THE AVERAGED, OVERDETERMINED AND GENERALISED LMS (AOGLMS) ALGORITHM

E. Alameda-Hernández[(a)]*, D. P. Ruiz*[(b)]*, D. Blanco*[(c)]*, D. C. McLernon*[(a)] *and M. C. Carrión*[(b)]

[(a)] School of Electronic and Electrical Engineering. University of Leeds. Leeds LS2 9JT. UK.
[(b)]Department of Applied Physics. University of Granada. 18071 Granada. Spain
[(c)]School of Engineering and Electronics. The University of Edinburgh. Edinburgh EH9 3JL. UK
E-mails: ealameda@ugr.es, druiz@ugr.es, David.Blanco@ee.ed.ac.uk, d.c.mclernon@ee.leeds.ac.uk

ABSTRACT

This contribution presents a new algorithm of the LMS family, derived from a novel orthogonality condition that holds for overdetermined problems that include an instrumental variable. This instrumental variable can be used to introduce higher-order statistics information. The convergence of the MSE for this new algorithm is theoretically studied, together with its superior performance when compared with other similar algorithms, under quite general hypotheses. The algorithm is then applied to the blind identification of moving average models; simulation results verify the analysis.

1. INTRODUCTION

Many applications of fundamental importance require the use of adaptive algorithms, like when solving problems with non-stationary statistics or in real-time implementations, as in communications, medicine, geophysics, etc.

The first approach of this kind was the Least Mean Square (LMS) algorithm [1], designed to minimise the Mean Squared Error (MSE) when performing a linear estimation. The solution to this problem of linear estimation, with the MSE as cost function, is attained when the estimation error and the vector of data used in the estimation are orthogonal, -i.e. the expected value of the product vanishes.

The same problem can be reformulated using a sum of squares, instead of the expected value operator, as the scalar product employed in the orthogonality condition. In this case an exact recursive solution can be found algebraically, giving rise to the Recursive Least Squares (RLS) algorithm. The RLS algorithm is more computationally demanding than the LMS algorithm, but as a reward, produces estimates asymptotically free of noise.

Following a similar procedure than when the RLS was obtained, if in the orthogonality condition the vector of data is replaced by an instrumental variable (IV) vector, the Recursive Instrumental Variable (RIV) algorithm is obtained, see [2] and references therein. The introduction of the IV allows for better properties of the estimated unknowns, like for example the benefits deriving from the use of higher-order statistics (HOS).

Another member of the RLS family can be added if the length of the vector of IV is larger than the number of unknowns. The overdetermined problem formulated in this case is solved via the Overdetermined Recursive Instrumental Variable (ORIV) algorithm [3]. In the LMS family, the equivalent to the RIV algorithm is the Generalised Least Mean Square (GLMS) algorithm [4], in the sense that can deal with IV's as well. But there is no mathematically-simple equivalent to the ORIV algorithm. The purpose of this work is to find such an algorithm.

This work was partially supported by the Ministerio de Ciencia y Tecnología of Spain under project TIC-2001-2902.

2. A NEW ORTHOGONALITY CONDITION

Let us assume that the solution w_0 to a given problem is obtained from

$$Rw_0 = r \quad (1)$$

and that this equation has to be solved adaptively and using a very reduced computational burden. The matrix R is the correlation matrix between a vector of instruments $\tilde{x}(n) = [\tilde{x}(n),\ \tilde{x}(n-1),\ \ldots,\ \tilde{x}(n-l+1)]^t$ and a vector of data $x(n) = [x(n),\ x(n-1),\ \ldots,\ x(n-q+1)]^t$, and the vector r is the correlation vector between the same IV vector and a desired response $d(n)$. In what follows $l > q$, so the matrix equation (1) is overdetermined.

The actual definition of $\tilde{x}(n)$, $x(n)$ and $d(n)$ depends on the problem to solve.

In real applications the correlations are not known, but they have to be estimated. Assuming ergodic series, the expected value operator can be approximated by temporal averages, so R and r can be estimated respectively by:

$$\Phi(n) = \sum_{i=1}^{n} \lambda^{n-i} \tilde{x}(i) x^t(i) \quad (2)$$

$$z(n) = \sum_{i=1}^{n} \lambda^{n-i} \tilde{x}(i) d(i) \quad (3)$$

The forgetting factor $0 < \lambda \leq 1$ must take a value less than 1 in non-stationary problems. Using these estimates the problem now turns to:

$$\Phi(n) w(n) = z(n) \quad (4)$$

Equation (4) can be solved recursively using the ORIV algorithm, shown in table 1.

From the update recursion for the weight vector $w(n)$ (11), another for the error vector $\Delta w(n) = w(n) - w_0$ can be derived:

$$\Delta w(n) = [I - K(n) X^t(n)] \Delta w(n-1) + K(n) \alpha_0(n) \quad (12)$$

where $\alpha_0(n) = v(n) - X^t(n) w_0$.

If in (12) the definition of the gain matrix $K(n) = \Gamma^{-1}(n) X(n) \Lambda^{-1}(n)$ is used, and then multiplied by $\Gamma(n)$, we

Table 1: The ORIV algorithm [3]

<u>Initial Conditions</u>
$\Phi^t(0) = \delta[I_{q\times q}\|0_{q\times l-q}]$; δ = arbitrary scalar
$\Gamma^{-1}(0) = \frac{1}{\delta^2}I_{q\times q}$; $w(0) = 0$; $z(0) = 0$
<u>Recursive process, for $n = 1,2...$</u>
Obtain $\{\tilde{x}(n), x(n), d(n)\}$ and construct $\tilde{x}(n)$ and $x(n)$
$X(n) = \left[\Phi^t(n-1)\tilde{x}(n) \;\; x(n)\right]$ (5)
$\lambda^2\Lambda(n) = \begin{pmatrix} -\tilde{x}^t(n)\tilde{x}(n) & \lambda \\ \lambda & 0 \end{pmatrix}$ (6)
$K(n) = \Gamma^{-1}(n-1)X(n)[\lambda^2\Lambda(n) + X^t(n)\Gamma^{-1}(n-1)X(n)]^{-1}$
$\Gamma^{-1}(n) = \frac{1}{\lambda^2}[\Gamma^{-1}(n-1) - K(n)X^t(n)\Gamma^{-1}(n-1)]$
$v(n) = \begin{pmatrix} \tilde{x}^t(n)z(n-1) \\ d(n) \end{pmatrix}$ (7)
$\alpha(n) = v(n) - X^t(n)w(n-1)$ (8)
$z(n) = \lambda z(n-1) + \tilde{x}(n)d(n)$ (9)
$\Phi(n) = \lambda\Phi(n-1) + \tilde{x}(n)x^t(n)$ (10)
$w(n) = w(n-1) + K(n)\alpha(n)$ (11)

get a useful expression:

$$\begin{aligned}\Gamma(n)\Delta w(n) =& \lambda^2\Gamma(n-1)\Delta w(n-1) + \\ &+ X(n)\Lambda^{-1}(n)\alpha_0(n)\end{aligned} \tag{13}$$

where the update expression for $\Gamma(n) = \Phi^t(n)\Phi(n)$ has been taken into account :

$$\Gamma(n) = \lambda^2\Gamma(n-1) + X(n)\Lambda^{-1}(n)X^t(n)$$

Expression (13) is a difference equation for $\Gamma(n)\Delta w(n)$, and because the total solution must tend to cero to verify the convergence in the mean (tracking) of the ORIV algorithm, the particular solution must vanish.

It can be shown that this last requisite is accomplished, assuming stationarity of the statistics involved in the problem, if

$$\mathrm{E}[X(n)\Lambda^{-1}(n)\alpha_0(n)] = 0 \tag{14}$$

That is, a new orthogonality condition is derived between the matrix of data $X(n)$, and the vector of error $\alpha_0(n)$, connected by the matrix $\Lambda(n)$. In particular, this orthogonality principle holds for overdetermined problems involving HOS; as such it is called the Overdetermined and Generalised Orthogonality Principle, and it must be satisfied by the solution of equation (1).

3. THE AOGLMS ALGORITHM

In this section, borrowing some ideas from the stochastic gradient methods, a new LMS-type algorithm is derived. The derivation is split into 3 steps:

3.1 Step 1:

Inserting the orthogonality condition (14) into an update recursion will guaranty convergence to the solution of (4):

$$w(n) = w(n-1) + \frac{\mu}{n}X(n)\Lambda^{-1}(n)\alpha(n) \tag{15}$$

where μ is the step-size, that as usual controls the convergence, and n assures normalization of the magnitudes under stationary statistics. The algorithm based on the update recursion (15) is the LMS/ORIV of [5].

3.2 Step 2:

To simplify the previous recursion, using (5), (6), (7) and (8), the gradient can be written as a sum of two vectors:

$$\begin{aligned}&X(n)\Lambda^{-1}(n)\alpha(n) = \\ &= \lambda x(n)\tilde{x}^t(n)[z(n-1) - \Phi(n-1)w(n-1)] \\ &\quad + \Phi^t(n)[\tilde{x}(n)d(n) - \tilde{x}(n)x^t(n)w(n-1)]\end{aligned} \tag{16}$$

Each of these vectors has a null expected value, which means that any of them, independently, can serve as a proper gradient to construct the corresponding stochastic gradient algorithm.

It can be proven that the one containing the averaged quantities $\Phi(n)$ and $z(n)$ in the difference, that is, the first summand, yields the algorithm providing the best estimates. So now we propose this new update recursion:

$$\begin{aligned}&w(n) = w(n-1) + \\ &+ \frac{\mu}{n}x(n)\tilde{x}^t(n)[z(n-1) - \Phi(n-1)w(n-1)]\end{aligned} \tag{17}$$

3.3 Step 3:

To reduce the operations per iteration, the update recursion of $\bar{e}(n) = z(n) - \Phi(n)w(n)$ in (17), called the averaged error vector, is obtained from (9), (10) and (17),

$$\bar{e}(n) = [\lambda - \frac{\mu}{n}\Phi(n)x(n)\tilde{x}^t(n)]\bar{e}(n-1) + \tilde{x}(n)e(n)$$

This last expression means that $\bar{e}(n)$ is a filtered version of $\tilde{x}(n)e(n)$ and also that most of the computational burden is in the product $\Phi(n)x(n)\tilde{x}^t(n)$. In order to reduce this computational burden, this product is replaced by a new free parameter of the algorithm f, a scalar.

With these final simplifications we get to the desired new algorithm, summarised in table 2.

This new algorithm is called Averaged Overdetermined and Generalised LMS algorithm (AOGLMS). The adjective '*generalised*' comes from the fact that can deal with IV's.

The simulation results will justify all the used simplifications.

4. CONVERGENCE ANALYSIS OF THE AOGLMS ALGORITHM

The function MSE(n) will be computed from the weight error autocorrelation matrix $R_\Delta(n) = \mathrm{E}[\Delta w(n)\Delta w^t(n)]$, as

Table 2: The AOGLMS algorithm

Initial Conditions
f, μ = arbitrary scalar; $w(0) = 0$; $\bar{e}(0) = 0$
Recursive process, for $n = 1,2...$
Obtain $\{\tilde{x}(n), x(n), d(n)\}$ and construct $\tilde{x}(n)$ and $x(n)$
$e(n) = d(n) - w^t(n-1)x(n)$ (18)
$\bar{e}(n) = f\bar{e}(n-1) + (1-f)\tilde{x}(n)e(n)$ (19)
$w(n) = w(n-1) + \mu x(n)\tilde{x}^t(n)\bar{e}(n)$ (20)

$\mathrm{MSE}(n) = \mathrm{Tr}\{R_\Delta(n)\}$. From (20) and using (18) and (19):

$$\begin{aligned}\Delta w(n) =& [I - \mu_n x(n)\tilde{x}^t(n)\tilde{x}(n)x^t(n)]\Delta w(n-1) - \\ &- \mu_n x(n)\tilde{x}^t(n)\sum_{i=1}^{n-1} f^{n-i}\tilde{x}(i)x^t(i)\Delta w(i-1) + \\ &+ \mu_n x(n)\tilde{x}^t(n)\sum_{i=1}^{n} f^{n-i}\tilde{x}(i)e_0(i)\end{aligned} \quad (21)$$

where $\mu_n = \mu(1-f)/(1-f^n)$ and $e_0(n) = d(n) - w_0^t x(n)$. Assume that for small μ the following idependence assumption (IA) holds:

(IA) The weight vector $w(n)$, and hence, the error vector, is independent of the vector of data $x(i)$ and of the instrumental variable $\tilde{x}(i)$ for $i = 1...n-1$.

This kind of soft independence assumption will simplify the mathematics; independence assumptions have been proven to yield quite accurate results. It can be shown from (21), working in the lower possible order in μ, that:

1. $R_\Delta(n, n-j) = 0$ for $j = 2...n$.
2. $R_\Delta(n, n-1) \sim R_\Delta(n-1)$.

where $R_\Delta(n,i) = \mathrm{E}[\Delta w(n)\Delta w(i)]$ and $a \sim b$ means that a and b behave nearly the same.

In this way, from (21), we can finally get:

$$\begin{aligned}R_\Delta(n) =& R_\Delta(n-1) - \mu_\infty \chi(0) R_\Delta(n-1) - \\ &- \mu_\infty R_\Delta(n-1)\chi(0) - \mu_\infty f\chi(0)R_\Delta(n-2) - \\ &- \mu_\infty f R_\Delta(n-2)\chi(0) + Nh(n)\end{aligned} \quad (22)$$

where $Nh(n)$ groups all the independent terms:

$$\begin{aligned}Nh(n) =& \mu_\infty \frac{1}{\sum_{\tau=0}^{n-1} f^\tau \chi(\tau)} \sum_{i=0}^{L} f^i \gamma(i) \sum_{i=0}^{L} f^i \gamma^t(i) + \\ &+ \mu_\infty \sum_{i=0}^{L} f^i \gamma(i) \sum_{i=0}^{L} f^i \gamma^t(i) \frac{1}{\sum_{\tau=0}^{n-1} f^\tau \chi(\tau)} \\ &+ \mu_\infty^2 \sum_{i,j=1}^{n} f^{n-i} f^{n-j} \mathrm{E}[\Psi_e(n,i)\Psi_e^t(n,j)]\end{aligned}$$

where the bias in the estimates has been considered:

$$\mathrm{E}[\Delta w(\infty)] = \frac{1}{\sum_{\tau=0}^{n-1} f^\tau \chi(\tau)} \sum_{i=0}^{L} f^i \gamma(i) \quad (23)$$

This expression for the bias can be obtained from expression (21), under IA assuming that the involved variables are uncorrelated if they are more than L time instants apart.

In previous expressions these definitions are considered under stationary statistics:

$\mathrm{E}[x(n)\tilde{x}^t(n)\tilde{x}(i)e_0(i)] \equiv \mathrm{E}[\Psi_e(n,i)] \equiv \gamma(n-i)$
$\chi(n-i) \equiv \mathrm{E}[x(n)\tilde{x}^t(n)\tilde{x}(i)x^t(i)]$

From a particular solution of (22), it can be shown that the Z-transform of $\mathrm{MSE}(n)$ is:

$$\mathrm{MSE}(Z) = \mathrm{Tr}\left\{\frac{1}{\mathrm{I} - (\mathrm{I} - 2\mu_\infty \chi(0))z^{-1} + 2\mu_\infty f\chi(0)z^{-2}} Nh(Z)\right\}$$

This expression is always valid provided that the zeros of the denominator lie inside the unit circle $\forall\, \lambda_i^\chi\ i = 1...p$, λ_i^χ is the i-th eigenvalue of $\chi(0)$. Given that $\lambda_i^\chi > 0\ \forall\, i$, because $\chi(0)$ is positive definite, we can always find a μ that makes the zeros lie inside the unit circle, -i.e. there is no convergence problem for the AOGLMS algorithm.

The noise terms that finally form the MSE are:

1. The noise coming from the bias (23).
2. Another term appearing due to the fact that although we are imposing $\mathrm{E}[\tilde{x}(n)e_0(n)] = 0$, its higher-order moments, -i.e. $\mathrm{E}[\tilde{x}(n)e_0(n)\tilde{x}^t(n)e_0(n)] \neq 0$, do not have to vanish.

5. COMPARISON WITH EXISTING ALGORITHMS

The novel AOGLMS algorithm will be applied to the blind identification of moving average (MA) systems following the paper by Giannakis and Mendel [6]. The MA process is defined by:

$$y(n) = \sum_{i=0}^{q} b_i s(n-i) + v(n)$$

The driving process $s(n)$ is independent and identically distributed, with one-sided exponential probability density function of zero mean, unit variance and skewness 2. Additive coloured Gaussian noise $v(n)$ is also present.

The AOGLMS algorithm will be compared with similar existing algorithms, already mention in this contribution:

1. The ORIV algorithm
2. The LMS/ORIV algorithm
3. The GLMS algorithm.

Two models are studied here, where the vector of coefficients $b = [b_0\ b_1\ ...\ b_q]$ is given by:

- Model A= $[1\ -0.8]$.
- Model B= $[1\ -1.1314\ 0.6400]$.

The learning curve for model A is shown in figure 1, the convergence time of all the algorithms has been set to approximately 30000 iterations. As expected the best results are given by ORIV, followed closely by LMS/ORIV and AOGLMS. It can be proven that the GLMS can not convergence for this model because the associated matrix R is not positive definite nor negative definite. This problem is not encountered in model B, as can be seen in figure 2. For model B the convergence time is set to 40000 iterations. The best results are again provided by ORIV, followed by LMS/ORIV, AOGLMS and GLMS respectively. A high SNR benefits the ORIV algorithm.

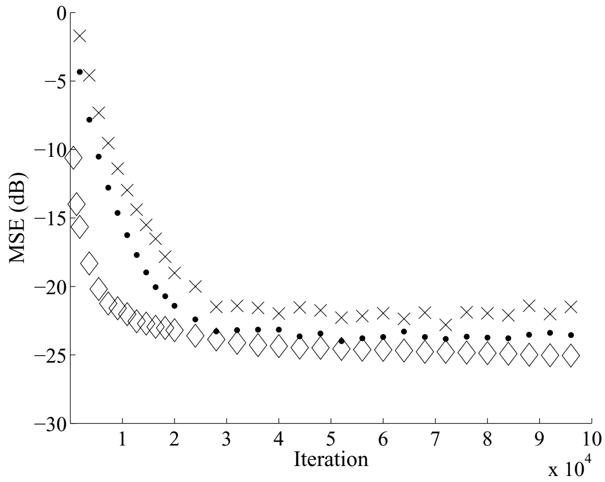

Figure 1: Learning curve for model A and SNR= 10 dB. ◇ ORIV, · LMS/ORIV with $\mu = 10^{-4.5}$ and × AOGLMS with $\mu = 10^{-4.40}$ and $f = 0.999$.

6. CONCLUSIONS

In this work a new orthogonality condition valid for overdetermined and generalised problems has been presented, and has been used to derive a new LMS-type algorithm, the AOGLMS, that can be applied to the same problem that the ORIV and LMS/ORIV algorithms. The convergence of the new algorithm has been studied theoretically under quite general assumptions. Simulation results show a behaviour of the novel AOGLMS close to the behaviour of ORIV and LMS/ORIV algorithms but with the further advantage of a much reduced computational burden.

REFERENCES

[1] B. Widrow and M.E. Hoff, "Adaptive switching circuits," *IRE WESCON Conv. Rec.*, vol. 4, pp. 96–104, 1960.

[2] A. Swami, "Fast transversal version of the RIV algorithm," *Int. Journ. Adaptive Control and Signal Processing*, vol. 10, pp. 267–281, 1996.

[3] B. Friedlander, "The overdetermined recursive instrumental variable method," *IEEE Trans. Automat. Contr.*, vol. Ac-29, no. 4, pp. 353–356, 1984.

[4] D. Aboutajdine, A. Adib, and A. Meziane, "Fast adaptive algorithms for AR parameters estimation using higher-order statistics," *IEEE Trans. Signal Processing*, vol. 44, no. 8, August 1996.

[5] Jitendra K. Tugnait, "A globally convergent adaptive blind equalizer based on second and fourth order statistics," in *Proceedings of ICC92*, 1992, pp. 1508–1512.

[6] G.B. Giannakis and J.M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 3, pp. 360–377, 1989.

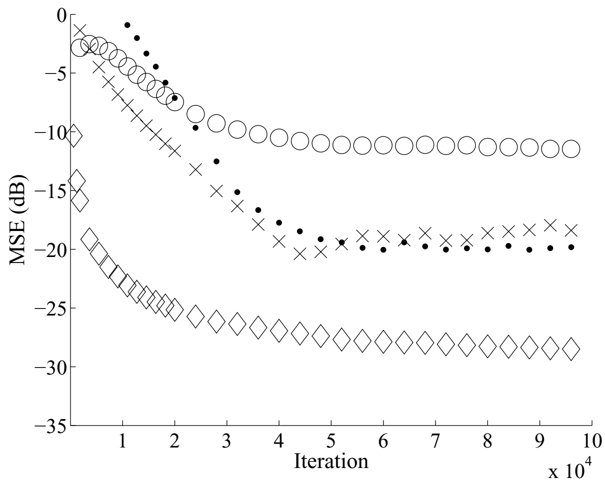

Figure 2: Learning curve for model B and SNR= 20 dB. ◇ ORIV, · LMS/ORIV with $\mu = 10^{-4.25}$, ∘ GLMS with $\mu = 10^{-3.5}$ and × AOGLMS with $\mu = 10^{-4.5}$ and $f = 0.999$.