

ROBUST NORMALISED LMS FILTERING

Malcolm D. Macleod

QinetiQ Ltd, St Andrews Rd, Malvern, UK, WR14 3PS

ABSTRACT

Several types of robust LMS adaptive filter algorithm have been proposed, to reduce misadjustment when impulsive noise is added to the input or reference signal. In many applications a normalised LMS algorithm is required, to increase convergence speed. This paper shows that if impulsive noise is present at the filter input (which is a realistic problem, for example, in some communications equalisers), the standard NLMS algorithm provides some robustness to this impulsive noise. New normalised LMS algorithms are then presented with improved robustness to impulsive input noise. An approximate theoretical analysis is confirmed by simulations. Finally, we show that, as for NLMS, simplified approximate arithmetic may be used to reduce the implementation cost of the new algorithms.

1. INTRODUCTION

A length- M FIR filter with input $x(k)$ and output $y(k)$ is defined by $y(k) = \mathbf{w}^T(k) \mathbf{x}(k)$ where T is the transpose, $\mathbf{w}(k) = [w_0(k) \ w_1(k) \ \dots \ w_{M-1}(k)]^T$ is the filter coefficient vector and $\mathbf{x}(k) = [x(k) \ x(k-1) \ \dots \ x(k-M+1)]^T$. The error between the filter output and reference signal $d(k)$ is $e(k) = d(k) - y(k)$ and in the LMS algorithm [1,2] the coefficient update at each sample time is given by $\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k) \mathbf{x}(k)$, where μ is the step size. This requires M individual coefficient updates

$$w_i(k+1) = w_i(k) + \mu e(k) x_i(k), \quad i = 1, \dots, M \quad (1)$$

where $x_i(k) = x(k-i)$ is the i^{th} element of $\mathbf{x}(k)$.

If impulsive noise (consisting of “spikes” or “outliers” of large amplitude λ and short duration) is added to $d(k)$ this adds spikes to $e(k)$, and causes coefficient misadjustment (proportional to λ) in (1). The effect of impulsive noise in the input signal $x(k)$ is significantly worse. If a spike of duration one sample and amplitude λ is added to $x(k)$ at sample m , the input signal becomes $x'(k) = x(k) + \lambda \delta(k-m)$. In the LMS algorithm each coefficient is then subjected to M erroneous updates, and severe misadjustment of LMS may occur if λ is large.

This can be a real problem, for example in some communications equaliser applications.

Two classes of “robust” LMS algorithm have been proposed, to reduce this problem. In Order Statistic LMS algorithms [1], *each* coefficient update in (1) is independently non-linearly filtered (for example using a median filter). In the second class, which includes the least mean M -estimate (LMM) algorithm [3], robust mixed norm (RMN) algorithm [4], and adaptive threshold nonlinear algorithm (ATNA) [5], the error $e(k)$ in (1) is replaced by a modified value. The performance of all these algorithms (and others) is compared in [3].

Section 2 shows that the conventional NLMS algorithm alone provides some robustness against impulsive input noise, and section 3 introduces novel normalised LMS algorithms which have increased robustness. These can be used on their own or in conjunction with the LMM algorithm to give even greater robustness. Section 4 shows that, as for NLMS, simplified approximate arithmetic may be used to reduce the implementation cost of the new algorithms.

2. ANALYSIS OF LMS AND NLMS

During the M samples following the input spike (that is, $k = m, \dots, m+M-1$), the output and the error become $y'(k) = y(k) + \lambda w_{k-m}(k)$ and $e'(k) = e(k) - \lambda w_{k-m}(k)$ respectively. Now to simplify analysis we assume that the filter coefficients do not change substantially during those samples. This is a poor assumption for LMS but, as we show later, a more reasonable assumption for NLMS, and the result proves useful for comparing algorithms. For LMS, the total update for the i^{th} coefficient during the M samples following the input spike would then be:

$$\mu \sum_{k=m}^{m+M-1} e'(k) x'_i(k) = \alpha_{1i} + \alpha_{2i} + \alpha_{3i} + \alpha_{4i} \quad (2)$$

$$\alpha_{1i} = \mu \sum_{k=m}^{m+M-1} e(k) x(k-i),$$

$$\text{where } \alpha_{2i} = -\mu \lambda \sum_{k=m}^{m+M-1} w_{k-m} x(k-i),$$

$$\alpha_{3i} = \mu \lambda e(m+i), \text{ and}$$

$$\alpha_{4i} = -\mu \lambda^2 w_i.$$

α_{1i} is the update which would occur with no spike present, and α_{3i} is always negligible in comparison with the other terms. If λ is large, α_{4i} is typically the largest term because it is proportional to λ^2 (although if w_i is small this may not be so). If it is therefore assumed, for simplicity, that the terms α_{4i} are the largest for all i . If the input signal is i.i.d. with variance σ_x^2 , then the resulting misadjusted filter output error variance is given by $\sigma_{em}^2 \approx \sum_i (-\mu\lambda^2 w_i)^2 \sigma_x^2 = \mu^2 \lambda^4 \|\mathbf{w}\|^2 \sigma_x^2$. If also the filter has converged prior to the input spike, so that its coefficient vector \mathbf{w} is close to optimal, then the variance of the reference signal $d(k)$ is $\sigma_d^2 \approx \|\mathbf{w}\|^2 \sigma_x^2$; hence

$$\sigma_{em}^2 \approx \mu^2 \lambda^4 \sigma_d^2, \quad (3)$$

so for large λ , severe misadjustment of LMS occurs, with variance approximately proportional to λ^4 (this is why the initial assumption that the coefficients do not change is poor).

In the LMS algorithm and all the un-normalised robust algorithms discussed in [3], the convergence speed falls as the input signal power falls. To overcome this, in the normalised LMS (NLMS) algorithm [1,2], the constant step size μ is replaced by the variable step size.

$$\mu_{N0} = \frac{\mu_A}{\|\mathbf{x}(k)\|^2 + a}. \quad (4)$$

Typically $\mu_A = 0.25$ to 0.5 . The small constant a prevents division by zero. Even for known input signal power, NLMS may converge faster than LMS [2], and its convergence speed is independent of input power.

Spikes in $d(k)$ do not affect μ_{N0} , so (if $\mu_{N0} \approx \mu$) their effect on the NLMS and LMS algorithms is very similar. However, a spike of amplitude λ added to the input $x(m)$, giving as before $x'(k) = x(k) + \lambda\delta(k-m)$, implies that $\|\mathbf{x}'(k)\| \gg \|\mathbf{x}(k)\|$ during the M samples for which $x(m)$ is in the filter stores ($k = m, \dots, m+M-1$); this increases the denominator of (4). If $\lambda \gg \|\mathbf{x}(k)\|$, then $\|\mathbf{x}'(k)\| \approx \lambda$ so the step size during these samples is approximately $\mu_{N0} \approx \mu_A / \lambda^2$, from (4). Hence from (3) the error variance after misadjustment is $\sigma_{em}^2 \approx \mu^2 \lambda^4 \sigma_d^2 \approx \mu_A^2 \sigma_d^2$, which is independent of λ . Because, for large λ , this misadjustment is much smaller than that for LMS, the original assumption (in section 2) that the coefficients do not change, on which our approximate analysis is based, is more reasonable for NLMS.

The ratio $\gamma = \sigma_d^2 / \sigma_{em}^2$, which is variance of the reference signal divided by the error variance, is the output SNR, so for the NLMS algorithm the output SNR after a large spike is given by $\gamma \approx 1 / \mu_A^2$. For the reasonable value $\mu_A = 0.25$, the output SNR would therefore be 12dB. If $\lambda \leq \|\mathbf{x}(k)\|$, the term α_{4i} becomes smaller than α_{2i} in (2), so as λ decreases the error variance

decreases and the output SNR with respect to misadjustment noise increases.

To confirm these predictions, an example based on that in [3] is used. An adaptive filter with $M = 9$ taps is used for system identification of a 9-tap FIR channel with coefficients $\mathbf{w} = [0.2 \ -0.4 \ 0.6 \ -0.8 \ 1.0 \ -0.8 \ 0.6 \ -0.4 \ 0.2]^T$. The channel input signal is a white zero-mean unit-variance Gaussian process, so the channel output variance σ_d^2 is $\|\mathbf{w}\|^2 = 3.4$. Gaussian noise is added to the reference signal $d(k)$ (i.e. the channel output) to give an SNR of 30dB. Isolated spikes are then added to the adaptive filter input $x(k)$, and the output SNR M samples after each spike is shown in Table 1. Because the theoretical analysis predicts behaviour which depends on the ratio $\lambda / \|\mathbf{x}(k)\|$, and $E(\|\mathbf{x}(k)\|) = \sqrt{M}\sigma_x$, the results in Table 1 are tabulated against the ratio $\lambda / (\sqrt{M}\sigma_x)$. The

presented results are all averages of 100 trials. It can be seen that, as predicted, the output SNR is close to 12dB for large λ , and increases as λ reduces. Results from a second example, using a channel and adaptive filter both of length $M = 19$, are also presented, to confirm that they are approximately independent of filter length M .

Fig. 1 shows the temporal evolution of the output noise to signal ratio (NSR) (σ_e^2 / σ_d^2), again averaged over 100 trials, following three input spikes of amplitudes λ given by $\lambda / (\sqrt{M}\sigma_x) = +8, 0$ and -8 dB.

3. ROBUST NORMALISED ALGORITHMS

To reduce misadjustment further, we may replace μ_{N0} by

$$\mu_{N1} = \frac{\mu_B}{M \max |\mathbf{x}(k)|^2 + a}, \quad (5)$$

where $\max |\mathbf{x}(k)|^2$ is the squared modulus of the largest-modulus element of $\mathbf{x}(k)$. The value of μ_B required in (5) to give the same convergence speed as NLMS depends on the input signal statistics; for a Gaussian input signal, $\mu_B \approx 3\mu_A$ is suitable. If λ is large, then during the M samples $k = m, \dots, m+M-1$ the step size in (5) is approximately $\mu_{N1} \approx \mu_B / (M\lambda^2) = 3\mu_{N0} / M$. Hence from (3) the error variance following misadjustment is $\sigma_{em}^2 = (3/M)^2 \mu_A^2 \sigma_d^2$. This is again independent of λ , but is smaller by a factor of $(M/3)^2$ than for NLMS. (This factor is 9.5 dB for $M = 9$, or 16 dB for $M = 19$.) As λ decreases, the ratio μ_{N1} / μ_{N0} rises from $3/M$ to approximately unity, so the performance of (5) converges towards that of NLMS (4). However, the minimum output SNR due to misadjustment following a single spike (which occurs when $\lambda \approx \|\mathbf{x}(k)\|$) is 18.4 dB for filter length $M = 9$, or 23.3 dB for $M = 19$, compared to the minimum for NLMS, which is approximately 13 dB, as

shown in Table 1. Fig. 1 also shows the reduced misadjustment obtained by using (5).

To achieve even smaller misadjustment, we may replace μ_{N0} by

$$\mu_{N2} = \frac{\mu_C (\text{mean} | \mathbf{x}(k) |)^2}{M \max | \mathbf{x}(k) |^4 + a}. \quad (6)$$

In this case, for a Gaussian input signal $\mu_C = 9\mu_A$ is a suitable choice. In the presence of a large spike of amplitude λ , the value $(\text{mean} | \mathbf{x}(k) |) \approx \lambda/M$, so the error variance is reduced with respect to NLMS by a factor of $(M^3/9)^2$. (This factor is 38 dB for $M = 9$ or 57 dB for $M = 19$.) The minimum output SNR due to misadjustment after a single spike (which occurs when $\lambda \approx 0.6 \| \mathbf{x}(k) \|$) is 21.7 dB for filter length $M = 9$, or 28.5 dB for $M = 19$, compared to the minimum of 13 dB for NLMS, as shown in Table 1. The improvement in SNR obtained using (6) rather than (5) can also be seen in Table 1, and for large spike amplitudes, the use of (6) completely suppresses misadjustment (indicated by underlining in Table 1). Fig. 1 shows the reduced NSR (i.e. increased SNR) obtained by using (6), and the complete suppression of misadjustment for the first two spikes in Fig. 1. However, the normalised step size μ_{N2} in (6) is noisier than μ_{N1} in (5), and some reduction in convergence speed is observed.

Another idea, that of suppressing adaptation completely based on a binary (yes/no) test to detect spikes in the input, was also investigated. However, the maximum output error variance of this approach was found to be no less than that achieved using (5).

4. IMPLEMENTATION ISSUES

Various methods have been proposed for reducing the cost of implementing the standard NLMS computation (4) in fixed-point arithmetic, by using certain approximations. In particular, division may be approximated by a binary shift [6],[7]; the same approximation may also be applied in (5) and (6). The squaring operation in (4) may also be replaced by a circuit for approximate squaring [6]; the simplest such circuit requires no hardware at all - it simply maps the bits of the input value to different locations in the output word. This method can also be used for the squaring and fourth power operations in (5) and (6).

An operation which appears in (5) and (6) but not in (4) is the maximum absolute value operator. For fixed-point arithmetic there is a simple approximation to this operation, as follows. The absolute values are first formed in the normal way (negating negative inputs) and then approximately squared using a method from [6]. The bitwise-OR of all the results is then formed. The result of the OR operation is never less than the true maximum (because the OR operation can only increase the number of ones in the positive binary number, not reduce it). It is also always less than twice the true maximum; the limiting

case is when the true maximum is exactly a power of two, say 2^k , and all the binary digits less significant than 2^k are set to one by the OR-ing operation, giving an output of $2^{k+1}-1$. Bitwise OR has a very low implementation cost.

5. DISCUSSION AND CONCLUSIONS

We have shown that in applications in which impulsive noise may be present in the filter input, the standard NLMS algorithm provides some robustness, unlike LMS. For NLMS the minimum output SNR due to misadjustment after a large input spike is $\gamma = \sigma_d^2 / \sigma_{em}^2 \approx 1 / \mu_A^2$, for example 12dB if $\mu_A = 0.25$.

We have introduced two new normalised LMS algorithms with improved robustness. For $M = 9$, their minimum output SNR is 5dB – 9dB higher than that of NLMS. For $M = 19$, their minimum output SNR is 10dB – 15dB higher than that of NLMS. In the case of the second new normalisation (6), the degree of improvement increases as the spike amplitude increases, so that complete suppression of the effects of large spikes is achieved.

Finally, we have shown that, as with NLMS, simplified approximate arithmetic may be used to reduce the implementation cost of the new normalisation formulae.

If there is also impulsive noise in the reference signal, then any of the above normalised algorithms could be combined with the LMM algorithm [3]. Since the LMM algorithm would then only have to reject the short spikes from the reference signal, it could employ a much shorter median filter than is required to reject the effect at the adaptive filter output of a spike at its input.

The new robust normalised LMS algorithms may be of use in applications such as communications system equalisers where a minimum output SNR of 20dB may be adequate. If the output SNR is required to be greater, then either conventional NLMS or one of the new normalised algorithms may be combined straightforwardly with the LMM algorithm [3] to provide both normalisation and greater robustness.

6. REFERENCES

- [1] P. M. Clarkson, *Optimal and Adaptive Signal Processing*, CRC Press Inc., Boca Raton, FL, 1993.
- [2] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, NJ, 1996.
- [3] Y. Zou, S.-C. Chan and T.-S. Ng, "Least Mean M-Estimate Algorithms for Robust Adaptive Filtering in Impulsive Noise", *IEEE Trans Circuits Syst. II*, vol. 47, pp. 1564–1569, Dec. 2000.
- [4] J. A. Chambers and A. Avlonitis, "A robust mixed-norm (RMN) adaptive filter algorithm," *IEEE Signal Processing Lett.*, vol. 4, no. 2, pp. 46–48, 1997.

[5] S. Koike, “Adaptive threshold nonlinear algorithm for adaptive filters with robustness against impulsive noise,” *IEEE Trans. Signal Processing*, vol. 45, no. 9, pp. 2391–2395, 1997.

[6] Cheng, S and Evans, J, “Implementation Of Signal Power Estimation Methods”, *IEEE Trans Circ Sys-II*, 44, 3, Mar 97, pp 240-250.

[7] Xue, P and Liu, B, “Adaptive Equalizer Using Finite-Bit Power-of-Two Quantizer”, *IEEE Trans ASSP*, 34, 6, Dec 86, pp1603-1611.

$\lambda / (\sqrt{M}\sigma_x)$ (dB)	12	8	4	0	-4	-8	-12
NLMS (4), M = 9	13.1	12.5	14.3	15.8	18.1	20.6	23.7
(5), M = 9	21.8	20.3	20.2	18.4	19.4	21.5	23.7
(6), M = 9	<u>34.4</u>	32.5	28.9	22.3	21.7	23.7	24.4
NLMS (4), M = 19	13.0	13.1	14.2	16.1	19.0	21.7	24.8
(5), M = 19	27.3	26.7	25.9	23.5	23.3	24.7	26.7
(6), M = 19	<u>35.0</u>	<u>35.1</u>	<u>34.8</u>	29.7	28.5	29.1	29.6

Table 1. Output SNR ($\gamma = \sigma_d^2 / \sigma_{em}^2$) (dB) after misadjustment for 3 normalised LMS algorithms, as a function of relative input spike amplitude (dB). underlining indicates that the effects of spike were completely suppressed in those cases.

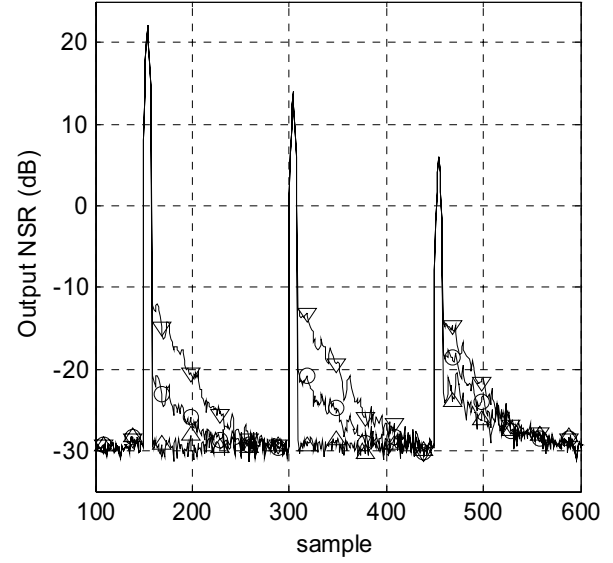


Fig 1. Output NSR (σ_e^2 / σ_d^2) for 3 normalised LMS algorithms: ∇ = NMLS; \circ = equation (5); Δ = equation (6). Learning curves averaged over 100 trials. $M = 9$. Input spike amplitudes +8, 0 and -8 dB w.r.t. $\sqrt{M}\sigma_x$, at samples number 150, 300 and 450.