Riemannian epsilon steepest descent method with Armijo line search

M. Mezghiche, N. Djeghaba, Z. Souici-Benhammadi

Abstract. In this paper we propose the epsilon steepest descent method to solve optimization problems in Riemannian manifolds. This algorithm accelerates the convergence of steepest descent method on Riemannian submanifolds of $\mathbb{R}^n$. We establish the convergence of our new algorithm with Armijo line search and we present a numerical experiment to minimize the Rayleigh quotient on the unit sphere.

M.S.C. 2010: 65M10, 53C21, 90C30.
Key words: Optimization on Riemannian manifolds; epsilon steepest descent method; Armijo inexact line search; geodesics.

1 Introduction

Consider the following minimization problem

$$\text{(P)} \qquad \min \{f(x), x \in M\},$$

where M is a Riemannian manifold, and f is a continuously differentiable function on M.

This optimization problem is solved by several manners. The oldest method of resolution is that of Cauchy called the method of the gradient or steepest descent. Thereafter, several developments were brought to this method in the Euclidian case and the Riemannian case.

In the Riemannian case, several geometrical objects are necessary for the resolution of the problem (P) such as the retractions and the research of the directions along geodesics [4]. The method generates a sequence of points x_k given by

$$x_{k+1} = R_{x_k}(\lambda_k v_k),$$

where $v_k \in T_{x_k}M$ is the direction of research, $T_{x_k}M$ is the tangent space at the point x_k on M, λ_k is the step of line search, and R is map of TM in M called retraction.

In Euclidian case, Benzine and Djeghaba [3] improved the algorithm of steepest descent by a new algorithm called epsilon-steepest descent algorithm. This last is

BSG Proceedings, vol. 25, 2018, pp. 45-58.

based on epsilon algorithm [2] and consists to calculate quantities with two indices $\varepsilon_j^{(k)}$ as follows

$$\varepsilon_{-1}^{(k)} = 0 \qquad \varepsilon_0^{(k)} = x_k \qquad k = 0, 1, ...$$
$$\varepsilon_{j+1}^{(k)} = \varepsilon_{j-1}^{(k+1)} + \frac{1}{\varepsilon_j^{(k+1)} - \varepsilon_j^{(k)}} \qquad j, k = 0, 1, ...$$

These quantities are organized in a table with two indexes (ε-table) such as the lower index denotes a column while the upper index denotes a descent diagonal.

$$\begin{array}{llllllllllll}
\varepsilon_{-1}^{(0)} = 0 & & & & & & & & & & & \\
 & \varepsilon_0^{(0)} = x_0 & & & & & & & & & & \\
\varepsilon_{-1}^{(1)} = 0 & & \varepsilon_1^{(0)} & & & & & & & & & \\
 & \varepsilon_0^{(1)} = x_1 & & \varepsilon_2^{(0)} & & & & & & & & \\
\varepsilon_{-1}^{(2)} = 0 & & \varepsilon_1^{(1)} & & \varepsilon_3^{(0)} & & & & & & & \\
 & \varepsilon_0^{(2)} = x_2 & & \varepsilon_2^{(1)} & & \varepsilon_4^{(0)} & & & & & & \\
\varepsilon_{-1}^{(3)} = 0 & & \varepsilon_1^{(2)} & & \varepsilon_3^{(1)} & & \varepsilon_5^{(0)} & & & & & \\
 & \varepsilon_0^{(3)} = x_3 & & \varepsilon_2^{(2)} & & \varepsilon_4^{(1)} & & \varepsilon_6^{(0)} & & & & \\
\varepsilon_{-1}^{(4)} = 0 & & \varepsilon_1^{(3)} & & \varepsilon_3^{(2)} & & \varepsilon_5^{(1)} & & \varepsilon_7^{(0)} & & & \\
 & \varepsilon_0^{(4)} = x_4 & & \varepsilon_2^{(3)} & & \varepsilon_4^{(2)} & & \varepsilon_6^{(1)} & & \varepsilon_8^{(0)} & & \\
. & & . & & . & & . & & . & & . & \\
 & . & & . & & . & & . & & . & & .
\end{array}$$

The table ε connects the quantities located at the four tops of a rhombus

$$\begin{array}{ccc}
 & \varepsilon_j^{(k)} & \\
\varepsilon_{j-1}^{(k+1)} \begin{array}{c}\nearrow \\ \searrow\end{array} & & \begin{array}{c}\searrow \\ \nearrow\end{array} \varepsilon_{j+1}^{(k)} \\
 & \varepsilon_j^{(k+1)} &
\end{array}$$

i.e. to calculate $\varepsilon_{j+1}^{(k)}$ we need to know $\varepsilon_{j-1}^{(k)}, \varepsilon_j^{(k+1)}$, and $\varepsilon_j^{(k)}$.

Now for the epsilon steepest descent method we use the algorithm of order two (column $\varepsilon_2^{(k)}$). Given a sequence $\{x_k\}_{k\in\mathbb{N}}$ the quantities $\varepsilon_2^{(k)}$ are obtained by

$$\varepsilon_2^{(k)} = x_{k+1} + \left[\frac{1}{x_{k+2} - x_{k+1}} - \frac{1}{x_{k+1} - x_k}\right]^{-1}.$$

In our work we shall generalize the epsilon steepest descent method with Armijo line search form Euclidean space to Riemannian submanifold of $\mathbb{R}^n$, we denoted R-epsilon steepest descent method. We show its convergence and we give an application with Rayleigh quotient on the unit sphere. This article is organized as follows: In Section 2 we give some preliminaries and geometrical concepts, in Section 3 we describe the R-steepest descent method [6], in Section 4 we present the R-epsilon steepest descent method with Armijo line search, and in the last section we show numerical tests on the unit sphere.

2 Preliminaries

Definition 2.1. Let M be a differentiable manifold, a Riemannian structure on M is a tensor field g of type (0,2) which for all $p \in M$ and $X, Y \in \chi(M)$ ($\chi(M)$ the set of C^∞ vectors fields on M) satisfies

(i) $g(X,Y) = g(Y,X)$,

(ii) $g_p : T_pM \times T_pM \to \mathbb{R}$ is positive definite.

Definition 2.2. A Riemannian manifold is a connected differentiable manifold with a Riemannian structure. For every $p \in M$, the Riemannian structure g provides an inner product on T_pM given by the nondegenerate symmetric bilinear form $g_p : T_pM \times T_pM \to \mathbb{R}$. We note

$$\langle X, Y\rangle = g_p(X,Y) \textit{ and } \|X\| = g_p(X,Y)^{1/2}, X, Y \in T_pM.$$

Let $t \mapsto \gamma(t), t \in [t_0, t_1]$, be a curve segment in M. The length of γ is defined by

$$L(\gamma) = \int_{t_0}^{t_1} g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))^{1/2} dt.$$

Because M is connected, any two points p and q in M can be joined by a curve. The infimum (minimum) of the length of all curve segments joining p and q yields a metric on M called the Riemannian metric and denoted by $d(p,q)$.

Definition 2.3. Let M be a Riemannian manifold with Riemannian structure g and $f : M \to \mathbb{R}$ a C^∞ function on M. The gradient of f at p, denoted $(\operatorname{grad} f)_p$, is the unique vector in T_pM such that

$$df_p(X) = \langle (\operatorname{grad} f)_p, X\rangle, \forall X \in T_pM.$$

Let M be an n-dimensional Riemannian manifold with Riemannian structure g, and $(U, x^1, ..., x^n)$ a coordinate chart on M. There exist n^2 functions g_{ij}, $1 \le i, j \le n$ on U such that

$$g = \sum_{i,j} g_{ij} dx^i \otimes dx^j.$$

Clearly $g_{ij} = g_{ji}$ for all i and j. Because g_p is nondegenerate for all $p \in U \subset M$, the symmetric matrix (g_{ij}) is invertible. The elements of its inverse are denoted by g^{kl}, i.e. $\sum_l^{il} g_{lj} = \delta^i_j$, where δ^i_j is the kronecker delta. Furthermore, given $f \in C^\infty(M)$, we have

$$df = \sum_i \left(\frac{\partial f}{\partial x^i}\right) dx^i.$$

Therefore, from the definition of $\operatorname{grad} f$, we see that

$$\operatorname{grad} f = \sum_{il} g^{il} \left(\frac{\partial f}{\partial x^l}\right) \frac{\partial}{\partial x^i}.$$

Definition 2.4. Let M be a differentiable manifold. An affine connection on M is a function $\nabla : \chi(M) \times \chi(M) \to \chi(M)$ such that for all $X, Y, V, W \in \chi(M), a, b \in \mathbb{R}$,and $f, g \in C^\infty(M)$ satisfies

1. $\nabla_V(aX + bY) = a\nabla_V X + b\nabla_V Y$,
2. $\nabla_{fX+gY}W = f\nabla_X W + g\nabla_Y$,
3. $\nabla_V(fW) = (Vf)W + f\nabla_V W$.

Definition 2.5. Let M be a differentiable manifold with affine connection ∇. Let $\gamma : I \to M$ be a smooth curve with tangent vectors $X(t) = \gamma'(t)$ where $I \subset \mathbb{R}$ is an open interval. γ is called a geodesic if

$$\nabla_X X = 0, \forall t \in I.$$

Theorem 2.1. *or every p in M and $X \neq 0$ in T_pM, there exists a unique geodesic $t \mapsto \gamma_X(t)$ such that*

$$\gamma_X(0) = p \text{ and } \gamma'_X(0) = X.$$

We define the exponential map $\exp_p : T_pM \to M$ by

$$\exp_p(X) = \gamma_X(1),\ \forall X \in T_pM$$

In particular, we have

$$\exp_p(tX) = \gamma_{tX}(1) = \gamma_X(t),\ \forall X \in T_pM. \quad (2.1)$$

Definition 2.6. Let $R : TM \to M$ be a smooth map, and R_x its restriction to T_xM. R is called retraction on M if it satisfies the following properties
1) $R_x(0_x) = x$, where 0_x denotes the zero element of T_xM.
2) With the canonical identification $T_{0_x}T_xM \simeq T_xM, R_x$ satisfies

$$DR_x(0_x) = id_{T_xM},$$

where $DR_x(0_x)$ denotes the derivative of R_x at 0_x, and id_{T_xM} the identity map on T_xM.

Definition 2.7. Let $f : M \to \mathbb{R}$ be a locally Lipschtiz function on a complete Riemannian manifold $M, v \in T_xM, g = -\frac{v}{\|v\|}$ is called a descent direction at x, if there exists $\alpha > 0$ such that

$$f(\exp_x(tg)) - f(x) \leq -t\|v\|, \forall t \in]0, \alpha[.$$

Algorithm of Armijo line search in Riemannian case[1]
Initialization step.
Given a function f on a Riemannian manifold M with retraction R, a point $x \in M$, and a tangent vector $v_k \in T_{x_k}M$, such that

$$\langle(\operatorname{grad} f)(x_k), v_k\rangle < 0,$$

and $\beta, \sigma \in]0, 1[$, $\bar{\lambda} > 0$.
Main Step. Find the smallest integer $i \geq 0$ such that

$$f(\exp_{x_k}(\beta^i\bar{\lambda}v_k)) - f(x_k) \leq \sigma\beta^i\bar{\lambda}\langle(\operatorname{grad} f)(x_k), v_k\rangle,$$

where $\lambda_k = \beta^i\bar{\lambda}$ is the Armijo step size.

3 Steepest descent method on Riemannian manifolds

Let $\Delta = \{x : (\text{grad}\, f)(x) = 0\}$ a set of desirable points on the Riemannian manifold M, and let $A(.) : M \to M$ an iteration map of the algorithm. Let $0 < \varepsilon \in \mathbb{R}$, and $\tilde{x} \in M, B(\tilde{x}, \varepsilon) := \{x \in M : d(x, \tilde{x}) < \varepsilon\}, \bar{B}(\tilde{x}, \varepsilon) := \{x \in M : d(x, \tilde{x}) \leq \varepsilon\}$ where $d(.,.)$ is the Riemannian distance defined on the Riemannian manifold M.

3.1 Description of the algorithm

Algorithm 1[7]

Step 1.(Initialization) choose an initial point $x_0 \in M$, put $k = 0$.
Step 2.(Principal) suppose that at the iteration k, we have x_k.
If $x_k \in \Delta$ stop.
Else, calculate $x_{k+1} = A(x_k) \in M$. Put $k = k + 1$ and go to step 2.

Remark 3.1. $x_{k+1} = A(x_k) = \exp_{x_k}(\lambda v_k)$, where $v_k \in T_{x_k}M$ is the direction of research and λ_k is the step of line search. If $v_k = -(\text{grad}\, f)(x_k)$ and λ_k is given by the Armijo line search, the algorithm 1 is called the gradient or steepest descent algorithm with Armijo line search.

3.2 Convergence

Theorem 3.1. *[7] Suppose that there exists a function $f : M \to \mathbb{R}$ such that*

(i) f is continuous on M;

(ii) $\forall x \notin \Delta, x \in M$ there exist $\varepsilon > 0, \delta > 0$ such that

$$f(y') - f(x') \leq -\delta, \forall x' \in B(x, \varepsilon), y' \in A(x') \subset M.$$

Then, if the sequence $\{x_k\}_{k\in\mathbb{N}}$ is generated by the algorithm 1, every accumulation point $\hat{x} \in M$ of $\{x_k\}_{k\in\mathbb{N}}$ is desirable i.e. $\hat{x} \in \Delta$.

4 R-epsilon steepest descent method on Riemannian submanifolds of $\mathbb{R}^n$ with Armijo line search

As in the Euclidian case, the steepest descent method in the Riemannian case becomes very slow when we approach the stationary point, therefore we will accelerate the convergence by the epsilon algorithm ([3],[5]).

4.1 Description of the algorithm

Algorithm 2 (R-epsilon steepest descent with Armijo line search)

Initialization. Choose an initial point $x_0 \in M, x_0 = (x_0^1, x_0^2, ..., x_0^i, ..., x_0^n)$.
Put $k = 0$ and go to principal step.
Principal step.

At the iteration k, we have $x_k = (x_k^1, x_k^2, ..., x_k^i, ..., x_k^n)$.

If $\|\text{grad}\, f(x_k)\| = 0$ stop, the minimum is the point x_k,

else put $r_k = x_k$;

calculate s_k, t_k successors of r_k using the steepest descent method with Armijo line search, such that

$$\begin{aligned} s_k &= R_{r_k}(-\lambda_k \,\text{grad}\, f(r_k)), \\ t_k &= R_{s_k}(-\beta_k \,\text{grad}\, f(s_k)), \end{aligned}$$

where λ_k and β_k are obtained by Armijo inexact line search.

If

$$s_k^i - r_k^i \neq 0, t_k^i - s_k^i \neq 0, \text{and } \frac{1}{t_k^i - s_k^i} - \frac{1}{s_k^i - r_k^i} \neq 0, \;\; i = 1, ..., n,$$

we put

$$\varepsilon_2^{k,i} = s_k^i + \left[\frac{1}{t_k^i - s_k^i} - \frac{1}{s_k^i - r_k^i} \right]^{-1}, \quad i = 1, ..., n$$

and

$$\varepsilon_2^{(k)} = \left(\varepsilon_2^{k,1}, \varepsilon_2^{k,2}, ..., \varepsilon_2^{k,i}, ..., \varepsilon_2^{k,n} \right).$$

If

$$\varepsilon_2^{(k)} \in M \quad \text{and} \quad f(\varepsilon_2^{(k)}) < f(t_k),$$

then put $x_k = \varepsilon_2^{(k)}$, $k = k + 1$ and go to principal step.

If

$$f(\varepsilon_2^{(k)}) \geq f(t_k) \text{ or } \varepsilon_2^{(k)} \notin M,$$

or

$$s_k^{i_0} - r_k^{i_0} = 0, \text{ or } t_k^{i_0} - s_k^{i_0} = 0, \text{or } \frac{1}{t_k^{i_0} - s_k^{i_0}} - \frac{1}{s_k^{i_0} - r_k^{i_0}} = 0,$$

$$i_0 \in \{1, ..., n\},$$

then put $x_k = t_k$, $k = k + 1$ and go to principal step.

4.2 Convergence

Let ϕ be a function defined by

$$\phi(\lambda) = f(\exp_x \lambda v_x) - f(x)$$

and let $\gamma_{v_x}(\lambda)$ be the geodesics in the direction v_x. We denote by

$$V_\lambda = \gamma'_{v_x}(\lambda)$$

the tangent vector field defined along $\gamma_{v_x}(\lambda)$ with $V_0 = v_x$.

Lemma 4.1. *[7] Let $\hat{x} \in M, \varepsilon > 0$ and $v_x \in D_x$ be a tangent vector at $x \in \bar{B}(\hat{x}, \varepsilon) \subset M$, ($D_x \subset T_x M$ with $\gamma(0) = x, \gamma'(0) = v$), $V_\lambda \in \chi(M)$ and $f \in C^\infty(M)$. Then*

$$\|v_x\| \geq c \,\|\text{grad}\, f(x)\|$$

implies that, for any λ, we have

$$\|V_\lambda\| \geq c\,\|\text{grad}\, f(x)\|\,. \tag{4.1}$$

As well,

$$\langle \text{grad}\, f(x), v_x\rangle \leq -\rho'\,\|\text{grad}\, f(x)\|\,\|v_x\|\,, \rho' > 0, \forall x \in B(\hat{x}, \varepsilon)$$

implies that, for any $\rho < \rho'$ fixed, there exists some $\bar{\lambda} > 0$ such that

$$\left\langle \frac{\partial f}{\partial x}(x), V_\lambda \right\rangle \leq -\rho\,\|\text{grad}\, f(x)\|\,(\|V_\lambda\|\,,\ \forall x \in \bar{B}(\hat{x}, \varepsilon), \forall \lambda \in [0, \bar{\lambda}]. \tag{4.2}$$

Theorem 4.2. *Let $f \in C^\infty(M), V_\lambda \in \chi(M)$ with $v_x \in D_x$ a tangent vector at $x \in M$ satisfying $\langle \text{grad}\, f(x), v_x\rangle < 0$, for $\text{grad}\, f(x) \neq 0$. Suppose that, for $\hat{x} \in M$ such that $\|\text{grad}\, f(\hat{x})\| \neq 0$, there exist $\varepsilon > 0, \rho' > 0$ and $c > 0$ such that $\forall x \in B(\hat{x}, \varepsilon)$*

$$\langle \text{grad}\, f(x), v_x\rangle \leq -\rho'\,\|\text{grad}\, f(x)\|\,\|v_x\|\,, \tag{4.3}$$

$$\|v_x\| \geq c\,\|\text{grad}\, f(x)\|\,. \tag{4.4}$$

Moreover, let $\{x_k\}_{k\in\mathbb{N}}$ be sequence generated by the algorithm 2. Then $x_k \to x^*$ implies that $\|\text{grad}\, f(x^*)\| = 0$.

Proof. According to R epsilon steepest descent algorithm, to find the successor x_{k+1} of x_k we need to use twice the steepest descent method with Armijo line search to calculate s_k and t_k i.e. we will check the theorem 3.1 for the both.

For the first one, f is continuous, thus it remain to show that (ii) of theorem 3.1 is satisfied, i.e $\forall x \in M, \text{grad}\, f(x) \neq 0, \exists\varepsilon > 0, m > 0$ such that

$$f(s_k) - f(x_k) \leq -m, \forall x_k \in B(\hat{x}, \varepsilon), s_k = \exp_{x_k}(\lambda v_{x_k}) \in M.$$

From (2.1)

$$\exp_{x_k}(\lambda v_x) = \gamma_{\lambda v_x}(1) = \gamma_{v_x}(\lambda), v_k \in T_{x_k}M$$

and we have

$$\frac{d}{d\lambda}(\exp_{x_k}(\lambda v_{x_k})) = \gamma'_{v_x}(\lambda) = V_\lambda.$$

Thus

$$\begin{aligned}
f(s_k) - f(x_k) &= f(\exp_{x_k}(\lambda v_{x_k})) - f(x_k) = \phi(\lambda) - \phi(0)\\
&= \int_0^1 \frac{d}{d(\lambda t)}(\phi(t\lambda))d(\lambda t)\\
&= \int_0^1 \frac{d}{d\lambda t}(f(\exp_{x_k}(\lambda t v_{x_k})) - f(x_k))\lambda dt\\
&= \int_0^1 \left[\frac{\partial}{\partial x} f(\exp_{x_k}(\lambda t v_{x_k}))\underbrace{\frac{d}{d\lambda t}(\exp_{x_k}(\lambda t v_{x_k}))}_{V_{\lambda t}} - \underbrace{\frac{d}{d\lambda t} f(x_k)}_{0}\right]\lambda dt\\
&= \int_0^1 \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})).V_{\lambda t}\lambda dt,
\end{aligned}$$

$$f(s_k) - f(x_k) = \int_0^1 \left\langle \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})), V_{\lambda t} \right\rangle \lambda dt, \tag{4.5}$$

$$\forall x_k \in M, \forall \lambda v_{x_k} \in T_{x_k} M.$$

By the mean theorem of integrals, we obtain that $\frac{\partial f}{\partial x}$ and $V_{\lambda t}$ are continuous in $[0,1]$, then there exist $\bar{t} \in [0,1]$ such that

$$\left\langle \frac{\partial f}{\partial x}(x_k), V_{\bar{t}\lambda} \right\rangle - \int_0^1 \left\langle \frac{\partial f}{\partial x}(x_k), V_{t\lambda} \right\rangle dt = 0. \tag{4.6}$$

We want to raise the following relation

$$f(s_k) - f(x_k) - \alpha\lambda \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle. \tag{4.7}$$

From (4.5) and (4.6), (4.7) becomes

$$f(s_k) - f(x_k) - \alpha\lambda \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle$$

$$= \lambda \left[\int_0^1 \left\langle \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k), V_{t\lambda} \right\rangle dt + (1-\alpha) \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle \right]$$

$$\leq \lambda \left[\sup_{t\in[0,1]} \left| \left\langle \frac{\partial f}{\partial x}(\exp(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k), V_{t\lambda} \right\rangle \right| + (1-\alpha) \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle \right], \tag{4.8}$$

$$\forall \alpha \in [0,1], \lambda v_{x_k} \in T_{x_k} M.$$

Now, we suppose that $\operatorname{grad} f(\hat{x}) \neq 0$. Then $\frac{\partial f}{\partial x}(\hat{x}) \neq 0$. Using the Cauchy-Schwartz inequality, we have

$$\left| \left\langle \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k), V_{t\lambda} \right\rangle \right| \leq \left\| \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k) \right\| \|V_{t\lambda}\|$$

and using (4.2) of Lemma 4.1, we obtain for any $0 < \rho < \rho'$

$$(1-\alpha) \left\langle \frac{\partial f}{\partial x_k}(x_k), V_{\bar{t}\lambda} \right\rangle \leq -\rho(1-\alpha) \|\operatorname{grad} f(x_k)\| \|V_{\lambda\bar{t}}\|$$

$$\forall x \in B(\hat{x}, \varepsilon), \forall \lambda \in [0, \bar{\lambda}] \text{for some } \bar{\lambda} > 0.$$

Then (4.8) becomes

$$f(s_k) - f(x_k) - \alpha\lambda \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle$$

(4.9)

$$\leq \lambda \left[\sup_{t\in[0,1]} \left(\left\| \frac{\partial f}{\partial x}(\exp(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k) \right\| \|V_{t\lambda}\| \right) - \rho(1-\alpha) \|\operatorname{grad} f(x_k)\| \|V_{\lambda\bar{t}}\| \right]$$

$$\forall x \in B(\hat{x},\varepsilon), \forall \lambda \in [0,\bar{\lambda}] \text{ for some } \bar{\lambda} > 0,$$

and from the isometry property of parallel translation $\|V_{t\lambda}\| = \|v_{x_k}\|$, we get

$$f(s_k) - f(x_k) - \alpha\lambda \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle$$

$$\leq \lambda \|v_{x_k}\| \left\{ \sup_{t\in[0,1]} \left\| \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k) \right\| - \rho(1-\alpha) \|\operatorname{grad} f(x_k)\| \right\}, \tag{4.10}$$

$$\forall x_k \in B(\hat{x},\varepsilon), \forall \lambda \in [0,\bar{\lambda}] \text{ with } \bar{\lambda} > 0.$$

We have $\alpha < 1$ and $\rho > 0$; then, for some $b > 0$,

$$\rho(1-\alpha) \|\operatorname{grad} f(x_k)\| > b > 0, \forall x_k \in \bar{B}(\hat{x},\varepsilon). \tag{4.11}$$

On the other side, $\frac{\partial f}{\partial x}(x)$ is uniformly continuous in $\bar{B}(\hat{x},\varepsilon)$, and therefore there exist an $a > 0$ such that if

$$\| \exp(\lambda t v_{x_k}) - x_k \| < a, \forall t \in [0,1], \lambda > 0.$$

Then

$$\left\| \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k) \right\| < b, x_k \in \bar{B}(\hat{x},\varepsilon),$$

and thus

$$\sup_{t\in[0,1]} \left\| \frac{\partial f}{\partial x}(\exp_{x_k}(\lambda t v_{x_k})) - \frac{\partial f}{\partial x}(x_k) \right\| < b, \;\; x_k \in \bar{B}(\hat{x},\varepsilon). \tag{4.12}$$

Further, from (4.11) and (4.12), (4.10) becomes

$$f(s_k) - f(x_k) - \alpha\lambda \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle \leq \underbrace{\lambda \|v_{x_k}\| (b-b)}_{0}$$

$$\Rightarrow f(s_k) - f(x_k) \leq \alpha\lambda \left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle, \forall x_k \in \bar{B}(\hat{x},\varepsilon). \tag{4.13}$$

Since $\langle \operatorname{grad} f(x_k), v_{x_k} \rangle < 0$ implies that $\left\langle \frac{\partial f}{\partial x}(x_k), V_0 \right\rangle < 0$ And $\left\langle \frac{\partial f}{\partial x}(x_k), V_\lambda \right\rangle$ is uniformly continuous in $\bar{B}(\hat{x},\varepsilon) \times [0,\bar{\lambda}]$, and therefore for any $\omega \in]0,1[$ there exist $0 < \lambda' < \lambda$, such that

$$\left\langle \frac{\partial f}{\partial x}(x_k), V_{\lambda\bar{t}} \right\rangle < \omega \left\langle \frac{\partial f}{\partial x}(x_k), V_0 \right\rangle.$$

Then

$$f(s_k) - f(x_k) \leq \alpha\lambda\omega \left\langle \frac{\partial f}{\partial x}(x_k), V_0 \right\rangle. \tag{4.14}$$

From (4.2) of Lemma 4.1 we have

$$\left\langle \frac{\partial f}{\partial x}(x_k), V_0 \right\rangle \leq -\rho \left\| \operatorname{grad} f(x_k) \right\| \left\| V_0 \right\|,$$

and from (4.1) of Lemma 4.1, we have

$$-\left\| V_0 \right\| \leq -c \left\| \operatorname{grad} f(x_k) \right\|.$$

Then

$$f(s_k) - f(x_k) \leq -\alpha\lambda\omega\rho c \left\| \operatorname{grad} f(x_k) \right\|^2, \forall x_k \in \bar{B}(\hat{x}, \varepsilon), \forall \lambda \in [0, \bar{\lambda}], \tag{4.15}$$

(4.11) gives

$$\rho(1-\alpha) \left\| \operatorname{grad} f(x_k) \right\| > b \Leftrightarrow \left\| \operatorname{grad} f(x_k) \right\|^2 > \frac{b^2}{\rho^2(1-\alpha)^2} /$$

Then

$$f(s_k) - f(x_k) \leq -\alpha\lambda\omega c \frac{b^2}{\rho(1-\alpha)^2},\ \forall x_k \in \bar{B}(\hat{x}, \varepsilon).$$

λ is given by Armijo algorithm $\lambda = \beta^{i_k}$ Then

$$f(s_k) - f(x_k) \leq -\alpha\beta^{i_k}\omega c \frac{b^2}{\rho(1-\alpha)^2},\ \forall x_k \in \bar{B}(\hat{x}, \varepsilon)$$

Put $m = \alpha\beta^{i_k}\omega c \frac{b^2}{\rho(1-\alpha)^2} > 0$; then we infer

$$f(s_k) - f(x_k) \leq -m,\ \forall x_k \in \bar{B}(\hat{x}, \varepsilon).$$

For the second claim, we do the same considering $t_k = \exp_{s_k}(\lambda v_{s_k})$ the successor of s_k, and we find

$$f(t_k) - f(s_k) \leq -m',\ \forall s_k \in \bar{B}(\hat{x}, \varepsilon).$$

Therefore

$$f(x_{k+1}) - f(x_k) = \underbrace{f(\varepsilon_2^{(k)}) - \ f(t_k)}_{\leq 0} + \underbrace{f(t_k) - f(s_k)}_{\leq -m'} + \underbrace{f(s_k) - f(x_k)}_{\leq -ma},$$

$$f(x_{k+1}) - f(x_k) \leq -(m + m'),$$

and hence $f(x_{k+1}) - f(x_k) \leq -\delta$ for $\delta = m + m'$. □

5 Application

5.1 Example 1: The unit sphere

We will apply the two methods steepest descent and ε-steepest descent to solve the following problem of minimization

$$\left\{ \min_{x \in S^{n-1}} f(x) \right\},$$

where $f : S^{n-1} \to \mathbb{R}$ is the Rayleigh quotient, $f(x) = x^T Ax$, and A is a symmetric matrix (not necessarily positive definite).

Consider the function $\bar{f} : \mathbb{R}^n \to \mathbb{R}, \bar{f}(x) = x^T Ax$, where f is its restriction to S^{n-1}, and note that S^{n-1} is a Riemannian submanifold of $\mathbb{R}^n$ with the Riemannian metric $\bar{g}(\xi, \eta) = \xi^T \eta$.

	S^{n-1}	$\mathbb{R}^n$
function	$f(x) = x^T Ax, x \in S^{n-1}$	$f(x) = x^T Ax, x \in \mathbb{R}^n$
metric	induced metric	$\bar{g}(\xi, \eta) = \xi^T \eta$
tangent space	$v \in \mathbb{R}^n, x^T v = 0$	$\mathbb{R}^n$
gradient	$\operatorname{grad} f(x) = P_x \operatorname{grad} f(x)$	$\operatorname{grad} f(x) = 2Ax$
restraction	$\exp_{x_k}(t.v_k)$	$x_{k+1} = x_k + t_k v_k$

1. We take the descent direction

$$v_k = -\operatorname{grad} f(x_k) = -2(Ax_k - x_k x_k^T Ax_k)$$

and the exponential map $\exp_{x_k} = x_k \cos(t) + \frac{v_k}{\|v_k\|} \sin(t)$, where $\|.\|$ is the Euclidian norm.

Numerical results

The following results are obtained using the steepest descent and R-epsilon steepest descent method with $MATLAB$, where $x_0 = [0.6, 0.8]^T$, $A = \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix}$, $\sigma = 0.1$, $\beta = 0.5, \bar{\lambda} = 1, \varepsilon = 0.00001$

Table 1: steepest descent:

x_k	iteration	$f(x_k)$	$\|\operatorname{grad} f(x_k)\|$
x_0	1	6.160000	3.760000
x_1	2	-2.148723	6.909940
x_2	3	-3.358457	2.565460
x_3	4	-3.524772	0.081704
x_4	5	-3.524797	0.075324
x_5	6	-3.524938	0.003190
x_6	7	-3.524938	0.001717
x_7	8	-3.524938	$7.3617.10^{-4}$
x_8	9	-3.524938	$4.9063.10^{-4}$
x_9	10	-3.524938	$1.2277.10^{-4}$
x_{10}	11	-3.524938	$3.0579.10^{-5}$
x_{11}	12	-3.524938	$7.7587.10^{-6}$

Table 2: R-epsilon steepest descent

x_k	iteration	$f(x_k)$	$\|\text{grad}\, f(x_k)\|$
x_0	1	6.160000	3.760000
x_1	2	-3.522032	0.341732
x_2	3	-3.524938	$2.0946.10^{-5}$
x_3	4	-3.524938	$1.2561.10^{-15}$

Comparison

According to the tables, it is clear that the method R-ε-steepest descent improves the method of R-steepest descent. For $x_0 = (0.6, 0.8)$, we obtain the optimal point $x^* = (-0.6710053, 0.741453)$ after 12 iterations using the first method while we obtains this result after 6 iterations using the second method.

2. Considering the same example with the retraction $R_{x_k}(v_k) = \frac{x_k + v_k}{\|x_k + v_k\|}$, we obtain the following results

Table 3: steepest descent

x_k	iteration	$f(x_k)$	$\|\text{grad}\, f(x_k)\|$
x_0	1	6.160000	3.760000
x_1	2	-3.478254	1.366731
x_2	3	-3.522032	0.341732
x_3	4	-3.524748	0.087431
x_4	5	-3.524937	0.022401
x_5	6	-3.524938	0.005740
x_6	7	-3.524938	0.001471
x_7	8	-3.524938	$3.7685.10^{-4}$
x_8	9	-3.524938	$9.6562.10^{-5}$
x_9	10	-3.524938	$2.4743.10^{-5}$
x_{10}	11	-3.524938	$6.3399.10^{-6}$

Table 4: R-epsilon steepest descent

x_k	iteration	$f(x_k)$	$\|\text{grad}\, f(x_k)\|$
x_0	1	6.160000	3.760000
x_1	2	-3.522032	0.341732
x_2	3	-3.524938	$2.0946.10^{-5}$
x_3	4	-3.524938	$1.2561.10^{-15}$

5.2 Example 2

We shall minimize the function $\bar{f} : \mathbb{R}^2 \to \mathbb{R}$, $\bar{f}(x) = x_1^2 + \frac{1}{2}x_2^2$, restricted to the unitary disc $S^1 = \{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$, i.e., we minimize the function $f : S^1 \to \mathbb{R}$,

$f(x) = x_1^2 + \frac{1}{2}x_2^2$, and we take the same retraction $R_{x_k}(v_k) = \frac{x_k+v_k}{\|x_k+v_k\|}$.

Numerical results

Table 5: steepest descent

x_0	nb.iteration	$f(x_k)$	$\|\text{grad}\, f(x_k)\|$
$(3,5)$	4	0.500000	$7.0773.10^{-9}$
$(0.7,2.5)$	3	0.500000	$6.8286.10^{-6}$
$(4,9.3)$	4	0.500000	$7.3685.10^{-12}$
$(1.7,0.6)$	6	0.500000	$4.2561.10^{-10}$
$(0.7,0.6)$	5	0.500000	$6.1556.10^{-12}$

Table 6: R-epsilon steepest descent

x_0	nb.iteration	$f(x_k)$	$\|\text{grad}\, f(x_k)\|$
$(3,5)$	3	0.500000	$7.0773.10^{-9}$
$(0.7,2.5)$	2	0.500000	$6.8286.10^{-6}$
$(4,9.3)$	3	0.500000	$7.3685.10^{-12}$
$(1.7,0.6)$	2	0.500000	$2.4825.10^{-16}$
$(0.7,0.6)$	3	0.500000	$6.1556.10^{-12}$

Acknowledgements. The authors acknowledge the editors of BSGP for their encouragements.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, New Jersey, 2008.

[2] C. Brezinski, *Acceleration of the Convergence in Numerical Analysis*, Lecture Notes in Mathematics 584, Springer Verlag, 1977.

[3] N. Djeghaba, B. Benzine, *Acceleration of the convergence of steepest descent method*, Demonstratio Mathematica, 39 (2006), 169-181.

[4] S. Kobayashi, K. Nomizu, *Foundations of Differential Geometry*, vol. 1, 15(1969), John Wiley & Sons, New York, London, 1963.

[5] N. Rahali, N. Djeghaba, R. Benzine, *Global convergence of the Armijo epsilon steepest descent algorithm*, Rev. Anal. Numér. Théor. Approx. 41, 2 (2012), 169-180.

[6] S.T. Smith, *Geometric Optimization Methods for Adaptive Filtering*, Harvard University, Cambridge, Massachusetts, 1993.

[7] Yang.Y, *Globally convergent optimization algorithms on Riemannian manifolds: uniform framework for unconstrained and constrained optimization*, Journal of Optimization of Theory and Applications 132, 2 (2007), 245-265.

Authors' address:

Maroua Mezghiche, Nacira Djeghaba and Zoubida Souici Benhammadi

Department of Mathematics, Badji Mokhtar Univ., Box 12, Annaba, 23000, Algeria.

E-mail: marouamezghiche@gmail.com
n.djeghaba@epst-annaba.dz ;
zsouicibenhammdi7@gmail.com