remarks:

- Well-known $f(x_{i+1}) - f(x_i) \leq -\dfrac{\|\nabla f(x_i)\|^2}{2L}$: useless in unbounded case
- $M = \mathbb{R}^d$ case: $f(x_{i+1}) - f(x_i) \leq -\dfrac{\|\nabla f(x_i)\|^2 + \|\nabla f(x_{i+1})\|^2}{2L}$

Proof of Lemma

$\tau_{x \to y}$: geodesic parallel transport

(gradient) − (gradient) = $\int$ (Hess)·(vector):

$$\begin{aligned}
\tau_{x_{i+1} \to x_i} \nabla f(x_{i+1}) &= \nabla f(x_i) + \int_0^1 \tau_{x_i^{(\lambda)} \to x_i} \nabla^2 f\left(x_i^{(\lambda)}\right) \tau_{x_i \to x_i^{(\lambda)}} \left(-\nabla f(x_i)/L\right) \mathrm{d}\lambda \\
&= \left(I - \int_0^1 \frac{1}{L} \tau_{x_i^{(\lambda)} \to x_i} \nabla^2 f\left(x_i^{(\lambda)}\right) \tau_{x_i \to x_i^{(\lambda)}} \mathrm{d}\lambda\right) \nabla f(x_i) \\
&=: B_i \nabla f(x_i) \\
& \boxed{B_i \succeq O, \|B_i\|_{\mathrm{op}} \leq 1}
\end{aligned}$$

($x_i^{(\lambda)}$: $\lambda : (1-\lambda)$ dividing point of $\overline{x_i x_{i+1}}$ ($x_i^{(0)} = x_i, x_{i+1}^{(1)} = x_{i+1}$))

Proof of Lemma

$$\boxed{\tau_{x_{i+1}\to x_i}\nabla f(x_{i+1}) = B_i \nabla f(x_i) \qquad (B_i \succeq O, \|B_i\|_{\mathrm{op}} \le 1)}$$

- $\|\nabla f(x_{i+1})\| = \|\tau_{x_{i+1}\to x_i}\nabla f(x_{i+1})\| \le \|B_i\|_{\mathrm{op}}\|\nabla f(x_i)\| \le \|\nabla f(x_i)\|$
- $$\begin{aligned} f(x_{i+1}) - f(x_i) &= \int_0^1 \left\langle -\nabla f(x_i)/L, \tau_{x_i^{(\lambda)}\to x_i}\nabla f\left(x_i^{(\lambda)}\right)\right\rangle \mathrm{d}\lambda && \text{(integral along geodesic)} \\ &\le \langle -\nabla f(x_i)/L, \tau_{x_{i+1}\to x_i}\nabla f(x_{i+1})\rangle && \text{(convexity)} \\ &= -\langle \nabla f(x_i), B_i \nabla f(x_i)\rangle /L && \text{(above equation)} \\ &\le -\langle \nabla f(x_i), B_i B_i \nabla f(x_i)\rangle /L && (B_i \succeq O, \|B_i\|_{\mathrm{op}} \le 1) \\ &= -\|\nabla \tau_{x_{i+1}\to x_i} f(x_{i+1})\|^2/L && \text{(above equation)} \\ &= -\|\nabla f(x_{i+1})\|^2/L && (\tau_{x\to y}\text{preserves norm}) \end{aligned}$$

Lemma (review)

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

Proof of Main Theorem (Sketch)

$$
\begin{aligned}
\lim_{i\to\infty} -(f(x_i) - f(x_0))/i &= \lim_{i\to\infty} f(x_i) - f(x_{i+1}) && \text{(Cesàro mean)} \\
&\geq \lim_{i\to\infty} \|\nabla f(x_i)\|^2/L && \text{(descent lemma)} \\
&\geq \left(\lim_{i\to\infty} \|\nabla f(x_i)\|/L\right) \inf_{x\in M} \|\nabla f(x)\| && (\|\nabla f(x_{i+1})\| \leq \|\nabla f(x_i)\|) \\
&= \left(\lim_{i\to\infty} \frac{1}{i}\sum_{j=0}^{i-1} d(x_j, x_{j+1})\right) \inf_{x\in M} \|\nabla f(x)\| && \text{(GD, Cesàro mean)} \\
&\geq \left(\lim_{i\to\infty} d(x_i, x_0)/i\right) \sup_{\xi\in M^\infty} -f^\infty(\xi) && \text{(tri. ineq. / weak duality)}
\end{aligned}
$$

On the other hand, $\lim_{i\to\infty} -\frac{f(x_i) - f(x_0)}{d(x_i, x_0)} \leq \sup_{\xi\in M^\infty} -f^\infty(\xi)$ (def. of f^∞)

Therefore, all inequalities are actually =. Especially:

- $\lim_{i\to\infty} \frac{f(x_i) - f(x_0)}{d(x_i, x_0)} = \inf_{\xi\in M^\infty} f^\infty(\xi) \implies x_i \to \xi^*$
- $\lim_{i\to\infty} \|\nabla f(x_i)\| = \inf_{x\in M} \|\nabla f(x)\| = \sup_{\xi\in M^\infty} -f^\infty(\xi)$

Table of Contents

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:

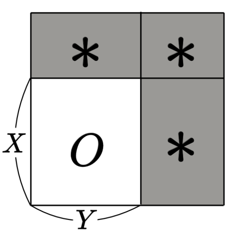
1. Update R s.t. $(RAC)\mathbf{1} = \mathbf{1}$ (row scaling)
2. Update C s.t. $(RAC)^\top\mathbf{1} = \mathbf{1}$ (column scaling)

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: $\mathcal{A} := (A_1, \dots, A_N) \in \mathbb{C}^{N(n\times n)}$
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 $0 \in \overline{\pi(G)v}$
- **stable** $:\iff \; 0 \notin \overline{\pi(G)v}$

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

Previous Researches: Threshold Approach

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

$$\text{minimize} \qquad F_v(x) \coloneqq \log\langle v, \pi(x)v\rangle = 2\log\|\pi(g)v\| \qquad (x = g^\dagger g)$$

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

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: Operator scaling case → Yes (later)

c.f. Strong duality $\implies$ Moment-weight inequality in GIT

(e.g. [Georgoulas, Robbin, and Salamon, 2021])

Implication of Main Theorem

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

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 opeartor scaling case [Franks, Soma, and Goemans, SODA'23]

Left-Right Action

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

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

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

c.f. Left-right torus action

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

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

Operator Scaling as Convex Optimization

Kempf–Ness function:

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

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]

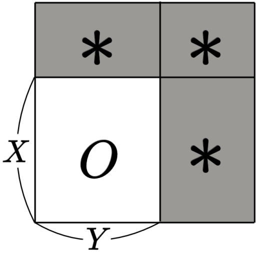
$\boxed{\textbf{Our result: GD limit} \Rightarrow \textbf{simultaneous block-triangular form}}$

Matrix scaling: Sinkhorn Limit ⇒ block-triangular [Hayashi, Hirai, and Sakabe, 2022+]

Unstability Certificate: Vanishing Subpace

Destabilizing 1-PSG ≈ **Vanishing subspace**:

$U A_k V =$
(for all k)



U, V: unitary
$\dim X + \dim Y \geq n + 1$

Algorithms:

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

Generalized DM Decomposition

(generalized) Dulmage–Mendelsohn decomposition [Hamada and Hirai, 2017]

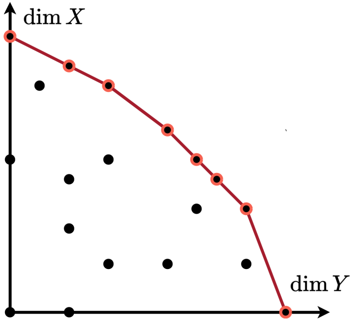
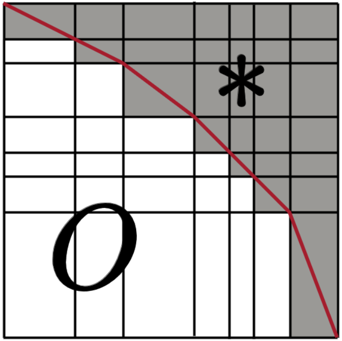


Figure of $(\dim X, \dim Y)$

$U^* A_k V^{*\dagger} =$
(for all k)



Corresponding DM decomposition

- Finest simultaneous block-triangular form
- Contains multiple maximal vanishing subspace

Gradient Descent Limit for Operator Scaling

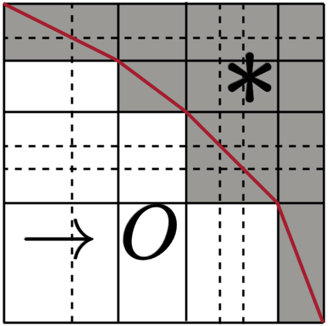
Theorem (this work)

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

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

$\sigma_i A \tau_i^\dagger \to$ **coarse DM decomposition (linear convergence)**

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



- **Coarse DM decomposition:** Unite blocks with same slopes (#col. / #row)
- 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 to GIT
  - Gradient descent limit = destabilizing 1-PSG
  - Operator scaling (left-right action): simultaneous block-triangular form
- Future works
  - Convergence rates
  - Rounding in general

Message: Unbounded convex optimization is meaningful problem!
New possibility to find a destabilizing 1-PSG in general!

Please check our arXiv! arXiv: 2404.09746

References I

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.

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.

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 $\mathrm{CAT}(0)$ 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.

References II

Hirai, H. (in press). "Convex analysis on Hadamard spaces and scaling problems". *Foundation of Computational Mathematics*. in press.

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.

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.

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.

References III

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.