

ON THE ESTIMATION OF COVARIANCE MATRICES FOR PHOTOGRAMMETRIC IMAGE COORDINATES

by Wolfgang Förstner and Ralf Schroth

Photogrammetric Institute
Stuttgart University

Content:

1. Introduction
2. Mathematical Model for Multi-Reseau Images
3. Applied Methods of Multivariate Analysis
 - 3.1 Estimation of covariance matrices in a multivariate Gauss-Markov model
 - 3.2 Approximation of covariance matrices using variance components
 - 3.3 Choice of basis matrices V_k
 - 3.4 Separation of approximated covariance matrices
 - 3.5 Principal component analysis
 - 3.6 Evaluation of approximations
4. Comparison with Other Methods
 - 4.1 Methods for the approximation of covariance matrices
 - 4.2 On Ebner's theory for the analysis of covariance matrices
5. Examples
 - 5.1 Approximation of a 5×5 submatrix of \hat{C}
 - 5.2 Approximation of \hat{C}_{xx} and \hat{C}_{yy}
 - 5.3 Separation of \hat{C}_{xx} and \hat{C}_{yy}
 - 5.4 Composition of \hat{C}
 - 5.5 Principal component analysis of \hat{C}
6. Discussion

Summary:

Photogrammetric net densification is based on a mathematical model, which takes into account the systematic errors of the image coordinates. The paper discusses possibilities for estimating covariance matrices or parameters of covariance matrices for photogrammetric image coordinates. The positive definiteness of the estimated matrices and their decomposition into products of smaller matrices are treated.

Zusammenfassung:

Der Netzverdichtung mit Hilfe photogrammetrischer Methoden liegt ein mathematisches Modell zugrunde, das die systematischen Fehler der Bildkoordinaten berücksichtigt. Der Artikel diskutiert die Möglichkeiten zur Schätzung von Kovarianzmatrizen oder von Parametern der Kovarianzmatrizen und ihre Zerlegung in Produkte kleinerer Matrizen zur Vereinfachung ihrer Struktur.

lectured by W. Förstner

1. Introduction

1.1 During the last 10 years the accuracy of photogrammetric point determination has reached a standard, which makes it attractive for high precision application as network densification or deformation measurements. One of the reasons for this development is the successful refinement of the mathematical model for the adjustment process, especially the refinement of the functional model by using self calibration technique (additional parameters) including checks to guarantee the reliability of the result. The accuracy then reaches standard deviations of 2-3 μ m in the image and high reliability with $\bar{s} < 5$, i.e. it can be guaranteed that the coordinates, photogrammetrically determined, are correct within 10 - 15 μ m tolerance (in the image), if proper tests and checks concerning gross and systematic errors, controllability and reliability are applied.

1.2 However, self calibration technique only is able to filter out the constant part of the systematic errors. Further refinement of the functional model using locally different sets of additional parameters to compensate varying systematic effects leads to miscellaneous difficulties in computation and evaluation. On the other side it seems to be impossible to quantify the different sources for the deviation of the idealized from the real beam passing atmosphere, optic, film development and measuring process to its representation - the image coordinates. Most of the systematic effects are varying with time, i.e. from image to image, leading to correlations within and between the images, even if the constant part of the systematic is conceived within the functional model.

1.3 The stochastic properties of photogrammetrically observed coordinates have already been analysed by Stark (1973) in a pilot study and recently by Schilcher (1980). They obtained moderate and high correlations within photogrammetric models and images. Schilcher demonstrated the possibility of compensating a rough functional model by using a refined stochastic model and vice versa (Ackermann, Schilcher, 1980). In both cases the estimation of a full covariance matrix was based on the residuals of a system with high redundancy. The residuals therefore were treated as true errors. This leads to difficulties with the positive definiteness of the estimated covariance matrix, which in both cases could be circumvented by either imposing restrictions on the structure of the covariance matrix or by modifying the estimation process.

1.4 The benefit of using a refined stochastic model was demonstrated by Ackermann (1976), who showed that in cases with high redundancy the precision of the points after an absolute orientation is strongly influenced (improved) by the (more realistic) stochastic model. It showed that point estimation is not so much influenced by errors in the stochastic model as interval estimation is. Especially all accuracy measures are directly dependent on the assumed stochastic properties of the observation.

Even if full covariance matrices will not be used in large block adjustments within the next years they will find an increasing market in small block application especially in industrial photogrammetry and in connection with on-line error detection, as the sensitivity of the statistical tests is improved.

1.5 The rigorous estimation of a covariance matrix for image coordinates has to cope with two difficulties:

- the data acquisition and
- the enormous computational effort.

In order to meet the conditions of a multivariate adjustment there should be a possibility of obtaining images with nearly the same exterior orientation. This only is possible in a laboratory test, which gives nonrealistic results. The application of variance component estimation requires the handling of full matrices with 10^6 and more elements if only 20 images with 25 points each are analysed.

In order to come to a theoretically satisfying solution, though, we restrict our investigation to a part of the total error budget. Reseau images give sufficient and ideally positioned information about a great part of the systematic errors such as flatness of the pressure plate, film deformation, measuring process etc. So we will not get any information about point definition, refraction or image motion. The fix frame of the reseau grid yields repeated realization of a regular point field being ideal for all kinds of multivariate analysis.

1.6 This paper is supposed to collect, discuss, modify and develop mathematical tools for the estimation, approximation and evaluation of covariance matrices for reseau images. The main problem is the positive definiteness of the estimated covariance matrices. The solution is based on a theorem by Pukelsheim and Styan (1979) which gives a sufficient condition for the estimated variance covariance matrix in a variance-component model for being positive, requiring the set of basis matrices to form a special Jordan algebra. The construction of such algebras is discussed in detail and tries to take into account the real structure of the data.

2. Mathematical Model for Multi-Reseau Images

The successive exposure of a reseau during a flight mission can be interpreted as a realization of a regular net at different epoches. This makes the application of the multivariate analysis feasible. It also allows a simultaneous treatment of all images and the estimation of full covariance matrices, if the adjacent mathematical model is chosen appropriately leading to a multivariate Gauss-Markov model (GMM).

We assume that all reseau images are generated under the same conditions, i.e. tail effects in the time series are avoided by e.g. omitting the first images of the film. As the calibration of the reseau only has influence on the constant part of the systematic, we assume that the calibration data are not known and should be estimated. This leads to the following functional model for the p images (cf. Fig. 1).

$$\underline{l}_k + \underline{v}_k = \underline{A} \hat{\underline{x}}_k + \hat{\underline{z}} \quad \text{for } k = 1, \dots, p \quad (2.1)$$

with the observed reseau coordinates \underline{l}_k , $o(\underline{l}_k) = n \times 1$, the corresponding residuals \underline{v}_k , the design matrix \underline{A} of order $o(\underline{A}) = n \times 3$ to define the datum of each image by a 3-parameter transformation and the mean vector $\underline{\hat{z}}$ of order $o(\underline{\hat{z}}) = n \times 1$ of the reseau coordinates.

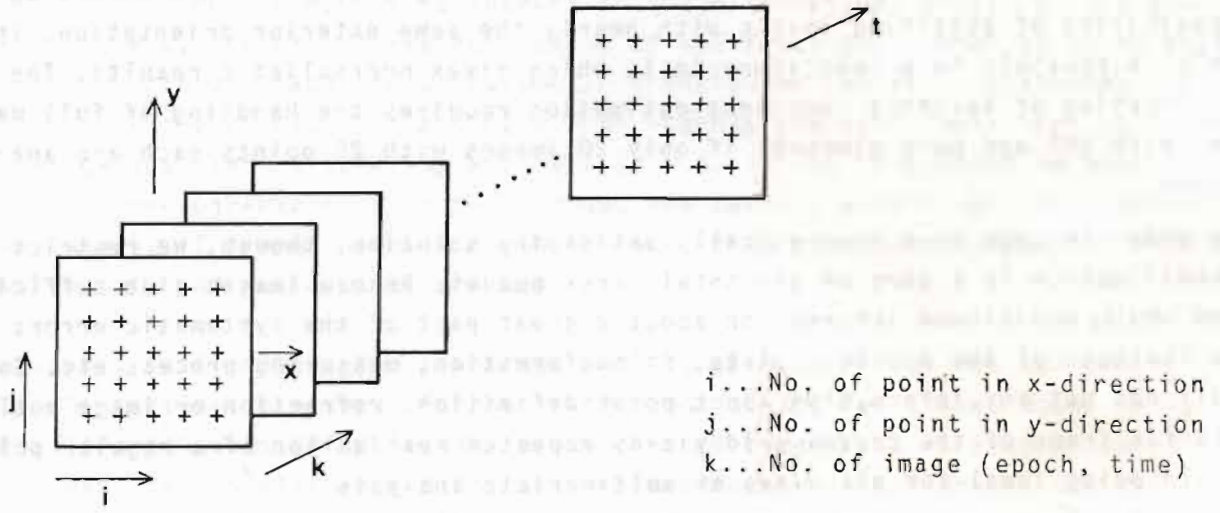


Fig. 1 : Model of multi-reseau images

If the vectors \underline{l}_k , \underline{v}_k , $\underline{\hat{x}}_k$ are collected in the matrices

$$\begin{aligned}\underline{L} &= (\underline{l}_1, \dots, \underline{l}_k, \dots, \underline{l}_p) \\ \underline{V} &= (\underline{v}_1, \dots, \underline{v}_k, \dots, \underline{v}_p) \\ \underline{\hat{X}} &= (\underline{\hat{x}}_1, \dots, \underline{\hat{x}}_k, \dots, \underline{\hat{x}}_p),\end{aligned}$$

we can rewrite eq. (2.1)

$$\underline{L} + \underline{V} = \underline{A} \underline{\hat{X}} + \underline{\hat{Z}} \underline{B}^T \quad (2.2)$$

with the vector $\underline{\hat{z}}$ (now generalized to a matrix) and the vector \underline{B} , $o(\underline{B}) = p \times 1$, consisting of 1's ($\underline{B} \equiv \underline{1}$, also generalized to a matrix). An equivalence to eq. (2.2) is

$$\text{vec} \underline{L} + \text{vec} \underline{V} = (\underline{I} \otimes \underline{A}) \text{vec} \underline{\hat{X}} + (\underline{B} \otimes \underline{I}) \text{vec} \underline{\hat{Z}} \quad (2.3)$$

with the vec operator mapping the columns of a matrix into a vector and the Kronecker product " \otimes ".

The corresponding stochastic model is

$$D(\underline{l}_k) = \underline{C} = \sigma_o^2 \underline{Q} \quad \text{for } k = 1, \dots, p \quad (2.4)$$

and

$$C(\underline{l}_k, \underline{l}_{k'}) = \sigma_{kk'} \underline{Q} = r_{kk'} \underline{C} \quad , \quad (2.5)$$

in which D and C represent the dispersion and the covariance operator, resp., σ_o^2 is the variance of unit weight, \underline{C} the covariance matrix, \underline{Q} the matrix of the weight coefficients. The covariance between different epoches are assumed to have the same structure as the dispersion matrix at one epoch; the $r_{kk'}$ are correspon-

ding correlations between the epochs. Eq. (2.5) makes the covariance of two observations l_{ik} and $l_{i'k'}$ in different epoches only dependent on the covariance between observation No. i and i' and the correlation between the epoch k and k' :

$$C(l_{ik}, l_{i'k'}) = r_{kk'} \cdot C_{ii'} = \sigma_{kk'} \cdot q_{ii'}$$

We can rewrite Eqs. (2.4) and (2.5) in

$$D(\text{vec } \underline{L}) = \underline{R} \otimes \underline{C} = \underline{\Sigma} \otimes \underline{Q} \quad (2.6)$$

with the correlation matrix $\underline{R} = (r_{kk'})$ and the covariance matrix $\underline{\Sigma} = (\sigma_{kk'})$.

Remark:

This multivariate model is an extended version of the usual multivariate GMM for homogeneous and complete observations. The term $\hat{\underline{Z}} \underline{B}^T$ contains common systematic errors (cf. Buck, 1977) as used in photogrammetric block adjustment. The introduction of a full covariance matrix \underline{C} for the image coordinates instead of a unit matrix is essential for the following analysis, but no real generalization of the normal GMM.

The regular point pattern of a reseau suggests a further decomposition of the dispersion matrix first leading to

$$\underline{C} = \begin{bmatrix} \underline{C}_{xx} & \underline{C}_{xy} \\ \underline{C}_{yx} & \underline{C}_{yy} \end{bmatrix} \quad (2.7)$$

If the covariance of two coordinates x_{ij} and $x_{i'j'}$ can be written as the product of separate contributions of the corresponding indexes (i.e. coordinates) i , i' , j and j' :

$$C(x_{ij}, x_{i'j'}) = \sigma_{ii'}^x \cdot \sigma_{jj'}^y \quad (2.8)$$

The dispersion matrix \underline{C}_{xx} becomes the form:

$$\underline{C}_{xx} = \underline{C}_{xx}^x \otimes \underline{C}_{xx}^y$$

Similar decomposition of \underline{C}_{xy} and \underline{C}_{yy} finally leads to

$$\begin{aligned} \underline{C} &= \begin{bmatrix} \underline{C}_{xx}^x \otimes \underline{C}_{xx}^y & \underline{C}_{xy}^x \otimes \underline{C}_{xy}^y \\ \underline{C}_{yx}^x \otimes \underline{C}_{yx}^y & \underline{C}_{yy}^x \otimes \underline{C}_{yy}^y \end{bmatrix} = \begin{bmatrix} \underline{C}_{xx}^x & \underline{C}_{xy}^x \\ \underline{C}_{yx}^x & \underline{C}_{yy}^x \end{bmatrix} * \begin{bmatrix} \underline{C}_{xx}^y & \underline{C}_{xy}^y \\ \underline{C}_{yx}^y & \underline{C}_{yy}^y \end{bmatrix} \\ &= \underline{C}^x * \underline{C}^y, \end{aligned} \quad (2.9)$$

where "*" is the generalized Hadamard product.

If this special structure could be found in real data - and we have found it - , this would lead to a very much simplified generation of a full covariance matrix and also to enormous savings in computing time for the estimation process. Keeping this in mind we want to discuss the existing procedures for estimating positive definite (p.d.) covariance matrices.

3. Applied Methods of Multivariate Analysis

This chapter collects the methods applied for estimation, approximation and evaluation of covariance matrices. It follows the steps of the calculation procedure which leads to the results presented in ch. 5. The analysis was based on a covariance matrix estimated in the multivariate model given above, in order to separate the estimation and the approximation process (cf. 3.1). This is justified by the fact, that using the estimated covariance matrix of the first iteration step for an improved estimation of the parameters leads only to neglectable changes of the covariance matrix in the second iteration step. The approximation itself uses the method of covariance component estimation which can be very much simplified if an estimated covariance matrix is given (ch. 3.2). The theorem of Pukelsheim and Styan (1979) gives sufficient conditions for the estimated covariance matrix to be positive definite. This condition requires the basis matrices to span a special Jordan algebra. Ch. 3.3 discusses the choice of different sets of basis matrices, which can serve for an approximation of \underline{C} with the structure of eq.(2.9). For this pilot study however, we approximated \underline{C}_{xx} and \underline{C}_{yy} separately and extrapolated \underline{C}_{xy} . The necessary separation of \underline{C}_{xx} and \underline{C}_{yy} is described in ch. 3.5.. The results are compared with a principal component analysis (ch. 3.5) and evaluated by different tests (ch. 3.6).

3.1 Estimation of the covariance matrix in a multivariate Gauss-Markov model

3.1.1 Estimation using true errors

The estimation of covariance matrix in a multivariate GMM is based on the residuals which are correlated and in general give not full information about the stochastic behavior of the observations. Therefore we first discuss the simple case, where there are no parameters to be estimated:

$$E(\underline{L}) = \underline{0} \quad , \quad D(\text{vec} \underline{L}) = \underline{R} \otimes \underline{C} = \underline{\Sigma} \otimes \underline{Q} \quad . \quad (311.1)$$

If \underline{R} is given, we obtain

$$\hat{\underline{C}} = \frac{\underline{L} \underline{R}^{-1} \underline{L}^T}{p} \quad \text{with } p \geq n \quad (311.2)$$

and if \underline{Q} is given

$$\hat{\underline{\Sigma}} = \frac{\underline{L}^T \underline{Q}^{-1} \underline{L}}{n} \quad \text{with } n \geq p \quad . \quad (311.3)$$

The estimated matrices are p.d. with probability 1, if their order is smaller than the rank of the corresponding quadratic forms. Thus with a given set of observations \underline{L} only one of both estimators can be used (except in the case $p = n$, where there is no redundancy!). A simultaneous estimation of \underline{R} and \underline{C} or $\underline{\Sigma}$ and \underline{Q} in general is not possible. If restrictions are imposed on one or both matrices during the estimation process (e.g. using variance components), one is only able to get information of \underline{R} and \underline{C} , which is statistically meaningful. This does of course not mean that it is realistic in all cases.

3.1.2 Estimation using residuals

The estimation of the covariance matrices \underline{C} or $\underline{\Sigma}$ in the generalized model eq. (2.2) uses the residuals \underline{V} . With the projection matrices

$$\underline{D}_A = \underline{I}_n - \underline{A} (\underline{A}^T \underline{C}^{-1} \underline{A})^{-1} \underline{A}^T \underline{C}^{-1} \quad \text{with } o(\underline{D}_A) = n \times n, r(\underline{D}_A) = n - r(\underline{A}) \quad (312.1)$$

and

$$\underline{D}_B = \underline{I}_p - \underline{B} (\underline{B}^T \underline{R}^{-1} \underline{B})^{-1} \underline{B}^T \underline{R}^{-1} \quad \text{with } o(\underline{D}_B) = p \times p, r(\underline{D}_B) = p - r(\underline{B}) \quad (312.2)$$

they can be written as

$$\underline{V} = -\underline{D}_A \underline{L} \underline{D}_B^T \quad (312.3)$$

(Proof cf. appendix A), with their covariance matrix

$$D(\text{vec} \underline{V}) = \underline{D}_B \underline{R} \underline{D}_B^T \otimes \underline{D}_A \underline{C} \underline{D}_A^T = \underline{R}_{VV} \otimes \underline{C}_{VV} \quad (312.4)$$

Using the first moments of the quadratic form $E(\underline{v}_i^T \underline{R}^{-1} \underline{v}_i)$ of two residuals \underline{v}_i and \underline{v}_j in all epochs, we can estimate the elements of covariance matrix \underline{C}_{VV} of the residuals (\underline{R} given)

$$\begin{aligned} E(\underline{v}_i^T \underline{R}^{-1} \underline{v}_i) &= \text{tr}(\underline{R}^{-1} E(\underline{v}_i \underline{v}_i^T)) = \text{tr}(\underline{R}^{-1} \underline{D}_B \underline{R} \underline{D}_B^T (\underline{C}_{VV})_{ii}) \\ &= (\underline{C}_{VV})_{ii}, \quad \text{tr}(\underline{R}^{-1} \underline{D}_B \underline{R} \underline{D}_B^T) \\ &= (\underline{C}_{VV})_{ii}, \quad (p - 1) \end{aligned} \quad (312.5)$$

leading to

$$\hat{\underline{C}}_{VV} = \frac{\underline{V} \underline{R}^{-1} \underline{V}^T}{p - 1} \quad \text{with } r(\hat{\underline{C}}_{VV}) = n - u \quad (312.6)$$

and similarly to (\underline{C} given)

$$\hat{\underline{R}}_{VV} = \frac{\underline{V}^T \underline{C}^{-1} \underline{V}}{n - u} \quad \text{with } r(\hat{\underline{R}}_{VV}) = p - 1 \quad (312.7)$$

As in the simple case $E(\underline{L}) = 0$, a simultaneous estimation of \underline{C}_{VV} and \underline{R}_{VV} is not possible. Moreover both matrices are singular. The rank deficiency of 1 and u corresponds to the single mean vector \underline{Z} and the $u=3$ unknown parameters for each epoch. Thus no estimations for \underline{R} and \underline{C} are available under this model.

Remark:

Eq. (312.7) converts to the standard equation $\hat{\underline{\Sigma}} = \underline{V}^T \underline{V} / (n - u)$ if $\underline{Z} \equiv \underline{0}$ and $\underline{C} = \underline{I}$. $\hat{\underline{\Sigma}}$ is p.d. with probability one in this case.

As a simultaneous estimation of \underline{R} and \underline{C} is not possible we restrict our analysis to the estimation of \underline{C} . The correlation matrix \underline{R} is assumed to be the identity matrix, i.e. no correlation between the images are assumed.

3.1.3 Distribution of \hat{C}_{VV}

As the estimated covariance matrices are singular they are not Wishart distributed. The similarity of the estimation process to the standard case suggests the introduction of a singular Wishart distribution which has no density function but only a characteristic function.

Definition 1 : (singular Wishart distribution)

The matrix $\hat{\Sigma}$ is said to follow a singular Wishart distribution $W(n, \Sigma)$, $o(\Sigma) = p$, $r(\Sigma) = r < p$ if there exists a matrix P that the transformed matrix $P \hat{\Sigma} P^T$ follows a regular Wishart distribution $W(n, P \Sigma P^T)$ with $r(P \Sigma P^T) = o(P \Sigma P^T) = r$.

Remark :

We will not change the notation to distinguish between the singular and regular distribution, as e.g. Searle (1971) does for the normal distribution. The results on the Wishart distribution with regular Σ are also valid for singular Σ if the rank of Σ is taken into account properly.

We now have the following

Theorem 1 : (Distribution of \hat{C}_{VV})

Given the model $L + V = A \hat{X} + \hat{Z} B^T$, $D(\text{vec } L) = I \otimes C$ ($o(L) = n \times p$, $o(A) = n \times u$, $r(A) = u$, $o(B) = p \times d$, $r(B) = d$, $o(\hat{X}) = u \times p$, $o(\hat{Z}) = n \times d$, $o(C) = n \times n$, $p > n$). With the least squares residuals (eq. 312.3) the matrix

$$\hat{C}_{VV} = \frac{1}{p-d} V V^T \quad (313.1)$$

follows a Wishart distribution with $n-u$ degrees of freedom

$$(n-u) \hat{C}_{VV} \sim W(n-u, C_{VV}).$$

Proof : The projection matrix D_A leads to the residuals V with respect to a weighted centre of gravity. The alternative projection matrix $D_{\bar{A}}$

$$D_{\bar{A}} = I - A (A^T E_r A)^{-1} A^T E_r \quad (313.2)$$

with

$$E_r = \text{blockdiag}(0, P_2), \quad P_2 = P_2^T, \quad r(P_2) = o(P_2) = u \quad (313.3)$$

leads to the residuals

$$\bar{V} = -D_{\bar{A}} D_A L D_B^T = -D_{\bar{A}} L D_B^T, \quad (313.4)$$

the last u of them being zero. The corresponding covariance matrix $\hat{C}_{\bar{V}\bar{V}}$ has the same rank deficiency as \hat{C}_{VV} , namely u . As the last u columns and rows of $C_{\bar{V}\bar{V}}$ are also zero, this can be interpreted as an adjustment with the last u observations used as datum parameters. If they are deleted from the adjustment, the reduced covariance matrix $C_{\bar{V}\bar{V}}(\text{red})$ has full rank and is Wishart distributed. It can be obtained by a transformation $P C_{\bar{V}\bar{V}} P^T$, where P contains the first $n-u$ rows of $D_{\bar{A}}$. q.e.d.

3.1.4 A positive definite covariance matrix for the observations

Given $\underline{R} = \underline{I}$, we are able to estimate $\hat{\underline{C}}_{vv}$ from eq. (312.6). But we want to have a positive definite covariance matrix for the observations. Though it is easy to modify a positive semidefinite (p.s.d.) matrix by adding $\mu^2 \underline{I}$, we search for a matrix $\hat{\underline{C}}$ which is as close to $\hat{\underline{C}}_{vv}$ as possible. In general the transformation from \underline{l}_k to \underline{v}_k

$$\underline{v}_k = -\underline{D}_A \underline{l}_k = -\underline{C}_{vv} \underline{C}^{-1} \underline{l}_k \quad (314.1)$$

is a similarity transformation according to Baarda (cf. v. Mierlo, 1978). The matrix \underline{D}_A in our case transforms the coordinates \underline{l}_k , which are measured in an arbitrary (a)-system, into the (m)-system of the centre of gravity, taking into account the unknown weight matrix \underline{C}^{-1} , as \underline{D}_A depends on \underline{C} . If we denote the unknown covariance matrix of the observations with $\hat{\underline{C}}$, one may require

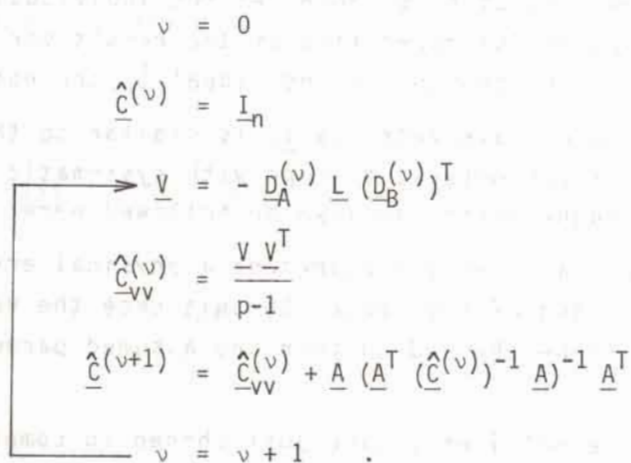
$$\hat{\underline{C}}_{vv} = \underline{D}_A \hat{\underline{C}} \underline{D}_A^T \quad (314.2)$$

But this equation is fulfilled if

$$\hat{\underline{C}} = \hat{\underline{C}}_{vv} + \underline{A} (\underline{A}^T \hat{\underline{C}}^{-1} \underline{A})^{-1} \underline{A}^T \quad (314.3)$$

as $\underline{D}_A = \hat{\underline{C}}_{vv} \hat{\underline{C}}^{-1}$. $\hat{\underline{C}}$ from eq. (314.3) is p.d. by construction. Of course this is one choice for $\hat{\underline{C}}$ as the solution of (314.2) is not unique. Eq. (314.3) describes an iterative procedure determining $\hat{\underline{C}}$. Also eq. (314.2) is only valid at the point of convergency. A refined notation would designate \underline{D}_A being stochastic.

The iteration procedure for the estimation of the matrix \underline{C} is summarized by the following flow chart:



We now simplify the structure of $\hat{\underline{C}}_{vv}$ and $\hat{\underline{C}}$ using variance components.

3.2 Approximation of covariance matrices using variance components

The estimation of the full covariance matrix imposes no restriction onto $\hat{\underline{C}}_{vv}$ or $\hat{\underline{C}}$, except being p.s.d or p.d. respectively. In order to simplify the application of a full covariance matrix we try to extract the structure of the covariance matrix. This can be done in several ways which are discussed in ch. 4.

Here we restrict to the powerful tool of variance components. The functional model for the approximation of $\hat{\underline{C}}$ is

$$\hat{\underline{C}} = \sum_{\kappa=1}^k t_{\kappa} \underline{V}_{\kappa} \quad (32.1)$$

The matrices \underline{V}_{κ} are given, they impose restrictions onto the structure of $\hat{\underline{C}}$. The factors t_{κ} are unknown. If we rewrite eq. (32.1)

$$\text{vec} \hat{\underline{C}} = \underline{D} \underline{t} \quad (32.2)$$

with $\underline{D} = (\text{vec } \underline{V}_1, \dots, \text{vec } \underline{V}_k)$, $\underline{t} = (t_1, \dots, t_k)^T$ the similarity of the variance component model with the GMM $\underline{l} = \underline{A} \underline{x}$ is revealed.

The corresponding stochastic model uses the dispersion matrix of $\text{vec } \hat{\underline{C}}$ which - in case $\hat{\underline{C}}$ is Wishart distributed - depends on the matrix of 4th moments

$$\underline{F} = \hat{\underline{C}} \otimes \hat{\underline{C}} \quad (32.3)$$

(cf. Pukelsheim, Styan, 1979).

The leastsquares estimator $\hat{\underline{t}}$ for \underline{t} thus given by

$$\hat{\underline{t}} = (\underline{D}^T \underline{F}^{-1} \underline{D})^{-1} \underline{D}^T \underline{F}^{-1} \text{vec} \hat{\underline{C}} \quad (32.4)$$

results in an approximated matrix

$$\underline{\bar{C}} = \sum_{\kappa=1}^k \hat{t}_{\kappa} \underline{V}_{\kappa} \quad (32.5)$$

This matrix is not necessarily positive definite. This reveals a great disadvantage of the method if used without care. Moreover the individual values \hat{t}_{κ} may be not positive thus making an interpretation of the result very doubtful. However the matrix $\underline{\bar{C}}$ might be positive, though the individual \hat{t}_{κ} are not.

The problem of choosing appropriate matrices \underline{V}_{κ} is similar to the problem of choosing appropriate additional parameters, to cope with systematic errors in a GMM. There are two lines of thought which can also be followed here:

- The parameters (i.e. matrices \underline{V}_{κ}) represent a physical error source, e.g. a cylindrical distortion of the image. In this case the result of an estimation process has to be checked whether the assumed parameter is realistic or not.
- The parameters or the matrices \underline{V}_{κ} are just chosen to compensate unknown effects by extending the mathematical model. In this case one must check whether the mathematical model is really able to determine the parameters in order to guarantee a reliable result. The parameters can be chosen arbitrarily, e.g. in order to simplify the evaluation of the parameters.

The already mentioned impossibility to model all sources properly and the following theorem support the second line of thought, where the matrices \underline{V}_{κ} are selected without having certain physical effects in mind.

Theorem 2 : (non-negativity-condition, Pukelsheim, Styan, 1979)

Consider independent and identically normally distributed random R^n -vectors \underline{z}_α with common mean $\underline{0}$ and common dispersion matrix $\sum t_k \underline{V}_k$, where $p \geq n$. Assume that the k decomposing matrices \underline{V}_k span a k -dimensional special Jordan algebra B . Define $G \subset R^k$ to be the region of those values t of the dispersion parameter such that $\sum t_k \underline{V}_k$ is positive definite, and assume $G \neq \emptyset$. Then:

- a) The maximum likelihood estimator for $\underline{t} \in G$ is almost surely equal to the uniform minimum variance unbiased estimator

$$\hat{\underline{t}} := (\underline{D}^T \underline{D})^{-1} \underline{D}^T \text{vec } \underline{C}, \tag{32.6}$$

where $\underline{D} := [\text{vec } \underline{V}_1, \dots, \text{vec } \underline{V}_k]$, and $\underline{C} = \sum z_\alpha z_\alpha^T / p$.

- b) The estimated dispersion matrix $\underline{\bar{C}} = \sum \hat{t}_k \underline{V}_k$ is nonnegative definite: in fact, if the sample dispersion matrix \underline{C} is positive definite, so is $\underline{\bar{C}}$.

(The explanation of a special Jordan algebra is given below.)

Remark :

- The theorem gives a sufficient condition for $\underline{\bar{C}}$ to be positive definite. It is not known, if it is also necessary (cf. Pukelsheim, 1980).
- The special structure of the set $\{\underline{V}_k\}$ does not only guarantee the positive definiteness of $\underline{\bar{C}}$ but also simplifies the estimation of \underline{t} drastically, as the fourth moments have not to be calculated.
- The theorem does not give the assertion for a given singular \underline{C} , that the approximated matrix $\underline{\bar{C}}$ has the same rank. This restricts the application, if one follows the statistical reasoning.

The vital assumption is , that the matrices \underline{V}_k span a special Jordan algebra which we will discuss now.

3.3 Choice of basis matrices \underline{V}_k

Let $\text{Sym}(n)$ be the set of all symmetric $n \times n$ matrices.

Definition 2 : (special Jordan algebra)

A subspace B of $\text{Sym}(n)$ is a special Jordan algebra, iff it is closed under the multiplication $\underline{A} \circ \underline{B} := \frac{1}{2}(\underline{A} \underline{B} + \underline{B} \underline{A})$.

Example 1 :

$n = 2$, $\text{Sym}(n)$ contains all symmetric matrices of order 2×2

$$B_1 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\} \quad \text{consists of all matrices which have the form } \begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

Obviously B_1 is a special Jordan algebra, as for arbitrary a, b, c, d :

$$\underline{A} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \underline{B} = \begin{bmatrix} c & d \\ d & c \end{bmatrix} \text{ follows } \frac{1}{2}(\underline{A} \underline{B} + \underline{B} \underline{A}) \in B_1. \text{ The space can also be spanned by}$$

the matrices $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ and $\begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}$ or even by $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This shows that

there is enough freedom in the choice of the basis matrices which span the algebra and the basis matrices must not necessarily be positive semidefinite.

In order to be able to construct further algebras one needs criteria to check the closedness. This can be done by the original formulation of the conditions by Seely (1971) who used a different terminology, namely the notion of quadratic subspaces.

Definition 3 : (quadratic subspace, Seely, 1971)

A subspace B of $\text{Sym}(n)$ is called a quadratic subspace iff $\underline{A}^2 \in B$ whenever $\underline{A} \in B$.

It can easily be proved that the algebra B_1 forms a quadratic subspace.

Seely gives several conditions for B to be a quadratic subspace (or a special Jordan algebra).

Lemma 1 : (conditions for B to be a quadratic subspace, Seely, 1971)

Let B be a subspace of $\text{Sym}(n)$ and let B_0 be an arbitrary spanning set for B then the following conditions are equivalent:

- a) $\underline{A} \in B \implies \underline{A}^2 \in B$
- b) $\underline{A}, \underline{B} \in B_0 \implies (\underline{A} + \underline{B})^2 \in B$
- c) $\underline{A}, \underline{B} \in B_0 \implies \underline{A}\underline{B} + \underline{B}\underline{A} \in B$ and
- d) $\underline{A} \in B \implies \underline{A}^k \in B$ for each finite integer $k \geq 1$.

Remark :

The third condition shows the connection with the special Jordan algebra. It is at the same time the easiest way to check whether a set of matrices spans a special Jordan algebra.

The following lemma is needed for the construction of algebras containing Kronecker products of matrices, which we want to apply for the approximation of \underline{C}_{xx} etc.

Lemma 2 : (commutative subspace, lemma 6 in Seely, 1971)

A necessary and sufficient condition for a subspace B to be a commutative quadratic subspace is the existence of a basis $\underline{V}_1, \underline{V}_2, \dots, \underline{V}_k$ for B such that each \underline{V}_i is idempotent and such that $\underline{V}_i \underline{V}_j = 0$ for $i \neq j$. Moreover, apart from the indexing such a basis for a commutative quadratic subspace is unique.

Remark :

- Commutativity of B implies $\underline{A}\underline{B} = \underline{B}\underline{A} \in B$ whenever $\underline{A}, \underline{B} \in B$
- B_1 is commutative as $B_1 = \{\underline{I}_2 - \underline{J}_2, \underline{J}_2\}$ with $(\underline{I}_2 - \underline{J}_2) \cdot \underline{J}_2 = \underline{0}$. (\underline{J}_2 cf. example 4)

As the reader will not be familiar with the possibilities, these lemmas give for the construction of special Jordan algebras, we will give several examples. To simplify the notion we will only refer to B as a Jordan algebra.

Example 2 : ($k \leq n$, commutative)

Let $\underline{c}_i, i = 1, \dots, k \leq n$ be k mutually orthogonal and normalized vectors. Then

$$B_2 = \{\underline{c}_i \underline{c}_i^T \mid i = 1, \dots, k\}$$

span a k -dimensional commutative Jordan algebra. If $k < n$ all matrices $\underline{A} \in B_2$ are singular. In order to be able to approximate a regular covariance matrix B_2 must be complemented by at least one regular matrix, e.g. a identity matrix (cf. example 4 which is a special case of $B_2 \cup \underline{I}$). If $k = n$, the identity matrix is element of B as $\sum_{i=1}^n \underline{c}_i \underline{c}_i^T = \underline{I}$.

Example 3 : ($k = 5$, commutative)

In order to approximate a 5×5 covariance matrix, e.g. \underline{C}_{xx}^x (cf. eq. 2.9) we chose the special vectors \underline{c}_i :

$$\underline{c}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \underline{c}_2 = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \quad \underline{c}_3 = \begin{bmatrix} 2 \\ -1 \\ -2 \\ -1 \\ 2 \end{bmatrix}, \quad \underline{c}_4 = \begin{bmatrix} -1 \\ 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad \underline{c}_5 = \begin{bmatrix} -1 \\ 4 \\ -6 \\ 4 \\ -1 \end{bmatrix},$$

leading to B_3 . They correspond to the first 5 orthogonal polynomials on the set of points $(-2, -1, 0, 1, 2)$ representing shift, scale and several deflections.

Example 4 : ($k = 2$, commutative)

Let $\underline{J}_n = \frac{1}{n} \underline{1} \underline{1}^T = \frac{1}{n} \underline{J}$ with $\underline{1}^T = (1, \dots, 1)$. Then

$$B_4 = \{\underline{I}_n, \underline{J}_n\}$$

is an extension of B_1 and also commutative. This algebra may be used for repeated measurements, though the correlated mean in this case is independent on the correlation coefficient, thus leading to a singular normal equation system (eq. 32.4), if based solely on repeated observations.

Example 5 : ($k = 5, n = n_1 + n_2$)

Let be all matrices of order $n \times n$ be partitioned according to $n_1 + n_2 = n$. With the matrices

$$\begin{aligned} \underline{V}_1 &= \begin{bmatrix} \underline{I}_{n_1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}, & \underline{V}_2 &= \begin{bmatrix} \underline{0} & \underline{0} \\ \underline{0} & \underline{I}_{n_2} \end{bmatrix}, & \underline{V}_3 &= \begin{bmatrix} \underline{J}_{n_1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}, \\ \underline{V}_4 &= \begin{bmatrix} \underline{0} & \underline{0} \\ \underline{0} & \underline{J}_{n_2} \end{bmatrix}, & \underline{V}_5 &= \underline{J}_n. \end{aligned}$$

$B_5 = \{\underline{V}_i \mid i = 1, \dots, 5\}$ spans a non commutative Jordan algebra. It can be used for covariance matrices of the following form, being extensions of B_4 towards two correlated groups of observations:

$$\begin{bmatrix} a & b & \dots & b & c & \dots & c \\ b & a & \dots & b & c & \dots & c \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ b & b & \dots & a & c & \dots & c \\ \hline c & c & \dots & c & d & e & \dots & e \\ c & c & \dots & c & e & d & \dots & e \\ . & . & . & . & . & . & . & . \\ c & c & \dots & c & e & e & \dots & d \end{bmatrix}$$

Example 6 : ($k = 9$, $n = n_1 + n_2 + 1$)

A further partitioning into 9 submatrices, one of them being a single element, leads to the following basis ($n_1 \geq 2$, $n_2 \geq 2$) of a non commutative Jordan algebra:

$$\underline{v}_1 = \begin{bmatrix} \underline{I}_{n_1} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \end{bmatrix}, \quad \underline{v}_2 = \begin{bmatrix} \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{I}_{n_2} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \end{bmatrix}, \quad \underline{v}_3 = \begin{bmatrix} \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & 1 \end{bmatrix},$$

$$\underline{v}_4 = \begin{bmatrix} \underline{J}_{n_1} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \end{bmatrix}, \quad \underline{v}_5 = \begin{bmatrix} \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{J}_{n_2} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \end{bmatrix}, \quad \underline{v}_6 = \begin{bmatrix} \underline{0} & \underline{0} \\ \underline{0} & \underline{J}_{n_1+n_2} \end{bmatrix},$$

$$\underline{v}_7 = \begin{bmatrix} \underline{J}_{n_1+n_2} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}, \quad \underline{v}_8 = \frac{1}{n_1+1} \begin{bmatrix} \underline{J}_{n_1} & \underline{0} & \underline{1} \\ \underline{0} & \underline{0} & \underline{0} \\ \underline{1}^T & \underline{0} & 1 \end{bmatrix}, \quad \underline{v}_9 = \underline{J}_{n_1+n_2+1}.$$

The smallest values for n_1 and n_2 together with a renumbering gives the following Jordan algebra B_6 which is used for the analysis of a 5×5 matrix. The basis matrices are:

$$\underline{v}_1 = \begin{bmatrix} 1 & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 1 \end{bmatrix}, \quad \underline{v}_2 = \begin{bmatrix} . & 1 & . & . & . \\ . & . & . & . & . \\ . & . & . & 1 & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix}, \quad \underline{v}_3 = \begin{bmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & 1 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix},$$

$$\underline{v}_4 = \begin{bmatrix} 1 & . & . & . & 1 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & . & . & . & 1 \end{bmatrix}, \quad \underline{v}_5 = \begin{bmatrix} . & . & . & . & . \\ . & 1 & . & 1 & . \\ . & . & . & . & . \\ . & 1 & . & 1 & . \\ . & . & . & . & . \end{bmatrix}, \quad \underline{v}_6 = \begin{bmatrix} 1 & 1 & . & 1 & 1 \\ 1 & 1 & . & 1 & 1 \\ . & . & . & . & . \\ 1 & 1 & . & 1 & 1 \\ 1 & 1 & . & 1 & 1 \end{bmatrix},$$

$$\underline{v}_7 = \begin{bmatrix} 1 & . & 1 & . & 1 \\ . & . & . & . & . \\ 1 & . & 1 & . & 1 \\ . & . & . & . & . \\ 1 & . & 1 & . & 1 \end{bmatrix}, \quad \underline{v}_8 = \begin{bmatrix} . & . & . & . & . \\ . & 1 & 1 & 1 & . \\ . & 1 & 1 & 1 & . \\ . & 1 & 1 & 1 & . \\ . & . & . & . & . \end{bmatrix}, \quad \underline{v}_9 = \begin{bmatrix} 1 & 1 & . & . & . \\ 1 & 1 & . & . & . \\ . & . & . & . & . \\ . & . & . & 1 & 1 \\ . & . & . & 1 & 1 \end{bmatrix}.$$

They impose the restriction onto \bar{C} to be also symmetric to the co-diagonal, i.e. they allow an approximation which yields the structure:

$$\bar{C} = \begin{bmatrix} & & & & & \\ & & & & & \\ & & a & b & c & d & e \\ & & b & f & g & h & d \\ & & c & g & i & g & c \\ & & d & h & g & f & b \\ & & e & d & c & b & a \\ & & & & & & \end{bmatrix}$$

This is useful for a symmetric pattern of observation as e.g. reseau coordinates are. This Jordan algebra is the first example in which B has a dimension $n < k < n(n+1)/2$!

Example 7 : (Kronecker product)

The following Jordan algebra, based on an idea of Pukelsheim (1981), is not commutative

$$B_7 = \{ \underline{A}_i \otimes \underline{A}_j \otimes \dots \otimes \underline{A}_k \mid \underline{A}_i \in B_1, \underline{A}_j \in B_m, \underline{A}_k \in B_n; B_1, B_m, B_n \text{ are special Jordan algebras } \in \text{Sym}(1), \text{Sym}(m), \text{Sym}(n) \text{ resp., all must be commutative except one} \}$$

Proof : using Lemma 1c (see appendix B)

This seemingly complicated Jordan algebra is used for the approximation of the 25×25 matrix $\underline{C}_{xx}^x \otimes \underline{C}_{xx}^y$ in eq. (2.9), see ch. 5.2.

If $\underline{C}_{xx}^x \in B_3$ and $\underline{C}_{xx}^y \in B_6$ the dimension of B_7 is $5 \times 9 = 45$, thus being much larger than 25, the rank of \underline{C}_{xx} . This allows a closer approximation than the combination \underline{C}_{xx}^x and $\underline{C}_{xx}^y \in B_3$. Nevertheless, the reduction from $25 \times 26 / 2 = 325$ elements is still large enough. A further reduction to 6 parameters will be discussed in the next chapter.

Example 8 : (Hadamard product)

The approximation of the full covariance matrix eq. (2.9) can be based on the following Jordan algebra

$$B_8 = \left\{ \begin{bmatrix} \underline{A}_i & \underline{A}_i^T & \underline{A}_i & \underline{B}_j^T \\ \underline{B}_j & \underline{A}_i^T & \underline{B}_j & \underline{B}_j^T \end{bmatrix} * \begin{bmatrix} \underline{C}_k & \underline{C}_k^T & \underline{C}_k & \underline{D}_1^T \\ \underline{D}_1 & \underline{C}_k^T & \underline{D}_1 & \underline{D}_1^T \end{bmatrix} \mid \underline{A}_i, \underline{B}_j, \underline{C}_k, \underline{D}_1 \text{ being 4 sets of mutually orthogonal normalized matrices} \right\}$$

Setting $\underline{B}_j = -\underline{C}_k = \frac{1}{\sqrt{n}} \underline{1}$ and $\underline{A}_i = \underline{D}_1 = \frac{1}{\sqrt{10}} (2 \ 1 \ 0 \ -1 \ -2)^T$ yields a base matrix representing a scale variation in a 5×5 reseau grid, whereas $\underline{A}_i = \frac{1}{\sqrt{34}} (4 \ 1 \ 0 \ 1 \ 4)^T$,

$\underline{B}_j = \frac{1}{\sqrt{10}} (2 \ 1 \ 0 \ -1 \ -2)^T$, $\underline{C}_k = \frac{1}{\sqrt{5}} (1 \ 1 \ 1 \ 1 \ 1)^T$, $\underline{D}_1 = \frac{1}{\sqrt{10}} (2 \ 1 \ 0 \ -1 \ -2)^T$ yields a variation of longitudinal tilt (rotation about the y-axis).

We have not applied this Jordan algebra for variance component estimation. But we have used the structure to extrapolate $\underline{\bar{C}}_{xy}$ from an approximation of $\underline{\bar{C}}_{xx}$ and $\underline{\bar{C}}_{yy}$ after extraction of the matrices \underline{A}_i , \underline{B}_j , \underline{C}_k and \underline{D}_l , which is described in ch. 3.4.3..

The following example only shows that the concept of variance components with a special Jordan algebra can be linked with the estimation of a covariance function of a discrete time series.

Example 9 : (zyclical Toeplitz matrices)

The Jordan algebra

$$B_9 = \left\{ \underline{V}^i = (v_{kl}^i) = \begin{cases} 1 & \text{for } |k-l| = i \text{ or } |k-l| = n+1-i \\ 0 & \text{otherwise} \end{cases} \mid i = 0, \dots, [n/2] \right\}$$

$[n/2] = \text{smallest integer} > n/2$

consists of all zyclical matrices with Toeplitz structure. It is not commutative. However the normal equation matrix (eq. 32.4, $\underline{F} = \underline{I}$) is $\underline{D}^T \underline{D} = 2 n \underline{I}$. The estimation is simply an averaging over all covariances with identical difference of indices: $(\underline{D}^T \text{vec} \hat{\underline{C}})_i = \sum_{\substack{|k-l|=i \\ \text{or } |k-l|=n+1-i}} c_{kl}$. The fact, that B_9 is a special Jordan algebra

proves the positivity of the estimated covariance function $\hat{\underline{t}} = (t_1, \dots, t_{[n/2]})$ of a discrete time series if periodicity is assumed.

Counterexample : (length dependent variance of distances)

The set

$$B = \{ \underline{I}, \text{diag}(s_i^2) \mid \text{not all } s_i \text{ identical} \}$$

is no special Jordan algebra, as the second matrix is not idempotent.

This might be an explanation for the existence of negative variance components in this special case. The conclusion can be

- that a length dependent variance is not possible without correlation between distances. In this case theory would unveil a too much simplified mathematical model,
- or
- that the condition, the set of base matrices spanning a special Jordan algebra, is not a necessary one.

The examples show a great variety of applicable basis sets for variance components. We are now ready to analyse the estimation process itself and check wheather it is possible to reach the desired structure of the estimated covariance matrix from eq. (2.9).

3.4 Separation of approximated covariance matrices

The aim is to use the special Jordan algebra B_7 for an approximation of the submatrices \underline{C}_{xx} and \underline{C}_{yy} , separate them into Kronecker products and then extrapolate \underline{C}_{xy} according to the structure of B_8 , if possible.

3.4.1 Variance components with basis B_7

The model for the variance component estimation is

$$\hat{\underline{C}}_{xx} = \sum_{ij} t_{ij} \underline{v}_{ij}, \quad (341.1)$$

with

$$\underline{v}_{ij} = \underline{v}_i^x \otimes \underline{v}_j^y, \quad \underline{v}_i^x \in B_x, \quad \underline{v}_j^y \in B_y \quad (341.2)$$

The specification of B_x and B_y is irrelevant in this context. $\hat{\underline{C}}_{yy}$ is treated similarly. If we introduce the matrix

$$\underline{I} = (t_{ij}), \quad (341.3)$$

the least squares solution for $\hat{\underline{I}}$ is:

$$\hat{\underline{I}} = \text{vec} \hat{\underline{I}} = (\underline{D}^T \underline{D})^{-1} \underline{D}^T \text{vec} \underline{C} = \underline{S}^{-1} \text{vec} \underline{Q} \quad (341.4)$$

with $\underline{S} = \underline{D}^T \underline{D}$ and $\text{vec} \underline{Q} = \underline{D}^T \text{vec} \hat{\underline{C}}$. As the matrices \underline{v}_{ij} have the special structure eq. (314.2) the solution can be rewritten

$$\hat{\underline{I}} = \underline{S}_x^{-1} \underline{Q} \underline{S}_y^{-1}, \quad (341.5)$$

with $\underline{S}_x = (\text{tr}(\underline{v}_i^x \underline{v}_k^x))$ and $\underline{S}_y = (\text{tr}(\underline{v}_j^y \underline{v}_l^y))$. (The proof uses $\text{tr}(\underline{A} \otimes \underline{B}) = \text{tr} \underline{A} \cdot \text{tr} \underline{B}$).

Thus the special structure of the basis matrices leads to a simplified normal equation system, reducing core requirement and computing time.

3.4.2 Approximation of $\hat{\underline{I}}$

The approximated matrix

$$\bar{\underline{C}}_{xx} = \sum_{ij} \hat{t}_{ij} (\underline{v}_i^x \otimes \underline{v}_j^y) \quad (342.1)$$

does not have the desired structure $\underline{C}_{xx}^x \otimes \underline{C}_{xx}^y$. Only if $\hat{\underline{I}}$ has rank 1, i.e. being separable into

$$\hat{\underline{I}} = \underline{a} \underline{b}^T, \quad (342.2)$$

with $\underline{a} = (a_i)$ and $\underline{b} = (b_j)$, $\bar{\underline{C}}_{xx}$ can be written as a Kronecker product

$$\bar{\underline{C}}_{xx} = \sum_{ij} a_i b_j (\underline{v}_i^x \otimes \underline{v}_j^y) = \sum_i a_i \underline{v}_i^x \otimes \sum_j b_j \underline{v}_j^y. \quad (342.3)$$

The rank of $\hat{\underline{I}}$ can be obtained by a singular value decomposition (SVD) (cf. Rao, 1973, p. 42) :

$$\hat{\underline{I}} = \sum_{i=1}^r \lambda_i \underline{a}_i \underline{b}_i^T, \quad r = r(\hat{\underline{I}}), \quad (342.4)$$

where the \underline{a}_i and the \underline{b}_i are sets of mutually orthogonal normalized vectors.

Generally $\hat{\underline{\Gamma}}$ will - at least numerically - have full rank. But if $\lambda_1 \gg \lambda_2 > \dots > \lambda_r$, $\hat{\underline{\Gamma}}$ can be approximated using the largest eigenvalue and the corresponding eigenvectors $\underline{a} = \underline{a}_1 \sqrt{\lambda_1}$ and $\underline{b} = \underline{b}_1 \sqrt{\lambda_1}$ leading to $\underline{\Gamma} = \underline{a} \underline{b}^T$.

This step, of course, disturbs the estimation process fundamentally, so that after the approximation of $\hat{\underline{\Gamma}}$ the positive definiteness of the covariance matrix can not be guaranteed any more:

$$\tilde{\underline{C}}_{xx} = \tilde{\underline{C}}_{xx}^x \otimes \tilde{\underline{C}}_{xx}^y = \sum_i \underline{a}_i \underline{v}_i^x \otimes \sum_j \underline{b}_j \underline{v}_j^y \quad (342.5)$$

It has to be checked separately.

If one of the two components $\tilde{\underline{C}}_{xx}^x$ and $\tilde{\underline{C}}_{xx}^y$ is not p. d. one might use only the positive parts, which can be obtained after a SVD yielding e. g.

$$\tilde{\underline{C}}_{xx}^x = \sum_i \underline{u}_i \underline{u}_i^T \lambda_i \quad (342.6)$$

by restricting the sum to the positive eigenvalues or even only to the largest one:

$$\dot{\underline{C}}_{xx}^x = \sum_{\lambda_i > 0} \underline{u}_i \underline{u}_i^T \lambda_i. \quad (342.7)$$

The resulting matrix now is p. s. d..

3.4.3 Extrapolation of \underline{C}_{xy}

Let us assume that the separation of \underline{C}_{xx} and \underline{C}_{yy} has led to an acceptable approximation, then we are able to extrapolate \underline{C}_{xy} using the structure of B_8 . To do so we use the positive parts of e. g. $\tilde{\underline{C}}_{xx}^x$ and $\tilde{\underline{C}}_{xx}^y$ leading to a decomposition

$$\dot{\underline{C}}_{xx}^x = \sum_{\lambda_i > 0} (\underline{u}_i \sqrt{\lambda_i}) (\underline{u}_i \sqrt{\lambda_i})^T = \underline{U} \underline{U}^T$$

and

$$\dot{\underline{C}}_{yy}^x = \sum_{\lambda_i > 0} (\underline{v}_i \sqrt{\lambda_i}) (\underline{v}_i \sqrt{\lambda_i})^T = \underline{V} \underline{V}^T.$$

With the matrices \underline{U} and \underline{V} the matrix $\dot{\underline{C}}_{xy}^x$ becomes the form

$$\dot{\underline{C}}_{xy}^x = \underline{U} \underline{V}^T.$$

$\dot{\underline{C}}_{xy}^y$ is approximated similarly. Thus the approximation of the matrix $\hat{\underline{C}}$ is complete.

3.5 Principal component analysis

The analytical simplification of the structure of $\hat{\underline{C}}$, as described in the last chapter can be circumvented by a principal component analysis, i. e. a SVD of $\hat{\underline{C}}$. This is a one step procedure, where standard programs are available. The disadvantage is the interpretation of the components, which has to be done by visual inspection.

The SVD of $\hat{\underline{C}}$

$$\hat{\underline{C}} = \underline{E} \underline{A} \underline{E}^T = \sum_i \lambda_i \underline{e}_i \underline{e}_i^T = \sum_i \underline{f}_i \underline{f}_i^T \quad (35.1)$$

leads to vectors \underline{f}_i , which are mutually orthogonal. Each \underline{f}_i can be interpreted as a variation of the coordinates l_j around their mean value with a standard deviation of f_{ij} . A graph of the \underline{f}_i admits an evaluation of the component and might give insight into physical effects.

On the other hand, as $\text{tr}(\hat{\underline{C}}) = \sum \lambda_i$, $\hat{\underline{C}}$ can be approximated restricting the sum eq. (35.1) to the k largest eigenvalues or to those for which the sum is greater than a certain percentage of $\text{tr}(\hat{\underline{C}})$:

$$\bar{\underline{C}} = \sum_{i=1}^k \lambda_i \underline{e}_i \underline{e}_i^T. \quad (35.2)$$

This matrix again is singular with $r(\bar{\underline{C}}) = k$, thus having the same disadvantages as $\hat{\underline{C}}$ consisting of the submatrices $\hat{\underline{C}}_{xx}^x$ etc. (eq. (342.7)).

3.6 Evaluation of the approximation

A statistical test for the variance components \hat{t}_k (eq. (32.4)) is available, if a commuting special Jordan algebra (ch. 3.3) is used and the basis matrices \underline{V}_k are mutually orthogonal idempotent matrices. If $r(\underline{V}_k) = 1$, then

$$\hat{t}_k \sim \chi^2 \sigma_k^2, \text{ with } \sigma_k^2 = \text{tr}(\underline{C} \underline{V}_k).$$

In all other cases the similarity of two covariance matrices \underline{C} and $\bar{\underline{C}}$, $\bar{\underline{C}}$ being an approximation of \underline{C} , can at least be checked using the spectrum $\{\lambda_i\}$ of $\bar{\underline{C}} \underline{C}^{-1}$, resulting from the general eigenvalue problem

$$|\bar{\underline{C}} - \lambda \underline{C}| = 0. \quad (36.1)$$

The following three criteria use the eigenvalues of (36.1)

3.6.1 Quotient criterion

The quotient

$$k = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (\lambda_{\min} = \min_{\lambda_i > 0} \lambda_i) \quad (36.1)$$

is a measure for the quality of the approximation, if $\bar{\underline{C}} = \underline{C}$, the quotient is 1. The

Raleigh-Quotient

$$\lambda_x = \frac{\underline{x}^T \bar{\underline{C}} \underline{x}}{\underline{x}^T \underline{C} \underline{x}} \quad (36.2)$$

is bounded, lying in the range between λ_{\min} and λ_{\max} . It can be interpreted as the quotient of two weight coefficients, the factor \sqrt{k} thus gives the maximum deviation between the standard deviations calculated with $\bar{\underline{C}}$ and \underline{C} resp., e. g. $k = 100$ indicates that the standard deviations of arbitrary functions $\underline{x}^T \underline{1}$ of the coordinates calculated with $\bar{\underline{C}}$ vary 10 times more than if they were calculated with \underline{C} .

This criterion only uses the extreme values of the spectrum, whereas the following criteria use all eigenvalues and adopt the matrix \underline{C} to be Wishart distributed. The tests are generalized to singular Wishart distributions, which can be proved by a proper S-transformation.

3.6.2 Test on the equality of \underline{C} and \underline{C}

If the $n \times n$ matrix \underline{C} , $r(\underline{C}) = r$, is Wishart distributed, $p\underline{C} \sim W(p, \underline{C})$, the test on $\underline{C} = \underline{C}$ uses the test statistic

$$T_1 = p \left(\sum_{\lambda_i \neq 0} (\ln \lambda_i + \lambda_i) - r \right) \quad (362.1)$$

(cf. Koch, 1980, p.138), which for $p \rightarrow \infty$ follows a $\chi^2(r(r+1)/2)$ -distribution.

3.6.3 The sphericity test

If the $n \times n$ matrix \underline{C} is Wishart distributed (cf. 3.6.2), the test on $\underline{C} = \sigma_0^2 \underline{C}$ with σ_0^2 unknown uses the test statistic

$$T_2 = \prod_{\lambda_i \neq 0} \lambda_i / \left(\sum_i \lambda_i / r \right)^r \quad (363.1)$$

(cf. Giri, 1977), with the approximation

$$P(- (p-1) \rho \log T_2 \leq z) = P(\chi_f^2 \leq z), \quad (363.2)$$

with $\rho = 1 - (2r^2 + r + 2) / (6r(p-1))$ and $f = r(r+1)/2 + 1$ for large r .

4. Comparison with other Methods

4.1 Methods for the approximation of covariance matrices

There exist several other methods for the approximation or analysis of covariance matrices. The properties of the most important ones are collected in table 1 in sequence of increasing strength of the assumptions, i. e. of decreasing flexibility.

Table 1 Methods for the approximation of a $n \times n$ covariance matrix \underline{C}

Method	Given Information	Free Parameters	Model ($k=1, \dots, k$)	Conditions/Abbreviations
Factor Analysis	-	$\underline{e}_k, \lambda_k, d_{ii}, k$	$\hat{\underline{C}} = \text{diag}(d_{ii}) + \sum_k \underline{e}_k \underline{e}_k^T \lambda_k$	$\underline{e}_k^T \underline{e}_{k'} = \delta_{kk'}$
Principal Component Analysis	-	$\underline{e}_k, \lambda_k, k$	$\hat{\underline{C}} = \sum_k \underline{e}_k \underline{e}_k^T \lambda_k$	$\underline{e}_k^T \underline{e}_{k'} = \delta_{kk'}$
Variance Component Estimation	\underline{v}_k	t_k, k	$\hat{\underline{C}} = \sum_k t_k \underline{v}_k$	-
Analysis of Covariance Matrices	c_{ij}, g_k	$\sigma_0^2, t_{kk'}, k$	$\hat{\underline{C}} = \sigma_0^2 \{ (c_{ij}) + \sum_{k,k'} g_k g_{k'}^T t_{kk'} \}$	special struct. of $(t_{kk'})$
Fourier Analysis	\underline{f}_k	k	$\hat{\underline{C}} = \sum_k \underline{f}_k \underline{f}_k^* \underline{f}_k^* \underline{C} \underline{f}_k$	$\underline{F} = (\sqrt{n} \underline{f}_k) = \text{Fouriermatrix}$ $= \exp(-j k 2\pi/n), j^2 = -1$

Variance component estimation reveals to be a compromise between the most flexible method of factor analysis, which can be interpreted as a generalization of variance components, and the unflexible method of Fourier analysis, which uses a special set of basis matrices, namely $\{\underline{V}_k\} = \{\underline{f}_k \underline{f}_k^*\}$ but forms a link to variance functions (cf. Ellenbeck, 1976; Schuh, 1981). Factor analysis, being numerically involving, has no unique solution, the conditions enforcing a unique solution making the interpretation difficult (cf. Press, 1972). These disadvantages are avoided by the principle component analysis; but this is purchased with an approximated matrix which always is singular, though non negative, and whose rank may be data dependent. Without a priori information factor and principle component analysis may give a good starting point, being a disadvantage if one searches certain structures. The only opponent to variance component estimation is Ebner's theory for the analysis of covariance matrices (1975), which has certain tempting properties, variance components do not have.

4.2 On Ebner's theory for the analysis of covariance matrices

We adapt Ebner's notation of (1975). The idea of his theory is to check whether the hypothesis

$$H_0 : \hat{\underline{C}} = \underline{C} + \underline{G} (\underline{G}^T \underline{I} \underline{G})^{-1} \underline{G}^T$$

holds true. \underline{C} describes the assumed covariance matrix of the coordinates, \underline{G} the effect of the filter parameters, i.e. the systematic effects, and \underline{I} weights these effects and is unknown. The column vectors \underline{g}_i of \underline{G} correspond to the vectors \underline{c}_i defining $\underline{V}_i = \underline{c}_i \underline{c}_i^T$. In order to check the choice of \underline{C} and \underline{G} separately, Ebner proposes to use two criteria. The corresponding tests are given below, modifying the criteria of Ebner.

The tests are based on two alternative hypotheses. The first one serves for checking the choice of \underline{C} , assuming the systematic effects are contained in \underline{G} :

$$H_1 : \hat{\underline{C}} = \underline{C}_1 + \underline{G} (\underline{G}^T \underline{I} \underline{G})^{-1} \underline{G}^T.$$

As the significance of the parameters \underline{G} is not known, they are eliminated by a S-transformation using \underline{P} , which contains the first n-u rows of the S-matrix $\underline{D}_G = \underline{I} - \underline{G} (\underline{G}^T \underline{E}_r \underline{G})^{-1} \underline{G}^T \underline{E}_r$ (with \underline{E}_r from eq. 313.3, u = number of filter parameters = r(G)). If $\hat{\underline{C}}$ is supposed to be Wishart distributed with p degrees of freedom, $\underline{P} \hat{\underline{C}} \underline{P}^T \sim W(p, \underline{P} \underline{C} \underline{P}^T)$ under H_0 . Thus the

$$\text{Criterion I: } \underline{P} \hat{\underline{C}} \underline{P}^T = \underline{P} \underline{C} \underline{P}^T$$

can be checked with the test given in ch. 3.6.2.

If H_0 is accepted, the second test compares H_0 with

$$H_2 : \hat{\underline{C}} = \underline{C}$$

to find out whether the filterparameters are significant. Exchanging the rôle of H_0 and H_2

$$\text{Criterion II: } \underline{H}_G \hat{\underline{C}} \underline{H}_G^T = \underline{H}_G \underline{C} \underline{H}_G^T,$$

with $\underline{H}_G = (\underline{G}^T \underline{C}^{-1} \underline{G})^{-1} \underline{G}^T \underline{C}^{-1}$, can be checked by the same test, as $\underline{H}_G \hat{\underline{C}} \underline{H}_G^T \sim W(p, \underline{H}_G \underline{C} \underline{H}_G^T)$ under H_2 , which plays the rôle of the null hypothesis. If criterion II is not fulfilled, the filter parameters are significant.

If the null-hypothesis is generalized to $H_0 : \hat{C} = \sigma_0^2 Q + G(G^T I G)^{-1} G^T$ the first criterion can be checked with the sphericity test of ch. 3.6.3.

Ebner's method has two advantages:

- it courses no difficulties with the positive definiteness of the matrices and
- it solves the problem of defining an appropriate coordinate system, as criterion I is simply a S-transformation which may include the datum parameters.

Though one is forced to assume a covariance matrix, which is not changed - only checked - by the analysis, this theory is worth to be investigated further.

5. Examples

The following examples show an application of the theory of ch. 3.. The data consist of $p = 76$ successive images of a film of the OEEPE Oberschwaben test (Ackermann, 1973). The estimation of the covariance matrix and its principal component analysis have been published by Reustlen (1980). His results are compared with those of the variance component estimation showing a high agreement which can be explained by the predominant influence of the varying scale factor.

5.1 Approximation of a 5x5 submatrix of \hat{C}

We begin with an approximation of a small submatrix to show the influence of different Jordan algebras and a preceding S-transformation onto the estimation process. Fig. 2 shows the covariance matrix of the x-coordinates of the 5 points in the first row.

$$\begin{bmatrix} 43.875 & 19.193 & -4.205 & -19.149 & -36.967 \\ & 13.909 & 1.076 & -7.406 & -17.698 \\ & & 9.470 & 5.138 & 3.104 \\ & & & 16.610 & 20.599 \\ \text{symmetrical} & & & & 46.851 \end{bmatrix} \quad (51.1)$$

Fig. 2 Estimated Covariance Matrix of Image Coordinates
(Submatrix of \hat{C})

Three types of spanning sets are used as basis matrices

- a.) the set $B(9)$, consisting of the 9 matrices given in example 6 (ch. 3.3)
- b.) the set $B(5)$, consisting of the 5 matrices, using the vectors \underline{c}_i from example 3: $B(5) = \{\underline{c}_i \underline{c}_i^T | i=1, \dots, 5\}$.
- c.) the set $B(4)$, consisting of only 4 matrices, also using the vectors \underline{c}_i from example 3: $B(4) = \{\underline{c}_i \underline{c}_i^T | i=1, \dots, 4\}$. This set, containing matrices with rank defekt $d \geq 1$, is intended to show an application of the extended theorem 2 in Pukelsheim, Styan (1979), which enables us to approximate covariance matrices having a given rank deficiency.

The approximation of the matrix (51.1) with the set $B(9)$ resulted in the matrix given in fig. 3

$$\begin{bmatrix} 45.363 & 19.896 & -0.550 & -18.824 & -36.967 \\ & 15.260 & 3.107 & -7.406 & -18.824 \\ & & 9.470 & 3.107 & -0.550 \\ & & & 15.260 & 19.896 \\ \text{symmetrical} & & & & 45.363 \end{bmatrix} \quad (51.2)$$

Fig. 3 Approximation of Covariance Matrix from fig. 2 using eq. (342.1)

The variance components are $\hat{\underline{t}}^T = (44.01, 8.94, -14.91, 10.37, -20.45, -3.01, 0.03, 38.32)$, which contain negative values. Nevertheless, the matrix (51.2) is p.d. and has the desired structure, being symmetrical also with respect to the co-diagonal.

Variance component estimation using B(9) is just averaging the corresponding co-variances, e.g. $45.363 = (43.875 + 46.851)/2$. Obviously the approximation is extremely good, which is proved by the quotient criterion $k = 1.54$.

In order to eliminate a possibly hidden datum parameter we transform the matrix (51.1) with a S-transformation using the S-matrix $\underline{S} = \underline{I} - \underline{J}_5 \in B(9)$, which eliminates a common shift. The quality of the approximation with the 3 sets are found in table 2.

set	k
B(9)	1.44
B(5)	1.88
B(4)	1.88

Table 2

Quality of Approximation of 5x5 matrix after a S-transformation

Remark :

- the S-transformation does not influence the quality of the approximation very much,
- the structure of the matrix allows a good approximation with already 5 basis matrices,
- as the rank of the matrix after the transformation is only 4, omitting the last basis matrix (i.e. the unit matrix) has no influence on the approximation.

The reason for this result is the predominant effect of the scale variation, represented by $\underline{V}_2 = \underline{c}_2 \underline{c}_2^T$, $\underline{c}_2^T = (2 \ 1 \ 0 \ -1 \ -2)$.

5.2 Approximation of $\hat{\underline{c}}_{xx}$ and $\hat{\underline{c}}_{yy}$

The 25x25 submatrices $\hat{\underline{c}}_{xx}$ and $\hat{\underline{c}}_{yy}$ were approximated without preceeding S-transformation, as this would have complicated the construction of the spanning set. Thus no variance component estimation in the strict sense is applied, because the matrices are not Wishart distributed.

Different Jordan algebras were applied according to eq. (34.2). The basic sets B_x and B_y for \underline{V}_i^x and \underline{V}_j^y and the corresponding quotients k are given in table 3. The quality of the approximation is moderate for $\hat{\underline{c}}_{xx}$ and differs significantly from the high quality of \underline{c}_{yy} . On the other hand the refinement of the Jordan algebra B_y leads to a significantly higher improvement (lines 1 and 3) of the approximation of $\hat{\underline{c}}_{xx}$ than the refinement of B_x (lines 1 and 2). This seems to suggest, that $\hat{\underline{c}}_{xx}$ has a more complicated structure than $\hat{\underline{c}}_{yy}$ and is more influ-

	B_x	B_y	$k = \lambda_{\max}/\lambda_{\min}$		v
			\underline{C}_{xx}	\underline{C}_{yy}	
1	(5)	(5)	25.1	8.83	25
2	(9)	(5)	24.1	8.58	45
3	(5)	(9)	19.8	8.88	45
4	(9)	(9)	18.4	8.97	81

Table 3

Quality of approximation of \underline{C}_{xx} and \underline{C}_{yy} using different sets for a variance component estimation and number v of free parameters.

enced by the y -coordinate of a point than by the x -coordinate. A physical explanation could be the transport of the film in x -direction, which might lead to an unhomogeneous deformation being unsymmetric in y -direction. This would also explain the good approximation of \underline{C}_{yy} .

The last line shows that a good approximation can be achieved though no Jordan algebra is used, as $B(9)$ is not commutative. The value $k = 8.97$ for the approximation of \underline{C}_{yy} is the largest in that column. This can be explained by the difference between the object function of the approximation (least squares) and of the evaluation function (quotient of eigenvalues).

The evaluation is supplemented by two statistical tests. It is assumed that \underline{C}_{xx} and \underline{C}_{yy} are regular Wishart distributed matrices and the approximations obtained with $B_x = B(5)$ and $B_y = B(9)$ (table 3, line 3) are given matrices. Though this assumption is not valid the tests give valuable information about the approximation. The first test against $H_0: \hat{\underline{C}}_{xx} = \underline{C}_{xx}$ (eq. 362.1) leads to a rejection of H_0 , whereas the second test against $H_0: \hat{\underline{C}}_{xx} = \sigma_0^2 \underline{C}_{xx}$ (σ_0^2 unknown) does not suggest a rejection (cf. table 4). The same holds true for the approximation of \underline{C}_{yy} . The results indicate that under a further (moderate) relaxation of the hypotheses the approximation would also be accepted

	$H_0: \hat{\underline{C}} = \underline{C}^{1)}$	$H_0: \hat{\underline{C}} = \sigma_0^2 \underline{C}$
$\hat{\underline{C}}_{xx}$	$2.66 > F_{0.001} = 1.26$	$230 < \chi_{0.001}^2 = 251$
$\hat{\underline{C}}_{yy}$	$1.34 > F_{0.001} = 1.26$	$137 < \chi_{0.001}^2 = 251$

Table 4 Tests of Quality

¹⁾ cf. Koch, 1980, p. 138.

5.3 Separation of \underline{C}_{xx} and \underline{C}_{yy}

The separation of \underline{C}_{xx} and \underline{C}_{yy} into Kronecker products is based on the SVD of $\hat{\underline{I}}$. The eigenvalues of $\hat{\underline{I}}$ are given in table 5.

$\hat{\underline{I}}_{xx}$	833.736	0.548	0.135	0.005	0.0004
$\hat{\underline{I}}_{yy}$	561.127	9.322	0.313	0.022	0.0002

Table 5 Eigenvalues of the Variance Components $\hat{\underline{I}}_{xx}$ and $\hat{\underline{I}}_{yy}$ of the Approximated Matrices \underline{C}_{xx} and \underline{C}_{yy}

The spectra show a significance dominance of the largest eigenvalues in both cases. The relative small 2nd eigenvalue of $\hat{\underline{I}}_{xx}$ allows a separation of $\hat{\underline{C}}_{xx} =$

$\tilde{\underline{C}}_{xx}^x \otimes \tilde{\underline{C}}_{xx}^y$ which is p.d. (eq. 342.5). However the separation of $\tilde{\underline{C}}_{yy}$ leads to an indefinite matrix $\tilde{\underline{C}}_{yy}^x$. Obviously the 2nd eigenvalue of $\tilde{\underline{C}}_{yy}$ contains information which is not neglectable. However the interpretaion of the corresponding variance matrix was not possible.

5.4 Composition of $\dot{\underline{C}}$

The further separation of $\tilde{\underline{C}}_{xx}$ and $\tilde{\underline{C}}_{yy}$ therefore uses only the components of $\tilde{\underline{C}}_{xx}^x$, $\tilde{\underline{C}}_{xx}^y$, $\tilde{\underline{C}}_{yy}^x$ and $\tilde{\underline{C}}_{yy}^y$ belonging to the largest eigenvalues of these matrices. The eigenvectors are given in table 6.

	$\dot{\underline{C}}_{xx}^x$	$\dot{\underline{C}}_{xx}^y$	$\dot{\underline{C}}_{yy}^y$	$\dot{\underline{C}}_{yy}^x$
	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>
1	2.318	2.650	4.392	1.418
2	2.318	1.333	2.200	1.471
3	2.318	0.000	0.000	1.385
4	2.318	-1.333	-2.200	1.471
5	2.318	-2.650	-4.392	1.418

Table 6

Eigenvalues a, b, c and d of the submatrices generating \underline{C}

They reveal a simple structure, which was already given in example 8 (ch. 3.3). If one assumes that \underline{C}_{xy} can be derived by using this structure, we obtain the matrix

$$\dot{\underline{C}} = \begin{bmatrix} \underline{a} & \underline{a}^T \otimes \underline{c} & \underline{c}^T \\ \underline{b} & \underline{a}^T \otimes \underline{d} & \underline{c}^T \\ \underline{b} & \underline{b}^T \otimes \underline{d} & \underline{d}^T \end{bmatrix} = \begin{bmatrix} \underline{a} \otimes \underline{c} \\ \underline{b} \otimes \underline{d} \end{bmatrix} \begin{bmatrix} \underline{a}^T \otimes \underline{c}^T, \underline{b}^T \otimes \underline{d}^T \end{bmatrix}$$

Because of symmetry, it only depends on $3 \times 4 = 12$ different elements to be compared with the 1275 elements of the whole matrix. $\dot{\underline{C}}$ describes a variation of the scale of a 25 point grid. Introducing independent equally weighted measuring errors with variance $\sigma_0^2 = (2 \mu m)^2$, which has been estimated from double measurements, enables an evaluation of $\dot{\underline{C}} + \sigma_0^2 \underline{I}$, as $r(\dot{\underline{C}}) = 1$ making a comparison of $\dot{\underline{C}}$ and $\hat{\underline{C}}$ not meaningful. The quality of the approximation seems to be not acceptable, as $k = 309$. But the principle component analysis gives no better results.

5.5 Principal component analysis of $\hat{\underline{C}}$

The principal component analysis of $\hat{\underline{C}}$ yields the broad spectrum shown in fig. 4. The first eigenvalue is dominant. On the other side there are also very small eigenvalues. The first two components are sketched in fig. 5 and 6 (next page). They show the already mentioned scale variation and a variation of the longitudinal tilt ϕ (rotation around the y-axis) including a cylindrical deformation. Both effects can be explained by the construction of either the camera or the copy-machine.

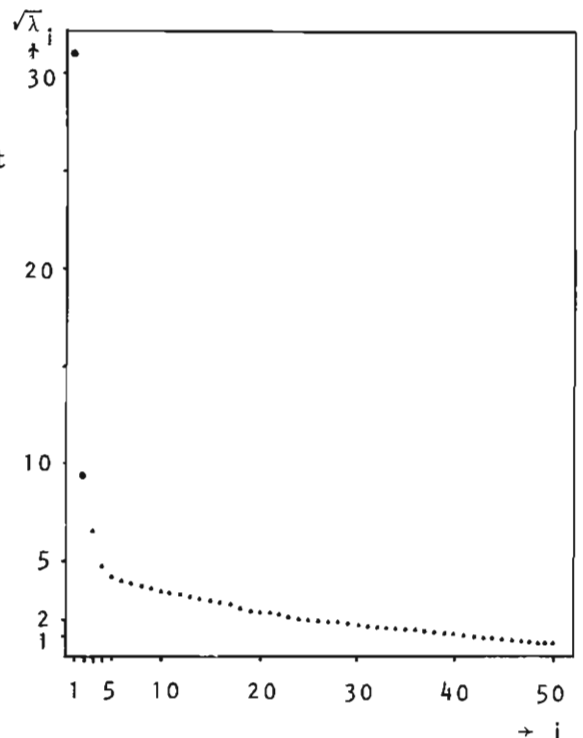


Fig. 4 Spectrum of $\hat{\underline{C}}$, $\sqrt{\lambda_{\max}} = 31$, $\sqrt{\lambda_{\min}} = 0.25$

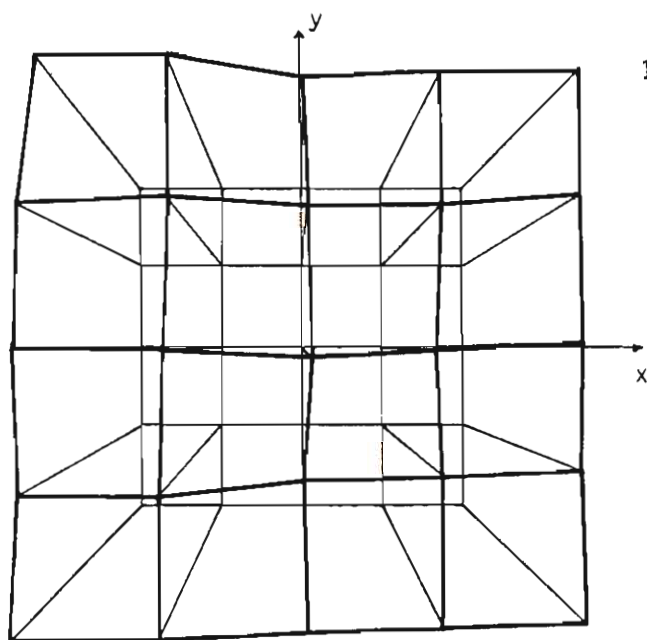


Fig. 5 1st Component of \hat{C}

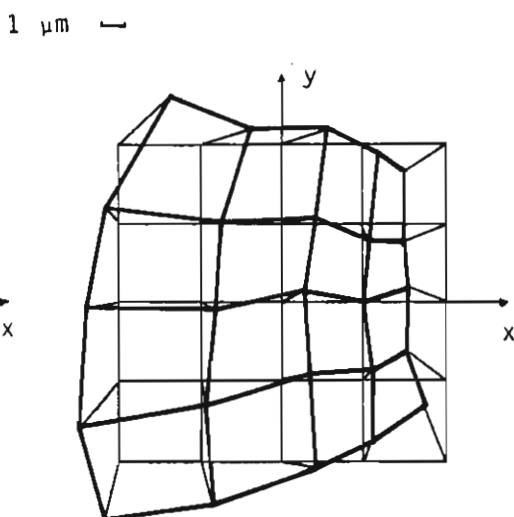


Fig. 6 2nd Component of \hat{C}

A comparison of this scale variation with that obtained by variance components reveals only a small difference of both vectors. The test of the 1st component of $\underline{C} + \sigma_0^2 \underline{I}$ against \hat{C} yields $k = 301$ (instead of 309). The vectors do not differ more than 7 % in length and direction. This is astonishing the more, as the approximation process using variance components seemed to loose a lot of information compared with the principal component analysis, and shows the power of the separation method. However, the ability to interpret also the second component (cf. fig. 6), which was not possible using variance components, demonstrates the higher flexibility to reveal hidden effects.

6. Discussion

The application of multivariate analysis for the estimation of covariance matrices of photogrammetric image coordinates required an adaption of the mathematical tools. The concept of the special Jordan algebra, guaranteeing the positive definiteness of the estimated covariance matrix, turned out to be an adequate method for this type of investigation. The results of the variance component estimation and the principal component analysis show a great coincidence. Thus the variance component estimation process in connection with the Jordan algebras is able to detect structures even if they are as complex as in this case.

The investigation, however, leaves some questions open:

1. The results have to be confirmed by further research using different data, different stochastic models, i. e. different sets of basis matrices and possibly different algorithms. This includes the estimation of the correlation matrix \underline{R} between the images and the influence of \underline{R} onto the estimation of \underline{C} .
2. The estimation of \underline{C}_{vv} takes care of the special definition of the coordinate system. A direct approximation of \hat{C}_{vv} would have been theoretically more satisfying, but was not tried as the 2nd. theorem of Pukelsheim and Styan (1979, not cited) does not guarantee the approximated matrix having the same null space as

\underline{C}_{VV} . An approximation method should be found which is able to handle singular matrices without the detour of approximating a substitute matrix with different rank. Perhaps this problem can be circumvented either by using the S-transformation with \underline{D}_A in the proof of theorem 1 or by Ebner's method, which simultaneously offers a statistical test.

3. The separation of a covariance matrix into Kronecker products leads to the serious problem of decomposing the matrix $\hat{\underline{I}}$, which here was tried to solve by a singular value decomposition. This is the sore spot in the decomposition of $\hat{\underline{C}}$, as the original guarantee for \underline{C} to be p.d. is lost. There might be, however, possibilities to approximate $\hat{\underline{I}}$ by a matrix of rank 1, which keeps the guarantee of positive definiteness for the corresponding approximation $\tilde{\underline{C}}$ of \underline{C} .

4. The assumption in theorem 2, that the basis matrices should span a special Jordan algebra, is only a necessary one. The basis matrices of the length dependent variance of distances not meeting this condition leave doubts about the necessity of the supposed condition.

5. The link of variance component estimation with the estimation of covariance functions needs further research. Only in the case of Fourier analysis an interpolation method for covariances exists. In this context special focus should be laid onto the approximation of singular covariance matrices by using covariance functions.

6. The methods discussed in this paper also could be used to estimate and approximate covariance matrices of geodetic observations. An example is the repeated measurement of distances or height differences to points which are spread over the horizon in various distances. In this case no net adjustment is necessary and the influence of direction and distance onto the variance covariance matrix can be separated.

Literature:

- Ackermann, F.: Testblock Oberschwaben. Program 1 - Results of Blockadjustments by Independent Models. OEEPE Publication Officielle N° 8, S. 15 - 21, 1973
- Ackermann, F.: Genauigkeitssteigerung durch Berücksichtigung der Korrelation bei der absoluten Orientierung des Bildpaars, DGK Reihe B, Heft 216, München 1976
- Ackermann, F. & Schilcher, M.: Auto- and Cross-Correlation of Image Coordinates. Pres. Paper, ISP-Commission III-Symposium, Moscow, Nachrichten aus dem Karten- und Vermessungswesen, Reihe B, Heft 36, S. 5 - 24, 1980
- Buck, U.: Numerische Auswertung und statistische Analyse von ingenieurphotogrammetrischen Deformationsmessungen, Wiss. Arb. d. Univ. Hannover, Heft 82, 1977
- Ebner, H.: Analysis of Covariance Matrices, ISP-Commission III-Symposium, Stuttgart, DGK Reihe B, Heft 214, S. 111 - 121, München 1975.
- Ellenbeck, H.: Untersuchung zum stochastischen Verhalten von Reseauaufnahmen, ISP Congress, Helsinki, 1976
- Giri, N. C.: Multivariate Statistical Inference. Academic Press, New York 1977
- Koch, K.-R.: Parameterschätzung und Hypothesentests. Dümmler Verlag, Bonn 1980
- v. Mierlo, J.: A Testing procedure for Analysing Geodetic Deformation Measurements, Pres. Paper to 11th Int. Symp. for Deformation Measurements, FIG, Bonn 1978
- Press, S. J.: Applied Multivariate Analysis. New York 1972
- Pukelsheim, F.: Linear Models and Convex Geometry: Aspects of Non-negative Variance Estimation, Extended Vers. of Pres. Paper at VIth Int. Conf. on Math. Stat., Poland, 1976 (1980)

Pukelsheim, F.: Personal Communication 1981

Pukelsheim, F. & Styan, G. P. H.: Nonnegative Definiteness of the Estimated Dispersion Matrix in a Multivariate Linear Model, Bull. de l'Académie Polonaise des Sciences, Serie des Sciences mathématiques, Vol. XXVII, No. 3-4, 1979

Rao, C. R.: Linear Statistical Inference and Its Applications. New York 1973

Reustlen, M.: Empirische Genauigkeitsuntersuchung von Bildkoordinaten, Diplomarbeit am Institut für Photogrammetrie der Universität Stuttgart, 1981

Schilcher, M.: Empirisch-statistische Untersuchung zur Genauigkeitsstruktur des photogrammetrischen Luftbilds. DGK Reihe C, Heft 262, München 1980

Schuh, H.: Spektralanalyse von Restverbesserungen im Meßbild, BuL Heft 1, S. 20 - 26, 1981

Searle, S. R.: Linear Models, New York, 1971

Seely, J.: Quadratic Subspaces and Completeness, Ann. of Math. Stat., Vol. 42, No. 2, p. 710-721, 1971

Stark, E.: Die Genauigkeitsstruktur im photogrammetrischen Einzelmodell, DGK Reihe C, Heft 193, München 1973

Appendix A: Proof of $\underline{V} = -\underline{D}_A \underline{L} \underline{D}_B^T$ (eq. 312.3)

Given: $E(\underline{L}) = \underline{A} \underline{X} + \underline{Z} \underline{B}^T$, $D(\text{vec} \underline{L}) = \underline{R} \otimes \underline{C}$, $E(\underline{L}) = \underline{L} + \underline{V}$

Proof: 1. Substitution: $\underline{W} = \underline{V} - \underline{Z} \underline{B}^T$

GMM given: $\underline{L} + \underline{W} = \underline{A} \underline{X}$, $D(\text{vec} \underline{L}) = \underline{R} \otimes \underline{C}$

Least squares (LS) solution: $\underline{W} = -\underline{D}_A \underline{L}$, $D(\text{vec} \underline{W}) = \underline{R} \otimes \underline{D}_A \underline{C} \underline{D}_A^T$

2. GMM given: $\underline{W} - \underline{V} = -\underline{Y} \underline{B}^T$, $D(\text{vec} \underline{W}) := \underline{R} \otimes \underline{C}_{\underline{W}\underline{W}}$

LS solution:

$$\begin{aligned} \text{vec} \underline{V} &= \text{vec} \underline{W} - (\underline{B} \otimes \underline{I}) \text{vec} \underline{Z} = \\ &= (\underline{I}_n \otimes \underline{I}_p - (\underline{B} \otimes \underline{I})(\underline{B}^T \otimes \underline{I})(\underline{R}^+ \otimes \underline{C}_{\underline{W}\underline{W}}^+)^+(\underline{B} \otimes \underline{I}))^+ (\underline{R}^+ \otimes \underline{C}_{\underline{W}\underline{W}}^+) \text{vec} \underline{W} \\ &= (\underline{I}_n \otimes \underline{I}_p) \text{vec} \underline{W} - \underline{B} (\underline{B}^T \underline{R}^{-1} \underline{B})^{-1} \otimes \underline{C}_{\underline{W}\underline{W}} \underline{C}_{\underline{V}\underline{V}}^+ \text{vec} \underline{W} \\ &= \text{vec} \underline{W} - \underbrace{\text{vec}(\underline{C}_{\underline{W}\underline{W}} \underline{C}_{\underline{W}\underline{W}}^+ \underline{W} \underline{B} (\underline{B}^T \underline{R}^{-1} \underline{B})^{-1} \underline{B}^T \underline{R}^{-1})}_{\equiv \underline{W}} \\ &= (\underline{I}_n \otimes \underline{I}_p - \underline{B} (\underline{B}^T \underline{R}^{-1} \underline{B})^{-1} \underline{B}^T \underline{R}^{-1} \otimes \underline{I}_p) \text{vec} \underline{W} \\ &= (\underline{D}_B \otimes \underline{I}_p) \text{vec} \underline{W} = \underline{W} \underline{D}_B^T \end{aligned}$$

3. With $\underline{W} = -\underline{D}_A \underline{L}$ follows $\underline{V} = -\underline{D}_A \underline{L} \underline{D}_B^T$

q.e.d.

Appendix B: Proof of

$$\underline{B} := (\underline{A}_1 \otimes \underline{A}_j \otimes \dots \otimes \underline{A}_k) \circ (\underline{A}_1, \otimes \underline{A}_j, \otimes \dots \otimes \underline{A}_k) \in B_7$$

Rearranging yields

$$\begin{aligned} \underline{B} &= (\underline{A}_1 \otimes \underline{A}_j \otimes \dots \otimes \underline{A}_k) (\underline{A}_1, \otimes \underline{A}_j, \otimes \dots \otimes \underline{A}_k) \\ &\quad + (\underline{A}_1, \otimes \underline{A}_j, \otimes \dots \otimes \underline{A}_k) (\underline{A}_1 \otimes \underline{A}_j \otimes \dots \otimes \underline{A}_k) \\ &= (\underline{A}_1 \underline{A}_1, \otimes \underline{A}_j \underline{A}_j, \otimes \dots \otimes \underline{A}_k \underline{A}_k) + (\underline{A}_1, \underline{A}_1 \otimes \underline{A}_j, \underline{A}_j \otimes \dots \otimes \underline{A}_k, \underline{A}_k) \end{aligned}$$

Now assume only B_1 to be non commutative, then

$$\begin{aligned} \underline{B} &= (\underline{A}_1 \underline{A}_1, \otimes \underline{A}_j \underline{A}_j, \otimes \dots \otimes \underline{A}_k \underline{A}_k) + (\underline{A}_1, \underline{A}_1 \otimes \underline{A}_j \underline{A}_j, \otimes \dots \otimes \underline{A}_k \underline{A}_k) \\ &= \underbrace{(\underline{A}_1 \underline{A}_1 + \underline{A}_1, \underline{A}_1)}_{\in B_1} \otimes \underbrace{\underline{A}_j \underline{A}_j}_{\in B_m} \otimes \dots \otimes \underbrace{\underline{A}_k \underline{A}_k}_{\in B_n} \end{aligned}$$

q.e.d.

SONDERDRUCK AUS:

PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM
ON GEODETIC NETWORKS AND COMPUTATIONS
OF THE INTERNATIONAL ASSOCIATION OF GEODESY
- COMMISSION X: CONTINENTAL NETWORKS -
Munich, August 31 to September 5, 1981

Volume VI:

Adjustment Procedures of Geodetic Networks

Veröffentlichungen der
Deutschen Geodätischen Kommission bei der
Bayerischen Akademie der Wissenschaften

Reihe B, Heft Nr. 258/VI, München 1982