

Overcoming the Independence Assumption in LMS Filtering

Markus Rupp¹ and Hans-Juergen Butterweck²

¹Inst. f. Nachrichtentechnik und Hochfrequenztechnik, Technische Universitaet Wien, Gusshausstr. 25-29, A-1040 Vienna, Austria, e-mail: mrupp@nt.tuwien.ac.at

²Eindhoven University of Technology, The Netherlands, e-mail: H.J.Butterweck@tue.nl

ABSTRACT

The learning process of the LMS algorithm remains understood only very poorly. Despite three decades of intensive research, very few results have been found to overcome the classical independence assumption in which the sequence of driving regression vectors is assumed to be statistically independent. While giving relatively precise results for processes of little correlation, the results obtained in other cases are far off from the true values. In this paper, a new approach is taken to investigate the *learning behavior* of the LMS algorithm using much milder conditions than in the classical independence theory. It is shown that our conditions lead to much better results, in particular for correlated driving processes when compared with the classical independence assumption.

1 INTRODUCTION

The independence assumption (IA)[1, 2, 3], i.e., the assumption that consecutive regression vectors in the LMS algorithm are considered statistically independent, leads like a red tape through history of stochastic gradient approaches. Nobody likes this somewhat obscure assumption, however, in practice hardly any other method has been so widely applied.

Many authors have tried loosening the conditions of the IA, using small adaptation gain [4], ODE approaches[5], neglecting statistics[6], finding alternative conditions[7], or computing the entire field of involved random variables [8] for MA and AR driving processes of fixed order. None of them could give a comprehensible answer why the IA works in many cases quite well.

In [9, 10] new recursive approaches (see further ahead in (8)-(10)) were shown to treat the LMS adaptation without imposing the IA. In [9] results for steady-state behavior were derived, which are more precise than those found under the classic IA. In his newest edition Haykin[11] decided thus to no longer present the IA at all and focus on this recursive scheme instead. However, while the latter provides excellent results for the steady state, the method has not been applied to the initial

learning of the LMS algorithm. In many applications, such initial learning is of greatest importance including step-size conditions for fastest convergence and stability.

2 CONCEPT

Consider the LMS updating rule for the $M \times 1$ weight vector \mathbf{w}_k given by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \mathbf{x}_k e_k \quad (1)$$

$$e_k = d(k) - \mathbf{x}_k^T \mathbf{w}_k. \quad (2)$$

For a tapped-delay line structure, the *regression vector* \mathbf{x}_k has the form

$$\mathbf{x}_k = [x(k), x(k-1), \dots, x(k-M+1)]^T, \quad (3)$$

with $x(k)$ denoting the driving process of the adaptive filter. Note that the values $x(k), \dots, x(k-M+1)$ are time shifted values of each other. The corresponding vectors \mathbf{x}_k thus can be constructed from its past value \mathbf{x}_{k-1} and one new element $x(k)$ so that successive vectors are *not statistically independent* as it is typically assumed in the independence assumption. Let now the *desired response* d_k be composed as

$$d(k) = \mathbf{x}_k^T \mathbf{w}_o + v_k, \quad (4)$$

where \mathbf{w}_o represents a reference $M \times 1$ model vector (interpretable as optimal Wiener solution, if the ACF matrix of the random vector process \mathbf{x}_k is regular) and v_k denotes *additive output noise*. Introducing the *weight error vector*

$$\boldsymbol{\epsilon}_k = \mathbf{w}_o - \mathbf{w}_k \quad (5)$$

allows to write (1) as

$$\boldsymbol{\epsilon}_{k+1} = \boldsymbol{\epsilon}_k - \mu \mathbf{x}_k e_k \quad (6)$$

$$= [\mathbf{I} - \mu \mathbf{x}_k \mathbf{x}_k^T] \boldsymbol{\epsilon}_k - \mu \mathbf{x}_k v_k. \quad (7)$$

Of interest is typically the evolution of $\text{trace}(\mathbf{E}[\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T])$ or $\text{trace}(\mathbf{E}[\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T \mathbf{R}])$ since $\text{trace}(\mathbf{E}[\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T \mathbf{R}]) + \sigma_v^2 \approx \mathbf{E}[e_k^2]$. Now note that two different scenarios are of interest. One is the evolution of $\boldsymbol{\epsilon}_k$ for a given but fixed system \mathbf{w}_o , while the other is its behavior for an ensemble of

systems \mathbf{w}_o . The latter is typically used in tracking analyses but also describes the average learning behavior of an ensemble of many constant systems \mathbf{w}_o .

The weight-error updating (7) can equivalently be written in an iterative form[9, 10]. With $\epsilon_k = \alpha_k + \beta_k + \gamma_k + \dots$ etc. the set of equations

$$\alpha_{k+1} = [\mathbf{I} - \mu\mathbf{R}]\alpha_k - \mu\mathbf{x}_k v_k \quad (8)$$

$$\beta_{k+1} = [\mathbf{I} - \mu\mathbf{R}]\beta_k + \mu[\mathbf{R} - \mathbf{x}_k \mathbf{x}_k^T]\alpha_k \quad (9)$$

$$\gamma_{k+1} = [\mathbf{I} - \mu\mathbf{R}]\gamma_k + \mu[\mathbf{R} - \mathbf{x}_k \mathbf{x}_k^T]\beta_k \quad (10)$$

$$\delta_{k+1} = [\mathbf{I} - \mu\mathbf{R}]\dots \quad (11)$$

satisfies (7), provided the iteration converges.

3 BASIC APPROXIMATION THEOREM

In the following, we will show that covariance matrices like $E[\alpha_k \alpha_k^T], E[\beta_k \beta_k^T], \dots$, can be approximated by polynomials in the input autocorrelation matrix \mathbf{R} , thus considerably simplifying the analysis of the LMS learning behavior.

We start with the Cayley Hamilton Theorem, stating that every matrix satisfies its own characteristic equation

$$r_0 \mathbf{I} + r_1 \mathbf{R} + \dots + r_{M-1} \mathbf{R}^{M-1} + \mathbf{R}^M = \mathbf{0}. \quad (12)$$

A direct consequence of this is that every matrix \mathbf{R}^p can be described in terms of a polynomial of maximum order $M - 1$, no matter how large p is.

Unfortunately, arbitrary matrices \mathbf{A} like $E[\alpha_k \alpha_k^T], E[\beta_k \beta_k^T]$, occurring in the LMS-update equations cannot be expected to be a polynomial in \mathbf{R} . If this would be the case, the following analysis would be strongly simplified and the Matrix-Riccati equations could be solved in closed form. However, it is possible to *approximate* any \mathbf{A} by a polynomial in \mathbf{R} , i.e., by the form

$$\mathbf{A} \approx e_0 \mathbf{I} + e_1 \mathbf{R} + \dots + e_{M-1} \mathbf{R}^{M-1}, \quad (13)$$

that is, by a linear combination of the vector

$$\mathbf{e} = [e_0, e_1, \dots, e_{M-1}]^T$$

with the parameters e_i chosen such that a particular distance between \mathbf{A} and the polynomial in \mathbf{R} is minimized. For example, \mathbf{A} can represent the initial (i.e. observed at $k = 0$) error covariance matrix $E[\epsilon_0 \epsilon_0^T]$ or partial terms of the form $E[\alpha_0 \alpha_0^T]$ or $E[\beta_0 \beta_0^T]$. The following theorem helps in evaluating the quality of such approximation.

Theorem 1: *Given symmetric matrices \mathbf{R} and \mathbf{A} , \mathbf{A} can be decomposed uniquely into two parts $\mathbf{A} = \mathbf{A}_S + \mathbf{A}_T$ so that \mathbf{A}_S is a polynomial in \mathbf{R} and $\text{trace}(\mathbf{A}_T \mathbf{R}^p) = 0$, for every value of p .*

Proof: Let an $(M - 1)$ th order polynomial in $\mathbf{R} \in \mathbf{R}^{M \times M}$ approximate some matrix $\mathbf{A} \in \mathbf{R}^{M \times M}$ in the sense that the minimum

$$\min_{\gamma_i} \left\| \mathbf{A} - \sum_{i=0}^{M-1} \gamma_i \mathbf{R}^i \right\|$$

is attained for some matrix norm. If we select the Frobenius norm, the problem simplifies to:

$$\min_{\gamma_i} \left\| \mathbf{A} - \sum_{i=0}^{M-1} \gamma_i \mathbf{R}^i \right\|_F^2 = \min_{\gamma_i} \sum_{k,l} \left| \mathbf{A}_{kl} - \sum_{i=0}^{M-1} \gamma_i (\mathbf{R}^i)_{kl} \right|^2. \quad (14)$$

The elements of \mathbf{A} as well as the elements of \mathbf{R}^i can be organized in a column vector of dimension $M^2 \times 1$, yielding $\mathbf{a} = \text{vec}(\mathbf{A})$ and $\mathbf{r}^{(i)} = \text{vec}(\mathbf{R}^i)$, respectively. The approximation problem then reads:

$$\min_{\gamma_i} \left\| \mathbf{A} - \sum_{i=0}^{M-1} \gamma_i \mathbf{R}^i \right\|_F^2 = \min_{\gamma_i} \left\| \mathbf{a} - \sum_{i=0}^{M-1} \gamma_i \mathbf{r}^{(i)} \right\|_2^2. \quad (15)$$

The set of vectors $\{\mathbf{r}^{(i)}\}$ span a sub-space of \mathbf{R}^{M^2} , say $S = \text{span}(\{\mathbf{r}^{(i)}\})$. The part $\mathbf{a}_S = \text{vec}(\mathbf{A}_S)$ of \mathbf{a} that can be represented by the set $\{\mathbf{r}^{(i)}\}$ is assumed to lie also in this subspace while the remainder \mathbf{a}_T lies in the complementary subspace T , such that $T \cup S = \mathbf{R}^{M^2}$ and $T \cap S = \{\mathbf{0}\}$.

Problem (15) is quadratic in the unknowns γ_i and can be solved by differentiation with respect to the γ_i . Arranging the values γ_i in a vector $\boldsymbol{\gamma}$ results in the following condition:

$$\begin{bmatrix} \mathbf{a}_T^T \mathbf{r}^{(0)} \\ \mathbf{a}_T^T \mathbf{r}^{(1)} \\ \vdots \\ \mathbf{a}_T^T \mathbf{r}^{(M-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}^{(0)T} \mathbf{r}^{(0)} & \mathbf{r}^{(0)T} \mathbf{r}^{(1)} & \dots & \mathbf{r}^{(0)T} \mathbf{r}^{(M-1)} \\ \mathbf{r}^{(1)T} \mathbf{r}^{(0)} & \mathbf{r}^{(1)T} \mathbf{r}^{(1)} & \dots & \mathbf{r}^{(1)T} \mathbf{r}^{(M-1)} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{r}^{(M-1)T} \mathbf{r}^{(0)} & \mathbf{r}^{(M-1)T} \mathbf{r}^{(1)} & \dots & \mathbf{r}^{(M-1)T} \mathbf{r}^{(M-1)} \end{bmatrix}. \quad (16)$$

Assuming that the right-hand matrix is regular, a unique solution for $\boldsymbol{\gamma}$ exists, leading to $\mathbf{A}_S = \sum \gamma_i \mathbf{R}^i$ as an approximation for \mathbf{A} .

Since expression (15) is quadratic in the variables γ_i , the error, or remainder term \mathbf{a}_T lies in an orthogonal subspace complementary to the one formed by the vectors $\mathbf{r}^{(i)}$, according to $\mathbf{a}_T^T \mathbf{r}^{(i)} = 0$ for every $i = 0, 1, \dots, M-1$. Now associate an $M \times M$ matrix \mathbf{A}_T with the corresponding $M^2 \times 1$ remainder $\mathbf{a}_T = \text{vec}(\mathbf{A}_T)$, so that $\mathbf{a} = \mathbf{a}_S + \mathbf{a}_T$ and $\mathbf{A} = \mathbf{A}_S + \mathbf{A}_T$. Then we can conclude that

$$\text{trace}(\mathbf{R}^p \mathbf{A}_T \mathbf{R}^q) = 0 \quad (17)$$

for arbitrary p, q . The proof is evident: $\text{trace}(\mathbf{A}_T \mathbf{R}^m) = \mathbf{a}_T^T \mathbf{r}^{(m)} = 0$. For values $m > M - 1$ the matrices can

be written in polynomial form of \mathbf{R} with smaller values of exponents and for the trace operator we have: $\text{trace}(\mathbf{ABC}) = \text{trace}(\mathbf{BCA})$. Thus, the trace of such a solution always yields zero even if weighted by \mathbf{R}^q . Note in particular, for $p = 0$, we obtain $\text{trace}(\mathbf{A}_T) = 0$. This result will be used frequently in the following sections.

4 ANALYSIS OF THE FIRST-ORDER TERM

Now consider the covariance matrix $E[\alpha_k \alpha_k^T]$ stemming from (8). This matrix is governed by the updating relation

$$E[\alpha_{k+1} \alpha_{k+1}^T] = [\mathbf{I} - \mu \mathbf{R}] E[\alpha_k \alpha_k^T] [\mathbf{I} - \mu \mathbf{R}] + \mu^2 \sigma_V^2 \mathbf{R}, \quad (18)$$

well-known as Matrix-Ricatti equation with some initial value $E[\alpha_0 \alpha_0^T] = \mathbf{A}_0$. The form (18) applies if and only if the cross term $E[v_k \mathbf{x}_k \alpha_k^T]$ vanishes, which occurs under the following mild assumption:

Assumption 1: The additive noise sequence v_k is statistically independent of the regression vector \mathbf{x}_k and is a zero-mean process with a white spectrum.

Proof of (18): The vectors α_k satisfying (8) are statistically only dependent upon previous products $\{v_{k-1} \mathbf{x}_{k-1}, v_{k-2} \mathbf{x}_{k-2}, \dots\}$. Due to Assumption 1 these are statistically independent of $\{v_k \mathbf{x}_k\}$ so that $E[v_k \mathbf{x}_k \alpha_k^T] = \mathbf{0}$.

A further consequence is that the Matrix-Ricatti equation (18) can now be solved explicitly. While in this case the solution could be given immediately, the analysis of further terms including $\{\beta, \gamma, \dots\}$ are much more involved and require a parametric description in terms of polynomials in \mathbf{R} . We therefore will start using this parametric formulation already for the first-order term $E[\alpha_0 \alpha_0^T] = a_0 \mathbf{I} + a_1 \mathbf{R} + \dots + a_{M-1} \mathbf{R}^{M-1}$, and later we will use $E[\beta_0 \beta_0^T] = b_0 \mathbf{I} + b_1 \mathbf{R} + \dots + b_{M-1} \mathbf{R}^{M-1}$, etc. correspondingly.

Lemma 1: Under Assumption 1, the Matrix-Ricatti Equation (18) of dimension $M \times M$ can be solved explicitly if $E[\alpha_0 \alpha_0^T]$ is a polynomial in \mathbf{R} . The solution can be described in parametrical form with M parameters.

Proof: If $E[\alpha_0 \alpha_0^T]$ is a polynomial in \mathbf{R} , the right-hand side of (18) and thus $E[\alpha_1 \alpha_1^T]$ is also a polynomial in \mathbf{R} . Continuation of this process preserves the same property for all consecutive $E[\alpha_k \alpha_k^T]$. Furthermore, due to the theorem of Cayley Hamilton, every exponent larger than $M - 1$ can be mapped into a set of terms in exponents ranging from zero to $M - 1$. Thus, (18) can be described as a mapping of a set of M coefficients $a_{i,k}$, $i = 0..M - 1$ to new coefficients $a_{i,k+1}$, $i = 0..M - 1$. Alternatively, the coefficients $a_{i,k}$ can be arranged into a vector

$$\mathbf{a}_k = [a_{0,k}, a_{1,k}, \dots, a_{M-1,k}]^T. \quad (19)$$

The Matrix-Ricatti equation (18) then describes how the vector \mathbf{a}_k evolves into the vector \mathbf{a}_{k+1} .

Lemma 2: If only $\text{trace}(E[\alpha_0 \alpha_0^T])$ (or terms of the form $\text{trace}(E[\alpha_0 \alpha_0^T] \mathbf{R}^p)$) is of interest, Assumption 1 is sufficient for the explicit solution of (18), i.e., the solution is not restricted for $E[\alpha_0 \alpha_0^T]$ to be a polynomial in \mathbf{R} .

Proof: Assume the matrix $E[\alpha_0 \alpha_0^T] = \mathbf{A}_0$ can be split into two terms $\mathbf{A}_0 = \mathbf{A}_{S,0} + \mathbf{A}_{T,0}$, $\mathbf{A}_{S,0} \in S$, $\mathbf{A}_{T,0} \in T$. After the first recursion step (18) we obtain for $k = 0$:

$$\mathbf{A}_1 = [\mathbf{I} - \mu \mathbf{R}] \mathbf{A}_0 [\mathbf{I} - \mu \mathbf{R}] + \mu^2 \sigma_V^2 \mathbf{R}. \quad (20)$$

Separating \mathbf{A} in its two components, we find

$$\mathbf{A}_{S,1} = [\mathbf{I} - \mu \mathbf{R}] \mathbf{A}_{S,0} [\mathbf{I} - \mu \mathbf{R}] + \mu^2 \sigma_V^2 \mathbf{R}, \quad (21)$$

$$\mathbf{A}_{T,1} = [\mathbf{I} - \mu \mathbf{R}] \mathbf{A}_{T,0} [\mathbf{I} - \mu \mathbf{R}], \quad (22)$$

for the transition at time instant $k = 0$ to $k = 1$, followed by similar transitions from k to $k + 1$. Thus, the part $\mathbf{A}_{S,k} \in S$ of \mathbf{A}_k remains in S while the complementary part remains in T . For the parts in T , it was already proven before, that terms in the trace of $\mathbf{A}_{T,k}$ remain zero: $\text{trace}([\mathbf{I} - \mu \mathbf{R}] \mathbf{A}_{T,0} [\mathbf{I} - \mu \mathbf{R}]) = \mathbf{0}$. Thus, if only trace terms of \mathbf{A}_k are of interest, as is the case here, it is sufficient to consider only the evolution of the polynomial approximation of \mathbf{A}_k , i.e., $\mathbf{A}_{S,k}$. Accordingly, in what follows the suffix S is omitted.

Lemma 3: The evolution of the parametric description \mathbf{a}_k for $E[\alpha_k \alpha_k^T] = a_{0,k} \mathbf{I} + a_{1,k} \mathbf{R} + \dots + a_{M-1,k} \mathbf{R}^{M-1}$ is governed by:

$$\mathbf{a}_{k+1} = [\mathbf{I} + 2\mu \mathbf{L} + \mu^2 \mathbf{Q}] \mathbf{a}_k + \mu^2 \sigma_V^2 \mathbf{e}_1 = \mathbf{A}^2 \mathbf{a}_k + \mu^2 \sigma_V^2 \mathbf{e}_1, \quad (23)$$

i.e., a set of linear equations containing an $M \times M$ matrix \mathbf{A}^2 and a driving vector $\mathbf{e}_1 = [0, 1, 0, 0, \dots, 0]^T$.

Proof: Since the initial covariance matrix can be written as $E[\alpha_0 \alpha_0^T] = a_{0,0} \mathbf{I} + a_{1,0} \mathbf{R} + \dots + a_{M-1,0} \mathbf{R}^{M-1}$, the left- and right-hand factor $(\mathbf{I} - \mu \mathbf{R})$ can equivalently be applied twice from the left. Every time $(\mathbf{I} - \mu \mathbf{R})$ is applied, the vector \mathbf{a}_k is mapped by the matrix \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 & \mu r_0 \\ -\mu & 1 & 0 & \dots & 0 & \mu r_1 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & 0 & -\mu & 1 & 0 & \mu r_{M-3} \\ 0 & \dots & 0 & -\mu & 1 & \mu r_{M-2} \\ 0 & \dots & \dots & 0 & -\mu & 1 + \mu r_{M-1} \end{bmatrix}. \quad (24)$$

Correspondingly, the matrix \mathbf{L} describing the linear

term is given by

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & \dots & \dots & 0 & r_0 \\ -1 & 0 & 0 & \dots & 0 & r_1 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & 0 & -1 & 0 & 0 & r_{M-3} \\ 0 & \dots & 0 & -1 & 0 & r_{M-2} \\ 0 & \dots & \dots & 0 & -1 & r_{M-1} \end{bmatrix} \quad (25)$$

and $\mathbf{Q} = \mathbf{L}^2$. Note that the eigenvalues of $-\mathbf{L}$ are those of \mathbf{R} . Finally, the noise term $\mu^2\sigma_V^2\mathbf{R}$ yields an additive term proportional to \mathbf{R} (thus, only modifying $a_{1,k}$), as represented by the vector $\mathbf{e}_1 = [0, 1, 0, 0, \dots, 0]^T$.

Lemma 4: *The steady-state value of $E[\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^T]$ written in its parametrical form is found as*

$$\mathbf{a}_\infty = \mu^2\sigma_V^2[\mathbf{I} - \mathbf{A}^2]^{-1}\mathbf{e}_1 = -\mu\sigma_V^2[\mathbf{L} + \mu\mathbf{Q}]^{-1}\mathbf{e}_1. \quad (26)$$

Proof: Directly by setting $\mathbf{a}_\infty = \mathbf{a}_{k+1} = \mathbf{a}_k$ in (23).

Note that the behaviour of the vector $\boldsymbol{\alpha}_k$ can be interpreted as the mean behaviour of the LMS system (just like a steepest descent algorithm). It thus has the meaning of the solution in the mean (obtained by invoking the IA) and is as in the standard LMS theory of little practical consequence. It is thus to expect that by using more and more of the terms $\boldsymbol{\beta}, \boldsymbol{\gamma}, \dots$ the solution will finally converge to the true solution reflecting the influence of all higher-order moments as well.

5 ANALYSIS OF HIGHER-ORDER TERMS

Due to the different driving terms, solving the matrix equations for the higher-order terms is not as simple as for the first-order term. The mixing process $v_k\mathbf{x}_k$ is now replaced by $[\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T]\boldsymbol{\alpha}_k, [\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T]\boldsymbol{\beta}_k, \dots$, leading to more sophisticated update rules. As an example we look at the evolution of $\boldsymbol{\beta}_k$ to $\boldsymbol{\gamma}_k$, reading:

$$\begin{aligned} E[\boldsymbol{\gamma}_{k+1}\boldsymbol{\gamma}_{k+1}^T] &= (\mathbf{I} - \mu\mathbf{R})E[\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T](\mathbf{I} - \mu\mathbf{R}) \quad (27) \\ &+ \mu^2E[(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)\boldsymbol{\beta}_k\boldsymbol{\beta}_k^T(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)] \\ &+ \mu E[(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)\boldsymbol{\beta}_k\boldsymbol{\gamma}_k^T(\mathbf{I} - \mu\mathbf{R})] \\ &+ \mu E[(\mathbf{I} - \mu\mathbf{R})\boldsymbol{\gamma}_k\boldsymbol{\beta}_k^T(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)]. \end{aligned}$$

Note that according to (8)-(10) the term $\boldsymbol{\beta}_k$ is a filtered version of $\boldsymbol{\alpha}_k$, $\boldsymbol{\gamma}_k$ is a filtered version of $\boldsymbol{\beta}_k$ and so on. Note also that the driving terms are weighted by $(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)$ causing the signals of the next higher order to be smaller than the previous ones. The following assumption is a manifestation of this observation.

Assumption 2: The higher-order terms $\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \dots$ are statistically independent of the driving process \mathbf{x}_k and of each other.

Note that, if compared to the classic IA such an assumption is a much weaker requirement than requiring the independence of the entire $\boldsymbol{\epsilon}_k$ to the regression vector \mathbf{x}_k . Also, the assumption only effects the driving terms and not the recursion terms. This is a major difference to the classical independence assumption. If an error is made by assuming statistical independence of the higher-order terms and the original driving process \mathbf{x}_k then such error impacts the solution only linearly. Also, since the first-order term carries the most energy and is already computed correctly, the remaining higher-order terms will only cause a small error. In summary such an IA is a very weak condition compared to the previously imposed classical IA.

As a consequence of Assumption 2 the two last terms in (27) disappear, thus leaving the simplified expression:

$$\begin{aligned} E[\boldsymbol{\gamma}_{k+1}\boldsymbol{\gamma}_{k+1}^T] &= (\mathbf{I} - \mu\mathbf{R})E[\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T](\mathbf{I} - \mu\mathbf{R}) \quad (28) \\ &+ \mu^2E[(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)E[\boldsymbol{\beta}_k\boldsymbol{\beta}_k^T](\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)]. \end{aligned}$$

In order to solve this equation, some assumption needs to be made on the driving process \mathbf{x}_k . We will, like most authors, follow the (real-valued) Gaussian assumption. We note, however, that other particular processes can also be applied and are suitable for this form of analysis (for example spherically invariant processes[12], complex valued Gaussian, white processes[13]). Note that no matter what random process is applied, the resulting driving term can again be decomposed into two components, one lying in S , the other in T , thus only the part in S contributes to the solution.

Assumption 3: The driving process \mathbf{x}_k follows a real-valued Gaussian statistic.

Utilizing Assumption 3 the driving process can also be computed explicitly. In such case, the driving term in (28) is given by

$$\begin{aligned} E[(\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)E[\boldsymbol{\beta}_k\boldsymbol{\beta}_k^T](\mathbf{R} - \mathbf{x}_k\mathbf{x}_k^T)] \quad (29) \\ = \mathbf{R}E[\boldsymbol{\beta}_k\boldsymbol{\beta}_k^T]\mathbf{R} + \mathbf{R}\text{trace}(\mathbf{R}E[\boldsymbol{\beta}_k\boldsymbol{\beta}_k^T]). \end{aligned}$$

This result allows for computing the steady-state behaviour. Details will be published elsewhere.

6 SIMULATION RESULTS

In the following simulations, an FIR system of order $M = 10$ is identified by the LMS algorithm of identical order. In the first two experiments, the system is changed on every run randomly with an ACF matrix $E[\mathbf{w}_o\mathbf{w}_o^T]$ being a specific instantiation of a polynomial in \mathbf{R} . The simulations are averaged over 20 different system instantiations.

In Figure 1 $E[\mathbf{w}_o\mathbf{w}_o^T] = \mathbf{I}$, and in Figure 2 $E[\mathbf{w}_o\mathbf{w}_o^T] = \mathbf{R}$. The driving process is a recursively filtered white

noise Gaussian Process (AR(1)) with recursion coefficient $r = 0.5$. The step-size was selected to be $\mu = 0.1/M = 0.01$, i.e., rather small. In Figure 1 an excellent match can be observed for the new theory as well as for the independence theory. Both, the convergence as well as the mismatch (expressed in terms of relative errors: $\|\epsilon_k\|_2^2/\|\mathbf{w}_o\|_2^2$) has been predicted very well.

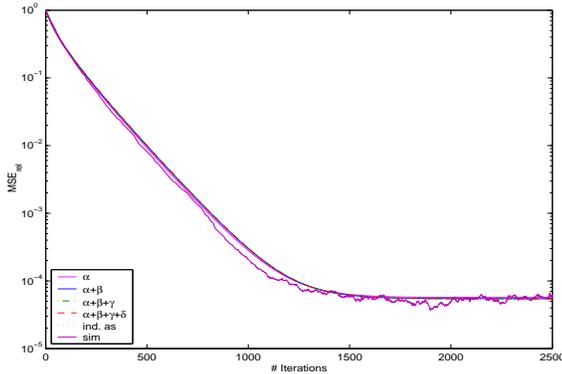


Figure 1: *Relative MSE for $E[\mathbf{w}_o \mathbf{w}_o^T] = \mathbf{I}$.*

In Figure 2 the independence assumption is not capable of predicting the convergence correctly nor the mismatch. The new theory however, predicts perfectly both properties. In both figures the value $E[\alpha_k \alpha_k^T]$ alone gives an excellent prediction of the behavior while utilizing more terms gives hardly a further improvement.

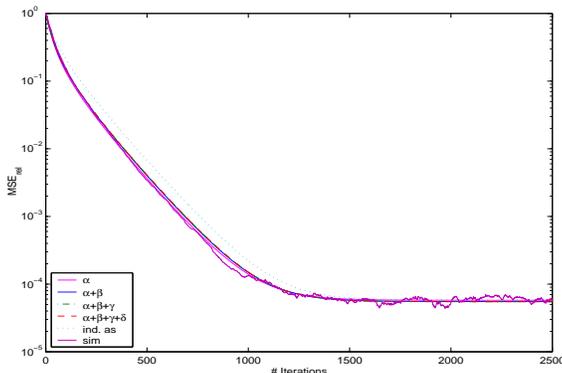


Figure 2: *Relative MSE for $E[\mathbf{w}_o \mathbf{w}_o^T] = \mathbf{R}$.*

In a final experiment depicted in Figure 3 a fixed system with $\mathbf{w}_o = [1, 1, 1, \dots, 1]$ was used for which we approximated the matrix by a polynomial in \mathbf{R} . We selected $r = 0.75$ and $\mu = 0.025$ which is close to the stability bound. The new theory predicts again the convergence curve quite well while the classical IA fails. However, the prediction is not as close as before and the utilization of more iterations gives the impression of deviating from the correct behaviour. Note, however, that the dashed curve is a better approximation in the beginning of the adaptation process.

7 CONCLUSION

A new method to determine the learning curves in adaptive filtering has been introduced. Compared with with

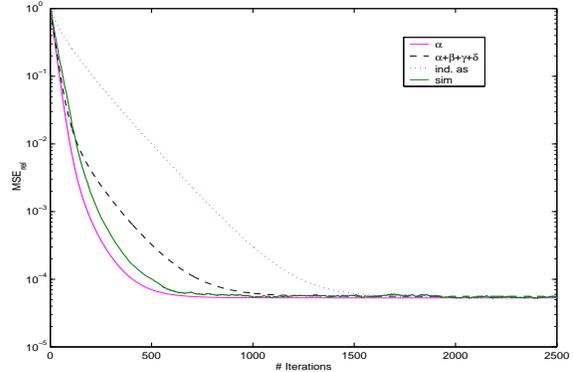


Figure 3: *Relative MSE for $\mathbf{w}_o = [1, 1, \dots, 1]$.*

the traditional method assuming statistical independence of regression vectors our method gives much better results, particularly for correlated processes.

References

- [1] J.E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963–993, 1979.
- [2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, NY: Prentice-Hall, Inc., 1985.
- [3] P.M. Clarkson, *Optimal and Adaptive Signal Processing*, CRC Press, 1993.
- [4] V. Solo, X. Kong, *Adaptive Signal Processing Algorithms, Stability and Performance*, Prentice Hall 95.
- [5] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Transactions on Automatic Control*, vol. 22, pp. 551-575, 1977.
- [6] M. Rupp, A.H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1428-1440, June 1996.
- [7] A.H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley 2003.
- [8] S.C. Douglas, W. Pan, Exact expectation analysis of the LMS adaptive filter, *IEEE Transactions on Signal Processing*, vol. 34, pp. 2863-2871, 1995.
- [9] H.J. Butterweck, "Iterative analysis of the steady-state weight fluctuations in LMS-type adaptive filters," *IEEE Transactions on Signal Processing*, vol. 47, is. 9, pp. 2558-2561, Sept. 1999.
- [10] H.J. Butterweck, "A wave theory of long adaptive filters," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, is. 6, pp. 739 -747, June 2001.
- [11] S. Haykin, *Adaptive Filter Theory*, NJ: Prentice Hall, fourth edition, 2002.
- [12] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Trans. Signal Processing*, vol. SP-41, no. 3, pp. 1149-1160, March 1993.
- [13] S.C. Douglas, M. Rupp *Chapter: Convergence Issues in Adaptive Filters* in *The DSP Handbook*, CRC Press, 1997.