ADAPTIVE ENVELOPE-CONSTRAINED FILTERING

Ba-Ngu Vo, Sumeetpal Singh and Vladislav Tadic

bv@ee.mu.oz.au, ssss@ee.mu.oz.au, vtadic@ee.mu.oz.au
Department of Electrical and Electronic Engineering,
The University of Melbourne, Vic. 3010, Australia

ABSTRACT

In the discrete-time Envelope-Constrained filtering problem, the gain of the filter is minimised subject to the constraint that the filter output to a prescribed input fits into a given envelope. In this paper, a novel adaptive algorithm for solving this problem based on stochastic optimisation is presented. The algorithm is simple to implement on-line and convergence is demonstrated in numerical examples. Under mild regularity assumptions convergence follows from standard stochastic approximation results.

1. INTRODUCTION

The (discrete-time) Envelope-Constrained (EC) filtering problem involves the design of a linear time-invariant (LTI) filter such that the response ψ to a specified excitation s fits into a prescribed template or envelope, ε^+, ε^- [1] as shown in Figure 1. The envelopes can arise either from practical considerations or from the standards set by certain regulatory bodies. For example, in pulse-compression for radar, the envelopes are selected to suppress side lobes while keeping the main lobe above a certain threshold [1]; in telecommunications pulse shapes used in transmission systems are specified using templates issued by standard bodies such as CCITT or ANSI (see e.g. [2]-[4]). For a technically oriented survey of the subject the reader is referred to [5].

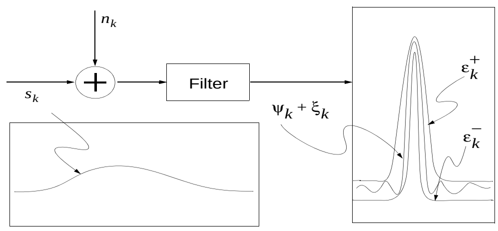

Fig. 1. Envelope Constrained Filter

The EC problem was formulated for FIR filters in [1] and IIR filters in [6]. The resulting optimisation problem can be solved off-line by Quadratic Programming (QP) via active set strategy [5]. Direct use of this algorithm would be inappropriate in applications where the parameters of the underlying signal model are either not known or varying with time. In such cases, it may be necessary to employ an adaptive filter with parameters that can be adjusted to their optimum value (see Figure 2). In the most common procedure, a test or training signal corrupted by noise $\widehat{s}$ is periodically used as the filter input. The filter response $\widehat{\psi}$ (to the training signal $\widehat{s}$) is checked against the boundaries ε^+, ε^- of the template. The result of the comparison is then processed in some way and fed back to adjust the filter coefficients. This process is repeated until, for practical purposes, convergence has occurred and the filter is ready to process data. Additional test pulses are then inserted into the data stream at regular intervals so that the filter can continue to be adjusted.

Adaptive algorithms for EC filters have been proposed in [7] and [8] based on the dual formulation. In [7] the problem was converted to an unconstrained non-smooth dual problem, which was then solved using subgradients, while in [8] the smooth constrained dual problem was solved using gradient flow. In [9], a modified penalty technique was used to approximate the primal problem as a smooth unconstrained problem which can be solved using descent methods. These algorithms converge under noise-free condition, however, for noisy input signals, no useful results on convergence has been established.

In this paper, we propose an adaptive EC filtering algorithm based on standard stochastic approximation techniques where convergence can be established under mild regularity assumptions using standard stochastic approximation results as in [10]. The proposed algorithm is based on solving the dual problem using stochastic gradient ascent with projection. Our algorithm is simple to implement and exhibits good convergence characteristic in our numerical examples.

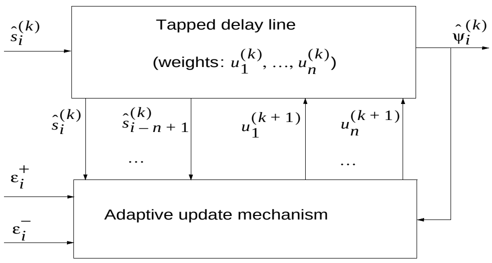

Fig. 2. An adaptive configuration for EC filter.

2. BACKGROUND

Let $\mathbf{u} = [u_1, ..., u_n]^T \in \mathbf{R}^n$ denote the vector of tap coefficients of the FIR filter, and $\mathbf{x} = [x_1, ..., x_m]^T \in \mathbf{R}^m$ denote the vector representing a finite support input signal. The response $\psi = [\psi_1, ..., \psi_N]^T \in \mathbf{R}^N, N = m + n - 1$, of the FIR filter $\mathbf{u}$ to the input $\mathbf{x}$ is given by

$$\psi = Y(\mathbf{x})\mathbf{u} \tag{1}$$

where $Y(\mathbf{x})$ is the following Toeplitz matrix

$$Y(\mathbf{x}) = \begin{bmatrix} x_1 & 0 & \cdots & 0 & 0 \\ \vdots & x_1 & & & \vdots \\ x_m & \vdots & & & 0 \\ 0 & x_m & & & 0 \\ 0 & 0 & & & x_1 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & x_m \end{bmatrix} \tag{2}$$

The mapping $\mathbf{x} \mapsto Y(\mathbf{x})$ is linear in $\mathbf{x}$. Hence, for a noise corrupted input $\mathbf{x} = \mathbf{s} + \mathbf{n}$, the filter output is $Y(\mathbf{s})\mathbf{u}+Y(\mathbf{n})\mathbf{u}$. In EC filtering, for a given signal $\mathbf{s}$ the noise-free response, $Y(\mathbf{s})\mathbf{u}$, is required to fit in a prescribed envelope. Using the filter norm for the cost function, the EC filtering problem can be posed as the following QP problem [1]

$$\begin{aligned} \min_{\mathbf{u}\in\mathbf{R}^n} f(\mathbf{u}) &= \mathbf{u}^T\mathbf{u} \\ \text{subject to } \epsilon^- &\leq Y(\mathbf{s})\mathbf{u} \leq \epsilon^+ \end{aligned} \tag{3}$$

where $\epsilon^-, \epsilon^+ \in \mathbf{R}^N$ are the upper and lower boundaries of the mask. Problem (3) involves a strictly convex cost and convex constraint, hence the solution is unique if one exists.

Assuming feasibility of primal problem, the dual problem is also a QP problem (see [1], [5], [8])

$$\begin{aligned} \max_{\mathbf{p}\in\mathbf{R}^{2N}} \phi(\mathbf{p}) &= \mathbf{p}^T A(\mathbf{s},\mathbf{s})\mathbf{p} + \mathbf{p}^T\mathbf{b} \\ \text{subject to } \mathbf{p} &\geq \mathbf{0} \end{aligned} \tag{4}$$

where for $\mathbf{w}, \mathbf{x} \in \mathbf{R}^m$

$$A(\mathbf{w},\mathbf{x}) \equiv -\frac{1}{4}\begin{bmatrix} I_N \\ -I_N \end{bmatrix} Y(\mathbf{w})Y^T(\mathbf{x})[I_N, -I_N] \tag{5}$$

$$\mathbf{b} = \begin{bmatrix} -\epsilon^+ \\ \epsilon^- \end{bmatrix} \tag{6}$$

For a fixed dual point $\mathbf{p}$, the corresponding primal solution which minimises the Lagrangian is given by

$$\mathbf{u}^o(\mathbf{p}) = -\frac{1}{2}Y^T(\mathbf{s})[I_N, -I_N]\mathbf{p} \tag{7}$$

If $\mathbf{p}^o$ is an optimal solution to Problem (4), then the optimal solution to Problem (3) is $\mathbf{u}^o(\mathbf{p}^o)$. Both the primal (3) and dual (4) problems can be solved off-line using standard tools such as QP via active set strategy [5].

3. STEEPEST ASCENT

Consider the general constrained problem

$$\max_{\theta\in\Theta} \phi(\theta) \tag{8}$$

where $\Theta \subseteq \mathbf{R}^d$ is a closed convex set. Let π_Θ denote the projection operator that maps every point $\mathbf{x} \in \mathbf{R}^d$ onto the unique point $\theta = \pi_\Theta(\mathbf{x}) \in \Theta$ defined as the closest point in Θ to $\mathbf{x}$. The projected steepest ascent algorithm to solve (8) generates the following iterates:

$$\theta^{(j+1)} = \pi_\Theta\left(\theta^{(j)} + \alpha^{(j)}\nabla\phi(\theta^{()j})\right) \tag{9}$$

where $\{\alpha^{(j)}\}$ is a positive sequence known as the step-size and $\nabla\phi$ is the gradient. The step-size sequence satisfies $\alpha^{(j)} \geq 0$, $\alpha^{(j)} \to 0$, $\sum \alpha^{(j)} = \infty$. Typically, $\alpha^{(j-1)} = 1/j$. The parameter $\theta^{(j)}$ is updated in the direction of increase of ϕ, followed by a projection into the feasible domain. Under suitable regularity conditions, the sequence $\{\theta^{(j)}\}$ converges to the maximizer of ϕ in Θ [10].

In the context of Problem (4) the gradient is given by

$$\nabla\phi(\mathbf{p}) = 2A(\mathbf{s},\mathbf{s})\mathbf{p} + \mathbf{b} \tag{10}$$

Note from (1), (7) and (5) that the gradient can be expressed in terms of the filter output and the envelope boundaries:

$$\nabla\phi(\mathbf{p}) = \begin{bmatrix} Y(\mathbf{s})\mathbf{u}^o(\mathbf{p}) - \epsilon^+ \\ \epsilon^- - Y(\mathbf{s})\mathbf{u}^o(\mathbf{p}) \end{bmatrix} \tag{11}$$

The projection $\mathbf{p} \mapsto \max(\mathbf{p}, \mathbf{0})$, where the max operator is taken component wise, can be easily implemented in hardware using simple circuitries.

The primal-dual update equations are then

$$\mathbf{u}^{(j)} = -\frac{1}{2}Y^T(\mathbf{s})[I_N, -I_N]\mathbf{p}^{(j)} \tag{12}$$

$$\mathbf{p}^{(j+1)} = \max\left(\mathbf{p}^{(j)} + \alpha^{(j)}\nabla\phi(\mathbf{p}^{(j)}), \mathbf{0}\right) \tag{13}$$

Since $\{\mathbf{p}^{(j)}\}$ converges to $\mathbf{p}^o$, it is clear that $\{\mathbf{u}^{(j)}\}$ converges to $\mathbf{u}^o(\mathbf{p}^o)$.

4. ADAPTIVE ALGORITHM

In an on-line scenario, exact knowledge of the noise-free signal $\mathbf{s}$ is not available. Adaptive algorithms periodically use noisy observations of $\mathbf{s}$ as training inputs to adjust the weights of the filter to solve Problem (3). One way of achieving this is to solve Problem (4) based on a stochastic counterpart of the primal–dual update equations (12) and (13).

A stochastic counterpart to (9) would correspond to the case where one does not have access to $\nabla\phi$ but only random unbiased observations $\widehat{\nabla\phi}(\theta^{(j)})$, i.e.

$$\mathbf{E}\{\widehat{\nabla\phi}(\theta^{(j)})\} = \nabla\phi(\theta^{(j)}).$$

We may relax the condition of unbiasedness to a condition concerning the convergence of the bias to zero. When (9) is iterated with $\widehat{\nabla\phi}(\theta^{(j)})$ instead, we still have convergence [10] (under suitable regularity conditions) to maximizer of ϕ.

When the jth training pulse is sent, the actual input to the EC filter can be modelled as $\widehat{\mathbf{s}}^{(j)} = \mathbf{s} + \mathbf{n}^{(j)}$, where $\mathbf{n}^{(j)}$ is a zero-mean i.i.d. noise process. Below, we describe 2 different techniques to estimate the gradient $\nabla\phi$, using the noisy training signals $\widehat{\mathbf{s}}^{(j)}$.

From (10) and (11), we see that gradient is known except for the terms involving the matrix $Y(\mathbf{s})Y^T(\mathbf{s})$. Hence, an estimator for the desired gradient can be found once we estimate $Y(\mathbf{s})$. Each time a noise corrupted training pulse $\widehat{\mathbf{s}}^{(j)}$ arrives, we have an unbiased observation $Y(\widehat{\mathbf{s}}^{(j)})$ of the matrix $Y(\mathbf{s})$. Let

$$Y(\widetilde{\mathbf{s}}^{(j)}) = Y\left(\frac{1}{j}\sum_{i=1}^{j}\widehat{\mathbf{s}}^{(i)}\right) = \frac{1}{j}\sum_{i=1}^{j} Y(\widehat{\mathbf{s}}^{(i)}) \qquad (14)$$

be the empirical average, where $\widetilde{\mathbf{s}}^{(j)}) = \frac{1}{j}\sum_{i=1}^{j}\widehat{\mathbf{s}}^{(i)}$. Then, by the Law of Large Numbers $\lim_{j\to\infty} Y(\widetilde{\mathbf{s}}^{(j)}) = Y(\mathbf{s})$ almost surely (a.s.). Hence,

$$\lim_{j\to\infty} Y(\widetilde{\mathbf{s}}^{(j)})Y^T(\widetilde{\mathbf{s}}^{(j)}) = Y(\mathbf{s})Y^T(\mathbf{s}) \quad \text{a.s.}$$

and the gradient estimator

$$\widehat{\nabla\phi}(\mathbf{p}^{(j)}) = 2A(\widetilde{\mathbf{s}}^{(j)}, \widetilde{\mathbf{s}}^{(j)})\mathbf{p}^{(j)} + \mathbf{b} \qquad (15)$$

is asymptotically unbiased. The estimation of $Y(\mathbf{s})$ in (14) can be implemented by the following recursion:

$$Y(\widetilde{\mathbf{s}}^{(j)}) = \frac{1}{j}[(j-1)Y(\widetilde{\mathbf{s}}^{(j-1)}) + Y(\widehat{\mathbf{s}}^{(j)})] \qquad (16)$$

The stochastic primal-dual update equations are then

$$\mathbf{u}^{(j)} = -\frac{1}{2}Y^T(\widetilde{\mathbf{s}}^{(j)})[I_N, -I_N]\mathbf{p}^{(j)} \qquad (17)$$

$$\mathbf{p}^{(j+1)} = \max\left(\mathbf{p}^{(j)} + \alpha^{(j)}\widehat{\nabla\phi}^{(j)}(\mathbf{p}^{(j)}), \mathbf{0}\right) \qquad (18)$$

As in [7] and [8] we could also use the following unbiased estimator of the gradient which requires 2 training pulses $\widehat{\mathbf{s}}^{(j)}$ and $\widehat{\mathbf{s}}'^{(j)}$ per iteration:

$$\widehat{\nabla\phi}(\mathbf{p}^{(j)}) = 2A(\widehat{\mathbf{s}}'^{(j)}, \widehat{\mathbf{s}}^{(j)})\mathbf{p}^{(j)} + \mathbf{b} \qquad (19)$$

Unbiasedness follows since the noise components of the training pulses $\widehat{\mathbf{s}}^{(j)}$ and $\widehat{\mathbf{s}}'^{(j)}$ are independent. Moreover, the gradient estimate can be obtained from the filter output, i.e. no explicit computation of the matrix $Y(\widehat{\mathbf{s}}'^{(j)})Y^T(\widehat{\mathbf{s}}^{(j)})$ in (19) is required. On arrival of $\widehat{\mathbf{s}}^{(j)}$, the filter $\mathbf{u}^{(j)}$ is updated by cross-correlating $\widehat{\mathbf{s}}^{(j)}$ with $[I_N, -I_N]\mathbf{p}^{(j)}$, i.e.

$$\mathbf{u}^{(j)} = -\frac{1}{2}Y^T(\widehat{\mathbf{s}}^{(j)})[I_N, -I_N]\mathbf{p}^{(j)} \qquad (20)$$

When the accompanying training pulse $\widehat{\mathbf{s}}'^{(j)}$ arrives, the output of $\mathbf{u}^{(j)}$ is $\widehat{\psi}^{(j)} \equiv Y(\widehat{\mathbf{s}}'^{(j)})\mathbf{u}^{(j)}$, which can be written as

$$\widehat{\psi}^{(j)} = -\frac{1}{2}Y(\widehat{\mathbf{s}}'^{(j)})Y^T(\widehat{\mathbf{s}}^{(j)})[I_N, -I_N]\mathbf{p}^{(j)} \qquad (21)$$

Hence, the gradient estimate (19) can be written as

$$\widehat{\nabla\phi}(\mathbf{p}^{(j)}) = \begin{bmatrix} \widehat{\psi}^{(j)} - \epsilon^+ \\ \epsilon^- - \widehat{\psi}^{(j)} \end{bmatrix} \qquad (22)$$

Even though the sequence of dual iterates converges to an optimal dual solution $\mathbf{p}^o$ a.s. (and hence the primal solution $\mathbf{u}^o(\mathbf{p}^o)$ is optimal), the primal iterate $\mathbf{u}^{(j)}$ has a very small probability of being feasible. This is because the optimal solution $\mathbf{u}^o(\mathbf{p}^o)$ lies on the boundary of the feasible set and is obtained using $\mathbf{s}$ in (7), whereas the primal iterate is obtained by using $\widehat{\mathbf{s}}^{(j)} \neq \mathbf{s}$ in (20).

To overcome this difficulty, the empirical average in (14) is used in the calculation of the primal iterate $\mathbf{u}^{(j)}$, but the gradient is obtained from the response $\widehat{\psi}^{(j)}$ of the updated filter $\mathbf{u}^{(j)}$ to the pulse $\widehat{\mathbf{s}}'^{(j)}$. More concisely, $\mathbf{u}^{(j)}$, $\widehat{\nabla\phi}(\mathbf{p}^{(j)})$ are computed according to (17) and (22) respectively. Under this update scheme, (22) can be written as

$$\widehat{\nabla\phi}(\mathbf{p}^{(j)}) = 2A(\widehat{\mathbf{s}}'^{(j)}, \widetilde{\mathbf{s}}^{(j)})\mathbf{p}^{(j)} + \mathbf{b}, \qquad (23)$$

which is clearly an unbiased estimator since the noise components of all the pulses $\widehat{\mathbf{s}}^{(1)}, \widehat{\mathbf{s}}^{(2)}, ..., \widehat{\mathbf{s}}^{(j)}$ and $\widehat{\mathbf{s}}'^{(j)}$ are independent.

To track variations in the noise-free input $\mathbf{s}$, windowing techniques can be incorporated in the estimation $Y(\widetilde{\mathbf{s}}^{(j)})$, e.g. a forgetting factor. Here we assume that the variation in $\mathbf{s}$ is slow in comparison with the convergence speed. We remark that a.s. convergence of the proposed algorithm can be establised under mild regularity conditions as in [10].

5. NUMERICAL STUDY

Consider a pulse compression example where the filter output to a Barker coded signal

$$\mathbf{s} = [1, 1, 1, 1, 1, -1, -1, 1, 1, -1, 1, -1, 1]^T$$

is required to fit in an envelope with a mainlobe peak of 0.69 ± 0.075 and sidelobe levels of ±0.025, i.e.

$$\epsilon^{+} = 0.001[\underbrace{25,...,25}_{19},765,\underbrace{25,...,25}_{19}]^T$$
$$\epsilon^{-} = 0.001[\underbrace{-25,...,-25}_{19},615,\underbrace{-25,...,-25}_{19}]^T$$

The training signal is corrupted by additive Gaussian i.i.d. noise with a standard deviation of 0.4, i.e. a signal to noise ratio, SNR = $10\log[(\text{peak signal})^2/(\text{noise variance})]$, of less than 8 dB.

The problem is solved using the 2 proposed adaptive algorithms with the stepsize sequence $\{1/(2l)\}$. Figure 3 plots the squared error between the adaptive filters and the optimal filter against the number of iterations. The filters responses to the noise-free signals after 1000 iterations are shown in Figure 4.

Observe that algorithm 1, which involves (17), (18) converges faster than algorithm 2 which involves (17) and (22) (using 2 test pulses per iteration). However, algorithm 2 has the advantage that the gradient estimate is given by the filter outputs and hence very little computation is required in the update. Both algorithms show significant improvement over those proposed in [7] and [8] in which the primal update does not converge at all.

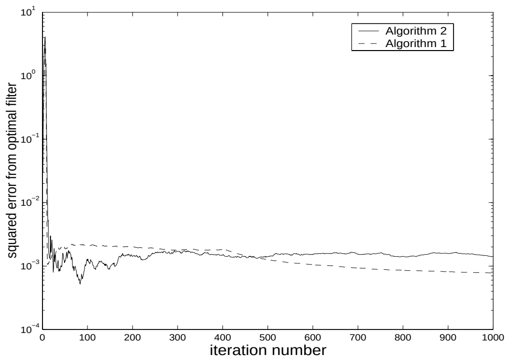

Fig. 3. Convergence.

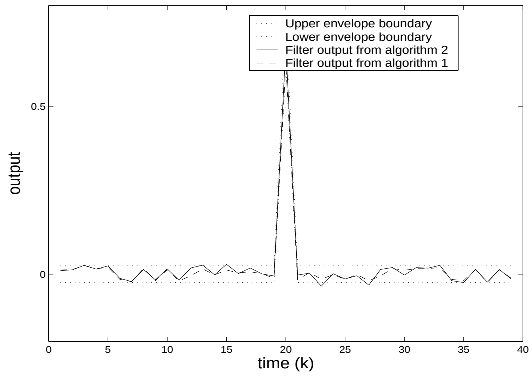

Fig. 4. EC filter outputs.

6. CONCLUSION

An adaptive algorithm for solving the EC filtering problem using noisy training signals has been proposed. The algorithm is based on solving the dual problem using stochastic optimisation techniques. An attractive feature of the proposed algorithm is the surprisingly simple implementation. Our examples have demonstrated good convergence characteristics. On a more theoretical note, convergence follows from standard stochastic approximation results [10].

7. REFERENCES

[1] R. Evans, T. Fortmann, and A. Cantoni, "Envelope constrained filter, Part I, theory and applications", *IEEE Trans. IT*, Vol. 23, pp. 421-434, 1977.

[2] Bell Communications, "DSX-3 Isolated Pulse Template and Equations", *Technical Reference TR-TSY-000499*, pp. 9-17, Issue 2, December 1988.

[3] CCITT, "Physical/Electrical characteristics of Hierarchical Digital Interfaces", *G.703*, Fascicle III, 1984.

[4] J. W. Lechleider, "Pulse envelope for digital systems", *Proc. IEEE. Int. Conf. Comm.*, Atlanta, GA, Vol.2, pp. 703-706, 1990.

[5] B. Vo, A. Cantoni and K. L. Teo, *Filter Design with Time Domain Mask Constraints: Theory and Applications*, Kluwer Academic Publishers, 2001.

[6] Z. Zang, B. Vo, A. Cantoni and K. L. Teo, "Iterative algorithms for constrained recursive filter design via digital Laguerre networks", *IEEE Trans. C&S I*, Vol. 46, No. 11, pp. 1342-1348, 1999.

[7] R. Evans, A. Cantoni, and T. Fortmann, "Envelope constrained filter, Part II, adaptive structures", *IEEE Trans. IT*, Vol. 23, pp. 435-444, 1977.

[8] J. Tsang, K. Teo and A. Cantoni, "Mean square convergence of adaptive envelope-constrained filtering", *IEEE Trans. SP*, Vol. 50, No. 6, pp. 1429-1437, 2002.

[9] B. Vo, A. Cantoni, and K. Teo, "A penalty approach to iterative algorithms for envelope constrained filter design", *IEEE Trans. SP*, Vol. 45, No. 7, pp. 1869-1873, 1997.

[10] G. Pflug, *Optimisation of Stochastic Models*, Kluwer Academic Publishers, 1996.