

Adaptive Optimization of Least Squares Tracking Algorithms: With Applications to Adaptive Antenna Arrays for Randomly Time-Varying Mobile Communications Systems

Robert Buche*
Mathematics Dept.
North Carolina State University
Raleigh, NC 27695
rtbuche@unity.ncsu.edu

Harold J. Kushner†
Applied Mathematics Department
Lefschetz Center for Dynamical Systems
Brown University, Providence, RI 02912
hjk@dam.brown.edu

February 9, 2003

Abstract

Adaptive antenna arrays are widely used for reducing the effects of interference and increasing capacity in mobile communications systems. The adaptation typically consists of updating the antenna weights by a recursive least-squares-type algorithm. We will add another adaptive loop that greatly improves the operation when the environment for the various links is randomly time-varying. The analysis is via stochastic approximation type arguments. Consider a single cell system with an (receiving) antenna array at the base station. Algorithms for tracking time varying parameters require a balance between the need to follow changes (implying a short memory) and the need to average the effects of disturbances (implying a long memory). Typical algorithms seek to recursively compute the antenna weights that minimize the weighted error function (at discrete times $kh, k = 1, 2, \dots$, for a small sampling interval

*This work was partially supported by National Science Foundation Grant ECS 9979250

†This work was partially supported by Contract DAAD-19-02-1-0425 from the Army Research Office and National Science Foundation Grant ECS 0097447.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 09 FEB 2003		2. REPORT TYPE		3. DATES COVERED 00-02-2003 to 00-02-2003	
4. TITLE AND SUBTITLE Adaptive Optimization of Least Squares Tracking Algorithms: With Applications to Adaptive Antenna Arrays for Randomly Time-Varying Mobile Communications Systems				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Brown University, Division of Applied Mathematics, 182 George Street, Providence, RI, 02912				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 23	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

$h) E \sum_{l=1}^k \alpha^{k-l} e_l^2$, where the e_l are the reception errors and $\alpha < 1$. This minimization is used only to get good weights, and the forgetting factor $\alpha < 1$ is used to allow tracking as conditions change. The actual performance of the algorithm is measured by the sample average bit error rate, and this depends heavily on the chosen value of α , as seen in simulations. The optimal value of α can change rapidly in time, perhaps significantly in seconds. Although the initial motivation arises in adaptive antennas, the method can be used to improve algorithms for tracking parameters of time varying systems, where similar issues are involved.

The additional adaptive loop tracks the optimal value of α . The adaptation can be based on a known pilot signal or it can be partially blind. The antenna weights and the value of α are adapted simultaneously. We give a stochastic-approximation-type algorithm for tracking the optimal α , which is based on an approximation to a natural “gradient descent” method. The algorithm is practical and can improve the performance considerably. The simulations under a variety of operating conditions show that the algorithm tracks the optimal weights and value of α very well. In terms of average bit error rates and for all of the scenarios tested, the new system always performs better (sometimes much better) than the original algorithm that uses any fixed value of α . Any particular application has special considerations. But the theoretical and simulation results show that the approach has great promise for significantly improving the systems.

Keywords: Stochastic approximation, tracking time-varying parameters, adaptive antennas, mobile communications, randomly time-varying channels.

1 Introduction

The adaptive antenna problem: Formulation. The method to be developed is applicable to the improvement of algorithms for the tracking of time-varying parameters, and can be a useful alternative to the method developed in [6] for the tracking of parameters of time-varying linear systems, but our development is confined to the adaptive antenna problem since that was the original motivation.

Adaptive antenna arrays are widely used to reduce the effects of interference and increase capacity in mobile communications systems. Algorithms for tracking time varying parameters require a balance between the need to follow changes (implying a short memory) and the need to average the effects of disturbances (implying a long memory). Consider a single cell system with an (receiving) antenna array at the base station. In applications, the adapted antenna weights are often computed by a recursive least-squares algorithm as follows [14, 15]. One recursively computes the antenna weights that (approximately) minimize the weighted error function (at discrete sampling times kh , for small $h > 0$ and $k = 1, 2, \dots$) $E \sum_{l=1}^k \alpha^{k-l} e_l^2$, where the e_l are some measure of the reception errors and $\alpha < 1$ is a “forgetting” factor [15]. The algorithm

is initialized using some initial block of data. The value of α is held less than unity to allow tracking of the optimal weights. This minimization is used only to get good weights, and the forgetting factor $\alpha < 1$ is used to allow tracking as conditions change. The actual performance of the algorithm is measured by the sample average bit error rate, and this depends heavily on the chosen value of α , as seen in simulations. The optimal values of the weights can change rapidly in time. The optimal value of α can also change rapidly in time, although less fast than the weights, but still perhaps significantly in a few seconds.

We will add another adaptive loop that adaptively tracks the optimal value of α and greatly improves the operation (as measured by average bit error rates) when the capacities of the various links are randomly time-varying. This additional loop (or algorithm) for adapting α is of the stochastic approximation type and is based on a natural “gradient descent” method. The algorithm is practical, and simulations under a variety of operating conditions show that the algorithm not only tracks the optimal value of α very well, but that the new system always performs better (sometimes much better), in the sense of sample average decision errors, than the original algorithm that uses any fixed value of α . The new adaptive algorithm for the parameter is analyzed via stochastic approximation theory [7]. The idea is motivated by a method for adapting the step size in stochastic approximation algorithms for the recursive identification of the values of time-varying parameters in linear systems; see [2], where the approach to adaptive adjustment of the step size in parameter tracking originated, and [6] for a detailed analysis of that case via general results in stochastic approximation, although the form of the algorithm and the systems and averaging issues are different here, due to the nature of the application. The individual driving terms in the stochastic approximation algorithm are not simply noise corrupted estimates of a gradient. But over many samples they average out to be such.

For notational simplicity only, we update both the weights and α at the same instants $kh, k = 1, 2, \dots$. Let K denote the number of antennas. Let $\bar{x}_{i,j,k}$ denote the complex (baseband) output of antenna i at time kh that is due to mobile j : it is the sum of components $\bar{r}_{i,j,k}$ and $\bar{n}_{i,j,k}$, due to the signal and noise, resp. Define $\bar{x}_{i,k} = \sum_j \bar{x}_{i,j,k}$, the output of antenna i due to all mobiles and noise. Define $\bar{r}_{j,k} = \{\bar{r}_{i,j,k}; i\}$, the vector-valued output of the K antennas due to the signal from mobile j . Define $\bar{X}_k = \{\bar{x}_{i,k}, i\}$. The weight vector depends on the mobile whose transmission is being tracked. Henceforth, we work with the algorithm for a fixed mobile, say, j_0 . (The procedure is duplicated for each mobile.) For the selected mobile j_0 , let $\bar{w}_{i,k}$ denote the complex weight assigned to antenna i at time $(k+1)h$ and define the vector $\bar{W}_k = \{\bar{w}_{i,k}, i \leq K\}$.

The weighted array output at $(k+1)h$ is $\Re\{\sum_i \bar{w}_{i,k}^* \bar{x}_{i,k+1}\} = \Re\{(\bar{W}_k^*)' \bar{X}_{k+1}\}$, where $\Re$ denotes the real part and $*$ denotes the complex conjugate, and this is used to estimate the transmitted signal symbol. *Henceforth, to simplify the notation, we concatenate the real and imaginary ($\Im$) components.* Let the unbarred quantities, such as $X_k = (\Re \bar{X}_k, \Im \bar{X}_k)$, etc., denote the concatenated values.

A typical algorithm for adapting the weights. Let $\{s_k\}$ denote a real-valued pilot or training sequence (from mobile j_0). It is assumed that the training sequence is known at the receiver. Simulations show that the algorithm to be presented also reduces the errors with only periodic use of a known pilot signal, with “blind” adaptation used in between. The development is confined to the known pilot signal case. For $\alpha \in (0, 1)$ and fixed weight vector W , define the errors $e_k(W) = s_k - W'X_k$ and the discounted error

$$J_k(\alpha, W) = \sum_{l=1}^k \alpha^{k-l} e_k(W)^2. \quad (1.1)$$

Typically $\alpha < 1$ to allow tracking of changing circumstances. A typical algorithm in applications [15] recursively computes the weight that minimizes (1.1) for some apriori fixed value of α . The weight $W_k(\alpha)$ that minimizes (1.1) is

$$W_k(\alpha) = P_k \sum_{l=1}^k \alpha^{k-l} X_l s_l, \quad P_k = Q_k^{-1}, \quad Q_k = \sum_{l=1}^k \alpha^{k-l} X_l X_l'. \quad (1.2)$$

Define $e_{k+1} = s_{k+1} - W_k' X_{k+1}$. The $W_k(\alpha)$ can be computed recursively by [8]

$$\begin{aligned} W_{k+1}(\alpha) &= W_k(\alpha) + L_{k+1}(\alpha) e_{k+1}, \\ L_{k+1}(\alpha) &= \frac{P_k(\alpha) X_{k+1}}{\alpha + X_{k+1}' P_k(\alpha) X_{k+1}}, \\ P_{k+1}(\alpha) &= \frac{1}{\alpha} \left[P_k(\alpha) - \frac{P_k(\alpha) X_{k+1} X_{k+1}' P_k(\alpha)}{\alpha + X_{k+1}' P_k(\alpha) X_{k+1}} \right]. \end{aligned} \quad (1.3)$$

Comments concerning (1.3). The procedure is initialized by using (1.2) for some value of $k > 2K$ and $\alpha = 1$. Then, under the usual conditions on the additive noise (Gaussian, nondegenerate covariance matrix, independent in time, independent of the signal), the matrix inverse P_k is well defined with probability one. Since $\alpha < 1$, the effect of the initialization wears off as $k \rightarrow \infty$. In applications, a dominant influence on the weights is the Doppler phase of the signal, which can change very rapidly, since the carrier wavelengths are very small. Owing to this and to the fact that $\alpha < 1$, the weights can change rapidly.

The term $e_l(W)$ is the difference between the desired signal and the weighted (with weight W) output of the array at lh and is used to compute the $W_k(\alpha)$ in (1.3). The term $e_{k+1}(W_k(\alpha))$ is the difference between the desired signal and the optimally weighted output of the array at $(k+1)h$. Of greater interest in applications is the bit or decision error. Suppose, for simplicity in the discussion, and to coordinate with the simulations, that the signal s_k takes the binary values ± 1 .¹ Then, for large k the pathwise average of the bit errors

$$\frac{1}{4k} \sum_{l=1}^k [\text{sign}[W_l'(\alpha) X_{l+1}] - s_l]^2 \quad (1.4)$$

¹Any finite signal set could be used.

is an appropriate measure of performance. The results of simulations of this algorithm for mobile communications were reported in [15].

On the value of α . The value of (1.4) can be quite sensitive to the value of the forgetting factor α , as can be seen from the numerical data in Section 5. If the mobiles are not moving and the variance of the additive noise is constant, then the optimal value of α will be unity. If the mobiles (particularly the one being tracked) are moving rapidly and the additive noise level is small, then the optimal value of α will be relatively small. In practice, the optimal value might vary rapidly, perhaps changing significantly many times per second, as the operating conditions change. Our aim is the development of a practical algorithm for adapting α ; i.e for tracking its current optimal value. It is based on an intuitively reasonable gradient descent idea. The simulations presented in Section 5 show the rapid response of the adaptive procedure.

Outline of paper. The adaptive algorithm, motivated by the work in [6] on tracking the parameters of linear time-varying systems, will be described in Section 2. The adaptive procedure finds the proper balance between the averaging of the noise effects (i.e., larger α) and the ability to track (i.e., smaller α). There are two levels of adaptation. One is that which estimates the optimal antenna weights, given α , and the other adapts α to optimize the overall performance. The precise model that defines the $\bar{x}_{j,k}$ will be given in Section 3. The simulation results in Section 5 clearly demonstrate the utility of the approach and behavior of the algorithm under a variety of challenging operating conditions, with different INR (interference to noise) and SINR (signal to interference plus noise) ratios. The algorithm is a form of stochastic approximation with “state-dependent” noise [7]. Its behavior can be well approximated by the solution of a mean ODE (ordinary differential equation), and a theoretical justification for the adaptation properties is in Section 4.

2 The Adaptive Algorithm for α

Simulations for the signal model of Section 3 indicate that for large k the function $\sum_{l=1}^k E e_l^2(W_l(\alpha))/k$ is strictly convex and continuously differentiable in α . The value increases sharply as α increases beyond its optimal value, and increases (although more slowly) as α decreases below its optimal value. It is somewhat insensitive to α around the optimal value. Finite difference estimators, as in (2.3), for the difference intervals that we use, provide excellent approximations to the derivatives.

The algorithm for adapting the weights and α . Let $\delta > 0$ be a small difference interval, let α_k denote the value of α at the k th update, and define $\alpha_k^\pm = \alpha_k \pm \delta/2$. The update formula for α_k will be given below. The algorithm

(1.3) is run for both $\alpha_k^\pm$. Define the two sets of recursions, for $\beta = +$ or $-$:

$$\begin{aligned} e_{k+1}^\beta &= s_{k+1} - [W_k^\beta]' X_{k+1}, \\ W_{k+1}^\beta &= W_k^\beta + L_{k+1}^\beta e_{k+1}^\beta, \\ L_{k+1}^\beta &= \frac{P_k^\beta X_{k+1}}{\alpha_k^\beta + X_{k+1}' P_k^\beta X_{k+1}}, \\ P_{k+1}^\beta &= \frac{1}{\alpha_k^\beta} \left[P_k^\beta - \frac{P_k^\beta X_{k+1} X_{k+1}' P_k^\beta}{\alpha_k^\beta + X_{k+1}' P_k^\beta X_{k+1}} \right]. \end{aligned} \quad (2.1)$$

For either case, $\beta = +$ or $-$, the solution to (2.1) can be written as

$$W_k^\beta = P_k \left[\sum_{l=1}^k \alpha_l^\beta \cdots \alpha_{k-1}^\beta X_l X_l' \right], \quad P_k^\beta = [Q_k^\beta]^{-1}, \quad Q_k^\beta = \sum_{l=1}^k \alpha_l^\beta \cdots \alpha_{k-1}^\beta X_l X_l' \quad (2.2)$$

where $\alpha_k \cdots \alpha_{k-1} = 1$. The computation is about twice what is required for the classical algorithm (1.3). Owing to the discounting in (2.2), the effects on W_k^β of the sample taken at time lh decreases geometrically as $k - l \rightarrow \infty$. W_k^β minimizes

$$\sum_{l=1}^k \alpha_l^\beta \cdots \alpha_{k-1}^\beta e_l^2(W).$$

Truncated weights. With appropriate initialization, under typical conditions on the noise, the inverse matrix in (2.2) is well defined with probability one, and very large values of the weights rarely occur in applications or in our simulations. If a large value did occur, it would be truncated or ignored. We will use truncated weights in the antennas and in the algorithm for updating α . For some large number L and $\beta = +$ or $-$, truncate the components of W_k^β to the interval $[-L, L]$. Denote the truncated values by $\hat{W}_k^\beta$. For large enough L , truncation will rarely occur. Define the errors $\hat{e}_k^\beta = s_{k+1} - \hat{W}_k^\beta X_{k+1}$.

For small $\mu > 0$, the adaptive algorithm for α is the finite difference form

$$\alpha_{k+1} = \alpha_k - \mu \frac{[\hat{e}_k^+]^2 - [\hat{e}_k^-]^2}{\delta}. \quad (2.3)$$

α_k is constrained to the interval $[\underline{\alpha}, \bar{\alpha}] \in (0, 1)$, where $\underline{\alpha}$ (resp., $\bar{\alpha}$) can be as close to zero (unity, resp.) as desired. The use of the finite difference in (2.3) avoids stability issues associated with the derivative form, and yields a more robust algorithm. The simulations were not sensitive to the value of δ if it was small (say, $\delta \leq .02$). In typical applications, the individual values of the $\hat{e}_k^\beta$ are not “noisy” estimates of the error; hence classical approaches to stochastic approximation would not be useful. But averaged over many samples, they “average out” to be such, as seen in Section 4. For decision making purposes, any value $\hat{W}_k^+$, $\hat{W}_k^-$, or a convex combination, can be used.

If the values of α_l are fixed at α and the process X_k is stationary, then for small $\delta > 0$ the stationary expectation of the coefficient of μ in (2.3) should be close to the derivative of the expectation with respect to α . The intuitive idea behind the algorithm is that, although the optimal value of α might change “fast,” it changes much more slowly than that of W , so that we are essentially in a stationary state. In this case, we clearly have a stochastic algorithm driven by a process whose values are estimates of the negative of a gradient. In applications, the process X_k can rarely be considered to be stationary or ergodic. For example, with the data model of Section 3, the Doppler phases are the dominant influences on the error. These phases can change rapidly over time. But, even then, as seen in Section 4, the desired averaging will occur. The overall conclusion is that, for small μ , the algorithm for adapting α behaves as a (finite difference approximation to a) gradient descent algorithm, which is what we are aiming for.

Comment on the properties of the algorithm for small μ . The algorithm (2.1) and (2.3) has two levels. The optimal weight W_k is tracked by (2.1), while the optimal forgetting factor α is tracked by (2.3). The actual stochastic approximation algorithm is (2.3), where the step size μ is small. The quantities $(X_k, \hat{W}_k^\pm, s_k, L_k^\pm, P_k^\pm)$ play the role of “state-dependent noise” [7]. Clearly, there are several time scales, and the “noise” has a complicated structure. Nevertheless, results from the theory of stochastic approximation in [7, Chapter 8] can be used to analyze the algorithm. In applications, the performance is a great deal less sensitive to the value of μ in (2.3) than the original algorithm (2.1) is to the choice of α . Under broad conditions, the algorithm performs well, with α_k being nearly optimal, and the bit error rate reduced in comparison with (1.3), with any fixed value of α . See Section 5.

The behavior of (2.3) for small μ is determined by a mean ODE, whose right hand side is a “local” average of the coefficient of μ in (2.3). Loosely speaking, since the α_k sequence varies much more slowly than do the “driving noises,” one can compute this local average by assuming that α_k is fixed. Let E_n denote the expectation conditioned on the data to sampling time nh . Suppose that there is a function $\bar{g}(\alpha)$ and m such that

$$\frac{1}{m} \sum_{l=n}^{n+m-1} E_n \frac{[\hat{e}_l^-]^2 - [\hat{e}_l^+]^2}{\delta} \approx \bar{g}(\alpha)$$

for large n and m , where α_l is held fixed at α . Then the mean ODE is $\dot{\alpha} = \bar{g}(\alpha)$ [7, Chapter 8]. See Section 4 for more detail.

3 The Physical Model for the Mobiles

We now specify the class of physical models more precisely. The averaging in the stochastic approximation argument depends on the structure of the observations $\{X_k\}$. We work with a class that is commonly used in adaptive antennas.

It covers interesting and important applications and well illustrates the essential ideas. The exact models for communications systems tend to be very complicated. There are issues of synchronization at many levels (e.g., chip, bit, packet), possible interference between sources, the type of modulation used, non ideal filters, whether matched or otherwise, and so forth. We will use a common representation (quite similar to that in [15]) for the sequence of bits at the output of the antennas, in baseband form and after processing, and supposing perfect synchronization.

In order to keep the analysis and simulations from getting too complicated and to focus on the essential issues of adaptation, the simulations were restricted to an array of three antennas and the mobiles move in two dimensions. The three antennas are evenly spaced with spacing $d > \lambda_j/2$ for all j , where λ_j is the carrier wavelength of mobile j . Nevertheless, the results concerning averaging and asymptotics hold for any number of antennas and for full three dimensional motion.

Line of sight (LOS) model. First consider the model where the received signals from the tracked and the main interfering mobiles are not scattered, but received on the “line of sight.” There are several strong interfering mobiles plus additive noise. The additive noise is Gaussian and white in time. Part of the (complex-valued) additive noise might be due to additional interferers with scattering. If the scattering is “uniform” (i.e., Rayleigh fading), then the associated noise component can be taken to be independent across the antenna elements since $d > \lambda_j/2$ for all j [4, 10]. All that we require is that the noise be nondegenerate in that the covariance matrix is non singular. In the simulations, it was taken to have independent components.

The tracked and interfering mobiles are in the far field so their transmitted electromagnetic wave can be assumed to be a plane wave at the antenna array. The signal amplitude at the receiver from mobile j at time k is $1/d_{j,k}^2$, where $d_{j,k}$ is the distance to a reference antenna in the array [4]. We assume a narrowband signal (carrier frequency $\gg$ signal bandwidth) so the signal does not change appreciably over the time that it takes to traverse the antenna array. The pilot signal s_k , for the tracked or desired user, is assumed known. It is finite-valued and i.i.d., and is independent of the signals from the other mobiles.² In practice, there would be either a training period or reference signals sent periodically, as part of the desired users synchronization signal.

²The independence is used to simplify some calculations. It is only required that the distribution of $\{s_{n+k+l}, l = 1, 2, \dots\}$ conditioned on the data to time nh converge weakly to the unconditional distribution as $k \rightarrow \infty$, uniformly in n .

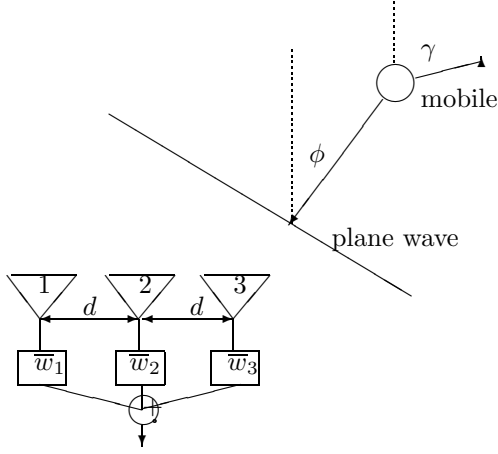


Figure 3.1. A Three Antenna Array.

The (complex-valued) antenna signature of a mobile corresponding to a plane wave of carrier wavelength λ arriving at an angle ϕ to the normal to the plane of the antennas (see Figure 3.1) is given by the vector (antenna 1 is the reference antenna)

$$c(\phi, \lambda) = \left[1, \exp\left(-i \frac{2\pi}{\lambda} d \sin \phi\right), \exp\left(-i \frac{2\pi}{\lambda} 2d \sin \phi\right) \right]. \quad (3.1)$$

There is an obvious extension for arbitrary K .

The Doppler frequency of mobile j at time kh is

$$\omega_{j,k}^d = -\frac{2\pi}{\lambda_j} v_{j,k} \cos(\phi_{j,k} - \gamma_{j,k}), \quad (3.2)$$

where $\gamma_{j,k}$ is the angle of travel of mobile j (see the figure), $v_{j,k}$ its speed, and $\phi_{j,k}$ the angle of arrival of its plane wave. The (complex-valued here) component of the received signal at the antenna array at sampling time kh due to mobile j is given by

$$\bar{r}_{j,k} = \frac{s_{j,k}}{d_{j,k}^2} [\exp i \psi_{j,k}^d] c(\phi_{j,k}, \lambda_j), \quad (3.3)$$

where $\psi_{j,k}^d = \psi_j(0) + \int_0^{kh} \omega_j^d(s) ds$, the Doppler phase at time kh , $s_{j,k}$ is the signal from mobile j , $\psi_j(0)$ is the initial phase, $d_{j,k}$ is the distance between mobile j and the array, and λ_j is the carrier wavelength for mobile j . Thus $s_k = s_{j_0,k}$, where we recall that j_0 denotes the tracked mobile.

Of particular interest is the typical case where the wave numbers $2\pi/\lambda_j$ are very large so that small variations in the mobility of the mobile can lead to large changes in the Doppler frequency.³ The model is close to what was used in [15].

³The signal X_k can be based either on TDMA or CDMA. In the latter case, it is measured after the matched filters, which use the signature of the desired user.

The scattering model. When scattering of the transmitted signal from many rapidly changing reflectors is important, a common model for the part of the antenna outputs that is due to the signal is a multipath form of (3.3), where each component is multiplied by a complex-valued random variable. For example,

$$\bar{r}_{j,k} = \sum_{l=0}^q \frac{h_{j,k-l}}{d_{j,k-l}^2} s_{j,k-l} [\exp i\psi_{j,k-l}^d] c_{j,k-l},$$

where $h_{j,k}$ is a complex gain, correlated in the time parameter k , and independent of the observation noise, the angle of arrival, distance to the array, and angle of travel.

There is no mathematical problem in handling the general delay model, but the simulations concerned only the so called flat fading form which is

$$\bar{r}_{j,k} = \frac{h_{j,k} s_{j,k}}{d_{j,k}^2} [\exp i\psi_{j,k}^d] c_{j,k}. \quad (3.4)$$

The complex-valued $h_{j,k}$ model the time-varying effects of many independent scattering objects. It is rarely the case in practice that more than one delay path is considered [13] and furthermore the case (3.4) is often used ([1, 3, 11], for example).

First, consider a “Rayleigh” channel. The autocorrelation function of the $h_{j,k}$, for each j , is usually assumed to be a zeroth order Bessel function of the first kind as in $R_j(kh) = J_0(\omega_j^d kh)$ [4]. The mathematical arguments that are required to justify the stochastic approximation averaging can be carried out under this assumption. But, for purposes of simulation, the autocorrelation function is often approximated by that of an autoregressive process [5, 13].⁴ Then the model for each channel is taken to be

$$h_{j,k+1} = \beta_{j,k} h_{j,k} + g_{j,k} \xi_{j,k}, \quad (3.5)$$

where $\xi_{j,k}$ is i.i.d. complex-Gaussian with variance unity and independent across the channels. The real number $g_{j,k}$ is chosen so that $E|h_{j,k}|^2 = 1$, assuming that the real number $0 < \beta_{j,k} < 1$ is fixed and the process stationary. For the Rayleigh channel, $\beta_{j,k}$ is chosen to roughly match the desired autocorrelation, and it will depend on the Doppler frequency [5].

A “Ricean” channel models the case where both the LOS and the scattering components are significant [9, 12]. In this case the term $h_{j,k}$ is replaced by $q_1 + q_2 h_{j,k}$, for positive constants q_i such that $q_1 + q_2 = 1$.

4 Convergence

Comments on Stochastic Approximation. Only simple results from SA theory are needed. Consider the vector-valued iteration $\theta_{n+1} = \theta_n + \mu Y_n$, where

⁴The reference assumed a constant Doppler frequency, and ours is time-varying.

μ is small positive number and the set $\{Y_n, n < \infty\}$ is uniformly integrable. Define the process $\theta^\mu(\cdot)$ by $\theta^\mu(t) = \theta_n$ for $t \in [n\mu, n\mu + \mu)$. Then $\{\theta^\mu(\cdot)\}$ is tight in the Skorohod topology [7]. Let E_n denote the expectation conditioned on the data to iterate n . Suppose that there is a continuous function $\bar{g}(\cdot)$ such that

$$E \left| \frac{1}{m} \sum_{l=n}^{n+m-1} E_n Y_l - \bar{g}(\theta_n) \right| \rightarrow 0 \quad (4.1)$$

uniformly in any finite $n\mu$ interval, as $m \rightarrow \infty, \mu \rightarrow 0$ and $\mu m \rightarrow 0$. Then all weak sense limits of $\theta^\mu(\cdot)$ satisfy the ODE $\dot{\theta} = \bar{g}(\theta)$, with $\theta(0) = \theta_0$. The full theory is in [7, Chapter 8].

Suppose that for each $\rho > 0$ there is $Y_{\rho,n}$ such that $\{Y_{\rho,n}, n < \infty\}$ is uniformly integrable and $\sup_n E|Y_n - Y_{\rho,n}| \leq \rho$ and let there be a continuous function $\bar{g}_\rho(\cdot)$ such that (4.1) holds for $Y_{\rho,n}$ and $\bar{g}_\rho(\cdot)$ used. Then $\dot{\theta} = \bar{g}_\rho(\theta) + \kappa$, where $|\kappa| \leq \rho$. If $\bar{g}_\rho(\cdot)$ converges to a continuous function $\bar{g}(\cdot)$, uniformly on each bounded set, then the mean ODE is $\dot{\theta} = \bar{g}(\theta)$. What is not quite standard in the averaging is that the effects of both noise and the “trigonometric” Doppler phase need to be accounted for.

Approximation of the algorithm (2.3). For each integer M , define $f_M(\cdot)$ by

$$\begin{aligned} f_M(X_l, \alpha_l; k-M \leq l \leq k) &= P_{M,k} \left[\sum_{l=k-M}^k \alpha_l \cdots \alpha_{k-1} X_l s_l \right], \\ P_{M,k} &= Q_{M,k}^{-1}, \quad Q_{M,k} = \sum_{l=k-M}^k \alpha_l \cdots \alpha_{k-1} X_l X_l'. \end{aligned} \quad (4.2)$$

Define $W_{M,k}^\beta = f_M(X_l, \alpha_l^\beta; k-M \leq l \leq k)$ for $\beta = +, -$. Under the assumptions on the additive noise, the matrix inverse will exist for all $\alpha_l \in [\underline{\alpha}, \bar{\alpha}]$, w.p.1, for large M ($> 2K$). In fact, for any $\rho > 0$,

$$\lim_{M \rightarrow \infty} \sup_k P \left\{ \left| W_{M,k}^\beta - W_k^\beta \right| \geq \rho \right\} = 0. \quad (4.3)$$

For any $\delta_0 > 0$, on the open X -set where the determinant of $Q_{M,k}$ is $> \delta_0$ for all possible α -values, $f_M(\cdot)$ is continuous in $(X_l, \alpha_l; k-M \leq l \leq k)$. The probability of the X -set where the determinant is $> \delta_0$ goes to one as $\delta_0 \rightarrow 0$. On the set where the determinant is zero, let the pseudoinverse replace the inverse, so that the “inverse” is defined and continuous for all $\{X_l, k-M \leq l \leq k\}$ and values of the α_l in the allowed range.

Now, truncate the components of $f_M(\cdot)$ to the interval $[-L, L]$, for large L , as was done to get the $\hat{W}_k$. Let $\hat{f}_M(\cdot)$ and $\hat{W}_{M,k}^\beta$ denote the truncations. Then, for each $\rho > 0$,

$$\lim_{M \rightarrow \infty} \sup_k P \left\{ \left| \hat{W}_{M,k}^\beta - \hat{W}_k^\beta \right| \geq \rho \right\} = 0. \quad (4.4)$$

Define $\hat{e}_{M,k+1}^\beta = s_{k+1} - [\hat{W}_{M,k}^\beta]' X_{k+1}$. Then

$$\sup_{M \rightarrow \infty} \sup_k E \left| [\hat{e}_k^\beta]^2 - [\hat{e}_{M,k}^\beta]^2 \right| = 0.$$

Let m (it can go to infinity and depend on μ) be such that $\mu m \rightarrow 0$ as $\mu \rightarrow 0$ and $m > \max\{M, 2K\}$. Let $0 \leq k - n \leq m$. Replacing $\alpha_l, \dots, \alpha_{k-1}$, in the $\hat{W}_{M,k}^\beta$ by α_n yields an error that goes to zero (in mean) uniformly in k as $\mu \rightarrow 0$. For $\beta = +, -$, we have

$$E \left| \left(\hat{e}_{M,k}^\beta \right)^2 - \left(\hat{e}_{M,k}^\beta(\alpha_n \text{ used}) \right)^2 \right| \rightarrow 0$$

uniformly in n, k, m , as $m\mu \rightarrow 0$. This replacement of the $\alpha_l, n \leq l \leq n + m$ by α_n will be used henceforth, whenever convenient. Consider $E_n \left[\hat{e}_{M,k}^\beta \right]^2$. Owing to the assumptions on the additive Gaussian noise, the conditional expectation is a smoothing operation. In fact, there are continuous functions $F_M(\cdot)$ (with components bounded uniformly in M , for each L) such that, modulo an error that goes to zero uniformly in n, m , in k such that $k - n \leq m$, and in the conditioning data, as $\mu \rightarrow 0$, with the replacement of the $\alpha_l, n \leq l \leq n + m$ by α_n ,

$$\begin{aligned} E_n \left[\left(\hat{e}_{M,k}^\beta \right)^2 \middle| \alpha_n^\beta, d_{j,l}, \phi_{j,l}, \psi_{j,l}^d; j, k - M \leq l \leq k \right] \\ = F_M \left(\alpha_n^\beta, d_{j,l}, \phi_{j,l}, \cos \psi_{j,l}^d, \sin \psi_{j,l}^d; j, k - M \leq l \leq k \right). \end{aligned} \quad (4.5)$$

Furthermore,

$$\sup_{M \rightarrow \infty} \sup_k E \left| [\hat{e}_k^\beta]^2 - [\hat{e}_{M,k}^\beta]^2 \right| = 0. \quad (4.6)$$

The limit in (4.6) holds uniformly in all non anticipative α -sequences whose values are confined to $[\underline{\alpha}, \bar{\alpha}]$. Thus, according to the comments on SA in the beginning of the section, to get the mean ODE, we can approximate by working with (4.5), with M fixed, but arbitrarily large.

An approximation for the d, ϕ . For slowly varying $d_{j,l}, \phi_{j,l}$, we can replace them in (4.5) by $d_{j,n}, \phi_{j,n}$, without loss of the essential features of the algorithm. For example, let the velocity be approximately 100 km/hr with $d_{j,0} = 1$ km. In one second there are typically over 1000 samples taken, and the distance has changed by about 3%, a very small fraction of the typical percentage change in the Doppler phase. Similarly, the angle of arrival usually changes very slowly in comparison with the Doppler phase.

Let $F_{M,n,k}^\beta$ denote the value of (4.5) with the substitution of $d_{j,n}, \phi_{j,n}$ for $d_{j,l}, \phi_{j,l}$. Suppose that there are continuous functions $\bar{f}_M(\cdot)$ such that

$$\lim_{m \rightarrow \infty} \sup_n E \frac{1}{m} \left| \sum_{l=n}^{n+m-1} E_n F_{M,n,l}^\beta - \bar{f}_M(\alpha_n^\beta, d_{j,n}, \phi_{j,n}) \right| = 0. \quad (4.7)$$

Then by the SA result at the beginning of the section, for large M the mean ODE is well approximated by

$$\dot{\alpha} = [\bar{f}_M(\alpha^-, d, \phi) - \bar{f}_M(\alpha^+, d, \phi)]/\delta \equiv \bar{g}_M(\alpha, d, \phi). \quad (4.8)$$

Details for (4.7) will now be filled in for some basic cases.

Proof of (4.7): The LOS model. Recall that the signals are assumed to be i.i.d. and binary-valued. The main issue in the verification of (4.7) concerns the effect of the Doppler frequency in (3.3). As noted above, we suppose that the $d_{j,k}, \phi_{j,k}$ are constant, with values d, ϕ , resp. Set $\alpha_n = \alpha$. First, suppose that the Doppler frequencies from all mobiles are constant and mutually incommensurate; that is $\omega_j^d(t) = \omega_j^d$ for all j . If $l - n$ is large enough (with $\mu(l - n)$ small), then $F_{M,n,l}^\beta$ in (4.7) depends only on $\alpha^\pm$ and on the values of the trigonometric arguments of $F_M(\cdot)$. It is not random. If $h\omega_j^d$ is small for all j , then there are many samples per cycle for each j . Then for large m the sum in (4.7), namely,

$$\frac{1}{m} \sum_{l=n}^{n+m-1} E_n F_{M,n,l}^\beta, \quad (4.9)$$

is very close to

$$\bar{f}_M(\alpha^\beta, d, \phi) = EF_M(\alpha^\beta, d, \phi, \cos(\psi_j - \omega_j^d lh), \sin(\psi_j - \omega_j^d lh); j, 0 \leq l \leq M), \quad (4.10)$$

where the ψ_j are independent and uniformly distributed on $[0, 2\pi]$. Then (4.7) will hold approximately, with an error that goes to zero as $h\omega_j^d \rightarrow 0$. Note that of $\omega_j^d = 100$ and $h = 10^{-4}$, then $h\omega_j^d = .01$, which tells us that the assumed smallness is not usually restrictive.

Now drop the assumption that h is small. Then the sampling of the Doppler phase (mod 2π) is “course,” and one has the possibility of the samples taking only a few values. But this will “rarely” occur in the following sense. For notational simplicity, and without loss of generality, let $\psi_j(0) = 0$. Then (4.5) is

$$F_M(\alpha^\beta, d, \phi, \cos \omega_j^d lh, \sin \omega_j^d lh; j, k - M \leq l \leq k).$$

For all but a null set of $\{\omega_j^d, j\}$, (4.10) is the limit of (4.9), as $m \rightarrow \infty$. There will be problems in special cases; e.g., where some $\omega_j^d h/2\pi$ is rational. But even there the approximation will be often good, barring special cases where the samples of the trigonometric functions take very few values.

Next suppose that the Doppler frequencies vary randomly as independent random walks. In particular, let the Doppler phase for mobile j at the l th sample be $\psi_j(0) + b_j lh + B_j(lh)$, where the $B_j(\cdot)$ are mutually independent Wiener processes, perhaps with different variances. Then for large $l - n$, for purposes of evaluating (4.7), the value of (4.9) can be well approximated by the representation

$$EF_M(\alpha^\beta, d, \phi, \cos \psi_{j,l}^d, \sin \psi_{j,l}^d; 0 \leq l \leq M, j),$$

where $\psi_{j,l}^d = \psi_j - b_j lh + B_j(lh)$, the ψ_j are mutually independent, uniformly distributed on $[0, 2\pi]$, and independent of the $B_j(\cdot)$.

The scattering cases. The analysis is nearly identical to that above. Write $h_{j,l} = h_{j,l}^R + ih_{j,l}^I$ in terms of the real and imaginary components. In (4.5), with the replacement of the $\alpha_l, n \leq l \leq n + m$ by α_n , condition the expectation on the $h_{j,l}$ also to yield

$$E_n \left[\left(\hat{e}_{M,k}^\beta \right)^2 \middle| \alpha_n^\beta, d_{j,l}, \phi_{j,l}, \psi_{j,l}^d, h_{j,l}; j, k - M \leq l \leq k \right].$$

Then the $\cos \psi_{j,l}^d$ in $F_M(\cdot)$ is replaced by the real part of $h_{j,l} e^{i\psi_{j,l}}$, namely, $h_{j,l}^R \cos \psi_{j,l}^d - h_{j,l}^I \sin \psi_{j,l}^d$, and the $\sin \psi_{j,l}^d$ is replaced by the imaginary part $h_{j,l}^R \sin \psi_{j,l}^d + h_{j,l}^I \cos \psi_{j,l}^d$. The rest of the details are very similar to those for the simpler LOS case and are omitted.

5 Simulations

Unless otherwise noted, the direction and velocity of each mobile evolved as a semi-Markov process, each moving independently of the others. They were constant for a short random interval, then there was sudden acceleration or deceleration in each coordinate, and so forth. Only the Doppler frequencies for the various mobiles are plotted, since they are the most important factors in the adaptation of α . The number of strong interfering mobiles (N_I) is either one or three. Mobile 2 is always the desired one, the one whose signal we are tracking. We used $\mu = .0008$, $\delta = .002$, $h = 4 \times 10^{-5}$ seconds, and the carrier frequency is 800×10^8 Hz. In each case, the signal amplitude at the receiver of each interfering mobile was approximately the same, and each was approximately one fourth that of the desired mobile. This represents a large interference. Changing the value of μ up or down by factor of four had little effect on the overall performance, demonstrating the insensitivity to μ (relative to that with respect to α) noted in Section 2. For example, for the run of Figure 1 dividing μ by four leads to a smoother graph for α , and the transient period for the discontinuity at $t = .6$ becomes about 0.1 second.

The LOS cases will be discussed first. Define, as usual (measured at the input to each antenna),

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

where P_{des} and P_i are the signal powers (at the antenna) of the desired and i th interfering mobile, resp., and σ^2 denotes the variance of the real and complex parts of the additive noise. The most important factor in the determination of the optimal value of α at any time is the Doppler shift, although the values are also affected by the SINR and INR. In each case, results from a single

run are presented. But the adaptation, errors, and relative performances, were consistent in all runs taken.

The simplest case is in Figure 1, where there is just one interferer and $\text{SINR} = 5.3$ db, $\text{INR} = 1.3$ db. The SINR is rather large, especially in view of the fact that it is measured after detection (e.g., in CDMA, after multiplying by the spreading sequence of the tracked mobile). Since the actual mobility and noise data can vary quite a bit, even over short time intervals, it is important not to ignore difficult cases. The top two figures plot the Doppler frequencies (in radians per second vs. time measured in seconds) of the two mobiles. For high Doppler frequencies, the optimal α is low, and conversely for low values, as expected. The algorithm tracks the optimal value very well. Mobile 2 starts with a high Doppler frequency (corresponding to a velocity of approximately 150 km/hr), which then decreases suddenly at $t = .6$ sec., then decreases more slowly, and finally increases slightly. The behavior of α is typical for all simulations with this mobility for mobile 2. It initially oscillates about $\alpha = .85$, which is very close to the optimal value for the associated Doppler frequency. Then, when the Doppler frequency drops to about 200 radians/sec, α increases quickly, and then continues to increase (on the average) as the Doppler frequency continues to drop. At $t = 1$, the Doppler frequency rises slightly and then remains constant. Except for the brief transient periods, the values of α are close to the optimal. When smaller μ is used, the paths of α are smoother, the transient period longer, all of which is intuitively reasonable, but the overall performance is very similar.

The sample number of bit errors divided by the number of samples to date (i.e., (1.4), the sample mean error probability) is in Figure 2. The fixed- α cases are for the classical mean square algorithms (1.3), with the chosen value of α used. The lowest (or best) line is for the adapted case, the one slightly above is for fixed $\alpha = .84$, and the highest for fixed $\alpha = .96$. Clearly, if a fixed value of α is to be used, then the errors depend heavily on its value. The adapted case was always the best, no matter what the mobilities. If the value of α is fixed, then a larger value is better for the second half of the run, which accounts for the decrease in the top curve there. In fact, for all cases in Figure 1 there were few errors after about $t = .6$.

For the case of Figure 1, and in general, the results are relatively insensitive to the Doppler frequencies of the interfering mobiles. This is true since the signals from the different mobiles are mutually independent: In particular, note that, in (1.2) and (2.1), the $s_{j,k}$ sequences for the non tracked mobiles are independent of that for the tracked mobile, which would make expressions such as (4.9) less sensitive to the nontracked mobiles than to the tracked mobile. In some cases, there is a fixed value of α such that (1.3) performs nearly as well as the adaptive algorithm. But even then, the optimal fixed value changes with the environment. While lower fixed values are better for the cases of Figures 1-4, the higher fixed values would be better if the tracked mobile moved more slowly (as in the case of Figure 9). Generally, the Doppler frequency will change many times in the course of a transmission, and any fixed value cannot be even nearly optimal all of the time.

An increase in the variance of the additive noise leads to larger errors and

a larger value of the optimal α , as expected, since it is optimal to do more “noise averaging.” There is a fairly short “memory” in these algorithms, so the randomness in the noise sequence has a significant affect on α . The behavior is smoother if smaller μ is used. Despite the fact that the α_n -plots are wilder than those in Figures 1 and 3, the adaptive algorithm still outperforms the classical algorithm.

Now, consider the case of Figure 3, where there are three strong interfering mobiles, SINR = 0.5 db, and INR = 7.9 db. Again, the optimal value of α was tracked well. The average bit errors are plotted in Figure 4. The error values for the fixed value $\alpha = .84$ were nearly as good as the adaptive algorithm. Again, there were few errors after about $t = .6$.

Another example of good tracking appears in Figure 5, where the Doppler frequency of the tracked mobile varies in a sawtooth fashion. The algorithm again produces nearly optimal α . The interfering mobiles are as in Figure 3. The bit error rates are in Figure 6. Clearly, the adaptive algorithm is better. In this case, for the fixed- α algorithm, the $\alpha = .96$ case starts off better than the $\alpha = .84$ case, since the Doppler frequency is relatively low. The comparison reverses itself as the Doppler frequency ranges from low to high and back again. See also Figures 7 and 8, where the Doppler frequency of the tracked mobile decreases in a linear manner. Figure 9 corresponds to the case where the tracked mobile is moving at a (slow !) speed of about 14.5 km/hour. Here, for the fixed- α algorithm, the larger value is better.

The results under Raleigh or Ricean scattering were similar, except that the errors were larger, due to the larger effective noise.

References

- [1] B.C. Banister and J.R. Zeidler. Tracking performance of the RLS algorithm applied to an antenna array in a realistic fading environment. *IEEE Trans. on Signal Processing*, 50:1037–1050, 2002.
- [2] J.-M. Brossier. *Egalization Adaptive et Estimation de Phase: Application aux Communications Sous-Marines*. PhD thesis, Institut National Polytechnique de Grenoble, 1992.
- [3] A. Chiani, A. Win, A. Zanella, and J.H. Winters. A simple and asymptotically tight upper bound on the symbol error probability of adaptive antennas with optimum combining. In *Proc. IEEE Conf. on Communications*, volume 3, pages 1429–1433, IEEE Press, New York, 2002.
- [4] W.C. Jakes. *Microwave Mobile Communications*. Wiley-Interscience, New York, 1974.
- [5] C. Komninakis, C. Fragouli, A. Sayed, and R. A. Wesel. Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *IEEE Transactions on Signal Processing*, 50:1065–1076, 2002.

- [6] H.J. Kushner and J. Yang. Analysis of adaptive step size SA algorithms for parameter tracking. *IEEE Trans. Automatic Control*, AC-40:1403–1410, 1995.
- [7] H.J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, Berlin and New York, 1997.
- [8] L. Ljung. *System Identification Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, 1986.
- [9] T.S. Rappaport. *Wireless Communications*. Prentice Hall, Englewood Cliffs, NJ, 1996.
- [10] J. Salz and J.H. Winters. Effect of fading correlation on adaptive arrays in digital mobile radio. *IEEE Trans. on Vehicular Technology*, 43:1049–1057, 1994.
- [11] S. Shamai and A.D. Wyner. Information-theoretic considerations for symmetric cellular, multiple-access fading channels—Part I. *IEEE Trans. on Information Theory*, 43:1877–1894, 1997.
- [12] M.K. Simon and M.S. Alouini. *Digital Communication over Fading Channels*. Wiley-Interscience, New York, 2000.
- [13] M.K. Tsatsanis, G.B. Giannakis, and G. Zhou. Estimation and equalization of fading channels with random coefficients. *Signal Processing*, 53:211–229, 1996.
- [14] G.V. Tsoulos, Editor. *Adaptive Antenna Arrays for Wireless Communications*. IEEE Press, New York, 2001.
- [15] J.H. Winters. Signal acquisition and tracking with adaptive arrays in digital mobile radio system IS-54 with flat fading. *IEEE Trans. Vehicular Technology*, 42:377–393, 1993.

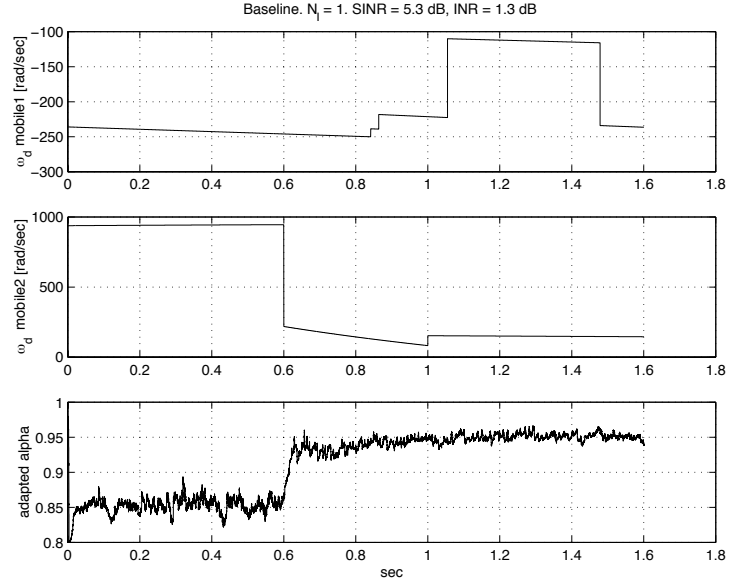


Figure 1: $N_I = 1$.

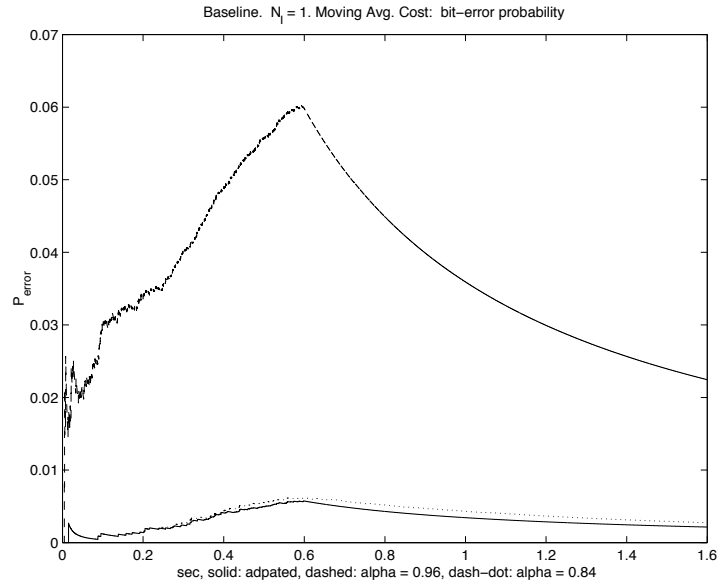


Figure 2: Average pathwise bit errors for Figure 1.

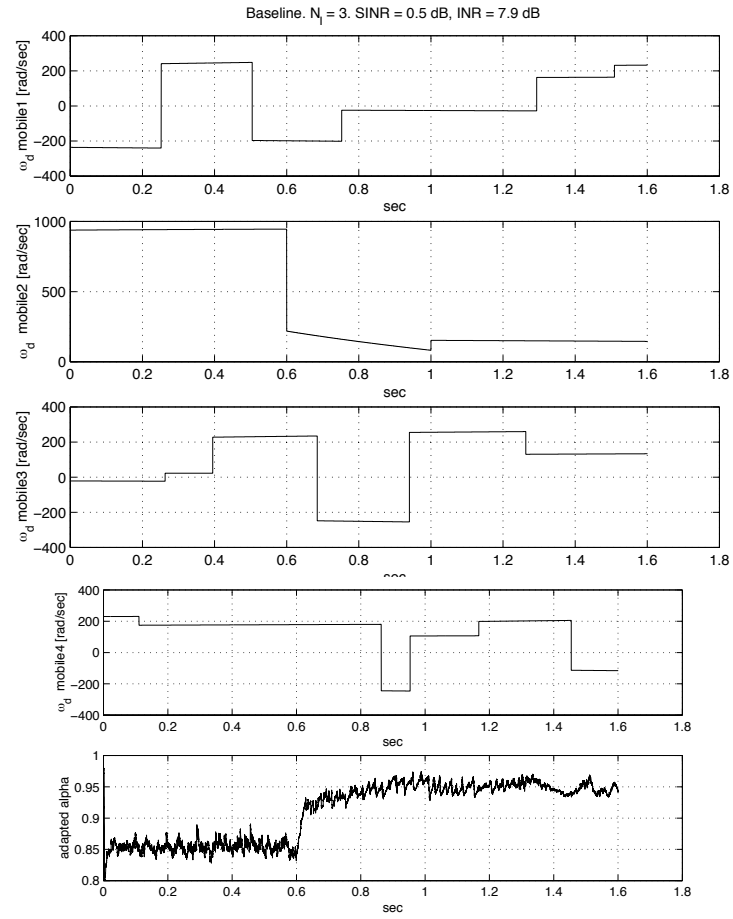


Figure 3: $N_I = 3$.

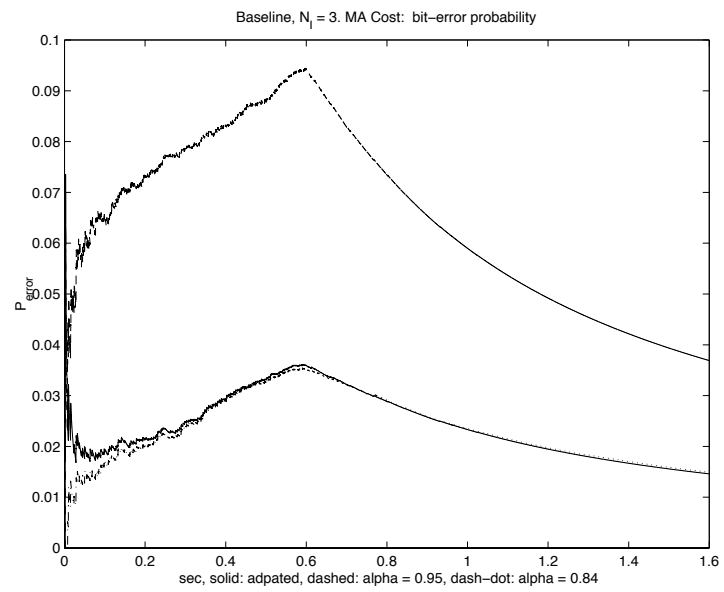


Figure 4: Average pathwise bit errors for Figure 3.

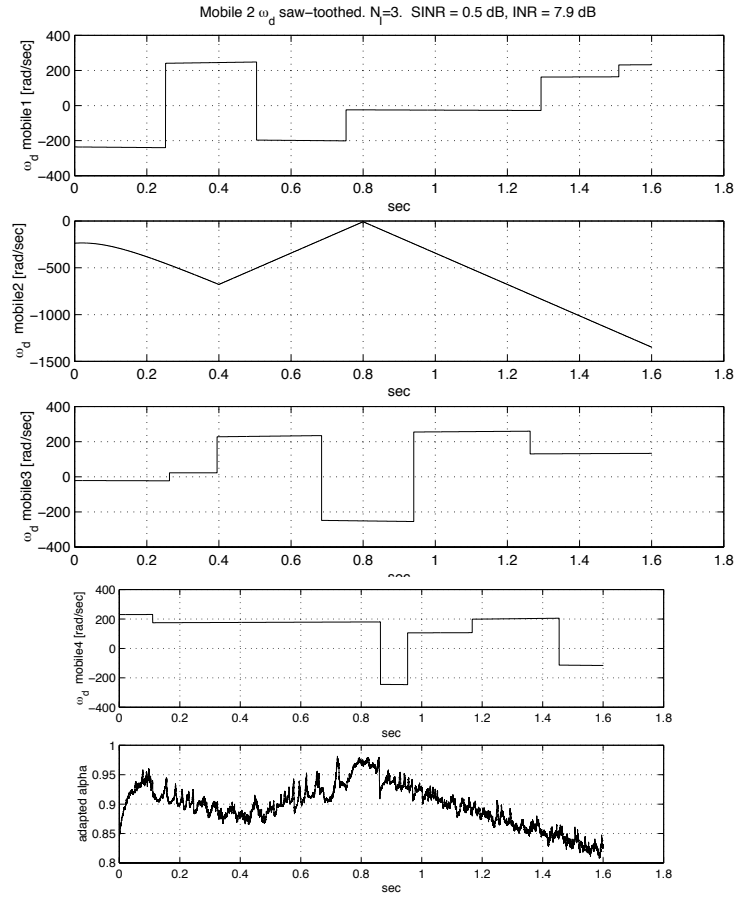


Figure 5: Sawtooth Doppler frequency, $N_I = 3$.

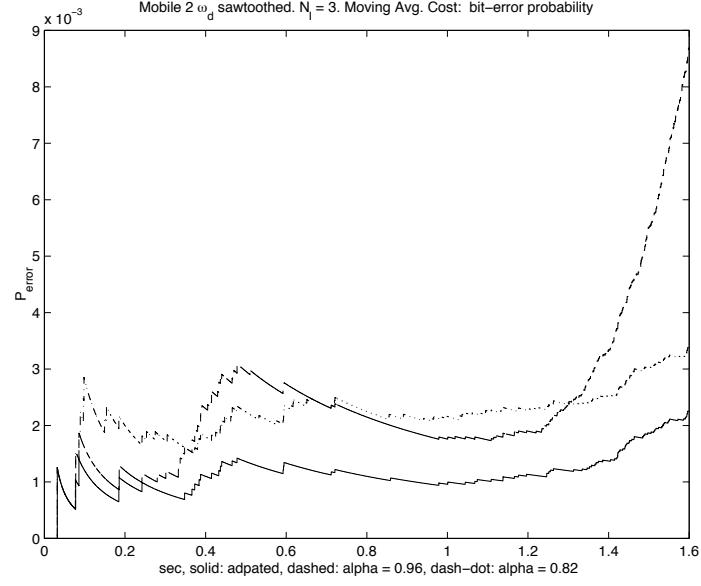


Figure 6: Pathwise average bit errors for Figure 5.

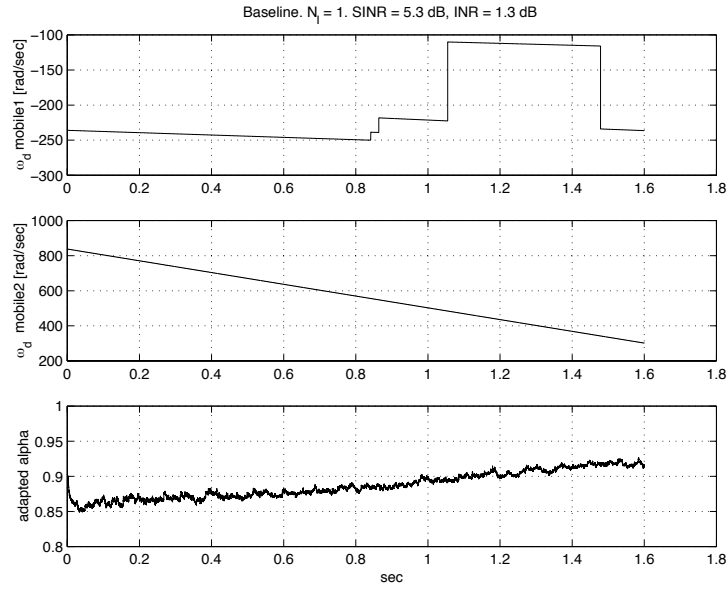


Figure 7: Linear Doppler frequency, $N_I = 1$.

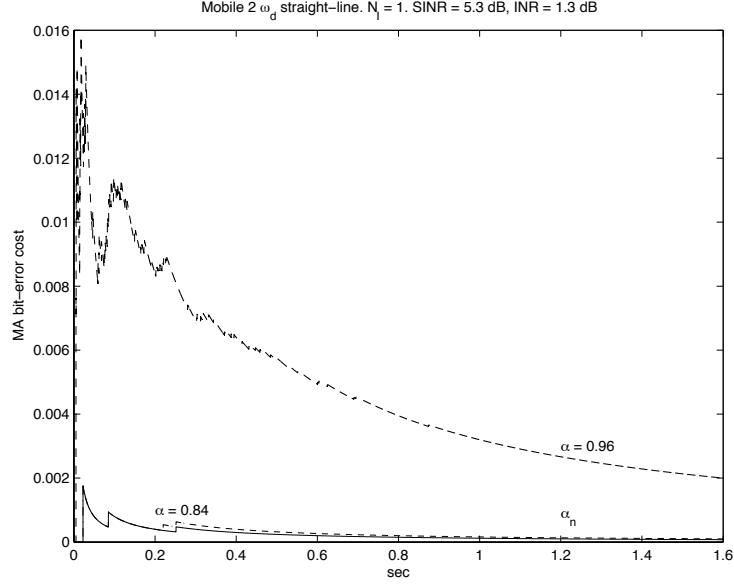


Figure 8: Pathwise average bit errors for Figure 7.

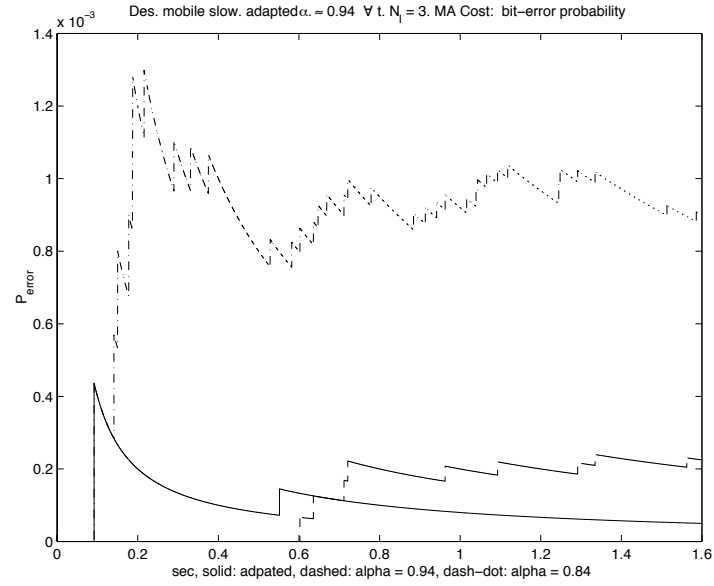


Figure 9: Desired mobile moves at 14.5 km/hr in constant direction, $N_I = 3$.