The Stability of Variable Step-Size LMS Algorithms

Saul B. Gelfand, Member, IEEE, Yongbin Wei, Student Member, IEEE, and James V. Krogmeier, Member, IEEE

Abstract—Variable step-size LMS (VSLMS) algorithms are a popular approach to adaptive filtering, which can provide improved performance while maintaining the simplicity and robustness of conventional fixed step-size LMS. Here, we examine the stability of VSLMS with uncorrelated stationary Gaussian data. Most VSLMS described in the literature use a data-dependent step-size, where the step-size either depends on the data before the current time (prior step-size rule) or through the current time (posterior step-size rule). It has often been assumed that VSLMS algorithms are stable (in the sense of mean-square bounded weights), provided that the step-size is constrained to lie within the corresponding stability region for the LMS algorithm. For a single tap filter, we find exact expressions for the stability region of VSLMS over the classes of prior and posterior step-sizes and show that the stability region for prior step-size coincides with that of fixed step-size, but the region for posterior step-size is strictly smaller than for fixed step-size. For the multiple tap case, we obtain bounds on the stability regions with similar properties.

The approach taken here is a generalization of the classical method of analyzing the exponential stability of the weight covariance equation for LMS. Although it is not possible to derive a weight covariance equation for general data-dependent VSLMS, the weight variances can be upper bounded by the solution of a linear time-invariant difference equation, after appropriately dealing with certain nonlinear terms. For prior step-size (like fixed step-size), the state matrix is symmetric, whereas for posterior step-size, the symmetry is lost, requiring a more detailed analysis. The results are verified by computer simulations.

Index Terms—Adaptive filtering, stability analysis, variable step-size.

I. INTRODUCTION

THERE HAS BEEN a lot of interest recently in variable step-size least mean square (VSLMS) algorithms [1], [3], [6], [8], [9], [11]. The idea is to adjust the step-size in a data dependent manner to improve the transient behavior. This translates into improved constant parameter learning as well as improved nonstationary parameter tracking, provided that the parameter variations do not lead to asymptotically optimal constant step-size rules for LMS (in the same way that certain nonstationary plant models such as random walk lead to asymptotically optimal constant gain matrices for the Kalman filter). Generally, the approach with VSLMS is to devise step-size rules that give large steps when the estimated error is large and small steps when the error is small, thereby avoiding the tradeoff between convergence rate and misadjustment for fixed step-size LMS. Furthermore, VSLMS can be designed to have comparable complexity and robustness (e.g., no divisions are required) to LMS, which has been difficult to achieve with fast least squares algorithms.

The weight updates in many of the VSLMS algorithms described in the literature take the form

\[ w(k+1) = w(k) + \alpha_k e_k x(k), \quad k = 0, 1, \ldots \]

where \( \alpha_k \) is the variable step-size, \( x(k) \) is the data vector, and \( e_k \) is the error signal. The step-size \( \alpha_k \) in the literature is typically computed in one of two ways: i) \( \alpha_k \) is a prior step-size that depends on \( x(i), e_i \) for \( i < k \), i.e., on the past data and errors; and ii) \( \alpha_k \) is a posterior step-size that depends on \( x(i), e_i \) for \( i \leq k \), i.e., on the current and past data and errors. Furthermore, a bound is enforced on \( \alpha_k \) to ensure stability (presumably): \( \alpha \leq \alpha_k \leq \bar{\alpha} \).

The analysis of these VSLMS algorithms in the literature typically proceeds in two steps [9], [11]. First, a rigorous stability analysis is attempted, apparently leading to conditions for mean-square (MS) bounded weights and bounded mean-square error (MSE), and second, an approximate performance analysis is carried out, including convergence to and characterization of the asymptotic weight mean, covariance, and MSE. That one can at least guarantee stability (MS bounded weights) rigorously would seem to support the performance analysis. Now, for stationary uncorrelated Gaussian data and fixed step-size \( \alpha \), it is known that a sufficient condition for MS bounded weights is that [4]

\[ 0 < \alpha < \frac{2}{3 \text{tr} R} \]

where \( R = E\{x(k)x(k)^T\} \) (we deal with the real data case here for simplicity). For the same data model but now with variable step-size \( \alpha_k \), it is tacitly assumed that a sufficient condition for MS bounded weights is that

\[ 0 < \alpha \leq \alpha_k \leq \bar{\alpha} < \frac{2}{3 \text{tr} R}. \]  

(1)

This assumption is straightforwardly correct if \( \alpha_k \) is deterministic, but the usual case is where the \( \alpha_k \) is data dependent. One wonders, for example, if the gap between \( \alpha \) and \( \bar{\alpha} \) plays a role in stability for such random step sizes.

In this paper, we give both necessary and sufficient conditions on the step-size sequence for the stability (MS bounded weights) of VSLMS in the classical setting of stationary uncorrelated Gaussian data. For the special case of a single tap filter, an exact characterization of the stability region is
obtained, whereby it is shown that (1) is a sufficient condition for MS bounded weights for prior step-size rules (which only depend on past data) but is not a sufficient condition for posterior step-size rules (which also depend on the current data). We also present an explicit example of unstable behavior in the latter case when (1) is satisfied. For multiple tap filters, bounds on the stability region are obtained. This work is a generalization of the results in [4] for stability of fixed step-size LMS algorithms and is important for rigorously establishing (at least under strong classical assumptions on the data) the stability of a class of adaptive algorithms that are of increasing interest in practical applications. It complements the approximate analysis of the VSLMS in [9], [11], which was shown to have value for the prediction of steady-state behavior.

A. Some Comments on the Assumptions, Methods, and Goals of the Analysis

The second-order stochastic stability analysis of the fixed step-size LMS algorithm has evolved over time. In the classical approach, LMS was analyzed under strong conditions such as stationary uncorrelated Gaussian data [4]. In this case, the weight error covariance equation is a linear time invariant (LTI) difference equation whose exponential stability can be directly examined. This exponential stability is essentially a necessary and sufficient condition for MS bounded weights. Subsequently, LMS was analyzed under weaker conditions on the data, such as stationary $M$-dependent data [10], stationary mixing data [7], and more general dependent-data models [5]. These works involve demonstrating the $p$th mean exponential stability of the (unforced) weight equation. This exponential stability for large enough $p$ is a sufficient but not necessary condition for MS bounded weights. We note that many works circumvent the difficulties associated with demonstrating exponential stability by assuming bounded data and/or bounded weights. In the latter case, the convergence and tracking behavior can be obtained by directly characterizing limit points and distributions. Such approaches may be the only viable approach for rigorous performance analysis of LMS (and more complex adaptive algorithms [2]) with general dependent-data models.

The classical approach to analyzing stability of LMS is constructive and is primarily concerned with conditions on the step-size, which ensure exponential stability (for stationary uncorrelated Gaussian data, a characterization of the step-size has actually been achieved [4]). Subsequent approaches are more qualitative and are primarily concerned with conditions on the data that ensure exponential stability for some small enough step-size (a characterization in terms of persistence of excitation condition has recently been obtained [5]). We observe that approaches that assume bounded data and/or bounded weights and give explicit step-size conditions for stability must be interpreted cautiously, especially if the conditions depend on the specific value of the data and/or weight bound as this can result in overly conservative step-sizes.

The upshot of this discussion is that in terms of predicting conditions on step-size that guarantee stability of LMS, the classical approach with (say) the assumption of stationary uncorrelated Gaussian data gives concrete results and is well-known to have empirical validity for more general data types. In this paper, we extend this approach to treat the stability of the variable step-size LMS (VSLMS) algorithm. Unlike LMS (with fixed step-size), even under the strong assumptions of stationary uncorrelated Gaussian data, an equation that recursively specifies the weight error covariance cannot be found, in general, for VSLMS with prior or posterior step-size sequences. This makes it difficult to analyze such algorithms in a constructive way, which yields bounds on step-size interval which guarantee stability (or instability). We shall circumvent this problem by examining how the stability of the VSLMS depends on an entire class of step-size sequences (specifically, the classes of prior or posterior step-sizes). We shall derive a linear time invariant (LTI) difference equation whose solution upper bounds componentwise the diagonal elements of the weight error covariance (i.e., the weight variances) uniformly over the specified class. The exponential stability of this equation is then sufficient to imply MS bounded weights for any step-size within the class. To get a necessary condition for MS bounded weights, we demonstrate a particular step-size sequence within the specified class such that the weight error covariance satisfies an unstable LTI difference equation that diverges for any initial condition. Some further discussion of the theoretical and practical aspects of our results is given in Section V.

II. THE SET-UP

Consider the VSLMS algorithm in the form

$$w(k + 1) = (I - \alpha_k x(k)x(k)^T)w(k) + \alpha_k d_k x(k)$$

$$w(0) = w_0$$

$$d_k = w^T x(k) + n_k$$  \hspace{1cm} (2)

[so that $c_k = (w^* - w(k))^T x(k) + n_k$] for $k = 0, 1, \ldots$. We shall assume the following.$^1$

A1) The regressors $\{x(k)\}$ are $d$-dimensional stationary independent Gaussian vectors with mean 0 and covariance $R$ (positive definite, with positive eigenvalues $\lambda_1, \ldots, \lambda_d$).

A2) The noise $\{n_k\}$ is a stationary independent sequence with mean 0 and variance $\sigma_n^2$.

A3) $\{x(k)\}, \{n_k\}$ and $w_0$ are independent.

The assumptions on the step-size sequences $\{\alpha_k\}$ are as follows. Let $X_k, E_k$ be the $c$-fields generated by $x(i), i \leq k$, and $e_i, i \leq k$, respectively. We shall consider both prior and posterior step-sizes $\alpha_k$ that are $X_{k-1} \vee E_{k-1}$ and $X_k \vee E_k$ measurable, respectively, and are constrained to lie within $[\alpha, \bar{\alpha}]$. We shall assume that $\alpha > 0$ to avoid certain technical problems in analyzing random step-size sequences with zero as a limit point. This is not a significant restriction since, as will be seen in the sequel, points $(\alpha, \bar{\alpha})$ with $\alpha = 0$ are either on the boundary or exterior to the region where the weights are MS bounded.

$^1$This system identification model is slightly more general than the adaptive filtering model of stationary independent Gaussian data (data vectors and desired signals).
Since we do not, in general, have a weight error covariance equation for random step-size, we choose to define stability directly in terms of MS bounded weights (rather than exponential stability). Let \( \mathcal{A} \) denote a particular class of step-size sequences \( \{\alpha_k\} \), e.g., \( \mathcal{A} \) could be fixed, deterministic, prior, or posterior step-sizes. We define the mean square (MS) stability region for step-size class \( \mathcal{A} \) as

\[
S_{\mathcal{A}} = \left\{(\alpha, \bar{\alpha}) : \mathbb{E}\left[\|w(k)\|^2\right] < \infty \text{ for all } \{\alpha_k\} \in \mathcal{A} \right\}
\]

such that \( 0 < \alpha \leq \alpha_k \leq \bar{\alpha} \) a.s. \hspace{1cm} (3)

This stability region for fixed step-sizes has been determined in [4] and is easily seen to coincide with the stability region for deterministic step-sizes. Here, we determine this stability region for two classes of random step-size sequences, i.e., prior and posterior step-sizes. Observe that the stability region \( S_{\mathcal{A}} \) as defined in (3) implies the following, in accordance with the discussion in Section I-A: If a step-size interval \( [\alpha, \bar{\alpha}] \) is in the stability region for step-size class \( \mathcal{A} \) i.e., \( (\alpha, \bar{\alpha}) \in S_{\mathcal{A}} \), then the weights are MS bounded for any step-size sequence \( \{\alpha_k\} \) in class \( \mathcal{A} \) with \( \alpha_k \in [\alpha, \bar{\alpha}] \); conversely, if a step-size interval \( [\alpha, \bar{\alpha}] \) is not in the stability region for step-size class \( \mathcal{A} \) i.e., \( (\alpha, \bar{\alpha}) \notin S_{\mathcal{A}} \), then the weights are not MS bounded for some step-size sequence \( \{\alpha_k\} \) in class \( \mathcal{A} \) with \( \alpha_k \in [\alpha, \bar{\alpha}] \).

Now, from [4] for the fixed step-size case \( \alpha_k = \alpha \), it is known that the weights are MS bounded if and only if \( 0 \leq \alpha < \alpha^* \), where

\[
\alpha^* = \sup \left\{ \alpha : 0 < \alpha < \frac{1}{\lambda_i}, i = 1, \ldots, d \right\}.
\]

It follows that

\[
S_{\text{posterior}} \subseteq S_{\text{prior}} \subseteq S_{\text{deterministic}} = S_{\text{fixed}} = \left\{(\alpha, \bar{\alpha}) : 0 < \alpha \leq \bar{\alpha} < \alpha^* \right\}
\]

where the inclusions are valid because deterministic step-size is a special case of prior step-size, which is in turn a special case of posterior step-size. We define the lower envelope of the MS stability region \( S_{\mathcal{A}} \) as

\[
\alpha_{\mathcal{A}}(\bar{\alpha}) = \inf\{(\alpha, \bar{\alpha}) : (\alpha, \bar{\alpha}) \in S_{\mathcal{A}}\}, \hspace{1cm} 0 < \alpha < \alpha^*
\]

\( \alpha_{\mathcal{A}}(\bar{\alpha}) \) is well-defined since \( (\alpha, \bar{\alpha}) \in S_{\mathcal{A}} \). When the class of step-size sequences under consideration is clear, we will refer to the MS stability region as \( S \) and its lower envelope as \( \alpha(\bar{\alpha}) \) (no subscripts). It turns out that the stability region and its lower envelope are independent of the initial condition \( w_0 \) (\( \mathbb{E}[\|w_0\|^2] < \infty \)).

To illustrate these definitions, in Fig. 1, we show the posterior step-size stability region for a particular \( d = 1 \) example (we shall return to this example in Section III-A; for now,

we are just interested in its qualitative features). From (5), the stability region for deterministic step-size is bounded by the triangle \( \{(\alpha, \bar{\alpha}) : 0 < \alpha \leq \bar{\alpha} < \alpha^* = \frac{2}{\lambda}\} \) (its lower envelope is identically zero). Furthermore, for this example, the stability region for prior step-size coincides with that for deterministic step-size (its lower envelope is again identically zero). However, for this example, the stability region for posterior step-size is strictly smaller than for deterministic step-size (its lower envelope is strictly positive), as is seen in Fig. 1.

Our goal is to characterize as tightly as possible the MS stability region for the classes of prior and posterior step-size rules. We first treat the single tap \( (d = 1) \) case since it is rather special (and far simpler), and we can give the exact stability region in this case.

### III. Stability Region for Single Tap Filter

For the case where \( d = 1 \), we have the following characterizations of the MS stability region for the VSLMS algorithm in (2). Here, \( \lambda = \mathbb{E}[x(k)^2] \) and \( \alpha^* = (2/3\lambda) \).

**Theorem 1**: The MS stability region for the class of prior step-size sequences is given by (for \( d = 1 \))

\[
S_{\text{prior}} = \{(\alpha, \bar{\alpha}) : 0 < \alpha \leq \bar{\alpha} < \alpha^* \}.
\]

**Theorem 2**: The MS stability region for the class of posterior step-size sequences is given by (for \( d = 1 \))

\[
S_{\text{posterior}} = \{(\alpha, \bar{\alpha}) : 0 < \alpha \leq \bar{\alpha} < \alpha^*, \rho < 1 \}
\]

where

\[
\rho = 1 - 2\mathbb{E}[a_k x(k)^2] + \mathbb{E}[a_k^2 x(k)^4]
\]

and

\[
a_k = \begin{cases} 
\alpha, & \text{if } \frac{\alpha + \bar{\alpha}}{2} x(k)^2 > 1 \\
\alpha_k, & \text{if } \frac{\alpha + \bar{\alpha}}{2} x(k)^2 \leq 1.
\end{cases}
\]

Furthermore, if \( \rho \geq 1 \) for some \( (\alpha, \bar{\alpha}) \), then \( \alpha_k = a_k \) is a posterior step-size sequence (with \( \alpha_k \in [\alpha, \bar{\alpha}] \)), which has unbounded weight variance.
Remark 1: $\rho = \rho(\alpha, \overline{\alpha})$ can be readily evaluated in terms of $\Phi_n(x) = \int_0^\infty t^{2n-1}e^{-t^2}dt$

$$\rho(\alpha, \overline{\alpha}) = 1 - 2\overline{\alpha}\lambda + 3\alpha^2\lambda^2 + 4(\overline{\alpha} - \alpha)\lambda \Phi_1 \left( \frac{1}{\sqrt{2}\lambda} \right) - 4(\overline{\alpha}^2 - \bar{\alpha}^2)\lambda^2 \Phi_2 \left( \frac{1}{\sqrt{2}\lambda} \right)$$

(9)

where $p = (\overline{\alpha} + \alpha)/2$.

Remark 2: It can be shown that the posterior step-size lower envelope is strictly positive for all $0 < \alpha < \alpha^*$ (see the example in Section III-A). Hence, the stability region for posterior step-size is strictly smaller than for deterministic step-size (for $d = 1$) whose lower envelope is identically zero. Note also that the stability region for prior step-size is the same as for deterministic step-size (for $d = 1$).

Next, we give the proofs of these results.

Proof of Theorem 1: Let $v(k) = u(k) - u^*$ and $f(\alpha) = E\{1 - \alpha x(k)^2\} = 1 - 2\alpha\lambda + 3\alpha^2\lambda^2$.

Then, using the assumptions on the data and step-size (see Section II), we can write

$$E\{v(k+1)^2\} = E\{f(\alpha_k)x(k)^2\} + E\{\alpha_k^2\} \lambda \sigma_n^2$$

$$\leq \sigma E\{v(k)^2\} + b$$

(10)

where

$$\sigma = \max_{\alpha \in [0,1]} f(\alpha)$$

and $b = \overline{\alpha}^2\lambda \sigma_n^2 > 0$. Now, it is seen that $f(\alpha)$ is maximized at $\alpha$ equal to

$$\alpha := \begin{cases} \overline{\alpha}, & \text{if } \frac{\alpha + \overline{\alpha}}{2} \lambda > \frac{1}{3} \\ \bar{\alpha}, & \text{if } \frac{\alpha + \overline{\alpha}}{2} \lambda \leq \frac{1}{3} \end{cases}$$

and therefore, $\sigma = f(\alpha)$.

Since $\sigma > 0$, we can replace the upper bound in (10) with equality to get

$$E\{v(k)^2\} = \sigma E\{v(k)^2\} + b$$

$$\xi(k+1) = \sigma \xi(k) + b$$

(11)

and $E\{v(k)^2\} \leq \xi(k)$ for all $k$. Now, (11) is exponentially stable if and only if $\sigma < 1$. In addition, this exponential stability implies $\sup_{k \geq 0} E\{v(k)^2\} \leq \sup_{k \geq 0} \xi(k) < \infty$. Hence, $S \supset \{(\alpha, \overline{\alpha}); 0 < \alpha \leq \overline{\alpha} < \alpha^*, \sigma < 1\}$, and we already know that $S \subset \{(\overline{\alpha}, \alpha^*); 0 < \overline{\alpha} \leq \alpha < \alpha^*\}$. It is easily checked that if $0 < \overline{\alpha} \leq \overline{\alpha} < \alpha^*$, then $\sigma < 1$. The theorem follows.

Proof of Theorem 2: Let $v(k) = u(k) - u^*$ and $f(\alpha, \lambda) = (1 - \alpha x^2)^2 = 1 - 2\alpha x^2 + \alpha^2 x^4$.

Then, using the assumptions on the data and step-size (see Section II) and Holder’s inequality, we have

$$E\{v(k+1)^2\} = E\{f(\alpha_k, x(k))v(k)^2\}$$

$$+ E\{2(1 - \alpha_k x(k)^2)v(k)\alpha_k x(k)n_k\}$$

$$+ E\{\alpha_k^2 x(k)^2\sigma_n^2\}$$

$$\leq \rho E\{v(k)^2\} + 2\rho^{1/2}(E\{v(k)^2\})^{1/2} + b$$

(12)

where

$$\rho = E\{\max_{\alpha \in [0,1]} f(\alpha, x(k))\}$$

and $b = \overline{\alpha}^2\lambda \sigma_n^2 > 0$. It is seen that $f(\alpha, x(k))$ is maximized at $\alpha$ equal to

$$a_k := \begin{cases} \overline{\alpha}, & \text{if } \frac{\alpha + \overline{\alpha}}{2} x(k)^2 > 1 \\ \alpha, & \text{if } \frac{\alpha + \overline{\alpha}}{2} x(k)^2 < 1 \end{cases}$$

and therefore, $\rho = E\{f(a_k, x(k))\}$.

We can write (12) as

$$E\{v(k+1)^2\} \leq \rho E\{v(k)^2\} + |O(E\{v(k)^2\})/\lambda^2| + b$$

(14)

where $O(\cdot)$ denotes a real-valued function such that $|O(\xi)| \leq L\xi$ for all $\xi$ with some positive constant. Now, given $\varepsilon > 0$, there exists $K > 0$ such that $|O(\xi^2/\lambda)| \leq \varepsilon K$ for all $\xi \geq 0$. Hence, for this choice of $\varepsilon$ and $K$, we have

$$E\{v(k+1)^2\} \leq (\rho + \varepsilon) E\{v(k)^2\} + K + b$$

(15)

Since $\rho + \varepsilon > 0$, we can replace the upper bound in (15) with equality to get

$$\xi(k+1) = (\rho + \varepsilon) \xi(k) + K + b$$

$$\xi(0) = E\{v(0)^2\}$$

(16)

and $E\{v(k)^2\} \leq \xi(k)$ for all $k$. However, given $\rho < 1$, we can choose $\varepsilon > 0$ such that $\rho + \varepsilon < 1$, which implies that (16) is exponentially stable. In addition, this exponential stability implies $\sup_{k \geq 0} E\{v(k)^2\} \leq \sup_{k \geq 0} \xi(k) < \infty$. Hence, $S \supset \{(\alpha, \overline{\alpha}); 0 < \alpha \leq \overline{\alpha} < \alpha^*, \rho < 1\}$, and we already know that $S \subset \{(\overline{\alpha}, \alpha^*); 0 < \overline{\alpha} \leq \alpha < \alpha^*\}$. Now, suppose $\rho \geq 1$, and let $\alpha_k = a_k$. Then, it is easy to see that

$$E\{v(k+1)^2\} = \rho E\{v(k)^2\} + c$$

where $c = E\{\alpha_k x(k)^2\sigma_n^2\} > 0$, and therefore, $E\{v(k)^2\} \rightarrow \infty$.

The theorem follows.

A. An Example

In Fig. 1, we plot the lower envelope of the $d = 1$ MS stability region for posterior step-size from Theorem 2 when $\lambda = 1$, and hence $\alpha^* = 2/3$ (the posterior step-size stability region lies above the envelope and below the line $\alpha = \overline{\alpha}$). Observe that the $d = 1$ MS stability region for posterior step-size is strictly smaller than for deterministic step-size (the deterministic step-size stability region lies above the line $\alpha = 0$ and below the line $\alpha = \overline{\alpha}$). We confirm the analysis via simulation with parameters $\lambda = 1$, $\sigma_n^2 = 0.01$, and the posterior step-size given in (8). We set $\overline{\alpha}$ to $1/3$ so that from Fig. 1, the smallest value of $\alpha$ for MS bounded weights is about 0.016. In Figs. 2 and 3, we show the MSE learning curve (averaged over 100 trials) for VSLMS with $\alpha = 0.08$ and $\alpha = 0.0032$, respectively, and LMS with $\alpha = 1/3$. The MSE of VSLMS in Fig. 3 appears to diverge, whereas that in Fig. 2 appears to be bounded, in accordance with the theory.
IV. Stability Region for Multiple Tap Filter

For the case where \( d > 1 \), we have the following approximations of the MS stability region for the VSLMS algorithm in (2). Let \( y_i(k) \) denote the \( i \)th component of \( \mathbf{y}(k) = \mathbf{M}^T \mathbf{x}(k) \), where \( \mathbf{M} \) is the orthogonal matrix such that \( \mathbb{E}\{\mathbf{y}(k)\mathbf{y}(k)^T\} = \mathbf{M}^T \mathbf{R} \mathbf{M} = \text{diag}(\lambda_1, \ldots, \lambda_d) \). Let \( i^* = \arg \max \lambda_i \cdot \alpha^* \) is given as in (4).

**Theorem 3**: The MS stability region for the class of prior step-size sequences satisfies

\[
S_{\text{prior}} \supset \left\{ (\alpha, \overline{\alpha}) : 0 < \alpha \leq \overline{\alpha} < \alpha^*; \rho_i \leq 1, i = 1, \ldots, d \right\}
\]

where

\[
\rho_i := 1 - 2 \mathbb{E}\{a_k y_i(k)^2\} + \mathbb{E}\{\sigma^2 y_i(k)^4\}
\]

\[
\frac{1}{\pi} (d - 1)(\overline{\alpha} - \alpha) \lambda_i + \frac{4}{\pi} (d - 1)(\overline{\alpha}^2 - \alpha^2) \lambda_i^2
\]

\[
\gamma_1 = \frac{1}{\pi} (\overline{\alpha}^2 - \alpha^2)(d - 1)
\]

\[
\gamma_2 = \frac{1}{\pi} (\overline{\alpha} - \alpha)
\]

**Remark 3**: The quantities \( \rho_i \) can be broken out according to the cases \( i = i^* \) and \( i \neq i^* \) and readily evaluated in terms of \( \Phi_0(x) = \left(\frac{2\sqrt{\pi}}{\sqrt{\pi}}\right) \int_0^x e^{-\xi^2} d\xi \) (similarly to the \( d = 1 \) case as in Remark 1). These evaluations are omitted here due to space limitations.

**Theorem 4**: The MS stability region for the class of posterior step-size sequences satisfies the following.

\[
S_{\text{posterior}} \subset \left\{ (\alpha, \overline{\alpha}) : 0 < \alpha \leq \overline{\alpha} < \alpha^*; \rho_i < 1, i = 1, \ldots, d \right\}
\]

where

\[
\rho_i := 1 - 2 \mathbb{E}\{a_k y_i(k)^2\} + \mathbb{E}\{\sigma^2 y_i(k)^4\}
\]

\[
\mu_i := \frac{\mathbb{E}\{\sigma^2 y_i(k)^4\}}{\mathbb{E}\{a_k^2\}}
\]

\[
\gamma_3 = \alpha^2 - \overline{\alpha}^2
\]

**Remark 4**: The quantities \( \rho_i \) can be broken out according to the cases \( i = i^* \) and \( i \neq i^* \) and readily evaluated in terms of \( \Phi_0(x) = \left(\frac{2\sqrt{\pi}}{\sqrt{\pi}}\right) \int_0^x e^{-\xi^2} d\xi \) (similarly to the \( d = 1 \) case as in Remark 1). These evaluations are omitted here due to space limitations.

Furthermore, if \( \max_{i \leq d} \rho_i \geq 1 \) or \( \sum_{i=1}^d (\overline{\alpha}^2 - \gamma_3) \mu_i^2 / (1 - \rho_i) \geq 1 \) for some \( \alpha_k \in [\overline{\alpha}, \alpha_i] \), then \( a_k = a_k \) is a posterior step-size sequence (with \( a_k \in [\overline{\alpha}, \alpha_i] \)), which has unbounded weight variance.

**Remark 5**: The quantities \( \mu_i \) can be broken out according to the cases \( i = i^* \) and \( i \neq i^* \) and readily evaluated in terms of \( \Phi_0(x) = \left(\frac{2\sqrt{\pi}}{\sqrt{\pi}}\right) \int_0^x e^{-\xi^2} d\xi \) (similarly to the \( d = 1 \) case as in Remark 1). These evaluations are omitted here due to space limitations.
Remark 4: The inclusion (17) corresponds to an upper bound on the prior step-size lower envelope; the inclusions (18) and (22) correspond to upper and lower bounds on the posterior step-size lower envelope, respectively. The lower bound on the prior step-size lower envelope can be positive for some $0<\overline{\alpha}<\alpha^*$ (see the example in Section IV-A). Hence, the stability region for posterior step-size can be strictly smaller than for deterministic step-size, whose lower envelope is identically zero. The upper bound on the prior step-size lower envelope can be positive for some $0<\overline{\alpha}<\alpha^*$ (see the example in Section IV-A). Note, however, that this does not imply that the stability region for prior step-size can be strictly smaller than for deterministic step-size. In fact, we do not know for $d>1$ whether the stability region for prior step-size can be strictly smaller than for deterministic step-size; we conjecture that the regions coincide (analogously to the $d=1$ case in Theorem 1).

Next, we give the proofs of Theorems 3 and 4.

Proof of Theorem 3: Let $\mathbf{v}(k) = \mathbf{M}^T(\mathbf{w}(k) - \mathbf{w}^*)$. Then, we have

$$
\mathbf{v}(k+1) = (\mathbf{I} - \alpha k \mathbf{v}(k) \mathbf{v}(k)^T) \mathbf{v}(k) + \alpha k \eta(k),
$$

(27)

Let $C(k) = (\epsilon_{ij}(k)) = E\{\mathbf{v}(k) \mathbf{v}(k)^T\}$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d) = E\{\mathbf{y}(k) \mathbf{y}(k)^T\}$. Using the assumptions on the data and the step-size (see Section II) and Gaussian moment factoring, we get

$$
C(k+1) = C(k) - (\Lambda \text{E}\{\alpha k \mathbf{v}(k) \mathbf{v}(k)^T\})
+ 2\Lambda \text{E}\{\alpha^2 k^2 \mathbf{v}(k) \mathbf{v}(k)^T\} \Lambda
+ \text{tr}(\Lambda \text{E}\{\alpha^2 k^2 \mathbf{v}(k) \mathbf{v}(k)^T\}) \Lambda
+ \text{E}\{\alpha^2 k^4 \mathbf{v}(k) \mathbf{v}(k)^T\},
$$

Let

$$
f_0(\alpha) = E\{(1 - \alpha \epsilon_{ij}(k)^2)^2\} = 1 - 2\alpha \lambda_j + 3\epsilon^2 \lambda_j^2,
$$

(28)

Writing out the diagonal components gives

$$
c_{ij}(k+1) = E\{f_0(\alpha) \epsilon_{ij}(k)^2\} + E\left\{\lambda_j \epsilon_{ij}(k)^2\right\}
+ 2\lambda_j E\{\alpha^2 \epsilon_{ij}(k)^2\} \lambda_j
+ \lambda_j \epsilon_{ij}(k)^2 + b_i
$$

(29)

where

$$
\sigma_i^2 = \max_{\alpha \in [0,1]} f_0(\alpha)
$$

and $b_i = \epsilon^2 \lambda_j^2 > 0$. Now, it is seen that $f_0(\alpha)$ is maximized at $\alpha$ equal to

$$
\alpha_i := \begin{cases}
\overline{\alpha}, & \text{if } \frac{\alpha + \overline{\alpha}}{2} \lambda_j > \frac{1}{3}
\alpha, & \text{if } \frac{\alpha + \overline{\alpha}}{2} \lambda_j \leq \frac{1}{3}
\end{cases}
$$

and therefore $\sigma_i^2 = f_0(\alpha_i)$. Rearranging (29), we get

$$
c_{ij}(k+1) \leq \overline{\alpha} \sigma_i^2(k) + \overline{\alpha}^2 \lambda_j \sum_j \lambda_j \epsilon_{ij}(k) + b_i
$$

(30)

where

$$
\overline{\sigma}_i = \sigma_i^2 - \overline{\alpha}^2 \lambda_j^2.
$$

Since $\overline{\sigma}_i \geq 0$ for all $i$ (as are the coefficients of all the other homogeneous terms), the upper bound in (30) can be replaced with equality to get

$$
\xi_i(k+1) = \overline{\sigma}_i \xi_i(k) + \overline{\alpha}^2 \lambda_j \sum_j \lambda_j \epsilon_{ij}(k) + b_i
$$

\xi_i(0) = c_{ij}(0)

(32)

and $c_{ij}(k) \leq \xi_i(k)$ for all $i$. In vector notation, this equation is

$$
\xi(k+1) = \Lambda \xi(k) + \mathbf{b}
$$

(33)

where

$$
\Lambda = \text{diag}(\overline{\sigma}_1, \ldots, \overline{\sigma}_d) + \overline{\alpha}^2 \mathbf{I}^T
$$

(34)

$$
\lambda = [\lambda_1, \ldots, \lambda_d]^T, \quad \mathbf{b} = [b_1, \ldots, b_d]^T.
$$

Now, (33) is exponentially stable if and only if the eigenvalues of $\Lambda$ all have magnitude less than one. In addition, this exponential stability implies $\sup_k \text{tr}(C(k)) \leq d \sup_k [\xi(k)] < \infty$. Let

$$
g = \sum_j \frac{\overline{\alpha}^2 \lambda_j^2}{1 - \overline{\alpha}^2 \lambda_j^2}
$$

(35)

Similarly to [4], it is shown that the eigenvalues of $\Lambda$ all have magnitude less than one if and only if $0 < \overline{\alpha} < \alpha^*$ and $g < \frac{1}{2}$, or equivalently, $\sigma_i^2 < 1 \forall i$ and $g < \frac{1}{2}$. Hence

$$
S = \{ \alpha, \overline{\alpha} : 0 < \alpha \leq \overline{\alpha} < \alpha^*, \sigma_i^2 < 1, i = 1, \ldots, d; g < 1 \}.
$$

It is easily checked that if $0 < \alpha \leq \overline{\alpha} < \alpha^*$, then $\sigma_i^2 < 1$ for all $i$. In addition, $\sigma_i^2 = \max_{\alpha \in [0,1]} f_0(\alpha)$, so that combining (28), (31), and (35), we obtain the following expression for $g$:

$$
g = \sum_{\alpha \in [0,1]} \frac{\overline{\alpha}^2 \lambda_j^2}{1 - \overline{\alpha}^2 \lambda_j^2} + \overline{\alpha}^2 \lambda_j^2
$$

Theorem follows directly. □

Proof of Theorem 4a): Define $\mathbf{v}(k), C(k)$, and $\Lambda$ as in the proof of Theorem 3. Using the assumptions on the step-size and the data (see Section II), we get

$$
C(k+1) = C(k) - \text{E}\{\alpha k \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{v}(k) \mathbf{v}(k)^T
+ \mathbf{v}(k) \mathbf{v}(k)^T \mathbf{y}(k) \mathbf{y}(k)^T\}
+ \text{E}\{\alpha^2 k^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{v}(k) \mathbf{v}(k)^T
+ \mathbf{v}(k) \mathbf{v}(k)^T \mathbf{y}(k) \mathbf{y}(k)^T\}
+ \text{E}\{\alpha \mathbf{y}(k) \mathbf{y}(k)^T (\mathbf{I} - \alpha k \mathbf{y}(k) \mathbf{y}(k)^T) \mathbf{v}(k) \mathbf{v}(k)^T
+ \mathbf{y}(k) \mathbf{v}(k)^T \mathbf{y}(k) \mathbf{v}(k)^T\}
+ \text{E}\{\alpha^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{n}_k\}
+ \text{E}\{\alpha^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{n}_k\},
$$

Let

$$
f(\alpha, y) = (1 - \alpha^2 y^2) = 1 - 2\alpha y^2 + \alpha^2 y^4.
$$
Writing out the diagonal components gives
\begin{align*}
c_{ii}(k+1) &= \mathbb{E}\{f(\alpha_k y(k))\sigma_i(k)^2\} \\
&= \mathbb{E}\{\alpha_k^2 y(k)^2 \sum_{j \neq i} y_j(k)^2 v_j(k)^2\} \\
&- 2p_k(k) + q_i(k) \\
&+ 2d_{ii}(k) + \mathbb{E}\{\alpha_k^2 y(k)^2 \sigma_i^2\} \\
&\leq \rho_1 \varepsilon_i(k) + \varepsilon \lambda_i \sum_{j \neq i} \lambda_j c_{jj}(k) - 2p_k(k) + q_i(k) \\
&+ 2d_{ii}(k) + b_i \\
\tag{36}
\end{align*}

where
\begin{align*}
\rho_1^2 &= \mathbb{E}\left\{ \max_{\alpha \in [\alpha_0, \alpha_1]} f(\alpha, y(k)) \right\} \\
\tag{37}
D(k) &= (d_{ij}(k)) = \mathbb{E}\{\alpha_k (I - \alpha_k y(k) y(k)^T)v(k)y(k)^T n_h\} \\
\tag{38}
p_i(k) &= \mathbb{E}\left\{ \alpha_k \sum_{j \neq i} y_j(k) y_j(k) v_i(k) v_j(k) \right\} \\
\tag{39}
q_i(k) &= \mathbb{E}\left\{ \alpha_k^2 \sum_{j \neq i} y_j(k)^2 y_j(k) v_i(k) v_j(k) \right\} \\
\tag{40}
\end{align*}
and \( b_i = \varepsilon \lambda_i \sigma_i^2 > 0 \). Now, it is seen that \( f(\alpha, y(k)) \) is maximized at \( \alpha \) equal to
\[ a_k \colon= \begin{cases} 
\bar{\alpha}, & \text{if } \frac{\alpha + \bar{\alpha}}{2} y_k^2(k) > 1 \\
\alpha, & \text{if } \frac{\alpha + \bar{\alpha}}{2} y_k^2(k) < 1 
\end{cases} \]
and therefore, \( \rho_1^2 = \mathbb{E}\{f(a_{k,i}, y(k))\} \).

The \( d_{ii}(k) \) term in (38) can be bounded as follows. Let \( \| \cdot \| \) denote the Euclidean matrix norm. We have
\begin{align*}
|d_{ii}(k)| &\leq \|D(k)\| \\
&\leq \alpha \mathbb{E}\{\|\mathbb{E}\{I - \alpha_k y(k) y(k)^T\}v(k)y(k)^T n_h\|\} \\
\end{align*}

Applying Euclidean matrix norm properties as well as Holder and triangle inequalities for the norm \( \mathbb{E}\{\cdot\}^{1/2} \) gives
\begin{align*}
|d_{ii}(k)| &\leq \alpha \mathbb{E}\{\|\mathbb{E}\{I - \alpha_k y(k) y(k)^T\}v(k)y(k)^T n_h\|\} \\
&\leq \alpha (1 + \alpha \mathbb{E}\{\|y(k)^2\|^{1/2}\}) \\
&\cdot \mathbb{E}\{\|v(k)^2\|^{1/2}\} \mathbb{E}\{\|y(k)^2\|^{1/2} \sigma_n\} \\
&\leq \alpha (1 + \alpha \mathbb{E}\{2 \tanh(A)^2 + (\tanh(A)^2)^{1/2}\}) \\
&\cdot (\tanh(A)^2 + \sigma_n \mathbb{E}\{\mathbb{E}\{C(k)^2\}\}^{1/2}) \\
\tag{41}
\end{align*}

where we have also used the assumptions on the data and the step-size and (in the last step) Gaussian moment factoring. The important observation (as seen below) is that \( |d_{ii}(k)| = O(\mathbb{E}\{\mathbb{E}\{C(k)^2\}\}^{1/2}) \).

We will need the following lower bound on \( p_i(k) \) and upper bound on \( q_i(k) \).

Proposition 1:  a)
\begin{align*}
p_i(k) &\geq \frac{1}{2\pi} (\bar{\alpha} - \bar{\alpha})(d-1) \lambda_i c_{ii}(k) \\
&+ \frac{1}{2\pi} (\bar{\alpha} - \bar{\alpha}) \sum_{j \neq i} \lambda_j c_{jj}(k). \\
\tag{42}
\end{align*}
b)
\begin{align*}
q_i(k) &\leq \frac{4}{\pi} (\bar{\alpha}^2 - \bar{\alpha}^2)(d-1) \lambda_i^2 c_{ii}(k) \\
&+ \frac{1}{\pi} (\bar{\alpha}^2 - \bar{\alpha}^2)(d-1) \lambda_i \sum_{j \neq i} \lambda_j c_{jj}(k). \\
\tag{43}
\end{align*}

Assume for the moment that Proposition 1 is true. Substituting (41)–(43) in (36), we get
\begin{align*}
c_{ii}(k+1) &\leq \rho_1^2 c_{ii}(k) + ((\bar{\alpha}^2 + \gamma_1) \lambda_i + \gamma_2) \\
&\cdot \sum_{j \neq i} \lambda_j c_{jj}(k) + \mathbb{E}\{\mathbb{E}\{C(k)^2\}\}^{1/2} + b_i \\
\tag{44}
\end{align*}

where \( \rho_1^2, \gamma_1, \gamma_2 \) are given in (19) and (20). Rearranging terms gives
\begin{align*}
c_{ii}(k+1) &\leq \rho_1^2 c_{ii}(k) + ((\bar{\alpha}^2 + \gamma_1) \lambda_i + \gamma_2) \\
&\cdot \sum_{j} \lambda_j c_{jj}(k) + \mathbb{E}\{\mathbb{E}\{C(k)^2\}\}^{1/2} + b_i \\
\tag{44}
\end{align*}
where
\[ \tilde{\rho}_1 = \rho_1^2 - (\bar{\alpha}^2 + \gamma_1) \lambda_i^2 - \gamma_2 \lambda_i. \]

Since \( \tilde{\rho}_1 \geq 0 \) for all \( i \) (as are the coefficients of all the other homogeneous terms), the upper bound in (44) can replaced with equality to get
\begin{align*}
\xi(k+1) &= \tilde{p}_i \xi(k) + ((\bar{\alpha}^2 + \gamma_1) \lambda_i + \gamma_2) \sum_{j} \lambda_j \xi_j(k) \\
&\quad + O\left(\left(\sum_{j} \xi_j(k)^{1/2}\right)^{1/2}\right) + b_i \\
\tag{45}
\end{align*}

and \( c_{ii}(k) \leq \xi_i(k) \) for all \( k \). We can write this equation in vector notation as
\begin{align*}
\xi(k+1) &= A \xi(k) + O(\xi(k)^{1/2}) + b \\
\tag{46}
\end{align*}
where
\[ A = \text{diag}(\tilde{\rho}_1, \ldots, \tilde{\rho}_l) + (\bar{\alpha}^2 + \gamma_1) \lambda \Lambda^T + \gamma_2 \lambda \Lambda^T \]
\[ \lambda = [\lambda_1, \ldots, \lambda_d]^T, \quad b = [b_1, \ldots, b_l]^T, \] and \( I \) is a \( (d \times d) \) dimensional column vector of all ones. Here, \( \mathbb{E}\{\mathbb{E}\{\cdot\}\} \) denotes a vector-valued function such that \( \mathbb{E}\{\mathbb{E}\{\xi(k)\}\} \leq L \mathbb{E}\{\xi(k)\} \) for all real-valued \( f(\cdot) \) and all vectors \( \xi \), where \( L \) is some positive constant. Taking Euclidean norms in (46) gives
\begin{align*}
\|\xi(k+1)\| \leq \|A\| \|\xi(k)\| + O(\|\xi(k)\|^{1/2}) + \|b\|. \\
\tag{47}
\end{align*}

Similarly to the proof of Theorem 2, (48) can be upper-bounded by the solution of an LTI difference equation, and if \( \|A\| < 1 \), then this equation is exponentially stable, which implies \( \sup_k \mathbb{E}\{C(k)\} \leq d \sup_k \mathbb{E}\{\xi(k)\} < \infty \). Now, for prior
step-size [see (34)], $A$ is symmetric, and hence, $\|A\|$ is the largest eigenvalue of $A$ in magnitude. In addition, there is a simple characterization of when the eigenvalues of $A$ all have magnitude less than 1 in terms of the eigenvalues of the correlation matrix $\mathbf{R}$ (equivalently, of $\Lambda$) and the step-size interval $[\alpha_1, \alpha_2]$. However, for posterior step-size [see (47)] $A$ is not symmetric.

To get a sufficient condition for the stability of (46) in terms of the eigenvalues of $A$, we can argue as follows. First, we show by induction that

$$
\xi(k+n) = A^n \xi(k) + O_n(\|\xi(k)\|^{1/2} + 1)
$$

where $O_n(\cdot)$ is a particular $O(\cdot)$ function that we subscript by $n$ for clarity. Now, (49) is true for $n=1$. Suppose (49) holds for $n = i - 1$. Then, from (46)

$$
\xi(k+i) = A^i \xi(k) + \Delta_i(k)
$$

where

$$
\Delta_i(k) = AO_{i-1}(\|\xi(k)\|^{1/2} + 1)
+ O_{i-1}(\|A^{i-1} \xi(k)\|^{1/2} + 1)^{1/2} + 1).
$$

Using the triangle inequality and properties of Euclidean matrix norms

$$
|\Delta_i(k)| \leq |A|O(\|\xi(k)\|^{1/2} + 1)
+ O((|A|^{i-1} \xi(k)) + O(|\xi(k)|^{1/2} + 1))^{1/2} + 1)
= O(|\xi(k)|^{1/2} + 1).
$$

Hence, $\Delta_i(k) = O(|\xi(k)|^{1/2} + 1)$, and (49) holds for $n = i$ and by induction for all $n$. Now, fix $n$, and set $k = mn + l$ ($m = 0, 1, \ldots, l = 0, 1, \ldots, n-1$) in (49). Taking norms, we have

$$
|\xi((m+1)n+1) - \xi((m+1)n) + O(nm + n) + O(|\xi((m+1)n+1)|^{1/2} + 1) + b)
$$

where $b$ is some positive constant. Similarly to the proof of Theorem 2, (50) can be upperbounded by the solution of an LTI difference equation, and if $|A^m| < 1$, then this equation is exponentially stable, which implies \( \sup_{n} \|C(mn + n)\| < \infty \) for \( l = 0, \ldots, n-1 \) and, hence, \( \sup_{n} \|C(kn)\| < \infty \). Furthermore, as is well known, this exists $n$ such that $|A^n| < 1$ if all the eigenvalues of $A$ are less than 1 in magnitude.

Let

$$
g = \sum_{i} (\alpha_i^2 + \gamma_i) \lambda_i^2 + \gamma_i \lambda_i.
$$

Similarly to [4], it is shown that the eigenvalues of $A$ have magnitude less than one if and only if $0 < \rho_i < 1$ for all $i$ and $g < 1$, or equivalently, $\rho_i < 1$ for all $i$ and $g < 1$. Hence $S = \{ (\alpha, \gamma) : 0 < \alpha \leq \alpha_1 < \alpha* \}$, $\rho_i < 1, i = 1, \ldots, d ; g < 1$.

Part a) of the Theorem follows.

It remains to prove Proposition 1. We will need the following lemma.

**Lemma 1:** Let $x, y, z$ be independent random variables with finite mean, and assume that $x, y$ both have zero median and zero mean. Then, we have the following.

i) $E\{xy \mid y > 0\} = E\{x \mid y > 0\} E\{y \mid y > 0\}$

ii) $E\{xy \mid y > 0\} = E\{x \mid y > 0\} E\{y \mid y > 0\}$

**Proof of Lemma 1:** i) By properties of conditional expectation

$$
E\{x \mid y > 0\} = E\{E\{x \mid y \} \mid y > 0\}
= E\{E\{x \mid y \} \mid y > 0\} + z I(z > 0) E\{x \mid y > 0\}.
$$

Since $x$ has zero mean and median, $E\{x \mid y > 0\} = -E\{x \mid y > 0\}$ and substituting this in (51) gives part i).

ii) Since $y$ has zero mean and median $E\{y \mid y > 0\} = E\{y \mid y > 0\}$ and therefore, part ii) follows from i) by taking $z = y$. □

**Proof of Proposition 1:** It will be notationally convenient to suppress the dependence of $\xi(k), \psi(k)$ on $k$. We assume that $A$ is stable.

a) Suppose $j \neq i$. Since $y_i, y_j$ are independent zero median random variables independent of $\psi_i, \psi_j$

$$
E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\} = E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}
= E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}.
$$

Now, from Lemma 1

$$
E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}
= E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}
= E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}.
$$

Hence, combining (52) and (53)

$$
E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}
= E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}.
$$

Now, using Holder’s inequality

$$
E\{o_{ik}y_{ik}y_{ij}y_{ij} > 0\}
\geq \frac{1}{2}(\alpha - \bar{c}) m_i m_j E\{\psi_i \psi_j\}.
$$

Substituting this in (39) gives

$$
p(k) \geq \frac{1}{2}(\alpha - \bar{c}) (d - 1) m_i^2 m_j^2 C_i(k)\sum_{j \neq i} m_j^2 C_j(k)).
$$

and part a) of the proposition follows by substituting for $m_i$ from (54).
**Theorem 4b):** Define $\mathbf{v}(k), \mathbf{C}(k)$, and $\Lambda$ as in the proof of Theorem 3. Consider the VSLMS algorithm in (27) with step-size $\alpha_k = \alpha_k$. Using the assumptions on the data and the choice of $\alpha_k$, we have

$$
\mathbf{C}(k + 1) = \mathbf{C}(k) - \left( E\{\alpha_k \mathbf{y}(k) \mathbf{y}(k)^T\} \right) \mathbf{C}(k) + \left( C(k) E\{\alpha_k \mathbf{y}(k) \mathbf{y}(k)^T\} \right) + E\{\alpha_k^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{C}(k) \mathbf{y}(k) \mathbf{y}(k)^T\} + \mathbf{B},
$$

where the constant term $\mathbf{B} = \langle b_{ij} \rangle = E\{\alpha_k^2 \mathbf{y}(k) \mathbf{y}(k)^T\} \sigma_e^2 > 0$. Writing out the diagonal components

$$
c_{ii}(k + 1) = \hat{p}_i c_{ii}(k) + (\bar{\sigma}^2 - \gamma_0) \mu_i \sum_{j \neq i} \mu_j c_{ij}(k) + b_{ii}
$$

Next, suppose $j \neq i$ and $j = i$. Since $y_{ji}, y_{ii}$ are independent zero-median random variables independent of $\nu_{ji}, \nu_{ii}$

$$
E\{\alpha_k^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{C}(k) \mathbf{y}(k) \mathbf{y}(k)^T\} = \frac{1}{2} (\bar{\sigma}^2 - \gamma_0) m_{3,i} m_i E\{\nu_{ji} \nu_{ii}\}. \quad (57)
$$

Now, from Lemma 1

$$
E\{\mathbf{y}(k) \mathbf{y}(k)^T \mathbf{C}(k) \mathbf{y}(k) \mathbf{y}(k)^T\} = E\{\mathbf{y}(k) \mathbf{y}(k)^T\} E\{\mathbf{C}(k)\} = E\{\mathbf{y}(k) \mathbf{y}(k)^T\} E\{\mathbf{C}(k)\} - E\{\mathbf{y}(k) \mathbf{y}(k)^T\} E\{\mathbf{C}(k)\}
$$

Hence, combining (57) and (58) gives

$$
E\{\alpha_k^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{C}(k) \mathbf{y}(k) \mathbf{y}(k)^T\} \leq \frac{1}{4} (\bar{\sigma}^2 - \gamma_0) m_{3,i} m_i E\{\nu_{ji} \nu_{ii}\}. \quad (59)
$$

Now, using Holder’s inequality

$$
E\{\alpha_k^2 \mathbf{y}(k) \mathbf{y}(k)^T \mathbf{C}(k) \mathbf{y}(k) \mathbf{y}(k)^T\} \leq \frac{1}{2} (\bar{\sigma}^2 - \gamma_0) m_{3,i} \lambda_i E\{\nu_{ji} \nu_{ii}\}.
$$

Substituting (56) and (60) in (40) gives (after some simplification)

$$
q_i(k) \leq \frac{1}{2} (\bar{\sigma}^2 - \gamma_0) \lambda_i (d - 1) \left( \frac{m_{3,i} \lambda_i}{\lambda_i} \right)^2 c_{ii}(k) + \sum_{j \neq i} m_{3,j} c_{ij}(k). \quad (60)
$$

and part b) of the proposition follows by substituting for $m_i$ and $m_{3,i}$ from (54) and (59).

This completes the proof of Theorem 4a).
smaller than for deterministic step-size (the deterministic step-size stability region lies above the line $\bar{\alpha} = 0$ and below the line $\bar{\alpha} = \bar{\varepsilon}$). In Figs. 6 and 7, we plot upper bounds on the lower envelope of the $d = 3$ MS stability region for prior step-size from Theorem 3 for the two choices of $\Lambda$ above (the prior step-size stability region is a superset of the region that lies above the upper bound on the envelope and below the line $\bar{\alpha} = \bar{\varepsilon}$). We confirm the analysis via simulation with parameters $\Lambda = \text{diag}(1.2, 10), \sigma^2_{\varepsilon} = 0.01$ and the posterior step-size given in (26). We set $\bar{\varepsilon}$ to 0.03 so that from Fig. 5, the smallest value of $\alpha$ for MS bounded weights lies between the lower bound 0.001 and the upper bound 0.01. In Figs. 6 and 7, we show the MSE learning curve (averaged over 100 trials) for VSLMS with $\alpha = 0.02$ and $\alpha = 0.0005$, respectively, as well as LMS with $\alpha = 0.03$. The MSE of VSLMS in Fig. 9 appears to diverge, whereas that in Fig. 8 appears to be bounded, in accord with the theory. Finally, we note that although we have not been able to prove that the deterministic and prior step-size stability regions coincide (except for $d = 1$), extensive simulation results support this conjecture.

V. FURTHER DISCUSSION AND CONCLUSIONS

Variable step-size LMS (VSLMS) algorithms can be applied in many situations to improve performance relative to LMS while maintaining the simplicity and robustness of the latter. Theoretical analysis of VSLMS has previously focused on steady-state performance (e.g., estimation of the weight covariance and MSE), implicitly assuming the existence of a stability property. A proper treatment of VSLMS stability (e.g., boundedness of weight variance and MSE) is important in so far as it gives the user of such an adaptive algorithm confidence in its ability to operate over a long period in an unknown environment.

To understand the analytical tools at our disposal for analyzing VSLMS, we reviewed some standard methods for analyzing LMS. First, there is analysis of LMS under strong classical
Suppose that the components corresponding to \( \lambda_i \) with magnitude greater than or equal to one in magnitude. Then, an eigenvalue of \( A \) is positive, and that there exists an eigenvalue of \( A \) which is in general false, and the stability issue is a more complex problem that requires careful analysis.

It is much harder to analyze VSLMS then ordinary LMS. The problem is that the data-dependent step-size does not admit a recursive weight covariance equation, even under the strong assumptions of stationary uncorrelated Gaussian data. To overcome this problem, we examined conditions on the step-size interval for stability over an entire class of step-size rules (prior or posterior step-size). A step-size interval is in the stable region for a specified step-size class if and only if the weights are MS bounded for all step-size rules that lie in that interval and belong to the class. For a single tap filter, we found the exact stability region for both posterior and prior step-size and showed that the posterior step-size region is strictly smaller than for deterministic step-size, whereas the prior step-size region coincides with that of deterministic step-size. For multiple taps, we found bounds on the stability region for both posterior and prior step-size and showed that the posterior step-size region is strictly smaller than for deterministic step-size, whereas the exact relationship between the prior and deterministic step-size regions is unknown (we conjecture that they coincide).

Our results apply to the stability of VSLMS algorithms over entire classes of step-size rules. They do not give precise estimates of stability regions for particular VSLMS algorithms and step-size rules. The goal of precisely predicting the stability region for a particular VSLMS algorithm is an interesting (and difficult) problem. One idea, which is based on the techniques developed in this paper, is to parameterize the class of posterior step-size sequences by a mixing parameter that reflects how strong the dependence on the current data is. Bounds similar to those developed in this paper might then be found in terms of the mixing parameter. By relating the mixing parameter to some "small" parameter that appears in a particular posterior step-size sequence, we can perhaps get a better bound (i.e., closer to a prior step-size bound).

**APPENDIX**

**Proposition 2:** Let \( z(k) \in \mathbb{R}^d \) be a solution of the linear difference equation

\[
\mathbf{z}(k+1) = \mathbf{A} \mathbf{z}(k) + \mathbf{p}, \quad k = 1, 2, \ldots
\]

where \( \mathbf{A} \in \mathbb{R}^{d \times d} \), \( \mathbf{p} \in \mathbb{R}^d \). Suppose that the components of \( \mathbf{A} \) and \( \mathbf{z}(1) \) are non-negative, that the components of \( \mathbf{p} \) are positive, and that there exists an eigenvalue of \( \mathbf{A} \) with magnitude greater than or equal to one. Then, \( \sup_k \| \mathbf{z}(k) \| = \infty \).

**Proof:** Let \( \mathbf{U} \) denote the orthogonal complement in \( \mathbb{R}^d \) of the subspace spanned by an eigenvector of \( \mathbf{A} \) corresponding to an eigenvalue of \( \mathbf{A} \) greater than or equal to one in magnitude. Let \( \mathbf{z}(k) = [z_1(k), \ldots, z_d(k)]^T \), \( \mathbf{p} = [p_1, \ldots, p_d] \), with \( z_i(k) \geq 0 \) and \( p_i > 0 \) for all \( i \). Fix the initial condition \( \mathbf{z}(1) \).
and let $z(k|p)$ denote the dependence of the solution $z(k)$ on $p$. Since the components of $A$ are non-negative, we have the following two facts: i) If $p_{i}^{1} \leq p_{i}^{2}$ for all $i$, then $z_{i}(k; p_{1}) \leq z_{i}(k; p_{2})$ for all $i$; and ii) if $\sup_{k} |z(k|p)| < \infty$, then $p \in U$.

Next, let $V = \{ p: p_{i} > 0, 1 \leq i \leq d; \sup_{k} |z(k|p)| < \infty \}$. Suppose $V \neq \emptyset$, and let $p_{0} \in V$ and $V_{0} = \{ p: 0 < p_{i} \leq p_{0,i}, 1 \leq i \leq d \}$. Using the facts cited above, we have $V_{0} \subset V \subset U$. However, $m_{d}(V_{0}) > 0$ while $m_{d}(U) = 0$, where $m_{d}(\cdot)$ is $d$-dimensional Lebesgue measure (volume), which is a contradiction. Hence, $V = \emptyset$.

**REFERENCES**


