TOWARDS A GENERAL THEORY OF ROBUST NONLINEAR FILTERING: SELECTION FILTERS *

Juan G. Gonzalez *Daniel L. Lau* *Gonzalo R. Arce*

Department of Electrical Engineering
University of Delaware, Newark, DE 19716, USA
gonzalez@ee.udel.edu

ABSTRACT

In this paper we introduce a general framework for edge preserving filters, derived from the powerful class of M-estimators. First, we show that under very general assumptions, *any* location estimator generates an edge preserving filter if we approximate the estimate by one of the input samples. Based on this premise, we propose the family of *S-estimators* or *S-filters*, as a *selection-type* class of filters arising from a computationally tractable "*s*electification" of location M-estimators. S-filters inherit the richness of the theory underlying the M-estimators framework, providing a very flexible family of robust estimators with edge preservation capabilities. Several properties of S-filters are studied. Sufficient and necessary conditions are given for an S-filter to present edge enhancing capabilities, and several novel filters within this framework are introduced and illustrated.

Data, figures and source code utilized in this paper are available at *http://www.ee.udel.edu/signals/robust/*

Keywords - Selection filters, M-estimators, S-estimators, S-filters, selectification, closification, edge preservation, impulse suppression, edge enhancing, edge sharpening.

I INTRODUCTION

The theory of nonlinear robust filtering has been motivated by the limitations of linear filters whenever the underlying processes are impulsive. Although linear methods are in general the optimal tools when signal statistics are Gaussian, it is well known that linear estimators suffer from significant performance degradation when noise distributions become heavy-tailed.

Signal smoothing and enhancing is a typical problem in which alternative nonlinear *robust* methods must be often used. This problem has motivated the development of a wide variety of nonlinear "low-pass" robust filtering techniques with the capability of dealing with impulsive noise while still maintaining acceptable performance [11]. However, due to the lack of a general theory underlying the problem of robust filtering, common procedures tend to be *ad hoc*, and most of the frameworks proposed are often very limited or too particular.

There is an important class of smoothing applications that requires, in addition to robustness, careful management and preservation of signal edges. This is the case for example in image processing, where the perception of edges is of fundamental importance to the human visual system. Again, linear smoothers show inappropriate for these kinds of applications because of the severe smearing to edges that they introduce.

A nonlinear filtering framework that has proven successful managing both *impulse resistivity* and *edge preservation*, is given by the class of median-based filters [11, 13]. *Possessing the above two properties, while still presenting relative computational simplicity, has been the key to the great acceptance and success of median-based filters in the signal processing community.* However, filters based on the median operator are not very flexible. They may cause edge jitter [4], streaking [3], and may remove important image details [1]. As it will be shown later in this paper, median-based filters tend to blur edges when smoothing noisy signals, and are not capable of performing edge enhancing operations. Also, although it is known that median-based filters are the optimal framework when the underlying statistics are Laplacian (i.e. biexponential), their smoothing ability can be significantly reduced if the noise distribution deviates from the Laplacian model.

Although several alternatives and generalizations have been proposed to circumvent the limitations of median filters [1, 2, 5, 10], there is the lack of a general theory underlying the problem of robust edge preserving smoothing. In this paper we introduce a solid framework for *selection-type*[1] filters, sufficiently powerful to attack the problem of robust smoothing in the presence of edges, and with enough flexibility to guarantee its success in many practical applications.

First, we establish a link that directly relates *any* location estimator (including the median), with a filtering framework satisfying the properties of impulse suppression and edge preservation. It is shown that this link can be easily built if we constrain the filter to be *selection-type*. This "selection" characteristic, possessed naturally by the median, is precisely what has made median-based techniques so popular in applications where edge and detail preservation are significantly important. Based on the above premise, we introduce the class of *S-estimators* or *S-filters*, as the class of selection-type filters derived from a *computationally tractable* "*s*electification" of location M-estimators.

The theory of M-estimators is a very general and mature field in robust statistics [7, 8]. For a set of samples $x_1, x_2, \cdots, x_N$, the M-estimator of location is defined, in general, as the value β which minimizes the sum $\sum_{i=1}^{N} \rho(x_i - \beta)$. The function ρ is usually known as the *cost function* associated with the estimator, and it plays a fundamental role in the understanding and characterization of

*THIS RESEARCH WAS FUNDED IN PART BY THE U. S. NATIONAL SCIENCE FOUNDATION UNDER GRANT MIP-9530923 AND THROUGH COLLABORATIVE PARTICIPATION IN THE ADVANCED TELECOMMUNICATIONS/INFORMATION DISTRIBUTION RESEARCH PROGRAM (ATIRP)CONSORTIUM SPONSORED BY THE U.S. ARMY RESEARCH LABORATORY UNDER COOPERATIVE AGREEMENT DAAL01-96-2-0002.

[1] We will refer to a filter as *selection-type* if its output is always equal to one of the input data. The filter operation can thus be seen as "selecting" one of the data from the input window according to some given rule.

M-estimators. The name "cost" is explained from the engineering interpretation that a "penalty" with value $\rho(x_i - \beta)$ shall be paid for the estimator to be away from sample x_i. Under this point of view, the M-estimator is the point β with the minimum sum of costs.

Table 1 shows several examples of M-estimators and the cost functions associated with them. Worth noting, the sample mean and the sample median can be seen, respectively, as the least squares and the least absolute value statistics, and thus they are members of the M-estimator family.

Estimator	$\rho(x)$
Mean	x^2
Median	$\|x\|$
Huber [8]	$\begin{cases} x^2 & \text{if } \|x\| \leq k \\ 2k\|x\| + k^2 & \text{if } x > k \end{cases}$
Myriad [5]	$\log(k^2 + x^2)$

Table 1. Several M-estimators and their corresponding associated cost functions.

II SELECTION ESTIMATORS AND THE SELECTIFICATION PROCEDURE

To begin, we introduce the concept of *closification* of a general location estimator (not necessarily in the M-estimator class), and show some important properties of the resulting "closified" estimator.

Definition 1 (Closified location estimator) *Given a set of samples $x_1, x_2, \cdots, x_N$, and a location estimator $\hat{\beta}$, we define $\hat{\beta}_C$, the* closified *version of $\hat{\beta}$, as the selection estimator that takes the sample* closest *to $\hat{\beta}$*

$$\hat{\beta}_C = \arg\min_{x_j} |\hat{\beta} - x_j|. \quad (1)$$

The procedure of transforming a location estimator in its closified version will be referred to as *closification.* Closified estimators follow an important "mode-type" property when the samples are restricted to binary[2] values:

Property 1 *Let $\hat{\beta}$ be a "reasonable" location estimator on a set of binary samples. Then, the closified estimator $\hat{\beta}_C$ can easily be calculated as the most repeated value in the sample.*

The above property comes from our empirical understanding of the job that a location estimator is supposed to do. Since we do not want to deal with the hassle of defining a "reasonable" location estimator, we do not offer any proof[3]. Nevertheless, the result is intuitively appealing. There is not any apparent reason why a *fair* estimate locating a set of binary samples, would be closer to the less repeated value than to the most repeated one.

The following two properties are a direct consequence of Property 1 for any filter based on closified estimators:

Property 2 (Edge preservation) *Step signals are preserved after the operation of the filter.*

Property 3 (Impulse suppression) *Constant pulses of length less than half the window size are suppressed by the filter.*

It is interesting to note that, whether for median or for closified filters, the reason for the appearance of the above two properties is exclusively explained by the "selection" characteristic of the estimators. Closification thus offers a simple and general procedure for providing edge preservation and impulse suppression capabilities to smoothing filters. These capabilities have been precisely the key for the great acceptance of median-type filtering in the signal processing community.

Nevertheless, in general, closification is a computationally expensive procedure. The direct computation of a closified estimator requires the prior calculation of the original estimator in addition to the minimization indicated in (1). In the following, searching for computational savings while still maintaining the properties of selection estimators, we introduce the concept of *selectification* in the context of cost function based M-estimation.

For the sake of mathematical tractability, and also to comply with our intuitive understanding of "reasonable" location estimation, we want to constrain our attention to cost functions with the following properties: (1) symmetry around zero, (2) increasing monotonicity on $(0, \infty)$, and (3) continuity.

Definition 2 (Selectified estimator) *Given a set of samples $x_1, x_2, \cdots, x_N$, and a cost function ρ satisfying the properties in the previous paragraph, we define the* selectified M-estimator *associated with ρ as*

$$\hat{\beta}_S = \arg\min_{x_j} \sum_{i=1}^{N} \rho(x_i - x_j). \quad (2)$$

We will refer to $\hat{\beta}_S$ as *the s*election estimator associated with ρ, and, as member of a general class, we will call it an *S-estimator*[4].

Note that the only difference between the definitions of S- and M-estimators, is the domain set for the minimization. While the M-estimator comes from an optimization over every possible real number, the S-estimator constrains the optimization over the finite set constituted by the samples. This is, in general, a valuable numerical advantage for small sample size applications. While the evaluation of closified M-estimators may require in most of the cases expensive numerical procedures, the complexity of any S-estimator is always on the order of $O(N^2)$ or less.

It can be easily shown that S-estimators, like closified estimators, follow Properties 1, 2 and 3. Important to note, S-estimators are, in general, different from closified M-estimators. An example in which closification and selectification lead to different results is illustrated in Fig. 1. Here, we have used the myriad [5] cost function, $\rho(x) = \log(k^2 + x^2)$, for the set of samples $x_1, x_2, \ldots, x_7$. As it is indicated, the M-estimate is the point $\hat{\beta}$ minimizing the sum of costs. The sample *closest* to $\hat{\beta}$, (x_3 in this case), corresponds to the closified estimator, whereas the sample with the *lowest* sum of costs (x_4 in this case), corresponds to the S-estimator.

There exist cost functions for which selectification and closification are equivalent procedures. The quadratic cost

[2] We use the term binary when the samples can only take either one of two possible values.

[3] To introduce a little more rigor, Property 1 by itself could be considered as a formal characterization of the class of *reasonable* estimators of location.

[4] The letter "S" is aimed to indicate the "*s*election" characteristic of the estimator. The term "S-estimator" has previously been used in statistics to denote a class of estimators based on the minimization of a scale statistic [12]. Since the scope of this work differs significantly from that of [12] using the same name should not give rise to confusion.

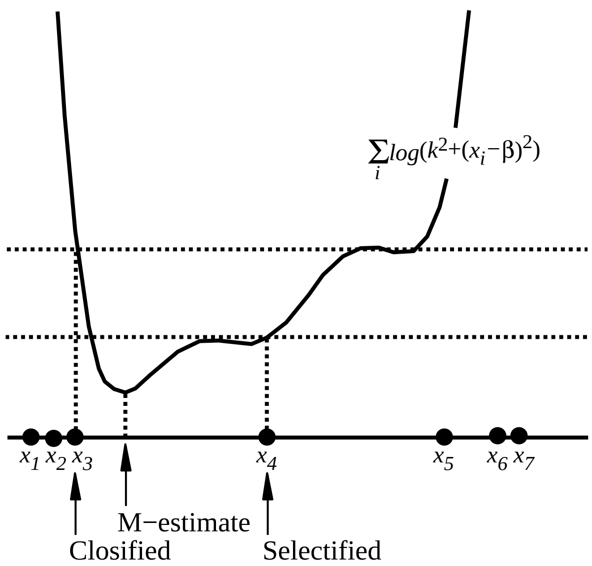

Figure 1. Illustration of closification and selectification procedures. While the closified estimate is the sample closest to the M-estimate, the selectified estimate is the sample with the lowest sum of costs.

function $\rho(x) = x^2$, associated with the mean, gives a particularly interesting example of this case. The S-filter derived from this cost function, provides an edge-preserving smoothing framework with very low computational complexity (on the order of $O(N)$). We call this filter the "Closest-to-mean" or CTM filter. A study of its properties and capabilities in near Gaussian environments is introduced in reference [9], also presented at this conference.

III PROPERTIES OF S-ESTIMATORS

We begin this section with a simple but surprising result:

Property 4 *All S-filters with window length* 3 *are equivalent, independently of the cost function* ρ.

Property 4 is a direct consequence of the following bounds, which we state without proof due to lack of space:

Property 5 *For any cost function* ρ*, the associated S-estimator* $\hat{\beta}_S$*, is bounded by*

$$x_{(2)} \le \hat{\beta}_S \le x_{(N-1)}, \tag{3}$$

where $x_{(i)}$ *denotes the i-th order statistic of the sample.*

Property 5 is of significant importance for smoothing applications, where in general, overshoots and/or undershoots in the filtered signal are not desired.

Although large window sizes are not common in most filtering applications, it is interesting to look at the asymptotic properties of S-estimators. It can be proven that, under very general conditions, S-estimators inherit the asymptotic behavior of their associated M-estimators. Thus, important results from M-estimation theory, such as consistency and asymptotic Gaussianity, can easily be extended to the family of S-estimators.

IV NATURAL SELECTION ESTIMATORS

M-estimators such as the median, possess the natural property of being selection-type without having to resort to the selectification procedure. We define any M-estimator with this property as a *natural S*-estimator. In the following, we introduce a powerful class of natural S- estimators.

Definition 3 *Let* $\hat{\beta}$ *be an M-estimator associated with the cost function* ρ*. We will call* $\hat{\beta}$ *a* totally descending *estimator if the derivative* ρ' *is monotonic decreasing on* $(0, \infty)$.

It can be shown that every totally descending estimator is natural selection. An important family of this type of estimators is formed by the "non-Hilbert" class of L_p estimators, associated with $\rho(x) = |x|^p$ for $0 < p < 1$. These estimators, which include the median in the limit case ($p = 1$), offer a computationally tractable alternative to robust filtering in impulsive environments, and possess very interesting properties that are worth further study. We conjecture that they form the only class of natural selection estimators which are scale invariant. In the limit, as $p \to 0$, the L_p class embraces the mode-myriad, a mode-type "zero-order" estimator with very high resistance to outliers [5, 6].

Another important class of S-estimators is characterized by the cost function

$$\rho(x) = \log(k^p + |x|^p), \tag{4}$$

where p is a positive constant, and $k \ge 0$ is a tuning parameter, intimately linked with the data variability.

For $p = 2$, expression (4) defines the class of *selection myriad* estimators, which has proven successful in the management of joint signal smoothing and edge enhancing [2, 5]. It can be shown that large values of k make the selection myriad to behave like the CTM estimator. Small values of k, on the contrary, resemble the behavior of the mode-myriad, indicated above.

For any $0 < p \le 1$, (4) defines a *totally descending* (and hence natural selection) estimator with a particular sensitivity to the values of the tuning constant k. When k is very large, it can be proven that the estimator behaves like the median, whereas for small values of k, the estimator resembles, again, the behavior of the mode-myriad.

A significant advantage of the family of estimators introduced in (4) is that it provides a very flexible filtering class with relatively low computational cost. It is easy to show that the estimator defined by (2) and (4) can be more efficiently calculated by

$$\hat{\beta}_S = \arg\min_{x_j} \prod_{i=1}^{N} (k^p + |x_i - x_j|^p), \tag{5}$$

which avoids the evaluation of the logarithms.

V EDGE ENHANCING FILTERS

It is well known that monotonic trends, which are the typical structure observed in blurred edges, stay invariant after the operation of the median filter. Thus, median type filters are significantly limited in image enhancing applications where sharpening of blurred edges is desired.

The S-filter family provides a rich framework in which edge enhancing can be easily managed. Carefully designing the cost function ρ, it is possible to obtain the desired levels of edge sharpening to be produced by the associated S-filter.

The following proposition, which we include without proof, characterizes the class of all S-filters that can perform signal edge enhancing.

Proposition 1 *An S-filter is provided with edge sharpening capabilities if and only if its associated cost function* ρ *has a region of nonconvexity.*

According to this, median and CTM filters cannot provide edge enhancing, whereas the class of totally descending filters can. Figure 2 illustrates the edge enhancing behavior of the filter associated with (4) ($p = 1$) as k is varied. For large values of k, this filter is equivalent to the median, which is unable to modify the monotonic edge. Reducing the value of k, progressively increases the edge enhancing capabilities of the filter. The above method can be successfully exploited in practice to induce edge enhancing features in median based filters. The parameter k would play the role of a *tuning constant* indicating the desired level of enhancing. This procedure is illustrated in the problem of smoothing the noisy blurred image of Fig. 3(left). The output of a 3×3 median filter is shown in Fig. 3(right). Obviously, the image blur cannot be overcome by the application of the median filter. Figure 4 shows how the edge enhancing capability of the filter improves as k is reduced.

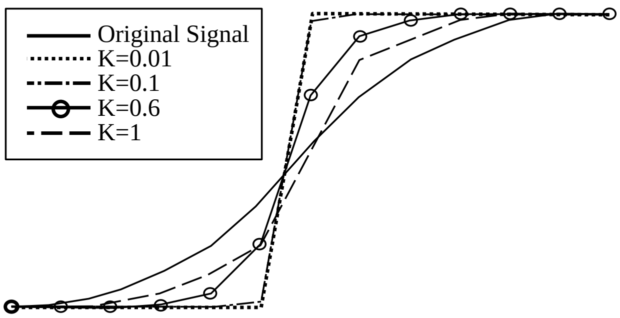

Figure 2. Providing edge enhancing capabilities to the median smoother via the family of estimators in (4). Smaller values of k induce increased edge sharpening capability. The median filter, corresponding to very large values of k, cannot modify the monotonic edge.

VI CONCLUSIONS

The class of S-filters, introduced in this paper, constitutes a powerful and flexible framework along the lines of the M-estimation philosophy. By constraining the minimization space of an M-estimator to be equal to the sample set, we transform the estimator in a *s*election-type (or S-)estimator with the features of edge preservation and impulse suppression. As in the case of M-estimation, careful cost function design is the key to control critical features such as filter robustness or edge enhancing capabilities. For example, we found that the performance of S- filters as edge enhancers can be directly characterized by the degree of nonconvexity in the associated cost functions. We also studied several properties of S-estimators, and introduced novel examples of estimators in this class. From these, it is worth highlighting the potential impact of the closest-to-mean (CTM) estimator, the "nonHilbert" L_p estimator, and the logarithmic cost function estimator introduced in expression (4). S-filters can play a significant role in applications such as image processing, where edge preservation and impulse suppression are features of paramount importance.

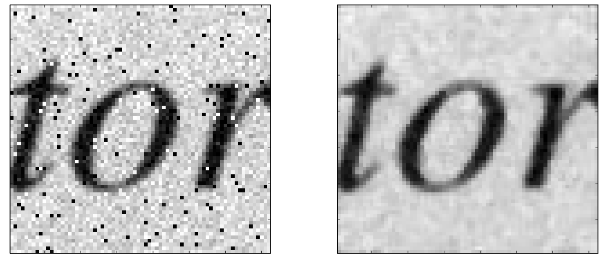

Figure 3. Smoothing of blurred image plus impulsive noise. (Left:) noisy image, and (right:) image smoothed with a 3×3 median filter.

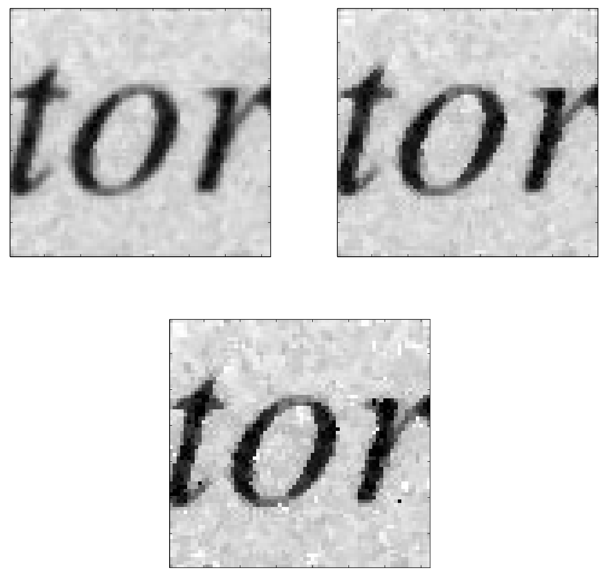

Figure 4. Edge enhancing introduced to the median filter by the logarithmic cost function in (4). (Top-left:) k=1.0, (top-right:) k=0.1, and (bottom:) k=0.001.

REFERENCES

[1] G. R. Arce and R. E. Foster. Detail-preserving ranked-ordered based filters for image processing. *IEEE Trans. in Acoustics, Speech, and Signal Processing*, 37(1), Jan. 1989.

[2] G. R. Arce, J. G. Gonzalez, and P. Zurbach. Myriad filters in imaging. In *Proc. of the ASILOMAR Conf. on Signals, Systems, and Computers*, Pacific Grove, California, USA, Nov. 1996.

[3] A. C. Bovik. Streaking in median filtered images. *IEEE Trans. in Acoustics, Speech, and Signal Processing*, ASSP-35:493–503, Apr. 1987.

[4] A. C. Bovik, T. S. Huang, and D. C. Munson. The effect of median filtering on edge estimation and detection. *IEEE Trans. in Pattern Analysis Machine Intelligence*, PA-MI9:181194, Mar. 1987.

[5] J. G. Gonzalez and G. R. Arce. Weighted myriad filters: A robust filtering framework derived from alpha-stable distributions. In *Proc. IEEE ICASSP'96*, Atlanta, Georgia, USA, 1996.

[6] J. G. Gonzalez, D. W. Griffith, and G. R. Arce. Zero-order statistics: A signal processing framework for very impulsive environments. In *Proc. of the IEEE Signal Processing Workshop of Higher Order Statistics*, Banff, Alberta, Canada, Submitted 1997.

[7] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* Wiley & Sons, New York, 1986.

[8] P. J. Huber. *Robust Statistics.* Wiley & Sons, New York, 1981.

[9] D. L. Lau and J. G. Gonzalez. The closest-to-mean filter: An edge preserving smoother for gaussian environments. In *Proc. IEEE ICASSP'97*, Munich, Germany, 1997.

[10] Y. H. Lee and S. A. Kassam. Generalized median filtering and related nonlinear filtering techniques. *IEEE Trans. on Signal Processing*, 33(3):672–683, June 1985.

[11] I. Pitas and A. N. Venetsanopoulos. *Non-linear Filters.* Kluwer, 1989.

[12] P. J. Rousseeuw and V. Yohai. *Robust regression by means of S-estimators.* Springer, New York, 1984.

[13] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. Weighted median filters: a tutorial. *IEEE Trans. on Circuits and Systems*, 41, May 1996.