NOVEL COST FUNCTION ADAPTATION ALGORITHM FOR ECHO CANCELLATION

***Colin F.N. Cowan*[1], *Corneliu Rusu*[2]**

[1]Dept. of Electrical and Electronic Engineering, The Queen's University of Belfast

ASHBY BUILDING, STRANMILLIS ROAD, BELFAST, BT9 5AH, U.K.

[2]Signal Processing Laboratory, Tampere University of Technology

P.O.BOX 553, SF-33101, TAMPERE, FINLAND

ABSTRACT

A new stochastic gradient algorithm for data echo cancellation, based on the cost function adaptation (CFA) is proposed. Qualities of the new adaptation algorithm as compared with that of the least mean square (LMS) and the least mean fourth (LMF) algorithms are demonstrated by means of simulations. Thus it is shown that continuous and automatic, adaptation of the error power yields a more satisfactory result. The cost function adaptation allows an increase in convergence rate and, at the same time, an improvement of residual error. The results were obtained with non-Gaussian binary sequences of data in presence of far-end signals in data echo-cancellers for full duplex digital data transmission over telephone lines.

1. INTRODUCTION

Adaptive echo cancellation in digital communication systems is one of the most intensively investigated fields of application for adaptive signal processing algorithms. So far, the Least-Mean-Square (LMS) adaptive algorithm has been the most commonly used approach due to ease of computation and optimality in the case of Gaussian noise statistics. For a digital echo canceller it is desirable to decrease the adaptation time, during which the transmission of useful data is not possible. Nevertheless, many other adaptive algorithms based upon non-mean-square error cost functions can also be choosen to increase the speed of convergence. Walach and Widrow have investigated the error of the power 4 as an alternative cost function and the Least-Mean-Fourth-Order (LMF) algorithm results [7]. Unfortunately, this algorithm has stability problems and it is relatively sensitive to noise due to the very large gradient terms which result for higher-order representation of errors.

Another alternative approach [6] shows that the use of the adaptive algorithm based on a cost function with the error power r higher than quadratic can be advantageous. Based on this cost function $J_r = E\left[|e_k|^r\right]$, the general form of the stochastic gradient algorithm with non-quadratic exponent can be computed using the simple recursive relation below

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu r \cdot |e_k|^{r-1} \cdot \operatorname{sgn}(e_k)\,, \tag{1}$$

where $\hat{\mathbf{h}}_k$ is the transpose of the vector of estimated filter coefficients at time sample k,

$$\hat{\mathbf{h}}_k = \left[\hat{h}_0, \hat{h}_1, \hat{h}_2, \ldots, \hat{h}_{N-1}\right]^t, \tag{2}$$

$\mathbf{x}_k$ is the transpose of the input observations vector

$$\hat{\mathbf{x}}_k = [x_k, x_{k-1}, x_{k-2}, \ldots, x_{k-N+1}]^t, \tag{3}$$

e_k is the error signal, N is the number of filter coefficients and μ denotes the adaptation step size that controls the stability and the rate of convergence of the algorithm. In the following section we shall derive a new kind of algorithm using (1) in which the power is adjusted using the value of e_k. Some comments are given in Section 3. Performance evaluation and experimental results from computer simulations are also shown in the last section.

2. COST FUNCTION ADAPTATION ALGORITHM

The signal-flow graph representation of data echo canceller is shown in Figure 1. The error signal in terms of the actual echo path output y_k, attenuated far-end signal f_k and synthetic echo signal $\hat{y}_k$ is

$$e_k = y_k + f_k - \hat{y}_k\,. \tag{4}$$

The output of the estimated echo path y_k can be written as

$$\hat{y}_k = \sum_{n=0}^{N-1} \hat{h}_n \cdot x_{k-n} = \hat{\mathbf{h}}_k^t \mathbf{x}_k. \tag{5}$$

We analyse the case of the cost function

$$J_r = E\left[|e_k|^r\right] \tag{6}$$

where the power is a function only of the error modulus:

$$r = r(|e_k|) \tag{7}$$

and we compute $\frac{\partial J_r}{\partial \mathbf{h}}$ with respect to each element of the general impulse response:

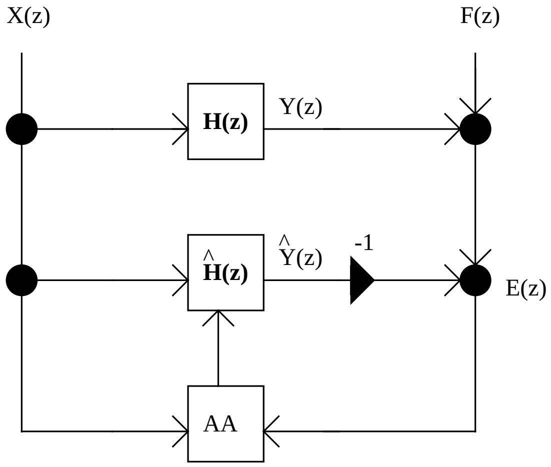

Figure 1. Signal-flow graph representation: AA-adaptive algorithm; H(z)-echo path transfer function; $\hat{H}(z)$ -adaptive filter transfer function.

$$\frac{\partial}{\partial \mathbf{h}}(J_r) = E\left[\frac{\partial}{\partial \mathbf{h}}\{|e_k|^r\}\right] = E\left[\frac{\partial}{\partial e_k}\{|e_k|^r\} \cdot \frac{\partial e_k}{\partial \mathbf{h}}\right] \tag{8}$$

(1)

Using the following formulae

$$\frac{\partial}{\partial e_k}\{|e_k|^r\} = r|e_k|^{r-1} + |e_k|^r \cdot \log|e_k| \cdot \frac{dr}{d|e_k|}; \tag{9}$$

$$\frac{\partial}{\partial \mathbf{h}} e_k = \frac{\partial}{\partial \mathbf{h}}\left(y_k + f_k - \mathbf{h}^t \mathbf{x}_k\right) = \mathbf{x}_k; \tag{10}$$

we have

$$\frac{\partial}{\partial \mathbf{h}}(J_r) = -E\left[|e_k|^{r-1}\left(r + |e_k| \cdot \log|e_k| \cdot \frac{dr}{d|e_k|}\right)\operatorname{sgn} e_k \mathbf{x}_k\right] \tag{11}$$

We follow the cost function adaptation algorithm when the power is choosen in such a manner that respects the relation:

$$r(|e_k|) + |e_k| \cdot \log|e_k| \cdot \frac{dr}{d|e_k|} = 0. \tag{12}$$

If we denote by $r_k = r(|e_k|)$ and $|e_k|_{dB}$ is the error modulus, measured in dBs, it results that for CFA selected the gradient is always zero, and the product $r_k|e_k|_{dB}$ is constant.

We can admit also that during one iteration, the power is constant. This is always true in practical problems. In this way CFA may be seen as a piecewise non-quadratic algorithm, in which the weights are computed with

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu \cdot r_k \cdot |e_k|^{r_k - 1} \cdot \operatorname{sgn}(e_k) \tag{13}$$

and the power is updated via

$$r_k = \frac{r_0|e_0|_{dB}}{|e_k|_{dB}}. \tag{14}$$

3. COMMENTS AND IMPLEMENTATION

First, the error must decrease and according with (14), we recommend starting with a large power r_0 and stopping with a small power r_∞. Anyway we must avoid large gradients which leads to divergence. In order to exploit both the noisy stability of the LMS algorithm and the initially faster convergence of LMF, we usually choose $r_0 = 4$ and $r_\infty = 2$.

However as in almost all adaptive systems, the adjustment of r depends on finite-time average characteristics rather than on instantaneous values of the error. A direct implementation of the previous relation (14) will lead to divergence.

In [2] the first decreasing smooth power-error algorithm was proposed:

$$r_k = -\frac{60}{20\log\frac{\sum_{j=k-9}^{k} e_j^2}{10}}. \tag{15}$$

Since the output mean-square error is actually fluctuating, the first 100 iterations used LMF and then the power is adjusted with (15). The restriction imposed reduces the general improvements of CFA algorithm. To avoid this condition, the power must be bounded during the adaptation time, otherwise the algortihm could diverge.

In an application such as data echo cancellation, the adjustment is usually done by calculating the mean value of tap-error vector. A normalised form of tap-error vector norm is used in the data echo cancellers

$$|p_k| = \frac{|\hat{\mathbf{h}}_k - \mathbf{h}_k|}{|\mathbf{h}_k|} \tag{16}$$

and, it is selected rather than the output mean-square error due to the high level of the far-end signal power.

Another CFA algorithm was proposed in [2]: the decreasing staircase power-error algorithm. This algorithm has the advantage of less computational effort by using only integer or square root powers in its implementation. The adaptive filter coeficients are updated with (13) and the power is computed with:

$$r_k = \begin{cases} 4, & |p_k| \le 20 \\ 3.5, & -25 \le |p_k| < -20 \\ 3, & -30 \le |p_k| < -25 \\ 2.5, & -36 \le |p_k| < -30 \\ 2, & |p_k| < -36 \end{cases} \tag{17}$$

We shall show in the following another new method to design better the CFA power-error relation.

Assuming that $\{x_k\}$ and $\{f_k\}$ are independent bipolar sequences, from the sets $\{1, -1\}$ and, $\{f, -f\}$ respectively, we have

$$e_k = y_k + f_k - \hat{y}_k = f_k + \sum_{n=0}^{N-1} \left(\hat{h}_n - h_n\right) \cdot x_{k-n} \tag{18}$$

and then

$$e_k^2 = f_k^2 + \sum_{i,j}^{N-1} \left(\hat{h}_i - h_i\right) \cdot \left(\hat{h}_j - h_j\right) \cdot x_{k-i} \cdot x_{k-j} + \\ +2 \cdot f_k \cdot \sum_{n=0}^{N-1} x_{k-n} \cdot \left(\hat{h}_n - h_n\right) \tag{19}$$

For $i \neq j$ we obtain

$$f_k^2 = f^2, x_k^2 = 1, E[x_{k-i}, x_{k-j}] = E[f_k, x_{k-n}] = 0 \tag{20}$$

and assuming that the channel is slowly varying, it results

$$E\left[e_k^2\right] = f^2 + E\left[\sum_{n=0}^{N-1} \left(\hat{h}_n - h_n\right)^2\right] = \\ = f^2 + E\left[\left|\mathbf{h}_k - \hat{\mathbf{h}}_k\right|^2\right] \tag{21}$$

We can approximate

$$\left|\hat{e}_k\right|_{dB} = 10 \cdot \log_{10}\left(E\left[e_k^2\right]\right) = \\ = 10 \cdot \log_{10}\left(f^2 + E\left[\left|\mathbf{h}_k - \hat{\mathbf{h}}_k\right|^2 / \left|\mathbf{h}_k\right|^2\right] E\left[\left|\mathbf{h}_k\right|^2\right]\right) \tag{22}$$

Inserting the previous formula in (14), it results that the computation of new power could be done with

$$r_k = r_0 \cdot \frac{\log\left(f^2 + E\left[\left|\mathbf{h}_0\right|^2\right] \cdot 10^{\left|p_0\right|_{dB}/20}\right)}{\log\left(f^2 + E\left[\left|\mathbf{h}_k\right|^2\right] \cdot 10^{\left|p_k\right|_{dB}/20}\right)} \tag{23}$$

where $\left|p_k\right|_{dB}$ is the normalised tap-error vector norm, measured in dBs. This is the main formula for the CFA power-error relation.

It is obviously that the necessary condition for r_k to be a decreasing function of $\left|p_k\right|_{dB}$ is

$$f^2 + E\left[\left|\mathbf{h}_0\right|^2\right] \cdot 10^{\left|p_0\right|_{dB}/20} > 1 \tag{24}$$

If the statistical characteristics of the channel are not available, a choice for $E\left[\left|\mathbf{h}_k\right|^2\right]$ could be $\left|\hat{\mathbf{h}}_k\right|^2$.

The convergence of CFA results if we assume that the power r_k is bounded:

$$1 < r_\infty < r_k \leq r_0, \forall k \in \dot{N} \tag{25}$$

and the proof is based on similar analysis with [6]. We can conclude that a large start power means stronger conditions for the convergence of CFA algorithm.

4. MODELLING AND SIMULATIONS

The simulator is a direct system modelling type where the echo canceller is trying to model the echo path. The data echo canceller is a 32 tap linear time varying FIR adaptive filter whose coefficients are updated regularly by the adaptation algorithm.

The echo-path model uses a single pole single zero digital filter. The transfer function of the echo path is

$$H(z) = \frac{z}{z-a} = \sum_{k=0}^{\infty} a^k \cdot z^{-k} \tag{26}$$

where $0 < a < 1$. We choose the feedback coefficient a of the echo path filter is in such a way that the power level of the impulse response will be attenuated by 60dB, at 32nd sample: $a = 0.80025$. Thereafter the series is truncated.

The near-end signal sequence x_k is modelled by a random bipolar sequence from the set $\{1, -1\}$. The attenuated far-end signal sequence f_k is also modelled by an independent random bipolar sequence from the set $\{f, -f\}$. The level of the attenuated far-end signal power is -15dB.

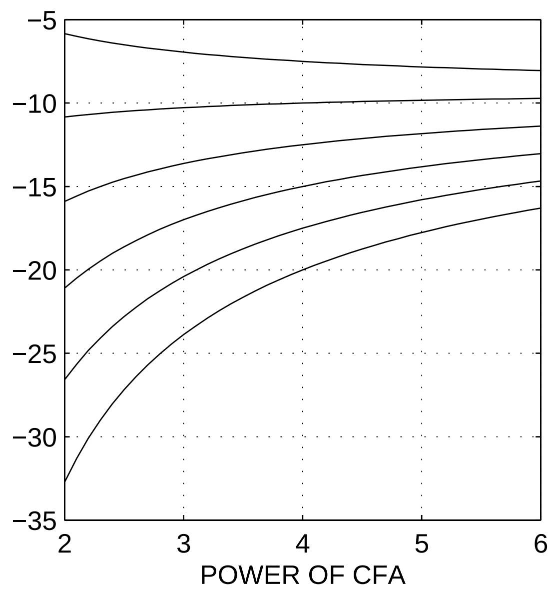

Figure 2. Initial conditions of CFA: normalised tap-error vector norm $|p_0|$ in dBs and the power.

Figure 2 shows the relation (23) in case of different $\left|p_0\right|_{dB}$ and r_0. It is clear now that an earlier start of CFA can affect the convergence.

We start with LMF and follow LMF since the normalised tap-error norm is less than the start value $\left|p_0\right|_{dB}$ corresponding to the power 4 ($\left|p_0\right|_{dB} < -10$, otherwise the numerator is negative). Then we continue with CFA updating the power with (23) and using $\left|\hat{\mathbf{h}}_k\right|^2$ instead of $E\left[\left|\mathbf{h}_k\right|^2\right]$. In this way we avoid the critical points of (12).

When $r_k = 2$ the CFA could be stopped or not. There are not major differences between the two cases: the steady-state is postponed and the speed of adaptation is improved (a little bit).

Applying (23) directly gives us the first final CFA algorithm: the

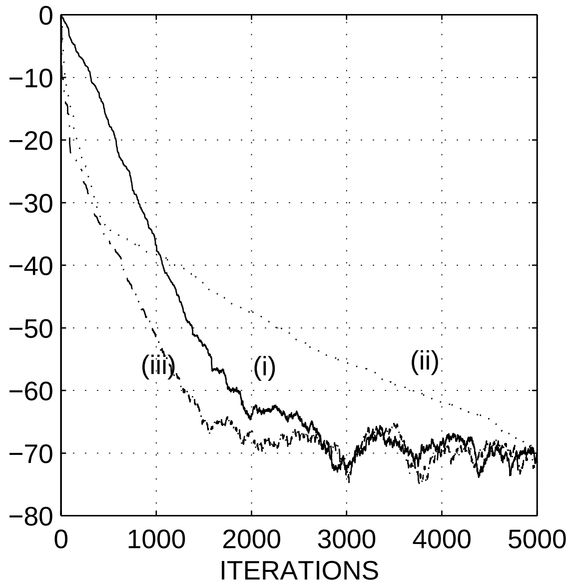

Figure 3. Normalised tap-error vector norm, dBs versus the number of iterations in the case of LMS (i), LMF (ii) and CFA(iii) with no average.

smooth decreasing error-power algorithm.The second final CFA algorithm implemented is the decreasing staircase error power algorithm and it has the advantage of less computational effort by using only integer or square root powers. For this second algorithm, we round the term from the right hand side of (23). Note that the results between the smooth and the staircase power error algorithms do not differ by much.

The plots shown in Fig. 3-4 were obtained for the smooth decreasing error-power algorithm with the parameters given by

$$r_0 = 4, r_\infty = 2, |e_0|_{dB} = -20dB \qquad (27)$$

and the step size μ was chosen as 0.001 in order to assure convergence of the LMF algorithm. We replace the statistical expectation operator E with an ensemble average of 20. The convergence level was set 20dB below the far-end signal level.

The plots indicate the advantage of the new implementation. First we have a fast convergence time, initially the same as the LMF and twice as fast as in the LMS case. Secondly, the results show an improvement in residual noise levels, below those achievable by the LMF alone.

5. CONCLUSION

A new algorithm for data echo cancellation based on cost function adaptation has been described. The power of the error function is adjusted according to an estimate of the error. The results show that this algorithm is better than LMS and LMF under the same conditions: providing a fast convergence time (smaller than LMF and almost twice as fast as in the LMS case) and, an improvement in residual noise levels.

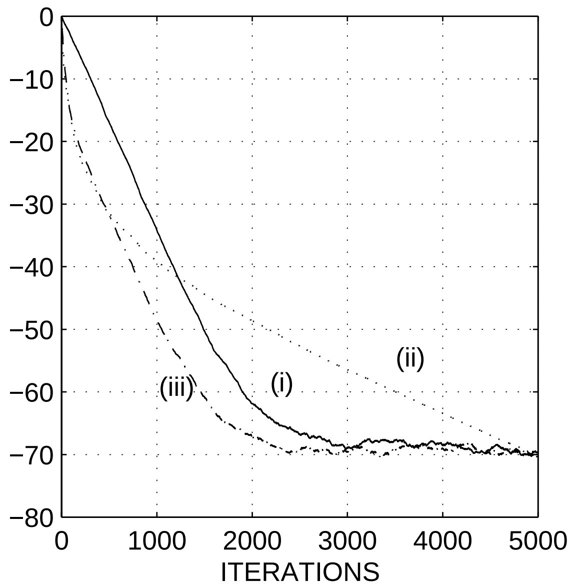

Figure 4. Normalised tap-error vector norm, dBs versus the number of iterations in the case of LMS (i), LMF (ii) and CFA(iii) with 20 averages.

6. REFERENCES

[1] Cowan C.F.N., and Grant P.M. *Adaptive Filters.* Prentice Hall, Englewood Cliffs, NJ, 1985.

[2] Cowan C.F.N. and Rusu C. “Adaptive echo cancellation using cost function adaptation". *Conference Digest of Fourth IMA International Conference on Mathematics in Signal Processing*, Warwick, UK, December, 1996.

[3] Haykin S. *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, NJ, 1991.

[4] Honig M.L. “Echo Cancellation of Voiceband Data Signals Using Recursive Least Squares and Stochastic Gradient Algorithms”. *IEEE Trans. Commun.* vol. COM-33, No.1, pp.65-73, Jan.1985.

[5] Messerschmitt D.G. “Echo cancellation in Speech and Data Transmission” *IEEE J. Selected Areas in Comm.*, vol. SAC-2, No.2. pp.283-297, Mar.1982.

[6] Shah S.A. and Cowan C.F.N. "Modified stochastic gradient algorithm using nonquadratic cost functions for data echo cancellation," *IEE.Proc.-Vis.Image Signal Process.*, vol.142, No. 3, pp.187-191, June 1995.

[7] Walach E. and WidrowB.“The least mean fourth (LMF) adaptive algorithm," *IEEE Trans.Inform.Theory*, vol.30, pp. 275-283, March 1984.