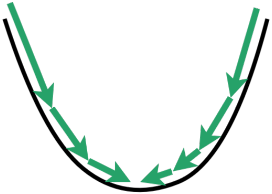
Gradient Descent for Unbounded Convex Functions on Hadamard Manifolds and Its Applications to Scaling Problems

Hiroshi Hirai: Nagoya University
◯ Keiya Sakabe: The University of Tokyo

FOCS 2024 @ Chicago, IL, USA
Oct. 30, 2024

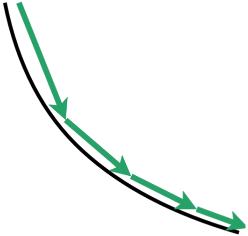
Unbounded Convex Optimization

Usual Convex Optimization



Convergence
Meaningful!

Unbounded Convex Optimization



Divergence
Meaningless?

Research Overview

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

Main Theorem

M: Hadamard manifold

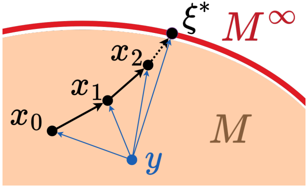
$f : M \to \mathbb{R}$, convex, L-smooth, $\inf_{x \in M} \|\nabla f(x)\| > 0$

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



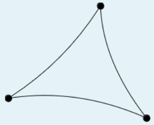
Hadamard Manifold

Hadamard manifold M

Riemannian manifold s.t.

- complete

$$\left.\begin{array}{l}\bullet\ \text{simply-connected}\\ \bullet\ \text{sectional curvature} \le 0\end{array}\right\}\ \mathrm{CAT}(0)$$



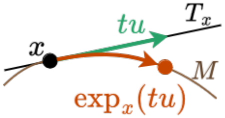
Examples: $\mathbb{R}^n$ (Euclidean sp.), $\mathbb{H}^n$ (hyperbolic sp.), P_n ($n \times n$ p.d. Hermitian matrices)

- ⋆ Unique geodesic between $\forall x, y \in M$
- ⋆ Any geodesic can be extended infinitely → 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**

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

Gradient descent

$$x_{i+1} \coloneqq \exp_{x_i}\left(-h \nabla f(x_i)\right) \qquad (h \leq 1/L : \text{step-size})$$

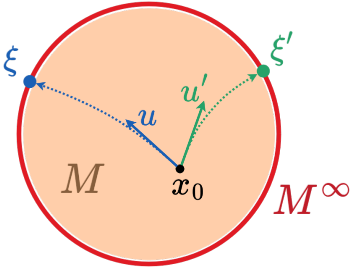
Boundary—Points at Infinity (See e.g. [Sakai, 1996])

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

Fix $x_0 \in M$

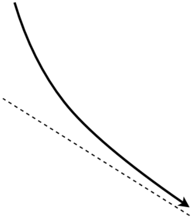
- $\forall x \in M \;:\; x = \exp_{x_0}(tu) \qquad (\exists! u \in S_{x_0}, t \in \mathbb{R}_+)$
- $\forall \xi \in M^\infty \colon\; \xi = \exp_{x_0}(\infty u) \qquad (\exists! u \in S_{x_0})$

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



Recession Function—A Tool for Unboundedness

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



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

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

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

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]

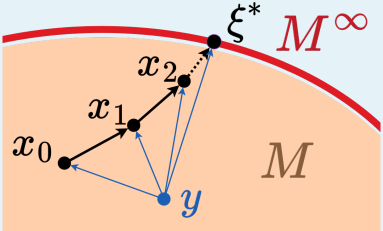
Main Theorem

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

Theorem (this work)

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

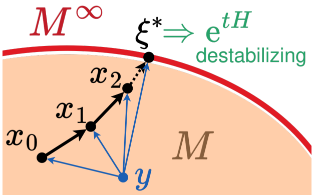
- $\inf_{x \in M} \|\nabla f(x)\| = \sup_{\xi \in M^\infty} -f^\infty(\xi)$ (strong duality)
- $x_i \to \xi^*$ in cone top. (ξ^*: minimizer of f^∞)
- $\|\nabla f(x_i)\| \to \inf_{x \in M} \|\nabla f(x)\|$



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

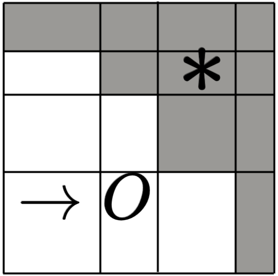
Application Overview

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



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

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



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

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

Unsolved: polynomial time?

Appears in computational algebra, quantum information, statistics, functional analysis, etc... [Bürgisser, Franks, Garg, Oliveira, Walter, and Wigderson, FOCS'19]

Previous Research: Optimization & Threshold

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**)
- $\inf_x F_v(x) > -\infty \iff 0 \notin \overline{\pi(G)v}$

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

Optimization algorithms (analyses in bounded case):

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

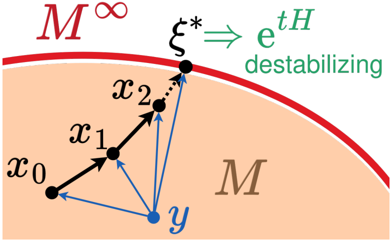
(c) $\exists$ 1-PSG $t \mapsto \mathrm{e}^{tH}$ s.t. $\lim_{t\to\infty} \pi\left(\mathrm{e}^{tH}\right) v = 0$ (**destabilizing** 1-PSG)

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

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

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?

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

Left-Right Action, Operator Scaling

Left-right action, Operator scaling
[Gurvits, 2004; Garg, Gurvits, Oliveira, and Wigderson, FOCS'16]

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

Algorithms (stable case):

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

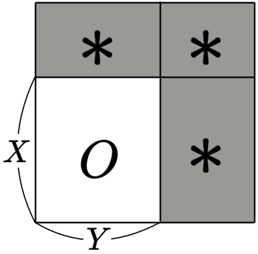
Open problem: Operator Sinkhorn limit in unstable case [Garg and Oliveira, 2018]

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

Unstability Certificate: Vanishing Subpace

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

$UA_kV =$
(for all k)



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

Algorithms:

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

Gradient Descent Limit of Operator Scaling

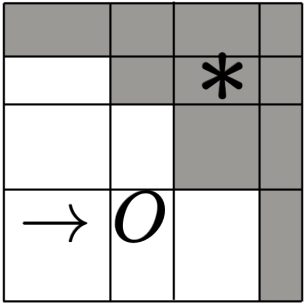
Theorem (this work)

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

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

$\sigma_i \mathcal{A} \tau_i^\dagger \to$ **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]

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 meaningful problem!
New possibility to find a destabilizing 1-PSG in general!

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". *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019*. IEEE, pp. 845–861. DOI: 10.1109/FOCS.2019.00055.

Derksen, H. and G. Kemper (2015). *Computational Invariant Theory*. Springer Berlin, Heidelberg. x+268.

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 (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". *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, 125, pp. 14–49.

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. Cham: Springer. 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 (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.

Hirai, H. (in press). "Convex analysis on Hadamard spaces and scaling problems". *Found. Comput. Math.* in press.

References II

Hirai, H., H. Nieuwboer, and M. Walter (2023). "Interior-point methods on manifolds: theory and applications". *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 (2018). "Constructive non-commutative rank computation is in deterministic polynomial time". *Comput. Complex.*, 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. Differ. Geom.*, 81.2, pp. 297–354.

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

Sakai, T. (1996). *Riemannian Geometry*. Providence, RI: American Mathematical Society. xiv+358. DOI: 10.1090/mmono/149.