Recursive Fréchet Mean Estimation

Cheng Wang[1] **Carlos J. Soto**[1,2,3]

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, Massachusetts, USA
[2] Humboldt University of Berlin, Berlin, Germany
[3] Zuse Institute Berlin, Berlin, Germany

Abstract

Estimating the mean of manifold-valued data is a central problem in modern statistics, yet it remains challenging due to the lack of a closed-form expression for the Fréchet mean. The gradient descent algorithm is widely used to approximate this quantity across various applications. Although generally effective, it can be computationally intensive for large datasets, as each iteration requires evaluating gradients with respect to the entire dataset. To address these limitations, we propose a tree-based, Recursive Fréchet Mean Estimator (RFME), tailored to data on manifolds. The proposed method leverages a hierarchical aggregation strategy to reduce computational complexity while preserving statistical accuracy. We establish the weak consistency of RFME with respect to the population Fréchet mean and discuss its computational properties. Through simulation studies and real-world applications, we demonstrate that RFME achieves competitive estimation accuracy with substantially improved efficiency. Moreover, as a generalization of the incremental Fréchet mean estimator, RFME also offers enhanced flexibility while maintaining practical advantages.

1 INTRODUCTION

In statistical inference, the center of a dataset is a fundamental measure of data location and is often of interest. The mean is a cornerstone of many statistical methods such as regression, principal component analysis, and convex analysis. In Euclidean space, the sample mean is canonically defined as $\bar{x}_n := \sum_{i=1}^n x_i/n$, a definition that aligns with related concepts such as barycenter, centroid, and expectation in probability theory. This definition, however, assumes a linear structure of the space with well-defined addition and multiplication operations, which do not generally hold for data residing on nonlinear manifolds.

With advancements in measurement techniques, manifold-valued data has become increasingly common in many research fields. Unlike data in a vector space (e.g., Euclidean space, $\mathbb{R}^n$), manifold-valued data typically comes from spaces without an additive structure, making the conventional sample mean ill-defined. Given a space equipped with a metric, the distance between points can be determined. The Fréchet mean, which identifies the mean as the point that minimizes the total squared-distance to all observations, provides an alternative definition. A standard approach for estimating the Fréchet mean is through the gradient descent method, as introduced by Karcher [1977] and Pennec [1999], with subsequent work by Le [2001] and Le [2004] establishing its applicability to complete, simply connected manifolds.

The Fréchet mean is a primary tool for the statistical analysis of manifold-valued data. It is widely used in areas including shape analysis [Bharath et al., 2018], computer vision [Fletcher and Joshi, 2007], and trajectory modeling [Su et al., 2014], where the observations reside in spaces such as the sphere, matrix spaces, and Lie groups. Beyond descriptive statistics, the Fréchet mean also plays a central role in extending classical statistical techniques to manifold settings. In particular, principal geodesic analysis [Fletcher et al., 2004] and k-means clustering [Tan et al., 2024] rely on Fréchet means to characterize dominant variation and to define cluster centroids on Riemannian manifolds.

Given the computational time and memory requirements of the gradient descent method, many studies have proposed incremental mean estimation techniques to address these challenges. In Euclidean space, the mean can be computed in an iterative manner as

$$\bar{x}_n = \frac{n-1}{n}\bar{x}_{n-1} + \frac{1}{n}x_n, \tag{1}$$

which allows each new observation to be incorporated without storing the entire dataset. This method has been general-

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:7103–7120.

ized to manifold-valued data, where the underlying space has a nonlinear structure. Sturm [2003] first introduced an inductive mean estimator for spaces of non-positive curvature (NPC), and proved its existence and consistency. Later, Arnaudon et al. [2012] presented a stochastic version of the algorithm, and Lim and Pálfia [2014a] extended the method as the weighted inductive mean on the cone of positive-definite Hermitian matrices.

The properties of the inductive mean have also been analyzed on various manifolds, such as the hypersphere [Salehian et al., 2015], Symmetric Positive-Definite (SPD) matrices equipped with the affine invariant Riemannian metric [Ho et al., 2013], SPD matrices equipped with the Stein metric [Salehian et al., 2013], tree-like shape spaces [Feragen et al., 2011], Grassmannian manifolds [Chakraborty and Vemuri, 2015], and the Stiefel manifold [Chakraborty and Vemuri, 2019]. The convergence of the inductive mean has been well-established on manifolds, primarily through its stochastic approximation procedure, while the deterministic proofs have also been provided to support practical applications [Lim and Pálfia, 2014b, Holbrook, 2012]. With the inductive mean, a variety of algorithms for learning with manifold-valued data have been developed, including Deep Neural Networks [Chakraborty et al., 2020], Recurrent Neural Networks [Chakraborty et al., 2018], nonlinear regression [Banerjee et al., 2016], and movement primitive learning [Daab et al., 2024].

In this paper, we introduce the Recursive Fréchet Mean Estimator (RFME), a tree-based inductive algorithm to approximate the mean of a dataset. This is akin to constructing a binary tree where the leaves correspond to the data points, intermediate nodes represent the means of their children, and thus the root of the tree provides the mean estimator for the entire dataset. RFME can be viewed as a generalization of the incremental Fréchet Mean Estimator [iFME, Salehian et al., 2015], with the key difference that it computes the mean in pairs at each step rather than following a fixed sequential order. We establish that RFME is a consistent estimator of the Fréchet mean and that it is more computationally efficient than the gradient descent approach. Additionally, our simulations on synthetic data and real-world cases suggest the proposed approach achieves comparable performance to iFME in terms of time consumption and convergence rate, while additionally offering the advantage of parallel computation.

The rest of the paper is organized as follows. Section 2 reviews the preliminaries of Riemannian manifolds, geodesics, and inductive means, which form the foundation of our proposed algorithm. Section 3 introduces our method and provides the theoretical guarantee for the convergence of RFME to the true Fréchet mean. Section 4 and 5 present experiments on both synthetic and real-world datasets to evaluate the performance of the algorithm, and Section 6 concludes the paper with a discussion.

2 BACKGROUND

2.1 RIEMANNIAN MANIFOLDS

Here we present a brief introduction into Riemannian manifolds and the necessary tools for optimization. For a more thorough handling of manifolds and differential geometry we refer to [Do Carmo, 1992] and [Lee, 2018] and for optimization on manifolds we refer to [Absil et al., 2008] and [Boumal, 2023].

Let $\mathcal{M}$ denote a d-dimensional complete, smooth Riemannian manifold. At each point $p \in \mathcal{M}$ we have an associated *tangent space*, $T_p\mathcal{M}$. Each tangent space is a vector space and consists of all elements tangent to the manifold $\mathcal{M}$ at p. The collection of all tangent spaces is referred to as the tangent bundle, $T\mathcal{M} = \{T_p\mathcal{M} | p \in \mathcal{M}\}$. Each tangent space has an associated inner product, $\langle \cdot, \cdot \rangle_p$, and the collection of all such inner products is the Riemannian metric tensor, $\{\langle \cdot, \cdot \rangle_p | p \in \mathcal{M}\}$. This inner product, in turn, allows us to consider geometric properties of the manifold such as angles and lengths. The length $\mathcal{L}(\cdot)$ of a path $\alpha : [0, 1] \to \mathcal{M}$ is $\mathcal{L}(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_{\alpha(t)}^{1/2} \mathrm{d}t$, where $\| \cdot \|_p^2 = \langle \cdot, \cdot \rangle_p$ and $\dot{\alpha}(t) = \frac{d}{dt}\alpha(t)$. The distance between two points p, q on the manifold is the length of the shortest path between them and denoted $d(p, q) := \inf_\alpha \mathcal{L}(\alpha)$. A path of minimal length is referred to as a *geodesic*.

Geodesics are not always unique, e.g. the antipodal point on a sphere, however all such geodesics have the same length. The set of such points, where the uniqueness fails, is known as the cut locus. The infimum distance from p to cut locus is the injectivity radius at p and the infimum of injectivity radius at all points is the injectivity radius of a Riemannian manifold, denoted as $\operatorname{inj} \mathcal{M}$. Given a geodesic α starting at p and initial velocity $\dot{\alpha}(0) = v$, the exponential map at p, denoted as $\exp_p : T_p\mathcal{M} \to \mathcal{M}$, is defined by $\exp_p v = \alpha(1)$. The map is a diffeomorphism in a sufficiently small neighborhood of p, bounded by the cut locus. Within this neighborhood, the inverse of exponential map, called the inverse exponential map or log map, is given by $\log_p q = v$, where $q = \alpha(1)$. The Riemannian distance then can also be given by $d(p, q) = \| \log_p q \|_p = \| \log_q p \|_q$.

Convexity on manifolds has been well established where the notion of straight lines is replaced with the notion of geodesics. A subset of a manifold $A \subset \mathcal{M}$ is said to be geodesically (strong) convex if for all $p, q \in A$ there exists a unique geodesic $\alpha : [0, 1] \to \mathcal{M}$ with $\alpha(0) = p, \alpha(1) = q$ such that $\alpha(t) \in A$ for all t. Given a function $E : \mathcal{M} \to \mathbb{R}$ and a geodesic α whose image is contained in $A \subset \mathcal{M}$, E is said to be geodesically convex in A if A is geodesically convex and $E \circ \alpha$ is convex. Note that $E \circ \alpha : [0, 1] \to \mathbb{R}$, so convexity is in the usual sense that $E(\alpha(t)) \leq (1 - t)E(p) + tE(q)$. Strict convexity is defined analogously by requiring strict convexity of $E \circ \alpha$. Strong convexity

is further specified with the condition $E(\alpha(t)) \leq (1-t)E(p) + tE(q) - \frac{\lambda}{2}t(1-t)\mathcal{L}(\alpha)^2$, which is also called geodesically λ-strong convex.

2.2 FRÉCHET AND KARCHER MEAN

Arguably, the most fundamental statistic of a dataset is the average or mean. On a manifold we do not use the typical definition of the mean $\bar{x} = 1/n \sum x_i$ as there is no guarantee this estimate will lie on the manifold. Further, the addition and multiplication are not intrinsically defined operations in a nonlinear geometric setting. For this reason, such an estimator is referred to as the *extrinsic* mean, when computed in an ambient Euclidean space and interpreted on the manifold. The classical *intrinsic* extension of the Euclidean mean is the Fréchet mean defined as

$$\mu \in \operatorname{argmin}_{x \in \mathcal{M}} \frac{1}{2} \sum_{i=1}^{n} d^2(x, x_i).$$

The Fréchet mean is thus the point on the manifold which minimizes the sum of squared distances from the data (i.e., the variance) [Fréchet, 1948]. This is sometimes referred to as the Karcher mean after Hermann Karcher who did extensive theoretical work on its convergence [Karcher, 1977]. Estimating this mean is typically done using a gradient descent approach. Letting $U(x) = \frac{1}{2}\sum_{i=1}^{n} d^2(x, x_i)$ be the variance energy, the gradient is $\nabla_x U(x) = -\sum_{i=1}^{n} \log_x x_i$ where $\log$ is the log map. We then update the estimate with exponential map as $\hat{x} = \exp_x(-\eta \nabla_x U(x))$ where η is a tuning parameter, the step size, and proceed to the next iteration. This optimization requires computing n many log maps per iteration which can be costly; for instance, on the space of symmetric positive-definite matrices with the affine Riemannian metric, the log map requires matrix inversion which is inherently costly. Note that some refer to the set of local minimizers of the variance function as the *Karcher Means*. We assume we have a unique global optimizer.

2.3 INCREMENTAL FRÉCHET MEAN ESTIMATOR

Computing the empirical Fréchet mean using gradient descent has a time complexity of $O(nk)$, with sample size n and number of iterations k. This can be computationally demanding, as it requires processing the entire dataset during each iteration. To address this issue, particularly in the context of streaming and incoming data, Salehian et al. [2013] proposed the following generalized form of an incremental algorithm:

$$\mu_1 = x_1$$
$$\mu_k = \operatorname{argmin}_{x \in \mathcal{M}}(\omega_k d^2(\mu_{k-1}, x) + (1-\omega_k) d^2(x_k, x)),$$

where μ_{k-1} is the $(k-1)$th mean estimator with weight ω_k. The weight ω_k allows a weighted average to be computed given the importance of the data point x_k. Here μ_k can be interpreted as an estimate of the empirical Fréchet mean of k data points. This estimator is consistent and benefits from not needing to recompute every time new data is introduced. Further, recomputing the Fréchet mean with the gradient descent method requires storing the entire dataset on the server, which imposes a substantial memory burden; the incremental algorithm alleviates this requirement.

In Euclidean space, the mean of two points always lies along the line segment connecting them; the analogous idea on Riemannian manifolds is that the mean of two points always lies along the geodesic between them. Moving along a geodesic from an estimate of the mean in the direction of a new data point is the idea behind the incremental Fréchet Mean Estimator [iFME, Salehian et al., 2015]. It is, thus, useful to characterize a general midpoint of a geodesic. Let α be a geodesic from $p \in \mathcal{M}$ to $q \in \mathcal{M}$, we further denote $\alpha_t(p, q) = \alpha(t)(p, q)$, where $t \in [0, 1]$.

The iFME is an algorithm to efficiently estimate the mean for streaming data and can be formulated for an equally weighted dataset as follows:

$$\hat{\mu}_1 = x_1$$
$$\hat{\mu}_k = \alpha_{\frac{1}{k}}(\hat{\mu}_{k-1}, x_k).$$

This construction extends equation 1 to the setting of Riemannian manifolds. The resulting estimator evolves along geodesics: at iteration k, the current estimator $\hat{\mu}_{k-1}$ is updated by moving a fraction $1/k$ of the geodesic distance toward the new data point x_k. Theoretical guarantees for this inductive procedure have been established in several geometric settings, such as non-positive curvature manifolds [NPC, Sturm, 2003] and on the sphere [Salehian et al., 2015].

3 RECURSIVE MEAN ALGORITHM

Suppose we have a dataset $D \subset \mathcal{M}$ with $D = \{x_1, x_2, \ldots, x_n\}$. We aim to combine elements of the dataset in a binary sort of fashion, thus it makes sense to consider $n = 2^k + 1 + r$ where $0 \leq r < 2^k$. This notation is typical in number theory where r refers to the remainder.

We organize the dataset as the leaves of a binary tree, with each internal node representing the weighted mean of its two children. Then the recursive relationship is defined as

$$\hat{\mu}_{a,b}^{(w_1+w_2)} = \alpha_\tau(\hat{\mu}_{a+1,2b-1}^{(w_1)}, \hat{\mu}_{a+1,2b}^{(w_2)}) \tag{2}$$

where $\tau = \frac{w_2}{w_1+w_2}$ is a weight indicating the step size on the geodesic. The indices may be a bit cumbersome, but the first subscript index refers to the layer from top to bottom, and the second index is the element index of that level. Generally, we have the layer index $a \in \{0, 1, 2, \ldots, k+1\}$, $k+2$ layers in total, and the node index b with $\max(b) \leq 2^a$

for each layer. The superscript is a weight with the leaves having unit weight (unless specified otherwise).

To incorporate all observations, we assume each data with unit weight so that the total weight of the estimate equals n, hence we have our Recursive Fréchet Mean Estimator (RFME) to be $\hat{\mu}_{0,1}^{(n)}$. To organize the data in a hierarchical structure, we first assign the data to the leaf nodes of a binary tree. When n is not a power of two, we perform pairwise merging at the leaf level by iteratively replacing adjacent pairs with their midpoint, while leaving the unpaired elements unchanged, until the number at the level equals 2^k, where k is the largest integer such that $2^k \leq n$. The tree then is extended upward in a bottom-up fashion, where each internal node is given by Equation 2 of its two children. The resulting structure is a complete binary tree at all levels except at the bottom level, which may be partially filled. A detailed description of the algorithm can be found in Algorithm 1.

This organization can be adapted to align with the structure of iFME if we construct the tree with $n+1$ many layers, where the bottom layer is x_1 and each subsequent layer includes one additional data point. As opposed to the iFME, the proposed algorithm can take advantage of parallel programming as, at any layer, each node is independent of every other node.

The top panel of Figure 1 illustrates an example of the recursive estimation process, where nodes belonging to the same layer are indicated by the same color. In this case $n = 2^3$, and we have that every $\tau = \frac{1}{2}$ which greatly simplifies the notation. The bottom panel of Figure 1 visually compares the empirical Fréchet mean estimator (eFM) using gradient descent method, incremental estimator (iFME), and the proposed recursive estimator (RFME). The samples are generated uniformly on the hemisphere of S^2, as discussed in Section 4.1.

Algorithm 1 Recursive Fréchet Mean Estimator

1: Given a set of data points $\{x_i\}_{i=1}^n$ with corresponding weights $\{w_i\}_{i=1}^n$ (default 1).
2: Assign $\hat{\mu}_{k,i}^{(w_i)} = x_i$, where $k = \lceil \log_2 n \rceil$.
3: Compute $\hat{\mu}_{k-1,j}^{(w'_j)} = \alpha_\tau(\hat{\mu}_{k,2j-1}^{(w_{2j-1})}, \hat{\mu}_{k,2j}^{(w_{2j})})$, where $\tau = \frac{w_{2j}}{w_{2j-1}+w_{2j}}$ and $w'_j = w_{2j-1} + w_{2j}$, for $j = 1, ..., n - 2^{k-1}$.
4: Assign $\hat{\mu}_{k-1,j}^{(w_j)} = \hat{\mu}_{k,l}^{(w_l)}$, for $j = n - 2^{k-1} + 1, ..., 2^{k-1}$ and $l = j + (n - 2^{k-1})$. Then set $k = k - 1$.
5: Compute $\hat{\mu}_{k-1,j}^{(w'_j)} = \alpha_\tau(\hat{\mu}_{k,2j-1}^{(w_{2j-1})}, \hat{\mu}_{k,2j}^{(w_{2j})})$, where $\tau = \frac{w_{2j}}{w_{2j-1}+w_{2j}}$ and $w'_j = w_{2j-1} + w_{2j}$, for $j = 1, ..., \frac{M}{2}$, $M = |\{\hat{\mu}_{k,i}^{(w_i)}\}|$.
6: If $k \neq 1$, set $k = k - 1$, and return to step 5.
7: Output $\hat{\mu}_{0,1}^{(\sum_{i=1}^n w_i)}$.

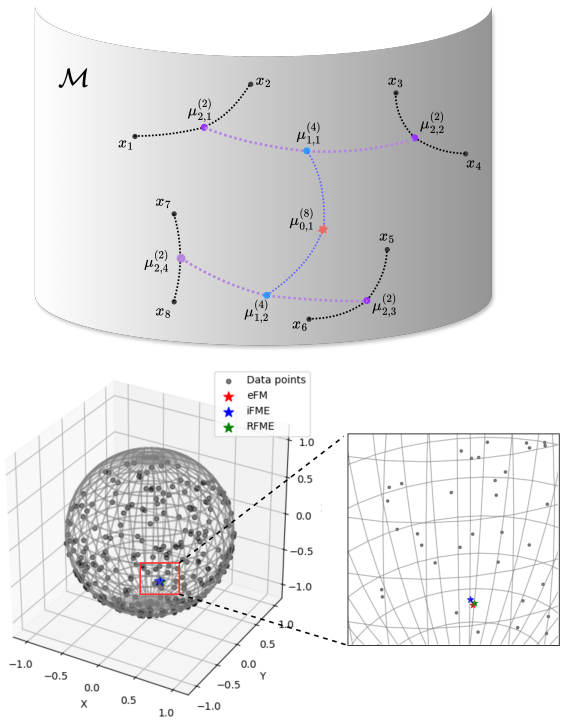

Figure 1: **Top:** Illustration of recursive mean estimation with a sample of size 2^3. **Bottom:** Comparison of the eFM via gradient descent method (red), iFME (blue) and RFME (green), using uniformly distributed data on hemisphere.

A related idea of obtaining the mean through pairwise merging was proposed by Cisneros-Velarde and Bullo [2022], who developed a pairwise distributed algorithm for computing the Wasserstein barycenter. The Wasserstein barycenter can be viewed as a special case of the Fréchet mean of probability measures under the optimal transport geometry. Their work investigates the computation of the barycenter in a distributed multi-agent setting, where each agent only has access to its local distribution and communicates with neighboring agents through a network represented by a (directed) graph. The proposed algorithm relies on the displacement interpolation, which defines the geodesic structure in the Wasserstein space. In contrast, our setting assumes a complete communication setting, where geodesics between any pair of data points are always available. While the Wasserstein space in Cisneros-Velarde and Bullo [2022] is a specific metric space, our framework considers this pairwise merging strategy in Riemannian manifolds, where the geometric structure is induced by the underlying Riemannian metric.

3.1 THEORETICAL GUARANTEES

Here we show that the estimator $\hat{\mu}_{0,1}^{(n)}$ is a consistent estimator of the expectation, $\mathbb{E}(X)$. Since we require a unique

minimizer for the variance energy $U(x)$, as defined in 2.2, we need bounds on the data. Therefore, we make the following assumption.

Assumption 3.1. Let $\mathcal{M}$ be a manifold with sectional curvature κ bounded above by $\kappa \leq \kappa'$, the dataset D lies in a geodesically convex ball, that is, $D \subset B_r(p)$ with $B_r(p)$ being an open ball centered at $p \in \mathcal{M}$ with radius r such that

$$r < \frac{1}{2}\min\{\operatorname{inj}\mathcal{M}, \frac{\pi}{\sqrt{\kappa'}}\}.$$

This condition guarantees the uniqueness of Fréchet mean [Karcher, 1977]. Specifically, for spaces with non-positive sectional curvature $\kappa \leq 0$, the global uniqueness is guaranteed. We consider unbiasedness in two regards. We note that the standard definition of expectation does not apply directly on manifolds. So, we first consider in Theorem 3.2 the pushforward density with the standard expectation and in Theorem 3.4 we consider energy based expectation.

Theorem 3.2 (Unbiasedness)**.** *Let $(\mathcal{M}, d)$ be a measurable space, and let $X : \Omega \to \mathcal{M}$ denote a random variable on $\mathcal{M}$, with a reference probability measure γ, absolutely continuous with respect to the Riemannian volume form (see Appendix A.2). The expectation of X is defined as $\mathbb{E}(X) = \arg\min_{p\in\mathcal{M}} \int_{\mathcal{M}} d^2(p, x)\gamma(dx)$ [Fréchet, 1948]. Let $\{x_i\}_{i\in\mathbb{N}}$ be a set of n i.i.d. samples from the distribution of X, then the recursive mean estimator $\hat{\mu}_{0,1}^{(n)}$ of the samples is an unbiased estimator of $\mathbb{E}(X)$.*

Proof. Let $f : \mathcal{M} \to \mathbb{R}$ be a density with reference measure γ and $x_i \sim f(x)$. There exists a density $h(u)$ defined on $T_p\mathcal{M}$ such that $h \propto \log_{p*} f$, i.e., the pushforward density onto the tangent space. Suppose that $\mu = \mathbb{E}_f(X) = \operatorname{argmin}_{p\in\mathcal{M}} \int_{\mathcal{M}} d^2(p, x)\gamma(dx)$, then we have that the mean satisfies the barycentric equation $\mathbb{E}_f(\log_\mu x) = 0$ [Pennec, 2006]. Thus, we have that if $\mathbb{E}_f(u) = 0$ where $u = \log_p x$, then $p = \mu$ [Chevallier et al., 2022]. Lastly, let $\tilde{x} := \log_p x$, then we have $\mathbb{E}_h(\tilde{x}) = \tilde{\mu}$.

Consider $\alpha : [0, 1] \to \mathcal{M}$ with $\alpha(0) = x_1, \alpha(1) = x_2$. Let $p, q \in \alpha$ and fix p, then on $T_p\mathcal{M}$ we have that $\tilde{q} = (1-t)\log_p x_1 + t\log_p x_2$ for some t since q also lies on the geodesic. So,

$$\begin{aligned}
\mathbb{E}_h(\tilde{q}) &= \mathbb{E}_h((1-t)\tilde{x}_1 + t\tilde{x}_2) \\
&= \mathbb{E}_h((1-t)\log_p x_1 + t\log_p x_2) \\
&= (1-t)\mathbb{E}_h(\log_p x_1) + t\mathbb{E}_h(\log_p x_2) \\
&= (1-t)\tilde{\mu} + t\tilde{\mu} \\
&= \tilde{\mu}
\end{aligned}$$

Letting $p \to q$ we then have that $\mathbb{E}_h(\tilde{q}) = 0$ as this makes the mean the center of the tangent space. Thus, all points on the geodesic between unbiased estimators are unbiased estimators. □

Observe that the Fréchet variance function [Fréchet, 1948] associated with the random variable X on $\mathcal{M}$ is

$$\mathbb{E}(d^2(p, X)) = \int_{\mathcal{M}} d^2(p, X)\gamma(dx),$$

which can be viewed as a stochastic version of the empirical variance energy $U(x) = \frac{1}{2}\sum_{i=1}^{n} d^2(x, x_i)$. The Fréchet variance of X, defined as $var(X) = \min_{p\in\mathcal{M}} \mathbb{E}(d^2(p, X))$, is attained at the Fréchet mean $\mathbb{E}(X)$. Therefore, the variance of X is given by

$$var(X) = \mathbb{E}(d^2(X, \mathbb{E}(X))).$$

An alternative definition of variance can be found in Appendix A.3.

Theorem 3.3 (Consistency)**.** *Let $var(\hat{\mu}_{0,1}^{(n)})$ be the variance of the recursive mean estimator, then we have $var(\hat{\mu}_{0,1}^{(n)}) \to 0$.*

Proof. By the convexity of the half-squared distance function, along the geodesic α, we have

$$\begin{aligned}
d^2(\alpha(t), p) \leq (1-t)d^2(\alpha(0), p) + td^2(\alpha(1), p) \\
-\lambda t(1-t)d^2(\alpha(0), \alpha(1))
\end{aligned}$$

for $p \in \mathcal{M}$ and $t \in [0, 1]$. Here

$$\lambda = \begin{cases} 1 & \kappa' \leq 0 \\ r\sqrt{\kappa'}\cot(r\sqrt{\kappa'}) & \kappa' > 0 \end{cases},$$

where κ' is the upper bound of the sectional curvature κ of $\mathcal{M}$ [Karcher, 1977, Afsari, 2011, Scieur et al., 2026]. See Appendix A.1 for further discussion.

Applying $\alpha(0) = x_1, \alpha(1) = x_2$, and $p = \mathbb{E}(X)$,

$$\begin{aligned}
d^2(\alpha(t), \mathbb{E}(X)) \leq (1-t)d^2(x_1, \mathbb{E}(X)) + td^2(x_2, \mathbb{E}(X)) \\
-\lambda t(1-t)d^2(x_1, x_2).
\end{aligned}$$

Taking the expectation does not change the inequality, thus

$$\begin{aligned}
\mathbb{E}(d^2(\alpha(t),&\mathbb{E}(X))) \leq \\
&(1-t)\mathbb{E}(d^2(x_1, \mathbb{E}(X))) + t\mathbb{E}(d^2(x_2, \mathbb{E}(X))) \\
&- \lambda t(1-t)\mathbb{E}(d^2(x_1, x_2)) \\
&\leq \max\{\mathbb{E}(d^2(x_1, \mathbb{E}(X))), \mathbb{E}(d^2(x_2, \mathbb{E}(X)))\} \\
&- \lambda t(1-t)\mathbb{E}(d^2(x_1, x_2)).
\end{aligned}$$

Therefore, replacing a pair of points with their weighted average strictly decreases the maximum expected squared distance, unless the points coincide. For the bottom layer, we have $\mathbb{E}(x_1) = \mathbb{E}(x_2) = \mathbb{E}(X)$, then $\mathbb{E}(d^2(\alpha(t), \mathbb{E}(X))) \leq var(X) - \lambda t(1-t)\mathbb{E}(d^2(x_1, x_2))$.

Let $l = \lceil \log_2 n \rceil - a$ denote the number of iterations needed to reach layer a, and define

$$V_l = \max_b (\mathbb{E}(d^2(\hat{\mu}_{a,b}^{(\cdot)}, \mathbb{E}(X)))$$

which represents the largest expected squared distance between the estimators of layer a and the true expectation. The inequality above yields $V_l \geq V_{l+1}$, implying that the sequence $\{V_l\}$ is non-increasing and typically strictly decreasing whenever the quadratic term is nonzero. Since $V_l \geq 0$, by the monotone convergence theorem, $V_l \to V_\infty \geq 0$, and we next show that $V_\infty = 0$.

Assume $V_\infty = L > 0$, then in the limit set, there exists at least one estimator $\hat{\nu}_1$ such that

$$\mathbb{E}(d^2(\hat{\nu}_1, \mathbb{E}(X))) = L.$$

By taking the weighted midpoint between $\hat{\nu}_1$ and another distinct estimator $\hat{\nu}_2$, we can further reduce the expected squared distance, which contradicts the assumption that we are at the limit. If no such $\hat{\nu}_2$ exists, then $\hat{\nu}_1$ must be biased, which violates the unbiasedness condition. Therefore, by contradiction, we conclude that $V_\infty = 0$, and consequently, $var(\hat{\mu}_{0,1}^{(n)}) \to 0$. □

From Theorems 3.2 and 3.3, we conclude that RFME is weakly consistent and that the recursive mean converges asymptotically to the expected mean.

Lastly, we demonstrate that choosing the weight $\tau = \frac{w_2}{w_1+w_2}$ for a weighted average is equivalent to selecting the corresponding step size along the geodesic $\alpha_\tau(\hat{\mu}_{a+1,2b-1}^{(w_1)}, \hat{\mu}_{a+1,2b}^{(w_2)})$.

Theorem 3.4. *Given two estimates of μ, μ_k and $\mu_{k'}$, and geodesic $\alpha_t(\mu_k, \mu_{k'})$, the minimizer of the Fréchet variance is attained at $t = \frac{k'}{k+k'}$. Here μ_k ($\mu_{k'}$ resp.) is an unbiased estimator for a sample of size k (k' resp.).*

Proof. See Appendix B. □

4 EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed RFME in comparison with iFME and the empirical Fréchet mean (eFM) estimator via the gradient descent method. As mentioned previously, iFME employs an add-one-at-a-time strategy, while the eFM estimator is derived by minimizing the overall squared distance to all data points. In the following simulations, we utilize code from Geomstats [Miolane et al., 2020].

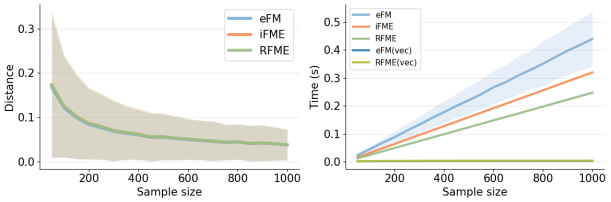

Figure 2: Accuracy and time comparisons of eFM, iFME, and RFME, with samples drawn from spherical uniform distribution. The shaded area denotes the confidence interval based on 2 standard deviations. Results are averaged over 1000 iterations with sample sizes ranging from 50 to 1000.

4.1 UNIT SPHERE

We first employ the uniform spherical sampling method, which, as the name suggests, selects points from the surface of a sphere such that each point has an equal probability of being chosen. To guarantee the uniqueness of the mean, we bound our sample to one hemisphere. The data generation procedure is constructed by sampling an azimuthal angle and a polar angle as

$$\phi \sim \mathcal{U}(0, \pi), \qquad \theta \sim \arccos \mathcal{U}(-1, 1),$$

where $\mathcal{U}$ denotes the uniform distribution. The sampled spherical coordinates are then transformed into Cartesian coordinates through

$$(x, y, z) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta).$$

This setting enables us to generate data with an expected intrinsic mean of $(0, 1, 0)$.

The performance of the eFM, iFME, and RFME is assessed in terms of estimation accuracy and computational efficiency. As illustrated in the left panel of Figure 2, where estimation accuracy is measured by the distance from the expected mean, all three methods show convergence toward the true value as the sample size increases, with progressively narrower confidence intervals. The differences in estimation errors among the three methods are marginal, and the confidence interval of RFME nearly overlaps with those of the other two approaches, indicating comparable statistical performance.

The right panel of Figure 2 compares the computational time of estimation methods across varying sample sizes. Benefiting from the structural properties, both eFM and our proposed method can be further accelerated through some computing techniques such as vectorization, which refers to performing operations on the entire dataset at once instead of processing elements sequentially via explicit loops, as in iFME. The comparison of non-accelerated implementations suggests that the eFM has the lowest efficiency, whereas RFME performs best. Moreover, the vectorized methods are

Table 1: Stability analysis. The average runtime (standard deviation) in milliseconds is reported over 1000 iterations for sample size n=100, 1000, 5000. To mitigate the impact of system-related anomalies, 5% of the most deviant observations are excluded.

Methods	$n = 100$	$n = 1000$	$n = 5000$
eFM	60.1(6.7)	572.6(62.5)	2925(326.9)
iFME	40.3(0.26)	403.9(1.2)	2020.5(8.6)
RFME	31.9(0.26)	319.7(1.1)	1597.3(6.5)
eFM(vec)	2.4(0.28)	3.6(0.40)	8.6(0.96)
RFME(vec)	3.0(0.02)	4.7(0.02)	8.2(0.06)

substantially more efficient than their non-accelerated counterparts. The numerical results of an independent runtime variability study, summarized in Table 1, indicate that vectorized eFM has better performance for small and moderate sample sizes. This can be attributed to the fact that the number of iteration steps eFM requires could vary with sample distribution, leading to greater variability in runtime, while RFME involves a relatively fixed number of steps, resulting in more stable computational cost. For larger sample sizes, however, RFME is expected to outperform the eFM in terms of overall efficiency. A more detailed comparison between vectorized approaches can be found in Appendix C.1.

4.2 SPECIAL ORTHOGONAL GROUP

The special orthogonal group, defined by

$$SO(n) = \{R \in \mathbb{R}^{n\times n} | R^T R = I, \det R = 1\},$$

consists of all orthogonal matrices with unit determinant and represents the group of orientation-preserving linear isometries of the Euclidean space. The group $SO(n)$ is a compact Lie group and also a smooth submanifold of Euclidean space with dimension $\frac{n(n-1)}{2}$. Associated with $SO(n)$ is its Lie algebra

$$\mathfrak{so}(n) = \{R \in \mathbb{R}^{x\times x} | R^T = -R\} = T_I SO(n),$$

which is the set of skew-symmetric matrices and follows from differentiating the defining relation $R^T R = I$ along smooth curves in $SO(n)$. One special example of $SO(n)$ is $SO(3)$, which is a three-dimensional Lie group that characterizes the set of all rotations of three-dimensional Euclidean space about the origin.

We model samples as random variables on $SO(3)$, generated by isotropic Gaussian perturbations around the mean rotation R_m. Specifically, data are drawn by first sampling noise in the Lie algebra

$$\epsilon_j \sim \mathcal{N}(0, \sigma^2 I_3),$$

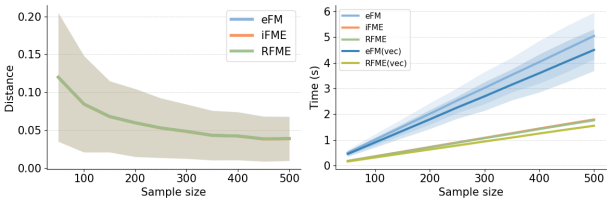

Figure 3: Accuracy and time comparisons of eFM, iFME, and RFME, with samples on $SO(3)$. The shaded area represents the confidence interval based on 2 standard deviations. Results are averaged over 500 iterations with sample sizes ranging from 50 to 500

with $\sigma = 0.5$, and mapping it to the group via the exponential map

$$R_j^\epsilon = \exp(\epsilon_j^\wedge),$$

where $\exp(\cdot)$ denotes the matrix exponential and $(\cdot)^\wedge$ is the skew-symmetric matrix operator (see Appendix C.2). Each observation is then obtained by

$$R_j = R_m R_j^\epsilon.$$

For the simulation, the identity matrix is chosen as the mean rotation, i.e., $R_m = I_3$. This construction induces a locally Gaussian distribution on $SO(3)$, and rotations are finally represented as 3×3 rotation matrices.

The evaluation of estimation accuracy, presented in the left panel of Figure 3, has behavior similar to that observed in the spherical setting. The three methods produce largely overlapping results, demonstrating convergence toward the expected value as the sample size increases, accompanied by progressively shrinking confidence intervals.

The right panel of Figure 3 reports the corresponding runtime comparison. In the matrix-valued settings, the computation of the geodesic, exponential, and log maps is computationally demanding, as it requires repeated matrix inversions and multiplications. Consequently, the eFM incurs substantially higher computational cost, and vectorization yields only limited improvement. In contrast, both iFME and RFME achieve much greater efficiency, with the vectorized RFME having the best overall performance among the methods considered.

4.3 PERMUTATION STUDY

To assess the sensitivity of the three estimators to sample ordering, here we conduct a permutation study of the eFM, iFME and RFME. Figure 4 presents the distance from the resulting estimators to the expected value under different permutations of the same dataset, on the unit sphere S^2 (top panel) and $SO(3)$ (bottom panel).

As shown in the figure, the empirical distributions of estimators are approximately Gaussian and share similar centers across methods. Among the three methods, eFM has the most concentrated distribution, while iFME shows the largest variability. Our proposed method has behavior on SO(3) similar to that of eFM and exhibits greater concentration than iFME in both settings. We note that the default starting point of eFM is the first observation, thus reordering the data will affect the resulting estimate. When the initial point is fixed independently of the sample ordering, the resulting estimate is expected to be invariant to permutations, yielding identical outputs across different orderings of the data.

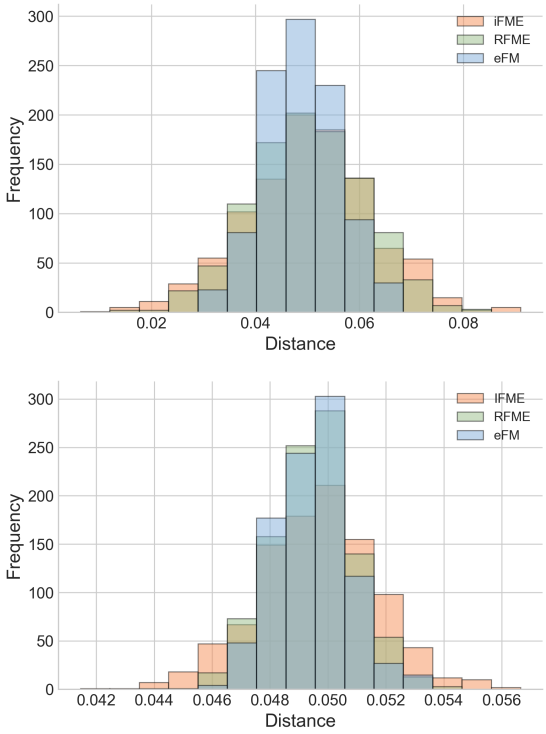

Figure 4: Permutation simulation of eFM, iFME and RFME on unit sphere S^2 (top) and $SO(3)$ (bottom). The x-axis denotes the distance to the expected mean. Sample sizes are set to 512 for S^2 and 128 for $SO(3)$, with 1000 shuffles in each setting.

5 REAL CASE STUDY

In statistical shape analysis, the geometric shape of objects are typically studied independently of their location, size, and orientation. This motivates the idea of defining the group action on the manifold. Given a Lie group G, the group action $G * \mathcal{M} \to \mathcal{M}$ is defined such that it satisfies both the identity and compatibility properties. Under this action, each element of the group corresponds to a smooth transformation of the manifold. For an element $p \in \mathcal{M}$, the orbit of p under a group G, denoted $[p]$, is defined as $[p] = \{g * p : g \in G\}$, and the quotient space $\mathcal{M}/G$ is the collection of all such orbits on the manifold $\mathcal{M}/G = \{[p] : p \in \mathcal{M}\}$.

Using the translation group $\mathbb{R}^n$, scaling group $\mathbb{R}^+$, and rotation group $SO(n)$, Kendall [1984] introduced the Kendall's shape space as a quotient space

$$\mathcal{L}_{n,k}/(\mathbb{R}^n \rtimes (\mathbb{R}^+ \times SO(n)),$$

where $\mathcal{L}_{n,k} = \{X \in \mathbb{R}^{n\times k} | \dim(\text{span}(X)) = n\}$ and $\rtimes$ is the semi-product of groups. In this formulation, the object's geometric shape corresponds to a point in the shape space.

Let $X \in \mathcal{L}_{n,k}$ be an ordered k-tuple representing the contour of an object, equipped with the Euclidean metric. After centering and normalizing, we get a representative $\hat{X}$ of its orbit $[\hat{X}]$ under $SO(n)$, where $\hat{X} = \frac{X - v1_k^T}{\|X - v1_k^T\|}$ and v is the column-wise mean. The orbit lies on a unit hypersphere $S^{n(k-1)-1}$, so that we can apply the spherical metric, e.g., the log map is given by

$$\log_{[\hat{X}1]}([\hat{X}_2]) = \frac{\theta}{\sin(\theta)}(O^*\hat{X}_2 - \cos(\theta)\hat{X}_1),$$

with $\theta = \arccos(\langle \hat{X}_1, O^*\hat{X}_2 \rangle)$, and $O^* = \text{argmax}_{O\in SO(n)} \langle \hat{X}_1, O\hat{X}_2 \rangle$ is the optimal rotation aligning $\hat{X}_2$ to $\hat{X}_1$.

In this section, we use two data sets to evaluate the models. The first one is the glass contour data, extracted from MPEG-7 image dataset [Thoum et al., 2008, Thourn and Kitjaidure, 2009]. The MPEG-7 dataset consists of 70 types of objects with 20 different shapes, and is widely used in the area of computer vision. Each contour in the dataset is represented by an ordered set of 100 landmarks with dimension 2.

The second set contains the corpus callosum shapes from the ADNI (Alzheimer's Disease Neuroimaging Initiative) database [Mueller et al., 2005]. The corpus callosum is a wide, thick bundle of nerve fibers that connects the two hemispheres of the brain. It helps the transmission of neural signals between the two sides and plays a key role in coordinating sensory perception, movement, and cognitive functions of people. The samples are extracted from the magnetic resonance imaging (MRI) scans, and represented by an ordered set of 51 landmarks in a 2D plane.

Figure 5 presents some representative samples from the real datasets, along with the corresponding estimators of Fréchet mean obtained by the three methods considered. Visually, the shaded, dashed, and dotted curves are nearly indistinguishable, with only minor differences in the region of high curvature, e.g., near corners of the shapes.

To quantitatively assess the estimation accuracy, we employ the sum-of-squared error (SSE) criterion, as the Fréchet mean corresponds to the minimum SSE. A table with numerical results is presented in Appendix C.3. For both datasets,

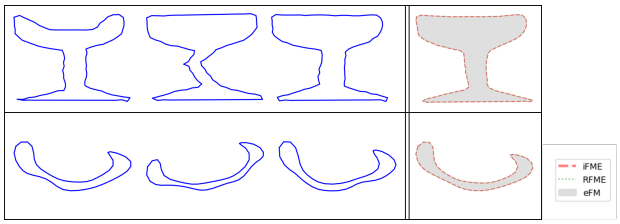

Figure 5: Three glass samples (out of 8) and corpus callosum samples (out of 409), and mean estimations from eFM (shaded), iFME (dashed line) and RFME (dotted line).

the SSE values produced by three methods are nearly identical with differences that are minimal and practically negligible. These results suggest that the three estimators achieve comparable estimation performance. For eFM, the gradient stopping tolerance is set to be 1×10^{-4}; using a smaller threshold could yield more accurate estimation but at the cost of increased computational time. Overall, both visual and numerical results indicate that RFME is an effective and efficient alternative to the traditional gradient-based method in a wide range of practical settings.

6 DISCUSSION

This paper introduces the Recursive Fréchet Mean Estimator (RFME), a tree-based inductive algorithm for approximating the barycenter of manifold-valued data. We establish that the RFME is a consistent estimator of the true Fréchet mean. Empirical evaluations from simulations and real-world case studies demonstrate that RFME achieves higher computational efficiency compared to existing methods.

Conceptually, RFME is a generalization of iFME. While iFME corresponds to a tree with only two nodes in each layer: one is the mean estimator and the other represents a data point, RFME allows for a more flexible structure. Similar to iFME, the proposed method remains sensitive to the ordering of the dataset. In terms of the computational complexity, both iFME and our proposed methods scale linearly with the number of data points, that is, $O(n)$. The Geomstats library has an implementation of iFME, which utilizes the functions for computing the geodesic. While replacing geodesic evaluations with the exponential and log maps could improve the efficiency and bring it closer to that of RFME, the inherently sequential structure prevents further acceleration. In contrast, RFME offers greater potential in both theory development and practical application.

In modern data analysis and machine learning, online learning has become increasingly important, as data often come in as a stream. This arises either because the full dataset is too large to be stored or processed at once, or because the data are generated continuously over time. This calls for adaptive systems that can update incrementally without revisiting the entire dataset. RFME fits into this setting well: it enables the current estimator to be updated with incoming data only, thereby reducing computational cost, memory usage, and storage requirements.

In addition, while this paper focuses on applying RFME to the sphere, $SO(3)$, and Kendall's shape space, which has positive curvature, the estimator itself relies solely on geodesics. As such, it can be readily generalized to a broader class of Riemannian manifolds, such as the Grassmannian manifold, the Stiefel manifold, and Lie groups. Future work may explore extending the method to more general metric or geometric settings, and establishing refined theoretical guarantees, such as non-asymptotic error bounds for finite samples.

References

P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

Bijan Afsari. Riemannian l^{p} center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.

Marc Arnaudon, Clément Dombry, Anthony Phan, and Le Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122(4):1437–1455, 2012.

Monami Banerjee, Rudrasis Chakraborty, Edward Ofori, Michael S. Okun, David E. Vaillancourt, and Baba C. Vemuri. A nonlinear regression technique for manifold valued data with applications to medical image analysis. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4424–4432, 2016. doi: 10.1109/CVPR.2016.479.

Karthik Bharath, Sebastian Kurtek, Arvind Rao, and Veerabhadran Baladandayuthapani. Radiologic image-based statistical shape analysis of brain tumours. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 67 (5):1357–1378, 2018.

Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

Rudrasis Chakraborty and Baba C. Vemuri. Recursive Fréchet mean computation on the grassmannian and its applications to computer vision. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4229–4237, 2015. doi: 10.1109/ICCV.2015.481.

Rudrasis Chakraborty and Baba C. Vemuri. Statistics on the Stiefel manifold: Theory and applications. *The Annals of Statistics*, 47(1):415 – 438, 2019. doi: 10.1214/18-AOS1692.

Rudrasis Chakraborty, Chun-Hao Yang, Xingjian Zhen, Monami Banerjee, Derek Archer, David Vaillancourt, Vikas Singh, and Baba Vemuri. A statistical recurrent model on the manifold of symmetric positive definite matrices. *Advances in neural information processing systems*, 31, 2018.

Rudrasis Chakraborty, Jose Bouza, Jonathan H Manton, and Baba C Vemuri. Manifoldnet: A deep neural network for manifold-valued data with applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (2):799–810, 2020.

Emmanuel Chevallier and Nicolas Guigui. A bi-invariant statistical model parametrized by mean and covariance on rigid motions. *Entropy*, 22(4):432, 2020.

Emmanuel Chevallier, Didong Li, Yulong Lu, and David Dunson. Exponential-wrapped distributions on symmetric spaces. *SIAM Journal on Mathematics of Data Science*, 4(4):1347–1368, 2022.

Pedro Cisneros-Velarde and Francesco Bullo. Distributed Wasserstein barycenters via displacement interpolation. *IEEE Transactions on Control of Network Systems*, 10(2): 785–795, 2022.

Tilman Daab, Noémie Jaquier, Christian Dreher, Andre Meixner, Franziska Krebs, and Tamim Asfour. Incremental learning of full-pose via-point movement primitives on riemannian manifolds. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2317–2323, 2024. doi: 10.1109/ICRA57147.2024.10610275.

Manfredo Perdigao Do Carmo. *Riemannian geometry*, volume 2. Springer, 1992.

Aasa Feragen, Søren Hauberg, Mads Nielsen, and François Lauze. Means in spaces of tree-like shapes. In *2011 International Conference on Computer Vision*, pages 736–746, 2011. doi: 10.1109/ICCV.2011.6126311.

P Thomas Fletcher and Sarang Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.

P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310, 1948.

Jeffrey Ho, Guang Cheng, Hesamoddin Salehian, and Baba Vemuri. Recursive Karcher expectation estimators and geometric law of large numbers. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31, pages 325–332. PMLR, 2013.

John Holbrook. No dice: a deterministic approach to the cartan centroid. *Journal of the Ramanujan Mathematical Society*, 27(4):509–521, 2012.

Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.

Huiling Le. Locating fréchet means with application to shape spaces. *Advances in Applied Probability*, 33(2): 324–338, 2001.

Huiling Le. Estimation of Riemannian barycentres. *LMS Journal of Computation and Mathematics*, 7:193–200, 2004. doi: 10.1112/S1461157000001091.

John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, 2018.

Yongdo Lim and Miklós Pálfia. Weighted inductive means. *Linear Algebra and its Applications*, 453:59–83, 2014a.

Yongdo Lim and Miklós Pálfia. Weighted deterministic walks for the least squares mean on hadamard spaces. *Bulletin of the London Mathematical Society*, 46(3):561–570, 2014b.

Nina Miolane, Nicolas Guigui, Alice Le Brigant, Johan Mathe, Benjamin Hou, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, Hatem Hajri, Yann Cabanes, Thomas Gerald, Paul Chauchat, Christian Shewmake, Daniel Brooks, Bernhard Kainz, Claire Donnat, Susan Holmes, and Xavier Pennec. Geomstats: A python package for Riemannian geometry in machine learning. *Journal of Machine Learning Research*, 21 (223):1–9, 2020.

Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics*, 15(4):869–877, 2005.

Xavier Pennec. Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In *NSIP*, volume 3, pages 194–198, 1999.

Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

Hesamoddin Salehian, Guang Cheng, Baba C. Vemuri, and Jeffrey Ho. Recursive estimation of the stein center of spd matrices and its applications. In *2013 IEEE International Conference on Computer Vision*, pages 1793–1800, 2013. doi: 10.1109/ICCV.2013.225.

Hesamoddin Salehian, Rudrasis Chakraborty, Edward Ofori, David Vaillancourt, and Baba C Vemuri. An efficient recursive estimator of the Fréchet mean on a hypersphere with applications to medical image analysis. *Mathematical Foundations of Computational Anatomy*, 3:143–154, 2015.

Damien Scieur, David Martínez-Rubio, Thomas Kerdreux, Alexandre d'Aspremont, and Sebastian Pokutta. Strongly convex sets in riemannian manifolds. In *The Fourteenth International Conference on Learning Representations*, 2026.

Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive. *Heat kernels and analysis on manifolds, graphs, and metric spaces*, 338:357, 2003.

Jingyong Su, Sebastian Kurtek, Eric Klassen, and Anuj Srivastava. Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *Annals of Applied Statistics*, 8(1):530–552, 2014.

Chao Tan, Huan Zhao, and Han Ding. Intrinsic k-means clustering over homogeneous manifolds. *Pattern Analysis and Applications*, 27(3):107, 2024.

Kosorl Thoum, Yuttana Kitjaidure, and Shozo Kondo. Affine invariant shape recognition based on multi-level of barycenter contour. In *2008 International Symposium on Communications and Information Technologies*, pages 145–149, 2008. doi: 10.1109/ISCIT.2008.4700171.

Kosorl Thourn and Yuttana Kitjaidure. Multi-view shape recognition based on principal component analysis. In *2009 International Conference on Advanced Computer Control*, pages 265–269, 2009. doi: 10.1109/ICACC.2009.69.

Geoffrey S Watson. Distributions on the circle and sphere. *Journal of Applied Probability*, 19(A):265–280, 1982.

Recursive Fréchet Mean Estimation (Supplementary Material)

Cheng Wang[1] **Carlos J. Soto**[1,2,3]

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, Massachusetts, USA
[2] Humboldt University of Berlin, Berlin, Germany
[3] Zuse Institute Berlin, Berlin, Germany

The source code for reproducing the simulation experiments and real-data analyses is available at: `https://github.com/Castiel578/RFME.git`.

A RIEMANNIAN GEOMETRY

A.1 STRONG GEODESIC CONVEXITY

Let $f : \mathcal{M} \to \mathbb{R}$ be a function on a Riemannian manifold. f is strongly convex (or λ-convex) if along every unique geodesic $\alpha : [0, 1] \to \mathcal{M}$:

$$f(\alpha(t)) \leq (1-t)f(\alpha(0)) + tf(\alpha(1)) - \frac{\lambda}{2}t(1-t)d^2(\alpha(0), \alpha(1)),$$

for some $\lambda > 0$. Equivalently,

$$\text{Hess}f(\alpha(t))[\dot{\alpha}(t), \dot{\alpha}(t)] \geq \lambda \|\dot{\alpha}(t)\|^2.$$

We then have the following result

Proposition A.1. *Let $f \in C^2(\mathcal{M})$ be λ-strongly geodesically convex, and $x^* \in U \subset \mathcal{M}$ be a minimizer of f in a geodesically convex open set, then $\forall x \in U$,*

$$\|\nabla f(x)\| \geq \lambda d(x, x^*).$$

Proof. Let α be a geodesic such that $\alpha(0) = x$ and $\alpha(1) = x^*$, we have $\|\dot{\alpha}(t)\| = d(x, x^*)$.

Taking the derivative of the function $f(\alpha(t))$,

$$\frac{d}{dt}f(\alpha(t)) = \langle \nabla f(\alpha(t)), \dot{\alpha}(t) \rangle_{\alpha(t)},$$

then

$$\frac{d^2}{dt^2}f(\alpha(t)) = \frac{d}{dt}\langle \nabla f(\alpha(t)), \dot{\alpha}(t) \rangle_{\alpha(t)} = \text{Hess}f(\alpha(t))[\dot{\alpha}(t), \dot{\alpha}(t)].$$

Since $\|\dot{\alpha}(t)\| = d(x, x^*)$ and $\text{Hess}f \succeq \lambda g$ with metric g,

$$\frac{d^2}{dt^2}f(\alpha(t)) \geq \lambda d^2(x, x^*).$$

Integrating both sides from 0 to 1 with respect to t, we have

$$\frac{d}{dt}f(\alpha(1)) - \frac{d}{dt}f(\alpha(0)) \geq \lambda d^2(x, x^*),$$

which is equivalent to

$$-\langle \nabla f(x), \dot{\alpha}(0) \rangle \geq \lambda d^2(x, x^*).$$

Using Cauchy-Schwarz inequality, we have

$$|\langle \nabla f(x), \dot{\alpha}(0) \rangle| \leq \|\nabla f(x)\| \|\dot{\alpha}(0)\| = \|\nabla f(x)\| d(x, x^*),$$

thus

$$\|\nabla f(x)\| d(x, x^*) \geq \lambda d^2(x, x^*).$$

Since $x \neq x^*$ a.s.,

$$\|\nabla f(x)\| \geq \lambda d(x, x^*).$$

□

Let $f(p) = \frac{1}{2}\mathbb{E}(d^2(p, x))$, and $\{x_i\} \subset B_r(p)$, with $r < \frac{1}{2}\min\{\text{inj}\mathcal{M}, \frac{\pi}{\sqrt{\kappa'}}\}$. For $m = \mathbb{E}(x) = \arg\min f(p)$, we have

$$\|\nabla f(p)\| \geq d(p, m) \cdot \min\{1, r\frac{cs_{\kappa'}(r)}{sn_{\kappa'}(r)}\},$$

see Karcher [1977], Afsari [2011]. Here

$$sn_\kappa(s) = \begin{cases} \frac{1}{\sqrt{\kappa}}\sin(\sqrt{\kappa}s) & \kappa > 0 \\ s & \kappa = 0 \\ \frac{1}{\sqrt{|\kappa|}}\sinh(\sqrt{|\kappa|}s) & \kappa < 0 \end{cases}, \qquad cs_\kappa(s) = \begin{cases} \cos(\sqrt{\kappa}s) & \kappa > 0 \\ 1 & \kappa = 0 \\ \cosh(\sqrt{|\kappa|}s) & \kappa < 0 \end{cases}.$$

Therefore,

$$\|\nabla f(p)\| \geq d(p, m) \cdot \begin{cases} 1 & \kappa' \leq 0 \\ r\sqrt{\kappa'}\cot(r\sqrt{\kappa'}) & \kappa' > 0 \end{cases}.$$

with

$$\lambda = \begin{cases} 1 & \kappa' \leq 0 \\ r\sqrt{\kappa'}\cot(r\sqrt{\kappa'}) & \kappa' > 0 \end{cases}.$$

A.2 RIEMANNIAN VOLUME FORM

Let $\mathcal{M}$ be a d-dimensional Riemannian manifold with metric g. In a local coordinate chart $(x^1, \cdots, x^d)$, the Riemannian volume form is given by

$$dV_g(x) = \sqrt{|\det(g)|}dx^1 \wedge \cdots \wedge dx^d$$

The volume form induces a measure on $\mathcal{M}$. Let γ be a probability measure on $\mathcal{M}$ that is absolutely continuous with respect to $dV_g(x)$, with density f. Then, for any measurable set $A \subseteq \mathcal{M}$, we have

$$\gamma(A) = \int_A \gamma(dx) = \int_A f(x)dV_g(x).$$

A.3 RIEMANNIAN VARIANCE

Let $f : \mathcal{M} \to \mathbb{R}$ be a density and $\mu = \mathbb{E}(X)$, then there exists a density $h(u)$ defined on $T_\mu\mathcal{M}$ and it satisfies

$$f(x)dV_g(x) = h(u)du$$

where du represents the Lebesgue measure on $T_\mu\mathcal{M}$ and $u = \log_\mu x$. Including a volume change term, f can be given by

$$f(x) = \frac{h(u)}{a(u)} \quad \text{where } a(u) = |\det(d\log_\mu x)|$$

see Chevallier and Guigui [2020]. $a(u)$ here measures the distortion of manifold in a neighborhood of μ compared to a flat space. For the second moment of density f, we have

$$\mathbb{E}_f(\|\log_\mu(x)\|^2) = \int_{x\in\mathcal{M}} \|\log_\mu(x)\|^2 f(x)dV_g(x) = \int_{u\in T_\mu\mathcal{M}} \|u\|^2 h(u)du = \mathbb{E}_h(\|u\|^2).$$

It can be checked that $\mathbb{E}_f(\log_\mu x) = \mathbb{E}_h(u) = 0 \in T_\mu\mathcal{M}$, thus if we define the Riemannian variance as $var(X) = \mathbb{E}_f(\|\log_\mu(x)\|^2)$ and $var(U) = \mathbb{E}_h(\|u\|^2)$, we have $var(X) = var(U)$.

B ADDITIONAL PROOF

Proof of Theorem 3.4. Recall that $U(x) = \frac{1}{2n}\sum_{i=1}^n d^2(x, x_i)$ is the sample variance energy which we aim to minimize. We have that $\mu_k = \text{argmin}_{x\in\mathcal{M}} \frac{1}{2k}\sum_{i=1}^k d^2(x, x_i)$ and $\mu_{k'} = \text{argmin}_{x\in\mathcal{M}} \frac{1}{2k'}\sum_{i'=1}^{k'} d^2(x, x_{i'})$. Let $\alpha(t)$ be a geodesic such that $\alpha(0) = \mu_k$ and $\alpha(1) = \mu_{k'}$. We aim to minimize the energy along the geodesic, i.e.,

$$\begin{aligned} U(\alpha(t)) &= \frac{1}{2(k+k')}\sum_{i=1}^{k+k'} d^2(\alpha(t), x_i) \\ &= \frac{1}{2(k+k')}\left[\sum_{i=1}^{k} d^2(\alpha(t), x_i) + \sum_{i'=1}^{k'} d^2(\alpha(t), x_{i'})\right]. \end{aligned}$$

To find the optimal point on the geodesic we consider the gradient. The gradient with respect to $\alpha(t)$, a fixed point on the geodesic, is

$$\nabla_{\alpha(t)}U(\alpha(t)) = -\frac{1}{(k+k')}\left[\sum_{i=1}^{k} \log_{\alpha(t)} x_i + \sum_{i'=1}^{k'} \log_{\alpha(t)} x_{i'}\right].$$

Since μ_k is the minimizer of the first summand and $\mu_{k'}$ is the minimizer of the second summand we thus have $-\frac{1}{(k+k')}\left[k\log_{\alpha(t)}\mu_k + k'\log_{\alpha(t)}\mu_{k'}\right]$. Evaluating this at $t = 0$ and noting $\log_{\alpha(0)}\mu_k = \log_{\mu_k}\mu_k = 0$ we get $-\frac{k'}{(k+k')}\log_{\mu_k}\mu_{k'}$ as desired. □

C ADDITIONAL RESULTS

C.1 SPHERE

Uniform Distribution:

Proposition C.1. *For the setting $\phi \sim \mathcal{U}(0, \pi)$ and $\theta \sim \arccos\mathcal{U}(-1, 1)$, the data generated by*

$$X = (x, y, z) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$$

has the intrinsic expectation $\mathbb{E}(X) = (0, 1, 0)$.

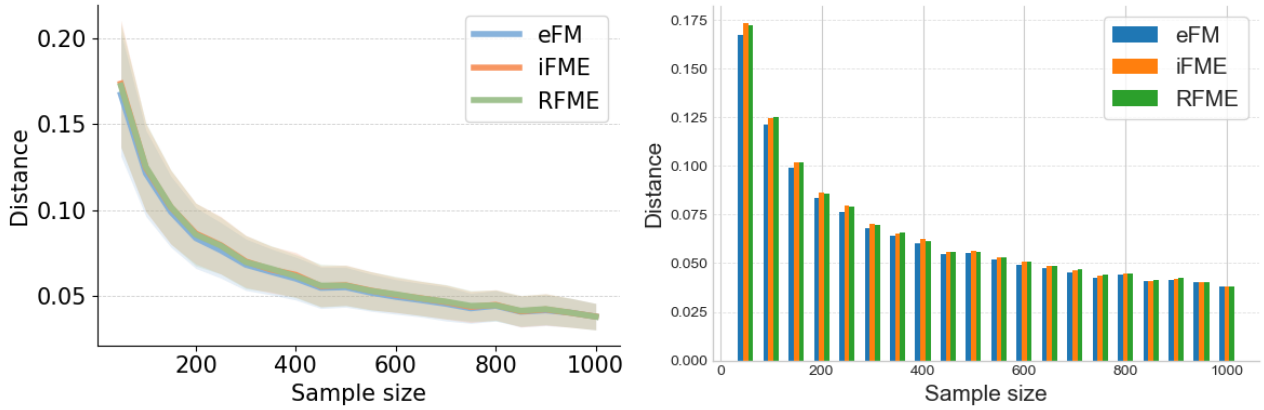

Figure 6: Accuracy comparisons of eFM, iFME, and RFME, on samples drawn from spherical uniform distribution. The shaded area denotes the confidence interval based on 2 standard errors. Results are averaged over 1000 iterations with sample sizes ranging from 50 to 1000.

Proof. Let

$$\mu = \arg\min_{p \in S^2} \mathbb{E}[d^2(p, X)]$$

with geodesic distance on S^2 as $d(p, q) = \arccos(\langle p, q \rangle)$.

Consider $\phi \sim \mathcal{U}(0, \pi)$ and $\cos\theta \sim \mathcal{U}(-1, 1)$, then we have $y = \sin\theta \sin\phi \geq 0$, and X is supported on the hemisphere $\{(x, y, z) \in S^2 : y \geq 0\}$. The distribution is symmetric with respect to y-axis, since the transformation $x \mapsto -x$ and $z \mapsto -z$ leave the distribution invariant. Therefore, the energy function

$$U(p) = \mathbb{E}[d^2(p, X)]$$

is invariant under the reflection with respect to the y-axis, and its minimizer must lie on the y-axis, i.e., $(0, \pm 1, 0)$.

It remains to determine the sign. For $\mu_1 = (0, -1, 0)$, we have

$$d(\mu_1, X) = \arccos(\langle \mu_1, X \rangle) = \arccos(-\sin\theta \sin\phi) \geq \frac{\pi}{2}.$$

Similarly, for $\mu_2 = (0, 1, 0)$, we have $d(\mu_2, X) \leq \frac{\pi}{2}$.

Since $\sin\theta \sin\phi > 0$ with positive probability, we have the strict inequality $\mathbb{E}(d^2(\mu_1, X)) > \mathbb{E}(d^2(\mu_2, X))$, thus the intrinsic Fréchet mean is $\mathbb{E}(X) = (0, 1, 0)$. □

The left panel of Figure 6 is a replicate of the left plot of Figure 2, with the confidence intervals constructed by 2 standard errors, whereas the right panel presents the same comparison as a bar plot.

von Mises-Fisher Distribution: An alternative evaluation framework is conducted to assess the performance of the proposed method under different levels of data dispersion. The von Mises-Fisher distribution, a Gaussian-like probability distribution on the $(p-1)$-sphere S^{p-1} in $\mathbb{R}^p$, has the density function

$$f_p(x; \mu, \varkappa) = C_p(\varkappa) \exp(\varkappa \mu^T x)$$

where $C_p(\varkappa)$ is the normalizing constant. For $p = 3$, this constant is given by $C_3(\varkappa) = \frac{\varkappa}{4\pi \sinh(\varkappa)}$ [Watson, 1982]. Here $\varkappa$ is the concentration parameter, with larger values indicating stronger concentration of data around μ, the mean direction. We fix the sample size at 1000 and repeat the experiment 1000 times, applying methods with various values of $\varkappa$.

As suggested in Figure 7, the vectorized approach has greater efficiency. The runtime of eFM increases as the data becomes less concentrated, while both iFME and RFME remain essentially the same, since the computational cost only depends on the sample size. Additionally, the von Mises-Fisher distribution does not constrain the samples to lie within a single hemisphere; it is expected to see the estimation accuracy of methods degrades when the data dispersion increases.

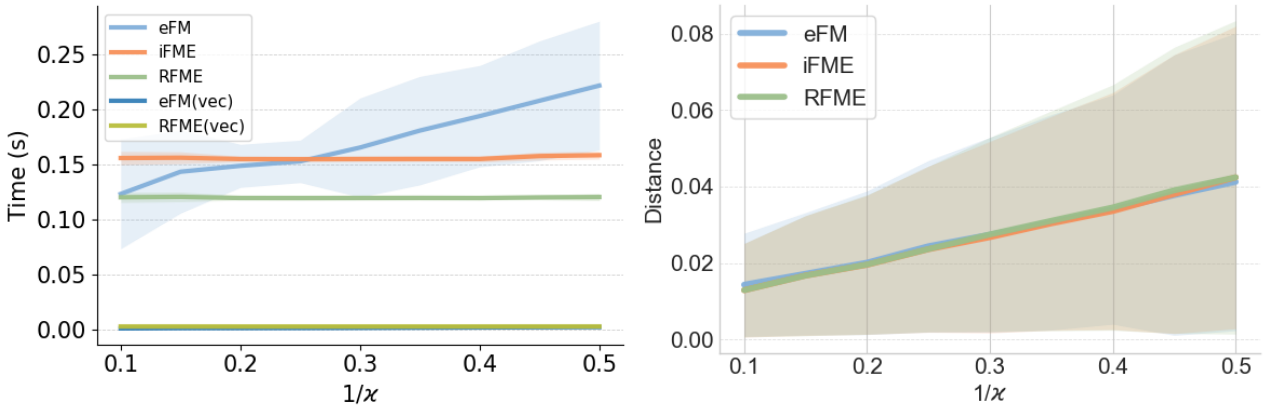

Figure 7: Time and accuracy comparisons of eFM, iFME, and RFME. The x-axis represents $1/\varkappa$, where the rightmost value corresponds to the most deviated dataset.

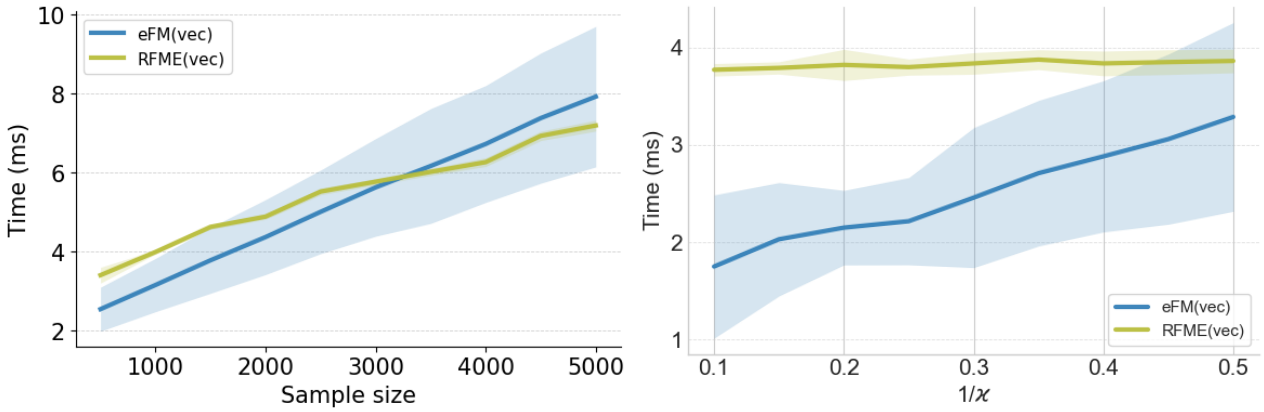

Figure 8: Time comparisons of vectorized eFM and RFME, with samples from uniform distribution (left) and Fisher distribution (right).

Figure 8 compares the runtime of vectorized eFM and RFME on samples drawn from uniform distribution (left) and von Mises-Fisher distribution (right). It is suggested that eFM is more efficient than RFME in most cases. This is expected, as the looping operations dominate the runtime in simple cases, such as the sphere, which also results in higher variability in computation time. However, when the data size is large or widely distributed, RFME becomes more efficient. Furthermore, for matrix manifolds like $SO(n)$, methods like iFME and RFME outperform eFM.

C.2 SPECIAL ORTHOGONAL GROUP

We here provide some details on generating Gaussian distributed samples on $SO(3)$. Specifically, the samples are generated as

$$R_j = R_m R_j^{\epsilon}$$

where we set the mean rotation $R_m = I_3$ and $R_j^{\epsilon} = \exp(\epsilon_j^{\wedge})$. The perturbation vector is set as $\epsilon_j \sim \mathcal{N}(0, 0.5^2 I_3) \subset \mathbb{R}^3$ and the skew-symmetric matrix operator $(\cdot)^{\wedge}$, defined as

$$\epsilon = \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \quad \longleftrightarrow \quad \epsilon^{\wedge} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix},$$

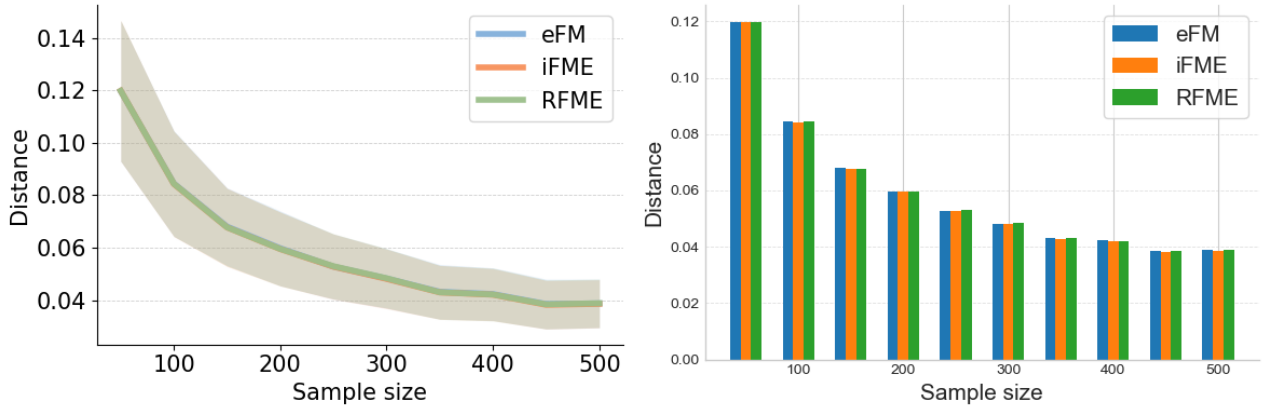

Figure 9: Accuracy comparisons of eFM, iFME, and RFME, on samples from $SO(3)$. The shaded area denotes the confidence interval based on 2 standard errors. Results are averaged over 1000 iterations with sample sizes ranging from 50 to 1000.

Table 2: Total error of eFM, iFME and RFME on datasets.

Dataset	eFM	iFME	RFME
glass	0.529	0.529	0.529
corpus callosum	24.260	24.259	24.258
hammer	1.202	1.202	1.202
hand	1.805	1.805	1.805

turn vectors from $\mathbb{R}^3$ into the Lie algebra $\mathfrak{so}(3)$. The matrix exponential is defined through its power series expansion as

$$\exp(\epsilon^\wedge) = \sum_{i=0}^{\infty} \frac{(\epsilon^\wedge)^k}{k!}.$$

For practical implementation, the exponential map admits a closed form expression given by Rodrigues' formula

$$\exp(\epsilon^\wedge) = I + \frac{\sin(\theta)}{\theta}\epsilon^\wedge + \frac{1-\cos(\theta)}{\theta^2}(\epsilon^\wedge)^2,$$

where $\theta = \|\epsilon\|$.

The left panel of Figure 9 is a replicate of the left plot of Figure 3, with the confidence intervals constructed by 2 standard errors, whereas the right panel presents the same comparison as a bar plot.

C.3 KENDALL'S SHAPE SPACE

Table 2 and Table 3 present the total error (TE) and sum-of-squared error (SSE) of estimators obtained from eFM, iFME and RFME on the MPEG-7 image dataset and ADNI database. The results indicate that the values are nearly identical across all three approaches, suggesting that the corresponding estimators show comparable performance and are statistically equivalent.

Figure 10 illustrates representative samples from the hammer and hand categories of the MPEG-7 dataset along with the estimators produced by each method. The visual comparison further supports the quantitative findings, demonstrating a high degree of similarity among the estimators.

Table 3: Sum-of-squared error of eFM, iFME and RFME on datasets.

Dataset	eFM	iFME	RFME
glass	0.041	0.041	0.041
corpus callosum	1.621	1.621	1.621
hammer	0.178	0.178	0.178
hand	0.411	0.411	0.411

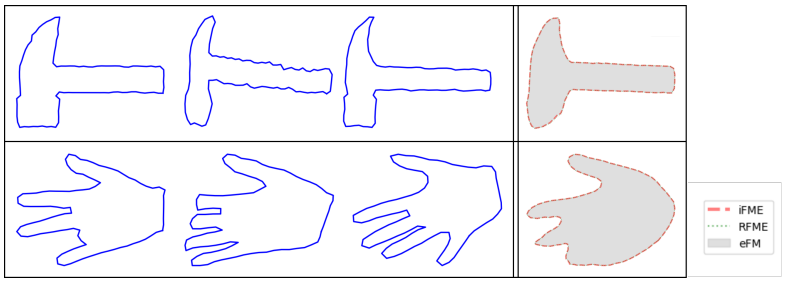

Figure 10: Hand samples and hammer samples, and mean estimations from eFM (shaded), iFME (dashed line) and RFME (dotted line).