

UNBIASED ADAPTIVE SYSTEM IDENTIFICATION FOR CORRELATED INPUT AND NOISE

Panagiotis Niavis

niavis@ceid.upatras.gr

George V. Moustakides

moustaki@upatras.gr

Department of Electrical & Computer Engineering
University of Patras, Greece

ABSTRACT

We consider the problem of adaptive system identification when the additive noise is colored, following an ARMA model, and correlated with the input signal. By first assuming exact knowledge of the ARMA coefficients we use the Kalman filter theory to develop a prototype adaptive estimation algorithm which is optimum in the case of uncorrelated input and noise and outperforms, considerably, the classical RLS. We then apply the prototype algorithm in the case of correlated input and noise and show that it provides unbiased estimates as opposed to classical RLS which is highly biased. In the final part of our article, motivated by our prototype algorithm, we propose an RLS-type algorithmic variant which estimates the ARMA coefficients at the same time with the system identification part. Simulations show that this alternative version is only slightly inferior to the prototype algorithm, which requires exact knowledge of the ARMA model, inheriting all its notable advantages.

Index Terms— Adaptive system identification, Adaptive filters, RLS

1. INTRODUCTION

There are several applications in practice where we encounter the problem of system identification with colored additive noise which is correlated with the input signal. A notable example that clearly falls under this class is echo cancellation in hearing-aids. In Fig. 1 we depict a graphical representation of such a system where $\{s_n\}$ is the “additive noise” (in system identification terminology) that represents the voice signal to be heard by the patient; $\{y_n\}$ is the signal measured at the microphone containing the echo due to the feedback path F (measured output); and finally $\{x_n\}$ is the signal coming out from the loudspeaker (measured input) which, due to the required processing by the hearing-aid DSP unit, is clearly correlated with the signal $\{y_n\}$ and therefore with $\{s_n\}$. The goal is to identify F and subtract the contribution of the feedback path from the measured signal $\{y_n\}$. This action is represented by the parallel branch involving the adaptive esti-

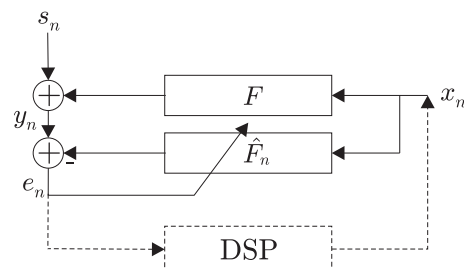


Fig. 1. Typical echo cancellation system for hearing aid where input (loudspeaker signal) $\{x_n\}$ and noise (acoustic input) $\{s_n\}$ are correlated due to DSP processing unit.

mate \hat{F}_n at time n of F .

We would like to point out that our focus in this article is on RLS and RLS-type adaptive algorithms. Even though these schemes are rarely used in real-time applications due to their high computational complexity, the reason we present them is because they can be used as prototypes for developing simpler algorithmic versions (i.e. LMS and LMS-like alternatives) that are more appropriate for this type of processing.

For the system identification problem depicted in Fig. 1 it is already known that the classical RLS is optimum in estimating F in the sense that, at every time instant, it possesses the smallest mean square error among all adaptive algorithms, *provided that the noise signal $\{s_n\}$ is white and uncorrelated with the input signal $\{x_n\}$* [1, Pages 485-490]. This interesting optimality property can be established by applying the Kalman filter theory, but for its validity it is also necessary the impulse response F to be random with known mean and covariance matrix.

Our primary goal in this work is to come up with an RLS-like algorithm capable of identifying the FIR impulse response F *without bias* for the case of an ARMA noise signal $\{s_n\}$, *by not requiring exact knowledge of the ARMA coefficients and no information about the dependency structure between input and noise*. We must mention that in [2, 3] a similar problem was considered, however our analysis is

more rigorous than in these references. Specifically, by using the Kalman filter theory, we prove optimality of a prototype algorithmic scheme when $\{s_n\}$ possesses a *known* ARMA structure and is uncorrelated with $\{x_n\}$, while in [2, 3] such optimality result is missing. Furthermore we prove, as in [2, 3], that when $\{s_n\}$ and $\{x_n\}$ are correlated, our prototype algorithm offers unbiased estimates and unlike the previous references we provide formulas for the asymptotic estimation error covariance matrix. Finally we develop an RLS-like algorithm capable of identifying the FIR filter *in parallel* with the ARMA model thus freeing us from any requirement of exact knowledge of the noise dynamics. We demonstrate that this algorithm is also asymptotically unbiased. Similar result does not exist in [2, 3].

The rest of the paper is organized as follows: Section 2 provides the formulation of the problem we are interested in under a general setup. In Section 3 we develop our adaptive algorithms and demonstrate that they enjoy several desirable characteristic properties. Finally, Section 4 contains our conclusions.

2. PROBLEM FORMULATION

Regarding notation, lower case letters denote scalars; upper case vectors and upper case, bold face matrices. Consider the following system

$$y_n = X_n^t F + s_n, \quad (1)$$

where F is an unknown, time invariant FIR system of length L ; $X_n = [x_n, \dots, x_{n-L+1}]^t$ with $\{x_n\}$ the measured input signal; $\{y_n\}$ the measured output and, as we mentioned in the Introduction, $\{s_n\}$ the additive noise. For $\{s_n\}$ we assume it is described by an ARMA(K, M) model of the form

$$s_n = -a_1 s_{n-1} - \dots - a_K s_{n-K} + w_n + b_1 w_{n-1} + \dots + b_M w_{n-M} \quad (2)$$

where $\{w_n\}$ is an unknown white sequence with variance σ_w^2 and K, M are known integers. We also assume that the ARMA process is stable suggesting that the roots of $A(z) = 1 + a_1 z^{-1} + \dots + a_K z^{-K}$ lie in the interior of the unit circle. For $B(z) = 1 + b_1 z^{-1} + \dots + b_M z^{-M}$, even though not necessary, we also make the same assumption suggesting that the filter $\frac{B(z)}{A(z)}$ is minimum phase.

Assuming, without loss of generality, that $M = K - 1$ and defining $S_n = [s_n, \dots, s_{n-K+1}]^t$, we can use the standard controllable canonical form and write (2) using state-space equations as

$$\begin{aligned} S_n &= \mathbf{A} S_{n-1} + \mathcal{E}_1 w_n \\ s_n &= \mathcal{B}^t S_n \end{aligned}$$

where

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_K \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}; \quad \mathcal{B} = \begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_{K-1} \end{bmatrix};$$

and \mathcal{E}_k denotes a vector with all elements 0 except the k th which is 1.

We can also define a state-space representation for the system in (2) where the state represents the desired FIR impulse response. Since the latter does not change in time the state equation is simply $F_n = \mathbf{I} F_{n-1}$ ($= F$) where \mathbf{I} denotes the identity matrix. Substituting s_n from the state-space equation in (1) and combining the two states F_n, S_n into a single state $[F_n; S_n]$ we can write (1) as the measurement equation of the following state-space equation

$$\begin{aligned} [F_n; S_n] &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} [F_{n-1}; S_{n-1}] + \mathcal{E}_{L+1} w_n \\ y_n &= [X_n; \mathcal{B}]^t [F_n; S_n]. \end{aligned} \quad (3)$$

In the next section, we provide adaptive estimates for the state process $\{[F_n; S_n]\}$ by applying the Kalman filter theory. This will give rise to efficient adaptive algorithms.

3. ADAPTIVE IDENTIFICATION ALGORITHMS

Suppose at each time instant n we are given the pair of observations (y_n, x_n) and we are interested in obtaining adaptive estimates $[\hat{F}_n; \hat{S}_n]$ for the state vector $[F_n; S_n]$. As we mentioned, first we are going to develop a prototype algorithm using the Kalman filter theory. This is the main focus of our next subsection.

3.1. Optimum adaptation for uncorrelated input and noise

Following the same idea that leads from Kalman filter to classical RLS, we first consider $\{s_n\}$ and $\{x_n\}$ as being uncorrelated retaining, however, the ARMA structure for $\{s_n\}$. This allows for the application of the Kalman filter theory to compute, adaptively, the optimum estimate $[\hat{F}_n; \hat{S}_n]$ which minimizes the mean square error at each time instant¹. The equations for the optimum adaptation are given in Table I, where σ_w^2 is the variance of the white process $\{w_n\}$. The algorithm is initialized with $[\hat{F}_0; \hat{S}_0]$ containing the mean of F concatenated with a zero vector representing the mean of S_0 . Similarly the error covariance matrix $\Sigma_{0|0}$ is initialized as a block diagonal where only the upper diagonal block is nonzero and equal to the covariance matrix of the random vector F .

¹If all processes are jointly Gaussian then the Kalman filter is optimum in the sense that it minimizes the mean square error. When the processes are not Gaussian then the Kalman filter is the best linear estimator.

Table I: Kalman Filter

From time $n - 1$: $\Sigma_{n-1|n-1}$, \hat{F}_{n-1} , \hat{S}_{n-1}

At time n , measure x_n , y_n and update $\Sigma_{n|n}$, \hat{F}_n , \hat{S}_n as follows:

$$\Sigma_{n|n-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \Sigma_{n-1|n-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^t \end{bmatrix} + \sigma_w^2 \mathcal{E}_{L+1} \mathcal{E}_{L+1}^t$$

$$K_n = \frac{\Sigma_{n|n-1} [X_n; \mathcal{B}]}{[X_n; \mathcal{B}]^t \Sigma_{n|n-1} [X_n; \mathcal{B}]}$$

$$\Sigma_{n|n} = (\mathbf{I} - K_n [X_n; \mathcal{B}]^t) \Sigma_{n|n-1}$$

$$\epsilon_n = y_n - [X_n; \mathcal{B}]^t \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} [\hat{F}_{n-1}; \hat{S}_{n-1}]$$

$$[\hat{F}_n; \hat{S}_n] = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} [\hat{F}_{n-1}; \hat{S}_{n-1}] + \epsilon_n K_n$$

The Kalman filter adaptation computes the optimum estimate for the whole state $[\hat{F}_n; \hat{S}_n]$. From Kalman filter theory we also know that any subset of the optimum estimator provides optimum estimates for the corresponding subset of the state. Therefore if in $[\hat{F}_n; \hat{S}_n]$ we limit ourselves to the upper part which is the estimate \hat{F}_n , then this constitutes the optimum estimator for F_n and, hence, for F at time n . Consequently, next, we present an algorithm that provides directly the estimates \hat{F}_n , by isolating the upper part from $[\hat{F}_n; \hat{S}_n]$. In order to be able to propose the desired algorithm we need to introduce two new processes $\{u_n\}$ and $\{\xi_n\}$ by filtering $\{x_n\}$ and $\{y_n\}$ respectively with the inverse of the filter that generates the ARMA process $\{s_n\}$. Specifically we define

$$\begin{aligned} u_n + \bar{U}_{n-1}^t B &= x_n + \bar{X}_{n-1}^t A \\ \xi_n + \bar{\Xi}_{n-1}^t B &= y_n + \bar{Y}_{n-1}^t A. \end{aligned} \quad (4)$$

where $\bar{X}_n = [x_n, \dots, x_{n-K+1}]^t$, $\bar{Y}_n = [y_n, \dots, y_{n-K+1}]^t$, $\bar{U}_n = [u_n, \dots, u_{n-K+1}]^t$, $\bar{\Xi}_n = [\xi_n, \dots, \xi_{n-K+1}]^t$, $A = [a_1, \dots, a_K]$ and $B = [b_1, \dots, b_K]$. We also define $U_n = [u_n, \dots, u_{n-L+1}]^t$. We then have the following theorem that yields the optimum adaptive estimator.

Theorem 1. *The adaptive identification algorithm that provides the optimum estimates of F at each time instant n has the following form:*

Table II: Prototype Algorithm

From time $n - 1$: \mathbf{P}_{n-1} , \hat{F}_{n-1}

At time n : Measure x_n, y_n

Update impulse response estimate as follows:

$$u_n = -\bar{U}_{n-1}^t B + x_n + \bar{X}_{n-1}^t A$$

$$\xi_n = -\bar{\Xi}_{n-1}^t B + y_n + \bar{Y}_{n-1}^t A$$

$$\bar{K}_n = \frac{\mathbf{P}_{n-1} U_n}{U_n^t \mathbf{P}_{n-1} U_n + 1}$$

$$e_n = \xi_n - U_n^t \hat{F}_{n-1}$$

$$\hat{F}_n = \hat{F}_{n-1} + e_n \bar{K}_n$$

$$\mathbf{P}_n = \frac{\mathbf{P}_{n-1} U_n U_n^t \mathbf{P}_{n-1}}{U_n^t \mathbf{P}_{n-1} U_n + 1}.$$

where \hat{F}_0 is initialized with the average of F and \mathbf{P}_0 with the covariance matrix of F divided by σ_w^2 .

Proof. The proof of Theorem 1 can be found in [4]. \square

The prototype adaptation has the extremely interesting characteristic of being able to outperform, considerably, the classical RLS where no inverse filtering is applied on the two measured signals $\{x_n\}$ and $\{y_n\}$. Actually the more colored the noise sequence $\{s_n\}$ is, the more we gain in performance when we apply the algorithm of Table II, instead of the classical RLS. Fig. 2 presents a typical example of the performance

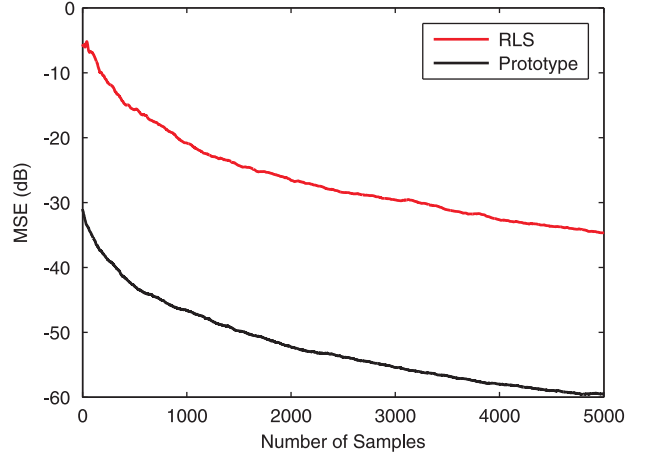


Fig. 2. Mean square estimation error for conventional RLS and our Prototype algorithm, when input and noise are uncorrelated.

of the two algorithms and, as we can observe, the algorithm of Table II can have far superior performance. Of course one can argue that this is in a sense expected since this algorithm has a more complete prior information regarding the noise sequence $\{s_n\}$ (knows exactly the ARMA coefficients) which is properly using to its advantage. However, in the last part of this section we are going to introduce an alternative scheme that requires the same prior knowledge as the classical RLS and behaves very closely to this optimal algorithm, thus eliminating this last practical weakness. Before presenting this interesting algorithmic variant, let us analyze the behavior of the two algorithms (Prototype and Classical RLS) when the noise sequence $\{s_n\}$ and the input sequence $\{x_n\}$ are correlated.

3.2. Correlated input and noise

As we argued in the Introduction, the case of correlated noise $\{s_n\}$ and input $\{x_n\}$ is a practically interesting problem since several echo cancellation problems exhibit this type of correlation between the two signals. To deal with this case one could analyze the problem assuming some form of correlation model between the two sequences $\{s_n\}$, $\{x_n\}$ and produce, again, an optimal algorithm. However in order to free

our analysis from any limitation imposed by such modeling, we decided, simply, to apply the Prototype and the RLS algorithm and analyze their convergence behavior. The reason we avoided adopting any correlation modeling between $\{s_n\}$ and $\{x_n\}$ is because in most practical systems the processing performed by the DSP unit (see Fig. 1) is extremely complicated and time varying, suggesting that our model will, most likely, be inefficient and inaccurate. The only mild assumption we impose regarding the correlation between the two sequences is the following:

Assumption 1: At every time instant n , x_n can be correlated with w_{n-1}, w_{n-2}, \dots but is uncorrelated with w_n, w_{n+1}, \dots

This is actually not a serious restriction since for the DSP unit we assume causal processing which suggests that it needs at least a unit delay to take the measurement y_n and process it in order to generate the sample x_n . We have now the following theorem that presents a very important property of our prototype algorithm and a serious weakness of classical RLS.

Theorem 2. *Under Assumption 1, the prototype adaptation depicted in Table II provides asymptotically unbiased estimates of F in the sense that $\lim_{n \rightarrow \infty} \hat{F}_n = F$, almost surely, while the classical RLS is biased.*

Proof. The detailed proof of Theorem 2 is presented in [4] along with analytic formulas for the bias of RLS and the asymptotic estimation error covariance matrix of the prototype algorithm. The proof of unbiasedness can also be found in [2, 3]. \square

Theorem 2 proves that the prototype algorithmic scheme continues to provide correct (unbiased) estimates as compared to the classical RLS which is biased. Furthermore, this desirable property is valid under the very mild restriction imposed by Assumption 1 without the need to have an actual model for the correlation between the two signals. Of course in order for this property to hold under such general condition, we had to sacrifice the optimality this algorithm enjoys in the uncorrelated noise/input case.

In Fig. 3 we present a typical example of the relative performance of the two algorithms for correlated noise/input sequences. As we can see our prototype adaptation has a mean square error that tends to 0, while the classical RLS, very quickly, converges to the wrong value. It is therefore clear that processing the two measured signals with the inverse filter, is crucial for enjoying the very important property of unbiased estimates. We should mention that in most existing hearing-aids the type of processing employed for echo cancellation produces biased estimates. As it turns out, this fact is the source behind several artifacts that were observed and demanded for special treatment with ad-hoc methods.

The prototype algorithm possesses notable advantages compared to the classical RLS. However, as we pointed out, these properties are attributed to the exact knowledge of the

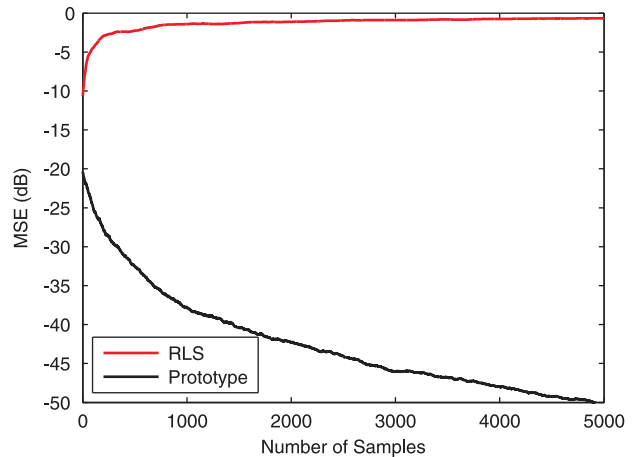


Fig. 3. Mean square estimation error for conventional RLS and our Prototype algorithm, when input and noise are correlated.

ARMA model of the noise sequence $\{s_n\}$. Of course, in practice, such prior information is difficult to obtain, basically because $\{s_n\}$ is mostly time-varying (as in the case of speech signals). In the next subsection we present a means for dealing with this serious practical problem.

3.3. An RLS-like adaptive algorithm

In this subsection we develop a variant of the prototype algorithm that does not require any prior knowledge of the ARMA model since it is capable of adaptively estimating the ARMA coefficients in parallel with the impulse response F . To produce the desired algorithm let us assume that \hat{F}_n is close to the true F . This suggests that (see Fig. 1) $e_n \approx s_n$. In fact e_n can be regarded as an efficient estimate \hat{s}_n of s_n . We can therefore write

$$e_n + E_{n-1}^t A = \hat{w}_n + \hat{W}_{n-1}^t B, \quad (5)$$

where $E_n = [e_n, \dots, e_{n-K+1}]^t$, $\hat{W}_n = [\hat{w}_n, \dots, \hat{w}_{n-K+1}]^t$ and \hat{w}_n is an estimate of w_n . Assuming at time $n-1$ that we have available estimates $\hat{A}_{n-1}, \hat{B}_{n-1}$ we can execute the adaptation depicted in Table II, using $\hat{A}_{n-1}, \hat{B}_{n-1}$ in place of the true values A, B . This will generate e_n . Furthermore, assuming that at time $n-1$ we also have E_{n-1}, \hat{W}_{n-1} we can compute \hat{w}_n from (5) as

$$\hat{w}_n = e_n + E_{n-1}^t \hat{A}_{n-1} - \hat{W}_{n-1}^t \hat{B}_{n-1}.$$

This equation allows for the parallel update of \hat{A}_n, \hat{B}_n using the classical RLS algorithm which, in combination with the algorithm in Table II, constitutes our final scheme. The complete adaptation is summarized in Table III.

Table III: Proposed Algorithm

From time $n - 1$: $\mathbf{P}_{n-1}, \hat{F}_{n-1}, \mathbf{Q}_{n-1}, \hat{A}_{n-1}, \hat{B}_{n-1}$

At time n : Measure x_n, y_n

Update impulse response estimate as follows:

$$u_n = -\bar{U}_{n-1}^t \hat{B}_{n-1} + x_n + \bar{X}_{n-1}^t \hat{A}_{n-1}$$

$$\xi_n = -\bar{\Xi}_{n-1}^t \hat{B}_{n-1} + y_n + \bar{Y}_{n-1}^t \hat{A}_{n-1}$$

$$\bar{K}_n = \frac{\mathbf{P}_{n-1} U_n}{U_n^t \mathbf{P}_{n-1} U_n + 1}$$

$$e_n = \xi_n - U_n^t \hat{F}_{n-1}$$

$$\hat{F}_n = \hat{F}_{n-1} + e_n \bar{K}_n$$

$$\mathbf{P}_n = \frac{\mathbf{P}_{n-1} U_n U_n^t \mathbf{P}_{n-1}}{U_n^t \mathbf{P}_{n-1} U_n + 1}$$

Update ARMA-coefficient estimate as follows:

$$G_n = \frac{\mathbf{Q}_{n-1} [E_{n-1}; -\hat{W}_{n-1}]}{[E_{n-1}; -\hat{W}_{n-1}]^t \mathbf{Q}_{n-1} [E_{n-1}; -\hat{W}_{n-1}] + 1}$$

$$\hat{w}_n = e_n + E_{n-1}^t \hat{A}_{n-1} - \hat{W}_{n-1}^t \hat{B}_{n-1}$$

$$[\hat{A}_n; \hat{B}_n] = [\hat{A}_{n-1}; \hat{B}_{n-1}] + \hat{w}_n G_n$$

$$\mathbf{Q}_n = \frac{\mathbf{Q}_{n-1} [E_{n-1}; -\hat{W}_{n-1}] [E_{n-1}; -\hat{W}_{n-1}]^t \mathbf{Q}_{n-1}}{[E_{n-1}; -\hat{W}_{n-1}]^t \mathbf{Q}_{n-1} [E_{n-1}; -\hat{W}_{n-1}] + 1}$$

For this algorithm we have the following theorem that addresses its convergence properties.

Theorem 3. *Under Assumption 1, the estimate \hat{F}_n of the proposed algorithm converges, almost surely, to the true impulse response F .*

Proof. The proof can be found in [4] along with closed form expressions for the asymptotic estimation error covariance matrix of the algorithm. \square

In order to test the efficiency of the proposed algorithm, we performed numerical simulations. We implemented a system, as the one depicted in Fig. 1, with the unknown impulse

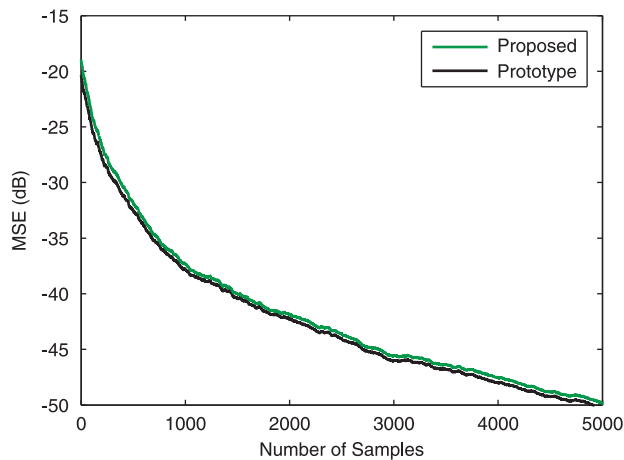


Fig. 4. Mean square estimation error for Prototype and Proposed algorithm, when input and noise are correlated.

response F drawn from a distribution with zero mean and large covariance matrix $\Sigma_F = \gamma \mathbf{I}$ with $\gamma \gg 1$. We selected the length of the impulse response to be $L = 8$. For the input signal $\{x_n\}$ we used i.i.d. Gaussian samples with mean 0 and unit variance. For the noise sequence $\{s_n\}$ we used an ARMA(2,2) model with the following coefficients: $a_1 = 0.6, a_2 = -0.3$ and $b_1 = -1.79, b_2 = 0.792$ while for the i.i.d. sequence $\{w_n\}$ we used $w_{n-1} = x_n$ in order to generate dependency between the two processes and satisfy Assumption 1.

We executed the two adaptations, namely the algorithm of Table II which knows the exact ARMA coefficients and the proposed appearing in Table III which estimates them. Fig. 4 depicts the corresponding relative performance. The curves we present are averages of 100 independent runs. As we can see, both algorithms have very similar performance with the proposed adaptation being slightly inferior, since the extra estimation required for the ARMA model inflicts a small increase in the final mean square error. Nevertheless the difference is not very notable and the two algorithms tend to have very comparable behavior.

4. CONCLUSION

In this paper we have developed an adaptive system identification algorithm for the case of additive colored noise which is also correlated with the input signal. The proposed algorithm, in parallel with system identification, performs an adaptive ARMA model estimation for the additive noise process. This fact increases significantly the overall convergence speed as compared to the classical RLS, assuring at the same time that the system identification part is unbiased. The latter property is not enjoyed by the classical RLS which provides highly biased estimates. Even though the focus in the current work is on RLS and RLS-like adaptations, the corresponding results can be easily adjusted to accommodate simpler algorithms as LMS and LMS-like alternatives that are more fitting for real-time applications.

5. REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory*, 4th edition, Prentice Hall, 2002.
- [2] J. Hellgren, "Analysis of feedback cancellation in hearing aids with filtered-X LMS and the direct method of closed loop identification," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 2, pp. 119-131, Feb. 2002.
- [3] A. Spriet et al., "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3749-3763, Oct. 2005.
- [4] P. Niavis and G. V. Moustakides, "Analysis of adaptive system identification for correlated input and noise," under preparation.