Novel Stochastic Gradient Adaptive Algorithm with Variable Length

Corneliu Rusu* and Colin F. N. Cowan†

Abstract — **The goal of this paper is to present a novel variable length LMS (Least Mean Square) algorithm, in which the length of the adaptive filter is always a power of two and it is modified using an error estimate. Unlike former variable length stochastic gradient adaptive techniques, the proposed algorithm works in non-stationary situations. The implementation of the adaptive filter is described and results of computer simulations are provided.**

1 Introduction

The design of traditional adaptive systems in signal processing applications has been strongly influenced by the need to provide an efficient implementation [1]. The result is often a non-optimal, but a simple design. From the stochastic gradient algorithms family, the most common one is the least mean square (LMS). The main reason for its popularity consists in its well known and used transversal structure, despite it is slow convergent compared to other more complicated algorithms. Much effort has been done to improve LMS performances and at the same time to keep the level of attractiveness in implementation, its robustness and reduced computational effort.

In the following we present a new variable length LMS algorithm. The length of the adaptive filter is chosen a power of two and it is changed by using an error estimate. Unlike former variable length stochastic gradient adaptive techniques, the proposed algorithm performs in abrupt change situations. This paper is organized as follows. In Section 2 we recall previous LMS approaches with variable length. The implementation of the adaptive filter is described in Section 3 where results of computer simulations are also presented.

2 Variable Length Algorithms

In some situations, there is a need to increase the filter length during adaptation. For example, the filter length of the LMS algorithm determines the equalizer performance, but it is difficult to decide a proper length because the filter length is related to channel characteristics. Thus the variable length adaptive algorithm in which the filter length varies is desirable. In this case the architecture needed to implement the adaptive filter is no longer fixed and therefore creates serious problems with regard to high speed hardware implementation [1].

We have to point out that by changing the length of the adaptive filter we modify the implementation of the adaptive filter by acting on the model of the system to be identified, i.e. on the adaptive filter structure. On the other hand, in the case of the variable step-size or variable error exponent methods our point of interest is only the numerical update routine of the adaptive filter coefficients, i.e. on the adaptive filter algorithm.

2.1 Previous Variable Length LMS Algorithms

In [2] Pritzker and Feuer have proposed the variable length stochastic gradient algorithm. In this algorithm the MSE (mean-square error) value is estimated at every iteration time, and if the error is equal to the minimum MSE for the specific filter length, the filter length is increased. Unfortunately their algorithm needs approximations of the minimal MSE of each filter length, and its performance variation largely depends on the MSE estimation.

Another variable length LMS algorithm has been proposed in [3] and it changes the filter length using the time constant concept. First several filter lengths are predermined. Then certain step-sizes and time constants are calculated in advance. A relationship is used which states that the product between the step-size and the adaptive filter length is two tenth of the inverse of the first element of the input correlation matrix. This variable length LMS algorithm compares the time-constant to the current iteration time and if they are equal, the filter length is increased.

We have found two possible complains in relation with these algorithms:

1. The parameters used in algorithm description depend in a certain manner on the adaptive process addressed;

*Tampere University of Technology, Tampere International Center for Signal Processing, Finland. Also with Technical University of Cluj-Napoca, Dept. of Electronics and Telecommunications, Romania. E-mail: corneliu.rusu@ieee.org.

†Dept. of Electrical and Electronics Engineering, The Queen's University of Belfast, Stranmillis Road, Belfast, BT9 5AH, U.K.

2. The update formulas for filter length provide only the increase of the filter length.

Such solution have apparently obvious advantages for stationary data. For non-stationary data such formulas might be inappropriate because of the inability to respond to changes in the optimum solution. Consequently the algorithms treat only the stationary case and only such type of simulations are actually presented in the corresponding works. Beside, the incrementation of the adaptive filter length is always with only one unit. According to [1] it has been demonstrated that it is faster to address cells in powers of two rather than groups in other sizes. Moreover impulse response of a FIR channel might be sparse. In these cases the incrementation by one could not be very advantageous and thus it will not improve all the time the performances of the adaptive filter.

3 The proposed variable length LMS

3.1 Framework

The setup used in this work is the data echo canceller. The adaptive FIR filter is trying to make a copy $\hat{y}(k)$, of the echo-path output $y(k)$, using the signal $x(k)$ as an input, based upon a measurement of the signal that remains after subtracting $\hat{y}(k)$ from the received signal $y(k) + f(k)$. The resulting error is $e(k) = y(k) + f(k) - \hat{y}(k)$. At the kth time sample, the adaptive filter has the impulse response given by

$$\hat{\boldsymbol{h}}(k) = [\hat{h}(0), \hat{h}(1), \ldots, \hat{h}(N_k - 1)]^t,$$

where N_k is the number of filter coefficients of the adaptive filter with variable length. The output of this filter can be written as

$$\hat{y}(k) = \sum_{n=0}^{N_k-1} \hat{h}(n)x(k-n).$$

3.2 Design Aspects

The analysis presented in [2] might provide several guiding lines in the design of any variable length adaptive algorithm:

Remark 1 . *The larger is the adaptive filter length, the slower convergence of the adaptive filter to Wiener filter is.*

Example 1. *Different adaptive filter length, but same step-size*

We consider the case of a channel having the model a single zero single pol ($p = 0.80025$) digital filter. Unlike other works, in this case the impulse response is not truncated, thus any special filter length has not been privileged. Moreover, the previous works [2, 3] can not be applied in this case. For the same reason, the performance measure selected is the mean-square error, though the normalized tap-error vector norm is more convenient to handle in the case of data echo cancellation. The output of the echo-path filter can be computed with the formula:

$$y(k) = py(k-1) + x(k),\ k \geq 1; \quad y(0) = 0.$$

The input data is binary ± 1 and the attenuated far-end signal is also binary, but subject of an attenuation of 20 dB. The learning curves have been obtained by doing 100 averages of the square error.

The results for the same step-size ($\mu = 0.001$), but for different length ($N = 4, 8, 16$) of the adaptive filter are presented in Fig. 1. We have performed the same simulations for bigger length ($N = 32, 64, 128$) and we conclude that LMS behaves almost the same as for $N = 32$ from convergence speed point of view. This performance is done essentially by the step-size value. Following this last note, we can see why the variable length LMS algorithms require variable step-size in implementation.

The level of gradient noise modifies when N increases in the following manner (Table 1):

1. First it decreases, because the IIR echo-path system is better identified when the number of terms is increasing;

2. Then it decreases as the misadjustment is depending on the length of the adaptive filter and echo-path model.

It means that there should be an optimum adaptive filter length and we are not obliged to increase indefinitely the adaptive filter length. Clearly this is suitable for hardware implementation.

Remark 2 . *Despite the fact that the step-size is increased (typically the increase is by an order of magnitude, beyond the stable range of the fixed length LMS) during early stages of adaptation, the level of gradient noise remains constant.*

Example 2. *Different adaptive filter length and related step-size*

N	4	8	16	32	64	128	256
$\mathcal{M}$	47	7.9	0.24	0.031	0.067	0.144	0.346

Table 1: Misadjustment $\mathcal{M}$ for different N and same step-size μ.

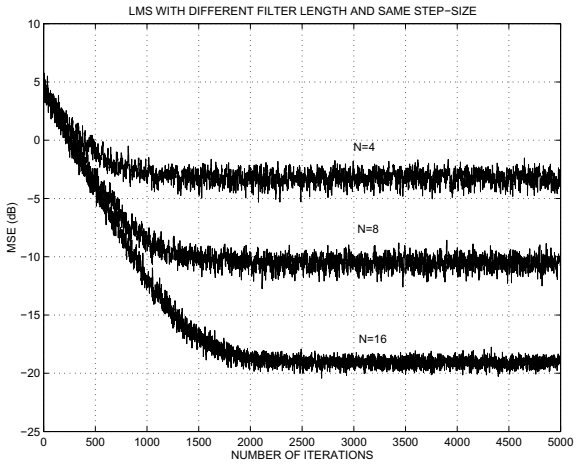


Figure 1: Learning curves of LMS with different adaptive filter lengths ($N = 4, 8, 16$) and same step-size $\mu = 0.001$.

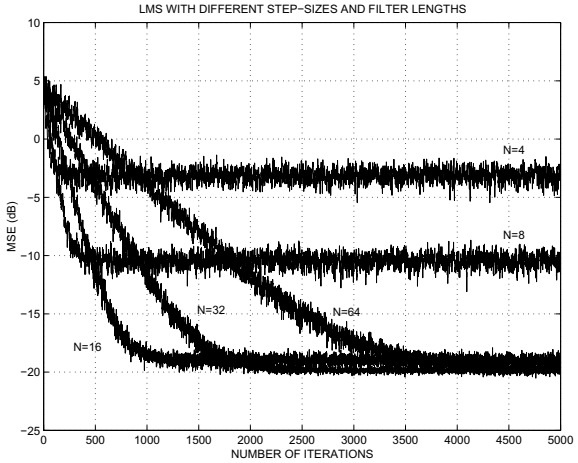


Figure 2: Learning curves of LMS with different adaptive filter lengths $N = 2^p$ and different step-sizes $\mu = 0.001 \cdot 2^{5-p} (p = 2, 3, 4, 5, 6)$.

We reconsider Example 1 with different step-size μ and different length of the adaptive filter ($N = 4, 8, 16, 32, 64, 128$). They are selected such that $\mu N = 0.032$. The results are presented in Figure 2. We conclude that the Remark 2 is valid only for a range of filter length (Table 2). If the step-size is not very small, the convergence speed is quite the same. Also LMS with $N = 32$ reaches the convergence level of -20 dB and it has the smallest length.

3.3 The Variable Length LMS Algorithm

The proposed variable length LMS algorithm consist of three different parts:

- adaptive filter length update;
- step-size update;
- adaptive filter coefficients update.

At every iteration we compute the error estimator using the recursive digital filter having the absolute error as input. We sample its output at W samples. We compare the sampled value of the recursive digital filter output with some given thresholds and the new adaptive filter length results. In the following we describe in details every step of the proposed variable length LMS algorithm.

1. The first part of the variable length LMS algorithm is filter length update

   - Compute the error estimator $\hat{e}(k)$

     $$\hat{e}(k) = \frac{L}{L+1}\hat{e}(k-1) + \frac{1}{L+1}|e(k)|, \quad k \geq 1;$$

   - At every W iterations ($k \equiv 0 \mod W$):
     - If $\hat{e}(k) > 2/N_k$, then $N_{k+1} = N_k/2$;
     - If $\hat{e}(k) < 4/N_k$, then $N_{k+1} = 2N_k$;
     - Otherwise $N_{k+1} = N_k$.

2. The second part of the algorithm is the step-size update. This is done by imposing the product $\mu_k N_k$ to be always constant.

3. The third part of the algorithm is the well-known LMS equation:

   $$\hat{\boldsymbol{h}}(k+1) = \hat{\boldsymbol{h}}(k) + 2\mu_k e_k [x(0), x(1), \ldots, x(N_{k+1}-1)]^t.$$

   In this relationship all the vectors should have dimension N_{k+1}. This means that if $N_k \neq N_{k+1}$, then they modify their structure in the following way:

   - If $N_k > N_{k+1}$, the weights of the vectors with indexes $N_k + 1$, $N_k + 2$, $\cdots$, N_{k+1} disappear;
   - If $N_k < N_{k+1}$, the added weights of the vectors are initialized to zero.

N	4	8	16	32	64
$\mathcal{M}$	48.31	8.123	0.2592	0.0316	0.0848

Table 2: Misadjustment $\mathcal{M}$ for different N and μ (Example 2).

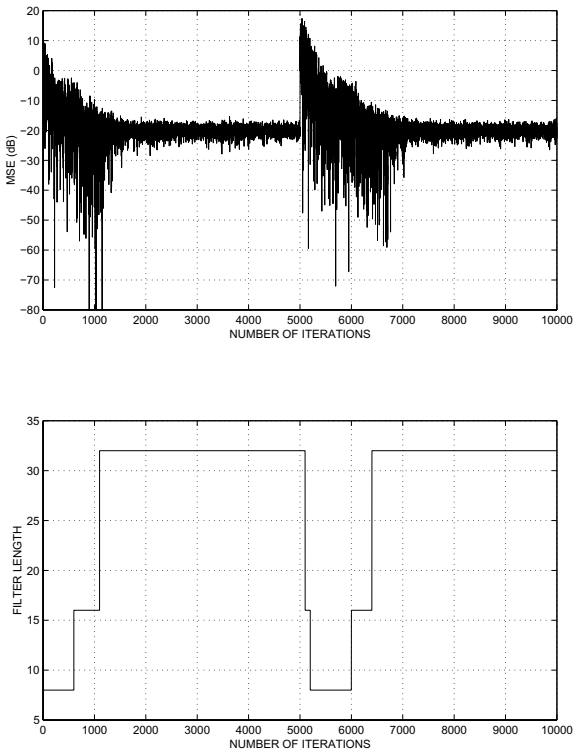


Figure 3: MSE and adaptive filter length of the proposed variable length LMS algorithm for an abrupt change situation.

3.4 Simulations

The results presented in this Section have been obtained using the following initializations: $N_0 = 8$, $\hat{e}(0) = 0.1$, $\mu_0 = 0.004$, $W = 100$ and $L = 100$.

Example 3. *The proposed variable length LMS algorithm works in abrupt change condition*

We reconsider the framework from Example 1, where the channel have had the model a single zero single pol digital filter. At the 5001 iteration the echo-path has a sign change. Thus the output of the echo-path filter can be computed with the formula:

$$y(k) = py(k-1) - x(k), \ k \geq 5001.$$

Fig. 3 shows clearly that the proposed variable length LMS algorithm is applicable to non-stationary situations. Moreover, we can see how the length of the filter modifies during the adaptation.

Example 4. *A short comparison with LMS*

Fig. 4 compares the proposed variable length LMS algorithm with the best performer LMS algorithm

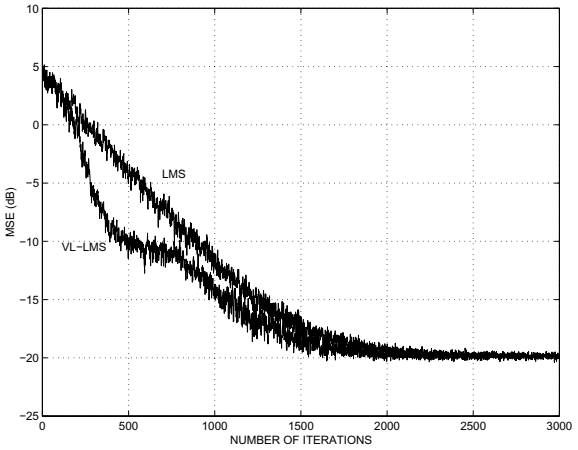


Figure 4: Learning curves of LMS and of the proposed variable length LMS (VL-LMS) algorithm (stationary).

shown in Examples 1 and 2 ($N = 32$), using the same framework and for the same conditions and performance measures. We can conclude that variable length LMS can achieve a better performance than LMS in stationary condition.

4 Conclusions

The novel variable length LMS algorithm we have presented in this paper behaves very well in both stationary and non-stationary conditions, and it offers an attractive implementation compared to the LMS. Despite the fact that the structure is not fixed, finally the computational load is reduced. This is a result of implementing LMS with a small number of coefficients at the early stages of adaptation.

References

[1] C. F. N. Cowan, R. F. Woods, J. P. Heron, P. Power, and F. J. Sweeney, "Advances in adaptive signal processing: totally adaptive systems," in *Proc. IFAC'98 Worshop on Adaptive Systems in Control and Signal Processing*, Glasgow, Scotland, U.K., Aug. 1998, pp. 185–194.

[2] Z. Pritzker and A. Feuer, "Variable length stochastic gradient algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 997–1001, Apr. 1991.

[3] Y. K. Won, R. H. Park, J. H. Park, and B.-U. Lee, "Variable LMS algorithms using the time constant concept," *IEEE Transactions on Consumer Electronics*, vol. 40, no. 3, pp. 655–661, Aug. 1994.