[4] S. Verdu and H. V. Poor, "Minimax linear observers and regulators for stochastic systems with uncertain second-order statistics," *IEEE Trans. Automat. Contr.*, vol. AC-29, June 1984.

[5] H. V. Poor and D. P. Looze, "Minimax state estimation for linear stochastic systems with noise uncertainty," *IEEE Trans. Automat. Contr.* vol. AC-26, Aug. 1981.

[6] J. C. Darragh and D. P. Looze, "Noncausal minimax linear state estimation for systems with uncertain second-order statistics," *IEEE Trans. Automat. Contr.*, vol. AC-29, June 1984.

[7] Y. A. Phillis, "Estimation and control of systems with unknown covariance and multiplicative noise," *IEEE Trans. Automat. Contr.*, vol. 34, Oct. 1989.

[8] A. J. Krener, "Kalman–Bucy and minimax filtering," *IEEE Trans. Automat. Contr.* vol. AC-25, Apr. 1980.

[9] A. V. Borisov and A. R. Pankov, "Optimal filtering in stochastic discrete time systems with unknown inputs," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 2461–2464, Dec. 1994.

[10] R. K. Mehra, "On the identification of variance and adaptive Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. AC-15, Apr. 1970.

[11] R. Bellman, *Introduction to Matrix Analysis.* New York: McGraw-Hill, 1970.

[12] G. S. Rodgers, *Matrix Derivatives.* New York: Marcel Dekker, 1980.

[13] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering for linear discrete time non-Gaussian systems," *SIAM J. Contr. Opt.*, vol. 34, no. 5, Sept. 1996.

[14] ——, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *IEEE Trans. Automat. Contr.*, vol. 42, Aug. 1997.

# Robust Estimation with Unknown Noise Statistics

Željko M. Đurović and Branko D. Kovačević

***Abstract*—The equivalence between the Kalman filter and a particular least squares regression problem is established and the regression problem is solved robustly using a statistical approach named $M$-estimation. $M$-robust estimators are derived for adaptive estimation of the unknown *a priori* state and observation noise statistics simultaneously with the system states. The feasibility of the approach is demonstrated with simulation.**

***Index Terms*—Adaptive filtering, non-Gaussian noise, robust estimation.**

## I. Introduction

The Kalman filter is one of the most important developments of linear estimation theory [1]. It is widely used and is also of great theoretical interest [2]. Unfortunately, the distribution of the noise arising in applications frequently deviates from the assumed Gaussian model, often being characterized by heavier tails and generating high-intensity noise realizations, named outliers, in the presence of which the performance of the Kalman filter can be very poor [3]. Thus, there appears to be considerable motivation for considering filters that are robustified to perform fairly well in non-Gaussian environments, especially in the presence of outliers. Robust statistical procedures provide formal methods to spot the outlying data points and reduce their influence [4]. Most of the contributions in this area have been directed toward censoring data [3], namely, if an observation differs sufficiently from its predicted value then it is discarded. This type of procedure suffers from several faults, but the principal one is that it is often hard to distinguish an outlier from a large, but not unnatural, deviation. Masreliez [5] described an approximate conditional mean approach, based on the assumption that the state prediction probability density function (p.d.f.) is Gaussian, and derived an explicit update formula. However, the procedure to evaluate the score function of the residual process involves convolution operations that are difficult to implement, even if the statistics of the residual process are available. Facing this difficulty, Masreliez and Martin [3] applied the score function of the min–max robust theory [4], namely, they first used a linear transformation to scale and symmetrize the p.d.f. of the residual process and then operated on the normalized residual with the min–max score function, which has to cut off the outliers. Unfortunately, such a transformation does not exist in the general case. Other min–max robust estimators were also proposed by Doraiswami [6] and Morris [7]. Morris assumed that the noise covariance matrices can be bounded and derived a min–max filter. However, this filter is still linear, and therefore it is susceptible to outliers. Doraiswami, on the other hand, assumed such constraints as partial probability, partial covariance, or a combination for the observables, which essentially represent a severe limitation on the tail behavior of the p.d.f. and limit the robust qualities of the filter. In order to avoid the min–max dilemma, where different noise models lead to different estimators for the same noise, Tsai and Kurz [8] proposed the $m$-interval polynomial approximation method for estimating the score function of the approximate conditional mean filter. This method is computationally more feasible than the ones oriented toward approximating the unknown residual p.d.f. [9], but it still remains a complex computation, especially in the case of the multidimensional residual process. The extension of the concept of Huber's $M$-estimation approach [4] to the problem of robust Kalman filtering has been discussed by Kovačević *et al.* [10]. This approach is emphasized over the other ones, since it is motivated by the maximum likelihood estimation, which makes it more natural and rather easy. However, it represents a minimization problem and iterative numerical methods are required for its solution. Moreover, the $M$-approach is conservative, and it not only has a lower efficiency in Gaussian noise, but also may degrade otherwise [8], [10]. Therefore, some kind of adaptation to the underlying noise conditions is necessary. An approach to overcome these difficulties, based on a weighted least squares approximation of the $M$-robust state estimates and $M$-robust adaptive estimation of the unknown noise statistics, simultaneously with the system states, is considered in this paper.

Manuscript received November 15, 1995. Recommended by Associate Editor, T. Runolfsson.

The authors are with the Faculty of Electrical Engineering, University of Belgrade, 11000 Belgrade, Yugoslavia (e-mail: kovacevic@buef31.etf.bg.ac.yu).

Publisher Item Identifier S 0018-9286(99)03022-6.

## II. Problem Formulation

The system under consideration is supposed to be represented by the state-space model

$$x(k+1) = F(k)x(k) + G(k)w(k) \tag{1}$$

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

where $x(k)$ is the state vector at time $t_k = kT$, $z(k)$ is the observation vector at time $t_k$, $w(k)$ is the state noise at time $t_k$ and $v(k)$ is the observation noise at time $t_k$, with $T$ being the uniform time step. Moreover, $w(k)$ and $v(k)$ are zero-mean white noises satisfying

$$E\left\{\begin{bmatrix} w(j) \\ v(j) \end{bmatrix}\begin{bmatrix} w(k)^T \\ v(k)^T \end{bmatrix}^T\right\} = \operatorname{diag}\{Q(k)\delta_{kj},\, R(k)\delta_{kj}\} \tag{3}$$

where $E\{\cdot\}$ is the mathematical expectation and $\delta_{kj}$ is the Kronecker's delta symbol ($\delta_{kj} = 0$ for $k \neq j$ and $\delta_{kj} = 1$ for $k = j$). Here $F(k)$ is the state transition matrix, $G(k)$ is the state noise transition matrix, and $H(k)$ is the observation matrix. Let $\hat{x}(k/l)$ ($l = k-1, k$) denote the linear least squares estimate of $x(k)$ given $\{z(j), j \le l\}$, and let $P(k/l)$ denote the corresponding error covariance matrix. Then the standard Kalman filter recursions are [1], [2]

$$\hat{x}(k+1/k) = F(k)\hat{x}(k/k) \tag{4}$$

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

$$v(k) = z(k) - H(k)\hat{x}(k/k-1) \tag{6a}$$

$$K(k) = P(k/k-1)H^T(k)[H(k)P(k/k-1)H^T(k) + R(k))^{-1} \tag{6b}$$

$$\hat{x}(k/k) = \hat{x}(k/k-1) + K(k)v(k) \tag{7a}$$

$$P(k/k) = [I - K(k)H(k)]P(k/k-1). \tag{7b}$$

It is also assumed that the initial state $x(0)$ is a random variable, independent of the future noises $w(k)$ and $v(k)$, with zero-mean and covariance matrix $P(0)$. Thus, the filter can be initialized with $\hat{x}(0/0) = 0$ and $P(0/0) = P(0)$. Here $v(k)$ is the zero-mean residual vector with the covariance matrix $E\{v(k)v^T(k)\} = H(k)P(k/k-1)H^T(k) + R(k)$.

The Kalman filter can also be thought of as the solution to a particular weighted least squares problem, namely, rewrite (1) and (2) as

$$\begin{bmatrix} I \\ H(k) \end{bmatrix} x(k) = \begin{bmatrix} F(k-1)\hat{x}(k-1/k-1) \\ z(k) \end{bmatrix} + E(k) \tag{8}$$

where $I$ is an identity matrix and

$$\varepsilon(k) = \begin{bmatrix} F(k-1)(x(k-1) - \hat{x}(k-1/k-1)) + G(k-1)w(k-1) \\ -v(k) \end{bmatrix} \tag{9}$$

$$E\{E(k)E^T(k)\} = \begin{bmatrix} P(k/k-1) & 0 \\ 0 & R(k) \end{bmatrix} = S(k)S^T(k) \tag{10}$$

where $P(k/k-1)$ is given by (5). The solution for $S(k)$ in (10) can be obtained by using UD factorization or Cholesky decomposition [11]. Multiplying (8) by $S^{-1}(k)$, one obtains

$$Y(k) = X(k)\beta(k) + \xi(k) \tag{11}$$

where

$$X(k) = S^{-1}(k)\begin{bmatrix} I \\ H(k) \end{bmatrix}, \qquad \xi(k) = -S^{-1}(k)E(k)$$

$$\beta(k) = x(k), \qquad Y(k) = S^{-1}(k)\begin{bmatrix} \hat{x}(k/k-1) \\ z(k) \end{bmatrix}. \tag{12}$$

The relation (11) is in the form of a standard linear least squares regression problem with $E\{\xi(k)\xi^T(k)\} = I$. The solution to (11) is given by [1], [2]

$$\hat{\beta}(k) = (X^T(k)X(k))^{-1}X^T(k)Y(k) \tag{13}$$

$$E\{(\beta(k) - \hat{\beta}(k))(\beta(k) - \hat{\beta}(k))^T\} = (X^T(k)X(k))^{-1}. \tag{14}$$

It can be shown that the Kalman filter recursions (4)–(7) are the solutions (13) and (14) to the linear regression problem given by (11) and (12) [10]; that is $\hat{\beta}(k) = \hat{x}(k/k)$ and $P(k/k) = \operatorname{cov}\{\hat{\beta}(k)\}$. Unfortunately, the Kalman filter is nonrobust. The purely data-oriented version of the word robust is the word resistant [12]. An estimate is called resistant if changing a small fraction of the data by large amounts results in little change to the estimate. This property is one of insensitivity to outliers. In addition, one may also insist that small changes in most of the data result in only small changes in the estimate. This requirement is one of insensitivity to rounding, grouping, and quantization errors. Moreover, the term robustness also has a probabilistic meaning, and at least three distinct probabilistic notations of robustness can be perceived. The oldest and most accessible one is that of efficiency robustness. Namely, an estimator is said to be efficiency robust if it has high efficiency at a nominal Gaussian model and high efficiency at a variety of strategically chosen non-Gaussian distributions [12]. Another notation is that of min–max robustness over a family of distributions. Typically, the family of distribution is infinite, and asymptotic variances are used as a quantitative performance measure [4]. Finally, the third form of robustness is qualitative robustness [13]. This is, in fact, a continuity requirement which is the probabilistic embodiment of the notion that small changes in the data should produce only small changes in the estimates. An important concept of this robustness is that of the influence curve [13], which measures the perturbation of the estimate caused by a single additional observation, the so-called contamination. Of course, qualitative robustness alone is hardly sufficient, since some ridiculous estimates, such as constants, independent of the data, are clearly qualitative robust. Therefore, efficiency robustness, in addition to qualitative robustness, is required. Unfortunately, the highly technical character of qualitative robustness makes it relatively inaccessible to applied workers.

## III. $M$-Robust Approach to Robustifying the Kalman Filter

In light of the above robustness concepts and the equivalence between the Kalman filter (4)–(7) and the linear regression solution (13) and (14), the Kalman filtering problem can be considered as a linear regression problem (11), which can be solved in a resistance and efficiency robust manner. This can be done by using the $M$-estimate [4], which is defined as the minimization problem

$$\hat{\beta}(k) = \arg\min_{\beta} J_n(\beta), \qquad J_n(\beta) = \sum_{i=1}^{n} \rho(y_i(k) - x_i^T(k)\beta) \tag{15}$$

where $y_i$(k) is the $i$th element of $Y(k)$, $x_i^T(k)$ is the $i$th row of $X(k)$, and $n$ is the dimension of $Y(k)$ in (11). Here $\rho(\cdot)$ is a robust score function that has to cut off the outliers. Particularly, if one chooses $\rho(\cdot)$ to be a quadratic function, the estimate (15) reduces to the least squares or Kalman filter solution (13). This algorithm is optimal when the underlying noise p.d.f. $p(\xi)$ is a Gaussian one [14]. However, due to the existence of some outliers, the real noise p.d.f. for large samples is no longer Gaussian. This means that we have to base the choice of $\rho(\cdot)$ on a p.d.f. that gives higher probability for large samples than a Gaussian one. Since comparatively high efficiency at the Gaussian samples is also desired, $\rho(z)$ should look like $z^2$ for small values of $z$ but increases more slowly for large values of $z$, in order to suppress the influence of outliers. Moreover, it is desirable that $\Psi(\cdot) = \rho'(\cdot)$, named the influence function [4], [12], be bounded and continuous. Boundedness ensures that no single observation can have an arbitrarily large influence on estimates, while continuity ensures that many rounding or quantization errors will not have major effects. This is, in fact, a resistant robustness requirement and corresponds, for example, to the choice of the Huber's score function [4]

$$\rho(z) = \begin{cases} \Delta|z| - \Delta^2/2, & |z| \geq \Delta \\ z^2/2, & |z| < \Delta \end{cases} \tag{16}$$

where $\Delta$ has to be chosen to give the desired efficiency at the Gaussian model. Other $\rho(\cdot)$ functions that are commonly used in robust estimation are given in [14]. Equating the first partial derivatives with respect to the elements of $\hat{\beta}(k)$, say $\hat{\beta}_j(k)$, $j = 1, \cdots, p$, equal to zero, we see that this is equivalent to finding the solution associated with the $p$ equations

$$\sum_{i=1}^{n} x_{ij}(k)\Psi(y_i(k) - x_i^T(k)\hat{\beta}(k)) = 0, \qquad j = 1, \cdots, p \tag{17}$$

where $x_{ij}(k)$ is the element in the $i$th row and $j$th column of $X(k)$. The solution of (17) becomes cumbersome because of the fact that $\Psi(\cdot)$ is nonlinear [10]. However, (17) can be expressed as a weighted least squares approximation

$$\sum_{i=1}^{n} x_{ij}(k)\omega_i(k-1)(y_i(k) - x_i^T(k)\hat{\beta}(k)) \approx 0, \qquad j = 1, 2, \cdots, p \tag{18}$$

where the weights $\omega_i(k-1)$ are given by

$$\omega_i(k-1) = \begin{cases} \Psi(y_i(k) - x_i^T(k)\hat{\beta}(k-1)) \\ \quad \cdot (y_i(k) - x_i^T(k)\hat{\beta}(k-1))^{-1}, & y_i(k) \neq x_i^T(k)\hat{\beta}(k-1) \\ 1, & y_i(k) = x_i^T(k)\hat{\beta}(k-1). \end{cases} \tag{19}$$

Now, the solution of (18) is given by

$$\hat{\beta}(k) = (X^T(k)\Omega(k-1)X(k))^{-1}X^T(k)\Omega(k-1)Y(k) \tag{20}$$

where $\Omega(k-1) = \mathrm{diag}\{\omega_1(k-1), \cdots, \omega_n(k-1)\}$, with $\mathrm{diag}(\cdot)$ being a diagonal matrix.

In this way, we are suggesting that the Kalman filter recursion (7a) at each step be replaced by an implicit update formula (20), based on the Huber's robust loss function (16). This entails more work at each step, but usually not so much more as to preclude its consideration. However, these estimates should be much better in practice than the least squares estimates.

The problem of updating the covariance can be dealt with in at least two different ways: 1) Use the standard Kalman recursions (5) and (7b). This is a reasonable strategy if the nonlinearity $\rho(\cdot)$ is almost like a quadratic function, similarly as in (16). 2) Use some robust alternatives to (14), instead of the recursion (7b). Several alternatives that achieve acceptable performance, especially when the number of equations is large compared to the number of variables, are discussed in [4] and [12]. Unfortunately, this approach is not suitable in the filtering problem under consideration so that we shall concentrate on the first one.

## IV. $M$-Robust Estimators for the Noise Statistics

The implementation of the Kalman filter, robustified or not, requires the first- and second-order moments of the noise processes to be known *a priori*. However, the *a priori* statistics for the stochastic disturbances in both the state and observation processes are not known exactly, owing to the modeling errors and the presence of outliers in the observation sequence. Therefore, it is desirable to estimate adaptively the unknown *a priori* noise statistics simultaneously with the system states. On the other hand, in most practical filtering applications the measurements are independent. These components may be processed one at a time, and the observation becomes one-dimensional, that is $z(k)$ in (2) is a scalar quantity. For the observation noise statistics, consider the measurement relation (2). The true state is unknown, so $v(\cdot)$ cannot be determined from (2), but an intuitive approximation for $v(i)$ is given by the quantity

$$\begin{aligned} r(i) &= z(i) - H(i)\hat{x}(i/i) \\ &= z(i) - H(i)\hat{\beta}(i), \qquad i = k-N+1, \cdots, k \end{aligned} \tag{21}$$

where $\hat{\beta}(i)$ is the robust $M$-estimate (20) at time $t_i = iT$, and $r(i)$ is defined as the observation noise sample. If $r(i)$ for $i = k-N+1, \cdots, k$ are assumed to be independent and identically distributed, a simple parameter estimation problem can be constructed [15]. Define a random variable $\mathcal{R}$ on the sample space $\Omega_r$ from which the data $r(i)$, $i = k-N+1, \cdots, k$ are obtained. Based on the empirical measurements, the mean $\overline{r}(k)$ and the variance $C_r(k)$ of the unknown distribution of $\mathcal{R}$ are to be estimated. Unfortunately, the sample mean and the sample variance, like the Kalman filter, lack robustness toward outliers [4]. Therefore, we propose to use the $M$-estimate of location parameter of the unknown distribution, as an alternative to the sample mean. As in (17), the $M$-robust estimate $\hat{r}(k)$ of $\overline{r}(k)$ is defined by [4]

$$\sum_{i=k-N+1}^{k} \Psi\left(\frac{r(i) - \hat{r}(k)}{\hat{d}(k)}\right) = 0 \tag{22}$$

where $\Psi(\cdot) = \rho'(\cdot)$ corresponds to the Huber nonlinearity in (16). Here $\hat{d}(k)$ is an estimate of the scale of the data $\{r(i)\}$. Although *ad hoc*, a popular robust statistics $\hat{d}$ is the median of the absolute median deviations

$$\hat{d}(k) = \text{median}\frac{|r(i) - \text{median}(r(i))|}{0.6745}, \qquad i = k-N+1, \cdots, k. \tag{23}$$

The divisor 0.6745 is used because then $\hat{d} \approx C_r^{1/2}$ if $N$ is large and if the sample actually arises from a normal distribution. The particular scheme of selecting $\hat{d}$ suggests appropriate values of the tuning constant $\Delta$ in (16), namely, because $\hat{d}$ is close to the standard deviation, $\Delta$ is usually taken to be approximately the value of 1.5 [4]. When $\Delta = 1.5$ we refer to this procedure as the (1.5)-Huber $M$-estimator. If the standard deviation is known, the asymptotic efficiency of this estimator under normal assumptions is greater than 95%, and in most heavy tailed situations it performs extremely well [4]. The estimating equation (22) is, of course, nonlinear, and an iterative procedure or some form of a weighted least squares approximation, similar to (20), can be used for its solution.

There are at least two possibilities for robust estimation of the observation noise variance $C_r$. One obvious choice is $\hat{C}_r = \hat{d}^2$. However, two points should be noted here. The first is that although $\hat{d}$ in (23) is robust, it turns out to be less efficient than some other robust estimates of variance. The second point is that $\hat{d}$ is a nuisance parameter in the computation of $\overline{r}$, and in this context the efficiency issue is not as crucial as in the estimation of the variance of the data, estimates of the latter being used in setting the Kalman filter gain. A more efficient estimator of the data variance should be based on the asymptotic variance formula for the location $M$-estimate (22). When $\hat{d} = C_r^{1/2}$, this formula is given by [4]

$$V = \lim_{N\to\infty} E\{N(\hat{r} - \overline{r})^2\} = \frac{C_r E\left\{\Psi^2\left[\dfrac{r(i) - \overline{r}}{C_r^{1/2}}\right]\right\}}{E^2\left\{\Psi'\left[\dfrac{r(i) - \overline{r}}{C_r^{1/2}}\right]\right\}}. \tag{24}$$

A natural estimate of $V$ in (24) is

$$\hat{V}(k)=\frac{\hat{d}^2(k)\left\{(1/N)\sum_{i=k-N+1}^{k}\Psi^2[(r(i)-\hat{r}(k))/\hat{d}(k)]\right\}}{\left\{(1/N)\sum_{i=k-N+1}^{k}\Psi'[(r(i)-\hat{r}(k))/\hat{d}(k)]\right\}^2}. \quad (25)$$

This $\hat{V}(k)$ would appear to be a reasonable estimate of $C_r(k)$. Therefore, given the estimates $\hat{r}(k)$ and $\hat{C}_r(k)$, it follows from (2) and (21) that the first-and second-order moments of the observation noise $\overline{v}(k)$ and $R(k)$ can be estimated as

$$\hat{\overline{v}}(k)=\hat{r}(k), \qquad \hat{R}(k)=\hat{C}_r(k)-HP(k/k)H^T. \quad (26)$$

Similarly, for the state noise statistics, consider the dynamic state relation (1) at a given time $t_k = kT$. The true states $x(k)$ and $x(k-1)$ are unknown, and $w(k-1)$ cannot be determined, but an intuitive approximation for $w(k-1)$ is

$$\begin{aligned} q(k) &= [G^T(k)G(k)]^{-1}G^T(k)(\hat{x}(k/k)-F(k)\hat{x}(k-1/k-1)) \\ &= [G^T(k)G(k)]^{-1}G^T(k)(\hat{\beta}(k)-F(k)\hat{\beta}(k-1)) \end{aligned} \quad (27)$$

where $q(k)$ is defined as the state noise sample at time $t_k = kT$, and $\hat{\beta}(l)$ ($l = k-1, k$) is the robust $M$-estimate (20) at time $t_l$. If $q(k)$ are assumed to be representative of $w(k)$, they may be considered independent and identically distributed. Again, defining a parameter estimation problem [15], let $Q$ be a random variable on the sample space $\Omega_q$ from which is obtained the data $\{q(i)\}$, $i = k-N+1, \cdots, k$. Based on these measurements, the mean $\overline{q}$(k) and the variance $C_q$(k) of the unknown distribution of $Q$ are to be estimated. Following the same steps as in the case of observation noise, an unbiased robust estimate $\hat{q}(k)$ of $\overline{q}(k)$ is given by (22), where $r(i)$ and $\hat{r}(k)$ are replaced by $q(i)$ and $\hat{q}(k)$, respectively. Moreover, (25) represents a reasonable robust estimate $\hat{C}_q(k)$ of $C_q(k)$ when $r(i)$ and $\hat{r}(k)$ are replaced by $q(i)$ and $\hat{q}(k)$, respectively. Furthermore, under (1) and (27), the first- and second-order moments of the state noise $\overline{w}(k)$ and $Q(k)$ can be estimated as

$$\begin{aligned} \hat{\overline{w}}(k) &= \hat{q}(k) \\ \hat{Q}(k) &= \hat{C}_q(k)-[G^T(k)G(k)]^{-1}[F(k)P(k-1/k-1)F^T(k) \\ &\quad -P(k/k)]. \end{aligned} \quad (28)$$

In summary, estimators for the first- and second-order moments of the noise processes are based upon the last $N$ of state and observation noise samples at time $t_k$, generated by the $M$-robust filter (20). This procedure requires some extra storage and a shifting operation on the noise samples. It is obvious that the noise variance estimators may become negative in numerical applications, especially at the beginning steps. For this reason, these quantities are always reset to the absolute values of their estimates.

## V. Experimental Results

A rigorous analysis of the convergence of the proposed algorithms is very difficult, owing to both the multivariable dynamic time-varying nature of the model and a nonlinear form of the robust algorithm itself. Also, it is extremely difficult to make a precise judgment on both their practical robustness, characterized by ability to reject outliers and insensitivity to *a priori* assumptions (nonlinearity form, noise statistics, and initial conditions), and their tracking capabilities. Therefore, extensive simulations have been undertaken, in order to clarify the properties of the algorithms and to make necessary comparisons. In the following characteristic examples, obtained on the three-state track model [16] with position only measurements given by [1], [2]

$$F=\begin{bmatrix}1 & T & T^2/2\\ 0 & 1 & T\\ 0 & 0 & 1\end{bmatrix}, \qquad G=\begin{bmatrix}0\\0\\1\end{bmatrix}, \qquad H=[1 \quad 0 \quad 0] \quad (29)$$

will be presented. Here $x(k)$ is the position, velocity, and acceleration state vector; $z(k)$ is the position measurement; $w(k)$ is the maneuver-noise; $v(k)$ is the measurement noise; $F$ is the transition matrix for uniform time step $T$ ($T = 4$ s) and $H$ is the measurement matrix [16]. The sequence $\{w(k)\}$ is adopted to be zero-mean white Gaussian with the unit variance, and the random variable $v(k)$ is generated from the heavy-tailed p.d.f.

$$p(\cdot) \sim (1-\varepsilon)\mathrm{N}(\cdot|0, 1)+\varepsilon\mathrm{N}(\cdot|0, \sigma_o^2), \qquad 0 \le \varepsilon \le 1, \qquad \sigma_o^2 > 1. \quad (30)$$

Here $N(\cdot|0, \sigma_o^2)$ denotes the normal p.d.f. with zero-mean and variance $\sigma_0^2$. In monopulse radars this heavy-tailed behavior is presented because of target glint [16], namely, the heavy-tailed aspect of the underlying noise p.d.f. is associated with the large glint spikes. These are in effect outliers, which contaminate the expected normal measurements and have a considerable influence on conventional linear estimates, such as the Kalman filter or least squares.

The following algorithms have been tested: Kalman filter, denoted as A1; Kalman filter with adaptive estimation of the noise statistics using the (1.5)-Huber procedure (22), denoted as A2; robust estimation procedure (20) with the nonlinearity (16), denoted as A3; robust estimation procedure (20) with adaptive estimation of the noise statistics based on the (1.5)-Huber procedure (22), denoted as A4.

Simulation results are compared in terms of the estimated noise statistics and cumulative estimation error (CEE):

$$\mathrm{CEE}(k)=\frac{1}{k}\sum_{i=1}^{k}\frac{||\hat{\beta}(i)-\beta(i)||}{||\beta(i)||} \quad (31)$$

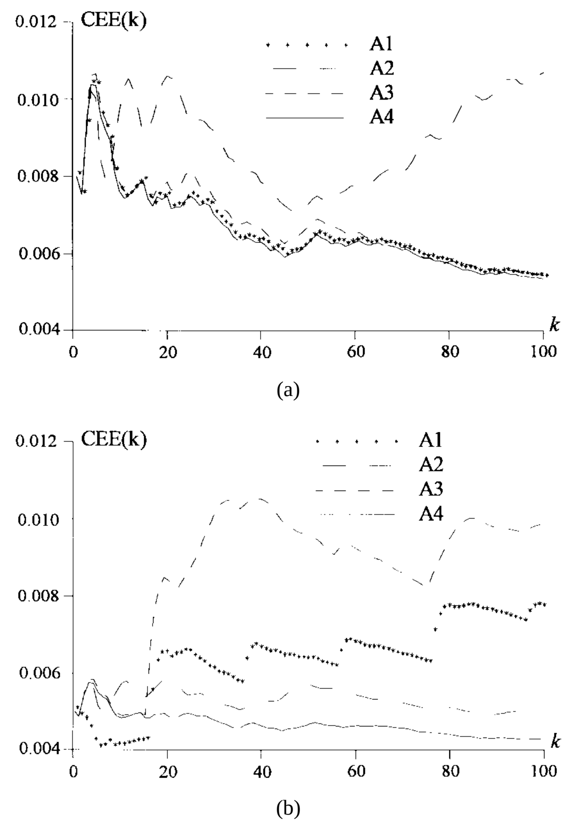


Fig. 1. Estimation error comparison of robust and nonrobust estimators in the case of: (a) Gaussian noise ($\varepsilon = 0$) and (b) contaminated Gaussian noise ($\varepsilon = 15\%$).

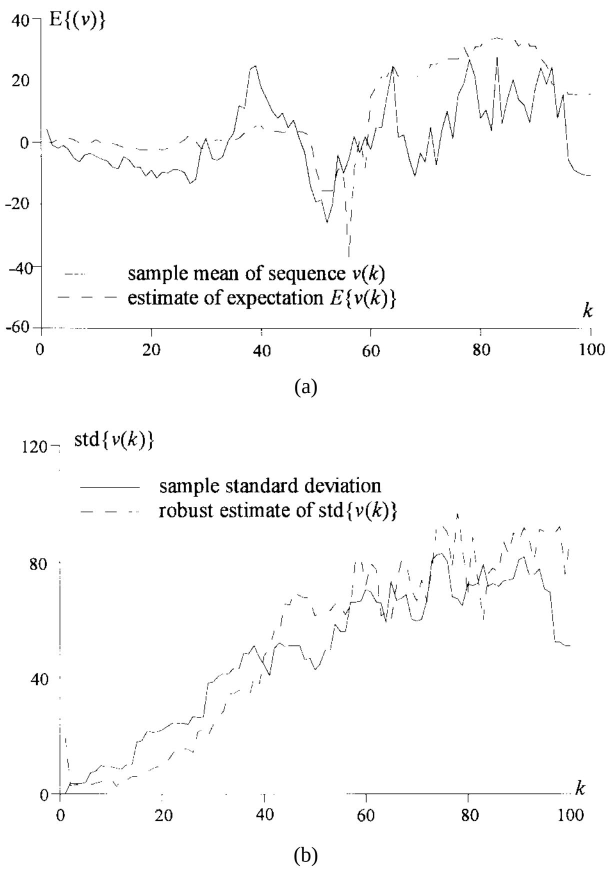


Fig. 2. Estimates of the measurement noise statistics: (a) the first-order moment (mean value) and (b) the second-order moment (standard deviation).

where $||\cdot||$ is the Euclidean norm. The CEE results are plotted in Fig. 1, while the estimated first- and second-order moments of the measurement noise process are illustrated in Fig. 2. The robust estimates of the measurement noise statistics are compared with the sample mean and the sample variance, obtained on the sliding window of size $N = 25$ when there are no outliers.

Note that the algorithm A1 is slightly better than the $M$-robust filter A3 in the case of Gaussian noise [Fig. 1(a)], and this is consistent with the optimality of the Kalman filter in the Gaussian case. Moreover, the adaptive filters, robustified or not, based on the $M$-robust estimation of the noise statistics, simultaneously with the system states, can improve the state estimation performance when *a priori* statistics of the Gaussian measurement noise are erroneous and there exist significant dynamical model errors, owing to the target maneuver [Fig. 1(b)]. Specifically, as the estimates of observation noise statistics become more accurate (Fig. 2), the filter adaptation provides the necessary corrections of the Kalman filter gains, which vary inversely with the observation noise variances, and this, in turn, results in good tracking capability. On the other hand, in the real situation concerned there is no state noise, that is, its variance is equal to zero; but in the presence of target maneuver the state model is inadequate, and we compensate for a such model error by introducing the state noise with the corresponding variance. Thus, the adaptive state noise variance estimates provide for the desired filter correction during the maneuver, when the state model is not adequate. Furthermore, the $M$-robust filters A3 and A4 are superior to the linear estimators A1 and A2 in the presence of measurement noise outliers in the form of glint spikes [Fig. 1(b)]. Finally, algorithm A4 provides the best performances (Fig. 1). The reason lies in the fact that only filter robustification, based on a suitably chosen nonlinear transformation of the normalized residuals, combined with its adaptation, using $M$-robust estimation of the unknown noise statistics simultaneously with the system states, provides a good compromise between noise immunity and tracking capability.

## VI. Conclusion

In this paper we have obtained feasible Kalman filter modifications that perform well in the presence of disturbance uncertainty and nonmodeled dynamics. We have shown in a realistic simulation problem of tracking maneuvering targets that the adaptive robustified Kalman filter, based on a suitably chosen nonlinear transformation of the normalized residuals and $M$-robust estimation of the unknown noise statistics, simultaneously with the system states, can improve the state estimation performance in the presence of erroneous *a priori* noise statistics and significant dynamical model errors. The improvement entails more work at each step, but usually not to preclude the consideration. On the contrary, these estimates should be much better in practice than the least square or Kalman filter estimate. Therefore, the proposed adaptive robustified estimator may offer a desirable alternative to other adaptive approaches, particularly those that involve state vector augmentation or estimation of the unknown residual distribution. Finally, the emphasized adaptive $M$-robustified procedure is the one that, with some modifications, can be used whenever the least squares or Kalman filter is used.

## Acknowledgment

The authors would like to express their sincere thanks to the referees for valuable suggestions which improved the final manuscript.

## References

[1] B. D. J. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1978.
[2] A. Gelb, *Applied Optimal Estimation.* Cambridge, MA: MIT Press, 1974.
[3] C. J. Masreliez and R. D. Martin, "Robust Bayesian estimation for linear model and robustifying the Kalman filter," *IEEE Trans. Automat. Contr.,* vol. 22, pp. 361–371, 1977.
[4] P. J. Huber, *Robust Statistics.* New York: Wiley, 1973.
[5] C. J. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," *IEEE Trans. Automat. Contr.,* vol. 22, pp. 107–111, 1975.
[6] R. Doraiswami, "A decision theoretic approach to parameter estimation," *IEEE Trans. Automat. Contr.,* vol. 21, pp. 860–865, 1976.
[7] J. M. Morris, "The Kalman filter: A robust estimator for some classes of linear quadratic problems," *IEEE Trans. Inform. Theory,* vol. 22, pp. 526–534, 1976.
[8] C. Tsai and L. Kurz, "An adaptive robustizing approach to Kalman filtering," *Automatica,* vol. 19, pp. 279–88, 1983.
[9] D. L. Alspach and W. Sorenson, "Recursive Bayesian estimation using Gaussian sums," *Automatica,* vol. 7, pp. 465–475, 1971.
[10] B. D. Kovačević, M. Đurović, S. T. Glavaški, "On robust Kalman filtering," *Int. J. Contr.,* vol. 3, pp. 547–562, 1992.
[11] G. J. Bierman, *Factorization Method for Discrete Sequential Estimation.* New York: Academic, 1977.
[12] J. W. Tukey, *Exploratory Data Analysis.* Reading, MA: Addison Wesley, 1977.
[13] F. R. Hampel, "A general qualitative definition of robustness," *Ann. Math. Stat.,* vol. 42, pp. 1887–1895, 1971.
[14] Ya. Z. Tsypkin, *Foundations of Informational Theory of Identification.* Nauka, 1984.
[15] K. A. Myers and B. D. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Trans. Automat. Contr.,* vol. 21, pp. 520–523, 1976.
[16] S. S. Blackman, *Multiple Target Tracking with Radar Applications.* Washington, DC: Artech House, 1986.