Working Paper 92-09
March 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)

# SQUARE ROOT KALMAN FILTER WITH CONTAMINATED OBSERVATIONS

T. Cipra*
R. Romera**
A. Rubio***

Abstract

The algorithm of square root Kalman filtering for the case of contaminated observations is described in the paper. This algorithm is suitable for the parallel computer implementation allowing to treat dynamic linear systems with large number of state variables in a robust recursive way.

Key words: Square root Kalman filter, robust, parallel algorithm.

* Statistics Department, Charles University of Prague.

** Artificial Intelligence Department, Facultad de Informática, Universidad Politécnica de Madrid; and Statistics and Econometrics Department, Universidad Carlos III de Madrid.

*** Mathematics Department, Universidad de Extremadura.

# SQUARE ROOT KALMAN FILTER WITH CONTAMINATED OBSERVATIONS

**T. Cipra, R. Romera, A. Rubio.**

***Abstract***: The algorithm of square root Kalman filtering for the case of contaminated observations is described in the paper. This algorithm is suitable for the parallel computer implementation allowing to treat dynamic linear systems with large number of state variables in a robust recursive way.

***Key words***: Square root Kalman filter, robust, parallel algorithm.

***Classification***: 62M20, 60G35, 93E11.

***1. Introduction***. This paper attempts to treat simultaneously two problems connected with practical implementation of the Kalman filtering:

(1) If the number n of state variables (the dimension of state vector) is large then the Kalman filter procedure is expensive requiring $0(n^3)$ operations for each state update. Applications with enormous number of state variables appear e.g. in aerodynamics including aircraft testing, medicine, robotics, seismology. Some applications in the framework of state-space modelling of time series with large numbers of state variables are given e.g. in [3]. However, the complexity of Kalman filter can be reduced by the parallel implementation on parallel computers that are used in practice with increasing popularity (see e.g. [1]). Nowadays there are numerous parallel Kalman filter algorithms suggested for various practical situations. One of possible approaches to this problem consists in the square root formulation of Kalman filter allowing to reduce the costs to 0(n) operations for each state update if the algorithm is implemented on a parallel machine (see e.g. [5], [7]).

(2) In practice the Kalman filter must frequently face to various forms of contaminated data. The occurrence of outliers and non-Gaussian distributions in the dynamic linear systems treated in practice has motivated a number of robust versions of Kalman filter (there are even suggestions concerning the robust Kalman filtering for nonlinear systems, see s.g. [2]). For instance, the approach to the Kalman filter robustification by [4] based on the M-estimation principle seems to provide good practical results.

With respect to the mentioned problems (1) and (2), this paper shows that it is not difficult to rewrite the robust Kalman filtering from [4] to the square root form. The robust Kalman filter [4] for models with contaminated observations is briefly reminded in Section 2, its square root form is described in Section 3 and the special case with scalar (contaminated) observations is considered in Section 4.

***2. Robust Kalman filter*** . Let us consider a dynamic linear system with contaminated observations of the form

$$x_{t+1} = F_t x_t + \omega_t, \qquad \omega_t \sim iid\ N(0, Q_t), \tag{2.1}$$

$$y_t = H_t x_t + v_t, \qquad v_t \sim iid\ \varepsilon\text{-}contaminated\ N(0, R_t), \tag{2.2}$$

where the residuals $\{\omega_t\}$ and $\{v_t\}$ are mutually independent. Moreover, some initial conditions are required to be fulfilled. The state equation (2.1) describes the development of an n-dimensional state vector $x_t$ in time while the observations equation (2.2) assigns the 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 at time t. The $\varepsilon$-contamination of the residual $v_t$ means that its normal distribution $N(0,R_t)$ with acceptable variances is contaminated by a small fraction $\varepsilon$ (e.g. $\varepsilon = 0.05$) of a symmetric distribution with heavy tails which enables to model outliers in observed data. In the standard normal case without contamination the Kalman filter provides recursive formulas for the minimum variance state estimator $\hat{x}_t^t = E(x_t | Y^t)$ and its covariance matrix $P_t^t = E\,[(x_t - \hat{x}_t^t)(x_t - \hat{x}_t^t)' \,/\, Y^t]$ using all available information $Y^t = \{ y_0, y_1, \ldots, y_t\}$ at time t.

In the contaminated case (2.1), (2.2), the work [4] replaces these formulas by the approximative ones of the form

$$\hat{x}_t^t = \hat{x}_t^{t-1} + P_t^{t-1} H_t' [H_t P_t^{t-1} H_t' + R_t^{1/2} W_t^{-1} R_t^{1/2}]^{-1} (y_t - H_t \hat{x}_t^{t-1}), \tag{2.3}$$

$$P_t^t = P_t^{t-1} - P_t^{t-1} H_t' [H_t P_t^{t-1} H_t' + R_t^{1/2} W_t^{-1} R_t^{1/2}]^{-1} H_t P_t^{t-1}, \tag{2.4}$$

where the predictive values $\hat{x}_{t+1}^t = E(x_{t+1} / Y^t)$ and $P_{t+1}^t = E[(x_{t+1} - \hat{x}_{t+1}^t)(x_{t+1} - \hat{x}_{t+1}^t)'/Y^t]$ for time t+1 at time t are constructed as

$$\hat{x}_{t+1}^{t}=F_t\,\hat{x}_t^{t}, \tag{2.5}$$

$$P_{t+1}^{t} = F_t P_t^t F'_t + Q_t. \tag{2.6}$$

The symbol $R_t^{1/2}$ denotes the square root matrix of $R_t$ and $W_t$ = diag $\{w_{1t},\dots,w_{mt}\}$ is the mxm diagonal matrix with

$$w_{jt} = \frac{\Psi_j(s_{jt}-b_{jt}\,\hat{x}_t^{t-1})}{s_{jt}-b_{jt}\,\hat{x}_t^{t-1}}, \tag{2.7}$$

where $\Psi_1,\dots\dots,\Psi_m$ are suitable robustifying psi-functions and

$$B_t = \begin{pmatrix} b_{1t} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_{mt} \end{pmatrix} = R_t^{-1/2}H_t, \qquad s_t = \begin{pmatrix} s_{1t} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ s_{mt} \end{pmatrix} = R_t^{-1/2}y_t \tag{2.8}$$

($R_t^{-1/2}$ is the inverse matrix of $R_t^{1/2}$ and $b_{jt}$'s are the 1 x n rows of the matrix $B_t$).

The formulas (2.3) - (2.6) can be easily rewritten to the predictive form

$$\hat{x}_{t+1}^{t} = F_t\hat{x}_t^{t-1}+K_t(y_t-H_t\,\hat{x}_t^{t-1}), \tag{2.9}$$

$$P_{t+1}^{t} = F_tP_t^{t-1}F_t'+Q_t-K_tR_{et}K_t', \tag{2.10}$$

where $R_{et}$ is the innovation covariance matrix of the form

$$R_{et} = H_tP_t^{t-1}H_t'+R_t^{1/2}W_t^{-1}R_t^{1/2} \tag{2.11}$$

and $K_t$ is the Kalman gain matrix of the form

$$K_t = F_tP_t^{t-1}H_t'R_{et}^{-1}. \tag{2.12}$$

Specially, if m=1 so that the observations $y_t$ are scalar in the observation equation and, in addition, if we use the Huber's psi-function $\Psi_H$ of the form

$$\psi_H(z) = \begin{cases} z & , \ |z| \le c, \\ c \ sgn(z) & , \ |z| > c \end{cases} \tag{2.13}$$

(it can be shown that this choice of psi-function in the case of ε-contaminated normal distribution provides robust estimates that are optimal in the min-max sense) then the approximative formula (2.9) can be replaced by the non-approximative one of the form

$$\hat{x}^t_{t+1} = F_t\hat{x}^{t-1}_t + r_{et}r_t^{-1/2}\psi_H(r_{et}^{-1}r_t^{1/2}(y_t - h_t\hat{x}^{t-1}_t))K_t, \tag{2.14}$$

where

$$r_{et} = h_t P^{t-1}_t h_t' + r_t, \tag{2.15}$$

$$K_t = r_{et}^{-1} F_t P^{t-1}_t h_t'. \tag{2.16}$$

The covariance matrix $P^t_{t+1}$ is let in the form

$$P^t_{t+1} = F_t P^{t-1}_t F_t' + Q_t - r_{et} K_t K_t'. \tag{2.17}$$

***3. Square root robust Kalman filter.*** The square root formulation of the classical Kalman filter (see e.g. [5], [6], [7]) takes advantage of the matrix factorization that can be written for a positive semidefinite matrix X as

$$X = LDL', \tag{3.1}$$

where L is a lower triangular matrix with units on its main diagonal and D is a diagonal matrix.

In the case of the robust Kalman filter from Section 2, its square root formulation will maintain the matrices $P^{t-1}_t$ and $R_{et}$ in the factorized form (3.1), i.e.

$$P^{t-1}_t = L_{pt}D_{pt}L_{pt}', \qquad R_{et} = L_{et}D_{et}L_{et}', \tag{3.2}$$

where $L_{pt}$, $L_{et}$ are lower triangular matrices with units on the main diagonal and $D_{pt}$, $D_{et}$ are diagonal matrices.

The input for the corresponding square root robust algorithm at time t involves the matrices $F_t$, $H_t$, $R_t$, $Q_t$, the observation vector $y_t$ and the predictive value $\hat{x}_t^{t-1}$ for time t.

Let us construct matrices

$$U = \begin{pmatrix} I_m, & H_t L_{pt}, & 0 \\ 0, & F_t L_{pt}, & I_n \end{pmatrix} \tag{3.3}$$

of the dimension (m+n) x (m+2n) ($I_m$ is the mxm identity matrix) and

$$V = \begin{pmatrix} R_t^{1/2} W_t^{-1} R_t^{1/2}, & 0, & 0 \\ 0, & D_{pt}, & 0 \\ 0, & 0, & Q_t \end{pmatrix} \tag{3.4}$$

of the dimension (m+2n) x (m+2n). The substantial procedure of the algorithm consists in the following factorization

$$UVU' = LDL', \tag{3.5}$$

where the (m+n) x (m+2n) matrix L and (m+2n) x (m+2n) matrix D are required to be of the same type as in (3.1). It is not difficult to show that the matrices L and D will have the form

$$L = \begin{pmatrix} L_{et}, & 0, & 0 \\ K_t L_{et}, & L_{p,t+1}, & 0 \end{pmatrix} \tag{3.6}$$

and

$$D = \begin{pmatrix} D_{et}, & 0, & 0 \\ 0, & D_{p,t+1}, & 0 \\ 0, & 0, & D^* \end{pmatrix} \tag{3.7}$$

where $D^*$ may be an arbitrary diagonal nxn matrix.

Then the output of the algorithm at time t contains

$$\hat{x}_{t+1}^t = F_t \hat{x}_t^{t-1} + (K_t L_{et}) L_{et}^{-1} (y_t - H_t \hat{x}_t^{t-1}), \tag{3.8}$$

$$P_{t+1}^t = L_{p,t+1} D_{p,t+1} L_{p,t+1}', \tag{3.9}$$

where the matrices $L_{et}$, $K_t L_{et}$, $L_{p,t+1}$, $D_{p,t+1}$ are taken from (3.6) and (3.7) (moreover, the matrices $L_{p,t+1}$, $D_{p,t+1}$ form the input for time t+1).

All procedures of this square root robust algorithm can be performed efficiently in the framework of the parallel implementation. For instance, an array of (m+n) x (m+2n) parallel processors using the scan-with-add operation is suitable for the factorization (3.5) (see [7]).

***4. The case of scalar observations.*** If the observations are scalar (m=1) and, in addition, the Huber's psi-function (2.13) is chosen then the robust Kalman filter from Section 2 reduces to the form (2.14)-(2.17).

The corresponding square root formulation will be more simple using the matrices

$$U = \begin{pmatrix} 1, & h_t L_{pt}, & 0 \\ 0, & F_t L_{pt}, & I_n \end{pmatrix} \tag{4.1}$$

of the dimension (1+n) x (1+2n) and

$$V = \begin{pmatrix} r_t, & 0, & 0 \\ 0, & D_{pt}, & 0 \\ 0, & 0, & Q_t \end{pmatrix} \tag{4.2}$$

of the dimension (1+2n)x(1+2n). Then the matrices L,D in the factorization (3.5) will have the form

$$L = \begin{pmatrix} 1, & 0, & 0 \\ K_t, & L_{p,t+1}, & 0 \end{pmatrix} \tag{4.3}$$

and

$$D = \begin{pmatrix} r_{et}, & 0, & 0 \\ 0, & D_{p,t+1}, & 0 \\ 0, & 0, & D^* \end{pmatrix} \tag{4.4}$$

providing the values $r_{et}$, $K_t$ for (2.14) and $L_{p,t+1}$, $D_{p,t+1}$ for (3.9) (and for input at time t+1).

The suggested square root robust Kalman filter seems to be suitable for the practical treatment of systems with large numbers of state variables and with contaminated observations. Its implementation on parallel computers is the object of continuing work.

## REFERENCES


[1] AKL, S. G.: The Design and Analysis of Parallel Algorithms, Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[2] CIPRA, T.: Note on approximate non-Gaussian filtering with nonlinear observation relation, Comment. Math. Univ. Carolinae, 31 (1990), 601-605.

[3] CIPRA, T., MOTYOKOVÁ, I.: Study on Kalman filter in time series analysis, Comment. Math. Univ. Carolinae 28 (1987), 549-563.

[4] CIPRA, T., ROMERA, R.: Robust Kalman filter and its application in time series analysis, Kybernetika 27 (1991), n° 6, 481-494.

[5] ITZKOWITZ, H.R., BAHETI, R.S.: Demonstration of square root Kalman filter on Warp parallel computer, Proceedings of the 1989 American Control Conference.

[6] KAILATH, T.: State-space modelling: square root algorithms, in : Systems and Control Encyclopedia (M.G. Singh, ed.), Pergamon Press, New York 1987, 2618-2622.

[7] PALIS, M.A., KRECKER, D.K.: Parallel Kalman filtering on the Connection Machine, GE Internal Technical Report, Advanced Technology Laboratories, GE Aerospace, Moorestown, New Jersey 1989.



Tomás Cipra, Dept. of Statistics, Charles University of Prague, Sokolovská 83, 186 00 Praha 8, Czechoslovakia.

Rosario Romera, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain, and Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe, 28903 Madrid, Spain.

Asunción Rubio, Dpto. de Matemáticas, Universidad de Extremadura, 10071 Cáceres, Spain.