

# Smoothing of Discontinuous Signals: The Competitive Approach

Maciej Niedźwiecki and William A. Sethares, *Member, IEEE*

**Abstract**—Discontinuous signals buried in noise cannot be recovered by linear filtering methods. This paper presents a new class of nonlinear filters in which sets of forward and backward linear predictors and smoothers compete with each other at each timestep. The winner of each competition is granted the right to produce the smoothed estimate at that timestep. This conceptually simple approach to nonlinear filtering, called the competitive smoothing approach, is justified by considering sets of Kalman filters (corresponding to the hypotheses used in the Bayesian framework) which are used to derive model credibility coefficients. These are shown to essentially “switch” between the various models. We argue that the concept of competitive smoothing is considerably more general than just the Kalman setting, and can be used with almost any predictors and/or smoothers. Several examples are presented which demonstrate the efficacy of the approach at both smoothing and preserving jump discontinuities. Comparisons are made with the other main nonlinear filtering approach, the median filter.

## I. INTRODUCTION

CONSIDER THE PROBLEM of estimating the unknown, possibly discontinuous signal  $x(\cdot)$  using the set of  $N$  observations

$$y(t) = x(t) + n(t), \quad t = 1, \dots, N \quad (1)$$

where  $\{n(t)\}$  is a white (i.i.d.) measurement noise. For “smooth” signals a number of linear techniques can be applied for the purpose of recovering  $\{x(t)\}$ , ranging from simple devices based on local averaging to more sophisticated tools such as Kalman-filter-based smoothing algorithms [1], [2].

Quite obviously, if the signal  $x(t)$  is subject to jump changes the linear smoothing procedures are not a good choice anymore, because all linear filters confuse and remove the high frequency components of the recovered signals along with the noise. For this reason, when applied to processes with jumps they cause a typical effect of “edge blurring.”

The celebrated median filter introduced by Tukey [3] seems to be the simplest nonlinear edge-preserving smoothing device and has been generalized to filters that incorporate rank-order operations (multilevel median filters, order statistic filters, stack filters) as well as to various data-dependent nonlinear schemes (gradient median filters, signal adaptive median filters, adaptive length median filters, trimmed mean filters)—the

Manuscript October 17, 1991; revised March 1, 1993. The associate editor coordinating the review of this paper and approving it for publication was Dr. Robert A. Gabel.

M. Niedźwiecki is with the Department of Automatic Control, Faculty of Electronics, Technical University of Gdańsk, Gdańsk, Poland.

W. A. Sethares is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA.

IEEE Log Number 9406837.

interested reader is referred to the survey paper of Coyle *et al.* [4] for a catalog of median-related filters.

The purpose of the present paper is to introduce and explore a new class of nonlinear edge-preserving filtering algorithms based on the concept of competitive smoothing. One simple version of the competitive smoother consists of two Kalman filters (running forward and backward in time, respectively) coupled in a nonlinear fashion. A brief review of standard Kalman smoothers is given in Section II to set notation and review relevant theoretical work. In contrast with the classical smoothing procedures, the competitive approach assumes that the signal measured at each instant  $t$  can be explained *either* in terms of its past *or* in terms of its future behavior. After placing this in the Bayesian framework of competing model hypotheses (Section III), the appropriate “decision rule” can be derived, resolving the competing elements at each timestep. Finally, Section IV extends the idea of competitive smoothing to a wider class of filters than just Kalman.

## II. REVIEW OF KALMAN-FILTER-BASED SMOOTHING ALGORITHMS

### A. Theoretical Development—From Whittaker to Kitagawa and Gersh

Suppose that the signal which has to be recovered from (1) is “smooth.” In a paper published in 1923, Whittaker [5] suggested that statistically satisfactory results to the signal reconstruction problem can be obtained by appropriately balancing a tradeoff between infidelity to data and infidelity to a smoothness constraint imposed upon  $x(\cdot)$ . By adopting the  $k$ th-order difference equation constraint on  $x(\cdot)$ , the penalty for unsmooth solutions can be expressed in the form

$$\sum_{t=1}^N [\nabla^k x(t)]^2 \quad (2)$$

where  $\nabla x(t) = x(t) - x(t-1)$  is the first-order difference and  $\nabla^k x(t) = \nabla(\nabla^{k-1} x(t))$ .

The estimates  $\hat{x}(1), \dots, \hat{x}(N)$  of the unknown signal samples can then be obtained by means of minimization (with respect to  $x(\cdot)$ ) of the cost function

$$\sum_{t=1}^N [y(t) - x(t)]^2 + \lambda \sum_{t=1}^N [\nabla^k x(t)]^2 \quad (3)$$

where  $\lambda$  is a “smoothness” parameter chosen by the investigator.

The next important step was due to Akaike [6], whose development can be regarded as the Bayesian reinterpretation of Whittaker's scheme (it appears, though, that Akaike was unaware of the problem stated by Whittaker). By regarding  $\lambda$  and  $k$  in (3) as "hyperparameters" with assigned prior distributions, Akaike derived a likelihood function  $L(\lambda, k)$  and obtained estimates of the smoothness parameters rather than choosing them arbitrarily or by means of "visual inspection."

Finally, a significant reduction of the computational complexity of the Whittaker-Akaike scheme was suggested by Kitagawa and Gersch [2]. By embedding the smoothness constraints into a state-space form signal model, Kitagawa and Gersch rephrased the original estimation problem as a problem of smoothing in the state space.

Consider the following fictitious model of signal variation

$$\nabla^k x(t) = \sum_{i=0}^k f_i x(t-i) = e(t) \quad (4)$$

where

$$f_i = (-1)^i \binom{k}{i}$$

and  $\{e(t)\}$  is a white noise sequence independent of  $\{n(t)\}$ .

Since

$$\nabla^k x(t) = 0$$

implies

$$x(t) = \sum_{i=0}^{k-1} a_i t^i$$

(4) can be regarded as a "perturbed" polynomial growth model.

Note that the model (4) and the system equation (1) can be rewritten in state space form as

$$z(t) = Fz(t-1) + Ge(t) \quad (5)$$

$$y(t) = G^T z(t) + n(t) \quad (6)$$

where

$$z(t) = \begin{bmatrix} x(t) \\ x(t-1) \\ \vdots \\ x(t-k+1) \end{bmatrix}$$

and

$$F = \begin{bmatrix} -f_1 & -f_2 & \cdots & -f_{k-1} & -f_k \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Consequently, the problem of estimation of the signal

$$x(t) = G^T z(t)$$

can be restated as a problem of reconstruction of the unknown state trajectory  $\{z(t)\}$ , and can be solved (under Gaussian assumptions) using the Kalman filter/smoothing methodology.

Note that if the noise processes are Gaussian, namely if

$$n(t) \sim \mathcal{N}(0, \rho) \quad e(t) \sim \mathcal{N}(0, \sigma)$$

the Kalman smoothing algorithm can be regarded as a recursive procedure that maximizes the following joint likelihood function

$$\prod_{t=1}^N p(n(t)) \prod_{t=1}^N p(e(t)) \propto \exp \left\{ -\frac{1}{2\rho} \sum_{t=1}^N [y(t) - x(t)]^2 \right\} \cdot \exp \left\{ -\frac{1}{2\sigma} \sum_{t=1}^N [\nabla^k x(t)]^2 \right\}$$

where  $\propto$  denotes proportionality, i.e., the one that minimizes the cost function (3) with

$$\lambda = \frac{\rho}{\sigma}$$

The main advantage of the state-space imbedding approach presented in [2] is that it significantly reduces the computational complexity of Whittaker's scheme (from  $O(N^3)$  to  $O(N)$ ) and it still allows computation of the likelihood of hyperparameters as required in the Akaike.

### B. Kalman Smoother as a Concatenation of Causal and Anticausal Kalman Filters

Out of a large number of different "realizations" of the fixed interval Kalman smoothers (see Anderson and Moore [7]), we will point to one which particularly suits our further purposes. It shows that the optimal smoother can be regarded a concatenation of two Kalman filters running forward and backward in time, respectively.

Let  $\mathcal{Y} = \{y(1), \dots, y(N)\}$  be the set of available measurements and let

$$\mathcal{Y}^-(t) = \{y(1), \dots, y(t)\}$$

$$\mathcal{Y}^+(t) = \{y(t), \dots, y(N)\}$$

denote the sets of "past and present" and "future and present" observations at a given time instant  $t \in [1, \dots, N]$ . It is known that the minimum variance estimate of unknown  $x(t)$  given all the data  $\mathcal{Y}$ , past and future, is given by (see e.g., Lewis [8])

$$\hat{x}(t | N) = E[x(t) | \mathcal{Y}] = \delta(t) [\delta^-(t | t)]^{-1} \hat{x}^-(t | t) + [\delta^+(t | t+1)]^{-1} \hat{x}^+(t | t+1) \quad (7)$$

where

$$\hat{x}^-(t | t) = E[x(t) | \mathcal{Y}^-(t)]$$

denotes the filter running forward in time,

$$\hat{x}^+(t | t+1) = E[x(t) | \mathcal{Y}^+(t+1)]$$

denotes the one-step-ahead predictor running backward in time,  $\delta^-(t | t)$  and  $\delta^+(t | t+1)$  are the corresponding error variances

$$\delta^-(t | t) = \text{var}[\hat{x}^-(t | t) | \mathcal{Y}^-(t)]$$

$$\delta^+(t | t+1) = \text{var}[\hat{x}^+(t | t+1) | \mathcal{Y}^+(t+1)]$$

and

$$\delta(t) = [[\delta^-(t|t)]^{-1} + [\delta^+(t|t+1)]^{-1}]^{-1}.$$

All quantities necessary to compute the smoothed estimate (7) can be obtained using two Kalman filters—one moving forward in time and the other one moving backward in time, respectively.

Let the symbols

$$\oplus \in \{-, +\}, \quad \ominus \equiv -\oplus \in \{+, -\}$$

denote binary operators/labels which will enable us to talk simultaneously about both forward and backward predictors or filters without repetition. For example, the notation

$$\hat{x}^\oplus(t \oplus 1 | t)$$

will be used to combine in one formula  $\hat{x}^-(t+1|t)$  and  $\hat{x}^+(t-1|t)$ .

Let

$$z^\oplus(t) = \begin{bmatrix} x(t) \\ x(t \oplus 1) \\ \vdots \\ x(t \oplus (k-1)) \end{bmatrix}. \quad (8)$$

Note that the forward-time signal evolution equation (4) has its backward time equivalent and that both can be written down in the form

$$\nabla_{\oplus}^k x(t) = e^\oplus(t) \quad (9)$$

where  $\nabla_{\oplus}$  denotes the forward/backward first order difference operator:  $\nabla_{\oplus} x(t) = x(t) - x(t \oplus 1)$  and  $e^-(t) = e(t)$ ,  $e^+(t) = -e(t+k)$ . The corresponding state space models are

$$z^\oplus(t) = Fz^\oplus(t \oplus 1) + Ge^\oplus(t) \quad (10)$$

$$y(t) = G^T z^\oplus(t) + n(t). \quad (11)$$

Finally note that the forward and backward Kalman filters/predictors for the system governed by (10) and (11) take the form

$$\begin{aligned} \hat{z}^\oplus(t|t) &= F\hat{z}^\oplus(t \oplus 1 | t \oplus 1) + V^\oplus(t|t)G\epsilon^\oplus(t) \\ \epsilon^\oplus(t) &= y(t) - G^T \hat{z}^\oplus(t|t \oplus 1) \\ \hat{z}^\oplus(t|t \oplus 1) &= F\hat{z}^\oplus(t \oplus 1 | t \oplus 1) \\ V^\oplus(t|t \oplus 1) &= FV^\oplus(t \oplus 1 | t \oplus 1)F^T + GG^T\lambda \\ V^\oplus(t|t) &= V^\oplus(t|t \oplus 1) \\ &\quad - \frac{V^\oplus(t|t \oplus 1)GG^TV^\oplus(t|t \oplus 1)}{1 + G^TV^\oplus(t|t \oplus 1)G} \end{aligned} \quad (12)$$

where  $V^\oplus(t|t)$ ,  $V^\oplus(t|t \oplus 1)$  are the normalized error covariance matrices

$$V^\oplus(t|t) = \text{cov}[\hat{z}^\oplus(t|t) | \mathcal{Y}^\oplus(t)]/\rho$$

$$V^\oplus(t|t \oplus 1) = \text{cov}[\hat{z}^\oplus(t|t \oplus 1) | \mathcal{Y}^\oplus(t \oplus 1)]/\rho$$

and (cf. (8))

$$\hat{x}^\oplus(t|t) = G^T \hat{z}^\oplus(t|t)$$

$$\hat{x}^\oplus(t|t \oplus 1) = G^T \hat{z}^\oplus(t|t \oplus 1).$$

### III. COMPETITIVE SMOOTHING

#### A. Basic Approach

One of the key assumptions of the ‘‘classical’’ approach to smoothing supposes that the underlying dynamic system is homogenous, implying that the forward-time and backward-time system descriptions (10), (11) must be equivalent. Due to the consistency of forward and backward models, the corresponding filters ‘‘cooperate’’ with each other and (7) can be interpreted as the optimal way of combining the estimates of  $x(t)$  obtained from two independent sets of measurements,  $\mathcal{Y}^-(t)$  and  $\mathcal{Y}^+(t+1)$  [8].

Consider now the case where, at a given time instant  $t$ , only one of the models (10) can describe appropriately the evolution of system parameters, i.e., where  $x(t)$  is consistent with *either* the forward *or* the backward model. The resulting estimation scheme will be referred to as *competitive smoothing*, because it forces the corresponding filters to compete with each other, rather than to cooperate.

Denote by  $T(t)$  the sliding window of width  $2M-1$ , centered at  $t$

$$T(t) = \{t-M+1, \dots, t, \dots, t+M-1\}$$

and let

$$T^\oplus(t \oplus i) = \{t \oplus i, \dots, t \oplus (M-1)\}$$

By  $\mathcal{Y}_T(t) = \{y(j), j \in T(t)\}$  we will denote the data ‘‘framed’’ by the window. Finally, let

$$\mathcal{Y}_T^\oplus(t \oplus i) = \{y(j), j \in T^\oplus(t \oplus i)\}.$$

Consider the following two mutually exclusive hypothetical models of parameter variation

$$\begin{aligned} \nabla_{\oplus}^k x(i) &= e^\oplus(i) \\ \mathcal{H}^\oplus(t) &: \text{var}[e^\oplus(t)] = \sigma \\ & i \in T^\oplus(t). \end{aligned} \quad (13)$$

$\mathcal{H}^+(t)$  is the forward model which attempts to explain the data at time  $t$  based on past observations, while  $\mathcal{H}^-(t)$  is the backwards model that attempts to explain the data at time  $t$  based on future observations. According to the competitive smoothing philosophy, either  $\mathcal{H}^-$  or  $\mathcal{H}^+$  (but not both) describe the present data. We look for the estimate of  $x(t)$  which minimizes the mean square risk function

$$E[(\hat{x}(t) - x(t))^2] \quad (14)$$

where the expectation is taken over  $\mathcal{H}^\oplus(t)$ ,  $\rho$  and all possible data sets  $\mathcal{Y}_T(t)$ .

Using the standard Bayesian approach one can show that the optimal solution is

$$\begin{aligned} \hat{x}(t|T) &= E[x(t) | \mathcal{Y}_T(t)] \\ &= \mu^-(t)E[x(t) | \mathcal{Y}_T(t), \mathcal{H}^-(t)] \\ &\quad + \mu^+(t)E[x(t) | \mathcal{Y}_T(t), \mathcal{H}^+(t)] \end{aligned} \quad (15)$$

where  $\mu^-(t)$  and  $\mu^+(t)$  are the posterior probabilities

$$\mu^\oplus(t) = p(\mathcal{H}^\oplus(t) | \mathcal{Y}_T(t))$$

which will also be referred to as *model credibility coefficients*.

We will need several assumptions in order to further transform (15). First, assume that in the case the hypothesis  $\mathcal{H}^\oplus(t)$

is true the data set  $\mathcal{Y}_T^\ominus(t \ominus 1)$  bears no information about  $y(t \oplus i), i \in T^\oplus(t)$

$$p(\mathcal{Y}_T^\oplus(t) | \mathcal{Y}_T^\ominus(t \ominus 1), \mathcal{H}^\oplus(t)) = p(\mathcal{Y}_T^\oplus(t) | \mathcal{H}^\oplus(t)). \quad (\text{A1})$$

According to (A1), the forward and backward models are assumed to be unrelated data generating mechanisms meeting at time  $t$ . This assumption is at the heart of the competitive smoothing philosophy and enforces the exclusivity of the two hypothetical models  $\mathcal{H}^-$  and  $\mathcal{H}^+$ . As will be shown, this causes the rule for combining the estimates to become (essentially) a “switching” rule, and will allow the algorithm to decipher jump discontinuities from noisy data in a very efficient manner. On the other hand, in situations when both the forward and backward predictors actually do describe the signal, the assumption causes suboptimal performance because it forces the estimation algorithm to ignore all data from the poorer model.

The second assumption

$$p(\mathcal{Y}_T^\oplus(t \oplus 1) | \mathcal{H}^\ominus(t)) = p(\mathcal{Y}_T^\ominus(t \ominus 1) | \mathcal{H}^\oplus(t)) \quad (\text{A2})$$

sets the “balance of uncertainties”—it postulates that exactly the same should be (un)known about the signal in the time interval  $T^-(t-1)$ , given that  $\mathcal{H}^+(t)$  is true, as about the signal in the interval  $T^+(t+1)$ , given that  $\mathcal{H}^-(t)$  is true. Failing (A2) is tantamount to a preference of one time direction over the other, to saying that more is known about the (say) forward causality than about backwards causality. *A priori*, there is no reason to prefer one over the other.

Our main task consists in combining in a rational way the results yielded by the forward and backward Kalman filters. Since we do not intend to modify the corresponding estimation algorithms, only to compare their performance in the analysis window  $T(t)$ , it is natural to choose prior densities for  $z^\oplus(t \oplus M)$  in the form

$$\begin{aligned} \pi(z^\oplus(t \oplus M) | \mathcal{H}^\oplus(t), \rho) \\ = \mathcal{N}(\hat{z}^\oplus(t \oplus M | t \oplus M), \rho V^\oplus(t \oplus M | t \oplus M)) \end{aligned} \quad (\text{A3})$$

where  $\hat{z}^\oplus$  and  $V^\oplus$  are yielded by the corresponding Kalman filters.

Assumption (A3) guarantees that the Kalman filters operated over the window  $T(t)$  are commensurate with the Kalman filters operated over the previous (and future) windows. This is done simply by choosing the “prior” densities in the window  $T(t)$  to be equal to the densities already calculated in the previous windows. Note that (A3) can be also viewed as a condition of “neutrality” of the decision-making Bayesian observer.

The last assumption concerns selection of prior distributions for  $\mathcal{H}^\oplus(t)$  and  $\rho$ . In order to arrive at results which depend entirely upon the experimental data we adopt the so-called noninformative priors, namely, we choose  $\pi(\mathcal{H}^\oplus(t), \rho)$  in the form

$$\begin{aligned} \pi(\mathcal{H}^\ominus(t), \rho) &= \pi(\mathcal{H}^\oplus(t))\pi(\rho) \\ \pi(\mathcal{H}^-(t)) &= \pi(\mathcal{H}^+(t)) = \frac{1}{2} \\ \pi(\rho) &\propto \frac{1}{\rho}, \quad \rho > 0. \end{aligned} \quad (\text{A4})$$

Note that  $\pi(\rho)$  in (A4) is an *improper* distribution. The noninformative prior (A4) is one standard choice (see Box and Tiao [9]) when nothing can be presumed known about the prior distribution. The particular form of (A4) is not crucial to the analysis, though it does lead to relatively simple closed form expressions for the credibility coefficients.

Under the assumptions (A1) and (A3)

$$E[x(t) | \mathcal{Y}_T(t), \mathcal{H}^\oplus(t)] = \hat{x}^\oplus(t | t).$$

In order to evaluate model credibility coefficients, observe that

$$\begin{aligned} \mu^\oplus(t) &= p(\mathcal{H}^\oplus(t) | \mathcal{Y}_T(t)) \\ &\propto p(\mathcal{Y}_T(t) | \mathcal{H}^\oplus(t))\pi(\mathcal{H}^\oplus(t)) \\ &= p(\mathcal{Y}_T^\oplus(t) | \mathcal{Y}_T^\ominus(t \ominus 1), \mathcal{H}^\oplus(t)) \\ &\quad \times p(\mathcal{Y}_T^\ominus(t \ominus 1) | \mathcal{H}^\oplus(t))\pi(\mathcal{H}^\oplus(t)) \end{aligned}$$

i.e., according to (A1) and (A2)

$$\mu^\oplus(t) \propto p(\mathcal{Y}_T(t) | \mathcal{H}^\oplus(t))\pi(\mathcal{H}^\oplus(t), \rho). \quad (16)$$

Observe also that the proportionality constant in (16) does not depend on  $\oplus$ .

Under Gaussian assumptions one can use the Kalman filter methodology to evaluate the conditional density

$$\begin{aligned} p(\mathcal{Y}_T^\oplus(t) | \mathcal{H}^\oplus(t)) &= \int_0^\infty p(\mathcal{Y}_T^\oplus(t) | \rho, \mathcal{H}^\oplus(t)) \\ &\quad \times \pi(\rho | \mathcal{H}^\oplus(t)) d\rho \end{aligned} \quad (17)$$

namely, it holds [10]

$$\begin{aligned} p(\mathcal{Y}_T^\oplus(t) | \rho, \mathcal{H}^\oplus(t)) &= 2\pi\rho^{-\frac{M}{2}} \left[ \prod_{i=0}^{M-1} \beta^\oplus(t \oplus i) \right]^{-\frac{1}{2}} \\ &\quad \cdot \exp \left[ -\frac{1}{2\rho} \sum_{i=0}^{M-1} \frac{[\epsilon^\oplus(t \oplus i)]^2}{\beta^\oplus(t \oplus i)} \right] \end{aligned} \quad (18)$$

where

$$\beta^\oplus(t \oplus i) = 1 + G^T V^\oplus(t | t \oplus i \oplus 1) G. \quad (19)$$

After substituting (14) into (13), carrying out the integration and combining with (12) one gets (for noninformative priors (A4))

$$\mu^\oplus(t) = C \left[ \prod_{i=0}^{M-1} \beta^\oplus(t \oplus i) \right]^{-\frac{1}{2}} \left[ \sum_{i=0}^{M-1} \frac{[\epsilon^\oplus(t \oplus i)]^2}{\beta^\oplus(t \oplus i)} \right]^{-\frac{M}{2}} \quad (20)$$

where  $C$  is the proportionality constant, independent of  $\oplus$ , which should be chosen so as to satisfy

$$\mu^-(t) + \mu^+(t) = 1.$$

Summing up, under assumptions (A1)–(A4) the competitive smoother takes the form

$$\hat{x}(t | T) = \mu^-(t)\hat{x}^-(t | t) + \mu^+(t)\hat{x}^+(t | t) \quad (21)$$

where the model credibility coefficients are given by (15).

## IV. GENERALIZATIONS AND SIMPLIFICATIONS

The competitive smoothing scheme of Section III consists of a pair of Kalman predictors in conjunction with a nonlinear rule for combining the estimates of the two predictors. The concept of competitive smoothing, however, is considerably more general than this specific scheme, and this section pursues several generalizations and simplifications.

Section IV-A extends the competitive smoothing idea to other predictors using the “prequential likelihood” rule of Dawid [11] and Section IV-B generalizes the basic scheme to multiple models which adjust their “memory” based on the rate of variation of the underlying signal.

Speaking from a practical perspective, one prefers schemes which are numerically simple, easy to implement, and robust to (erroneous) assumptions about the unknown underlying signal, even if they are not fully justified theoretically. We will show that for a sufficiently long decision windows the weighting rule based on (20) and all the similar rules to be derived in this section *de facto* reduce to the rule which says to choose the predictor with the smallest prediction error accumulated over the window of competition. This intuitively appealing “switching rule” can be exploited in several ways. Using the switching approach, Section IV-C suggests the incorporation of a true smoothing element into the competition. This improves the performance of the algorithm in situations when the forward and backward predictors both adequately describe the underlying motion of the signal and, more importantly, substantially reduces the negative effect of “parameter jitter” caused by occasional switching between the competing algorithms.

## A. Competition Among a More General Class of Predictors

Let  $\hat{x}^\oplus(t | t \oplus 1)$  be any forward-time/backward-time predictor of  $x(t)$  based on the past/future observation history, and let  $\epsilon^\oplus(t)$  denote the corresponding prediction error

$$y(t) = \hat{x}^\oplus(t | t \oplus 1) + \epsilon^\oplus(t). \quad (22)$$

Can we judge the credibility of the competing predictors  $\hat{x}^-(\cdot | \cdot)$  and  $\hat{x}^+(\cdot | \cdot)$  without invoking explicit models of signal variation such as the generalized random walk model (4)? The answer is yes. View (22) as an implicit model of generating data  $\mathcal{Y}$ . Since for the optimal predictor  $\hat{x}^\oplus(\cdot | \cdot)$ , prediction errors  $\epsilon^\oplus(\cdot)$  form a white noise sequence, the likelihood of a data set  $\mathcal{Y}_T^\oplus(t)$  for the model (22) can be defined as

$$\begin{aligned} L(\mathcal{Y}_T^\oplus(t); \hat{x}^\oplus(\cdot | \cdot)) &= p(\mathcal{Y}_T^\oplus(t) | \hat{x}^\oplus(\cdot | \cdot)) \\ &= \prod_{i=0}^{M-1} p(\epsilon^\oplus(t \oplus i)) \end{aligned} \quad (23)$$

provided that the sequence  $\{\epsilon^\oplus(t)\}$  is independent. Dawid [11] called (23) the *prequential likelihood function* in contrast to the Fischer or classical likelihood. Note that the predictor  $\hat{x}^\oplus(\cdot | \cdot)$  and the innovation sequence  $\{\epsilon^\oplus(t)\}$  play the same role in the prequential approach as the parameter vector and measurement noise in the classical approach.

Define implicit hypothetical models of signal variation

$$\begin{aligned} y(i) &= \hat{x}^\oplus(i | i \oplus 1) + \epsilon^\oplus(i) \\ \mathcal{H}^\oplus(t) &: \text{var}[\epsilon^\oplus(t)] = \rho \\ & i \in T^{\epsilon^\oplus}(t) \end{aligned} \quad (24)$$

then

$$p(\mathcal{Y}_T^\oplus(t) | \mathcal{H}^\oplus(t)) = L(\mathcal{Y}_T^\oplus(t); \hat{x}^\oplus(\cdot | \cdot)).$$

Under assumptions (A1)–(A4), and with a Gaussian assumption on the innovation sequence  $\epsilon^\oplus(t) \sim \mathcal{N}(0, \rho)$ , one has

$$\begin{aligned} \mu^\oplus(t) &\propto \int_0^\infty (2\pi\rho)^{-\frac{M}{2}} \exp\left\{-\frac{1}{2\rho} \sum_{i=0}^{M-1} [\epsilon^\oplus(t \oplus i)]^2\right\} \frac{1}{\rho} d\rho \\ &\propto \left(\sum_{i=0}^{M-1} [\epsilon^\oplus(t \oplus i)]^2\right)^{-\frac{M}{2}}. \end{aligned} \quad (25)$$

The natural extension of (21) to an arbitrary predictor based competitive smoother becomes

$$\hat{x}(t | T) = \mu^-(t)\hat{x}^-(t | t-1) + \mu^+(t)\hat{x}^+(t | t+1) \quad (26)$$

where the predictor credibility coefficients are evaluated according to (25).

It is not difficult to show that the “prequential rule” (25) is asymptotically equivalent to the Bayesian rule (20). Since

$$\lim_{t \rightarrow \infty} V^-(t | t-1) = \lim_{t \rightarrow -\infty} V^+(t | t+1) = V_\infty \quad (27)$$

where  $V_\infty$  is the steady state solution to (12), for all values of  $t$  such that both  $t$  and  $N-t$  are “sufficiently large” it holds

$$\beta^-(t) = \beta^+(t) = 1 + G^T V_\infty G$$

and the Bayesian rule (20) reduces to (23).

It is interesting to note that the prequential decision rule can be extended to non-Gaussian innovation sequences. Suppose, for example, that the distribution of  $\epsilon^\oplus(t)$  is double exponential

$$p(\epsilon^\oplus(t)) = \frac{1}{2\rho} \exp\left\{-\frac{|\epsilon^\oplus(t)|}{\rho}\right\}. \quad (28)$$

Then, instead of (cred coefficients), the credibility coefficients can be calculated

$$\begin{aligned} \mu^\oplus(t) &\propto \int_0^\infty (2\rho)^{-M} \exp\left\{-\frac{1}{\rho} \sum_{i=0}^{M-1} |\epsilon^\oplus(t \oplus i)|\right\} \frac{1}{\rho} d\rho \\ &\propto \left(\sum_{i=0}^{M-1} |\epsilon^\oplus(t \oplus i)|\right)^{-M}. \end{aligned} \quad (29)$$

One can expect the decision rule (23) to be more robust than (21). In particular, it should be much less sensitive to outliers in the data (since the double exponential distribution is “heavily tailed,” it is frequently used to characterize disturbances that contain outliers [12]).

### B. Extension to Multiple Models

The coefficient  $\lambda = \sigma/\rho$  in (12) should reflect the rate of variation of the recovered signal. If a suitable value of  $\lambda$  is not known *a priori*, or if it is time-dependent, a number of Kalman filter-based estimation algorithms, each set for a different value of  $\lambda$ , can be run in parallel and combined appropriately. This technique can be readily adopted for use in competitive smoothing.

Consider  $K$  forward/backward estimation algorithms of the form (12) characterized by  $K$  different variance ratios  $\lambda_1, \lambda_2, \dots, \lambda_K$ , and the corresponding set of mutually exclusive and equiprobable hypothetical models  $\mathcal{H}_j^\oplus(t), \pi(\mathcal{H}_j^\oplus(t)) = 1/2K, j = 1, 2, \dots, K$  of the form (13). Minimization of the mean square risk function (14) under assumptions (A1)–(A4) (with the obvious modifications) leads to the following generalized version of (20), (21):

$$\hat{x}(t|T) = \sum_{j=1}^K [\mu_j^-(t)\hat{x}_j^-(t|t) + \mu_j^+(t)\hat{x}_j^+(t|t)] \quad (30)$$

$$\mu_j^\oplus(t) \propto \left[ \prod_{i=0}^{M-1} \beta_j^\oplus(t \oplus i) \right]^{-\frac{1}{2}} \left[ \sum_{i=0}^{M-1} \frac{[\epsilon_j^\oplus(t \oplus i)]^2}{\beta_j^\oplus(t \oplus i)} \right]^{-\frac{M}{2}} \quad (31)$$

where the quantities  $\hat{x}_j^\oplus(t|t), \epsilon_j^\oplus(t),$  and  $\beta_j^\oplus(t)$  are yielded by the  $j$ th forward/backward Kalman filter. In addition to recovering sudden signal changes, the extended competitive smoother (10), (31) is capable of adjusting its “memory” (i.e., smoothing range) to the locally observed “rate of variation” of the underlying signal. Based on sensitivity and robustness analyses of parallel estimation algorithms, one can argue that the parameters  $\lambda_1, \lambda_2, \dots, \lambda_K$  should form a geometric progression [10].

The extension of the prequential decision rule (25) to multiple models is also straightforward. The corresponding competitive smoother takes the form

$$\hat{x}(t|T) = \sum_{j=1}^K [\mu_j^-(t)\hat{x}_j^-(t|t-1) + \mu_j^+(t)\hat{x}_j^+(t|t+1)] \quad (32)$$

$$\mu_j^\oplus(t) \propto \left( \sum_{i=0}^{M-1} [\epsilon_j^\oplus(t \oplus i)]^2 \right)^{-\frac{M}{2}} \quad (33)$$

Finally, we note that for sufficiently long decision windows (i.e., for sufficiently large  $M$ ), the weighting rule based on (33) will *de facto* reduce to the following switching rule

$$\hat{x}(t|T) = \hat{x}_j^o(t|t \circ 1) \quad (34)$$

where

$$\{\mathbf{o}, \mathbf{j}\} = \arg \min_{\oplus, j} \sum_{i=0}^{M-1} [\epsilon_j^\oplus(t \oplus i)]^2 \quad (35)$$

i.e., to a rule that chooses the predictor that has most recently performed the best.

### C. Incorporation of a True Smoothing Element

The preceding scheme is not a true smoother in the sense that it does not incorporate knowledge of the signal from both before and after the present time in making its current estimate. Rather, it uses knowledge either before or after the present time. This is inextricably linked to the hypothesis (10) and the assumptions that either the forward or backward predictor (but not both) can explain the present data. In regions where both forward and backward models actually are linked to the signal motion (for instance, when the underlying signal is a ramp or other slowly varying signal) then one might expect a certain degradation in the behavior of the competitive scheme when compared to a standard Kalman smoother.

To see this clearly, consider a pair of order zero predictors trying to describe a ramp signal. The forward predictor lags behind the ramp while the backward predictor leads the ramp. Switching between them results in a jagged estimate that bounces from one to the other due to small noises on the signal—see Figs. 1(c) and 2(c) of the example presented in Section V.

Of course, this can be fixed (as in Section I) by incorporating higher order predictors, but the same problem will then reappear with higher order signals. For instance, first-order predictors would solve the problem with ramps, but would be susceptible to a similar jitter when confronted with parabolic signals. We believe that there is a better solution.

Suppose that instead of competing just a forward and a backward predictor, a true smoother was added into the competition. If the competition was conducted fairly, then the forward and backward predictors would “win” the competition at and near discontinuities as before, while the smoother would tend to “win” in the neighborhood of slowly varying parameters. Ideally, we would return to the hypotheses (13) to incorporate the new signal model, appropriately revise the assumptions (A1–A4) to include the smoother, and then rederive the model credibility coefficients (20). Unfortunately, due to technical complexities, we have been unable to derive such a scheme rigorously within the Bayesian framework. Nonetheless, we can argue heuristically how such a scheme should be structured.

As in the derivation of Section III, the model credibility coefficients will undoubtedly become “switching” rules in which either the forward predictor, the backward predictor, or the smoother will be chosen at each time step, depending on the character of the underlying signal near that time. As with the predictors, the smoothers will operate over a window of (user defined) time  $W$  and the prediction errors will be averaged over a window of (user defined) length  $M$  in order to calculate the model credibility coefficients (or, equivalently, to “choose” which predictor/smoothers to use).

A few implementation issues are worth discussing:

- 1) When comparing the forward and backward predictors at a given time step, it is necessary to use different blocks of length  $M$ . The forward prediction errors are gathered on the block  $[t, \dots, t - M + 1]$  while the backward prediction errors are gathered over the block

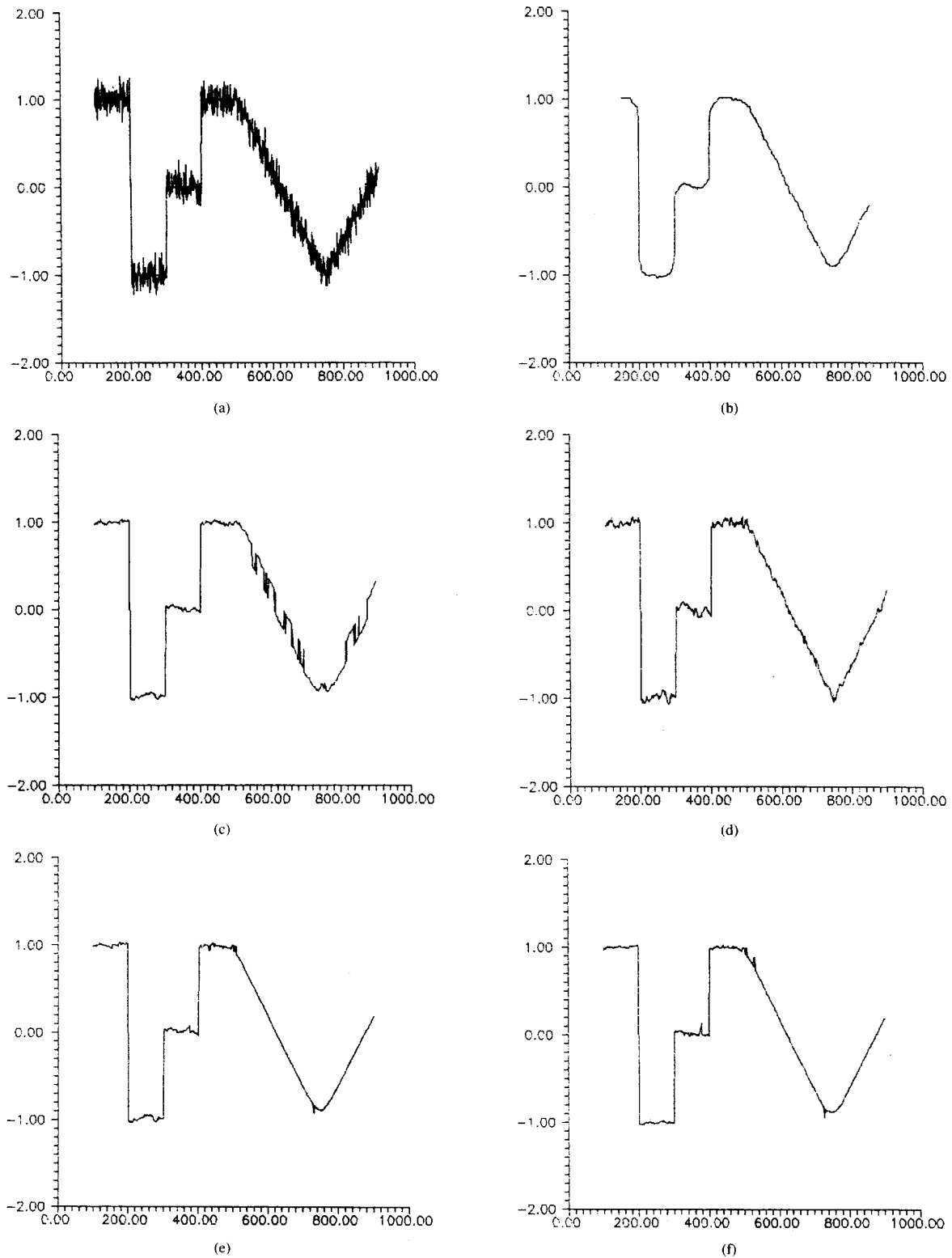


Fig. 1. (a) Test signal corrupted by a Gaussian noise with standard deviation 0.1. The results of applying different nonlinear filters: (b) median filter; (c) a pair of competing first-order Kalman filters; (d) a pair of competing second-order Kalman filters; (e) a pair of competing first order Kalman predictors and a true smoother; (f) a pair of competing averagers and a smoother.

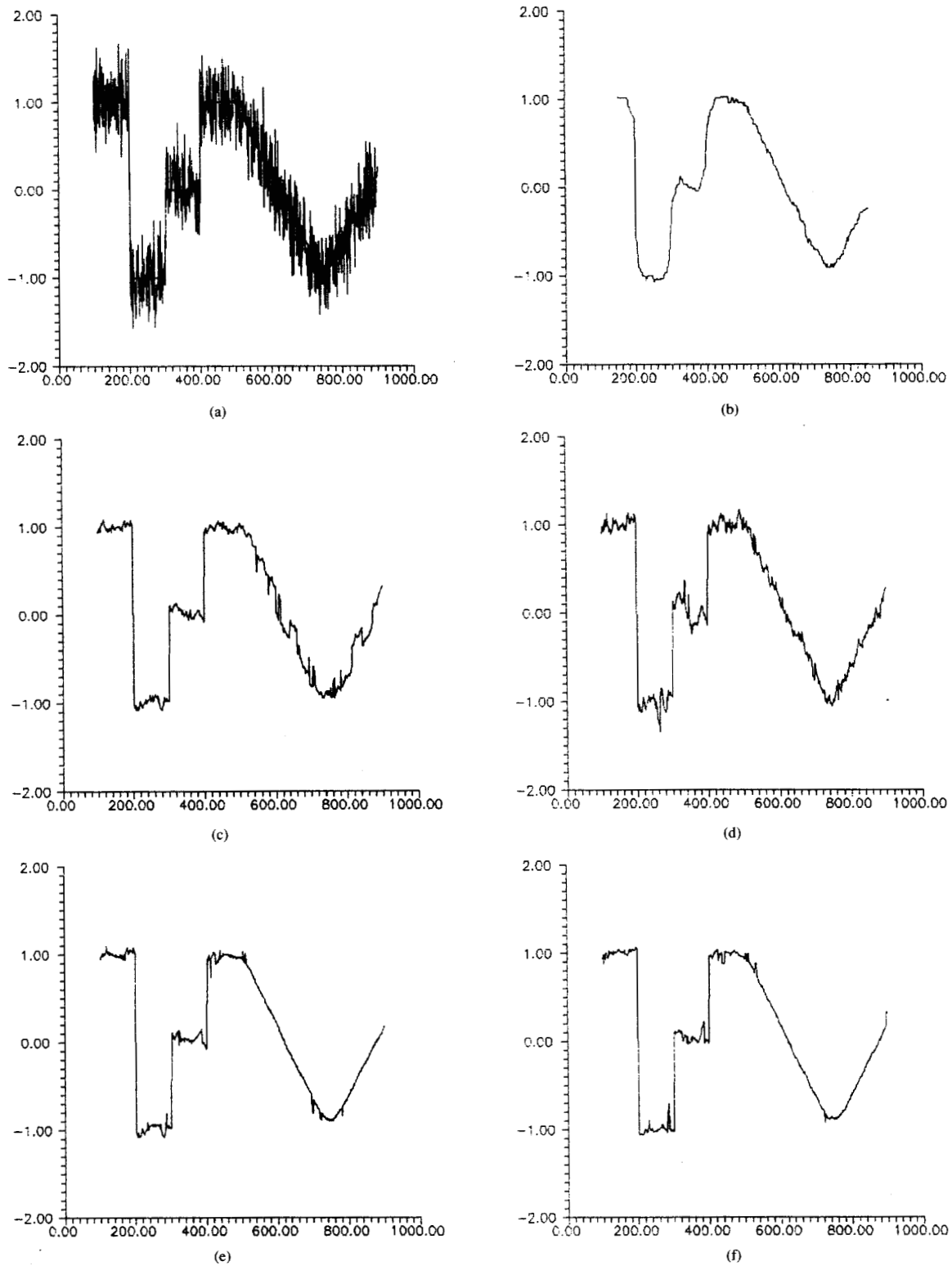


Fig. 2. (a) Test signal corrupted by a Gaussian noise with standard deviation 0.25. The results of applying different nonlinear filters: (b) median filter; (c) a pair of competing first order Kalman filters; (d) a pair of competing second order Kalman filters; (e) a pair of competing first-order Kalman predictors and a true smoother; (f) a pair of competing averages and a smoother.

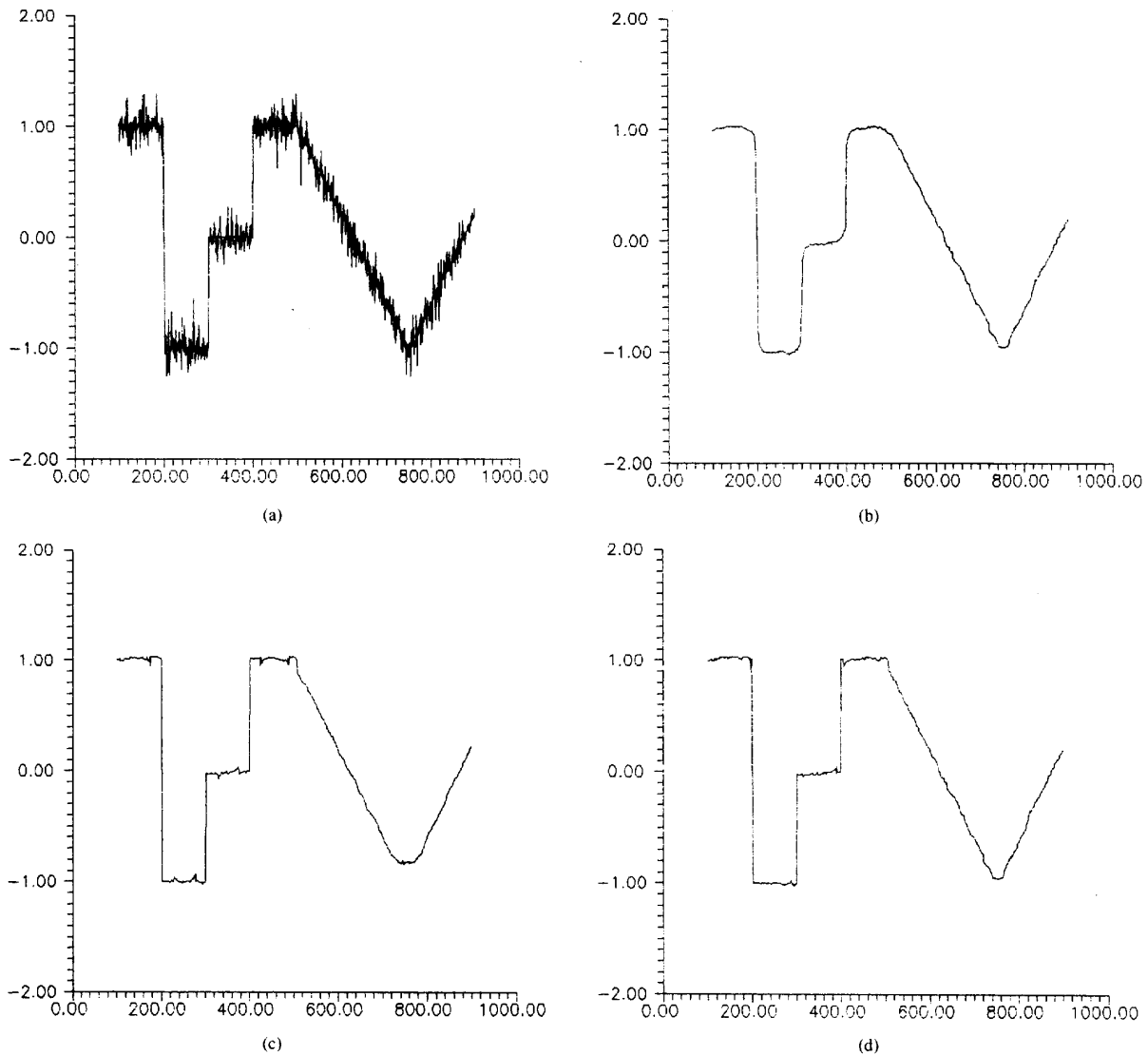


Fig. 3. (a) Test signal corrupted by an exponential noise with standard deviation 0.1. The results of applying different nonlinear filters: (b) median filter; (c) a pair of competing median filters/predictors and a linear combiner; (d) a pair of competing median filters/predictors and a "holey" median smoother.

$[t, \dots, t + M - 1]$ . Thus

$$J^\oplus(t) = \sum_{i=0}^{M-1} [\epsilon^\oplus(t \oplus i)]^2 \quad (36)$$

defines the summed square prediction error used previously for the purpose of predictor selection, cf. (35).

With the smoother, the user has more freedom to choose the window. For instance, one might choose

$$J_s(t) = \sum_{i=-K}^K e^2(t + i)$$

where  $K = (M - 1)/2$  (for odd  $M$ ). A better idea, however, is to evaluate the smoother over windows coincident with the windows of  $J^-(t)$  and  $J^+(t)$ . This provides better comparisons with the forward and back-

ward predictors since the noise sequences encountered in calculating (say)  $J^-(t)$  for the forward predictor and in calculating  $J_s^-(t)$  for the smoother are *identical*. This is like choosing a pair of smoothers, one designed to compete fairly with the forward predictor and one designed to compete fairly with the backward predictor—see [13] for a more detailed discussion of this issue.

- 2) It is unfair to use the sample that is being estimated when making that estimate (imagine using a window of length one—the "smoother" would give zero prediction error). This suggests that it is necessary to use a "holey" smoother, one which ignores the data at the present time in calculating the smoothed estimate.

In consideration of the above, and coinciding with the desire to keep the algorithm as simple as possible, we have chosen

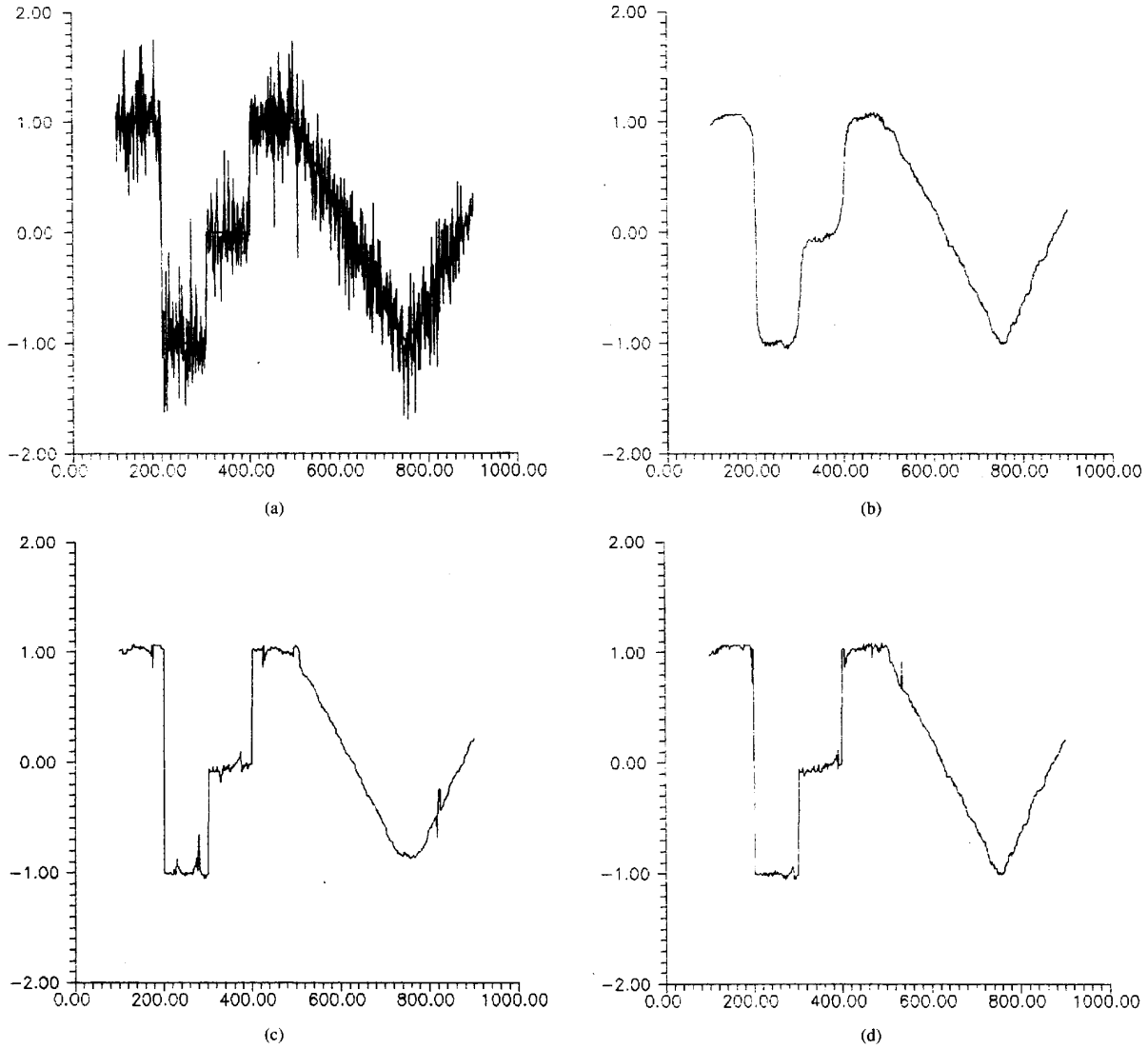


Fig. 4. (a) Test signal corrupted by an exponential noise with standard deviation 0.25. The results of applying different nonlinear filters: (b) median filter; (c) a pair of competing median filters/predictors and a linear combiner; (d) a pair of competing median filters/predictors and a "holey" median smoother.

to implement a holey smoother obtained after averaging the forward and backward estimates

$$\hat{x}_s(t) = \frac{1}{2}(\hat{x}^-(t|t-1) + \hat{x}^+(t|t+1)).$$

The corresponding prediction (extrapolation) error at time  $t$  is

$$\epsilon_s(t) = y(t) - \hat{x}_s(t).$$

These prediction errors are then gathered over the two windows

$$J_s^\oplus(t) = \sum_{i=0}^{M-1} \epsilon_s^2(t \oplus i) \quad (37)$$

which compete fairly against the windows of summation in (36) via the switching rule that chooses the final "competitively

smoothed" estimate as

$$\hat{x}(t) = \begin{cases} \hat{x}^-(t|t-1) & \text{if } J_*(t) = J^-(t) \\ \hat{x}_s(t) & \text{if } J_*(t) = J_s^-(t) \text{ or } J_s^+(t) \\ \hat{x}^+(t|t+1) & \text{if } J_*(t) = J^+(t) \end{cases} \quad (38)$$

where  $J_*(t) = \min[J^-(t), J_s^-(t), J_s^+(t), J^+(t)]$ . Clearly, (38) is the natural extension of (34), (35).

Of course, the competing idea is not limited to the use of Kalman predictors. Virtually any predictive method (for instance averagers, least mean square predictors, signed error algorithms) can be used. The details of the algorithms we have implemented (including our favorite choice of windows for the holey smoother) are given in the next section along with several simulations, and with some comments regarding jitter and switching artifacts.

One can easily envision using multiple smoothers in the same way that multiple Kalman predictors are competed in Section III-C, or one could mix averagers and LMS algorithms and Kalman predictors of various lengths. In such multiple schemes, however, there is the danger that large numbers of competing schemes will increase the number and severity of erroneous “decisions” based on pathological noise sequences.

#### D. Coping With Outliers

Median filters have two important features which make them particularly useful in signal processing applications

- 1) they are capable of reducing the measurement noise without significantly distorting the step-like signal features;
- 2) they are capable of rejecting outliers from the data (such as occasional spikes or large disturbances produced by “heavy-tailed” noise sources).

So far our developments were focused on the jump-preservation problem only. A natural question arises whether it is possible to design competitive smoothers which could also (i.e., in addition to jump recovery) deal with outliers. Quite obviously, all competitive smoothers based on linear predictors do not meet the above requirement simply because their component filters are not capable of rejecting outliers. However, the problem can be easily ameliorated if the smoother is composed of suitably chosen *nonlinear* filters.

As an example, we will show how one can incorporate median filters into the smoothing framework developed in Section IV-C.

Let  $A_K = (a_1, \dots, a_K)$  be any sequence of  $K$  real numbers and let  $\tilde{A}_K = (\tilde{a}_1, \dots, \tilde{a}_K)$  denote the same sequence rearranged in the nondecreasing order (i.e.,  $\tilde{a}_1 \leq \tilde{a}_2 \leq \dots \leq \tilde{a}_K$ ). The median of  $A_K$  will be defined as follows

$$\begin{aligned} \text{med}[a_1, \dots, a_K] &= \text{med}[\tilde{a}_1, \dots, \tilde{a}_K] \\ &= \begin{cases} \tilde{a}_{k+1} & \text{if } K = 2k + 1 \\ \frac{\tilde{a}_k + \tilde{a}_{k+1}}{2} & \text{if } K = 2k \end{cases} \end{aligned}$$

Note that if  $K$  is an odd number the median is simply the middle value of  $A_K$ , i.e., it is both the  $(k+1)$ st largest and the  $(k+1)$ st smallest element in  $A_K$ .

The simple median-based competitive smoother can be obtained by following the general recommendations of Section IV-C, namely, by combining the switching rule (38) with the filters

$$\hat{x}^\oplus(t | t \oplus 1) = \text{med}[x(t \oplus 1), \dots, x(t \oplus K)] \quad (39)$$

and

$$\begin{aligned} \hat{x}_s(t) &= \frac{1}{2} (\text{med}[x(t-1), \dots, x(t-K)] \\ &\quad + \text{med}[x(t+1), \dots, x(t+K)]). \end{aligned} \quad (40)$$

Instead of the linear combiner (40) some other forms of the “holey” smoother can be also used such as the “holey” median filter

$$\hat{x}_s(t) = \text{med}[x(t-k), \dots, x(t-1), x(t+1), \dots, x(t+k)]. \quad (41)$$

In both cases mentioned above, competitive smoothing yields considerably better results than the conventional median smoothing (see Section V).

*Remark:* We note that the “full” median filter

$$\text{med}\{x(t-k), \dots, x(t-1), x(t), x(t+1), \dots, x(t+k)\}$$

can be easily modified to yield (42) as its byproduct. For this reason, if  $K = 2k + 1$  only one modified median filter has to be swept across the data  $x(1), \dots, x(N)$  to yield  $\hat{x}^-(t | t-1)$ ,  $\hat{x}^+(t | t+1)$  and  $\hat{x}_s(t)$ , given by (39) and (41), for all values of  $t$ .

Even though both median-based competitive smoothers described above are completely *ad hoc* they certainly illustrate potentials of the “competitive approach” advocated in this paper. It should be also clear that the competitive approach to smoothing does not attempt to replace median or any other order statistical filters. In fact a carefully designed competitive smoother can be used to *enhance* performance of such filters whenever their application is justified.

## V. SIMULATIONS AND EXAMPLES

To illustrate the behavior of the competitive smoothing algorithm, consider a “true” signal that consists of several jumps and ramps corrupted by a Gaussian noise with standard deviation 0.10 (in Fig. 1(a)) and with standard deviation 0.25 (in Fig. 2(a)). As a benchmark, Figs. 1(b) and 2(b) show the result of smoothing with a single pass of a median filter. Figs. 1(c) and 2(c) show the result of an incorrect application of a pair of competing first-order Kalman predictors. Note especially the jagged switching artifacts on the ramp portion of the curve. This can be fixed in two ways. Section III suggests using a higher order Kalman model, and Figs. 1(d) and 2(d) demonstrate this approach with a pair of second-order predictors. In contrast, Figs. 1(e) and 2(e) utilize a pair of first-order predictors plus a true smoother as suggested in Section IV. Finally, Figs. 1(f) and 2(f) show the application of a pair of backward and forward averagers competing against a smoother. Several remarks are in order.

- A) The noise sequence in Fig. 2 is simply a scaled version of the noise sequence in Fig. 1. Thus features in the smoothed versions which are not simple scalings of each other are directly attributable to the nonlinearity of the smoothing process.
- B) In all cases, the model weighting factors were so close to the switching rule that virtually no differences were evident in the simulations. Thus, all of the competitive smoothing simulations were done using the simplified “switching rule” (38).
- C) The competitive schemes must choose a window over which the prediction errors will be gathered, and a smoother must be chosen for the simulations in (e) and (f). Since the results appeared relatively insensitive to the choice of prediction error window (15 was almost indistinguishable from 30), a window of  $M = 20$  was chosen for all simulations. The smoother was chosen as the simple average of the forward and backward predictors as in Section IV.

D) Perhaps the most difficult part of these simulations was to choose the “free parameters” so that the pictures present a fair comparison. These free parameters are the window length of the median filter, the parameter  $\lambda$  (the effective memory) of the Kalman filter, and the window length for the averager. There are several possibilities:

- 1) Run the algorithms on a “true” signal that is constant, and equalize the sums of squared errors by suitable choices of the window sizes and the parameter  $\lambda$ .
- 2) With a constant “true” signal, equalize the summed square error as in (1) but using only the “forward” parts of the predictors.
- 3) Choose the free parameters so that the effective window lengths (memory) of the various algorithms are equivalent.

Using 1), the smoother portion of the competing algorithms always “wins” the competition. This not only gives an overly optimistic estimate of how well the competitive schemes perform, but also requires the median length to be excessively long. With the noise as in Fig. 1, the median would have to be 90 timesteps long to match an averager window of 30. Criterion 2) is appealing primarily because the forward predictor portions of the algorithms are linear and hence the noise attenuation is a standard calculation. The median filter can be handled using the approximation that (with Gaussian noise) the median must be  $3/2$  as long as a linear filter for equal noise performance. It is unclear, however, how much the forward predictors alone can tell about the behavior of the full competitive scheme. Criterion (3) makes sense from a different point of view. For most edge-preserving schemes, some *a priori* information about the expected time between jumps is needed. One can choose the smoothing window based on this knowledge, and the tradeoffs are relatively straightforward. Longer windows tend to give better noise attenuation while short windows tend to capture smaller details.

Fortunately, criteria 2) and 3) give results that are approximately the same. Set a window size of 30 for the averager (30 window + 20 to sum the prediction error = 50, half of the expected time between jumps). This gives a window of 45 for the median (which is close to 50). The forgetting factor for the first order Kalman predictor can then be set to  $2/(\sqrt{1-\lambda}) = 30$ , which gives (approximately)  $1 - \lambda = 0.004$ . These are the values used in the simulations. Based on similar considerations, the forgetting constant for the second order Kalman predictor was set at the level  $\lambda = 0.0001$ .

- E) The median filter tends to round the edges near the jumps, while the competing smoothers tend to leave clear, sharp edges.
- F) The competing schemes tend to be “roughest” at about half a window length away from jumps. In the immediate vicinity of a jump, either the forward or the backward predictor are clear winners. Far from jumps,

the smoother is the clear winner (this is why the ramp in Fig. 1(e) and (f) is so smooth). The worst place for the competing scheme is in the middle ground where switching is likely between roughly equal predictors/smoothers. This occasionally results in spikes, as can be seen in Fig. 1(e) and (f) just before the jump at time 300, and just after the jump at time 400 in Fig. 1(e). Since such spikes are often isolated, one might profitably post filter by a very short median to “clean up” such artifacts.

- G) The switching spikes of F) also demonstrate the potential danger in having too many models compete. As the number of competing predictors and smoothers increases, artifacts such as these spikes become more likely. Thus, multiple model schemes as in Section III should be restricted to use on signals which are truly multiscaled, or on signals about which no *a priori* knowledge is available.
- H) The ramp portion of Figs. 1(c) and 2(c) shows the necessity of incorporating either the higher order predictor or the smoother (or both). If, however, the signal is known to be piecewise constant, then little will be gained by adding the smoother (which would increase the likelihood of switching problems) or of using higher order models (which are more sensitive to jumps).
- I) Just as the median filter is often used repetitively to help sharpen the edges of a signal, the competing schemes can be used repetitively to further smooth the signal. Successive passes of competing smoothers continue to average out noises while still passing the jumps unchanged.
- J) In Figs. 1(f) and 2(f), simple averagers are used to make forward and backward predictions and smoothed estimates. Obviously, there are no significant differences between Fig. 1(e) and (f) and between Fig. 2(e) and (f). This substantiates our claim that the crux of the competitive smoothing approach is in the competition of appropriate predictors and smoothers, and not on the particular form (i.e., Kalman) that these predictors assume.
- K) On the ramp portion of the signal, Figs. 1(d) and 2(d) successfully ameliorate the negative effects of switching, since both the forward and backward predictors are able to closely match the ramp. As expected, the jump discontinuities at times 200, 300, and 400 are correctly resolved. Overall, however, the behavior of this smoother is far from satisfactory. Noises cause the second-order predictors to make quite jittery estimates, and the competition of two jittery estimates is still jittery (since higher order predictors are usually highpass filters they may be extremely sensitive to noise). Comparing Figs. 1(d) and 2(d) to Figs. 1(e) and 2(e) suggests that the addition of a true smoother is considerably more effective than the use of higher order predictors as a way to guarantee smoothness of recovered signals.

The quantitative results of comparison of different filters are summarized in Table I. For each filter the mean value of the

TABLE I

ENSEMBLE AVERAGE OF THE RESIDUAL SUM OF SQUARES (ALL RESULTS ARE SCALED BY  $10^{-3}$ ) OBTAINED FOR FIVE DIFFERENT NONLINEAR FILTERS: MEDIAN FILTER (MED), A PAIR OF COMPETING FIRST-ORDER KALMAN PREDICTORS AND A SMOOTHER (KAL1&SMO), A PAIR OF COMPETING AVERAGERS AND A SMOOTHER (AVE&SMO), A PAIR OF COMPETING SECOND-ORDER KALMAN PREDICTORS (KAL2), A PAIR OF COMPETING SECOND-ORDER KALMAN PREDICTORS, AND A SMOOTHER (KAL2&SMO)

	MED	KAL1&SMO	AVE&SMO	KAL2	KAL2&SMO
$\rho=0.10$	1.46	0.68	0.68	1.33	1.21
$\rho=0.25$	8.07	3.85	3.83	8.18	7.21

TABLE II

ENSEMBLE AVERAGE OF THE RESIDUAL SUM OF SQUARES (ALL RESULTS ARE SCALED BY  $10^{-3}$ ) OBTAINED FOR THREE DIFFERENT NONLINEAR FILTERS: MEDIAN FILTER (MED), A PAIR OF COMPETING MEDIAN FILTERS/PREDICTORS AND A LINEAR COMBINER (SUPMED1), A PAIR OF COMPETING MEDIAN FILTERS/PREDICTORS AND A "HOLEY" MEDIAN SMOOTHER (SUPMED2)

	MED	SUPMED1	SUPMED2
$\rho=0.10$	1.41	1.16	0.50
$\rho=0.25$	6.65	4.11	3.16

residual sum of squares

$$J = \sum_{100}^{900} (x(t) - \hat{x}(t | T))^2$$

shown in Table I is an ensemble average obtained for 100 different realizations of the noise sequence (superimposed on the same "true" signal). The results show clearly advantages of the low-order competitive smoothers over the classical median filter; they also demonstrate the relative inefficiency of smoothers based on higher order models.

The last set of simulation experiments was designed to check performance of median-based competitive smoothers. The test signal remained unchanged except that the Gaussian noise was replaced with the exponential noise (which is "heavily-tailed" and hence is frequently used to model outliers).

Results obtained for a classical median filter ( $K = 45$ ), competitive smoother based on two median filters/predictors and a linear combiner (41) ( $K = 45$ ) and a competitive smoother based on two median filters/predictors and a "holey" median smoother (42) ( $K = 45$ ,  $k = 22$ ) are shown in Figs. 1 (for a noise with standard deviation 0.1) and 2 (for a noise with standard deviation 0.25). Finally, Table II compares the average performance of the three filters described above (obtained for 100 realizations of the noise sequence). The advantages of using the competitive approach are pretty clear, especially if the second version of the competitive smoother, based on (42), is utilized. Experiments (not reported here) with the purely impulsive noise led to identical conclusions.

## VI. CONCLUSION

The competitive smoothing idea has been applied to the filtering of discontinuous signals buried in noise. The method is ideal for preserving jump discontinuities while attenuating noise. Although the competitive idea has only been formally justified in the Kalman filter setting, we believe that there is good evidence that the scheme is useful with almost any set

of well chosen predictors and smoothers. A full analysis of the performance of such nonlinear filters is nontrivial, but the conceptual simplicity of the scheme gives hope that it may be possible to better understand its behavior. It is shown that a carefully designed competitive smoother can yield much better results than the classical median filter. For this reason it can be considered a useful addition to the signal processors arsenal of nonlinear smoothing tools.

## REFERENCES

- [1] J. P. Norton, "Optimal smoothing in the identification of linear time-variable systems," *Proc. IEEE*, vol. 122, pp. 663-668, 1975.
- [2] G. Kitagawa and W. Gersch, "A smoothness priors time-varying AR coefficient modelling of nonstationary covariance time series," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 48-56, 1985.
- [3] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1971.
- [4] E. J. Coyle, J-H Lin, and M. Gabbouj, "Optimal stack filtering and the estimation and structural approaches to image processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2037-2066, 1989.
- [5] E. T. Whittaker, "On a new method of graduation," in *Proc. Edinburgh Math. Soc.*, 1923, vol. 41, pp. 63-75.
- [6] H. Akaike, "Likelihood and the Bayes procedure," in *Bayesian Statistics*, J. M. Bernardo *et al.*, Eds. Valencia, Spain: University Press, 1980, pp. 141-166.
- [7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [8] F. L. Lewis, *Optimal Estimation*. New York: Wiley, 1986.
- [9] G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, 1973.
- [10] M. Niedźwiecki, "Identification of nonstationary stochastic systems using parallel estimation schemes," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 329-334, 1990.
- [11] A. P. Dawid, "Present position and potential developments: Some personal views, statistical theory, the prequential approach," *J. Roy. Statist. Soc. A*, vol. 147, pp. 278-292, 1984.
- [12] J. Astola and Y. Neuvo, "Optimal median type filters for exponential noise distributions," *Signal Processing*, vol. 17, pp. 95-104, 1989.
- [13] M. Niedźwiecki and W. A. Sethares, "New filtering algorithms based on the concept of competitive smoothing," in *23rd ISICIE Symp. Stochastic Syst. Theory and its Appl.* (Osaka, Japan), 1991, pp. 129-132.



Maciej Niedźwiecki received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland and the Dr. Hab. degree from the Technical University of Warsaw in 1977, 1981 and 1991, respectively.

He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986-1989. He is now with the Department of Automatic Control, Faculty of Electronics, Technical University of Gdańsk. His main areas of research interest include system identification, signal processing, and adaptive systems.



William A. Sethares received the B.A. degree in mathematics from Brandeis University, Waltham, MA, and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY.

He has worked at the Raytheon Company as a Systems Engineer and is currently on the faculty of the Department of Electrical and Computer Engineering at the University of Wisconsin-Madison. His research interests include adaptive systems in signal processing, communications, control, and electronic music.