A Proximal Regularization of the Steepest Descent Method in Riemannian Manifold

J.X. da Cruz Neto, O.P.Ferreira and L.R.Lucambio Perez

Abstract

We extend the steepest descent method to solve optimization problems in Riemannian manifolds. Same proof's techniques used in R^n can be modified to prove the existence of cluster points, that such cluster points are critical points and that if the manifold has non-negative sectional curvature and the function is convex then the generated sequence is convergent.

Mathematics Subject Classification: 49M10, 90C30, 53C21
Key words: Convex programming, steepest descent method, Riemannian manifolds.

1 Introduction

The steepest descent method is one of the oldest and simplest procedures for minimization of a real function defined on R^n. It is also the departure point for many other more sophisticated optimization algorithms. Despite its simplicity and notoriety was only in 1995 that its convergence theory was completed - see [8] and [2] -. In [8] the stepsize t_k is calculated by means the addition of a term of proximal regularization at the one-dimensional search. In [2] other two different way for the calculus of the stepsize were introduced. For the first one it is necessary that Lipschitz constant for the gradient of the objective function be known, which may possible to take t_k from a given sequence satisfying some related constraints. The second way is an Armijo-type search. With respect to minimize a convex function over a convex subset of R^n it is known the so called projected subgradient method. Its completed analysis of convergence can be found in [1]. Actually, in [1], the authors proved the convergence of the generated sequence to a solution point for problems in Hilbert spaces.

From another point of view, an extended class of non- convex constrained minimization problems can be seen as minimization problems in Riemannian manifolds. The study of the extension of known optimization methods to solve minimization problems over Riemannian manifolds was the subject of various works -see [4], [5], [7], [10], [11] and their references.

The gradient method as defined in [2] was modified in [4], [5] and [10] in order to use it in the solution of optimization problems in Riemannian manifolds. By means

Balkan Journal of Geometry and Its Applications, Vol.4, No.2, 1999, pp. 1-8

of these modifications, the authors of [4], [5] and [10] obtained the same convergence results as in [2] when the sectional curvature of the Riemannian manifold is non-negative.

The extension of the subgradient method to solve optimization problems in Riemannian manifolds is studied in [7]. The authors obtain the same convergence results as in [3] when the Riemannian manifolds is complete and with non-negative sectional curvature.

In this paper we modify the gradient method with proximal regularization in the one-dimensional search of the stepsize, i.e., as defined in [8], aiming its use in the solution of optimization problems in Riemannian manifolds. We will obtain the same results of convergence, namely, if the problem has solution, the objective function is convex and sectional curvature is non-negative, then the sequence generated by our method converges to a solution.

2 Basic concepts

In this section, we introduce some fundamental properties and notations of Riemannian manifolds. Throughout this paper, all manifolds are smooth and connected. All functions and vector fields are also assumed to be smooth. These basic facts can be find in any introductory book on Riemannian Geometry for example [6] and [9].

Given a manifold M, denote by $\mathcal{X}(M)$ the space of vector fields over M and by T_xM the tangent space of M at x and by $\mathcal{F}(M)$ the ring of functions over M. Let M be endowed with a Riemannian metric $\langle , \rangle$, with corresponding norm denoted by $|\ |$, so that M is now a *Riemannian* manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining points x and y in M, i.e., such that $\gamma(a) = x$ and $\gamma(b) = y$, by $l(\gamma) = \int_a^b |\gamma'(t)|dt$, and, moreover, by minimizing this length functional over the set of all such curves we obtain a distance $d(p, q)$ which induces the original topology on M. Also, the metric induces a map $f \in \mathcal{F}(M) \mapsto \text{grad } f \in \mathcal{X}(M)$ which associates to each f its *gradient* via the rule $\langle \text{grad } f, X\rangle = df(X)$, $X \in \mathcal{X}(M)$. The chain rule generalizes to this setting in the usual way: $(f \circ \gamma)'(t) = \langle \text{grad } f(\gamma(t)), \gamma'(t)\rangle$. In particular, if f assumes either a maximum or a minimum value at a point $x \in M$, then grad $f(x) = 0$. More generally, points where grad f vanishes are called *critical points* of f.

Let ∇ be the Levi-Civita connection associated to $(M, \langle , \rangle)$. If γ is a curve joining points x and y in M, then, for each $t \in [a, b]$, ∇ induces an isometry (relative to $\langle , \rangle$) $P_\gamma(t) : T_xM \to T_{\gamma(t)}M$, the so-called *parallel transport* along γ from x to $\gamma(t)$. When the reference to a curve joining x and y is not necessary, we use the notation P_{xy}. A vector field V along γ is said to be *parallel* if $\nabla_{\gamma'} V = 0$. If γ' itself is parallel we say that γ is a *geodesic*. The geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second order nonlinear ordinary differential equation, hence γ is determined by its position and velocity at one point as far as it is defined. It is easy to check that $|\gamma'|$ is constant. We say that γ is *normalized* if $|\gamma'| = 1$. The restriction of a geodesic to a closed bounded interval is called a *goedesic segment*. A geodesic segment joining p and q in M is said to be *minimal* if its length equals $d(p, q)$.

A Riemannian manifold is *complete* if geodesics are defined for any values of t. Hopf-Rinow's theorem asserts that if this is the case then any pair of points, say x

and y, in M can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, (M,d) is a complete metric space, and bounded and closed subsets are compact. *In this paper, all manifolds are assumed to be complete.* The *exponential map* $exp_x : T_xM \to M$ is defined by $exp_x v = \gamma_v(1,x)$, where $\gamma(.) = \gamma_v(.,x)$ is the geodesic by it position x and velocity v at one point as far as it is defined. In this case, we can prove that, $exp_x tv = \gamma_v(t,x)$ for any values of t.

One of the fundamental objects of Riemannian manifolds is the *curvature tensor* R defined for $X, Y, Z \in \mathcal{X}(M)$ by

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z,$$

where $[\,,\,]$ is the Lie bracket. Clearly, R is a tensor of type (3,1). Given $x \in M$ and a plane $\sigma \subset T_xM$, the quantity

$$K(u,v) = \langle R(u,v)v, u\rangle / (||u||^2||v||^2 - \langle u, v\,\rangle^2)$$

does not depend on the basis $\{u,v\} \subset \sigma$. Hence, $K(u,v) = K(\,\sigma)$ depends only on σ and is called the *sectional curvature* of σ at x. In the section 4.2 of this paper, we will be mainly interested in Riemannian manifolds for which $K(\sigma) \geq 0$ for any σ. Such manifolds are referred to as *manifolds with nonnegative curvature*. A fundamental geometric property of this class of manifolds is that the distance between geodesics issuing from one point is, at least locally, bounded from above by the distance between the corresponding rays in the tangent space. A global formulation of this general principle is the *law of cosines* that we now pass to describe.

A *geodesic hinge* in M is a pair of normalized geodesics segment γ_1 and γ_2 such that $\gamma_1(0) = \gamma_2(0)$ and at least one of them, say γ_1, is minimal. From now on $l_1 = l(\gamma_1)$, $l_2 = l(\gamma_2)$, $l_3 = d(\gamma_1(l_1), \gamma_2(l_2))$ and $\alpha = \sphericalangle(\gamma_1'(0), \gamma_2'(0))$.

Theorem 2.1 *(Law of cosines) In a complete Riemannian manifold with nonnegative curvature, with the notation introduced above, we have*

$$l_3^2 \leq l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha. \tag{1}$$

Proof. see [4] and [5]. □

We say that $f : M \to R$ is *convex* if, for each geodesic $\gamma : R \to M$, $f \circ \gamma : R \to R$ is convex as a real function, namely,

$$f(\gamma((1-\lambda)a + \lambda b)) \leq (1-\lambda) f(\gamma(a)) + \lambda f(\gamma(b)),$$

for any $\lambda \in [0,1]$. We state now a necessary and sufficient conditions for convexity.

Theorem 2.2 *A function $f : M \to R$ is convex if and only if, for any $p \in M$ and any geodesic $\gamma : [0, +\infty) \to R$ such that $\gamma(0) = p$, we have*

$$f(\gamma(t)) - f(p) \geq t\langle grad\, f(p), \gamma'(0)\rangle. \tag{2}$$

Proof. See, for example, [5] and [11] .

□

Perhaps the most important consequence of this theorem is the following.

Corollary 1 *If $f : M \to R$ is convex then all its critical points are global minimum points.*

Proof. Immediately.

□

3 The Proximal regularization

Let M be a complete Riemannian manifold. We will consider the optimization problem

$$\min_{x\in M} f(x), \tag{3}$$

where $f: M \to R$ is a continuously differentiable function. The steepest descent method for the problem (3) is given bellow.

Algorithm 1 (Classical steepest descent Method).

Take $x_0 \in M$.

For $k = 0, 1, \ldots$ define

$$\begin{aligned} t_k &= \operatorname{argmin}_{t\geq 0} f(exp_{x_k}(-t\,\operatorname{grad} f(x_k))), \\ x_{k+1} &= exp_{x_k}(-t_k \operatorname{grad} f(x_k)). \end{aligned}$$

Therefore t_k is a minimizer of the restriction of the f on the geodesic starting at x_k with tangent vector $-\operatorname{grad} f(x_k)$. By introduction of a *proximal regularization* in the search of the stepsize t_k we modify the steepest descent method and replace it by the following algorithm:

Algorithm 2 (Regularized steepest descent Method).

Our Algorithm 2 requires an exogenous sequence, $\{\lambda_k\}$, of real numbers, such that, for all k, $\lambda^{'} \leq \lambda_k \leq \lambda^{''}$, where $0 < \lambda^{'} \leq \lambda^{''}$.

Initialization Step. Take $x_0 \in M$.

Iterative Step. For $k = 0, 1, \ldots$ define

$$\varphi_k(t) = f(exp_{x_k}(-t\operatorname{grad} f(x_k))) + t^2\lambda_k ||\operatorname{grad} f(x_k)||^2, \tag{4}$$

calculate

$$t_k = \operatorname{argmin}_{t\geq 0} \varphi_k(t) \tag{5}$$

and set

$$x_{k+1} = exp_{x_k}(-t_k \operatorname{grad} f(x_k)). \tag{6}$$

We assume that problem (3) has solutions. Set $f^* = \min_{x\in M} f(x)$.

Proposition 1. *Let $\{x_k\}$ given by the Algorithm 2. The sequence $\{x_k\}$ is well defined and, for all k*

$$\langle \operatorname{grad} f(x_{k+1}) , P_{x_k\, x_{k+1}} \operatorname{grad} f(x_k) \rangle = 2\lambda_k\, t_k ||\operatorname{grad} f(x_k)||^2, \tag{7}$$

where $P_{x_k\, x_{k+1}}$ is the parallel transport along $exp_{x_k}(-t\operatorname{grad} f(x_k))$ from x_k to x_{k+1}.

Proof. By induction, suppose that x_k is known. We have two cases. In the first case $\operatorname{grad} f(x_k) = 0$. In this case $\varphi_k(t) = f(x_k)$, for all t, and therefore any $t \geq 0$ is solution of (5) and by (6) holds $x_{k+1} = x_k$. In second case $\operatorname{grad} f(x_k) \neq 0$. In this case

$$\varphi_k(t) \geq f^* + t^2\lambda_k ||\operatorname{grad} f(x_k)||^2 \tag{8}$$

for all $t \geq 0$. Taking limit in (8) we get $\lim_{t\to\infty} \varphi_k(t) = \infty$, then (5) has solution and x_{k+1} is defined. Since $\varphi_k'(t_k) = 0$, from chain rule and (6), the equality (7) holds. □

From now on $\{x_k\}$ and $\{t_k\}$ refers to the sequences generated by Algorithm 2.

4 Convergence analysis

With our results, which we present next, we show that the techniques to solve optimization problems in R^n from [8] can be extended to solve problems in Riemannian manifolds.

4.1 Weak convergence

Without hypothesis concerning the curvature of Riemannian manifold M and convexity of f, we will prove that if x_0 belongs to a bounded level set of f, then $\{x_k\}$ *converges weakly*, namely, it is bounded, the distance between consecutive iterates goes to zero and all its cluster points are critical points.

Theorem 4.1 *For all k,*

$$f(x_{k+1}) \leq f(x_k) - \lambda_k t_k^2 ||grad\, f(x_k)||^2. \tag{9}$$

Furthermore,

i) *The sequence $\{f(x_k)\}$ is decreasing and convergent. In particular, if level set $M^0 = \{x \in M\,;\, f(x) \leq f(x_0)\}$ is bounded, then the sequence $\{x_k\} \subset M^0$ is bounded too.*

ii) $\sum_{k=0}^{\infty} t_k^2 ||grad\, f(x_k)||^2 < \infty$. *Moreover,* $\lim_{k\to\infty} d(x_k, x_{k+1}) = 0$.

iii) *If $\overline{x}$ is a cluster point of the sequence $\{x_k\}$, then $grad\, f(\overline{x}) = 0$.*

Proof. Definition of t_k at (5) implies (9). Item (i) is an immediate consequence of (9).

To prove item (ii) observe that

$$\begin{aligned} \sum_{k=0}^{n} t_k^2 ||grad\, f(x_k)||^2 &\leq \frac{1}{\lambda'}\big(f(x_0) - f(x_{k+1})\big) \\ &\leq \frac{1}{\lambda'}(f(x_0) - f^*), \end{aligned}$$

and that $d(x_k, x_{k+1}) \leq t_k ||\text{grad}\ f(x_k)||$.

Now, let $\bar{x}$ a cluster point of sequence $\{x_k\}$ and $\{x_{k_j}\}$ the subsequence of $\{x_k\}$ which converges to $\bar{x}$. From item (ii) the sequence $\{x_{k_j+1}\}$ converges also to $\bar{x}$.

From Proposition 1,

$$\langle \text{grad}\, f(x_{k_j+1}), P_{x_{k_j} x_{k_j+1}} \text{grad}\, f(x_{k_j}) \rangle = 2\lambda_{k_j} t_{k_j} ||\text{grad}\, f(x_{k_j})||^2. \tag{10}$$

The Riemannian metric, the parallel transport and gradient field are continuous, and then

$$\lim_{j\to\infty} \langle\, \text{grad}\, f(x_{k_j+1})\,,\, P_{x_{k_j}\, x_{k_j+1}} \text{grad}\, f(x_{k_j})\,\rangle = ||\text{grad}\, f(\bar{x})||^2. \tag{11}$$

Theorem 4.1 item (ii) implies that $\lim_{j\to\infty} t_{k_j} ||\text{grad}\, f(x_{k_j})|| = 0$. Then

$$\lim_{j\to\infty} 2\lambda_{k_j}\, t_{k_j} ||\text{grad}\, f(x_{k_j})||^2 \;=\; 0 \tag{12}$$

because $\{\lambda_{k_j}\}$ is bounded.

From (10), (11) and (12) it follows that $\text{grad}\, f(\bar{x}) = 0$ and item (iii) is proved. □

4.2 Full convergence

To achieve the full convergence of $\{x_k\}$ we need convexity of f and the nonnegativity of the sectional curvature of M, but we don't need the existence of bounded level set of f.

Proposition 2. *Let M be a Riemannian manifold with nonnegative curvature. Let $f\colon M \to \mathbf{R}$ be a convex function. Then, for any $y \in M$ we have*

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + t_k^2 ||grad\, f(x_k)||^2 + 2t_k(f(y) - f(x_k)) \tag{13}$$

for all k.

Proof. See, for example, [4], [5], [10].

Lemma 1 *Let $(M\,,\,d)$ be a complete metric space. If the sequence $\{y_k\} \subset M$ has a cluster point $\overline{y}$ satisfying*

$$d^2(y_{k+1}\,,\,\overline{y}) \;\leq\; d^2(y_k, \overline{y}) \;+\; \epsilon_k, \tag{14}$$

for some sequence $\{\epsilon_k\}$, such that $\epsilon_k \geq 0$ and $\sum_{k=0}^{\infty} \epsilon_k \;<\; \infty$, then $\lim_{k\to\infty} y_k = \overline{y}$.

Proof. See [5].

Consider the set $\mathcal{O} = \{\ddagger \in \mathcal{M} \,:\, \{(\ddagger) \leq \inf_{\|\to\infty} \{(\S^{\|})\}$. Because we are assuming that problem (3) has solutions, then $\mathcal{O}$ is nonempty and, by item (i) of Theorem 4.1 and continuity of f any cluster point of $\{x_k\}$ is in $\mathcal{O}$.

Theorem 4.2 *Let M be a Riemannian manifold with non-negative curvature. Let $f\colon M \to \mathbf{R}$ be a convex function. Then the sequence $\{x_k\}$ converges to a minimizer point of f .*

Proof. Take y in $\mathcal{O}$. From Proposition 2 we have

$$d^2(x_{k+1}\,,\,y) \;\leq\; d^2(x_k\,,\,y) + t_k^2 ||\mathrm{grad}\, f(x_k)||^2 \tag{15}$$

for all k. The least inequality implies

$$d^2(x_{k+1}\,,\,y) \;\leq\; d^2(x_0\,,\,y) + \sum_{j=0}^{k} t_k^2 ||\mathrm{grad}\, f(x_k)||^2 \tag{16}$$

for all k. From Proposition 4.1 item (ii) it follows that the sequence $\{x_k\}$ is bounded. Take $\overline{x}$ a cluster point of $\{x_k\}$ and observe that the equation (16) holds for $y = \overline{x}$. By equation (16) and setting $\epsilon_k \;=\; t_k^2 ||\mathrm{grad}\, f(x_k)||^2$ in the Lemma 1 we have $\lim_{k\to\infty} x_k = \overline{x}$. Therefore, from Theorem 4.1 item (iii), convexity of function f and Corollary 1 it follows that $\overline{x}$ its minimizer.

□

5 Final remarks

The algorithm proposed in [8] solves unconstrained convex problem in R^n. Our algorithm 2 solves constrained non-convex problem in R^n when the constraint set is a Riemannian manifold. Then, we can say that algorithm 2 generalizes, in a certain sense, the algorithm presented in [8], namely, if the Riemannian manifold is the Euclidean space R^n then both algorithms are the same. The algorithm in [1] solves constrained convex problems in Hilbert spaces, in particular, in the R^n. While the problems solved by [1] have to be convex , the problems solved by our algorithm are convex in the Riemannian sense, but they may be non-convex in the usual sense. We remark that the techniques used in the full convergence proof impose restrictions about the manifold, that is, we need non-negative curvature of the Riemannian manifold. It remains to remove this hypothesis.

References

[1] Alber, Y. I., Iusem, A. N. and Solodov, M. V., *On the projected subgradient method for nonsmooth convex optimization in a Hilbert space*, Mathematical Programming 81 (1998) pp 25-35.

[2] Burachik, R. S., Graña Drummond, L. M., Iusem, A. N. and Svaiter, B. F., *Full convergence of the steepest descent method with inexact line searches*, Optimization 32 (1995) pp 137-146.

[3] Correa, R. and Lemaréchal, C., *Convergence of Some Algorithms for Convex Minimization*, Mathematical Programming, Vol. 62 (1993) pp. 267–275.

[4] da Cruz Neto, J. X. and Oliveira, P. R., *Geodesic Methods in Riemannian Manifolds*, Universidade Federal do Rio de Janeiro, Research Report 95–10 PESC/COPPE (1995).

[5] da Cruz Neto, J. X., Lima, L. L. and Oliveira, P. R., *Geodesic Algorithm in Riemannian Manifolds*, Balkan Journal of Geometry and Its Applications, Vol. 3, 2 (1998) pp. 89 - 100.

[6] do Carmo, M. P., *Riemannian Geometry*, Boston, Birkhauser (1992).

[7] Ferreira, O. P. and Oliveira, P. R., *Subgradient Algorithm on Riemannian Manifolds*, Journal of Optimization Theory and Applications, 97,1 (1998), pp 93-104.

[8] Iusem, A. N. and Svaiter, B. F., *A Proximal Regularization of the Steepest Descent Method*, Recherche opérationnelle/Operations Research 29,2 (1995) pp 123-130.

[9] Sakai, T., *Riemannian Geometry*, Translations of Mathematical Monographs, vol. 149, American Mathematical Society, Providence, R.I. (1996).

[10] Udrişte, C., *Convex Functions and Optimization Methods on Riemannian Manifolds*, Mathematics and its Applications, Vol. 297, Kluwer Academic Publishers (1994).

[11] Udrişte, C., *Sufficient Decrease Principle on Riemannian Manifolds*, Balkan Journal of Geometry and Its Applications, 1,2 (1996), pp 111-123.

J.X. da Cruz Neto, DM/CCN/ Universidade Federal do Piauí, Campus da Ininga, 64.049-550, Teresina, PI, Brazil, e-mail: jxavier@ufpi.br ,

O.P. Ferreira, IME/Universidade Federal de Goiás, Campus Samambaia, Caixa Postal 131, CEP 74001-970, Goiânia, GO, Brazil, e-mail: orizon@mat.ufg.br.

L.R.Lucambio Perez, IME/Universidade Federal de Goiás, Campus Samambaia, Caixa Postal 131, CEP 74001-970, Goiânia, GO, Brazil, e-mail:lrlp@mat.ufg.br