MINIMAX STACK FILTERING IN A PARAMETERIZED ENVIRONMENT

Moncef Gabbouj[†] and Edward J. Coyle[‡]

[†] Research Institute of Information Technology
Tampere University of Technology
P.O. Box 553, SF--33101 Tampere, Finland
[‡] School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907 USA

ABSTRACT

In this paper, a new optimization algorithm based on the minimax error criterion is presented. The algorithm assumes a parameterized environment where the input to the stack filter can be modelled by some parameterized stochastic process.

First, it is assumed that the signal parameters are fixed while the noise parameters vary over a pre-determined range. The algorithm's task is to pick an optimal stack filter, among the set of all stack filters, which minimizes the worst effect of the noise in the minimax sense. This optimization scheme is illustrated through an example.

Later, the optimization algorithm is generalized to allow parameter variations in the underlying input signal itself. This scheme can for instance deal with data-dependent noise. The optimization goal remains the same, that is, to pick an optimal stack filter which minimizes the worst effect of noise and signal parameter variations in the minimax sense.

Both algorithms can be solved using a 0-1 linear program. A Linear program would suffice where soft decisions are acceptable. In this case, the stacking property of the resulting filter is still preserved provided that it be interpreted in the probabilistic sense.

I. Introduction

In many estimation problems, the underlying input signal process and the corrupting noise process are often described each by some parameterized stochastic model. Variations in the model parameters usually have various effects on the performance of the estimation process. The aim of this paper is to develop an optimization algorithm to pick an optimal stack filter, among the class of stack filters, which minimizes the worst effect of parameter variations in the minimax sense.

Previous optimizations algorithms over the class of stack filters were all based on the mean absolute error criterion, i.e., optimization in L_1 [1-4] where the underlying processes are assumed to have fixed parameters. Though this error criterion might prove useful in image processing since it leads to the conditional median as the optimal estimator [1], there do exist other error criteria, such as the MSE and the maximum absolute deviation.

In this paper, the latter is chosen because a minimax solution guarantees that the errors associated with the approximation fall within specified limits where these limits were made as small as possible with no increase in the complexity of the calculations involved.

The most interesting factor in this optimization process is, perhaps, the fact that it can be accomplished using a linear program (LP). A similar approach was used in [5] where the aim was to pick the best stack filter which minimizes the maximum absolute deviation between the output signal and some desired input.

This paper is organized as follows. In the next section, the optimization theory is presented. It is shown that the best stack filter which minimizes the worst effect of noise in the minimax sense can be found via a linear program. In section III, an application example is shown. In section IV, the optimization scheme is further generalized to allow variations in the parameters of the signal model. Section V contains some conclusions.

II. Minimax Optimization in a Parameterized Stochastic Model

Suppose that the noise corrupting the signal comes from some parameterized stochastic process. The minimax problem will then select the best filter that will perform in a robust manner in the presence of various noise parameters.

Assume that the noise process $N(t)$ belongs to a certain family of stochastic processes parameterized by the set of parameters $\lambda \in \Lambda$. The set of parameters is at this point arbitrary, but finite. By conditioning on the value of λ, one would like to determine the positive Boolean function $f(\cdot)$ which minimizes the worst effect of the noise under different values of λ.

An instance of a minimax problem is:

$$\min_{f \in B_n} \max_{\bar{\lambda} \in \Lambda} E\left[|S(t) - S_f(W_\lambda(t))| \mid \lambda = \bar{\lambda} \right], \quad (13)$$

where $S_f(\cdot)$ is a stack filter based on the n-variable positive Boolean function $f \in B_n$, $S(t)$ is the M+1-valued input signal, and $W_\lambda(t)$ is the window state at time t.

0-7803-0593-0/92 $3.00 1992 IEEE

Using the threshold decomposition and the stacking property, rewrite Eq. (13) as follows.

$$\min_{f \in B_n} \max_{\bar{\lambda} \in \Lambda} \sum_{l=1}^{M} E\left[|s_l - f(w_{\lambda l})| \mid \bar{\lambda} = \lambda \right]. \tag{14}$$

Consider the expression inside the summation and use successive conditioning on $w_{\lambda l}$.

$$E\left[|s_l - f(w_{\lambda l})| \mid \lambda = \bar{\lambda} \right] =$$
$$E_W\left[E\left[|s_l - f(\bar{w}_{\lambda l})| \mid \lambda = \bar{\lambda}, w_{\lambda l} = \bar{w}_{\lambda l} \right] \right]. \tag{15}$$

Now, the inner expectation is a function of $\bar{\lambda}$ and $\bar{w}_{\lambda l}$. From previous analysis in the MAE case, Eq. (15) is equivalent to:

$$E_W\left[\alpha^{(l)}(\bar{\lambda},\bar{w}) P_f(1 \mid \bar{w},\bar{\lambda}) + \beta^{(l)}(\bar{\lambda},\bar{w}) \right], \tag{16}$$

where

$$\alpha^{(l)}(\bar{\lambda},\bar{w}) = 1 - 2\pi(1 \mid \bar{w},\bar{\lambda}) \quad and \quad \beta^{(l)}(\bar{\lambda},\bar{w}) = \pi(1 \mid \bar{w},\bar{\lambda}).$$

Taking the expectation over W reduces to:

$$\sum_{\bar{w} \in Q_W} \left[\alpha^{(l)} \pi(\bar{w}) P_f(1 \mid \bar{w},\bar{\lambda}) + \beta_{\bar{\lambda}}^{(l)} \pi(\bar{w}) \right] =$$
$$\sum_{i=1}^{2^n} \bar{\alpha}_i^{(l)} P_f(1 \mid \bar{w}_{\bar{\lambda} i}) + \bar{\beta}_{\bar{\lambda}}^{(l)}, \tag{17}$$

where

$$\bar{\alpha}_i^{(l)} = \alpha^{(l)}(\bar{\lambda},\bar{w}_i)\pi(\bar{w}_i) \quad and \quad \bar{\beta}_{\bar{\lambda}}^{(l)} = \sum_{i=1}^{2^n} \beta^{(l)}(\bar{\lambda},\bar{w}_i)\pi(\bar{w}_i).$$

Now, Eq. (14) can be rewritten as:

$$\min_{f \in B_n} \max_{\bar{\lambda} \in \Lambda} \sum_{l=1}^{M} \left[\sum_{i=1}^{2^n} \bar{\alpha}_i^{(l)} P_f(1 \mid \bar{w}_{\bar{\lambda} i}) + \bar{\beta}_{\bar{\lambda}}^{(l)} \right], \tag{18}$$

which is equivalent to:

$$\min_{f \in B_n} \max_{\bar{\lambda} \in \Lambda} \sum_{i=1}^{2^n} \hat{\alpha}_i P_f(1 \mid \bar{w}_{\bar{\lambda} i}) + \hat{\beta}_{\bar{\lambda}}, \tag{19}$$

where $\hat{\alpha}_i = \sum_{l=1}^{M} \bar{\alpha}_i^{(l)}$ and $\hat{\beta}_{\bar{\lambda}} = \sum_{l=1}^{M} \bar{\beta}_{\bar{\lambda}}^{(l)}$.

Now, suppose that $\Lambda \subset \mathcal{E}^d$,i.e., $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_d\}^d$; $\lambda_i \in R$. In Eq. (19) above, let

$$X_k = \sum_{i=1}^{2^n} \hat{\alpha}_{\bar{\lambda}_k,i} P_f(1 \mid \bar{w}_{\bar{\lambda}_k,i}) + \hat{\beta}_{\bar{\lambda}_k}, \tag{20}$$

then the new LP which will find the optimal stack filter for the multiple level input has the following structure:

$$\begin{array}{ll} \min & u \\ s.t. & X_k \le u \quad k=1,\ldots,d \\ & AX \le b \\ & X, u \ge 0. \end{array} \tag{21}$$

A is the matrix containing the stacking constraints, see [2].

Remark: The size of the matrix A is $O(n2^n)$, thus if $d \le n2^n$, then the minimax LP and the MAE LP have the same order of complexity. Furthermore, if $d \ll 2^n$, then the multiple level minimax LP contains fewer inequality straints than the previous binary minimax LP, see [5

However, one must mention that the overhead for t multiple level case is much more involved.

III. Applications

In this section, some examples will be treated in which some stochastic models for the signal and the noise are assumed.

III.1. Description of the Signal

In the following optimization examples, the signal $s(t)$, t=0,1,... is an irreducible, discrete-time Markov chain with state space $Q_s = \{q_1, q_2, \ldots, q_M\}$ and state transition matrix

$$P_s = \begin{bmatrix} p_{11} & p_{1M} \\ p_{21} & p_{2M} \\ \vdots & \vdots \\ p_{M1} & p_{MM} \end{bmatrix}.$$

Since all the information about the statistics of the different threshold levels is contained in the coefficients of Eq. (20), one can just consider a binary valued signal as the input to the filter. This neither reduces nor increases the complexity of the LP that finds the optimal stack filter, but it does reduce the overhead.

Once the state transition matrix P_s of the signal has been determined, the following system of linear equations is solved:

$$h(j) = \sum_{i \in E} h(i) P_s(i,j); \quad j \in E, \tag{22}$$

with

$$\sum_{j \in E} h(j) = 1, \tag{23}$$

where E is the index set of the states in Q_s. This produces the limiting probabilities of the states of the signal which are all recurrent non-null.

III.2. Description of the Noise

The noise n(t), t=0,1,2,..., corrupting the signal, is also an irreducible, discrete-time Markov chain with state space $Q_n = \{X, Y, Z\}$ and state transition matrix P_n parameterized by a parameter set $\lambda = (\lambda_1,\lambda_2) \in \Lambda \subset \mathcal{E}^2$.

$$P_n = \begin{bmatrix} \lambda_1 & \lambda_2 & 1-\lambda_1-\lambda_2 \\ \lambda_2 & \lambda_1 & 1-\lambda_1-\lambda_2 \\ 1-\lambda_1-\lambda_2 & \lambda_2 & \lambda_1 \end{bmatrix},$$

where $0 \le \lambda_1,\lambda_2 \le 1$ and $0 \le \lambda_1 + \lambda_2 \le 1$. Again, the noise is assumed to be independent of the signal. The states of the noise in Q_n act upon the signal in the following manner:

If the noise is in state:

X _it does not interfere with the signal,
Y _it drives the signal to its highest value,
Z _it drives the signal to its lowest value.

Solving the system of linear equations in (22) with condition (23), the following limiting probabilities for the rameterized noise process are obtained:

$$v(X)=\frac{\dfrac{(\lambda_1+2\lambda_2-1)\lambda_2}{1-\lambda_1+\lambda_2}+1-\lambda_1-\lambda_2}{2-2\lambda_1-\lambda_2},$$

$$v(Y)=\frac{\lambda_2}{1-\lambda_1+\lambda_2},$$

and

$$v(Z)=1-v(X)-v(Y).$$

III.3. Description of the Received Process

The process which appears at the input of the binary stack filter is called r(t), t=0,1,2,... This received binary signal is an irreducible finite state Markov chain; furthermore, it is assumed to be a memoryless function of the signal and the noise processes. Since the signal and the noise are independent, the received process is the output of a memoryless function operating on a Markov process. The state space Q_r associated with this process is the product space $Q_s \times Q_n$ and its transition matrix is

$$P_r = P_s \otimes P_n,$$

see [2], where $\otimes$ denotes the Kronecker product for matrices.

III.4. Constructing the Binary Window Process

At any time t, there is a sequence of n bits in the filter's window, (sometimes, n is referred to as the window width of the filter). Therefore, the process which specifies the transitions between all the possible n-bit sequences (there are 2^n of them) is constructed from the elementary single sample definition of the received binary process.

Let Q_w denote the state space of the window process, then $Q_w=(Q_r)^n$, that is Q_r crossed with itself n times. Label the states in Q_w as follows: $Q_w=\{q_1,q_2,\ldots,q_{m^n}\}$, where m is the number of states in Q_r and each q_i is a binary sequence of n bits. The state transition matrix P_w was found in [2] to be

$$P_w=\sum_{i=1}^{m} 1_{m,i} \otimes (1_{m,i}^T \cdot P_r) \otimes I_{m^{n-2}} \otimes 1_m, \tag{24}$$

where

$1_{m,i}$ is a column vector with a 1 in the i'th entry.
P_r is an $(m \times m)$ square matrix, the state transition matrix for the received process.
1_m is a column vector of m 1's.
$I_{m^{n-2}}$ is the $(m^{n-2} \times m^{n-2})$ identity matrix.
T denotes the transpose operation.

The limiting probabilities of the binary window process with the above state transition matrix are arranged in vector $\bar{\pi}$

$$\bar{\pi}=[\pi(w_1),\pi(w_2),\cdots,\pi(w_{m^n})],$$

where $\pi(w_i)=\lim_{t\to\infty} P(w(t)=w_i)$.

More specifically, consider the following state space of the input signal taken from [6],

$$Q_s=\{(0,a),(0,b),(0,c),(1,a),(1,b),(1,c)\},$$

and the state transition matrix is

$$P_s=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0.4 & 0 & 0 & 0 & 0 & 0.6 \end{bmatrix}.$$

A careful examination of P_s reveals that any signal, that is a Markov chain, with the above state transition matrix must be a root signal to a window width five median filter.

Solving the system of linear equations in (22) with condition (23), one obtains the limiting probability vector h whose components are:
h(0,a) = 0.12245 , h(0,b) = 0.12245 , h(0,c) = 0.20408
h(1,a) = 0.12245 , h(1,b) = 0.12245 , h(1,c) = 0.30612.

III.5. The Received Signal

The state space of the received signal is $Q_r = Q_s \times Q_n$ with $6 \times 3 = 18$ states. The noise process was described earlier. The parameterized state transition matrix of the received process is

$$P_r=P_s \otimes P_n=\begin{bmatrix} 0 & P_n & 0 & 0 & 0 & 0 \\ 0 & 0 & P_n & 0 & 0 & 0 \\ 0 & 0 & 0.4P_n & 0.6P_n & 0 & 0 \\ 0 & 0 & 0 & 0 & P_n & 0 \\ 0 & 0 & 0 & 0 & 0 & P_n \\ 0.4P_n & 0 & 0 & 0 & 0 & 0.6P_n \end{bmatrix}.$$

Then, again the window process is formed and the limiting probabilities of the corresponding window state space Q_w are computed as was outlined in [6]. Next, the cost coefficients are generated and the LP is used to solve for an optimal filter.

III.6. Examples

Let the parameter set be further constrained as follows,

$$\frac{1}{2} \le \lambda_1 < 1, \quad 0 < \lambda_2 < \frac{1}{2}, \quad \frac{1}{2} \le \lambda_1 + \lambda_2 < 1,$$

and divide the unit interval into 10 equal intervals. Removing all combinations of λ_1,λ_2 which produce reducible Markov chains, one is left with the following values of the parameters:

$$\Lambda=\{(0.5,0.1),(0.5,0.2),(0.5,0.3),(0.5,0.4),(0.6,0.1),$$
$$(0.6,0.2),(0.6,0.3),(0.7,0.1),(0.7,0.2),(0.8,0.1)\}.$$

According to Eq. (21), 25 additional constraints will be added to the stacking constraints to form the over all constraint matrix. Once again, one is confronted with randomization. The additional constraints have real valued coefficients and the over all constraint matrix is no longer totally unimodular [7]. However, simulation results show that when the number of constraints added to the stacking constraints is much smaller than 2^n, then only few decisions will be soft decisions. In the simulations shown below ecision variables take on exactly one of these three value x, 1 with $0 \le x \le 1$ in most cases.

As was previously done in the MAE case, one ca modify one of the cost coefficients, for instance, $C(\vec{w}_j,1,$ and get different optimal filters.

Table 1: Sub-Optimal ww 5 filters for various cost factors.

$C(\bar{w}_j,1,0)$	$f(x_1,x_2,x_3,x_4,x_5)$	*W.MNME*
0.1-0.4	0	0.050-0.220
0.415 to 0.45	$x_1x_2x_3+x_2x_3x_4+x_3x_4x_5$ or $x_1x_2x_3+x_3x_4$	0.228 to 0.235
0.4625	$x_1x_2x_3+x_3x_4$	0.237
0.475 to 0.5	$x_1x_2x_3+x_3x_4$ or $x_2x_3+x_3x_4$	0.239 to 0.243
0.55 to 0.65	$x_2x_3+x_3x_4x_5$ or x_3	0.253 to 0.269
0.75-1.0	x_3	0.287-0.333
1.25	$x_3+x_2x_4x_5+x_1x_2x_4$	0.382
1.5	x_3 or x_2+x_3	0.435
2.5	1	0.450

The first and the last entry in the above table are both hard or binary decisions. Entries with a single positive Boolean expression correspond to tertiary decisions with values 0,1, and $0.9 \le x \le 1$ which have been rounded up. The top expressions in the multiple function entries were obtained by ceiling soft decisions to 1 while the bottom ones were obtained by flooring soft decisions to 0. Note that the ranges of the error in the table above correspond to the optimal value of the objective of the optimal solution with soft decisions.

IV. Model Generalization

An immediate generalization of the above minimax LP model for treating the multiple level case is the following.

Suppose that the input signal itself belongs to some parametrized class of a stochastic process. This is actually more realistic than assuming specific values of the model for the input signal. These parameters are usually estimated before an optimization procedure is adopted. These values may cover a whole range of numbers and picking one of them up would only be arbitrary. So, what if one could incorporate this parameter estimation into the estimation of the signal itself. Then one would hopefully have a better estimate of the input signal.

To avoid a more cumbersome notation, let us assume that the input signal comes from a stochastic process parameterized again by a parameter set Λ. Now Λ contains the noise parameters as well as the signal parameters. An instance of a minimax problem could be:

$$\min_{f} \max_{\bar{\lambda}} E\left[|S_\lambda(t) - S_f(W_\lambda(t))| \,|\, \lambda = \bar{\lambda} \right].$$

Note that the only difference between this equation and Eq. [illegible] is the subscript on the multi-valued input signal.

Using the threshold decomposition, the stack [illegible] property and successively conditioning on the window process state, one obtains an expression similar to Eq. (15) except for the thresholded version of the input signal s_l which will have a second subscript λ.

Working out the steps as in the previous section, the final form of the minimax LP would look like:

$$\begin{aligned} \min \quad & u \\ s.t. \quad & X_k \le u \quad k=1,\ldots,d \\ & AX \le b \\ & X, u \ge 0, \end{aligned}$$

where d is the cardinality of the parameter set Λ and A is the matrix containing the stacking constraints. As for the complexity of this LP compared to that of the minimum mean absolute error, as was stated previously, as long as $d \le n2^n$, both LP's have essentially the same order of complexity.

V. Conclusions

In this paper, a new optimization theory for stack filters was presented. The minimax criterion was used to select the best stack filter which minimizes the worst effect of the noise in a parameterized stochastic model. Several examples were tabulated.

The problem of randomization appeared. However, it was less severe in this case than it was in [5] since all soft decisions made by the LP assumed in most cases the same values. Once again, sub-optimal solutions would have to be accepted unless a 0-1 LP is used.

The optimization algorithm was later extended to allows variations in the input signal parameters. The algorithm's task is again to pick the best stack filter in the minimax sense which minimizes the worst effect of noise and signal parameter variations.

LIST OF REFERENCES

[1] E.J. Coyle, J.-H. Lin, and M. Gabbouj, "Optimal stack filtering and the estimation and structural approaches to image processing," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 37, Dec. 1989, pp. 2037-2066.

[2] E.J. Coyle and J.-H. Lin, "Stack filters and the mean absolute error criterion," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-36, no. 8, pp. 1244-1254, Aug. 1988.

[3] E.J. Coyle, "Rank order operators and the mean absolute error criterion," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 63-76, Jan. 1988.

[4] 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*, vol. 38, pp. 955-968, June 1990.

[5] M. Gabbouj and E.J. Coyle, "Minimax Optimization over the Class of Stack Filters," *Proceedings of the SPIE Visual Communications and Image Processing'90*, Oct. 1-4, 1990, Lausanne, Switzerland, pp. 143-154.

[6] M. Gabbouj, "Optimal Stack Filter Examples and Positive Boolean Functions," *MS Thesis*, Purdue University, West Lafayette, IN, Dec. 1986.

[7] M. Gabbouj and E.J. Coyle, "On the LP which Finds a MMAE Stack Filter," *IEEE Trans. on Signal Processing*, Vol. 39, Nov. 1991, pp. 2419-2424.