

Expectation and Maximization Algorithm for Estimating Parameters of a Simple Partial Erasure Model

Tsai-Sheng Kao and Mu-Huo Cheng

Abstract—The identification of the model parameters of a high-density recording channel generally requires solution of nonlinear equations. In this paper, we apply the expectation and maximization (EM) algorithm to realize the maximum likelihood estimation of the parameters of a simple partial erasure model, including the reduction parameters and the isolated transition response. The algorithm that results from this approach iteratively solves two least-squares problems and, thus, realization is simple. Computer simulations verify the feasibility of the EM algorithm, and show that the proposed algorithm has fast convergence and the resulting estimator is asymptotically efficient.

Index Terms—Expectation and maximization algorithm, least-squares methods, maximum-likelihood estimation, Monte Carlo methods, simple partial erasure model.

I. INTRODUCTION

AS RECORDING densities grow in magnetic storage, nonlinear distortions become the primary factors to limit the detector performance. Nonlinear distortions in magnetic storage are mainly the nonlinear transition shift and the partial erasure. A magnetic recording channel model that represents nonlinear distortions accurately enables one to design an improved detector [1]–[3]. Several models have been presented to characterize the nonlinear distortions [4]–[6]. It is known that if the channel model is more complicated, the complexity of detector for the model may be prohibitively higher. A simple partial erasure model [7], which is simplified from the transition-width-reduction model in [4], is shown to preserve sufficient accuracy and is much simpler than the other models, and, thus, is often adopted to design the detector.

Accurate estimation of model parameters is crucial in designing a detector of high performance such that high recording density can be achieved. Techniques widely used for identifying nonlinear distortions are the echo extraction method using pseudorandom binary sequence input [8], the autocorrelation method [9], the frequency domain nonlinear measurement method [10], and an adaptive technique using the stochastic gradient [11]. These methods, however, either require the input with a special pattern, or have to compute the gradient vector and determine the step size; thus, the complexity is often high and the convergence is not assured.

In this paper, we focus on the simple partial erasure model and apply the expectation and maximization (EM) algorithm [12]–[14] for estimating the model parameters, including the reduction parameters and samples of the isolated transition response. The EM algorithm ensures the convergence and obtains the maximum-likelihood (ML) estimates of the channel parameters [15]. At each iteration, the EM algorithm is required to solve two least-squares problems. Thus, the EM algorithm for this problem has simple realization. Simulation results also demonstrate that fast convergence speed of the proposed algorithm is obtained.

The rest of paper is arranged as follows. In Section II, a simple partial erasure model is reviewed and the sampled output in terms of the model parameters is formulated. In Section III, the problem is first formulated as the joint ML estimation of the model parameters. Then, the EM algorithm is applied to solve the joint ML estimation iteratively. In Section IV, the feasibility of the EM algorithm is verified by computer simulations. Finally, a conclusion is given in Section V.

II. SIMPLE PARTIAL ERASURE CHANNEL MODEL

Assume that the nonlinear transition shift has been eliminated by the precompensation technique [16]. In terms of the simple partial erasure model, the playback signal of a high-density recording channel can be expressed as

$$y(t) = \sum_k r_k b_k h(t - kT) + n(t) \quad (1)$$

where the effective transition-width ratio r_k is determined by the neighboring transitions and is given by

$$r_k = \begin{cases} 1, & b_{k-1} = b_{k+1} = 0 \\ \gamma_1, & |b_{k-1}| \neq |b_{k+1}| \\ \gamma_2, & |b_{k-1}| = |b_{k+1}| = 2. \end{cases} \quad (2)$$

The nonreturn-to-zero-inverted (NRZI) modulated input b_k with the period T may be of values $\{+2, 0, -2\}$. The term $h(t)$ represents the isolated transition response, which is often modeled as the Lorentzian function

$$h(t) = \frac{1}{1 + \left(\frac{2t}{PW_{50}}\right)^2} \quad (3)$$

where the normalized density PW_{50} quantifies the recorded bit density. Note that here we assume $h(t)$ is unknown and its samples are to be estimated. The last term $n(t)$ in (1) represents the

Manuscript received February 6, 2002; revised July 26, 2002.

The authors are with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: mhcheng@cc.nctu.edu.tw).

Digital Object Identifier 10.1109/TMAG.2002.806343

measurement noise. The simple partial erasure model characterizes the channel nonlinearity by the reduction parameters γ_1 and γ_2 ; when $\gamma_1 = \gamma_2 = 1$, the model is reduced to a linear model. The simple partial erasure model commonly sets $\gamma_2 = \gamma_1^2$. Here, we use two independent variables to represent the model; thus, the modeling flexibility is enhanced.

The sampled output $y_m = y(mT)$ can be obtained from (1)

$$y_m = \sum_k r_{m-k} b_{m-k} h(kT) + n(mT) \quad (4)$$

$$= \sum_{k=-L}^L r_{m-k} b_{m-k} h_k + n_m \quad (5)$$

where $h_k = h(kT)$, $n_m = n(mT)$, and L is chosen such that $h(kT)$ for $k > L$ are small enough to be neglected.

Denote $\mathbf{h} = [h_{-L}, \dots, h_0, \dots, h_L]^T$, $d_m = r_m b_m$, and $\mathbf{d}_m = [d_{m+L}, \dots, d_m, \dots, d_{m-L}]^T$, where the subscript T represents the transpose operation; then, the sampled output (5) can be rewritten as follows:

$$y_m = \mathbf{d}_m^T \mathbf{h} + n_m. \quad (6)$$

This formulation illustrates the relation between the output y_m and the isolated transition response \mathbf{h} . We can obtain another expression emphasizing the effect of reduction parameters to represent the sampled output as

$$y_m = x_m + \mathbf{c}_m^T \boldsymbol{\lambda} + n_m \quad (7)$$

where $\boldsymbol{\lambda} = [\gamma_1, \gamma_2]^T$, $\mathbf{c}_m = [\alpha_m, \beta_m]^T$, and

$$x_m = \sum_{k=-L}^L b_{m-k} f_1(b_{m-k-1}, b_{m-k+1}) h_k \quad (8)$$

where

$$f_1(b_{m-k-1}, b_{m-k+1}) = \begin{cases} 1, & b_{m-k-1} = b_{m-k+1} = 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$\alpha_m = \sum_{k=-L}^L b_{m-k} f_2(b_{m-k-1}, b_{m-k+1}) h_k \quad (10)$$

where

$$f_2(b_{m-k-1}, b_{m-k+1}) = \begin{cases} 1, & |b_{m-k-1}| \neq |b_{m-k+1}| \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$\beta_m = \sum_{k=-L}^L b_{m-k} f_3(b_{m-k-1}, b_{m-k+1}) h_k \quad (12)$$

where

$$f_3(b_{m-k-1}, b_{m-k+1}) = \begin{cases} 1, & |b_{m-k-1}| = |b_{m-k+1}| = 2 \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Note that the formulation (7) expresses the relation between the output y_m and the reduction parameters $\boldsymbol{\lambda}$. The two different expressions (6) and (7) of the output sample will be used in the

application of the EM algorithm. In (6), the reduction parameters $\boldsymbol{\lambda}$ are buried in the data vector \mathbf{d}_m , while in (7), the isolated transition response \mathbf{h} are buried in the coefficient vector \mathbf{c}_m .

The problem considered here is to estimate the isolated transition response \mathbf{h} and the reduction parameters $\boldsymbol{\lambda}$ from the given N samples y_1, \dots, y_N under the assumption that all concerned b_k s are known. In Section III, we shall derive how the EM algorithm is applied to obtain the estimates of \mathbf{h} and $\boldsymbol{\lambda}$.

III. JOINT ML ESTIMATION OF MODEL PARAMETERS VIA EM ALGORITHM

In this section, the ML estimation of the model parameters is first formulated. Then, the direct solution using this formulation is shown to require for solving highly nonlinear and intractable equations. Finally, the EM algorithm for realizing the estimator is derived and discussed.

A. ML Estimator

Using the representation (6) to express the N sample data in a matrix form, we obtain

$$\mathbf{y} = D\mathbf{h} + \mathbf{n} \quad (14)$$

where the sample data vector $\mathbf{y} = [y_1, \dots, y_N]^T$, the noise vector $\mathbf{n} = [n_1, \dots, n_N]^T$, and the matrix $D = [\mathbf{d}_1, \dots, \mathbf{d}_N]^T$. Assume that the noise sample data are identically independent Gaussian distribution with zero-mean and of variance σ^2 . The likelihood function is thus derived

$$p(\mathbf{y} | \mathbf{h}, \boldsymbol{\lambda}) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(\frac{-1}{2\sigma^2} \|\mathbf{y} - D\mathbf{h}\|_2^2\right) \quad (15)$$

where $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$. The ML estimates of \mathbf{h} and $\boldsymbol{\lambda}$ can be obtained directly by solving the equations derived by setting the derivatives of the logarithm of the likelihood function (15) with respect to \mathbf{h} and $\boldsymbol{\lambda}$ to zeros. Note that the reduction parameters $\boldsymbol{\lambda}$ are buried in the data matrix D , so this approach is required to solve nonlinear and intractable equations. Therefore, a better algorithm should be developed.

B. EM Algorithm for Estimating Model Parameters

The EM algorithm, developed in 1976 [12], is well known for its power to solve nonlinear intractable ML problems via iterations of two simple steps: the expectation step (E-step), and the maximization step (M-step). Besides the ensured convergence, the EM algorithm often demands low realization complexity. Here, we apply the EM algorithm for finding the ML estimates of \mathbf{h} and $\boldsymbol{\lambda}$. The E-step is formulated for estimating the reduction parameters $\boldsymbol{\lambda}$, then the M-step is used to obtain the estimate of \mathbf{h} .

Using the terminology in [12], we set the reduction parameters $\boldsymbol{\lambda}$ as the hidden data; the measurement data \mathbf{y} is incomplete because we cannot obtain the ML estimate of \mathbf{h} via \mathbf{y} without knowing the hidden data $\boldsymbol{\lambda}$. The main theme of the EM algorithm is on the following auxiliary function:

$$Q(\mathbf{h} | \hat{\mathbf{h}}) = E[\log p(\mathbf{y}, \boldsymbol{\lambda} | \mathbf{h}) | \mathbf{y}, \hat{\mathbf{h}}] \quad (16)$$

$$= \int_{\boldsymbol{\lambda}} \log p(\mathbf{y}, \boldsymbol{\lambda} | \mathbf{h}) p(\boldsymbol{\lambda} | \mathbf{y}, \hat{\mathbf{h}}) d\boldsymbol{\lambda} \quad (17)$$

where $E(\cdot)$ denotes the expectation operator. Given an initial estimate $\hat{\mathbf{h}}$ of the parameter \mathbf{h} , the E-step in the EM algorithm evaluates the auxiliary function $Q(\mathbf{h}|\hat{\mathbf{h}})$, yielding the byproduct $\hat{\boldsymbol{\lambda}}$, the estimate of $\boldsymbol{\lambda}$. The M-step then finds the estimate of \mathbf{h} for maximizing the auxiliary function $Q(\mathbf{h}|\hat{\mathbf{h}})$. This estimate of \mathbf{h} in the M-step is used again in the E-step as $\hat{\mathbf{h}}$ for evaluating the auxiliary function; then the M-step follows. Such iteration continues until the estimates converge. The convergence value of the estimate $\hat{\mathbf{h}}$ in the M-step and the byproduct $\hat{\boldsymbol{\lambda}}$ in the E-step are the desired results. It is shown in [15] that at each iteration in the EM algorithm, these estimated parameters are obtained so that the likelihood function in (15) is nondecreasing, and, thus, the convergence is guaranteed.

As seen from (17), different realization algorithms of the EM algorithm emerged when different distributions of the hidden data are assumed. Here, we assume that $\boldsymbol{\lambda}$ is an unknown, deterministic vector, then the detailed algorithm and its derivation of using the EM algorithm for our problem are briefly discussed as follows.

Step 1. Initialization:

Given an initial estimate $\hat{\mathbf{h}}$.

Step 2. E-step (Expectation):

Using the representation (7) and the estimate $\hat{\mathbf{h}}$, we can express the N measurements in the matrix form

$$\mathbf{y} = \hat{\mathbf{x}} + \hat{\mathbf{C}}\boldsymbol{\lambda} + \mathbf{n} \quad (18)$$

where $\hat{\mathbf{C}} = [\hat{c}_1, \dots, \hat{c}_N]^T$, $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_N]^T$, and \hat{c}_k , \hat{x}_k are obtained using (9)–(13) with h_k replaced by \hat{h}_k . Then the estimate of $\boldsymbol{\lambda}$ is derived by the least-squares method, yielding

$$\hat{\boldsymbol{\lambda}} = (\hat{\mathbf{C}}^T \hat{\mathbf{C}})^{-1} \hat{\mathbf{C}}^T (\mathbf{y} - \hat{\mathbf{x}}). \quad (19)$$

Since $\boldsymbol{\lambda}$ is assumed deterministic and the noise sample data are independent normal distributed, using the representation (6) and replacing $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\lambda}}$ we obtain the Q function (17)

$$Q(\mathbf{h}|\hat{\mathbf{h}}) = K(\mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \hat{\mathbf{D}}^T \mathbf{h} + \mathbf{h}^T \hat{\mathbf{D}}^T \hat{\mathbf{D}} \mathbf{h}) \quad (20)$$

where K is a constant and $\hat{\mathbf{D}}$ is obtained using (5), (6) and (14) with $\boldsymbol{\lambda}$ replaced by $\hat{\boldsymbol{\lambda}}$.

Step 3. M-step (Maximization):

Once the Q is obtained, the M-step is to find the estimate \mathbf{h} for maximizing the function, resulting in

$$\hat{\mathbf{h}} = (\hat{\mathbf{D}}^T \hat{\mathbf{D}})^{-1} \hat{\mathbf{D}}^T \mathbf{y}. \quad (21)$$

The EM-algorithm iteratively executes the E-step and M-step until convergence. The convergence is often checked by evaluating the measure of the difference between the estimated parameters of successive iterations. When the measure is less than a predetermined value, the iteration terminates. The computational flow chart of the EM algorithm for estimating the parameters of a simple partial erasure model is depicted in Fig. 1.

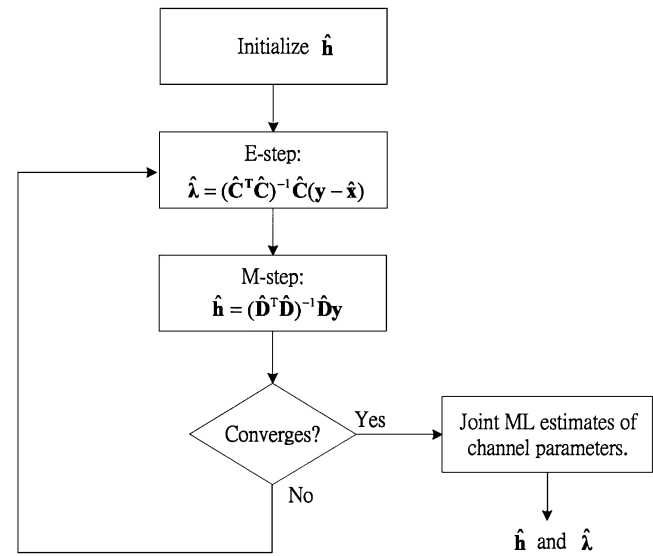


Fig. 1. Computational flowchart of the EM algorithm.

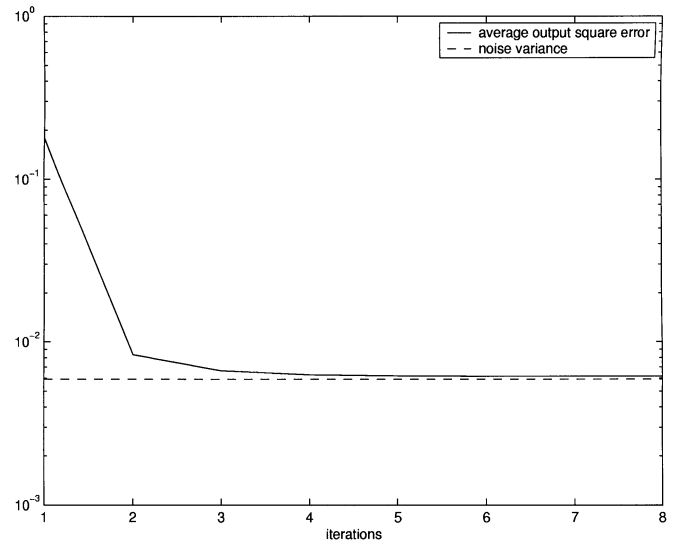


Fig. 2. Averaged output square error versus iterations of a simulation run.

IV. COMPUTER SIMULATIONS

In this section, the EM algorithm is demonstrated to accurately estimate the parameters of a simple partial erasure model by computer simulations. Assume that the model has a Lorentzian isolated transition response with $PW_{50} = 2$ and the reduction parameters $\gamma_1 = 0.7$ and $\gamma_2 = 0.49$. The number of samples for modeling the isolated transition response is set to be 21, i.e., $L = 10$. The data b_k s are obtained by NRZI encoding of a random signal of equal probability on two values $\{1, -1\}$. The number of the sample data is $N = 1000$. The signal power, given the model settings, can be analytically obtained to be 0.5921. The signal-to-noise ratio (SNR), defined as $10 \log (E[(y_k - n_k)^2]/E[n_k^2])$, is set to be 20 dB; hence, the noise variance is equal to 0.005921.

The EM algorithm discussed above with zeros as the initial estimate is simulated. The simulation results of one typical run are shown in Figs. 2–4. Fig. 2 depicts the average square output error, defined as $\varepsilon = \sum_{k=1}^N (y_k - \hat{y}_k)^2 / N$, where \hat{y}_k is the output

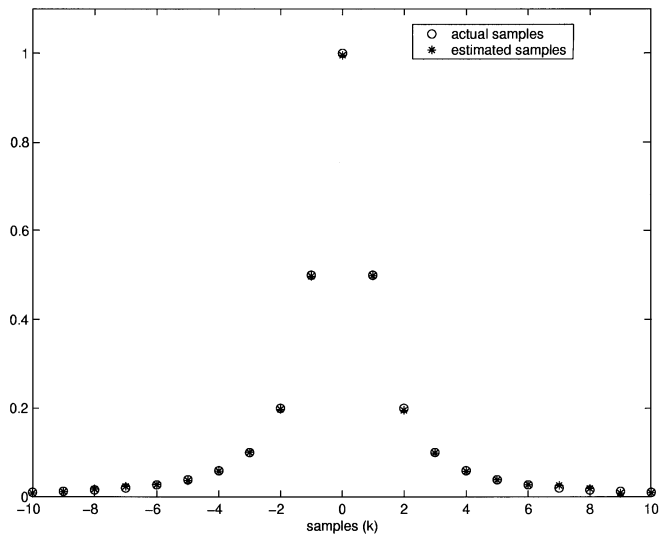


Fig. 3. Samples of the actual isolated transition response and its estimates obtained at the eighth iteration of a simulation run.

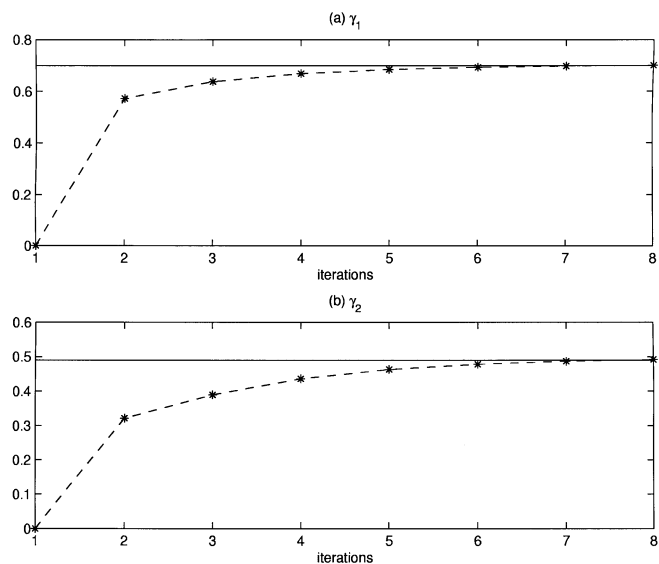


Fig. 4. (a) $\hat{\gamma}_1$ versus iterations of a simulation run (b) $\hat{\gamma}_2$ versus iterations of a simulation run.

of the identified model versus the iteration. Note that the y axis is in log-scale and the noise power is also shown in a dashed line. This result indicates that the algorithm converges rapidly in a few iterations, and the convergent average square error is close to the noise variance. The algorithm terminates at the eighth iteration when the predetermined error measure, defined as the 2-norm of the difference between the present and the previous estimates of $\hat{\mathbf{h}}$, is set to 10^{-3} . The samples of the actual isolated transition response \mathbf{h} and its estimates $\hat{\mathbf{h}}$, obtained at the eighth iteration, are shown in \circ and in $*$, respectively, in Fig. 3. The estimated reduction parameters $\hat{\gamma}_1$ and $\hat{\gamma}_2$ at each iteration are also depicted in $*$ in Fig. 4(a) and (b). From these simulations, the estimated samples of the isolated transition response and the reduction parameters are very close to the true ones, which verifies the proposed method.

TABLE I
ESTIMATED ERROR VARIANCE OF EACH REDUCTION PARAMETER AND ITS CRAMÉR-RAO BOUND UNDER VARIOUS SNRS

SNR	estimated variance(γ_1)	CR-bound(γ_1)	estimated variance(γ_2)	CR-bound(γ_2)
10 dB	2.20×10^{-4}	1.55×10^{-4}	6.92×10^{-4}	4.90×10^{-4}
20 dB	2.21×10^{-5}	1.55×10^{-5}	7.17×10^{-5}	4.90×10^{-5}
30 dB	2.36×10^{-6}	1.55×10^{-6}	8.32×10^{-6}	4.90×10^{-6}

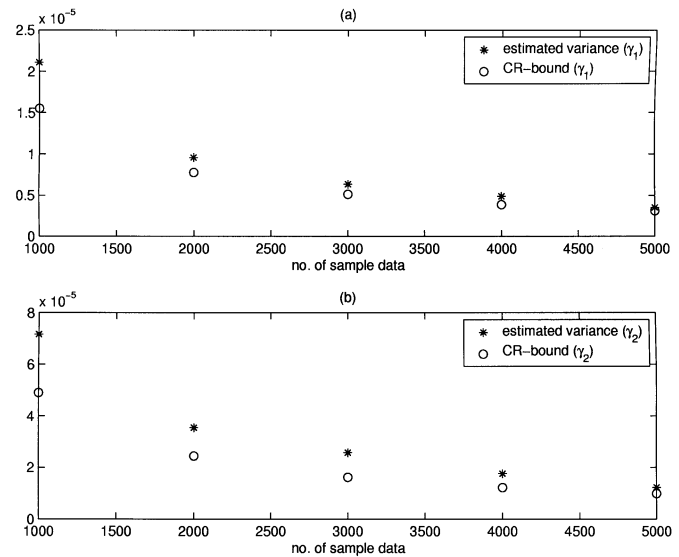


Fig. 5. (a) Estimated error variance of γ_1 versus the number of data samples. (b) Estimated error variance of γ_2 versus the number of data samples.

We also demonstrate by simulation that the joint ML estimator of the model parameters is asymptotically efficient. For simplicity, we shall only show the error variance of each reduction parameter, obtained by Monte Carlo simulations using the average of 100 independent runs. The simulation results under various SNRs are listed in Table I. As shown in the table, each error variance is in the same order as the Cramér-Rao bound. Also shown in Fig. 5, under SNR of 20 dB increasing the number of data samples decreases the discrepancy between the obtained error variance and the Cramér-Rao bound. When the data sample number equals 5000, the error variances of γ_1 and γ_2 are nearly equal to their corresponding Cramér-Rao bounds, respectively. These results show that the proposed algorithm for estimating the parameters of a simple partial erasure model is asymptotically efficient.

V. CONCLUSION

The EM algorithm has been successfully applied for estimating the parameters of a simple partial erasure model, including the reduction parameters and the isolated transition response. This approach not only avoids solving the nonlinear equations but provides an effective way to identify accurately the model parameters. The resulting algorithm is an iteration of solving two least-squares problems; hence, its realization is simple. Simulation results also show that the convergence is fast and the resulting estimator is asymptotically efficient. This algorithm can estimate the model parameters rapidly and accurately and, thus, is expected to improve the performance of high-density magnetic recording.

REFERENCES

- [1] S. Choi, S. Ong, C. You, D. Hong, and J. Cho, "Performance of neural equalizers on partial erasure model," *IEEE Trans. Magn.*, vol. 33, pp. 2788–2790, Sept. 1997.
- [2] M. P. C. Fossorier, T. Mori, and H. Imai, "Simplified trellis diagram for a simple partial erasure model," *IEEE Trans. Magn.*, vol. 34, pp. 320–322, Jan. 1998.
- [3] S. Choi and D. Hong, "Performance of RBF equalizer in data storage channels," in *Int. Joint Conf. Neural Networks*, vol. 2, 1999, pp. 1077–1080.
- [4] T. Yamauchi and J. M. Cioffi, "A nonlinear model in the thin film disk recording systems," in *Int. Conf. Magn.*, vol. EC-10, Apr. 1993.
- [5] W. E. Ryan and N. H. Yeh, "Viterbi detector for Pr4-equalized magnetic recording channels with transition-shift and partial erasure nonlinearities," *IEEE Trans. Magn.*, vol. 32, pp. 3950–3952, Sept. 1996.
- [6] R. Hermann, "Volterra modeling of digital magnetic saturation recording channels," *IEEE Trans. Magn.*, vol. 26, pp. 2125–2127, Sept. 1990.
- [7] I. Lee, T. Yamauchi, and J. M. Cioffi, "Performance comparison of receivers in a simple partial erasure model," *IEEE Trans. Magn.*, vol. 30, pp. 1465–1469, July 1994.
- [8] D. Palmer, P. A. Ziperovich, R. Wood, and T. Howell, "Identification of nonlinear write effect using pseudo-random sequences," *IEEE Trans. Magn.*, vol. 23, pp. 2377–2379, Sept. 1987.
- [9] X. Che and P. A. Ziperovich, "A time-correction method of calculating nonlinearities utilizing pseudo-random sequences," *IEEE Trans. Magn.*, vol. 30, pp. 4239–4241, Nov. 1994.
- [10] Y. S. Tang and C. Tsang, "A technique for measuring nonlinear bit shift," *IEEE Trans. Magn.*, vol. 27, pp. 5316–5318, Nov. 1991.
- [11] Y. S. Cho and N. J. Lee, "A technique for nonlinear distortion in high-density magnetic recording channels," *IEEE Trans. Magn.*, vol. 34, pp. 40–44, Jan. 1998.
- [12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, ser. B, vol. 39, no. 1, pp. 1–38, 1997.
- [13] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, pp. 47–60, Nov. 1996.
- [14] R. Perry, W. A. Berger, and K. Buckley, "EM algorithms for sequence estimation over random ISI channels," in *Asilomar Conf. Signals, Systems and Computers*, vol. 1, 1999, pp. 295–299.
- [15] C. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
- [16] S. X. Wang and A. M. Taratorin, *Introduction to Magnetic Information Storage Technology*. New York: Academic, 1999, pp. 291–293.

Tsai-Sheng Kao was born in Taipei, Taiwan, R.O.C., in 1975. He received the B.S. and M.S. degrees in electrical and control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1997 and 1999, respectively. He is currently a doctoral student in the Department of Electrical and Control Engineering at National Chiao Tung University.

His research interests include magnetic storage systems, parameter estimation, and digital signal processing.

Mu-Huo Cheng received the B.S. degree in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, R.O.C., in 1980, the M.S. degree in electronic engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1987, and the Ph.D. degree from the Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, PA, in 1995.

Since 1995, he has been an Associate Professor in the Department of Electrical and Control Engineering at National Chiao Tung University. His current research interests include adaptive signal processing, detection, and estimation.