Indian Institute of Technology Bombay

RnD Project

Statistical Shape Analysis

Authors:
Nilesh Kulkarni
Sushant Hiray

Supervisor:
Prof. Suyash Awate

A thesis submitted in fulfillment of the requirements for RnD Project

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

January 2015

Acknowledgements

We would like to take this opportunity to express our deepest gratitude towards our guide Prof. Suyash Awate for his motivating support, guidance and valuable suggestions.

Abstract

A primary goal of statistical shape analysis is to describe the variability of a population of geometric objects. A standard technique for computing such descriptions is principal component analysis. However, principal component analysis is limited in that it only works for data lying in a Euclidean vector space. While this is certainly sufficient for geometric models that are parameterized by a set of landmarks or a dense collection of boundary points, it does not handle more complex representations of shape. In this report techniques such as Principal Geodesic Analysis (PGA) and Principal Nested Spheres(PNS) are presented and subsequently demonstrated and discussed in depth through a biometric case study of hand data set.

Contents

Chapter 1

Principal Component Analysis

1.1 Introduction

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. [1]

The main purposes of a principal component analysis are the analysis of data to identify patterns, reduce the dimensions of the dataset with minimal loss of information and extract various features from the data.

1.2 Understanding the PCA approach

We can summarize the approach in the following steps:

- Consider our dataset consists of d-dimensional samples.
- Compute the d-dimensional mean
- Compute the covariance matrix of the data set.
- Compute the eigenvalue and correspondingly the eigen vectors of the covariance matrix.
- Sort the eigen vectors in the decreasing order of their eigen values.
- Choose m vectors corresponding to m biggest eigen values and form a $d \times m$ matrix
- Use this $d \times m$ matrix to transform the samples onto the new subspace using the equation $y = W^T x$

1.2.1 Problem Statement

This project focuses on estimating the mean shape, given various shapes. The given shapes can be rotated, scaled, translated versions of each other with some error (introduced possibly during annotations).
To test our code, we used 2 types of data:

- Automatically generated data comprising of ellipses with differing major-axis, minor-axis and rotation w.r.t the X-Y axos
- A dataset comprising of annotated images. It consists of 56 points per image and 40 such images.

1.3 Algorithm Used

To completely understand the algorithms used, we will first define some terms.

- **Shape** is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.
- The **Shape Space** is the set of all possible shapes of the object in question
- **The Procrustes distance** is a least-squares type shape metric that requires two aligned shapes with one-to-one point correspondence

1.3.1 Procrustes Analysis

The alignment of images involves four steps:

- Compute the centroid of each shape.
- Re-scale each shape to have equal size.
- Align w.r.t. position the two shapes at their centroids.
- Align w.r.t. orientation by rotation

The Procrustes Distance between 2 images x_1 and x_2 is given by:

$$P_d^2 = \sum_{j=1}^{n} [(x_{j1} - x_{j2})^2 + (y_{j1} - y_{j2})^2]$$

The centroid of the shape is found out as follows:

$$(X, Y) = (\frac{1}{n}\sum_{j=1}^{n} x_j, \frac{1}{n}\sum_{j=1}^{n} y_j)$$

To align 2 shapes together, Single Value Decomposition is applied.

- Arrange the size and position aligned x_1 and x_2 as $n \times 2$ matrices.
- Calculate the SVD, UDVT, of $x_1^T x_2$ in order to maximize the correlation between the two sets of landmarks.
- The rotation matrix needed to optimally superimpose x_1 upon x_2 is then VU^T. i.e VU^T is the rotation matrix.

1.3.2 Generalized Procrustes Analysis

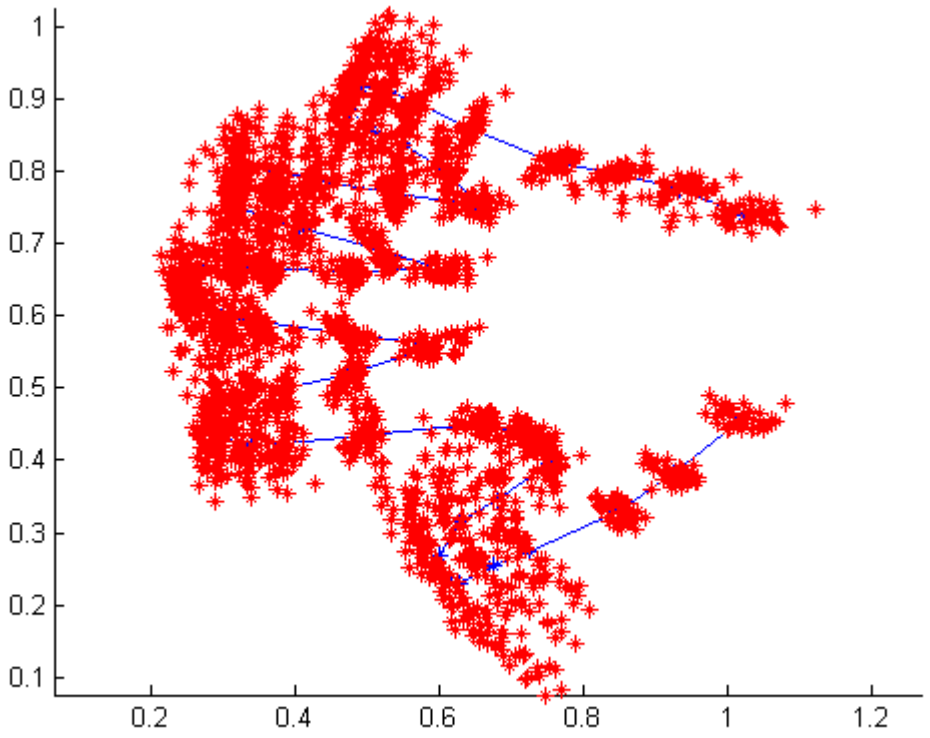
This is essentially an iterative approach using the Procrustes Method. It is done as follows:

- Choose an initial estimate of the mean shape (we chose the first shape in the set).
- Align all the remaining shapes to the mean shape.
- Re-calculate the estimate of the mean from the aligned shapes.
- If the estimated mean has changed return to step 2.

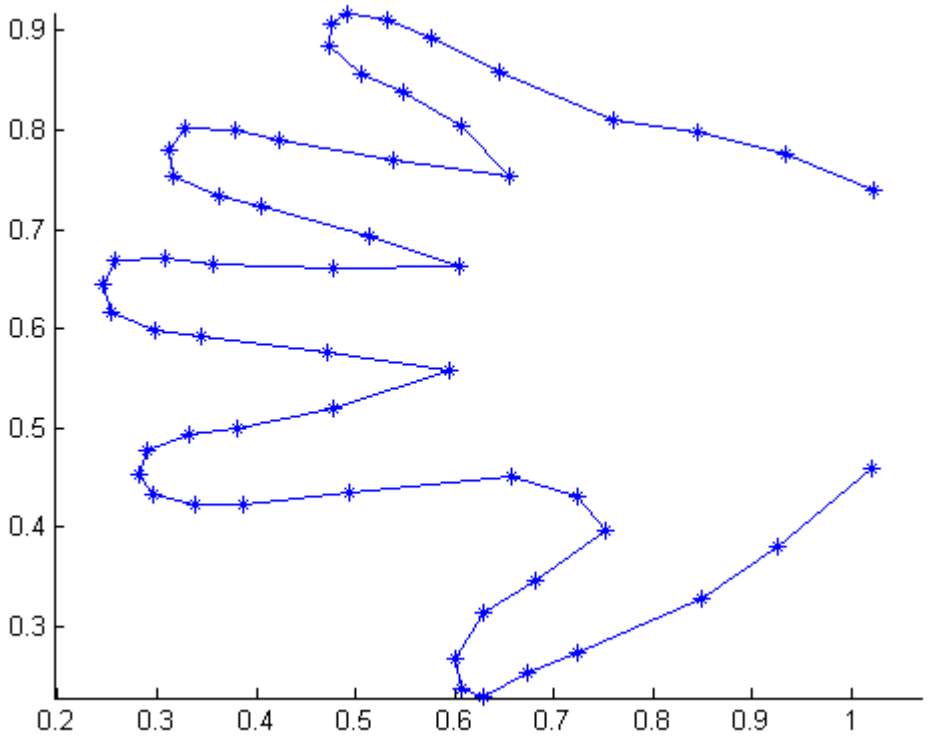
1.4 Experiments

The following subsections will show in detail the result of using the algorithm.

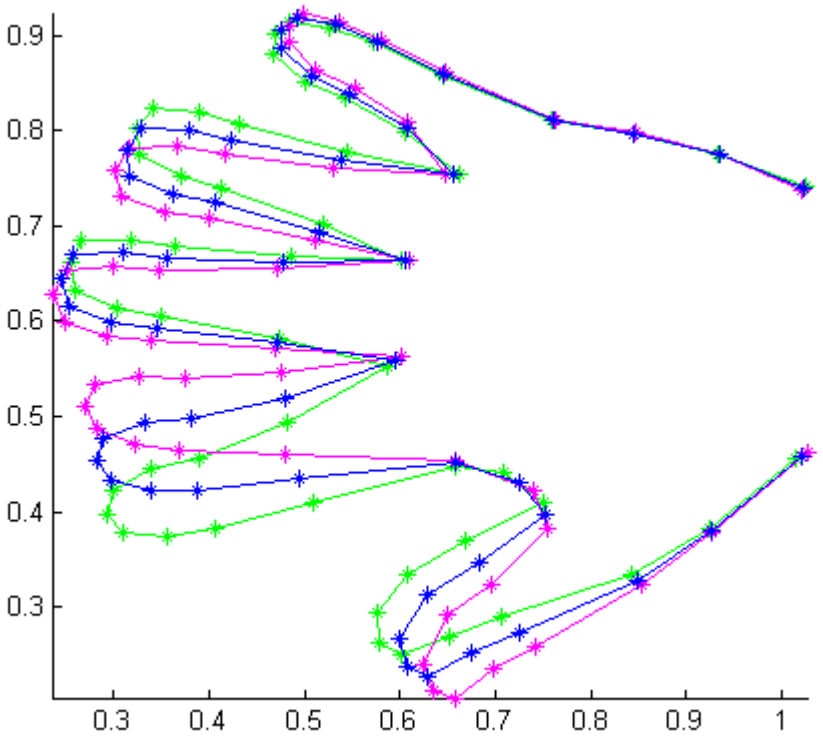
1.4.1 Hand Data



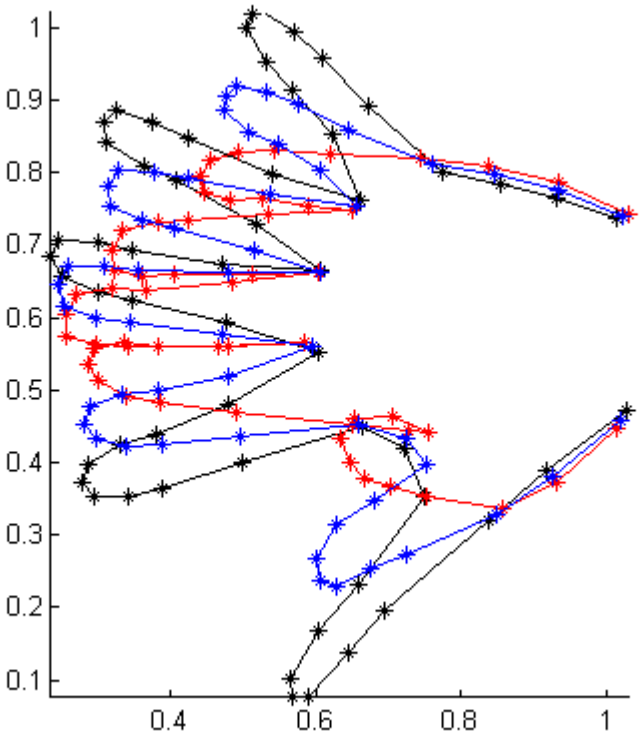
This image shows the Scaled Data Points for all the annotated hand images in red and the final mean data in blue. The mean shape has been joined together just for further clarity.



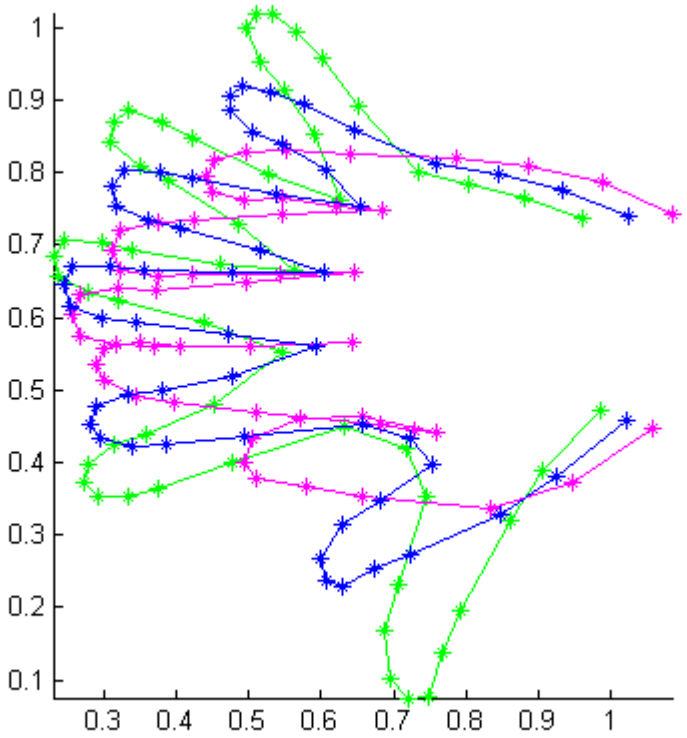
The mean hand shape which was buried beneath the scaled data points in previous image.



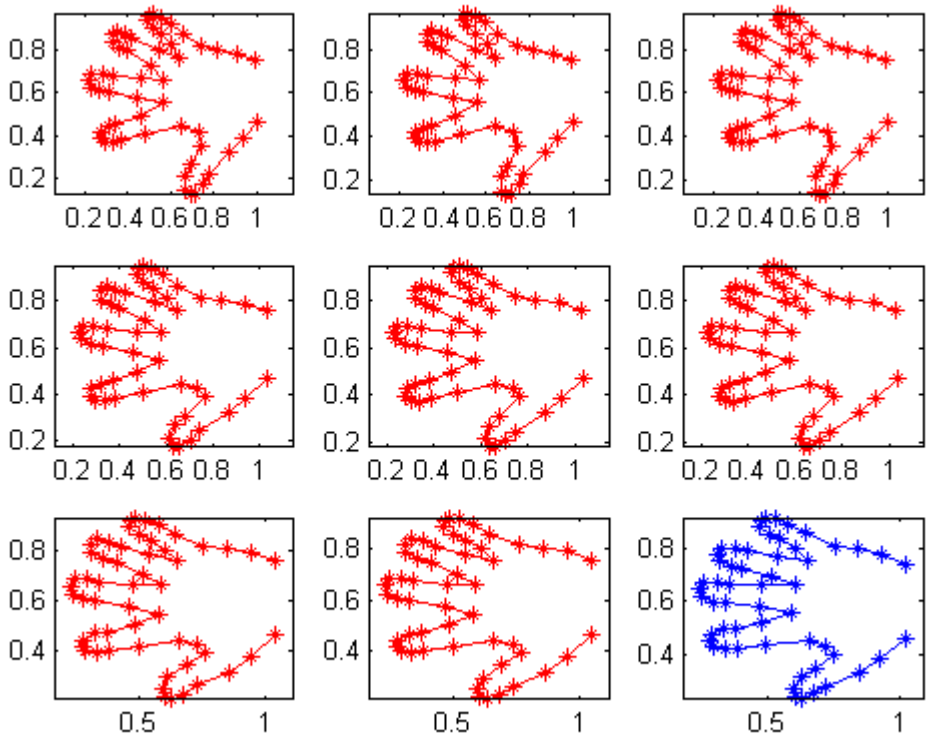
Blue colored is the mean hand. The green color is $Mean - \sqrt{\lambda}$ and the Magenta is $Mean + \sqrt{\lambda}$ This is the mode1 image formed corresponding to the largest eigen value.



Blue colored is the mean hand. The red color is $Mean - \sqrt{\lambda}$ and the black is $Mean + \sqrt{\lambda}$. This image is corresponding to the second highest eigen value. As we can see, the red color variant tries to describe fingers held together whereas the black variant describes fingers stretching outwards.



Blue colored is the mean hand. The green color is $Mean - \sqrt{\lambda}$ and the Magenta is $Mean + \sqrt{\lambda}$. This image is corresponding to the 3^{rd} highest eigen value. In continuance with the previous description, the green colored variant is a bit more wider and the magenta variant is a bit more close knit.



The 1^{st} 7 images show annotated Hand images, the last image shows the Calculated Mean Hand Shape. The scales of all the images are same and hence we can compare the given data with the final Mean Shape.

Chapter 2

Principal Geodesic Analysis

2.1 Improving Over PCA

A primary goal of statistical shape analysis is to describe the variability of a population of geometric objects. In the previous section we described the usage of principal component analysis for statistical shape analysis.

Principal component analysis works best for data lying in a Euclidean vector space. Although this works very good for basic geometric models, it does not handle complex representation of shapes. While this is certainly sufficient for geometric models that are parameterized by a set of landmarks or a dense collection of boundary points, it does not handle more complex representations of shape. The primary reason being the fact that we consider linear distances while using PCA.

In this section we will look into the method of principal geodesic analysis, a generalization of principal component analysis to the manifold setting. Instead of using linear distances, PGA uses geodesic distances.

2.2 Basic Definitions

Definition 2.1. Geodesic is a locally length-minimizing curve. In the plane, the geodesics are straight lines. On the sphere, the geodesics are great circles (like the equator)

Definition 2.2. N-sphere: For any natural number n, an n-sphere of radius r is defined as the set of points in (n + 1)-dimensional Euclidean space which are at distance r from

a central point, where the radius r may be any positive real number.

$$S^n = \{x \in \mathbf{R}^{n+1} \colon \|x\| = r\} \tag{2.1}$$

Definition 2.3. Exponential Map Consider a tangent vector in T_pS^2 as a vector $v = (v_1, v_2, 0)$ in the x-y plane. Then the exponential map is given by

$$Exp_p(v) = \left(v_1 \frac{sin\|v\|}{\|v\|}, v_2 \frac{sin\|v\|}{\|v\|}, cos\|v\|\right) \tag{2.2}$$

where $\|v\| = \sqrt{v_1^2 + v_2^2}$

Definition 2.4. Log Map The log map for a point $x = (x_1, x_2, x_3) \in S^2$ is given by

$$Log_p(x) = \left(x_1 . \frac{\theta}{sin(\theta)}, x_2 . \frac{\theta}{sin(\theta)}\right) \tag{2.3}$$

where $\theta = arccos(\theta)$ is the distance from the base point p to the point x.

Definition 2.5. The intrinsic mean of a collection of points $x_1, ..., x_N \in M$ as the minimizer in M of the sum-of-squared Riemannian distances to each point.

$$\mu = \arg\min_{x \in M} \sum_{i=1}^{N} d(x, x_i)^2 \tag{2.4}$$

2.3 Understanding Principal Geodesic Analysis

The goal while using PGA is to find a sequence of nested geodesic submanifolds that maximize the projected variance of the data.

Let $T\mu M$ denote the tangent space of M at the intrinsic mean μ of the x_i. Let $U \in T\mu M$ be a neighborhood of 0 such that projection is well-defined for all geodesic submanifolds of $Exp_\mu(U)$. The principal geodesic submanifolds are defined by first constructing an orthonormal basis of tangent vectors $v_1, ..., v_d \in T_\mu M$ that span the tangent space $T\mu M$. These vectors are then used to form a sequence of nested subspaces $V_k = span(v_1, ..., v_k) \cap U$. The principal geodesic submanifolds are the images of the V_k under the exponential map: $H_k = Exp_\mu(V_k)$.

The first principal direction is chosen to maximize the projected variance along the corresponding geodesic:

$$v_1 = \arg\max_{\|v\|=1} \sum_{i=1}^{N} \|Log_\mu(\pi_H(x_i))\|^2 \tag{2.5}$$

where $H = Exp_\mu(span(\{v\}) \cap U)$

The remaining principal directions are then defined recursively as

$$v_k = \arg\max_{\|v\|=1} \sum_{i=1}^{N} \|Log_\mu(\pi_H(x_i))\|^2 \tag{2.6}$$

where $H = Exp_\mu(span(\{v_1, ..., v_{k-1}, v\}) \cap U)$

2.3.1 Algorithm

Finding the Intrinsic Mean:

Pseudo Code

```
Initialise mu0 = x1 (Any data point)
LOOP (dmu > threshold):
    dmu = average of log map of each data point wrt muj
    muj1 = exponential map of dmu wrt muj
```

Finding the principal modes of variations

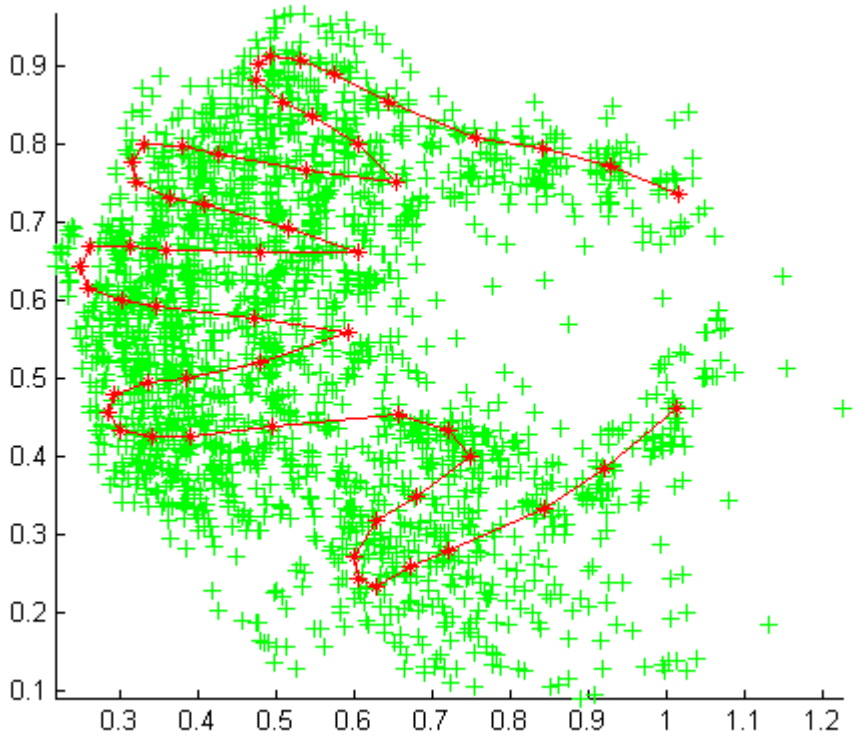
Pseudo Code

```
    mu = Intrinsic mean of the data points (Above algorithm)
    {ui} = log map of data point xi wrt mu
    S = averaged norm of {ui}
    Find the eigen vectors corresponding to this S using the normal PCA
```

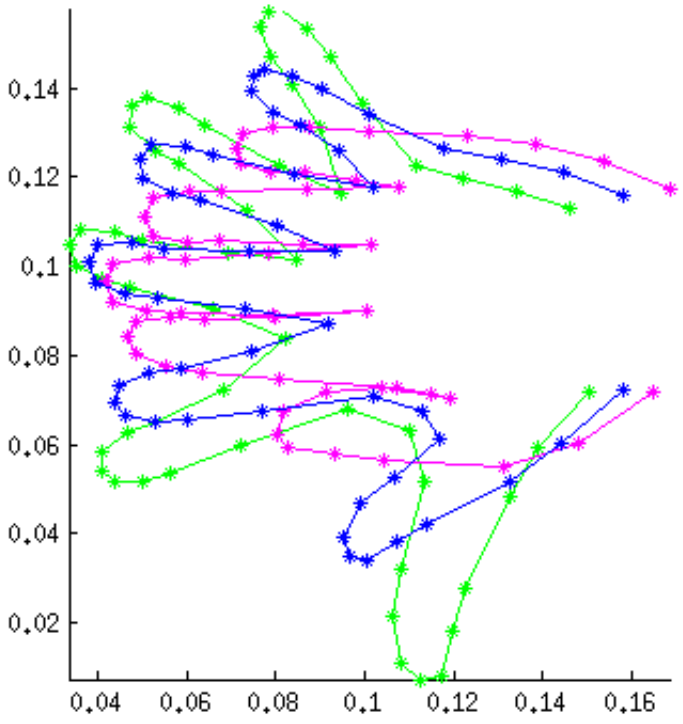
2.4 Experiments

The following subsections will show in detail the result of using the algorithm.

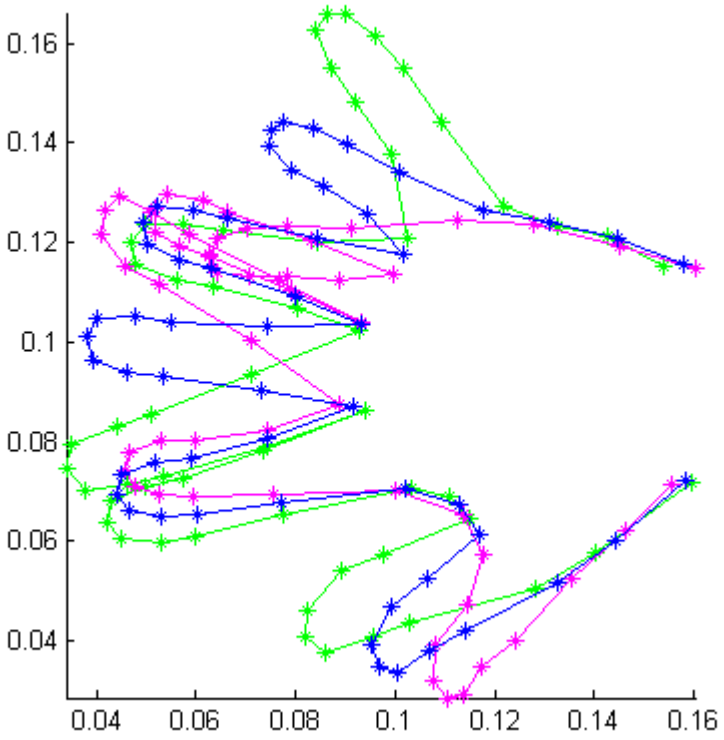
2.4.1 Hand Data



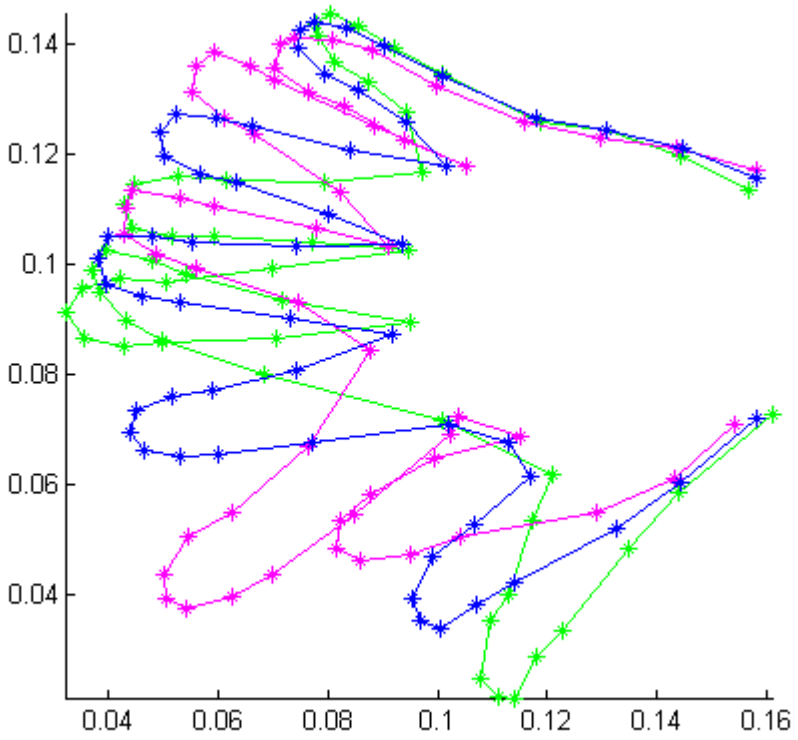
This image shows the Scaled Data Points for all the annotated hand images in green and the final mean data in red. The mean shape has been joined together just for further clarity.



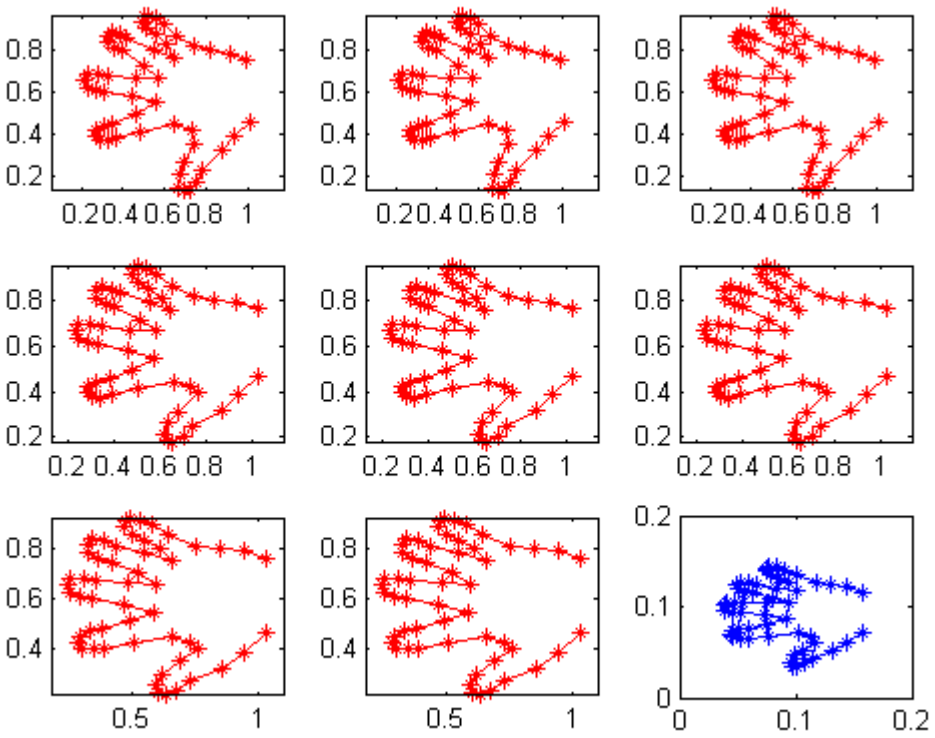
Blue colored is the mean hand. The green color is $Mean - \sqrt{\lambda}$ and the Magenta is $Mean + \sqrt{\lambda}$ This is the mode1 image formed corresponding to the largest eigen value.



Blue colored is the mean hand. The green color is $Mean - \sqrt{\lambda}$ and the Magenta is $Mean + \sqrt{\lambda}$. This image is corresponding to the second highest eigen value. As we can see, the Majenta color variant tries to describe fingers held together on the left side whereas the green variant describes fingers held together on the right side.



Blue colored is the mean hand. The green color is $Mean - \sqrt{\lambda}$ and the Magenta is $Mean + \sqrt{\lambda}$. This image is corresponding to the 3^{rd} highest eigen value. In continuance with the previous description, the green colored variant is a bit more wider and the magenta variant is a bit more close knit.



The 1^{st} 7 images show annotated Hand images, the last image shows the Calculated Mean Hand Shape. The scales of all the images are same and hence we can compare the given data with the final Mean Shape.

Chapter 3

Principal Nested Spheres

Principal Nested Spheres(PNS) is another approach to data dimensionality reduction. It is significantly different than PCA and PGA as disscussed in chapter 1 and 2 respectively. The main idea on PNS is based on using a Top-Down approach in finding the modes of variations. The most important aspect being PNS tries to find sub-spaces that fit the data best contrary to the others that try to find modes along which the data varies.

Before going to the PNS we need to discuss some prelimanaries.

3.1 Geometry of Nested Spheres

For a unit sphere S^d, the geodesic joining any two points is a great circle joining the two points. A natural distance on S^d is the geodesic distance function $\rho_d(.,.)$ defined as the length of the shortest great circle segment joinin[width=100px,keepaspectratio=true]g $x, y \in S^d$, $\rho(x,y) = cos^{-1}(x^T y)$.

Definition 3.1. A subsphere A_{d-1} in S^d is defined by an axis $v \in S^d$ and a distance $r \in [0, \frac{\pi}{2}]$ as follows

$$A_{d-1}(v,r) = \{x \in S^d : \rho v, x = r\} \tag{3.1}$$

More details of the Definition 1 from [2]

The transformation equations are as follows

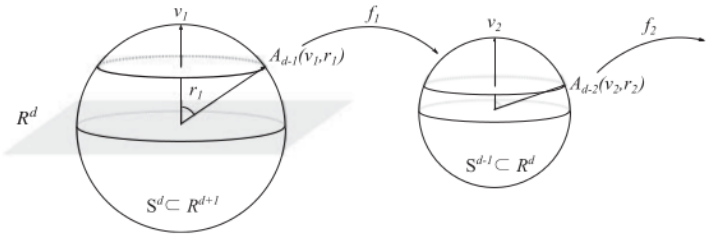


FIGURE 3.1: The subspehere $A_{d-1}(v_1, r_1)$ in S^d and its relation to S^{d-1}, through the transformation f_1. This is recursively done until we reach dimension 2. Figure from [2]

$$
\begin{aligned}
f_k(x) &= \frac{1}{sin(r_k)} R^-(v_k)x, \\
f_k^{-1}(x^\dagger) &= R^T(v_k) \begin{pmatrix} sin(r_k)x^\dagger \\ cos(r_k) \end{pmatrix}
\end{aligned}
\tag{3.2}
$$

where $R(v_k)$ is a rotation matrix $(m+1) \times (m+1)$ that moves v_k to north pole. $R^-(v_k)$ is the $m \times (m+1)$ matrix consisting on m rows from $R(v_k)$.

3.2 Finding the best fitting subsphere

Let $x_1, x_2, \ldots, x_n$ be a sample in space $S^d d \geq 2$. We have residual errors $\xi_i(v, k)$ of a subsphere $A_{d-1}(v, k)$ w.r.t to S^d as the residual error of sample point between x_i and $x_i^\dagger$ (which the corresponding point on $A_{d-1})(v, k)$. $\xi = \rho_d(v, x) - r$. Now, the task is to find the best fitting $\hat{A_{d-1}} = A_{d-1}(\hat{v}, \hat{k})$ which tries to minimize the summation of above errors.

$$
\sum_{i=1}^{N} \xi_i = \sum_{i=1}^{N} \{\rho(x_i, \hat{v_1}) - r\}^2 \tag{3.3}
$$

where, $v_1 \in S^d$ and $r \in [0, \frac{\pi}{2}]$
Now we have define the projection of the point x in S^d on $\hat{A_{d-1}}$ by the following results

$$
Px; A_{d-1}(v, r) = \frac{sin(r)x + sin\{\rho_d(x, v) - r\}v}{sin\rho_d(x, v)} \tag{3.4}
$$

3.2.1 Sequence of Principal Nested Spheres

The procedure to find the sample principal nested spheres consists of iteratively trying to find the best fitting subspheres and then mapping all the data points on to it. The d-1 dimension subsphere $\hat{A_{d-1}}$ is the best fitting subsphere. Now using the data that is projected on $\hat{A_{d-1}}$ we try t find another subsphere $\in S^{d-2}$ which best fits the data,

called $\hat{A_{d-2}}$. This procedure is continued iteratively until we reach a d =2. At each stage the dimension of the data decreases by 1, and when the dimension is 2 we try to find the intrinsic mean as defined in 2.5

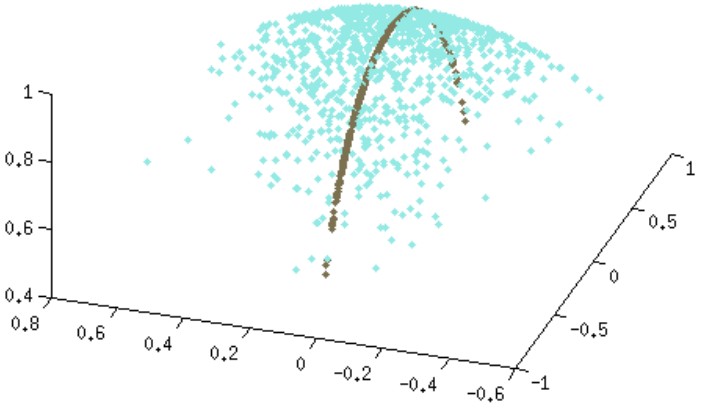


FIGURE 3.2: PNS fitting data in R^3, the light color represnts data, while the dark is the fitted greated circle $\in S^1$

3.3 Computing the best Sub-Sphere

Now, computing the best fitting subsphere is not a convex problem because the function we are trying to minimize. To understand the same consider

The computation of PNS requires an iterative applications of minimization, projection and transformation. The fornulae for projection [3.4] and transformation [3.2] are aleardy mentioned above. Here first try to inutively analyse the problem. The least constrained optimization problem mentioned [3.3] as we know is a non linear constrained minmization. The method to solve this is an iterative strategy involving two phases.

- Outer Iterations
- Inner Iterations

To solve this problem we need tow borrow the definitions of log map [2.4] and exponential map [2.3].
The algorithm first tries to find a suitable point of tangency, which is also the axis of

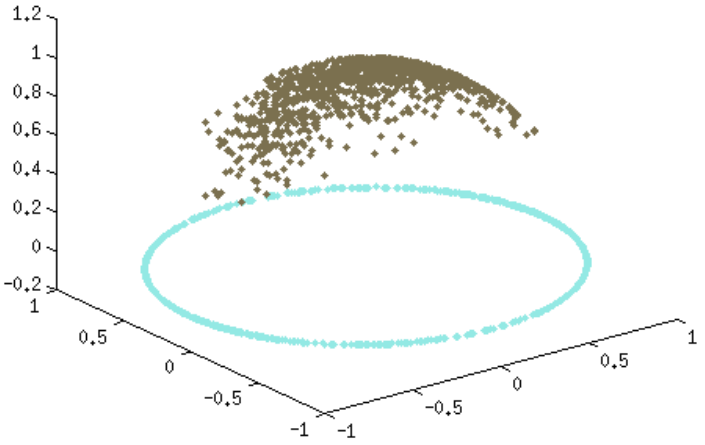


FIGURE 3.3: PNS fitted a local minima to the data if we dont use intelligent intialization

the fitted subsphere. Given a candidate v_0, the data is mapped to the tangent space $T_{v_0}S^m$ using log map. With $x_i^\dagger = Log_{v_0}(x_i)$, the inner loops finds

$$\min_{v^\dagger, r} \sum_{i=1}^{n} (||x_i^\dagger - v^\dagger||) - r)^2, \tag{3.5}$$

is a non-linear least squares regression problem that can be solved numerically. We applied gradient descent for this problem with following intialization.

Pseudo Code

```
LOOP:
    Eig = Eigen Vectors for the Data
    v0 = Initailize with Eigen vector of least Eigen Value
    TangentMapData = Porject Data to Tangent space with v0 the axis
    (v',r') = optimize(TangentMapData) // Finds the optimal v' and r'
    v1 = v' - (v0 . v')v0 // Only consider componenet of v' perpendicular to v0.
    v0 = exp_map(v0,v1) // update for v0
```

The above loop is executed until the residual error becomes constant or there is no change in the value v_0.

Lets us analyse each and every aspect the steps we have taken.

- Initialization of v_0 : The initialization of v_0 is done via trying to find the intializlize it with the subspace that captures teh most of the variance. So, the subspace that v_0 helps us capture is the set of all the vectors that are perpendicular to it. Hence, we have to ensure that when we initialize we are going to eliminate a possible subspace. So, we choose to eliminate that mode of variation with captures the least data approimately given by the PCA.
- After initializing v_0 the next step is to perfrom gradient descent optimization with adaptable step size to update $v^{\dagger}$. While applying gradient descent the r is kept constant and only v is varried there seems to be no change in $v^{\dagger}$ or objective (termination criteria).
- Soving for r has a closed form solution, the mean of all the points.

3.3.1 Understanding subspaces

The PNS algorithm is iterative in nature, given a data of dimension (d+1) it recursively tries to find sub spaces $S^d, S^{d-1}, \ldots, S^2$. And then it finds the geodesic mean **??**. So, for every subspace we have a mapping which try to represent described by (v_i, r_i), $i \in [d, 2]$.

3.4 From Subspaces to Modes of Variation

The final objective of any data demensionality reduction algorithm is to try find the modes of variation. There is not a much use if we tell a application specific person regarding these spheres, the end result that we are concerned would be the directions in which the data varies. So, we need a methodology to analyse the modes from these subspaces.

3.4.1 PNS with Greater Circles

Greater circle is a subspace withing S^d with $r = \frac{\pi}{2}$. We analysed the PNS algorithm for greater circles and dervied a formulation to find a modes of variation. Greater Circle play a special role in optimization 3.5. It makes analysis and visualization easy at it reduces the number of spaces where we have to search.

Consider finding the first 2 modes of variation, the first mode of variation will be the geodesic mean, now the since the modes of variation has have to be orthonormal to each other. The second mode of variation will be passing through the geodesic mean.

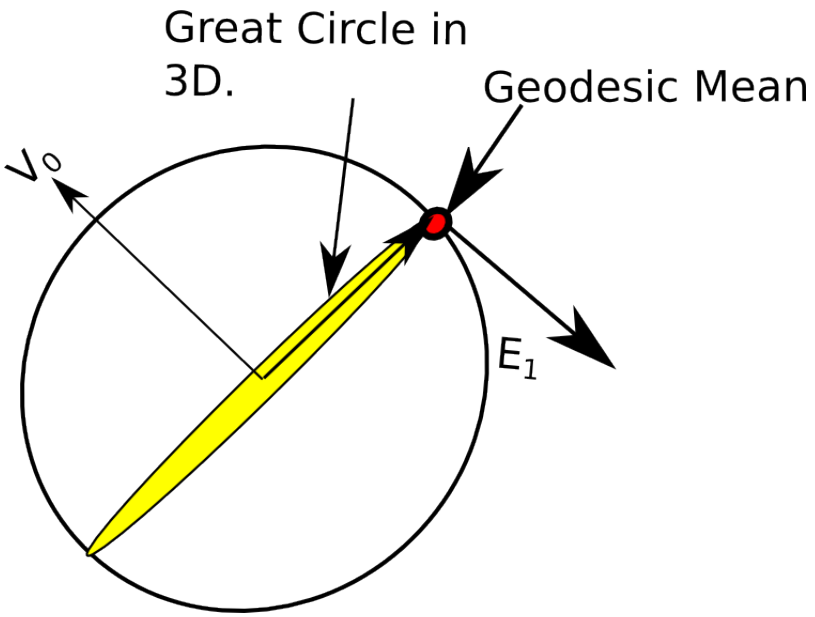


FIGURE 3.4: The figure show modes of variation S^2, The yellow ellipse represents the S^1 sub space which captures the data best. Red dot represents the Geodesic Mean(GM)

Now we reproject the data from S^1 to S^2 and we see that we can find the vector in tangent space at GM, which is represented by E_1. Now, to find the higher modes of variation we rotate this space according the the inverse mapping and then we continue doing this until we reach the final subspace S^d.

3.4.2 Analysing Variances of Subspaces

We did an interesting experiment by analysing the variances captured by the subspaces and we got the variances captured by these subspaces non-increasing at each level. Inuitively speaking the first subspace which we compute captures most of the variance and the remaining spaces are part of this and will contribute a non-zero amount to the variance.

So when you consider a data actually distributed on say a greater circle but you just try to add dimensionality to it by rotating and increasing componenets. In such a case PNS ignores all the rotation and in the first iteration of PNS one can find a good fititng sphere. So, all the lower subspaces capture zero variances.

Bibliography

[1] Wikipedia. Principal component analysis — wikipedia, the free encyclopedia, 2014. URL `http://en.wikipedia.org/wiki/Principal_component_analysis`. [Online; accessed 25-November-2014].

[2] Sungkyu Jung, Ian L. Dryden, and J. S. Marron. Analysis of principal nested spheres. *Biometrika*, 2012. doi: 10.1093/biomet/ass022. URL `http://biomet.oxfordjournals.org/content/early/2012/07/24/biomet.ass022.abstract`.