2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Stochastic Approximation for Adapting Weighted Least-Squares Algorithms: Applications in Wireless Adaptive Antenna Arrays with Time-Varying Channels and Periodically Blind Signaling.

Robert T. Buche*
NC State University, Dept. of Mathematics
Raleigh, NC 27695, USA
rtbuche@unity.ncsu.edu

***Abstract*— We consider an added level of adaptation to *classical* weighted least-squares (WLS) by adapting the forgetting factor α using a stochastic approximation (SA) algorithm, thus obtaining the *modified* WLS algorithm. The SA adapts α to the random operating conditions and leads to improved performance over the case where α is fixed. Modified WLS has wide applicability in systems operating in random environments which are difficult to predict. We focus on a wireless adaptive antenna array application where one wishes to detect the signal from a desired user (i.e. "reference signal") where the operating conditions are changing due to the mobility and channel variations. In [1], the SA for adapting α was analyzed and the modified WLS algorithm was investigated in simulations of the line-of-sight (LOS) channel case where the reference signal is known. Here we extend the simulation study to cases where the channel is Ricean and Rayleigh and also consider the case where the known reference signal is replaced by a periodically blind signal. The great benefit of the modified WLS is demonstrated. The periodically blind signaling case leads to new issues in the stochastic approximation analysis which we discuss.**

I. INTRODUCTION

A. Motivation, outline.

With the growing use and demands of applications in wireless, there is much research surrounding increasing the spectral efficiency [bits/sec/unit bandwidth] and using antenna arrays is a promising approach. This is the original motivation for our work but the method described has wide applicability to time-varying systems.

We will describe the *classical* WLS method with an added layer of adaptation. In particular, a stochastic approximation (SA) is used to adapt the forgetting factor and we call the entire algorithm *modified* WLS. We investigate the modified WLS for obtaining optimal antenna weights used to recover the signal from a desired mobile (the "reference" signal) where the other mobiles are regarded as interferers. A popular alternative for increasing the spectral efficiency with antenna arrays is space-time coding which requires channel knowledge and this is difficult to obtain accurately. In our method, channel information is not needed; indeed, the algorithm adapts to the changing environment (operating conditions). In [1] the algorithm and analysis of the SA adapting the forgetting factor was done for the line-of-sight (LOS) case and when the reference signal is known. Here, we extend the discussion the the case where the channel is Ricean or Rayleigh, also adding the case where the signal is periodically blind (i.e., where the reference signal is only known for short intervals periodically spaced).

Outline. In section II we give our wireless antenna array model. In section III we give the *classical* WLS algorithm and then give the *modified* WLS form, specifying the SA for adapting α. In section IV we highlight the SA analysis for the case where the reference signal is known; more detail is in [1]. In this section we also comment on new issues to consider when the signaling is periodically blind. In section V, the main focus of the paper, we present a collection of simulations where the channel is Rayleigh or Ricean and the signaling is either periodically blind or the reference signal is known. The simulation further characterize the algorithm and shows clearly its benefit.

*Work supported under a NCSU Faculty Research and Professional Development Award.

II. MODEL

Figure 1 gives the model for our motivating wireless application. In the figure, $\bar{x}$ denotes the complex represen-

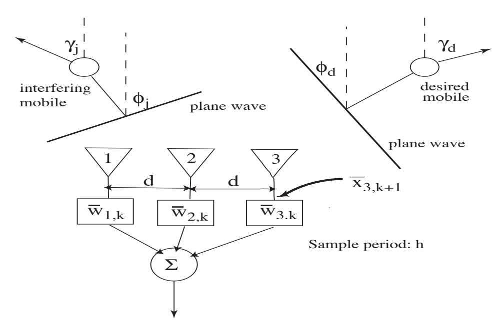

Fig. 1. Adaptive antenna array: desired mobile and interferer.

tation of the antenna signal and $\bar{w}$ is the corresponding weight. The actual signal after combining is given by $\Re\{\sum_i \bar{w}_{i,k}\bar{x}_{i,k+1}\} = W_k' X_{k+1}$. The antennas are assumed to be spaced sufficiently far apart so the scattering components at the antenna are mutually independent. The desired mobile transmits the signal we wish to detect. To simulate Rayleigh or Ricean channel conditions, the signal from the mobile is multiplied by a complex-valued gain H_k ([3]) where at time k,

$$H_k = (c + (1-c)h_k), \quad 0 \le c \le 1, \qquad (1)$$

where h_k is given by an AR(1) process driven by complex Gaussian noise. The coefficient of this AR(1) process is obtained by autocorrelation matching to Jake's model, given by the zeroth-order Bessel function of the first kind. The value of c specifies the relative strength of the LOS component and scattering component; $c = 0$ corresponds Rayleigh, $c = 1$ is LOS, otherwise it is Ricean. We consider the flat fading form for the received signal from mobile j at time k

$$\bar{r}_{j,k} = \frac{H_{j,k} s_{j,k}}{d_{j,k}^2} \left[\exp i\psi_{j,k}^d\right] c_{\phi_j,\lambda_k}, \tag{2}$$

where ψ^d is the Doppler phase, λ is the carrier wavelength, ϕ is the angle of arrival to the antenna array, d is the distance to the antenna array, s is the signal sent from the mobile, and $c(\phi, \lambda)$ is the spatial signature (antenna 1 is the reference antenna)

$$c(\phi,\lambda) = \left[1,\ \exp\left(-i\,\frac{2\pi}{\lambda} d \sin\phi\right),\ \exp\left(-i\,\frac{2\pi}{\lambda} 2d \sin\phi\right)\right].$$

III. Algorithm

First we briefly describe the *classical* WLS algorithm (see [4], for example) then describe our *modified* WLS. Let $\mathcal{S} = \{s_k\}$, $s_k \in \{\pm 1\}$ denote the reference signal. At time k, wish to find the weights to minimize

$$J_k(\alpha, W) = \sum_{l=1}^{k} \alpha^{k-l} e_k(W)^2, \ \ e_k(W) = s_k - W' X_k, \tag{3}$$

and where the forgetting factor $\alpha \in (0, 1)$ for tracking. The solution ("one-shot" form) is given by

$$W_k(\alpha) = P_k \sum_{l=1}^{k} \alpha^{k-l} X_l s_l,\ P_k = Q_k^{-1},\ Q_k = \sum_{l=1}^{k} \alpha^{k-l} X_l X_l'.$$

For slow mobility/small noise, we expect α to be relatively large (incorporating much of the distant past in the calculation of the current weight) compared to the fast mobility/high noise case. The W_k can be computed recursively where the form is that in (6). Instead of the cost in (3) we are interested in the bit-error performance and so will consider in the simulations in section V the moving average (MA) bit-error cost

$$\frac{1}{4k} \sum_{i=1}^{k} \left[\text{sgn}\left[W_l'(\alpha) X_{l+1}\right] - s_l\right]^2. \tag{4}$$

The modified WLS is given by (5) and (6). The SA used to adapt α, for small $\mu > 0$, is given by

$$\alpha_{k+1} = \alpha_k - \mu \frac{[e_{k+1}^+]^2 - [e_{k+1}^-]^2}{\delta}; \tag{5}$$

δ and the e_k are defined below. The SA has a finite difference form which gives a good approximation of the derivative of the error with respect to α. This is used in conjuction with in the following recursive form for WLS. Let $\delta > 0$ be small and $\alpha_k^{\pm} = \alpha_k \pm \delta/2$. For $\beta = \pm$, run the following for $\beta = +$ and $\beta = -$ to get the finite difference in (5):

$$\begin{aligned} e_{k+1}^{\beta} &= s_{k+1} - [W_k^{\beta}]' X_{k+1}, \\ W_{k+1}^{\beta} &= W_k^{\beta} + L_{k+1}^{\beta} e_{k+1}^{\beta}, \\ L_{k+1}^{\beta} &= \frac{P_k^{\beta} X_{k+1}}{\alpha_k^{\beta} + X_{k+1}' P_k^{\beta} X_{k+1}} \\ P_{k+1}^{\beta} &= \frac{1}{\alpha_k^{\beta}} \left[P_k^{\beta} + \frac{P_k^{\beta} X_{k+1} X_{k+1}' P_k^{\beta}}{\alpha_k^{\beta} + X_{k+1}' P_k^{\beta} X_{k+1}}\right]. \end{aligned} \tag{6}$$

Note that the solution to (6) is given by

$$\begin{aligned} &W_k^{\beta} = P_k \left[\textstyle\sum_{l=1}^{k} \alpha_l^{\beta} \cdots \alpha_{k-1}^{\beta} X_l s_l\right],\ P_k^{\beta} = \left[Q_k^{\beta}\right]^{-1}, \\ &Q_k^{\beta} = \textstyle\sum_{l=1}^{k} \alpha_l^{\beta} \cdots \alpha_{k-1}^{\beta} X_l X_l', \quad (\alpha_k \cdots \alpha_{k-1} = 1). \end{aligned} \tag{7}$$

The noise structure in (5) is complex; it is a "non-Markov state-dependent noise" [2] and is given by $\xi_k = (X_k, W_k^{\pm}, s_k, L_k^{\pm}, P_k^{\pm})$. Further discussion of the algorithm is in [1].

IV. Stochastic Approximation Analysis

The averaging in the SA analysis of (5) is nonstandard due to the non-Markovian form of the noise ξ_k and the trigonometric terms due to the Doppler phase which can be fast-varying. Developed theory can, however, be used in the SA analysis [2].

A. Reference signal known case

Highlights of the stochastic approximation analysis in [1] for the LOS case is given with a comment on the Ricean and Rayleigh case given at the end of this subsection.

Consider the stochastic approximation algorithm given by

$$\theta_{n+1} = \theta_n + \mu Y_n, \tag{8}$$

where the step size $\mu > 0$ is small and $\{Y_n, n < \infty\}$ is uniformly integrable. Let $\theta^{\mu}(\cdot)$ be defined by a piecewise constant interpolation given by $\theta^{\mu}(t) = \theta_n$ for $t \in [n\mu, n\mu{+}\mu)$. Then $\{\theta^{\mu}(\cdot)\}$ is tight in the Skorohod topology [5]. Let E_n be the conditional expectation conditioned on the data up to (but not including) iterate n. We wish to show that the average of the conditional expectation converges (in mean) to a continuous function $\bar{g}(\cdot)$, i.e.

$$\lim_{m\to\infty,\mu\to 0, m\mu\to 0} \frac{1}{m} E \left| \sum_{l=n}^{n+m-1} E_n Y_l - \bar{g}(\theta_n) \right| \to 0 \tag{9}$$

uniformly over finite intervals of length $n\mu$. Then all weak sense limits of $\theta^{\mu}(\cdot)$ satisfy the ODE $\dot{\theta} = \bar{g}(\theta)$ with $\theta(0) = \theta_0$. This is the ODE method for SA using weak convergence methods. In the SA analysis, the past values in (7) to calculate the weights W_k will be truncated to have length M and the weights will be truncated to lie in the interval $[-L,\ L]$, for large L. We justify working with the truncated process since one can show that as $M \to \infty$

and $L \to \infty$ the truncated processes converge (in mean) to the corresponding untruncated processes. This is used in conjunction with the result that if there is a uniformly integrable $\{Y_{\rho,n}, n < \infty\}$ such that $\sup_n E|Y_n - Y_{\rho,n}| < \rho$ then there is a $\bar{g}_\rho(\theta)$ giving $\dot{\theta} = \bar{g}_\rho(\theta) + \kappa$, where $|\kappa| \le \rho$. As the truncation levels are increased, $\rho \downarrow 0$ and the limit the ODE becomes $\dot{\theta} = \bar{g}(\theta)$. For our problem we take $Y_k = \left(\hat{e}^\beta_{M,k}\right)^2$, where the M denotes that M past values are used in calculating the weights and the hat means that the weights are truncated to $[-L, L]$. Note $m\mu \to 0$ as $\mu \to 0$ and $m \to \infty$ so we can replace the α_k's for calculating the weights by a single value α_n to get

$$E\left|\left(\hat{e}^\beta_{M,k}\right) - \left(\hat{e}^\beta_{M,k}\,\alpha_n \text{ used})^2\right)^2\right| \to 0.$$

Assuming that the noise is Gaussian and independent of the signal, in [1] it is shown that as $\mu \to 0$, the conditional expectation can be written

$$E_n\left[\left(\hat{e}^\beta_{M,k}\right)^2 |\alpha^\beta_n, d_{j,l}, \phi_{j,l}\psi^d_{j,l};\, j, k-M \le l \le k\right] = F_M(\alpha^\beta_n, d_{j,l}, \phi_{j,l}, \cos\phi^d_{j,l}, \sin\phi^d_{j,l};\, j, k-M \le l \le k), \quad (10)$$

where $F_M(\cdot)$ is continuous and where the parameters in the conditioning are defined in section II. Finally it is shown in [1] that there are continuous $\bar{f}_M(\cdot)$ such that (compare to equation (9))

$$\lim_{m\to\infty} \sup_n E\frac{1}{m}\left|\sum_{l=n}^{n+m-1} E_n F^\beta_{M,n,l} - \bar{f}_M(\alpha^\beta_n, d_{j,n}, \phi_{j,n})\right| = 0.$$

For $F_{M,n,l}$ in (11), the $d_{j,l}$ and $\phi_{j,l}$ are kept constant at $d_{j,n}$ and $\phi_{j,n}$ (hence the subscript n in $F_{M,n,l}$) due to their slow variation compared to the Doppler phase. If there are many samples taken over a cycle of the Doppler phase (i.e. when $w^d h$ small, where w^d is the Doppler frequency), then (11) holds with an error going to zero as the sampling rate increases. If the sampling is slow with few samples over a Doppler cycle, then (11) holds except on a set of Lesbegue measure zero, where $\omega^d h/2\pi$ is rational. The $\bar{f}_M(\alpha^\beta, d, \phi)$ is used to approximate the limit ODE corresponding to (5), see [1]. The averaging for the Ricean and Rayleigh case are similar; we just have a scattering coefficient added in the averaging but its introduction does not change the analysis.

B. Comments: periodically blind case.

During the blind part of the signaling an estimate of the transmitted signal is obtained using the antenna weight W. For fixed signal X the error is given by

$$\text{sgn}\,(W'X) - W'X,$$

where $\text{sgn}\,(W'_{k-1}X_k)$ is the estimate of s_k. This is discontinuous in W with the set of discontinuities depending on the the signal set, $\mathcal{S}$. During this blind part of the transmisision, the estimate of s_k is not independent of the noise ξ_k and one may not be able to assert (10) in the analysis for obtaining the ODE limit.

In [7], stochastic approximations with discontinuous forcing terms are analyzed where Y_n in (8) is replaced by the form $V(\theta, Y_n)$ where the "vector field" V is discontinuous in the SA parameter θ. Conditions on the manifold of singularities for V living in the space $\mathcal{X} \times \mathcal{Y}$ (where $\theta \in \mathcal{X}$ and $Y_n \in \mathcal{Y}$) are given which result in a Lipschitz condition holding for the vector field giving the class of discontinuities for which one can, with the addition of a strong enough mixing condition, obtain a limit ODE. The results in [7], while impressive, are not directly applicable to our analysis. The space $\mathcal{Y}$ in our case is much more complex as it is contains the complicated state dependent noise ξ. More fundalmentally, a w.p1. convergence result is established in [7] whereas we are applying the weak convergence method. But the ideas in [7], motivates future work investigating the conditions for which one can get (10) for some continuous $F_M(\cdot)$ or some more general form. In the latter case, one may have that $\bar{f}_M$ is replaced by a differential inclusion $G_M(\cdot)$ in (11) with the metric now meaning distance to $G_M(\cdot)$. For now, in the simulation results in section V one can see the often dramatic affects on the evolution of the adapted alpha under a periodically blind signal compared to the reference signal known case.

V. Simulation Results

The model in the section II was simulated under varying channel, signaling, and mobility. The simulations clearly demonstrate the usefulness of adapting the forgetting factor to the changing operating conditions. In the simulations, the MA bit-error cost is given by (4). We also note that

$$\text{SINR} = 10\log\frac{P_{des}}{\sum_{i=1}^{N_I} P_i + 2\sigma^2}, \quad \text{INR} = 10\log\frac{\sum_{i=1}^{N_I} P_i}{2\sigma^2},$$

where P_{des} and P_i, resp., are the signal powers (at the antenna) of the desired and ith interfering mobile, respectively.

In this section we examine the moving average (MA) bit-error cost. We compare the adapted alpha case to constant $\alpha = 0.96$ or $\alpha = 0.84$. Some LOS cases are discussed in [1] when the reference signal is known. Here we include Ricean and Rayleigh channels as well as a couple of mobility models. We then discuss the results for cases when the reference signal is composed of blocks containing a pilot sequence followed by unknown transmissions and so the adaptation is done under a periodically blind signal. We also consider a case where the channel transitions between Ricean and Rayleigh.

A. Reference signal known

We first consider cases where the reference signal is known; in subsection V-B, we treat the periodically blind signal case. We consider the case were the mobility results in the Doppler frequency in Figure 2, where the desired mobile (whose signal we wish to estimate) is mobile 2.

For this mobility, we found the MA bit-error for a LOS (Figure 3) and Ricean (Figure 6) channel (the errors used for calculating the MA bit-error in Figure 3 are shown in Figure 4). For the Ricean case we chose $c = 2/3$ in (1). For both cases we see that $\alpha = 0.96$ leads to a large number of errors over $t = [0, 0.6]$ seconds, but becomes a better choice over the rest of the interval where the Doppler frequency of the desired mobile decreases. This is reasonable since in "fast" dynamics, one would want to weight the distant past less in obtaining a current estimate. The corresponding adapted alpha for the LOS and Ricean case are in Figures 5 and Figure 7, respectively. In Figure 5 the value of alpha is low for higher Doppler frequency and high for lower Doppler frequency. In the interval $t \in [0.6, 1]$ there is initial brief transient and then alpha increases steadily corresponding to the gradual decrease in Doppler frequency; the adapted alpha tracks the optimal value for minimizing the cost under the operating conditions. The adapted alpha is insensitive to the Doppler frequency changes in mobiles 1, 3, and 4 (see Figure 2). In the Ricean case (Figure 7) , the oscillations of alpha is due to the scattering component in the signal.

In Figure 9, the channel is Ricean but the mobility of the desired mobiles is reduced (see Figure 8), resulting in the Doppler frequency for the desired mobile having the same form but lower values (i.e., abrupt change in Doppler frequency at $t = 0.6$ sec, where the initial Doppler frequency is about 240 rad/sec instead of 900 rad/sec: compare Figure 8 and Figure 2). In this case $\alpha = 0.96$ is a good choice for the forgetting factor instead of $\alpha = 0.84$ as in Figure 6.

For the same case but now with a Rayleigh channel, $\alpha = 0.84$ is a better choice for the constant forgetting factor; see Figure 10. This illustrates that α is not just sensitive to Doppler frequency but depends on the channel characteristics. In all the plots so far, the adapted alpha gave the best (or nearly so) response.

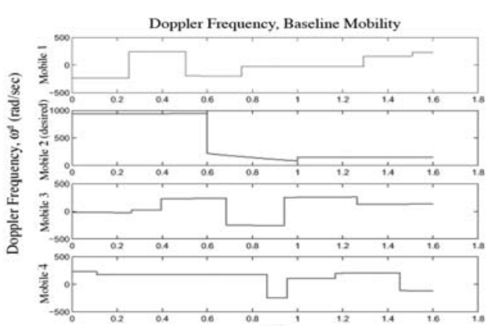

Fig. 2. "Baseline" Doppler frequency. Mobile 2 is the desired mobile.

B. Reference signal Periodically Blind

When the known reference signal is replaced by a periodically blind signal, similar trends are seen in the cost as in subsection V-A. In our simulations we used blocks of 162 bits where the first 28 bits of each block was a pilot signal

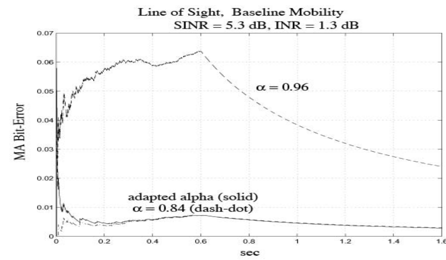

Fig. 3. Moving-average bit error, LOS channel, Baseline Mobility.

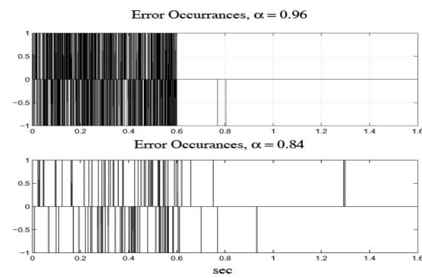

Fig. 4. Error occurrences for Figure 3.

and over the rest of the block blind adaptation was used. The adapted alpha plots associated with the periodically blind case have significantly different features than when a known reference signal is used.

In Figure 11 the LOS case corresponding to Figure 3 (where the reference signal is known) is given and here we see the cost is shifted up and the adapted alpha still gives the best response. The adaptive alpha plot for Figure 11 is in Figure 12, where, for comparison, we also include the adaptive alpha plot for the case where the reference signal is known. The costs in Figure 13 is for the conditions used for Figure 3 but with the slower mobility of the desired mobile in Figure 8. The adapted alpha corresponding to Figure 13 is in Figure 14. Between these cost figures (Figure 11 and Figure 13) notice the switch in the best constant alpha case. In the adapted alpha plot Figure 12, notice that the response between the reference signal known case and the periodically blind case differ and is more oscillatory for the periodically blind case. This seems reasonable given that the blind adaptation is taking place over a larger portion of Doppler cycles as compared to Figure 14 with slower

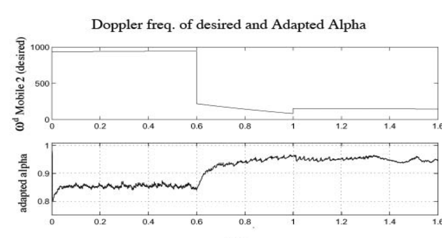

Fig. 5. Adapted alpha corresponding to Figure 3.

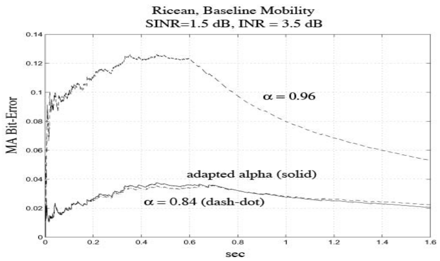

Fig. 6. Moving-average bit error, Ricean channel. Baseline Mobility.

Fig. 9. Moving-average bit error. Ricean channel. "Slower Mobility."

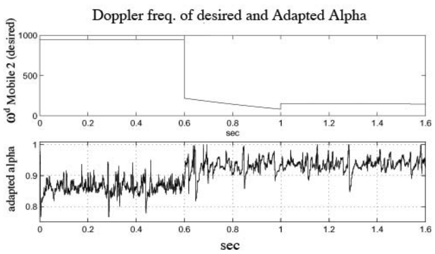

Fig. 7. Adapted alpha corresponding to Figure 6.

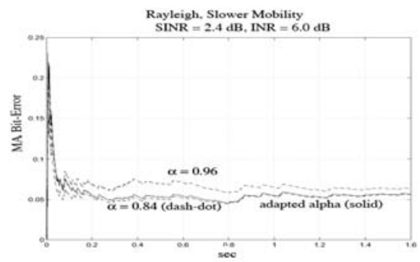

Fig. 10. Moving-average bit error, Rayleigh channel, Slower Mobility.

mobility where the α response is nearly identical.

The Ricean case in Figure 15 corresponds to Figure 6 but with periodically blind signaling. Note the cost increases in Figure 15 and the relative performance of the different cases becomes closer (this observation is discussed below, where it is seen in other cases). The adapted alpha case corresponding to Figure 15 is in Figure 16; note the sporadic "instabilities" where the hard limits $\alpha_u = 0.5$ and $\alpha_l = 1.0$ are hit. However, simulations show that the cost is insensitive to these "instabilities" as the cost is essentially invariant to values in $\alpha_l \in [0.5, 0.8]$.

Figure 18 corresponds to Figure 10 (where the reference signal is known) but with periodically blind signaling. Here the error performance is close for all cases (i.e. constant alpha cases and the adapted alpha case) illustrating the trade-off between slower mobility (favoring higher α) and fast variations due to the Rayleigh fading (favoring smaller α). It also illustrates that the nature of the periodically blind signal (ie. the relative size of the pilot signal in a block) affects the cost since the relative performance of the constant alpha cases and the adapted alpha case become closer seen by comparing in Figure 18 and Figure 10. This trend is also seen the LOS case (compare Figures 3 and 11) and the Ricean case (compare Figures 6 and 15).

The benefit of adapting alpha for the periodically blind case is also seen when the channel characteristics change. This is illustrated in Figure 17 where the channel starts as Ricean with $c = 2/3$ and then at $t = 0.4$ seconds, the channel becomes Rayleigh. This may seem unrealistic but we are interested in the performance of the algorithm under challenging operating conditions. The mobility is as in Figure 8 except after $t = 0.6$ seconds, the desired mobile stops moving. In the interval $t = [0.4, 0.6]$ note that the $\alpha = 0.96$ case becomes worse (due to the increased scattering) rising to the curve for the $\alpha = 0.84$ case, but the adapted alpha case remains the best.

REFERENCES

[1] R. Buche, H.J. Kushner, *Adaptive optimization of tracking algorithms: application to adaptive antenna arrays for randomly-varying*

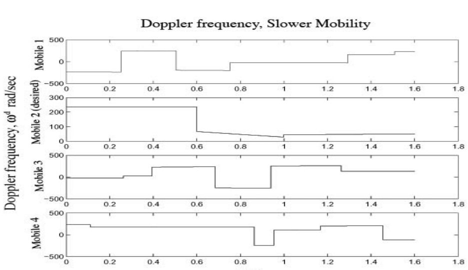

Fig. 8. "Slower Mobility." Doppler frequency. Mobile 2 is the desired mobile.

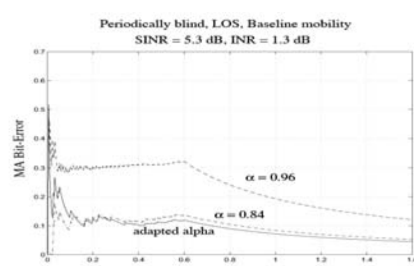

Fig. 11. Moving-average bit error, Los channel, Baseline Mobility. Periodically Blind.

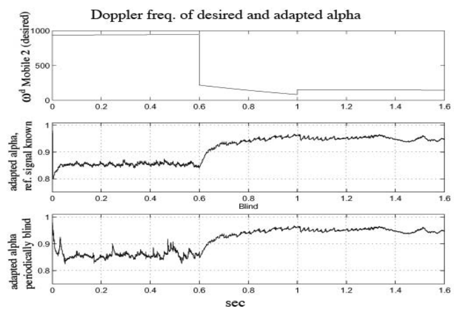

Fig. 12. Adapted alpha corresponding to Figure 11.

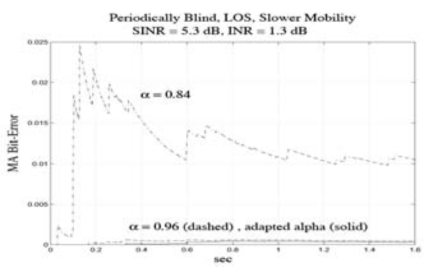

Fig. 13. Moving-average bit error, Los channel, Slower Mobility. Periodically Blind.

mobile communications, Proceedings of the 2003 IEEE Conference on Decision and Control, Vol. 5, pp. 5320-5327.

[2] H.J. Kushner and G. Yin, *Stochastic approximation and recursive algorithms and applications,* Springer, New York, 2003.

[3] C. Komninakis, C. Fragouli, A. H. Sayed, R.D. Wesel, "Multi-Input Multi-Output Fading Channel Tracking and Equalization Using Kalman Estimation, " *IEEE Transactions on Signal Processing*, Vol. 50, No. 5, May 2000.

[4] L. Ljung, *System Identification,* Prentice Hall, New Jersey, 1987.

[5] S.N. Ethier and T.G. Kurtz, *Markov Processes*, Wiley, New York, 1986.

[6] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[7] A. Benveniste, M. Goursat, and G. Ruget, "Analysis of Stochastic Approximation Schemes with Discontinuous Dynamics and Dependent Forcing Terms with Applications to Data Communication Algorithms", *IEEE Transactions on Automatic Control*, Vol. 25, No. 6, 1980.

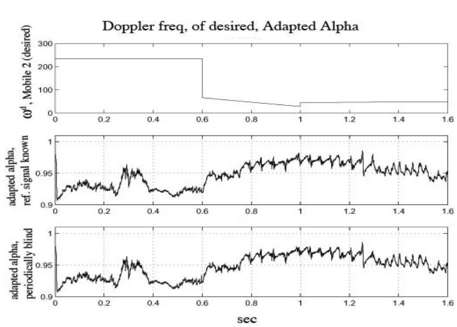

Fig. 14. Adapted alpha, corresponding to Figure 13.

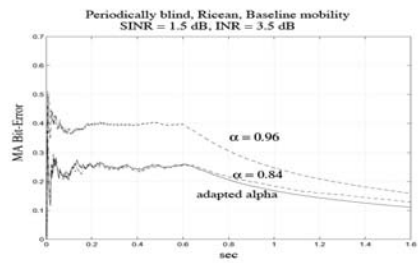

Fig. 15. Moving-average bit error, Ricean channel, Baseline Mobility. Periodically Blind.

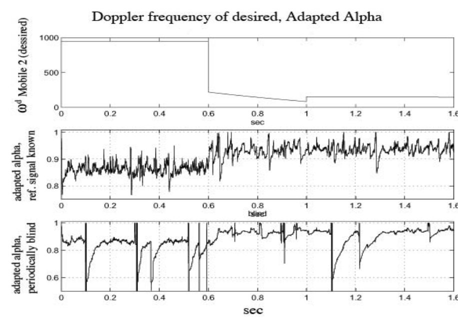

Fig. 16. Adaptive alpha corresponding to Figure 15

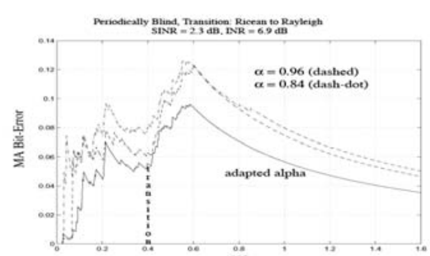

Fig. 17. Switching channels: Ricean to Rayleigh, slower mobility. Periodically Blind.

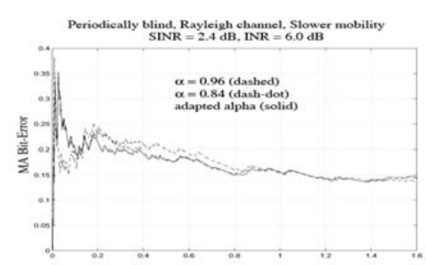

Fig. 18. Moving-average bit error, Rayleigh channel, Slower Mobility. Periodically Blind.