Gradient Estimation for Stochastic Optimization of Optical Code-Division Multiple-Access Systems: Part II—Adaptive Detection

Narayan B. Mandayam, *Member, IEEE,* and Behnaam Aazhang, *Senior Member, IEEE*

***Abstract*—In this sequel, we develop infinitesimal perturbation analysis (IPA)-based stochastic gradient algorithms for deriving optimum detectors with the average probability of bit error being the objective function that is minimized. Specifically, we develop both a class of linear as well as nonlinear (threshold) detectors. In the linear scheme, the receiver despreads the received optical signal with a sequence that minimizes the average bit-error rate. In the case of the threshold detector, the detection threshold for the photoelectron count is optimized to achieved minimum average bit-error rate. These algorithms use maximum likelihood estimates of the multiple-access interference based on observations of the photoelectron counts during each bit interval, and alleviate the disadvantage of previously proposed schemes that require explicit knowledge of the interference statistics. Computer-aided implementations of the detectors derived here are shown to outperform the correlation detector. Sequential implementations of the adaptive detectors that require no preamble are also developed, and make them very viable detectors for systems subject to temporal variations.**

***Index Terms*—Infinitesimal perturbation analysis, linear detectors, minimum probability of error detection, stochastic gradient algorithms, threshold detectors.**

I. Introduction

As pointed out in Part I (see [1]) of this paper, the multiple-access nature of optical CDMA systems renders intractable the optimization of such systems with respect to minimizing the average probability of bit error. This applies to optimum detection as well, where the design of minimum average probability of bit-error detectors is not possible unless explicit assumptions are made on the multiple-access interference. As a result, the existing single-user and multiuser detection schemes [2]–[7] are quite difficult to implement in that they require perfect estimates of the multiple access interference in the channel. Based on the DEDS formulation for optical CDMA systems developed in [1], we develop infinitesimal perturbation analysis (IPA)-based stochastic gradient algorithms for minimum probability of bit-error detection. We develop a class of threshold as well as linear detectors that do not require explicit information about the multiple-access interference.

Specifically, in the class of threshold detectors, we develop a one-shot detector for optimum single-user detection, where the optimum threshold for photodetection is derived without any assumptions on the interference statistics. We develop an optimization scheme using the hybrid Newton's method where the average probability of error is minimized with respect to the detection threshold. The gradients required for this optimization procedure are evaluated using IPA estimates averaged over several bit intervals. Moreover, this gradient algorithm assumes that the correlated intensities can be observed over each of the bit intervals. As an extension, we develop an adaptive detector for optimum single-user detection where no assumptions are made on the interference statistics. This scheme is based on a stochastic gradient algorithm using IPA, which requires the desired user to transmit a stochastic training sequence or a preamble of bits and the detection threshold is recursively updated from bit to bit. This scheme uses maximum likelihood estimates of the multiple-access interference based on observations of the photoelectron count at the output of the correlator. In the class of linear detectors, we develop several implementations of a scheme where the received intensities in the channel are linear filtered using an optical tapped delay line [4], [5]. The weights of this optical linear filter are chosen to minimize the average probability of bit error.

The adaptive detectors are shown to outperform the conventional correlation detector, as well as to alleviate the disadvantages of the previously proposed detection schemes [2]–[7], that require knowledge of the multiple-access interference. The adaptive algorithms are shown to converge, and also allow very simple recursive update structures that make them very attractive detectors for practical implementation in optical CDMA channels. Finally, we also present a sequential algorithm that does not require the desired user to transmit a preamble. In this scheme, bit estimates of the desired user are also incorporated in the stochastic gradient algorithm.

The paper is organized as follows. We describe the system in Section II, followed by the threshold detectors in Section III. The linear detectors are presented in Section IV, and the conclusions in Section V.

II. System Description

We will consider an optical CDMA system (time-encoded) where an optical encoder maps each bit of information into

Manuscript received May 5, 1996; revised October 25, 1996. This paper was supported by the Advanced Technology Program of the Texas Higher Education Coordinating Board under Grant 003604-018. This paper was presented in part at the 1994 Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 1994.

N. B. Mandayam is with the Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08855-0909 USA.

B. Aazhang is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892 USA.

Publisher Item Identifier S 0733-8716(97)02680-2.

a very high rate optical sequence, which is then coupled into a single-mode fiber channel. At the receiver end, the optical pulse sequence is correlated to a stored replica of itself (correlation process), and the resulting sequence serves as the intensity that excites photoelectrons at the output of the photodetector. These photoelectrons are compared to a threshold at the comparator for data recovery. Consider a K-user optical CDMA system where the users employ codes from a family of optical pulse sequences of length N_c. If user k is sending bit i in the time interval $[0, T)$, then the intensity of the modulated light will be $\lambda_i^{(k)}(t)$ where

$$\lambda_i^{(k)}(t) = \sum_{n=1}^{N_c} \lambda_i^{(k)}(n)\Pi_{T_c}(t - nT_c), \qquad i = 0, 1; \quad \text{for} \quad t \in [0, T) \tag{1}$$

and $\Pi_{T_c}(t)$ is a unit rectangular pulse of duration T_c, and $\boldsymbol{\lambda}_i^{(k)} = [\lambda_i^{(k)}(1), \cdots, \lambda_i^{(k)}(N_c)]$ is a signature sequence of length $N_c = T/T_c$ with each $\lambda_i^{(k)}(n) \in \{0, \lambda_k\}$. This gives rise to the following two hypotheses at the receiver of the desired user (taken to be user 1) in the time interval $[0, T)$ as

$$\begin{aligned} H_0&: \tilde{\lambda}_0^{(1)}(t) = \lambda_0^{(1)}(t) + \sum_{k=2}^{K} \lambda_b^{(k)}(t) \\ H_1&: \tilde{\lambda}_1^{(1)}(t) = \lambda_1^{(1)}(t) + \sum_{k=2}^{K} \lambda_b^{(k)}(t) \end{aligned} \tag{2}$$

where $\tilde{\lambda}_i^{(1)}(t)$ is the sum of the intensities in the channel under hypothesis H_i, due to the first user and $K-1$ interferers, and in $\lambda_b^{(k)}(t)$ the symbol b denotes the information of the kth user. The receiver corresponding to user 1 has a replica of the signature sequence assigned to user 1, and the light in the channel due to the user 1 and other $K-1$ interferers is correlated with this replicated signature sequence. Without loss of generality, we assume that each user is employing on–off keying, and hence at the first receiver, the intensities are correlated with $\boldsymbol{\lambda}_1^{(1)}$ since $\boldsymbol{\lambda}_0^{(1)} = \mathbf{0}$. For convenience, let $\lambda_1 = 1$, and $\lambda = \langle \boldsymbol{\lambda}_1^{(1)}, \boldsymbol{\lambda}_1^{(1)} \rangle$. Then the correlated intensity at the nth chip interval, i.e., the nth observation from the sample path, is given by $\alpha_n + \beta_n$, where α_n is the contribution of the desired user sequence to the (auto) correlation process with $\alpha = \sum_{n=1}^{N_c} \alpha_n$ being either λ or 0, depending on whether the desired user is transmitting a 1 or a 0, and β_n is the contribution of the interfering users to the (cross) correlation and is given by

$$\beta_n = \sum_{k=2}^{K} \Lambda_{k,n}, \qquad n = 1, \cdots, N_c \tag{3}$$

where $\Lambda_{k,n}$ is either λ_k or 0 with probability $p_{k,n}$ and $1-p_{k,n}$, respectively, $\forall k = 2, \cdots, K, \forall n = 1, \cdots, N_c$. The above probabilities depend on the cross-correlation properties of the codes, and also the relative delays of the users. From observing such a sample path, we can compute the sample performance function under hypothesis H_i (i.e., the probability of error from a single sample path) as

$$P_{e,i} = \sum_{r \in \Delta_{1-i}} \mathcal{K}(i, r), \qquad i = 0, 1 \tag{4}$$

where Δ_i is the decision region for the photoelectron count r that hypothesis H_i is true, and the kernel $\mathcal{K}(i, r)$ is given by

$$\mathcal{K}(i, r) = \frac{(i\lambda + \beta)^r}{r!} \exp\left(-(i\lambda + \beta)\right) \tag{5}$$

where the symbol β is given as $\beta = \sum_{n=1}^{N_c} \beta_n$. The average probability of bit error is given (for binary symmetric hypothesis) by

$$\overline{P}_e = \tfrac{1}{2}\,\overline{P}_{e,0} + \tfrac{1}{2}\,\overline{P}_{e,1} \tag{6}$$

where $\overline{P}_{e,i} = E[P_{e,i}]$, $i = 0, 1$. As in [1] and [8], we observe the sequence of correlated intensities at the desired user, and this constitutes the nominal sample path. Based on the observation of a single sample path, we first compute the sample performance function, and then, according to appropriate perturbation generation and propagation rules, we compute the sample derivative. Even though the above formulation has been presented for chip-synchronous codes, a similar analysis can be performed for chip-asynchronous codes as well. In this case, the observation is over a period $[0, T)$, which may not correspond to exactly N_c observations over each of the chips.

III. Threshold Detection

Before developing a class of threshold detectors, we begin by illustrating how IPA can be used for estimating the sensitivity of the average probability of error to the detection threshold. These estimates are shown to be unbiased, and are further used to develop an optimum one-shot single-user detector for systems with nondeterminant interference statistics. This scheme is based on finding the optimum threshold that minimizes the average probability of error for the desired user. We use a gradient algorithm that uses IPA estimates based on observation of the sample realizations at the output of the correlator. This algorithm assumes that the sample path $(\alpha_n + \beta_n), \forall n = 1 \cdots N_c$ can be observed directly. However, in a practical setting, due to photon detection, the output of the correlator is in the form of photoelectron counts. To overcome this drawback, the above results are extended to develop an adaptive single-user detector [9] that uses an IPA-based stochastic gradient algorithm for minimizing the average probability of error. The adaptive algorithm proposed here requires the desired user to transmit a stochastic training sequence or a preamble of bits. Furthermore, this adaptive scheme incorporates maximum likelihood estimates of the multiple-access interference based on observations of the photoelectron count during each bit interval. Finally, we develop a sequential detector that uses bit estimates of the desired user during each bit interval in the stochastic gradient algorithm for minimum probability of detection error. This precludes the preamble of bits to be transmitted by the desired user.

A. Optimum Threshold Detector

We now present a scheme for optimal detection (i.e., minimum average probability of error with respect to interference statistics) when the interferer codes are unknown or, in other words, the statistics of the output of the correlator are unknown, i.e., the probabilities $p_{k,n}$ defined in (3) are not known $\forall k = 2, \cdots, K, \forall n = 1, \cdots, N_c$. The correlation detector [10], [11], is a suboptimum detector in the sense that it is optimal if only the desired user is present in the system. If the input statistics are known, then the optimum threshold for the likelihood ratio detector is given by

$$\gamma = \arg_r \left\{ \frac{E[\mathcal{K}(1,r)]}{E[\mathcal{K}(0,r)]} = 1 \right\} \tag{7}$$

and has been referred to in the literature as the optimized single-user detector [2], [12]. However, if the interference statistics are unknown, the expectation in (7) cannot be evaluated. We would like to estimate the optimum threshold for detection as

$$\begin{aligned} \gamma^* &= \arg \min_{\gamma} \overline{P}_e \\ &= \arg \min_{\gamma} \left\{ \sum_{r=0}^{\gamma} E[\mathcal{K}(1, r)] + \sum_{r=\gamma+1}^{\infty} E[\mathcal{K}(0, r)] \right\} \end{aligned} \tag{8}$$

which is a minimization of a function of one variable, and can be solved by application of the hybrid Newton's method [13, pp. 32–35] as

$$\gamma_{m+1} = \gamma_m - \mu \frac{d\overline{P}_e}{d\gamma} \tag{9}$$

where $d\overline{P}_e/d\gamma$ is a finite-difference derivative and μ determines the rate of convergence of the above method. We propose to apply IPA to estimate the finite-difference derivative in (9). This overcomes the difficulty of evaluating the expectation, and the derivative is evaluated from the sample path itself. Since the algorithm in (9) is not adaptive, we use IPA estimates averaged over several bit intervals to compute $d\overline{P}_e/d\gamma$. However, we need to show that the IPA estimate of $d\overline{P}_e/d\gamma$ is an unbiased estimate. Clearly, the sample performance function $P_{e,i}$ in (4) is a discontinuous function of γ since γ can take only integer values. Therefore, the interchange of the expectation and derivative operator is not possible in this case. To allow such an interchange, we will now introduce a real variable δ such that $\gamma = \lfloor \delta \rfloor$. Now, the sample performance function in (4) can be rewritten as

$$\begin{aligned} P_e = \sum_{r=0}^{\lfloor \delta \rfloor -1} \mathcal{K}(1, r) + \sum_{r=\lfloor \delta \rfloor +1}^{\infty} \mathcal{K}(0, r) + \frac{(\lambda+\beta)^{\delta}}{\Gamma(\delta+1)} \\ \cdot \exp\left(-(\lambda+\beta)\right) \end{aligned} \tag{10}$$

where the factorial has been replaced by the Gamma function $\Gamma(\cdot)$. Note that we have used the *a priori* probability of each hypothesis in the definition of the average probability of error in (6). An alternate formulation could use the *a priori* probabilities to weight the above sample performance function instead. If the desired user transmits a preamble of bits, then by observing the sequence of correlated intensities, $(\alpha_n + \beta_n), \forall n = 1 \cdots N_c$, we can form the sample performance function in the above equation. We now present a proposition due to Ho and Cao [14, p. 65] on the unbiasedness of IPA estimates for discontinuous sample functions.

Proposition 1: Consider a sample performance function $L(\theta, \xi) = L_1(\theta, \xi) + L_2(\theta, \xi)$ such that $L_1(\theta, \xi)$ is a continuous and bounded function of θ, while $L_2(\theta, \xi)$ is a piecewise constant function of θ. Then the IPA estimate of the sensitivity of $L(\theta, \xi)$ with respect θ is unbiased if and only if the probability that $L(\theta, \xi)$ jumps in $[\theta, \theta + \Delta\theta]$ is $o(\Delta\theta)$, and the average height of the jump at θ equals zero.

The sample performance function in (10) is now comprised of three functions, two of which are piecewise constant functions of δ, and one of them is a continuous and bounded function of δ.

Proposition 2: The IPA estimate of the sensitivity of the probability of bit error to the detection threshold is unbiased.

Proof: The sample performance function in (10) jumps only at the discontinuities of the first two terms in (10). These jumps occur exactly once in every unit interval. Thus, the measure of the set of points of the jumps is zero. Therefore, the probability of a jump in $[\delta, \delta + \Delta\delta]$ is of the order of $P_{[\delta, \delta+\Delta\delta]} = o(\Delta\delta)$. Furthermore, if $d(\delta, h)$ denotes the probability density function for a jump of height h at δ, the average height of a jump at δ is given as

$$\int hd(\delta, h) \le |h|_{\max} P_{[\delta, \delta+\Delta\delta]} \le 2o(\Delta\delta)$$

where $|h|_{\max}$ is the maximum jump height and is bounded by two. This is due to the fact that $|P_e(\delta + \Delta\delta) - P_e(\delta)| \le 2$. The proof is now complete via Proposition 1. □

Thus, we have an IPA estimation-based method for designing optimal threshold detectors for optical CDMA systems. This scheme is not adaptive in the sense that the derivative during each step of the algorithm is evaluated by using IPA estimates averaged over several sample realizations. However, this technique could be used to design *a priori* the optimum threshold via discrete-event simulations of the users in the system.

B. Adaptive Single-User Detector

Having developed the above optimum threshold detector, we now extend the results from the previous section in order to develop an adaptive single-user detector that updates the threshold sequentially from bit to bit. In this algorithm, the IPA estimates are evaluated at each step of the algorithm from the corresponding sample realization without any averaging being used. We use a stochastic gradient algorithm that uses IPA estimates to sequentially update the detector threshold as follows. The threshold value at the mth bit is given as

$$\delta_m = \delta_{m-1} - \mu_m \frac{dP_e}{d\delta} \tag{11}$$

where $(dP_e/d\delta)$ is estimated using IPA from the corresponding sample realization. Furthermore, using Proposition 2 and the analyticity of the sample performance function P_e, it can be shown that

$$\frac{dP_e}{d\delta} = \frac{\{(\lambda+\beta)^{\delta-1}\,\delta - \boldsymbol{\Psi}(\delta+1)\,(\lambda+\beta)^{\delta}\}}{\Gamma(\delta+1)} \cdot \exp\left(-(\lambda+\beta)\right) \tag{12}$$

where $\boldsymbol{\Psi}(x)$ is the Digamma function [15]. Thus, we have a simple recursive update structure for implementing the above adaptive detector. We now state the following proposition on the convergence of the above adaptive algorithm.

Proposition 3: The adaptive algorithm in (11) converges in the sense that $\lim_{m\to\infty} \delta_m \xrightarrow{a.s.} \delta^*$, where

$$\delta^* = \arg \min_{\delta} \overline{P}_e.$$

The proof of the above proposition is presented in the Appendix. In terms of the practical implementation of the above detector, the desired user is required to transmit a training sequence or a preamble of bits. Furthermore, due to photon detection, the output of the correlator sampled at every bit interval is expressed as photoelectron counts which are conditionally Poisson. The correlated intensity $(\alpha+\beta)$ acts as the intensity of the Poisson process. Therefore, the above adaptive algorithm is easily implemented using maximum likelihood estimates of the correlated intensity based on observations of the photoelectron counts during each bit interval. This estimate, combined with the knowledge of the desired user's preamble, enables the construction of the sample derivative in (12). If r_m denotes the observed photoelectron count in the mth bit interval, then the sample gradient in (12) is computed as

$$\frac{dP_e}{d\delta} = \frac{\{(\lambda+\hat{\beta})^{\delta-1}\,\delta - \boldsymbol{\Psi}(\delta+1)\,(\lambda+\hat{\beta})^{\delta}\}}{\Gamma(\delta+1)} \cdot \exp\left(-(\lambda+\hat{\beta})\right) \tag{13}$$

where $\hat{\beta} = r_m - b_{1,m}\lambda$, and $b_{1,m}$ is the mth bit in the preamble transmitted by the desired user. Note that the detector requires knowledge of the desired user's code sequence since $\lambda = \langle \boldsymbol{\lambda}_1^{(1)}, \boldsymbol{\lambda}_1^{(1)} \rangle$.

As an extension of the above algorithm, we propose a sequential detection scheme which precludes the use of a preamble by the desired user. In this scheme, bit estimates of the desired user made during each interval are incorporated in computing the sample derivative required for the stochastic update algorithm in (11). Specifically, the sample gradient in (13) is computed with $\hat{\beta}$ being given as $\hat{\beta} = r_m - \hat{b}_{1,m}\lambda$, where

$$\hat{b}_{1,m} = \begin{cases} 1, & \text{if } r_m > \delta_{m-1} \\ 0, & \text{if } r_m \le \delta_{m-1} \end{cases}$$

where δ_{m-1} is the threshold value from the $(m-1)$th iteration. Therefore, in this sequential scheme, the desired user is not required to transmit a training sequence. All of the adaptive detectors presented here are initialized with the threshold for the correlation detector. In the following, the performance of both the one-shot optimal detector as well as the adaptive single-user detectors is presented.

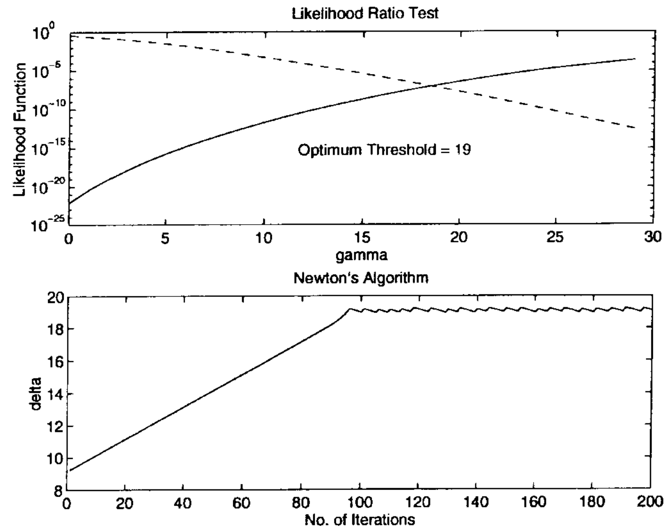

Fig. 1. Optimum detection for nondeterminant interferer codes. A 15-user optical CDMA system using optical codes of $\{500, 50\}$ with probability of an interferer of 0.1. In the top figure, $\gamma^* = 19$ which corresponds to the optimum threshold via the likelihood ratio rule. The results of Newton's algorithm using IPA estimates of $(d\overline{P}_e/d\gamma)$ is shown versus the number of iterations. It is seen that the algorithm converges to $\gamma^* = 19$.

C. Numerical Results

For the purposes of illustrating the threshold detection schemes, we consider an optical CDMA system where the interferer codes are $\{0, 1\}$ optical pulse sequences. The results of the optimal one-shot scheme developed using IPA estimates are shown in Fig. 1. For this example, a 15-user optical CDMA system employing optical pulses of length 500 and weight 50 with interferer probability of 0.1 is considered. The threshold for optimum detection is computed using the hybrid Newton's method described in (9). In this scheme, the IPA estimates at each step were computed by averaging over 100 trials. The parameter μ was chosen to be 0.01. The threshold via Newton's algorithm converges to $\gamma^* = 19$. This is exactly the optimum threshold computed via the likelihood ratio test given in (7). As an example of a system where the interference statistics cannot be determined, we consider an optical CDMA system where the users employ prime sequences [16]. In this case, the results of the IPA analysis via Newton's method are shown at the bottom of Fig. 2. As a means of cross validation, we see that the zero crossing is seen to occur at $\gamma = 36$. For prime sequences, evaluation of the likelihood ratio rule in (7) is not straightforward. Instead, at the top of Fig. 2, the Monte Carlo simulations of the probability of error are plotted versus the threshold for detection. It is seen that the minimum error probability once again occurs at $\gamma = 36$, which corresponds exactly to the threshold found from IPA analysis. The parameter μ was again set to 0.01. Thus, we have a method that is robust to nondeterminant interferer codes, and can be used as a one-shot detector for a wide variety of systems employing optical pulse sequences. We illustrate the performance of the adaptive single-user detector for OCDMA systems in Fig. 3. An 11-user system employing codes of length 1000 and weight 30, with the probability of an interferer being 0.2, is considered. It is seen that the adaptive

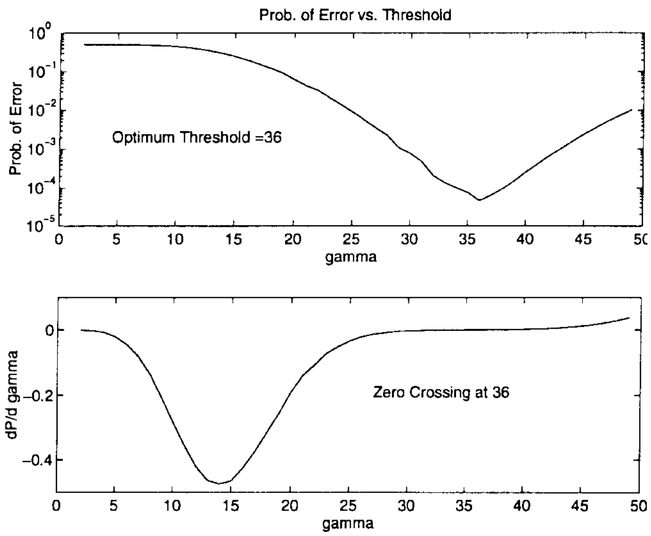

Fig. 2. Optimum detection for nondeterminant interferer codes. A 15-user optical CDMA system using prime sequences from GF (31). In the bottom figure, the IPA estimate of $(d\overline{P}_e/d\gamma)$ is plotted against γ. The zero crossing is at $\gamma^* = 36$, which corresponds to the optimum threshold via the brute-force simulations of the probability of bit error shown in the top figure for different thresholds.

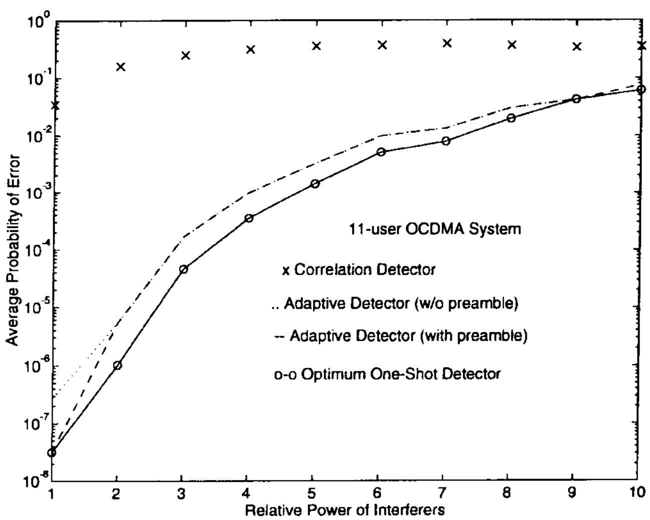

Fig. 3. Performance of the adaptive single-user detector. An 11-user optical CDMA system using optical pulse sequences of length 1000, weight 30, and probability of an interferer 0.2. The performance of the adaptive detector with and without a preamble is shown versus the correlation detector and the optimum one-shot detector for increasing power of the interferers.

detector is fairly resistant to unequal received power effects, unlike the correlation detector which shows a degradation in performance with increasing interferer powers. Both of the schemes with and without preamble are seen to closely mimic the performance of the optimum one-shot detector. The convergence of the adaptive detector is shown in Fig. 4 against the number of iterations, where the step sizes μ_m were chosen to be identically 0.01.

IV. Linear Detection

In this section, we develop a linear detection scheme for optical CDMA systems with the objective of achieving minimum probability of bit error for the desired user. For full bandwidth utilization in an optical CDMA system, it is required to have all-optical encoding and decoding since electronic bandwidth is limited. Many systems have been proposed that implement a correlation detector that is realized using a fiber tapped delay line (TDL) [4], [5]. However, as pointed out earlier, the correlation detector's performance is severely degraded by an increasing number of users in the system. To overcome this drawback, all-optical linear multiuser detectors have been proposed [6], [7], which were implemented using strictly 0–1 weights for the filter taps (i.e., the weights of the fiber TDL are either 0 or 1). However, these require explicit knowledge of the multiple-access interference. To overcome this drawback, we propose an IPA-based stochastic gradient algorithm to implement an adaptive linear detector. In this scheme, the weights of the fiber TDL are varied from bit to bit to adapt to the multiple-access interference, with the objective of achieving minimum probability of bit error. Further, the weights are constrained to take values in [0, 1], and are not necessarily restricted to be strictly 0–1 sequences.

A. Adaptive Linear Detector

We will first establish the following preliminaries in order to formulate the appropriate sample performance function and the sample gradients. Following the notation in Section II, the sample performance function under hypothesis H_i is given as

$$P_{e,i} = \sum_{r \in \Delta_{1-i}} \mathcal{H}(i, r), \qquad i = 0, 1 \tag{14}$$

where Δ_i is the decision region for the photoelectron count r that hypothesis H_i is true, and the kernel $\mathcal{H}(i, r)$ is given by

$$\mathcal{H}(i, r) = \frac{\left(\boldsymbol{h}^\top \left(\boldsymbol{\lambda}_i^{(1)} + \sum_{k=2}^{K} \boldsymbol{\lambda}_b^{(k)}\right)\right)^r}{r!} \cdot \exp\left(-\boldsymbol{h}^\top \left(\boldsymbol{\lambda}_i^{(1)} + \sum_{k=2}^{K} \boldsymbol{\lambda}_b^{(k)}\right)\right) \tag{15}$$

where $\boldsymbol{h}$ is the vector of the taps of the linear filter. Therefore, $\boldsymbol{h}$ has N_c elements, each of which corresponds to the weight of the fiber TDL for the appropriate chip interval. Note that $\mathcal{H}(i, r)$ is exactly the same as the sample performance function in (4), but we use a different notation to highlight the fact that the linear filter taps are now being considered as the parameter vector. The linear detection scheme is based on finding the best vector $\boldsymbol{h}^*$ such that

$$\begin{aligned}\boldsymbol{h}^* &= \arg\min_{\boldsymbol{h}} \overline{P}_e \\ &= \arg\min_{\boldsymbol{h}} \left\{\sum_{r=0}^{\gamma} E[\mathcal{H}(1, r)] + \sum_{r=\gamma+1}^{\infty} E[\mathcal{H}(0, r)]\right\}. \end{aligned} \tag{16}$$

Since we want to establish a scheme that is not dependent on explicit information about the multiple-access interference, we use an IPA-based stochastic gradient algorithm as

$$\boldsymbol{h}_{m+1} = \boldsymbol{h}_m - \mu_m \nabla_{\boldsymbol{h}} P_e \tag{17}$$

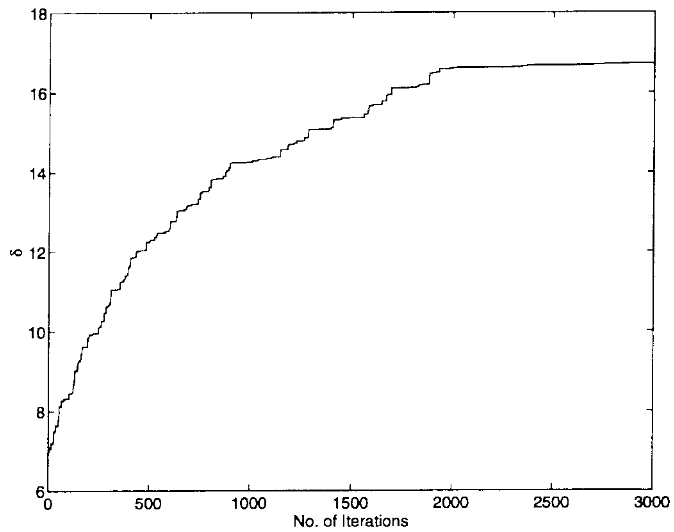

Fig. 4. Performance of the adaptive single-user detector. A three-user optical CDMA system using optical pulse sequences of length 1000, weight 40, and probability of an interferer 0.2. The convergence of the adaptive detector threshold is shown against the number of iterations for the case of constant step size.

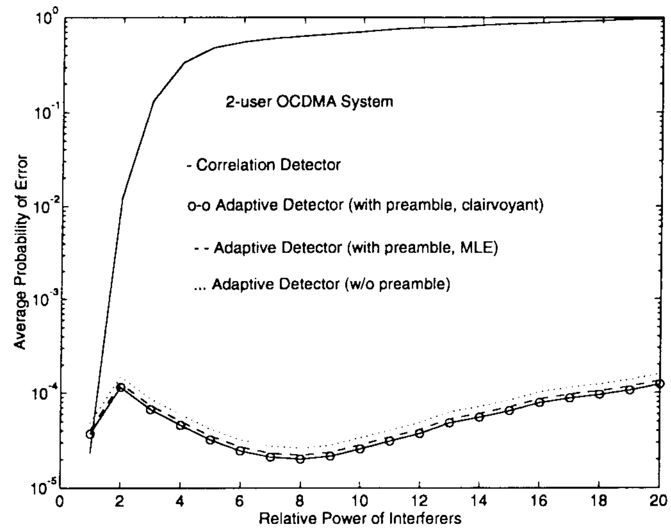

Fig. 5. Performance of the adaptive linear detectors. A two-user optical CDMA system using optical pulse sequences of length 32, weight 4, and probability of an interferer 0.5. The performance of the adaptive detectors with and without a preamble is shown versus the correlation detector for increasing power of the interferers.

where $\nabla_{\boldsymbol{h}} P_e$ is an IPA estimate that is evaluated during each bit interval directly from the sample path. It is easily shown that the IPA estimate is unbiased, and furthermore, the sample derivative is once again very easily computed, yielding a simple recursive structure as follows. The partial derivatives $\partial \mathcal{H}(i, r)/\partial h_n$ follow a very simple recursion as

$$\frac{\partial \mathcal{H}(i,r)}{\partial h_n} = [\mathcal{H}(i, r-1) - \mathcal{H}(i, r)] \left(\boldsymbol{\lambda}_i^{(1)} + \sum_{k=2}^{K} \boldsymbol{\lambda}_b^{(k)} \right)_n \tag{18}$$

where $(\boldsymbol{\lambda}_i^{(1)} + \sum_{k=2}^{K} \boldsymbol{\lambda}_b^{(k)}]_n$ is the received signal in the nth chip interval. Using (18), it follows that

$$\frac{\partial P_e}{\partial h_n} = -\mathcal{H}(1, \gamma) \left(\boldsymbol{\lambda}_1^{(1)} + \sum_{k=2}^{K} \boldsymbol{\lambda}_b^{(k)} \right)_n + \mathcal{H}(0, \gamma) \left(\sum_{k=2}^{K} \boldsymbol{\lambda}_b^{(k)} \right)_n. \tag{19}$$

As in the case of the threshold detection schemes, the above sample gradient can be constructed if the desired user transmits a preamble or a stochastic training sequence. The maximum likelihood estimation of the intensities follows exactly as before. A sequential implementation that uses bit estimates made in each interval to form the sample performance functions also follows as in the case of the threshold detection scheme.

Proposition 4: The adaptive algorithm in (17) converges in the sense that $\lim_{m \to \infty} \boldsymbol{h}_m \xrightarrow{a.s.} \boldsymbol{h}^*$, where

$$\boldsymbol{h}^* = \arg \min_{\boldsymbol{h}} \overline{P}_e.$$

The proof of the above proposition and the conditions required for convergence are presented in the Appendix.

B. Numerical Results

A two-user optical CDMA system is considered (see Fig. 5) where the users use codes of length 32 with weights being 4. The correlation detector is seen to degrade rapidly as the interferer power gets stronger relative to that of the desired user. The adaptive linear detector using a preamble and perfect knowledge of the correlated intensities (which is why we have named it clairvoyant) is shown to be fairly near–far resistant to unequal received power effects. The schemes using joint maximum likelihood estimation and no preamble are also seen to closely mimic the performance of the clairvoyant detector. In Fig. 6, in order to illustrate the algorithm dynamics, a two-user synchronous system is considered with the users employing optical pulse sequences of length 10. The filter taps of the linear filter are shown at convergence, where it is seen that for the chips where the interferer contributes to the correlation, the filter taps are nearly zero, thereby reducing the multiple-access interference. To further illustrate the dynamics of the algorithm, the cross correlation between the interference vector and the linear filter taps are shown in Fig. 7 as a function of the number of iterations of the algorithm. It is seen that at every step, the algorithm dynamics are such that it reduces this correlation value. For all of the above cases, the step sizes required for the algorithm were set to 0.01. Using variable step sizes can accelerate the convergence of the above algorithms; however, this may not be attractive from a practical standpoint. The cross-correlation value due to the correlation detector is shown for reference, which serves as the initial point for the algorithm.

V. Conclusions

Here, a DEDS formulation was used to develop infinitesimal perturbation analysis for estimating the sensitivity of the average probability of error to the threshold for photodetection. These estimates were used to develop an optimum one-shot

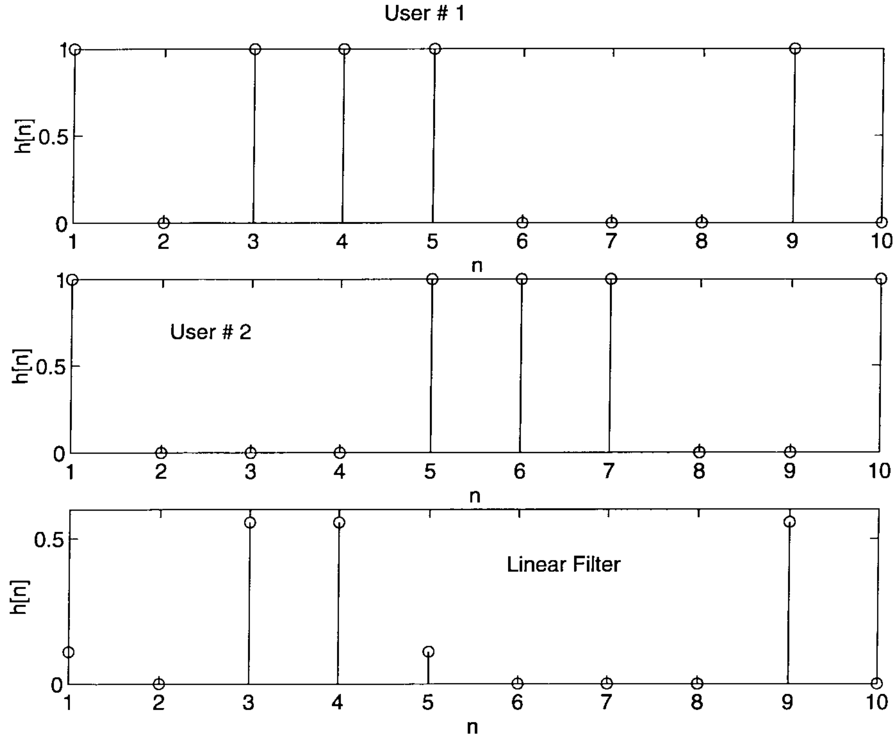

Fig. 6. Dynamics of the adaptive linear detection algorithm. As an example, two-user optical CDMA system using optical pulse sequences of length $N_c = 10$ is shown. The linear filter tap values $h[n]$ are shown versus $n = 1, \cdots, N_c$ at convergence for the case when the interferer (user 2) is ten times stronger than the desired user 1.

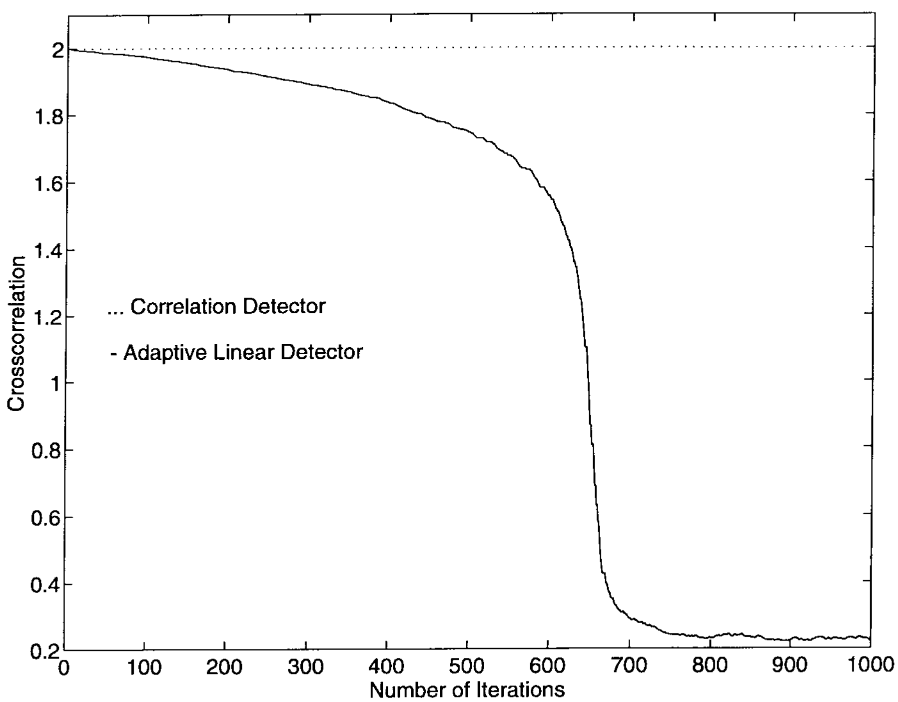

Fig. 7. Dynamics of the adaptive linear detection algorithm. The cross-correlation values between the interference vector and the linear filter taps are shown as a function of the number of iterations of the algorithm. It is seen that at every step, the algorithm dynamics are such that the cross-correlation value is reduced away from that of the correlation detector.

threshold detector, as well as adaptive threshold detectors that minimized the average probability of error by choosing the optimum threshold for photodetection. A class of linear detectors was also developed where the filter taps were adaptively updated so as to minimize the average probability of bit error. These detectors alleviate the disadvantage of the previously proposed single-user and multiuser schemes that require explicit information about the multiple-access interference. Owing to the simplicity of the adaptive detectors proposed here, these IPA-based stochastic gradient algorithms can be used as a viable alternative to the correlation detector as well as multiuser schemes. These adaptive algorithms can be employed for systems subject to temporal variations due to dispersion effects and a variable number of users in the channel.

APPENDIX

A. Stochastic Gradient Algorithms Using IPA

Consider a discrete-event dynamic system that is represented by the (vector) pair $(\boldsymbol{\theta}, \boldsymbol{\zeta})$ with $J(\boldsymbol{\theta}) = E[L(\boldsymbol{\theta}, \boldsymbol{\zeta})]$ as the expected performance measure of interest. It is required to find the parameter $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$, which can be solved via an algorithm of adaptation [17] where the nth step of the iterative algorithm is given as

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - \gamma_n \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_{n-1}, \boldsymbol{\zeta}_n) \tag{20}$$

where $\boldsymbol{\theta}_n$ is the nth update of the parameter vector, $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_{n-1}, \boldsymbol{\zeta}_n)$ is the gradient evaluated using infinitesimal perturbation analysis, and γ_n is the step size. For such an algorithm, we are interested in the convergence of the adaptive algorithm to the minimizer. We now present the following lemma [17, p. 56] on the convergence of such an algorithm.

Lemma 1 (Tsypkin): The sufficient conditions for the algorithm of adaptation in (20) to converge almost surely, i.e., $\lim_n \boldsymbol{\theta}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}^*$, are

$$\gamma_n > 0, \sum_{n=1}^{\infty} \gamma_n = \infty, \sum_{n=1}^{\infty} \gamma_n^2 < \infty \tag{21}$$

$$\inf_{\epsilon < \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < 1/\epsilon} E[(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\zeta})] > 0; \tag{22}$$

$$E[\nabla_{\boldsymbol{\theta}}^\top L(\boldsymbol{\theta}, \boldsymbol{\zeta}) \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\zeta})] \le \alpha(1 + \boldsymbol{\theta}^\top \boldsymbol{\theta}) \tag{23}$$

where $\boldsymbol{\theta}^* = \arg\{\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta})\}$, and the operation $\boldsymbol{\theta}^\top$ denotes the transpose of the vector. □

A very simple physical and geometric meaning of the above conditions is as follows. The condition in (21) requires that the rate of decrease in γ_n is such that the variance of the estimate of $J(\boldsymbol{\theta})$ is reduced to zero. The condition in (22) requires that the gradient near the optimum value is always negative, and hence defines the behavior of the surface $E[\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta})]$ near zero. This is, in a sense, a requirement on the convexity of the objective function near the minimum. Finally, the condition in (23) implies that the mathematical expectation of the quadratic form $E[\nabla_{\boldsymbol{\theta}}^\top L(\boldsymbol{\theta}, \boldsymbol{\zeta}) \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\zeta})]$ has a rate of increase with $\boldsymbol{\theta}$ smaller than in a parabola of second degree.

B. Convergence of the Threshold Detection Algorithm

We now sketch a proof of Proposition 3, i.e., we will show the convergence of the stochastic gradient algorithm in (11) for adaptive optimum single-user detection of optical CDMA systems. For convenience, we rewrite the algorithm given as

$$\delta_n = \delta_{n-1} - \mu_n \frac{dP_e}{d\delta} \tag{24}$$

where the sample performance function is given as in (10)

$$P_e = \sum_{r=0}^{\lfloor \delta \rfloor - 1} \mathcal{K}(1, r) + \sum_{r=\lfloor \delta \rfloor + 1}^{\infty} \mathcal{K}(0, r) + \frac{(\lambda + \beta)^\delta}{\Gamma(\delta + 1)} \cdot \exp(-(\lambda + \beta)) \tag{25}$$

and the algorithm is said to converge when

$$\lim_{n \to \infty} \delta_n \xrightarrow{\text{a.s.}} \delta^*$$

where $\delta^* = \arg\{\min_\delta \overline{P}_e\}$. By Lemma 1, the sufficient conditions for the algorithm in (24) to converge almost surely are given as

$$\mu_n > 0, \sum_{n=1}^{\infty} \mu_n = \infty, \sum_{n=1}^{\infty} \mu_n^2 < \infty; \tag{26}$$

$$\inf_{\epsilon < \|\delta - \delta^*\| < 1/\epsilon} E\left[(\delta - \delta^*) \frac{dP_e(\delta, \beta)}{d\delta}\right] > 0; \tag{27}$$

$$E\left[\left(\frac{dP_e(\delta, \beta)}{d\delta}\right)^2\right] \le \alpha(1 + \delta^2) \tag{28}$$

where $P_e(\delta, \beta)$ is as given in (25) with $\beta = \sum_{n=1}^N \beta_n$. The first of the above sufficient conditions is satisfied by choosing the update step size μ_n appropriately. As in the previous section, the second condition is a convexity requirement on the average probability of error, and is usually satisfied if the interference statistics are such that their likelihood ratio is monotonic. This translates to the fact that the likelihood ratio test for detecting between the two hypotheses yields decision regions that are simply connected. In the absence of this condition, the algorithm will converge to a local minimum. We will now prove the third condition stated above. We observe that the derivative of the sample performance function is given as

$$\frac{dP_e(\delta, \beta)}{d\delta} = \frac{d}{d\delta}\left\{\frac{(\lambda + \beta)^\delta \exp(-(\lambda + \beta))}{\Gamma(\delta + 1)}\right\} \tag{29}$$

which can be written as

$$\frac{dP_e(\delta, \beta)}{d\delta} = \frac{\{(\lambda + \beta)^{\delta - 1} \delta - \boldsymbol{\Psi}(\delta + 1)(\lambda + \beta)^\delta\}}{\Gamma(\delta + 1)} \cdot \exp(-(\lambda + \beta)) \tag{30}$$

where $\boldsymbol{\Psi}(x)$ is the Digamma function defined as

$$\boldsymbol{\Psi}(x) = \frac{d}{dx} \ln(\Gamma(x)).$$

Therefore, we can write

$$E\left[\left(\frac{dP_e(\delta, \beta)}{d\delta}\right)^2\right] = E\left[\left\{\frac{\exp(-(\lambda + \beta))(\lambda + \beta)^\delta}{\Gamma(\delta + 1)}\right\}^2 \cdot \left\{\frac{\delta}{\lambda + \beta} - \boldsymbol{\Psi}(\delta + 1)\right\}^2\right] \tag{31}$$

which can be bounded from above as

$$E\left[\left(\frac{dP_e(\delta, \beta)}{d\delta}\right)^2\right] \le E\left[\left\{\frac{\exp(-(\lambda + \beta))(\lambda + \beta)^\delta}{\Gamma(\delta + 1)}\right\}^2 \left\{\frac{\delta}{(\lambda + \beta)}\right\}^2\right], \quad \forall\, \delta > 0.462 \tag{32}$$

where we have used the fact that $\boldsymbol{\Psi}(\delta + 1) > 0, \forall\, \delta > 0.462$, i.e., for all values of the detection threshold $\gamma = \lfloor \delta \rfloor$ greater than zero. This is the range that the detection threshold actually takes, since the adaptive algorithm is initialized with the correlation detector threshold which is greater than zero, even when there are no interferers in the channel owing to the dark current. The above equation can be further rewritten as

$$E\left[\left(\frac{dP_e(\delta,\beta)}{d\delta}\right)^2\right] \le E\left[\left\{\frac{\delta}{\lambda+\beta}\right\}^2\right] \tag{33}$$

where we have used the fact that $\{(\exp(-(\lambda+\beta))(\lambda+\beta)^\delta/\Gamma(\delta+1))\}^2 \le 1$. A simple algebraic manipulation of (33) results in

$$E\left[\left(\frac{dP_e(\delta,\beta)}{d\delta}\right)^2\right] \le (1+\delta^2)\alpha, \qquad \alpha > 0 \tag{34}$$

where $\alpha = \min\{1/(\lambda+\beta)^2\}$, and this quantity is always bounded below by zero, since we are considering a finite number of users in the system with intensities always less than infinity. Thus, the condition in (28) is satisfied.

C. Convergence of the Linear Detection Algorithm

Here, we will sketch a proof of Proposition 4, i.e., we will list the conditions under which the stochastic gradient algorithm in (17) converges. The algorithm is given as

$$\boldsymbol{h}_{m+1} = \boldsymbol{h}_m - \mu_m \nabla_{\boldsymbol{h}} P_e \tag{35}$$

where the sample performance function P_e is given as

$$P_e = \sum_{r=0}^{\gamma} \mathcal{H}(1,r) + \sum_{r=\gamma+1}^{\infty} \mathcal{H}(0,r) \tag{36}$$

with $\mathcal{H}(1,r)$ as given in (15). Here, we will prove the convexity condition stipulated as (by Lemma 1)

$$\inf_{\epsilon<\|\boldsymbol{h}-\boldsymbol{h}^*\|<1/\epsilon} E[(\boldsymbol{h}-\boldsymbol{h}^*)^\top \nabla_{\boldsymbol{h}} P_e] > 0. \tag{37}$$

Without loss of generality, let $\boldsymbol{h} = \boldsymbol{h}^* + \delta\boldsymbol{v}$, where $\delta > 0$, and $\boldsymbol{v}$ is a unit vector. Then the left-hand side of (37) is given as

$$E[-\delta\boldsymbol{v}^\top \nabla_{\boldsymbol{h}} P_e] = E\left[\delta\left(\sum_{n=1}^{N_c}\left(\boldsymbol{\lambda}_1^{(1)} + \sum_{k=2}^{K}\boldsymbol{\lambda}_b^{(k)}\right)_n\right)\mathcal{H}(1,\gamma) - \delta\left(\sum_{n=1}^{N_c}\left(\sum_{k=2}^{K}\boldsymbol{\lambda}_b^{(k)}\right)_n\right)\mathcal{H}(0,\gamma)\right]. \tag{38}$$

If the threshold is chosen optimally, i.e., according to a likelihood ratio test, it follows for equiprobable hypothesis that

$$\mathcal{H}(1,\gamma) \approx \mathcal{H}(0,\gamma) \tag{39}$$

which implies that (37) is satisfied. The quadratic condition cannot be explicitly checked for in this case, however, the stochastic gradient algorithm is based on a Newton-type descent method, and hence appropriate choice of the step size will guarantee the algorithm to proceed so as to ensure that the local Taylor's series approximation holds at every step.

References

[1] N. B. Mandayam and B. Aazhang, "Gradient estimation for stochastic optimization of optical code-division multiple access systems: Part I—Generalized sensitivity analysis," this issue, pp. 731–741.

[2] M. Brandt-Pearce and B. Aazhang, "Unequal received power effects on single-user and multi-user detection of optical code division multiple access systems," in *Proc. 22nd Annu. Conf. Inform. Sci. Syst.*, Princeton Univ., Princeton, NJ, Mar. 1992.

[3] S. Verdu, "Multiple-access channels with point-process observations: Optimum demodulation," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 642–651, Sept. 1986.

[4] J. Hui, "Pattern code modulation and optical decoding—A novel code-division multiplexing technique for multifiber networks," *IEEE J. Select. Areas Commun.*, vol. SAC-3, pp. 916–927, Nov. 1985.

[5] P. R. Prucnal, M. A. Santoro, and S. K. Sehgal, "Ultrafast all-optical synchronous multiple access fiber network," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 1484–1493, Dec. 1986.

[6] D. Brady, "Asymptotic multiuser efficiency for optical channels," in *Proc. 21st Annu. Conf. Inform. Sci. Syst.*, Mar. 1991.

[7] M. Brandt-Pearce, "All-optical linear detection for optical CDMA communication systems," in *Proc. Int. Symp. Inform. Theory*, Trondheim, Norway, June 1994.

[8] N. B. Mandayam and B. Aazhang, "Generalized sensitivity analysis of optical code division multiple access systems," in *1993 Conf. Inform. Sci. Syst.*, Mar. 1993.

[9] ——, "An adaptive single-user detector for optical code division multiple access systems," in *1994 Ann. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994.

[10] J. A. Salehi, "Code division multiple access techniques in optical fiber networks: Part I—Fundamental principles," *IEEE Trans. Commun.*, vol. 37, no. 8, pp. 824–833, Aug. 1989.

[11] A. W. Lam and A. M. Hussain, "Performance analysis of code division multiple access optical communications with avalanche photodetectors," in *Proc. Conf. Inform. Sci. Syst.*, Mar. 1989.

[12] M. Brandt-Pearce and B. Aazhang, "Performance analysis of single-user and multi-user detectors for optical code division multiple access communication systems," *IEEE Trans. Commun.*, vol. 43, pp. 435–444, Feb./Mar./Apr. 1995.

[13] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall Series in Computational Mathematics). Englewood Cliffs, NJ: Prentice-Hall, 1983.

[14] Y.-C. Ho and X.-R. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems*. Boston: Kluwer Academic, 1991.

[15] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1970.

[16] P. A. Davies and A. A. Shaar, "Prime sequences: Quasioptimal sequences for OR channel code division multiplexing," *Electron Lett.*, vol. 19, pp. 888–890, Oct. 13, 1983.

[17] Y. Z. Tsypkin, *Adaptation and Learning in Automatic Systems*. New York: Academic, 1971.

Narayan B. Mandayam (M'94), for a photograph and biography, see this issue, p. 741.

Behnaam Aazhang (S'81–M'85–SM'91) for a photograph and biography, see this issue, p. 741.