Kybernetika

Tomáš Cipra; Ma.Rosario Romera Ayllón

Robust Kalman filter and its application in time series analysis

Kybernetika, Vol. 27 (1991), No. 6, 481--494

Persistent URL: `http://dml.cz/dmlcz/124292`

KYBERNETIKA — VOLUME 27 (1991), NUMBER 6

ROBUST KALMAN FILTER AND ITS APPLICATION IN TIME SERIES ANALYSIS

TOMÁŠ CIPRA, ROSARIO ROMERA

A method of robustification of the Kalman filter is suggested in the paper. In general, the method provides approximative recursive formulas for robust estimation of the state but in some special cases exact recursive formulas can be derived. The steady model and the AR(1) model are investigated in more details including a simulation study and the strong consistency of the recursive formulas for the robust estimation of the autoregressive parameter.

1. INTRODUCTION

The Kalman filter is a useful instrument for recursive treatment of dynamic linear systems (see e.g. [2]) including some popular time series model (nowadays there are even various non-linear generalizations of the Kalman filter).

Let us consider a dynamic system of the form

$$x_t = F_t x_{t-1} + w_t\,, \tag{1.1}$$

$$y_t = H_t x_t + v_t\,, \tag{1.2}$$

where

$$\mathsf{E}w_t = 0\,,\quad \mathsf{E}v_t = 0\,,\quad \mathsf{E}(w_s w_t') = \delta_{st} Q_t\,,\quad \mathsf{E}(v_s v_t') = \delta_{st} R_t\,,\quad \mathsf{E}(w_s v_t') = 0 \tag{1.3}$$

and some initial conditions are fulfilled. The state equation (1.1) describes behavior of an n-dimensional state vector x_t in time while the observation equation (1.2) describes relation of the unobservable state x_t to an m-dimensional observation vector y_t. The matrices F_t, H_t, Q_t, R_t of appropriate dimensions are supposed to be known.

The Kalman filter gives recursive formulas for construction of the linear minimum variance estimator $\hat{x}_t^t$ of the state x_t and for its error covariance matrix $P_t^t = \mathsf{E}(x_t - \hat{x}_t^t)(x_t - \hat{x}_t^t)'$ in a current time period t using all previous information $Y^t = \{y_0, y_1, \ldots, y_t\}$. These formulas have the form

$$\hat{x}_t^t = \hat{x}_t^{t-1} + P_t^{t-1} H_t'(H_t P_t^{t-1} H_t' + R_t)^{-1}(y_t - H_t \hat{x}_t^{t-1})\,, \tag{1.4}$$

$$P_t^t = P_t^{t-1} - P_t^{t-1} H_t'(H_t P_t^{t-1} H_t' + R_t)^{-1} H_t P_t^{t-1}\,, \tag{1.5}$$

where

$$\hat{x}_t^{t-1} = F_t \hat{x}_{t-1}^{t-1}, \tag{1.6}$$

$$P_t^{t-1} = F_t P_{t-1}^{t-1} F_t' + Q_t \tag{1.7}$$

are predictive values constructed for time t at time $t-1$.

The standard Kalman filter supposes normal distributions of the residuals w_t and v_t, i.e.

$$w_t \sim N(0, Q_t), \quad v_t \sim N(0, R_t). \tag{1.8}$$

Then $\hat{x}_t^t$ is even the minimum variance estimator of the state since it holds

$$\hat{x}_t^t = \mathsf{E}(x_t \mid Y^t). \tag{1.9}$$

However, the assumption of the normal residuals is not frequently fulfilled in practice where one must face various forms of contamination of data. Therefore robustification of the Kalman filter is very important from the practical point of view. Various robust modifications of the Kalman filter have been suggested in the literature (see e.g. [6], [13], [14], [15], [17], [21]). Some of them are connected with difficulties when they are applied practically (e.g. the approach in [14] assumes that one can construct such linear transform T_t that the transformed residual process $T_t(y_t - H_t\hat{x}_t^{t-1})$ has some special distributional properties although the transformation T_t depends on the distribution of the residuals which is not apriori known).

In Section 2 a robust modification of the Kalman filter is suggested which seems to be simple from the numerical point of view. The robustification is based on the methodology of the M-estimators (see e.g. [11]) and, in general, it gives approximative recursive formulas for robust estimation of the state. Some special cases which enable to construct exact recursive formulas are described in Section 3. Numerical examples are given in Section 4. The strong consistency of the recursive formulas for robust estimation of the autoregressive parameter in the model AR(1) is proved in the Appendix.

2. ROBUST KALMAN FILTER

It is known (see e.g. [3]) that the current state estimate $\hat{x}_t^t$ in (1.4) can be derived from the predictive values $\hat{x}_t^{t-1}$ and P_t^{t-1} in (1.6) and (1.7), when a current value y_t is observed, by the following minimization procedure

$$\hat{x}_t^t = \operatorname{argmin} \{(\hat{x}_t^{t-1} - x_t)' (P_t^{t-1})^{-1} (\hat{x}_t^{t-1} - x_t) + \\ + (y_t - H_t x_t)' R_t^{-1} (y_t - H_t x_t)\}, \tag{2.1}$$

where argmin is taken over $x_t \in \mathbb{R}^n$. The procedure (2.1) can be looked upon as the weighted least squares method and it is equivalent to the (non-weighted) least squares method in the linear regression model

$$\begin{pmatrix} (P_t^{t-1})^{-1/2} \hat{x}_t^{t-1} \\ R_t^{-1/2} y_t \end{pmatrix} = \begin{pmatrix} (P^{t-1})^{-1/2} \\ R_t^{-1/2} H_t \end{pmatrix} x_t + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}, \tag{2.2}$$

where the symbol $D^{-1/2}$ denotes the square root matrix of an inverse matrix D^{-1} and the residuals ε_t and η_t fulfil

$$\mathsf{E}\varepsilon_t = 0\,, \quad \mathsf{E}\eta_t = 0\,, \quad \operatorname{var}\begin{pmatrix}\varepsilon_t\\ \eta_t\end{pmatrix} = I\,. \tag{2.3}$$

Moreover, if one compares (2.2) with (1.1) and (1.2) then one obtains

$$\varepsilon_t = (P_t^{t-1})^{-1/2}\,(\hat{x}_t^{t-1} - F_t x_{t-1} - w_t)\,, \quad \eta_t = R_t^{-1/2} v_t \tag{2.4}$$

so that a possible contamination of w_t results in a contamination of ε_t without affecting η_t and, similarly, a possible contamination of v_t results in a contamination of η_t without affecting ε_t (in the current time period t).

Let us rewrite the model (2.2) separately for particular rows as

$$\begin{aligned} p_{it} &= a_{it} x_t + \varepsilon_{it}\,, \quad i = 1, \dots, n\,,\\ s_{jt} &= b_{jt} x_t + \eta_{jt}\,, \quad j = 1, \dots, m\,, \end{aligned} \tag{2.5}$$

where

$$(P_t^{t-1})^{-1/2}\,\hat{x}_t^{t-1} = \begin{pmatrix} p_{1t}\\ \vdots \\ p_{nt}\end{pmatrix}, \quad R_t^{-1/2} y_t = \begin{pmatrix} s_{1t}\\ \vdots \\ s_{mt}\end{pmatrix},$$

$$(P_t^{t-1})^{-1/2} = \begin{pmatrix} a_{1t}\\ \vdots \\ a_{nt}\end{pmatrix}, \quad R_t^{-1/2} H_t = \begin{pmatrix} b_{1t}\\ \vdots \\ b_{mt}\end{pmatrix},$$

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1t}\\ \vdots \\ \varepsilon_{nt}\end{pmatrix}, \quad \eta_t = \begin{pmatrix} \eta_{1t}\\ \vdots \\ \eta_{mt}\end{pmatrix}.$$

The model (2.5) has such form that the corresponding least squares method

$$\hat{x}_t^t = \operatorname{argmin}\,\Big\{\sum_{i=1}^{n} (p_{it} - a_{it} x_t)^2 + \sum_{j=1}^{m} (s_{jt} - b_{jt} x_t)^2\Big\} \tag{2.6}$$

can be easily robustified replacing (2.6) by

$$\hat{x}_t^t = \operatorname{argmin}\,\Big\{\sum_{i=1}^{n} \varrho_{1i}(p_{it} - a_{it} x_t) + \sum_{j=1}^{m} \varrho_{2j}(s_{jt} - b_{jt} x_t)\Big\} \tag{2.7}$$

(argmin is taken over $x_t \in \mathbb{R}^n$), where ϱ_{1i} and ϱ_{2j} are suitable robustifying functions with derivatives ψ_{1i} $(i = 1, \dots, n)$ and ψ_{2j} $(j = 1, \dots, m)$ used in the methodology of M-estimation. According to (2.4) the application of the robustifying functions ϱ_{1i} suppresses consequences of a contamination of w and, similarly, the application of the robustifying functions ϱ_{2j} suppresses consequence of a contamination of v.

The normal equations for $\hat{x}_t^t$ corresponding to (2.7) have the form

$$\sum_{i=1}^{n} a'_{it}\,\psi_{1i}(p_{it} - a_{it}\hat{x}_t^t) + \sum_{j=1}^{m} b'_{jt}\,\psi_{2j}(s_{jt} - b_{jt}\hat{x}_t^t) = 0 \tag{2.8}$$

and can be solved explicitly only in some special cases (see Section 3). In general, one can use the following approximative normal equations

$$\sum_{i=1}^{n} w_{1it} a'_{it}(p_{it} - a_{it}\hat{x}_t^t) + \sum_{j=1}^{m} w_{2jt} b'_{jt}(s_{jt} - b_{jt}\hat{x}_t^t) = 0\,, \tag{2.9}$$

where the weights w_{1it} $(i = 1, \ldots, n)$ and w_{2jt} $(j = 1, \ldots, m)$ are defined as

$$\begin{aligned} w_{1it} &= \frac{\psi_{1i}(p_{it} - a_{it}\hat{x}_t^{t-1})}{p_{it} - a_{it}\hat{x}_t^{t-1}}\,, \\ w_{2jt} &= \frac{\psi_{2j}(s_{jt} - b_{jt}\hat{x}_t^{t-1})}{s_{jt} - b_{jt}\hat{x}_t^{t-1}}\,. \end{aligned} \tag{2.10}$$

The equations (2.9) follow from (2.8) if we approximate $\hat{x}_t^t$ by $\hat{x}_t^{t-1}$. They can be considered as a recursive variant of the normal equations from the IWLS (Iterated Weighted Least Squares) method which is a popular algorithm for numerical calculation of M-estimates (see e.g. [11], [22]).

Using the approximation (2.9) one obtains after some algebraic treatment the following robust modification of the recursive formulas (1.4) and (1.5)

$$\begin{aligned} \hat{x}_t^t = \hat{x}_t^{t-1} &+ (P_t^{t-1})^{1/2}\, W_{1t}^{-1}(P_t^{t-1})^{1/2}\, H'_t[H_t(P_t^{t-1})^{1/2}\, W_{1t}^{-1}(P_t^{t-1})^{1/2}\, H'_t + \\ &+ R_t^{1/2} W_{2t}^{-1} R_t^{1/2}]^{-1}\,(y_t - H_t\hat{x}_t^{t-1})\,, \end{aligned} \tag{2.11}$$

$$\begin{aligned} P_t^t = (P_t^{t-1})^{1/2}\, & W_{1t}^{-1}(P_t^{t-1})^{1/2} - (P_t^{t-1})^{1/2}\, W_{1t}^{-1}(P_t^{t-1})^{1/2}\,. \\ &. H'_t[H_t(P_t^{t-1})^{1/2}\, W_{1t}^{-1}(P_t^{t-1})^{1/2}\, H'_t + R_t^{1/2} W_{2t}^{-1} R_t^{1/2}]^{-1}\,. \\ &. H_t(P_t^{t-1})^{1/2}\, W_{1t}^{-1}(P_t^{t-1})^{1/2}\,, \end{aligned} \tag{2.12}$$

where $\hat{x}_t^{t-1}$, P_t^{t-1} are given in (1.6), (1.7) and

$$W_{1t} = \operatorname{diag}\{w_{11t}, \ldots, w_{1nt}\}\,, \quad W_{2t} = \operatorname{diag}\{w_{21t}, \ldots, w_{2mt}\}\,.$$

3. SPECIAL CASES

In this section some special cases of the model (1.1), (1.2) are given which enable to find the explicit solution of the normal equations (2.8). In this way one obtains the non-approximative robust modification of the recursive formula (1.4) for the state estimate $\hat{x}_t^t$. On the other hand, the derivation of the non-approximative recursive formula for the corresponding error covariance matrix P_t^t is usually so difficult that we recommend to use the classical formula (1.5).

One of the most frequent types of contaminated data are ε-contaminated normal data in which a normal distribution with an acceptable variance is contaminated by a small fraction ε (e.g. $\varepsilon = 0{\cdot}05$) of a symmetric distribution with heavy tails (it is the source of so called outliers). For such data with ε-contaminated distribution $N(0, 1)$ (the unit variance can be achieved by means of standardization) the Huber's function ψ_H of the form

$$\psi_H(z) = \begin{cases} z & \text{for } |z| \leqq c \\ c \operatorname{sgn}(z) & \text{for } |z| > c \end{cases} \tag{3.1}$$

gives robust estimates of location which are optimal in the min-max sense, i.e. which have the minimal variance over the least favorable distributions (see e.g. [11], [14]). The constant c depends on ε (e.g. one recommends $c = 1.645$ for $\varepsilon = 0.05$). Considering its practical importance we confine ourselves in the following text to the ψ functions of the type (3.1). In the case without contamination we shall use the classical least squares approach with the function

$$\psi_{LS}(z) = z\,. \tag{3.2}$$

Some of the following models are very popular in time series analysis.

(a) Kalman filter with contaminated scalar observations:

$$x_t = F_t x_{t-1} + w_t\,, \tag{3.3}$$

$$y_t = h_t x_t + v_t\,. \tag{3.4}$$

It is a special case of (1.1)–(1.3) with $m = 1$ and

$$w_t \sim N(0, Q_t)\,, \quad v_t \sim \varepsilon\text{-contaminated} \quad N(0, r_t)\,. \tag{3.5}$$

The normal equations (2.8) with $\psi_{1i} = \psi_{LS}$ $(i = 1, \ldots, n)$ and $\psi_{21} = \psi_H$ give the following robust recursive formulas

$$\hat{x}_t^t = \hat{x}_t^{t-1} + P_t^{t-1} h_t' r_t^{-1/2} \psi_H\left(\frac{r_t^{1/2}(y_t - h_t \hat{x}_t^{t-1})}{h_t P_t^{t-1} h_t' + r_t}\right), \tag{3.6}$$

$$P_t^t = P_t^{t-1} - \frac{P_t^{t-1} h_t' h_t P_t^{t-1}}{h_t P_t^{t-1} h_t' + r_t}\,, \tag{3.7}$$

where $\hat{x}_t^{t-1}$ and P_t^{t-1} are given in (1.6) and (1.7). The formula (3.6) can be rewritten as

$$\hat{x}_t^t = \hat{x}_t^{t-1} + \begin{cases} \dfrac{P_t^{t-1} h_t'}{h_t P_t^{t-1} h_t' + r_t} (y_t - h_t \hat{x}_t^{t-1}) \\ \qquad \text{for } |y_t - h_t \hat{x}_t^{t-1}| \leqq c r_t^{-1/2} (h_t P_t^{t-1} h_t' + r_t) \\ P_t^{t-1} h_t' r_t^{-1/2} c \operatorname{sgn}(y_t - h_t \hat{x}_t^{t-1}) \\ \text{otherwise}\,. \end{cases} \tag{3.8}$$

If one uses a general function ψ instead of ψ_H then according to (2.11) and (2.12)

$$\hat{x}_t^t = \hat{x}_t^{t-1} + \frac{P_t^{t-1} h_t'}{h_t P_t^{t-1} h_t' + r_t/w_t} (y_t - h_t \hat{x}_t^{t-1})\,, \tag{3.9}$$

$$P_t^t = P_t^{t-1} - \frac{P_t^{t-1} h_t' h_t P_t^{t-1}}{h_t P_t^{t-1} h_t' + r_t/w_t}\,, \tag{3.10}$$

where

$$w_t = \frac{\psi(r_t^{-1/2}(y_t - h_t \hat{x}_t^{t-1}))}{r^{-1/2}(y_t - h_t \hat{x}_t^{t-1})}. \tag{3.11}$$

The case described in (a) is applicable e.g. in the situation when one estimates recursively regression parameters x_t in a linear regression model (3.4) with contaminated observations y_t (in the simplest case one can put $F_t = I$, $Q_t = 0$, $r_t = \sigma^2$).

(b) Filtering in steady model with contaminated observations:

$$x_t = x_{t-1} + w_t, \tag{3.12}$$

$$y_t = x_t + v_t, \tag{3.13}$$

where in addition to (1.3)

$$w_t \sim N(0, q_t), \quad v_t \sim \varepsilon\text{-contaminated} \quad N(0, r_t). \tag{3.14}$$

The one-dimensional process y_t ($n = 1$), which presents a one-dimensional random walk x_t ($m = 1$) observed with an error v_t, is called the steady model and has useful applications in practical time series analysis (see e.g. [8]). According to (3.6) and (3.7) one obtains the following robust recursive formulas for the filtered values of the process y_t

$$\hat{x}_t^t = \hat{x}_{t-1}^{t-1} + (P_{t-1}^{t-1} + q_t)\, r_t^{-1/2}\, \psi_H\left(\frac{r_t^{1/2}(y_t - \hat{x}_{t-1}^{t-1})}{P_{t-1}^{t-1} + q_t + r_t}\right), \tag{3.15}$$

$$P_t^t = \frac{(P_{t-1}^{t-1} + q_t)\, r_t}{P_{t-1}^{t-1} + q_t + r_t}. \tag{3.16}$$

(c) Recursive estimation in autoregressive model AR(p) with innovation outliers:

$$x_t = x_{t-1}, \tag{3.17}$$

$$y_t = h_t x_t + v_t, \tag{3.18}$$

where $h_t = (y_{t-1}, \ldots, y_{t-p})$ and in addition to (1.3)

$$v_t \sim \varepsilon\text{-contaminated} \quad N(0, \sigma^2). \tag{3.19}$$

The one-dimensional process y_t ($n = p$, $m = 1$) is called the autoregressive process with innovation outliers (see e.g. [12], [22]). According to (3.6) and (3.7) one obtains the following robust recursive formulas for the parameter estimates

$$\hat{x}_t^t = \hat{x}_{t-1}^{t-1} + P_{t-1}^{t-1} h_t' \sigma^{-1}\, \psi_H\left(\frac{\sigma(y_t - h_t \hat{x}_{t-1}^{t-1})}{h_t P_{t-1}^{t-1} h_t' + \sigma^2}\right), \tag{3.20}$$

$$P_t^t = P_{t-1}^{t-1} - \frac{P_{t-1}^{t-1} h_t' h_t P_{t-1}^{t-1}}{h_t P_{t-1}^{t-1} h_t' + \sigma^2}. \tag{3.21}$$

Specially in the model AR(1) one has $h_t = y_{t-1}$ so that

$$\hat{x}_t^t = \hat{x}_{t-1}^{t-1} + P_{t-1}^{t-1} y_{t-1} \sigma^{-1} \psi_H \left(\frac{\sigma(y_t - y_{t-1} \hat{x}_{t-1}^{t-1})}{P_{t-1}^{t-1} y_{t-1}^2 + \sigma^2} \right), \tag{3.22}$$

$$P_t^t = \frac{P_{t-1}^{t-1} \sigma^2}{P_{t-1}^{t-1} y_{t-1}^2 + \sigma^2}. \tag{3.23}$$

Various convergence theorems which are mostly based on approaches of the stochastic approximation can be proved for the previous recursive formulas (see also [4], [5], [9], [10], [16], [18], [19], [23] and others). For demonstration in the Appendix we shall give the proof of the following assertion.

Theorem. Let in the model AR(1)

$$y_t = \vartheta y_{t-1} + v_t, \quad t = \ldots, -1, 0, 1, \ldots \tag{3.24}$$

an estimate $\hat{x}_t$ of the parameter ϑ be given by means of the recursive formulas

$$\hat{x}_t = \hat{x}_{t-1} + P_t y_{t-1} \sigma^{-1} \psi_H \left(\frac{\sigma(y_t - y_{t-1} \hat{x}_{t-1})}{P_t y_{t-1}^2 + \sigma^2} \right), \quad t = 1, 2, \ldots, \tag{3.25}$$

$$P_t = \frac{P_{t-1} \sigma^2}{P_{t-1} y_{t-1}^2 + \sigma^2}, \quad t = 1, 2, \ldots \tag{3.26}$$

with initial (random) values $\hat{x}_0$ and P_0. Let the following assumptions be fulfilled

$$|\vartheta| < 1; \tag{3.27}$$

$$v_t \sim \text{iid}, \quad \mathsf{E} v_t = 0, \quad \text{var } v_t = \sigma^2 \quad (0 < \sigma^2 < \infty); \tag{3.28}$$

the distribution of v_t is symmetric such that $F_v(-\varepsilon) < F_v(\varepsilon)$ for each $\varepsilon > 0$; (3.29)

$$\mathsf{E} \hat{x}_0^2 < \infty, \quad P_0 > 0 \quad \text{a.s.}, \quad \hat{x}_0, P_0, v_t \quad \text{are independent}. \tag{3.30}$$

Then

$$\hat{x}_t \to \vartheta \quad \text{a.s.} \tag{3.31}$$

Remark 1. Under the assumptions of Theorem y_t is a stationary AR(1) process such that the common distribution of its white noise has a positive mass concentrated about zero (specially, this assumption is fulfilled for an ε-contaminated normal distribution with zero mean). If comparing (3.25) and (3.26) with (3.22) and (3.23) Theorem uses the simpler denotation $\hat{x}_t$ and P_t instead of $\hat{x}_t^t$ and P_t^t. Moreover, the formula (3.25) contains the more actual value P_t instead of P_{t-1}.

Remark 2. Theorem stays valid if we replace σ^2 in (3.25) and (3.26) by random variables σ_t^2 such that $0 < k \leqq \sigma_t^2 \leqq K < \infty$ a.s. (k, K are constants). Therefore in applications one can use these formulas with a recursive estimate σ_t^2 of σ^2 which is trimmed in a suitable way.

Remark 3. Other generalizations are possible. E. g., one can use a more general function ψ fulfilling

$$|\psi(x)| \leqq c\,, \quad -\infty < x < \infty\,; \tag{3.32}$$

$$b \int_{-\infty}^{\infty} \psi(u(b + x))\, \mathrm{d}F_v(x) \geqq 0\,, \quad -\infty < b < \infty\,, \quad (2\sigma)^{-1} \leqq u \leqq \sigma^{-1}\,; \tag{3.33}$$

if $b_t \int_{-\infty}^{\infty} \psi(u_t(b_t + x))\, \mathrm{d}F_v(x) \to 0$ for an arbitrary sequence

$$(2\sigma)^{-1} \leqq u_t \leqq \sigma^{-1} \quad \text{then} \quad b_t \to 0\,. \tag{3.34}$$

It is also possible to reformulate the problem for non-linear autoregressive models of the type

$$y_t = \vartheta f_{t-1}(Y^{t-1}) + v_t \tag{3.35}$$

(f_t are suitable non-linear functions, see [1]) replacing y_{t-1} by $f_{t-1}(Y^{t-1})$ in (3.25), (3.26) and assuming in addition to (3.27)–(3.30)

$$\mathrm{E}[f_t(Y^t)]^2 \leqq K < \infty\,, \quad t = 0, 1, \ldots; \tag{3.36}$$

$$\sum_{t=0}^{n} [f_t(Y^t)]^2/n \to d \quad \text{a.s.} \tag{3.37}$$

(K, d are constants).

4. NUMERICAL EXAMPLE

Example 1. Peña and Guttman [17] have demonstrated their method of robust Kalman filtering based on calculation of posterior probabilities in mixtures of normal distributions be means of a simulation example for the steady model (3.12), (3.13) with $w_t \sim N(0, 1)$ and $v_t \sim N(0, 4)$. In Table 1 their results are compared with the ones provided by the non-robustified Kalman filter (1.4)–(1.7) and by the robust formulas (3.15), (3.16) with $c = 1{\cdot}645$. Obviously the outlier $y_{20} = 35{\cdot}00$ is not suppressed in the non-robust filtering ($\hat{x}_{20}^{20} = 16{\cdot}76$ is strongly biased) while the both robust methods give acceptable filtered values ($\hat{x}_{20}^{20} = 6{\cdot}04$ and $6{\cdot}87$). The results provided by the both robust methods are comparable but the recursive formulas (3.15), (3.16) are simpler numerically.

Example 2. The process AR(1) of the form $y_t = 0{\cdot}5 y_{t-1} + v_t$ with innovation outliers v_t has been generated for various heavy-tailed distributions contaminating the normal distribution. Table 2 contains the results of the recursive estimation of the parameter $\vartheta = 0{\cdot}5$ for $v_t \sim 0{\cdot}9N(0, 1) + 0{\cdot}1C(0, 3)$ and $v_t \sim 0{\cdot}95N(0, 1) + 0{\cdot}05R(-25, 25)$, where $C(a, b)$ denotes the Cauchy distribution with density $f(x) = (1/(\pi b))\,[1 + ((x - a)/b)^2]^{-1}$ and $R(a, b)$ denotes the uniform distribution on (a, b). The robust recursive estimates $\hat{x}_t^t = \hat{x}_t$ according to (3.25), (3.26) with $\sigma^2 = 1$, $\hat{x}_0 = 0$, $P_0 = 1$ are compared with the non-robust estimates according to (1.4)–(1.7). If an observation y_{t_0} is a distinct outlier then it can deviate unpleasantly the subsequent non-robust estimates $\hat{x}_t^t$ ($t \geqq t_0$) while the robust estimates are not

Table 1. Kalman filter in the simulated steady model (3.12), (3.13) with $w_t \sim N(0, 1)$, $v_t \sim \sim N(0, 4)$.

t	Simulated values (3.12), (3.13)		Non-robust filter (1.4)–(1.7)		Robust filter Peña and Guttman [17]		Robust filter (3.15), (3.16)	
	x_t	y_t	$\hat{x}_t^t$	P_t^t	$\hat{x}_t^t$	P_t^t	$\hat{x}_t^t$	P_t^t
1	10·00	8·65	9·66	4·0	9·66	8·7	9·66	4·0
2	9·83	7·28	8·34	2·2	8·01	3·0	8·34	2·2
3	9·98	7·44	7·94	1·8	7·73	2·0	7·94	1·8
4	8·99	11·13	9·25	1·6	9·12	1·9	9·25	1·6
5	9·36	11·18	10·02	1·6	9·97	1·7	10·02	1·6
6	8·50	5·45	8·22	1·6	8·46	2·3	8·22	1·6
7	8·90	6·17	7·42	1·6	7·45	1·9	7·42	1·6
8	8·20	3·92	6·05	1·6	6·07	1·9	6·05	1·6
9	8·47	12·32	8·50	1·6	6·91	3·6	8·16	1·6
10	7·46	6·95	7·90	1·6	6·93	2·2	7·69	1·6
11	6·49	10·46	8·90	1·6	8·40	2·0	8·77	1·6
12	7·34	9·54	9·15	1·6	8·88	1·7	9·07	1·6
13	7·82	7·07	8·33	1·6	8·16	1·7	8·29	1·6
14	7·06	8·17	8·27	1·6	8·16	1·6	8·24	1·6
15	6·85	5·59	7·22	1·6	7·18	1·6	7·21	1·6
16	5·67	5·99	6·74	1·6	6·71	1·6	6·73	1·6
17	3·69	7·29	6·95	1·6	6·94	1·6	6·95	1·6
18	3·37	5·94	6·56	1·6	6·55	1·6	6·56	1·6
19	3·25	1·96	4·76	1·6	5·10	2·2	4·76	1·6
20	2·81	35·00	16·76	1·6	6·04	3·1	6·87	1·6
21	2·36	−0·62	9·86	1·6	4·81	5·5	4·76	1·6
22	2·46	4·13	7·62	1·6	4·40	2·5	4·51	1·6
23	0·82	−0·84	4·32	1·6	2·61	3·4	2·42	1·6
24	0·24	2·78	3·72	1·6	2·70	2·1	2·56	1·6
25	1·62	1·93	3·02	1·6	2·37	1·8	2·32	1·6
26	1·46	0·45	2·02	1·6	1·60	1·7	1·59	1·6
27	1·96	2·54	2·22	1·6	1·97	1·6	1·96	1·6
28	2·62	−0·95	0·98	1·6	0·86	1·7	0·82	1·6
29	2·95	2·69	1·65	1·6	1·58	1·6	1·55	1·6
30	1·40	−0·89	0·66	1·6	0·63	1·6	0·60	1·6
31	2·84	2·83	1·51	1·6	1·48	1·6	1·47	1·6

affected. E.g., the outlier $y_{479} = -140{\cdot}97$ in the model with $v_t \sim 0{\cdot}9N(0, 1) + 0{\cdot}1C(0, 3)$ produces the non-robust estimate $\hat{x}_{479}^{479} = 0{\cdot}950$ and deviates the subsequent estimates so that ..., $\hat{x}_{500}^{500} = 0{\cdot}689$, ..., $\hat{x}_{550}^{550} = 0{\cdot}680$, ... against the robust estimates $\hat{x}_{479}^{479} = 0{\cdot}500$, ..., $\hat{x}_{500}^{500} = 0{\cdot}507$, ..., $\hat{x}_{550}^{550} = 0{\cdot}506$, ... Similarly, the outlier $y_{89} = 15{\cdot}55$ in the model with $v_t \sim 0{\cdot}95N(0, 1) + 0{\cdot}05R(-25, 25)$ produces the non-robust estimates $\hat{x}_{89}^{89} = 0{\cdot}313$, $\hat{x}_{90}^{90} = 0{\cdot}330$, ..., $\hat{x}_{100}^{100} = 0{\cdot}312$, ... against the robust estimates $\hat{x}_{89}^{89} = 0{\cdot}504$, $\hat{x}_{90}^{90} = 0{\cdot}499$, ..., $\hat{x}_{100}^{100} = 0{\cdot}487$, ...

Table 2. Recursive estimation in the simulated models AR(1) of the form $y_t = 0{\cdot}5y_{t-1} + v_t$ with innovation outliers.

t	$v_t \sim 0{\cdot}9N(0,1) + 0{\cdot}1C(0,3)$ non-robust $\hat{x}_t^t$	robust $\hat{x}_t^t$	$v_t \sim 0{\cdot}95N(0,1) + 0{\cdot}05R(-25,25)$ non-robust $\hat{x}_t^t$	robust $\hat{x}_t^t$
10	−0·310	−0·256	0·700	0·486
20	0·351	0·316	0·661	0·504
30	0·365	0·339	0·651	0·502
40	0·288	0·274	0·622	0·488
50	0·266	0·277	0·595	0·471
60	0·563	0·517	0·452	0·466
70	0·481	0·490	0·421	0·466
80	0·481	0·491	0·511	0·517
90	0·481	0·490	0·330	0·499
100	0·481	0·490	0·312	0·487
150	0·496	0·495	0·357	0·479
200	0·496	0·495	0·368	0·480
250	0·493	0·494	0·368	0·477
300	0·493	0·494	0·395	0·479
350	0·493	0·494	0·405	0·480
400	0·493	0·494	0·434	0·488
450	0·494	0·495	0·439	0·491
500	0·689	0·507	0·438	0·489
550	0·680	0·506	0·452	0·488
600	0·677	0·506	0·458	0·491
650	0·677	0·506	0·477	0·497
700	0·672	0·507	0·476	0·496
750	0·671	0·507	0·476	0·496
800	0·668	0·507	0·480	0.499
850	0·667	0·508	0·480	0·498
900	0·666	0·508	0·480	0·497

APPENDIX: PROOF OF THEOREM

Lemma 1. Let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \ldots \subset \mathscr{F}$ be a sequence of σ-algebras in a probability space $(\Omega, \mathscr{F}, P)$. Let $z_t, \beta_t, \zeta_t, \eta_t$ $(t = 0, 1, \ldots)$ be non-negative $\mathscr{F}$-measurable random variables such that

$$\mathsf{E}(z_t \mid \mathscr{F}_{t-1}) \leqq (1 + \beta_{t-1})\, z_{t-1} + \zeta_{t-1} - \eta_{t-1}, \quad t = 1, 2, \ldots; \tag{A.1}$$

$$\sum_{t=0}^{\infty} \beta_t < \infty \quad \text{a.s.}, \quad \sum_{t=0}^{\infty} \zeta_t < \infty \quad \text{a.s.} \tag{A.2}$$

Then the sequence z_t converges a.s. and

$$\sum_{t=0}^{\infty} \eta_t < \infty \quad \text{a.s.} \tag{A.3}$$

Proof. See [20].

Lemma 2. Let in the model (3.24) an estimate $\hat{x}_t$ of the parameter ϑ be given by means of the recursive formulas

$$\hat{x}_t = \hat{x}_{t-1} + a_{t-1} y_{t-1} \psi_H(u_{t-1}(y_t - y_{t-1}\hat{x}_{t-1})) \,, \quad t = 1, 2, \dots \tag{A.4}$$

with an initial (random) value $\hat{x}_0$. Here a_t and u_t $(t = 0, 1, \dots)$ are $\mathscr{F}_t$-measurable random variables for $\mathscr{F}_t = \sigma\{\hat{x}_0, v_t, v_{t-1}, \dots\}$ fulfilling

$$0 < a_t^{(1)} \leqq a_t \leqq a_t^{(2)} \,, \quad \sum_{t=0}^{\infty} a_t^{(1)} = \infty \,, \quad \sum_{t=0}^{\infty} (a_t^{(2)})^2 < \infty \,; \tag{A.5}$$

$$0 < k \leqq u_t \leqq K < \infty \quad \text{a.s.} \tag{A.6}$$

for deterministic sequences $a_t^{(1)}$, $a_t^{(2)}$ and constants k, K. Let the assumptions (3.27) to (3.30) be fulfilled. Then

$$\hat{x}_t \to \vartheta \quad \text{a.s.} \tag{A.7}$$

Proof. Put

$$\tilde{x}_t = \hat{x}_t - \vartheta \,.$$

Then (A.4) can be rewritten to the form

$$\tilde{x}_t = \tilde{x}_{t-1} - a_{t-1} y_{t-1} \psi_H(u_{t-1}(y_{t-1}\tilde{x}_{t-1} - v_t)) \,.$$

Hence one obtains

$$\tilde{x}_t^2 \leqq \tilde{x}_{t-1}^2 - 2a_{t-1} y_{t-1} \tilde{x}_{t-1} \psi_H(u_{t-1}(\tilde{x}_{t-1} - v_t)) + (a_{t-1}^{(2)} c y_{t-1})^2 \tag{A.8}$$

and for the conditional expectations

$$\mathsf{E}(\tilde{x}_t^2 \mid \mathscr{F}_{t-1}) \leqq \tilde{x}_{t-1}^2 - 2a_{t-1} y_{t-1} \tilde{x}_{t-1} \, \mathsf{E}\{\psi_H(u_{t-1}(y_{t-1}\tilde{x}_{t-1} - v_t)) \mid \mathscr{F}_{t-1}\} + (a_{t-1}^{(2)} c y_{t-1})^2 \,.$$

Let us apply Lemma 1 for $z_t = \tilde{x}_t^2$, $\beta_{t-1} = 0$, $\zeta_{t-1} = (a_{t-1}^{(2)} c y_{t-1})^2$, $\eta_{t-1} = 2a_{t-1} y_{t-1} \tilde{x}_{t-1} \, \mathsf{E}\{\psi_H(u_{t-1}(y_{t-1}\tilde{x}_{t-1} - v_t)) \mid \mathscr{F}_{t-1}\}$. The only problem may be to verify that $\eta_t \geqq 0$ a.s.: Let us denote

$$\varphi(b, u) = \mathsf{E}_v \psi_H(u(b + v)) = \int_{-\infty}^{\infty} \psi_H(u(b + x)) \, \mathrm{d}F_v(x) \,, \quad -\infty < b < \infty \,, \quad k \leqq u \leqq K \,. \tag{A.9}$$

Then due to the assumptions (3.28) and (3.29) it even holds

$$b \, \varphi(b, u) > 0 \,, \quad b \neq 0 \,, \quad k \leqq u \leqq K \,, \tag{A.10}$$

which guarantees specially the non-negativeness of η_t.

According to Lemma 1 there exists a (finite) random variable $\tilde{x}$ such that

$$x_t \to \tilde{x} \quad \text{a.s.} \tag{A.11}$$

From (A.8) it follows for an arbitrary n

$$\tilde{x}_n^2 \leqq \tilde{x}_0^2 - 2\sum_{t=1}^{n} a_{t-1} y_{t-1} \tilde{x}_{t-1} \psi_H(u_t(y_{t-1}\tilde{x}_{t-1} - v_t)) + c^2 \sum_{t=1}^{n} (a_{t-1}^{(2)} y_{t-1})^2$$

and hence

$$2\sum_{t=1}^{\infty} \mathsf{E}\{a_{t-1}y_{t-1}\tilde{x}_{t-1}\psi_H(u_{t-1}(y_{t-1}\tilde{x}_{t-1} - v_t))\} \leqq$$

$$\leqq \mathsf{E}\tilde{x}_0^2 + (c\sigma_y)^2 \sum_{t=1}^{\infty} (a_{t-1}^{(2)})^2 ,$$

where $\sigma_y^2 = \operatorname{var} y_t = \mathsf{E}y_t^2$. Therefore according to (A.5) one has

$$\sum_{t=1}^{\infty} a_{t-1}^{(1)} \mathsf{E}\{y_{t-1}\tilde{x}_{t-1}\psi_H(u_{t-1}(y_{t-1}\tilde{x}_{t-1} - v_t))\} < \infty .$$

Since $\sum a_t^{(1)} = \infty$ a subsequence must exist such that

$$\sum_{t=1}^{\infty} \mathsf{E}\{y_{t_j-1}\tilde{x}_{t_j-1}\psi_H(u_{t_j-1}(y_{t_j-1}\tilde{x}_{t_j-1} - v_{t_j}))\} < \infty .$$

Hence

$$y_{t_j-1}\tilde{x}_{t_j-1}\, \mathsf{E}\{\psi_H(u_{t_j-1}(y_{t_j-1}\tilde{x}_{t_j-1} - v_{t_j})) \mid \mathscr{F}_{t_j-1}\} \to 0 \quad \text{a.s.}$$

or equivalently by means of the denotation (A.9)

$$y_{t_j-1}\tilde{x}_{t_j-1}\varphi(y_{t_j-1}\tilde{x}_{t_j-1}, u_{t_j-1}) \to 0 \quad \text{a.s.}$$

Due to (A.10) it implies

$$y_{t_j-1}\tilde{x}_{t_j-1} \to 0 \quad \text{a.s.} \tag{A.12}$$

Further one can write

$$v_{t_j}\tilde{x}_{t_j-1} = y_{t_j}(\tilde{x}_{t_j-1} - \tilde{x}_{t_j}) + y_{t_j}\tilde{x}_{t_j} - \vartheta y_{t_j-1}\tilde{x}_{t_j-1} . \tag{A.13}$$

Since y_{t_j} are identically distributed and the limit relations (A.11) and (A.12) hold all three summands on the right-hand side of (A.13) converge in probability to zero, i.e.

$$v_{t_j}\tilde{x}_{t_j-1} \to 0 \quad \text{in probability} . \tag{A.14}$$

Due to independence of $\tilde{x}_{t_j-1}$ and v_{t_j}, where v_{t_j} are identically distributed, and due to (A.11) it implies finally

$$\tilde{x}_t \to 0 \quad \text{a.s.}$$

Proof of Theorem. It holds

$$P_t = (P_{t-1}^{-1} + y_{t-1}^2/\sigma^2)^{-1} = [P_0^{-1} + (y_0^2 + \ldots + y_{t-1}^2)/\sigma^2]^{-1} .$$

Hence one obtains that the $\mathscr{F}_t$-measurable random variable

$$u_t = \sigma/(P_{t+1}y_t^2 + \sigma^2) , \quad t = 0, 1, \ldots$$

fulfils

$$(2\sigma)^{-1} \leqq u_t \leqq \sigma^{-1}$$

and further due to the properties of the process y_t (see [7], p. 210, Thm. 6)

$$tP_t \to \sigma^2/\sigma_y^2 \quad \text{a.s.} \tag{A.15}$$

Let us choose arbitrary $\varepsilon > 0$ and $0 < \delta < \sigma^2/\sigma_y^2$. With respect to (A.15) there exists

t_0 such that

$$\mathsf{P}(\sup_{t \geqq t_0} |tP_t - \sigma^2/\sigma_y^2| < \delta) > 1 - \varepsilon$$

and hence

$$\mathsf{P}(\bigcap_{t \geqq t_0} [|tP_t - \sigma^2/\sigma_y^2| < \delta]) > 1 - \varepsilon\,.$$

Put

$$\bar{x}_t = \begin{cases} \hat{x}_t & \text{for} \quad t = 0, 1, \ldots, t_0 - 1 \\ \bar{x}_{t-1} + P_t y_{t-1} \sigma^{-1} \psi_H(u_{t-1}(y_t - y_{t-1}\bar{x}_{t-1})) & \text{for} \quad t \geqq t_0\,, \quad |tP_t - \sigma^2/\sigma_y^2| < \delta \\ \bar{x}_{t-1} + t^{-1}(\sigma/\sigma_y^2)\, y_{t-1} \psi_H(u_{t-1}(y_t - y_{t-1}\bar{x}_{t-1})) & \text{for} \quad t \geqq t_0\,, \quad |tP_t - \sigma^2/\sigma_y^2| \geqq \delta\,. \end{cases}$$

Then according to Lemma 2 with $a_t^{(1)} = t^{-1}(\sigma^2/\sigma_y^2 - \delta)\,\sigma^{-1}$, $a_t^{(2)} = t^{-1}(\sigma^2/\sigma_y^2 + \delta)\,.$ $.\,\sigma^{-1}$, $k = (2\sigma)^{-1}$, $K = \sigma^{-1}$ it holds

$$\bar{x}_t \to \vartheta \quad \text{a.s.}$$

Finally one can write

$$\mathsf{P}(\hat{x}_t \to \vartheta) \geqq \mathsf{P}(\bigcap_{t \geqq t_0} [\bar{x}_t = \hat{x}_t] \cap [\bar{x}_t \to \vartheta]) = \mathsf{P}(\bigcap_{t \geqq t_0} [\bar{x}_t = \hat{x}_t]) \geqq$$
$$\geqq \mathsf{P}(\bigcap_{t \geqq t_0} [|tP_t - \sigma^2/\sigma_y^2| < \delta]) > 1 - \varepsilon\,.$$

Since $\varepsilon > 0$ can be arbitrary it must be $\hat{x}_t \to \vartheta$ a.s. □

ACKNOWLEDGEMENT

We thank to Prof. M. Hušková and Dr. P. Lachout for helpful comments on the proof of Theorem.

(Received July 11, 1970.)

REFERENCES

[1] K. K. Aase: Recursive estimation in non-linear time series models of autoregressive type. J. Roy. Statist. Soc. Ser. B *45* (1983), 228–237.

[2] B. D. O. Anderson and J. B. Moore: Optimal Filtering. Prentice-Hall, Englewood Cliffs, New Jersey 1979.

[3] A. E. Bryson and J. C. Ho: Applied Optimal Control. J. Wiley, New York 1975.

[4] K. Campbell: Recursive computation of M-estimates for the parameters of a finite autoregressive process. Ann. Statist. *10* (1982), 442–453.

[5] J. E. Englund: Multivariate Recursive M-estimators of Location and Scatter for Dependent Sequences. Research Report, University of Lund and Lund Institute of Technology 1988.

[6] A. A. Ershov and R. S. Liptser: Robust Kalman filter in discrete time. Automat. Remote Control *39* (1978), 359–367.

[7] E. J. Hannan: Multiple Time Series. J. Wiley, New York 1970.

[8] P. J. Harrison and C. F. Stevens: Bayesian forecasting. J. Roy. Statist. Soc. Ser. B *38* (1976), 205—247.

[9] U. Holst: Convergence of a recursive stochastic algorithm with m-dependent observations. Scand. J. Statist. *7* (1980), 207—215.

[10] U. Holst: Convergence of a recursive robust algorithm with strongly regular observations. Stochastic Process. Appl. *16* (1984), 305—320.

[11] P. J. Huber: Robust Statistics. J. Wiley, New York 1981.

[12] R. D. Martin: Robust estimation for time series autoregressions. In: Robustness in Statistics (R. L. Launer and G. N. Wilkinson, eds.), Academic Press, New York 1979, pp. 147—176.

[13] C. J. Masreliez: Approximate non-Gaussian filtering with linear state and observation relations. IEEE Trans. Automat. Control *AC-20* (1975), 107—110.

[14] C. J. Masreliez and R. D. Martin: Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Automat. Control *AC-22* (1977), 361—371.

[15] R. J. Meinhold and N. D. Singpurwalla: Robustification of Kalman filter models. J. Amer. Statist. Assoc. *84* (1989), 479—486.

[16] M. Pantel: Adaptive Verfahren der stochastischen Approximation. Dissertation, Universität Essen 1979.

[17] D. Peña and J. Guttman: Optimal collapsing of mixture distributions in robust recursive estimation. Comm. Statist. Theory Methods *18* (1989), 817—833.

[18] B. T. Polyak and Ya. Z. Tsypkin: Adaptive estimation algorithms: convergence, optimality, stability (in Russian). Avtomat. Telemekh. (1979), 3, 71—84.

[19] B. T. Polyak and Ya. Z. Tsypkin: Optimal methods of estimation of autoregressive parameters under incomplete information (in Russian). Tekh. kibernet. (1983), 1, 118—126.

[20] H. Robbins and D. Siegmund: A convergence theorem for non negative almost supermartingales and some applications. In: Optimizing Methods in Statistics (J. S. Rustagi, ed.), Academic Press, New York 1971, pp. 233—257.

[21] L. D. Servi and Y. C. Ho: Recursive estimation in the presence of uniformly distributed measurement noise. IEEE Trans. Automat. Control *AC-26* (1981), 563—565.

[22] N. Stockinger and R. Dutter: Robust Time Series Analysis: A Survey. Supplement to Kybernetika vol. *23* (1987).

[23] Yu. Sh. Verulava: Convergence of a stochastic approximation algorithm for estimating an autoregressive parameter (in Russian). Avtomat. Telemekh. (1981), 7, 115—119.

RNDr. Tomáš Cipra, CSc., matematicko-fyzikální fakulta UK (Faculty of Mathematics and Physics — Charles University), Sokolovská 83, 186 00 Praha 8. Czechoslovakia.

Prof. Rosario Romera, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla de Monte, 28660 Madrid. Spain.