THE LMS ALGORITHM UNDER ARBITRARY LINEARLY FILTERED PROCESSES

Markus Rupp

Institute of Telecommunications, Vienna University of Technology
Gusshausstr. 25/e389, 1040, Vienna, Austria
phone: + (431) 58801 38901, fax: + (431) 58801 38999, email: mrupp@.nt.tuwien.ac.at
web: *www.nt.tuwien.ac.at/∼rupp*

ABSTRACT

In this paper the mean square convergence of the LMS algorithm is shown for a large class of linearly filtered random driving processes. In particular this paper contains the following contributions: i) The parameter error vector covariance matrix can be decomposed into two parts, a first part that exists in the modal space of the driving process of the LMS filter and a second part, existing in its orthogonal complement space, not contributing to the performance measures (misadjustment, mismatch) of the algorithm. ii) The LMS updates force the initial values of the parameter error vector covariance matrix to remain essentially in the modal space of the driving process and components of the orthogonal complement die out. iii) The impact of additive noise is shown to contribute only to the modal space of the driving process independent of the noise statistic and thus defines the steady-state of the filter. In particular it will be shown that the joint fourth order moment $m_x^{(2,2)}$ of the decorrelated driving process is a more relevant parameter for the step-size bound and not as often believed the second order moment $m_x^{(2)}$.

1. INTRODUCTION

The famed Least Mean Square (LMS) algorithm [1] is the most successful adaptive algorithm. It can be found at million numbers in electrical echo compensators, in telephone switches as well as in the form of adaptive equalizers. With a fixed step-size, starting at initial value $\mathbf{w}_0$, the LMS algorithm is given by

$$\begin{aligned} e_k &= d_k - \mathbf{u}_k^T \mathbf{w}_k &&= v_k + \mathbf{u}_k^T(\mathbf{w}-\mathbf{w}_k) && (1)\\ \mathbf{w}_{k+1} &= \mathbf{w}_k + \mu \mathbf{u}_k e_k &&; k=0,1,2,.. && (2)\end{aligned}$$

Here, a reference model $d_k = \mathbf{w}^T\mathbf{u}_k + v_k$ has been introduced as it is common for a system identification problem, assuming that an optimal solution $\mathbf{w} \in R^{M\times 1}$ exists. It is further assumed that the observed system output is additively disturbed by real-valued, zero-mean, noise $v_k \in R$ of variance σ_v^2. Regression vector $\mathbf{u}_k \in R^{M\times 1}$, M denoting the order of the filter. The algorithm starts with some initial value $\mathbf{w}_0$, trying to improve its estimate $\mathbf{w}_k \in R^{M\times 1}$ with every time instant k. All signals are formulated as real-valued which makes the derivations easier to follow. In most cases it is straightforward to extend the results to complex-valued processes.

While deterministic approaches have proven l_2-stability for any kind of driving signal $\mathbf{u}_k$ [2–4], results from stochastic approaches are restricted to specific classes of random processes (unfiltered Independent Identically Distributed (IID) [5], Gaussian [6, 7], and Spherically Invariant Random Processes (SIRP) [8]). Nevertheless, such stochastic analysis is useful since it provides information about how the speed of convergence and the steady-state error depend on the step-size μ. The resulting stability bounds [5–9] are typically conservative:

$$\mu_{\text{classic}} \leq \frac{2}{3\text{tr}[\mathbf{R}_{\text{uu}}]} \quad (3)$$

and are based on second order moments of the autocorrelation matrix $\mathbf{R}_{\text{uu}} = E[\mathbf{u}_k\mathbf{u}_k^T]$. Furthermore, the derivation of this bound for stability in the mean square sense of arbitrary filter structures (e.g. FIR) is based on the so called

Independence Assumption (IA): Regression vector $\mathbf{u}_k$ is statistically independent of the past regression vectors, i.e., $\{\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, ..., \mathbf{u}_0\}$.

A consequence of such assumption is that parameter estimate $\mathbf{w}_k$ (as well as parameter error vector $\mathbf{w}-\mathbf{w}_k$) is independent of $\mathbf{u}_k$ and thus $E[\mathbf{u}_k\mathbf{u}_k^T(\mathbf{w}-\mathbf{w}_k)(\mathbf{w}-\mathbf{w}_k)^T\mathbf{u}_k\mathbf{u}_k^T] = E[\mathbf{u}_k\mathbf{u}_k^T E[(\mathbf{w}-\mathbf{w}_k)(\mathbf{w}-\mathbf{w}_k)^T]\mathbf{u}_k\mathbf{u}_k^T] = E[\mathbf{u}_k\mathbf{u}_k^T\mathbf{K}_k\mathbf{u}_k\mathbf{u}_k^T]$ where the parameter covariance matrix

$$\mathbf{K}_k = E[(\mathbf{w}-\mathbf{w}_k)(\mathbf{w}-\mathbf{w}_k)^T] \quad (4)$$

(in literature this is also often called the weight error covariance matrix) was applied. The IA holds exactly for the linear combiner case in which the succeeding regression vectors are statistically independent of each other (see for example in multiple sidelobe canceller (MSLC) applications [10]) while in practice the LMS algorithm is mostly run on transversal filters in which the regression vectors exhibit a shift dependency. Nevertheless the shift property of the regression vector will be considered here as part of the process generation, that is $\mathbf{u}_k = [u_k, u_{k-1}, ..., u_{k-M+1}]^T$. At the same time, our derivations are exactly correct only for the linear combiner, or assuming the IA for arbitrary structures. For very long filters, it will be shown that the IA is the only requirement, no matter what input statistic is given. The major advantages of the IA are that the evolution of the parameter error covariance matrix $\mathbf{K}_k$ can be computed and thus the learning curves of the algorithm can be derived. With this knowledge also comes the steady-state values and furthermore the derivation of practical step-size bounds.

In this paper the classic stochastic approaches from Horowitz and Senne [6] and Feuer and Weinstein [7] are pursued and extended in numerous directions, relying on an MA driving process similar to [11]. Symmetric matrices as they appear in the form of the parameter error vector covariance matrix $\mathbf{K}_k$ can be decomposed into two complementary subspaces [12], i.e.,

$$\begin{aligned}\mathbf{K}_k &= b_0\mathbf{I} + b_1\mathbf{R}_{\text{uu}} + ... + b_{M-1}\mathbf{R}_{\text{uu}}^{M-1} + \mathbf{K}_k^{\perp} \\ &= P(\mathbf{R}_{\text{uu}}) + \mathbf{K}_k^{\perp}. \quad (5)\end{aligned}$$

Here, $P(\mathbf{R}_{\text{uu}})$ denotes a polynomial in $\mathbf{R}_{\text{uu}}$ and $\mathbf{K}_k^{\perp}$ is an element of its orthogonal complement. It turns out that only members in subspace $P(\mathbf{R}_{\text{uu}})$ contribute to the error performance measures (see (41) for mismatch: $\text{tr}[\mathbf{K}_k\mathbf{R}_{\text{uu}}^0]$, see (42) for misadjustment: $\text{tr}[\mathbf{K}_k\mathbf{R}_{\text{uu}}]$) of the algorithm as only terms of $\text{tr}[\mathbf{K}_k\mathbf{R}_{\text{uu}}^l]$ are of interest while the complementary part $\text{tr}[\mathbf{K}_k^{\perp}\mathbf{R}_{\text{uu}}^l] = 0$ and thus does not contribute to the performance measures. In Section 2 it is demonstrated that an initial parameter error vector covariance matrix $\mathbf{K}_0 = P(\mathbf{R}_{\text{uu}})$ is forced by the LMS algorithm to remain a member of the modal space of $\mathbf{R}_{\text{uu}}$. However this rule is not true in a strict sense for arbitrary driving processes and requires some mild approximations to make it a more general statement. In Section 3 our considerations are complemented by adding noise terms

and finally all elements are merged to a strong statement about a large class of linearly filtered random processes of moving average type. This class certainly includes linearly filtered IID processes but even some particular statistically dependent terms can also be included so that SIRPs are covered as well. A crucial parameter to describe dynamical as well as stability behavior turns out to be the joint fourth order moment $m_{\mathrm{x}}^{(2,2)} = E[x_k^2 x_l^2]$; for $l \neq k$ of the corresponding decorrelated (white)[1] driving process. Some conclusions in Section 4 round out the paper.

The notation $A[x_k]$ is used to describe a linear operator on a scalar input and $A[\mathbf{x}_k]$ on a vector input. As the linear operator $A[\cdot]$ in our contribution is limited to a linear time-invariant filter, it can equivalently be described by a convolution $A[x_k] = \sum_{m=0}^{P} a_m x_{k-m}$ with the coefficients a_m describing the impulse response of the filter. Consequently, $A[\mathbf{x}_k] = \sum_{m=0}^{P} a_m \mathbf{x}_{k-m}$. Equivalently, such convolution can be described by a linear transformation applying an upperright Toeplitz matrix $\mathbf{A} \in R^{M\times(M+P)}$ to an input vector $\mathbf{x}_k \in R^{(M+P)\times 1}$. In this case the output vector $\mathbf{u}_k = \mathbf{A}\mathbf{x}_k$ is of dimension $R^{M\times 1}$. If the vector $\mathbf{x}_k = [x_k, x_{k-1}, \ldots x_{k-M-P+1}]^T$ exhibits a shift property, so does the corresponding output $\mathbf{u}_k = [u_k, u_{k-1}, \ldots u_{k-M+1}]^T$. The variable x_k denotes a white process while u_k denotes the corresponding filtered process throughout this paper. Furthermore, two other linear operators on square matrices will be used: 1) $\Lambda = \mathrm{diag}[\mathbf{L}]$ on a matrix $\mathbf{L}$ results in a diagonal matrix Λ whose diagonal entries are identical to the diagonal of $\mathbf{L}$ and all other entries are zero; 2) $\mathrm{tr}[\mathbf{L}]$ indicates the trace of the matrix, i.e., $\mathrm{tr}[\mathbf{L}] = \sum_{m=1}^{M} \mathbf{L}_{mm}$.

2. MODAL SPACE OF THE LMS ALGORITHM

Let us consider the classical LMS analysis [5–7] utilizing the IA as stated in the introduction. For simplicity, noise is ignored at this point as interest is only in the evolution of $\mathbf{K}_k$. In a first step the following homogeneous equation

$$\begin{aligned}\mathbf{K}_1 &= E[(\mathbf{I}-\mu\mathbf{u}_0\mathbf{u}_0^T)\mathbf{K}_0(\mathbf{I}-\mu\mathbf{u}_0\mathbf{u}_0^T)^T] \\ &= \mathbf{K}_0 - \mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}_0 - \mu\mathbf{K}_0\mathbf{R}_{\mathrm{uu}} \\ &\quad +\mu^2 E[\mathbf{u}_0\mathbf{u}_0^T\mathbf{K}_0\mathbf{u}_0\mathbf{u}_0^T]\end{aligned} \tag{6}$$

is obtained to illustrate the behavior. Let us start with two examples.

Example 1: Assume first a specific solution for a Gaussian random process and for $\mathbf{K}_0 = \mathbf{R}_{\mathrm{uu}}^0 = \mathbf{I}$, that is $\mathbf{K}_0$ is member of modal space $\mathscr{R}_{\mathrm{u}}$ of $\mathbf{R}_{\mathrm{uu}}$. If it is for example assumed that initial parameter estimate $\mathbf{w}_0 = \mathbf{0}$ and an average over many possible systems $\mathbf{w}$ is performed, $\mathbf{K}_0 = E[\mathbf{w}\mathbf{w}^T] = \mathbf{I}$ can be a realistic assumption. If on the other hand a-priori knowledge on the set of systems is present, other values may be more realistic. In the first step

$$\mathbf{K}_1 = \mathbf{I} - 2\mu\mathbf{R}_{\mathrm{uu}} + \mu^2(2\mathbf{R}_{\mathrm{uu}}^2 + \mathbf{R}_{\mathrm{uu}}\mathrm{tr}[\mathbf{R}_{\mathrm{uu}}]), \tag{7}$$

is obtained that is $\mathbf{K}_1$ is a second order polynomial in $\mathbf{R}_{\mathrm{uu}}$ and thus in the modal space of $\mathbf{R}_{\mathrm{uu}}$. Assume now that $\mathbf{K}_k$ develops into an arbitrary polynomial in $\mathbf{R}_{\mathrm{uu}}$. How does it change from time instant k to $k+1$?

$$\begin{aligned}\mathbf{K}_{k+1} &= \mathbf{K}_k - 2\mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k + \mu^2(2\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k\mathbf{R}_{\mathrm{uu}} \\ &\quad +\mathbf{R}_{\mathrm{uu}}\mathrm{tr}[\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k]).\end{aligned} \tag{8}$$

In other words, it remains a polynomial in $\mathbf{R}_{\mathrm{uu}}$. The same is in fact true if $\mathbf{K}_0$ is any polynomial in $\mathbf{R}_{\mathrm{uu}}$. It can thus be concluded that the LMS update equation under a real-valued Gaussian process forces the parameter error vector covariance matrix $\mathbf{K}_0 = P(\mathbf{R}_{\mathrm{uu}})$ to evolve into a polynomial in the modal space of $\mathbf{R}_{\mathrm{uu}}$. Terms of the orthogonal complement are never generated.

[1] The terms *white* and *decorrelated* will be used interchangeably in the following.

Example 2: Let us now assume that the initial covariance matrix is entirely from the orthogonal complement $\mathscr{R}_{\mathrm{u}}^{\perp}$, that is $\mathbf{K}_0 = \mathbf{K}^{\perp}$. In the first step

$$\begin{aligned}\mathbf{K}_1 &= \mathbf{K}^{\perp} - \mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}^{\perp} - \mu\mathbf{K}^{\perp}\mathbf{R}_{\mathrm{uu}} \\ &\quad +\mu^2 E[\mathbf{u}_0\mathbf{u}_0^T\mathbf{K}^{\perp}\mathbf{u}_0\mathbf{u}_0^T] = \mathbf{K}_1^{\perp}\end{aligned} \tag{9}$$

is obtained. As $\mathrm{tr}[\mathbf{K}_1]$ is of interest, $\mathrm{tr}[\mathbf{K}_1^{\perp}] = 0$ is found. Thus all terms originating from $\mathbf{K}^{\perp}$ have no impact on $\mathrm{tr}[\mathbf{K}_1]$ (or on $\mathrm{tr}[\mathbf{K}_1\mathbf{R}_{\mathrm{uu}}]$). Now in the next step:

$$\begin{aligned}\mathbf{K}_2 &= \mathbf{K}_1 - \mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}_1 - \mu\mathbf{K}_1\mathbf{R}_{\mathrm{uu}} + \mu^2 E[\mathbf{u}_1\mathbf{u}_1^T\mathbf{K}_1\mathbf{u}_1\mathbf{u}_1^T] \\ &= \mathbf{K}_1^{\perp} - \mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}_1^{\perp} - \mu\mathbf{K}_1^{\perp}\mathbf{R}_{\mathrm{uu}} + 2\mu^2\mathbf{R}_{\mathrm{uu}}\mathbf{K}_1^{\perp}\mathbf{R}_{\mathrm{uu}} \\ &= \mathbf{K}_2^{\perp}.\end{aligned} \tag{10}$$

Part $\mathbf{K}_1^{\perp}$ from the orthogonal complement space thus only contributes to this space as $\mathbf{K}_2^{\perp}$ but has no influence in the modal space of $\mathbf{R}_{\mathrm{uu}}$. Thus, any component from the orthogonal complement will remain there and will not generate a component in the modal space of $\mathbf{R}_{\mathrm{uu}}$.

A general $\mathbf{K}_0$ will be a linear combination as shown in (5). Take for example a fixed system $\mathbf{w}$ to be identified. In this case $\mathbf{K}_0 = \mathbf{w}\mathbf{w}^T$. This value can be decomposed into $P(\mathbf{R}_{\mathrm{uu}})$ in the modal space of $\mathbf{R}_{\mathrm{uu}}$ and a component $\mathbf{K}^{\perp}$ from its orthogonal complement. As the polynomial evolves, it will stay in the modal space and contribute to the learning performance terms while the perpendicular terms will not contribute to the algorithm's performance curves under the trace operation. This also allows a description of the evolution of the individual components, starting with $\mathbf{K}_k = \mathbf{K}_k^{\parallel} + \mathbf{K}_k^{\perp}$, with $\mathbf{K}_k^{\parallel} \in \mathscr{R}_{\mathrm{u}}$ and $\mathbf{K}_k^{\perp} \in \mathscr{R}_{\mathrm{u}}^{\perp}$ a set of homogeneous equations is obtained

$$\begin{aligned}\mathbf{K}_{k+1}^{\parallel} &= \mathbf{K}_k^{\parallel} - \mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k^{\parallel} - \mu\mathbf{K}_k^{\parallel}\mathbf{R}_{\mathrm{uu}} \\ &\quad +\mu^2(2\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k^{\parallel}\mathbf{R}_{\mathrm{uu}} + \mathbf{R}_{\mathrm{uu}}\mathrm{tr}[\mathbf{K}_k^{\parallel}\mathbf{R}_{\mathrm{uu}}]).\end{aligned} \tag{11}$$

$$\begin{aligned}\mathbf{K}_{k+1}^{\perp} &= \mathbf{K}_k^{\perp} - \mu\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k^{\perp} - \mu\mathbf{K}_k^{\perp}\mathbf{R}_{\mathrm{uu}} \\ &\quad +2\mu^2\mathbf{R}_{\mathrm{uu}}\mathbf{K}_k^{\perp}\mathbf{R}_{\mathrm{uu}},\end{aligned} \tag{12}$$

which in turn allows the formulation of a first statement for Gaussian driving processes.

Lemma 2.1 *Assume driving process $\mathbf{u}_k = \mathbf{A}\mathbf{x}_k$, a linearly filtered white Gaussian process $\mathbf{x}_k$ with $E[\mathbf{x}_k\mathbf{x}_k^T] = \mathbf{I}_{M+P}$, and $\mathbf{A}$ an upperright Toeplitz matrix for linearly filtering. Under the IA the initial parameter error vector covariance matrix $\mathbf{K}_0$ of the LMS algorithm evolves into 1) a polynomial in $\mathbf{A}\mathbf{A}^T = \mathbf{R}_{\mathrm{uu}}$ of the modal space of $\mathbf{R}_{\mathrm{uu}}$, solely responsible for the mismatch and the misadjustment of the algorithm and 2) a part in its orthogonal complement that has no impact on the performance measures.*

As such examples are rather intuitive for the particular case of a Gaussian driving process, (spherically invariant process as a generalization of Gaussian processes can be included straightforwardly), it is of interest what can be said about larger classes of driving processes. To achieve this goal, a few considerations with respect to the driving process are required.

Driving Process: The properties of Lemma 2.1 are not only maintained by Gaussian random processes but by a much larger class of driving processes. It will be shown that these properties hold for random processes that are constructed by a linearly filtered white, zero mean random process $u_k = A[x_k] = \sum_{m=0}^{P} a_m x_{k-m}$, whose only conditions are that:

Driving Process Assumptions (A1):

$$
\begin{aligned}
m_{\mathrm{x}}^{(2)} &= E[x_k^2] = 1 && (13)\\
m_{\mathrm{x}}^{(2,2)} &= E[x_k^2 x_l^2] \leq c_2 < \infty; k \neq l && (14)\\
m_{\mathrm{x}}^{(4)} &= E[x_k^4] \leq c_3 < \infty && (15)\\
m_{\mathrm{x}}^{(1,1,1,1)} &= E[x_k x_l x_m x_n] = 0; k \neq l \neq m \neq n && (16)\\
m_{\mathrm{x}}^{(2,1,1)} &= E[x_k^2 x_m x_n] = 0; k \neq m \neq n && (17)\\
m_{\mathrm{x}}^{(1,3)} &= E[x_k x_l^3] = 0; k \neq l && (18)\\
m_{\mathrm{x}} &= E[x_k] = 0. && (19)
\end{aligned}
$$

The last four conditions (16)-(19) are listed here for completeness. They exclude processes that do not have a zero mean in some sense and have been assumed in most of the literature, even though not often explicitly mentioned. Linearly filtering such processes will preserve the zero mean properties (16)-(19). These processes include certainly real-valued Gaussian and SIRP ($3m_{\mathrm{x}}^{(2,2)} = m_{\mathrm{x}}^{(4)} = 3$), as well as IID processes $\left(m_{\mathrm{x}}^{(2,2)} = \left(m_{\mathrm{x}}^{(2)}\right)^2\right)$. Constructing vectors $\mathbf{x}_k = [x_k, x_{k-1}, ..., x_{k-N+1}]^T$ the following second and forth order expressions are found:

$$
\begin{aligned}
E[\mathbf{x}_k \mathbf{x}_k^T] &= \mathbf{I}_N, && (20)\\
E[\mathbf{x}_k \mathbf{x}_k^T \mathbf{x}_k \mathbf{x}_k^T] &= (m_{\mathrm{x}}^{(4)} + (M-1) m_{\mathrm{x}}^{(2,2)}) \mathbf{I}_N. && (21)
\end{aligned}
$$

Correspondingly the linearly filtered vectors read $\mathbf{u}_k = \mathbf{A}\mathbf{x}_k$ with an upperright Toeplitz matrix $\mathbf{A}$ of dimension $M \times N$. The impulse response of the coloring filter is given by $a_0, a_1, ...a_P$ and appears on every row of $\mathbf{A}$ starting with a_0 on its main diagonal. In general driving process vector $\mathbf{x}_k$ is longer than $\mathbf{u}_k$, depending on the order P of the impulse response[2].

Lemma 2.2 *Assume driving process $u_k = A[x_k]$ to originate from a linearly filtered white random process x_k so that $\mathbf{u}_k = \mathbf{A}\mathbf{x}_k$ with $\mathbf{x}_k^T = [x_k, x_{k-1}, ..., x_{k-N+1}]$, $\mathbf{A}$ denoting an upperright Toeplitz matrix with the correlation filter impulse response and x_k satisfying conditions (13)-(21). Parameter error vector covariance matrix $\mathbf{K}_0 = \mathbf{K}_0^{\parallel} + \mathbf{K}_0^{\perp}$ of the LMS algorithm essentially (with error of order $\mathcal{O}(\mu^2/M)$) evolves into a polynomial in $\mathbf{A}\mathbf{A}^T$ in the modal space of $\mathbf{R}_{\mathrm{uu}}$ while terms in its orthogonal complement $\mathbf{K}^{\perp}$ remain there or die out.*

Note that this formulation may associate that this is only true for linearly filtered processes of moving average (MA) type. As no condition on the order P of such process is imposed, P (and thus $N = M + P$) can become arbitrarily large and thus autoregressive processes (AR) or combinations (ARMA) can be resembled as well (e.g., see [9](Chapter 2.7)).

Proof: The proof proceeds in two steps: first, rewriting (6) for $\mathbf{K}_0 = \mathbf{I}$ to get to know the most important terms and mathematical steps based on a simpler formulation, and then refining the arguments for arbitrary values of $\mathbf{K}_k$ to $\mathbf{K}_{k+1}$.

For $\mathbf{K}_0 = \mathbf{I}$ and recalling that $\mathbf{R}_{\mathrm{uu}} = E[\mathbf{u}_k \mathbf{u}_k^T] = \mathbf{A} E[\mathbf{x}_k \mathbf{x}_k^T] \mathbf{A}^T = \mathbf{A}\mathbf{A}^T$ the following is obtained:

$$
\mathbf{K}_1 = \mathbf{I} - 2\mu \mathbf{A}\mathbf{A}^T + \mu^2 \mathbf{A} E[\mathbf{x}_k \mathbf{x}_k^T \mathbf{A}^T \mathbf{A} \mathbf{x}_k \mathbf{x}_k^T] \mathbf{A}^T. \qquad (22)
$$

On the main diagonal of the $M \times M$ matrix $\mathbf{A}\mathbf{A}^T$ identical elements are found: $\sum_{i=0}^P |a_i|^2$, thus $\mathrm{tr}[\mathbf{A}^T\mathbf{A}] = \mathrm{tr}[\mathbf{A}\mathbf{A}^T] = M \sum_{i=0}^P |a_i|^2$, with P denoting the filter order of the MA process.

[2] Alternatively, the IA can be removed by employing particular processes in which each element of the regression vector $\mathbf{u}_k = [u_{k,1}, u_{k,2}, ..., u_{k,M}]^T$ is generated by a filtered version of individual processes $x_{k,1}, ..., x_{k,M}$. As such processes seem artificial, this approach is not followed here.

Due to properties (13)-(21) of driving process x_k

$$
\begin{aligned}
&E[\mathbf{x}_k \mathbf{x}_k^T \mathbf{L} \mathbf{x}_k \mathbf{x}_k^T]_{ij} && (23)\\
&= \begin{cases} m_{\mathrm{x}}^{(2,2)} (\mathbf{L}_{ij} + \mathbf{L}_{ji}) & ; i \neq j \\ m_{\mathrm{x}}^{(4)} \mathbf{L}_{ii} + m_{\mathrm{x}}^{(2,2)} \sum_{k \neq i} \mathbf{L}_{kk} & ; i = j \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
E[\mathbf{x}_k \mathbf{x}_k^T \mathbf{L} \mathbf{x}_k \mathbf{x}_k^T] &= m_{\mathrm{x}}^{(2,2)} (\mathbf{L} + \mathbf{L}^T) + m_{\mathrm{x}}^{(2,2)} \mathrm{tr}[\mathbf{L}] \mathbf{I}_{M+P}\\
&+ \left(m_{\mathrm{x}}^{(4)} - 3 m_{\mathrm{x}}^{(2,2)}\right) \mathrm{diag}[\mathbf{L}] && (24)
\end{aligned}
$$

is found, where $\mathrm{diag}[\mathbf{L}]$ denotes a diagonal matrix with the diagonal terms of a matrix $\mathbf{L}$ as entries. For spherically invariant random processes (including Gaussian) the term $(m_{\mathrm{x}}^{(4)} - 3m_{\mathrm{x}}^{(2,2)})$ for real-valued signals vanishes and thus the problem can be solved classically. In our particular case $\mathbf{L} = \mathbf{A}^T\mathbf{A} \in \mathbf{R}^{(M+P)\times(M+P)}$ with $\mathrm{tr}[\mathbf{A}^T\mathbf{A}] = M \sum_{i=0}^P |a_i|^2$, $\mathrm{diag}[\mathbf{A}\mathbf{A}^T] = \sum_{i=0}^P |a_i|^2 \mathbf{I}_{M+P} = \mathrm{tr}[\mathbf{A}^T\mathbf{A}]/M \mathbf{I}_{M+P}$. One problematic term remains however: $\mathrm{diag}[\mathbf{A}^T\mathbf{A}]$. At this point the following is proposed with an identity matrix $\mathbf{I}_{M+P}$ of the corresponding dimension:

Approximation A2: $\mathrm{diag}[\mathbf{A}^T\mathbf{A}] \approx \frac{\mathrm{tr}[\mathbf{A}^T\mathbf{A}]}{M} \mathbf{I}_{M+P}$.

Note the approximation would be exact (up to the dimension) if there would be the term $\mathrm{diag}[\mathbf{A}\mathbf{A}^T]$ instead of $\mathrm{diag}[\mathbf{A}^T\mathbf{A}]$. The approximation can be interpreted as replacing each of the diagonal elements of $\mathbf{A}^T\mathbf{A}$ by their average value $\mathrm{tr}[\mathbf{A}^T\mathbf{A}]/M$. Consider the relative difference matrix

$$
\begin{aligned}
\mathbf{\Delta}_{\varepsilon} &= \left(\frac{\mathrm{tr}[\mathbf{A}^T\mathbf{A}]}{M}\right)^{-1} \left[\mathrm{diag}[\mathbf{A}^T\mathbf{A}] - \frac{\mathrm{tr}[\mathbf{A}^T\mathbf{A}]}{M} \mathbf{I}_{M+P}\right]\\
&= \frac{M}{\mathrm{tr}[\mathbf{A}^T\mathbf{A}]} \mathrm{diag}[\mathbf{A}^T\mathbf{A}] - \mathbf{I}_{M+P} && (25)
\end{aligned}
$$

of dimension $(M+P) \times (M+P)$. Its P diagonal terms at the beginning and end of the diagonal remain non-zero while those terms in the middle (whose range can be substantially large if $M \gg P$) are zero. The first P elements on the diagonal are for example given by $\mathbf{\Delta}_{\varepsilon,ii} = -\sum_{m=i}^P |a_m|^2 / \sum_{m=0}^P |a_m|^2; i = 1..P$. It is worth comparing the long filter derivations by Butterweck [13] that exclude border effects at the beginning and ending of the matrices. Our approximation can thus be interpreted along the same lines of approximations, just originating from a different approach. With this error term $\mathbf{\Delta}_{\varepsilon}$, we find

$$
\begin{aligned}
&E[\mathbf{x}_k \mathbf{x}_k^T \mathbf{A}^T \mathbf{A} \mathbf{x}_k \mathbf{x}_k^T] && (26)\\
&= 2 m_{\mathrm{x}}^{(2,2)} \mathbf{A}^T\mathbf{A}\\
&\quad + (m_{\mathrm{x}}^{(2,2)} + (m_{\mathrm{x}}^{(4)} - 3m_{\mathrm{x}}^{(2,2)})/M) \mathrm{tr}[\mathbf{A}^T\mathbf{A}] \mathbf{I}_{M+P}\\
&\quad + (m_{\mathrm{x}}^{(4)} - 3m_{\mathrm{x}}^{(2,2)}) \mathbf{\Delta}_{\varepsilon}\\
&= 2 m_{\mathrm{x}}^{(2,2)} \mathbf{A}^T\mathbf{A} + \gamma_{\mathrm{x}} m_{\mathrm{x}}^{(2,2)} \mathrm{tr}[\mathbf{A}^T\mathbf{A}] \mathbf{I}_{M+P}\\
&\quad + (\gamma_{\mathrm{x}} - 1) m_{\mathrm{x}}^{(2,2)} \mathrm{tr}[\mathbf{A}^T\mathbf{A}] \mathbf{\Delta}_{\varepsilon},
\end{aligned}
$$

is obtained with a newly introduced *pdf-shape* correction value

$$
\gamma_{\mathrm{x}} = 1 + \left(\frac{m_{\mathrm{x}}^{(4)}}{m_{\mathrm{x}}^{(2,2)}} - 3\right) \frac{1}{M}, \qquad (27)
$$

a value that depends on the statistics of process x_k. The term $\frac{m_{\mathrm{x}}^{(4)}}{m_{\mathrm{x}}^{(2,2)}} - 3$ is similar to the excess kurtosis $\frac{E[|x-m_{\mathrm{x}}|^4]}{E[|x-m_{\mathrm{x}}|^2]^2} - 3 = \frac{m_{\mathrm{x}}^{(4)}}{(m_{\mathrm{x}}^{(2)})^2} - 3$. Processes with negative excess kurtosis are often referred to as sub-Gaussian processes while a positive excess kurtosis leads to so-called super-Gaussian processes. This (slightly abused) terminology will be used correspondingly to discriminate the term $\frac{m_{\mathrm{x}}^{(4)}}{m_{\mathrm{x}}^{(2,2)}} - 3$.

Thus sub-Gaussian processes in this sense take on γ_x values smaller than one while super-Gaussian processes have values larger than one. However, it is also noted that our approximation error $\mathbf{\Delta}_\varepsilon$ has an impact only in case $\gamma_x \neq 1$ which vanishes not only for Gaussian pdfs but also with growing filter order M! Note further that the term in the LMS algorithm where Approximation A2 applies, is proportional to μ^2. It thus has no impact for small step-sizes but certainly on the stability bound. A first conclusion thus is that the error on the parameter error vector covariance matrix due to this approximation is of $\mathcal{O}(\mu^2)$. Furthermore, the approximation error term is proportional to $\gamma_x - 1$ that is proportional to $1/M$. Approximation A2 can thus be concluded to cause an error of the parameter error vector covariance matrix of order $\mathcal{O}(\mu^2/M)$. The consequence that the applied approximation is negligible for large filter order M as well as for Gaussian-type processes is reflected in Lemma 2.2 by the wording "essentially". This means that in extreme cases (small M and far away from Gaussian) indeed a very small proportion can leak into the complementary space. At the first update with $\mathbf{K}_0 = \mathbf{I}$

$$\begin{aligned}\mathbf{K}_1 &= \mathbf{I} - 2\mu\mathbf{A}\mathbf{A}^T + 2\mu^2 m_x^{(2,2)}(\mathbf{A}\mathbf{A}^T)^2 \\ &\quad + \mu^2 m_x^{(2,2)}\gamma_x \mathrm{tr}[\mathbf{A}^T\mathbf{A}]\mathbf{A}\mathbf{A}^T + \mathcal{O}(\mu^2/M)\end{aligned} \tag{28}$$

is obtained, a polynomial in $\mathbf{A}\mathbf{A}^T$.

Now the proof starts for general updates from $\mathbf{K}_k$ to $\mathbf{K}_{k+1}$. While the first terms that are linear in μ are straightforward, the quadratic part in μ needs more attention.

$$\begin{aligned}&E[\mathbf{u}_k\mathbf{u}_k^T\mathbf{K}_k\mathbf{u}_k\mathbf{u}_k^T] = \\ &= \mathbf{A}E[\mathbf{x}_k\mathbf{x}_k^T\mathbf{A}^T\mathbf{K}_k\mathbf{A}\mathbf{x}_k\mathbf{x}_k^T]\mathbf{A} \\ &= m_x^{(2,2)}\mathbf{A}\left(2\mathbf{A}^T\mathbf{A}\mathbf{K}_k\mathbf{A}^T\mathbf{A} + \mathbf{A}^T\mathbf{A}\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_k]\right)\mathbf{A}^T \\ &+ (m_x^{(4)} - 3m_x^{(2,2)})\mathbf{A}\mathrm{diag}[\mathbf{A}^T\mathbf{K}_k\mathbf{A}]\mathbf{A}^T.\end{aligned} \tag{29}$$

Here the same approximation method as before in A2 is imposed, that is $\mathrm{diag}[\mathbf{A}^T\mathbf{K}_k\mathbf{A}] \approx \mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_k]\mathbf{I}_{M+P}/M$ resulting in

$$\begin{aligned}E[\mathbf{u}_k\mathbf{u}_k^T\mathbf{K}_k\mathbf{u}_k\mathbf{u}_k^T] &= 2m_x^{(2,2)}\mathbf{A}\mathbf{A}^T\mathbf{K}_k\mathbf{A}^T\mathbf{A} \\ &+ \gamma_x m_x^{(2,2)}\mathbf{A}\mathbf{A}^T\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_k] + \mathcal{O}(\mu^2/M)\end{aligned} \tag{30}$$

and obtaining eventually

$$\begin{aligned}\mathbf{K}_{k+1} &= \mathbf{K}_k - \mu\mathbf{A}\mathbf{A}^T\mathbf{K}_k - \mu\mathbf{K}_k\mathbf{A}\mathbf{A}^T \\ &+ \mu^2 m_x^{(2,2)}\left(2\mathbf{A}\mathbf{A}^T\mathbf{K}_k\mathbf{A}\mathbf{A}^T + \gamma_x\mathbf{A}\mathbf{A}^T\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_k]\right) \\ &\quad + \mathcal{O}(\mu^2/M).\end{aligned} \tag{31}$$

Now this can be split in two parts into modal space $\mathbf{K}^{\|}$ and in its orthogonal complement $\mathbf{K}^{\perp}$ as in (11) and (12) before and the following is obtained:

$$\begin{aligned}\mathbf{K}_{k+1}^{\|} &= \mathbf{K}_k^{\|} - \mu\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\|} - \mu\mathbf{K}_k^{\|}\mathbf{A}\mathbf{A}^T \\ &+ \mu^2 m_x^{(2,2)}\left(2\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\|}\mathbf{A}\mathbf{A}^T + \gamma_x\mathbf{A}\mathbf{A}^T\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\|}]\right) \\ &\quad + \mathcal{O}(\mu^2/M)\end{aligned} \tag{32}$$

$$\begin{aligned}&= \mathbf{K}_k^{\|} - 2\mu\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\|} \\ &+ \mu^2 m_x^{(2,2)}\left(2[\mathbf{A}\mathbf{A}^T]^2\mathbf{K}_k^{\|} + \gamma_x\mathbf{A}\mathbf{A}^T\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\|}]\right) \\ &\quad + \mathcal{O}(\mu^2/M),\end{aligned} \tag{33}$$

$$\begin{aligned}\mathbf{K}_{k+1}^{\perp} &= \mathbf{K}_k^{\perp} - \mu\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\perp} - \mu\mathbf{K}_k^{\perp}\mathbf{A}\mathbf{A}^T \\ &\quad + 2\mu^2 m_x^{(2,2)}\mathbf{A}\mathbf{A}^T\mathbf{K}_k^{\perp}\mathbf{A}\mathbf{A}^T + \mathcal{O}(\mu^2/M).\end{aligned} \tag{34}$$

The consequence of this statement is that the parameter error vector covariance matrix is forced by the driving process to remain only in the modal space of the driving process. This is not only true for its initial values but at every time instant k. The components of the orthogonal complement remain in there or die out. This statement will be addressed next in the context of step-size bounds for stability.□

3. INFLUENCE OF NOISE AND COMPLETE LMS LEARNING BEHAVIOR

So far, additive noise v_k has been neglected in the evolution of $\mathbf{K}_k$. Adding noise in our reference model ($d_k = \mathbf{w}^T\mathbf{u}_k + v_k$) introduces an additional term $\mu^2\mathbf{R}_{uu}\sigma_v^2 = \mu^2\mathbf{A}\mathbf{A}^T\sigma_v^2$ in the evolution of $\mathbf{K}_k$ and thus defines the inhomogeneous equations. Independent of the noise statistics, this additional term lies also in the modal space of the driving process and thus will not change our previous statements, as long as the IA holds. Therefore components of the orthogonal complement die out as long as $0 < \mu < 1/[m_x^{(2,2)}\lambda_{\max}]$, $\lambda_{\max}$ denoting the largest eigenvalue of $\mathbf{R}_{uu}$[3]. As we will see later, the step-size bound for the component $\mathbf{K}^{\|}$ of the modal space is smaller and thus all terms in the orthogonal complement will die out for the step-size range of interest. As terms in the orthogonal complement die out, $\mathbf{K}_\infty = \lim_{k\to\infty}\mathbf{K}_k$ is expected to exist only in the modal space of $\mathbf{R}_{uu}$. Computing the steady-state solution for $k \to \infty$ and omitting the approximation error terms $\mathcal{O}(\mu^2/M)$, simplicity leads to

$$\begin{aligned}\mathbf{K}_\infty &= \mathbf{K}_\infty - 2\mu\mathbf{A}\mathbf{A}^T\mathbf{K}_\infty + \mu^2\sigma_v^2\mathbf{A}\mathbf{A}^T \\ &+ \mu^2 m_x^{(2,2)}\left(2(\mathbf{A}\mathbf{A}^T)^2\mathbf{K}_\infty + \gamma_x\mathbf{A}\mathbf{A}^T\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_\infty]\right),\end{aligned} \tag{35}$$

or equivalently

$$\begin{aligned}&2\mathbf{A}\mathbf{A}^T\mathbf{K}_\infty - 2\mu m_x^{(2,2)}(\mathbf{A}\mathbf{A}^T)^2\mathbf{K}_\infty \\ &\quad - \mu m_x^{(2,2)}\gamma_x\mathrm{tr}[\mathbf{A}\mathbf{A}^T\mathbf{K}_\infty]\mathbf{A}\mathbf{A}^T = \mu\sigma_v^2\mathbf{A}\mathbf{A}^T.\end{aligned} \tag{36}$$

Since $\mathbf{K}_\infty$ exists only in the modal space of $\mathbf{A}\mathbf{A}^T$ diagonalizing both by the same unitary matrix[4] leads to $\mathbf{Q}\mathbf{K}_\infty\mathbf{Q}^T = \Lambda_K$ and $\mathbf{Q}\mathbf{A}\mathbf{A}^T\mathbf{Q}^T = \Lambda_u$. Thus, we have

$$2\Lambda_u\Lambda_K - \mu m_x^{(2,2)}(2\Lambda_u^2\Lambda_K - \gamma_x\Lambda_u\mathrm{tr}[\Lambda_u\Lambda_K]) = \mu\sigma_v^2\Lambda_u. \tag{37}$$

Stacking the diagonal values of the matrices into vectors: $\Lambda_u\mathbf{1} = \lambda_u, \Lambda_K\mathbf{1} = \lambda_K$, $\lambda_u^T = [\lambda_1, \lambda_2, ..., \lambda_M]$, the following is obtained

$$\left[2\Lambda_u - 2\mu m_x^{(2,2)}\Lambda_u^2 - \mu m_x^{(2,2)}\gamma_x\lambda_u\lambda_u^T\right]\lambda_K = \mu\sigma_v^2\lambda_u. \tag{38}$$

resulting in the well-known form [8][Eqn. (3.15)]:

$$\begin{aligned}\lambda_K &= \mu\sigma_v^2\left[2\Lambda_u - 2\mu m_x^{(2,2)}\Lambda_u^2 - \mu m_x^{(2,2)}\gamma_x\lambda_u\lambda_u^T\right]^{-1}\lambda_u \\ &= \beta_\infty\left[2\Lambda_u - 2\mu m_x^{(2,2)}\Lambda_u^2\right]^{-1}\lambda_u,\end{aligned} \tag{39}$$

with the definition

$$\beta_\infty = \frac{2\mu\sigma_v^2}{2 - \mu m_x^{(2,2)}\gamma_x\sum_i \frac{\lambda_i}{1-\mu m_x^{(2,2)}\lambda_i}} \tag{40}$$

[3]The procedure to obtain such result is the same as explained in the following paragraph for $\mathbf{K}^{\|}$, just much simpler as the trace terms do not appear.

[4]Even if A2 is not satisfied and $\mathbf{K}_\infty$ had a component in the orthogonal complement space, the method of applying $\mathbf{Q}$ can be used. Although then $\mathbf{K}_\infty$ is not diagonalized and Λ_K is not of diagonal form, for the performance measures only its diagonal terms are of importance and will be considered later on.

obtained by employing the matrix inversion lemma [9]: $[P(\Lambda_u) + \lambda_u\lambda_u^T]^{-1}\lambda_u = 1/[1+\lambda_u^T P^{-1}(\Lambda_u)\lambda_u]P^{-1}(\Lambda_u)\lambda_u$. The final steady-state system mismatch is thus given by

$$\mathrm{tr}[\mathbf{K}_\infty] \quad = \quad \mathbf{1}^T\lambda_K = \frac{\mu\sigma_v^2\sum_i \frac{1}{1-\mu m_x^{(2,2)}\lambda_i}}{2-\mu m_x^{(2,2)}\gamma_x\sum_i \frac{\lambda_i}{1-\mu m_x^{(2,2)}\lambda_i}} \qquad (41)$$

and the misadjustment

$$\mathscr{M} = \frac{\lambda_u^T\lambda_K}{\sigma_v^2} = \frac{\mu\sum_i \frac{\lambda_i}{1-\mu m_x^{(2,2)}\lambda_i}}{2-\mu m_x^{(2,2)}\gamma_x\sum_i \frac{\lambda_i}{1-\mu m_x^{(2,2)}\lambda_i}}. \qquad (42)$$

The only difference between this and the classic solution for SIRPs [8] is the term γ_x that contains influences of the fourth order moments $m_x^{(4)}$ and $m_x^{(2,2)}$ as well as $m_x^{(2,2)}$ explicitly. After showing several aspects of the LMS solution, a more general statement can be formulated for the transient and steady-state behavior of the algorithm.

Theorem 3.1 *Assuming driving process $u_k = A[x_k]$ to originate from a linearly filtered white random process x_k with properties (13)-(21), any parameter error vector covariance matrix $\mathbf{K}_0$ evolves essentially into two parts: a polynomial in $\mathbf{A}\mathbf{A}^T$, stemming from its decomposition onto the modal space of $\mathbf{R}_{uu}$ and a second part $\mathbf{K}_k^\perp$ of its orthogonal complement. The LMS update affects these two parts independently of each other.*

The proof is straightforward by applying all previous results. In other words, the complementary subspace part $\mathbf{K}_k^\perp$ has no impact on the LMS performance measures and can thus be neglected not only for Gaussian but for a large class of linearly filtered random processes. A consequence of this theorem is that the step-size bound can be derived either from (41) or by Gershgorin's circle theorem from the matrix in (38). The result is identical and conservative:

$$0 < \mu \leq \frac{2}{m_x^{(2,2)}\left(2\lambda_{\max} + \gamma_x\mathrm{tr}[\mathbf{R}_{uu}]\right)}. \qquad (43)$$

Depending on the statistics of the driving process a more or less conservative bound is obtained. It is worth to distinguish sub- and super Gaussian cases. For sub-Gaussian distributions, $\gamma_x < 1$ while for super-Gaussian $\gamma_x > 1$. The step-size bound thus varies with the distribution type more or less by $\mathrm{tr}[\mathbf{R}_{uu}]$ in the bound (43). For SIRPs (and thus Gaussian) distributions as well as for very long filters $\gamma_x = 1$ and thus

$$0 < \mu \leq \frac{2}{3m_x^{(2,2)}\mathrm{tr}[\mathbf{R}_{uu}]} \leq \frac{2}{m_x^{(2,2)}\left(2\lambda_{\max} + \mathrm{tr}[\mathbf{R}_{uu}]\right)}. \qquad (44)$$

This result is identical to (3) for Gaussian processes as $m_x^{(2,2)} = 1$.

Note that the results are conservative. For a statistically white driving process, for example, an exact bound leads to $\mu \leq 2/[m_x^{(2,2)}\mathrm{tr}[\mathbf{R}_{uu}](\gamma_x + 2/M)]$, thus a significantly larger bound and still depending on the distribution by the value of γ_x.

Further note that the components in the orthogonal complement space indeed vanish as argued at the beginning of Section III. Take for example Eqn. (12) or (34) as the evolution of the orthogonal complement. It is straightforward to show that for the given step-size range in (43) or (44) the components $\mathbf{K}_k^\perp$ vanish as long as there is no new components induced by violating Assumption A2.

4. CONCLUSION

In this contribution a stochastic analysis of second order moments in terms of the parameter error covariance matrix has been shown for the LMS algorithm under the large class of linearly filtered, random driving processes. While results were only known for few statistics, in this contribution the large class of linearly filtered white processes with arbitrary statistics is treated. Particularly interesting is that the parameter error covariance matrix is essentially being forced to remain in the modal space of the driving process $\mathbf{R}_{uu}$, independent of the correlation and the pdf of the driving process. In addition to the independence assumption some mild approximation is required for this derivation, causing minor errors only for short filters and pdfs being very different from Gaussian. Even such mild approximation is no longer required once the filter order grows to very large values. All results have been validated by MC simulations but due to limited space will be reported elsewhere.

Acknowledgment

This work has been funded by the NFN SISE project S10609 (National Research Network Signal and Information Processing in Science and Engineering).

REFERENCES

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

[2] A.H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive gradient algorithms via the small gain theorem," in *Proc. SPIE 1995*, San Diego, USA, July 1995, pp. 458–469.

[3] M. Rupp and A.H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Transactions on SP*, vol. 44, no. 6, pp. 1428–1439, Jun. 1996.

[4] A.H. Sayed and M. Rupp, "Robustness issues in adaptive filtering," in *The DSP Handbook*. CRC Press, 1998.

[5] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM J. Res. Develop.*, vol. 16, no. 6, pp. 546–555, 1972.

[6] L.L. Horowitz and K.D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Transactions on Signal Processing*, vol. 29, pp. 722–736, June 1981.

[7] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP–33, no. 1, pp. 222–230, Feb. 1985.

[8] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Transactions on SP*, vol. 41, no. 3, pp. 1149–1160, Mar. 1993.

[9] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice–Hall, Inf. and System Sciences Series, 1986.

[10] R. Nitzberg, "Normalized LMS algorithm degradation due to estimation noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 740 –750, Nov. 1986.

[11] S.C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2863 –2871, Dec. 1995.

[12] M. Rupp and H.-J. Butterweck, "Overcoming the independence assumption in LMS filtering," in *Proc. of Asilomar Conference*, Nov. 2003, pp. 607–611.

[13] H.-J. Butterweck, "A wave theory of long adaptive filters," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 6, pp. 739 –747, June 2001.