Austrian Journal of Statistics
2025, Volume 54, 150–168.
http://www.ajs.or.at/
doi:10.17713/ajs.v54i3.2077

Robust Filtering for Linear State-space Models with Non-propagating Outliers Following a Mixture of Gaussian Distributions

Bernhard Spangl
BOKU University, Vienna, Austria

Abstract

The problem of recursive filtering in linear state-space models is considered. The solution to this problem is the classical Kalman filter which is optimal in the sense that it minimizes the variance of the estimated states, if the error processes of the state and observation equations are both Gaussian. However, the Kalman filter is well known to be sensitive to outliers, so robustness is an issue. Two approximate conditional-mean (ACM) type filters for vector-valued observations are proposed that generalize existing univariate filters of similar type to the multivariate case. These new ACM-type filters are compared by simulations in a multivariate setting with additive outliers to the classical Kalman filter and the robust least squares (rLS) filter, another approach robustifying the Kalman filter. Additionally, different settings of tuning parameters and their impact are investigated. The results of the simulation experiments show that in the presence of additive outliers the multivariate ACM-type filters not only outperform the classical Kalman filter, as expected, but they also outperform the rLS filter.

Keywords: additive outliers, ACM-type filter, rLS filter, robust Kalman filtering, state-space models.

1. Introduction

In 2005 Fritz Leisch was invited to give a seminal talk on behalf of the R Core Team with the title 'R: The Language, Common Pitfalls & Underused Features' at the 1st International Workshop on Robust Statistics and R in Treviso, Italy. The development of the statistical software R, in which Fritz Leisch had a major role, enables statisticians in general, but also in the area of robust statistics, to implement theoretical ideas and to make them available to a broader user community. To highlight just one of the many developments in the area of robust statistics, a paper on robust fitting of mixtures by Neykov, Filzmoser, Dimova, and Neytchev (2007) builds on Fritz Leisch's seminal and most cited work on finite mixture models (Leisch 2004). This exchange of ideas was not just one way but went both ways as there exists a paper, too, in which Fritz proposes how to tackle the EM-estimation of mixtures of regression models in the presence of outliers (Leisch 2008). To his memory we dedicate this article.

In the following, let $\boldsymbol{y}_t$, $t = 1, \ldots, n$, denote an observed q-dimensional, vector-valued process which is assumed to be a linear transformation of an unobserved p-dimensional signal $\boldsymbol{x}_t$ with some noise added. Then the *state-space model* can be defined as follows:

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{w}_t \ , \\ \boldsymbol{y}_t &= \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{v}_t \ , \end{aligned} \tag{1}$$

where $\boldsymbol{x}_t$ is the unobserved p-dimensional vector called *state vector*. The first equation in (1) is named *state equation* and the second *observation equation*. It is assumed that $\boldsymbol{w}_t$ has dimension p, $\boldsymbol{\Phi}$ is a $p \times p$ matrix and $\boldsymbol{H}$ is a $q \times p$ matrix. Further, $\boldsymbol{x}_t$ is independent of future values of $\boldsymbol{w}_t$, and $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are zero mean independent and identically distributed (iid) sequences which are mutually independent but could be non-Gaussian. A more general definition of state-space models considering correlated errors as well as more complex models including exogenous variables or selection matrices can be found in Durbin and Koopman (2012) or Shumway and Stoffer (2025).

Generally, three types of outliers may be considered: innovation outliers (IO), additive outliers (AO), and substitutive outliers (SO). In our context, innovation outliers will change the error process $\boldsymbol{w}_t$ of the state equation. Hence, they are propagating as they affect all subsequent observations. Additive outliers will change the error process $\boldsymbol{v}_t$ of the observation equation whereas substitutive outliers will change the observation $\boldsymbol{y}_t$ directly. Hence, additive outliers and substitutive outliers are non-propagating as they have no effect on subsequent observations. Innovation outliers and additive outliers were first introduced by Fox (1972). Here we focus on a special case of non-Gaussian distributed $\boldsymbol{v}_t$, i.e., on additive outliers. The original additive outlier model (AO model) is a commonly used model for outliers in the case of univariate time series. It consists of a stationary core process to which occasional outliers have been added. For the methods presented here, it is convenient to model the distribution of $\boldsymbol{v}_t$ by a contaminated multivariate normal distribution

$$\mathcal{CN}_q(\gamma, \boldsymbol{\mu}, \boldsymbol{R}, \boldsymbol{\mu}^{(c)}, \boldsymbol{R}^{(c)}) = (1-\gamma)\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{R}) + \gamma\mathcal{N}_q(\boldsymbol{\mu}^{(c)}, \boldsymbol{R}^{(c)}) \ , \tag{2}$$

where γ is the amount of contamination. In addition, $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal distribution with d-dimensional mean vector $\boldsymbol{\mu}$ and $d \times d$ covariance matrix $\boldsymbol{\Sigma}$. The value of γ defines the percentage of additive outliers within the data. For natural reasons γ is expected to be less than 50%. Although additive outliers were originally introduced for univariate time series, we will still use this term because we refer to situations where the observation process $\boldsymbol{y}_t$ is disturbed by a contamination at certain times t.

Based on Masreliez's result (Masreliez 1975) an approximate conditional-mean (ACM) type filter for vector-valued observations is proposed which generalizes Martin's result (Martin 1979) to the multivariate case. By extensive simulation studies we show that the suggested new filter performs better in a multivariate setting with additive outliers than the robust least squares (rLS) filter, another approach robustifying the Kalman filter, proposed by Ruckdeschel (2001).

The outline of the paper is as follows: In the next section the new multivariate approximate conditional-mean type filters are described in detail, whereas the rLS filter proposed by Ruckdeschel (2001) is summarized in Section 3. In Section 4 the outline of the simulation study is given and the results are presented. In Section 5 further applications of the proposed filters are discussed and additional remarks are given. Finally, Section 6 summarizes and concludes the paper. In Appendix A the classical Kalman filter and smoother is briefly reviewed and the notation that is used in the paper is specified.

2. Approximate conditional-mean type filtering

The robust filter described in this section is an approximate conditional-mean (ACM) type filter motivated by Masreliez's result (Masreliez 1975).

2.1. Masreliez's theorem

If $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are Gaussian a straightforward calculation of $\boldsymbol{x}_{t|t} = E(\boldsymbol{x}_t|\boldsymbol{Y}_t)$ with $\boldsymbol{Y}_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ yields the Kalman filter recursion equations (see, for example, Jazwinski 1970). For non-Gaussian $\boldsymbol{v}_t$ the calculation of the exact $\boldsymbol{x}_{t|t}$ is difficult. However, Masreliez (1975) made the simplifying assumption that the state prediction density $f_{\boldsymbol{x}_t}(\cdot|\boldsymbol{Y}_{t-1})$ is Gaussian, i.e.,

$$\boldsymbol{x}_t|\boldsymbol{Y}_{t-1} \sim \mathcal{N}_p(\boldsymbol{x}_{t|t-1}, \boldsymbol{P}_{t|t-1}) \ .$$

Note that in the Gaussian case the prediction and filtering error covariance matrices $\boldsymbol{P}_{t|t-1}$ and $\boldsymbol{P}_{t|t}$ do not depend upon $\boldsymbol{Y}_{t-1}$ and $\boldsymbol{Y}_t$, respectively. (An exact definition of $\boldsymbol{P}_{t|t-1}$ and $\boldsymbol{P}_{t|t}$ is given in Appendix A.) However, one should not expect $\boldsymbol{P}_{t|t-1}$ and $\boldsymbol{P}_{t|t}$ to be independent of the data in general, and in fact it turns out that in Masreliez's result both, $\boldsymbol{P}_{t|t-1}$ and $\boldsymbol{P}_{t|t}$, depend upon the data in an intuitively appealing manner.

For the following ACM filter theorem it is assumed that the observations are generated by (1) and that $\boldsymbol{\Phi}$, the covariance matrix $\boldsymbol{Q}$ of the $\boldsymbol{w}_t$ and the density $f_{\boldsymbol{v}_t}$ of the $\boldsymbol{v}_t$ are known in advance.

Theorem 2.1. *(Masreliez 1975). If $\boldsymbol{x}_t|\boldsymbol{Y}_{t-1} \sim \mathcal{N}_p(\boldsymbol{x}_{t|t-1}, \boldsymbol{P}_{t|t-1})$, $t \geq 1$, then $\boldsymbol{x}_{t|t} = E(\boldsymbol{x}_t|\boldsymbol{Y}_t)$, $t \geq 1$, is generated by the recursions*

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top \Psi_t(\boldsymbol{y}_t) \ , \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top \Psi_t'(\boldsymbol{y}_t)\boldsymbol{H}\boldsymbol{P}_{t|t-1} \ , \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{\Phi}\boldsymbol{P}_{t|t}\boldsymbol{\Phi}^\top + \boldsymbol{Q} \ , \end{aligned} \tag{3}$$

where $\Psi_t(\boldsymbol{y}_t)$ is a q-dimensional vector with components

$$(\Psi_t(\boldsymbol{y}))_i = -(\partial/\partial y_i) \log f_{\boldsymbol{y}_t}(\boldsymbol{y}|\boldsymbol{Y}_{t-1})$$

and is usually called the score function *for the observation prediction density $f_{\boldsymbol{y}_t}(\cdot|\boldsymbol{Y}_{t-1})$, and $\Psi_t'(\boldsymbol{y}_t)$ is a $q \times q$ matrix with elements*

$$(\Psi_t'(\boldsymbol{y}))_{ij} = (\partial/\partial y_j)(\Psi_t(\boldsymbol{y}))_i \ .$$

If $f_{\boldsymbol{y}_t}(\cdot|\boldsymbol{Y}_{t-1})$ is Gaussian, it is easy to verify that Masreliez's filter reduces to the Kalman filter. In this case we have

$$\begin{aligned} \Psi_t(\boldsymbol{y}_t) &= (\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R})^{-1}(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}) \ , \\ \Psi_t'(\boldsymbol{y}_t) &= (\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R})^{-1} \ , \end{aligned}$$

where $\boldsymbol{R}$ is the covariance of the observation noise $\boldsymbol{v}_t$.

Although Masreliez (1975) did not specify initial conditions for the above recursions, appropriate $\boldsymbol{x}_{1|0}$ and $\boldsymbol{P}_{1|0}$ may be set to $\boldsymbol{x}_{1|0} = E(\boldsymbol{x}_1) = \boldsymbol{\mu}_1$ and $\boldsymbol{P}_{1|0} = \mathrm{Cov}(\boldsymbol{x}_1) = \boldsymbol{\Sigma}_1$, i.e., the unconditional mean and covariance of $\boldsymbol{x}_1$ (cf. Martin 1981). However, in order to agree with the definition of the classical Kalman filter recursions we will specify the initial conditions for the above recursions by setting $\boldsymbol{x}_{0|0} = E(\boldsymbol{x}_0) = \boldsymbol{\mu}_0$ and $\boldsymbol{P}_{0|0} = \mathrm{Cov}(\boldsymbol{x}_0) = \boldsymbol{\Sigma}_0$ (see, for example, Durbin and Koopman 2012; Shumway and Stoffer 2025). This will lead to slightly different results only for the first few $\boldsymbol{P}_{t|t-1}$'s and $\boldsymbol{P}_{t|t}$'s.

2.2. Masreliez's filter for state-space models with additive outliers

Assume in the following that the observation noise $\boldsymbol{v}_t$ is contaminated by additive outliers according to a contaminated multivariate normal distribution given in (2).

Proceeding from Masreliez's theorem (cf. Section 2.1) it is also still assumed for the state prediction density $f_{\boldsymbol{x}_t}(\cdot|\boldsymbol{Y}_{t-1})$ that $\boldsymbol{x}_t|\boldsymbol{Y}_{t-1} \sim \mathcal{N}_p(\boldsymbol{x}_{t|t-1}, \boldsymbol{P}_{t|t-1})$. Then the observation prediction density $f_{\boldsymbol{y}_t}(\cdot|\boldsymbol{Y}_{t-1})$ in Masreliez's theorem is obtained by convolving the prediction density $f_{\boldsymbol{z}_t}(\cdot|\boldsymbol{Y}_{t-1})$ of $\boldsymbol{z}_t|\boldsymbol{Y}_{t-1} \sim \mathcal{N}_q(\boldsymbol{H}\boldsymbol{x}_{t|t-1}, \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top)$ for $\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t$ with the observation noise density $f_{\boldsymbol{v}_t}$.

Now, if the $\boldsymbol{v}_t$'s have a contaminated multivariate normal distribution (2) with $\boldsymbol{\mu} = \boldsymbol{\mu}^{(c)} = \boldsymbol{0}$ convolution of $f_{\boldsymbol{z}_t}$ with $f_{\boldsymbol{v}_t}$ gives

$$\boldsymbol{y}_t|\boldsymbol{Y}_{t-1} \sim (1-\gamma)\mathcal{N}_q(\boldsymbol{H}\boldsymbol{x}_{t|t-1}, \boldsymbol{R}_t) + \gamma\mathcal{N}_q(\boldsymbol{H}\boldsymbol{x}_{t|t-1}, \boldsymbol{R}_t^{(c)}) \ ,$$

with $\boldsymbol{R}_t = \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R}$ and $\boldsymbol{R}_t^{(c)} = \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R}^{(c)}$. We can also write the observation prediction density $f_{\boldsymbol{y}_t}$ as

$$f_{\boldsymbol{y}_t}(\cdot|\boldsymbol{Y}_{t-1}) = g_t(. - \boldsymbol{H}\boldsymbol{x}_{t|t-1}) \ ,$$

where g_t is obtained by convolution, i.e., it is the density function of the distribution

$$\mathcal{N}_q(\boldsymbol{0}, \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top) * \mathcal{F}_{\boldsymbol{v}_t} \ ,$$

and $\mathcal{F}_{\boldsymbol{v}_t}$ denotes the distribution of the observation noise $\boldsymbol{v}_t$. To ease notation we will further on set $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}$.

Thus, Masreliez's filter recursions (3) become

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\Psi_t(\boldsymbol{\varepsilon}_t) \ , \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\Psi_t'(\boldsymbol{\varepsilon}_t)\boldsymbol{H}\boldsymbol{P}_{t|t-1} \ , \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{\Phi}\boldsymbol{P}_{t|t}\boldsymbol{\Phi}^\top + \boldsymbol{Q} \ , \end{aligned} \tag{4}$$

where $\Psi_t(\boldsymbol{\varepsilon}_t)$ is a q-dimensional vector with components

$$(\Psi_t(\boldsymbol{y}))_i = -(\partial/\partial y_i)\log g_t(\boldsymbol{y})$$

and $\Psi_t'(\boldsymbol{\varepsilon}_t)$ is a $q \times q$ matrix with elements

$$(\Psi_t'(\boldsymbol{y}))_{ij} = (\partial/\partial y_j)(\Psi_t(\boldsymbol{y}))_i \ .$$

As an illustrative example we plot the score function of a contaminated bivariate normal distribution at time t for two combinations of values of γ, $\boldsymbol{R}_t$ and $\boldsymbol{R}_t^{(c)}$:

Situation A

$$\gamma = 0.1 \ , \quad \boldsymbol{R}_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \quad \boldsymbol{R}_t^{(c)} = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix} \ .$$

Situation B

$$\gamma = 0.1 \ , \quad \boldsymbol{R}_t = \begin{pmatrix} 2 & -0.2 \\ -0.2 & 2 \end{pmatrix} \ , \quad \boldsymbol{R}_t^{(c)} = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix} \ .$$

The resulting score functions are shown in Figures 1 and 2. In all four sub-plots the x-y plane is the domain of either the contaminated bivariate normal density function or its score function. In plot (a) the graph of the contaminated bivariate normal density function is displayed. Plot (b) shows the logarithm of the contaminated bivariate normal density function as contour plot, where the ellipses connect points of the same value, together with the corresponding score function, which is in the case of $q = 2$ a two-dimensional vector field of gradients. In the lower row of Figures 1 and 2 the graphs of the first and second coordinate

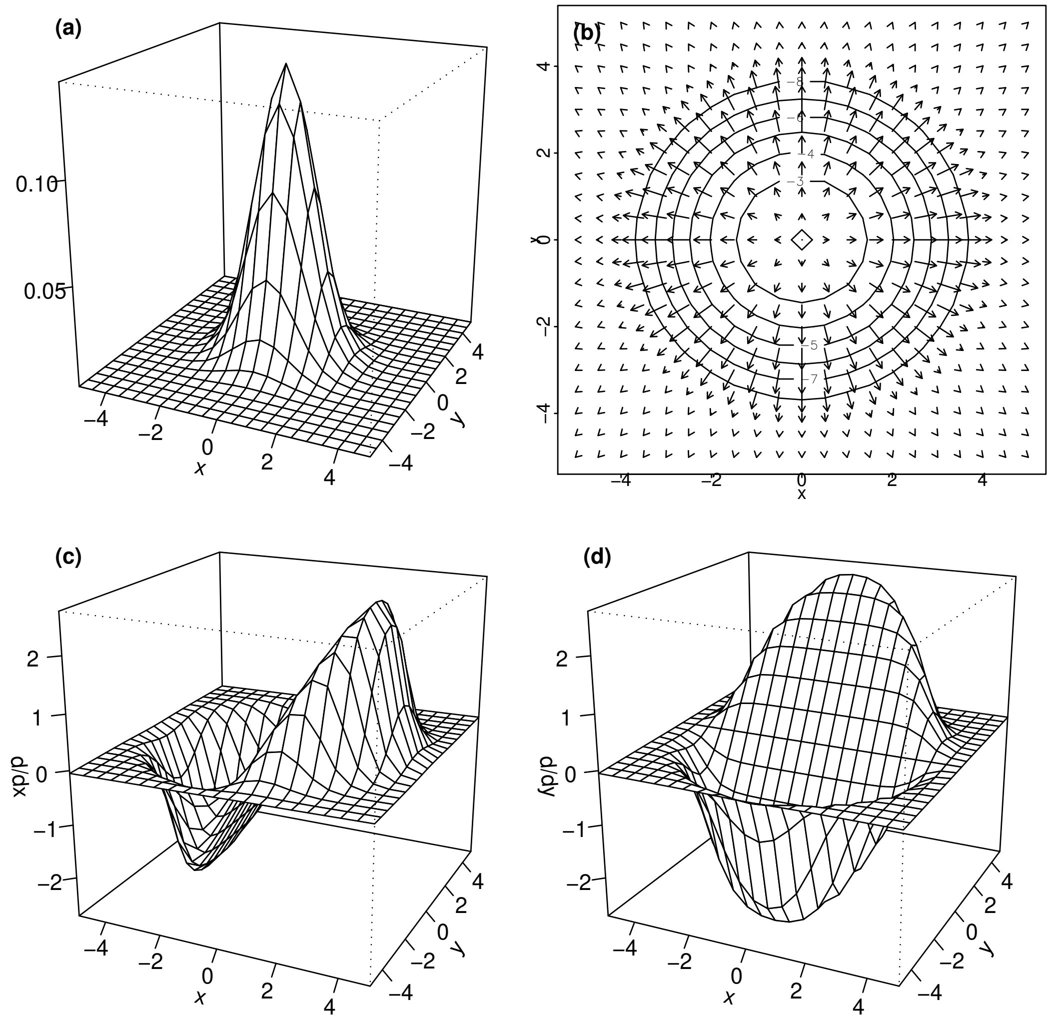

Figure 1: The contaminated bivariate normal distribution (a) and the corresponding score function of Situation A as two-dimensional vector field of gradients (b) as well as its first (c) and second (d) coordinate plotted separately.

of the corresponding score function are shown. The first coordinate is plotted on the vertical axis in plot (c) and the second one in plot (d).

Now the density function g_t can be represented as

$$g_t(\boldsymbol{\varepsilon}_t) = |\boldsymbol{S}_t| g(\boldsymbol{S}_t \boldsymbol{\varepsilon}_t) \ , \tag{5}$$

where g is the density function of the distribution

$$\mathcal{N}_q(\mathbf{0}, \boldsymbol{A}_t) * \mathcal{F}_{\boldsymbol{v}_t, \boldsymbol{B}} \ ,$$

with

$$\mathcal{F}_{\boldsymbol{v}_t, \boldsymbol{B}}(\boldsymbol{u}) = \mathcal{F}_{\boldsymbol{v}_t}(\boldsymbol{B}_t \boldsymbol{u}) \ ,$$

and $\boldsymbol{S}_t$, $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ are appropriately specified.

It is easily proven that (5) is valid if $\mathcal{F}_{\boldsymbol{v}_t}$ is Gaussian or a mixture of Gaussian distributions with component-specific mean vectors $\mathbf{0}$ and arbitrary covariance matrices. As in the univariate case, this is generally not true if $\mathcal{F}_{\boldsymbol{v}_t}$ is non-Gaussian.

Hence, in the case of $\mathcal{F}_{\boldsymbol{v}_t}$ being a contaminated multivariate normal distribution (2) with $\boldsymbol{\mu} = \boldsymbol{\mu}^{(c)} = \mathbf{0}$, if we set

$$\boldsymbol{S}_t = \boldsymbol{V}_t \boldsymbol{D}_t^{-1/2} \boldsymbol{V}_t^{\top}$$

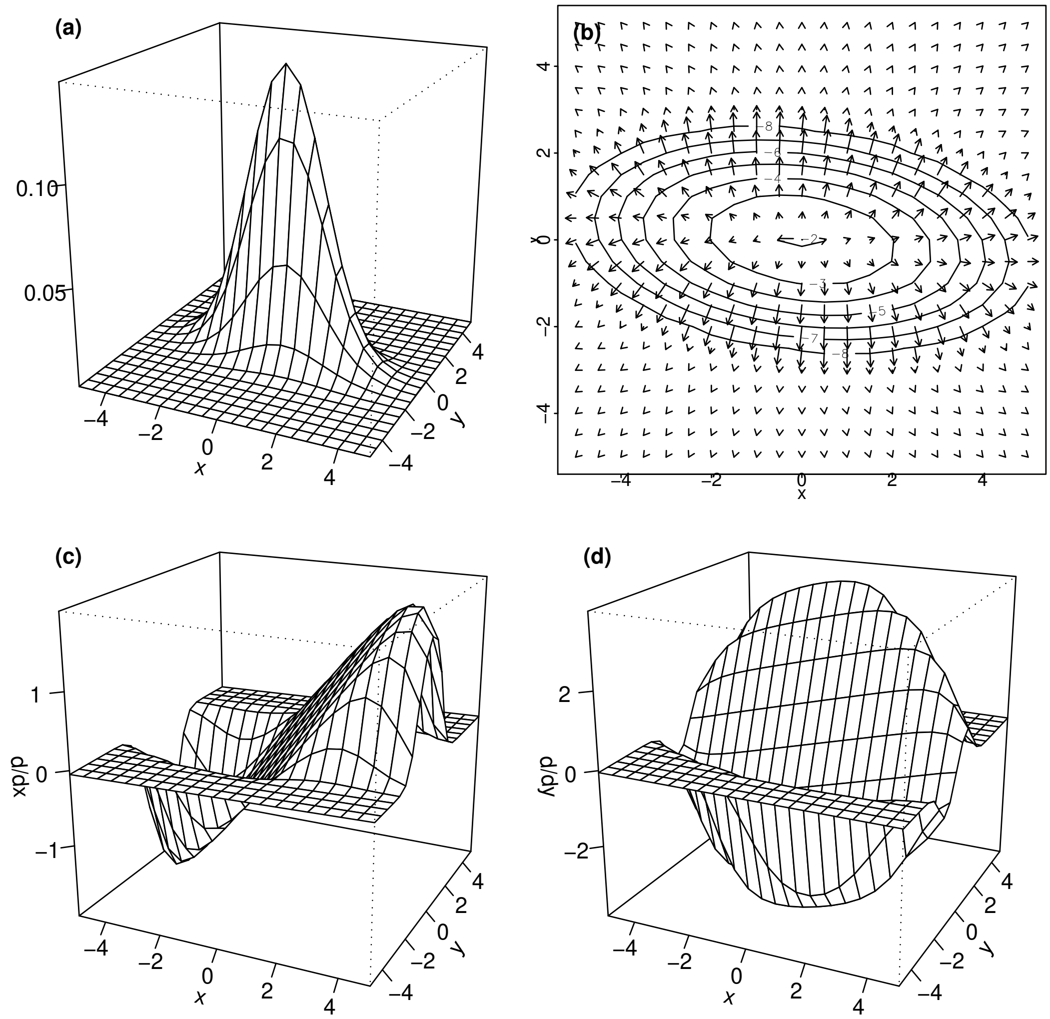

Figure 2: The contaminated bivariate normal distribution (a) and the corresponding score function of Situation B as two-dimensional vector field of gradients (b) as well as its first (c) and second (d) coordinate plotted separately.

with $\boldsymbol{R}_t = \boldsymbol{V}_t \boldsymbol{D}_t \boldsymbol{V}_t^\top = \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R}$, and further,

$$\boldsymbol{A}_t = \boldsymbol{S}_t \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top \boldsymbol{S}_t \ , \quad \text{and} \quad \boldsymbol{B}_t = \boldsymbol{S}_t^{-1} \ ,$$

we get

$$\begin{aligned} g(\boldsymbol{u}) &= \varphi_q(\boldsymbol{u};\boldsymbol{0},\boldsymbol{A}_t) * \Big[(1-\gamma)|\boldsymbol{B}_t|\varphi_q(\boldsymbol{B}_t\boldsymbol{u};\boldsymbol{0},\boldsymbol{R}) + \gamma|\boldsymbol{B}_t|\varphi_q(\boldsymbol{B}_t\boldsymbol{u};\boldsymbol{0},\boldsymbol{R}^{(c)})\Big] \\ &= (1-\gamma)\Big[\varphi_q(\boldsymbol{u};\boldsymbol{0},\boldsymbol{A}_t) * \varphi_q(\boldsymbol{u};\boldsymbol{0},(\boldsymbol{B}_t^{-1})^\top \boldsymbol{R}\boldsymbol{B}_t^{-1})\Big] + \\ &\qquad\qquad \gamma\Big[\varphi_q(\boldsymbol{u};\boldsymbol{0},\boldsymbol{A}_t) * \varphi_q(\boldsymbol{u};\boldsymbol{0},(\boldsymbol{B}_t^{-1})^\top \boldsymbol{R}^{(c)}\boldsymbol{B}_t^{-1})\Big] \\ &= (1-\gamma)\varphi_q(\boldsymbol{u};\boldsymbol{0},\boldsymbol{I}_q) + \gamma\varphi_q(\boldsymbol{u};\boldsymbol{0},\boldsymbol{S}_t\boldsymbol{R}_t^{(c)}\boldsymbol{S}_t) \ , \end{aligned}$$

where $\varphi_d(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the density function of the d-dimensional multivariate normal distribution $\mathcal{N}_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$. To emphasize the fact that the $(1-\gamma)$-weighted part of the density function g is the density of the q-dimensional standard normal distribution $\mathcal{N}_q(\boldsymbol{0},\boldsymbol{I}_q)$ we drop the subscript t. The matrix $\boldsymbol{D}_t$ denotes the diagonal matrix containing the eigenvalues of $\boldsymbol{R}_t$ and $\boldsymbol{V}_t$ is the matrix of the corresponding eigenvectors.

If $\gamma = 0$, i.e., in the pure Gaussian situation, the idea is simply to standardize $\boldsymbol{\varepsilon}_t$ by $\boldsymbol{S}_t$. In cases where only few outliers are expected g_t may be still approximated by a multivariate

Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{R}_t = \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R}$ as before.

Note that $\boldsymbol{S}_t$ is simply the square root of the inverse of $\boldsymbol{R}_t$, i.e., $\boldsymbol{S}_t\boldsymbol{S}_t = \boldsymbol{R}_t^{-1}$. Note further that $\boldsymbol{S}_t = \boldsymbol{V}_t\boldsymbol{D}_t^{-1/2}\boldsymbol{V}_t^\top$ is symmetric and is just a distortion of the q-dimensional space without any rotation.

Based on the same considerations as before an approximation of the score function Ψ_t can be obtained:

$$\Psi_t(\boldsymbol{\varepsilon}_t) \cong \boldsymbol{S}_t\Psi(\boldsymbol{S}_t\boldsymbol{\varepsilon}_t) \ , \tag{6}$$

and

$$\Psi_t'(\boldsymbol{\varepsilon}_t) \cong \boldsymbol{S}_t\Psi'(\boldsymbol{S}_t\boldsymbol{\varepsilon}_t)\boldsymbol{S}_t \ ,$$

where $\Psi(\boldsymbol{S}_t\boldsymbol{\varepsilon}_t)$ is a q-dimensional vector with components

$$(\Psi(\boldsymbol{y}))_i = -(\partial/\partial y_i)\log g(\boldsymbol{y})$$

and $\Psi'(\boldsymbol{S}_t\boldsymbol{\varepsilon}_t)$ is a $q \times q$ matrix with elements

$$(\Psi'(\boldsymbol{y}))_{ij} = (\partial/\partial y_j)(\Psi(\boldsymbol{y}))_i \ .$$

This yields simplified versions of (4):

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\boldsymbol{S}_t\Psi(\boldsymbol{S}_t\boldsymbol{\varepsilon}_t) \ , \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\boldsymbol{S}_t\Psi'(\boldsymbol{S}_t\boldsymbol{\varepsilon}_t)\boldsymbol{S}_t\boldsymbol{H}\boldsymbol{P}_{t|t-1} \ , \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{\Phi}\boldsymbol{P}_{t|t}\boldsymbol{\Phi}^\top + \boldsymbol{Q} \ . \end{aligned} \tag{7}$$

2.3. ACM-type filters for state-space models with additive outliers

Since $\mathcal{F}_{\boldsymbol{v}_t}$ will rarely be known in practice, we follow on the same lines as proposed by Martin (1979). Therein he suggests how to define an approximate conditional-mean (ACM) type filter for autoregressive models of order p. We now generalize his results and define a multivariate approximate conditional-mean (ACM) type filter for state-space models with vector-valued observations.

Like for univariate M-estimators the score function Ψ is replaced by a good robustifying psi-function ψ, which is bounded and continuous and leaves "small" vectors unchanged. Again, let

$$(\boldsymbol{S}_t\boldsymbol{S}_t)^{-1} = \boldsymbol{R}_t = \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R} \ ,$$

where $\boldsymbol{R}$ is the covariance matrix of the uncontaminated multivariate Gaussian distribution of the observation noise $\boldsymbol{v}_t$. Then, noting that $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}$, the recursions (7) are

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\boldsymbol{S}_t\psi(\boldsymbol{S}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1})) \ , \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\boldsymbol{S}_t\psi'(\boldsymbol{S}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}))\boldsymbol{S}_t\boldsymbol{H}\boldsymbol{P}_{t|t-1} \ , \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{\Phi}\boldsymbol{P}_{t|t}\boldsymbol{\Phi}^\top + \boldsymbol{Q} \ , \end{aligned} \tag{8}$$

with $\boldsymbol{x}_{t|t-1} = \boldsymbol{\Phi}\boldsymbol{x}_{t-1|t-1}$. Further, $\psi(\boldsymbol{S}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}))$ is a q-dimensional vector and $\psi'(\boldsymbol{S}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}))$ denotes a $q \times q$ matrix with elements

$$(\psi'(\boldsymbol{y}))_{ij} = (\partial/\partial y_j)(\psi(\boldsymbol{y}))_i \ .$$

In the univariate setting psi-functions are defined in a way that they leave small values unchanged whereas large values will be bounded or even set to zero as for redescending psi-functions. One way to extend this approach to the multivariate case is to apply it to the

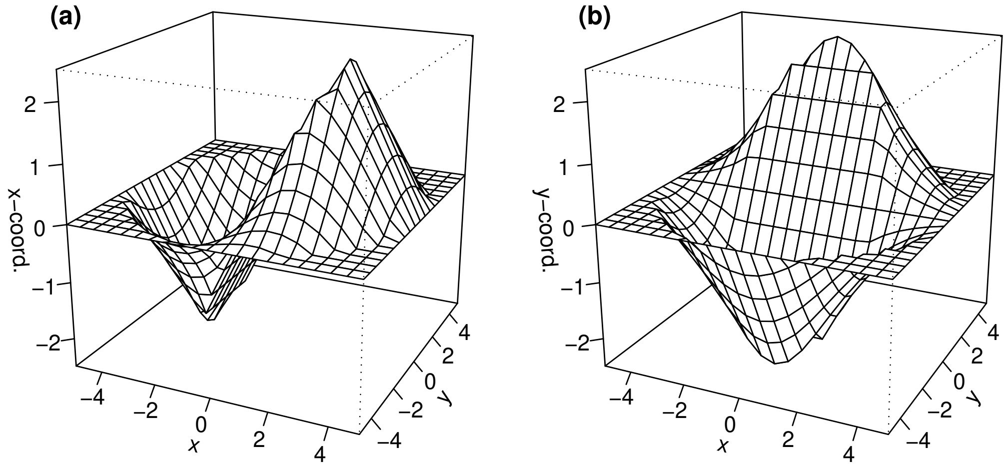

Figure 3: First (a) and second (b) coordinate of the approximated score function of the contaminated bivariate normal distribution of Situation A.

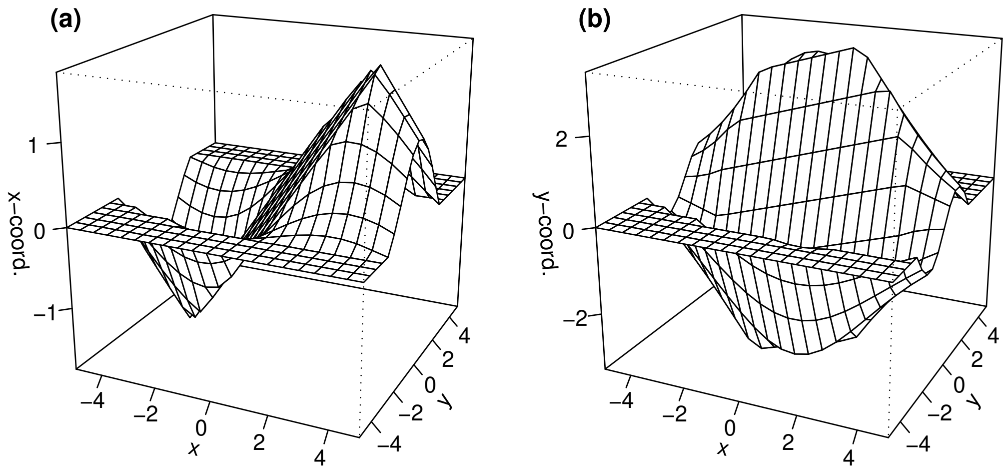

Figure 4: First (a) and second (b) coordinate of the approximated score function of the contaminated bivariate normal distribution of Situation B.

length of a vector, i.e., the direction of the vector remains unchanged and only its length is altered. We therefore suggest to define a multivariate analogue of Hampel's three-part redescending psi-function in the following way:

$$\psi_{HA}^{(m)}(\boldsymbol{s}) = \begin{cases} \boldsymbol{s} & \text{if} \quad \|\boldsymbol{s}\| \leq a \\ \frac{a}{\|\boldsymbol{s}\|}\boldsymbol{s} & \text{if} \quad a < \|\boldsymbol{s}\| \leq b \\ \frac{a}{c-b}(c - \|\boldsymbol{s}\|)\frac{\boldsymbol{s}}{\|\boldsymbol{s}\|} & \text{if} \quad b < \|\boldsymbol{s}\| \leq c \\ \boldsymbol{0} & \text{if} \quad c < \|\boldsymbol{s}\| \; . \end{cases} \tag{9}$$

Furthermore, we propose to use the multivariate analogue of Hampel's two-part redescending psi-function where we set $a = b$ in the robustifying psi-function ψ of the above filter recursions. Figures 3 and 4 show the approximated score functions of Figures 1 and 2 using the approximation (6) and the multivariate analogue of Hampel's two-part redescending psi-function with parameters $a = b = 2.5$ and $c = 5.0$. Plot (a) shows the graph of the first coordinate of the approximation of the score function originally displayed in plot (c), whereas plot (b) corresponds to the original plot (d).

Hence, the ACM filter for state-space models based on the multivariate analogue of Hampel's two-part redescending psi-function has the appealing feature that $\boldsymbol{x}_{t|t} = \boldsymbol{\Phi}\boldsymbol{x}_{t-1|t-1}$ and $\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1}$ by virtue of $\psi^{(m)}_{HA}(\boldsymbol{s}) = \psi'^{(m)}_{HA}(\boldsymbol{s}) = \boldsymbol{0}$ if $\|\boldsymbol{s}\| > c$. This characteristic is exactly as expected from an outlier-rejection rule in the filtering context.

2.4. An alternative ACM-type filter for state-space models with additive outliers

We note that the weighting in the correction step of the univariate ACM-type filter as well as of the multivariate one is a discontinuous function if using Hampel's redescending psi-function ψ_{HA} or its multivariate analogue $\psi^{(m)}_{HA}$. For the univariate case we have explicitly implemented the first derivative of Hampel's psi-function, whereas for the multivariate case the calculation of the Jacobian matrix is accomplished via numerical differentiation.

To illustrate why this can be problematic, we give an example in the following. Let $\Delta\boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1} = (-3.33, 3.73)^\top$ and compute the first derivative of Hampel's multivariate psi-function $\psi^{(m)}_{HA}$ with $a = b = 2.5$ and $c = 5.0$ using numerical differentiation. The resulting Jacobian matrix is given as

$$(\psi^{(m)'}_{HA}(\Delta\boldsymbol{y}_t))_{ij} = (\partial/\partial y_j)(\psi^{(m)}_{HA}(\Delta\boldsymbol{y}_t))_i = \begin{pmatrix} 1.473131 \times 10^{-05} & -3.093866 \times 10^{-05} \\ -1.650248 \times 10^{-05} & 3.465155 \times 10^{-05} \end{pmatrix} ,$$

although it should be equal to $\boldsymbol{0}_{2\times 2}$ as $\|\Delta\boldsymbol{y}_t\| = \|(-3.33, 3.73)^\top\| = 5.00018 > c$. Moreover, note that the Jacobian matrix is not symmetric anymore which yields non-symmetric prediction and filtering error covariance matrices.

To avoid discontinuities in the univariate case, the first derivative of the psi-function may be replaced by a continuous weight function

$$w(r) = \psi_{HA}(r)/r \tag{10}$$

as already proposed by Martin and Thomson (1982).

To avoid discontinuities in the case of vector-valued observations Gandhi and Mili (2010) proposed to apply psi-functions as well as weight functions coordinate-wise which will bound the effect of the residuals but will change their direction in space.

An alternative approach may be to reformulate the multivariate analogue of Hampel's redescending psi-function given in (9) as follows:

$$\psi^{(m)}_{HA}(\boldsymbol{s}) = \psi^{*}_{HA}(\|\boldsymbol{s}\|)\,\boldsymbol{s} \quad , \text{ where}$$

$$\psi^{*}_{HA}(r) = \begin{cases} 1 & \text{if} \quad r \le a \\ \frac{a}{r} & \text{if} \quad a < r \le b \\ \frac{a}{c-b}(c-r)\frac{1}{r} & \text{if} \quad b < r \le c \\ 0 & \text{if} \quad c < r \ . \end{cases}$$

Note that in $\mathbb{R}^+$ the function $\psi^{*}_{HA}(r)$ is equal to the continuous weight function $w(r)$ given in (10). Now, to ease notation we will set $\Delta\boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}$ and propose a new correction step replacing the recursions (8) by

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\boldsymbol{R}_t^{-1}\psi^{*}_{HA}((\Delta\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}\Delta\boldsymbol{y}_t)^{1/2})\Delta\boldsymbol{y}_t \ , \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top\boldsymbol{R}_t^{-1}\psi^{*}_{HA}((\Delta\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}\Delta\boldsymbol{y}_t)^{1/2})\boldsymbol{H}\boldsymbol{P}_{t|t-1} \ , \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{\Phi}\boldsymbol{P}_{t|t}\boldsymbol{\Phi}^\top + \boldsymbol{Q} \ . \end{aligned}$$

We will refer to this variant as alternative ACM-type filter (ACM2) in the remainder of the manuscript. Note that $\|\boldsymbol{S}_t\Delta\boldsymbol{y}_t\| = (\Delta\boldsymbol{y}_t^\top\boldsymbol{S}_t\boldsymbol{S}_t\Delta\boldsymbol{y}_t)^{1/2} = (\Delta\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}\Delta\boldsymbol{y}_t)^{1/2}$ is the Mahalanobis norm of $\Delta\boldsymbol{y}_t$ with respect to the multivariate Gaussian distribution $\mathcal{N}_q(\boldsymbol{0}, \boldsymbol{R}_t)$. This avoids the calculation of $\boldsymbol{S}_t$, the square root of the inverse of $\boldsymbol{R}_t$, too. A similar approach was used by Boudt and Croux (2010) to bound the effect of returns in a multivariate GARCH model.

3. The robust least squares (rLS) filter algorithm

Ruckdeschel (2001) proposed an alternative robustified version of the Kalman filter which is briefly reviewed in the following. For details the reader is referred to Ruckdeschel (2000, 2001).

The idea is simply to reduce in the correction step (A.3) of the classical Kalman filter the influence of an observation $\boldsymbol{y}_t$ that is affected by an additive outlier. Instead of $\boldsymbol{K}_t\boldsymbol{\varepsilon}_t$ with $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}$ a Huberized version of it is used, i.e.,

$$H_b(\boldsymbol{K}_t\boldsymbol{\varepsilon}_t) = \boldsymbol{K}_t\boldsymbol{\varepsilon}_t \min\{1, \frac{b}{\|\boldsymbol{K}_t\boldsymbol{\varepsilon}_t\|}\} ,$$

so that the obtained result will be equal to the one of the classical Kalman filter, if $\|\boldsymbol{K}_t\boldsymbol{\varepsilon}_t\|$ is not larger than b. If on the other hand $\|\boldsymbol{K}_t\boldsymbol{\varepsilon}_t\|$ exceeds the value of b, the direction of $\boldsymbol{K}_t\boldsymbol{\varepsilon}_t$ will remain unchanged but it will be projected on the q-dimensional ball with radius b, i.e., its influence is bounded by the value of b.

Hence, these modifications almost yield the classical Kalman filter recursions with the only exception of replacing the first line of the correction step in (A.3) by

$$\boldsymbol{x}_{t|t} \;=\; \boldsymbol{x}_{t|t-1} + H_b(\boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1})) .$$

Ruckdeschel (2001) proved that the rLS filter is SO-optimal under certain side conditions. The term SO stands for substitutive outlier and means that, instead of disturbing $\boldsymbol{v}_t$, contamination affects $\boldsymbol{y}_t$ directly, replacing it by an arbitrarily distributed variable $\boldsymbol{y}'_t$ with some low probability.

Note that the calibration, i.e., finding b to a given loss of efficiency, has to be done beforehand.

4. Simulation study

To test the performance of the multivariate ACM-type filters three different state-space models with different additive outlier situations and different amounts of contamination were simulated and the results were compared to those of the rLS filter.

Example 4.1. For the first state-space model the following hyperparameters were used:

$$\boldsymbol{\mu}_0 = (0,0)^\top , \qquad \boldsymbol{\Sigma}_0 = \boldsymbol{0}_{2\times 2} ,$$
$$\boldsymbol{\Phi} = \begin{pmatrix} 0.5 & 0.3 \\ 0.6 & 0.5 \end{pmatrix} , \quad \boldsymbol{Q} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} ,$$
$$\boldsymbol{H} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} , \quad \boldsymbol{R} = \begin{pmatrix} 2 & -0.2 \\ -0.2 & 0.5 \end{pmatrix} ,$$

where $\boldsymbol{0}_{p\times p}$ denotes the $p\times p$ zero matrix.

The $\boldsymbol{v}_t$'s of the observation process were simulated from a contaminated bivariate normal distribution (2) defined by

$$(1-\gamma)\mathcal{N}_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{R}) + \gamma\mathcal{N}_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}) .$$

Example 4.2. For the second state-space model the following hyperparameters were used:

$$\boldsymbol{\mu}_0 = (20, 0)^\top \,, \qquad \boldsymbol{\Sigma}_0 = \mathbf{0}_{2\times 2} \,,$$
$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \,, \qquad \boldsymbol{Q} = \begin{pmatrix} 0 & 0 \\ 0 & 9 \end{pmatrix} \,,$$
$$\boldsymbol{H} = \begin{pmatrix} 0.3 & 1 \\ -0.3 & 1 \end{pmatrix} \,, \qquad \boldsymbol{R} = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix} \,.$$

Note that the first coordinate of the above state process is a random walk and therefore non-stationary, whereas the second coordinate is just white noise.

Here the $\boldsymbol{v}_t$'s were simulated from a contaminated bivariate normal distribution with a contaminating distribution given by

$$(1-\gamma)\mathcal{N}_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{R}) + \gamma\mathcal{N}_2(\begin{pmatrix} 25 \\ 30 \end{pmatrix}, \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}) \,.$$

Note that the mean of the contaminating normal distribution, $\boldsymbol{\mu}^{(c)}$, is now unequal to zero. In the original work by Martin (1979) it was assumed that $\boldsymbol{\mu}^{(c)}$ is equal to zero, this assumption can be further relaxed.

Example 4.3. The third state-space model is taken from Becker (2023). A six-dimensional Kalman filter is designed to estimate the vehicle's location in the x-y-plane. We further assume constant acceleration dynamics. The system state $\boldsymbol{x}_t$ is defined by:

$$\boldsymbol{x}_t = (x_t, \dot{x}_t, \ddot{x}_t, y_t, \dot{y}_t, \ddot{y}_t)^\top \,.$$

The vector $\boldsymbol{y}_t$ contains the observed x and y coordinates of the vehicle at time t, $t = 1, \ldots, 35$. The following hyperparameters were used:

$$\boldsymbol{\mu}_0 = (0,0,0,0,0,0)^\top \,, \quad \boldsymbol{\Sigma}_0 = \mathbf{0}_{6\times 6} \,,$$
$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & \Delta t & 0.5\Delta t^2 & 0 & 0 & 0 \\ 0 & 1 & \Delta t & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \Delta t & 0.5\Delta t^2 \\ 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0.5 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0.5 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \,,$$
$$\boldsymbol{Q} = \begin{pmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & \frac{\Delta t^2}{2} & 0 & 0 & 0 \\ \frac{\Delta t^3}{2} & \Delta t^2 & \Delta t & 0 & 0 & 0 \\ \frac{\Delta t^2}{2} & \Delta t & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & \frac{\Delta t^2}{2} \\ 0 & 0 & 0 & \frac{\Delta t^3}{2} & \Delta t^2 & \Delta t \\ 0 & 0 & 0 & \frac{\Delta t^2}{2} & \Delta t & 1 \end{pmatrix} \sigma_a^2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 1 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 1 & 1 \end{pmatrix} 0.2^2 \,,$$
$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \,, \quad \boldsymbol{R} = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix} \,,$$

where $\Delta t = 1s$ is the time between successive measurements and σ_a^2 is a random variance in acceleration.

Here the $\boldsymbol{v}_t$'s were simulated from a contaminated bivariate normal distribution with a contaminating distribution given by

$$(1-\gamma)\mathcal{N}_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{R}) + \gamma\mathcal{N}_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}) \,.$$

For all state-space models the contamination γ was varied from 0% to 20% by steps of 5% and 400 realizations for each level of contamination were simulated.

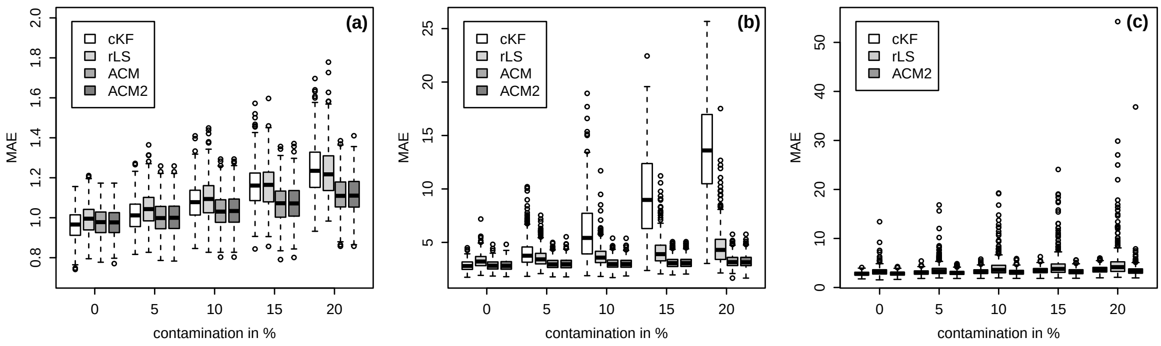

Figure 5: Boxplots of the median absolute errors of Example 4.1 (a), Example 4.2 (b), and Example 4.3 (c). The rLS filter (rLS) was calibrated to a given loss of efficiency $\delta = 0.1$ for all examples with a clipping value $b = 3.7$ (Examples 4.1 and 4.3) and $b = 2.4$ (Example 4.2), respectively. For the ACM-type filters (ACM, ACM2) the tuning parameters of the multivariate analogue of Hampel's three-part redescending psi-function were set to $a = b = 2.5$ and $c = 5.0$ for all examples.

Then the multivariate state processes of each simulated realization of all state-space models were computed using the multivariate ACM-type filters and the rLS filter proposed in Section 2 and by Ruckdeschel (2001), respectively. For the ACM-type filters the multivariate analogue of Hampel's three-part redescending psi-function with different combinations of the parameters a, b, and c was used for all examples. We set the tuning parameters $(a, b, c)^\top$ equal to $(2.5, 2.5, 5.0)^\top$, $(2.6, 2.6, 3.6)^\top$, and $(3.0, 3.0, 7.0)^\top$, respectively. These combinations have already been used in the literature for Martin's original ACM-type filter which is limited to the univariate case (Martin 1979). Additional details may be found in Martin (1979), Martin and Thomson (1982), and Insightful (2001). The rLS filter was calibrated to a given loss of efficiency $\delta = 0.1$ for all models and the clipping value b was set to $b = 3.7$ in Example 4.1 and Example 4.3 and to $b = 2.4$ in Example 4.2. As we do not know the amount of contamination in practical applications, setting $\delta = 0.1$ seems to be a reasonable trade-off between protecting against outliers and still achieving a relatively high efficiency. Moreover, also the classical Kalman filter estimates were calculated.

Finally, the median absolute error, denoted by MAE, between the true state process and the different filter estimates was computed, i.e.,

$$MAE_{\widehat{\boldsymbol{x}}_{t|t}} = \text{median}\|\boldsymbol{x}_t - \widehat{\boldsymbol{x}}_{t|t}\| \quad ,$$

where $\boldsymbol{x}_t$ and $\widehat{\boldsymbol{x}}_{t|t}$ denote the true state vector and the filter estimate, respectively. Additionally, we compute the standard error of the median. It is well known that the distribution of the median of a sample of size n, from a population having the density function $f(z)$, is asymptotically normal distributed with mean m and variance

$$\frac{1}{4nf(m)^2} \,,$$

where m is the median of the population (see, for example, Cramér 1946, p. 369). A kernel density estimator was used to estimate $f(z)$.

Regarding the computation time the rLS filter and the alternative ACM-type filter (ACM2) perform slightly better than the ACM-type filter (ACM) once the clipping value b is fixed. This is due to the fact that additional computations of the Jacobian matrix of Hampel's multivariate psi-function have to be done within the correction step of the ACM-type filter.

In Figure 5 the results of the simulation study are presented. For the first simulation experiment (Example 4.1) plot (a) displays the median absolute errors between the true state vector

Table 1: Median MAE and corresponding standard error of the median (given in parentheses) of Example 4.1

		0% (s.e.)	5% (s.e.)	10% (s.e.)	15% (s.e.)	20% (s.e.)
cKF		.97 (.01)	1.01 (.01)	1.08 (.01)	1.16 (.01)	1.23 (.01)
rLS	$b = 3.7$	1.00 (.00)	1.04 (.01)	1.09 (.01)	1.16 (.01)	1.22 (.01)
ACM	$a = b = 2.5,\ c = 5.0$	.98 (.00)	1.00 (.01)	1.03 (.01)	1.07 (.01)	1.11 (.01)
ACM	$a = b = 2.6,\ c = 3.6$	.98 (.00)	1.00 (.01)	1.04 (.01)	1.07 (.01)	1.11 (.01)
ACM	$a = b = 3.0,\ c = 7.0$	.97 (.01)	**.99 (.01)**	**1.03 (.01)**	**1.07 (.01)**	**1.11 (.01)**
ACM2	$a = b = 2.5,\ c = 5.0$	.98 (.00)	1.00 (.01)	1.03 (.01)	1.07 (.01)	1.11 (.01)
ACM2	$a = b = 2.6,\ c = 3.6$	.98 (.01)	1.00 (.01)	1.04 (.01)	1.08 (.01)	1.12 (.01)
ACM2	$a = b = 3.0,\ c = 7.0$	**.97 (.01)**	.99 (.01)	1.03 (.00)	1.07 (.01)	1.11 (.01)

Table 2: Median MAE and corresponding standard error of the median (given in parentheses) of Example 4.2

		0% (s.e.)	5% (s.e.)	10% (s.e.)	15% (s.e.)	20% (s.e.)
cKF		2.80 (.03)	3.76 (.07)	5.42 (.18)	8.97 (.29)	13.60 (.32)
rLS	$b = 2.4$	3.21 (.04)	3.42 (.05)	3.58 (.05)	3.90 (.06)	4.30 (.09)
ACM	$a = b = 2.5,\ c = 5.0$	2.82 (.03)	2.95 (.03)	2.97 (.03)	3.06 (.03)	3.16 (.04)
ACM	$a = b = 2.6,\ c = 3.6$	2.82 (.03)	2.97 (.03)	3.02 (.03)	3.08 (.03)	3.19 (.04)
ACM	$a = b = 3.0,\ c = 7.0$	2.79 (.03)	2.95 (.03)	**2.94 (.03)**	3.05 (.03)	3.19 (.04)
ACM2	$a = b = 2.5,\ c = 5.0$	2.81 (.03)	2.96 (.03)	2.97 (.03)	3.06 (.03)	**3.15 (.04)**
ACM2	$a = b = 2.6,\ c = 3.6$	2.82 (.03)	2.97 (.03)	3.01 (.03)	3.09 (.03)	3.17 (.04)
ACM2	$a = b = 3.0,\ c = 7.0$	**2.78 (.03)**	**2.93 (.03)**	2.96 (.03)	**3.05 (.03)**	3.17 (.04)

and each filter estimate for all levels of contamination. Similarly, the median absolute errors for the second and third state-space model (Examples 4.2 and 4.3) are seen in plot (b) and (c), respectively. In general, the MAE increases as the amount of contamination increases. This is especially visible in plots (a) and (b) of Figure 5. As expected, this difference is largest for the classical Kalman filter estimates. However, it is also larger for the filter estimates obtained by the rLS filter than for those of the ACM-type filters.

Note that we omitted the results of the ACM-type filter (ACM) of Example 4.3 in plot (c) of Figure 5 as for some simulated processes the discontinuity of the multivariate analogue of Hampel's psi-function yields non-symmetric prediction and filtering error covariance matrices. This seems to be an issue especially in higher dimensions.

Moreover, because of the fact that the rLS filter was calibrated to a given loss of efficiency $\delta = 0.1$ it yields larger errors in the case of no contamination compared to the other methods. This has already been noted by Ruckdeschel (2001) and is especially visible in plot (b) of Figure 5. Furthermore, it can be seen that according to the median absolute errors the ACM-type filter performs better than the rLS filter for all contamination levels.

Additionally to Figure 5 the median MAE of Examples 4.1, 4.2 and 4.3 are listed in Tables 1, 2 and 3, respectively. For each state-space model, all filters, and each level of contamination the median MAE together with their corresponding standard errors (given in parentheses) were calculated and the best method for each contamination level is highlighted in bold.

We would expect that the classical Kalman filter always performs best in the case of no contamination; this is, according to Tables 1, 2 and 3, only true for Example 4.3 (cf. Table 3). However, it should be noted that in the case of no contamination the classical Kalman filter and both variants of the ACM-type filter perform equally well for Examples 4.1 and 4.2 (cf. Tables 1 and 2). In the case of outliers, regardless of the level of contamination, the ACM-type filter performs best. Additionally, there is no major difference in the performance of both

Table 3: Median MAE and corresponding standard error of the median (given in parentheses) of Example 4.3

		0% (s.e.)	5% (s.e.)	10% (s.e.)	15% (s.e.)	20% (s.e.)
cKF		**2.81 (.03)**	3.03 (.03)	3.22 (.04)	3.44 (.04)	3.59 (.05)
rLS	$b = 3.7$	3.14 (.04)	3.31 (.05)	3.58 (.07)	3.78 (.08)	4.15 (.09)
ACM2	$a = b = 2.5$, $c = 5.0$	2.86 (.03)	2.98 (.03)	3.03 (.04)	**3.18 (.04)**	**3.33 (.04)**
ACM2	$a = b = 2.6$, $c = 3.6$	2.87 (.03)	2.99 (.03)	3.10 (.04)	3.20 (.04)	3.35 (.05)
ACM2	$a = b = 3.0$, $c = 7.0$	2.82 (.03)	**2.98 (.03)**	**3.03 (.03)**	3.20 (.03)	3.36 (.05)

ACM-type filter variants regarding the tuning parameters a, b, and c. The results in Tables 1, 2 and 3 suggest to set the tuning parameters a, b, and c equal to $(a, b, c)^\top = (3.0, 3.0, 7.0)^\top$, if one expects a low to moderate contamination of outliers, and to $(a, b, c)^\top = (2.5, 2.5, 5.0)^\top$ in situations with an expected moderate to high contamination of outliers. As we usually do not know the amount of outliers in practical applications the default values for a, b, and c are set to $(a, b, c)^\top = (2.5, 2.5, 5.0)^\top$.

5. Discussion

It is worth noting that, following the classical Kalman filter routine, the robustness adjustments only alter the correction step (A.3) of the Kalman filter recursions. Thus, our proposed filter procedures are strictly recursive and can therefore be applied to more complex models such as state-space models including exogenous variables or models with time-varying transition and observations matrices, $\boldsymbol{\Phi}_t$ and $\boldsymbol{H}_t$, respectively (see, for example, Durbin and Koopman 2012; Shumway and Stoffer 2025). An extension to time-varying parameter (TVP) regression models (see, for example, Lubik and Matthes 2015; Hauzenberger 2021; Hauzenberger, Huber, Koop, and Onorante 2022; Lucchetti and Valentini 2024) is straightforward as these models can be represented as state-space models, too. Additionally, the proposed filters can easily be used to also robustify the Extended Kalman filter or the Unscented Kalman filter for non-linear state-space models (see, for example, Wan and van der Merwe 2001).

The proposed robustification of the Kalman filter may be linked to score-driven models that are also called dynamic conditional score models (Harvey 2013) or generalized autoregressive score models (Creal, Koopman, and Lucas 2013). A score-driven model can be regarded as providing an approximation to the solution for the corresponding parameter-driven unobserved component model (Harvey 2022). The unobserved component model is in state-space form, and as such, it may be handled by the Kalman filter. The main ingredient in the score-driven approach is the replacement of the one-step ahead prediction error in the Kalman filter by a variable that is proportional to the score of an assumed conditional distribution of the observation at time t. Thus, this setup allows the score-driven model to guard against outliers.

As already mentioned in Section 1 outliers may be propagating or non-propagating, which induces the somewhat conflicting goals of tracking and attenuation. Here, we focus on additive outliers, i.e., we are interested in estimating the underlying clean state process $\boldsymbol{x}_t$ in situations where the observed signal $\boldsymbol{y}_t$ is contaminated by non-propagating outliers. This is a typical filtering application. On the other hand, in situations where the state process $\boldsymbol{x}_t$ is affected by innovation outliers we want to follow the state process as quickly as possible. This typical tracking application was studied in Ruckdeschel, Spangl, and Pupashenko (2014).

Note that the proposed psi-functions allow you to identify outliers on the fly. The psi-function will downweight an outlying observation if the length of $\Delta\boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}$ exceeds a certain value. Thus, it will downweight all coordinates simultaneously. However, there are situations where only one or more coordinates of an observation are suspect whereas the others are fine.

By downweighting such an observation entirely we may lose valuable information residing in its uncontaminated coordinates. Hence, another contamination model may be considered, that of cellwise outliers (cf. Alqallaf, Van Aelst, Yohai, and Zamar 2009). In an autoregressive model AR(p) of order p outlying observations will typically lead to cellwise outliers. Raymaekers and Rousseeuw (2024), to their knowledge, have been the first who applied cellwise robust methods to AR(p) models.

Although the estimation of the model parameters is not the objective of this paper, as we assume they are known, we would nonetheless like to make a few comments on their estimation. The model parameters may be estimated by maximum likelihood (see, for example, Durbin and Koopman 2012). However, it is well known that maximum likelihood estimators can be very sensitive to outliers (see, for example, Neykov *et al.* 2007). To overcome this problem robustified maximum likelihood estimators such as the weighted trimmed likelihood estimator (TLE; Vandev and Neykov 1998) or the weighted maximum likelihood estimator (WLE; Markatou, Basu, and Lindsay 1998) have been proposed. The latter was used to robustly estimate the unknown parameters of an autoregressive-moving average model by Agostinelli (2003). Instead of maximizing the loglikelihood directly by means of iterative numerical procedures we suggest to use an EM-type algorithm proposed by Shumway and Stoffer (1982). This EM algorithm is not robust per se but it can easily be robustified by using robust filters and smoothers. This approach using the rLS filter has been studied by Ruckdeschel *et al.* (2014).

In general, the number of model parameters grows at a quadratic rate with the process dimension. However, knowing the physical properties underlying the process being modeled can reduce the number of parameters that need to be estimated. Hence, for practical applications the transition and observation matrices will often be sparse and the covariance matrices will have a diagonal or block-diagonal structure.

6. Conclusions

Based on the work of Masreliez (1975) and Martin (1979) new multivariate approximate conditional-mean (ACM) type filters for state-space models with vector-valued observations are developed which generalize Martin's results (Martin 1979).

The results of the simulation experiments show that the multivariate ACM-type filters perform very well compared to the rLS filter proposed by Ruckdeschel (2001). Additionally, the results suggest to set the tuning parameters a, b, and c equal to $(a, b, c)^\top = (3.0, 3.0, 7.0)^\top$, if one expects a low to moderate contamination of outliers, and to $(a, b, c)^\top = (2.5, 2.5, 5.0)^\top$ in situations with an expected moderate to high contamination of outliers. Moreover, the ACM-type filters yield remarkably good results in situations where the $\boldsymbol{v}_t$'s have a contaminated multivariate normal distribution as well as in outlier situations where the mean of the contaminating distribution is additionally unequal to zero.

The adaptation to other kinds of outliers, namely innovation outliers and their joint appearance with additive outliers, are the focus of ongoing research.

Acknowledgments

The computation and all graphics were done with R (cf. R Core Team 2024). The rLS filter algorithm was implemented by Peter Ruckdeschel. An R package, containing the multivariate ACM-type filter algorithms and the rLS filter algorithm, exists and is available at https://R-forge.R-project.org/projects/robkalman/.

Additionally, the author thanks two anonymous referees and the editors for their valuable and helpful comments.

A. The classical Kalman filter

The primary aim of any analysis using state-space models as defined by (1) is to find estimators of the underlying unobserved signal $\boldsymbol{x}_t$, given the data $\boldsymbol{Y}_s = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_s\}$, up to time s. If $s < t$, $s = t$ or $s > t$, the problem is called *prediction*, *filtering* or *smoothing*, respectively.

In addition, the obtained estimators T_t of $\boldsymbol{x}_t$ given $\boldsymbol{Y}_s$ should be best in the sense of the minimum mean-squared error. The solution is the conditional mean of $\boldsymbol{x}_t$ given $\boldsymbol{Y}_s$, i.e.,

$$T_t(\boldsymbol{Y}_s) = E(\boldsymbol{x}_t|\boldsymbol{Y}_s) \ ,$$

and will further on be denoted by $\boldsymbol{x}_{t|s}$.

If the problem is restricted to the class of linear estimators, the solution to these problems is accomplished via the *Kalman filter* and *smoother* (cf. Kalman 1960; Kalman and Bucy 1961).

In the following we will first focus on the Kalman filter. Its recursions can be split into three steps:

1. Initialization ($t = 0$):

$$\boldsymbol{x}_{0|0} = \boldsymbol{\mu}_0 \ , \quad \boldsymbol{P}_{0|0} = \boldsymbol{\Sigma}_0 \ , \tag{A.1}$$

   where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the p-dimensional unconditional mean and $p \times p$ covariance matrix of $\boldsymbol{x}_0$.

2. Prediction ($t \geq 1$):

$$\begin{aligned} \boldsymbol{x}_{t|t-1} &= \boldsymbol{\Phi}\boldsymbol{x}_{t-1|t-1} \ , \\ \boldsymbol{P}_{t|t-1} &= \boldsymbol{\Phi}\boldsymbol{P}_{t-1|t-1}\boldsymbol{\Phi}^\top + \boldsymbol{Q} \ . \end{aligned} \tag{A.2}$$

3. Correction ($t \geq 1$):

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}) \ , \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t\boldsymbol{H}\boldsymbol{P}_{t|t-1} \ , \\ \text{with} \quad \boldsymbol{K}_t &= \boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^\top + \boldsymbol{R})^{-1} \ . \end{aligned} \tag{A.3}$$

The $p \times q$ matrix $\boldsymbol{K}_t$ is called the *Kalman gain*. The $p \times p$ matrix $\boldsymbol{P}_{t|t-1}$ is the conditional prediction error covariance matrix,

$$\boldsymbol{P}_{t|t-1} = E\Big((\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})^\top|\boldsymbol{Y}_{t-1}\Big) \ ,$$

and the conditional filtering error covariance matrix $\boldsymbol{P}_{t|t}$ is given by

$$\boldsymbol{P}_{t|t} = E\Big((\boldsymbol{x}_t - \boldsymbol{x}_{t|t})(\boldsymbol{x}_t - \boldsymbol{x}_{t|t})^\top|\boldsymbol{Y}_t\Big) \ .$$

The $p \times p$ matrix $\boldsymbol{Q}$ and the $q \times q$ matrix $\boldsymbol{R}$ denote the covariance matrices of $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$, respectively.

We now state the recursions of the Kalman smoother based on Kalman filtering results:

4. Smoothing ($t \leq n$):

$$\begin{aligned} \boldsymbol{x}_{t-1|n} &= \boldsymbol{x}_{t-1|t-1} + \boldsymbol{J}_{t-1}(\boldsymbol{x}_{t|n} - \boldsymbol{\Phi}\boldsymbol{x}_{t-1|t-1}) \ , \\ \boldsymbol{P}_{t-1|n} &= \boldsymbol{P}_{t-1|t-1} - \boldsymbol{J}_{t-1}(\boldsymbol{P}_{t|n} - \boldsymbol{P}_{t|t-1})\boldsymbol{J}_{t-1}^\top \\ \text{with} \quad \boldsymbol{J}_{t-1} &= \boldsymbol{P}_{t-1|t-1}\boldsymbol{\Phi}^\top\boldsymbol{P}_{t|t-1}^{-1} \ . \end{aligned} \tag{A.4}$$

Moreover, let $f_{\boldsymbol{x}_t}(\cdot|\boldsymbol{Y}_{t-1})$ denote the state prediction density, i.e., the density of $\boldsymbol{x}_t$ conditioned on prior observations $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}$. Similarly, $f_{\boldsymbol{y}_t}(\cdot|\boldsymbol{Y}_{t-1})$ is the observation prediction density conditioned on the past observations.

References

Agostinelli C (2003). "Robust Time Series Estimation via Weighted Likelihood." In R Dutter, P Filzmoser, U Gather, PJ Rousseeuw (eds.), *Developments in Robust Statistics*, pp. 1–16. Physica Verlag, Heidelberg. doi:10.1007/978-3-642-57338-5_1.

Alqallaf F, Van Aelst S, Yohai VJ, Zamar RH (2009). "Propagation of Outliers in Multivariate Data." *The Annals of Statistics*, **37**(1), 311–331. doi:10.1214/07-aos588.

Becker A (2023). "Kalman Filter from the Ground Up." Self-published. URL https://www.kalmanfilter.net/.

Boudt K, Croux C (2010). "Robust M-Estimation of Multivariate GARCH Models." *Computational Statistics & Data Analysis*, **54**(11), 2459–2469. doi:10.1016/j.csda.2009.11.007.

Cramér H (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton.

Creal D, Koopman SJ, Lucas A (2013). "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics*, **28**(5), 777–795. doi:10.1002/jae.1279.

Durbin J, Koopman SJ (2012). *Time Series Analysis by State Space Methods*. 2nd edition. Oxford University Press, New York. doi:10.1093/acprof:oso/9780199641178.001.0001.

Fox AJ (1972). "Outliers in Time Series." *Journal of the Royal Statistical Society B*, **34**(3), 350–363. doi:10.1111/j.2517-6161.1972.tb00912.x.

Gandhi MA, Mili L (2010). "Robust Kalman Filter Based on a Generalized Maximum-Likelihood-Type Estimator." *IEEE Transactions on Signal Processing*, **58**(5), 2509–2520. doi:10.1109/tsp.2009.2039731.

Harvey AC (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press, Cambridge. doi:10.1017/cbo9781139540933.

Harvey AC (2022). "Score-Driven Time Series Models." *Annual Review of Statistics and Its Application*, **9**(2022), 321–342. doi:10.1146/annurev-statistics-040120-021023.

Hauzenberger N (2021). "Flexible Mixture Priors for Large Time-varying Parameter Models." *Econometrics and Statistics*, **20**, 87–108. doi:10.1016/j.ecosta.2021.06.001.

Hauzenberger N, Huber F, Koop G, Onorante L (2022). "Fast and Flexible Bayesian Inference in Time-varying Parameter Regression Models." *Journal of Business & Economic Statistics*, **40**(4), 1904–1918. doi:10.1080/07350015.2021.1990772.

Insightful (2001). *S-PLUS 6 for Windows: Guide to Statistics, Vol. 2*. Insightful Corp., Seattle.

Jazwinski A (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Kalman RE (1960). "A New Approach to Linear Filtering and Prediction Problems." *ASME Journal of Basic Engineering*, **82**(1), 35–45. doi:10.1115/1.3662552.

Kalman RE, Bucy RS (1961). "New Results in Filtering and Prediction Theory." *ASME Journal of Basic Engineering*, **83**(1), 95–108. doi:10.1115/1.3658902.

Leisch F (2004). "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R." *Journal of Statistical Software*, **11**(8), 1–18. doi:10.18637/jss.v011.i08.

Leisch F (2008). "Modelling Background Noise in Finite Mixtures of Generalized Linear Regression Models." In P Brito (ed.), *Compstat 2008–Proceedings in Computational Statistics*, pp. 385–396. Physica Verlag, Heidelberg. doi:10.1007/978-3-7908-2084-3_32.

Lubik TA, Matthes C (2015). "Time-varying Parameter Vector Autoregressions: Specification, Estimation, and an Application." *Economic Quarterly*, **Q4**, 323–352. doi:10.21144/eq1010403.

Lucchetti RJ, Valentini F (2024). "Linear Models with Time-varying Parameters: A Comparison of Different Approaches." *Computational Statistics*, **39**(7), 3523–3545. doi:10.1007/s00180-023-01452-3.

Markatou M, Basu A, Lindsay BG (1998). "Weighted Likelihood Equations with Bootstrap Root Search." *Journal of the American Statistical Association*, **93**(442), 740–750. doi:10.1080/01621459.1998.10473726.

Martin RD (1979). "Approximate Conditional-Mean Type Smoothers and Interpolators." In T Gasser, M Rosenblatt (eds.), *Smoothing Techniques for Curve Estimation*. Springer-Verlag, Berlin. doi:10.1007/bfb0098493.

Martin RD (1981). "Robust Methods for Time Series." In DF Findley (ed.), *Applied Time Series II*. Academic Press, New York. doi:10.1016/b978-0-12-256420-8.50027-7.

Martin RD, Thomson DJ (1982). "Robust-Resistant Spectrum Estimation." *Proceedings of the IEEE*, **70**(9), 1097–1115. doi:10.1109/proc.1982.12434.

Masreliez CJ (1975). "Approximate Non-Gaussian Filtering with Linear State and Observation Relations." *IEEE Transactions on Automatic Control*, **20**(1), 107–110. doi:10.1109/tac.1975.1100882.

Neykov N, Filzmoser P, Dimova R, Neytchev P (2007). "Robust Fitting of Mixtures Using the Trimmed Likelihood Estimator." *Computational Statistics & Data Analysis*, **52**(1), 299–308. doi:10.1016/j.csda.2006.12.024.

Raymaekers J, Rousseeuw PJ (2024). "Challenges of Cellwise Outliers." *Econometrics and Statistics*. doi:10.1016/j.ecosta.2024.02.002. In press.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.

Ruckdeschel P (2000). "Robust Kalman Filtering." In W Härdle, Z Hlávka, S Klinke (eds.), *XploRe—Application Guide*, pp. 483–516. Springer-Verlag. doi:10.1007/978-3-642-57292-0_18.

Ruckdeschel P (2001). *Ansätze zur Robustifizierung des Kalman-Filters*. Ph.D. thesis, Bayreuther Mathematische Schriften, Mathematisches Institut, Universität Bayreuth, Bayreuth.

Ruckdeschel P, Spangl B, Pupashenko D (2014). "Robust Kalman Tracking and Smoothing with Propagating and Non-propagating Outliers." *Statistical Papers*, **55**(1), 93–123. doi:10.1007/s00362-012-0496-4.

Shumway RH, Stoffer DS (1982). "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm." *Journal of Time Series Analysis*, **3**(4), 253–264. doi:10.1111/j.1467-9892.1982.tb00349.x.

Shumway RH, Stoffer DS (2025). *Time Series Analysis and Its Applications: With R Examples*. Springer-Verlag, New York. doi:10.1007/978-3-031-70584-7.

Vandev DL, Neykov NM (1998). "About Regression Estimators with High Breakdown Point." *Statistics*, **32**(2), 111–129. doi:10.1080/02331889808802657.

Wan EA, van der Merwe R (2001). "The Unscented Kalman Filter." In S Haykin (ed.), *Kalman Filtering and Neural Networks*. John Wiley & Sons, New York. doi:10.1002/0471221546.ch7.

Affiliation:

Bernhard Spangl
Institue of Statistics
BOKU University
A-1180 Vienna, Austria
E-mail: bernhard.spangl@boku.ac.at
URL: https://boku.ac.at/nwnr/stat

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 54 *Submitted:* 2024-10-25
2025 *Accepted:* 2025-04-07