Noname manuscript No.
(will be inserted by the editor)

A modified Armijo for damping a Riemannian Newton-type method

Marcio Antônio de Andrade Bortoloti ·
Teles Araújo Fernandes

Received: date / Accepted: date

Abstract In this work, we present a strategy of linesearch to find singularities of vector fields defined on Riemannian manifolds. We adapt the Armijo rule for preventing a too small step length on the Newton direction, when it exists. Otherwise, we keep the Armijo rule on a safeguard direction. This strategy improves the damped Riemannian Newton method. We employ the proposed algorithm to minimize a joint diagonalization problem on the Stiefel manifold, and to find singularities of a non-conservative vector field on the sphere. Our numerical study shows the presented strategy has a good performance.

Keywords Armijo linesearch · Global convergence · Riemannian Newton method

Mathematics Subject Classification (2020) 90C30 · 49M15 · 65K05

1 Introduction

In this work, we propose an adapted Armijo linesearch for globalizing Newton method, to find a point $p \in \mathbb{M}$ satisfying

$$X(p) = 0, \tag{1}$$

where $X : \mathbb{M} \to T\mathbb{M}$ is a differentiable vector field, $\mathbb{M}$ is a finite dimensional Riemannian manifold, and $T\mathbb{M}$ is the tangent bundle of $\mathbb{M}$. Notice that finding singularities of gradient vector fields on Riemannian manifolds, which includes finding local minimizers, is a particular case to problem (1).

Marcio Antônio de Andrade Bortoloti
DCET/UESB, CP-95, CEP 45083-900-Vitória da Conquista, Bahia, Brazil
E-mail: mbortoloti@uesb.edu.br

Teles Araújo Fernandes
DCET/UESB, CP-95, CEP 45083-900-Vitória da Conquista, Bahia, Brazil
E-mail: telesfernandes@uesb.edu.br

In the 1970s, or earlier, iterative methods on manifolds arise in the context of optimizing a real-valued function (see [10]). In recent years, there has been a growing interest for the development of numerical algorithms for manifolds, see [3,15]. There are many numerical problems posed in manifolds arising in various natural contexts, for instance, finding the largest eigenvalue of a symmetric matrix may be posed as maximizing Rayleigh's quotient defined on a sphere, invariant subspace computations, and matrix diagonalization problems. These problems may be naturally posed on Riemannian manifolds. Hence, we can use the specific underlying geometric and algebraic structures to reduce significantly the computational cost of finding the zeros of a vector field. **It is well-known that Newton's method generetes a sequence that converges quadratically to a zero of a vector field. However, it a local convergence, by nature. In other words, Newton's method will converge to the zero of a vector field, when provided it a good enough starting guess, see [12,1,9,8]. Unfortunately, it is usual to expend significant computational effort to obtain a good enough starting guess. To overcome this drawback of Newton's method for** (1)**, a strategy have been introduced is to combine Newton's method with global methods for unconstrained optimization to produce a global method for** (1)**, (see [4]).** In cases where the objective function is twice continuously differentiable and strictly convex, the Newton direction is a descent one of the objective function. As a consequence, by adjusting the step size in the Newton direction using the Armijo rule, for instance, we can ensure global convergence of this method. This is a strategy of dumping the Newton step size to globalize this method which is known as damped Newton's method, see [7,2,5]. For a recent study, see [3,4]. **It is well-known that, if the Newton direction is not generated by selecting a starting point close to a solution, the step length, obtained by Armijo linesearch, may be quite small.** To fix this issue, in this paper we present an adapted Armijo linesearch in order to produce a satisfactory step length on the Newton direction, when it exists. Otherwise, we keep the classical Armijo linesearch on a safeguard direction for preventing the premature stopping of algorithm.

SERÁ ALTERADO APÓS NOVA SEÇÃO: To analyze the numerical behavior of the proposed algorithm, we have considered problems on Stiefel and sphere manifolds. The results show that the proposed algorithm has a good performance when compared with Algorithm 3 presented in [3].

The remainder of this paper is organized as follows. In the next section, we present some basic concepts and preliminary results, Section 3 presents the main algorithm and a convergence analysis. In Section 4, numerical experiments are presented, and finally, we conclude this paper with some comments.

2 Preliminaries

In this section, we briefly introduce notations, definitions, and auxiliary results we will use in the rest of the paper. Some basic concepts used here can be found in many introductory books on Riemannian geometry, for example, [6, 13]. We denote $\mathbb{M}$ as a *finite dimensional Riemannian manifold*, $T_p\mathbb{M}$ the *tangent space* of $\mathbb{M}$ at p, and $T\mathbb{M} = \bigcup_{p\in\mathbb{M}} T_p\mathbb{M}$ the *tangent bundle* of $\mathbb{M}$. The corresponding norm associated to the Riemannian metric $\langle \cdot\ ,\ \cdot\rangle$ is denoted by $\|\cdot\|$. We denote $\|\cdot\|_F$ as the Frobenious norm. The Riemannian distance between p and q in $\mathbb{M}$ is given by $d(p, q)$, which induces the original topology on $\mathbb{M}$, namely, $(\mathbb{M}, d)$ that represents a complete metric space. The open ball of radius $r > 0$ centred at p is defined as $B_r(p) := \{q \in \mathbb{M} : d(p, q) < r\}$. Let $\Omega \subseteq \mathbb{M}$ be an open set and denote by $\mathcal{X}(\Omega)$ the *space of differentiable vector fields* on Ω. Let ∇ be the Levi-Civita connection associated to $(\mathbb{M}, \langle\cdot\ ,\ \cdot\rangle)$. The covariant derivative of $X \in \mathcal{X}(\Omega)$, determined by ∇, defines at each $p \in \Omega$ a linear map $\nabla X(p) : T_p\mathbb{M} \to T_p\mathbb{M}$ given by $\nabla X(p)v = \nabla_Y X(p)$, where Y is a vector field such that $Y(p) = v$. For $f : \mathbb{M} \to \mathbb{R}$ a twice-differentiable function, the Riemannian metric, induces the mappings $f \mapsto \mathrm{grad} f$ and $f \mapsto \mathrm{Hess} f$. Its *gradient* and *hessian* can be associated to the Riemannian metric by

$$\langle \mathrm{grad} f, X\rangle := df(X), \qquad \langle \mathrm{Hess} f X, X\rangle := d^2 f(X, X), \qquad \forall\ X \in \mathcal{X}(\Omega),$$

respectively. The last two equalities imply $\mathrm{Hess} f X = \nabla_X \mathrm{grad} f$, for all $X \in \mathcal{X}(\Omega)$. The *norm of a linear map* $A : T_p\mathbb{M} \to T_p\mathbb{M}$ is defined by $\|A\| := \sup\{\|Av\|\ :\ v \in T_p\mathbb{M},\ \|v\| = 1\}$. A vector field V along a differentiable curve γ in $\mathbb{M}$ is said to be *parallel* iff $\nabla_{\gamma'} V = 0$. For each $t \in [a, b]$, the operator ∇ induces an isometry relative to Riemannian metric $\langle\cdot,\cdot\rangle$, $P_{\gamma,a,t} : T_{\gamma(a)}\mathbb{M} \to T_{\gamma(t)}\mathbb{M}$, defined by $P_{\gamma,a,t}\, v = V(t)$, where V is the unique vector field on γ such that $\nabla_{\gamma'(t)}V(t) = 0$ and $V(a) = v$. Further, we can notice $P_{\gamma,\,b_1,\,b_2} \circ P_{\gamma,\,a,\,b_1} = P_{\gamma,\,a,\,b_2}$ and $P_{\gamma,\,b,\,a} = P^{-1}_{\gamma,\,a,\,b}$. As long as there is no confusion, we will consider the notation P_{pq} instead of $P_{\gamma,\,a,\,b}$, when γ is the unique segment of curve joining p and q. In the next, we present a function which is used to establish the proposed algorithm.

Definition 1 Let $X : \mathbb{M} \to T\mathbb{M}$ be a differentiable vector field. The merit function is defined by

$$\varphi(p) = \frac{1}{2}\|X(p)\|^2. \tag{2}$$

The next definition introduces the concept of retraction, which has been introduced by [11]. It is a generalization of the classical exponential mapping.

Definition 2 A *retraction* on a manifold $\mathbb{M}$ is a smooth mapping R defined from the tangent bundle $T\mathbb{M}$ onto $\mathbb{M}$ with the following properties: If R_p denotes the restriction of R to $T_p\mathbb{M}$, then

(i) $R_p(0_p) = p$, where 0_p denotes the origin of $T_p\mathbb{M}$;
(ii) With the canonical identification $T_{0_p}T_p\mathbb{M} \simeq T_p\mathbb{M}$, $R'_p(0_p) = I_p$, where I_p is the identity mapping on $T_p\mathbb{M}$, and R'_p denotes the diferential of R_p.

Remark 1 Since $R'_p(0_p) = I_p$, by the Inverse Function Theorem, R_p is a local diffeomorphism (see [13, Cap. 2, Sec. 2]). Hence, we can define the *injectivity radius* of $\mathbb{M}$ at p with respect to R by

$$i_p := \sup\left\{r > 0 \ : \ R_{p|_{B_r(0_p)}} \text{ is a diffeomorphism}\right\},$$

where $B_r(0_p) := \{v \in T_p\mathbb{M} :\| v - 0_p \|< r\}$. In case $\bar{p} \in \mathbb{M}$. Then the above definition implies that if we consider δ such that $0 < \delta < i_{\bar{p}}$, then $V_\delta(\bar{p}) := R_{\bar{p}} B_\delta(0_{\bar{p}})$ is an open set. Moreover, for all $p \in V_\delta(\bar{p})$ the curve segment $\gamma_{\bar{p}p}(t) = R_{\bar{p}}\left(tR_{\bar{p}}^{-1}p\right)$ joining $\bar{p}$ to p belongs to $V_\delta(\bar{p})$.

The first auxiliary result of this work establishes that, if $\nabla X(\bar{p})$ is nonsingular then there exists a neighborhood of $\bar{p}$ such that ∇X is also nonsingular. Its proof can be found in [8, Lemma 3.2].

Lemma 1 *Consider that ∇X is a continuous operator at $\bar{p} \in \mathbb{M}$. Then, $\lim_{p\to\bar{p}} \|P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})\| = 0$. Moreover, if $\nabla X(\bar{p})$ is nonsingular then there exists $0 < \delta < \delta_{\bar{p}}$ such that, for each $p \in B_\delta(\bar{p})$, $\nabla X(p)$ is nonsingular and $\left\|\nabla X(p)^{-1}\right\| \leq 2\left\|\nabla X(\bar{p})^{-1}\right\|$.*

Consider the constant value δ given by Lemma 1. Denote $\bar{\delta} := \delta$, R a retraction, and X a differentiable vector field. We define *Newton's iterate mapping* as follows:

$$\begin{aligned} N_{R,X} : B_{\bar{\delta}}(\bar{p}) &\to \mathbb{M} \\ p &\mapsto R_p(-\nabla X(p)^{-1}X(p)). \end{aligned}$$

Lemma 2 *Let $\bar{p} \in \Omega$ such that $X(\bar{p}) = 0$. If ∇X is continuous at $\bar{p}$ and $\nabla X(\bar{p})$ is nonsingular, then there exists $0 < \hat{\delta} < \delta_{\bar{p}}$ such that, $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(\bar{p}) \subset \Omega$ and $\lim_{p\to\bar{p}} \varphi\left(N_{R,X}(p)\right) / \| X(p) \|^2 = 0$. As a consequence, there is a $\delta > 0$ such that, for all $\sigma \in (0, 1/2)$ and $\delta < \hat{\delta}$ it holds*

$$\varphi\left(N_{R,X}(p)\right) \leq \varphi\left(p\right) + \sigma\left\langle \text{grad } \varphi(p), -\nabla X(p)^{-1}X(p)\right\rangle, \qquad \forall\ p \in B_\delta(\bar{p}).$$

Proof See [3, Lemma 4.3].

We finish this section with the following result whose proof can be found in [3, Lemma 4.2].

Lemma 3 *If ∇X is continuous at $\bar{p} \in \Omega$ and $\nabla X(\bar{p})$ is nonsingular, then there exists $0 < \hat{\delta} < \delta_{\bar{p}}$ such that $B_{\hat{\delta}}(\bar{p}) \subset \Omega$, $\nabla X(p)$ is nonsingular for $p \in B_{\hat{\delta}}(\bar{p})$. Moreover, for all $p \in B_{\hat{\delta}}(\bar{p})$ the vector $v = -\nabla X(p)^{-1}X(p)$ is the unique solution of the linear equation $X(p) + \nabla X(p)v = 0$, and for $0 \leq \theta < 1/\operatorname{cond}(\nabla X(\bar{p}))$ we have*

$$\langle \text{grad } \varphi(p), v\rangle \leq -\theta \| \text{grad } \varphi(p)\|\|v\|, \qquad \forall\ p \in B_{\hat{\delta}}(\bar{p}).$$

3 Globalization of Newton Method

In this section, we present an adapted Armijo linesearch to globalize the Newton method to find a point $p \in \mathbb{M}$ satisfying $X(p) = 0$, where $X : \mathbb{M} \to T\mathbb{M}$ is a differentiable vector field. In [4,3], damped Newton methods were presented to solve this problem. In such works, the idea is to use the Newton method to minimize (2), taking its gradient as a safeguard direction, when Newton direction does not exist, and adjusting the step size by using Armijo linesearch for both directions. Using the same idea, we establish an adapted Armijo linesearch procedure to damping the Newton direction keeping the classical Armijo on safeguard direction. The following algorithm presents formally the strategy.

Algorithm 1:

Input: $\sigma \in (0, 1/2)$, $\theta \in [0, \gamma)$, $\alpha_{min} > 0$, and $p_0 \in \mathbb{M}$.

1 Set $k := 0$.

2 Compute $v_k \in T_{p_k}M$ such that

$$X(p_k) + \nabla X(p_k)v_k = 0. \tag{3}$$

3 If v_k exists, compute

$$\alpha_k := \max\left\{2^{-j} : \varphi\left(R_{p_k}(2^{-j}v_k)\right) \leq (1 + \sigma\theta 2^{1-j})\varphi\left(p_k\right), j \in \mathbb{N}\right\}. \tag{4}$$

Otherwise, go to 5.

4 If $\alpha_k \geq \alpha_{min}$ go to 6.

5 Set the search direction as

$$v_k = -\operatorname{grad} \varphi(p_k) = -\nabla X(p_k)^* X(p_k) \tag{5}$$

and compute the step size by the rule

$$\alpha_k := \max\left\{2^{-j} \ : \ \varphi\left(R_{p_k}(2^{-j}v_k)\right) \leq \varphi(p_k) - \sigma 2^{-j}\| \operatorname{grad} \varphi(p_k)\|^2, \ j \in \mathbb{N}\right\}. \tag{6}$$

6 Set $p_{k+1} := R_{p_k}(\alpha_k v_k)$.

Before presenting our convergence analysis for the proposed algorithm, we describe its characteristics. First, we compute v_k as a solution of (3) if any, and then we check if it satisfies (4) for some not too small step length. In this case, we use it as a search direction in **Step 6**. When v_k does not satisfy either (3) or $\alpha_k \geq \alpha_{min}$ we set $v_k = -\operatorname{grad} \varphi(p_k)$ as the search direction with the Armijo linesearch given by (6). Then, we set it to compute the next iterate in **Step 6**. We would like to point out that by Lemma 3, $\gamma = 1/\operatorname{cond}(\nabla X(\bar{p}))$.

Nos experimentos numéricos devemos informar como foi obtido o γ!!!

The next two results establish properties for the sequence generated by Algorithm 1. The first one assures the existence of the step size on the safeguard direction. Its proof can be found in [4, Lemma 3]. The second one assures the existence of the step size on the Newton direction.

Lemma 4 *Assume that $p \in \mathbb{M}$ with $X(p) \neq 0$. If $v \neq 0$ is given by* (5) *then* $\langle \text{grad } \varphi(p), v \rangle < 0$.

Lemma 5 *Let $\bar{p} \in \Omega \subset \mathbb{M}$ and R be a retraction. If ∇X is continuous at $\bar{p}$ and $\nabla X(\bar{p})$ is nonsingular then, there exists $0 < \hat{\delta} < \delta_{\bar{p}}$ such that $\nabla X(p)$ is nonsingular and the vector $v = -\nabla X(p)^{-1} X(p)$ is the unique solution of the linear equation $X(p) + \nabla X(p)v = 0$ for all $p \in B_{\hat{\delta}}(\bar{p})$. Moreover, if φ is given by* (2) *then, there are $\gamma > 0$, $\theta \in [0, \gamma)$, and $\sigma \in (0, 1/2)$ such that*

$$\varphi\left(R_p(v)\right) \leq (1 + 2\sigma\theta)\varphi\left(p\right), \qquad \forall p \in B_{\hat{\delta}}(\bar{p}).$$

Proof Since ∇X is continuous at $\bar{p} \in \Omega$ and $\nabla X(\bar{p})$ is nonsingular, Lemma 1 ensures the existence of a $\hat{\delta}$ such that $0 < \hat{\delta} < \delta_{\bar{p}}$ and $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(\bar{p})$. Moreover, the vector $v = -\nabla X(p)^{-1} X(p)$ is the unique solution of the linear equation $X(p) + \nabla X(p)v = 0$, for all $p \in B_{\hat{\delta}}(\bar{p})$. It follows from Lemma 2 that

$$\varphi\left(R_p(v)\right) \leq \varphi\left(p\right) + \sigma \left\langle \text{grad } \varphi(p), v \right\rangle, \qquad \forall\ p \in B_{\hat{\delta}}(\bar{p}).$$

Taking $\gamma = 1/\operatorname{cond}(\nabla X(\bar{p}))$, Lemma 3 ensures that for $\theta \in [0, \gamma)$ we have

$$\langle \text{grad } \varphi(p), v \rangle \leq -\theta \| \text{grad } \varphi(p) \| \|v\|, \qquad \forall\ p \in B_{\hat{\delta}}(\bar{p}).$$

Moreover, from Cauchy-Schwarz inequality and (2), we obtain

$$-\| \text{grad } \varphi(p) \| \|v\| \leq 2\varphi\left(p\right).$$

From three last inequalities, we conclude $\varphi\left(R_p(v)\right) \leq (1 + 2\sigma\theta)\varphi\left(p\right), \forall p \in B_{\hat{\delta}}(\bar{p})$.

We finish this section by presenting the following theorem which establishes convergence of a sequence generated by Algorithm 1. Taking into account Lemma 5, the proof of this theorem can be made using the same idea in [3, Theorem 4.2]. For this reason, we do not present it here.

Theorem 1 *Let $\mathbb{M}$ be a Riemannian manifold, R a retraction on $\mathbb{M}$, $X : \mathbb{M} \to T\mathbb{M}$ be a continuously differentiable vector field, and $\{p_k\}$ a sequence generated by Algorithm 1. Then, if $\bar{p} \in \mathbb{M}$ is an accumulation point of $\{p_k\}$, $\bar{p}$ is a critical point of φ. Moreover, if $\nabla X(\bar{p})$ is nonsingular, the convergence of $\{p_k\}$ to $\bar{p}$ is superlinear and $X(\bar{p}) = 0$.*

4 Numerical Experiments

In this section, we present two examples to show some numerical properties of Algorithm 1. The first one is a minimization problem on the Stiefel manifold that is related to applications in the independent component analysis (ICA). The second is an academic problem of finding a zero of a non-conservative vector field on the sphere. We compare the computational performance of Algorithm 1 with Algorithm 3, proposed in [3]. All problems were stopped when the norm of vector field reached the tolerance of 10^{-6}. When on a problem an algorithm has exceeded the maximum number of 2000 iterates we declare it has not been able to solve the problem. **We have taken the step lengths, given by** (4)**, always greater than** $\alpha_{min} = 10^{-5}$**.** All Matlab codes are freely available in the following web site: "https://github.com/mbortoloti/damped-newton-new-linesearch/".

4.1 A joint diagonalization problem on the Stiefel manifold

We analyse Algorithm 1 to solve the following joint diagonalization problem defined on the Stiefel manifold:

Problem 1 Given $A_1, \cdots, A_N$ symmetric matrices, find $P \in St(n,p) = \{P \in \mathbb{R}^{n\times p};\ P^TP = I_p\}$, with $p \leq n$, that minimizes

$$f(P) = -\sum_{l=1}^{N} \left\| \mathrm{diag}\left(P^T A_l P\right) \right\|_F^2,$$

where $\mathrm{diag}(Y)$ is a diagonal matrix generated by the diagonal entries of Y.

In [14] was proposed a Newton-type method using a vectorization strategy to solve this problem. Taking into acount this idea, we implement Algorithm 1 and perform numerical experiments considering, for each $l = 1, \cdots, N$, the matrices $A_l = PL_lP^T$, with $P \in St(n,n)$ and L_l a diagonal matrix whose non zero elements are given by $\lambda_1 > \cdots > \lambda_n$. In this case, an optimal solution is $X^* = PI_{n,p}$, where $I_{n,p}$ is the $n \times p$ identity matrix. Initial guesses were generated by considering $X_0 = \mathrm{qf}(X^* + \rho\ X_{rand})$ where, X_{rand} is a $n \times p$ random matrix and $\rho = 10^{-4}, 10^{-3}, \cdots, 10^2$. We assume $N = 5$. For $n = 10$ fixed and $p = 2, 3, 4, 5$ were considered 10 random initial guesses for each combination of (n,p). The sequences were obtained by using the retraction $R_P(V) = \mathrm{qf}(P + V)$, where qf denotes the Q factor of QR decomposition. We take $\sigma = 10^{-4}$ and $\theta = 0.9$. In Figure 1, it can be seen that Algorithm 1 is faster than Algorithm 3 proposed in [3].

4.2 A Non-Conservative Vector Field on the Sphere

In this section, we aim to explore the behaviour of Algorithm 1 applied to find a singularity of a non conservative vector field defined on the sphere. The problem is formally described as the following:

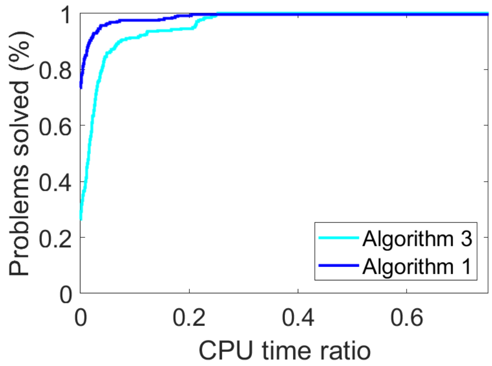

Fig. 1 Performance profile comparing the algorithms 1 and 3 (given in [3]) to solve the Problem 1.

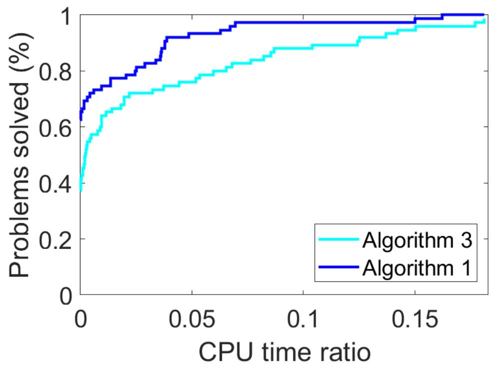

Fig. 2 Performance profile comparing the algorithms 1 and 3 to solve the Problem 2.

Problem 2 Find $p \in \mathbb{M}$ such that $X(p) = 0$, where $\mathbb{M} = (S^{n-1}, \langle , \rangle)$ is equipped with the canonical metric, $X(p) := Q(p - p^*) - \langle p, Q(p - p^*)\rangle p$ with Q a skew-symmetric matrix, and a fixed $p^* \in \mathbb{M}$, see [4].

Numerical study was developed for dimensions $n = 100, 200, 300, 400, 500$. For each dimension, we consider a fixed skew-symmetric matrix $Q = (A - A^T)/2$, where A is a randomly generated matrix. It were taken 15 random initial guesses on $\mathbb{M}$, for each dimension, and the parameters of linesearches were $\sigma = 10^{-3}$ and $\theta = 10^{-1}$. Performance profiles are depicted on Figure 2 comparing Algorithm 1 and Algorithm 3, proposed in [3]. We can notice Algorithm 1 solved all problems. On the other hand, Algorithm 3 did not solve all problems because the step length became quite small. This suggests Algorithm 1 has contributed to fix this issue.

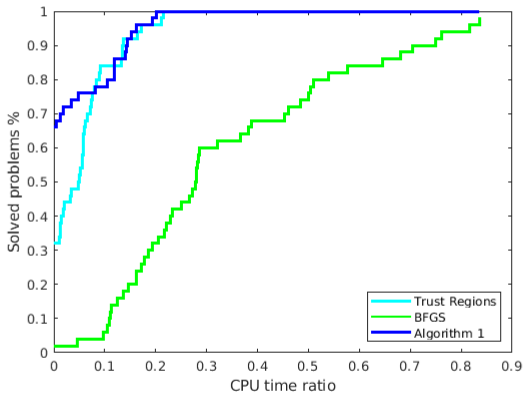

Fig. 3 Performance profiles comparing Algorithm 1, Trust Region and BFGS methods for the Problem 3 .

4.3 An academic function on Symmetric Positive Definite Matrices

In this section we present a performance of Algorithm 1 in the search for critical points of an academic function. We compared the performance of the proposed algorithm with the well-known BFGS, [xx] and Trust Region, [xx]. These algorithms were executed using the Matlab Tool Box Manopt,[xx].

To present the problem, we need consider $\mathbb{P}^n_{++}$, the manifold of symmetric positive definite matrices of order n,

Problem 3 Let $\mathbb{P}^n_{++}$ be the manifold of symmetric positive definite matrices of order n. Find $P^* \in \mathbb{P}^n_{++}$, such that $\operatorname{grad} f(P^*) = 0$, where $f : \mathbb{P}^n_{++} \to \mathbb{R}$ is given by $f(X) = a\log(\det X) + b\,\mathrm{trace}\,(X^{-1})$, with $a, b > 0$, $\det P$ is the determinant of P, and $\operatorname{grad} f(P)$ is the riemannian gradient of f.

For numerical study we have considered $a = 5.0$ and $b = 1.0$. It were considered tests for $n = 100, 200, 300, 400, 500$, and were taken 10 random initial guesses on $\mathbb{P}^n_{++}$ for each n. The parameters of linesearch were chosen as $\sigma = 10^{-3}$ and $\theta = 0.9$.

First, we compare the performance of Algorithm 1 with BFGS and Trust Region methods. It can be seen in Figure 3 that the proposed algorithm have solved all problems faster than the other considered methods.

Now, we present the gradient norm of function f against iteration numbers to compare Algorithm 1 with BFGS and Trust Region methods. It was considered the function f with $a = 10.0$ and $b = 1.0$ and $n = 500$. We take the same initial guess for the methods. In Figure 4 it can be seen that the proposed algorithm reached the desired tolerance for the gradient norm with fewer iterations than the Trust Region and BFGS methods.

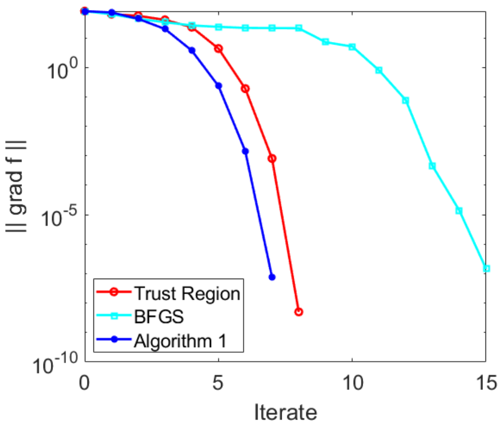

Fig. 4 Gradient norm as function of iteration number of Algorithm 1, BFGS and Trust Region for Problem 3 with $a = 10.0$, $b = 1.0$ and $n = 500$.

5 Conclusions

In this work, we have presented a strategy of linesearch to improve the globalization of Riemaniann Newton method, given in [3]. This strategy relaxes the Armijo rule for preventing a step length quite small on the Newton direction. From a numerical point of view, we would like to note the step length given in (4) appeared to be sensitive to the inverse of the conditioning number of the covariant derivative of considered vector field. It would be interesting to develop a study to obtain a further improvement of the Algorithm 1. Numerical experiments were developed on the Stiefel and sphere manifolds. They show that Algorithm 1 presents good performance when compared with Algorithm 3 presented in [3].

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ (2008). DOI 10.1515/9781400830244. URL http://dx.doi.org/10.1515/9781400830244
2. Bertsekas, D.P.: Constrained optimization and Lagrange multiplier methods. Computer Science and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London (2014)
3. Bortoloti, M.A.A., Fernandes, T.A., Ferreira, O.P.: An efficient damped Newton-type algorithm with globalization strategy on Riemannian manifolds. J. Comput. Appl. Math. **403**, Paper No. 113853, 15 (2022). DOI 10.1016/j.cam.2021.113853. URL https://doi.org/10.1016/j.cam.2021.113853
4. Bortoloti, M.A.A., Fernandes, T.A., Ferreira, O.P., Yuan, J.: Damped Newton's method on Riemannian manifolds. J. Global Optim. **77**(3), 643–660 (2020). DOI 10.1007/s10898-020-00885-0. URL https://doi.org/10.1007/s10898-020-00885-0
5. Burdakov, O.: Some globally convergent modifications of Newton's method for solving systems of nonlinear equations. In: Soviet mathematics-Doklady, vol. 22, pp. 376–378 (1980)

6. do Carmo, M.P.: Riemannian geometry. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA (1992). DOI 10.1007/978-1-4757-2201-7. URL http://dx.doi.org/10.1007/978-1-4757-2201-7. Translated from the second Portuguese edition by Francis Flaherty
7. Dennis Jr., J.E., Schnabel, R.B.: Numerical methods for unconstrained optimization and nonlinear equations, *Classics in Applied Mathematics*, vol. 16. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1996). DOI 10.1137/1.9781611971200. URL http://dx.doi.org/10.1137/1.9781611971200. Corrected reprint of the 1983 original
8. Fernandes, T.A., Ferreira, O.P., Yuan, J.: On the Superlinear Convergence of Newton's Method on Riemannian Manifolds. J. Optim. Theory Appl. **173**(3), 828–843 (2017). DOI 10.1007/s10957-017-1107-2. URL http://dx.doi.org/10.1007/s10957-017-1107-2
9. Ferreira, O.P., Silva, R.C.M.: Local convergence of Newton's method under a majorant condition in Riemannian manifolds. IMA J. Numer. Anal. **32**(4), 1696–1713 (2012). DOI 10.1093/imanum/drr048. URL http://dx.doi.org/10.1093/imanum/drr048
10. Luenberger, D.G.: The gradient projection method along geodesics. Management Sci. **18**, 620–631 (1972)
11. Manton, J.H.: Optimization algorithms exploiting unitary constraints. IEEE Trans. Signal Process. **50**(3), 635–650 (2002). DOI 10.1109/78.984753. URL https://doi.org/10.1109/78.984753
12. Ortega, J.M.: Numerical analysis. A second course. Academic Press, New York-London (1972). Computer Science and Applied Mathematics
13. Sakai, T.: Riemannian geometry, *Translations of Mathematical Monographs*, vol. 149. American Mathematical Society, Providence, RI (1996). Translated from the 1992 Japanese original by the author
14. Sato, H.: Riemannian Newton-type methods for joint diagonalization on the Stiefel manifold with application to independent component analysis. Optimization **66**(12), 2211–2231 (2017). DOI 10.1080/02331934.2017.1359592. URL https://doi.org/10.1080/02331934.2017.1359592
15. Sato, H.: Riemannian Optimization and Its Applications. Springer Nature (2021)