Sion's Minimax Theorem in Geodesic Metric Spaces and a Riemannian Extragradient Algorithm

author names withheld

Under Review for OPT 2023

Abstract

Deciding whether saddle points exist or are approximable for nonconvex-nonconcave problems is usually intractable. We take a step toward understanding a broad class of nonconvex-nonconcave minimax problems that *do remain* tractable. Specifically, we study minimax problems in geodesic metric spaces. The first main result of the paper is a geodesic metric space version of Sion's minimax theorem; we believe our proof is novel and broadly accessible as it relies on the *finite intersection property* alone. The second main result is a specialization to geodesically complete Riemannian manifolds, for which we analyze first-order methods for smooth minimax problems.

The full version of this paper is currently "in press" at the SIAM Journal on Optimization.

1. Introduction

We study minimax optimization problems of the form

$$\min_{x \in X} \max_{y \in Y} \quad f(x, y), \tag{1}$$

where the constraint sets X and Y lie in geodesic metric spaces, and f is a suitable bifunction. Problem (1) generalizes the standard Euclidean minimax problem where $X \subseteq \mathbb{R}^m, Y \subseteq \mathbb{R}^n$. Minimax problems as such have drawn great attention recently, e.g., in generative adversarial networks [8], robust learning [6, 14], multiagent reinforcement learning [3], adversarial training [9], etc.

A common goal of solving minimax problems is to find global saddle points[1]. A pair (x^*, y^*) is a *saddle point* if x^* is a minimum of $f(\cdot, y^*)$ and y^* is a maximum of $f(x^*, \cdot)$. In game theory, a saddle point is a special Nash equilibrium [15] for a two-player game. When f is convex-concave (i.e., convex in x and concave in y), existence of saddle points is guaranteed by Sion's minimax theorem [17], and their computation is often tractable (e.g., [16]). But without the convex-concave structure, saddle points may fail to exist, or even when they exist, computing them can be intractable [4]. Even computing local saddle points with linear constraints is PPAD-complete [5]. Therefore, it is natural to pose the following question:

Which nonconvex-nonconcave minimax problems admit saddle points, and can we compute them?

While at this level of generality this question is unlikely to admit satisfactory answers, it motivates us to pursue a more nuanced study, and to seek tractable subclasses of problems or alternative optimality criteria—e.g., the works [7, 10, 12] explore this topic and establish novel optimality criteria

1. Without further qualification, we refer to global saddle points as saddle points in this paper.

for nonconvex-nonconcave problems. We instead explore a rich subclass of nonconvex-nonconcave problems that do admit saddle points: minimax problems over *geodesic metric spaces* [2]. We provide sufficient conditions that ensure existence of saddle points by establishing a metric space analog of Sion's theorem. An informal statement of our first main result is as follows:

Theorem 1 (Informal; see Thm. 3) *Let X, Y be geodesically convex subsets of geodesic metric spaces $\mathcal{M}$ and $\mathcal{N}$, and let X be compact. If a bifunction $f : X \times Y \to \mathbb{R}$ is geodesically (quasi)-convex-concave and (semi)-continuous, then the equality* $\sup_{y\in Y} \min_{x\in X} f(x, y) = \min_{x\in X} \sup_{y\in Y} f(x, y)$ *holds.*

If we further assume that both X and Y are compact, then there exists a saddle point $(x^*, y^*) \in X \times Y$. Later in the paper, we will address computability of saddle points by focusing on the special case of Riemannian manifolds, for which we exploit the available differentiable structure to obtain first-order algorithms for the *Riemannian minimax* problem

$$\min_{x\in\mathcal{M}} \max_{y\in\mathcal{N}} \ f(x, y), \tag{P}$$

where $\mathcal{M}, \mathcal{N}$ are finite-dimensional complete and connected Riemannian manifolds, while $f : \mathcal{M} \times \mathcal{N} \to \mathbb{R}$ is a smooth geodesically convex-concave bifunction. When the manifolds in (P) are Euclidean, first-order methods such as optimistic gradient descent-ascent and extragradient can find saddle points efficiently [13, 16]. But in the Riemannian case, the extragradient steps do not succeed by merely translating Euclidean concepts into their Riemannian counterparts. We must account for the distortion caused by nonlinear geometry; to that end, we introduce an additional correction that offsets the distortion and thereby helps us obtain a Riemannian corrected extragradient (RCEG) algorithm, for which our main result is stated below.

Theorem 2 (Informal; see Thm. 5) *Under suitable conditions on the Riemannian manifolds $\mathcal{M}$, $\mathcal{N}$, RCEG admits a curvature-dependent $\mathcal{O}(\sqrt{\tau}/\epsilon)$ convergence to an ϵ-saddle point for geodesically convex-concave problems, where τ is a constant determined by curvature of the manifolds.*

Our analysis enables us to efficiently solve minimax problems in nonlinear spaces. This is verified by conducting experiments on SPD bilinear functions.

2. Preliminaries

Metric (geodesic) geometry. In a metric space $(\mathcal{M}, d_{\mathcal{M}})$, a path $\gamma : [0, 1] \to \mathcal{M}$ joining $x, y \in \mathcal{M}$ is called a *geodesic* if it is of constant speed and locally minimizing. $\mathcal{M}$ is called a *geodesic metric space* if any two points $x, y \in \mathcal{M}$ are joined by a geodesic. A non-empty set $X \subset \mathcal{M}$ is called a *geodesically convex set*, if every (not necessarily unique) geodesic connecting two points in X lies completely within X. A function $f : \mathcal{M} \to \mathbb{R}$ is geodesically convex, if for any $x, y \in \mathcal{M}$ and $t \in [0, 1]$, for any geodesic γ with $\gamma(0) = x$ and $\gamma(1) = y$, the following inequality holds: $f(\gamma(t)) \leq (1 - t)f(x) + tf(y)$. Moreover, we say f is geodesically quasi-convex if $f(\gamma(t)) \leq \max\{f(x), f(y)\}$; (concavity and quasi-concavity are defined by considering $-f$).

Riemannian geometry. An n-dimensional *manifold* is a topological space that is *locally* Euclidean. A smooth manifold is referred as a *Riemannian* manifold if it is endowed with a Riemannian metric $\langle\cdot, \cdot\rangle_x$ on the tangent space $T_x\mathcal{M}$, for each $x \in \mathcal{M}$. The metric induces a norm on the tangent space, denoted $\|\cdot\|_x$; we usually omit x when it causes no confusion.

A curve $\gamma : [0, 1] \to \mathcal{M}$ on Riemannian manifold is a geodesic if it is locally length-minimizing and of constant speed. An exponential map defines a mapping from tangent space $T_x\mathcal{M}$ to $\mathcal{M}$ as $\mathtt{Exp}_x(v) = \gamma(1)$, where γ is the geodesic with $\gamma(0) = x$ and $\gamma'(0) = v$. If geodesic is unique between any two points, we can define the inverse map as $\mathtt{Log}_x : \mathcal{M} \to T_x\mathcal{M}$. The exponential map also induces the Riemannian distance as $d_{\mathcal{M}}(x, y) = \|\mathtt{Log}_x(y)\|$. A *parallel transport* $\Gamma_x^y : T_x\mathcal{M} \to T_y\mathcal{M}$ provides a way of comparing vectors between different tangent spaces that preserves inner product, i.e., $\langle u, v\rangle_x = \langle \Gamma_x^y u, \Gamma_x^y v\rangle_y$ for $u, v \in T_x\mathcal{M}$. Unlike Euclidean space, a Riemannian manifold is not always flat and sectional curvature κ characterizes the distortion of geometry.

We will restrict our discussion to connected and complete manifolds, which admits at least a geodesic between any two points [11]. Hence, it is a geodesic metric space and inherits the definition of geodesically convex sets and geodesically convex/concave functions in geodesic space.

3. Main theorem: Minimax in Nonlinear Geometry

In Euclidean space, Sion's minimax theorem guarantees strong duality for convex-concave minimax problems. In this section, we establish an analog of Sion's theorem in geodesic metric spaces.

We consider the general form of (P) in geodesic metric spaces, i.e., $\mathcal{M}, \mathcal{N}$ are geodesic metric spaces, $f|_{X\times Y}$ is a geodesically (quasi-)convex-concave bifunction restricted to compact convex subset $X \subseteq \mathcal{M}$ and convex subset $Y \subseteq \mathcal{N}$. We present below our main theorem.

Theorem 3 (Sion's theorem in geodesic metric space) *Let $(\mathcal{M}, d_{\mathcal{M}})$ and $(\mathcal{N}, d_{\mathcal{N}})$ be geodesic metric spaces. Suppose $X \subseteq \mathcal{M}$ is a compact and geodesically convex set, and $Y \subseteq \mathcal{N}$ is a geodesically convex set. If the bifunction $f : X \times Y \to \mathbb{R}$ satisfies: (1) $f(\cdot, y)$ is geodesically-quasi-convex and lower semi-continuous; and (2) $f(x, \cdot)$ is geodesically-quasi-concave and upper semi-continuous, then we have the equality*

$$\min_{x\in X} \sup_{y\in Y} f(x, y) = \sup_{y\in Y} \min_{x\in X} f(x, y).$$

We specialize to the Riemannian minimax problem (P) and obtain immediately next corollary.

Corollary 4 *Suppose that $\mathcal{M}$ and $\mathcal{N}$ are finite-dimensional complete and connected Riemannian (sub)-manifolds. If subsets X, Y and the bifunction f satisfy the condition in Thm. 3, and additionally, Y is also compact, then the following min-max identity holds:*

$$\min_{x\in X} \max_{y\in Y} f(x, y) = \max_{y\in Y} \min_{x\in X} f(x, y).$$

By Cor. 4 we deduce that there is at least one saddle point (x^*, y^*) such that: $\min_{x\in X} f(x, y^*) = f(x^*, y^*) = \max_{y\in Y} f(x^*, y)$. If f is geodesically convex-concave, the minimax problem (P) can be tackled by closing the *duality gap*, defined for a given pair $(\hat{x}, \hat{y})$ as

$$\mathrm{gap}_f(\hat{x}, \hat{y}) := \max_{y} f(\hat{x}, y) - \min_{x} f(x, \hat{y}).$$

The duality gap then serves as an optimality criterion as in the Euclidean setup.

4. Riemannian Minimax Algorithms and Analysis

In this section we present our algorithm for minimax optimization of a geodesically convex-concave bifunction f on Riemannian manifolds under a suitable smoothness assumption. Building upon the aforementioned optimality criterion, we establish convergence rate of our algorithm via a non-asymptotic analysis. To this end, we assume the following regularity conditions.

Assumption 1 *The gradients of f are geodesically L-smooth, i.e., for any two pairs (x, y) and $(x', y') \in \mathcal{M} \times \mathcal{N}$, the gradient satisfies the bounds*

$$\|\nabla_x f(x, y) - \Gamma_{x'}^{x} \nabla_x f(x', y')\| \leq L \left(d_{\mathcal{M}}(x, x') + d_{\mathcal{N}}(y, y')\right),$$
$$\|\nabla_y f(x, y) - \Gamma_{y'}^{y} \nabla_y f(x', y')\| \leq L \left(d_{\mathcal{M}}(x, x') + d_{\mathcal{N}}(y, y')\right).$$

Assumption 2 *The bifunction $f(x, y)$ is geodesically convex-concave in (x, y).*

Further, we require the curvature of $\mathcal{M}$ and $\mathcal{N}$ to be bounded in range $[\kappa_{\min}, \kappa_{\max}]$. An additional bound on the diameter is necessary when positive curvature is involved, i.e., $\kappa_{\max} > 0$. It allows us to (1) use *comparison inequalities* (see [1, 19].), and (2) to ensure that the geodesic is unique between any two points [11], so that we can use the *inverse exponential map* $\texttt{Log}$.

Assumption 3 *The sectional curvatures of $\mathcal{M}, \mathcal{N}$ lie in the range $[\kappa_{\min}, \kappa_{\max}]$ with $\kappa_{\min} \leq 0$. Moreover, if $\kappa_{\max} > 0$, the diameter of the corresponding manifold is upper bounded by $\pi/\sqrt{\kappa_{\max}}$.*

Riemannian corrected extragradient. We present a Riemannian extragradient method with a correction term (RCEG) for geodesically convex-concave f (see Algorithm 1). We overload manifold operations to allow compact notation for the Riemannian gradient step of pair $(x, y) \in \mathcal{M} \times \mathcal{N}$:

$$\texttt{Exp}_{(x,y)}(u, v) := (\texttt{Exp}_x(u),\ \texttt{Exp}_y(v)). \tag{2}$$

We use a geodesic averaging scheme [18, 19] in Algorithm 1: at each iteration we calculate

$$(\overline{w}_{t+1}, \overline{z}_{t+1}) = \texttt{Exp}_{(\overline{w}_t, \overline{z}_t)}\left(\tfrac{1}{t+1} \cdot \texttt{Log}_{\overline{w}_t}(w_{t+1}),\ \tfrac{1}{t+1} \cdot \texttt{Log}_{\overline{z}_t}(z_{t+1})\right). \tag{3}$$

The following theorem shows that the averaged output of RCEG achieves a curvature-dependent convergence rate for smooth convex-concave f on Riemannian manifolds.

Algorithm 1: Riemannian Corrected Extragradient (RCEG)

Input: objective f, initialization (x_1, y_1), step-size η

$w_1 \leftarrow x_1$, $z_1 \leftarrow y_1$

for $t = 1, 2, \ldots, T$ **do**

  $(w_t, z_t) \leftarrow \texttt{Exp}_{(x_t, y_t)}(-\eta \nabla_x f(x_t, y_t), \eta \nabla_y f(x_t, y_t))$

  $(x_{t+1}, y_{t+1}) \leftarrow \texttt{Exp}_{(w_t, z_t)}(-\eta \nabla_x f(w_t, z_t) + \texttt{Log}_{w_t}(x_t), \eta \nabla_y f(w_t, z_t) + \texttt{Log}_{z_t}(y_t))$

end

Output: geodesic averaging scheme $(\overline{w}_T, \overline{z}_T)$ as in (3)

Theorem 5 *Suppose Assumptions 1–3 hold, and the iterations remain in subdomains[2] of bounded diameter $D_{\mathcal{M}}$ and $D_{\mathcal{N}}$. Let (x_t, y_t, w_t, z_t) be the sequence generated by Algorithm 1 initialized at*

2. The condition allows an upper-bound for distortion (*cf.* τ) and is regular in Riemannian optimization literature [1, 19].

$x_1 = w_1$, $y_1 = z_1$. *Then, using a step-size* $\eta = \frac{1}{2L\sqrt{\tau}}$*, the following bound holds for* T*:*

$$\max_{y \in \mathcal{N}} f(\overline{w}_T, y) - \min_{x \in \mathcal{M}} f(x, \overline{z}_T) \leq \frac{d^2_{\mathcal{M}}(x_1, x^*) + d^2_{\mathcal{N}}(y_1, y^*)}{\eta T},$$

with $(\overline{w}_T, \overline{z}_T)$ *obtained via averaging in* (3)*, and* $\tau = \tau([\kappa_{\min}, \kappa_{\max}], \max(D_{\mathcal{M}}, D_{\mathcal{N}}))$.

Thm. 5 is a natural nonlinear extension of the known result on extragradient method in the Euclidean setting. We notice that, different from Riemannian minimization algorithms (e.g., [19]), whenever the lower and upper bound of curvature coincide, the curvature-free convergence rate is retrieved.

5. Toy example: SPD Bilinear Function

We provide a synthetic test problem to illustrate the better convergence property of RCEG over Riemannian gradient descent-ascent (RGDA) methods. As the direct generalization of its Euclidean counterpart, RGDA formulates the following iteration

$$(x_{t+1}, y_{t+1}) = \texttt{Exp}_{(x_t, y_t)}(-\eta \nabla_x f(x_t, y_t), \eta \nabla_y f(x_t, y_t))$$

and is not guaranteed to convergence for convex-concave objectives. We empirically verify this by utilizing $f(x, y) = \langle \texttt{Log}_x(x_0), \texttt{Log}_y(y_0) \rangle_F$, where x, y belong to the same SPD manifold $\mathcal{P}(n)$ and $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product. Then f formalizes an analogy of Euclidean bilinear function. The result in Fig. 1 illustrates that, while our RCEG is convergent, similar to its Euclidean counterpart, the naive RGDA method can diverge for certain geodesically convex-concave objectives.

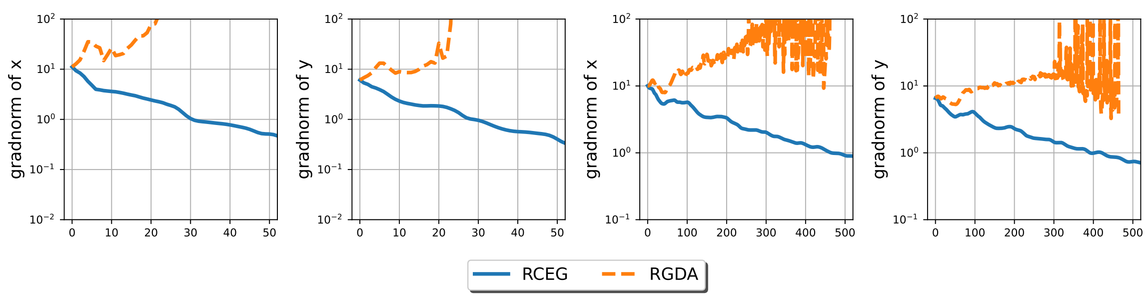

Figure 1: Comparison between RGDA and RCEG for bilinear objective. While RCEG is convergent, the RGDA method is divergent for minimax problem $f(x, y) = \langle \texttt{Log}_x(x_0), \texttt{Log}_y(y_0) \rangle_F$, where x, y are defined on $\mathcal{P}(100)$. We utilize a step-size $\eta = 0.2$.

6. Conclusion

In this work, we provide a new perspective into nonconvex-nonconcave minimax optimization and game theory by considering geodesic convex-concave problems in non-linear geometries. First, we provide an analog of Sion's theorem on geodesic metric spaces. Second, we provide novel and efficient minimax algorithm for a different class of geodesic convex-concave games on geodesically complete Riemannian manifolds. We believe our work takes a significant step towards understanding the properties of minimax problems in non-linear geometry, and should help inform the study of many structured learning problems on manifolds.

References

[1] Foivos Alimisis, Antonio Orvieto, Gary Bécigneul, and Aurelien Lucchi. A continuous-time perspective for modeling acceleration in Riemannian optimization. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1297–1307. PMLR, 2020.

[2] Dmitri Burago, Iu D Burago, Yuri Burago, Sergei Ivanov, Sergei V Ivanov, and Sergei A Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, Providence, RI, 2001.

[3] Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Trans. Syst. Man Cybern. Syst.*, 38(2):156–172, 2008.

[4] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM J. Comput.*, 39(1):195–259, 2009.

[5] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1466–1478, 2021.

[6] Laurent El Ghaoui and Hervé Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, 18(4):1035–1064, 1997.

[7] Tanner Fiez, Lillian J Ratliff, Eric Mazumdar, Evan Faulkner, and Adhyyan Narang. Global convergence to local minmax equilibrium in classes of nonconvex zero-sum games. In *Advances in Neural Information Processing Systems*. Curran Associates, 2021.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. Curran Associates, 2014.

[9] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *preprint, arXiv:1412.6572*, 2014.

[10] Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *Proceedings of the International Conference on Machine Learning*, pages 4880–4889. PMLR, 2020.

[11] John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer, Cham, Switzerland, 2006.

[12] Oren Mangoubi and Nisheeth K Vishnoi. Greedy adversarial equilibrium: an efficient alternative to nonconvex-nonconcave min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 896–909, 2021.

[13] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1497–1507. PMLR, 2020.

[14] Andrea Montanari and Emile Richard. Non-negative Principal Component Analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inform. Theory*, 62(3):1458–1484, 2015.

[15] John F Nash. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[16] Arkadi Nemirovski. Prox-method with rate of convergence $o(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[17] Maurice Sion. On general minimax theorems. *Pacific J. Math.*, 8(1):171–176, 1958.

[18] Nilesh Tripuraneni, Nicolas Flammarion, Francis Bach, and Michael I Jordan. Averaging stochastic gradient descent on riemannian manifolds. In *Proceedings of the Conference on Learning Theory*, pages 650–687. PMLR, 2018.

[19] Hongyi Zhang and Suvrit Sra. First-order methods for geodesically convex optimization. In *Proceedings of the Conference on Learning Theory*, pages 1617–1638. PMLR, 2016.