Robust Kalman Filter Based on a Generalized Maximum-Likelihood-Type Estimator

Mital A. Gandhi, Student Member, IEEE, and Lamine Mili, Senior Member, IEEE

Abstract—A new robust Kalman filter is proposed that detects and bounds the influence of outliers in a discrete linear system, including those generated by thick-tailed noise distributions such as impulsive noise. Besides outliers induced in the process and observation noises, we consider in this paper a new type called structural outliers. For a filter to be able to counter the effect of these outliers, observation redundancy in the system is necessary. We have therefore developed a robust filter in a batch-mode regression form to process the observations and predictions together, making it very effective in suppressing multiple outliers. A key step in this filter is a new prewhitening method that incorporates a robust multivariate estimator of location and covariance. The other main step is the use of a generalized maximum likelihood-type (GM) estimator based on Schweppe’s proposal and the Huber function, which has a high statistical efficiency at the Gaussian distribution and a positive breakdown point in regression. The latter is defined as the largest fraction of contamination for which the estimator yields a finite maximum bias under contamination. This GM-estimator enables our filter to bound the influence of residual and position, where the former measures the effects of observation and innovation outliers and the latter assesses that of structural outliers. The estimator is solved via the iteratively reweighted least squares (IRLS) algorithm, in which the residuals are standardized utilizing robust weights and scale estimates. Finally, the state estimation error covariance matrix of the proposed GM-Kalman filter is derived from its influence function. Simulation results revealed that our filter compares favorably with the $\mathcal{H}_\infty$-filter in the presence of outliers.

Index Terms—Impulsive noise, Kalman filter, non-Gaussian filter, outliers, prewhitening, robust statistics.

I. INTRODUCTION

Based on the seminal papers by Kalman and Bucy [1], [2], the classical Kalman filter (KF) has been applied in various fields such as control systems, reliability engineering, and wireless communications; for example, see [3]–[13]. The system’s dynamics and observations are represented by a state space model, which contains the system process and observation noises [14] that affect the state predictions and observations, respectively. The KF is a maximum-likelihood estimate assuming a linear model, quadratic performance criterion, and Gaussian probability distribution for the system process and observation noises. In practical systems though, the assumed model is only an approximate one and the two types of noises may follow a thick-tailed, non-Gaussian probability distribution, inducing innovation and observation outliers. In addition, we introduce and consider a new type called structural outliers that are caused by errors in the state transition and observation matrices.

Innovation and observation outliers arise naturally in many areas of engineering. Examples of these occurrences include hardware discontinuities in digital control systems [15], faults in the sensors of a control system including target estimation and tracking in aerospace applications [16], [17], and co-channel interference and fading in wireless communications [18]–[21], just to name a few. On the other hand, structural outliers may occur due to modeling errors such as computational floating point errors in hardware or model mismatch induced by a time-varying bias in the model. In the presence of all these outliers, the KF may provide a strongly biased solution or even diverge [2].

To handle these outliers, several nonlinear methods have been proposed in the literature [22]–[40]. A review of the methods developed for signal processing applications up to 1985 can be found in the survey paper authored by Kassam and Poor [22]. In 1970, Bucy and Senne [23] initiated one of the earliest nonlinear filters to handle non-Gaussian noise. However, it assumes the noise probability distribution function (PDF) is known a priori and is computationally intensive with increasing order of state variables. In the mid-1970s, Masreliez and Martin [24], [25] pioneered the application of robust statistics to handle symmetric, $\epsilon$-contaminated Gaussian noise in the observations and system process by means of separate filters and stochastic approximation. For the particular case of observation outliers, various methods have been proposed in the literature, namely Christensen and Soliman’s filter based on the least absolute value criterion [26]; Doblinger’s adaptive KF scheme [27]; and Durovic, Durgaaprasad, and Kovacevic’s filters utilizing the maximum likelihood-type (M-)estimator based robust KF [28]–[30]. However, none of these methods iterate when solving the underlying nonlinear estimator at each recursion time step, effectively assuming that the predictions are accurate. Therefore, the filters may yield unreliable results when observation and innovation outliers occur simultaneously. Similarly, assuming the observations are accurate would also lead to erroneous estimates. Hence, a filter is needed that does not rely completely on either the predictions or observations; instead, it should process them simultaneously and solve the underlying estimator iteratively. Finally, the classical KF error covariance matrix has been inaccurately retained in these
filters. Indeed, it needs to be replaced by that of the underlying nonlinear estimator. The only exception is the method proposed by Durovic and Kovacevic [30], which uses the covariance matrix for M-estimators from Huber [31].

In contrast to methods dealing with arbitrarily large outliers, such as those discussed above, the H_{\infty}-filter, an approach stemming from robust control [32]–[39] minimizes the worst-case estimation error averaged over all samples by treating modeling errors and uncertainties as unknown-but-bounded noise. However, it breaks down in the presence of randomly occurring outliers of any type, whereas our generalized maximum-likelihood Kalman filter (GM-KF) is capable of directly suppressing them. In this respect, these two approaches are complementary to each other as will be demonstrated in Section V.

The Kalman filter proposed by Chan et al. [40] also takes a minimax approach on the norm of the residuals to provide robust estimates. The filter replaces the least squares with an M-estimator and uses semidefinite programming to derive the estimates, but it handles only unknown-but-bounded system uncertainties, not structural outliers. The filter does not withstand innovation outliers either, making it very vulnerable to all but observation outliers. In this paper, we propose a general regression framework for estimating the states of a dynamic system via robust Kalman filters. In this framework, any estimator whose covariance matrix can be calculated may be used to derive a filter. We now describe its three key steps, which are creating a redundant observation vector, performing robust prewhitening, and estimating the state vector via a robust estimator. Observation redundancy is required for an estimator to be capable of suppressing the outliers, and can be achieved in practice by simply placing more sensors in the system. To process simultaneous observations, we convert the classical recursive approach into a batch-mode regression form [41] in the first step of the GM-KF. When the signal is treated in this setting, observation and innovation outliers can be seen as vertical outliers and structural outliers as bad leverage points [41]. The latter is defined as one whose projection in the factor space is distant from the bulk of the data points. Because an M-estimator is not robust to bad leverage points, we use a GM-estimator instead [31], [41]–[43].

The second step consists of applying a prewhitening procedure that utilizes a robust estimator of location and covariance such as the projection statistics (PS). This procedure robustly uncorrelates the noise when outliers are present in the predictions and observations. In the third step, the unconstrained nonlinear optimization in the GM-estimator is solved using the iteratively re-weighted least squares (IRLS) algorithm, yielding robust estimates suppressing all types of outliers. Based on the work of Hampel et al. [43] and Ferholz [44], the relationship between the influence function of the GM-estimator and its asymptotic covariance matrix is established, allowing us to derive the asymptotic state estimation error covariance matrix of the GM-KF.

This paper is organized as follows. Section II contains background material on the nature of the outlying noise, robust statistics, and filtering techniques. The GM-KF is developed in Section III. The filter’s error covariance matrix at the correction step is derived in Section IV. In Section V, the filter is shown to have a positive breakdown point through simulations estimating the longitudinal dynamics of a helicopter. Conclusions are drawn in Section VI.

II. BACKGROUND

A. Outlier Description and Characteristics

In a general sense, outliers are observations that do not follow the pattern of the majority of the data. Two types of outliers have been considered in the literature. They are known under different names, as indicated in Table I. The first type occurs via observation noise \( \epsilon_k \), affecting the vector \( \tilde{z}_k \) expressed as

\[
\tilde{z}_k = H_k x_k + \epsilon_k.
\]

The second type affects the propagated state \( \tilde{x}_k \) and occurs via the process noise \( w_k \) in the system dynamics model given by

\[
\tilde{x}_k = F_{k-1} \tilde{x}_{k-1} + w_{k-1}.
\]

For consistency with the type of noises that causes the outlier, we call these observation and innovation outliers, respectively. It is noted that these names have been used in [46]. Furthermore, we also consider structural outliers generated by errors in \( H_k \) and \( F_{k-1} \) in (1) and (2).

Formally, we use an \( \varepsilon \)-contaminated model to induce a topological neighborhood around the target distribution \( F \), yielding a probability distribution \( G \) given by \( G = (1 - \varepsilon) F + \varepsilon H \) [31], where \( H \) is an unknown distribution for the outliers. Based on this model, we now discuss measures that quantify the robustness of an estimator, namely the influence function, maximum bias curve, and breakdown point.

B. Robust Statistical Theory

1) Influence Function of an Estimator: The foundation of modern robust statistics was initiated by Huber [31], which was later expanded with the approach based on influence functions by Hampel et al. [43]. For the regression problem given by (1), where \( \varepsilon \) is assumed to be deterministic and \( k \) is fixed, the asymptotic influence function, defined as the Gâteaux derivative [31], is given by

\[
IF(x, H_k; F) = \lim_{\varepsilon \to 0} \frac{T((1 - \varepsilon) F + \varepsilon H_k) - T(F)}{\varepsilon}
\]
where $h_{j}$ is the $j$th column vector of the matrix $H_{F}^{T}$. This function is a measure of the local sensitivity of an estimator $T(G)$ to an arbitrary infinitesimal contamination at $\varepsilon$, i.e., $H = \Delta_{\varepsilon}$. In the location case, it is the slope of the tangent of the bias curve at $\varepsilon = 0$ in the direction of $\Delta_{\varepsilon}$, whose supremum is termed gross-error sensitivity. The influence function of the GM-estimator plays a very important role in developing the GM-KF as it is essential in assessing its robustness and in deriving its error covariance matrix expressed as $\sum = E_{F}[H(\Delta_{\varepsilon})]^{T}$. The influence function of the GM-estimator is derived in Section IV along with this relationship following the approach of von Mises functionals by Hampel et al. [43] and Ferndholz [44].

2) Asymptotic Maximum Bias of an Estimator: While the influence function assesses the effects of infinitesimal fraction of contamination [43], the upper bound of the bias of an estimator for varying levels of contamination, $0 \leq \varepsilon < \varepsilon^{*}$, is indicated by the maximum bias curve. Indeed, for the $\varepsilon$-contaminated model given by $G = (1 - \varepsilon)F + \varepsilon H$, the maximum asymptotic bias of the estimator $\hat{\theta}$ in its functional form $T$ at any such $G$ is defined as $b_{\text{max}}(\varepsilon; F) = \sup_{H} |T(G) - \theta|$, where the supremum is taken over all $H$ and the estimator $T$ is assumed to be Fisher consistent at $F$, that is, $T(F) = \theta$ asymptotically. It is desired for an estimator’s maximum bias under contamination to be not much larger than the minimum possible maximum bias curve. For example, Huber [31] showed that the sample median attains the minimax bias in the location case. The associated maximum bias curve is derived in Rousseeuw and Croux [48].

3) Robustness Versus Statistical Efficiency: Besides the influence function and maximum bias curve, the breakdown point, $\varepsilon^{*}$, is another concept for assessing the robustness of an estimator. In general, both bounded influence and a positive breakdown point under contamination are very important for an estimator’s bias stability, where the latter is defined as the largest fraction of contamination $\varepsilon$ that an estimator can tolerate. The highest possible asymptotic value for $\varepsilon^{*}$ is 0.5 and is reached by the sample median in the location case. In contrast, the sample mean has a breakdown point $\varepsilon^{*} = 0$ since a single outlier can lead to an arbitrarily biased estimate. Under the assumption of general position, which requires every set of $n$ row vectors of the matrix $H$ in the linear regression model, $Z = H \beta + \epsilon$ to be linearly independent, the maximum fraction of outliers that any regression equivariant estimator can handle with $m_{i}$ observations and $m$ state variables is given by $\varepsilon^{*}_{\text{max}} = [(m_{i} - n)/2]/m_{i}$, where $[\cdot]$ denotes the integer part [47]. Clearly, using more observations will increase this upper bound, which asymptotically tends to 50%. In contrast to these concepts of breakdown point and influence function from robustness theory is the Fisherian concept of statistical efficiency at a given probability distribution [49]. An asymptotically efficient estimator is one that attains the Cramér–Rao lower bound, the theoretical lower bound of the variance that any unbiased estimator may have at that probability distribution. A tradeoff exists between the breakdown point of an estimator and its statistical efficiency; for example, the sample median attains the highest possible breakdown point, but has a statistical efficiency of only 64% at the Gaussian distribution as shown by Hampel et al. [43].

This robustness–efficiency tradeoff needs to be properly addressed when designing the solution to an estimation problem. This is exactly what we have done with the design of the GM-KF. First of all, the GM-estimator has a breakdown point of up to 35% in linear regression for a known error covariance matrix [14]. Second, redundancy is leveraged in the filter by combining the predictions with the current observations to form the larger data set, which further increases the filter’s breakdown point. Finally, unlike the Mallows-type estimators that downweight both good and bad leverage points, we use the Schweppe-type method that downweights only the latter, and therefore, attains a 95% statistical efficiency for the Huber $\rho$-function with $c = 1.57$. As shown in Section V-A, this high efficiency at the Gaussian distribution has been verified through simulations. Thus, the Schweppe-type Huber GM-estimator’s bounded influence to residuals and position, along with a positive breakdown point that is enhanced by a redundant measurement vector, allows the GM-KF to produce more reliable state estimates in the presence of outliers while also providing high statistical efficiency at the Gaussian distribution.

C. Review of Filtering Techniques

1) Classical Kalman Filter: A discrete linear Gauss–Markov system can be described by means of a dynamic state equation and an observation equation with conditions on noise and initial values [4]–[6]. Let the state of this system be a stochastic vector $x_{k}\in \mathbb{R}^{n \times 1}$. At every time $k$, suppose that $x_{k}$ is observed via an observation vector $z_{k}\in \mathbb{R}^{m \times 1}$. Let the dynamics of $x_{k}$ and $z_{k}$ be described for $k\in \mathbb{N}$ by (1) and (2) where $w_{k}\in \mathbb{R}^{n \times 1}$ is the system model error vector; $\epsilon_{k}\in \mathbb{R}^{m \times 1}$ is the observation error vector; $F_{k}\in \mathbb{R}^{n \times n}$ is the state transition matrix; and $H_{k}\in \mathbb{R}^{m \times n}$ is the observation matrix at time $k$. The system and observation errors are generally assumed to be zero-mean, Gaussian, white, and uncorrelated with each other, that is, $w_{k} \sim N(0, W_{k})$ and $\epsilon_{k} \sim N(0, R_{k})$ where $W_{k}$ and $R_{k}$ are given covariance matrices. The objective is to infer knowledge about the state vector given the model and observations until time $k$, where the initial state vector, $x_{0}$, is assumed to be a random vector normally distributed as $N(\bar{x}_{0}, \Sigma_{0})$. The equations of the recursive KF are derived as

\begin{align}
\hat{x}_{k|k-1} & = F_{k-1}\hat{x}_{k-1|k-1} \\
\sum_{k|k-1} & = F_{k-1}\sum_{k-1|k-1}F_{k-1}^{T} + W_{k-1} \\
K_{k} & = \sum_{k|k-1}H_{k}^{T}(H_{k}\sum_{k|k-1}H_{k}^{T} + R_{k})^{-1} \\
\hat{x}_{k|k} & = \hat{x}_{k|k-1} + K_{k}(z_{k} - H_{k}\hat{x}_{k|k-1}) \\
\sum_{k|k} & = \sum_{k|k-1} - K_{k}H_{k}\sum_{k|k-1}.
\end{align}

(4) (5) (6) (7) (8)

Its limitations are actually well-known [4]–[6]. First, it assumes that the true system model is known with certainty; however, structural errors are likely in practice and may degrade the filter’s performance. Second, being a linear estimator, computation is very fast and recursive. However, the filter has a breakdown point of zero since a single outlier from any of the three noise sources can carry the filter’s bias over all bounds. The $H_{\infty}$-filter discussed in the next section is intended to overcome some of these limitations.
2) $H_{\infty}$-Filtering: The following brief description of the $H_{\infty}$-filter follows the work of Simon [32]. The reader is referred to [32]–[39] for further details. Research on the technique was first introduced in the frequency domain by Grimble [36]. The basic premise underlying this filter is to minimize the worst-case estimation error in a system. Conceptually, the method can be seen as a multiple-input-multiple-output linear time-invariant (LTI) filter that is characterized by an $m \times n$ matrix $P$ of transfer functions, where the component $P_{ij}$ is a transfer function relating the $i$th output to the $j$th input [50].

Using this matrix, the LTI system’s input and output power spectra $S$ are related as $S(f_{\text{input}}) = |P(f)|^2 S(f_{\text{output}})$, where $|P(f)|^2$ means that each element of the matrix $P$ is squared. Considering the processes $Y_k$ and $\hat{X}_k$ as inputs driving the state space model given in (1) and (2), the objective function for the $H_{\infty}$-filter, given by

$$J_H = \frac{\sum_{k=0}^{N-1} ||X_k - \hat{X}_k||^2_\bar{P}_k}{||X_0 - \hat{X}_0||^2_\bar{P}_0 + \sum_{k=0}^{N-1} (||Y_k||^2_{\bar{L}_k+1} + ||\hat{X}_k||^2_{\bar{L}_k}^{-1})}$$

(9)

can then be seen as limiting the transfer function norm induced from the exogenous signals at the input to the estimation error at the output. In other words, the filter provides a way to limit the frequency response of the estimator through the transfer function matrix [32]. Note that $\bar{P}_P$, $\bar{W}_P$, $\bar{L}_k$, and $B_k$ in (9) are symmetric, positive-definite matrices chosen by the designer [32].

A game theoretic strategy along with the method of dynamic constrained optimization using Lagrange multipliers is employed to design the filter. The first step is to find the optimizing values of $Y_k$ and $\hat{X}_0$ that maximize the objective function, subject to the constraint of the model equations. Using dynamic constrained optimization, the optimal values are expressed as $\hat{X}_k^* = W_k^* \hat{X}_0^*$ and $\hat{X}_0^* = \hat{X}_0 + \rho_0 \hat{X}_0^*$. Given that $W_k$ and $\hat{X}_0$ have been set to their optimized values, the next step is to find the stationary points with respect to $\hat{X}_k$ and $\hat{X}_0$. In this manner, the optimal solution to the filter is given as follows [32]:

$$K_k = P_k [I - \gamma K_k P_k + H_k^T \bar{R}_k^{-1} H_k P_k^{-1}]^{-1} H_k^T \bar{R}_k^{-1}$$

(10)

$$\hat{X}_{k+1} = F_k \hat{X}_k + E_k F_k K_k (\hat{X}_0 - H_k \hat{X}_k)$$

(11)

$$P_{k+1} = F_k P_k [I - \gamma K_k P_k + H_k^T \bar{R}_k^{-1} H_k P_k^{-1}]^{-1} F_k^T + W_k$$

(12)

An advantage of the filter lies in its ability to handle unknown-but-bounded system model uncertainty. Simon [32] has shown that the classical Kalman filter outperforms the $H_{\infty}$-filter when the noise follows assumptions; however, the latter may perform better, for example, when a constant bias exists in the mean of the noise, which is a violation of one of the standard Kalman filter assumptions. The $H_{\infty}$-filter’s design matrices lead to the optimizing solution, but a priori knowledge of the magnitude of the process noise $W_k$, observation noise $\epsilon_k$, and the initial estimation error in the application is required to properly choose them [32]. The filter’s sensitivity to the selection of these matrices can be a weakness in this approach. Particularly, the filter breaks down in the presence of outliers since the design matrices of the $H_{\infty}$-filter, just like the covariance matrix in the classical KF, cannot accommodate well the outliers induced by the thick tails of a noise distribution. It will be seen later in this paper that the $H_{\infty}$-filter performs very poorly in the presence of just one observation or innovation outlier. Thus, as will be shown in Section V, the filter is found to be complementary to the GM-KF in handling unknown deterministic bias or minimizing the average worst-case estimation error.

III. DEVELOPING THE ROBUST GM-KALMAN FILTER

A. Linear Regression Framework

We now develop the proposed GM-KF. Fig. 1 depicts the key steps of the method. Beginning with a prediction or initial state vector, we first convert the linear dynamic system in (1) and (2) to a batch mode linear regression form so that multiple observations can be processed simultaneously, giving a higher redundancy and breakdown point at each time step. Using a relation between the true state and its prediction, namely, $\hat{X}_{k|k-1} = \hat{X}_k + \hat{X}_{k|k-1}$, we obtain

$$\hat{X}_{k|k-1} = Hi_k \hat{x}_k + [\epsilon_k]$$

(13)

where $\hat{X}_{k|k-1}$ is the error between the true state and its prediction, and $I$ is the identity matrix. This batch-form linear regression is expressed in a compact form as

$$\tilde{y}_k = \tilde{Y}_k \hat{x}_k + \epsilon_k.$$  

(14)

The covariance matrix of the error $\tilde{y}_k$ is given by

$$\tilde{Y}_k = \begin{bmatrix} R_k & 0 \\ 0 & \sum_{k=1}^{N-1} \end{bmatrix} = \bar{L}_k \bar{L}_k^T$$

(15)

where the term $\bar{L}_k$ may be obtained by Cholesky decomposition, $\tilde{Y}_k$ is assumed to be a known noise covariance of $\epsilon_k$, and $\sum_{k=1}^{N-1}$ is the filter error propagation due to prediction given by (5).

Note that the incoming observations are assumed to be synchronized in time to realize the maximum benefits of redundancy, especially if the state is highly varying. If this is not the case, the delayed observations quickly become outlying data.
compared to the rest of the observations and predictions, which are recognized by the GM-KF and suppressed accordingly.

B. Robust Prewhitening

It is now desired to uncorrelate the data in the linear regression. However, applying the matrix \( L_k \) given by (15) in the prewhitening step with outliers present in the data would lead to negative effects on the data [51], [52]. For example, Fig. 2 depicts several points around (0, 0) sampled from a Gaussian distribution that is contaminated by a group of outliers around (15, 15). Fig. 3 shows that prewhitening has artificially shifted the outliers closer to the point cloud, resulting in a higher uncertainty in the distance measures.

Clearly, we need to detect and handle the outliers in the regression first. One classical outlier detection method is the Mahalanobis distance, which utilizes the nonrobust sample mean and sample standard deviation as estimators of location and scale. In this work, we use the robust PS estimator [53], [54], which employs instead the sample median and median-absolute-deviation of the data points \( \tilde{u} \) on the direction of all possible vectors \( \tilde{u} \) as robust estimates of location and scale, respectively. Formally, it is expressed as

\[
PS_i = \max_{||u||=1} \frac{|\tilde{u}^T u - \text{med}_j(\tilde{u}^T u)|}{1.4826 \text{med}_k |\tilde{u}^T u - \text{med}_j(\tilde{u}^T u)|},
\]

(16)

A large fraction of outliers can be handled due to the robust estimates in the expression above. Indeed, given that the general position assumption is satisfied, the breakdown point [55] of the PS attains the maximum, \( \varepsilon_{\text{PS max}} = [(m_k - n - 1)/2]/m_k \).

But, because it is not practical to consider every single vector \( u \) as described above, Gasko and Donoho [56] proposed to investigate only those directions originating from the coordinate-wise median of the point cloud and passing through each of the data points, yielding a total of \( m_k \) directions to be examined. The PS value is then the worst one-dimensional projection of the point’s distance to the cloud. Computing the distances using this algorithm is very fast, even in high dimensions, but with the loss of affine equivariance [41], [57]. Yet, this is not a shortcoming in our application as we use the distances as a diagnostic tool to identify the vertical outliers and leverage points before any transformation is applied to the data.

One may apply the PS algorithm at each time step \( k \) to the matrix \( \tilde{H}_k \) in (14) but this would only detect structural outliers. Instead, the PS values are computed using the vector \( \tilde{z}_k \) because it captures the effects of all three types of outliers. Actually, the PS algorithm requires that the data points to be multidimensional; hence, we compute the PS values using a matrix \( Z \) of dimension \( (m + n) \times 2 \) that contains the column vectors \( \tilde{z}_k \) and \( \tilde{z}_k - 1 \). The \( j \)th point is considered an outlier if \( PS^2_j > \tilde{b} \), where \( PS^2_j \) is the statistic of the \( j \)th element of the vector \( \tilde{z}_k \) and \( \tilde{b} \) is set equal to \( \chi^2_{2m_k} \), where \( d = 2 \). The motivation for this threshold stems from the fact that \( PS^2 \) follows roughly the \( \chi^2 \) distribution, as shown by Rousseeuw and Van Zomeren [58], when the underlying data points follow the Gaussian distribution and \( m_k/n \geq 5 \). The latter condition of redundancy in \( Z \) may be readily satisfied with \( m_k = m + n \) observations and only \( n = 2 \) dimensions. Finally, the PS is also able to avoid masking effects caused by multiple simultaneous outliers, as demonstrated in Fig. 2 by the very tight inner covariance ellipse.

After computing the PS values for the elements of \( \tilde{z}_k \), the outliers must now be downweighted using a weight function [42] given by \( \omega_i = \min(1, d^2/PS^2_i) \), where we pick \( d = 1.5 \) to yield good statistical efficiency at the Gaussian distribution without increasing the bias too much under contamination [31], [43]. The meaning of \( \omega_i \) will be apparent in the next section. By downweighting the outliers instead of completely removing them, the procedure is able to maintain good statistical efficiency at the Gaussian distribution while providing robustness [43]. Finally, we may perform prewhitening by multiplying \( L_k^{-1} \) into (14) on the left-hand side,

\[
(L_k^{-1}) \tilde{z}_k = (L_k^{-1}) \tilde{H}_k \tilde{z}_k + (L_k^{-1}) \tilde{\epsilon}_k
\]

(17)
which can be put into the following form:

$$ y_k = A_k x_k + \eta_k, $$  

(18)

### C. Robust Filtering Based on GM-Estimation

Since the outliers have been downweighted, one may ask if the least squares estimator can now be utilized to obtain the state solution. Actually, the answer is negative because structural errors in $F$ and $H$ can still negatively affect the filter solution through the error covariance matrix $\Sigma$ and state estimation equations. Therefore, a robust technique is still needed that computes a weight matrix $Q$ online while solving the filter. The class of M-estimators is robust but only effective against the observation and innovation outliers’ influence in residual. It is desired to bound the influence of position also, i.e., outlying $\varrho_k$ in $y_k = D_k x_k + \eta_k$, where $D_k$ is the $i$th column vector of the matrix $A_k^T$. Of the several proposals available in the literature, we use the Schweppe-type GM-estimator to solve for $\varrho_k$. It consists in dividing $r_i$ and multiplying $\varrho_k$ by the weight $\varpi_i$ to intertwine the influence of residual and position such that the total influence is bounded. Schweppe’s proposal has been shown to have a positive breakdown point by Maronna et al. [14].

Formally, the GM-estimator is defined as that which minimizes the objective function

$$ J(\varrho) = \sum_{i=1}^{m} \varpi_i^2 \rho(r_{Si}) $$  

(19)

where $\rho(\cdot)$ represents a nonlinear function of standardized residuals; $r_{Si} = r_i/\varpi_i$, with the residuals $r_i = y_i - D_k^T x$, and the robust scale estimate $s$ is the median absolute deviation, defined as $s = 1.4826 \text{median}[|r_i|]$. The constant 1.4826 is a correction factor for Fisher consistency at the Gaussian distribution [43]. This means that $s$ tends asymptotically to $\sigma$ when the observations follow $N(\mu, \sigma^2)$. Note that one cannot simply recompute the PS using the vector $\varrho_k$ to obtain the weights $\varpi_k$ in (19). Particularly, because the outliers in $\varrho_k$ have already been downweighted, corresponding PS values will not contribute accurate information about the remaining structural outliers to the desired weight matrix $Q$.

To overcome this, we just use the values already computed for $\varpi_i$, and achieve computational efficiency at the same time. Returning to (19), we use the Huber $\rho$-function, given by

$$ \rho(r_{Si}) = \begin{cases} \frac{1}{2} r_{Si}^2, & \text{for } |r_{Si}| < c \\ c^2 / 2, & \text{elsewhere} \end{cases} $$  

(20)

as it exhibits $L_2$-norm properties for small residuals and to $L_1$-norm properties for large residuals. Effectively, the $\rho$-function has high efficiency at Gaussian noise but a bounded and continuous influence function for outliers. Again, we pick $c = 1.5$ as the threshold in (20). If the noise PDF $f$ is known, then choosing the $\rho$-function as $\rho(r) = -\ln f(r)$ will yield the maximum-likelihood solution at that PDF for independent and identically distributed data. For the Gaussian distribution, $\rho(r) = r^2 / 2$ gives the optimal solution.

The GM-estimator is obtained by setting the partial derivatives of the objective function in (19) to zero, yielding

$$ \frac{\partial J(\varrho)}{\partial \varrho} = \sum_{i=1}^{m} \varpi_i \frac{\partial \rho(r_{Si})}{\partial r_{Si}} \psi(r_{Si}) = 0. $$  

(21)

The above is clearly a system of nonlinear equations with the $\psi$-function given by $\psi(r_{Si}) = \partial \rho(r_{Si})/\partial r_{Si}$.

### D. Solving for the Robust State Estimates Using IRLS

This system of nonlinear equations is solved using the IRLS algorithm [31], [42]. Multiplying and dividing the $\psi$-function in (21) by $r_{Si}$, and defining the scalar weight function as $q(r_{Si}) = \psi(r_{Si})/r_{Si}$, allows us to put (21) into the matrix form expressed as $A_k^T Q u_k = A_k \hat{x}_k$, where $Q = \text{diag} \{ q(r_{Si}) \}$. Solving for the state estimate using IRLS then yields

$$ u^{(v+1)}_{k|k} = (A_k^T Q^{(v)} A_k)^{-1} A_k^T Q^{(v)} y_k. $$  

(22)

In contrast to the direct noniterative computation of the linear Kalman filter, the IRLS algorithm or Newton’s method is necessary to solve the nonlinear estimator in the GM-KF. It is a price to be paid for robustness because iterating updates the weights online to reflect variations or unexpected impulses in the predictions and observations, which the Kalman filter does not do. In other words, the standardized residuals are applied via the diagonal weight matrix $Q$ to assign weights to the predictions and observations simultaneously and online. Nevertheless, a very fast performance and convergence has been noticed for the GM-KF, generally converging within five iterations at each time step. One final equation is needed to complete this robust filter: the update to the filter error covariance matrix. While the forward-prediction error covariance matrix, given by (5), remains the same as the classical KF, the update filter error covariance matrix needs to be revised.

### IV. ERROR COVARIANCE MATRIX OF THE GM-KF

#### A. Influence Function of the GM-Estimator

The new filter error covariance matrix is derived in this section. First, the influence function of the GM-KF [42], [59], [60] is established for the regression form given by (18). Assuming the system matrices $F_k$ and $H_k$ are deterministic and independent of the residual error vector $x$, the cumulative probability distribution function of $x$ is denoted by $\Phi(x)$. The GM-estimator $\hat{x}$ provides an estimate for $x$ by processing the redundant measurement vector $y$ and solving the implicit equation in (21). This equation is written in compact form as $\sum_{i=1}^{m} \lambda_{k}(x, \varrho_k, \varpi) = 0$, where

$$ \lambda_{k}(x, \varrho_k, \varpi) = \varrho_k \varpi_k \psi(r_{Si}) $$  

(23)

and $\varrho_k$ is the $i$th column vector of the deterministic matrix $A_k^T$ in (18). In terms of the asymptotic functional form of the estimator $\hat{J}(G)$, it is expressed as

$$ \int \lambda(x, \varrho_k, \hat{J}(G)) dG = 0. $$  

(24)
We begin the derivation of the asymptotic influence function of the estimator $T(G)$, given by
\[ \Pi(I, a; \Phi) = \left. \frac{\partial T(G)}{\partial I} \right|_{I=0} = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)\Phi + \varepsilon \Delta_I) - T(\Phi)}{\varepsilon} \]
(25)
where $F = \Phi$ and $H = \Delta_I$, by substituting $G = (1-\varepsilon)\Phi + \varepsilon \Delta_I$ into (24), yielding
\[ \int \Delta(I, a, T(G))d((1-\varepsilon)\Phi + \varepsilon \Delta_I) \\
= \int \Delta(I, a, T(G))d\Phi + \varepsilon \int \Delta(I, a, T(G))d\Delta_I = \Omega. \]
(26)
Differentiating (26) with respect to $\varepsilon$ gives
\[ \frac{\partial}{\partial \varepsilon} \int \Delta(I, a, T(G))d\Phi + \int \Delta(I, a, T(G))d\Delta_I = \Omega. \]
(27)
Evaluating (27) at $\varepsilon = 0$ makes the last term equal to zero, and assuming regularity conditions and Fisher consistency at $\Phi$, given by $\int \Delta(I, a, T(\Phi))d\Phi = \Omega$, further reduces it to
\[ \int \frac{\partial}{\partial \varepsilon} \Delta(I, a, T(G))d\varepsilon + \int \Delta(I, a, T(G))d\Delta_I = \Omega. \]
(28)
After applying the chain rule to the first term and the sifting property of the Dirac impulse to the second term, solving for the influence function gives
\[ \Pi(I, a; \Phi) = \left. \frac{\partial T(I)}{\partial I} \right|_{I=0} = -\left( \int \frac{\partial}{\partial \varepsilon} \Delta(I, a, \varepsilon) |_{\varepsilon=\Phi} \right)^{-1} \Delta(I, a, T(\Phi)), \]
(29)
Substituting (23) into (29) and expanding the partial derivative yields
\[ \Pi(I, a; \Phi) = -\left. \int \frac{\partial \psi(r_{si})}{\partial r_{si}} \frac{\partial r_{si}}{\partial \Phi} d\Phi \right|_{\Phi=T(\Phi)}^{-1} \frac{\partial \psi(r_{si})}{\partial \Phi} \]
(30)
which reduces to
\[ \Pi(I, a; \Phi) = \left. \int \frac{\partial \psi(r_{si})}{\partial r_{si}} a^T d\Phi \right|_{\Phi=T(\Phi)}^{-1} a^T \psi(r_{si}). \]
(31)
Substituting (32) into (31) gives the following expression of the influence function:
\[ \Pi(I, a; \Phi) = \frac{\psi(r_{si})}{E_{\Phi}[\psi'(r_{si})]} (A^T A)^{-1} a. \]
(33)

B. Relationship Between the Influence Function and the Covariance Matrix of an Estimator

Following Fernholz [44], we derive the relationship of the influence function to the asymptotic covariance matrix of the estimation error vector. The Taylor expansion of the functional form of the estimator $T$ gives $T(G) = T(\Phi) + T'(G - \Phi) + \text{rem}(G - \Phi)$, and if the influence function exists, then we get the von Mises expansion given by
\[ T(G) - T(\Phi) = \int \Pi(I, a; \Phi) d(G - \Phi) + \text{rem}(G - \Phi). \]
(34)
Since $\int \Pi(I, a; \Phi) d\Phi = 0$ due to Fisher consistency at the distribution $\Phi$, we have
\[ T(G) - T(\Phi) = \int \Pi(I, a; \Phi) dG + \text{rem}(G - \Phi). \]
(35)
For $G = F_m(\mathbf{z})$, where $F_m(\mathbf{z})$ is the empirical distribution function given by $F_m(\mathbf{z}) = (1/m) \sum_{i=1}^{m} u(\mathbf{z} - r_i)$, and $u(\cdot)$ is the unit step function, the integral term in (35) becomes
\[ \int \Pi(I, a; \Phi) dF_m = \frac{1}{m} \sum_{i=1}^{m} \Pi(I, a; F_m). \]
(36)
Hence, we have
\[ \sqrt{m} [\Pi(T(F_m) - T(\Phi))] = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \Pi(I, a; F_m) + \sqrt{m} \text{rem}(F_m - \Phi). \]
(37)
If the remainder term converges in probability to zero, then by virtue of the central limit theorem, the probability distribution of $\sqrt{m} [\Pi(T(F_m) - T(\Phi))$ tends to $N(0, \Sigma_k)$, with an asymptotic covariance matrix expressed as
\[ \Sigma_k = \frac{E_T [\Pi T^T]}{E_T [\psi^2(r_{si})]} (A^T A)^{-1} (A^T Q_{\psi} A)(A^T A)^{-1} \]
(38)
where $Q_{\psi} = \text{diag} (\psi_{\psi}^2)$.

V. SIMULATION RESULTS

Simulations have been carried out to evaluate the performance of the GM-KF from both the efficiency and the robustness viewpoint. Specifically, we first assess the impact that observation redundancy has on the state estimate mean-square error (MSE). Then, we investigate the robustness of our filter when applied to three models, namely a vehicle and aircraft tracking model, both based on the global positioning...
system (GPS), and a helicopter’s dynamic model. Comparison to the $H_{\infty}$-filter is performed for the first model. The step-by-step description of the GM-KF algorithm is summarized in Table II.

A. Mean-Square Error of the GM-KF State Estimates

To see the benefits of observation redundancy, we consider a GPS-based vehicle tracking controller that is governed by a dynamic model with the following transition and observation matrices:

$$
E = \begin{bmatrix}
1 & 0 & T & 0 \\
0 & 1 & 0 & T \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix},
$$

$$
H_1 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix},
$$

$$
H_2 = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}^T.
$$

This model is provided with four or eight observations when $H_1$ or $H_2$ are applied, respectively, and is characterized by a state vector $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$, which contains the horizontal position, $x_1$, the vertical position, $x_2$, and the associated velocities, $x_3$ and $x_4$. A 10-Hz sampling rate is assumed, giving $T = 0.1$ s as the sampling period. No outliers are induced in this example.

The MSE values resulting from the classical KF and the GM-KF are displayed in Tables III and IV for $H_1$ and $H_2$, respectively. Specifically, we notice that the filter’s MSE is mainly determined by the observation with the lowest variance. For example, its MSE equals 19.3 when a single observation with noise variance $\sigma_{z_k}^2 = 10$ is processed, whereas the MSE reduces to 5.4 when multiple observations are processed and at least one of them has a variance $\sigma_{z_k}^2 < 10$. Note that in both cases the system process noise variance is $\sigma_{w_k}^2 = 10$. Thus, we may say that from an efficiency view point, multiple observations are advantageous in that the one with the lowest noise variance drives the MSE value.

The relative efficiencies of the GM-KF and the KF can be viewed with respect to the number of redundant observations processed by the filter. Particularly, as observed in Table IV, the average ratio of the MSE values of the KF and GM-KF gives a relative efficiency of 85%. Similarly, it is observed in Fig. 4 that the relative efficiency of the GM-KF approaches 95% as the number of redundant observations is increased. This indicates that for the Gaussian case, the asymptotic MSE of our filter is equal to $1/0.95 = 1.05$ times larger than that of the KF, which reaches the Cramér–Rao lower bound.

B. GPS-Based Vehicle Tracking Controller

Next, we evaluate the robustness of our filter using the transition matrix given in (39) with $H_2$ as the observation matrix, yielding $m = 8, n = 4$, and $m_y = 12$. The model assumptions are as follows: 10-Hz sampling rate, giving $T = 0.1$ s as the sampling period; Gaussian observation noise with a known $4 \times 4$ diagonal covariance matrix $R$ with elements equal to 0.1; and Gaussian system noise with a known $4 \times 4$ covariance matrix $W$ equal to the identity matrix. To investigate the resistance of our filter and of an $H_{\infty}$-filter in this model, a single observation outlier is induced in $z_{2,30}$, i.e., the second element of $z$ at $t = 30$ s, via a noise value of $\epsilon_{2,30} = -100$. Thus, when the

---

**TABLE II**

**STEPS OF THE GM-KF**

1. Predict the state and covariance matrix using (4) and (5), respectively.
2. Form a redundant observation vector as shown in (13).
3. Identify the outliers using the Projection Statistics algorithm.
4. Obtain weights using $\alpha_i^2 = \text{min}[1, \theta_i^2/P_{i}^2]$, and apply them to $\tilde{z}_k$.
5. Compute $\hat{R}_k$ given by (15).
6. Perform data prewhitening, as shown in (17).
7. Solve the final linear regression, given by (18) using the solution in (22).
8. Update the covariance matrix using (38).

**TABLE III**

**KF AND GM-KF MSE FOR A SINGLE OBSERVATION PER STATE VARIABLE**

<table>
<thead>
<tr>
<th>System process noise variance</th>
<th>Noise variance of the single-observation case</th>
<th>KF MSE</th>
<th>GM-KF MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>4.94</td>
<td>9.46</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>13.3</td>
<td>19.3</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>33.7</td>
<td>58.9</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>15.9</td>
<td>35.0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>61.4</td>
<td>73.5</td>
</tr>
</tbody>
</table>

**TABLE IV**

**KF AND GM-KF MSE FOR TWO OBSERVATIONS PER STATE VARIABLE**

<table>
<thead>
<tr>
<th>System process noise variance</th>
<th>Noise variances of the two-observation case</th>
<th>KF MSE</th>
<th>GM-KF MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1, 1</td>
<td>3.52</td>
<td>3.96</td>
</tr>
<tr>
<td>10</td>
<td>10, 10</td>
<td>4.65</td>
<td>5.39</td>
</tr>
<tr>
<td>100</td>
<td>1, 10</td>
<td>12.6</td>
<td>14.1</td>
</tr>
<tr>
<td>100</td>
<td>10, 10</td>
<td>4.89</td>
<td>6.1</td>
</tr>
<tr>
<td>100</td>
<td>10, 10, 1</td>
<td>14.9</td>
<td>17.9</td>
</tr>
</tbody>
</table>
vehicle is located at position of $x_1 = 325$ m and $x_2 = 100$ m, the observed value is $x_1 = 325$ m and $x_2 = 0$ m. As observed in Fig. 5, our GM-KF can withstand this outlier whereas the H$_{\infty}$-filter’s estimate is pulled far away from the true position. For the H$_{\infty}$-filter, the parameter $\gamma$ is set to 0.025. $B_k$ is a $4 \times 4$ diagonal matrix with elements equal to 0.025, and $B$ and $W$ are the same as above.

A comparison of the proposed GM-KF with the H$_{\infty}$-filter should also encompass the computational costs of the algorithms that solve them. Using the above model, we have provided in Table V the amount of time required to compute the estimates using these two filters. Table V also presents the tradeoff between the number of redundant observations and computational costs for the GM-KF. Clearly, the GM-KF takes more computational effort than the H$_{\infty}$-filter. However, the average time per step is on the order of 2–4 ms, a delay that may be acceptable in many applications. While such increases in computational costs are often a price to be paid for robustness, they are to be minimized wherever possible. It should be noted that these computing time estimates are derived using unoptimized, experimental code implemented in MATLAB; hence, we only expect that this delay would be less when implemented in a more suitable, noninterpretable, programming language.

![Fig. 5. Metered position (dashed black line) versus GM-KF estimate (solid line) and an H$_{\infty}$-filter estimate (dotted line) of a vehicle’s position with one observation outlier.](image-url)

### TABLE V
**COMPUTATIONAL COST OF GM-KF VERSUS H$_{\infty}$-FILTER FOR DIFFERENT OBSERVATION REDUNDANCIES**

<table>
<thead>
<tr>
<th>Filter Name</th>
<th>Time to Process 60 Time Steps (ms)</th>
<th>Average Time Per Step (ms)</th>
<th># of Redundant Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM-KF</td>
<td>68.0</td>
<td>1.13</td>
<td>2</td>
</tr>
<tr>
<td>H$_{\infty}$</td>
<td>70.5</td>
<td>1.18</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>74.2</td>
<td>1.24</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>79.5</td>
<td>1.33</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>84.4</td>
<td>1.41</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>6.40</td>
<td>0.11</td>
<td>0</td>
</tr>
</tbody>
</table>

### TABLE VI
**LIST OF SYMBOLS**

- $x_k$: State vector at time step $k$
- $z_k$: Observation vector at time step $k$
- $F_k$: State transition matrix at time step $k$
- $H_k$: Observation matrix at time step $k$
- $G_k$: Observation noise vector at time step $k$
- $w_k$: System process noise vector at time step $k$
- $v_k$: Target probability distribution
- $G$: $e$-contaminated probability distribution
- $H$: Contamination probability distribution
- $v$: Amount of contamination
- $T(G)$: Functional form of the estimator at $G$
- $r_k$: $\ell_k$ residual
- $h_k$: Transpose of the $k$th row of matrix $H$
- $\Delta_k$: Infinitesimal contamination at $r_k$
- $\Sigma_{h_k}$: Predicted error covariance matrix at time step $k$ given data up to time step $k-1$
- $\Sigma_{u_k}$: Filter error covariance matrix at time step $k$ given data up to time step $k$
- $\mathbf{I}^F$: Influence function
- $\tilde{x}_{k|k-1}$: Predicted state vector at time step $k$ given data up to time step $k-1$
- $\hat{x}_{k|k}$: Filtered state vector at time step $k$ given data up to time step $k$
- $b_{max}(\varepsilon; F)$: Maximum bias of $\tilde{x}$ at $F$ as a function of $\varepsilon$
- $m_k$: Total number of observations in $z_k$
- $n_k$: Number of state variables in $x_k$
- $R_k$: Observation noise covariance matrix at time step $k$
- $W_k$: System process noise covariance matrix at time step $k$
- $K_k$: Kalman filter gain at time step $k$
- $m_l$: Power spectrum
- $P_l$: Matrix of transfer functions
- $J_k$: Objective function for the H$_{\infty}$-filter
- $B_k$: Design matrix in the H$_{\infty}$-filter
- $\gamma$: Design parameter in the H$_{\infty}$-filter
- $\delta_{k,t-1}$: Error between predicted and true state vectors
- $\tilde{z}_{k,t}$: Redundant observation vector at time step $k$
- $\tilde{B}_k$: Combined system and observation matrix at time step $k$ corresponding to $\tilde{z}_{k,t}$
- $\tilde{z}_{k,t}$: Noise vector corresponding to $\tilde{z}_{k,t}$
- $\tilde{R}_k$: Noise covariance matrix corresponding to $\tilde{z}_{k,t}$
- $L_k$: Cholesky decomposition of $\tilde{A}_k$
- $P_l$: Projection Statistic for the $l$th data point
- $U$: Unit vector
- $q(\cdot)$: Weight function used in a GM-estimator
- $\sigma_l$: Weights derived from the projection statistics
- $Y$: Observation vector after prewhitening
- $A_k$: Combined system and observation matrix corresponding to $Y$
- $\eta_k$: Noise vector corresponding to $Y$
- $\phi_k$: Transpose of the $k$th row of the matrix $A_k$
- $J(\cdot)$: Objective function for a GM-estimator
- $\rho(\cdot)$: $\rho$-function used in GM-estimator
- $F_{\ell_k}$: $\ell_k$-scaled residuals
- $\psi(\cdot)$: Psi-function of a GM-estimator
- $\Phi$: Target Gaussian probability distribution
- $Q$: Diagonal weight matrix with diagonal elements, $q(\cdot)$
- $Q_{\ell_k}$: Diagonal weight matrix with diagonal elements, $\sigma_l^2$

Authorized licensed use limited to: to IEEEExplore provided by Virginia Tech Libraries. Downloaded on May 17,2010 at 22:31:09 UTC from IEEE Xplore. Restrictions apply.
C. GPS-Based Aircraft Tracking Model

Next, we consider a tracking problem using GPS data for an aircraft in a circular maneuvering exercise. The state transition matrix, \( E \), is given by

\[
E = \begin{bmatrix}
1 & 0 & T & 0 \\
0 & 1 & 0 & T \\
0 & 0 & 1 & 1 \\
0 & 0 & -9 & 1
\end{bmatrix}.
\]

The observation matrix is equal to \( H_2 \) given by (39). For this model, the maximum number of outliers that any equivariant estimator can handle is \( m_t = n_t + 2 = 16 \). Extensive simulations were carried out in which observation, innovation, and structural outliers were introduced simultaneously at a given time step and successively over several time steps. Through these simulations, it is found that the GM-KF is able to suppress up to three concurrent outliers but not four, implying that it has a breakdown point of \( \varepsilon^* = 3/12 = 0.25 \), which is a little bit less than the theoretical maximum.

As an example, consider a scenario in which \( W \) and \( \Sigma_{400} \) are set equal to the identity matrix; \( R \) contains 0.75 on the diagonal elements and zero otherwise; and three outliers occur from \( t = 15 \) s to \( t = 18 \) s, one innovation and two observations. The first element of the predicted vector \( \hat{x}_{14t} \) is replaced by \( \hat{x}_{14t} = [-16.6, -20.7] \) at \( t = 15 \) s and \( t = 16 \) s, and the observation \( z \) is corrupted by outliers such that the first two elements of the original vectors from \( t = 15 \) s to \( t = 18 \) s, given by

\[
\begin{bmatrix}
\frac{2}{15} & \frac{4}{15} \\
\frac{2}{15} & \frac{4}{15}
\end{bmatrix}
\]

are replaced by

\[
\begin{bmatrix}
\frac{2}{15} & \frac{4}{15} \\
\frac{2}{15} & \frac{4}{15}
\end{bmatrix}
\]

while the observation matrix is the same as \( H_2 \) in (39), and the state vector \( x = \begin{bmatrix} x_1 x_2 x_3 \end{bmatrix} \) contains the horizontal and vertical velocities, pitch rate, and pitch angle, respectively.

D. GPS-Based Aircraft Dynamic Model

Now, we consider a model that represents the dynamic behavior of a helicopter under typical loading and flight conditions at airspeed of 135 knots [61]. In the discrete time, the transition matrix, \( E \), is given by

\[
E = \begin{bmatrix}
0.9964 & 0.002579 & -0.0001258 & -0.04597 \\
0.002579 & 0.9964 & -0.0001258 & -0.04597 \\
0.002579 & 0.9387 & -0.01879 & -0.3834 \\
0.000492 & 0.03388 & 0.9383 & 0.1302 \\
0.000492 & 0.03388 & 0.9383 & 0.1302
\end{bmatrix}
\]

Fig. 6 depicts the measured signal, true state sequence, and the GM-KF estimated values of this example. As observed in the figure, the GM-KF suppresses these three outliers, providing a robust solution near the true state sequence. On average, the filter converges within four iterations of the IRLS algorithm in this scenario, with a convergence criterion of \( \| \hat{x}_{k+1} - \hat{x}_k \| < 0.01 \), where \( k \) is the iteration number.

Fig. 7 depicts an example in which two structural outliers occur simultaneously from \( t = 15 \) s to \( t = 36 \) s in the following manner: \( F_{333} = 10 r_1 F_{333} \) and \( H_{333} = 3 r_2 H_{333} \), where \( r_1 \) and \( r_2 \) are random numbers drawn from the uniform distribution. Clearly, the GM-KF is able to suppress concomitant structural outliers due to redundancy and the use of a robust estimation procedure. More importantly, this example represents a problem of time-varying bias caused by model uncertainties, which the GM-KF solves efficiently and online without \textit{a priori} knowledge of the contamination distribution.

Fig. 8 depicts another example in which all three types of outliers occur simultaneously, namely one structural, one observation, and one innovation outlier. The state transition matrix is corrupted as described above, but only from \( t = 15 \) s to \( t = 18 \) s. In addition, the third element of the observation
vector $\mathbf{x}$ is corrupted at each time step by an observation outlier, i.e., \begin{align*}
[\hat{z}^3_{12}, \hat{z}^3_{16}, \hat{z}^3_{17}, \hat{z}^3_{18}] = [18.4, 21.4, 58.6] \text{ replaces } [\hat{z}^3_{15}, \hat{z}^3_{16}, \hat{z}^3_{17}, \hat{z}^3_{18}] = [-0.04, 0.84, 3.49].
\end{align*}
Finally, the predicted vector contains an innovation outlier in the third element, which is induced by replacing
\begin{equation}
[\hat{x}^3_{15|14}, \hat{x}^3_{16|15}, \hat{x}^3_{17|16}, \hat{x}^3_{18|17}] = [0.14, 2.53, 0.41, 3.89] \quad (44)
\end{equation}
by
\begin{equation}
[\hat{x}^3_{15|14}, \hat{x}^3_{16|15}, \hat{x}^3_{17|16}, \hat{x}^3_{18|17}] = [54.4, 20.8, 45.5, 85.2]. \quad (45)
\end{equation}

Extensive Monte Carlo simulations have confirmed that the GM-KF has a breakdown point of 25%. On average, the filter converged within five iterations of the IRLS algorithm under the same convergence criteria. Consequently, the GM-KF may be utilized in the platform’s control system to obtain reliable state estimates when there are up to 25% of gross errors of all three types. By contrast, as seen in Fig. 9, the $H_\infty$-filter cannot withstand even a single observation outlier.

VI. Conclusion

In this paper, a batch-mode GM-Kalman filter is proposed that performs well in Gaussian noise and under contamination due to observation, innovation, and structural outliers. A new error covariance matrix for the filter was developed using the influence function. The GM-KF’s efficiency and robustness to concurrently occurring outliers were investigated through various simulations. In particular, from an efficiency viewpoint, we noticed that multiple observations are advantageous in the estimation as the one with the lowest noise variance drives the MSE value. The statistical efficiency for the GM-KF relative to the classical KF turns out to be 85% in our simulations. From a robustness perspective, the GM-KF was shown to suppress all three types of outliers in various scenarios with a breakdown point of 25%. The filter’s robustness under contamination was also compared to the weak performance of the $H_\infty$-filter. In summary, we may say that batch-mode estimation designed into the GM-KF is much more desirable than recursive estimation, as the redundancy in the observations leads to a more efficient and robust filter.

Many aspects of the GM-KF are open to further research. The $\rho$-function can be modified to obtain different convergence rate, robustness, and efficiency properties in the filter. For example, replacing the Huber $\rho$-function with the strictly convex logistic function allows one to iterate by means of the Newton method instead of the IRLS algorithm. The filter convergence rate would then be quadratic instead of linear. Besides variations to the GM-KF, completely new filters can also be developed by replacing the GM-estimator with another regression estimator of choice, such as the MM-estimator [14]. Finally, the framework can be extended to develop a robust filter for nonlinear systems, e.g., by reformulating the extended Kalman filter.

REFERENCES
