Partial Update LMS Algorithms

Mahesh Godavarti, *Member, IEEE,* and Alfred O. Hero, III, *Fellow, IEEE*

***Abstract*—Partial updating of LMS filter coefficients is an effective method for reducing computational load and power consumption in adaptive filter implementations. This paper presents an analysis of convergence of the class of Sequential Partial Update LMS algorithms (S-LMS) under various assumptions and shows that divergence can be prevented by scheduling coefficient updates at random, which we call the Stochastic Partial Update LMS algorithm (SPU-LMS). Specifically, under the standard independence assumptions, for wide sense stationary signals, the S-LMS algorithm converges in the mean if the step-size parameter μ is in the convergent range of ordinary LMS. Relaxing the independence assumption, it is shown that S-LMS and LMS algorithms have the same sufficient conditions for exponential stability. However, there exist nonstationary signals for which the existing algorithms, S-LMS included, are unstable and do not converge for any value of μ. On the other hand, under broad conditions, the SPU-LMS algorithm remains stable for nonstationary signals. Expressions for convergence rate and steady-state mean-square error of SPU-LMS are derived. The theoretical results of this paper are validated and compared by simulation through numerical examples.**

***Index Terms*—Exponential stability, max partial update, partial update LMS algorithms, periodic algorithm, random updates, sequential algorithm, set-membership.**

I. Introduction

The least mean-squares (LMS) algorithm is a popular algorithm for adaptation of weights in adaptive beamformers using antenna arrays and for channel equalization to combat intersymbol interference. Many other application areas of LMS include interference cancellation, echo cancellation, space time modulation and coding, signal copy in surveillance, and wireless communications. Although there exist algorithms with faster convergence rates like RLS, LMS is popular because of its ease of implementation and low computational costs [18], [20], [25].

Partial updating of the LMS adaptive filter has been proposed to reduce computational costs and power consumption [13], [14], [22], which is quite attractive in the area of mobile computing and communications. Many mobile communication devices have applications like channel equalization and echo cancellation that require the adaptive filter to have a very large number of coefficients. Updating the entire coefficient vector is costly in terms of power, memory, and computation and is sometimes impractical for mobile units.

Manuscript received September 17, 2003; revised July 8, 2004. This work was completed while M. Godavarti was a Ph.D. candidate at the University of Michigan, Ann Arbor, under the supervision of Prof. A. O. Hero. This work was presented in part, at ICASSP 1999, Phoenix, AZ; SAM 2000, Boston, MA; and ICASSP 2001, Salt Lake City, UT. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tulay Adali.

M. Godavarti is with the Ditech Communications, Inc., Mountain View, CA 94043 USA.

A. O. Hero, III, is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA.

Digital Object Identifier 10.1109/TSP.2005.849167

Two types of partial update LMS algorithms are prevalent in the literature and have been described in [11]. They are referred to as the "Periodic LMS algorithm" and the "Sequential LMS algorithm." To reduce computation needed during the update part of the adaptive filter by a factor of P, the Periodic LMS algorithm (P-LMS) updates all the filter coefficients every Pth iteration instead of every iteration. The Sequential LMS (S-LMS) algorithm updates only a fraction of coefficients every iteration. Another variant referred to as "Max Partial Update LMS algorithm" (Max PU-LMS) has been proposed in [1], [9], and [10]. Yet another variant known as the "set-membership partial-update NLMS algorithm" (SMPU-NLMS) based on data-selective updating appears in [8]. The algorithm combines the ideas of set-membership normalized algorithms with the ideas of partial update algorithms. These variants have data dependent updating schedules and therefore can have faster convergence, for stationary signals, than P-LMS and S-LMS algorithms that have data independent updating schedules. However, for nonstationary signals, it is possible that data dependent updating can lead to nonconvergence. This drawback is illustrated by comparing Max PU-LMS and SMPU-NLMS to the regular LMS and proposed SPU-LMS algorithms through a numerical example. SPU-LMS is similar to P-LMS and S-LMS algorithms in the sense that it also uses data independent updating schedules. Thus, while analytical comparison to Max PU-LMS and SMPU-NLMS algorithms would be interesting, comparisons are limited to S-LMS and P-LMS.

In [11], for stationary signals, convergence conditions were derived for the convergence of S-LMS under the assumption of small step-size parameter (μ), which turned out to be the same as those for the standard LMS algorithm. Here, bounds on μ are obtained that hold for stationary signals and arbitrary fixed sequence of partial updates. First, under the standard independence assumptions, it is shown that for stationary signals first order stability of LMS implies first order stability of S-LMS. However, the important characteristic of S-LMS, which is shared by P-LMS as well, is that the coefficients to be updated at an iteration are pre-determined. It is this characteristic which renders P-LMS and S-LMS unstable for certain signals and which makes an alternative random coefficient updating approach attractive.

In this paper, we propose a new partial update algorithm in which the subset of the filter coefficients that are updated each iteration is selected at random. The algorithm, referred to as the Stochastic Partial Update LMS algorithm (SPU-LMS), involves selection of a subset of size N/P coefficients out of P possible subsets from a fixed partition of the N coefficients in the weight vector. For example, filter coefficients can be partitioned into even and odd subsets and either even or odd coefficients are randomly selected to be updated in each iteration. Conditions on the step-size parameter are derived that ensure convergence in the mean and the mean square sense for stationary signals, for deterministic signals, and for the general case of mixing signals.

Partial update algorithms can be contrasted against another variant of LMS known as the Fast Exact LMS (FE-LMS) [4]. Here also, the updates are done every Pth instead of every iteration (P has to be much smaller than N, the filter length, to realize any computational savings [4]). However, the updates after every Pth iteration result in exactly the same filter as obtained from LMS with P updates done every iteration. Therefore, the algorithm suffers no degradation with respect to convergence when compared to the regular LMS. A generalized version of Fast Exact LMS appears in [5] where the Newton transversal filter is used instead of LMS.

When convergence properties are considered the FE-LMS algorithm is more attractive than the PU-LMS algorithm. However, PU-LMS algorithms become more attractive when the available program and data memory is limited. The computational savings in FE-LMS come at the cost of increased program memory, whereas PU-LMS algorithms require negligible increase in program size and in some implementations might reduce the data memory required. Moreover, in FE-LMS the reduction in number of execution cycles is offset by the additional cycles needed for storing the data in intermediate steps. Finally, the computational savings for the FE-LMS algorithm are realized for a time-series signal. If the signal happens to be the output of an array, that is the output of an individual antenna is the input to a filter tap, then the method employed in [4] to reduce computations no longer holds.

The main contributions of this paper can be summarized as follows.

- For stationary signals and arbitrary sequence of updates, it is shown, without the independence assumption, that S-LMS has the same stability and mean-square convergence properties as LMS.
- Signal scenarios are demonstrated for which the prevalent partial update algorithms do not converge.
- A new algorithm is proposed, called the Stochastic Partial Update LMS Algorithm (SPU-LMS), that is based on randomizing the updating schedule of filter coefficients that ensures convergence.
- Stability conditions for SPU-LMS are derived for stationary signal scenarios, and it is demonstrated that the steady-state performance of the new algorithm is as good as that of the regular LMS algorithm.
- A persistence of excitation condition for the convergence of SPU-LMS is derived for the case of deterministic signals, and it is shown that this condition is the same as for the regular LMS algorithm.
- For the general case of mixing signals, it is shown that the stability conditions for SPU-LMS are the same as that of LMS. The method of successive approximation is extended to SPU-LMS and the results used to show that SPU-LMS does not suffer a degradation in steady-state performance.
- It is demonstrated through different examples that for nonstationary signal scenarios, as might arise in echo cancellation in telephone networks or digital communication systems, partial updating using P-LMS and S-LMS might be undesirable as these are not guaranteed to converge. SPU-LMS is a better choice because of its guaranteed convergence properties.

The organization of the paper is as follows. First, in Section II, a brief description of the sequential partial update algorithm is given. The algorithm is analyzed for the case of stationary signals under independence assumptions in Section II-A. The rest of the paper deals with the new algorithm. A brief description of the algorithm is given in Section III, and its analysis is given in Sections III-A (uncorrelated input and coefficient vectors), B (deteriministic signals), and C (correlated input and coefficient vectors). It is shown that the performance of SPU-LMS is very close to that of LMS in terms of stability conditions and final mean squared error. Section IV discusses the performance of the new algorithm through analytical comparisons with the existing partial udpate algorithms and through numerical examples (Section IV-A). In particular, Section IV demonstrates, without the independence assumption, the exponential stability and the mean-square convergence analysis of S-LMS for stationary signals and of P-LMS for the general case of mixing signals. Finally, conclusions and directions for future work are indicated in Section V.

II. Sequential PU-LMS Algorithm

Let $\{x_{i,k}\}$ be the input sequence, and let $\{w_{i,k}\}$ denote the coefficients of an adaptive filter of odd length, N. Define

$$W_k = [w_{1,k}\ w_{2,k}\ \dots\ w_{N,k}]^T$$
$$X_k = [x_{1,k}\ x_{2,k}\ x_{3,k}\ \dots\ x_{N,k}]^T$$

where the terms defined above are for the instant k and T denotes the transpose operator. In addition, Let d_k denote the desired response. In typical applications, d_k is a known training signal which is transmitted over a noisy channel with unknown FIR transfer function.

In the stationary signal setting, the offline problem is to choose an optimal W such that

$$\begin{aligned}\xi^{(W)} &= E\left[(d_k - y_k)(d_k - y_k)^*\right]\\ &= E\left[(d_k - W^H X_k)(d_k - W^H X_k)^*\right]\end{aligned}$$

is minimized, where a^* denotes the complex conjugate of a, and $W^H = (W^T)^*$ denotes the complex conjugate transpose of W. The solution to this problem is given by

$$W_{\text{opt}} = R^{-1} r \tag{1}$$

where $R = E[X_k X_k^H]$, and $r = E[d_k^* X_k]$. The minimum attainable mean square error $\xi^{(W)}$ is given by

$$\xi_{\min} = E\left[d_k d_k^*\right] - r^H R^{-1} r.$$

For the following analysis, we assume that the desired signal d_k satisfies the following relation:[1][11]

$$d_k = W_{\rm opt}^H X_k + n_k \tag{2}$$

where X_k is a zero mean complex circular Gaussian[2] random vector, and n_k is a zero mean circular complex Gaussian (not necessarily white) noise, with variance $\xi_{\min}$, uncorrelated with X_k.

Assume that the filter length N is a multiple of P. For convenience, define the index set $S = \{1, 2, \ldots, N\}$. Partition S into P mutually exclusive subsets of equal size $S_1, S_2, \ldots, S_P$. Define $\mathcal{I}_i$ by zeroing out the jth row of the identity matrix I if $j \notin S_i$. In that case, $\mathcal{I}_i X_k$ will have precisely N/P nonzero entries. Let the sentence "choosing S_i at iteration k" stand to mean "choosing the weights with their indices in S_i for update at iteration k."

The S-LMS algorithm is described as follows. At a given iteration k, one of the sets $S_i, i = 1, \ldots, P$ is chosen in a predetermined fashion, and the update is performed. Without loss of generality, it can be assumed that at iteration k, the set $S_{k\%P+1}$ is chosen for update, where $k\%P$ denotes the operation "k modulo P."

$$w_{k+1,j} = \begin{cases} w_{k,j} + \mu e_k^* x_{k,j}, & \text{if } j \in S_{k\%P+1} \\ w_{k,j}, & \text{otherwise} \end{cases}$$

where $e_k = d_k - W_k^H X_k$. The above update equation can be written in a more compact form

$$W_{k+1} = W_k + \mu e_k^* \mathcal{I}_{k\%P+1} X_k. \tag{3}$$

In the special case of odd and even updates $P = 2$, S_1 consists of all odd indices and S_2 of all even indices.

Define the coefficient error vector as

$$V_k = W_k - W_{\rm opt}$$

which leads to the following coefficient error vector update for S-LMS when k is even

$$V_{k+1} = \left(I - \mu \mathcal{I}_1 X_k X_k^H\right) V_k + \mu n_k \mathcal{I}_1 X_k$$

and the following when k is odd:

$$V_{k+1} = \left(I - \mu \mathcal{I}_2 X_k X_k^H\right) V_k + \mu n_k \mathcal{I}_2 X_k.$$

[1]Note that the model assumed for d_k is same as assuming d_k and X_k are jointly Gaussian sequences. Under this assumption, d_k can be written as $d_k = W_{\rm opt}^H X_k + m_k$, where $W_{\rm opt}$ is as in (1) and $m_k = d_k - W_{\rm opt}^H X_k$. Since $E[m_k X_k] = E[X_k d_k] - E[X_k X_k^H] W_{\rm opt} = 0$ and m_k and X_k are jointly Gaussian, we conclude that m_k and X_k are independent of each other which is same as model (2).

[2]A complex circular Gaussian random vector consists of Gaussian random variables whose marginal densities depend only on their magnitudes. For more information, see [21] or [24. p. 198].

A. Analysis: Stationary Signals, Independent Input, and Coefficient Vectors

Assuming that d_k and X_k are jointly WSS random sequences, we analyze the convergence of the mean coefficient error vector $E[V_k]$. We make the standard assumptions that V_k and X_k are independent of each other [3]. For the regular full update LMS algorithm, the recursion for $E[V_k]$ is given by

$$E[V_{k+1}] = (I - \mu R) E[V_k] \tag{4}$$

where I is the N-dimensional identity matrix, and $R = E[X_k X_k^H]$ is the input signal correlation matrix. The well-known necessary and sufficient condition for $E[V_k]$ to converge in (4) is given by [18]

$$\rho(I - \mu R) < 1$$

where $\rho(B)$ denotes the spectral radius of B ($\rho(B) = \max|\lambda_i(B)|$). This leads to

$$0 < \mu < \frac{2}{\lambda_{\max}(R)} \tag{5}$$

where $\lambda_{\max}(R)$ is the maximum eigen-value of the input signal correlation matrix R. Note that this need not translate to be the necessary and sufficient condition for the convergence of $E[V_k]$ in actuality as (4) has been obtained under the independence assumption which is not true in general.

Taking expectations under the same assumptions as above and using the independence assumption on the sequences X_k, n_k, which is the independence assumption on X_k and V_k, we obtain, when k is even

$$\begin{aligned} E[V_{k+1}] &= (I - \mu \mathcal{I}_1 R) E[V_k] \\ E[V_{k+2}] &= (I - \mu \mathcal{I}_2 R) E[V_{k+1}] \end{aligned}$$

and when k is odd

$$\begin{aligned} E[V_{k+1}] &= (I - \mu \mathcal{I}_2 R) E[V_k] \\ E[V_{k+2}] &= (I - \mu \mathcal{I}_1 R) E[V_{k+1}]. \end{aligned}$$

Simplifying the above two sets of equations, we obtain, for even-odd S-LMS when k is even

$$E[V_{k+2}] = (I - \mu \mathcal{I}_2 R)(I - \mu \mathcal{I}_1 R) E[V_k] \tag{6}$$

and when k is odd

$$E[V_{k+2}] = (I - \mu \mathcal{I}_1 R)(I - \mu \mathcal{I}_2 R) E[V_k]. \tag{7}$$

It can be shown that under the above assumptions on X_k, V_k and d_k, the convergence conditions for even ($\rho((I - \mu \mathcal{I}_2 R)(I - \mu \mathcal{I}_1 R)) < 1$) and odd update equations ($\rho((I - \mu \mathcal{I}_1 R)(I - \mu \mathcal{I}_2 R)) < 1$) are identical. We therefore focus on (6). It will be shown that if $\rho(I - \mu R) < 1$, then $\rho((I - \mu \mathcal{I}_2 R)(I - \mu \mathcal{I}_1 R)) < 1$.

Now, if instead of just two partitions of odd and even coefficients ($P = 2$), there are any number of arbitrary partitions

$(P \geq 2)$, and then, the update equations can be similarly written as above, with $P > 2$. Namely

$$E[V_{k+P}] = \prod_{i=1}^{P} \left(I - \mu \mathcal{I}_{(i+k)\%P+1} R\right) E[V_k]. \tag{8}$$

$\mathcal{I}_i, i = 1, \ldots, P$ is obtained from I, which is the identity matrix of dimension $N \times N$, by zeroing out some rows in I such that $\sum_{i=1}^{P} \mathcal{I}_i = I$.

We will show that for any arbitrary partition of any size $(P \geq 2)$, S-LMS converges in the mean if LMS converges in the mean. The case $P = 2$ follows as a special case. The intuitive reason behind this fact is that both the mean update equation for LMS $E[V_{k+1}] = (I - \mu R)E[V_k]$ and the mean update equation for S-LMS $E[V_{k+1}] = (I - \mu \mathcal{I}_{k\%P+1} R)E[V_k]$, $i = 1, \ldots, P$ try to minimize the mean squared error $E[V_k^H]RE[V_k]$. This error term is a quadratic bowl in the $E[V_k]$ coordinate system. Note that LMS moves in the direction of the negative gradient $-RE[V_k]$ by retaining all the components of this gradient in the $E[V_k]$ coordinate system, whereas S-LMS discards some of the components at every iteration. The resulting direction, in which S-LMS updates its weights, obtained from the remaining components can be broken into two components: one in the direction of $-RE[V_k]$ and one perpendicular to it. Hence, if LMS reduces the mean squared error, then so does S-LMS.

The result is stated formally in Theorem 2, and the following theorem is used in proving the result.

Theorem 1—[19, Prob. 16, p. 410]: Let B be an arbitrary $N \times N$ matrix. Then, $\rho(B) < 1$ if and only if there exists some positive definite $N \times N$ matrix A such that $A - B^H AB$ is positive definite. Here, $\rho(B)$ denotes the spectral radius of B $(\rho(B) = \max_{1,\ldots,N} |\lambda_i(B)|)$.

Theorem 2: Let R be a positive definite matrix of dimension $N \times N$ with $\rho(R) = \lambda_{\max}(R) < 2$; then, $\rho(\prod_{i=1}^{P}(I - \mathcal{I}_i R)) < 1$, where $\mathcal{I}_i, i = 1, \ldots, P$ are obtained by zeroing out some rows in the identity matrix I such that $\sum_{i=1}^{P} \mathcal{I}_i = I$. Thus, if X_k and d_k are jointly wide sense stationary, then S-LMS converges in the mean if LMS converges in the mean.

Proof: Let $\mathbf{x}_0 \in \mathbb{C}^N$ be an arbitrary nonzero vector of length N. Let $\mathbf{x}_i = (I - \mathcal{I}_i R)\mathbf{x}_{i-1}$. In addition, let $\mathbf{P} = \prod_{i=1}^{P}(I - \mathcal{I}_i R)$.

First, we will show that $\mathbf{x}_i^H R \mathbf{x}_i \leq \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \alpha \mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1}$, where $\alpha = (1/2)(2 - \lambda_{\max}(R)) > 0$.

$$\begin{aligned}\mathbf{x}_i^H R \mathbf{x}_i &= \mathbf{x}_{i-1}^H (I - R\mathcal{I}_i) R (I - \mathcal{I}_i R) \mathbf{x}_{i-1} \\ &= \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \alpha \mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1} - \beta \mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1} \\ &\quad + \mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathcal{I}_i R \mathbf{x}_{i-1}\end{aligned}$$

where $\beta = 2 - \alpha$. If we can show $\beta R \mathcal{I}_i R - R \mathcal{I}_i R \mathcal{I}_i R$ is positive semi-definite, then we are done. Now

$$\beta R \mathcal{I}_i R - R \mathcal{I}_i R \mathcal{I}_i R = \beta R \mathcal{I}_i \left(I - \frac{1}{\beta} R\right) \mathcal{I}_i R.$$

Since $\beta = (1 + \lambda_{\max}(R)/2) > \lambda_{\max}(R)$, it is easy to see that $I - (1/\beta)R$ is positive definite. Therefore, $\beta R \mathcal{I}_i R - R \mathcal{I}_i R \mathcal{I}_i R$ is positive semi-definite, and

$$\mathbf{x}_i^H R \mathbf{x}_i \leq \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \alpha \mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1}.$$

Combining the above inequality for $i = 1, \ldots, P$, we note that $\mathbf{x}_P^H R \mathbf{x}_P < \mathbf{x}_0^H R \mathbf{x}_0$ if $\mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1} > 0$ for at least one i, $i = 1, \ldots, P$. We will show by contradiction that is indeed the case.

If not, then $\mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1} = 0$ for all i, $i = 1, \ldots, P$. Since $\mathbf{x}_0^H R \mathcal{I}_1 R \mathbf{x}_0 = 0$, this implies $\mathcal{I}_1 R \mathbf{x}_0 = \mathbf{0}$. Therefore, $\mathbf{x}_1 = (I - \mathcal{I}_1 R)\mathbf{x}_0 = \mathbf{x}_0$. Similarly, $\mathbf{x}_i = \mathbf{x}_0$ for all i, $i = 1, \ldots, P$. This, in turn, implies that $\mathbf{x}_0^H R \mathcal{I}_i R \mathbf{x}_0 = 0$ for all i, $i = 1, \ldots, P$, which is a contradiction since $R(\sum_{i=1}^{P} \mathcal{I}_i)R$ is a positive-definite matrix, and $0 = \sum_{i=1}^{P} \mathbf{x}_0^H R \mathcal{I}_i R \mathbf{x}_0 = \mathbf{x}_0^H R(\sum_{i=1}^{P} \mathcal{I}_i) R \mathbf{x}_0 \neq 0$.

Finally, we conclude that

$$\mathbf{x}_0^H \mathbf{P}^H R \mathbf{P} \mathbf{x}_0 = \mathbf{x}_P^H R \mathbf{x}_P < \mathbf{x}_0^H R \mathbf{x}_0.$$

Since $\mathbf{x}_0$ is arbitrary, we have $R - \mathbf{P}^H R \mathbf{P}$ to be positive definite so that applying Theorem 1, we conclude that $\rho(\mathbf{P}) < 1$.

Finally, if LMS converges in the mean, we have $\rho(I - \mu R) < 1$ or $\lambda_{\max}(\mu R) < 2$, which, from the above proof, is sufficient to conclude that $\rho(\prod_{i=1}^{P}(I - \mu \mathcal{I}_i R)) < 1$. Therefore, S-LMS also converges in the mean. ■

Remark 1: Note that it is sufficient for $\mathcal{I}_i$ to be such that $\sum_{i=1}^{P} \mathcal{I}_i$ is positive definite. That means that the subsets updated at each iteration need not be mutually exclusive.

Remark 2: It is interesting to note that in the proof above if

1) we choose $\alpha = (1/2)(2 - \lambda_{\max}(I_i R I_i)) > 0$ and $\beta = 2 - \alpha$ for each i;
2) we write $\beta R \mathcal{I}_i R - R \mathcal{I}_i R \mathcal{I}_i R$ as $\beta R \mathcal{I}_i (\mathcal{I}_i - (1/\beta) \mathcal{I}_i R \mathcal{I}_i) \mathcal{I}_i R$ instead of as $\beta R \mathcal{I}_i (I - (1/\beta)R) \mathcal{I}_i R$

then it can be shown that for stationary signals the sequential algorithm enjoys a more lenient condition on μ for convergence in the mean: $0 < \mu < (2/\max_i\{\lambda_{\max}(\mathcal{I}_i R \mathcal{I}_i)\})$. This condition is more lenient than that of regular LMS: $0 < \mu < (2/\lambda_{\max}(R))$.

With a little extra effort, a tighter bound on the spectral radius of $\prod_{i=1}^{P}(I - \mu \mathcal{I}_i R)$ can be demonstrated.

Theorem 3: Fix $\mu^* < 2/\lambda_{\max}(R)$, and let $\mathcal{I}_i$ be such that $\sum_{i=1}^{P} \mathcal{I}_i = I$. Then, there exists a constant $0 < \alpha_{\mu^*}$ dependent only on μ^* such that $\rho(\prod_{i=1}^{P}(I - \mu \mathcal{I}_i R))$ is contained within a circle of radius $(1 - \mu \alpha_{\mu^*})$ for all $0 < \mu < \mu^*$.

Proof: Let $\mathbf{x}_0 \in \mathbb{C}^N$ be an arbitrary nonzero vector of length N as before. Let $\mathbf{x}_i = (I - \mu \mathcal{I}_i R)\mathbf{x}_{i-1}$ and $\mathbf{P}(\mu) = \prod_{i=1}^{P}(I - \mu \mathcal{I}_i R)$.

From the proof of Theorem 2, we have, for $i = 1, \ldots, P$

$$\begin{aligned}\mathbf{x}_i^H R \mathbf{x}_i &\leq \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \mu \left(1 - \mu \frac{\lambda_{\max}(R)}{2}\right) \\ &\quad \times \mathbf{x}_{i-1}^H R \mathcal{I}_i R \mathbf{x}_{i-1} \\ &= \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \mu \left(1 - \mu \frac{\lambda_{\max}(R)}{2}\right) \mathbf{x}_{i-1}^H R^{\frac{1}{2}} R^{\frac{1}{2}} \\ &\quad \cdot \mathcal{I}_i R^{\frac{1}{2}} R^{\frac{1}{2}} \mathbf{x}_{i-1} \\ &\leq \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \alpha_i \lambda_{\min}^* \mu \left(1 - \mu \frac{\lambda_{\max}(R)}{2}\right) \\ &\quad \times \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} \\ &\leq \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1} - \alpha_i \lambda_{\min}(R) \mu \left(1 - \mu \frac{\lambda_{\max}(R)}{2}\right) \\ &\quad \cdot \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1}\end{aligned}$$

with $\lambda_{\min}^* = \min\{\lambda(R^{1/2}\mathcal{I}_i R^{1/2}) > 0\} \geq \lambda_{\min}(R)$, and $\alpha_i = (\mathbf{y}_i')^H \mathbf{y}_i' / \mathbf{x}_{i-1}^H R \mathbf{x}_{i-1}$. $\mathbf{y}_i'$ is defined as $\mathbf{y}_i' = \mathcal{P}_i(R^{1/2}\mathbf{x}_{i-1})$, where $\mathcal{P}_i(\mathbf{x})$ denotes the projection of $\mathbf{x}$ onto the nonzero eigenspace of $R^{1/2}\mathcal{I}_i R^{1/2}$.

Next, consider $\hat{\mathbf{x}}_i = R^{1/2}\mathbf{x}_i$, $i = 0, 1, \ldots, P$. Then, the update equation for $\hat{\mathbf{x}}_i$ is $\hat{\mathbf{x}}_i = (I - R^{1/2}\mathcal{I}_i R^{1/2})\hat{\mathbf{x}}_{i-1}$. Let $\mathbf{y}_i'$ be as before and $\mathbf{y}_i = \mathcal{P}_i(\hat{\mathbf{x}}_0)$.

Let

$$\mathbf{z}_i = \hat{\mathbf{x}}_i - \hat{\mathbf{x}}_{i-1} = -\mu R^{\frac{1}{2}} \mathcal{I}_i R^{\frac{1}{2}} \hat{\mathbf{x}}_{i-1}.$$

Then

$$\mathbf{z}_i^H \mathbf{z}_i \leq \mu^2 \lambda_{\max}^2(R) \left(\mathbf{y}_i'\right)^H \mathbf{y}_i' = 4 \left(\mathbf{y}_i'\right)^H \mathbf{y}_i'.$$

In addition

$$\mathbf{y}_i = \mathcal{P}_i(\hat{\mathbf{x}}_0) = \mathcal{P}_i\left(\sum_{j=1}^{i-1} \mathbf{z}_j\right) + \mathcal{P}_i(\hat{\mathbf{x}}_{i-1}) = \sum_{j=1}^{i-1} \mathcal{P}_i(\mathbf{z}_j) + \mathbf{y}_i'$$

for $i = 1, \ldots, P$. Next, denoting $|\mathbf{z}|$ for $\sqrt{\mathbf{z}^H \mathbf{z}}$ and making use of the facts that $\hat{\mathbf{x}}_i^H \hat{\mathbf{x}}_i = \mathbf{x}_i^H R \mathbf{x}_i \leq \mathbf{x}_0^H R \mathbf{x}_0 = \hat{\mathbf{x}}_0^H \hat{\mathbf{x}}_0$ and $|\mathcal{P}(\mathbf{z}_j)| \leq |\mathbf{z}_j|$, we obtain for $i = 1, \ldots, P$

$$\begin{aligned} |\mathbf{y}_i| &\leq \sum_{j=1}^{i-1} |\mathbf{z}_j| + |\mathbf{y}_i'| \leq 2 \sum_{j=1}^{i-1} \left|\mathbf{y}_j'\right| + |\mathbf{y}_i'| \\ &= 2 \sum_{j=1}^{i-1} \sqrt{\alpha_j} |\hat{\mathbf{x}}_{i-1}| + \sqrt{\alpha_i} |\hat{\mathbf{x}}_i| \\ &\leq \left(2 \sum_{j=1}^{i} \sqrt{\alpha_j}\right) |\hat{\mathbf{x}}_0|. \end{aligned}$$

Therefore, at least one of α_j is greater than or equal to $(1/4P^3)/(\lambda_{\min}(R)/\lambda_{\max}(R))$. Otherwise, $\lambda_{\min}(R)\hat{\mathbf{x}}_0^H \hat{\mathbf{x}}_0 \leq \hat{\mathbf{x}}_0^H R \hat{\mathbf{x}}_0 \leq \lambda_{\max}(R) \sum_{i=1}^{P} \mathbf{y}_i^H \mathbf{y}_i < \lambda_{\min}(R)\hat{\mathbf{x}}_0^H \hat{\mathbf{x}}_0$, which is a contradiction.

Next, choosing $\alpha_{\mu^*} = (1/8P^3)(1 - \mu^*(\lambda_{\max}(R)/(2))(\lambda_{\min}^2(R)/\lambda_{\max}(R))$ and noting that $(1/4P^3)(1 - \mu(\lambda_{\max}(R)/2))(\lambda_{\min}^2(R)/\lambda_{\max}(R)) > 2\alpha_{\mu^*}$ for all $0 < \mu < \mu^*$, we obtain

$$\begin{aligned} \mathbf{x}_P^H R \mathbf{x}_P &\leq (1 - \mu 2\alpha_{\mu^*}) \mathbf{x}_j^H R \mathbf{x}_j \\ &\leq (1 - \mu\alpha_{\mu^*})^2 \mathbf{x}_0^H R \mathbf{x}_0. \end{aligned}$$

This leads to

$$(1 - \mu\alpha_{\mu^*})^{-2} \mathbf{x}_P^H R \mathbf{x}_P \leq \mathbf{x}_0^H R \mathbf{x}_0.$$

Finally, using Theorem 1, we conclude that $\rho((1 - \mu\alpha_{\mu^*})^{-1}\mathbf{P}(\mu)) \leq 1$ or $\rho(\mathbf{P}(\mu)) \leq (1 - \mu\alpha_{\mu^*})$. ■

Remark 3: If we assume that R is block diagonal such that $R = \sum_{i=1}^{P} \mathcal{I}_i R \mathcal{I}_i$ with $\sum_{i=1}^{P} \mathcal{I}_i = I$, then an even tighter bound on $\rho(\mathbf{P}(\mu))$ can be obtained. In this case, $\mathcal{I}_i R = \mathcal{I}_i R \mathcal{I}_i$ and $\mathbf{P}(\mu)$ turns out to be simply

$$\begin{aligned} \prod_{i=1}^{P} (I - \mu \mathcal{I}_i R) &= \prod_{i=1}^{P} (I - \mu \mathcal{I}_i R \mathcal{I}_i) \\ &= I - \mu \sum_{i=1}^{P} \mathcal{I}_i R \mathcal{I}_i = I - \mu R. \end{aligned}$$

III. Stochastic PU-LMS Algorithm

The description of SPU-LMS is similar to that of S-LMS (Section II). The only difference is as as follows. At a given iteration, k, for S-LMS one of the sets S_i, $i = 1, \ldots, P$ is chosen in a *predetermined fashion*, whereas for SPU-LMS, one of the sets S_i is sampled at *random* from $\{S_1, S_2, \ldots, S_P\}$ with probability $1/P$ and subsequently the update is performed. i.e.,

$$w_{k+1,j} = \begin{cases} w_{k,j} + \mu e_k^* x_{k,j} & \text{if } j \in S_i \\ w_{k,j} & \text{otherwise} \end{cases} \tag{9}$$

where $e_k = d_k - W_k^H X_k$. The above update equation can be written in a more compact form

$$W_{k+1} = W_k + \mu e_k^* I_k X_k \tag{10}$$

where I_k now is a random matrix chosen with equal probability from $\mathcal{I}_i$, $i = 1, \ldots, P$ (recall that $\mathcal{I}_i$ is obtained by zeroing out the jth row of the identity matrix I if $j \notin S_i$).

A. Analysis: Stationary Stochastic Signals, Independent Input, and Coefficient Vectors

For the stationary signal analysis of SPU-LMS, the desired signal d_k is assumed to satisfy the same conditions as in Section II, namely, $d_k = W_{\mathrm{opt}}^H X_k + n_k$. In this section, we make the usual assumptions used in the analysis of standard LMS [3]: We assume that X_k is a Gaussian random vector and that X_k and $V_k = W_k - W_{\mathrm{opt}}$ are independent. I_k and X_k are independent of each other by definition. We also assume, in this section, for tractability, that $R = E[X_k X_k^H]$ is block diagonal such that $\sum_{i-1}^{P} \mathcal{I}_i R \mathcal{I}_i = R$.

For convergence-in-mean analysis, we obtain the following update equation conditioned on a choice of S_i.

$$\begin{aligned} E[V_{k+1}|S_i] &= (I - \mu I_k R) E[V_k|S_i] \\ &= (I - \mu \mathcal{I}_i R) E[V_k|S_i] \end{aligned}$$

which after averaging over all choices of S_i gives

$$E[V_{k+1}] = \left(I - \frac{\mu}{P} R\right) E[V_k]. \tag{11}$$

To obtain the above equation, we have made use of the fact that the choice of S_i is independent of V_k and X_k. Therefore, μ has to satisfy $0 < \mu < (2P/\lambda_{\max})$ to guarantee convergence in mean.

For the convergence-in-mean square analysis of SPU-LMS, the convergence of the error variance $E[e_k e_k^*]$ is studied as in [20]. Under the independence assumptions, we obtain $E[e_k e_k^*] = \xi_{\min} + \mathrm{tr}\{RE[V_k V_k^H]\}$, where $\xi_{\min}$ is as defined earlier.

First, conditioned on a choice of S_i, the evolution equation of interest for $\mathrm{tr}\{RE[V_k V_k^H]\}$ is given by

$$\begin{aligned} RE\left[V_{k+1}V_{k+1}^H|S_i\right] = &RE\left[V_k V_k^H|S_i\right] - 2\mu R \mathcal{I}_i RE\left[V_k V_k^H|S_i\right] \\ &+ \mu^2 \mathcal{I}_i R \mathcal{I}_i E\left[X_k X_k^H A_k X_k X_k^H|S_i\right] + \mu^2 \xi_{\min} R \mathcal{I}_i R \mathcal{I}_i \end{aligned} \tag{12}$$

where $A_k = E[V_k V_k^H]$. Let $U_k = QV_k$, where Q satisfies $QRQ^H = \Lambda$. Applying the definition of U_k to (12), we obtain the equation

$$G_{k+1} = \left(I - \frac{2\mu}{P}\Lambda + \frac{\mu^2}{P}\Lambda^2 + \frac{\mu^2}{P}\Lambda^2 \mathbf{1}\mathbf{1}^T\right) G_k + \frac{\mu^2}{P}\xi_{\min}\Lambda^2 \mathbf{1} \tag{13}$$

where G_k is a vector of diagonal elements of $\Lambda E[U_k U_k^H]$, and $\mathbf{1}$ is an $N \times 1$ vector of ones.

It is easy to obtain the following necessary and sufficient conditions (following the procedure of [20]) for convergence of $E[V_k V_k^H]$ from (12)

$$0 < \mu < \frac{2}{\lambda_{\max}}$$
$$\eta(\mu) \stackrel{\text{def}}{=} \sum_{i=1}^{N} \frac{\mu\lambda_i}{2-\mu\lambda_i} < 1 \tag{14}$$

which are independent of P and identical to that of LMS. As pointed out in Section II-A, the above conditions have been obtained under the independence assumption that are not valid in general.

The integrated MSE difference [20]

$$\mathcal{J} = \sum_{k=0}^{\infty} [\xi_k - \xi_\infty] \tag{15}$$

introduced in [12] is used as a measure of the convergence rate and $M(\mu) = (\xi_\infty - \xi_{\min})/\xi_{\min}$ as a measure of misadjustment. It is easily established that the misadjustment takes the form

$$M(\mu) = \frac{\eta(\mu)}{1-\eta(\mu)} \tag{16}$$

which is the same as that of the standard LMS. Thus, we conclude that random update of subsets has no effect on the final excess mean-squared error.

Finally, it is straightforward to show (following the procedure of [12]) that the integrated MSE difference is

$$\mathcal{J} = P\,\text{tr}\left\{[2\mu\Lambda - \mu^2\Lambda^2 - \mu^2\Lambda^2\mathbf{1}\mathbf{1}^T]^{-1}(G_0 - G_\infty)\right\} \tag{17}$$

which is P times the quantity obtained for standard LMS algorithm. Therefore, we conclude that for block diagonal R, random updating slows down convergence by a factor of P without affecting the misadjustment. Furthermore, it can be easily verified that a much simpler condition $0 < \mu < (1/\text{tr}\{R\})$, provides a sufficient region for convergence of SPU-LMS and the standard LMS algorithm.

B. Analysis: Deterministic Signals

Here, we follow the analysis for LMS, albeit extended to complex signals, which is given in [25, pp. 140–143]. We assume that the input signal X_k is bounded, that is $\sup_k(X_k^H X_k) \le B < \infty$, and that the desired signal d_k follows the model

$$d_k = W_{\text{opt}}^H X_k$$

which is different from (2) in that d_k is assumed to be perfectly predictable from X_k.

Define $V_k = W_k - W_{\text{opt}}$ and $e_k = d_k - W_k^H X_k$.

Lemma 1: If $\mu < 2/B$, then $\overline{e_k^2} \to 0$ as $k \to \infty$. Here, $\overline{\{\cdot\}}$ indicates statistical expectation over all possible choices of S_i, where each S_i is chosen randomly with equal probability from $\{S_1, \ldots, S_P\}$.

Proof: See Appendix I.

For a positive definite matrix A_k, we say that A_k converges exponentially fast to zero if there exists a γ, $0 < \gamma < 1$ and a positive integer K such that $\text{tr}\{A_{k+K}\} \le (1-\gamma)\text{tr}\{A_k\}$ for all k. $\text{tr}\{A\}$ denotes the trace of the matrix A.

Theorem 4: If $\mu < 2/B$ and the signal satisfies the following persistence of excitation condition, for all k, there exist $K < \infty$, $\alpha_1 > 0$ and $\alpha_2 > 0$ such that

$$\alpha_1 I < \sum_{i=k}^{k+K} X_i X_i^H < \alpha_2 I \tag{18}$$

then $\overline{V_k}^H \overline{V_k} \to 0$, and $\overline{V_k^H V_k} \to 0$ exponentially fast.

Proof: See Appendix I.

Condition (18) is identical to the persistence of excitation condition for standard LMS [25]. Therefore, the sufficient condition for exponential stability of LMS is enough to guarantee exponential stability of SPU-LMS.

C. Analysis: Correlated Input and Coefficient Vectors

In this section, the performance of LMS and SPU-LMS is analytically compared in terms of stability and misconvergence when the uncorrelated input and coefficient signal vectors assumption is invalid. Unlike the analysis in Section III-A, where the convergence analysis and the performance analysis could be tackled with the same set of equations, here, the stability and performance analyzes have to be done separately. For this, we employ the theory, which is extended here to circular complex random variables developed in [16] for stability analysis and [2] for final mean-squared error analysis. Our analysis holds for the broad class of signals that are ϕ-mixing. Mixing conditions provide a very general and powerful way to describe the rate of decay of the dependence between pairs of samples as the sample times are moved farther apart. Such conditions are much weaker than conditions on the rate of decay of the autocorrelation function, which are restricted to second-order analysis and Gaussian processes. For this reason, general mixing conditions, such as the ϕ-mixing condition defined in Appendix III, have been widely used in adaptive signal processing and adaptive detection [2], [7], [16], [17], [23] to analyze convergence of algorithms for dependent processes. We adopt this framework in this paper (see Appendices II and IV for detailed proofs and definitions) and summarize the results in this section.

The analysis in Section III-A is expected to hold for small μ, even when the uncorrelated input and coefficient vectors assumption is violated. It is, however, not clear for what values of μ the results from Section III-A are valid. The current section makes the dependence of the value of μ explicit and concludes that stability and performance of SPU-LMS are similar to that of LMS.

Result 1 (Stationary Gaussian Process): Let x_k be a stationary Gaussian random process such that $E[x_k x_{k-l}] = r_l \to 0$ as $l \to \infty$ and $X_k = [x_k\ x_{k-1}\ \ldots\ x_{k-n+1}]$, then for any $p \ge 1$, there exist constants $\mu^* > 0$, $M > 0$, and $\alpha \in (0,1)$ such that for all $\mu \in (0, \mu^*]$ and for all $t \ge k \ge 0$

$$\left[E\left\|\prod_{j=k+1}^{t}\left(I - \mu I_j X_j X_j^H\right)\right\|^p\right]^{\frac{1}{p}} \le M(1-\mu\alpha)^{t-k}$$

if and only if the input correlation matrix $E[X_k X_k^H] = R_{xx}$, is positive definite.

Proof: See Appendix II.

It is easily seen from the extension of [16] to complex signals that the LMS algorithm requires the same necessary and sufficient conditions for convergence (see Appendix II). Therefore, the necessary and sufficient conditions for convergence of SPU-LMS are identical to those of LMS.

The analysis in Result 1 validates the analysis in Section III-A, for similar input signals, where the analysis was done under the independence assumption. In both cases, we conclude that necessary and sufficient condition for convergence is that the covariance matrix be positive definite. Although Section III-A also gives some bounds on the step-size parameter μ, it is known they are not very reliable as the analysis is valid only for very small μ.

The mean squared analysis on $V_k \stackrel{\text{def}}{=} W_k - W_{\text{opt}}$ is based on the analysis in [2], which follows the method of successive approximation. The details of the extension of this method to SPU-LMS are provided in Appendix IV. The analysis in this section is performed by assuming that

$$d_k = X_k^H W_{\text{opt}} + n_k.$$

The effectiveness of the method is illustrated in Results 2 and 3, where the steady-state performance of the two algorithms is compared for two simple scenarios where the independence assumption is clearly violated.

Result 2 (i. i. d. Gaussian Process): Let $X_k = [x_k\ x_{k-1}\ \dots\ x_{k-N+1}]^T$, where N is the length of the vector X_k. $\{x_k\}$ is a sequence of zero mean i.i.d Gaussian random variables. Let σ^2 denote the variance of x_k and σ_v^2 denote the variance of n_k. It is assumed that n_k is a white i.i.d. Gaussian noise. Then, for LMS, we have

$$\lim_{k\to\infty} E\left[V_k V_k^H\right] = \mu^2\left[\frac{\sigma_v^2}{2\mu}I + \frac{N\sigma^2\sigma_v^2}{4}I + C\mu^{\frac{1}{2}}I\right] \tag{19}$$

and for SPU-LMS, assuming N to be a multiple of P and sets S_i to be mutually exclusive, we have

$$\lim_{k\to\infty} E\left[V_k V_k^H\right] = \mu^2\left[\frac{\sigma_v^2}{2\mu}I + \frac{\frac{(N+1)P-1}{P}\sigma^2\sigma_v^2}{4}I + C\mu^{\frac{1}{2}}I\right].$$

Note that the constant C in the final mean square expression for SPU-LMS is the same as that for LMS. Therefore, for large N, it can be seen that SPU-LMS is marginally worse than LMS in terms of misadjustment.

Proof: See Appendix IV-A.

It will be interesting to see how the results above compare to the results obtained under the independence assumptions analysis in Section III-A. From (13), we obtain the vector of diagonal elements of $\lim_{k\to\infty} E[V_k V_k^H]$ $\mathcal{V}_d$ to be

$$\mathcal{V}_d = \mu^2\left[\frac{\sigma_v^2}{2\mu}\mathbf{1} + \frac{(N+1)\sigma^2\sigma_v^2}{4}\mathbf{1}\right] + O(\mu^4)\mathbf{1}$$

for both LMS and SPU-LMS, where $\mathbf{1}$ is an $N \times 1$ vector of ones. The analysis in this section gives

$$\mathcal{V}_d = \mu^2\left[\frac{\sigma_v^2}{2\mu}\mathbf{1} + \frac{N\sigma^2\sigma_v^2}{4}\mathbf{1}\right] + O\left(\mu^{\frac{3}{2}}\right)\mathbf{1}$$

for LMS and

$$\mathcal{V}_d = \mu^2\left[\frac{\sigma_v^2}{2\mu}\mathbf{1} + \frac{\frac{(N+1)P-1}{P}\sigma^2\sigma_v^2}{4}\mathbf{1}\right] + O\left(\mu^{\frac{3}{2}}\right)\mathbf{1}$$

for SPU-LMS.

Result 3 (Spatially Uncorrelated Temporally Correlated Process): Let X_k be given by

$$X_k = \kappa X_{k-1} + \sqrt{1-\kappa^2}U_k$$

where U_k is a vector of circular Gaussian random variables with unit variance. Here, in addition, it is assumed that n_k is a white i.i.d. Gaussian noise with variance σ_v^2. Then, for LMS, we have

$$\lim_{k\to\infty} E\left[V_k V_k^H\right] = \mu^2\left[\frac{\sigma_v^2}{2\mu}I + \frac{N\sigma_v^2}{4}I + C\mu^{\frac{1}{2}}I\right] \tag{20}$$

and for SPU-LMS, assuming N to be a multiple of P and sets S_i to be mutually exclusive, we have

$$\lim_{k\to\infty} E\left[V_k V_k^H\right] = \mu^2\left[\frac{\sigma_v^2}{2\mu}I + \frac{\sigma^2}{4}\left[N+1-\frac{1}{P}\right]I + C\mu^{\frac{1}{2}}I\right].$$

Here, in addition, for large N, SPU-LMS is marginally worse than LMS in terms of misadjustment.

Proof: See Appendix IV-B.

Here, in addition, the results obtained above can be compared to the results obtained from the analysis in Section III-A. From (13), we obtain $\mathcal{V}_d$ to be

$$\mathcal{V}_d = \mu^2\left[\frac{\sigma_v^2}{2\mu}\mathbf{1} + \frac{(N+1)\sigma_v^2}{4}\mathbf{1}\right] + O(\mu^4)\mathbf{1}$$

for both LMS and SPU-LMS. The analysis in this section gives

$$\mathcal{V}_d = \mu^2\left[\frac{\sigma_v^2}{2\mu}\mathbf{1} + \frac{N\sigma_v^2}{4}\mathbf{1}\right] + O\left(\mu^{\frac{3}{2}}\right)\mathbf{1}$$

for LMS and

$$\mathcal{V}_d = \mu^2\left[\frac{\sigma_v^2}{2\mu}\mathbf{1} + \left(N+1-\frac{1}{P}\right)\frac{\sigma^2}{4}\mathbf{1}\right] + O\left(\mu^{\frac{3}{2}}\right)\mathbf{1}$$

for SPU-LMS.

Therefore, the analysis in this section highlights differences in the convergence of LMS and SPU-LMS that would not have been apparent from the analysis in Section III-A. It can be noted that for small N the penalty for assuming independence is not insignificant (especially for SPU-LMS). However, for large N, the independence assumption analysis manages to yield a reliable estimate, even for larger values of μ, in spite of the fact that the assumption is clearly violated.

IV. Discussion and Examples

It is useful to compare the performance of the new algorithm to those of the existing algorithms by performing the analyses of Sections III-A, B, and C on the periodic Partial Update LMS

Algorithm (P-LMS) and the sequential Partial Update LMS Algorithm (S-LMS). To do that, we first need the update equation for P-LMS, which is as follows:

$$W_{k+P} = W_k + \mu e_k^* X_k.$$

We begin with comparing the convergence-in-mean analysis of the partial update algorithms. Combining P-iterations, we obtain for LMS $V_{k+P} = (I - \mu R)^P V_k$, for P-LMS $V_{k+P} = (I - \mu R)V_k$, for SPU-LMS $V_{k+P} = (I - (\mu/P)R)^P V_k$, and, finally, for S-LMS (assuming $R = \sum_{i=1}^{P} \mathcal{I}_i R \mathcal{I}_i$) $V_{k+P} = (I - \mu R)V_k$. Therefore, the rate of decay of all the partial update algorithms is P times slower than that of LMS.

For convergence-in-mean square analysis of Section III-A, we will limit the comparison to P-LMS. The convergence of Sequential LMS algorithm has been analyzed using the small μ assumption in [11]. Under this assumption for stationary signals, using the independence assumption, the conditions for convergence turn out to be the same as that of SPU-LMS. For P-LMS using the method of analysis described in [20], it can be inferred that the conditions for convergence are identical to standard LMS. That is, (14) holds also for P-LMS. In addition, the misadjustment factor remains the same. The only difference between LMS and P-LMS is that the integrated MSE $\mathcal{J}$ (15) for P-LMS is P times larger than that of LMS. Therefore, we again conclude that the behavior of SPU-LMS and P-LMS algorithms is very similar for stationary signals.

However, for deterministic signals the difference between P-LMS and SPU-LMS becomes evident from the persistence of excitation condition. The persistence of excitation condition for P-LMS is [11] as follows: For all k and for all j, $1 \le j < N/P$, there exist $K < \infty$, $\alpha_1 > 0$ and $\alpha_2 > 0$ such that

$$\alpha_1 I < \sum_{i=k\frac{N}{P}+j}^{(k+K)\frac{N}{P}+j} X_i X_i^H < \alpha_2 I. \tag{21}$$

Since any deteriministic signal satisfying (21) also satisfies (18) but not vice-versa, it can be inferred that (21) is stricter than that for SPU-LMS (18).

Taking this further, using the analysis in Appendix II, for mixing signals, the persistence of excitation condition can similarly be shown to be the following: There exists an integer $K > 0$ and a constant $\delta > 0$ such that for all $k \ge 0$ and for all j, $1 \le j < N/P$

$$\sum_{i=k\frac{N}{P}+j}^{(k+K)\frac{N}{P}+j} E\left[X_i X_i^H\right] \ge \delta I.$$

Here, in addition, it can be seen that this condition is stricter than that of SPU-LMS (25). In fact, in Section IV-A, signals are constructed, based on the persistence of excitation conditions for SPU-LMS and P-LMS, for which P-LMS is guaranteed not to converge, whereas SPU-LMS will converge.

The analysis of Appendix II can be extended to S-LMS if an additional requirement of stationarity is imposed on the excitation signals. For such signals, it can be easily seen that the necessary and sufficient conditions for statibility of LMS, SPU-LMS and P-LMS are exactly the same and are sufficient for exponential stability of S-LMS (see Appendix III for details).

In addition, applying the analysis of Appendix IV used to derive Results 2 and 3, it can be easily seen that the final error covariance matrix for P-LMS is same as that of LMS [see (19) and (20)]. Exactly the same results can be obtained for S-LMS as well by combining the results of Appendix III with the analysis in Appendix IV restricted to stationary ϕ-mixing signals.

For nonstationary signals, the convergence of S-LMS is an open question, although analysis for some limited cyclo-stationary signals has been performed in [15]. In this paper, we show through simulation examples that this algorithm diverges for certain nonstationary signals and, therefore, should be employed with caution.

In summary, for stationary signals all three algorithms (P-LMS, S-LMS, and SPU-LMS) enjoy the same convergence properties as LMS. It is for nonstationary signals that S-LMS and P-LMS might fail to converge, and it is for such signals that the advantage of SPU-LMS comes to the fore. SPU-LMS enjoys the same convergence properties as LMS, even for nonstationary signals, in the sense that it is guaranteed to converge for all signals that LMS converges for.

A. Numerical Examples

In the first two examples, we simulated an m-element uniform linear antenna array operating in a multiple signal environment. Let A_i denote the response of the array to the ith plane wave signal:

$$A_i = \Big[e^{-\mathbf{j}\left(\frac{m}{2}-\tilde{m}\right)\omega_i}\ e^{-\mathbf{j}\left(\frac{m}{2}-1-\tilde{m}\right)\omega_i} \cdots \\ e^{\mathbf{j}\left(\frac{m}{2}-1-\tilde{m}\right)\omega_i}\ e^{\mathbf{j}\left(\frac{m}{2}-\tilde{m}\right)\omega_i}\Big]^T$$

where $\tilde{m} = (m+1)/2$ and $\omega_i = 2\pi D \sin\theta_i/\lambda$, $i = 1, \ldots, M$. θ_i is the broadside angle of the ith signal, D is the interelement spacing between the antenna elements, and λ is the common wavelength of the narrowband signals in the same units as D and $(2\pi D/\lambda) = 2$. The array output at the kth snapshot is given by $X_k = \sum_{i=1}^{M} A_i s_{k,i} + n_k$, where M denotes the number of signals, the sequence $\{s_{k,i}\}$ the amplitude of the ith signal, and n_k the noise present at the array output at the kth snapshot. The objective, in both the examples, is to maximize the SNR at the output of the beamformer. Since the signal amplitudes are random, the objective translates to obtaining the best estimate of $s_{k,1}$, which is the amplitude of the desired signal, in the MMSE sense. Therefore, the desired signal is chosen as $d_k = s_{k,1}$.

Example 1: In the first example (see Fig. 1), the array has four elements, and a single planar waveform with amplitude $s_{k,1}$ propagates across the array from direction angle $\theta_1 = 0$. The amplitude sequence $\{s_{k,1}\}$ is a binary phase shifty keying (BPSK) signal with period four taking values on $\{-1,1\}$ with equal probability. The additive noise n_k is circular Gaussian with variance 0.25 and mean 0. In all the simulations for SPU-LMS, P-LMS, and S-LMS, the number of subsets for partial updating P was chosen to be 4, that is, a single coefficient is updated at each iteration. It can be easily determined from (14) that for Gaussian and independent signals, the necessary and sufficient condition for convergence of the update equations for LMS and SPU-LMS under the independence assumptions analysis is $\mu < 0.225$. Fig. 2 shows representative trajectories of the empir-

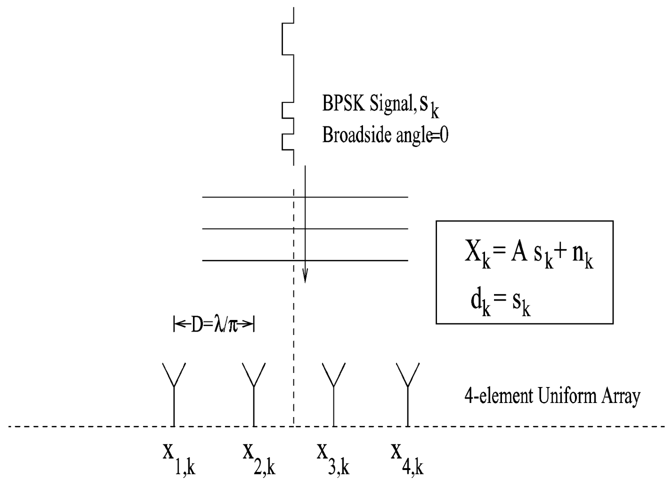


Fig. 1. Signal scenario for Example 1.

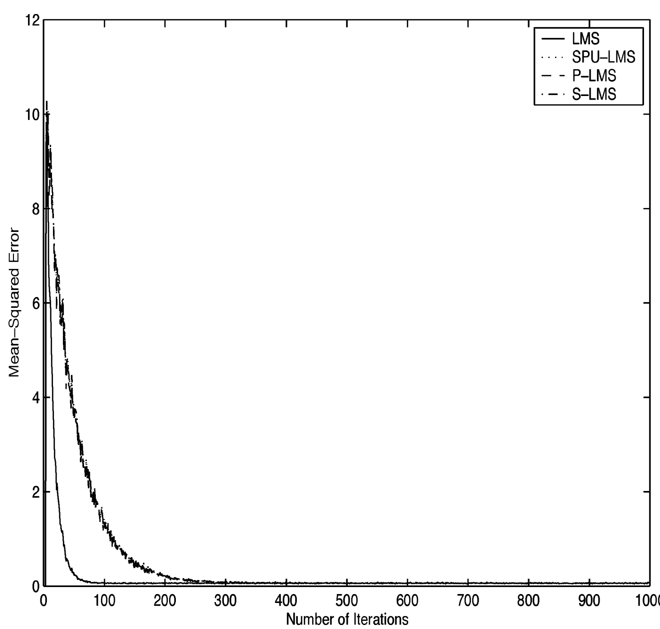


Fig. 2. Trajectories of MSE for Example 1.

ical mean-squared error for the LMS, SPU-LMS, P-LMS, and S-LMS algorithms averaged over 100 trials for $\mu = 0.01$. All algorithms were found to be stable for the BPSK signals, even for μ values greater than 0.225. It was only as μ approached 0.32 that divergent behavior was observed. As expected, LMS and SPU-LMS were observed to have similar μ regions of convergence. It is also clear from Fig. 2 that, as expected, SPU-LMS, P-LMS, and S-LMS take roughly four times longer to converge than LMS.

Example 2: In the second example, we consider an eight-element uniform linear antenna array with one signal of interest propagating at angle θ_1 and three interferers propagating at angles θ_i, $i = 2, 3$, and 4 (see Fig. 3). The array noise n_k is again mean 0 circular Gaussian but with variance 0.001. Signals are generated, such that $s_{k,1}$ is stationary and $s_{k,i}$, $i = 2, 3$, and 4 are cyclostationary with period four, which make both S-LMS and P-LMS nonconvergent. All the signals were chosen to be independent from time instant to time

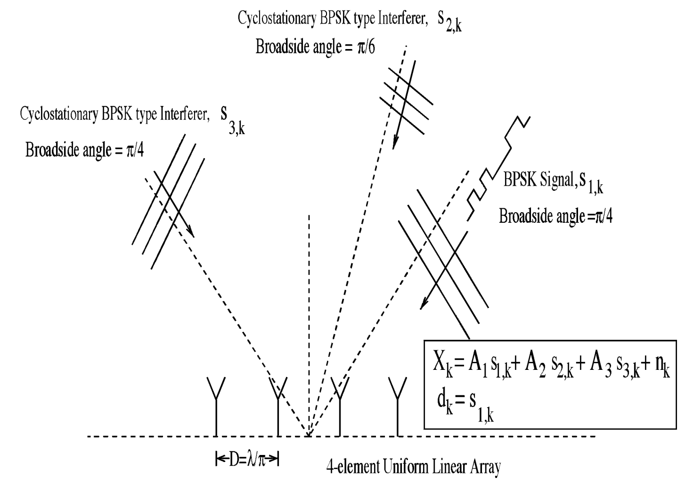


Fig. 3. Signal scenario for Example 2.

instant. First, we found signals for which S-LMS does not converge by the following procedure. Make the small μ approximation $I - \mu\sum_{i=1}^{P} I_i E[X_{k+i}X_{k+i}^H]$ to the transition matrix $\prod_{i=1}^{P}(I - \mu I_i E[X_{k+i}X_{k+i}])$, and generate sequences $s_{k,i}$, $i = 1, 2, 3$, and 4 such that $\sum_{i=1}^{P} I_i E[X_{k+i}X_{k+i}^H]$ has roots in the negative left half plane. This ensures that $I - \mu\sum_{i=1}^{P} I_i E[X_{k+i}X_{k+i}^H]$ has roots outside the unit circle. The sequences found in this manner were then verified to cause the roots to lie outside the unit circle for all μ. One such set of signals found was that $s_{k,1}$ is equal to a BPSK signal with period one taking values in $\{-1, 1\}$ with equal probability. The interferers $s_{k,i}$, $i = 2, 3$, and 4 are cyclostationary BPSK type signals taking values in $\{-1, 1\}$ with the restriction that $s_{k,2} = 0$ if $k\%4 \neq 1$, $s_{k,3} = 0$ if $k\%4 \neq 2$ and $s_{k,4} = 0$ if $k\%4 \neq 3$. Here, $a\%b$ stands for a modulo b. θ_i, $i = 1, 2, 3$, and 4 are chosen such that $\theta_1 = 1.0388$, $\theta_2 = 0.0737$, $\theta_3 = 1.0750$, and $\theta_4 = 1.1410$. These signals render the S-LMS algorithm unstable for all μ.

The P-LMS algorithm also fails to converge for the signal set described above, irrespective of μ and the choice of θ_1, θ_2, θ_3, and θ_4. Since P-LMS updates the coefficients every fourth iteration, it sees at most one of the three interfering signals throughout all its updates and, hence, can place a null at, at most, one signal incidence angle θ_i. Fig. 4 shows the envelopes of the e_k^2 trajectories of S-LMS and P-LMS for the signals given above with the representative value $\mu = 0.03$. As can be seen, P-LMS fails to converge, whereas S-LMS shows divergent behavior. SPU-LMS and LMS were observed to converge for the signal set described above when $\mu = 0.03$.

Example 3: In the third example, consider a four-tap filter ($N = 4$) with a time series input, that is, $X_k = [x_k\ x_{k-1}\ x_{k-2}\ x_{k-3}]^T$. The input, the filter coefficients, and the desired output are all real valued. In this example, the goal is to reconstruct the transmitted BPSK signal s_k from the received signal x_k at the receiver using a linear filter. x_k is a distorted version of s_k when s_k passes through a linear channel with transfer function given by $1/(1+0.4z^{-1}-0.26z^{-2}-0.2z^{-3})$. The receiver noise n_k is a zero mean Gaussian noise with variance 0.01. s_k is a signal with symbol duration of four samples. The desired output d_k is now simply given by $d_k = s_k$. The

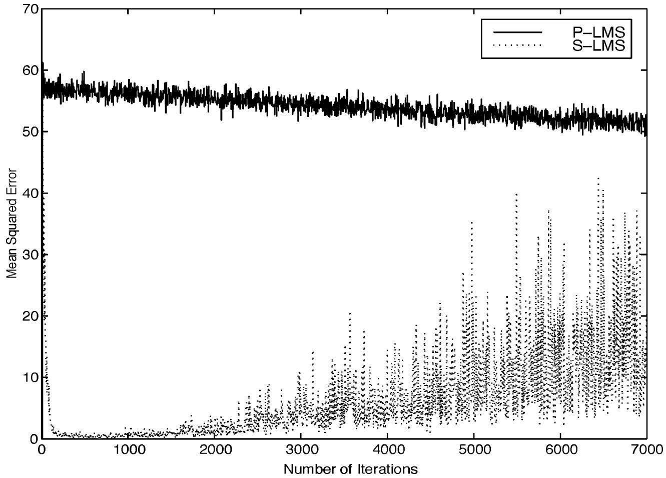


Fig. 4. Trajectories of MSE for Example 2.

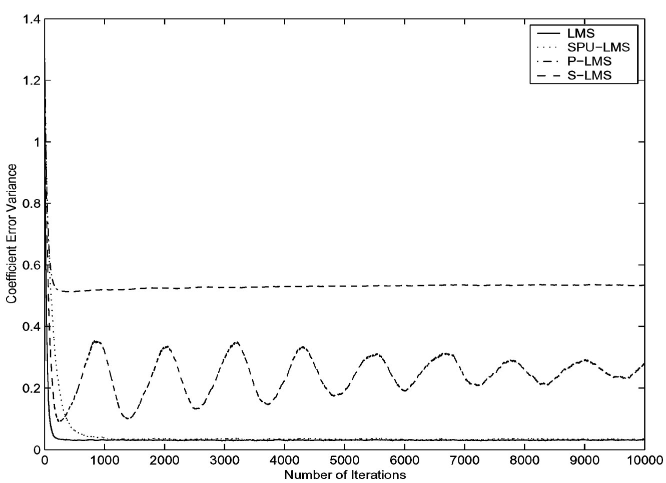


Fig. 5. Trajectories of MSE for LMS, SPU-LMS, P-LMS, and S-LMS for Example 3.

update is such that one coefficient is updated per iteration, i.e., $P = 4$. In this case, the coefficient error variance is plotted rather than the mean squared error as this is a better indication of system performance. Fig. 5 shows the trajectories of coefficient-error variance for LMS, SPU-LMS, P-LMS, and S-LMS for a representative value of $\mu = 0.01$, respectively. As can be seen, P-LMS and S-LMS fail to converge, whereas LMS and SPU-LMS do converge.

Example 4: In the fourth example, we show a nonstationary signal for which Max PU-LMS and SMPU-NLMS algorithms do not converge. For algorithmic details of these two algorithms and their analysis, see [8]. The two algorithms can be made to not converge by first constructing deteriministic signals for which their behavior is the same as that of S-LMS and then finding a candidate among such signals for which S-LMS diverges.

Consider a four-tap filter with time series input $X_k = S_k = [s_k \; s_{k-1} \; s_{k-2} \; s_{k-3}]^T$. The goal in this example is to obtain the best estimate of $W_{\rm opt} = [1 \; 0.4 \; -0.26 \; -0.204]$ from $d_k = W_{\rm opt}^T X_k + n_k$ and X_k, where n_k is a Gaussian random variable with zero mean and variance of 0.01. The update is such that one coefficient is updated per iteration, i.e., $P = 4$.

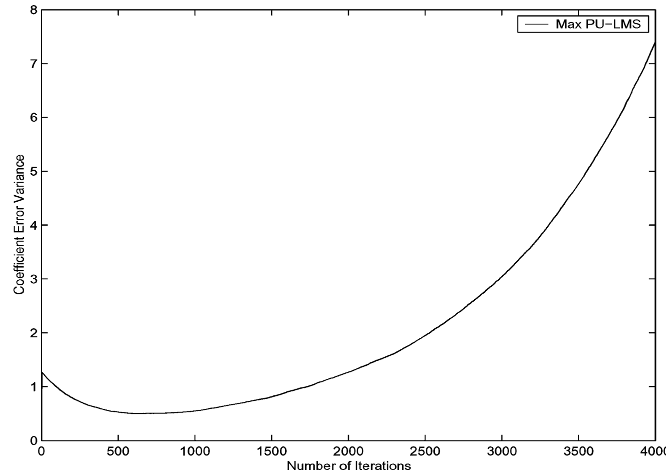


Fig. 6. Trajectory of MSE for Max PU-LMS for Example 4.

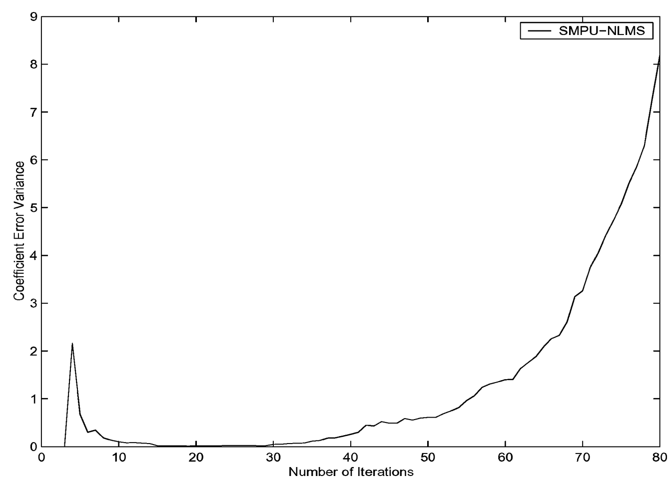


Fig. 7. Trajectory of MSE for SMPU-NLMS for Example 4.

s_k is chosen to be a deteriministic sequence of the following form $s_k = b_{k\%4}$, where $\{b_0, b_1, b_2, b_3\}$ is a fixed sequence satisfying $|b_0| < |b_1| < |b_2| < |b_3|$. Such a restriction on s_k and b_k ensures that SMPU-NLMS in updating only one coefficient per iteration ends up updating the coefficients in a sequential manner. For this signal, Max PU-LMS also updates the coefficients in a sequential manner, and its behavior is exactly that of S-LMS. The values b_k and $k = 0, \ldots, 3$ were chosen such that $\sum_{i=1}^{4} I_i S_{4*k+i+2} S_{4*k+i+2}^T$ for all k has eigenvalues in the left half plane. That means that the small μ approximation of the S-LMS update matrix $\prod_{i=1}^{4}(I - \mu I_i S_{4*k+i+2} S_{4*k+i+2}^T)$ has eigenvalues outside the unit circle. For such input signals, there is a good likelihood that SMPU-NLMS will diverge along with S-LMS and Max PU-LMS. A signal for which the three algorithms have been observed to diverge has $b_0 = 0.1924$, $b_1 = -0.5364$, $b_2 = -0.5521$, and $b_3 = 0.6087$.

Here also, the coefficient error variance is plotted rather than the mean squared error. Figs. 6 and 7 show the trajectory of coefficient-error variance for MAX PU-LMS for a representative value of $\mu = 0.01$ and for SMPU-NLMS for a representative value of $\gamma = 0.01$ (for a description of γ, see [8]), respectively. Fig. 8 shows the corresponding trajectories for LMS and SPU-LMS, again for a representative value of $\mu = 0.01$. As

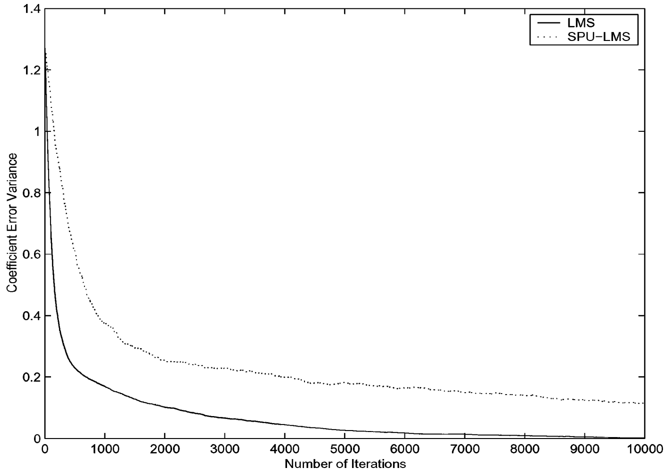


Fig. 8. Trajectories of MSE for LMS and SPU-LMS for Example 4.

can be seen, Max PU-LMS and SMPU-NLMS fail to converge, while SPU-LMS and LMS do.

V. Conclusion and Future Work

In this paper, the sequential partial update LMS algorithm has been analyzed, and a new algorithm based on randomization of filter coefficient subsets for partial updating of filter coefficients has been proposed.

For S-LMS, stability bounds on step-size parameter μ for wide sense stationary signals have been derived. It has been shown that if the regular LMS algorithm converges in mean, then so does the sequential LMS algorithm for the general case of arbitrary but fixed ordering of the sequence of partial coefficient updates. Relaxing the assumption of independence, for stationary signals, stability and second-order (mean square convergence) analysis of S-LMS has been performed. The analysis was used to establish that S-LMS has similar behavior as LMS.

In the context of nonstationary signals the poor convergence properties of S-LMS and Periodic LMS have been demonstrated, and as a result, a new algorithm SPU-LMS with better performance has been designed. For SPU-LMS the conditions on step-size for convergence-in-mean and mean-square were shown to be equivalent to those of standard LMS. It was verified by theory and by simulation that LMS and SPU-LMS have similar regions of convergence. It was also shown that the Stochastic Partial Update LMS algorithm has the same performance as P-LMS and S-LMS for stationary signals but can have superior performance for some cyclo-stationary and deterministic signals. It was also demonstrated that the randomization of filter coefficient updates does not increase the final steady-state error as compared to the regular LMS algorithm.

The idea of random choice of subsets proposed in this paper can be extended to include arbitrary subsets of size N/P and not just subsets from a particular partition. No special advantage is immediately evident from this extension, however.

In the future, tighter bounds on the convergence rate of the mean update equation of S-LMS for stationary signals can be established for the general case of input correlation matrix R. Necessary and sufficient conditions for the convergence of the algorithm for the general case of mixing-signals still need to be derived. These can be addressed in the future.

In addition, it can be investigated whether performance analysis of Max PU-LMS and SMPU-NLMS algorithms mentioned in Section I can be performed using the techniques employed in this paper. Special emphasis should be laid on nonstationary signal performance because, as has been shown through a numerical example, these algorithms can diverge for such signals.

Appendix I
Proofs of Lemma 1 and Theorem 4

Proof of Lemma 1: First note that $e_k = -V_k^H X_k$. Next, consider the Lyapunov function $\mathcal{L}_{k+1} = \overline{V_{k+1}^H V_{k+1}}$, where $\overline{\{\cdot\}}$ is as defined in Lemma 1. Averaging the following update equation for $V_{k+1}^H V_{k+1}$:

$$V_{k+1}^H V_{k+1} = V_k^H V_k - \mu \mathrm{tr}\left\{V_k V_k^H X_k X_k^H \mathcal{I}_i\right\} \\ -\mu \mathrm{tr}\left\{V_k V_k^H \mathcal{I}_i X_k X_k^H\right\} + \mu^2 \mathrm{tr}\left\{V_k V_k^H X_k X_k^H \mathcal{I}_i X_k X_k^H\right\}$$

over all possible choices of $S_i, i = 1, \ldots, P$, we obtain

$$\mathcal{L}_{k+1} = \mathcal{L}_k - \frac{\mu}{P}\mathrm{tr}\left\{\overline{V_k V_k^H} X_k \left(2 - \mu X_k X_k^H\right) X_k^H\right\}.$$

Since $\sup_k (X_k^H X_k) \leq B < \infty$, the matrix $(2I - \mu X_k X_k^H) - (2I - \mu BI)$ is positive definite. Therefore

$$\mathcal{L}_{k+1} \leq \mathcal{L}_k - \frac{\mu}{P}(2 - \mu B)\mathrm{tr}\left\{\overline{V_k V_k^H} X_k X_k^H\right\}.$$

Since $\mu < 2/B$

$$\mathcal{L}_{k+1} \leq \mathcal{L}_k - \mathrm{tr}\left\{\overline{V_k V_k^H} X_k X_k^H\right\}.$$

Noting that $\overline{e_k^2} = \mathrm{tr}\{\overline{V_k V_k^H} X_k X_k^H\}$, we obtain

$$\mathcal{L}_{k+1} + \sum_{l=0}^{k} \overline{e_k^2} \leq \mathcal{L}_0$$

and since $\mathcal{L}_0 < \infty$, we have $\overline{e_k^2} = O(1/k)$ and $\lim_{k\to\infty} \overline{e_k^2} = 0$. ■

Before proving Theorem 4, we need Lemmas 2 and 3. We reproduce the proof of Lemma 2 from [25] using our notation because this enables us to understand the proof of Lemma 3 better.

Lemma 2—[25, Lemma 6.1 p. 143–144]: Let X_k satisfy the persistence of excitation condition in Theorem 4, and let

$$\Pi_{k,k+D} = \begin{cases} \prod_{l=k}^{k+D} \left(I - \frac{\mu}{P} X_l X_l^H\right), & \text{if } D \geq 0 \\ 1, & \text{if } D < 0 \end{cases}$$

and

$$\mathcal{G}_k = \sum_{l=0}^{K} \Pi_{k,k+l-1}^H X_{k+l} X_{k+l}^H \Pi_{k,k+l-1}$$

where K is as defined in Theorem 4. Then, $\mathcal{G}_k - \eta I$ is a positive definite matrix for some $\eta > 0$ and $\forall k$.

Proof: The proof is by contradiction. Then, for some vector ω such that $\omega^H\omega = 1$, suppose that we do not have $\omega^H \mathcal{G}_k \omega \le c^2$, where c is any arbitrary positive number.

Then

$$\sum_{l=0}^{K} \omega^H \Pi_{k,k+l-1}^H X_{k+l} X_{k+l}^H \Pi_{k,k+l-1} \omega \le c^2 \Rightarrow$$
$$\omega^H \Pi_{k,k+l-1}^H X_{k+l} X_{k+l}^H \Pi_{k,k+l-1} \omega \le c^2 \quad \text{for } 0 \le l \le K.$$

Choosing $l = 0$, we obtain $\omega^H X_k X_k^H \omega \le c^2$ or $\|\omega^H X_k\| \le c$.

Choosing $l = 1$, we obtain

$$\left\| \omega^H \left(I - \frac{\mu}{P} X_k X_k^H \right) X_{k+1} \right\| \le c.$$

Therefore

$$\begin{aligned} \|\omega^H X_{k+1}\| &\le \left\| \omega^H \left(I - \frac{\mu}{P} X_k X_k^H \right) X_{k+1} \right\| \\ &\quad + \frac{\mu}{P} \|\omega^H X_k\| \left\| X_k^H X_{k+1} \right\| \\ &\le c + \frac{\mu}{P} Bc = c \left(1 + \frac{2}{P} \right). \end{aligned}$$

Choosing $l = 2$, we obtain

$$\left\| \omega^H \left(I - \frac{\mu}{P} X_k X_k^H \right) \left(I - \frac{\mu}{P} X_{k+1} X_{k+1}^H \right) X_{k+2} \right\| \le c.$$

Therefore

$$\begin{aligned} \|\omega^H X_{k+2}\| &\le \left\| \omega^H \left(I - \frac{\mu}{P} X_k X_k^H \right) \right. \\ &\quad \left. \cdot \left(I - \frac{\mu}{P} X_{k+1} X_{k+1}^H \right) X_{k+2} \right\| \\ &\quad + \frac{\mu}{P} \left\| \omega^H X_k X_k^H X_{k+2} \right\| \\ &\quad + \frac{\mu}{P} \left\| \omega^H X_{k+1} X_{k+1}^H X_{k+2} \right\| \\ &\quad + \frac{\mu^2}{P^2} \left\| \omega^H X_k X_k^H X_{k+1} X_{k+1}^H X_{k+2} \right\| \\ &\le O(c). \end{aligned}$$

Proceeding along similar lines, we obtain $\|\omega^H X_{k+l}\| \le Lc$ for $l = 0, \ldots, K$, where L is some constant. This implies $\omega^H \sum_{l=k}^{k+K} X_l X_l^H \omega \le (K+1) L^2 c^2$. Since c is arbitrary, we obtain that $\omega^H \sum_{l=k}^{k+K} X_l X_l^H \omega < \alpha_1$, which is a contradiction. ■

Lemma 3: Let X_k satisfy the persistence of excitation condition in Theorem 4, and let

$$\mathcal{P}_{k,k+D} = \begin{cases} \prod_{l=k}^{k+D} \left(I - \mu I_l X_l X_l^H \right), & \text{if } D \ge 0 \\ 1, & \text{if } D < 0 \end{cases}$$

where I_l is a random masking matrix chosen with equal probability from $\{\mathcal{I}_i, i = 1, \ldots, P\}$, and let

$$\Omega_k = \sum_{l=0}^{K} \overline{\mathcal{P}_{k,k+l-1}^H X_{k+l} X_{k+l}^H \mathcal{P}_{k,k+l-1}}$$

where K is as defined in Theorem 4, and $\overline{\{\cdot\}}$ is the average over randomly chosen I_l. Then, $\Omega_k - \gamma I$ is a positive definite matrix for some $\gamma > 0$ and $\forall k$.

Proof: The proof is by contradiction. Then, for some vector ω such that $\omega^H\omega = 1$, suppose that we do not have $\omega^H \Omega_k \omega \le c^2$, where c is any arbitrary positive number.

Then

$$\sum_{l=0}^{K} \omega^H \overline{\mathcal{P}_{k,k+l-1}^H X_{k+l} X_{k+l}^H \mathcal{P}_{k,k+l-1}} \, \omega \le c^2 \Rightarrow$$
$$\omega^H \overline{\mathcal{P}_{k,k+l-1}^H X_{k+l} X_{k+l}^H \mathcal{P}_{k,k+l-1}} \, \omega \le c^2 \quad \text{for } 0 \le l \le K.$$

Choosing $l = 0$, we obtain $\omega^H X_k X_k^H \omega \le c^2$ or $\|\omega^H X_k\| \le c$.

Choosing $l = 1$, we obtain

$$\omega^H \overline{\left(I - \mu X_k X_k^H I_k \right) X_{k+1} X_{k+1}^H \left(I - \mu I_k X_k X_k^H \right)} \, \omega \le c^2.$$

Therefore

$$\begin{aligned} &\omega^H X_{k+1} X_{k+1}^H \omega - \frac{\mu}{P} \omega^H X_k X_k^H X_{k+1} X_{k+1}^H \omega \\ &\quad - \frac{\mu}{P} \omega^H X_{k+1} X_{k+1}^H X_k X_k^H \omega + \frac{\mu^2}{P} \omega^H X_k X_k^H \\ &\quad \times \left[\sum_{i=0}^{P} \mathcal{I}_i X_{k+1} X_{k+1}^H \mathcal{I}_i \right] X_k X_k^H \omega \le c^2. \end{aligned}$$

Now

$$\begin{aligned} \left\| \omega^H X_k X_k^H X_{k+1} X_{k+1}^H \omega \right\| &\le \|\omega^H X_k\| \|X_k\| \\ &\quad \cdot \left\| X_{k+1}^H X_{k+1} \right\| \|\omega\| \\ &\le c B^{\frac{3}{2}} \end{aligned}$$

and

$$\left\| \omega^H X_k X_k^H \left[\sum_{i=0}^{P} \mathcal{I}_i X_{k+1} X_{k+1}^H \mathcal{I}_i \right] X_k X_k^H \omega \right\| \le c^2 P B^2.$$

Therefore, $\omega^H X_{k+1} X_{k+1}^H \omega = O(c)$, which implies $\|\omega^H X_{k+1}\| = O(c^{1/2})$. Proceeding along the same lines, we obtain $\|\omega^H X_{k+1}\| = O(c^{1/L})$ for $l = 0, \ldots, K$ for some constant L. This implies $\omega^H \sum_{l=k}^{k+K} X_l X_l^H \omega = O(c^{2/L})$. Since c is arbitrary, we obtain that $\omega^H \sum_{l=k}^{k+K} X_l X_l^H \omega < \alpha_1$, which is a contradiction. ■

Now, we are ready to prove Theorem 4.

Proof of Theorem 4: First, we will prove the convergence of $\overline{V}_k^H \overline{V}_k$. We have $\overline{V}_{k+1} = (I - (\mu/P) X_k X_k^H) \overline{V}_k$. Proceeding as before, we obtain the following update equation for $\overline{V}_k \overline{V}_k^H$

$$\begin{aligned} &\overline{V}_{k+K+1}^H \overline{V}_{k+K+1} \\ &\quad = \overline{V}_{k+K}^H \overline{V}_{k+K} - 2 \frac{\mu}{P} \overline{V}_{k+K}^H X_{k+K} X_{k+K}^H \overline{V}_{k+K} \\ &\qquad + \frac{\mu^2}{P^2} \overline{V}_{k+K}^H X_{k+K} X_{k+K}^H X_{k+K} X_{k+K}^H \overline{V}_{k+K} \\ &\quad \le \overline{V}_{k+K}^H \overline{V}_{k+K} - \frac{\mu}{P} \overline{V}_{k+K}^H X_{k+K} X_{k+K}^H \overline{V}_{k+K}. \end{aligned}$$

The last step follows from the fact that $\mu < 2/B$. Using the update equation for $\overline{V_k}$ repeatedly, we obtain

$$\overline{V}_{k+K+1}^H \overline{V}_{k+K+1} \le \overline{V}_k^H \overline{V}_k - \frac{\mu}{P} \overline{V}_k^H \mathcal{G}_k \overline{V}_k.$$

From Lemma 2, we have

$$\overline{V}_{k+K+1}^H \overline{V}_{k+K+1} \le \left(1 - \frac{\mu}{P} \eta \right) \overline{V}_k^H \overline{V}_k$$

which ensures exponential convergence of $\mathrm{tr}\{\overline{V}_k \overline{V}_k^H\}$.

Next, we prove the convergence of $\overline{V_k^H V_k}$. First, we have the following update equation for $\mathrm{tr}\{\overline{V_k V_k^H}\}$

$$\mathrm{tr}\left\{\overline{V_{k+K+1}V_{k+K+1}^H}\right\} \leq \mathrm{tr}\left\{\overline{V_{k+K}V_{k+K}^H}\right\} - \frac{\mu}{P}\mathrm{tr}\left\{X_{k+K}X_{k+K}^H\overline{V_{k+K}V_{k+K}^H}\right\}. \quad (22)$$

Using (22) and

$$\overline{V_{k+1}V_{k+1}^H} = \overline{\left(I - \mu I_k X_k X_k^H\right)\overline{V_k V_k^H}\left(I - \mu X_k X_k^H I_k\right)}$$

repeatedly, we obtain the following update equation:

$$\mathrm{tr}\left\{\overline{V_{k+K+1}V_{k+K+1}^H}\right\} \leq \mathrm{tr}\left\{\overline{V_k V_k^H}\right\} - \mathrm{tr}\left\{\Omega_k \overline{V_k V_k^H}\right\}.$$

From Lemma 3, we have

$$\mathrm{tr}\left\{\overline{V_{k+K+1}V_{k+K+1}^H}\right\} \leq (1 - \mu\gamma)\mathrm{tr}\left\{\overline{V_k V_k^H}\right\}$$

which ensures the exponential convergence of $\mathrm{tr}\{\overline{V_k V_k^H}\}$. ■

APPENDIX II
STABILITY ANALYSIS FOR MIXING SIGNALS

The results in this section are an extension of analysis in [16] to SPU-LMS with complex input signals. Notations are the same as those used in [16]. Let $\|A\| \stackrel{\text{def}}{=} \{\sum_{i,j}|a|_{ij}^2\}^{1/2} = \|A\|_F$ be the Frobenius norm of the matrix A. This is identical to the definition used in [2]. Note that in [16], $\|A\| \stackrel{\text{def}}{=} \{\lambda_{\max}(AA^H)\}^{1/2} = \|A\|_S$ is the spectral norm of A. Since for a $m \times n$ matrix A, $\|A\|_S \leq \|A\|_F \leq \max\{m,n\}\|A\|_S$, the results in [16] could also have been stated with the definition used here.

A process ϵ_k is said to be ϕ-mixing if there is a function $\phi(l)$ such that $\phi(l) \to 0$ as $l \to \infty$ and

$$\sup_{A \in \mathcal{M}_{-\infty}^k(X), B \in \mathcal{M}_{k+l}^{\infty}(\epsilon)} |P(B|A) - P(B)| \leq \phi(l)$$

$\forall m \geq 0$, $k \in (-\infty, \infty)$, where $\mathcal{M}_i^j(\epsilon)$, $-\infty \leq i \leq j \leq \infty$ is the σ algebra generated by $\{\epsilon_k\}$, $i \leq k \leq j$.

Let X_k be the input signal vector generated from the following process:

$$X_k = \sum_{j=-\infty}^{\infty} A(k,j)\epsilon_{k-j} + \psi_k \quad (23)$$

with $\sum_{j=-\infty}^{\infty}\sup_k \|A(k,j)\| < \infty$. $\{\psi_k\}$ is a d-dimensional deterministic process, and $\{\epsilon_k\}$ is a general m-dimensional ϕ-mixing sequence. The weighting matrices $A(k,j) \in \mathcal{R}^{d \times m}$ are assumed to be deterministic.

Define the index set $S = \{1, 2, \ldots, N\}$ and $\mathcal{I}_i$, as in Section III. Let I_j be a sequence of i.i.d $d \times d$ masking matrices chosen with equal probability from $\mathcal{I}_i$, $i = 1, \ldots, P$.

Then, we have the following theorem which is similar to Theorem 2 in [16].

Theorem 5: Let X_k be defined by (23) in Appendix III, where $\{\epsilon_k\}$ is a ϕ-mixing sequence such that it satisfies for any $n \geq 1$ and any increasing integer sequence $j_1 < j_2 < \ldots < j_n$

$$E\left[\exp\left(\beta \sum_{i=1}^n \|\epsilon_{j_i}\|^2\right)\right] \leq M\exp(Kn) \quad (24)$$

where β, M, and K are positive constants. Then, for any $p \geq 1$, there exist constants $\mu^* > 0$, $M > 0$, and $\alpha \in (0,1)$ such that for all $\mu \in (0, \mu^*]$ and for all $t \geq k \geq 0$

$$\left[E\left\|\prod_{j=k+1}^{t}\left(I - \mu I_j X_j X_j^H\right)\right\|^p\right]^{\frac{1}{p}} \leq M(1-\mu\alpha)^{t-k}$$

if and only if there exists an integer $h > 0$ and a constant $\delta > 0$ such that for all $k \geq 0$

$$\sum_{i=k+1}^{k+h} E\left[X_i X_i^H\right] \geq \delta I. \quad (25)$$

Proof: The proof is just a slightly modified version of the proof of Theorem 2 derived in [16, pp. 763–769, Sec. IV]. The modification takes into account that F_k in [16] is $F_k = X_k X_k^H$, whereas it is $F_k = I_k X_k X_k^H$ in the present context. ■

Note that [16, Th. 2] can be stated as a corollary to Theorem 5 by setting $I_j = I$ for all j. In addition, note that Condition (25) has the same form as Condition (18).

For Result 1, which is just a special case of Theorem 5, it is enough [16] to observe the following.

1) Gaussian X_k is obtained from (23) by choosing $A_k = 0$ and $\psi_k = 0$ for all k and ϵ_k to be Gaussian.
2) The Gaussian signal sequence as described in Result 1 is a phi-mixing sequence.
3) The Gaussian signals satisfy the condition in (24).
4) For stationary signals, $E[X_i X_i^H] = R_{xx}$ for all values of i, and, therefore, the following condition:

- There exists an integer $h > 0$ and a constant $\delta > 0$ such that for all $k \geq 0$

$$\sum_{i=k+1}^{k+h} E\left[X_i X_i^H\right] \geq \delta I \quad (26)$$

simply translates to R_{xx} being positive definite.

APPENDIX III
S-LMS STABILITY ANALYSIS FOR STATIONARY MIXING SIGNALS

The results in this section are an extension of analysis in [16] to S-LMS with stationary complex input signals. Notations are the same as those used in Appendix II. Let ϵ_k, X_k, ψ_k, and $A(k,j)$ be as defined in Appendix II.

In this section, we will place an additional restriction of stationarity on ϵ_k. Define the index set $S = \{1, 2, \ldots, N\}$ and $\mathcal{I}_i$ as in Section III. Then, Theorem 3 means that $F_i = \mathcal{I}_{i\%P+1}X_i X_i^H$ satisfies the following property of averaged exponential stability.

Lemma 4: Letting $F_i = \mathcal{I}_{i\%P+1}X_i X_i^H$, then F_i is averaged exponentially stable. That is, there exist constants $\mu^* > 0$, $M > 0$, and $\alpha > 0$ such that for all $\mu \in (0, \mu^*]$ and for all $t \geq k \geq 0$

$$\left\|\prod_{j=k+1}^{t}\left(I - \mu E[F_j]\right)\right\| = \left\|\prod_{j=k+1}^{t}\left(I - \mu\mathcal{I}_{j\%P+1}R\right)\right\| \leq M(1-\mu\alpha)^{t-k}.$$

Proof: From Theorem 3, we know that there exist $\mu^* > 0$, $M_0 > 0$, and $\gamma > 0$ such that for all t, $k > 0$

$$\left\| \prod_{j=k+1}^{k+tP} (I - \mu \mathcal{I}_{j\%P+1} R) \right\| \le M_0 (1-\mu\gamma)^t.$$

Note that

$$\|I - \mu \mathcal{I}_{j\%P+1} R\| \le \|I\| + \mu \|\mathcal{I}_{j\%P+1}\| \|R\| \le M'$$

for some $M' > 0$ and for all $\mu \in (0, \mu^*]$ and $j = 1, \ldots, P$. Letting $\lambda = (1-\mu\gamma)^{1/P}$, then

$$\left\| \prod_{j=k+1}^{k+tP+l} (I - \mu \mathcal{I}_{j\%P+1} R) \right\| \le \frac{M_0 \lambda^{tP+l} (M')^l}{\lambda^l}.$$

Noting that $0 < \mu\gamma < 1$, we have

$$\begin{aligned}(1-\mu\gamma)^{\frac{1}{P}} &= \left(1 - \frac{\mu}{P}\gamma + O\left((\mu\gamma)^2\right)\right)\\ &< (1-\mu\alpha)\end{aligned}$$

for some $\alpha > 0$. This leads to

$$\left\| \prod_{j=k+1}^{t} (I - \mu \mathcal{I}_{j\%P+1} R) \right\| \le M(1-\alpha\mu)^{t-k}$$

where $M = M_0 \max\{1, (M'/\lambda^*)^P\}$, and $\lambda^* = (1-\mu^*\gamma)^{1/P}$. ■

Using Lemma 4 and following the analysis of [16], we have the following theorem, which is similar to [16, Th. 2].

Theorem 6: Let X_k be defined by (23), where $\{\epsilon_k\}$ is a stationary ϕ-mixing sequence such that it satisfies, for any $n \ge 1$

$$E\left[\exp\left(\beta n \|\epsilon_k\|^2\right)\right] \le M \exp(Kn) \tag{27}$$

where β, M, and K are positive constants. Then, for any $p \ge 1$, there exist constants $\mu^* > 0$, $M > 0$, and $\alpha \in (0,1)$ such that for all $\mu \in (0, \mu^*]$ and for all $t \ge k \ge 0$

$$\left[E\left\| \prod_{j=k+1}^{t} \left(I - \mu \mathcal{I}_{j\%P+1} X_j X_j^H\right) \right\|^p\right]^{\frac{1}{p}} \le M(1-\alpha\mu)^{t-k}$$

if $R_{xx} = E[X_j X_j^H]$ is positive definite.

The corresponding result for LMS obtained from the extension of the analysis in [16] to complex signals is the following.

Result 4 (LMS Stability: Stationary Process): Let X_k be defined by (23), where $\{\epsilon_k\}$ is a stationary ϕ-mixing sequence such that it satisfies, for any $n \ge 1$

$$E\left[\exp\left(\beta n \|\epsilon_k\|^2\right)\right] \le M \exp(Kn) \tag{28}$$

where β, M, and K are positive constants. Then, for any $p \ge 1$, there exist constants $\mu^* > 0$, $M > 0$, and $\alpha \in (0,1)$ such that for all $\mu \in (0, \mu^*]$ and for all $t \ge k \ge 0$

$$\left[E\left\| \prod_{j=k+1}^{t} \left(I - \mu X_j X_j^H\right) \right\|^p\right]^{\frac{1}{p}} \le M(1-\alpha\mu)^{t-k}$$

if and only if $R_{xx} = E[X_j X_j^H]$ is positive definite.

Therefore, exponential stability of LMS implies exponential stability of S-LMS.

The application of Theorem 6 to X_k obtained from a time-series signal is illustrated below.

Result 5 (Stationary Gaussian Process): Let x_k be a stationary Gaussian random process such that $E[x_k x_{k-l}] = r_l \to 0$ as $l \to \infty$, and $X_k = [x_k\ x_{k-1}\ \ldots\ x_{k-n+1}]$; then, for any $p \ge 1$, there exist constants $\mu^* > 0$, $\alpha \in (0,1)$, and $M > 0$ such that for all $\mu \in (0, \mu^*]$ and for all $t \ge k \ge 0$

$$\left[E\left\| \prod_{j=k+1}^{t} \left(I - \mu \mathcal{I}_{j\%P+1} X_j X_j^H\right) \right\|^p\right]^{\frac{1}{p}} \le M(1-\alpha\mu)^{t-k}$$

if the input correlation matrix $E[X_k X_k^H] = R_{xx}$ is positive definite.

APPENDIX IV
PERFORMANCE ANALYSIS FOR MIXING SIGNALS

The results in this section are an extension of analysis in [2] to SPU-LMS with complex signals. The results enable us to predict the steady-state behavior of SPU-LMS without the standard uncorrelated input and coefficient vectors assumption employed in Section III-A. Moreoever, the two lemmas in this section state that the error terms for LMS and SPU-LMS are bounded above by the same constants. These results are very useful for comparison of steady-state errors of SPU-LMS and LMS in the sense that the error terms are of the same magnitude. A couple of examples using the analysis in this section were presented in Section III-C as Results 2 (see details in Appendix IV-A) and 3 (see details in Appendix IV-B), where the performance of SPU-LMS and LMS was compared for two different scenarios.

We begin the mean square error analysis by assuming that

$$d_k = X_k^H W_{\text{opt}} + n_k.$$

Then, we can write the evolution equation for the tracking error $V_k \stackrel{\text{def}}{=} W_k - W_{\text{opt}}$ as

$$V_{k+1} = \left(I - \mu P_k X_k X_k^H\right) V_k + \mu X_k n_k$$

where $P_k = I$ for LMS and $P_k = I_k$ for SPU-LMS.

In general, V_k obeys the following inhomogeneous equation:

$$\delta_{k+1} = (I - \mu F_k)\delta_k + \xi_k, \quad \delta_0 = 0$$

where δ_k can be represent by a set of recursive equations as follows:

$$\delta_k = J_k^{(0)} + J_k^{(1)} + \ldots + J_k^{(n)} + H_k^{(n)}$$

where the processes $J_k^{(r)}$, $0 \le r < n$, and $H_k^{(n)}$ are described by

$$\begin{aligned}J_{k+1}^{(0)} &= (I - \mu \bar{F}_k) J_k^{(0)} + \xi_k\\ J_0^{(0)} &= 0\\ J_{k+1}^{(r)} &= (I - \mu \bar{F}_k) J_k^{(r)} + \mu Z_k J_k^{(r-1)}\\ J_k^{(r)} &= 0, \quad 0 \le k < r\\ H_{k+1}^{(n)} &= (I - \mu F_k) H_k^{(n)} + \mu Z_k J_k^{(n)}\\ H_k^{(n)} &= 0, \quad 0 \le k < n\end{aligned}$$

where $Z_k = F_k - \bar{F}_k$, and $\bar{F}_k$ is an appropriate deterministic process, which is usually chosen as $\bar{F}_k = E[F_k]$. In [2], under appropriate conditions, it was shown that there exists some constant $C < \infty$ and $\mu_0 > 0$ such that for all $0 < \mu \leq \mu_0$, we have

$$\sup_{k\geq 0}\left\|H_k^{(n)}\right\|_p \leq C\mu^{\frac{n}{2}}.$$

Now, we modify the definition of weak dependence as given in [2] for circular complex random variables. The theory developed in [2] can be easily adapted for circular random variables using this definition. Let $q \geq 1$ and $X = \{X_n\}_{n\geq 0}$ be a $(l \times 1)$ matrix-valued process. Let $\beta = (\beta(r))_{r\in N}$ be a sequence of positive numbers decreasing to zero at infinity. The complex process $X = \{X_n\}_{n\geq 0}$ is said to be (δ, q)-weak dependent if there exist finite constants $C = \{C_1, \ldots, C_q\}$, such that for any $1 \leq m < s \leq q$ and m-tuple $k_1, \ldots, k_m$ and any $(s-m)$-tuple $k_{m+1}, \ldots, k_s$, with $k_1 \leq \ldots \leq k_m < k_m + r \leq k_{m+1} \leq \ldots \leq k_s$, it holds that

$$\sup\left|\text{cov}\left(f_{k_1,i_1}\left(\tilde{X}_{k_1,i_1}\right)\cdot\ldots\cdot f_{k_m,i_m}\left(\tilde{X}_{k_m,i_m}\right),\right.\right.$$
$$\left.\left. f_{k_{m+1},i_{m+1}}\left(\tilde{X}_{k_{m+1},i_{m+1}}\right)\cdot\ldots\cdot f_{k_s,i_s}\left(\tilde{X}_{k_s,i_s}\right)\right)\right| \leq C_s\beta(r)$$

where the supremum is taken over the set $\{1 \leq i_1, \ldots, i_s \leq l, f_{k_1,i_1}, f_{k_2,i_2} \ldots f_{k_m,i_m}\}$, and $\tilde{X}_{n,i}$ denotes the ith component of $X_n - E(X_n)$. The set of functions $f_{n,i}()$ over which the sup is being taken are given by $f_{n,i}(\tilde{X}_{n,i}) = \tilde{X}_{n,i}$ and $f_{n,i}(\tilde{X}_{n,i}) = \tilde{X}_{n,i}^*$.

Define $\mathcal{N}(p)$ from [2] as follows:

$$\mathcal{N}(p) = \left\{\epsilon : \left\|\sum_{k=s}^{t} D_k\epsilon_k\right\|_p \leq \rho_p(\epsilon)\left(\sum_{k=s}^{t}|D_k|^2\right)^{\frac{1}{2}} \right.$$
$$\forall 0 \leq s \leq t \text{ and } \forall D = \{D_k\}_{k\in N}$$
$$\left.(q \times l) \text{ deterministic matrices}\right\}$$

where $\rho_p(\epsilon)$ is a constant, depending only on the process ϵ and the number p.

F_k can be written as $F_k = P_k X_k X_k^H$, where $P_k = I$ for LMS and $P_k = I_k$ for SPU-LMS. It is assumed that the following hold true for F_k. For some $r, q \in N, \mu_0 > 0$, and $0 < \alpha < 1/\mu_0$

- **F1** (r, α, μ_0): $\{F_k\}_{k\geq 0}$ is is L_r-exponentially stable. That is

$$\left[E\left\|\prod_{j=k+1}^{t}(I - \mu F_j)\right\|^r\right]^{\frac{1}{r}} \leq M(1-\mu\alpha)^{t-k}.$$

- **F2** (α, μ_0): $\{F_k\}_{k\geq 0}$ is averaged exponentially stable. That is

$$\left\|\prod_{j=k+1}^{t}(I - \mu E[F_j])\right\| \leq M(1-\mu\alpha)^{t-k}.$$

Conditions **F3** and **F4** stated below are trivially satisfied for $P_k = I$ and $P_k = I_k$.

- **F3** (q, μ_0): $\sup_{k\in N}\sup_{\mu\in(0,\mu_0]}\|P_k\|_q < \infty$, and $\sup_{k\in N}\sup_{\mu\in(0,\mu_0]}|E[P_k]| < \infty$.
- **F4** (q, μ_0): $\sup_{k\in N}\sup_{\mu\in(0,\mu_0]}\mu^{-1/2}\|P_k - E[P_k]\|_q < \infty$.

The excitation sequence $\xi = \{\xi_k\|_{k\geq 0}$ [2] is assumed to be decomposed as $\xi_k = M_k\epsilon_k$, where the process $M = \{M_k\}_{k\geq 0}$ is a $d \times l$ matrix-valued process, and $\epsilon = \{\epsilon_k\}_{k\geq 0}$ is a $(l \times 1)$ vector-valued process that verifies the following assumptions:

- **EXC1**: $\{M_k\}_{k\in Z}$ is $\mathcal{M}_0^k(X)$-adapted,[3] and $\mathcal{M}_0^k(\epsilon)$ and $\mathcal{M}_0^k(X)$ are independent.
- **EXC2**(r, μ_0): $\sup_{\mu\in(0,\mu_0]}\sup_{k\geq 0}\|M_k\|_r < \infty$, $(r > 0, \mu_0 > 0)$.
- **EXC3**(p, μ_0): $\epsilon = \{\epsilon_k\}_{k\in N}$ belongs to $\mathcal{N}(p)$, $(p > 0, \mu_0 > 0)$.

The following theorems from [2] are relevant.

Theorem 7 ([2, Th.]): Let $n \in N$, and let $q \geq p \geq 2$. Assume **EXC1**, **EXC2**$(pq/(q-p), \mu_0)$, and **EXC3**(p, μ_0). For $a, b, \alpha > 0$, $a^{-1} + b^{-1} = 1$, and some $\mu_0 > 0$, assume in addition that **F2**(α, μ_0), **F4**(aqn, μ_0), and

- $\{G_k\}_{k\geq 0}$ is $(\beta, (q+2)n)$ weakly dependent and $\sum(r+1)^{((q+2)n/2)-1}\beta(r) < \infty$;
- $\sup_{k\geq 0}\|G_k\|_{bqn} < \infty$.

Then, there exists a constant $K < \infty$ (depending on $\beta(k)$, $k \geq 0$ and on the numerical constants p, q, n, q, b, μ_0, and α but not otherwise on $\{X_k\}$, $\{\epsilon_k\}$ or on μ), such that for all $0 < \mu \leq \mu_0$, for all $0 \leq r \leq n$

$$\sup_{s\geq 1}\left\|J_s^{(r)}\right\|_p \leq K\rho_p(\epsilon)\sup_{k\geq 0}\|M_k\|_{\frac{pq}{(q-p)}}\mu^{\frac{(r-1)}{2}}.$$

Theorem 8 ([2, Th. 2]): Let $p \geq 2$, and let $a, b, c > 0$ such that $1/a + 1/b + 1/c = 1/p$. Let $n \in N$. Assume **F1**(a, α, μ_0), and

- $\sup_{s\geq 0}\|Z_s\|_b < \infty$;
- $\sup_{s\geq 0}\|J_s^{(n+1)}\|_c < \infty$.

Then, there exists a constant $K' < \infty$ (depending on the numerical constants a, b, c, α, μ_0, and n but not on the process $\{\epsilon_k\}$ or on the stepsize parameter μ) such that for all $0 < \mu \leq \mu_0$

$$\sup_{s\geq 0}\left\|H_s^{(n)}\right\|_p \leq K'\sup_{s\geq 0}\left\|J_s^{(n+1)}\right\|_c.$$

We next show that if LMS satisfies the assumptions above (assumptions in [2, Sec. 3.2]), then so does SPU-LMS. Conditions **F1** and **F2** follow directly from Theorem 5. It is easy to see that **F3** and **F4** hold easily for LMS and SPU-LMS.

Lemma 5: The constant K in Theorem 7 calculated for LMS can also be used for SPU-LMS.

Proof: Here, all that is needed to be shown is that if LMS satisfies the conditions (**EXC1**), (**EXC2**), and (**EXC3**), then so does SPU-LMS. Moreover, the upper bounds on the norms for LMS are also upper bounds for SPU-LMS. That easily follows because $M_k^{LMS} = X_k$, whereas $M_k^{SPU-LMS} = I_kX_k$ and $\|I_k\| \leq 1$ for any norm $\|\cdot\|$. ∎

Lemma 6: The constant K' in Theorem 8 calculated for LMS can also be used for SPU-LMS.

[3]A sequence of random variables X_i is called adapted with respect to a sequence of σ-fields $\mathcal{F}_i$ if X_i is $\mathcal{F}_i$ measurable [6].

Proof: First, we show that if for LMS $\sup_{s\geq 0}\|Z_s\|_b < \infty$, then so it is for SPU-LMS. First, note that for LMS, we can write $Z_s^{LMS} = X_sX_s^H - E[X_sX_s^H]$, whereas for SPU-LMS

$$\begin{aligned} Z_s^{SPU-LMS} &= I_sX_sX_s^H - \frac{1}{P}E\left[X_sX_s^H\right] \\ &= I_sX_sX_s^H - I_sE\left[X_sX_s^H\right] \\ &\quad + \left(I_s - \frac{1}{P}I\right)E\left[X_sX_s^H\right]. \end{aligned}$$

This means that $\|Z_s^{SPU-LMS}\|_b \leq \|I_s\|_b\|Z_s^{LMS}\|_b + \|I_s-(1/P)I\|_b\|E[X_sX_s^H]\|_b$. Therefore, since $\sup_{s\geq 0}\|_bE[X_sX_s^H]\|_b < \infty$ and $\sup_{s\geq 0}\|Z_s^{LMS}\|_b < \infty$, we have

$$\sup_s \left\|Z_s^{SPU-LMS}\right\|_b < \infty.$$

Since all conditions for Theorem 2 have been satisfied by SPU-LMS in a similar manner, the constant obtained is also the same. ■

A. I.I.D Gaussian Input Sequence

In this section, we assume that $X_k = [x_k\ x_{k-1}\ \ldots\ x_{k-N+1}]^T$, where N is the length of the vector X_k. $\{x_k\}$ is a sequence of zero mean i.i.d Gaussian random variables. We assume that $w_k = 0$ for all $k \geq 0$. In that case

$$\begin{aligned} V_{k+1} &= \left(I - \mu P_kX_kX_k^H\right)V_k + \mu X_kn_k \quad \text{with} \\ V_0 &= -W_{\text{opt},0} = W_{\text{opt}} \end{aligned}$$

where for LMS, we have $P_k = I$ and $P_k = I_k$ in case of SPU-LMS. We assume n_k is a white i.i.d. Gaussian noise with variance σ_v^2. We see that since the conditions (24) and (25) are satisfied for Theorem 5, both LMS and SPU-LMS are exponentially stable. In fact, both have the same α exponent of decay. Therefore, conditions **F1** and **F2** are satisfied.

We rewrite $V_k = J_k^{(0)} + J_k^{(1)} + J_k^{(2)} + H_k^{(2)}$. Choosing $\bar{F}_k = E[F_k]$, we have $E[P_kX_kX_k^H] = \sigma^2 I$ in the case of LMS and $(1/P)\sigma^2 I$ in the case of SPU-LMS. By Theorems 7 and 8 and Lemmas 5 and 6, we can upperbound both $|J_k^{(2)}|$ and $|H_k^{(2)}|$ by exactly the same constants for LMS and SPU-LMS. In particular, there exists some constant $C < \infty$ such that for all $\mu \in (0, \mu_0]$, we have

$$\begin{aligned} \sup_{t\geq 0}\left|E\left[J_t^{(1)}\left(J_t^{(2)} + H_t^{(2)}\right)^H\right]\right| &\leq C\|X_0\|_{\frac{r(r+\delta)}{\delta}}\rho_r^2(v)\mu^{\frac{1}{2}} \\ \sup_{t\geq 0}\left|E\left[J_t^{(0)}H_t^{(2)}\right]\right| &\leq C\rho_r(v)\|X_0\|_{\frac{r(r+\delta)}{\delta}}\mu^{\frac{1}{2}}. \end{aligned}$$

Next, for LMS we concentrate on

$$\begin{aligned} J_{k+1}^{(0)} &= (1-\mu\sigma^2)J_k^{(0)} + \mu X_kn_k \\ J_{k+1}^{(1)} &= (1-\mu\sigma^2)J_k^{(1)} + \mu\left(\sigma^2 I - X_kX_k^H\right)J_k^{(0)} \end{aligned}$$

and for SPU-LMS we concentrate on

$$\begin{aligned} J_{k+1}^{(0)} &= \left(1 - \frac{\mu}{P}\sigma^2\right)J_k^{(0)} + \mu I_kX_kn_k \\ J_{k+1}^{(1)} &= \left(1 - \frac{\mu}{P}\sigma^2\right)J_k^{(1)} + \mu\left(\frac{\sigma^2}{P}I - I_kX_kX_k^H\right)J_k^{(0)}. \end{aligned}$$

After tedious but straightforward calculations (following the procedure in [2]), we obtain for LMS

$$\begin{aligned} \lim_{k\to\infty} E\left[J_k^{(0)}\left(J_k^{(0)}\right)^H\right] &= \mu^2\left[\frac{\sigma_v^2}{\mu(2-\mu\sigma^2)}I\right] \\ \lim_{k\to\infty} E\left[J_k^{(0)}\left(J_k^{(1)}\right)^H\right] &= 0 \\ \lim_{k\to\infty} E\left[J_k^{(0)}\left(J_k^{(2)}\right)^H\right] &= 0 \\ \lim_{k\to\infty} E\left[J_k^{(1)}\left(J_k^{(1)}\right)^H\right] &= \mu^2\left[\frac{N\sigma^2\sigma_v^2}{(2-\mu\sigma^2)^2}I\right] \\ &= \mu^2\left[\frac{N\sigma^2\sigma_v^2}{4}I + O(\mu)I\right] \end{aligned}$$

which yields $\lim_{k\to\infty} E[V_kV_k^H] = \mu^2[(\sigma_v^2/2\mu)I + (N\sigma^2\sigma_v^2/4)I + O(\mu^{1/2})I]$ and for SPU-LMS, we obtain

$$\begin{aligned} \lim_{k\to\infty} E\left[J_k^{(0)}\left(J_k^{(0)}\right)^H\right] &= \mu^2\left[\frac{\sigma_v^2}{\mu\left(2-\frac{\mu}{P}\sigma^2\right)}I\right] \\ \lim_{k\to\infty} E\left[J_k^{(0)}\left(J_k^{(1)}\right)^H\right] &= 0 \\ \lim_{k\to\infty} E\left[J_k^{(0)}\left(J_k^{(2)}\right)^H\right] &= 0 \\ \lim_{k\to\infty} E\left[J_k^{(1)}\left(J_k^{(1)}\right)^H\right] &= \mu^2\left[\frac{\frac{(N+1)P-1}{P}\sigma^2\sigma_v^2}{\left(2-\frac{\mu}{P}\sigma^2\right)^2}I\right] \\ &= \mu^2\left[\frac{\frac{(N+1)P-1}{P}\sigma^2\sigma_v^2}{4}I + O(\mu)I\right] \end{aligned}$$

which yields $\lim_{k\to\infty} E[V_kV_k^H] = \mu^2[(\sigma_v^2/2\mu)I + ((((N+1)P-1)/P)\sigma^2\sigma_v^2/4)I + O(\mu^{1/2})I]$.

B. Temporally Correlated Spatially Uncorrelated Array Output

In this section, we consider X_k given by

$$X_k = \kappa X_{k-1} + \sqrt{1-\kappa^2}U_k$$

where U_k is a vector of circular Gaussian random variables with unit variance. Similar to Appendix IV-A, we rewrite $V_k = J_k^{(0)} + J_k^{(1)} + J_k^{(2)} + H_k^{(2)}$. Since we have chosen $\bar{F}_k = E[F_k]$, we have $E[P_kX_kX_k^H] = I$ in the case of LMS and $(1/P)I$ in the case of SPU-LMS. Again, conditions **F1** and **F2** are satisfied because of Theorem 5. By [2] and Lemmas 1 and 2, we can upperbound both $J_k^{(2)}$ and $H_k^{(2)}$ by exactly the same constants for LMS and SPU-LMS. By Theorems 7 and 8 and Lemmas 5 and 6, we have that there exists some constant $C < \infty$ such that for all $\mu \in (0, \mu_0]$, we have

$$\begin{aligned} \sup_{t\geq 0}\left|E\left[J_t^{(1)}\left(J_t^{(2)} + H_t^{(2)}\right)^H\right]\right| &\leq C\|X_0\|_{\frac{r(r+\delta)}{\delta}}\rho_r^2(v)\mu^{\frac{1}{2}} \\ \sup_{t\geq 0}\left|E\left[J_t^{(0)}H_t^{(2)}\right]\right| &\leq C\rho_r(v)\|X_0\|_{\frac{r(r+\delta)}{\delta}}\mu^{\frac{1}{2}}. \end{aligned}$$

Next, for LMS, we concentrate on

$$\begin{aligned} J_{k+1}^{(0)} &= (1-\mu)J_k^{(0)} + \mu X_kn_k \\ J_{k+1}^{(1)} &= (1-\mu)J_k^{(1)} + \mu\left(I - X_kX_k^H\right)J_k^{(0)} \end{aligned}$$

and for SPU-LMS, we concentrate on

$$J_{k+1}^{(0)} = \left(1 - \frac{\mu}{P}\right) J_k^{(0)} + \mu I_k X_k n_k$$
$$J_{k+1}^{(1)} = \left(1 - \frac{\mu}{P}\right) J_k^{(1)} + \mu \left(\frac{1}{P} I - I_k X_k X_k^H\right) J_k^{(0)}.$$

After tedious but straighforward calculations (following the procedure in [2]), we obtain for LMS

$$\lim_{k\to\infty} E\left[J_k^{(0)} \left(J_k^{(0)}\right)^H\right] = \mu^2 \left[\frac{\sigma_v^2}{\mu(2-\mu)} I\right]$$
$$\lim_{k\to\infty} E\left[J_k^{(0)} \left(J_k^{(1)}\right)^H\right] = -\mu^2 \left[\frac{\kappa^2 \sigma_v^2 N}{2(1-\kappa^2)} I + O(\mu) I\right]$$
$$\lim_{k\to\infty} E\left[J_k^{(0)} \left(J_k^{(2)}\right)^H\right] = \mu^2 \left[\frac{\kappa^2 \sigma_v^2 N}{4(1-\kappa^2)} I + O(\mu) I\right]$$
$$\lim_{k\to\infty} E\left[J_k^{(1)} \left(J_k^{(1)}\right)^H\right] = \mu^2 \left[\frac{(1+\kappa^2) \sigma_v^2 N}{4(1-\kappa^2)} I + O(\mu) I\right]$$

which leads to $\lim_{k\to\infty} E[V_k V_k^H] = \mu^2[(\sigma_v^2/2\mu)I + (N\sigma_v^2/4)I + O(\mu^{1/2})I]$, and for SPU-LMS, we obtain

$$\lim_{k\to\infty} E\left[J_k^{(0)} \left(J_k^{(0)}\right)^H\right] = \mu^2 \left[\frac{\sigma_v^2}{\mu\left(2-\frac{\mu}{P}\right)} I\right]$$
$$\lim_{k\to\infty} E\left[J_k^{(0)} \left(J_k^{(1)}\right)^H\right] = -\mu^2 \left[\frac{\kappa^2 \sigma_v^2 N}{2(1-\kappa^2)P} I + O(\mu) I\right]$$
$$\lim_{k\to\infty} E\left[J_k^{(0)} \left(J_k^{(2)}\right)^H\right] = \mu^2 \left[\frac{\kappa^2 \sigma_v^2 N}{4(1-\kappa^2)P} I + O(\mu) I\right]$$
$$\lim_{k\to\infty} E\left[J_k^{(1)} \left(J_k^{(1)}\right)^H\right] = \mu^2 \left[\frac{\sigma_v^2}{4}\left[\frac{N}{P}\frac{1+\kappa^2}{1-\kappa^2} + (N+1)\frac{P-1}{P}\right] I + O(\mu) I\right]$$

which leads to $\lim_{k\to\infty} E[V_k V_k^H] = \mu^2[(\sigma_v^2/2\mu)I + (\sigma^2/4)[N+1-(1/P)]I + O(\mu^{1/2})I]$.

REFERENCES

[1] T. Aboulnasr and K. Mayyas, "Selective coefficient update of gradient-based adaptive algorithms," *Signal Process.*, vol. 47, no. 5, pp. 1421–1424, May 1999.

[2] R. Aguech, E. Moulines, and P. Priouret, "On a perturbation approach for the analysis of stochastic tracking algorithms," *SIAM J. Control Optimization*, vol. 39, no. 3, pp. 872–899, 2000.

[3] T. S. Alexander, *Adaptive Signal Processing: Theory and Applications*. New York: Springer-Verlag, 1986.

[4] J. Benesty and P. Duhamel, "A fast exact least mean square adaptive algorithm," *Signal Process.*, vol. 40, no. 12, pp. 2904–2920, Dec. 1992.

[5] K. Berbidis and S. Theodoridis, "A new fast block adaptive algorithm," *Signal Process.*, vol. 47, no. 1, pp. 75–87, Jan. 1999.

[6] P. Billingsley, *Probability and Measure*, ser. Probability and Mathematical Statistics. New York: Wiley, 1995.

[7] M. C. Campi, "Performance of RLS identification algorithms with forgetting factor: A ϕ-mixing approach," *J. Math. Syst., Estim. Control*, vol. 7, pp. 29–53, 1997.

[8] P. S. R. Diniz and S. Werner, "Partial-update NLMS algorithms with data-selective updating," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 938–949, Apr. 2004.

[9] S. C. Douglas, "Analysis and implementation of the Max-NLMS adaptive filter," in *Proc. ASIMOLAR Conf.*, vol. 1, 1996, pp. 6591–6663.

[10] ——, "A family of normalized LMS algorithms," *IEEE Signal Process. Lett.*, vol. 1, no. 3, pp. 49–51, Mar. 1994.

[11] ——, "Adaptive filters employing partial updates," *IEEE Trans. Circuits Syst.*, vol. CAS-II, pp. 209–216, Mar. 1997.

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

[13] M. J. Gingell, B. G. Hay, and L. D. Humphrey, "A block mode update echo canceller using custom LSI," in *Proc. GLOBECOM Conf.*, vol. 3, Nov. 1983, pp. 1394–1397.

[14] M. Godavarti, "Implementation of a G.165 line echo canceller on Texas Instruments' TMS320C3x and TMS320C54x chips," in *Proc. ICSPAT Conf.*, Sep. 1997, pp. 65–69.

[15] M. Godavarti and A. O. Hero III, "Stability analysis of the sequential partial update LMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6, May 2001, pp. 3857–3860.

[16] L. Guo, L. Ljung, and G. J. Wang, "Necessary and sufficient conditions for stability of LMS," *IEEE Trans. Autom. Control*, vol. 42, no. 6, pp. 761–770, Jun. 1997.

[17] D. R. Halverson and G. L. Wise, "Memoryless discrete-time detection of phi-mixing signals in phi-mixing noise," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 415–417, Mar. 1984.

[18] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[20] 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, Jun. 1981.

[21] E. J. Kelly and K. M. Forsythe, "Adaptive Detection and Parameter Estimation for Multidimensional Signal Models," Lincoln Lab., Mass. Inst. Technol., Tech. Rep. 848, April 1989.

[22] D. Messerschmitt, D. Hedberg, C. Cole, A. Haoui, and P. Winship, Digital Voice Echo Canceller With a TMS32020, Application Report: SPRA129, Texas Instruments, Dallas, TX.

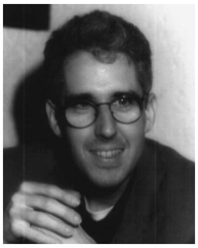
[23] G. Moustakides and J. Thomas, "Min-max detection of weak signals in phi-mixing noise," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 5, pp. 529–537, May 1984.

[24] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1991.

[25] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

Mahesh Godavarti (M'01) was born in Jaipur, India in 1972. He received the B.Tech degree in electrical and electronics engineering from the Indian Institute of Technology, Madras, India, in 1993 and the M.S. degree in electrical and computer engineering from the University of Arizona, Tucson, in 1995. He was at the University of Michigan, Ann Arbor, from 1997 to 2001, from which he received the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering.

Currently, he is a Staff DSP Engineer with Ditech Communications, Mountain View, CA, where he is researching new algorithms for speech enhancement. His research interests include topics in speech and signal processing, communications, and information theory.



Alfred O. Hero, III (F'98) was born in Boston, MA, in 1955. He received the B.S. degree (summa cum laude) from Boston University in 1980 and the Ph.D. from Princeton University, Princeton, NJ, in 1984, both in electrical engineering.

While at Princeton, he held the G.V.N. Lothrop Fellowship in Engineering. Since 1984, he has been a Professor with the University of Michigan, Ann Arbor, where he has appointments in the Department of Electrical Engineering and Computer Science, the Department of Biomedical Engineering, and the Department of Statistics. He has held visiting positions at I3S University of Nice, Sophia-Antipolis, France, in 2001; Ecole Normale Supérieure de Lyon, Lyon, France, in 1999; Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1999; Scientific Research Labs of the Ford Motor Company, Dearborn, MI, in 1993; Ecole Nationale Superieure des Techniques Avancees (ENSTA) and Ecole Superieure d'Electricite, Paris, in 1990; and Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, from 1987 to 1989. His research has been supported by NIH, NSF, AFOSR, NSA, ARO, ONR, DARPA, and by private industry in the areas of estimation and detection, statistical communications, bioinformatics, signal processing, and image processing.

Dr. Hero served as Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 1995 to 1998, and again in 1999, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2002. He was Chairman of the Statistical Signal and Array Processing (SSAP) Technical Committee from 1997 to 1998 and Treasurer of the Conference Board of the IEEE Signal Processing Society. He was Chairman for Publicity of the 1986 IEEE International Symposium on Information Theory (Ann Arbor, MI) and General Chairman of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing (Detroit, MI). He was co-chair of the 1999 IEEE Information Theory Workshop on Detection, Estimation, Classification, and Filtering (Santa Fe, NM) and the 1999 IEEE Workshop on Higher Order Statistics (Caesaria, Israel). He chaired the 2002 NSF Workshop on Challenges in Pattern Recognition. He co-chaired the 2002 Workshop on Genomic Signal Processing and Statistics (GENSIPS). He was Vice President (Finance) of the IEEE Signal Processing Society from 1999 to 2002. He was Chair of Commission C (Signals and Systems) of the U.S. National Commission of the International Union of Radio Science (URSI) from 1999 to 2002. He has been a member of the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society since 1999. He is also a member of the SEDD Review Panel of the U.S. National Research Council. He is currently President-Elect of the IEEE Signal Processing Society. He is a member of Tau Beta Pi, the American Statistical Association (ASA), the Society for Industrial and Applied Mathematics (SIAM), and the U.S. National Commission (Commission C) of URSI. He received the 1998 IEEE Signal Processing Society Meritorious Service Award, the 1998 IEEE Signal Processing Society Best Paper Award, and the IEEE Third Millennium Medal. In 2002, he was appointed IEEE Signal Processing Society Distinguished Lecturer.