Exact Principal Geodesic Analysis for Data on SO(3)

Salem Said, Nicolas Courty, Nicolas Le Bihan, Stephen J. Sangwine

EXACT PRINCIPAL GEODESIC ANALYSIS FOR DATA ON $SO(3)$

Salem SAID(1)*, Nicolas COURTY*(2)*, Nicolas LE BIHAN*(1) *and Stephen J. SANGWINE*(3)

(1): GIPSA-Lab, Dept. Images and Signal, Grenoble, France, (2): VALORIA, Univ. de Bretagne Sud, Vannes, France
(3): Dept. of Electronic Systems Engineering, Univ. of Essex, Colchester, United Kingdom
Salem.said@gipsa-lab.inpg.fr, Nicolas.Courty@univ-ubs.fr,
Nicolas.le-bihan@gipsa-lab.inpg.fr, S.sangwine@ieee.org

ABSTRACT

PGA, or Principal Geodesic Analysis, is an extension of the classical PCA (Principal Component Analysis) to the case of data taking values on a Riemannian manifold. In this paper a new and original algorithm, for the exact computation of the PGA of data on the rotation group $SO(3)$, is presented. Some properties of this algorithm are illustrated, with tests on simulated and real data, and its possible applications are finally discussed.

1. INTRODUCTION

PGA, or Principal Geodesic Analysis, is a recently proposed statistical framework [1], mainly aimed at analyzing the variability of data on a Riemannian manifold. It is an extension of the classical PCA [2] which performs the same task for data in a Euclidean vector space, but fails to do so when the data belong to a more general Riemannian manifold [3].

PGA has found applications in diffusion tensor magnetic resonance imaging (DT-MRI) [3] and shape analysis [4]. In this paper, we consider the PGA of data belonging to the rotation group $SO(3)$ [5]. Such data appear in the study of plate tectonics, continuum modelling of DNA chains [6], robotics [7] and computer graphics [8, 9]. PGA can be used to analyze the variability of data on $SO(3)$, giving a better understanding of their structure. In particular, we will see in this paper that PGA can be used to reduce the dimensionality of data on $SO(3)$. Analyzing the variability of data is an interesting starting point for compression, feature extraction and segmentation applications. In the case of data in a vector space, PCA is widely used for such applications. PGA can be a basis for similar applications for manifold-valued data. We hope to investigate such applications in the field of computer graphics in future works.

All existing algorithms for PGA [1, 3, 4] are based on linear approximations. The main contribution of this paper is a new algorithm that performs *exact* PGA for data on $SO(3)$. We believe this is important for a better understanding of this new framework.

The outline of this paper is the following: In Section 2, unit quaternions are introduced, as a tool for describing rotations, and necessary pre-requisits on their geometry are given. In Section 3, PGA is introduced and our algorithm for computing exact PGA is described. In Section 4 the convergence of our algorithm is illustrated with examples from simulations and from real data. In particular, we consider the PGA of a simulated Gaussian distribution on $SO(3)$ [10] and also of real data obtained from motion capture. In Section 5, we conclude on the possible applications of PGA to the context of signal processing.

2. THEORETICAL BACKGROUND

There are many ways of representing rotations [5, 11]. Here, the use of unit quaternions [12] is preferred to other representations. Unit quaternions offer a compact notation and are easier to manipulate using computers. In order to deal with the ambiguity arising from the fact that the group S^3 of unit quaternions is in a 2-to-1 homomorphism with the rotation group $SO(3)$ [5], a pre-step of hemispherization of the unit-quaternion valued data can be carried out [10]. In the following, data are supposed to be available in the form of a hemispherized distribution of unit quaternions. Under this assumption, we can consider that we are working on the group S^3. The following two paragraphs introduce concepts from the Riemannian geometry of S^3 that arise in the definition of PGA.

2.1 Riemannian metric on S^3

The quaternion sphere S^3 is a Lie group isomorphic to $SU(2)$. It is formed by unit quaternions [5, 12]. Any unit quaternion $q \in S^3$ can be written:

$$q = \cos\angle(q) + \sin\angle(q)\hat{q} \tag{1}$$

Where $\angle(q) \in [0, 2\pi]$ and $\hat{q}$ is a pure quaternion that can always be chosen in the half space given by $\mathscr{S}(\hat{q}\mathbf{k}) \leq 0$. We will always make this choice. [1] We are interested in the biinvariant Riemannian geometry of S^3 [13]. This Riemannian geometry is identical to the one induced from the ambient Euclidian space $\mathbb{R}^4$ [13, 14]. The Riemannian metric on S^3 can be used to give it the topology of a metric space with a metric distance function [15] given by:

$$d(x,y) = \arccos\left[\mathscr{S}(x\bar{y})\right] \text{ for } x, y \in S^3 \tag{2}$$

Note that $d(x,y) \in [0,\pi]$ for any two points $x, y \in S^3$. An important property is that between any two points $x, y \in S^3$ there passes a geodesic of minimal length $d(x,y)$. This geodesic is unique if $d(x,y) < \pi$.

2.2 Geodesics of S^3

A geodesic of the biinvariant Riemannian metric of S^3 passing through a point q in a direction W (a unit pure quaternion such that $\mathscr{S}(W\mathbf{k}) \leq 0$) is of the form [13, 14]:

$$c(s) = qe^{Ws},\ s \in [0, 2\pi] \tag{3}$$

In particular, a geodesic passing through the identity 1 is given by:

$$c_W(s) = \cos(s) + \sin(s)W,\ s \in [0, 2\pi] \tag{4}$$

The parameter s is the distance from the identity 1 to the current point along the geodesic curve c_W.

The projection $\pi_W(x)$ of a point $x \in S^3$ on a geodesic through the identity c_W is defined as the closest point on c_W to x [1]. We call the distance $d(x, c_W) \equiv d(x, \pi_W(x))$ the distance of the point x to the geodesic c_W. Finding $\pi_W(x)$ is an optimization problem. Since the geodesics of S^3 are compact, the solution of this optimization problem always exists. It is given by the following proposition which can be proved by direct calculation:

Proposition 1 *For any $x \in S^3$ and any geodesic c_W,*

[1] The notation of basic quaternion operations is as usual. $\mathscr{S}(q)$ and $\mathscr{V}(q)$ are the scalar and vector parts, the bar denotes conjugation, $\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ are the usual quaternion basis, etc.

- *If* $\mathscr{S}(x)+W\mathscr{S}(W\bar{x}) \neq 0$ *then the projection* $\pi_W(x)$ *is unique and given by:*

$$\pi_W(x) = c_W(s^*)\text{ , where } s^* = \angle(\mathscr{S}(x)+W\mathscr{S}(W\bar{x})) \quad (5)$$

- *If* $\mathscr{S}(x)+W\mathscr{S}(W\bar{x}) = 0$ *then the function* $D(s)$ *is a constant equal to* $\pi/2$.

The projection $\pi_W(x)$ *is thus almost always unique. It is given by the unit quaternion along the projection of x on the plane Span* $\{1, W\}$.

3. PRINCIPAL GEODESIC ANALYSIS: PGA

3.1 Introduction to PGA

PGA is defined in analogy with classical Principal Component Analysis (PCA) and constitutes an extension of it. There are several variants of PGA [1]. Here, the one given in [4] is adopted, as it uses the group structure more clearly. Given a distribution of data $\{q_i \in S^3\}_{i=0}^{N-1}$, the goal of PGA, like PCA, is to find a set of geodesic directions [2], called principal geodesic directions or principal geodesics, that best represent the variability of the data and that allow the reconstruction of the data with good precision. Unlike the linear case of PCA, the number of principal geodesics is not, in general, limited by the dimension of the space where the data are taken (in our case $dim(S^3) = 3$).

Here we will consider that the data are centered, *i.e.* that their center of mass (or Karcher mean) is the identity $1 \in S^3$. This requires, first of all, that the data be sufficiently localized for its center of mass to be well-defined [13, 16]. This is verified whenever the points q_i can be inscribed in a Riemannian sphere of radius strictly less than $\pi/2$. It is then sufficient to multiply the data, on the left or on the right, by the inverse of its center of mass, in order to make them centered.

Given the first k principal geodesic directions, $v_1, v_2, ..., v_k$ it is possible to represent the data as a product of projections on its principal geodesic directions [4]:

$$q_i = \pi_1(q_i)\pi_2(q_i^{(1)})...\pi_k(q_i^{(k-1)})q_i^{(k)} \quad (6)$$

Where $\pi_j(.)$ is the projection on the geodesic c_{v_j} and $q_i^{(k)}$ is the reconstruction error at order k. As explained above, the principal geodesic directions will have the two interesting properties of representing the variability of the data and of allowing to reconstruct it with good precision. It will be seen, experimentally, that the reconstruction error decreases rapidly with k. This means that the first few principal geodesic directions are usually sufficient for a reasonably precise description of the data.

The following procedure, taken from [4], explains how to find, for any $k > 0$, the first k principal geodesic directions:

1. The first principal geodesic direction v_1 is defined as generating the geodesic c_{v_1} that is closest, in mean distance, to all the data:

$$v_1 = argmin_v\{\sum_i d(q_i, c_v)\} \quad (7)$$

2. The following principal geodesics are defined recursively ($1 \leq j \leq k$):

$$v_{j+1} = argmin_v\{\sum_i d(q_i^{(j)}, c_v)\} \quad (8)$$

Where $q_i^{(j)}$ is the reconstruction error $q_i^{(j)} = (\pi_j(q_i))^{-1}q_i^{(j-1)}$ as defined in formula (6).

The recursivity of this procedure appears in the fact that the $(j+1)$th principal direction is the first principal geodesic direction of $\{q_i^{(j)}\}$. In this sense, we only need to master two operations in order to carry out PGA. These two operations are finding the first principal geodesic direction and projecting. Projection can be carried out quite directly, using proposition 2.2 and the operation of finding the first principal geodesic direction will be discussed in the next paragraph 3.2. Let us rewrite the procedure given above in a way that shows its recursive nature more clearly.

Assume that the operations of projection $\pi_V(Q) = \{\pi_v(q_i)\}$ and of finding the first principal geodesic direction, V, of the data $Q = \{q_i\}$ are given primitives. Use the notation, $V = FPD(Q)$, for the operation of finding the first principal geodesic direction. In order to find the principal geodesic directions of the data Q we use the following recursive function PGA:

Algorithm 1 Compute list of principal geodesic directions $PGA(Q) = (V1, V2, .., Vk, ..)$ from data $Q = \{q_i\}$

```
Function PGA(Q)
if ∑_i d(q_i, 1) < ε then
   return ();
else
   V = FPD(Q)
   return (V, PGA((π_V(Q))^{-1} Q))
end if
```

where ε is a default precision, () is an empty list, and $(V, PGA((\pi_V(Q))^{-1}Q))$ denotes the concatenation of vector V with the result list of $PGA((\pi_V(Q))^{-1}Q))$.

Hence, PGA is a procedure that can be used to describe data $\{q_i \in S^3\}_{i=0}^{N-1}$. It describes this data in terms of a set of principal geodesic directions. These directions describe the variability of the data and also allow to reconstruct it. The number of principal geodesic directions is not limited.

3.2 Finding the first principal geodesic direction

In order to carry out PGA, the operation of finding the first principal geodesic direction of data needs to be performed several times. This operation consists in finding the closest geodesic to the data *wrt.* mean distance. This is a kind of "geodesic regression", in the sense that it gives a geodesic that best represents the variability of a set of samples. Here, an analytical characterization of the first principal geodesic direction, as well as a gradient algorithm for finding it numerically, is given.

3.2.1 Analytical characterization

Given a dataset $\{q_i \in S^3\}$, then the first principal geodesic direction is given in equation (7), where the minimum should be found under the constraint that v is a unit vector. This optimization problem can be seen as searching for the minimum of the following function:

$$\begin{array}{c} \mathscr{D} : S^2 \to \mathbb{R}_+ \\ \mathscr{D}(v) = \frac{1}{N}\sum_i d(q_i, c_v) \end{array} \quad (9)$$

A necessary condition for a unit vector $V \in S^2$ to give the first principal geodesic direction of data q_i can be found through standard use of Lagrange multipliers [17], as explained in the following proposition:

Proposition 2 *If* V *is the first principal geodesic direction of data* q_i *then* V *solves the following equation:*

$$V \propto \sum_i sin(s_i^*)\mathscr{V}(q_i) \quad (10)$$

Where $s_i^* = \angle(\mathscr{S}(q_i) + v\mathscr{S}(V\bar{q}_i))$*, as in proposition 2.2. Or equivalently, the equation:*

$$V \propto \sum_i \frac{\mathscr{S}(V\bar{q}_i)}{\sqrt{\mathscr{S}^2(q_i) + \mathscr{S}^2(V\bar{q}_i)}}\mathscr{V}(q_i) \quad (11)$$

In other words, if $\mathscr{D}(v)$ *has a minimum then it reaches it at* V *which is the unit vector along the weighted mean of the vector parts of the data given in formulae (10) and(11).*

[2] In the case of PCA, these are just straight lines

3.2.2 *Gradient Algorithm*

Finding the first principal geodesic direction of the data q_i has been reduced to minimizing the function $\mathscr{D}(v)$. This is a real-valued function on the unit sphere S^2. An adequately initialized gradient algorithm on the sphere can be an effective way of finding the global minimum of this function. Gradient algorithms on various Riemannian manifolds, including spheres are given in [18].

We use a simplified version of the Riemannian gradient descent method given in [18]. It consists in considering $\mathscr{D}(v)$ as a function of vectors $v \in \mathbb{R}^3$ and applying a usual gradient descent method to it. However, at every step of the gradient descent method we project the gradient on the tangent space of the sphere $S^2 = \{||v|| = 1\}$ and normalize the new vector obtained this way. This algorithm is an approximation of the Riemannian gradient descent algorithm, to the second order of the gradient step.

The update rule of this algorithm is the following:

$$V_{k+1} = \frac{V_k - \lambda \nabla_{\|} \mathscr{D}(V_k)}{||V_k - \lambda \nabla_{\|} \mathscr{D}(V_k)||} \tag{12}$$

Where λ is the gradient step and $\nabla_{\|}\mathscr{D}(V_k)$ is the component of the gradient of $\mathscr{D}(V_k)$ in the tangent plane to the shere S^2. The gradient itself is given by:

$$\nabla \mathscr{D}(V_k) = \frac{1}{N} \sum_i \frac{-\mathscr{S}(V_k \bar{q_i})}{\sqrt{\mathscr{S}^2(q_i) + \mathscr{S}^2(V_k \bar{q_i})}} \mathscr{V}(q_i) \tag{13}$$

As usual, the efficiency of this minimization is subject to the initial choice of V_0. Choosing V_0 close to the solution vector V guarantees a fast convergence. Of course there is no unique way to do this. Here, we choose an initialization based on the following reasoning: if V achieves the minimum of $\mathscr{D}(v)$ then, by definition of the first principal geodesic direction, the data q_i lies relatively close to the geodesic c_V. Based on this assumption, we can consider that $\mathscr{S}^2(q_i) + \mathscr{S}^2(V\bar{q_i}) \approx 1$, for any q_i. By inserting this into equation (10) we obtain:

$$V \propto \sum_i \sin \angle(q_i) \mathscr{V}(q_i) \tag{14}$$

This value can be used to initialize the gradient descent. If the data q_i lies on a geodesic c_W, then it is easy to verify that the quaternion V given by formula (14) solves equation (10), that is $V = W$. In other words, formula (14) will directly give the vector minimizing $\mathscr{D}(v)$. In this case, the initialization ensures the convergence of the algorithm in one step.

A measure of how closely packed the dataset is around the first principal geodesic direction V is given by the reconstruction error q_i^1 as given in formula (8), that is:

$$q_i^1 = (\pi_V(q_i))^{-1} q_i \tag{15}$$

The reconstruction error measures how far from the first principal geodesic each data sample q_i is. If the whole dataset lies on a geodesic, this will be the first principal geodesic and we will have $\frac{1}{N}\sum_i d(q_i^1, 1) = 0$. In general, the average:

$$R_1 = \frac{1}{N} \sum_i d(q_i^1, 1) \tag{16}$$

measures the dispersion of the data away from the first principal geodesic. The smaller it is, the more aligned the data is along the first principal direction.

4. EXAMPLES AND CONVERGENCE

In this section, some examples of PGA, applied to simulated data and also to real data from motion capture, are given. They will be used to discuss the convergence of PGA and how it describes the variability of data. The convergence of PGA is measured by the reconstructon error at order k (see formula (8)), $q_i^{(k)} = (\pi_k(q_i))^{-1} q_i^{(k-1)}$. This quantity expresses the precision to which the data can be reconstructed using the first k principal geodesic directions. A global measure of this precision is given by the average:

$$R_k = \frac{1}{N} \sum_i d(q_i^k, 1) \tag{17}$$

If $R_k = 0$, then the first k principal geodesic directions describe the data completely. R_k is decreasing with k. This means that the precision of the description of the data given by PGA improves with every new principal geodesic direction. Since the R_k form a decreasing sequence of positive numbers, we can conclude that the data can be described to any desired degree of precision by using a sufficient number of principal geodesics. We give two simple examples of data with low dimensionality in paragraph 4.1. Then, in paragraph 4.2, we move on to the PGA of a Gaussian distribution. Finally, in paragraph 4.3, an example of real data from motion capture is studied.

4.1 Data with low dimensionality

Here two examples are discussed of data that can be completely described using a low number of principal geodesic. By low number, it is meant a number inferior to the dimension of the initial group S^3 where the data are taken. This happens when the data are confined to a submanifold of S^3 of dimension 1 or 2. These two examples show how PGA, like PCA, can be used to reduce the dimensionality of the data.

In the first example, $R_1 = 0$. The data can be described completely using the first principal geodesic. Simply, this means that all the data lie on some geodesic c_W.

Figure 1 shows how, in such a case, the mean distance of the data to the geodesic generated by V_k -see paragraph 3.2.2 equation (12)- decreases during the gradient descent. This mean distance decreases to 0, which shows that the data belong to the first principal geodesic so that $R_1 = 0$. Note that in this example the gradient descent was initialized at a random value. The initialization given in formula (14) was not used. This initialization would have lead to the algorithm converging in one step.

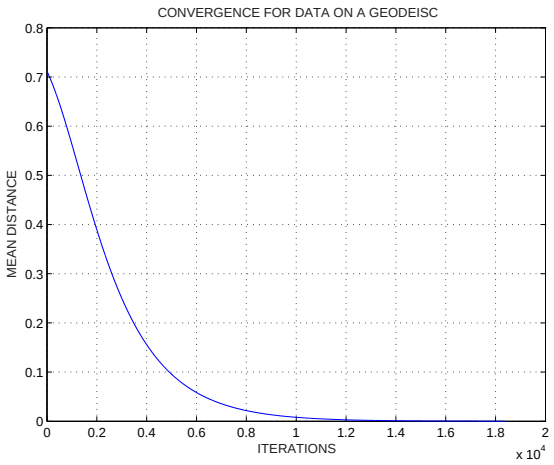

Figure 1: Gradient descent for data on a geodesic

We also considered an example where $R_1 \neq 0$ and $R_2 = 0$. The data lies, in this case, on a non degenerate two dimensional submanifold of S^3 and it can be completely described using its first two principal geodesic directions. By characterizing the variability of the data as being essentially along a 2-dimensional submanifold and not on all of S^3, PGA allows a reduction of its dimensionality.

When performing PGA in such a case, we will use the gradient descent algorithm of paragraph 3.2.2 twice. Indeed we have to find two principal geodesic directions. Figure 2 shows the convergence of PGA in this case. The first decreasing part of the graph corresponds to the first application of the gradient descent algorithm. When the first principal direction is found, at the end of this decreasing part, the search for the second principal geodesic direction starts. This search converges in one step. Indeed, the curve immediately goes to zero. This happens thanks to the initialization given by formula (14).

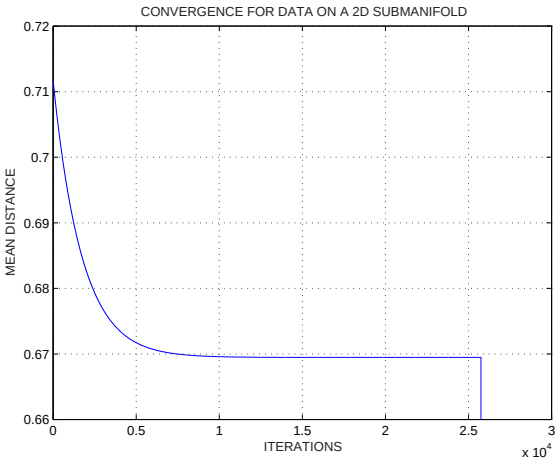

Figure 2: Gradient descent for data on a two dimensional submanifold

4.2 PGA of a Gaussian distribution

In the two examples of the last paragraph, the structure of the data was fairly simple. A low number of principal geodesics was sufficient to describe it. When the data has a more complex structure, PGA needs a relatively high number of principal geodesics in order to describe the data. In this section, the results of applying PGA to a Gaussian distribution are presented. It is seen that such a distribution requires a high number of principal geodesics to describe it.

Here, the definition given in [10] of a Gaussian distribution on S^3 is adopted. This approach defines the Gaussian distribution as the one with maximum entropy for a given covariance [16]. Its probability density function is given by:

$$p(q) = k.exp\left(\frac{-\langle log(q)|\Gamma|log(q)\rangle}{2}\right) \qquad (18)$$

Where k is a renormalization constant and Γ is a symmetric matrix called the concentration matrix of the distribution. $|log(q)\rangle$ denotes the vector corresponding to the pure quaternion $log(q)$ and $\langle log(q)|\Gamma|log(q)\rangle$ is the quadratic form defined by Γ. It is shown in [10] how to generate data following this distribution.

Figure 3 shows that a reasonably precise representation of the data generated following a Gaussian distribution can be achieved using 5 principal geodesics. This representation is however not exact. We do not have $R_5 = 0$. The 5 decreasing parts of this graph correspond to 5 calls to the gradient descent algorithm, one for every principal geodesic. It is interesting to note that we have found the first three principal geodesic directions of the generated data to coincide with the three eigendirections of the matrix Γ. It is still not clear how general this result is.

4.3 PGA on motion capture data

In this last example we tested the PGA on real data acquired from motion capture. Those data were acquired using an optical *VICON*

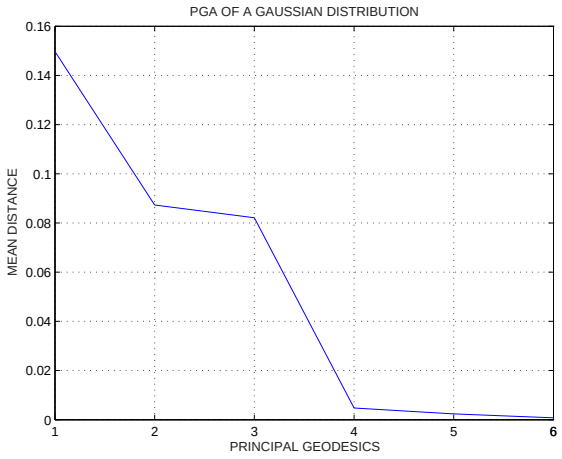

Figure 3: PGA of a Gaussian distribution

system at the rate of 30 frames per second. The rotation time series corresponds to the global rotation of a body during a 300 frames animation. The results of the reconstruction error (distance to the original data) using the k first principal geodesics are presented in Figure 4. It can be seen from Figure 4 that the reconstruction with

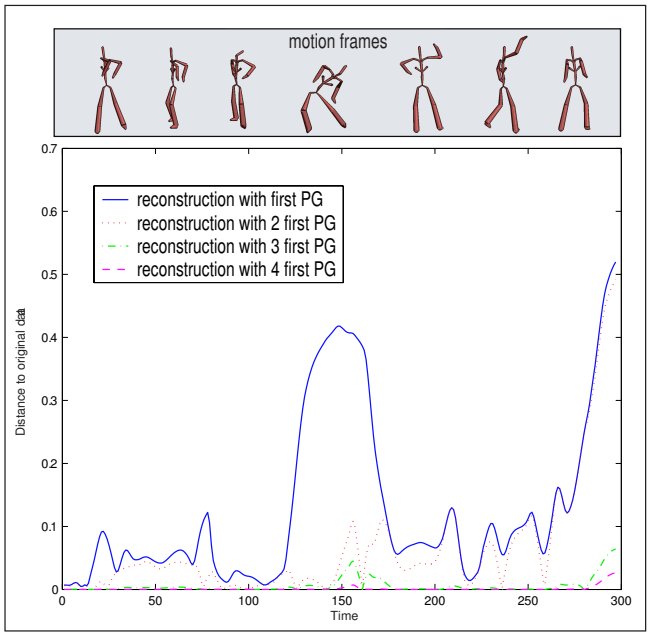

Figure 4: Reconstruction error using a limitated number of principal geodesics (PG). This example has been computed using data acquired from motion capture

only one geodesic is not very accurate, especially when the motion is rather complicated, such as depicted in the 4th frame (see the motion frames in the upper part of Figure 4). When increasing the number of principal geodesics, the error (distance to the original data) decreases rapidly (already for 2 principal geodesics), even in the case of complicated motions. It is possible to envisage an automatic segmentation of the motion by looking at the derivative of this reconstruction error [19]. This type of applications constitutes the perspectives of our work and illustrates the potential applications of PGA.

5. CONCLUSIONS AND PERSPECTIVES

In this paper, a new algorithm for the exact calculation of PGA for data on the rotation group $SO(3)$ was introduced. The goal was to gain a better understanding of this new statistical framework and to start exploring its potential applications. The algorithm we introduced was used to study the convergence of PGA and its ability to represent the variability of data. This task is performed by the classical PCA for data in a vector space. We have found that PGA extends several of the functionalities of PCA to the case of data on a Riemannian manifold -in particular $SO(3)$. Indeed, PGA can detect data with low dimensionality and reduce its dimensionality. It can represent the variability of the data using a set of principal geodesic directions and reconstruct it with any desired precision by using a sufficient number of principal geodesics. We applied PGA to a simulated Gaussian distribution on $SO(3)$ and to real data from motion capture. We have found that a higher number of principal geodesics is needed to describe these kinds of data, which have a more complicated structure.

In the case of data from motion capture, we have seen how PGA can reconstruct a time sequence of rotations. The number of principal geodesics needed to reconstruct a certain part of the time sequence with a given precision becomes higher when this part contains a more complicated motion. In other words, increasing the number of principal geodesics increases the level of detail in the reconstruction PGA gives of the data.

The ability of PGA to reduce the dimensionality of data, to represent its variability and to reconstruct it with different levels of precision makes it an interesting basis for compression, feature extraction and segmentation applications. Indeed, PCA performs these same tasks for vector-valued data and is used in such applications. The algorithm given in this paper, which makes it possible to apply PGA to data on the rotation group $SO(3)$, can be seen as a starting point for the investigation of such applications. We hope to further investigate the applications of PGA in computer graphics in future works.

REFERENCES

[1] P. T. Fletcher, *Statistical variability in nonlinear spaces: Applications to shape analysis and DT-MRI*. PhD Thesis, Chapell Hill 2004.

[2] I. T. Jolliffe, *Principal Component Analysis*. Springer series in Statistics, Second Edition, October 2002.

[3] P. T. Fletcher, and S. Joshi, "Riemannian geometry for the statistical analysis of diffusion tensor data," *Signal Processing*, vol. 87, pp. 250–262, Jun. 2006.

[4] P. T. Fletcher, and S. Joshi, *Statistics of shape via Principal Geodesic Analysis on Lie groups*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 95–101, 2003.

[5] S. L. Altmann, *Rotations, quaternions and double groups*. Dover Publications, 2005.

[6] M. Moakher, *Means and averaging in the group of rotations*, SIAM J. Matrix Anal. Appl., vol. 24, pp. 1–16, Jun. 2002.

[7] J. Kuffner, *Effective sampling and distance metrics for 3d rigid body path planning*. Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA 2004), IEEE, May, New Orleans, 2004.

[8] K. Shoemake, *Quaternions and* 4×4 *matrices*. Graphics Gems II, J. Arvo, Ed. Academic Press, 1991, pp. 351–354.

[9] J. Lee, and S. Y. Shin, *General construction of time-domaine filters for orientation data*, IEEE Transactions on Visualization and Computer Graphics, 8(2):119– 128, April - June 2002.

[10] M. P. Johnson, *Exploiting quaternions to support expressive interactive character motion*. PhD Thesis, Massachusetts Institute of Technology, 2003.

[11] G. S. Chirikjian and A. B. Kyatkin , *Engineering applications of non commutative harmonic analysis*. CRC Press, September 2000, Chapter 5.

[12] J. P. Ward, *Quaternions and Cayley Numbers, algebra and applications*. Kluwer Academic Publishers, 1997.

[13] J. H. Manton, "A globally convergent algorithm for computing the center of mass on compact Lie groups," in *ICARCV*, 2004.

[14] X. Pennec, *Bi-invariant means in Lie groups. Application to Left-invariant polyaffine transformations*. INRIA Research Report No. 5885, April, 2006.

[15] Y. Choquet-Bruhat, C. De Witt-Morette ,M. Dillard-Bleick, *Analysis, manifolds and physics*. North-Hollad 7th Edition, September 2004, Chapter 5.

[16] X. Pennec, *Probabilities and statistics on Riemannian manifolds: A geometric approach*. INRIA Research Report No. 5093, January 2004.

[17] S. Lang, *Calculus of several variables*. Springer, February 2005, pp.135–143.

[18] C. J. Taylor and D. J. Kriegman, *Minimization on the Lie group* $SO(3)$ *and related manifolds*. Yale University, Technical report no. 9405, April 1994.

[19] J. Barbic and A. Safonova and J. Pan and C. Faloutsos and J. Hodgins and N. Pollard *Segmenting Motion Capture Data into Distinct Behaviors*. Proc. of the Graphics Interface 2004 Conference (GI 2004), 185–194, May 17-19, 2004, London, Ontario, Canada.