

Blind Iterative Channel Identification and Equalization

R. R. Lopes and J. R. Barry

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250

Abstract — We propose an iterative solution to the problem of blindly and jointly identifying the channel response and transmitted symbols in a digital communications system. The proposed algorithm iterates between a *symbol estimator*, which uses tentative channel estimates to provide soft symbol estimates, and a *channel estimator*, which uses the symbol estimates to improve the channel estimates. The proposed algorithm shares some similarities with the Expectation-Maximization (EM) algorithm but with lower complexity and better convergence properties. Specifically, the complexity of the proposed scheme is linear in the memory of the equalizer, and it avoids most of the local maxima that trap the EM algorithm.

I. INTRODUCTION

An important class of algorithms for blind channel identification is based on the iterative strategy depicted in Fig. 1 [1–9]. In these algorithms, an initial channel estimate is used by a symbol estimator to provide tentative soft estimates of the transmitted symbol sequence. These estimates are used by a channel estimator to improve the channel estimates. The improved channel estimates are then used by the symbol estimator to improve the symbol estimates, and so on.

Iterative techniques for channel identification have three important advantages. First, the channel estimates are capable of approaching the maximum-likelihood (ML) estimates. Second, blind identification of the channel allows the receiver to use decision-feedback equalization, ML sequence detection via the Viterbi algorithm, or maximum a posteriori (minimum symbol-error rate) detection via the BCJR algorithm [10]. Such detectors offer excellent performance even when the channel has a spectral null, where equalizers based on linear filters (using the constant-modulus algorithm, for example)

are known to fail. Third, iterative schemes easily exploit, in a nearly optimal way, any *a priori* information the receiver may have about the transmitted symbols. This *a priori* information may arise because of pilot symbols, a training sequence, precoding, or error-control coding. Thus, iterative channel estimators are well-matched to applications involving turbo equalization [4][11] or semi-blind channel estimation [12]. By exploiting the existence of coding, a blind iterative detector is able to benefit from the immense coding gains that typify today's powerful codes (based on turbo coding or low-density parity-check codes, for example). These codes allow reliable communication with an SNR that is lower than ever before, which only exacerbates the identification problem for traditional algorithms that ignore the existence of coding.

Most prior work in blind iterative channel identification can be tied to the EM algorithm of [13]. The first such detector [1] used hard decisions produced by a Viterbi sequence detector. The first use of EM with soft symbol estimates (as produced by BCJR) was proposed in [2]. It is a direct application of EM (as are [3] and [4]), and hence is guaranteed to converge to a local maximum of the joint channel-sequence likelihood function. An adaptive version [5] of EM was applied to the identification problem in [6]. The algorithms proposed in [7–9] are modifications of EM.

The algorithm we propose differs from prior work in its lower complexity and improved convergence. Specifically, we propose a symbol estimator based on a DFE, whose complexity increases linearly with the equalizer memory, as opposed to exponentially with channel memory. We also propose an extended-window channel estimator which greatly reduces the probability of misconvergence to a nonglobal local maximum of the joint likelihood function.

II. PROBLEM STATEMENT

Consider the transmission of K independent and identically distributed zero-mean symbols a_k with unit energy $E[a_k^2] = 1$, belonging to some alphabet \mathcal{A} , across a dispersive channel with memory μ and additive white Gaussian noise. The received signal at time k can be written as

$$r_k = \mathbf{h} \mathbf{a}_k + n_k, \quad (1)$$

where $\mathbf{h} = (h_0, h_1, \dots, h_\mu)$ represents the channel impulse response, $\mathbf{a}_k = (a_k, a_{k-1}, \dots, a_{k-\mu})^T$, and $n_k \sim \mathcal{N}(0, \sigma^2)$ is the

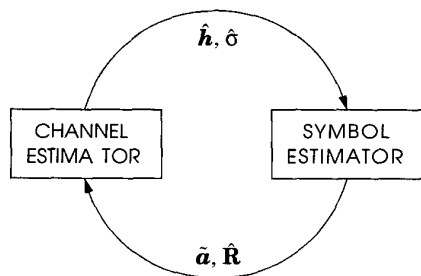


Figure 1. Blind iterative channel identification.

noise. Let $\mathbf{a} = (\alpha_0, \alpha_1, \dots, \alpha_{K-1})$ and $\mathbf{r} = (r_0, r_1, \dots, r_{N-1})$ denote the input and output sequences, respectively, where $N = K + \mu$. The blind channel identification problem is to estimate \mathbf{h} and σ relying solely on the received signal \mathbf{r} and knowledge of the probability distribution function for \mathbf{a} .

III. EM AND BLIND CHANNEL IDENTIFICATION

The EM algorithm [13] provides an iterative solution to the blind identification problem that fits the paradigm of Fig. 1. It generates a sequence of channel estimates with non-decreasing likelihood, and under general conditions converges to a local (not necessarily global) maximum of the likelihood function. Application to blind channel identification was first proposed in [2]. The *channel estimator* (see Fig. 1) for the $(i+1)$ -st iteration of the EM algorithm is defined by

$$\hat{\mathbf{h}}^{(i+1)} = \mathbf{R}_i^{-1} \mathbf{p}_i, \quad (2)$$

$$\hat{\sigma}_{(i+1)}^2 = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E} \left[|r_k - \hat{\mathbf{h}}^{(i+1)} \mathbf{a}_k|^2 \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)} \right], \quad (3)$$

where

$$\mathbf{R}_i = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E} [\mathbf{a}_k \mathbf{a}_k^T \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)}], \quad (4)$$

$$\mathbf{p}_i = \frac{1}{N} \sum_{k=0}^{N-1} r_k \mathbb{E} [\mathbf{a}_k \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)}]. \quad (5)$$

The *symbol estimator* provides the values of $\tilde{\mathbf{a}}_k^{(i)} = \mathbb{E} [\mathbf{a}_k \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)}]$ and $\mathbb{E} [\mathbf{a}_k \mathbf{a}_k^T \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)}]$ that are required by (3) through (5). The symbol estimator of the EM algorithm uses the BCJR algorithm [10], which has complexity exponential in the channel memory, and is based on the tentative channel estimates fed from the channel estimator.

Note that \mathbf{R}_i and \mathbf{p}_i can be viewed as estimates of the *a posteriori* autocorrelation matrix of the transmitted sequence and the cross-correlation vector between the transmitted and receive sequences, respectively. Thus, (2) is very similar to the least-squares solution [16], in which these quantities are replaced by the actual sample autocorrelation matrix and cross-correlation vector.

IV. THE CHANNEL AND SYMBOL ESTIMATORS

We first propose a low-complexity channel estimator that avoids the matrix inversion of (2). From the channel model it is clear that $h_n = \mathbb{E}[r_k \mathbf{a}_{k-n}]$. But

$$\mathbb{E}[r_k \mathbf{a}_{k-n}] = \mathbb{E} \left[r_k \mathbb{E} [\mathbf{a}_{k-n} \mid r_k] \right] = \mathbb{E} \left[r_k \mathbb{E} [\mathbf{a}_{k-n} \mid \mathbf{r}] \right]. \quad (6)$$

Note that the channel estimator has no access to $\mathbb{E} [\mathbf{a}_k \mid \mathbf{r}]$, which requires exact channel knowledge. However, based on the iterative paradigm of Fig. 1, at the i -th iteration it does

have access to $\tilde{\mathbf{a}}_k^{(i)} = \mathbb{E} [\mathbf{a}_k \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)}]$. Replacing this value in (6), and also replacing time average for ensemble average, leads to the following channel estimator:

$$\hat{h}_n^{(i+1)} = \frac{1}{N} \sum_{k=0}^{N-1} r_k \tilde{\mathbf{a}}_{k-n}^{(i)} \quad (7)$$

for $n = 0, 1, \dots, \mu$. The idea of estimating the channel by correlating the received signal with the transmitted sequence has been explored in [17], but the algorithm proposed in that work relies on a training sequence.

Combining the estimates (7) into a single vector, we find that $\hat{\mathbf{h}}^{(i+1)} = (\hat{h}_0^{(i+1)}, \dots, \hat{h}_\mu^{(i+1)}) = \mathbf{p}_i$. Thus, we may view (7) as a simplification of the EM estimate $\mathbf{R}_i^{-1} \mathbf{p}_i$; it avoids matrix inversion by replacing \mathbf{R}_i by \mathbf{I} . This is reasonable given that \mathbf{R}_i is an *a posteriori* estimate of the autocorrelation matrix of the transmitted vector, which is known to be the identity.

As for estimating the noise variance, let $\hat{\sigma}_k^{(i)}$ be an estimate of the transmitted sequence, chosen as the element of \mathcal{A} closest to $\tilde{\mathbf{a}}_k^{(i)}$. We propose to compute $\hat{\sigma}$ using:

$$\hat{\sigma}_{(i+1)}^2 = \frac{1}{N} \sum_{k=0}^{N-1} |r_k - \hat{\mathbf{h}}^{(i)} \hat{\mathbf{a}}_k^{(i)}|^2. \quad (8)$$

Since this channel estimator no longer needs \mathbf{R}_i , the symbol estimator need only produce $\tilde{\mathbf{a}}^{(i)} = \mathbb{E} [\mathbf{a} \mid \mathbf{r}; \hat{\mathbf{h}}^{(i)}, \hat{\sigma}^{(i)}]$. Although the BCJR algorithm can compute $\tilde{\mathbf{a}}^{(i)}$ exactly, its high complexity (exponential in μ) is often prohibitive.

We now propose a low-complexity symbol estimator based on a DFE, restricting our attention to the binary alphabet $\mathcal{A} = \{-1, 1\}$. Consider filtering the channel output with a minimum mean-square error DFE (MMSE-DFE), resulting in an equalizer output z_k that is approximately free of intersymbol interference (ISI). In this case, we could write $z_k \approx \mathbf{A} \mathbf{a}_k + v_k$, where \mathbf{A} is the gain of the equivalent ISI-free channel, and the equivalent "noise" term v_k is approximately zero mean and Gaussian with some variance σ_v^2 . If this model were valid, the output of the symbol estimator could be written as $\mathbb{E} [\mathbf{a}_k \mid \mathbf{r}] = \tanh(\mathbf{A} z_k / \sigma_v^2)$. Thus, we propose to use $\hat{\mathbf{h}}^{(i)}$ and $\hat{\sigma}^{(i)}$ to compute the MMSE-DFE coefficients, and to filter the received sequence with this equalizer. The symbol estimator then outputs $\tilde{\mathbf{a}}_k = \tanh(\mathbf{A} z_k / \sigma_v^2)$. The complexity of this symbol estimator is independent of channel memory, and increases linearly with the length of the equalizer.

As for computing \mathbf{A} and σ_v^2 , we propose that the symbol estimator be equipped with the following internal iteration:

$$A_{i+1} = \frac{1}{N} \sum_{k=0}^{N-1} z_k \tanh(\mathbf{A}_i z_k / \sigma_{v,i}^2) \quad (9)$$

$$\sigma_{v,i+1}^2 = \frac{1}{N} \sum_{k=0}^{N-1} (z_k - A_{i+1} \text{sign}(z_k))^2, \quad (10)$$

which are repeated until a stop criterion is met. These iterations result from the application of the simplified channel estimator (7), (8) to a 1-tap channel. Usually, only a few repetitions are needed, and A_i is initialized to $A_0 = \sigma_{v,0}^2$.

V. THE EXTENDED-WINDOW ALGORITHM

Misconvergence is a common characteristic of iterative channel identification algorithms. In fact, it is very easy to find an example in which both EM and the proposed algorithm, as defined so far, fail to identify the channel.

To illustrate the problem of misconvergence, consider $\mathbf{h} = [1 \ 2 \ 3 \ 4 \ 5]$. The proposed algorithm converges to $\hat{\mathbf{h}}^{(20)} = [2.179 \ 3.073 \ 4.108 \ 5.092 \ 0.120]$ after 20 iterations, with $K = 1000$ bits, $\text{SNR} = \|\mathbf{h}\|^2/\sigma^2 = 20$ dB, with initialization $\hat{\mathbf{h}}^{(0)} = [1 \ 0 \ 0 \ 0 \ 0]$ and $\hat{\sigma}_{(0)} = 1$, and with $N_f = 15$ forward and $N_b = 5$ feedback DFE coefficients. The estimate is seen to roughly approximate a shifted or delayed version of the actual channel. This misconvergence stems from the delay mismatch between \mathbf{h} and the initialization $\hat{\mathbf{h}}^{(0)}$. The iterative scheme cannot compensate for this delay. In fact, after convergence, $\text{sign}(\tilde{a}_k)$ is a good estimate for a_{k-1} , and h_0 can be well estimated by correlating r_k with \tilde{a}_{k+1} . But because the delay n in (7) is limited to the narrow window $\{0, \dots, \mu\}$, (7) never computes this correlation. This observation leads us to propose the extended-window algorithm, in which (7) is computed for a broader range of n .

To determine how much the correlation window must be extended, consider two extreme cases. First, suppose $\mathbf{h} \approx [1 \ 0 \ 0 \ 0 \dots \ 0]$ and $\hat{\mathbf{h}}^{(i)} \approx [0 \dots \ 0 \ 0 \ 0 \ 1]$. In this case, the symbol estimator output \tilde{a}_k is such that $\text{sign}(\tilde{a}_k) \approx a_{k-\mu}$. Thus, to estimate h_0 we must compute (7) for $n = -\mu$. Likewise, if $\mathbf{h} \approx [0 \dots \ 0 \ 0 \ 0 \ 1]$ and $\hat{\mathbf{h}}^{(i)} \approx [1 \ 0 \ 0 \ 0 \dots \ 0]$, the symbol estimator output \tilde{a}_k is such that $\text{sign}(\tilde{a}_k) \approx a_{k+\mu}$, and so to estimate h_μ we must compute (7) for $n = 2\mu$. These observations suggest the extended-window algorithm, which computes

$$\mathbf{g}_n = \frac{1}{N} \sum_{k=0}^{N-1} r_k \tilde{a}_{k-n}^{(i)} \quad (11)$$

for $n \in \{-\mu, \dots, 2\mu\}$. By doing this, we ensure that $\mathbf{g} = [g_{-\mu}, \dots, g_{2\mu}]$ has $\mu + 1$ entries that estimate the desired correlations $E[r_k a_{k-n}]$, for $n \in \{0, \dots, \mu\}$. Its remaining terms are an estimate of $E[r_k a_{k-n}]$ for $n \notin \{0, \dots, \mu\}$, and hence should be close to zero. We may thus define the channel estimate by $\hat{\mathbf{h}}^{(i+1)} = [g_v, \dots, g_{v+\mu}]$, where the delay parameter $v \in \{-\mu, \dots, \mu\}$ is chosen so that $\hat{\mathbf{h}}^{(i+1)}$ represents the $\mu + 1$ consecutive coefficients of \mathbf{g} with highest energy.

The noise variance may again be estimated using (8), but accounting for the delay v by using $\hat{a}_k = \text{sign}(\tilde{a}_{k-v}^{(i)})$.

VI. THE IMPACT OF $\hat{\sigma}$

It is interesting to note that while substituting the actual values of \mathbf{h} or \mathbf{a} for their estimates will always improve the performance of the iterative algorithm, the same is not true for σ . Indeed, substituting σ for $\hat{\sigma}$ will often result in performance degradation.

Intuitively, one can think of $\hat{\sigma}$ as playing two roles: in addition to measuring σ , it also acts as a *measure of reliability* in the channel estimate $\hat{\mathbf{h}}$. Consider a decomposition of the observation vector:

$$\mathbf{r} = \mathbf{a} * \hat{\mathbf{h}} + \mathbf{a} * (\mathbf{h} - \hat{\mathbf{h}}) + \mathbf{n}, \quad (12)$$

where $*$ denotes convolution. The second term represents the contribution to \mathbf{r} from the estimation error. By using $\hat{\mathbf{h}}$ to model the channel in the BCJR algorithm, we are in effect lumping the estimation error with the noise. Combining the two results in an effective noise sequence with variance larger than σ^2 . It is thus appropriate that $\hat{\sigma}$ should exceed σ whenever $\hat{\mathbf{h}}$ differs from \mathbf{h} . Alternatively, it stands to reason that an unreliable channel estimate should translate to an unreliable (*i.e.*, with small magnitude) symbol estimate, regardless of how well $\mathbf{a} * \hat{\mathbf{h}}$ matches \mathbf{r} . A large value of $\hat{\sigma}$ in BCJR ensures this.

Fortunately, (8) measures the energy of both the second and the third term in (12). If $\hat{\mathbf{h}}$ is a poor channel estimate, $\tilde{\mathbf{a}}$ will also be a poor estimate for \mathbf{a} , and convolving $\tilde{\mathbf{a}}$ and $\hat{\mathbf{h}}$ will produce a poor match for \mathbf{r} . Thus, (8) will produce a large estimated noise variance.

An interesting and natural question is if the $\hat{\sigma}$ produced by (8) is large enough. We will show in the next section that increasing $\hat{\sigma}$ beyond (8) helps avoid misconvergence.

VII. SIMULATION RESULTS

As a first test of the extended-window algorithm, we have simulated the transmission of $K = 1000$ bits over the channel $\mathbf{h} = [-0.2287, 0.3964, 0.7623, 0.3964, -0.2287]$ from [7], with $\text{SNR} = 18$ dB, $N_f = 15$ forward and $N_b = 5$ feedback DFE coefficients, and 3 inner iterations. To stress the fact that the proposed algorithm is not sensitive to initial conditions, we initialized $\hat{\mathbf{h}}$ randomly using $\hat{\mathbf{h}}^{(0)} = \mathbf{u} \hat{\sigma}_{(0)} / \|\mathbf{u}\|$, where $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\hat{\sigma}_{(0)}^2 = \sum_{k=0}^{N-1} r_k^2 / 2N$. (This implies an initial estimated SNR of 0 dB.). The curves shown in Fig. 2 are the average of 100 independent runs of this experiment. Only the convergence of \hat{h}_0 , \hat{h}_2 and \hat{h}_3 is shown. (The behavior of \hat{h}_1 and \hat{h}_4 is similar to that of \hat{h}_3 and \hat{h}_0 , respectively, but we show only those with worse convergence.) The shaded regions around the channel estimates correspond to plus and minus one standard deviation. For comparison, we show the average behavior of the EM algorithm in Fig. 3. Unlike the good

performance of the extended window algorithm, the EM algorithm even fails to converge in the mean to the correct estimates. This happens because the EM algorithm can get trapped in local maxima of the likelihood function [13], while the extended-window does not.

To further support the claim that the proposed algorithm avoids most of the local maxima of the likelihood function that trap the EM algorithm, we ran both these algorithms on 1,000 random channels of memory $\mu = 4$, generated as $\mathbf{h} = \mathbf{u} / \|\mathbf{u}\|$, where $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The estimates were initialized to $\hat{\sigma}_{(0)}^2 = \sum_{k=0}^{N-1} r_k^2 / 2N$ and $\hat{\mathbf{h}}^{(0)} = [\hat{\sigma}_{(0)}, 0, \dots, 0]$. We used SNR = 18 dB, $K = 1000$ bits, $N_f = 40$, $N_b = 10$, and 3 inner iterations. In Fig. 4 we show the rate of success of the algorithms versus iteration. The algorithms were deemed successful if they produced fewer than 3 bit errors in a block. It is again clear that the extended-window algorithm has a much better

performance than the EM algorithm. This can also be seen in Fig. 5, where we show histograms of the norm of the estimation errors (the difference between the channel vector and its estimate) for the extended-window and the EM algorithms, computed after 70 iterations. We see that while the proposed algorithm produces a negligible number of errors with norm above 0.2, about 45% of the errors produced by the EM algorithm have norm above 0.2.

Some interesting observations can be made by considering the unsuccessful runs of the extended-window algorithm. Consider, for instance, $\mathbf{h} = [0.4823, 0.0194, -0.5889, 0.6285, -0.1586]$, which was not correctly estimated in the previous experiment. First note that this channel does not present severe ISI. In fact, none of the channels the algorithm failed to identify introduce severe ISI. Instead, their common characteristic is the presence of 3 or more coefficients of

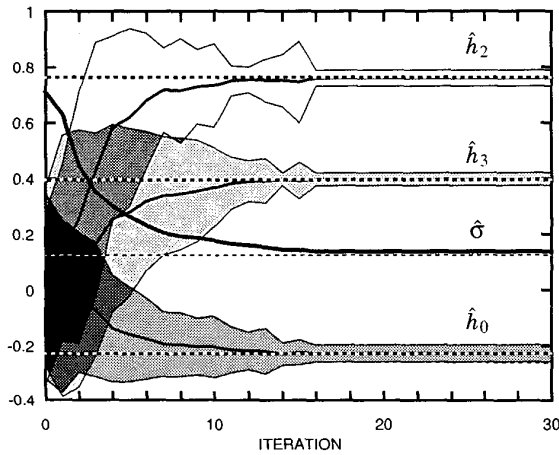


Figure 2. Estimates of $\mathbf{h} = [-0.2287, 0.3964, 0.7623, 0.3964, -0.2287]$, produced by the extended-window algorithm.

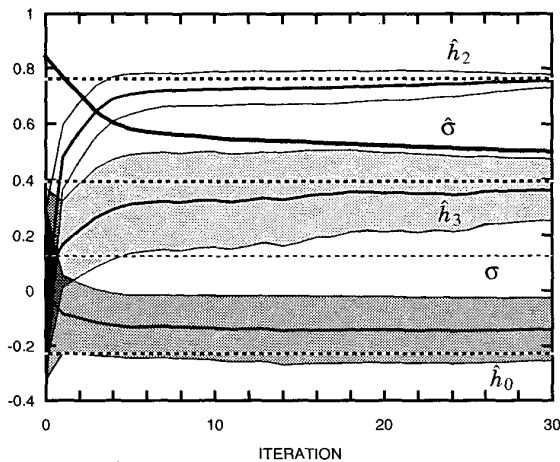


Figure 3. EM estimates of same channel as Fig. 2.

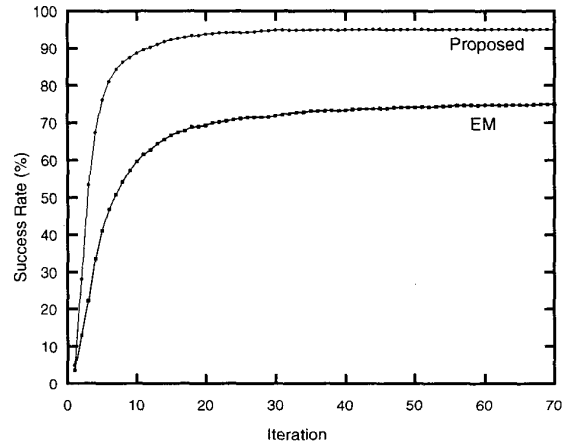


Figure 4. Success rate for extended window and EM algorithms for random channels over an ensemble of 1000.

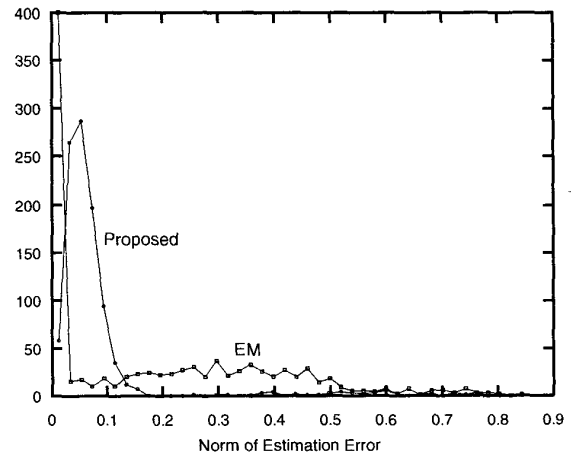


Figure 5. Histograms of estimation errors for the extended-window and the EM algorithms.

approximately the same magnitude. The convergence behavior is also a strong function of the transmitted block length K . For this particular channel, with SNR = 18 dB, $K = 1887$ bits, $N_f = 40$, $N_b = 10$, and 3 inner iterations, the proposed algorithm fails, yielding BER = 30% after convergence. If K is changed to $K = 1888$, then the algorithm is successful, yielding no errors after convergence. Thus, for this particular channel and others we have tested, there is a cutoff block length K above which the algorithm is successful, and below which the algorithm fails.

The impact of $\hat{\sigma}$ was also analyzed. For $\mathbf{h} = [0.4823, 0.0194, -0.5889, 0.6285, -0.1586]$, $K = 1000$, SNR = 13 dB, $N_f = 40$, $N_b = 10$, 3 inner iterations, and with $\hat{\sigma}$ computed according to (8), the proposed algorithm fails, yielding BER = 30% after convergence. However, if we pass $\alpha\hat{\sigma}$ to the symbol estimator instead of $\hat{\sigma}$, then a suitable choice for α can lead to successful convergence. In Fig. 6, we plot the BER after 60 iterations versus α . We see that increasing α can improve the performance of the algorithm, but increasing it too much will again cause the algorithm to fail. In spite of this encouraging observation, we were not able to find a strategy that significantly impacts the success rate over a broad range of channels. For instance, keeping $\alpha = 3$ and rerunning the experiment with the 1000 random channels, we obtained a success rate of 95.6%, as opposed to 95.3% obtained with the regular algorithm ($\alpha = 1$). Thus, finding the optimal value for α for a broad class of channels remains an open problem.

VIII. CONCLUSIONS

We presented an iterative channel identification technique with complexity linear in the number of channel coefficients. We have shown that this technique can be seen as a reduced-complexity version of the EM algorithm. A key feature of the proposed algorithm is its extended window, which greatly improves the convergence behavior of the algorithm, avoiding most of the local maxima of the likelihood function that trap the EM algorithm. A small percentage of channels still cause difficulty, but a smarter initialization strategy (as opposed to the identity initialization proposed here) will likely help. Nevertheless, the problem of devising an iterative strategy that is guaranteed to always avoid misconvergence, regardless of initialization, remains open.

IX. REFERENCES

[1] M. Feder and J. A. Catipovic, "Algorithms for Joint Channel Estimation and Data Recovery — Application to Equalization in Underwater Communications," *IEEE J. Oceanic Eng.*, vol. 16, no. 1, pp. 42–5, Jan. 1991.
 [2] G. K. Kaleh and R. Vallet, "Joint Parameter Estimation and Symbol Detection for Linear or Nonlinear Unknown Channels," *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2406–2413, July 1994.

[3] C. Anton-Haro, J. A. R. Fonollosa and J. R. Fonollosa, "Blind Channel Estimation and Data Detection Using Hidden Markov Models," *IEEE Trans. Sig. Proc.*, vol. 45, no. 1, pp. 241–246, Jan 1997.
 [4] J. Garcia-Frias and J. D. Villaseñor, "Blind Turbo Decoding and Equalization," *IEEE Vehic. Techn. Conference*, vol. 3, pp. 1881–1885, 1999.
 [5] V. Krishnamurthy, J. B. Moore, "On-Line Estimation of Hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure," *EEE Trans. Sig. Proc.*, vol. 41, no. 8, pp. 2557–2573, Aug 1993.
 [6] L. B. White, S. Perreau, P. Duhamel, "Reduced Complexity Blind Equalization for FIR Channel Input Markov Models", *IEEE International Conference on Communications*, vol. 2, pp.993–997, 1995.
 [7] M. Shao and C. L. Nikias, "An ML/MMSE Estimation Approach to blind Equalization", *ICASSP*, vol. 4, pp. 569–572, 1994.
 [8] H. A. Cirpan and M. K. Tsatsanis, "Stochastic Maximum Likelihood Methods for Semi-Blind Channel Equalization," *Signal Processing Letters*, vol. 5, no. 1, pp. 1629–1632, Jan 1998.
 [9] B. P. Paris, "Self-Adaptive Maximum-Likelihood Sequence Estimation," *IEEE Global Telecommunications Conf.*, vol. 4, pp. 92–96, 1993.
 [10] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. on Info. Theory*, pp. 284–287, March 1974.
 [11] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, A. Glavieux, "Iterative Correction of Intersymbol Interference: Turbo Equalization," *European Trans. Telecom.*, vol. 6, no.5, pp.507–511, Sept.–Oct. 1995.
 [12] J. Ayadi, E. de Carvalho and D. T. M. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," *IEEE ICASSP*, vol. 6, pp. 3185–3188, 1998.
 [13] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm", *Journal of the Royal Statistics Society*, vol. 39, no. 1, pp 1–38, 1997.
 [14] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes(1)," *IEEE Intern. Conf. on Communications*, pp. 1064–70, Geneva, May 1993.
 [15] M. Varanasi and B. Aazhang, "Multistage Detection in Asynchronous Code-Division Multiple-Access Communications," *IEEE Transactions on Communications*, vol. 38, pp. 509–519, April 1990.
 [16] Poor, H. V., *An Introduction to Signal Detection and Estimation*, Second Edition, Springer-Verlag, 1994.
 [17] C. A. Montemayor, P. K. Flikkema, "Near-Optimum Iterative Estimation of Dispersive Multipath Channels," *IEEE Vehicular Technology Conference*, vol. 3, pp. 2246–2250, 1998.

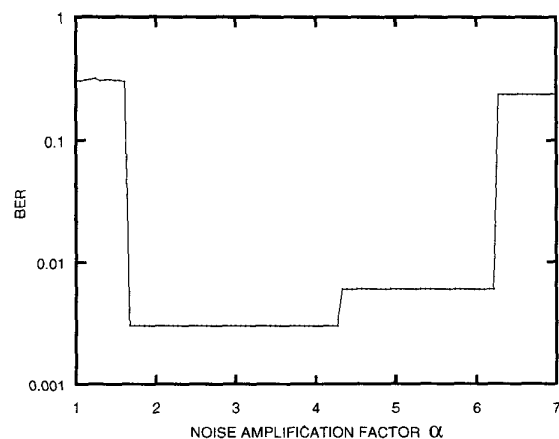


Figure 6. Fixing $\hat{\sigma}$ artificially high improves performance.