LS DETECTION GUIDED NLMS ESTIMATION OF SPARSE SYSTEMS

John Homer

School of Info. Tech. & Elec. Engg.
Univ. of Queensland, Brisbane, Australia.
Email: homerj@itee.uq.edu.au

Iven Mareels

Dept. of Elec. & Electronic Engg.
Univ. of Melbourne, Melbourne, Australia.
Email: i.mareels@unimelb.edu.au

ABSTRACT

In various estimation problems the system being estimated may be represented by a sparse parameter vector, such that only a 'small' number of the vector elements are 'significant' or 'active'. In this paper we propose an NLMS estimator which incorporates a least squares based active parameter criterion; such that NLMS adaptation is applied only to those system parameters detected as being active. This results in a significant improvement in convergence rates, as compared to the standard NLMS estimator. Importantly, for sparse systems, the computational cost of the newly proposed detection guided NLMS estimator is only slightly greater than that of the standard NLMS estimator.

1. INTRODUCTION

We define a sparse system as one in which the system is represented by a finite parameter set; and many of the parameters within this set are 'insignificant' or 'inactive'. Sparsely parametrized systems exist in many applications, such as temporal and spatial acoustic echo paths within hands-free telephony [1]. Estimation of such sparsely parametrised systems, via the common parallel configuration of Figure 1, is often conducted using the popular normalised least mean square (NLMS) estimator [2]. The standard NLMS approach involves adaptation of each and every parameter during each sample interval. However, this approach is plagued by slow convergence rates when the system has a long parameter representation. Highly correlated input signals, such as the speech signals within communication systems, tend to further worsen the convergence rates [3].

An approach to combat these effects is to NLMS estimate only the 'active' parameters. The key to this approach lies in the use of a suitable activity criterion for accurately determining whether any particular parameter is active. Following on from the work of Homer *et. al.* [4], [5], we propose an activity criterion which is based on the minimisation of a structurally consistent least squares (SC-LS) cost function. Our proposed LS detection guided NLMS estimator applies the activity criterion during each sample interval to determine the active parameters, and subsequently updates the NLMS estimate only for those parameters. Simulations indicate that, for sparsely parametrised systems, the proposed estimator converges considerably more quickly than the standard NLMS estimator. Furthermore, for sparse systems, the computational cost of the proposed estimator is only slightly greater than that of the standard NLMS estimator.

Based on optimising (minimising) the asymptotic parameter estimation error, a parameter is 'significant' or 'active' ... and should be NLMS estimated ... if its magnitude lies above the NLMS adaptive noise level [4]. Otherwise, the parameter is 'insignificant' or 'inactive' and should not be NLMS estimated. Simulations indicate that the proposed NLMS estimator typically detects (and estimates) all active parameters, while it typically ignores (does not estimate) the inactive parameters. In summary, the proposed NLMS estimator, in comparison to the standard NLMS estimator provides significantly improved convergence rates, without compromising the asymptotic estimation performance, and with only slightly increased computational costs. Additional simulations (not reported here) indicate that the proposed estimator also has a fast tracking capability of time varying systems.

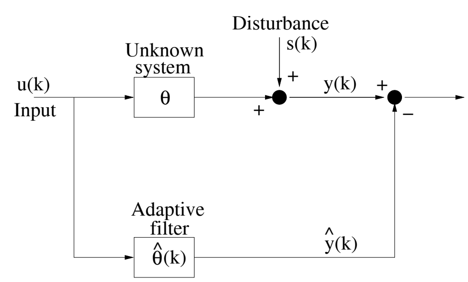

Fig. 1. System estimation configuration

2. SYSTEM ESTIMATION CONFIGURATION

The configuration we consider throughout this paper is shown in Figure 1. We assume that the unknown system is a MISO linear, time invariant system which is modeled by a parametric vector $\theta = \{\theta_j\}_{j=0}^{n-1} = [\theta_0, \theta_1, ..., \theta_{n-1}]^T$. We assume all signals are sampled. At sampling instant k: $U(k) = \{u_j(k)\}_{j=0}^{n-1}$ is the signal input vector to the unknown system and the estimator; an additive disturbance, $s(k)$, occurs at the output of the unknown system; and $y(k)$ is the observed output from the unknown system. The observed system output is given by: $y(k) = U(k)^T\theta + s(k)$. The output from the estimator is $\hat{y}(k) = U^T(k)\hat{\theta}(k)$ where $\hat{\theta} = [\hat{\theta}_0 \ \hat{\theta}_1 \ ... \ \hat{\theta}_{n-1}]^T$ is an estimate of the parameter vector θ. The adaptive NLMS estimator equation is:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{\mu}{U^T(k)U(k) + \epsilon}U(k)e(k), \quad (1)$$

where: $e(k) = y(k) - \hat{y}(k)$; μ and ϵ are small positive constants.

Remark1: The above system configuration, with the following change in notation $u_j(k) \rightarrow u(k-j)$, is also applicable to one-dimensional temporal signal-channels, such as temporal acoustic echo channels. Such channels are SISO systems.

In addition to the above we assume the following. (i) The elements of the input signal vector are samples of zero mean, bounded, wide sense stationary process(es) of variance σ_u^2. (ii) The disturbance signal is a zero mean, bounded, wide sense stationary white process of variance σ_s^2. (iii) The disturbance signal is uncorrelated with the input signal vector. (iv) The unknown system, θ, is sparsely active:

$$\theta = [\eta(M_1), \theta_{t1}, \eta(M_2), \theta_{t2}, ..., \theta_{tm}, \eta(M_m)]$$

where: $m \ll n$; $0 \leq t_1 < t_2 < ... < t_m \leq n-1$; $M_i = t_i - t_{i-1} - 1$; θ_{ti} =active parameter; $\eta(M_i)$ =vector containing M_i inactive parameters. We define an *active* parameter as a parameter with a magnitude greater than the NLMS adaptive noise level [4]: $|\theta_{ti}| > \sqrt{\mu\sigma_s^2/(\hat{m}\sigma_u^2)}$, where $\hat{m}$ is the estimate of m. Each of the remaining parameters is defined as an *inactive* parameter.

3. DETECTION GUIDED NLMS ESTIMATION

The activity criterion within our detection guided estimator is derived from the following structurally consistent least squares based cost function [4]:

$$V_{SCLS}(N) = V_{LS}(N) + m\sigma_y^2 logN \quad (2)$$

where: $V_{LS}(N) = \sum_{k=1}^{N}[y(k) - \hat{\theta}^T U(k)]^2$; σ_y^2 = variance of $y(k)$; m = unknown number of active parameters; $\hat{\theta}$ = estimate which contains only m *nonzero* parameters. In general, minimisation of $V_{SCLS}(N)$ requires examination and comparison of a very large number C_m^n of parameter sets: $C_m^n = \sum_{m=1}^{n} \frac{n!}{(n-m)!m!}$.

The development of the activity criterion begins by assuming the input signal vector has uncorrelated elements. Then for sufficiently large N we may approximate V_{SCLS} of (2) by [4]:

$$\tilde{V}_{SCLS} = \sum_{k=1}^{N} y^2(k) - \sum_{i=1}^{m}[X_{ti}(N) - \sigma_y^2 logN] \quad (3)$$

$$X_{ti}(N) = \frac{[\sum_{k=1}^{N} y(k)u_{ti}(k)]^2}{\sum_{k=1}^{N} u_{ti}^2(k)}. \quad (4)$$

It is apparent that $\tilde{V}_{SCLS}$ is minimised by (and hence the indices of the active parameters correspond to) those indices $t_i = j$ which satisfy:

$$X_j(N) > T(N) \quad (5)$$

$$\text{where } T(N) = \sigma_y^2 logN \approx \frac{logN}{N}\sum_{k=1}^{N} y^2(k).$$

Note, this relatively simple criterion for determining the active parameter indices stems from the parameters within $\tilde{V}_{SCLS}$ being decoupled.

The criterion of (5) is not applicable to the case in which the input signal vector elements are correlated. This is because the correlation causes coupling between the parameters within the $X_j(N)$ numerator term $\sum_{k=1}^{N} y(k)u_j(k)$. As such, $X_j(N)$ is dependent not only on parameter θ_j but also on the neighbouring parameters.

To reduce the coupling between neighbouring parameters, we propose the following three modifications.

Modification 1: Replace $X_j(N)$ by:

$$XX_j(N) = \frac{[\sum_{k=1}^{N}\{e(k) + \hat{\theta}_j(k)u_j(k)\}u_j(k)]^2}{\sum_{k=1}^{N} u_j(k)^2}. \quad (6)$$

Modification 2: Replace $T(N)$ by:

$$TT(N) = \frac{logN}{N}\sum_{k=1}^{N} e^2(k). \quad (7)$$

Modification 3: Include the exponentially forgetting operator: $W_N(k) = (1-\gamma)^{N-k}$, $0 < \gamma \ll 1$ within the summation terms of $XX_j(N)$ and $TT(N)$.

The purpose of Modification 1 is based on the following. The cause of parameter coupling in $X_j(N)$ arises from the following numerator term:

$$numX_j(N) \triangleq \frac{1}{N}\sum_{k=1}^{N} y(k)u_j(k)$$
$$= \frac{1}{N}[\sum_{i=1}^{N}\sum_{p \neq j}(\theta_p u_p(k)u_j(k)$$
$$+ \theta_j u_j(k)u_j(k) + s(k)u_j(k)].$$

The first component in the summation is the cause of parameter coupling. This becomes more significant with an increase in the correlation amongst the input signal elements.

The equivalent numerator term of $XX_j(N)$ is:

$$\begin{aligned} numXX_j(N) &\triangleq \frac{1}{N}[\sum_{k=1}^{N}\{e(k)+\hat{\theta}_j(k)u_j(k)\}u_j(k)] \\ &= \frac{1}{N}[\sum_{k=1}^{N}\sum_{p\neq j}(\theta_p-\hat{\theta}_p(k))u_p(k)u_j(p) \\ &+ \theta_j u_j(k)u_j(k)+s(k)u_j(k)]. \end{aligned}$$

Here the parameter coupling effect of the first term should be significantly weakened, assuming $\hat{\theta}_p(k)$ converges towards θ_p (for $p = 0, 1, ..., n-1$.)

Modification 2 stems from the realisation that for *inactive* taps (and assuming $\hat{\theta}_{inactive} \approx 0$) the numerator term $numXX_j(N)$ is approximately:

$$numXX_j(N) \approx \frac{1}{N}[\sum_{k=1}^{N} e(k)u_j(k)].$$

Consequently, combining this with the LS theory on which the original activity criterion (5) is based, then suggests this second proposed modification. This reasoning for Modification 2, however, is only relevant if the system estimation error vector $\Delta\theta(k) \triangleq \theta - \hat{\theta}(k)$ is non-time varying. Clearly, this is not the case. Modification 3 reduces the effect of the time varying nature of $\Delta\theta(k)$.

Remark2: The inclusion of Modification 3 also improves the applicability of the proposed detection guided estimator to time varying systems. This capability of the proposed estimator is not explored in this paper.

The proposed LS detection guided NLMS estimator is as follows.

For each parameter index j:

1. (a) Initialise $r_j(0) = \epsilon_1,\ 0 < \epsilon_1 \ll \sigma_u^2$. (b) Initialise: $p_j(0) = q(0) = a(0) = 0$.

2. At each sample interval N:

$$\begin{aligned} p_j(N) &= (1-\gamma)p_j(N-1) \\ &\quad +[e(N)+\hat{\theta}_j(N)u_j(N)]u_j(N) \\ r_j(N) &= (1-\gamma)r_j(N-1)+u_j^2(N) \\ q(N) &= (1-\gamma)q(N-1)+e^2(N) \\ a(N) &= (1-\gamma)a(N-1)+1 \\ XX_j(N) &= p_j^2(N)/r_j(N) \\ TT(N) &= q(N)log\{a(N)\}/a(N). \end{aligned}$$

3. If $XX_j(N) > TT(N)$ then label j as an active parameter index ti; otherwise label j as an inactive parameter index.

4. For each detected active parameter index update the NLMS estimate:

$$\hat{\theta}_j(N) = \hat{\theta}_j(N-1) + \frac{\mu e(k)u_j(N)}{\sum_{ti} u_{ti}^2(N)+\epsilon}$$

where $\sum_{ti}$ =summation over all detected active tap indices.

5. For each identified inactive parameter index reset the NLMS estimate to zero.

We measure computational complexity by the number of multiplications per sample interval (MPSI). The MPSI required depends on whether we are considering a MISO system or a temporal SISO system (See Remark1).

(i) For the MISO system, the standard NLMS estimator requires $3n+2$ MPSI. The proposed NLMS estimator requires $6n+2m+4$ MPSI.

(ii) For the temporal SISO system, the standard NLMS estimator requires $2n+3$ MPSI. The proposed NLMS estimator requires $4n+m+5$ MPSI.

(Note: The above MPSI numbers assume the values of $a(N)$ and $log\{a(N)\}/a(N)$ are available from a look-up table.) Hence, for sufficiently long sparse systems ($n \gg 1$, $n \gg m$), the computational cost of the proposed detection guided NLMS estimator is essentially twice that of the standard NLMS estimator.

4. SIMULATIONS

Simulations were conducted to compare the performance of the standard NLMS estimator with that of the newly proposed LS detection guided NLMS estimator. The systems considered were based on one-dimensional temporal signal-channels; such that θ corresponded to the impulse response vector of the signal-channel: $\theta \to \Theta(z^{-1}) = [\theta_0+\theta_1 z^{-1}+ ...+\theta_{n-1}z^{-n+1}]$. Accordingly, the system has a single input signal stream $u(k)$ and the input signal vector $U(k)$ corresponds to: $U(k) = [u(k),\ u(k-1),\ ...,\ u(k-n+1)]$.

The simulations involved the following: input signals $u(k)$ described by the correlated signal model:

$$u(k) = w(k)/[1-0.8z^{-1}],$$

where $w(k)$ is a discrete white zero mean unit variance Gaussian process; disturbance signals $s(k)$ described by a discrete white zero mean unit variance Gaussian process; constants $\mu = 0.01,\ \epsilon = 0.1,\ \epsilon_1 = 0.001,\ \gamma = 0.0001$. Zero initial conditions $\hat{\theta}(k=0) = 0(100)$ were employed for each of the estimators.

Two different system parameter vectors were considered, each having a length of $n = 100$. Their active parameters were as follows.

$\theta_1:\ \theta_{30} = 0.2;\ \theta_{41} = -4;\ \theta_{52} = -0.04;\ \theta_{82} = 10.$

$\theta_2:\ \theta_{11:15} = randn(5);\ \theta_{56:65} = randn(10).$

where *randn(G)* denotes a vector of G samples of a unit variance zero mean white Gaussian process. The first system parameter vector contains 4 sparsely separated active parameters; the second contains 15 active parameters in the form of two sparsely separated 'clusters'. All other parameters are inactive; that is their magnitudes lie below the NLMS adaptive noise level. A log plot of the parameter magnitudes for both systems is shown in Figure 2.

The (asymptotic) NLMS adaptive noise level [4]

$$lim_{k\to\infty}|\Delta\theta_j(k)| = \sqrt{\mu\sigma_s^2/(\hat{m}\sigma_u^2)}, \forall j$$

for each of the two system simulation conditions is: (i) System1: 0.03 (for $\hat{m} = 4$); (ii) System2: 0.0155 (for $\hat{m} = 15$). These levels are indicated by a horizontal line in the respective log plots of Figure 2. Note: (i) System1 has one of its active parameters (θ_{52}) very near the adaptive noise level; (ii) System2 has one of its active parameters (θ_{70}) relatively near the adaptive noise level, and has several of its inactive parameters lying just under the adaptive noise level.

Figure 3 shows the simulation results for System1. Fig. 3(a) shows the plot over time (sample number) of the number of parameters, $\hat{m}$, detected as being active....using the proposed activity criterion. Fig. 3(b) shows the corresponding log plot over time of the parameter vector estimation error $||\theta - \hat{\theta}(k)||^2$ where $||.||$ denotes the Euclidean norm. The plots are the average of 10 equivalent simulations. Figure 4 shows the corresponding results for System2.

These two figures (together with an examination of the actual detected parameter indices) indicate the following.
(a) For both systems the proposed activity criterion converges relatively quickly towards the true set of active parameters. For both systems the activity criterion slightly over-estimates the true number of active parameters; moreso for system2, but this is to be expected since several of the inactive parameters lie just below the adaptive noise level.
(b) For both systems the proposed NLMS estimator converges ($\hat{\theta}(k) \to \theta$) significantly more quickly than the standard NLMS estimator; moreso for system1, but this is to be expected since system1 has fewer active parameters. Note, for each system the proposed estimator has an asymptotic parameter estimation error close to that theoretically expected for the NLMS estimator: $lim_{k\to\infty}||\theta - \hat{\theta}(k)||^2 = \mu\sigma_s^2/\sigma_u^2 = 0.0036$.

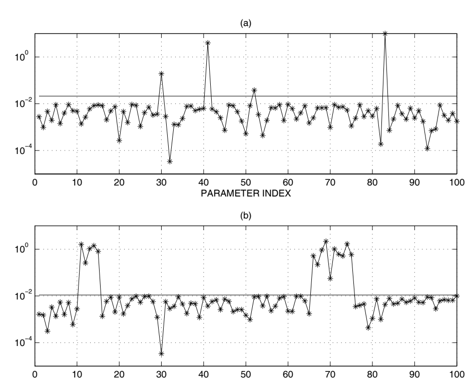

Fig. 2. Parameter magnitudes for (a) System1; (b) System2

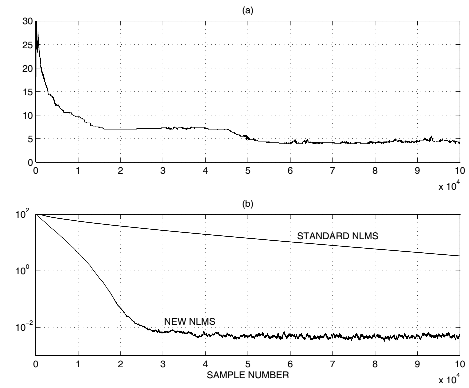

Fig. 3. Simulation results for System1

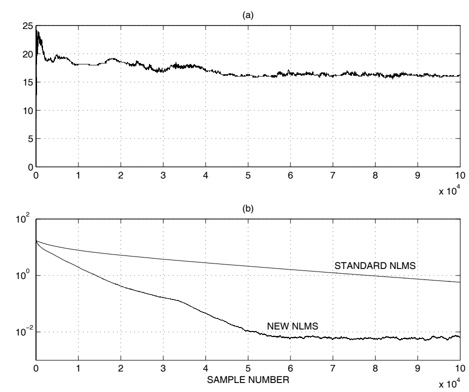

Fig. 4. Simulation results for System2

5. REFERENCES

[1] J.R. Casar-Corredera and J.A. Alcazar-Fernandez, "An acoustic echo canceller for teleconference systems", *Proc. Int. Conf. Acoustic, Speech, Signal Processing (ICASSP86)*, Tokyo, Japan, pp. 1317-1320, 1986.

[2] B. Widrow, J.M. McCool, M.G. Larimore and C.R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter", *Proc. of the IEEE*, Vol. 64, No. 8, pp. 1151-1162, 1976.

[3] J. Homer, R.R. Bitmead, I.M.Y. Mareels, "Quantifying the effects of dimension on the convergence rate of the LMS adaptive FIR estimator", *IEEE Trans. on Signal Proc.*, Vol. 46, pp. 2611-2615, Oct. 1998.

[4] J. Homer, I.M.Y. Mareels, R.R. Bitmead, B. Wahlberg and F. Gustafsson, "Improved LMS estimation via structural detection", *IEEE Trans. on Signal Proc.*, Vol. 46, pp. 2651-2663, Oct. 1998.

[5] J. Homer, "Detection guided NLMS estimation of sparsely parametrized channels", *IEEE Trans. on Circuits and Systems II*, Vol. 47, pp. 1437-1442, Dec. 2000.