inter•noise 96
25th Anniversary Congress – Liverpool
INCE: 75

PERFORMANCE OF MULTI-CHANNEL FEEDFORWARD ADAPTIVE SYSTEMS

J Minkoff

Lucent Technologies/Bell Laboratories, Whippany, NJ 07981 USA

1. INTRODUCTION

In this paper we investigate the performance capability yielded by the filtered-X LMS algorithm in multi-channel, feedforward adaptive systems. Recently[1,2] , this algorithm has been applied to active noise cancellation in a multi-channel system employing inputs consisting of multiple sinusoidal tones. This formulation, and the results, are applicable to deterministic signals only. In what follows an alternative generalized formulation is developed yielding results that are applicable to arbitrary wideband signals. Results include expressions for optimum multi-channel adaptive-filter transfer functions and the associated minimized cost functions, expressed entirely in terms of measurable spectral densities involving deterministic or random reference and target signals. The results also yield a maximum achievable upper performance bound on error cancellation for multi-channel feedforward adaptive systems.

2. MULTI-CHANNEL FEEDFORWARD ADAPTIVE SYSTEMS

The analysis will be carried out in frequency space, for which a generalized multi-channel feedforward adaptive system is shown in Figure 1. In the figure **W** denotes the L × K matrix of, usually, KL adaptive-filters (some of the entries could be zero) connecting the K reference signals to the L secondary sources (actuators) and **P** denotes the M × L matrix of forward transfer functions between the L secondary sources and the M error sensors. The inputs to each error sensor consist of a disturbance or target signal D_i, which is to be cancelled or at least reduced, plus a cancelling, or error-r

[1] Elliot, S.J., Stothers, I.M., and Nelson, P.A., "A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration" IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-35, No. 10, Oct. 1987, 1423-1434.

[2] Elliot, S.J., Boucher, C.E. and Nelson, P.A., "The Behavior of a Multiple Channel Active Control System", IEEE Trans. on Signal Processing, Vol. 40, No. 5, May 1992, 1041-1052.

eduction signal produced by the adaptive system, which yields an error signal E_i. In the weight-iteration stage ΔW – the discrete elements of the adaptive-filter impulse responses are commonly referred to as weights – combinations of reference signals and error signals are cross correlated periodically at some rate yielding correction factors that are added to the previous-interval's weights. Equivalently, there is an iteration in frequency space in which a correction factor is added to the adaptive-filter transfer function of the preceding interval. In the filtered-X LMS algorithm, the transfer-function matrix **P** must be inserted into the forward reference-signal path prior to ΔW in order to "compensate"[3] for the aforementioned physical structure **P** between the secondary sources and the error sensors.

The cost function to be considered is the sum of the two quadratic forms

$$\xi = \vec{E}^\dagger \, \alpha \, \vec{E} + (\mathbf{W}\vec{X})^\dagger \, \eta \, (\mathbf{W}\vec{X}) \tag{1}$$

where, to insure convergence, α and η must be Hermitian and positive definite but need not be diagonal. Also, both α and η can vary with frequency.

Now, using the above definitions, the squared error ξ can be written as

$$\xi = \left(\vec{D}^\dagger - \vec{X}^\dagger \, \mathbf{W}^\dagger \, \mathbf{P}^\dagger\right) \alpha \left(\vec{D} - \mathbf{P}\,\mathbf{W}\vec{X}\right) + \vec{X}^\dagger \, \mathbf{W}^\dagger \, \eta \, \mathbf{W}\vec{X} \tag{2}$$

Taking $\partial\xi/\partial W = 0$ yields the weight–iteration matrix (3)

$$\Delta \mathbf{W} = 2\,\mathbf{P}^\dagger \, \alpha \, \vec{E}\,\vec{X}^\dagger + 2\eta\,\mathbf{W}\vec{X}\vec{X}^\dagger$$

and the weight-iteration procedure for the filtered-X LMS algorithm is for the k+1 st iteration

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2\mu \mathbf{P}^\dagger \alpha \, \vec{E}\,\vec{X}^\dagger - 2\mu\eta\,\mathbf{W}(k)\vec{X}\vec{X}^\dagger \tag{4}$$

where μ is the step-size parameter which, in general, need not be the same for each filter.

Now let us assume that the system has converged to the steady state. We wish to determine an expression for the W_{ij} that minimize the cost function. The necessary condition is

$$\mathbf{P}^\dagger \, \alpha \, \vec{E}\,\vec{X}^\dagger - \eta\,\mathbf{W}\vec{X}\vec{X}^\dagger = 0 \quad \text{or}$$

$$\mathbf{P}^\dagger \, \alpha \, \mathbf{P}\,\mathbf{W}\vec{X}\vec{X}^\dagger + \eta\,\mathbf{W}\vec{X}\vec{X}^\dagger = \mathbf{P}^\dagger \, \alpha \, \vec{D}\,\vec{X}^\dagger \tag{5}$$

It is assumed that the fluctuations in the reference and target signals will be symmetrical about their mean values, assumed to be zero, in which case the fluctuations in the W_{ij} will also be symmetrical about their mean values $\overline{W}_{ij}$. Hence, with these assumptions, all the relevant probability density functions are symmetrical with respect to the various means (e.g. Gaussian distributions),

[3] Widrow, B. and Stearns, Samuel D., *Adaptive Signal Processing*, Prentice Hall, Englewood, NJ 1985.

and, by applying the expectation operator E: $E[(W-\overline{W})\,(\vec{X}\vec{X}^{\dagger})] = 0$. Now define the matrices $\mathbf{T} = E(\vec{D}\,\vec{X}^{\dagger}) = \overline{\vec{D}\vec{X}}^{\dagger}$ $\mathbf{S} = E(\vec{X}\,\vec{X}^{\dagger})$ and application of the expectation operator to (5) yields

$$(\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P})\,\overline{\mathbf{W}}\,\mathbf{S} + \boldsymbol{\eta}\,\overline{\mathbf{W}}\mathbf{S} = \mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{T} \text{ and } \overline{\mathbf{W}} = \left[\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P} + \boldsymbol{\eta}\right]^{-1}\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{T}\,\mathbf{S}^{-1} \qquad (6)$$

which shows that $\mathbf{S}$ must be non-singular for a unique solution.

Now, given the optimum $\overline{\mathbf{W}}$ we wish to evaluate the resulting minimized cost function. From (2)

$$\xi = \vec{D}^{\dagger}\,\boldsymbol{\alpha}\,\vec{D} - \vec{D}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P}\,\overline{\mathbf{W}}\,\vec{X} - \vec{X}^{\dagger}\,\overline{\mathbf{W}}^{\dagger}\,\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\vec{D} + \qquad (7)$$

$$\vec{X}^{\dagger}\,\overline{\mathbf{W}}^{\dagger}\,\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P}\,\overline{\mathbf{W}}\,\vec{X} + \vec{X}^{\dagger}\,\overline{\mathbf{W}}^{\dagger}\,\boldsymbol{\eta}\,\overline{\mathbf{W}}\,\vec{X}^{\dagger}$$

This form for ξ, which is similar to that derived in [1], is adequate for deterministic signals, but not convenient for random signals because application of an expectation operation does not directly yield spectral densities. In order to cast the result in a more suitable form we make use of an alternative way of writing bilinear forms. Namely, it is trivial to show that if $\vec{X}$ is M×1, $\vec{Y}$ is N×1 and $\mathbf{C}$ is M×N then: $\vec{X}^{\dagger}\,\mathbf{C}\vec{Y} = \mathrm{tr}\,(\vec{X}^{*}\,\vec{Y}^{T}\,\mathbf{C}^{T})$ where tr is the trace operator. And by applying the expectation operator to (7), referring to (3) and (5), and using $\mathbf{tr}\,(\mathbf{A}\,\mathbf{B}) = \mathbf{tr}\,(\mathbf{B}\,\mathbf{A}) = \mathbf{tr}\,(\mathbf{A}\,\mathbf{B})^{T}$

$$\bar{\xi} = \vec{D}^{\dagger}\,\boldsymbol{\alpha}\,\vec{D} - \mathbf{tr}\,\left[\boldsymbol{\alpha}\,\mathbf{P}\,\overline{\mathbf{W}}\,\mathbf{T}^{\dagger}\right] \qquad (8)$$

But, referring to (6), and letting $\mathbf{M} = \mathbf{P}\,(\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P} + \boldsymbol{\eta})^{-1}\,\mathbf{P}^{\dagger}$ we obtain

$$\bar{\xi} = \vec{D}^{\dagger}\,\boldsymbol{\alpha}\,\vec{D} - \mathbf{tr}\,\left[\boldsymbol{\alpha}\,\mathbf{M}\,\boldsymbol{\alpha}\,\mathbf{T}\,\mathbf{S}^{-1}\,\mathbf{T}^{\dagger}\right] \qquad (9)$$

3. MAXIMUM ACHIEVABLE PERFORMANCE

Maximum performance is achieved when no constraint is placed on power consumption (i.e., $\boldsymbol{\eta} = \mathbf{0}$), in which case we deal with the reduced cost function. $\bar{\xi}_R = \vec{E}^{\dagger}\,\boldsymbol{\alpha}\,\vec{E}$. It could be of interest to let $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ vary with frequency over the working bandwidth, which would enable tailoring of the cost function to whatever specification might be required. For example, it would be possible, say, to maximize error reduction in certain bands by setting $\boldsymbol{\eta} = 0$ in those bands, setting $\boldsymbol{\alpha} = 0$ in other bands to minimize power consumption, etc. This however raises the issue of the rank of $\mathbf{P}$. That is, since $\boldsymbol{\eta}$ has been assumed to be positive definite then $[\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P}+\boldsymbol{\eta}]^{-1}$ in (6) always exists and $\overline{\mathbf{W}}$ is always uniquely defined. However, if $\boldsymbol{\eta} = \mathbf{0}$, then $(\mathbf{P}^{\dagger}\,\boldsymbol{\alpha}\,\mathbf{P})^{-1}$ does not exist unless $\mathbf{P}$ is full rank. We now consider some of the effects on system operation when $\boldsymbol{\eta} = \mathbf{0}$, which also raises the issue of the dependence of the system characteristics on whether the transfer-function matrix $\mathbf{P}$ is square or rectangular.

If $\eta = 0$, the reduced cost function becomes

$$\bar{\xi}_R = \vec{D}^{\dagger}\,\alpha\,\vec{D} - \mathrm{tr}\left[\alpha M_o\,\alpha\,T S^{-1}\,T\right] \text{ where } M_o = P\left[P^{\dagger}\alpha P\right]^{-1}P^{\dagger} \tag{10}$$

But if **P** is square we have simply $\overline{W} = P^{-1}\,T S^{-1}$ with $M_o\alpha = I$ and

$$\bar{\xi}_R = \vec{D}^{\dagger}\,\alpha\,\vec{D} - \mathrm{tr}\left[\alpha\,T S^{-1}\,T^{\dagger}\right] \tag{11}$$

Comparison of (10) with (11) shows that the minimum mean square error for the reduced cost function for square **P**-matrix systems depends only on the statistical properties of the reference and target signals, whereas for rectangular systems the forward transfer functions P_{ij} also come into play. As will be shown, with sufficiently high coherence between reference and target signals, the mean-square errors can ideally be made arbitrarily small, whereas, in the rectangular case, because of the **P** dependence, even under these conditions small errors are not guaranteed. A striking example of this is presented in the results of the computer experiments in Sec. 4.

3.1 Square and Rectangular System

Specification of a reduced cost function determines the number of detection points. For any given ξ_R, it will now be proved that maximum performance – in terms of minimizing the reduced cost function – is achieved when the number of secondary sources is equal to the number of detection points (square system). This is a basic property of the filtered-X LMS algorithm. However, it is not always possible to take advantage of this property, since there can certainly be applications in which the use of square systems would not be advisable. Relative practical applicability of square and rectangular systems is not considered further here.

To prove the optimality of square over rectangular systems – subject to the aforementioned qualifications – consider a system with a desired number M detection points and L secondary sources. Given an arbitrary M×1 disturbance vector $\vec{D}$, the purpose of the adaptation process is to find an L×1-dimensional secondary source vector $\vec{\Sigma} = W\vec{X}$ such that $PW\vec{X} = \vec{D}$ (12)

where **P** is M×L. If M = L and **P** is non-singular then $\vec{\Sigma} = W\vec{X} = P^{-1}\vec{D}$ and multiplying from the right by $\vec{X}^{\dagger}$, applying the expectation operator, referring to (12), and assuming that **S** is non-singular, yields the foregoing result

$$\overline{W} = P^{-1}\,TS^{-1}. \tag{13}$$

On the other hand if M > L (it is assumed here that the number of secondary sources is not greater than the number of disturbances), then (12) in general represents an inconsistent set of equations for which there is no solution; the Σ_is are overdetermined. This situation can be described geometrically as follows[4]. Eqn. (12) represents an attempt to express the M-dimensional vector $\vec{D}$ as a linear combination of the L columns of **P** with weights Σ_i which are the elements of $\vec{\Sigma}$. Hence, for arbitrary $\vec{D}$, if L < M the dimension L of the column space of **P** will in general not be large enough to accomplish this. Unless by fortuitous circumstance $\vec{D}$ happens to be in the column space of **P**, the best we can hope for is to find that $\vec{\Sigma}_0$ such that $\mathbf{P}\vec{\Sigma}_0$ is the projection of $\vec{D}$ in the column space of **P**. This yields the minimum error under the circumstances because the error vector is perpendicular to the projection of $\vec{D}$, but by the Pythagorean theorem, the system performance will be degraded from that achieved by (13), which proves the assertion. Referring to (6), with $\eta = 0$, the solution for W is obtained in this case by requiring for arbitrary $\vec{Y}$ that $(\mathbf{P}\vec{Y})^\dagger \,\boldsymbol{\alpha}\,(\vec{D} - \mathbf{P}\,\mathbf{W}\vec{X}) = \vec{Y}^\dagger\, \mathbf{P}^\dagger\, \boldsymbol{\alpha}\,(\vec{D} - \mathbf{P}\,\mathbf{W}\vec{X}) = 0$ and therefore $\mathbf{P}^\dagger \boldsymbol{\alpha} \mathbf{P}\,\mathbf{W}\vec{X} = \mathbf{P}^\dagger \boldsymbol{\alpha}\vec{D}$. And again, multiplying from the right by $\vec{X}^\dagger$, applying the expectation operator etc., yields the foregoing solution for the rectangular case. This illustrates the fundamentally algebraic character of the multidimensional convergence process.

Now considering a square system with M secondary sources and disturbances and K reference signals, and diagonal **S** it can be shown that maximum performance is given by

$$\xi = \sum_{j=1}^{M} \alpha_j \overline{|D_j|^2} \left(1 - \sum_{i=1}^{K} \gamma_{ij}^2\right) \tag{14}$$

where the coherences γ_{ij} between reference signals X_i and disturbance D_j are given by

$$\gamma_{ij}^2 = \frac{|\overline{X_i^* D_j}|^2}{\overline{|X_i|^2}\;\overline{|D_j|^2}} = \frac{|T_{ij}|^2}{\overline{|X_i|^2}\;\overline{|D_j|^2}}$$

4. RESULTS OF COMPUTER EXPERIMENTS

In order to validate the results of Section 3, experiments were run using a computer simulation of a feedforward adaptive system, such as is diagrammed in Figure 1, with K = L = M = 2. The reference and target signals consisted of band-limited Gaussian noise.

We first consider a square well-conditioned system with $|P_{11}| = |P_{22}| = 80$ dB and $|P_{12}| = |P_{22}|$ either 70 dB or 46 dB and diagonal **S** and **T** with $\gamma_{11}^2 = .95$, $\gamma_{22}^2 = .90$ and $\gamma_{12}^2 = \gamma_{21}^2 = 0$. Referring to (14) we should expect reductions in D_1 and D_2 by factors of nominally $10\log(1-\gamma_{11}^2)^{-1} = 13$ dB and $10\log(1-\gamma_{22}^2)^{-1} = 10$ dB, independent of the values of the off-diagonal transfer-function magnitude.

The results for these cases are shown in Figure 2 for $|P_{12}| = |P_{21}| = 70$ dB and in Figure 3 for $|P_{12}| = |P_{21}| = 46$ dB.

[4] Strang, G., *Linear Algebra and Its Applications*, Ch. 3, Second Edition, Academic Press.

The results agree to within tenths of a dB, which supports the prediction of **P**-independence for square systems. The values of error reduction are also in good agreement with predictions, to within ~1 dB for Tgt 1 and ~1.5 dB for Tgt 2.

We next consider a rectangular case. In this case $\mathbf{P} = [P_{11}\, P_{12}]^\dagger$. And since both **S** and **T** are diagonal, (9) becomes

$$\bar{\xi} = \overline{|D_1|^2}\left[1 - \frac{|P_{11}|^2 \gamma_{11}^2}{P_{11}|^2 + |P_{12}|^2}\right] + \overline{|D_2|^2}\left[1 - \frac{|P_{12}|^2 \gamma_{22}^2}{|P_{11}|^2 + |P_{12}|^2}\right] \tag{15}$$

In effect, the P_{ij} dependence reduces the effective coherence. In this example, if $|P_{11}| = |P_{12}|$ the maximum error reduction would be 3 dB. If, as would probably be more likely, $|P_{11}|$ and $|P_{12}|$ were not equal, the performance for the target signal corresponding to the larger P_{ij} improves at the expense of the other. If one of the P_{ij}'s is much larger than the others, the system reduces effectively to a single channel. In these experiments, $|P_{11}|^2 = 10^8$ and $|P_{12}|^2$ is either $10^{4.6}$ or 10^7. Hence, using (15), the equations corresponding to (14) for reduction of D_1 and D_2 respectively are

$\underline{P_{12} = 46 \text{ dB}}$

$$10\log\left[1-(.99)(.95)\right]^{-1} = 13 \text{ dB} \quad 10\log\left[1-(4\times10^{-4})(.90)\right]^{-1} = 0 \text{ dB}$$

$\underline{P_{12} = 70 \text{ dB}}$

$$10\log\left[(1-(.909)(.95))\right]^{-1} = 8.7 \text{ dB} \quad 10\log\left[(1-(.090)(.90))\right]^{-1} = .4 \text{ dB}$$

The results of the computer experiments for these two cases are presented in Figures 4 and 5 showing extremely close agreement with these predictions, and of course the dependence on **P** for rectangular systems.

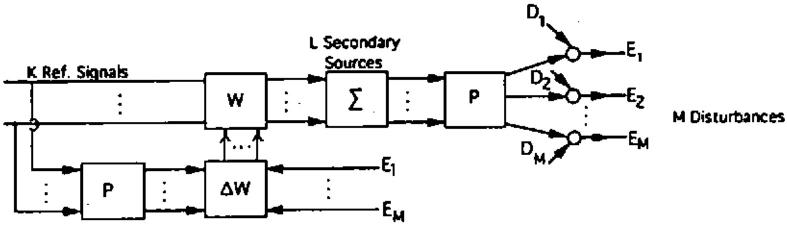
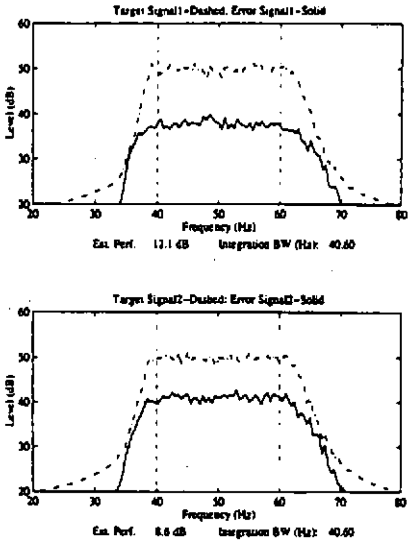
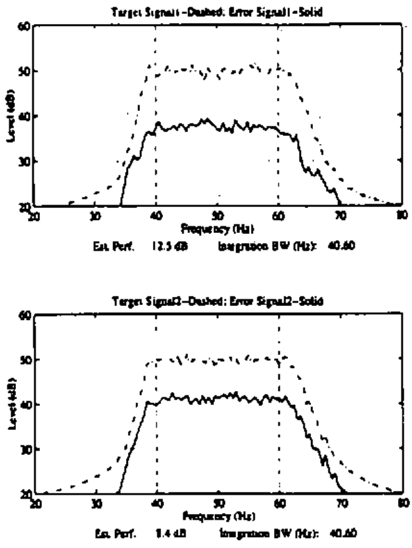


Fig. 1 Multidimensional Feedforward Adaptive System



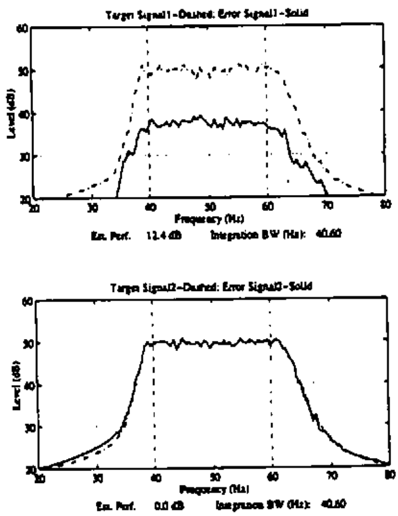
2X2 System, Well Conditioned, 70-dB Off-Diagonal TF s

Fig. 2 Performance of Square Well Conditioned System - K = 2, L = 2, M = 2



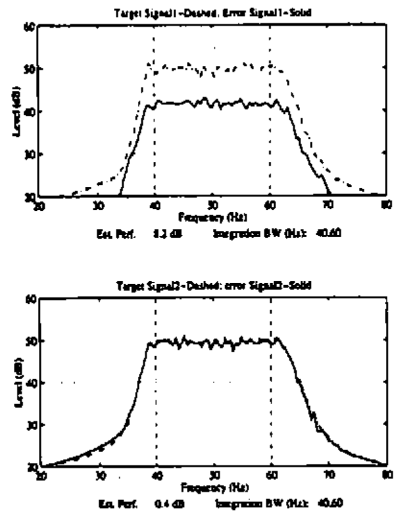
2X2 System, Well Conditioned, 40-dB Off-Diagonal TF s

Fig. 3 Performance of Square Well Conditioned System



1X2 System, 46 dB Off-Diagonal TF's

Fig. 4 Performance of Rectangular System - K = 2, L = 1, M = 2



1X2 System, 10 dB Off-Diagonal TF's

Fig. 5 Performance of Rectangular System - K = 2, L = 1, M = 2