Nonlinear Filters Satisfying Restricted Linearity Conditions

Ronald K. Pearson
ProSanos Corporation
Harrisburg, PA, USA

Jaakko Astola
Moncef Gabbouj
Institute of Signal Processing
Tampere University of Technology
Tampere, FINLAND

Abstract

This paper introduces the class of $\mathcal{K}$-linear filters, which are linear when restricted to a specified cone $\mathcal{K}$ of input sequences, but which can exhibit very nonlinear behavior for input sequences that do not belong to this set. An important special case is the class of *monotone-linear* filters, which are linear when restricted to monotone input sequences. It is shown that all positive L-filters belong to this class. Other results presented here include a general procedure for generating new $\mathcal{K}$-linear filters from known examples, and a closely related, simple linearization strategy that is useful in characterizing their nonlinearity.

1 Introduction

Linear digital filters are extremely important in practice, in large part because their linearity provides the basis for a variety of useful qualitatitive characterizations and design rules. The problems of analyzing and designing nonlinear digital filters are much harder, in general, because we normally have no exploitable alternative to linearity in developing these results. It is, however, possible to obtain interesting and useful classes of nonlinear filters by requiring them to satisfy the axioms of linearity, not for all real-valued input sequences $\{x_k\}$, but only for a restricted subset $\mathcal{S}$ of these sequences. More specifically, a *cone* of sequences is any set $\mathcal{K}$ of sequences satisfying the following condition [17, p. 28]:

$$\{x_k\}, \{y_k\} \in \mathcal{K} \;\Rightarrow\; \{\alpha x_k + \beta y_k\} \in \mathcal{K}, \qquad (1)$$

for all real $\alpha, \beta \geq 0$. Given a cone $\mathcal{K}$, define the filter $\mathcal{F}$ to be $\mathcal{K}$*-linear* if the following two conditions hold:

a. the cone $\mathcal{K}$ is an invariant set for $\mathcal{F}$, meaning $\{x_k\} \in \mathcal{K} \Rightarrow \mathcal{F}\{x_k\} \in \mathcal{K}$;

b. the behavior of the filter $\mathcal{F}$ is positive-linear when restricted to $\mathcal{K}$, meaning that for all $\{x_k\}, \{y_k\} \in \mathcal{K}$ and all $\alpha, \beta > 0$, we have:

$$\mathcal{F}\{\alpha x_k + \beta y_k\} = \alpha \mathcal{F}\{x_k\} + \beta \mathcal{F}\{y_k\}.$$

Note that while any linear filter $\mathcal{L}$ satisfies the second of these conditions on any cone $\mathcal{K}$, the first is generally not satisfied without imposing some restrictions on $\mathcal{L}$. For the cone $\mathcal{I}$ of increasing pseudocausal sequences introduced in the next section, it can be shown [14] that $\mathcal{L}$ must belong to the class of *externally positive systems,* defined as those with nonnegative impulse responses [6, p. 9], a restriction that has important behavioral consequences [4].

2 Monotone-linear Filters

To make these results more concrete, consider the cone $\mathcal{I}$ of *increasing pseudocausal sequences*, defined as follows: $\{x_k\} \in \mathcal{I}$ if and only if there exists some finite $k^* \leq 0$ such that $x_k = 0$ for all $k < k^*$ and $x_{k+1} \geq x_k$ for all $k \geq k^*$. The corresponding class of $\mathcal{K}$-linear filters will be called *monotone-linear filters* and it has the following useful characterization. Define the unit step at time ℓ as the function

$$s_\ell(k) = \begin{cases} 1 & k \geq \ell \\ 0 & k < \ell, \end{cases} \qquad (2)$$

and note that $s_\ell(k) = s_0(k-\ell)$. Any sequence $\{x_k\} \in \mathcal{I}$ can be written as:

$$\begin{aligned} x_k &= x_{k^*} s_{k^*}(k) + (x_{k^*+1} - x_{k^*}) s_{k^*+1}(k) + \cdots \\ &= \sum_{j=-\infty}^{\infty} (x_j - x_{j-1}) s_0(k-j), \end{aligned} \qquad (3)$$

where $x_j - x_{j-1} \geq 0$ for all j. Thus, if $\mathcal{F}$ is monotone-linear,

$$\mathcal{F}\{x_k\} = \sum_{j=-\infty}^{\infty} (x_j - x_{j-1})\mathcal{F}\{s_0(k-j)\}. \quad (4)$$

If $\mathcal{F}$ is also shift-invariant, define its unit step response as $f_k = \mathcal{F}\{s_0(k)\}$, leading to the following result for any $\{x_k\} \in \mathcal{I}$:

$$\mathcal{F}\{x_k\} = \sum_{j=-\infty}^{\infty} (x_j - x_{j-1}) f_{k-j}. \quad (5)$$

While this result is well-known for linear filters—without the restriction to increasing pseudocausal sequences—it is important to emphasize that Eq. (5) holds on $\mathcal{I}$ for the entire class of monotone-linear filters, which includes some extremely nonlinear members, as discussed in Secs. 3 and 4.

3 The class of L-filters

The class of *L-filters* is defined by [5, 11]:

$$\mathcal{F}\{x_k\} = \sum_{j=-K}^{K} a_j x_{(j)}, \quad (6)$$

where $\{x_{(j)}\}$ is the rank-ordered sequence obtained from the windowed input sequence $\{x_{k-K}, \ldots, x_k, \ldots, x_{k+K}\}$, satisfying:

$$x_{(-K)} \leq \cdots \leq x_{(0)} \leq \cdots \leq x_{(K)}. \quad (7)$$

(If $x_i = x_j$ for some i, j in the data window, the original relative ordering of i and j is retained in constructing this rank-ordered sequence.)

If we require $a_j \geq 0$ for all j in Eq. (6), the resulting L-filter is monotone-linear. Specifically, note that any sequence $\{x_k\} \in \mathcal{I}$ already satisfies the rank-order conditions of Eq. (7): that is, $x_{(j)} = x_{k-j}$ for $j = -K, \ldots, 0, \ldots, K$. Thus, for all increasing input sequences, the L-filter defined in Eq. (6) is equivalent to the positive linear FIR filter:

$$\mathcal{F}\{x_k\} = \sum_{j=-K}^{K} a_j x_{k-j}. \quad (8)$$

The resulting class of positive L-filters includes both the linear unweighted moving average filter, $\mathcal{U}_K$, obtained by setting $a_j = 1/(2K+1)$ for all j in Eq. (6) and the median filter, $\mathcal{M}_K$, obtained by setting $a_0 = 1$ and $a_j = 0$ for all $j \neq 0$. Other members of this class include the trimmed mean filters and the Winsorized trimmed mean filters [2, 3, 11].

4 Nonlinear cascade structures

For any cone $\mathcal{K}$, the class of $\mathcal{K}$-linear filters is closed under cascade interconnection: if filters $\mathcal{F}$ and $\mathcal{G}$ both belong to the $\mathcal{K}$-linear class, then so does the cascade $\mathcal{F} \circ \mathcal{G}$, obtained by first applying filter $\mathcal{G}$ to the input sequence and then applying filter $\mathcal{F}$ to the result. In particular, since $\mathcal{G}\mathcal{K} \subset \mathcal{K}$ and $\mathcal{F}\mathcal{K} \subset \mathcal{K}$, it follows that $\mathcal{F} \circ \mathcal{G}\mathcal{K} \subset \mathcal{K}$, establishing Condition (a.) for the cascade. Similarly, note that

$$\begin{aligned} \mathcal{F} \circ \mathcal{G}\{\alpha x_k + \beta y_k\} &= \mathcal{F}[\alpha\mathcal{G}\{x_k\} + \beta\mathcal{G}\{y_k\}] \\ &= \alpha\mathcal{F} \circ \mathcal{G}\{x_k\} + \beta\mathcal{F} \circ \mathcal{G}\{y_k\}, \end{aligned} \quad (9)$$

establishing Condition (b.) for the cascade.

This observation leads to the following particularly simple construction, that is nevertheless powerful enough to lead to an interesting class of nonlinear filters. First, recall that a sequence $\{r_k\}$ is a *root sequence* for the filter $\mathcal{F}$ if $\mathcal{F}\{r_k\} = \{r_k\}$, and let $\mathcal{R}_\mathcal{F}$ denote the set of all root sequences for the filter $\mathcal{F}$. Since the filter $\mathcal{F}$ is trivially $\mathcal{K}$-linear if $\mathcal{K} \subset \mathcal{R}_\mathcal{F}$, it follows that any cascade interconnection of nonlinear filters containing $\mathcal{K}$ in their root sets and linear filters that leave $\mathcal{K}$ invariant defines a $\mathcal{K}$-linear filter. More specifically, the nonlinear cascade construction considered here consists of filters of the general form:

$$\mathcal{F} = \mathcal{L}_1 \circ \mathcal{N}_1 \circ \mathcal{L}_2 \circ \mathcal{N}_2 \circ \cdots \circ \mathcal{L}_n \circ \mathcal{N}_n, \quad (10)$$

where all linear components $\mathcal{L}_i$ satisfy the invariance condition:

$$\mathcal{L}_i\{x_k\} \in \mathcal{K} \text{ for all } \{x_k\} \in \mathcal{K}, \quad (11)$$

and all nonlinear components $\mathcal{N}_i$ satisfy the root condition:

$$\mathcal{N}_i\{x_k\} = \{x_k\} \text{ for all } \{x_k\} \in \mathcal{K}. \quad (12)$$

Note that any cascade of admissible linear and nonlinear components may be written in the form of Eq. (10) by defining certain components $\mathcal{L}_i$ or $\mathcal{N}_i$ to be identity filters for which all sequences are roots.

For the case of monotone-linear filters, the linear components satisfying Eq. (11) were completely characterized in Sec. 1 as those with nonnegative impulse response coefficients. A well-known example of a nonlinear filter satisfying the root condition (12) on $\mathcal{I}$ is the median filter, whose root set contains all monotone input sequences [7]. Consequently, arbitrary cascades of positive-linear filters with median filters belong to the monotone-linear filter class, despite their overall nonlinearity. Obviously, this condition is also met by any nonlinear filter whose root set contains that of the median filter $\mathcal{M}_K$ for some K. One example is the *Hampel filter* [15], a member of the class of *decision-based filters* [2]. This is a data cleaning filter based on the following idea: an outlier detection procedure is applied to the moving data window $\{x_{k-K}, \ldots, x_k, \ldots, x_{k+K}\}$ to determine whether the central element, x_k, is an outlier with respect to the data window or not. If so, this central data element is replaced with a more representative value, in this case the median of the $2K+1$ values in the data window; if x_k is not an outlier within this moving neighborhood, no change is made: the filter output is simply x_k. More specifically, x_k is declared to be an outlier and replaced with m_k if

$$|x_k - m_k| > tS_k, \tag{13}$$

where m_k is the median of the data window values, S_k is the median absolute deviation scale estimate [10, p. 107], defined by:

$$S_k = 1.4826 \text{ median}\{|x_{k-K} - m_k|, \ldots, |x_{k+K} - m_k|\}, \tag{14}$$

and t is a non-negative tuning parameter. For convenience in the following discussions, this filter will be denoted $\mathcal{H}_{K,t}$ to note that its performance depends on the window half-width parameter K and the threshold parameter t. Two useful observations are first, that if $\{r_k\}$ is a root sequence of the filter $\mathcal{H}_{K,t}$ for some fixed $t \geq 0$, then it is also a root sequence for $\mathcal{H}_{K,t'}$ for any $t' > t$. This fact follows from the observation that if $\{r_k\}$ is a root sequence of $\mathcal{H}_{K,t}$, then r_k never satisfies condition (13), implying:

$$|r_k - m_k| \leq tS_k \leq t'S_k, \tag{15}$$

so that $\{r_k\}$ is also a root sequence of $\mathcal{H}_{K,t'}$. The second key observation is that when $t = 0$, the Hampel filter $\mathcal{H}_{K,0}$ reduces to the median filter $\mathcal{M}_K$. Thus, it follows as a corollary that the root sequence set $\mathcal{R}_K$ of the median filter $\mathcal{M}_K$ is contained in the root sequence sets for $\mathcal{H}_{K,t}$ for all $t \geq 0$.

Finally, note that the root sets for many other nonlinear filters contain the increasing pseudocausal sequence set $\mathcal{I}$ as a subset even if they do not contain the entire root set for any median filter $\mathcal{M}_K$. Two specific examples include the FMH filter [9]:

$$\begin{aligned} y_k = \mathcal{F}\{x_k\} &= \text{median}\{L_-(K), x_{(0)}, L_+(K)\}, \\ L_-(K) &= \frac{x_{k-K} + \cdots + x_{k-1}}{K}, \\ L_+(K) &= \frac{x_{k+1} + \cdots + x_{k+K}}{K}, \end{aligned} \tag{16}$$

and the LUM smoothing filter [8]:

$$y_k = \mathcal{F}\{x_k\} = \text{median}\{x_{(-s)}, x_k, x_{(s)}\}, \tag{17}$$

for some integer $0 < s < K$. Note that for the FMH filter, if $x_k \leq x_{k+1}$ for all k, it follows that $L_-(K) \leq x_k = x_{(0)} \leq L_+(K)$, while for the LUM filter, $x_k \leq x_{k+1}$ for all k implies that

$$x_{(-s)} = x_{k-s} \leq x_k =\leq x_{k+s} = x_{(s)}. \tag{18}$$

Thus, in both cases, $y_k = x_k$, demonstrating that all increasing pseudocausal sequences belong to the root sets for both of these filters.

5 Linearization strategies

It is common in engineering practice to analyze nonlinear systems by linearizing them, seeking a linear approximation that captures their dominant behavior. This has the advantage of allowing us to bring some of our considerable intuition about and analysis results for linear systems to bear on nonlinear problems. In the case of smooth (i.e., infinitely differentiable) functions, this approach is based on the existence of a Taylor series expansion about some fixed reference value. Unfortunately, for most of the nonlinear filters of interest here, the nonlinear functions on which they are based are non-smooth, making Taylor series-based linearization completely inapplicable. In particular, Aczél, Gronau and Schwaiger [1, Prop. 9] have shown that any function $f : R^n \to R$ satisfying the positive homogeneity condition:

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda f(x_1, x_2, \ldots, x_n) \quad (19)$$

for all $\lambda > 0$ is differentiable in the neighborhood of $\mathbf{0}$ if and only if $f(\cdot)$ is the linear function:

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \alpha_i x_i, \quad (20)$$

for arbitrary real constants $\{\alpha_i\}$. This result extends to arbitrary points $\mathbf{x}_0 \in R^n$ if the function $f : R^n \to R$ is also translation-invariant [1, Prop. 9]:

$$f(x_1 + c, x_2 + c, \ldots, x_n + c) = f(x_1, x_2, \ldots, x_n) + c, \quad (21)$$

for all real constants c, subject to the additional constraint that the constants $\{\alpha_i\}$ sum to 1. Since most of the popular nonlinear filters in the signal processing literature satisfy both conditions (19) and (21), it follows that the notions of Taylor series linearization are not useful for the filters of interest here.

An alternative linearization approach that is applicable to nonlinear filters based on functions satisfying conditions (19) and (21) is that of Mallows [12]. Unfortunately, this procedure is computational rather than analytic in character, depending on the specific nonlinear filter considered and the reasonableness of an additive decomposition of the input sequence into Gaussian and non-Gaussian random sequences.

The special characteristics of the nonlinear filter classes of interest here lead to a much simpler linearization strategy. Specifically, since these filters are linear when restricted to some cone $\mathcal{K}$ of input sequences, we seek a linear extension of the filter to the class $\mathcal{C}$ of *all* input sequences. In cases where this linear extension is unique, we adopt it as the linearization of the $\mathcal{K}$-linear filter $\mathcal{F}$. In the case of monotone-linear filters, this idea leads to an immediate result: since a monotone-linear filter $\mathcal{F}$ is uniquely specified on $\mathcal{I}$ by its unit step response, the linearization of $\mathcal{F}$ is defined as the linear filter $\mathcal{L}$ whose unit step response is equal to that of $\mathcal{F}$. For the class of positive L-filters, this linearization corresponds to replacing the rank-ordered samples $x_{(i)}$ in the data window with the corresponding time-ordered samples x_{k-i} for $i = -K, \ldots, K$.

For the $\mathcal{K}$-linear filters based on the cascade structure discussed in Sec. 4, the strategy just described also leads to an immediate linearization result. Specifically, since the nonlinear components in this cascade structure all contain $\mathcal{K}$ in their root sets, these components may be regarded as identity maps on $\mathcal{K}$. Thus, the unique linear extensions of these cascade filters are obtained by simply omitting these nonlinear components. That is, the linearization of the $\mathcal{K}$-linear filter defined in Eq. (10) is simply

$$\mathcal{L} = \mathcal{L}_1 \circ \mathcal{L}_2 \circ \cdots \circ \mathcal{L}_n. \quad (22)$$

6 An example

In the nonlinear dynamic modeling literature, the cascade connection of a memoryless nonlinearity followed by a linear dynamic model is called a *Hammerstein model*, while the interconnection of the same two components in the opposite order is called a *Wiener model* [13]. Consequently, a monotone-linear filter obtained as the cascade interconnection of a nonlinear filter followed by a positive-linear filter will be termed a Hammerstein structure, while the opposite interconnection will be termed a Wiener structure. Since the linearizations of these two structures are the same, their difference in behavior illustrates the importance of interconnection order for nonlinear filter cascades. More specifically, consider the Hammerstein and Wiener structures based on the unweighted linear average filter $\mathcal{U}_3$ and the median filter $\mathcal{M}_3$, as defined in Sec. 3. In the Hammerstein structure, the median filter acts first and the result is smoothed by the unweighted linear average filter, while in the Wiener structure this order is reversed.

Fig. 1 shows two sequences: the line overlaid with open circles is the common input sequence $\{u_k\}$ applied to both the Hammerstein and Wiener structures, consisting of a sinlge exponential decay, contaminated with 10 randomly spaced outliers, all of

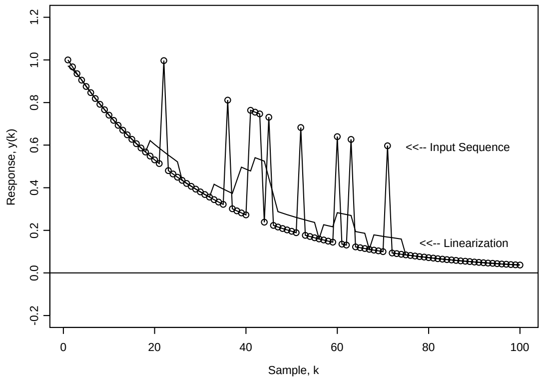

Figure 1: Contaminated exponential decay input sequence (line overlaid with open circles) and linearized filter response (solid line) for both the Hammerstein and Wiener filter structures described in Fig. 2.

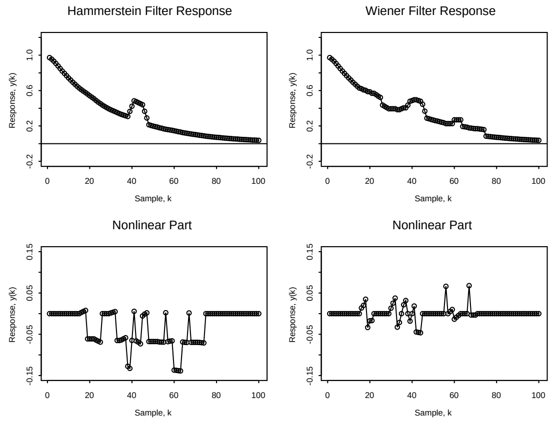

Figure 2: Responses of the Hammerstein (upper left) and Wiener (upper right) filter responses to the contaminated exponential input sequence from Fig. 1, with the nonlinear part of each response shown in the plot below.

the same magnitude. The solid line not overlaid with points is the response of the linear unweighted average filter $\mathcal{U}_3$, representing the common linearization of both the Hammerstein and Wiener structures. Note that except in the vicinity of the outliers, the linear filter response is essentially identical to the input sequence.

Fig. 2 shows the responses of the Hammerstein structure (upper left plot) and the Wiener structure (upper right plot) to the impulse-contaminated exponential decay input sequence shown in Fig. 1. The lower two plots in this figure show the *nonlinear parts* of these filter responses, defined as the difference between the filter response and its linearization. Comparing these bottom two plots, it is clear that the nonlinear part of the Hammerstein filter's response is generally larger in magnitude than that of the Wiener filter, consistent with the fact that the Wiener filter's response shows a much stronger resemblence to the linear filter response shown in Fig. 1 than the Hammerstein filter's response does.

7 Summary

This paper has defined the general class of $\mathcal{K}$-linear filters, which behave linearly on a specified cone $\mathcal{K}$ of input sequences but which are nonlinear on the space of all real-valued input sequences. A general cascade interconnection strategy is presented in Sec. 4 for constructing members of this filter class, based on linear filters $\mathcal{L}$ for which $\mathcal{K}$ is invariant (i.e., $\mathcal{L}\mathcal{K} \subset \mathcal{K}$) and nonlinear filters $\mathcal{N}$ whose root set contains $\mathcal{K}$. Specific nonlinear filters meeting this root condition are discussed for the monotone-linear case where $\mathcal{K}$ corresponds to $\mathcal{I}$, the cone of increasing input sequences. It is also shown that the class of positive L-filters belongs to the monotone-linear filter class, including special cases like the median filter and trimmed mean filters. An interesting open question is whether there exist other simple nonlinear filter structures in the monotone-linear class that are neither L-filters nor nonlinear cascade structures (e.g., clone-based constructions analogous to the FMH filter structure [16]).

Another useful result presented here is the fact that the class of monotone-linear filters is completely characterized on the cone $\mathcal{I}$ of increasing input sequences by the unit step response, extending the well-known result for the class of linear filters. One advantage of this result is that it provides the basis for a well-defined linearization procedure for these filters: the linearization $\mathcal{L}_{\mathcal{F}}$ of a monotone-linear filter $\mathcal{F}$ is simply the linear filter defined by the unit step response of $\mathcal{F}$. In the case of L-filters, this linearization is obtained by replacing the order statistics $\{x_{(j)}\}$ constructed for the moving data window with the samples $\{x_{k-j}\}$ on which this window is based, since these two sequences are equal for all $\{x_k\} \in \mathcal{I}$. For the nonlinear cascade structures considered here, the linearization is obtained by simply omitting the nonlinear filter components from the cascade. An important practical advantage of this linearization procedure is its simplicity, relative to computationally-based alternatives like the procedure of Mallows [12]; also, it was noted in Sec. 5 that the Taylor-series expansions on which most engineering linearizations of nonlinear systems are based are not applicable here since the filters involved are based on non-smooth functions.

References

[1] J. Aczel, D. Gronau, and J. Schwaiger, "Increasing Solutions of the Homogeneity Equation and of Similar Equations," *J. Math. Anal. Appl.*, v. 182, 1994, pp. 436–464.

[2] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, CRC Press, 1997.

[3] J.B Bednar and T.L. Watt, "Alpha-trimmed means and their relationship to median filters," *IEEE Trans. Acoustics, Speech, Signal Proc.*, v. 32, 1984, pp. 145–153.

[4] L. Benvenuti, L. Farina and B.D.O. Anderson, "Filtering through combination of positive filters," *IEEE Trans. Circuits Systems*, v. 46, 1999, pp. 1431–1440.

[5] A.C. Bovik, T.S. Huang, and D.C. Munson, Jr., "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acoustics, Speech, Signal Proc.*, v. 31, 1983, pp. 1342–1350.

[6] L. Farina and S. Rinaldi, *Positive Linear Systems: Theory and Applications*, Wiley, 2000.

[7] N.C. Gallagher and G.L. Wise, "A Theoretical Analysis of the Properties of Median Filters," *IEEE Trans. Acoustics, Speech, Signal Proc.*, v. 29, 1981, pp. 1136–1141.

[8] R.C. Hardie and C.G. Boncelet, "LUM filters: a class of rank order based filters for smoothing and sharpening," *IEEE Trans. Signal Proc.*, v. 41, 1993, pp. 1061–1076.

[9] P. Heinonen and Y. Neuvo, "FIR-Median hybrid filters," *IEEE Trans. Acoustics Speech Signal Proc.*, v. 35, 1987, pp. 832–838.

[10] P.J. Huber, *Robust Statistics*, Wiley, 1981.

[11] Y.H. Lee and S.A. Kassam, "Generalized median filtering and related nonlinear filtering techniques," *IEEE Trans. Acoustics, Speech, Signal Proc.*, v. 33, 1985, pp. 672–683.

[12] C.L. Mallows, "Some theory of nonlinear smoothers," *Ann. Statist.*, v. 8, 1980, pp. 695–715.

[13] R.K. Pearson, *Discrete-Time Dynamic Models*, Oxford, 1999.

[14] R.K. Pearson, "Empirical models that exhibit monotone responses to monotone inputs," *Proc. European Control Conference*, Karlsruhe, Germany, August, 1999, (CD-ROM).

[15] R.K. Pearson, "Outliers in process modeling and identification," *IEEE Trans. Control System Technology*, v. 10, 2001, pp. 55–63.

[16] R.K. Pearson, M. Gabbouj, and J. Astola, "Nonlinear FIR Filter Clones," *Proc. Nonlinear Signal Image Processing Conf. (NSIP'03)*, 2003.

[17] G.M. Ziegler, *Lectures on Polytopes*, Springer-Verlag, 1995.