Tina Memo No. 2011-001
Internal Report

Measurement Covariance Structure for Pseudo Landmarks

N.A. Thacker, H. Ragheb

Last updated
25 / 1 / 2011

ISBE, Medical School,
University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT, UK.

Measurement Covariance Structure for Pseudo Landmarks

N.A. Thacker and H. Ragheb
ISBE, Medical School,
University of Manchester,
Manchester, M13 9PT, UK.
neil.thacker@manchester.ac.uk

Abstract

In previous documents we have outlined an approach for the analysis of 2 and 3D shape vectors using measurement covariances for the situation where the measurements are known to be independent. The purpose of this document is to explain the mathematics required to extend this approach to "pseudo-landmarks", where information is provided at a given 2 or 3D location only in 1D. We show that the use of a pseudo-inverse together with generalisation of the measurement process is sufficient to solve the task.

Introduction

We wish to determine a measurement covariance C for a set of measured locations which define a measurement vector, for the purposes of constructing an approximate weighted linear model. In previous work we explained that for 2D independent landmarks the appropriate covariance could be estimated from the residuals from the model, such that each landmark contributed a 2x2 matrix to the diagonal of C.

In many situations, landmarks are defined in such a way that the measured locations depend upon the other landmarks, for example using two points to describe a reference line from which one dimension of location is defined. In these cases the required C matrix is no longer simply 2x2 diagonal, but has off diagonal correlation terms. Under these circumstances we can no longer simply zero the off-diagonal terms in C. In addition we expect the uninformative components of 2 and 3D measurements to destabilise the estimation of both shape alignments and C. In order to reliably estimate (and represent) C we therefore need to know how the structure of these correlations is determined by the definition of these "pseudo-landmarks" and then treat these factors appropriately.

The approach taken here is to analyse the simplest 1D system and then use basic mathematical arguments to extend our understanding to 2 and 3D. The key issues to be addressed are those of scientific reproducibility, we wish to construct a method for measurement of arbitrary shape, which is not dependant upon the specific choice of landmarks. In particular, we need to show that landmarks (or specific directional measurements pertaining to them) which contain no additional information do not modify the result of subsequent statistical analysis.

1 Method

1.1 1 D Example

We start by considering two independent landmark points A and B in 1D with the points vector $X = (A, B)$ and the 2x2 covariance matrix

$$C_2 = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix}$$

Imagine now that A and B are part of a shape description based upon examples of real world objects. Given the arbitrary nature of landmark selection, we could just have easily selected a third measurement, which (unknown to us) was entirely defined by the first two. What will be the effect on our model and statistical analysis of the existence of the additional point?

We proceed by adding a third point D, where $D = f(X) = (A + B)/2$, that is correlated with A and B and therefore provides no new information. Error propagation can now be used to compute the new 3x3 covariance matrix C_3 for the new points vector $Y = (A, B, D)$, i.e.

$$C_3 = \frac{df(x)^T}{dx} C_2 \frac{df(x)}{dx} \tag{1}$$

Although only the first derivative term is taken into account from the Taylor series, this is exact as the case we consider here is linear.

$$\frac{df(x)}{dx} = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{pmatrix}$$

$$\frac{df(x)^T}{dx} C_2 = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \\ \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_B^2 \end{pmatrix}$$

So that

$$C_3 \;=\; \begin{pmatrix} \sigma_A^2 & 0 & \frac{1}{2}\sigma_A^2 \\ 0 & \sigma_B^2 & \frac{1}{2}\sigma_B^2 \\ \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_B^2 & \frac{1}{4}\sigma_A^2 + \frac{1}{4}\sigma_B^2 \end{pmatrix}$$

With regard to statistical equivalence of the model, note that it is misleading to expect equivalent covariance matrices for X and Y, it is the probability density function that should be consistent so that;

$$\begin{gathered} X^T C_2^{-1} X = \chi^2(X) \quad ; \quad p(X) = \exp[-\chi^2(X)] \\ Y^T C_3^{-1} Y = \chi^2(Y) \quad ; \quad p(Y) = \exp[-\chi^2(Y)] = p(X) \end{gathered} \tag{2}$$

It is straightforward to compute C_2^{-1}, and so, we find that $X^T C_2^{-1} X = A^2/\sigma_A^2 + B^2/\sigma_B^2$. However to compute C_3^{-1}, even for a numerical example, we find that C_3 is now singular and only a pseudo inverse may be computed. For example if $\sigma_A = \sigma_B = 10$ then

$$C_3 = \begin{pmatrix} 100 & 0 & 50 \\ 0 & 100 & 50 \\ 50 & 50 & 50 \end{pmatrix}$$

It can be shown that the pseudo inverse matrix C_3^{-1} is valid using the eigen vectors and values of C_3. Using Matlab, the eigen vectors V and values λ of C_3 are given by:

$$V = \begin{pmatrix} 0.4082 & -0.4082 & 0.8165 \\ -0.7071 & 0.7071 & 0 \\ -0.5774 & -0.5774 & -0.5774 \end{pmatrix}$$

where each row is a vector and $\lambda = [0, 100, 150]$. the pseudo inverse of C_3 is given by

$$C_3^{-1} = 0.00001 \begin{pmatrix} 72 & -28 & 22 \\ -28 & 72 & 22 \\ 22 & 22 & 22 \end{pmatrix}$$

If we then compute $Y^T C_3^{-1} Y$, we find that it is equivalent to $X^T C_2^{-1} X$, i.e. $A^2/100 + B^2/100$. This means that equivalent χ^2 values (and therefore probability densities) are obtained for the X and Y sets of points. Although the eigen values inform us that the covariance matrix is singular (there is an unnecessary dimension in the measurement vector), simple inspection of the eigen vectors is not sufficient to determine that we can eliminate the third component of Y.

Although this numerical example works well, equivalence of the χ^2 depends upon exact cancellation of terms during calculation of the eigen vectors. Certainly, for an appropriate selection of eigen vectors (those spanning the theoretical correlating dimensions of the generating model), the mathematical definitions are sufficient to ensure equivalent density models in general for any linear process defining a pseudo-landmark. As

$$X^T C_2^{-1} X = X^T \frac{df(x)^T}{dx} \left(\frac{df(x)^T}{dx} C_2 \frac{df(x)}{dx}\right)^{-1} \frac{df(x)}{dx} X = Y^T C_3^{-1} Y$$

However, in practical situations the introduction of many correlated measurements and noise, would be expected to make reliable estimation of C and it's pseudo-inverse difficult. An uninformative measurement strictly has an infinite variance, even though empirical observation of the constrained value may indicate a high accuracy. It is therefore important to note that if we know a measurement conveys no additional information we can simply delete it from the representation vector and covariance matrix. As well as removing this potential contradiction, we can expect additional benefits due to; more clearly identified uninformative parameters, reduction in inverted matrix size, and the reduction of numerical problems associated with singular matrices. This may all be viewed as common sense, but it useful to know that there are reasons to expect that it should work in practice.

1.2 The 2D and 3D Case

We can now use the above argument to extend the idea from 1D to 2D. When working with curves in 2D, the 2x2 covariance matrix corresponding to each pseudo landmark will need to be represented so that the lack of information in particular directions is made explicit. If we can represent a 2D or 3D pseudo-landmark so that the unmeasured components align with the co-ordinates, then the null measurement dimensions can simply be eliminated from the representation vector without changing the underlying density model. Fortunately this is easier to arrange than we might expect.

Although linear shape models are generally constructed for 2D and 3D measured points in a fixed co-ordinate system, there is nothing to stop us estimating the linear correlations between measurements in any fixed set of individual landmark co-ordinates, provided that the relationships between these different co-ordinates are fixed across the data set. Thus for a pseudo-landmark we are free to rotate the measurement process so that the non-measured directions align with the co-ordinate axes for that landmark. The corresponding dimensions can then simply be eliminated from the measurement vector in the way illustrated above, without affecting statistical analysis.

2 Conclusions

A procedure exists for the definition of landmarks and pseudo-landmarks, which permits the construction of consistent shape density models and associated measurement covariances. If there are dependent (pseudo) landmarks among the independent landmarks in the shape data such that the dependency is a function of the other measured landmarks $d_n = f(x_i, y_i, x_j, y_j)$, we can choose to shorten each shape vector. Specifically, the shape vector $(x_1, y_1, x_2, y_2, x_i, y_i, x_j, y_j, ..., x_n, y_n)$ may be reduced to $(x_1, y_1, x_2, y_2, x_i, y_i, x_j, y_j, ..., d_n)$ where n, is a pseudo landmark. Here, d_n is the corresponding measurement in the direction perpendicular to the constraint direction provided by independent landmarks. By eliminating all unmeasured dimensions in this way all remaining covariance terms must be either 2x2 or 1x1 diagonal in C. This once more allows us to estimate C from samples while enforcing the known independence properties by zero-ing non-correlating terms. The only remaining correlations seen in data will be those due to biological variation, and much more easily accommodated during the building of a weighted linear model.

Once the mathematical relationships and measurement approach have been specified it may appear to some that all of this should have been obvious. Of course we can seek to build a linear model for any specification of measurement, not just those in a fixed co-ordinate frame. However, what some biologists may find surprising is the implication that there is nothing intrinsically problematic with defining pseudo-landmarks for a shape analysis. The fact that these points continue to cause problems in analysis is entirely due to the attempt to make a 2 or 3D measurement when only a 1D measurement is possible.