RECURSIVE MMSE ESTIMATION OF WIRELESS CHANNELS BASED ON TRAINING DATA AND STRUCTURED CORRELATION LEARNING

Gerald Matz

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
phone: +43 1 58801 38916, fax: +43 1 58801 38999, email: g.matz@ieee.org

ABSTRACT

We present a novel block-recursive technique for training based channel estimation that augments MMSE channel estimation with structured correlation learning to obtain the required statistics. We provide formulations for both stationary and nonstationary wireless channels along with a robust initialization procedure. The method is efficient, applicable in a variety of communications schemes (MIMO, multiuser, OFDM), and has performance superior to that of RLS from which it differs in an important way. Simulation examples show nearly optimum performance with particular gains for strongly correlated channels.

1. INTRODUCTION

We consider the problem of estimating the coefficients of a wireless fading channel on a block-by-block basis using training data. The simplest approach in such a context, not requiring any statistical prior knowledge, is the *least squares* (LS) channel estimator [1]. Exploiting channel statistics (i.e., long-term channel properties) via a (linear) *minimum mean square error* (MMSE) approach [1] allows to achieve smaller estimation errors. However, MMSE estimation is highly dependent on accurate knowledge of channel statistics. Several approaches have been proposed for the acquisition or specification of the required second-order statistics, e.g. specification of a "representative" default channel statistics, which obviously suffers from potential mismatch of the true and assumed statistics. Alternatively, channel statistics estimated from channel pre-estimates (e.g., the LS estimate) can be used [2]. In [3], we proposed *structured* correlation learning (see [1]) to obtain accurate channel statistics which also leads to superior channel estimation performance.

The main contributions of this paper build upon [3] and can be summarized as follows:

- In contrast to the batch procedure in [3], we provide an efficient recursive implementation of MMSE estimation including structured correlation learning; we also show how additional structure of the channel correlation matrix (i.e., uncorrelated channel coefficients) can be taken into account in a systematic way.
- We introduce a simple modification rendering the method applicable in nonstationary environments and we propose a robust initialization for the correlation learning process.
- We relate our method to RLS channel estimation and point out an important difference which explains the superior performance of our method.
- Finally, we illustrate the application of our approach to different contexts including flat-fading *multi-input multi-output* (MIMO) systems, multiuser uplinks over frequency-selective channels, *and orthogonal frequency division multiplexing* (OFDM) systems transmitting over doubly selective channels.

Funding by FWF grant J2302.

2. BACKGROUND

System Model. Our generic discrete system model for the kth block is given by (see Section 6 for specific applications)

$$\mathbf{y}[k] = \mathbf{S}\,\mathbf{h}[k] + \mathbf{n}[k]\,. \tag{1}$$

Here, $\mathbf{h}[k]$ is the $M \times 1$ channel coefficient vector to be estimated, $\mathbf{S}$ is an $N \times M$ matrix containing the known training data, $\mathbf{y}[k]$ is the $N \times 1$ received vector induced by the training data, and $\mathbf{n}[k]$ is the $N \times 1$ noise vector. We assume that $N \geq M$ and that $\mathbf{S}$ has full-rank, i.e., $\mathcal{S} = \operatorname{span}\{\mathbf{S}\}$ is an M-dimensional subspace of $\mathbb{C}^N$. The zero-mean vectors $\mathbf{n}[k]$ and $\mathbf{h}[k]$ are mutually independent and statistically characterized by[1] $\mathcal{E}\{\mathbf{n}[k]\,\mathbf{n}^H[k']\} = \sigma^2[k]\,\delta[k-k']\,\mathbf{I}$ and $\mathcal{E}\{\mathbf{h}[k]\,\mathbf{h}^H[k']\} = \mathbf{R_h}[k]\,\delta[k-k']$, respectively. The latter implies that subsequent training blocks are separated by more than the channel's coherence time; thus, the channel coefficients can be estimated in a block-by-block fashion without performance loss.

LS and MMSE Channel Estimators. The simplest channel estimator for the model (1) is the LS estimator [1]

$$\hat{\mathbf{h}}_{\mathrm{LS}}[k] = \mathbf{S}^{\#}\mathbf{y}[k] = \mathbf{h}[k] + \mathbf{S}^{\#}\mathbf{n}[k],$$

where $\mathbf{S}^{\#} = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$ denotes the pseudo-inverse of $\mathbf{S}$. Apart of the training data contained in $\mathbf{S}$, the LS estimator does not require any (statistical) prior knowledge. To exploit long-term channel properties (i.e., channel statistics), usually a (linear) MMSE channel estimator is used [1]:

$$\hat{\mathbf{h}}[k] = \mathbf{R_h}[k]\mathbf{S}^H(\mathbf{S}\mathbf{R_h}[k]\mathbf{S}^H + \sigma^2[k]\mathbf{I})^{-1}\mathbf{y}[k] \tag{2a}$$

$$= \mathbf{S}^{\#}\,(\mathbf{I} - \sigma^2[k]\mathbf{R}_{\mathbf{y}}^{-1}[k])\,\mathbf{y}[k]\,. \tag{2b}$$

Here, $\mathbf{R_y}[k] = \mathcal{E}\{\mathbf{y}[k]\,\mathbf{y}^H[k]\}$. In practice, the correlation matrices $\mathbf{R_h}[k]$, $\mathbf{R_y}[k]$ and the noise variance σ^2 have to be

[1] Here, $\mathcal{E}\{\cdot\}$ denotes expectation (ensemble averaging) and superscript ^{H} denotes hermitian transposition.

specified. In a stationary context ($\mathbf{R}_\mathbf{y}[k] = \mathbf{R}_\mathbf{y}$, $\mathbf{R}_\mathbf{h}[k] = \mathbf{R}_\mathbf{h}$), either default specifications (i.e., $\mathbf{R}_\mathbf{h} = \mathbf{I}$) or robust designs based on least favorable correlations [4] are used; alternatively, the correlation matrices $\mathbf{R}_\mathbf{y}$ and $\mathbf{R}_\mathbf{h}$ are replaced with the respective sample correlation matrices [2]:

$$\hat{\mathbf{R}}_\mathbf{y}[k]=\frac{1}{k}\sum_{i=1}^{k}\mathbf{y}[i]\mathbf{y}^H[i], \quad \hat{\mathbf{R}}_\mathbf{h}^{\mathrm{LS}}[k]=\frac{1}{k}\sum_{i=1}^{k}\hat{\mathbf{h}}_{\mathrm{LS}}[i]\hat{\mathbf{h}}_{\mathrm{LS}}^H[i]\,, \quad (3)$$

Although these sample correlations can be shown to be related as $\hat{\mathbf{R}}_\mathbf{h}^{\mathrm{LS}}[k] = \mathbf{S}^\#\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{S}^{\#H}$, it turns out that using them in (2a) and (2b), respectively, gives *different* MMSE channel estimators.

3. STRUCTURED CORRELATION LEARNING

The sample correlations $\hat{\mathbf{R}}_\mathbf{h}^{\mathrm{LS}}$ and $\hat{\mathbf{R}}_\mathbf{y}$ do not take into account the correlation structure

$$\mathbf{R}_\mathbf{y}[k] = \mathbf{S}\mathbf{R}_\mathbf{h}[k]\mathbf{S}^H + \sigma^2[k]\mathbf{I}\,. \quad (4)$$

This motivates us to propose a novel approach for *structured correlation learning* [1] in order to estimate $\mathbf{R}_\mathbf{y}[k]$ and $\mathbf{R}_\mathbf{h}[k]$ in a way consistent with (4).

Nonstationary Case. We derive an algorithm for structured correlation learning in nonstationary scenarios where $\mathbf{R}_\mathbf{h}[k]$, $\mathbf{R}_\mathbf{y}[k]$, and $\sigma^2[k]$ depend indeed on the block (time) index k. This extends the stationary case considered in [3]. With structured correlation estimation (cf. [1]), the channel correlation $\mathbf{R}_\mathbf{h}[k]$ and the noise variance $\sigma^2[k]$ are being considered as parameters that are estimated via an LS approach from an "observed" receive correlation matrix $\hat{\mathbf{R}}_\mathbf{y}[k]$ by taking into account (4):

$$\big(\hat{\mathbf{R}}_\mathbf{h}[k], \hat{\sigma}^2[k]\big) = \arg\min_{\substack{\mathbf{R}_\mathbf{h}[k]\\ \sigma^2[k]}} \big\|\hat{\mathbf{R}}_\mathbf{y}[k] - (\mathbf{S}\mathbf{R}_\mathbf{h}[k]\mathbf{S}^H + \sigma^2[k]\mathbf{I})\big\|^2 \quad (5)$$

($\|\cdot\|$ denotes the Frobenius norm). Since we are interested in nonstationary scenarios, we propose to include exponential forgetting in the sample correlation of $\mathbf{y}[k]$, i.e.,

$$\hat{\mathbf{R}}_\mathbf{y}[k] = (1-\lambda)\sum_{i=1}^{k}\lambda^{k-i}\mathbf{y}[i]\mathbf{y}^H[i]\,. \quad (6)$$

Here, $0 < \lambda < 1$ denotes the forgetting factor. According to [1], (5) amounts to solving the linear equations ($\mathrm{tr}\{\cdot\}$ denotes trace)

$$\mathbf{S}^H\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{S} \;=\; \mathbf{S}^H\mathbf{S}\,\hat{\mathbf{R}}_\mathbf{h}[k]\mathbf{S}^H\mathbf{S} + \hat{\sigma}^2[k]\mathbf{S}^H\mathbf{S} \quad (7)$$

$$\mathrm{tr}\{\hat{\mathbf{R}}_\mathbf{y}[k]\} \;=\; \mathrm{tr}\{\mathbf{S}\hat{\mathbf{R}}_\mathbf{h}[k]\mathbf{S}^H\} + N\hat{\sigma}^2[k] \quad (8)$$

for $\hat{\mathbf{R}}_\mathbf{h}[k]$ and $\hat{\sigma}^2$. Pre- and post-multiplying (7) by $(\mathbf{S}^H\mathbf{S})^{-1}$ yields the channel correlation estimate

$$\hat{\mathbf{R}}_\mathbf{h}[k] = \mathbf{S}^\#\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{S}^{\#H} - \hat{\sigma}^2[k](\mathbf{S}^H\mathbf{S})^{-1} \quad (9a)$$

$$= \mathbf{S}^\#\big(\hat{\mathbf{R}}_\mathbf{y}[k] - \hat{\sigma}^2[k]\mathbf{I}\big)\mathbf{S}^{\#H}\,, \quad (9b)$$

where the second expression follows from $(\mathbf{S}^H\mathbf{S})^{-1} = \mathbf{S}^\#\mathbf{S}^{\#H}$. Plugging this expression into (8) leads to

$$N\hat{\sigma}^2[k] = \mathrm{tr}\{\hat{\mathbf{R}}_\mathbf{y}[k] - \mathbf{P}\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{P} + \hat{\sigma}^2[k]\mathbf{P}\}$$

$$= \mathrm{tr}\{\mathbf{P}^\perp\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{P}^\perp\} + M\hat{\sigma}^2[k]\,. \quad (10)$$

Here, $\mathbf{P} = \mathbf{S}\mathbf{S}^\#$ and $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$ denote the orthogonal projection matrices on $\mathcal{S}$ and $\mathcal{S}^\perp$, respectively. From (10) we obtain the noise variance estimate (cf. (6))

$$\hat{\sigma}^2[k] = \frac{1}{N-M}\,\mathrm{tr}\{\mathbf{P}^\perp\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{P}^\perp\} \quad (11a)$$

$$= \frac{1-\lambda}{N-M}\sum_{i=1}^{k}\lambda^{k-i}\big\|\mathbf{P}^\perp\mathbf{y}[i]\big\|^2. \quad (11b)$$

Note that $\mathbf{P}^\perp\mathbf{y}[k]$ can be interpreted as estimate of the noise within the $(N-M)$-dimensional noise subspace $\mathcal{S}^\perp$.

Using (9a) and (11a), the structured correlation estimate of $\mathbf{R}_\mathbf{y}[k]$ is given by

$$\hat{\mathbf{R}}_\mathbf{y}^{(s)}[k] = \mathbf{S}\hat{\mathbf{R}}_\mathbf{h}[k]\mathbf{S}^H + \hat{\sigma}^2[k]\mathbf{I} = \mathbf{P}\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{P} + \hat{\sigma}^2[k]\mathbf{P}^\perp. \quad (12)$$

It is seen that with this estimate, $\hat{\mathbf{R}}_\mathbf{y}[k]$ is retained within $\mathcal{S}$ while within the noise-only subspace $\mathcal{S}^\perp$ it is replaced by the more accurate estimate $\hat{\sigma}^2[k]\mathbf{P}^\perp$.

Stationary Case. The results obtained in [3] for stationary scenarios (i.e., $\mathbf{R}_\mathbf{h}[k]$, $\mathbf{R}_\mathbf{y}[k]$, and $\sigma^2[k]$ are *independent* of k) are consistent with our foregoing development. In particular, (9), (11), and (12) are still valid, however, with the exponential forgetting sample correlation $\hat{\mathbf{R}}_\mathbf{y}[k]$ in (6) replaced by the conventional sample correlation in (3). The noise variance estimate in this case reads (cf. (11b))

$$\hat{\sigma}^2[k] \;=\; \frac{1}{N-M}\,\frac{1}{k}\sum_{i=1}^{k}\big\|\mathbf{P}^\perp\mathbf{y}[i]\big\|^2. \quad (13)$$

Additional Structure. The structured correlation learning approach discussed above can be further refined if additional structural constraints on the channel correlation $\mathbf{R}_\mathbf{h}[k]$ are known. As an example, we consider the special case where the channel coefficients $\mathbf{h}[k]$ are uncorrelated, i.e., $\mathbf{R}_\mathbf{h}[k] = \mathrm{diag}\{\rho_1^2[k], \ldots, \rho_M^2[k]\}$ is a diagonal matrix consisting of the channel powers $\rho_i^2[k] = \mathcal{E}\{\,|h_i[k]|^2\}$ ($h_i[k]$ is the ith element of $\mathbf{h}[k]$). To simplify the discussion, we assume $\mathbf{S}$ to be orthogonal. Denoting the ith column of $\mathbf{S}$ by $\mathbf{s}_i$, the receive correlation matrix can here be written as

$$\mathbf{R}_\mathbf{y}[k] = \sum_{i=1}^{M}\rho_i^2[k]\,\mathbf{s}_i\mathbf{s}_i^H + \sigma^2[k]\mathbf{I}\,.$$

Defining, $\boldsymbol{\rho}_\mathbf{h}^2[k] = \big[\rho_1^2[k]\ldots\rho_M^2[k]\big]^T$, the structured correlation learning problem reads

$$\big(\hat{\boldsymbol{\rho}}_\mathbf{h}^2[k], \hat{\sigma}^2[k]\big) = \arg\min_{\substack{\boldsymbol{\rho}_\mathbf{h}^2[k]\\ \sigma^2[k]}}\left\|\hat{\mathbf{R}}_\mathbf{y}[k] - \left[\sum_{i=1}^{M}\rho_i^2[k]\,\mathbf{s}_i\mathbf{s}_i^H + \sigma^2[k]\mathbf{I}\right]\right\|^2, \quad (14)$$

which amounts to solving the system of linear equations

$$\mathbf{s}_i^H\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{s}_i = \hat{\rho}_i^2[k] + \hat{\sigma}^2[k],$$

$$\mathrm{tr}\{\hat{\mathbf{R}}_\mathbf{y}[k]\} = \sum_{i=1}^{M}\hat{\rho}_i^2[k] + \hat{\sigma}^2[k]\,N\,.$$

This can be shown to lead to the channel power estimates

$$\hat{\rho}_i^2[k] = \mathbf{s}_i^H\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{s}_i - \hat{\sigma}^2[k],$$

and to the same noise variance estimate as in (11). It is furthermore easily seen that

$$\hat{\mathbf{R}}_\mathbf{h}[k] = \mathrm{diag}\{\hat{\boldsymbol{\rho}}_\mathbf{h}^2[k]\} = \mathrm{diag}\{\mathbf{S}^H\hat{\mathbf{R}}_\mathbf{y}[k]\mathbf{S}\} - \hat{\sigma}^2[k]\mathbf{I}\,.$$

Comparing with (9) and recalling our assumption $\mathbf{S}^H\mathbf{S}=\mathbf{I}$, we see that the additional structure of having uncorrelated channel coefficients is simply enforced by setting all non-diagonal elements of $\mathbf{S}^H\hat{\mathbf{R}}_{\mathbf{y}}[k]\mathbf{S}$ to zero.

4. RECURSIVE MMSE ESTIMATION

In this section, we show that by using structured correlation estimates in the MMSE channel estimator (2), we can arrive at an efficient recursive implementation that performs block-wise MMSE estimation while adaptively learning the underlying channel statistics.

4.1. Subspace Formulation

Plugging $\hat{\mathbf{R}}_{\mathbf{h}}[k]$, $\hat{\mathbf{R}}_{\mathbf{y}}^{(s)}[k]$, and $\hat{\sigma}^2[k]$ into (2a) and (2b), respectively, leads to two different but equivalent expressions for an improved MMSE channel estimator:

$$\begin{aligned}\hat{\mathbf{h}}[k] &= \hat{\mathbf{R}}_{\mathbf{h}}\mathbf{S}^H(\mathbf{S}\hat{\mathbf{R}}_{\mathbf{h}}\mathbf{S}^H+\hat{\sigma}^2[k]\mathbf{I})^{-1}\mathbf{y}[n] \\ &= \mathbf{S}^{\#}\Big(\mathbf{I}-\hat{\sigma}^2[k]\big[\hat{\mathbf{R}}_{\mathbf{y}}^{(s)}\big]^{-1}\Big)\mathbf{y}[n]\,. \end{aligned} \tag{15}$$

The rest of the paper will be based on the second formulation of the MMSE estimator, which allows to show that in fact all calculations required for channel estimation can be performed within the M-dimensional subspace $\mathcal{S}$ (obviously, $\mathcal{S}^{\perp}$ contains only uncorrelated noise and thus needs to be considered only for noise variance estimation). Consider an orthonormal matrix $\mathbf{U}$ whose columns form an orthonormal basis for $\mathcal{S}$. Defining the transformed length-M observation vector $\mathbf{x}[k]=\mathbf{U}^H\mathbf{y}[k]=\mathbf{C}\mathbf{h}[k]+\mathbf{U}^H\mathbf{n}[k]$ with correlation matrix $\mathbf{R}_{\mathbf{x}}[k]=\mathcal{E}\{\mathbf{x}[k]\,\mathbf{x}^H[k]\}$ and the transformed $M\times M$ training data matrix $\mathbf{C}=\mathbf{U}^H\mathbf{S}$, it can indeed be shown that (15) is equivalent to

$$\hat{\mathbf{h}}[k] = \mathbf{C}^{-1}\big(\mathbf{I}-\hat{\sigma}^2[k]\hat{\mathbf{R}}_{\mathbf{x}}^{-1}[k]\big)\,\mathbf{x}[k]\,. \tag{16}$$

where in the stationary case

$$\hat{\mathbf{R}}_{\mathbf{x}}[k]=\frac{1}{k}\sum_{i=1}^{k}\mathbf{x}[i]\,\mathbf{x}^H[i] \tag{17}$$

and in the nonstationary case

$$\hat{\mathbf{R}}_{\mathbf{x}}[k]=(1-\lambda)\sum_{i=1}^{k}\lambda^{k-i}\mathbf{x}[i]\,\mathbf{x}^H[i]\,. \tag{18}$$

We will show below that the estimator (16) has a complexity of $\mathcal{O}(M^2)$ operations per block. This follows from the fact that $\hat{\mathbf{R}}_{\mathbf{x}}^{-1}[k]$ can be computed from $\hat{\mathbf{R}}_{\mathbf{x}}^{-1}[k-1]$ with complexity $\mathcal{O}(M^2)$ using Woodbury's identity [1]. Furthermore,

$$\big\|\mathbf{P}^{\perp}\mathbf{y}[k]\big\|^2=\big\|\mathbf{y}[k]\big\|^2-\big\|\mathbf{P}\mathbf{y}[k]\big\|^2=\big\|\mathbf{y}[k]\big\|^2-\big\|\mathbf{x}[k]\big\|^2$$

implies that the noise variance update (cf. (11b) and (13)) amounts to $\mathcal{O}(N+M)$ operations.

4.2. Recursive Implementation

We next discuss an efficient block-recursive implementation of the subspace version (16) of our MMSE channel estimator. For clarity of presentation, we make the simplifying assumption that the noise variance is independent of k and exactly known, i.e., $\hat{\sigma}^2[k]=\sigma^2$. We will show below that highly accurate noise variance estimates justifying this assumption can be obtained according to (13) during the initialization phase of the recursive estimator. For the subsequent development, we reformulate the channel estimate (16) via an MMSE (Wiener) filter as $\hat{\mathbf{h}}[k]=\mathbf{W}[k]\,\mathbf{x}[k]$ where

$$\mathbf{W}[k]=\mathbf{C}^{-1}\big(\mathbf{I}-\sigma^2\hat{\mathbf{R}}_{\mathbf{x}}^{-1}[k]\big)\,. \tag{19}$$

Stationary Case. We first consider a stationary channel ($\mathbf{R}_{\mathbf{h}}[k]=\mathbf{R}_{\mathbf{h}}$, $\mathbf{R}_{\mathbf{x}}[k]=\mathbf{C}\mathbf{R}_{\mathbf{h}}\mathbf{C}^H+\sigma^2\mathbf{I}$). From (17) we have

$$\hat{\mathbf{R}}_{\mathbf{x}}[k]=\left(1-\frac{1}{k}\right)\hat{\mathbf{R}}_{\mathbf{x}}[k-1]+\frac{1}{k}\mathbf{x}[k]\,\mathbf{x}^H[k]\,.$$

This is a rank-one update of $\hat{\mathbf{R}}_{\mathbf{x}}[k-1]$ that allows to use Woodbury's identity [1] to obtain a recursion for the (scaled) inverse correlation $\mathbf{Q}[k]=\big[k\hat{\mathbf{R}}_{\mathbf{x}}[k]\big]^{-1}$ required in (19):

$$\mathbf{Q}[k]=\mathbf{Q}[k-1]-\mathbf{Q}[k-1]\,\mathbf{x}[k]\,\mathbf{g}^H[k]\,. \tag{20}$$

Here, we used the Kalman-type gain vector

$$\mathbf{g}[k]=\frac{\mathbf{Q}[k-1]\,\mathbf{x}[k]}{1+\mathbf{x}^H[k]\,\mathbf{Q}^H[k-1]\,\mathbf{x}[k]}. \tag{21}$$

Plugging (20) into (16), it can be shown that

$$\mathbf{W}[k]=\frac{k}{k-1}\Big(\mathbf{W}[k-1]+\mathbf{e}[k]\mathbf{g}^H[k]\Big)-\frac{1}{k-1}\mathbf{C}^{-1}\,, \tag{22}$$

with the "noisy" channel estimation error

$$\mathbf{e}[k]=\hat{\mathbf{h}}_{\mathrm{LS}}[k]-\hat{\mathbf{h}}[k|k-1] \tag{23}$$

and the predicted channel vector

$$\hat{\mathbf{h}}[k|k-1]=\mathbf{W}[k-1]\,\mathbf{x}[k]\,, \tag{24}$$

obtained by applying the MMSE filter from block k–1 to the observation in the kth block. For large k, it is seen that the recursion (22) approximately simplifies to the rank-one RLS-type update $\mathbf{W}[k]\approx\widetilde{\mathbf{W}}[k]=\mathbf{W}[k-1]+\mathbf{e}[k]\mathbf{g}^H[k]$. However, for small k, the MMSE filter $\mathbf{W}[k]$ is a linear combination of $\widetilde{\mathbf{W}}[k]$ and the LS estimation filter[2] $\mathbf{C}^{-1}$.

By applying (22) to $\mathbf{x}[k]$, we obtain

$$\hat{\mathbf{h}}[k] = \frac{k}{k-1}\frac{1}{1+\alpha^2[k]}\,\hat{\mathbf{h}}[k|k-1]+\frac{\alpha^2[k]-\frac{1}{k-1}}{1+\alpha^2[k]}\hat{\mathbf{h}}_{\mathrm{LS}}[k]\,.$$

Here, we used the shorthand notation $\alpha^2[k]=\mathbf{x}^H[k]\mathbf{Q}[k-1]\mathbf{x}[k]$. For large k, this equation shows that the MMSE estimate (approximately) is a convex combination of $\hat{\mathbf{h}}[k|k-1]$ (which involves all statistical information accumulated up to block $k-1$) and the current LS estimate $\hat{\mathbf{h}}_{\mathrm{LS}}[k]$ (which involves no statistical knowledge at all). In particular, if $\mathbf{x}[k]$ is poorly represented in $\hat{\mathbf{R}}_{\mathbf{x}}[k]$ (as will typically be the case in the initial learning phase), $\alpha^2[k]$ will be very large and $\hat{\mathbf{h}}[k]$ will be closer to $\hat{\mathbf{h}}_{\mathrm{LS}}[k]$. In contrast, if $\mathbf{x}[k]$ is strongly represented in $\hat{\mathbf{R}}_{\mathbf{x}}[k]$ (typically after convergence), $\alpha^2[k]$ is small and $\hat{\mathbf{h}}[k]\approx\hat{\mathbf{h}}[k|k-1]$ (i.e., basically $\mathbf{W}[k]\approx\mathbf{W}[k-1]$).

Nonstationary Case. We next extend the foregoing block-recursive MMSE filter implementation to channels whose correlation matrix $\mathbf{R}_{\mathbf{h}}[k]$ changes gradually from block to block. Obviously, $\mathbf{R}_{\mathbf{x}}[k]=\mathbf{C}\mathbf{R}_{\mathbf{h}}[k]\mathbf{C}^H+\sigma^2\mathbf{I}$ also depends

[2]Note that $\hat{\mathbf{h}}_{\mathrm{LS}}[k]=\mathbf{S}^{\#}\mathbf{y}[k]=\mathbf{C}^{-1}\mathbf{U}^H\mathbf{y}[k]=\mathbf{C}^{-1}\mathbf{x}[k]$.

explicitly on the block index k. In that case, the correlation estimate allows for the recursive computation (cf. (18))

$$\hat{\mathbf{R}}_{\mathbf{x}}[k] = \lambda\,\hat{\mathbf{R}}_{\mathbf{x}}[k-1] + (1-\lambda)\mathbf{x}[k]\,\mathbf{x}^H[k]\,.$$

Defining the (scaled) inverse correlation $\mathbf{Q}[k]=(1-\lambda)\hat{\mathbf{R}}_{\mathbf{x}}^{-1}[k]$ and again using Woodbury's identity results in[3]

$$\mathbf{Q}[k] = \frac{1}{\lambda}\big[\mathbf{Q}[k-1] - \mathbf{Q}[k-1]\,\mathbf{x}[k]\,\mathbf{g}^H[k]\big]\,, \tag{25}$$

with the Kalman-type gain vector

$$\mathbf{g}[k] = \frac{\mathbf{Q}[k-1]\,\mathbf{x}[k]}{\lambda + \mathbf{x}^H[k]\,\mathbf{Q}^H[k-1]\,\mathbf{x}[k]}. \tag{26}$$

Plugging (25) into (16), we obtain in a similar way as for the stationary case the filter recursion

$$\mathbf{W}[k] = \frac{1}{\lambda}\Big(\mathbf{W}[k-1] + \mathbf{e}[k]\mathbf{g}^H[k]\Big) - \frac{1-\lambda}{\lambda}\mathbf{C}^{-1}$$

and a corresponding expression for the channel estimate:

$$\hat{\mathbf{h}}[k] \;=\; \frac{1}{\lambda}\frac{1}{1+\alpha^2[k]}\,\hat{\mathbf{h}}[k|k-1] + \frac{\alpha^2[k] - \frac{1-\lambda}{\lambda}}{1+\alpha^2[k]}\hat{\mathbf{h}}_{\mathrm{LS}}[k]\,.$$

Here, the "noisy" channel estimation error $\mathbf{e}[k]$ and the predicted channel vector $\hat{\mathbf{h}}[k|k-1]$ are defined as before (cf. (23) and (24), respectively). It is seen that the results for the nonstationary case can easily be obtained from those for the stationary case by formally replacing $1/k$ by $1-\lambda$ and vice versa. Hence, all interpretations provided in the stationary case for large k apply to the nonstationary case with λ close to 1 (the usual case) as well.

4.3. Complexity

We next provide an analysis of the computational complexity of our block-recursive MMSE filter. The first stage consists of the transformation $\mathbf{x}[k] = \mathbf{U}^H\mathbf{y}[k]$ which requires $\mathcal{O}(NM)$ operations. Then the LS estimate $\hat{\mathbf{h}}_{\mathrm{LS}}[k]=\mathbf{C}^{-1}\mathbf{x}[k]$ and the predicted channel vector $\hat{\mathbf{h}}[k|k-1] = \mathbf{W}[k-1]\mathbf{x}[k]$ are calculated using $\mathcal{O}(M^2)$ operations each. The calculation of the Kalman gain vector $\mathbf{g}[k]$ is as well dominated by the $\mathcal{O}(M^2)$ complexity of the matrix-vector product $\tilde{\mathbf{x}}[k] = \mathbf{Q}[k-1]\mathbf{x}[k]$. The $\mathcal{O}(M^2)$ outer product of the intermediate vector $\tilde{\mathbf{x}}[k]$ and the Kalman gain vector $\mathbf{g}[k]$ determines the complexity of the update equation for $\mathbf{Q}[k]$. The update of the MMSE filter $\mathbf{W}[k]$ amounts to the outer product of $\mathbf{e}[k]$ and $\mathbf{g}[k]$ and requires $\mathcal{O}(M^2)$ operations as well. Finally, the MMSE filtering step requires also $\mathcal{O}(M^2)$ operations.

4.4. Initialization

Regarding the initialization of the filter recursion, we note that $\hat{\mathbf{R}}_{\mathbf{x}}[k]$ is invertible only for $k \geq M$. Hence, the recursion for $\mathbf{Q}[k]$ is actually meaningful only for $k > M$. In our simulations we observed $\hat{\mathbf{R}}_{\mathbf{x}}[k]$ to be statistically unstable even for block indices up to $2M \ldots 3M$. This is even more true for $\hat{\mathbf{R}}_{\mathbf{x}}^{-1}[k]$ and results in estimates $\hat{\mathbf{h}}[k]$ that perform worse than the LS estimate. Intuitively, using poor (i.e., mismatched) correlation estimates in the nominally optimum MMSE estimator (that tries to exploit the full second-order statistics) leads to larger errors than using the rather robust LS estimate that does not attempt to exploit any statistics at all.

As a remedy for this problem, we suggest to use the LS estimator while learning $\mathbf{R}_{\mathbf{x}}[k]$ in the first M blocks and afterwards switch to the MMSE recursions. The first M blocks can also be used to learn the noise variance σ^2. Our simulations indicated that typically another $M \ldots 2M$ blocks are required in order for $\hat{\mathbf{R}}_{\mathbf{x}}[k]$ and $\mathbf{Q}[k]$ to be stable enough to allow the MMSE estimator to outperform the LS estimator.

To improve initial convergence, we propose a robust initialization that uses only the diagonal $\mathbf{D}_{\mathbf{x}}[k] = \mathrm{diag}\{\hat{\mathbf{R}}_{\mathbf{x}}[k]\}$ (i.e., the mean power of the channel coefficients) for $k = 1, \ldots, M$. Based on $\mathbf{D}_{\mathbf{x}}[M]$, the full correlation and its inverse are acquired only for $k > M$. We note that the diagonal matrix $\mathbf{D}_{\mathbf{x}}[k]$ can be stably inverted even for $k < M$ and it typically allows to outperform the LS estimator already within the $(M+1)$th block. In essence, using only $\mathbf{D}_{\mathbf{x}}[k]$ in the MMSE filter design amounts to a robust MMSE estimator [5] which exploits knowledge of the channel coefficient powers only but ignores potential correlations of the channel coefficients. Such an approach is known to be much less susceptible to errors (mismatch) in the estimated statistics. Indeed, we observed in our simulations that this robust initialization features improved performance (cf. Section 7).

5. RELATION TO RLS

The structure of the block-recursive MMSE estimator and its $\mathcal{O}(M^2)$ complexity are strongly reminiscent of the RLS algorithm [6]. We thus next provide a comparison with an RLS approach to acquire and track the optimal estimation filter. We note that direct RLS or Kalman tracking of the channel vector $\mathbf{h}[k]$ is not feasible since channel realizations in different blocks are assumed uncorrelated.

To derive the RLS recursions [6], we define $\hat{\mathbf{h}}_{\mathrm{RLS}}[k] = \mathbf{W}_{\mathrm{RLS}}[k]\mathbf{x}[k]$ and minimize the respective errors

$$\varepsilon^2 = \frac{1}{k}\sum_{i=1}^{k}\|\mathbf{h}[i] - \hat{\mathbf{h}}_{\mathrm{RLS}}[i]\|^2, \quad \varepsilon^2 = \sum_{i=1}^{k}\lambda^{k-i}\|\mathbf{h}[i] - \hat{\mathbf{h}}_{\mathrm{RLS}}[i]\|^2,$$

for the stationary and nonstationary case. It can be shown that these minimization problems amount to solving a system of linear equations for the optimal RLS filter $\mathbf{W}_{\mathrm{RLS}}[k]$ by inverting the sample correlation of the observation (cf. (17) and (18)). Doing this inversion in a recursive fashion leads to the following update equation for the RLS filter:

$$\mathbf{W}_{\mathrm{RLS}}[k] = \mathbf{W}_{\mathrm{RLS}}[k-1] + \big(\mathbf{h}[k] - \hat{\mathbf{h}}_{\mathrm{RLS}}[k|k-1]\big)\mathbf{g}^H[k]\,. \tag{27}$$

Here, the Kalman gain $\mathbf{g}^H[k]$ is defined as before (cf. (21) and (26)) and $\hat{\mathbf{h}}_{\mathrm{RLS}}[k|k-1] = \mathbf{W}_{\mathrm{RLS}}[k-1]\,\mathbf{x}[k]$ denotes the RLS channel prediction. Unfortunately, the true channel vector $\mathbf{h}[k]$ in the RLS update (27) is not available; hence, it needs to be replaced with the LS estimate $\hat{\mathbf{h}}_{\mathrm{LS}}[k]$ in practice.

Comparing the RLS and MMSE updates at time k assuming $\mathbf{W}_{\mathrm{RLS}}[k-1] = \mathbf{W}[k-1]$, it is seen that the MMSE filter $\mathbf{W}[k]$ is related to the RLS filter $\mathbf{W}_{\mathrm{RLS}}[k]$ via a scaling and an offset proportional to $\mathbf{C}^{-1}$. Consider the affine matrix space $\mathbf{C}^{-1} + \gamma(\mathbf{W}_{\mathrm{RLS}}[k] - \mathbf{C}^{-1})$. Obviously, $\mathbf{W}_{\mathrm{RLS}}[k]$ is obtained with $\gamma = 1$. In contrast, the MMSE filter is obtained with $\gamma = k/(k-1) > 1$ (stationary case) or $\gamma = 1/\lambda > 1$ (nonstationary case). Thus, when moving on the

[3] To simplify the discussion, we slightly abuse notation and use the same symbols for the stationary and the nonstationary case.

affine space connecting $\mathbf{C}^{-1}$ and $\mathbf{W}_{\mathrm{RLS}}[k]$ the MMSE filter lies beyond $\mathbf{W}_{\mathrm{RLS}}[k]$. These slightly larger "hops" are responsible for the improved convergence and performance of our MMSE approach as compared to the RLS approach (see the simulations section). In part, the inferior performance of the RLS approach can be attributed to the fact that one is forced to replace $\mathbf{h}[k]$ in (27) with $\hat{\mathbf{h}}_{\mathrm{LS}}[k]$. We again emphasize that the block-recursive algorithm (27) attempts to track the optimum channel estimation filter and is different from symbol-recursive Kalman and RLS algorithms tracking directly the channel (e.g. [7, 8]).

6. APPLICATION EXAMPLES

We next illustrate how the above general developments can be applied to specific communication schemes.

Flat-Fading MIMO Systems. Consider a block-fading narrowband MIMO system with M_t transmit and M_r receive antennas. We assume that within each block L transmit vectors $\mathbf{s}_1, \ldots, \mathbf{s}_L$ are dedicated to training. The training data transmission within the kth block can be modeled as

$$\mathbf{Y}[k] \;=\; \mathbf{S}_0 \mathbf{H}[k] + \mathbf{N}[k]\,,$$

with the $L \times M_t$ training matrix $\mathbf{S}_0 = [\mathbf{s}_1 \ldots \mathbf{s}_L]^T$, and the $M_t \times M_r$ channel matrix $\mathbf{H}[k]$. $\mathbf{Y}[k]$ and $\mathbf{N}[k]$ denote the $L \times M_r$ receive and noise matrix, respectively. Our generic model (1) is obtained with[4] $\mathbf{y}[k] = \mathrm{vec}\{\mathbf{Y}[k]\}$, $\mathbf{S} = \mathbf{I}_{M_t} \otimes \mathbf{S}_0$, $\mathbf{h}[k] = \mathrm{vec}\{\mathbf{H}[k]\}$, and $\mathbf{n}[k] = \mathrm{vec}\{\mathbf{N}[k]\}$. We thus have $N = LM_r$ and $M = M_t M_r$, i.e., $N > M$ amounts to $L > M_t$.

Multiuser Uplink. Let L users synchronously communicate training sequences $\mathbf{s}_l$ $(l = 1, \ldots, L)$ of length $N - M_c - 1$ over frequency selective channels $\mathbf{h}_l[k]$ to a base station. The channel impulse response vectors $\mathbf{h}_l[k]$ have length M_c. It is straightforward to show that the length-N receive signal $\mathbf{y}[k]$ induced by the training data is given by

$$\mathbf{y}[k] = \sum_{l=1}^{L} \mathbf{S}_l \mathbf{h}_l[k] + \mathbf{n}[k]\,,$$

where $\mathbf{S}_l$ are $N \times M_c$ Toeplitz matrices induced by the training sequences $\mathbf{s}_l$. This expression can be rewritten as in (1) by setting $\mathbf{h}[k] = \left[\mathbf{h}_1^T[k] \ldots \mathbf{h}_L^T[k]\right]^T$ and $\mathbf{S} = [\mathbf{S}_1 \ldots \mathbf{S}_L]$. In this case, $M = LM_c$ and the condition $N \geq M$ requires that the training sequence length is larger than the total number of channel coefficients. Note that single user systems are a simple special case obtained with $L = 1$.

Pilot-Symbol-Assisted OFDM. Let us consider OFDM transmission over a doubly selective fading channel. Each block consists of L_t OFDM symbols and L_f subcarriers in order to transmit symbols $s_{m,l}[k]$ $(m = 1, \ldots, L_t,\ l = 1, \ldots, L_f)$. The received OFDM demodulator sequence is $y_{m,l}[k] = \tilde{H}_{m,l}[k]\, s_{m,l}[k] + n_{m,l}[k]$, where $n_{m,l}[k]$ is additive Gaussian noise and $\tilde{H}_{m,l}[k]$ denotes the channel coefficients. Assuming a maximum delay of M_τ samples and a maximum (normalized) Doppler of $M_\nu/(L_t L_f)$, there is

$$\tilde{H}_{m,l}[k] = \sum_{\tau=-M_\tau}^{M_\tau-1} \sum_{\nu=-M_\nu}^{M_\nu-1} H_{\tau,\nu}[k]\, e^{j2\pi\left(\frac{\tau l}{L_f} - \frac{\nu m}{L_t}\right)}$$

[4]The $\mathrm{vec}\{\cdot\}$ operation amounts to stacking the columns of a matrix into a vector and $\otimes$ denotes the Kronecker product [9].

We note that $M = 4M_\tau M_\nu$ is proportional to the channel's spreading factor [10] and that any other sparse basis expansion model for $\tilde{H}_{m,l}[k]$ could be used instead of the Fourier expansion.

Within each block, $N = N_t N_f$ pilot symbols are transmitted on the regular rectangular lattice (mP_t, lP_f), with $N_t = L_t/P_t$ and $N_f = L_f/P_f$ pilot positions in time and frequency, respectively. The pilot induced receive sequence

$$\mathbf{y}[k] = \mathrm{vec}\left\{ \begin{pmatrix} y_{P_1,P_2}[k] & \cdots & y_{N_t P_1, P_2}[k] \\ \vdots & \ddots & \vdots \\ y_{P_1, N_f P_2}[k] & \cdots & y_{N_t P_1, N_f P_2}[k] \end{pmatrix} \right\}$$

can then be shown to be given by (1) with

$$\mathbf{h}[k] = \mathrm{vec}\left\{ \begin{pmatrix} H_{-M_\tau,-M_\nu}[k] & \cdots & H_{M_\tau-1,-M_\nu-1}[k] \\ \vdots & \ddots & \vdots \\ H_{-M_\tau,M_\nu}[k] & \cdots & H_{M_\tau-1,M_\nu-1}[k] \end{pmatrix} \right\}.$$

and $\mathbf{S} = \mathrm{diag}\{\mathbf{s}\}\left[\mathbf{F}_{L_t \times M_\nu}^{(P_t)} \otimes \mathbf{F}_{L_f \times M_\tau}^{(P_f)H}\right]$. Here, the pilot symbol vector $\mathbf{s}$ is defined analogously to $\mathbf{y}[k]$ and $\mathbf{F}_{a \times b}^{(c)}$ is obtained by picking the first and last b columns and every cth row from the $a \times a$ DFT matrix.

7. SIMULATION RESULTS

In this section, we illustrate the performance of the proposed method via numerical simulation examples.

Stationary MIMO System. First consider estimation of a 4×4 block-fading stationary MIMO channel with spatial correlation characterized by the Kronecker correlation matrix $\mathbf{R}_{\mathbf{h}} = \mathbf{R}_t \otimes \mathbf{R}_r$ with transmit and receive correlation $(\mathbf{R}_t)_{ij} = (\mathbf{R}_r)_{ij} = \rho^{|i-j|/2}$. We simulated the stationary version (i.e., no forgetting) of our block-recursive estimator assuming perfect knowledge of the noise variance. An orthogonal BPSK training matrix with $L = 8$ was used and the SNR was set to 7 dB. Fig. 1(a) shows the performance (normalized MSE) of our block recursive estimator and the LS estimator versus block index k for various ρ (the first 16 blocks required to obtain a non-singular $\hat{\mathbf{R}}_{\mathbf{x}}$ are not shown). After a short convergence phase (about $2M_t M_r = 32$ blocks), our estimator has learned the channel correlation reliably enough to outperform the LS estimator. The performance gains are at least ≈ 0.8 dB (for $\rho = 0$) and can be as large as ≈ 5.5 dB (for $\rho = 0.9$). It is furthermore seen that the performance of our recursive estimator quickly approaches the theoretical MMSE (achieved with the true channel statistics).

Nonstationary OFDM System. Our second example pertains to an OFDM system with $L_f = 64$ subcarriers and a block length of $L_t = 16$ OFDM symbols. The $N = 16$ QPSK pilots per block are $P_f = 16$ subcarriers and $P_t = 4$ OFDM symbols apart and suffice to estimate a doubly selective channel with $M_\nu = M_\tau = 4$. In contrast to the prevailing WSSUS assumption [11], we assume that the individual multipath components $H_{\tau,\nu}[k]$ are correlated [12]. Furthermore, the path loss changes such that the receive SNR gradually decreases by ≈ 8 dB. Fig. 1(b) shows the normalized MSE versus block index k achieved with the proposed method using forgetting factor $\lambda = 0.99$ and conventional initialization (labeled MMSE_0) and robust initialization (labeled MMSE_1). The first $M = 16$ blocks (not shown) are

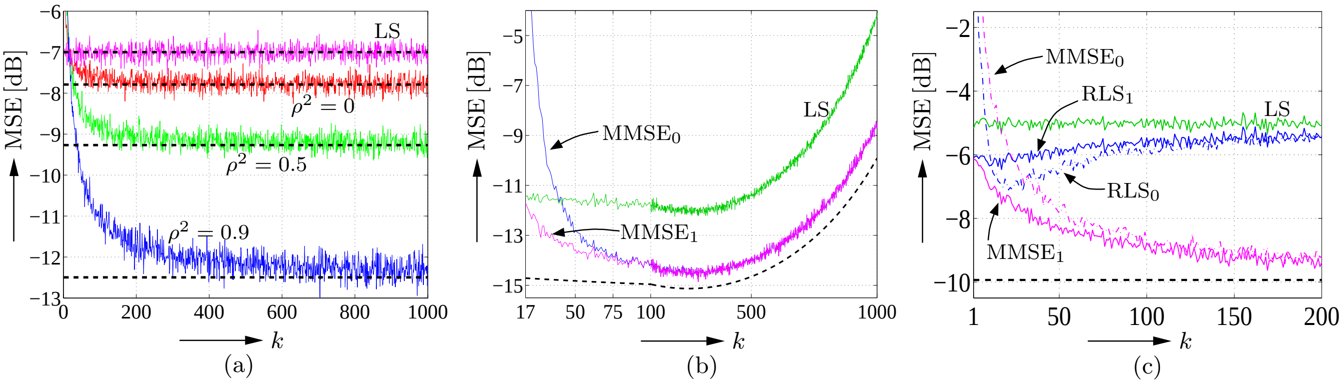

Figure 1: *Normalized MSE versus block index for (a) a 4×4 flat fading MIMO system with different correlation parameters ρ, (b) an OFDM system transmitting over a nonstationary doubly selective channel, (c) a 4×4 MIMO system with $\rho = 0.8$. The MSE achieved by the LS estimator (labeled 'LS') and the theoretical MSE (dashed line) are always shown for comparison.*

used for initialization and for estimation of the noise variance. The remaining part of the initial 100 blocks is shown on a finer time-scale for better visibility of the convergence behavior. Note that the theoretical MMSE (dashed line) increases by about 4 dB due to the decrease in receive SNR. Again, the performance of our method is close to optimal performance (i.e., we succeed in tracking the nonstationary statistics) and in general much better (up to 5 dB) than that of LS estimation. During the initial convergence phase, robust correlation learning is seen to outperform conventional learning which is even inferior to LS estimation (due to inversion of unreliable correlation estimates).

Comparison with RLS. Finally, Fig. 1(c) shows a comparison of our block-recursive MMSE estimator with RLS for a stationary 4×4 correlated MIMO channel with $\rho = 0.8$ (see above), again using conventional and robust initialization (indicated by subscripts $_0$ and $_1$, respectively). The SNR in this case was 5 dB. For conventional initialization it is seen that RLS initially converges faster than our MMSE estimator but degrades rapidly to LS performance for $k > 20$. In contrast, the MMSE estimator gradually converges to the theoretical MMSE. For robust initialization, both RLS and MMSE start out with significantly smaller MSE than with conventional initialization. Again, the MMSE estimator approaches optimal performance while RLS performance converges towards LS performance which is inferior to our method by almost 5 dB.

8. CONCLUSIONS

We introduced a block-recursive training-based scheme for wireless channel estimation that uses structured learning of the long-term channel properties (statistics). Using a robust initialization, our scheme allows for quick acquisition and tracking of channel statistics, leading to a performance which is close to the theoretical MMSE. We provided an insightful interpretation of our method by comparing it with an (inferior) RLS approach. The complexity of our algorithm scales at most with the square of the number of channel coefficients. Our algorithm attempts to learn the full channel correlation matrix and is thus particularly suited to MIMO channels with spatial correlation and to non-WSSUS channels [12] with correlated delay-Doppler components.

REFERENCES

[1] L. L. Scharf, *Statistical Signal Processing*. Reading (MA): Addison Wesley, 1991.

[2] F. Dietrich and W. Utschick, "Pilot-assisted channel estimation based on second-order statistics," *IEEE Trans. Signal Processing*, vol. 53, pp. 1178–1193, March 2005.

[3] N. Czink, G. Matz, D. Seethaler, and F. Hlawatsch, "Improved MMSE estimation of correlated MIMO channels using a structured correlation estimator," in *Proc. IEEE SPAWC-05*, (New York), June 2005.

[4] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Comm.*, vol. 46, pp. 902–915, July 1998.

[5] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, March 1985.

[6] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs (NJ): Prentice Hall, 1991.

[7] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1065–1076, May 2002.

[8] D. Schafhuber, G. Matz, and F. Hlawatsch, "Kalman tracking of time-varying channels in wireless MIMO-OFDM systems," in *Proc. 37th Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove, CA), pp. 1261–1265, Nov. 2003.

[9] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits and Systems*, vol. CAS-25, pp. 772–781, Sept. 1978.

[10] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 3rd ed., 1995.

[11] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Comm. Syst.*, vol. 11, pp. 360–393, 1963.

[12] G. Matz, "Characterization of non-WSSUS fading dispersive channels," in *Proc. IEEE ICC-2003*, (Anchorage, AK), pp. 2480–2484, May 2003.