修士論文 要旨

Optimization over Orbit Closures of Real Reductive Lie Group Actions

数理情報学専攻 48206241 Zhan Zhiyuan
指導教員 Professor Hiroshi Hirai

1 Introduction

Sinkhorn [5] introduced the matrix scaling problem, that is, given a nonnegative matrix A to determine whether there are positive diagonal matrices X, Y such that XAY is doubly stochastic or approximately doubly stochastic. This problem has many applications, such as the existence of perfect matching in a bipartite graph, polynomial identity testing, and so on. Gurvits [3] generalized this problem to the operator scaling problem, which is applied to the non-commutative polynomial identity testing, the Brascamp–Lieb inequalities, and so on.

These two problems are special cases of a classic problem arising from the invariant theory. Let (π, V) be an action of an algebraic group G, that is, $\pi(g) \in GL(V)$ for any $g \in G$, where V is a vector space. The problem is to determine whether 0 is in the orbit closure $\overline{\pi(G)v}$ of $v \in V$, called the null cone membership. Besides above scaling problems, the tensor scaling problem, Horn's problem, and many other problems, are concrete cases of the null cone membership. This problem has another view. For any such action and nonzero $v \in V$, there is the Kempf-Ness function $f_v \colon G \to \mathbb{R}$. 0 is not in the orbit closure v if and only if f_v is bounded below, and if and only if 0 is in $\overline{\nabla f_v(G)}$, the closure of the image of the gradient. Therefore, it is related to some optimization problem.

2 Previous Research and This Work

For a complex reductive algebraic group G acting on a complex vector space V, Bürgisser et al. [2] transformed this problem into optimizing the Kempf-Ness function $f_v(g) := \log \|\pi(g)v\|^2$. Moreover, let $K = G \cap U(n)$ be a maximal compact subgroup of G, then the norm $\|\cdot\|$ on V is K-invariant. Therefore, f_v is defined on G/K, which has a standard Riemannian structure. The null cone membership is related to an optimization problem on a Riemannian manifold. They also showed f_v is geodesically convex and L-smooth for some $L > 0$. If 0 is not in the orbit closure of v, then clearly $f_{v,\inf} := \inf f_v > -\infty$. The norm minimization problem is to find $g \in G$ such that $f_v(g)$ is closed to $f_{v,\inf}$ and the scaling problem is to find $g \in G$ such that $\nabla f_v(g)$ is closed to 0. They applied the first order and second order algorithms to optimize f_v to consider these two problems. Because of the smoothness and convexity of f_v, the iteration complexity of these algorithms has an explicit formula. Furthermore, for some special cases, they showed these algorithms can be applied to solve the null cone membership.

This work is an extension of [2]. We consider the above problems for real reductive Lie group actions and extend some of the results in [2] to the real case without considering the complex structure. Let $G \subset G(n, \mathbb{R})$ be a real reductive Lie group and $K = G \cap O(n)$, the orthogonal group. Then $P \simeq G/K$, where $P = G \cap P(n)$, the set of all positive definite matrices. For any action of G on V, there is also an inner product $\langle \cdot, \cdot \rangle$ that is K-invariant. Therefore, we define the Kempf-Ness function $f_v \colon P \to \mathbb{R}$. Then the norm minization problem is to minimize f_v on P. Moreover, we equip P with the standard Riemannian structure. The scaling problem is to minimize the gradient ∇f_v on P.

Contributions and Results:

(1) f_v defined on P is geodesically convex and L-smooth for $L > 0$.

(2) Because of the smoothness of f_v, applying the Riemannian gradient descent (RGD) algorithm to f_v solves the scaling problem.

(3) By extending the result in the complex case [2],

$$f_v(x) - f_{v,\inf} \geqslant C \|\nabla f_v(x)\|_x ,$$

the RGD algorithm solves the norm minimization problem.

(4) We also analyze a general scaling problem on P by employing the results in [4].

3 Optimization on Riemannian Manifold

Let (π, V) be an action of real reductive Lie group G and $P = G \cap P(n)$. Defining the Kempf-Ness function $f_v(x) = \log\langle v, \pi(x)v\rangle$ for any $x \in P$.

Theorem 1. *f_v is geodesically convex on P.*

Given $\varepsilon > 0$ and $v \in V$ such that $f_{v,\inf} > -\infty$,

- Scaling Problem: find $x_s \in P$, such that
$$\|\nabla f_v(x_s)\|_{x_s} < \varepsilon.$$
- Norm Minimization Problem: find $x_n \in P$, such that
$$f_v(x_n) - f_{v,\inf} < \varepsilon.$$

These problems are related to optimizing f_v on P.

We apply the RGD algorithm to f_v, that is, let $x_0 = I$ and updating $x_1, \cdots, x_T$ by

$$x_{t+1} = x_t^{\frac{1}{2}} e^{-\eta x_t^{-\frac{1}{2}} \nabla f_v(x_t) x_t^{-\frac{1}{2}}} x_t^{\frac{1}{2}},$$

and then returning x_s such at $\|\nabla f_v(x_s)\|_{x_s}$ attaches the minimum of all $\|\nabla f_v(x_t)\|_{x_t}$. In order to find an appropriate T and η in RGD, we extend the definition of weight norm $N(\pi)$ in [2] to the real case and prove the smoothness of f_v.

Theorem 2. *f_v is $N(\pi)^2$-smooth, that is,*

$$\left|\frac{d^2}{dt^2} f\left(x^{\frac{1}{2}} e^{tX} x^{\frac{1}{2}}\right)\right| \leqslant N(\pi)^2 \|X\|_F^2,$$

for any $x \in P$ and any matrix X such that $e^X \in P$.

Therefore, for any $\varepsilon > 0$, by setting

$$\eta = \frac{1}{N(\pi)^2},\ T > \frac{2N(\pi)^2}{\varepsilon^2}\left(\log\|v\|^2 - f_{v,\inf}\right)$$

in the RGD algorithm, it returns x_s such that $\|\nabla f_v(x_s)\|_{x_s} < \varepsilon$ by [1].

For the norm minimization problem, we find a relation between $f_v(x) - f_{v,\inf}$ and $\|\nabla f_v(x)\|_x$, which is an extension of the complex case [2]. We define a parameter $\gamma(\pi)$, called the weight margin.

Theorem 3. *It holds that*

$$f_v(x) - f_{v,\inf} \geqslant \log\left(1 - \frac{\|\nabla f_v(x)\|_x}{\gamma(\pi)}\right)^{-1}.$$

Therefore, for any $0 < \varepsilon < \log 2$, if $\|\nabla f_v(x)\|_x$ is less than $\frac{1}{2}\gamma(\pi)\varepsilon$, then $f_v(x) - f_{v,\inf} < \varepsilon$.

For any nonzero $v \in V$, defining $\mu(v) = \nabla f_v(I)$, called the moment map. The moment polytope is

$$\Delta(v) := \left\{s\left(\mu(w)\right) : w \in \overline{\pi(G)\cdot v}\right\},$$

where $s\left(\mu(w)\right) = \operatorname{diag}\left(\lambda_1, \cdots, \lambda_n\right)$ with $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and all λ_i are eigenvalues of $\mu(w)$. Let $p \in \Delta(v)$. The p-scaling problem is to find $g \in G$ such that $\|s\left(\mu(\pi(g)v)\right) - p\|_F < \varepsilon$. Hirai [4] showed $f_v + b_p$ is bounded below for $p \in \Delta(v)$, where b_p is the Busemann function defined on P. They also showed

$$\|\nabla(f_v + b_p)(x)\|_x = \left\|\mu(\pi(x^{\frac{1}{2}})v) - ks(X)k^T\right\|_F,$$

where $x^{\frac{1}{2}} = bk$ is the RQ decomposition. Therefore, optimizing $f_v + b_p$ can solve the p-scaling problem.

4 Conlusions

This work optimized the Kempf-Ness function by the RGD algorithm. This method solves the scaling problem and the norm minimization problem of real reductive Lie group actions, because of the smoothness of the Kempf-Ness function. The p-scaling problem can also be viewed as an optimization problem by applying the Busemann function.

References

[1] N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, Cambridge, to appear.

[2] 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. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 845–861, 2019.

[3] L. Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004.

[4] H. Hirai. Convex analysis on Hadamard spaces and scaling problems. *arXiv preprint arXiv:2203.03193*, 2022.

[5] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics*, 35(2):876–879, 1964.