BAYESIAN CHANNEL ESTIMATION FOR DOUBLY CORRELATED MIMO SYSTEMS

Ansgar Scherb and Karl-Dirk Kammeyer

Department of Communications Engineering, University of Bremen
Otto-Hahn-Allee, D-28359 Bremen, Germany
{scherb, kammeyer}@ant.uni-bremen.de

ABSTRACT

In this paper we derive a Bayesian estimator for doubly correlated MIMO channels. The Bayesian estimator has clearly superior normalized mean squared error performance compared to parametric approaches especially when the channel is strongly correlated. However, since the computational costs may exceed practical limits we present a class of fix point algorithms significantly reducing the numerical effort. The convergence behavior of the fixed point algorithm is examined and a regularization is proposed in order to guarantee general convergence.

1. INTRODUCTION

In the last two decades the research on multiple-input multiple-output (MIMO) systems has attracted a lot of attention due to its high channel capacity at comparatively low bandwith consumption [1]. However, the issue of channel estimation (CE) has emerged as one of the bottlenecks in coherent MIMO transmission, since the amount of pilot symbols is proportionally increasing with the number of transmit antennas. Thus, especially in rapidly varying environments, where the channel impulse response has to be tracked in short periods, a large fraction of the available bandwidth has to be occupied by pilot symbols. From the information theoretic point of view the capacity degradation by reason of pilot assisted channel estimation was investigated in [2–4]. The impact of inaccurate channel state information (CSI) may become even worse when applying space time signal processing as e.g. beamforming [5] or Alamouti [6, 7].

In order to refine the accuracy of channel estimates without increasing the training overhead, numerous semiblind CE methods have been proposed, e.g. [8]. Also noncoherent space time coding has been introduced in [9, 10], where the channel need not to be explicitly identified. However, most of these approaches require a high computational effort or suffer from a performance degradation and, thus, are not capable for practical applications.

These arguments motivate the need of more advanced techniques for pilot assisted CE. In [11, 12] the issue of optimum pilot constellations in terms of a minimum mean squared error (MSE) is addressed. Another source of improvement might be found by exploring the long term properties of time varying channels. In [13] it had been observed that the channel covariance matrix is very slowly changing in time. Thus, the covariance matrix may be tracked over a long time and treated as a-priori knowledge for the current CE. If the covariance matrix is rank deficient, the channel is constraint on a certain subspace. Performance improvements as well as computational simplification can be achieved by matching the channel onto this subspace as shown in [14]. The gain achieved by incorporating covariance knowledge generally depends on the strength of channel correlations. Hence, on the one hand correlations are cumbersome in terms of channel capacity, on the other hand channel estimation may benefit therefrom. In this paper we presume the Kronecker model which is a well verified assumption for spatially correlated MIMO channels [15]. In [16] on the basis of this model statistical pilot shaping is suggested. However, the computational costs of Bayesian CE are much higher for correlated than for non-correlated channels. In this paper a class of fixpoint algorithms is presented having low complexity but delivering estimates close to the maximum a-posteriori probability bound.

The outline of this paper is as follows. In Section 2 we introduce the system model. The basics of parametric and Bayesian channel estimation are presented in Section 3. Subsequently, a class of fixpoint algorithms is proposed in Section 4, and numerical results are shown in Section 5. Finally, the paper is concluded in Section 6.

2. SYSTEM MODEL

Throughout the paper boldface lowercase letter denote column vectors, boldface uppercase letters denote matrices and $\mathbf{I}_M$ is the identity matrix of size $M \times M$. The superscript "T" stands for transpose and "H" for Hermitian (conjugate transpose). vec $(\mathbf{A})$ stacks the columns of $\mathbf{A}$ in a column vector, $\otimes$ is the Kronecker product operator. $\|\mathbf{A}\|_F$ denotes the Frobenius norm and $\|\mathbf{A}\|_2$ the L2-norm of a matrix $\mathbf{A}$.

We consider a MIMO communication link with M receive and N transmit antennas. Let $\mathbf{S} \in \mathbb{C}^{N\times K}$ be the matrix of transmitted pilot symbols of length $K \geq N$ with $\mathrm{E}\{[\mathbf{S}]^2_{n,k}\} = 1$ is the mean power per symbol. Without loss of generality we consider an orthogonal pilot design, i.e. $\mathbf{SS}^H = K\mathbf{I}_N$ (Orthogonal pilot design has been shown to be optimum for uncorrelated MIMO channels [11, 12]. In fact this is not necessarily true for correlated MIMO channels.). The received signal is given by

$$\mathbf{R} = \mathbf{HS} + \mathbf{V}, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{M\times N}$ is the channel matrix whose entries $[\mathbf{H}]_{m,n}$ characterize the transmission path from the n-th transmit antenna to the m-th receive antenna and $\mathbf{V} \in \mathbb{C}^{M\times K}$ is additive white identically independently distributed (i.i.d.) noise of mean power $E\{|[V]_{m,k}|^2\} = \sigma_v^2$. The channel coefficients are assumed to be random variables characterized by

$$\mathbf{H} = \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}}\mathbf{H}_{\mathrm{W}}\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \tag{2}$$

where $[\mathbf{H}_{\mathrm{W}}]_{m,n} \in \mathcal{CN}(0,1)\ \forall\, m = 1,\cdots,M;\ n = 1,\cdots,N$ is i.i.d. and $\mathbf{\Phi}_{\mathrm{R}}$ and $\mathbf{\Phi}_{\mathrm{T}}$ denote the spatial correlation matrix at receiver and transmitter, respectively.

In order to derive the Bayesian channel estimator we rewrite Eq. (1) as

$$\mathbf{r} = \bar{\mathbf{S}}\mathbf{h} + \mathbf{v}, \tag{3}$$

where $\mathbf{r} = \mathrm{vec}\,(\mathbf{R})$, $\mathbf{v} = \mathrm{vec}\,(\mathbf{V})$, $\mathbf{h} = \mathrm{vec}\,(\mathbf{H})$ and $\bar{\mathbf{S}} = \mathbf{S}^T \otimes \mathbf{I}_M$ with "$\otimes$" denoting the Kronecker product[1]. The probability density function (PDF) of the channel is given by

$$\begin{aligned} p(\mathbf{h}) &= \frac{\exp\left(-\mathbf{h}^H(\mathbf{\Phi}_{\mathrm{T}}^{-T} \otimes \mathbf{\Phi}_{\mathrm{R}}^{-1})\mathbf{h}\right)}{\pi^{MN}(\det \mathbf{\Phi}_{\mathrm{T}})^M(\det \mathbf{\Phi}_{\mathrm{R}})^N} \\ &= \frac{\exp\left(-\|\mathbf{\Phi}_{\mathrm{R}}^{-\frac{1}{2}}\mathbf{H}\mathbf{\Phi}_{\mathrm{T}}^{-\frac{1}{2}}\|_F^2\right)}{\pi^{MN}(\det \mathbf{\Phi}_{\mathrm{T}})^M(\det \mathbf{\Phi}_{\mathrm{R}})^N} \end{aligned} \tag{4}$$

and the likelihood to obtain $\mathbf{r}$ under the condition that $\bar{\mathbf{S}}$ was transmitted over the channel $\mathbf{h}$ is given by

$$p(\mathbf{r}|\mathbf{h},\bar{\mathbf{S}}) = \frac{\exp\left(-(\mathbf{r}-\bar{\mathbf{S}}\mathbf{h})^H(\mathbf{r}-\bar{\mathbf{S}}\mathbf{h})/\sigma_v^2\right)}{(\pi\sigma_v^2)^{MK}} \tag{5}$$

3. PARAMETRIC AND BAYESIAN CHANNEL ESTIMATION

In common literature (e.g. [17]) estimation methods are classified into *parametric* and *Bayesian* approaches. A standard parametric approach is the best linear unbiased estimator (BLUE) which is often referred to as least squares channel estimation. It is readily known that BLUE satisfies the maximum likelihood (ML) criterion, i.e.

$$\hat{\mathbf{h}}_{\mathrm{BLUE}} = \arg\max_{\mathbf{h}} p(\mathbf{r}|\mathbf{h},\bar{\mathbf{S}}) = \frac{1}{K}\bar{\mathbf{S}}^H\mathbf{r} \tag{6}$$

which is equivalent to

$$\hat{\mathbf{H}}_{\mathrm{BLUE}} = \frac{1}{K}\mathbf{RS}^H. \tag{7}$$

[1]Note that $\mathrm{vec}\,(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \cdot \mathrm{vec}\,(\mathbf{B})$.

In contrast to parametric methods the Bayesian approach treats the desired parameters as random variable with a-priori known statistics. Clearly, the a-priori PDF of the channel $p(\mathbf{h})$ is assumed to be perfectly known at the receiver. Thus, we may find a Bayesian estimator by incorporating $p(\mathbf{h})$ in (6). The obtained estimator is optimum in the sense of the maximum a-posteriori (MAP) criterion. The Bayesian approach can be expressed by

$$\begin{aligned} \hat{\mathbf{h}}_{\mathrm{Bay}} &= \arg\max_{\mathbf{h}} p(\mathbf{h}|\mathbf{r},\bar{\mathbf{S}}) \\ &= \arg\max_{\mathbf{h}} p(\mathbf{h})p(\mathbf{r}|\mathbf{h},\bar{\mathbf{S}}) \\ &= \underbrace{(K\mathbf{I}_{MN} + \sigma_v^2(\mathbf{\Phi}_{\mathrm{T}}^{-T} \otimes \mathbf{\Phi}_{\mathrm{R}}^{-1}))}_{\mathbf{\Theta}}{}^{-1}\bar{\mathbf{S}}^H\mathbf{r}. \end{aligned} \tag{8}$$

Note that $\mathbf{\Theta}$ has only Kronecker structure (i.e. can be expressed in terms of $\mathbf{A}^T \otimes \mathbf{B}$) if either the transmit or the receive correlation matrix $\mathbf{\Phi}_{\mathrm{T}}$ and $\mathbf{\Phi}_{\mathrm{R}}$, respectively, is an identity matrix. Otherwise Eq. (8) cannot be converted into a compact matrix form similar to (7). In fact, due to matrix inversion of dimension $MN \times MN$ the computational costs of the Bayesian approach for doubly correlated MIMO channels are of order $\mathcal{O}((MN)^3)$ which significantly exceeds the numerical effort required for BLUE.

4. FIX POINT ALGORITHM

In this section we derive a class of fixpoint algorithms for Bayesian channel estimation. First, the non-regularized fixpoint algorithm (NRFA) is presented in Section 4.1. Although NRFA is not reasonable for practical application due to a lack of convergence, it is the basis for the fixpoint approaches presented in the subsequent sections. In order to guarantee general convergence we introduce the concept of regularization. The "linearly regularized fixpoint algorithm" (LRFA) is presented in section Section 4.2 and an alternative approach named as "projectively regularized fixpoint algorithm" (PRFA) is presented in Section 4.3. Since the two regularized approaches LRFA and PRFA provide opposite benefits, in Section 4.4 a combination termed as "linearly and projectively regularized fixpoint algorithm" (LPRFA) is proposed.

4.1. NRFA

Left multiplying (8) by $\mathbf{\Theta}/K$ we obtain

$$\begin{aligned} &\hat{\mathbf{h}}_{\mathrm{Bay}} + \frac{\sigma_v^2}{K}(\mathbf{\Phi}_{\mathrm{T}}^{-T} \otimes \mathbf{\Phi}_{\mathrm{R}}^{-1})\hat{\mathbf{h}}_{\mathrm{Bay}} = \frac{1}{K}\bar{\mathbf{S}}^H\mathbf{r} \\ \Rightarrow &\hat{\mathbf{h}}_{\mathrm{Bay}} = \frac{1}{K}\bar{\mathbf{S}}^H\mathbf{r} - \frac{\sigma_v^2}{K}(\mathbf{\Phi}_{\mathrm{T}}^{-T} \otimes \mathbf{\Phi}_{\mathrm{R}}^{-1})\hat{\mathbf{h}}_{\mathrm{Bay}}. \end{aligned} \tag{9}$$

In contrast to (8) this expression can be converted into compact matrix form by

$$\begin{aligned}\hat{\mathbf{H}}_{\text{Bay}} &= \frac{1}{K}\mathbf{R}\mathbf{S}^H - \frac{\sigma_v^2}{K}\mathbf{\Phi}_{\text{R}}^{-1}\hat{\mathbf{H}}_{\text{Bay}}\mathbf{\Phi}_{\text{T}}^{-1} \\ &= \underbrace{\hat{\mathbf{H}}_{\text{BLUE}} - \frac{\sigma_v^2}{K}\mathbf{\Phi}_{\text{R}}^{-1}\hat{\mathbf{H}}_{\text{Bay}}\mathbf{\Phi}_{\text{T}}^{-1}}_{f(\hat{\mathbf{H}}_{\text{Bay}})}, \end{aligned} \quad (10)$$

where the r.h.s. can be expressed as function of the desired channel impulse response $\hat{\mathbf{H}}$. Apparently, the MAP-solution $\hat{\mathbf{H}}_{\text{Bay}}$ is a fixpoint of this function (i.e. a point that is mapped onto itself by the function). On the basis of this expression we deduce the fix point Algorithm 1. The numerical costs of

Algorithmus 1 Fix Point Algorithm

initialize $\hat{\mathbf{H}}_0 = \mathbf{0}$ and $i = 0$
repeat
  calculate $\hat{\mathbf{H}}_{i+1} = f(\hat{\mathbf{H}}_i)$
  set $i = i + 1$
until $\|\hat{\mathbf{H}}_i - \hat{\mathbf{H}}_{i-1}\|_F^2 <$ threshold

one iteration are given by $\mathcal{O}(N^3 + M^3)$ due to the inversion of $\mathbf{\Phi}_{\text{R}}$ and $\mathbf{\Phi}_{\text{T}}$. Thus, compared to (8) even with a moderate number of iterations the computational complexity is significantly reduced. Note that the result of the first iteration $\hat{\mathbf{H}}_1$ is identical to the BLUE solution. In order to explore the convergence behavior of NRFA, we recapitulate some common results of functional analysis [18]:

Definition: Let $f : X \to Y$ be a function defined on a metric space. Then this function is said to be Lipschitz-continuous in the subset $Z \subseteq X$ if there exist a nonnegative real number L such that for each $a, b \in Z$ the condition

$$\|f(a) - f(b)\|_Y \leq L\|a - b\|_X \quad (11)$$

holds. The smallest value L satisfying (11) is called Lipschitz-constant.

Banachs fixpoint theorem: A function $f : X \to X$ has exactly one fixpoint $f(x_\infty) = x_\infty$ if it is Lipschitz continuous in X and the corresponding Lipschitz constant is $L < 1$.

In fact, the value L is also an indicator for the convergence behavior of fixpoint algorithms. By replacing $f(x_{n-1}) = x_n$ and $f(x_\infty) = x_\infty$ in (11) it can easily be seen that the progress per iteration is lower bounded by

$$\|x_{n-1} - x_\infty\| \leq L\|x_n - x_\infty\|. \quad (12)$$

As a consequence, general convergence can be guaranteed for $L < 1$.

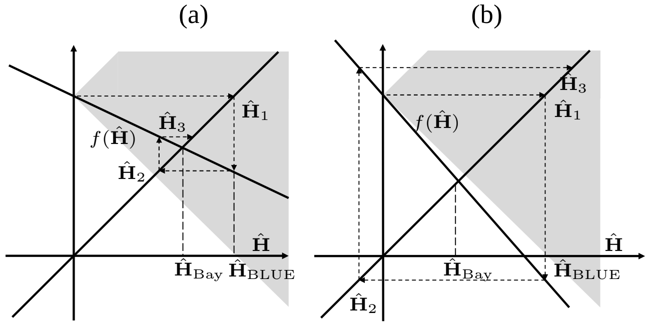

Fig. 1. Convergence behavior of NRFA at (a) $L_{\text{NRFA}} < 1$ und (b) $L_{\text{NRFA}} > 1$.

The Lipschitz constant of $f(\hat{\mathbf{H}})$ can be calculated by

$$\begin{aligned} L_{\text{NRFA}} &= \max_{\hat{\mathbf{H}}_a, \hat{\mathbf{H}}_b} \frac{\|f(\hat{\mathbf{H}}_a) - f(\hat{\mathbf{H}}_b)\|_2}{\|\hat{\mathbf{H}}_a - \hat{\mathbf{H}}_b\|_2} \\ &= \max_{\hat{\mathbf{H}}_a, \hat{\mathbf{H}}_b} \frac{\sigma_v^2}{K} \frac{\|\mathbf{\Phi}_{\text{R}}^{-1}(\hat{\mathbf{H}}_a - \hat{\mathbf{H}}_b)\mathbf{\Phi}_{\text{T}}^{-1}\|_2}{\|\hat{\mathbf{H}}_a - \hat{\mathbf{H}}_b\|_2} \\ &= \frac{\sigma_v^2}{K\lambda_{\min,\text{T}}\lambda_{\min,\text{R}}}, \end{aligned} \quad (13)$$

where $\lambda_{\min,\text{T}}$ and $\lambda_{\min,\text{R}}$ are the minimum eigenvalues of $\mathbf{\Phi}_{\text{T}}$ and $\mathbf{\Phi}_{\text{R}}$, respectively. Apparently, the convergence behavior of NRFA becomes crucial at low SNR and when the channel is strongly correlated. As illustrated in Fig. 1 convergence can be guaranteed as long as $f_{\text{NRFA}}(\hat{\mathbf{H}})$ is within the gray colored region, whereas otherwise NRFA diverges. Note that the intersection of $f_{\text{NRFA}}(\hat{\mathbf{H}})$ and the y-axis is always at $f(\mathbf{0}) = \hat{\mathbf{H}}_{\text{BLUE}}$ independently from L_{NRFA}.

4.2. LRFA

The main idea behind regularization is to replace $f(\hat{\mathbf{H}})$ in (10) by a function which is as close as possible to the original function while its Lipschitz constant is bounded by a predefined value $L_0 < 1$. Such a linear regularized function is given by

$$f_{\text{LPRA}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}_{\text{BLUE}} - a\frac{\sigma_v^2}{K}\mathbf{\Phi}_{\text{R}}^{-1}\hat{\mathbf{H}}\mathbf{\Phi}_{\text{T}}^{-1} \quad (14)$$

where

$$a = \min\left(1, \frac{L_0 K}{\sigma_v^2}\lambda_{\min,\text{T}}\lambda_{\min,\text{R}}\right) \quad (15)$$

is a reel nonnegative parameter ensuring $L_{\text{LRFA}} \leq L_0$. Note, that while the regularization is active, i.e. $a < 1$, the fix point $\hat{\mathbf{H}}_{\text{LRFA}} = f(\hat{\mathbf{H}}_{\text{LRFA}})$ is shifted by a small fraction from the desired value $\hat{\mathbf{H}}_{\text{Bay}}$ towards $\hat{\mathbf{H}}_{\text{BLUE}}$. An illustration of this effect is shown in Fig. 2. Thus, by adjusting L_0 properly we may trade off between performance and convergence speed. However, the worst case parameter $a = 0$ still delivers the BLUE solution. A drawback of LRFA is that if either the transmit or the receive correlation matrix is rank deficient, i.e.

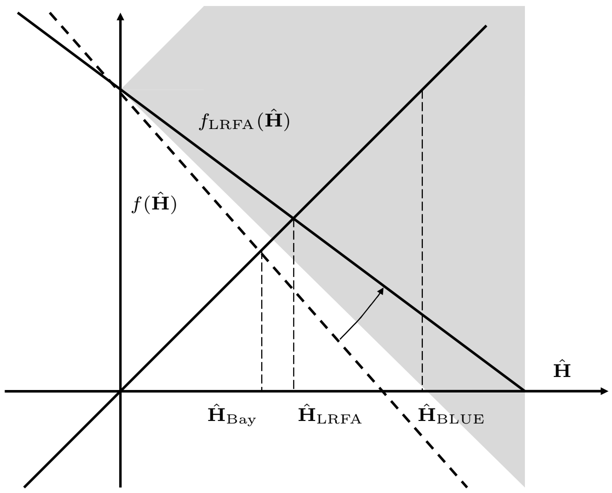

Fig. 2. Convergence behavior of linear regularized fixpoint algorithm.

$\lambda_{\min,\mathrm{T}} = 0$ or $\lambda_{\min,\mathrm{R}} = 0$, due to $a = 0$ it always delivers the BLUE solution.

4.3. PRFA

In contrast to the previously discussed LRFA this method is also able to deal with rank deficient correlation matrices in a smart way. Implying the singular value decompositions

$$\mathbf{\Phi}_{\mathrm{T}} = \mathbf{U}\mathbf{\Lambda}_{\mathrm{T}}\mathbf{U}^H \text{ and } \mathbf{\Phi}_{\mathrm{R}} = \mathbf{V}\mathbf{\Lambda}_{\mathrm{R}}\mathbf{V}^H$$

the PDF of the channel given in (4) can be rewritten by

$$\begin{aligned} p(\mathbf{H}) &\propto \exp(-\|\mathbf{\Phi}_{\mathrm{R}}^{-\frac{1}{2}}\mathbf{H}\mathbf{\Phi}_{\mathrm{T}}^{-\frac{1}{2}}\|_F^2) \\ &= \exp(-\|\mathbf{V}\mathbf{\Lambda}_{\mathrm{R}}^{-\frac{1}{2}}\bar{\mathbf{H}}\mathbf{\Lambda}_{\mathrm{T}}^{-\frac{1}{2}}\mathbf{U}^H\|_F^2) \\ &= \exp(-\sum_{i,j}|\bar{h}_{i,j}|^2\lambda_{i,\mathrm{T}}^{-1}\lambda_{j,\mathrm{R}}^{-1}), \end{aligned} \tag{16}$$

where $\bar{\mathbf{H}} = \mathbf{V}^H\mathbf{H}\mathbf{U}$ is the rotated channel matrix whose entries $\bar{h}_{i,j}$ are independently distributed and subsequently called the (i,j)-th eigen-component of the channel. From (16) follows that the product of the eigenvalues $\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}}$ corresponds to the variance of $\bar{h}_{i,j}$. Hence, when $\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}}$ is very small the contribution of the corresponding eigen-component $\bar{h}_{i,j}$ can be neglected for channel estimation. Since even those eigen-components are responsible for the bad convergence behavior of NRFA, it seems to be evident to eliminate them from the fixpoint iteration. Therefore, we define the projection matrices

$$\mathbf{Q}_{I,\mathrm{T}} = \sum_{i=1}^{I}\mathbf{u}_i\mathbf{u}_i^H \quad \text{und} \quad \mathbf{Q}_{J,\mathrm{R}} = \sum_{j=1}^{J}\mathbf{v}_j\mathbf{v}_j^H$$

and the pseudo inverse of the rank reduced correlation matrices

$$\mathbf{\Phi}_{I,\mathrm{T}}^{\mathrm{inv}} = \sum_{i=1}^{I}\lambda_{i,\mathrm{T}}^{-1}\mathbf{u}_i\mathbf{u}_i^H \quad \text{und} \quad \mathbf{\Phi}_{J,\mathrm{R}}^{\mathrm{inv}} = \sum_{j=1}^{J}\lambda_{j,\mathrm{R}}^{-1}\mathbf{v}_j\mathbf{v}_j^H,$$

where $\mathbf{u}_i$ and $\mathbf{v}_j$ are the i-th column of $\mathbf{U}$ corresponding to the eigenvalue $\lambda_{i,\mathrm{T}}$ and, respectively, the j-th column of $\mathbf{V}$ corresponding to the eigenvalue $\lambda_{j,\mathrm{R}}$. Each eigen-component according to $(i,j) > (I,J)$ is removed from the instantaneous channel estimate by right and left multiplying (10) by $\mathbf{Q}_{I,\mathrm{T}}$ and $\mathbf{Q}_{J,\mathrm{R}}$. Thus, the PRFA is obtained by

$$\begin{aligned} f_{\mathrm{PRFA}}(\hat{\mathbf{H}}) &= \mathbf{Q}_{J,\mathrm{R}}(\hat{\mathbf{H}}_{\mathrm{BLUE}} - \frac{\sigma_v^2}{K}\mathbf{\Phi}_{\mathrm{R}}^{-1}\hat{\mathbf{H}}\mathbf{\Phi}_{\mathrm{T}}^{-1})\mathbf{Q}_{I,\mathrm{T}} \\ &= \mathbf{Q}_{J,\mathrm{R}}\hat{\mathbf{H}}_{\mathrm{BLUE}}\mathbf{Q}_{I,\mathrm{T}} - \frac{\sigma_v^2}{K}\mathbf{\Phi}_{J,\mathrm{R}}^{\mathrm{inv}}\hat{\mathbf{H}}\mathbf{\Phi}_{I,\mathrm{T}}^{\mathrm{inv}}. \end{aligned} \tag{17}$$

The Lipschitz-constant of $f_{\mathrm{PRFA}}(\hat{\mathbf{H}})$ is given by

$$L_{\mathrm{PRFA}} = \frac{\sigma_v^2}{K\lambda_{I,\mathrm{T}},\lambda_{J,\mathrm{R}}} \leq \frac{\sigma_v^2}{K\lambda_{\min,\mathrm{T}},\lambda_{\min,\mathrm{R}}}. \tag{18}$$

In order to determine appropriate dimensions (I,J) we propose to solve following optimization problem: Choose (I,J) such that the energy (variance) of participating channel eigen-components in (17) becomes maximum under the condition $L_{\mathrm{PRFA}} < L_0$, i.e.

$$(I,J) = \arg\max_{\bar{I},\bar{J}} \sum_{i=1}^{\bar{I}}\sum_{j=1}^{\bar{J}}\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}} \quad \text{s.t.} \quad \lambda_{\bar{I},\mathrm{T}}\lambda_{\bar{J},\mathrm{R}} > \frac{\sigma_v^2}{KL_0}. \tag{19}$$

As it will be shown in section 5 PRFA is superior to LRFA when the channel is very strongly correlated. Conversely, PRFA is not robust in the sense that its worst case estimate is bounded by $\hat{\mathbf{H}}_{\mathrm{BLUE}}$.

4.4. LPRFA

Merging the regularization techniques applied in LRFA and PRFA we obtain the LPRFA which is defined by the function

$$f_{\mathrm{LPRFA}}(\hat{\mathbf{H}}) = \mathbf{Q}_{J,\mathrm{R}}\hat{\mathbf{H}}_{\mathrm{ML}}\mathbf{Q}_{I,\mathrm{T}} - a_{I,J}\frac{\sigma_v^2}{K}\mathbf{\Phi}_{J,\mathrm{R}}^{\mathrm{inv}}\hat{\mathbf{H}}\mathbf{\Phi}_{I,\mathrm{T}}^{\mathrm{inv}}. \tag{20}$$

The parameter

$$a_{I,J} = \min\left(1, \frac{L_0 K}{\sigma_v^2}\lambda_{I,\mathrm{T}}\lambda_{J,\mathrm{R}}\right) \tag{21}$$

is upper bounding the Lipschitz constant L_{LPRFA} to L_0. Appropriate values for (I,J) can be found by minimizing the mean squared error between true and estimated channel (see Appendix), i.e.

$$\begin{aligned} (I,J) &= \arg\min_{\bar{I},\bar{J}} \mathrm{E}\left(\|\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LPRFA}}(\bar{I},\bar{J})\|_F^2\right) \\ &= \arg\max_{\bar{I},\bar{J}} \sum_{i=1}^{\bar{I}}\sum_{j=1}^{\bar{J}} \frac{(\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}} - (1-2a_{\bar{I},\bar{J}})\frac{\sigma_v^2}{K})\lambda_{i,\mathrm{T}}^2\lambda_{j,\mathrm{R}}^2}{(\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}} + a_{\bar{I},\bar{J}}\frac{\sigma_v^2}{K})^2}. \end{aligned} \tag{22}$$

5. NUMERICAL RESULTS

In the simulations we modeled the transmit and receive correlation matrix by

$$[\mathbf{\Phi}_{\mathrm{T}}]_{m,n} = \rho_{\mathrm{T}}^{|m-n|} \quad \text{and} \quad [\mathbf{\Phi}_{\mathrm{R}}]_{m,n} = \rho_{\mathrm{R}}^{|m-n|}, \qquad (23)$$

respectively, where the coefficients $0 \leq \rho_{\mathrm{T}}, \rho_{\mathrm{R}} \leq 1$ determine the degree of correlation. All curves in Fig. 3 are plotted over the receive correlation in the range $0.5 \leq \rho_{\mathrm{R}} < 1$, whereas the transmit correlation was kept constant at $\rho_{\mathrm{T}} = 0.3$. Note that the correlation matrix $\mathbf{\Phi}_{\mathrm{R}}$ equals to identity if $\rho_{\mathrm{R}} = 0$ whereas it tends to be the all-ones matrix for ρ_{R} being close to one. In fact, the fully correlated case $\rho = 1$ is not covered within the simulations since then $\mathbf{\Phi}_{\mathrm{R}}$ is singular. Fig. 3 (a), (c) and (e) illustrate the influence of the correlation on the normalized mean squared error (NMSE) $\mathrm{E}\{\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2\}/\mathrm{E}\{\|\mathbf{H}\|_F^2\}$ of LRFA, PRFA and LPRFA, respectively, at 12 dB SNR. For the reason of comparability the NMSE curves of BLUE, the ideal Bayesian approach and NRFA are plotted in each of these figures. Obviously, the NMSE performance of BLUE is not affected by correlations, whereas especially in case of strong correlations the Bayesian estimator considerably benefits from the pre-knowledge on the transmit and receive correlation matrix. Due to the fact of being the MAP estimator the Bayesian estimator displays the lowest attainable NMSE-values and may, therefore, serve as reference. The NRFA works very well up to the point at approximately $\rho_{\mathrm{R}} = 0.8$, where the Lipschitz constant exceeds the admissible range. Beyond this threshold NRFA diverges. General convergence is attained by each of the regularized fixpoint algorithms. LRFA is robust in the sense that its NMSE performance is lower bounded by BLUE and it is monotonically increasing for growing ρ_{R}. Nevertheless, the gap between the ideal Baysian approach and LRFA becomes large at ρ_{R} close to one. Conversely, an abrupt performance degradation can be observed for PRFA whenever a channel eigencomponent is faded out, whereas PRFA clearly outperforms LRFA for very strong correlations. LPRFA shows excellent performance over the whole range. The effect of L_0 on the computational complexity in terms of the number of required iterations until convergence is illustrated in the Figures 3 (b), (d) and (f). Especially for LPRFA the performance penalty obtained by choosing L_0 small is minor, whereas the numerical costs are considerably reduced. For the considered 8×8 MIMO system they amounts to approximately 1% of those of the ideal Baysian approach.

6. CONCLUSIONS

In this paper we have presented a class of fixpoint algorithms estimating doubly correlated MIMO channels with regard to a-priori knowledge of the correlation matrices. Especially the approach termed as LPRFA has shown NMSE performance close to that of the ideal MAP solution at comparatively low computational complexity.

Appendix

The mean squared error (MSE) between true and estimated channel for LPRFA can be analytically calculated by

$$\begin{aligned}
&\mathrm{E}\left(\|\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LPRFA}}(I,J)\|_F^2\right) = \mathrm{E}\left(\|\mathbf{h} - \hat{\mathbf{h}}_{\mathrm{LPRFA}}(I,J)\|_2^2\right) \\
&= \underbrace{\mathrm{tr}\,\mathrm{E}\left(\mathbf{h}\mathbf{h}^H\right)}_{\mathrm{tr}\mathbf{\Phi} = MN} \\
&+ \mathrm{tr}\left(\mathbf{Q}_{I,J}\left(\mathbf{I}_{MN} + a_{I,J}\frac{\sigma_v^2}{K}\mathbf{\Phi}_{I,J}^{\mathrm{inv}}\right)^{-2} \underbrace{\mathrm{E}\left(\hat{\mathbf{h}}_{\mathrm{BLUE}}\hat{\mathbf{h}}_{\mathrm{BLUE}}^H\right)}_{\mathbf{\Phi} + \frac{\sigma_v^2}{K}\mathbf{I}_{MN}}\right) \\
&- 2\mathrm{tr}\,\Re\left\{\mathbf{Q}_{I,J}\left(\mathbf{I}_{MN} + a_{I,J}\frac{\sigma_v^2}{K}\mathbf{\Phi}_{I,J}^{\mathrm{inv}}\right)^{-1} \underbrace{\mathrm{E}\left(\mathbf{h}\hat{\mathbf{h}}_{\mathrm{BLUE}}^H\right)}_{\mathbf{\Phi}}\right\} \\
&= MN - \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}} - (1 - 2a_{I,J})\frac{\sigma_v^2}{K})\lambda_{i,\mathrm{T}}^2\lambda_{j,\mathrm{R}}^2}{(\lambda_{i,\mathrm{T}}\lambda_{j,\mathrm{R}} + a_{I,J}\frac{\sigma_v^2}{K})^2},
\end{aligned} \qquad (24)$$

where $\mathbf{Q}_{I,J} = \mathbf{Q}_{I,\mathrm{T}}^T \otimes \mathbf{Q}_{J,\mathrm{R}}$, $\mathbf{\Phi} = \mathbf{\Phi}_{\mathrm{T}}^T \otimes \mathbf{\Phi}_{\mathrm{R}}$ and $\mathbf{\Phi}_{I,J}^{\mathrm{inv}} = (\mathbf{\Phi}_{I,\mathrm{T}}^{\mathrm{inv}})^T \otimes \mathbf{\Phi}_{J,\mathrm{R}}^{\mathrm{inv}}$. Note that by setting $a_{I,J} = 1$ the MSE of PRFA, and by setting $I = N$ and $J = M$ the MSE of LRFA is obtained.

7. REFERENCES

[1] I. E.Telatar. *Capacity of Multi-antenna Gaussian Channels*. European Transactions on Telecommunications, Vol. 10, No. 6, pp. 585-595, November, 1999.

[2] B. Hassibi, and B. M. Hochwald. *How Much Training is Needed in Multiple-Antenna Wireless Links?*. IEEE Transactions on Information Theory, Vol. 49, No. 4, pp. 951-5963, April, 2003.

[3] H. Vikalo, B. Hassibi, B. M. Hochwald and T. Kailath. *On the Capacity of Frequency selective Channels in Training-Based Transmission Schemes*. IEEE Transactions on Signal Processing, Vol. 52, No. 9, pp. 951-5963, April, 2004.

[4] L. Zheng, and D. N. C. Tse. *Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel*. IEEE Transactions on Information, Vol. 48, No. 2, pp. 359-383, April, 2002.

[5] C. R. Murthy, A. K. Jagannatham, and B. D. Rao. *Training-Based and Semiblind Channel Estimation for MIMO Systems with Maximum Ratio Transmission*.

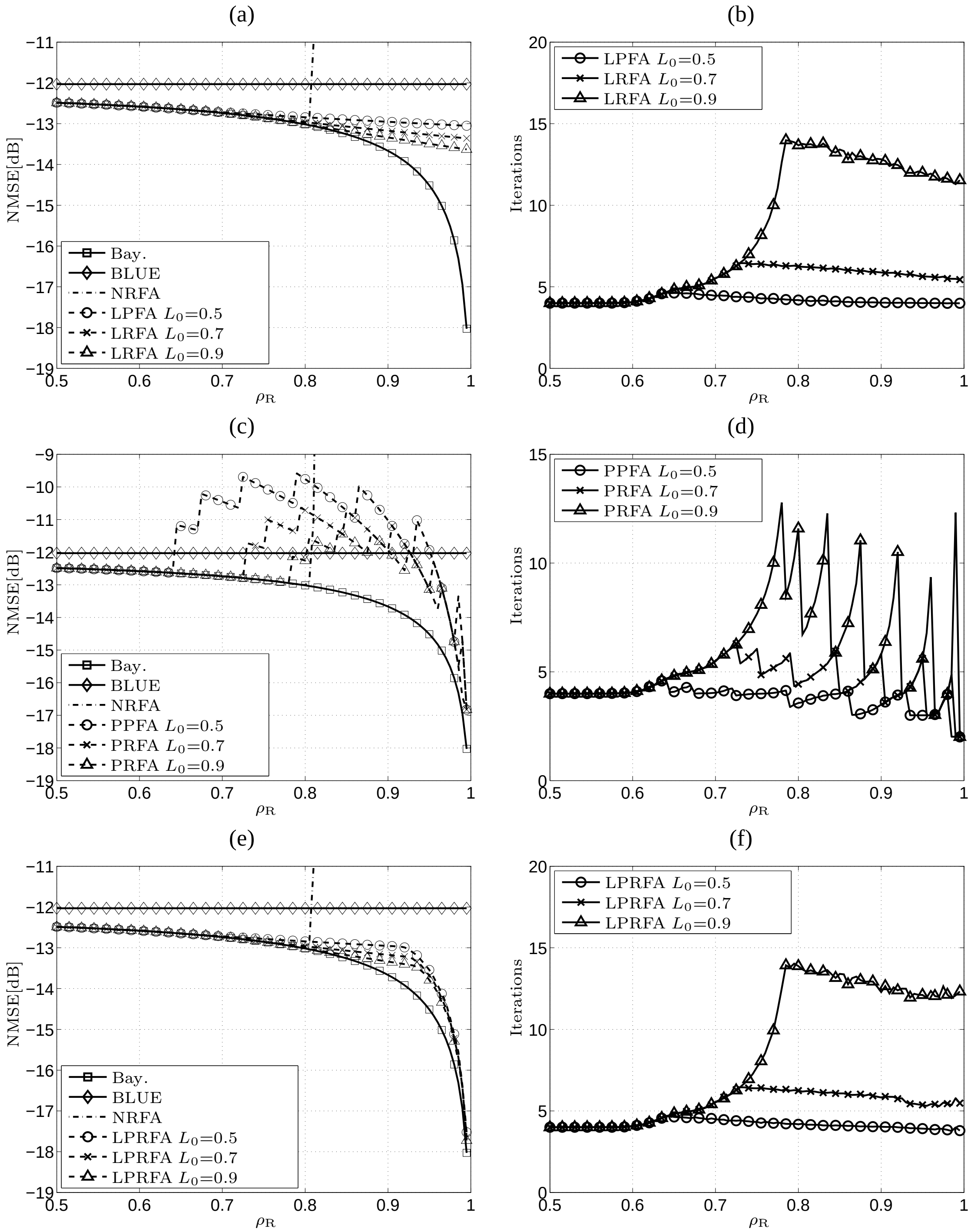

Fig. 3. NMSE vs. $\rho_{\rm R}$ of LRFA (a) PRFA (c) and LPRFA (e) for $(N = 8 \times M = 8)$ system with $K = 8$ pilot symbols at 12 dB SNR. Corresponding mean number of required iterations until convergence vs. $\rho_{\rm R}$ for LRFA (b) PRFA (d) and LPRFA (f).

IEEE Transaction on Signal Processing, Vol. 54, No. 7, pp. 2546 - 2558, July, 2006.

[6] R. Seeger, L. Broetje, and K. D. Kammeyer. *Investigation of Orthogonal Space-Time Block Codes using a MIMO Hardware Demonstrator*. Winter International Symposium on Information and Communication Technologies (WISICT 2004), Cancun, Mexiko, January, 2004

[7] L. Wenyu, and N. C. Beaulieu. *Effects of channel-estimation errors on receiver selection-combining schemes for Alamouti MIMO systems with BPSK*. IEEE Transactions on Communications, Vol. 54, No. 1, pp. 169 - 178, January, 2006.

[8] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong. *Signal Processing Advances in Wireless & Mobile Communications: Trends in Channel Estimation and Equalization, Vol. 1*. Prentice Hall PTR, Saddle River, NJ, 2000.

[9] M. B. Hochwald, and W. Sweldens. *Differential Unitary SpaceTime Modulation*. IEEE Transactions on Communications, Vol. 46, No. 2, pp. 543-564, March, 2000.

[10] B. M. Hochwald, and T. Marzetta. *Unitary space-time modulation for multiple-antenna communicationsin Rayleigh flat fading*. IEEE Transactions on Information Theory, Vol. 48, No. 12, pp. 2041-2052, December, 2000.

[11] X. Ma, L. Yang, and G. B. Giannakis. *Optimal Training for MIMO Frequency-Selective Fading Channels*. IEEE Transactions on Wireless Communications, Vol. 4, No. 2, pp. 453 - 466, May, 2005.

[12] M. Biguesh, and A. B. Gershman. *Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals*. IEEE Transactions on Signal Processing, Vol. 54, pp. 884-893, March, 2006.

[13] M. Paetzold. *Mobilfunkkanäle*. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, 1999.

[14] F. A. Dietrich and W. Utschick. *Pilot-Assisted Channel Estimation Based on Second-Order Statistics*. IEEE Transaction on Signal Processing, vol. 53(3), pp. 1178-1193, March, 2005.

[15] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen. *A Stochastic MIMO Radio Channel Model with Experimental Validation*. IEEE Jourtnal on Selected Areas of Communications, Vol. 20, pp. 1211-1226, August, 2001.

[16] M. Kiesling, J. Speidel,and Y. Chen. *MIMO Channel Estimation in Correlated Fading Environments*. IEEE Vehicular Technology Conference, October, 2003.

[17] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. PTR Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[18] V. I. Istratescu. *Fixed Point Theory, An Introduction*. D. Reidel, Netherlands, 1981.