KALMAN FILTERING ALGORITHM IN PRESENCE OF OUTLIERS

V.I. LOBACH
Belarusian State University
Minsk, BELARUS
e-mail: lobach@bsu.by

Abstract
A Kalman Filtering algorithm which is robust to observational outliers is developed by assuming that the measurement error may come from either one of two normal distributions and that transition between these distribution is governed by a Markov Chain. The state estimate is obtained as a weighted average of the estimates from the two parallel filters where the weights are the posterior probabilities. The impotents obtained by this Robust Kalman Filter in the presence of outliers is demonstrated with examples.

1 Introduction
The Kalman Filter is well known recursive estimator for the state of a linear system and has been used in the fields of forecasting and control. It has been derived as a least squares estimator, and also, under the assumption of normality as a Bayesian estimate. However, as with most least squares estimators, it is very sensitive to observational outliers. This sensitivity to outliers is a major draw back of the filter. Bad observations arising from periodic sensor problems can seriously bias the filter estimates for a considerable period of time thereafter.

In this paper a Bayesian approach is used to derive a recursive state estimator when occasional spurious observations will arise. The new state estimator is shown to have the structure of two parallel Kalman Filters in which the final state estimate is a weighted average of the estimators from the two filters.

2 The State Variable Model
Consider the state variable model given by

\[ x_t = Fx_{t-1} + G\alpha_t, \]  
\[ z_t = Hx_t + \varepsilon_t, \]

where \( x_t \) is an \((n \times 1)\) vector of state variables defining the dynamic behavior of a system, \( F, G \) are known matrices (possibly time varying) and \( H \) is a known row vector of the appropriate dimensions. The measurement \( z_t \) taken at time \( t \) is related linearly to the states through the measurement equation (2.2). The state and observation noise sequences \( \{\alpha_t\} \) and \( \{\varepsilon_t\} \) are usually assumed to be i.i.d. Gaussian random variables with known variances \( \sigma_\alpha^2 \) and \( \sigma_\varepsilon^2 \) respectively. Under the assumption that the system
that the conditional density with zero mean and variance and that the transition probability matrix is a Markov Chain (MC). Suppose that the initial probability \( (\pi_0, \pi_1) \) leads to a real change in the state of the system or in the variance of \( \alpha_t \) (an observational outlier) which has no effects on the true state \( (x_t) \), but will have a pronounced effect on the Kalman Filter estimates. In the following section we use a Bayesian approach to develop a robust filter to handle observational outliers.

3 Derivation of the Robust Filter

We characterize the observational outliers by replacing \( \varepsilon_t \) in equation (2.2) by \((1 - \tilde{\alpha}_t)\varepsilon_t + \tilde{\alpha}_t \gamma_t\), where \( \{\tilde{\alpha}_t, t = 0, 1, \ldots\} \) is a sequence of \( \{0, 1\} \) random variables following a Markov Chain (MC). Suppose that the initial probability \((\pi_0, \pi_1)\) of MC is known and that the transition probability matrix is \( \{q_{t-1, t}\} \). \( \gamma_t \) is a Gaussian random variable with zero mean and variance \( \sigma_\gamma^2 \) which is very large compared with \( \sigma_\varepsilon^2 \). We also assume that the conditional density \( p(x_t | z_{0}^{t-1}) \) is a Gaussian distribution with mean \( \hat{x}_{t-1} \) and covariance matrix \( P_{t-1}. \)

Then

\[
p(x_t, i_t | z_0^{t-1}) = \sum_{i_{t-1}=0}^{1} \int p(x_{t-1}, i_{t-1} | z_0^{t-1}) \cdot p(x_t, i_t | x_{t-1}, i_{t-1}, z_0^{t-1}) dx_{t-1},
\]

which on further simplification yields [1], [2]:

\[
p(x_t, i_t | z_0^{t-1}) = (2\pi)^{-\frac{n}{2}} \sum_{i_{t-1}=0}^{1} q_{i_{t-1}, i_t} k_{i_{t-1}} \left| P_{i_{t-1}} \right|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (x_t - \hat{x}_{i_{t-1}})^T P_{i_{t-1}}^{-1} (x_t - \hat{x}_{i_{t-1}}) \right\},
\]

where \( q_{i_{t-1}, i_t} = P \{i_t | i_{t-1}, z_0^{t-1}\} \), \( P_{i_{t-1}} = FP_{i_{t-1}}F^T + GG^T \sigma_\alpha^2 \), \( \hat{x}_{i_{t-1}} = F\hat{x}_{i_{t-1}} \).

By Bayes theorem

\[
p(x_t, i_t | z_0^{t}) = \frac{p(x_t, i_t | z_0^{t-1}) \cdot p(z_t | x_t, i_t, z_0^{t-1})}{\sum_{i_{t-1}=0}^{1} \int p(x_{t-1}, i_{t-1} | z_0^{t-1})p(z_t | x_{t-1}, i_{t-1}, z_0^{t-1}) dx_{t-1}} =
\]

\[
= C \cdot (2\pi)^{-\frac{n}{2}} \left( \sum_{i_{t-1}=0}^{1} q_{i_{t-1}, i_t} \cdot k_{i_{t-1}} \right) \left| P_{i_{t-1}} \right|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (x_t - \hat{x}_{i_{t-1}})^T P_{i_{t-1}}^{-1} (x_t - \hat{x}_{i_{t-1}}) \right\} \times \exp \left\{ -\frac{1}{2} (z_t - Hx_t)^T \sigma_{i_t}^{-2} (z_t - Hx_t) \right\},
\]
where
\[ \sigma_i^2 = (1 - i_t)\sigma_e^2 + i_t\sigma_\gamma^2, \]
\[ C^{-1} = \sum_{i_t=0}^{1} \int p(x_t, i_t | z_0^{(-1)}) \cdot p(z_t | x_t, i_t, z_0^{(-1)}) dx_t. \]
This can be simplified
\[ p(x_t, i_t | z_0^t) = k_{i_t} (2\pi)^{-n/2} |P_{t|t}^{(i_t)}|^{-1/2} \exp \left\{ -\frac{1}{2} \left( x_t - \hat{x}_{t|t}^{(i_t)} \right) \cdot \left( P_{t|t}^{(i_t)} \right)^{-1} \cdot \left( x_t - \hat{x}_{t|t}^{(i_t)} \right) \right\}, \]
where
\[ \hat{x}_{t|t}^{(i_t)} = \hat{x}_{t|t-1} + R_{i_t}^{-1} P_{t|t-1} H^T (z_t - H \hat{x}_{t|t-1}), \tag{3.1} \]
\[ P_{t|t}^{(i_t)} = P_{t|t-1} - R_{i_t}^{-1} P_{t|t-1} H P_{t|t-1} H^T, \tag{3.2} \]
\[ R_{i_t} = \sigma_i^2 + H P_{t|t-1} H^T, \tag{3.3} \]
and
\[ k_{i_t} = C \cdot (\sqrt{2\pi}\sigma_{i_t})^{-1} \frac{1}{2} \left| P_{t|t}^{(i_t)} \right| \frac{1}{2} \left( \sum_{i_{t-1}=0}^{1} q_{i_{t-1},i_t} k_{i_{t-1}} \right) \left| P_{t|t-1} \right|^{-1/2} \times \]
\[ \exp \left\{ -\frac{1}{2} (z_t - H \hat{x}_{t|t-1})^T R_{i_t}^{-1} (z_t - H \hat{x}_{t|t-1}) \right\}. \tag{3.4} \]

The \( k_{i_t} \) are the posterior probabilities of the current observation at time \( t \) being an outlier \( (i_t = 1) \) or not being outlier \( (i_t = 0) \). \( C \) is a normalizing constant. Note that the term \( \sum_{i_{t-1}=0}^{1} q_{i_{t-1},i_t} k_{i_{t-1}} = P_{t|i_t} \) is the prior probability of being in state \( i_t \) at time \( t \).

Thus
\[ p(x_t | z_0^t) = \sum_{i_t=0}^{1} k_{i_t} N \left( \hat{x}_{t|t}^{(i_t)}, P_{t|t}^{(i_t)} \right), \tag{3.5} \]
where \( N \left( \hat{x}_{t|t}^{(i_t)}, P_{t|t}^{(i_t)} \right) \) denotes a normal distribution with mean \( \hat{x}_{t|t}^{(i_t)} \) and covariance matrix \( P_{t|t}^{(i_t)} \). For computational simplicity we approximate the posterior density (3.5) by a single Gaussian distribution with matching moments. We write this posterior density as
\[ p(x_t | z_0^t) = N(\hat{x}_t, P_t), \]
where \( \hat{x}_t = \sum_{i_t=0}^{1} k_{i_t} \hat{x}_{t|t}^{(i_t)} \), \( P_t = \sum_{i_t=0}^{1} k_{i_t} P_{t|t}^{(i_t)} \).

In the proposed robust filter (3.1) – (3.4) the gain corresponding to \( i_t \) is given by
\[ K_{i_t} = R_{i_t}^{-1} \hat{P}_{t|t-1} H^T, \]
where \( R_{i_t} \) is given in (3.3). Thus the state estimate is computed as a linear combination of two parallel filters having a lower and a higher gain depending on the current observation \( z_t \).
4 Simulation Examples

4.1 ARIMA(0,1,1) model with outliers

Consider the following special case of the system of equations given by (2.1), (2.2).

\[ x_t = x_{t-1} + \alpha_t, \]  
\[ z_t = x_t + \varepsilon_t, \]  

where \( \{\alpha_t\} \) and \( \{\varepsilon_t\} \) are i.i.d. Gaussian random variables with mean zero and variances \( \sigma^2_\alpha \) and \( \sigma^2_\varepsilon \) respectively. This is equivalent to an ARIMA(0,1,1) model:

\[ \nabla z_t = (1 - \theta B)b_t \]  

for \( \theta > 0 \), where \( \theta \) is the moving average (MA) parameter, is a backshift operator such that \( Bz_t = z_{t-1} \) and \( b_t \) is an i.i.d. sequence of Gaussian random variables with mean zero and variance \( \sigma^2_b \). The relationships among the parameters in the state model form (4.1), (4.2) and difference equation (4.3) are given in [3] as

\[ (1 - \theta^2)/\theta = \sigma^2_\alpha/\sigma^2_\varepsilon, \quad \sigma^2_b = \sigma^2_\alpha/(1 - \theta)^2. \]

The procedure in the previous section is illustrated with time series (4.3). The subsequent 100 observations were then considered as the sample period and outliers were introduced at the 25th, 50th, 65th and 75th observations by adding \( N(0, 1.75) \) random variables to them. Initially the ordinary Kalman filter (with no outlier protection) was run with \( \hat{x}_0 = 17 \) (the first observation in the sample period). \( \hat{P}_0 = 0.009, \hat{\sigma}^2_\alpha = 0.009 \) and \( \hat{\sigma}^2_\varepsilon = 0.071 \). The modeling results show that the ordinary Kalman filter estimates are sensitive to spurious observations. For example the predictions after spurious observation at \( t = 50 \) remain too high for a considerable period of time before aligning with the data again. A similar behavior can be noticed at time points subsequent to the other discrepant observation.

The robust filter was run assuming \( \hat{x}_0 = 17, \hat{P}_0 = 0.009, \hat{\sigma}^2_\varepsilon = 0.071, \hat{\sigma}^2_\gamma = 1.75, \pi_0 = 0.9, \pi_1 = 0.1 \). The modeling results show that this filter is robust to observational outliers. For example, the prediction at \( t = 50 \) and at the subsequent points do not seem to be affected by the aberrant observation at \( t = 50 \) unlike in the previous case.

In addition, the robust filter also provides the posterior probability \( R_t \) of any observation being an outlier.

4.2 ARIMA(1,1,0) model with outliers

Consider the model

\[ (1 - \varphi B)(1 - B)y_t = \alpha_t, \]  

where \( \varphi \) is the autoregressive parameter and \( \{\alpha_t\} \) is a white noise sequence with mean zero and variance \( \sigma^2_\alpha \). Let the measurement process be defined by

\[ z_t = y_t + \varepsilon_t, \]  

where \( \{\varepsilon_t\} \) is a sequence of i.i.d. Gaussian random variables with mean zero and variance \( \sigma^2_\varepsilon \). For \( \varphi > 0 \) this model represents a very slowly drifting process. Predictions
for any lead time obtained at a particular time point \( t = \tau \) in this series is a linear function of \( z_\tau \) and \( z_{\tau - 1} \). Thus, if there is an observational outlier at \( \tau \) it strongly influences the predictions. Thus robust Kalman Filtering and outlier detection are quite relevant in this case.

We shall now incorporate occasional outliers by replacing \( \varepsilon_t \) by \((1 - i_t)\varepsilon_t + i_t\gamma_t\), where \( \{\gamma_t\} \) and \( \{i_t\} \) are as defined before. A state space equivalent of this ARIMA(1,1,0) observational outlier model can be written as

\[
x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & \varphi \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \alpha_t, \quad (4.6)
\]

\[
z_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + (1 - i_t)\varepsilon_t + i_t\gamma_t,
\]

where \( x_t \) is a \((2 \times 1)\) state vector. Then the state and covariance predictions at time \( t \) can be given by

\[
\hat{x}_{t|t-1} = \begin{pmatrix} 1 & 0 \\ 1 & \varphi \end{pmatrix} (\hat{x}_{t-1} - \varphi \hat{x}_{2t-1})^T,
\]

\[
\hat{P}_{t|t-1} = \begin{pmatrix} 1 & 0 \\ 1 & \varphi \end{pmatrix} \hat{P}_{t-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \sigma^2_{\alpha},
\]

\[
\hat{z}_{t|t-1} = \begin{pmatrix} 0 & 1 \end{pmatrix} \hat{x}_{t|t-1} = \hat{x}_{1t-1} + \varphi\hat{x}_{2t-1}.
\]

A simulation study was conducted using 100 observations from the model (4.4) with \( \varphi = 0.8 \), \( \sigma_{\alpha} = 1 \), \( \sigma_\varepsilon = 5 \). As in the previous example the outliers were introduced at the points \( t = 25, 50, 65, 75 \) using a Gaussian distribution with \( \sigma_\gamma = 25 \). The initial values of the states which were used to generate the series were \( \hat{x}_{10} = 20 \) and \( \hat{x}_{20} = 150 \) respectively. The ordinary Kalman filter was initially used to estimate the state and to make predictions. The aberrant observations seem to have a strong influence on the predictions. Secondly, the robust Kalman filter was run on the assumption that \( P\{i_t = 1\} = 0.1 \). Unlike before, the predictions are not influenced in this case. Again the four large outliers have been clearly detected.

References


