ANALYSIS OF COVARIANCE MATRICES

SUMMARY
The paper starts with a representation of the concept of inner accuracy which was introduced by P. Meissl in 1962 and which is applied frequently in Geodesy. Proceeding from this concept a theory is developed, allowing for a rigorous analysis of covariance matrices. By this theory any given covariance matrix can be disintegrated into a covariance matrix of simpler structure and the effect of a set of filter parameters. An example shows how the analysis works and demonstrates the power of the theory.

THE CONCEPT OF INNER ACCURACY
In 1962 P. Meissl introduced a concept into Geodesy which allows to filter the effect of an arbitrary set of parameters out of a given covariance matrix. The accuracy remaining after filtering is called inner accuracy [1], [2], [3].

In Geodesy the filter parameters most of the time are restricted to shifts, rotations and eventually a scale factor. In this case the inner accuracy is the accuracy being liberated from the effect of shifts, rotations and scale. Moreover most of the geodetic applications of the theory of inner accuracy are related to free adjustments [4], [5], [6].

It should be emphasized, however, that the concept of inner accuracy is neither restricted to free adjustments nor to a certain number or type of filter parameters.

FORMULATION
We start from a random vector \( x \) and the associated covariance matrix \( M \). The vector \( x \) is split into the expectation vector \( E[x] \) and the increment vector \( dx \):

\[
x = E[x] + dx
\]

(1)

From \( dx \) we separate the effect of the filter parameters \( dt \), being represented by the filter matrix \( G \). The remaining vector is called \( dx \). The associated covariance matrix we call \( Q \).

\[
dx = dx - Gdt
\]

(2)

The filter parameters \( dt \) now are determined so that the trace of \( Q \) becomes a minimum.

\[
\text{tr}Q \rightarrow \text{Min}
\]

(3)

The derivation of the vector \( dt \) and of the corresponding covariance matrix \( R \) is given in [2]. The results are:

\[
dt = (G^TG)^{-1}G^Tdx
\]

(4)

\[
R = (G^TG)^{-1}G^TMG(G^TG)^{-1}
\]

(5)

Inserting (4) into (2) we get (I = unit matrix):

\[
dx = (I - G(G^TG)^{-1}G^T)dx
\]

(6)

\[
Q = (I - G(G^TG)^{-1}G^T)M(I - G(G^TG)^{-1}G^T)
\]

(7)

The covariance matrix \( Q \) represents the inner accuracy of the random vector \( x \) which remains when the effect of the filter parameters \( dt \) is eliminated. The results (4) and (6) are identical with the results of a least squares adjustment, which fits the vector \( x \) onto the vector \( E[x] \) using the parameters \( dt \) and minimizing the sum of squares of the residuals \( dx \):

\[
dx^Tdx \rightarrow \text{Min}
\]

(8)
From there it follows that the minimum conditions (3) and (8) are equivalent. This is shown in detail in proof 1 (see appendix). At the same time $Q$ is identified as the covariance matrix of the residuals.

**GENERALIZATION**

By introducing a weight matrix $P$ the minimum conditions (8) and (3) can be generalized to:

$$d\bar{x}^T P d\bar{x} \rightarrow \text{Min}$$

$$\text{tr} Q P \rightarrow \text{Min} \quad (9)$$

The conditions (9) and (10) are equivalent again (see proof 2, appendix). Analogue to the more general minimum condition (10) equations (4) to (7) are generalized to:

$$dt = (G^T P G)^{-1} G^T P dx$$

$$R = (G^T P G)^{-1} G^T P M (I-P G (G^T P G)^{-1} G^T) \quad (11)$$

The conditions (9) and (10) are equivalent again (see proof 2, appendix). Analogue to (15) the covariance matrix $M$ of the random vector $dx$ can be represented as:

$$M = \begin{bmatrix} Q & U \end{bmatrix} \begin{bmatrix} I & G \\ U^T & R \end{bmatrix} \begin{bmatrix} I \\ G^T \end{bmatrix} = Q + GU^T + UG^T + GRG^T \quad (16)$$

The submatrix $U$ of the common covariance matrix of the components $d\bar{x}$ and $dt$ can be obtained by applying the law of error propagation to equations (13) and (11):

$$U = (I-G (G^T P G)^{-1} G^T P M P G (G^T P G)^{-1})$$

The existence of $U$ demonstrates that $d\bar{x}$ and $dt$ are correlated with each other.

**A THEORY FOR ANALYSIS OF COVARIANCE MATRICES**

In a previous paper a theory was presented, which proceeds from the concept of inner accuracy and allows for a rigorous analysis of a given covariance matrix [7]. In the present paper this theory is derived slightly different and the whole problem is treated more comprehensively. By analysis we understand a rigorous disintegration of $M$ into a covariance matrix $K$ with a structure as simple as possible and the effect of a set of filter parameters, represented by a coefficient matrix $G$.

**DERIVATION OF THE THEORY**

We search for a random vector $dx_K$ with the associated covariance matrix $K$ and for a filter matrix $G$ which allow for a rigorous separation of the given random vector $dx$ according to:

$$dx = dx_K + G \Delta t \quad (18)$$

At the same time we dispose of the weight matrix $P$, which was arbitrary up to now and set:

$$P \triangleq K^{-1} \quad (19)$$

The vector $dx_K$ and the covariance matrix $K$ can be represented analogue to equations (15) and (16).
\[ \text{Equation (24) is equivalent with the condition:} \]
\[ Q_K = 0 \]  
\[ \text{(25)} \]
\[ \text{Using equations (22) and (23) we convert the condition (25) into the more practicable form:} \]
\[ K = Q + GSG^T \]
\[ \text{(26)} \]
\[ \text{Combining equation (16) and (21) and considering condition (25) the covariance matrix } M \text{ now can be disintegrated analogue to the separation of } dx \text{ in (18):} \]
\[ M = K + [I \ G] [Q \ U] [R - S] [G^T] = K + GUT + UGT + GTG^T \]
\[ \text{with} \]
\[ T = R - S \]
\[ \text{(27)} \]
\[ \text{R and S are positiv semidefinit matrices, but the difference matrix } T \text{ is not necessarily positiv semidefinit. Equation (27) represents the aspired analysis of the covariance matrix } M. \text{ The meaning is that } M \text{ can be expressed rigorously by the covariance matrix } K \text{ and the effect of filter parameters with the coefficient matrix } G. \text{ As can be shown condition (26) is necessary and adequate for the validity of equation (27) (see proof 4, appendix). Therefore (26) can be used as a proper criterion regarding the choice of } K \text{ and } G. \]

**CRITERION I**

The given covariance matrix } M \text{ can be analysed rigorously according to equation (27) if and only if the chosen matrices } K \text{ and } G \text{ fulfil condition (26). The performance of the analysis can be simplified essentially by replacing the choice of } K \text{ by the choice of the weight matrix } P \text{ which is related to } K \text{ according to equation (19). Then the covariance matrix } K \text{ needed in criterion I is estimated as follows:} \]
\[ E \left[ \bar{\sigma}_0^2 \right] = E \left[ dx^T P dx \right] / r = trQP / r \]
\[ K = E \left[ \bar{\sigma}_0^2 \right] P^{-1} \]
\[ \text{(28)} \]
\[ \text{Equation (28) is the expectation of the variance factor being computed from the residuals } dx. \text{ The redundancy } r \text{ is determined by the number } n \text{ of random variables minus the number } u \text{ of filter parameters. The proof of equation (28) and a discussion of equation (29) is given in the appendix (proof 5).} \]
As soon as criterion I is fulfilled the question appears whether all filter parameters effect the analysis or whether some of them can be omitted without effecting criterion I. Therefore we look for a criterion which detects filter parameters without influence over the analysis. For that purpose we split the vector $\Delta t$ into the components $\Delta t_1$ and $\Delta t_2$ and represent equation (27) accordingly as:

$$
M = K + [G_1 G_2] + [R_{11} R_{12} R_{12} R_{22}] - [S_{11} S_{12} S_{12} S_{22}]
$$

$$
= K + G_{11} T_{11} G_1 + G_{12} T_{12} G_2 + G_{22} T_{22} G_2
$$

with

$$
T_{11} = R_{11} - S_{11}
$$
$$
T_{12} = R_{12} - S_{12}
$$
$$
T_{22} = R_{22} - S_{22}
$$

The aspired criterion can be formulated as follows:

CRITERION II

The analysis (30) is not effected by the filter parameters $\Delta t_2$ and can be represented without putting up $G_2$ if and only if the following equations (31) are valid

$$
T_{22} = 0
$$
$$
U_2 = 0
$$

(see proof 6, appendix).

SPECIAL CASES OF THE ANALYSIS

Equation (27) represents the general case of an analysis of the given covariance matrix $M$. Beside this various special cases of the analysis are possible. Two of them, being of particular interest and appearing frequently shall be treated in detail.

SPECIAL CASE A : $U_2 = 0$, $T_{22} = \text{positiv semidefinit}$

With that equation (27) is simplified considerably to:

$$
M = K + G T
$$

(27a)

Considering (27a) together with equation (18) we see that here $T$ is the covariance matrix of the filter parameters $\Delta t$. Moreover from (27a) it follows that $d x_K$ and $\Delta t$ are not correlated with each other. In this case equation (18) can be interpreted as a separation of the random vector $d x$ into the independent components of noise and signal, being used in collocation [8].

SPECIAL CASE B : $U_2 = 0$, $T_{22} = \text{positiv semidefinit}$

With that equation (30) is simplified to:

$$
M = K + G_1 U_1 + G_2 U_2 + G_1 T_1 G_1 + G_2 T_2 G_2
$$

(30a)

Equation (30a) can be split properly into

$$
M = K + G_1 U_1 + G_1 T_1 G_1 + G_2 T_2 G_2
$$

(30b)

$$
M = K + G_2 T_2 G_2
$$

(30c)
The covariance matrix $\overline{M}$ differs from $M$ due to the effect of the filter parameters $\Delta t_1$ only (see proof 7, appendix). If the filter parameters $\Delta t_1$ are of no particular interest, equation (30a) therefore can be replaced by the much simpler disintegration (30c). Considering equation (30c) we see that here $\Sigma_2$ is the covariance matrix of the filter parameters $\Delta t_2$. Moreover $\Delta t_2$ and $\delta x\bar{K}$ are not correlated with each other.

**PERFORMANCE OF THE ANALYSIS**

The following block diagram shows the steps of the analysis and their sequence. The analysis starts with a proper choice of the weight matrix $P$ and the filter matrix $G$, representing the stochastic model and the functional model of the analysis. The necessity to assume a proper mathematical model a priori we know from least squares adjustment and regression analysis respectively. Of great importance in this context is the fact that the suitability of $P$ and $G$ can be checked rigorously by criterion I.

Equation (26) which is used in criterion I is identical with the basic equation of a posteriori variance and covariance estimation, given in [9]. Therefore the corresponding procedures can be used successfully to estimate $K$. Most of the time it will be sufficient to assume uncorrelated random variables $\delta x\bar{K}$ and to estimate their weights only. Concerning the choice of the filter matrix $G$ use can be made of the fact that filter parameters without influence over the analysis are detected by criterion II. Therefore it can be recommended to start the analysis with putting up relatively many filter parameters. Of course they have to be linear independent. As filter parameters often the coefficients of regression polynomials will be used.

With the practical application frequently it will not be possible to fulfill criterion I rigorously. In this case, suitable statistical test procedures have to be applied to decide whether criterion I is fulfilled or not. If equation (26) being used in criterion I isn't valid exactly we must not apply criterion II rigorously. The question whether some of the filter parameters don't effect the analysis in this case again has to be answered by applying suitable statistical tests.

**COMMENT**

The analysis of a covariance matrix $M$ according to equation (27) has to be discriminated from a decomposition of a covariance matrix $\Sigma$ by factor analysis [10]. This method of multivariate analysis is characterized by:

$$\Sigma = \Lambda \Lambda^T + D$$  \hspace{1cm} (32)

$\Lambda$ is the factor loading matrix. The number of columns of $\Lambda$ is fixed usually. $D$ is a diagonal matrix. Equation (32) is less general than (27) because no covariance term corresponding with $U$ is existing in (32). As opposed to the analysis (27) where $P$ and $G$ are chosen a priori and improved if necessary, in case of factor analysis $\Lambda$ and $D$ are estimated directly. The corresponding estimation procedures are relatively complicated and depend on the assumption of normal distributed variables. The estimation of $\Lambda$ and $D$ is followed by an interpretation of the factor loadings of $\Lambda$. With the analysis treated here this step is avoided completely because the meaning of the filter parameters is given a priori.
APPLICATIONS OF THE THEORY

The concept presented in this paper is a suitable tool to analyse any given covariance matrix, obtained theoretically or empirically. This shall be demonstrated by the following analysis of the theoretical covariance matrix of the \( z \) coordinates of a photogrammetric model.

For that purpose we suppose vertical wide angle photography. The base length we assume as \( b = 1 \) and the flying height as \( h = 153/92 \). The 8 model points have the same heights and are distributed regularly (see figure 2). Points 3 and 5 are control points in planimetry and height, point 2 is an additional height control point (free adjustment).

The image coordinates we assume as uncorrelated observations of variance 1. Putting up a rigorous least squares adjustment according to the bundle method we obtain the covariance matrix \( M \) of the 8 model heights as a submatrix of the complete inverse of the normal equation matrix:

\[
M = \begin{bmatrix}
8.30 & * & * & * & 1.38 & 1.38 \\
* & * & * & * & * & * \\
* & * & * & * & * & * \\
* & * & 17.30 & -0.70 & 5.02 & 0.52 \\
* & * & * & * & * & * \\
1.38 & * & 5.02 & 0.52 & 8.86 & 1.51 \\
1.38 & * & 0.52 & * & 5.02 & 1.51 & 8.86 \\
\end{bmatrix}
\]

The variances and covariances belonging to the height control points 2, 3 and 5 are zero of course. Due to the existing symmetry the heights 4 and 6 as well as 7 and 8 are of equal accuracy.

The analysis of the covariance matrix \( M \) we start assuming \( P = I \) for the weight matrix and putting up 6 filter parameters according to a regression polynomial of degree 2 in the model coordinates \( x \) and \( y \). With that we obtain the following filter matrix \( G \):

\[
G = \begin{bmatrix}
1 & x_1 & y_1 & x_1^2 & x_1 y_1 & y_1^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_8 & y_8 & x_8^2 & x_8 y_8 & y_8^2 \\
\end{bmatrix}
\]

The first three filter parameters allow for a shift in \( z \) and for tilts in \( x \) and \( y \) direction. These parameters are needed for levelling the model. The other three filter parameters are put up arbitrary. Performing the analysis we obtain:

\[
E [c_0^2] = 5.53 \\
K = 5.53 I
\]

This covariance matrix \( K \) and the chosen filter matrix \( G \) fulfil criterion I. That means that \( K \) and \( G \) allow for a rigorous disintegration of \( M \) according to equation (27). The matrices \( U \) and \( T \) follow as:
Because criterion I is fulfilled criterion II can be applied rigorously. Doing this we see that the last filter parameter $\Delta t_6$ going with $y^2$ has no influence over the analysis of $M (T_{66} = 0, U_6 = 0)$.

If we collect the first three levelling parameters in the vector $\Delta t_1$ and the last three in the vector $\Delta t_2$ and if we divide the matrices $G$, $U$ and $T$ correspondingly we see that the premises of special case B are given:

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix},$$

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix},$$

$$M = K + G T_{22} G^T.$$

$M$ differs from $\bar{M}$ due to the effect of the levelling parameters $\Delta t_1$ only. We are allowed to replace $M$ by $\bar{M}$ because $M$ is arbitrary with respect to these three parameters due to the arbitrary choice of the three height control points 2, 3 and 5. If we fix these three heights, we obtain different results for $U_1$, $T_{11}$ and $T_{12}$ but we get $U_2 = 0$ again and $T_{22}$ remains unchanged.

From these facts it follows that $\bar{M}$ can be represented by $K = 5.53 I$ and the effect of two filter parameters, going with $x^2$ and $xy$ respectively. These two parameters are uncorrelated with each other and their variances are 5.53 and 3.47 respectively.

The results of this analysis can be interpreted as follows: The covariance matrix $K$ describes the accuracy of the model heights without the effect of the orientation parameters of the bundle adjustment. This accuracy is obtained keeping the orientation parameters of both photos fixed. Then all model heights get the same accuracy and are not correlated with each other. The variance in $z$ is $2.153^2/92^2 = 5.53$.

Among the orientation parameters of the images 1 and 2 the only one of interest here are those which lead to model deformations in $z$ being not compensated by the filter parameters $\Delta t_1$ of model levelling. These orientation parameters are $\phi_1$ or $\phi_2$ causing a cylinder shaped deformation in $z$ and $\omega_1$ or $\omega_2$ causing a twisted model. The filter parameters going with $x^2$ and $xy$ are able to compensate those deformations rigorously. The filter parameter going with $y^2$ is not needed at all and gets a variance of zero accordingly.
ACKNOWLEDGMENT

The author is grateful to Prof. Meissl, Graz, to Prof. Grafarend, Bonn and to Mr. Pope, Rockville for fruitful and encouraging discussions.

REFERENCES

Proof 1
Replacing (8) by the expectation $E\left[dx^Tdx\right]$ and considering
$E\left[dx\right] = 0$ we get:
$E\left[dx^Tdx\right] = E\left[tr(dx^Tdx)\right] = E\left[tr(dx^tdx^T)\right] = tr(E\left[dx^tdx^T\right]) = trQ$

Proof 2
Analogous to proof 1 we obtain:
$E\left[dx^TPdx\right] = E\left[tr(dx^TPdx)\right] = E\left[tr(dx^tdx^TP)\right] = tr(E\left[dx^tdx^TP\right]) = trQ_P$

Proof 3
From (18) follows (24):
$dx_K = (I-G(G^TPG)^{-1}G^TP)dx_K = (I-G(G^TPG)^{-1}G^TP)(dx-G\Delta t) = dx$
From (24) follows (18):
$dx = dx + Gdt = dx_K + Gdt = dx_K + G(dt-dt_K) = dx_K + G\Delta t$

Proof 4
(26) is identical with (25). From (25) follows (27) directly.
From (27) follows (26):
$K = M-GU^T - UG^T - GRG^T + GSG^T = Q + GSG^T$

Proof 5
(28) follows from proof 2:
$E\left[\sigma_0^2\right] = E\left[dx^T Pdx\right] / r = trQ_P / r$

(29) is an approximate estimation of $K$. The rigorous relation
between $K$ and $P$ should be:
$K = E\left[\sigma_0^2\right]_K P^{-1}$
with $E\left[\sigma_0^2\right]_K$ being computed analogous to above as:
$E\left[\sigma_0^2\right]_K = E\left[dx^T K P dx_K\right] / r = trQ_K P / r$

A rigorous estimation of $K$ using $E\left[\sigma_0^2\right]_K$ is impossible because
$K$ itself is needed for the determination of $E\left[\sigma_0^2\right]_K$. Therefore
$E\left[\sigma_0^2\right]$, being determinable replaces $E\left[\sigma_0^2\right]_K$ in (29). The better
the choice of $P$ and $G$ the closer is $E\left[\sigma_0^2\right]$ to $E\left[\sigma_0^2\right]_K$. As soon
as $K$, computed from (29) fulfills criterion 1 we obtain:
$E\left[\sigma_0^2\right] = trQ_P / r = trQ_K P / r = E\left[\sigma_0^2\right]_K$
The rigorous validity of 
\[ K = E \left[ a^2 \right] K P^{-1} \]
can be shown starting from \( K = cP^{-1} \) and proving \( c = E \left[ a^2 \right] K \)

\[ E \left[ a^2 \right] K = trQKP/r = tr(KP-G(G^T K^{-1} G)^{-1} G^T P)/r \]

\[ = (trKP-tr(G^T K^{-1} G)^{-1} G^T P)/r = c(n-u)/r = c \]

Proof 6

From (30) and (31) follows \((Q-QK)_1 = 0\), which is necessary and adequate for the validity of (30) using \( G_1 \) only

\[ (Q-QK)_1 = (I-G_1(G_1^T P G_1)^{-1} G_1^T P)(M-K)(I-PG_1(G_1^T P G_1)^{-1} G_1^T P) \]

\[ = (I-G_1(G_1^T P G_1)^{-1} G_1^T P)(G_1 U_1 + U_1 G_1 + G_1 T_1 + G_1 T_2 G_2^T + G_2 T_2 G_1^T) \]

\[ (I-PG_1(G_1^T P G_1)^{-1} G_1^T P) = 0 \]

For proving that (31) follows from (30) and \((Q-QK)_1 = 0\) we separate \((G^T P G)^{-1}\) into:

\[ (G^T P G)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \]

\( U_2 \) and \( T_{22} \), appearing in (30) we represent explicitely as:

\[ U_2 = (I-G(G^T P G)^{-1} G^T P)(M-K)P(G_1 A_{12} + G_2 A_{22}) \]

\[ = (I-G_1 A_{11} G^T P - G_1 A_{12} G^T P - G_2 A_{11} G^T P - G_2 A_{12} G^T P)(M-K)(P G_1 A_{12} + P G_2 A_{22}) \]

\[ T_{22} = (A_{12} G_1 P + A_{22} G_2 P)(M-K)(P G_1 A_{12} + P G_2 A_{22}) \]

In the following proofs we consider:

\[ \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} G_1^T P G_1 & G_1^T P G_2 \\ G_2^T P G_1 & G_2^T P G_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \]

From 0 = \((Q-QK)_1\) follows \( U_2 = 0 \):

\[ 0 = (I-G_1 A_{11} G^T P - G_2 A_{12} G^T P - G_1 A_{12} G^T P - G_2 A_{22} G^T P) \]

\[ (Q-QK)_1(P G_1 A_{12} + P G_2 A_{22}) \]

\[ = (I-G_1 A_{11} G^T P - G_2 A_{12} G^T P - G_1 A_{12} G^T P - G_2 A_{22} G^T P)(I-G_1(G_1^T P G_1)^{-1} G_1^T P) \]

\[ (M-K)(P G_1 A_{12} + P G_2 A_{22}) \]

\[ = (I-G_1 A_{11} G^T P - G_2 A_{12} G^T P - G_1 A_{12} G^T P - G_2 A_{22} G^T P)(M-K) \]

\[ (P G_1 A_{12} + P G_2 A_{22}) = U_2 \]
From $0 = (Q - Q_K)_1$ follows $T_{22} = 0$:

\[
0 = (A_{12}^T G_{11}^T P + A_{22}^T G_{22}^T P)(Q - Q_K)_1 (P G_{12} + P G_{22})
\]

\[
= (A_{12}^T G_{11}^T P + A_{22}^T G_{22}^T P)(I - G_{11}^T (G_{11}^T P G_{11}^T)^{-1} G_{11}^T P) (M - K) (I - P G_{11}^T (G_{11}^T P G_{11}^T)^{-1} G_{11}^T P) (P G_{12} + P G_{22})
\]

\[
= (A_{12}^T G_{11}^T P + A_{22}^T G_{22}^T P) (M - K) (P G_{12} + P G_{22}) = T_{22}
\]

**Proof 7**

We put up $G_1$ only and prove that $M$ and $\overline{M}$ lead to the same covariance matrix of residuals:

\[
(Q - Q)_1 = (I - G_1^T (G_{11}^T P G_{11}^T)^{-1} G_{11}^T P) (M - \overline{M}) (I - P G_1^T (G_{11}^T P G_{11}^T)^{-1} G_{11}^T P)
\]

\[
= (I - G_1^T (G_{11}^T P G_{11}^T)^{-1} G_{11}^T P) (G_{11} U_{11}^T + U_{11} G_{11}^T + G_{11} T_{11} G_{11}^T + G_{11} T_{12} G_{12}^T + G_{22}^T G_{11}^T)
\]

\[
(I - P G_1^T (G_{11}^T P G_{11}^T)^{-1} G_{11}^T P) = 0
\]

From $(Q - Q)_1 = 0$ follows that $M$ and $\overline{M}$ differ due to the effect of $\Delta t_1$ only.