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

Hiroshi Hirai (Nagoya University), **Keiya Sakabe** (Ruhr University Bochum)

FOCS 2024, arXiv:2404.09746

Motivations

- What does gradient descent do for unbounded objective functinons?
- What is the gradient descent limit of unscalabe operator scaling?

arXiv link:

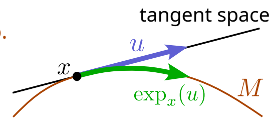
General Theory: Unbounded Case in Convex Optimization

Setting: Geodesically Convex Optimization

M: Hadamard manifold (i.e., complete, simply-connected, curvature ≤ 0)

$f : M \to \mathbb{R}$: geodesically convex (i.e., convex along any geodesic) & L-smooth

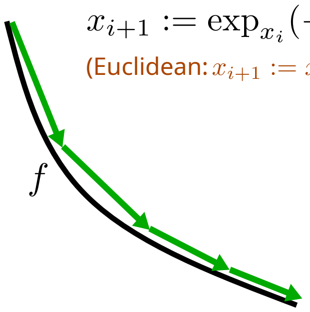
Exponential map is diffeo. (Cartan–Hadamard)



Gradient Descent:

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

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



Question: What can we obtain from (x_i) when f is lower-unbounded?

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

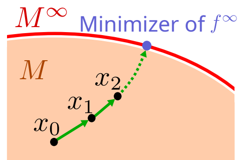
Main Theorem

Strong duality (new purely math statement)

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

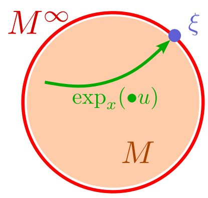
Gradient norm is minimized.

Divergence behavior characterization: (x_i) converges to the minimizer of f^∞



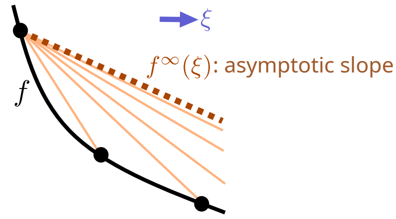
Technical Tools:

Cone Topology (compactification)



point on M^∞ = limit of geodesic

Recession Function [KLM'09, Hir'24]



$$f_{\mathcal{A}}(x, y) := \log \sum_k \operatorname{Tr} x A_k y A_k^\dagger$$ (x, y: positive definite, det = 1)

Special case of *non-commutative optimization* [BFGOWW'19]

Application: Operator Scaling

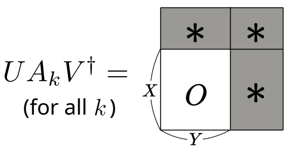
$\mathcal{A} = (A_1, \ldots, A_N)$: Kraus operators of a CP map $\rho \mapsto \sum_k A_k \rho A_k^\dagger$

(ε-)doubly-stochastic $\overset{\text{def.}}{\iff}$ $\sum_k A_k A_k^\dagger \approx_\varepsilon I,\ \sum_k A_k^\dagger A_k \approx_\varepsilon I.$

Operator Scaling [Gur'04, AGLOW'17]

Given $\mathcal{A} = (A_1, \ldots, A_N) \in \mathbb{C}^{N(n\times n)}$, can one make $g\mathcal{A}h^\dagger := (gA_1h^\dagger, \ldots)$ doubly-stochastic with $g, h \in \mathrm{GL}_n$?

Unscalable $\iff$ $\inf \|\nabla f_{\mathcal{A}}\| > 0$ $\iff$ a **vanishing subspace** exists:

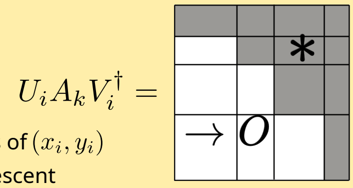


$\dim X + \dim Y \geq n + 1$

U, V : unitary

Our Result

Gradient descent for unbouded $f_{\mathcal{A}}$ gives a simultaneous block-triangular form:



U_i, V_i : eigenbases of (x_i, y_i)

(x_i, y_i): gradient descent

Maximal vanishing subspaces emerge!

Future Work

- Obtain convergence rates
- Other algorithms (second-order, Sinkhorn, ...)
- More applications?

References

[AGLOW'17]: Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira, and A. Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. STOC2018, pp. 172–181, 2018.
[BFGOWW'19]: P. Bürgisser, C. Franks, A. Garg, R. Oliveira, M. Walter, and A. Wigderson. Towards a theory of non-commutative optimization: geodesic 1st and 2nd order methods for moment maps and polytopes. FOCS 2019, pp. 845–861, 2019.
[Gur'04]: L. Gurvits. Classical complexity and quantum entanglement. J. Comput. System Sci., 69(3):448–484, 2004.
[Hir'24]: H. Hirai. Convex analysis on Hadamard spaces and scaling problems. Found. Comput. Math., 24(6):1979–2016, 2024.
[KLM'09]: M. Kapovich, B. Leeb, and J. Millson. Convex functions on symmetric spaces, side lengths of polygons and the stability inequalities for weighted configurations at infinity. J. Differential Geom., 81(2):297–354, 2009.

名古屋大学 NAGOYA UNIVERSITY

RUHR UNIVERSITÄT BOCHUM RUB

科研費 KAKENHI

erc