University of Alberta

Euclidean Distance Matrix Analysis of Landmarks Data: Estimation of Variance

by

Liangyuan Hu ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science
in
Statistics

Department of Mathematical and Statistical Sciences

Edmonton, Alberta
Spring 2007

Your file Votre référence
ISBN: 978-0-494-29968-5
Our file Notre référence
ISBN: 978-0-494-29968-5

ABSTRACT

"Form" of an object consists of "size" and "shape". The form of an object is the characteristic that remains invariant under a group of transformations comprising of translation, rotation, and/or reflection. Landmarks are commonly used to quantitatively represent form of an object. Estimation of the mean form and variability around it is important in fields such as evolutionary quantitative genetics, surgery, protein science, etc. Due to the presence of nuisance parameters of rotation and translation, general covariance matrix is known to be not identifiable. But certain structured covariance matrices can be shown to be identifiable. In this thesis, we provide conditions under which the covariance matrix of landmarks data is identifiable. Furthermore, we provide a computationally simple approach based on Euclidean Distance Matrix Analysis (EDMA) to estimate the mean form and covariance matrix. It is shown that this estimator is consistent, i.e., as sample size increases, it converges to the true covariance matrix. We use simulations to check the validity of the theoretical results. Further we use the method to estimate variability in the mouse mandibles using landmark data.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr. Subhash Lele, for his guidance, support, suggestions and patience throughout my studies and in preparation of this thesis. His knowledge, inspiration and thoughtful comments truly helped me in all the time of research for and writing of this thesis.

Special thanks go to Dr. N.G.Narasimha Prasad and Dr. Tarek El-Bialy, for their time spent in reading my thesis and participating on my defense committee.

Thank you to Dr. Yau Shu Wong for his encouragement to me. Without his trust in me, I would not have been able to study in this area.

My gratitude extends to the Department of Mathematical and Statistical Sciences faculty members who taught me courses during my studies.

My final thanks go to my husband, M.J., for this patiencé, understanding, support and commitment throughout the past two years of my studies and preparation of this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

"Form" consists of "size" and "shape". The analysis of the form of an object is important in biological sciences and is useful in many fields such as medical statistics, surgery, genetics, protein science (Godzik 1996) and evolutionary biology (Gould 1977; Lele and Ritchersmeier, 2001). Practical analysis of biological form is hindered by the complexity of quantifying an entire form. One approach to reducing this complexity is to consider a few biologically important points on the form under consideration and assume that the configuration of these points approximates the underlying form adequately for the problem at hand (Dryden and Mardia, 1998; Lele and Ritchersmeier, 2001; Lele and McCulloch, 2002). Such points are called "landmarks". With current imaging technology, landmark data can be easily and accurately recorded. For example, suppose we want to calculate landmark coordinates on a mouse mandible. A few landmarks on the mouse mandible such as: Coronoid tip, Inferior incisor alveolus and Superior incisor alveolus may be chosen to represent the form of the mouse mandible (Cheverud, et al., 1997). The loci of landmarks on the mandible of one mouse are shown in Figure 1.1. The description of these landmarks is in Table 4.1. We fix an object on a digitizer and the coordinate of each landmark with respect to the chosen coordinate system is computed using the digital video data collection system. In this way, we can get the landmarks data matrix. Suppose we selected n landmarks on the form of the mouse mandible, then the landmarks data matrix for one subject may look like:

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \cdots & \cdots \\ x_n & y_n \end{bmatrix}$$

where x, y denote the 2 coordinates.

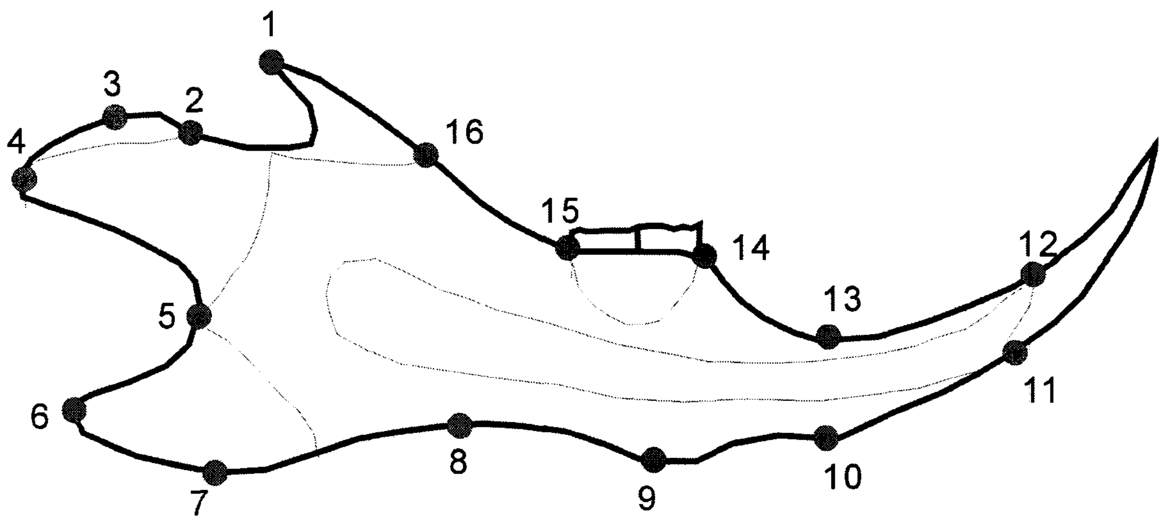

Figure 1.1: Diagram of a mouse mandible indicating the locations of 16 landmarks.

Assuming the configuration of these landmarks represents the form of mouse mandible adequately, the question is how to measure the variability among individuals that are represented by these 2D landmark data. In statistical studies, when analyzing landmarks data, variability is particularly difficult to characterize, because data on an individual is collected in a coordinate system specific to the orientation of that individual during data collection. This makes the problem statistically challenging. It is known that the general variance parameter is non-identifiable (Lele, 1993; Lele, 2002). In this thesis, we provide conditions under which a structured matrix variance is identifiable, further we provide a simple approach based on Euclidean Distance Matrix Analysis (EDMA) to estimate these parameters consistently. Other statistical methods have been developed to solve this particular question. One of the most widely used methods is that of Generalized Procrustes Analysis (GPA). Another recently developed method is Hierarchical Generalized Procrustes Analysis (HGPA) (Theobald and Wuttke, 2006). We compare EDMA based estimators with these two methods in Chapter 5.

The outline of the thesis is as follows. Chapter 2 gives a brief introduction to landmark coordinate data, the perturbation model and the algorithm for estimating mean form and covariance matrix. Chapter 3 evaluates the performance of EDMA with simulated data. In Chapter 4 we present an analysis of the mouse mandible data. In Chapter 5, we compare the performance of EDMA to that of HGPA with simulated data. Chapter 6 concludes the thesis.

Chapter 2

The perturbation model and EDMA

Suppose we have K landmarks on a D-dimensional objects, then we construct a $K \times D$ matrix whose jth row consists of the D coordinates of the jth landmark. Usually D is either 2 or 3 and K is assumed to be larger than D. We use X_i to denote the $K \times D$ matrix of coordinates for the ith individual.

2.1 Perturbation Model

Landmarks data are commonly modeled using the perturbation model. (Goodall, 1991 and Lele, 1993). The perturbation model may be thought of as representing the following process. To generate a random geometrical object or equivalently, a K point configuration in D-dimensional Euclidean space, nature first chooses a mean form (represented by matrix M) and perturbs the elements of this matrix by adding noise to this mean form according to a matrix-valued Gaussian distribution. The K point configuration so obtained is then rotated and/or reflected by an unknown angle and translated by an unknown amount. Such perturbed, translated, rotated, and/or reflected K point configurations constitute our data.

The above description can be put in a mathematical form as follows. Let M denote the $K \times D$ landmark coordinate matrix corresponding to the mean form. Let E_i be the $K \times D$ matrix representing the error for the ith individual and we assume E_i is Gaussian with mean matrix $\mathbf{0}$ and variance-covariance matrix $\Sigma_K \otimes \Sigma_D$ where Σ_K is a $K \times K$ positive definite matrix representing the variance among elements within the same column of E_i and Σ_D is a $D \times D$ positive definite matrix representing the variance among elements within the row of E_i. Let Γ_i be an $D \times D$ orthogonal matrix representing rotation and/or reflection of ($M + E_i$), and t_i a $K \times D$ matrix with identical rows representing translation. Then the landmark coordinate matrix corresponding to the ith individual may be represented as $X_i = (M + E_i)\Gamma_i + t_i$. It then follows that

$$X_i \sim MN_{K \times D}(M\Gamma_i + t_i, \Sigma_K, \Gamma_i^T \Sigma_D \Gamma_i)$$

for $i = 1, 2, \ldots$, n. Here "MN" stands for "matrix normal" (Gupta, A.K. and Nagar, D.K., 1999). Parameters of interest are (M, Σ_K, Σ_D) and (Γ_i, t_i) are the nuisance parameters.

2.2 Eliminating the nuisance parameters

To estimate the mean form M and the variance-covariance matrix Σ_K and Σ_D, we need to eliminate the nuisance parameters first. So we consider transforming the data in such a way that the distribution of the transformed data is independent of the nuisance parameters. Lele (1993) and Lele & McCulloch (2002) use a maximal invariant statistics $T(\cdot)$ to eliminate nuisance parameters. They define the maximal invariant as follows. Let S denote the space of all $K \times D$ matrices and let $T(\cdot)$ be a function defined on this space such that for X and X^* in S, $T(X) = T(X^*)$ if and only if X^* is just a rotation, translation, and/or reflection of X. Then $T(\cdot)$ is called a maximal invariant defined on the space S under the group of rotation, translation and reflection of X.

Let $$H = I - \frac{1}{K}(\mathbf{1}^T\mathbf{1})$$ where $\mathbf{1} = (1,1,\cdots 1)$ a $1 \times K$ row vector

be a $K \times K$ centering matrix. Let $X^C = HX$, then the columns of X^C sum to zero. The following theorem gives a maximal invariant of X, a $K \times D$ matrix of landmark coordinates.

Theorem 2.1. $T(X) = HXX^T H^T$ is a maximal invariant statistic, where X is a $K \times D$ matrix.

Proof:

1) $T(X)$ is invariant.

$$T(X\Gamma + t) = H(X\Gamma + t)(\Gamma^T X^T + t^T)H^T = HX\Gamma\Gamma^T H^T = T(X)$$

since t has identical rows and then $Ht = \mathbf{0}$.

2) $T(X)$ is maximal invariant.

To show that it is a maximal invariant, we need to show that, given $T(X)$, it can be mapped back to a unique orbit in the original space. This can be proved using the fact that $T(X)$ is a centered inner product matrix and so there exists a unique (up to rotation, translation, reflection) mapping from the centered inner product matrix to a coordinate matrix (Lele 1991; Lele 1993; Lele and McCulloch, 2002).

Furthermore, it follows from standard multivariate normal distribution theory (Arnold, 1981, Chap. 17, Sect. 3) that if $\Sigma_D = I$

$$B_i = T(X_i) = X_i^C (X_i^C)^T \sim \text{Wishart}_K(D, \Sigma_K^*, MM^T)$$

that is, the random variables B_i s are $K \times K$ matrices and have a Wishart distribution independent of nuisance parameters, where $\Sigma_K^* = H\Sigma_K H^T$ is a $K \times K$ non-negative definite matrix of rank $K-1$ corresponding to the variance of the columns of X_i^C. Lele (1993) shows that Σ_K^* and $M^C(M^C)^T$ are identifiable and provides a consistent estimator of Σ_K^* and $M^C(M^C)^T$ based on the method of moments. Note that $T(M) = M^C(M^C)^T = HMM^T H$ is a centered inner product matrix corresponding to the mean form M. The second point of the proof of *Theorem 1* establishes that estimation of $M^C(M^C)^T$ are equivalent to estimating the mean form. In other words, given $M^C(M^C)^T$ one can construct M (up to translation, rotation, and reflection).

2.2.1 The Estimation of $M^C(M^C)^T$ and Σ_K^*

We use the following notation (Lele, 1993):

(i) $F(X) = [F_{lm}]_{\substack{l=1,2,\dots,K \\ m=1,2,\dots,K}}$ where F_{lm} is the Euclidean distance between landmarks l and m. Euclidean distance is the straight line distance between two points that can be measured by a ruler.

(ii) $Eu(X) = [F_{lm}^2] = [e_{lm}]$ denotes the matrix of squared distances.

(iii) $B(X) = X^C(X^C)^T$ denotes the centered inner product matrix.

(iv) Let $\Sigma_K = [\sigma_{lm}]_{\substack{l=1,2,\ldots,K \\ m=1,2,\ldots,K}}$ be the variance-covariance matrix and,

$Eu(M) = [\delta_{lm}]_{\substack{l=1,2,\ldots,K \\ m=1,2,\ldots,K}}$ be the Euclidean distance matrix corresponding to the mean form M.

The following theorems lead to the consistent moment estimator for δ_{lm}'s. The proof follows from the consistency of the sample moments and the consistency of a continuous function of sample moments (Chung, 1974). See also Johnson and Kotz (1970, Chap.28) for properties of noncentral χ^2 distribution.

Theorem 2.2. $e_{l,m} \sim \phi_{lm}\chi_D^2(\delta_{lm}/\phi_{lm})$ that is, squared Euclidean distances between pairs of landmarks have a non-central χ^2 distribution with D degrees of freedom, noncentrality parameter δ_{lm} and scaling parameter ϕ_{lm} ,where $\phi_{lm} = \sigma_{ll} + \sigma_{mm} - 2\sigma_{lm}$.

Theorem 2.3. For a two-dimensional object,

$$E(e_{l,m}) = 2\phi_{l,m} + \delta_{l,m} = \alpha_1$$
$$Var(e_{l,m}) = 4\phi_{l,m}^2 + 4\delta_{l,m}\phi_{l,m} = \alpha_2$$

and

$$\alpha_1^2 - \alpha_2 = (\delta_{l,m})^2 \qquad (1)$$

We can then equate the sample moments to the population moments to obtain a moment estimator for δ_{lm}.

Theorem 2.4. Let e^i_{lm} denote the squared Euclidean distance between landmarks l and m in the i th object.

$$\text{Let} \quad \overline{e}_{l,m} = \frac{1}{n}\sum_{i=1}^{n} e^i_{lm}$$

$$S^2(e_{l,m}) = \frac{1}{n}\sum_{i=1}^{n}(e^i_{l,m} - \overline{e}_{l,m})^2$$

and

$$\hat{\delta}_{l,m} = [(\overline{e}_{l,m})^2 - S^2(e_{l,m})]^{1/2} \tag{2}$$

Then as $n \to \infty$,

$$\hat{\delta}_{l,m} \to \delta_{l,m} \text{ in probability}$$

We can also obtain the moment estimator of δ_{lm} for three-dimensional objects.

Theorem 2.5.

$$E(e_{l,m}) = 3\phi_{l,m} + \delta_{l,m} = \beta_1$$
$$Var(e_{l,m}) = 6\phi^2_{l,m} + 4\delta_{l,m}\phi_{l,m} = \beta_2$$

and

$$\beta_1^2 - \frac{3}{2}\beta_2 = (\delta_{l,m})^2 \tag{3}$$

Theorem 2.6. Using the same notation as in Theorem 4, and

$$\hat{\delta}_{l,m} = [(\overline{e}_{l,m})^2 - 1.5S^2(e_{l,m})]^{1/2} \tag{4}$$

It follows that as $n \to \infty$,

$$\hat{\delta}_{l,m} \to \delta_{l,m} \text{ in probability}$$

Next theorem utilizes the estimators of δ_{lm} to obtain a consistent estimator of the variance-covariance parameter Σ_K^*. The proof follows from Arnold (1981, Theorem 17.6) and consistency of moments and consistency of continuous function of moments (Chung, 1974).

Theorem 2.7. $E(B(X)) = D\Sigma_K^* + B(M)$ and

$$\hat{\Sigma}_K^* = \frac{1}{D}[\frac{1}{n}\sum_{i=1}^{n} B(X_i) - \hat{B}(M)] \to \Sigma_K^* \text{ in probability}$$

Following the theorems, the algorithm of obtaining $\hat{M}$ and Σ_K^* can be shown as below:

Step 1. Calculate $B(M) = -\frac{1}{2}H\{Eu(M)\}H$.

Step 2. Calculate the eigenvalues and eigenvectors of $B(M)$. Let the eigenvalues be $\lambda_1 > \lambda_2 > \cdots > \lambda_K$ and the corresponding eigenvectors be $h_1, h_2, \cdots, h_K$.

Step 3. The estimator of the centered mean form $\hat{M}^C$ is given by:

For a two-dimensional object: $\hat{M}^C = [\sqrt{\lambda_1}h_1, \sqrt{\lambda_2}h_2]$

For a three-dimensional object: $\hat{M}^C = [\sqrt{\lambda_1}h_1, \sqrt{\lambda_2}h_2, \sqrt{\lambda_3}h_3]$

Step 4. Caluculate $B(X) = -\frac{1}{2}H\{Eu(X)\}H$.

Step 5. The estimator of Σ_K^* is given by $\hat{\Sigma}_K^* = \frac{1}{D}[\frac{1}{n}\sum_{i=1}^{n} B(X_i) - B(M)] \to \Sigma_K^*$.

This shows that $\Sigma_K^* = H\Sigma_K H^T$ is identifiable and estimable.

2.2.2 Conditions for Identifiability of Σ_K

However, biologically Σ_K is of interest. Unfortunately, mapping from Σ_K^* to Σ_K is non-unique because the centering matrix H is singular and hence is not invertible (recall that $\Sigma_K^* = H\Sigma_K H$). To make this mapping unique, we need to impose conditions on Σ_K. In the following, we provide such conditions.

Since the identifiability of the parameters is independent of the choice of the centering matrix (Lele and McCulloch, 2002), we define another centering matrix L of rank $K-1$ for the sake of mathematical convenience. Let L be a $(K-1)\times K$ matrix whose first column consists of -1s and the rest of the matrix is an identity matrix of dimension $(K-1)\times(K-1)$. Let $\tilde{\Sigma}_K = L\Sigma_K L^T$.

Note that Σ_K is a symmetric $K \times K$ matrix of full rank K while $\tilde{\Sigma}_K$ is a $(K-1)\times(K-1)$ matrix of rank $K-1$. One can not go from $\tilde{\Sigma}_K$ to Σ_K in a unique fashion. For example, if

$$\Sigma_K = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

and,

$$\tilde{\Sigma}_K = L\Sigma_K L^T = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} \\ \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{33} - 2\sigma_{13} \end{bmatrix}$$

There are six unknown elements in Σ_K : σ_{11} , σ_{22} , σ_{33} , $\sigma_{12} = \sigma_{21}$, $\sigma_{13} = \sigma_{31}$ and $\sigma_{23} = \sigma_{32}$, but only three equations in $\tilde{\Sigma}_K$:

$$\sigma_{11} + \sigma_{22} - 2\sigma_{12} = a$$
$$\sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} = b$$
$$\sigma_{11} + \sigma_{33} - 2\sigma_{13} = c$$

where a,b,c are some known constant. We can not identify these unknown elements from a system of the three linear equations. Hence Σ_K is non identifiable.

Generally, there are $K(K+1)/2$ unknown elements in Σ_K while we can only get $K(K-1)/2$ linear equations from $\tilde{\Sigma}_K$, so Σ_K is not identifiable. In order to identify Σ_K , we need to put constrains on the structure of it: at least K covariance elements in Σ_K are zero so that the number of unknown elements in Σ_K is reduced to at most $K(K-1)/2$.

For such constructed Σ_K , there is a unique transformation function of $\tilde{\Sigma}_K$ such that $\Sigma_K = f(\tilde{\Sigma}_K)$. We use the following notation:

$vec(X)$: a column vector formed by stringing out the columns of matrix X , one after the other (Magnus, J.R and Neudecker, H., 1999, Chap. 2, Sect. 4).

For example,

$$X = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix},$$

then $$vec(X) = \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix}.$$

Since Σ_K and $\tilde{\Sigma}_K$ are symmetric, stringing out the columns is equivalent to stringing out the rows. Then generally,

$$\Sigma_K = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \cdots & \sigma_{1K} \\ & \sigma_{22} & \cdots & \cdots & \sigma_{2K} \\ & & & \cdots & \cdots \\ & & & & \sigma_{KK} \end{bmatrix}_{(K \times K)}$$

$$vec(\Sigma_K) = \left(\sigma_{11}, \cdots, \sigma_{1K}, \sigma_{21}, \cdots \sigma_{2K}, \cdots\cdots \sigma_{KK}\right)^T_{(K^2 \times 1)}$$

$$\tilde{\Sigma}_K = \begin{bmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \cdots & \cdots & \sigma_{11} + \sigma_{2K} - \sigma_{12} - \sigma_{1K} \\ & \sigma_{11} + \sigma_{33} - 2\sigma_{13} & \cdots & \cdots & \sigma_{11} + \sigma_{3K} - \sigma_{13} - \sigma_{1K} \\ & & \cdots & \cdots & \cdots \\ & & & & \sigma_{11} + \sigma_{KK} - \sigma_{1K} - \sigma_{1K} \end{bmatrix}_{(K-1)\times(K-1)}$$

$$
vec(\tilde{\Sigma}_K) = \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} \\ \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} \\ \sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14} \\ \vdots \\ \sigma_{11} + \sigma_{2K} - \sigma_{12} - \sigma_{1K} \\ \sigma_{11} + \sigma_{32} - \sigma_{12} - \sigma_{13} \\ \sigma_{11} + \sigma_{33} - 2\sigma_{13} \\ \vdots \\ \sigma_{11} + \sigma_{3K} - \sigma_{13} - \sigma_{1K} \\ \vdots \\ \vdots \\ \sigma_{11} + \sigma_{K2} - \sigma_{12} - \sigma_{1K} \\ \sigma_{11} + \sigma_{K3} - \sigma_{13} - \sigma_{1K} \\ \vdots \\ \vdots \\ \sigma_{11} + \sigma_{KK} - \sigma_{1K} - \sigma_{1K} \end{pmatrix}_{(K-1)^2 \times 1}
$$

Let A denote a matrix such that $Avec(\Sigma_K) = vec(\tilde{\Sigma}_K)$. A then has such a form:

$$
\begin{pmatrix}
1 & -2 & 0 & \cdots & \cdots & 0 & 0 & \boxed{1 \ \cdots \ \cdots \ 0} & & \\
1 & -1 & -1 & \cdots & \cdots & 0 & \vdots & \ddots & & \\
1 & -1 & 0 & -1 & \cdots & 0 & \vdots & \ddots & & \textcircled{\mathbf{0}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & \ddots & & \\
1 & -1 & 0 & \cdots & \cdots & -1 & 0 & \boxed{0 \ \cdots \ \cdots \ \cdots \ 1} & & \\
1 & -1 & -1 & \cdots & \cdots & 0 & 0 & & 1 \ \cdots \ \cdots \ \cdots \ 0 & \\
1 & 0 & -2 & \cdots & \cdots & 0 & \vdots & & \ddots & \mathbf{0} \\
1 & 0 & -1 & \cdots & \cdots & 0 & \vdots & \mathbf{0} & \ddots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & & \vdots & & \ddots & \\
1 & 0 & -1 & \cdots & \cdots & -1 & 0 & & 0 \ \cdots \ \cdots \ \cdots \ 1 & \\
\vdots & & & & & & \vdots & \vdots & & \\
\vdots & & & & & & \vdots & \vdots & & \\
\vdots & & & & & & \vdots & \vdots & & \\
1 & -1 & 0 & \cdots & \cdots & -1 & 0 & & & 1 \ \cdots \ \cdots \ 0 \\
1 & 0 & -1 & \cdots & \cdots & -1 & \vdots & & & \ddots \\
1 & 0 & 0 & -1 & \cdots & -1 & \vdots & & \mathbf{0} & \ddots \\
\vdots & & & & \ddots & & \vdots & & & \ddots \\
1 & 0 & 0 & \cdots & \cdots & -2 & 0 & & & 0 \ \cdots \ \cdots \ 1
\end{pmatrix}_{(K-1)^2 \times K^2}
$$

There are $(K-1)\times K$ sub-matrices of dimension $(K-1)\times K$ in A, where

$$
\begin{matrix}
1 & \cdots & \cdots & \cdots & 0 \\
 & \ddots & & & \\
 & & \ddots & & \\
 & & & \ddots & \\
0 & \cdots & \cdots & \cdots & 1
\end{matrix}
$$

denotes a $(K-1)\times K$ submatrix, and

$\mathbf{0}$ denotes a $(K-1)\times K$ submatrix comprising of all 0 elements.

Since both Σ_K and $\tilde{\Sigma}_K$ are symmetric matrix, we can string out only the upper triangle of the matrices Σ_K and $\tilde{\Sigma}_K$ to get revised $vec(\Sigma_K)$ and $vec(\tilde{\Sigma}_K)$ of dimension $\frac{K(K+1)}{2}\times 1$ and $\frac{K(K-1)}{2}\times 1$ respectively. One can reduce A by removing the rows and columns correspondingly. The reduced matrix A is then of dimension $\frac{K(K-1)}{2}\times\frac{K(K+1)}{2}$. Thus $Avec(\Sigma_K)=vec(\tilde{\Sigma}_K)$ gives a system of $\frac{K(K-1)}{2}$ equations in $\frac{K(K+1)}{2}$ unknowns. This system does not have a unique solution. We need to put constrains on Σ_K in order to get a unique solution of the equation $Avec(\Sigma_K)=vec(\tilde{\Sigma}_K)$. Suppose we impose constrains on Σ_K such that at least K covariance elements of are zero. Then we can reduce the size of the system by at least K.

Theorem 2.8.

If (i) A is a known $m\times n$ matrix and has full rank n, $m>n$

(ii) x is a $n\times 1$ vector containing n unknown non zero elements,

(iii) b is a known $m\times 1$ vector,

Then the equation $Ax=b$ has a unique least square solution $x=(A^T A)^{-1}A^T b$.

Proof: It follows from the generalized inverse of matrices theorem (Rao and Mitra, 1971, Chapter 2 or C.W.Groetsch, 1977, 113-116).

Theorem 2.9. Σ_K is identifiable if the corresponding reduced matrix A of dimension $\frac{K(K-1)}{2}\times q$ has full rank and $\frac{K(K-1)}{2}\geq q$.

Proof:

In constrains imposed Σ_K, there are at least K zero covariance elements. We delete all zero elements in $vec(\Sigma_K)$, and then delete the corresponding columns in matrix A. Then the reduced matrix A is full rank and the number of columns in it is at most $\frac{K(K-1)}{2}$. From now on, we refer the final version of A, $vec(\Sigma_K)$ and $vec(\tilde{\Sigma}_K)$ as A, $vec(\Sigma_K)$ and $vec(\tilde{\Sigma}_K)$. $vec(\Sigma_K)$ contains q non-zero elements. Matrix A is of dimension $\frac{K(K-1)}{2} \times q$.

(1) If $\frac{K(K-1)}{2} = q$, then A is a full ranked square matrix, the equation $Avec(\Sigma_K) = vec(\tilde{\Sigma}_K)$ has a unique exact solution: $vec(\Sigma_K) = A^{-1}vec(\tilde{\Sigma}_K)$.

(2) If $\frac{K(K-1)}{2} > q$, then following *Theorem 2.8*, $Avec(\Sigma_K) = vec(\tilde{\Sigma}_K)$ has a unique least square solution $vec(\Sigma_K) = (A^T A)^{-1} A^T vec(\tilde{\Sigma}_K)$.

Hence, $vec(\Sigma_K)$ is identifiable. And since it is simply a stacking of Σ_K, Σ_K is identifiable.

We now illustrate the result for two different model structures: diagonal structure assuming the landmarks are independent and off-diagonal structure with some 0 covariance assuming correlation exists between some landmarks.

(1) Diagonal: $K = 3$ **and** $D = 2$

Let $$\Sigma_K = \begin{bmatrix} \sigma_{11} & & \\ & \sigma_{22} & \\ & & \sigma_{33} \end{bmatrix}$$

Then $$\tilde{\Sigma}_K = L\Sigma_K L^T = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & & \\ & \sigma_{22} & \\ & & \sigma_{33} \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} + \sigma_{22} & \sigma_{11} \\ & \sigma_{11} + \sigma_{33} \end{bmatrix}$$

and $$vec(\Sigma_K) = \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \end{bmatrix}, \qquad vec(\tilde{\Sigma}_K) = \begin{bmatrix} \sigma_{11} + \sigma_{22} \\ \sigma_{11} \\ \sigma_{11} + \sigma_{33} \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

A is square and has full rank 3, then $Avec(\Sigma_K) = vec(\tilde{\Sigma}_K)$ has a unique solution $vec(\Sigma_K) = A^{-1}vec(\tilde{\Sigma}_K)$. Hence Σ_K is identifiable.

(2) Off-diagonal: $K = 4$ and $D = 3$

Let $$\Sigma_K = \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & \sigma_{23} & 0 \\ 0 & \sigma_{32} & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{pmatrix}$$

Then

$$\tilde{\Sigma}_K = L\Sigma_K L^T = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & \sigma_{23} & 0 \\ 0 & \sigma_{32} & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{pmatrix} \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{pmatrix} \sigma_{11}+\sigma_{22} & \sigma_{11}+\sigma_{23} & \sigma_{11} \\ & \sigma_{11}+\sigma_{33} & \sigma_{11} \\ & & \sigma_{11}+\sigma_{44} \end{pmatrix}$$

and $$vec(\Sigma_K) = \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{33} \\ \sigma_{44} \end{bmatrix}, \qquad vec(\tilde{\Sigma}_K) = \begin{bmatrix} \sigma_{11}+\sigma_{22} \\ \sigma_{11}+\sigma_{23} \\ \sigma_{11} \\ \sigma_{11}+\sigma_{33} \\ \sigma_{11} \\ \sigma_{11}+\sigma_{44} \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}_{6\times 5}$$

Matrix A has full rank 5 and its row dimension 6 is greater than its column dimension 5, and then $Avec(\Sigma_K) = vec(\tilde{\Sigma}_K)$ has a unique least square solution:

$$vec(\Sigma_K) = (A^T A)^{-1} A^T vec(\tilde{\Sigma}_K).$$

Hence Σ_K is identifiable.

2.3 The Algorithm to Estimate Σ_K

Section 2.2.2 gives the conditions under which Σ_K is a unique transformation of $\tilde{\Sigma}_K$: $vec(\Sigma_K) = (A^T A)^{-1} A^T vec(\tilde{\Sigma}_K)$. Thus if $\tilde{\Sigma}_K$ is estimable, then Σ_K which satisfies the condition in *Theorem 2.9* is estimable.

Following theorems give a consistent estimator of $\tilde{\Sigma}_K$.

Theorem 2.10. There exists a one-one transformation function such that $\tilde{\Sigma}_K = f(\Sigma_K^*)$.

Proof:

Recall that $\Sigma_K^* = H\Sigma_K H$ and $\tilde{\Sigma}_K = L\Sigma_K L^T$. There exists a matrix Y such that $YH = L$, by solving equation $YH = L$, we can get the structure of matrix Y :

$$\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ -1 & 0 & \cdots & \cdots & 1 & 0 \\ -2 & -1 & -1 & \cdots & -1 & 0 \end{pmatrix}_{(K-1)\times K}$$

and hence $\tilde{\Sigma}_K = L\Sigma_K L^T = YH\Sigma_K HY^T = Y\Sigma_K^* Y^T$.

Theorem 2.11. $\hat{\tilde{\Sigma}}_K = Y\hat{\Sigma}_K^* Y^T$ and $\hat{\tilde{\Sigma}}_K \xrightarrow{P} \tilde{\Sigma}_K$.

Proof:

We give a consistent estimator of Σ_K^* in section 2.2.1. That is,

$$\hat{\Sigma}_K^* \xrightarrow{P} \Sigma_K^*$$

Since $\tilde{\Sigma}_K = Y\Sigma_K^* Y^T$, then

$$\hat{\tilde{\Sigma}}_K = Y\hat{\Sigma}_K^* Y^T \xrightarrow{P} Y\Sigma_K^* Y^T = \tilde{\Sigma}_K$$

This follows from the consistency of moments and the consistency of continuous function of moments (Chung, 1974).

Theorem 2.12. For constrains imposed Σ_K which satisfies the conditions in *Theorem 2.9,* $\hat{\Sigma}_K \xrightarrow{P} \Sigma_K$.

Proof:

(1) If A is a full rank square matrix, then $vec(\Sigma_K) = A^{-1} vec(\tilde{\Sigma}_K)$. It follows from the consistency of moments and the consistency of continuous function of moments (Chung, 1974) that

$$vec(\hat{\Sigma}_K) = A^{-1} vec(\hat{\tilde{\Sigma}}_K) \xrightarrow{P} A^{-1} vec(\tilde{\Sigma}_K) = vec(\Sigma_K)$$

since $vec(\hat{\tilde{\Sigma}}_K) \xrightarrow{P} vec(\tilde{\Sigma}_K)$ from *theorem 2.11*. Note that $vec()$ is simply a stacking operation of a matrix.

(2) If the row dimension of A is greater than the column dimension of A, i.e. $\frac{K(K-1)}{2} > q$, then the unique least square solution is

$$vec(\Sigma_K) = (A^T A)^{-1} A^T vec(\tilde{\Sigma}_K).$$

Again, following (Chung, 1974), $vec(\hat{\Sigma}_K) \xrightarrow{P} vec(\Sigma_K)$.

Given the model structure of Σ_K and considering the way of forming $vec(\Sigma_K)$, one can simply transform $vec(\hat{\Sigma}_K)$ back to $\hat{\Sigma}_K$ and $\hat{\Sigma}_K \xrightarrow{P} \Sigma_K$. Hence we get a consistent estimator of the covariance matrix of landmarks data.

Chapter 3

Performance Study of EDMA using Simulated Data

In this chapter, we study the performance of the estimators derived in Chapter 2 by using simulation.

3.1 Simulation Design

We consider two different model structures of Σ_K and two different dimensions $D = 2$ and $D = 3$. First we consider the diagonal model structure with 2 dimensional data. Then we consider the diagonal model structure with 3 dimensional data. We then move on to off-diagonal model structure first with 2 dimensional data and later with 3 dimensional data.

For each case, we randomly generate 5,000 matrix normal random variables $X_i^{'}s$ with known mean form and known variance-covariance matrix. Following are the steps of generating landmarks data.

Step 1: generate a $K \times D$ standard normal matrix $Z \sim MN(\mathbf{0}, I)$

Step 2: find cholesky decomposition C such that $\Sigma_K = CC^T$

Step 3: $X = (CZ + M)I, X \sim MN(M, CC^T = \Sigma_K, I)$

Step 4: estimate Σ_K

Step 5: repeat step 4 100 times and hence get $100\,\hat{\Sigma}_{Ki}$'s, where $i = 1,2,\cdots 100$

3.2 Simulation Results

(1) Diagonal matrix case: K=3 D=2

The true mean form matrix is $M = \begin{pmatrix} 2.70 & 4.72 \\ 7.07 & -2.36 \\ -1.53 & 2.59 \end{pmatrix}$.

The true covariance matrix is $\Sigma_K = \begin{pmatrix} .87 & 0 & 0 \\ 0 & .59 & 0 \\ 0 & 0 & .42 \end{pmatrix}$ $\Sigma_D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

And the true centered inner product matrix corresponding to the mean form M is

$$M^C (M^C)^T = \begin{pmatrix} 9.427 & -12.512 & 3.085 \\ & 34.771 & -22.259 \\ & & 19.173 \end{pmatrix}$$

Mean of the 100 estimates of $M^C(M^C)^T$ and Σ_K obtained in *Step 5* are as follows:

$$\hat{M}_C \hat{M}_C^T = \begin{pmatrix} 9.44 & -12.50 & 3.06 \\ & 34.71 & -22.21 \\ & & 19.15 \end{pmatrix} \qquad \hat{\Sigma}_K = \begin{pmatrix} .856 & 0 & 0 \\ 0 & .569 & 0 \\ 0 & 0 & .422 \end{pmatrix}$$

This shows that EDMA based estimators are consistent and asymptotically unbiased.

(2) Diagonal matrix case: K=4 D=3

The true mean form is $M = \begin{pmatrix} 2.70 & 4.72 & 7.07 \\ -2.36 & -1.53 & 2.59 \\ 8.62 & 1.10 & 2.63 \\ 4.98 & 7.43 & 5.21 \end{pmatrix}$

The true covariance matrix is $\Sigma_K = \begin{pmatrix} .87 & 0 & 0 & 0 \\ 0 & .59 & 0 & 0 \\ 0 & 0 & .42 & 0 \\ 0 & 0 & 0 & .63 \end{pmatrix}$ $\Sigma_D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

And the true centered inner product matrix corresponding to the mean form M is

$$M^C(M^C)^T = \begin{pmatrix} 11.083 & -8.206 & -12.009 & 9.132 \\ & 57.242 & -18.737 & -30.299 \\ & & 32.762 & -2.015 \\ & & & 23.182 \end{pmatrix}$$

Mean of the 100 estimates of $M^C(M^C)^T$ and Σ_K obtained in *Step 5* are as follows:

$$\hat{M}^C(\hat{M}^C)^T = \begin{pmatrix} 11.03 & -8.16 & -12.02 & 9.15 \\ & 57.09 & -18.73 & -30.20 \\ & & 32.78 & -2.03 \\ & & & 23.07 \end{pmatrix} \quad \hat{\Sigma}_K = \begin{pmatrix} .89 & 0 & 0 & 0 \\ 0 & .56 & 0 & 0 \\ 0 & 0 & .43 & 0 \\ 0 & 0 & 0 & .64 \end{pmatrix}$$

This shows that EDMA based estimators are consistent and asymptotically unbiased.

(3) Off-diagonal matrix case: K=4, D=3

The true mean form is $M = \begin{pmatrix} 2.70 & 4.72 & 7.07 \\ -2.36 & -1.53 & 2.59 \\ 8.62 & 1.10 & 2.63 \\ 4.98 & 7.43 & 5.21 \end{pmatrix}$

The true covariance matrix $\Sigma_K = \begin{pmatrix} .78 & 0 & 0 & 0 \\ 0 & .66 & .40 & 0 \\ 0 & .40 & .87 & 0 \\ 0 & 0 & 0 & .53 \end{pmatrix}$ $\Sigma_D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

And the true centered inner product matrix corresponding to the mean form M is

$$M^C(M^C)^T = \begin{pmatrix} 11.083 & -8.206 & -12.009 & 9.132 \\ & 57.242 & -18.737 & -30.299 \\ & & 32.762 & -2.015 \\ & & & 23.182 \end{pmatrix}$$

Mean of the 100 estimates of $M^C(M^C)^T$ and Σ_K obtained in *Step 5* are as follows:

$$M^C(M^C)^T = \begin{pmatrix} 11.086 & -8.224 & -11.953 & 9.091 \\ & 57.164 & -18.663 & -30.277 \\ & & 32.659 & -2.044 \\ & & & 23.230 \end{pmatrix}$$

$$\hat{\Sigma}_K = \begin{pmatrix} .77 & 0 & 0 & 0 \\ 0 & .69 & .43 & 0 \\ 0 & .43 & .88 & 0 \\ 0 & 0 & 0 & .54 \end{pmatrix}$$

This shows that EDMA based estimators are consistent and asymptotically unbiased.

(4) Off-diagonal matrix case: K=5, D=2

The true mean form is $M = \begin{pmatrix} 2.70 & 4.72 \\ 7.07 & 6.36 \\ 8.53 & 2.59 \\ 10.62 & 6.70 \\ 13.68 & 8.98 \end{pmatrix}$

The true covariance matrix is $\Sigma_K = \begin{pmatrix} .66 & 0 & 0 & 0 & 0 \\ 0 & .58 & 0 & .39 & 0 \\ 0 & 0 & .47 & 0 & 0 \\ 0 & .39 & 0 & .73 & 0 \\ 0 & 0 & 0 & 0 & .82 \end{pmatrix}$ $\quad \Sigma_D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

And the true centered inner product matrix corresponding to the mean form M is

$$M^C(M^C)^T = \begin{pmatrix} 35.19 & 7.88 & 3.71 & -13.18 & -33.61 \\ & 2.34 & -1.62 & -2.64 & -5.96 \\ & & 10.76 & -2.70 & -10.15 \\ & & & 5.10 & 13.42 \\ & & & & 36.30 \end{pmatrix}$$

Mean of the 100 estimates of $M^C(M^C)^T$ and Σ_K obtained in *Step 5* are as follows:

$$\hat{M}_C\hat{M}_C^T = \begin{pmatrix} 35.16 & 7.86 & 3.73 & -13.11 & -33.63 \\ & 2.32 & -1.60 & -2.62 & -5.96 \\ & & 10.87 & -2.72 & -10.28 \\ & & & 5.06 & 13.40 \\ & & & & 36.48 \end{pmatrix}$$

$$\hat{\Sigma}_K = \begin{pmatrix} .68 & 0 & 0 & 0 & 0 \\ 0 & .57 & 0 & .38 & 0 \\ 0 & 0 & .48 & 0 & 0 \\ 0 & .38 & 0 & .73 & 0 \\ 0 & 0 & 0 & 0 & .83 \end{pmatrix}$$

This shows that EDMA based estimators are consistent and asymptotically unbiased.

For all of the four cases, EDMA gives fairly accurate, consistent estimators of both the mean form and the covariance matrix of landmarks data. How good does it perform on the real world data? We analyze the mouse mandible data in next Chapter.

Chapter 4

Analysis of Variability in Mouse Mandibles

The evolution of morphological structures by natural selection depends on the availability of genetic variation for the traits in question. Particularly for multidimensional features such as shape, the response to selection depends critically on the patterns of genetic and phenotypic variation. Therefore estimation of those covariance matrices has long been a central part of evolutionary quantitative genetics (Klingenberg and Leamy, 2001).

The rodent mandible is composed of several parts that are morphologically recognizable and have distinct developmental origins, and it has long been used as a model for genetics, development, and evolution of complex morphological structures (Atchley and Hall, 1991; Hall 1999). Following data were provided by Professor Jim Cheverud of department of Anatomy and Neurobiology, Washington University School of Medicine, Saint Louis, Missouri (U.S.A). The two-dimensional data coordinates of the mandibular landmarks were obtained from lateral views of the right hemi-mandible using a digital video data collection system (Cheverud et al., 1997). Distances were measured in millimeters (mms).

The data are the positions of 16 landmarks on the mandibles of mice. Figure 1.1 shows a diagram of a mouse mandible, where black circles indicate the locations of 16 landmarks. Description of these landmarks is provided in Table 4.1. Figure 4.1 shows the configuration of 16 landmarks for one subject.

Table 4.1: 16 mouse mandible landmarks

Landmark Number	Landmark name
1	Coronoid tip
2	Anterior condylar facet
3	Posterior condylar facet
4	Inferior condylar facet
5	Deepest point of mandibular notch
6	Posterior angular process
7	Inferior angular process
8	Anterior angular process
9	Posterior inferior corpus
10	Anterior inferior corpus
11	Inferior incisor alveolus
12	Superior incisor alveolus
13	Deepest point of incisive notch
14	Anterior molar alveolus
15	Posterior molar alveolus
16	Coronoid base

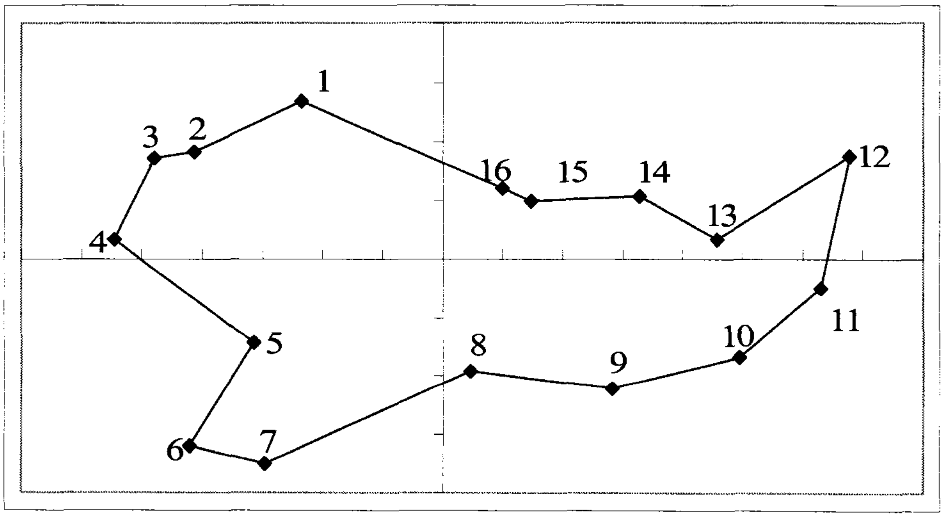

Figure 4.1: The configuration of 16 mandible landmarks (measured in mm)

To estimate the mean form and variance parameter, we assume that Σ_D is an identity matrix and Σ_K is a diagonal matrix, i.e. these 16 mandible landmarks vary independently with each other. Following are the estimators of Σ_K and M in Table 4.2.

$$\hat{\Sigma}_K = diag\,(.042,.017,.024,.007,.004,.038,.041,.036,.056,.068,.102.105,.072,.024,.024,.016)$$

Table 4.2: The estimated mean form matrix of the 16 landmarks data

Landmark number	Coordinate x	Coordinate y
1	-2.5642147	-2.9238803
2	-4.3598175	-2.1062128
3	-5.0189366	-2.0386707
4	-5.7237441	-0.6476102
5	-3.4644795	1.1693193
6	-4.5855937	2.9159961
7	-3.3321319	3.2438449
8	0.1369185	1.7688938
9	2.4780984	2.1174301
10	4.6209991	1.6495556
11	6.0020391	0.5078821
12	6.5324078	-1.7178527
13	4.2879408	-0.3604548
14	3.0344353	-1.1350108
15	1.2218793	-1.1121225
16	0.7341997	-1.3311070

Visualizing of the estimated mean form is shown in Figure 4.2. The circles around the landmarks points denote the variability around them. The radiant of the circle is $r = 2\sqrt{\hat{\sigma}_{ii}}$, where $i = 1,2,\cdots,16$.

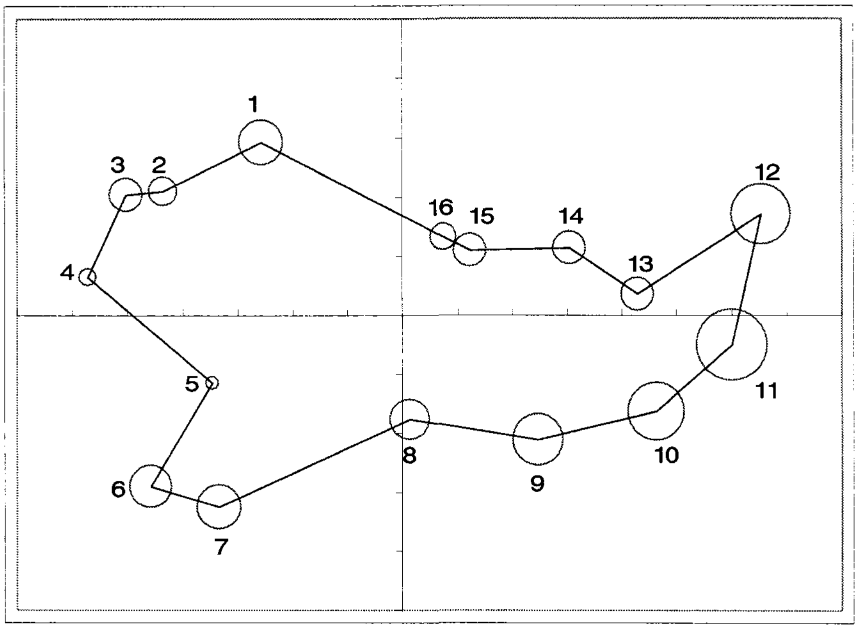

Figure 4.2: Estimated Mean Form (measured in mm)

From Figure 4.2, we can see that the landmarks on the incisor alveolus region (the right part of the mandible) generally have bigger variability that the landmarks on the angular process region (the left part of the mandible). This may indicate that the tissues or organisms around the incisor alveolus region tend to vary while the ones around the angular process region tend to remain stable among mouse individuals. However, we are not equipped to provide full and proper biological interpretation at this time. We leave that to our biologist colleagues.

Chapter 5

Comparison of EDMA with General Procrustes Analysis Method

Aside from EDMA, there are a few other methods that have been developed to estimate the mean form and variability of landmarks. One of the most widely used methods is Generalized Procrustes Analysis (GPA) (Goodall, 1991; Theobald, 2006).

In this chapter, we first provide a brief overview of GPA. We then compare the estimators of M and Σ_K based on GPA with the method developed in this thesis.

5.1 Generalized Procrustes Analysis

GPA employs superimposition to estimate form or shape difference. The superimposition technique involves three steps (Richtsmeier et al., 2001)

1) Fix one of the mean forms in a particular orientation and call it the reference object.

2) Translate and rotate the other mean form so that it matches the reference object according to some criterion.

3) Study the magnitude and direction of differences between forms at each landmark.

Different criteria for matching provide different superimpositions. The least squares criterion leads to a GPA. Goodall (1991) provide an iterative algorithm to estimate M , Γ_i , t_i , Σ_D and Σ_K . The mean form and the vaiance-covariance parameters are estimated by

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i$$

$$\hat{\Sigma}_D = \frac{1}{nK}\sum_{i=1}^{n}(\tilde{X}_i - \hat{M})^T(\tilde{X}_i - \hat{M})$$

$$\hat{\Sigma}_K = \frac{1}{nD}\sum_{i=1}^{n}(\tilde{X}_i - \hat{M})(\tilde{X}_i - \hat{M})^T / tr(\hat{\Sigma}_D)$$

where $\tilde{X}_i$ are least square superimposition matrices.

Theobald and Wuttke (2006) provide a modified version of this algorithm. They claim that these estimators are consistent. However, notice that the number of observations is n and the number of parameters is $2n+3$. Following Neyman and Scoff (1948), Lele (1993) showed that these estimators are in fact inconsistent.

5.2 Comparison of EDMA with GPA

We use simulations to compare the performance of EDMA estimators with the GPA estimators as described by Theobald and Wuttke (2006). We randomly generated 5,000 matrix normal random variables $X_i^{'}s$ with known mean form and known variance-covariance matrix. Here we only consider the diagonal model structure for $D = 3$. And the eigenvalues of Σ_K are generated from inverse gamma distribution to guarantee the invertibility of the estimated landmark covariance matrix (Theobald and Wuttke, 2006).

True $M^C(M^C)^T$ and Σ_K and the estimators of $M^C(M^C)^T$ and Σ_K by EDMA and GPA are given in Table 5.1.

Table 5.1 Comparison of estimators by EDMA and GPA with the truth.

	EDMA	Truth	GPA
$M^C(M^C)^T$	$\begin{bmatrix} 75.46 & 24.85 & 25.18 & 24.99 & -150.48 \\ 24.85 & 75.24 & -25.48 & -25.15 & -49.45 \\ 25.18 & -25.48 & 74.92 & -24.70 & -49.92 \\ 24.99 & -25.15 & -24.70 & 75.11 & -50.25 \\ -150.48 & -49.45 & -49.92 & -50.25 & 300.10 \end{bmatrix}$	$\begin{bmatrix} 75 & 25 & 25 & 25 & -150 \\ 25 & 75 & -25 & -25 & -50 \\ 25 & -25 & 75 & -25 & -50 \\ 25 & -25 & -25 & 75 & -50 \\ -150 & -50 & -50 & -50 & 300 \end{bmatrix}$	$\begin{bmatrix} 76.05 & 25.40 & 24.28 & 25.64 & -151.37 \\ 25.40 & 75.25 & -25.44 & -25.05 & -50.16 \\ 24.28 & -25.44 & 75.87 & -25.77 & -48.94 \\ 25.64 & -25.05 & -25.77 & 76.02 & -50.83 \\ -151.3 & -50.16 & -48.94 & -50.83 & 301.30 \end{bmatrix}$
Σ_K	$\begin{bmatrix} .496 & & & & \\ & .726 & & & \\ & & 1.978 & & \\ & & & .316 & \\ & & & & .382 \end{bmatrix}$	$\begin{bmatrix} 0.50 & & & & \\ & 0.71 & & & \\ & & 1.87 & & \\ & & & 0.35 & \\ & & & & 0.39 \end{bmatrix}$	$\begin{bmatrix} 0.38 & -0.02 & -0.03 & -0.05 & -0.04 \\ -0.02 & 0.44 & 0.09 & -0.09 & -0.04 \\ -0.03 & 0.09 & 1.88 & 0.05 & -0.05 \\ -0.05 & -0.09 & 0.05 & 0.44 & -0.03 \\ -0.04 & -0.04 & -0.05 & -0.03 & 0.01 \end{bmatrix}$

Remarks:

EDMA gives fairly accurate, consistent estimators of both the mean form and the covariance matrix of landmarks data. The mean form estimator given by GPA is OK, but the estimator of the covariance matrix is not close to the truth. Especially, notice that $\hat{\sigma}_{55} \approx 0$ by GPA (purple circle in Table 5.1). GPA uses a singular matrix to estimate a non-singular matrix.

Theobald and Wuttke (2006) claim that as $K \to \infty$ and $N \to \infty$, the modified GPA gives consistent estimators of M and $\Sigma = \Sigma_K \otimes \Sigma_D$, but no complete mathematical proof is provided. We feel that the condition that the number of landmarks converging to infinity is an unrealistic condition in practice. For small number of landmarks, we have shown that GPA estimators are inconsistent, whereas EDMA estimators are shown to be consistent.

Chapter 6

Summary

In this thesis, we discussed the general concepts of "form" and "landmarks", and explained why estimation of the variance of landmarks data is important in many fields such as evolution of organism and genetics, etc.

Due to the presence of nuisance parameters of rotation and translation, general covariance matrix is known to be not identifiable. But certain structured covariance matrices can be shown to be identifiable. In this thesis, we provided conditions under which the covariance matrix of landmarks data is identifiable. Furthermore, we provided a computationally simple approach based on Euclidean Distance Matrix Analysis (EDMA) to estimate the mean form and covariance matrix. It is shown that the estimators are consistent.

We studied the performance of EDMA estimators with the simulated data and the mouse mandibles data. We showed that EDMA gives consistent and asymptotically unbiased estimators of both the mean form and the covariance matrix. We also compared EDMA with GPA and showed that the estimator of the covariance matrix based on GPA is inconsistent.

Bibliography

Arnold, S.F. (1981). Theory of Linear Models and Multivariate Analysis. New York: John Wiley and Sons.

Bookstein, F.L. (1986). Size and Shape Spaces for Landmark Data in Two Dimensions, Statistical Science, 1, 181-242.

Bookstein, F.L. (1991). Morphometric tools for landmark data: geometry and biology. Cambridge: Cambridge University Press.

Campbell, G. (1986). Comments on "Size and Shape Spaces for Landmark Data in Two Dimensions" by F.L. Bookstein, Statistical Science, 1, 227-228.

Cheverud, J.M and Routman, E.J. and Irschick, D.J. (1997) Pleiotropic Effects of Individual Gene Loci on Mandibular Morphology, Evolution, 51(6), 2006-2016.

Chung, K.L. (1994). A Course in Probability Theory. New York: Academic Press.

Dryden, I.L. and Mardia, K.V. (1998). Statistical Shape Analysis. New York: Wiley.

Godzik, A. (1996). The Structural Alignment Between Two Proteins: Is There A Unique Answer? Protein Science, 5, 1325–1338.

Goodall, C.R. (1991). Procrustes Methods in Statistics Analysis Shape, Journal of the Royal Statistical Society Series B - Statistical Methodology, 53, 285-339.

Gould, S.J. (1977). Ontogeny and Phylogeny. Cambridge, MA: Harvard University Press.

Groetsch, C.W. (1977). Generalized Inverses of Linear Operators: Representation and Approximation. New York: Dekker.

Gupta, A.K. and Nagar, D.K. (1999). Matrix Variate Distributions. Chapman and Hall: CRC press.

Johnson, N. and Kotz, S. (1970). Distributions in Statistical – Continuous Univariate Distributions 2. Boston: Houghton Mifflin Company.

Klingenberg, C.P. and Leamy, L. (2001). Quantitative Genetics of Geometric Shape in the Mouse Mandible, Evolution, 55(11), 2342-2352.

Lande, R. (1979). Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry, Evolution, 33, 402-416.

Lele, S. (1991). Some Comments on Coordinate Free and Scale Invariant Method in Morphometrics, American Journal of Physical Anthropology, 85, 407-418.

Lele, S. (1993). Euclidean Distance Matrix Analysis (EDMA): Estimation of Mean Form and Mean Form Difference, Mathematical Geology, 25, 573-601.

Lele, S. and McCulloch, C. (2002). Invariance, Identifiability, and Morphometrics, Journal of the American Statistical Association, 97, 796-804.

Lele, S. and Richtsmeier, J.T. (1991). Euclidean Distance Matrix Analysis: A Coordinate free Approach for Comparing Biological Shapes, American Journal of Physical Anthropology, 86, 415-427.

Lele, S. and Richtsmeier, J.T. (1992). On Comparing Biological Shapes: Detection of Influential Landmarks, American Journal of Physical Anthropology, 87, 49-65.

Lele, S. and Richtsmeier, J.T. (2001). An invariant approach to statistical analysis of shapes. Interdisciplinary studies in statistics series. London: Chapman and Hall - CRC press.

Magnus, J.R. and Neudecker, H. (1999). Matrix Differential Calculus with Applications in Statistics and Econometrics 2. New York: John Wiley & Sons, Inc.

Mardia, K.V. and Dryden, I. (1989). The Statistical Analysis of Shape Data, Biometrika, 76, 271-282.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations, Econometrica, 16, 1-32.

Rao, C.R. and Mitra, S.K. (1971). Generalized Inverse of Matrices and its Applications. New York: John Wiley & Sons, Inc.

Rao, C. and Suryawanshi, S. (1996). Statistical analysis of shape of objects based on landmark data, Proceedings of the National Academy of Sciences of the United States of America, 93, 12132-12136.

Reyment, R., Blackith, R. and Cambell, N. (1984). Multivariate morphometrics. London: Academic Press.

Richtsmeier, J.T., DeLeon, V.B. and Lele, S. (2002). The Promise of Geometric Morphometrics, Yearbook of Physical Anthropology, 45, 63-91.

Theobald, D. and Wuttke, D. (2006). Empirical Bayes Hierarchical Models for Regularizing Maximum Likelihood Estimation in the Matrix Gaussian Procrustes Problem, Proceedings of the National Academy of Sciences of the United States of America, 103, 18521-18527.

Appendix: Program written in R

```
#######generate matrix normal random variables X's######

n<-5000

Z<-matrix(nrow=n*K,ncol=D)

for (i in 1:n){

Z[((i-1)*K+1):(i*K),]<-matrix(round(rnorm(K*D),3),nrow=K,ncol=D)}

C<-chol(SigmaK)

X<-matrix(nrow=n*K,ncol=D)

for (i in 1:n){

X[((i-1)*K+1):(i*K),]<- crossprod(C,Z[((i-1)*K+1):(i*K),])+Meanform}

vartest<- function(x,y,z,w) {

X<-read.table(x)

n<-y

K<-z

D<-w

######### estimate centered mean form and SigmaKstar##########

#step1: calculate Euclidean distance for X

d<-function(x,y){

sum((x-y)^2)}

EuX<-matrix(nrow=n*K,ncol=K)

for (i in 1:n){

 for (j in ((i-1)*K+1):(i*K)){

  for (l in ((i-1)*K+1):(i*K)){
```

```
   EuX[j,(l-(i-1)*K)]<-d(X[j,],X[l,])}}}
#EuX


#step 2 calculate Eu(M)
EuM<-matrix(nrow=K,ncol=K)
for (i in 1:K){
 for (j in 1:K){sum1<-0
  for (l in 1:n){
  sum1<-sum1+EuX[((l-1)*K+i),j]}
  mean.squ.eudis<-sum1/n
  sum2<-0
  for (l in 1:n){
  sum2<-sum2+(EuX[((l-1)*K+i),j]-mean.squ.eudis)^2}
  var.squ.eudis<-sum2/n
if (D==2) EuM[i,j]<-sqrt((mean.squ.eudis)^2-var.squ.eudis)
else EuM[i,j]<-sqrt((mean.squ.eudis)^2-1.5*var.squ.eudis)}}
EuM


#step 3 calculate B(M)
I<-diag(1,K)
ones<-array(rep(1,K),c(1,K))
H<-I-(1/K)*crossprod(ones,ones)
BM<-(-1/2)*H%*%EuM%*%H
```

```
BM


#step 4 calculate eigenvalues and eigenvectors of B(M)

eigen(BM)


#step 5 estimate centred mean form M

est.M<-matrix(nrow=K,ncol=D)

for (i in 1:D){

est.M[,i]<-sqrt(eigen(BM)$values[i])*eigen(BM)$vectors[,i]}


#step 6 estimate SigmaKstar=H*SigmaK*H

BX<-matrix(nrow=n*K,ncol=K)

for (i in 1:n){

BX[((i-1)*K+1):(i*K),]<-(-1/2)*H%*%EuX[((i-1)*K+1):(i*K),]%*%H}

a<-0

for (i in 1:n){

a<-a+BX[((i-1)*K+1):(i*K),]}

est.SigmaKstar<-(a/n-BM)/D


# compare HM with est.M ; HMt(M)H with est.M*t(est.M); H*SigmaK*H with
est.SigmaKstar

H%*%Meanform; est.M

H%*%Meanform%*%t(Meanform)%*%H; est.M%*%t(est.M)
```

```
H%*%SigmaK%*%H; est.SigmaKstar

################Use maple to solve Y such that YH=L#################
fcol<-array(c(rep(-1,K-2),-2),c(K-1,1))
identity<-diag(1,K-2,K-1)
augment<-array(c(rep(-1,K-2),0),c(1,K-1))
Y<-cbind(fcol,rbind(identity,augment))
Y

#  from est.SigmaKstar to est.SigmaK~
est.SigmaKtilde<-Y%*%est.SigmaKstar%*%t(Y); est.SigmaKtilde
L<-cbind(c(rep(-1,K-1)),diag(1,K-1))
#L%*%SigmaK%*%t(L)#

######### model for diagonal matrix #################

c<-as.vector(est.SigmaKtilde)
A<-matrix(nrow=(K-1)^2,ncol=K)
iden<-diag(1,K-1)
for (i in 1:(K-1)){
A[(i-1)*K+1,]<- c(1,iden[i,])
}
for (i in 1:(K-2)){
```

```
A[((i-1)*K+2):(i*K),]<-matrix(rep(c(1,rep(0,K-1)),K-1),nrow=K-1,byrow=T)}

###A###

est.vect<- round((solve(t(A)%*%A))%*%t(A)%*%c,3)

est.SigmaK<-diag(c(est.vect))


####################general case######################

c<-as.vector(est.SigmaKtilde)

b.ori<-as.vector(SigmaK)

m<-0;b.dis<-rep(0,K*(K+1)/2)

for(i in 1:K){

for(j in i:K){

m=m+1;

b.dis[m]<-SigmaK[i,j]}}

#b.dis

d<-0

for (i in 1:(K*(K+1)/2)){

if( b.dis[i]!=0) d<-c(d,b.dis[i])}

b<-d[-1];b

c.dis<-rep(0,K*(K-1)/2)

q<-0

for(i in 1:(K-1)){

for(j in i:(K-1)){

q=q+1;
```

```
c.dis[q]<-est.SigmaKtilde[i,j]}}
iden<-diag(1,K-1)
A<-matrix(nrow=(K-1)^2,ncol=K^2)
A[,K+1]<-0
for (i in 1:(K-1)){
A[,1]<-1
A[((i-1)*(K-1)+1):(i*(K-1)),2:K]<-diag(-1,K-1)
A[((i-1)*(K-1)+1):(i*(K-1)),i+1]<--1
A[(i-1)*K+1,i+1]<--2
A[((i-1)*(K-1)+1):(i*(K-1)),(i*K+2):((i+1)*K)]<-iden
A[((i-1)*(K-1)+1):(i*(K-1)),c(-(1:(K+1)),-((i*K+2):((i+1)*K)))]<-0}
pcol<-rep(0,K*(K-1)/2)
t<-1
for (i in 1:(K-1)){
for (j in 1:i){
pcol[t]<-i*K+j
t<-t+1}}
B<-matrix(nrow=(K-1)^2,ncol=K*(K+1)/2)
pcol.new<-c(pcol,0)
i=1;ColB=1
for (ColA in 1:K^2){
if (ColA==pcol.new[i]) i<-i+1 else {B[,ColB]<-A[,ColA];ColB<-ColB+1}
}
```

```
prow<-rep(0,(K-2)*(K-1)/2)
t<-1
for (i in 1:(K-2)){
for (j in 1:i){
prow[t]<-i*(K-1)+j
t<-t+1}}
C<-matrix(nrow=K*(K-1)/2,ncol=K*(K+1)/2)
prow.new<-c(prow,0)
i=1;RowC=1
for (RowB in 1:(K-1)^2){
if (RowB==prow.new[i]) i<-i+1 else {C[RowC,]<-B[RowB,];RowC<-RowC+1}
}

Q<-0
for (i in 1:(K*(K+1)/2)){
if (b.dis[i]!=0)
Q<-cbind(Q,C[,i])}
A.aug<-Q[,-1]
A.aug
#########calculate the rank of a matrix(by using SVD)################
matrix.rank <- function(A, eps=.Machine$double.eps){
    sv. <- abs(svd(A)$d)
    sum((sv./max(sv.))>eps)
```

```
}

matrix.rank(A.aug)
Kstar<-length(b);Kstar
if (nrow(A.aug)!=ncol(A.aug)) {est.vect<- solve(A.aug,c.dis)} else
 if (ncol(A.aug)==Kstar) {est.vect<-
(solve(t(A.aug)%*%A.aug))%*%t(A.aug)%*%c.dis} else cat(" SigmaK is not
identifiable\n")
est.vect

##general case est.SigmaK##
est.SigmaK<-matrix(nrow=K,ncol=K)
t=1
for (i in 1:K){
for (j in i:K){
if (SigmaK[i,j]==0) {est.SigmaK[i,j]<-0;est.SigmaK[j,i]<-0}
else{est.SigmaK[i,j]<-est.vect[t];est.SigmaK[j,i]<-est.vect[t];t<-t+1}
}}
out<-new.env()
out$MMt<-BM
out$M<-est.M
out$SigmaK<-est.SigmaK
output<-as.list(out)}
```