A DESCENT ALGORITHM FOR QUASICONVEX OPTIMIZATION OVER RIEMANNIAN MANIFOLD

S.S. Souza
P. S. M. Santos
J. X. da Cruz Neto

Universidade Federal do Piauí
Campus Min. Petrônio Portela, DM - CCN, SG-4, Teresina, Brazil
sissy, psergio, jxavier@ufpi.br

ABSTRACT

In this work we analyze an algorithm based on descent directions for solving the problem of minimizing quasiconvex functions over a Riemannian manifold. The step size is obtained by Armijo's rule with backtracking procedure. A numerical experiment is reported.

KEYWORDS: Descent methods, Quasiconvex optimization, Riemannian manifolds, Mathematical Programming.

1 Introduction

Let M be a complete Riemannian manifold with finite dimension. We consider a method for solving the problem

$$\min_{x \in M} f(x). \tag{1}$$

We assume that $f \colon M \to R$ is a continuously differentiable quasiconvex function.
As argued in Qi et al (2010), optimization on manifolds finds applications in two broad classes of situations: Classical equality-constrained optimization problems where the constraints specify a submanifold of $\mathbb{R}^n$; and problems where the objective function has continuous invariance properties that we want to eliminate for various reasons, e.g., efficiency, consistency, applicability of certain convergence results, avoid failure of certain algorithms due to degeneracy.

Fields of application include computer vision, signal processing, motion and structure estimation, see, for instance, Absil et al (2008), Adler et al (2002), Lee (2005) and references therein. Among the oldest papers that study optimization problems over Riemannian manifolds highlight the variants of Luenberger's basic algorithm, given in Luenberger (1972), presented by Cruz Neto et al in Cruz Neto et al(1998) and Cruz Neto et al(1999). In Luenberger(1972), the author presented an algorithm on geodesics to obtain convergence results for the gradient projection method. In Cruz Neto et al(1998) and Cruz Neto et al(1999), they extended the steepest descent method with Armijo's rule. An interesting remark is these geodesic algorithms exploit better the intrinsic properties of the constraint set and the objective function, as discussed by Yang in Yang(2007).
In this work we propose a descent method based in sufficient directions for solving problem (1). For this purpose we generalize the definition of sufficient descent directions given in Dussault(2000) for Euclidean spaces.
The structure of the work is simple. In Section 2, we present some basic facts in setting of Riemannian manifolds. In Section 3, we define the algorithm and analyze its convergence. Finally, we report a numerical experiment in Section 4.

2 Preliminaries

In this section, we recall some necessary properties and definitions in settings of Riemannian manifolds. These basics fact can be found for example in do Carmo(1992).
We assume throughout this work that all manifolds are smooth and connected and all functions and vector fields are smooth.
Given a manifold M, we denote by T_xM the tangent space of M at x. Let M be a Riemannian manifold with Riemannian metric given by $\langle , \rangle$ and let $\| \cdot \|$ its associated norm. We recall that the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining points x and y in M, it is defined by $l(\gamma) = \int_a^b |\gamma'(t)| dt$. We denote by $d(x, y)$ the distance obtained by minimizing the length functional over the set of all such curves.
The parallel transport along γ from $\gamma(t_0)$ to $\gamma(t_1)$, denoted by P_{γ,t_0,t_1} is an application $P_{\gamma,t_0,t_1} : T_{\gamma_{t_0}} M \to T_{\gamma_{t_1}} M$ given by $P_{\gamma,t_0,t_1} = V(t_1)$ where V is the unique vector field along γ such that $\frac{DV}{dt} = 0$ and $V(t_0) = v$, where $\frac{DV}{dt}$ denotes the covariant derivative for the vector field V along γ. If γ' itself is parallel we say that γ is a geodesic. The restriction of a geodesic to a closed bounded interval is called a geodesic segment. A geodesic segment joining x and y in M is said to be minimal if its length equals $d(x, y)$. Henceforth, for $x, y \in M$, γ_{xy} denotes the geodesic $\gamma_{xy} : [0, 1] \to M$ joining the points x and y, that is, $\gamma_{xy}(0) = x$ and $\gamma_{xy}(1) = y$. We denote by Γ the set of all geodesic arcs γ_{xy} from x to y.
In this paper, we assume that M is a complete manifold, that is, all geodesics are defined

for any values of t, which implies that any pair of points in M can be joined by a (not necessarily unique) minimal geodesic segment.
The exponential map, $exp_x : T_xM \to M$, is defined by $exp_x(v) = \gamma_v(1, x)$, where $\gamma(\cdot) = \gamma_v(\cdot, x)$ is the geodesic by it position x and velocity v at one point as far as it is defined, which implies that $exp_x(tv) = \gamma_v(t, x)$ for any values of t.
Throughout this paper we consider M with nonnegative sectional curvature. A fundamental geometric property of this class of manifolds is the called law of cosines, whose proof can be encountered in Cruz Neto et al(1998). For described it, we recall a geodesic hinge in M is a pair of normalized geodesic segments γ_{xy} and γ_{xz} (that is, $\gamma_{xy}(0) = \gamma_{xz}(0)$) such that at least one of them, say γ_{xy}, is minimal.

Theorem 1 *(Law of cosines) In a complete Riemannian manifold with nonnegative curvature it holds*

$$l^2 \leq l_{xy}^2 + l_{xz}^2 - 2l_{xy}l_{xz}\cos\omega,$$

where $l_{xy} = l(\gamma_{xy})$, $l_{xz} = l(\gamma_{xz})$, $l = d(\gamma_{xy}(l_{xy}), \gamma_{xz}(l_{xz}))$ *and* $\omega = \sphericalangle(\gamma'_{xy}(0), \gamma'_{xz}(0))$.

We will need the concept of normal cone of a subset of a manifold at a point. We recall a vector $\xi \in T_xM$ is said to be a subgradient of f at x when $f(\gamma_{xy}(t)) \geq f(x) + t\langle \xi, \gamma'_{xy}(0)\rangle$, for any $t \geq 0$. Let C be a closed subset of M at $x \in C$. Let $\delta_C(x)$ be the indicator function. We called the normal cone of C at x to the set

$$N(x, C) = \partial\delta_C(x).$$

To conclude this section we recall some properties about quasiconvex function and quasi-Fejér convergence in the setting of the Riemannian geometry.
A function $f : M \to I\!R$ is called quasiconvex when for all $x, y \in M$, $f(\gamma_{xy}(t)) \leq \max\{f(x), f(y)\}$, for all $\gamma_{xy} \in \Gamma$, $t \in (0, 1)$. In a geometrical point of view, the quasiconvex functions are characterized by total convexity of their sublevel sets $\{x \in M \ |f(x) \leq c\}$. One main difference between convex and quasiconvex functions is that the quasiconvex functions are not continuous within their domain and directional derivatives are not necessarily defined. For differentiable quasiconvex functions, the following theorem gives a first-order characterization.

Theorem 2 *Let* $f : M \to I\!R$ *be a differentiable quasiconvex function on a complete Riemannian manifold* M *and let* $x, y \in M$. *Then*

$$f(x) \leq f(y) \quad \Longrightarrow \quad \langle gradf(y), \gamma'_{xy}(0)\rangle \leq 0,$$

where $\gamma_{xy} \in \Gamma$.

Given a (X, d) a complete metric space. A sequence $\{y^k\}$ of X is called quasi-Fejér convergent to a set $U \subset X$ if for all $u \in U$ there exists a real numbers sequence $\{\epsilon_k\}$ such that $\epsilon_k \geq 0$, $\sum_{k=0}^{\infty} \epsilon_k < +\infty$ and $d^2(y^{k+1}, u) \leq d^2(y^k, u) + \epsilon_k$. We will use also the following property, whose proof is similar to that encountered in [**?**, Theorem 1] by using the distance d instead of the Euclidean norm.

Theorem 3 *In a complete metric space* (X, d), *if* $\{y^k\}$ *is quasi-Fejér convergent to a nonempty set* $U \subset X$ *then* $\{y^k\}$ *is bounded. If furthermore a cluster point* $\bar{y}$ *of* $\{y^k\}$ *belongs to* U, *then* $\{y^k\}$ *converges and* $\lim_{k\to+\infty} y^k = \bar{y}$.

3 The sufficient descent method

For Euclidean spaces, as defined by Dussault in (Dussault, 2000), d_x is said to be a sufficient descent direction if there exist two positive constants γ_0 and γ_1, independents of x, such that

$$\begin{array}{rcl} d_x^T \nabla f(x) & \leq & -\gamma_0 \|\nabla f(x)\|_2^2 \\ \|d_x\|_2 & \leq & \gamma_1 \|\nabla f(x)\|_2, \end{array}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. We extend the definition of sufficient descent direction for optimization on Riemannian manifolds as follows.
We consider positive real numbers α and β such that $d^k \in T_{x^k}M$ satisfies

$$\langle \mathrm{grad} f(x^k), d^k \rangle \leq -\alpha \|\mathrm{grad} f(x^k)\|^2 \quad (2)$$

$$\|d^k\| \leq \beta \|\mathrm{grad} f(x^k)\|, \quad (3)$$

where the inner product and the norm are given by a Riemannian metric. In this case, we called d^k by sufficient descent directions.
In order to describe our method we denote $L_\lambda = \{x \in M : f(x) \leq \lambda\}$. In addition, given a point x^k, we define

$$D_k = \{d \in T_{x^k}M : \ -d \in N(x^k, L_{f(x^k)}),\ d \text{ satisfies } (2)-(3)\}. \quad (4)$$

Algorithm 1 *(Sufficient Descent Algorithm)*
Initialization: *Let $x^0 \in M$ and set real numbers $\theta \in (0,1)$ and $\tau \in (0,\ 1/2)$*
Main Step: *Given $x^k \in M$ and $d^k \in D_k$,*
If $\mathrm{grad} f(x^k) = 0$ stop.
Else, find

$$x^{k+1} = exp_{x^k}(\theta_k d^k), \quad (5)$$

where

$$\theta_k = argmax\{\theta^j : f(exp_{x^k}(\theta^j d^k)) - f(x^k) \leq \tau\theta^j \langle \mathrm{grad} f(x^k), d^k\rangle, j \in I\!N\}. \quad (6)$$

Remark 1 *We note that D_k is a nonempty set since f is a quasiconvex function, which (by Theorem 2) implies that $d^k = -\mathrm{grad} f(x^k) \in D_k$. Therefore, the sequence generated by (5) always exists due to the existence of θ_k (the proof of this existence is very simple and it can be encountered for example in (Cruz Neto et al, 1998)).*

From now on we consider that the Algorithm 1 generates an infinite sequence, $\{x^k\}$, of iterates. Let $U = \{x \in M : f(x) < f(x^k),\ \forall k \in I\!N\}$.

Lemma 1 *For Algorithm 1, it holds*

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + \theta_k^2 \|d^k\|^2 - 2\theta_k \langle \gamma'_{x^k\,x}(0), d^k\rangle,$$

for all $x \in U$.

Proof:
Let $x \in U$ be an arbitrary point. Suppose that $\gamma_{x^k x}$ and $\gamma_{x\, x^{k+1}}$ are minimal geodesic segments, $\gamma_{x^k\, x^{k+1}} \in \Gamma$ is such that $\gamma'_{x^k\, x^{k+1}}(0) = \theta_k d^k$ and moreover ω is the angle between $-\gamma'_{x^k\, x}(0)$ and $\gamma'_{x^k\, x^{k+1}}(0)$. From Theorem 1, we have

$$\begin{array}{rcl} d^2(x^{k+1}, x) & \leq & d^2(x^k, x) + \theta_k^2 \|d^k\|^2 + 2d(x^k, x)\theta_k \|d^k\| \cos(\omega) \\ d^2(x^{k+1}, x) & \leq & d^2(x^k, x) + \theta_k^2 \|d^k\|^2 - 2d(x^k, x)\theta_k \|d^k\| \cos(\pi - \omega) \\ & = & d^2(x^k, x) + \theta_k^2 \|d^k\|^2 - 2\theta_k \langle \gamma'_{x^k\, x}(0), d^k\rangle. \end{array}$$

■

Lemma 2 *For Algorithm 1, we have*

(a) $\sum_{k=0}^{\infty} \theta_k^2 \|d^k\|^2 < +\infty$;

(b) *the sequence* $\{f(x^k)\}$ *is nonincreasing.*

Proof:
Part (a). Since θ_k satisfies the Armijo's criterion and d^k is a sufficient descent direction, we get

$$\begin{array}{rcl} f(x^{k+1}) - f(x^k) & \leq & \tau\theta_k\langle \mathrm{grad} f(x^k), d^k\rangle \\ & \leq & -\tau\theta_k\alpha\|\mathrm{grad} f(x^k)\|^2 \\ & \leq & -\frac{\tau\theta_k\alpha}{\beta^2}\|d^k\|^2. \end{array} \tag{7}$$

Hence,

$$\theta_k\|d^k\|^2 \leq \frac{\beta^2}{\alpha\tau}[f(x^k) - f(x^{k+1})].$$

By definition of θ_k, last inequality becomes

$$\theta_k^2\|d^k\|^2 \leq \frac{\beta^2}{\alpha\tau}[f(x^k) - f(x^{k+1})]$$

and thus

$$\sum_{k=0}^{l} \theta_k^2\|d^k\|^2 \quad \leq \quad \frac{\beta^2}{\alpha\tau}[f(x^0) - f(x^{l+1})] \quad \leq \quad \frac{\beta^2}{\alpha\tau}[f(x^0) - f^*],$$

where f^* is the optimum value of the problem (1). Then, $\sum_{k=0}^{\infty} \theta_k^2\|d^k\|^2 < \infty$.

Part (b). It follows immediately from the inequality (7). ■

Remark 2 *We recall that the set* D_k *is defined by (4) for* $\lambda = f(x^k)$*. Note that if* $d^k \in D_k$ *then* $-d^k \in N(x^k, L_{f(x^k)})$*. It is a known result that* $\xi \in \partial\delta_C(x)$ *if, and only if,* $\langle \xi, \gamma'_{xy}(0)\rangle \leq 0$*. Thus,* $\langle d^k, \gamma'_{x^k y}(0)\rangle \geq 0$ *for any* $\gamma_{x^k y} \in \Gamma$*.*

Next we establish quasi-Fejér convergence of the sequence of the iterates.

Proposition 1 *The sequence* $\{x^k\}$ *is quasi-Fejér convergent to* U*.*

Proof:
Let $\bar{x} \in U$. Since $d^k \in D_k$, we have $\langle \gamma'_{x^k\,\bar{x}}(0), d^k\rangle \geq 0$ for all $\gamma_{x^k\,\bar{x}} \in \Gamma$. So, from Lemma 1, we obtain

$$\begin{array}{rcl} d^2(x^{k+1}, \bar{x}) & \leq & d^2(x^k, \bar{x}) + \theta_k^2\|d^k\|^2 - 2\theta_k\langle\gamma'_{x^k\,\bar{x}}(0), d^k\rangle \\ & \leq & d^2(x^k, \bar{x}) + \theta_k^2\|d^k\|^2. \end{array}$$

Last inequality and Lemma 2 imply the result. ■

Proposition 2 *The sequence* $\{x^k\}$ *is convergent.*

Proof:
From Proposition 1 we obtain that $\{x^k\}$ is quasi-Fejér convergent to U. Therefore, by Theorem 3, $\{x^k\}$ is bounded. Let $\{x^l\}$ a subsequence of $\{x^k\}$ which converges to $\bar{x}$. By Theorem 3 it is enough to show that $\bar{x} \in U$. From continuity of f, $\lim_{l\to+\infty} f(x^l) = f(\bar{x})$ and due to the Part (b) from Lemma 2 we have that the whole sequence $\{f(x^k)\}$ converges to $f(\bar{x})$. So, $f(\bar{x}) < f(x^k)$ for all $k \in I\!N$ and thus $\bar{x} \in U$. ■

Proposition 3 *The sequence $\{gradf(x^k)\}$ converges to zero.*

Proof:
By first inequality in (7) we have

$$-\tau\theta_k\langle \mathrm{grad} f(x^k), d^k\rangle \ \leq \ f(x^k) - f(x^{k+1}).$$

By summing both members in the last inequality we obtain the convergence of the series of the left hand side and consequently

$$\lim_{k\to+\infty} \theta_k\langle \mathrm{grad} f(x^k), d^k\rangle = 0. \tag{8}$$

We proceed to prove that $\lim_{k\to+\infty}\inf\langle \mathrm{grad} f(x^k), d^k\rangle = 0$ by contradiction. Suppose that there exists $\bar{\eta} < 0$ such that

$$\lim\inf_{k\to+\infty}\langle \mathrm{grad} f(x^k), d^k\rangle = \bar{\eta}. \tag{9}$$

By Proposition 2 we have that $\lim_{k\to+\infty} x^k = \bar{x} \in U$. The definition of θ_k in (6) implies that θ_k converges to zero and moreover that there exists $\delta > 0$ such that $\forall t_k \in (\theta_k, \delta\theta_k]$ it holds

$$f(x^k) + \tau t_k\langle \mathrm{grad} f(x^k), d^k\rangle \leq f(\exp_{x^k}(t_k d^k)). \tag{10}$$

Consider the function $\phi : [0, t_k] \to I\!R$ given by $\phi(t) = f(\exp_{x^k}(td^k))$. By applying Mean Value Theorem on ϕ, for such k, there exists $\bar{t}_k \in [0, t_k]$ such that

$$f(\exp_{x^k}(t_k d^k)) - f(x^k) = t_k\langle \mathrm{grad} f(\exp_{x^k}(\bar{t}_k d^k)), P_{\bar{\gamma}_k,0,\bar{t}_k}(d^k)\rangle, \tag{11}$$

where $P_{\bar{\gamma}_k,0,\bar{t}_k}(d^k)$ is the parallel transport of d^k along geodesic $\bar{\gamma}_k$ such that $\bar{\gamma}_k(0) = x^k$ and $\bar{\gamma}_k{}'(0) = d^k$. Combining (10) and (11), we obtain

$$\tau\langle \mathrm{grad} f(x^k), d^k\rangle \leq \langle \mathrm{grad} f(\exp_{x^k}(\bar{t}_k d^k)), P_{\bar{\gamma}_k,0,\bar{t}_k}(d^k)\rangle.$$

Passing to the lim inf as $k \to +\infty$ in inequality above and taking in account the continuity of $\mathrm{grad} f$, exponential map and parallel transport we get $\tau\bar{\eta} \leq \bar{\eta}$, which implies that $\tau \geq 1$ and we have a contradiction. Therefore, $\lim_{k\to+\infty}\langle \mathrm{grad} f(x^k), d^k\rangle = 0$. Then, by (2), $\lim_{k\to+\infty} \mathrm{grad} f(x^k) = 0$. The proof is complete. ■

The following theorem is a consequence of the last result.

Theorem 4 *The sequence $\{x^k\}$ converges to a stationary point.*

Proof:
By Proposition 3, if $\bar{x}$ is a limit point of $\{x^k\}$ then $\mathrm{grad} f(\bar{x}) = 0$ and the proof is complete.
■

4 Numerical experiments

In this section we present a numerical experiment to illustrates the performance of the Sufficient Descent Algorithm (SDA) for solving the minimization of a quasiconvex function on a Hypercube. In this example we compare the implementation of the steepest descent method given in Papa Quiroz et al(2008) with a SDA implementation where the sufficient directions are rotations of the gradient direction.

The algorithm was coded in SCILAB 5.1.1 on a 2GB RAM Dual Core Pentium notebook.

We denote *Iter(k)* the number of iterations and by *Call.Armijo* the number of steps in Armijo's search.

We consider the problem

$$\min\{f(x): 0 \le x \le e\},$$

where $x = (x_1, \ldots, x_n)$, $e = (1, \ldots, 1) \in \mathbb{R}^n$. As suggested in Papa Quiroz et al(2008), take the connected and complete Riemannian manifold $M = ((0,1)^n, X^{-2}(I-X)^{-2})$, then the iteration of SDA becomes

$$x_i^{k+1} = \frac{1}{2}\left\{1 + \tanh\left(\frac{1}{2}x_i^k(1-x_i^k)d_i^k\theta_k + \frac{1}{2}\log\frac{x_i^k}{1-x_i^k}\right)\right\},\ i = 1,2,\ldots,n$$

where $\theta_k = 2^{-j_k}$, j_k is the least positive integer such that

$$f(x^{k+1}) \le f(x^k) - \tau\theta_k\|v^k\|^2,$$

$v^k = -X_k^2(I-X_k)^2 d^k$ and $\tau \in (0,1)$.

Next we present one experiment with stop test $d(x^{k-1}, x^k) \le 10^{-6}$ where $d(x,y)$ is the geodesic distance between points x and y, as defined by

$$d(x,y) = \left\{\sum_{i=1}^{n}\left[\log\left(\frac{y_i}{1-y_i}\right) - \log\left(\frac{x_i}{1-x_i}\right)\right]^2\right\}^{\frac{1}{2}}.$$

The function $f(x) = \sqrt{-log\,(x_1(1-x_1)x_2(1-x_2))}$ is quasiconvex in M and has an unique minimal point at $x^* = (0.5, 0.5)$ with $f(x^*) = 2\sqrt{\log 2} \approx 1.665109222$. First, we use $\tau = 0.1$ and ϑ_k varying between 55^0 and 60^0 with d^k given by a rotation of $-\text{grad} f(x^k)$, that is, $d^k = -R(\vartheta_k)\text{grad} f(x^k)$. We show the SDA behavior in Table 1. Next, we use $\tau = 0.1$, $\alpha_k = 1, \beta_k = 2$, $\vartheta = 60^0$ and $d^k = -R(\vartheta)\text{grad} f(x^k)$ to compare SDA with SDM, by considering the same starting guesses given in Papa Quiroz et al(2008), and those results are shown in Table 2.

Table 1: Behaviour with variable rotate matrix

Iter(k)	Call.Armijo	x^k	$f(x^k)$	$d(x^{k-1}, x^k)$	ϑ_k
0	–	$(0.100000, 0.900000)$	2.19451	–	–
1	1	$(0.866669, 0.222506)$	1.97800	5.3336894	55.31
2	1	$(0.466446, 0.541680)$	1.66855	2.4569044	57.21
3	1	$(0.504297, 0.525710)$	1.66592	0.1646317	58.02
4	1	$(0.516183, 0.517268)$	1.66578	0.0583701	58.46
⋮	⋮	⋮	⋮	⋮	⋮
20	1	$(0.500584, 0.500243)$	1.66510	0.0007996	59.64
⋮	⋮	⋮	⋮	⋮	⋮
44	1	$(0.500000, 0.500000)$	1.66510	$0.79866e-007$	59.84

Table 2: Comparison with SDM

x^0	Scheme	Iter(k)	Call.Armijo	x^k	$d(x^{k-1}, x^k)$
$(0.45, 0.51)$	SDM	65	65	$(0.499999, 0.500000)$	$9.27003e-007$
	SDA	33	34	$(0.500000, 0.500000)$	$9.21483e-007$
$(0.40, 0.60)$	SDM	71	71	$(0.499999, 0.500001)$	$9.93398e-007$
	SDA	45	45	$(0.500000, 0.500000)$	$8.11867e-007$
$(0.10, 0.90)$	SDM	85	85	$(0.499999, 0.500001)$	$8.92053e-007$
	SDA	52	52	$(0.500000, 0.500000)$	$8.09453e-007$
$(0.20, 0.30)$	SDM	79	79	$(0.499999, 0.499999)$	$8.79813e-007$
	SDA	45	45	$(0.499999, 0.499999)$	$7.87057e-007$
$(0.70, 0.60)$	SDM	75	75	$(0.500001, 0.500001)$	$8.82938e-007$
	SDA	47	47	$(0.500000, 0.500000)$	$8.00870e-007$

5 Final remarks

In this paper we have presented a sufficient descent algorithm, with the stepsize chosen by an Armijo's criterion, for quasiconvex problems on Riemannian manifolds. Under mild assumptions, we have established full convergence of the sequence of the iterates to a stationary point. A preliminary numerical experiment on the quasiconvex problem indicates that significant gains in the number of iterations can be achieved when we compare our method with an implementation of the Steepest Descent Method found in the literature.

References

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

Adler R. L., Dedieu J.-P., Margulies J. Y., Martens M., Shub M. (2002), Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA Journal of Numerical Analysis*, 22, 359-390.

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

Cruz Neto, J. X., Lima, L. L. and Oliveira, P. R. (1998), Geodesic algorithm in Riemannian manifolds, *Balkan Journal of Geometry its Applications*, 3, 89-100.

Cruz Neto, J.X., Ferreira, O.P., Lucambio Perez, L.R. (1999), A proximal regularization of the steepest descent method in Riemannian manifolds, *Balkan J. Geom. Appl.*, 4, 1-8.

do Carmo, M. P., *Riemannian Geometry*, Birkhauser, Boston, 1992.

Dussault, J.P. (2000), Convergence of implementable descent algorithms for unconstrained optimization, *J. Optim. Theory Appl.*, 104, 739-750.

Lee, P. Y., Geometric Optimization for Computer Vision, *PhD thesis*, Australian National University, 2005.

Luenberger, D.G. (1972), Linear and nonlinear programming, *Management Science*, 18, 620-631.

Papa Quiroz, E. A., Quispe, E. M. and Oliveira, P. R. (2008), Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemmanian manifolds,*Journal of Mathematical Analysis and Aplications*, 341, 467-477.

Qi, C., Gallivan, K. and Absil, P. -A., Riemannian BFGS Algorithm with Applications, To appear in *Recent Trends in Optimization and its Applications in Engineering*, Eds. M. Diehl, F. Glineur, and W. Michiels, Springer.

Yang, Y. (2007), Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization, *J. Optim. Theory Appl.*, 132, 245-265.