DELFT UNIVERSITY OF TECHNOLOGY
REPORTS OF THE FACULTY OF GEODESY
MATHEMATICAL AND PHYSICAL GEODESY

88.2

PERFORMANCE ANALYSIS OF KALMAN FILTERS

by

P.J.G. Teunissen
M.A. Salzmann

Delft, The Netherlands
September 1988

PERFORMANCE ANALYSIS OF KALMAN FILTERS

Abstract

Methods for the analysis of the performance of Kalman filters are considered in the paper. The methods are all based on the innovation sequence which has well defined statistical properties if the filter is optimal. Local and global teststatistics are presented and discussed. A global slippage teststatistic is introduced.This test, which has batch type properties, is given in a recursive form.

1. Introduction

With the advent of powerful microcomputers sophisticated algorithms for kinematic positioning are used more frequently. It is common practice to process data from different sensors simultaneously in a so-called integrated navigation system to obtain a best estimate of position. The algorithm most often used in these integrated navigation systems is the well known Kalman filter. Some typical examples of the application of Kalman filters in navigation systems is given in [1],[2],[3],[4], and [5].
In this paper we restrict ourselves to the performance analysis of Kalman filters. To obtain useful (positioning) results using an integrated navigation system it is important that the performance of the underlying filter is at an optimum. Methods for the detection of departures from optimality are all based on the so-called innovation sequence. The innovation sequence of an optimal filter has precisely defined characteristics which can be compared with the output of an actually implemented Kalman filter. Under normal conditions the innovation sequence is a zero mean Gaussian white noise sequence with known covariance. Independence can be tested with the so-called run test or reverse arrangements test, or with tests based on the autocorrelation function. These tests are all non parametric or distribution free. Tests for the covariance are based on the Wishart distribution. See e.g. [6], [7], and [8].
In this paper we focus on slippage tests, which basically test the zero mean of the innovation sequence. We present the local overall model test and the one-dimensional local slippage test. These tests are very easily implemented and we belief that every software package should at least have these two types of tests included. We also present the global overall model test statistic which is a simple weighted mean of the local overall model teststatistic. Finally we introduce a new powerful teststatistic, which is called the one-dimensional global slippage teststatistic. It is a teststatistic given in recursive form which has batch type properties.

The contents of the paper is as follows. The Kalman filter and the assumptions underlying its model are briefly described in the next section. In section 3 the innovation sequence is introduced and its characteristics are outlined. The local and global teststatistics are derived in sections 4 and 5 respectively. Section 6 contains some conclusions. Two appendices are attached to the paper. Appendix A contains some

necessary theory on adjustment and hypothesis testing. Appendix B gives a simple derivation of the Kalman filter using the principles of least squares.

2. The Linear Kalman Filter

In this section we present and briefly discuss the mathematical model and recursive relations of the linear discrete time Kalman filter. For a more extensive discussion the reader is referred to the literature, see e.g. [9] , [10]. The mathematical model which forms the basis of the Kalman filter is

$$(1a)\quad \underline{x}_k = \Phi_{k,k-1}\underline{x}_{k-1} + \underline{q}_k$$

$$(1b)\quad \underline{y}_k = A_k\underline{x}_k + \underline{e}_k$$

Equation (1a) represents the **dynamic model** and equation (1b) represents the **measurement model.** The dynamic model is a linear vector difference equation. The independent variable t, which is often time, can assume the values $t_0 \le t_1 \le \ldots \le t_N$, where the t_i are not necessarily equidistant. The state of the system at t_k is given by the n-dimensional vector $\underline{x}_k$. The underscore indicates that the state vector $\underline{x}_k$ is a vector random variable.
In many applications (1a) is derived from the linearization or linear perturbation equations relating to a dynamic system, so that the nxn matrix $\Phi_{k,k-1}$ may be assumed to be a known state transition matrix with the following properties:

$$\Phi_{k,k} = I \quad \text{for all } k$$
$$\Phi_{k,l}\Phi_{l,m} = \Phi_{k,m}\,.$$

The initial state $\underline{x}_0$ is considered to be a vector random variable with a Gaussian distribution and the known statistics

$$E\{\underline{x}_0\} = x_0$$
$$E\{(\underline{x}_0-x_0)(\underline{x}_0-x_0)^*\} = P_{0|0}$$

The operator E{.} denotes the mathematical expectation and $(.)^*$ denotes transpose.
The n-dimensional vector random variable $\underline{q}_k$ represents the dynamic system noise. It is assumed to have a Gaussian distribution with the known statistics

$$E\{\underline{q}_k\} = 0 \quad \text{for all } k$$
$$E\{\underline{q}_k\underline{q}_l^*\} = Q_k\delta_{kl}\,,$$

where δ_{kl} is the Kronecker delta, i.e. $\delta_{kl}=1$ for k=l and $\delta_{kl}=0$ otherwise.
The measurement model (1b) states that at each time t_k there are m_k measurements collected in the m_k-dimensional observation vector $\underline{y}_k$ available. The vector of observations is linearly related to the state through the known $m_k$xn designmatrix A_k and is corrupted by additive measurement noise $\underline{e}_k$.
The vector random measurement noise $\underline{e}_k$ is assumed to have a Gaussian distribution with the known statistics

$$E\{\underline{\varepsilon}_k\} = 0 \quad \text{for all } k$$
$$E\{\underline{\varepsilon}_k \underline{\varepsilon}_l^*\} = R_k \delta_{kl} ,$$

The above defined covariance matrices $P_{0|0}$, Q_k, and R_k are all assumed to be positive definite and thus invertible.
Finally it is assumed that the random sequences $\underline{q}_k$ and $\underline{\varepsilon}_k$ are mutually uncorrelated and also uncorrelated with the initial state:

$$E\{\underline{q}_k \underline{\varepsilon}_l^*\} = 0 \quad \text{for all } k,l$$
$$E\{\underline{q}_k \underline{x}_0^*\} = 0 \quad \text{for all } k$$
$$E\{\underline{\varepsilon}_k \underline{x}_0^*\} = 0 \quad \text{for all } k$$

The mathematical model described above provides the basis for all succeeding discussion.

Depending on the application one has in mind, one might wish to obtain an estimate of the state at a certain time k which depends on all observations taken up to and including time k+l. If l<0 the process is called **prediction**. The state estimate depends then only on the observations taken prior to the desired time of estimation. If l=0 the process is called **filtering**. In this case the state estimate depends on all the observations taken prior to and at time k. Finally if l>0 the process is called **smoothing**. The state estimate depends then on observations taken prior to, on, and after time k. (see also fig. 1). It will be clear that under normal conditions smoothed estimates are more precise and reliable than filtered estimates. Similarly filtered estimates are more precise and reliable than predicted estimates.

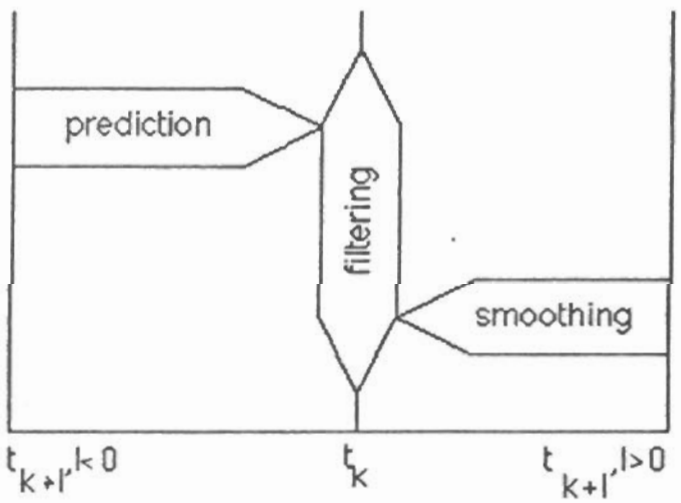

Fig. 1. Relation between prediction, filtering, and smoothing of the state at time k

Since we have real time applications of the estimation problem in mind, we shall restrict ourselves in the following to recursive prediction and filtering. The problem we stand for is to determine an estimate of the state at time k that is a linear combination of an estimate of the state at time k-1 and the observations at time k. Furthermore the estimate must be "best" in a certain sense. Kalman [11] was the first to solve this problem for the continuous time model using the minimum mean square error criterion. When Kalman's

method of derivation is applied to the discrete time model the so-called linear discrete time Kalman filter is obtained. It basically consists of two parts: the **time update** which gives the predicted state, and the **measurement update** which gives the filtered state.
The time update of the state and its covariance matrix are given as

$$(2a)\quad \hat{x}_{k|k-1} = \Phi_{k,k-1}\hat{x}_{k-1|k-1}$$

$$(2b)\quad P_{k|k-1} = \Phi_{k,k-1}P_{k-1|k-1}\Phi^{*}_{k,k-1} + Q_k \quad ,$$

and the measurement updates of the state and its covariance matrix are given as

$$(3a)\quad \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - A_k\hat{x}_{k|k-1})$$

$$(3b)\quad P_{k|k} = (I - K_kA_k)P_{k|k-1}$$

where

$$(4)\quad K_k = P_{k|k-1}A^{*}_k(A_kP_{k|k-1}A^{*}_k + R_k)^{-1}$$

is the so-called Kalman gain matrix.
Equation (2a) gives the best estimate of the state at time k in the minimum mean square error sense using all observations prior to time k, whereas equation (3a) gives the best estimate of the state using both $\hat{x}_{k|k-1}$ and y_k.
Although Kalman used the minimum mean square error principle to derive his equations, it can be shown that if the model is linear and all vector random variables are Gaussian the methods of maximum likelihood, maximum a posteriori and least squares lead to identical results. Since surveyors and hydrographers are probably the most familiar with the principle of least squares, a simple derivation of the above Kalman filter equations based on this principle is given in appendix B.

3. The Innovation Sequence

As was pointed out in the previous section the recursive Kalman filter produces optimal estimates of the state vector with well defined statistical properties. These estimates are only optimal, however, as long as the given assumptions underlying the mathematical model hold. Misspecifications in the dynamic model and/or the measurement model will invalidate the results of estimation and thus also any conclusion based on them. It is therefore of crucial importance to have ways to verify the validity of the assumed mathematical model.
An important role in the process of model testing is played by the so-called **innovation sequence**. The innovation sequence is defined as the difference between the actual system output and the predicted output based on the predicted state. Thus the innovation sequence is given by

$$(5)\quad \underline{v}_k = \underline{y}_k - A_k\underline{\hat{x}}_{k|k-1} \quad , k=1,2, \ldots.$$

It is called the innovation sequence since it represents the new information brought in by the latest observation vector. This can be seen from the measurement update equation

(3a), which shows that the filtered state is a linear combination of the predicted state and the innovation.
Under normal conditions the innovation is "small" and corresponds to random fluctuations in the output since all the systematic trends are eliminated by the model. If, however, the model is misspecified the innovation is "large" and contains systematic trends because the model no longer represents the physical system adequately.
Under normal conditions the innovation sequence has well defined statistical properties. It can be shown, see e.g. [12], that if the model is valid, the innovation sequence is a zero mean Gaussian white noise sequence with known covariance:

(6a) $E\{\underline{v}_k\} = 0$ for all k

(6b) $E\{\underline{v}_k \underline{v}_l^*\} = Q_{vk}\delta_{kl}$,

where

(7) $Q_{vk} = R_k + A_k P_{k|k-1} A_k^*$

is the covariance matrix of the innovation $\underline{v}_k$.
These properties can be used to test the innovation sequence for zero mean, whiteness and a given covariance. A sequence is called white if it consists of a sequence of uncorrelated random variables.
In the following we will restrict ourselves to misspecifications in the mean of the innovation sequence. That is, we will only consider slippage tests. The necessary teststatistics are all functions of the innovations and are optimized to detect a particular misspecification in the assumed mathematical model.

4. Local Slippage Tests

In this section we present some methods of hypothesis testing as applied to the linear discrete time Kalman filter. For a brief review of the theory of hypothesis testing the reader is referred to appendix A.
We consider the **local overall model (LOM) test** and the one-dimensional **local slippage (LS) test.** By local we mean that the tests when performed at time k only depend on the predicted state at time k and the observations at time k. Observations taken after time k have no effect on these tests. The influence of observations taken prior to time k is only felt indirectly via the predicted state.
The purpose of the LOM test is to detect misspecifications in the mathematical model occurring at time k. Since misspecifications in the model which have occurred prior to time k do affect the predicted state, the LOM test could detect them. The LOM test is however not optimal for these past misspecifications. This case will be considered in the next section.
The teststatistic of the LOM test is given as

(8) $$\underline{T}_k = \frac{\underline{v}_k^* Q_{v_k}^{-1} \underline{v}_k}{m_k} ,$$

where m_k is the number of observations taken at time k.
The decision that a misspecification in the model has occurred at time k is made once

$$T_k \geq \chi^2_\alpha(m_k,0) ,$$

where $\chi^2_\alpha(m_k,0)$ is the upper α probability point of the Chi-squared distribution $\chi^2(m_k,0)$ with m_k degrees of freedom.
After a misspecification has been detected with the above test, one can try to diagnose the type of misspecification. This is done with the one dimensional LS test. The corresponding teststatistic is given as

$$(9)\quad \underline{w}_k = \frac{c^* Q_{v_k}^{-1} \underline{v}_k}{(c^* Q_{v_k}^{-1} c)^{1/2}} .$$

The choice of the c-vector depends on the type of misspecification one wants to test. With (9) one can try to detect a possible slippage in the mean of the predicted state, or in the mean of the measurement noise, or in both. For instance, if one suspects sensor failures or outlying observations one can follow the **datasnooping** approach by choosing m_k c-vectors of the form

$$(10)\quad \underset{m_k x 1}{c_i} = (0,.....,0,\ \underset{1}{1},\underset{(i-1),i,(i-1)}{0,.....1})^* \quad \text{for } i=1,....,m_k \ .$$

The observation for which the w-teststatistic is a maximum can then be considered as the most likely outlying observation.
The **detectability** of the misspecifications is very much governed by the structure of the inverse of the covariance matrix of the vector of innovations. For instance, sensor failures are better detectable the larger the diagonal elements of Q_{vk}^{-1} are. One is better able to discriminate between different sensor failures the smaller the off-diagonal elements of Q_{vk}^{-1} are. When designing a Kalman filter one should in case of datasnooping thus aim at a diagonal dominant matrix Q_{vk}^{-1}.

5. Global Slippage Tests in Recursive Form

In the preceding section we considered local slippage tests. There it was shown that the teststatistic $\underline{T}_k$ and $\underline{w}_k$ can be computed once the innovation vector $\underline{v}_k$ and its covariance matrix Q_{vk} are available. The local tests can therefore be executed in real time parallel with the computations of the Kalman filter update equations. This has the advantage that corrective action can be taken in real time.
A disadvantage is, however, that the tests discussed above are local and therefore may not be able to detect global unmodelled trends. In particular if the local redundancy m_k is small or if the matrix Q_{vk}^{-1} is not diagonal dominant enough, various misspecifications may pass unnoticed.
One possible remedy is of course to keep all the collected data stored in the computer's memory. With a batch type of algorithm one can then process all the data for the detection and localization of model misspecifications. It will be clear, since smoothing is involved, that the resulting tests are more powerful than the local tests discussed above. There are however two disadvantages to such an approach. Firstly the property of recursiveness is lost due to the batch type of processing. Secondly no real time corrective action can be taken once a misspecification is detected. This, however, may not be as serious as it sounds. In practice a delay may be acceptable if it is small enough. That is, in practice it may be more important to detect a misspecification with a possible delay than not to detect

it all! In any case, the delay is the price one has to pay for the inclusion of smoothing. With these remarks in mind it thus seems worthwhile to look for teststatistics in recursive form which have batch type properties.
First we consider the **global overall model (GOM) test.** The GOM teststatistic is given as

$$(11)\quad \underline{T}^k = \frac{\sum_{i=1}^{k} \underline{v}_i^{*} Q_{v_i}^{-1} \underline{v}_i}{\sum_{i=1}^{k} m_i} = \frac{\sum_{i=1}^{k} m_i \underline{T}_i}{\sum_{i=1}^{k} m_i} .$$

It is the weighted mean of the LOM teststatistics $\underline{T}_i$ and is therefore very easily implemented in a computer program. One way to derive the GOM teststatistic is to consider a set up of batch type for the Kalman filter. Equation (11) follows, however, immediately once one recognizes that the LOM teststatistics are mutually independent because of the Gaussian white noise property of the innovation sequence.
The decision that a misspecification has occurred is made once

$$T^k \geq \chi^2_\alpha(\sum_{i=1}^{k} m_i, 0) ,$$

where $\chi^2_\alpha(\Sigma_i m_i,0)$ is the upper α probability point of the Chi-squared distribution $\chi^2(\Sigma_i m_i,0)$ with $\Sigma_i m_i$ degrees of freedom (i=1,..,k). If the GOM test comes to rejection misspecifications at times $1 \leq l \leq k$ may have occurred.
The type of misspecification that has occurred is diagnosed by the one dimensional **global slippage (GS) test.** Let us restrict ourselves to the case of a sensor failure that has occurred at time l. Thus we consider a slippage in the mean of the observation vector of the form:

$$(12a)\quad E\{\underline{y}_k | H_a\} = E\{\underline{y}_k | H_0\} + c_{l,i}^k \nabla \quad ,$$

where

$$(12b)\quad c_{l,i}^k = \begin{cases} 0 \text{ for } k < l \\ c_i \text{ for } k \geq l \end{cases} ,$$

and the c_i-vector is of the form as given in equation (10) with m_k constant for all k, (i.e. m_k=m). Note that with (12) at each time k the number of alternative hypotheses considered is k times m.
The computation of the GS teststatistic is more complicated than that of the GOM teststatistic. Unlike the computation of the GOM teststatistic one cannot compute the GS teststatistic as the sum or weighted mean of the LS teststatistics $\underline{w}_k$. This is because the misspecification $c^k_{l,i}\nabla$ of (12) not only influences the innovation directly via the observation vector, but also indirectly via the predicted state. We therefore need to know how $c^k_{l,i}$ affects the innovation sequence.

The propagation of $c^k_{l,i}$ on the innovation sequence follows from the time and measurement update equations of the Kalman filter

$$(13)\quad \begin{cases} \bar{c}^k_{l,i} = c^k_{l,i} - A_k x_{k,l}, & x_{k,l}= 0 \text{ for } k \le l \\ x_{k+1,l} = \Phi_{k+1,k}(x_{k,l} + K_k \bar{c}^k_{l,i}) . \end{cases}$$

These are **recursive** equations which show how a sensor failure starting at time l influences the innovations at time k≥l. With (13) we are now able to write down the GS teststatistic as

$$(14)\quad \underline{w}^k_{l,i} = \frac{\sum_{j=1}^{k} \bar{c}^{j\bullet}_{l,i} Q_{v_j}^{-1} \bar{v}_j}{\left(\sum_{j=1}^{k} \bar{c}^{j\bullet}_{l,i} Q_{v_j}^{-1} \bar{c}^{j}_{l,i}\right)^{1/2}} .$$

Note that for k=l the teststatistic $\underline{w}^k_{l,i}$ reduces to that of the LS teststatistic because $\bar{c}^k_{l,i}=c^k_{l,i}=c_i$. As with the LS teststatistic the denominator of (14) is a measure of detectability. It should be large enough to have a proper detectability probability. In a similar way one can compute the scalars which measure the discriminability between the different alternative hypotheses.

In principle one has to compute the GS teststatistic for l=1,2, ... , k and the different c-vectors considered. This is shown in fig. 2a. The failing sensor i and the failure time l are determined by the maximum of the k times m values of $\underline{w}^k_{l,i}$.

Although the computations necessary to obtain the $w^k_{l,i}$ are still somewhat involved, they are much less than when done in a batch mode. This is mainly due to the recursive form of (13) which parallels the Kalman update equations. One can reduce the number of computations by introducing a moving window of length N and constraining l to k-N+1≤l≤k. In this way the teststatistic $\underline{w}^k_{l,i}$ need not be computed for l<k-N+1. This is shown in fig. 2b. A further reduction is achieved by constraining l to k-N+1≤ l ≤k-M. This is shown in fig. 2c. The parameters N and M have to be chosen in such a way that the information used in the $\underline{w}^k_{l,i}$'s is both not too little and not computationally excessive. The choice of the parameters N and M is thus primarily determined by the measures of detectability and discriminability. But they also depend on the time delay in detection one is willing to accept.

The above teststatistic (14) has been derived for the alternative hypotheses considered in (12). In a similar way the recursive teststatistic for e.g. outlying observations or slippages in the mean of the state vector can be derived. Also the cases where ∇ is known or where the common variancefactor of the covariance matrices is unknown can be derived analogously.

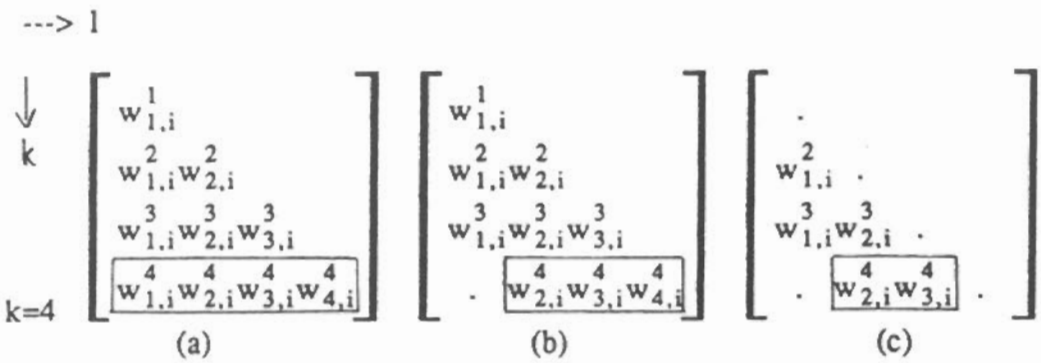

Fig. 2. The teststatics $\underline{w}^k_{l,i}$ with (a) no window; (b) a moving window with N=3, M=0; and (c) a moving window with N=3, M=1.

6. Conclusions

In this paper methods for the performance analysis of Kalman filters are discussed. We focussed our attention on methods for the detection of misspecifications in the means of observation vectors and state vectors. All teststatistics are functions of the innovations.
A distinction is made between local and global teststatistics. The local LOM and LS tests presented in section 4 can be implemented very easily. We recommend these tests to be available in every software product incorporating a Kalman filter.
In section 5 it is shown that the global GOM teststatistic is simply a weighted mean of the LOM teststatistics. As such it is very useful for detecting unmodelled global trends.
A possible misspecification detected with the GOM test can be diagnosed using the GS teststatistic. In spite of the fact that the GS has batch type properties, it can be computed in a recursive form, thereby facilitating its use for filter applications, if one is willing to accept a small time delay. Its implementation should be considered especially in those situations where detectability with the local slippage tests is poor due to a lack of local redundancy.
Furthermore we stressed the importance of the measures of detectability and discriminability for the design of a Kalman filter.

Acknowledgements

This paper was written while the authors were on leave at the departments of Surveying Engineering of the Universities of Calgary, Alberta and New Brunswick, Fredericton, Canada respectively. The hospitality received is gratefully acknowledged. The first author also gratefully acknowledges the funding received from the Netherlands Organization for Scientific Research (N.W.O). Financial assistance for the second author was also provided by a strategic grant entitled "Marine Geodesy Applications" from the Natural Sciences and Engineering Research Council of Canada, held by Dr. D.E. Wells.

7. References

1. Cox, D.B. (1980). Integration of GPS with Inertial Navigation Systems, Special Issue of *Navigation*. .The Institute of Navigation, Washington D.C., pp. 144-153.
2. Wong, R.V.C. (1982). A Kalman filter-smoother for an inertial survey system of the local level type. Publ. 20001, Div. of Surveying Engineering. Univ. of Calgary.
3. LaChapelle, G. (1983). Use of Kalman filtering for GPS aided marine navigation. Geodesy in Transition. Publ. 60002, Div. of Surveying Engineering. Univ. of Calgary.
4. Bakker, G., J.C. de Munck, and G.L. Strang van Hees (1985). Course on Radiopositioning. Lecture Notes. Delft University of Technology, Faculty of Geodesy, 14-17 January.
5. Napier, M.E., D. Parker (1986). Application of an inertial-based integrated positioning system for underwater surveys. *Proceedings of HYDRO86*, Special Publication No. 18, The Hydrographic Society. pp. 19.1-19.13.
6. Morrison, D.F. (1976). *Multivariate Statistical Methods*. 2nd. ed. McGraw-Hill, New York.
7. Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.
8. Bendat, J.S. and A.G. Piersol (1986). *Random Data: Analysis and Measurement Procedures*.2nd. edition. Wiley-Interscience, New York.
9. Gelb, A. (ed.) (1974). *Applied Optimal Estimation*. MIT Press, Cambridge MA.
10. Anderson, B.D.O and J.B. Moore (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ.
11. Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering*, Vol. 82, pp 34-45.
12. Kailath, T. (1968). An innovations approach to least squares estimation, part I: Linear filtering in additive white noise. *IEEE Trans. on Automatic Control*, Vol. AC-13(6), pp. 646-655.
13. Baarda, W. (1968). *A Testing Procedure for Use in Geodetic Networks*. Neth. Geod. Comm., Publ. on Geodesy, New Series, Vol. 2, No.5, Delft.
14. Teunissen, P.J.G. (1984). Quality control in geodetic networks. In: *Optimization and Design of Geodetic Networks*. Eds. E. Grafarend and F. Sanso. pp. 526-546.
15. Kok, J.J. (1984). *On Datasnooping and Multiple Outlier Testing*. NOAA Technical Report NOS NGS 30. U.S. Department of Commerce, Rockville, Md.
16. Pope, A.J. (1976). *The Statistics of Outliers and the Detection of Outliers*. NOAA Technical Report NOS 65 NGS 1.
17. Teunissen, P.J.G. (1986). Adjusting and testing with the models of the affine and similarity transformation, *Manuscripta Geodaetica*, Vol. 11, pp. 214-225.

Appendix A: Linear Least Squares and Slippage Tests

In this appendix a brief review is given of some adjustment and testing theory. For a more in depth discussion of these topics the reader is referred to [13], [14], and [15].

Linear Least Squares

Consider the linear model of observation equations

$$(A1)\quad \underset{mx1}{E\{\underline{y}\}} = \underset{mxn}{A}\ \underset{nx1}{x}\ , \quad \underset{mxm}{cov\{\underline{y}\}} = \sigma^2 Q_y$$

where $\underline{y}$ is the m-dimensional random vector of observational variates; E{.} denotes mathematical expectation; A is the designmatrix of order mxn; x is the n-dimensional vector of unknown parameters; cov{$\underline{y}$} is the covariance matrix of $\underline{y}$ and σ^2 is the variance factor of unit weight. The underscore in $\underline{y}$ indicates the random character of the observational variates. It will be assumed that $\underline{y}$ has a Gaussian distribution, i.e. $\underline{y}$ is normally distributed with mean Ax and covariance matrix $\sigma^2 Q_y$. The matrices A and Q_y are assumed to be of full rank, and m is assumed to be larger than n.

Application of the least squares criterium

$$\underset{x}{\text{minimize}}\ (y-Ax)^* Q_y^{-1} (y-Ax)$$

gives the linear least squares estimators of the parameters x and mean E{$\underline{y}$} respectively

$$(A2)\quad \begin{cases} \hat{\underline{x}} = (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \underline{y} \\ \hat{\underline{y}} = A\hat{\underline{x}} \end{cases}$$

The covariance matrices of $\hat{\underline{x}}$ and $\hat{\underline{y}}$ follow from applying the error propagation law to (A2). This gives:

$$(A3)\quad \begin{cases} Q_{\hat{x}} = \sigma^2 (A^* Q_y^{-1} A)^{-1} \\ Q_{\hat{y}} = \sigma^2 A Q_{\hat{x}} A^* \end{cases}$$

It can be shown that the linear least squares estimators of (A2) are optimal in the sense that they are maximum likelihood estimators as well as minimum variance linear unbiased estimators, provided the Gaussian assumption and model (A1) hold.
In the context of model validation an important role is played by the least squares residual vector $\hat{\underline{e}}$. This random vector is defined as:

$$(A4)\quad \hat{\underline{e}} = \underline{y} - \hat{\underline{y}} = (I - A Q_{\hat{x}} A^* Q_y^{-1})\, \underline{y}$$

Its covariance matrix reads

(A5) $Q_{\hat{e}} = Q_y - Q_{\hat{y}}$.

Slippage Tests

The objective of hypothesis testing is to decide, on the basis of an observed sample of $\underline{y}$, whether model (A1) is correct. Model (A1) will be called our null hypothesis H_0. It will be clear that our null hypothesis may be violated in many different ways: $\underline{y}$ may not be normally distributed, $\underline{y}$ may not have a mean Ax, $\underline{y}$ may not have a covariance matrix $\sigma^2 Q_y$, etc. Here we will restrict ourselves to misspecifications in the mean Ax. That is, we will concentrate on whether the mean of $\underline{y}$ can be represented as Ax or not. More specifically we will assume that the misspecifications, if present, are of an additive nature. Thus we will oppose the null hypothesis

$$H_0 : \underset{mx1}{E\{\underline{y}\}} = \underset{mxn}{A}\ \underset{nx1}{x} \quad , \ \underset{mxm}{\operatorname{cov}\{\underline{y}\}} = \sigma^2 Q_y$$

to one or more relaxed alternative hypothesis H_a of the type

$$H_a : \underset{mx1}{E\{\underline{y}\}} = \underset{mxn}{(A} \quad \underset{mxp}{C)} \ \underset{(n+p)x1}{\begin{bmatrix} x \\ \nabla \end{bmatrix}} , \ \underset{mxm}{\operatorname{cov}\{\underline{y}\}} = \sigma^2 Q_y \ ,$$

where ∇ is a p-dimensional vector of explanatory variables taking care of the postulated misspecifications in H_0.

The Overall Model Test $\underline{T}$

If the mxp matrix C of the alternative hypothesis H_a is chosen such that the matrix (A:C) is square and invertible, then p=m-n and no particular restrictions are imposed on the mean of $\underline{y}$ under H_a. The function of the least squares residuals $\hat{\underline{e}}$ which in this case can be used as a teststatistic for deciding whether H_0 should be accepted or not is given as

$$(A6) \quad \underline{T} = \frac{\hat{\underline{e}}^{\bullet} Q_y^{-1} \hat{\underline{e}}}{\sigma^2 (m-n)} \ .$$

The random variable $\underline{T}$ has a so-called Chi-squared distribution $\chi^2(m-n,\lambda)$ with m-n degrees of freedom and noncentrality parameter λ, where

$$(A7) \quad \lambda = \sigma^{-2} \nabla^{\bullet} C^{\bullet} Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C \nabla \ .$$

The decision rule whether to accept or reject H_0 becomes then

$$\text{(A8)} \quad \begin{cases} \text{reject } H_0 \text{ if } T \geq \chi^2_\alpha(m\text{-}n,0) \\ \text{accept } H_0 \text{ otherwise.} \end{cases}$$

This test is called the **overall model test** since no restrictions are placed on the mean of $\underline{y}$ under H_a. The Chi-squared distribution can be found in standard textbooks on probability theory. $\chi^2_\alpha(m\text{-}n,0)$ is the upper α probability point of the Chi-squared distribution $\chi^2(m\text{-}n,0)$. α is the probability of rejecting H_0 when in fact it is true. It is called the level of significance and is given by

$$\alpha = \text{Prob}\,(\,\chi^2(m\text{-}n,0) \geq \chi^2_\alpha(m\text{-}n,0)\,)$$

By fixing α one can compute $\chi^2_\alpha(m\text{-}n,0)$ and execute test (A8). The test is then said to have size α.

The one-dimensional Slippage Test $\underline{w}$; ∇ unknown

If the number of columns of the mxp matrix C is chosen to be equal to one, then p=1 and the matrix C becomes a vector. This vector is denoted by c. In this case the number of restrictions placed on the mean of $\underline{y}$ under H_a is one less than the number of restrictions placed on the mean of $\underline{y}$ under H_0. The corresponding teststatistic is given by

$$\text{(A9)} \quad \underline{w} = \frac{c^* Q_y^{-1} \underline{\hat{e}}}{\sigma (c^* Q_y^{-1} Q_e Q_y^{-1} c)^{1/2}} .$$

The random variable $\underline{w}$ has a normal distribution $N(\pm\lambda^{1/2},1)$, with a variance of one and a mean of $\pm\lambda^{1/2}$, where the sign is that of ∇ and

$$\text{(A10)} \quad \lambda = \sigma^{-2} \nabla^* c^* Q_y^{-1} Q_e Q_y^{-1} c \nabla .$$

Note that since c is a vector, ∇ is a scalar. The decision rule to accept or reject H_0 on the basis of the formulated alternative hypothesis becomes then

$$\text{(A11)} \quad \begin{cases} \text{reject } H_0 \text{ if } |\underline{w}| \geq N_{1/2\alpha}(0,1) \\ \text{accept } H_0 \text{ otherwise} . \end{cases}$$

So far the null hypothesis H_0 has been opposed to one alternative hypothesis H_a. In order, however, to specify the alternative hypothesis, one has to specify the vector c. But this can be a difficult task indeed, since it is generally not known beforehand what misspecifications, if any, will be present in H_0. Hence, in practice choices have to be made. Let us therefore assume that a set of one-dimensional misspecifications has been defined. Furthermore it will be assumed that if a misspecification occurs, it is one of the set defined.

The problem we face is then to decide whether a misspecification of H_0 has occurred and if so which one. The search for possible misspecifications can then be done as follows. First decide on the basis of the overall model test whether a misspecification occurred at all. If so, then look for the most likely alternative hypothesis in the predefined set. The most likely alternative hypothesis is the one for which $|\underline{w}|$ takes a maximum value. Once the possible misspecification in H_0 has been identified, corrective action can be taken. In order to be able to further identify the source of the misspecification in H_0 it can be useful to compute the linear least squares estimator of ∇ under the accepted alternative hypothesis. This estimator is computed as

$$(A12)\quad \hat{\underline{\nabla}} = \frac{\sigma \underline{w}}{(c^{\bullet} Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c)^{1/2}} .$$

An important example of the search for possible misspecifications in H_0 is given by the search for outlying observations. In this case the predefined set of alternative hypothesis consists of m (the number of observations) alternative hypotheses, each of which models an outlier in one observation at a time. For the *i* th observation the c vector then takes the simple form

$$(A13)\quad \underset{mx1}{c} = (0,.....,0,\ 1,0,.....1)^{*} \qquad .$$
$$\qquad\qquad\quad 1 \quad (i-1),i,(i-1) \quad m$$

This very successful technique for the search of outlying observations is known as **datasnooping** (see [13], [14], [15]).

Above it was assumed that the variance factor of unit weight σ^2 is known. If it is unknown, teststatistics similar to (A9) can be derived. The overall model test, however, then fails to exist. For more details see [14], [15], [16], and [17].

The one-dimensional slippage test $\Delta\underline{w}$; ∇ known

Although the possibility that one is able to formulate an alternative hypothesis with ∇ known very rarely occurs in geodetic applications, there are some exceptions. One such exception is given by the problem of lane slip identification in offshore positioning. If ∇ is known a more powerful test than (A11) can be derived. The corresponding test for deciding between H_0 and H_a reads

$$(A14)\quad \begin{cases} \text{reject } H_0 \text{ if } \Delta\underline{w} < 0 \\ \text{accept } H_a \text{ otherwise} \end{cases} ,$$

where the teststatistic $\Delta\underline{w}$ is given as

$$(A15)\quad \Delta\underline{w} = \frac{\nabla^{\bullet} c^{\bullet} Q_y^{-1} Q_{\hat{e}} Q_y^{-1} (2\underline{y} - c\nabla)}{(\nabla^{\bullet} c^{\bullet} Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c \nabla)^{1/2}} .$$

Measures for the detectability of alternative hypotheses

In the field of hypothesis testing two important questions are: 1. How well can we detect a particular alternative hypothesis? and 2. How well can we discriminate between different alternative hypotheses?
If the c-vectors of all alternative hypotheses are normalized to the same value, one can take as a measure of **detectability** the scalar

$$(A16)\quad r_{\nabla}^2 = \frac{\sigma^2}{c^* Q_y^{-1} Q_e Q_y^{-1} c} \ .$$

The larger r_{∇}^2 is, the less detectable the corresponding alternative hypothesis.
Note that r_{∇}^2 equals the variance of $\hat{\nabla}$ in (A12). This gives an interpretation to the measure of detectability in terms of the estimability of ∇ under H_a. In particular if the variance of $\hat{\nabla}$ is infinite the corresponding alternative hypothesis is not detectable.
As a measure of **discriminability** between two alternative hypothesis H_{ai} and H_{aj} one can take

$$(A17)\quad s_{ij} = \frac{c_i^* Q_y^{-1} Q_e Q_y^{-1} c_j}{\sqrt{c_i^* Q_y^{-1} Q_e Q_y^{-1} c_i}\sqrt{c_j^* Q_y^{-1} Q_e Q_y^{-1} c_j}} \ .$$

Note that s_{ij} equals the correlation coefficient between the estimators $\underline{\hat{\nabla}}_i$ and $\underline{\hat{\nabla}}_j$. It also equals the correlation coefficient between the teststatistics $\underline{w}_i$ and $\underline{w}_j$. The larger s_{ij} is, the less one is able to discriminate between H_{ai} and H_{aj} .
For an in depth treatment of the necessary theory, which comes under the general heading of **reliability theory**, the reader is referred to the referenced literature.

Appendix B: The Linear Kalman Filter; A Possible Derivation

There exist various derivations of the linear Kalman filter. These derivations are based on principles like least squares, minimum mean square error, maximum likelihood, and maximum a posteriori. In general the use of different principles leads to different estimators. However, in case of linear systems where the probability density functions are assumed to be Gaussian all the above mentioned estimation methods yield the same estimator. Thus, the framework used to discuss such systems reduces to one of personal preference. Since the principle of least squares is probably the one which surveyors and hydrographers are the most familiar with, our derivation of the linear Kalman filter is based on this principle.

The linear model of observation equations from which the linear Kalman filter can be derived is given as

$$(B1)\quad E\left\{\begin{bmatrix}\hat{\underline{x}}_{k-1|k-1}\\ \underline{d}_k\\ \underline{y}_k\end{bmatrix}\right\} = \begin{bmatrix} I & 0\\ -\Phi_{k,k-1} & I\\ 0 & A_k\end{bmatrix}\begin{bmatrix}x_{k-1}\\ x_k\end{bmatrix};\ \begin{bmatrix} P_{k-1|k-1} & 0 & 0\\ 0 & Q_k & 0\\ 0 & 0 & R_k\end{bmatrix}$$

Note that this model is of the same form as the linear model of observation equations (A1) given in appendix A.
The Gaussian random vector $\hat{\underline{x}}_{k-1|k-1}$, with covariance matrix $P_{k-1|k-1}$, is our estimator of the state x_{k-1} at time k-1. It summarizes all the information available at time k-1 about state x_{k-1}.
The Gaussian random vector $\underline{d}_k$, with covariance matrix Q_k, is our estimator of the difference between the state x_k and the propagated state $\Phi_{k|k-1}x_{k-1}$. Would we know the dynamic model perfectly, we would set both the mean $E\{\underline{d}_k\}$ and covariance matrix Q_k equal to zero. Due to all sorts of random disturbances, however, one is in practice usually not able to model the dynamics of the system completely. This is why the difference between the state and propagated state is modelled as a random vector.
The Gaussian random vector $\underline{y}_k$, with covariance matrix R_k, is an estimator of the observational variates at time k. Its mean is related to the state x_k through the design matrix A_k.
In order to estimate we need sample values. In practice we have only samples available for $\hat{\underline{x}}_{k-1|k-1}$ and $\underline{y}_k$. The sample of $\hat{\underline{x}}_{k-1|k-1}$ is given by the best estimate of x_{k-1} at time k-1 and the sample of $\underline{y}_k$ is given by the observations. There is, however, no sample available for $\underline{d}_k$. Since the difference between the state and propagated state is considered to be small, the random vector $\underline{d}_k$ is treated as a pseudo observational variate for which the sample value can be taken equal to zero.

Prediction

Let us first consider the least squares estimation of the state without the use of the observations $\underline{y}_k$. Model (B1) reduces then to

$$(B2)\quad E\left\{\begin{bmatrix}\hat{\underline{x}}_{k-1|k-1}\\ \underline{d}_k\end{bmatrix}\right\} = \begin{bmatrix} I & 0\\ -\Phi_{k,k-1} & I\end{bmatrix}\begin{bmatrix}x_{k-1}\\ x_k\end{bmatrix};\ \begin{bmatrix} P_{k-1|k-1} & 0\\ 0 & Q_k\end{bmatrix}$$

Note that there is no redundancy since the model contains 2n equations with 2n unknowns. Thus the available estimate $\hat{x}_{k-1|k-1}$ of x_{k-1} cannot be improved upon. Due to the lack of redundancy in (B2) the least squares estimator of x_k, which we shall denote by $\hat{\underline{x}}_{k|k-1}$, simply follows from inverting the designmatrix of (B2). Thus

$$(B3)\quad \hat{\underline{x}}_{k|k-1} = \Phi_{k,k-1}\hat{\underline{x}}_{k-1|k-1} + \underline{d}_k$$

Application of the error propagation law gives for the covariance matrix of $\underline{x}_{k|k-1}$:

$$(B4)\quad P_{k|k-1} = \Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1} + Q_k\ .$$

Since the sample value of $\underline{d}_k$ is taken equal to zero it follows from (B3) that the least squares estimate of x_k based on model (B2) is given by

$$(B5) \quad \hat{x}_{k|k-1} = \Phi_{k,k-1}\hat{x}_{k-1|k-1}$$

Equations (B4) and (B5) constitute the well known **time update equation** of the linear Kalman filter. They are given as equations (2a) and (2b) in section 2.

Filtering

Let us now consider the least squares estimation of the state with the observations $\underline{y}_k$ included. In this case model (B1) applies. Since there is redundancy (the redundancy equals the dimension of the vector of observations) the available estimate $\hat{x}_{k-1|k-1}$ of x_{k-1} can now be improved. This improvement is, however, part of smoothing (i.e. one uses the observations of time k to estimate the state at time k-1) and is not considered in the Kalman filter. We therefore eliminate the state x_{k-1} from model (B1). This gives

$$(B6) \quad E\left\{\begin{bmatrix}\Phi_{k,k-1}\hat{\underline{x}}_{k-1|k-1}+\underline{d}_k\\ \underline{y}_k\end{bmatrix}\right\} = \begin{bmatrix}I\\ A_k\end{bmatrix}x_k;\begin{bmatrix}\Phi_{k,k-1}P_{k-1|k-1}\Phi^{\bullet}_{k,k-1}+Q_k & 0\\ 0 & R_k\end{bmatrix}$$

With (B3) and (B4) this can also be written as

$$(B7) \quad E\left\{\begin{bmatrix}\hat{\underline{x}}_{k|k-1}\\ \underline{y}_k\end{bmatrix}\right\} = \begin{bmatrix}I\\ A_k\end{bmatrix}x_k;\begin{bmatrix}P_{k|k-1} & 0\\ 0 & R_k\end{bmatrix}$$

Straightforward application of the least squares algorithm gives for the least squares estimator of x_k, denoted by $\hat{\underline{x}}_{k|k}$:

$$(B8) \quad \hat{\underline{x}}_{k|k} = (P^{-1}_{k|k-1} + A^{\bullet}_k R^{-1}_k A_k)^{-1}(P^{-1}_{k|k-1}\hat{\underline{x}}_{k|k-1} + A^{\bullet}_k R^{-1}_k \underline{y}_k)$$

Application of the error propagation law gives for the covariance matrix of $\hat{\underline{x}}_{k|k}$:

$$(B9) \quad P_{k|k} = (P^{-1}_{k|k-1} + A^{\bullet}_k R^{-1}_k A_k)^{-1}$$

Equations (B8) and (B9) constitute the so-called **measurement update equations** of the linear Kalman filter. An alternative form of the measurement update equations can be obtained by invoking the following matrix inversion lemma:

$$(B10) \quad (C^{-1} + B^{\bullet}D^{-1}B)^{-1} = C - CB^{\bullet}(D + BCB^{\bullet})^{-1}BC \ ,$$

where C and D are symmetric matrices. The identity (B10) is easily verified by multiplying the right hand side of (B10) with $[C^{-1}+B^{*}D^{-1}B]$. Application of the matrix inversion lemma (B10) to (B8) and (B9) gives after some arrangements for the filtered state

(B11) $\hat{\underline{x}}_{k|k} = \hat{\underline{x}}_{k|k-1} + P_{k|k-1}A_k^{\bullet}(R_k+A_kP_{k|k-1}A_k^{\bullet})^{-1}(\underline{y}_k - A_k\hat{\underline{x}}_{k|k-1})$,

and for its covariance matrix

(B12) $P_{k|k} = P_{k|k-1} - P_{k|k-1}A_k^{\bullet}(R_k+A_kP_{k|k-1}A_k^{\bullet})^{-1}A_kP_{k|k-1}$.

The measurement update equations (B11) and (B12) are the ones which are ussually presented in the literature. They are given as equations (3a) and (3b) in section 2.

The local teststatistics for the Kalman filter which are presented in section 4 can be derived by applying the results of appendix A to model (B6). In a similar way the global teststatistics for the Kalman filter presented in section 5 can be derived from applying the results of appendix A to the batch mode form of model (B1).