Fast System Identification Using Prominent Subspace LMS

Rongshan Yu*, Ying Song, Milashini Nambiar

*Signal Processing Department, Institute for Infocomm Research, A*STAR, Singapore.*

Abstract

In this paper, we propose prominent subspace least-mean-square (PS-LMS) algorithms for fast identification of time-varying system. It is shown that the dimensionality of system identification can be dramatically reduced if the unknown system is sparse in the sense that its parameter set has a skewed statistical distribution when expressed in a proper basis. In such cases, the system identification can be effectively carried out in a prominent subspace without introducing significant modeling error. A PS-LMS algorithm, that exploits this property is proposed first. The proposed algorithm can significantly improve the convergence speed of the traditional LMS algorithm if the unknown systems are sparse and have long impulse responses in the time-domain. To reduce the modeling error of PS-LMS introduced by dimension reduction, an enhanced PS-LMS (PS-LMS+) algorithm is further proposed. It is shown that PS-LMS+ is able to reduce the modeling error of PS-LMS while preserving its fast convergence property. Finally, experiments

*Corresponding author.

Email addresses: ryu@i2r.a-star.edu.sg (Rongshan Yu), sying@i2r.a-star.edu.sg (Ying Song), nambiarm@i2r.a-star.edu.sg (Milashini Nambiar)

were conducted to compare the performances of PS-LMS and PS-LMS+ with those of conventional LMS, recursive least squares (RLS), proportional normalized LMS (PNLMS), improved PNLMS (IPNLMS) and $\mu-$law PNLMS (MPNLMS) algorithms on systems of different levels of sparseness in transform domain, and the results confirm the superiority of the proposed algorithms for sparse systems.

Keywords:
Adaptive filter, least-mean-square (LMS), system identification, singular value decomposition.

---

1. Introduction

System identification (Ljung, 1999) has been a central issue in various application areas such as control, channel equalization, echo cancellation, active noise cancellation (ANC), image restoration, and seismic system identification. In many of these applications, the system can be characterized by a linear finite impulse response (FIR) filter. In such cases, the system identification problem is simplified to the problem of identifying the unknown impulse response from the system input-output data, for which solutions are readily given by adaptive filtering techniques (Haykin, 2001). Among them, Widrow and Hoff's least-mean-square (LMS) algorithm (Widrow, 1970) and its variations, such as normalized LMS (NLMS) (Nagumo and Noda, 1967), have been widely used due to their simplicity and robustness. Despite their popularity, however, LMS algorithms in their simplest forms suffer from several important limitations and drawbacks, namely, unsatisfactory convergence rates and asymptotic performance, particularly when the unknown

system has a long impulse response or inputs are highly colored (Widrow et al., 1976).

Recently, a significant amount of effort has been devoted to improving the convergence behavior of LMS for systems with sparse impulse responses, i.e., impulse responses that have only a few large coefficients interspersed among many negligible ones. Typical examples of such systems include circuit echo paths within telephony networks (Sondhi and Berkley, 1980), room acoustic echo paths (Allen and Berkley, 1979), and mobile radio channels (Failli, 1989). Several approaches have been proposed to exploit this prior information of system sparsity to improve the performance of LMS, including systems that apply a subset selection scheme based on either statistical detection of active taps (Kawamura and Hatori, 1986; Homer et al., 1998; Li et al., 2006) or subsequential partial updating (Etter, 1985; Godavarti and Hero, 2005) during the filtering process, systems that assign proportional step sizes of different taps according to their magnitudes (Gay, 1998; Duttweiler, 2000), and systems that introduce an ℓ_1 or ℓ_0 norm penalty on the coefficients into the quadric cost function of standard LMS (Chen et al., 2009; Gu et al., 2009).

In this paper, we study a different source of sparsity of the unknown system, namely, sparsity in the transform domain, to improve the convergence behavior of LMS. We observe that due to the presence of physical or electronic constraints, some natural systems may have a limited degree of freedom compared to its length. These systems are sparse in the sense that their parameter sets have concise representations when expressed in a proper basis even if they are spread out in the time-domain. A typical ex-

ample of such a system is the ANC headset (Kuo and Morgan, 1996), where the variations of the electro-acoustic path between the loudspeaker and the error microphone are by and large constrained by the design and wearing position of the headset (Kuo et al., 2006). It is apparent that these kinds of constraints will introduce significant dependencies among the parameters of the unknown system. As a result, the distributions of the parameters may be highly skewed, and only have significant variations in a subspace of reduced dimension. We call this subspace the prominent subspace of the unknown system. Based on this observation, we further introduce two new LMS algorithms, namely, the prominent subspace LMS (PS-LMS) and the enhanced prominent subspace LMS (PS-LMS+), to exploit this sparsity and improve the performance of the LMS algorithm for system identification. In the proposed algorithm, the prominent subspace of the system of interest is first established empirically based on off-line measurements of the response of the system. Subsequently, in PS-LMS, the LMS adaptation is performed only in the prominent subspace to reduce the dimensionality of the problem for faster convergence. It is shown that PS-LMS can effectively improve the convergence behavior of LMS, and that the modeling error introduced due to dimension reduction is small for sparse systems. Meanwhile, PS-LMS will introduce modeling error due to the dimension reduction operation in the LMS adaptation. To mitigate the modeling error issue, in PS-LMS+ the LMS adaptation is performed in both prominent subspace and its complement, albeit with adaptation factors determined so as to optimize the convergence rate. It is shown that PS-LMS+ is able to achieve small modeling errors compared to PS-LMS and meanwhile preserves its fast convergence

property.

The PS-LMS and PS-LMS+ algorithms proposed in this paper are conceptually different from reduced-rank adaptive filtering technologies based on either eigen-decomposition (Scharf and Tufts, 1987; Gabriel, 1986; Haimovich and Bar-Ness, 1991; Goldstein and Reed, 1997), or the multistage Wiener filter (Goldstein et al., 1998; Honig and Goldstein, 2002) despite their similarities in the adaptive filtering operation. In those prior-art works, the adaptive filtering is performed on a subspace spanned by either a subset of the eigenvectors of the covariance matrix of observed inputs, or on the Krylov subspace obtained by a successive orthonormalization process. The focus was thus to identify a subspace such that the mean square error (MSE) is minimized, and prior knowledge of the distribution of the unknown system's parameters is not assumed. On the contrary, the proposed PS-LMS and PS-LMS+ algorithms are based on the assumption that the parameters of unknown time-varying system are drawn from a non-white process, and the adaptive filtering operation is then performed in a subspace spanned by a subset of the eigenvectors of the covariance matrix of the system parameters. The proposed algorithms improve the convergence behavior of LMS algorithms even when the input process is white, which is impossible for traditional fast LMS methods that rely on the existence of correlations among the input vectors.

Certain results in the paper were published in preliminary form in (Yu et al., 2012). In this paper, the design principles of PS-LMS and PS-LMS+ are illustrated more clearly by contrasting them to traditional deficient-length LMS (Mayyas, 2005), PNLMS (Gay, 1998; Duttweiler, 2000), IPNLMS (Ben-

esty and Gay, 2002) and MPNLMS (Deng and Doroslovacki, 2005), and the merits of both algorithms are analyzed in more detail. Furthermore, the simulation part has been expanded to include realistic scenarios such as those from echo-path models (G.168, 2007). In addition, the adaptation step sizes in PS-LMS+ are determined systematically rather than empirically, and a low complexity implementation of PS-LMS+ is also given. The rest of this paper is organized as follows: Section 2 introduces the concept of the prominent subspace and the design of the PS-LMS for sparse system identification. Section 3 illustrates how to further reduced modelling error by introducing enhanced PS-LMS. In Section 4, the three use case experiments results are presented to compare the performances of PS-LMS and PS-LMS+ with the conventional LMS, RLS, PNLMS, IPNLMS and MPNLMS algorithms. Finally, conclusions are made in Section 5.

2. Prominent Subspace LMS

In this section, we introduce prominent subspace concept and provide the design of the PS-LMS.

2.1. Prominent Subspace of a Sparse System

We start from a general case where the set of all possible realizations of an unknown variable system H is characterized by a $L-$order FIR filter coefficient vector $\mathbf{h} \triangleq [h_1, h_2, \cdots, h_L]^T$ in the Euclidean space $\mathbb{R}^L$. Eigendecomposition of the system covariance matrix $\mathbf{R_h} \triangleq E_{\mathbf{h}}[\mathbf{h}\mathbf{h}^T]$, where $E_{\mathbf{h}}[\cdot]$ is statistical expectation with respect to the distribution of $\mathbf{h}$, yields

$$\mathbf{R_h} = \mathbf{V}\mathbf{\Lambda_h}\mathbf{V}^T, \tag{1}$$

where $\mathbf{V} \triangleq [\mathbf{v}_1, \ldots, \mathbf{v}_{_L}]$ denotes an orthogonal matrix of eigenvectors $\{\mathbf{v}_i \in \mathbb{R}^L\}$, and $\mathbf{\Lambda_h} \triangleq \text{diag}\,[\lambda_1, \ldots, \lambda_{_L}]$ denotes a diagonal matrix of eigenvalues of $\mathbf{R_h}$. Due to possible constraints $\mathbf{h}$ may not be uniformly distributed in $\mathbb{R}^L$. As a result, the eigenvalues of the system covariance matrix are in general not equal. Without loss of generality, we assume the eigenvalues are sorted in descending order in (1), thus $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{_L}$.

The unknown system $\mathbf{h}$ can be expanded in the orthonormal basis $\{\mathbf{v}_i\}_{i=1}^L$ as:

$$\mathbf{h} = \mathbf{V}\mathbf{c}, \tag{2}$$

where $\mathbf{c}$ is the projection of the coefficient vector $\mathbf{h}$ onto the row space of $\mathbf{V}$, i.e., $\mathbf{c} \triangleq \mathbf{V}^T\mathbf{h}$. Clearly, it is possible to perform system identification within a subspace of $\mathbb{R}^L$ in which the unknown system has significant variances to reduce the dimensionality of the identification problem without introducing significant modeling error. More formally, it can be shown that the average modeling error is minimized if the system identification is performed in the subspace spanned by the N eigenvectors $\{\mathbf{v}_i\}_{i=1}^N$ that correspond to the N largest eigenvalues of the covariance matrix $\mathbf{R}_h$.

Denoting $\mathbf{V}_{_N} \triangleq [\mathbf{v}_{_1}, \ldots, \mathbf{v}_{_N}]$, the least squares (LS) modeling of $\mathbf{h}$ is simply its projection onto the span of $\mathbf{V}_{_N}$:

$$\bar{\mathbf{h}} = \mathbf{V}_{_N}\mathbf{V}_{_N}^T\mathbf{h}. \tag{3}$$

The average modeling error $\bar{J}$ due to the projection is thus the variance of $\mathbf{h}$

in the orthogonal complement of span$\mathbf{V}_{N}$, which is given by:

$$\begin{aligned} \bar{J} &\triangleq E\left[\|\mathbf{h}-\bar{\mathbf{h}}\|^2\right] \\ &= \mathrm{tr}\left(\widetilde{\mathbf{V}}_{\mathrm{N}}^{\mathrm{T}}\mathbf{R}_{\mathrm{h}}\widetilde{\mathbf{V}}_{\mathrm{N}}\right) \\ &= \sum_{i=N+1}^{L}\lambda_i, \end{aligned} \tag{4}$$

where $\widetilde{\mathbf{V}}_{N} \triangleq [\mathbf{v}_{N+1},\ldots,\mathbf{v}_{L}]$. Clearly, for systems that are sparse in the sense that $\lambda_i \approx 0$, $N < i \leq L$, the system identification can be effectively performed in span $\mathbf{V}_{N}$ without introducing significant modeling error. We will call the subspace $\Psi \triangleq \mathrm{span}V_{N}$ the prominent subspace of the unknown system H, and system models $\{\bar{\mathbf{h}} \in \Psi\}$ N-sparse because they have at most N nonzero entries if they are expanded in $\mathbf{V}$.

2.2. LMS Adaptation in Prominent Subspace

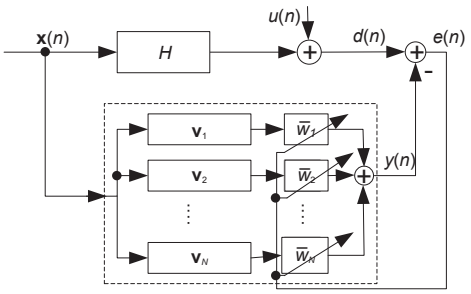


Figure 1: Block diagram of PS-LMS for system identification.

We will now turn our attention to the problem of performing system identification in the prominent subspace. As shown in Fig. 1, let $\mathbf{x}(n) \triangleq [x(n), x(n-1), \cdots, x(n-L+1)]^T$ be the input signal and $d(n)$ be the observed output signal of the unknown system. system identification is now achieved

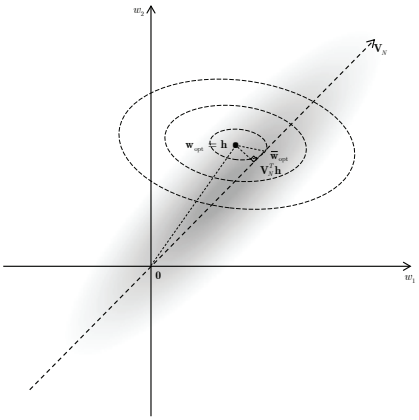


Figure 2: System identification bias of Wiener solution. The MMSE solution $\bar{\mathbf{w}}_{\text{opt}}$ is, in general, not equal to the LS modeling of system in the primary subspace.

by adjusting the weight coefficients $\{\bar{w}_i\}_{i=1}^N$ of the orthonormal basis $\{\mathbf{v}_i\}_{i=1}^N$ such that the MSE of the error signal $e(n)$ is minimized. Denote $\bar{\mathbf{w}}(n) \triangleq [\bar{w}_1(n), \ldots, \bar{w}_N(n)]^T$ and $\bar{\mathbf{x}}(n) \triangleq \mathbf{V}_N^T \mathbf{x}_(n)$, the error signal $e(n)$ is given by:

$$e(n) = d(n) - \hat{d}(n), \tag{5}$$

where $\hat{d}(n) \triangleq \bar{\mathbf{x}}(n)^T \bar{\mathbf{w}}(n)$ is the output of the LMS filter. The Wiener solution that minimizes the MSE of $e(n)$ is thus:

$$\bar{\mathbf{w}}_{\text{opt}} = \mathbf{R}_{NN}^{-1} \mathbf{p}_N, \tag{6}$$

where $\mathbf{R}_{NN}$ and $\mathbf{p}_N$ are, respectively, the covariance matrix of $\bar{\mathbf{x}}(n)$ and the cross-correlation vector between $\bar{\mathbf{x}}(n)$ and $d(n)$ defined as follows:

$$\mathbf{R}_{NN} \triangleq E[\bar{\mathbf{x}}(n)\bar{\mathbf{x}}^T(n)], \tag{7}$$

and,

$$\mathbf{p}_N \triangleq E[d(n)\bar{\mathbf{x}}(n)]. \tag{8}$$

The Wiener solution $\bar{\mathbf{w}}_{\text{opt}}$ is an MMSE projection of the unknown system to the prominent subspace Ψ with respect to the error signal $e(n)$. Therefore, it is in general different from the LS modeling of $\mathbf{h}$ in Ψ, which is given by $\bar{\mathbf{c}} \stackrel{\Delta}{=} \mathbf{V}_N^T\mathbf{h}$, due to the coupling of the input signal in the prominent subspace Ψ and its orthogonal complement. These two different projections are illustrated in Fig. 2. More specifically, we note that

$$\begin{aligned} d(n) &= \mathbf{x}^T(n)\mathbf{h} + u(n) \\ &= \mathbf{x}^T(n)\mathbf{V}\mathbf{c} + u(n) \\ &= \begin{bmatrix} \bar{\mathbf{x}}^T(n) & \tilde{\mathbf{x}}^T(n) \end{bmatrix} \begin{bmatrix} \bar{\mathbf{c}} \\ \tilde{\mathbf{c}} \end{bmatrix} + u(n) \\ &= \bar{\mathbf{x}}^T(n)\bar{\mathbf{c}} + \tilde{\mathbf{x}}^T(n)\tilde{\mathbf{c}} + u(n), \end{aligned} \tag{9}$$

where, $\tilde{\mathbf{x}}(n)$ and $\tilde{\mathbf{c}}$ are, respectively, the projections of input signal $\mathbf{x}(n)$ and system coefficient vector $\mathbf{h}$ onto the row space of $\widetilde{\mathbf{V}}_N$, i.e., $\tilde{\mathbf{x}}(n) \stackrel{\Delta}{=} \widetilde{\mathbf{V}}_N^T\mathbf{x}(n)$ and $\tilde{\mathbf{c}} \stackrel{\Delta}{=} \widetilde{\mathbf{V}}_N^T\mathbf{h}$. Replacing (9) into (8) and noting that $E[u(n)\mathbf{x}(n)] = \mathbf{0}$ we have

$$\mathbf{p}_N = \mathbf{R}_{NN}\bar{\mathbf{c}} + \mathbf{R}_{N\tilde{N}}\tilde{\mathbf{c}}, \tag{10}$$

where $\mathbf{R}_{N\tilde{N}} \stackrel{\Delta}{=} E[\bar{\mathbf{x}}(n)\tilde{\mathbf{x}}^T(n)]$ is the covariance matrix of $\bar{\mathbf{x}}(n)$ and $\tilde{\mathbf{x}}(n)$. It now follows from (6) that

$$\bar{\mathbf{w}}_{\text{opt}} = \bar{\mathbf{c}} + \mathbf{R}_{NN}^{-1}\mathbf{R}_{N\tilde{N}}\tilde{\mathbf{c}}. \tag{11}$$

Clearly, it can be seen that $\bar{\mathbf{w}}_{\text{opt}}$ is in general deviated from the LS estimation of unknown system in Ψ. However, it can be seen that the LS estimation of the unknown system is given by the Wiener solution when $\mathbf{R}_{N\tilde{N}} = 0$, which is

typically the case when the input signal $x(n)$ is white. The Wiener solution is exact when the system is N-sparse, i.e., when $\tilde{\mathbf{c}} = \mathbf{0}$ regardless of the colorness of the input signal.

Similar to the conventional LMS algorithm, the optimal Wiener solution can be found adaptively through the steepest descent algorithm using measured gradients of squares of $e(n)$. The resulting prominence subspace LMS (PS-LMS) algorithm is given as follows:

$$\bar{\mathbf{w}}(n+1) \;\; = \;\; \bar{\mathbf{w}}(n) + \bar{\mu}\bar{\mathbf{x}}(n)e(n). \tag{12}$$

Here $\bar{\mu}$ is the step size of the LMS adaptation. By noting that the time-domain system model is given by $\mathbf{w}(n) = \mathbf{V}_N\bar{\mathbf{w}}(n)$, the PS-LMS algorithm in (12) can be equivalently expressed in the time-domain as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \bar{\mu}\mathbf{V}_{_N}\mathbf{V}_{_N}^T\mathbf{x}(n)e(n), \tag{13}$$

which clearly demonstrates that the PS-LMS algorithm is obtained by projecting the gradient vector $\mathbf{x}(n)e(n)$ of conventional LMS onto Ψ during the adaptation process. Moreover, supposing the unknown coefficients are sparse in the time domain, i.e., the orthogonal basis $\mathbf{v}_i$ is the natural basis for a discrete time signal $\mathbf{e}_i$, the proposed algorithm is essentially the deficient-length LMS (Mayyas, 2005). In this respect, the proposed algorithm can be seen as a generalization of the deficient-length LMS where the adaptive filter length is less than that of the unknown system in a transform domain. In addition, following the mean square stability analysis in (Mayyas, 2005) it can be seen that mean-square convergence of the proposed algorithm is guaranteed if the step size $\bar{\mu}$ satisfies

$$0 < \bar{\mu} < \frac{2}{3\mathrm{tr}(\mathbf{R}_{_{\mathrm{NN}}})}. \tag{14}$$

It can be shown that the expected modeling error of PS-LMS is given as follows:

$$\begin{aligned} J_{\text{PS-LMS}} &= E\left[\|\mathbf{h} - \mathbf{w}(n)\|^2\right] \\ &= \bar{J} + J_{\Delta}(n), \end{aligned} \tag{15}$$

where

$$\bar{J} \triangleq E[\|\mathbf{h} - \mathbf{V}_{N}\bar{\mathbf{w}}_{\text{opt}}\|^2] \tag{16}$$

is the the modeling bias due to the projection onto prominent subspace, and

$$J_{\Delta}(n) \triangleq E[\|\mathbf{V}_{N}\bar{\nu}(n)\|^2] \tag{17}$$

is the modeling error due to the gradient noise of LMS algorithm. Here $\bar{\nu}(n) \triangleq \bar{\mathbf{w}}_{\text{opt}} - \bar{\mathbf{w}}(n)$ is the weight-error vectors. Replacing (11) into (16) it can be shown that $\bar{J}$ is given by:

$$\begin{aligned} \bar{J} &= E[\|(\mathbf{V}_{\tilde{N}} - \mathbf{V}_{N}\mathbf{R}_{NN}^{-1}\mathbf{R}_{N\tilde{N}})\tilde{\mathbf{c}}\|^2] \\ &= \text{tr}[\mathbf{B}\mathbf{B}^T\mathbf{R}_{\tilde{\mathbf{c}}}], \end{aligned} \tag{18}$$

where $\mathbf{B} \triangleq \mathbf{V}_{\tilde{N}} - \mathbf{V}_{N}\mathbf{R}_{NN}^{-1}\mathbf{R}_{N\tilde{N}}$, and $\mathbf{R}_{\tilde{\mathbf{c}}} \triangleq E[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T]$ is the covariance matrix of $\tilde{\mathbf{c}}$. Furthermore, if the observed data vector is white it can be further shown that $\mathbf{B}\mathbf{B}^T = \mathbf{I}$ and thus we have

$$\bar{J} = \sum_{i=N+1}^{L} \lambda_i. \tag{19}$$

On the other hand, denoting $\mathbf{R}_{\bar{\nu}}(n) = E[\bar{\nu}(n)\bar{\nu}(n)^T]$, it now follows from the standard LMS analysis that under mild conditions $J_{\Delta}(n)$ at steady-state is

given by:

$$\begin{aligned} J_{\Delta} &= \mathrm{tr}[\mathbf{R}_{\bar{\nu}}(n)] \\ &= \frac{1}{2}\bar{\mu}N\bar{\epsilon}_{\min}, \end{aligned} \tag{20}$$

where $\epsilon_{\min}$ is the minimum mean square error produced by the Wiener filter.

It can be seen for the above analysis that under the constraint of same LMS adaptation rate, increasing the dimension of the prominent subspace in PS-LMS may not necessarily lead to a smaller modeling error, which is proportional to the dimension of the LMS algorithm. It is thus an important design consideration to identify the optimal dimensions of the prominent subspace when it is practised in real-life applications.

3. Enhanced Prominent Subspace LMS

3.1. LMS with Fast Adaptation in Prominent Subspace

One of the limitations of PS-LMS is that it only converges to the projection of the system transfer function onto the prominent subspace, which may not be desirable for identification applications that are sensitive to modeling error. To overcome this limitation, the PS-LMS algorithm can be modified by adding small but non-zero gradient descent vectors in the null space of $\mathbf{V}_N$, which leads to the enhanced PS-LMS (PS-LMS+) algorithm given as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu'\mathbf{V}\mathbf{G}\mathbf{V}^T\mathbf{x}(n)e(n), \tag{21}$$

where $\mu' > 0$, and $\mathbf{G} \triangleq \mathrm{diag}(g_1, g_2, \ldots, g_L)$ is a diagonal matrix such that $\sum_i g_i = 1$. Furthermore, we require $g_i > g_{\min}, 0 < i \leq L$ where $g_{\min}$ is a

small positive value to ensure PS-LMS+ converging to the full-rank Wiener solution.

It can be seen that (21) can be equivalently written as:

$$\check{\mathbf{w}}(n+1) = \check{\mathbf{w}}(n) + \mu' \mathbf{G}\check{\mathbf{x}}(n)e(n), \tag{22}$$

where

$$\check{\mathbf{w}}(n) \triangleq \mathbf{V}^T \mathbf{w}(n), \tag{23}$$

and

$$\check{\mathbf{x}}(n) \triangleq \mathbf{V}^T \mathbf{x}(n). \tag{24}$$

From (22), it can be seen that PS-LMS+ is analogous to the PNLMS algorithm (Gay, 1998; Duttweiler, 2000) such that the available adaptation "power" as determined by g_i, $0 < i \leq L$ is distributed unevenly over the directions defined by $\mathbf{V}$. In this respect, PS-LMS+ can be taken as a generalization of PNLMS in the transform domain, but with an important difference: In PS-LMS+, the distribution of g_i's is determined based on *a priori* knowledge regarding the sparsity of the statistical distribution of the unknown system; while in PNLMS, such knowlege is not used. Instead, the distribution of g_i in PNLMS is determined according to the intermediate estimated tap-weight, which may be taken as *a posteriori* knowledge of the distribution of the unknown system's transfer function.

3.2. Optimal Adaptation Step Size for PS-LMS+

We are now turn our attention to the problem of determining the optimal selection of g_i's for fast system convergence. Intuitively, larger g_i's should be

used in the directions of prominent subspace where unknown system under identification are likely to have larger coefficients so that PS-LMS+ can converge more quickly. Here, we propose an algorithm to determine g_i towards the desired convergence behavior of PS-LMS+ given the priori knowledge of the distribution of the unknown system.

Consider the weight-error vector of PS-LMS+ in the transform domain after n iterations, which is denoted by $\check{\nu}(n) \triangleq \check{\mathbf{w}}_{\text{opt}} - \check{\mathbf{w}}(n)$ where $\check{\mathbf{w}}_{\text{opt}} \triangleq \mathbf{V}^T\mathbf{w}_{\text{opt}}$. Assuming an initial guess $\mathbf{w}(0) = 0$, the initial error is given by $\check{\nu}(0) = \check{\mathbf{w}}_{\text{opt}}$. From (22) it is straightforward to show:

$$\check{\nu}(n+1) = \check{\nu}(n) - \mu'\mathbf{G}\check{\mathbf{x}}(n)\left[\check{\mathbf{x}}^T(n)\check{\nu}(n) + u(n)\right]. \tag{25}$$

Let $\xi(n) = E[\check{\nu}(n)]$. Assuming that the input signal is white and has unit variance, taking the expectation of both sides of (25), it can be shown that the following relationship holds for the expected values of weight-error vectors:

$$\xi(n+1) = \xi(n) - \mu'\mathbf{G}\xi(n) = (\mathbf{I} - \mu'\mathbf{G})\xi(n), \tag{26}$$

which implies:

$$\xi(n) = (\mathbf{I} - \mu'\mathbf{G})^n\check{\mathbf{w}}_{\text{opt}}. \tag{27}$$

Now, we want to determine a set of $\{g_i\}_{i=1}^{L}$ such that the PS-LMS+ algorithm approaches as quickly as possible to a pre-determined condition. For convenience, we consider mean square ℓ_∞ norm of weigh-error vector in the orthonormal basis $\mathbf{V}$ here although it is also possible to use other types of norms as well. The mean square ℓ_∞ norm of weigh-error vector is given

as follows

$$\begin{aligned} E_h[\|\xi(n)\|_\infty^2] &= E_h[\max_i\{(1-\mu' g_i)^{2n}|\mathbf{v}_i^T\mathbf{h}|^2\}] \\ &= \max_i (1-\mu' g_i)^n E_h[|\mathbf{v}_i^T\mathbf{h}|^2] \\ &= \max_i (1-\mu' g_i)^{2n}\lambda_i. \end{aligned}$$

The problem of determining $\{g_i\}_{i=1}^L$ can be formulated as an optimization problem that minimizes some $y \in \mathbb{R}$ such that $E[\|\xi(y)\|_\infty^2] \leq \epsilon^2$ for some $\epsilon > 0$, which can be written as follows:

$$\begin{aligned} \min_{g_i, y} \quad & y; \\ \text{subject to} \quad & \sum g_i = 1; \\ & g_i \geq g_{\min}, \quad 0 < i \leq L; \\ \text{and} \quad & (1-\mu' g_i)^{2y}\lambda_i \leq \epsilon^2, \quad 0 < i \leq L. \end{aligned} \tag{28}$$

The problem of (28) is, unfortunately, not a convex optimization problem. However, it can be shown the constraint on y above can be rewritten as:

$$y \geq \frac{1}{2}\frac{\log(\epsilon^2/\lambda_i)}{\log(1-\mu' g_i)}, \forall i : \lambda_i > \epsilon^2, \tag{29}$$

which can be further rewritten as:

$$\log y \geq \log(-\log(\frac{\epsilon^2}{\lambda_i})) - \log(-\log(1-\mu' g_i)) + \log\frac{1}{2}, \forall i : \lambda_i > \epsilon^2. \tag{30}$$

Therefore, the above optimization can be reformulated as

$$\begin{aligned} \min_{g_i} \quad & \max_{\{i:\lambda_i > \epsilon^2\}} \log(-\log(\frac{\epsilon^2}{\lambda_i})) - \log(-\log(1-\mu' g_i)); \\ \text{subject to} \quad & \sum g_i = 1; \\ \text{and} \quad & g_i \geq g_{\min}, 0 < i \leq L. \end{aligned} \tag{31}$$

It can be easily verified the above optimization problem is convex when $\mu' < 0.5$, and hence it can be solved efficiently by numerical methods such as the interior-point method (Nemirovski and Todd, 2008).

3.3. Convergence of PS-LMS+

It can be shown that PS-LMS+ converges to the full-rank Wiener solution in mean under mild conditions. Specifically, following the analysis in (Duttweiler, 2000) it can been shown that convergence of PS-LMS+ is guaranteed if the following conditions are held:

1. $x(n)$ is an independent and identically distributed Gaussian process.
2. $\mathbf{w}(n)$, $x(n)$, and $u(n)$ are mutually independent.
3. $0 < \mu' < \mu_{\max}$ where $\mu_{\max}$ is the solution of the following equation:

$$\mu \sum_{n=1}^{L} \frac{g_i \sigma_x^2}{2 - 2\mu g_i \sigma_x^2} = 1. \tag{32}$$

Furthermore, it can be shown that under the same conditions the expected modeling error of PS-LMS+ at steady state is given by:

$$\begin{aligned} J_{\text{PS}-\text{LMS+}} &= E[\|\mathbf{w}_{\text{opt}} - \mathbf{w}(\infty)\|^2] \\ &= E[\|\mathbf{V}\check{\nu}(\infty)\|^2] \\ &= \text{tr}[\mathbf{R}_{\check{\nu}}(\infty)] \\ &= \frac{\Phi}{1-\Phi}\frac{\sigma_u^2}{\sigma_x^2}, \end{aligned} \tag{33}$$

where

$$\Phi \triangleq \sum_{i=1}^{L} \frac{\mu' g_i \sigma_x^2}{2 - 2\mu' g_i \sigma_x^2}. \tag{34}$$

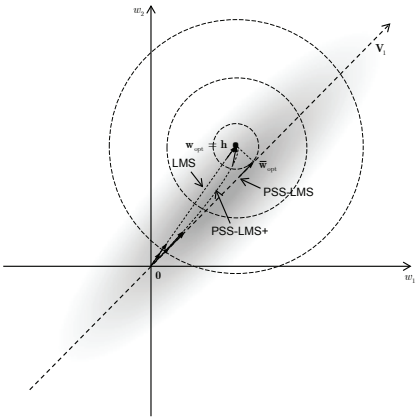


Figure 3: Convergence paths of LMS, PS-LMS, and PS-LMS+ algorithms. The shaded area indicates the distribution of the unknown system. Here we assume that the observed data is statistically white and hence the equal error contours of the MSE surface are circles.

Note that $\sum_{i=1}^{L} g_i = 1$ by construction. If μ' is small in comparison to one, or if all the g_n are small in comparison to one, then Φ is approximately equal to $\mu'\sigma_x^2/2$ and

$$J_{\text{PS}-\text{LMS}+} \approx \frac{\mu'\sigma_x^2}{2-\mu'\sigma_x^2}\frac{\sigma_u^2}{\sigma_x^2} \approx \frac{1}{2}\mu'\sigma_u^2. \tag{35}$$

Therefore, the modeling error at steady state is relevant only to the value of μ', and is independent from the distribution of g_i.

The different convergence behaviors of LMS, PS-LMS, and PS-LMS+ when applied to a system with skewed distribution is illustrated with a simple example in Fig. 3. This figure shows the convergence paths of LMS, PS-LMS and PS-LMS+ in a two-dimensional space. The probability distribution of the unknown system is indicated by the shaded area in the figure. It can be seen that the distribution is rather skewed, and most system coefficients reside near a one-dimensional prominent subspace indicated by $\mathbf{V}_1$. For a given unknown system $\mathbf{h}$, the LMS algorithm will converge to the unknown

system following the steepest descent path statistically. For the PS-LMS algorithm, the convergence is constrained onto the prominent subspace so it will only converge to a bias modeling of the unknown system. However, a faster convergence rate is expected for PS-LMS since a larger adaptation factor can be used due to the dimension reduction of PS-LMS. The PS-LMS+ algorithm will also converge to the unknown system, while following an accelerated convergence path since a larger adaptation factor is used in the prominent subspace. It is also interesting to note that the main difference of these LMS algorithms lies in their different choices of adaptation factors: the PS-LMS algorithm is obtained from PS-LMS+ if one sets $g_i = 0$ for $i \geq N$. On the other hand, if all the g_is are equal, the PS-LMS+ algorithm will be reduced to the normal standard LMS algorithm.

3.4. Complexity Analysis

The PS-LMS algorithm is preferably implemented in the transform domain (12) to exploit its reduced dimension feature for implementation cost reduction. Compared to the standard LMS algorithm, N additional linear filtering operations on the time-domain observed data are required in order to produce the transformed observed signal $\bar{\mathbf{x}}(n)$. Once they are obtained, the PS-LMS could then be implemented as a standard dimension-N LMS algorithm. The complexity of PS-LMS can be further reduced if fast algorithms are used to implement the linear filtering operation.

Brute-force implementation of PS-LMS+ requires L full-length filter operations in addition to a standard L-tap LMS adaptation operation, which is in generation too complex for real-life applications when L is large. However, if the adaptation step sizes in the non-prominent subspace are equal,

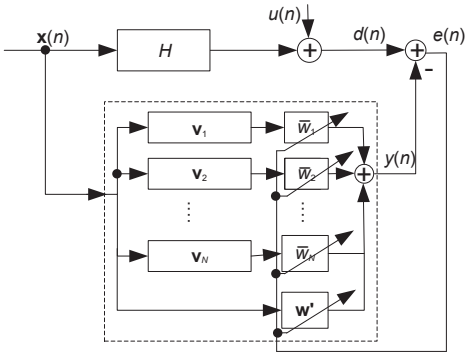


Figure 4: Implementation of PS-LMS+ using parallel PS-LMS and LMS algorithms.

Table 1: Comparison of computational complexity

Algorithm	Multiplications	Divisions	Additions
LMS	2L+1	0	2L
PNLMS	4L+4	1	4L-2
RLS	$2L^2$+7L+5	L^2+4L+3	$2L^2$+6L+4
PS-LMS	NL+2N+1	0	NL+N-1
PS-LMS+	(N+2)L+2N+2	0	(N+2)L+N-2

the adaptation algorithm in (36) can be implemented as parallel operating N-tap LMS in the transform domain and a standard LMS algorithm. To show this, we first rewrite (21) as

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) + \mu' \mathbf{V}_{N} \mathbf{G}_{N} \mathbf{V}_{N}^{T} \mathbf{x}(n) e(n) \\ &\quad + \mu' g_{N+1} (\mathbf{I} - \mathbf{V}_{N} \mathbf{V}_{N}^{T}) \mathbf{x}(n) e(n), \end{aligned} \tag{36}$$

where $\mathbf{G}_{N}$ is the diagonal matrix consisting of the first N entries of $\mathbf{G}$, and it is assumed that $g_{N+1} = g_{N+2} = \ldots = g_{L}$. We can decompose the weight vector $\mathbf{w}(n)$ as follows: $\mathbf{w}(n) = \mathbf{w}'(n) + \mathbf{V}_{N} \bar{\mathbf{w}}(n)$, where $\bar{\mathbf{w}}(n) = \mathbf{V}_{N}^{T} \mathbf{w}(n)$.

Then (36) can be readily rewritten as two parallel LMS algorithms:

$$\begin{bmatrix} \mathbf{w}'(n+1) \\ \bar{\mathbf{w}}(n+1) \end{bmatrix} = \begin{bmatrix} \mathbf{w}'(n) \\ \bar{\mathbf{w}}(n) \end{bmatrix} + \begin{bmatrix} \mu' g_{N+1} \mathbf{x}(n) e(n) \\ \mu' (\mathbf{G}_N - g_{N+1} \mathbf{I}) \mathbf{V}_N^T \mathbf{x}(n) e(n) \end{bmatrix}, \tag{37}$$

The output of PS-LMS+ can be similarly implemented as the summation of those from these two LMS algorithms as follows:

$$\hat{d}(n) = \bar{\mathbf{x}}(n)^T \bar{\mathbf{w}}(n) + \mathbf{x}(n)^T \mathbf{w}'(n), \tag{38}$$

which is illustrated in Fig. 4. The condition that $g_{N+1} = g_{N+2} = \ldots = g_L$ for some specific N can be enforced by incorporating it directly in the optimization problem (28). Alternatively, it is also possible to select an appropriate $g_{\min}$ such that the optimal solution from (28) satisfies $g_{N+1} = g_{N+2} = \ldots = g_L = g_{\min}$.

The computational complexities of LMS, RLS, PS-LMS, and PS-LMS+ are summarized in Table 1. As shown in this table, LMS has the lowest computational complexity and RLS, in general, has a much higher computational complexity. For both PS-LMS and PS-LMS+, the computational complexity increases linearly with the dimensionality of the prominent subspace N. It can also be observed that although the computational complexities of PS-LMS and PS-LMS+ are higher than that of LMS, for a sparse approximate system that can be effectively represented in a prominent subspace with a small N, the complexity increment is only moderate and justifiable considering the improved convergence performance.

4. Experiment Results

In this section, we present numerical simulation results on three different use cases to verify the performance of PS-LMS and PS-LMS+. We first introduce the impulse responses from three different use cases used in our experiments. We further show the results of eigenvector decomposition using ANC headset as an example. After that, we discuss the relationship between the dimension of the prominent subspace and the modeling error for PS-LMS algorithm. Finally, we compare the convergence behaviors of PS-LMS and PS-LMS+ with several algorithms in terms of convergence rate, such as LMS, RLS, PNLMS and its improved versions.

4.1. Experiments Setup

In the experiments, three different systems were used for performance comparison. They include ANC models representing the impulse responses of the electro-acoustic path between the loudspeaker and the microphone of an active noise control headset, echo models representing the impulse responses of telephone network echo path and random models whose impulses were generated with independent and identical Gaussian distributed data.

The ANC models are obtained by collecting $M = 640$ sets of FIR filter coefficients measured from the electro-acoustic path of an ANC headset when it was worn by different human subjects at normal wearing positions and positions that deviated slightly from the normal wearing position. The measurements were conducted in a quiet listening room. The sampling rate is 8 kHz and a 64 order FIR filter is used.

The echo models are generated numerically based on the ITU-T G.168

Recommendation (G.168, 2007) on speech echo cancellers. The echo path is simulated by linear digital filters with the impulse responses given by:

$$h_e(k) = 10^{-\mathrm{ERL}/20} K_i m_i(k - \delta_t), i = 1, \ldots, 8, \tag{39}$$

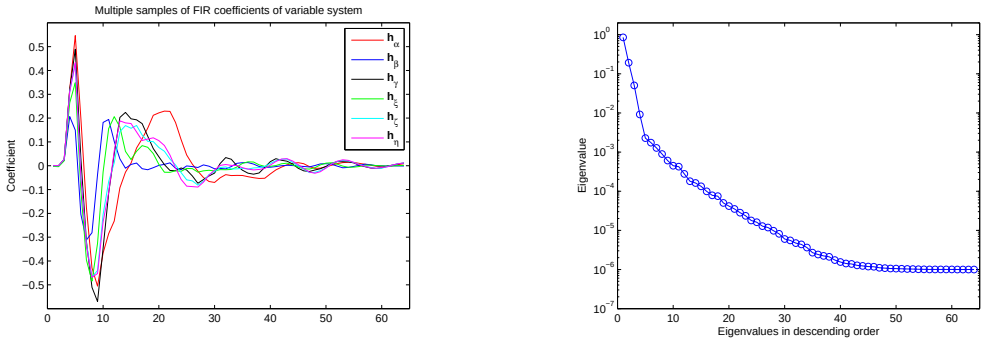
which is the delayed and attenuated version of the sequences $m_i(k), i = 1, \ldots, 8$. Here ERL is echo return loss and δ_t is the delay.

In our experiment, the echo models were generated using sequences 1-4 with uniformly distributed ERL and δ_t: ERL $\sim$ U($\mathrm{ERL}_{\min}, 2\mathrm{ERL}_{\min}$) and $\delta_t \sim U(0, 6)$ where $U(a, b)$ denotes uniform distribution between a and b and $\mathrm{ERL}_{\min}$ is minimum ERL with white noise input case for every sequence. The order of the resulting FIR filters is 134. After that, according to the dispersion time distribution in the ITU standard, we removed filters with dispersion times more than 12.5 ms, which corresponding to 100 samples for 8 kHz sampling rate. A total of $M = 640$ FIR filters were generated in our experiments.

Finally, for the random models, the tap-weights of FIR filters were generated using an independent and identical Gaussian distributed source with zero mean. The length of the filters is 64 and in total $M = 640$ filters were used in our experiments.

4.2. Eigenvector Decomposition

We assume that $\mathbf{h}_m, m = 1, \ldots, M$ are M sets of FIR filter coefficients collected from a time-varying system. The covariance matrix of the FIR coefficients from the system is then estimated by replacing the statistical



(a) (b)

Figure 5: (a) Examples of FIR filters from ANC models; (b) Eigenvalues of $\mathbf{R_h}$ in descend order for ANC models.

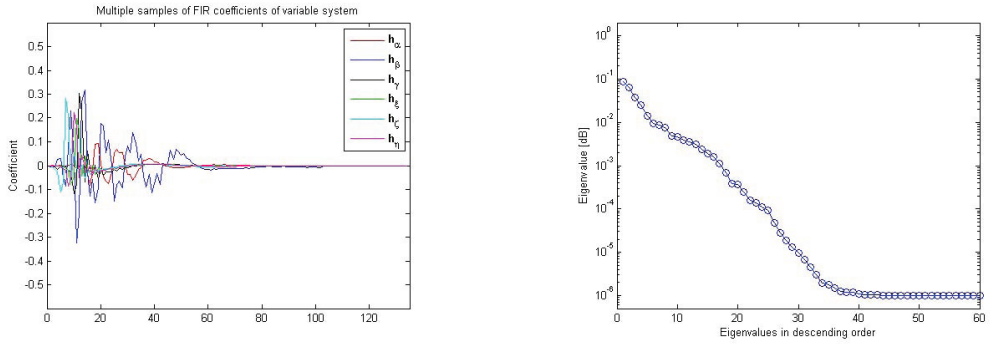
expectation with the time-average of all the observations:

$$\mathbf{R_h} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{h}_m\mathbf{h}_m^T. \tag{40}$$

Consequently, the eigenvalues and associated eigenvectors can be estimated from $\mathbf{R_h}$ as:

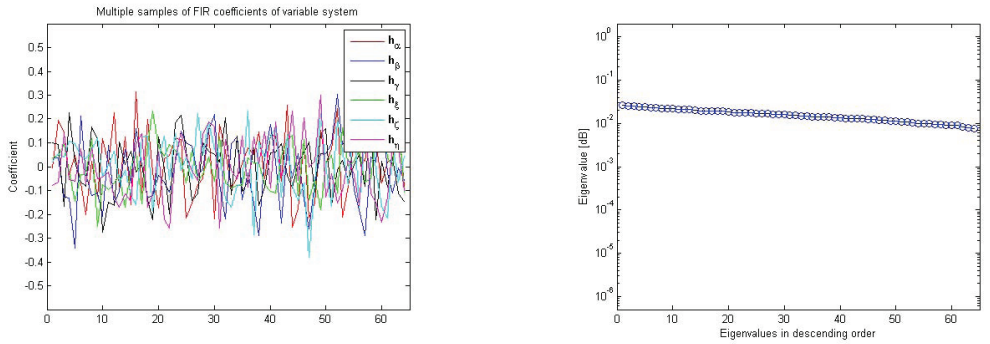
$$\mathbf{R_h}\mathbf{V} = \lambda\mathbf{V}. \tag{41}$$

The resulting eigenvalues of three different models are illustrated in Fig. 5, 6 and 7 for ANC, echo and random models respectively with examples of time-domain FIR filter coefficients from those models. It can be seen that these three systems show different levels of skewness in the distribution of eigenvalues. In particular, the sparseness of ANC models is very obvious as the first few largest eigenvalues are much larger than the rest of the eigenvalues. Clearly, it is possible to perform system identification in the prominent subspace spanned by the eigenvectors associated with those eigenvalues without introducing significant modeling error. The eigenvectors corresponding



(a) (b)

Figure 6: (a) Examples of FIR filters from echo models; (b) Eigenvalues of $\mathbf{R_h}$ in descend order for echo models.



(a) (b)

Figure 7: (a) Examples of FIR filters from random models; (b) Eigenvalues of $\mathbf{R_h}$ in descend order for random models

to first eight largest eigenvalues of ANC system are shown in Fig. 8 together with their corresponding eigenvalues. On the other hand, the echo models shows slightly moderate skewness in the distribution of eigenvalues, where a prominent subspace of higher dimensions are needed to capture the prominent components of an echo model. The eigenvalues of random models show

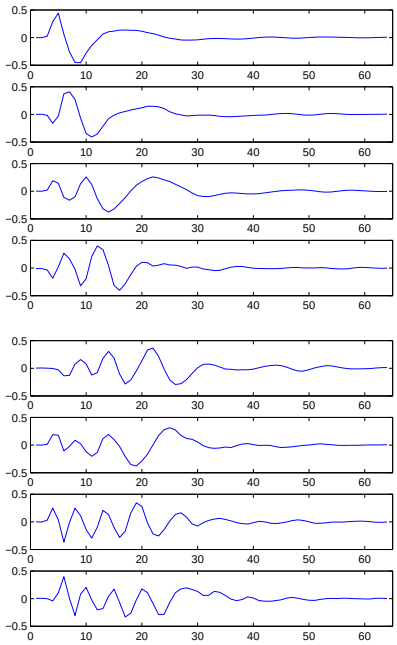


Figure 8: Eight eigenvector filters associated with eight largest eigenvalues. From top to bottom the eigenvalues are $\lambda_1 = 0.8525$, $\lambda_2 = 0.1918$, $\lambda_3 = 0.0503$, $\lambda_4 = 0.0091$, $\lambda_5 = 0.0023$, $\lambda_6 = 0.0017$, $\lambda_7 = 0.0013$, $\lambda_8 = 0.0009$.

the least skewness in the distribution. In fact, in this case, since the distribution of eigenvalues is very disperse in the transform domain, there shouldn't be any advantage of performing system identification in the prominent subspace.

It is interesting to note that although a system that is sparse in time-domain will be sparse in transform domain as well provided a proper basis is found, the converse is not necessarily true. A system that is sparse in the frequency domain can still have a impulse response that is very disperse in the time domain. To verify this, we measured the spareness levels of the

Table 2: Mean and range of sparseness levels in time-domain and transform-domain. Results are shown in "mean (min, max)"

	$\mathcal{S}(\mathbf{h})$	$\mathcal{S}(\mathbf{c})$
ANC	0.5703 (0.4419, 0.6668)	0.8820 (0.8209, 0.9421)
Echo	0.6038 (0.4323, 0.7893)	0.8131 (0.7203, 0.9223)
Random	0.2259 (0.1688, 0.2697)	0.2396 (0.2652, 0.4480)

above plants using the the spareness measurement proposed in (Hoyer, 2004):

$$\mathcal{S}(\mathbf{h}) = \frac{\sqrt{L} - \frac{\sum |h(i)|}{\sqrt{\sum h(i)^2}}}{\sqrt{L} - 1}, \tag{42}$$

and the results are compared to their spareness levels in the frequency-domain $\mathcal{S}(\mathbf{c})$ where $\mathbf{c} = \mathbf{V}^T\mathbf{h}$. The comparison results are given in Table 2, where the mean and the range of the levels of sparseness in both the time and transform domains are listed. From these results, it can be seen that most ANC and Echo filters are highly sparse after transformation although they are not sparse in the time-domain. On the other hand, as expected, the random filters are not sparse in both domains.

4.3. PS-LMS Modeling Error vs. Dimension

We further investigated the relationship between the modeling error of the PS-LMS algorithm and the dimension of the prominent subspace using an example from ANC models. In these experiments, we assume that both the input signal $x(n)$ and noise signal $u(n)$ are white noises with variance $\sigma_x^2 = 1$ and $\sigma_u^2 = 0.01$, respectively.

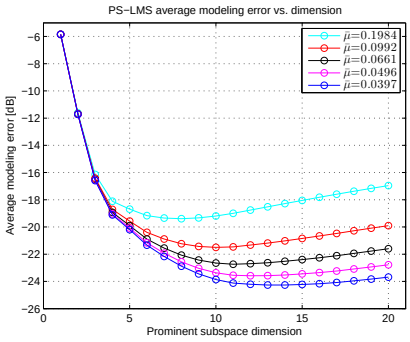


Figure 9: PS-LMS modeling error vs. dimension.

The results are given in Fig. 9, which clearly demonstrate the trade-off between the adaptation step size and the dimension of prominent subspace in PS-LMS. It can be observed that when the step size $\bar{\mu} = 0.1984$, average modeling error initially decreases with an increase in the dimension N and reaches its minimum -19.7 dB at $N = 7$. Beyond this point, increasing the dimension of the prominent subspace doesn't further reduce modeling error due to the presence of gradient noise. It is also observed that when the dimension is increased from $N = 1$ to $N = 2$, the average modeling error reduces significantly from -5.8 dB to -11.5 dB. A similar reduction of average modeling error is obtained when N is increased from 2 to 4. However, when N is increased from 4 to 5, the average modeling error reduction is not significant. Likewise, there is little reduction in the average modeling error when N is increased from 5 to 7. It can also been observed that for smaller LMS adaptation step sizes, using a higher dimensional prominent subspace in PS-LMS algorithm reduces the average modeling error. For example, if the adaptation step size is reduced to $\bar{\mu} = 0.0992$, the minimum average modeling error reaches -21.8 dB at dimension $N = 9$. Of course, both

decreasing the LMS adaptation step size and increasing the dimension of prominent subspace in PS-LMS algorithm will lead to slower convergence behavior which may not be desirable in real-life applications.

4.4. Effect of varying error ϵ on convergence of PS-LMS+

We investigated the effect of varying ϵ on the convergence behaviour of PS-LMS+ using a system constructed by cascading two randomly chosen FIR filters from ANC models. In this experiment, g_{min} was set to 0.01 and the same settings for input and noise signals from previous experiments were assumed. Fig. 10 shows that increasing the value of ϵ increases the rate of initial convergence of PS-LMS+, at the expense of convergence to steady state. For $\epsilon = 0.8$, however, PS-LMS+ converges to -15 dB within the first 200 iterations, and to steady state within the same number of iterations as for lower values of ϵ. Therefore, having the flexibility to modify the value of ϵ allows us to fine-tune the performance of PS-LMS+, and achieve a desirable trade-off between initial convergence and convergence to steady state. The distributions of g_i's under different ϵ are given in Fig. 11, which shows that larger ϵ will lead to more "adaptation energy" being allocated to directions of prominent subspace, thus increasing the initial convergence of PS-LMS+ at the expense of slower convergence speed at other directions.

4.5. Convergence Behavior

We now illustrate the convergence behavior of the PS-LMS and PS-LMS+ algorithms, comparing their performance with that of the conventional LMS, RLS, PNLMS, IPNLMS and MPNLMS algorithms. In our experiments, the unknown systems were constructed by cascading two randomly chosen FIR

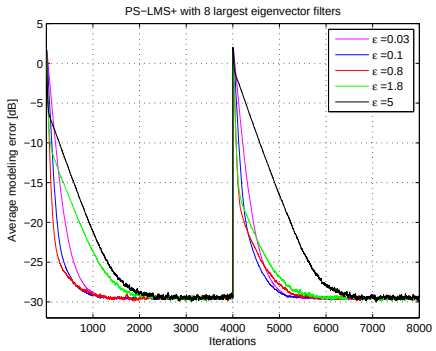


Figure 10: Convergence of PSLMS+ for different ϵ, with fixed $g_{min} = 0.01, \mu = 0.1984$

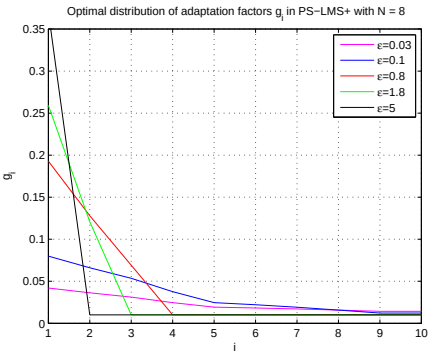


Figure 11: Optimal distribution of g_i for different ϵ, with fixed $g_{min} = 0.01, \mu = 0.1984$

filters from the M observations, i.e., the unknown system was started with a randomly picked transfer function $\mathbf{h}_\alpha$ and then switched to another randomly picked transfer function $\mathbf{h}_\beta$ at time t_0.

Fig.12 compares the performances of the different algorithms in terms of their convergence speed when they are applied to the constructed time-varying system of ANC model when SNR = 20 dB. The results are shown after averaging 200 independent runs. The x-axis of Fig. 12 represents the number of iterations and the y-axis represents the average modeling er-

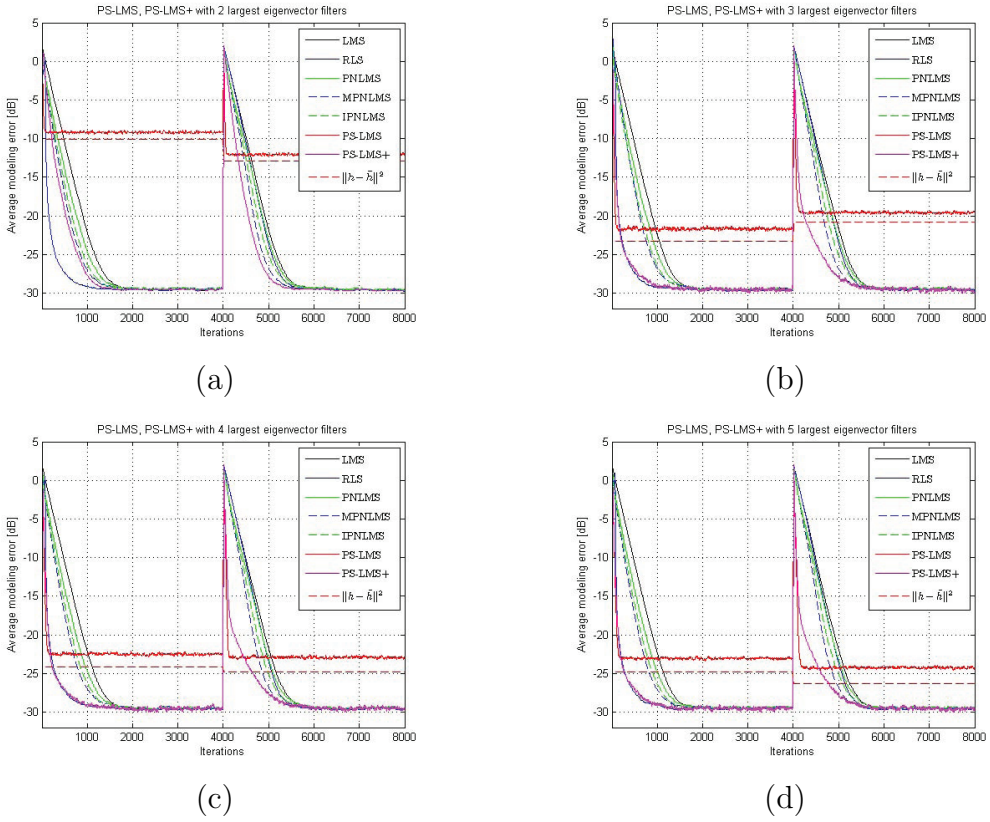


Figure 12: Comparison of identification performances of LMS, RLS, PNLMS, MPNLMS, IPNLMS, PS-LMS and PS-LMS+ when SNR=20 dB with (a) $N = 2$, (b) $N = 3$, (c) $N = 4$, and (d) $N = 5$.

ror in dB. The step size of LMS was set to $\mu = 0.0031$, which yields a modeling error of approximately -30 dB when the algorithm converges. To facilitate performance comparisons, the adaptation step size of PS-LMS+ was set to $L \cdot \mu = 0.1984$ and the optimal distribution of adaptation power g_i were obtained from the optimization algorithm with the constraint that $g_{N+1} = g_{N+2} = \ldots = g_L$. The adaptation step sizes of PS-LMS were adjusted according to N as $\bar{\mu} = \frac{L}{N}\mu$ such that they achieve similar steady-state

performances. For the same purpose, the step-sizes of PNLMS, IPNLMS and MPNLMS were selected as $L \cdot \mu = 0.1984$ and the forgetting factor and the initial value of the inverse covariance matrix of RLS were set to 0.9967 and $100\mathbf{I}$ respectively so that similar steady-state performances are achieved for those algorithms. The proportionality (or activation) ρ and the initialization parameters δ of PNLMS are empirically adjusted to $\rho = 0.3$ and $\delta = 0.01$ for the best convergence performance. The optimum step-size modify factor is 1000 for MPNLMS (Deng and Doroslovacki, 2005) and step-size modify factor is -0.5 for IPNLMS (Benesty and Gay, 2002).

In Fig. 12(a), we demonstrate the convergence behavior of the PS-LMS and PS-LMS+ algorithms with the dimension $N = 2$. Fig. 12(a) shows that the LMS, RLS, PNLMS, IPNLMS and MPNLMS algorithms converge with average steady-state modeling error ≈ -29.5 dB after 2000, 1200, 1900, 1800 and 2000 iterations, respectively; while PS-LMS takes only 50 iterations to converge. However the PS-LMS introduces much larger steady-state modeling error ≈ -9 dB, which is expected when N is too small. For comparison, the modeling errors of PS-LMS ($\| \mathbf{h} - \bar{\mathbf{h}} \|^2$) due to dimension reduction are also indicated in the figures. PS-LMS+ takes 250 iterations to converge to ≈ -10 dB and then further converges to ≈ -29.5 dB steady-state modeling error after 1600 iterations. Similar convergence behavior is observed for LMS, PS-LMS and PS-LMS+ when the unknown time-varying system is switched from $\mathbf{h}_\alpha$ to $\mathbf{h}_\beta$, although PS-LMS and PS-LMS+ demonstrate slightly better convergence behavior, with PS-LMS reaching ≈ -12 dB after 70 iterations and PS-LMS+ approaching -15 dB after 300 iterations and further converging to -29.5 dB after 1400 iterations. It is also observed that RLS achieves a

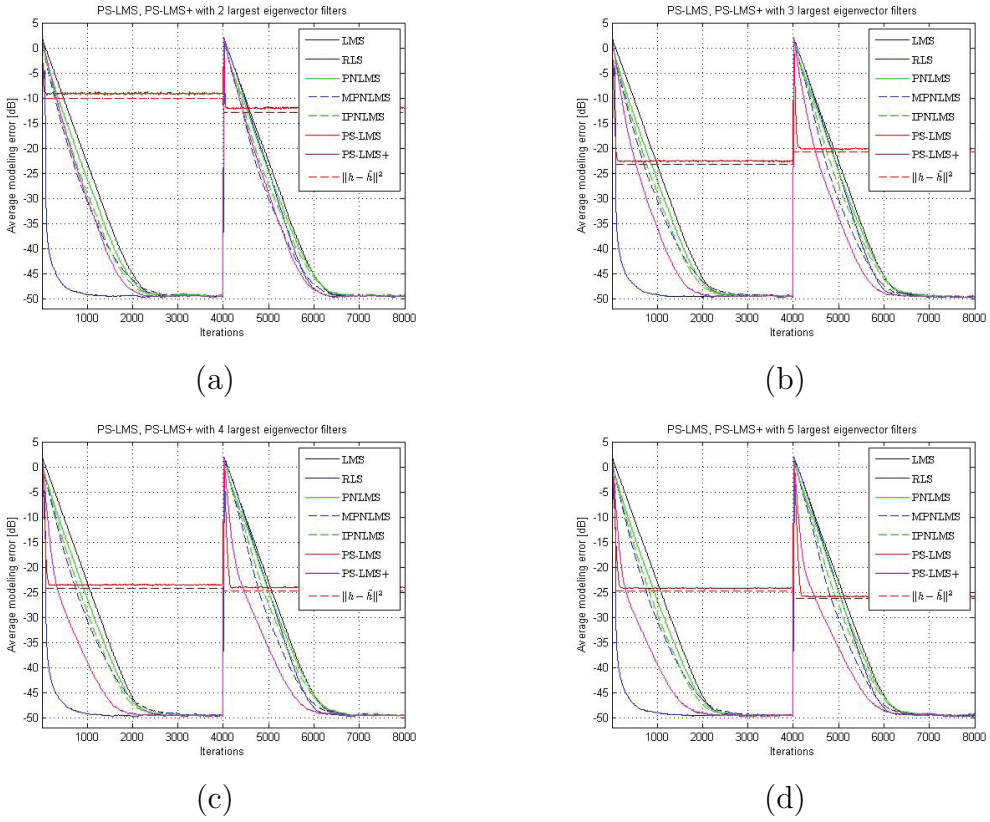


Figure 13: Comparison of identification performance of LMS, RLS, PNLMS, MPNLMS, IPNLMS, PS-LMS and PS-LMS+ when SNR=40 dB with (a) $N = 2$, (b) $N = 3$, (c) $N = 4$, and (d) $N = 5$.

good initial convergence performance at a stationary environment. However, when operating in a non-stationary environment, it doesn't perform as well as LMS in terms of tracking capability(Haykin et al., 1997). On the other hand, the result of MPNLMS shows that it has better convergence performance at $\mathbf{h}_\alpha$ and similar convergence performance at $\mathbf{h}_\beta$ when compared to LMS.

In Fig. 12(b), we assess the convergence behavior of PS-LMS and PS-

LMS+ when the dimension of prominent subspace is increased to $N = 3$. It is observed that PS-LMS now takes about 150 iterations with an average modeling error ≈ -21.5 dB at steady-state, which is significantly lower than the modeling error when $N = 2$. PS-LMS+ takes about 200 iterations to converge with modeling error ≈ -24 dB and further converges to -29.5 dB after 1200 iterations, which is comparable to the more complex RLS. Similar convergence behavior is observed when the variable system is switched from $\mathbf{h}_\alpha$ to $\mathbf{h}_\beta$.

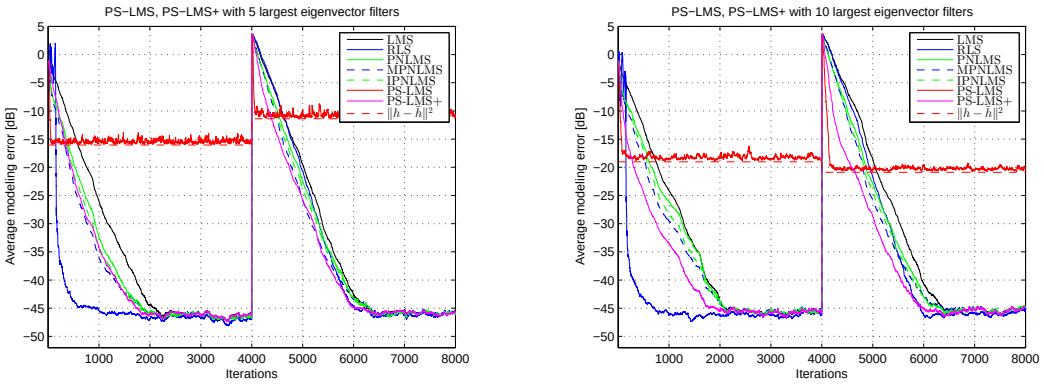
In Fig. 12(c), it can be seen that PS-LMS and PS-LMS+ demonstrate slightly improved initial convergence performance compared to when $N = 3$. However, when the system is switched to $\mathbf{h}_\beta$, an obvious improvement of the convergence behavior is obtained for both PS-LMS and PS-LMS+. From Fig. 12(d), it can be seen that the performance increment from increasing the dimension of the prominent subspace starts to diminish when N is sufficiently large. Further extending the dimension of the prominent subspace doesn't improve the performance. Instead, it reduces the convergence speed and introduces unnecessary additional computational complexity for PS-LMS.

In Fig. 13, we illustrate the convergence behavior of the PS-LMS and PS-LMS+ algorithms with SNR = 40 dB. It can be seen that at this SNR setting, the modeling error at steady state has been reduced to approximately -50 dB. As the result, the gaps between modeling errors of PS-LMS ($\|\mathbf{h} - \bar{\mathbf{h}}\|^2$) and those of other full-rank algorithms are larger compared to previous experiment. However, PS-LMS+ still converges to the same modeling errors of other full-rank algorithms with fewest numbers of iterations. The initial convergence speed of PS-LMS+ is slightly slower compared to previous ex-

periment (SNR = 20 dB) since more adaptation power was allocated to the non prominent subspace by the optimization algorithm in adaption to the lower noise floor.

We performed the same comparison for echo models at SNR = 40 dB. In this experiment, necessary adjustments were made to the parameters of the adaptive algorithms to maintain the same steady-state modeling error. For example, the step size of PNLMS, IPNLMS and MPNLMS was adjusted to $L \cdot \mu = 0.4154$ since the order of the FIR filters is 134. The results are shown in Fig. 14 (a) and (b) for $N = 5$ and $N = 10$ respectively. It can be seen from the results that at $N = 5$, performing adaptation on the prominent subspace doesn't introduce significant advantage to PS-LMS+ since a significant energy still resides in the non prominent system subspace, which is also evident from the gap between the modeling errors of PS-LMS and those of other full-rank algorithms. However, if the dimension of prominent subspace increases to $N = 10$, the performance of PS-LMS+ improves, and it outperforms other algorithms in terms of convergence speed in both initial adaption and when the system switch from $\mathbf{h}_\alpha$ to $\mathbf{h}_\beta$.

It is of interest to verify the robustness of the proposed algorithms for outliers that violate the sparseness assumption, i.e., systems with parameters that are non-sparse in the transform domain. To this end, we repeated the previous experiments using FIR filters from the random models. Clearly, in this case, both $\mathbf{h}_\alpha$ and $\mathbf{h}_\beta$ are non-sparse in either time-domain or any transform domain. Fig. 15 shows the convergence behavior of LMS, RLS, PNLMS, PS-LMS and PS-LMS+ using the same parameters as in ANC experiments at SNR = 40 dB. Fig. 15(a) and Fig. 15(b) show the convergence



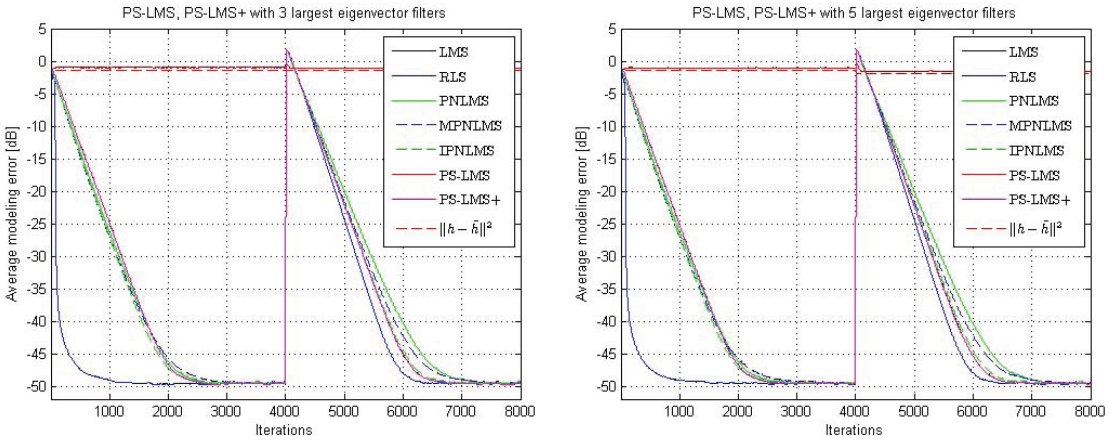
(a) (b)

Figure 14: Comparison of identification performances of LMS, RLS, PNLMS, MPNLMS, IPNLMS, PS-LMS and PS-LMS+ for echo model when SNR=40 dB with (a) $N = 5$ and (b) $N = 10$.

behaviors of the PS-LMS and PS-LMS+ algorithms with $N = 3$ and $N = 5$, respectively. It was observed that in such a case, PS-LMS yields poor convergence performances, whereas PS-LMS+ shows its robustness, converging to -49.5 dB at the same rate as LMS. This is expected since the optimal distribution of g_i for a non-sparse system is uniform and hence in this case PS-LMS+ degenerates to a normal LMS.

5. Conclusions

In this paper, two new adaptive algorithms, namely, PS-LMS and PS-LMS+ are proposed for fast identification of systems that are sparse in the transform domain. PS-LMS constrains the LMS adaptation in the prominent subspace of the unknown system, thus effectively improves the convergence speed of the LMS algorithm at the cost of slightly larger modeling errors in steady-state. On the other hand, PS-LMS+ combines the virtues of both



(a) (b)

Figure 15: Comparison of identification performances of LMS, RLS, PNLMS, MPNLMS, IPNLMS, PS-LMS and PS-LMS+ for random models when SNR = 40 dB with (a) $N = 3$ and (b) $N = 5$.

PS-LMS and LMS by using a larger LMS adaption steps in the prominent subspace, thus improving the convergence speed of the LMS algorithm without compromising its steady-state performance. The performances of PS-LMS and PS-LMS+ for system identification are compared with conventional LMS, RLS, PNLMS, IPNLMS and MPNLMS. It is found that the proposed algorithms provide faster convergence speed with the cost of slightly higher computational complexity comparing to conventional LMS. On the other hand, when comparing to RLS which is highly complex, the proposed algorithms achieves similar convergence speed for stationary system as well as faster tracking capability if the system is non-stationary. Simulation results also show that PS-LMS+ is more robust to outliers where the parameters of unknown system are non-sparse after prominent system subspace decompo-

sition compared to PS-LMS.

The proposed algorithms is suitable for applications where the plant has a transfer function that is either unknown to the control system, or time-varying during operation, and meanwhile the transfer function has a skew distribution (i.e., some impulse responses have larger probabilities to occur than others) that can be identified using off-line measurements. A typical application of the proposed algorithms is the modeling of electro-acoustic path between the loudspeaker and the error microphone for ANC headset. It can be easily extended to other systems that have the characteristics described above.

References

Allen, J.B., Berkley, D.A., 1979. Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Amer. 65, 943–950.

Benesty, J., Gay, S., 2002. An improved PNLMS algorithm, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2009, pp. 1881–1884.

Chen, Y., Gu, Y., Hero, A., 2009. Sparse LMS for system identification, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2009.

Deng, H., Doroslovacki, M., 2005. Improving convergence of the PNLMS algorithm for sparse impulse response identification. IEEE Signal Processing Letters , 181–184.

Duttweiler, D., 2000. Proportionate normalized least-mean-squares adaptation in echo cancelers. IEEE Trans. Speech and Audio Processing 8, 508–518.

Etter, D., 1985. Identification of sparse impulse response systems using an adaptive delay filter, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 1985.

Failli, R., 1989. Digital land mobile communication: final report. Office for official publications of the European Communities.

G.168, I.T., 2007. Annex d: Echo-path models for testing of speech echo cancellers. ITU-T Recommendation G.168 .

Gabriel, W., 1986. Using spectral estimation techniques in adaptive processing antenna systems. IEEE Trans. Antennas and Propagation 34, 291 – 300.

Gay, S., 1998. An efficient, fast converging adaptive filter for network echo cancellation, in: the Thirty-Second Asilomar Conference on Signals, Systems Computers, 1998.

Godavarti, M., Hero, A., 2005. Partial update LMS algorithms. IEEE Trans. Signal Processing 53, 2382–2399.

Goldstein, J.S., Reed, I.S., 1997. Reduced-rank adaptive filtering. IEEE Trans. Signal Processing 45, 492 –496.

Goldstein, J.S., Reed, I.S., Scharf, L.L., 1998. A multistage representation of

the wiener filter based on orthogonal projections. IEEE Trans. Information Theory 44, 2943 –2959.

Gu, Y., Jin, J., Mei, S., 2009. ℓ_0 norm constraint lms algorithm for sparse system identification. IEEE Signal Process. Lett. 16, 774–777.

Haimovich, A.M., Bar-Ness, Y., 1991. An eigenanalysis interference canceler. IEEE Trans. Signal Processing 39, 76 –84.

Haykin, S., 2001. Adaptive Filter Theory. Prentice Hall. 4 edition.

Haykin, S., Sayed, A., Zeidler, J., Yee, P., Wei, P., 1997. Adaptive tracking of linear time-variant systems by extended RLS algorithms . IEEE Trans. signal processing 45, 1118 – 1128.

Homer, J., Mareels, I., Bitmead, R.R., Wahlberg, B., Gustafsson, A., 1998. LMS estimation via structural detection. IEEE Trans. Signal Processing 46, 2651–2663.

Honig, M.L., Goldstein, J.S., 2002. Adaptive reduced-rank interference suppression based on the multistage Wiener filter. IEEE Trans. Communications 50, 986 –994.

Hoyer, P.O., 2004. non-negative matrix factorization with sparseness constraints. journal of machine learning research , 1457–1469.

Kawamura, S., Hatori, M., 1986. A tap selection algorithm for adaptive filters, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 1986.

Kuo, S.M., Mitra, S., Gan, W.S., 2006. Active noise control system for headphone applications. IEEE Transactions on Control Systems Technology 14, 331–335.

Kuo, S.M., Morgan, D.R., 1996. Active Noise Control Systems: Algorithms and DSP Implementations. Wiley, New York.

Li, Y., Gu, Y., Tang, K., 2006. Parallel NLMS Filters with Stochastic Active Taps and Step-Sizes for Sparse System Identification, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2006.

Ljung, L., 1999. System identification: theory for the user. Prentice Hall. 2 edition.

Mayyas, K., 2005. Performance analysis of the deficient length LMS adaptive algorithm. IEEE Tran. Signal Process. , 2727–2734.

Nagumo, J., Noda, A., 1967. A learning method for system identification. IEEE Trans. Automatic Control 12, 282–287.

Nemirovski, A.S., Todd, M.J., 2008. Interior-point methods for optimization. Acta Numerica , 191–234.

Scharf, L., Tufts, D., 1987. Rank reduction for modeling stationary signals. IEEE Trans. Acoustics, Speech and Signal Processing 35, 350 – 355.

Sondhi, M., Berkley, D., 1980. Silencing echoes on the telephone network. Proceedings of the IEEE 68, 948–963.

Widrow, B., 1970. Adaptive filters, in: Kalman, R.E., DeClaris, N. (Eds.), Aspects of Network and System Theory. Holt, Rinehart and Winston, New York, pp. 563–587.

Widrow, B., McCool, J.M., Larimore, M.G., Johnson, C.R., 1976. Stationary and nonstationary learning characteristics of the LMS adaptive filter. Proc IEEE 64, 1151–1162.

Yu, R., Song, Y., Rahardja, S., 2012. LMS in prominent system subspace for fast system identification, in: Proc. IEEE International workshop on Statistical Signal Processing (SSP).