REGULAR ARTICLE

Highly resistant gradient descent algorithm for computing intrinsic mean shape on similarity shape spaces

H. Fotouhi · M. Golalizadeh

Abstract Among many algorithms, gradient descent algorithm (GDA) is a simple tool to derive an optimal quantity in dealing with an optimization problem in the linear space. Apart from the initial value, the step size has a great impact on the convergence rate of this algorithm. Its affect on the geometric structure of the consecutive configurations is more crucial if one works with an optimization problem in the statistical shape analysis. In other words, if the step size of the GDA is not properly tuned, the geometry might not be preserved while the algorithm is moving forward to reach an optimal mean shape. In order to improve the performance of the GDA, we introduce a dynamic step size and a new criterion both to check the geometry in each step of the algorithm and to accelerate the convergence rate. These lead to a new robust algorithm on deriving the intrinsic mean on the shape space. We compare the performance of our proposed procedure to the usual GDA using a real shape data accompanied with simulation studies.

1 Introduction

It is known from elementary statistics that the mean of a set of data lying on an Euclidean space is the minimizer of the sum of squared Euclidean distance of a fixed point to the observations on hand. However, if the data under study are not the elements of an Euclidean space, the well-known Pythagorean theorem cannot be directly used any more and some adaptions are required. Generally speaking, a proper Riemannian

H. Fotouhi · M. Golalizadeh (✉)
Tarbiat Modares University, Tehran, Iran
e-mail: golalizadeh@modares.ac.ir

metric needs to be considered if the observed data belong to a curved manifold (Pennec 1999). Nonetheless, this is not the case for every curved manifold. Instead, as discussed in Bhattacharya and Patrangenaru (2003) and Bhattacharya and Patrangenaru (2005), one can obtain a metric via embedding the manifold in a higher dimensional Euclidean space. This results in an extrinsic analysis using the metric inherited from the ambient Euclidean space. Then, one can derive various summary statistics including the sample mean and variance. Further, making some statistical inferences such as building a confidence region or predicting a new value will also be possible.

One of the new areas in the statistical setting which was under consideration over the last three decades or so, is the statistical shape analysis (Dryden and Mardia 1998). Frankly speaking, it deals with any phenomenon in which the geometrical feature is the main interest of the research. Mathematically, having an object lying in Euclidean space, its shape is defined as what is left after removing the location, scale and rotations effects [called similarity transformation (SO)] out from that object (Kendall 1984). Hence, shape data are considered as the points on a finite dimensional nonlinear differential manifold (Grenander 1994). It is also known that shape space can be viewed as the quotient of a Riemannian manifold.

A common Riemannian metric in the statistical shape analysis is the Procrustes distances and the mean shape is defined as the minimizer of the squared Procrustes distances (Kendall 1984). However, to derive a measure of variability among shapes and particularly to perform multivariate statistical analysis, the usual standard statistical tools, available on Euclidean space, cannot be directly utilized. So, among many methods to obtain the shape variance, the non-Euclidean shape space is approximated by a linearized space at the vicinity of the mean shape and then the Principal Component Analysis (PCA) is invoked there (Dryden and Mardia 1998). Recently, a new tool which is called Principal Geodesic Analysis (PGA) was proposed to, directly, evaluate the variability on the curved manifolds including the shape space (Fletcher et al. 2004). The core basis of this method, which works well in some particular spaces, is mainly based upon GDA. Among many performance parameters of the PGA, the step size and the threshold value are key components on both accelerating the algorithm and guaranteeing the convergence to the optimum object.

In this paper, we will demonstrate how, at each stage of the GDA, an non-tuned step size both increases the time of convergence and fails to preserve geometrical structures of the objects before reaching the intrinsic mean shape. This is due to the fact that the optimum choice of step size parameter not only accelerate the convergence rate but also maintain the geometrical structure of the shape under study. It further helps the user to control the results in each step of the algorithm before reaching the optimum object which is the intrinsic mean shape in our case. Later, by introducing a new criterion for checking geometrical structure of objects, we propose a more sensible algorithm which works well in the shape analysis context. The performance of our proposed method is compared to the usual GDA in estimating the intrinsic mean shape of a real data set and also a simulation study.

This paper is organized as follows. A brief review of statistical shape analysis along with a popular measure of similarity are presented in Sect. 2. Then, statistical analysis of manifold data and generalization of GDA to the manifold valued statistics are given in Sect. 3. Later, in Sect. 4, we first highlight non-stability of the standard GDA in

preserving the geometrical structure of objects. Then, we propose a new procedure, called robust gradient descent algorithm (RGDA), which is more resistant than the standard GDA. Performance of employing both the GDA and RGDA on the real data set and also in a simulation study is evaluated in Sect. 5. The paper ends with some concluding remarks.

2 A brief review of shape analysis

Based upon the progress on technology, new fields of sciences have merged into various discipline. Statistical shape analysis is one of the new, and active area of multivariate statistics which deals with geometrical structures of objects. Although the historical background of this field goes as far back as to Galileo, it was formally defined and introduced to the statistical communities by late (Kendall 1977). He defines the shape of an object as the whole geometrical information left when the location, scale and rotation effects are filtered out from that object. Thereafter many attitudes to the shape analysis were raised and the application of this new statistical field has motivated many researchers to employ it in other sciences. Some instances are statistical shape analysis of brain (Free et al. 2001), discovering variability of DNA molecules (Dryden 2002), medial image processing (Fletcher et al. 2003), facial gender classification (Wu et al. 2007) and study of vertebral fractures (Sommer et al. 2010). Other applications of this subject can be found in Dryden and Mardia (1998).

In the statistical shape analysis, one is usually encountered with objects or configurations. As expected, the original objects are usually available in two or three dimensions. However, to give an insight of the shape analysis, here we provide the mathematical backgrounds for the 2-dimensional (2D) objects. To start, let us consider a configuration identified by $k \geq 3$ key points in two dimensions.

These points are called the landmarks (see, e.g. Dryden and Mardia 1998) and are set on the interior of the objects or on its outline (see, e.g. Bookstein 1991). However, these cases are not of any interest in this paper.

Suppose we set k landmarks on an object lying on 2D Euclidean space. It is common to write the coordinate of these landmarks in a $k \times 2$ matrix known as configuration matrix. Clearly, this configuration can also be identified by a $k \times 1$ dimensional complex vector. In our case, the landmarks are the pairs of (x, y), i.e. 2-D Cartesian coordinates. Following Kendall (1984), we assume that the landmarks do not totally coincide as this odd case won't define a proper shape. It is worth to mention that various approaches to derive the shape of an object led to different shape coordinate systems. Among them, Kendall (1977) and Bookstein (1986) coordinates are well known shape coordinate systems. The researchers who are new to the statistical shape analysis prefer to work with latter one at least because of its correct geometrical view (Dryden and Mardia 1998).

One of the main objective of the statistical shape analysis is to evaluate how close the objects are to each other. Obviously, the optimal matching is done using filtering out the translation, scale, and rotation effects. According to the statistical shape analysis terminologies (see, e.g. Dryden and Mardia 1998), superimposition of two objects, through filtering out similarity transformation, as close as possible is known as the

full Ordinary Procrustes Analysis (OPA), and the resulted distance is called the full Procrustes distance, denoted by $d_F(.,.)$. Accordingly, the squared distance between A and B, two centred $k \times 2$ configuration matrices (it can be done using, for instance, the centering matrix), are defined as

$$d_F^2(\mathbf{A}, \mathbf{B}) = \inf_{\boldsymbol{\Gamma} \in SO(2),\ \beta \in \mathbb{R}^+,\ \boldsymbol{\gamma} \in \mathbb{R}^2} \| \mathbf{B} - \beta \boldsymbol{\Gamma} \mathbf{A} - 1_k \boldsymbol{\gamma}^T \|^2,$$

where $\beta > 0$ is a scale parameter, $\boldsymbol{\Gamma}$ is an 2×2 rotation matrix, and $\boldsymbol{\gamma}$ is an 2×1 location vector. Also, $SO(2)$ is the special orthogonal group of 2×2 rotation matrices, i.e. $\boldsymbol{\Gamma}$ satisfying equalities $\boldsymbol{\Gamma}^T \boldsymbol{\Gamma} = \boldsymbol{\Gamma} \boldsymbol{\Gamma}^T = I_2$ and $|\boldsymbol{\Gamma}| = 1$.

To clarify the concept of OPA we here provide a simple example. Consider the triangles **A**, **B** and **C** with the following configuration matrices:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

It is clearly seen that **A** and **B** are equilateral and **C** is right-angled (also isosceles). In fact, the configuration matrix **B** has already been registered in Bookstein coordinate system, i.e. its last row is shape coordinate, and this will be used in the section dealing with simulation study.

Difference between these triangles can be easily computed by hand or using the function `procOPA` available under the library `shapes` (Dryden 2011), in the statistical computing software (R Development Core Team 2012). Based upon the outputs, we have

$$d_F^2(\mathbf{A}, \mathbf{C}) = 2.5 > d_F^2(\mathbf{B}, \mathbf{C}) = 1.25 > d_F^2(\mathbf{A}, \mathbf{B}) = 0.$$

Consequently, two triangles **A** and **B** are more similar than **A** and **C** or **B** and **C**.

An extension of OPA to more configurations is simply defined through Procrustes rotation (Mardia et al. 1979) and is called Generalized Procrustes Analysis (GPA) or full Procrustes fitting (Dryden and Mardia 1998). Accordingly, the full Procrustes mean shape or Procrustes mean shape is a solution of the optimization problem

$$\arg \min_{\mathbf{b}} \sum_{i=1}^{N} d_F^2(\mathbf{x}_i, \mathbf{b}),$$

where $\mathbf{x}_i,\ i = 1, \ldots, N$ and $\mathbf{b}$ are the shape configurations in terms of the Kendall's shape coordinate system. Note that a closed form solution of this optimization problem exists (see, e.g. Result 3.2 on page 44 of Dryden and Mardia 1998).

In general, the shape space is a finite dimensional nonlinear Riemannian manifold. So statistical analysis of shapes is led to study of statistics on manifold. In order to perform statistical analysis on manifold, it is common to approximate the manifold by a linearized space and perform most of the analysis there and then project the result back to the manifold. For example, to explore the variability on the shape space, the PCA is

usually employed on the tangent space where the Euclidean property can be invoked (Dryden and Mardia 1998). The embedding was also recommended in some texts (see, e.g. Hendriks and Landsman 1998; Patrangenaru 1998; Bhattacharya and Patrangenaru 2003, 2005; Micheas and Dey 2005). Although these approaches might provide some reasonable answers to the questions on hand, in some circumstances the induced errors are high (Huckemann and Hotz 2009) and it sounds the better choice is to perform statistical analysis directly on the manifold. The advantages of this will be consistency in representation, dimensionality reduction, and accuracy in measurements (Sommer et al. 2010). Along with some useful definitions, a comprehensive treatment of how to perform statistical analysis directly on manifolds is given in Pennec (2006).

3 GDA and its extension on shape space

The GDA or classical steepest descent method, which was first proposed by Cauchy (1847), is one of the oldest methods to minimize any general nonlinear function. Using the exact gradient vector, the GDA starts from an appropriate arbitrary initial value to calculate the minimum of one-dimensional or multi-dimensional functions. It goes in the positive direction of the gradient vector at each stage of the algorithm. The convergence rate of this algorithm is highly dependent on the step size and in a lesser extend on the initial value.

There were many efforts to improve the efficiency of the GDA method. To name some, we could mention (Bershad 1987; Matthews and Xie 1993; Douglas and Pan 1995; Liu 2001; Meza 2010). These led to a newfound interest in the steepest descent method, both from a theoretical and practical viewpoints. It was emphasized all over these papers that although the gradient direction will guarantee the convergence, a wrong step size might decrease the convergence rate. More cares are needed if the space under study is non-Euclidean, or generally a curved manifold. In this paper, we provide some ideas to explain where this situation will arise in employing the the GDA in the shape analysis. However, in this section we review some backgrounds on utilizing the GDA in the shape space.

Let M be a manifold and $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be N observations from this manifold. A solution to the optimization problem

$$\arg\min_{\mathbf{a}\in M} \sum_{i=1}^{N} d^2(\mathbf{x}_i, \mathbf{a}), \tag{1}$$

is defined as the intrinsic sample mean (say $\widehat{\boldsymbol{\mu}}$), where $d^2(\mathbf{x}_i, \mathbf{a})$ indicates the squared Riemannian distance between the i-th data point and a fixed m-vector $\mathbf{a}$ (Karcher 1977). According to Fletcher et al. (2004), the Riemannian distance between two points $\mathbf{x}, \mathbf{y} \in M$, is defined as the minimum length over all possible smooth curves between $\mathbf{x}$ and $\mathbf{y}$. Using an idea originally proposed in Pennec (1999), it was demonstrated in Fletcher et al. (2004) that a solution to (1) is gained via employing the GDA in the manifold M. Also, an iterative technique based upon the Newton's method was proposed, in Groisser (2004), to derive a solution of (1) and referred it to as the

Riemannian center of mass. The counterpart of this sample mean in the population (in statistical sense) is called the intrinsic mean which was first treated in Bhattacharya and Patrangenaru (2003). Technically, these two terms are different but for the sake of brevity, omitting the word 'sample', we use the intrinsic mean henceforth. Further, we indicate the sample mean with $\boldsymbol{\mu}$ instead of $\widehat{\boldsymbol{\mu}}$ thereafter. It is worth to mention that this quantity was called an intrinsic Frechet mean in the statistical shape analysis community (see, e.g. Le 2001).

Let $T_{\boldsymbol{\mu}}M$, be the tangent space at the intrinsic mean $\boldsymbol{\mu}$. Then, each point $\mathbf{x}_i \in M$, can be reached by a vector $\mathbf{w}_i \in T_{\boldsymbol{\mu}}M$, via $\mathrm{Exp}_{\boldsymbol{\mu}}\mathbf{w}_i = \mathbf{x}_i$, where $\mathrm{Exp}_{\boldsymbol{\mu}}$ is the exponential map which maps straight lines through the origin $T_{\boldsymbol{\mu}}M$, to geodesics on M passing through $\boldsymbol{\mu}$. Note that the map sending $\mathbf{x}_i \in M$, to $\mathbf{w}_i \in T_{\boldsymbol{\mu}}M$, is known as the logarithm map and denoted by $\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x}_i$. Due to the Euclidean feature of $T_{\boldsymbol{\mu}}M$, a metric satisfying common distance conditions can be defined on it and then its corresponding metric on M will be obtained using a proper transformation. For more details, the reader can consults (Pennec 2006).

Following Fletcher et al. (2004), the intrinsic mean can also be defined as the solution to the optimization problem

$$\boldsymbol{\mu} = \arg \min_{\mathbf{w} \in M} \sum_{i=1}^{N} \parallel \mathrm{Log}_{\mathbf{w}}(\mathbf{x}_i) \parallel^2, \tag{2}$$

where $\mathrm{Log}_{\mathbf{w}}(.)$ is logarithm map with starting point $\mathbf{w}$ belonging to M. Then, provided the observations $\mathbf{x}_i$, $i = 1, \ldots, N$ lie in a strongly convex neighborhood, a unique solution is guaranteed (Karcher 1977). The solution is obtained via employing the GDA which takes successive steps in the negative gradient direction where the gradient is proportional to

$$- \sum_{i=1}^{N} \mathrm{Log}_{\boldsymbol{\mu}}(\mathbf{x}_i).$$

Particularly, in order to obtain $\boldsymbol{\mu}$, starting from an initial estimate of the mean (say $\boldsymbol{\mu}_0$), and using the negative gradient direction vector, the algorithm iterates through the equality

$$\boldsymbol{\mu}_{j+1} = \mathrm{Exp}_{\boldsymbol{\mu}_j}\left(\frac{\tau}{N} \sum_{i=1}^{N} \mathrm{Log}_{\boldsymbol{\mu}_j}(\mathbf{x}_i) \right), \tag{3}$$

where τ is the step size, a positive constant value. Note that this algorithm only converges locally. Hence, care must be taken in the right choices of the initial value and the step size. Valuable suggestions on how to choose some feasible values for $\boldsymbol{\mu}_0$, and τ, are given in Fletcher et al. (2004).

To have a taste on how this problem works, we provide the ideas for a simple manifold here. Particularly, we consider a hypersphere of unit radius in $2(k-1)$ real dimensions (written as S^{2k-3}), which is a pre-shape space of the non-coincident k point set configurations in 2D Euclidean space and denoted by S_k^2, (see, e.g. Dryden and

Mardia 1998). Following Buss and Fillmore (2001), the exponential map, acted on this unit hypersphere, to map a tangent vector $\mathbf{v} \in T_{\mathbf{w}} S^{2k-3}$, to the point on S^{2k-3}, with the speed $||\mathbf{v}||$, passing through the point $\mathbf{w} \in S^{2k-3}$, along the geodesic $\gamma_{\mathbf{v}}(t)$, with the initial velocity $\mathbf{v}$, is given by $\text{Exp}_{\mathbf{w}}(\mathbf{v}) = \gamma_{\mathbf{v}}(1)$. However, to the best our knowledge, there is not any explicit formula for the logarithm map for this case. Technically, we require such expression in terms of the shape coordinate systems. However, we can derive it first in terms of data points on the pre-shape space $S_k^2 = S^{2k-3}$. Below, we derive the exponential map for this space. Note that our data are $2(k-1)$-vectors, but, for ease of notations, we write them as m-vectors, i.e. $2(k-1) = m$, that means we are assuming the data are observations on $S^{2k-3} = S^{m-1}$.

Let the fixed starting point (it is called the base point in Fletcher et al. 2004) for the logarithm map is an m-vector $\mathbf{p} = (0, 0, \ldots, 0, 1)^T \in S^{m-1}$, which is the North pole on the unit hypersphere S^{m-1}. Then, according to Buss and Fillmore (2001), for a point $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_m)^T \in S^{m-1}$, the logarithm map is given by

$$\mathbf{Log_p}(\mathbf{x}) = \left(x_1 \frac{\theta}{\sin \theta}, x_2 \frac{\theta}{\sin \theta}, \ldots, x_{m-1} \frac{\theta}{\sin \theta}, 1 \right), \tag{4}$$

where $\theta = \arccos x_m$ indicates the spherical distance of point x from the center of the unit hypersphere. It can be seen from Eq. (2) with $M = S^{m-1}$, that the starting point should be an arbitrary starting point rather than the fixed $\mathbf{p}$. So, in order to derive the intrinsic mean shape using (3), we require the logarithm map given by (4) in terms of an arbitrary point $\mathbf{w} \in S^{m-1}$. To gain this, we use the method of projecting a point from an unit radius hypersphere onto a hyperplane.

Let the equality $x_1^2 + x_2^2 + x_3^2 + \cdots + x_m^2 = r^2$ represents the equation of an $(m-1)$-dimensional hypersphere with radius r, centre at the origin and denoted by the surface equation

$$f(\mathbf{x}, r) = f\big((x_1, x_2, x_3, \ldots, x_m)^T, r\big) = x_1^2 + x_2^2 + x_3^2 + \cdots + x_m^2 - r^2 = 0. \tag{5}$$

Then, re-writing Eq. (5) in terms of $\mathbf{x}_{(-m)}$, the tangent plane to the hypersphere at an arbitrary point $\mathbf{w} = (w_1, w_2, w_3, \ldots, w_m)^T$ is given by

$$(x_1 - w_1) \frac{\partial x_m}{\partial x_1}\bigg|_{\mathbf{x}_{(-m)} = \mathbf{w}_{(-\mathbf{m})}} + \cdots + (x_{m-1} - w_{m-1}) \frac{\partial x_m}{\partial x_{m-1}}\bigg|_{\mathbf{x}_{(-m)} = \mathbf{w}_{(-\mathbf{m})}} - (x_m - w_m) = 0.$$

where

$$\mathbf{x}_{(-i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)^T, \quad \mathbf{w}_{(-i)} = (w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_m)^T.$$

Simple mathematical manipulation turn this latter equation to $\langle \mathbf{x}, \mathbf{w} \rangle = r^2$, where $\langle ., . \rangle$ represents the inner product. On the other hand, the projection of the point $\mathbf{w}$, set on the tangent plane onto the hyperplane with the equation

$$\langle \mathbf{a}, \mathbf{x} \rangle + c = a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots a_m x_m + c = 0,$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_m)^T$, and c is a constant, is the point $\mathbf{w} - \lambda\mathbf{a}$, where

$$\lambda = \frac{\langle \mathbf{a}, \mathbf{w} \rangle}{\parallel \mathbf{a} \parallel}.$$

Hence, via transferring the point $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_m)^T \in S^{m-1}$ to the tangent plane of the hypersphere at point $\mathbf{p} = (0, 0, 0, \ldots, 1)^T$, using (4) and then projecting the resulted point onto the hyperplane passing through $\mathbf{w}$, which also belongs to S^{m-1}, we could derive the coordinates of the logarithm map with an arbitrary point. The map is given by

$$\mathrm{Log}_{\mathbf{w}}\left(\mathbf{x}\right) = (x_1 \frac{\theta}{\sin\theta} + \frac{w_1}{w_m}\lambda, \ldots, x_{m-1}\frac{\theta}{\sin\theta} + \frac{w_{m-1}}{w_m}\lambda, 1 + \lambda), \qquad (6)$$

where

$$\lambda = w_m \left(-x_1 w_1 \frac{\theta}{\sin\theta} - \cdots - x_{m-1}w_{m-1}\frac{\theta}{\sin\theta} + 1 - w_m\right).$$

Now, substituting the expression (6) into Eq. (3) and using the equality $\mathbf{Exp}_{\mathbf{w}}(\mathbf{v}) = \gamma_{\mathbf{v}}(1)$, along with proper adaptions to the notations provide an iterative algorithm to obtain the intrinsic mean on the shape space with the pre-shape space S_k^2. We employ this algorithm in the simulations as well as real application studies.

General comments on strong sensitivity of the algorithm to the initial value, $\boldsymbol{\mu}_0$ and step size τ and also describing a procedure on how to implement the algorithm are discussed by Fletcher et al. (2004). However, since preserving geometry is crucial in the statistical shape analysis, we believe that more care must be taken in choosing $\boldsymbol{\mu}_0$ and τ, while invoking the algorithm (3). We run this algorithm for various $\boldsymbol{\mu}_0$ and τ, in the shape analysis and saw losing the geometry of the objects to gain the intrinsic mean shape. This phenomenon shall be illustrated later in a simulation study and using some real data. To overcome this problem, we propose a robust algorithm and provide some insights to choose some feasible start values in obtaining the intrinsic mean shape in the next section.

4 Robust gradient descent algorithm to derive mean shape

As mentioned earlier, in some circumstances employing the GDA may be highly affected by nontuned step size value. Particularly, while dealing with the shape data and aiming to obtain the intrinsic mean shape, it may effect the geometrical objects generated in each stage of running the GDA. Since the geometrical form is vital in the statistical shape analysis, we aim to extend the GDA to overcome such possible problem. Particularly, we propose an algorithm to compute the intrinsic mean through implementing the tuned GDA aiming not to loose the geometry of the object under study. We call it the Robust gradient descent algorithm (RGDA). Our method consists of a modification of the GDA accompanied with statistical shape criteria. Our main objective is to extend the GDA in such way that the geometry of the intrinsic mean

shape could be more stable when the algorithm proceeds. Moreover, we set a new step on the procedure of the algorithm such that it could be more resistant to the possible odd intrinsic mean shapes during iterating the the GDA.

Recalling Eq. (2), the intrinsic mean shape leads to a real matrix which has the least sum of squared Riemannian distance from entire observations. Furthermore, it has the least Procrustes sum of square from each observation, if the Procrustes distance is used as the metric. Hence, we impose another criterion to have a compromise value between these two measures. We call it the Mean Ordinary Procrustes Sum of Square (MOSS) which, in each stage of the GDA, measures the distance of generated candidate mean from each observation. Generally, consider configuration matrices $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\boldsymbol{\mu}$, (all $k \times m$ matrices of coordinates from k points in m dimensions). Then following P. 84 in Dryden and Mardia (1998), we have

$$\text{MOSS}(\mathbf{X}_1, \ldots, \mathbf{X}_n, \boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \text{OSS}(\mathbf{X}_i, \boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^{n} ||\boldsymbol{\mu}||^2 \sin^2 \rho(\mathbf{X}_i, \boldsymbol{\mu}),$$

where $\rho(\mathbf{X}_i, \boldsymbol{\mu})$ is the Procrustes distance between $\mathbf{X}_i$ and $\boldsymbol{\mu}$. In fact, this is the mean of OPA's but we keep our notations just for consistency. Using this value, we could check how close we are to the intrinsic mean shape. Obviously, to do this we require another threshold to inspect the MOSS. It is denoted by φ in this paper. Apart from these criteria, we are interested on imposing a robust quantity to improve the performance of our algorithm. It is well known that, unlike the mean, the median is robust to the outliers. Hence, to speed up the algorithm we could consider the median of two quantities arising from two consecutive stages of the RGDA, i.e. one the GDA passed and another the current step. According to our empirical studies, this last criterion works very well in getting to the mean shape as quick as possible. Note that, the median used in our study is just entrywise which is a drawback of our proposed algorithm. Following Fletcher et al. (2009), one can use the geometric median on Riemannian manifold which is halfway point along the geodesic and is expected to performs better than our median.

Another key factor affecting the performance of the GDA is the initial value $(\boldsymbol{\mu}_0)$, of the algorithm. For the shape analysis case, we recommend choosing the Procrustes mean, denoted by $\boldsymbol{\mu}_P$, as the initial value. This is due to its strong similarity to the shape observations. Moreover, it induces the well localization of the subsequent means while the GDA goes forward. Note that the distance used for proceeding the algorithm is Riemannian distance as mentioned after Eq. (1). Lastly, in some situations the fixed threshold value (ε) might also affect the performance of the GDA. This shall simply be presented in our graphical representation in the next section.

Bearing in mind our discussion given above and defining our geometry checking function as

$$\text{GCF}(\boldsymbol{\mu}_P, \boldsymbol{\mu}_q) = \text{MOSS}(\mathbf{x}_1, \ldots, \mathbf{x}_N, \boldsymbol{\mu}_P) - \text{MOSS}(\mathbf{x}_1, \ldots, \mathbf{x}_N, \boldsymbol{\mu}_q),$$

the algorithm to derive the intrinsic mean shape is as follows:

Algorithm: Robust Gradient Descent

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in S_k^m$.

Output: $\boldsymbol{\mu} \in M$, the intrinsic mean shape.

Set $j = 1$ and, starting from an initial, say $\boldsymbol{\mu}_0$, follow the following steps:

Step 1: Set, $\Delta\boldsymbol{\mu} = \frac{\tau}{N}\sum_{i=1}^{N}\text{Log}_{\boldsymbol{\mu}_{j-1}}(\mathbf{x}_i)$, and $\boldsymbol{\mu}_j = \text{Exp}_{\boldsymbol{\mu}_{j-1}}(\Delta\boldsymbol{\mu})$.

Step 2: If $\parallel \text{GCF}(\boldsymbol{\mu}_P, \boldsymbol{\mu}_j) \parallel \leq \varphi$, go to Step 3.

        Else, set $\boldsymbol{\mu}_{new}$=median $\{\boldsymbol{\mu}_{j-1}, \boldsymbol{\mu}_j\}$,

           $\Delta\boldsymbol{\mu} = \frac{\tau}{N}\sum_{i=1}^{N}\text{Log}_{\boldsymbol{\mu}_{new}}(\mathbf{x}_i)$,

           $\boldsymbol{\mu}_j = \text{Exp}_{\boldsymbol{\mu}_{new}}(\Delta\boldsymbol{\mu})$,

        and go back to Step 2.

Step 3: If $\parallel \Delta\boldsymbol{\mu} \parallel < \varepsilon$, then $\boldsymbol{\mu}_j$ is optimum and stop the algorithm.

        Else, set $j = j + 1$ and proceed the algorithm from Step 1.

In the next section we implement this algorithm to a real data. We also consider some simulation studies to evaluate variant aspects of the GDA and RGDA methods.

5 Simulation studies and real data analysis

In this section, we compare performance of the GDA and RGDA using some simulation studies as well as real data examples. The relevant parameters in these algorithms are altered in such way that more features of the methods could be taken into account. However, there might be some particular situations in which the new proposed algorithm might fail to converge during the iterations of the algorithm. We provide some ideas on how to tackle those cases in the statistical shape analysis settings.

Our first simulation study is concerned about a simple shape case; triangle shape. Let assume the triangle **B** in Sect. 2 is the intrinsic population mean of 200 random triangles and the aim is to obtain the intrinsic sample mean shape using both the GDA and RGDA procedures. To have shape data for this case, the Cartesian coordinates of the random triangles are simulated using the multivariate normal with the Cartesian coordinate of **B**, stacked in a single column, as the mean vector and $\sum_\mathbf{B} = 10^{-4}\text{diag}\{4, 2, 5, 3, 5, 4\}$, as the covariance matrix of this distribution, where diag is used to represent a diagonal matrix with determined elements. Clearly, at this end, we have 200 matrices with dimensions 3×2. Then, we derive the shape coordinates, say Bookstein coordinates, of these objects. The resulted shape distribution is called the offset normal shape density (see P. 130 on Dryden and Mardia 1998). Since the triangle **B** is the Euclidean mean, we prefer to consider the Frechet mean to evaluate the performance of GDA and RGDA procedures because for large sample size, along with geodesic distance, the Frechet mean is a consistent estimator of the intrinsic population mean (see Bhattacharya and Patrangenaru 2003, 2005). To consider variant aspects of the algorithms, we set the step size (τ) at 0.1, 0.4, 0.7 and 1 and threshold value (ε) at 1e-6, 1e-5 and 1e-3. The parameter φ in the RGDA is set to 3. Then, we record the number of the iterations that the RGDA and GDA methods were repeated before converging to the intrinsic mean shape. Note that the iteration for the RGDA refers to the outer loop, i.e. updating of $\boldsymbol{\mu}_j$ to $\boldsymbol{\mu}_{j+1}$, rather

Table 1 The numbers of the iterations in which the RGDA and GDA were repeated before converging to the intrinsic mean shape

τ	ε		
	1e-6	1e-5	1e-3
0.1	4 (15)	4 (10)	3 (7)
0.4	5 (16)	4 (12)	3 (9)
0.7	6 (18)	5 (15)	4 (10)
1	9 (21)	6 (15)	4 (12)

The original data are some simulated triangles. Those numbers for the GDA are reported inside of the brackets

Table 2 The time (in second) of executing the RGDA and GDA to reach the intrinsic mean shape

Those numbers for the GDA are reported inside of the brackets

τ	ε		
	1e-6	1e-5	1e-3
0.1	2.92 (0.16)	2.97 (0.11)	2.81 (0.10)
0.4	3.12 (0.17)	2.96 (0.13)	2.75 (0.10)
0.7	3.17 (0.20)	3.08 (0.15)	2.93 (0.10)
1	3.60 (0.22)	3.19 (0.16)	2.92 (0.12)

than nested ones. The results are reported in Table 1. The values inside of the brackets are those for the GDA. As can be seen from table, less iterations are taken by the RGDA than the GDA to converge. Also, as expected, for fixed τ, to increase ε leads to less iteration of convergence in both methods to the the intrinsic mean shape. However, nothing can be said for altering τ while ε is fixed. Generally, if one is looking for small iteration, it is recommended to take τ and ε, small and big, respectively.

We also recorded the entire run time (executed on a 3.6 GHz PC) of these two procedures to converge the intrinsic mean. The run time reported here is the CPU time while executing the R code for each of the procedure. Generally, the GDA is faster than RGDA, which might be a drawback of our proposed algorithm. However, we lost the speed on the cost of keeping the geometry and reaching to the real mean as close as possible in implementing the RGDA. This is mainly because of internal loop on which repeats till get the right shape object (Table 2).

In addition to evaluate the numbers of iterations, we are concerned about the geometry of the triangles while algorithms are proceeding for various values of the step size (τ). In particular, we are interested on how far the intrinsic mean shapes are from the true mean (which is considered the sample Frechet mean as discussed earlier) using two procedures. Hence, we set $\varepsilon = $ 6e-3 and τ to 0.1, 0.4, 0.7 and 1. The intrinsic mean shapes derived using both algorithms along with the shape of the Frechet mean as the true mean, are plotted in Fig. 1. Furthermore, we reported d_F^2, and the values of τ in each panel. The triangles with the solid, dotted and dotted-dashed lines are representing the true mean and the intrinsic mean shapes given by the GDA and RGDA
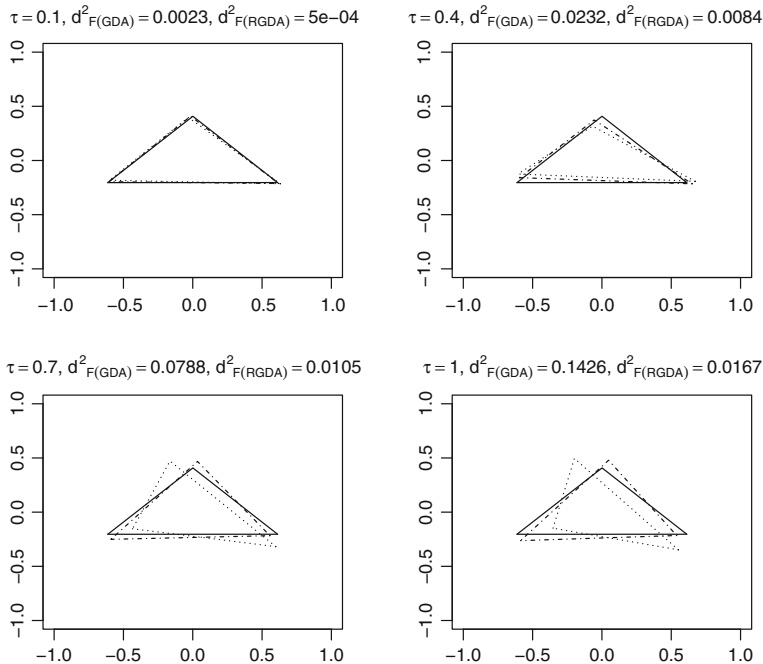
Fig. 1 Impact of τ on obtaining the intrinsic mean shape using both the GDA and RGDA procedures. The *triangles with the solid, dotted and dotted-dashed lines* are representing the true mean and the intrinsic mean shape given by the GDA and RGDA procedures, respectively. Although as τ increases d_F^2 increases in employing both methods, but the RGDA outperforms the GDA. Nonetheless, as τ increases the intrinsic mean shapes get farther from the true mean.

procedures, respectively. As the figure shows, the RGDA outperforms the GDA in most cases in terms of the minimum d_F^2. That means the RGDA is more close to the mean shape than the GDA. Furthermore, if τ increases the intrinsic mean shapes derived using both procedures would be farther from the true mean shape, i.e. the bigger τ the bigger d_F^2 gets. So, one clear message of this simulation is that if the aim is to get the intrinsic mean shape more similar to the mean shape, it is recommended to keep τ small.

We were also interested on studying the combinational affects of τ and ε on the quantity of MOSS while deriving the intrinsic mean shape through employing the RGDA and GDA methods. Based upon the above simulation study and our experiences in other empirical investigations, we set $\varepsilon = 6e - 5$ and allow τ to vary on the interval $(0, 2]$. However, since the plot would be so messy for entire range of τ we divided the output into two single plot based on $\tau \in (0, 1]$ and $\tau \in (1, 2]$. The results were plotted in Fig. 2. As can be seen from this figure, the RGDA has smaller values throughout the interval τ. Particularly, although there is no difference between two approaches for small τ's e.g. say $\tau \in (0, 0.1)$, the MOSS using the GDA procedure gets bigger and bigger for the rest of τ's, indicating overall better performance of the RGDA. As a short result, we can state that one should switch to the RGDA method when the GDA is worse in terms of the minimum MOSS.
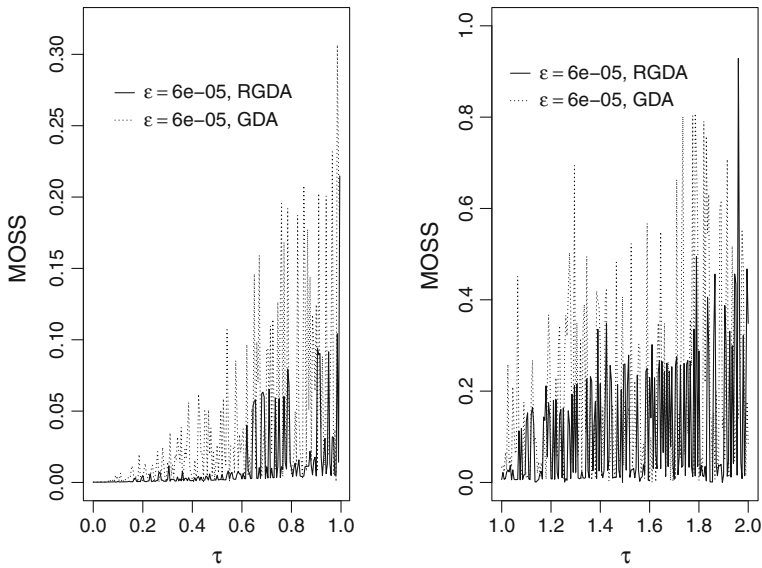
Fig. 2 Combinational affects of ε, and τ on the MOSS quantity to get the intrinsic mean shape using the RGDA and GDA methods. The *solid and dashed lines* are the values of the MOSS quantities in employing, respectively, the RGDA and GDA methods to derive the intrinsic mean shape. Unlike the GDA, the MOSS quantities for the RGDA method is quite small throughout the range of τ. So, it can be said that the RGDA outperforms the GDA.

We further investigated the impact of the φ on performance of the RGDA. So setting $\varepsilon = 6e - 5$, and $\tau = 0.1, 0.4$, we run the RGDA algorithm for simulated triangles when φ varies on the interval (Grenander 1994; Fletcher et al. 2004). Then, the value of MOSS were recorded. Figure 3 shows that there is not a clear pattern for MOSS in varying φ for fixed τ. Nonetheless, for fixed φ, bigger τ leads to smaller MOSS, although there are some exceptions due to random mean generation.

Now, we shall look at some real data sets and attempt to explore variant aspects of our proposed algorithm. Also, our aim is to compare the proposed approach with common GDA.

The first considered data set is the brain shape data. The data were already analyzed by some researchers (see e.g. Free et al. 2001) to investigate other aspects of the shape analysis. The data are saved under the name of `brains` and can be loaded from the library `shapes` in the statistical software (R Development Core Team 2012). In the sense of shape analysis, 24 landmarks are located in 58 adult healthy brains of gorilla. In order to evaluate our methods, we set the step size (τ) equal to 0.1, 0.4, 0.7, 1 and the fixed threshold value (ε) to 0.01, 0.06, 0.1, 0.6. Initially, We would like to derive the intrinsic mean shape using the GDA method and check whether or not the method converges. We have recorded the number of iterations that the algorithm takes in the case of convergence. However, there were some cases in which the approach was never converged. We wrote "NC", standing for Not Converged for these situations. Table 3 shows the results for this scenario. As can be seen, the method will usually converge at the first iteration if ε, and τ get big and small values, respectively. There
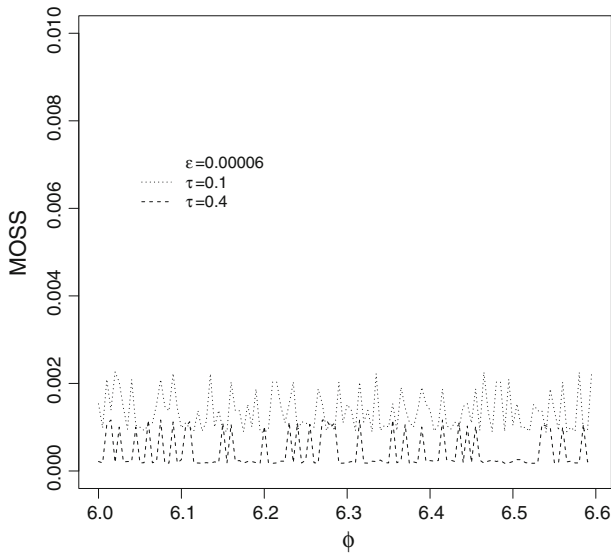
Fig. 3 Affect of φ on performance of the RGDA aiming on a proper calibration. The *dotted and dashed lines* are the values of the MOSS quantities in employing the RGDA method to derive the intrinsic mean shape when τ equals to 0.1 and 0.4, resp. As can be seen, there is not any particular pattern for the value of the MOSS while φ varies and τ is kept fixed.

Table 3 Results of performing the GDA approach for brain data set

	ε			
τ	0.01	0.06	0.1	0.6
0.1	1	1	1	1
0.4	NC	2	1	1
0.7	NC	2	2	2
1	NC	NC	NC	2

The numbers inside of the cells indicate the number of iteration that the GDA proceeds to converge to the intrinsic mean. If the method does not converge, the word "NC" is written

is no guarantee to derive the intrinsic mean shape for small values of ε, while τ is increasing. If one wishes to set ε, as small as possible, which is expected in real applications, it is recommended to keep τ, small too so that the intrinsic mean could be guaranteed by the GDA method.

It might be of great interest to check the geometry for the situations in which the method leads to a solution. Hence, we further compared the intrinsic mean shape returned by the GDA method with the Procrustes mean shape using the d_F^2, quantity, in the cases in which the algorithm converges. Figure 4 shows configurational plot of both the intrinsic and Procrustes mean shapes for the case which ε, was fixed at 0.6 and τ varies as before. We provided the values of d_F^2, as well τ, in each panel. Moreover, the Procrustes and intrinsic mean shapes were indicated by solid circle and
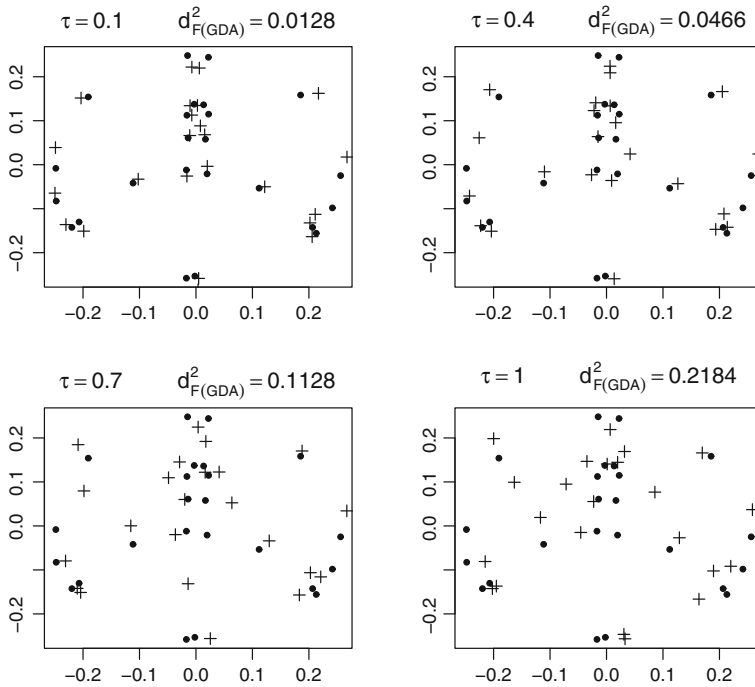
Fig. 4 Impact of τ on performance of the GDA method with ε, set at 0.6. The Cartesian coordinates of the *Procrustes and intrinsic* mean shapes are indicated by *solid circle* and $+$ sign, respectively. It is seen that as τ increases d_F^2 increases too, indicating non-stability of the geometrical structure of the intrinsic mean shape given by the GDA methods.

$+$ sign, respectively. As figure shows, the bigger τ the bigger d_F^2, gets. It means as τ increases, the intrinsic mean shape goes further, in terms of shape distance, from the Procrustes mean shape, and so non-stability occurs for the larger τ, while ε, is fixed at some small values.

We followed the same procedure to evaluate the performance of the RGDA. For ease of repetition, we omit some explanations concerning the values for the fixed parameters. The exception is the parameter for checking the geometry on performing the RGDA, i.e. φ, which is fixed at 3.4. The Table 4 shows the number of iterations that the RGDA takes to converge to the intrinsic mean shape. As can be seen, all combinations of ε, and τ are led to an optimal solution. Hence, unlike the GDA, we get an improvement in terms of converging to the intrinsic mean shape while using the RGDA method.

To check how the geometry is preserved, we plotted the configurational figure of the intrinsic mean shape given by the RGDA method along with the Procrustes mean in Fig. 5. Like the GDA case, as τ increases the value of d_F^2, increases too. Moreover, assuming the Procrustes mean for this data set is well describing the average geometrical feature of all 58 adult healthy brains, the intrinsic mean shape returned by the RGDA method is a better representative of the mean shape than that given by GDA. In other words, from Figs. 4 and 5, the d_F^2, arising from the RGDA is small in

Table 4 Results on performing RGDA for brain data set

	ε			
τ	0.01	0.06	0.1	0.6
0.1	1	1	1	1
0.4	5	2	1	1
0.7	5	2	2	1
1	7	4	5	3

The numbers inside of the cells indicate the number of iterations that the RGDA pass to reach the intrinsic mean shape. It is seen that, unlike GDA, the RGDA method will converge at maximum iteration of 7 for all combinations of τ and ε
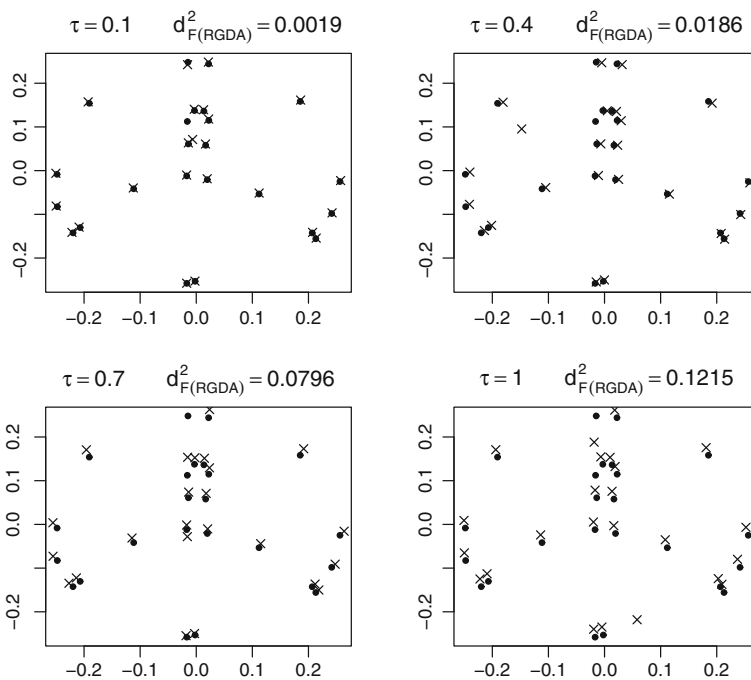

Fig. 5 Impact of τ on performance of the RGDA method with ε, set at 0.6. The Cartesian coordinates of the Procrustes and intrinsic mean shapes are indicated by *solid circle* and $+$ signs, respectively. Although, as τ increases d_F^2, increases too, its increment is less than that of the GDA method.

comparison with that given from the GDA for all τ. So, once again there is a great improvement on employing the RGDA to derive the intrinsic mean shape.

We were also concerned about the performance of the RGDA method in comparison with the GDA based upon the MOSS criterion. So, for ε, fixed at 0.6, while τ varies in the interval (0, 1], we derived the MOSS for the RGDA and GDA methods. The results were plotted in Fig. 6. As expected, to increase τ causes an steady increment on the MOSS quantity resulted from both methods. However, the increment for the RGDA
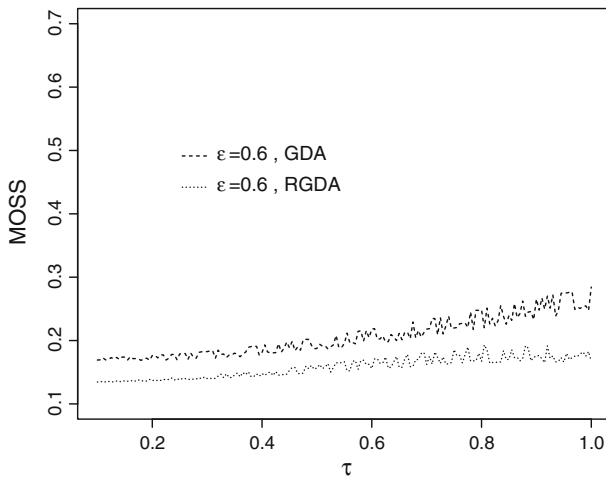
Fig. 6 Impact of ε, on performance of the RGDA and GDA methods with varying τ. *Dashed and dotted lines* are representing the MOSS quantities for the GDA and RGDA methods, respectively. Throughout the range of τ, the RGDA outperforms GDA for $\varepsilon = 0.6$.

method is less in comparison with that of the GDA method throughout the range of τ. We had the same experience for other values of ε, while τ varies, but we confine ourselves to these combinations for ease of repetition. Hence, as a general result in studying the brains data, we recommend to use the RGDA instead of the GDA when the aim is to get the intrinsic mean shape.

Another data set was also considered to evaluate the GDA and RGDA methods. It is a part of a larger study on assessing the effects of selection for body weight on the shape of mouse vertebra using three groups of mice. Further shape analysis of these data is provided in Dryden and Mardia (1998). Here, 30 control group of mouse vertebra with six landmarks in two dimensions is considered. Like our previous example, these data, called `qcet2.dat`, are also available on the library `shapes`. We followed, relatively, the same procedure as in studying the previous data set to compare the GDA and RGDA methods. So, for sake of brevity, we do not explain some unnecessary details. The important issue is on choosing the proper values for parameters τ and ε, which should be done heuristically. Hence, we set these parameters as $\tau = 0.001, 0.004, 0.1, 0.4, 0.7, 1$ and $\varepsilon = $ 1e-6,1e-5,1e-3. Table 5 shows how many iterations two methods passed before converging to the intrinsic mean shape. Note that the values for the GDA methods are inside of the brackets. In almost all cases, the RGDA outperforms the GDA. So, the RGDA is faster than the GDA in converging to the intrinsic mean shape.

We were further studied the impact of parameters ε and τ on performance of both methods based upon the MOSS criterion and using the mouse vertebra data set. Again, we get the same results in terms of better performance of the RGDA method in comparison with the GDA method.

	$\varepsilon/6$		
τ	1e-6	1e-5	1e-3
0.001	4 (18)	1 (13)	1 (2)
0.004	5 (19)	4 (15)	1 (2)
0.1	14 (27)	11 (21)	8 (8)
0.4	20 (36)	13 (30)	6 (9)
0.7	34 (35)	30 (29)	5 (15)
1	30 (33)	19 (31)	12 (17)

Table 5 The numbers of the iterations in which the RGDA and GDA were repeated before converging to the intrinsic mean shape for the mouse vertebra data

Those numbers for the GDA method are written inside of the brackets

6 Conclusion

To derive the statistical measures of centrality and variability is an essential stage of any statistical analysis. When the observations take their values on a linear space, so many procedures have been proposed to derive these quantities. Nowadays, statistics is encountered with data belonging to some non-linear spaces, mainly, because of great progress in technology. Recently, new fields of sciences bring along themselves such data from vast disciplines. Statistical shape analysis is an example of such new subject, which deals with geometrical aspects of the data on hand. In particular, its main interest lies down on whole geometrical information about an object when the similarity effects has already been removed out from the object.

The GDA, as a traditional method, was extended to derive a measure of variability on the manifold valued statistics. However, it suffers from fatal problems when one deals with geometrical observations such as shape data. Particularly, the geometry might not be preserved when the GDA goes forward. In addition, when using this method, the convergence to the intrinsic mean shape might either take a long iterations or even not be guaranteed while improper values for the step size and threshold are considered. We proposed new algorithm to overcome these obstacles in the statistical shape analysis area. Our method, called RGDA, outperforms better in comparison with the GDA in various scenarios. Further, our simulation study indicated that not only the geometry is preserved but also it is robust to possible odd configuration produced at any stage of the algorithm. We further discovered that the GDA might not be able to converge the intrinsic mean while there is not such problems in employing the RGDA method. Both simulation study and real data investigations support setting the small step size and big threshold value if the aim is to derive the intrinsic mean using both methods. However, unlike the GDA, the RGDA guarantees the optimal solution in all combinations of these parameters.

We considered the landmarks-based shape analysis in this paper. Other views to the shape analysis grow up more recently in real life applications. They are set-theory and functional based methods (Stoyan and Stoyan 1994). Hence, employing the GDA and RGDA in the statistical shape analysis from these prospective are interesting topic for future research. In addition, aiming to speed up the algorithms in converging to the intrinsic mean shape, calibrations of the step size and threshold parameters will be

interesting topic to study in future. Moreover, to perform the ANOVA for the shape data, similar to what was done in Figueiredo (2008), using shape variability measures with the roots in conjunction with these methods will be worth investigating.

References

Bershad NJ NJ (1987) On the optimum gain in LMS adaptation. IEEE Trans Acoust Speech Sig Process 35:1065–1068

Bhattacharya R, Patrangenaru V (2003) Large sample theory of intrinsic and extrinsic sample means on manifolds. I. Ann Stat 31:1–29

Bhattacharya R, Patrangenaru V (2005) Large sample theory of intrinsic and extrinsic sample means on manifolds. II. Ann Stat 33:1225–1259

Bookstein FL (1986) Size and shape spaces for landmark data in two dimensions (with discussion). Stat Sci 1:181–242

Bookstein FL (1991) Morphometric tools for landmarks data. Cambridge University Press, Cambridge

Buss SR, Fillmore JP (2001) Spherical average and application to spherical splines and interpolation. ACM Trans Graphics 20:95–126

Cauchy A (1847) Methodes generales pour la resolution des systemes dequations simultanees. Comp Rend Acad Sci Paris 25:536–538

Douglas SC, Pan W (1995) Exact expectation analysis of the LMS adaptive filter. IEEE Trans Sig Process 43:2863–2871

Dryden IL, Kume A, Le HL, Wood ATA, Laughton C (2002) Size-and-shape analysis of DNA molecular dynamics simulations. In: Mardia KV, Aykroyd RG, and McDonnell P (eds) Proceedings in statistics of large datasets. LASR2002, University of Leeds, pp 23–26

Dryden IL, Mardia KV (1998) Statistical shape analysis. Wiley, Chichester

Dryden IL (2011) Shapes package. R Foundation for Statistical Computing, Vienna Contributed package

Figueiredo A (2008) Two-way ANOVA for the watson distribution defined on the hypersphere. Stat Papers 49:363–376

Fletcher PT, Lu C, Joshi S (2003) Statistics of shape via principal geodesic analysis on lie groups. Proceedings of the IEEE conference on computer vision and pattern recognition, pp 95–101

Fletcher PT, Lu C, Pizer SM, Joshi S (2004) Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans Med Imaging 23:995–1005

Fletcher PT, Venkatasubramanian S, Joshi S (2009) The geometric median on Riemannian manifolds with application to robust atlas estimation. Neuroimage 45:143–152

Free SL, O'Higgins P, Maudgil DD, Dryden IL, Lemieux L, Fish DR, Shorvon SD (2001) Lanmark-based morphometrics of the normal adult brain using MRI. Neuroimage 13:801–813

Grenander G (1994) General pattern theory. Clarendon Press, Oxford

Groisser D (2004) Newton's method, zeroes of vector fields, and the Riemannian center of mass. Adv Appl Math 33:95–135

Hendriks H, Landsman Z (1998) Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. J Multivar Anal 67:227–243

Huckemann S, Hotz T (2009) Principal component geodesic for planar shape space. J Multivar Anal 100:699–714

Karcher H (1977) Riemannian center of mass and mollifier smoothing. Commun Pure Appl Math 30: 509–541

Kendall DG (1977) The diffusion of shape. Adv Appl Probab 9:428–430

Kendall DG (1984) Shape manifolds, procrustean metrics and complex projective spaces. Bull Lond Math Soc 16:81–121

Le H (2001) Locating frechet means with application to shape spaces. Adv Appl Probab 33:324–338

Liu J (2001) A novel adaptation scheme in the NLMS algorithm echo cancellation. IEEE Sig Process Lett 8:20–22

Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic Press, London

Matthews VJ, Xie Z (1993) A stochastic gradient adaptive gilter with gradient adaptive step size. IEEE Trans Sig Process 41:2075–2087

Meza JC (2010) Steepest descent. Wiley Interdiscip Rev 2:719–722

Micheas AC, Dey DK (2005) Modelling shape distributions and inferences for assessing difference in shapes. J Multivar Anal 92:257–280

Patrangenaru V (1998) Asymptotic statistics on manifolds. Ph.D. Dissertation, Indiana Univ

Pennec X (1999) Probabilities and statistics on riemannian manifolds: basic tools for geometric measurements. In: Cetin AE, Akarun L, Ertuzun A, Gurcan MN, Yardimci Y (eds) Proceding of nonlinear signal and image processing (NSIP99), 1, 194–198, June 20–23, Antalya, Turkey, IEEE-EURASIP

Pennec X (2006) Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. J Math Imaging Vis 25:127–154

R Development Core Team (2012) R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, 3–900051-07-0. http://www.R-project.org

Sommer S, Lauze F, Hauberg S, Nielson M (2010) Manifold value statistics, exact principal geodesic analysis and the effect of linear approximation. Lect Note Comput Sci 6316:43–56

Stoyan D, Stoyan H (1994) Fractals. Random shapes and point fields. Methods of geometrical statistics. Wiley, Chichester

Wu J, Smith WAP, Hancock ER (2007) Weighted principal geodesic analysis for facial gender classification. Prog Pattern Recognit Image Anal Appl 4756:331–339