A NOVEL VARIABLE TAP-LENGTH ALGORITHM FOR LINEAR ADAPTIVE FILTERS

Y. Gong and C. F. N. Cowan

Department of Electrical & Electronics Engineering
The Queen's University of Belfast
Stranmillis Road, Belfast BT9 5AH
United Kingdom

ABSTRACT

The tap-length is an important structural parameter of the linear FIR adaptive filter. Although the optimum tap-length that balances the performance and complexity varies with scenarios, most current adaptive filters fix the tap-length at some compromise value, making them inefficient to implement especially in time varying scenarios. In this paper, we propose a novel gradient search based variable tap-length algorithm using the concept of the pseudo fractional tap-length, and show that the new algorithm can converge to the optimum tap-length in the mean. Results of computer simulations are also provided to verify the analysis in this paper.

1. INTRODUCTION

In linear adaptive filters, the tap-length, or the number of the taps, is an important parameter that significantly influences the performance of the adaptive filter. On the one hand, the tap-length needs to be long enough to ensure good performance since the minimum mean squared error (MMSE) is a monotonic non-increasing function of the tap-length. On the other hand, the tap-length can not be too long as it otherwise increases the adaptation noise. Moreover, even without the adaption noise, it is still not suitable to have an unnecessarily long filter in view of complexity, since the improvement of the MMSE performance due to the tap-length increase always becomes trivial when the tap-length is long enough. Therefore, there exists an *optimum* tap-length that balances the conflicting requirements of performance and complexity. In most adaptive filters, however, the tap-length is usually fixed at some compromise value determined by observation of typical scenarios, implying that often the filter is too long and sometimes inadequate for the severity of conditions.

There are several *variable tap-length* algorithms available in the literature. Most of them (e.g [1, 2, 3]), however, target more at improving the convergence rate of the LMS algorithm than at searching for the optimum tap-length. A more relevant work is a recent paper by Riera-Palou *et al.* [4], in which the original TDL structure of an FIR equalizer is partitioned into several segments and the tap-length can be adjusted by one segment being added to, or removed from, the filter according to the difference of the output error levels from the last two segments. The idea is that if the tap-length is long enough, the last two segments have similar levels of the output error. However, because the difference of the instantaneous output error levels from the last two segments may not always reflect the difference of the MMSE with equivalent tap-lengths, this method does not always converge to the optimum tap-length. Moreover, the key parameters such as the number of the segments must be carefully chosen for different applications, making the algorithm inflexible in application.

In this paper, we will propose a novel variable tap-length algorithm using a gradient search method with instantaneous inputs. The proposed algorithm is based on the observation that though the MSE cost function with respect to the tap-length is difficult, if not impossible, to obtain, the relationship between the MSE and tap-length can be revealed in an ad-hoc manner since the tap-length is only a one dimensional parameter. Moreover, although the tap-length must be an integer, we can apply the concept of the pseudo *fractional* tap-length to make instantaneous length adaption possible, where the true tap-length is the integer part of the fractional tap-length.

This research was sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/S00217/01.

2. OPTIMUM TAP-LENGTH

This section gives definitions of the optimum and suboptimum tap-lengths for an FIR adaptive filter.

It is well known, e.g. [5], that the *converged MSE* to which the LMS algorithm converges is given by:

$$\xi_N(\infty) = \xi_{N,\ \mathrm{opt}} + \xi_{N,\ \mathrm{excess}}, \tag{1}$$

where $\xi_{N,\ \mathrm{opt}}$ is the MMSE, $\xi_{N,\ \mathrm{excess}}$ is the excess MSE which is defined as:

$$\xi_{N,\ \text{excess}} = \mathcal{M}_N \cdot \xi_{N,\ \text{opt}}, \tag{2}$$

where $\mathcal{M}_N$ is the misadjustment, and the subscript N denotes the parameters relating to the filter with tap-length N.

For better convergence behavior, many applications use the normalized-LMS which has been shown to have a constant level of misadjustment under different scenarios [6, 7]. Then its $\xi_N(\infty)$ and $\xi_{N,\ \text{opt}}$ only differ by a factor of constant level. Further considering that $\xi_{N,\ \text{opt}}$ is a monotonic non-increasing function of N, we have that as N increases, $\xi_N(\infty)$ keeps decreasing until either it starts to increase or the decrease becomes negligible no matter how large N increases. With this observation, we have the following definitions.

Definition 1 *If defining $\Delta_N = \xi_N(\infty) - \xi_{N+1}(\infty)$ as the decrease of the converged MSE when the tap-length is increased from N to $N+1$, then the "optimum tap-length" is the smallest N_o that satisfies:*

$$\Delta_N \leqslant \mathcal{E} \qquad \text{for all } \ N \geqslant N_o, \tag{3}$$

where $\mathcal{E}$ is a predetermined value according to the system requirements, N and N_o are positive integers. ■

Δ_N is basically the gradient of the function of the converged MSE with respect to the tap-length at N and it may be negative due to the adaptation noise. $\mathcal{E}$ is usually a small positive number. In many scenarios we may have a *suboptimum* tap-length which is defined in Definition 2.

Definition 2 *If a positive integer M satisfies $M < N_o$ and:*

$$\Delta_M \leqslant \mathcal{E}, \tag{4}$$

then the "suboptimum tap-length" is M, where N_o and $\mathcal{E}$ are defined in Definition 1. If a group of concatenated integers $M, M+1, \cdots, M+L-1$, but neither $M-1$ nor $M+L$, are all suboptimum tap-lengths, then the set of $M, \cdots, M+L-1$ is called one "suboptimum tap-length set", and $L+1$ is called the "length" of the suboptimum tap-length set.

During the next section, we will derive a novel variable tap-length algorithm that can find N_o adaptively.

3. TAP-LENGTH ADAPTATION

In applications, $\xi_N(\infty)$ is usually not available and can be obtained by the exponential average as:

$$\bar{\xi}(i) = \lambda\bar{\xi}(i-1) + (1-\lambda)e^2(i), \tag{5}$$

where λ is a forgetting factor which is set close to one.

Defining n_f as the pseudo tap-length which can take fractional values, we have the following adaptation rule:

$$n_f(i+1) = (n_f(i) - \alpha) + \beta \cdot \left[\bar{\xi}_{\text{p}} - \bar{\xi}(i)\right], \tag{6}$$

where the true tap-length $N(i) = \lfloor n_f(i) \rfloor$, $\lfloor . \rfloor$ truncates the embraced value to the nearest integer, $\bar{\xi}_{\text{p}}$ is the estimated converged MSE for the previous tap-length $N(i-1)$, $\bar{\xi}(i)$ is obtained by (5), β is the step-size for $n_f(i)$ adaptation and α is called a *leaky factor* which is used to prevent $n_f(i)$ from increasing to an undesirably large value.

Initially we set $\alpha \ll \beta$, $\bar{\xi}_{\text{p}} = \mathcal{P}$ and $N(0) = n_f(0) = N_s$, where $\mathcal{P} > \xi_N(\infty)$ for all N, and $N_s < N_o$ (e.g. $N_s = 1$). Then $n_f(i)$ starts to increase from N_s. At the so-called *changing time* that $|n_f(i) - N(i)| \geqslant 1$, we let $\bar{\xi}_{\text{p}} = \bar{\xi}(i)$ and append one zero-tap to the current filter. At non-changing time, both $\bar{\xi}_{\text{p}}$ and $N(i)$ remain unchanged.

Starting at one changing time, the filter converges towards the MMSE corresponding to the new tap-length. If β is small enough and $\alpha \ll \beta$, we can have $\text{E}[\bar{\xi}(i)] = \xi_{\lfloor n_f(i) \rfloor}(\infty)$ before the next changing time. Then taking expectations on both sides of (6) gives:

$$\text{E}[n_f(i+1)] = (\text{E}[n_f(i)] - \alpha) + \beta \cdot \left[\xi_{\lfloor n_f(i) \rfloor - 1}(\infty) - \xi_{\lfloor n_f(i) \rfloor}(\infty)\right], \tag{7}$$

where i corresponds to the converged periods of the adaptive filter. It is clear from (7) that $\text{E}[n_f(i)]$ keeps increasing until (note $N(i) = \lfloor n_f(i) \rfloor$):

$$\xi_{N(i)-1}(\infty) - \xi_{N(i)}(\infty) \leqslant \frac{\alpha}{\beta}. \tag{8}$$

Obviously in (6), the tap-length can only be increased. Similarly we can also construct the recursion to decrease the tap-length which is given by:

$$n_f(i+1) = (n_f(i) - \alpha) - \beta \cdot \left[\bar{\xi}_{\text{p}} - \bar{\xi}(i)\right], \tag{9}$$

where the true tap-length $N(i) = \lfloor n_f(i) \rfloor + 1$. Initially we set $\alpha \ll \beta$, $n_f(0) > N_o$, and $\bar{\xi}_{\text{p}} = \mathcal{P}$ which is same as that for (6). Then $n_f(i)$ starts to decrease from N_b. The changing time is also defined as the time when $|n_f(i) - N(i)| > 1$, at which we let $\bar{\xi}_{\text{p}} = \bar{\xi}(i)$ and remove the last tap from the adaptive filter. Similar to the analysis for (6), we have that if β is small enough and $\alpha \leqslant \beta$, $\text{E}[n_f(i)]$ keeps decreasing until:

$$\xi_{N(i)-1}(\infty) - \xi_{N(i)}(\infty) > \frac{\alpha}{\beta}. \tag{10}$$

It is clear from (8) and (10) that if there are no suboptimum tap-lengths, both (6) and (9) can converge to the optimum tap-length in the mean, where $\mathcal{E}$ in Definition 1 is set to be α/β.

Some systems may have suboptimum tap lengths. Fortunately, both (6) and (9) are adapted based on the instantaneous values of $\bar{\xi}(i)$, the variance of which can be regarded as random disturbance to the search procedure. If such disturbance is much larger than the variation of the sub-optimum tap-lengthes, the search algorithm can escape from the sub-optima. Or we may increase the adjusting step-size of the tap-length, i.e. adjust N by the value of K ($K > 1$) at every changing time. Then the tap-length adaptation converges to a value within $[N_0, N_o + K - 1]$ rather than the exact value of N_o. This still satisfies the tap-length requirement for most systems since they usually do not require a highly accurate value of the optimum tap-length.

Both (6) and (9) can only search for the tap-length in one direction. Therefore unless β is very small which however implies slow convergence rate, the search may fail due to the inaccurate estimate of $\xi_N(\infty)$. However, noting that (6) and (9) differ only by a sign factor, we can merge them into one recursion. Then we have the basic adaptive tap-length algorithm as shown below.

For every i=1,2,3, $\cdots$
  $\bar{\xi}(i) = \lambda\bar{\xi}(i-1) + (1-\lambda)e^2(i)$
  $n_f(i+1) = (n_f(i) - \alpha) + \beta\,\gamma\left[\bar{\xi}_{\mathrm{p}} - \bar{\xi}(i)\right]$ (11)
  if $|n_f(i) - N(i)| \geqslant K$
    $\bar{\xi}_{\mathrm{p}} = \bar{\xi}(i)$, $N(i) = \langle n_f(i)\rangle$
    $\Delta_N = N(i) - N(i-1)$, $\gamma = \mathrm{sign}(\Delta_N)$
    if $\Delta_N > 0$
      Append Δ_N zero-taps to $\mathbf{w}(i)$
    End
    if $\Delta_N < 0$
      Truncate the last $|\Delta_N|$ taps from $\mathbf{w}(i)$
    End
  End
  if $|n_f(i) - N(i)| < K$ $N(i) = N(i-1)$ End

In the above procedure, the true tap-length $N(i) = \langle n_f(i)\rangle$, where $\langle.\rangle$ rounds the embraced value to the nearest integer. K is the step-size for the tap-length adjustment. $\gamma(i)$ is called the *direction factor*, with which (6) and (9) are merged into one recursive equation. Specifically, at the changing time i, if $N(i) - N(i-1) > 0$, we have $\gamma(i) = 1$ and (6) is actually applied, as otherwise we have $\gamma(i) = -1$ and (9) is used.

Initially we may set $N(0) = N_s$ which is a small integer and $\gamma_0 = 1$. Then the tap-length has a small value at the beginning of the adaptation and increases to a larger value later. This arrangement is obviously appropriate for the convergence of the adaptive filter [2, 3].

In many applications, the power of the desired signal is always normalized to one, implying that the MMSE is usually smaller than one and is thus better represented in a logarithmic scale than in a linear scale. Therefore the adaptation rule of (11) may be modified as:

$$n_f(i+1) = (n_f(i) - \alpha) + \beta\,\gamma(i)\left[\log\bar{\xi}_{\mathrm{p}} - \log\bar{\xi}(i)\right]. \quad (12)$$

And the initialization of $\bar{\xi}_{\mathrm{p}}$ can be set as 1.

4. NUMERICAL SIMULATIONS

In this section, we apply the proposed variable tap-length algorithm to the application of adaptive system modelling as shown in Fig. 1, where $s(i)$ is white Gaussian signal with variance one, $H(z) = 0.35 + z^{-1} + 0.35z^{-2}$, $\eta(i)$ is white noise with SNR 20dB, the logarithm version of recursion (12) is used for the tap-length adaptation. For comparison, we investigate the scenarios when $W_o(z) = W_1(z)$ and $W_o(z) = W_2(z)$ respectively, where:

$$W_1(z) = \frac{1 + 0.2z^{-8}}{1 - 0.7z^{-1}}, \qquad W_2(z) = \frac{1}{1 - 0.3z^{-1}}. \quad (13)$$

The impulse response of $W_o(z)$ is not truncated, and thus any special filter length has not been privileged.

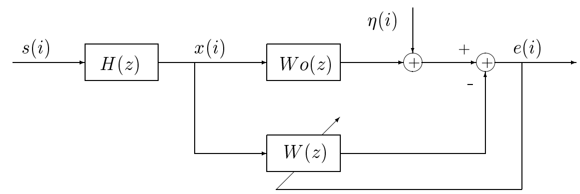


Fig. 1. Block diagram of the adaptive system modelling simulation.

The curves of the converged MSE $\xi_N(\infty)$ with respect to the tap-length N are shown in Fig. 2. The $\mathcal{E}$ defined in Definition 1 is set as 0.04. Then when $W_o(z) = W_1(z)$, the optimum tap-length $N_o = 15$, the suboptimum tap-length set is $(6, 7)$, and when $W_o(z) = W_2(z)$, $N_o = 4$ and no sub-optima are present.

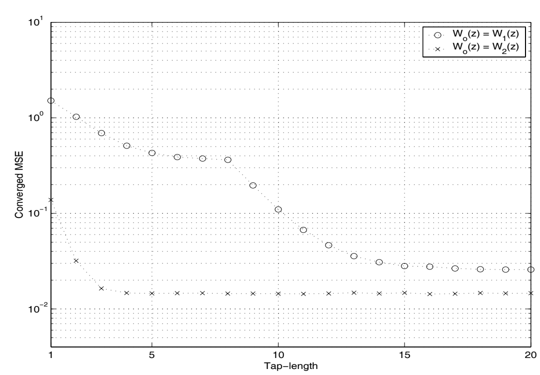


Fig. 2. The curves of the converged MSE and Δ_N with respect to the tap-length.

Fig. 3 shows the learning curves of $n_f(i)$ with different initializations for one typical simulation *run*, where $W_o(z) = W_1(z)$, $\beta = 0.1$, $\alpha = 0.004$ so that $\alpha/\beta = \mathcal{E} = 0.04$, $K = 2$ (i.e. N is adjusted by two every time), and

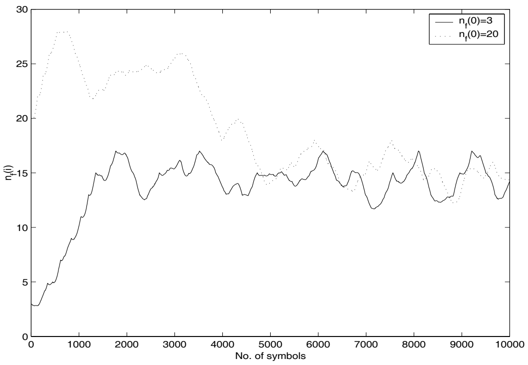


Fig. 3. The learning curves of $n_f(i)$ with different initializations.

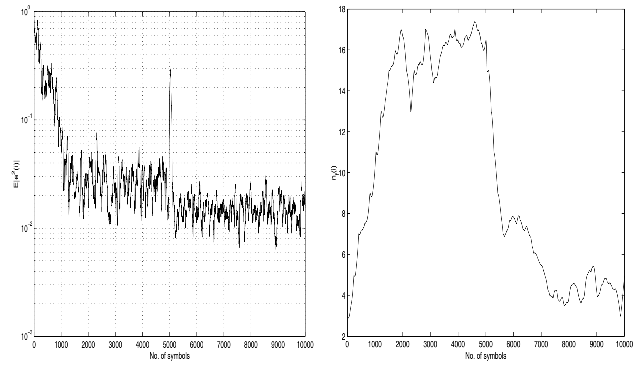
initially $\gamma = 1$ and $\bar{\xi}_{\rm p} = 1$. It is clearly shown in Fig. 3 that $n_f(i)$ converges to around the optimum tap-length for both $n_f(0) = 3$ and $n_f(0) = 20$.

Because the transient behavior of the tap-vector adaptation may initially drive $n_f(i)$ away from N_o, the initial set up of the algorithm may significantly affect the convergence behavior of $n_f(i)$ though it does not influence the final convergence of $n_f(i)$. Therefore in Fig. 3, when $n_f(0) = 3$ and $\gamma = 1$, we observe the "good" initialization that $n_f(i)$ converges to around N_o in about $1,500$ symbols, and when $n_f(0) = 20$ and $\gamma = 1$, we observe the "bad" initialization that $n_f(i)$ converges after around $5,000$ symbols. This problem maybe overcome by resetting the algorithm every time the channel varies sharply, the detail of which is however beyond the scope of this paper and may left to future study.

Fig. 4 (a) and (b) shows the learning curves of the MSE and $n_f(i)$ in a time varying scenario respectively, where $W_o(z) = W_1(z)$ when the numbers of transmitted symbols $i < 5000$, and $W_o(z) = W_2(z)$ when $i \geqslant 5000$. Both learning curves are based on one simulation run, and Fig. 4 (a) is obtained by averaging the MSE learning curve with a rectangular smoothing window of size 50. It is clearly shown in Fig. 4 that the proposed algorithm successfully tracks the channel variations.

5. CONCLUSIONS

This paper defines the optimum and suboptimum tap-lengths for the FIR adaptive filter and proposes a novel gradient search based variable tap-length algorithm using the concept of fractional tap-length. We show that the proposed algorithm can converge to the optimum tap-length in the mean with low complexity, and provide computer simulation results to verify the analysis. Finally we point out that the tap-length adaptation when used in the system modelling application can also be regarded as the channel order estimate, where the optimum tap-length is just the effective channel order.



(a) Learning curve of the MSE.
(b) Learning curve of $n_f(i)$.

Fig. 4. Track the variation of $W_o(z)$.

6. REFERENCES

[1] Y. Gu, K. Tang, H. Cui, and W. Du, "Convergence analysis of a deficient-length LMS filter and optimal-length to model exponential decay impulse response," *IEEE Signal Processing letters*, vol. 10, no. 1, pp. 4 – 7, Jan. 2003.

[2] C. Rusu and C. F. N. Cowan, "Novel stochastic gradient adaptive algorithm with variable length," in *European Conference on Circuit Theory and Design (ECCTD'01)*, Espoo, Finland, Aug. 2001.

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

[4] F. Riera-Palou, J. M. Noras, and D. G. M. Cruickshank, "Linear equalisers with dynamic and automatic length selection," *Electronic Letters*, vol. 37, no. 25, pp. 1553 – 1554, Dec. 2001.

[5] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 1996.

[6] N. J. Bershard, "Analysis of the normalized LMS algorithm with Gussian inputs," *IEEE Trans. on Acoust., Speech & Signal Processing*, vol. ASSP-34, pp. 793–806, Aug. 1986.

[7] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data," *IEEE Trans. on Information Theory*, vol. 34, no. 4, pp. 680–691, July 1988.