OPTIMIZATION ERUDITORUM
Volume 3 (2026), No. 1, 20–33
https://doi.org/10.69829/oper-026-0301-ta03
Tulipa Opera Scholarum

THE STOCHASTIC PROJECTED GRADIENT METHOD FOR OPTIMIZATION PROBLEMS ON HADAMARD MANIFOLDS

XIAO-BO LI[1] AND YU-KE ZHONG[1,*]

[1] *School of Sciences, Civil Aviation Flight University of China, 618300, People's Republic of China*

Abstract. In this paper, the stochastic projected gradient method for optimization problems is proposed on Hadamard manifolds, which utilize the Riemannian metric and curvature information. Under some specified assumptions on compactness and the problem geometry, the convergence rate of the stochastic projected gradient method for optimization problems is presented. Furthermore, assume in addition that f is σ-strongly convex for some $\sigma > 0$, some convergence results are also established on Hadamard manifolds.

Keywords. The stochastic projected gradient method, Convergence results, Optimization problems, Hadamard manifolds.

1. Introduction

The paradigm of optimization has been fundamentally reshaped by the geometry of the underlying space. For recent years, the implicit assumption has been a Euclidean space, where the setting is globally linear, and possesses a constant metric. Algorithms like stochastic gradient descent [3, 6, 20, 21], and its their variants(e.g.,Adam) have achieved remarkable success in this Euclidean framework, enabling the training of deep neural networks and large-scale statistical models. However, many high-impact problems are inherently non-Euclidean. For example, covariance matrix estimation, low-rank matrix completion, computational geometry and computer vision. In these cases, a naive Euclidean update $x_{k+1} = x_k - \eta_k g_k$ (where g_k is a stochastic gradient) is invalid, as x_{k+1} will almost certainly leave the manifold. Riemannian optimization addresses this by performing updates in the tangent space and then mapping the point back onto the manifold via a retraction, a process that inherently respects the geometry of the constraint set.

The stochastic component is equally vital in machine learning. Objective functions are often a finite sum over data points, i.e., $f(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell(\theta; x_i)$. For large N, computing the full gradient is computationally prohibitive. A stochastic gradient $g_k \approx \nabla f(\theta_k)$, estimated from a mini-batch, offers a computationally cheap, albeit noisy, alternative. Moreover, in online learning scenarios, where data arrives sequentially, the true objective may be an expectation over an unknown data distribution, making stochastic optimization the only feasible approach. In the Euclidean setting, some methods handle complex constraints, which can reduce computational expense. For instance, Necoara et al. [20] extended random projection ideas to problems with composite objective functions (a smooth part plus a possibly non-smooth part), combining random coordinate descent for updating the solution with respect to the objective and random projection for handling constraints. Based on the results of [20], Necoara et al. [21] introduced a rigorous analysis of the random projection method for large-scale convex optimization problems with numerous linear constraints: by randomly selecting a small block of constraints and

*Corresponding author.
E-mail address: lixiaobo.double@163.com (X.-B. Li), 18989132187@163.com (Y.-K. Zhong)
2020 Mathematics Subject Classification: 58C99; 90C26; 90C90.
Accepted: February 08, 2026.

performing a projection onto the hyperplane defined by that block, the authors proved convergence rates for both the objective function value and the overall algorithm. Furthermore, Bertsekas [6] proposed a general framework that encompasses random projection methods, and analyzed stochastic (sub)gradient and proximal gradient methods where the constraint set is complex. In their work, "random projection" emerges as a special case: the algorithm extracts a subgradient from a randomly selected constraint and takes a step that maintains feasibility with respect to that constraint.

In recent years, extending the theories and methods of optimization from Euclidean spaces to Riemannian manifolds has attracted significant attention in recent years; see, e.g., [1, 4, 5, 7–19, 22–24]. The extension of stochastic optimization algorithms from Euclidean space to Riemannian manifolds is a vibrant area of research. A more common and directly analogous theme is Riemannian stochastic gradient descent and Riemannian optimization, where its "stochasticity" stems from the selection of a random subset of data points. For instance, Riemannian stochastic gradient methods have been developed in [7, 24], which utilize the Riemannian metric and curvature information to iteratively minimize the objective function. Bonnabel [7] formally introduced and analyzed stochastic gradient descent on Riemannian manifolds, establishing a general framework for optimizing functions defined on manifolds when using only noisy (stochastic) gradient estimates. The algorithm is a direct generalization of euclidean stochastic gradient method: at each step, it takes a step in the direction of the stochastic Riemannian gradient, followed by a retraction back onto the manifold. The paper proves almost sure convergence to a critical point, establishing a rigorous theoretical basis for all subsequent stochastic manifold optimization algorithms. Furthermore, Zhang [24] analyzed algorithms for geodesically convex optimization on manifolds, which is the natural generalization of convexity. While it covers full-gradient methods, its significance for stochastic methods is profound. It introduces and analyzes Riemannian stochastic variance reduced gradient, a key algorithm for finite-sum problems. Riemannian stochastic variance reduced gradient reduces the variance of the stochastic gradients, leading to a much faster convergence rate than standard Riemannian stochastic variance reduced gradient. This work bridges the gap between modern, fast Euclidean stochastic algorithms and their Riemannian counterparts, providing both algorithm designs and non-asymptotic convergence guarantees.

In this paper, the Riemannian stochastic projected gradient method is proposed on Hadamard manifolds, which integrates ideas from optimization theory with Riemannian geometry to efficiently solve optimization problems on Hadamard manifolds. Since Riemannian manifolds generally lack a linear structure, conventional techniques in the Euclidean space cannot be applied and new techniques have to be proposed. Our results are distinguished by the following key aspects: First, we introduce the stochastic projected gradient method for optimization problems on Hadamard manifolds, and its convergence and complexity results are established, which generalizes some algorithm results in [6, 20] from $\mathbb{R}^n$ to Hadamard manifolds; Second, Riemannian manifolds encode intrinsic geometric properties, a generalized version of the theorem would respect these properties, ensuring that the stochastic projected gradients and stepsize rules align with the manifold's geometry; Third, from a theoretical point, extending stochastic gradient convergence results to Riemannian manifolds deepens our understanding of optimization in non-Euclidean settings. It requires refining concepts such as projection, gradient noise, and stepsize analysis, such generalizations enrich the theory of stochastic optimization, bridging it with differential geometry.

This work is organized as follows: In Section 2, some necessary definitions and concepts are provided on Hadamard manifolds. In Section 3, the stochastic projected gradient method for optimization problems is presented on Hadamard manifolds. In Section 4, under some specified assumptions on compactness and the problem geometry, the convergence property and the complexity result of the stochastic projected gradient method are obtained on Hadamard manifolds.

2. Preliminaries

In this section, some standard definitions and results from Riemannian manifolds are recalled, which can be found in some introductory books on Riemannian geometry; see for example [11, 17].

Let M be a finite-dimensional differentiable manifold and $x \in M$. The tangent space of M at x is denoted by T_xM and the tangent bundle of M by $TM = \bigcup_{x\in M} T_xM$. $\langle , \rangle_x$ is denoted by the inner product on T_xM with the associated norm $\|\cdot\|_x$. If there is no confusion, then the subscript x is omitted. If M is endowed with a Riemannian metric g, then M is a Riemannian manifold. Given a piecewise smooth curve $\gamma : [t_0, t_1] \to M$ joining x to y, that is, $\gamma(t_0) = x$ and $\gamma(t_1) = y$, the length of γ by $l(\gamma) = \int_a^b \|\gamma'(t)\|dt$ can be defined. Minimizing this length functional over the set of all curves, a Riemannian distance $d(x, y)$ which induces the original topology on M is obtained.

A Riemannian manifold is complete if for any $x \in M$, all geodesic emanating from x are defined for all $t \in \mathbb{R}$. By Hopf-Rinow theorem [23], any pair of points $x, y \in M$ can be joined by a minimal geodesic. The exponential mapping $\exp_x : T_xM \to M$ is defined by $\exp_x v = \gamma_v(1, x)$ for each $v \in T_xM$, where $\gamma(\cdot) = \gamma_v(\cdot, x)$ is the geodesic starting x with velocity v, that is, $\gamma(0) = x$ and $\gamma'(0) = v$. It is easy to see that $\exp_x tv = \gamma_v(t, x)$ for each real number t.

Let M be a Hadamard manifold, i.e., a complete, simply-connected Riemannian manifold with non-positive sectional curvature. The key property of Hadamard manifold is (CN) Inequality / Comparison Theorem: For any geodesic triangle $\triangle pqr$ in M with comparison triangle $\triangle \bar{p}\bar{q}\bar{r}$ in $\mathbb{R}^2$, the angles at corresponding vertices satisfy $\angle pqr \leq \angle \bar{p}\bar{q}\bar{r}$. A crucial consequence is the cosine inequality: for a triangle with vertices p, q, r, $a = d(p, q), b = d(p, r), c = d(q, r)$, and $A = \angle qpr$, we have: $c^2 \geq a^2 + b^2 - 2ab\cos A$.

Definition 2.1. [2] Let M be a Hadamard manifold. A subset C of M is said to be convex if, for any $x, y \in C$, the geodesic γ joining the endpoints x and y is contained in C; that is, if $\gamma : [0, 1] \to M$ is a geodesic such that $\gamma(0) = x$ and $\gamma(1) = y$, then $\gamma_{x,y}(t) = \exp_x(t \exp_x^{-1} y) \in C$ for any $t \in [0, 1]$.

Definition 2.2. [2] Let M be a Hadamard manifold and let $C \subseteq M$ be convex. A function $\phi : C \to \mathbb{R}$ is said to be convex if, for any $x, y \in C$,

$$\phi(\gamma_{x,y}(t)) \leq t\phi(x) + (1 - t)\phi(y),\ t \in [0, 1],$$

where $\gamma_{x,y}(t) = \exp_y t(\exp_y^{-1} x)$ for every $t \in [0, 1]$.

Definition 2.3. [2] Let M be a Hadamard manifold. Suppose that $f : M \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Let $x \in \mathrm{dom} f$ and $v \in T_xM$. The directional derivative in direction v and the subdifferential f are, respectively, defined by

$$f^{'}(x; v) = \lim_{t\to 0^+} \frac{f(\exp_x tv) - f(x)}{t},$$

and

$$\partial f(x) := \{u \in T_xM : \langle u, v\rangle \leq f^{'}(x; v), \forall v \in T_xM\}.$$

Let $C \subset M$ be a non-empty, closed, and convex set. The projection operator $P_C : M \to C$ is defined by: $P_C(x) = \arg\min_{y\in C} d(y, x)$. The non-positive curvature ensures that $P_C(x)$ exists and is unique for all $x \in M$. A fundamental characterization of the projection is the following variational inequality.

Lemma 2.4. Let M be a Hadamard manifold. Let $x \in M$ and $p = P_C(x)$. Then, for every $q \in C$, we have

$$\langle \exp_p^{-1} x,\ \exp_p^{-1} q\rangle \leq 0.$$

Proof. Let $q \in C$ be arbitrary, and let $\gamma : [0, 1] \to C$ be the unique geodesic from $p = \gamma(0)$ to $q = \gamma(1)$. Since C is convex, we know $\gamma(t) \in C, t \in [0, 1]$. We consider the following function:

$$\phi(t) = \frac{1}{2} d^2(\gamma(t), x).$$

Since $p = P_C(x)$ and $\phi(0) = \frac{1}{2} d^2(p, x)$, it follows that $\phi'(0) \geq 0$. The derivative of the distance squared function on a Hadamard manifold is given by

$$\phi'(0) = -\langle \exp_p^{-1} x,\, \gamma'(0) \rangle.$$

Since $\gamma'(0) = \exp_p^{-1} q$, we have

$$-\langle \exp_p^{-1} x,\, \exp_p^{-1} q \rangle \geq 0,$$

which is equivalent to the stated inequality. □

Lemma 2.5. Let M be a Hadamard manifold. The projection operator P_C is non-expansive. That is, for any $x, y \in M$,

$$d(P_C(x), P_C(y)) \leq d(x, y).$$

Proof. Let $x, y \in M$ be arbitrary. Define $p = P_C(x), q = P_C(y)$. For the point x and its projection p , taking the other point in C to be q , from Lemma 2.4, we have

$$\langle \exp_p^{-1} x,\, \exp_p^{-1} q \rangle \leq 0. \tag{2.1}$$

For the point y and its projection q, taking the other point in C to be p , from Lemma 2.4, we get

$$\langle \exp_q^{-1} y,\, \exp_q^{-1} p \rangle \leq 0. \tag{2.2}$$

Now, consider the geodesic triangle $\triangle pqx$. Denote its side lengths: $a = d(p, q), b = d(p, x), c = d(q, x)$. Let $A = \angle qpx$ be the angle at vertex p. By the definition of the angle and the Riemannian metric, we know

$$\cos A = \frac{\langle \exp_p^{-1} q,\, \exp_p^{-1} x \rangle}{\| \exp_p^{-1} q \| \, \| \exp_p^{-1} x \|} = \frac{\langle \exp_p^{-1} q,\, \exp_p^{-1} x \rangle}{a\, b}.$$

From inequality (2.1), we know $\cos A \leq 0$.

We now apply the cosine inequality for Hadamard manifolds to triangle $\triangle pqx$:

$$c^2 \geq a^2 + b^2 - 2ab \cos A.$$

Since $\cos A \leq 0$, it follows that $-2ab \cos A \geq 0$. Therefore, we can obtain the inequality:

$$d^2(q, x) = c^2 \geq a^2 + b^2 = d^2(p, q) + d^2(p, x). \tag{2.3}$$

By a symmetric argument, applying the cosine inequality to triangle $\triangle pqy$ and using inequality (2.2), it follows that

$$d^2(p, y) \geq d^2(p, q) + d^2(q, y). \tag{2.4}$$

We now use a fundamental distance inequality in CAT(0) spaces, which is a consequence of the (CN) comparison property [2]. For any four points p, q, x, y in a Hadamard manifold, the following holds:

$$d^2(p, y) + d^2(q, x) \leq d^2(p, x) + d^2(q, y) + 2d(p, q)d(x, y). \tag{2.5}$$

Substitute the lower bounds from (2.3) and (2.4) into (2.5), we know

$$\left(d^2(p, q) + d^2(q, y)\right) + \left(d^2(p, q) + d^2(p, x)\right) \leq d^2(p, x) + d^2(q, y) + 2d(p, q)d(x, y).$$

Then we have

$$2d^2(p, q) \leq 2d(p, q)d(x, y).$$

If $d(p,q)=0$, the theorem holds trivially. Otherwise, we can divide both sides by $2d(p,q)$ to obtain the desired result:

$$d(p,q) \leq d(x,y).$$

That is,

$$d(P_C(x), P_C(y)) \leq d(x,y).$$

□

3. The Stochastic Projected Gradient Method on Hadamard Manifolds

In this section, we assume that M is a Hadamard manifold. The main model that will be discussed is

$$\min_{x \in C} f(x), \tag{3.1}$$

where $C \subseteq M$ is nonempty, closed, and convex.

Assumption 1.

(i) $f : M \to \mathbb{R} \cup \{+\infty\}$ is proper, closed, and convex;

(ii) $C \subseteq \operatorname{intdom}(f)$;

(iii) The optimal set of (3.1) is nonempty and denoted by X^*, the optimal value of problem (3.1) is denoted by f^*.

Now, the stochastic projected gradient method on Hadamard manifolds is introduced as follows.

Algorithm The stochastic projected gradient method on Hadamard manifolds

Initialization: pick $x^0 \in C$ arbitrarily.

General step: for any $k = 0, 1, \dots,$ compute the following steps:

pick a step size $t_k > 0$, and a random vector $g^k \in T_{x^k}M$;

set $x^{k+1} = \operatorname{proj}_C(\exp_{x^k}(-t_k g^k))$.

Stopping criteria: If $g^k = 0$, terminate the algorithm.

Since the vectors g^k are random vectors, so are the iterate vectors x^k. The assumption on the random vectors g^k are given below.

Assumption 2.

(i) For any $k \geq 0$, $E(g^k|x^k) \in \partial f(x^k)$;

(ii) There exists a constant $L_f > 0$ such that for any $k \geq 0$,

$$E(\|g^k\|^2|x^k) \leq L_f^2.$$

It follows from (i) of Assumption 2 that g^k is an unbiased estimator of a subgradient at x^k. That is,

$$f(z) \geq f(x^k) + \langle E(g^k|x^k), \exp_{x^k}^{-1} z\rangle, \ \forall z \in \operatorname{dom} f.$$

Lemma 3.1. [3] Let $d \in \mathbb{R}$. Then for any $k \geq 2$,

$$\frac{d + \sum_{n=[k/2]}^{k} \frac{1}{n+1}}{\sum_{n=[k/2]}^{k} \frac{1}{\sqrt{n+1}}} \leq \frac{4(d + \log(3))}{\sqrt{k+2}}.$$

4. The Convergence Result

In this section, under some assumptions, the analysis of the stochastic projected gradient method for optimization problems is introduced on Hadamard manifolds. The sequence of best achieved function values, which is defined by $f_{best}^k = \min_{n=0,1,\dots,k} f(x^n)$. It is easy to check that the sequence $\{f_{best}^k\}_{k\geq 0}$ is non-increasing.

Theorem 4.1. (Convergence of the stochastic projected gradient). Suppose that Assumptions 1 and 2 hold. Let $\{x^k\}_{k\geq 0}$ be the sequence generated by Algorithm with positive stepsizes $\{t_k\}_{k\geq 0}$.

(i) If $\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \to 0$ as $k\to\infty$, then $E(f_{best}^k)\to f^*$ as $k\to\infty$.

(ii) Assume that C is compact. Let Θ be an upper bound on the half-squared diameter of C:

$$\max_{x,y\in C}\frac{1}{2}d^2(x,y)\leq \Theta.$$

If $t_k=\frac{\sqrt{2\Theta}}{L_f\sqrt{k+1}}$, then for any $k\geq 2$,

$$E(f_{best}^k)-f^*\leq \frac{\delta L_f\sqrt{2\Theta}}{\sqrt{k+2}},$$

where $\delta = 2(1+\log(3))$.

Proof. (i) For any $k\geq 0$, $x^*\in X^*$, $x^{n+1}=\text{proj}_C(\exp_{x^n}(-t_ng^n))$, $x^*=\text{proj}_C(x^*)$, it follows from Lemma 2.5 that

$$d^2(x^{n+1},x^*)=d^2(\text{proj}_C(\exp_{x^n}(-t_ng^n)),\text{proj}_C(x^*))\leq d^2(\exp_{x^n}(-t_ng^n),x^*).$$

In Euclidean space, for the points $x, y, \exp_x(v)$, we have the law of cosine equality:

$$d^2(\exp_{x^n}(-t_ng^n),x^*)=d^2(x^n,x^*)+\|-t_ng^n\|^2-2\langle -t_ng^n,\exp_{x^n}^{-1}x^*\rangle.$$

Conversely, on a Hadamard manifold, because triangles are "skinnier than Euclidean triangles," the corresponding distance satisfies an inequality rather than an equality, this implies that

$$d^2(\exp_{x^n}(-t_ng^n),x^*)\leq d^2(x^n,x^*)+\|-t_ng^n\|^2-2\langle -t_ng^n,\exp_{x^n}^{-1}x^*\rangle.$$

By Assumptions 1 and 2, we know

$$\begin{aligned} E(d^2(x^{n+1},x^*)|x^n) &= E(d^2(\exp_{x^n}(-t_ng^n),x^*)|x^n)\\ &\leq t_n^2E(\|g^n\|^2|x^n)+2t_nE(\langle g^n|x^n,\exp_{x^n}^{-1}x^*\rangle)+d^2(x^n,x^*)\\ &\leq t_n^2E(\|g^n\|^2|x^n)+2t_n(f^*-f(x^n))+d^2(x^n,x^*)\\ &\leq t_n^2L_f^2+2t_n(f^*-f(x^n))+d^2(x^n,x^*).\end{aligned}$$

Taking expectation with respect to x^n, it holds that

$$E(d^2(x^{n+1},x^*))\leq E(d^2(x^n,x^*))-2t_n(E(f(x^n))-f^*)+t_n^2L_f^2.$$

Summing over $n=m,m+1,\dots,k$, we obtain

$$E(d^2(x^{k+1},x^*))\leq E(d^2(x^m,x^*))-2\sum_{n=m}^k t_n(E(f(x^n))-f^*)+L_f^2\sum_{n=m}^k t_n^2.$$

Thus,

$$\sum_{n=m}^k t_n(f(x^n)-f^*)\leq \frac{1}{2}[E(d^2(x^m,x^*))+L_f^2\sum_{n=m}^k t_n^2],$$

which implies that

$$\sum_{n=m}^k t_n(\min E(f(x^n))-f^*)\leq \frac{1}{2}[E(d^2(x^m,x^*))+L_f^2\sum_{n=m}^k t_n^2].$$

Since $E(\min\{X_1,X_2,\dots,X_p\})\leq \min_{i=1,2,\dots,p}E(X_i)$, it holds that

$$E(f_{best}^k)-f^*\leq \frac{E(d^2(x^m,x^*))+L_f^2\sum_{n=m}^k t_n^2}{2\sum_{n=m}^k t_n}. \tag{4.1}$$

Let $m = 0$ in (4.1). Then

$$E(f_{best}^k) - f^* \leq \frac{E(d^2(x^0, x^*)) + L_f^2 \sum_{n=0}^k t_n^2}{2\sum_{n=0}^k t_n}.$$

If $\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \to 0$, then $E(f_{best}^k) \to f^*$ as $k \to \infty$.

To show the validity of (ii). Using (4.1) with $m = [k/2]$, and the bound $\max_{x,y\in C} \frac{1}{2}d^2(x,y) \leq \Theta, t_n = \frac{\sqrt{2\Theta}}{L_f\sqrt{n+1}}$, it follows that

$$E(f_{best}^k) - f^* \leq \frac{L_f\sqrt{2\Theta}}{2} \cdot \frac{1 + \sum_{n=[k/2]}^k \frac{1}{n+1}}{\sum_{n=[k/2]}^k \frac{1}{\sqrt{n+1}}},$$

which, combined with Lemma 3.1, yields the desires result. □

In Algorithm , if $g^k \in \partial f(x^k)$, then Algorithm reduces to the projected gradient method on Hadamard manifolds.

Lemma 4.2. Suppose that Assumption 1 holds. Let $\{x_k\}$ be the sequence generated by Algorithm with $g^k \in \partial f(x^k)$. Then for any $x^* \in X^*$ and $k \geq 0$,

$$d^2(x^{k+1}, x^*) \leq d^2(x^k, x^*) - 2t_k(f(x^k) - f^*) + t_k^2\|g^k\|^2.$$

Proof. Similar to the proof of Theorem 4.1, we have

$$\begin{aligned} d^2(x^{k+1}, x^*) &= d^2(\text{proj}_C(\exp_{x^k} - t_k g^k), \text{proj}_C x^*) \\ &\leq d^2(\exp_{x^k} - t_k g^k, \exp_{x^k}\exp_{x^k}^{-1} x^*) \\ &\leq t_k^2\|g^k\|^2 + \|\exp_{x^k}^{-1} x^*\|^2 + 2t_k\langle g^k, \exp_{x^k}^{-1} x^*\rangle \\ &\leq t_k^2\|g^k\|^2 + d^2(x^k, x^*) + 2t_k(f^* - f(x^k)). \end{aligned}$$

□

Theorem 4.3. Suppose that Assumptions 1 and 2 hold. Assume that C is convex and compact. Let Θ be an upper bound on the half-squared diameter of C:

$$\max_{x,y\in C} \frac{1}{2}d^2(x,y) \leq \Theta.$$

Let $\{x^k\}_{k\geq 0}$ be the sequence generated by Algorithm with $g^k \in \partial f(x^k)$ and stepsize chosen as either

$$t_k = \frac{\sqrt{2\Theta}}{L_f\sqrt{k+1}}$$

or

$$t_k = \begin{cases} \frac{\sqrt{2\Theta}}{\|g^k\|\sqrt{k+1}}, g^k \neq 0, \\ \frac{\sqrt{2\Theta}}{L_f\sqrt{k+1}}, g^k = 0. \end{cases}$$

Then for all $k \geq 2$,

$$f_{best}^k - f^* \leq \frac{\delta L_f\sqrt{2\Theta}}{\sqrt{k+1}},$$

where $\delta = 2(1 + \log(3))$.

Proof. By Lemma 4.2, for any $n \geq 0$,

$$\frac{1}{2}d^2(x^{n+1}, x^*) \leq \frac{1}{2}d^2(x^n, x^*) - t_n(f(x^n) - f^*) + \frac{1}{2}t_n^2\|g^n\|^2.$$

Summing the above inequality over $n = [k/2], [k/2] + 1, \ldots, k$, it holds that

$$\begin{aligned}\sum_{n=[k/2]}^{k} t_n(f(x^n) - f^*) &\leq \frac{1}{2}d^2(x^{[k/2]}, x^*) - \frac{1}{2}d^2(x^{k+1}, x^*) + \sum_{n=[k/2]}^{k} \frac{t_n^2}{2}\|g^n\|^2 \\ &\leq \Theta + \sum_{n=[k/2]}^{k} \frac{t_n^2}{2}\|g^n\|^2 \\ &\leq \Theta + \Theta \sum_{n=[k/2]}^{k} \frac{1}{n+1}. \qquad (4.2)\end{aligned}$$

Since $t_n \geq \frac{\sqrt{2\Theta}}{L_f(n+1)}$ and $f(x^n) \geq f_{best}^k$ for all $n \leq k$, it follows that

$$\sum_{n=[k/2]}^{k} t_n(f(x^n) - f^*) \geq (\sum_{n=[k/2]}^{k} \frac{\sqrt{2\Theta}}{L_f\sqrt{n+1}})(f_{best}^k - f^*). \qquad (4.3)$$

Therefore, combining (4.2) and (4.3) yields that

$$f_{best}^k - f^* \leq \frac{L_f\sqrt{2\Theta}}{2} \cdot \frac{1 + \sum_{n=[k/2]}^{k} \frac{1}{n+1}}{\sum_{n=[k/2]}^{k} \frac{1}{\sqrt{n+1}}},$$

which combined with Lemma 3.1, yields the desired result. □

Example 4.4. (minimization of sum of convex functions) Consider the optimization problem

$$(P) \ \min_{x \in C} f(x) = \sum_{i=1}^{m} f_i(x),$$

where $f_1, f_2, \ldots, f_m : M \to \mathbb{R} \cup \{+\infty\}$ are proper closed and convex functions. Suppose that C is convex, compact and Assumption 1 holds. In addition, assume that for any $i = 1, 2, \ldots, m$, there exists a constant $\tilde{L}_{f_i}$ for which

$$\|g\| \leq \tilde{L}_{f_i}, \forall g \in \partial f_i(x), x \in C,$$

where $\tilde{L}_{f_i}$ is a Lipschitz constant of f_i over C. The first method to solve (P) is the projected gradient method on Hadamard manifolds.

Algorithm The projected gradient method on (P)

Initialization: pick $x^0 \in C$ arbitrarily.
General step: for any $k = 0, 1, \ldots,$ choose $g_i^k \in \partial f_i(x^k)$, and compute:

$$x^{k+1} = \text{proj}_C(\exp_{x^k}(-\frac{\sqrt{2\Theta}}{\|\sum_{i=1}^{m} g_i^k\|\sqrt{k+1}} \sum_{i=1}^{m} g_i^k)).$$

By Theorem 4.3, for any $k \geq 2$, we know

$$f_{best}^k - f^* \leq \frac{\delta \tilde{L}_f\sqrt{2\Theta}}{\sqrt{k+2}},$$

where $\delta = 2(1+\log(3))$ and $\tilde{L}_f$ is a Lipschitz constant of f over C . For $k \geq N_1 = \max\{\frac{2\delta^2 \tilde{L}_f^2 \Theta}{\epsilon^2} - 2, 2\}$, an ϵ-optimal solution is obtained. Since the computation of $\sum_{i=1}^m g_i^k$ might be too expensive in cases m is large, the stochastic projected subgradient method at iteration k can be employed, the unbiased estimate of g_i^k is defined as

$$g^k = m g_{i_k}^k,$$

where i_k is randomly picked from $\{1, 2, \ldots, m\}$ via a uniform distribution. Obviously,

$$E(g^k | x^k) = \sum_{n=1}^m \frac{1}{m} m g_i^k = \sum_{n=1}^m g_i^k \in \partial f(x^k).$$

Also,

$$E(\|g^k\|^2 | x^k) = \sum_{n=1}^m \frac{1}{m} m^2 \|g_i^k\|^2 \leq m \sum_{i=1}^m \tilde{L}_{f_i} \equiv L_f^2.$$

The stochastic projected gradient method on (P) takes the following form.

Algorithm The stochastic projected gradient method on (P)

Initialization: pick $x^0 \in C$ arbitrarily .

General step: for any $k = 0, 1, \ldots,$ compute the following steps:

pick $i_k \in \{1, 2, \ldots\}$ randomly via a uniform distribution and $g_{i_k}^k \in \partial f_{i_k}(x^k)$;

compute $x^{k+1} = \text{proj}_C(\exp_{x^k}(-\frac{\sqrt{2\Theta} m}{L_f \sqrt{k+1}} g_{i_k}^k))$, where $L_f = \sqrt{m \sum_{i=1}^m \tilde{L}_{f_i}^2}$.

From Theorem 4.1, it follows that

$$E(f_{best}^k) - f^* \leq \frac{\delta \sqrt{m \sum_{i=1}^m \tilde{L}_{f_i}^2} \sqrt{2\Theta}}{\sqrt{k+2}}.$$

In particular, $N_2 = \max\left\{\frac{2\delta^2 m \Theta \sum_{i=1}^m \tilde{L}_{f_i}^2}{\epsilon^2} - 2, 2\right\}$ iterations are sufficient in order to ensure that an ϵ-optimal solution.

In the following, assume that f is σ-strongly convex for some $\sigma > 0$, the convergence of the stochastic projected gradient is obtained on Hadamard manifolds.

Theorem 4.5. (convergence of the stochastic projected gradient for strongly convex functions) Suppose that Assumptions 1 and 2 hold. Assume in addition that f is σ-strongly convex for some $\sigma > 0$. Let $\{x^k\}_{k \geq 0}$ be the sequence generated by Algorithm with positive stepsizes $t_k = \frac{2}{\sigma(k+1)}$.

(i) Then for any $k \geq 0$,

$$E(f_{best}^k) - f^* \leq \frac{2L_f^2}{\sigma(k+1)}.$$

(ii) Define $x^k = \exp_{x^*}(\sum_{n=0}^k \alpha_n^k \exp_{x^*}^{-1} x^n)$, where $\alpha_n^k = \frac{2n}{k(k+1)}$. Then

$$E(f(x^k)) - f^* \leq \frac{2L_f^2}{\sigma(k+1)}.$$

Proof. For any $x^* \in X^*$ and $n \geq 0$, similar to the proof of Theorem 4.1, it follows that

$$\begin{aligned} E(d^2(x^{n+1}, x^*)|x^n) &= E(d^2(\text{proj}_C \exp_{x^n}(-t_n g^n), \text{proj}_C x^*)|x^n) \\ &\leq E(d^2(\exp_{x^n}(-t_n g^n)), \exp_{x^n}(\exp_{x^n}^{-1} x^*))|x^n) \\ &\leq t_n^2 E(\|g^n\|^2|x^n) + 2t_n\langle E(g^n|x^n), \exp_{x^n}^{-1} x^*\rangle + d^2(x^n, x^*). \end{aligned} \tag{4.4}$$

Obviously, f is σ-strongly convex on M if and only if $f \circ \exp_x$ is σ-strongly convex on $T_x M$ for every $x \in M$. Then we obtain from [3] that

$$f(\exp_{x^n}(\exp_{x^n}^{-1} x^*)) \geq f(\exp_{x^n} 0_{x^n}) + \langle E(g^n|x^n), \exp_{x^n}^{-1} x^*\rangle + \frac{\sigma}{2}\|\exp_{x^n}^{-1} x^*\|^2.$$

That is,

$$f(x^*) \geq f(\exp_{x^n} 0_{x^n}) + \langle E(g^n|x^n), \exp_{x^n}^{-1} x^*\rangle + \frac{\sigma}{2} d^2(x^n, x^*).$$

Plugging the above into (4.4), it follows that

$$E(d^2(x^{n+1}, x^*)|x^n) \quad \leq \quad t_n^2 E(\|g^n\|^2|x^n) + 2t_n(f^* - f(x^n)) + (1 - \sigma t_n) d^2(x^n, x^*).$$

Dividing by $2t_n$, and using $E(\|g^n\|^2|x^n) \leq L_f^2$ yields

$$f(x^n) - f(x^*) \leq \frac{t_n}{2} L_f^2 + \frac{1}{2}(t_n^{-1} - \sigma) d^2(x^n, x^*) - \frac{1}{2t_n} E(d^2(x^{n+1}, x^*)|x^n). \tag{4.5}$$

Let $t_n = \frac{2}{\sigma(n+1)}$ in (4.5). Then

$$f(x^n) - f(x^*) \leq \frac{1}{\sigma(n+1)} L_f^2 + \frac{\sigma(n-1)}{4} d^2(x^{n+1}, x^*) - \frac{\sigma(n+1)}{4} E(d^2(x^{n+1}, x^*)|x^n).$$

Multiplying the above by n and taking expectation with respect to x^n yields

$$n(E(f(x^n)) - f^*) \leq \frac{\sigma n(n+1)}{4} E(d^2(x^n, x^*)) - \frac{\sigma n(n+1)}{4} E(d^2(x^{n+1}, x^*)) + \frac{n}{\sigma(n+1)} L_f^2.$$

Summing over $n = 0, 1, \ldots, k$,

$$\sum_{n=0}^{k} n(E(f(x^n)) - f^*) \leq -\frac{\sigma}{4} k(k+1) E(d^2(x^{k+1}, x^*)) + \frac{L_f^2}{\sigma} \sum_{n=0}^{k} \frac{n}{n+1} \leq \frac{L_f^2 k}{\sigma}. \tag{4.6}$$

Since $E(f(x^n)) \geq E(f_{best}^k)$ for all $n = 0, 1, \ldots, k$, it follows that

$$\sum_{n=0}^{k} n(E(f_{best}^k) - f^*) \leq \frac{L_f^2 k}{\sigma},$$

that is,

$$\frac{k(k+1)}{2}(E(f_{best}^k) - f^*) \leq \frac{L_f^2 k}{\sigma}.$$

Then,

$$E(f_{best}^k) - f^* \leq \frac{2L_f^2}{\sigma(k+1)}.$$

(ii) Divide (4.6) by $\frac{k(k+1)}{2}$ to obtain

$$\sum_{n=0}^{k} \alpha_n^k (E(f(x^n) - f^*) \leq \frac{2L_f^2}{\sigma(k+1)}.$$

This implies that

$$\sum_{n=0}^{k} \alpha_n^k (E(f(\exp_{x^*}(\exp_{x^*}^{-1} x^n)) - f^*) \leq \frac{2L_f^2}{\sigma(k+1)}.$$

That is,

$$\sum_{n=0}^{k} \alpha_n^k (E(\hat{f}(\exp_{x^*}^{-1} x^n) - f^*) \leq \frac{2L_f^2}{\sigma(k+1)},$$

where $\hat{f} = f \circ \exp_{x^*}$. By Jensen's inequality, it follows that

$$\begin{aligned} E(f(x^k)) - f^* &= E(f(\exp_{x^*} \sum_{n=0}^{k} \alpha_n^k (\exp_{x^*}^{-1} x^n)) - f^* = E(\hat{f}(\sum_{n=0}^{k} \alpha_n^k (\exp_{x^*}^{-1} x^n)) - f^* \\ &\leq \sum_{n=0}^{k} \alpha_n^k (E(\hat{f}(\exp_{x^*}^{-1} x^n) - f^*) \leq \frac{2L_f^2}{\sigma(k+1)}. \end{aligned}$$

□

Example 4.6. Let

$$M = H^2 := \{(x_1, x_2) \in \mathbb{R}^2 | x_2 > 0\}$$

be the upper half-plane model of hyperbolic space. From [1, 10, 12], the upper half-plane has constant negative curvature $K = -1$. Thus, H^2 is a Hadamard manifold. Taking $\bar{x} = (\bar{x}_1, \bar{x}_2) \in H^2$, we get $T_{\bar{x}} H^2 = \mathbb{R}^2$. The hyperbolic metric in the upper half-plane model is $ds^2 = \frac{d\bar{x}_1^2 + d\bar{x}_2^2}{\bar{x}_2^2}$. For points (x_1, x_2) and (y_1, y_2) in $\mathbb{H}^2$, the distance is:

$$d = \operatorname{arcosh}\left(1 + \frac{(y_1 - x_1)^2 + (y_2 - x_2)^2}{2x_2 y_2}\right).$$

For a point $p = (x_1, x_2)$ and tangent vector $v = (v_1, v_2)$:

$$\exp_p(v) = \left(x_1 + x_2 \frac{v_1}{\|v\|} \sinh(\|v\|), \quad x_2 \cosh(\|v\|) + \frac{x_2 v_2}{\|v\|} \sinh(\|v\|)\right),$$

where $\|v\| = \sqrt{v_1^2 + v_2^2}/x_2$ is the hyperbolic norm. Furthermore, For points $p = (x_1, x_2)$ and $q = (y_1, y_2)$,

$$\log_p(q) = \frac{d(p,q)}{\sinh(d(p,q))} \left(x_2(y_1 - x_1), \quad x_2 y_2 \cosh(d(p,q)) - x_2^2\right).$$

We consider the following optimization problem:

$$\min_{x \in M} f(x) = (x_1^2 + (x_2 - 2)^2)/x_2^2.$$

Let $x^0 = (1.0, 3.0)^T$ be the initial point. The tolerance, given by the norm of gradient of the function is set to be 10^{-8}. From Figure 1, we know the stochastic projected gradient method offers superior initial convergence speed, making it effective for quick solutions.

Example 4.7. Let

$$M = \mathcal{P}(n) := \{P \in \mathbb{R}^{n \times n} : P = P^T, P \succ 0\}$$

be the set of symmetric positive definite matrices. From [1,10,12], under the affine-invariant metric, $\mathcal{P}(n)$ has non-positive sectional curvature $K \leq 0$. Thus, $\mathcal{P}(n)$ is a Hadamard manifold. At any point $P \in \mathcal{P}(n)$, the tangent space consists of all $n \times n$ symmetric matrices U, V : $T_P \mathcal{P}(n) = \{U \in \mathbb{R}^{n \times n} : U = U^T\}$.

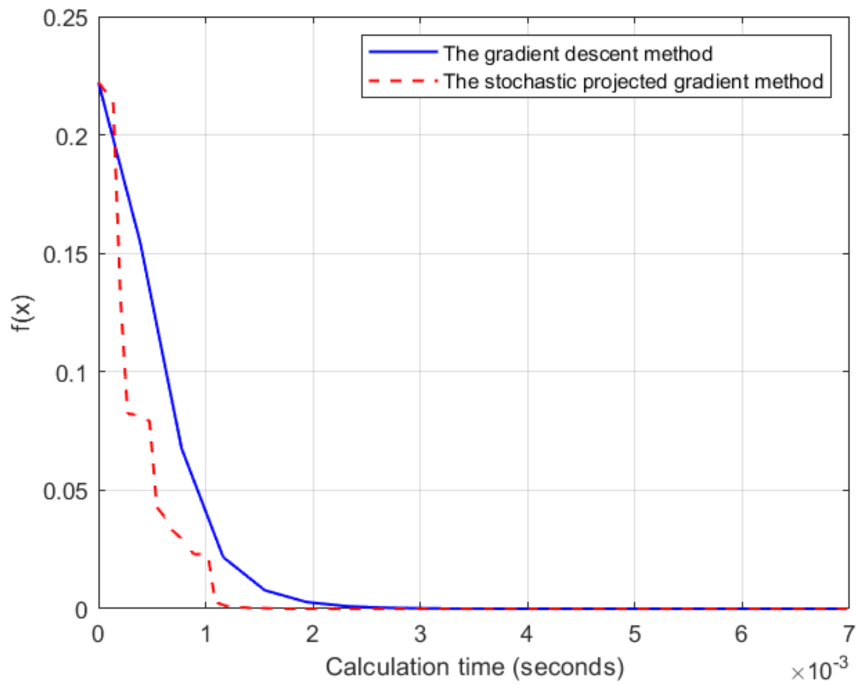

FIGURE 1. The comparison of convergence speed

Because the manifold $\mathcal{P}(n)$ itself is composed of symmetric matrices, so tangent vectors along curves on the manifold must also be symmetric.

For any two tangent vectors $U, V \in T_P\mathcal{P}(n)$, their inner product at P is defined as $g_P(U,V) = \operatorname{tr}(P^{-1}UP^{-1}V)$. Geodesics generalize the notion of "straight lines" on a manifold and are the paths generated by the exponential map. The geodesic curve connecting P and Q is given by:

$$\gamma(t) = P^{1/2}(P^{-1/2}QP^{-1/2})^t P^{1/2}, \quad t \in [0,1].$$

If $P = I$, this simplifies to $\gamma(t) = Q^t$. The exponential map $\exp_P(U)$ is as follows:

$$\exp_P(U) = P^{1/2}\exp(P^{-1/2}UP^{-1/2})P^{1/2}.$$

The inverse of the exponential map, which calculates the tangent vector pointing from P to Q is as follows:

$$\log_P(Q) = P^{1/2}\log(P^{-1/2}QP^{-1/2})P^{1/2}.$$

Now, we consider the minimization of the following optimization problem over $\mathcal{P}(n)$:

$$\min_{X\in\mathcal{P}(n)} f(X) = \frac{1}{2}X^TAX - b^TX,$$

where A is a symmetric positive definite matrix.

Let $n = 10000$ and b be a random vector. The tolerance, given by the norm of gradient of the function is set to be 10^{-6}. In Figure 2, we obtain the stochastic projected gradient method has an advantage when computational time is limited (in the early to middle stages). Furthermore, The problem is likely large-scale and ill-conditioned, leading to extremely large objective function values.

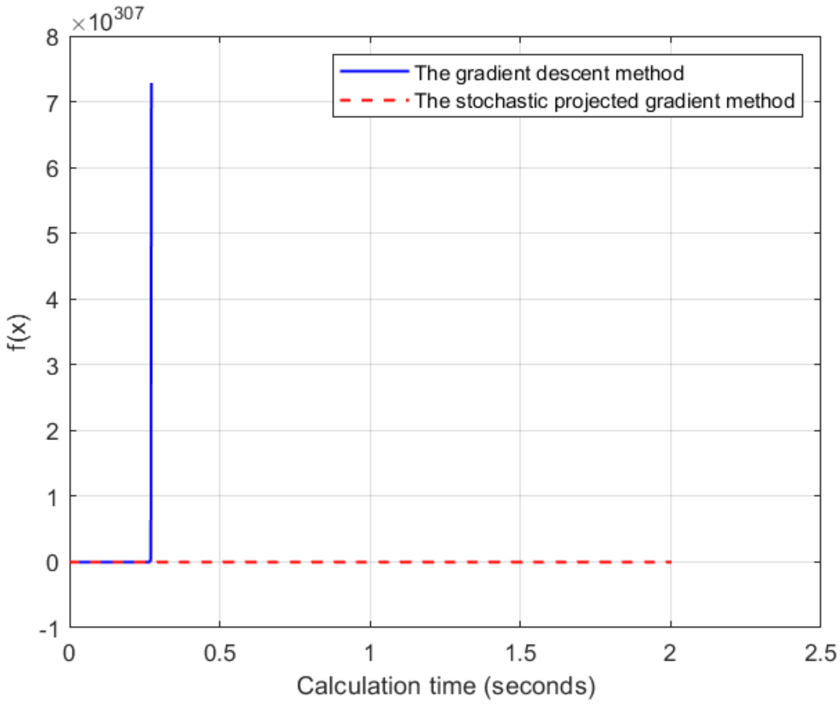

Figure 2. The comparison of convergence speed

5. Conclusions

In this context, the stochastic projected gradient method for optimization problems is introduced on Hadamard manifolds. Under some specific assumptions, the convergence result of the stochastic projected gradient method is established on Hadamard manifolds. Future research directions involve extending these ideas to general Riemannian manifold. Furthermore, it is interesting to analyze the algorithms performance for non-smooth objective functions on manifolds, potentially combining concepts from Riemannian optimization and non-smooth analysis.

Statements and Declarations

The authors declare that they have no conflict of interest, and the manuscript has no associated data.

References

[1] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.
[2] M. Bačák. *Convex Analysis and Optimization on Hadamard Spaces*. De Gruyter, Berlin, 2014.
[3] A. Beck. *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, 2017.
[4] G. C. Bento, O. P. Ferreira, and P. R. Oliveira. Unconstrained steepest descent method for multicriteria optimization on riemannian manifolds. *Journal of Optimization Theory and Applications*, 154(1):88–107, 2012.
[5] G. C. Bento, O. P. Ferreira, and P. R. Oliveira. Proximal point method for a special class of nonconvex functions on hadamard manifolds. *Optimization*, 64(2):289–319, 2015.
[6] D. P. Bertsekas. Subgradient, and proximal methods for convex optimization: A survey. ArXiv preprint arXiv:1507.01030, 2017.
[7] S. Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.
[8] N. Boumal, B. Mishra, and P. A. Absil. Manopt, a matlab toolbox for optimization on manifolds version 4.0. 2016. https://www.manopt.org.

[9] N. Boumal, B. Mishra, P. A. Absil, and R. Sepulchre. Manopt, a matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15(42):1455–1459, 2014.
[10] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33. American Mathematical Society Providence, 2001.
[11] I. Chavel. *Riemannian Geometry-A Modern Introduction.* Cambridge University Press, London, 1994.
[12] J. X. da Cruz Neto and P. R. Oliveira. Geodesic methods in riemannian manifolds. Research report, Universidade Federal do Rio de Janeiro, PEXC/COPPE, 1995.
[13] S. L. Feng, W. Huang, L. L. Song, S. Ying, and T. Zeng. Proximal gradient method for nonconvex and nonsmooth optimization on hadamard manifolds. *Optimization Letters*, 16(8):2277–2297, 2022.
[14] W. Huang. *Optimization Algorithms on Riemannian Manifolds with Applications.* PhD thesis, Florida State University, 2013.
[15] W. Huang, K. A. Gallivan, and P.-A. Absil. A broyden class of quasi-newton methods for riemannian optimization. *SIAM Journal on Optimization*, 25(3):1660–1685, 2015.
[16] C. L. Ji, Y. H. Fu, and P. He. Adaptive Riemannian stochastic gradient descent and reparameterization for Gaussian mixture model fitting. In B. Yanıkoğlu and W. Buntine, editors, *Proceedings of the 15th Asian Conference on Machine Learning*, volume 222, pages 534–549. PMLR, 2024.
[17] W. Klingenberg. *A Course in Differential Geometry.* Springer, Berlin, 1978.
[18] X. B. Li, X. C. Ge, and K. Tu. The generalized conditional gradient method for composite multiobjective optimization problems on riemannian manifolds. *Journal of Nonlinear and Variational Analysis*, 7(5):839–857, 2023.
[19] D. Martínez-Rubio, C. Roux, C. Criscitiello, and S. Pokutta. Accelerated methods for riemannian min-max optimization ensuring bounded geometric penalties. In Y. Li, S. Mandt, S. Agrawal, and E. Khan, editors, *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, pages 280–288. PMLR, 2025.
[20] I. Necoara and A. Patrascu. A random coordinate descent algorithm for optimization problems with composite objective function and linear constraints. *Computational Optimization and Applications*, 57(2):307–337, 2014.
[21] I. Necoara and A. Patrascu. A random projection method for convex optimization with linear constraints. *Mathematics of Operations Research*, 43(4):1057–1095, 2018.
[22] W. Ring and B. Wirth. Optimization methods on riemannian manifolds and their application to shape space. *SIAM Journal on Optimization*, 22(2):596–627, 2012.
[23] T. Sakai. *Riemannian Geometry.* American Mathematical Society, Providence, RI, 1996.
[24] H. Y. Zhang and S. Sra. First-order methods for geodesically convex optimization. In V. Feldman, A. Rakhlin, and O. Shamir, editors, *29th Annual Conference on Learning Theory*, pages 1617–1638, Columbia University, New York, USA, 2016. PMLR.