Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

An Adaptive Riemannian Spectral Gradient Method Without Cost Function Evaluations

Harry Oviedo[1]
Universidad Adolfo Ibáñez, Av. Diag. Las Torres 2640, Santiago de Chile, Chile.

Abstract. This work addresses a spectral correction for the adaptive weighted normalized gradient method in the minimization of smooth functions on a Riemannian manifold. The proposed algorithm incorporates a Barzilai-Borwein-type stepsize in order to accelerate the Riemannian gradient procedure. The new approach is globally convergent and does not require evaluating the objective function throughout the iterative process. Our preliminary numerical experiments show that the use of spectral correction improves the performance of the adaptive Riemannian gradient method.

Keywords. Riemannian Optimization, Gradient Method, Steplength Selection, Adaptive Methods

1 Introduction

The field of optimization on Riemannian manifolds has experienced significant growth in recent years, driven by its broad applicability across various domains, including signal processing [1, 9, 19, 20], pattern recognition [9, 11, 19], matrix completion [1, 5, 9], deep learning [9], eigenvalue computation [5, 9, 16, 17], energy minimization [9, 15], among others. On Riemannian manifolds, the traditional gradient descent algorithm is extended to the Riemannian gradient descent, where the updates are performed along the manifold's geometry rather than in the ambient Euclidean space.

Traditionally, the selection of stepsizes in the gradient method is accomplished using line-search [12]. This approach determines suitable stepsize by evaluating the objective function multiple times until a sufficient decrease condition is satisfied. More recently, researchers have developed adaptive stepsize rules that rely solely on gradient evaluations in the context of Euclidean optimization. In particular, the method WNGrad [22] is a gradient method in which the stepsize at each iteration is obtained as the inverse of a weighted sum of the squared norms of all the gradients already computed by the method. Remarkably, in [22] it was shown that WNGrad performs at most $\mathcal{O}(\epsilon^{-2})$ iterations to generate an ϵ-approximate stationary point of the objective function. Adaptive gradient descent methods have gained widespread popularity in the field of deep learning and optimization due to their ability to automatically adapt the stepsize (also called *learning rate*) without conducting a line-search and avoiding cost function evaluations, leading to improved convergence and generalization. Indeed, Grapiglia et. al. in [8] generalized the WNGrad method to minimize smooth functions on a Riemannian manifold.

On the other hand, it is well-known that spectral stepsizes are often able to improve the numerical performance of gradient methods while maintaining low computational cost, which makes them attractive for solving large-scale problems, see [4, 6, 13, 18]. In this work, we propose an adaptive gradient method, which maintains the algorithmic structure of the general procedure presented in [8], and incorporates a spectral stepsize to accelerate the numerical performance of the Grapiglia's

[1]harry.oviedo@uai.cl

method.

The rest of this paper is organized as follows. In Section 2, we establish the optimization problem to be addressed and also describe the algorithm developed in [8]. In subsection 2.1, we present the proposal which is based on the incorporation of spectral correction within the framework of the Grapiglia's method. Some numerical experiments are reported in Section 3. Finally, in Section 4, we give our conclusions.

2 The Riemannian WNgrad Method

We begin this section by posing the optimization problem to be solved. In particular, we focus on the following problem,

$$\min f(x) \quad \text{s.t.} \quad x \in \mathcal{M}, \tag{1}$$

where $f : \mathcal{M} \to \mathbb{R}$ is a bounded from below smooth function (not necessarily convex) and $\mathcal{M}$ is a Riemannian manifold connected and complete.

Recently, in [8] was proposed a low-cost gradient method to solve (1), namely the adaptive Riemannian gradient method (ARWNGrad). ARWNGrad is a generalization of the weight normalized gradient method (WNGrad) introduced by Wu et. al. in [22] for minimizing a continuously differentiable function on the Euclidean space $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function and $x_0 \in \mathbb{R}^n$ an initial point, the WNGrad procedure computes the iterates by using the following update scheme

$$x_{k+1} = x_k - \frac{1}{b_k}\nabla f(x_k), \quad \text{with} \quad b_k = b_{k-1} + \frac{1}{b_{k-1}}\|\nabla f(x_k)\|^2. \tag{2}$$

Remarkably, WNGrad does not require evaluating the objective function. In addition, it was shown in [22] that WNGrad performs at most $\mathcal{O}(\epsilon^{-2})$ iterations to generate an ϵ-approximate stationary point of the objective function. Several works have extended and analyzed the gradient method to minimize smooth function on manifolds, see [1, 3, 7, 9, 14–17, 21]. However, none of these works corresponds to the generalization of the iterative scheme (2). Following the ideas exposed in [22], Grapiglia et. al. [8] quickly extended the WNGrad method to the Riemannian framework. To accomplish this task, the iteration scheme (2) is changed to

$$x_{k+1} = \exp_{x_k}\left(-\frac{1}{b_k}\operatorname{grad} f(x_k)\right), \quad \text{with} \quad b_k = b_{k-1} + \frac{1}{b_{k-1}}\|\operatorname{grad} f(x_k)\|^2, \tag{3}$$

where for $x \in \mathcal{M}$, $T_x\mathcal{M}$ denotes the tangent space of $\mathcal{M}$ at x, $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ denotes the exponential map at x, this is the function such that $\exp_x(v) = \gamma(1)$ and γ is a geodesic defined in an interval containing $[0, 1]$ with $\gamma(0) = x$ and $\dot{\gamma}(0) = v$, for $v \in T_x\mathcal{M}$. Notice that the sequence of stepsizes $\{\frac{1}{b_k}\}$ in (3) (an also in (2)) is monotonically decreasing and therefore when k grows the stepsize $\frac{1}{b_k}$ can be very small, which makes method (3) progress very slowly throughout the process. In order to break this pattern, in [8] the authors proposed a flexible version of the (3) method that allows the method to use larger stepsizes under certain conditions. Following the development in [8], the ARWNGrad algorithm is presented in Algorithm 1.

DOI: 10.5540/03.2026.012.01.0244

Algorithm 1 Adaptive Riemannian Gradient Method

Require: Given $x_0 \in \mathcal{M}$, and parameters $\hat{c} \in \mathbb{N} - \{0\}$, $b_{\min} > 0$, $b_0 \geq b_{\min}$ and $\alpha \in [0, 1)$. Set $t_0 = \frac{1}{b_0}$, $w_0 = \|\text{grad}\, f(x_0)\|$, $c_0 = 0$ and $k = 0$.
while $\|\text{grad}\, f(x_k)\| \neq 0$ **do**
  $x_{k+1} = \exp_{x_k}(-t_k \text{grad}\, f(x_k))$.
  if $\|\text{grad}\, f(x_{k+1})\| \leq \alpha\, w_k$ **then**
    if $c_k + 1 = \hat{c}$ **then**
      $(b_{k+1}, w_{k+1}, c_{k+1}) = (\hat{b}_k, \|\text{grad}\, f(x_{k+1})\|, 0)$, where $\hat{b}_k \in [b_{\min}, b_k]$.
    else if $c_k + 1 < \hat{c}$ **then**
      $(b_{k+1}, w_{k+1}, c_{k+1}) = \Big(b_k + \frac{\|\text{grad}\, f(x_{k+1})\|^2}{b_k}, w_k, c_k + 1\Big)$.
    end if
  else
    $(b_{k+1}, w_{k+1}, c_{k+1}) = \Big(b_k + \frac{\|\text{grad}\, f(x_{k+1})\|^2}{b_k}, w_k, 0\Big)$.
  end if
  Define $t_{k+1} = \frac{1}{b_{k+1}}$.
  $k \leftarrow k + 1$.
end while

In Algorithm 1 the terms $\hat{c}$, $b_{\min}$, b_0 and α are global parameters which must be selected by the user (they can even be chosen randomly), clearly satisfying their corresponding conditions as described at the beginning of Algorithm 1. In the section on numerical experiments (see Section 3), we comment on how they will be chosen. Observe that when $\alpha = 0$, Algorithm 1 reduces to the Riemannian extension of WNGrad presented in (3). The above algorithm is globally convergent (under Lipschitzian assumptions on the Riemannian gradient) and computes an ϵ-approximate stationary point of the objective function in at most $\mathcal{O}(\epsilon^{-2})$ gradient evaluations when $\alpha = 0$. For $\alpha > 0$ the Algorithm 1 has a worst case complexity bound of $\mathcal{O}(|\, log(\epsilon)\, |\, \epsilon^{-2})$ gradient evaluations, for details, see [8].

2.1 An ARWNGrad With Spectral Correction

In this subsection, we present the contribution of this paper. In particular, our goal is to incorporate a Barzilai-Borwein-type stepsize into the Algorithm 1. Notice that Algorithm 1 leaves an open window for the selection of the term $\hat{b}_k$, which directly impacts the calculation of the stepsize t_{k+1}. In [8] the authors proposed to compute this scalar by the following relation

$$\hat{b}_k = \min\left\{b_0, \max\left\{b_{\min}, \frac{b_k}{3\hat{c}}\right\}\right\}.$$

Although this formula showed numerical advantages versus the method given by (3) (Algorithm 1 with $\alpha = 0$), this heuristic does not take advantage of quasi-Newton information. In view of this, we propose to calculate $\hat{b}_k$ as follows

$$\hat{b}_k = \min\left\{b_k, \max\left\{b_{\min}, \frac{1}{\tau_k^{BB}}\right\}\right\} \quad \text{with} \quad \tau_k^{BB} = \frac{\|s_{k-1}\|}{\|y_{k-1}\|}, \tag{4}$$

where the tangent vectors $s_{k-1}, y_{k-1} \in T_{x_k}\mathcal{M}$ are given by

$$s_{k-1} = \mathcal{T}_{x_{k-1}\to x_k}(-t_{k-1}\text{grad}\, f(x_{k-1})) = -t_{k-1}\mathcal{T}_{x_{k-1}\to x_k}(\text{grad}\, f(x_{k-1})), \tag{5}$$

and

$$y_{k-1} = \text{grad}\, f(x_k) - \mathcal{T}_{x_{k-1}\to x_k}(\text{grad}\, f(x_{k-1})) = \text{grad}\, f(x_k) - \frac{1}{t_{k-1}} s_{k-1}, \tag{6}$$

respectively. In the above equations, the map $\mathcal{T}_{x_k \to x_{k+1}} : T_{x_k}\mathcal{M} \to T_{x_{k+1}}\mathcal{M}$ is a vector transport, see Definition 8.1.1 in [1]. Note that the stepsize τ_k^{BB} is the geometric mean of the Riemannian version of the spectral stepsizes

$$\tau_k^{BB1} = \frac{\|s_{k-1}\|^2}{\langle s_{k-1}, y_{k-1}\rangle} \quad \text{and} \quad \tau_k^{BB2} = \frac{\langle s_{k-1}, y_{k-1}\rangle}{\|y_{k-1}\|^2}, \tag{7}$$

which were originally derived by Barzilai-Borwein [2] by imposing a quasi-Newton property in the context of minimization on $\mathbb{R}^n$ and extended to the Riemannian setting by Iannazzo et. al. [10]. Although the spectral stepsizes (7) were introduced in [10] to accelerate the standard Riemannian gradient method [1], they were used in combination with a non-monotone line-search, so the scheme proposed in [10] requires evaluating the objective function at least once per iteration. By contrast, in this work we propose the use of spectral stepsizes within the framework of Algorithm 1, which does not need to evaluate the objective function.

3 Computational Results

In this section, we compare the numerical behavior of the proposal in two groups of experiments. The purpose of these experiments is to illustrate numerically that the incorporation of spectral correction in Algorithm 1 can sometimes lead to better efficiency, in terms of the number of iterations and CPU time. We performed computational experiments comparing the following methods

- **ARMIJO**: the Riemannian gradient method equipped with Armijo line-search.
- **RSGM**: the Riemannian spectral gradient method equipped with the nonmonotone Zhang and Hager's line-search in combination with the stepsize τ_k^{BB} presented in (4).
- **RWNG**: Algorithm 1 with $\alpha = 0$ and $b_0 = \|\text{grad}\, f(x_0)\|$.
- **ARWNG**: Algorithm 1 with $\alpha = 0.9$, $\hat{c} = 5$, $\hat{b}_k = \min\{\hat{b}_{\max}, \max\{10^{-4}, b_k/(3\hat{c})\}\}$, where $\hat{b}_{\max} = b_0 = \|\text{grad}\, f(x_0)\|$.
- **ARWNGs**: Algorithm 1 with $\alpha = 0.9$, $\hat{c} = 5$, $\hat{b}_k$ according to (4), where $b_{\min} = 10^{-4}$.

All these procedures were implemented in MATLAB (R2016a) and the experiments were performed on an 11th Gen Intel(R) Core(TM) i5-1135G7, CPU 2.40GHz with 8 GB RAM. In stop all these methods using the stopping condition $\|\text{grad}\, f(x_k)\| \leq 10^{-4}$ or with the maximum number of iterations $k = 100000$. In the two groups of experiments, we consider optimization problems over the Riemannian manifold $\mathbb{P}_{++} = \{X \in \mathbb{R}^{n\times n} : X^\top = X,\ X \succ 0\}$, that is, the set of all the n-by-n symmetric positive definite matrices. The exponential map $\exp_X : T_X\mathcal{M} \to \mathcal{M}$ associated with this manifold is given by $\exp_X(V) = X^{1/2}\exp_m(X^{-1/2}VX^{-1/2})X^{1/2}$, where $\exp_m$ denotes the exponential matrix. In all the tables reported in this section, *Iter*, *Nfe*, *Time* and *Grad* denote the averaged of number of iterations, the averaged of number of objective function evaluations, the averaged total computing time in seconds and the averaged gradient norm $\|\text{grad}\, F(\hat{X})\|_F$ where $\hat{X}$ is the optimum estimated by the method, respectively.

3.1 Problem Class 1

We consider the following manifold constrained optimization problem

$$\min_{X\in\mathbb{P}_{++}} F(X) := \ln(\det(X))^2 - \ln(\det(X)). \tag{8}$$

The above problem was taken from [8]. In our first experiment, we solve (8) for $n = 100$. For this value of n, the codes were run for 100 different randomly generated initial points. The starting points were assembled following the approach described in subsection 4.1 in [8]. The results are reported in Table 1.

Table 1: Numerical results associated with Problem Class 1 with $n = 100$.

Method	**Iter**	**Nfe**	**Time**	**Grad**
ARMIJO	32	257	1.6366	6.2583e-5
RSGM	9.82	10.82	0.0652	1.0220e-5
RWNG	100000	0	625.249	0.3761
ARWNG	280.76	0	1.7617	5.4098e-5
ARWNGs	63.5	0	0.3938	1.7758e-5

From Table 1, we observe that all algorithms successfully solve the problem class 1, except for the RWNG method which stopped with the maximum number of iterations. Among the remaining four methods, we notice that RSGM was the most efficient both in terms of number of iterations and CPU time. Furthermore, we clearly observe that our ARWNGs outperforms the Algorithm 1 because it was more efficient than the two procedures developed in [8].

3.2 Problem Class 2

Now, we consider the following minimization problem

$$\min_{X \in \mathbb{P}_{++}} F(X) := \frac{1}{2} \sum_{j=1}^{m} \|\ln(X^{-1/2} A_j X^{-1/2})\|_F^2, \tag{9}$$

where $A_1, \ldots, A_m$ are data matrices. The unique global optimum of the problem (9) is called **Karcher mean** and is defined as the geometric mean of matrices, see [10]. This problem also appears in [8]. In the second group of experiments, we select $n \in \{20, 100\}$ and $m = 10$. For all pairs (n, m), we solve (9) for 100 different randomly generated starting points with all the methods. Again, the initial points were constructed exactly as described in subsection 4.2 in [8]. The numerical results are collected in Table 2.

Table 2: Numerical results associated with Problem Class 2.

Method	**Iter**	**Nfe**	**Time**	**Grad**
	$(n, m) = (20, 10)$			
ARMIJO	38.02	153.33	2.8418	8.0460e-5
RSGM	5.77	6.77	0.1056	3.4181e-5
RWNG	8.68	0	0.0843	5.2236e-5
ARWNG	32.61	0	0.3181	6.1549e-5
ARWNGs	7.22	0	0.0697	4.1748e-5
	$(n, m) = (100, 10)$			
ARMIJO	131.53	528.33	145.877	4.7892e-5
RSGM	7.81	8.81	2.1216	4.4351e-5
RWNG	48.24	0	6.8020	8.8383e-5
ARWNG	92.89	0	13.1205	6.8357e-5
ARWNGs	11.72	0	1.6526	5.7537e-5

From Table 2 we notice that ARWNGs was superior to the rest of the approaches and surprisingly solved the problem (9) in less computational time than RSGM. We also note that between the

two procedures developed in [8], the more basic version of Algorithm 1 (RWNG) performed better than the elaborate version (ARWNG). Nonetheless, our spectral correction strategy improves the performance of Algorithm 1. The ARMIJO method solved the problem very slowly.

4 Conclusion

In this work, we have addressed the problem of minimizing a continuously differentiable function over a Riemannian manifold. To solve this problem, we consider a recently introduced adaptive Riemannian gradient method, called ARWNGrad, which does not need to execute line-search throughout the entire iterative process. In particular, the stepsizes are determined by a closed-formula. The considered approach preserves all the theoretical guarantees of the standard Riemannian gradient method and its algorithmic simplicity. With the aim of improving the numerical behavior of ARWNGrad, we have incorporated a spectral correction in the computation of the stepsize based on the geometric mean of the Barzilai-Borwein stepsizes, maintaining the algorithmic structure of the ARWNGrad and its established theory. We have conducted some numerical experiments to evaluate the proposal in two optimization problems defined on the set of all the n-by-n positive definite matrices. Our numerical results clearly show that the proposed procedure outperforms the original versions of the ARWNGrad method.

References

[1] P-A. Absil, R. Mahony, and R. Sepulchre. **Optimization algorithms on matrix manifolds**. Princeton University Press, 2009. ISBN: 9780691132983.

[2] J. Barzilai and J. M. Borwein. "Two-point step size gradient methods". In: **IMA journal of numerical analysis** 8.1 (1988), pp. 141–148. DOI: 10.1093/imanum/8.1.141.

[3] G. C. Bento, O. P. Ferreira, and J. G. Melo. "Iteration-complexity of gradient, subgradient and proximal point methods on Riemannian manifolds". In: **Journal of Optimization Theory and Applications** 173.2 (2017), pp. 548–562. DOI: 10.1007/s10957-017-1093-4.

[4] E. G. Birgin, J. M. Martínez, and M. Raydan. "Spectral projected gradient methods: review and perspectives". In: **Journal of Statistical Software** 60 (2014), pp. 1–21. DOI: 10.18637/jss.v060.i03.

[5] J-P. Chehab, H. Oviedo, and M. Raydan. "Optimization schemes on manifolds for structured matrices with fixed eigenvalues". In: **Computational Optimization and Applications** (2024), pp. 1–26. DOI: 10.1007/s10589-024-00630-3.

[6] D. Di Serafino, V. Ruggiero, G. Toraldo, and L. Zanni. "On the steplength selection in gradient methods for unconstrained optimization". In: **Applied Mathematics and Computation** 318 (2018), pp. 176–195. DOI: 10.1016/j.amc.2017.07.037.

[7] O. P. Ferreira, M. S. Louzeiro, and L. Prudente. "Gradient method for optimization on Riemannian manifolds with lower bounded curvature". In: **SIAM Journal on Optimization** 29.4 (2019), pp. 2517–2541. DOI: 10.1137/18M1180633.

[8] G. N. Grapiglia and G. Stella. "An Adaptive Riemannian Gradient Method Without Function Evaluations". In: **Journal of Optimization Theory and Applications** 197.3 (2023), pp. 1140–1160. DOI: 10.1007/s10957-023-02227-y.

[9] J. Hu, X. Liu, Z. Wen, and Y. Yuan. "A brief introduction to manifold optimization". In: **Journal of the Operations Research Society of China** 8 (2020), pp. 199–248. DOI: 10.1007/s40305-020-00295-9.

[10] B. Iannazzo and M. Porcelli. "The Riemannian Barzilai–Borwein method with nonmonotone line search and the matrix geometric mean computation". In: **IMA Journal of Numerical Analysis** 38.1 (2018), pp. 495–517. DOI: 10.1093/imanum/drx015.

[11] E. Kokiopoulou, J. Chen, and Y. Saad. "Trace optimization and eigenproblems in dimension reduction methods". In: **Numerical Linear Algebra with Applications** 18.3 (2011), pp. 565–602. DOI: 10.1002/nla.743.

[12] J. Nocedal and S. J. Wright. **Numerical optimization**. New York: Springer, 1999.

[13] H. Oviedo. "A cyclic delayed weighted steplength for the gradient method". In: **Ricerche di Matematica** 73.2 (2024), pp. 873–885. DOI: 10.1007/s11587-021-00646-5.

[14] H. Oviedo. "Global convergence of Riemannian line search methods with a Zhang-Hager-type condition". In: **Numerical Algorithms** 91.3 (2022), pp. 1183–1203. DOI: 10.1007/s11075-022-01298-8.

[15] H. Oviedo. "Implicit steepest descent algorithm for optimization with orthogonality constraints". In: **Optimization Letters** 16.6 (2022), pp. 1773–1797. DOI: 10.1007/s11590-021-01801-5.

[16] H. Oviedo, O. Dalmau, and H. Lara. "Two adaptive scaled gradient projection methods for Stiefel manifold constrained optimization". In: **Numerical Algorithms** 87 (2021), pp. 1107–1127. DOI: 10.1007/s11075-020-01001-9.

[17] H. Oviedo and H. Lara. "Spectral residual method for nonlinear equations on Riemannian manifolds". In: **Computational and Applied Mathematics** 40.7 (2021), p. 238. DOI: 10.1007/s40314-021-01630-3.

[18] M. Raydan. "The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem". In: **SIAM Journal on Optimization** 7.1 (1997), pp. 26–33. DOI: 10.1137/S1052623494266365.

[19] U. Rodriguez and O. Dalmau. "Symmetric nonnegative matrix factorization with elastic-net regularized block-wise weighted representation for clustering". In: **Pattern Analysis and Applications** 25.4 (2022), pp. 807–817. DOI: 10.1007/s10044-022-01062-7.

[20] U. Rodríguez and O. Dalmau. "Block-wise weighted sparse representation-based classification". In: **Signal, Image and Video Processing** 14.8 (2020), pp. 1647–1654. DOI: 10.1007/s11760-020-01700-9.

[21] S. T. Smith. **Optimization techniques on Riemannian manifolds**. Online. Accessed on 16/04/2025, arXiv preprint arXiv:1407.5965. 2014.

[22] X. Wu, R. Ward, and L. Bottou. **Wngrad: Learn the learning rate in gradient descent**. Online. Accessed on 16/04/2025, arXiv preprint arXiv:1803.02865. 2018.

