On the Performance of too Short Adaptive FIR Filters

Daniël W.E. Schobben and Piet C.W. Sommen

Abstract— **Performance analyses of adaptive algorithms such as LMS and RLS often rely on the assumption that the input signal is stationary. Also, it often is assumed, that the adaptive finite impulse response (FIR) filter is long enough to make a good approximation of the unknown system. In many practical situations these assumptions do not hold and the instantaneous misadjustment of the adaptive filter can grow large [1]. In this paper this effect is investigated and two methods are presented to improve the performance of the adaptive filter.**

Keywords— **Short Adaptive Filter, Performance.**

I. Introduction

An adaptive filter convergences to the Wiener solution if it's input signal is stationary. For an adaptive FIR filter which length is shorter than the impulse response of the unknown system, the Wiener solution relies on the statistics of the input signal to achieve optimal performance. If the statistics of the input signal change, the performance of the adaptive filter deteriorates. The goal of the paper is to investigate how large this deterioration is and what can be done about it. Therefore, a performance analysis is presented for a too short adaptive FIR filter with a non-stationary input signal. Simulations are included which show that this performance can degrade significantly as a result of the non-stationary input signal. In addition it is shown that this problem can alleviated in two different ways.

A character which denotes a vector will be underlined. Superscripts denote vector or matrix dimensions, a matrix with one superscript is square. Complex conjugate transpose is denoted as $.^H$ and the expectation operator is denoted by $E\{.\}$. The $N \times N$ matrix identity and the $K \times L$ zero matrix will be denoted by I^N and $0^{K,L}$ respectively.

Consider the length N adaptive filter $\underline{w}^N[k]$ (Figure 1) which is trying to minimize the residual signal $r[k] = \tilde{e}[k] - (\underline{w}^N[k])^H \underline{x}^N[k]$, with $\underline{x}^N[k] = (x[k-N+1], \ldots, x[k])^T$. The residual is defined as $r[k] = \tilde{e}[k] - (\underline{w}^N[k])^H \underline{x}^N[k]$. Minimizing $E\{|r[k]|^2\}$ yields the Wiener solution [2] [3] $\underline{w}^N_{\mathrm{opt}} = (R^N_x)^{-1} \underline{\rho}^N_{x\tilde{e}}$, where $\tilde{e}[k] = s[k] + \sum_{i=0}^{M-1} h_i x[k-i]$, $R^N_x = E\{\underline{x}^N[k](\underline{x}^N[k])^T\}$ and $\underline{\rho}^N_{x\tilde{e}} = E\{\underline{x}^N[k]\tilde{e}[k]\}$. Assuming that s and x are independent of each other, the j^{th} entry of this cross-correlation vector equals ($j \in [0, \ldots, N-1]$)

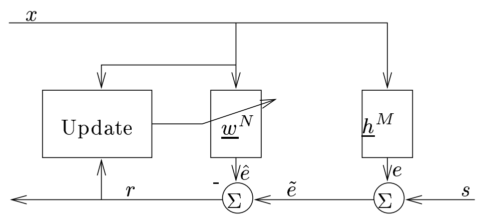

Fig. 1. Adaptive filter

$$\begin{aligned} (\underline{\rho}^N_{x\tilde{e}})_j &= E\left\{ x[k-N+1+j] \sum_{i=0}^{M-1} h_i x[k-i] \right\} , \\ &= \sum_{i=0}^{M-1} h_i E\left\{ x[k-N+1+j] x[k-i] \right\} , \\ &= \sum_{i=0}^{M-1} h_i \rho_{N-1-i-j} , \end{aligned} \tag{1}$$

with $\rho_i = E\{x[k]x[k-i]\}$. For $M > N$ the cross-correlation vector can now be written as

$$\underline{\rho}^N_{x\tilde{e}} = \begin{pmatrix} \rho_{-M+N} & \cdots & \rho_{-1} & \rho_0 & \cdots & \rho_{N-1} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ \rho_{-M+1} & \cdots & \rho_{-N} & \rho_{-N+1} & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} h_{M-1} \\ \vdots \\ h_0 \end{pmatrix} . \tag{2}$$

Observe that for $M = N$, $\underline{w}^N_{\mathrm{opt}} = (R^N_x)^{-1} R^N_x \underline{h}^N = \underline{h}^N$. For $M > N$ however,

$$\underline{w}^N_{\mathrm{opt}} = \left(\tilde{R}^{N,M-N} | \mathrm{I}^N \right) \underline{h}^M , \tag{3}$$

with

$$\tilde{R}^{N,M-N} = (R^N_x)^{-1} \begin{pmatrix} \rho_{-M+N} & \cdots & \rho_{-1} \\ \vdots & & \vdots \\ \rho_{-M+1} & \cdots & \rho_{-N} \end{pmatrix} . \tag{4}$$

The authors are with the Eindhoven University of Technology, P.O.Box 513, 5600 MB Eindhoven, the Netherlands. E-mail: D.W.E.Schobben@ele.tue.nl, P.C.W.Sommen@ele.tue.nl

If $\rho_i \neq 0$ for $-M < i < 0$, the N optimal adaptive weights are obtained from the first N weights of the unknown system plus a convolution of ρ_i and h_j, with $-M+1 \leq i < 0$ and $N \leq j < M$. The latter contribution depends on the statistics of x. As a result, the misadjustment of the adaptive filter deteriorates momentarily when the statistics of the input signal change.

II. Final misadjustment

In the following analyses, it is assumed that there is no double talk, i.e. $s[k]=0$. The final misadjustment $J_{\text{min}} = E\{|r[k]|^2\}$ is defined as

$$\begin{aligned} J_{\text{min}} &= E\{(\underline{h}^M - \underline{\tilde{w}}^M[k])^H \underline{x}^M[k])^2\}\,, \\ &= (\underline{h}^M - \underline{\tilde{w}}^M[k])^H R_x^M (\underline{h}^M - \underline{\tilde{w}}^M[k]), \end{aligned} \quad (5)$$

with

$$\underline{\tilde{w}}^M[k] = \begin{pmatrix} \underline{0}^{M-N} \\ \underline{w}^N[k] \end{pmatrix}. \quad (6)$$

Now let

$$\underline{h}^M = \begin{pmatrix} \underline{h}_\Delta^{M-N} \\ \underline{h}_{\Delta 2}^N \end{pmatrix}. \quad (7)$$

If $\underline{w}^N[k]$ equals the Wiener solution, i.e. $\underline{w}^N[k] = \underline{h}_{\Delta 2}^N + \tilde{R}^{N,M-N} \underline{h}_\Delta^{M-N}$, the final misadjustment equals

$$J = (\underline{h}_\Delta^{M-N})^H \cdot \begin{pmatrix} \mathrm{I}^{M-N} \\ -\tilde{R}^{N,M-N} \end{pmatrix}^H R_x^M \begin{pmatrix} \mathrm{I}^{M-N} \\ -\tilde{R}^{N,M-N} \end{pmatrix} \underline{h}_\Delta^{M-N}. \quad (8)$$

However, if the filter has converged to the Wiener solution and R_x changes, the misadjustment changes from (8) to

$$J = (\underline{h}_\Delta^{M-N})^H \cdot \begin{pmatrix} \mathrm{I}^{M-N} \\ -\tilde{R}^{N,M-N} \end{pmatrix}^H R_{\text{new}}^M \begin{pmatrix} \mathrm{I}^{M-N} \\ -\tilde{R}^{N,M-N} \end{pmatrix} \underline{h}_\Delta^{M-N}. \quad (9)$$

The $\tilde{R}^{N,M-N}$ are based on the old statistics. If the adaptive filter would not converge to the Wiener solution, but would make an unbiased estimate of the first N coefficients of the unknown system instead, the final misadjustment would be

$$J_{\text{min}} = (\underline{h}_\Delta^{M-N})^H R_x^{M-N} \underline{h}_\Delta^{M-N}. \quad (10)$$

The filter would no longer have to re-converge in the case that R_x changes to R_{new} and the misadjustment would immediately equal the new final misadjustment $(\underline{h}_\Delta^{M-N})^H R_{\text{new}}^{M-N} \underline{h}_\Delta^{M-N}$. However, this would be at the expense of a higher misadjustment.

III. Dynamic Behavior

In this section, the dynamic behavior of the adaptive filter is considered. The LMS algorithm can be written as

$$\begin{aligned} \underline{d}^M[k+1] &= \left(\mathrm{I}^M - 2\alpha \begin{pmatrix} \underline{0}^{M-N} \\ \underline{x}^N[k] \end{pmatrix} (\underline{x}^M[k])^H\right) \underline{d}^M[k] \\ &- 2\alpha \begin{pmatrix} \underline{0}^{M-N} \\ \underline{x}^N[k] \end{pmatrix} s[k]\,, \end{aligned} \quad (11)$$

with the difference weight vector defined as $\underline{d}^M[k] = \underline{h}^M - \underline{\tilde{w}}^M[k]$. The entries of the difference vector that change in time $\underline{d}^N[k] = \underline{h}_{\Delta 2}^N - \underline{w}^N[k]$ are obtained from (11)

$$\begin{aligned} \underline{d}^N[k+1] &= \left(\mathrm{I}^N - 2\alpha \underline{x}^N[k] (\underline{x}^N[k])^H\right) \underline{d}^N[k] \\ &-2\alpha \underline{x}^N[k] \left(s[k] + (\underline{h}_\Delta^{M-N})^H \underline{x}^{M-N}[k-N]\right), \end{aligned} \quad (12)$$

By assuming that the input signal is almost stationary for a certain time interval, i.e. $\underline{d}^N[k]$ changes fast compared to the possible changes in the input signals statistics. Also it is assumed that the adaptation step size α is small, so that the input signal vector and the difference vector $\underline{d}^N[k]$ may be separated under the expectation operator. The expectation of the weight difference vector becomes

$$\begin{aligned} E\{\underline{d}^N[k+1]\} &= \left(\mathrm{I}^N - 2\alpha R_x^N\right) E\{\underline{d}^N[k]\} \\ &-2\alpha \left(0^{N,M-N} \mathrm{I}^N\right) R_x^M \begin{pmatrix} \mathrm{I}^{M-N} \\ 0^{N,M-N} \end{pmatrix} \underline{h}_\Delta^{M-N}, \end{aligned} \quad (13)$$

so that

$$\begin{aligned} E\{\underline{d}^N[k]\} &= \left(\mathrm{I}^N - 2\alpha R_x^N\right)^k \left(\underline{d}^N[0] + \tilde{R}^{N,M-N} \underline{h}_\Delta^{M-N}\right) \\ &-\tilde{R}^{N,M-N} \underline{h}_\Delta^{M-N}. \end{aligned} \quad (14)$$

Consider the case that the adaptive filter has converged and the input signal's statistics changes at time $k=0$, then the difference vector becomes

$$\begin{aligned} E\{\underline{d}^N[k]\} &= \left(\mathrm{I}^N - 2\alpha R_{\text{new}}^N\right)^k \left(\tilde{R}_{\text{new}}^{N,M-N} - \tilde{R}^{N,M-N}\right) \underline{h}_\Delta^{M-N} \\ &-\tilde{R}_{\text{new}}^{N,M-N} \underline{h}_\Delta^{M-N}. \end{aligned} \quad (15)$$

At $k=0$, this vector is $\left(\tilde{R}_{\text{new}}^{N,M-N} - \tilde{R}^{N,M-N}\right) \underline{h}_\Delta^{M-N}$ away from it's new final solution. This contribution vanishes with $\left(\mathrm{I}^N - 2\alpha R_{\text{new}}^N\right)^k$. In the end, the difference vector reaches its new final solution, i.e. $E\{\underline{d}^N[\infty]\} = -\tilde{R}_{\text{new}}^{N,M-N} \underline{h}_\Delta^{M-N}$.

IV. Experiments

The acoustical impulse response used in the simulations is a measured room impulse response sampled at 48kHz and is plotted is Figure 2. The first 120 samples approximately equal zero, so that the direct path is 2.5 ms. It is assumed that the impulse response equals zero after 500 samples. An adaptive filter is used to estimate the first 300 coefficients of the room impulse response. This is a reasonable choice since the delay in the direct path and the reverberation time are not known in advance. It is assumed that there is no double talk ($\sigma_s^2 = 0$) in the simulations.

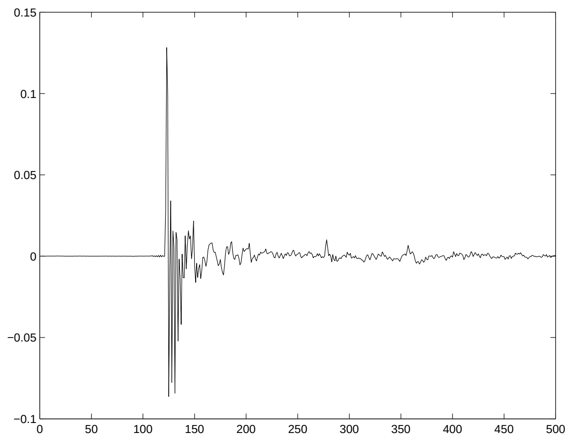

Fig. 2. Acoustical impulse response

First an AR(1) input signal is considered, so that $x[k] = i[k] + ax[k-1]$ with $i[k]$ a white noise sequence with $E\{(i[k])^2\} = 1$ and $E\{i[k]\} = 0$. The previous and the new parameter of the AR(1) process are (a, a_{new}) shown in the first and in the second column of Table I. Column 3 to 6 show the final misadjudgment given by (8) (old Wiener solution), (9) (old Wiener solution, new statistics $R_x^M = R_{\text{new}}^M$), (8) (new Wiener solution), (10) (old unbiased solution), (10) (new unbiased solution). These final misadjudgements are all normalized by the reference signal power $E\{(\tilde{e}[k])^2\} = (\underline{h}^M)^T R^M \underline{h}^M$, with $R^M = R_x^M$ or $R^M = R_{\text{new}}^M$. It follows from this table that $J'_{\text{OldWiener}}$ can become larger than $J_{\text{NewNobais}}$. This means that the term $\tilde{R}^{N,M-N}\underline{h}^{M-N}$ in (3) can decrease the adaptive filter performance when the statistics of the input signal change. It can easily be shown that the Wiener term only affects one adaptive weight $(\underline{w}^N)_0$ when the input signal is an AR(1) process. This can also be seen from Figure 3 where a part of the Wiener solution corresponding to the is plotted which corresponds to the first entry of Table I ($a = 0.9, a_{\text{new}} = 0.1$).

Next, simulations are presented for a more complex ARMA process. The previous and the new parameters of the ARMA process are shown in the first and in the second column of Table II. For example, the input signal for the first simulation was generated from $x[k] - 0.1x[k-1] + x[k-2] + 0.5x[k-3] = i[k] + 0.9i[k-1]$, with $i[k]$ a white noise sequence. The Wiener solution corresponding to this first simulation is plotted in Figure 4. The figure shows that the Wiener solution again equals the impulse response except for the last few coefficients. From Table II it follows that for the ARMA signals considered, the deterioration for changing statistics is not too large, for the impulse response considered.

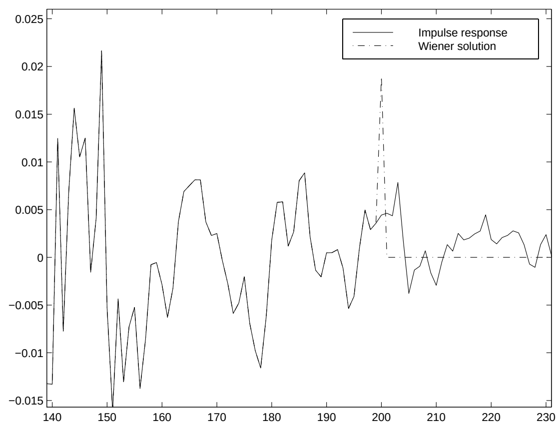

Fig. 3. Wiener solution for an AR(1) process

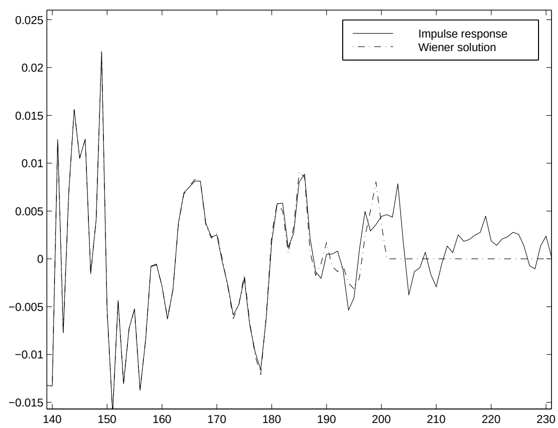

Fig. 4. Wiener solution for an ARMA process

V. Improving the performance of the too short adaptive filter

A solution to the problem of increased errors due to a non-stationary input signal can be to increase the adaptation step-size for the last few weights. Large errors can still occur for a short time, but the adaptive filter re-converges faster when a non-stationarity is encountered. The optimal distribution of these adaptation step-sizes are dependent of the input signal and the impulse response. Thus, the step-sizes can be optimized for a particular application.

When non-stationarities occur very often, it can be desirable to estimate the first N coefficients of the unknown system, instead of using the Wiener solution. This can be achieved in the following way. The adaptive filter size is increased to $N+P$ weights, so that the first N weights of the adaptive filter resemble the corresponding weights of the unknown system. The adaptive filter can now generate two signals. The first being obtained from the first N tabs, and the second being obtained from all $N+P$ tabs. The second signal is now used to calculate the residual signal which controls the adaptive weights. The first signal is used to generate the residual signal which is sent to the far end. This signal is no longer affected by non-stationarities in the input signal. A small overhead is required however for the P additional weights.

TABLE I

Simulation results for an AR(1) process

a	a_{new}	$J_{\text{OldWiener}}$	$J'_{\text{OldWiener}}$	$J_{\text{NewWiener}}$	$J_{\text{OldNobais}}$	$J_{\text{NewNobais}}$
0.9	0.1	0.0789	0.0228	0.0195	0.0835	0.0195
0.5	-.05	0.0319	0.0120	0.0110	0.0322	0.0111
0.99	-0.99	0.2283	3.1177	0.0105	0.4613	0.0169

TABLE II

Simulation results for an ARMA process

old	new	$J_{\text{OldWiener}}$	$J'_{\text{OldWiener}}$	$J_{\text{NewWiener}}$	$J_{\text{OldNobais}}$	$J_{\text{NewNobais}}$
[1 -0.1 1 0.5],[1 0.9]	[1 -0.2 0 .4],[1 .1]	0.0149	0.0116	0.0111	0.0161	0.0112
rand(N/8,1),[1 0.9]	rand(N/8,1),[1 .5]	0.0564	0.0651	0.0649	0.0611	0.0696
rand(N/8,1),[1 0.9]	[1 0.1],[1]	0.0398	0.0214	0.0193	0.0423	0.0193

VI. Conclusion

An analysis of a too short adaptive filter is presented in this paper. It is shown that the Wiener solution which performs well for a stationary input signal, is momentary impaired when the input signal is non-stationary. Simulations where done which confirm this and show that the problem emerges from only a few adaptive weights. A solution to the problem is to increase the adaptation stepsize parameters associated with these weights. In this way the misadjustment error can still grow large at the moment that the input signal's statistics change, but the adaptive filter reconvergences much faster. An other solution is to omit the few adaptive weights that cause the problem in the adaptive filter. The adaptive filter no longer has to reconverge when a non-stationarity occurs.

References

[1] R.D. Poltman, "Stochastic gradient algorithm for system identification using adaptive fir-filters with too low number of coefficients," *IEEE Trans. Circuits and Systems*, vol. 35, no. 2, pp. 247–250, February 1988.

[2] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs : Prentice Hall, 2nd edition, 1991.

[3] B. Widrow and S.D. Stearns, *Adaptive signal processing*, Englewood Cliffs, N.J. : Prentice-Hall, 1985.