Geodesic Algorithms in Riemannian Geometry

J. X. da Cruz Neto, L. L. de Lima and P. R. Oliveira

Abstract

We introduce some variants of Luenberger's geodesic method in the framework of Riemannian Geometry. Global convergence and convergence rate estimates are obtained under additional assumptions on the underlying geometry, namely, by assuming that the Riemannian manifold has non-negative curvature and that the objective function is convex.

Mathematics Subject Classification: 49M10, 90C30, 53C21
Key words: Steepest descents methods, manifolds of nonnegative curvature, convexity, convergence rate.

1 Introduction

Steepest descent methods, in which the concept of gradient is associated to a given metric, constitute a generalization of a wide variety of primal algorithms spanning from Cauchy and Newton in unconstrained Nonlinear Programming to the classical projective and the Dikin and Karmarkar algorithms as well (see [1]). At a first level, i.e., when dealing with *unconstrained* problems, the conceptual transparence of the method quickly leads to neat proofs for the global convergence and sharp bounds for the rate of convergence. If one restrict ourselves, however, to *constrained* problems, namely, problems like

$$\min\, g(x) \;\text{ s.t. }\; h(x) = 0, \tag{1}$$

where $g : R^n \to R$ is the objective function and $h : R^n \to R^m$, $m < n$, specifies the manifold of constraints $\mathbf{M} = \{x \in R^n; h(x) = 0\}$, the situation is not so simple. First of all, the unconstrained descent direction $-\operatorname{grad} g(x)$, $x \in \mathbf{M}$, is not in general tangent to $\mathbf{M}$ (we assume throughout this paper that the rank of dh is everywhere maximal so that $\mathbf{M}$ is a regular submanifold of R^n). This can be promptly fixed by assigning a new descent direction $p(x)$ given by orthogonal projection of $-\operatorname{grad} g(x)$ over $T_x\mathbf{M}$, the tangent space to $\mathbf{M}$ at x. But then a new problem arises since moving x in the direction of $p(x)$ soon leaves $\mathbf{M}$, thus violating the constraint. A way of

Balkan Journal of Geometry and Its Applications, Vol.3, No.2, 1998, pp. 89-100

overcoming this difficulty is to use some ad hoc procedure to project the segment determined by $p(x)$ back to $\mathbf{M}$ so that constraint is restored.

The asymptotic analysis of this methodology was carried out by Luenberger in the pioneering paper [2]. His idea was to move feasible points x along the unique geodesic $x(t)$, $t \geq 0$, of $\mathbf{M}$ such that $x(0) = x$ and $x'(0) = p(x)$. Luenberger remarked that all the ad hoc methods referred to above have the same asymptotic behaviour as his method and then went on to provide, under suitable convexity assumptions, the necessary theoretical analysis by following essentially the same steps as in the unconstrained case.

The purpose of this paper is to introduce some variants of Luenberger's geodesic method in a totally intrinsic framework, namely, by following the direction determined by the gradient of the objective function f relative to some Riemannian metric defined over some smooth manifold $\mathbf{M}$ and then to discuss global convergence and asymptotic rate behavior under additional assumptions on the underlying geometry. The motivation for carrying this program out is basically twofold:

- From our viewpoint, Luenberger's direction $p(x)$ is given by $-\operatorname{grad} f(x)$, where f is the restriction of g to $\mathbf{M}$ and the gradient now is computed relative to the induced metric on $\mathbf{M}$ viewed as a Riemannian submanifold of R^n, so that Luenberger's results are at least partially recovered from ours. There are, however, some important differences between the approaches that deserve some comments. First, Luenberger proves global convergence and convergence rate estimates under the assumption that the Lagrangian L naturally associated to the problem (see section 6 below) is *strongly* convex in the sense that its Hessian H^L, when restricted to $\mathbf{M}$, satisfies $a \leq H^L \leq b$ for some constants $0 < a < b$. In our intrinsic approach, we are able to prove global convergence (Theorem 4 below) under the much weaker convexity assumption $H^f \geq 0$ on f but assuming that $\mathbf{M}$ has *nonnegative* Riemannian curvature as well. The main ingredient in our convergence proof is the *law of cosines* (see Theorem 1 below), a geometric inequality about geodesics which is characteristic to such a class of manifolds. Once we have this convergence result, we can easily recover Luenberger's convergence rate estimates by assuming in addition that f is strongly convex, i.e., $a \leq H^f \leq b$ for constants a and b as above (H^f is now computed intrinsically) and then reducing everything to Luenberger's argument after isometrically imbedding in R^n a small piece of our Riemannian manifold $\mathbf{M}$ containing the minimum of f (Theorem 6 below.) We remark that neither the constraints manifold $\{x \in R^n; h(x) = 0)\}$ has in general nonnegative curvature nor an abstract Riemannian manifold $\mathbf{M}$ can in general be isometrically embedded in R^n in such a way that $\mathbf{M} = \{x \in R^n; h(x) = 0\}$ unless we restrict ourselves to small neighborhoods of a point in $\mathbf{M}$. In other words, Luenberger's and our approach complement each other in a sense.

- It has been realized in recent years that certain interior points algorithms in Linear Programming [1], [3] and in Convex Programming [12]-[15] can be investigated via this intrinsic approach in the sense that it is possible to define a Riemannian metric on the interior of the feasible region in such a way that the gradient of the objective function relative to this metric coincides, up to a sign, with the descent vector field associated to the algorithm. Moreover, these geometries in general are so well behaved that the otherwise difficult problem of explicitly integrating the geodesic equations can be easily solved. This implies in particular that our convergence results regarding geodesic algorithms are well suited to handle this important class of interior points algorithms. As a matter of fact, our approach suggests that, these ideal conditions granted, these problems can be solved with just *one* iteration. We shall not address these questions here, however, since a more complete investigation regarding the use of geodesic methods in Linear Programming will appear in a forthcoming paper [1].

We notice that an extensive treatment of geodesic algorithms in Riemannian geometry appears in [4], [5], [10] - [15].

This paper is organized as follows. In Section 2, we review some basic facts on Riemannian Geometry. We give no proofs since the exposition is meant basically for fixing notation. Section 3 is also informative and contains basic results on convex functions defined on Riemannian manifolds; for completeness, sketches of proofs are included. In Section 4, we introduce two intrinsic variants of Luenberger's geodesic method, whose convergence analysis is discussed in Section 5.

2 Preliminaries on Riemannian geometry

Throughout this paper, unless otherwise stated, all manifolds are smooth and connected. All functions and vector fields are also asssumed to be smooth. As general references for this section, see [5] and [6].

Given a manifold $\mathbf{M}$, denote by $\mathcal{X}(\mathbf{M})$ the space of vector fields over $\mathbf{M}$ and by $\mathcal{F}(\mathbf{M})$ the ring of functions over $\mathbf{M}$. Let $\mathbf{M}$ be endowed with a Riemannian metric $\langle , \rangle$, with corresponding norm denoted by $| \ |$, so that $\mathbf{M}$ is now a *Riemannian* manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a,b] \to \mathbf{M}$ joining points p and q in $\mathbf{M}$, i.e., such that $\gamma(a) = p$ and $\gamma(b) = q$, 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 $\mathbf{M}$. Also, the metric induces a map $f \in \mathcal{F}(\mathbf{M}) \mapsto \text{grad } f \in \mathcal{X}(\mathbf{M})$ which associates to each f its *gradient* via the rule $\langle \text{grad } f, X \rangle = df(X)$, $X \in \mathcal{X}(\mathbf{M})$. 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 $p \in \mathbf{M}$ then grad $f(p) = 0$. More generally, points where grad f vanishes are called *critical points* of f.

Let ∇ be the Levi-Civita connection associated to $(\mathbf{M}, \langle , \rangle)$. If γ is a curve joining points p and q in $\mathbf{M}$, then, for each $t \in [a,b]$, ∇ induces an isometry (relative to $\langle , \rangle$) $P_\gamma(t) : T_p\mathbf{M} \to T_{\gamma(t)}\mathbf{M}$, the so-called *parallel transport* along γ. 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 $\mathbf{M}$ 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 p and q, in $\mathbf{M}$ can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, $(\mathbf{M}, d)$ is a complete metric space and bounded and closed subsets are compact. *In this paper, all manifolds are assumed to be complete.*

The fundamental local invariant of Riemannian manifolds is the *curvature tensor* R defined for $X, Y, Z \in \mathcal{X}(\mathbf{M})$ 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 (1,3). Given $p \in \mathbf{M}$ and a plane $\sigma \subset T_p\mathbf{M}$, the quantity

$$K(x,y) = \frac{\langle R(x,y)y, x\rangle}{|x|^2|y|^2 - \langle x, y\ \rangle^2}$$

does not depend on the basis $\{x, y\} \subset \sigma$. Hence, $K(x,y) = K(\ \sigma)$ depends only on σ and is called the *sectional curvature* of σ at p. In 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 $\mathbf{M}$ is a pair of normalized geodesics γ_1 and γ_2 such that $\gamma_1(0) = \gamma_2(0)$ and at least one of them, say γ_1, is minimal. Set

$$l_1 = l(\gamma_1), \quad l_2 = l(\gamma_2), \quad l_3 = d(\gamma_1(l_1), \gamma_2(l_2))$$

and $\alpha = \angle(\gamma_1'(0), \gamma_2'(0))$. We then have the following

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

$$l_3^2 \leq l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha.$$

For the sake of completeness, we include a proof here. Given a geodesic hinge $(\gamma_1, \gamma_2, \alpha)$ as above, consider a corresponding geodesic hinge in the plane $(\overline{\gamma}_1, \overline{\gamma}_2, \overline{\alpha})$ such that, with self-explanatory notation, $\overline{l}_i = l_i, i = 1, 2$, and $\overline{\alpha} = \alpha$. Since geodesics in the plane are straight lines, the usual law of cosines applied to this plane hinge gives

$$\overline{l}_3^2 = l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha.$$

Now, Toponogov's theorem ((6)) implies $l_3 \leq \overline{l}_3$ and the result follows.

Given $f : \mathbf{M} \to R$, its *Hessian* is the bilinear symmetric form on $\mathcal{X}(\mathbf{M})$ given by

$$H^f(X,Y) = \langle \nabla_X \text{grad } f, Y \rangle.$$

A straightforward computation shows that $H^f(X,Y) = (XY - \nabla_X Y)f$. At critical points of f, we have $(\nabla_X Y)f = 0$ so that $H^f(X,Y) = XY(f)$. As usual, given $a \in R$ we use $H^f \geq a$ as a shorthand for the condition that $H^f(X,X) \geq a|X|^2$ for any X. Similarly,

$$H^f \leq b \quad \text{if} \quad -H^f = H^{-f} \geq -b.$$

3 Convexity in Riemannian manifolds

For further information regarding convex functions on Riemannian manifolds, we refer to [5].

We say that $f : \mathbf{M} \to R$ is *convex* if, for each geodesic $\gamma : R \to \mathbf{M}$, $f \circ \gamma : R \to \mathbf{M}$ 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 now state some well- known necessary and suficient conditions for convexity.

Theorem 3.1 *(First order condition for convexity) A function $f : \mathbf{M} \to R$ is convex if and only if, for any $p \in \mathbf{M}$ and any geodesic $\gamma : [0, +\infty) \to R$ such that $\gamma(0) = p$, we have*

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

Proof. If f is convex, define $h : [0, +\infty) \to R$ by $h(t) = f(\gamma(t))$. Hence, h is convex and hence satisfies $h(t) - h(0) \geq th'(t)$. But $h''(t) \geq 0$ hence $h'(t) \geq h'(0)$. By the chain rule, $h'(0) = \langle \text{grad } f(p), \gamma'(0) \rangle$, as desired. The converse is obvious. ■

Perharps the most important consequence of this theorem is the following

Corollary 3.1 *If $f : \mathbf{M} \to R$ is convex, then all of its critical points are global minimum points. In particular, if $\mathbf{M}$ is compact, then f is constant.*

Proof. Let p be a critical point. Thus, grad $f(p) = 0$. Given $q \in \mathbf{M}$, $q \neq p$, there exists, by Hopf-Rinow's theorem, a geodesic segment $\gamma : [a,b] \to \mathbf{M}$ joining p and q. It follows from (2) that $f(\gamma(b)) - f(\gamma(a)) \geq 0$, hence $f(q) \geq f(q)$, as desired. ■

Theorem 3.2 *(Second order condition for convexity) A function $f : \mathbf{M} \to R$ is convex if and only its Hessian H^f is positive semi- definite.*

Proof. Assume f convex and consider $p \in \mathbf{M}$ and $v \in T_p\mathbf{M}$. Let γ be the unique geodesic such that $\gamma(0) = p$ and $\gamma'(0) = v$. Set, as usual, $h = f \circ \gamma$. Clearly, $h''(t) \geq 0$ for any t. Now, the chain rule implies $h'(t) = \langle \text{grad } f(\gamma(t)), \gamma'(t) \rangle$ and hence (ommiting t)

$$0 \leq h'' = \langle \nabla_{\gamma'} \text{grad } f(\gamma), \gamma' \rangle + \langle \text{grad } f(\gamma), \nabla_{\gamma'}\gamma' \rangle = \langle \nabla_{\gamma'} \text{grad } f(\gamma), \gamma' \rangle = H^f_\gamma(\gamma', \gamma'),$$

since $\nabla_{\gamma'}\gamma' = 0$. Set $t = 0$ and the result follows. The converse statement is proved analogously. ■

4 Geodesic descent algorithms

The concept of geodesic descent was introduced in the literature by Luenberger ([2]) in order to analyse the rate of convergence of the method of gradient projection as applied to the nonlinear programming problem

$$\min\ g(x) \quad \text{s.t.} \quad h(x) = 0.$$

Here, $g : R^n \to R$ is the objective function and $h : R^n \to R^m$, $m < n$, defines the (in general non-linear) constraints. Luenberger's method works as follows:

1. Given x_k, $k \geq 1$, feasible, compute the projection p_k of $-grad\ g(x_k)$ over the tangent plane to the manifold of constraints $\mathbf{M} = \{x \in R^n; h(x) = 0)\}$;
2. Determine the unique geodesic $x(t), t \geq 0$, of $\mathbf{M}$ such that $x(0) = x_k$ and $x'(0) = p_k$;
3. Minimize $g(x(t))$, $t > 0$, obtaining t_k and set $x_{k+1} = x(t_k)$.

The idea underlying the use of the geodesic descent method as an approximation to the gradient projection method is the well-known fact that geodesics behave locally like the segments obtained by gradient projection if $\mathbf{M}$ were flat. Furthermore, as the process converges, Luenberger argues that the distance between points obtained by the two methods goes to zero *faster* than the respective stepsize t_k. Thus, the asymptotic rate of convergence of both methods is the same. From this viewpoint, interest in the geodesic method would be merely theoretical in the sense that the analysis of the rate of convergence is best accomplished if one works intrinsically, i.e., following the geodesics in $\mathbf{M}$, rather than with the projected gradient direction. Furthermore, from a practical viewpoint, the geodesic method has the serious drawback that, due to its nonlinear character, solutions to the geodesic equations are rarely available in closed form. Hence, step 2 above involves a considerable additional numerical effort. However, in recent years, it has been shown that descent vector fields associated to certain interior points algorithms in Linear Programming [3], [1], and in Convex Programming [12] - [15] are gradient fields relative to suitable Riemannian metrics defined on the interior of the feasible set. Moreover, it so happens that, due to the high degree of symmetry of these metrics, the geodesic equations are often explicitly solved. This suggests that geodesic algorithms can be effectively used as computational schemes in Interior Points Approach. This has been our main motivation for working toward a convergence analysis of geodesic algorithms in Riemannian manifolds. As remarked in the Introduction, some of these issues will be treated also in a forthcoming paper ([4]).

We now present two intrinsic variants of Luenberger's basic algorithm. Given a geodesic γ denote by $l(t)$ the length of the segment between $\gamma(0)$ and $\gamma(t)$. Observe that $l(t) = \mu t$ for some $\mu > 0$.

Definition 4.1. If $\Gamma > 0$, $f : \mathbf{M} \to R$ is said to have Γ -Lipschitzian gradient if for any p, $q \in \mathbf{M}$ and any geodesic segment $\gamma : [0, a] \to R$ joining p and q we have

$$|\text{grad}\ f(\gamma(t)) - P_\gamma \text{grad}\ f(p)| \leq \Gamma l(t),$$

for any $t \in [0, a]$.

The first two steps of our algorithms are as follows:

1. Given x_k, $k \geq 1$ feasible, compute $p_k = -\operatorname{grad} f(x_k)$.

2. Determine the unique geodesic $x(t)$, $t \geq 0$, of $\mathbf{M}$ such that $x(0) = x_k$ and $x'(0) = p_k$.

The third step in Luenberger's algorithm described above is replaced generating the following intrinsic algorithms:

ALGORITHM A (fixed step)

Given $\delta_1 > 0$ and $\delta_2 > 0$ such that $\delta_1\Gamma + \delta_2 < 1$, where Γ is the Lipschitz constant associated to *grad* f, choose

$$t_k \in (\delta_1, \frac{2}{\Gamma}(1 - \delta_2)).$$

ALGORITHM B (Armijo search)

Choose $t_k = 2^{-i_k}\bar{t}$, where $\bar{t} > 0$ is given and i_k is the least positive natural number such that

$$f(x(t_k)) \leq f(x_k) - \beta t_k^2 |\text{grad } f(x_k)|^2, \tag{3}$$

with $\beta \in (0, 1)$.

5 Convergence analysis

The following result will be useful later.

Lemma 5.1 *If $\gamma : [a, b] \to \mathbf{M}$ is a geodesic joining p and q in $\mathbf{M}$ and $f : \mathbf{M} \to R$ is given, then*

$$\begin{aligned} f(q) &= f(p) + a\langle grad\ f(p), \gamma'(0)\rangle \\ &\quad + \int_0^a \langle grad\ f(\gamma(t)) - P_{\gamma(t)}(t)grad\ f(p), P_\gamma(t)\gamma'(0)\rangle\, dt. \end{aligned}$$

Proof. If $h = f \circ \gamma$ then the chain rule implies

$$h'(t) = \langle \text{grad } f(\gamma(t)), P_\gamma(t)\gamma'(0)\rangle,$$

since γ' is parallel along γ. On the other hand,

$$\langle \text{grad } f(p), \gamma'(0)\rangle = \langle P_\gamma(t)\text{grad } f(p), P_\gamma(t)\gamma'(0)\rangle,$$

since $P_\gamma(t)$ is an isometry. Now, just apply the fundamental theorem of calculus to h. ∎

We start our analysis by proving a quasi-convergence result for a not necessarily convex function f under some additional compactness assumption on its sub-level sets $\mathbf{M}^a = \{x \in \mathbf{M}; f(x) \leq a\}$, $a \in R$. In the sequel, we shall remove this latter assumption by assuming convexity of f and non-negative curvature for $\mathbf{M}$. We remark that quasi-convergence for (1) under Armijo search has been first proved in [7].

Theorem 5.1 *Let $\{x_k\}$ be a sequence of points generated either by algorithm A or by algorithm B. Then:*
i) there exists a constant β such that

$$f(x_{k+1}) \leq f(x_k) - \beta t_k^2 |grad\ f(x_k)|^2. \tag{4}$$

In particular, $\{f(x_k)\}$ is non-increasing;
ii) if $\mathbf{M}^{f(x_1)}$ is bounded, then $\{x_k\}$ is quasi-convergent to the set of critical points of f in the sense that:

1. *$\{x_k\}$ is bounded;*
2. *$\lim_{x\to+\infty} d(x_{k+1}, x_k) = 0$;*
3. *any accumulation point of $\{x_k\}$ is a critical point of f.*

Proof. Let us consider first the case of algorithm A. Choose γ such that $\gamma'(0) = -\operatorname{grad} f(x_k)$. From lemma 5 above, we have

$$\begin{aligned} f(x_{k+1}) &= f(x_k) - t_k|\operatorname{grad} f(x_k)|^2 \\ &\quad + \int_0^{t_k} \langle \operatorname{grad} f(\gamma(t)) - P_\gamma(t)\operatorname{grad} f(x_k), -P_\gamma \operatorname{grad} f(x_k))\rangle\, dt. \end{aligned}$$

Using the Cauchy-Schwartz inequality, the Γ-Lipschitz condition and the isometry character of P_γ, we get

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) - t_k|\operatorname{grad} f(x_k)|^2 + |\operatorname{grad} f(x_k)|^2 \Gamma \frac{t_k^2}{2} \\ &\leq f(x_k) - (\frac{1}{t_k} - \frac{\Gamma}{2})|\operatorname{grad} f(x_k)|^2 t_k^2. \end{aligned}$$

Since $\delta_1 < t_k < \dfrac{2}{\Gamma}(1-\delta_2)$, we have

$$0 < \frac{\Gamma}{2}(\frac{1}{1-\delta_2}) - \frac{\Gamma}{2} = \frac{\delta_2 \Gamma}{2(1-\delta_2)} = \beta < \frac{1}{t_k} - \frac{\gamma}{2},$$

and i) follows.

Since $\{f(x_k)\}$ is non-increasing and $\mathbf{M}^{f(x_1)}$ is compact, $ii1$) follows readily. Moreover, there exists $\lim_{k\to+\infty} f(x_k) = l^* \leq f(x_k)$, $\forall k$. Thus, from (4) we obtain

$$\sum_{k=0}^{s} t_k^2 |\operatorname{grad} f(x_k)|^2 \leq \frac{1}{\beta} \sum_{k=0}^{s} ((f(x_k) - f(x_{k+1})) \leq \frac{1}{\beta}(f(x_0) - l^*). \tag{5}$$

Hence, $\lim_{k\to+\infty} t_k |\operatorname{grad} f(x_k)| = 0$ and, since $\delta_1 < t_k$, we get

$$\lim_{k\to+\infty} |\operatorname{grad} f(x_k)| = 0. \tag{6}$$

Now, $d(x_{k+1}, x_k) \leq t_k|\text{grad } f(x_k)|$ implies

$$\lim_{k\to+\infty} d(x_{k+1}, x_k) \leq \lim_{k\to+\infty} t_k|\text{grad } f(x_k)| = 0$$

and this proves $ii2$). To prove $ii3$), let x^* be an accumulation point of $\{x_k\}$, i.e., there exists a subsequence $\{x_{k_j}\}$ such that $\lim_{j\to+\infty}\{x_{k_j}\} = x^*$. Since grad f is continuous, we infer

$$\lim_{k\to+\infty} |\text{grad } f(x_k)| = |\text{grad } f(x^*)|.$$

But (6) implies $|grad\ f(x^*)| = 0$, hence x^* is a critical point. This concludes the proof for algorithm A. As for algorithm B, note that (4) is satisfied by construction of $\{x_k\}$, so the proof is the same as above. ■

The following theorem is the heart of this section. It will allows us to give proofs of the global convergence of algorithms A and B.

Theorem 5.2 *Let $f : \mathbf{M} \to R$ convex, where $\mathbf{M}$ has nonnegative curvature. Then, for any $y \in \mathbf{M}$ the following inequality holds*

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + t_k^2|grad\ f(x_k)|^2 + 2t_k(f(y) - f(x_k)), \tag{7}$$

where $\{x_k\}$ is the sequence generated either by algorithm A or B and t_k is the stepsize.

Proof. Consider the geodesic hinge $(\gamma_1, \gamma_2, \alpha)$, where γ_1 is a minimal geodesic segment joining x_k to y, γ_2 is a minimal geodesic segment joining x_k to x_{k+1} and $\alpha = \angle(\gamma_1'(0), -\text{grad } f(x_k))$. By the law of cosines (Theorem 1) we have

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + t_k^2|\text{grad } f(x_k)|^2 - 2t_k d(x_k, y)|\text{grad } f(x_k)| \cos\theta.$$

But

$$\cos(\pi - \theta) = -\cos\theta$$

and

$$\langle \text{grad } f(x_k), \gamma'(0)\rangle = |\text{grad } f(x_k)| \cos(\pi - \theta).$$

Putting this into (7) and using the first order convexity condition (Theorem 2), the result follows. ■

Remark. Inequality (7) has recently been applied in the setting of subgradient algorithms in Riemannian geometry ([8]).

We have seen above that, even assuming a Lipschitz condition on *grad* f, we have been able to prove global convergence in a weak sense only under some additional compactness assumption on f. As noticed above, in the following two theorems, we shall prove global convergence for algorithms A and B under suitable conditions on $\mathbf{M}$ and f.

Theorem 5.3 *Let $f : \mathbf{M} \to R$ be convex with $\mathbf{M}$ a manifold of nonnegative curvature. Then the sequence generated either by algorithm A or B converges globally to a minimum point, should it exist.*

Proof. If $a = \min f(x) > -\infty$ then $\mathbf{M}^a \neq \emptyset$. Let $y \in \mathbf{M}^a$. Inequality (7) gives

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + t_k^2 |\text{grad } f(x_k)|^2. \tag{8}$$

Now, (5) implies

$$\sum_{k=1}^{\infty} t_k^2 |\text{grad } f(x_k)|^2 < +\infty, \tag{9}$$

thus $\{x_k\}$ is bounded by (8). From Hopf-Rinow's theorem, this sequence has some accumulation point x^*. Since $\{f(x_k)\}$ in non-increasing (Theorem 6, *i*)) and f is continuous, we get

$$\lim_{k\to\infty} f(x_k) = f(x^*), \quad f(x^*) \leq f(x_k), \ \forall k. \tag{10}$$

From (8) and (9) above we conclude that $\{x_k\}$ is such that the assumptions of the elementary lemma below holds with $\delta_k = t_k^2 |\text{grad } f(x_k)|^2$. Hence, $\lim_{k\to+\infty} x_k = x^*$. On the other hand, we know that $0 < \delta_1 \leq t_k$ and $\lim_{k\to\infty} t_k |\text{grad } f(x_k)| = 0$. Thus, $\lim_{k\to+\infty} |\text{grad } f(x_k)| = 0$. Since grad f is continuous, we conclude grad $f(x^*) = 0$. From Corollary 3, x^* is a minimum. ■

Lemma 5.2 *Let* $\mathbf{M}$ *be a complete metric space. If the sequence* $\{x_k\} \subset \mathbf{M}$ *has an accumulation point* x^* *satisfying*

$$d^2(x_{k+1}, x^*) \leq d^2(x_k, x^*) + \delta_k,$$

where $\delta_k \geq 0$ *and* $\sum_{k=1}^{\infty} \delta_k < +\infty$*, then* $\lim_{k\to+\infty} x_k = x^*$.

As remarked in the Introduction, it is proved in [2] that the convergence rate of the constrained geometric descent method which solves (1) is given by the Kantorovitch ratio $(b-a)^2/(b+a)^2$, if the associated Lagrangian L (see definition below), when restricted to $\mathbf{M}$, is strongly convex, i.e., $a \leq H^L \leq b$ for $0 < a < b$. We use his argument in order to prove a similar result for our algorithms.

Theorem 5.4 *Let* $f : \mathbf{M} \to R$ *be strongly convex, i.e.,* $a \leq H^f \leq b$ *for constants* a *and* b *as above. Then the convergence rate of algorithms* A *and* B *is also given by the Kantorovitch ratio.*

Proof. Using Nash Theorem ([9]), we can isometrically imbed in some R^n a small piece $\mathcal{U}$ of $\mathbf{M}$ containing a minimum point for f. If $\epsilon > 0$ is small enough then the set of normal segments of radius ϵ centered at points of $\mathcal{U}$ determine a tubular neighborhood $\mathcal{V}$ of $\mathcal{U}$. Clearly, $\mathcal{V}$ has a natural coordinate system given by $y = (x,t) \in \mathcal{U} \times B_\epsilon(0)$, where $B_\epsilon(0) \subset R^m$ is an ϵ-ball (here, $n-m$, $m < n$, is the dimension of $\mathbf{M}$). We identify $(x, 0)$ with x. Define $g : \mathcal{V} \to R$ and $h : \mathcal{V} \to R^m$ by $g(x,t) = f(x)$ and $h(x,t) = t$. It is obvious that $\mathcal{U} = \{y \in \mathcal{V}; h(y) = 0\}$ is a regular submanifold of $\mathcal{V}$ and f is the restriction of g to $\mathcal{U}$. Introduce also the *Lagrangian* $L : \mathcal{V} \times R^m \to R$ by $L(y,\lambda) = g(y) + \langle \lambda, h(y) \rangle$. Since g is constant and h is linear along the x-fibers $\{(x,t); t \in (-\epsilon,\epsilon)\}$, a straightforward computation implies $H^L_{(x,0)} = H^f_x$. Thus, our convexity assumption translates into Luenberger's and everything follows now from Luenberger's result referred to above. ■

6 Conclusions

Luenberger proposed a steepest descent method as an elegant device for analysing the convergence rate estimates for the classical projected gradient algorithm. In this article, we extend this methodology to the the more general framework of Riemannian manifolds with nonnegative curvature. The obvious inclusion of the null curvature case has applications to the metrics associated to affine directions such as Dikin's affine-scale and Karmarkar's projective direction.

References

[1] J.X. da Cruz Neto and P.R. Oliveira, (1995), *A unified view of primal methods through Riemannian metrics*, Technical Report, PESC-COPPE, UFRJ, Brazil, submitted.

[2] D.G. Luenberger, (1972), *The gradient projection method along geodesics*, Management Sci., 18 620-631.

[3] D.A. Bayer and J.C. Lagarias, (1989), *The non-linear geometry of Linear Programming I. Affine and projective scaling trajectories*, Trans. Amer. Math. Soc. 314, 499-526.

[4] J.X. da Cruz Neto and de L.L. Lima, *Geodesic descent methods for constrained linear programming problems*, in preparation.

[5] C. Udriste, (1994), *Convex functions and optimization methods in Riemannian geometry*, Kluwer Academic Publishers.

[6] M.P. Do Carmo, (1991), *Riemannian Geometry*, Birkhausen.

[7] D. Gabay, (1982), *Minimizing a differentiable function over a diferentiable manifold*, J. of Opt. Th. and Applic., 37, 177-219.

[8] O.P. Ferreira and P.R. Oliveira, *Subgradient algorithm on Riemannian manifolds*, J. of Opt. Th. and Applic., 97, 1, 1998.

[9] M. Gromov, (1986), *Partial differential relations*, Ergebnisse Math., Springer Verlag, Berlin.

[10] C.Udrişte, *Convergence of minimization methods on Riemannian manifolds*, International Workshop on Differential and Its Applications, Bucharest, July 25-30, Scientific Bulletin, Polytehnica University of Bucharest, 55, 3-4 (1993).

[11] C.Udrişte, *Minimization algorithms on Riemannian manifolds*, Proceedings of the $23-rd$ Conference on Geometry and Topology, Cluj- Napoca, Romania, 1995, 185-193.

[12] C.Udrişte, *Optimization methods on Riemannian manifolds*, Proceedings of IRB International Workshop, Monteroduni, Italy, August 8-12, 1995, Algebras, Groups and Geometries 14 (1997), 339-359.

[13] C.Udrişte, *Riemannian convexity in programming (I)*, Proceedings of 25-th National Conference on Geometry and Topology, AL.I.Cuza University of Iassy, Romania, September 18-23, 1995; An. St. Univ.AL.I. Cuza, Iassy, 42(1996), 123-136.

[14] C.Udrişte, *Sufficient decrease principle on Riemannian manifolds*, Workshop on Differential Geometry, Global Analysis and Lie Algebras, Aristotle University of Thessaloniki, December 13-16, 1995; Balkan Journal of Geometry and Its Applications, 1,2(1996), 111-123.

[15] C.Udrişte, *Riemannian convexity in programming (II)*, Balkan Journal of Geometry and Its Applications, 1,1(1996), 99-100.

Departamento de Matemática
Universidade Federal do Piauí
Teresina, Brasil
jxavier@ufpi.br

Departamento de Matemática
Universidade Federal do Ceará
Fortaleza, Brasil
levi@mat.ufc.br

Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brasil
poliveir@cos.ufrj.br