Statistical Analysis of Kalman Filters by Conversion to Gauss-Helmert Models with Applications to Process Noise Estimation

Arne Petersen, Reinhard Koch
Institute of Computer Science, Christian-Albrechts-University of Kiel
{apetersen,rk}@mip.informatik.uni-kiel.de

Abstract

This paper introduces a reformulation of the extended Kalman Filter using the Gauss-Helmert model for least squares estimation. By proving the equivalence of both estimators it is shown how the methods of statistical analysis in least squares estimation can be applied to the prediction and update process in Kalman Filtering. Especially the efficient computation of the reliability (or redundancy) matrix allows the implementation of self supervising systems. As an application an unparameterized method for estimating the variances of the filters process noise is presented.

1. Introduction

Since its introduction by R. E. Kalman in 1960 the Kalman Filter has become widely used in signal processing. Its applications range from simple estimation of constants to complex navigation filters using multi-sensor fusion. This is due to the fact, that a well tuned Kalman Filter produces excellent results with a rather small computational effort. To deduce such a filter its tuning and therefore its statistical analysis is crucial. Although numerous techniques have been proposed to analyze the fusion of the predicted state and the observations (e.g. residual-tests for outlier detection, see chapter 2.2.4.6.3 in [5]), there is a lack on concepts for the analysis of the whole filter.

For least squares estimators (LSE) the situation is different. A wide theory of methods for system-analysis is available (e.g. see [4]) but the applications of LSEs to real-time problems are rather rare. Therefore this paper aims at the reformulation of the extended Kalman Filter using the Gauss-Helmert model (GHM, see [5] for details) as one of the most general LSEs. This way access to the theory for GHMs is granted to the Kalman Filter. The reliability analysis for example can be used to determine the effect of single observations on the estimation, as explained more detailed in [1].

2. Models and notation

In this section the notation for the Kalman Filter and the corresponding GHM are introduced. The notations are similar to those in [6] for the Kalman Filter and [5] for the GHM respectively. There can be found more detailed information about both methods also.

2.1. The Kalman Filter

The Kalman Filter estimates a sequence $\hat{x}_k$ of n-dimensional system states together with their covariance matrices P_k from a given sequence of m-dimensional observations y_k. This is accomplished using the prediction $x_k = f_k(x_{k-1}, w_{k-1})$ and the measurement $y_k + v_k = h_k(x_k)$ equations. Here $w_k \sim \mathcal{N}(0, Q_k)$ is the zero-mean process noise and $v_k \sim \mathcal{N}(0, R_k)$ the zero-mean observation noise. In the following x_k^- denotes the a priori and $\hat{x}_k$ the a posteriori state estimate. Furthermore let:

$$F_k = \frac{\partial f_k(x,0)}{\partial x}|_{x=\hat{x}_{k-1}} \quad H_k = \frac{\partial h_k(x)}{\partial x}|_{x=x_k^-}$$

$$W_k = \frac{\partial f_k(\hat{x}_{k-1}, w)}{\partial w}|_{w=0}$$

Starting with the initial values x_0 and P_0 the filter alternating predicts the a priori state:

$$\begin{aligned} x_k^- &= f_k(\hat{x}_{k-1}, 0) \\ P_k^- &= F_k P_{k-1} F_k^T + W_k Q_{k-1} W_k^T \end{aligned}$$

and updates this estimate using:

$$\begin{aligned} K_k &= P_k^- H_k^T \left(H_k P_k^- H_k^T + R\right)^{-1} \\ \hat{x}_k &= x_k^- + K_k \cdot \left(y_k - h_k(x_k^-)\right) \\ P_k &= \left(I - K_k H_k\right) P_k^- \end{aligned}$$

2.2. The Gauss-Helmert model

In the GHM a set of parameters $\hat{p}$ and their error covariance matrix P_p are estimated from a given set of observations l having the error covariance matrix P_l. The parameters and the observations are connected through s implicit constraints $g(\hat{p}, l + \hat{v}) = 0_s$. Here $\hat{v}$ represents the zero-mean residual vector that may be computed during the estimation process. To solve the given equation system a parameter initialization p^0 as well as gs Jacobians are needed. Let:

$$A = \frac{\partial g(p,l)}{\partial p}|_{p=p^0} \quad B = \left(\frac{\partial g(p^0,l)}{\partial l}|_l\right)^T$$

The first parameter update can now be computed using:

$$\begin{aligned} P_p &= \left(A^T \left(B^T P_l B\right)^{-1} A\right)^{-1} \\ \Delta p &= -P_p A^T \left(B^T P_l B\right)^{-1} g(p^0, l) \end{aligned}$$

For non-linear constraints it is necessary to compute the solution iteratively. As will be shown only the first iteration step is needed for the estimation of the extended Kalman Filter so we won't go into detail on how to iterate. Using the notation from section 2.1 we now define:

$$\begin{aligned} p_k^0 &= x_k^- = f_k(\hat{x}_{k-1}, 0) \\ l_k = \begin{pmatrix} \hat{x}_{k-1} \\ w_{k-1} \\ y_k \end{pmatrix}, P_{l_k} &= \begin{pmatrix} P_{k-1} & 0 & 0 \\ 0 & Q_{k-1} & 0 \\ 0 & 0 & R_k \end{pmatrix} \quad (1) \\ g_k(p_k, l_k) &= \begin{pmatrix} f_k(\hat{x}_{k-1}, w_{k-1}) - p_k \\ y_k - h_k(p_k) \end{pmatrix} \end{aligned}$$

Since the observation w_{k-1} for the process noise is unknown we assume $w_{k-1} = 0$ like it is done in the prediction step in the Kalman Filter. For the design matrices one can verify that:

$$A_k = \begin{pmatrix} -I_n \\ -H_k \end{pmatrix} \qquad B_k = \begin{pmatrix} F_k^T & 0 \\ W_k^T & 0 \\ 0 & I_m \end{pmatrix} \quad (2)$$

3. Proof of equivalence

In this section we will prove that solving the above defined GHM is equivalent to the combined prediction and update for step k in the Kalman Filter. To verify this equivalence we have to deduce two equations needed at a later stage at first. Setting $S := H_k P_k^- H_k^T + R_k$ that is $K_k = P_k^- H_k^T S^{-1}$ we can compute the gain K_k for the Kalman Filter by:

$$\begin{aligned} K_k &= K_k \left(H_k P_k^- H_k^T R_k^{-1} + I - H_k P_k^- H_k^T R_k^{-1}\right) \\ &= P_k^- H_k^T S^{-1} \left(S R_k^{-1} - H_k P_k^- H_k^T R_k^{-1}\right) \\ &= P_k^- H_k^T R_k^{-1} - P_k^- H_k^T S^{-1} H_k P_k^- H_k^T R_k^{-1} \\ &= \left(I - K_k H_k\right) P_k^- H_k^T R_k^{-1} \\ &= P_k H_k^T R_k^{-1} \quad (3) \end{aligned}$$

Furthermore the projected observation covariance in the GHM using the design matrices from (2) reduces to:

$$\begin{aligned} &\left(B_k^T P_{l_k} B_k\right)^{-1} \\ &= \begin{pmatrix} \left(F_k P_{k-1} F_k^T + W_k Q_{k-1} W_k^T\right) & 0 \\ 0 & R_k \end{pmatrix}^{-1} \\ &= \begin{pmatrix} P_k^{-\,-1} & 0 \\ 0 & R_k^{-1} \end{pmatrix} \quad (4) \end{aligned}$$

Utilizing these equalities we can now deduce the isomorphism between both models. The first derivation proves the equality of the error covariance for the estimated parameters $\hat{p}$ in the GHM with the one for the filtered state $\hat{x}_k$ from the Kalman Filter. The indices k are omitted in the following for better readability.

$$\begin{aligned} P_p &= \left(A^T \left(B^T P_l B\right)^{-1} A\right)^{-1} \\ &= \left(A^T \begin{pmatrix} P^{-\,-1} & 0 \\ 0 & R^{-1} \end{pmatrix} A\right)^{-1} \\ &= \left(P^{-\,-1} + H^T R^{-1} H\right)^{-1} \\ &\overset{*}{=} P^- - P^- H^T \left(H P^- H^T + R\right)^{-1} H P^- \\ &= \left(I_n - KH\right) P^- = P \end{aligned}$$

Here equality $(*)$ follows from the inversion formula given on page 26 in [5].

In the second derivation we show the equality of the computed updates for the GHM parameters $\hat{p}$ with the filter state $\hat{x}$ by exploiting equalities (4) and (3).

$$\begin{aligned} \Delta p &= -P_p A^T \left(B^T P_l B\right)^{-1} g(p^0, l) \\ &= -P \left(-I_n \mid -H^T \right) \begin{pmatrix} P^{-\,-1} & 0 \\ 0 & R^{-1} \end{pmatrix} g(p^0, l) \\ &= \left(P P^{-\,-1} \mid P H^T R^{-1} \right) \begin{pmatrix} 0 \\ y - h(x^-) \end{pmatrix} \\ &= P H^T R^{-1} \left(y - h(x^-)\right) = K \left(y - h(x^-)\right) \end{aligned}$$

Finally by using this identity we can verify that:

$$\hat{p}_k = p_k^0 + \Delta p_k = x_k^- + K_k \left(y_k - h(x_k^-)\right) = \hat{x}_k$$

4. Statistical analysis of Kalman Filters

This section provides the link between the statistical analysis in GHMs and Kalman Filters. After proving the equivalence of the Kalman Filter and the defined GHM it is evident that an isomorphic GHM can be given for any extended Kalman Filter. Observe that in definition (1) the a posteriori state $\hat{x}_{k-1}$, the process noise w_{k-1} and the observations y_k as well as the according covariance matrices are used as observations to estimate $\hat{x}_k$. This way the GHM can be used for statistical analysis not only of the filters observations but also of its history and process noise.

The computation of the reliability - sometimes called redundancy - matrix $\mathcal{R}$ allows the analysis of the contribution of the single observations in the GHM to its overall redundancy (see [1] or [5]). For the given GHM this matrix can be computed efficiently due to the sparseness of the involved matrices. By this means the influence of $\hat{x}_{k-1}$,w_{k-1} and y_k on the estimation of $\hat{x}_k$ can be analyzed. From this conclusions can be drawn to the effectiveness of the filtering process. Assume a reliability number (diagonal entry of $\mathcal{R}$) reaches zero for a component of w_{k-1}. Then this component has no contribution to the redundancy of the estimation process. From this follows that this part of the process noise and thus the error it introduces cannot be corrected by the observations and the state history.

Further work on the exploitation of $\mathcal{R}$ can be found in [1] and [2]. There applications for the reliability matrix in Gauss-Markov models are given for sensitivity analysis and advanced outlier detection respectively. Since the estimation in a GHM can be transformed into a Gauss-Markov model (see [4] or [1]) the proposed methods may be applied to the presented GHM with some small adaptations.

Another application is the estimation of variance components. This method allows the estimation of scaling factors for the covariances of clusters of the used observations in the GHM. By this means the variances of the process noise for a Kalman Filter can be estimated given a sufficient large dataset. An algorithm using this estimation method is presented in the following section.

5. Example: Process noise estimation

The identification of the process noise for a given Kalman Filter is a hard problem in the most cases. Work has been done to estimate the noise parameters for a filter using basically two different approaches. One is to create an adaptive filter (see [3]), the other is to maximize a likelihood function with respect to the noise parameters (see [7]). The former approach increases the computational effort during the filtering process and the latter is only applicable for special process models.

In the following the presented model-transformation will be used for offline estimation of the variances of a Kalman Filters process noise. This does not increase the computational effort for the filter and can be applied to every Kalman Filter defined in the model presented in 2.1. Therefor it is assumed that the process noise is stable and statistically independent component-by-component, that is the covariance matrices $Q_k \equiv diag(\sigma_1^2, ..., \sigma_r^2) =: Q$ are diagonal and constant. This restriction is acceptable because known variations and correlations can be incorporated using the system functions f_k. Furthermore a given sequence of observations $y_1, ..., y_N$ with corresponding covariances and a filter initialization x_o, P_0 are required.

At first we define a simple Kalman Filter for estimating a one dimensional movement. As prediction model a constant velocity is assumed so that a two-dimensional state using position and velocity is applicable. As measurements two independent observations of the position and one of the velocity are used, that is y_k is three-dimensional. Finally the model functions are:

$$\begin{aligned} f_k(x_{k-1}, w_{k-1}) &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot x_{k-1} + w_{k-1} \\ h_k(x_k) &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot x_k \end{aligned}$$

Secondly the proposed GHM needs to be stacked. That is the estimation of an improved state sequence $x_1, ..., x_N$ is done as global adjustment. Therefor we define a new GHM:

$$p = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \quad p^0 = \begin{pmatrix} x_1^- \\ \vdots \\ x_N^- \end{pmatrix}$$

$$l = \begin{pmatrix} x_0 \\ w_1 \\ \vdots \\ w_N \\ y_1 \\ \vdots \\ y_N \end{pmatrix} \quad g(p, l) = \begin{pmatrix} f_1(x_0, 0) - x_1 \\ \vdots \\ f_N(x_{N-1}, 0) - x_N \\ y_1 - h_1(x_1) \\ \vdots \\ y_N - h_N(x_N) \end{pmatrix}$$

$$P_l = diag\left(P_0, Q \cdots Q, R_1 \cdots R_N\right)$$

Where again $w_k = 0 \ \forall k$ is assumed. Observe that for the minimal case $N = 1$ this is the same model as in 2.2.

Thirdly a pre-run with the Kalman Filter is needed from which the parameter initialization (the filters a priori states) and the filters Jacobians are retrieved. For this filter run the given observations and an arbitrary diagonal matrix $Q_k \equiv Q$ for the process noise are used.

Afterwards the method for covariance component estimation (CCE, see [4] or [5]) in GHMs is applied for one iteration. Therefor the observations l are divided

Dataset number	Runtime in Sec.	σ_1^2 initial	σ_2^2 initial	$\hat{\sigma}_1^2$ estimated	$\hat{\sigma}_2^2$ estimated	RMS pos. initial	RMS pos. improved	RMS vel. initial	RMS vel. improved
1	69	1e-4	1e-6	0.184	0.001	5.49	1.072	0.4	0.204
	84	100	100	0.184	0.001	1.838	1.072	0.315	0.204
	54	0.15	5e-4	0.184	0.001	1.195	1.072	0.244	0.204
2	64	1e-4	1e-6	0.193	0.0011	5.49	1.029	0.403	0.209
	83	100	100	0.193	0.0011	1.815	1.029	0.324	0.209
	52	0.15	5e-4	0.193	0.0011	1.168	1.029	0.249	0.209

Table 1. Examples for the estimated process noise with different datasets and start variances.

into three groups. One group for the first components of the process noise $w_1^{(1)}, ..., w_N^{(1)}$, one for the second components $w_1^{(2)}, ..., w_N^{(2)}$ and one for the remaining observations. When parts of the process noise are not observable, that is their reliability number is 0, they can be grouped with other observable parts. From the first CCE iteration three scaling factors for the covariances of these groups result. With them the corresponding covariance matrices are scaled and afterwards used for a re-run of the Kalman Filter. With the resulting a priori states as initialization a new CCE iteration is started. This iteration technique is continued until all scaling factors reach one, that is the covariances don't change anymore. The resulting values for $\hat{Q} = diag(\hat{\sigma}_1, \hat{\sigma}_2)$ are the estimates for the filters process noise covariance.

This way the filter can be tuned for the characteristics of a dataset without the need of guesses for the unknown process noise. The proposed experiments have been performed on two synthetical datasets for $N = 1000$ time-steps and are presented in table 1. It can be seen that the estimated values for σ_1 and σ_2 are independent of the initial matrix Q and clearly reduce the root mean square error (RMS) for the filtered state sequence. The influence of the observation noise on the estimated variances is measurable for the used number of observations but decreases with increasing dataset size. A cutout for the plot of positions for the first experiment (first line in the table) is visualized in figure 1.

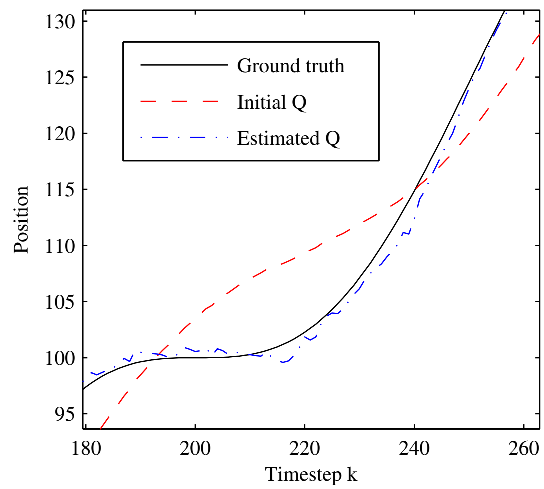

Figure 1. A cutout from the plot of ground truth and kalman-filtered positions.

6. Conclusions

It has been shown how an arbitrary extended Kalman Filter can be transformed into a Gauss-Helmert model for least squares estimation. This way the existing methods of statistical analysis for least squares estimations can be easily applied to the Kalman Filter.

Furthermore the given example application has proven that the covariance component estimation can be used to efficiently estimate the noise parameters from a given set of observations. For the estimation no initial guess for the noise parameters is needed. This way a powerful tool for filter tuning is provided.

References

[1] J. D. Cothren. *Reliability in Constrainded Gauss-Markov Models*. PhD thesis, Ohio State University, 2004.

[2] X. Ding and R. Coleman. Multiple outlier detection by evaluating redundancy contributions of observations. *Journal of Geodesy*, 1996.

[3] C. Hide, T. Moore, and M. Smith. Adaptive kalman filtering algorithms for integrating gps and low cost ins. *Position Location and Navigation Symposium*, 2004.

[4] K.-R. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, 2 edition, 1999.

[5] J. C. McGlone, W. Förstner, and B. Wrobel. *Manual of Photogrammetry*, chapter 2 Mathematical Concepts in Photogrammetry. asprs, 2004.

[6] G. Welch and G. Bishop. An introduction to the kalman filter. Technical report, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 2006.

[7] K.-V. Yuen, K.-I. Hoi, and K.-M. Mok. Selection of noise parameters for kalman filter. *Earthquake Engineering and Engineering Vibration*, 2007.