Information and Inference: A Journal of the IMA (2023) **00**, 1–16
https://doi.org/DOI HERE

A Worst-case Complexity Analysis for Riemannian Non-monotone Line-search Methods

HARRY OVIEDO*
Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Av. Diag. Las Torres 2640, 7941169, Metropolitan Region, Chile
*H. Oviedo: harry.oviedo@uai.cl

[]

In this paper we deal with non-monotone line-search methods to minimize a smooth cost function on a Riemannian manifold. In particular, we study the number of iterations necessary for this class of algorithms to obtain ε-approximated stationary points. Specifically, we prove that under a regularity Lipschitz-type condition on the pullbacks of the cost function to the tangent spaces of the manifold and other mild assumptions, the Riemannian non-monotone line-search methods generates points with Riemannian gradient norm smaller than ε in $\mathscr{O}(\varepsilon^{-2})$ iterations. Our worst-case complexity includes a wide variety of known non-monotone strategies existing in the literature. Additionally, we establish the global convergence for this family of methods. The bounds obtained in our analysis agree with the bounds known for line-search methods in the field of unconstrained nonlinear optimization and hence generalizes previous work.

Keywords: Riemannian optimization; Non-monotone line search; Worst-case complexity; Global convergence.

1. Introduction

We address the problem of computing a local minimum of a cost function on a Riemannian manifold, which can be modeled as

$$\min_{x \in M} F(x), \tag{1.1}$$

where $F : M \to \mathbb{R}$ is a continuously differentiable function and the pair $(M, \langle \cdot, \cdot \rangle)$ forms a Riemannian manifold with associated Riemannian metric $\langle \cdot, \cdot \rangle$. For now on, the local inner product $\langle \cdot, \cdot \rangle_x$ represents the restriction of the Riemannian metric to the tangent space T_xM, for all $x \in M$. In addition, given $x \in M$ an arbitrary point, $\nabla_M F(x)$ will denotes the Riemannian gradient of the objective function F evaluated at x. Solving this problem for some particular objective functions and manifolds is very useful in some applications as for example in dictionary learning [4, 27], Brockett cost minimization [4], simultaneous localization and mapping [4], Low-rank nearest correlation estimation [6], Max cut [6], Cryo-EM [6], Deep learning [6], independent component analysis [28], matrix completion [26, 29, 30], data clustering and image segmentation [31, 32], among others.

Note that Problem (1.1) reduces to the well-known unconstrained non-linear optimization problem [2] when M is an Euclidean space, as for example when $M = \mathbb{R}^n$. Nonetheless, in the more general case, the presence of the manifold constraint makes (1.1) a challenging problem, because the set M is typically nonconvex and hence (1.1) could have multiple minimizers.

In order to deal with this problem, we consider an important family of numerical procedures so-called the *Riemannian line-search methods* [1, 4]. This family of methods includes among its members, for example, the Riemannian gradient descent method, the Riemannian quasi-Newton methods, the Riemannian Newton method, some Riemannian conjugate gradient methods, among others. In this paper, we are interesting in develop the worst-case complexity analysis for this rich collection of methods. The goal of this kind of rigorous study, is to establish an estimate of the maximum number of iterations that a certain algorithm requires to obtain a ε-stationary point of the objective function, in the Riemannian sense.

In the special case when M is an Euclidean space, there are several works that have investigated the worst-case complexity analysis of the line-search methods, we suggest the reader see [5, 13, 16]. Recently in [5] Grapiglia and Sachs established the worst-complexity analysis of a generalized family of non-monotone line-search schemes in the Euclidean context. Essentially, in [5] it is proved that this family of methods computes a point $x \in M = \mathbb{R}^n$ with $||\nabla F(x)|| \leq \varepsilon$ in $\mathcal{O}(\varepsilon^{-2})$ iterations, under some assumptions. Nevertheless, in the Riemannian setting, there is little literature that has carried out this type of analysis. Much of the existing theory related to manifold constrained optimization is primarily dedicated to establishing the global convergence to stationary point of some iterative methods, see [1, 4, 6, 7, 17]. Here the word "global" refers to the fact that the convergence result does not depend on the particular selection of the starting point with which the numerical method begins to iterate.

In the seminal paper [12], Boumal et. al. studied the global rates of convergence for the gradient method and the trust-region method in the Riemannian framework. In particular, the authors in [12] demonstrated that the Riemannian gradient method generates points $x \in M$ with $||\nabla_M F(x_k)||_{x_k} \leq \varepsilon$ in $\mathcal{O}(\varepsilon^{-2})$ iterations for constant step-size, and also for step-sizes determined with the backtracking heuristic equipped with the Armijo rule. Additionally, in [12] the authors proved that the number of iterations required by the Riemannian trust-region method to compute an ε-critical point is of order $\mathcal{O}(\varepsilon^{-3})$. The theoretical results obtained in [12] correspond to monotone methods, that is, they construct a sequence of feasible points $\{x_k\} \subset M$ such that their associated sequence of objective values $\{F(x_k)\}$ is monotonically decreasing. Furthermore, excluding the results related to the Riemannian trust-region method, the study developed in [12] only considers the search direction $d_k = -\nabla_M F(x_k)$, and thus the analysis is very limited.

In this paper, we consider a whole family of Riemannian non-monotone line-search methods and analyze their complexity and global convergence. To carry out this study, we consider a very general algorithmic framework, which involves generic search directions $d_k \in T_{x_k}M$, and generic non-monotone terms. So, the designed algorithm has the capacity to generate sequences $\{x_k\} \subset M$ such that their corresponding sequence of objective values $\{F(x_k)\}$ is non-monotone, which differs from [12]. We prove that the number of iterations required by the proposed non-monotone algorithm to return a point satisfying the relation $||\nabla_M F(x_k)||_{x_k} \leq \varepsilon$ is $\mathcal{O}(\varepsilon^{-2})$, which coincides with the known bound for the Euclidean case [5, 13], and matches with the same bound obtained in [12] for the Riemannian gradient method. Therefore, our analysis can be regarded as an extension of the results presented in [5, 12].

This manuscript is organized as follows. In Section 2, we present a generic Riemannian non-monotone line-search method which will be the object of study. The worst-case complexity estimates are given in Section 3. Section 4 is dedicated to the global convergence analysis of the proposed algorithm. Some conclusions are provided in Section 5.

2. Non-monotone line-search algorithm

The main tool for extending the line-search methods from the Euclidean to the Riemannian context is the mapping called *retraction*. The rigorous definition of retraction appears in [1]. This class of mechanisms allows us to map vectors from the tangent space T_xM of the manifold at $x \in M$ to points on M. Particularly, the Riemannian line-search algorithms construct a sequence of iterates $\{x_k\} \subset M$ using the following recursive scheme, starting with a given $x_0 \in M$,

$$x_{k+1} = R_{x_k}(\tau_k d_k), \quad \forall k, \tag{2.1}$$

where $R_{x_k} : T_{x_k}M \to M$ is a retraction, $d_k \in T_{x_k}M$ is a tangent descent direction and $\tau_k = \alpha_k \delta^l > 0$ is the step-size, where l is some non-negative integer, $\delta \in (0,1)$ is a constant selected by the user and $\alpha_k > 0$ for all $k \geq 0$. In the Euclidean case, that is, when $M = \mathbb{R}^n$, the usual retraction is $R_{x_k}(\tau_k d_k) = x_k + \tau_k d_k$. In this context the complexity analysis of the non-monotone line-search methods has been studied in [5, 13]. In general, the total amount of work required by non-monotone line-search methods, in the worst case, to attain a prescribed tolerance $\varepsilon > 0$ is of the type $\mathscr{O}(\varepsilon^{-2})$. However, the results presented in all the works [5, 13] do not apply to minimization problems on a Riemannian manifold.

The terminus *non-monotone*, in this class of algorithms, refers to the fact that the step-size $\tau_k = \alpha_k \delta^l$ is selected by satisfying some non-monotone condition related to the objective function. One of the most used rules for selecting τ_k in (2.1) is the non-monotone line-search condition of Zhang and Hager [15], who propose to select $\tau_k = \alpha_k \delta^l$ choosing l as the smallest non-negative integer such that the condition

$$F(R_{x_k}(\alpha_k \delta^l d_k)) \leq C_k + c_1 \alpha_k \delta^l \langle \nabla_M F(x_k), d_k \rangle_{x_k}, \tag{2.2}$$

is satisfied for all $k \geq 0$, where $C_k = \frac{\eta_{k-1} Q_{k-1} C_{k-1} + F(x_k)}{Q_k}$, for all $k \geq 1$, with $Q_k = \eta_{k-1} Q_{k-1} + 1$, $\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $Q_0 = 1$ and $C_0 = F(x_0)$. In fact, in [15] this strategy is introduced for the Euclidean case. The condition (2.2) corresponds to its extension to the Riemannian setting. This rule has been used in many methods for optimization on manifolds, see [6, 7, 8, 9, 10, 11].

The Zhang and Hager's non-monotone rule can be generalized as

$$F(R_{x_k}(\alpha_k \delta^l d_k)) \leq C_k + c_1 \alpha_k \delta^l \langle \nabla_M F(x_k), d_k \rangle_{x_k}, \tag{2.3}$$

for all $k \geq 0$, where C_k is the convex combination between C_{k-1} and $F(x_k)$, that is, $C_k = \theta_k C_{k-1} + (1 - \theta_k) F(x_k)$, where $\theta_k \in [0,1]$ for all k with $\{\theta_k\}$ convergent. In fact, if we choose $\theta_k = \frac{\eta_{k-1} Q_{k-1}}{Q_k}$ in (2.3) then we recover the Zhang and Hager's non-monotone rule (2.2).

Another non-monotone condition widely used in the literature is the Grippo's non-monotonic strategy [18]. This strategy proposes computing $\tau_k = \alpha_k \delta^l$ with l being the smallest non-negative integer such that

$$F(R_{x_k}(\alpha_k \delta^l d_k)) \leq \max_{0 \leq i \leq m(k)} [F(x_{k-i})] + c_1 \alpha_k \delta^l \langle \nabla_M F(x_k), d_k \rangle_{x_k}, \tag{2.4}$$

is verified for all $k \geq 0$, where $m(0) = 0$ and $0 \leq m(k) \leq \min\{m(k-1)+1, N\}$ for $N \in \mathbb{N}$ fixed. This technique has been used in [19, 20], for solving Riemannian constrained optimization problems.

We can also consider more basic non-monotone rules, such as

$$F(R_{x_k}(\alpha_k\delta^l d_k)) \leq T_k + c_1\alpha_k\delta^l\langle\nabla_M F(x_k), d_k\rangle_{x_k}, \tag{2.5}$$

for all $k \geq 0$, where $T_k = \frac{M}{k+1}$ with $M > 0$ is a constant. There are other non-monotone strategies available in the Euclidean case, for example see [21, 22, 23, 24], which can be easily generalized to the Riemannian context. We do not list them here because they have not been widely used in the solution of optimization problems on manifolds. However, our analysis includes these techniques.

Now, we present a generic line-search algorithm to deal with optimization problems on Riemannian manifolds.

Algorithm 1 Riemannian Non-monotone Line–Search Method

Require: $x_0 \in M$, $\alpha_0 > 0$, $\delta, c_1 \in (0,1)$, $k := 0$.

1: Choose a non-null tangent descent direction $d_k \in T_{x_k}M$.

2: Select the non-monotone term $\nu_k \geq 0$.

3: Compute $l \geq 0$ as the smallest non-negative integer such that

$$F(R_{x_k}(\alpha_k\delta^l d_k)) \leq (F(x_k) + \nu_k) + c_1\alpha_k\delta^l\langle\nabla_M F(x_k), d_k\rangle_{x_k} \tag{2.6}$$

and

$$\alpha_k\delta^l||d_k||_{x_k} \leq \varphi_k. \tag{2.7}$$

4: Set $l_k = l$; $x_{k+1} = R_{x_k}(\alpha_k\delta^{l_k}d_k)$; $\alpha_{k+1} = \alpha_k\delta^{l_k-1}$; $k \leftarrow k+1$; and to Step 1.

Given a Riemannian manifold M, it is not always possible to build a retraction defined on the entire tangent bundle $TM := \cup_{x\in M}T_xM$. For example, for the Symplectic Stiefel manifold $M \equiv Sp(2p,2n) := \{X \in \mathbb{R}^{2n,2p} : X^\top J_{2n}X = J_{2p}\}$, the retraction based on the Cayley transform developed in [25], is only defined for tangent vectors in a certain neighborhood of the null vector $0_x \in T_xM$, for details see [25]. For this reason, in Algorithm 1, we are assuming that the retraction $R_{x_k}(\cdot)$ that appears in (2.6) is only defined locally, that is, for each k there exists a radius $\varphi_k > 0$ such that $R_{x_k} : T_{x_k}M \cap \mathbb{B}(0_{x_k}, \varphi_k) \to M$ is well-defined. Here $\mathbb{B}(0_{x_k}, \varphi_k)$ represents the closed ball $\mathbb{B}(0_{x_k}, \varphi_k) = \{z_k \in T_{x_k}M : ||z_k||_{x_k} \leq \varphi_k\}$, for all $k \geq 0$. Additionally, when $\varphi_k = \infty$ for all k, we say that the retraction $R(\cdot)$ is defined globally.

In addition, in the algorithm above, ν_k is the term that controls the degree of non-monotonicity of the proposed algorithm. Note that this term depends on k, which means that we can adjust it at each iteration. Thus, Algorithm 1 is very flexible since it can be implemented with different non-monotone terms, and therefore it is not only limited to use the non-monotone strategy of Zhang and Hager or another of the existing ones.

On the other hand, observe that by selecting $\nu_k = C_k - F(x_k)$, in Algorithm 1, we recover the Zhang and Hager's non-monotone rule (2.3). Similarly, choosing $\nu_k = \max_{0\leq i\leq m(k)}[F(x_{k-i})] - F(x_k)$, we obtain the Grippo's non-monotone condition (2.3). The same argument applies to the rest of the non-monotone strategies.

3. Complexity analysis

In this section we present the complexity analysis of Algorithm 1 under the assumption $\varphi_{\inf} := \inf_k \varphi_k > 0$. As mentioned in [12], if the injectivity radius of the manifold is positive then there exist retractions verifying the relation $\inf_{x\in M} \varphi(x) > 0$, which is the case when M is a compact manifold, see [14].

In order to carry out the worst-case complexity analysis of Algorithm 1, we require the assumptions listed below.

- **A1**: [Lower bound]. There exists a constant $F_{low} \in \mathbb{R}$ such that $F(x) \geq F_{low}$, for all $x \in M$.
- **A2**: [Restricted Lipschitz-type gradient for pullbacks]. There exists $L_g \geq 0$ such that, for all iterate x_k generated by Algorithm 1, the pullback $\hat{F}_k(\cdot) := F \circ R_{x_k}(\cdot)$ verifies

$$| \hat{F}_k(\xi) - (F(x_k) + \langle \nabla_M F(x_k), \xi \rangle_{x_k}) | \leq \frac{L_g}{2} ||\xi||^2_{x_k}, \tag{3.1}$$

  for all $\xi \in T_{x_k}M$ such that $||\xi||_{x_k} \leq \varphi_k$.
- **A3**: [Direction properties]. There exists three constants $c_2 \in (0,1]$, $0 < c_3 < c_4 < \infty$ such that the tangent search direction z_k satisfies

$$\langle \nabla_M F(x_k), d_k \rangle_{x_k} \leq -c_2 ||\nabla_M F(x_k)||_{x_k} ||d_k||_{x_k}, \quad \text{and} \quad c_3 ||\nabla_M F(x_k)||_{x_k} \leq ||d_k||_{x_k} \leq c_4 ||\nabla_M F(x_k)||_{x_k}, \tag{3.2}$$

  for all $k \geq 0$.
- **A4**: There exists a constant $D > 0$ such that

$$||d_k||_{x_k} \leq D, \quad \forall k \geq 0. \tag{3.3}$$

Additionally, we will consider an extra assumption related to the sequence ν_k that calibrates the degree of the non-monotonicity of the Algorithm 1.

- **A5**: $\lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \nu_k = 0$.

It is straightforward to proof that if the sequence $\{\nu_k\}$ is such that $\sum_{k=0}^{\infty} \nu_k < \infty$ then **A5** is verified. Nevertheless, the assumption **A5** can be satisfied for sequences $\{\nu_k\}$ such that its corresponding series is not convergent. For example, let $\nu_k = \frac{M}{k+1}$ for all $k \geq 0$, then observe that this sequence satisfy **A5** but $\sum_{k=0}^{\infty} \nu_k$ is divergent. In addition, observe that if M is a compact manifold then **A3** implies **A4**.

As mentioned in [12], if M is a compact Riemannian manifold embedded in some Euclidean space $\mathscr{E}$, $R_{x_k}(\cdot)$ is a retraction globally defined and $F : \mathscr{E} \to \mathbb{R}$ has Lipschitz continuous gradient in the convex hull of M, then the composition $\hat{F}_k(\cdot) = F \circ R_{x_k}(\cdot)$ verifies the regularity condition **A2** for any sequence of point $\{x_k\} \subset M$ and with $\varphi_k = \infty$, for all k. This result is stated in Lemma 4 [12], and is valid for several Riemannian manifold that arise frequently in applications, such as the Stiefel manifold, the oblique manifold, the Grassmannian manifold, the unit sphere, among others.

The abovementioned Lipschitz continuity type assumption **A2** gives us an upper bound for the composition $\hat{F}_k(\cdot)$ for all k:

$$F(R_{x_k}(\xi)) \leq F(x_k) + \langle \nabla_M F(x_k), \xi \rangle_{x_k} + \frac{L_g}{2} ||\xi||^2_{x_k}, \tag{3.4}$$

for all $\xi \in T_{x_k}M \cap \mathbb{B}(0_{x_k}, \varphi_k)$.

In addition, according to the assumption **A3**, it follows easily that

$$\langle \nabla_M F(x_k), d_k \rangle_{x_k} \leq -c_5 ||d_k||^2_{x_k}, \tag{3.5}$$

where $c_5 = (c_2 c_3)/(c_4^2) > 0$.

The following lemma establishes that it is always possible to find values $t > 0$ such that (2.6) is satisfied and $td_k \in \mathrm{Dom}(R_{x_k})$.

Lemma 1. *Suppose that the assumptions **A2-A4** hold. Then the properties*

$$F(R_{x_k}(td_k)) \leq (F(x_k) + \nu_k) + c_1 t \langle \nabla_M F(x_k), d_k \rangle_{x_k}, \tag{3.6}$$

and

$$td_k \in \mathrm{Dom}(R_{x_k}), \tag{3.7}$$

are fulfilled, for all $t \in (0, \tilde{\alpha}]$. *Here* $\tilde{\alpha} := \min\{\frac{2(1-c_1)}{L_c}, \frac{\varphi_{inf}}{D}\}$, *where* $L_c = \frac{L_g}{c_5} > 0$.

Proof Firstly, let $t \in (0, \tilde{\alpha}]$ be an arbitrary value in this interval. So, we have $0 < t \leq \tilde{\alpha} \leq \frac{\varphi_{\inf}}{D}$. Now, using **A4** we get

$$\begin{aligned} ||td_k||_{x_k} &= t||d_k||_{x_k} \\ &\leq \frac{\varphi_{\inf}}{D}||d_k||_{x_k} \\ &\leq \frac{\varphi_k}{D}||d_k||_{x_k} \\ &\leq \varphi_k, \end{aligned}$$

this means that $td_k \in \mathrm{Dom}(R_{x_k})$ for each $t \in (0, \tilde{\alpha}]$. Moreover, it follows from assumption **A2** that

$$F(R_{x_k}(td_k)) \leq F(x_k) + t\langle \nabla_M F(x_k), d_k \rangle_{x_k} + \frac{L_g}{2} t^2 ||d_k||^2_{x_k}, \quad \forall t \in (0, \tilde{\alpha}]. \tag{3.8}$$

By using the assumption **A3**, which implies that (3.5), in (3.8) we arrive at

$$\begin{aligned} F(R_{x_k}(td_k)) &\leq F(x_k) + t\langle \nabla_M F(x_k), d_k \rangle_{x_k} - \frac{L_c}{2} t^2 \langle \nabla_M F(x_k), d_k \rangle_{x_k} \\ &= F(x_k) + \left[1 - \frac{L_c}{2} t\right] t \langle \nabla_M F(x_k), d_k \rangle_{x_k}, \end{aligned} \tag{3.9}$$

which is valid for all $t \in (0, \tilde{\alpha}]$. Additionally, since $0 < t \leq \tilde{\alpha} \leq \frac{2(1-c_1)}{L_c}$, we have that

$$1 - \frac{L_c}{2} t \geq c_1. \tag{3.10}$$

Substituting (3.10) in (3.9) and keeping in mind that $\langle \nabla_M F(x_k), d_k \rangle_{x_k} < 0$, we obtain

$$\begin{aligned} F(R_{x_k}(td_k)) &\leq F(x_k) + c_1 t \langle \nabla_M F(x_k), d_k \rangle_{x_k} \\ &\leq (F(x_k) + \nu_k) + c_1 t \langle \nabla_M F(x_k), d_k \rangle_{x_k}, \end{aligned}$$

for all $t \in (0, \tilde{\alpha}]$, where the last inequality is a consequence of the fact that $\nu_k \geq 0$. □

Remark 1. *The previous lemma implicitly establishes the well definition of Algorithm 1. In fact, from Lemma 1 we know that (3.6) and (3.7) are verified for any $t \in (0, \tilde{\alpha}]$. Moreover, since $\delta \in (0,1)$, it is always possible to find an integer $l \geq 0$ such that $\alpha_k \delta^l \in (0, \tilde{\alpha}]$, and therefore Algorithm 1 can not cycle in Step 3.*

Remark 2. *When the retraction R_{x_k} is globally defined, the condition (3.7) is always satisfied for any $t \in \mathbb{R}$. In this case, Lemma 1 is true replacing $\tilde{\alpha}$ by with $\bar{\alpha} = \frac{2(1-c_1)}{L_c}$ and without assuming **A4**.*

A lower bound for the sequence $\{\alpha_k\}$ is provided by the lemma below.

Lemma 2. *Suppose that the assumptions **A2**-**A4** hold. Then,*

$$\alpha_k \geq \underline{\alpha} := \min\{\alpha_0, \tilde{\alpha}\}, \tag{3.11}$$

for all $k \geq 0$, where $\tilde{\alpha}$ is the constant that appears in Lemma 1.

Proof The proof is by induction. Clearly the present lemma is valid for $k = 0$. Now, let us suppose that the inequality (3.11) is fulfilled for some $k \geq 0$. Notice that if $l_k = 0$ then Step 4 and the inductive hypothesis lead to

$$\alpha_{k+1} = \delta^{l_k - 1} \alpha_k = \delta^{-1} \alpha_k > \alpha_k \geq \underline{\alpha},$$

and thus (3.11) is verified for $k+1$, in this particular case. So, to complete the proof, it is enough to show that α_{k+1} satisfied (3.11) when $l_k \geq 1$.

Observe that if $l_k \geq 1$ then we have two possibilities, either $\alpha_k \delta^{l_k} \in (0, \tilde{\alpha}]$ or $\alpha_k \delta^{l_k} > \tilde{\alpha}$. Let us analyze the first case:

$$\alpha_k \delta^{l_k} \in (0, \tilde{\alpha}]. \tag{3.12}$$

Since $\delta \in (0,1)$, from Step 4, we have

$$\alpha_{k+1} = \alpha_k \delta^{l_k - 1} > \alpha_k \delta^{l_k} > 0. \tag{3.13}$$

In view of (3.12) and (3.13) we have again two cases: either $\alpha_{k+1} \in (\alpha_k \delta^{l_k}, \tilde{\alpha}]$ or $\alpha_{k+1} > \tilde{\alpha}$. We affirm that the first of these two cases is false. Indeed, if $\alpha_{k+1} \in (\alpha_k \delta^{l_k}, \tilde{\alpha}] \subset (0, \tilde{\alpha}]$ then by Lemma 1 we known that (3.6) and (3.7) hold for $t := \alpha_{k+1} = \alpha_k \delta^{l-1}$. Keeping in mind that $l - 1 \geq 0$, then we have found a non-negative integer "$l-1$", which is less than l, such that $\alpha_{k+1} = \alpha_k \delta^{l-1}$ satisfies (2.6)-(2.7), but this contradicts the definition of l in Algorithm 1. Therefore the unique valid possibility is that $\alpha_{k+1} > \tilde{\alpha}$.

Note that in this case (3.11) is verified for $k+1$ trivially.

The remaining case to analyze is when $\alpha_k\delta^{l_k} > \tilde{\alpha}$ with $l_k \geq 1$. It follows from the fact $\delta \in (0,1)$ and Step 4 that

$$\alpha_{k+1} = \alpha_k\delta^{l_k-1} > \alpha_k\delta^{l_k} > \tilde{\alpha} \geq \underline{\alpha},$$

thus in this final case the property (3.11) is also valid for $k+1$. □

The following technical result establishes an upper bound associated with the total number of function evaluations. Its demonstration is similar to the proof of Theorem 3 in [13], and we briefly write the proof here to make the paper self-contained.

Lemma 3. *Suppose that the assumptions **A2-A4** are fulfilled at each iteration. In Algorithm 1, let's denote by N_k the total number of function evaluations up to the k-th iteration. Then,*

$$N_k \leq 2(k+1) + \frac{1}{\log(\delta)}(\log(\underline{\alpha}) - \log(\alpha_0)),$$

where $\underline{\alpha}$ is the constant that appears in the previous lemma.

Proof Theorem 3 in [13] applies here, since the proof only uses $\alpha_{k+1} - \delta^{l_k-1}\alpha_k = 0$ and the relation $\alpha_k \geq \underline{\alpha}$, for all k. □

For the rest of the analysis, we define

$$\kappa_c := \min\left\{c_1c_2c_3\delta\underline{\alpha}, \frac{2\delta c_1c_2c_3c_5(1-c_1)}{L_g}\right\}. \tag{3.14}$$

Observe that, given $\varepsilon > 0$, under the assumption **A5**, we will denote by $K(0,\varepsilon)$ to any non-negative integer such that

$$K \geq K(0,\varepsilon) \quad \Rightarrow \quad \frac{1}{K}\sum_{k=0}^{K-1}\nu_k \leq \frac{1}{2}\kappa_c\varepsilon^2, \tag{3.15}$$

where κ_c is the constant defined in (3.14).

The theorem below states an upper bound on the number of iterations necessary by Algorithm 1 to achieve the precision $||\nabla_M F(x_k)||_{x_k} \leq \varepsilon$ in the residual.

Theorem 1. *Let $\{x_k\}$ be a sequence generated by Algorithm 1 and $\varepsilon > 0$. Suppose that the assumptions **A1–A5** hold and let K be a non-negative integer such that*

$$K \geq \max\left\{K(0,\varepsilon), \frac{2(F(x_0) - F_{low})}{\kappa_c\varepsilon^2}\right\}. \tag{3.16}$$

Then,

$$\min_{k\in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon,$$

where $I_K = \{0,1,2,\ldots,K-1\}$.

Proof Using the Lipschitz-type assumption **A2**, **A3** and Lemma 2, we obtain

$$\begin{aligned} \nu_k + [F(x_k) - F(x_{k+1})] &\geq -c_1 \alpha_k \delta^{l_k} \langle \nabla_M F(x_k), d_k \rangle_{x_k} \\ &\geq c_1 c_2 c_3 \delta \alpha_k \delta^{l_k - 1} ||\nabla_M F(x_k)||^2_{x_k} \\ &= c_1 c_2 c_3 \delta \alpha_{k+1} ||\nabla_M F(x_k)||^2_{x_k} \\ &\geq c_1 c_2 c_3 \delta \underline{\alpha} ||\nabla_M F(x_k)||^2_{x_k} \\ &\geq \kappa_c ||\nabla_M F(x_k)||^2_{x_k}. \end{aligned}$$

Summing up both sides of this last inequality from $k = 0$ to $k = K - 1$ and, we have

$$\begin{aligned} \sum_{k=0}^{K-1} \kappa_c ||\nabla_M F(x_k)||^2_{x_k} &\leq \sum_{k=0}^{K-1} [f(x_k) - f(x_{k+1})] + \sum_{k=0}^{K-1} \nu_k \\ &\leq F(x_0) - F(x_K) + \sum_{k=0}^{K-1} \nu_k \\ &\leq F(x_0) - F_{low} + \sum_{k=0}^{K-1} \nu_k, \end{aligned}$$

leading to

$$\kappa_c K \left(\min_{k \in I_K} ||\nabla_M F(x_k)||^2_{x_k} \right) \leq F(x_0) - F_{low} + \sum_{k=0}^{K-1} \nu_k.$$

The inequality above gives us

$$\min_{k \in I_K} ||\nabla_M F(x_k)||^2_{x_k} \leq \frac{F(x_0) - F_{low}}{K \kappa_c} + \frac{1}{\kappa_c} \frac{1}{K} \sum_{k=0}^{K-1} \nu_k. \tag{3.17}$$

Since the index K satisfies (3.16), we immediately have $K \geq K(0, \varepsilon)$. Hence, by applying (3.15) in (3.17) we get

$$\min_{k \in I_K} ||\nabla_M F(x_k)||^2_{x_k} \leq \frac{F(x_0) - F_{low}}{K \kappa_c} + \frac{\varepsilon^2}{2}. \tag{3.18}$$

On the other hand, again in view of (3.16) we obtain

$$\frac{F(x_0) - F_{low}}{K \kappa_c} \leq \frac{\varepsilon^2}{2}.$$

Combining this last result with (3.18), we conclude that

$$\min_{k \in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon. \tag{3.19}$$

This completes the proof. □

Below we present a consequence of Theorem 1.

Corollary 1. *Assume that **A1–A4** are fulfilled and that the series $\sum_{k=0}^{\infty} \nu_k$ is convergent. Let $\{x_k\}$ be a sequence generated by Algorithm 1 and $\varepsilon > 0$. If*

$$K \geq \frac{2}{\kappa_c \varepsilon^2} \left(\max \left\{ \sum_{k=0}^{\infty} \nu_k, F(x_0) - F_{low} \right\} \right), \tag{3.20}$$

then

$$\min_{k \in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon,$$

with $I_K := \{0, 1, 2, \ldots, K-1\}$.

Proof From the iteration procedure, we have that

$$0 \leq \frac{1}{K} \sum_{k=0}^{K-1} \nu_k \leq \frac{1}{K} \sum_{k=0}^{\infty} \nu_k, \quad \forall K \geq 1.$$

Since $\sum_{k=0}^{\infty} \nu_k$ is a convergent series, then we must have that **A5** hold. Furthermore, from the hypothesis (3.20) we know that

$$K \geq \frac{2}{\kappa_c \varepsilon^2} \left(\sum_{k=0}^{\infty} \nu_k \right),$$

which implies that

$$\frac{\kappa_c \varepsilon^2}{2} \geq \frac{1}{K} \sum_{k=0}^{\infty} \nu_k \geq \frac{1}{K} \sum_{k=0}^{K-1} \nu_k.$$

This means that if we choose $K(0,\varepsilon) = 2\kappa_c^{-1} \varepsilon^{-2} \left(\sum_{k=0}^{\infty} \nu_k \right)$ then the condition (3.15) holds. So, the inequality (3.20) can be formulated as:

$$K \geq \left(\max \left\{ K(0,\varepsilon), \frac{2(F(x_0) - F_{low})}{\kappa_c \varepsilon^2} \right\} \right).$$

Finally, by applying Theorem 1 we obtain the desired result. □

Corollary 1 applies to summable non-monotone line–searches, that is, those line–searches strategies such that the sequence of non-monotone terms $\{\nu_k\}$ satisfies that $\sum_{k=0}^{\infty} \nu_k$ is convergent. As mentioned in [5], several non-monotone techniques, for instance, the non-monotone rule of Zhang and Hager [15] and the non-monotone rule of Ahookhosh et. al. [24] (for details, see [16]); verified this property.

In addition, observe that in the case when $\{\nu_k\}$ is summable, Corollary 1 provides a worst-case complexity bound of $\mathcal{O}(\varepsilon^{-2})$ iterations, which agrees with the established bound for the case of summable non-monotone line-searches in the Euclidean context (when $M = \mathbb{R}^n$), see [5]. The following corollary states a worst-case complexity bound for non summable sequences.

Corollary 2. *Assume that **A1–A4** are fulfilled and that* $\lim_{k\to\infty} \nu_k = 0$. *Let* $V > 0$ *be a positive constant such that* $\nu_k \leq V$, *for all* $k \geq 0$. *Given* $\varepsilon > 0$ *and* $\tau > 0$, *let* $k_0(\tau)$ *be a positive integer such that* $\nu_k \leq \tau$ *if* $k \geq k_0(\tau)$. *Let* $\{x_k\}$ *be any sequence generated by Algorithm 1. If*

$$K \geq \max\left\{\frac{2k_0(\tau/2)V}{\tau}, 1 + k_0(\tau/2), \frac{2(F(x_0) - F_{low})}{\kappa_c \varepsilon^2}\right\}, \tag{3.21}$$

for $\tau = (\kappa_c \varepsilon^2)/2$, *it follows that*

$$\min_{k\in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon, \tag{3.22}$$

with $I_K = \{0, 1, 2, \ldots, K-1\}$. *Moreover, if* $\nu_k = \frac{M}{k}$, *for all* $k > 0$, *with* $M > 0$ *constant, then (3.22) holds if:*

$$K \geq \max\left\{\frac{16M^2}{\kappa_c^2 \varepsilon^4}, 1 + \frac{4M}{\kappa_c \varepsilon^2}, \frac{2(F(x_0) - F_{low})}{\kappa_c \varepsilon^2}\right\}. \tag{3.23}$$

Proof Let $\tau > 0$. First, denote by $s(\tau) := k_0(\tau/2) > 0$. If

$$K \geq \max\left\{\frac{2sV}{\tau}, 1 + k_0(\tau/2), \frac{2(F(x_0) - F_{low})}{\kappa_c \varepsilon^2}\right\},$$

then we have

$$\begin{aligned}
\frac{1}{K}\sum_{k=0}^{K-1} \nu_k &= \frac{1}{K}\sum_{k=0}^{s(\tau)-1} \nu_k + \frac{1}{K}\sum_{k=s(\tau)}^{K-1} \nu_k \\
&\leq \frac{1}{K}\sum_{k=0}^{s(\tau)-1} V + \frac{1}{K}\sum_{k=s(\tau)}^{K-1} \frac{\tau}{2} \\
&\leq \frac{s(\tau)V}{K} + \frac{\tau}{2} \\
&\leq \tau.
\end{aligned}$$

Hence, the assumption **A5** is fulfilled: In addition, observe that by selecting

$$K(0,\varepsilon) = \max\left\{\frac{2s(\tau)V}{\tau}, 1 + s(\tau)\right\},$$

with $\tau = (\kappa_c \varepsilon^2)/2$, the condition (3.15) is verified. Now, note that, with these selections, (3.21) can be rewritten as:

$$K \geq \max\left\{K(0,\varepsilon), \frac{2(F(x_0) - F_{low})}{\kappa_c \varepsilon^2}\right\}.$$

Consequently, all the hypotheses of Theorem 1 are satisfied. Therefore, we obtain

$$\min_{k\in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon, \tag{3.24}$$

which proves the first part.

On the other hand, suppose that $\nu_k = \frac{M}{k}$, for all $k \geq 1$, where M is a given positive constant. Notice that the sequence $\{\nu_k\}$ has the following properties:

$$\nu_k \to \infty, \quad \text{and} \quad \nu_k \leq M, \forall k,$$

and given $\tau > 0$ we also have

$$\nu_k \leq \tau \quad \Leftrightarrow \quad k \geq \frac{M}{\tau}.$$

Thus, in this specific case, we have that $k_0(\tau) = \frac{M}{\tau}$ and $V = M > 0$. Therefore, the condition (3.21) becomes (3.23), and so we conclude that (3.22) is also valid in this case. □

4. Global convergence

In this section, we analyze the global convergence of the non-monotone Algorithm 1. The results presented in this section follow directly from the complexity analysis given in the previous section. It is worth mentioning that the global convergence for some non-monotone line–search schemes has been studied in previous works, see [1, 4, 7]. Nonetheless, the results provided here are more general since we do not consider particular formulas for the sequence $\{\nu_k\}$.

From now on, we assume, without loss of generality, that Algorithm 1 never reaches $||\nabla_M F(x_k)||_{x_k} = 0$. Otherwise, the procedure has been successfully stopped at a finite iteration and there is nothing to prove. Our first result is valid for sequences $\{\nu_k\}$ summable.

Theorem 2. *Suppose that the assumptions **A1–A4** hold and that $\sum_{k=0}^{\infty} \nu_k < \infty$. Let $\{x_k\}$ be an infinite sequence generated by Algorithm 1. Then,*

$$\lim_{k\to\infty} ||\nabla_M F(x_k)||_{x_k} = 0.$$

Proof It follows from the assumption **A3** that

$$\langle \nabla_M F(x_k), d_k \rangle_{x_k} \leq -c_2 c_3 ||\nabla_M F(x_k)||^2_{x_k}. \tag{4.1}$$

On the flip side, from the construction of Algorithm 1, we have

$$\begin{aligned} F(x_{k+1}) &\leq F(x_k) + c_1 \alpha_k \delta^l \langle \nabla_M F(x_k), d_k \rangle_{x_k} + \nu_k \\ &= F(x_k) + c_1 \delta \alpha_{k+1} \langle \nabla_M F(x_k), d_k \rangle_{x_k} + \nu_k. \end{aligned} \tag{4.2}$$

Combining (4.1) with (4.2), using Lemma 2 and (3.14), we arrive at

$$\begin{aligned} F(x_{k+1}) &\leq F(x_k) - c_1 c_2 c_3 \delta \alpha_{k+1} ||\nabla_M F(x_k)||^2_{x_k} + \nu_k \\ &\leq F(x_k) - c_1 c_2 c_3 \delta \underline{\alpha} ||\nabla_M F(x_k)||^2_{x_k} + \nu_k \\ &\leq F(x_k) - \kappa_c ||\nabla_M F(x_k)||^2_{x_k} + \nu_k, \end{aligned}$$

or equivalently,

$$||\nabla_M F(x_k)||_{x_k}^2 \leq \frac{1}{\kappa_c}[F(x_k) - F(x_{k+1})] + \frac{1}{\kappa_c}\nu_k.$$

Summing up this last inequality for $k = 0, \dots, K-1$ and considering the assumption **A1**, we have

$$\begin{aligned} \sum_{k=0}^{K-1} ||\nabla_M F(x_k)||_{x_k}^2 &\leq \frac{1}{\kappa_c}\sum_{k=0}^{K-1}[F(x_k) - F(x_{k+1})] + \frac{1}{\kappa_c}\sum_{k=0}^{K-1}\nu_k, \\ &= \frac{1}{\kappa_c}[F(x_0) - F(x_K)] + \frac{1}{\kappa_c}\sum_{k=0}^{K-1}\nu_k \\ &\leq \frac{1}{\kappa_c}[F(x_0) - F(x_{low})] + \frac{1}{\kappa_c}\sum_{k=0}^{K-1}\nu_k. \end{aligned} \tag{4.3}$$

Taking limits in (4.3) and keeping in mind that $\{\nu_k\}$ is summable, we obtain

$$\sum_{k=0}^{\infty} ||\nabla_M F(x_k)||_{x_k}^2 \leq \frac{F(x_0) - F(x_{low})}{\kappa_c} + \frac{1}{\kappa_c}\sum_{k=0}^{\infty}\nu_k < \infty,$$

which implies that

$$\lim_{k\to\infty} ||\nabla_M F(x_k)||_{x_k}^2 = 0,$$

demonstrating the theorem. □

A weaker global convergence result than the one established in Theorem 2 is presented below. In particular, this result applies to sequences $\{\nu_k\}$ that are not necessarily summable.

Theorem 3. *Suppose that the assumptions **A1–A4** hold. Let $\{x_k\}$ be an infinite sequence generated by Algorithm 1. If* $\lim_{k\to\infty}\nu_k = 0$, *then*

$$\liminf_{k\to\infty} ||\nabla_M F(x_k)||_{x_k} = 0.$$

Proof The proof is by contradiction. If the theorem is not true, there exists a constant $\varepsilon > 0$ such that

$$||\nabla_M F(x_k)||_{x_k} > \varepsilon, \quad \forall k \geq 0. \tag{4.4}$$

Thus, since the sequence $\{\nu_k\} \subset \mathbb{R}_+$ tends to zero, for the ε above, there exist two constants $V \in \mathbb{R}$ and $k_0(\frac{\kappa_c\varepsilon^2}{4}) \in \mathbb{N}$ such that $\nu_k \leq V$ for any $k \geq 0$, and $\nu_k \leq \frac{\kappa_c\varepsilon^2}{4}$, for all $k \geq k_0(\frac{\kappa_c\varepsilon^2}{4})$. Now, let us select $K \in \mathbb{N}$ such that

$$K \geq \max\left\{\frac{4Vk_0(\frac{\kappa_c\varepsilon^2}{4})}{\kappa_c\varepsilon^2}, 1 + k_0\left(\frac{\kappa_c\varepsilon^2}{4}\right), \frac{2(F(x_0) - F_{low})}{\kappa_c\varepsilon^2}\right\}.$$

It follows from Corollary 2 that

$$\min_{k\in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon, \tag{4.5}$$

where $I_k = \{0, 1, \dots, K-1\}$. Combining (4.4) with (4.5) we obtain

$$\varepsilon < \min_{k\in I_K} ||\nabla_M F(x_k)||_{x_k} \leq \varepsilon,$$

which is a contradiction. Therefore the theorem is true. □

The following result applies to a collection of non-monotone line-searches based on the Grippo's non-monotone strategy (2.4).

Corollary 3. *Suppose that the assumptions **A1-A4** hold. Let $\{x_k\}$ be an infinite sequence generated by Algorithm 1 substituting the rule* (2.6) *by*

$$F(R_{x_k}(\alpha_k\delta^l d_k)) \leq R_k + c_1\alpha_k\delta^l\langle\nabla_M F(x_k), d_k\rangle_{x_k}, \tag{4.6}$$

where

$$F(x_k) \leq R_k \leq \max_{0\leq i\leq m(k)} F(x_{k-i}), \tag{4.7}$$

with $m(0) = 0$ and $0 \leq m(k) \leq \min\{m(k-1)+1, N\}$ for $N \in \mathbb{N}$ being a constant predetermined by the user. If the level set $\mathscr{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded, then

$$\liminf_{k\to\infty} ||\nabla_M F(x_k)||_{x_k} = 0.$$

Proof Applying the steps of the proof of Lemma 2 in [3] and considering (4.7), one can obtain that

$$\lim_{k\to\infty} F(x_k) = \lim_{k\to\infty} \max_{0\leq i\leq m(k)} F(x_{k-i}).$$

In addition, observe that if we choose $\nu_k = R_k - F(x_k) \geq 0$, for all k in Algorithm 1, we recover the inequality (4.6). Keeping in mind this observation and noticing that

$$\lim_{k\to\infty} \nu_k = \lim_{k\to\infty}[R_k - F(x_k)] = \lim_{k\to\infty}\left[\max_{0\leq i\leq m(k)} F(x_{k-i}) - F(x_k)\right] = 0.$$

Lastly, the result follows as a consequence of Theorem 3. □

5. Conclusions

In this work we focused on the optimization problem on a Riemannian manifold. Particularly, we considered a generic Riemannian non-monotone line-search algorithm to address this problem. We derived bounds on the total number of iterations needed by this class of iterative methods to achieve ε-approximated critical points under some assumptions. In our Algorithm 1 the degree of non-monotonicity is managed by a generic sequence $\{\nu_k\} \subset \mathbb{R}_+$.

When the feasible set M is a Euclidean space, in previous works [5, 13], were proved that this family of procedures require at most $\mathscr{O}(\varepsilon^{-2})$ iterations to obtain a point that approximately verify the first–order optimality condition. In addition, in the case when M is a Riemannian manifold this bound was demonstrated in [12] for the Riemannian gradient descent with $\nu_k = 0$, for all k. In this framework, our analysis obtained the same bound, but considering general tangent descent directions and non-negative non-monotone terms $\nu_k \geq 0$. In particular, our study is valid for sequences $\{\nu_k\} \subset \mathbb{R}_+$ such that its series $\sum_k^\infty \nu_k$ is convergent, and also for some sequences $\{\nu_k\} \subset \mathbb{R}_+$ such that its series $\sum_k^\infty \nu_k$ is divergent but satisfies $\nu_k \to 0$. Therefore, our analysis includes a great variety of non-monotone schemes existing in the literature. Finally, we established the global convergence of the algorithm considering these two types of non-negative sequences.

Acknowledgement

The author was financially supported by the Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, through the FES startup package for scientific research.

REFERENCES

1. Absil P., Mahony R, Sepulchre R (2009) Optimization algorithms on matrix manifolds. Princeton university press.
2. Luenberger D. and Ye Y., et. al. (1984) Linear and nonlinear programming, vol 2, Springer.
3. Amini K., Ahookhosh M. and Nosratipour H. (2014) An inexact line search approach using modified nonmonotone strategy for unconstrained optimization. Numerical Algorithms, vol. 66, pp. 49–78.
4. Boumal N. (2023) An introduction to optimization on smooth manifolds. Cambridge University Press.
5. Grapiglia G. and Sachs E. (2021) A generalized worst-case complexity analysis for non-monotone line searches. Numerical Algorithms, vol. 87, pp. 779–796.
6. Hu J., Liu X., Wen Z. and Yuan Y. (2020) A brief introduction to manifold optimization. Journal of the Operations Research Society of China, vol. 8, pp. 199–248.
7. Oviedo H. (2022) Global convergence of Riemannian line search methods with a Zhang-Hager-type condition. Numerical Algorithms, vol. 91(3), pp. 1183–1203.
8. Dalmau O. and Oviedo H. (2018) Projected nonmonotone search methods for optimization with orthogonality constraints. Computational and Applied Mathematics, vol. 37(3), pp. 3118–3144.
9. Oviedo H., Lara H. and Dalmau O. (2019) A non-monotone linear search algorithm with mixed direction on stiefel manifold. Optimization Methods and Software, vol. 34(2), pp. 437–457.
10. Oviedo H., Dalmau O. and Lara H. (2021) Two adaptive scaled gradient projection methods for stiefel manifold constrained optimization. Numerical Algorithms, vol. 87(2), pp. 1107–1127.
11. Oviedo H. (2022) Implicit steepest descent algorithm for optimization with orthogonality constraints. Optimization Letters, vol. 16(6), pp. 1773–1797.
12. Boumal N., Absil P. and Cartis C. (2019) Global rates of convergence for nonconvex optimization on manifolds. IMA Journal of Numerical Analysis, vol. 39(1), pp. 1–33.
13. Grapiglia G. and Sachs E. (2017) On the worst-case evaluation complexity of non-monotone line search algorithms. Computational Optimization and applications, vol. 68, pp. 555–577.
14. Chavel I. (2006) Riemannian geometry: a modern introduction. Cambridge university press, vol. 98.
15. Zhang H. and Hager W. (2004) A nonmonotone line search technique and its application to unconstrained optimization. SIAM journal on Optimization, vol. 14(4), pp. 1043–1056.
16. Sachs E. and Sachs S. (2011) Nonmonotone line searches for optimization algorithms. Control and Cybernetics, vol. 40(4), pp. 1059–1075.
17. Udriste C. (1994) Convex functions and optimization methods on Riemannian manifolds. Springer Science and Business Media.
18. Grippo L., Lampariello F. and Lucidi S. (1986) A nonmonotone line search technique for Newton's method. SIAM journal on Numerical Analysis, vol. 23(4), pp. 07–716.
19. Iannazzo B. and Porcelli M. (2018) The Riemannian Barzilai–Borwein method with nonmonotone line search and the matrix geometric mean computation. IMA Journal of Numerical Analysis, vol. 38(1), pp. 495–517.
20. Zhu X. (2017) A Riemannian conjugate gradient method for optimization on the Stiefel manifold. Computational optimization and Applications, vol. 67, pp. 73–110.
21. Yu Z. and Pu D. (2008) A new nonmonotone line search technique for unconstrained optimization. Journal of Computational and Applied Mathematics, vol. 219(1), pp. 134–144.
22. Huang S., Wan Z. and Chen X. (2015) A new nonmonotone line search technique for unconstrained optimization. Numerical Algorithms, vol. 68(4), pp. 671–689.
23. Hu S., Huang Z. and Lu N. (2010) A non-monotone line search algorithm for unconstrained optimization. Journal of Scientific Computing, vol. 42(1).

24. Ahookhosh M., Amini K. and Bahrami S. (2012) A class of nonmonotone Armijo-type line search method for unconstrained optimization. Optimization, vol. 61(4), pp. 387–404.
25. Gao B., Son N., Absil P. and Stykel T. (2021) Riemannian optimization on the symplectic Stiefel manifold. Optimization, vol. 31(2), pp. 1546–1575.
26. Lara, H., Yuan, J. and Oviedo H. (2014). Matrix completion via a low rank factorization model and an augmented lagrangean succesive
27. Rodríguez-Domínguez, U., Dalmau, O. (2020). Hierarchical discriminative deep dictionary learning. IEEE Access, 8, 142680-142690.
28. Absil, P. A., Gallivan, K. A. (2006, May). Joint diagonalization on the oblique manifold for independent component analysis. In 2006 IEEE
29. Dai, W., Kerman, E., Milenkovic, O. (2012). A geometric approach to low-rank matrix completion. IEEE Transactions on Information Theory, 58(1), 237-247.
30. Boumal, N., Absil, P. A. (2011). RTRMC: A Riemannian trust-region method for low-rank matrix completion. Advances in neural information processing systems, 24.
31. Meinhold R., Hayes T. and Cahill N. (2016) Efficiently computing piecewise flat embeddings for data clustering and image segmentation. In: IEEE MIT Undergraduate Research Technology Conference (URTC), pp. 1–4.
32. Yu Y., Fang C. and Liao Z. (2015) Piecewise flat embedding for image segmentation. In: Proceedings of the IEEE international conference on computer vision, pp. 1368–1376.