INFORMATION AND COMPUTATION 79, 163-192 (1988)

Outlier Resistant Filtering and Smoothing*

HARALAMPOS TSAKNAKIS

Alphatech, Inc.,
Burlington, Massachusetts 01803-4901

AND

P. PAPANTONI-KAZAKOS

Electrical Engineering Department,
Thornton Hall, University of Virginia,
Charlottesville, Virginia 22901

We consider a stationary Gaussian information process transmitted through an additive noise channel. We assume that the noise and information processes are mutually independent, and we model the noise process as nominally Gaussian with additive outliers. For the above system model, we first develop a theory for outlier resistant filtering and smoothing operations. We then design specific such nonlinear operations, and we study their performance. The performance criteria are the asymptotic mean squared error at the Gaussian nominal model, the breakdown point, and the influence function. We find that the proposed operations combine excellent performance at the nominal model with strong resistance to outliers.

I. INTRODUCTION

In filtering and smoothing, information carrying data are extracted from noisy observations. The formalization and solution of the filtering and smoothing problems are well established, when the joint process that characterizes the relationship between information and noise data sequences is statistically well known (see Kalman, 1960, 1963; Kolmogorov, 1941; Wiener, 1949), or parametrically known. Linear filtering and smoothing operations are then by far the most widely used, due to their simplicity in implementation. In practice, however, the occurrence of occasional extremely erroneous data values, called outliers, are frequently observed. Furthermore, linear data operations are notoriously nonresistant to such outliers, inducing dramatic performance instabilities. The purpose of this

* This work was supported by the U. S. Air Force Office of Scientific Research, under Grants AFOSR-83-0229 and AFOSR-87-0224.

0890-5401/88 $3.00

paper is to establish a theory for outlier resistant filtering and smoothing procedures and to provide such specific data operations for Gaussian information processes and additive, nominally Gaussian, noise processes. The initial steps of our presentation are based on the theory of quantitative robustness (see Boente, Fraiman, and Yohai, 1982; Cox, 1978; Hampel, 1971; Papantoni–Kazakos and Gray, 1979; Papantoni–Kazakos, 1981, 1987, 1984a, 1984b). Our approaches on pertinent performance criteria are as those in Hampel *et al.* (1986).

Problems of nonlinear filtering are considered in the paper by Masreliez and Martin (1977). In particular, the above authors present a robustification procedure for Kalman filters operating on the outputs of linear dynamical systems. Discussion of their results and comparisons with ours are given in Sections 4 and 6 of this paper.

A general theory and methodology for nonlinear smoothers, acting on stationary processes, is developed by Mallows (1980). The issue of primary concern there is the decomposition of smoothers into linear and nonlinear parts and the study of their properties. Furthermore, the problem of outlier resistance is examined, using as the indicator of resistance an extension of Hampel's concept of the breakdown point. However, explicit design issues are not undertaken. Relevant results on the design and analysis of specific outlier resistant filters and smoothers for stationary processes can be found in Tsaknakis (1986).

In Section 2 of this paper, we first present a formalization of the filtering and smoothing problem under consideration. Then, we define outlier resistance for filtering and smoothing operations and present certain sufficient conditions for resistance of such operations. In Section 3, a two person game formalization is adopted for fixed finite length operations and the corresponding least favorable structure is derived. In Section 4, the above structure is used for the design of a causal recursive filtering operation when the nominal information process is autoregressive and the nominal noise process is i.i.d. Then the asymptotic properties of the resulting operation are studied on a stationary environment, in terms of asymptotic outlier resistance, asymptotic stationarity, and asymptotic mean square error at the nominal model.

In Section 5, we define the breakdown point and the influence function of a filtering or smoothing operation. Both these quantities are defined in such a way as to reflect important sensitivity aspects of the mean square error, induced by the filtering or smoothing operation, to the action of outliers. Then, we continue with the explicit evaluation and study of the breakdown point and influence function of the filter presented in Section 4. Section 6 is devoted to the numerical evaluation and comparison of the proposed filter in relationship to an existing one, for specific numerical examples. Finally, in Section 7, we briefly present some conclusions.

2. Preliminaries

We consider real-valued discrete-time information and noise stochastic processes, denoted respectively by $\{X_n, n \in Z\}$, $\{W_n, n \in Z\}$, where Z is the set of integers. The observation process, $\{Y_n, n \in Z\}$ is given by the equation

$$Y_n = X_n + W_n, \qquad n \in Z. \tag{1}$$

It will be assumed that the information and noise processes are independent. Then a complete statistical description of the model (1) is provided by the probability measures of $\{X_n\}$, $\{W_n\}$, denoted by μ_s, μ_N, respectively. The probability measures of the observation process $\{Y_n\}$, denoted by μ_Y, is expressed as the convolution $\mu_Y = \mu_s * \mu_N$ and the joint probability measure of $\{Y_n, X_n\}$, denoted by μ, is expressed as the product $\mu(\{Y_n, X_n\}) = \mu_s(\{X_n\})\, \mu_N(\{Y_n - X_n\})$. Assuming that $\{X_n\}$ has finite second-order moments, let us consider the minimum mean square estimation of the information process value X_0 given a finite length l observation sequence $\{y_i, y_{i+l-1}\}$, denoted as y^l for short. If $i+l-1 \leqslant 0$, we refer to causal filtering or simply filtering. If $i+l-1 > 0$, we refer to noncausal filtering or smoothing. Given the measure μ, the minimum mean square estimator, $\hat{X}_0$, of X_0 is the conditional expectation

$$\hat{X}_0(y^l) = E\{X_0/y^l, \mu\} \tag{2}$$

which is a function of the sequence y^l whose specific form is determined by μ. If μ is Gaussian, $\hat{X}_0(y^l)$ is an affine transformation of y^l. The induced by $\hat{X}_0$ mean square error is denoted by $e(\mu, \hat{X}_0)$ and is a functional of μ and $\hat{X}_0$ given by the expression

$$e(\mu, \hat{X}_0) = E\{[X_0 - \hat{X}_0(Y^l)]^2/\mu\}. \tag{3}$$

The occurrence of occasional erroneous values in the noise process $\{W_n\}$, called outliers, induces uncertainties in the description of the measure μ_N. That induces further uncertainties in the measures μ and μ_Y. The initial issue here is the qualitative characterization of those uncertainties. A particularly useful tool for describing uncertainties of probability measures is the Prohorov distance (see Hampel, 1971; Boente *et al.*, 1982; Papantoni-Kazakos, 1987, 1984b), whose definition is given below.

Let $\rho(\cdot, \cdot)$ be a metric in R^n, and let v_1, v_2 be probability measures defined on the Borel field of (R^n, ρ). Let N be the class of all joint measures, v, whose marginals are v_1 and v_2. The Prohorov distance $\Pi_\rho(v_1, v_2)$ is defined as

$$\Pi_\rho(v_1, v_2) = \inf_{v \in N} \inf\{\delta > 0 : v(\alpha, \beta : \rho(\alpha, \beta) > \delta) \leqslant \delta\}, \tag{4}$$

where α, β denote elements of R^n.

The selection of the metric $\rho(\cdot, \cdot)$ reflects the pattern according to which the outliers corrupt the nominal process. For the purpose of this paper, we select a metric which corresponds to outliers occurring in batches, or bursts of size m, m being a fixed design parameter. Such a metric is defined as (see Papantoni-Kazakos, 1984a, 1984b):

For $\alpha, \beta \in R^n$, let $\tilde{\alpha}, \tilde{\beta}$ be sequences generated by repetitions of α and β. Also, let $\tilde{\alpha}_j^k$ denote $(\tilde{\alpha}_j, ..., \tilde{\alpha}_k)$, $j \leqslant k$. Then, we define the metric $\rho_{n,m}(\cdot, \cdot)$ in R^n as

$$\rho_{n,m}(\alpha, \beta) = \inf\{\delta > 0 : n^{-1}[\underset{i=1,\ldots,n}{\#\quad i} : \gamma_m(\tilde{\alpha}_i^{i+m-1}, \tilde{\beta}_i^{i+m-1}) > \delta] \leqslant \delta\}, \tag{5}$$

where, the auxiliary metric $\gamma_m(\cdot, \cdot)$ is defined as

$$\gamma_m(\alpha', \beta') = m^{-1} \sum_{i=1}^{m} |\alpha'_i - \beta'_i|, \qquad \alpha', \beta' \in R^m. \tag{6}$$

In the sequel, we use the Prohorov distance in (4) with the metric (5) to give a formal definition of outlier resistant estimators. Let μ_{oN}, μ_{oY} denote the nominal measures of the noise and observation processes respectively, and μ_o denote the nominal joint measure of the observation and information processes. Also, let $\mu_{os} = \mu_s$ be the fixed information process measure.

An estimator $\hat{X}_0(y^l)$ of X_0 from the observation sequence y^l is called *outlier resistant* or *qualitatively robust* at μ_{oN} if $\forall \eta > 0$, there is an $\varepsilon > 0$ such that

$$\Pi_{\rho_{n,m}}(\mu_{oN}, \mu_N) < \varepsilon \qquad \text{implies } |e(\mu_o, \hat{X}_0) - e(\mu, \hat{X}_0)| < \eta$$

for every n. Notice that μ_o and μ are fully determined from μ_{oN} and μ_N.

Considering stationary and ergodic processes, the limit $\lim_{n\to\infty} \Pi_{\rho_{n,m}}(\mu_{oN}, \mu_N)$ is equal to the Prohorov distance $\Pi_{\gamma_m}(\mu_{oN}, \mu_N)$. Since the Prohorov distance $\Pi_{\gamma_m}(\cdot, \cdot)$ metrizes the weak topology of the probability measure on (R^m, γ_m), an estimator $\hat{X}_0(y^l)$ of length $l \leqslant m$ is resistant, if it is pointwise continuous and bounded. Such estimators are constructed in Section 3. However, for $l > m$, these conditions are no longer sufficient; appropriate resistant estimators of asymptotically large length are constructed in Section 4.

Consider now the m-dimensional restriction of the nominal measure μ_{oN}, and let it be denoted by μ_{oN}^m. Furthermore, assume that μ_{oN}^m is absolutely continuous with density f_{oN}^m. Then, the ε-contaminated class of densities

$$\mathscr{F}_N^m(\varepsilon) = \{f_N^m = (1-\varepsilon) f_{oN}^m + \varepsilon h^m,\ h^m \text{ arbitrary } m\text{-dimensional density}\} \tag{7}$$

is contained in the class $\Pi_{\gamma_m}(\mu_{oN}, \mu_N) \leqslant \varepsilon$ of measures μ_N, for any ε, $0<\varepsilon<1$. The constant ε can be interpreted as frequency of outlier occurrence.

The class $\mathscr{F}_N^m(\varepsilon)$ of noise densities induces the following class of joint m-dimensional densities of the observation and information processes.

$$\begin{aligned}\mathscr{F}^m(\varepsilon) = \{f^m : f^m(y^m, x^m) \\ = (1-\varepsilon) f_{os}^m(x^m) f_{oN}^m(y^m - x^m) + \varepsilon f_{os}^m(x^m)\, h^m(y^m - x^m),\ h^m \text{ arbitrary}\}.\end{aligned} \tag{8}$$

Class $\mathscr{F}^m(\varepsilon)$ contains all the necessary statistical information for constructing estimators of length at most m and will be used in the forthcoming section as the model for statistical contamination.

3. Construction of Filtering and Smoothing Operations—Step 1

In this section we derive a finite length robust estimator of the information process given observation sequence of length $l \leqslant m$, where m corresponds to outlier patterns, as discussed in the previous section. The derivation is based on a two-person game formulation of the estimation problem, with payoff function the induced mean square error. To fix ideas, suppose that X_0 is to be estimated from a length l observation sequence y_i^{i+l-1}, denoted as y^l for short (assume $i \leqslant 0 \leqslant i+l-1$). The joint density of x_0 and y^l, denoted by $f(x_0, y^l)$, belongs to an ε-contaminated class obtained from the appropriate restriction of the more general class $\mathscr{F}^m(\varepsilon)$, as defined in (8). We assume that the information process is a fixed zero mean Gaussian process and that the nominal noise process is also zero-mean Gaussian. Given an estimator $\hat{X}_0(y^l)$ and a joint density $f(x_0, y^l)$, the mean square error $e(f, \hat{X}_0)$ of $\hat{X}_0$ at f is the expectation $E\{[X_0 - \hat{X}_0(Y^l)]^2/f\}$. The objective is to find a density-estimator pair $(f^*, \hat{X}_0^*)$ that constitutes a saddle point, i.e.,

$$e(f, \hat{X}_0^*) \leqslant e(f^*, \hat{X}_0^*) \leqslant e(f^*, \hat{X}_0) \tag{9}$$

for every $\hat{X}_0$ measurable and $f \in \mathscr{F}^m(\varepsilon)$

Unfortunately, a saddle point solution of the above game for the class $\mathscr{F}^m(\varepsilon)$ does not exist. In particular, the quantity $\inf_{\hat{X}_0} \sup_{f \in F^m(\varepsilon)} e(f, \hat{X}_0)$ is strictly larger than $\sup_{f \in F^m(\varepsilon)} \inf_{\hat{X}_0} e(f, \hat{X}_0)$ and the latter supremum with respect to f cannot be attained in $\mathscr{F}^m(\varepsilon)$. This is due to the non-tightness of $\mathscr{F}^m(\varepsilon)$ which allows probability masses to escape to infinity. For this reason we consider an enlargement of the class $\mathscr{F}^m(\varepsilon)$ to include all densities of the form (we denote the enlarged class by the same symbol):

$$\mathscr{F}^m(\varepsilon) = \{f^m : f^m(y^m, x^m) = (1-\varepsilon) f^m_{os}(x^m) f^m_{oN}(y^m - x^m) + \varepsilon f^m_{os}(x^m)\, h^m(y^m)$$
$$h^m \text{ arbitrary } m\text{-dimensional density}\}. \tag{10}$$

The enlarged class $\mathscr{F}^m(\varepsilon)$ in (10) is equivalent to considering outliers affecting the observation process directly, not via the additive noise process, as is the case with the class in (8). However, the minimax value of the game for the class in (8) is the same as the minimax value for the class in (10). Furthermore, a saddle point solution of the game (9) always exists within the class $\mathscr{F}^m(\varepsilon)$ in (10). From now on we consider only the class $\mathscr{F}^m(\varepsilon)$ as defined in (10), and we seek the saddle point solution of the game in this class.

From the results in Papantoni-Kazakos (1984a) we conclude that the saddle point of the game can be found by solving

$$\sup_{f \in F^m(\varepsilon)} \inf_{\hat{X}_0} e(f, \hat{X}_0). \tag{11}$$

The expression $\inf_{\hat{X}_0} e(f, \hat{X}_0)$ represents the minimum mean square error at the density f and can be written as

$$\sigma_o^2 - I(f)$$

where, $\sigma_o^2 = E(X_0^2)$ is the fixed variance of X_0, and

$$I(f) = E\{E^2\{X_0/Y^l, f\}/f\} = \int_{R^l} \frac{(\int_{R^1} x_0 f(x_0, y^l)\, dx_0)^2}{\int_{R^1} f(x_0, y^l)\, dx_0}\, dy^l. \tag{12}$$

Considering the form of $f(x_0, y^l)$ in terms of the nominal and contaminating densities, as derived from (10), and the zero mean assumption of the nominal densities, the quantity $I(f)$ can be written as a functional of the l-dimensional restriction of the density of the observation sequence y^l. Let us denote the latter density by $f_Y(y^l)$. After some algebra, we obtain

$$I(f) = I(f_Y) = \int_{R^l} \frac{((1-\varepsilon) f_{oY}(y^l)(P^{\mathrm{T}} y^l))^2}{f_Y(y^l)}\, dy^l, \tag{13}$$

where, $f_{oY}(y^l)$ is the nominal density of Y^l at the vector point y^l, given by the convolution of the information density f^l_{os} and nominal noise density f^l_{oN}, and the inner product $P^{\mathrm{T}} y^l$ is the optimal linear estimator of X_0 from

y^l under *nominal conditions* (i.e., for $\varepsilon = 0$). The density $f_Y(y^l)$ belongs to the class $\mathscr{F}^l_Y(\varepsilon)$, obtained from $F^m(\varepsilon)$ as follows

$$\mathscr{F}^l_Y(\varepsilon) = \{f_Y(y^l) : f_Y(y^l) = (1-\varepsilon) f^l_{os}(y^l) * f^l_{oN}(y^l) + \varepsilon h^l(y^l)\}.$$

Problem (11) can now be reduced to

$$\inf_{f_Y \in F^l_Y(\varepsilon)} I(f_Y). \tag{14}$$

Although the class of densities $\mathscr{F}^k_Y(\varepsilon)$ is not tight (therefore not compact) in the weak topology of all probability measures on the Borel σ-field of the metric space (R^l, γ_l), the infimum in (14) is attained in $\mathscr{F}^l_Y(\varepsilon)$. Furthermore, there is a unique member of $\mathscr{F}^l_Y(\varepsilon)$ attaining that infimum, under the nominal assumptions discussed before. The above assertions together with the explicit form of the infimum and the corresponding estimator, constitute the statements of Theorem 1 below, whose proof is in the Appendix.

Let $\phi(x)$ and $\Phi(x)$ be the zero mean unit variance Gaussian density and cumulative distribution, respectively. Let $H(\lambda, z)$, $\lambda > 0$ be the Huber function defined as

$$H(\lambda, z) = \max(-\lambda, \min(\lambda, z)). \tag{15}$$

Finally, let r be the nominal variance of the linear form $P^T Y^l$, i.e., $r = E\{(P^T Y^l)^2\} = \int_{R^l} (P^T y^l)^2 f_{oY}(y^l)\, dy^l$. Then, we express Theorem 1 as follows.

THEOREM 1. (i) *There is a unique saddlepoint solution* $(f^*, \hat{X}_0^*)$ *of the game* (9).

(ii) *The saddlepoint observation density* f^*_Y *and estimator* $\hat{X}_0^*$ *are given by the equations*

$$f^*_Y(y^l) = \begin{cases} (1-\varepsilon) f_{oY}(y^l), & \text{for} \quad |P^T y^l| \leqslant \lambda \\[2ex] (1-\varepsilon) \dfrac{|P^T y^l|}{\lambda} f_{oY}(y^l), & \text{for} \quad |P^T y^l| > \lambda \end{cases} \tag{16}$$

$$\hat{X}_0^*(y^l) = H(\lambda, P^T y^l), \tag{17}$$

where

$$\lambda = c\sqrt{r}$$

$$c : \Phi(c) + c^{-1}\phi(c) = \frac{2-\varepsilon}{2(1-\varepsilon)}. \tag{18}$$

We note that the estimator $\hat{X}_0^*$ in (17) above, is a truncated version of the linear, nominally optimal mean square estimator $P^{\mathrm{T}}y^l$. The truncation constant λ is proportional to the square root of the quantity r which is the variance gain in estimating X_0 from y^l under nominal conditions ($\varepsilon = 0$). The proportionality factor c tends to infinity for $\varepsilon \to 0$. In the latter case, the estimator (17) becomes identical to the nominally optimal mean square estimator.

There are interesting similarities and differences between the estimator $\hat{X}_0^*$ in (17) and the classical robust parameter estimator of Huber (1964). Both estimators introduce the same form of nonlinearity to limit the influence of bad observations. However, while Huber's estimator applies the nonlinearity on each one of the observation data, the estimator derived here applies a similar nonlinearity on a linear combination of the observation data. Furthermore, the form of the least favorable density derived in (16) has heavier tails than the Gaussian by the linear factor $|P^{\mathrm{T}}y^l|$, while, in the robust parameter estimation problem, the corresponding least favorable density has much heavier exponential tails. Regarding these comparisons, it should be pointed out that Huber's result is based on the maximum likelihood estimation of the unknown mean of a contaminated distribution, while the result of Theorem 1 is based on a Bayesian estimation of a random process corrupted by contaminated noise and with the mean square as performance criterion.

Regarding qualitative robustness, we note that for any ε, $0 < \varepsilon < 1$, the estimator $\hat{X}_0^*$ is both continuous and bounded satisfying thus the conditions for outlier resistance stated in the previous section for finite length estimators.

The mean square error induced by $\hat{X}_0^*$ at the least favorable density f_Y^* is equal to $\sigma_o^2 - I(f_Y^*)$. This is the largest possible error within the class $F^m(\varepsilon)$ and by substitution we obtain

$$e(f_Y^*, \hat{X}_0^*) = \sigma_o^2[1 - (1 - \varepsilon)(2\Phi(c) - 1)\, q^2],$$

where, $q = \sigma_o^{-1}\sqrt{r}$. Let $e(f_o, \hat{X}_0^*)$ be the mean square error induced by the robust estimator $\hat{X}_0^*$ at the nominal Gaussian density. Also, let e^o be the nominally optimal mean square error. Then, after some computations we obtain

$$e^o = \sigma_o^2(1 - q^2)$$

$$e(f_o, \hat{X}_0^*) = e^o - 2r(\Phi(-c)(1 + c^2) - c\phi(c)).$$

The second term of the right-hand side of the above equation is always positive and represents the performance loss that is incurred if the robust nonlinear estimator $\hat{X}_0^*$ is applied, instead of the linear nominally optimal one.

4. Construction of Filtering and Smoothing Operations—Step 2

We now consider the case when the number of observation data is larger than the parameter m. For this case and for arbitrary nominal information and noise processes, results concerning the design and study of appropriate nonlinear filtering and smoothing operations can be found in Tsaknakis (1986). For the purpose of this paper we will focus on autoregressive Gaussian information processes and white Gaussian nominal noise processes.

Let the nominal information and observation processes $\{X_n\}$, $\{Y_n\}$ be given by the equations

$$\begin{aligned} X_n &= a_1 X_{n-1} + a_2 X_{n-2} + \cdots + a_k X_{n-k} + V_n \\ Y_n &= X_n + W_n, \end{aligned} \tag{19}$$

where, $\{V_n\}$, $\{W_n\}$ are mutually independent, i.i.d., and zero mean Gaussian, with variances σ_x^2 and σ_w^2, respectively. Upon defining

$$\mathbf{U}_n^{\mathrm{T}} = [X_n, X_{n-1}, ..., X_{n-k+1}]$$

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_k \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots 1 & 0 \end{bmatrix} \tag{20}$$

$$B^{\mathrm{T}} = [1 \quad 0 \quad \cdots \quad 0],$$

the nominal model can be described in the following vector form

$$\begin{aligned} \mathbf{U}_n &= A\mathbf{U}_{n-1} + BV_n \\ Y_n &= B^{\mathrm{T}}\mathbf{U}_n + W_n. \end{aligned} \tag{21}$$

Writing the system (19) in the vector form (21) has the advantage of the recursive Kalman filtering relationships for the nominal model. We want to estimate x_0 given observations $y_0, y_{-1}, ..., y_{-l+1}$ for any value of l, when the observation process is corrupted by outliers occurring in batches of size m. When $l \leqslant m$, we apply the minimax estimator derived in the previous section. When $l > m$, we consider estimating the entire vector $\mathbf{u}_0$ given the above measurements, and we define the following recursive estimator

$$\hat{\mathbf{u}}_{0,l} = A^m \hat{\mathbf{u}}_{-m,l} + \mathbf{g}_m \left(\sum_{i=-m+1}^{0} \mathbf{b}_{il}(y_i - B^T A^{m+i} \hat{\mathbf{u}}_{-m,l}) \right). \tag{22}$$

In (22), $\hat{\mathbf{u}}_{0,l}$, $\hat{\mathbf{u}}_{-m,l}$, denote the estimates of the vectors $\mathbf{u}_0$, $\mathbf{u}_{-m}$ given observation data $(y_0, y_{-1}, ..., y_{-l+1})$, $(y_{-m}, y_{-m-1}, ..., y_{-m-l+1})$, respectively. Also, $\{\mathbf{b}_{il}, i=0, ..., -l+1\}$ are the vector-valued coefficients of the linear m-step recursion of the Kalman filter operation on the system (21). Finally, the vector-valued function $\mathbf{g}_m$ is defined as

$$\mathbf{g}_m(\mathbf{x}) = [H(\lambda_{0,m}, x_1), H(\lambda_{-1,m}, x_2), ..., H(\lambda_{-k+1,m}, x_k)]^{\mathrm{T}}$$
$$\mathbf{x} = [x_1, ..., x_k]^{\mathrm{T}}, \tag{23}$$

where

$$\lambda_{-j,m} = c[r_{-j,m}]^{1/2}$$
$$c: \Phi(c) + c^{-1}\phi(c) = \frac{2-\varepsilon}{2(1-\varepsilon)} \tag{24}$$

$r_{-j,m}$: variance gain in estimating x_{-j} given $\{y_i, -m+1 \leqslant i \leqslant 0\}$ under nominal conditions.

$H(\cdot, \cdot)$: the Huber function as defined in (15).

From (22), and in view of the definitions (23) and (24), it is evident that every scalar nonlinearity is applied to linear combinations of at most m observation data. Furthermore, if $\varepsilon \to 0$, the positive constants $\{\lambda_{-j,m}, j=0, ..., -k+1\}$ tend to infinity and the estimator in (22) becomes identical to the optimal at the nominal Gaussian model estimator. For $\varepsilon > 0$, the above constants are all finite and they determine the amount of limiting which is introduced in each entry of the innovations term of the Kalman filter.

A filter similar to (22) was earlier considered by Masreliez and Martin (1977), for the case $m = 1$. The above authors applied the nonlinearity on a transformed version of the innovations process. However, their analysis was based on an ad hoc assumption that the process formed by the residuals is Gaussian. Then, using this assumption, they derived a covariance recursion, thus avoiding the problem of nested nonlinearities in the actual nonlinear recursion. Later on, we will numerically demonstrate the performance of the above filter as compared with that in (22), as analyzed by the methods we present in the sequel.

Here, we are primarily interested in the study of the asymptotic properties of the estimator in (22) when the number of observations tends to infinity, and the nominal information process is stationary. The condition for stationarity of the latter process is that all the roots of the polynomial equation

$$\lambda^k - a_1\lambda^{k-1} - \cdots - a_k = 0$$

have magnitudes less than one.

The first issue is the asymptotic outlier resistance of the estimator. Theorem 2 below, whose proof is in the Appendix, establishes that property.

THEOREM 2. *Let $\{X_n\}$ in (19) have finite variance and be stationary. Then, the filter in (22) is asymptotically ($l \to \infty$) outlier resistant for mutually independent m-size batches of outliers.*

The next issue is the asymptotic stationary of the filter itself when $l \to \infty$. In order to study that we consider the residual process

$$\begin{aligned}\mathbf{U}_0 - \hat{\mathbf{U}}_{0,l} = {} & \mathbf{U}_0 - A^m\mathbf{U}_{-m,l} + A^m(\mathbf{U}_{-m} - \hat{\mathbf{U}}_{-m,l}) \\ & - \mathbf{g}_m\left(\sum_{i=-m+1}^{0} \mathbf{b}_{il}((Y_i - B^{\mathrm{T}}A^{m+i}\mathbf{U}_{-m,l}) \right. \\ & \left. + \mathbf{B}^{\mathrm{T}}A^{m+i}(\mathbf{U}_{-m} - \hat{\mathbf{U}}_{-m,l})\right). \end{aligned} \tag{25}$$

For l going to infinity along multiples of m, the above residual process becomes asymptotically stationary. This will be shown by establishing a more general result regarding the asymptotic stationarity of Markov processes with Euclidean state space. The latter result is expressed in Theorem 3 below, whose proof is in the Appendix. In the sequel we denote $\|\mathbf{x}\| \triangleq \max_1 |x_i|$ for $\mathbf{x} = (x_1, ..., x_n)^{\mathrm{T}}$.

THEOREM 3. *Let $f(\mathbf{x}, \mathbf{v}): R^k \times R^l \to R^k$ be measureable. Let $\{\mathbf{X}_n, n \geqslant 0\}$ be a stochastic process in R^k defined by*

$$\mathbf{X}_{n+1} = f(\mathbf{X}_n, \mathbf{V}_n), \qquad n \geqslant 0, \tag{26}$$

where $\{\mathbf{V}_n, n \geqslant 0\}$ is an i.i.d. process in R^l, independent of $\mathbf{X}_0$ with distribution $P(\cdot)$. Then, if there is a positive ζ, such that $\zeta < 1$ and

$$\int_{R^l} \| f(\mathbf{x}, \mathbf{v}) - f(\mathbf{x}', \mathbf{v})\| \, dP(\mathbf{v}) \leqslant \|\mathbf{x} - \mathbf{x}'\| \cdot \zeta, \qquad \forall \mathbf{x}, \mathbf{x}' \in R^k \tag{27}$$

the process $\{\mathbf{X}_n\}$ is asymptotically stationary.

The residual process (25) satisfies the conditions of Theorem 3. This can be shown by using the properties of the nonlinearity $\mathbf{g}_m(\cdot)$, namely that $\|\mathbf{g}_m(x) - \mathbf{g}_m(x')\| \leqslant \|x - x'\|$, and certain standard properties of the linear filtering coefficients $\{\mathbf{b}_{il}\}$ and the stationary matrix A. As a result, the marginal probability density of the residuals converges weakly to a steady state density. The covariance of that steady state density is what we call the asymptotic mean square error induced by the filter (22) at the nominal Gaussian model. In fact, it is even true, as a result of Theorem 3, that the sequence of covariances of the residual process converges to the steady state covariance.

The computation of the steady state covariance is an important component in the study of the asymptotic properties of the proposed filter. It is interesting to point out that the deviation of the robust filter from the nominally optimal linear filter builds up as the number l of observations increases, and we would like to see what the performance is for an asymptotically large number of observations, as compared to the nominally optimal asymptotic performance. The difference in performance will clearly exhibit the price that one has to pay for achieving robustness in this context.

Due to the nature of the nonlinear residual recursion, the computation of the asymptotic covariance is a difficult and tedious task. As analytic, or closed form, expressions seem impossible to obtain, we approached the problem by deriving upper and lower bounds. The derivation was based on the asymptotic stationarity of the residual process, which implies $\lim_{l \to \infty} E\{(\mathbf{U}_0 - \hat{\mathbf{U}}_{0,l})^2\} = \lim_{l \to \infty} E\{(\mathbf{U}_{-m} - \hat{\mathbf{U}}_{-m,l})^2\}$, and the approximation of the square of the second term in (25) by upper and lower quadratic bounds in terms of $\mathbf{U}_{-m} - \hat{\mathbf{U}}_{-m,l}$. The bounds were finally obtained by solving two fixed point matrix equations of the form $X = A^m X (A^m)^T + G(X)$. The two bounds are found to be tight enough and approaching each other as the design parameter m becomes larger, at the exponential rate $|\mu_{\max}(A)|^{2m}$, where $\mu_{\max}(A)$ is the largest magnitude eigenvalue of the matrix A. As a result, a reasonably good estimate of the asymptotic covariance was obtained. We defer the discussion of this issue until Section 6, where the above results are numerically demonstrated and analyzed.

5. Breakdown Point; Influence Function

Let us consider the frequently observed in practice case of independent and additive outliers. In particular, let the noise sequence $\{\ldots, W_{-1}, W_0, W_1, \ldots\}$ be such that each of its elements is generated by the nominal Gaussian noise process, with probability $1-\delta$, and it is instead equal to some deterministic value, v, with probability δ, $0 \leqslant \delta \leqslant 1$. Let the value v occur with probability δ, independently per noise datum. Given the above outlier model, given some asymptotic filtering or smoothing operation, $\hat{X}_0$, let $e(f_0, \delta, v, \hat{X}_0)$ denote the induced mean squared error.

That is, if f_o represents the overall nominal Gaussian model, then, $e(f_o, \delta, v, \hat{X}_0) = E\{(X_0 - \hat{X}_0)^2 \mid f_o, \delta, v\}$. Let us denote, $e(f_o, \delta, \hat{X}_0) \triangleq \lim_{v \to \pm\infty} e(f_o, \delta, v, \hat{X}_0)$, and let there exist some value δ^*, $0 \leqslant \delta^* \leqslant 1$, such that,

$$e(f_o, \delta, \hat{X}_0) > E\{X_0^2 \mid f_o\}; \qquad \forall \delta > \delta^*$$

$$e(f_o, \delta, \hat{X}_0) \leqslant E\{X_0^2 \mid f_o\}; \qquad \forall \delta \leqslant \delta^*.$$

Then, the value δ^* is called the *breakdown point* of the asymptotic operation $\hat{X}_0$. The breakdown point clearly represents the maximum frequency of independent, asymptotically large in amplitude outliers that the operation $\hat{X}_0$ can tolerate, before it becomes worthless, that is, before it starts inducing mean squared error that is larger than that induced when no observation data are available. We note that the breakdown points of the nominally optimal linear filtering and smoothing operations are easily found to equal *zero*.

Let us now consider a generalization of the outlier model presented above. In particular, let us consider the case where independent, size m blocks of outliers may occur. Then, each block occurs with probability δ, and it consists of a value v per datum in the block. Given some filtering or smoothing operation $\hat{X}_0$, we then denote the induced mean squared error, $e_m(f_o, \delta, v, \hat{X}_o)$. Denoting by $e(f_o, \hat{X}_o)$ the mean squared error in the absence of the above outlier model, we denote, $J_{m,\delta}(v) \triangleq e_m(f_o, \delta, v, \hat{X}_0) - e(f_o, \hat{X}_0)$. We call $J_{m,\delta}(v)$ the *variation function at* δ. Given δ, the variation function exhibits the difference between the mean squared error, when the outlier value is v and the frequency of the outlier blocks is δ, and the mean squared error in the absence of outliers. We call $I_{m,\delta}(v) \triangleq \delta^{-1} J_{m,\delta}(v)$, the *normalized variation function at* δ, and we call $I_m(v) \triangleq \lim_{\delta \to 0} I_{m,\delta}(v)$ the *influence function*. The influence function is the slope of the variation function at $\delta = 0$, and it exhibits the effect of the outlier value v, at asymptotically small outlier frequencies δ.

Regarding the computation of the breakdown point and the influence function $I_m(v)$ of the filtering operation in (22), an approach similar to that used for the asymptotic variance was adopted. In particular, upper and lower bounds were computed for both quantities. These bounds approach each other at the same exponential rate $|\mu_{\max}(A)|^{2m}$.

The influence function, $I_m^o(v)$, of the nominally optimal linear filter was also computed for comparison. The latter has a closed form expression which is a quadratic function of the outlier value v,

$$I_m^o(v) = \sum_{i=0}^{\infty} [(I - C) A^m]^i [v^2 \boldsymbol{\mu}\boldsymbol{\mu}^T - \sigma_w^2 N][(A^T)^m (I - C)^T]^i,$$

where

$$C = \sum_{i=-m+1}^{0} b_i B^{\mathrm{T}} A^i, \; (b_i = \lim_{l \to \infty} b_{il})$$

$$\boldsymbol{\mu} = \sum_{i=-m+1}^{0} b_i$$

$$N = \sum_{i=-m+1}^{0} \mathbf{b}_i \mathbf{b}_i^{\mathrm{T}}.$$

6. Numerical Results

In this section we present some numerical results regarding the asymptotic performance of the filtering operation in (22), for two special cases of the nominal model (20).

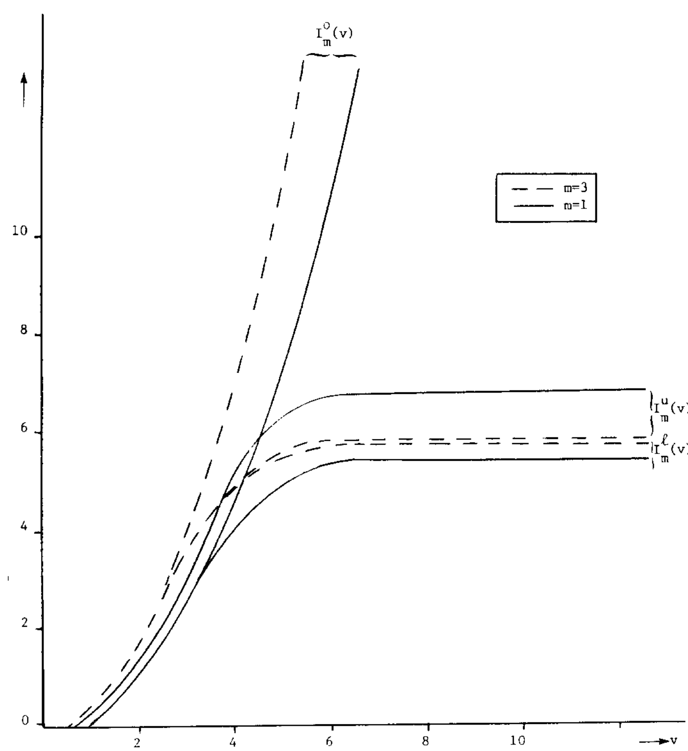

Fig. 1. Bounds on the influence function, Model 1. Causal filtering operation in (22); $\varepsilon = 0.002$; $I_m^0(v)$: influence function induced by the optimal at the nominal model filter.

TABLE I

Bounds on the Asymptotic Mean Squared Error, at the Nominal Model.

ε \ m	1	2	3	4	5	6
0.002	0.53167	0.53284	0.53333	0.53346	0.53350	0.53351
	0.66941	0.56629	0.54159	0.53552	0.53401	0.53364
0.01	0.53488	0.53963	0.54136	0.54183	0.54195	0.54198
	0.67108	0.57247	0.54945	0.54385	0.54246	0.54211
0.1	0.58157	0.608293	0.61640	0.01851	0.61904	0.61917
	0.70620	0.63797	0.62328	0.62032	0.61949	0.61929
0.15	0.60983	0.64401	0.65401	0.65659	0.65723	0.65740
	0.72961	0.67249	0.66099	0.65832	0.65767	0.65750
0.25	0.66941	0.71376	0.72608	0.72921	0.72999	0.73019
	0.78026	0.73998	0.73249	0.73080	0.73039	0.73028
0.3	0.70079	0.74848	0.76146	0.76474	0.76556	0.76576
	0.80718	0.77357	0.76758	0.76626	0.76594	0.76586
0.4	0.76727	0.81887	0.83243	0.83582	0.83667	0.83688
	0.86426	0.84156	0.83795	0.83719	0.83701	0.83697

Note. Model 1. Causal filtering operation in (22). Asymptotic mean squared error induced by the optimal at the nominal model causal filter = 0.53112. Upper lines: lower bounds.

TABLE II

Bounds on the Breakdown Point

ε \ m	1	2	3	4	5	6
0.002	0.09928	0.06814	0.04932	0.03788	0.03056	0.02556
	0.14352	0.07476	0.05048	0.03811	0.03060	0.02557
0.01	0.14699	0.10040	0.07274	0.05597	0.04522	0.03786
	0.20676	0.10942	0.07433	0.05628	0.04528	0.03788
0.1	0.32204	0.21602	0.15715	0.12180	0.09898	0.08326
	0.40878	0.22978	0.15974	0.12228	0.09908	0.08328
0.15	0.38225	0.25595	0.18674	0.14516	0.11824	0.09962
	0.47011	0.27034	0.18937	0.14568	0.11835	0.09964
0.25	0.48129	0.32349	0.23761	0.18576	0.15194	0.12840
	0.56488	0.33815	0.24036	0.18631	0.15205	0.12842
0.3	0.52466	0.35423	0.26119	0.20477	0.16783	0.14203
	0.60450	0.36875	0.26395	0.20533	0.16795	0.14206
0.4	0.60417	0.41329	0.30739	0.24246	0.19955	0.16937
	0.67478	0.42723	0.31012	0.24301	0.19967	0.16940

Note. Model 1. Causal filtering operation in (22). Independent per datum outliers. Upper lines: lower bounds.

TABLE III
Bounds on the Breakdown Point

ε \ m	1	2	3	4	5	6
0.002	0.09928	0.13164	0.14080	0.14315	0.14315	0.14389
	0.14352	0.14394	0.14394	0.14394	0.14394	0.14394
0.01	0.14699	0.19073	0.20274	0.20580	0.20656	0.20676
	0.20676	0.20686	0.20682	0.20682	0.20682	0.20682
0.1	0.32204	0.38537	0.40125	0.40520	0.40618	0.40643
	0.40878	0.40676	0.40653	0.40651	0.40651	0.40651
0.15	0.38225	0.44639	0.46212	0.46601	0.46698	0.46723
	0.47011	0.46759	0.46733	0.46731	0.46731	0.46731
0.25	0.48129	0.54234	0.55688	0.56045	0.56134	0.56156
	0.56488	0.56195	0.56166	0.56164	0.56164	0.56164
0.3	0.52466	0.58298	0.56672	0.60010	0.60093	0.60114
	0.60450	0.60152	0.60124	0.60121	0.60121	0.60121
0.4	0.60417	0.65577	0.66775	0.67067	0.67140	0.67158
	0.67478	0.67193	0.67166	0.67164	0.67164	0.67164

Note. Model 1. Causal filtering operation in (22). Independent size-m batches of outliers. Upper lines: lower bounds.

TABLE IV
Bounds on the Asymptotic Mean Squared Error at the Nominal Model.

ε \ m	1	2	3	4	5	6
0.002	0.55402	0.57594	0.59937	0.61040	0.61445	0.61566
	0.83214	0.68407	0.63361	0.62154	0.61764	0.61658
0.01	0.57548	0.62504	0.66518	0.68180	0.68763	0.68936
	0.86214	0.73994	0.70200	0.69383	0.69109	0.69035
0.1	0.62110	0.69155	0.72865	0.74097	0.74499	0.74615
	0.89436	0.79589	0.76040	0.75110	0.74788	0.74698
0.15	0.65204	0.72942	0.74120	0.77011	0.77401	0.77432
	0.94013	0.83110	0.79568	0.78320	0.77516	0.77501
0.25	0.69875	0.73479	0.76678	0.78133	0.79002	0.79012
	0.95182	0.86264	0.80203	0.79312	0.79202	0.79136
0.3	0.73478	0.73930	0.78033	0.79300	0.80400	0.80511
	0.96067	0.91011	0.86481	0.82414	0.80923	0.80547
0.4	0.73510	0.74902	0.79087	0.81142	0.82267	0.82320
	0.97033	0.91437	0.86690	0.83571	0.82610	0.82359

Note. Model 2. Causal filtering operation in (22). Asymptotic mean squared error induced by the optimal at the nominal model causal filter = 0.54731. Upper lines: lower bounds.

TABLE V

Bounds on the Breakdown Point

ε \ m	1	2	3	4	5	6
0.002	0.07594	0.05802	0.04513	0.035395	0.028765	0.02411
	0.13890	0.07501	0.04960	0.035980	0.029010	0.02486
0.01	0.11029	0.08225	0.06334	0.04958	0.04030	0.03380
	0.20020	0.11510	0.08010	0.05156	0.0450	0.03388
0.1	0.25689	0.18640	0.14353	0.11313	0.09248	0.07790
	0.39537	0.22540	0.15003	0.11804	0.09424	0.07823
0.15	0.32899	0.23552	0.18083	0.14286	0.11705	0.09878
	0.47563	0.27100	0.20242	0.14811	0.11829	0.09890
0.25	0.47094	0.33123	0.25350	0.20121	0.16563	0.12747
	0.60225	0.39693	0.26089	0.20541	0.16735	0.12784
0.3	0.53838	0.37811	0.28952	0.23047	0.19020	0.16150
	0.65802	0.42004	0.29457	0.23215	0.19082	0.16195
0.4	0.66166	0.47002	0.36191	0.29019	0.24090	0.20548
	0.75106	0.52401	0.39102	0.30016	0.24210	0.20602

Note. Model 2. Causal filtering operation in (22). Independent per datum outliers. Upper lines: lower bounds.

TABLE VI

Bounds on the Breakdown Point

ε \ m	1	2	3	4	5	6
0.002	0.07594	0.11269	0.12939	0.13424	0.13578	0.13622
	0.13890	0.13995	0.14500	0.13952	0.13595	0.13682
0.01	0.11029	0.15774	0.17826	0.18406	0.18592	0.18643
	0.20020	0.19500	0.19851	0.18820	0.18683	0.18682
0.1	0.25689	0.33813	0.37173	0.38136	0.38444	0.38530
	0.39537	0.39220	0.39104	0.38804	0.38740	0.38607
0.15	0.32899	0.41556	0.45031	0.46022	0.46336	0.46424
	0.47563	0.47215	0.46903	0.46630	0.46502	0.46482
0.25	0.47094	0.55275	0.58401	0.59287	0.59561	0.59820
	0.60225	0.60112	0.60039	0.60004	0.59970	0.59918
0.3	0.53838	0.61325	0.64137	0.64932	0.65175	0.65244
	0.65802	0.65720	0.65695	0.65530	0.65398	0.65307
0.4	0.66166	0.71912	0.74020	0.74616	0.74795	0.74845
	0.75106	0.75083	0.75010	0.74970	0.74912	0.74887

Note. Model 2. Causal filtering operation in (22). Independent size-m batches of outliers. Upper lines: lower bounds.

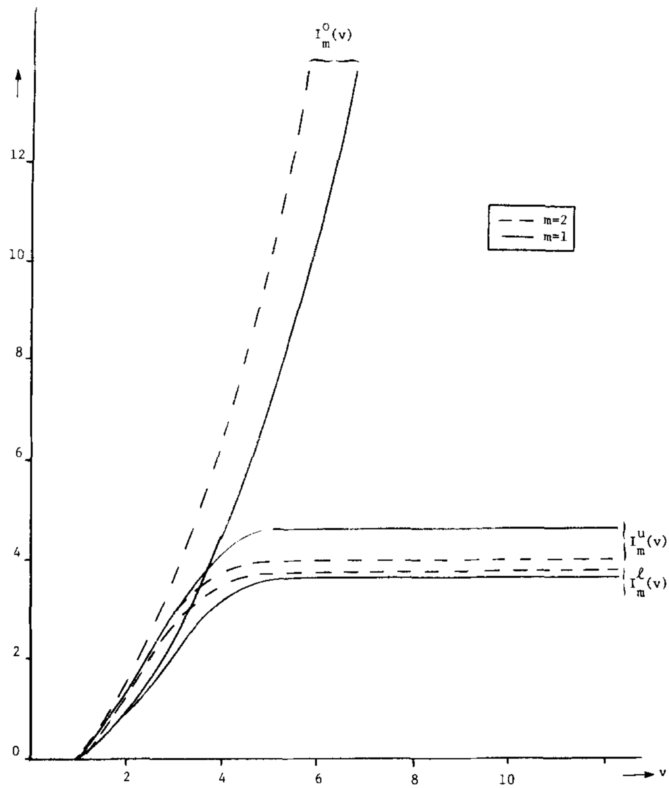

FIG. 2. Bounds on the influence function, Model 1. Causal filtering operation in (22); $\varepsilon = 0.01$; $I_m^0(v)$: influence function induced by the optimal at the nominal model filter.

MODEL 1. First-order autoregressive with autoregressive parameter $\alpha = 0.5$, and $\sigma_x^2 = \sigma_w^2 = 1$.

MODEL 2. Third-order autoregressive with $a_1 = 0.6$, $a_2 = 0.07$, $a_3 = -0.06$, and $\sigma_x^2 = \sigma_w^2 = 1$.

Tables I, II, and III and Figs. 1 and 2 exhibit the performance of the filtering operation in (22), for various values of the design parameters ε and m, when the nominal model is Model 1. When the nominal model is instead Model 2, the corresponding performance is exhibited in Tables IV, V, and VI and Fig. 3. Tables II and V correspond to independent per datum outliers, while Tables III and VI correspond to independent m-size batches of outliers.

TABLE VII

Comparison of Asymptotic Mean Square Error Bounds between Filtering Operation in (22) and the Filter by Masreliez and Martin

ε	A	B
0.002	0.53167	0.53283
	0.66941	0.66992
0.01	0.53488	0.53934
	0.67108	0.67378
0.1	0.58157	0.60346
	0.70620	0.72394
0.15	0.60983	0.63695
	0.72961	0.75215
0.25	0.66941	0.70312
	0.78026	0.80876
0.3	0.70079	0.73546
	0.80718	0.83740
0.4	0.76727	0.80487
	0.86426	0.89603

Note. Model 1. Optimal at the nominal: 0.53112: A: Causal filtering operation in (22), $m = 1$; B: Filter by Masrelize and Martin. Upper lines: lower bounds.

TABLE VIII

Comparison of Breakdown Point Bounds

ε	A	B
0.002	0.09928	0.12240
	0.14352	0.17464
0.01	0.14699	0.17853
	0.20676	0.24633
0.1	0.32204	0.36890
	0.40878	0.45648
0.15	0.38225	0.42999
	0.47011	0.51629
0.25	0.48129	0.52707
	0.56488	0.60631
0.3	0.52466	0.56851
	0.60450	0.64324
0.4	0.60417	0.64312
	0.67478	0.70798

Note. Model 1: A: Causal filtering operation in (22), $m = 1$; B: Filter by Masrelize and Martin. Upper lines: lower bounds.

TABLE IX

Comparison of Asymptotic Mean Square Error Bounds between Filtering Operation in (22) and the Filter by Masreliez and Martin.

ε	A	B
0.002	0.53284	0.53283
	0.56629	0.66992
0.01	0.53963	0.53934
	0.57247	0.67378
0.1	0.60829	0.60346
	0.63797	0.72394
0.15	0.64401	0.63695
	0.67249	0.75215
0.25	0.71376	0.70312
	0.73998	0.80876
0.3	0.74848	0.73646
	0.77357	0.83740
0.4	0.81887	0.80487
	0.84156	0.89603

Note. Model 1. Optimal at the nominal error: 0.53112: A: Filtering operation in (22), $m = 2$; B: Filter by Masreliz and Martin. Upper lines: lower bounds.

TABLE X

Comparison of Breakdown Point Bounds

ε	A	B
0.002	0.13164	0.12240
	0.14394	0.17464
0.01	0.19073	0.17853
	0.20686	0.24633
0.1	0.38537	0.36890
	0.40676	0.45648
0.15	0.44693	0.42999
	0.46759	0.51629
0.25	0.54234	0.52707
	0.56195	0.60631
0.3	0.58298	0.56851
	0.60152	0.64324
0.4	0.65577	0.64312
	0.67193	0.70798

Note. Model 1. Size-m batches of outliers: A: Causal filtering operation in (22), $m = 2$; B: Filter by Masrelize and Martin. Upper lines: lower bounds.

Both the upper and the lower bounds of the asymptotic at the nominal mean squared error (Tables I and IV) are monotonically increasing when the contamination parameter ε increases, for any fixed m. Moreover, for fixed ε, particularly for small values of ε, the upper bounds of the asymptotic error decrease sharply when m increases, while the corresponding lower bounds experience relatively smaller variations with m. Regarding the breakdown point (Tables II, III, V, and VI), we first observe that both upper and lower bounds increase when ε increases, for any fixed m. For the case of independent per datum outliers, the upper and lower bounds of the breakdown point decrease when m increases. On the contrary, when independent m-size batches of outliers are acting, the lower bounds of the corresponding breakdown point increase with m, while the upper bounds remain practically constant. Finally, the upper and lower bounds of the influence function of the filtering operation in (22) are

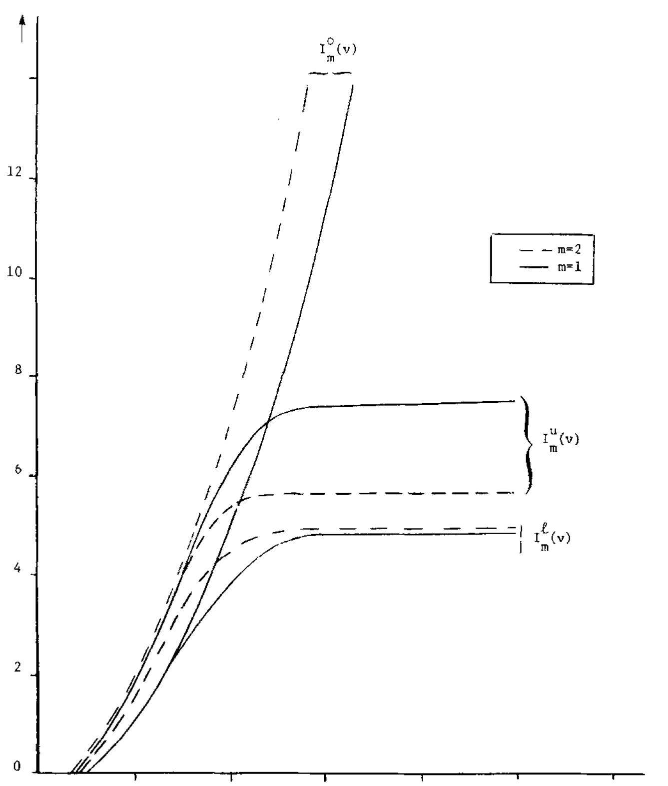

FIG. 3. Bounds on the influence function, Model 2. Causal filtering operation in (22); $\varepsilon = 0.01$; $I_m^0(v)$: influence function induced by the optimal at the nominal model filter.

always monotonically increasing and bounded, as can be seen from Figs. 1, 2, and 3. They both reach certain saturation points depending on ε and m, and, for fixed m, these saturation points are decreasing when ε increases. In all the above cases and for all values of ε, the upper and lower bounds tend to become equal for large m, permitting thus a more accurate evaluation of the performance measures of the filtering operation in (22).

The filtering operation in (22) can combine close to optimal at the nominal model performance, together with good protection against outliers. In addition, this operation is more appropriate for protection against independent batches of outliers. Similar results are drawn when the order of the nominal autoregressive model in (20) is some arbitrary integer k.

Using the concepts and methods that we developed in previous sections, we analyzed the asymptotic performance of the filter proposed by

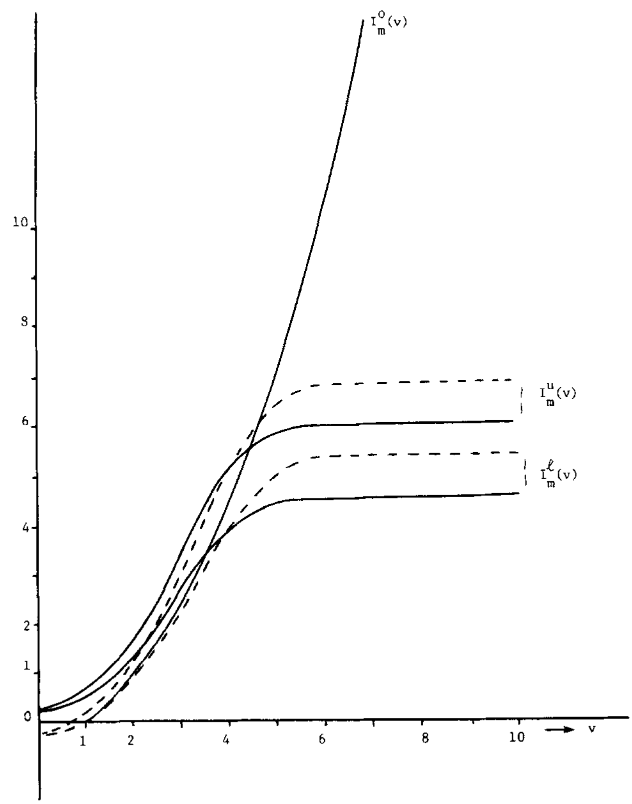

FIG. 4. Bounds on the influence function, Model 1: $m = 1$; $\varepsilon = 0.002$; - - - causal filtering operation in (22); ——— filter by Masreliez and Martin.

Masreliez and Martin when it operates on a stationary environment. In Tables VII and VIII, the asymptotic mean square error bounds and the breakdown point bounds of the latter filter are shown (column B) versus the corresponding bounds for the filter in (22) presented here. In Figs. 4 and 5 the same comparison is made for the influence functions of the two filters. Both filters were assumed to operate on the same process which was taken here to be Model 1, and for $m = 1$. It is observed that the mean square error bounds of the filter (22) are uniformly better than those of the Masreliez and Martin filter (Table VII), at the expense of lower breakdown points (Table VIII) and higher saturation points of the influence functions. However, for $m = 2$, it can be clearly seen from Tables IX and X, that the breakdown points of the filter (22) improve considerably while the mean square error remains small, especially for low contamination levels.

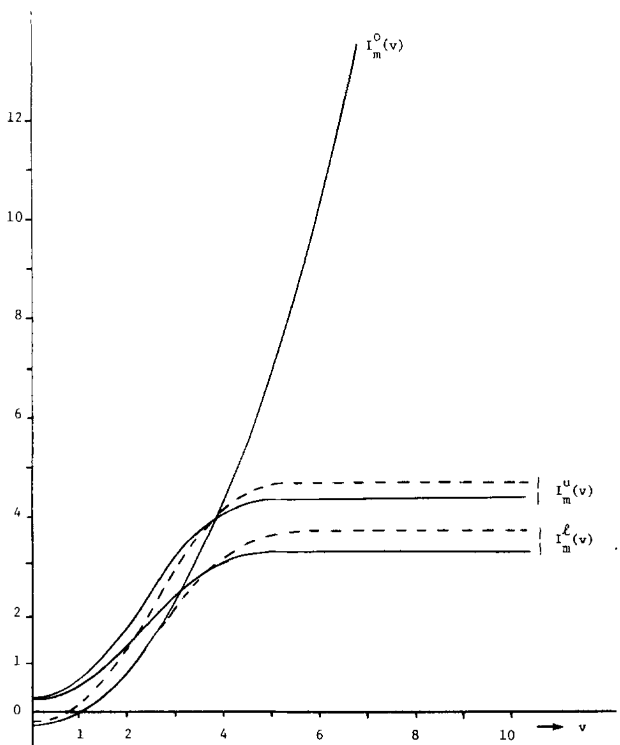

FIG. 5. Bounds on the influence function, Model 1: $m = 1$; $\varepsilon = 0.01$; – – – causal filtering operation in (22); ——— filter by Masreliez and Martin.

7. Conclusions

We designed and analyzed nonlinear filtering and smoothing operations that were found to provide effective resistance to outliers and simultaneously good performance at the nominal Gaussian model. The proposed estimators can be easily implemented, being only slightly more complex (in implementation) than the usual linear estimators. However, the analysis and evaluation of their asymptotic performance were considerably more involved than that for linear estimators, both from a theoretical and a computational point of view.

Due to the nonlinear recursion which is involved in (22), an exact covariance recursion is not possible. So, it was necessary to study the entire functional recursion of probability distributions. Then, we proved asymptotic stationary of the residual process by establishing a more general result concerning the asymptotic stationarity of Markov processes with Euclidean state space.

For the proposed estimators, strong robustness and good performance at the nominal are conflicting requirements. The more robust an estimator is, the worse performance it has, and vice versa. The trade-off between robustness and performance has to be adjusted for each particular problem by appropriately varying the design parameters ε and m, according to the specific requirements and the available knowledge about the underlying situation.

Appendix: Proof of Theorem 1

We first prove that if the optimization problem

$$\inf_{f_Y \in F_Y^l(\varepsilon)} I(f_Y)$$

has a solution, it is unique. Indeed, let f_1, f_2 be two l-dimensional densities in $F_Y^l(\varepsilon)$ attaining the infimum. Then, since $I(\cdot)$ is convex, any density f_δ of the form

$$f_\delta = (1-\delta) f_1 + \delta f_2, \qquad 0 \leqslant \delta \leqslant 1$$

must attain the same infimum. Thus, $I(f_\delta)$ is constant for $0 \leqslant \delta \leqslant 1$. It is implied that

$$0 = \frac{d^2 I(f_\delta)}{d\delta^2} = 2 \int_{R^l} \frac{((1-\varepsilon) f_{oY}(y^l)(P^{\mathrm{T}} y^l))^2 \, (f_2(y^l) - f_1(y^l))^2}{(f_\delta(y^l))^3} \, dy^l, \quad \text{(A.1)}$$

where the differentiation under the integral sign is justified by the dominated convergence theorem (observe that $f_\delta \geqslant (1-\varepsilon) f_{oY} > 0$). From (A.1) we conclude that $f_1 = f_2$ a.e. (dy^l), since $P \neq 0$ and the set where $P^T y^l = 0$ is a proper subspace of R^l.

We now prove

$$I(f_Y^*) \leqslant I(f_Y), \qquad \forall f_Y \in \mathscr{F}^l_Y(\varepsilon) \tag{A.2}$$

Let $E_\lambda = \{y^l : |P^T y^l| \leqslant \lambda\}$ and E_λ^c its complement in R^l. Since $f_Y^*(y^l) = (1-\varepsilon) f_{oY}(y^l) \max\{1, |P^T y^l|/\lambda\}$, we have the relationships:

$$\begin{aligned} I(f_Y^*) - I(f_Y) &= \int_{E_\lambda} \frac{[(1-\varepsilon) P^T y^l f_{oY}(y^l)]^2 (f_Y(y^l) - (1-\varepsilon) f_{oY}(y^l))}{f_Y(y^l) f_Y^*(y^l)} dy^l \\ &\quad + \int_{E_\lambda^c} \frac{[(1-\varepsilon) P^T y^l f_{oY}(y^l)]^2 (f_Y(y^l) - f_Y^*(y^l))}{f_Y(y^l) f_Y^*(y^l)} dy^l \\ &\leqslant \lambda^2 \int_{R^l} \frac{f_Y^*(y^l)}{f_Y(y^l)} (f_Y(y^l) - f_Y^*(y^l))\, dy^l \\ &= \lambda^2 \left(1 - \int_{R^l} \frac{(f_Y^*(y^l))^2}{f_Y^*(y^l)} dy^l\right) \leqslant 0. \end{aligned}$$

The inequality in (A.2) follows from the above relationships. The expressions in (18), determining the value of the constant λ, evolve from the requirement that $\int_{R^l} f_Y^*(y^l)\, dy^l = 1$.

Finally, the form of the robust estimator $\hat{X}_o^*(y^l)$ is equal to the conditional expectation $E(X_0/y^l)$ at the least favorable density $f^*(x_o, y^l)$,

$$\hat{X}_0^*(y^l) = \int_{R^l} x_0 \frac{f^*(x_0, y^l)}{f_Y^*(y^l)} dx_0, \tag{A.3}$$

where

$$f^*(x_0, y^l) = (1-\varepsilon) f_o(x_0, y^l) + \varepsilon f_{os}(x_0)\, h^*(y^l). \tag{A.4}$$

Substituting (A.4) into (A.3) and recalling that $\int_{R^l} x_0 f_{os}(x_0)\, dx_0 = 0$ and $\int_{R^l} f_o(x_0, y^l)\, dx_o = f_{oY}(y^l)\, P^T y^l$, we obtain

$$\begin{aligned} \hat{X}_o^*(y^l) &= \frac{(1-\varepsilon) P^T y^l f_{oY}(y^l)}{f_Y^*(y^l)} \\ &= \begin{cases} P^T y^l, & \text{for } |P^T y^l)| \leqslant \lambda \\ \lambda \operatorname{sgn}(P^T y^l), & \text{for } |P^T y^l| > \lambda \end{cases} = H(\lambda, P^T y^l), \end{aligned}$$

Proof of Theorem 2. The operation in (22) has the general form, $\hat{X}_n = \sum_i \alpha_i \hat{X}_i + g(y^m, \sum_i \alpha_i \hat{X}_i, \{\hat{X}_j\})$, where, for some bounded λ,

$$g(x) = \begin{cases} x; & |x| \leqslant \lambda \\ \lambda \operatorname{sgn} x; & |x| > \lambda \end{cases},$$

and where $|\sum_i a_i| \leqslant c$, for some given $c > 0$. Therefore,

$$|\hat{X}_n| \leqslant \lambda \left[1 + \left| \sum_i a_i \right| \right] \leqslant \lambda(c+1); \forall n. \tag{A.5}$$

Let $E_{\mu_o}\{[X_n - \hat{X}_n]^2\}$ denote the mean squared error induced by the estimate $\hat{X}_n$, when the Gaussian nominal observation process is acting. Let $E_m\{[X_n - \hat{X}_n]^2\}$ be the same error, when some process in class $\mathscr{F}^m$ is acting instead. Let y^n and z^n denote sequences that are respectively generated by the processes μ_o and μ. Given some set A^n in R^n, and in connection with (A.5) and the Schwartz inequality, we have

$$\begin{aligned} E_\mu\{[X_n - \hat{X}_n]^2 \mid z^n \in A^n\} &= E_{\mu_o}\{X_n^2 \mid z^n \in A^n\} - 2E_\mu\{X_n \hat{X}_n \mid z^n \in A^n\} \\ &\quad + E_\mu\{[\hat{X}_n]^2 \mid z^n \in A^n\} \\ &\leqslant c + 2E^{1/2}\{X_n^2 \mid z^n \in A^n\}\, E^{1/2}\{[\hat{X}_n]^2 \mid z^n \in A^n\} + E_\mu\{[\hat{X}_n]^2 \mid z^n \in A^n\} \\ &\leqslant c + 2\lambda c^{1/2}(c+1) + \lambda^2(c+1)^2 = [c^{1/2} + \lambda(c+1)]^2 \triangleq C. \end{aligned} \tag{A.6}$$

Due to (A.6) and considering ergodic and stationary observation processes in conjunction with $\mathscr{F}^m$, we obtain: given $\eta > 0$, there exists n_o, such that

$$\begin{aligned} &\forall n > n_o; E_\mu\{[X_n - \hat{X}_n]^2\} \\ &\quad \leqslant (1 - \varepsilon + \eta)\, E\{[X_n - \hat{X}_n]^2 \mid [\# i : \gamma_m(z_{i+1}^{i+m}, y_{i+1}^{i+m}) > \varepsilon] \\ &\quad \leqslant n\varepsilon, y^n \varepsilon R^n\} + \varepsilon C, \end{aligned} \tag{A.7}$$

where, for independent m-size outliers, there exists some $\varepsilon_o > 0$, such that

$$\begin{aligned} &E\{[X_n - \hat{X}_n]^2 \mid [\# i : \gamma_m(z_{i+1}^{i+m}, y_{i+1}^{i+m}) > \varepsilon] \leqslant n\varepsilon, \quad y^n \in R^n\} \\ &\quad \leqslant E_{\mu_o}\{[X_n - \hat{X}_n]^2\} + \varepsilon C; \qquad \forall \varepsilon < \varepsilon_0, \forall n > n_0. \end{aligned} \tag{A.8}$$

From (A.7) and (A.8) we conclude: Given $\eta = \varepsilon/2$, there exist n_o and $\varepsilon > 0$, such that

$$|\varepsilon_\mu\{[X_n - \hat{X}_n]^2\} - E_{\mu_o}\{[X_n - \hat{X}_n]^2\}|$$
$$\leqslant \varepsilon\left(2 - \frac{\varepsilon}{2}\right) C + \frac{\varepsilon}{2} E_{\mu_o}\{[X_n - \hat{X}_n]^2\} \leqslant \varepsilon \frac{5}{2} C \triangleq \delta; \qquad \forall n > n_o, \forall \varepsilon < \varepsilon_o.$$

Thus, given $\delta > 0$, there exist, n_o, and $\varepsilon : 0 < \varepsilon < \min(\varepsilon_o, 2\delta/5C)$, such that

$$\Pi_{n,\rho_n}(\mu_o, \mu) < \varepsilon \qquad \text{implies } |E_\mu\{[X_n - \hat{X}_n]^2\} - E_{\mu_o}\{[X_n - \hat{X}_n]^2\}| < \delta; \quad \forall n > n_o$$

The proof of theorem is now complete.

Proof of Theorem 3. From (26) we conclude that $\{\mathbf{X}_n\}$ is a Markov process. Thus, to prove asymptotic stationary, it suffices to show that, given any distribution for $\mathbf{X}_0$, the distribution of $\mathbf{X}_n$ converges weakly to a unique distribution in R^k, as $n \to \infty$.

Let $\mu_o(\mathbf{x})$ be an arbitrary density function, $\forall \mathbf{x} \in R^k$. Let then the sequence $(\mu_n(\mathbf{x}), n \geqslant 0)$ be defined as

$$\mu_{n+1}(\mathbf{x}) = \int_{R^k} \Lambda(\mathbf{x}, \boldsymbol{\omega})\, \mu_n(\boldsymbol{\omega})\, d\boldsymbol{\omega}, \tag{A.9}$$

where $\Lambda(\mathbf{x}, \boldsymbol{\omega})$ denotes the conditional density function of $\mathbf{x}$, given $\boldsymbol{\omega}$, when $\mathbf{x} = f(\boldsymbol{\omega}, \mathbf{v})$, and where $\boldsymbol{\omega}$ is independent of $\mathbf{v}$, and $p(\mathbf{v})$ is the density function of $\mathbf{v}$ at $\mathbf{v} \in R^l$. Let us now define the sequence $\{\Lambda^{(n)}(\mathbf{x}, \boldsymbol{\omega}), n \geqslant 1\}$, as

$$\begin{aligned} \Lambda^{(1)}(\mathbf{x}, \boldsymbol{\omega}) &= \Lambda(\mathbf{x}, \boldsymbol{\omega}) \\ \Lambda^{(n+1)}(\mathbf{x}, \boldsymbol{\omega}) &= \int_{R^k} \Lambda^{(n)}(\mathbf{x}, \mathbf{z})\, \Lambda^{(1)}(\mathbf{z}, \boldsymbol{\omega})\, d\mathbf{z}. \end{aligned} \tag{A.10}$$

Then, we can write

$$\mu_n(\mathbf{x}) = \int_{R^k} \Lambda^{(n)}(\mathbf{x}, \boldsymbol{\omega})\, \mu_0(\boldsymbol{\omega})\, d\boldsymbol{\omega}. \tag{A.11}$$

To show weak convergence of the sequence $\{\mu_n(\mathbf{x})\}$, we need to show that there exists a density function $\mu(\mathbf{x})$; $\mathbf{x} \in R^k$, such that, for any continuous and bounded function, $g(\mathbf{x})$ in R^k, we have

$$\int_{R^k} g(\mathbf{x})\, \mu_n(\mathbf{x})\, d\mathbf{x} \xrightarrow[n \to \infty]{} \int_{R^k} g(\mathbf{x})\, \mu(\mathbf{x})\, d\mathbf{x}. \tag{A.12}$$

Let us define the sequence $\{g_n(\mathbf{x}), n \geqslant 0\}$, as

$$g_0(\mathbf{x}) = g(\mathbf{x})$$
$$g_n(\mathbf{x}) = \int_{R^k} \Lambda(\mathbf{z}, \mathbf{x})\, g_{n-1}(\mathbf{z})\, d\mathbf{z}. \tag{A.13}$$

Then

$$\int_{R^k} g(\mathbf{x})\, \mu_n(\mathbf{x})\, d\mathbf{x} = \int_{R^k} g_n(\mathbf{x})\, \mu_0(\mathbf{x})\, d\mathbf{x}. \tag{A.14}$$

Let us define

$$u_n \triangleq \sup_{\delta > 0} \left(\delta^{-1} \sup_{\|\boldsymbol{\omega} - \mathbf{z}\| < \delta} | g_n(\mathbf{z}) - g_n(\boldsymbol{\omega}) | \right). \tag{A.15}$$

Without lack of generality, we will assume that the quantities $\{u_n\}$ are all finite. (This is true if, for example, the functions $g_n(\mathbf{x})$ satisfy a Lipshitz condition.) From (26) and (A.13) we obtain

$$g_n(\mathbf{x}) = \int_{R^k} g_{n-1}(f(\mathbf{x}, \mathbf{v}))\, p(\mathbf{v})\, d\mathbf{v}. \tag{A.16}$$

From (A.15) and (A.16) we conclude

$$\begin{aligned} u_n &\leqslant \int_{R^k} \left\{ \sup_{\delta > 0} \left(\delta^{-1} \sup_{\|\mathbf{x} - \boldsymbol{\omega}\| < \delta} | g_{n-1}(f(\mathbf{x}, \mathbf{v})) - g_{n-1}(f(\boldsymbol{\omega}, \mathbf{v}))| \right) \right\} p(\mathbf{v})\, d\mathbf{v} \\ &\leqslant \int_{R^k} h(\mathbf{v})\, p(\mathbf{v}) \left\{ \sup_{\delta > 0} \left([\delta h(\mathbf{v})]^{-1} \sup_{\|\mathbf{x} - \boldsymbol{\omega}\| < \delta h(\mathbf{v})} | g_{n-1}(\mathbf{x}) - g_{n-1}(\boldsymbol{\omega})| \right) \right\} d\mathbf{v} \\ &= u_{n-1} \int_{R^k} h(\mathbf{v})\, p(\mathbf{v})\, d\mathbf{v} = \zeta u_{n-1}, \qquad \zeta < 1. \end{aligned} \tag{A.17}$$

From (A.17), we conclude that $u_n \to 0$, as $n \to \infty$, and that $g_n(\mathbf{x}) \to g(\mathbf{x}) = \text{constant}$ on R^k, as $n \to \infty$. Thus,

$$\int_{R^k} g(\mathbf{x})\, \mu_n(\mathbf{x})\, d\mathbf{x} = \int_{R^k} g_n(\mathbf{x})\, \mu_0(\mathbf{x})\, d\mathbf{x} \xrightarrow[n \to \infty]{} \text{constant}. \tag{A.18}$$

Due to (27), the sequence $\{\mu_n(\mathbf{x})\}$ is tight. Thus, there exists a subsequence $\{\mu_{n_i}(\mathbf{x})\}$, and a density function $\mu(\mathbf{x})$ in R^k, such that for every continuous and bounded function $g(\mathbf{x})$, we have

$$\int_{R^k} \mu_{n_i}(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x} \xrightarrow[n_i \to \infty]{} \int \mu(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x}. \tag{A.19}$$

From (A.18) and (A.19) immediately follows that

$$\int_{R^k} g(\mathbf{x})\,\mu_n(\mathbf{x})\,d\mathbf{x} \xrightarrow[n\to\infty]{} \int_{R^k} g(\mathbf{x})\,\mu(\mathbf{x})\,d\mathbf{x}$$

and the proof of the theorem is now complete.

RECEIVED May 15, 1987; ACCEPTED February 16, 1988

REFERENCES

BOENTE, G., FRAIMAN, R., AND YOHAI, V. (1982), "Qualitative robustness for general stochastic processes," Dept. of Statistics, Tech. Report, No. 26, Univ. of Washington, Seattle.

BUSTOS, O., FRAIMAN, R., AND YOHAI, V. (1984), Asymptotic behavior of the estimates based on residual autocovariances for ARMA models *in* "Robest and Nonlinear Time Series Analysis," Lecture Notes in Statistics Vol. 26, pp. 26–49, Springer-Verlag, New York.

COX, D. (1978), "Metrics on Stochastic Processes and Qualitative Robustness," Dept. of Statistics Techn. Report No. 23, Univ. of Washington, Seattle.

HAMPEL, F. R. (1971), A general qualitative definition of robustness, *Ann. Math. Statist.* **42**, 1887–1895.

HAMPEL, F. R. (1974), The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.* **69**, 383–394.

HAMPEL, F. R., *et al.* (1986), "Robust Statistics. The Approach Based on Influence Functions," Wiley, New York.

HUBER, P. J. (1964), Robust estimation of a location parameter, *Ann. Math. Statist.* **35**, 73–101.

HUBER, P. J. (1981, "Robust Statistics" Wiley, New York.

KALMAN, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Engrg* (ASME) **82D**, 35–45.

KALMAN, R. E. (1963), New methods in Weiner filtering, *in* "Proceedings, First Symposium on Engineering Applictions of Random Function Theory and Probability," Chap. 9, Wiley, New York.

KOLMOGOROV, A. N. (1941), Stationary sequences in Hilbert Space, *Bull. Math. Univ. Moscow* **2**, 40.

MALLOWS, C. J. (1980), Some theory of nonlinear smoothers, *Ann. Statist.* **8** No. 4, 695–715.

MARTIN, R. D., and Yohai, V. J. (1984), "Influence Curves for Time Series," Dept. of Statistics Technical Report No. 51. Univ. of Washington, Seattle.

MARTIN, R. D., AND ZEH, J E. (1977), Determining the character of time series outliers, *in* "Proceedings, American Statistical Assoc."

MASRELIEZ, C. J., AND MARTIN, R. D. (1977), robust bayesian estimation for the linear model and robustifying the Kalman filter, *IEEE Trans. Automat Control* **AC-22**, 361–371.

PAPANTONI-KAZAKOS, P. AND GRAY, R. M. (1979), Robustness of estimators on stationary observations, *Ann. Probab.* **7**, 989–1002.

PAPANTONI-KAZAKOS, P. (1981), Sliding block encoders that are rho-bar continuous functions of their input, *IEEE Trans. Inform. Theory* **IT-27**, 372–376.

PAPANTONI-KAZAKOS, P. (1987), Qualitative robustness in time series, *Inform. and Computation* **72**, No. 3, 239–269.

PAPANTONI-KAZAKOS, P., (1984a), A game theoretic approach to robust filtering, *Inform. and Control* **60**, Nos. 1–3, 168–191.

PAPANTONI-KAZAKOS, P., (1984b), Some aspects of qualitative robustness in time series, *in* Lecture Notes in Statistcs Vol. 26, pp. 218–230, Springer-Verlag, New York.

TSAKNAKIS, H. (1986), "Robust Methods in Time Series," Ph. D. dissertation, Univ. of Connecticut.

WHITTLE, P. (1953), The analysis of multiple stationary time series *J. Roy. Statist. Soc. Ser. B* **15**, 125–139.

WIENER, N. (1949), "Extrapolation, Interpolation and Smoothing of Stationary Time Series, MIT Press, Cambridge.

Printed in Belgium