# Manifold Valued Statistics, Exact Principal Geodesic Analysis and the Effect of Linear Approximations

Stefan Sommer[1], François Lauze[1], Søren Hauberg[1], and Mads Nielsen[1,2]

[1] Dept. of Computer Science, Univ. of Copenhagen, Denmark
sommer@diku.dk
[2] Nordic Bioscience Imaging, Herlev, Denmark

**Abstract.** Manifolds are widely used to model non-linearity arising in a range of computer vision applications. This paper treats statistics on manifolds and the loss of accuracy occurring when linearizing the manifold prior to performing statistical operations. Using recent advances in manifold computations, we present a comparison between the non-linear analog of Principal Component Analysis, Principal Geodesic Analysis, in its linearized form and its exact counterpart that uses true intrinsic distances. We give examples of datasets for which the linearized version provides good approximations and for which it does not. Indicators for the differences between the two versions are then developed and applied to two examples of manifold valued data: outlines of vertebrae from a study of vertebral fractures and spacial coordinates of human skeleton end-effectors acquired using a stereo camera and tracking software.

**Keywords:** manifolds, Riemannian metrics, linearization, manifold valued statistics, Principal Geodesic Analysis (PGA), Geodesic PCA.

## 1 Introduction

This paper treats the effect of linearization when using the non-linear analog of Principal Component Analysis, Principal Geodesic Analysis (PGA, [1]), to estimate the variability in sets of manifold valued data. Until recently, PGA has been performed by linearizing the manifold, which distorts intrinsic distances, but with the introduction of more powerful computational tools [2], PGA can now be computed with true intrinsic distances. We show how simple and fast indicators allow us to approximate the differences between linearized PGA and exact PGA with true intrinsic distances and evaluate the effect of the linearization.

As a test case for the indicators, we perform a comparison between two manifold valued datasets: outlines of vertebrae from a study of vertebral fractures, and human skeleton end-effectors in spatial coordinates recorded using a stereo camera and tracking software. We will show that linearized PGA provides a reasonable approximation in only one of the experiments and that the indicators allow us to predict this before doing the time-intensive computation of exact PGA with intrinsic distances.

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part VI, LNCS 6316, pp. 43–56, 2010.

### 1.1 Motivation

A wide variety of problems in computer vision possess non-linear structure and are therefore naturally modeled using Riemannian geometry. In diffusion tensor imaging [3,4,5], for image segmentation [6] and registration [7], shape spaces [8], and human motion modeling [9,10], Riemannian manifolds have been used to enforce consistency in data, provide dimensionality reduction, and define more accurate metrics. The wide applicability of manifolds in modeling problems has created the need for statistical tools for manifold data.

Generalizing linear statistical operations to manifolds [1,11,12,13] provides examples of the theoretical and computational problems arising when departing from familiar Euclidean spaces. The tools developed when pursuing this have been used successfully for a range of computer vision applications, and the area is the subject of active research [2,13]. Depending on the level of approximation used in the computations, manifold statistics can be hard to carry out in practice because operations such as finding distances and performing optimization do not admit the closed-form solutions often found in Euclidean spaces [1].

One way of doing manifold statistics is projecting the set of manifold valued data points to the tangent space of a mean point of the manifold. The vector space structure of the tangent space brings back convenient Euclidean statistics, but the distortion of the distances between the data points inherent in the linearization may however lead to sub-optimal solutions to the statistical problems. In contrast to this, some statistical operations can be carried out with true intrinsic manifold distances giving a true picture of the data [2,13]. This, however, often comes at the cost of increased computational complexity and requires conditions on the locality of data.

Because of the trade-offs between convenient linearization and exact modeling, we seek for ways to evaluate the extent of the distortion between the linearized data and true manifold data; we are interested in determining if performing statistics with intrinsic distances offers significant advantages over the linearized approach. Such knowledge has the potential of saving substantial computation time and to improve results of statistical operations.

### 1.2 Related Work

The mathematical aspects of manifolds are covered extensively in the literature with [14,15] providing good references. Numerical and computational aspects of interest in a general setting are considered in the theoretical papers [16,17] while more specific shape related applications are proposed in [18,19,20].

Both the mathematical community, e.g. [11], and more applied fields, computer vision in particular [1,12], have worked with different aspect of statistics on manifolds. A recent wave of interest by statisticians [21,13] has created new methods with strong links to tools developed in computer vision [13].

The manifold generalization of linear PCA, PGA, was first introduced in [22], but it was formulated in the form most widely used in [1]. It has subsequently been used for several applications. To mention a few, the authors in [1,4] study

variations of medial atoms, [23] uses a variation of PGA for facial classification, [24] presents examples on motion capture data, and [20] applies PGA to vertebrae outlines. The algorithm presented in [1] for computing PGA with linearization has been most widely used. In contrast to this, [24] computes PGA as defined in [22] without approximations, but only for a specific manifold, the Lie group SO(3). By using ODE formulations of geodesics and taking derivatives, [2] provides algorithms for computing PGA without approximations on wide classes of manifolds.

Geodesic PCA (GPCA, [13,21]) is in many respects close to PGA but optimizes for the placement of the center point and minimizes projection residuals along geodesics instead of maximizing variance in geodesic subspaces. GPCA uses no linear approximation, but it is currently only computed on spaces where explicit formulas for geodesics exist and on quotients of such spaces.

### 1.3 Content and Outline

In the next section, we discuss the benefits of using manifolds in modeling, manifold valued statistics, and linearization. Then, in section 3, we consider in detail the specific case of Principal Geodesic Analysis and use synthetic examples to explain the differences between linearized PGA and exact PGA with true intrinsic distances. We progress to developing indicators of these differences, and, in section 4, we compare linearized and intrinsic PGA on real-life examples of manifold valued datasets and analyze the power of the indicators. The paper thus contributes by

(1) developing simple and fast indicators of the difference between linearized PGA and exact PGA that show the effect of linearization,
(2) giving examples of the differences between linearized PGA and exact PGA on real-life datasets from computer vision,
(3) and showing the power of the indicators when applied to the datasets.

## 2 Manifolds and Manifold Valued Statistics

The interest in manifolds as modeling tools arises from the non-linearity apparent in a variety of problems. We will in the following exemplify this by considering the pose of a human skeleton captured by e.g. a tracking system or motion capture equipment. Consider the position of a moving hand while the elbow and the rest of the body stay fixed. The hand cannot move freely as the length of the lower arm restricts it movement. Linear vector space structure is not present; if we multiply the position of the hand by a scalar, the length of the arm would in general change in order to accommodate the new hand position. Even switching to an angular representation of the pose of the elbow joint will not help; angles have inherent periodicity, which is not compatible with vector space structure.

Though the space of possible hand positions is not linear, it has the structure of a manifold since it possesses the property that it locally can be approximated

by a vector space. Furthermore, we can, in a natural way, equip it with a Riemannian metric [14], which allows us to make precise notions of length of curves on the space and intrinsic acceleration. This in turns defines the Riemannian manifold equivalent of straight lines: geodesics. The length of geodesics connecting points defines a distance metric on the manifold.

### 2.1 Benefits from Modeling Using Manifolds

The main advantages of introducing manifolds in modeling are as follows: consistency in representation, dimensionality reduction, and accuracy in measurements. Consistency ensures the modeled object satisfies the requirements making up the manifold; when moving the position of the hand on the manifold, we are certain the length of the lower arm is kept constant. Such requirements reduce the number of degrees of freedom and hence provide dimensionality reduction. Consistency and dimensionality reduction are therefore closely linked.

Accuracy is connected to the distance measure defined by the Riemannian metric. A reasonable measure of the distance between two positions of the hand will be the length of the shortest curve arising when moving the hand between the positions. Such a curve will, in this example, be a circular arc, and, in the manifold model, the distance will be the length of the arc. In the vector space model, however, the distance will be the length of the straight line connecting the hand positions and, hence, will not reflect the length of an allowed movement of the hand. The manifold model therefore gives a more accurate distance measure.

### 2.2 Linearizing the Manifold

By linearizing the manifold to the tangent space of a mean point, we can in many applications ensure consistency, but not accuracy, in statistical operations. Let $M$ be a manifold and $\{x_1, \ldots, x_N\}$ a dataset consisting of points on the manifold. An intrinsic mean [11] is defined as a solution to the optimization problem

$$\mu = \operatorname{argmin}_q \sum_{i=1}^{N} d(x_i, q)^2 \tag{1}$$

with $d(x_i, q)$ denoting the manifold distance between the $i$th data point and the mean candidate $q$.

Each point $p$ of a manifold has a connected linear space called the tangent space and denoted $T_pM$. The dimension of $T_pM$ is equal to the dimension of the manifold, which, as in the vector space case, specifies the number of degrees of freedom. Vectors in the tangent space are often mapped back to the manifold using the exponential map, $\mathrm{Exp}_p$, which maps straight lines trough the origin of $T_pM$ to geodesics on $M$ passing $p$.

If we consider the tangent space of an intrinsic mean, $T_\mu M$, we can represent $x_i$ by vectors $w_i$ in $T_\mu M$ such that $\mathrm{Exp}_\mu w_i = x_i$.[1] The map that sends $x_i \in M$

[1] See Figure 1 for an example of a 2-dimensional manifold with sampled elements of the tangent space of the mean and corresponding points on the manifold.

to $w_i \in T_\mu M$ is called the logarithm map and denoted $\mathrm{Log}_\mu$. The vector space structure of $T_\mu M$ allows us to use standard statistical tools on $\{w_1, \ldots, w_N\}$. We could for example infer some distribution in $T_\mu M$, sample a vector $v$ from it, and project the result back to a point $p$ on the manifold so that $p = \mathrm{Exp}_\mu v$. It is important to note that consistency is ensured in doing this; $p$ will be on the manifold and hence satisfy the encoded requirements. Turning to the example of hand positions, we have found a consistent way of sampling hand positions without violating the fixed length of the lower arm.

The above procedure can be seen as a way of linearizing the manifold around the intrinsic mean $\mu$ because the tangent space $T_\mu M$ provides a first order approximation of the manifold around $\mu$. Yet, distances between vectors in $T_\mu M$ do not always reflect the manifold distances between the corresponding points on the manifold: distances between $w_i$ and the origin of $T_\mu M$ equal the distances $d(x_i, \mu)$, but the inter-point distances $d(x_i, x_j)$ are not in general equal to the tangent space distances$\|w_i - w_j\|$. Accuracy may therefore be lost as a result of the approximation. In short, linearization preserves consistency but may destroy accuracy.

## 3 Principal Geodesic Analysis

Principal Component Analysis (PCA) is widely used to model the variability of datasets of vector space valued data and provide linear dimensionality reduction. PCA gives a sequence of linear subspaces maximizing the variance of the projection of the data or, equivalently, minimizing the reconstruction errors. The $k$th subspace is spanned by an orthogonal basis $\{v^1, \ldots, v^k\}$ of principal components $v^i$.

PCA is dependent on the vector space structure and hence cannot be performed on manifold valued datasets. Principal Geodesic Analysis was developed to overcome this limitation. PGA centers its operations at a point $\mu \in M$ with $\mu$ usually being an intrinsic mean of the dataset $\{x_1, \ldots, x_N\}$, and finds geodesic subspaces, which are images $S = \mathrm{Exp}_\mu V$ of linear subspaces $V$ of the tangent space $T_\mu M$. A projection operator $\pi_S$ is defined by letting $\pi_S(x)$ be a point in $S$ closest to $x$. The $k$th geodesic subspace $S_k$ is then given as $\mathrm{Exp}_\mu(V_k)$, $V_k = \mathrm{span}\,\{v^1, \ldots, v^k\}$, where the principal directions $v^i$ are given recursively by

$$\begin{aligned} v^i &= \mathrm{argmax}_{\|v\|=1, v \in V_{i-1}^\perp} \frac{1}{N} \sum_{j=1}^{N} d(\mu, \pi_{S_v}(x_j))^2 \ , \\ S_v &= \mathrm{Exp}_\mu(\mathrm{span}\,(V_{i-1}, v)) \ . \end{aligned} \tag{2}$$

The term being maximized is the sample variance, the expected value of the squared distance to $\mu$. PGA therefore extends PCA by finding geodesic subspaces in which variance is maximized.

Since the projection $\pi_{S_k}(x)$ is hard to compute, PGA is traditionally approximated by linearizing the manifold. The data $x_1, \ldots, x_N$ are projected to $T_\mu M$ using $\mathrm{Log}_\mu$, and regular PCA is performed on $w_i = \mathrm{Log}_\mu x_i$. Equation (2) then becomes

$$v^i \approx \text{argmax}_{\|v\|=1, v \in V_{i-1}^{\perp}} \frac{1}{N} \sum_{j=1}^{N} \left( \langle w_j, v \rangle^2 + \sum_{l=1}^{k-1} \langle w_j, v^l \rangle^2 \right) . \tag{3}$$

We can define a normal distribution $\mathcal{N}$ in $T_\mu M$ using the result of the PCA procedure, and, in doing so, we have performed the procedure described in section 2.2. We will refer to PGA with the approximation as *linearized* PGA. PGA as defined by (2) without the approximation will be referred to as *exact* PGA. Advances in manifold computations allow exact PGA to be computed on the Lie group SO(3) [24] and, more recently, on wide classes of manifolds [2].

Replacing maximization of the sample variances $d(\mu, \pi_{S_v}(x_j))^2$ by minimization of the squared reconstruction errors $d(x_j, \pi_{S_v}(x_j))^2$, we obtain another manifold extension of PCA and thus an alternate definition of PGA:

$$v^i = \text{argmin}_{\|v\|=1, v \in V_{i-1}^{\perp}} \frac{1}{N} \sum_{j=1}^{N} d(x_j, \pi_{S_v}(x_j))^2 . \tag{4}$$

In contrast to vector space PCA, the two definitions are not equivalent. It can be shown that, in some cases, solutions to (2) will approach parts of the manifold where the cost function is non differentiable, a problem we have not encountered when solving for (4). We are currently working on a paper giving a theoretical treatment of this phenomenon and other differences between the definitions. The latter formulation is chosen for Geodesic PCA to avoid similar instabilities of variance maximization [13]. In correspondence with this, we will use (4) in the rest of the paper, but we stress that this choice is made only to avoid instabilities in (2) and that all computations presented can be performed using the former definition with only minor changes to the optimization algorithms [2].

### 3.1 Linearized PGA vs. Exact PGA

Computing the projection map $\pi_S$ is particularly time-intensive causing the computation of exact PGA to last substantially longer than linearized PGA. To give an example, computing linearized PGA for one of the datasets later in this paper takes 5 seconds with a parallelized Matlab implementation, and computing exact PGA for the same example requires approximately 10 minutes. This time penalty makes it is worth considering the actual gain of computing exact PGA. We will in this section give examples of low dimensional manifolds on which it is possible visually to identify the differences between the methods.

We consider surfaces embedded in $\mathbb{R}^3$ and defined by the equation

$$S_c = \{(x, y, z) | cx^2 + y^2 + z^2 = 1\} \tag{5}$$

for different values of the scalar $c$. For $c > 0$, $S_c$ is an ellipsoid and equal to the sphere $\mathbb{S}^2$ in the case $c = 1$. The surface $S_0$ is a cylinder and, for $c < 0$, $S_c$ is an hyperboloid. Consider the point $p = (0, 0, 1)$ and note that $p \in S_c$ for all $c$. The curvature of $S_c$ at $p$ is equal to $c$. Note that in particular for the cylinder case

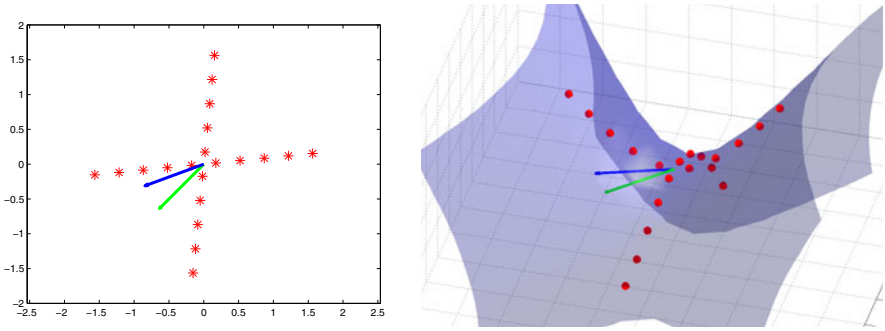


**Fig. 1.** $T_pS_{-2}$ with sampled points and first principal components (blue exact PGA, green linearized PGA) (left) and $S_{-2}$ with projected points and first principal components (blue exact PGA (2), green linearized PGA) (right)

the curvature is zero; the cylinder locally has the geometry of the plane $\mathbb{R}^2$ even though it informally seems to curve.

We evenly distribute 20 points along two straight lines through the origin of the tangent space $T_pS_c$, project the points from $T_pS_c$ to the surface $S_c$, and perform linearized and exact PGA. Since linearized PCA amounts to Euclidean PCA in $T_pS_c$, the first principal component divides the angle between the lines for all $c$. In contrast to this, the corresponding residuals and the first principal component found using exact PGA are dependent on $c$. Table 1 shows the angle between the principal components found using the different methods, the average squared residuals and differences between squared residuals for different values of $c$. Let us give a brief explanation of the result. The symmetry of the sphere and the dataset causes the effect of curvature to even out in the spherical case $S_1$. The cylinder $S_0$ has local geometry equal to $\mathbb{R}^2$ which causes the equality between the methods in the $c = 0$ case. The hyperboloids with $c < 0$ are non-symmetric causing a decrease in residuals as the first principal component approaches the hyperbolic axis. This effect increases with curvature causing the the first principal component to align with this axis for large negative values of $c$.

It is tempting to think that increasing absolute curvature causes increasing differences between the methods. Yet, redoing the experiment with the lines rotated by $\pi/4$ making them symmetric around the $x$ and $y$ axes will produce vanishing differences. Curvature in itself, therefore, does not necessarily imply

**Table 1.** Differences between methods for selected values of $c$

| **c:** | **1** | **0.5** | **0** | **-0.5** | **-1** | **-1.5** | **-2** | **-3** | **-4** | **-5** |
|---|---|---|---|---|---|---|---|---|---|---|
| angle (°): | 0.0 | 0.1 | 0.0 | 3.4 | 14.9 | 22.2 | 24.8 | 27.2 | 28.3 | 28.8 |
| lin. sq. res.: | 0.251 | 0.315 | 0.405 | 0.458 | 0.489 | 0.508 | 0.520 | 0.534 | 0.539 | 0.541 |
| exact sq. res.: | 0.251 | 0.315 | 0.405 | 0.458 | 0.478 | 0.482 | 0.485 | 0.489 | 0.491 | 0.492 |
| diff (%): | 0.0 | 0.0 | 0.0 | 0.1 | 2.3 | 5.1 | 6.7 | 8.4 | 8.9 | 9.0 |

large differences, and the actual differences are hence dependent on both curvature and the dataset.

### 3.2 The Difference Indicators

The projection $\pi_S$ is in (3) approximated using the orthogonal projection in the tangent space $T_\mu M$. We let $\tau_S$ denote the difference in residuals arising when using the two projections and aim at approximating $\tau_S$ to give an estimate of the gain in precision obtained by using true projections. The subspaces optimizing (4) and (3) will in general differ due to the different projection methods and the fact that residuals are approximated by tangent space distances in (3). We let $\rho$ denote the difference in residuals between the projection of the data to the two subspaces, and we aim at approximating $\rho$ to indicate the gain in accuracy when computing exact PGA.

We start by giving precise definitions for $\tau_S$ and $\rho$ before deriving the indicators $\tilde{\tau}_S$ and $\sigma$ of their values. The term indicators is used to emphasize expected correlation between the values of e.g. $\tau_S$ and the indicator $\tilde{\tau}_S$ but with no direct expression for the correlation.

Assume $v_1, \ldots, v_{k-1}$ are principal components and let $v \in T_\mu M$ be such that $v_1, \ldots, v_{k-1}, v$ constitues an orthonormal basis. Let the geodesic subspace $S_v$ be given by $\mathrm{Exp}_\mu \mathrm{span}\{v_1, \ldots, v_{k-1}, v\}$, and let $w_j = \mathrm{Log}_\mu x_j$ for each element of the dataset $\{x_1, \ldots, x_N\}$. We denote by $\hat{\pi}_{S_v}(x_j)$ the point on the manifold corresponding to the orthogonal tangent space projection of $w_j$, i.e.

$$\hat{\pi}_S(x_j) = \mathrm{Exp}_\mu \left( \langle w_j, v \rangle \, v + \sum_{l=1}^{k-1} \langle w_j, v^l \rangle \, v^l \right) , \tag{6}$$

and define the average projection difference

$$\tau_S = \frac{1}{N} \sum_{j=1}^{N} \left( d(x_j, \hat{\pi}_{S_v}(x_j))^2 - d(x_j, \pi_{S_v}(x_j))^2 \right) . \tag{7}$$

Let now $v$ be an exact PGA principal geodesic component computed using (4) and let $\hat{v}$ be a linearized PGA principal component computed using (3). We let $S_v$ and $S_{\hat{v}}$ denote the geodesic subspaces corresponding to $v$ and $\hat{v}$. The average residual difference is then given by

$$\rho = \frac{1}{N} \sum_{j=1}^{N} \left( d(x_j, \pi_{S_{\hat{v}}}(x_j))^2 - d(x_j, \pi_{S_v}(x_j))^2 \right) . \tag{8}$$

Note that both $\tau_S$ and $\rho$ are positive since $\pi_{S_v}$ minimizes residuals and $v$ minimizes (4).

### 3.3 The Projection Difference

Since $\pi_{S_v}(x_j)$ is the point in $S_v$ closest to $x_j$, the differences expressed in each term of (7) measure the difference between $f(\hat{\pi}_{S_v}(x_j))$ and $f(y_j)$ with $y_j \in S_v$

minimizing the map $f(y) = d(x_j, y)^2$. The gradient $\nabla_y f$ vanishes in such a minimum leading us to approximate the difference by the norm of the gradient at $\hat{\pi}_{S_v}(x_j)$. The gradient is readily evaluated since it is given by the component of $-2\mathrm{Log}_{\hat{\pi}_{S_v}(x_j)}(x_j)$ in the tangent space of $S_v$ [11]. We use this to approximate $\tau_S$ by

$$\tau_{S_v} \approx \tilde{\tau}_{S_v} = \frac{2}{N} \sum_{j=1}^{N} \|\nabla_{\hat{\pi}_{S_v}(x_j)} f\| \tag{9}$$

and note that each term of the sum, and therefore the entire indicator $\tilde{\tau}_{S_v}$, is inexpensive to compute.

### 3.4 The Residual Difference

We now heuristically derive an indicator $\sigma$ that is correlated with $\rho$. The correlation will be confirmed later by the experiments. Assume for a moment that distances in the tangent space $T_\mu M$ approximate the true manifold distances well. The residual sums $\frac{1}{N}\sum_{j=1}^{N} d(x_j, \pi_{S_{\hat{v}}}(x_j))^2$ and $\frac{1}{N}\sum_{j=1}^{N} d(x_j, \pi_{S_v}(x_j))^2$ will then be close to identical since $v$ is chosen to minimize the latter sum, and $\hat{v}$ is chosen to minimize the sum of tangent space residuals. The difference $\rho$ will therefore be close to zero. Conversely, assume that distances in the tangent space differ greatly from the true manifold distances. On constant curvature spaces like the sphere $S_1$, these distance differences will generally be uniformly distributed causing the linearized principal component $\hat{v}$ to be close to $v$ and $\rho$ therefore close to zero. On the contrary, the distance differences will vary on spaces with non-constant curvature like $S_{-1}$ where $\hat{v}$ in general is far from $v$ causing $\rho$ to be large. We therefore expect $\rho$ to be correlated with the standard deviation $\sigma$ of the differences between the tangent space residual approximations and the actual orthogonal projection residuals,

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left( \|w_j - \mathrm{Log}_\mu(\hat{\pi}_{S_{\hat{v}}})\| - d(x_j, \hat{\pi}_{S_{\hat{v}}}(x_j)) - \mu \right)^2}\,, \tag{10}$$

with $\mu$ the mean value of the scalars $\|w_j - \mathrm{Log}_\mu(\hat{\pi}_{S_{\hat{v}}})\| - d(x_j, \hat{\pi}_{S_{\hat{v}}}(x_j))$. We use $\sigma$, which again is fast to compute, to indicate the size of $\rho$.

## 4 Experiments

We present experiments on the synthetic data of section 3.1 and on two real-life datasets for two purposes: the experiments will show examples where computing exact PGA results in increased accuracy as well as examples where linearized PGA performs well, and the power of the indicators developed in section 3 will be explored.

When investigating the correlation between the indicator $\tilde{\tau}_{S_{\hat{v}}}$ and the projection difference $\tau_{S_{\hat{v}}}$, we let $\hat{v}$ be the first principal component computed using linearized PGA. In addition, we compare the residual difference $\rho$ with the indicator $\sigma$.

### 4.1 Synthetic Data

We test the indicators on the manifolds $S_c$ with the synthetic data described in section 3.1. Figure 2 shows $\tau_S$ as a function of the indicator $\tilde{\tau}_{S_{\hat{v}}}$ and $\rho$ as a function of the indicator $\sigma$ for each value of $c$. For both graphs, we see correlation between the indicators and actual differences. For $c = 1$ and $c = 0.5$, $\sigma$ is relatively high compared to $\rho$ stressing that the indicators only give approximations and that, if full precision is required, exact PGA should be computed.

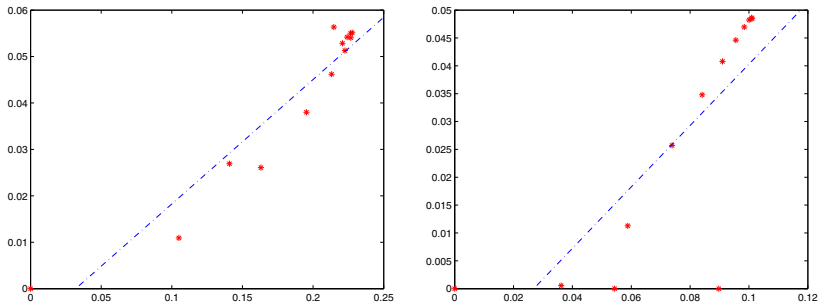


**Fig. 2.** Synthethic data: Projection difference $\tau_{S_{\hat{v}}}$ as a function of the indicator $\tilde{\tau}_{S_{\hat{v}}}$ with the broken line fitted to the points (left) and residual difference $\rho$ as a function of the indicator $\sigma$ with the broken line fitted to the points (right)

### 4.2 Vertebrae Outlines

In this experiment, we consider outlines of vertebrae obtained in a study of vertebral fractures. The dataset of 36 lateral X-rays have been manually annotated by medical experts to identify the outline of the vertebra of each image. To remove variability in the number and placement of points, a resampling is performed to ensured constant inter-point distances. With this equidistance property in mind, the authors in [20] define a submanifold of $\mathbb{R}^{2n}$ on which the outlines naturally reside. We give a brief review of the setup but refer to the paper for details. The equidistance constraint is encoded using a map $F : \mathbb{R}^{2n} \to \mathbb{R}^{n-2}$ with components

$$F^i(P_1, ..., P_n) = d_{i+2,i+1} - d_{i+1,i}, \quad i = 1, .., n-2 \tag{11}$$

with $n$ the number of points and $d_{i,j} = (x_i - x_j)^2 + (y_i - y_j)^2$ the squared distances between points $P_i$ and $P_j$. The constraint is satisfied for a vertebra outline $c = \{P_1, \ldots, P_n\}$ if $F(c) = 0$. An additional constraint is added to remove scaling effects by ensuring the outline reside on the unit sphere. The preimage $A_n = F^{-1}(0)$ is then a submanifold of $\mathbb{R}^{2n}$, the space of equidistant vertebra outlines. We choose 8 random outlines from the dataset and perform linearized PGA and exact PGA. The experiment consists of 20 such selections,

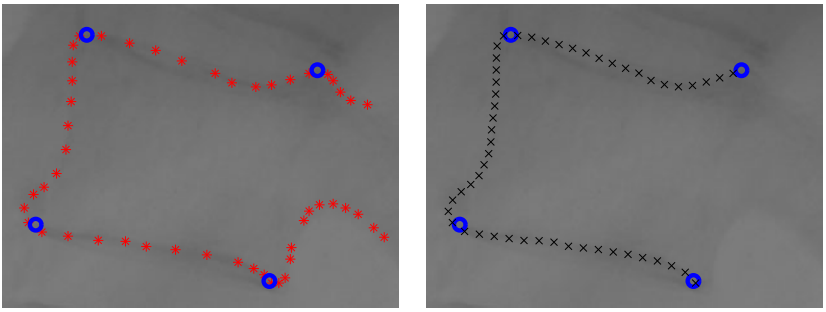

**Fig. 3.** Manually annotated vertebrae outline (left) and resampled outline (right)

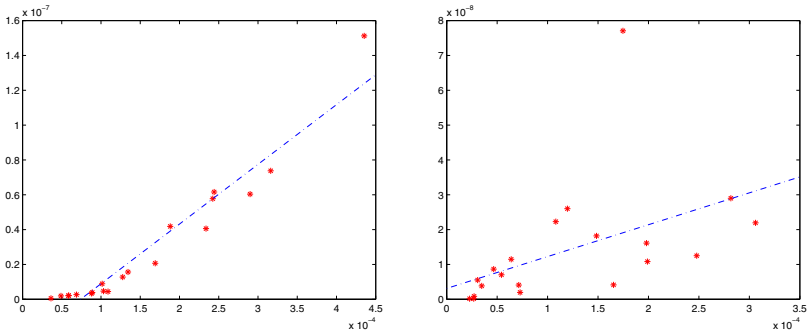


**Fig. 4.** Vertebrae outlines: Projection difference $\tau_{S_{\hat{v}}}$ as a function of the indicator $\tilde{\tau}_{S_{\hat{v}}}$ (left) and residual difference $\rho$ as a function of the indicator $\sigma$ (right)

and, for each selection, the entities $\tau_{S_{\hat{v}}}$, $\tilde{\tau}_{S_{\hat{v}}}$, $\rho$ and $\sigma$ are computed and plotted in Figure 4. Though we visually see correlation between the indicators and their respective associated values in the figures, not only are the correlations low, as the indicators and their values have significantly different orders of magnitude, but in reality, both the indicators and the associated values are in the order of the computation tolerance, i.e close to zero from a numerical point of view. As small indicators should imply small values, we can conclude that the indicators works as required and that, for the example of vertebra outlines, doing statistics on the manifold $A_n$ is helpful in keeping the data consistent, i.e. the equidistance constraint satisfied, but provides little added accuracy.

### 4.3 Human Poses

In this experiment, we consider human poses obtained using tracking software. A consumer stereo camera[2] is placed in front of a test person, and the tracking

[2] `http://www.ptgrey.com/products/bumblebee2/`

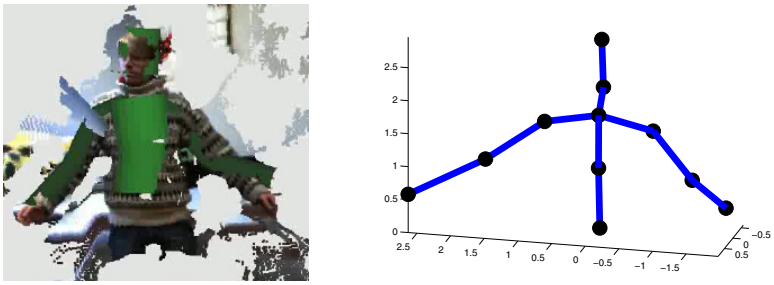


**Fig. 5.** Camera output superimposed with tracking result (left) and a tracked pose with 11 end-effectors marked by thick dots (right)

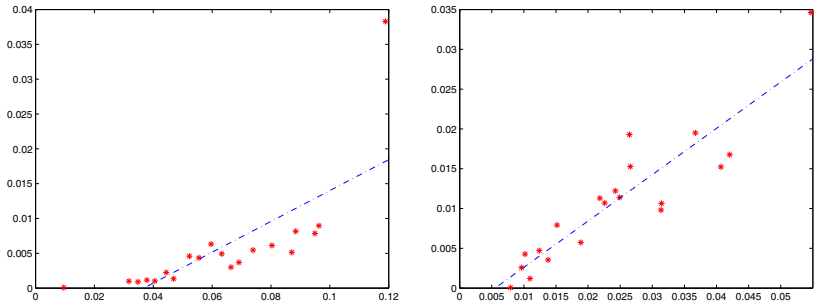


**Fig. 6.** Human poses: Projection difference $\tau_{S_{\hat{v}}}$ as a function of the indicator $\tilde{\tau}_{S_{\hat{v}}}$ (left) and residual difference $\rho$ as a function of the indicator $\sigma$ (right)

software described in [10] is invoked in order to track the pose of the persons upper body. The recorded poses are represented by the human body end-effectors; the end-points of each bone of the skeleton. The placement of each end-effector is given spatial coordinates so that an entire pose with $k$ end-effectors can be considered a point in $\mathbb{R}^{3k}$. To simplify the representation, only the end-effectors of a subset of the skeleton are included, and, when two bones meet at a joint, their end-points are considered one end-effector. Figure 5 shows a human pose with 11 end-effectors marked by thick dots.

The fact that bones do not change length in short time spans gives rise to a constraint for each bone; the distance between the pair of end-effectors must be constant. We incorporate this into a pose model with $b$ bones by restricting the allowed poses to the preimage $F^{-1}(0)$ of the map $F : \mathbb{R}^{3k} \to \mathbb{R}^b$ given by

$$F^i(x) = \|e_{i_1} - e_{i_2}\|^2 - l_i^2 \ , \tag{12}$$

where $e_{i_1}$ and $e_{i_2}$ denote the spatial coordinates of the end-effectors and $l_i$ the constant length of the $i$th bone. In this way, the set of allowed poses constitute a $3k - b$-dimensional implicitly represented manifold.

We record 26 poses using the tracking setup, and, amongst those, we make 20 random choices of 8 poses and perform linearized PGA and exact PGA. For each experiment, $\tau_{S_{\hat{v}}}$, $\tilde{\tau}_{S_{\hat{v}}}$, $\rho$, and $\sigma$ are computed and plotted in Figure 6. The indicators provide a good picture of the projection and residual differences, which are significantly greater than for the vertebra experiment. The indicators and the corresponding true values are now at the same order of magnitude, and the correlation between the indicators and the values they correspond to is therefore significant. The maximal increase in average squared residuals is 1.53 percent with individual squared point residuals changing up to 30.7 percent.

## 5 Conclusion

In this paper, we have explored the differences between exact PGA and its widely used simplification, linearized PGA. We have developed simple indicators of the loss of accuracy when using the linearized PGA instead of exact PGA. As shown on real-life examples of manifold valued datasets, these indicators provide meaningful insight into the accuracy of the linearized method. The experiments, in addition, show that linearization is in some cases a good and fast approximation, but exact PGA offers better accuracy for other applications.

We are currently working on deriving formal arguments for the correlation between $\sigma$ and $\rho$. In the future, we plan to apply the developed indicators to the many uses of PGA, which have previously been computed using the linearized approach, to test whether exact PGA can provide significant increases in accuracy and hence more precise modeling. In order to make better decisions on whether to use linearized or exact PGA, it will be useful to find thresholds for the values of $\tilde{\tau}_{S_{\hat{v}}}$ and $\sigma$ dependent on the sought for precision. Future research will hopefully lead to such thresholds.

## References

1. Fletcher, P., Lu, C., Pizer, S., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging 23, 995–1005 (2004)
2. Sommer, S., Lauze, F., Nielsen, M.: The differential of the exponential map, jacobi fields, and exact principal geodesic analysis (2010) (submitted)
3. Fletcher, P.T., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Processing 87, 250–262 (2007)
4. Fletcher, P.T., Joshi, S.: Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In: Sonka, M., Kakadiaris, I.A., Kybic, J. (eds.) CVAMIA/MMBIA 2004. LNCS, vol. 3117, pp. 87–98. Springer, Heidelberg (2004)
5. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. Int. J. Comput. Vision 66, 41–66 (2006)
6. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision 22, 61–79 (1995)

7. Pennec, X., Guttmann, C., Thirion, J.: Feature-based registration of medical images: Estimation and validation of the pose accuracy. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 1107–1114. Springer, Heidelberg (1998)
8. Kendall, D.G.: Shape manifolds, procrustean metrics, and complex projective spaces. Bull. London Math. Soc. 16, 81–121 (1984)
9. Sminchisescu, C., Jepson, A.: Generative modeling for continuous Non-Linearly embedded visual inference. In: ICML, pp. 759–766 (2004)
10. Hauberg, S., Sommer, S., Pedersen, K.S.: Gaussian-like spatial priors for articulated tracking. In: Daniilidis, K. (ed.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 425–437. Springer, Heidelberg (2010)
11. Karcher, H.: Riemannian center of mass and mollifier smoothing. Communications on Pure and Applied Mathematics 30, 509–541 (1977)
12. Pennec, X.: Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. J. Math. Imaging Vis. 25, 127–154 (2006)
13. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic PCA for riemannian manifolds modulo isometric lie group actions. Statistica Sinica 20, 1–100 (2010)
14. do Carmo, M.P.: Riemannian geometry. Mathematics: Theory & Applications. Birkhauser, Boston (1992)
15. Lee, J.M.: Riemannian manifolds. Graduate Texts in Mathematics, vol. 176. Springer, New York (1997); An introduction to curvature
16. Dedieu, J., Nowicki, D.: Symplectic methods for the approximation of the exponential map and the newton iteration on riemannian submanifolds. Journal of Complexity 21, 487–501 (2005)
17. Noakes, L.: A global algorithm for geodesics. Journal of the Australian Mathematical Society 64, 37–50 (1998)
18. Klassen, E., Srivastava, A.: Geodesics between 3D closed curves using Path-Straightening. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 95–106. Springer, Heidelberg (2006)
19. Schmidt, F., Clausen, M., Cremers, D.: Shape matching by variational computation of geodesics on a manifold. In: Pattern Recognition, pp. 142–151. Springer, Berlin (2006)
20. Sommer, S., Tatu, A., Chen, C., Jørgensen, D., de Bruijne, M., Loog, M., Nielsen, M., Lauze, F.: Bicycle chain shape models. In: MMBIA/CVPR 2009, pp. 157–163 (2009)
21. Huckemann, S., Ziezold, H.: Principal component analysis for riemannian manifolds, with an application to triangular shape spaces. Advances in Applied Probability 38, 299–319 (2006)
22. Fletcher, P., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on lie groups. In: CVPR 2003, vol. 1, p. I-95 – I-101 (2003)
23. Wu, J., Smith, W., Hancock, E.: Weighted principal geodesic analysis for facial gender classification. In: Progress in Pattern Recognition, Image Analysis and Applications, pp. 331–339. Springer, Berlin (2008)
24. Said, S., Courty, N., Bihan, N.L., Sangwine, S.: Exact principal geodesic analysis for data on so(3). In: EUSIPCO 2007 (2007)