Outliers Rejection in Kalman Filtering—Some New Observations

Z. Berman

Berman Consulting and Training Ltd., Israel

email: berman.consult@gmail.com

***Abstract*— A standard outlier-rejection scheme applied in Kalman filtering, based on the acceptance/rejection gate for measurement innovation, is discussed in this paper. The main idea behind this approach is based on assumptions that measurements can be "normal", as described in the measurement model and "abnormal" outliers that are generated by a totally different model. The goal of the acceptance/rejection gate is to accept normal measurements and reject abnormal ones. A concrete and simple case of range estimation in the presence of multipath outliers is thoroughly analyzed. The results are both nontrivial (even surprising) and important for designers of such rejection schemes who may use them as guidance for efficient design. The first observation is that the outlier-rejection scheme may provide worse results than the scheme with no rejection at all. This is because there is a positive, albeit relatively low probability that the system will enter and remain in a mode in which outliers are accepted and normal measurements are rejected. In this case, the estimation errors become very big and have a significant influence on the total standard deviations (even if their probability of occurrence is low). The main and very important conclusion is that outlier-rejection schemes cannot be applied without a proper recovery scheme that prevents the system from remaining "stuck" in normal-measurement rejection mode. In this paper, three different recovery schemes are proposed:**

- **a one-sided rejection scheme (only applicable to multipath-type outliers)**
- **a Kalman-filter reset scheme**
- **a set of parallel Kalman filters, where the output is provided by the filter with minimal innovation size.**

The design and performance analysis of each recovery scheme are described. The conclusion is that the performance of the recovery schemes is very close to the case without any outliers at all, up to very high (0.45) multipath-occurrence probability.

Keywords—Kalman filter, outlier, measurement rejection

I. The Outlier Rejection Scheme—An Overview

In today's advanced navigation systems, proper sensor integration seems to be the key to success. Traditionally, Kalman filters developed for linear, stochastic systems were used for sensor integration. In practice, however, sensors can produce unexpected anomalies in their measurements, for example, in the GNSS (due to sporadic interference or significant multipath) magnetometers affected by environmental magnetic variance, false fixes from feature-matching systems and many others. The common engineering procedure used to overcome such phenomena is outlier rejection (also called innovation filtering, spike filtering, measurement gating, reasonability testing or prefiltering; for overview and details see Chapter 17 in [1]). The main idea for outlier rejection is to define an acceptance/rejection gate for every measurement. The goal is to reject measurements created by outliers and accept normal measurements (created by the assumed nominal measurement model). In this section, the outlier-rejection scheme is presented in the context of Kalman-filter implementation.

Every Kalman filter is based on two equations: propagation (1) and measurement (2):

$$X(k+1) = A(k)X(k) + w(k) \tag{1}$$

$$z(k) = H(k)X(k) + v(k) \tag{2}$$

where $X(k)$ is the state vector, $A(k)$ is the transition matrix, $w(k)$ is the process noise, $z(k)$ is the measurement, $H(k)$ is the measurement matrix, and $v(k)$ is the measurement noise. In 1960, Rudolf Kalman proposed a recursive; optimal estimation algorithm for this problem (with several additional assumptions) called the Kalman filter [2]. This algorithm has been well documented (see, for example, [3],[4],[5]).

Aside from the model described in (1) and (2), the Kalman filter requires three additional statistical quantities: P_0 – the covariance of the state vector $X(0)$, $Q(k)$ – the covariance of the process noise $w(k)$ and $R(k)$ – the covariance of the measurement noise $v(k)$. In addition, zero mean, independence and Gaussian distribution are assumed for all random variables.

The equations of the Kalman filter are described below (assuming that the propagation rate and measurement rate are not necessarily equal).

Initialization

$$\begin{aligned} X^e(0) &= 0 \\ P(0) &= P_0 \end{aligned} \tag{3}$$

Propagation

$$X^e(k+1) = A(k)X^e(k) \tag{4}$$

$$P(k+1) = A(k)P(k)A^T(k) + Q(k) \tag{5}$$

Measurement

If the measurement step is true:

$$\xi(k) = z(k) - H(k)X^e(k) \tag{6}$$

$$S(k) = H(k)P(k)H^T(k) + R(k) \quad (7)$$

$$K(k) = P(k)H^T(k)\left(S(k)\right)^{-1} \quad (8)$$

$$X^{e+}(k) = X^e(k) + K(k)\xi(k) \quad (9)$$

$$P^+(k) = \left(I - K(k)H(k)\right)P(k)\cdot$$
$$\cdot\left(I - K(k)H(k)\right)^T + K(k)R(k)K^T(k) \quad (10)$$

$$X^e(k) = X^{e+}(k) \quad (11)$$

$$P(k) = P^+(k) \quad (12)$$

Else

Go to propagation

End

The quantity $\xi(k)$ is called innovation or residual and plays a crucial role in the outlier -rejection scheme. Since $z(k) = H(k)X(k) + v(k)$, where $X(k)$ is the true value of the state vector, and $v(k)$ is the measurement noise with covariance matrix $R(k)$, it is easy to verify that $S(k)$ is the innovation covariance. Moreover, if all assumptions for the Kalman filter are fulfilled, the innovation has zero mean and Gaussian distribution.

Now, let us assume that some outliers in the measurement equation can occur; this effect is usually modeled as follows:

$$z(k) = H(k)X(k) + v(k) + I(k)m(k) \quad (13)$$

where $I(k)$ describes the probability of outlier appearance, and $m(k)$ is its size. The outlier-appearance probability is usually low, but its size $m(k)$ is large. The idea behind outlier rejection is to reject measurements with $I(k)=1$ (outliers) and accept measurements with $I(k)=0$ (normal measurements). Detection of the hypothesis $I(k)=0$ is based on the observation that then, the innovation $\xi(k)$ has a Gaussian distribution with zero mean and covariance $S(k)$. The standard condition to detect this hypothesis is:

$$\left|\xi(k)\right| \le \kappa\sqrt{S(k)} \quad (14)$$

where with $\kappa = 2.5, 3$, condition (14) successfully detects 98.7%, 99.7% of the cases with $I(k)=0$; the loss of a few percent of valid measurements usually has a negligible effect on performance. Therefore, to obtain the standard outlier-rejection scheme, the following line should be added, between (7) and (8):

If $\left|\xi(k)\right| > \kappa\sqrt{S(k)}$, go to propagation (15)

At first glance this appears to work well: for $\left|\xi(k)\right| > \kappa\sqrt{S(k)}$, the measurement is rejected, if $I(k)=0$, the influence on performance is minor, and if $I(k)=1$, appropriate action is taken. For $\left|\xi(k)\right| \le \kappa\sqrt{S(k)}$, the measurement is accepted, if $I(k)=0$ appropriate action is taken, and if $I(k)=1$, the outlier size is probably small, so no severe damage is expected. This type of reasoning gives rise to a decrease in the value of κ such that the rejection rate of valid measurements is acceptable, but the acceptance rate of actual outliers ($I(k)=1$) is decreased.

However, it turns out that something very important is missing in this reasoning. This paper deals with this missing aspect, and the patient reader will encounter this absent consideration in Section III. Ways to correct the outlier-rejection scheme are presented in Section IV. To demonstrate this matter, an example case study is thoroughly analyzed

II. THE CASE STUDY—DESCRIPTION

The purpose of this section is to present the performance results obtained when a standard outlier-rejection scheme is applied to a case study. The case study involves a range measurement with multipath effect. The goal is to estimate the receiver bias. The state variables are the receiver biases random constant and Markov process.

The propagation system model is as follows:

$$X(k+1) = \begin{bmatrix} 1 & 0 \\ 0 & \alpha_m \end{bmatrix} X(k) + \begin{bmatrix} 0 \\ w(k) \end{bmatrix} \quad (16)$$

where $X_1(k)$ is a random constant with 1 m standard deviation, and $X_2(k)$ is a Gaussian Markov process with 0.3 m standard deviation and 30 s time constant. The corresponding system-modeling parameters are given by:

$$\alpha_m = \exp\left(-\frac{dt}{\tau_m}\right) \text{ and } \sigma_w = \sigma_m\sqrt{1-\alpha_m^{\,2}}$$

where dt is integration time, τ_m is the Markov process time constant, σ_w is the standard deviation of the process noise $w(k)$, and σ_m is the standard deviation of the Markov process.

The measurement model is given by:

$$z(k) = HX(k) + v(k) + I(k)m(k) \quad (17)$$

with $H = [1 \ \ 1]$; $v(k)$ is the Gaussian measurement noise with 0.3 m standard deviation. $I(m)m(k)$ is the multipath effect, modeled as follows:

$I(m)$ is a binary random variable such that $\text{Prob}\{I(k)=1\} = p, \text{Prob}\{I(k)=0\} = 1-p$ and $m(k)$ is uniformly distributed on the interval $\left[0, m\right]$. The multipath probability of occurrence is treated here as a parameter in the range [0÷0.45]; the maximal multipath value is set to $m=3$; the test duration is 300 s, and the measurement rate is 1 Hz. All random variables are independent.

In practical situations, the multipath model is not known, and therefore it is not used as a design parameter, and is only used for system-performance evaluations.

The Kalman filter is designed for this scheme, assuming no multipath effect. Then an outlier-rejection procedure is added, as described in section I. Notice that in principle, the outlier-rejection scheme converts a linear Kalman filter to a nonlinear algorithm, so covariance analysis is not applicable. The analysis is carried out by the Monte Carlo method (3,000 runs per result), where the estimated range error is defined as the sum of the estimated biases: random constant and Markov process, namely:

$$\delta X(k)=\left(X_1^e(k)+X_2^e(k)+\right)-\left(X_1(k)+X_2(k)\right) \qquad (18)$$

where $X_1^e(k), X_2^e(k)$ are the estimated values. Table I presents the results of this analysis with two parameters: p – the probability of outlier appearance and κ – treated here as a design parameter for the outlier-rejection scheme. The far-right column describes results of a standard Kalman filter, without any rejection scheme $(\kappa=\infty)$. The rms value is used instead of the standard deviation, because due to nonlinearity of the applied algorithm, the mean value is not necessarily zero.

TABLE I
RMS ESTIMATION ERRORS FOR A STANDARD OUTLIER-REJECTION SCHEME

p	rms range error (m)				
	$\kappa=2$	$\kappa=2.5$	$\kappa=3$	$\kappa=4$	$\kappa=\infty$
0	0.21[2]	0.13	0.13	0.13	0.13
0.05	0.25[2]	0.19[2]	0.16[2]	0.13[1]	0.15
0.1	0.30[2]	0.23[2]	0.18	0.14[1]	0.20
0.15	0.36[2]	0.27[2]	0.19	0.15[1]	0.26
0.20	0.41[2]	0.32	0.22	0.17[1]	0.32
0.25	0.42[2]	0.33	0.26	0.19[1]	0.39
0.30	0.46[2]	0.38	0.30	0.21[1]	0.46
0.35	0.49	0.43	0.32	0.25[1]	0.53
0.40	0.54	0.47	0.35	0.28[1]	0.61
0.45	0.59	0.52	0.39	0.33	0.68

The results provided in Table I are undoubtedly surprising. First, notice that a standard rejection scheme with $\kappa=2.5$ sometimes provides results that are inferior to those obtained without any rejection scheme (see the orange cells, superscript 2). Second, notice that the best results in Table I (green cell, superscript 1) are for high κ values $(\kappa=4)$ and not, as expected, for the low κ values. Finally, note that performance degradation, even for the best case, is highly significant with respect to the value 0.13 m, obtained without any multipath effect. In section III, we explain these effects.

III. DYNAMIC, NONLINEAR EFFECTS DURING OUTLIER REJECTION

Proper analysis of the problem described in section II begins with proper interpretation of the nonlinear properties of the algorithm. It is well known that the output of a nonlinear algorithm driven by Gaussian-distributed inputs is not necessarily Gaussian-distributed. So the first and second moment analysis, based on mean, standard deviation, or rms values, can be misleading and should not be used. Fig. 1 describes the histogram of the range error for a typical Monte Carlo set of runs.

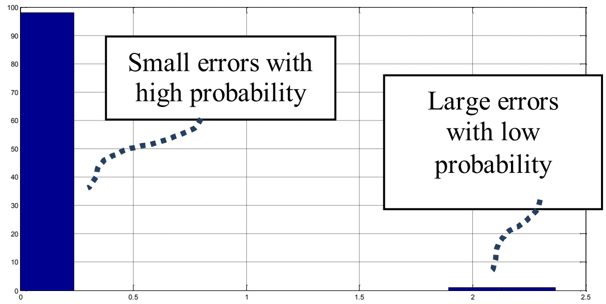

Fig. 1. Typical range error histogram.

Our problem can be described as follows:

- Usually (with relatively high probability), the system works well (very small estimation errors).
- Sometimes (with low probability), the system works very badly (very big estimation errors).

Now, the goal is to explain how those rare but bad cases are created.

Consider the acceptance/rejection gate:

$|\xi(k)|\le\kappa\sqrt{S(k)}$

For normal measurements it is given by:

$|\xi(k)|=\left|H(k)X(k)+v(k)-H(k)XH^e(k)\right|\le\kappa\sqrt{S(k)}$

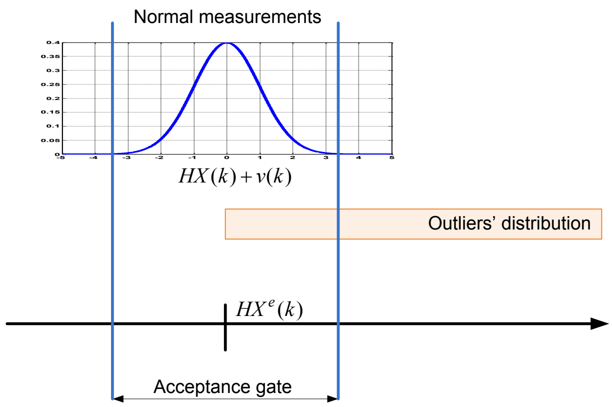

Fig. 2. Visualization of the outlier-rejection scheme—normal case.

The acceptance gate is around $HX^e(k)$ and as long as it is close to $HX(k)+v(k)$, the case is normal, most of the normal measurements are accepted, and most of the outliers are rejected (Fig. 2). But now, let us assume that several positive outliers were accepted, and increased the value of $HX^e(k)$, which is the center of the acceptance gate. Fig. 3 describes this situation.

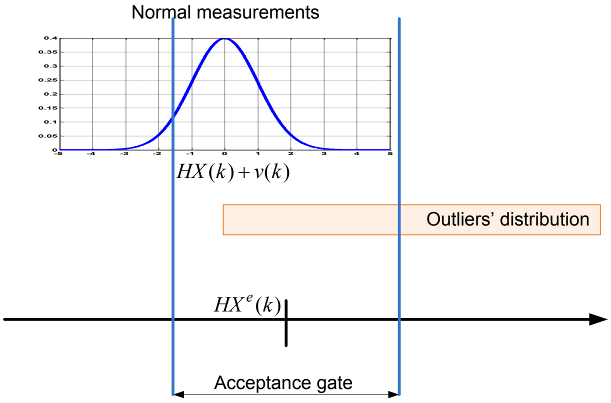

Fig. 3. Visualization of the outlier rejection scheme—intermediate case.

For this intermediate case, two effects may further increase $HX^e(k)$. The first is that more outliers are accepted, all of them positive. The second is that part of the negative tail of the normal measurements is now beyond the acceptance gate, so some measurements related to negative errors will be rejected, while all measurements related to positive errors will be accepted, causing further positive movement of $HX^e(k)$. Eventually, in some rare instances, $HX^e(k)$ can move so far away that the underlying acceptance gate will not accept normal measurements (!). Fig. 4 describes this situation. Notice that the "big" estimation errors of 2 m, as described in Fig. 1, indeed cause the acceptance gate to move beyond the normal measurement distribution.

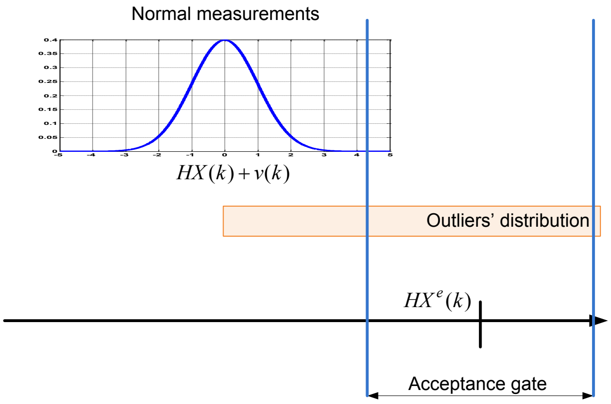

Fig. 4. Visualization of the outlier rejection scheme—fail mode.

In the case described in Fig. 4, the center of the acceptance gate is so far away from the center of the normal measurement distribution $HX(k)+v(k)$ that any normal measurement cannot be accepted. The Kalman-filter measurements will use only outliers. The real problem with this case is that if it does occur, the system cannot recover from the situation.

Notice that this effect can be described in very general terms. The following effects are needed to obtain the fail case:

- Some accepted outliers move the gate center $HX^e(k)$ significantly in one direction.
- Non-symmetry of the acceptance gate with respect to normal measurement distribution moves the acceptance gate center further in the same direction. This effect can be called positive feedback or unstable mode in the outlier-rejection scheme. This process may end with the fail mode—the center of the acceptance gate lies beyond the normal measurement distribution.

The author's conjecture is that the fail mode has a non-zero, positive appearance probability as long as normal measurements and outlier distributions overlap. This is the case for all known, to the author, outlier-rejection problems. Of course, there may be some practical cases in which the probability of going into fail mode is very low. Since the stochastic model for outliers is usually unknown, this probability cannot be analyzed in practice.

It turns out that after understanding the fail-mode mechanism, several recovery algorithms can be addressed. The next section describes a few of them, tested and verified for the case study presented herein.

IV. Fail-Mode Recovery Algorithms

In this section three recovery algorithms are proposed:

- One-sited gate
- System reset
- Multi-filter configuration

It should be noted that a variety of recovery algorithms can be proposed, but most of them are modifications of the proposed ones.

The one-sited gate approach is based on the fact that multipath error can only be positive. The acceptance/rejection criteria are similar to criterion (14), but instead of using $|\xi(k)|$, the value $\xi(k)$ is used:

$$\xi(k) \le \kappa\sqrt{S(k)} \tag{19}$$

In this way, it is impossible to obtain a case in which all normal measurements are rejected. Table II describes the rms range errors of the outlier-rejection scheme after this modification.

TABLE II
The RMS Estimation Errors for One-Sited Outlier-Rejection Scheme

p	rms range error (m)				
	$\kappa=2$	$\kappa=2.5$	$\kappa=3$	$\kappa=4$	$\kappa=\infty$
0	0.19	0.13	0.13	0.13	0.13
0.05	0.19	0.13	0.13	0.13	0.15
0.1	0.19	0.13	0.13	0.13	0.20
0.15	0.19	0.13	0.13	0.14	0.26
0.20	0.19	0.13	0.14	0.15	0.32
0.25	0.19	0.13	0.14	0.17	0.39
0.30	0.19	0.14	0.15	0.19	0.46
0.35	0.19	0.15	0.16	0.21	0.53

0.40	0.19	0.15	0.18	0.24	0.61
0.45	0.20	0.16	0.20	0.28	0.68

Table II shows significantly improved performance relative to the standard rejection scheme presented in Table I. The lowest errors are achieved for $\kappa = 2.5$; the most important achievement is that even for a high probability of outlier appearance, the performance of the one-sited rejection scheme is very similar to that without any outliers at all (see 0.13 m error for $p = 0$).

The sign of the outliers is not always known or constant. In this case, a different recovery scheme should be applied. A recovery scheme that does not assume any knowledge about the outlier sign is based on the following two principles:

- Detect the fail mode by detecting high rejection rate
- Recover from this mode by Kalman-filter re-initialization (system reset)

This algorithm requires some tuning, to set the threshold and window for the rejection-rate measurement. The results presented in Table III were obtained for the following application. Two counters are defined: one for continuous rejections (after every acceptance this counter is reset) termed C_{cont}, and a total counter that counts rejection during a 20-s window termed S_{cont}. System reset is defined as filter reinitialization: $X^e(k^*) = 0, P(k^*) = P_0$ where k^* is the reset cycle.

The condition for reset is defined by

$$C_{cont} \geq 4 \wedge S_{cont} > 1.3p \cdot 20 \wedge time < 240 \quad (20)$$

The reasoning behind this condition is as follows. It is not expected that four continuous rejections will be obtained for a normal case, even with the maximal $p = 0.45$. During 20 s, the expected rejection count is $20p$; 30% above this indicates that one has strayed from the normal case, at least to an intermediate state, or even to the fail mode. The last condition avoids system initialization during the last 60 s. This condition was added because for very late reset, the system does not have enough time to converge its estimations to proper values. Of course, one can propose different conditions that might work as well or perhaps even better. In any case, the recommended starting point for any recovery scheme for an outlier-rejection scheme is to use a continuous counter or total counter and add conditions for which the reset is no longer effective. As shown in Table III, this algorithm works very well for the case study.

TABLE III
THE RMS ESTIMATION ERRORS FOR RESET-TYPE RECOVERY

p	rms range error (m)				
	$\kappa = 2$	$\kappa = 2.5$	$\kappa = 3$	$\kappa = 4$	$\kappa = \infty$
0	0.19	0.13	0.13	0.13	0.13
0.05	0.19	0.13	0.13	0.13	0.15
0.1	0.19	0.13	0.13	0.13	0.20
0.15	0.19	0.13	0.13	0.14	0.26
0.20	0.19	0.13	0.14	0.15	0.32
0.25	0.19	0.14	0.14	0.17	0.39
0.30	0.20	0.14	0.15	0.19	0.46
0.35	0.20	0.15	0.16	0.21	0.53
0.40	0.20	0.16	0.18	0.24	0.61
0.45	0.20	0.17	0.20	0.28	0.68

The best performance is achieved for $\kappa = 2.5$ and up to $p = 0.2$, the range errors with a recovery scheme based on reset are equal to the case without any outliers at all. For a higher probability of outlier appearance, the performance degradation is minor.

The last recovery scheme presented here is based on a multi-model Kalman filter, sometimes called filter bank. The main idea is to run several Kalman filters with different assumptions and to use a weighted output with respect to residual sizes of the different filters. To design a recovery algorithm, five Kalman filters are applied with different initial conditions, and the output is taken from the filter with the lowest residual error. The five different initial conditions are as follows: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0.3 \end{bmatrix}, \begin{bmatrix} 1 \\ -0.3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -1 \\ -0.3 \end{bmatrix}$. During the last 20 s, the innovation squares are summed; the filter with the minimal sum of innovation squares is selected to provide the output. Table IV presents the results of this algorithm.

TABLE IV
THE RMS ESTIMATION ERRORS FOR FIVE PARALLEL KALMAN FILTERS

p	rms range error (m)				
	$\kappa = 2$	$\kappa = 2.5$	$\kappa = 3$	$\kappa = 4$	$\kappa = \infty$
0	0.14	0.14	0.14	0.14	0.13
0.05	0.14	0.14	0.14	0.14	0.15
0.1	0.14	0.14	0.15	0.15	0.20
0.15	0.14	0.15	0.15	0.17	0.26
0.20	0.14	0.16	0.16	0.19	0.32
0.25	0.15	0.16	0.18	0.21	0.39
0.30	0.15	0.17	0.19	0.23	0.46
0.35	0.16	0.18	0.21	0.26	0.53
0.40	0.17	0.20	0.23	0.30	0.61
0.45	0.18	0.21	0.25	0.34	0.68

This time, the best estimations are achieved for $\kappa = 2$; they are almost the same as for the previous two algorithms.

V. SUMMARY

This papers shows that a "naïve" outlier-rejection procedure can be very disappointing. The reason is that the system may get into a rare fail mode that is related to high rejection rate and very big estimation errors. The presented recovery schemes solved this problem and exhibited estimation errors that almost eliminated the effect of the outliers.

The system analyzed here was a very simple case study. The analysis results revealed that a Kalman filter with a rejection scheme is a nonlinear dynamic system that may exhibit very surprising behavior, at least relative to linear, Kalman filter-based systems. The issue is the risk of getting stuck in a mode where the "good" measurements are rejected. This situation can occur not only due to overlapping outlier and measurement-noise distributions, as described in Section III, but also due to temporal sensor failure (see discussion in Chapter 17 of [1]), or inappropriate sensor error models used for the Kalman filter that cause the Kalman-filter covariance to diverge from the actual error covariance.

It should be noted that outliers can appear in any practical sensor-integration scheme. The list of possible reasons is very long. We can therefore agree that an outlier-rejection scheme is a desirable feature for any sensor-integration scheme. The author believes that the conclusions presented here are highly relevant for every Kalman filter that integrates inertial sensors.

The most important lesson learned here is that every outlier-rejection scheme can go into fail mode and therefore, a recovery scheme is a necessary supplement. A recovery scheme based on reset is perhaps the most popular, effective, and simple algorithm. As mentioned earlier, temporal failure may cause outlier-like behavior, but with absolutely unknown stochastic and functional characteristics. A reset-based recovery scheme sets a limit for the maximal allowable rejection rate without any further assumptions about outlier characteristics. Therefore, it is useful under the most general, unexpected conditions. A one-sited recovery scheme and multi-model Kalman-filter recovery schemes may work very efficiently. For both of these schemes, some a priori knowledge about the outliers' characteristics is required.

Acknowledgment

I would like to thank Prof. Dorota A. Grejner Brzezinska for inspiring me to delve into the analysis of nonlinear effects in Kalman filtering. In addition, I use this opportunity to describe my appreciation to Prof. Jason Rife and Prof. Demoz Gebre-Egziabher for encouraging and inspiring discussions.

References

[1] P.D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, 2nd ed. Norwood, MA Artech House, 2013.

[2] R.E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[3] S. Maybeck, *Stochastic Models, Estimation, and Control*. New York, Academic Press, 1982.

[4] A. Gelb, *Applied Optimal Estimation*. Cambridge: M.I.T. Press, 1974.

[5] M.S. Grewal, and A.P. Andrews, *Kalman Filtering Theory and Practice Using MATLAB*, 3rd ed. New Yersey, Wiley, 2008.