On the approximation of the Riemannian barycenter

Simon Mataigne[1][0009−0004−1970−8102], P.-A. Absil[1][0000−0003−2946−4178], and Nina Miolane[2][0000−0002−1200−9024]

[1] ICTEAM Institute, UCLouvain, Louvain-la-Neuve, Belgium.
simon.mataigne@uclouvain.be, pa.absil@uclouvain.be
[2] UC Santa Barbara, Electrical and Computer Engineering, Santa Barbara, CA.
ninamiolane@ucsb.edu

Abstract. We present a method for computing an approximate Riemannian barycenter of a collection of points lying on a Riemannian manifold. Our approach relies on the use of theoretically proven under- and over-approximations of the Riemannian distance function. We compare it to Riemannian steepest descent on the exact objective function of the Riemannian barycenter and to an approach that approximates the Riemannian logarithm using lifting maps. Experiments are conducted on the Stiefel manifold.

Keywords: Riemannian barycenter · Karcher mean · Fréchet mean · Riemannian center of mass · Bounds · Stiefel manifold.

1 Introduction

Computing the Riemannian barycenter of a dataset lying on a Riemannian manifold is a fundamental problem in statistics on manifolds. See, e.g., [14] and references therein for applications. We outline the key aspects of this problem.

Let $\mathcal{M}$ be a complete manifold equipped with a Riemannian metric g and $d_g : \mathcal{M}^2 \to \mathbb{R}$ be the distance function induced by the Riemannian metric; see section 2 for details. Given a collection of points $\{x_i\}_{i=1}^N \subset \mathcal{M}$ and a parameter $\nu \geq 1$, let

$$f_\nu : \mathcal{M} \to \mathbb{R} : x \mapsto \left(\sum_{i=1}^{N} d_g(x, x_i)^\nu \right)^{\frac{1}{\nu}} \quad \text{and} \quad x^* \in \operatorname*{arg\,min}_{x \in \mathcal{M}} f_\nu(x). \tag{1}$$

Conditions regarding the existence and uniqueness of x^* can be found, e.g., in [3]. The point x^* is called a *Riemannian l^ν-barycenter* of $\{x_i\}_{i=1}^N$. Classical choices for ν are $\nu = 1, 2$, or $\nu \to \infty$ (minimax problem). For $\nu = 2$, x^* generalizes the notion of Euclidean mean on Riemannian manifolds; it is then often referred to as a Fréchet mean or a Karcher mean [11]. For $\nu = 1$, x^* generalizes the median.

Solving (1) usually relies on gradient-based optimization algorithms, which require the computation of *Riemannian logarithms* at each iteration [4]. However, there are important manifolds lacking tractable algorithms that are guaranteed to compute the Riemannian logarithm within prescribed accuracy, e.g.,

the Stiefel manifold of orthonormal p-frames in $\mathbb{R}^n$, $n > p$, defined by

$$\mathrm{St}(n,p) \coloneqq \{X \in \mathbb{R}^{n\times p} \mid X^\top X = I_p\},$$

where I_p is the $p \times p$ identity matrix. See [1,7,8] and related works for applications of the Stiefel manifold. To tackle this computational bottleneck of (1), an approach, described in section 4, approximates the Riemannian logarithm by *lifting maps* [10,6].

Our approach replaces the exact objective function f_ν in (1) by an approximate objective function h_ν that is based on theoretically guaranteed under- and over-approximations of f_ν. The estimated Riemannian l^ν-barycenter is then obtained by computing

$$\widehat{x}^* \in \arg\min_{x\in\mathcal{M}} h_\nu(x). \tag{2}$$

For h_ν well chosen, $\widehat{x}^*$ can be obtained at significantly lower computational cost than x^*. A computable bound on $\frac{f_\nu(\widehat{x}^*)-f_\nu(x^*)}{f_\nu(x^*)}$ is given in Theorem 1 and numerical comparisons are conducted in section 5.

2 Preliminaries on manifolds

We briefly introduce essential notions of Riemannian geometry. Details can be found, e.g., in [1,15]. A Riemannian metric g, *metric* for short, is a family $\{g_x : T_x\mathcal{M}\times T_x\mathcal{M} \to \mathbb{R}\}_{x\in\mathcal{M}}$ of symmetric positive definite bilinear forms that depend smoothly on the location x. For all $\xi \in T_x\mathcal{M}$, the norm induced by the metric is $\|\xi\|_g \coloneqq \sqrt{g_x(\xi,\xi)}$ and the length $l_g(\gamma)$ of every continuously differentiable curve $\gamma : [0,1] \to \mathcal{M}$ is given by $l_g(\gamma) = \int_0^1 \left\|\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\right\|_g \mathrm{d}t$. If $\mathcal{M}$ is a complete Riemannian manifold, it follows from the Hopf-Rinow theorem [15, Chap. III, Thm. 1.1] that the distance function induced by g satisfies

$$d_g(x,\widetilde{x}) = \min\{l_g(\gamma) \mid \gamma : [0,1] \to \mathcal{M}, \gamma(0) = x,\ \gamma(1) = \widetilde{x}\}.$$

A *geodesic* $\gamma_g : \mathbb{R} \to \mathcal{M}$ is a locally minimizing curve with constant speed $v > 0$, i.e., for all $t \in \mathbb{R}$, there is $\varepsilon > 0$ such that for all $s \in \mathbb{R}$, if $|t-s| < \varepsilon$, then $d_g(\gamma_g(t),\gamma_g(s)) = |t-s|v$.

The Riemannian exponential at $x \in \mathcal{M}$ is the function $\mathrm{Exp}_x : T_x\mathcal{M} \to \mathcal{M}$ mapping ξ to the point reached at unit time by the geodesic with starting point x and initial velocity ξ, see, e.g., [15, Chap. II, Sec. 2]. The Riemannian logarithm is defined in [12,13]:

Definition 1. *Let $x, \widetilde{x} \in \mathcal{M}$ where $\mathcal{M}$ is a complete Riemannian manifold endowed with a metric g. The Riemannian logarithm $\mathrm{Log}_x(\widetilde{x})$ is a set-valued function returning all $\xi \in T_x\mathcal{M}$ such that*

$$\mathrm{Exp}_x(\xi) = \widetilde{x} \text{ and } \|\xi\|_g = d_g(x,\widetilde{x}). \tag{3}$$

The curves $[0,1] \ni t \mapsto \mathrm{Exp}_x(t\xi)$ are then called minimal geodesics.

It is known that the subset of $\mathcal{M}$ where Log_x is not a singleton is included in the *cut locus* $\mathcal{C}_x$ and is a zero measure set of $\mathcal{M}$ [15, Chap. III, Lem. 4.4]. In numerical algorithms, it is thus harmless to consider that Log_x returns a singleton. The Riemannian logarithm is key to the Riemannian l^ν-barycenter because, if $x \notin \cup_{i=1}^N \mathcal{C}_{x_i}$(and $x \notin \{x_i\}_{i=1}^N$ if $\nu < 2$), then $\mathrm{grad}_x\left(d_g(x,x_i)^\nu\right) = -\nu d_g(x,x_i)^{\nu-2}\mathrm{Log}_x(x_i)$ [4, Eq. 2.8], where grad_x denotes the Riemannian gradient. If $\nu \geq 2$ and if $x^* \notin \cup_{i=1}^N \mathcal{C}_{x_i}$, it can be deduced that $\mathrm{grad}_x(f_\nu(x^*)^\nu) = 0$ if and only if [4, Sec. 2.1.5]

$$\sum_{i=1}^{N} d_g(x^*,x_i)^{\nu-2}\mathrm{Log}_{x^*}(x_i) = 0_x. \tag{4}$$

The stationarity condition (4) is illustrated in Figure 1.

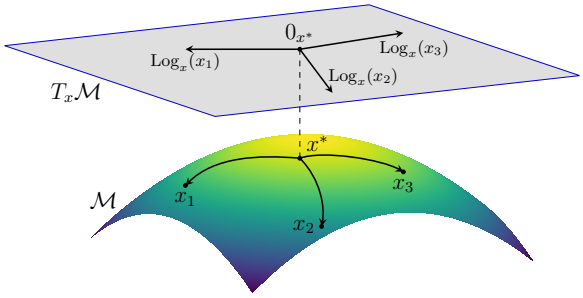

Fig. 1. An artist view of the Riemannian l^ν-barycenter. The figure represents a collection of points $\{x_i\}_{i=1}^3 \subset \mathcal{M}$ and their Riemannian l^ν-barycenter x^*. If $\nu \geq 2$, x^* must be a first-order stationary point of (1), i.e., satisfies (4).

3 Minorizing and majorizing the objective function

Assume one knows two functions $\widehat{m}_g, \widehat{M}_g : \mathcal{M}^2 \to \mathbb{R}$ such that for all pairs $x, \widetilde{x} \in \mathcal{M}$, we have

$$\widehat{m}_g(x,\widetilde{x}) \leq d_g(x,\widetilde{x}) \leq \widehat{M}_g(x,\widetilde{x}). \tag{5}$$

If $\mathcal{M}$ is embedded in the Euclidean space $\mathbb{R}^M$, then $\widehat{m}_g$ and $\widehat{M}_g$ can often be chosen as functions of the Euclidean distance $\|x-\widetilde{x}\|_{\mathrm{E}}$. We give explicit examples of such functions on the Stiefel manifold $\mathrm{St}(n,p)$, $n \geq 2p$, endowed with the so-called β-metric [12, Eq. 2.1] in [12, Thm. 7.1]:

$$\widehat{m}_\beta(X,\widetilde{X}) = \min\{1,\sqrt{\beta}\}2\sqrt{p}\arcsin\left(\frac{\|X-\widetilde{X}\|_{\mathrm{E}}}{2\sqrt{p}}\right),$$

$$\text{and}\quad \widehat{M}_\beta(X,\widetilde{X}) = \max\{1,\sqrt{\beta}\}\begin{cases} 2\arcsin\left(\frac{\|X-\widetilde{X}\|_{\mathrm{E}}}{2}\right) & \text{if } \|X-\widetilde{X}\|_{\mathrm{E}} \leq 2, \\ \frac{\pi}{2}\|X-\widetilde{X}\|_{\mathrm{E}} & \text{otherwise.} \end{cases}$$

Let us define under- and over-approximations of f_ν from (1) by

$$\widehat{f}_\nu(x) \coloneqq \left(\sum_{i=1}^{N} \widehat{m}_g(x, x_i)^\nu\right)^{\frac{1}{\nu}} \quad \text{and} \quad \widehat{F}_\nu(x) \coloneqq \left(\sum_{i=1}^{N} \widehat{M}_g(x, x_i)^\nu\right)^{\frac{1}{\nu}}.$$

It is clear by (5) that for all $x \in \mathcal{M}$, we have $\widehat{f}_\nu(x) \leq f_\nu(x) \leq \widehat{F}_\nu(x)$.

A choice for h_ν in (2) that offers theoretical guarantees is the minimization of $\widehat{f}_\nu$. Indeed, by setting $h_\nu \coloneqq \widehat{f}_\nu$, it follows by the optimality properties of a Riemannian l^ν-barycenter x^* and its approximation $\widehat{x}^* \in \operatorname{argmin}_{x\in\mathcal{M}} \widehat{f}_\nu(x)$ that the following sequence of inequalities holds:

$$\widehat{f}_\nu(\widehat{x}^*) \leq \widehat{f}_\nu(x^*) \leq f_\nu(x^*) \leq f_\nu(\widehat{x}^*) \leq \widehat{F}_\nu(\widehat{x}^*). \tag{6}$$

By (6), we can bound from above the functional relative error between $f_\nu(x^*)$ and $f_\nu(\widehat{x}^*)$. This is shown in Theorem 1.

Theorem 1. *Let $\{x_i\}_{i=1}^N \subset \mathcal{M}$, f_ν and x^* be defined as in (1). Moreover, let $\widehat{x}^* \in \operatorname{argmin}_{x\in\mathcal{M}} \widehat{f}_\nu(x)$, then*

$$\frac{f_\nu(\widehat{x}^*) - f_\nu(x^*)}{f_\nu(x^*)} \leq \frac{\widehat{F}_\nu(\widehat{x}^*) - \widehat{f}_\nu(\widehat{x}^*)}{\widehat{f}_\nu(\widehat{x}^*)}. \tag{7}$$

Proof. By (6), we have $f_\nu(\widehat{x}^*) - f_\nu(x^*) \leq \widehat{F}_\nu(\widehat{x}^*) - \widehat{f}_\nu(\widehat{x}^*)$. Moreover, it also holds that $\widehat{f}_\nu(\widehat{x}^*) \leq f_\nu(x^*)$. This concludes the proof.

The upper bound from Theorem 1 only depends on $\widehat{f}_\nu, \widehat{F}_\nu$ and $\widehat{x}^*$, which can all be computed at low computational cost. Moreover, the upper bound (7) can be small in practice, as exemplified in section 5.

For any function h_ν in (2), the distance $d_g(x^*, \widehat{x}^*)$ between a barycenter x^* and its approximation $\widehat{x}^*$ can be bounded from above by the computable quantity $N^{-\frac{1}{\nu}} 2\widehat{F}_\nu(\widehat{x}^*)$, as shown in Theorem 2.

Theorem 2. *Let $\{x_i\}_{i=1}^N \subset \mathcal{M}$, f_ν and x^* be defined as in (1). Then, for all $x \in \mathcal{M}$, we have $d_g(x^*, x) \leq N^{-\frac{1}{\nu}} 2 f_\nu(x) \leq N^{-\frac{1}{\nu}} 2\widehat{F}_\nu(x)$.*

Proof. By the triangle inequality, for $i = 1, ..., N$, we have $d_g(x^*, x) \leq d_g(x^*, x_i) + d_g(x_i, x)$. Therefore, by usual ν-norm inequalities in vector spaces, it follows that

$$\begin{aligned} d_g(x^*, x) &\leq \frac{1}{N}\left(\sum_{i=1}^{N} d_g(x^*, x_i) + \sum_{i=1}^{N} d_g(x_i, x)\right) \\ &\leq \frac{1}{N}\left[N^{1-\frac{1}{\nu}}\left(\sum_{i=1}^{N} d_g(x^*, x_i)^\nu\right)^{\frac{1}{\nu}} + N^{1-\frac{1}{\nu}}\left(\sum_{i=1}^{N} d_g(x, x_i)^\nu\right)^{\frac{1}{\nu}}\right] \\ &= N^{-\frac{1}{\nu}}(f_\nu(x^*) + f_\nu(x)) \leq N^{-\frac{1}{\nu}} 2 f_\nu(x). \end{aligned}$$

To conclude, it is enough to acknowledge that $f_\nu(x) \leq \widehat{F}_\nu(x)$.

In particular, Theorem 2 yields $d_g(x^*, \widehat{x}^*) \leq N^{-\frac{1}{\nu}} 2\widehat{F}_\nu(\widehat{x}^*)$. Therefore, setting $h_\nu = \widehat{F}_\nu$ in (2) and thus $\widehat{x}^* \in \operatorname{argmin}_{x\in\mathcal{M}} \widehat{F}_\nu(x)$ minimizes the upper bound on $d_g(x^*, \widehat{x}^*)$ from Theorem 2.

Similarly, the quantity $d_g(x^*, \widehat{x}^*)$ can be bounded from below. However, in comparison with Theorem 2, the lower bound from Theorem 3 can not be obtained without knowing x^*.

Theorem 3. *Let $\{x_i\}_{i=1}^N \subset \mathcal{M}$, f_ν and x^* be defined as in (1). Then, for all $x \in \mathcal{M}$, we have $d_g(x^*, x) \geq N^{-1}(f_\nu(x) - f_\nu(x^*))$.*

Proof. By the reverse triangle inequality, for $i = 1, ..., N$, we have $d_g(x^*, x) \geq |d_g(x^*, x_i) - d_g(x_i, x)|$. Therefore, it follows that

$$\begin{aligned} d_g(x^*, x) &\geq \frac{1}{N}\sum_{i=1}^N |d_g(x_i, x) - d_g(x^*, x_i)| \geq \frac{1}{N}\left(\sum_{i=1}^N |d_g(x_i, x) - d_g(x^*, x_i)|^\nu\right)^{\frac{1}{\nu}} \\ &\geq \frac{1}{N}\left[\left(\sum_{i=1}^N |d_g(x, x_i)|^\nu\right)^{\frac{1}{\nu}} - \left(\sum_{i=1}^N |d_g(x^*, x_i)|^\nu\right)^{\frac{1}{\nu}}\right] \qquad (8) \\ &= N^{-1}(f_\nu(x) - f_\nu(x^*)), \end{aligned}$$

where (8) holds by the reverse triangle inequality.

In view of Theorem 3, we have $d_g(x^*, \widehat{x}^*) \geq N^{-1}(f_\nu(\widehat{x}^*) - f_\nu(x^*))$.

To conclude this section, we briefly discuss the case of linear bounds $\widehat{m}_g$ and $\widehat{M}_g$ in terms of the Euclidean distance, i.e., there are $l_M \geq l_m > 0$ such that $\widehat{m}_g(x, \widetilde{x}) = l_m\|x - \widetilde{x}\|_{\mathrm{E}}$ and $\widehat{M}_g(x, \widetilde{x}) = l_M\|x - \widetilde{x}\|_{\mathrm{E}}$. Then, by choosing $\widehat{x}^* \in \operatorname{argmin}_{x\in\mathcal{M}} \widehat{f}_\nu(x)$, Theorem 1 yields

$$\frac{f_\nu(\widehat{x}^*) - f_\nu(x^*)}{f_\nu(x^*)} \leq \frac{l_M - l_m}{l_m} = \frac{l_M}{l_m} - 1.$$

For example, on the Stiefel manifold $\mathrm{St}(n, p)$, $n \geq 2p$, we have $l_m = \min\{1, \sqrt{\beta}\}$ and $l_M = \max\{1, \sqrt{\beta}\}\frac{\pi}{2}$ [12, Cor. 7.2]. Moreover, if $\nu = 2$, it is easily shown that $\widehat{x}^*$ is a projection of $\overline{x} = \frac{1}{N}\sum_{i=1}^N x_i$ on $\mathcal{M}$. Indeed, we have

$$\widehat{x}^* \in \operatorname*{argmin}_{x\in\mathcal{M}} \sum_{i=1}^N \|x - x_i\|_{\mathrm{E}}^2 = \operatorname*{argmin}_{x\in\mathcal{M}} \|x - \overline{x}\|_{\mathrm{E}}^2 =: \mathrm{Proj}_{\mathcal{M}}(\overline{x}).$$

Finally, $\mathrm{Proj}_{\mathcal{M}}(\overline{x})$ can be efficiently computed on many manifolds such as the Grassmann manifold, the Stiefel manifold and the manifold of fixed-rank matrices. On the Stiefel manifold, it is given by the orthogonal factor of the polar decomposition. However, nonlinear bounds in (5) usually offer better approximations of the true Riemannian l^ν barycenter x^*.

4 Lifting maps as approximations of the Riemannian logarithm

Retractions offer a framework in Riemannian optimization to approximate the Riemannian exponential at low computational cost [1]. A retraction R is a smooth family of mappings $\{R_x : T_x\mathcal{M} \to \mathcal{M}\}_{x\in\mathcal{M}}$ such that for all $x \in \mathcal{M}$, $\xi \in T_x\mathcal{M}$, $t \geq 0$, the retraction approximates the Riemannnian exponential at first order: $R_x(t\xi) = \mathrm{Exp}_x(t\xi) + \mathrm{o}(t)$.

Similarly, *lifting maps*, also called tangent-bundle maps, approximate the Riemannian logarithm [9,16,5]. A lifting map L is a smooth family of mappings $\{L_x : \mathcal{M} \to T_x\mathcal{M}\}_{x\in\mathcal{M}}$ such that for all $x \in \mathcal{M}, \xi \in T_x\mathcal{M}$ and $t \geq 0$ small enough, $L_x(\mathrm{Exp}_x(t\xi)) = t\xi + \mathrm{o}(t)$. As a result,

$$d_g(x, \mathrm{Exp}_x(t\xi)) = \|L_x(\mathrm{Exp}_x(t\xi))\|_g + \mathrm{o}(t). \tag{9}$$

In a neighborhood $\mathcal{V} \subset \mathcal{M}$ of x, a lifting map can be obtained by taking the inverse of a retraction, i.e., $L_x \coloneqq R_x^{-1}$. An example of such a lifting map is given by the orthogonal projection on the tangent space $T_x\mathcal{M}$: $L_x(\widetilde{x}) \coloneqq \mathrm{Proj}_{T_x\mathcal{M}}(\widetilde{x} - x)$. This map is locally the inverse of the so-called *orthographic retraction* [10].

To approximate the Riemannian l^ν-barycenter, the approach from [10,9,6] is to consider the limit of the fixed-point iteration

$$x^{k+1} = R_{x^k}\left(\sum_{i=1}^{N} \nu \|L_{x^k}(x_i)\|_g^{\nu-2} L_{x^k}(x_i)\right) \text{ for } k = 0, 1, 2...., \tag{10}$$

which approximates the stationarity condition (4). However, for $d_g(x, \widetilde{x})$ large enough, $L_x(\widetilde{x})$ can be a very poor approximation of $\mathrm{Log}_x(\widetilde{x})$, as illustrated in Figure 2. An additional caveat is the difficulty to define the mapping $L_x(\widetilde{x})$ for all $\widetilde{x} \in \mathcal{M}$. For example, on the Stiefel manifold, the inverses of the polar retraction, the QR retraction and the Cayley retraction have guaranteed existence only in a neighborhood of x [10].

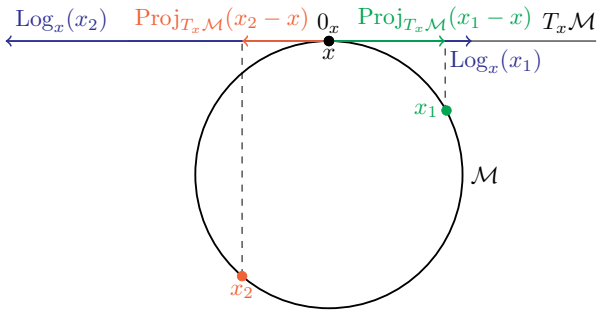

Fig. 2. While the lifting map approximates well the Riemannian logarithm for points x and x_1 that are close enough, they can differ significantly if the distance between the points x and x_2 is large.

5 Numerical experiments on the Stiefel manifold

We conduct numerical experiments on the Stiefel manifold $\mathrm{St}(n,p)$ endowed with the β-metric [12] for different values of N, ν, n, p and β. The code to reproduce these experiments is available at https://github.com/smataigne/Approximate_barycenter.jl.

We compare three algorithms: (A1) is the Riemannian gradient descent (RGD) on $(f_\nu)^\nu$, (A2) is the RGD on $(\widehat{f}_\nu)^\nu$ and (A3) is the fixed point iteration from (10) with R_X being the popular QR retraction and L_X being the orthogonal projection on $T_X\mathrm{St}(n,p)$. The Riemannian logarithms in (A1) are numerically estimated using [17, Alg. 2 and 4]. The sequences generated by (A1), (A2) and (A3) are respectively denoted by $\{X^k\}_{k=1}^*$, $\{Y^k\}_{k=1}^*$ and $\{Z^k\}_{k=1}^*$ with $X^0 = Y^0 = Z^0$.

The stopping criterion is as follows: (A1) runs until $|f_\nu(X^*)^\nu - f_\nu(X^{*-1})^\nu| \leq \varepsilon f_\nu(X^0)^\nu$, (A2) runs until $|\widehat{f}_\nu(Y^*)^\nu - \widehat{f}_\nu(Y^{*-1})^\nu| \leq \varepsilon \widehat{f}_\nu(Y^0)^\nu$ and (A3) runs until $|g_\nu(Z^*)^\nu - g_\nu(Z^{*-1})^\nu| \leq \varepsilon g_\nu(Z^0)^\nu$ where we define the function $g_\nu(Z)^\nu := \sum_{i=1}^N \|\mathrm{Proj}_{T_Z\mathrm{St}(n,p)}(X_i - Z)\|_\beta^\nu$. We choose $\varepsilon = 10^{-6}$.

We consider spread and clustered datasets $\{X_i\}_{i=1}^N$: spread matrices are sampled with uniform distribution on $\mathrm{St}(n,p)$ while clustered matrices are sampled such that their pairwise distances do not exceed the known (when $\beta = 1$ [18]) or suspected (when $\beta = \frac{1}{2}$ [2]) value of the injectivity radius.

The results of the experiments are given in Table 1, where they are compared with the bounds of Theorem 1 and Theorem 2. In these experiments, we can observe that (i) both (A2) and (A3) provide quantitatively accurate approximations of the barycenter computed by (A1), (ii) the bounds are satisfied and (iii) the bounds give a better estimate of the error measures when the dataset is clustered.

$f_\nu(X^*)$	$f_\nu(Y^*)$	$f_\nu(Z^*)$	$\frac{f_\nu(X^*)-f_\nu(Y^*)}{f_\nu(X^*)} \leq \frac{\widehat{F}_\nu(Y^*)-\widehat{f}_\nu(Y^*)}{\widehat{f}_\nu(Y^*)}$	$d_g(X^*,Y^*) \leq N^{-\frac{1}{\nu}} 2\widehat{F}_\nu(Y^*)$
Clustered dataset, $(N,n,p,\beta,\nu) = (5,50,20,1,1)$				
4.384	4.384	4.384	1.114e−6 $\leq$ 0.03344	0.001613 $\leq$ 1.811
Spread dataset, $(N,n,p,\beta,\nu) = (5,50,20,1,2)$				
11.59	11.66	11.66	0.005723 $\leq$ 0.4886	0.7046 $\leq$ 15.03
Clustered dataset, $(N,n,p,\beta,\nu) = (10,100,40,0.5,1)$				
8.426	8.427	8.427	9.751e−5 $\leq$ 0.4707	0.01176 $\leq$ 1.953
Spread dataset, $(N,n,p,\beta,\nu) = (10,100,40,0.5,2)$				
23.50	24.23	24.23	0.03116 $\leq$ 1.076	3.274 $\leq$ 23.52

Table 1. Average results over 10 runs of the performances of algorithms (A1), (A2) and (A3) for the computation of an approximate Riemannian l^ν-barycenter.

In Figure 3, we show convergence curves of algorithms (A1), (A2) and (A3) for $(N,n,p,\beta,\nu) = (5,100,40,\frac{1}{2},2)$ on a spread dataset. We also show the evolution of the running times as the size p of $\mathrm{St}(140,p)$ increases. We can observe that the computational costs of (A2) and (A3) are significantly smaller than the

cost of (A1) (by a factor larger than 100) and that (A2) and (A3) have similar computational costs. For large values of N, n and p, (A1) is not tractable.

In conclusion, both (A2) and (A3) are computationally efficient approaches providing accurate approximations of the Riemannian l^ν-barycenter. They are thus worthwhile alternatives to (A1). Moreover, (A2) benefits from theoretical guarantees (Theorem 1) for spread datasets where the behavior of (A3) is not known in general.

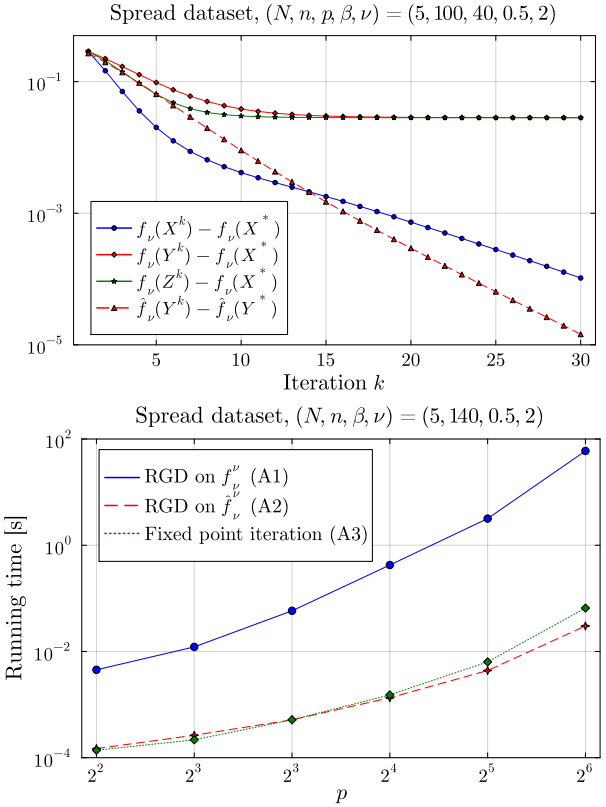

Fig. 3. At the top: convergence curves of algorithms (A1), (A2) and (A3) to obtain the results of Table 1. For each algorithm, the evolution of f_ν is represented with solid lines. The dashed line represents, for $\{Y^k\}_{k=1}^*$, the evolution of the optimized objective function $\widehat{f}_\nu$. At the bottom: a comparison of the evolution of the running time of the three algorithms as the size p of $\mathrm{St}(140, p)$ increases using BenchmarkTools.jl.

Acknowledgments. Simon Mataigne is a Research Fellow of the Fonds de la Recherche Scientifique - FNRS. This work was supported by the Fonds de la Recherche Scientifique - FNRS under grant no T.0001.23. Nina Miolane acknowledges funding from the NSF Career 2240158.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Absil, P.-A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ (2008), http://press.princeton.edu/titles/8586.html
2. Absil, P.-A., Mataigne, S.: The Ultimate Upper Bound on the Injectivity Radius of the Stiefel Manifold. SIAM Journal on Matrix Analysis and Applications **46**(2), 1145–1167 (2025), https://doi.org/10.1137/24M1644808
3. Afsari, B.: Riemannian L^p center of mass: Existence, uniqueness, and convexity. Proceedings of the American Mathematical Society **139**(2), 655–673 (2011), http://www.jstor.org/stable/41059320
4. Afsari, B., Tron, R., Vidal, R.: On the convergence of gradient descent for finding the Riemannian center of mass. SIAM Journal on Control and Optimization **51**(3), 2230–2260 (2013), https://doi.org/10.1137/12086282X
5. Barbero-Liñán, M., de Diego, D.M.: Retraction maps: A seed of geometric integrators. Foundations of Computational Mathematics **23**(4), 1335–1380 (2023), https://doi.org/10.1007/s10208-022-09571-x
6. Bouchard, F., Laurent, N., Said, S., Le Bihan, N.: Beyond R-barycenters: an effective averaging method on Stiefel and Grassmann manifolds (2025), https://arxiv.org/abs/2501.11555
7. Chakraborty, R., Vemuri, B.C.: Statistics on the Stiefel manifold: Theory and applications. The Annals of Statistics **47**(1), 415 – 438 (2019), https://doi.org/10.1214/18-AOS1692
8. Edelman, A., Arias, T.A., Smith, S.T.: The Geometry of Algorithms with Orthogonality Constraints. SIAM Journal on Matrix Analysis and Applications **20**(2), 303–353 (1998), https://doi.org/10.1137/S0895479895290954
9. Fiori, S., Kaneko, T., Tanaka, T.: Tangent-bundle maps on the Grassmann manifold: Application to empirical arithmetic averaging. IEEE Transactions on Signal Processing **63**(1), 155–168 (2015), https://doi.org/10.1109/TSP.2014.2365764
10. Kaneko, T., Fiori, S., Tanaka, T.: Empirical arithmetic averaging over the compact Stiefel manifold. IEEE Transactions on Signal Processing **61**, 883–894 (02 2013), https://doi.org/10.1109/TSP.2012.2226167
11. Karcher, H.: Riemannian Center of Mass and so called Karcher mean (2014), https://arxiv.org/abs/1407.2087
12. Mataigne, S., Absil, P.-A., Miolane, N.: Bounds on the geodesic distances on the Stiefel manifold for a family of Riemannian metrics (2024), https://arxiv.org/abs/2408.07072
13. Mataigne, S., Zimmermann, R., Miolane, N.: An Efficient Algorithm for the Riemannian Logarithm on the Stiefel Manifold for a Family of Riemannian Metrics. SIAM Journal on Matrix Analysis and Applications **46**(2), 879–905 (2025), https://doi.org/10.1137/24M1647801
14. Pennec, X.: Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. Journal of Mathematical Imaging and Vision **25**(1), 127–154 (2006), https://doi.org/10.1007/s10851-006-6228-4, a preliminary appeared as INRIA RR-5093, January 2004.
15. Sakai, T.: Riemannian Geometry. Fields Institute Communications, American Mathematical Society (1996), https://books.google.be/books?id=ODDyngEACAAJ
16. Zhu, X., Sato, H.: Riemannian conjugate gradient methods with inverse retraction. Computational Optimization and Applications **77**(3), 779–810 (2020), https://doi.org/10.1007/s10589-020-00219-6

17. Zimmermann, R., Hüper, K.: Computing the Riemannian logarithm on the Stiefel manifold: Metrics, methods, and performance. SIAM Journal on Matrix Analysis and Applications **43**(2), 953–980 (2022), https://doi.org/10.1137/21M1425426
18. Zimmermann, R., Stoye, J.: The Injectivity Radius of the Compact Stiefel Manifold under the Euclidean Metric. SIAM Journal on Matrix Analysis and Applications **46**(1), 298–309 (2025), https://doi.org/10.1137/24M1663818