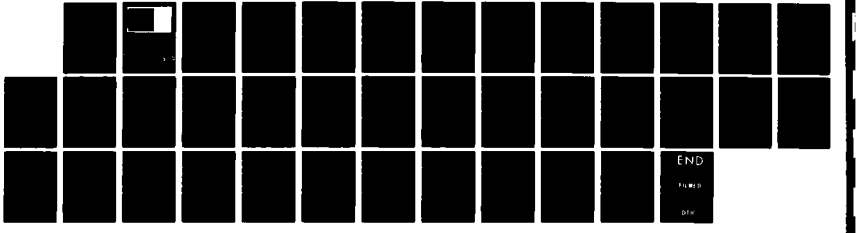
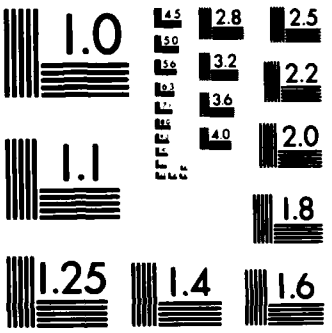
AD-A149 044 ROBUST KALMAN FILTERING AND ITS APPLICATIONS(U) 1/1
WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER
I GUTTMAN ET AL. OCT 84 MRC-TSR-2766 DAAG29-80-C-0041
UNCLASSIFIED F/G 12/1 NL





MRC Technical Summary Report # 2766

ROBUST KALMAN FILTERING AND ITS APPLICATIONS

Irwin Guttman and Daniel Peña

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

October 1984

(Received August 24, 1984)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

NSERC
Ottawa, Ontario
Canada

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ROBUST KALMAN FILTERING AND ITS APPLICATIONS

Irwin Guttman* and Daniel Peña**

Technical Summary Report #2766

October 1984

ABSTRACT

This paper presents a robust Kalman filtering algorithm that is obtained assuming a scale contaminated normal distribution for the noise of the measurement equation. The mixture of normals obtained as a posterior distribution is approximated at each stage by a normal distribution with the same mean and variance. The resulting algorithm is simple, has a straightforward interpretation and seems to provide useful robust estimators in several statistical problems that are briefly reviewed. Originator-supplied

AMS (MOS) Subject Classifications: 62M20, 60G35, 62F35

Key Words: Kalman Filter, Robustness, mixtures of normals, and

Work Unit Number 4 - Statistics and Probability

* Department of Statistics, University of Toronto, Toronto, Ontario M5S 1A1.

** ETSII, Universidad Politécnica de Madrid, Spain.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. The first author acknowledges support from the NSERC of Canada under Grant No. A8743 and the second author to the United States-Spanish Joint Committee for Education and Cultural Affairs.

SIGNIFICANCE AND EXPLANATION

Real data sets almost always contain outlying (extreme) observations and outliers are particularly damaging in on line control situations in which the data is processed recursively. Thus, an extremely bad value can distort the whole mechanism of control and make the process very unstable.

In this paper, we offer a relatively simple model and obtain a procedure to deal with the above problem. To represent the appearance of bad observations, a scale contaminated normal distribution has been assumed for the measurement error.

In fact, we have shown in this paper, how a Bayesian approach allows the development of a simple recursive estimation algorithm that has the desired property of "filtering" bad (i.e., extreme) observations. Indeed, extreme values are downweighted by their posterior probability of being spurious, and the estimates of parameters are updated, recursively, accordingly.

Finally, we apply our model to the case of exponential smoothing with contaminated error, and show that the parameter estimates obtained from the resulting algorithm are a weighted combination of certain 2^n smoothing schemes. The application of the procedure to a broad range of statistical estimation problems is briefly discussed.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

ROBUST KALMAN FILTERING AND ITS APPLICATIONS

Irwin Guttman* and Daniel Peña**

1. Introduction and Summary

Kalman (1960) introduced a method of updating knowledge about the "state" of process parameters, say $\underline{\theta}$, at time t, using a least squares procedure. This method, now known as Kalman filtering, has wide applicability, from on-line process control in industry to applications in economics. Kalman's results are reproducible using a Bayes approach with normal theory, conditional on known values of variances and co-variances involved.

One aspect of the filtering process is that it is sensitive to extreme observations. Indeed, one or more wild observations can make the Kalman filter unstable. This a well recognized result in both the Statistical and Engineering literatures, and is discussed in the use of a Kalman-type filtering scheme that takes into account the possibility of spuriously generated observations giving rise to extreme observations. This filtering scheme automatically examines the possibility that the current observation is spurious, and if the evidence points to this, downweights that observation in the filter, and does the opposite for seemingly "good" (i.e., non spurious) observations.

We develop this filter in Section 3, after reviewing the standard Kalman filter in Section 2. Section 4 sketches a number of applications, and the use of our filter in these areas. Finally, Section 5 provides some discussion of our results.

* Department of Statistics, University of Toronto, Toronto, Ontario M5S 1A1.

** ETSII, Universidad Politécnica de Madrid, Spain.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. The first author acknowledges support from the NSERC of Canada under Grant No. A8743 and the second author to the United States-Spanish Joint Committee for Education and Cultural Affairs.

2. The Standard Kalman Filter

The results due to Kalman (1960) may be derived and approached from the Bayesian point of view. Suppose that we observe, at time t, a response vector $\underset{\sim}{y}_t$, say of order $(p\times 1)$, and that this random response vector is such that

$$\underset{\sim}{y}_t = A_t\underset{\sim}{\theta}_t + \underset{\sim}{\varepsilon}_y,\ \underset{\sim}{\varepsilon}_y \sim N_{\underset{\sim}{\varepsilon}_y}(\underset{\sim}{0},C) \tag{2.1}$$

where A_t is a $(p\times r)$ matrix of known coefficients, $\underset{\sim}{\theta}_t$ is a $(r\times 1)$ vector of unknown process parameters, and C is a $(p\times p)$ positive definite matrix, assumed known. The vector $\underset{\sim}{\theta}_t$ is referred to as the current (i.e., at time t) state of the process parameters $\underset{\sim}{\theta}$. As t varies, the states are also assumed to have a linear structure, viz, for given $\underset{\sim}{\theta}_{t-1}$, V,

$$\underset{\sim}{\theta}_t = \Omega_t\underset{\sim}{\theta}_{t-1} + \varepsilon_\theta,\ \underset{\sim}{\varepsilon}_\theta \sim N_{\underset{\sim}{\varepsilon}_\theta}(\underset{\sim}{0},V) \tag{2.2}$$

where Ω_t is a $r\times r$ known matrix, and V is a $(r\times r)$ positive definite matrix. Finally, it is also assumed that we have prior information about $\underset{\sim}{\theta}_{t-1}$, given $\underset{\sim}{y}_{t-1}$, $\underset{\sim}{y}_{t-2},\ldots$. This assumption is sometimes referred to as the Inductive Hypothesis, and says that prior to observing $\underset{\sim}{y}_t$, and given $\underset{\sim}{y}_{t-1},\underset{\sim}{y}_{t-2},\ldots$, that the distribution of $\underset{\sim}{\theta}_{t-1}$, given $\underset{\sim}{y}_{t-1},\ldots$ has structure

$$\underset{\sim}{\theta}_{t-1} = \underset{\sim}{\mu}_{t-1} + \underset{\sim}{\varepsilon}_{\theta_{t-1}},\ \underset{\sim}{\varepsilon}_{\theta_{t-1}} \sim N(\underset{\sim}{0},V_{t-1}) \tag{2.3}$$

where V_{t-1} is a $(r\times r)$ positive definite matrix. We shall see below that $\underset{\sim}{\mu}_{t-1}$ is a function of the y's.

Now using the above assumptions, we may repidly deduce that the prior for $\underset{\sim}{\theta}_t$, given $\underset{\sim}{y}_{t-1}$ is

$$N_{\underset{\sim}{\theta}_t}(\underset{\sim}{\mu}_{t:t-1},V_{t:t-1}) \tag{2.4}$$

where

$$\begin{aligned}\underset{\sim}{\mu}_{t:t-1} &= \Omega_t\,\underset{\sim}{\mu}_{t-1}\\ V_{t:t-1} &= V + \Omega_t V_{t-1}\Omega_t'\ .\end{aligned} \tag{2.4a}$$

(The subscript "t:t-1" refers to the fact that we are at time t and have observed $\underset{\sim}{y}_{t-1}$.) The quick and easy way to see (2.4) - (2.4a), is as follows: We have, for given $\underset{\sim}{\theta}_{t-1}$,

$$\underset{\sim}{\theta}_t = \Omega_t \underset{\sim}{\theta}_{t-1} + \underset{\sim}{\varepsilon}_{\theta_t} \tag{2.5}$$

from (2.2). But $\underset{\sim}{\theta}_{t-1}$ is itself random, so that using (2.3) in (2.5), we have

$$\begin{aligned} \underset{\sim}{\theta}_t &= \Omega_t[\mu_{t-1} + \underset{\sim}{\varepsilon}_{\theta_{t-1}}] + \underset{\sim}{\varepsilon}_{\theta_t} \\ &= \Omega_t \mu_{t-1} + \Omega_t \underset{\sim}{\varepsilon}_{\theta_{t-1}} + \underset{\sim}{\varepsilon}_{\theta_t} \cdot \end{aligned} \tag{2.6}$$

Because of the nomality assumptions, and assuming independence of $\underset{\sim}{\theta}_{t-1}$, $\underset{\sim}{\theta}_t$, we have that: Given $\underset{\sim}{y}_{t-1}$,

$$\underset{\sim}{\theta}_t \sim N_{\underset{\sim}{\theta}_t}(\Omega_t \mu_{t-1}, \Omega_t V_{t-1} \Omega_t' + V) \quad , \tag{2.7}$$

as claimed in (2.4) - (2.4a). It is convenient to write the result (2.4) or (2.7) as

$$\underset{\sim}{\theta}_t = \mu_{t:t-1} + \underset{\sim}{\varepsilon}_{t:t-1} \quad , \quad \underset{\sim}{\varepsilon}_{t:t-1} \sim N(\underset{\sim}{0}, V_{t:t-1}) \quad . \tag{2.8}$$

We may now also deduce, using assumptions (2.1), (2.2) and (2.3), with the result (2.4), the predictive distribution of the yet unobserved $\underset{\sim}{y}_t$, given $\underset{\sim}{y}_{t-1}$. For from (2.1), we have

$$\underset{\sim}{y}_t = A_t \underset{\sim}{\theta}_t + \underset{\sim}{\varepsilon}_y \tag{2.9}$$

where $\underset{\sim}{\varepsilon}_y \sim N(\underset{\sim}{0}, C)$, and from (2.8), given $\underset{\sim}{y}_{t-1}$, we have that

$$\begin{aligned} \underset{\sim}{y}_t &= A_t(\mu_{t:t-1} + \underset{\sim}{\varepsilon}_{t:t-1}) + \underset{\sim}{\varepsilon}_y \\ &= A_t \mu_{t:t-1} + A_t \underset{\sim}{\varepsilon}_{t:t-1} + \underset{\sim}{\varepsilon}_y \end{aligned} \tag{2.10}$$

which implies, for given $\underset{\sim}{y}_{t-1}$, that

$$\underset{\sim}{y}_t \sim N(A_t \mu_{t:t-1}, A_t V_{t:t-1} A_t' + C) \quad . \tag{2.11}$$

We let the predictive variance of (2.11) be denoted by R_t, that is,

$$R_t = C + A_t V_{t:t-1} A_t' \quad . \tag{2.11a}$$

Now the results (2.7) and (2.11) give the distributions of $\underset{\sim}{\theta}_t$ and $\underset{\sim}{y}_t$ <u>before</u> observing $\underset{\sim}{y}_t$, namely $p(\underset{\sim}{\theta}_t|\underset{\sim}{y}_{t-1})$ and $p(\underset{\sim}{y}_t|\underset{\sim}{y}_{t-1})$, respectively. Now when we observe $\underset{\sim}{y}_t$, we are in the position to deduce the posterior for $\underset{\sim}{\theta}_t$, given $\underset{\sim}{y}_t$, for are in the position of having (2.1) as the sampling distribution of $\underset{\sim}{y}_t$, given $\underset{\sim}{\theta}_t$, and (2.4) as the prior for $\underset{\sim}{\theta}_t$, (given $\underset{\sim}{y}_{t-1}$ etc). We remark that once this posterior is obtained, it plays the role of (2.3) for the next stage, to be discussed below. Now using Bayes' Theorem with (2.1) and (2.4) as the necessary ingredients yields: Given $\underset{\sim}{y}_t$,

$$\underset{\sim}{\theta}_t \sim N(\underset{\sim}{\mu}_{t:t} \ , \ V_{t:t}) \tag{2.12}$$

where

$$\underset{\sim}{\mu}_{t:t} = \underset{\sim}{\mu}_{t:t-1} + V_{t:t-1}A_t'M_t^{-1}(\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1}) \tag{2.12a}$$

and

$$V_{t:t}^{-1} = A_t'C^{-1}A_t + V_{t:t-1}^{-1} \tag{2.12b}$$

so that, as is easily verified,

$$V_{t:t} = V_{t:t-1} - V_{t:t-1}A_t'M_t^{-1}A_tV_{t:t-1} \tag{2.12c}$$

where, we recall that M_t is given at (2.11a).

The derivation of these results is given in Appendix I. Notice the updating pattern from $\underset{\sim}{\mu}_{t:t-1}$, the prior mean, to $\underset{\sim}{\mu}_{t:t}$ contained in (2.12a). Indeed, (2.12a) implies that the current information $\underset{\sim}{\mu}_{t:t}$ about the process parameters $\underset{\sim}{\theta}_t$, given $\underset{\sim}{y}_t$, is the prior information of $\underset{\sim}{\theta}_t$ given $\underset{\sim}{y}_{t-1}$, $\underset{\sim}{\mu}_{t:t-1}$ plus an updating term, obtained by "filtering" the deviation of $\underset{\sim}{y}_t$ from its predictive expectation - see (2.11) - by use of the matrix

$$KF_t = V_{t:t-1}A'M_t^{-1} \ . \tag{2.13}$$

Indeed the matrix KF_t of (2.13) is referred to as the Kalman Gain matrix, and we note that an alternative form - see Appendix I - is

$$KF_t = V_{t:t}A'C^{-1} \ .$$

Note too the update of $V_{t:t}$ contained in (2.12b) - for example, we update $V_{t:t-1}^{-1}$, the precision of the prior of $\underset{\sim}{\theta}_t$, given $\underset{\sim}{y}_{t-1}$, by adding the precision $A_t'C^{-1}A_t$ of the regression process parameter $\underset{\sim}{\theta}_t$ of (2.1) to obtain $V_{t:t}^{-1}$, etc. To enter the next stage, we replace (2.3) with (2.12) by setting $\underset{\sim}{\mu}_{t:t} = \underset{\sim}{\mu}_t$ and $V_{t:t} = V_t$, and make the obvious modifications in (2.1) and (2.2), and repeat the process.

To start the Kalman filter, prior conditions for the state vector $\underset{\sim}{\theta}$ must be made, say $\underset{\sim}{\mu}_0$ for its expectation, and the matrix V_0 for its variance - covariance structure. Once declared, the estimate for $\underset{\sim}{\theta}_1$, the state of nature for the process that yields $\underset{\sim}{y}_1$, is

$$\hat{\underset{\sim}{\theta}}_{1:0} = \Omega_1 \underset{\sim}{\mu}_0 = \underset{\sim}{\mu}_{1:0} \tag{2.14}$$

and its variance - covariance matrix is

$$V_{1:0} = \Omega_1 V_0 \Omega_1' + V \quad . \tag{2.14a}$$

The forecast of the new observation $\underset{\sim}{Y}_1$ (see (2.11)) is

$$\hat{\underset{\sim}{Y}}_1 = A_1 \mu_{1:0} \tag{2.15}$$

with variance - covariance (see (2.11) - (2.11a))

$$M_1 = C + A_1 V_{1:0} A_1' \quad . \tag{2.15a}$$

When $\underset{\sim}{Y}_1$ is actually observed, we update as follows - the expected state of nature goes from $\mu_{1:0}$ to $\mu_{1:1}$, where

$$\mu_{1:1} = \mu_{1:0} + V_{1:0} A_1' M_1^{-1} (\underset{\sim}{Y}_1 - A_1 \mu_{1:0}) \tag{2.16}$$

and the affiliated variance - covariance is updated to $V_{1:1}$, where

$$V_{1:1}^{-1} = V_{1:0}^{-1} + A_1' C^{-1} A_1 = (V_{1:0} - V_{1:0} A_1' M_1^{-1} A_1 V_{1:0})^{-1} \quad . \tag{2.16a}$$

As indicated befoe, we now set

$$\mu_1 = \mu_{1:1} \quad \text{and} \quad V_1 = V_{1:1} \tag{2.17}$$

and inquire about $\underset{\sim}{\theta}_2$, prior and posterior to seeing $\underset{\sim}{Y}_2$, etc. . We have

$$\underset{\sim}{Y}_2 = A_2 \underset{\sim}{\theta}_2 + \underset{\sim}{\varepsilon}_y, \ \underset{\sim}{\varepsilon}_y \sim N(\underset{\sim}{0}, C) \quad , \tag{2.18}$$

with

$$\underset{\sim}{\theta}_2 = \Omega_2 \underset{\sim}{\theta}_1 + \underset{\sim}{\varepsilon}_\theta, \ \underset{\sim}{\varepsilon}_\theta \sim N(\underset{\sim}{0}, V) \quad , \tag{2.18a}$$

while

$$\underset{\sim}{\theta}_1 = \mu_1 + \underset{\sim}{\varepsilon}_{\theta_1} \quad , \quad \underset{\sim}{\varepsilon}_{\theta_1} \sim N(\underset{\sim}{0}, V_1) \quad . \tag{2.18b}$$

The last two statements may be combined so that we have, given $\underset{\sim}{Y}_1$,

$$\underset{\sim}{\theta}_2 \sim N(\Omega_2 \mu_1, \ \Omega_2 V_1 \Omega_2' + V)$$

or (2.18c)

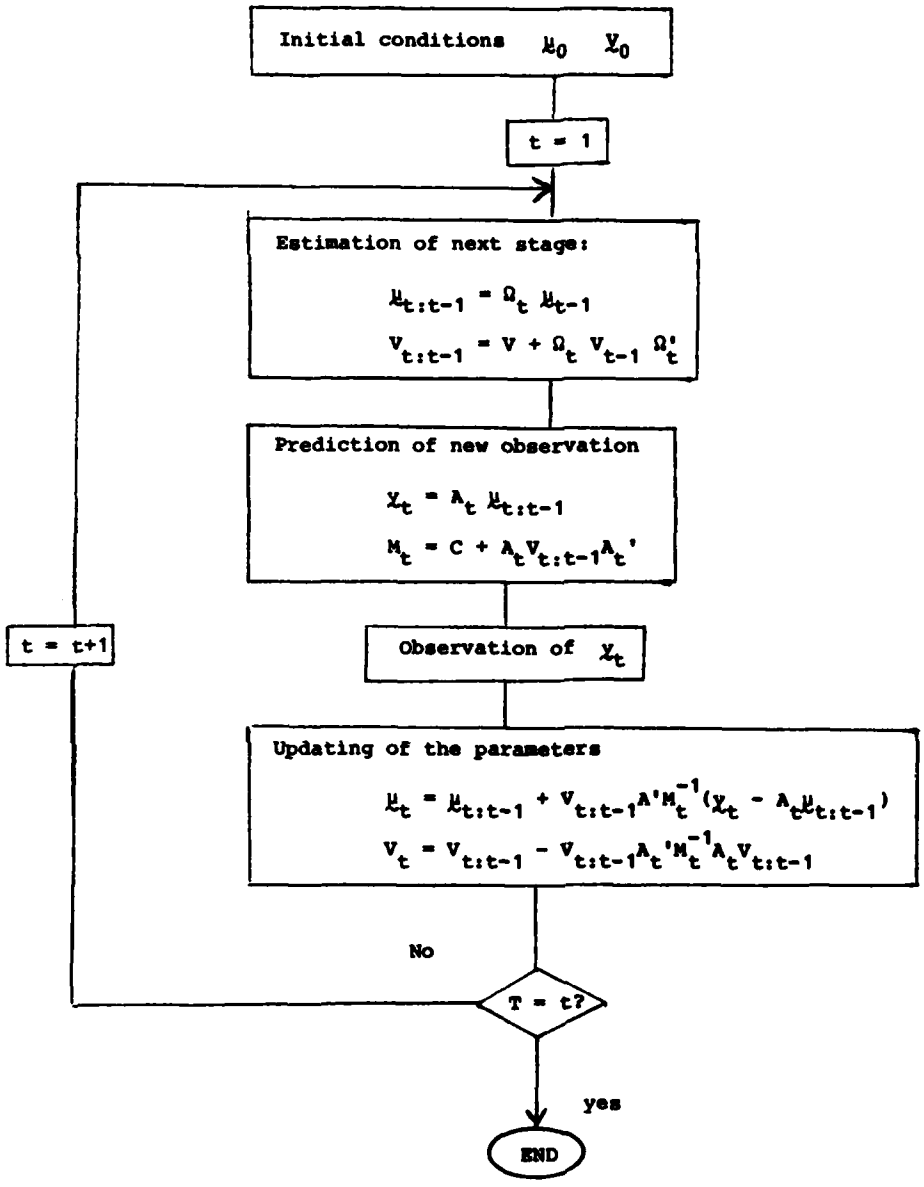
$$\underset{\sim}{\theta}_2 \sim N(\mu_{2:1} \ , \ V_{2:1}) \quad .$$

The distribution (2.18) is now used with (2.18c) to find the posterior of $\underset{\sim}{\theta}_2$, given $\underset{\sim}{Y}_2$, which is

$$\underset{\sim}{\theta}_2 \sim N(\mu_{2:2} \ , \ V_{2:2}) \tag{2.19}$$

where $\mu_{2:2}$ is given by (2.12a) with $t = 2$ and $V_{2:2}$ is determined by (2.12b), etc. , and the loop continues in the same way for $t = 3,4,\ldots$. This is pictoralized in Figure 2.1 for the general case of having reach state $t - 1$ and observed $\underset{\sim}{Y}_{t-1}$, etc. .

Figure 2.1


Initial conditions μ_0 Σ_0
t = 1
Estimation of next stage:
$\mu_{t:t-1} = \Omega_t \mu_{t-1}$
$V_{t:t-1} = V + \Omega_t V_{t-1} \Omega_t'$
Prediction of new observation
$\chi_t = A_t \mu_{t:t-1}$
$M_t = C + A_t V_{t:t-1} A_t'$
Observation of χ_t
Updating of the parameters
$\mu_t = \mu_{t:t-1} + V_{t:t-1} A' M_t^{-1} (\chi_t - A_t \mu_{t:t-1})$
$V_t = V_{t:t-1} - V_{t:t-1} A_t' M_t^{-1} A_t V_{t:t-1}$
t = t+1
No
T = t?
yes
END

There are several comments to be made about the foregoing Kalman filter. Essentially, the whole process is a least squares procedure, as the reader will not doubt have known or guessed. Least squares estimates are well known to be non-robust to outlying observations (see Andrews et al (1972)), and in the Kalman filter case this could make the whole procedure unstable, with devastating consequences in some situations, such as line-process control of mass produced items. Then too, when $p \geq 3$, it is well known that least squares estimators are not admissible (see Stein (1956)).

Finally, the assumption (2.1) that implies that ε_y's come from the same distribution is much too strong in practice. Much evidence exists that shows that sets of data almost always contain a small proportion of observations that have been spuriously generated (i.e., not in the manner intended) giving rise to extreme or outlying observations (see the general discussion in the paper by Box and Tiao (1968) and Guttman (1973)). For these reasons, we replace the assumption (2.1) by a more realistic sampling model, and investigate what form the ensuing "Kalman filtering" process will take in the following section.

3. A Robust Algorithm

3.1. A Different Sampling Model.

As indicated in the previous section, the oft-made assumption (2.1) is highly suspect, and skepticism about this assumption often points to the question of the possible effect of outliers on procedures developed using (2.1) in general, and for us, in particular on Kalman Filtering. Outliers are feared mainly due to the fact that they may have been generated spuriously, thus biasing results. This is a well recognized concern in the engineering literature, and for example, the papers of Alspach and Sorenson (1971), Masreliez (1975), Masreliez and Martin (1977) and Tsai and Kurz (1983) use a different approach then ours to meet this problem. Of course, the problem of how to deal with outliers in other situations and the general problem of "robustness" of various procedures is the focus of much of the current Statistical literature - see the references cited in Section 2, for example.

Because of the above general concern, it is desired to establish procedures that are robust to outliers in that they accomodate the appearance of abberant observations appropriately - roughly speaking, giving small weight to observations that seem spuriously generated, and large weight to seemingly "good" observations.

A spuriousness that gives rise to outliers often means that the error distributions involved have tails heavier than those of the normal distribution, we will generalize below the method of accomodating outliers used by Box and Tiao (1968), and replace the assumption (2.1) by the so-called Scaled-Contaminated Model (SCM). This model was introduced into statistical practice by Jeffreys (1961) and has been used by Box and Tiao (1968) to robustify estimation in the standard linear model, by Abraham and Box (1979) to accomodate outliers in time series, etc. Indeed, Cheng and Box (1980), have shown that the SCM model represents a sensible modeling in many situations where spuriousness is feared.

The SCM model, simply stated is that

$$\underset{\sim}{Y}_t = A_t \underset{\sim}{\theta}_t + \underset{\sim}{\varepsilon}_y \tag{3.1}$$

where

$$\underset{\sim}{\varepsilon}_y \sim \alpha_1 N(\underset{\sim}{0}, C_1) + \alpha_2 N(\underset{\sim}{0}, C_2) \quad .$$

In (3.1), we assume that the known constants α_1, α_2 are such that $\alpha_2 = 1-\alpha_1$, and that $\alpha_2 \in (0,.15)$, as is common in most applications. Further, we also assume C_1 and C_2 are known, and such that by any measure, C_2 is larger than C_1. For example, it could be that (for σ_y^2 a known scalar),

$$C_1 = \sigma_y^2 I \quad , \quad C_2 = k^2\sigma_y^2 I, \; k^2 > 1 \quad . \tag{3.2}$$

The prescription (3.1) says that with small probability α_2, $\underset{\sim}{\chi}_t$ is generated spuriously from $N(A_t \underset{\sim}{\theta}_t, C_2)$, etc. .

In addition to the assumption (3.1), we also, in this section, make the assumptions (2.2) and (2.3) of the previous section, and inquire into the question of how the 3 assumptions (3.1), (2.2) and (2.3) affect the updating procedure discussed in the introductory section. For convenience we list the assumptions used in this section at the point:

(i) See (3.1) - (3.2)

(ii) $\underset{\sim}{\theta}_t = \Omega_t \underset{\sim}{\theta}_{t-1} + \underset{\sim}{\varepsilon}_\theta \quad , \quad \underset{\sim}{\varepsilon}_\theta \sim N(\underset{\sim}{0}, V)$ (3.3)

(iii) $\underset{\sim}{\theta}_{t-1} = \underset{\sim}{\mu}_{t-1} + \underset{\sim}{\varepsilon}_{\underset{\sim}{\theta}_{t-1}} \quad , \quad \underset{\sim}{\varepsilon}_{\underset{\sim}{\theta}_{t-1}} \sim N(\underset{\sim}{0}, V_{t-1})$, for given $\underset{\sim}{\chi}_{t-1}$.

The assumptions (ii) and (iii) of course, give rise to the result (2.4), namely that $\underset{\sim}{\theta}_t$, given $\underset{\sim}{\chi}_{t-1}$ has distribution (the prior of $\underset{\sim}{\theta}_t$) (2.4), viz

$$N_{\underset{\sim}{\theta}_t}(\underset{\sim}{\mu}_{t:t-1}, V_{t:t-1}) \tag{3.4}$$

where

$$\underset{\sim}{\mu}_{t:t-1} = \Omega_t \underset{\sim}{\mu}_{t-1}, \quad V_{t:t-1} = V + \Omega_t V_{t-1} \Omega_t' \quad . \tag{3.4a}$$

We can now determine the predictive distribution $h(\cdot|\underset{\sim}{\chi}_{t-1})$, say, of $\underset{\sim}{\chi}_t$, given $\underset{\sim}{\chi}_{t-1}$. Formally, this is defined as

$$h(\underset{\sim}{\chi}_t|\underset{\sim}{\chi}_{t-1}) = \int_{\underset{\sim}{\theta}_t} f(\underset{\sim}{\chi}_t|\underset{\sim}{\theta}_t) p(\underset{\sim}{\theta}_t|\underset{\sim}{\chi}_{t-1}) d\underset{\sim}{\theta}_t \tag{3.5}$$

and here, f is dictated by (3.1), while p is obtained from (2.4). The result of doing the integration (3.5), as proved in Appendix II, is as follows:

$$h(\underset{\sim}{\chi}_t|\underset{\sim}{\chi}_{t-1}) \sim \alpha_1 N_{\underset{\sim}{\chi}_t}(A_t \underset{\sim}{\mu}_{t:t-1}, M_{t,1}) + \alpha_2 N_{\underset{\sim}{\chi}_t}(A_t \underset{\sim}{\mu}_{t:t-1}, M_{t,2}) \tag{3.6}$$

with

$$M_{t,i} = C_i + A_t V_{t:t-1} A_t' \quad , \quad i = 1,2 \quad . \tag{3.6a}$$

We note that

$$E(\underset{\sim}{Y}_t|\underset{\sim}{Y}_{t-1}) = A_t \underset{\sim}{\mu}_{t:t-1}, \; V(\underset{\sim}{Y}_t|\underset{\sim}{Y}_{t-1}) = A_t V_{t:t-1} A_t' + \sum_{j=1}^{2} \alpha_j C_j \quad . \tag{3.6b}$$

We note that we may write

$$V(\underset{\sim}{Y}_t|\underset{\sim}{Y}_{t-1}) = \alpha_1 M_{t,1} + \alpha_2 M_{t,2} \quad . \tag{3.6c}$$

3.2. The resulting recursive algorithm.

The results (3.4) and (3.6) are statements that can be made before seeing $\underset{\sim}{Y}_t$. Once $\underset{\sim}{Y}_t$ is observed, we are in the position of being able to find the (posterior) distribtuion of $\underset{\sim}{\theta}_t$, given $\underset{\sim}{Y}_t$. For the prior of $\underset{\sim}{\theta}_t$, we have the distribution specified by (3.4), while the (sampling) distribution of $\underset{\sim}{\theta}_t$ is given by (3.1). The result of using Bayes' Theorem with these ingredients, as proved in Appendix II, is that the posterior of $\underset{\sim}{\theta}_t$, given $\underset{\sim}{Y}_t$ is such that

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{Y}_t) \sim \sum_{i=1}^{2} \alpha_{t,i} N_{\underset{\sim}{\theta}_t}(\underset{\sim}{\mu}_{t:t}^{(i)} \; , \; V_{t:t}^{(i)}) \tag{3.7}$$

where

$$\underset{\sim}{\mu}_{t:t}^{(i)} = \underset{\sim}{\mu}_{t:t-1} + V_{t:t-1} A_t' M_{t,i}^{-1} (\underset{\sim}{Y}_t - A_t \underset{\sim}{\mu}_{t:t-1}) \tag{3.7a}$$

$$\begin{aligned} V_{t:t}^{(i)} &= (V_{t:t-1}^{-1} + A_t' C_i^{-1} A_t)^{-1} \\ &= V_{t:t-1} - V_{t:t-1} A_t' M_{t,i}^{-1} A_t V_{t:t-1} \end{aligned} \tag{3.7b}$$

and where

$$\alpha_{t,i} = \frac{\alpha_i f(\underset{\sim}{Y}_t|A_t \underset{\sim}{\mu}_{t:t-1}, M_{t,i})}{\sum_{j=1}^{2} \alpha_j f(\underset{\sim}{Y}_t|A_t \underset{\sim}{\mu}_{t:t-1}, M_{t,j})} \tag{3.7c}$$

with f denoting the density of the Normal multivariate distribution, so that, in general

$$f(\underset{\sim}{\chi}|\underset{\sim}{\eta}, M) = (2\pi)^{-p/2} |M|^{-1/2} \exp - \tfrac{1}{2}(\underset{\sim}{\chi} - \underset{\sim}{\eta})' M^{-1} (\underset{\sim}{\chi} - \underset{\sim}{\eta}) \quad . \tag{3.7d}$$

The reader will not doubt recognize that the denominator of (3.7c) is, on using (3.7d), the predictive density h of $\underset{\sim}{Y}_t$, given $\underset{\sim}{Y}_{t-1}$, stated in (3.6). Indeed, using (3.7d) in (3.7c) yields

$$\alpha_{t,1} = \left[1 + \frac{\alpha_2}{\alpha_1}\left(\frac{|M_{t,1}|}{|M_{t,2}|}\right)^{1/2} \exp\{\tfrac{1}{2}(\underset{\sim}{y}_t - A_t\,\underset{\sim}{\mu}_{t:t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1})(\underset{\sim}{y}_t - A_t\,\underset{\sim}{\mu}_{t:t-1})\}\right]^{-1} \tag{3.7e}$$

$$\alpha_{t,2} = 1 - \alpha_{t,1} \quad .$$

From (3.7) we easily find (see Appendix II for proofs)

$$\begin{aligned} E(\underset{\sim}{\theta}_t|\underset{\sim}{y}_t) &= \underset{\sim}{\mu}_{t:t-1} + V_{t:t-1}A_t'[\alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1}](\underset{\sim}{y}_t - A_t\,\underset{\sim}{\mu}_{t:t-1}) \\ &= \underset{\sim}{\mu}_{t:t} \\ V(\underset{\sim}{\theta}_t|\underset{\sim}{y}_t) &= V_{t:t-1} - V_{t:t-1}A_t'B_tA_tV_{t:t-1} = V_{t:t} \end{aligned} \tag{3.8}$$

where

$$B_t = \alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1} - \alpha_{t,1}\alpha_{t,2}(M_{t,1}^{-1} - M_{t,2}^{-1})(\underset{\sim}{y}_t - A_t\,\underset{\sim}{\mu}_{t:t-1})(\underset{\sim}{y}_t - A_t\,\underset{\sim}{\mu}_{t:t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1}) \quad . \tag{3.8a}$$

The quantities $\alpha_{t,1}$ and $\alpha_{t,2} = (1-\alpha_{t,1})$ are the posterior probabilities that $\underset{\sim}{y}_t$ has come from the intended source (i.e., $N(A_t\,\underset{\sim}{\theta}_t, C_1)$) and the spurious source (i.e., $N(A_t\,\underset{\sim}{\theta}_t, C_2)$), respectively. The posterior expectation is made up of two parts (c.f. with (2.12a)): the prior expectation $\underset{\sim}{\mu}_{t:t-1}$ of $\underset{\sim}{\theta}_t$, given $\underset{\sim}{y}_{t-1}$ (see (3.4)) and a deviation of $\underset{\sim}{y}_t$ from its (marginal) predictive expectation $A_t\,\underset{\sim}{\mu}_{t:t-1}$ (see (3.6b)), but this time the filtering (gain) matrix is the weighted sum of two Kalman gain matrices, where the weights are the estimates $\alpha_{t,i}$ just commented on. Put another way, the gain matrix involves the weighted sum of the predictive precisions that would be involved if sampling was from either $N(A_t\,\underset{\sim}{\theta}_t, C_1)$ or $N(A_t\,\underset{\sim}{\theta}_t, C_2)$, with weights that are the probabilities that $\underset{\sim}{y}_t$ was so sampled. For more introspection about the updated variance - covariance $V_{t:t}$, we first invite the reader to inspect the updated $V_{t,t}^{(i)}$'s, and to acquaint themselves with the details of how these are used by reading the proof of (3.8a) in Appendix II.

We remark that to continue this procedure, we now use the following scheme

Set $\underset{\sim}{\mu}_t = \underset{\sim}{\mu}_{t:t}$ (see (3.8)) and $V_t = V_{t:t}$ (see (3.8a)) . (3.9)

Let $$p(\underset{\sim}{\theta}_t|\underset{\sim}{\chi}_t) \simeq N(\underset{\sim}{\mu}_t, V_t) \quad . \tag{3.10}$$

We may now enter the next stage - we replace (iii) of (3.3) with:

Given $\underset{\sim}{\chi}_t$, $$\underset{\sim}{\theta}_t = \underset{\sim}{\mu}_t + \underset{\sim}{\varepsilon}_{\theta_t} \sim N(\underset{\sim}{\theta}, V_t) \tag{3.11}$$

and making the obvious modifications in (3.1) and (ii) of (3.3) - t is replaced by (t+1) - we repeat the above process.

As in the previous section, to start this Kalman Filter, we must state prior conditions for the state vector $\underset{\sim}{\theta}$, say $\underset{\sim}{\mu}_0$ for its expectation, and the matrix V_0 for its variance - covariance structure. Once these are declared, then using (3.4) we have that the estimate for $\underset{\sim}{\theta}_1$ before seeing $\underset{\sim}{\chi}_1$ is

$$\hat{\underset{\sim}{\theta}}_{1:0} = \Omega_1 \underset{\sim}{\mu}_0 = \underset{\sim}{\mu}_{1:0} \tag{3.12}$$

and the associated variance - covariance matrix is

$$V_{1:0} = V + \Omega_1 V_0 \Omega_1' \quad . \tag{3.12a}$$

The forecast of the new observation $\underset{\sim}{\chi}_1$ is (see (1.5b))

$$\hat{\underset{\sim}{\chi}}_1 = A_1 \underset{\sim}{\mu}_{1:0} = A_1 \Omega_1 \underset{\sim}{\mu}_0 \tag{3.13}$$

with associated variance - covariance matrix (see (3.6b))

$$\sum_{j=1}^{2} \alpha_j C_j + A_1 V_{1:0} A_1' \tag{3.13a}$$

(c.f. with (2.15a)).

When $\underset{\sim}{\chi}_1$ is actually observed, we update as follows - the expected state of nature goes from $\underset{\sim}{\mu}_{1:0}$ to $\underset{\sim}{\mu}_{1:1}$, where (see (3.8))

$$\underset{\sim}{\mu}_{1:1} = \underset{\sim}{\mu}_{1:0} + V_{1:0} A_1' [\alpha_{t,1} M_{1,1}^{-1} + \alpha_{t,2} M_{1,2}^{-1}] (\underset{\sim}{\chi}_1 - A_1 \underset{\sim}{\mu}_{1:0}) \tag{3.14}$$

where the matrices $M_{1,i}$ are defined in (3.6a), and where the $\alpha_{1,i}$ are given by (3.7c). The associated variance - covariance is (see (3.8a))

$$V_{1:1} = V_{1:0} - V_{1:0} A_1' B_1 A_1 V_{1:0} \tag{3.14a}$$

where

$$\begin{aligned} B_1 &= \alpha_{1,1} M_{1,1}^{-1} + \alpha_{1,2} M_{1,2}^{-1} \\ &\quad - \alpha_{1,1}\alpha_{1,2} (M_{1,1}^{-1} - M_{1,2}^{-1}) (\underset{\sim}{\chi}_1 - A_1 \underset{\sim}{\mu}_{1:0}) (\underset{\sim}{\chi}_1 - A_1 \underset{\sim}{\mu}_{1:0})' (M_{1,1}^{-1} - M_{1,2}^{-1}) \quad . \end{aligned} \tag{3.14b}$$

As mentioned in this section, we now set

$$\mu_1 = \mu_{1:1} \quad \text{and} \quad V_1 = V_{1:1} \tag{3.15}$$

and take, for the distribution of $\underset{\sim}{\theta}_1$, given $\underset{\sim}{\chi}_1$, $p(\underset{\sim}{\theta}_1|\underset{\sim}{\chi}_1)$ given by

$$p(\underset{\sim}{\theta}_1|\underset{\sim}{\chi}_1) = f(\underset{\sim}{\theta}_1|\mu_1,V_1) \tag{3.16}$$

that is, $\underset{\sim}{\theta}_1 \sim N(\mu_1,V_1)$, given $\underset{\sim}{\chi}_1$. This may be combined with (ii) of (3.3) for $t = 2$ to yield

$$p(\underset{\sim}{\theta}_2|\underset{\sim}{\chi}_1) = f(\underset{\sim}{\theta}_2|\mu_{2:1},V_{2:1}) \tag{3.17}$$

where

$$\mu_{2:1} = \Omega_2 \mu_1 \quad , \quad V_{2:1} = V + \Omega_2 V_1 \Omega_2' \quad . \tag{3.17a}$$

Now the distribution of $\underset{\sim}{\chi}_2$, given $\underset{\sim}{\theta}_2$, is of course specified by (3.1), that is

$$p(\underset{\sim}{\chi}_2|\underset{\sim}{\theta}_2,C_1,C_2,\alpha_1,\alpha_2) = \alpha_1 f(\underset{\sim}{\chi}_2|A_2\ \underset{\sim}{\theta}_2,C_1) + \alpha_2 f(\underset{\sim}{\chi}_2|A_2\ \underset{\sim}{\theta}_2,C_2) \tag{3.18}$$

and (3.18) may be combined with (3.17) to produce the posterior of $\underset{\sim}{\theta}_2$, given $\underset{\sim}{\chi}_2$ which is

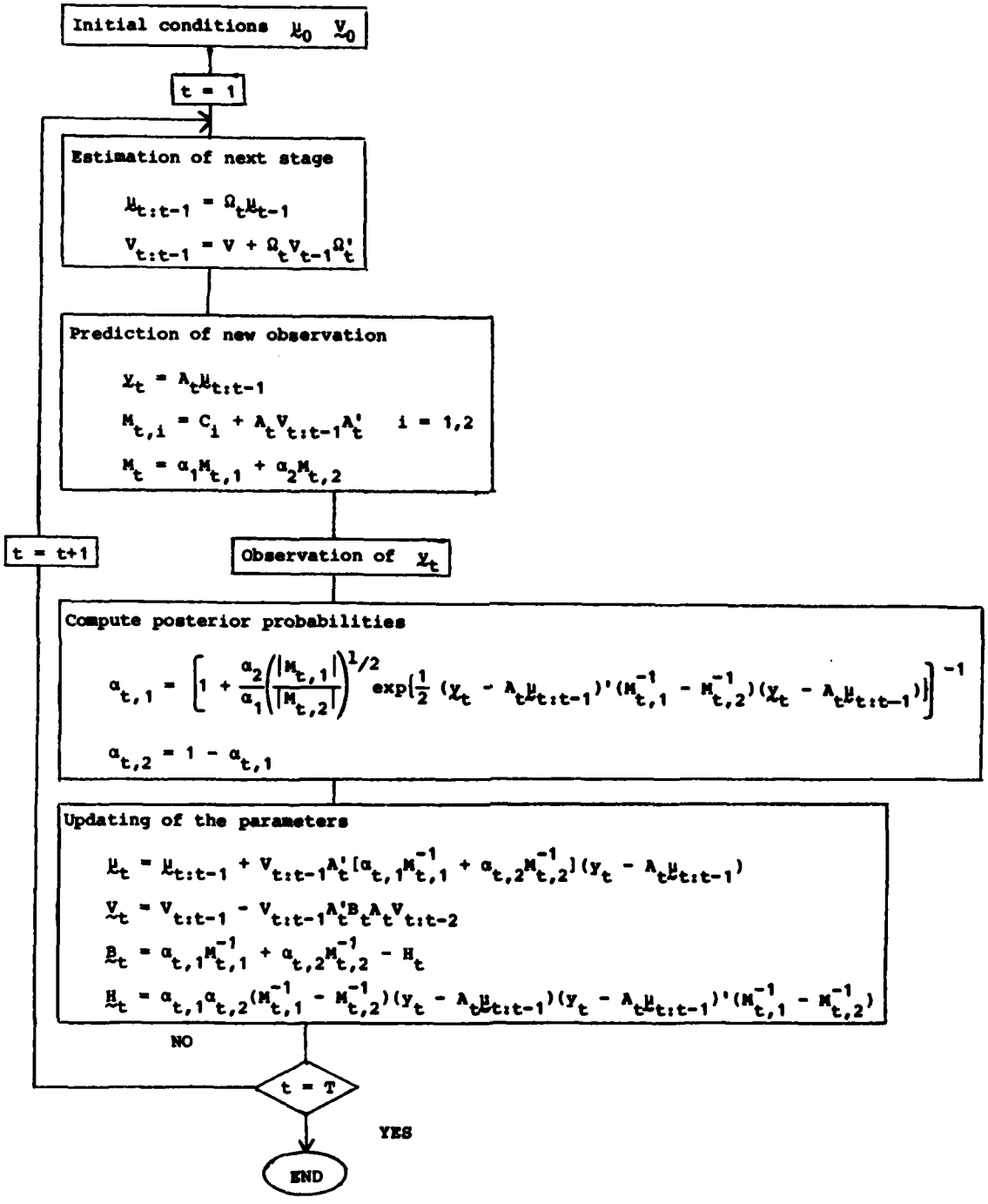
$$p(\underset{\sim}{\theta}_2|\underset{\sim}{\chi}_2) = \sum_{i=1}^{2} \alpha_{2,i} f(\underset{\sim}{\theta}_2|\mu_{2:2}^{(i)},V_{2:2}^{(i)}) \tag{3.19}$$

where $\mu_{2:2}^{(i)}$ and $V_{2:2}^{(i)}$ are found from (3.7a) - (3.7b). The probability $\alpha_{2,1}$ is of course (see (3.7c))

$$\alpha_{2,1} = \frac{\alpha_1 f(\underset{\sim}{\chi}_2|A_2\ \mu_{2:1},M_{2,1})}{\sum_{j=1}^{2} \alpha_i f(\underset{\sim}{\chi}_2|A_2\ \mu_{2:1},M_{2,i})} \tag{3.20}$$

and $\alpha_{2,2} = 1-\alpha_{2,1}$, and the loop continues in this way for $t = 3,4,\ldots$ etc. This is pictoralized in Figure 3.1 for the general case.

Figure 3.1


Initial conditions $\underline{\mu}_0$ $\underline{V}_0$
t = 1
Estimation of next stage
$\underline{\mu}_{t:t-1} = \Omega_t \underline{\mu}_{t-1}$
$V_{t:t-1} = V + \Omega_t V_{t-1} \Omega_t'$
Prediction of new observation
$\underline{y}_t = A_t \underline{\mu}_{t:t-1}$
$M_{t,i} = C_i + A_t V_{t:t-1} A_t'$ i = 1,2
$M_t = \alpha_1 M_{t,1} + \alpha_2 M_{t,2}$
t = t+1
Observation of $\underline{y}_t$
Compute posterior probabilities
$\alpha_{t,1} = \left[1 + \frac{\alpha_2}{\alpha_1}\left(\frac{|M_{t,1}|}{|M_{t,2}|}\right)^{1/2} \exp\{\frac{1}{2}(\underline{y}_t - A_t\underline{\mu}_{t:t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1})(\underline{y}_t - A_t\underline{\mu}_{t:t-1})\}\right]^{-1}$
$\alpha_{t,2} = 1 - \alpha_{t,1}$
Updating of the parameters
$\underline{\mu}_t = \underline{\mu}_{t:t-1} + V_{t:t-1} A_t' [\alpha_{t,1} M_{t,1}^{-1} + \alpha_{t,2} M_{t,2}^{-1}](y_t - A_t \underline{\mu}_{t:t-1})$
$\underline{V}_t = V_{t:t-1} - V_{t:t-1} A_t' B_t A_t V_{t:t-2}$
$\underline{B}_t = \alpha_{t,1} M_{t,1}^{-1} + \alpha_{t,2} M_{t,2}^{-1} - H_t$
$\underline{H}_t = \alpha_{t,1}\alpha_{t,2}(M_{t,1}^{-1} - M_{t,2}^{-1})(y_t - A_t\underline{\mu}_{t:t-1})(y_t - A_t\underline{\mu}_{t:t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1})$
NO
t = T
YES
END

4. Scope of the New Filter

The previous section develops a filter which may be used in many various situations. To illustrate this, we discuss several instances where the new filter of Section 3 may be applied.

4.1. Bayesian Forecasting

One of the most well known applications of the Kalman filter in Statistics is to that of Bayesian forecasting, as developed by Harrison and Stevens (1976). These authors used the standard formulation developed in the engineering literature for the linear control problem that we have presented in our Section 2, calling it the dynamic linear model. They showed that most statistical models can be considered under this general framework, which thus allows an unified approach, namely, by using the Kalman filter for recursive estimation of the parameters. Harrison and Stevens (1976) then applied these ideas to Bayesian forecasting.

As our robust filtering includes the standard algorithm as a particular case, the robust version of the algorithm we have developed can be applied to any of the forecasting models discussed by these authors.

To illustrate the behavior of the algorithm we shall discuss its applications to one of the most widely used model for forecasting, the so-called Steady Model. This model is a particular case of the general formulation with $A_t = 1$, $\Omega_t = 1$ and with y_t a scalar. Then, $\mu_{t:t-1} = \mu_{t-1}$ and $V_{t:t-1} = V + V_{t-1}$. The updating equations for the standard recursive estimation of this model are

$$\mu_t = \mu_{t-1} + \left(\frac{V+V_{t-1}}{C+V+V_{t-1}}\right)(y_t - \mu_{t-1}) \tag{4.1}$$

$$V_t = \frac{C(V+V_{t-1})}{C+V+V_{t-1}} \quad . \tag{4.1a}$$

After some iteration the system will reach a stable state in which $V_t = V_{t-1} = a$. Calling $\theta = C[c+v+a]^{-1}$, we see that (4.1) can be written as

$$\mu_t = \theta\, \mu_{t-1} + (1-\theta)y_t \tag{4.2}$$

and applying it recursively and assuming $\mu_0 = 0$, we obtain

$$\mu_t = \sum_{i=1}^{t-1} \theta^i(1-\theta)y_{t-i} \quad . \tag{4.3}$$

We have, then, from (4.3) the result that the steady model is equivalent to the IMA (1,1) model, which is equivalent to exponential smoothing.

To apply the robust Kalman filtering algorithm we have developed, we assumed $c_1 = \sigma^2$, $c_2 = k^2\sigma^2$; then

$$\mu_t = \mu_{t-1} + (V + V_{t-1})\left[\frac{\alpha_{t,1}}{\sigma^2+V+V_{t-1}} + \frac{\alpha_{t,2}}{k^2\sigma^2+V+V_{t-1}}\right](y_t - \mu_{t-1})$$

and assuming as before a stable state with $V_t = V_{t-1} = a$, then calling $\theta_1 = \sigma^2(v^2 + v + a)^{-1}$ and $\theta_2 = (k^2\sigma^2)(k^2\sigma^2 + v + a)^{-1}$, we have

$$\mu_t = [\alpha_{t,1}\theta_1 + \alpha_{t,2}\theta_2)\mu_{t-1} + [(1-\theta_1)\alpha_{t,1} + (1-\theta_2)\alpha_{t,2}]y_t \quad .$$

This expression shows that μ_t is obtained as a weighted combination of μ_{t-1} and y_t, as in the standard case, but now the weights are changing in every iteration and depend on the probabilities $\alpha_{t,1}$ and $\alpha_{t,2}$. Setting $\theta_t = \alpha_{t1}\theta_1 + \alpha_{t2}\theta_2$, we can write

$$\mu_t = \theta_t\mu_{t-1} + (1-\theta_t)y_t \tag{4.4}$$

and now the expression of μ_t as a function of the observations is

$$\mu_t = \sum_{i=0}^{n} \frac{(1-\theta_{t-i})}{\theta_{t-i}} y_{t-i}\left(\prod_{j=0}^{i} \theta_{t-j}\right) \quad . \tag{4.5}$$

To interpret this equation let us denote by $w(r;i,j,k,..h)$ the probability that r observations came from the "spurious" or "bad" population at times i, j, k,..h, where there are r symbols i, j,..,h. Then

$$w(r;i,j,k,..h) = \Big(\prod_{\substack{t=1 \\ t\neq i,j,..,h}}^{n} \alpha_{t,1}\Big)\alpha_{i2}\alpha_{j2}\alpha_{k2}\cdots\alpha_{h2} \tag{4.6}$$

and let us call $\hat{\mu}_t(r;i,j,k,..h)$ the smoothing associated with this combination of observations (see (4.3)). This smoothing is the result of assuming

$$\mu_t = \theta_1\mu_{t-1} + (1-\theta_1)y_t \tag{4.7}$$

for observations y_t, $t = 1,\ldots n$ but $t \neq i,j,k,..h$, and

$$\mu_t = \theta_2\mu_{t-1} + (1-\theta_2)y_t \tag{4.7a}$$

for $t = i,j,k,..h$. Then, it is straightforward to show that

$$\mu_t = \Sigma w(r;i,j,..h)\hat{\mu}(r;i,j,..h) \qquad (4.8)$$

where here the summation is over all the 2^t possible combinations of observations chosen from $y_1,\ldots,y_t$. The main advantage of our algorithm (4.5) is that the 2^t exponential smoothing factors are the result of the recursive relationship (4.4) and are not computed separately but globally and here in a very direct way.

4.2. Robust Linear Regression

As mentioned previously, Box and Tiao (1968) assumed a scale contaminated normal distribution for the noise of a regression model and used Bayesian estimation methods to obtain a "robust" estimation procedure that downweights suspicious observations in the linear model. Additionally, Chen and Box (1979) showed that, given appropriate values to the parameters of the noise distributions, the Box-Tiao weights can reproduce functions for downweighting residuals using M-estimators that have been proposed on empirical grounds by Andrews et al (1972). Although this approach provides a general way to deal with outliers in the linear model, the computations needed are cumbersome, because the posterior distribution of the parameter vector is a weighted average of 2^n posterior distributions. Box and Tiao (1968) suggested that one need compute only the first few leading terms, but with a large set of data the computations are still heavy, and there are no clear rules about how many terms we would need to obtain a proper approximation. Little (1983) has explored the relationship between the Bayesian weights and some influence measures proposed for the linear model and has used this relationship to suggest an algorithm to determine which of the weights that matter, and their subsequent computation. Here again, however, the computations are still heavy.

We will show in this section how we can compute the posterior mean in a simple way using the robust algorithm we have suggested.

The regression can be written as a state space model as

$$\begin{aligned} y_t &= x_t' \beta_t + \epsilon_t \\ \beta_t &= \beta_{t-1} = \beta \end{aligned} \qquad (4.9)$$

where y_t is now scalar in the observation equation and the regression parameter vector is

constant over time. Assuming that the noise ε_t has a scale contaminated normal distribution

$$\varepsilon_t \sim (1-\alpha)N(0,\sigma^2) + \alpha N(0,k^2\sigma^2) \quad , \quad k^2 > 1 \quad , \tag{4.10}$$

for fixed values of α, k^2 and σ^2, we then have a particular case of our previous formulation. In practice, however, neither of these three parameters are known. Although we can assume α and k^2 known and make a sensitivity study afterwards (see Box and Tiao (1968)), the problem of estimating σ^2 remains. There are two possible solutions.

The first is to compute a robust estimator of $\hat{\sigma}^2$ and use this value in the computations. The second is to iterate in the estimation of the robust value until convergence. Starting with a preliminary robust estimate of σ^2, say $\hat{\sigma}^2_{(1)}$ obtained from the residuals of the least square fit that leads to the estimate $\hat{\beta}_{(1)}$, we compute an estimate $\hat{\beta}_{(2)}$ using the robust Kalman filtering algorithm with the value $\hat{\sigma}^2_{(1)}$ for the variance. Then, we compute the new residuals and a new robust estimator $\hat{\sigma}^2_{(2)}$ for the variance. Using this estimator a new application of the algorithm is made which provides a new estimator $\hat{\beta}_{(3)}$ and, consequently, a new set of residuals. From them a new robust variance estimator $\hat{\sigma}^2_{(3)}$ is built and the process is repeated until convergence.

In this particular case, writing $V_t = \sigma^2 \dot{V}_t$, and $m_{it} = \sigma^2 \dot{m}_{it}$ the computations needed are:

$$\dot{m}_{1t} = (1 + \underset{\sim}{x}_t' V_{t-1} \underset{\sim}{x}_t) = (1 + h_{(i)}) \tag{4.11}$$

$$\dot{m}_{2t} = (k^2 + h_{(i)}) \quad , \tag{4.12}$$

$$\alpha_{t,1} = \left[1 + \left(\frac{\alpha}{1-\alpha}\right)\sqrt{\frac{\dot{m}_{1t}}{\dot{m}_{2t}}}\exp\frac{1}{2}\left\{\left(\frac{y_t - \underset{\sim}{x}_t'\hat{\underset{\sim}{\beta}}_{t-1}}{\sigma_{(t)}}\right)^2\left(\frac{1}{\dot{m}_{1t}} - \frac{1}{\dot{m}_{2t}}\right)\right\}\right]^{-1} \tag{4.13}$$

$$\hat{\underset{\sim}{\beta}}_{(t)} = \hat{\underset{\sim}{\beta}}_{(t-1)} + \dot{\underset{\sim}{v}}_{t-1}\,\underset{\sim}{x}_t'\left(\frac{\alpha_{t,1}}{\dot{m}_{1t}} + \frac{\alpha_{t,2}}{\dot{m}_{2t}}\right)e_{(t)} \tag{4.14}$$

$$e_{(t)} = y_t - \underset{\sim}{x}_t'\hat{\underset{\sim}{\beta}}_{(t-1)} \tag{4.15}$$

$$\dot{V}_t = \dot{V}_{t-1} - \dot{V}_{t-1}\,\underset{\sim}{x}_t\, b_t\, \underset{\sim}{x}_t'\,\dot{V}_{t-1} \tag{4.16}$$

where $B_t = b_t$ is given by

$$b_t = \frac{\alpha_{t,1}}{\dot{m}_{1t}} + \frac{\alpha_{2,t}}{\dot{m}_{2t}} - \alpha_{t,1}\alpha_{t,2}\left(\frac{1}{\dot{m}_{1t}} - \frac{1}{\dot{m}_{2t}}\right)^2\left(\frac{e_{(t)}}{\hat{\sigma}_{(t)}}\right)^2 \quad . \tag{4.17}$$

A simple alternative for a robust estimator of scale is the median absolute deviation from the sample median $\tilde{y}$, given by

$$\text{MAD} = \underset{t}{\text{median}}\{|y_t - \tilde{y}|\} \quad .$$

This estimator is known to be more efficient for the contaminated normal distribution than the sample variance, but is biased. A robust unbiased estimator of σ^2 has been provided by Tukey et al (1977) and is given by

$$\hat{\sigma}_r^2 = (.64)^{-1}\underset{i}{\text{median}}\{e_i\} \tag{4.18}$$

where the e_i are the residuals from the least squares fit. The final covariance matrix for the parameters will be $V_t = \hat{\sigma}_r^2 \dot{V}_t$, where $\dot{V}_t$ is computed using (4.16).

Equation (4.14) shows that the estimate $\hat{\beta}_{(t)}$ is a linear combination of estimators that would have been obtained assuming that ε_t is distributed as $N(0,\sigma^2)$ and $N(0,k^2\sigma^2)$, respectively, with weights that are the posterior probabilities $\alpha_{t,i}$'s, discussed before. As for the variance, instead of the least squares expression:

$$\dot{V}_t^{-1} = (\dot{V}_{t-1}^{-1} + \underset{\sim}{x}_t \underset{\sim}{x}_t')^{-1} \tag{4.19}$$

where $V_t^{-1} = X'X$, now equation (4.11) imposes an adaptive formulation:

$$\dot{V}_t^{-1} = (\dot{V}_{t-1}^{-1} + a_t \underset{\sim}{x}_t \underset{\sim}{x}_t')^{-1} \tag{4.20}$$

where $a_t = b_t/(1 - h_{(i)}b_t)$ depends on the posterior probabilities of the observation being an outlier. For instance, it is straightforward to see that if $\alpha_{t,1} = 1$, $a_t = 1$, and if $\alpha_{t,2} = 1$, $a_t = k^{-2}$.

This algorithm provides a useful computational device to estimate $\underset{\sim}{\beta}$ when the set of data is large. However, when the sample is small there are two problems that make this approximation a crude one. First, the distribution for $\underset{\sim}{\beta}$ at each state is not normal, but a t-distribution when σ^2 is unknown. Second, with a small set of data the ordering of the observations can influence the results. These two problems are not important for a large data set because (1) the t-distribution, when the number of degrees of freedom is large, is very well approximated by the normal and (2) the order of the observations will

not affect much as long as the number of observations is large and stable state has been reached.

The above is, in contrast with the standard Kalman Filter in which σ^2 is not needed, while here the variances m_{it} are required to compute the posterior probabilities that an observation has been spuriously generated, and for the variance of the estimation itself.

4.3. Autoregressive Time Series Estimation

Assuming in the previous case that $x_t' = (y_{t-1},\ldots,y_{t-p})$, the model reduces to the state space estimation of an autoregressive process. Abraham and Box (1979) have studied inference in this type of model when there is a small probability that "bad" observations occur. The exact Bayesian solution is again difficult to compute because it involves, as in the regression case, the computation of 2^n distributions.

Although in this case the previous approach can be used to obtain a robust estimate of the parameters when all the data has been collected, the algorithm can also be used as a robust procedure for on-line estimation when the observations are received sequentially. In this latter case however, a way to compute an estimate of σ^2 given the observed data is needed because at every instant t only observations $y_1,\ldots,y_t$, will be available.

To solution we suggest is to start with a robust estimate of the variance, and update this estimation when the new residual $y_t - \hat{y}_t$ is computed. The algorithm will be defined by the same equations but now (4.13) and (4.17) are:

$$\alpha_{1t} = \left[1 + \left(\frac{\alpha}{1-\alpha}\right)\sqrt{\frac{\dot{m}_{1t}}{\dot{m}_{2t}}}\exp\frac{1}{2}\left\{\left(\frac{y_t - \underset{\sim}{x}_t'\hat{\beta}_{t-1}}{\hat{\sigma}_t}\right)^2\left(\frac{1}{\dot{m}_{1t}} - \frac{1}{\dot{m}_{2t}}\right)\right\}\right]^{-1} \tag{4.21}$$

$$b_t = \frac{\alpha_{1t}}{\dot{m}_{1t}} + \frac{\alpha_{2t}}{\dot{m}_{2t}} - \alpha_{1t}\alpha_{2t}\left(\frac{1}{\dot{m}_{1t}} - \frac{1}{\dot{m}_{2t}}\right)^2\left(\frac{e_{(t)}}{\hat{\sigma}_t}\right)^2 \tag{4.22}$$

with

$$e_{(t)} = y_t - \hat{\beta}_{t1}y_{t-1} - \ldots - \hat{\beta}_{tp}y_{t-p} \tag{4.23}$$

$$\hat{\sigma}_t = \underset{t}{\text{Median}}\{e_{(t)}\}/.64 \quad . \tag{4.24}$$

4.4. Multivariate estimation

In the previous examples the observed vector $\underset{\sim}{y}_t$ was a scalar but the algorithm is also very simple when $\underset{\sim}{y}_t$ is a vector. As an example, we will revise only the multivariate regression model. The standard formulation is

$$\begin{aligned} \underset{\sim}{y}_t &= \underset{\sim}{H}_t \underset{\sim}{\beta}_t + \underset{\sim}{\varepsilon}_t \\ \underset{\sim}{\beta}_t &= \underset{\sim}{\beta}_{t-1} = \underset{\sim}{\beta} \end{aligned} \tag{4.25}$$

with

$$\underset{\sim}{H}_t = \begin{bmatrix} \underset{\sim}{x}_t' & \underset{\sim}{0}' \cdots \underset{\sim}{0}' \\ & \ddots \\ \underset{\sim}{0}' \cdots & \underset{\sim}{x}_t' \end{bmatrix} \qquad \underset{\sim}{\beta}_t = \begin{bmatrix} \beta_{1t} \\ \vdots \\ \beta_{kt} \end{bmatrix} \tag{4.26}$$

where $\underset{\sim}{x}_t'$ is the $1 \times p$ vector of explanatory variables and $\underset{\sim}{\beta}_{it}$ is the $p \times 1$ vector of parameters linked to component y_{it} of $\underset{\sim}{y}_t$. The noise $\underset{\sim}{\varepsilon}_t$ is assumed to have a mixed distribution and we can impose different structures depending on the particular problem on hand. For instance if

$$C_1 = \begin{bmatrix} \sigma_{11} & \sigma_{12} \cdots \sigma_{1k} \\ \sigma_{k1} \cdots & \sigma_{kk} \end{bmatrix} \tag{4.27}$$

the presence of a multivariate outlier can be modelled assuming that they come from a distribution with covariance matrix

$$C_2 = \begin{bmatrix} k_1^2\sigma_{11} & & \\ & \ddots & \\ & & k_k^2\sigma_{kk} \end{bmatrix} \quad , \quad k_j^2 > 1 \quad , \tag{4.28}$$

which implies that the components of an outlier are unrelated. This is referred to as external structure for spurious observations. Another example is $C_2 = h^2 C_1$, $h^2 > 1$, so that the components of outlying observations are related in the same way as observations from the good or intended source, but variances of outlying components are larger. This is referred to as internal multivariate structure for spurious observations.

5. Conclusions

Real data sets almost always contain outlying (extreme) observations, and outliers are particularly damaging in on line control situations in which the data is processed recursively. Thus, an extremely bad value can distort the whole mechanism of control and make the process very unstable. In Industrial practice, for instance, all types of ad hoc procedures have been developed to cope with this situation (see Mah and Tamhane (1982) and Crowe et al (1983)), but a general methodology is needed.

In this paper, we offer a relatively simple model and obtain a procedure to deal with the above problem. To represent the appearance of bad observations, a scale contaminated normal distribution has been assumed for the measurement error. We have chosen this model because other authors have demonstrated that its use provides sensible solutions to other statistical problems, for example, in linear model estimation.

In fact, we have shown in this paper, how a Bayesian approach allows the development of a simple recursive estimation algorithm that has the desired property of "filtering" bad (i.e., extreme) observations. Indeed, extreme values are downweighted by their posterior probability of being spurious, and the estimates of the parameters are updated, recursively, accordingly.

Finally, we apply our model to the case of exponential smoothing with contaminated error, and show that the parameter estimates obtained from the resulting algorithm are a weighted combination of certain 2^n smoothing schemes. The application of this procedure to a broad range of statistical estimation problems is briefly discussed.

6. Acknowledgements

This research was sponsored by the United States Army under Contract No. DAAG29-80-C-0041, and by NSERC of Canada under Grant No. A8743, and, additionally by the United States-Spanish Joint Committee for Education and Cultural Affairs. The authors are very grateful to Tom Leonard, University of Wisconsin-Madison, for valuable discussions and comments on this paper.

Appendix I

As mentioned in Section 2, we wish, in the Appendix, to prove the results (2.12). We have that the prior for $\underset{\sim}{\theta}_t$ is $N_{\underset{\sim}{\theta}_t}(\underset{\sim}{\mu}_{t:t-1}, V_{t:t-1})$, where $\underset{\sim}{\mu}_{t:t-1}$ and $V_{t:t-1}$ are defined at (2.4a), and from (2.1), we have that the density of $\underset{\sim}{Y}_t$ is $N_{\underset{\sim}{Y}_t}(A_t \underset{\sim}{\theta}_t, C)$. Hence, the posterior of $\underset{\sim}{\theta}_t$, given $\underset{\sim}{Y}_t$, is such that

$$p(\underset{\sim}{Y}_t|\underset{\sim}{\theta}_t) \propto \exp -\frac{1}{2}\{[\underset{\sim}{\theta}_t - \underset{\sim}{\mu}_{t:t-1}]' V_{t:t-1}^{-1}[\underset{\sim}{\theta}_t - \underset{\sim}{\mu}_{t:t-1}] + [\underset{\sim}{Y}_t - A_t \underset{\sim}{\theta}_t]' C^{-1}[\underset{\sim}{Y}_t - A_t \underset{\sim}{\theta}_t]\} \quad . \tag{AI.1}$$

Now the expression in the braces of the exponent in AI.1 is easily seen to be

$$\underset{\sim}{\theta}_t'[V_{t:t-1}^{-1} + A_t'C^{-1}A_t]\underset{\sim}{\theta}_t - 2\underset{\sim}{\theta}_t'[V_{t:t-1}^{-1}\underset{\sim}{\mu}_{t:t-1} + A_t' C^{-1}\underset{\sim}{Y}_t] + \text{const.}^{(1)} \tag{AI.2}$$

which in turn, on completing the square, has the form

$$[\underset{\sim}{\theta}_t - V_{t:t}\underset{\sim}{b}_t]' V_{t:t}^{-1}[\underset{\sim}{\theta}_t - V_{t:t}\underset{\sim}{b}_t] + \text{const.}^{(1)} - \text{const.}^{(2)} \tag{AI.3}$$

where $\text{const}^{(j)}$ are functions of A_t, C, $V_{t:t-1}$ and $\underset{\sim}{Y}_t$, all of which are assumed known. Also, we have written $V_{t:t}^{-1}$ for the matrix of the quadratic form in (AI.2) and (AI.3), that is

$$V_{t:t}^{-1} = V_{t:t-1}^{-1} + A_t' C^{-1} A_t \tag{AI.4}$$

and we have, clearly, that $\underset{\sim}{b}_t$ is given by

$$\underset{\sim}{b}_t = V_{t:t-1}^{-1}\underset{\sim}{\mu}_{t:t-1} + A_t' C^{-1}\underset{\sim}{Y}_t \quad . \tag{AI.4a}$$

Hence, $p(\underset{\sim}{Y}_t|\underset{\sim}{\theta}_t)$ is such that

$$p(\underset{\sim}{Y}_t|\underset{\sim}{\theta}_t) \propto \exp -\tfrac{1}{2}[\underset{\sim}{\theta}_t - V_{t:t}\underset{\sim}{\beta}_t]' V_{t:t}^{-1}[\underset{\sim}{\theta}_t - V_{t:t}\underset{\sim}{b}_t] \tag{AI.5}$$

which is to say, that a posteriori, $\underset{\sim}{\theta}_t \sim N(\underset{\sim}{\mu}_{t:t}, V_{t:t})$, where

$$\underset{\sim}{\mu}_{t:t} = V_{t:t}\underset{\sim}{b}_t \tag{AI.6}$$

and $V_{t:t}$ is such that its inverse is as defined in (AI.4). It remains to show that $V_{t:t}$ is given by (2.12c) and that (AI.6) may be rewritten as in (2.12a). For the latter, we have, from (AI.6), on using (AI.4a), that

$$\underset{\sim}{\mu}_{t:t} = V_{t:t}[V_{t:t-1}^{-1}\underset{\sim}{\mu}_{t:t-1} + A_t' C^{-1}\underset{\sim}{Y}_t] \tag{AI.7}$$

or

$$\underset{\sim}{\mu}_{t:t} = V_{t:t}[(V_{t:t-1}^{-1} + A_t' C^{-1}A_t)\underset{\sim}{\mu}_{t:t-1} + A_t'C^{-1}(\underset{\sim}{Y}_t - A_t \underset{\sim}{\mu}_{t:t-1})] \tag{AI.7a}$$

and using (AI.4) we have

$$\mu_{t:t} = V_{t:t}[V_{t:t}^{-1}\mu_{t:t-1} + A_t' C^{-1}(\chi_t - A_t \mu_{t:t-1})] \qquad \text{(AI.7b)}$$

so that

$$\mu_{t:t} = \mu_{t:t-1} + V_{t:t} A_t' C^{-1}(\chi_t - A_t \mu_{t:t-1}) \quad . \qquad \text{(AI.7c)}$$

To show the equivalence of (AI.7c) with (2.12a), we must show that

$$V_{t:t}A_t'C^{-1} = V_{t:t-1}A_t'M_t^{-1} \qquad \text{(AI.8)}$$

where M_t is as defined at (2.11a). But

$$V_{t:t} = V_{t:t} V_{t:t-1}^{-1} V_{t:t-1} \qquad \text{(AI.9)}$$

and again using (AI.4), we have

$$V_{t:t} = V_{t:t}[V_{t:t}^{-1} - A_t' C^{-1} A_t]V_{t:t-1} \qquad \text{(AI.9a)}$$

or

$$V_{t:t} = V_{t:t-1} - V_{t:t} A_t' C^{-1} A_t V_{t:t-1} \qquad \text{(AI.9b)}$$

which is to say

$$V_{t:t} + V_{t:t} A_t' C^{-1} A_t V_{t:t-1} = V_{t:t-1} \quad . \qquad \text{(AI.9c)}$$

Post multiplying by A_t' on both sides of (AI.9c), we may write

$$V_{t:t} A_t' C^{-1}C + V_{t:t} A_t' C^{-1}A_tV_{t:t-1}A_t' = V_{t:t-1}A_t' \qquad \text{(AI.9d)}$$

or

$$V_{t:t} A_t' C^{-1}(C + A_tV_{t:t-1}A_t') = V_{t:t-1} A_t' \qquad \text{(AI.9e)}$$

and from (2.11a)

$$V_{t:t} A_t' C^{-1} = V_{t:t-1} A_t' M_t^{-1} \quad , \qquad \text{(AI.9f)}$$

and (AI.8), and hence (2.12a), is now demonstrated.

Finally, we wish to derive the result (2.12c). We have from (2.12b) that

$$V_{t:t}^{-1} = V_{t:t-1}^{-1}[I + V_{t:t-1} A_t' C^{-1} A_t] \quad . \qquad \text{(AI.10)}$$

Hence

$$V_{t:t} = [I + V_{t:t-1} A_t' C^{-1} A_t]^{-1}V_{t:t-1} \qquad \text{(AI.10a)}$$

and using a well known identity for inverses of matrices of he form $I + EF$, viz

$$(I+EF)^{-1} = I - E(I+FE)^{-1}F \qquad \text{(AI.11)}$$

with E identified as $V_{t:t-1} A_t' C^{-1}$ and F as A_t, we find

$$V_{t:t} = \{I - V_{t:t-1} A_t' C^{-1}[I + A_tV_{t:t-1}A_t'C^{-1}]^{-1}A_t\}V_{t:t-1} \qquad \text{(AI.12)}$$

so that

$$V_{t:t} = \{I - V_{t:t-1} A_t' [C + A_t V_{t:t-1} A_t']^{-1} A_t'\} V_{t:t-1} \qquad \text{(AI.12a)}$$

or, on using the definiton of M_t given in (2.11a),

$$V_{t:t} = V_{t:t-1} - V_{t:t-1} A_t' M_t^{-1} A_t V_{t:t-1} \qquad \text{(AI.13)}$$

which indeed is the result (2.12c).

Appendix II

This appendix is devoted to the derivation of results (3.6), (3.7) and (3.8).

We have from (3.5), with f given by (3.1) and p given by (2.3), that the predictive distribution of $\underset{\sim}{Y}_t$, given $\underset{\sim}{Y}_{t-1}$, is such that

$$h(\underset{\sim}{Y}_t|\underset{\sim}{Y}_{t-1}) = \sum_{i=1}^{2} a_i \int_{\underset{\sim}{\theta}_t} f(\underset{\sim}{Y}_t|A_t\ \underset{\sim}{\theta}_t, C_i) f(\underset{\sim}{\theta}_t|\underset{\sim}{\mu}_{t:t-1}, V_{t:t-1}) d\underset{\sim}{\theta}_t \qquad \text{(AII.1)}$$

where the density f is defined in (3.7d). But the integral operation is clearly equivalent to the argument used in (2.10) with C_i replacing C so that from (2.10) - (2.11) we have

$$h(\underset{\sim}{Y}_t|\underset{\sim}{Y}_{t-1}) = \sum_{i=1}^{2} a_i f(\underset{\sim}{Y}_t|A_t\ \underset{\sim}{\mu}_{t:t-1}, M_{t,i}) \qquad \text{(AII.2)}$$

where $M_{t,i}$ is as defined in (3.6a) - see (2.11a), and the result (3.6) is now proved.

To derive the results (3.7), we note that (3.4) gives the prior for $\underset{\sim}{\theta}_t$, while the distribution specified by (3.1) dictates the likelihood of $\underset{\sim}{Y}_t$, given $\underset{\sim}{\theta}_t$. Hence the posterior of $\underset{\sim}{\theta}_t$, given $\underset{\sim}{Y}_t$ is such that

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{Y}_t) \propto \sum_{i=1}^{2} a_i f(\underset{\sim}{Y}_t|A_t\ \underset{\sim}{\theta}_t, C_i) f(\underset{\sim}{\theta}_t|\underset{\sim}{\mu}_{t:t-1}, V_{t:t-1}) \quad . \qquad \text{(AII.3)}$$

In common to the work of Appendix I, the summand of (AII.3) requires the "completion of a square" operation to form a quadratic form in $\underset{\sim}{\theta}_t$. However, since the "constants" left over in this operation depend on "i", they cannot be absorbed by the constant of proportionality. Now for the i^{th} term, we have in the exponent, apart from the $-\frac{1}{2}$,

$$(\underset{\sim}{Y}_t - A_t\ \underset{\sim}{\theta}_t)' C_i^{-1} (\underset{\sim}{Y}_t - A_t\ \underset{\sim}{\theta}_t) + (\underset{\sim}{\theta}_t - \underset{\sim}{\mu}_{t:t-1})' V_{t:t-1}^{-1} (\underset{\sim}{\theta}_t - \underset{\sim}{\mu}_{t:t-1}) \qquad \text{(AII.4)}$$

and from the work in Appendix I, we see that (AII.4) may be written as

$$\begin{aligned} &[\underset{\sim}{\theta}_t - V_{t:t}^{(i)}\ \underset{\sim}{b}_t^{(i)}]' V_{t:t}^{(i)\,-1} [\underset{\sim}{\theta}_t - V_{t:t}^{(i)}\ \underset{\sim}{b}_t^{(i)}] + \underset{\sim}{Y}_t' C_i^{-1}\ \underset{\sim}{Y}_t \\ &+ \underset{\sim}{\mu}_{t:t-1}' V_{t:t-1}^{-1}\ \underset{\sim}{\mu}_{t:t-1} - \underset{\sim}{b}_t^{(i)\prime}\ V_{t:t}^{(i)}\ \underset{\sim}{b}_t^{(i)} \end{aligned} \qquad \text{(AII.5)}$$

where

$$\begin{aligned} V_{t:t}^{(i)} &= \{V_{t:t-1}^{-1} + A_t' C_i^{-1} A_t\}^{-1} \\ &= V_{t:t-1} - V_{t:t-1} A_t' M_{t,i}^{-1} A_t V_{t:t-1} \end{aligned} \qquad \text{(AII.6)}$$

and $M_{t,i} = C_i + A_t V_{t:t-1} A_t'$ (c.f. with (AI.4) and (AI.13)), and where

$$\underset{\sim}{b}_t^{(i)} = V_{t:t-1}^{-1} \underset{\sim}{\mu}_{t:t-1} + A_t' C_i^{-1} \underset{\sim}{y}_t \tag{AII.7}$$

(c.f. with (AI.4a)).

Now the last three terms in (AII.5) may be written as

$$\underset{\sim}{y}_t'[C_i^{-1} - C_i^{-1} A_t V_{t:t}^{(i)} A_t' C_i^{-1}]\underset{\sim}{y}_t + \underset{\sim}{\mu}_{t:t-1}'[V_{t:t-1}^{-1} - V_{t:t-1}^{-1} V_{t:t}^{(i)} V_{t:t-1}]$$
$$- 2\underset{\sim}{y}_t' C_i^{-1} A_t V_{t:t}^{(i)} V_{t:t-1}^{-1} \underset{\sim}{\mu}_{t:t-1} \tag{AII.8}$$

and it is straightforward to show that (AII.8) may be written as

$$\underset{\sim}{y}_t' M_{t,i}^{-1} \underset{\sim}{y}_t + \underset{\sim}{\mu}_{t:t-1}' A_t' M_{t,i}^{-1} A_t \underset{\sim}{\mu}_{t:t-1} - 2\underset{\sim}{y}_t' M_{t,i}^{-1} A_t \underset{\sim}{\mu}_{t:t-1} \tag{AII.9}$$

which is of course,

$$(\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1})' M_{t,i}^{-1} (\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1}) \quad .$$

Hence $p(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t)$ of (AII.3) is such that

$$p(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t) \propto \sum_{i=1}^{2} \alpha_i |C_i^{-1}|^{1/2} |V_{t:t-1}^{-1}|^{1/2} \exp - \frac{1}{2}(\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1})'$$
$$M_{t,i}^{-1}(\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1}) \times \exp - \frac{1}{2} (\underset{\sim}{\theta}_t - \underset{\sim}{\mu}_{t:t}^{(i)})' V_{t:t}^{(i)\,-1} (\underset{\sim}{\theta}_t - \underset{\sim}{\mu}_{t:t}^{(i)}) \tag{AII.10}$$

where we have written $\underset{\sim}{\mu}_{t:t}^{(i)} = V_{t:t}^{(i)} \underset{\sim}{b}_t^{(i)}$, and using the work in Appendix I, we find

$$\underset{\sim}{\mu}_{t:t}^{(i)} = \underset{\sim}{\mu}_{t:t-1} + V_{t:t}^{(i)} A_t' C_i^{-1} (\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1}) \tag{AII.11}$$

or

$$\underset{\sim}{\mu}_{t:t}^{(i)} = \underset{\sim}{\mu}_{t:t-1} + V_{t:t-1} A_t' M_{t,i}^{-1} (\underset{\sim}{y}_t - A_t \underset{\sim}{\mu}_{t:t-1}) \quad . \tag{AII.11a}$$

Using the definition of the multivariate normal density given in (3.7d), we now easily find that

$$p(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t) = K \sum_{i=1}^{2} \alpha_i \frac{|C_i^{-1}|^{1/2} |V_{t:t-1}^{-1}|^{1/2}}{|M_{t,i}^{-1}|^{1/2} |V_{t:t}^{(i)\,-1}|^{1/2}} f(\underset{\sim}{y}_t | A_t \underset{\sim}{\mu}_{t:t-1}, M_{t,i})$$
$$\times f(\underset{\sim}{\theta}_t | \underset{\sim}{\mu}_{t:t}^{(i)}, V_{t:t}^{(i)}) \tag{AII.12}$$

and it is easy to see that the determinants involved are such that the indicated products and ratios involved is 1. Finally, integrating with respect to $\underset{\sim}{\theta}_t$ yields

$$K^{-1} = \sum_{i=1}^{2} \alpha_i \, f(\underset{\sim}{y}_t | A_t \, \underset{\sim}{\mu}_{t:t-1}, M_{t,i}) = h(\underset{\sim}{y}_t | \underset{\sim}{y}_{t-1}) \tag{AII.13}$$

so that using this in (AII.12) we now have

$$p(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t) = \sum_{i=1}^{2} \alpha_{t,i} \, f(\underset{\sim}{\theta}_t | \underset{\sim}{\mu}_{t:t}^{(i)}, V_{t:t}^{(i)}) \tag{AII.14}$$

with $\alpha_{t,i}$ as advertised in (3.7c).

Now using (AII.14), we can calculate moments. We have

$$\begin{aligned} E(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t) &= \sum_{i=1}^{2} \alpha_{t,i} \int \underset{\sim}{\theta}_t \, f(\underset{\sim}{\theta}_t | \underset{\sim}{\mu}_{t:t}^{(i)}, V_{t:t}^{(i)}) d\underset{\sim}{\theta}_t \\ &= \sum_{i=1}^{2} \alpha_{t,i} \, \underset{\sim}{\mu}_{t:t}^{(i)} \end{aligned} \tag{AII.15}$$

and using (AII.11a), which is the result (3.7a), we have

$$E(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t) = \underset{\sim}{\mu}_{t:t-1} + \sum_{1}^{2} \alpha_{t,i} V_{t:t-1} A_t' M_{t,i}^{-1} (\underset{\sim}{y}_t - A_t \, \underset{\sim}{\mu}_{t:t-1}) \tag{AII.16}$$

which may be expressed as in (3.8), or as above in (AII.15).

To find the variance - covariance, we first determine $E(\underset{\sim}{\theta}_t \, \underset{\sim}{\theta}_t' | \underset{\sim}{y}_t)$ and then use the identity

$$V(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t) = E(\underset{\sim}{\theta}_t \, \underset{\sim}{\theta}_t' | \underset{\sim}{y}_t) - [E(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t)] [E(\underset{\sim}{\theta}_t | \underset{\sim}{y}_t)]' \quad . \tag{AII.17}$$

Now from (AII.14), we have

$$\begin{aligned} E(\underset{\sim}{\theta}_t \, \underset{\sim}{\theta}_t' | \underset{\sim}{y}_t) &= \sum_{i=1}^{2} \alpha_{t,i} E(\underset{\sim}{\theta}_t \, \underset{\sim}{\theta}_t' | \underset{\sim}{\mu}_{t:t}^{(i)}, V_{t:t}^{(i)}) \\ &= \sum_{i=1}^{2} \alpha_{t,i} [V_{t:t}^{(i)} + \underset{\sim}{\mu}_{t:t}^{(i)} \, \underset{\sim}{\mu}_{t:t}^{(i)'}] \quad . \end{aligned} \tag{AII.18}$$

Now $V_{t,t}^{(i)}$ is given in (AII.6) so that we have

$$E(\underset{\sim}{\theta}_t \underset{\sim}{\theta}_t' | \underset{\sim}{Y}_t) = V_{t:t-1} - \sum_{i=1}^{2} \alpha_{t,i} V_{t:t-1} A_t' M_{t,i}^{-1} A_t V_{t:t-1}$$

$$+ \sum_{i=1}^{2} \alpha_{t,i} \underset{\sim}{\mu}_{t:t}^{(i)} \underset{\sim}{\mu}_{t:t}^{(i)\,\prime} \quad . \tag{AII.19}$$

Substituting (AII.16) in (AII.19), and doing some straightforward but tedious algebra, and remembering that $\alpha_{t,2} = 1-\alpha_{t,1}$, we find

$$V(\underset{\sim}{\theta}_t | \underset{\sim}{Y}_t) = E(\underset{\sim}{\theta}_t \underset{\sim}{\theta}_t' | \underset{\sim}{Y}_t)$$

$$- \{\underset{\sim}{\mu}_{t:t-1} + \sum_{1}^{2} \alpha_{t,1} V_{t:t-1} A_t' M_{t,i}^{-1} (\underset{\sim}{Y}_t - A_t \underset{\sim}{\mu}_{t:t-1})\} \tag{AII.20}$$

$$\times \{\underset{\sim}{\mu}_{t:t-1} + \Sigma\, \alpha_{t,i} V_{t,t-1} A_t' M_{t,i}^{-1} (\underset{\sim}{Y}_t - A_t \underset{\sim}{\mu}_{t:t-1})\}',$$

takes the form advertised in (3.8a).

REFERENCES

(1) Abraham, B. and Box, G. E. P. (1979), "Bayesian Analysis of Some Outlier Problems in Time Series," Biometrika, 66, pp. 229-236.

(2) Alspach, D. L. and Sorenson, H. W. (1971), "Recursive Bayesian Estimation Using Gaussian Sums," Automatica, 6, pp. 465-479.

(3) Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972), Robust Estimates of Location, Princeton University Press, Princeton, New Jersey.

(4) Aoki, M. (1967), Optimization of Stochastic Systems, Academic Press, New York.

(5) Box, G. E. P. and Tiao, G. C. (1968), "A Bayesian Approach to Some Outlier Problems," Biometrika, 55, pp. 119-129.

(6) Bryson, A. E., Jr. and Ho, Y.-C. (1975), Applied Optimal Control, Hemisphere Publishing Corporation, New York.

(7) Chen, G. and Box, G. E. P. (1979), "Further study of robustification via a Bayesian approach", Technical Report #1998, Mathematics Research Center, University of Wisconsin-Madison.

(8) Cheng, G. and Box, G. E. P. (1980), "Implied Assumptions for Some Proposed Robust Estimators," Mathematics Research Center Technical Summary Report #1997, University of Wisconsin-Madison.

(9) Crowe, C. M. et al (1983), "Reconciliation of Process Flow Rates by Matrix Projection," AIChE Journal, 29, 6, pp. 881-888.

(10) Guttman, I. (1973), "Care and Handling of Univariate or Multivariate Outliers in Detecting Spuriosity - A Baysian Approach," Technometrics, Vol. 15, pp. 723-738.

(11) Harrison, P. J. and Stevens, C. F. (1976), "Bayesian Forecasting," (with discussion), Journal of the Royal Statistical Society, B, 38, pp. 205-247.

(12) Jeffreys, H. (1961), Theory of Probability, third edition, Oxford: Clarendon Press.

(13) Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," Transactions of ASME, Vol. 82D, p. 35.

(14) Little, J. K. (1983), "Regression Diagnostics and the Bayesian Analysis of the Scale-Contaminated Normal Model," Technical Report 713, Department of Statistics, University of Wisconsin-Madison.

(15) Mah, R. S. H. and Tamhane, A. C. (1982), "Detection of Gross Errors in Process Data," AICHE Jorunal, 28, 5, pp. 828-830.

(16) Masreliez, C. J. (1975), "Approximate Non-Gaussian Filtering With Linear State and Observation Relations," IEEE Trans. on Automatic Control, A-20, pp. 107-110.

(17) Masreliez, C. J. and Martin, R. D. (1977), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," IEEE Trans. on Automatic Control, AC-22, 3, pp. 361-371.

(18) McWorter, A., Spivey, W. A. and Wrobleski, W. J. (1976), "Sensitivity Analysis of Varying Parameter Econometric Models," International Statistical Review, Vol. 44, #2, pp. 265-282.

(19) Plackett, R. L. (1950), "Some Theorems on Least Squares," Biometrika, 37, pp. 149-157.

(20) Stein, C. (1956), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," 3rd Berk Symp. Math. Stat. Prob., 1, 197-206.

(21) Tsai, C. and Kurz, L. (1983), "An Adaptive Robustizing Approach to Kalman Filtering," Automatica, 19, 3, pp. 279-288.

IG/DP/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER #2766	2. GOVT ACCESSION NO. AD-A149 044
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Robust Kalman Filtering and Its Applications	5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Irwin Guttman and Daniel Peña	8. CONTRACT OR GRANT NUMBER(s) A8743 DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below	12. REPORT DATE October 1984
	13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

NSERC
Ottawa, Ontario
Canada

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Kalman Filter, Robustness, mixtures of normals

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper presents a robust Kalman filtering algorithm that is obtained assuming a scale contaminated normal distribution for the noise of the measurement equation. The mixture of normals obtained as a posterior distribution is approximated at each stage by a normal distribution with the same mean and variance. The resulting algorithm is simple, has a straightforward interpretation and seems to provide useful robust estimators in several statistical problems that are briefly reviewed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

2-85

DTIC