ASYMPTOTIC ANALYSIS OF THE UNDERDETERMINED RECURSIVE LEAST-SQUARES ALGORITHM

B. Baykal, O. Tanrikulu and A. G. Constantinides
Sig. Proc. & Dig. Systs. Sec., Dept. of EE. Eng.,
Imperial College of Sci., Tech. and Med., London SW7 2BT, UK.
Email: b.baykal@ic.ac.uk, Fax: + 44 171 594 6234

ABSTRACT

The asymptotic analysis of the Underdetermined Recursive Least-Squares (URLS) algorithm is performed. In particular, the behaviour of the weight-error correlation matrix is investigated and the misadjustment is calculated. For highly correlated input signals the misadjustment is shown to be inversely proportional to the minimum eigenvalue of the underdetermined order autocorrelation matrix. Simulations are included to justify the conclusions.

1 INTRODUCTION

Adaptive filtering is a prominent research and application area in digital signal processing. It embodies a large number of techniques which have their own merits and drawbacks from different perspectives. The NLMS and RLS algorithms can be classified as two extreme cases from a linear algebra point of view. Both techniques employ inversions (inverse of the sample autocorrelation matrix in the RLS algorithm, and the inverse of the norm square of the input vector) which play decisive role in the overall performance of the algorithms. Specifically, it can be observed that the NLMS algorithm solves only one equation with respect to the N unknowns, where N is the adaptive filter length. In the sliding window covariance formulation of the RLS algorithm, a constant number of equations L which is larger than or equal to the adaptive filter size, is used to estimate N adaptive filter coefficients. The other two types of the RLS formulations, namely Exponential Windowed (EW) and Growing Memory Covariance (GMC) RLS algorithms solve an increasing number of equations across time. Apart from the NLMS and RLS algorithms, the last option, $1 \leq L \leq N$, has previously received little exposure [1]. This particular case, namely the URLS algorithm is investigated in this paper. Similar to the NLMS and RLS algorithms, an autocorrelation matrix must be propagated in time, the order of which is underdetermined with respect to the filter order. The URLS family algorithms have an inherent whitening effect on the input signal which may be considered as "proportional to the number of equations solved L to estimate the unknown system" in which L is also referred as the *prediction order*. Thus, the URLS algorithm lends itself to applications in which one has a priori knowledge about the spectral characteristics of the input signal and/or high additive noise is present. A representative example is the acoustical echo cancellation problem in hands-free telephony. The absence of a handset permits the microphone to pick up the loudspeaker signal, which is retransmitted to the far-end speaker and perceived as echo. The length of an echo path may be in the order of 100 to 4000 samples depending on the environment. Noise can also be captured by microphones particularly in a noisy environment or if the gain of the microphone is high. In noisy measurements, the optimality of least-squares estimates is lost and the convergence of the algorithm must be controlled via a stepsize, which is exemplified by decreasing the stepsize of the NLMS algorithm used in echo cancellation part of the hands-free telephony equipment [2]. The properties of speech signals suggest that a prediction order of 10-12 yields sufficient decorrelation. Thus, using an RLS type algorithm would be a waste of resources if the echo path is several hundred taps long. Alternatively, the desirable properties of the URLS algorithm provide an attractive solution.

2 URLS ALGORITHM

A derivation of the URLS algorithm which utilizes the *principal of minimal disturbance* [3] can be found in [4]. The URLS algorithm is defined as

$$W_{N,m,N,k} = W_{N,m,N,k-1} + \mu \varepsilon^H_{m,m,N,k} U^{-1}_{m,N,k} X^H_{m,N,k}, \quad (1)$$

where $W_{N,m,N,k}$ is the tap weight vector, $X_{m,N,k}$ is the $N \times m$ input signal matrix, μ is the step-size, $\varepsilon^H_{m,m,N,k} \triangleq d^H_{m,k} - W_{N,m,N,k-1} X_{m,N,k}$ and $U_{m,N,k} \triangleq X^H_{m,N,k} X_{m,N,k}$. The complexity of this algorithm is $O(mN) + O(m^2)$.

3 MISADJUSTMENT OF THE URLS ALGORITHM

The weight error vector recursion of the URLS algorithm is

$$v_k = v_{k-1} + \mu \left(\epsilon_{opt} - v_{k-1} X_k\right) K^H_k \quad (2)$$

where $v_k \triangleq W_k - W_{opt}$ and ϵ_{opt} is assumed to be white, independent of the input and referred as the minimum attainable error in the Wiener solution. Regrouping the terms in (2) yields

$$v_k = v_{k-1} \left(I - \mu P_k\right) + \mu \epsilon_{opt} K^H_k. \quad (3)$$

The weight error correlation matrix is

$$\begin{aligned} v^H_k v_k &= \left(I - \mu P_k\right) v^H_{k-1} v_{k-1} \left(I - \mu P_k\right) \\ &+ \mu \left(I - \mu P_k\right) v^H_{k-1} \epsilon_{opt} K^H_k + \mu K_k \epsilon^H_{opt} v_{k-1} \left(I - \mu P_k\right) \\ &+ \mu^2 K_k \epsilon^H_{opt} \epsilon_{opt} K^H_k. \end{aligned} \quad (4)$$

When the simplifications are made we obtain

$$v_k^H v_k = v_{k-1}^H v_{k-1} - \mu P_k v_{k-1}^H v_{k-1} - \mu v_{k-1}^H v_{k-1} P_k + \mu^2 P_k v_{k-1}^H v_{k-1} P_k + \mu^2 K_k \epsilon_{opt}^H \epsilon_{opt} K_k^H + \underbrace{\mu (I - \mu P_k) v_{k-1}^H \epsilon_{opt} K_k^H}_{M_1} + \underbrace{\mu K_k \epsilon_{opt}^H v_{k-1} (I - \mu P_k)}_{M_2} \quad (5)$$

The contribution from the term $\mu^2 E\left\{P_k v_{k-1}^H v_{k-1} P_k\right\}$ is smaller than those of the first three terms on the right hand side of (5) when the step-size μ is small, hence, it can be neglected [5]. Also, as the number of iterations tends to infinity, $v_k^H v_k \approx v_{k-1}^H v_{k-1}$ and therefore

$$\mu\left[P_k v_{k-1}^H v_{k-1} + v_{k-1}^H v_{k-1} P_k\right] = \underbrace{\mu^2 K_k \epsilon_{opt}^H \epsilon_{opt} K_k^H}_{N} + M_1 + M_2. \quad (6)$$

which is a Lyapunov equation [6]. Under the assumption that P_k is strictly stable, the solution $Z \triangleq v_{k-1}^H v_{k-1}$ can be written as [6]

$$Z = \int_{t=0}^{\infty} \exp\left[P_k t\right]\left\{-\left(N + M_1 + M_2\right)\right\} \exp\left[P_k t\right] dt \quad (7)$$

In fact, P_k is not strictly stable because some of its eigenvalues are zero which arises from the fact that P_k is a projection matrix. Hence, an explicit form cannot be found without imposing further assumptions.

The matrices M_1 and M_2 include correlations of v_{k-1} and X_k, ϵ_{opt} and X_k computed as time averages. When the independence assumptions are invoked, i.e, X_k, ϵ_{opt} and v_{k-1} are assumed to be independent random variables, for large m and N, the contributions from M_1 and M_2 are approximately zero if the signals are assumed to be ergodic. If the expected values of both sides of (6) are evaluated, the contributions from M_1 and M_2 are identically zero due to independence assumptions. In this case, the assumption that m and N must be large is not needed. We can also show that $E\left\{v_{k-1}\right\} = 0$, i.e, $W_k \to W_0$ as $k \to \infty$.

Therefore, we can write the matrix Z in (6) as the solution of the following equation

$$\mu[P_k Z + Z P_k] = \mu^2 K_k \epsilon_{opt}^H \epsilon_{opt} K_k^H, \quad (8)$$

i.e, only the contribution from N is considered. The solution is given as

$$Z = \frac{\mu}{2} K_k \epsilon_{opt}^H \epsilon_{opt} K_k^H, \quad (9)$$

which is valid for large m and N values. When the expected values of both sides are evaluated

$$E\left\{Z\right\} = \frac{\mu}{2} E\left\{K_k \epsilon_{opt}^H \epsilon_{opt} K_k^H\right\}, \quad (10)$$

the asymptotical behaviour of the weight error correlation matrix is obtained, which should be valid for any m and N. The independence of ϵ_{opt} and the input leads to

$$E\left\{Z\right\} = \frac{\mu}{2} \sigma^2 E\left\{K_k K_k^H\right\} \quad (11)$$

where σ^2 is the minimum obtainable mean-square power. The misadjustment $\mathcal{M}$ is defined as

$$\mathcal{M} = \lim_{k \to \infty} \frac{E\left\{\left(v_k^H X_k\right)^2\right\}}{\sigma^2}, \quad (12)$$

which can be rewritten as

$$\mathcal{M} = \frac{\mathrm{tr}\left[\breve{U}_N E\left\{Z\right\}\right]}{\sigma^2} \quad (13)$$

by invoking the independence assumption [7], where $\breve{U}_N$ is the expected value of the N-th order autocorrelation matrix of the input.

The general misadjustment expression in (13) is not very informative which motivates the development of an approximate expression. Let us return to (8). Evaluation of expected values of both sides yields

$$\mu[E\left\{P_k\right\} E\left\{Z\right\} + E\left\{Z\right\} E\left\{P_k\right\}] = \mu^2 \sigma^2 E\left\{K_k K_k^H\right\}, \quad (14)$$

where σ^2 is the minimum obtainable mean-square power. We have also imposed the fundamental independence assumption. When P_k and K_k are written in terms of the input, we have

$$E\left\{X_k U_k^{-1} X_k^H\right\} E\left\{Z\right\} + E\left\{Z\right\} E\left\{X_k U_k^{-1} X_k^H\right\} = \mu \sigma^2 E\left\{X_k U_k^{-2} X_k^H\right\}. \quad (15)$$

When $N \gg m$, the approximation $E\left\{X_k U_k^{-1} X_k^H\right\} \approx E\left\{X_k \left(E\left\{U_k\right\}\right)^{-1} X_k^H\right\}$ is valid [8]. $E\left\{U_k\right\} \approx N \breve{U}_m$ is also true, where $\breve{U}_m$ is the m-th order autocorrelation matrix of the input. Thus, (15) simplifies to

$$E\left\{X_k \breve{U}_m^{-1} X_k^H\right\} E\left\{Z\right\} + E\left\{Z\right\} E\left\{X_k \breve{U}_m^{-1} X_k^H\right\} = \mu \sigma^2 N^{-1} E\left\{X_k \breve{U}_m^{-2} X_k^H\right\}. \quad (16)$$

The relationship in (16) is a Lyapunov equation [6], the solution of which gives the asymptotical behaviour of the weight error correlation matrix Z as the number of iterations tends to infinity. Hence,

$$E\left\{X_k Q_m \Lambda_m^{-1} Q_m^H X_k^H\right\} E\left\{Z\right\} + E\left\{Z\right\} E\left\{X_k Q_m \Lambda_m^{-1} Q_m^H X_k^H\right\} = \mu \sigma^2 N^{-1} E\left\{X_k Q_m \Lambda_m^{-2} Q_m^H X_k^H\right\}, \quad (17)$$

where $\breve{U}_m = Q_m \Lambda_m Q_m^H$ so that the autocorrelation matrix $\breve{U}_m$ is decomposed into its eigenvalues Λ_m and eigenvectors Q_m. The only difference between the matrices in the expected value operations is that we have Λ_m^{-1} on the left hand side and Λ_m^{-2} on the right hand side. If they were approximated as scalar multiples of each other, we would have

$$AE\left\{Z\right\} + E\left\{Z\right\} A = \alpha A, \quad (18)$$

where α is a scalar and A is an approximate matrix obtained from $E\left\{X_k Q_m \Lambda_m^{-1} Q_m^H X_k^H\right\}$ and $E\left\{X_k Q_m \Lambda_m^{-2} Q_m^H X_k^H\right\}$. The solution of (18) is given as $Z = (\alpha/2) I$. Hence, if $E\left\{X_k Q_m \Lambda_m^{-1} Q_m^H X_k^H\right\}$ is approximated as a scalar multiple of $E\left\{X_k Q_m \Lambda_m^{-2} Q_m^H X_k^H\right\}$ or vice versa, we can obtain a tractable solution, i.e, we have to minimize

$$\left\| E\left\{X_k Q_m \Lambda_m^{-1} Q_m^H X_k^H\right\} - c E\left\{X_k Q_m \Lambda_m^{-2} Q_m^H X_k^H\right\} \right\|_2 \quad (19)$$

with respect to the scalar c. (19) is equivalent to

$$\text{Approximation I:} \qquad \text{Minimize} \quad \left\|\Lambda_m^{-1} - c \Lambda_m^{-2}\right\|_2. \quad (20)$$

Similar to the approach above, the other alternative is

$$\text{Approximation II:} \qquad \text{Minimize} \quad \left\| \Lambda_m^{-2} - c\Lambda_m^{-1} \right\|_2 . \quad (21)$$

Let λ and $\tilde{\lambda}$ be the column vectors which consist of the inverse eigenvalues and their squares respectively. Thus, the solutions of the approximation I in (20) and the approximation II in (21) are given as

$$c = \frac{\tilde{\lambda}^H \lambda}{\tilde{\lambda}^H \tilde{\lambda}}, \quad (22a)$$

$$e = \frac{\left[\sum_{i=1}^m \lambda_i^{-2}\right]\left[\sum_{i=1}^m \lambda_i^{-4}\right] - \left[\sum_{i=1}^m \lambda_i^{-3}\right]^2}{\sum_{i=1}^m \lambda_i^{-4}} \quad (22b)$$

and

$$c = \frac{\lambda^H \tilde{\lambda}}{\lambda^H \lambda}, \quad (23a)$$

$$e = \frac{\left[\sum_{i=1}^m \lambda_i^{-2}\right]\left[\sum_{i=1}^m \lambda_i^{-4}\right] - \left[\sum_{i=1}^m \lambda_i^{-3}\right]^2}{\sum_{i=1}^m \lambda_i^{-2}} \quad (23b)$$

respectively, where the associated approximation errors e are also shown. Hence, by comparing the denominators of the approximation errors, we can conclude that the approximation I is better if $\sum_{i=1}^m \lambda_i^{-4} > \sum_{i=1}^m \lambda_i^{-2}$ is satisfied. This condition is true for high-eigenvalue-spread autocorrelation matrices. Thus, we can safely assume that for highly correlated input, the approximation I yields better results with $c = \frac{\tilde{\lambda}^H \lambda}{\lambda^H \lambda}$. We have also illustrated this fact in Figure 1 where a representative example of a highly correlated input is chosen. The inverse eigenvalues of a 5-th order autocorrelation matrix are 0.5, 0.9, 1.8, 2, 5.8. In Figure 1(a) the axes are scaled so that a comparison can be made with Figure 1(b). Figure 1(c) is essentially same as Figure 1(a). Figures 1(b) and 1(c) correspond to approximations in (21) and (20) respectively. The sums of the lengths of the vertical lines from the inverse eigenvalues to the line or parabola represent the modelling errors. Clearly, the fitting of a quadratic to the eigenvalues results in better approximation, c.f, Figure 1(a) or 1(c).

Thus, by taking the approximation I in (20) into account, (17) becomes

$$\begin{aligned} & c\left[E\left\{X_k Q \Lambda^{-2} Q^H X_k^H\right\} E\{Z\}\right. \\ & \left. + E\{Z\} E\left\{X_k Q \Lambda^{-2} Q^H X_k^H\right\}\right] \\ & = \mu\sigma^2 N^{-1} E\left\{X_k Q \Lambda^{-2} Q^H X_k^H\right\}, \end{aligned} \quad (24)$$

the solution of which is given as, analogous to (18),

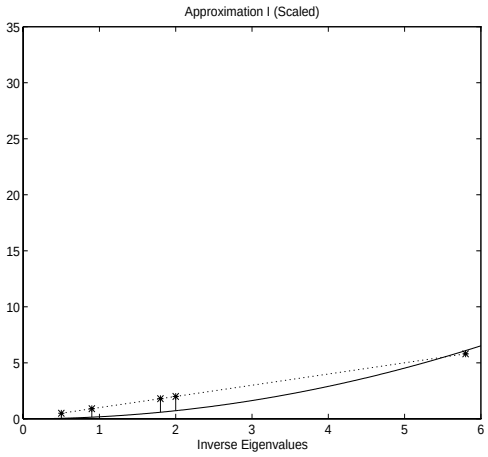
$$E\{Z\} = E\left\{v_{k-1}^H v_{k-1}\right\} \approx \frac{\mu\sigma^2 N^{-1}}{2c} I. \quad (25)$$

More explicitly, we have

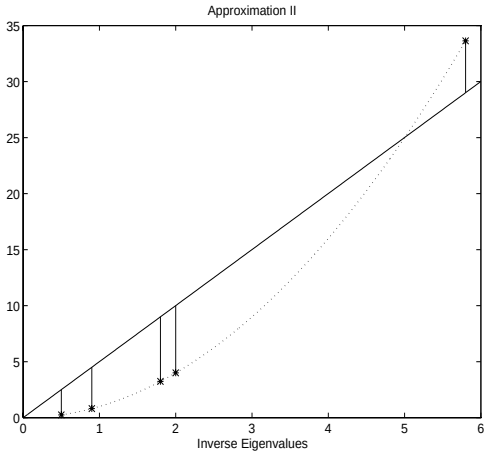
$$E\left\{v_{k-1}^H v_{k-1}\right\} \approx \frac{\mu\sigma^2 N^{-1} \sum_{i=1}^m \lambda_i^{-4}}{2\sum_{i=1}^m \lambda_i^{-3}} I, \quad (26)$$

which can be further approximated if the smallest eigenvalue is dominant, so that

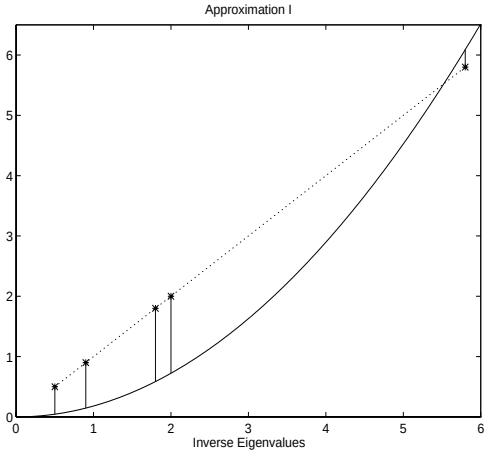
$$E\left\{v_{k-1}^H v_{k-1}\right\} \approx \frac{\mu\sigma^2 N^{-1} \lambda_{min}^{-1}}{2} I. \quad (27)$$



(a) Fitting of a parabola to a line (scaled).



(b) Fitting of a line to a parabola.



(c) Fitting of a parabola to a line.

Figure 1: Comparison of eigenvalue approximations for the misadjustment.

Hence, the weight error correlation matrix is approximately diagonal in the steady state. The misadjustment from (13) is

$$\mathcal{M} \approx \frac{1}{\sigma^2} \text{ tr} \left(\breve{U}_N \frac{\mu\sigma^2}{2N\lambda_{min}} I \right). \tag{28}$$

After some algebra, we can obtain

$$\mathcal{M} \approx \frac{\mu\breve{u}}{2\lambda_{min}} \tag{29}$$

where $\breve{u}$ is the diagonal element of the autocorrelation matrix, i.e, the power of the input signal.

It is observed that the misadjustment is inversely proportional to the minimum eigenvalue of the m-th order autocorrelation matrix which leads to the conclusion that the URLS algorithm produces higher misadjustment than the NLMS algorithm for the same step-size and adaptive filter length with highly correlated inputs. If the input signal is white, the URLS algorithms of all orders are equivalent to the NLMS algorithm.

4 SIMULATIONS

The above conclusions are experimentally verified by the following experiment. The pole locations of the autoregressive input are $\left[\begin{array}{ccc} -0.7 & 0.65 \mp 0.7j & -0.8 \mp 0.4j \end{array} \right]$ in which there are a couple of poles very close to the unit circle. Therefore, the autocorrelation matrix has high eigenvalue spread. The identification experiment is performed by the NLMS algorithm, and the URLS algorithms with $m = 2, \ldots, 5$. The input signal has unity power which gives $\breve{u} = 1$, and $\lambda_{min} = 1$ for the NLMS algorithm since the eigenvalue of the first order autocorrelation matrix is equal to the input signal power. The adaptive filter length is chosen as 96 and the noise power as 1. The stepsize is set to 0.01 to reduce the effects of the neglected terms in the analysis. The experimental misadjustment is calculated as the mean of the last 10,000 samples of the resulting Monte Carlo simulation of 20 runs and 500,000 samples. The theoretical misadjustment is calculated from (29). The results are presented in Table 1.

		Misadjustment $\mathcal{M}$	
	λ_{min}	Experimental	Theoretical
NLMS, $m = 1$	1.000000	0.005999	0.005000
URLS, $m = 2$	0.190023	0.023527	0.026312
URLS, $m = 3$	0.137765	0.031536	0.036293
URLS, $m = 4$	0.134287	0.032922	0.037233
URLS, $m = 5$	0.067878	0.041778	0.073661

Table 1: The misadjustments for various order URLS algorithms with $N = 96$, $\sigma^2 = 1$, $\mu = 0.01$, $\breve{u} = 1$.

The experimental results also show the trend of increase in the misadjustment as the order of the URLS algorithm is increased. The discrepancies between theoretical and experimental results can be ascribed mostly to the fact that the adaptive filter length may not be enough to approximate the true autocorrelation matrix.

The above theoretical result about the NLMS algorithm is in agreement with the analysis in [7]. More detailed techniques about the NLMS algorithm can be found in [9, 10].

Another promising way to analyze the URLS algorithm would be to assume the input signal is a sinusoid buried in wide-band noise so that the eigenvalue spread is high and the sinusoid leads to tractable expressions when projection matrices are involved. In fact, the analysis of the filtered-X algorithm in [11] takes the above assumption as the starting point.

5 CONCLUSIONS

The asymptotic convergence properties of the URLS algorithm are investigated. The weight error correlation matrix is shown to be diagonal at the steady-state. Two alternatives are examined one of which leads to an accurate misadjustment expression for highly correlated input signals. The theoretical findings are verified in comprehensive experiments.

References

[1] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol. 67-A, no. 5, pp. 19–27, 1984.

[2] J. C. Magan, Editor, "Basic methods for echo cancellation, deliverable 3.123.1," tech. rep., Commission of the European Communities (FREETEL no.6166), July 1993.

[3] B. Widrow and M. Lehr, "30 years of adaptive neural networks: Perceptron, madaline and backpropagation," *Proceedings of the IEEE, Special Issue on Neural Networks I*, vol. 78, 1990.

[4] B. Baykal, *Underdetermined Recursive-Least Squares Adaptive Filtering*. PhD thesis, Imperial College, London, July 1995.

[5] E. Walach and W. B., "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Transactions on Information Theory*, vol. IT-30, pp. 275–283, 1984.

[6] J. E. Slotine and W. Li, *Applied Nonlinear Control*. Prentice-Hall, 1991.

[7] S. Haykin, *Adaptive Filter Theory, 2nd Ed.* Prentice-Hall, 1991.

[8] D. Slock, "The block underdetermined covariance (BUC) fast transversal filter (FTF) algorithm for adaptive filtering," in *Proceedings of the ASILOMAR-92*, pp. 550–554, 1992.

[9] D. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions On Signal Processing*, vol. 41, no. 9, pp. 2811–2825, 1993.

[10] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated gaussian data," *IEEE Transactions On Information Theory*, vol. 34, no. 4, pp. 680–691, 1988.

[11] M. Mboup, M. Bonnet, and N. Bershad, "LMS coupled adaptive prediction and system identification: A statistical model and transient mean analysis," *IEEE Transactions On Signal Processing*, vol. 42, pp. 2607–2615, 1994.