Subgradient Algorithm on Riemannian Manifolds

O. P. Ferreira and P. R. Oliveira

Communicated by F. Giannessi

Abstract. The subgradient method is generalized to the context of Riemannian manifolds. The motivation can be seen in non-Euclidean metrics that occur in interior-point methods. In that frame, the natural curves for local steps are the geodesics relative to the specific Riemannian manifold. In this paper, the influence of the sectional curvature of the manifold on the convergence of the method is discussed, as well as the proof of convergence if the sectional curvature is nonnegative.

Key Words. Nondifferentiable optimization, convex programming, subgradient methods, Riemannian manifolds.

1. Introduction

Tools from Riemannian geometry have been used in mathematical programming to obtain both theoretical results and practical algorithms (see Refs. 1–7). Recent studies by Helmke and Moore (Ref. 8) and Udriste (Ref. 9) provide some examples and a vast bibliography.

The subgradient method is one of the classical algorithms of nondifferentiable optimization, discovered by Shor (Ref. 10) in the early sixties, and is always an object of study; see Refs. 11–13 and their references. This motivates us to study the matter in the context of Riemannian geometry.

The subgradient algorithm is proposed to solve the problem with constraints

$$\min_{x \in M} f(x),$$

1This research was partially supported by CAPES/MEC and CNPq, Brazil.
2Professor, Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia, Goiás, Brazil.
3Professor, Programa de Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
where $M$ is a (connected) complete manifold and $f$ a real-valued convex function defined on $M$. Several definitions and results of Riemannian geometry will be necessary for the construction of the algorithm, which will allow us to see it as an unconstrained method; basic geometry references are for example Refs. 14–17.

The proof of convergence given here is inspired by the works of Burachik et al. (Ref. 18) and Corrêa and Lemaréchal (Ref. 11). It is also important to note, as in $M = \mathbb{R}^n$, the fundamental importance of the inequality

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + t_k^2 + 2[t_k/\|s_k\|][f(y) - f(x_k)],$$

first obtained by Kiwiel (Ref. 19), and generalized by da Cruz Neto and Oliveira (Ref. 20) for manifolds with nonnegative sectional curvature.

2. Basic Concepts

Let $M$ be a (connected) complete Riemannian manifold. Let $T_xM$ be the tangent space to $M$ at $x$. The exponential map $\exp_x$ is defined on $T_xM$ by

$$v \mapsto \exp_x v := \gamma(1, x, v),$$

where $\gamma$ is the geodesics of $M$ such that

$$\gamma(0, x, v) = x \quad \text{and} \quad \gamma'(0, x, v) = v.$$

When the reference to the point $x$ is not necessary or is implicit, the notation $\gamma_v$ means that

$$\gamma_v(0) = v.$$

If $\|v\| = 1$, the geodesic $\gamma_v$ is said to be parameterized by arc length or normalized.

If $\exp_x : V \to U$ where $V \subset T_xM$ and $U \subset M$, is a diffeomorphism, then $U$ is called a normal neighborhood of $x$. If

$$B_\epsilon(0) := \{v \in T_xM/\|v\| < \epsilon\}$$

is such that $B_\epsilon(0) \subset V$, we call $\exp_x B_\epsilon(0) := B_\epsilon(x)$ the normal ball or geodesics with center $x$ and radius $\epsilon$.

Given $x$ and $x' \in M$, the distance from $x$ to $x'$ is

$$d(x, x') := \inf_{\alpha_{x'x}} l(\alpha_{x'x}),$$

(1)
where \( a_{xx} : [a, b] \rightarrow M \) is a piecewise smooth curve joining \( x \) to \( x' \), that is,

\[
a_{xx}(a) = x, \quad a_{xx}(b) = x',
\]

and

\[
l(a_{xx}) := \int_a^b \| a'_{xx}(t) \| \, dt
\]

is the length of \( a_{xx} \).

**Theorem 2.1** (Hopf and Rinow). Let \( M \) be a (connected) Riemannian manifold. The following statements are equivalent:

(a) for every \( x \in M \), \( \exp_x \) is defined on all \( TXM \), that is, \( M \) is complete;

(b) \( (M, d) \) is complete as a metric space, where \( d \) is defined in (1); namely, any Cauchy sequence of \( M \) is a convergent sequence.

Furthermore, each of the above statements implies that:

(c) any two points \( x, x' \in M \) can be joined by a geodesics of length \( l(\gamma) = d(x, x') \); the geodesics with this property is called minimal.

**Proof.** This can be found in Ref. 15.

**Definition 2.1.** A real-valued function \( f \) defined on a complete Riemannian manifold \( M \) is said to be a convex function if \( f \) is convex when restricted to any geodesics of \( M \), which means that

\[
(f \circ \gamma)(ta + (1-t)b) \leq tf(\gamma(a)) + (1-t)f(\gamma(b))
\]

holds for any \( a, b \in \mathbb{R} \) and \( 0 \leq t \leq 1 \).

Some properties related to convex functions on Riemannian manifolds can be found in Refs. 9, 16, 17.

**Definition 2.2.** A real-valued function \( f \) defined on a complete Riemannian manifold \( M \) is said to be Lipschitzian if there exists a constant \( L(M) = L > 0 \) such that

\[
|f(x) - f(x')| \leq Ld(x, x'), \tag{2}
\]

for all \( x, x' \in M \). Besides this global concept, if it is established that, for all \( x_0 \in M \), there exists \( L(x_0) \geq 0 \) and \( \delta = \delta(x_0) > 0 \) such that Inequality (2) occurs, with \( L = L(x_0) \), for all \( x \) and \( x' \in B_\delta(x_0) := \{x \in M | d(x_0, x) < \delta\} \), then \( f \) is called locally Lipschitzian.
Theorem 2.2. Let $f$ be a convex function. Then, $f$ is locally Lipschitzian.

Proof. See Refs. 16 and 9.

Definition 2.3. Let $f$ be a convex function and $x \in M$. A vector $s \in T_x M$ is said to be a subgradient of $f$ at $x$ if, for any geodesics $\gamma$ of $M$ with $\gamma(0) = x$, $$(f \circ \gamma)(t) \geq f(x) + t \langle s, \gamma'(0) \rangle,$$ for any $t \geq 0$. The set of all the subgradients of $f$ at $x$, denoted by $\partial f(x)$, is called the subdifferential of $f$ at $x$.

Theorem 2.3. Let $f$ be a convex function. Then, for every $x \in M$, $\partial f(x)$ is nonempty, convex, and compact.

Proof. See Refs. 16 and 9.

Lemma 2.1 (Gauss). Let $x \in M$, $v \in T_x M$ such that $\exp_x v$ is defined and $u \in T_x M \approx T_0(T_x M)$. Then, $$\langle (d \exp_x)_v (u), (d \exp_x)_v (u) \rangle = \langle v, u \rangle.$$ 

Proof. See Refs. 14 and 15.

Theorem 2.4 (Toponogov). Let $M$ be a complete Riemannian manifold with sectional curvature $K \geq H$. Let $\gamma_1$ and $\gamma_2$ be segments of normalized geodesics in $M$ with $\gamma_1(0) = \gamma_2(0)$. Let us indicate by $M^2(H)$ a manifold of dimension 2 with constant curvature $H$. Admit that the geodesics $\gamma_1$ is minimal and that, if $H > 0$, $l(\gamma_2) \leq \pi / \sqrt{H}$. Consider two normalized geodesics $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ in $M^2(H)$, such that $\tilde{\gamma}_1(0) = \tilde{\gamma}_2(0)$, $l(\gamma_i) = l(\tilde{\gamma}_i) = l_i$, $i = 1, 2,$ and $$(\tilde{\gamma}_1(0), \tilde{\gamma}_2(0)) = (\gamma_1'(0) \gamma_2'(0)).$$ Then, $$d(\gamma_1(l_i), \gamma_2(l_2)) \leq d(\tilde{\gamma}_1(l_i), \tilde{\gamma}_2(l_2)).$$ 

Proof. See Refs. 14 and 15.

Corollary 2.1. Let $M$ be a complete Riemannian manifold with sectional curvature $K \geq 0$. If $\gamma_{v_1}$ and $\gamma_{v_2}$ are normalized geodesics such that $\gamma_{v_1}(0) = \gamma_{v_2}(0)$, then $$d(\gamma_{v_1}(t_1), \gamma_{v_2}(t_2)) \leq \| t_2 v_2 - t_1 v_1 \|.$$
Proof. With the notation of Theorem 2.4, $M^2(H)$ is the subspace generated by the vectors $v_1$ and $v_2$, $H=0$, $\gamma_1 = \gamma_{v_1}$, $\gamma_2 = \gamma_{v_2}$, $\bar{\gamma}_1(t) = tv_1$, and $\bar{\gamma}_2(t) = tv_2$. We observe too that, in this case, we have no hypothesis on $\gamma_2 = \gamma_{v_2}$. When this identification is made, the proof is immediate from Theorem 2.4.

3. Definition of the Problem and the Algorithm

Let $M$ be a (connected) complete Riemannian manifold, and let $f$ be a convex function. The set $\mathcal{C}^*$ denotes the set of minimizers of $f$ and

$$f^* = \inf_{x \in M} f(x)$$

denotes its infimal value. The problem is to estimate $f^*$ and also to find a point of $\mathcal{C}^*$, if such point exist, that is, $\mathcal{C}^* \neq \emptyset$.

Algorithm 3.1. Conceptual Subgradient Algorithm. The sequence $\{t_k\}$ is given, with $t_k > 0$ for $k = 1, 2, \ldots$

Step 0. Initialize. Choose $x_1 \in M$ and obtain $s_1 \in \partial f(x_1)$. Make $k = 1$.

Step 1. If $s_k = 0$, stop. Otherwise, calculate the geodesics $\gamma_{s_k}$ with $\gamma_{s_k}(0) = x_k$, $\gamma_{s_k}'(0) = v_k$, $v_k = -s_k/\|s_k\|$.

Step 2. Make $x_{k+1} = \gamma_{s_k}(t_k)$.

Step 3. Obtain $s_{k+1} \in \partial f(x_{k+1})$. Make $k = k + 1$, and go to Step 1.

4. Preliminary Results

As is known, the sequence obtained by Algorithm 3.1 does not decrease the function. So a $\{t_k\}$ sequence should be chosen such that the respective $\{x_k\}$ sequence approaches the set $\mathcal{C}^*$, as is usual in the case where $M = \mathbb{R}^n$.

As in Ref. 13, for $\mathbb{R}^n$, we shall obtain some preliminary properties, in particular an upper estimate for the $t_k$'s.

Theorem 4.1. Let $x^* \in \mathcal{C}^* \neq \emptyset$, and let $B_\delta(x^*)$ be a normal ball. If $x_k \in B_\delta(x^*)$, $x_k \in B_\delta(x^*)$, then $\delta_k > 0$ exists such that, by choosing $0 < t_k < \delta_k$ in Algorithm 3.1,

$$d(x_{k+1}, x^*) < d(x_k, x^*).$$

Proof. Let $\gamma_\delta$ be the geodesic parameterized by the arc length, given by Theorem 2.1 (c), such that $\gamma_\delta(0) = x_k$ and $\gamma_\delta(t^*) = x^*$, with
Consider the normal sphere $S^*(x^*)$, boundary of the normal ball $B^*(x^*)$, which by the Gauss lemma is a submanifold of codimension 1 with

$$T_{x_k}S = \{ u \in T_{x_k}M : \langle u, v \rangle = 0 \}.$$

Let $s_k \in \partial f(x_k)$. By definition,

$$f \circ \gamma_s(t) \geq f(x_k) + t \langle s_k, v \rangle,$$

and at $t = t^*$,

$$f(x_a) - f(x_k) \geq t^* \langle s_k, v \rangle.$$

As $x_k \neq 0^*$, it follows that $f(x_k) < f(x_a)$, which implies $\langle s_k, v \rangle < 0$. So, $\delta_k > 0$ exists such that $\gamma_s(t) \in B^*(x_a)$ for all $0 < t < \delta_k$, where $v_k = - s_k / \| s_k \|$. Therefore, for all $0 < t_k < \delta_k$,

$$d(\gamma_s(t_k), x^*) < d(x_k, x^*),$$

and this proves the result. 

It is a consequence of the Hadamard theorem (Refs. 14 and 15) that, if $M$ is complete simply connected and has sectional curvature $K \leq 0$, then $\exp_\cdot : T_xM \rightarrow M$ is a global diffeomorphism and this implies that we can take $B_{\delta}(x^*) = M$. But Theorem 4.1 does not give an estimate for $\delta_k$ in terms of $x_k$ and $x^*$, as happens when $M = \mathbb{R}^n$. Nonetheless, this estimate can be obtained if $K \geq 0$, as follows.

Intuitively, an idea can already be had of the influence of the sectional curvature $K$ on the behavior of the sequence $\{x_k\}$ defined by Algorithm 3.1. First observe that, if $K > 0$, the geodesics tend to approximate one another, and the contrary occurs if $K < 0$. This suggests that we can go further along the geodesics without distancing ourselves from $x^*$, on a nonnegative curvature than on a negative one.

**Lemma 4.1.** See Ref. 20. Let $\{x_k\}$ be the sequence generated by Algorithm 3.1. If $M$ has sectional curvature $K \geq 0$, then for all $y \in M$,

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + t^2_k + 2t_k \| s_k \| \| f(y) - f(x_k) \|,$$

for all $k \in \mathbb{N}$.

**Proof.** Let $\gamma_{v_1}$ be the minimizing geodesics, that is, for $v_1$ such that $\| v_1 \| = 1$, we have $\gamma_{v_1}(0) = x_k$, $\gamma_{v_1}(t_1) = y$, with $t_1 = d(x_k, y)$. Also, let $\gamma_{v_2}$ be the geodesics such that $v_2 = v_k = - s_k / \| s_k \|$, $\gamma_{v_2}(0) = x_k$ and $\gamma_{v_2}(t_k) = x_{k+1}$,
with \( t_k = t_2 \). Then, it follows from Corollary 2.1 and Definition 2.1 that
\[
\begin{align*}
    d^2(x_{k+1}, y) & \leq \| -t_k[s_k/\|s_k\|] - t_1v_1 \|^2 \\
    & = t_1^2 + t_2^2 + 2[t_k/\|s_k\|] \langle s_k, t_1v_1 \rangle \\
    & \leq d^2(x_k, y) + t_1^2 + 2[t_k/\|s_k\|][f(y) - f(x_k)].
\end{align*}
\]

Consider the set
\[ \mathcal{O} = \{ z \in \mathcal{M} / f(z) \leq \inf_{x_k} f(x_k) \}. \]

**Corollary 4.1.** Let \( \{ x_k \} \) be the sequence generated by Algorithm 3.1, and let \( K \geq 0 \) be the sectional curvature of \( \mathcal{M} \). For all \( z \in \mathcal{O} \), we have
\[
    d^2(x_{k+1}, z) \leq d^2(x_k, z) + t_k^2,
\]
for all \( k \in \mathbb{N} \).

**Proof.** This follows immediately from the definition of \( \mathcal{O} \) and by Lemma 4.1. \( \Box \)

**Theorem 4.2.** Let \( x_* \in \mathcal{O}^* \neq \emptyset \), and let \( \{ x_k \} \) be the sequence generated by Algorithm 3.1. If \( \mathcal{M} \) has a curvature \( K \geq 0 \) and \( x_k \notin \mathcal{O}^* \), then
\[
    d(x_{k+1}, x_*) < d(x_k, x_*),
\]
for all
\[
    0 < t_k < [2/\|s_k\|][f(x_k) - f(x_*)]. \tag{3}
\]

**Proof.** In Lemma 4.1, take \( y = x_* \). Then,
\[
    d^2(x_{k+1}, x_*) \leq d^2(x_k, x_*) + t_k^2 + 2[t_k/\|s_k\|][f(x_*) - f(x_k)].
\]
As \( x_* \neq x_k \), it follows that, for all
\[
    0 < t_k < [2/\|s_k\|][f(x_k) - f(x_*)],
\]
we have
\[
    t_k^2 + 2[t_k/\|s_k\|][f(x_*) - f(x_k)] < 0,
\]
and this concludes the proof. \( \Box \)

**Definition 4.1.** A sequence \( \{ y_k \} \) in the complete metric space \((\mathcal{M}, d)\) is Fejér convergent to a set \( \mathcal{W} \subset \mathcal{M} \) if, for every \( w \in \mathcal{W} \), there exists a sequence
Theorem 4.3. Let \{y_k\} be a sequence in the complete metric space \((M, d)\). If \{y_k\} is quasi-Fejer convergent to a nonempty set \(W \subset M\), then \{y_k\} is bounded. If furthermore, a cluster point \(y\) of \{y_k\} belongs to \(W\), then \(\lim_{k \to \infty} y_k = y\).

Proof. This is the same proof as in Ref. 18, replacing \(\| \cdot \|\) by \(d\).

5. Convergence of the Algorithm

Theorem 5.1. Let \(\{x_k\}\) be the sequence generated by Algorithm 3.1, and let \(K \geq 0\) be the sectional curvature of \(M\). If the sequence \(\{t_k\}\) is chosen to satisfy
\[
\sum_{k=0}^{\infty} t_k = +\infty, \quad \sum_{k=0}^{\infty} t_k^2 < +\infty,
\]
then
\[
\liminf_{k \to \infty} f(x_k) = f^*.
\]
In addition, if \(\emptyset^* \neq \emptyset\), then the sequence \(\{x_k\}\) converges to a point \(x^* \in \emptyset^*\).

Proof. By contradiction, suppose that
\[
\liminf_{k \to \infty} f(x_k) < f^*.
\]
On the one hand, this implies that \(\emptyset \neq \emptyset\). Besides, from (5) and Corollary 4.1, \(\{x_k\}\) is quasi-Fejer convergent to \(\partial f\), where \(\epsilon_k = t_k\). Therefore, \(\{x_k\}\) is bounded and consequently \(\{s_k\}\), where \(s_k \in \partial f(x_k)\), is also bounded. Let us say that \(\|s_k\| < C_0\), for all \(k \in \mathbb{N}\), where \(C_0 > 0\). On the other hand, given \(z \in \partial f\), there exist \(C_1 > 0\) and \(k_0 \in \mathbb{N}\) such that
\[
f(z) < f(x_k) - C_1, \quad \text{for all } k > k_0.
\]
For this \(z\), it follows from Lemma 4.1 that
\[
d^2(x_{k+1}, z) \leq d^2(x_k, z) + t_k(t_k - 2C_1/C_0).
\]
From hypothesis (5), $t_k \to 0$, and so we may suppose that $k_0$ is such that $t_k < (C_1/C_0)$, for all $k > k_0$. Substituting this into (6), we have

$$d^2(x_{k+1}, z) \leq d^2(x_k, z) - (C_1/C_0)t_k,$$

for all $k > k_0$. Adding (7), we have

$$\sum_{j=k_0}^{k} t_j \leq (C_0/C_1)(d^2(x_{k_0}, z) - d^2(x_{j+k_0}, z)),$$

for all $l$. This contradicts (4), as $d(x_{l+k_0}, z)$ is bounded. It follows from the first part that $\{f(x_k)\}$ possesses a decreasing monotonous subsequence $\{f(x_{k_j})\}$ such that

$$\lim_{j \to \infty} f(x_{k_j}) = f^*.$$

Without loss of generality we shall suppose that the sequence $\{f(x_k)\}$ is decreasing, monotonous, and converges to $f^*$. Being bounded, the sequence $\{x_k\}$ possesses a convergent subsequence $\{x_{k_j}\}$. Let us say that

$$\lim_{j \to \infty} x_{k_j} = x^*,$$

which by the continuity of $f$ implies

$$f(x^*) = \lim_{j \to \infty} f(x_{k_j}) = f^*,$$

and so $x^* \in \mathcal{E}$. Thus, $\{x_k\}$ has an cluster point $x^* \in \mathcal{E}$; as $\{x_k\}$ is quasi-Fejér convergent to $\mathcal{E}$, it follows from Theorem 4.3 that the sequence $\{x_k\}$ converges to $x^*$.

There follow two examples of complete Riemannian manifolds where the geodesics has an explicit formula. For more details on these examples, see Refs. 4, 7, and 20.

**Example 5.1.** The set

$$M = \text{int}(\mathbb{R}^*_+ \cap \{x \in \mathbb{R}^n / x > 0\}),$$

with the affine-scaling metric $g = (g_{ij})$, where $g_{ij}(x) = \delta_{ij}/x_i x_j$ is a complete Riemannian manifold, with $K \equiv 0$ and tangent plane on point $x \in M$ equal to $T_x M = \mathbb{R}^n$. The geodesic $\gamma: \mathbb{R} \to M$ of $M$ with initial conditions

$$\gamma(0) = a \in M \quad \text{and} \quad \gamma'(0) = b \in T_a M,$$
is

\[ t \mapsto \gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t)), \]

with

\[ \gamma_i = a_i \exp((a_i/b_i) t), \quad i = 1, \ldots, n. \]

Thus, the problem \( \min f(x) \), such that \( x \geq 0 \), can be solved by Algorithm 3.1, by using this last expression.

**Example 5.2.** The set

\[ M = \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1, x_i > 0 \right\}, \]

with the projective metric

\[ g = (g_{ij}), \quad \text{where } g_{ij}(x) = (1/x_i x_j) [\delta_{ij} - 1/n], \]

is a complete Riemannian manifold, with \( K \equiv 0 \) and tangent plane at point \( x \in M \) given by

\[ T_x M = \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 0 \right\}. \]

The geodesic \( \gamma : \mathbb{R} \to M \) of \( M \) with initial conditions equal to

\[ \gamma(0) = a \in M \quad \text{and} \quad \gamma'(0) = b \in T_a M \]

is

\[ t \mapsto \gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t)), \]

with

\[ \gamma_i(t) = a_i \exp((b_i/a_i - b_i/a_i) t) \left( a_i + \sum_{k=2}^n a_k \exp((b_k/a_k - b_k/a_k) t) \right). \]

Therefore, the problem

\[ \min f(x), \text{ such that } \sum_{i=1}^n x_i = 1, x \geq 0, \]

can be solved by Algorithm 3.1, by using this last expression.
6. Conclusions

In Section 4, one sees the need of the hypothesis on the sectional curvature of the manifold, since the role of Corollary 2.1 is central in the proof presented, which is only valid in the case of positive curvature. As geodesics, in the presence of positive curvature, tend to grow closer, it is intuitive to believe that it is possible to improve the estimate (3) in terms of the curvature. Proof of the convergence without a hypothesis about the curvature also remains open. Finally, we observe that Algorithm 3.1 solves the constrained problem

\[
\min_{x \in M} f(x),
\]

where \( M \) is a connected, complete Riemannian manifold of nonnegative sectional curvature, and the geodesics of \( M \) are available or easily approximated.

References


