Assesment of the efficiency of the LMS algorithm based on spectral information

(Invited Paper)

Aaron Flores and Bernard Widrow
ISL, Department of Electrical Engineering, Stanford University, Stanford CA, USA
Email: {aflores,widrow}@stanford.edu

***Abstract*— The LMS algorithm is used to find the optimal minimum mean-squared error (MMSE) solutions for a wide variety of problems. Unfortunately, its convergence speed depends heavily on its initial conditions when the autocorrelation matrix R of its input vector has a high eigenvalue spread. In many applications such as system identification or channel equalization, R is Toeplitz. In this paper we exploit the Toeplitz structure of R to show that when the weight vector is initialized to zero, the convergence speed of LMS is related to the similarity between the input PSD and the power spectrum of the optimum solution.**

I. Introduction

The LMS algorithm is widely used to find the minimum mean-squared error (MMSE) solutions for a variety of linear estimation problems, and its efficiency has been extensively studied during the last decades [1]–[18]. The convergence speed of LMS depends on two factors: the eigenvalue distribution of the input autocorrelation matrix R, and the chosen initial condition. Modes corresponding to small eigenvalues converge much more slowly than those corresponding to large eigenvalues, and the initialization of the algorithm determines how much excitation is received by each of these modes. In practice, it is not known how the chosen initial conditions excite each of the modes, making the LMS speed of convergence difficult to predict. Our analysis of the convergence speed is concerned with applications where the input vector comes from a tapped delay line, inducing a toeplitz structure in the R matrix. Using the toeplitz nature of R, we show that the speed of convergence of the LMS algorithm can be estimated from its input power spectral density and the spectrum of the difference between the initial weight vector and the optimum solution. The speed of convergence of LMS was qualitatively assessed by comparing it to its ideal counterpart, the LMS/Newton algorithm, which is often used as a benchmark for adaptive algorithms [17], [18]. In section II we describe both the LMS and LMS/Newton algorithms and derive approximations to their learning curves on which our transient analysis is based. In section III we define the performance metric used to evaluate LMS speed of convergence, and in section IV we show how that metric can be estimated from spectral information. We illustrate our result with simulations in section V and summarize our conclusions and future work in Section VI.

II. The LMS and LMS/Newton algorithms

Adaptive algorithms such as LMS and LMS/Newton are used to solve linear estimation problems like the one depicted in Fig. 1, where the input vector $\boldsymbol{x}_k = [x_{1k} x_{2k} \cdots x_{Lk}]^T$ and desired response $d_k \in \Re$ are jointly stationary random processes, $\boldsymbol{w}_k = [w_{1k} w_{2k} \cdots w_{Lk}]^T$ is the weight vector, $y_k = \boldsymbol{x}_k^T \boldsymbol{w}_k$ is the output, and $\epsilon_k = d_k - y_k$ is the error. The Mean Square Error (MSE) is defined as $\xi_k = E[\epsilon_k^2]$ and it is a quadratic function of the weight vector. The optimal weight vector that minimizes ξ_k is given by $\boldsymbol{w}^* = \boldsymbol{R}^{-1}\boldsymbol{p}$, where $\boldsymbol{R} = E[\boldsymbol{x}_k \boldsymbol{x}_k^T]$ is the input autocorrelation matrix (assumed to be full rank), and $\boldsymbol{p} = E[\boldsymbol{x}_k d_k]$ is the crosscorrelation vector. The minimum MSE (MMSE) obtained using $\boldsymbol{w}^*$ is denoted by ξ^*.

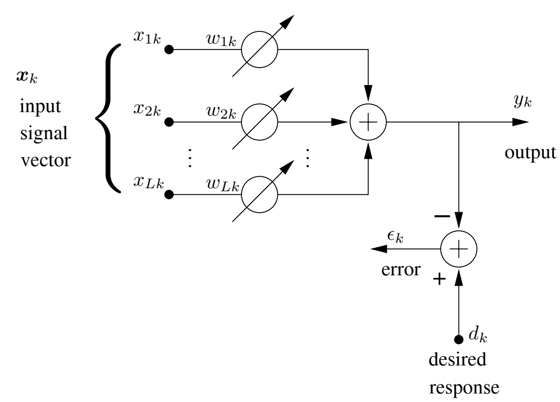

Fig. 1. Adaptive Linear Combiner

A. The LMS algorithm

Often in practice $\boldsymbol{R}^{-1}\boldsymbol{p}$ cannot be calculated due to the lack of knowledge of the statistics $\boldsymbol{R}$ and $\boldsymbol{p}$. However, when samples of $\boldsymbol{x}_k$ and d_k are available, they can be used to iteratively adjust the weight vector to obtain an approximation of $\boldsymbol{w}^*$. The simplest and most widely used algorithm for this is LMS [1]. It performs instantaneous gradient descent adaptation of the weight vector:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + 2\mu\epsilon_k \boldsymbol{x}_k. \tag{1}$$

The step size parameter is μ and the initial weight vector $\boldsymbol{w}_0$ is arbitrarily set by the user. The MSE sequence ξ_k corresponding to the sequence of adapted weight vectors $\boldsymbol{w}_k$ is commonly known as the learning curve.

Next we derive an approximation for ξ_k that is the basis of our speed of convergence analysis. Since $\boldsymbol{R}$ is symmetric, we can diagonalize it as $\boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{\Lambda}\boldsymbol{Q}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues λ_i of $\boldsymbol{R}$, and $\boldsymbol{Q}$ is a real orthonormal matrix with the corresponding eigenvectors of $\boldsymbol{R}$ as columns. Let $\boldsymbol{v}_k \triangleq \boldsymbol{w}_k - \boldsymbol{w}^*$ be the weight vector deviation from the optimal solution and $\boldsymbol{v}'_k \triangleq \boldsymbol{Q}^T\boldsymbol{v}_k$. Let $\boldsymbol{F}_k$ be the vector obtained from the diagonal of $E[\boldsymbol{v}'_k{\boldsymbol{v}'_k}^T]$, and $\boldsymbol{\lambda} \triangleq [\lambda_1\lambda_2\ldots\lambda_L]^T$. Assuming the random processes $\{\boldsymbol{x}_k, d_k\}$ to be independent, it can be shown [3] that the MSE at time k can be expressed as

$$\xi_k = \xi^* + \boldsymbol{\lambda}^T\boldsymbol{F}_k. \tag{2}$$

If we further assume $\boldsymbol{x}_k$ to be gaussian, it was shown in [5] and [3] that $\boldsymbol{F}_k$ obeys the following recursion

$$\boldsymbol{F}_{k+1} = (I - 4\mu\boldsymbol{\Lambda} + 8\mu^2\boldsymbol{\Lambda}^2 + 4\mu^2\boldsymbol{\lambda}\boldsymbol{\lambda}^T)\boldsymbol{F}_k + 4\mu^2\xi^*\boldsymbol{\lambda}, \tag{3}$$

where $\boldsymbol{F}_k$ is shown [5] to converge if $\mu < \frac{1}{3\,\mathrm{Tr}\,(R)}$. This condition in μ allows us to approximate (3) by

$$\boldsymbol{F}_{k+1} \approx (I - 4\mu\boldsymbol{\Lambda})\boldsymbol{F}_k + 4\mu^2\xi^*\boldsymbol{\lambda}. \tag{4}$$

Denoting by $\mathbf{1} \in \Re^L$ a vector of ones and using (4),

$$\boldsymbol{F}_k \approx (I - 4\mu\boldsymbol{\Lambda})^k(\boldsymbol{F}_0 - \mu\xi^*\mathbf{1}) + \mu\xi^*\mathbf{1}, \tag{5}$$

Substituting (5) in (2), we obtain the following approximation for the LMS learning curve

$$\xi_k \approx \xi_\infty + \boldsymbol{\lambda}^T(I - 4\mu\boldsymbol{\Lambda})^k(\boldsymbol{F}_0 - \mu\xi^*\mathbf{1}). \tag{6}$$

where $\xi_\infty \triangleq \lim_{k\to\infty}\xi_k \approx \xi^*(1 + \mu\,\mathrm{Tr}\,(\boldsymbol{R}))$.

B. The LMS/Newton algorithm

The LMS/Newton algorithm [6] is an ideal variant of the LMS algorithm that uses $\boldsymbol{R}$ to "whiten" its input. Although most of the time it cannot be implemented in practice due to the lack of knowledge of $\boldsymbol{R}$, it is of theoretical importance as a benchmark for adaptive algorithms [17], [18]. The LMS/Newton algorithm is the following

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + 2\mu\lambda_{\text{avg}}\boldsymbol{R}^{-1}\epsilon_k\boldsymbol{x}_k \tag{7}$$

where $\lambda_{\text{avg}} \triangleq \frac{\sum_{i=1}^N \lambda_i}{n}$. It is well known [6] that the LMS/Newton algorithm is equivalent to the LMS algorithm with learning rate $\mu\lambda_{\text{avg}}$ and $\boldsymbol{x}_k$ previously whitened and normalized such that $\Lambda = I$. Therefore, assuming the same independence and gaussian statistics of $\boldsymbol{x}_k$ and d_k as with LMS, we can replace μ by $\mu\lambda_{\text{avg}}$ and Λ by I in (6) to obtain the following approximation for the LMS/Newton learning curve

$$\xi_k \approx \xi_\infty + (1 - 4\mu\lambda_{\text{avg}})^k\boldsymbol{\lambda}^T(\boldsymbol{F}_0 - \mu\xi^*\mathbf{1}) \tag{8}$$

where the asymptotic MSE ξ_∞ is the same as for LMS.

III. LMS Transient Efficiency

Examining (6) and (8) we can see that the learning curve for LMS is a sum of geometric sequences (modes) with $1 - 4\mu\lambda_i$ as geometric ratios, whereas for LMS/Newton it consists of a single geometric sequence with geometric ratio $1 - 4\mu\lambda_{\text{avg}}$. If all the eigenvalues of $\boldsymbol{R}$ are equal, the learning curves for LMS and LMS/Newton are the same. Generally, the eigenvalues of $\boldsymbol{R}$ are not equal, and LMS consists of multiple modes, some faster than LMS/Newton and some slower than it as depicted in Fig. 2

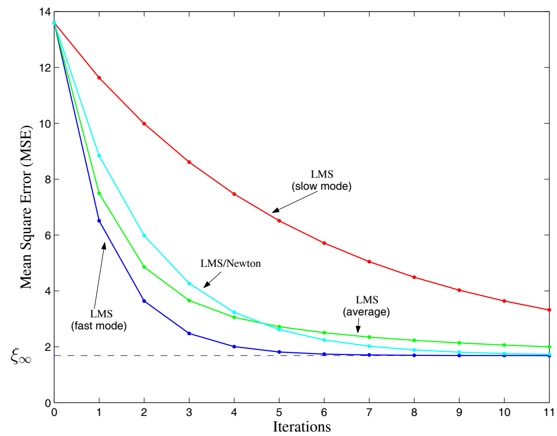

Fig. 2. LMS and LMS/Newton modes

To evaluate the speed of convergence of a learning curve we use the natural measure suggested in [5],

$$J \triangleq \sum_{k=0}^{\infty}\xi_k - \xi_\infty. \tag{9}$$

Small values of J indicate fast convergence and large values indicate slow convergence. Using (6) and (8) to obtain J for LMS and LMS/Newton respectively we get

$$J^{LMS} \approx \frac{\mathbf{1}^T\boldsymbol{F}_0 - \mu L\xi^*}{4\mu} = \frac{\boldsymbol{v}_0^T\boldsymbol{v}_0}{4\mu} - \frac{L\xi^*}{4} \tag{10}$$

$$J^{LMS/Newton} \approx \frac{\boldsymbol{\lambda}^T\boldsymbol{F}_0 - \mu\lambda_{\text{avg}}L\xi^*}{4\mu\lambda_{\text{avg}}} = \frac{\boldsymbol{v}_0^T\boldsymbol{R}\boldsymbol{v}_0}{4\mu\lambda_{\text{avg}}} - \frac{L\xi^*}{4} \tag{11}$$

To compare the speed of convergence of LMS to the one of LMS/Newton, we define what we call the LMS Transient Efficiency

$$\text{LMS Tansient Efficiency} \triangleq \frac{J^{LMS/Newton}}{J^{LMS}} \tag{12}$$

If the Transient Efficiency is bigger/smaller than one, then LMS performs better/worse than LMS/Newton in the sense of the J metric. Assuming that the initial weight vector $\boldsymbol{w}_0$ is far enough from $\boldsymbol{w}^*$ such that $\boldsymbol{v}_0^T\boldsymbol{v}_0 \gg \mu L\xi^*$ and $\frac{\boldsymbol{v}_0^T\boldsymbol{R}\boldsymbol{v}_0}{\mu\lambda_{\text{avg}}} \gg \mu L\xi^*$, we obtain from (10), (11) and (12)

$$\text{LMS Transient Efficiency} \approx \frac{1}{\lambda_{\text{avg}}}\frac{\boldsymbol{v}_0^T\boldsymbol{R}\boldsymbol{v}_0}{\boldsymbol{v}_0^T\boldsymbol{v}_0} \tag{13}$$

Since (13) is invariant to scaling of $\boldsymbol{v}_0$ and $\boldsymbol{R}$, we can assume without loss of generality $\|\boldsymbol{v}_0\| = \lambda_{\text{avg}} = 1$, obtaining the following compact expression:

$$\text{LMS Transient Efficiency} \approx \boldsymbol{v}_0^T \boldsymbol{R} \boldsymbol{v}_0 \tag{14}$$

This is the first contribution in this paper and it will be used to obtain an approximation of the LMS Transient Efficiency in terms of spectra in the next section.

IV. LMS Transient Efficiency in terms of spectral information

In this section we show that in the case that $\boldsymbol{R}$ is toeplitz, the LMS Transient Efficiency can be expressed in terms of the Fourier spectrum of $\boldsymbol{v}_0$ and the input power spectral density. The toeplitz structure of $\boldsymbol{R}$ arises from applications where the input vector $\boldsymbol{x}_k$ comes from a Tapped Delay Line (TDL) as shown in Fig. 3, i. e. $\boldsymbol{x}_k = [x_k \; x_{k-1} \; x_{k-2} \; \ldots \; x_{k-L+1}]^T$, where x_k is a stationary scalar random process with autocorrelation sequence $\phi_{xx}[n] \stackrel{\triangle}{=} E[x_k \, x_{k+n}]$. The input autocorrelation matrix is therefore given by $\boldsymbol{R}_{k,l} = \phi_{xx}[k-l]$.

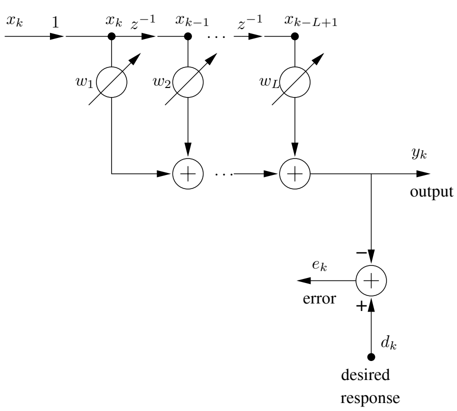

Fig. 3. Tapped Delay Line

We define the Input Power Spectral Density vector $\boldsymbol{\Phi_{xx}} \in \Re^M$ as the M uniformly spaced samples of the DTFT of the N-point truncation of the input autocorrelation function $\phi_{xx}[n]$, i. e.

$$\boldsymbol{\Phi_{xx}}[m] \stackrel{\triangle}{=} \sum_{n=-N+1}^{N-1} \phi_{xx}[n] e^{-2\pi j \frac{mn}{M}} \tag{15}$$
$$m = 0, 1, \ldots M-1, \quad N \geq L, \quad M \geq N + L - 1.$$

In order to show how $\boldsymbol{\Phi_{xx}} \in \Re^M$ is related to the LMS Transient Efficiency, the following Lemma will be needed

Lemma 1 (Spectral Factorization of $\boldsymbol{R}$): Let $\boldsymbol{\Psi}$ be a $M \times M$ diagonal matrix with $\boldsymbol{\Psi}_{m,m} = \boldsymbol{\Phi_{xx}}[m]$ as its diagonal entries, and let $\boldsymbol{U}$ be a $L \times M$ matrix with elements $\boldsymbol{U}_{l,m} = \frac{1}{\sqrt{M}} e^{2\pi j \frac{ml}{M}}$ with $0 \leq l \leq L-1$ and $0 \leq m \leq M-1$, then $\boldsymbol{R}$ can be factored in the following way

$$\boldsymbol{R} = \boldsymbol{U} \boldsymbol{\Psi} \boldsymbol{U}^* \tag{16}$$

Proof: Let $\boldsymbol{T} \stackrel{\triangle}{=} \boldsymbol{U} \boldsymbol{\Psi} \boldsymbol{U}^*$, then

$$\boldsymbol{T}_{k,l} = \sum_{m=0}^{M-1} \frac{1}{\sqrt{M}} e^{2\pi j \frac{km}{M}} \boldsymbol{\Psi}_{m,m} \frac{1}{\sqrt{M}} e^{2\pi j \frac{-ml}{M}} \tag{17}$$
$$= \frac{1}{M} \sum_{m=0}^{M-1} \boldsymbol{\Phi_{xx}}[m] e^{2\pi j \frac{m(k-l)}{M}} \tag{18}$$
$$= \frac{1}{M} \sum_{m=0}^{M-1} \sum_{n=-N+1}^{N-1} \phi_{xx}[n] e^{-2\pi j \frac{mn}{M}} e^{2\pi j \frac{m(k-l)}{M}} \tag{19}$$
$$= \sum_{n=-N+1}^{N-1} \phi_{xx}[n] \frac{1}{M} \sum_{m=0}^{M-1} e^{-2\pi j \frac{m(k-l-n)}{M}} \tag{20}$$
$$= \sum_{n=-N+1}^{N-1} \phi_{xx}[n] \delta[k-l-n] \tag{21}$$
$$= \phi_{xx}[k-l] = \boldsymbol{R}_{k,l} \tag{22}$$

where the step from (20) to (21) was done using the following inequalities: $-L+1 \leq k-l \leq L-1$, $-N+1 \leq n \leq N-1$, and $M \geq N + L - 1$. ■

Let $\boldsymbol{V}_0 \stackrel{\triangle}{=} \boldsymbol{U}^* \boldsymbol{v}_0$ be the M-point DFT of $\boldsymbol{v}_0$, i. e.

$$\boldsymbol{V}_0[m] \stackrel{\triangle}{=} \frac{1}{\sqrt{M}} \sum_{n=0}^{L-1} \boldsymbol{v}_0[n] e^{-2\pi j \frac{mn}{M}} \quad m = 0, 1, \ldots M-1, \tag{23}$$

Using Lemma 1 and (23) we obtain

$$\boldsymbol{v}_o^T \boldsymbol{R} \boldsymbol{v}_0 = \boldsymbol{v}_o^T \boldsymbol{U} \boldsymbol{\Psi} \boldsymbol{U}^* \boldsymbol{v}_0 = \sum_{m=0}^{M-1} \Phi_{xx}[m] |\boldsymbol{V}_o[m]|^2, \tag{24}$$

furthermore, inserting (24) into (14), we arrive to the following approximation for the LMS Transient Eficciency in terms of the input psd and the spectrum of $\boldsymbol{v}_o$

$$\text{LMS Transient Efficiency} \approx \sum_{m=0}^{M-1} \Phi_{xx}[m] |\boldsymbol{V}_0[m]|^2 \tag{25}$$

We should note that, although the independence assumption under which the approximation in (14) was derived is violated when a tapped delay line is used, it has been observed to hold in simulations. Since $\boldsymbol{v}_0$ and $\boldsymbol{R}$ are scaled such that $\|\boldsymbol{v}_0\| = \lambda_{\text{avg}} = 1$, we can show that

$$\frac{\sum_{m=0}^{M-1} \Phi_{xx}[m]}{M} = 1, \quad \sum_{m=0}^{M-1} |\boldsymbol{V}_0[m]|^2 = 1 \tag{26}$$

These normalizations on $\Phi_{xx}[m]$ and $|\boldsymbol{V}_0[m]|^2$ allow us to interpret the right hand side of (25) as a weighted average of $\Phi_{xx}[m]$ where the weights are given by $|\boldsymbol{V}_0[m]|^2$. Therefore, if $\Phi_{xx}[m]$ is large in the frequency ranges where $|\boldsymbol{V}_0[m]|^2$ is also large, the LMS Transient Efficiency will likely be bigger than one, implying that LMS will outperform LMS/Newton under the J metric. On the other hand, if $\Phi_{xx}[m]$ tends to be small in the frequency ranges where $|\boldsymbol{V}_0[m]|^2$ is large, the LMS Transient Efficiency will likely be less than one indicating that LMS will perform worse than LMS/Newton in the J metric sense.

An important application of (25) is to the special case when the weight vector is initialized to zero $\boldsymbol{w}_o = 0$, hence $\boldsymbol{v}_o = -\boldsymbol{w}^*$, which results in

$$|\boldsymbol{V}_0[m]|^2 = |\boldsymbol{W}^*[m]|^2 \tag{27}$$

where $\boldsymbol{W}^*[m]$ is the M-point Discrete Fourier Transform of the optimal or Wiener solution $\boldsymbol{w}^*$ as defined in (23). Therefore, when the weight vector is initialized to zero, the speed of convergence of the LMS algorithm can be estimated by the weighted average of the input psd with the weights given by the spectrum of the Wiener solution. This observation can be very useful for system identification applications, where prior approximate knowledge of the input psd and frequency response of the plant to be identified may be used to estimate LMS's performance before implementing it. For example, if the plant has a low-pass frequency response and the input psd has very little power in the frequency range above the cut-off frequency, LMS will converge very fast. On the other hand, if the plant has a high-pass nature and the input psd is small for high frequencies, LMS will be very slow. In channel equalization applications the spectrum of the optimum solution is the opposite of the channel frequency response; hence, if we initiliaze the weight vector to zero, (25) will be small implying a slow convergence of LMS, a phenomenom often observed when using LMS in equalization or adaptive inverse control problems. We illustrate these ideas with two examples in the next section.

V. Simulations

Consider the system identification problem depicted in Fig. 4. The additive plant noise is independent of the input. Both, the adaptive filter and the plant consists of 16 taps, so the spectrum of the optimum solution coincides with the plant frequency response. The weight vector is initialized to zero and adapted for 800 iterations. The learning rate is set so that $\mu \operatorname{Tr}(R) = 0.05$. The power spectrum of the input and the frequency response of the plant to be identified are shown in Fig. 5, where a value of $M = 200$ was used to calculate $\Phi_{xx}[m]$ and $|\boldsymbol{W}^*[m]|^2$. The eigenvalue spread for this exercise was 5709. By inspection of the input psd and the frequency response of the plant we expect the LMS Transient Efficiency to be bigger than one, indicating that LMS should perform better than LMS/Newton, which is confirmed by their learning curves shown in Fig. 6. The estimate for the LMS Transient efficiency obtained from the simulation was 2.064, very close to the approximation of 2.069 given by (25).

To illustrate what happens if we use zero initial conditions when the input psd and the Wiener solution spectrum are very distinct, consider the simple equalization problem depicted in Fig. 7. The input to the channel is a white signal, hence the input psd for the adaptive filter is given by the magnitude squared of the channel frequency response. The channel frequency response and spectrum of the wiener solution for the adaptive equalizer are shown in Fig. 8 where a value of $M = 200$ was used. The eigenvalue spread of the input to the adaptive filter was 75.4. As intuitively expected, the input

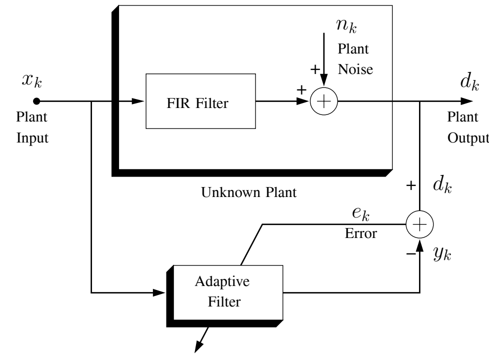

Fig. 4. System Identification Example

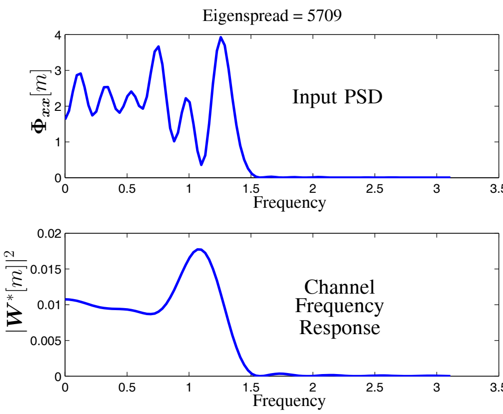

Fig. 5. Input Power Spectrum and Plant Frequency Response

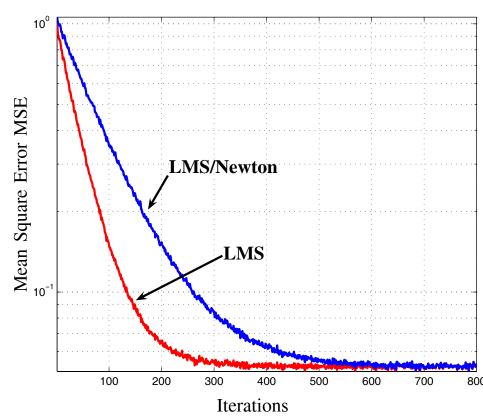

Fig. 6. Learning curves for System Identification example

psd and Wiener spectrum have opposite shapes resulting in a small LMS Transient Efficiency,indicating that LMS should perform worse than LMS/Newton, which was corroborated by their learning curves shown in Fig. 9. The estimate for the LMS Transient efficiency obtained from the simulation was 0.132, very close to the approximation of 0.137 given by (25).

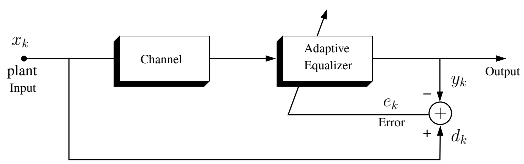

Fig. 7. Equalization Example

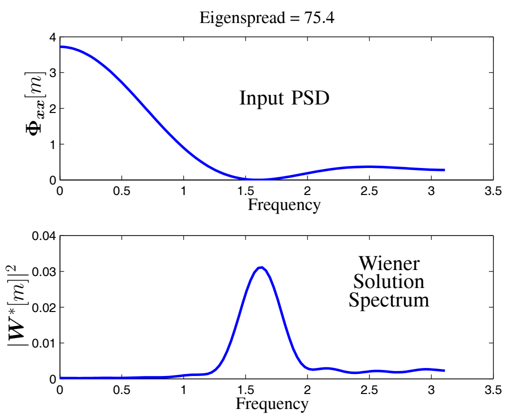

Fig. 8. Frequency response of plant and optimum equalizer

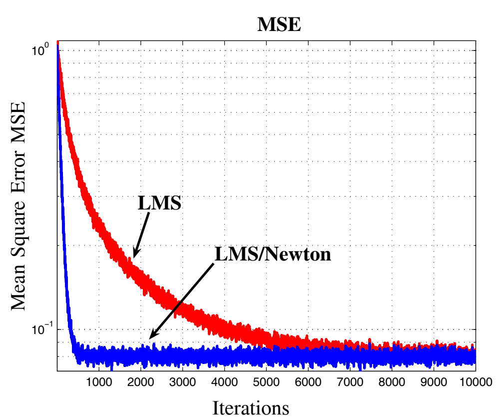

Fig. 9. Learning curves for equalization example

VI. Conclusions and Future Work

We have shown that when the LMS algorithm is used to train a transversal adaptive filter with jointly stationary input and desired response signals, its transient performance can be assesed by the inner product between the input psd and the spectrum of the initial weight vector deviation from the Wiener solution. This implies that when the initial weight vector is set to zero, the transient efficiency of the LMS algorithm can be predicted from approximate knowledge of the input psd and the frequency response of the Wiener solution. We described how this theory can be applied to system identification and equalization tasks; however, our results can be useful in predicting the LMS performance in any application where the input autocorrelation matrix is toeplitz and spectral information about the input and optimum solution is available.

An extension of this analysis is being made to nonstationary conditions, where the spectral characteristics of the input and optimum solutions keep a general shape, say both being always low pass. A similar analysis based on the Mean Square Deviation (MSD) instead of the MSE is also currently investigated.

References

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

[2] B. Widrow, J. M. McCool, M. G. Larimore, and J. C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter.," *Proceedings of the IEEE*, vol. 64, no. 8, pp. 1151 – 62, Aug. 1976.

[3] L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-29, no. 3, pp. 722–736, June 1981.

[4] B. Widrow and E. Walach, "On the statistical efficiency of the lms algorithm with nonstationary inputs.," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 211 – 221, 1984.

[5] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-33, no. 1, pp. 222–229, Feb. 1985.

[6] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, Upper Saddle River, NJ, 1985.

[7] F. M. Boland and J. B. Foley, "Stochastic convergence of the lms algorithm in adaptive systems.," *Signal Processing*, vol. 13, no. 4, pp. 339 – 52, DEC 1987.

[8] B. F. Boroujeny, "On statistical efficiency of the lms algorithm in system modeling.," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1947 – 51, MAY 1993.

[9] D. R. Morgan, "Slow asymptotic convergence of lms acoustic echo cancelers.," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 2, pp. 126 – 36, MAR 1995.

[10] S. C. Douglas and W. M. Pan, "Exact expectation analysis of the lms adaptive filter.," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2863 – 71, DEC 1995.

[11] B. Hassibi, A. H. Sayed, and T. Kailath, "h^∞ optimality of the lms algorithm.," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 267 – 80, FEB 1996.

[12] M. Reuter and J. R. Zeidler, "Nonlinear effects in lms adaptive equalizers.," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1570 – 9, JUN 1999.

[13] K. J. Quirk, L. B. Milstein, and J. R. Zeidler, "A performance bound for the lms estimator.," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 1150 – 8, MAY 2000.

[14] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters.," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314 – 24, FEB 2001.

[15] H. J. Butterweck, "A wave theory of long adaptive filters.," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 6, pp. 739 – 47, JUN 2001.

[16] O. Dabeer and E. Masry, "Analysis of mean-square error and transient speed of the LMS adaptive algorithm," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1873–1894, July 2002.

[17] B. Widrow and M. Kamenetsky, "Statistical efficiency of adaptive algorithms.," *Neural Networks*, vol. 16, no. 5/6, pp. 735 – 44, JUN-JUL 2003.

[18] S. Haykin and B. Widrow, *Least-mean-square adaptive filters*, Wiley-Interscience, Hoboken, N.J., 2003.