

Robust Fréchet Mean and PGA on Riemannian Manifolds with Applications to Neuroimaging

Monami Banerjee¹, Bing Jian² and Baba C. Vemuri^{1*}

¹ University of Florida, Gainesville, USA

² Google, Mountain View, CA, USA.

Abstract. In this paper, we present novel algorithms to compute robust statistics from manifold-valued data. Specifically, we present algorithms for estimating the robust Fréchet Mean (FM) and performing a robust exact-principal geodesic analysis (ePGA) for data lying on *known* Riemannian manifolds. We formulate the minimization problems involved in both these problems using the minimum distance estimator called the L_2E . This leads to a nonlinear optimization which is solved efficiently using a Riemannian accelerated gradient descent technique. We present competitive performance results of our algorithms applied to synthetic data with outliers, the corpus callosum shapes extracted from OASIS MRI database, and diffusion MRI scans from movement disorder patients respectively.

1 Introduction

In this age of data deluge, manifold-valued features are widespread in Science and Engineering disciplines. As the amount of data to be processed grows by leaps and bounds, there is an obvious need to reduce the dimensionality of the data and provide some sort of statistics. To this end, Principal Component Analysis (PCA) has been employed as the main workhorse for data lying in vector spaces. However, when the input data reside on a smooth manifold, the nonlinear generalization called, Principal Geodesic Analysis (PGA) [7] is often employed. PGA in [7] makes use of the concept of linearization by first finding the intrinsic mean of the data lying on the smooth manifold and then makes use of the inverse Riemannian exponential (Exp) map, also called the Riemannian Log map, to map the data from the manifold to the tangent space anchored at the intrinsic mean of the data on the manifold. Then it performs PCA of this Log mapped data and projects the principal vectors back on to the manifold using the Exp map, obtaining the principal geodesic submanifolds. In order for this algorithm to work, it is assumed that the Riemannian Exp and Log maps exist in the desired neighborhood and can be computed efficiently. Although this linearized version of PGA is computationally efficient, it lacks in accuracy of the computed principal components when the input manifold-valued data have a large variance. In order to cope with this issue, Sommer et al. [26] proposed to solve the problem without

* This research was funded in part by, the NSF grant IIS-1525431 to BCV and the UFII fellowship to MB.

resorting to the aforementioned linearization. They called their algorithm, *exact-PGA*. In *exact-PGA* [26], the cost function being optimized involves minimization of the projection distance, also called the reconstruction error. This is a hard nonlinear optimization problem and a general efficient solution is lacking to date. Recently, Chakraborty et al. [3], proposed an efficient solution to the exact-PGA problem for constant curvature manifolds with several applications. There are several variants of the PGA algorithm in literature and we briefly discuss a few here. Authors in [23] presented a technique to compute the principal geodesics (without approximation) but only for the special Lie group, $SO(3)$. Geodesic PCA (GPCA) [15] solves a different optimization function namely, optimizing the projection error along the geodesics. GPCA does not use a linear approximation, but is restricted to manifolds where a closed form expression for the geodesics exists. More recently, a probabilistic version of PGA called PPGA was presented in [28], which is a nonlinear version of PPCA [27]. As an alternative to PGA, Hauberg [11] introduced a Riemannian version of the well known principal curves algorithm restricted to complete Riemannian manifolds. However, none of these methods are robust to outliers in the data.

In this paper, we present a statistically robust formulation for estimating the Fréchet mean (FM) [9] as well as for computing the *exact-PGA*. Our notion of robustness here implies relative “insensitivity (stickiness)” of the location of the estimated FM to outliers in the data. Further, our work here is restricted to smooth Riemannian manifolds and does not consider stratified spaces. For “sticky” and “non-sticky” FM of data in stratified spaces, we refer the reader to [13]. Our formulation makes use of the well known M-estimator in statistics called L_2E introduced by Scott [24]. In finding the FM from data, one minimizes the so called sum of squared geodesic distances between the unknown FM and the given data samples. Whereas, for estimating the principal components in the *exact-PGA* algorithm, one minimizes the sum of squared projection errors, defined by the geodesic distances between the data points and the geodesics emanating from the FM as a function of direction. In [8], authors proposed a geometric median formulation on Riemannian manifolds and have shown robustness of the median. *Though mean and median are distinct statistics, for the sake of completeness, we will compare our robust mean formulation with the intrinsic median.*

The robust formulation for estimating the FM and computing the PGA involves casting the aforementioned geodesic distance minimization costs in both problems into the L_2E based M-estimator framework. This is precisely what is achieved here in this paper. Note that there are no robust PGA methods in existing literature and the theory and algorithm presented here are the first to the best of our knowledge. Hence, we compare our work to one other method of achieving robust PGA, which is our own modification of the conventional non-robust PGA. This modification involves replacing the PCA in the tangent space at the FM performed in the conventional PGA [7] by an existing robust PCA algorithm [12].

The rest of this paper is organized as follows: Section 2 contains the theoretical formulation of the robust FM and PGA respectively. In Section 3, we present

several examples on synthetic data, shapes extracted from OASIS MRI database, and diffusion MRI scans from movement disorder patients respectively. In Section 4, we draw conclusions.

2 Robust Statistics

In this section, first we present a novel formulation of, the robust FM followed by the robust PGA. In both these formulations, we will use the well known and statistically robust minimum distance estimator called the integral squared error (a.k.a. the L_2 error) denoted by L_2E [24].

2.1 Robust FM on a Riemannian manifold: Using the L_2E

Computing the FM is a commonly encountered task in many application problems including but not limited to shape analysis [16], directional statistics [19], diffusion tensor analysis [22,20,17] and many others. In the following we will first define the FM and then present a formulation for computing it robustly from data corrupted with outliers.

Let (\mathcal{M}^m, g) be a Riemannian manifold equipped with a Riemannian metric g [6], where $m = \dim(\mathcal{M})$. Given a point $p \in \mathcal{M}$ and the tangent vector $\mathbf{v} \in T_p\mathcal{M}$, there exists a unique geodesic α such that $\alpha(0) = p$ and $\alpha' = \mathbf{v}$. In general, its existence however is only guaranteed locally. The Riemannian exponential map Exp at the point $p \in \mathcal{M}$ is a locally diffeomorphic map on to the neighborhood of p and is defined as $\text{Exp}_p(\mathbf{v}) = \alpha(1)$, where, α is defined over $[0, 1]$. If $B(p)$ denotes the largest such neighborhood, then, the inverse of the Exp map, called the Riemannian Log map is well defined in this neighborhood. Let $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbf{R}$ be the distance induced by the metric g on \mathcal{M} . Then, for $p, q \in \mathcal{M}$, $d(p, q) = \|\text{Log}_p(q)\|_p$. Given the data $X = \{x_i\}_{i=1}^N \subset \mathcal{M}$, the FM μ is defined by the following minimization [9], $\mu = \text{argmin}_{x \in \mathcal{M}} \sum_{i=1}^N d^2(x, x_i)$. The existence and uniqueness of the FM in general is only guaranteed within a geodesic ball of a certain radius [1,21]. As is usually the case in literature [7,26,3], we will also assume that the input data lie within this geodesic ball, so that FM exists and is unique. Our assumption about the data including the outliers lying inside a geodesic ball of an appropriate radius is in line with earlier work reported in [8].

We now propose a robust formulation to compute FM on a Riemannian manifold based on the well known robust L_2E estimator [24]. It is easy to show that minimizing the sum-of-squared function defined in the FM computation is equivalent to maximizing the likelihood of the distances (of sample points from the FM) being randomly drawn from a one-dimensional half-normal distribution. It is well known that maximum likelihood estimation (MLE) is well suited for estimation problems in which the model is a good descriptor for the data. However, it well known that the ML estimates are highly biased if the data contain outliers [2]. In [2], Basu et al. defined a single parameter family of divergences between distributions, termed the *density power divergence*. This family of divergences

is controlled by a single parameter and includes the KL (Kullback-Leibler) divergence and the L_2 distance as limiting cases. Recognizing that the minimum density power divergence estimators can be interpreted as a particular case of M-estimators [14], Basu et al. [2] have shown that the L_2 estimator (L_2E) corresponding to L_2 distance is superior to MLE in terms of robustness. Scott [24] also exploited the applicability of L_2E to parametric modeling and demonstrated its robustness behavior and nice properties of practical importance.

In the parametric case, given the random variable $\epsilon \in \mathbf{R}$ with unknown density $g(\epsilon)$, and a model $f(\epsilon|\boldsymbol{\theta})$, with a parameter vector $\boldsymbol{\theta}$, we can write the L_2E criterion as, $L_2(f, g) = \int [f(\epsilon|\boldsymbol{\theta}) - g(\epsilon)]^2 d\epsilon$. Note that, in the expansion of this integral, the term $\int g(\epsilon)^2 d\epsilon$ does not depend on $\boldsymbol{\theta}$ and $\int f(\epsilon|\boldsymbol{\theta}) g(\epsilon) d\epsilon = E_g[f(\epsilon|\boldsymbol{\theta})]$ is the so called expected height of the density which can be approximated by the estimator $\frac{1}{N} \sum_{i=1}^N f(\epsilon_i|\boldsymbol{\theta})$. Hence, the proposed estimator minimizing the L_2 distance will be,

$$\hat{\boldsymbol{\theta}}_{L_2E} = \arg \min_{\boldsymbol{\theta}} \left(\int f(\epsilon|\boldsymbol{\theta})^2 d\epsilon - \frac{2}{N} \sum_{i=1}^N f(\epsilon_i|\boldsymbol{\theta}) \right). \quad (1)$$

In the case of model being a Gaussian or mixture of Gaussians, we have a closed form expression for the integral in the bracketed quantity in (1) and hence can avoid numerical integration which severely limits the practical applications not only in computation time but also in accuracy.

Noting that the L_2E criterion does not require that the model $f(\epsilon|\mu)$ be a density, Scott [25] suggested a method for outlier detection and clustering by partial mixture modeling. A partial mixture model basically advocates the use of a small number of components in a ‘‘full’’ N -component Gaussian mixture model, thereby under fitting the outlier-corrupted data. One of the advantages of this model is that the weight w provides us the fraction of the data to which the component has been fitted. For example, using just a single component, this model will account for the largest fraction (cluster) of the data and thus accounting for the inliers (assuming that the outliers correspond to a smaller fraction). This is appropriate in our work because, our focus is on finding the FM (and later, PGA) of a single cluster of data possibly corrupted with outliers.

In our context, $\epsilon = d(\mu, x) \in [0, \infty)$. Note that, $\epsilon \in \mathbf{R}$ depends on the FM, $\mu \in \mathcal{M}$. Inspired by the ideas describe above, to model the density of the geodesic distances from the FM to the data, we chose a partial mixture of half-normal densities with mean $\sqrt{2}\sigma/\sqrt{\pi}$, and variance, $\sigma^2(1 - 2/\pi)$ (here σ is unknown parameter of the half-normal density), i.e., $f(\epsilon|\mu) = w \phi(\epsilon|\mu|\sigma^2)$. For simplicity of notation, we will drop the μ from $\epsilon(\mu)$ in the rest of the paper. Here, $\phi(\epsilon|\sigma^2)$ is the half-normal density and w denotes the weight of the partial mixture.

Lemma 1. *If the model density $f(\epsilon|\mu) = w \phi(\epsilon|\sigma^2)$, then the L_2E criterion is given by,*

$$L_2E(\mu, w, \sigma^2) = \frac{w^2}{\sqrt{\pi}\sigma^2} - \frac{2\sqrt{2}w}{\sqrt{\pi}N\sigma} \sum_{i=1}^N \exp \left\{ -\frac{d^2(\mu, x_i)}{2\sigma^2} \right\}. \quad (2)$$

Proof.

$$\begin{aligned} \int f(\epsilon|\mu)^2 d\epsilon &= \int w^2 \phi^2(\epsilon|\sigma^2) d\epsilon = w^2 \int \left[\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \right]^2 d\epsilon \\ &= w^2 \int \left[\exp\left(-\frac{\epsilon^2}{\sigma^2}\right) \right] d\epsilon = \frac{w^2}{\sqrt{\pi}\sigma^2}. \end{aligned} \quad (3)$$

Substituting $\epsilon_i = d(\mu, x_i)$, we get $f(\epsilon_i|\mu) = \frac{\sqrt{2}w}{\sqrt{\pi}\sigma} \exp\left\{-\frac{d^2(\mu, x_i)}{2\sigma^2}\right\}$. Now, using Eq. 1, we get the desired result in Eq. 2. ■

Estimation of the parameters is then achieved by minimizing this L_2E criterion with respect to the parameters. We derived an analytic expression for gradient of the L_2E cost (not shown here due to lack of space) and employed it in a variant of the accelerated gradient descent (AGD) [10] adopted to Riemannian manifolds. This leads to the optimal set of parameters, $\hat{\mu}_{L_2E} \in \mathcal{M}$, $\hat{w} > 0$, $\hat{\sigma}^2 > 0$. In addition to the robust FM estimate, $\hat{\mu}_{L_2E}$, the partial weight \hat{w} also indicates the fraction of the data being treated as outliers. In the following, we state and prove the robustness of the FM estimator formulated above.

Theorem 1. *The L_2E formulation to compute FM in Eq. 2 is robust to outliers.*

Proof. Let x_j be an outlier, for some j , i.e., $d(\mu, x_j)$ is very large. The *influence function* [14] of $L_2E(\mu, w, \sigma^2)$ is proportional to $\frac{\partial L_2E(\mu, w, \sigma^2)}{\partial d(\mu, x_j)}$. If we can show that as $d(\mu, x_j) \rightarrow \infty$, $\frac{\partial L_2E(\mu, w, \sigma^2)}{\partial d(\mu, x_j)} \rightarrow 0$, we can then claim that our formulation to compute FM, i.e., the L_2E criterion $L_2E(\mu, w, \sigma^2)$ is robust. Now, $\frac{\partial L_2E(\mu, w, \sigma^2)}{\partial d(\mu, x_j)} = \frac{-4\sqrt{2}w}{\sqrt{\pi}N\sigma} \exp\left\{-\frac{d^2(\mu, x_j)}{2\sigma^2}\right\} \text{Log}_{x_j} \mu$. So, in the limit as $d(\mu, x_j) \rightarrow \infty$, $\frac{\partial L_2E(\mu, w, \sigma^2)}{\partial d(\mu, x_j)} \rightarrow 0$, i.e., our formulation is robust. ■

2.2 Principal Geodesic Analysis (PGA)

The goal of PGA is to find a set of $r < m$ orthogonal basis vectors of $T_\mu\mathcal{M}$, called principal vectors $\{\mathbf{v}_j\}_{j=1}^r$, such that the data variance along the geodesic submanifold spanned by these principal vectors is maximized [3,7]. An alternative definition of PGA [26] involves minimizing the reconstruction error, $\sum d^2(x_i, \hat{x}_i)$, where \hat{x}_i is the i^{th} reconstructed data point in the principal submanifold spanned by the basis vectors $\{\mathbf{v}_j\}_{j=1}^r$. These two formulations result in the same solution in \mathbf{R}^n but not so on a general Riemannian manifold. In [3], the principal vectors $\{\mathbf{v}_j\}$ are defined recursively by,

$$\mathbf{v}_j = \underset{\|\mathbf{v}\|=1, \mathbf{v} \in V_{j-1}^\perp}{\text{argmin}} \frac{1}{N} \sum_{i=1}^N d^2(x_i, \Pi_{S_j}(x_i)), \quad S_j = \text{Exp}_\mu(\text{span}\{V_{j-1}, \mathbf{v}_j\}), \quad (4)$$

where, $V_{j-1} = \{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$. S_j is the submanifold spanned by $V_j = \{V_{j-1}, \mathbf{v}_j\}$, and $\Pi_{S_j}(x)$ is the point in S_j closest to $x \in \mathcal{M}$. In this paper, we will use this alternative formulation to define a robust formulation of PGA on a Riemannian manifold.

2.3 Robust PGA on a Riemannian manifold: Using L₂E

Equipped with a robust FM formulation and the basic PGA, we are now ready to propose a formulation for the robust PGA. Let the i^{th} reconstructed data point be denoted by \hat{x}_i , then using an approach analogous to the one used to define robust FM, we model the density for the distance between x and \hat{x} by a partial mixture of half-normal density with $\sqrt{2}\sigma'/\sqrt{\pi}$ mean and variance $\sigma'^2(1 - 2/\pi)$ (here σ' is unknown). Let $\epsilon' = d(x, \hat{x})$, using the same notation as in Eq. 4, $\hat{x} = \Pi_{S_j}(x)$ and $S_j = \text{Exp}_\mu(\text{span}\{V_{j-1}, \mathbf{v}_j\})$. Now to get the set of parameters $\{\hat{\mathbf{v}}_j\}, \hat{w}' > 0, \hat{\sigma}'^2 > 0$, we minimize the following L₂E criterion:

$$\text{L}_2\text{E}(\{\mathbf{v}_j\}, w', \sigma'^2) = \frac{w'^2}{\sqrt{\pi}\sigma'^2} - \frac{2\sqrt{2}w'}{\sqrt{\pi}N\sigma'} \sum_{i=1}^N \exp\left\{-\frac{d^2(\hat{x}_i, x_i)}{2\sigma'^2}\right\}, \quad (5)$$

with an added constraint to ensure that the principal vectors $\{\mathbf{v}_j\}$ are mutually orthogonal. In order to minimize the above function, we need an analytic expression for the projection, $\Pi_{S_j}(x)$. This analytic expression can either be exact or an approximation. We now present the derivations of the projection operator $\Pi_{S_j}(x)$.

Various Forms of the Projection Operator In this section, we present a method to approximate $\Pi_{S_j}(x)$ on a Riemannian manifold \mathcal{M} . On manifolds with constant sectional curvature, we resort to the exact analytic form of $\Pi_{S_j}(x)$ derived in [3]. Let $\hat{x} = \Pi_{S_j}(x)$, then \hat{x} can be expressed as $\text{Exp}_\mu(\sum_j c(x, \mathbf{v}_j) \mathbf{v}_j)$ where the coefficient function $c: \mathcal{M} \times T_\mu\mathcal{M} \rightarrow \mathbf{R}$ can be defined as $c(x, \mathbf{v}_j) = \text{sgn}(g_\mu(\mathbf{v}_j, \text{Log}_\mu x)) d(\mu, \Pi_{\text{span}\{\mathbf{v}_j\}}(x))$, where $\Pi_{\text{span}\{\mathbf{v}_j\}}(x)$ returns the closest point of x on the geodesic of dim-1 submanifold spanned by \mathbf{v}_j . We use $\text{sgn}(g_\mu(\mathbf{v}_j, \text{Log}_\mu x))$ to define $c(x, \mathbf{v}_j)$, as the coefficient can be negative as well. Since, on a general Riemannian manifold, $\Pi_{\text{span}\{\mathbf{v}_j\}}(x)$ is the solution of a hard optimization problem [26], here we approximate $c(x, \mathbf{v}_j)$ by $c(x, \mathbf{v}_j) = g_\mu(\text{Log}_\mu x, \mathbf{v}_j)$.

Moving on to the case of non-zero constant curvature manifolds, it is possible to derive the exact analytic expression of the projection operator as was shown in [3]. These analytic expressions will considerably reduce the computational complexity involved in computing the projection operator. Equipped with these closed form expressions for $\Pi_{\text{span}\{\mathbf{v}_j\}}(x)$ on constant curvature manifolds, we can compute $c(x, \mathbf{v}_j)$ analytically as $c(x_i, \mathbf{v}_j) = \text{sgn}(g_\mu(\mathbf{v}_j, \text{Log}_\mu x_i)) d(\mu, \Pi_{\text{span}\{\mathbf{v}_j\}}(x_i))$ on constant curvature manifolds. Thus, we get $\hat{x}_i = \text{Exp}_\mu\left(\sum_j c(x_i, \mathbf{v}_j) \mathbf{v}_j\right)$, for all i .

Theorem 2. *The L₂E formulation to compute the PGs in Eq. 5 is robust to outliers.*

Proof. We first observe that the minimization of $\text{L}_2\text{E}(\{\mathbf{v}_j\}, w', \sigma'^2)$ in Eq. 5 is equivalent to the maximization of

$$\text{L}_2\tilde{\text{E}}(\{\mathbf{v}_j\}, w', \sigma'^2) = \frac{w'^2}{\sqrt{\pi}\sigma'^2} - \frac{2\sqrt{2}w'}{\sqrt{\pi}N\sigma'} \sum_{i=1}^N \exp\left\{-\frac{d^2(\hat{x}_i, \mu)}{2\sigma'^2}\right\}$$

This follows from the fact that minimization of reconstruction error is equivalent to maximization of variance of the reconstructed point. This equivalence relation is exploited in literature [26,3]. Now, assume that x_j is an outlier for some j , i.e., $d(x_j, \mu)$ is very large. We can see that, for all \mathbf{v} , $c(x_j, \mathbf{v}) = g_\mu(\mathbf{Log}_\mu x_j, \mathbf{v})$ is very large as norm of $\mathbf{Log}_\mu x_j$ is very large. So, $d(\hat{x}_j, \mu) = \|(\sum_k c(x_j, \mathbf{v}_k) \mathbf{v}_k)\|$ is also very large, where $\|\cdot\|$ is taken with inner product g_μ . Now, the influence function of $L_2\tilde{E}(\{\mathbf{v}_j\}, w', \sigma'^2)$ is proportional to $\frac{\partial L_2\tilde{E}(\{\mathbf{v}_j\}, w', \sigma'^2)}{\partial d(\hat{x}_j, \mu)}$. Using calculations analogous to those in the proof of Theorem 1, we can see that as $d(\hat{x}_j, \mu) \rightarrow \infty$, $\frac{\partial L_2\tilde{E}(\{\mathbf{v}_j\}, w', \sigma'^2)}{\partial d(\hat{x}_j, \mu)} \rightarrow 0$, i.e., our formulation to compute the PGs is robust. ■

Similar to the L_2E FM, we derived an analytic gradient for equation 5, and employed the manifold extension of a variant of AGD in [10]. However, due to lack of space, we do not include them here.

3 Experiments

In this section, we present results for data lying on two Riemannian manifolds namely: the hypersphere, \mathbf{S}^m (with canonical metric), and the symmetric positive definite matrix manifold, $SPD(m)$ (with $GL(m)$ invariant metric). In all experiments, we randomly perturb some fraction of the data points in order to create outliers. We performed two sets of experiments for each data set which are described below:

- We compared our Robust L_2E -FM, μ^* , with the conventional FM, $\bar{\mu}$, and the Fréchet Median (FMe), $\tilde{\mu}$, [9,8] of the outlier added data set. Let the FM of the original data, i.e., without the outliers, be denoted by μ . Then, we compared $d(\mu, \mu^*)$ with $d(\mu, \tilde{\mu})$ and $d(\mu, \bar{\mu})$ respectively, where d is the geodesic distance on the manifold where the data reside. We also computed and compared the sample variances (s^2) with μ^* , $\tilde{\mu}$ and $\bar{\mu}$ using the same geodesic distance.
- We compared the proposed L_2E -PGA, with a robust extension of the PGA algorithm in [7], where instead of PCA in the tangent space anchored at the FM, we use a robust PCA on the Log-mapped data in the tangent space anchored at the FMe, $\tilde{\mu}$, of the given data. The specific robust PCA algorithm used in this context is the one in [12], which uses trimmed-Grassmann averages to compute the principal components in the PCA algorithm. Robustness is achieved via the use of ℓ_1 -norm. Hence, we call this the Grassmann-median PGA or simply GMPGA. We measured the reconstruction error, E_r , to assess performance of the methods. The original data without the outliers was used to compute the reconstruction error using just the leading principal vector. For the sake of completeness, a similar comparison with PGA [7] is also reported.

In both the above cases, we also report the computation time for each method. Now, we will separately discuss the results for the two commonly encountered manifolds namely, the \mathbf{S}^m and the $SPD(m)$.

3.1 Robust FM and PGA for Data on the Hypersphere \mathbf{S}^m

In this section, we first present the results for a synthetic data set on \mathbf{S}^2 . In this data, we first generated 1000 samples on \mathbf{S}^2 by perturbing points along a chosen direction $(0, 1, 0)^t$, with the perturbation following a log-normal distribution in the tangent space at the north pole. This created a band of points along the aforementioned direction (see Fig. 1). Then we randomly select 5%, 10%, 15% and 20% of the data (percentage of outliers is denoted by ς), and overwrite them with points similarly generated but along a different direction vector anchored at the north pole, specifically the vector $(0, 1, 1)^t$. This produces the said amount of outliers as shown in Fig. 1. We then compared the performances of L_2E -PGA, GMPGA and PGA on this data set.

From the plots, it is evident that the L_2E -PGA outperforms the competition. The conventional PGA fails to detect outliers as expected. The detailed comparison results with GMPGA and PGA are shown in Table 1 (right). In table 1 (on the left), we also present the comparative performances of L_2E -FM and $\tilde{\mu}$. We can see that for FM computation, L_2E -FM outperforms the non-robust FM, $\tilde{\mu}$. In the case of L_2E -PGA, it takes comparable time but gives the best performance compared to the two competitors.

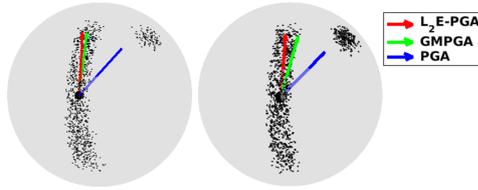


Fig. 1. Robust FM computation for the synthetic data on \mathbf{S}^2 with, *Left:* 10% and *Right:* 20% outliers.

ς (%)	μ^*			$\tilde{\mu}$			$\tilde{\mu}$			L_2E -PGA		GMPGA		PGA	
	$d(\mu, \mu^*)$	s^2	t(s)	$d(\mu, \tilde{\mu})$	s^2	t(s)	$d(\mu, \tilde{\mu})$	s^2	t(s)	E_r	t(s)	E_r	t(s)	E_r	t(s)
5	0.04	0.24	2.01	0.06	0.25	0.01	0.05	0.25	0.32	0.01	3.63	0.01	2.65	0.09	1.67
10	0.07	0.24	2.84	0.12	0.26	0.01	0.08	0.25	0.37	0.01	3.56	0.01	4.71	0.10	1.69
15	0.12	0.25	3.15	0.17	0.27	0.01	0.14	0.26	0.23	0.01	4.92	0.02	4.81	0.10	1.71
20	0.16	0.26	3.35	0.23	0.29	0.01	0.19	0.27	0.31	0.01	9.62	0.04	7.07	0.11	3.90

Table 1. Synthetic data results on \mathbf{S}^2 for, *Left:* FM and *Right:* PGA.

The poor performance of conventional PGA in the presence of outliers is not at all surprising because it can not cope with outliers. GMPGA however demonstrates comparable performance to our L_2E -PGA in the case of a small fraction of outliers but not for larger fractions.

OASIS data [18]: We now compare the performance of the L_2E based FM and PGA with the competing methods on publicly available OASIS data. This dataset consists of T1-MR brain scans of subjects with ages in the range from 18 to 96 including individuals with early Alzheimer’s Disease. We have identified an individual to be *Young* (with age between 10 to 40), *Middle Aged* (with age between 40 to 70) and *Old* (with age between 70 to 100).

ς (%)	μ^*			$\tilde{\mu}$			$\tilde{\mu}$			L_2E -PGA		GMPGA		PGA	
	$d(\mu, \mu^*)$	s^2	t(s)	$d(\mu, \tilde{\mu})$	s^2	t(s)	$d(\mu, \tilde{\mu})$	s^2	t(s)	E_r	t(s)	E_r	t(s)	E_r	t(s)
5	0.09	1.13	4.69	0.19	1.16	0.01	0.16	1.14	0.07	1.08	1.04	1.11	0.85	1.11	0.78
10	0.14	1.14	6.83	0.26	1.17	0.01	0.23	1.15	0.07	1.08	3.91	1.13	0.93	1.13	0.80
15	0.25	1.15	21.76	0.47	1.26	0.01	0.37	1.19	0.08	1.10	4.60	1.24	0.97	1.13	0.80
20	0.35	1.19	22.11	0.48	1.27	0.01	0.43	1.25	0.08	1.11	4.60	1.25	2.63	1.13	0.81

Table 2. OASIS data results on CP^{250} for, *Left:* FM and *Right:* PGA

We randomly picked 4 brain scans from within each decade, totalling 36 brain images. From each brain scan, we segmented the corpus callosum (CC) region. Then, we take a set of landmark points from the boundary of the CC shape and map it to *Kendall's shape space* [16], which is a complex projective space, $\mathbf{C}P^{250}$. The results are shown in the Table 2. Similar to before, this result indicates a superior performance of the L_2E -FM in comparable time. In the case of PGA, L_2E -PGA yields a smaller reconstruction error compared to GMPGA and the conventional (non-robust) PGA. This superior performance could be attributed to the fact that unlike in GMPGA, there is no linearization operation in the L_2E -PGA.

Further, the L_2E formulation has no tuning parameters and is less prone to local minima due to the presence of a natural scale parameter σ' which is automatically adjusted starting at a large initial value permitting global search and gradually decreasing to small σ' that permits a precise approximation. For each of the three classes of individuals, i.e., Young, Middle aged and Old, we also present the reconstructed shape using first 34 principal geodesics in Fig. 2. In terms of shape reconstruction,

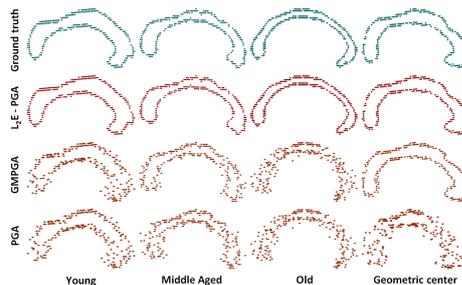


Fig. 2. Reconstruction results from the OASIS data

PGA performs the worst, and L_2E -PGA performs the best. The last column of this figure depicts the mean shape computed using L_2E -FM, the conventional FM and FMe. As expected, FMe is better than FM but our L_2E -FM results in the “best” mean shape.

3.2 Robust FM and PGA on $SPD(m)$

Movement disorder data: In this experiment, the data consists of HARDI acquisitions from patients with Parkinsons disease (PD), essential tremor (ET). The goal here is to perform robust PGA and demonstrate the power of this representation via a depiction of the reconstruction error. All the HARDI data for full brain were acquired on a 3T Phillips MR scanner using a single-shot spin echo EPI sequence, with the following acquisition parameters: repetition time=7748ms, echo time=86ms, flip angle=90, # of diffusion gradients: 64, field of view = 224 224 mm, in-plane resolution = 2 mm iso-tropic, slice-thickness=2 mm, SENSE factor=2. Data from 22 control and 26 PD patients were acquired using the above HARDI acquisition protocol. Distortions due to eddy currents and head motion was corrected by using the FSL software.

From previous studies, it is well known that, the basal ganglia region of the brain is significantly affected by Parkinson’s disease, we chose our ROI for analysis to be this region in the brain. The image volume size we work with here is $(112 \times 112 \times 60)$ for each diffusion gradient direction.

Along with the diffusion images, for each data, we also have a mask defining the following region of interests (ROIs) in basal ganglia: left and right anterior substantia nigra, left and right posterior substantia nigra, left and right thalamus, and left and right putamen. We first constructed an atlas of the control population [5], and then affinely registered all the s_0 images, with the s_0 image of the atlas, to bring them to a common co-ordinate system. For each image, we used the rotation, computed from the affine matrix, to re-orient the 64 gradient directions. We also used the affine matrices to warp the ROI masks, so that they will match the registered images. We then non-rigidly registered each image to the atlas using [4]. From this, we computed the Cauchy Deformation Tensor (CDT) in each voxel as $\sqrt{JJ^T}$, where J is the Jacobian from the non-rigid registration.

ς (%)	μ^*			$\hat{\mu}$			$\tilde{\mu}$		
	$d(\mu, \mu^*)$	s^2	t(s)	$d(\mu, \hat{\mu})$	s^2	t(s)	$d(\mu, \tilde{\mu})$	s^2	t(s)
5	1.40	0.01	3.11	31.63	998.32	1.99	16.26	264.14	4.07
10	1.47	0.01	10.64	42.94	1841.98	2.39	25.91	672.47	4.16
15	1.57	0.01	12.41	65.24	4294.63	2.70	53.97	2916.183	4.01
20	1.66	0.01	26.72	88.79	7889.97	2.88	73.74	4543.60	4.23

L ₂ E-PGA		GMPGA		PGA	
E_r	t(s)	E_r	t(s)	E_r	t(s)
0.19	151.96	1.93	43.47	8.26	10.50
0.22	153.74	7.96	51.82	6.63	22.49
0.22	156.50	13.74	59.39	13.06	43.00
0.22	157.21	14.96	62.68	16.79	58.99

Table 3. Movement disorder data results on the product manifold of SPD(3) for, *Top*: FM and *Bottom*: PGA

We first constructed an atlas of the control population [5], and then affinely registered all the s_0 images, with the s_0 image of the atlas, to bring them to a common co-ordinate system. For each image, we used the rotation, computed from the affine matrix, to re-orient the 64 gradient directions. We also used the affine matrices to warp the ROI masks, so that they will match the registered images. We then non-rigidly registered each image to the atlas using [4]. From this, we computed the Cauchy Deformation Tensor (CDT) in each voxel as $\sqrt{JJ^T}$, where J is the Jacobian from the non-rigid registration.

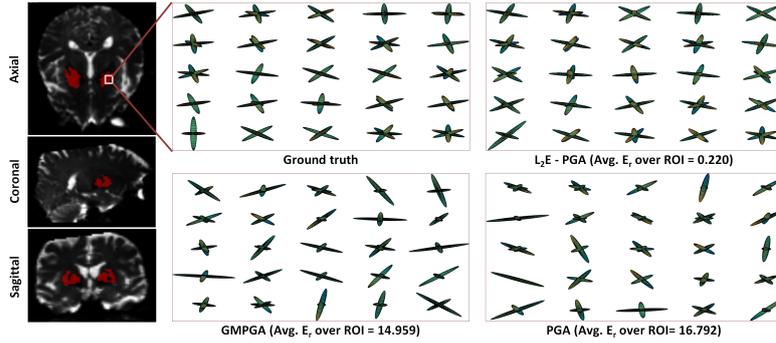


Fig. 3. CDT reconstruction results for the Movement disorder data

Since CDT is a symmetric positive definite (SPD) matrix, the CDT field of each image lies on a product SPD manifold. We constructed a combined ROI mask, from the 22 initial ROI masks of the control population. In this combined ROI mask, a voxel value is set to 1, if there is more than $\eta\%$ overlap of the initial ROIs, else it is set to 0. To emphasize the effect of the PD in the basal ganglia, among the $112 \times 112 \times 60$ voxels in each image, we considered the CDTs only in the voxels of the combined ROI mask. We chose $\eta = 50$ to get 864 voxels. It is not uncommon to misclassify between patients with essential tremor (ET) and PD patients, hence, we naturally have data samples from ET patients as an outlier. As before, we have reported results in Table 3 by varying the percentage of outliers present in the data, i.e., varying mis-labelled samples with ET. The results in Table 3 indicates superior E_r values from our formulation compared to the competitors. In Fig. 3, we have shown comparison results of reconstructed

CDTs (with 20% outliers) of a 5×5 region inside the ROI (region colored red in the figure). Our visualization of the CDT (3×3 SPD matrices) presented here is as follows: eigen values are used as lengths of the axis of the ellipsoid and eigen vectors give the orientation of each ellipsoid. It is clear from the figure that our robust PGA formulation yields a better reconstruction.

4 Conclusions

In this paper, we presented novel algorithms to compute the robust FM, dubbed the L_2E -FM, and robust PGA, dubbed the L_2E -PGA, for data on Riemannian manifolds. In both these problems, we formulated the minimizations involved using an M-estimator called the L_2E . One of the key advantages of the proposed L_2E based formulation is that it is free of tuning parameters. Further, unlike the conventional PGA which uses a linear approximation of the manifold in the neighborhood of the FM, L_2E -PGA uses the exact-PGA cost function which yields more accurate results even in the case of data with large variance. Through an extensive set of synthetic and real experiments, we showed that our L_2E formulation achieves robustness in computing both the FM and PGA. Further, since there are no other robust PGA methods in literature to compare with, we developed the GMPGA method, which performs GMPCA (Grassmann averaging to compute PCA) [12] in the tangent space anchored at the FM. Finally, we presented experiments on MRI data from the publicly available OASIS database and diffusion MRI scans of movement disorder patients as well as synthetic data on the sphere with varying amounts of outliers to demonstrate superior performance of our robust algorithms in comparison to the competing methods.

Acknowledgements: Authors thank Drs. Vaillancourt, Okun and Ofori of the University of Florida, for providing us the movement disorder data used here.

References

1. Afsari, B.: Riemannian L_p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* 139(2), 655–673 (2011)
2. Basu, A., Harris, I.R., Hjort, N.L., Jones, M.: Robust and efficient estimation by minimising a density power divergence. *Biometrika* 85(3), 549–559 (1998)
3. Chakraborty, R., Seo, D., Vemuri, B.C.: An efficient exact-pga algorithm for constant curvature manifolds. *IEEE CVPR* (2016)
4. Cheng, G., Vemuri, B.C., Carney, P.R., Mareci, T.H.: Non-rigid registration of high angular resolution diffusion images represented by gaussian mixture fields. In: *MICCAI*. pp. 190–197 (2009)
5. Cheng, G., Vemuri, B.C., Hwang, M.S., Howland, D., Forder, J.R.: Atlas construction from high angular resolution diffusion imaging data represented by gaussian mixture fields. In: *ISBI*. pp. 549–552 (2011)
6. DoCarmo, M.P.: *Riemannian Geometry*. Birkhauser (1992)
7. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE TMI* 23(8), 995–1005 (2004)

8. Fletcher, P.T., Venkatasubramanian, S., Joshi, S.: The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage* 45(1), S143–S152 (2009)
9. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. In: *Annales de l'institut Henri Poincaré*. vol. 10, pp. 215–310 (1948)
10. Ghadimi, S., Lan, G.: Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Math. Programming, Ser. A* 156(1), 59–99 (2016)
11. Hauberg, S.: Principal curves on riemannian manifolds. *IEEE TPAMI* 38(9), 1915–1921 (2015)
12. Hauberg, S., Feragen, A., Black, M.J.: Grassmann averages for scalable robust PCA. In: *CVPR*. pp. 3810–3817 (2014)
13. Hotz, T., Huckemann, S., Le, H., Marron, J.S., Mattingly, J.C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V., Skwerer, S., et al.: Sticky central limit theorems on open books. *The Annals of Applied Probability* 23(6), 2238–2258 (2013)
14. Huber, P.J.: *Robust statistics* (2011)
15. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic pca for riemannian manifolds modulo isometric lie group actions. *Statistica Sinica* (2010)
16. Kendall, D.G.: Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* 16(2), 81–121 (1984)
17. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing. *JMIV* 25(3), 423–444 (2006)
18. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, Y.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* pp. 1498–1507 (2007)
19. Mardia, K., Dryden, I.: Shape distributions for landmark data. *Advances in Applied Probability* pp. 742–755 (1989)
20. Moakher, M., Batchelor, P.G.: Symmetric positive-definite matrices: From geometry to applications and visualization. In: *VPTF*, pp. 285–298 (2006)
21. Pennec, X.: Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *JMIV* 25(1), 127–154 (2006)
22. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. *IJCV* 66(1), 41–66 (2006)
23. Said, S., Courty, N., Le Bihan, N., Sangwine, S.J.: Exact principal geodesic analysis for data on $SO(3)$. In: *EUSIPCO-2007*. pp. 1700–1705 (2007)
24. Scott, D.W.: Parametric statistical modeling by minimum integrated square error. *Technometrics* 43(3), 274–285 (2001)
25. Scott, D.W.: Outlier detection and clustering by partial mixture modeling. In: *COMPSTAT'04*. pp. 453–464 (2004)
26. Sommer, S., Lauze, F., Hauberg, S., Nielsen, M.: Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations. In: *ECCV*, pp. 43–56 (2010)
27. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B* 61(3), 611–622 (1999)
28. Zhang, M., Fletcher, P.T.: Probabilistic principal geodesic analysis. In: *Advances in Neural Information Processing Systems*. pp. 1178–1186 (2013)