Robust Methods for Estimation Problems

Ankan Bansal
10095
Department of Electrical Engineering
Email: ankan@iitk.ac.in

Atul Kumar Sinha
10171
Department of Electrical Engineering
Email: atulkr@iitk.ac.in

***Abstract*—Estimation in conventional signal processing is often based on strong assumptions on the probability distribution of the sensor noise, stationarity, linearity, or independence and identical distribution of random variables. Many a time engineering problems rely on Gaussian distribution of the data and noise. This is sometimes well justified and enables a simple derivation of optimal estimators. But sometimes, these assumptions do not hold in practice and nominal optimality becomes useless. A slight deviation from the assumed distribution may result in drastic degradation of the estimator. Hence there is a need for statistically robust methods of estimation. In this paper, we elaborate on some of the important robust methods for estimation problems for both single channel and multi-channel cases, particulary the generalized M-estimators and the sequential M-estimators. Further, robust Kalman filter based estimation is also highlighted.**

I. Introduction

Statictical signal processing often relies on strong assumptions about the signal and noise characteristics e.g. the Gaussian or any other specific model, which are nominal and may, at best, be only approximately valid most of the time. In practice one rarely has perfect knowledge about the statictical description of the signal or the noise processes. In particular, measurement campaigns have confirmed the presence of impulsive (heavy-tailed) noise, which can cause optimal signal processing techniques, especially the ones derived using the nominal Gaussian probability model, to be biased or to even break down. The occurrence of impulsive noise has been reported, for example, in outdoor mobile communication channels, due to switching transients in power lines or automobile ignition, in radar and sonar systems as a result of natural or man-made electromagnetic and acoustic interference and in indoor wireless communication channels, owing, e.g., to microwave ovens and devices with electromechanical switches, such as electric motors in elevators, printers, and copying machines. In geolocation position estimation and tracking, nonline-of-sight (NLOS) signal propagation, caused by obstacles such as buildings or trees, results in outliers in the measurements, to which conventional position estimation methods are very sensitive. In classical short-term load forecasting, the prediction accuracy is adversely influenced by outliers, which are caused by nonworking days or exceptional events such as strikes or natural disasters.

These situations naturally enforces the question, how sensitive is the performace of an estimator to deviations in the signal and noise characteristics from those for which the scheme is designed. Unfortunately, it turns out that in many cases nominally optimum signal processing schemes can suffer a drastic degradation in performace even for apparently small deviations from the nominal assumptions. This motivates the search for robust signal processing techniques; that is, techniques with good performance under any nominal conditions and acceptable perforamce for signal and noise conditions other that the nominal which can range over the whole of allowable classes of possible characteristics.

The rest of the paper is organised as follows. In section II, we discuss the basic concepts of robustness followed by a discussion on robust estimators for single-channel data and multi-channel data in sections III and IV respectievely. Robust Kalman Filters are discussed in section V. Finally, we conclude the article in section VI.

II. Basic Concepts of Robustness

Some basic definitions and concepts are presented in this section.

A. *The Breakdown Point (BP)*

The BP is used to characterize the quantitative robustness of an estimator. It indicates the maximal fraction of outliers (highly deviating samples) in the observations, which an estimator can handle without breaking down. The BP takes values between 0 % and 50 %, where a higher BP value corresponds to larger quantitative robustness. The BP of the sample mean is zero, which means that a single outlier may throw the estimator completely off, while for the sample median it is 50 %.

B. *The Influence Function (IF)*

The IF describes the bias impact of infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination.

$$IF(z;\hat{\theta},F_\theta) = \lim_{\epsilon\to 0} \frac{\hat{\theta}_\infty(G) - \hat{\theta}_\infty(F_\theta)}{\epsilon} = \left.\frac{\partial\hat{\theta}_\infty(G)}{\partial\epsilon}\right|_{\epsilon\to 0}, \quad (1)$$

where $\hat{\theta}_\infty(F_\theta)$ and $\hat{\theta}_\infty(G)$ are the asymptotic values of the estimator when the data is distributed, following, respectively, F_θ and the contaminated distribution $G = (1-\epsilon)F_\theta + \epsilon\Delta_z$ being the point-mass probability on z and ϵ the fraction of contamination. The IF is depicted with respect to z, the position of the infinitesimal contamination. If the IF is bounded and continuity, the estimator is considered to be qualitatively robust against infinitesimal contamination.

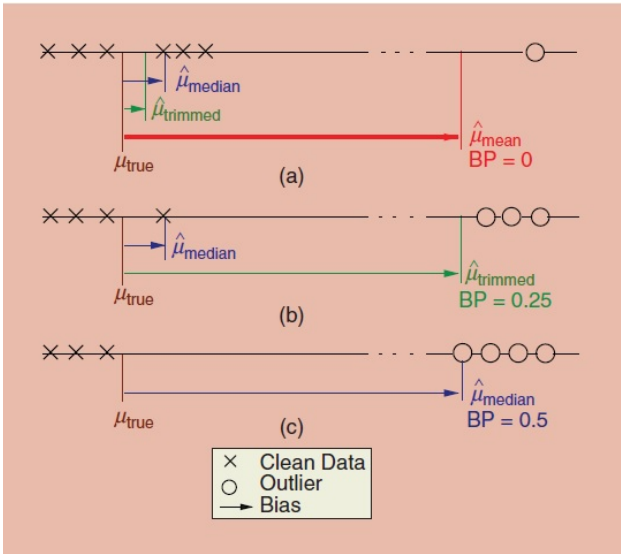

Fig. 1. [1] The bias and BP of three traditional estimators of location μ, i.e., (a) the sample mean, (b) the α-trimmed mean, and (c) the sample median. "Clean observations are depicted as crosses and outliers as circles. The BP of the sample mean is zero, which means that a single outlier has an unbounded effect on its bias. The α-trimmed mean, where $\alpha = 0.25$ for this example, resists one outlier by ignoring the largest and smallest 25% of the data. The BP of the sample median equals the highest possible value of 50%, which means that its bias remains bounded even in situations when up to half of the observations, in this case three, are replaced by arbitrarily large values.

C. *The Maximum Bias Curve (MBC)*

The MBC gives information on the bias affected by a specific amount of contamination. It plots the absolute value of the maximum possible asymptotic bias $b_{\hat{\theta}}(F,\theta) = \hat{\theta}_\infty(F_\theta) - \theta$ of an estimator $\hat{\theta}$ with respect to the fraction of contamination ϵ, whereby the most robust one minimizes the MBC.

$$MBC(\epsilon, \theta) = \max\{|b_{\hat{\theta}}(F,\theta)| : F \in \mathcal{F}_{\epsilon,\theta}\}, \tag{2}$$

where $\mathcal{F}_{\epsilon,\theta} = \{(1-\epsilon)F_\theta + \epsilon G\}$ is an ϵ - neighbourhood of distributions around the nominal distribution F_θ with G being an arbitrary contaminating distribution.

III. Robust Estimators for Single-Channel Data

Consider the observation model:

$$Y_n = \mu + V_n \quad n = 1, ..., N, \tag{3}$$

where Y_n, V_n are random variables and y_n, v_n are their realizations for $n = 1, ..., N$. The goal is to estimate the value of μ from observations y_n, $n = 1, ..., N$.

Many engineering problems can be formulated as in (3) and the problem is known as estimation of location [6]. The LSE of location, i.e., the sample mean, gives all observations the same weight. This gives the best statistical performance if the Gaussian assumption is fulfilled. However in many practical situations the noise process is non-Gaussian. Then for a robust location estimate we should weigh the observations $y_n, n = 1, ..., N$ such that we give more weight to data that is close to the measurement model as compared to the one that is unlikely to occur. Keeping this in mind the M-estimator was introduced [5]. M-estimators result in a non linear IF.

A. *M-Estimator*

M-estimators are a generalization of MLE. They can resist outliers without preprocessing the data. In the location case, they solve

$$\sum_{n=1}^{N} \psi(Y_n - \hat{\mu}) = 0, \tag{4}$$

where ψ is the derivative of the loss function ρ, which is to be minimized. The MLE under the Gaussian noise assumption corresponds to an M-estimator if $\rho(X) = X^2$, which results in a linear ψ-function. By selection of an appropriate ψ-function, M-estimators achieve high efficiency under the nominal model as well as qualitative robustness and a high BP. For M-estimators, the IF at a nominal Gaussian, is proportional to their ψ-function.

B. *Linear Regression Models*

Many engineering problems in very diverse areas can be formulated as a linear regression for which the parameters of interets are sought.The linear regression model is given by

$$Y_n = \boldsymbol{h}_n^\top \boldsymbol{\theta} + V_n, \; n = 1, ..., N, \tag{5}$$

where we assume that the predictors $\boldsymbol{h}_n = (h_{1n}, ...h_{pn})^\top$, the errors V_n and data points $(Y_n, \boldsymbol{h}_n^\top)$ for $n = 1, ..., N$, are i.i.d. random variables, $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^\top$ are the unknown parameters if interest, and $\boldsymbol{h}_n$ and V_n are mutually independent.

The LSE is not robust against any type of outliers that contribute in an unbounded fashion to the bias of the estimator. To overcome this problem, several robust estimators of $\boldsymbol{\theta}$ have been proposed [5].

M-estimators gain their robustness compared to the MLE under the Gaussian noise assumption, which coincides with the LSE, by downweighting vertical outliers with a bounded $\psi = \rho\prime$. For the LSE $\rho(x) = x^2$, while $\rho(x) = |x|$ gives the l_1-estimator. The M-estimator is a solution of

$$\sum_{n=1}^{N} \psi\left(\frac{\hat{v}_n}{\hat{\sigma}_V}\right) \boldsymbol{h}_n = \mathbf{0}, \tag{6}$$

with residuals $\hat{v}_n = y_n - \boldsymbol{h}_n^\top \hat{\boldsymbol{\theta}}$ and corresponding robust scale estimate $\hat{\sigma}_V$. The solution is computed by iteratively reweighted least-squares. When M-estimates are computed in this way, they are regression equivariant, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{H}, \mathbf{Y}+\mathbf{Hu}) = \hat{\boldsymbol{\theta}}(\mathbf{H}, \mathbf{Y}) + \mathbf{u}$, where $\mathbf{u}$ is any $p \times 1$ vector, scale equivariant, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{H}, \sigma\mathbf{Y}) = \sigma\hat{\boldsymbol{\theta}}(\mathbf{H}, \mathbf{Y})$, where σ is any scalar, and affine equivariant, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{HA}, \mathbf{Y}) = \mathbf{A}^{-1}\hat{\boldsymbol{\theta}}(\mathbf{H}, \mathbf{Y})$, where $\mathbf{A}$ is any $p \times p$ nonsingular matriH.

Another common problem in signal processing, which may contain vertical outliers, is the estimation of the parameters of a sinusoid in contaminated noise. Robust estimation of complex sinusoids with known frequencies based on sensor measurements $Y_{n,n=1,...,N}$, can be formulated as a linear regression problem. The observations follow the model

$$Y_n = \sum_{k=1}^{K} \theta_k e^{j\omega_k n} + V_n \quad n = 1, 2, ..., N, \tag{7}$$

where θ_k is the unknown complex amplitude of the sinusoid with known frequencies $\omega_k, k = 1, ..., K$ and $V_n, n = 1, ..., N$ is the contaminated Gaussian i.i.d. circular complex-valued noise. The goal is to estimate θ_k for $k = 1, 2, ..., K$. By separating the real from the imaginary parts, (7) can be reformulated in the form of (5). Using an M-estimator will provide robustness against possible vertical outliers. In [4], the authors introduce a semiparametric approach, where they estimate the distribution of Vn to robustly estimate $\theta_k, k = 1, ..., K$. They show the advantage of the semiparametric estimator over the M-estimator.

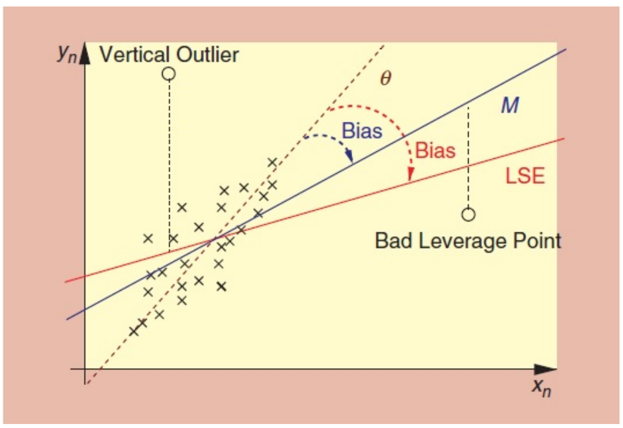

Fig. 2. [1] Different types of outliers in regression and their effect on the bias of the LSE and on an M-estimator. The LSE is not robust against any type of outliers. M-estimators are robust against vertical outliers, but they become nonrobust with a BP equal to zero in the presence of leverage points.

IV. Robust Estimators for Multi-channel Data

Many key signal processing applications use multiple sensors to acquire multichannel data from measurement systems such as multiple-input, multiple-output (MIMO) communication systems as well as phased array and MIMO radar systems. The measurements are vector valued and often realizations of random signals and noise are described by a multivariate probability distribution function.

One may attempt to derive robust signal processing techniques for multichannel data by simply applying robust methods developed for scalar signals to each signal component independently. This approach is unsuitable as it may lead to unexpected results.Moreover, none of the data vector components may be outlying alone but the multivariate observations they form may be far away from the majority of the data. Such outliers would not be detected if robust signal processing is performed on individual single components, which also ignores the underlying correlation among the vector components. In such cases estimators that are affine equivariant are highly desirable.

Robust estimation of the multivariate location parameter is needed in many multichannel filtering problems such as noise attenuation and outlier (impulsive noise) removal in color images, multichannel biomedical measurements, and multimodal imaging as well as sensor array signal processing.

A. *The Minimum Volume Ellipsoid Estimator and the Minimum Covariance Determinant Estimator*

The minimum volume ellipsoid (MVE) estimate of location is defined as the center of the minimum volume ellipsoid covering at least h out of N data points. The minimum covariance determinant (MCD) estimate of location is found by computing the arithmetic mean for h out of N points for which the determinant of the sample covariance matrix is minimal. Given h samples that are not outlying, estimation is then based on the clean data points. The corresponding covariance matrix estimates are the sample covariance matrix of these points. Both of these estimates employ a robust Mahalanobis distance, which is obtained by replacing the sample mean $\hat{\boldsymbol{\mu}}$ and the sample covariance $\hat{\boldsymbol{\Sigma}}$ by their robust counterparts.

B. *The $S-$, τ-and MM-Estimator of Location and Scatter*

The S-estimator of location and scatter is found by minimizing the determinant of an M-estimate of the covariance matrix. The intuition behind the multivariate S-estimator is to minimize a robust measure of the distance, i.e., the determinant of an M-estimator of the covariance matrix. The BP of S-estimators is independent of the dimensionality of the data, unlike M-estimators for which the BP decreases.

C. *The M-estimator of Covariance*

The M-estimator of covariance based on a sample $\mathbf{z}_1, ..., \mathbf{z}_N$ in $\mathbb{C}^K$ is the positive definite Hermitian (PDH) $K \times K$ matrix $\hat{\boldsymbol{\Sigma}}$, which solves

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{n=1}^{N} w(\mathbf{z}_n^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{z}_n) \mathbf{z}_n \mathbf{z}_n^H, \tag{8}$$

where $w : [0, \infty) \to \mathbb{R}$ is a weighting function. A robust weighting function $\dot{w()}$ is descending to zero, i.e., a highly deviating observation $\mathbf{z}_n$ with large $\|\hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{z}_n\|^2 = \mathbf{z}_n^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{z}_n$ is given less weight.

Using the complex t-distribution as a heavy-tailed model, we obtain t_v M-estimators as an example of M-estimators, with the weight function

$$w(s) = w_v(s) = \frac{2K + v}{v + 2s}, \tag{9}$$

where $v > 0$. This M-estimator is also the MLE of the data is complex K vector-valued t_v-distributed. The estimates may be computed with the following iterative formula:

$$\hat{\boldsymbol{\Sigma}}_{K+1} = \frac{1}{N} \sum_{n=1}^{N} w_v(\mathbf{z}_n^H \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{z}_n) \mathbf{z}_n \mathbf{z}_n^H, \tag{10}$$

which converges to the unique solution $\hat{\boldsymbol{\Sigma}}$ under mild regularity conditions on the data

D. Sequential M-estimator

Consider an alternative categorisation of the system in which the observations are measured sequentially,

$$\boldsymbol{y}[n] = \mathbf{H}[n]\boldsymbol{\theta} + \boldsymbol{v}[n] \tag{11}$$

where $\boldsymbol{y}[n] = [y_1, y_2, \ldots, y_n]^\top \in \mathcal{R}^n$ is the observation vector at time $n, \mathbf{H}[n] = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n]^\top \in \mathcal{R}^{n\times P}$ is the corresponding system matrix and $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_P]^\top \in \mathcal{R}^P$ is the vector of P unknown parameters to be estimated, and $\boldsymbol{v}[n] = [v_1, v_2, \ldots, v_n]^\top \in \mathcal{R}^n$ is the noise vector with a non-Gaussian density function $g(x)$ described by the ϵ-contaminated model as

$$g(v) = (1-\epsilon)f_G(v, \sigma^2) + \epsilon\mathcal{I}(v) \tag{12}$$

where ϵ represents the contamination percentage, σ^2 is the variance of the Gaussian noise, f_G is the zero-mean Gaussian density, and $\mathcal{I}$ is some unknown symmetric function representing the impulsive part of the noise density. It is of practical interest to compute the estimate of the unknown parameter vector $\boldsymbol{\theta}$ for a small number of observations and then update the estimate whenever a new observation is recorded. When the noise is purely gaussian, i.e. $\epsilon = 0$, the following sequential LS algorithm is optimum in the maximum-likelihood sense.

$$\begin{aligned}
\hat{\boldsymbol{\theta}}[n+1] &= \hat{\boldsymbol{\theta}}[n] + \mathbf{K}[n+1]\left(\boldsymbol{y}[n+1] - \boldsymbol{h}_{n+1}^\top\hat{\boldsymbol{\theta}}[n]\right) \\
&\quad \ldots && (13)\\
\mathbf{K}[n+1] &= \frac{\boldsymbol{\Sigma}[n]\boldsymbol{h}_{n+1}}{\sigma_n^2 + \boldsymbol{h}_{n+1}^T\boldsymbol{\Sigma}[n]\boldsymbol{h}_{n+1}} && (14)\\
\boldsymbol{\Sigma}[n+1] &= \left(\mathbf{I} - \mathbf{K}[n+1]\boldsymbol{h}_{n+1}^\top\right)\boldsymbol{\Sigma}[n] && (15)\\
\boldsymbol{\Sigma}[n] &= \left(\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n]\right)^{-1} && (16)
\end{aligned}$$

As with the general case, slight departure from the nominal assumptions can render this estimator useless and thus a robust sequential estimator is called for. In the following sections, we elaborate on a seqential M-estiamtion technique under the framework of robust statistics when the noise in not Gaussian.

*1) **Algorithm**:* Here, we give a summary of the algorithm.

- Let $\hat{\boldsymbol{\theta}}[n]$ and $\hat{\boldsymbol{\theta}}[n-1]$ be the known estimates of $\boldsymbol{\theta}$ at the n^{th} observation.
- Define $\gamma(x) = -\partial\psi(x)/\partial x$, and compute $\mathbf{C}_{\hat{\boldsymbol{\theta}}[n]}[n]$ and $\mathbf{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]$, where

$$\begin{aligned}
\mathbf{C}_{\boldsymbol{\theta}}[n+1] &= diag\left(\gamma(z_1(\boldsymbol{\theta})), \ldots, \gamma(z_3(\boldsymbol{\theta}))\right) && (17)\\
z_{n+1}(\boldsymbol{\theta}) &= y_{n+1} - \boldsymbol{h}_{n+1}^\top\boldsymbol{\theta} && (18)
\end{aligned}$$

- Compute the difference, $\delta\mathbf{C}_{\hat{\boldsymbol{\theta}}[n]} = \mathbf{C}_{\hat{\boldsymbol{\theta}}[n]}[n] - \mathbf{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]$.
- Compute (only once),

$$\mathbf{J}[n] = \left(\mathbf{H}^\top[n]\mathbf{C}_{\hat{\boldsymbol{\theta}}[n]}\mathbf{H}[n]\right)^{-1} \tag{19}$$

- At $(n+1)^{th}$ observation, compute

$$\mathbf{F}_{\hat{\boldsymbol{\theta}}[n]}[n] = \mathbf{J}[n]\left(\mathbf{I} - \left(\mathbf{H}^T[n]\delta\mathbf{C}_{\hat{\boldsymbol{\theta}}[n]}\mathbf{H}[n]\right)\mathbf{J}[n]\right) \tag{20}$$

- Update the vector

$$\begin{aligned}
&\mathbf{J}[n+1] = \mathbf{F}_{\hat{\boldsymbol{\theta}}[n]}[n] - \\
&\frac{\gamma\left(y_{n+1} - \boldsymbol{h}_{n+1}^\top\hat{\boldsymbol{\theta}}[n]\right)\mathbf{F}_{\hat{\boldsymbol{\theta}}[n]}[n]\boldsymbol{h}_{n+1}\boldsymbol{h}_{n+1}^\top\mathbf{F}_{\hat{\boldsymbol{\theta}}[n]}[n]}{1 + \gamma\left(y_{n+1} - \boldsymbol{h}_{n+1}^\top\hat{\boldsymbol{\theta}}[n]\right)\boldsymbol{h}_{n+1}^\top\mathbf{F}_{\hat{\boldsymbol{\theta}}[n]}[n]\boldsymbol{h}_{n+1}}
\end{aligned} \tag{21}$$

- Update the parameter estimates, as follows

$$\hat{\boldsymbol{\theta}}[n+1] = \hat{\boldsymbol{\theta}}[n] - \mathbf{J}[n+1]\boldsymbol{h}_{n+1}\psi\left(y_{n+1} - \boldsymbol{h}_{n+1}^\top\hat{\boldsymbol{\theta}}[n]\right) \tag{22}$$

*2) **Initialization**:* One must start with some robust estimate of the unknown parameters for a given number of observations $n_0 \leq N$, to use the sequential updates, i.e., by solving the following M-equations

$$\boldsymbol{g}_{\boldsymbol{\theta}}[n_0] = \mathbf{H}^\top[n_0]\psi\left(\boldsymbol{y}[n_0] - \mathbf{H}[n_0]\hat{\boldsymbol{\theta}}[n_0]\right) = \mathbf{0}_P \tag{23}$$

Numerical procedures for solving M-equations can be found, for example, in [5], [10]. For identifiability, we require $n_0 > P$, where P is the number of parameters. If the total number of observations $N \gg P$, it is recommended to select $n_0 \geq 4P$ [7], [8]. However, the main problem with the aforementioned initialization is that one still needs to solve the nonlinear-equations. It is therefore more preferable to look for some low-cost linear but reasonably robust technique. For that reason, we propose to use the LS solution but over the trimmed data instead of the full data. The main idea is to remove outliers, leading to slightly less informative data but more suitable for linear estimation [1].

*3) **Discussion**:* It is notable that while the sequential LS algorithm yields the exact updates, the proposed sequential M-estimation technique only provides approximate updates. This is because of the non-linearity of the score function, $\psi(x)$. We have only considered stationary processes in this paper. However, extension to nonstationary processes is straightforward. Also, the assumption of a known the variance of the Gaussian noise can be relaxed in practice by some adaptive mechanism (see [5] and [9] for a number of possibilities).

V. Robust Kalman Filter

A. Background

The system dynamics and observations are represented by a state space model, which contains the system process and observation noises that affect the state predictions and observations, respectively, as follows

$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{s}_k + \boldsymbol{e}_k \quad (24)$$

$$\boldsymbol{s}_k = \mathbf{F}_{k-1} \boldsymbol{s}_{k-1} + \boldsymbol{w}_{k-1} \quad (25)$$

where $\boldsymbol{e}_k$ and $\boldsymbol{w}_k$ are observation noise and process noise respectievely. For consistency with the type of noises that cause the outliers, we call these observation and innovation outliers, respectively. Formally, we use an ϵ-contaminated model to induce a topological neighborhood around the target distribution as in (12).

To handle these outliers, several nonlinear methods have been proposed in the literature. For the particular case of observation outliers, various methods have been proposed in the literature, namely Christensen and Solimans filter based on the least absolute value criterion; Doblingers adaptive KF scheme; and Durovic, Durgaprasad, and Kovacevics filters utilizing the maximum likelihood-type (M-)estimator based robust KF. However, none of these methods iterate when solving the underlying nonlinear estimator at each recursion time step, effectively assuming that the predictions are accurate which may yield unreliable results when observation and innovation outliers occur simultaneously. 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. The Kalman filter proposed by Chan et al 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.

Here we present an algorithm based on 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. Apart from the observation and innovation outliers, we also consider structural outliers generated by errors in $\mathbf{H}_k$ and $\mathbf{F}_{k-1}$ in (24) and (25). 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.

B. Algorithm Description

*1) **Step I : Linear Regression Framework**:* Beginning with a prediction or initial state vector, we first convert the linear dynamic system in (24) and (25) 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, $\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{s}_k + \boldsymbol{\delta}_{k|k-1}$, we obtain

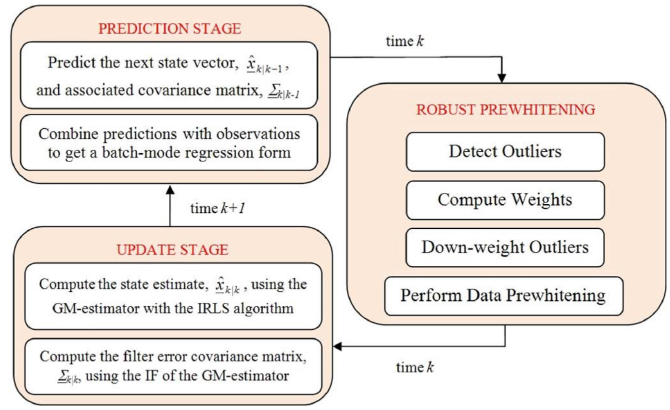

Fig. 3. [2] Flow chart depicting the key steps of the GM-KF method: state prediction, robust prewhitening, and state estimate update.

$$\begin{bmatrix} \boldsymbol{y}_k \\ \hat{\boldsymbol{s}}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_k \\ \mathbf{I} \end{bmatrix} \boldsymbol{s}_k + \begin{bmatrix} \boldsymbol{e}_k \\ \boldsymbol{\delta}_{k|k-1} \end{bmatrix} \quad (26)$$

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

$$\tilde{\boldsymbol{y}}_k = \tilde{\mathbf{H}}_k \boldsymbol{s}_k + \tilde{\boldsymbol{e}}_k \quad (27)$$

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

$$\tilde{\mathbf{R}}_k = \begin{bmatrix} \mathbf{R}_k & 0 \\ 0 & \boldsymbol{\Sigma}_{k|k-1} \end{bmatrix} = \mathbf{L}_k \mathbf{L}_k^\top \quad (28)$$

where the term $\mathbf{L}_k$ may be obtained by Cholesky decomposition, $\mathbf{R}_k$ is assumed to be a known noise covariance of $\boldsymbol{e}_k$, and $\boldsymbol{\Sigma}_{k|k-1}$ is the filter error propagation.

*2) **Step II : Robust Prewhitening**:* It is now desired to uncorrelate the data in the linear regression. However, applying the matrix $\mathbf{L}_k$ given by (28) in the prewhitening step with outliers present in the data would lead to negative effects on the data. Thus, we need to detect and handle the outliers in the regression first. One classical outlier detection method is the Mahalanobis distances, 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 which employs instead the sample median and median-absolute-deviation of the data points $\boldsymbol{h}$ on the direction of all possible vectors $\boldsymbol{u}$ as robust estimates of location and scale, respectively. Formally, it is expressed as

$$PS_i = \max_{\|\boldsymbol{u}\|=1} \frac{\left| \boldsymbol{h}_i^\top \boldsymbol{u} - \mathrm{med}_j \left(\boldsymbol{h}_j^\top \boldsymbol{u} \right) \right|}{1.486 \, \mathrm{med}_k \left| \boldsymbol{h}_k^\top \boldsymbol{u} - \mathrm{med}_j \left(\boldsymbol{h}_j^\top \boldsymbol{u} \right) \right|} \quad (29)$$

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 of the PS attains the maximum, $\epsilon^*_{\max} = [(m_t - n - 1)/2]/m_t$.

After computing the PS values for the elements of $\tilde{\boldsymbol{y}}_k$, the outliers must now be downweighted using a weight function given by $\varpi_i = \min\left(1, d^2/PS_i^2\right)$, where we pick $d = 1.5$ to yield good statistical efficiency at the Gaussian distribution without increasing the bias too much under contamination. The meaning of ϖ 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. Finally, we may perform prewhitening by multiplying $\mathbf{L}_k^{-1}$ into (27) on the left-hand side,

$$(\mathbf{L}_k^{-1})\tilde{\boldsymbol{y}}_k = (\mathbf{L}_k^{-1})\tilde{\mathbf{H}}_k \boldsymbol{s}_k + (\mathbf{L}_k^{-1})\tilde{\boldsymbol{e}}_k \tag{30}$$

which can then be put into the following form:

$$\boldsymbol{z}_k = \mathbf{A}_k \boldsymbol{s}_k + \boldsymbol{\eta}_k \tag{31}$$

*3) **Step III : 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 $\mathbf{F}$ and $\mathbf{H}$ can still negatively affect the filter solution through the error covariance matrix $\boldsymbol{\Sigma}$ and state estimation equations. Therefore, a robust technique is still needed that computes a weight matrix $\mathbf{Q}$ 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 $\boldsymbol{a}_i$ in (31), where $\boldsymbol{a}_i$ is the i^{th} column vector of the matrix. Of the several proposals available in the literature, here we explain the Schweppe-type GM-estimator [11] to solve for $\boldsymbol{s}_k$.

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

$$J(\boldsymbol{s}) = \sum_{i=1} \varpi_i^2 \rho(r_i/s\varpi_i) \tag{32}$$

where $\rho(.)$ represents a nonlinear function of standardized residuals; with the residuals r_i defined as $r_i = z_i - \boldsymbol{a}_i^\top \hat{\boldsymbol{s}}$; and the robust scale estimate is the median absolute deviation, defined as $s = 1.4826\text{median}_i|r_i|$. The constant 1.4826 is a correction factor for Fisher consistency at the Gaussian distribution. The Huber ρ-function is a good choice for the ρ function as it exhibits L_2-norm properties for small residuals and to L_1-norm properties for large residuals. The GM-estimator is obtained by setting the partial derivatives of the objective function in (32) to zero. This yields a system of nonlinear equations which can be solved using the IRLS algorithm yielding

$$\hat{\boldsymbol{s}}_{k|k}^{(v+1)} = \left(\mathbf{A}_k^\top \mathbf{Q}^{(v)} \mathbf{A}_k\right)^{-1} \mathbf{A}_k^\top \mathbf{Q}^{(v)} \boldsymbol{y}_k \tag{33}$$

In contrast to the direct noniterative computation of the linear Kalman filter, the IRLS algorithm or Newtons 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.

VI. Conclusion

In this paper we saw some methods for robust estimation. In many cases our usual assumptions like Gaussian nature of data, stationarity, linearity etc. do not hold in practical situations. So there is a need to develop statistics which do not rely on such assumptions, i.e. are robust to these. We saw some robust estimation methods for both single channel and multichannel data. We also highlighted a robust Kalman filter based estimation technique.

References

[1] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, “Robust estimation in signal processing,” IEEE Signal Proc. Magazine, May 2012.

[2] M. A. Gandhi and L. Mili, “Robust Kalman filter based on a generalized maximum-likelihood-type estimator,” IEEE Trans. Signal Processing, vol. 58, no. 5, pp. 25092520, 2010.

[3] D. S. Pham and A. M. Zoubir, “A sequential algorithm for robust parameter estimation,” IEEE Signal Processing Lett., vol. 12, no. 1, pp. 2124, 2004.

[4] U. Hammes, E. Wolsztynski, and A. M. Zoubir, “Transformation-based robust semiparametric estimation,” IEEE Signal Processing Lett., vol. 15, no. 1, pp. 845848, 2008.

[5] P. J. Huber and E. M. Ronchetti, “Robust Statistics” Hoboken, NJ: Wiley, 2009.

[6] S. A. Kassam and H. V. Poor, “Robust techniques for signal processing: A survey,” Proc. IEEE, vol. 73, no. 3, pp. 433481, 1985.

[7] “Robust regression: Asymptotics, conjectures and Monte Carlo,” Annals Statist., vol. 1, no. 5, pp. 799821, 1973

[8] D. S. Pham, A. M. Zoubir, and R. F. Brcich,“Adaptive ML signal detection in non-Gaussian channels, in Proc. IEEE Workshop on Statistical Signal Processing (SSP03), St. Louis, MO, 2003, pp. 4548.

[9] P. J. Rousseeuw and C. Croux, “Alternatives to the median absolute deviation, J. Amer. Statist. Assoc., vol. 88, no. 424, pp. 12731283, Dec 1993

[10] X. Wang and H. V. Poor,Robust multiuser detection in non-Gaussian channels, “IEEE Trans. Signal Processing,” vol. 47, no. 2, pp. 289305, Feb. 1999

[11] R. A. Maronna, R. D. Martin, and V. J. Yohai, “Robust Statistics: Theory and Methods”, New York: Wiley, 2006