

Semiblind BLUE Channel Estimation With Applications to Digital Television

Christopher Pladdy, Serdar Özen, Sreenivasa M. Nerayanuru, Michael Zoltowski, and Mark Fimoff, *Member, IEEE*

Abstract—A semiblind iterative algorithm to construct the best linear unbiased estimate (BLUE) of the channel impulse response (CIR) vector h for communication systems that utilize a periodically transmitted training sequence within a continuous stream of information symbols is devised. The BLUE CIR estimate for the general linear model $y = Ah + w$, where w is the correlated noise, is given by the Gauss–Markoff theorem. The covariance matrix of the correlated noise, which is denoted by $C(h)$, is a function of the channel that is to be identified. Consequently, an iteration is used to give successive approximations $h^{(k)}$, $k = 0, 1, 2, \dots$ to h_{BLUE} , where $h^{(0)}$ is an initial approximation given by the correlation processing, which exists at the receiver for the purpose of frame synchronization. A function $F(h)$ for which h_{BLUE} is a fixed point is defined. Conditions under which h_{BLUE} is the unique fixed point and for which the iteration proposed in the algorithm converges to the unique fixed point h_{BLUE} are given. The proofs of these results follow broadly along the lines of Banach fixed-point theorems.

Index Terms—Banach fixed-point theorem, equalization, fixed point iterations, Gauss–Markoff theorem, semiblind channel estimation.

I. INTRODUCTION

RELIABLE communication often requires accurate estimation of the channel impulse response (CIR) to facilitate channel equalization. Semiblind algorithms exploit information used by blind methods (for example, the statistics of unknown data symbols) as well as information from known training symbols. We devise a semiblind iterative algorithm to construct the best linear unbiased estimate (BLUE) of the channel in the case where we have the general linear model for the received data, i.e., where the noise is not white, since we consider the effects of noise due to unknown data, where such noise is correlated. The algorithm that we propose is

designed to work within the constraints imposed by the ATSC digital TV 8-VSB [1] modulation system. This standard uses a training sequence of length 728, which is repeated every frame of 260 416 data symbols. Delay spreads observed from data capture in urban multipath settings give channel lengths of 400–500 symbols. Such channels are too long to form an overdetermined system for an accurate least squares (LS) estimate of the channel without considering correlated noise due to unknown data, hence, the use of the statistics of the unknown data that surround the training sequence, which makes our algorithm a semiblind algorithm. In a mobile vehicular setting (e.g., digital TV reception in a minivan), the physical channel will have changed substantially when the next training burst arrives. For this reason, we require a channel estimation method that yields a reliable channel estimate based on one burst of training data. To achieve this, our method considers an overdetermined system that must account for correlated noise due to transmission data. The BLUE CIR estimate, which we propose, allows for such accurate channel estimation, even in a mobile setting. We give conditions under which the iterative method is guaranteed to converge by considering the method as a function iteration converging to the unique fixed point of the function. See the Banach fixed-point theorem function iteration and contraction mappings in [2]–[4] for contraction mappings in signal processing and image reconstruction.

For communications systems utilizing a periodically transmitted training sequence, LS-based channel estimation and correlation-based channel estimation algorithms have been the two most commonly used approaches [5]. Both methods use a stored copy of the training sequence at the receiver. Most channel estimation schemes do not account for the baseline noise term, which occurs due to the correlation of the stored copy of the training sequence with the unknown symbols adjacent to the transmitted training sequence [5], [6]. We take into account the statistics of these unknown data symbols, which gives rise to the fact that our algorithm is semiblind, using not only known training symbols but also the statistics of unknown adjacent information symbols [7]. Several recent papers consider different aspects of semiblind channel estimation. Notably, de Carvalho and Slock [8] consider identifiability conditions for blind and semiblind finite-impulse response (FIR) multi-channel estimation. Medles *et al.* [9] consider linear-prediction-based semiblind estimation of multiple-input–multiple-output (MIMO) FIR channels. Lasaulce [10] considers a semiblind method for code-division multiple-access (CDMA) systems, while Buchoux [11] considers the performance of semiblind subspace-based methods. Rontogiannis *et al.* [12] consider the semiblind problem via a separable LS approach.

Manuscript received April 21, 2004; revised May 31, 2005 and July 19, 2005. The review of this paper was coordinated by Prof. Z. Xu.

C. Pladdy was with Zenith Electronics R&D Center, Lincolnshire, IL 60069-3630 USA. He is now with MITRE Corporation, Leavenworth, KS 66048 USA (e-mail: cpladdy@mitre.org).

S. Özen was with Zenith Electronics R&D Center, Lincolnshire, IL 60069-3630 USA. He is now with the Department of Electrical Engineering, Izmir Institute of Technology-Urla, 35430 Izmir, Turkey (e-mail: serdarozen@iyte.edu.tr).

S. M. Nerayanuru was with Zenith Electronics R&D Center, Lincolnshire, IL 60069-3630 USA. He is now with Motorola Inc., Libertyville, IL 60048 USA.

M. Zoltowski is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: mikedz@ecn.purdue.edu).

M. Fimoff was with Zenith Electronics R&D Center, Lincolnshire, IL 60069-3630 USA.

Digital Object Identifier 10.1109/TVT.2006.878553

Although the simulation examples that we draw upon are concerned with the ATSC digital TV 8-VSB system, the theory and algorithm that we present could be applied, with minor modifications, to any digital communication system with linear modulation that employs a periodically transmitted training sequence. The novel algorithm that we present is targeted at systems that are desired to work with channels having long delay spreads L_d . In particular, we consider the case where $(NT + 1)/2 < L_d < NT$, where N is the number of training symbols, T is the symbol period, and NT is the duration of the available training sequence. For instance, the 8-VSB digital TV system has 728 training symbols, whereas the delay spreads of the terrestrial channels have been observed to be at least 400–500 symbols long.

II. BASEBAND DATA TRANSMISSION MODEL

The baseband symbol rate sampled receiver pulse-matched filter output is given by

$$y[n] \equiv y(t)|_{t=nT} = \sum_k I_k h[n - k] + \nu[n] \quad (1)$$

$$= \sum_k I_k h[n - k] + \sum_k \eta[k] q^*[-n + k] \quad (2)$$

where

$$I_k = \left\{ \begin{array}{ll} a_k, & 0 \leq k \leq N - 1 \\ d_k, & N \leq k \leq N' - 1 \end{array} \right\} \in A \equiv \{a_1, \dots, a_M\} \quad (3)$$

is the M -ary complex-valued training sequence $A \subset C$ [13]–[15]. $\{a_k\}$ denotes the first N known (training) symbols within a frame of length N' , and $\{d_k\}$ denotes the remaining $N' - N$ random data within the frame. $\nu(t) = \eta(t) * q^*(-t)$ denotes the complex (colored) noise process after the receiver (pulse) matched filter, with $\eta(t)$ being a zero-mean white Gaussian noise process with spectral density σ_η^2 per real and imaginary part. $h(t)$ is the complex-valued impulse response of the composite dynamic channel, including the pulse shaping transmit filter $q(t)$, the physical CIR $c(t)$, and the receive filter $q^*(-t)$, and is given by

$$h(t) = p(t) * c(t) = \sum_{k=-K}^L c_k p(t - \tau_k). \quad (4)$$

$p(t) = q(t) * q^*(-t)$ is the convolution of the transmit and receive filters, where $q(t)$ has a finite support of $[-T_q/2, T_q/2]$, and the span of the transmit and receive filters T_q is an integer multiple of symbol period T . $\{c_k\} \subset C$ denotes the complex-valued physical channel gains, and $\{\tau_k\}$ denotes the multipath delays or the time of arrivals (TOA). It is assumed that the time variations of the channel are slow enough that $c(t)$ can be assumed to be a static intersymbol interference (ISI) channel, at least throughout the training period. We also note that for the 8-VSB system, the transmitter pulse shape is the Hermitian symmetric root-raised cosine pulse, which implies that $q(t) = q^*(-t)$. In the sequel, we will denote both the transmit and receive filters by $q[n] \equiv q(t)|_{t=nT}$. In addition, the sampled

matched filter output signal $y[n]$ will be used extensively in vector form; therefore, we introduce the notation $y_{[n_1:n_2]} = [y[n_1], \dots, y[n_2]]^T \in R^{n_2 - n_1 + 1}$. Similarly, $\eta_{[n_1:n_2]} = [\eta[n_1], \dots, \eta[n_2]]^T \in R^{n_2 - n_1 + 1}$, and $\nu_{[n_1:n_2]} = [\nu[n_1], \dots, \nu[n_2]]^T \in R^{n_2 - n_1 + 1}$.

Without loss of generality, the symbol rate sampled complex-valued composite CIR $h[n]$ can be written as a finite dimensional vector

$$h = [h[-N_a], \dots, h[-1], h[0], h[1], \dots, h[N_c]]^T \in R^{N_a + N_c + 1} \quad (5)$$

where N_a and N_c denote the number of anticausal and causal taps, respectively. $(N_a + N_c + 1)$ is the total memory of the channel.

III. CHANNEL ESTIMATION VIA CORRELATION

Channel estimation via correlation is typically performed by correlating the received signal with a copy of the training sequence stored at the receiver. Cox [16] initially considered this problem for outdoor channels, while Devasirvatham [17] considered indoor channels, and Parsons *et al.* [18] reduced the complexity of the algorithm for implementation. The correlation used for initial synchronization is considered in [19]. Other works that consider channel estimation via correlation are [20]–[22].

We denote the channel estimate obtained via correlation

$$h_{\text{corr}} = [h_{\text{corr}}[-N_a], \dots, h_{\text{corr}}[-1], h_{\text{corr}}[0], h_{\text{corr}}[1], \dots, h_{\text{corr}}[N_c]]^T$$

which is given by

$$h_{\text{corr}}[k] = \sum_{n=k}^{N-1+k} y[n] a_{n-k}.$$

Collecting together the $(N_a + N_c + 1)$ of these equations, the correlation channel estimate is given in matrix vector form by

$$h_{\text{corr}} = \hat{A} y_{[-N_a:N_c+N-1]}$$

where

$$\hat{A} = T \left\{ [a_0, 0_{N_a+N_c}^T]^T, [a_0, \dots, a_{N-1}, 0_{N_a+N_c}^T] \right\} \\ = \begin{bmatrix} a_0 & a_1 & \cdots & a_{N-1} & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & a_{N-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_{N-1} \end{bmatrix}$$

is the $(N_a + N_c + 1) \times (N_a + N_c + N)$ Toeplitz convolution matrix with first column $[a_0, 0_{N_a+N_c}^T]^T$ and first row $[a_0, \dots, a_{N-1}, 0_{N_a+N_c}^T]$.

The received signal $y_{[-N_a:N_c+N-1]}$ is given by

$$y_{[-N_a:N_c+N-1]} = Xh + v_{[-N_a:N_c+N-1]}$$

where $v_{[-N_a:N_c+N-1]} = \widehat{Q}\eta_{[-N_a-L_q:N_c+N-1-L_q]}$, where $\eta_{[-N_a-L_q:N_c+N-1-L_q]}$ is additive white Gaussian noise (AWGN), and $\widehat{Q} \in R^{(N_a+N_c+N) \times (N_a+N_c+N+2L_q)}$ is given by

$$\widehat{Q} = \begin{bmatrix} q^T & 0 & \cdots & 0 \\ 0 & q^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q^T \end{bmatrix}$$

and

$$q^T = [q[+L_q], \dots, q[0], \dots, q[-L_q]] \in R^{2L_q+1}$$

where q denotes the symbol rate sampled receiver pulse-matched filter. The matrix X given by

$$X = T \left\{ x_{[0:N_a+N_c+N-1]}, x_{[0:-N_c-N_a]}^T \right\}$$

is the $(N_a + N_c + N) \times (N_a + N_c + 1)$ Toeplitz matrix with first column $x_{[0:N_a+N_c+N-1]}$ and first row $x_{[0:-N_c-N_a]}^T$. The values of x_n are given by

$$x_n = \begin{cases} a_n, & \text{for all } 0 \leq n \leq N-1 \\ d_n, & \text{for all other } n \end{cases}.$$

Hence

$$X = \begin{bmatrix} a_0 & d_{-1} & \cdots & d_{-N_c-N_a} \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & a_1 & \ddots & d_{-1} \\ a_{N-1} & \vdots & \ddots & a_0 \\ d_N & a_{N-1} & \ddots & a_1 \\ \vdots & \ddots & \ddots & \vdots \\ d_{N_a+N_c+N-1} & \cdots & d_N & a_{N-1} \end{bmatrix}$$

$$= (\widehat{A})^T + (\widehat{D})^T$$

where

$$(\widehat{D})^T = \begin{bmatrix} 0 & d_{-1} & \cdots & d_{-N_c-N_a} \\ 0 & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & d_{-1} \\ 0 & \vdots & \ddots & 0 \\ d_N & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ d_{N_a+N_c+N-1} & \cdots & d_N & 0 \end{bmatrix}.$$

Hence, using (3), we have

$$h_{\text{corr}} = \widehat{A}y_{[-N_a:N_c+N-1]} \\ = \widehat{A}(\widehat{A})^T h + \widehat{A}(\widehat{D})^T h + \widehat{A}v_{[-N_a:N_c+N-1]}$$

If the training sequence has been sent in isolation of transmission data (purely a sounding sequence), then the term $\widehat{A}(\widehat{D})^T h$ does not occur. In our case, the training sequence is embedded in the transmission data, and this term does appear. Loosely

speaking, the general idea of the correlation estimate is that due to the pseudonoise properties of the training sequence, $\widehat{A}(\widehat{A})^T$ is very close to the identity matrix, so we see that the correlation estimate approximates the actual channel, modulo the noise term $\widehat{A}v_{[-N_a:N_c+N-1]}$. This situation is worsened in the case of the training sequence embedded in transmission data due to the appearance of the term $\widehat{A}(\widehat{D})^T h$.

IV. FORMULATION OF THE LS CHANNEL ESTIMATION PROBLEM

Based on (1) and assuming that $N \geq (N_a + N_c + 1)$, we can write the pulse-matched filter output that corresponds “only” to the known training symbols compactly in vector notation as

$$y_{[N_c:N-N_a-1]} = \widetilde{A}h + \nu_{[N_c:N-N_a-1]} \\ = \widetilde{A}h + \widetilde{Q}\eta_{[N_c-L_q:N-N_a-1+L_q]}$$

where

$$\widetilde{A} = T \left\{ [a_{N_c+N_a}, \dots, a_{N-1}]^T, [a_{N_c+N_a}, \dots, a_0] \right\}$$

$$= \begin{bmatrix} a_{N_c+N_a} & a_{N_c+N_a-1} & \cdots & a_0 \\ a_{N_c+N_a+1} & a_{N_c+N_a} & \cdots & a_1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{N-1} & a_{N-2} & \cdots & a_{N-1-N_c-N_a} \end{bmatrix} \quad (6)$$

is the $(N - N_a - N_c) \times (N_a + N_c + 1)$ Toeplitz convolution matrix with first column $[a_{N_c+N_a}, \dots, a_{N-1}]^T$ and first row $[a_{N_c+N_a}, \dots, a_0]$, and $\nu_{[N_c:N-N_a-1]} = \widetilde{Q}\eta_{[N_c-L_q:N-N_a-1+L_q]}$ is the colored noise at the receiver matched filter output, where $\widetilde{Q} \in R^{(N-N_a-N_c) \times (N-N_a-N_c+2L_q)}$ is given by

$$\widetilde{Q} = \begin{bmatrix} q^T & 0 & \cdots & 0 \\ 0 & q^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q^T \end{bmatrix}$$

where q is the symbol rate sampled receiver pulse-matched filter, as previously defined.

The value of N , as imposed by the 8-VSB modulation system, is 704 (or 728 if the 24-mode symbols are incorporated into the training symbols). The channel length $N_a + N_c + 1$ is determined to be in excess of 500 symbols from the field test data. We use a channel length of 576 or 512. These values of parameters ensure that the LS system previously described is badly underdetermined and does not give a reliable CIR estimate.

V. FORMULATION OF THE OVERDETERMINED LS CHANNEL ESTIMATION PROBLEM

To write an overdetermined LS system, we write the pulse-matched filter output that includes “all” the contributions from

the known training symbols, including the output that includes adjacent unknown random data. This system is written as

$$\begin{aligned} y_{[-N_a:N+N_c-1]} &= (A + D)h + \nu_{[-N_a:N+N_c-1]} \\ &= Ah + Dh + Q\eta_{[-N_a-L_q:N+N_c-1+L_q]} \end{aligned} \quad (7)$$

where

$$A = T \left\{ \left[a_0, \dots, a_{N-1}, \underbrace{0, \dots, 0}_{N_a+N_c} \right]^T, \left[a_0, \underbrace{0, \dots, 0}_{N_a+N_c} \right] \right\} \quad (8)$$

is the $(N + N_a + N_c) \times (N_a + N_c + 1)$ Toeplitz matrix with first column $[a_0, \dots, a_{N-1}, 0, \dots, 0]^T$ and first row $[a_0, 0, \dots, 0]$, and

$$D = T \left\{ \left[\underbrace{0, \dots, 0}_N, d_N, \dots, d_{N_a+N_c+N-1} \right]^T, \left[0, \underbrace{d_{-1}, \dots, d_{-N_a-N_c}}_{\text{data from previous frame}} \right] \right\} \quad (9)$$

is a Toeplitz matrix that includes adjacent random information symbols only, both prior to and after the training sequence. We will only use the statistics of this random data (since the actual values are unknown) and use these statistics to compute the covariance matrix of the correlated noise, which will be used in the solution for the BLUE estimate of h . The data symbols $d_{-1}, \dots, d_{-N_a-N_c}$ denote the unknown information symbols transmitted at the end of the previous frame. $\nu_{[-N_a:N+N_c-1]} = Q\eta_{[-N_a-L_q:N+N_c-1+L_q]}$ is the colored noise at the receiver matched filter output, where $Q \in R^{(N+N_a+N_c) \times (N+N_a+N_c+L_q)}$ is given by

$$Q = \begin{bmatrix} q^T & 0 & \dots & 0 \\ 0 & q^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q^T \end{bmatrix}. \quad (10)$$

To compute the covariance matrix for the noise contribution $Dh + Q\eta_{[-N_a-L_q:N+N_c-1+L_q]}$ to the received vector $y_{[-N_a:N+N_c-1]}$, it is advantageous to rewrite the term Dh .

We define $d \in R^{N+2(N_c+N_a)}$ by

$$d = [d_{-N_c-N_a}, \dots, d_{-1}, 0_{1 \times N}, d_N, \dots, d_{N+N_c+N_a-1}]^T$$

and $H \in R^{(N+N_a+N_c) \times (N+2(N_c+N_a))}$ is the Toeplitz channel convolution matrix

$$H = \begin{bmatrix} \bar{h}^T & 0 & \dots & 0 \\ 0 & \bar{h}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{h}^T \end{bmatrix} \quad (11)$$

where $\bar{h}^T = [h[N_c], \dots, h[1], h[0], h[-1], \dots, h[-N_a]] \in R^{N_c+N_a+1}$ is the time reversal of the channel vector h^T . Then, we note that

$$Dh = Hd$$

and while it is difficult to obtain a closed-form expression for the expected value

$$E[Dh(Dh)^H] = E[Dh h^H D^H]$$

we can obtain a closed-form expression for the expected value

$$\begin{aligned} E[Hd(Hd)^H] &= HE[dd^H]H^H \\ &= \sigma_d^2 HSH^H \end{aligned}$$

where $E[dd^H] = \sigma_d^2 S$, and $S \in R^{(N+2(N_c+N_a)) \times (N+2(N_c+N_a))}$ is given by

$$S = \begin{bmatrix} I_{(N_c+N_a)} & 0_{(N_c+N_a) \times N} & 0_{(N_c+N_a)} \\ 0_{N \times (N_c+N_a)} & 0_N & 0_{N \times (N_c+N_a)} \\ 0_{(N_c+N_a)} & 0_{(N_c+N_a) \times N} & I_{(N_c+N_a)} \end{bmatrix}. \quad (12)$$

Note that this expression for matrix S is derived theoretically and is used exactly in this form in the algorithm. No estimation of S is required so that issues concerning finite block length, which would give rise to off-diagonal elements in S , do not arise. Since $Dh = Hd$, we may then rewrite (7) as

$$y_{[-N_a:N+N_c-1]} = Ah + Hd + Q\eta_{[-N_a-L_q:N+N_c-1+L_q]}.$$

VI. GENERAL LINEAR MODEL AND BLUE ESTIMATOR

A. General Linear Model for the Received Data

At the end of the previous section, we concluded that the received data vector $y_{[-N_a:N+N_c-1]}$ may be written in a form that is referred to as the ‘‘general linear model’’ (see [23] or [24]), where ‘‘general’’ refers to the fact that the noise may be correlated. The overall noise term $w = Hd + Q\eta_{[-N_a-L_q:N+N_c-1+L_q]}$ has a covariance matrix given by

$$\begin{aligned} \text{cov}(w) &= \frac{1}{2} E[ww^H] \\ &= \frac{\sigma_d^2}{2} HSH^H + \frac{\sigma_\eta^2}{2} QQ^H \\ &= \frac{\sigma_d^2}{2} \left(HSH^H + \frac{\sigma_\eta^2}{\sigma_d^2} QQ^H \right) \end{aligned}$$

since the data symbols $d[\cdot]$ and the AWGN term $\eta[\cdot]$ are uncorrelated. Using the fact that due to the choice of transmitter and receiver pulse shaping, the product QQ^H is very nearly equal to $\|q\|^2 I$, and $\sigma_v^2 = \sigma_\eta^2 \|q\|^2$, then we can give the further approximation

$$\text{cov}(w) \simeq \frac{\sigma_d^2}{2} \left(HSH^H + \frac{\sigma_v^2}{\sigma_d^2} I \right)$$

where I is the $(N + N_a + N_c) \times (N + N_a + N_c)$ identity matrix. This approximation is often useful in simulations. We note that since $H = H(h)$ from (11), then $\text{cov}(w)$ is a function of h , i.e., $\text{cov}(w)$ is a function of the CIR vector, which we are attempting to estimate.

We will define

$$C(h) = HSH^H + \frac{\sigma_\eta^2}{\sigma_d^2} QQ^H \quad (13)$$

so that $\text{cov}(w) = (\sigma_d^2/2)C(h)$. We often use the approximation

$$C(h) \simeq HSH^H + \frac{\sigma_\nu^2}{\sigma_d^2} I$$

especially in simulations.

B. Solution of the General Linear Model and the Gauss–Markoff Theorem

The BLUE estimate, which is the minimum variance unbiased estimator for the general linear model, is given by the Gauss–Markoff theorem (see [23]–[25] or [26]), which states in part that if the data observed can be modeled as

$$y = Ah + \eta$$

where y is an $N \times 1$ vector of observations, A is a known $N \times p$ observation matrix (with $N > p$ and $\text{rank } p$), h is a $p \times 1$ vector of parameters to be estimated, and η is an $N \times 1$ noise vector with mean and variance given by $E[\eta] = 0$ and $\text{var}(\eta) = C$, respectively, then the best linear unbiased estimator of h is

$$\hat{h}_{\text{BLUE}} = (A^T C^{-1} A)^{-1} A^T C^{-1} y$$

and the covariance matrix of \hat{h}_{BLUE} is

$$C_{\hat{h}_{\text{BLUE}}} = (A^T C^{-1} A)^{-1}.$$

In the case where the noise is white (i.e., $\text{var}(\eta) = \sigma^2 I$), this expression collapses to the standard LS solution of the normal equations, i.e., $\hat{h}_{\text{BLUE}} = (A^T A)^{-1} A^T y$, and the covariance matrix of \hat{h}_{BLUE} is $C_{\hat{h}_{\text{BLUE}}} = \sigma^2 (A^T A)^{-1}$.

VII. ITERATIVE ALGORITHM FOR FINDING THE BLUE CHANNEL ESTIMATE

For notational ease, from now on, we write y for $y_{[-N_a:N+N_c-1]}$. Using the Gauss–Markoff theorem, we see that the BLUE channel estimator for (7) is

$$h_{\text{BLUE}} = \left(A^T (C(h))^{-1} A \right)^{-1} A^T (C(h))^{-1} y \quad (14)$$

but as already noted, the right-hand side depends on the CIR vector h , which we are attempting to estimate. Note that the expression on the right-hand side of (14) is independent of the factor $\sigma_d^2/2$, which appears in the expression $\text{cov}(w) = (\sigma_d^2/2)C(h)$. This suggests using an iterative algorithm to

compute h_{BLUE} . Given an initial estimate $h^{(0)}$, for the CIR, we use the iteration

$$h^{(k+1)} = \left(A^T \left(C \left(h^{(k)} \right) \right)^{-1} A \right)^{-1} A^T \left(C \left(h^{(k)} \right) \right)^{-1} y \quad (15)$$

for $k = 0, 1, 2, \dots$. The covariance matrix at the k th stage is formed using

$$C(h^{(k)}) = \frac{\sigma_d^2}{2} H \left(h^{(k)} \right) S H \left(h^{(k)} \right)^H + \frac{\left(\sigma_\nu^{(k)} \right)^2}{2} I. \quad (16)$$

The covariance matrix depends on the channel noise variance σ_ν^2 (where $\sigma_\nu^2 = \|q\|^2 \sigma_\eta^2$). We estimate σ_ν^2 at the k th stage of the algorithm as $\left(\sigma_\nu^{(k)} \right)^2$, which is given by

$$\left(\sigma_\nu^{(k)} \right)^2 = \frac{1}{(N - N_a - N_c)} \left\| y_{[N_c:N-N_a-1]} - \tilde{A} h^{(k)} \right\|_2^2 \quad (17)$$

where \tilde{A} is given by (6). Note that \tilde{A} does not contain any unknown data but is the “short” matrix that contains only training values so that our estimate of the AWGN variance is not “contaminated” by energy due to the data variance, as would be the case if we mistakenly use matrix A in this estimate.

The data variance σ_d^2 is computed theoretically using knowledge of how the particular data is modeled as a random variable. For example, in the case of the ATSC 8-VSB system, the data is modeled as a uniform random variable on the set $\{\pm 1, \pm 3, \pm 5, \pm 7\}$, for which $\sigma_d^2 = 21$.

Hence, at the k th stage in the algorithm, we compute a new estimate for the CIR vector $h^{(k)}$. Using $h^{(k)}$, we form the new covariance matrix using (16) and then find the inverse $(C(h^{(k)}))^{-1}$. Finally, we compute $h^{(k+1)}$ using (15). Formally, we have the following algorithm:

Input:

Stored training vector: $t = [a_0, a_2, \dots, a_{N-1}]^T$

Vector of received values: y

Initial channel estimate $h^{(0)}$ (from correlation)

Matrix S

for $k = 0, 1, 2, \dots$

 Compute the noise estimate $\left(\sigma_\nu^{(k)} \right)^2$ using (17)

 Formation of the channel matrix $H(h^{(k)})$

 Formation of covariance matrix using (16)

 Compute $(C(h^{(k)}))^{-1}$

 Compute channel estimate $h^{(k+1)}$ using (15)

end for

Practical low-complexity methods for solving (15) in the **for** loop must be determined. We are presently working on such low-complexity approximations to this part of the algorithm [27]–[30].

TABLE I
BRAZILIAN CHANNEL PROFILES

Name	Delays (in symbols)	Path Gain (Relative)
Brazil-C	-0.957839182	0.7263
	0	1
	3.55153854	0.6457
	15.25009125	0.9848
	24.03207745	0.7456
Brazil-D	29.16566498	0.8616
	0	1
	5.16587424	0.65575
	22.27783266	0.75697
	31.2104902	0.87482
Brazil-E	61.45237898	1.01565
	62.20573564	0.7379
	0	1
	10.762238	1
	21.524476	1

VIII. SIMULATIONS FOR FINDING THE BLUE CHANNEL ESTIMATE

We show the simulation results for channels used in the high-definition TV (HDTV) literature, which are known as the Brazilian channels. We specifically use the Brazilian channels identified as C, D, and E described in Table I. We also give results for Brazilian channels A2, A3, and A5. In the interests of space, we do not explicitly give their delays and path gains in the table, but they are shown in Fig. 1. All the channels may be found in [1].

A. 2-Norm of Error

We show the simulation results for four iterations of the algorithm to demonstrate its performance. The 2-norm of the error is given at each iteration. The norm of the error is seen to be decreasing at each iteration. In practice, a stopping criterion for the iterative algorithm would be used [31]. We use two distinct initial approximations to the CIR in the simulation of the algorithm. ‘‘Correlation’’ in Table II denotes starting the algorithm from the correlation estimate for the CIR, whereas ‘‘aBLUE’’ denotes starting the algorithm from the initial approximation of the CIR vector as a delta function, i.e., for each CIR vector $h \in R^{N_a+N_c+1}$, we approximate h by $h_\delta = [0, \dots, 0, 1, 0, \dots, 0]^T \in R^{N_a+N_c+1}$, where the 1 appears in the $(N_a + 1)$ th position. More information concerning this approximation can be found in [27]–[29].

Note that the error values for the correlation estimate and aBLUE estimate are the initial values given in Table II for the iteration $k = 0$. Hence, we see that the iterations $k = 1, 2, 3$ improve on the correlation and aBLUE estimates.

We give the variance of the BLUE estimate, as given by the Gauss–Markoff theorem. We have $\text{var}(h_i^{\text{BLUE}}) = ((A^T C^{-1} A)^{-1})_{ii}$. We treat this quantity as a baseline noise value that the algorithm can approach. In Table III, we give the quantity $\|V\|$, which denotes

$$\left\| \left((A^T C^{-1} A)^{-1} \right)_{ii} \right\|_2$$

which is the 2-norm of the diagonal of matrix $(A^T C^{-1} A)^{-1}$, which we consider to be a vector.

We give the values for the norms of the Jacobian matrices for the function F given by (18) for each of the channels considered in Table IV. The norms considered are the 2-norm, 1-norm, ∞ -norm, and Frobenius norm (which is denoted as F-norm in the table). We note that in each case, it is the 2-norm of the Jacobian that is less than 1. This guarantees convergence of the proposed algorithm (see Section IX) and occurs for the initial approximation to the CIR vector h , which is sufficiently close to the actual value of h . We emphasize that we have encountered many channels with the norm of the Jacobian matrix that is not less than 1 with respect to any of the matrix norms, which we consider not to result in nonconvergence of the algorithm. The condition of the mapping, being a contraction mapping, is sufficient but is not necessary for convergence (see Section IX).

B. Symbol Error Rate for Brazil-D

In Fig. 2, we show the symbol error rate for a decision feedback equalizer (DFE) [32], [33], in operation with simulated 8-VSB data transmitted through the Brazil-D channel for a simulated Doppler channel for the frequency range of 0–75 Hz. The figure shows the symbol error rate plotted against the Doppler frequency in hertz, as the DFE taps evolve using an adaptive update [such as least mean square (LMS)] [34], [35]. We do not give the specific architecture used here. The plots are given for the three separate cases, namely 1) initializing the DFE with the iterative BLUE estimate after one iteration of our algorithm (denoted as BLUE in the figure), 2) initializing the DFE with the approximate BLUE estimate given in [27], (denoted as aBLUE in the figure), and 3) initializing the DFE with the correlation estimate of the CIR (denoted as XCORR in Fig. 2). The improved initial channel estimate obtained using the iterative BLUE estimate allows for operating at a higher Doppler rate than merely using the correlation estimate as the initialization for the channel. The initial channel estimate is used to compute minimum-mean-square-error (MMSE) DFE tap-weights, which are then updated adaptively as the channel evolves in time.

IX. PROOF OF CONVERGENCE OF THE ITERATIVE METHOD

A. Contraction Mappings and Fixed-Point Theorems

For simplicity of notation, we write $y = y_{[-N_a:N+N_c-1]}$. We consider the convergence of the iterative method by considering the iteration of the following function.

Definition 1: Let $F = F_{y,A} : R^{N_a+N_c+1} \rightarrow R^{N_a+N_c+1}$ be defined by

$$F(h) = F_{y,A}(h) = (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} y. \quad (18)$$

Since h_{BLUE} is given by the expression on the right-hand side of (18), then we see that h_{BLUE} is a fixed point of function F [2]. We give conditions under which F has a unique fixed point h_{BLUE} and for which the iteration converges to this unique fixed point h_{BLUE} . Background material concerning the notation used, differentials, partial derivatives, and the

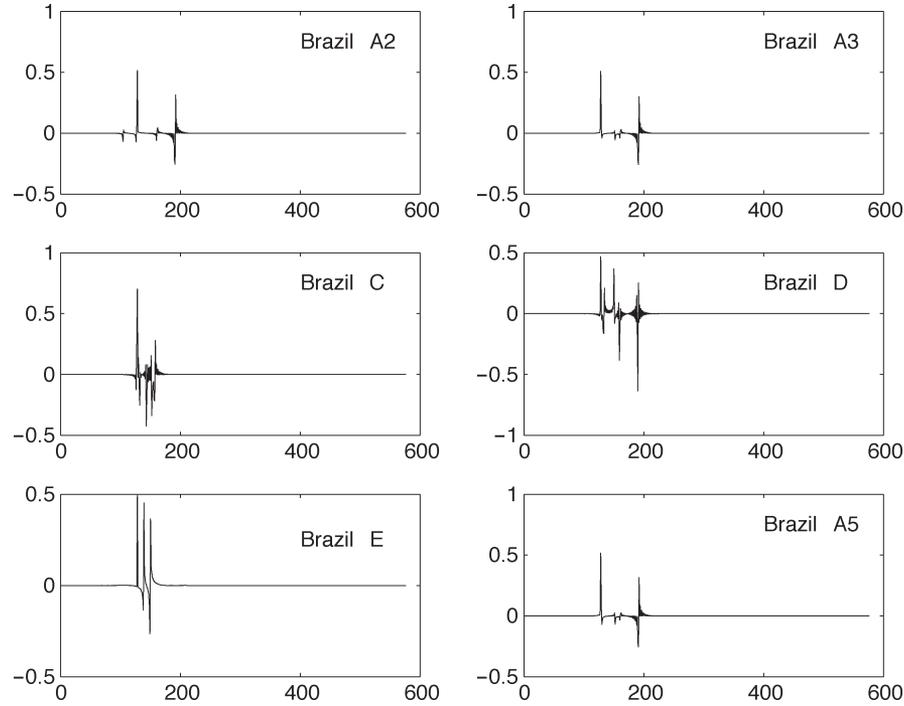


Fig. 1. Brazilian channels A2, A3, C, D, E, and A5.

TABLE II
ERROR 2-NORM

Channel	Initial.	Iter. 0	Iter. 1	Iter. 2	Iter. 3	Iter. 4
A2 18dB	Corr.	0.8967	0.4379	0.3772	0.3687	0.3644
A2 18dB	aBLUE	0.9089	0.4160	0.3696	0.3631	0.3636
A2 28dB	Corr.	0.8748	0.3512	0.2391	0.2223	0.2131
A2 28dB	aBLUE	0.9089	0.2369	0.2082	0.1974	0.1956
A3 18dB	Corr.	0.8838	0.4427	0.3837	0.3743	0.3674
A3 18dB	aBLUE	0.8348	0.3934	0.3647	0.3649	0.3664
A3 28dB	Corr.	0.8676	0.3568	0.2427	0.2306	0.2136
A3 28dB	aBLUE	0.8348	0.2060	0.2004	0.1937	0.1924
C 18dB	Corr.	0.8840	0.4777	0.4376	0.4323	0.4250
C 18dB	aBLUE	1.3575	0.4393	0.4196	0.4096	0.4153
C 28dB	Corr.	0.8619	0.3815	0.2823	0.2253	0.2374
C 28dB	aBLUE	1.3575	0.2731	0.2629	0.2584	0.2534
D 18dB	Corr.	0.8007	0.4348	0.3482	0.3252	0.3149
D 18dB	aBLUE	1.6033	0.4111	0.3467	0.3208	0.3139
D 28dB	Corr.	0.7837	0.1441	0.1264	0.1215	0.1173
D 28dB	aBLUE	1.6033	0.3022	0.2092	0.1555	0.1287
E 18dB	Corr.	0.8805	0.4430	0.3611	0.3473	0.3386
E 18dB	aBLUE	1.4147	0.4067	0.3618	0.3399	0.3369
E 28dB	Corr.	0.8669	0.3523	0.2095	0.1609	0.1458
E 28dB	aBLUE	1.4147	0.2326	0.1839	0.1761	0.1475
A5 18dB	Corr.	0.8827	0.4368	0.3829	0.3736	0.3679
A5 18dB	aBLUE	0.8948	0.4038	0.3652	0.3658	0.3668
A5 28dB	Corr.	0.8680	0.3471	0.2455	0.2317	0.2182
A5 28dB	aBLUE	0.8948	0.2227	0.2094	0.2049	0.2022

TABLE III
BASELINE NOISE

Ch.	A2	A3	C	D	E	A5
$\ V\ $	0.1347	0.1330	0.1594	0.1408	0.1511	0.1700

Jacobian mapping and matrix representation can be found in [36, Sec. 8.2]. We state the following theorems, which outline the conditions for convergence of the iterative method.

Theorem 1: Let U be a convex open set in $R^{N_a+N_c+1}$ and $F_{y,A} : U \rightarrow R^{N_a+N_c+1}$, as defined in (18). We let $\|\cdot\|_p$ denote

TABLE IV
NORM $DF_{y,A}$

Channel	Norm			
	$2 - norm$	$1 - norm$	$\infty - norm$	$F - norm$
Brazil-A2 18dB	0.5033	1.9830	2.3420	2.0819
Brazil-A3 18dB	0.4380	1.8924	2.2633	2.0063
Brazil-C 18dB	0.7590	2.6366	2.9145	2.6654
Brazil-D 18dB	0.4825	2.0219	2.6750	2.2276
Brazil-E 18dB	0.8260	2.9078	3.4205	2.7770
Brazil-A5 18dB	0.4533	1.9487	2.3511	2.0704
Brazil-A2 28dB	0.4980	1.9836	2.3403	2.0791
Brazil-A3 28dB	0.4368	1.9023	2.2592	2.0026
Brazil-C 28dB	0.7709	2.6632	2.9291	2.6723
Brazil-D 28dB	0.5045	2.0382	2.6833	2.2386
Brazil-E 28dB	0.8211	2.8801	3.3942	2.7633
Brazil-A5 28dB	0.4524	1.9617	2.3617	2.0701

the vector p -norm (for $p = 1$ or 2) and use this notation to denote the matrix p -norm (again, for $p = 1$ or 2), where whether we are using the vector or matrix norm is clear from the context. If $F_{y,A}$ is differentiable at each point of U and if

$$\|DF_{y,A}(h)\|_p \leq M, \quad \forall h \in U \quad (19)$$

(where $DF_{y,A}$ denotes the Jacobian matrix of $F_{y,A}$), then we have

$$\|F_{y,A}(h_1) - F_{y,A}(h_2)\|_p \leq M \|h_1 - h_2\|_p, \quad \forall h_1, h_2 \in U. \quad (20)$$

Note that $DF_{y,A} \in R^{(N_a+N_c+1) \times (N_a+N_c+1)}$ is given by

$$\begin{aligned} DF_{y,A} &= [D_1 F_{y,A}, \dots, D_L F_{y,A}] \\ &= \left[\frac{\partial F_{y,A}(h)}{\partial h_1}, \dots, \frac{\partial F_{y,A}(h)}{\partial h_L} \right]. \end{aligned}$$

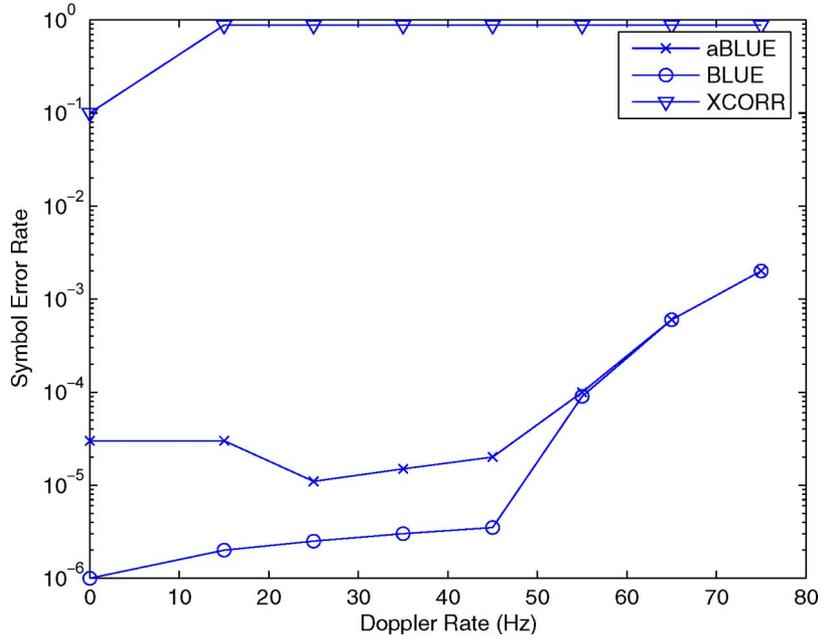


Fig. 2. DFE SER for increasing Doppler: Brazil-D. (Color version available online at <http://ieeexplore.ieee.org>.)

Proof: The proof for 2-norms follows directly from the results in [36]. The proof for 1-norms is based on the proofs in [36, Th. 8.17, Corollary 8.18, Corollary 8.19], with the 2-norms replaced by 1-norms. Kantorovich and Akilov [37] give more general results for metric spaces. ■

We have used the 1-norm for ease of computation concerning our specific example. Specifically, the matrix 1-norm affords a nice formula for computation as the maximum column sum [31], i.e., for $A = [a_{ij}] \in R^{N \times N}$, $\|A\|_1 = \max_{1 \leq j \leq N} \sum_{i=1}^N |a_{ij}|$. There is no corresponding easy-to-compute formula for the matrix 2-norm. However, when using Matlab for simulations, we may also compute the 2-norm of the Jacobian matrix with relative ease. A function that satisfies an inequality of the form (20), where $M < 1$, is called contraction mapping [36], [37]. The following theorem may be found in a general form in [37].

Theorem 2: Let p be equal to 1 or 2. Let U be a convex open set in $R^{N_a+N_c+1}$ and $F_{y,A} : U \rightarrow R^{N_a+N_c+1}$, as defined in (18). If $F_{y,A}$ is differentiable at each point of U and if

$$\|DF_{y,A}(h)\|_p \leq M, \quad \forall h \in U \quad (21)$$

and $M < 1$, then function $F_{y,A}$ has a unique fixed point h_0 on U , and the iteration

$$h^{(k+1)} = F_{y,A}(h^{(k)}), \quad k = 0, 1, 2, \dots$$

used in the iterative BLUE algorithm produces a sequence $\{h^{(k)}\}_{k=0}^{\infty}$, which converges to this unique fixed point h_0 of $F_{y,A}$, i.e.,

$$\lim_{k \rightarrow \infty} h^{(k)} = h_0$$

and $h_{\text{BLUE}} = h_0$, where h_{BLUE} is the BLUE of the CIR vector.

Proof: The proof is straightforward and is given in the Appendix. ■

B. Computation of the Jacobian Matrix

We state the following lemma concerning the differentiation of the inverse of a matrix.

Lemma 3: For any matrix $B = B(\theta) \in R^{N \times N}$ (which depends on a vector parameter $\theta = [\theta_1, \dots, \theta_K] \in R^K$ such that there exists an open set $U \subset R^K$) such that B is nonsingular and differentiable for all $\theta \in U$, then we have for any i , $1 \leq i \leq K$, and for any $\theta \in U$

$$\frac{\partial B^{-1}(\theta)}{\partial \theta_i} = -B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \theta_i} B^{-1}(\theta).$$

Proof: Differentiating both sides of $B(\theta)B^{-1}(\theta) = I$ with respect to θ_i and using the product rule, we obtain $(\partial B(\theta)/\partial \theta_i)B^{-1}(\theta) + B(\theta)(\partial B^{-1}(\theta)/\partial \theta_i) = O_N$, from which the result follows. ■

Concerning $C(h)$, we have the following.

Proposition 4: $C(h)$ is a differentiable function of h for all $h \in R^{N_a+N_c+1}$. $C(h)$ is a positive definite matrix for all $h \in R^{N_a+N_c+1}$. Consequently, $C(h)$ is invertible for all $h \in R^{N_a+N_c+1}$, and $C^{-1}(h)$ is differentiable for all $h \in R^{N_a+N_c+1}$.

Proof: The fact that $C(h)$ is a differentiable function of h for all $h \in R^{N_a+N_c+1}$ is clear from (13). $C(h)$ is a positive definite matrix for all $h \in R^{N_a+N_c+1}$ since for all $h \in R^{N_a+N_c+1}$ and for all nonzero $x \in R^{N_a+N_c+N}$, we have

$$\begin{aligned} \langle C(h)x, x \rangle &= \langle HSH^H x, x \rangle + \frac{\sigma_\eta^2}{\sigma_d^2} \langle QQ^H x, x \rangle \\ &= \|SH^H x\|^2 + \frac{\sigma_\eta^2}{\sigma_d^2} \|Q^H x\|^2 \\ &\geq \frac{\sigma_\eta^2}{\sigma_d^2} \|Q^H x\|^2 \\ &> 0, \quad \text{for all } x \in R^{N_a+N_c+N}, x \neq 0 \end{aligned}$$

since Q^H is of full column rank, provided that $\sigma_\eta^2 \neq 0$. We have used the fact that $S = S^T S$. Hence, $C(h)$ is a positive definite matrix for all $h \in R^{N_a+N_c+1}$ and, consequently, is nonsingular for all $h \in R^{N_a+N_c+1}$. $C^{-1}(h)$ is differentiable for all $h \in R^{N_a+N_c+1}$, following Lemma 3. ■

We notice that the proof hinges on the fact that the AWGN variance σ_η^2 is nonzero. This is often the case and is encountered in engineering applications under the term diagonal loading (see, e.g., [38]), which is a method to improve matrix conditioning, and in mathematical applications, a similar role is played by a regularization parameter [25]. Concerning the differentiability of F , we have

Proposition 5: The function F given by (18) is differentiable for all $h \in R^{N_a+N_c+1}$.

Proof: Since $C(h)$ is nonsingular and differentiable for all $h \in R^{N_a+N_c+1}$, then $A^T C(h)^{-1} A$ is nonsingular and differentiable for all $h \in R^{N_a+N_c+1}$; therefore, $(A^T C(h)^{-1} A)^{-1}$ is differentiable by Lemma 3. Hence, F is a product of differentiable functions and is differentiable for all $h \in R^{N_a+N_c+1}$. ■

The partial derivatives of the function F are given by the following Proposition.

Proposition 6: For $1 \leq i \leq N_a + N_c$ and for $F_{y,A}$, as defined in (18), we have

$$\frac{\partial F_{y,A}(h)}{\partial h_i} = A_C(h) \frac{\partial C(h)}{\partial h_i} C(h)^{-1} \{A A_C(h) - I\} y$$

where $A_C(h)$ is defined by

$$A_C(h) = (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1}.$$

Proof: Differentiate, using the product rule [39], with respect to h_i , the expression for $A_C(h)y$, and use Lemma 3 thrice. Details are given in the Appendix. ■

Concerning the Jacobian matrix, we have the following.

Proposition 7: The Jacobian matrix $DF_{y,A}(h)$ for function $F_{y,A}(h)$ is given by

$$\begin{aligned} DF_{y,A}(h) &= [D_1 F_{y,A}(h), \dots, D_{(N_a+N_c+1)} F_{y,A}(h)] \\ &= \left[\frac{\partial F_{y,A}(h)}{\partial h_1}, \dots, \frac{\partial F_{y,A}(h)}{\partial h_{(N_a+N_c+1)}} \right] \end{aligned} \quad (22)$$

where the column vector $\partial F_{y,A}(h)/\partial h_i$ is given by

$$\frac{\partial F_{y,A}(h)}{\partial h_i} = \left[\frac{\partial (F_{y,A}(h))_1}{\partial h_i}, \dots, \frac{\partial (F_{y,A}(h))_L}{\partial h_i} \right]^T \quad (23)$$

where

$$F_{y,A}(h) = [(F_{y,A}(h))_1, \dots, (F_{y,A}(h))_L]^T \in R^{N_a+N_c+1}.$$

Proof: Equation (22) is given after [36, Prop. 8.21], whereas (23) is given by [36, Prop. 6]. ■

Concerning convergence with respect to the different topologies generated by different norms, it is a well-known result that all norms are equivalent on a finite dimensional vector space and, thus, generate the same topology [31], [37], [40].

Consequently, given a sequence $\{x_n\}_{n=1}^\infty \subset V$, which converges to a limit point x with respect to the topology gen-

erated by the norm $\|\cdot\|$ (i.e., $\lim_{n \rightarrow \infty} \|x_n - x\| = 0$), then this sequence converges to the same limit point x with respect to the topology generated by the norm $\|\cdot\|_*$ (i.e., $\lim_{n \rightarrow \infty} \|x_n - x\|_* = 0$). Hence, convergence with respect to a particular norm implies convergence with respect to any equivalent norm. Therefore, it is irrelevant what norm we use to show convergence of the iterative algorithm, as convergence with respect to a particular norm implies convergence with respect to the topology generated by any given norm.

X. CONCLUSION

We have proposed a semiblind iterative algorithm which works within the ATSC DTV standard to achieve BLUE channel estimation. We claim that the algorithm generalizes to any communication system that utilizes a periodically transmitted training sequence embedded in data. The BLUE channel estimate is given by solving Gauss–Markoff equations. Due to our system constraints, to formulate an overdetermined system, we encounter the added complication that the noise covariance matrix is actually a function of the CIR, which we are trying to find. This suggests an iterative algorithm, where an initial approximation to the CIR is used to generate a noise covariance matrix and, hence, a better approximation to the CIR. We give conditions under which the iteration is represented by a contraction mapping and, thus, converges with a linear convergence of errors. We present simulation examples demonstrating superior channel estimation of the iterative algorithm over other previously used methods for Brazilian channels [1]. We present simulation examples that demonstrate the superior performance of an adaptive DFE for a Doppler channel after initialization using the iterative algorithm. In each case, for the simulations, we make a comparison between the correlation method of channel estimation and the approximate method given in [27].

APPENDIX

A. Proof of Theorem 2

We break the proof into several parts.

1) $\{h^{(k)}\}_{k=0}^\infty$ is a Cauchy sequence in $R^{N_a+N_c+1}$.

For arbitrary positive integer k and for $p = 1$ or 2 , we have

$$\begin{aligned} \left\| h^{(k)} - h^{(k-1)} \right\|_p &= \left\| F_{y,A} \left(h^{(k-1)} \right) - F_{y,A} \left(h^{(k-2)} \right) \right\|_p \\ &\leq M \left\| h^{(k-1)} - h^{(k-2)} \right\|_p \\ &\leq \dots \\ &\leq M^{k-1} \left\| h^{(1)} - h^{(0)} \right\|_p. \end{aligned}$$

For M , as given in (21), and for arbitrary $\varepsilon > 0$, we select a positive integer k such that

$$M^k < \frac{(1-M)\varepsilon}{\left\| h^{(1)} - h^{(0)} \right\|_p}.$$

Then, for any positive integer l , we have

$$\begin{aligned}
\|h^{(k+l)} - h^{(k)}\|_p &\leq \|h^{(k+l)} - h^{(k+l-1)}\|_p \\
&\quad + \|h^{(k+l-1)} - h^{(k+l-2)}\|_p \\
&\quad + \dots + \|h^{(k+1)} - h^{(k)}\|_p \\
&\leq \{M^{k+l-1} + M^{k+l-2} + \dots + M^k\} \\
&\quad \times \|h^{(1)} - h^{(0)}\|_p \\
&= \frac{M^k - M^{k+l-1}}{1 - M} \|h^{(1)} - h^{(0)}\|_p \\
&\leq M^k \left(\frac{\|h^{(1)} - h^{(0)}\|_p}{1 - M} \right) \\
&< \varepsilon.
\end{aligned}$$

Hence, $\{h^{(k)}\}_{k=0}^{\infty}$ is a Cauchy sequence in $R^{N_a+N_c+1}$.

- 2) Since $R^{N_a+N_c+1}$ is complete and $\{h^{(k)}\}_{k=0}^{\infty}$ is a Cauchy sequence in $R^{N_a+N_c+1}$, then by the definition of a complete topological space, there exists h_0 such that $\lim_{k \rightarrow \infty} h^{(k)} = h_0$.
- 3) h_0 , which is defined as the limit of the Cauchy sequence $\{h^{(k)}\}_{k=0}^{\infty}$, is the unique fixed point of function $F_{y,A}$. We have

$$\begin{aligned}
F_{y,A}(h_0) &= F_{y,A} \left(\lim_{k \rightarrow \infty} h^{(k)} \right) \\
&= \lim_{k \rightarrow \infty} F_{y,A}(h^{(k)}) \\
&= \lim_{k \rightarrow \infty} h^{(k+1)} \\
&= h_0
\end{aligned}$$

where we have used the definition of h_0 as $\lim_{k \rightarrow \infty} h^{(k)}$, the fact that $F_{y,A}$ is continuous, and the definition of the iteration $h^{(k+1)} = F_{y,A}(h^{(k)})$. Hence, h_0 is a fixed point of function $F_{y,A}$. To show the uniqueness of this fixed point, assume by way of contradiction that there are two such distinct fixed points a and b , i.e., $F_{y,A}(a) = a$, $F_{y,A}(b) = b$, and $a \neq b$. Then, since a and b are fixed points of $F_{y,A}$

$$\begin{aligned}
\|a - b\|_p &= \|F_{y,A}(a) - F_{y,A}(b)\|_p \\
&\leq M \|a - b\|_p \text{ by (20)} \\
&< \|a - b\|_p \text{ since } M < 1
\end{aligned}$$

which is a contradiction. Hence, no such two distinct fixed points exist; therefore, the fixed point h_0 of $F_{y,A}$ is unique.

- 4) Clearly, the sequence $\{h^{(k)}\}_{k=0}^{\infty}$, which our iterative process generates, converges to the unique fixed point h_0 of $F_{y,A}$ since h_0 was defined as the limit point of the sequence. The BLUE channel estimate is defined to be the solution of

$$h = (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} y$$

i.e., it is the unique fixed point of function $F_{y,A}$, in which case, $h_0 = h_{\text{BLUE}}$. ■

B. Proof of Proposition 6

We have

$$\begin{aligned}
\frac{\partial F(h)}{\partial h_i} &= \frac{\partial}{\partial h_i} \left((A^T C(h)^{-1} A)^{-1} \right) A^T C(h)^{-1} y \\
&\quad + (A^T C(h)^{-1} A)^{-1} A^T \frac{\partial}{\partial h_i} (C(h)^{-1}) y \\
&= - (A^T C(h)^{-1} A)^{-1} A^T \frac{\partial}{\partial h_i} (C(h)^{-1}) \\
&\quad \times A (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} y \\
&\quad - (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} \\
&\quad \times \left(\frac{\partial}{\partial h_i} C(h) \right) C(h)^{-1} y \\
&= (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} \left(\frac{\partial}{\partial h_i} C(h) \right) \\
&\quad \times C(h)^{-1} A (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} y \\
&\quad - (A^T C(h)^{-1} A)^{-1} A^T C(h)^{-1} \\
&\quad \times \left(\frac{\partial}{\partial h_i} C(h) \right) C(h)^{-1} y
\end{aligned}$$

which gives the result. ■

REFERENCES

- [1] ATSC Digital Television Standard, A/53. (Sep. 1995). [Online]. Available: <http://www.atsc.org/standards.html>
- [2] E. Zeidler, *Applied Functional Analysis Applications to Mathematical Physics: Applied Mathematical Sciences*, vol. 108. New York: Springer-Verlag, 1995.
- [3] D. H. Griffel, *Applied Functional Analysis*. Mineola, NY: Dover, 2002.
- [4] C. Byrne, "A unified treatment of some iterative algorithms in signal processing and image reconstruction," *Inverse Probl.*, vol. 20, no. 1, pp. 103–120, Feb. 2004.
- [5] H. Arslan and G. E. Bottomly, "Channel estimation in narrowband wireless communication systems," *Wireless Commun. Mobile Comput.*, vol. 1, no. 2, pp. 201–219, Apr.–Jun. 2001.
- [6] H.-K. Song, "A channel estimation algorithm using sliding window approach and tuning algorithm for MLSE," *IEEE Commun. Lett.*, vol. 3, no. 7, pp. 211–213, Jul. 1999.
- [7] E. de Carvalho and D. T. M. Slock, "Semi-blind methods for FIR multichannel estimation," in *Signal Processing Advances in Wireless and Mobile Communications: Trends in Channel Estimation and Equalization*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [8] —, "Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1053–1064, Apr. 2004.
- [9] A. Medles, D. T. M. Slock, and E. De Carvalho, "Linear prediction based semi-blind estimation of MIMO FIR channels," in *Proc. 3rd IEEE Signal Process. Workshop Signal Process. Adv. Wireless Communications*, Taoyuan, Taiwan, R.O.C., Mar. 20–23, 2001, pp. 58–61.
- [10] S. Lasaulce, P. Loubaton, and E. Moulines, "A semi-blind channel estimation technique based on second-order blind method for CDMA systems," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1894–1904, Jul. 2003.
- [11] V. Buchoux, O. Cappé, E. Moulines, and A. Gorokhov, "On the performance of semi-blind subspace based channel estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1750–1759, Jun. 2004.
- [12] A. Rontogiannis, A. Marava, K. Berberidis, and J. Palicot, "Semi-blind estimation of multi-path channel parameters via a separable least squares approach," in *Proc. 14th Int. Conf. DSP*, Jul. 1–3, 2002, vol. 1, pp. 49–52.

- [13] S. Özen, M. D. Zoltowski, and M. Fimoff, "A novel channel estimation method: Blending correlation and least squares based approaches," in *Proc. ICASSP*, 2002, vol. 3, pp. 2281–2284.
- [14] S. Özen, W. Hillery, M. D. Zoltowski, S. M. Nerayanuru, and M. Fimoff, "Structured channel estimation based decision feedback equalizers for sparse multipath channels with applications to digital TV receivers," in *Proc. Rec. 36th Asilomar Conf. Signals, Syst. and Comput.*, vol. 1, pp. 558–564.
- [15] C. Pladdy, S. Özen, S. M. Nerayanuru, M. D. Zoltowski, and M. Fimoff, "Best linear unbiased channel estimation for frequency selective multipath channels with long delay spreads," in *Proc. Rec. Veh. Technol. Conf.—Fall*, Orlando, FL, Oct. 2003, pp. 1238–1242.
- [16] D. C. Cox, "Delay Doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment," *IEEE Trans. Antennas Propag.*, vol. AP-20, no. 5, pp. 625–635, Sep. 1972.
- [17] D. Devasirvatham, "Time delay spread and signal level measurements of 850 MHz radio waves in building environment," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 11, pp. 1300–1305, Nov. 1986.
- [18] J. Parsons, D. Demery, and A. Turkmani, "Sounding techniques for wideband mobile radio channels: A review," *Proc. Inst. Electr. Eng.—I*, vol. 138, no. 5, pp. 437–446, Oct. 1991.
- [19] W. W. Jibrail and A.-R. Houmadi, "Acquisition of direct sequence spread spectrum signals using sliding correlators," *Int. J. Electron.*, vol. 71, no. 5, pp. 733–743, 1991.
- [20] J. Ruprecht, "Maximum likelihood estimation of multipath channels," Ph.D. dissertation, Swiss Federal Inst. Technol., Zurich, Switzerland, 1989.
- [21] Z. Kostić, M. Sezan, and E. Titlebaum, "Estimation of the parameters of a multipath channel using set-theoretic convolution," *IEEE Trans. Commun.*, vol. 40, no. 6, pp. 1006–1011, Jun. 1992.
- [22] C. A. Montemayor and P. G. Flikkema, "Near-optimum iterative channel estimation of dispersive multipath channels," in *Proc. Rec. Veh. Technol. Conf.*, 1998, pp. 2246–2250.
- [23] S. L. Kay, *Fundamentals of Statistical Signal Processing*, vol. 1. Upper Saddle River, NJ: Prentice-Hall, 1993.
- [24] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Upper Saddle River, NJ: Prentice-Hall, 1995.
- [25] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM, 2002.
- [26] A. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA: SIAM, 1996.
- [27] S. Özen, M. Fimoff, C. Pladdy, S. Nerayanuru, and M. D. Zoltowski, "Approximate best linear unbiased channel estimation for frequency selective multipath channels with long delay spreads," in *Proc. Asilomar Conf. Signals and Syst.*, Monterey, CA, Nov. 2003, pp. 1122–1127.
- [28] C. Pladdy, S. M. Nerayanuru, M. Fimoff, P. Ding, S. Özen, and M. D. Zoltowski, "Taylor series approximation for low complexity semi-blind best linear unbiased channel estimates for the general linear model with applications to DTV," in *Proc. Asilomar Conf. Signals and Syst.*, Monterey, CA, Nov. 2004, pp. 2208–2212.
- [29] C. Pladdy, S. M. Nerayanuru, M. Fimoff, S. Özen, and M. D. Zoltowski, "Taylor series approximation of semi-blind best linear unbiased channel estimates for the general linear model," in *Proc. MILCOM Conf.*, Nov. 2004, pp. 1509–1514.
- [30] —, "Taylor series approximation of semi-blind BLUE channel estimates with applications to DTV," *Inverse Probl. Sci. Eng.*, to be published.
- [31] D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing*. Pacific Grove, CA: Brooks/Cole, 1991.
- [32] J. Proakis, *Digital Communications*. New York: McGraw-Hill, 2001.
- [33] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. New York: Plenum, 2001.
- [34] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [35] B. Farhang-Boroujeny, *Adaptive Filters Theory and Applications*. New York: Wiley, 1999.
- [36] A. Browder, *Mathematical Analysis: An Introduction*. New York: Springer-Verlag, 1996.
- [37] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*. New York: Pergamon, 1964.
- [38] J. D. Hiemstra, M. E. Weippert, H. N. Nguyen, and J. S. Goldstein, "Insertion of diagonal loading into the multistage Wiener filter," in *Proc. Sensor Array and Multichannel Signal Process. Workshop*, Aug. 4–6, 2002, pp. 379–382.
- [39] S. Lang, *Undergraduate Analysis*, 2nd ed. New York: Springer-Verlag, 1997.
- [40] J. Munkres, *Topology*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.



Christopher Pladdy was born in Cardiff, Wales, U.K., in 1965. He received the B.Sc. degree in mathematics from Cardiff University, in 1987, the M.S. degree in mathematics and Ph.D. degree in applied mathematics from the University of Alabama, Birmingham, in 1991 and 1994, respectively, and the M.S. degree in electrical engineering with an emphasis in signal processing and digital communication from the University of Florida, Gainesville, in 2001. His dissertation was in the areas of functional analysis and mathematical physics.

From 1994 to 1998, he was an Assistant Professor of mathematics with Nicholls State University, Thibodaux, LA, and from 1998 to 2001, he was an Adjunct Lecturer in mathematics with the University of Florida, while working toward the M.S. degree in electrical engineering. He joined Zenith Electronics R&D Center, Lincolnshire, IL, in 2001, as a Research Engineer, working primarily on equalization of terrestrial high-definition TV signals, including mobile reception of such signals. Since September 2005, he has been a Lead Operations Research Analyst with the MITRE Corporation, Leavenworth, KS.



Serdar Özen was born in Ankara, Turkey, in 1970. He received the B.S.E.E. degree from Bogaziçi University, Istanbul, Turkey, in 1992, the M.S.E.E.x and M.S. degrees in mathematics from Rensselaer Polytechnic Institute, Troy, NY, both in 1996, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 2003.

From 1992 to 1993, he was a Project Engineer with the Ministry of National Defense, Undersecretariat for Defense Industries, Ankara, Turkey. He was a Visiting Research Engineer from 2001 to 2003 and a Senior Research engineer from 2003 to 2004 with the Zenith Electronics R&D Center, Lincolnshire, IL. He is currently an Assistant Professor with the Department of Electrical and Electronics Engineering, Izmir Institute of Technology, Izmir, Turkey. He is also the Coordinator of the Wireless Communication Networks and Multimedia Research Center, Izmir Institute of Technology, leading several industrial contracts for national and European Union projects. His main research interests are in the area of estimation theory applications for digital communications, wireless systems, and signal processing.



Sreenivasa M. Nerayanuru was born in Nandyal, Andhra Pradesh, India. He received the B.Tech. degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Kakinada, India, in 1998 and the M.S. degree in electrical engineering with emphasis on digital signal processing from the University of Missouri-Rolla, in 2000.

He joined Zenith Electronics R&D Center, Lincolnshire, IL, as an Engineer, working primarily on improved reception of terrestrial high-definition television signals, including equalization, mobile reception, channel estimation, noise prediction, and interference cancellation. Since August 2005, he has been a Senior Electronics Engineer with Motorola Inc., Libertyville, IL.



Michael Zoltowski, biography not available at the time of publication.



Mark Fimoff (M'76) received the degree from the University of Illinois, Urbana-Champaign.

He is currently a member of the Technical Staff at Techwell R&D, Chicago, IL. He was with Zenith from 1983 to 2005, where he contributed to the development of the eight-level vestigial sideband modulation digital television transmission standard and was involved in digital communications system development as well as digital image processing. He holds 40 patents, has 30 patents pending, and has a number of publications in these areas. His current

research interests include adaptive equalization and channel estimation.