

FAST k -MEANS CLUSTERING IN RIEMANNIAN MANIFOLDS VIA FRÉCHET MAPS: APPLICATIONS TO LARGE-DIMENSIONAL SPD MATRICES

JI SHI, NICOLAS CHARON, ANDREAS MANG, DEMETRIO LABATE, AND ROBERT AZENCOTT

Department of Mathematics, University of Houston, Houston, TX, US

ABSTRACT. We introduce a novel, efficient framework for clustering data on high-dimensional, non-Euclidean manifolds that overcomes the computational challenges associated with standard intrinsic methods. The key innovation is the use of the p -Fréchet map $F^p : \mathcal{M} \rightarrow \mathbb{R}^\ell$ – defined on a generic metric space $\mathcal{M}$ – which embeds the manifold data into a lower-dimensional Euclidean space $\mathbb{R}^\ell$ using a set of reference points $\{r_i\}_{i=1}^\ell$, $r_i \in \mathcal{M}$. Once embedded, we can efficiently and accurately apply standard Euclidean clustering techniques such as k -means. We rigorously analyze the mathematical properties of F^p in the Euclidean space and the challenging manifold of $n \times n$ symmetric positive definite matrices $SPD(n)$. Extensive numerical experiments using synthetic and real $SPD(n)$ data demonstrate significant performance gains: our method reduces runtime by up to two orders of magnitude compared to intrinsic manifold-based approaches, all while maintaining high clustering accuracy, including scenarios where existing alternative methods struggle or fail.

1. INTRODUCTION

Cluster analysis is one of the most common tasks in modern data science. Given a set of observations in some data space, being able to identify specific subgroups based on similarity patterns across the dataset is often a key stepping stone for data exploration and further analysis. Unlike classification, clustering is a fundamentally unsupervised problem as no labels are available a priori. Thus, the extraction of clusters can only be achieved by relying on some notion of proximity between data points, which is typically measured via a certain distance defined on the data space. Perhaps the most widely used general clustering method is the well-known k -means algorithm, in which the partition of the dataset is built to minimize its resulting dispersion. Although this problem cannot be solved exactly in polynomial time, several greedy approximate schemes have been proposed, in particular Lloyd’s celebrated k -means algorithm [19], which is widely used in applications. One clear upside of Lloyd’s algorithm is its simplicity: the whole scheme consists of a cluster assignment step followed by an update of the cluster centroids, which are iterated until stabilization of the clusters. The first step is typically performed by assigning each data point to the closest current cluster centroid as measured by the distance in the data space. The second step involves the computation of the centroid of each of the clusters at the current iteration. For data living in an Euclidean space, the latter simply reduces to computing the mean of each cluster, and thus both steps of Lloyd’s scheme have closed-form updates, making the algorithm very fast to run even on spaces of large dimensions.

However, some obvious difficulties arise when, instead, data points live in a non-Euclidean manifold $\mathcal{M}$, for which the computation of the distance between two given points may require solving

Date: November 13, 2025.

2020 Mathematics Subject Classification. 62H30, 53Z50.

Key words and phrases. Clustering, Fréchet map k -means, Manifold data, Symmetric positive definite matrices.

This work was partly supported by the National Science Foundation (NSF) through the grant DMS-2438562 (NC) and DMS-2145845 (AM) and by the Simons Foundation grant MPS-TSM-00002738 (DL). Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the NSF. This work was completed in part with resources provided by the Research Computing Data Core at the University of Houston.

a geodesic boundary value problem. More importantly, even though theoretical extensions of the notion of centroid or mean (known as Fréchet or Karcher mean) exist for a certain class of manifolds $\mathcal{M}$, these are highly non-trivial and almost always involve numerically costly optimization on $\mathcal{M}$. Due to the need for repeated evaluations of such distances and means in Lloyd’s approach, its direct adaptation to the manifold setting, often referred to as intrinsic k-means [32], can easily become prohibitively expensive, even more so on high-dimensional manifolds. Yet manifold data is increasingly common in many applications. In robotics, for instance, pose configurations are typically modeled as products of elements in the manifold of 3D rotations. In signal/image analysis, one is often interested in correlation or covariance matrices of signals, which naturally live in the manifold of symmetric positive definite matrices $SPD(n)$. The fields of shape analysis and computer vision also consider data points (such as curves or surfaces) in inherently nonlinear spaces. We refer the reader to recent surveys such as [26] for more highlights on the growing importance of manifold data, and its associated challenges.

This paper introduces an efficient approach for k-means clustering of potentially large-dimensional non-Euclidean data, which we coin *Fréchet Map Clustering (FMC)*. The core idea is to map the data into some (preferably smaller) Euclidean space in which standard Euclidean clustering techniques can be applied at much lower cost compared to the original manifold $\mathcal{M}$. Specifically, our proposed approach relies on the special family of p -Fréchet maps $F^p : \mathcal{M} \rightarrow \mathbb{R}^\ell$ defined on a generic metric space $\mathcal{M}$ with distance $d_{\mathcal{M}}$ and parameterized by a set of reference points in $\mathcal{M}$. While this is an established strategy in modern machine learning with architectures such as autoencoders that are designed to learn a latent space data embedding tailored to a given task, we follow here a different paradigm by focusing on a particular and more restrictive, albeit more interpretable, class of mappings for which no prior training phase is needed.

Although in principle FMC can be applied to broad classes of manifolds or metric spaces, our main focus in this work is on the Riemannian manifold of $n \times n$ symmetric positive definite (**SPD**) matrices $SPD(n)$. SPD matrices, in particular, correlation matrices, appear in a range of applications, including high-dimensional statistics, image analysis, multi-sensor monitoring, and communication networks. In neuroimaging, for instance, correlation matrices of high dimensions derived from diffusion tensor imaging or functional magnetic resonance imaging are often employed to model the strength of neural connections between different brain sites and to assess brain function in normal and disease states [13, 14, 35]. In many of these problems, one is faced with the task of clustering large sets of SPD matrices, for example, high-dimensional correlation matrices, and we shall validate the potential of the FMC approach in this setting.

1.1. Contributions. We present a new method for efficiently clustering large-dimensional non-Euclidean data by embedding them into a Euclidean space using p -Fréchet maps, with a focus on the cases $p = 1$ and $p = 2$. The key contributions of this work are the following.

- We proved several fundamental properties of the p -Fréchet map F^p (for general p and, in particular, for the specific choices of $p = 1, 2$) when defined on a Euclidean space. These include the differentiability of F^p and control bounds on the induced distortion, as well as sufficient conditions under which F^p is a diffeomorphism onto its image. We further analyze the question of separability of the image of balls under Fréchet maps, a key property when performing clustering in the image space.
- We examined the extension of this theoretical analysis to the significantly more challenging case of the $SPD(n)$ manifold, highlighting in particular the important questions remaining open in this setting.
- We proposed a principled, yet practical strategy for selecting the reference points that control the map F^p while maximizing stability and clustering accuracy.
- We empirically evaluated the proposed method against existing approaches for clustering data on SPD manifolds using both synthetic data and experimental data.

Our results confirm that **the proposed methodology can efficiently cluster data in Riemannian manifolds without significantly sacrificing accuracy**. The proposed approach is up to **two orders of magnitude faster** than the standard intrinsic k -means on the manifold $\mathcal{M}$, and we obtain **runtimes that are competitive with similar strategies** operating in the tangent space, while **producing high-accuracy clustering results more consistently**.

1.2. Limitations. Despite having established some clear advantages of the FMC framework, there are unresolved issues that will be the subject of future work, including the following.

- Although we have a rather complete picture of the mathematical properties of Fréchet maps in the Euclidean case, there are important theoretical gaps to generalize some of those results to $SPD(n)$ and to the broader class of *Cartan–Hadamard manifolds*.
- We empirically observed that the choice of reference points has a significant impact on the performance of FMC. Although we have established a principled heuristic for selecting reference points, which yields very competitive results, additional work is required to derive strategies with rigorous theoretical guarantees.

1.3. Related Work. In recent years, there has been significant interest in the design of geometry-aware methods for classifying or clustering data in non-Euclidean, high-dimensional spaces. Due to the potential intricacies and numerical cost associated with some basic operations in manifolds (the estimation of a mean, for instance), a common workaround consists of embedding data points into some Euclidean space in a way that preserves as much of the original structure in the dataset as possible. Some very popular strategies involve the use of metric multidimensional scaling (MDS) [22] or t-distributed stochastic neighbor embedding (t-SNE) [16]. A downside of both of these methods, however, is that they require the computation of all distances between every pair of data points, which can represent a non-negligible amount of computations in the manifold case. Furthermore, recent works such as [6] have suggested that t-SNE embeddings may, in some situations, generate artefactual clusters.

Another approach, which is more closely related to ours, especially with regard to $SPD(n)$, is the *log-Euclidean* framework of [4, 25]. In its standard form, the log-Euclidean setting consists of a linearized tangent space approximation of the metric at a specific template point (usually the Fréchet mean of the dataset). This results in a fast-to-compute and often efficient data embedding strategy to perform clustering. We demonstrate that FMC is competitive with the log-Euclidean framework in terms of runtime and clustering accuracy. More importantly, we show empirically that it maintains good clustering accuracy in settings where the performance of the log-Euclidean framework deteriorates.

Lastly, we will also compare our approach to the generic *Intrinsic Riemannian Clustering*, i.e., the direct transposition of Lloyd’s algorithm to the SPD manifold setting. For the computation of the clusters’ Fréchet means, in addition to the standard gradient descent method from [2], we also consider alternative faster approximation schemes such as the *iterative centroid method proposed* in [17, 10] based on the recursive scheme from [31].

1.4. Layout. We provide background material in Section 2. We start by briefly discussing the problem of k -means clustering for manifold data in Section 2.1. We introduce the general notion of Fréchet mapping in Section 2.2. We explore the properties of the Fréchet mapping for $SPD(n)$ in Section 3. This includes the structure of $SPD(n)$ (see Section 3.1), properties of F^p for $SPD(n)$ (see Section 3.2), a comparison of different clustering methods in $SPD(n)$ (see Section 3.3). We describe our approach for selecting reference points for F^p in Section 3.4. Numerical results are reported in Section 4. We conclude with Section 5.

2. BACKGROUND AND GENERAL FRAMEWORK

Before we provide some theoretical insights, we introduce the problem setting as well as notations, and define the Fréchet map more precisely.

2.1. k-means clustering for manifold data. Clustering is a fundamental problem of machine learning. From a set of observations in a data manifold $\mathcal{M}$ its goal is to partition these observations into a set of meaningful clusters $\mathcal{P} = (CL_1, \dots, CL_k)$, where each $CL_j, j = 1, \dots, k$, is determined by measuring the proximity between observations using a distance on $\mathcal{M}$.

Among existing clustering approaches, the k -means algorithm [20] remains among the most widely used methods due to the simplicity of its formulation and its ability to easily adapt to different types of data. Given a finite set of observations $\mathcal{D} = \{x_i\}_{i=1}^n \subset \mathcal{M}$ and a fixed number $k \in \mathbb{N}$ of target clusters, the k -means algorithm looks for a partition $\mathcal{P}$ that minimizes the *total dispersion* defined by

$$(1) \quad \text{totdisp}(\mathcal{P}) = \sum_{i=1}^k \text{disp}(CL_i) = \sum_{i=1}^k \frac{1}{|CL_i|} \sum_{x \in CL_i} d_{\mathcal{M}}(x, c_i)^2,$$

where $d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ denotes the distance on the data manifold being considered, $c_i \in \mathcal{M}$ is the centroid or “mean” of the cluster CL_i in the partition $\mathcal{P}$, and $|CL_i|$ is the cardinality of the cluster CL_i . To make the above definition precise, one needs to make more specific assumptions on the structure of the data manifold $\mathcal{M}$. A typical setup is to consider a Riemannian manifold $\mathcal{M}$ with a corresponding Riemannian distance $d_{\mathcal{M}}$, in which case one can extend the notion of Euclidean mean via the so-called Fréchet mean (also known as the Fréchet mean). Leaving aside for now the question of the existence and uniqueness of the Fréchet mean in a Riemannian manifold, the total dispersion in Equation (1) can be interpreted as the sum of the variances of each cluster.

Although there is only a finite number of possible partitions $\mathcal{P}$ of $\mathcal{D}$, this number grows exponentially with the number of observations, so that finding a global minimum of $\text{totdisp}(\mathcal{P})$ is an NP-hard problem. A practical alternative to address this problem is the greedy iterative approach known as Lloyd’s algorithm [19], which is generically referred to as the k -means algorithm. The classical Lloyd’s algorithm in $\mathbb{R}^m$ is given as Algorithm 1.

Algorithm 1 Lloyd’s algorithm [19].

- 1: **Input:** A set of points $\mathcal{D} = \{x_1, \dots, x_N\} \subset \mathbb{R}^m$ and an initialization of the cluster centroids $C = \{c_1, \dots, c_k\} \subset \mathbb{R}^m$
 - 2: stop $\leftarrow$ false
 - 3: **while** $\neg$ stop **do**
 - 4: Assign each $x \in \mathcal{D}$ to the cluster with closest center, i.e., $x \in CL_i$ for

$$(2) \quad i = \arg \min_{j=1, \dots, k} d(x, c_j),$$
 where d is the Euclidean distance in $\mathbb{R}^m$
 - 5: Recalculate the cluster centroids C by setting for each $i = 1, \dots, k$,

$$(3) \quad c_i \leftarrow \frac{1}{|CL_i|} \sum_{x \in CL_i} x$$
 - 6: stop $\leftarrow$ check convergence
 - 7: **end while**
 - 8: **Output:** Cluster centroids $C = \{c_1, \dots, c_k\}$ and corresponding clusters $CL_1, \dots, CL_k$
-

Remark 1. Since the cluster assignment in Algorithm 1 is based on the proximity to the cluster centroids c_i ’s, it leads to a partition of the whole space $\mathbb{R}^m$ delimited by $k(k+1)/2$ mediatrices

hyperplanes $\{x : \|x - c_i\| = \|x - c_j\|\}$ for all pairs of cluster centers (c_i, c_j) . This shows that the algorithm separates the different clusters through affine hyperplanes in $\mathbb{R}^m$ that are the mediatrices of the cluster centroids.

Lloyd’s algorithm in $\mathbb{R}^m$ can be shown to converge, but not necessarily to a global minimizer of the total dispersion. In practice, it is common to run the algorithm for multiple different initializations of the centroids and ultimately select the solution with the lowest total dispersion, which generally provides a good estimate of the true solution.

However, extending Lloyd’s algorithm from $\mathbb{R}^m$ to a more general Riemannian manifold $\mathcal{M}$ is not as straightforward. Specifically, the centroid computation, which is simply an arithmetic mean in Equation (3), must be replaced by a Fréchet mean on $\mathcal{M}$ and typically requires additional conditions on the manifold or the dataset $\mathcal{D}$.

Indeed, the existence and uniqueness of the Fréchet mean of a set of points in $\mathcal{M}$ is not guaranteed for general Riemannian manifolds. This can be ensured either by assuming that data points in $\mathcal{D}$ are sufficiently concentrated or by considering a more specific structure for $\mathcal{M}$. A particularly well-suited class of manifolds in that regard is the *Cartan–Hadamard manifolds*, which are the simply connected complete Riemannian manifolds with non-positive sectional curvature. In this case, one can ensure the existence and uniqueness of geodesics between any two points in $\mathcal{M}$ as well as the existence and uniqueness of Fréchet means for any set of points in $\mathcal{M}$, cf. [29, 27, 1]. Cartan–Hadamard manifolds encompass many interesting examples of data manifolds beyond Euclidean spaces that are found in applications, including the space of SPD matrices, which will be the focus of this paper.

Despite the Cartan–Hadamard manifold providing an adequate theoretical setting to extend the k -means algorithm to manifolds, there are still important practical difficulties compared to the Euclidean case. For some of those manifolds, an explicit expression of the geodesic distance may not be available, in which case one needs to solve a geodesic search problem to compute a distance between two points. More importantly, even in cases where the distance can be evaluated in a closed form (such as $SPD(n)$, c.f. Section 3.1 below), there is in general no closed-form solution for the Fréchet mean of a given set of points in $\mathcal{M}$. As a result, to find a centroid in Equation (3), one needs to find an approximate solution of

$$(4) \quad c_i = \operatorname{argmin}_{p \in \mathcal{M}} \sum_{x \in CL_i} d_{\mathcal{M}}(x, p)^2.$$

In a Cartan–Hadamard manifold, (4) is a convex problem which can be tackled using various optimization strategies, the simplest one being the standard gradient descent method, c.f. [2]. However, due to the large number of Fréchet means that need to be estimated throughout the iterations of k -means, this computation could become prohibitively costly for clustering. Some alternative strategies to solve eq. (4) include stochastic gradient descent [3, 7] or the recursive barycenter scheme [31, 17]. Even these approaches are usually not sufficient on their own to make k -means clustering efficiently tractable for large-dimensional manifolds. One of the key contributions of our work is a novel methodology that addresses this challenge.

2.2. The Fréchet mapping. In this section, we present the main idea underpinning the proposed approach.

2.2.1. General setting and basic properties. The p -Fréchet map F^p on a generic metric space $\mathcal{M}$ with distance $d_{\mathcal{M}}$ is defined as follows.

Definition 1. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space and fix a set of ℓ points $\{r_1, \dots, r_{\ell}\} \subset \mathcal{M}$. For $p \in \mathbb{R}$, $p \geq 1$, we define the p -Fréchet map associated with the list of reference points $r = (r_1, \dots, r_{\ell})$

by

$$\begin{aligned} F_r^p : \mathcal{M} &\rightarrow \mathbb{R}_+^\ell \\ x &\mapsto (d_{\mathcal{M}}(r_1, x)^p, \dots, d_{\mathcal{M}}(r_\ell, x)^p). \end{aligned}$$

In some cases, when the dependency on the reference point or order p does not need to be emphasized, we shall abbreviate the notation for the Fréchet map to F^p or even simply F .

In this work, we will be interested in two particular values of p , namely $p = 1$ and $p = 2$. The case $p = 1$ corresponds to the standard definition of the Fréchet map introduced in the field of discrete geometry [8, 21], where it has played a particular role in the construction of quasi-isometric embeddings of finite metric spaces into Euclidean spaces. The case $p = 2$ corresponds to the squared Fréchet mapping and offers the advantage of being differentiable everywhere when $\mathcal{M}$ is a smooth Riemannian manifold, together with certain convexity properties. We note that the general p -Fréchet map is well-defined for any metric space $(\mathcal{M}, d_{\mathcal{M}})$ as the only requirement is that the distance between any two points exists.

Following the idea of performing data clustering by applying the k-means algorithm in $\mathbb{R}_+^\ell$ on the images of the data points via F_r , it is important to investigate the properties of the p -Fréchet maps. Our first observation is that, in a general metric space $\mathcal{M}$, for any choice of reference points, the map F_r^p is locally Lipschitz.

Proposition 1. *Let F_r^p , $p \geq 1$, be a p -Fréchet map on a metric space $(\mathcal{M}, d_{\mathcal{M}})$ associated with a list of reference points $r = (r_1, \dots, r_\ell)$ in $\mathcal{M}$. For $p = 1$, the Fréchet map F_r^1 is a globally Lipschitz map on $\mathcal{M}$. If for some $x_0 \in \mathcal{M}$ and $\delta > 0$, one has $r_1, \dots, r_\ell \in B(x_0, \delta)$ then for any $x, x' \in B(x_0, \delta)$, it holds that $\|F_r^p(x) - F_r^p(x')\|_\infty \leq p 2^{p-1} \delta^{p-1} d_{\mathcal{M}}(x, x')$ for any $p \geq 1$.*

Proof. From the triangle inequality, we immediately see that $|d(x, r_i) - d(x', r_i)| \leq d(x, x')$ for all $i = 1, \dots, \ell$, and thus $\|F_r^1(x) - F_r^1(x')\|_\infty \leq d_{\mathcal{M}}(x, x')$ which proves that F_r^1 is 1-Lipschitz on $\mathcal{M}$. For $p \geq 1$, we observe that for each i :

$$\begin{aligned} |d_{\mathcal{M}}(x, r_i)^p - d_{\mathcal{M}}(x', r_i)^p| &\leq p \max\{d_{\mathcal{M}}(x, r_i), d_{\mathcal{M}}(x', r_i)\}^{p-1} |d_{\mathcal{M}}(x, r_i) - d_{\mathcal{M}}(x', r_i)| \\ &\leq p \max\{d_{\mathcal{M}}(x, r_i), d_{\mathcal{M}}(x', r_i)\}^{p-1} d_{\mathcal{M}}(x, x'). \end{aligned}$$

Since $d_{\mathcal{M}}(x, r_i) \leq d_{\mathcal{M}}(x, x_0) + d_{\mathcal{M}}(x_0, r_i) \leq 2\delta$ and similarly for x' , we obtain the stated upper bound. $\square$

Proposition 1 guarantees that the distortion induced by F_r^p from the original distance in $\mathcal{M}$ to the Euclidean norm in the image space $\mathbb{R}_+^\ell$ remains controlled for data within a geodesic ball and even globally controlled for $p = 1$.

When $\mathcal{M}$ is a smooth Cartan–Hadamard manifold, it is well-known that each squared distance function $x \mapsto d_{\mathcal{M}}(p, x)^2$, for $p \in \mathcal{M}$, is differentiable on $\mathcal{M}$ and its Riemannian gradient is given by $-2 \log_x p$, where $\log_x : \mathcal{M} \rightarrow T_x \mathcal{M}$ is the logarithm map of $\mathcal{M}$ at the foot point x (which induces a diffeomorphism between $\mathcal{M}$ and $T_x \mathcal{M}$); c.f. [24]. This implies the following result.

Lemma 1. *Let $\mathcal{M}$ be a Cartan–Hadamard manifold with Riemannian distance $d_{\mathcal{M}}$. Any Fréchet map F_r^p on $(\mathcal{M}, d_{\mathcal{M}})$ with a list of reference points $r = (r_1, \dots, r_\ell)$ is differentiable on $\mathcal{M}$ for $p \geq 2$ and on $\mathcal{M} \setminus \{r_1, \dots, r_\ell\}$ for $1 \leq p < 2$ with the Riemannian Jacobian matrix given, on those respective sets, by:*

$$DF_r^p(x) = (-pd(r_i, x)^{p-2} \log_x r_1, -pd(r_i, x)^{p-2} \log_x r_2, \dots, -pd(r_i, x)^{p-2} \log_x r_\ell).$$

In particular, F_r^1 is not differentiable at the reference points as opposed to F_r^2 . We illustrate in the following section useful important properties of the p -Fréchet maps with $p = 1, 2$ in the context of the clustering problem considered in this paper when $\mathcal{M}$ is a Euclidean space.

2.2.2. p -Fréchet maps on Euclidean space. In this section, we examine the situation in which $\mathcal{M}$ is the space $\mathbb{R}^m$ equipped with the usual Euclidean metric. Although this case is not of practical interest when it comes to the clustering framework we propose, this analysis provides useful insights in preparation to Section 3.2.

To simplify the exposition, we focus on the squared Fréchet map F_r^2 and mention, in remarks, the case of F_r^1 when relevant. If $r = (r_1, \dots, r_\ell) \in \mathbb{R}^{m \times \ell}$ are given reference points, the associated Fréchet map is $F_r^2(x) = (\|x - r_1\|^2, \dots, \|x - r_\ell\|^2)^\top$ with differential $DF_r^2(x) = 2(x - r_1, \dots, x - r_\ell)$. From this, we immediately see that the rank of $DF_r^2(x)$ at $x \in \mathbb{R}^m$ is the dimension of $\text{Span}(x - r_1, \dots, x - r_\ell)$. If $DF_r^2(x)$ is full-rank, by the local inverse function theorem, F_r^2 is locally invertible from a neighborhood of x to its image. In this case, $DF_r^2(x)$ is rank deficient if and only if the reference points lie in some affine subspace of dimension strictly smaller than m passing through x . It follows, conversely, that $DF_r^2(x)$ is full rank when either one of the following two conditions is satisfied:

- (1) $\ell \geq m + 1$ and the affine hull of the reference points is such that $\text{Aff}(r_1, \dots, r_\ell) = \mathbb{R}^m$.
- (2) $\ell = m$, the reference points are affinely independent (i.e., $\text{Aff}(r_1, \dots, r_\ell)$ is of dimension $m - 1$) and $x \notin \text{Aff}(r_1, \dots, r_\ell)$.

The first situation corresponds to having more reference points than the dimension of the space placed in a “generic position.” In this case, F_r is locally invertible at each point $x \in \mathbb{R}^m$. The second situation consists of taking a generic configuration of exactly m reference points. In this case, the Fréchet map is locally invertible at each point outside of the $(m - 1)$ -dimensional affine hull of the reference points.

The next natural question is whether F_r^2 is also a globally injective map. We reason by contradiction; if we assume that F_r^2 is not injective on $\mathbb{R}^m$, then there exist $x \neq x'$ such that $F_r^2(x) = F_r^2(x')$, i.e., $\|x - r_i\| = \|x' - r_i\|$ for any $i = 1, \dots, \ell$. Geometrically, this means that $r_1, \dots, r_\ell$ all lie on the mediatrix hyperplane $H = \{z \in \mathbb{R}^m : \|x - z\| = \|x' - z\|\}$. This immediately implies that with $\ell \geq m + 1$ and the reference points in generic position, i.e., $\text{Aff}(r_1, \dots, r_\ell) = \mathbb{R}^m$ as above, the Fréchet map F_r^2 is necessarily injective. Let $H^0 = \text{Aff}(r_1, \dots, r_\ell)$ denote the affine hull of affinely independent reference points $r_1, \dots, r_\ell$. When $\ell = m$, we know that H^0 is an affine subspace of dimension $m - 1$ that divides $\mathbb{R}^m$ into two disjoint halfspaces H^- and H^+ . In this case, it is easy to see that any two points x, x' , which are symmetric with respect to the hyperplane H^0 , satisfy $F_r^2(x) = F_r^2(x')$ and thus F_r^2 cannot be globally injective on $\mathbb{R}^m$. However, F_r^2 is injective on both halfspaces H^- and H^+ : if not, using the same mediatrix argument as previously, the reference points would all lie at the intersection of H_0 and some other transverse hyperplane, which is impossible by assumption.

Summarizing the previous paragraphs, we have obtained the following general property for Fréchet maps on Euclidean spaces.

Theorem 1. *Let F_r^2 be a Fréchet maps on $\mathbb{R}^m$ with reference points $r = (r_1, \dots, r_\ell) \in \mathbb{R}^{m \times \ell}$. The following holds:*

- (1) *if $\ell \geq m + 1$ and $\text{Aff}(r_1, \dots, r_\ell) = \mathbb{R}^m$ then F_r^2 is a diffeomorphism from $\mathbb{R}^m$ to its image $F_r^2(\mathbb{R}^m) \subset \mathbb{R}^\ell$;*
- (2) *if $\ell = m$ and the reference points are affinely independent, then F_r^2 is a diffeomorphism on each of the two halfspaces delimited by the affine hyperplane $H^0 = \text{Aff}(r_1, \dots, r_m)$.*

Under either of the two configurations of reference points described above, one could ask what is specifically the image of the Fréchet map and how to derive its inverse map. This amounts to reconstructing the position of a point given its distances to the fixed set of reference points. This problem is known as the *multilateration* problem in the literature; it is connected to applications in GPS positioning. Geometrically, it can be seen as finding the intersection of ℓ spheres centered at the reference points. Figure 1(a) depicts an illustration in $\mathbb{R}^3$ with $\ell = 3$ reference points, and the

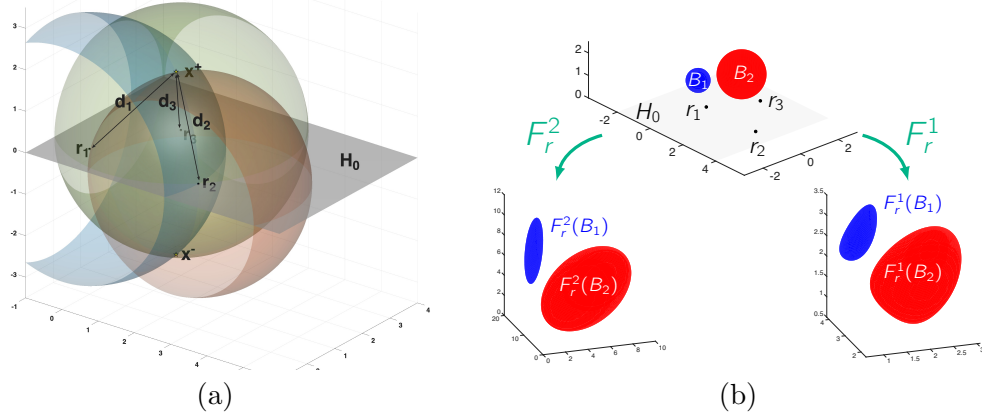


FIGURE 1. Illustration of the Fréchet map in the Euclidean space $\mathcal{M} = \mathbb{R}^3$ with $\ell = 3$ reference points. The left panel (a) shows the two symmetric points with the same given image $d = (d_1, d_2, d_3)$ by the Fréchet map. The right panel shows the images of two disjoint balls in the upper halfspace by the Fréchet maps F_r^2 and F_r^1 .

two resulting solutions that are symmetric with respect to the hyperplane H^0 . Mathematically, the problem can be cast as a system of quadratic equations on the coordinates of the point $x \in \mathbb{R}^m$. This problem can be approached in various ways [5]. Since the proposed Fréchet map clustering approach does not explicitly require the computation of the inverse map, we do not elaborate further on this point. However, we note that some partial derivation for $\ell = m$ can be found as part of the proof of Section B.

Remark 2. *Theorem 1 and the above statements remain nearly identical for the 1-Fréchet map. Specifically, when $\ell \geq m + 1$, F_r^1 is a homeomorphism from $\mathbb{R}^m$ to $F_r^1(\mathbb{R}^m)$ and a diffeomorphism on the subset $\mathbb{R}^m \setminus \{r_1, \dots, r_\ell\}$. For $\ell = m$, the result is the same as point 2 of the theorem.*

A direct consequence of Theorem 1 is that, under the stated assumptions, two disjoint subsets $S_1, S_2 \subset \mathbb{R}^m$ have disjoint images in $\mathbb{R}^\ell$ under the Fréchet map. Yet, as it comes to using Fréchet maps in combination with k -means, the key question is to determine under which conditions $F_r(S_1)$ and $F_r(S_2)$ are in addition separable by a hyperplane of $\mathbb{R}^\ell$. This is a necessary condition for k -means applied to $F_r^2(S_1) \cup F_r^2(S_2)$ to have a solution that perfectly clusters those two sets (see Remark 1). This can be guaranteed in the case of $\ell = m$ reference points thanks to the following result.

Theorem 2. *Let F_r^2 be a Fréchet map on $\mathbb{R}^m$ with reference points $r = (r_1, \dots, r_m) \in \mathbb{R}^{m \times m}$ and assume that the reference points $\{r_1, \dots, r_m\}$ are affinely independent in $\mathbb{R}^m$. Then the image $F_r^2(\mathbb{R}^m)$ is the interior of a paraboloid of $\mathbb{R}^m$. Moreover, if B is a ball contained in either H^- or H^+ , its image $F_r^2(B)$ is an ellipsoid. Consequently, given two disjoint balls B_1 and B_2 both included in one of the two half-spaces H^- or H^+ , their images $F_r^2(B_1)$ and $F_r^2(B_2)$ are separable by a hyperplane in $\mathbb{R}^m$.*

The proof is included in Section B. This result shows that, despite the distortion induced by the Fréchet map, F_r^2 still preserves the convexity of balls in $\mathcal{M}$ and thus their separability by hyperplanes in the image space. Figure 1(b) shows a particular example for the case $m = 3$. We point out that this is specific to the case of the 2-Fréchet map: for the 1-Fréchet map, the image of a ball no longer has a simple ellipsoidal geometry, as also shown in Figure 1(b). We can still show the following weaker result, which is proved in Section C:

Theorem 3. *Let A be a compact subset of $\mathbb{R}^m$. Assume that $r = (r_1, \dots, r_m)$ are m reference points in $\mathbb{R}^m$ where $\{r_1, \dots, r_m\}$ are affinely independent in $\mathbb{R}^m$, and such that either $A \subset H^-$ or*

$A \subset H^+$. If B is any ball of radius $\rho > 0$ with $B \subset A$ and the reference points satisfy the condition:

$$\frac{\rho}{d(r, A)} < \frac{1 - (m - 1)\mu}{\sqrt{m}}$$

where $d(r, A)$ is the distance from the reference point set to A and μ denotes the largest mutual coherence over A i.e. $\mu = \max_{x \in A} \max_{i \neq j} \left| \left\langle \frac{x - r_i}{\|x - r_i\|}, \frac{x - r_j}{\|x - r_j\|} \right\rangle \right|$, then $F_r^1(B)$ is convex in $\mathbb{R}^m$.

We see that the above sufficient condition for convexity is satisfied for reference points chosen sufficiently far from A while also requiring those reference points to be sufficiently spread apart in order to control the mutual coherence μ . However, it is likely that less restrictive conditions could be obtained, but we leave it an open question for future work.

The main problem we are interested in addressing is whether a similar picture can be obtained for Fréchet maps on manifolds $\mathcal{M}$. Some analogous analysis can be done with manifolds of constant negative curvature, namely the hyperbolic spaces $\mathcal{M} = \mathbb{H}^m$. For the sake of brevity, we will not detail this case as it would require the introduction of many additional definitions and notations. In the next section, we shall instead focus on the manifold of SPD matrices of size $n \times n$, which we denote by $SPD(n)$, that are connected to our applications of interest. As we show below, this class of manifolds already involves many key challenges in generalizing the results of Theorem 1 and Theorem 2, respectively.

3. SPD MATRICES

In this section, we focus on the manifold $SPD(n)$. This is a very insightful case to consider, as they are Cartan–Hadamard spaces in which explicit expressions for the geodesics and the exponential and logarithm maps are available. However, the curvature is not constant, which leads to considerable challenges in the context of the clustering framework considered in this work. Furthermore, several applications involve objects living in $SPD(n)$, sometimes with a large dimension n , so that the Fréchet map framework we propose in this paper is especially relevant in this setting.

3.1. The manifold $SPD(n)$ and its Riemannian structure. We start by introducing some basic definitions and notation about $SPD(n)$. First, as an open subset of the space of symmetric matrices $Sym(n)$, $SPD(n)$ can be viewed as a submanifold (of dimension $n(n + 1)/2$) of $Sym(n)$. Its tangent space at any $P \in SPD(n)$ can be identified with $Sym(n)$ itself.

Being in addition a convex subset of $Sym(n)$, it may seem logical to equip $SPD(n)$ with the restriction of the standard Euclidean (i.e., the Frobenius) distance between symmetric matrices. However, it is known that this simple metric can result in unwanted effects when computing averages and, by extension, when performing clustering of SPD matrices. It also lacks some fundamental invariance with the group action of the affine group $GL(n)$ by conjugation. This invariance is a rather natural property when considering SPD matrices representing the covariance of processes, since it encodes the independence of the distance to the choice of coordinate system. Thus, defining adequate metrics on the manifold $SPD(n)$ has been an important topic of research and has led to many different mathematical constructions. As our focus here is on Riemannian metrics and Cartan–Hadamard manifolds, we first discuss the widely used affine invariant metric proposed originally in [30]. It can be obtained as follows: First, one starts with the Euclidean metric on the tangent space at the identity matrix Id . For any $V, W \in T_{\text{Id}}SPD(n) \approx Sym(n)$, we define

$$\langle V, W \rangle_{\text{Id}} = \text{Tr}(V^T W) = \text{Tr}(VW).$$

The idea is then to extend it to the whole $SPD(n)$ via affine invariance. This means enforcing that the action of the affine group on $SPD(n)$, given by $P \mapsto A^T P A$ for any $A \in GL(n)$, is by isometry. Then, if $P \in SPD(n)$, one can write $P = P^{1/2} P^{1/2} = (P^{1/2})^T \text{Id} P^{1/2}$, where $P^{1/2} \in SPD(n)$ is the

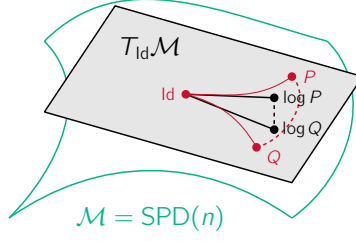


FIGURE 2. Illustration of the Riemannian vs log-Euclidean metrics. We show the tangent space $T_{\text{Id}}\mathcal{M}$ and two points P and Q on the manifold $\mathcal{M}$. Riemannian distances are shown in red. Euclidean distances in $T_{\text{Id}}\mathcal{M}$ are shown in black. We project points P and Q to $T_{\text{Id}}\mathcal{M}$ using a logarithmic map.

SPD square root of P . In addition, the invariance property of the metric implies necessarily that

$$(5) \quad \langle V, W \rangle_P = \langle P^{-1/2}VP^{-1/2}, P^{-1/2}WP^{-1/2} \rangle_{\text{Id}} = \text{Tr}(P^{-1}VP^{-1}W).$$

The properties of this metric have been extensively studied; see, for instance, [25] for a synthesis of those results. We recapitulate the main expressions used in the rest of the paper. First, one can show that the resulting Riemannian distance can be computed explicitly between any two matrices $P, Q \in \text{SPD}(n)$ and is given by

$$(6) \quad d(P, Q)^2 = \text{Tr}(\log(P^{-1/2}QP^{-1/2})^2).$$

Likewise, the constant-speed geodesic $\gamma : [0, 1] \rightarrow \text{SPD}(n)$ from P to Q is given by

$$(7) \quad \gamma(t) = P^{1/2} \exp(t \log(P^{-1/2}QP^{-1/2})) P^{1/2} = P^{1/2} (P^{-1/2}QP^{-1/2})^t P^{1/2}.$$

As above, $\log$ and $\exp$ denote the matrix logarithm and exponential for symmetric positive definite matrices. Furthermore, the Riemannian exponential and logarithm maps on $\text{SPD}(n)$ have the following expressions, which hold for any $P, Q \in \text{SPD}(n)$ and any $V \in \text{Sym}(n)$:

$$(8) \quad \begin{aligned} \exp_P(V) &= P^{1/2} \exp(P^{-1/2}VP^{-1/2}) P^{1/2} \\ \log_P(Q) &= P^{1/2} \log(P^{-1/2}QP^{-1/2}) P^{1/2}. \end{aligned}$$

It follows that the manifold $\text{SPD}(n)$ is complete for the affine-invariant metric. This is because the exponential map is always well-defined; it is a diffeomorphism from $\text{Sym}(n)$ to $\text{SPD}(n)$. Moreover, it can be shown (c.f., [25], Theorem 3.3) that all sectional curvatures of $\text{SPD}(n)$ are non-positive. This property makes $\text{SPD}(n)$ a Cartan–Hadamard manifold.

Although the above affine-invariant metric provides a relatively simple Riemannian structure on $\text{SPD}(n)$, there are many alternative (typically non-Riemannian) metrics (or even divergences) that have been considered in the literature. We do not provide a comprehensive list of such metrics but specifically mention the case of the *log-Euclidean* framework introduced in [4], since it is a natural approach to compare to the Fréchet mapping idea introduced in this work. In its standard form, the log-Euclidean metric can be seen as an approximation of the affine-invariant Riemannian distance centered at the identity matrix Id (or, more generally, at a chosen template point in $\text{SPD}(n)$). Given two matrices $P, Q \in \text{SPD}(n)$, we can represent them as elements of the tangent space at Id using the $\log$ map at Id , which is simply the usual matrix logarithm. One can then compare the resulting symmetric matrices $\log P \in \text{Sym}(n)$ and $\log Q \in \text{Sym}(n)$ based on the metric on $T_{\text{Id}}\text{SPD}(n)$ (i.e., the Euclidean metric on $\text{Sym}(n)$). This leads to the log-Euclidean distance

$$d_{\text{LE}}(P, Q) = \|\log P - \log Q\|_F.$$

In other words, the log-Euclidean distance on $SPD(n)$ is obtained as the Frobenius norm between the logarithms of P and Q . The general idea is illustrated in Figure 2. Although this metric loses the full affine invariance of Equation (6), it retains the invariance with respect to the action of similarities of $\mathbb{R}^n$. The most relevant aspect of the log-Euclidean framework is that it allows for significantly faster computation compared to the geodesic distance. For example, performing k -means clustering can be reduced to applying standard Euclidean k -means to the logarithm of the data matrices, resulting in computational complexity comparable to the standard k -means algorithm in Euclidean spaces.

3.2. Properties of the Fréchet map on $SPD(n)$. Similarly to the discussion on the Euclidean case in Section 2.2.2, here we focus our analysis on the squared Fréchet map ($p = 2$) and only state the corresponding results for the 1-Fréchet map in remarks, as the arguments are similar. For brevity, throughout this section we drop from the notation the reference point set and the power p and simply write the squared Fréchet map as:

$$F : P \in SPD(n) \mapsto (d(R_1, P)^2, d(R_2, P)^2, \dots, d(R_\ell, P)^2) \in \mathbb{R}_+^\ell,$$

with d being the Riemannian distance defined in Equation (6); we denote the reference points, which are matrices in $SPD(n)$, as $R_1, \dots, R_\ell$. It follows from Lemma 1 that F is differentiable and the Riemannian differential is given, for any $P \in SPD(n)$ and any $V \in Sym(n)$, by:

$$(9) \quad DF(P) \cdot V = \begin{pmatrix} \langle -2 \log_P(R_1), V \rangle_P \\ \vdots \\ \langle -2 \log_P(R_\ell), V \rangle_P \end{pmatrix} = \begin{pmatrix} -2 \operatorname{Tr}(P^{-1/2} \log(P^{-1/2} R_1 P^{-1/2}) P^{-1/2} V) \\ \vdots \\ -2 \operatorname{Tr}(P^{-1/2} \log(P^{-1/2} R_\ell P^{-1/2}) P^{-1/2} V) \end{pmatrix}.$$

3.2.1. Local invertibility of F . Based on Equation (9), we investigate the local invertibility of the Fréchet map depending on the position of the set of reference points. We recall that F is *locally invertible* at a point $P \in SPD(n)$ when $DF(P)$ is of rank m . This also means that F is a diffeomorphism from a certain neighborhood of P to the image of that neighborhood. Viewing $(\log_P R_1, \dots, \log_P R_\ell)$ as a matrix in $\mathbb{R}^{m \times \ell}$ with $m = n(n+1)/2$ being the dimension of $Sym(n)$, we introduce the set

$$\Gamma_{R_1, \dots, R_\ell} = \{P \in SPD(n) : \operatorname{rank}(\log_P R_1, \dots, \log_P R_\ell) = m\}.$$

We note that $\Gamma_{R_1, \dots, R_\ell}$ is an open subset of $SPD(n)$, which is automatically empty for $\ell < m$. From Equation (9) it follows that $\Gamma_{R_1, \dots, R_\ell}$ is precisely the set of points P for which $DF(P)$ is full rank. We can describe this set in an alternative way. To do so, let us denote by $S_{R_1, \dots, R_\ell}$ the complement of $\Gamma_{R_1, \dots, R_\ell}$, i.e., the set of all $P \in SPD(n)$ such that $DF(P)$ is rank deficient. For any $P \in S_{R_1, \dots, R_\ell}$, one has that the tangent vectors $(\log_P(R_1), \dots, \log_P(R_\ell))$ lie in a strict subspace of $Sym(n)$. In other words, there exists $V \in Sym(n)$, $V \neq 0$, such that the reference points $R_1, \dots, R_\ell$ all belong to $\mathcal{H}_{P, V^\perp} = \{\exp_P(W) : W \in Sym(n), \operatorname{Tr}(VW) = 0\}$. The set $\mathcal{H}_{P, V^\perp}$ can be thought of as one possible Riemannian equivalent of an affine hyperplane passing through P and with tangent vectors orthogonal to V at P . Similar to the Euclidean space, we see that $S_{R_1, \dots, R_\ell}$ is the reunion of the “Riemannian geodesic hyperplanes” $\mathcal{H}_{P, V^\perp}$ for $P \in SPD(n)$ and non-zero $V \in Sym(n)$ that contain all the reference points. We obtain the following result:

Proposition 2. *Any Fréchet map F on $SPD(n)$ is locally invertible on*

$$\Gamma_{R_1, \dots, R_\ell} = SPD(n) \setminus \bigcup \{\mathcal{H}_{P, V^\perp} : R_1, \dots, R_\ell \in \mathcal{H}_{P, V^\perp}\}.$$

Unfortunately, unlike for the Euclidean case, it is significantly more difficult to characterize which Riemannian hyperplanes contain a given set of reference points. Thus, we cannot give a more explicit geometric description of $\Gamma_{R_1, \dots, R_\ell}$. Nevertheless, we obtain the following result.

Theorem 4. *Let F be a Fréchet map on $SPD(n)$ with reference points $R_1, \dots, R_\ell$ where $\ell \geq m$, where $m = \dim(SPD(n)) = n(n+1)/2$. If the reference points $R_1, \dots, R_\ell$ are such that there exists $Q \in SPD(n)$ for which $\{\log_Q R_1, \dots, \log_Q R_\ell\}$ is of rank m , then the set $S_{R_1, \dots, R_\ell}$ is of Lebesgue measure zero in $SPD(n)$. In other words, F is locally invertible on the open subset $\Gamma_{R_1, \dots, R_\ell}$, which has full measure in $SPD(n)$.*

Proof. Let us first consider the case $\ell = m$. In that case, $S_{R_1, \dots, R_m}$ is equivalently the set of all $P \in SPD(n)$ such that $\det(\log_P R_1, \dots, \log_P R_m) = 0$. We introduce the mapping $\psi : SPD(n) \rightarrow \mathbb{R}$ defined by

$$\psi(P) = \det(\log_P R_1, \dots, \log_P R_m).$$

Now, given the expression of the Riemannian logarithm Equation (8), since the matrix logarithm and square root are both real analytic functions on $SPD(n)$ and the determinant of a $m \times m$ matrix is a polynomial function of its coefficients, it directly follows that ψ is real analytic. Also, by assumption, ψ is not identically zero (since $\psi(Q) \neq 0$) and thus, from standard results on zero sets of real analytic functions [23], the zero level set of ψ , in other words $S_{R_1, \dots, R_m}$, is of vanishing Lebesgue measure in $SPD(n)$.

For $\ell > m$, given some reference points $R_1, \dots, R_\ell$, we see that $P \in S_{R_1, \dots, R_\ell}$ if and only if all $m \times m$ subdeterminants of the matrix $(\log_P R_1, \dots, \log_P R_m)$ are equal to zero. With the assumption of the theorem, one of the $m \times m$ subdeterminants of $(\log_Q R_1, \dots, \log_Q R_m)$ is necessarily non-vanishing. Therefore, we can use the same argument as above and deduce that $S_{R_1, \dots, R_\ell}$ is an intersection of sets at least one of which is of measure zero. Thus $S_{R_1, \dots, R_\ell}$ is also of measure zero. $\square$

We stress that the condition on the reference points in Theorem 4 is very generic. For instance, with $Q = \text{Id}$, it holds as soon as $(\log R_1, \dots, \log R_\ell)$ is a full rank matrix. This is true for almost all choices of reference points (when considering the product Lebesgue measure of $Sym(n)^\ell$). The proof also yields the fact that the set $S_{R_1, \dots, R_\ell}$ is actually a real analytic subvariety of $Sym(n)$. In the case where $\ell > m$, one could expect (as in the Euclidean case) that—for almost all choices of ℓ reference points—the Fréchet map is locally invertible on the whole space $SPD(n)$. However, due to the lack of a simple characterization of geodesic hyperspaces in $SPD(n)$, we are only able to obtain the statement under the following stronger assumption.

Theorem 5. *Let F be a Fréchet map on $SPD(n)$ with reference points $R_1, \dots, R_\ell$ where $\ell \geq 2m$, where $m = \dim(SPD(n)) = n(n+1)/2$. For almost all choices of reference points in $SPD(n)$, one has $S_{R_1, \dots, R_\ell} = \emptyset$. In other words, the associated Fréchet map is locally invertible on $SPD(n)$.*

Proof. Thanks to the above description of $S_{R_1, \dots, R_\ell}$, we know that $S_{R_1, \dots, R_\ell} \neq \emptyset$ if and only if there exist $P \in SPD(n)$ and V a unit Frobenius norm matrix of $Sym(n)$ such that $R_1, \dots, R_\ell \in \mathcal{H}_{P, V^\perp}$. Let us denote by $\overline{Sym}(n)$ the unit sphere of $Sym(n)$ and its tangent bundle $T\overline{Sym}(n) = \{(V, W) \in Sym(n) \times Sym(n) : \text{Tr}(VW) = 0\}$ and by $T\overline{Sym}(n)^{\otimes \ell}$ the ℓ -times tangent bundle. We introduce the mapping

$$G : SPD(n) \times T\overline{Sym}(n)^{\otimes \ell} \rightarrow SPD(n)^\ell \\ (P, V, W_1, \dots, W_\ell) \mapsto (\exp_P(PW_1P), \dots, \exp_P(PW_\ell P)).$$

We note that for any $P \in SPD(n)$ and $(V, W_1, \dots, W_\ell) \in T\overline{Sym}(n)^{\otimes \ell}$, we have for all $i = 1, \dots, \ell$ that $\langle V, PW_iP \rangle_P = \text{Tr}(P^{-1}VP^{-1}PW_iP) = \text{Tr}(VW_i) = 0$ and thus the matrices $\exp_P(PW_1P), \dots, \exp_P(PW_\ell P) \in \mathcal{H}_{P, V^\perp}$. Therefore, the set of all ℓ -tuples of reference points for which $S_{R_1, \dots, R_\ell} \neq \emptyset$ is contained in the image $G(SPD(n) \times T\overline{Sym}(n)^{\otimes \ell})$. As G is a differentiable map and $SPD(n) \times T\overline{Sym}(n)^{\otimes \ell}$ is a manifold of dimension $2m - 1 + \ell m - \ell$, which is strictly smaller than ℓm (the dimension of $SPD(n)^\ell$) when $\ell \geq 2m$, we deduce that for almost all reference points $R_1, \dots, R_\ell$, we have $S_{R_1, \dots, R_\ell} = \emptyset$. $\square$

An open question that we leave to future investigation is to determine whether the above still holds for a number of reference points ℓ between $m + 1$ and $2m$, as it does for Fréchet maps in Euclidean space.

Remark 3. When considering the 1-Fréchet map F^1 instead of its squared version, the statement of Theorem 4 still holds provided that there exists $Q \in SPD(n) \setminus \{R_1, \dots, R_\ell\}$ satisfying the theorem's condition. Similarly, in Theorem 5, the conclusion simply becomes that F^1 is a locally invertible map on $SPD(n) \setminus \{R_1, \dots, R_\ell\}$.

3.2.2. *The case $\ell = m$.* A primary interest in this paper is to identify Fréchet maps F with the smallest possible number of reference points in an attempt to limit the computational cost together with the dimension of the output space for F . We examine more closely the case $\ell = m$ and attempt to describe more precisely the real analytic subvariety $S_{R_1, \dots, R_m}$ for certain specific configurations of reference points. When $\ell = m$, we recall that $P \in S_{R_1, \dots, R_m}$ if and only if $\{\log_P R_1, \dots, \log_P R_m\}$ are linearly dependent.

A first observation is that $S_{R_1, \dots, R_m}$ can be, in fact, connected to the intrinsic notion of *exponential barycentric subspace* that was introduced for general Riemannian manifolds and studied in [24]. The exponential barycentric subspace of $R_1, \dots, R_m$ is defined by

$$EBS(R_1, \dots, R_m) = \left\{ P : \exists \lambda \in \mathcal{P}_m^*, P \text{ is a critical point of } \sigma(Q; \lambda)^2 \doteq \sum_{i=1}^m \lambda_i d(R_i, Q)^2 \right\},$$

where $\mathcal{P}_m^*$ denotes the set of $\lambda = (\lambda_i) \in \mathbb{R}^m$ such that $\sum_{i=1}^m \lambda_i = 1$. This can be interpreted as the set of all the weighted Fréchet barycenters of the R_i 's. Using again Lemma 1, the criticality condition is equivalent to $\sum_{i=1}^m \lambda_i \log_P R_i = 0$. The latter implies in particular that $\{\log_P R_1, \dots, \log_P R_m\}$ are linearly dependent. We deduce that

$$EBS(R_1, \dots, R_m) \subseteq S_{R_1, \dots, R_m}.$$

In contrast to the Euclidean case, these two sets are not necessarily equal as $S_{R_1, \dots, R_m}$ also contains all the critical points of the functions $\sigma(Q; \lambda)^2$, i.e., the solutions of $\sum_{i=1}^m \lambda_i \log_P R_i = 0$ for the non-zero $\lambda \in \mathbb{R}^m$ with $\sum_{i=1}^m \lambda_i = 0$.

Although [24] provides more explicit characterizations of *EBS* in constant curvature manifolds (such as the sphere and hyperbolic plane), there is unfortunately no known corresponding simple description in the case of $SPD(n)$. We conjecture that a possible obstruction lies in the fact that $EBS(R_1, \dots, R_m)$ or $S_{R_1, \dots, R_m}$, unlike affine hyperplanes in a Euclidean space, are not necessarily totally geodesic. In what follows, we say that a submanifold $S \subset SPD(n)$ is *totally geodesic* if it is geodesically complete (i.e., $\exp_P(V) \in S$ for any $P \in S$ and $V \in T_P S$) and if any geodesic of S is also a geodesic in $SPD(n)$. In symmetric spaces such as $SPD(n)$, totally geodesic submanifolds can be characterized via the Lie triple system condition (see Theorem 7.2 in [15] or [34]). This condition allows us to fully classify all totally geodesic submanifolds of codimension 1 in $SPD(n)$, as stated below.

Proposition 3. For $n \geq 3$, the totally geodesic submanifolds of $SPD(n)$ of dimension $m - 1$ are exactly the subsets $SPD_r(n) = \{P \in SPD(n) : \det(P) = r\}$ for $r > 0$. Furthermore, there is a well-defined projection $\Pi_r : SPD(n) \rightarrow SPD_r(n)$ given by $\Pi_r(P) = \sqrt[n]{r/\det(P)} P$ for all $P \in SPD(n)$.

It is easily verified that $SPD_r(n)$ is totally geodesic based on the expression of geodesics Equation (7). Showing that these are, in fact, the only ones based on the Lie triple system condition is a little more involved. Since, to our knowledge, there is no statement of this result in the literature, a proof is detailed in Section D, in which we also derive a corresponding result for the special case of $n = 2$.

Based on Proposition 3, let us now consider the situation in which all m reference points belong to a totally geodesic hypersurface $SPD_r(n)$ for some $r > 0$. We then ask the question whether

$EBS(R_1, \dots, R_m)$ or $S_{R_1, \dots, R_m}$ precisely coincide with $SPD_r(n)$ in this case. A partial answer is given by the following result.

Theorem 6. *Assume that $R_1, \dots, R_m \in SPD_r(n)$ for some $r > 0$. Then*

$$EBS(R_1, \dots, R_m) \subseteq SPD_r(n) \subseteq S_{R_1, \dots, R_m}.$$

Moreover, these three spaces are all equal unless the reference points lie on a submanifold of the form $\exp_P(H)$ for $P \in SPD_r(n)$ and H an affine subspace of $Sym(n)$ of dimension at most $m - 2$. Under this condition, the Fréchet map associated to $R_1, \dots, R_m$ is an immersion on each of the halfspaces $SPD_r^+(n) = \{P : \det(P) > r\}$ and $SPD_r^-(n) = \{P : \det(P) < r\}$.

The proof can be found in Section E. We stress that one question that remains unaddressed in Theorem 6 is whether the condition of equality between the exponential barycentric subspace $S_{R_1, \dots, R_m}$ and $SPD_r(n)$ is truly generic. Indeed, while it can be interpreted as a form of affine independence of the reference points within the submanifold $SPD_r(n)$, it still involves checking all possible foot points $P \in SPD_r(n)$, unlike the notion of affine independence that is introduced in [24]. We leave it to future investigations to establish if, for instance, almost all configurations of m reference points in $SPD_r(n)$ do satisfy this condition. This issue notwithstanding, the result of Theorem 6 provides a picture in part reminiscent of the Euclidean case: choosing reference points located on a totally geodesic hypersurface $SPD_r(n)$ results in a Fréchet map that is a local diffeomorphism on each of the two halfspaces delimited by $SPD_r(n)$.

Remark 4. *We emphasize that we have mainly focused the discussion on the local injectivity of the Fréchet map on $SPD(n)$. Some natural follow-up questions are to determine which configurations of reference points further result in F being a globally injective map, what is then the image of F , and how to obtain the inverse map, as in the Euclidean situation discussed in Section 2.2.2. To our knowledge, these questions become significantly more difficult in $SPD(n)$. Our preliminary exploration of the simple case of $\ell = 3$ reference points in the three-dimensional $SPD(2)$ manifold has shown that finding the inverse of F , i.e., reconstructing a matrix $P \in SPD(2)$ from its distances to the three reference points, can be reduced to solving a single analytic equation on the determinant of P . However, the precise structure of solutions of this equation, even in this basic case, remains elusive. Numerical evidence suggests a quite different picture from the Euclidean case given by Theorem 1. For these reasons, we leave such issues for future investigation.*

3.3. k -means clustering methods in SPD. In this section, we describe several possible strategies for applying k -means on datasets of SPD matrices, which will be compared numerically in the next section.

The first and most straightforward approach, which we henceforth refer to as the *Intrinsic Riemannian Clustering (IRC)*, is the direct transposition of the classical Lloyd's algorithm, reported in Section 2 as Algorithm 1, to the SPD manifold setting. This transposition requires replacing the Euclidean distance with the Riemannian distance Equation (6) on SPD and computing the cluster centroids in step 5 Algorithm 1 with the Fréchet mean, rather than the arithmetic mean. For a given set $\{X_1, \dots, X_N\} \subset SPD(n)$, computing the Fréchet mean requires solving the minimization problem

$$(10) \quad \bar{P} = \arg \min_{P \in SPD(n)} \left\{ f(P) = \frac{1}{N} \sum_{i=1}^N d(X_i, P)^2 \right\}.$$

Due to the aforementioned properties of SPD, this is a convex problem with a unique solution. It can be solved numerically via Riemannian gradient descent in $SPD(n)$ with gradient $g(P) = -(2/N) \sum_{i=1}^N \log_P X_i$. The gradient can be directly expressed as described in Section 3.2. We adopt the standard scheme from [2] which is summarized in Algorithm 2 below.

Algorithm 2 Gradient descent method to estimate the Fréchet mean (based on [2]).

- 1: **Input:** A set of points $X_1, \dots, X_N \in SPD(n)$ and an initialization $P_0 \in SPD(n)$
 - 2: $P^{(0)} \leftarrow P_0, t \leftarrow 0, \text{stop} \leftarrow \text{false}$
 - 3: **while** $\neg \text{stop}$ **do**
 - 4: $g(P^{(t)}) \leftarrow -(2/N) \sum_{i=1}^N \log_{P^{(t)}} X_i$
 - 5: $\eta^{(t)} \leftarrow \text{compute step size (constant or adaptative)}$
 - 6: $P^{(t+1)} \leftarrow \exp_{P^{(t)}}(-\eta^{(t)} g(P^{(t)}))$
 - 7: $t \leftarrow t + 1$
 - 8: $\text{stop} \leftarrow \text{check convergence}$
 - 9: **end while**
 - 10: **Output:** Fréchet mean $\bar{P} = P^{(t)}$
-

Since repeated evaluations of Fréchet means is the main computational bottleneck of the k -means method in manifolds, some algorithmic strategies have been proposed to reduce computational cost, most notably the *Iterative Centroid Method (ICM)* proposed in [10, 17], which is based on the *recursive barycenter* scheme suggested in [31]. This approach approximates the Fréchet mean of a set $X_1, \dots, X_N \in SPD(n)$ from the iterations

$$P_1 = X_1,$$

$$P^{(t+1)} = (P^{(t)})^{1/2} ((P^{(t)})^{-1/2} X^{(t+1)} (P^{(t)})^{-1/2})^{1/(t+1)} (P^{(t)})^{1/2}$$

for $t = 1, \dots, N-1$. In other words, starting from the first element X_1 in the set, one keeps moving the mean estimate along the geodesic connecting it to the next element with a step decreasing as the inverse of the iteration number. Thus, this approach involves computing N geodesics, which is comparable to a single iteration of the full gradient descent method of Algorithm 2 (albeit one can parallelize the latter). Note that this scheme is also related to a form of stochastic gradient descent method for Equation (10). In the following, we refer to this specific k -means algorithm, which uses this recursive scheme to estimate the Fréchet mean at each iteration, as *Approximate Riemannian Clustering (ARC)*.

As discussed above, in this work, we consider an alternative approach for performing k -means on manifold data that involves a prior embedding of the data into some Euclidean space. Our proposed approach, called *Fréchet Map Clustering (FMC)*, uses a Fréchet map and is summarized in Algorithm 3. As discussed above, the properties of the Fréchet map and, hence, the properties of the algorithm, are fundamentally linked to the number and selection of the reference points. We discuss and evaluate different strategies for reference point selection in Section 4.

Algorithm 3 Fréchet Map Clustering (FMC)

- 1: **Input:** A dataset $\mathcal{D} = \{X_1, \dots, X_N\} \subset SPD(n)$, the number of clusters $k \in \mathbb{N}$.
 - 2: **Parameters:** A set $\{R_1, \dots, R_\ell\}$ of reference points in $SPD(n)$, order $p = 1, 2$.
 - 3: Define the Fréchet map $F^p : P \mapsto F(P) = (d(P, R_1)^p, \dots, d(P, R_\ell)^p) \in \mathbb{R}^\ell$ and compute the image of $\mathcal{D}$ by F^p : $F^p(\mathcal{D}) \leftarrow \{F^p(X_1), \dots, F^p(X_N)\} \subset \mathbb{R}^\ell$.
 - 4: Apply the k -means algorithm in $\mathbb{R}^\ell$ to the set $F^p(\mathcal{D})$ to partition the dataset into k clusters $H_1, \dots, H_k \subset \mathbb{R}^\ell$.
 - 5: Identify the corresponding clustering $(CL_1, \dots, CL_k)$ in $SPD(n)$ by a re-labeling the data points which defines $CL_r = \{X_j \in \mathcal{D} \mid F^p(X_j) \in H_r\}$.
 - 6: **Output:** A partition of $\mathcal{D}$ into the clusters $(CL_1, \dots, CL_k)$.
-

Since computing means and distances in the Euclidean setting is very fast compared to the analog manifold computations, the critical advantage of the FMC algorithm over the methods

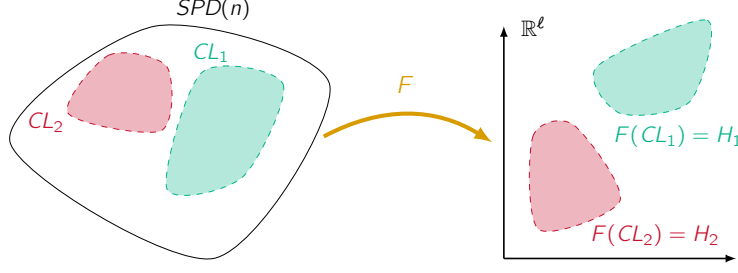


FIGURE 3. FMC algorithm. The Fréchet map F takes a finite set $\mathcal{D} \subset SPD(n)$ into $\mathbb{R}^\ell$. Next, the k -means algorithm in $\mathbb{R}^\ell$ is applied to partition the set $F(\mathcal{D})$ into k clusters $\{H_1, \dots, H_k\}$. Finally, a simple re-labeling of the data is to identify the corresponding clusters $(CL_1, \dots, CL_k)$ of $\mathcal{D}$ in $SPD(n)$.

described above is to carry out the iterative steps of k -means in a Euclidean space rather than a Riemannian manifold $\mathcal{M}$. This is illustrated in Figure 3. The computationally most expensive step in Algorithm 3 is step 3, which requires computing ℓr distances $d(X, R_i)$ in $SPD(n)$. Each such distance computation requires at most $\mathcal{O}(n^3)$ floating point operations, so that the total cost of step 3 is $\mathcal{O}(\ell N n^3)$, where N is the number of data points and ℓ is the size of the reference set in $SPD(n)$.

Note that the injectivity of the Fréchet mapping F is not necessary to implement FCM since step 5 is simply a re-labeling of the original data points.

For comparison, we also consider below the *Log-Euclidean Clustering (LEC)*, originally proposed in [4], which also maps the data from $SPD(n)$ into a Euclidean space. LEC follows steps similar to Algorithm 3, the main difference being that the Fréchet map F is replaced with the log map $P \mapsto \log_R P \in Sym(n)$ for a given reference point R (typically $R = \text{Id}$). Subsequently, the k -means algorithm is applied in the Euclidean space $Sym(n) \approx \mathbb{R}^{n(n+1)/2}$ and the cluster points are then mapped back into $SPD(n)$.

3.4. Reference point selection. We found that the performance of the FMC algorithm (Algorithm 3) depends significantly on the positioning of the reference points $\{R_i\}_{i=1}^\ell$, $\ell \in \mathbb{N}$, making the selection of reference points a critical component for deployment. Our goal is a strategy that is computationally fast, cluster-agnostic, and consistently ensures high clustering accuracy. While the theoretical analysis of the previous sections gives a certain insight into how the choice of reference points influence the properties of the resulting Fréchet map, these results were mostly focused on the situation of a number of reference points ℓ equal or larger than the manifold's dimension m . Here, we will instead introduce heuristic approaches in which ℓ is chosen sometimes much smaller than m . Although this a priori breaks the invertibility of the Fréchet map, we empirically found that the overall clusters structure of the data can still be preserved sufficiently and lead to robust clustering results for the FMC method. We will specifically investigate two heuristics: a random selection approach and a principled strategy that requires tuning two scalar hyperparameters. Section 3.5 details our numerical validation of these strategies.

3.4.1. Random reference point selection. The simplest method of selecting reference points for Algorithm 3 is random selection. While placing them randomly on a hypersphere enclosing the dataset was explored, it yielded suboptimal results due to our general lack of knowledge about the dataset's geometry. Therefore, we adopted a more practical strategy in which we randomly select the reference points directly from the dataset itself.

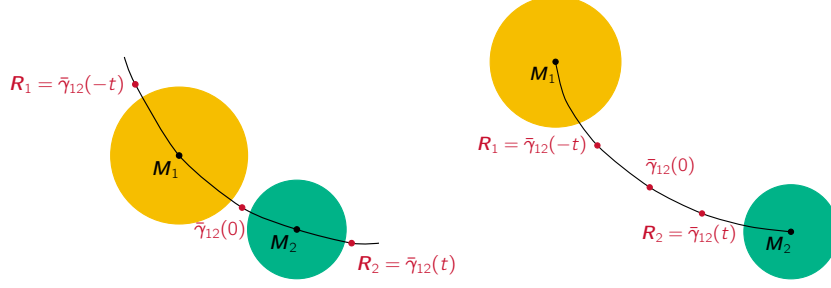


FIGURE 4. Illustration of the proposed strategy to select reference points for the Fréchet map F . Left: Close case. We select the reference points R_1 and R_2 outside the segment that connects the Fréchet means M_1 and M_2 of the clusters CL_1 and CL_2 . Right: Far case. We select the reference points R_1 and R_2 inside the segment that connects the centers M_1 and M_2

3.4.2. Reference point selection based on approximate Fréchet means. Although using random point selection is appealing due to its simplicity, it might lead to inconsistent clustering results. Consequently, we designed a more principled approach that takes into account the geometry of our problem. In summary, our idea is to place the reference points $\{R_m\}_{m=1}^\ell$ on the geodesic $\gamma_{M_i M_j}(t)$ that runs through the centroids, i.e., the Fréchet means M_i and M_j , of each pair of clusters CL_i and CL_j . Unfortunately, this requires knowing the clusters (true labels), which obviously defeats the purpose. However, since we are solely interested in placing reference points, it is sufficient to compute an approximate solution of the clustering problem so that we can compute approximate Fréchet means $\tilde{M}_i$, $\tilde{M}_j$ and a geodesic $\gamma_{\tilde{M}_i \tilde{M}_j}(t)$ running through those points, where

$$(11) \quad \gamma_{\tilde{M}_i \tilde{M}_j}(t) := \tilde{M}_i^{1/2} \exp(t \tilde{M}_i^{-1/2} \tilde{M}_j \tilde{M}_i^{-1/2}) \tilde{M}_i^{1/2}, \quad t \in \mathbb{R}.$$

The next tasks we face are where to place the reference points along this geodesic and how many reference points per cluster pair CL_i and CL_j to select to accurately cluster the data. We elaborated the following strategy on the basis of empirical observations.

- (1) For each pair of approximate Fréchet means $\tilde{M}_i$, $\tilde{M}_j$, we place *two* reference points along the geodesic connecting $\tilde{M}_i$ and $\tilde{M}_j$.
- (2) We place the reference points *between* $\tilde{M}_i$ and $\tilde{M}_j$ if the clusters are “far” from each other (“far case”); we place the reference points *outside* the geodesic curve connecting $\tilde{M}_i$ and $\tilde{M}_j$, if the clusters are “close” (“close case”). This is shown in Figure 4.

This strategy requires us not only to provide an estimate for the clusters $\{CL_i\}_{i=1}^k$ but also to estimate the radius $\rho_i > 0$ of each cluster CL_i . Moreover, we need to quantify what we mean by the cluster pair being “far” or “close” and determine where along the geodesic to place the two reference points. The following remarks explain the rationale of our strategy.

Remark 5. In an attempt to embed the data into a Euclidean space with the lowest possible dimension, we initially selected only one reference point per pair $\{CL_i, CL_j\}$. However, we observed heuristically that the clustering performance of the FMC algorithm could be unstable with respect to permutations of the location of this reference point (e.g., depending on the reference point being located on the far side of cluster CL_i or CL_j). If the number of clusters is small, doing permutations is a viable strategy since computing the permutations is not expensive, and selecting the best reference point assignment based on, e.g., dispersion or some criterion related to the downstream decisions based on the clustering performance, does not add significant runtime. However, if the

number of clusters increases, this permutation approach becomes expensive. We observed that selecting two reference points per pair $\{CL_i, CL_j\}$ makes the FMC algorithm significantly more stable. A potential downside is that this approach embeds the data in a higher-dimensional space.

Remark 6. We introduced the “far” and “close” cases since we observed that computing long geodesics $\gamma(t)$ can introduce numerical instability. To avoid computing long geodesics, we place the reference points between the clusters if the approximate Fréchet means $\tilde{M}_i$ and $\tilde{M}_j$ are “far” from each other relative to the Riemannian radii ρ_i and ρ_j of the clusters. To differentiate these two cases and, thus, determine the position of the reference points, we assign the coordinate along the geodesic by choosing

$$(12) \quad t_{ij} = \begin{cases} t_{\text{close}} & \text{if } \frac{d(\tilde{M}_i, \tilde{M}_j)}{\tilde{\rho}_i + \tilde{\rho}_j} < \epsilon_d, \quad i, j = 1, \dots, k, \quad i \neq j, \\ t_{\text{far}} & \text{otherwise,} \end{cases}$$

for the “close case” and the “far case,” respectively. See fig. 4 for an illustration. Here, $\tilde{\rho}_i$ and $\tilde{\rho}_j$ represent the approximate Riemannian radii of the convex hull of $\widetilde{CL}_i$ and $\widetilde{CL}_j$, respectively, and $\epsilon_d > 0$, $t_{\text{close}} > 1$, and $0 \leq t_{\text{far}} < 1$ are user defined parameters. The parameter t_{ij} controls how far along the geodesic running through $\tilde{M}_i$ and $\tilde{M}_j$ we place the reference points.

Next, we address the question of how to approximate CL_i , M_i , and ρ_i . While the simplest approach is to carry out the approximation directly in Euclidean space (using the embedding of $SPD(n)$ in $\mathbb{R}^m$), this tends to lead to inconsistent clustering results. A better approximation is obtained using the log-Euclidean framework, in which we compute the approximate clusters $\{\widetilde{CL}_i\}_{i=1}^k$ by applying k -means in Euclidean space after embedding the data $\mathcal{D}$ into the tangent space using matrix logarithms $\log_{\text{Id}} : SPD(n) \rightarrow \mathbb{R}^{n(n+1)/2}$ (we evaluate the matrix logarithm at Id in all our experiments). Subsequently, we use the clusters $\widetilde{CL}_i$ to compute the Fréchet means $\tilde{M}_i$ on the manifold using ICM [10, 17] (see also Section 3.3). To find $\tilde{\rho}_i$ we draw $N_\rho \ll N$ samples X_j from $\widetilde{CL}_i \subset \mathcal{D} = \{X_1, \dots, X_N\}$ for each approximate cluster $\widetilde{CL}_i$. Given $\tilde{M}_i$ and $\tilde{\rho}_i$, we determine the location of the reference points based on the criterion Equation (12). To place the reference points, we compute the midpoint $\bar{M}_{ij} = \gamma_{\tilde{M}_i \tilde{M}_j}(1/2)$ of the geodesic that connects $\tilde{M}_i$ and $\tilde{M}_j$ using Equation (11). Given $\bar{M}_{ij}$, we position the reference points along the geodesic at locations $\bar{\gamma}_{ij}(\pm t_{ij})$ based on the auxiliary geodesic

$$(13) \quad \bar{\gamma}_{ij}(t) := \bar{M}_{ij}^{1/2} \exp(t \bar{M}_{ij}^{-1/2} M_j \bar{M}_{ij}^{-1/2}) \bar{M}_{ij}^{1/2}$$

defined by the midpoint $\bar{M}_{ij}$. We summarize this approach in Algorithm 4.

Remark 7. Using the log-Euclidean framework to identify approximate clusters $\{\widetilde{CL}_i\}_{i=1}^k$ as a way to select reference points $\{R_m\}_{m=1}^\ell$ may seem counterintuitive, since, in some cases, this approach already solves the clustering problem. However, we show in Section 4 that our framework yields more accurate and more consistent results than the log-Euclidean framework at a comparable computing cost.

By the description above, the selection of the reference point for our FCM algorithm requires assigning the values of the hyperparameters: t_{close} , t_{far} , N_ρ , and ϵ_d . We discuss the sensitivity of our FCM algorithm with respect to changes in these hyperparameters in Section 4.

3.5. Performance Evaluation. In order to assess and compare clustering methods, we report different performance scores that we briefly recap below. The first criterion is the overlap between the estimated clusters and some *ground truth* labels, assuming, in the context of method validation, that these are known. For classification tasks, accuracy is a key metric to evaluate model performance. One popular approach to report accuracy is based on the so-called confusion matrix,

Algorithm 4 Reference point selection based on approximate Fréchet means. We use the Fréchet means of approximate clusters $\{\widetilde{CL}_i\}_{i=1}^k$ found via the log-Euclidean framework to define an auxiliary geodesic $\bar{\gamma}_{ij}$ that is used to position two reference points per cluster pair.

- 1: **Input:** A set of points $X_1, \dots, X_N \in SPD(n)$, the number of clusters k , and hyperparameters $t_{\text{close}}, t_{\text{far}}, N_\rho$, and ϵ_d
 - 2: $\mathcal{E} \leftarrow \{\log_{\text{Id}}(X_1), \dots, \log_{\text{Id}}(X_N)\}$
 - 3: $\{\widetilde{CL}_i\}_{i=1}^k \leftarrow$ apply Euclidean k -means to $\mathcal{E}$
 - 4: $\{\tilde{M}_i\}_{i=1}^k \leftarrow$ compute Fréchet means on $SPD(n)$ for each cluster $\widetilde{CL}_i$ using ICM
 - 5: $\{\tilde{\rho}_i\}_{i=1}^k \leftarrow$ estimate Riemannian radius $\tilde{\rho}_i$ for each $\widetilde{CL}_i$ by drawing N_ρ random samples X_j within $\widetilde{CL}_i$; compute 90% quantile ρ_i of the distances $d(\tilde{M}_i, X_j)$
 - 6: $\{t_{ij}\} \leftarrow$ given $\tilde{M}_i, \tilde{M}_j, \tilde{\rho}_i, \tilde{\rho}_j$, determine t_{ij} based on Equation (12) for each $i, j = 1, \dots, k, i \neq j$
 - 7: $\{\bar{M}_{ij}\} \leftarrow$ compute midpoints

$$\bar{M}_{ij} := \gamma_{\tilde{M}_i \tilde{M}_j}(1/2)$$
 for each pair of distinct clusters $\{\widetilde{CL}_i, \widetilde{CL}_j\}, i, j = 1, \dots, k, i \neq j$, via Equation (11)
 - 8: $\{R_m\}_{m=1}^\ell \leftarrow$ given t_{ij} and $\{\bar{M}_{ij}\}$ compute two reference points (in symmetric location) based on the auxiliary geodesic $\bar{\gamma}_{ij}$ Equation (13) evaluated at $\pm t_{ij}$
 - 9: **Output:** Reference points $\{R_m\}_{m=1}^\ell \subset SPD(n)$
-

which compares true labels with predicted labels. In clustering scenarios, since there is no *a priori* correspondence between the cluster indices and the original labels, one needs to first solve an optimal assignment problem to find the best permutation of the estimated clusters that match them to *ground truth* groups. This is classically done using, e.g., the *Hungarian method* [18]. We apply this exact approach to obtain correspondences between clusters and *ground truth* labels from which we can then compute the confusion matrix between the two sets of clusters. We then define the clustering accuracy as the proportion of data samples assigned to the correct group.

A second quantitative measure of clustering is the total Riemannian dispersion, which is the objective function of the intrinsic Riemannian k -means method presented in Section 3.3. If $\hat{\mathcal{P}} = (\hat{CL}_1, \dots, \hat{CL}_k)$ are the output clusters for any given method, the Riemannian dispersion is given by Equation (1), where $d_{\mathcal{M}}$ is the affine-invariant Riemannian distance on $SPD(n)$. To provide an easier-to-interpret score, whenever some *ground truth* groups $\mathcal{P} = (CL_1, \dots, CL_k)$ are known for the considered example, we also report the normalized dispersion:

$$(14) \quad \overline{\text{totdisp}}(\hat{\mathcal{P}}) = \frac{\text{totdisp}(\hat{\mathcal{P}})}{\text{totdisp}(\mathcal{P})}.$$

We note that $\mathcal{P}$ may not necessarily correspond to the minimal total dispersion among all partitions in general. Consequently, the above normalized total dispersion may, in some cases, be smaller than one. As such, this should be understood as a mere renormalization to make the obtained values of the dispersion more interpretable across different simulations.

4. NUMERICAL EXPERIMENTS

In this section, we present extensive numerical experiments using both synthetic and real-world data in $SPD(n)$ to illustrate our FMC algorithm.

4.1. Reference Point Selection. Our first set of experiments aimed to validate the strategies for the selection of reference points in the FMC algorithm we presented in Section 3.4 using synthetic data.

TABLE 1. FCM results for a synthetic benchmark dataset of k balls with random centers C_i and random radii ρ_i . For each choice of ℓ (number of reference points), we generate 100 benchmark sets of disjoint balls. We report the average clustering accuracy for varying ℓ and k (ground truth data is available).

k	ℓ	mean (%)	std	min (%)	10th (%)
2	2	99.28%	0.010	96.00%	98.00%
	4	99.83%	0.004	98.00%	99.00%
	6	99.92%	0.003	98.00%	100.00%
3	3	94.54%	0.090	68.00%	76.70%
	6	98.46%	0.040	74.00%	97.90%
	9	99.74%	0.006	97.00%	99.00%
4	6	94.57%	0.070	73.00%	81.70%
	12	98.96%	0.015	91.00%	97.90%
	18	99.69%	0.005	98.00%	99.00%
5	10	95.26%	0.040	79.00%	88.70%
	20	98.48%	0.012	95.00%	97.00%
	30	98.95%	0.007	96.00%	98.00%

For these experiments, we generated multiple balls B_i , $i = 1, \dots, k$, in $SPD(4)$ by randomly sampling their centers C_i and assigning corresponding radii ρ_i drawn uniformly from $[0.8, 1.2]$. The centers are retained only if their normalized pairwise distances satisfy

$$(15) \quad d_{\text{low}} \leq \frac{d(C_i, C_j)}{\rho_i + \rho_j} \leq d_{\text{up}}, \quad 1 \leq i \neq j \leq k,$$

for some user-defined parameters $d_{\text{low}}, d_{\text{up}}$. As normalized pairwise distances close to 1 make the balls harder to separate, in our experiments below, we selected these parameters close to 1. For each ball, we generated 4000 samples uniformly distributed within their boundary.

4.1.1. Random reference point selection. Purpose: To assess the performance of the FCM algorithm using the random reference point selection strategy presented in Section 3.4.1, for various choices of the number of clusters k .

Setup: We generated multiple benchmark synthetic data, as described above, with the number of balls $k \in \{2, 3, 4, 5\}$. For each value of k , we generated 100 instances using a varying number of reference points ℓ chosen randomly within the dataset.

Results: We report in Table 1 the result of the FCM algorithm. The table includes the mean accuracy (averaged over the 100 runs), its standard deviation, the minimum accuracy (i.e., the worst performance observed), and the 10th percentile (the accuracy threshold below which the lowest 10% of runs fall).

Conclusion: For a fixed k , increasing ℓ steadily improves the mean accuracy. Random reference point selection does not guarantee very high clustering accuracy, especially as the manifold dimension increases. Specifically: (1) for fixed k , increasing ℓ steadily improves mean accuracy and reduces variability (higher minima and 10th percentiles); (2) as k increases, a larger ℓ is required to achieve near-perfect accuracy; (3) for $k = 5$, the performance plateaus below 99% even at $\ell = 30$, indicating diminishing returns and a clear gap relative to $k \leq 4$.

4.1.2. Reference point selection based on approximate Fréchet means. Purpose: To assess the performance of the FMC algorithm using the reference point selection approach in Section 3.4.2 for various choices of the hyperparameters t_{close} and t_{far} .

Setup 1: We created a benchmark synthetic dataset using $k = 5$ balls in $SPD(4)$ and generated 100 random configurations with $d_{\text{low}} = 1.1$ and $d_{\text{up}} = 3$. Following Equation (12) with $\epsilon_d = 2.5$

(found heuristically to yield satisfactory results), we vary

$$t_{\text{close}} \in \{1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\} \quad \text{and} \quad t_{\text{far}} \in \{0.30, 0.35, 0.40\}.$$

Results: For each of the 100 data configurations with any choice of t_{close} and t_{far} , the FMC algorithm achieved a mean clustering accuracy of 100% and a mean normalized dispersion ratio of 1; standard deviation was negligible in both cases.

Setup 2: We constructed 500 benchmark datasets with $k = 2$ balls in $SPD(4)$. For the “close case” we set $d_{\text{low}} = 1.10$ and $d_{\text{up}} = 1.15$. Similarly to the experiment above, we selected

$$t_{\text{close}} \in \{1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}.$$

For the “far case” we set $d_{\text{low}} = 2.50$ and $d_{\text{up}} = 5.00$ and choose $t_{\text{far}} \in \{0.30, 0.35, 0.40\}$.

Since our algorithm relies on random sampling to estimate the Riemannian radii, we repeated the experiment for each choice of t_{far} and t_{close} 100 times. For each choice, we set the reference points to $R_1 = \bar{\gamma}_{12}(-t_{ij})$ and $R_2 = \bar{\gamma}_{12}(-t_{ij})$, respectively, as described above.

Results: Across all runs, we consistently observed that the clustering accuracy and the normalized dispersion are equal to 1.

Conclusion: The FMC algorithm achieved a very high accuracy and normalized dispersion for all choices of t_{close} and t_{far} considered in this experiment.

4.1.3. *Comparison of reference point selection strategies.* **Purpose:** To compare the performance of the FMC algorithm under the two strategies for the selection of reference points presented in Section 3.4.

Setup: We adopted the same strategy outlined in the above experiment to generate the benchmark data sets in $SPD(4)$. We considered two strategies to select the reference points. For the principled reference point selection strategy of Section 3.4.2, we followed the same approach as in the above experiment and selected 20 reference points after fixing $t_{\text{close}} = 2$ and $t_{\text{far}} = 0.35$. For the simpler random selection strategy presented in Section 3.4.1, the 20 reference points were selected randomly from the dataset.

Results: The FMC algorithm in combination with the principled reference point selection strategy of Section 3.4.2 achieved a mean clustering accuracy of 100% (with negligible standard deviation). In contrast, when we used the simpler random selection strategy of Section 3.4.1, mean accuracy dropped to 98.48% with a standard deviation of 0.012. If the number of reference points increases to 30, the mean accuracy is 98.95% with a standard deviation of 0.007.

Conclusion: The reference-point placement informed by the approximate Fréchet means of Section 3.4.2 is more sample-efficient and stable than naive random selection.

4.2. **A more challenging configuration.** **Purpose:** To examine the impact of choosing 1– vs 2– Fréchet map in the FCM approach and compare the FCM approach with the log-Euclidean (LEC) embedding.

Setup: We considered 4 disjoint balls of fixed radii $r = 1$ in $SPD(4)$, placed in a challenging configuration. Namely, we generated the center of the first ball according to $C_1 = \exp(V_1)$ where V_1 is a random matrix in $Sym(4)$ (with each entry drawn from a centered normal distribution) which is then rescaled so that $\|V_1\|_F = 12$. In other words, C_1 is obtained by moving from Id in a random direction up to a distance of 12 from Id . We then selected the second center as $C_2 = \exp(V_2)$ where $V_2 = V_1 + \Delta$ is a small random perturbation of V_1 with Δ being a random matrix of $Sym(4)$ with entries drawn from a centered normal distribution of variance 0.1. To ensure that the resulting Riemannian balls $B(C_1, 1)$ and $B(C_2, 1)$ remain disjoint, we only kept those two centers when $d(C_1, C_2) > 2$. The centers of the last two balls were then set to $C_3 = C_1^{-1} = \exp(-V_1)$ and $C_4 = C_2^{-1} = \exp(-V_2)$. This ensures, on the one hand, that all 4 balls are disjoint from one another since $d(C_3, C_4) = d(C_1, C_2) > 2$ (the equality itself follows from the expression of the

TABLE 2. Comparison of 1-Fréchet map (1-FCM), squared Fréchet map (2-FCM), and the log-Euclidean clustering (LEC) methods. We report the mean accuracy (ground truth labels are available) and dispersion for the three methods over 250 draws of four balls in $SPD(4)$, selected randomly according to the scheme described in Section 4.2. We report the mean accuracy and the mean dispersion along with their standard deviations σ (in brackets).

Method	Mean Accuracy (std σ)	Mean Dispersion (std σ)
1-FMC (ours)	89.61% (17.27%)	2.11 (2.25)
2-FMC (ours)	69.52% (20.64%)	5.58 (4.39)
LEC	62.08% (6.44%)	10.39 (6.26)

TABLE 3. Comparison of FCM, ARC, and IRC on synthetic data sampled from four disjoint balls in $SPD(20)$. We report the runtime (in seconds), the accuracy, and the normalized dispersion (from left to right).

Method	Runtime	Speedup	Accuracy	Normalized Dispersion
IRC	832	—	94.1%	1.17
ARC	513	8.8×	91.6%	1.26
2-FCM (ours)	58	14.3×	100.0%	1.00

distance Equation (6)), while also making the configuration of those balls symmetric and thus centered around Id.

We generated 250 realizations of the above scheme, and, for each ball, generated 5000 samples drawn uniformly inside it according to the same approach described in the above experiments. We then evaluated the clustering accuracy and normalized dispersion of the 1-Fréchet and 2-Fréchet map (1-FCM and 2-FCM) methods (using the reference point selection strategy described in Section 3.4.2 and deployed in the numerical example above), as well as the LEC approach.

Results: Results are reported in Table 2.

Conclusion: The 1-FCM method achieved the best accuracy and dispersion, outperforming the 2-FCM and LEC approaches. We explain the poor performance of LEC (worst performance) by the fact that geodesic distances are not well approximated in the log-Euclidean setting for points far away from Id, as is the case here. The squared Fréchet map method leads to slightly improved accuracy and dispersion, yet still appears to suffer from the distortion induced by the mapping. This experiment shows that the Fréchet map method, especially the 1-Fréchet map, produces much lower metric distortion (cf. Proposition 1) compared to LEC. The difference becomes particularly evident when the manifold data configuration is not well reflected by its tangent space approximation.

4.3. Efficiency comparison with IRC and ARC. Purpose: To compare the proposed FCM algorithm against the IRC and ARC algorithms on high-dimensional SPD data.

Setup: We generated four disjoint balls in $SPD(20)$ where the radius of each ball is randomly selected within the interval $[0.8, 1.2]$. We sampled 4000 points uniformly within each ball. For the implementation of the FCM algorithm, we generated the reference points using the strategy presented in Section 3.4.2 and already utilized in the above numerical experiments.

Results: Table 3 reports the clustering performance of the FCM, IRC, and ARC algorithms. Performance numbers, including average accuracy, normalized dispersion, and running time, are averages over 50 replications.

Conclusion: The FCM algorithm with $t_{\text{close}} = 5$ and $t_{\text{far}} = 0.35$ achieves clustering accuracy comparable to IRC and ARC, while being approximately fourteen times faster than IRC and nine times faster than ARC.

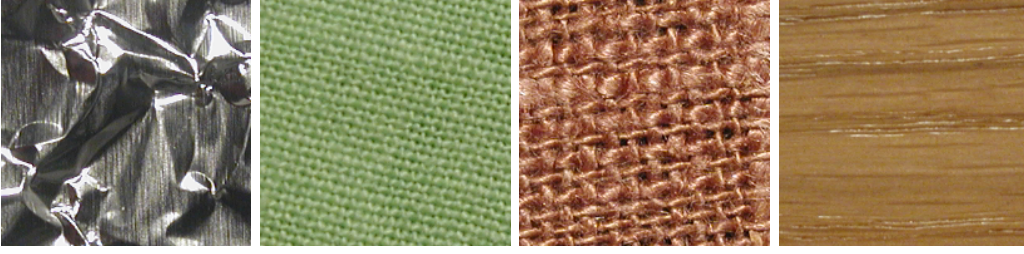


FIGURE 5. Representative datasets from the four textures (aluminum foil, cotton, linen, and wood) considered in this study. These images are taken from the KTH-TIPS2b dataset.

TABLE 4. Comparison of FCM, ARC, and IRC on the KTH-TIPS2b texture dataset. We report the runtime (in seconds), the speedup, the clustering accuracy (ground truth textures are available), and the normalized dispersion.

Method	Runtime	Speedup	Accuracy	Normalized Dispersion
IRC	174	—	77.42%	0.68
ARC	66	2.6×	77.67%	0.69
1-FCM	7	24.9×	76.33%	0.69
2-FCM	8	21.8×	76.15%	0.69
LEC	3	58.0×	75.84%	0.69

4.4. FCM algorithm on the texture dataset. Purpose: To test the performance of the FCM algorithm on real data. Here, we consider images of texture.

Setup: We select four categories (see Figure 5) from the KTH-TIPS2b texture dataset [9]. Each category has 432 images. We resize the original images to 128×128 pixels. We generate covariance descriptors as follows. For each pixel location (u, v) in the 128×128 image, where $1 \leq u, v \leq 128$, we compute a 23-dimensional feature vector

$$x_{u,v}^m = [r_{u,v}, g_{u,v}, b_{u,v}, |G_{u,v}^{0,0}(m)|, \dots, |G_{u,v}^{4,3}(m)|]^T \in \mathbb{R}^{23},$$

where $G_{u,v}^{o,s} \in \mathbb{R}$, $0 \leq o \leq 4$, $0 \leq s \leq 3$ are the Gabor filter coefficients of the image $m \in \{1, \dots, 1728\}$ centered at (u, v) defined in Section 2.1 of [33]. Next, for each image, we compute the 23×23 covariance matrix

$$C_m = \frac{1}{N-1} \sum_{u=1}^{128} \sum_{v=1}^{128} (x_{u,v}^m - \mu_{u,v}^m)(x_{u,v}^m - \mu_{u,v}^m)^T \in \mathbb{R}^{23,23},$$

where $\mu_{u,v}$ is the mean of feature vector $x_{u,v} \in \mathbb{R}^{23}$. In summary, we obtain a set of 1728 covariance matrices C_k , $k = 1, \dots, 1728$ of size 23×23 .

We applied four clustering methods: the proposed FCM algorithm with $t_{\text{close}} = 5$ and $t_{\text{far}} = 0.35$, and the algorithms ARC, IRC, and LEC.

Results: Table 4 reports the results (averaged over 50 repetitions to get more precise estimates), including the average runtime, accuracy, and normalized dispersion.

Conclusion: The FCM algorithm achieves results comparable to those of the ARC and IRC methods in terms of clustering performance. The FMC approach is significantly faster than ARC and IRC, with LEC being slightly faster than FCM, but not as accurate. LEC is faster than FCM but slightly less accurate.

4.5. FCM algorithm on the COBRE resting-state fMRI dataset. Purpose: To test the performance of the FCM algorithm on real data. Here, we consider an fMRI dataset.

TABLE 5. Comparison of FCM, ARC, LEC, and IRC on the COBRE dataset. We report (from left to right) the runtime (in seconds), the speedup compared to IRC, the accuracy, and the normalized dispersion.

Method	Runtime	Speedup	Accuracy	Normalized Dispersion
IRC	11 976	—	60.96%	0.95
ARC	203	60×	61.73%	0.88
1-FCM (ours)	20	600×	61.05%	0.88
2-FCM (ours)	20	600×	61.73%	0.88
LEC	18	600×	61.25%	0.88

Setup: We used the COBRE resting-state fMRI dataset from [11], which contains two diagnostic categories, schizophrenia (SCZ) and healthy control (Control), for a total of 146 subjects in total (72 SCZ and 74 Control). Using the released precomputed functional connectivity features with 197-voxel region-of-interest resolution, we calculated one 197×197 SPD matrix per subject.

Each subject’s connectivity matrix is provided as an upper-triangular matrix without the diagonal. We recovered the full symmetric matrix A by completion, setting the diagonal entries to 1 (correlation convention), and adding a small perturbation $\epsilon > 0$ to the identity to ensure strict positive-definiteness. That is, we obtained SPD matrices as

$$A \leftarrow \frac{1}{2}(A + A^T) + \epsilon \text{Id}$$

with $\epsilon = 1.00\text{e-}6$.

We applied four clustering approaches on the COBRE dataset: our proposed FCM algorithm with $t_{\text{close}} = 5$ and $t_{\text{far}} = 0.35$, and the algorithms ARC, IRC, and LEC.

Results: Table 5 reports the mean performance over 10 repeated runs, including average accuracy, normalized dispersion, runtime, and speedup factor.

Conclusion: All algorithms achieve comparable accuracy, while FCM and LEC exhibit significantly lower computational cost. We note that the accuracy for all approaches is fairly low in this particular example, while the normalized dispersions are all below 1, which reflects the fact that the two groups in this dataset are not necessarily well-separated, at least based on the k-means approach.

5. CONCLUSION AND FUTURE PERSPECTIVES

We introduced and evaluated a new variant of the classical k-means method for clustering data on non-Euclidean spaces. Our approach is simple to implement, easy to parallelize, and fast, while able to closely match the results of the intrinsic k-means scheme on the manifold with some adequate choice of reference points. The main observations are:

- The runtime of the proposed algorithms is at least one order of magnitude faster than performing clustering on the manifold without sacrificing accuracy.
- The performance of the proposed approach is consistent across a series of synthetic and real-data examples. Although the Log-Euclidean embedding approach remains slightly faster, its accuracy can deteriorate for certain data configurations. We demonstrated empirically that the accuracy of our method remains more stable in such situations.
- We investigated some theoretical properties of Fréchet maps that could provide a fundamental mathematical underpinning to our proposed FMC clustering approach. Although one can derive a quite complete picture for Fréchet maps on Euclidean spaces, we found that a similar analysis becomes much more challenging on the $SPD(n)$ spaces and a fortiori on more general Cartan-Hadamard manifolds. However, we anticipate that this preliminary exploration will draw the attention of the applied mathematics community and pave the way for future work on this topic.

From a wider perspective, we believe that Fréchet maps could be further leveraged as a general parametric class of mappings on manifolds for applications beyond the clustering problem considered in this work, and be used within e.g. autoencoder architectures to estimate optimal latent feature space embeddings for manifold datasets.

ACKNOWLEDGMENTS

The authors thank David Pierucci for his help in the characterization of totally geodesic hypersurfaces of $SPD(n)$, as well as Alice Barbora Tumpach and Xavier Pennec for several stimulating discussions on topics related to this work.

APPENDIX A. HARDWARE AND SOFTWARE LIBRARIES

All experiments were executed on the Carya Cluster—a modern computing system hosted by the Research Computing Data Core of the Hewlett-Packard Enterprise Data Science Institute of the University of Houston. Carya hosts a total of 9,984 Intel CPU cores and 327,680 Nvidia GPU cores integrated within 188 compute and 20 GPU nodes, equipped with Intel Xeon G6252 CPUs and NVIDIA V100 GPUs. All nodes are equipped with solid-state drives for local high-performance data storage. Our code was implemented in `Python` version 3.12.3. Some of the modules used in our code are based on `scikit-learn` version 1.5.1 and `scikit-image` version 0.24.0. Upon acceptance of this article, our code will be released on GitHub at <https://github.com/jishi24/Frechet-Clustering>.

APPENDIX B. PROOF OF THEOREM 2

Let us first examine the characterization of the image of the Fréchet map F_r . Let $d = (d_1, \dots, d_m) \in \mathbb{R}^m$ such that $d = F_r(x)$ for some $x \in \mathbb{R}^m$, i.e., $d_i = \|x - r_i\|^2$, $i = 1, \dots, m$. By translation invariance, we may assume that $r_m = 0$. Then, due to the assumption made on the reference points, the vectors $r_i = r_i - r_m$ for $i = 1, \dots, m-1$ are linearly independent. Their Gram matrix $G = (r_i^\top r_j) \in \mathbb{R}^{(m-1) \times (m-1)}$ is thus positive definite. We then define the vectors z, b , and u in $\mathbb{R}^{m-1}$ by $z_i = d_i - d_m = \|r_i\|^2 - 2r_i^\top x$, $b_i = \|r_i\|^2$, and $u_i = r_i^\top x$ for all $i = 1, \dots, m-1$. It results that $u = \frac{1}{2}(b - z)$.

We now aim to express x with respect to u and d_m . To that end, we write $x = x_{H^0} + sn$, where $s \in \mathbb{R}$, H^0 is the hyperplane spanned by the r_i 's, n the unit normal vector to that hyperplane, and x_{H^0} is the projection of x onto H^0 . Since u is the vector of inner products of x and the r_i 's for $i = 1, \dots, m-1$, one immediately has that $x_{H^0} = \sum_{i=1}^{m-1} \alpha_i r_i$ with $\alpha = G^{-1}u$, from which we also get $\|x_{H^0}\|^2 = u^\top G^{-1}u$. Then $\|x\|^2 = \|x_{H^0}\|^2 + s^2$, i.e., $s^2 = d_m - u^\top G^{-1}u$. It follows that a vector d in $\mathbb{R}^m$ belongs to the image $F_r(\mathbb{R}^m)$ if and only if $u^\top G^{-1}u \leq d_m$; that is:

$$F_r(\mathbb{R}^m) = \left\{ d \in \mathbb{R}^m : \begin{pmatrix} \|r_1\|^2 + d_m - d_1 \\ \vdots \\ \|r_{m-1}\|^2 + d_m - d_{m-1} \end{pmatrix}^\top G \begin{pmatrix} \|r_1\|^2 + d_m - d_1 \\ \vdots \\ \|r_{m-1}\|^2 + d_m - d_{m-1} \end{pmatrix} \leq 4d_m \right\}.$$

We note that the above expression is the equation of the interior of a paraboloid in $\mathbb{R}^m$, proving the first claim in Theorem 2. The above equations also provide the inverse of F_r for d in the above set, which is obtained by taking $s = \pm \sqrt{d_m - u^\top G^{-1}u}$, with the two possible solutions corresponding to two symmetric points x^-, x^+ on each side of the hyperplane H^0 as illustrated in Figure 1.

Assume now that $B(x_0, \rho_0)$ is a ball such that $B(x_0, \rho_0) \subset H^+$ (the same argument applies in H^-). Denoting $x_0^\perp = n^\top x_0$ the component of x_0 normal to H^0 which is also the distance from x_0 to the hyperplane, we have $x_0^\perp > \rho_0$. Using the notations introduced previously, the condition

that $x = F_r^{-1}(d)$ belongs to $B(x_0, \rho_0)$ is equivalent to:

$$\begin{aligned} \rho_0^2 &\geq \|x - x_0\|^2 = d_m - 2x^\top x_0 + \|x_0\|^2 \\ &= d_m - 2x_{H^0}^\top x_0 - 2sn^\top x_0 + \|x_0\|^2 \\ &= d_m - 2 \sum_{i=1}^{m-1} (G^{-1}u)_i r_i^\top x_0 - 2\sqrt{d_m - u^\top G^{-1}u} x_0^\perp. \end{aligned}$$

Let us define the vector $u_0 := (r_1^\top x_0, \dots, r_{m-1}^\top x_0)^\top \in \mathbb{R}^{m-1}$. Then the second term on the right-hand side can be rewritten according to:

$$\sum_{i=1}^{m-1} (G^{-1}u)_i r_i^\top x_0 = u_0^\top G^{-1}u.$$

Therefore, $x \in B(x_0, \rho_0)$ if and only if

$$\sqrt{d_m - u^\top G^{-1}u} x_0^\perp \geq \frac{d_m}{2} - u_0^\top G^{-1}u + \frac{\|x_0\|^2 - \rho_0^2}{2},$$

which, by squaring both sides, is also equivalent to

$$(x_0^\perp)^2 (d_m - u^\top G^{-1}u) \geq \left(\frac{d_m}{2} - u_0^\top G^{-1}u + \frac{\|x_0\|^2 - \rho_0^2}{2} \right)^2.$$

This in turn can be expressed as

$$(16) \quad (x_0^\perp)^2 u^\top G^{-1}u + \left(\frac{d_m}{2} - u_0^\top G^{-1}u \right)^2 + g(d) \leq C,$$

where $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is some linear function of d and $C \in \mathbb{R}_+$ is a constant that we do not explicitly specify to keep the expression compact. As (u, d_m) is itself a linear invertible function of d , the above describes the interior of a quadric of $\mathbb{R}^m$. Furthermore, since $x_0^\perp > 0$ and G^{-1} is positive definite, we see that $(u, d_m) \mapsto (x_0^\perp)^2 u^\top G^{-1}u + \left(\frac{d_m}{2} - u_0^\top G^{-1}u \right)^2$ is a positive definite quadratic form on $\mathbb{R}^m$. Thus, the set of d that satisfies Equation (16) is the interior of an ellipsoid. We conclude that $F_r(B)$ is the interior of an ellipsoid of $\mathbb{R}^m$ (intersected with the interior of the paraboloid from the previous paragraph).

Finally, the last statement in the theorem is simply a consequence of the Hahn–Banach theorem since $F_r(B_1)$ and $F_r(B_2)$ are two convex subsets of $\mathbb{R}^m$ (as the intersections of the inside of a paraboloid and an ellipsoid) and are also disjoint owing to the injectivity of F_r on H^+ given by Theorem 1.

APPENDIX C. PROOF OF THEOREM 3

Unlike the case of 2-Fréchet maps, it is not as simple to geometrically characterize the image of a ball under F_r^1 . Instead, our proof of convexity relies on showing the positivity of the curvature of the image set boundary. Let us assume, without loss of generality, that $A \subset H^+$. First of all, from Remark 2, we know that F_r^1 is a diffeomorphism on the halfspace H^+ . Let B be a ball in A , of radius $\rho > 0$, and $S = \partial B$ the boundary sphere. We will use the following expression for the second fundamental form of the image $F_r^1(S) \subset \mathbb{R}_+^m$, which follows from standard results in Riemannian geometry of embedded submanifolds:

Lemma 2. *For any $x \in S$, and $u, v \in T_x S$, the second fundamental form of the hypersurface $F_r^1(S)$ satisfies*

$$(17) \quad \mathbb{I}_{F_r^1(x)}(DF_r^1(x) \cdot u, DF_r^1(x) \cdot v) = \langle D^2 F_r^1(x)(u, v) + \frac{1}{\rho} \langle u, v \rangle DF_r^1(x) \cdot n, n' \rangle,$$

where n denotes the unit outward normal vector to S at x and n' the unit outward normal to $F(S)$ at $F(x)$.

Now, given two linearly independent tangent vectors u, v to S at x and their pushforward $u' = DF_r^1(x) \cdot u, v' = DF_r^1(x) \cdot v \in T_{F_r^1(x)} F_r^1(S)$, the sectional curvature of $F_r^1(S)$ associated to the plane spanned by (u', v') is classically given by (c.f. [12]):

$$(18) \quad K(u', v') = \frac{\mathbb{I}(u', u')\mathbb{I}(v', v') - \mathbb{I}(u', v')^2}{\|u'\|^2\|v'\|^2 - \langle u', v' \rangle^2}$$

where we dropped the dependency in x and $F_r^1(x)$ to lighten notation. Since the denominator is positive for any linearly independent vectors u, v , we just need to show that, under the adequate conditions on the reference points, the second fundamental form remains positive definite. To do so, we need the following expressions for the first and second-order differentials of the Fréchet map:

$$DF_r^1(x) \cdot u = \begin{pmatrix} \langle \frac{x-r_1}{\|x-r_1\|}, u \rangle \\ \vdots \\ \langle \frac{x-r_m}{\|x-r_m\|}, u \rangle \end{pmatrix}, \quad D^2F_r^1(x)(u, v) = \begin{pmatrix} \frac{1}{\|x-r_1\|} \langle \Pi_1 u, v \rangle \\ \vdots \\ \frac{1}{\|x-r_m\|} \langle \Pi_m u, v \rangle \end{pmatrix}$$

where Π_i , for $i = 1 \dots, m$, denotes the orthogonal projection onto the hyperplane perpendicular to $x - r_i$. Then, using (17) and the fact that $n' = \frac{DF_r^1(x)^{-T}(n)}{\|DF_r^1(x)^{-T}(n)\|}$, we see that for all $x \in S \subset A$ and tangent vector $u \in T_x S$:

$$(19) \quad \begin{aligned} \mathbb{I}(u', u') &= \frac{1}{\rho} \|u\|^2 \langle DF_r^1(x) \cdot n, n' \rangle + \langle D^2F_r^1(x)(u, u), n' \rangle \\ &= \frac{1}{\rho} \|u\|^2 (n^T DF_r^1(x)^{-1} DF_r^1(x)^{-T} n) + \sum_{i=1}^m \frac{n'_i}{\|x - r_i\|} u^T \Pi_i u \\ &\geq \frac{\sigma_{\min}^2(DF_r^1(x))}{\rho} \|u\|^2 - \frac{\sqrt{m}}{d(r, A)} \|u\|^2 \end{aligned}$$

where $d(r, A) = \min_{i=1, \dots, m} d(r_i, A)$ denotes the distance of the reference point set to the compact A . We note that the right hand side term can be made arbitrarily small by selecting reference points far enough from A . On the other hand, one still needs to control the decrease of the first term and specifically of $\sigma_{\min}^2(DF_r^1(x))$, the smallest singular value of the first differential of F_r^1 . Given the expression of DF_r^1 computed above, this corresponds to the smallest eigenvalue of the Gram matrix for the m unit vectors $\left(\frac{x-r_1}{\|x-r_1\|}, \dots, \frac{x-r_m}{\|x-r_m\|} \right)$. By the Gershgorin circle theorem, this eigenvalue is bounded from below by

$$\begin{aligned} \sigma_{\min}^2(DF_r^1(x)) &\geq 1 - \max_{i=1, \dots, m} \sum_{j \neq i} \left| \left\langle \frac{x-r_i}{\|x-r_i\|}, \frac{x-r_j}{\|x-r_j\|} \right\rangle \right| \\ &\geq 1 - \max_{i=1, \dots, m} (m-1) \max_{j \neq i} \left| \left\langle \frac{x-r_i}{\|x-r_i\|}, \frac{x-r_j}{\|x-r_j\|} \right\rangle \right| \\ &\geq 1 - (m-1) \max_{j \neq i} \left| \left\langle \frac{x-r_i}{\|x-r_i\|}, \frac{x-r_j}{\|x-r_j\|} \right\rangle \right|. \end{aligned}$$

Now, by taking the minimum of these lower bounds over the compact set A , we get that for all $x \in A$, $\sigma_{\min}^2(DF_r^1(x)) \geq 1 - (m-1)\mu$. Thus, going back to (19):

$$\mathbb{I}(u', u') \geq \left(\frac{1 - (m-1)\mu}{\rho} - \frac{\sqrt{m}}{d(r, A)} \right) \|u\|^2.$$

Therefore, the second fundamental form is positive definite as soon as $\frac{\rho}{d(r,A)} < \frac{1-(m-1)\mu}{\sqrt{m}}$. In that case, all sectional curvatures $K(u', v')$ in (18) are positive. Since $F_r^1(S)$ is the diffeomorphic image of a sphere, it is a complete manifold and, thus, by the result of [28], we deduce that $F_r^1(B)$ is convex.

APPENDIX D. PROOF OF PROPOSITION 3

Here we derive the complete list of the totally geodesic hypersurfaces of $SPD(n)$ for $n \geq 3$ as stated in the proposition, and also cover the case $n = 2$ for completeness. Let us first state an equivalent to the Lie triple system condition characterizing totally geodesic submanifolds in $SPD(n)$:

Theorem 7 (Corollary 1.2 in [34]). *Any totally geodesic submanifold of $SPD(n)$ is of the form $S = Q \exp(H)Q$ with $Q \in GL(n)$ and H a subspace of $Symm(n)$ that satisfies:*

$$[V, [V, W]] \in H, \quad \text{for any } V, W \in H,$$

where $[V, W] \doteq VW - WV$ is the usual Lie Bracket on matrices.

We further point out that the above Lie triple system condition is also equivalent (via a simple linearity argument) to the fact that $[[U, V], W] \in H$ for any $U, V, W \in H$. Based on this theorem, we can thus first focus on determining which codimension 1 subspaces H satisfy the above Lie triple system condition.

• **Characterization of H :** let H be a $(m-1)$ -dimensional subspace of $Symm(n)$ such that $[V, [V, W]] \in H$ for any $V, W \in H$. There exists $A \in Symm(n)$ with $A \neq 0$ for which $H = \{V \in Symm(n) : \text{Tr}(AV) = 0\}$, i.e., H is the orthogonal subspace to A . We may diagonalize A as $A = R^T D R$ with $D = \text{diag}(d_i)_{i=1, \dots, n}$ and $R \in O(n)$. Then $\text{Tr}(AV) = 0$ is equivalent to $\text{Tr}(D R V R^T) = 0$ so we can write $H = R^T H_D R$, where H_D is the space of symmetric matrices orthogonal to D . It is then clear that H is totally geodesic if and only if H_D is totally geodesic since $P \mapsto R^T P R$ is an isometry of $SPD(n)$.

We first treat the case $n \geq 3$. Since $D \neq 0$, without loss of generality, we assume that $d_1 \neq 0$. Let us denote by $\{E_{ij}\}_{i,j=1, \dots, m}$ the canonical basis of $\mathbb{R}^{m \times m}$, i.e., $E_{ij} = (\delta_{\{k=i, l=j\}})_{k,l}$. Then one can check that the following is a basis for $H_D = \{V \in Symm(n) : \text{Tr}(DV) = 0\}$:

$$\mathcal{B} = \left\{ (E_{ij} + E_{ji})_{i \neq j}, \left(\frac{d_i}{d_1} E_{11} - E_{ii} \right)_{i \geq 2} \right\}.$$

In particular, if H_D is totally geodesic then for any i, j with $i \neq 1, j \neq 1$ and $i \neq j$ (which exist since $n \geq 3$), we must have:

$$(20) \quad \left[E_{1,j} + E_{j,1}, \left[\frac{d_i}{d_1} E_{1,1} - E_{i,i}, E_{1,j} + E_{j,1} \right] \right] \in H_D.$$

After calculations, we find that:

$$\begin{aligned} \left[\frac{d_i}{d_1} E_{1,1} - E_{i,i}, E_{1,j} + E_{j,1} \right] &= \frac{d_i}{d_1} (E_{1,j} - E_{j,1}) \\ \left[E_{1,j} + E_{j,1}, \left[\frac{d_i}{d_1} E_{1,1} - E_{i,i}, E_{1,j} + E_{j,1} \right] \right] &= \frac{d_i}{d_1} [E_{1,j} + E_{j,1}, E_{1,j} - E_{j,1}] \\ &= 2 \frac{d_i}{d_1} (E_{j,j} - E_{1,1}). \end{aligned}$$

from which it follows that condition Equation (20) is equivalent to $2 \frac{d_i d_j}{d_1} - 2 \frac{d_i d_1}{d_1} = 0$, i.e., $d_j = d_1$ for any $j \neq 1$. Thus, the matrix D is of the form λId for some $\lambda \neq 0$ and so is $A = R^T D R$. It follows that H is the space of matrices with zero trace.

In the case $n = 2$, without loss of generality, one may write, up to a scaling factor, $D = \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}$, where $d \in \mathbb{R}$ remains to be determined. We see that a basis for the linear subspace H_D is then given by:

$$\mathcal{B} = \left\{ B_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} d & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

By the remark below, Theorem 7, and the bilinearity of the Lie bracket operation, the triple Lie bracket condition is also equivalent to $U, V, W \in \mathcal{B} \implies [U, [V, W]] \in H_D$. Computation of the different possible triple Lie brackets leads to:

$$\begin{aligned} [B_1, B_2] &= B_1 B_2 - B_2 B_1 = \begin{pmatrix} 0 & -d-1 \\ d+1 & 0 \end{pmatrix}, \\ [B_1, [B_1, B_2]] &= B_1 [B_1, B_2] - [B_1, B_2] B_1 = \begin{pmatrix} 2d+2 & 0 \\ 0 & -2d-2 \end{pmatrix}, \\ [B_2, [B_1, B_2]] &= B_2 [B_1, B_2] - [B_1, B_2] B_2 = - \begin{pmatrix} 0 & (d+1)^2 \\ (d+1)^2 & 0 \end{pmatrix}. \end{aligned}$$

In particular, if H_D is totally geodesic then $[B_1, [B_1, B_2]] \in H_D$, i.e., $\text{Tr}(D[B_1, [B_1, B_2]]) = 0$, which gives the condition $d = -1$ or $d = 1$. Conversely, for $d = -1$ or $d = 1$, the triple Lie bracket condition is clearly satisfied, and we obtain a totally geodesic space. This in turn means that either $D = \text{Id}$ or $D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, which is the matrix of the reflection by the x axis. In the first case, H is again the space of matrices of $\text{Symm}(n)$ with zero trace. In the second case, we see that $H_D = \{a\text{Id} + bJ : a, b \in \mathbb{R}\}$ where $J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, from which one gets that $H = \left\{ R^\top \begin{pmatrix} a & b \\ b & a \end{pmatrix} R : a, b \in \mathbb{R} \right\}$.

• **Classification of totally geodesic hypersurfaces:** when H is the subspace of zero trace matrices, $\exp(H)$ is the precisely submanifold $\text{SPD}_1(n)$ of SPD matrices of determinant 1. Then for any $Q \in \text{GL}(n)$, one can easily check that $Q \exp(H) Q = \text{SPD}_r(n)$ where $r = \det(Q)^2$. Based on the above characterization of H and Theorem 7, we deduce that for $n \geq 3$, the totally geodesic hypersurfaces of $\text{SPD}(n)$ are exactly the $\text{SPD}_r(n)$ for $r > 0$.

For $n = 2$, in addition to the $\text{SPD}_r(2)$, we also have the totally geodesic submanifolds $Q \exp(H) Q$ for $Q \in \text{GL}(2)$ and the subspaces $H = \left\{ R^\top \begin{pmatrix} a & b \\ b & a \end{pmatrix} R : a, b \in \mathbb{R} \right\}$ with $R \in O(2)$. Furthermore, for any $a, b \in \mathbb{R}$,

$$\exp(R^\top (a\text{Id} + bJ) R) = R^\top \exp(a\text{Id} + bJ) R = e^a R^\top \begin{pmatrix} \cosh(b) & \sinh(b) \\ \sinh(b) & \cosh(b) \end{pmatrix} R.$$

Moreover, we can see that the set of all matrices $e^a \begin{pmatrix} \cosh(b) & \sinh(b) \\ \sinh(b) & \cosh(b) \end{pmatrix}$ for $a, b \in \mathbb{R}$ is equal to the following cone in $\text{SPD}(2)$:

$$\mathcal{C} = \left\{ \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} : \alpha > |\beta| \right\}.$$

We deduce that the second family of totally geodesic submanifolds of dimension 2 in $\text{SPD}(2)$ are the $Q^\top \mathcal{C} Q$ for $Q \in \text{GL}(2)$, i.e., the translated versions of the cone $\mathcal{C}$. Note that with the particular choice $Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, one has $Q^\top \mathcal{C} Q = \left\{ \begin{pmatrix} \alpha + \beta & 0 \\ 0 & \alpha - \beta \end{pmatrix} : \alpha > |\beta| \right\}$ which is exactly the set of all the diagonal matrices of $\text{SPD}(2)$ and is an obvious example of a totally geodesic submanifold of dimension 2.

• **Projection onto $SPD_r(n)$:** lastly, we prove the expression of the projection $\Pi_r(P)$ of $P \in SPD(n)$ onto the totally geodesic hypersurface $SPD_r(n)$. From Theorem 4.2 in [34], the projection map exists and is continuous on $SPD(n)$. First, we see that $\Pi_r(P) \doteq \sqrt[n]{\frac{r}{\det(P)}}P \in SPD_r(n)$, where we denote $\lambda(P) = \sqrt[n]{\frac{r}{\det(P)}} > 0$. We note that the tangent space to the submanifold $SPD_r(n)$ at $\Pi_r(P)$ is the subspace of symmetric matrices V satisfying $\text{Tr}(\Pi_r(P)^{-1}V) = 0$ which is the same as the set of $V \in \text{Symm}(n)$ such that $\text{Tr}(P^{-1}V) = 0$. Consequently, the orthogonal vectors to that subspace, with respect to the metric $\langle \cdot, \cdot \rangle_{\Pi_r(P)}$, are the sP for $s \in \mathbb{R}$ since for any $V \in T_{\Pi_r(P)}SPD_r(n)$:

$$\begin{aligned} \langle sP, V \rangle_{\Pi_r(P)} &= \text{Tr}(\Pi_r(P)^{-1}(sP)\Pi_r(P)^{-1}V) \\ &= \frac{s}{\lambda(P)^2} \text{Tr}(P^{-1}PP^{-1}V) = \frac{s}{\lambda(P)^2} \text{Tr}(P^{-1}V) = 0. \end{aligned}$$

Now, we have in addition that:

$$\begin{aligned} \log_{\Pi_r(P)} P &= \Pi_r(P)^{1/2} \log(\Pi_r(P)^{-1/2}P\Pi_r(P)^{-1/2})\Pi_r(P)^{1/2} \\ &= \lambda(P)P^{1/2} \log(\lambda(P)^{-1}P^{-1/2}PP^{-1/2})P^{1/2} \\ &= \lambda(P)P^{1/2} \log(\lambda(P)^{-1}\text{Id})P^{1/2} \\ &= -(\lambda(P) \log \lambda(P))P. \end{aligned}$$

from which we deduce that $\log_{\Pi_r(P)} P$ is orthogonal to $T_{\Pi_r(P)}SPD_r(n)$. If Q is any point in $SPD_r(n)$, the geodesic triangle $\Delta Q\Pi_r(P)P$ thus has a right angle at $\Pi_r(P)$. Since $SPD(n)$ is a non-positively curved space, we have the classical inequality:

$$d(P, Q)^2 \geq d(P, \Pi_r(P))^2 + d(\Pi_r(P), Q)^2.$$

Therefore $\Pi_r(P) = \arg \min_{Q \in SPD_r(n)} d(P, Q)^2$ showing that $\Pi_r(P)$ is the projection of P onto $SPD_r(n)$.

APPENDIX E. PROOF OF THEOREM 6

Without loss of generality, we can reduce the proof to the case $r = 1$ and assume that $R_1, \dots, R_m$ are all SPD matrices of unit determinant.

• $SPD_1(n) \subseteq S_{R_1, \dots, R_m}$: for any $P \in SPD_1(n)$, the log vectors $\log_P R_1, \dots, \log_P R_m$ must all belong to the tangent space at P of $SPD_1(n)$ since the latter is totally geodesic. This tangent space, which is the space of all symmetric matrices V satisfying $\text{Tr}(P^{-1}V) = 0$, being of dimension $m - 1$, we deduce that $\log_P R_1, \dots, \log_P R_m$ are linearly dependent and thus that $P \in S_{R_1, \dots, R_m}$.

• $EBS(R_1, \dots, R_m) \subseteq SPD_1(n)$: let $P \in EBS(R_1, \dots, R_m)$ so that there exists $\lambda \in \mathbb{R}^m$ for which $\sum_{i=1}^m \lambda_i = 1$ and $\sum_{i=1}^m \lambda_i \log_P R_i = 0$, and let $d = \det(P)$. For each $i = 1, \dots, m$, we may decompose $\log_P R_i$ into its tangential and normal components to the submanifold $SPD_d(n)$ which writes specifically: $\log_P R_i = \alpha_i P + V_i$ with $\alpha_i \in \mathbb{R}$ and V_i satisfying $\text{Tr}(P^{-1}V_i) = 0$. We then have:

$$\begin{aligned} R_i &= \exp_P(\log_P R_i) = P^{1/2} \exp(P^{-1/2}(\alpha_i P + V_i)P^{-1/2})P^{1/2} \\ &= P^{1/2} \exp(\alpha_i \text{Id} + P^{-1/2}V_i P^{-1/2})P^{1/2} \\ &= e^{\alpha_i} P^{1/2} \exp(P^{-1/2}V_i P^{-1/2})P^{1/2}. \end{aligned}$$

Since $\det(R_i) = 1$, we thus get that

$$1 = e^{n\alpha_i} \det(P) \exp(\text{Tr}(P^{-1/2}V_i P^{-1/2})) = e^{n\alpha_i} \det(P) \exp(\text{Tr}(P^{-1}V_i)) = e^{n\alpha_i} d.$$

We deduce that $\alpha_i = \alpha = -\frac{1}{n} \log(d)$ and so, for all $i = 1, \dots, m$, $\log_P R_i = \alpha P + V_i$. Therefore, $\alpha (\sum_{i=1}^m \lambda_i) P + \sum_{i=1}^m \lambda_i V_i = \alpha P + \sum_{i=1}^m \lambda_i V_i = O$. Since P is orthogonal to $\sum_{i=1}^m \lambda_i V_i$, it follows that necessarily $\alpha = 0$, in other words $d = 1$ and $P \in SPD_1(n)$.

• We prove the last statement of the theorem by contraposition. Assume first that $P \in SPD_1(n) \setminus EBS(R_1, \dots, R_m)$, then one deduces as above that $\log_P R_1, \dots, \log_P R_m$ are linearly dependent i.e. $\sum_{i=1}^m \lambda_i \log_P R_i = 0$, with in addition $\sum_{i=1}^m \lambda_i = 0$. This means that $\{\log_P R_1, \dots, \log_P R_m\}$ is a set of affinely dependent vectors in $T_P SPD_1(n)$ and thus belongs to some affine subspace of dimension at most $m - 2$ in $Symm(n)$ which leads to the conclusion in this case.

On the other hand, if $P \in S_{R_1, \dots, R_m} \setminus SPD_1(n)$, we have $\sum_{i=1}^m \lambda_i \log_P R_i = 0$ for some $\lambda \in \mathbb{R}^m \setminus \{0\}$. Using the same decomposition as previously, we write $\log_P R_i = \alpha_i P + V_i$ with $\text{Tr}(P^{-1} V_i) = 0$. Once again, we can show as above that for all i , $\alpha_i = \alpha = -\log(d)/n$, where $d \neq 1$ is the determinant of P , from which it follows that:

$$(21) \quad \sum_{i=1}^m \lambda_i \log_P R_i = \alpha \left(\sum_{i=1}^m \lambda_i \right) P + \sum_{i=1}^m \lambda_i V_i = 0.$$

Now, since P is orthogonal (with respect to $\langle \cdot, \cdot \rangle_P$) to $\sum_{i=1}^m \lambda_i V_i$ and $\alpha \neq 0$, this implies that $\sum_{i=1}^m \lambda_i = 0$ and $\sum_{i=1}^m \lambda_i V_i = 0$. Let us now introduce $\Pi(P)$ the projection of P onto $SPD_1(n)$, which by Proposition 3 is $\Pi(P) = d^{-1/n} P$. Then we have for each $i = 1, \dots, m$:

$$\begin{aligned} \log_{\Pi(P)} R_i &= \Pi(P)^{1/2} \log(\Pi(P)^{-1/2} R_i \Pi(P)^{-1/2}) \Pi(P)^{1/2} \\ &= d^{-1/n} P^{1/2} \log(d^{1/n} P^{-1/2} R_i P^{-1/2}) P^{1/2} \\ &= d^{-1/n} \log(d^{1/n} \text{Id}) + d^{-1/n} P^{1/2} \log(P^{-1/2} R_i P^{-1/2}) P^{1/2} \\ &= d^{-1/n} \frac{\log(d)}{n} \text{Id} + d^{-1/n} \log_P R_i. \end{aligned}$$

Now using the fact that $\sum_{i=1}^m \lambda_i = 0$ and (21), we get:

$$\sum_{i=1}^m \lambda_i \log_{\Pi(P)} R_i = d^{-1/n} \sum_{i=1}^m \log_P R_i = 0.$$

This implies that $\log_{\Pi(P)} R_1, \dots, \log_{\Pi(P)} R_m$ are affinely dependent in $T_P SPD_1(n)$ and thus, as in the previous case, that the reference points all belong to $\exp_P(H)$ for some affine subspace H of dimension at most $m - 2$.

REFERENCES

- [1] B. Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.
- [2] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.
- [3] M. Arnaudon, C. Dombry, A. Phan, and L. Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122(4):1437–1455, 2012.
- [4] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.
- [5] S. Bancroft. An algebraic solution of the GPS equations. *IEEE Transactions on Aerospace and Electronic Systems*, (1):56–59, 2007.
- [6] N. Bergam, S. Snoeck, and N. Verma. t-SNE Exaggerates Clusters, Provably. *arXiv preprint arXiv:2510.07746*, 2025.
- [7] Silvere Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.
- [8] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985.

- [9] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *Tenth IEEE International Conference on Computer Vision*, volume 2, pages 1597–1604. IEEE, 2005.
- [10] C. Cury, J. A. Glaunes, and O. Colliot. Template estimation for large database: a diffeomorphic iterative centroid method using currents. In *International Conference on Geometric Science of Information*, pages 103–111. Springer, 2013.
- [11] C. Dansereau. Cobre (for machine learning), 2015.
- [12] M. P. Do Carmo and J. Flaherty Francis. *Riemannian geometry*, volume 2. Springer, 1992.
- [13] I. L. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, pages 1102–1123, 2009.
- [14] R. Hekmati, R. Azencott, W. Zhang, Z. D Chu, and M. J. Paldino. Localization of epileptic seizure focus by computerized analysis of fMRI recordings. *Brain Informatics*, 7(1):1–13, 2020.
- [15] S. Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 80. Academic Press, 1979.
- [16] G. E. Hinton and S. Roweis. Stochastic neighbor embedding. *Advances in Neural Information Processing Systems*, 15, 2002.
- [17] J. Ho, G. Cheng, H. Salehian, and B. Vemuri. Recursive Karcher expectation estimators and geometric law of large numbers. In *Artificial Intelligence and Statistics*, pages 325–332. PMLR, 2013.
- [18] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.
- [19] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- [20] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, pages 281–298. University of California press, 1967.
- [21] J. Matousek. *Lectures on discrete geometry*, volume 212. Springer Science & Business Media, 2013.
- [22] A. Mead. Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 41(1):27–39, 1992.
- [23] B. S. Mityagin. The zero set of a real analytic function. *Mathematical Notes*, 107:529–530, 2020.
- [24] X. Pennec. Barycentric Subspace Analysis on Manifolds. *The Annals of Statistics*, 46(6A):2711–2746, 2018.
- [25] X. Pennec. Manifold-valued image processing with SPD matrices. In *Riemannian Geometric Statistics in Medical Image Analysis*, pages 75–134. 2020.
- [26] X. Pennec, S. Sommer, and T. Fletcher. *Riemannian geometric statistics in medical image analysis*. Academic Press, 2019.
- [27] P. Petersen. *Riemannian geometry*, volume 171. Springer, 2006.
- [28] R. Sacksteder. On hypersurfaces with no negative sectional curvatures. *American Journal of Mathematics*, 82(3):609–630, 1960.
- [29] K. Shiga. Hadamard manifolds. *Geometry of Geodesics and Related Topics*, 3:239–282, 1984.
- [30] L. T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, pages 211–223, 1984.
- [31] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Heat kernels and analysis on manifolds, graphs, and metric spaces*, 338:357, 2003.
- [32] C. Tan, H. Zhao, and H. Ding. Intrinsic k-means clustering over homogeneous manifolds. *Pattern Analysis and Applications*, 27(3):107, 2024.
- [33] J. Y. Tou, Y. H. Tay, and P. Y. Lau. Gabor filters as feature images for covariance matrix on texture classification problem. In *International Conference on Neural Information Processing*, pages 745–751. Springer, 2008.
- [34] A. B. Tumpach and G. Laroitonda. Totally geodesic submanifolds in the manifold SPD of symmetric positive-definite real matrices. *Information Geometry*, 7(Suppl 2):913–942, 2024.
- [35] K. You and H.-J. Park. Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity. *NeuroImage*, 225:117464, 2021.