ELSEVIER

Signal Processing 53 (1996) 211–229

SIGNAL PROCESSING

Estimation and equalization of fading channels with random coefficients[1]

Michail K. Tsatsanis[a,*], Georgios B. Giannakis[b], Guotong Zhou[c]

[a]Electrical Engineering Department, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[b]Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442, USA
[c]School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 1 March 1995; revised 15 April 1996

Abstract

The time-varying tap coefficients of frequency-selective fading channels are typically modeled as random processes with low-pass power spectra. However, traditional adaptive techniques usually make no assumption on the channel's time variations and hence do not exploit this information. In this paper, Kalman filtering methods are derived to track the channel by employing a multichannel autoregressive description of the time-varying taps in a decision-feedback equalization framework. Fitting a model to the variations of the channel's taps is a challenging task because the tap coefficients are not observed directly. Higher-order statistics are employed in this paper in order to estimate the model parameters from input/output data. Consistency of the proposed method is shown, and some illustrative simulations are presented.

Zusammenfassung

Die zeitvarianten Impulsantwortkoeffizienten von Kanälen mit frequenzselektivem Schwund werden typisch als Zufallsprozesse mit Tiefpaß–Leistungsspektren modelliert. Traditionelle Adaptionsverfahren treffen jedoch keine Annahme über die zeitliche Kanalvariation and nützen daher diese Information nicht aus. In dieser Arbeit werden Kalmanfilterverfahren zur Kanalverfolgung hergeleitet, indem eine mehrkanalige autoregressive Beschreibung der zeitvarianten Koeffizienten in einem entscheidungsrückgekoppelten Entzerrer eingesetzt wird. Die Anpassung eines Modells an die Variationen der Kanalkoeffizienten stellt eine anspruchsvolle Aufgabe dar, da die Koeffizienten nicht direkt beobachtet werden. In dieser Arbeit werden Statistiken höherer Ordnung eingesetzt, um die Modellparameter aus Eingangs/Ausgangsdaten zu schätzen. Die Konsistenz des vorgeschlagenen Verfahrens wird gezeigt und einige Simulationen werden zur Illustration vorgestellt.

Résumé

Les coefficients à variation temporelle de canaux d'étouffement sélectifs en fréquence sont typiquement modélisés comme des processus aléatoires caractérisés par des spectre de puissance passe-bas. Cependant, les techniques adaptatives traditionnelles ne font habituellement aucune hypothèse sur les variations temporelles du canal, et n'exploitent donc pas cette information. Dans ce papier, des méthodes de filtrage de Kalman sont dérivées pour suivre le canal en utilisant une

[1] This work was supported by the National Science Foundation Grant NSF-MIP 9424305.

* Corresponding author.

PII S0165-1684(96)00087-4

description multicanal autorégressive des coefficients dans un modèle d'égalisation par rétroaction de décisions. Ajuster un modèle aux variations des coefficients du canal est une tache ardue, parce que les coefficients ne sont pas observés directement. Ici, des méthodes statistiques d'ordres supérieurs sont employées pour estimer les paramètres du modèle à partir des données en entrée/sortie. La consistence de la méthode proposée est montrée, et quelques simulations sont présentées en guise d'illustration.

1. Introduction

Frequency-selective fading is encountered in several communications applications and presents a major impeding factor for high-speed digital transmission. It may occur because of random changes in the reflection medium (e.g., in ionospheric channels [20, Chapter 7]) or due to transmitter/receiver motion (e.g., TDMA[2] mobile radio channel, appearing in cellular telephony [2, 8]). Frequency selective fading is also common in underwater communications due to random local changes in the salinity and temperature of the ocean [14].

Fast adaptive algorithms, of the recursive least-squares (RLS) family, are typically employed in fading environments to track the channel's variations and equalize the received symbols [3, 23]. Periodic retraining and (when possible) diversity combining are commonly used to guard against deep fades and runaway effects.

In order to analyze the performance of these algorithms under fading conditions, a number of stochastic models have been used to describe the channel's time evolution [15; 14; 18; 20, Chapter 7] Each tap coefficient is usually considered as a random process with respect to time, as opposed to a constant in time-invariant channels. The statistical description, or average characteristics of the taps, reflect the average characteristics of the channel. Similarly, in general linear regression problems arising in systems theory, statistics and economics, random coefficient models have been used to account for unmodeled dynamics or perturbations of the system under study. In most of these applications, however, the problem of interest is to estimate the mean of the random coefficients, rather than track their time variations [4, 7, 12, 21, 11]. On the contrary, in communications applications, tracking of the time varying (TV) coefficients is of utmost importance. Moreover, a richer structure exists in fading channel models, compared with simpler i.i.d.[3] perturbations often adopted in regression theory [21]. The channel taps are usually assumed to be low-pass circular complex Gaussian processes [6, p. 89] in order to model progressive time variations (slow fading) [17; 16; 20; Chapter 7]. They may have either zero mean (Rayleigh fading) or non-zero mean (Rician fading), depending on the presence or absence of a line-of-sight path [18].

The information on the channel dynamics, provided by this stochastic framework, can be exploited to develop more accurate tracking algorithms. However, in most cases, these models have only been used for analyzing and simulating *existing* algorithms in fading environments, rather than improving them.

An exception to this rule is the work of Iltis and Fuxjaeger, who have proposed Kalman filtering ideas to track the code delay, Doppler shift and multipath parameters in a direct-sequence, spread-spectrum communication link [13, 14, 9, 10]. They model the TV tap coefficients and code delay as autoregressive (AR) processes with respect to time (uncorrelated with each other), and then employ an extended Kalman filter to adaptively estimate them. Their method takes advantage of the channel's time evolution model (assumed to be AR), but no method is provided to estimate its parameters. This limits its applicability to channels with fixed and a priori known structure. Hence, it cannot be used in channels whose statistics change slowly with time, as for example in mobile radio communications, where the territory slowly changes due to the receiver/transmitter motion. Finally, the assumption of uncorrelated tap coefficients further limits the applicability of the method, since it is not valid in several applications [14, 8].

Fitting a model to describe a system's time evolution, and estimating its parameters can be a

[2] Time-division multiple access.

[3] Independent, identically distributed.

challenging problem since the time-varying parameters are not directly observed. In the system identification literaure, only very simple i.i.d. or random walk models have been considered in order to simplify the problem [5]. Whenever more accurate models have been used, their parameters have been arbitrarily assigned [13]. It has been successfully argued that even a crude, approximate model of the time variations is better than no model at all [5]. However, a more accurate estimation of the model parameters can only be expected to improve the performance of these approaches.

In this paper we bridge this modelling gap by deriving novel higher-than second-order statistics (HOS) methods for estimating the model parameters. We address the estimation and equalization problem for general fading channels, not restricted to spread spectrum communications. The TV channel taps are modeled as general, multichannel AR, circular complex Gaussian processes, thus allowing Raleigh or Rician fading with correlated tap coefficients. Conditions for the identifiability of the model from input/output statistics are derived and consistency of the proposed method is shown. Based on the estimated model, a Kalman filtering procedure is proposed to track the TV tap coefficients. Finally, the TV tap estimates are used to decode the transmitted symbols, either in a Viterbi decoder (maximum likelihood input estimation), or using simpler decision-feedback schemes.

By combining a model-fitting procedure with Kalman-tracking algorithms, we introduce a novel viewpoint to the fading channel equalization problem. We broaden the applicability of Kalman based equalization methods which are expected to outperform traditional adaptive LMS and RLS techniques, since the latter do not exploit the tap's variation dynamics.

The rest of the paper is organized as follows. In Section 2, the problem statement is presented and some modeling aspects of fading channels are discussed. In Section 3, Kalman-based channel estimation methods are developed, while in Section 4, the necessary model parameters are obtained using higher-order statistics. The novel algorithm is discussed in detail in Section 5, and some illustrative simulations are presented in Section 6.

2. Modeling the multipath fading channel

Digital transmission through a fading channel is usually accomplished via the setup of Fig. 1. Discrete-time, complex symbols are transmitted every T_s seconds through a system consisting of the transmitter's pulse shaping filter, the TV multipath channel, and the receiver's matched filter and sampler. We assume symbol spaced sampling in this paper, i.e., every T_s seconds, although the extension to fractionally spaced receivers is an interesting future direction.

Let $h_c(t;\tau)$ be the continuous-time, TV impulse response (notice the explicit dependence on t) of the overall system, including the multipath channel and the transmitter/receiver filters. Then, the system of Fig. 1 is equivalent to Fig. 2. The received signal $y_c(t)$ is [4] (see Fig. 2)

$$y_c(t) = \sum_{k=0}^{\infty} w(k) h_c(t; t - kT_s) + v_c(t). \quad (1)$$

After the sampler, the discrete-time received signal is

$$y(n) \stackrel{\triangle}{=} y_c(t)\Big|_{t=nT_s} = \sum_{k=0}^{\infty} w(k) h_c(nT_s; nT_s - kT_s) + v_c(nT_s). \quad (2)$$

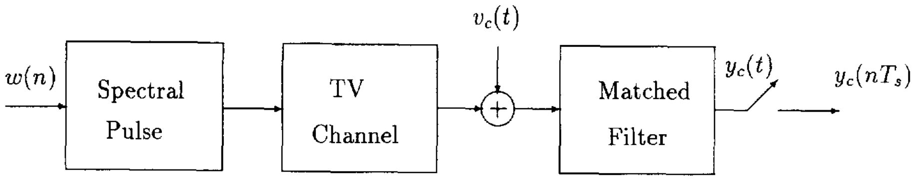

Fig. 1. Actual setup.

[4] We use the subscript c to denote continuous-time signals.

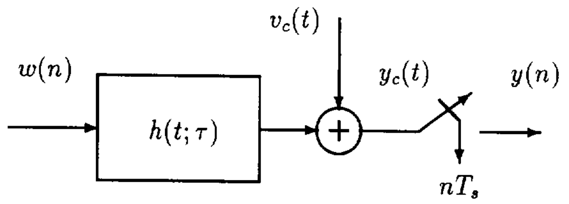

Fig. 2. Equivalent channel model.

If we define $v(n) \triangleq v_c(nT_s), h(n;k) \triangleq h_c(nT_s;kT_s)$, Eq. (2) becomes

$$y(n) = \sum_{k=0}^{\infty} w(k)h(n;n-k) + v(n) = \sum_{k=0}^{\infty} h(n;k)w(n-k) + v(n), \quad (3)$$

where the last equality is due to a change of variables.

The impulse response $h(n;k)$ geneally extends to infinity, but in communications applications it is common practice to truncate it at some order q, yielding the discrete-time model

$$y(n) = \sum_{k=0}^{q} h(n;k)w(n-k) + v(n). \quad (4)$$

In many applications the number q may be a rather small number. Signal propagation studies have shown that the delay spread rarely exceeds one symbol period in the North American Cellular Standard [23,8], while it can be up to five ($q = 5$) symbol periods for the European standard GSM [2,22].

Eq. (4) provides a generic input/output relationship, which describes the fading channel. However, unless the time variations of $h(n;k)$ are further specified, the model (4) is incomplete, and of rather limited use. Indeed, if $h(n;k)$ can arbitrarily vary with n, then there exist infinite different channel responses, which satisfy (4), even in the absence of noise and when $y(n)$ and $w(n)$ are available. The approximation of $h(n;k)$ by constants has proven to be unsatisfactory for modeling fading channels. Instead, the tap coefficients have been typically modeled as circular complex Gaussian random processes (w.r.t. n), by appealing to central limit theorem arguments. Processes with low-pass spectral characteristics are usually considered to model the relatively progressive time variations (slow fading).

A number of additional assumptions have been used in the past to simplify the model of Eq. (4). In general regression problems, $h(n;k)$'s have been considered to be i.i.d. sequences [21], an assumption not valid in communications applications. The *uncorrelated scattering* assumption has been extensively used, although it is not valid for several fading channels [14,17]. Under this assumption, the tap coefficients are uncorrelated with each other, i.e., $E\{h(n;k_1)h^*(n+\tau;k_2)\} = 0$ for $k_1 \neq k_2, \forall\tau$. Different channel paths indeed exhibit uncorrelated fading, but mixing those paths toegether in the receiver filters results in correlated $h(n;k)$'s in general. Under the uncorrelated scattering assumption, simple AR(1) models have been considered in [13], $h(n;k) = \alpha h(n-1;k) + u(n)$, where the parameter α is arbitrarily chosen close to unity, so that $h(n;k)$ represents a generic low-pass process for each k; $u(n)$ was an i.i.d. Gaussian sequence.

In this work we consider the random vector

$$\boldsymbol{h}(n) \triangleq [h(n;0),\ldots,h(n;q)]^{\mathrm{T}}, \quad (5)$$

as a general multichannel process and do not restrict ourselves to the uncorrelated scattering case. The vector $\boldsymbol{h}(n)$ may have zero or non-zero mean depending on whether the direct line-of-sight path is blocked or not (e.g., as in ionospheric links). It turns out that the mean of $\boldsymbol{h}(n)$ is relatively simple to estimate, as explained in Section 4. In order to simplify the notation, however, the main results are presented assuming the mean of $\boldsymbol{h}(n)$ to be zero, while the details of the non-zero mean case are delineated in Sections 4 and 6. In Section 4, we derive algorithms to estimate the spectral characteristics of $\boldsymbol{h}(n)$ from the data $y(n)$ and $w(n)$. In this way, the channel is adjusted to the fading environment, providing a more accurate description. We will use a parameteric spectral estimation approach, thereby modeling the vector process $\boldsymbol{h}(n)$ as a multichannel AR process of order p,

$$\boldsymbol{h}(n) = \sum_{l=1}^{p} \boldsymbol{A}(l)\boldsymbol{h}(n-l) + \boldsymbol{u}(n), \quad (6)$$

where $\boldsymbol{u}(n)$ is an i.i.d. circular complex Gaussian vector process with correlation matrix[5] $\boldsymbol{R}_{uu}(\tau) \triangleq E\{\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n+\tau)\} = \sigma_u^2 \boldsymbol{I}\delta(\tau)$. The choice of the order p represents a tradeoff between the accuracy of the model

[5] Superscript H denotes Hermitian transpose.

(6) and the increased variance of the estimated parameter matrices $A(l)$. For low-pass processes, however, even a low-order AR model may be sufficient [13]. In [25], a simplified version of (4) and (6) was studied with $q = 0$, in a different context.

The following assumptions are imposed on (4)–(6):

AS1: $v(n)$ and $u(n)$ are zero mean, i.i.d., circular complex Gaussian processes, uncorrelated with each other, with variance σ_v^2 and $\sigma_u^2 I$, respectively.

AS2: $w(n)$ is a zero mean, i.i.d. sequence with values drawn from a finite complex constellation with variance σ_w^2. The constellation is also assumed to be circularly symmetric, i.e., $E\{w^2(n)\} = 0$.

AS3: The transmitted signal satisfies $E\{|w(n)|^2\} \neq 0$ and $[E\{|w(n)|^2\}]^2 \neq E\{|w(n)|^4\}$.

Regarding the models (4) and (6), the goals of this work are:

1. To estimate the unknown coefficient matrices $\{A(l)\}_{l=1}^{p}$.
2. To derive minimum variance estimators of the tap coefficients $h(n;k)$, based on estimates of the matrices $A(l)$.
3. To ultimately decode $w(n)$, given the data $y(n)$ and the estimated channel $\hat{h}(n;k)$.

In order to motivate the use of the proposed model, let us concentrate on point 2 above, and derive channel tracking algorithms based on Eqs. (4) and (6). This will also motivate the estimation algorithms for the model parameters $A(l)$, developed in Section 4.

3. Tracking the channel coefficients

In this section, we are interested in deriving minimum variance estimators for the coefficients $h(n;k)$, under the model (4) and (6). For the time being the matrices $A(l)$ are considered known (see Section 4 for ways of estimating them).

The proposed method is based on Kalman filtering ideas and yields an adaptive algorithm, which can be implemented on-line. Like most adaptive algorithms, it requires a training period to adjust to the channel, after which one switches to a decision-directed mode. During the training period, the transmitted symbols are known to the receiver, while in the decision-directed mode, the previously decoded symbols $\hat{w}(n)$ are used in their place.

Let $w(n)$, $y(n)$ be given (training mode) in (4) and (6), and let the determinant $|A(z)|$ have zeros inside the unit circle, where $A(z) = I - \sum_{l=1}^{p} A(l)z^{-l}$. If we define $\mathbf{h}(n) \triangleq [h^{\mathrm{T}}(n),\ldots,h^{\mathrm{T}}(n-p+1)]^{\mathrm{T}}$, then Eq. (6) can be written in state space form

$$\mathbf{h}(n+1) = \begin{bmatrix} A(1) & A(2) & \cdots & A(p) \\ I & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I & 0 \end{bmatrix} \mathbf{h}(n) + \begin{bmatrix} u(n) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{7}$$

Let $\mathscr{A}$ be the block square matrix in the RHS of (7), then the state equation becomes

$$h(n+1) = \mathscr{A}h(n) + Ju(n), \tag{8}$$

where $J \triangleq [I, 0, \ldots, 0]$. In order to obtain the measurement equation, define $\mathbf{w}(n) \triangleq [w(n), w(n-1), \ldots, w(n-q)]^{\mathrm{T}}$ and $\mathbf{w}(n) \triangleq [\mathbf{w}^{\mathrm{T}}(n), \mathbf{0}^{\mathrm{T}}, \ldots, \mathbf{0}^{\mathrm{T}}]^{\mathrm{T}}$. Then Eq. (4) can be written as

$$y(n) = \mathbf{w}^{\mathrm{T}}(n)h(n) + v(n). \tag{9}$$

We point out that because $h(n)$ is time varying, it cannot be identified using RLS on Eq. (9) alone.

Eqs. (8) and (9) offer a state-space representation of the fading channel model with transition matrix $\mathscr{A}$ (assumed known in this section). Based on this representation, the minimum variance estimator for the state vector, i.e., the conditional expectation of $\mathbf{h}(n)$, given $\{w(k), y(k)\}_{k=0}^{n-1}$, can be computed using the Kalman filter [1]. The recursions are summarized in Table 1 (see also [1, p. 44]). Matrices $K(n)$ and $\Sigma(n|n)$ denote the Kalman gain and the covariance of the state vector $\mathbf{h}(n)$ given data $\{w(i), y(i)\}_{i=0}^{n}$, respectively.

Extended Kalman filtering ideas were employed in [9] in a spread-spectrum environment. The state vector there consisted of the code-delay, the Doppler spread and the tap coefficients (assumed to be uncorrelated). However, no method for estimating the state transition matrix $\mathscr{A}$ was given. In this paper, we apply state-space approaches to general fading channel environments allowing for correlated coefficients, and as explained in the next section, we couple the Kalman filter with algorithms which estimate $\mathscr{A}$.

Table 1
Kalman-based, channel-tracking algorithm

Initialization $\hat{\boldsymbol{h}}(0\|-1)=\mathbf{0}$, $\Sigma(0\|-1)=\mathbf{0}$	
Recursion:	$\hat{\boldsymbol{h}}(n+1\|n)=[\mathscr{A}-\boldsymbol{K}(n)\mathbf{w}^{\mathrm{T}}(n)]\hat{\mathbf{h}}(n\|n-1)+\boldsymbol{K}(n)y(n)$
	$\boldsymbol{K}(n)=\mathscr{A}\Sigma(n\|n-1)\mathbf{w}^*(n)\times[\mathbf{w}^{\mathrm{T}}(n)\Sigma(n\|n-1)\mathbf{w}^*(n)+\sigma_v^2]^{-1}$
	$\Sigma(n+1\|n)=\mathscr{A}[\Sigma(n\|n-1)-\Sigma(n\|n-1)\mathbf{w}^*(n)[\mathbf{w}^{\mathrm{T}}(n)\Sigma(n\|n-1)\mathbf{w}^*(n)+\sigma_v^2]^{-1}\times\mathbf{w}^{\mathrm{T}}(n)\Sigma(n\|n-1)]\mathscr{A}^{\mathrm{H}}+\sigma_u^2\boldsymbol{J}\boldsymbol{J}^{\mathrm{H}}$

After the training period, the Kalman filter of Table 1 should also be combined with a decision feedback or Viterbi equalizer to provide the decoded symbols $\hat{w}(n)$, needed in the Kalman recursion. These issues are discussed in Section 5.

4. Determining the channel's statistics

In this section we employ higher-order statistics to estimate the state transition matrix $\mathscr{A}$ in (8), or equivalently the parameter matrices $\boldsymbol{A}(l), l=1,\ldots,p$, in (6), given the data $y(n)$ and $w(n)$. It is well known in multichannel time-series analysis that the AR parameters are uniquely identified by the correlation matrices $\boldsymbol{R}_{hh}(\tau)=E\{\boldsymbol{h}(n)\boldsymbol{h}^{\mathrm{H}}(n+\tau)\}$. Indeed, by postmultiplying (6) by $\boldsymbol{h}^{\mathrm{H}}(n-\tau)$ and taking expected values of both sides, we arrive at the multichannel normal equations [19, Chapter 9]

$$\boldsymbol{R}_{hh}^{\mathrm{H}}(\tau)=\sum_{l=1}^{p}\boldsymbol{A}(l)\boldsymbol{R}_{hh}^{\mathrm{H}}(\tau-l)+\sigma_u^2\delta(\tau)\boldsymbol{I},\quad \tau=0,\ldots,p. \tag{10}$$

By solving those equations, the parameter matrices $\boldsymbol{A}(l)$ are obtained, provided that some estimates of the correlation matrices $\boldsymbol{R}_{hh}(\tau)$ are available. However, $\boldsymbol{h}(n)$ is part of the state vector $\mathbf{h}(n)$, and is not directly measured. Hence, the estimation of its correlation matrix is not a trivial task. In the sequel we employ higher-order statistics to estimate $\boldsymbol{R}_{hh}(\tau)$. In particular, we seek estimates of the (i,j) element of the matrix (cf. (5)) for $\tau=0,\ldots,p$,

$$[\boldsymbol{R}_{hh}(\tau)]_{i,j}=r_h(\tau;i-1,j-1),\quad i,j=1,\ldots,q+1, \tag{11}$$

$$r_h(\tau;i,j)\triangleq E\{h(n;i)h^*(n+\tau;j)\}. \tag{12}$$

4.1. Least-squares methods

Let us consider the conditional expectation of $y(n)y^*(n+\tau)$, given the data $\{w(l),y(l)\}_{l=n-q}^{n}$. Then from (4) and the independence between $h(n;k)$ and $w(n)$ we obtain

$$\begin{aligned}E\{y(n)y^*(n+\tau)|\boldsymbol{w}(n)\}&=\sum_{k_0,k_1=0}^{q}E\{h(n;k_0)h^*(n+\tau;k_1)\}\\&\qquad\times w(n-k_0)w^*(n+\tau-k_1)+\sigma_v^2\delta(\tau)\\&=\sum_{k_0,k_1=0}^{q}r_h(\tau;k_0,k_1)w(n-k_0)w^*(n+\tau-k_1)\\&\quad+\sigma_v^2\delta(\tau).\end{aligned} \tag{13}$$

We wish to solve for the parameters $r_h(\tau;k_0,k_1)$ from (13). Note that the conditional expectation is a time-varying statistic which depends on n. If (13) is written for a fixed τ and $n=0,\ldots,N-1$, then it provides a set of linear equations involving $r_h(\tau;k_0,k_1)$.

Let us consider the general linear expression

$$E\{y(n)y^*(n+\tau)|\boldsymbol{w}(n)\}=\sum_{k_0,k_1=0}^{q}\theta_\tau^{(k_0,k_1)}w(n-k_0)w^*(n+\tau-k_1)+\sigma_v^2\delta(\tau), \tag{14}$$

for some parameters $\theta_\tau^{(k_0,k_1)}$, or in vector form

$$E\{y(n)y^*(n+\tau)|\boldsymbol{w}(n)\}=\phi^{\mathrm{H}}(n;\tau)\boldsymbol{\theta}_\tau+\sigma_v^2\delta(\tau)+e(n;\tau), \tag{15}$$

where the parameter vector

$$\boldsymbol{\theta}_\tau\triangleq[\theta_\tau^{(0,0)},\ldots,\theta_\tau^{(0,q)},\theta_\tau^{(1,0)},\ldots,\theta_\tau^{(q,q)}]^{\mathrm{T}} \tag{16}$$

is indexed according to $\{(k_1,k_2)_{k_1=0}^{q}\}_{k_2=0}^{q}$ and $\phi^{\rm H}(n;\tau)$ is a row vector defined accordingly as

$$\phi^{\rm H}(n;\tau) \triangleq [w(n)w^*(n+\tau),\ldots,w(n)w^*(n+\tau-q), \\ w(n-1)w^*(n+\tau),\ldots, \\ w(n-q)w^*(n+\tau-q)]. \tag{17}$$

Notice that (13) represents a particular solution of (15) with

$$\boldsymbol{\theta}_\tau = \boldsymbol{\theta}_{\tau,\rm true}, \\ \boldsymbol{\theta}_{\tau,\rm true} \triangleq [r_h(\tau;0,0),\ldots,r_h(\tau;0,q), \\ r_h(\tau;1,0),\ldots,r_h(\tau;q,q)]. \tag{18}$$

We intend to solve (15) via LS using an approximation of the LHS, in an effort to recover $\boldsymbol{\theta}_{\tau,\rm true}$. We propose to use the instantaneous approximation $y(n)y^*(n+\tau)$, as an estimate of the random variable $E\{y(n)y^*(n+\tau)|\boldsymbol{w}(n)\}$. It can easily be checked that this estimate is unbiased, i.e., $E\{e(n;\tau)\}=0$, where $e(n;\tau)=y(n)y^*(n+\tau)-E\{y(n)y^*(n+\tau)|\boldsymbol{w}(n)\}$ is the estimation error. Also the error is uncorrelated with the regressors $w(n-k_0)w^*(n+\tau-k_1)$; i.e., $E\{e(n;\tau)w(n-k_0)w^*(n+\tau-k_1)\}=0$. Hence, the LS solution of the linear regression

$$y(n)y^*(n+\tau) = \phi^{\rm H}(n;\tau)\boldsymbol{\theta}_\tau + \sigma_v^2\delta(\tau) + e(n;\tau), \tag{19}$$

for $n=0,\ldots,N-1$, will yield an unbiased estimate of $\boldsymbol{\theta}_{\tau,\rm true}$ under some further assumptions. The following proposition quantifies this claim and shows strong consistency of the proposed method under some conditions.

Proposition 1. *Let $y(n), w(n)$ be related by* (4) *and* (6) *under assumptions* (AS1)–(AS3). *If $\hat{\boldsymbol{\theta}}_{\tau,N}$ denotes the solution of* (19) *for $n=0,\ldots,N-1$, then under these assumptions, $\hat{\boldsymbol{\theta}}_{\tau,N} \to \boldsymbol{\theta}_\tau^{(0)}$ as $N\to\infty$ in the mean square sense and with probability one, where $\boldsymbol{\theta}_\tau^{(0)}$ is defined in* (16).

Proof. See Appendix A.

Some remarks are now in order:

1. The fact that an instantaneous approximation is used in (19) may raise some questions on the variance of the resulting estimator and the accuracy of the proposed method. It should be kept in mind, however, that the approximate relationship in (19) holds for every time point $n=0,\ldots,N-1$. Thus, N such equations will be used in this regression problem and the approximation errors will be averaged out in the LS solution.
2. It turns out it is not hard to show that the solution of (19) converges to the solution of (15) for large N, i.e., the approximation involved in (19) is immaterial as far as consistency is concerned. The crucial question, however, on which the success of the method depends is whether (15) has a unique solution, in which case $\boldsymbol{\theta}_{\tau,\rm true}$ is identifiable. An equivalent question is whether the regressors $w(n-k_0)w^*(n+\tau-k_1)$ are linearly independent, or become 'linearly independent' in some sense as N grows large. As Proposition 1 shows, this can be guaranteed under certain assumptions.
3. It appears as if only second-order information is present in the LS problem of (19). A more careful consideration, however, reveals that in the process of solving the LS problem, fourth-order auto- and cross-moments are implicitly computed (see Appendix A). Proposition 1 shows that, under some conditions, the statistics of the TV taps can be recovered from these moments. A short discussion on those assumptions follows.

Classical analysis of digital transmission through a fading medium models $h(n;k)$ as zero mean random variables as in (AS1). In certain applications, however, like cellular communications, a direct non-fading path may also exist, superimposed on the fading path. In this case, the coefficients $h(n;k)$ have non-zero means (Rician fading). However, we may treat the $h(n;k)$'s as zero mean without loss of generality in the current framework. Indeed, if the overall channel is

$$\bar{h}(n;k) = \bar{h}_m(k) + h(n;k), \tag{20}$$

where $\bar{h}_m(k)$ is a constant mean (in general $\neq 0$) and $E\{h(n;k)\}=0$, then by substituting (20) into (4), we obtain

$$\bar{y}(n) = \sum_{k=0}^{q} \bar{h}_m(k)w(n-k) + y(n), \tag{21}$$

$$y(n) = \sum_{k=0}^{q} h(n;k)w(n-k) + v(n). \tag{22}$$

The constants $\bar{h}_m(k)$ can be consistently estimated from the LS solution of Eq. (21) and $y(n)$ can be recovered. Hence, Eq. (22) becomes identical to the modeling Eq. (4), once $y(n)$ has been recovered from $\bar{y}(n)$.

The assumption on the circular symmetry of the constellation (AS2) is also not crucial for our developments. It is included for technical convenience as most constellations fulfill it. On the contrary, (AS3) is necessary to obtain consistent estimates. It turns out (see Appendix A) that correlations of the form $r_h(\tau; k, k+\tau)$, $k = 0, \ldots, q$, cannot be estimated if (AS3) does not hold. The corresponding regressors in (19) become dependent, and only the quantity $\sum_{k=0}^{q} r_h(\tau; k, k+\tau)$ can be obtained.

(AS3) holds true for most constellations, with the exception of certain constant modulus schemes such as 4-PSK and 8-PSK. In such cases, the easiest way to obviate the problem is to estimate $r_h(\tau; k_0, k_1)$ only for $\tau > q$, where the troublesome regressors do not appear. Then, the normal equations (10) can be solved for $\tau = q+1, \ldots, q+p+1$, mimicking the approach used to estimate AR parts of ARMA models. A different approach is to estimate directly the higher-order moments of the data and obtain closed-form solutions, as explained in the sequel.

4.2. Closed-form solutions

Cumulants of order k are defined as combinations of moments up to order k [6, p. 19]. In this paper, only third- and fourth-order cumulants will be used, which for zero-mean random variables are defined as

$$\mathrm{cum}\{x_1, x_2, x_3\} = E\{x_1 x_2 x_3\}, \tag{23}$$

$$\mathrm{cum}\{x_1, x_2, x_3, x_4\} = E\{x_1 x_2 x_3 x_4\} - E\{x_1 x_2\}E\{x_3 x_4\} - E\{x_1 x_2\}E\{x_2 x_4\} - E\{x_1 x_4\}E\{x_2 x_3\}. \tag{24}$$

Second-order cumulants coincide with the covariance, i.e.,

$$\mathrm{cum}\{x_1, x_2\} = E\{x_1 x_2\} - E\{x_1\}E\{x_2\}. \tag{25}$$

The simplest method for estimating the lags $r_h(\tau; k, k+\tau)$, which may not be recoverable through the LS solution, employs third-order cumulants. Consider the quantity $E\{y(n)y^*(n+\tau_1)w(n+\tau_2)\}$. Substituting $y(n)y^*(n+\tau_1)$ using (4) we obtain

$$E\{y(n)y^*(n+\tau_1)w(n+\tau_2)\} = \sum_{k_0, k_1 = 0}^{q} E\{h(n; k_0)h^*(n+\tau_1; k_1)\} \times E\{w(n-k_0)w^*(n+\tau_1-k_1)w(n+\tau_2)\}, \tag{26}$$

due to the mutual independence among $h(n; k), w(n)$ and $v(n)$. Under the i.i.d. assumption for $w(n)$ (26) reduces to

$$E\{y(n)y^*(n+\tau_1)w(n+\tau_2)\} = \gamma_{3w} E\{h(n; -\tau_2)h^*(n+\tau_1; \tau_1-\tau_2)\}, \tag{27}$$

where $\gamma_{3w} \triangleq E\{w^2(n)w^*(n)\}$. Eq. (27) shows that the desired tap correlations coincide, within a scaling factor, with the triple correlations in the LHS. For $\tau_2 = -k_0$, $\tau_1 = \tau$, (27) becomes

$$E\{y(n)y^*(n+\tau)w(n-k_0)\} = \gamma_{3w} r_h(\tau; k_0, \tau+k_0), \tag{28}$$

providing a closed-form solution for $r_h(\tau; k_0, \tau+k_0)$. Hence the simple estimator

$$\hat{r}_h(\tau; k_0, \tau+k_0) = \frac{1}{\gamma_{3w}} \left[\frac{1}{N} \sum_{n=0}^{N-1} y(n)y^*(n+\tau)w(n-k_0) \right] \tag{29}$$

can be used, provided that $\gamma_{3w} \neq 0$. This approach is not applicable to non-skewed (symmetrically distributed) input signals which have $\gamma_{3w} = 0$. This may happen whenever $w(n)$ is drawn from an equiprobable, symmetric constellation. Hence, caution should be used during the training period, in selecting a training pattern that is non-equiprobable, and therefore skewed.

When $\gamma_{3w} = 0$, closed-form expressions for the desired tap correlations can be derived using fourth-order statistics. If we introduce product processes $y_2(n; \tau_1) \triangleq y(n)y^*(n+\tau_1)$, and $w_2(n; \tau_2, \tau_3) \triangleq w(n+\tau_2)w^*(n+\tau_3)$, then the cross-cumulant $c_{y_2 w_2}(\tau_1, \tau_2, \tau_3)$,

$$c_{y_2 w_2}(\tau_1, \tau_2, \tau_3) \triangleq \mathrm{cum}\{y_2(n; \tau_1), w_2(n; \tau_2, \tau_3)\}, \tag{30}$$

can be expressed as

$$c_{y_2w_2}(\tau_1,\tau_2,\tau_3) = \sum_{k_0,k_1=0}^{q} \mathrm{cum}\{h(n;k_0)h^*(n+\tau_1;k_1)w(n-k_0) \times w^*(n+\tau_1-k_1), w(n+\tau_2)w^*(n+\tau_3)\} \quad (31)$$

using (4) and the multilinearity property of cumulants [6, p. 19]. Substituting the cumulant in the RHS of (31) from (25) we obtain

$$c_{y_2w_2}(\tau_1,\tau_2,\tau_3) = \sum_{k_0,k_1=0}^{q} E\{h(n;k_0)h^*(n+\tau_1;k_1)\} \times[E\{w(n-k_0)w^*(n+\tau_1-k_1)w(n+\tau_2) \times w^*(n+\tau_3)\} - E\{w(n-k_0)w^*(n+\tau_1-k_1)\} \times E\{w(n+\tau_2)w^*(n+\tau_3)\}]. \quad (32)$$

Substituting the fourth-order moment in the RHS from (24) and using the fact that $E\{w^2(n)\}=0$ we obtain

$$c_{y_2w_2}(\tau_1,\tau_2,\tau_3) = \sum_{k_0,k_1=0}^{q} r_h(\tau_1;k_0,k_1)[\mathrm{cum}\{w(n-k_0), w^*(n+\tau_1-k_1), w(n+\tau_2), w^*(n+\tau_3)\} + E\{w(n-k_0)w^*(n+\tau_3)\}E\{w(n+\tau_2) \times w^*(n+\tau_1-k_1)\}]. \quad (33)$$

If $w(n)$ is i.i.d., then (33) reduces to

$$c_{y_2w_2}(\tau_1,\tau_2,\tau_3) = \gamma_{4w} r_h(\tau_1;-\tau_2,\tau_1-\tau_2)\delta(\tau_2-\tau_3) + \sigma_w^4 r_h(\tau_1;-\tau_3,\tau_1-\tau_2), \quad (34)$$

where $\gamma_{4w} = \mathrm{cum}\{w(n),w(n),w^*(n),w^*(n)\}$. Hence, for $\tau_1=\tau$, $\tau_3=-k_0$, $\tau_2=\tau-k_1$, we obtain the closed-form expression

$$c_{y_2w_2}(\tau_1,\tau_2,\tau_3) = \begin{cases} \sigma_w^4 r_h(\tau;k_0,k_1), & k_1 \neq k_0+\tau, \\ (\gamma_{4w}+\sigma_w^4) r_h(\tau;k_0,k_1), & k_1 = k_0+\tau. \end{cases} \quad (35)$$

From (35), we obtain closed-form solution for the estimation of $r_h(\tau;k_0,k_1)$ using fourth-order statistics as follows:

$$r_h(\tau;k_0,k_1) = \begin{cases} \sigma_w^{-4} c_{y_2w_2}(\tau;\tau-k_1,-k_0), & k_1 \neq k_0+\tau, \\ (\gamma_{4w}+\sigma_w^4)^{-1} c_{y_2w_2}(\tau;-k_0,-k_0), & k_1 = k_0+\tau. \end{cases} \quad (36)$$

5. Equalization algorithm – discussion

In previous sections, a Kalman filtering procedure was developed to track the TV tap coefficients, based on some modeling parameters. Methods to estimate those parameters were also given. All those pieces have to be linked together now and be coupled with a specific equalization technique, in order to eventually compose a practical system. In this section, we discuss the implementation details of a complete equalization algorithm based on the novel approach.

We assume the typical scenario of periodic retraining, found in mobile wireless communications (usually in the order of 10% of the transmission time [23]). Hence, at each training period, the estimates of the model parameters from the previous ones have to be improved. In order to avoid instabilities, we do not update the model parameters during the operating mode. In the operating mode, only the Kalman filter is run, in parallel with a decision feed-back equalizer to jointly track the channel and equalize the received symbols. The equalizer used has orders L_1 and L_2 for the feedforward and feedback branch, respectively (DF(L_1,L_2)), and is of the form

$$\hat{w}(n) = \sum_{j=-L_1}^{0} g(n;j)y(n-j) + \sum_{j=1}^{L_2} g(n;j)\hat{w}(n-j). \quad (37)$$

Once the channel $h(n;k)$ has been estimated, the optimal parameters $g(n;k)$ of the feedforward part, which minimize $J(L_1,L_2) = E\{|w(n) - \hat{w}(n)|^2|\boldsymbol{h}(n)\}$ are given by the solution of the following equations (see, e.g., [20, p. 595]): [6]

$$\sum_{j=-L_1}^{0} \psi_{lj}\, g(n;j) = h^*(n;-l), \quad l = -L_1,\dots,0, \tag{38}$$

where

$$\psi_{lj} = \sigma_w^2 \sum_{k=0}^{-l} h^*(n;k)h(n;k+l-j) + \sigma_v^2\delta(l-j), \quad l,j = -L_1,\dots,0. \tag{39}$$

The coefficients of the feedback filter are given in terms of the coefficients computed in (38) as

$$g(n;k) = -\sum_{j=-L_1}^{0} g(n;j)h(n;k-j), \quad k = 1,\dots,L_2. \tag{40}$$

The steps of the algorithm are explained in detail next.

Training mode. Assume that some estimates of the tap correlations $\hat{\boldsymbol{\theta}}_\tau$, the AR matrices $\hat{\boldsymbol{A}}(l)$ and the channel coefficients $\hat{\boldsymbol{h}}(n)$ are available from past iterations. The first time, initialize to zero.

Step 1. Collect the training data $y(n)$ and $w(n)$, $n = 0,\dots,N_{\text{train}} - 1$.

Step 2. Form Eq. (19) and update the current $\hat{\boldsymbol{\theta}}_\tau$ using RLS.

Step 3. Solve the normal equations (10) to obtain a new estimate $\hat{\boldsymbol{A}}(l), l = 1,\dots,p$. In order to save computations, an adaptive gradient method to solve (10) and update $\hat{\boldsymbol{A}}(l)$ can be used.

Step 4. Run the Kalman filter (Table 1) to estimate $\hat{\boldsymbol{h}}(n)$ based on the updated $\hat{\boldsymbol{A}}(l)$.

[6] In order to simplify Eqs. (38)–(40), we locally approximate the channel by a TI one, i.e. $h(n+\tau;k) \approx h(n;k)$ for $\tau \in [n-L_1, n+L_2]$.

Operating mode. Assume $\hat{\boldsymbol{\theta}}_\tau$, $\hat{\boldsymbol{A}}(l)$, $\hat{\boldsymbol{h}}(n)$ have been initialized during the training period. Then, for every new data point $y(n)$:

Step 1. Solve Eqs. (38) and (39) to obtain the equalizer coefficients $g(n;k)$. Adaptive gradient methods may be used to solve the equations, to reduce computations.

Step 2. Equalize the data using (37).

Step 3. Decode the equalized $\hat{w}(n)$ and update $\hat{\boldsymbol{h}}(n)$ using the Kalman recursion.

A number of different variations on this algorithm may be used. For example, a Viterbi decoder may replace the decision feedback one. Alternatively, the parameter matrices may be updated during the operational mode as well based on the decoded symbols. Finally, the closed form solutions of Section 4.2 may be used as an alternative to estimate $\hat{\boldsymbol{\theta}}_\tau$. The complete study of all the variations however, is outside the scope of this paper. In the next section, some representative simulations are given in order to compare the proposed method with existing techniques.

6. Simulations

The purpose of this section is to demonstrate our proposed algorithm and compare it with existing algorithms.

6.1. Description of the problem

We consider a random channel $\bar{\boldsymbol{h}}(n) \triangleq [h(n;0), h(n;1)]^{\mathrm{T}}$ (of order $q = 1$), whose mean is given by $\bar{\boldsymbol{h}}_m = [1 + 0.2\mathrm{j}, -0.5 + 0.5\mathrm{j}]^{\mathrm{T}}$. The equivalent mean removed channel, $\boldsymbol{h}(n) = \bar{\boldsymbol{h}}(n) - \bar{\boldsymbol{h}}_m$, is modeled as a multi-channel AR(1) process ($p = 1$),

$$\boldsymbol{h}(n) \triangleq \begin{bmatrix} h(n;0) \\ h(n;1) \end{bmatrix} = \boldsymbol{A}(1)\boldsymbol{h}(n-1) + \boldsymbol{u}(n), \tag{41}$$

where $\boldsymbol{u}(n)$ is a zero-mean i.i.d. circular complex Gaussian process with correlation matrix $\boldsymbol{R}_{uu}(\tau) = \sigma_u^2\delta(\tau)\boldsymbol{I}$, and

$$\boldsymbol{A}(1) = \begin{bmatrix} 0.3 & -0.8 \\ -0.5 & 0.3 \end{bmatrix}. \tag{42}$$

We first generate $\boldsymbol{h}(n)$, obtain $\bar{\boldsymbol{h}}(n) = \boldsymbol{h}(n) + \bar{\boldsymbol{h}}_{\mathrm{m}}$, and then simulate the noisy channel output as

$$\bar{y}(n) = \bar{h}(n;0)w(n) + \bar{h}(n;1)w(n-1) + v(n), \quad (43)$$

where the input symbol $w(n)$ is drawn from a 16-QAM constellation, and the additive noise $v(n)$ is a zero-mean i.i.d. circular complex Gaussian process with variance σ_v^2.

We shall illustrate and evaluate three different algorithms: our proposed method, the Kalman-based method of [13], and the plain RLS algorithm. For all three methods, a periodic re-training period of 16 samples is used, alternating with an operating period of 128 samples. In all cases the receiver starts decoding the data after an initial training of 1024 points, or an equivalent of 64 idle frames.

6.2. Description of the algorithms

Let us explain first what happens during the initial training mode. The objective here is to obtain estimates of the $\boldsymbol{A}(1)$ matrix, the mean of the channel $\bar{\boldsymbol{h}}_{\mathrm{m}}$, and the channel $\bar{\boldsymbol{h}}(n)$ itself, from the input $w(n)$ and the output $\bar{y}(n)$. Our first step is to solve for $\bar{\boldsymbol{h}}_{\mathrm{m}} \triangleq [\bar{h}_{\mathrm{m}}(0), \bar{h}_{\mathrm{m}}(1)]^{\mathrm{T}}$ from

$$\bar{y}(n) = \bar{h}_{\mathrm{m}}(0)w(n) + \bar{h}_{\mathrm{m}}(1)w(n-1) + y(n), \quad (44)$$

via least squares. Recursive estimates of $\bar{\boldsymbol{h}}_{\mathrm{m}}$ can be obtained by employing RLS, after which we can form the zero-mean equivalent output process $y(n)$ of (44). Note that the KF procedure of Table 1 is based on the zero-mean processes $y(n)$ and $\boldsymbol{h}(n)$.

Next, we wish to solve for the $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ vectors (cf. (18)) which contain correlation information of the zero-mean channel $\boldsymbol{h}(n)$. For our problem here ($p = q = 1$), we have that

$$\boldsymbol{\theta}_0 \triangleq \begin{bmatrix} r_h(0;0,0) \\ r_h(0;0,1) \\ r_h(0;1,0) \\ r_h(0;1,1) \end{bmatrix}, \qquad \boldsymbol{\theta}_1 \triangleq \begin{bmatrix} r_h(1;0,0) \\ r_h(1;0,1) \\ r_h(1;1,0) \\ r_h(1;1,1) \end{bmatrix}. \quad (45)$$

Using (19) for $\tau = 0$ and $\tau = 1$ we obtain

$$|y(n)|^2 = [|w(n)|^2, w(n)w^*(n-1), w(n-1)w^*(n), |w(n-1)|^2]\boldsymbol{\theta}_0 + \sigma_v^2 + e(n;0), \quad (46)$$

and

$$y(n)y^*(n+1) = [w(n)w^*(n+1), |w(n)|^2, w(n-1)w^*(n+1), w(n-1)w^*(n)]\boldsymbol{\theta}_1 + e(n;1), \quad (47)$$

which can be solved recursively for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ via RLS. Since

$$\boldsymbol{R}_{hh}(0) \triangleq \begin{bmatrix} r_h(0;0,0) & r_h(0;0,1) \\ r_h(0;1,0) & r_h(0;1,1) \end{bmatrix}, \quad \boldsymbol{R}_{hh}(1) \triangleq \begin{bmatrix} r_h(1;0,0) & r_h(1;0,1) \\ r_h(1;1,0) & r_h(1;1,1) \end{bmatrix}, \quad (48)$$

they are available once $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are found, and $\boldsymbol{A}(1)$ can be solved according to (cf. (10))

$$\boldsymbol{R}_{hh}^{\mathrm{H}}(1) = \boldsymbol{A}(1)\boldsymbol{R}_{hh}^{\mathrm{H}}(0). \quad (49)$$

After $\boldsymbol{A}(1)$ is estimated, we run the KF following the steps of Table 1 to find $\boldsymbol{h}(n)$. Note that in our case here, $\mathscr{A} = \boldsymbol{A}(1), \boldsymbol{h}(n) = \boldsymbol{h}(n)$, and $\boldsymbol{J} = \boldsymbol{I}$. Once $\boldsymbol{h}(n)$ is found, we then compute $\bar{\boldsymbol{h}}(n)$ by incorporating the recursive estimate of $\bar{\boldsymbol{h}}_{\mathrm{m}}$.

The estimates $\boldsymbol{A}(1)$, $\bar{\boldsymbol{h}}_{\mathrm{m}}$ and $\bar{\boldsymbol{h}}(n)$ are necessary quantities for the operating mode. In the operating mode, we first use the DFE to obtain preliminary symbol estimates $\tilde{w}(n)$. These estimates are subsequently transformed into $\hat{w}(n)$ using the nearest-neighbor criterion, where $\hat{w}(n)$ are valid 16-QAM symbols. For our particular problem here ($q = 1$), the DFE employs the current channel output $\bar{y}(n)$ and the previously decoded symbol $\hat{w}(n-1)$ to infer

$$\tilde{w}(n) = g(n;-1)\bar{y}(n+1) + g(n;0)\bar{y}(n) + g(n;1)\hat{w}(n-1). \quad (50)$$

The feedforward coefficients can be computed as follows [20, p. 595]:

$$\begin{bmatrix} g(n;-1) \\ g(n;0) \end{bmatrix} = \begin{bmatrix} \bar{h}^*(n;0)\bar{h}(n;1) & \sigma_v^2/\sigma_w^2 + |\bar{h}(n;0)|^2 \\ \sigma_v^2/\sigma_w^2 + |\bar{h}(n;0)|^2 + |\bar{h}(n;1)|^2 & \bar{h}^*(n;1)\bar{h}(n;0) \end{bmatrix}^{-1} \begin{bmatrix} \bar{h}^*(n;0) \bar{h}^*(n;1) \end{bmatrix}, \quad (51)$$

and the feedback coefficient is found from

$$g(n;1) = -g(n;0)\bar{h}(n;1). \tag{52}$$

Note that in the operating mode, $\hat{w}(n)$ replaces $w(n)$ in (44) in order to obtain $y(n)$ and run the KF to update $\boldsymbol{h}(n)$. Quantities $\boldsymbol{A}(1)$ and $\bar{\boldsymbol{h}}_{\mathrm{m}}$ are kept the same as at the end of the training mode.

It is well known that 'run-away' effect may incur to the DFE. A brief retraining mode is hence introduced after each operating mode in order to cope with a possible run-away sequence, and to update the estimates of $\boldsymbol{A}(1)$ and $\bar{\boldsymbol{h}}_{\mathrm{m}}$, in case the latter two are slowly varying as well.

In the Kalman procedure of [13], the matrix $\hat{\boldsymbol{A}}(1)$ is a priori set to

$$\hat{\boldsymbol{A}}(1) = \begin{bmatrix} 0.99 & 0 \\ 0 & 0.99 \end{bmatrix}. \tag{53}$$

In the training mode, one only estimates $\bar{\boldsymbol{h}}_{\mathrm{m}}$ using RLS and updates $\boldsymbol{h}(n)$ using KF, whereas in the operating mode, $\boldsymbol{h}(n)$ (and hence $\bar{\boldsymbol{h}}(n)$) is updated using KF based on the a priori chosen value of $\boldsymbol{A}(1)$ and the estimate of $\bar{\boldsymbol{h}}_{\mathrm{m}}$. The DFE is then implemented to recover the symbols $w(n)$.

In the plain RLS method, one first estimates a constant vector $\boldsymbol{g} \triangleq [g(-1), g(0), g(1)]'$ during the training mode, by minimizing $E\{|w(n) - \hat{w}(n)|^2\}$, where

$$w(n) = g(-1)\bar{y}(n+1) + g(0)\bar{y}(n) + g(1)w(n-1), \tag{54}$$

via RLS and based on the available data $w(n)$ and $\bar{y}(n)$. In the operating mode, the preliminary symbol estimate

$$\tilde{w}(n) = g(-1)\bar{y}(n+1) + g(0)\bar{y}(n) + g(1)\hat{w}(n-1) \tag{55}$$

is first obtained, and then translated into $\hat{w}(n)$ using the nearest-neighbor criterion, $\hat{w}(n)$ being a valid 16-QAM symbol.

6.3. Simulation results

In our simulations, the average signal-to-noise ratio (SNR) is defined as

$$\mathrm{SNR} \triangleq 10 \log_{10} \frac{E\{|\bar{y}(n) - v(n)|^2\}}{E\{|v(n)|^2\}}. \tag{56}$$

For our first example, we used $\sigma_u^2 = 0.002$ and $\sigma_v^2 = 0.01$, which yields SNR = 25.75 dB. In Figs. 3(a)–(d) we show elementwise the recursive estimate of the real part, and in Figs. 3(e)–(h) the imaginary part, of $\boldsymbol{A}(1)$ in dotted lines. The true values are indicated by the dashed line. The figure shows that after the first 200 samples, the estimates converge close to the true system's parameters.

Following the initial training period, there was an operating period, and Figs. 4(a)–(d) and Figs. 4(e)–(h) show, respectively, the tracking of the same channel taps using our method and that of [13]. The latter seems to have relatively worse tracking performance due to the inaccurate guess of the $\boldsymbol{A}(1)$ matrix, but the difference was not significant for this particular operating period. We repeated the retraining-operating modes for nine additional times, and Fig. 5 shows a run-away phenomenon occurring for the method of [13]. The final symbol error rate in $\hat{w}(n)$ was 0 for our method, 0.0236 for that of [13], and 0.0173 for the plain RLS.

Next, we wish to illustrate and compare the performance of the algorithms in terms of their abilities to equalize the symbols. In Fig. 6(a) we show the constellation of 128 received symbols $\bar{y}(n)$ with $\sigma_u^2 = 0.005$ and $\sigma_v^2 = 1 \times 10^{-5}$. Figs. 6(b)–(d) show the equalized symbols using the proposed, that of [13], and the plain RLS method, respectively. The proposed method is seen to have the best equalizing ability.

Finally, we wish to evaluate the performance of the algorithms by plotting the respective error probabilities as a function of the SNR. We chose $\sigma_u^2 = 0.004$, and varied σ_v^2 so that the SNR was changing. Fig. 7 shows the error rate in $\hat{w}(n)$ for the proposed method (solid line), the method of [13] (dashed line), and the plain RLS method (dotted line). Fifty independent realizations were carried out at each SNR, from which the error rate was calculated based on five operating periods per realization. The performance of the proposed method surpassed that of [13] and plain RLSs at about 18 and 21dB, respectively, and is much superior at high SNRs. This figure also shows that the severe and time-varying ISI is the major impeding factor in this channel rather than the effects of the additive noise.

Some remarks and observations are now in order:

1. When σ_u^2 is smaller, i.e., the channel has smaller variations, the plain RLS method gains more

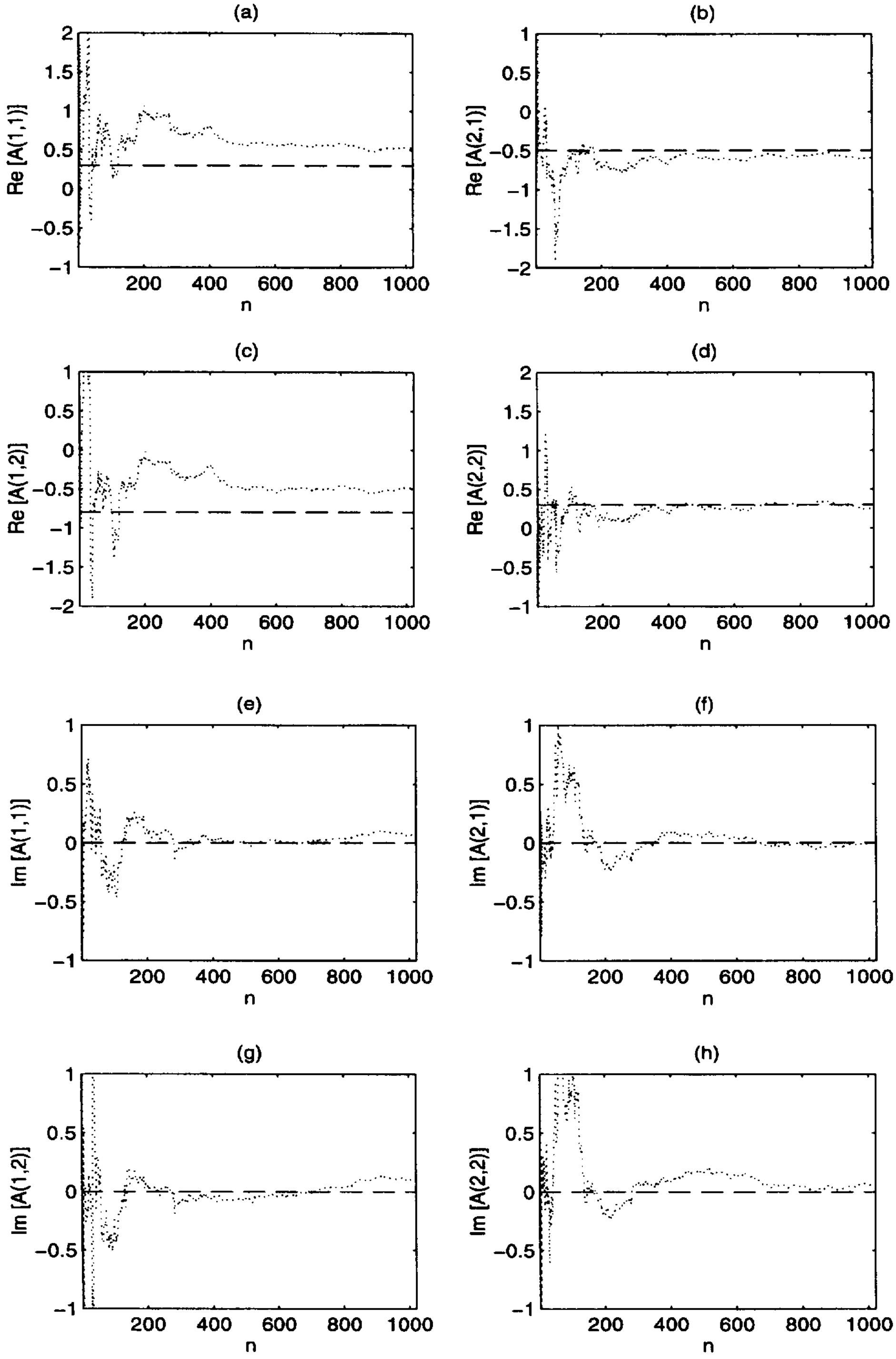

Fig. 3. Estimation of the A matrix: real and imaginary parts.

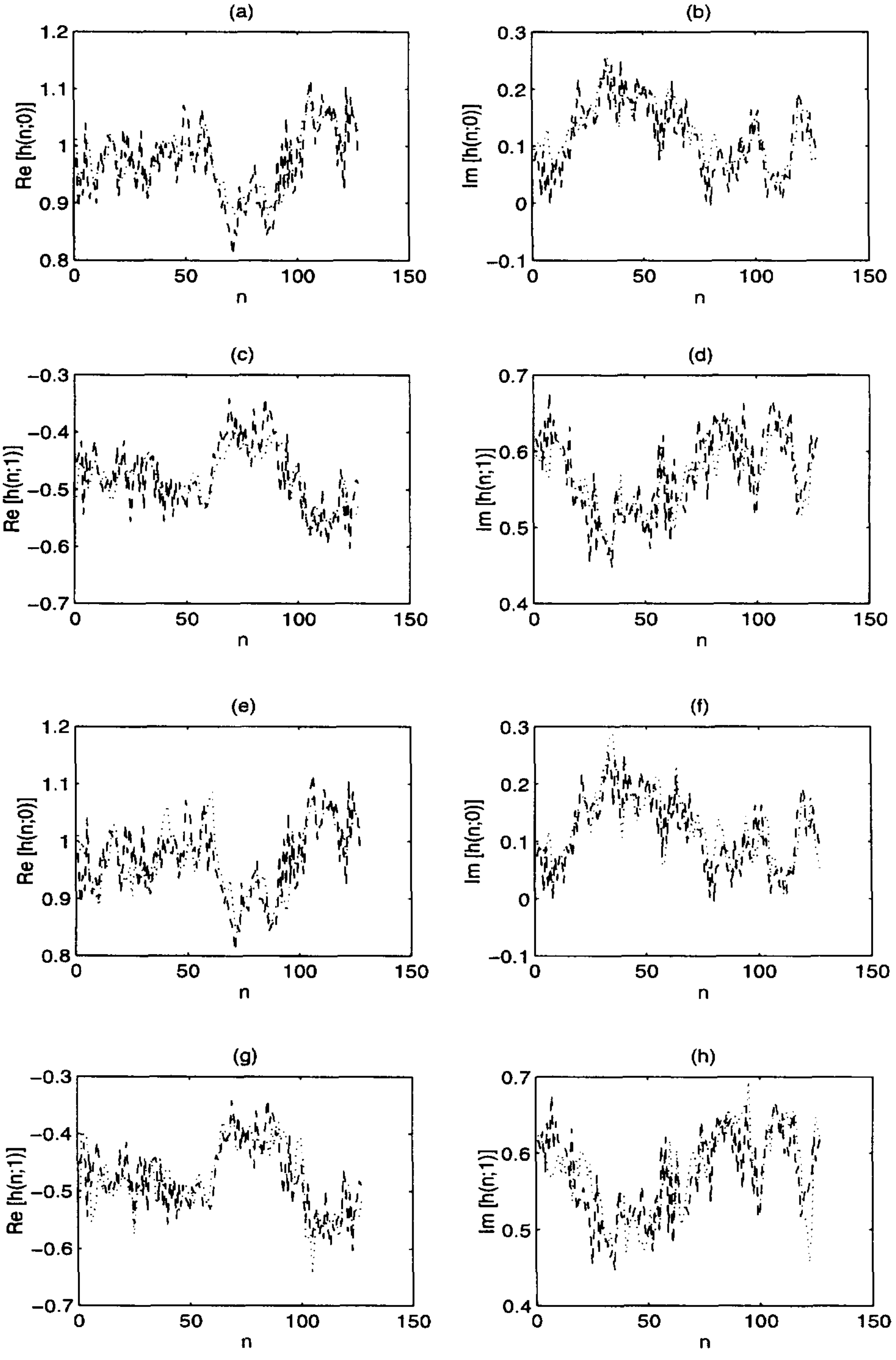

Fig. 4. Channel estimation: (a)–(d) proposed method; (e)–(f) method of [13].

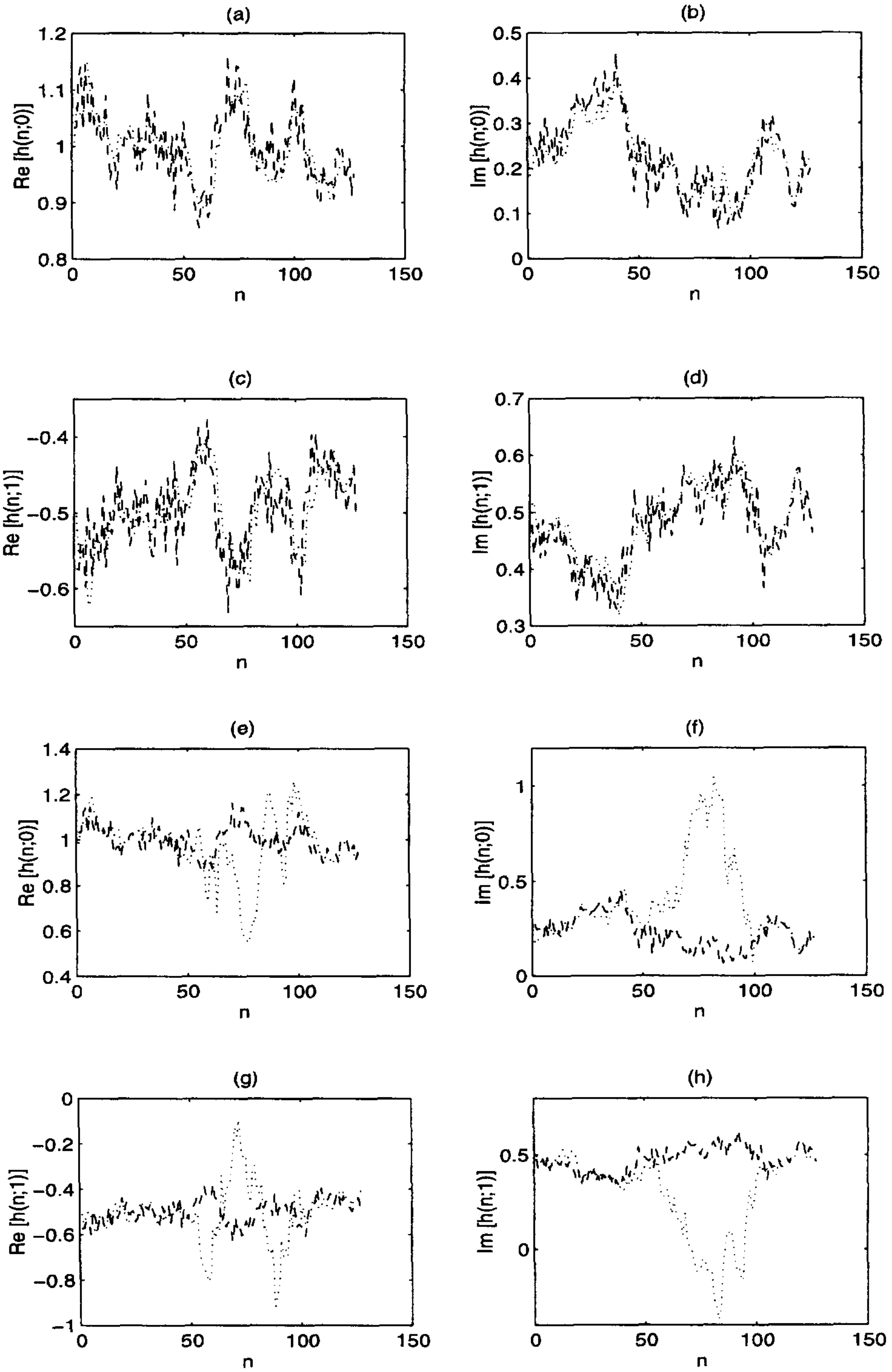

Fig. 5. Channel estimation: (a)–(d) proposed method; (e)–(f) method of [13].

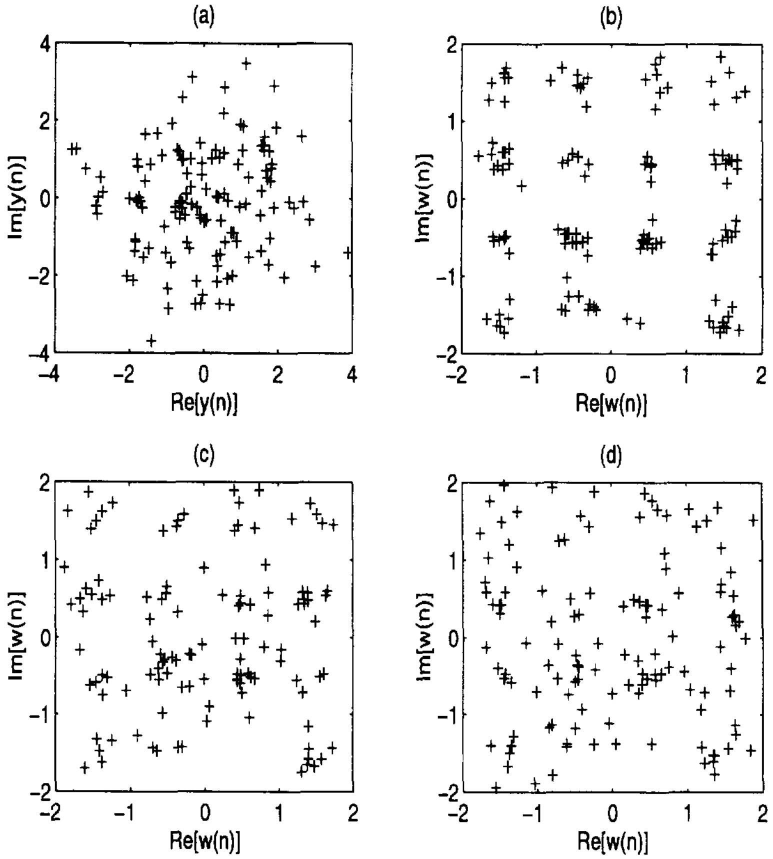

Fig. 6. Received (a) and equalized symbols: (b) proposed; (c) method of [13]; (d) plain RLS.

advantage. This is consistent with the fact that the g equalizer vector is assumed constant in the RLS framework.

2. When σ_u^2 is larger; i.e., the channel has more variations, the proposed method shows a clear advantage over either the method of [13] or the plain RLS method. This is due to the fact that the proposed algorithm is designed to cope with channel variations.
3. Both the proposed method and that of [13] are reasonably robust w.r.t. error in the $A(1)$ estimate/guess. It turns out that when the true $A(1)$ matrix is diagonal, the guess in (53) performed quite well even for cases where the true $A(1)$ was much different. This corroborates the assertion of [5, 13] that a rough guess is better than no guess at all. However, when $A(1)$ is a full matrix, the guess in (53) often leads to poor performance, and in some cases, worse than that of the plain RLS algorithm.

7. Conclusions

The most important message this work conveys, is that whenever some a priori information on the fading channel's behavior is available, it should be taken into account to improve the equalizer's performance. If the channel is described by randomly varying coefficients, a stochastic model for the time variations may allow the use of Kalman tracking procedures. In this work, we show how to estimate the parameters of this time-variation model using HOS, and how to integrate the estimator in a Kalman-based equalizer.

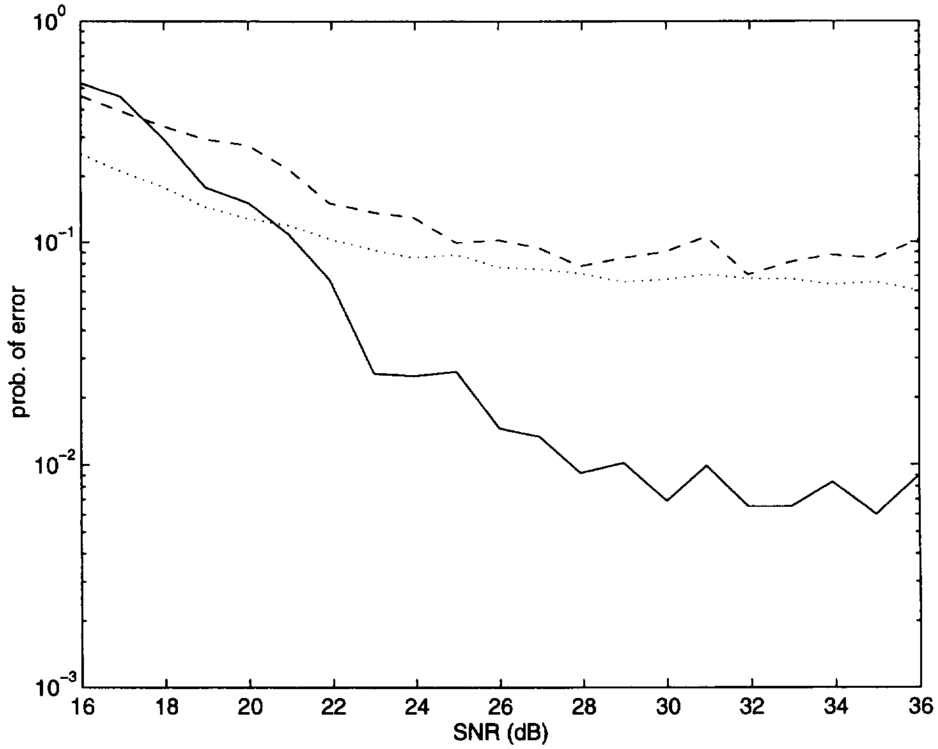

Fig. 7. Performance evaluation of the algorithms.

Appendix A. Proof of Proposition 1

The LS solution of (19) is given by

$$\hat{\theta}_{\tau,N} = \left[\frac{1}{N}\sum_{n=0}^{N-1}\phi(n;\tau)\phi^{\mathrm{H}}(n;\tau)\right]^{-1} \times \left[\frac{1}{N}\sum_{n=0}^{N-1}\phi(n;\tau)y(n)y^{*}(n+\tau)\right]. \qquad (A.1)$$

By substituting $y(n)y^{*}(n+\tau)$ in (A.1) from (19) we obtain

$$\hat{\theta}_{\tau,N} - \theta_{\tau} = \left[\frac{1}{N}\sum_{n=0}^{N-1}\phi(n;\tau)\phi^{\mathrm{H}}(n;\tau)\right]^{-1} \times \left[\frac{1}{N}\sum_{n=0}^{N-1}\phi(n;\tau)\varepsilon(n;\tau)\right], \qquad (A.2)$$

where $\varepsilon(n;\tau) \triangleq \sigma_v^2\delta(\tau) + e(n;\tau)$. Hence, we need to show that the estimation error $\hat{\theta}_{\tau,N} - \theta_{\tau}$ converges to zero as $N \to \infty$. Under assumptions (AS1) and (AS2), we can show that the regression error $e(n;\tau)$ is a zero-mean process with finite memory, and finite moments of any order. Then, the bracketed terms in (A.2) converge to their expected values as $N \to \infty$, both in the mean square sense and with probability one (see [24, Chapter 7]):

$$\hat{\theta}_{\tau,N} - \theta_{\tau} \to R_{\phi\phi}^{-1}R_{\phi e} \quad \text{as } N \to \infty \text{ w.p.l.} \qquad (A.3)$$

Using the fact that $e(n;\tau)$ is uncorrelated with the regressors, i.e., $R_{\phi e} = 0$, we conclude from (A.3) that $\hat{\theta}_{\tau,N}$ is consistent if and only if $R_{\phi\phi}$ is invertible.

We now show that $R_{\phi\phi}$ is invertible under (AS3). Let $i = k_0(q+1)+k_1, j = k_2(q+1)+k_3$, and order the regressors in (17) so that the (i,j) element of $R_{\phi\phi}$ is

$$[R_{\phi\phi}]_{i,j} = E\{w^{*}(n-k_0)w(n+\tau-k_1) \times w(n-k_2)w^{*}(n+\tau-k_3)\}. \qquad (A.4)$$

Eq. (A.4) can be written as

$$\begin{aligned}[\boldsymbol{R}_{\phi\phi}]_{i,j} &= \mathrm{cum}\{w^*(n-k_0), w(n+\tau-k_1),\\ &\qquad w(n-k_2), w^*(n+\tau-k_3)\}\\ &\quad +E\{w^*(n-k_0)w(n+\tau-k_1)\}\\ &\quad \times E\{w(n-k_2)w^*(n+\tau-k_3)\}\\ &\quad +E\{w^*(n-k_0)w(n-k_2)\}\\ &\quad \times E\{w(n+\tau-k_1)w^*(n+\tau-k_3)\}\\ &= \gamma_{4w}\delta(k_0-k_2)\\ &\quad \times\delta(k_0-k_1+\tau)\delta(k_0-k_3+\tau)\\ &\quad +\sigma_w^4\delta(k_0-k_1+\tau)\delta(k_2-k_3+\tau)\\ &\quad +\sigma_w^4\delta(k_0-k_2)\delta(k_1-k_3), \qquad (A.5)\end{aligned}$$

where the last equality is due to the i.i.d. nature of $w(n)$. From (A.5) we can see that for columns where $k_2 \neq k_3 - \tau$ only the third term is nonzero, yielding $[\boldsymbol{R}_{\phi\phi}]_{i,j} = \sigma_w^4\delta(i-j)$, i.e., portions of a diagonal matrix. For the columns where $k_2 = k_3 - \tau$, however, (A.5) becomes

$$\begin{aligned}[\boldsymbol{R}_{\phi\phi}]_{i,j} &= (\gamma_{4w}+\sigma_w^4)\delta(k_0-k_2)\delta(k_1-k_3)\\ &\quad +\sigma_w^4\delta(k_0-k_1+\tau). \qquad (A.6)\end{aligned}$$

Hence, those columns have nonzero entries (equal to σ_w^4), whenever $k_1 = k_0 + \tau$, with the exception of the diagonal entry ($k_2 = k_0$, $k_3 = k_1$), which equals $\gamma_{4w} + 2\sigma_w^4 = m_{4w}$.

Let us collect those columns (with $k_3 = k_2 + \tau$) separately, to examine whether they are linearly independent. These columns are independent of the remaining diagonal ones, since they have nonzero entries only in the places where the diagonal columns are zero. Those nonzero entries occur only for $i = k_0(q+2) + \tau$, i.e., they are separated by $q + 2$ zeros. If, for simplicity, in the notation we drop the common zero entries, we obtain the matrix

$$\tilde{\boldsymbol{R}}_{\phi\phi} = \begin{bmatrix} m_{4w} & \sigma_w^4 & \cdots & \sigma_w^4 \\ \sigma_w^4 & m_{4w} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_w^4 \\ \sigma_w^4 & \cdots & \sigma_w^4 & m_{4w} \end{bmatrix}. \qquad (A.7)$$

In order to check if its columns are linearly independent, we compute its eigenvalues. This matrix is circulant and the eigenvalues are given by the DFT of the first row [6, p. 73],

$$\begin{aligned}\lambda_k &= m_{4w} + \sum_{n=1}^{M-1}\sigma_w^4 \mathrm{e}^{-\mathrm{j}2\pi kn/M}\\ &= m_{4w} - \sigma_w^4 + \sigma_w^4\delta(k), \quad k = 0,\ldots,M-1, \quad (A.8)\end{aligned}$$

where M denotes the size of the matrix. Since $m_{4w} \neq 0$, $\lambda_k = 0$ only if $m_{4w} = \sigma_w^4$, which is impossible due to (AS3). Hence, $\boldsymbol{R}_{\phi\phi}$ has full rank and $\hat{\boldsymbol{\theta}}_{\tau,N}$ is a strongly consistent estimator of $\boldsymbol{\theta}_\tau$.

References

[1] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[2] G. D'Aria, F. Muratore and V. Palestini, "Simulation and performance of the pan-European land mobile radio system", *IEEE Trans. Vehicular Technol.*, Vol. 41, No. 2, May 1992, pp. 177–189.

[3] G. D'Aria, R. Piermarini and V. Zingarelli, "Fast adaptive equalizers for narrow-band TDMA mobile radio", *IEEE Trans. Vehicular Technol.*, Vol. 40, No. 2, May 1991, pp. 392–404.

[4] M. Athans, R. Ku and S.B. Gerswin, "The uncertainty threshold principle: Some fundamental limitations of optimal decision making under dynamic uncertainty", *IEEE Trans. Automatic Control*, Vol. 22, June 1977, pp. 491–495.

[5] T. Bohlin, "Four cases of identification of changing systems", in: R.K. Mehra and D.G. Lainiotis, eds., *System Identification: Advances and Case Studies*, Mathematics in Science and Engineering, Vol. 126, Academic Press, New York, 1976.

[6] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-day, San Francisco, CA, 1981.

[7] H.F. Chen and P.E. Caines, "On the adaptive control of a class of signals with random parameters and disturbances", *Automatica*, Vol. 21, November 1985, pp. 737–741.

[8] *EIA/TIA Interim Standard IS-54, Cellular System Dual-Mode Mobile Station-Base Station Compatibility Standard*, Electronic Industries Association, May 1990.

[9] A.W. Fuxjaeger and R.A. Iltis, "Adaptive parameter estimation using parallel Kalman filtering for spread spectrum code and Doppler tracking", *IEEE Trans. Commun.*, Vol. 42, No. 6, June 1994, pp. 2227–2230.

[10] A.W. Fuxjaeger and R.A. Iltis, "Acquisition of timing and Doppler shift in a direct sequence spread spectrum system", *IEEE Trans. Commun.*, Vol. 42, No. 10, October 1994, pp. 2870–2879.

[11] C. Hildreth and J.P. Houck, "Some estimators for a linear model with random coefficients", *Amer. Statist. Assoc. J.*, June 1986, pp. 584–595.

[12] M. Horvath, "Parameter estimation for linear discrete time models with random coefficients", *IEEE Trans. on*

Automatic Control, Vol. 36, No. 10, October 1991, pp. 1213–1215.

[13] R.A. Iltis, "Joint estimation of PN code delay and multipath using extended Kalman filter", *IEEE Trans. Commun.*, Vol. 38, No. 10, October 1990, pp. 1677–1685.

[14] R.A. Iltis and A.W. Fuxjaeger, "A digital DS spread-spectrum receiver with joint channel and Doppler shift estimation", *IEEE Trans. Commun.*, Vol. 39, No. 8, August 1991, pp. 1255–1267.

[15] W.C. Jakes, Jr., *Microwave Mobile Communications*, Wiley, New York, 1974.

[16] M.C. Jeruchim, P. Balaban and K.S. Shanmugan, *Simulation of Communication Systems*, Plenum Press, New York, 1992.

[17] D.W. Matolak, Variable-complexity trellis decoding and detection for a statistically-known, time-varying dispersive channel, Ph.D. Thesis, University of Virginia, Charlottesville, 1994.

[18] S.J. Papantoniou, Modeling the mobile radio channel, Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich, 1990.

[19] M.B. Priestley, *Spectral Analysis and Time-Series*, Academic Press, London, 1981.

[20] J. Proakis, *Digital Communications*, McGraw Hill, New York, 1989.

[21] B. Raj, "Linear regression with random coefficients: The finite sample and convergence properties", *J. Amer. Statist. Assoc.*, Vol. 70, No. 349, March 1975, pp. 127–137.

[22] T.S. Rappaport, S.Y. Seidel and R. Singh, "900-MHz multipath propagation measurements for U.S. digital cellular radiotelephone", *IEEE Trans. Vehicular Technol.*, Vol. 39, No. 2, May 1990, pp. 132–139.

[23] R. Sharma, W.D. Grover and W.A. Krzymien, "Forward-error-control (FEC)-assisted adaptive equalization for digital cellular mobile radio", *IEEE Trans. Vehicular Technol.*, Vol. 42, No. 1, February 1993, pp. 94–102.

[24] T. Söderström and Petre Stoica, *System Identification*, Series in Systems and Control Engineering, Prentice-Hall, London, 1989.

[25] A. Swami, "Multiplicative noise models: Parameter estimation using cumulants", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 355–373.