

RESEARCH

Open Access



Approximate Kalman filtering by both M-robustified dynamic stochastic approximation and statistical linearization methods

Miloš Pavlović^{1,2}, Zoran Banjac^{2*}  and Branko Kovačević¹

*Correspondence:
zoran.banjac@vlatacom.com

¹ School of Electrical Engineering,
University of Belgrade, Belgrade,
Serbia

² Vlatacom Institute of High
Technology, Belgrade, Serbia

Abstract

The problem of designing a robustified Kalman filtering technique, insensitive to spiky observations, or outliers, contaminating the Gaussian observations has been presented in the paper. Firstly, a class of M-robustified dynamic stochastic approximation algorithms is derived by minimizing at each stage a specific time-varying M-robust performance index, that is, general for a family of algorithms to be considered. The gain matrix of a particular algorithm is calculated at each stage by minimizing an additional criterion of the approximate minimum variance type, with the aid of the statistical linearization method. By combining the proposed M-robust estimator with the one-stage optimal prediction, in the minimum mean-square error sense, a new statistically linearized M-robustified Kalman filtering technique has been derived. Two simple practical versions of the proposed M-robustified state estimator are derived by approximating the mean-square optimal statistical linearization coefficient with the fixed and the time-varying factors. The feasibility of the approaches has been analysed by the simulations, using a manoeuvring target radar tracking example, and the real data, related to an object video tracking using short-wave infrared camera.

Keywords: Impulsive noise, Kalman filtering, Non-Gaussian noise, Nonlinear filtering, Outliers, Robust estimation, Statistical linearization, Stochastic approximation, Tracking

1 Introduction

One of the most important contributions to the estimation theory is the optimal linear Kalman filter. The simplicity of the optimal Kalman filter is contained in its linear predictor–corrector structure, making this result attractive from a practical point of view [1–7]. The Kalman filter produces optimum on the average by minimizing the expectation of a scalar-valued penalty, score or loss function, having the random estimation error as the argument. Such a criterion function that is symmetric, convex and equal to zero for the zero-valued argument is known as the admissible one [3, 5]. Kalman filter is the optimal state estimator within a class of the admissible score functions, and as a consequence, it also represents an optimal estimator in the minimum variance sense

[1–7]. To obtain the optimal performance by the Kalman filter, it is necessary to provide a correct a priori description of the system state dynamics and the statistics of random observations. In this sense, if a system state dynamics and the associated observations are confined to severe nonlinearities that cannot be described properly by linearization, and/or if the underlying stochastic sequences are not Gaussian, the Kalman filter may degrade its performance [8, 9]. In general, under nonlinear state dynamics and/or non-Gaussian observations, the design of an optimal state estimator can be rather cumbersome [1–5]. Therefore, there exists an interest in a class of estimation procedures that is not optimal, concerning some statistical performance measure, but produces the bounded total estimation error. A family of dynamic stochastic approximation procedures offers a reasonable choice, since it produces fairly well results in many applications, including parameter and state estimation, optimization, pattern classification and signal processing [10–14]. In this sense, any Kalman filter with an erroneous gain sequence, owing to departure from the theoretically optimal conditions in practice, may be considered as a dynamic stochastic approximation algorithm. On the other hand, it is commonly assumed that real measurements are approximately Gaussian distributed, due to the central limit theorem of the statistics [5]. Moreover, statistical analysis of the numerous industrial and scientific observations has shown that these contain, as a rule, five to ten percentages of outliers [15]. Therefore, in many practical situations, the real probability distribution function (pdf) of the random observations is similar in the middle to the assumed Gaussian one, but differs from it by the heavier tails, generating the spiky observations, or the outliers, contaminating the mainly Gaussian distributed observations [15–20]. Particularly, the optimal Kalman filter is sensitive to outliers, due to its linear dependence upon observations, or it is non-robust. Therefore, there is also an additional practical interest in designing a class of robust filtering techniques that can cope with outliers.

A simple concept of the robustness is the so-called censored data, where the measurement data that differ sufficiently from the predicted values are discarded [15]. Such type of the robust procedure suffers from several faults, and the principle one is that it is often hard to distinguish an outlier from a large, but not unnatural, deviation. Therefore, to handle the outliers more efficiently, several robust procedures are proposed in the statistical literature [15–19]. Particularly, the Huber's M-robust estimator is frequently applied, because it approximates the optimal maximum likelihood (ML) estimator [17]. Thus, it can be understood and implemented easily by the practical workers. In this sense, many various combinations of the M-robust estimator and the optimal Kalman filter, or the linear least-square estimator, have been proposed in the literature [20–37]. In general, any estimation procedure is a combination of the criterion to be minimized, the model of the variables to be estimated and an estimation algorithm [7]. In this sense, the proposed robust estimators in the selected literature may be classified into the two groups. The first one is a family of the non-recursive or offline robust schemes, where the Kalman filtering problem is recast as a linear regression problem, which is solved by the M-robust estimator [20–27]. The posed optimization problem is nonlinear, and an iterative numerical method is required to solve it. Thus, the standard or the simplified Newton's method, as well as the iteratively reweighted least-square method, are recommended [22, 24, 27, 38]. A such derived robust estimator is in a batch-mode regression

form, processing the observations and the predictions simultaneously, that makes it very effective in suppressing the outliers. However, the robustness in these estimators is achieved at the cost of the increasing computational requirements. In general, a non-recursive, or offline, estimator may also be used in a real-time application by introducing a one-step rectangular sliding window of a proper length [20]. The basic problems in choosing the window length are related to time-varying parameter changes, together with the influence of outliers contaminating the observations. In general, a smaller parameter estimates variance is obtained at a longer window length, as a consequence of a larger averaging of the measurement data. However, this is in collision with the requirement to follow possible time-varying changes in the parameters to be estimated. Moreover, a short window length may result in unreliable parameter estimates, because of a high order of the underlying parameter regression model. Furthermore, a bias, of shift, in the parameter estimates is unavoidable since the sliding window permanently encompasses the observations contaminated by outliers. In this sense, the M -robust procedures are efficient in suppressing the influence of outliers, thus reducing significantly the bias and the variance of the robust estimates. Finally, as mentioned above, a non-recursive robust estimator is rather computationally complex, and an increase in the computational complexity basically depends on the number of necessary iterations to solve the parameter regression problem. Therefore, to solve the posed problems, it is more natural to use a recursive robust procedure than the non-recursive one. In this sense, starting from the computational considerations, the second group represents a family of the estimators that calculate an estimate recursively, because of the practical requirements to online or real-time signal processing. Such derived robust recursive estimator represents an acceptable balance between the computation efforts and the practical robustness performance [20, 31–35]. A new member of this family has been proposed in this article. The mentioned recursive robust estimators differ from the newly proposed one by the level of models that are used for the model-based signal processing. In this sense, the above estimators are based on the black-box models that have a parametric or polynomial form (FIR, AR, ARMA, etc.) [5–7]. Moreover, black-box models are basically used as a data prediction mechanism, and the estimated parameters may be used to extract limited physical information. However, the newly proposed recursive robust estimator is derived from the true model-based technique, using lumped physical model structure characterized by a state-space representation. Such a true model-based approach is incorporating the mathematical models of both the physical phenomenology, or system state dynamics, and measurement process, including noise, into the estimation process to extract the desired information [1–7]. This, in turn, produces better estimator performance than the black-box model-based estimation techniques. In general, the computational requirements depend on the order of the underlying state-space model, and for a not too large number of dynamic system states there are no significant additional requirements to the computational resources. Moreover, a recursive weighted least squares-type estimator, representing a combination of the Huber's M -robust estimator with a specific linear form of the dynamic stochastic approximation procedure, has been proposed recently to redesign the measurement-update recursion in the optimal Kalman filter [36, 37]. Here, the resulting state update recursion is still linear in the observations, but insensitivity to outliers is achieved by using a nonlinear weighting

factor in the Kalman gain calculation. Such a quasi-linear robust state estimator produces worse estimation performance than the proposed true nonlinear robust estimator in this article. The last one treats the outlier more severely by both the nonlinear residual processing and the Kalman gain calculation using the nonlinear weighting factor. In addition, many suboptimal nonlinear state estimators have been designed by applying the Taylor series expansion to describe a nonlinear system state dynamic [1–7]. Another frequently used method is the statistical approximation, generally producing a better nonlinearity approximation than the Taylor series method [1–3]. The simplest form of such method is known as the statistical linearization. Here, the linear approximation of the nonlinearity is used and, analogously to the estimation problem, the mean-square error (MSE) criterion is minimized to calculate the underlying coefficients. This, in turn, assumes that the pdf of the nonlinearity random argument is known in advance, and the Gaussian one is adopted frequently. Moreover, the statistical linear approximation can often be made for an adopted pdf, in such a manner that the calculated coefficients provide for a more accurate result, in the statistical sense, than the truncated Taylor series of a high order. Therefore, the statistical linearization method has a potential advantage for designing a suboptimal nonlinear filter [1–3]. Also, the Huber's M -robust approach has been proposed to making a suboptimal nonlinear filter more robust [28–30].

In this article, a new combination of the Huber's M -robust estimator and the nonlinear dynamic stochastic approximation algorithm of the approximate minimum variance type has been proposed. In this sense, the Huber's M -robust concept is utilized to design a family of the M -robustified dynamic stochastic approximation procedures, by minimizing at each stage the general time-varying M -robust performance index, based on the Huber's M -robust score function. To produce fast convergence, the gain matrix of a particular algorithm is derived by step-by-step minimization of an approximate minimum variance-type criterion. The posed nonlinear optimization problem is solved approximately, by using the statistical linearization method. Furthermore, by approximating, at each stage, the mean-square optimal statistical linearization coefficient by the average slope of the Huber's M -robust influence function, representing the first derivative of the underlying score function, a new feasible statistically linearized M -robustified dynamic stochastic approximation procedure is derived. Moreover, by approximating the average slope of the Huber's influence function with the current sample, an adaptive version of the proposed robust recursive state estimator has been obtained. Starting from the optimal Kalman filter structure, in which the prediction and the correction terms are independent, regarding the state estimate given the predicted one and vice versa, the derived robust recursive state estimator is used to redesign the correction phase, making the Kalman filter more robust. The practical robustness of the designed versions of the statistically linearized M -robustified Kalman filter has been analysed by both the simulations, using an example of single target radar tracking under an impulsive noise environment, and the real data, concerning an object tracking in a video sequence, generated by the short-wave infrared camera.

The paper organization is the next. A brief description of the Kalman filtering technique, and some discussion on the robustness issues, are presented in Sect. 2. Section 3 is devoted to the synthesis of a new statistically linearized M -robustified Kalman filtering technique, by both the M -robustified dynamic stochastic approximation algorithm

of the approximately minimum variance type and the statistical linearization method. Moreover, both the fixed and the time-varying suitable approximations of the mean-square optimal statistical linearization coefficients are considered in Sect. 3. Experimental results obtained by both the simulations, using a manoeuvring target radar tracking scenario, and the real data, related to an object video tracking using the short-wave infrared camera are presented in Sect. 4. The concluding remarks are given in Sect. 5. The complete derivation of the proposed statistically linearized M-robustified Kalman filtering technique is given in Appendix 1, while the derivation of the optimal statistical linearization coefficients is presented in Appendix 2.

2 Problem formulation

Let us consider a linear dynamic stochastic system which is represented by the first-order linear difference state vector equation

$$x_{k+1} = F_k x_k + G_k w_k \tag{1}$$

and the linear algebraic measurement vector equation

$$y_{k+1} = H_{k+1} x_{k+1} + v_{k+1} \tag{2}$$

where x_k is the state vector, y_k is the observation vector, w_k is the zero-mean state noise or disturbance vector with covariance matrix Q_k , and v_k is the zero-mean observation noise vector with the covariance matrix R_k , at the discrete time index, k . Moreover, the time-varying matrices F , G and H are also known in advance for each discrete time index, k .

Here, the initial random state vector, x_0 , is the Gaussian one with known both the mean value, m_0 , and the corresponding covariance matrix, P_0 . Also, it is assumed that the zero-mean white Gaussian noise sequences, $\{w_k\}$ and $\{v_k\}$, are mutually uncorrelated, and uncorrelated with the initial state, x_0 , for all discrete time indices, k .

Let $\hat{x}_{k|l} = E\{x_k | Y^l\}$, ($l = k - 1, k$) denote the optimal linear least-square estimates of the state, x_k , given the observations $Y^l = \{y_j, j \leq l\}$, where $E\{\cdot | \cdot\}$ is underlying conditional expectation, and let $P_{k|l} = E\{\tilde{x}_{k|l} \tilde{x}_{k|l}^T\}$ denote the corresponding covariance matrix of the estimation error, $\tilde{x}_{k|l} = x_k - \hat{x}_{k|l}$. Then, the standard Kalman filter recursions are given by [1–7].

1) Time update (prediction phase):

$$\hat{x}_{k+1|k} = E\{x_{k+1} | Y^k\} = F_k \hat{x}_{k|k} \tag{3}$$

$$P_{k+1|k} = E\{\tilde{x}_{k+1|k} \tilde{x}_{k+1|k}^T\} = F_k P_{k|k} F_k^T + G_k Q_k G_k^T \tag{4}$$

2) Measurement update (correction, estimation or filtering phase):

$$\hat{x}_{k+1|k+1} = E\{x_{k+1} | Y^{k+1}\} = \hat{x}_{k+1|k} + K_{k+1} \varepsilon_{k+1}; \varepsilon_{k+1} = y_{k+1} - H_{k+1} \hat{x}_{k+1|k} \tag{5}$$

$$\begin{aligned} K_{k+1} &= P_{k+1|k} H_{k+1}^T S_{k+1}^{-1} \\ P_{k+1|k+1} &= [I - K_{k+1} H_{k+1}] P_{k+1|k} \end{aligned} \quad (6)$$

$$S_{k+1} = E \left\{ \varepsilon_{k+1} \varepsilon_{k+1}^T \right\} = H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1} \quad (7)$$

The Kalman filter is initialized with $\hat{x}_{0|0} = m_0$, $P_{0|0} = P_0$. The Kalman filter optimality is contained in its suitable predictor–corrector form, and the associated calculation of the gain matrix, K , [1, 5]. However, as mentioned before, the Kalman filter is non-robust, in the sense of its sensitivity to spiky observations, bad data or outliers. In the statistical literature, there exist at least four definitions of robustness, [15–20]. The two of them, named the qualitative and the min–max robustness, respectively, are based on the strong mathematical treatments, [17, 19]. The other two, the so-called resistant and the efficiency robustness, are primarily oriented towards data, and are based on the empirical reasoning, [15, 16, 18]. Roughly speaking, the resistant robustness means that an estimator eliminates successfully the influence of outliers, while the efficiency robustness denotes that an estimator provides for an acceptable estimation quality under both the pure Gaussian observation, and the Gaussian one contaminated by outliers. Both robustness features designate the practical robustness, and are emphasized by the practitioners. Also, although there exist several robust estimation procedures in the statistical literature, the Huber’s M-robust approach is preferable, since it originates from the optimal maximum likelihood (ML) concept, making it more natural and easier to implement, [17]. In this sense, an estimator must not be exactly the optimal ML estimator, but has to approximate the optimal one in such a manner to achieve the practical robustness goals. It should be noted that the min–max robust estimation is exactly the optimal ML estimation based on the loss or score function, $\rho(\cdot) = -\ln p_0(\cdot)$, named the likelihood function, with $p_0(\cdot)$ being the worst-case pdf within the given pdf’s class. The worst-case pdf contains the minimal information about a variable to be estimated and minimizes the Cramer–Rao lower bound. This represents a non-classical variational problem that can be solved exactly only for the static models, when the posed problem reduces to minimizing the Fisher information, [17, 20]. In addition, the qualitative robustness is based on the Hampel’s definition of the influence function, as a suitable measure of the robustness capacity [19]. In this sense, the influence function represents the first derivative, or the slope, of the robust score function, ρ , used to define the Huber’s M-robust performance index. [17, 19].

As mentioned before, any Kalman filter whose gain differs from the optimal one, owing to the errors in the presumed noise statistics, or due to an inadequate representation of the system state dynamics, can be viewed as a dynamic stochastic approximation algorithm, [1, 10–13]. Therefore, this algorithm may represent a suitable substitution to an optimal estimation technique, when the assumptions on which the latter one is based are not fulfilled in practice. Starting from the practical limitations of the linear optimal Kalman filter, this approach can be applied further to making the optimal Kalman filter more robust.

3 Statistically linearized M-Robustified Kalman filtering

As mentioned above, the Huber’s M-robust approach combined with the dynamic stochastic approximation method may be used to computing a robust recursive state estimates of the dynamic stochastic system represented by (1), assuming the scalar observations, in (2). Also, a case of the multidimensional measurements, in (2), may be considered in the same manner by processing the individual observations one at a time. This approach assumes that the components of the measurement vector, in (2), can be processed sequentially, as the uncorrelated scalar observations. In this sense, one has to redefine the measurement vector, in (2), to making the corresponding measurement errors, or the noise vector components, to be mutually uncorrelated. This, in turn, results in a diagonal form of the measurement uncertainty covariance matrix, R_k . A suitable stable numerical decomposition method that is frequently used in practice is the Cholesky factorization, or its modification named the UD-decomposition, [2, 3].

The Huber’s M-robust estimator minimizes an empirical average loss, being defined by the nonlinear score function, ρ , to estimate the constant parameters in a linear regression problem, [17]. To apply this robust approach to a dynamic system recursive state estimation, the time-varying M-robust performance index is introduced, instead of the M-robust performance measure in the form of the empirical average loss, that is

$$J_k(\bar{x}) = E \left\{ \rho \left(\frac{\varepsilon_k(\bar{x})}{s_k} \right) \middle| \bar{x}, Y^k \right\} \tag{8}$$

with $E \{ \cdot | \bar{x}, Y^k \}$ being the conditional expectation under the known predicted state, \bar{x} , at the present stage, k , as well as the known observations up to the current stage, Y^k , where y is a scalar system output in (2), [12, 20]. Starting from (8), one can define a family of the dynamic stochastic approximation recursive estimators, minimizing the M-robust performance index (8) at each stage, k

$$\hat{x}_k = \bar{x}_k - \Gamma_k g_k(\bar{x}_k); \quad g_k(\bar{x}_k) = \nabla_{\bar{x}} J_k(\bar{x}_k) \tag{9}$$

where Γ_k is the matrix gain, and \bar{x}_k is a one-step prediction, to take into account for the changes in the current state, x_k , in (1). The term, $\nabla_{\bar{x}} J_k(\cdot)$ in (9), designates the gradient vector of the scalar-valued deterministic M-robust criterion in (8). Taking into account (8), one obtains

$$\nabla_{\bar{x}} J_k(\bar{x}_k) = \frac{\partial J_k(\bar{x}_k)}{\partial \bar{x}} = -\frac{1}{s_k} E \left\{ \psi \left(\frac{\varepsilon_k}{s_k} \right) \middle| \bar{x}_k, Y^k \right\} H_k^T \tag{10}$$

with $\psi(\cdot)$ being the first derivative, named the influence function, of the robust score function, $\rho(\cdot)$, in (8). Moreover, the term, $\partial(\cdot)/\partial x = \{ \partial(\cdot)/\partial x_1 \cdots \partial(\cdot)/\partial x_n \}^T$, denotes the partial derivative operator, where x is the $n \times 1$ column vector.

Analogously to (5), the measurement prediction residual, or the innovation, is defined by

$$\varepsilon_k = \varepsilon(\bar{x}_k) = y_k - H_k \bar{x}_k \tag{11}$$

where y_k is the scalar system output, and H_k is the observation vector in (2), at the stage, k . In addition, s_k is the normalizing (scaling) factor that provides for the scale-invariant state estimates, and represents an estimate of the corresponding standard deviation.

In general, the conditional expectation in (10) is indeterminable and, analogously to the dynamic stochastic approximation approach, can be approximated by the current sample, [1, 3]. Thus, the unknown expectation, in (10), can be estimated at each stage, k , by the current realization of the underlying random argument. This, in turn, results in the stochastic gradient vector representation

$$m_k = -\frac{1}{s_k} \psi\left(\frac{\varepsilon_k}{s_k}\right) H_k^T \tag{12}$$

Furthermore, by replacing (12) with (9), a family of the M-robustified dynamic stochastic approximation recursive state estimators takes the form

$$\hat{x}_k = \bar{x}_k + \frac{1}{s_k} \Gamma_k H_k^T \psi\left(\frac{\varepsilon_k}{s_k}\right) \tag{13}$$

In words, the posed optimization problem (8) reduces to finding the solution of the equation, $g_k(\cdot) = 0$, at each stage, k , with g_k , in (9), being the so-called regression function. Since this function is unknown, it is replaced by the random sample realization, (12), and the resulting estimation scheme (13) is known as the dynamic stochastic approximation algorithm, [1, 10–12].

The role of an admissible score function, ρ , in (8) is to provide for the practical robustness of the estimation procedure (13). To achieve such performance, the M-robust influence function, $\psi = \rho'$, has to be a bounded and continuous function [15–19]. This, in turn, produces that both the single and the grouped outliers will not have a significant impact on the state estimates (13), satisfying the resistant robustness requirement. Additionally, to obey the efficiency robustness feature, the estimation procedure (13) has to perform fairly well under both the pure Gaussian observations and the Gaussian one contaminated by outliers. The Huber’s M-robust score function, ρ_H , that is quadratic in the middle, but increases slowly than the quadratic one in the tails, obeys both the practical robustness requirements [17]. The corresponding M-robust influence function is the monotonously non-decreasing saturation-type nonlinearity, given by

$$\psi_H(z) = \rho'_H(z) = \min(|z|, \Delta) \operatorname{sgn}(z); \quad \Delta = 1.5 \tag{14}$$

where Δ is the tuning constant that controls the efficiency robustness. The choice $\Delta = 1.5$ often produces an acceptable result, and such procedure is known as the Huber’s 1.5 M-robust approach [17]. Nonlinear data processing using the saturation function (14) is known in the statistical literature as the winsorization [15–19]. As mentioned before, statistical analysis has shown that the various measurement data contain, as a rule, 5 to 10 percentage of outliers [15]. In this sense, it is frequently assumed that the observations are generated by the Gaussian mixture pdf

$$p_\delta(\cdot) = (1 - \delta)N(\cdot|0, \sigma_n^2) + \delta N(\cdot|0, \sigma_o^2); \quad 0 \leq \delta \leq 1, \sigma_n^2 = 1, \sigma_o^2 \gg \sigma_n^2 \tag{15}$$

where δ is the contamination degree, and σ_n^2 is the unit variance of the mostly observations generated by the standard zero-mean Gaussian pdf with the unit variance,

$N(\cdot|0, 1)$, while σ_o^2 is a large variance of outliers, generated by the zero-mean normal pdf, $N(\cdot|0, \sigma_o^2)$. Such pdf is also known as the δ -contaminated normal one [15–20].

Particularly, for the pdf class (15), with an arbitrarily zero-mean symmetric contaminating pdf, instead of the Gaussian one, $N(\cdot|0, \sigma_o^2)$, the worst-case pdf, p_0 , in the sense of the minimal Fisher information, is the Gaussian one in the middle and the Laplace, or double exponential, one in the tails. The influence function, $\psi = \rho'$, of the associated likelihood function, $\rho(\cdot) = -\ln p_0(\cdot)$, is the saturation-type nonlinearity, in (14), [17]. Examples of the pdf’s classes commonly used in engineering problems, and the derivation of the worst-case pdf within the prespecified class, are presented in the literature, [17, 20].

The role of the matrix gain, Γ_k , in (13) is to control the convergence speed. At this moment, the gain, Γ_k , is not connected to any assumption about the random state to be estimated, and the corresponding noise sequences. Therefore, to link the optimal Kalman filter with the recursive robust state estimator (13), an additional optimization criterion of the approximate minimum variance type is introduced

$$J_1(\Gamma_k) = Trace P_k; P_k = E\{\tilde{x}_k(+)\tilde{x}_k^T(+)\}; \tilde{x}_k(+) = x_k - \hat{x}_k \tag{16}$$

where the matrix P_k is the estimation error covariance at the stage k , with *Trace* being the matrix trace. Minimization of the scalar, deterministic criterion, in (16), at each stage, k , with respect to the gain matrix Γ_k , represents a complex nonlinear problem, and an approximate optimal solution can be obtained by using the statistical linearization technique, [1–3]. Starting from an odd ψ -function, in (14), where its random argument, z , is a sample from a zero-mean white scaled measurement residual sequence, $\{\varepsilon_k/s_k\}$ in (13), with a symmetric pdf belonging to the class (15), the application of the statistical linearization method results in the following approximation of the influence function

$$\psi\left(\frac{\varepsilon_k}{s_k}\right) \approx \alpha \frac{\varepsilon_k}{s_k}; \alpha = \frac{1}{\sigma_z^2} E\{\psi(z)z\} \tag{17}$$

with α being the mean-square optimal statistical linearization coefficient, while σ_z^2 is the variance of the random argument z , (for more details, see Appendix 2).

Particularly, if $\psi(\cdot)$ in (17) is the saturation function (14), the coefficient, α , is dependent on the linear segment of $\psi(\cdot)$, the saturation threshold, Δ , and the variance, σ_z^2 . In general, for the small σ_z -values, in comparison with the Δ -values, the probability of saturation occurrence is low, resulting in the α -values close to one. Moreover, for the higher σ_z -values, the α -values are significantly smaller than one, due to a larger probability of the saturation existence. Therefore, for the prespecified ψ -function, in (14), and the different σ_z -values, a set of the α -coefficients from the interval [0, 1] is obtained. Furthermore, since the normalized residual, in (17), has the unit standard deviation that is smaller than the threshold, Δ , the corresponding coefficient, α , is close to one. By substituting (17) in (13), one obtains the following relation for the statistically linearized M-robustified Kalman state estimates, instead of the recursions (5)–(7),

$$\hat{x}_k = \bar{x}_k + K_k \varepsilon_k; K_k = s_k^{-2} \alpha \Gamma_k H_k^T \tag{18}$$

Here, the standard deviation, s_k , of the measurement residual, ε_k , in (18) may be defined by the associated variance in (7), yielding

$$s_k = S_k^{1/2} = \left(H_k M_k H_k^T + R_k \right)^{1/2} \tag{19}$$

with the matrix, M_k , being the prediction error covariance defined by (4), that is

$$M_k = E \left\{ \tilde{x}_k(-) \tilde{x}_k^T(-) \right\}; \tilde{x}_k(-) = x_k - \bar{x}_k \tag{20}$$

Particularly, for the measurement noise model, in (15), the underline observation noise variance in (19) is given by

$$R_k = (1 - \delta) \sigma_n^2 + \delta \sigma_o^2 \tag{21}$$

In general, the contamination degree, δ , is not exactly known in practice and cannot be determined adequately from the measurement residuals, [15–20]. As mentioned before, a reasonable choice in practice is to adopt the δ -value in advance within the interval from 0.05 to 0.1, corresponding to 5 to 10 percentage of outliers in the Gaussian distributed measurement data. Furthermore, the standard deviation of outliers, σ_o , is also unknown, but it is significantly greater than the unit nominal standard deviation, σ_n , of the mainly zero-mean Gaussian noise samples, in (15). Taking into account (1), (2), (11), (18)–(20), together with the convenient simplifications, one obtains an approximate optimal solution, by minimizing the adopted criterion, in (16), (for more details, see Appendix 1)

$$\Gamma_k = M_k; P_k = (I - K_k H_k) M_k \tag{22}$$

In addition, starting from (13)–(22), the statistically linearized M-robustified dynamic stochastic approximation recursive state estimator is defined by

$$\hat{x}_k = \bar{x}_k + K_k s_k \psi_H \left(\frac{\varepsilon_k}{s_k} \right); K_k = s_k^{-2} \alpha M_k H_k^T \tag{23}$$

Here, the residual, ε_k , scaling factor, s_k , Huber’s influence factor, ψ_H , and the coefficient, α , are defined by (11), (14), (17) and (19), respectively, while the estimation error covariance matrix, P_k , is given in (22). The recurrent relations (20), (22) and (23), are similar to the measurement-update recursions, (5)–(7), in the filtering stage of the linear optimal Kalman filter. Since the design of the prediction and the estimation processes in the optimal Kalman filter are independent, the last one may be robustified by combining the recursive robust estimation process in (22) and (23), instead of the measurement-update recursions (5)–(7), with the one-step mean-square optimal prediction in (3), (4), to derive a new M-robustified version of the optimal Kalman filter. In this sense, the time-update recurrent relations, (3) and (4), in the prediction stage of the linear optimal Kalman filter define also the recursive prediction process in the M-robustified statistically linearized Kalman filter. Thus, for the one-step prediction, \bar{x}_k in (23), and the corresponding prediction error covariance matrix, M_k in (20) and (23), the same recursions as in (3) and (4) are obtained, that is

$$\bar{x}_k = F_{k-1}\hat{x}_{k-1}; M_k = F_{k-1}P_{k-1}F_{k-1}^T + G_{k-1}Q_{k-1}G_{k-1}^T \tag{24}$$

Here, F_k and G_k are the system state transition matrix and the state noise matrix, in (1), with the state noise covariance matrix, Q_k .

Unfortunately, the measurement noise statistics are not exactly known in many applications, and in such circumstances the mean-square optimal linearization coefficient α in (17) is indeterminable. Therefore, the optimal coefficient, in (17), may be approximated by the fixed coefficient, α_f , defined by the relation

$$\alpha_f = E\left\{\frac{\psi(z)}{z}\right\}; \frac{\psi(z)}{z} \approx \psi'(z) \tag{25}$$

where $\psi'(\cdot)$ is the first derivate, or the slope, of the ψ -function.

Particularly, for the Huber’s ψ_H -function in (14), the relation (25) may serve to explain the physical meaning of the fixed coefficient, α_f , and to estimate its value. Starting from (14) and (25), one gets

$$\alpha_f = E\{\psi'_H(z)\} = \int_{|z|\leq\Delta} p(z)dz = 1 - \delta \tag{26}$$

where $p(\cdot)$ is the unknown measurement noise pdf, in (15). Here is assumed that the real pdf of the scaled residual, ε/s in (23), also belongs to the given pdf class, in (15), (for more details, see Appendix 2). The integral, in (26), is equal to the probability that the observations are generated by the nominal standard Gaussian pdf, corresponding to the linear part of the ψ_H -function in (14), with the slope, ψ'_H , being equal to one. In accordance with (15), the underlying probability may be estimated by (26), using the assumed contamination degree, δ . It should be noted that the calculation of the fixed coefficient, α_f , in (26) may be also based on the worst-case pdf, p_0 , within the given class (15), resulting in

$$\alpha_f = E\{\psi'_H(z)\} = \int_{|z|\leq\Delta} p_0(z)dz = 2(1 - \delta)erf(\Delta) \tag{27}$$

where erf is the error function, [17, 20]. This solution is asymptotically equal to (26), since the value of the erf function is close to 0.5 for a large enough argument.

Thus, the relations (26), or (27), define the fixed and feasible approximations of the optimal statistical linearization coefficient, in (17), representing an approximation of the Huber’s M-robust influence function average slope, where the influence function, ψ_H , is defined by (14).

Thus, in the absence of outliers, corresponding to the zero-valued contamination degree, δ , the α value, in (26) or (27), is equal to one, reducing the robust gain, K_k in (23), to the optimal Kalman gain, in (6). Since the robust influence function, ψ_H , in (14), operates in its linear regime, corresponding to the linear influence function of the optimal Kalman filter, (3)–(7) the robust recursive estimator, (23), (24), performs as the optimal Kalman filter. However, in the presence of outliers, the fixed factor, α_f in (26) or (27), decreases with the contamination degree increased values, decreasing further the values of the robust gain matrix, K_k . in (23). Since the influence function, ψ_H in (14), now

operates in its saturation regime, the combination of these two effects suppresses the influence of outliers to the robust recursive estimates, in (23).

On the other hand, the variable approximation of the optimal statistical linearization coefficient, in (25), is given by

$$\alpha(\varepsilon_k/s_k) = \alpha_k = \begin{cases} \frac{\psi_H(\varepsilon_k/s_k)}{\varepsilon_k/s_k} & \text{for } \varepsilon_k \neq 0 \text{ and } s_k \neq 0 \\ 1 & \text{for } \varepsilon_k = 0 \text{ or } s_k = 0 \end{cases} \quad (28)$$

where the expectation, in (26), is replaced by the current sample. This approximation represents the current slope of the Huber's M-robust influence function, in (14), being approximately equal to zero or one. Thus, in the absence of outliers, the variable coefficient, α_k in (28), has the unit value, so the robust gain matrix, K_k in (23), is reduced to the optimal Kalman gain, in (6). Since the Huber's influence function, in (14), operates in the linear regime, the robust recursive estimator, in (23), performs as the optimal linear Kalman filter. On the other hand, in the presence of outliers, the variable α_k -coefficient, in (28), is close to zero-value, decreasing significantly the robust gain, K_k in (23), while the influence function, ψ_H , now is confined to its saturation regime, thus suppressing more efficiently the influence of outliers, in comparison with the application of the fixed coefficient, in (26) or (27).

In summary, the proposed statistically linearized M-robustified Kalman filtering algorithm consists of the time update, in (24), and the measurement update given by (11), (14), (19), (22), (23) and (26), or (28). It belongs to a class of recursive stochastic procedures, and some theoretical analysis of the estimates convergence is very difficult, due to a nonlinear form of the robust recursive state estimator by itself and a time-varying system state dynamics. Therefore, further analysis is based on both the simulations, using a manoeuver target radar tracking example, and the real data, related to an object tracking in the video sequence, generated by the short-wave infrared camera.

4 Experimental results and discussion

As mentioned above, the simulation example is related to the radar tracking problem in conditions close to reality. In this sense, the radar measurements usually consist of range, azimuth and elevation angles, since the observation noises are uncoupled in the spherical coordinate system (SCS). However, a requirement for simple filtering implies the desirability of the uncoupled filtering in the Cartesian coordinate system (CCS) [40]. Thus, the CCS would employ the three independent Kalman filters in each of the coordinates, (x, y, z) . In addition, if the sampling period is larger than the target manoeuver time constant, the computationally convenient reduction to the three independent two-state (position and velocity components) Kalman filters in the (x, y, z) directions is recommended, because the Kalman gains associated with the acceleration terms are rather small [40]. As a consequence, the state noise covariance, Q , has to be chosen so to compensate for the missing acceleration terms. Bearing in mind that the measurements when transformed from the SCS to the CCS are no more uncoupled, the proposed approach represents a trade-off between a potential performance loss and computational feasibility. Since the obtained simulation results are very similar in each

of the (x, y, z) - CCS directions, only the simulation results related to the x -CCS axis are presented in the sequel.

The first simulation task is to model the system state dynamics, using the kinematic equations of motion [40]. Thus, if x_k, v_k and a_k indicate the target position, velocity and acceleration, respectively, at the discrete time, $t_k = kT, k = 0, 1, \dots$, x -CCS axis, with T being the uniform sampling period, and assuming that the acceleration is constant over the sampling interval, $t_k \leq t \leq t_{k+1}$, one obtains, by integrating the acceleration twice time over the given interval, the following set of equations (the equivalent equations may be written for the y and z -CCS directions)

$$\begin{aligned} x_t &= x_k + v_k(t - t_k) + \frac{1}{2}a_k(t - t_k)^2 \\ v_t &= v_k + a_k(t - t_k) \\ a_t &= a_k \end{aligned} \tag{29}$$

A particular target position trajectory may be obtained from (29), by defining in advance a piece-wise constant acceleration profile. The model (29) with the zero-valued acceleration term is known as the constant velocity (CV) one. Thus, any target movement that cannot be represented by the CV-model may be considered as the target manoeuvre [40]. An example of such target position trajectory, used in the simulations, is presented in Fig. 1. The measurement sequence is simulated using the linear position sensor, represented by

$$y_k = x_k + v_k \tag{30}$$

where the zero-mean white measurement noise sequence, $\{v_k\}$, is confined to the pdf (15).

In a mono-pulse radar, such heavy-tailed feature of the underlying observation noise pdf is associated with the large target glint spikes, representing the outliers [40]. A sample of the random variable, v_k , with such pdf may be generated by firstly taking a sample, u , belonging to the (0,1)-uniform pdf. Thus, if the sample, u , is greater than the δ -value, the sample, v_k , is generated from the standard zero-mean Gaussian pdf with the unit variance; otherwise, the sample, v_k , is generated from the contaminating zero-mean Gaussian pdf with the assumed huge variance $\sigma_o^2 \gg 1$. Observations are generated in a separate computer program from a set of the true kinematic equations

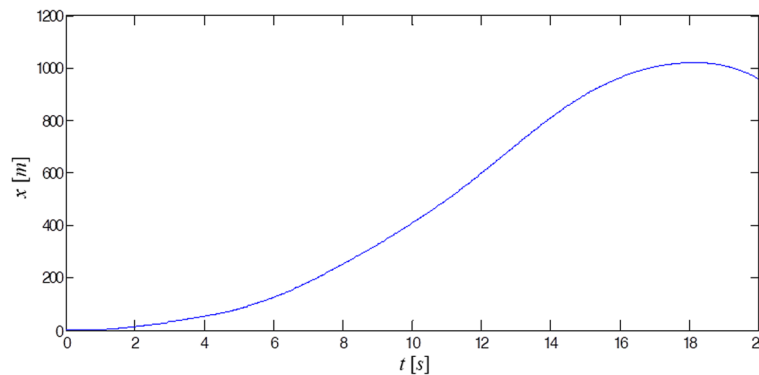


Fig. 1 The exact target position trajectory used in simulations

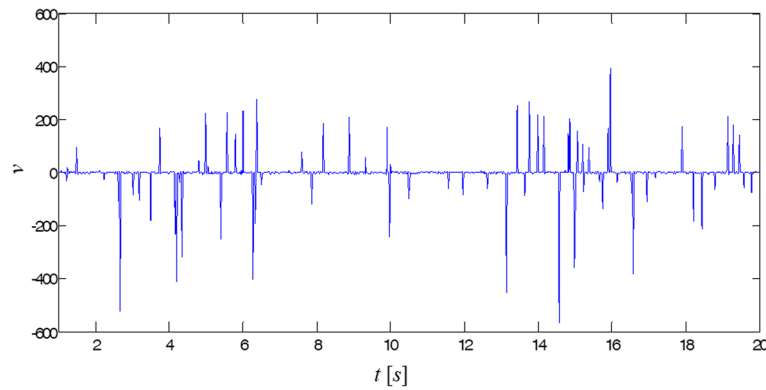


Fig. 2 A typical observation noise sample, from δ -contaminated Gaussian pdf, Eq. (15)

of the target motion, (29) and (30), and the previously obtained noise sample, v_k . A typical observation noise record is depicted in Fig. 2. Besides, the filter world is represented by the two-dimensional, discrete, time-invariant state-space model in the form (1), (2), given by

$$\begin{aligned} x_{k+1} &= Fx_k + Gw_k; \\ y_k &= Hx_k + v_k \end{aligned}; F = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, H = [1 \ 0] \tag{31}$$

The state transition matrix, F , follows directly from (29) by taking $t = t_{k+1}$ and neglecting the acceleration term, while the observation or information vector, H , follows from (30). The zero-mean white state noise sequence, $\{w_k\}$, is introduced artificially to compensate for the unmodelled system dynamics, associated with the unknown target manoeuvre. The variances of the noise sequences, $\{v_k\}$ and $\{w_k\}$, are given by $Q = 0.1$ and $R = 1$, respectively. Moreover, the uniform time-step $T = 0.02s$ is used. The following algorithms have been compared: the linear optimal Kalman filter (3)–(7), designated as A1; the statistically linearized M-robustified Kalman filter (11), (14), (17), (19), (22)–(24) with the variable statistical linearization coefficient in (28), designated as A2; the statistically linearized M-robustified Kalman filter (11), (14), (17), (19), (22)–(24) with the fixed statistical linearization coefficient in (26), designated as A3; and the quasi-linear approximation of the algorithm A2, based on the linear residual transformation, in (23), instead of the nonlinear one in (14), together with application of the same nonlinear residual processing in computing the adaptive gain matrix, K_k , as in A2, designated as A4.

Here, the initial state estimate, \hat{x}_0 , and the corresponding covariance, P_0 , are calculated using a suboptimal procedure based on the first two observations, [40]

$$\hat{x}_0 = \begin{bmatrix} y_2 \\ \frac{y_2 - y_1}{T} \end{bmatrix}; P_0 = \begin{bmatrix} 1 & 1/T \\ 1/T & Q + 2/T^2 \end{bmatrix} \tag{32}$$

The performances of the analysed filters are compared both in terms of the estimated and the true position profiles, as well as the cumulative estimation error criterion

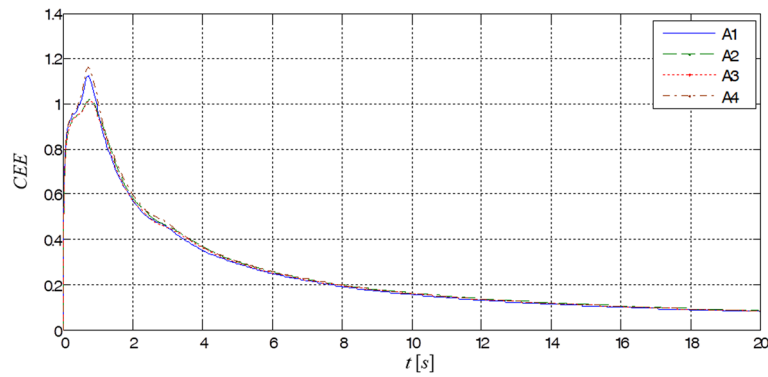
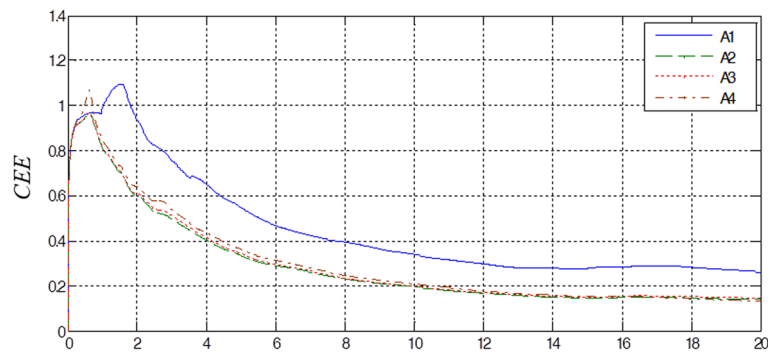
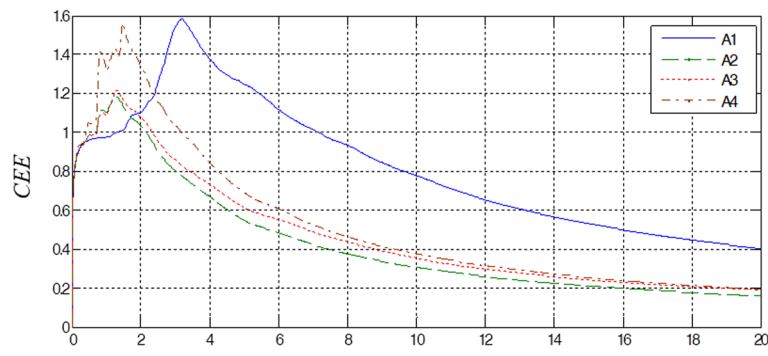


Fig. 3 Comparison by CEE measure (33) of different algorithms under pure Gaussian observations, Eq. (15) with $\delta = 0$



(a)



(b)

Fig. 4 Algorithms comparison by CEE criterion (33) under δ -contaminated Gaussian observation noise pdf (15): **a** $\delta = 0.07$; **b** $\delta = 0.15$

$$CEE(k) = \frac{1}{k} \sum_{i=1}^k \frac{\|\hat{x}_k - x_k\|^2}{\|x_k\|^2} \tag{33}$$

with $\|\cdot\|$ being the Euclidean norm, where the true target position trajectory, x_k , is depicted in Fig. 1. The CEE criterion values obtained for different algorithms and different measurement noise realizations in (15) are presented in Figs. 3 and 4. The results plotted in Fig. 3 have shown that the robustified Kalman filters A2-A4 satisfy the

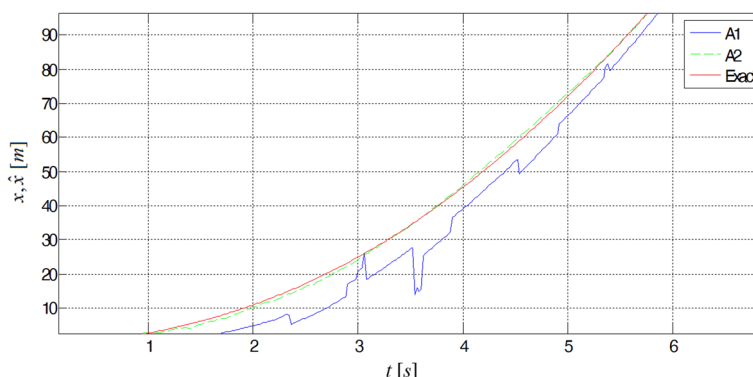


Fig. 5 The parts of the true and the estimated target trajectories, generated by the algorithms A1 and A2

efficiency robustness requirement, since the obtained values of the criterion (33) for these algorithms are not significantly larger, in comparison with the optimal Kalman filter, A1, under the pure Gaussian observations. In addition, the state estimators A2–A4 also satisfy the resistant robustness requirement, producing significantly smaller values of the criterion (33) than the optimal Kalman state estimator, A1, in the presence of outliers within the Gaussian observations, as depicted in Fig. 4. The parts of the true and the estimated target trajectories, generated by the algorithms A1 and A2, are depicted in Fig. 5. Similar results are obtained for the algorithms A3 and A4. However, an analysis of the estimator performances using the true and the estimated profiles is not suitable, since the target positions on the trajectories are expressed in much larger units, in comparison with the values of the underlying estimation errors. Therefore, the adopted *CEE* criterion in (33) is a more suitable factor of goodness, concerning the estimation quality. In this sense, the simulation results presented in Figs. 3, 4, 5 have shown that the proposed robust filters A2–A4 obey the practical robustness requirements.

Moreover, extensive Monte Carlo simulations have shown that the robustified versions A2–A4 of the optimal linear Kalman filter, A1, perform fairly well for the contamination degree $\delta \leq 0.3$, since for greater δ -values the observation noise model (15) is no more adequate. Furthermore, the best performances are obtained for the algorithm A2, owing to the common effects of the nonlinear residual transformation, in (14), and the calculation of the gain matrix, in (23), using an adaptive robustifying linearization coefficient, in (28). In words, these effects result in the values of the gain matrix large enough to produce a good tracking feature, but also small enough to provide for the noise reduction. The algorithm A4 produces a slightly worst result than A2. A disadvantage of the algorithm A3 is the application of the fixed linearization coefficient, in (26), depending on the exactly unknown contamination degree, δ , in (15) that cannot be estimated properly from the residuals, [15, 17]. Additionally, in the presence of outliers, the fixed linearization coefficient, in (26), reduces the gain matrix values, in (23), but the so-obtained gain factor values are larger than the gain values generated by the adaptive factor, in (28). This, in turn, makes the underlying state estimates more sensitive to impulsive noise, or outliers, in comparison with the algorithms A2 and A4. In this sense, although the algorithm A4 is linear in the observations, as A1, it utilizes the nonlinear robust data

processing in calculating the gain matrix, as in A2, suppressing efficiently the influence of outliers.

The second part of the experimental results is devoted to the real data, concerning to object tracking in the video sequence using the short-wave infrared camera. In this sense, the goal of the video tracking is an estimation of the location of a moving object in the video sequence. For the experimental analysis of a single moving object tracking in the video sequence, a kernelized correlation filter (KCF), [41], is used as a basic tracker since it is the one of the fastest trackers that does not require the graphics processing unit for real-time processing, [42]. In video tracking, the occlusions are among the most challenging problems, [43]. Although the KCF algorithm performs very well under the regular conditions, its performance decreases in the presence of occlusions. In this sense, when the tracked object disappears, due to the full occlusion, the KCF tracker will get stuck at the position of occlusion and continue to track the background, as the object of interest. To overcome this problem, the prediction and the estimation of the object's motion dynamics are required. Thus, the KCF tracker is synced with the Kalman filter to improve the tracking performance when the object is occluded. Furthermore, during the occlusion period, the tracked object may perform a manoeuvre. However, in a case of the object manoeuvre under the occlusion, it may be happened that the object is not re-detected after the occlusion, [44]. In this sense, the search area (window) used by the KCF tracker is not sufficient for object re-detection after occlusion, [41]. Therefore, when the occlusion is detected, the extended search area is used for the possible object re-detection after occlusion. Here, for occlusion detection, the peak-to-sidelobe ratio (PSR) metrics is used, [44]. The extended object search area is implemented by replicating the search windows around the central one, [45]. The dynamics of the tracked object, and those of the central search window position, are estimated by the Kalman filter, defining at the same time the central position of the extended search area. In this way, by estimating the dynamics of the object's motion, and by expanding the object search area, it is possible to overcome occlusions, and re-detect the object after occlusion, providing for continued tracking.

The object being tracked is represented by a bounding box, which is defined with the centre (x_c, y_c) in the image plane, and the corresponding height and width. To approximate the inter-frame object position (bounding box centre) displacements, the linear constant velocity model, in (31), is applied in the two directions, x and y , yielding the state-space model in the form (1), (2). Thus, the system state vector, X , and the corresponding system matrices, F , G , and H , are given by

$$X = \begin{bmatrix} x_c \\ \dot{x}_c \\ y_c \\ \dot{y}_c \end{bmatrix}; F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}; H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}; G = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (34)$$

with \dot{x} , \dot{y} , being the first derivative, or the velocity, of the state vector components, x and y , respectively, and T is equal to one. Thus, the two independent two-state (position and velocity components) Kalman filters in the x and y directions are used, in (34). Kalman filter, defined by (34), is initialized on the first video sequence frame with the ground-truth object position. The corresponding estimation error covariance matrix, P_0 ,

the state noise covariance matrix, Q , and the observation noise covariance matrix, R , are defined as follows:

$$P_0 = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}; Q = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}; R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (35)$$

Particularly, the object tracking in infrared imagery is considered in the sequel. For this purpose, the two characteristic video sequences are recorded, using the “Vlatacom electro-optical surveillance system,” [46]. The video sequences are recorded using the short-wave infrared camera, with the resolution of 576×504 pixels and the 25 frames per second. The recorded scenarios cover the typical urban scenes in real-life surveillance applications, including the single moving object and the various type of occlusions, such as static and moving, partial or complete, and short-term and long-term occlusions. The objects of interest for tracking are pedestrians. The first video sequence contains 825 frames with the partial and the full static occlusions. The second one has 225 frames, and the tracked object is occluded by the partial moving and the full static occlusions.

Starting from the two recorded sequences, and analysing the moving object tracking, using the combination of the improved KCF algorithm, with an extended search area and the standard Kalman filter, it can be concluded that the appearance of different types of occlusions results occasionally in large intensity errors, which may be treated as outliers. Thus in Fig. 6 are shown the position errors for the vertical, y , and the horizontal, x , directions. The position measurement errors, which deviate significantly from the majority of the population in the central cluster, in Fig. 6, present the outliers, caused by various occlusions. As mentioned before, the standard Kalman filter has the linear influence function, so that it is sensitive to outliers or non-robust. As a consequence, these errors in the measurement data can lead to the object loss and the tracking failure. Therefore, the M -robustified statistically linearized Kalman filter with the variable linearization coefficient, denoted above as A_2 , is proposed to suppress the influence of outliers in the video tracking applications.

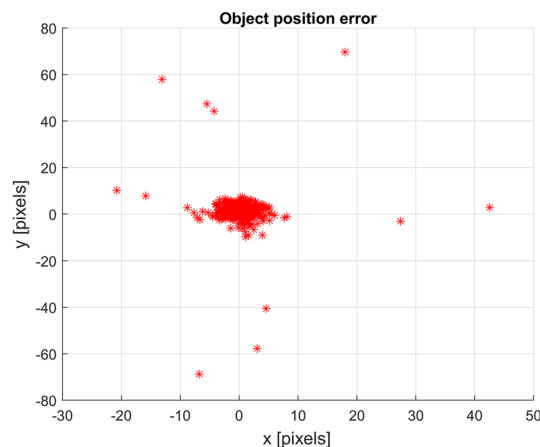


Fig. 6 Object position errors on the created video sequences with occasionally large errors caused by the occlusions, representing outliers

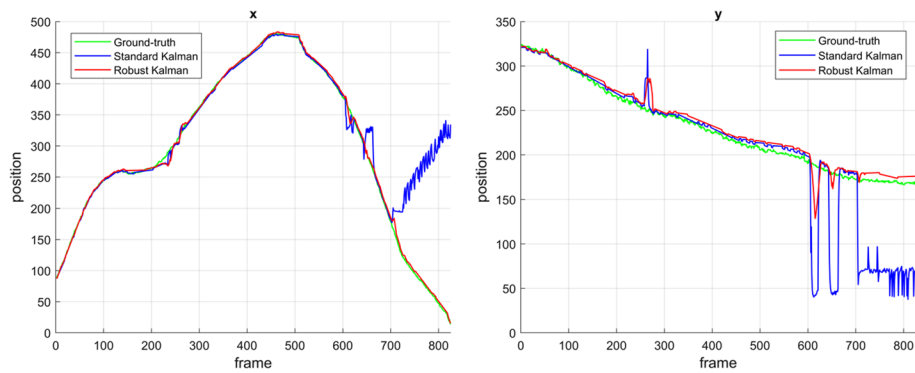


Fig. 7 Comparison of different algorithms in the presence of the occlusions in the first test video sequence; object position error in the horizontal direction (x)—left; and object position error in the vertical direction (y)—right

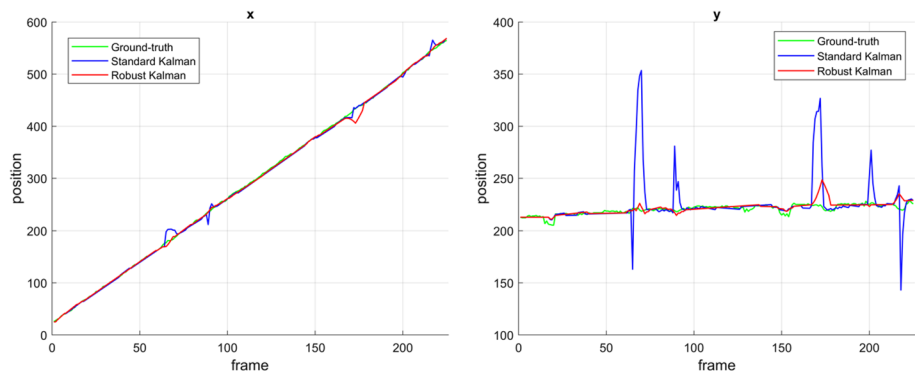


Fig. 8 Comparison of different algorithms in the presence of the occlusions in the second test video sequence; object position error in the horizontal direction (x)—left; and object position error in the vertical direction (y)—right



Fig. 9 The first test video sequence. Pedestrian tracking in scenario with partial and full static occlusions. Comparison of bounding boxes obtained by standard Kalman filter (blue box) and robust Kalman filter (red box) to ground-truth (green box) bounding box



Fig. 10 The second test video sequence. Pedestrian tracking in scenario with partial moving and full static occlusions. Comparison of bounding boxes obtained by standard Kalman filter (blue box) and robust Kalman filter (red box) to ground-truth (green box) bounding box

The position errors of the different algorithms, based on the combination of the improved KCF tracker with the standard and the M-robustified adaptive Kalman filters, are shown in Fig. 7 (the first recorded video sequence) and in Fig. 8 (the second recorded video sequence). To clearly demonstrate tracking performance on real data in Fig. 9 (the first video sequence) and Fig. 10 (the second video sequence) are shown frames from these sequences with ground-truth bounding boxes, and bounding boxes generated by the standard and the robust Kalman filter. Figures 9 and 10 show that in the case of occlusions, the algorithm using the standard Kalman filter significantly deviates from the ground-truth position. In the scenario in Fig. 9, there is a complete loss of the object of interest, which can also be seen in Fig. 7. In the scenario in Fig. 10, although there is no loss of the object, at the moment of occlusion, the error of the algorithm is very large, which is confirmed by the graphic in Fig. 8. In the system with a camera on the pan-tilt [46], these large position errors may lead to a sudden movement of the system and the loss of the object from the field of view. On the other hand, the robust Kalman filter approach successfully overcomes occlusions and continues tracking more smoothly. The obtained results, based on the real data, confirm the earlier derived conclusions from the simulation results. The presented experimental results also indicate a possibility of designing an efficient robust tracking system in the video surveillance applications, being a combination of the KCF tracker and the proposed adaptive M-robustified version of the optimal Kalman filter.

5 Conclusion

Kalman filter produces the optimal state estimates of a linear dynamic stochastic system in the presence of the Gaussian distributed both the random input, the so-called state noise and the additive measurement noise. The optimality of Kalman filter is related to its predictor-corrector recursive form and the computation of the gain sequence. However, presence of the erroneous noise statistics and/or miss modelling may cause significant deviations from the theoretically optimal performances.

Starting from these practical limitations of the optimal Kalman filter, a new class of the statistically linearized M-robustified Kalman filtering algorithms has been proposed in this article. The proposed robust algorithms are feasible and provide for the recursive dynamic system state estimation. The article also produces the complete derivation of the algorithms to be handled. In this sense, the time-update recursion is designed accordingly to the optimal Kalman filter, and the measurement-update recursion is designed as the nonlinear dynamic stochastic approximation procedure, generated by minimizing at each stage the generalized time-varying Huber's M-robust performance index. Thus, the optimal Kalman filter robustification is obtained by nonlinear transformation of the scaled residuals through Huber's M-robust influence function. Analogously to the standard Kalman filter, the robust recursive estimator gain matrix is computed from an additional optimization procedure of the minimum variance type. The posed nonlinear optimization process applies the statistical linearization technique to provide for a suboptimal robust version of the Kalman gain matrix. Since the determination of the mean-square optimal statistical linearization coefficient assumes the exact knowledge of the observation noise statistics, both the fixed and the variable approximations of the

optimal coefficient are proposed. Thus, the fixed approximation of this coefficient represents an approximation of the average slope of the Huber's M -robust influence function that is estimated further by the assumed probability of outliers occurrence. A variable version of such fixed coefficient is obtained by approximating the expectation by the current sample, resulting in the present slope of M -robust influence function.

Theoretical convergence analysis of the proposed robust algorithms is difficult, due to both their nonlinear forms and a time-varying multidimensional system dynamics. Therefore, practical robustness of the derived state estimators, including the resistant and the efficiency robustness, is analysed by simulations, using a single manoeuvring target tracking example. The experimental results also allow understanding of the algorithms operations, with and without outliers, where each case is accomplished by an adequate robust gain matrix. Additionally, it indicates to the conclusion that both the nonlinear transformation of the scaled measurements residuals, using the Huber's M -robust influence function, and the robustified computation of the gain matrix, applying an adaptive statistical linearization coefficient, provide for a good compromise between the tracking performance and the noise immunity. In words, the variable coefficient is adapted properly to the nonlinear form of the M -robust influence function, reducing the effects of outliers. Moreover, the fixed statistical linearization coefficient results in a slower decrease in the gain matrix, in comparison with the variable one. This, in turn, eliminates the effects of outliers worse than in the case of the variable statistical linearization coefficient. A quasi-linear approximation of the proposed statistically linearized M -robustified Kalman state estimator, based on the linear residual transformation, together with the adaptive nonlinear residual processing in calculating the robust gain matrix, produces a slightly worse performance than the starting nonlinear robust estimator. Moreover, the experimental results based on real data, concerning a video tracking, using short-wave infrared camera, are also analysed. The real data consist of the two recorded video sequences, representing the typical urban scenes in the real-life surveillance applications including the pedestrian as an object of interest in the scenarios with various types of occlusions (static or moving, partially or complete, short and long term). The application of the proposed video tracker, being the combination of the improved kernelized correlation filter and the M -robustified statistically linearized Kalman filter, provides an efficient, robust method for tracking of manoeuvring object in the presence of occlusions. These results are in accordance with the conclusion derived from the simulations and indicate to a possibility of designing an efficient robust video tracking algorithm.

The proposed statistically linearized M -robustified filtering technique can be also applied to some redescending influence function that may be better in eliminating the influence of outliers. However, the robust score function associated with such influence function, and determining the M -robust performance measure to be minimized, is not the convex one. Therefore, there could be convergence problems during the robust filter initialization. The problem may be circumvented by applying the two-step estimation procedure, where in the first step the proposed M -robust version of the Kalman filter, based on the Huber's monotonously non-decrease influence function, is applied. This, in turn, generates the good initial guesses to the M -robust version of the Kalman filter, based on a redescending M -robust influence function, in the second step.

Finally, an approximation of the nonlinearities in the motion and measurement equations of a nonlinear stochastic dynamic system by the statistical linearization technique can be combined with the proposed statistically linearized M-robustified Kalman filter to obtain a robust recursive state estimator of the approximate minimum variance type. Although these equations look like those of the proposed statistically linearized M-robustified Kalman filter, they are much more complex. In this sense, much auxiliary computation is needed to obtain the various vectors and matrices of the corresponding expectations, defining the statistical linearization coefficients. Thus, the degree of difficulty in calculating the underlying coefficients is a significant argument in deciding whether to apply this method in practice. However, calculated values of the required quantities can be stored for “look-up” during the state estimation, greatly reducing the computation efforts.

In summary, since the most industrial and scientific data contain unavoidable outliers, owing to metre and communication errors, incomplete measurements, errors in mathematical models, etc., the proposed recursive robust estimator may be applied in different problems, including system identification, state estimation, signal processing and adaptive control.

Appendix 1 Derivation of the statistically linearized M-robustified versions of the Kalman filter (algorithms A2, A3 and A4)

Starting from (1) and (2), one obtains for the prediction error in (20) the relation

$$\tilde{x}_k(-) = F_{k-1}\tilde{x}_{k-1}(+) + G_{k-1}w_{k-1} \tag{36}$$

where the estimation error, $\tilde{x}_k(+)$, is defined by (16), while the variable, w_k , is a sample from a zero-mean white state noise, in (1). Furthermore, taking into account (1), (2), (11) and (18), the estimation error in (16) is given by

$$\tilde{x}_k(+) = \tilde{x}_k(-) - K_k H_k \tilde{x}_k(-) - K_k v_k \tag{37}$$

where the variable, v_k , is a sample from the zero-mean white measurement noise, in (2). Analogously to the linear optimal Kalman filter, the initial condition defined in Sect. 2 guarantees that the prediction and the estimation errors are unbiased at each stage, k , resulting in

$$E\{\tilde{x}_k(-)\} = 0; \quad E\{\tilde{x}_k(+)\} = 0; \quad k = 1, 2, \dots \tag{38}$$

The proof is based on the mathematical induction, [1–5]. In addition, under the assumed hypothesis in Sect. 2, the random errors, $\tilde{x}_k(+)$ and $\tilde{x}_k(-)$, are uncorrelated with the state and the measurement noises, producing the underlying cross-covariances

$$E\{\tilde{x}_{k-1}(+)w_{k-1}^T\} = 0, E\{\tilde{x}_k(-)v_k^T\} = 0 \tag{39}$$

where 0 denotes a zero matrix. Starting from (A1), (A2) and (A4), one obtains for the prediction error covariance in (20) the relation (24), that is identical to the time update in (3), (4). Moreover, due to (A2) and (A4), the estimation error covariance in (16) can be calculated at each stage by the recursive relation

$$P_k = M_k - K_k H_k M_k - M_k H_k^T K_k^T + R_k K_k K_k^T \tag{40}$$

where R_k is the scalar measurement noise variance. By substituting the gain, K_k , from (18) into (A5), and applying the matrix trace operation, one obtains for the approximate minimum variance criterion, in (16), the following expression

$$J_1(\Gamma_k) = \text{Trace } M_k - 2s_k^{-2}\alpha \text{Trace } \Gamma_k H_k^T H_k M_k + s_k^{-4}\alpha^2 R_k \text{Trace } \Gamma_k H_k^T H_k \Gamma_k \tag{41}$$

where the scaling factor, s_k , is defined by (19). In deriving (A6) is used the fact that the third matrix term in (A5) represents the transpose matrix of the second one, yielding the same matrix trace. Further step in the derivation is based on the usage of the rules for the partial derivate of trace of product of matrices, [1–5].

$$\frac{\partial}{\partial A} \text{Trace } BAC = B^T C^T; \frac{\partial}{\partial A} \text{Trace } ABA^T = 2AB; \text{ if } B = B^T \tag{42}$$

By comparing (A7) with (A6), one concludes that $B = I$, $A = \Gamma_k$ and $C = H_k^T H_k M_k$ for the second term in (A6), while $A = \Gamma_k$ and $B = H_k^T H_k$ for the third one. Taking into account these equivalences, one obtains, by partially differentiating (A6) and equating the resulting matrix equation with the zero matrix, the following algebraic equation

$$\frac{\partial J_1(\Gamma_k)}{\partial \Gamma_k} = -2s_k^{-2}\alpha M_k H_k^T H_k + 2s_k^{-4}\alpha^2 R_k \Gamma_k H_k^T H_k = 0 \tag{43}$$

The matrix equation (A8) requires further simplifications, to generate a feasible sub-optimal solution for Γ_k . Firstly, since the value of the optimal statistical linearization coefficient, α , in (17) is close to one, one can replace α^2 in the second term of (A8) by α . Moreover, by using (19), it further follows

$$s_k^{-2}R_k = \left(R_k^{-1} H_k M_k H_k^T + 1 \right)^{-1} \approx 1 \tag{44}$$

Namely, the first term in brackets of (A9) is proportional to the uncertainty in the prediction, expressed by the covariance, M_k , in (20) and (24), but inversely proportional to the measurement noise average power, R_k . Moreover, the proposed nonlinear filter, based on the winsorization technique, in (14) and (23), obeys the efficiency robustness requirement. In this sense, it is almost efficient as the optimal linear Kalman filter under the pure Gaussian observations, but retains a good efficiency under the existence of outliers within the Gaussian observations. Therefore, the estimation error covariance, P , in (24) is rather small, so that the uncertainty in the prediction, M_k , is directly proportional to the state noise average power, Q . Moreover, the measurement noise variance, R , determined by (21), is significantly larger than the state noise variance, Q . As a consequence, the equation (A9) reduces to the unit value, as it is expressed by its right-hand side. Under the adopted two approximations, the relation (A8) takes the form

$$\frac{\partial J_1(\Gamma_k)}{\partial \Gamma_k} \approx -2s_k^{-2}\alpha (M_k - \Gamma_k) H_k^T H_k = 0 \tag{45}$$

Bearing in mind (18) and (A10), it further follows

$$M_k = \Gamma_k; K_k = s_k^{-2} \alpha M_k H_k^T \tag{46}$$

Finally, by replacing (A11) into (A5), one obtains the estimation error covariance expressed by

$$P_k = M_k - \alpha s_k^{-2} M_k H_k^T H_k M_k - \alpha s_k^{-2} M_k H_k^T H_k M_k + \alpha^2 s_k^{-4} R_k M_k H_k^T H_k M_k \tag{47}$$

By applying the mentioned two approximations, $\alpha^2 \approx \alpha$ and (A9), as well as by subtracting the gain, K , from (A11) into (A12), the last relation reduces to (22).

Finally, starting from the approximation of α by α^2 in (18), the relation (23) is obtained by substituting Γ_k from (22) into (18), and including in the so-obtained equation the ψ -function, in (14), instead of its statistical linear approximation in (17). Thus, if one uses the variable α in (28), having the value close to zero or one, the above approximation of α with α^2 is reasonable. On the other hand, the fixed approximation of α , in (26) or (27), is equal to the probability of the regular observations, and for a small or moderate contamination degree, δ in (15), the probability value is close to one, justifying the above approximation. Of course, this value decreases with the contamination degree increased values, reducing the gain factor values, K_k in (23), suppressing the influence of outliers.

A class of statistically linearized M-robustified Kalman filtering algorithms, with α being the free parameter, is defined by the prediction recursions (24), and the estimation, or filtering, recursions defined by (11), (14), (19), (22) and (23). Particular robust algorithm is defined by choosing a suitable approximation of the indeterminable mean-square optimal statistical linearization coefficient, α , in (17). Thus, the choice of the fixed approximation, in (26) or (27), of the optimal coefficient, α , in (17) results in the algorithm A3. Furthermore, one obtains the algorithm A2 by choosing the variable approximation, in (28), of the optimal coefficient, α , in (17). The algorithm A4 follows from the algorithm A2 by replacing the nonlinear M-robust influence function, ψ , in (23) with the linear one, but the robust mechanism to generating the gain matrix, K , in (23), remains unchanged, as in A2. The derivation of the proposed recursive robust algorithms is thus completed.

Appendix 2 Derivation of the optimal statistical linearization coefficients

Statistical linearization is a type of the statistical approximation techniques, where the basic principle is to approximate a given vector-valued function, $\psi(z)$, of a random vector argument, z , by the linear matrix form

$$\psi(z) = \alpha z + \beta \tag{48}$$

Here, the parameters, α and β , are some matrix coefficients that have to be determined. Analogously to the estimation problem, by defining the function representation error

$$e = \psi(z) - \beta - \alpha z \tag{49}$$

these coefficients may be calculated by minimizing the mean-square error (MSE) criterion

$$J(\alpha, \beta) = \text{Trace} E\{ee^T\} = E\{e^T e\} = E\{\|e\|^2\} \tag{50}$$

Here, *Trace* is the matrix trace, with $\|\cdot\|$ being the Euclidian norm, while $E\{\cdot\}$ denotes the mathematical expectation, [1–5]. Substituting (B2) in (B3), and changing the order of linear operators, *Trace* and $E\{\cdot\}$, together with the application of the rules in (A7), one obtains

$$\frac{\partial J(\alpha, \beta)}{\partial \beta} = 2E\{\psi(z) - \beta - \alpha z\} \tag{51}$$

By setting (B4) equal to zero, it further follows

$$\beta = E\{\psi(z)\} - \alpha E\{z\} \tag{52}$$

Moreover, by substituting the coefficient, β , from (B5) into (B2) and (B3), and differentiating again the so-obtained equation with respect to the matrix α , using the rules (A7), one obtains

$$\frac{\partial J(\alpha)}{\partial \alpha} = 2E\{\alpha \tilde{z}\tilde{z}^T + [E\{\psi(z)\} - \psi(z)]\tilde{z}^T\}; \tilde{z} = E\{z\} - z \tag{53}$$

Setting (B6) equal to zero, and solving the resulting equation, it further follows

$$\alpha = \left[E\{\psi(z)z^T\} - E\{\psi(z)\}E\{z^T\} \right] P_z^{-1}; P_z = E\{\tilde{z}\tilde{z}^T\} \tag{54}$$

If $\psi(z)$ is a vector-valued function of the multidimensional argument, z , then the statistically linearized solutions in (B5) and (B7) require evaluations of the multidimensional integrals, following from the definitions of the corresponding multidimensional mathematical expectations. This assumes the joint pdf for the components of random vector, z , to be given in advance. Moreover, the most frequently adopted joint pdf is the multidimensional Gaussian one. The computation may be much simplified for nonlinearities with a small number of argument variables. Particularly, if both the random variable ψ and the random variable z are scalars with the zero-mean values, then β in (B5) and α in (B7) reduce to the scalar-valued deterministic quantities, given by

$$\alpha = \frac{E\{\psi(z)z\}}{\sigma_z^2}; \beta = 0 \tag{55}$$

with σ_z^2 being the variance of the random argument, z . More precisely, the equation (B8) assumes that the argument, z , is a zero-mean random variable with a symmetric pdf, while ψ is an odd real function of the scalar argument, z .

Particularly, the second assumption is fulfilled for the Huber’s robust influence function, $\psi(\cdot)$, in (14). Moreover, the random argument, z , in (B9) corresponds to the normalized scaled measurement residual, $\tilde{\varepsilon}_k = \varepsilon_k/s_k$, in (17), where the residual, ε_k , is given by (11), with s_k being its standard deviation calculated in (19). Therefore, by substituting the current measurement, y_k , from (2) into (11), one gets

$$\tilde{\varepsilon}_k = \varepsilon_k/s_k = (H_k \tilde{x}_k(-) + v_k)/s_k; s_k = \left(H_k M_k H_k^T + R_k \right)^{1/2}; R_k = 1 \tag{56}$$

where the prediction error, $\tilde{x}_k(-)$, in (A1) is a zero-mean random variable with the covariance M_k , computed by (24). Analogously to the optimal Kalman filter, the zero-mean error, $\tilde{x}_k(-)$, in (B9) is the Gaussian distributed if both the zero-mean white Gaussian noises, w_k and v_k , and the Gaussian initial state vector, x_0 , are also Gaussian distributed. In this sense, one can write

$$\tilde{x}_k(-) \sim N(\cdot|0, M_k) \tag{B10}$$

where $N(\cdot|a, b)$ denotes the Gaussian pdf with the mean a and the covariance b . Moreover, starting from (B10), the random variable, $H_k\tilde{x}_k(-)/s_k$, in (B9) has the zero-mean Gaussian pdf, defined by

$$H_k\tilde{x}_k(-)/s_k \sim N\left(\cdot|0, H_kM_kH_k^T/s_k^2\right) \tag{57}$$

Furthermore, a zero-mean observation noise, v_k , is confined to the Gaussian mixture pdf in (15), yielding the unit nominal variance, R_k , in (B9). This, in turn, results that the scaled random variable, v_k/s_k , in (B9) has the following Gaussian mixture pdf

$$v_k/s_k \sim (1 - \delta)N\left(\cdot|0, 1/s_k^2\right) + \delta N\left(\cdot|0, \sigma_0^2/s_k^2\right) \tag{58}$$

Thus, normalized residual, $\tilde{\varepsilon}_k$, in (B9) is defined by the sum of the zero-mean Gaussian random variables, so that its conditional pdf given the past observations, $p(\tilde{\varepsilon}_k|Y^{k-1})$, is defined by the convolution between the underlying Gaussian pdfs, to which the random addends in (B9) are confined. Additionally, since the convolution between the Gaussian pdfs yields also the Gaussian one, with the corresponding mean and covariance, one gets from (B9), (B11) and (B12) the conditional pdf of the scaled residual given the past measurement

$$p\left(\tilde{\varepsilon}_k|Y^{k-1}\right) = N\left(\cdot|0, H_kM_kH_k^T/s_k^2\right) \otimes \left\{ (1 - \delta)N\left(\cdot|0, 1/s_k^2\right) + \delta N\left(\cdot|0, \sigma_0^2/s_k^2\right) \right\} \tag{59}$$

where \otimes denotes the convolution integral, [5]. Since the convolution is a linear operator, the relation (B13) can be rewritten

$$p\left(\tilde{\varepsilon}_k|Y^{k-1}\right) = (1 - \delta)N\left(\cdot|0, \left(H_kM_kH_k^T + 1\right)/s_k^2\right) + \delta N\left(\cdot|0, \left(H_kM_kH_k^T + \sigma_0^2\right)/s_k^2\right) \tag{60}$$

Taking into account (19), the variance of the second normal pdf in (B14) can be approximated by

$$\frac{\sigma_0^2\left(\sigma_0^{-2}H_kM_kH_k^T + 1\right)}{R_k\left(R_k^{-1}H_kM_kH_k^T + 1\right)} \approx \sigma_0^2; R_k = 1 \tag{61}$$

The right-hand side approximation in (B15) follows from the approximate relation (A9), yielding

$$\sigma_0^{-2}H_kM_kH_k^T + 1 \approx R_k^{-1}H_kM_kH_k^T + 1 \approx 1 \tag{62}$$

By substituting (B15) into (B14), one obtains

$$p\left(\tilde{\varepsilon}_k | Y^{k-1}\right) = (1 - \delta)N\left(\tilde{\varepsilon}_k | 0, 1\right) + \delta N\left(\tilde{\varepsilon}_k | 0, \sigma_0^2\right) \quad (63)$$

Since (B17) is a symmetric pdf, the first assumption under which the expression (B8) is derived is also fulfilled, justifying the application of (B8) to the relation (17). Thus, the derivation is completed.

Abbreviations

A1	Linear optimal Kalman filter algorithm
A2	Statistically linearized M-robustified Kalman filter algorithm with the variable statistical linearization coefficient
A3	Statistically linearized M-robustified Kalman filter algorithm with the fixed statistical linearization coefficient
A4	Quasi-linear approximation of algorithm A2, based on the linear residual transformation, together with application of the nonlinear residual processing in computing the gain matrix
CCS	Cartesian coordinate system
CEE	Cumulative estimation error criterion
CV	Constant velocity
KCF	Kernelized correlation filter
ML	Maximum likelihood
MSE	Mean square error
pdfP	Probability distribution function
PSR	Peak-to-sidelobe ratio
SCS	Spherical coordinate system
SWIR	Short-wave infrared

Acknowledgements

The authors are grateful for the valuable comments and suggestions of the unknown reviewers that improved the final version of the manuscript.

The authors are also grateful to the Vlatacom Institute of High Technologies for using real datasets for infrared video surveillance.

Author contributions

MP designed the work, analysed and interpreted the data and drafted the manuscript. ZB participated in the design of the study, performed the experiments and analysis and helped to draft the manuscript. BK performed in the theoretical analyse and helped to produce the final version of the manuscript. All the authors read and approved the final manuscript.

Funding

This paper is an outcome of activities under project #157 VMSIS3_Advanced supported by the Vlatacom Institute of High Technologies, Belgrade, Serbia.

Availability of data and materials

Data are available on request from the authors.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 29 September 2022 Accepted: 6 June 2023

Published online: 21 June 2023

References

1. A. Gelb (ed.), *Applied optimal estimation*, Analytic Sciences Corporation (MIT Press, Cambridge, MA, 2010)
2. M.S. Grewal, A.P. Andrews, *Kalman filtering theory and practice using Matlab* (Wiley, Hoboken, NJ, 2015)
3. R. Stengel, *Stochastic optimal control* (Wiley, New York, 1986)
4. F. van der Heijden, B. Lei, G. Xu, F. Ming, Y. Zou, D. de Ridder, D.M. Tax, *Classification, parameter estimation, and state estimation: an engineering approach using Matlab* (Wiley, Hoboken, NJ, 2017)

5. Kovačević B., & Đurović Ž., *Fundamentals of stochastic signals, systems and estimation theory with worked examples* (Springer, Berlin, 2011)
6. M. Verhaegen, V. Verdult, *Filtering and system identification: a least squares approach* (Cambridge University Press, Cambridge, 2012)
7. J.V. Candy, *Model-based signal processing* (Wiley, Hoboken, NJ, 2006)
8. C. Price, An analysis of the divergence problem in the Kalman filter. *IEEE Trans. Autom. Control* **13**(6), 699–702 (1968). <https://doi.org/10.1109/tac.1968.1099031>
9. P. Hanlon, P. Maybeck, Characterization of Kalman filter residuals in the presence of mismodeling. *IEEE Trans. Aerosp. Electron. Syst.* **36**(1), 114–131 (2000). <https://doi.org/10.1109/7.826316>
10. A.E. Albert, L.A. Gardner, *Stochastic approximation and nonlinear regression* (The MIT Press, Cambridge, MA, 1967)
11. G. Saridis, Z. Nikolic, K. Fu, Stochastic approximation algorithms for system identification, estimation, and decomposition of mixtures. *IEEE Trans. Syst. Sci. Cybernet.* **5**(1), 8–15 (1969). <https://doi.org/10.1109/tssc.1969.300238>
12. S. Stanković, B. Kovačević, Analysis of robust stochastic approximation algorithms for process identification. *Automatica* **22**(4), 483–488 (1986). [https://doi.org/10.1016/0005-1098\(86\)90053-1](https://doi.org/10.1016/0005-1098(86)90053-1)
13. Tsytkin, I., *Adaptation and learning in automatic systems* (Academic Press, New York, 1971)
14. J.M. Mendel, *Adaptive, learning and pattern recognition systems: theory and applications* (Acad. Press, New York, 2012)
15. V. Barnett, T. Lewis, *Outliers in statistical data* (Wiley, Chichester, 2000)
16. W.N. Venables, B.D. Ripley, *Modern applied statistics with S* (Springer, New York, 2011)
17. P.J. Huber, E.M. Ronchetti, *Robust statistics* (Wiley, Hoboken, NJ, 2009)
18. R.R. Wilcox, *Introduction to robust estimation and hypothesis testing* (Academic Press, Amsterdam, 2017)
19. F.R. Hampel, E.N. Ronchetti, P.J. Rousseeuw, W.A. Stahel, *Robust statistics: the approach based on influence functions* (Wiley, Hoboken, NJ, 1986)
20. B. Kovačević, M. Milosavljević, M. Veinović, M. Marković, *Robust digital processing of speech signals*. Springer (2017)
21. C. Boncelet, B. Dickinson, An approach to robust Kalman filtering, in *The 22Nd IEEE Conference On Decision And Control* (1983). <https://doi.org/10.1109/cdc.1983.269847>
22. B. Kovačević, Ž. Đurović, S. Glavaški, On robust Kalman filtering. *Int. J. Control* **56**(3), 547–562 (1992). <https://doi.org/10.1080/00207179208934328>
23. Ž. Đurović, B. Kovačević, Robust estimation with unknown noise statistics. *IEEE Trans. Autom. Control* **44**(6), 1292–1296 (1999). <https://doi.org/10.1109/9.769393>
24. G. Chang, M. Liu, M-estimator-based robust Kalman filter for systems with process modeling errors and rank deficient measurement models. *Nonlinear Dyn.* **80**(3), 1431–1449 (2015). <https://doi.org/10.1007/s11071-015-1953-0>
25. L. Chang, B. Hu, G. Chang, A. Li, Robust derivative-free Kalman filter based on Huber's M-estimation methodology. *J. Process Control* **23**(10), 1555–1561 (2013). <https://doi.org/10.1016/j.jprocont.2013.05.004>
26. D. deMenezes, D. Prata, A. Secchi, J. Pinto, A review on robust M-estimators for regression analysis. *Comput. Chem. Eng.* **147**, 107254 (2021). <https://doi.org/10.1016/j.compchemeng.2021.107254>
27. M.A. Gandhi, L. Mili, Robust Kalman filter based on a generalized maximum-likelihood-type estimator. *IEEE Trans. Signal Process.* **58**(5), 2509–2520 (2010). <https://doi.org/10.1109/tsp.2009.2039731>
28. J. Valluru, S. Patwardhan, L. Biegler, Development of robust extended Kalman filter and moving window estimator for simultaneous state and parameter/disturbance estimation. *J. Process Control* **69**, 158–178 (2018). <https://doi.org/10.1016/j.jprocont.2018.05.008>
29. M. Murata, H. Nagano, K. Kashino, Unscented statistical linearization and robustified Kalman filter for nonlinear systems with parameter uncertainties, in *2014 American Control Conference* (2014). <https://doi.org/10.1109/acc.2014.6858583>
30. K. Li, B. Hu, L. Chang, Y. Li, Robust square-root cubature Kalman filter based on Huber's M-estimation methodology. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **229**(7), 1236–1245 (2014). <https://doi.org/10.1177/0954410014548698>
31. Y. Zou, S. Chan, T. Ng, Robust M-estimate adaptive filtering. *IEE Proc. Vis. Image Signal Process.* **148**(4), 289 (2001). <https://doi.org/10.1049/ip-vis:20010316>
32. Z.D. Banjac, B.D. Kovacevic, M.M. Milosavljevic, M.D. Veinovic, Local echo canceler with optimal input for true full-duplex speech scrambling system. *IEEE Trans. Signal Process.* **50**(8), 1877–1882 (2002). <https://doi.org/10.1109/tsp.2002.800415>
33. Z. Banjac, Kovačević, B., Veinović, M., & Milosavljević, M., Robust least mean square adaptive FIR filter algorithm. *IEE Proc. Vis. Image Signal Process.* **148**(5), 332–336 (2001). <https://doi.org/10.1049/ip-vis:20010594>
34. B. Kovačević, Z. Banjac, M. Milosavljević, *Adaptive digital filters* (Springer, Berlin, 2013)
35. B. Kovačević, Z. Banjac, I.K. Kovačević, Robust adaptive filtering using recursive weighted least squares with combined scale and variable forgetting factors. *EURASIP J. Adv. Signal Process.* (2016). <https://doi.org/10.1186/s13634-016-0341-3>
36. Z. Banjac, Z. Durovic, B. Kovacevic, Approximate Kalman filtering using robustified dynamic stochastic approximation method, in *2018 26th Telecommunications Forum (TELFOR)* (2018). <https://doi.org/10.1109/telfor.2018.8611817>
37. Z. Banjac, B. Kovačević, Robustified Kalman filtering using both dynamic stochastic approximation and M-robust performance index. *Tehnicki Vjesnik - Technical Gazette* **29**(3), 907–914 (2022). <https://doi.org/10.17559/tv-20200929143934>
38. S.C. Chapra, R.P. Canale, *Numerical methods for engineers* (McGraw-Hill Education, New York, NY, 2015)
39. T. Young, R. Westerberg, Error bounds for stochastic estimation of signal parameters. *IEEE Trans. Inf. Theory* **17**(5), 549–557 (1971). <https://doi.org/10.1109/tit.1971.1054696>
40. S.S. Blackman, R. Popoli, *Design and analysis of modern tracking systems* (Artech House, Norwood, MA, 1999)

41. J.F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 583–596 (2015). <https://doi.org/10.1109/tpami.2014.2345390>
42. S. Javed, M. Danelljan, F.S. Khan, M.H. Khan, M. Felsberg, J. Matas, Visual object tracking with discriminative filters and siamese networks: a survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(5), 6552–6574 (2022). <https://doi.org/10.1109/TPAMI.2022.3212594>
43. R. Xia, Y. Chen, B. Ren, Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter. *J. King Saud Univ. Comput. Inf. Sci.* **34**(8), 6008–6018 (2022). <https://doi.org/10.1016/j.jksuci.2022.02.004>
44. M. Zolfaghari, H. Ghanei-Yakhdan, M. Yazdi, Real-time object tracking based on an adaptive transition model and extended Kalman filter to handle full occlusion. *Vis. Comput.* **36**(4), 701–715 (2020). <https://doi.org/10.1007/s00371-019-01652-3>
45. J. Shin, H. Kim, D. Kim, J. Paik, Fast and robust object tracking using tracking failure detection in kernelized Correlation Filter. *Appl. Sci.* **10**(2), 713 (2020). <https://doi.org/10.3390/app10020713>
46. N. Latinović, I. Popadić, B. Tomić, A. Simić, P. Milanović, S. Nijemčević, M. Perić, M. Veinović, Signal processing platform for long-range multi-spectral electro-optical systems. *Sensors* **22**(3), 1294 (2022). <https://doi.org/10.3390/s22031294>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
