ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20200929143934
Original scientific paper

Robustified Kalman Filtering Using Both Dynamic Stochastic Approximation and M-Robust Performance Index

Zoran BANJAC*, Branko KOVAČEVIĆ

Abstract: In this paper, the problem of designing the feasible Kalman filter under a non-Gaussian stochastic environment characterized by spiky noise or outliers has been considered. Firstly, the similarity among a class of dynamic stochastic approximation algorithms and the standard Kalmanfilter is found. Moreover, the particular dynamic stochastic approximation algorithm is derived by minimizing the generalized M-robust performance index. The adopted robust criterion represents the conditional expectation of a suitably chosen non-linear transformation of the measurement residuals, given the predicted system states and the observation sequence. The standard Kalman filtering time-update recursions are used to account for the predicted changes in the system states at each stage. Furthermore, the speed of algorithm convergence is improved by choosing the gain matrix from the minimization of an additional criterion at each stage, resulting in an approximately minimum variance algorithm. A target tracking scenery is simulated to demonstrate the practical robustness of the proposed state estimator.

Keywords: impulsive noise; Kalman filtering; non-Gaussian noise; nonlinear filters; robust estimation; state estimation

1 INTRODUCTION

The state estimation problem is one of determining the vector valued function of measurements, in order to estimate the value of system state for some fixed time instance, [1-4]. In general, the state is a multidimensional stochastic process, which is not directly accessible to observations. A common approach is based on the application of a scalar-valued admissible loss, being a symmetric, non-decreasing function of the random estimation error, that is equal to zero for the zero arguments, [3]. An estimate that minimizes the mean value of random loss is the optimal one. The fundamental result of estimation theory states that the optimal estimate for all admissible loss functions is the conditional mathematical expectation, assuming a symmetric and convex conditional probability distribution of random states given the measurements, [3]. Particularly, if the mean square error (MSE) performance index is adopted the only requirement is that the continuous conditional probability density function (pdf) exists at almost all arguments, [1-4]. Another important case is related to the Gaussian random processes, when the conditional expectation is a linear function of the measurements. Furthermore, if the state estimator design is restricted to a class of linear functions then only the first and second order moments define the optimal estimator, [3].

Particularly, one of the most significant achievements in the linear estimation is Kalman filtering. Except for the theoretical interest, it is also important for the practical applications, [1-4]. Kalman filter provides for optimal state estimates under the conditions that the state vector satisfies a linear discrete-time stochastic difference equation driven by the Gaussian random excitation, while the observation vector is linearly dependent upon the system states, and is corrupted by the additive Gaussian noise. In addition, since the linear Kalman filter is optimal for all admissible loss functions, it is also optimal in the minimum variance sense. Furthermore, one of the most significant practical features of the Kalman filter is its recursive form, a property that makes it extremely useful in real-time signal processing, [1-4].

On the other hand, for non-linear system dynamics or non-Gaussian noise the optimal estimator design may be too complex. In such situation the optimal linear Kalman filter may cease to work. Therefore, one might be also interested in a class of estimators which are not optimal in any statistical sense, but which provide for bounded statistical errors in practice. Particularly, a class of dynamic stochastic approximation algorithms performs quite well in many applications related to estimation optimisation and pattern classification, [5-9].

Additionally, in many practical applications the real noise distribution differs from the supposed normal one by heavier tails. This, in turn, generates the spiky noise realizations named outliers [9]. Since the Kalman filter is a linear function of observations, it is susceptible to outliers. Therefore, it is of great practical interest to designing robustified Kalman filtering techniques that are able to cope with impulse noise or outliers. In such circumstances, robust statistical methods provide suitable tools to spot bad data points and suppress their effects [9-11]. Particularly, the M-robust statistical approach is widely used in practice, since it represents an approximation of the maximum likelihood estimation that is easy to implement, [10]. In recent literature, there exist a number of articles devoted to robustifying the Kalman filter using the M-robust approach, [12-20].

In this article a new simple but efficient robustified version of Kalman filter has been proposed. This approach is based on an extension of the Huber's M-robust estimation concept to the problem of robustifying dynamic stochastics approximation methods, in the sense of its insensitivity to impulsive noise or outliers. The proposed M-robust algorithm is combined with both the mean-square optimal prediction, and calculation of the gain matrix from the minimization of an additional criterion, in order to derive an approximately minimum variance robust state estimator, the so-called M-robustified Kalman filter.

The paper is organised as follows. A brief description of optimality and robustness in linear dynamic system state estimation is given in Section 2. Section 3 is dedicated to the Huber's M- robust estimation, and its application to robustifying dynamic stochastics approximation recursive algorithm. Robustified Kalman filter using both M-robustified dynamic stochastic approximation and mean-square optimal state prediction is overviewed in Section 4, while its complete derivation is given in appendix A. An alternative relation for the gain matrix of robustified Kalman filer is derived in appendix B. Experimental analysis is presented in Section 5, while conclusions are presented in Section 6.

2 PROBLEM FORMULATION

Let us assume that an abstract linear discrete stochastic system is given by the state-space model

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

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

Here $x(k)$ is the random state vector, $y(k)$ is the observation or measurement vector, $w(k)$ is the state noise or disturbance, and $v(k)$ is the additive measurement noise, at the discrete time indexed by k. Furthermore, $w(k)$ and $v(k)$ are zero-mean noises that are uncorrelated by themselves and mutually, satisfying

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

with $E\{\cdot\}$ being the mathematical expectation, δ_{kj} denotes the Kronecker's delta symbol, and $\operatorname{diag}\{\cdot,\cdot\}$represents the block-diagonal matrix. In addition, $F(k)$, $G(k)$, and $H(k)$ are the given matrices. If $\hat{x}(k\,|\,l)$, $l=k-1,k$, is the linear optimal estimate of the state, $x(k)$, when the measurements, $\{y(j), j\le l\}$, are given, while $P(k|l)$ denotes the underlying estimation error covariance matrices, and then the classical Kalman filter equations are the following, [1-4]

Time update (prediction stage):

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

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

Measurement update (estimation or correction stage):

$$\varepsilon(k)=y(k)-H(k)\hat{x}(k\,|\,k-1) \tag{6}$$

$$K(k)=P(k\,|\,k-1)H^T(k)\cdot \cdot\left[H(k)P(k\,|\,k-1)H^T(k)+R(k)\right]^{-1} \tag{7}$$

$$\hat{x}(k\,|\,k)=\hat{x}(k\,|\,k-1)+K(k)\varepsilon(k) \tag{8}$$

$$P(k\,|\,k)=\left[I-K(k)H(k)\right]P(k\,|\,k-1) \tag{9}$$

with I being the identity matrix. It is also assumed that the initial state vector, $x(0)$, is a random vector with zero-mean, $E\{x(0)\}=m_0=0$, and the corresponding covariance matrix $P_0=E\{x(0)x^T(0)\}$. Moreover, the initial state is also uncorrelated with the future noises $w(k)$ and $v(k)$. Thus, the filter can be initialized with $\hat{x}(0\,|\,0)=m_0$ and $P(0\,|\,0)=P_0$. This, in turn, results in the unbiased state estimate at each step, that is $E\{\hat{x}(k\,|\,k)\}=E\{x(k)\}$, [3, 4]. It should be also noted that the measurement residual or innovation, $\varepsilon(k)$, represents the zero-mean, $E\{\varepsilon(k)\}=0$, uncorrelated random sequence with the covariance matrix, [3, 4]

$$E\left\{\varepsilon(k)\varepsilon^T(i)\right\}=S(k)\delta_{ki}; \quad S(k)=H(k)P(k\,|\,k-1)H^T(k)+R(k) \tag{10}$$

In most applications the elements of the observation vector, $y(k)$, are independent, and they can be analysed sequentially, one by one. As a consequence, $y(k)$ and $\varepsilon(k)$ are zero-mean, scalar, real random variables, having the variances $R(k)$ and $S(k)$, respectively. The important feature of the procedure is the requirement for decorrelation of the observation vector components, providing a diagonal covariance matrix of the observation noise. This task can be performed by using Choleskey factorization or UD decomposition, [1, 4]. Moreover, in a number of practical situations, the state disturbance, $w(k)$, is also assumed to be zero-mean scalar real random variable having the variance, $Q(k)$, [1-4].

Unfortunately, if the noise statistics are erroneous or unknown, the design of underlying optimal estimate can be impossible [3, 4, 8, 20]. Thus, the Kalman estimator is susceptible to departures of the noise statistics from the adopted Gaussian one, or it is not robust [12-18]. Formal definitions of robustness are given in the literature [9-11], but these are inconvenient for practical workers. In this sense, simple intuitive and empirical definitions are more suitable. Particularly, the two types of such definitions are popular in practice, the so-called resistant and efficiency robustness. Thus, an estimation procedure is robust in the resistant sense if it stays finite when some of the observations become too large in positive or negative direction. On the other hand, an estimator is efficiency robust if it possesses a high efficiency for the Gaussian distribution, but it also remains acceptably efficient when the real distribution differs from the assumed Gaussian one by heavier tails. This, in turn, generates the outlying data points contaminating the normally distributed observations. Therefore, the practical robustness combines both the resistant and efficiency features. Since the Kalman filter is linear in observations, it is not robust in the practical sense. In addition, the Huber's M-robust approach is preferred in practice, because it is based on the maximum likelihood approach, making it easy to understand and implement. This approach is used in the sequel to designing M-robustified dynamic stochastic approximation, in order to apply it for measurement update in the predictor-corrector structure of the Kalman filter, resulting in its M-robustified version.

3 M-ROBUST APPROACH TO DESIGNING A CLASS OF STOCHASTIC APPROXIMATION ALGORITHMS

Many robust schemes are derived by considering a problem of parameter estimation using linear scalar observations

$$y_k = h_k^T x_k + v_k \tag{11}$$

Here x is unknown parameter vector to be estimated, h_k is the given time-varying vector, and v_k is zero-mean white measurement noise. A situation of multidimensional observations can be handled by processing sequentially the individual components of y_k, one by one. Moreover, the M-robust statistical approach is based on the minimization of the performance index [9-11]

$$J(\hat{x}) = E\left\{\rho\left(\varepsilon_k(\hat{x})\right)\right\};\ \varepsilon_k(\hat{x}) = y_k - h_k^T \hat{x} \tag{12}$$

with $\rho(\cdot)$ being a penalty or loss function, reducing the influence of outliers. Since it is commonly supposed that the observation noise is confined to the normal distribution, large efficiency at the normal noise is required. For this reason, $\rho(x)$ function has to be quadratic in the middle. Furthermore, as a rule, the measurements data contain 5 to 10 percentage of outliers, [9]. As a consequence, it is also required that its derivative named the influence function, $\Psi(\cdot) = \rho'(\cdot)$ be bounded and continuous, [10, 11]. Bounded feature obeys the requirement that a particular outlier may not have a significant influence on estimates. On the other hand, the continuity feature ensures that grouped, or patchy, outliers will not produce large impact. This is the requirement for resistant feature. Both properties are fulfilled for the saturation type nonlinearity, the so-called Huber's loss, [10]

$$\rho(x) = \begin{cases} \Delta|x| - \Delta^2/2; & |x| \ge \Delta \\ x^2/2; & |x| < \Delta \end{cases} \tag{13}$$

where Δ is the free parameter to be adopted. The choice of this parameter has to provide for desired efficiency at the normal noise, but also to remain high efficiency in the presence of outliers. Experimental results have shown that the choice of $\Delta = 1.5$ satisfies the efficiency robustness requirements. This estimation procedure is called the Huber's 1.5-M-robust estimator [10]. By differentiating the performance index (12) with respect to the elements $\hat{x}_i$, $i = 1, 2, \ldots n$, of $\hat{x}$, a nonlinear algebraic system of the n equations is obtained, corresponding to the vector equation

$$g(\hat{x}) = \nabla_{\hat{x}} J(\hat{x}) = -E\left\{\Psi\left(\varepsilon_k(\hat{x})\right)\right\} h_k = 0 \tag{14}$$

The relation (14) may be used as the generator of robust stochastic approximation recursive scheme. It should be noted that stochastic approximation was firstly proposed as the iterative procedure for finding a root, x_0, of the algebraic nonlinear equation $g(x) = 0$, when $g(x_k)$ cannot be obtained from estimate, $\hat{x}_k$, of x, [5]. This is equivalent to the deterministic problem of obtaining a solution for a known function, $g(x)$, which can be treated by any of classical iterative numerical methods, represented uniquely by [21].

$$\begin{aligned} &\hat{x}_{k+1} = \hat{x}_k + \gamma_k g_k;\ g_k = g(\hat{x}_k) \\ &sgn(\gamma_k) = -\mathrm{sgn}(g'_k);\ g'_k = g'(\hat{x}_k) \end{aligned} \tag{15a}$$

where sgn$(\cdot)$ is the sign function, $g'(\cdot)$ is the first derivative of $g(\cdot)$ and γ_k is gain factor, which controls the convergence rate, [5]. Different numerical algorithms follow from the choice of the gain factor, [21]. In a stochastic environment g_k cannot be evaluated adequately, and has to be replaced by a noisy measurement, m_k, yeilding the stochastic version of (15a)

$$\hat{x}_{k+1} = \hat{x}_k + \gamma_k m_k;\ m_k = g_k + e_k \tag{15b}$$

Because m_k has a stochastic error or measurement noise, e_k, the relation Eq. (15b) is named stochastic approximation. This relation Eq. (15b) generates the sequence of estimates that in some probabilistic sense converge to x_0 if the gain, γ_k, is in the proper direction in Eq. (15a), [5]. The application of Eq. (15b) to the multivariable problem Eqs. (11) to (14) requires to define the equivalent relations to the corresponding scalar ones, resulting in

$$m_k = -\Psi(\varepsilon_k) h_k;\ \varepsilon_k = y_k - h_k^T \hat{x}_k \tag{16}$$

The Eq. (16) is a direct consequence of Eq. (14), when the indeterminate mean-value, $E\{\Psi(\cdot)\}$ is estimated by the current sample. Here, Ψ is derivative of, ρ in Eq. (13), yielding

$$\Psi(x) = \min(|x|, \Delta)\,\mathrm{sgn}(x);\ \Delta = 1.5 \tag{17}$$

Such type of statistical data processing is known as winsorization, [9-11]. Thus, the multivariable stochastic approximation based on M-robust approach is defined by Eqs. (15b), (16) and (17), where the scalar gain factor, γ_k, is a free quantity to be adopted. Although stochastic approximation methods, as well as the M-robust estimators are primarily developed for estimating constant parameters, analogous methods can be applied for estimating variables that are time-varying. This results in the M-robust dynamic stochastic approximation algorithm that represents a basis for robustifying the standard Kalman filter.

4 M-STATISTICAL APPROACH TO ROBUSTIFYING KALMAN FILTER USING DYNAMIC STOCHASTIC APPROXIMATION

The problem of constant parameter vector estimation, based on the M-robustified dynamic stochastic approximation, can be easily extended to estimating the states, x_k, of dynamic system, described by Eqs. (1), (2). Such an algorithm is a time-varying version of the parameter estimator defined by Eqs. (15b), (16) and (17), and is represented by

$$\hat{x}_k = \overline{x}_k - \Gamma_k \nabla_{\overline{x}} J_k\left(\overline{x}_k\right) \tag{18}$$

where Γ_k is the matrix gain factor, corresponding to the scalar one, γ_k, in Eq. (15b), and $\overline{x}_k$ is used to account for the predicted change in x_k, at each stage, k. Since the variation in x_k is governed by the stochastic difference Eq. (1), where w_k is zero-mean white noise, one concludes, by analogy with the time update Eq. (4) in the Kalman filter, that the one step MSE-optimal prediction is given by

$$\overline{x}_k = F_{k-1}\hat{x}_{k-1} \tag{19}$$

Here the prediction error covariance matrix, M_k, is given by Eq. (5), that is

$$\begin{aligned} M_k &= E\left\{\tilde{x}_k(-)\tilde{x}_k^{\,T}(-)\right\};\ \tilde{x}_k(-) = x_k - \overline{x}_k \\ M_k &= F_{k-1}P_{k-1}F_{k-1}^T + G_{k-1}Q_{k-1}G_{k-1}^T \end{aligned} \tag{20}$$

where the estimation error covariance matrix is defined by

$$P_k = E\left\{\tilde{x}_k(+)\tilde{x}_k^{\,T}(+)\right\};\ \tilde{x}_k(+) = x_k - \hat{x}_k \tag{21}$$

Moreover, $\nabla_{\overline{x}} J_k(\cdot)$ denotes the gradient vector of the modified performance index Eq. (12), representing the conditional expectation of the nonlinearly transformed measurement residuals, ε_k, given the predicted state, $\overline{x}$, and the measurement sequence, $Y^k = \{y_1, y_2, ..., y_k\}$, i.e.

$$J_k(\overline{x}) = E\left\{\rho\left(\varepsilon_k(\overline{x})/d_k\right)\middle|\overline{x}, Y^k\right\};\ \varepsilon_k(\overline{x}) = y_k - H_k\overline{x} \tag{22}$$

with ρ being the robust loss function in Eq. (13). The M-robust criterion Eq. (22) is general in the sense of the class of algorithms to be handled. Here, the normalising term, d_k, provides for the scale-invariant estimates. The dynamic stochastic approximation algorithm can be obtained by minimizing the M-robust criterion (22) at each stage, where the standard Kalman filter time-update recursions Eqs. (19), (20) are used to predict the system states. By analogy with the relation Eq. (14), one obtains from Eq. (22)

$$g_k = \nabla_{\overline{x}} J_k(\overline{x}) = -\frac{1}{d_k} E\left\{\psi\left(\varepsilon_k/d_k\right)\middle|\overline{x}_k, Y^k\right\} H_k^T \tag{23}$$

where Ψ is defined by Eq. (17) and the measurement residual, or innovation, is given by Eq. (22), that is

$$\varepsilon_k = \varepsilon_k\left(\overline{x}_k\right) = y_k - H_k\overline{x}_k \tag{24}$$

The unknown conditional mathematical expectation, in Eqs. (22), (23) can be approximated by the single random realisation

$$m_k = -\frac{1}{d_k}\Psi\left(\varepsilon_k/d_k\right)H_k^{\,T} \tag{25}$$

By substituting Eq. (25) into Eq. (18), further follows

$$\hat{x}_k = \overline{x}_k + \frac{1}{d_k}\Psi\left(\varepsilon_k/d_k\right)\Gamma_k H_k^T \tag{26}$$

or in the more convenient linear measurement form,

$$\hat{x}_k = \overline{x}_k + K_k\varepsilon_k \tag{27}$$

Here the new gain matrix in Eq. (27) is defined by

$$K_k = d_k^{-2}\omega_k\Gamma_k H_k^T \tag{28}$$

with the robust normalising penalty factor

$$\omega_k = \begin{cases} \dfrac{\Psi\left(\varepsilon_k/d_k\right)}{\varepsilon_k/d_k} & \text{for } \varepsilon_k \neq 0 \text{ and } d_k \neq 0 \\ 1 & \text{for } \varepsilon_k = 0 \text{ and/or } d_k = 0 \end{cases} \tag{29}$$

The scale factor, d_k, is a suitable estimate of the standard deviation of residual, ε_k, in Eqs. (22) to (29). This estimate can be generated by using the standard Kalman filtering calculation of residual variance in Eq. (10), that is

$$d_k = S_k^{\,1/2} = \left(H_k M_k H_k^T + R_k\right)^{1/2} \tag{30}$$

The robust state estimator is defined by Eqs. (19), (20), (24), (27) to (30), with unspecified gain matrix, Γ_k, in Eqs. (18) and (28). Starting from the desired fast-tracking performances, Γ_k may be obtained by minimizing at each stage an additional criterion

$$J_1\left(\Gamma_k\right) = TraceP_k; \tag{31}$$

where $Trace$ is the matrix trace operation. The posed optimisation problem is non-linear and can be solved by using convenient approximations. Starting from Eqs. (1), (2), (24), (27) to (30), one obtains by approximately minimizing Eq. (31) the suboptimal solution $\Gamma_k = M_k$, yielding further (for more details, see the Appendix A)

$$P_k = (I - K_k H_k) M_k; \Gamma_k = M_k \tag{32}$$

Thus, the proposed M-robust Kalman filter obeys the recursive predictor-corrector structure of the standard Kalman filtering. However, the prediction stage is defined by the standard Kalman filtering time-update recursions, while the correction stage is given by the M-robustified dynamic stochastic approximation algorithm. The corresponding recursions are given by:

Time update: Eq. (19), Eq. (20)

Measurement update: Eq. (17), Eq. (24), Eqs. (27) to (30), Eq. (32)

The optimality of Kalman filter is contained in its structure and calculation of the gain matrix. There is an intuitive logic behind the equations for the filter gain matrix, robustified or not. It follows from the matrix inversion lemma an alternative form of the Eq. (7)

$$K(k) = P(k \mid k) H_k^T R_k^{-1} \tag{33}$$

where $P(k \mid k)$ is defined by Eq. (9), [3, 4]. It can be concluded from Eq. (33) that each element of the filter gain matrix, K, is essentially the ratio between the statistical measures of the uncertainty in the state estimates, as is determined by P, and the uncertainty in a measurement, that is defined by R. In other words, the gain matrix, K, is proportional to the uncertainty in the estimate and inversely proportional to the average power of the noise. Thus, if measurement noise is large (R is large) and state estimation errors are small (P is small), the innovation, ε, in Eq. (6) is due chiefly to the observation noise and only small changes in the state estimates should be made through a small value of K. In addition, a small measurement noise, due to small R, and a large uncertainty in the state estimates, due to large P, suggest that the innovation or the measurement residual, ε, contains considerable information about errors in the states. This, in turn, will be used for strong corrections to the state estimates, owing to large K. Similarly, the gain matrix Eq. (28) of the robustified Kalman filter can be approximately expressed as (see Appendix B)

$$K_k = \omega_k P_k H_k^T R_k^{-1} \tag{34}$$

where ω_k is defined by Eq. (29), and P_k is given by Eq. (32). It should be noted that the weighted term ω in Eq. (29) represents an approximation to the first derivative of the Ψ-function in Eq. (17). This means that $\omega(x) \approx 1$ for $|x| \le \Delta$ and $\omega(x) \approx 0$ for $|x| > \Delta$. In other words, for the most observations belonging to the linear part of the influence function, Ψ, corresponding to the pure normal noise, the robust weighting term, ω, is equal to one, and the Eq. (34) reduces to the optimal gain matrix in Eq. (33). On the other hand, in the saturation part of Ψ-function, corresponding to the presence of outliers, the robust term ω is close to zero, yielding a small value of K in Eq. (34). This, in turn, produces a small correction to the state estimate and, as a consequence, suppresses the influence of outliers in the measurement sequence.

A theoretical convergence analysis of the proposed robust estimate represents a rather complex technical task, since the system in question is given by a dynamic, multi-variable, time-varying model, while the robust state estimator has a nonlinear form. However, it is often possible to achieve an estimation error that is statistically bounded, [4, 19]. Moreover, the proposed robust filtering technique is obtained by using approximations and intuitive reasoning. For this reason, further practical verifications are needed.

5 EXPERIMENTAL RESULTS

The performances of the derived robust filter are analysed from both practical robustness and tracking performances, using the manoeuvre target tracking scenery. The experimental results are generated upon two-dimensional state-space representation, using linear position sensor, [22]

$$x_{k+1} = F x_k + G w_k;\ y_k = H_k x_k + v_k$$
$$F = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix};\ G = \begin{bmatrix} 0 \\ 1 \end{bmatrix};\ H = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{35}$$

Here x_k is the target state vector containing its position and velocity, y_k is the discrete measurement of position, w_k is the state noise, representing the target manoeuvre, and v_k is the additive measurement noise. In addition, the matrix F is obtained by applying the uniform sampling with the period $T = 4$ s. The white noise, $\{w_k\}$, is zero-mean Gaussian with the variance $Q = 0.01$. The model Eq. (35) is inadequate during target manoeuvre and the state noise $\{w_k\}$ is introduced to compensate the effects of mismodeling. Moreover, zero-mean white noise $\{v_k\}$ is generated using the Gaussian mixture pdf, representing

$$p(\cdot) = (1-\alpha) N(\cdot \mid 0,1) + \alpha N(\cdot \mid 0, \sigma_o^2); \tag{36}$$

where $0 \le \alpha < 1$, $\sigma_o^2 > 1$, and $N(\cdot|0, \sigma)$ represents the zero-mean normal pdf with the variance σ^2.

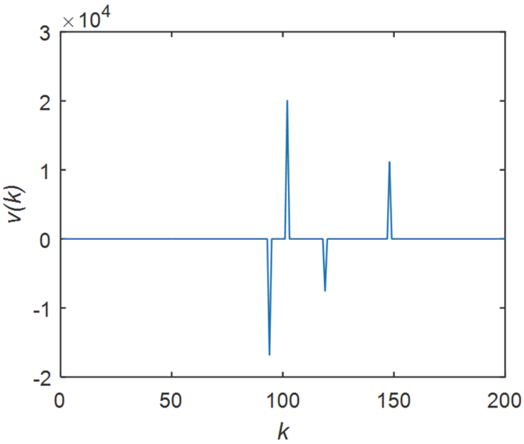

Figure 1 The heavy-tailed noise record

The heavy tailed nature of the noise pdf Eq. (36) is due to the target glint in mono-pulse radars, [22]. This, in turn, results in outliers contaminating assumed Gaussian observations. The random variable with the pdf (36) can be simulated by firstly generating a sample, u, from the uniform pdf on the interval (0, 1). If $u > \alpha$ then v is generated by independent sample from the pdf $N\left(\cdot \mid 0,1\right)$. Otherwise, it is a sample from the pdf $N\left(\cdot \mid 0,\sigma_o^2\right)$. The heavy-tailed noise realization is presented in Fig. 1.

The performances of the standard Kalman filter, designated as A1; are compared with the proposed robustified state estimator, designated as A2.

Experimental results are presented by two norms of estimation error, defined by

$$n_1\left(k\right) = \frac{\left\|\hat{x}_k - x_k\right\|}{\left\|x_k\right\|};\ n_2\left(k\right) = \frac{1}{k}\sum_{i=1}^{k} n_1\left(i\right) \tag{37}$$

where $\left\|\cdot\right\|$ is the Euclidean norm. Fig. 2 presents the true target position generated by the well-known cinematic equations of motion, outside the filter model Eq. (35), [22]. Here the true initial state vector $x_0^T = \left[3, 0.3\right]$.

Filter initialization requires the first guess $\hat{x}_0$, together with the corresponding covariance matrix P_0, to be adopted in advance. A common practical approach is to use the first two position measurements for providing the initial guesses, [22].

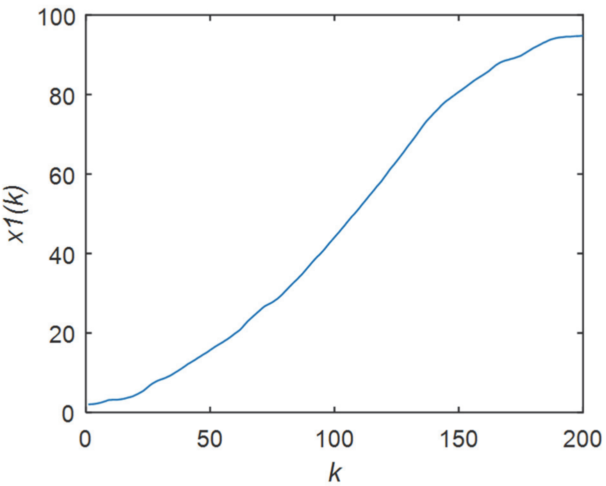

Figure 2 The true target position trajectory

$$\hat{x}_0 = \begin{bmatrix} y_2 \\ \dfrac{y_2 - y_1}{T} \end{bmatrix};\ P_0 = \begin{bmatrix} 1 & 1/T \\ 1/T & Q + 2/T^2 \end{bmatrix} \tag{38}$$

The n_2 norm (37), obtained for dissimilar noise sequences, Eq. (36) is presented in Figs. 3 and 4.

The obtained results have shown that the Kalman filter, A1, gives smaller total estimation error than the proposed robust filter, A2, upon the pure normal observations. Robust filter, A2, provides better estimation quality than the linear one, A1, upon the Gaussian mixture pdf in Eq. (36). Namely, Gaussian mixture family generates the outliers by the contaminating second term in Eq. (36). The influence of outliers results in the larger values of residual variance Eq. (10), and as a consequence reduces the Kalman filter gains, Eq. (7).

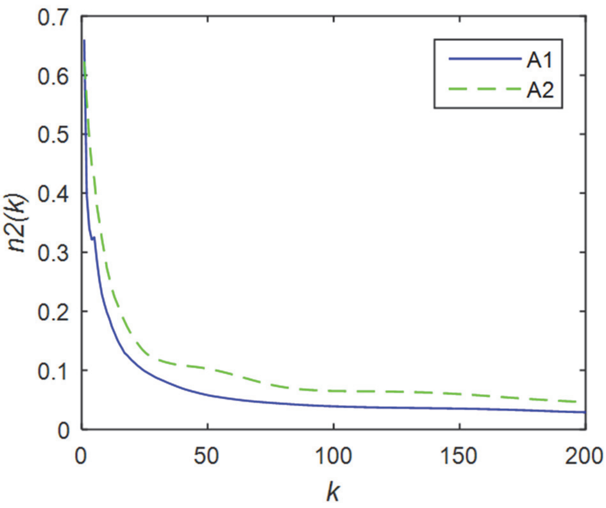

Figure 3 Comparison of different algorithms upon Gaussian noise (α = 0)

On the other hand, the robust filter, A2, gives smaller values of the residual variance Eq. (30), producing high enough values of the gain matrix, Eq. (28). This, in turn, gives a good tracking capability. At the same time, the obtained values of the gain matrix Eq. (28) are small enough to provide for immunity to outliers. Furthermore, the robust gain matrix in Eq. (28) differs from the optimal one in Eq. (7) by the robust weighted term, ω, in Eq. (29). As mentioned before, this term represents a suitable approximation of the first derivate of robust influence function, Ψ, in Eq. (17). Thus, it has the unity value for small and moderate arguments, corresponding to the pure Gaussian observations without outliers. As a consequence, for the most of observations the robustified Kalman filter, A2, behaves as the optimal one, A1, satisfying the efficiency robustness property. On the other hand, for huge absolute values of arguments, corresponding to bad data, or outliers, contaminating the measurements, the robust term ω is close to zero, reducing the value of robust gain matrix in Eq. (28). For this reason, the influence of outliers to the state estimate is significantly reduced, providing for the resistant robustness feature. Simulations have shown that the robust algorithm works quite well for $\alpha \leq 0.3$. Thus, for higher α the noise model Eq. (36) is no more adequate.

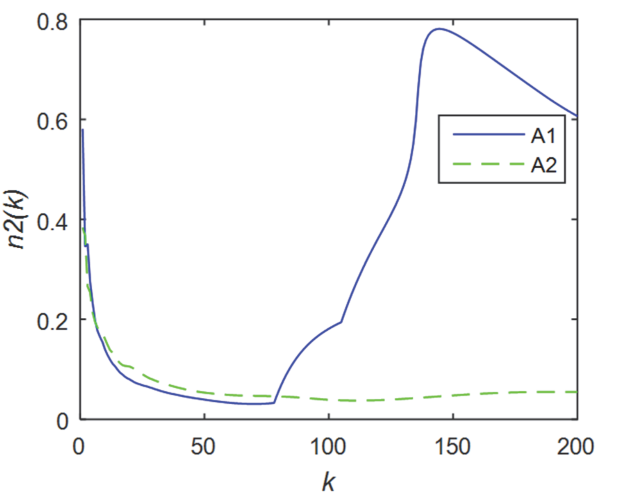

Figure 4 Comparison of different algorithms upon contaminated Gaussian noise (α = 0.1)

6 CONCLUSIONS

Kalman filter provides optimal estimates of system states in the case of a linear dynamic system, which is driven by the normal stochastic process and in presence of the Gaussian additive measurement noise. Presence of outliers in measurements usually yields poor performance of the Kalman filter, i.e. Kalman filter is not a robust estimator. The paper presents a new feasible algorithm of a robustified Kalman filter. The algorithm introduces elements of dynamic stochastic approximation algorithms equipped with an M-robust performance index, in the structure of the Kalman filter. Thus, the time update is performed according to the standard Kalman filtering algorithm, while the measurement update is performed as the M-robust stochastic approximation algorithm. Robustification of the Kalman filter is achieved by non-linear processing of measurement residuals. In order to improve convergence of the algorithm, additional optimisation procedure is introduced in computation of the estimator gain. The optimisation process yields the state estimation similar to a minimum variance procedure. The algorithm is feasible and yields recursive computation of the system state estimates. The paper provides detailed derivation of the algorithm. Further, an extensive analysis of the algorithm is performed. The analysis gives deep insight into the algorithm operation in absence and presence of the outliers. Each situation is accompanied with an appropriate estimator gain. Thus, presence of the outliers is reflected on the estimator gain, as a consequence of the robustification procedure applied to the standard Kalman filter. Performed analysis provides clear and unambiguous implementation of the algorithm. Additionally, emphasized M-robustified Kalman filtering technique offers an alternative to the standard Kalman filter in the situations characterized by erroneous noise statistics and mismodelling. Moreover, in many nonlinear tasks the robustified dynamic stochastic approximation recursive method is rather easyto implement.

Appendix A: Derivation of the M-robust Kalman filter

The prediction error is given by Eqs. (1), (19) and (20), that is

$$\tilde{x}_k(-) = x_k - \bar{x}_k = F_{k-1}\tilde{x}_{k-1}(+) + G_{k-1}w_{k-1} \tag{A1}$$

where, due to Eqs. (1), (2), (21), (24) and (27), the estimation error, $\tilde{x}_k(+)$, is

$$\tilde{x}_k(+) = x_k - \hat{x}_k = \tilde{x}_k(-) - K_k H_k \tilde{x}_k(-) - K_k v_k \tag{A2}$$

with $\{v_k\}$ being a sequence of zero-mean random variables in Eq. (3). If the initial guess, $\hat{x}_0(+) = E\{x(0)\}$, one concludes by the mathematical induction that, [3, 4]

$$E\{\tilde{x}_k(-)\} = 0,\ E\{\tilde{x}_k(+)\} = 0 \tag{A3}$$

In addition, one obtains form Eqs. (21), (A1) and (A2) that the prediction error covariance matrix, M_k, is given by Eq. (20). This expression is derived under the assumption Eq. (3), yielding

$$E\{\tilde{x}_{k-1}(+) w_{k-1}^T\} = 0,\ E\{\tilde{x}_k(-) v_k^T\} = 0 \tag{A4}$$

Moreover, it follows from Eqs. (3), (A2) and (A4) that the estimation error covariance matrix Eq. (21) is given by

$$\begin{aligned} P_k &= E\{\tilde{x}_k(+)\tilde{x}_k^T(+)\} \\ &= M_k - K_k H_k M_k - M_k H_k^T K_k^T + R_k K_k K_k^T \end{aligned} \tag{A5}$$

By substituting Eqs. (28) and (30) into (A5), the relation Eq. (31) reduces to

$$\begin{aligned} J_1(\Gamma_k) &= Trace\, M_k - 2d_k^{-2}\omega_k Trace\,\Gamma_k H_k^T H_k M_k \\ &\quad + d_k^{-4}\omega_k^2 R_k Trace\,\Gamma_k H_k^T H_k \Gamma_k \end{aligned} \tag{A6}$$

from which it follows

$$\begin{aligned} \frac{\partial J_1(\Gamma_k)}{\partial \Gamma_k} &= -2d_k^{-2}\omega_k M_k H_k^T H_k \\ &\quad + 2d_k^{-4}\omega_k^2 R_k \Gamma_k H_k^T H_k = 0 \end{aligned} \tag{A7}$$

Due to Eq. (17), the term ω in Eq. (29) can be approximated by the integer variable with zero and unity values, so that $\omega^2 \approx \omega$. For this reason, the relation Eq. (A7) can be approximated as

$$\frac{\partial J_1(\Gamma_k)}{\partial \Gamma_k} \approx -2d_k^{-2}\omega_k\left(M_k - d_k^{-2}R_k\Gamma_k\right)H_k^T H_k = 0 \tag{A8}$$

from which it follows

$$\Gamma_k = d_k^2 R_k^{-1} M_k = \left(R_k^{-1} H_k M_k H_k^T + 1\right) M_k \approx M_k \tag{A9}$$

since in the saturation part of the Ψ-function, corresponding to the existence of outliers in the measurement data, the variance, R_k, is large, while in the linear part of the Ψ-function, characterized by the pure Gaussian measurement noise, the robust estimator performs as the optimal one, yielding a small error, M_k. Using Eqs. (A9), (28) and (A5) further follows

$$\begin{aligned} P_k &= M_k - 2d_k^{-2}\omega_k M_k H_k^T H_k M_k \\ &\quad + R_k d_k^{-4}\omega_k^2 M_k H_k^T H_k M_k \\ &\approx M_k - d_k^{-2}\omega_k M_k H_k^T H_k M_k \end{aligned} \tag{A10}$$

which represents an alternative form of Eq. (32), so that the derivation is completed.

Appendix B: Derivation of the gain matrix in Eq. (34)

By substituting Eq. (30) in Eq. (28), one obtains

$$K_k = \omega_k M_k H_k^T \left(H_k M_k H_k^T + R_k \right)^{-1} \tag{B1}$$

Furthermore, by replacing Eq. (B1) into Eq. (32), and since ω_k in Eq. (29) can be approximated by zero-one variable, further follows

$$\omega_k P_k = \omega_k M_k - \frac{\left(\omega_k M_k H_k^T H_k \omega_k M_k \right)}{H_k \omega_k M_k H_k^T + \omega_k R_k} \tag{B2}$$

Taking into account the matrix inversion lemma [1-3], one obtains from Eq. (B2)

$$P_k^{-1} = M_k^{-1} + H_k^T R_k^{-1} H_k \tag{B3}$$

Using Eqs. (B3), (B1) can be rewritten as

$$\begin{aligned} K_k &= \frac{P_k \left(M_k^{-1} + H_k^T R_k^{-1} H_k \right) \omega_k M_k H_k^T}{H_k M_k H_k^T + R_k} \\ &= \frac{P_k H_k^T \left(R_k^{-1} R_k + R_k^{-1} H_k M_k H_k^T \right) \omega_k}{H_k M_k H_k^T + R_k} \end{aligned} \tag{B4}$$

which is identical to Eq. (34).

7 REFERENCES

[1] Candy, J. V. (2010). *Model-based signal processing*. Hoboken, N.J, IEEE Press.

[2] Verhaegen, M. & Verdult, V. (2012). *Filtering and system identification: a least squares approach*. Cambridge, Cambridge University Press.

[3] Kovačević B. & Đurović Ž. (2011). *Fundamentals of stochastic signals, systems and estimation theory: With worked examples*. Berlin, Springer.

[4] Grewal, M. S. & Andrews, A. P. (2015). *Kalman filtering theory and practice using Matlab*. John Wiley & Sons.

[5] Albert, A. E. & Gardner, L. A. (1967). *Stochastic approximation and nonlinear regression*. MIT Press (MA).

[6] Mendel, J. M. (2012). *Adaptive, learning and pattern recognition systems: theory and applications*. Academic Press.

[7] Heijden, F. van der., Lei, B., Xu, G., Ming, F., Zou, Y., de Ridder, D., & Tax, D. M. (2017). *Classification, parameter estimation, and state estimation: An engineering approach using MATLAB*. John Wiley & Sons.

[8] Banjac, Z., Đurović, Ž., & Kovačević, B. (2018). Approximate Kalman Filtering Using Robustified Dynamic Stochastic Approximation Method. *Telecommunications Forum (TELFOR*), 353-356. https://doi.org/10.1109/TELFOR.2018.8611817

[9] Barnett, V. & Lewis, T. (1994). *Outliers in statistical data*. Wiley-Blackwell.

[10] Huber, P. J. & Ronchetti, E. M. (2011). *Robust statistics*. John Wiley & Sons. https://doi.org/10.1007/978-3-642-04898-2_594

[11] Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing*. Academic Press. https://doi.org/10.1016/B978-0-12-804733-0.00001-9

[12] Gandhi, M. A. & Mili, L. (2010). Robust Kalman Filter Based on a Generalized Maximum-Likelihood-Type Estimator. *IEEE Transactions on Signal Processing*, *58*(5), 2509-2520. https://doi.org/10.1109/TSP.2009.2039731

[13] Li, K., Hu, B., Chang, L., & Li, Y. (2014). Robust square-root cubature Kalman filter based on Huber's M-estimation methodology. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *229*(7), 1236-1245. https://doi.org/10.1177/0954410014548698

[14] Vlahović, N. & Djurovic, Z. (2020). Robust tracking of moving objects using thermal camera and speeded up robust features descriptor. *International Journal of Adaptive Control and Signal Processing*, *35*(4), 549-566. https://doi.org/10.1002/acs.3212

[15] Zhao, Y., Qi, W., Liu, P., & Lin, J. (2020). M-estimation-Based robust Kalman filter algorithm for three-dimensional AOA target tracking. *2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICSP)*. https://doi.org/10.1109/icicsp50920.2020.9231976

[16] Kovačević B., Banjac, Z., & Milosavljević M. (2013). *Adaptive Digital Filters*. Berlin, Heidelberg: Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33561-7

[17] Kovačević, B., Đurović, Ž., & Glavaški, S. (1992). On robust Kalman filtering. *International Journal of Control*, *56*(3), 547-562. https://doi.org/10.1080/00207179208934328

[18] Đurovic, Ž. & Kovačević, B. (1999). Robust estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, *44*(6), 1292-1296. https://doi.org/10.1109/9.769393

[19] Kovačević, B., Milosavljević, M. M., Veinović, M., & Marković, M. (2018). *Robust Digital Processing of Speech Signals*. Cham, Springer International Publishing.

[20] Hanlon, P. & Maybeck, P. (2000). Characterization of Kalman filter residuals in the presence of mismodeling. *IEEE Transactions on Aerospace and Electronic Systems*, *36*(1), 114-131. https://doi.org/10.1109/7.826316

[21] Chapra, S. C. & Canale, R. P. (2015). *Numerical methods for engineers*. New York, NY, McGraw-Hill Education.

[22] Blackman, S. S. & Popoli, R. (1999). *Design and analysis of modern tracking systems*. Norwood, MA, Artech House.

Contact information:

Zoran BANJAC, PhD, System Architect
(Corresponding author)
Institute Vlatacom,
5 Milutina Milankovica Blvd. 11070 New Belgrade, Serbia
E-mail: zoran.banjac@vlatacom.com

Branko KOVAČEVIĆ, PhD, Full Professor
School of Electrical Engineering, University of Belgrade,
Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia
E-mail: kovacevic_b@etf.bg.ac.rs