Nonlinear Filter Design: Methodologies and Challenges

Sari Peltonen, Moncef Gabbouj and Jaakko Astola
Signal Processing Laboratory, Tampere University of Technology
P.O. Box 553, 33101 Tampere, Finland
Jaakko.Astola@tut.fi

Abstract

Linear filtering techniques have serious limitations in dealing with signals that have been created or processed by a system exhibiting some degree of nonlinearity or, in general, situations where the relevance of information cannot be specified in frequency domain. In image processing, many of these characteristics are often present, and it is no wonder that image processing is the field where nonlinear filtering techniques have first shown clear superiority over linear filters.

Since nonlinear filters are all of those filters that are not linear, there is a large variety of different filters in use, and no common theory can exist. This makes filter design challenging, and optimization is meaningful only after restricting the class. It can be done in several conceptually different ways and, in this paper, we consider these techniques and the optimization methods that go together with the particular restriction. We review polynomial and rational filter classes and optimization of stack filters under structural constraints and statistical constraints.

1. Signals and signal properties requiring nonlinear processing

In general, a signal is referred to as stationary if its statistical properties do not change over time or space. In many applications, stationarity in the wide sense is often assumed and it refers to processes with constant mean, finite power and having autocorrelation functions, which depend only on the time or space difference. Usually, it is also assumed that only (joint) second order statistics of the desired signal and the observed signal are known a priori. Thus, we do not have the data to determine the parameters of nonlinear estimators and the use of linear filters is most natural. Deterministically, one often uses the terms smooth or slowly varying signals. This means that when viewed in frequency domain the signal contains only low frequency components. Thus all high frequency disturbances can be removed by simple low-pass filtering. Such signals and processes preclude sharp transitions (details and edges often encountered in image and image sequences) and impulses (e.g. in heavy-tailed noise). When viewed in frequency domain these signals contain perhaps all frequencies and no frequency selective filtering is able to separate the desired signal and the noise. Optimal linear filtering will basically use more those frequencies where the relative energy of the signal is high compared to noise. In such cases and without resorting to sophisticated techniques (e.g. data-dependent tricks), linear processing will blur those transitions and will fail to remove impulses (as it tends to locally spread its energy). On the other hand, we may have some additional information available. For instance, if we know that the signal is locally monotonic and the noise is impulsive with certain low occurrence probability, a simple median filtering may restore the signal with practically no errors.

This phenomenon is even more prevailing in multivariate signals, e.g. color images, image sequences and satellite images, where the different channels may have very nontrivial joint statistics. Such dependence is totally ignored by any marginal (or component-wise) processing.

Optimal (Wiener) linear filtering in the most widely used form assumes that the observed and the desired signals are jointly Gaussian and that the noise corrupting process is uncorrelated with the desired signal. Such assumptions are hardly valid in any practical application involving images or color images. The need to develop more suitable processing techniques to deal with such signals and applications is clearly motivated.

The theory of linear filtering is based on the theory of linear spaces. The components of the signal belong to the real or the complex field and we have well defined order or magnitude concepts available. This fits well to the case where the signal originates from a sensor with linear characteristics. On the other hand, in many image processing tasks the signal is just two-valued (black and white) and the order is more or less arbitrary. Or there are several gray levels but the scales are not linear. The relationship of sensor output to subjective sense of brightness, for instance, can be quite nontrivial. One can ask if the use of theories based on linear spaces is really the right thing to do. In biology we encounter signals where there is no clear structure in the set of "values" that the signal components take. In statistical terms we are dealing with classified data. The signal processing methods should be designed for this situation. In practice, identifying the set with some subset of integers and so imposing structure may work reasonably well but the danger is that it will induce properties that were not actually present in the original data. If we use methods that assume very little of the model of the data and the processing when there in reality is a structured model behind our signal processing task we are not fully exploiting the possibilities but we cannot obtain results that are impossible in reality. On the other hand, if we are using methods that are based on assumptions that do not hold, there is danger that the results we get are grossly wrong. Typical example would be the use of moving average to filter out impulsive noise.

Thus it is evident that we need a large variety of methods so that in each case we are able to fully utilize the structure of the data. The need to use nonlinear filters is evident in many applications but the design process is difficult since the filters often are conceptually difficult to understand and tools for their comparison are inadequate. Experienced signal processing professionals know quite well the underlying implicit assumptions of different applications in addition to the strong and weak points of different nonlinear filter classes, and are thus able to pick a working method for a particular task. The ultimate goal of course would be to find the optimal filter, but due to large variety of different nonlinear filters without any common theory, optimization becomes meaningful only after restricting the filter class. This is the reason why we in this paper concentrate on the following filter classes: polynomial, rational and stack filters.

The paper is organized as follows. In Section 2 polynomial and rational filter classes are considered and in Section 3 the class of stack filters. For these filter classes we discuss how by restricting the filter class an optimal filter can be found and mention some application areas of the filters. After that we discuss the challenges in nonlinear filter design. Discussion and some conclusions are drawn in Section 5. The reference list given at the end of the paper is by no means complete, as the paper is not intended to be a tutorial in nonlinear filter design.

2. Polynomial and rational filters

Volterra systems are mainly used in communications and also in other widely different applications, such as control, hydrology, turbulence, motion of vessels and population studies, see [16]. As straightforward extensions (to higher kernel order) of linear filters Volterra systems are conceptually easier to assimilate. This makes it possible to extend many results obtained for the analysis and design of linear systems in relatively easy ways to the case of polynomial systems. If information of the joint higher order moments of the desired and observed signal is available, a polynomial filter can be directly optimized in the mean square sense. Usually there is no underlying model that would provide this information and direct estimation from data has proved to be unreliable.

An advantage of Volterra filters is that they are nonlinear with respect to their input but linear with respect to their coefficients. Due to this feature they have good properties for many applications. The worst obstacles for the use of polynomial filters are the computational complexity and the numerical sensitivity of polynomials.

Rational function is the ratio of two polynomials and thus rational filters are direct generalizations of polynomial filters. So many similar considerations as for the polynomial filters apply also for the rational filters. As the class is much wider and from approximation theory we know that the approximation power of rational functions is much higher than that of polynomial functions, it also follows that the design is more difficult and robustness problems are more pronounced. Rational filters also possess good extrapolation capabilities and they can be trained using a linear algorithm.

Rational filters were used by Leung and Haykin [13] based on the work of Walsh [31] for signal detection and estimation, and were later applied in image filtering and enhancement by Ramponi [23, 24]. Cheikh *et al.* [3] extended these filters to vector rational filters and applied them to color image interpolation. Median-rational and vector median-rational hybrid filters were developed in [10] and [11], respectively, for the purpose of restoring images and color images corrupted by Gaussian and impulsive type noises.

3. Optimal filtering over the class of stack filters

Many nonlinear signal processing methods have their origin in statistics. In fact, the median filter was first introduced in statistics for smoothing economical time series [30]. It soon became evident that the median filter performs very well especially in image processing applications where sharp transitions are common. Particularly in urban or other "man-made" scenes we almost always have sharp edges and these edges usually are the most important information in the image. Attempts to retain sharp edges in linear filtering lead to "ringing" effects that are often more disturbing than noise.

In applications involving images, image sequences and color images, order statistics and their close kin morphological filters have by far been the most prominent and successful classes of nonlinear techniques, see [2], [4], [7], [5], [22], and [26].

One of the greatest limitations of order statistics filters is the fact that they are "smoothers". Without additional processing or combinations, their use remains limited to restoration applications, in which they excel especially in the presence of heavy-tailed noise (to be removed) and important signal details (to be preserved). General Boolean filters and morphological filters with non-flat structuring elements do not suffer from such shortcomings; however, they do not benefit from the stacking property which unifies all subclasses of stack filters; ranked order, median, weighted median and weighted order statistics filters. The stacking property says that the Boolean function that defines the filter is positive (or increasing as is the standard term in mathematical morphology). There is usually no underlying physical model that would demand the filter to be increasing. The power of increasing filters comes from the fact that concept narrows the filter class in a way that fits well to design processes.

Since each positive Boolean function (PBF) defines a stack filter, a simple exhaustive search through the class is not feasible. For instance, for the window size N the number of different stack filters is greater than $2^{2^N}/N$, [1, 27]. Fortunately, the properties of stack filters may be used to develop computationally efficient algorithms for the design of optimal stack filters. The result of optimization depends on the closeness criterion (or cost function) used for determining which filter restores the noisy signal closest to the ideal signal. Most commonly used closeness criteria are the mean absolute error (MAE), the mean square error (MSE) and minimax error. It is also useful, bearing in mind the application at hand, to restrict the class of possible stack filters by giving constraints to the optimization. The nature of these constraints can be either structural or statistical and more detailed information about them than what will given in here can be found for example in [19].

We consider here few of the several different approaches to optimize stack filters. Gabbouj and Coyle [6] first developed an approach called optimal filtering under structural constraints. Especially in many image processing problems it is not enough to know the "average" behavior of the filter; we need to be sure that it will handle certain signal segments in a prescribed way. This can be achieved using so-called structural constraints, the goal of which is to preserve some desired signal details, for example, pulses in 1-D signals, or lines in images, and to remove undesired signal patterns. In this approach such a stack filter is picked from the class of window width N stack filters that the average MAE per time unit between the filter's output and the desired signal is minimized. The MAE criterion reduces the estimation error of the stack filter to the sum of the decision errors incurred by the Boolean filters on each level of the threshold decomposition architecture. Now, a linear program can be used to determine if any stack filter of a particular window width has the desired structure-preserving of deleting properties. If some do, then the linear program selects the one minimizing the MAE between its output and the desired signal.

A basic statistical descriptor that can be used to study the noise attenuation properties and in the optimization of stack filters is the output distribution. Output distribution can be expressed by using coefficients A_i which give the number of those binary input vectors with i ones for which the output of the PBF equals zero. In [12] optimality theory of stack filters was extended and various new constraints were introduced. The optimization was performed in the mean square sense by finding a solution of the integer linear programming task of minimizing the output variance under the constraints for A_i and then determining a stack filter with the obtained coefficients A_i, if it exists. It must be emphasized that usually there is no guarantee that such a stack filter exists. However, once we have the target coefficients A_i the search for the optimal stack filter is simpler. We also can take a stack filter that has coefficients A_i close to the solution of the optimization problem and then check to see if its filtering behavior is satisfactory.

Properties of the coefficients A_i induce some basic constraints that need always to be satisfied but in a more meaningful situation we have additional constraints on the coefficients A_i that arise, for example, from requirements that the filter must have a certain degree of robustness and also be able to preserve details of a prescribed type. The constraints that give detail preservation can be given, for example, by fixing predetermined values of the defining Boolean function. These structural constraints induce new constraints for the coefficients A_i.

The rank selection probabilities give an intuitively appealing way of constraining a stack filter. For instance, a certain amount of robustness is guaranteed if we require that the rank selection probabilities corresponding to the largest and the smallest samples are equal to zero. This will give a stack filter that is "trimmed" in the same way as an L-filter with the coefficients corresponding to a number of the largest and smallest coefficients equal to zero. Constraints on the rank selection probabilities translate immediately into constraints on A_i. Similarly from breakdown points and breakdown probabilities we can achieve new constraints for the coefficients A_i.

In this optimization approach it is possible to use other constraints than those in [12] as long as they induce constraints to the coefficients A_i. This enables the use of recently introduced output distributional influence function (ODIF) [18] which is a tool designed for analyzing the robustness of the finite length filters. The ODIFs for the expectation and variance of a stack filter by using coefficients

A_i given in [17] are the most useful for the optimization purposes.

An adaptive filtering approach was developed by Lin *et al.* [14]. One more approach would be training based on a representative training set containing the ideal signal and the corrupted signal. A training framework for the optimal nonlinear filter design problem unsing MAE cost and fast procedures for obtaining optimal solutions have been developed for stack filters [29]. A more robust version of this optimization approach is given in [20] where a new robustness cost based on the ODIF is combined with the old MAE cost.

If we have information of the possible desired signal form expressed in Boolean vectors in such a form that it does not conflict the positivity of the defining Boolean function, designing optimal increasing filter becomes straightforward, [6], [12], [14], [32], and [33]. In some problems, notably in document image processing, noise is loosely speaking signal dependent binary union and intersection noise and increasing filters turn out to perform quite badly [15]. Here one must give up positivity and the penalty is that the large number of parameters and non-robust behavior of an unconstrained Boolean function makes the design of filters with large window sizes impossible. Recently there have emerged new ways to constrain the function leading to much better performance for large window sizes [25].

4. Challenges in filter design

What makes a good class of nonlinear filters? The answer is more indicative than concrete. The following suggestions might help in such a process:

- A class that is capable of modeling the underlying system in the application at hand. Because we never know the correct model parameters, but they have to be estimated from data, among the classes of filters that are capable of modeling the system one should choose one that is robust with respect to errors in the model. Especially when the filter parameters are found by a training procedure, the more there are parameters the more data is needed. In principle, there are methods based on stochastic complexity to analyze these concepts but no ready-made guidelines are yet available.
- A class that is simple to represent, easy to analyze and design. This is also related to the robustness aspect above in the sense that these properties enable us to verify the desired behavior.
- A class that is easy to implement. A simple example is low order polynomial filter that can be implemented very much along the lines of linear filters. Also there are many efficient algorithms for implementing median type filters.

Practical considerations:

- The ultimate goal of filtering: preprocessing for a detection or an estimation system, enhancement or restoration for human processing
- Amount and form of data available, is there model, training material
- Selection of the filter class
- Optimization

A unified and efficient framework for nonlinear filter design remains one of the most challenging tasks in this field. Even though we cannot hope to obtain a framework as powerful as the techniques for designing linear filters we should be able to build a methodology that would tie together the conditions and assumptions of the problem, the major nonlinear filter classes, relevant cost functions and accessible optimization algorithms. It is clear that the methodology must be able to deal with both statistical and deterministic aspects of the problem and filters. This framework cannot be obtained by one step (leap) but it will emerge as the result of incremental steps from the joint efforts of the signal processing community. However it is good to keep the ultimate goal in mind while solving problems for more immediate demands. Few attempts have been made to this end, see for instance [9] and [34]. Here we consider some problems whose solutions will clearly take us forward on this path.

We all agree that it would be important to be able to devise a feasible optimization procedure with a suitable cost function, even for a specific application, e.g. image restoration.

The unification of two or more existing filter classes will undoubtfully increase the modeling power of the framework. Therefore, it would be of great interest to determine the class of problems (signals) that can be solved (represented) by the new framework.

A related problem is that of the filter structure, or more specifically, the filter size. An often asked question is how large should the filter size be. Most of the answers have been try and see type. In [28], a solution was proposed to this problem, in which both the optimization and the filter structure were combined in a recursive manner.

Another equally important challenge to the nonlinear signal and image processing community is to develop new and attracting applications. Next to a mature theory (still developing), interesting applications would be the driving force to open up new frontiers in the field. Most of the current applications remain in the areas of signal (1- and M-D) restoration, enhancement, edge detection and interpolation. Recently, stack and Boolean filters were successfully used as predictors in a DPCM lossless image compression scheme, [21]. More such endeavors are needed in other areas such as speech analysis and processing, telecommunications and data analysis and communication.

5. Discussion and conclusions

The past two decades have witnessed the development of a number of successful nonlinear classes equipped with strong analytical tools, allowing the designer to solve a number of interesting problems, particularly in the area of image processing. Among these, one can cite the class of stack filters, [6], [14], and [29]; weighted median filters [32] and [33]; morphological filters [4] and [26]; polynomial filters [16]; and order statistics filters [22].

On the other hand, it is evident from the previous discussion that a number of important open problems and challenges remain to be solved. Among these, we mentioned a unifying framework involving two or more existing classes of nonlinear filters. It would be important to identify useful classes of signals and systems that can be modeled by this new and unifying framework. The hybrid approach, where two or more classes are combined, has proven to be useful in a number of applications involving conflicting constraints (e.g. removal of different types of noises and preservation of details) see e.g. [8], [10], and [11]. Although this is not what is meant by unification, it is nevertheless an indication of the additional capability resulting from combining different filter classes.

Finally, the applications areas must be further developed and diversified in order to attract users as well as young researchers to devote more efforts in the field.

References

[1] S. Agaian, J. Astola, and K. Egiazarian. *Binary Polynomial Transforms and Nonlinear Digital Filters*. Marcel Dekker, New York, 1995.

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

[3] F. A. Cheikh, L. Khriji, M. Gabbouj, and G. Ramponi. Color image interpolation using vector rational filters. In *Proc. SPIE Conference on Nonlinear Image Processing IX*, San Jose, CA, 24–30 January 1998.

[4] E. R. Dougherty and J. Astola. *An Introduction to Nonlinear Image Processing*, volume TT 16. SPIE Press, 1994.

[5] M. Gabbouj. Weighted median filtering – striking analogies to FIR filters. In C. Toumazou, editor, *Circuits and Systems Tutorials*, pages 5–21. IEEE ISCAS 94, Oxford, UK, 1994.

[6] M. Gabbouj and E. J. Coyle. Minimum mean absolute error stack filtering with structural constraints and goals. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(6):955–968, June 1990.

[7] M. Gabbouj, E. J. Coyle, and N. C. Gallagher, Jr. An overview of median and stack filtering. *Circuits, Systems, and Signal Processing, Special issue on Median and Morphological Filtering*, 11(1):7–45, 1992.

[8] P. Heinonen and Y. Neuvo. FIR-median hybrid filters with predictive FIR substructures. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 36:892–899, June 1988.

[9] V. G. Kamat, E. R. Dougherty, and J. Barrera. Bayesian multiresolution filter design. In *Proc. SPIE Nonlinear Image Processing XI*, volume 3961, pages 22–33, San Jose, California, 24–25 January 2000.

[10] L. Khriji and M. Gabbouj. Median-rational hybrid filters for image restoration. *Electronic Letters*, 34(10):977–979, May 1998.

[11] L. Khriji and M. Gabbouj. A new class of multichannel image processing filters: vector median-rational hybrid filters. *IEICE Trans. on Information and Systems*, E82-D(12):1589–1596, December 1999.

[12] P. Kuosmanen. *Statistical Analysis and Optimization of Stack Filters*. Number 77 in Acta Polytechnica Scandinavica, Electrical Engineering Series. Helsinki, 1994.

[13] H. Leung and S. Haykin. Detection and estimation using an adaptive rational function filter. *IEEE Trans. on Signal Processing*, 42(12):3366–3376, December 1994.

[14] J. H. Lin, T. M. Sellke, and E. J. Coyle. Adaptive stack filtering under the mean absolute error criterion. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(6):938–954, June 1990.

[15] R. P. Loce and E. R. Dougherty. *Enhancement and restoration of digital documents*. SPIE Optical Engineering Press, Bellingham, Washington, 1997.

[16] V. J. Mathews and G. L. Sicuranza. *Polynomial Signal Processing*. Wiley, 2000.

[17] S. Peltonen and P. Kuosmanen. Robustness of nonlinear filters for image processing. *Journal of Electronic Imaging*, 2001.

[18] S. Peltonen, P. Kuosmanen, and J. Astola. Output distributional influence function. In *Proc. of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, volume 1, pages 33–37, Antalya, Turkey, 20–23 June 1999.

[19] S. Peltonen, P. Kuosmanen, K. Egiazarian, M. Gabbouj, and J. Astola. Analysis and optimization of weighted order statistic and stack filters. In S. Mitra and G. Sicuranza, editors, *Nonlinear Image Processing*, chapter 1, pages 1–26. Academic Press, 2001.

[20] S. Peltonen, I. Tăbuş, J. Astola, E. Dougherty, and N. Hirata. Robust optimization of stack filters. In *Proc. of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Baltimore, Maryland USA, 3–6 June 2001.

[21] D. Petrescu, I. Tăbuş, and M. Gabbouj. Prediction capabilities of Boolean and stack filters for lossless image compression. *Multidimensional Systems and Signal Processing*, 10(2):161–187, April 1999.

[22] I. Pitas and A. N. Venetsanopoulos. *Nonlinear Digital Filters, Principles and Applications*. Kluwer Academic Publishers, 1990.

[23] G. Ramponi. The rational filter for image smoothing. *IEEE Signal Processing Letters*, 3(3):63–65, 1996.

[24] G. Ramponi. Image processing using rational functions. In *Proc. of Cost 254 Workshop*, Budapest, Hungary, 6–7 February 1997.

[25] O. V. Sarca, E. R. Dougherty, and J. T. Astola. Secondarily constrained Boolean filters. *Signal Processing*, 71(3):247–263, December 1998.

[26] J. Serra, editor. *Image Analysis and Mathematical Morphology*. Academic Press, 1988.

[27] I. Shmulevich, T. M. Sellke, M. Gabbouj, and E. J. Coyle. Stack filters and free distributive lattices. In *Proc of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, pages 927–930, Halkidiki, Greece, 1995.

[28] I. Tăbuş and M. Gabbouj. Fast order-recursive algorithms for optimal stack filter design. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2391–2394, Detroit, MI, 8–12 May 1995.

[29] I. Tăbuş, D. Petrescu, and M. Gabbouj. A training framework for stack and Boolean filtering – Fast optimal design procedures and robustness case study. *IEEE Trans. on Image Processing, Special issue on nonlinear image processing*, 5(6):809–826, June 1996.

[30] J. W. Tukey. Nonlinear (nonsuperposable) methods for smoothing data. In *Congr. Tec. EASCON-74*, page 673, 1974 (abstract only).

[31] J. L. Walsh. The existence of rational functions of best approximation. *Trans. Am. Math. Soc.*, 33:668–689, 1931.

[32] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo. Optimal weighted median filters under structural constraints. *IEEE Trans. on Signal Processing*, 43(3):591–604, March 1995.

[33] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. Weighted median filters: A tutorial. *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, 43:157–192, March 1996.

[34] B. Zeng, M. Gabbouj, and Y. Neuvo. A unified method for rank order, stack and generalized stack filters based on classical Bayes decision. *IEEE Trans. on Circuits and Systems*, 38(9):1003–1020, September 1991.