SIGNAL PROCESSING V: Theories and Applications
L. Torres, E. Masgrau, and M.A. Lagunas (eds.)

SECOND-ORDER STATISTICAL ANALYSIS OF TWO CONSTRAINED LMS ALGORITHMS

J.C. PESQUET, O. MACCHI and G. TZIRITAS

Laboratoire des Signaux et Systèmes
ESE, Plateau de Moulon, 91192 Gif sur Yvette Cédex, France

This paper investigates the first and second-order statistical properties of two constrained LMS algorithms, the usual Leakage Factor (LF) algorithm and a new one called the Stabilizing Factor (SF) algorithm. They have proved to be especially useful in some specific situations where the LMS algorithm misbehaves.
It is shown that the weight estimates are biased. Then, we prove that convergence speed is increased by the LF algorithm while it is unmodified by the SF one. Moreover, lower and upper bounds of the variance of estimates are given. A substancial reduction of this variance may be reached compared with the LMS case. Finally, a bias-variance trade-off is achieved which is especially interesting when the signal to noise ratio is low and/or an *a priori* knowledge is available about the weigths to be identified.

1. INTRODUCTION

The LMS algorithm is a well-known method to adapt transversal filters [1]. It allows to estimate a discrete signal $y(n)$ on the basis of an N dimensional vector $\mathbf{X}(n)$ according to:

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \mu(y(n) - \mathbf{H}(n)^T\mathbf{X}(n))\mathbf{X}(n),\ \mu > 0 \tag{1.1}$$

where $\mathbf{H}(n)$ is the vector of weights of the transversal filter.

Nevertheless, this algorithm is known to lead to misbehaviours in some specific situations. Such a problem is reported in [2], concerning the digital implementation of a fractionally spaced equalizer. The correlation matrix of the observation being ill-conditioned, the estimation becomes strongly sensitive to round-off errors. Another example is Adaptive Differential Pulse Code Modulation (ADPCM), when a backward updating of the predictor is used. It was evidenced, both for speech [3] and for images [4], that the LMS algorithm may yield the instability of the decoder prediction filter, in the presence of transmission errors.

A possible solution to these problems is to introduce a regularization term into the criterion to be optimized. This leads to a compromise between a minimum output error and a smoothing constraint property specific to the problem. According to the form of the constraint, different kinds of soft-constrained gradient algorithms are obtained. In this paper, two of them are considered.

The updating equation of the first one is:

$$\mathbf{H}_L(n+1) = (1-\mu\alpha)\,\mathbf{H}_L(n) + \mu(y(n) - \mathbf{H}_L(n)^T\mathbf{X}(n))\mathbf{X}(n)\,,\ \alpha > 0\,. \tag{1.2}$$

This form corresponds to the usual Leakage Factor algorithm and will be denoted LF. A few results about this algorithm can be found in [5] and [6].

The second algorithm will be called the LMS with a Stabilizing Factor (SF) and is defined by:

$$\mathbf{H}_S(n+1) = \mathbf{H}_S(n) + \mu[y(n) - (1+\beta)\mathbf{H}_S(n)^T\mathbf{X}(n)]\mathbf{X}(n)\,,\ \beta > 0\,. \tag{1.3}$$

The demonstration was made that this algorithm improves the robustness of ADPCM systems [7].

Subsequently, the properties of these algorithm are analyzed in the stationary case and a parallel is drawn between them. The eigenvalues of the correlation matrix $\mathfrak{R} \triangleq E\{\mathbf{X}(n)\mathbf{X}(n)^T\}$ are denoted $(\lambda_i)_{1\le i\le N}$, the minimum and maximum values corresponding respectively to indexes m and M. It is further assumed that $\mathfrak{R}$ is non singular but λ_m can be chosen as small as desired.

2. RELATIONSHIPS OF THE SOFT-CONSTRAINED ALGORITHMS WITH THE LMS ALGORITHM

The LMS algorithm is the stochastic gradient technique to minimize the mean square error involved in linear estimation:

$$J(\mathbf{H}) = E\{(y(n) - \mathbf{H}^T\mathbf{X}(n))^2\}\ . \tag{2.1}$$

The soft-constrained algorithms can be obtained in the same way from modified criteria. For the LF algorithm, the criterion of minimization is generalized as follows:

$$J_L(\mathbf{H}) = J(\mathbf{H}) + \alpha\mathbf{H}^T\mathbf{H} \tag{2.2}$$

and, for the SF algorithm, it becomes

$$J_S(\mathbf{H}) = J(\mathbf{H}) + \beta \mathbf{H}^T \mathrm{E}\{\mathbf{X}(n)\,\mathbf{X}(n)^T\}\mathbf{H}\ . \qquad (2.3)$$

The additional term in (2.2) (resp. (2.3)) is a regularizing constraint. This means that it becomes of primary importance in the criterion when $\|\mathbf{H}\|$ (resp. $\mathbf{H}^T\mathbf{X}(n)$) takes too large values and therefore it will favour "reasonable" estimations.

Moreover, it is useful to note that the SF algorithm is directly related to the LMS algorithm. Indeed, if modified vector of observations and adaptation step-size are defined according to

$$\mathbf{X}'(n) = (1+\beta)\mathbf{X}(n) \qquad (2.4a)$$

$$\mu' = \frac{\mu}{1+\beta} \qquad (2.4b)$$

the SF algorithm reduces to the LMS algorithm.

3. FIRST-ORDER STATISTICAL ANALYSIS

3.1. Assumption

Subsequently, it is considered that

(H1) $\mathbf{H}(n)$ *and* $\mathbf{X}(n)$ *are independent..*

From a physical point of view, this hypothesis is not necessarily valid in all applications. Nevertheless, sophisticated mathematical studies such as ODE methods or statistical analysis [8] have shown that it is a good approximation, especially if μ is small.

3.2. LF algorithm

Taking the expectation of each side of (1.2) and using (H1) leads to:

$$\mathrm{E}\{\mathbf{H}_L(n+1)\} = [\mathbf{I} - \mu(\alpha\mathbf{I} + \mathfrak{R})]\,\mathrm{E}\{\mathbf{H}_L(n)\} + \mu\mathfrak{R}\tilde{\mathbf{H}} \qquad (3.2.1)$$

where $\tilde{\mathbf{H}}$ is the best mean square linear estimate of the weights.
Then, if

$$\mu < \frac{2}{\alpha + \lambda_M} \qquad (3.2.2)$$

$\mathrm{E}\{\mathbf{H}_L(n)\}$ converges exponentially toward

$$\tilde{\mathbf{H}}_L = \mathfrak{R}(\alpha\mathbf{I} + \mathfrak{R})^{-1}\tilde{\mathbf{H}}\ . \qquad (3.2.3)$$

Therefore, the price to be paid for the introduction of the leakage factor is a bias on the estimates of weights.
Equation (3.2.1) shows also that the convergence factor associated to the i^{th} eigenvector of the correlation matrix is

$$\tau_L(i) = 1 - \mu(\alpha + \lambda_i)\ . \qquad (3.2.4)$$

Thus, when $\mathfrak{R}$ is ill-conditioned and the m^{th} eigenvector corresponds to a near singular direction ($\lambda_m \cong 0$), a minimum convergence factor $1 - \mu\alpha$ is secured. On the contrary, the LMS algorithm does not allow to forget the initial values in the singular directions.

3.3. SF algorithm

Owing to the modifications (2.4), it is easily shown that if

$$\mu < \frac{2}{(1+\beta)\lambda_M} \qquad (3.3.1)$$

$\mathrm{E}\{\mathbf{H}_S(n)\}$ converges toward the biased estimate

$$\tilde{\mathbf{H}}_S = (1+\beta)^{-1}\tilde{\mathbf{H}}\ . \qquad (3.3.2)$$

The convergence factors of the algorithm are

$$\tau_S(i) = 1 - \mu(1+\beta)\lambda_i\ . \qquad (3.3.3)$$

As the influence of β upon λ_i is multiplicative, $\tau_S(i)$ remains close to 1 when λ_i is negligible and the behaviour of the algorithm is comparable to the LMS behaviour.

4. SECOND-ORDER STATISTICAL ANALYSIS OF THE LF ALGORITHM

4.1. Framework

Let us define the deviation vector $\mathbf{V}_L(n) \triangleq \mathbf{H}_L(n) - \tilde{\mathbf{H}}_L$. It comes from (1.2) that

$$\mathbf{V}_L(n+1) = [\mathbf{I} - \mu(\alpha\mathbf{I} + \mathbf{X}(n)\mathbf{X}(n)^T)]\mathbf{V}_L(n) + \mu b(n)\mathbf{X}(n) - \mu\mathbf{C}(n) \qquad (4.1.1a)$$

$$\mathbf{C}(n) = (\mathbf{X}(n)\mathbf{X}(n)^T - \mathfrak{R})(\tilde{\mathbf{H}}_L - \tilde{\mathbf{H}}) \qquad (4.1.1b)$$

where $b(n) \triangleq y(n) - \tilde{\mathbf{H}}^T\mathbf{X}(n)$.
This Section is devoted to the study of $\mathfrak{R}_L(n) \triangleq \mathrm{E}\{\mathbf{V}_L(n)\mathbf{V}_L(n)^T\}$. To this purpose, the following assumptions are made:
(H2) $\mathbf{X}(n)$ *is a zero-mean Gaussian vector ;*
(H3) $b(n)$ *is zero-mean with variance* σ_b^2, *i.i.d. and independent of* $\mathbf{X}(n)$.
Hypothesis (H2) is convenient to take into account the fourth-order statistics of the observation. In the same time, (H3) helps simplifying the analysis without being very restrictive in practice.

4.2. Transient behaviour

The transient part $\mathbf{V}_L^0(n)$ of $\mathbf{V}_L(n)$ may be defined by

$$\mathbf{V}^0_{\rm L}(0) = \mathbf{V}_{\rm L}(0) \quad (4.2.1a)$$

$$\mathbf{V}^0_{\rm L}(n+1) = [\mathbf{I} - \mu(\alpha\mathbf{I} + \mathbf{X}(n)\mathbf{X}(n)^{\rm T})]\mathbf{V}^0_{\rm L}(n) \quad (4.2.1b)$$

Combining (H1), (H2) and (4.2.1) allows to find the recurrence satisfied by $\mathcal{R}^0_{\rm L}(n) \triangleq {\rm E}\{\mathbf{V}^0_{\rm L}(n)\mathbf{V}^0_{\rm L}(n)^{\rm T}\}$. It is then straightforward to show that

$$\theta_{\rm L}(m)\,{\rm tr}\,\mathcal{R}^0_{\rm L}(n) \le {\rm tr}\,\mathcal{R}^0_{\rm L}(n+1) \le \theta_{\rm L}(M)\,{\rm tr}\,\mathcal{R}^0_{\rm L}(n) \quad (4.2.2)$$

with $\theta_{\rm L}(i) \triangleq \mu^2\lambda_i({\rm tr}\mathcal{R} + \lambda_i) + [1 - \mu(\alpha + \lambda_i)]^2$.
The sequences bounding ${\rm tr}\mathcal{R}^0_{\rm L}(n)$ are exponentially vanishing if

$$\mu < \frac{2(\alpha + \lambda_M)}{\lambda_M({\rm tr}\,\mathcal{R} + \lambda_M) + (\alpha + \lambda_M)^2} \,. \quad (4.2.3)$$

This condition is more restrictive than (3.2.2) and is close to the convergence condition as it is seen from reference [9] in the case of the LMS algorithm ($\alpha = 0$). It must be emphasized that the leakage factor narrows the interval into which μ can be chosen.

Moreover, when $\mu = [\lambda_i({\rm tr}\,\mathcal{R} + \lambda_i) + (\alpha + \lambda_i)^2]^{-1}(\alpha + \lambda_i)$, $\theta_{\rm L}(i)$ reaches a minimum value which is a decreasing function of α. Therefore, the leakage factor allows to improve the convergence speed of the algorithm.

4.3. Steady state performances

Finding the asymptotic covariance matrix $\mathcal{R}_{\rm L}$ of $\mathbf{V}_{\rm L}(n)$, when n tends to $+\infty$, is a more difficult problem. By using (H1) and (H2), (4.1.1) leads to

$$(\alpha\mathbf{I} + \mathcal{R})\mathcal{R}_{\rm L} + \mathcal{R}_{\rm L}(\alpha\mathbf{I} + \mathcal{R}) - \mu[{\rm tr}(\mathcal{R}\mathcal{R}_{\rm L})\mathcal{R} + \mathcal{R}\mathcal{R}_{\rm L}(\alpha\mathbf{I} + \mathcal{R}) + (\alpha\mathbf{I} + \mathcal{R})\mathcal{R}_{\rm L}\mathcal{R} + \alpha^2\mathcal{R}_{\rm L}] = \mu\,(\sigma_b^2\mathcal{R} + \mathcal{R}_{\rm C}). \quad (4.3.1)$$

with $\mathcal{R}_{\rm C} \triangleq {\rm E}\{\mathbf{C}(n)\mathbf{C}(n)^{\rm T}\}$. In the Karhunen-Loève basis where $\mathcal{R}$ is transformed in a diagonal matrix, the above linear equation takes a simpler form and the exact expression of ${\rm tr}\,\mathcal{R}_{\rm L}$ may be evaluated. Unfortunately, this expression is not easily tractable and it is useful to find lower and upper bounds. After some calculations, it may be proved that

$$v_{\rm L}(m) \le {\rm tr}\,\mathcal{R}_{\rm L} \le v_{\rm L}(M) \quad (4.3.2a)$$

where

$$v_{\rm L}(i) \triangleq v^b_{\rm L}(m) + v^{\rm C}_{\rm L}(m) \quad (4.3.2b)$$

$$v^b_{\rm L}(i) \triangleq \frac{\mu N\lambda_i\sigma_b^2}{2(\alpha + \lambda_i) - \mu[\lambda_i({\rm tr}\,\mathcal{R} + \lambda_i) + (\alpha + \lambda_i)^2]} \quad (4.3.2c)$$

$$v^{\rm C}_{\rm L}(i) \triangleq \frac{\mu\alpha^2(N+1)\lambda_i^2\|\tilde{\mathbf{H}}\|^2}{(\alpha + \lambda_m)(\alpha + \lambda_M)\{2(\alpha + \lambda_i) - \mu[\lambda_i({\rm tr}\,\mathcal{R} + \lambda_i) + (\alpha + \lambda_i)^2]\}} \,. \quad (4.3.2d)$$

Two points are worth being noted about these inequalities:
(i) if $\alpha = 0$, the inequalities reduce to results consistent with those already known for the LMS algorithm [9] ;
(ii) if $\mathcal{R} = \sigma^2\mathbf{I}$, the inequalities obviously become equalities.

5. SECOND-ORDER STATISTICAL ANALYSIS OF THE SF ALGORITHM

5.1. Framework

The deviation vector to be considered is now $\mathbf{V}_{\rm S}(n) \triangleq \mathbf{H}_{\rm S}(n) - \tilde{\mathbf{H}}_{\rm S}$. According to Section 2, the study of $\mathcal{R}_{\rm S}(n) \triangleq {\rm E}\{\mathbf{V}_{\rm S}(n)\mathbf{V}_{\rm S}(n)^{\rm T}\}$ under assumptions (H2) and (H3) may be derived from the results available for the LMS algorithm.

5.2. Transient behaviour

Owing to (4.2.3) and (2.4) the covariance matrix $\mathcal{R}^0_{\rm S}(n)$ of the transient part $\mathbf{V}^0_{\rm S}(n)$ of $\mathbf{V}_{\rm S}(n)$ satisfy

$$\theta_{\rm S}(m)\,{\rm tr}\,\mathcal{R}^0_{\rm S}(n) \le {\rm tr}\mathcal{R}^0_{\rm S}(n+1) \le \theta_{\rm S}(M)\,{\rm tr}\,\mathcal{R}^0_{\rm S}(n) \quad (5.2.1)$$

where $\theta_{\rm S}(i) \triangleq \mu^2(1+\beta)^2\lambda_i({\rm tr}\mathcal{R} + \lambda_i) + [1 - \mu(1+\beta)\lambda_i]^2$.
The condition of convergence is given by

$$\mu < \frac{2}{(1+\beta)({\rm tr}\,\mathcal{R} + 2\lambda_M)} \quad (5.2.2)$$

and, when $\mu = [(1+\beta)({\rm tr}\,\mathcal{R} + 2\lambda_M)]^{-1}$, $\theta_{\rm S}(i)$ reaches a minimum which is independent of β. This means that the SF algorithm is of no use to increase the convergence speed.

5.3 Steady state performances

According to (4.3.2) and (2.4), the asymptotic covariance matrix $\mathcal{R}_{\rm S}$ of $\mathbf{V}_{\rm S}(n)$ is such that

$$v_{\rm S}(m) \le {\rm tr}\,\mathcal{R}_{\rm S} \le v_{\rm S}(M) \quad (5.3.1a)$$

where

$$v_{\rm S}(i) \triangleq \frac{\mu N\sigma_b^2}{(1+\beta)[2 - \mu(1+\beta)({\rm tr}\,\mathcal{R} + 2\lambda_i)]} \,. \quad (5.3.1b)$$

6. BIAS-VARIANCE TRADE-OFF

6.1. SF algorithm

This case is first studied as it is easier. The overall mean square error of the algorithm is then asymptotically $\|\tilde{\mathbf{H}}_{\rm S} - \tilde{\mathbf{H}}\|^2 + {\rm tr}\,\mathcal{R}_{\rm S}$.

According to (3.3.2), the bias term is given by

$$\|\tilde{H}_S - \tilde{H}\|^2 \triangleq b_S = [\beta(1+\beta)^{-1}]^2 \|\tilde{H}\|^2 \ . \qquad (6.1.1)$$

Hence, it increases with β. In the same time, the expression (5.3.1b) shows that $v_S(i)$ decays for lower values of β, the minimum corresponding to a division of the variance of the LMS algorithm by a factor close to $[2\mu(\mathrm{tr}\mathbf{R} + 2\lambda_i)]^{-1}$

Furthermore, it is straightforward to prove that $b_S(\beta) + v_S(i)$ is minimum for a value $\beta(i)$ of the stabilizing factor which achieves the bias-variance trade-off. For instance, if the adaptation step-size is chosen in order to maximize the convergence speed (see Section 5.2), then $\beta(i) = [(\mathrm{tr}\mathbf{R} + 2\lambda_i)\|\tilde{H}\|^2]^{-1} N\sigma_b^2$ and the mean square error is decreased by a factor $1 + \beta(i)$ compared with the LMS algorithm. The improvement appears important mainly when the signal to noise ratio is low.

6.2. LF algorithm

The overall mean square error is now $\|\tilde{H}_L - \tilde{H}\|^2 + \mathrm{tr}\,\mathbf{R}_L$. According to (3.2.3), we have

$$b_L(M) \le \|\tilde{H}_L - \tilde{H}\|^2 \le b_L(m) \qquad (6.2.1)$$

where $b_L(i) = [\alpha(\lambda_i + \alpha)^{-1}]^2 \|\tilde{H}\|^2$. This last quantity is also an increasing function of α.

Since $v_L(i)$ has a complicated form, futher informations can be hardly found from an analytical study. Nevertheless computer simulations of this function shows that behaviours similar to the SF ones are obtained. In particular, an interesting bias-variance trade-off is reached, when the signal to noise ratio is low.

7. USE OF AN *A PRIORI* KNOWLEDGE

Sometimes, an *a priori* knowledge $\mathbf{H}_0$ is available concerning the weights which must be identified. This knowledge may be introduced in the algorithm to assess better performances. If criteria (2.2) and (2.3) are modified in the following way

$$J_L'(\mathbf{H}) = J(\mathbf{H}) + \alpha(\mathbf{H} - \mathbf{H}_0)^T(\mathbf{H} - \mathbf{H}_0) \qquad (7.1a)$$

$$J_S'(\mathbf{H}) = J(\mathbf{H}) + \beta(\mathbf{H} - \mathbf{H}_0)^T E\{\mathbf{X}(n)\mathbf{X}(n)^T\}(\mathbf{H} - \mathbf{H}_0) \qquad (7.1b)$$

the algorithms become

$$\mathbf{H}_L'(n+1) = (1 - \mu\alpha)\,\mathbf{H}_L'(n) + \mu(y(n) - \mathbf{H}_L'(n)^T\mathbf{X}(n))\mathbf{X}(n) + \mu\alpha\mathbf{H}_0 \qquad (7.2a)$$

$$\mathbf{H}_S'(n+1) = \mathbf{H}_S'(n) + \mu[y(n) - (1+\beta)\mathbf{H}_S'(n)^T\mathbf{X}(n) + \beta\,\mathbf{H}_0^T\mathbf{X}(n)]\mathbf{X}(n)\ . \qquad (7.2b)$$

Then, the limits of the average estimated weights are

$$\tilde{\mathbf{H}}_L' = \mathbf{R}(\alpha\mathbf{I} + \mathbf{R})^{-1}\tilde{\mathbf{H}} + \alpha(\alpha\mathbf{I} + \mathbf{R})^{-1}\mathbf{H}_0 \qquad (7.3a)$$

$$\tilde{\mathbf{H}}_S' = (1+\beta)^{-1}\tilde{\mathbf{H}} + \beta(1+\beta)^{-1}\mathbf{H}_0\ . \qquad (7.3b)$$

Therefore, α and β may be interpreted as weighting factors between the *a priori* knowledge and the best mean square estimate. It is easily checked that the results of the previous Section are still valid if $\|\tilde{\mathbf{H}}\|^2$ is replaced by $\|\tilde{\mathbf{H}} - \mathbf{H}_0\|^2$. Hence a good choice of $\mathbf{H}_0$ such that $\|\tilde{\mathbf{H}} - \mathbf{H}_0\| < \|\tilde{\mathbf{H}}\|$, is beneficial to the bias-variance trade-off.

8. CONCLUSIONS

In this paper, we have studied the LF and SF algorithms which are two constrained LMS algorithms. We have shown that the LF algorithm allows to improve the transient behaviour while the SF algorithm has no effect on it. We have evidenced that these methods lead to biased estimates of the weights but that, in the same time, the variance of the estimation may be decreased. Then, it is possible to achieve a bias-variance trade-off which is especially interesting when the signal to noise ratio is low and/or an *a priori* knowledge is available about the weights.

Note that the above investigations do not take into account the better stability properties of the SF algorithm [7].

REFERENCES

[1] B. Widrow and S.D. Stearns, "Adaptive Signal Processing", Prentice-Hall, pp. 99-116, 1985.

[2] R.D. Gitlin, H.C. Meadors, J.R. and S.B. Weinstein, "The tap-leakage algorithm: an algorithm for the stable operation of a digitally implemented fractionally spaced adaptive equalizer", Bell Syst. Tech. J., Vol. 61, No. 8, pp. 1817-1839, Oct. 1982.

[3] M. Bonnet, O. Macchi, M. Jaïdane-Saïdane, "Theoretical Analysis of the ADPCM CCITT Algorithm", IEEE Trans. Commun., June 1990.

[4] G. Tziritas and J.C. Pesquet, "A hybrid image coder: adaptive intra-interframe predictor using motion compensation", Proc. ICASSP 89, pp. 1878-1881, Glascow, May 1989.

[5] A. Segalen and G. Demoment, "Constrained LMS algorithm", Electron Lett., Vol. 18, No. 5, pp. 226-227, March 1982.

[6] M. Bellanger, "Adaptive digital filters and signal analysis", Dekker, pp. 112-114, 1987.

[7] J.C. Pesquet, G. Tziritas, O. Macchi, "Modified LMS algorithms for robust ADPCM", Proc. ICAASP90.

[8] O. Macchi and Eweda , "Second-order convergence analysis of stochastic adaptive linear filtering", IEEE Trans. Automat. Contr., Vol AC 28, No. 1, pp. 76-85, Jan. 1983.

[9] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data", IEEE Trans. Acoustics, Speech, Signal Processing, Vol. ASSP 33, No. 1,pp. 222-230, Feb. 1985.