Augmenting the Training Sequence Part in Semiblind Estimation for MIMO Channels

Abdelkader Medles, Dirk T.M. Slock
Eurecom Institute 2229 route des Crêtes, B.P. 193
06904 Sophia Antipolis Cedex
FRANCE
Email: {abdelkader.medles,dirk.slock}@eurecom.fr

Abstract— **The Multiple Input Multiple Output (MIMO) channel results from the use of Multiple Transmit and Multiple Receive antennas, which allows to achieve high spectral efficiency by spatial multiplexing. The high number of coefficients in the channel response (number of TX antennas by number of RX antennas by delay spread) allows to achieve high diversity and to improve the outage capacity, but at the same time represents a challenge for channel estimation as it imposes the use of a longer Training Sequence(TS) leading to a rate loss. In this paper, we augment the TS artificially by including the blind part (unknown symbols) information and the non pure training information, this allows to reduce the TS length needed for channel estimation and hence to save rate. We use semiblind approaches that exploit both training and blind information. These techniques have a complexity not immensely much higher than that of training based techniques. For the flat channel case, the technique we present achieves the Cramer-Rao Bound. In the frequency-selective channel case we use a quadratic semiblind criterion that combines a training based least-squares criterion with a blind criterion based on linear prediction.**

I. Introduction

Consider linear digital modulation over a linear channel with additive Gaussian noise. Assume that we have N_{tx} transmitters and N_{rx} receiving channels. If the channel is assumed to be flat, then the (symbol rate) sampled received signal at discrete time k can be written as:

$$y_i(k) \;=\; \sum_{j=1}^{N_{tx}} h_{ij} a_j(k) \;+\; v_i(k) \qquad (1)$$

where the $a_j(k)$ are the transmitted symbols from source j, h_{ij} is the (overall) channel response from transmitter j to receiver i and $v_i(k)$ is the additive noise at the same receiver. The discrete-time RX signal can be represented in vector form as:

$$\mathbf{y}_k = \sum_{j=1}^{N_{tx}} \mathbf{h}_j a_j(k) + \mathbf{v}_k = \mathbf{H}\,\mathbf{a}_k + \mathbf{v}_k$$

$$\mathbf{y}_k = \begin{bmatrix} y_1(k) \\ \vdots \\ y_{N_{rx}}(k) \end{bmatrix},\ \mathbf{v}_k = \begin{bmatrix} v_1(k) \\ \vdots \\ v_{N_{rx}}(k) \end{bmatrix},\ \mathbf{h}_j = \begin{bmatrix} h_{1j} \\ \vdots \\ h_{N_{rx}j} \end{bmatrix},$$

$$\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_{N_{tx}}],\ \mathbf{a}_k = [a_1(k) \cdots a_{N_{tx}}(k)]^T\ .$$

Superscripts ^{T}, ^{H} denote transpose and Hermitian transpose respectively.
The multichannel aspect leads to a signal subspace when $N_{tx} < N_{rx}$, since $\mathbf{y}_k = \mathbf{H}\,\mathbf{a}_k + \mathbf{v}_k$. The existence of this signal subspace has led to the development of a wealth of blind channel estimation techniques over the last decade. Some of these techniques are relatively simple due to the modeling of the unknown input symbols as either deterministic unknowns (deterministic input case) or uncorrelated random variables (without exploiting their finite alphabet nature). The latter (uncorrelated) case is also called the Gaussian input case because (only) second-order statistics are exploited. However, when $N_{tx} \geq N_{rx}$, there is no deterministic blind channel information and only limited Gaussian information. Even when $N_{tx} < N_{rx}$, most of these blind techniques are not very robust and leave channel ambiguities. These ambiguities can range from a simple scalar ambiguity factor for Single-Input Multiple-Output (SIMO) channels to a square $N_{tx} \times N_{tx}$ matrix for Multiple-Input Multiple-Output (MIMO) channels [1], [2].
On the other hand, all current standardized communication systems employ some form of known inputs to allow channel estimation. The channel estimation performance in those cases can always be improved by a semiblind approach which exploits both training and blind information. The training information allows to resolve the blind ambiguities and robustifies the channel estimates. The purpose of this paper is to introduce semiblind techniques of which the complexity is not immensely much higher than that of training based techniques.

To use the second-order statistics the input samples are modeled as i.i.d. white centered Gaussian inputs $\mathbf{a}_k \sim \mathcal{CN}(0, \sigma_a^2 \mathbf{I}_{N_{tx}})$, the independent noise is considered to be i.i.d. white Gaussian $\mathbf{v}_k \sim \mathcal{CN}(0, \sigma_v^2 \mathbf{I}_{N_{rx}})$ and $SNR = \frac{\sigma_a^2}{\sigma_v^2} = \rho$. The received signal frame contains two parts:

- Training Sequence of N_{TS} pilot symbol vectors. The training received signal follows a non-zero mean Gaussian distribution: $\mathbf{y}_k^{TS}/\mathbf{H} \sim \mathcal{CN}(\mathbf{H}\mathbf{a}_k^{TS}, \sigma_v^2 \mathbf{I}_{N_{rx}})$.
- Blind part of N_B data symbol vectors. These follow a zero mean Gaussian distribution: $\mathbf{y}_k/\mathbf{H} \sim \mathcal{CN}(0, \sigma_v^2\,\mathbf{I}_{N_{rx}} + \sigma_a^2\,\mathbf{H}\mathbf{H}^H)$.

Below we assume the noise power σ_v^2 to be known by the receiver.

II. Maximum Likelihood Channel Estimator

The Maximum Likelihood (ML) Channel Estimator, is the one that maximize the Log Likelihood (LL) of the total received signal:

$$\begin{aligned} LL(\mathbf{H}) &= \ln p(\mathbf{Y}/\mathbf{H}) \\ &= constant \\ &- {\sigma_v^2}^{-1} \sum_{k=1}^{N_{TS}} (\mathbf{y}_k^{TS} - \mathbf{H}\mathbf{a}_k^{TS})^H (\mathbf{y}_k^{TS} - \mathbf{H}\mathbf{a}_k^{TS}) \\ &- \sum_{k=1}^{N_B} \mathbf{y}_k^H (\sigma_v^2 \mathbf{I}_{N_{rx}} + \sigma_a^2 \mathbf{H}\mathbf{H}^H)^{-1} \mathbf{y}_k \\ &- N_B \ln\det(\sigma_v^2 \mathbf{I}_{N_{rx}} + \sigma_a^2 \mathbf{H}\mathbf{H}^H) \\ &= constant \\ &+ \underbrace{-{\sigma_v^2}^{-1} \mathrm{tr}\{(\mathbf{Y}^{TS} - \mathbf{H}\mathbf{A}^{TS})^H (\mathbf{Y}^{TS} - \mathbf{H}\mathbf{A}^{TS})\}}_{LL_{TS}(\mathbf{H})} \\ &+ \underbrace{-N_B \mathrm{tr}\{\mathbf{R}^{-1}(\mathbf{H})\hat{\mathbf{R}}\} - N_B \ln\det \mathbf{R}(\mathbf{H})}_{LL_B(\mathbf{H})} \end{aligned}$$

where $\mathbf{Y}^{TS} = [\mathbf{y}_1^{TS} \ldots \mathbf{y}_{N_{TS}}^{TS}]$, $\hat{\mathbf{R}} = \frac{1}{N_B} \sum_{k=1}^{N_B} \mathbf{y}_k \mathbf{y}_k^H$ and $\mathbf{R}(\mathbf{H}) = \sigma_v^2 \mathbf{I}_{N_{rx}} + \sigma_a^2 \mathbf{H}\mathbf{H}^H$. $LL_{TS}(\mathbf{H})$ and $LL_B(\mathbf{H})$ are the Log Likelihood of the blind and the training parts. The ML channel estimate is then:

$$\widehat{\mathbf{H}}_{ML} = \arg\max_{\mathbf{H}} LL(\mathbf{H})$$

A. Information Matrix Issues

Let the Singular Value Decomposition (SVD) of the channel be: $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{Q} = \mathbf{W}\mathbf{Q}$ where $\mathbf{U}$(resp. $\mathbf{Q}$) is a $N_{rx} \times \min\{N_{rx}, N_{tx}\}$ (resp. $\min\{N_{rx}, N_{tx}\} \times N_{tx}$) unitary matrix, i.e $\mathbf{U}^H\mathbf{U} = \mathbf{I}$ (resp. $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$). Let $\mathbf{W} = \mathbf{W}(\alpha)$ and $\mathbf{Q} = \mathbf{Q}(\beta)$ be two bijective real parameterizations: $\mathbf{H} = \mathbf{H}(\alpha, \beta)$. The blind part contains no information on $\mathbf{Q}$. The Fischer information matrix is then:

$$\begin{aligned} J(\mathbf{H}) &= -\mathrm{E}_{\mathbf{Y}} \frac{\partial}{\partial \mathbf{h}} \left(\frac{\partial \ln p(\mathbf{Y}/\mathbf{H})}{\partial \mathbf{h}}\right)^T \\ &= J_B(\mathbf{H}) + J_{TS}(\mathbf{H}) \\ &= M_1 \, J_B(\alpha) \, M_1^T + J_{TS}(\mathbf{H}) \end{aligned}$$

where $M_1 = \frac{\partial \alpha^T}{\partial \mathbf{h}} = \frac{\partial \mathbf{w}^T}{\partial \mathbf{h}} \frac{\partial \alpha^T}{\partial \mathbf{w}} = (\mathsf{M}(\mathbf{Q}^H) \otimes \mathbf{I}_{N_{rx}}) \frac{\partial \alpha}{\partial \mathbf{w}}$, $\mathbf{h} = [\Re(vec(\mathbf{H}))^T \, \Im(vec(\mathbf{H}))^T]^T$, $\mathbf{w} = [\Re(vec(\mathbf{W}))^T \, \Im(vec(\mathbf{W}))^T]^T$ and $\mathsf{M}(\mathbf{M}) = \begin{bmatrix} \Re(\mathbf{M}) & -\Im(\mathbf{M}) \\ \Im(\mathbf{M}) & \Re(\mathbf{M}) \end{bmatrix}$ for any matrix $\mathbf{M}$. $\mathrm{E}_{\mathbf{Y}}$ denotes expectation w.r.t. $\mathbf{Y}$. $J_B(\mathbf{H})$ and $J_{TS}(\mathbf{H})$ are the Fischer information matrices of the blind and the training parts. $J_{TS}(\mathbf{H})$ can be evaluated easily: $J_{TS}(\mathbf{H}) = \frac{2}{\sigma_v^2} \mathsf{M}\left(\mathbf{A}^{TS}\mathbf{A}^{TS^H}\right) \otimes \mathbf{I}_{N_{rx}}$.

The MSE error of any unbiased channel estimate satisfies:

$$\mathrm{E}||\tilde{\mathbf{H}}||^2 \geq \mathrm{E}_{\mathbf{H}} \, CRB = \mathrm{tr}\, \mathrm{E}_{\mathbf{H}} J^{-1}(\mathbf{H}),$$

where $CRB = \mathrm{tr} J^{-1}(\mathbf{H})$ is the Cramer-Rao Bound on the estimate of the channel for a given channel realization. We use $\mathrm{E}_{\mathbf{H}}$ to average over a possible statistical distribution of the channel.

For the design of the TS, the following theorem gives a useful result:

Theorem 1 : For statistical channel $\mathbf{H} = \mathbf{W}\mathbf{Q}$ with $\mathbf{Q}$ uniformly distributed over the Grassmann manifold, the minimum of $\mathrm{E}_{\mathbf{H}} \, CRB$ is achieved by a white training sequence: $\mathbf{A}^{TS} \, \mathbf{A}^{TS^H} \propto \mathbf{I}$.

Proof :

Let $\mathbf{U}^{TS}\mathbf{D}^{TS}\mathbf{U}^{TS^H}$ be the eigen decomposition of $\mathbf{A}^{TS}\mathbf{A}^{TS^H}$, then

$$\mathrm{E}_{\mathbf{H}} CRB = \mathrm{E}_{\mathbf{H}} \; \mathrm{tr} \, [\tfrac{2}{\sigma_v^2} \mathbf{D}^{TS} \otimes \mathbf{I}_{2\,N_{rx}} + ([\mathsf{M}(\mathbf{U}^{TS^H}) \otimes \mathbf{I}_{N_{rx}}] M_1) \, J_B(\alpha) \, ([\mathsf{M}(\mathbf{U}^{TS^H}) \otimes \mathbf{I}_{N_{rx}}] M_1)^T]^{-1}.$$

$$[\mathsf{M}(\mathbf{U}^{TS^H}) \otimes \mathbf{I}_{N_{rx}}] M_1 = [\mathsf{M}(\left(\mathbf{Q}\mathbf{U}^{TS}\right)^H) \otimes \mathbf{I}_{N_{rx}}] \frac{\partial \alpha^T}{\partial \mathbf{w}}.$$

Given that $\mathbf{U}^{TS}$ is unitary, $\mathbf{Q}\mathbf{U}^{TS}$ has the same uniform distribution as $\mathbf{Q}$. On the other hand, $J_B(\alpha)$ is independent of $\mathbf{Q}$. Hence we can then conclude that the CRB is independent of $\mathbf{U}^{TS}$ and that:

$$\mathrm{E}_{\mathbf{H}} \, CRB = \mathrm{E}_{\mathbf{H}} \, \mathrm{tr} \left(M_1 \, J_B(\alpha) \, M_1^T + \frac{2}{\sigma_v^2} \mathbf{D}^{TS} \otimes \mathbf{I}_{2\,N_{rx}} \right)^{-1}.$$

The second step is to show that $\mathrm{E}_{\mathbf{H}} \, CRB = f(\mathbf{D}^{TS})$ is a convex function over the connex set $\mathbf{D}^{TS} \geq 0$. Let $\mathbf{D}^{TS} = diag(d_1^{TS}, \ldots, d_{N_{tx}}^{TS})$ and $\mathbf{C}$ the Hessian of $f(\mathbf{D})$ ($\mathbf{C}_{i,j} = \frac{\partial^2 f}{\partial d_i \, \partial d_j}$), then it can be shown that for any real positive vector $\mathbf{x} = [x_1 \cdots x_{N_{tx}}]^T \geq 0$ (let $\mathbf{X} = diag(x_1, \ldots, x_{N_{tx}})$): $\mathbf{x}^T\mathbf{C}\mathbf{x} =$

$$\frac{8}{\sigma_v^4} \mathrm{E}_{\mathbf{H}} \mathrm{tr} \left(J^{-2}(\mathbf{H})(\mathbf{X} \otimes \mathbf{I}_{2\,N_{rx}}) J^{-1}(\mathbf{H})(\mathbf{X} \otimes \mathbf{I}_{2\,N_{rx}}) \right) \geq 0$$

which follows from the fact that $J(\mathbf{H}) = M_1 \, J_B(\alpha) \, M_1^T + \frac{2}{\sigma_v^2} \mathbf{D}^{TS} \otimes \mathbf{I}_{2\,N_{rx}}$ is symmetric positive definite. This shows the convexity of $\mathrm{E}_{\mathbf{H}} \, CRB = f(\mathbf{D}^{TS})$ over the connex set $\mathbf{D}^{TS} \geq 0$. Then $\mathrm{E}_{\mathbf{H}} \, CRB$ has a global minimum under a power constraint expressed on the trace of $\mathbf{D}^{TS}$. This leads us to express the Lagrangian of this optimization problem:

$$L(\mathbf{D}^{TS}, \lambda) = f(\mathbf{D}^{TS}) + \lambda \left(\mathrm{tr}(\mathbf{D}^{TS}) - P_{constraint}\right)$$

$$\frac{\partial L}{\partial d_i} = -\frac{2}{\sigma_v^2} \mathrm{E}_{\mathbf{H}} \mathrm{tr} \left(J^{-2}(\mathbf{H}) \, (\mathbf{I}_i \otimes \mathbf{I}_{2\,N_{rx}}) \right) + \lambda = 0$$

where $\mathbf{I}_i$ is the matrix with 1 at the i^{th} diagonal element and zeros elsewhere. The solution to these equations corresponds to $\mathbf{D}^{TS}$ being a multiple of identity. Hence $\mathbf{A}^{TS} \, \mathbf{A}^{TS^H} \propto \mathbf{I}$ achieves the global minimum of $\mathrm{E}_{\mathbf{H}} \, CRB$. This proves the theorem. □

Asymptotic Behavior. If $\left(\mathbf{A}^{TS}\mathbf{A}^{TS^H}\right)^{-1}$ exists (persistently exciting training sequence), and $N_B >> \rho \, N_{TS}$, the Cramer-Rao Bound verifies:

$$\mathrm{E}_{\mathbf{H}} \, CRB = \mathrm{tr} \; \mathrm{E}_{\mathbf{H}} \left\{ J_{TS}^{-1}(\mathbf{H}) \, \mathbf{P}^{\perp}_{J_{TS}^{-\frac{1}{2}}(H) M_1} \right\} + O(\frac{1}{N_B})$$

where $O(\frac{1}{N_B})$ denotes a quantity of the order of $\frac{1}{N_B}$ The CRB is dominated by the part of the channel resulting from the projection on the orthogonal complement of $J_{TS}^{-\frac{1}{2}}(H)M_1$. This corresponds to the channel part that cannot be identified blindly, and hence gets identified only by the training.

Semiblind Method. The results above motivate us to propose the following method:

1- Estimate $\mathbf{U}$ and $\mathbf{D}$ from the Blind Part:

$$\hat{\mathbf{U}}\hat{\mathbf{D}} = \hat{\mathbf{W}} = \arg\max_{\mathbf{W}} LL_B(\mathbf{W})$$

2- Estimate $\mathbf{Q}$ from the Training Sequence Part:

$$\hat{\mathbf{Q}} = \arg\max_{\mathbf{Q}} LL_{TS}(\mathbf{Q}/\mathbf{W} = \hat{\mathbf{W}})$$

3- Form $\widehat{\mathbf{H}} = \hat{\mathbf{W}}\,\hat{\mathbf{Q}}$.

This method is further elaborated in the following section.

III. Gaussian Semi-Blind (GSB) Approach

The approach just described belongs to the Gaussian category because the blind information it exploits involves symbol second-order statistics. It is also semi-blind since blind and training based parts get combined.

Solution Blind Part. We write the eigen decompositions of the true and the estimated covariance matrices of the signal as: $\mathbf{R} = \mathbf{U}\,(\sigma_v^2 \mathbf{I}_{\min\{N_{tx},N_{rx}\}} + \sigma_a^2 D^2)\mathbf{U}^H + \sigma_v^2 \mathbf{U}^{\perp}\mathbf{U}^{\perp H}$, $\hat{\mathbf{R}} = \mathbf{U}_e \mathbf{S}_e \mathbf{U}_e^H$ where the subscript e denotes sample estimates, and $\mathbf{U}^{\perp}$ provides an orthonormal basis for the orthogonal complement of $\mathbf{U}$.

The Blind LL Part (up to a constant) is then:

$$LL_B(\mathbf{H}) = -N_B \mathrm{tr}\{\mathbf{R}^{-1}\hat{\mathbf{R}}\} - N_B \ln\det(\sigma_v^2\mathbf{I} + \sigma_a^2 D^2)\ .$$

Theorem 2 : The solution of the Blind Part is:

- $\hat{\mathbf{U}}$ corresponds to the $\min\{N_{tx}, N_{rx}\}$ dominant eigenvectors in $\mathbf{U}_e$
- $\hat{\mathbf{D}}$ matches the $\min\{N_{tx}, N_{rx}\}$ dominant eigenvalues of $\frac{1}{\sigma_a}\left(\lfloor \mathbf{S}_e - \sigma_v^2 \mathbf{I}_{N_{rx}} \rfloor_+\right)^{1/2}$
- $\hat{\mathbf{W}} = \hat{\mathbf{U}}\hat{\mathbf{D}}$

where $\lfloor . \rfloor_+$ takes the positive semidefinite part of its Hermitian argument.

Proof:

We first derive the solution for the unitary factor and then for the diagonal factor. We rewrite the parametric covariance matrix as: $\mathbf{R} = \mathbf{U}_R\mathbf{S}_R\mathbf{U}_R^H$, where $\mathbf{S}_R = diag(s_{R,1}, \ldots, s_{R,N_{rx}})$ in which the $s_{R,i}$ are organized in increasing order (introduce also similarly $\mathbf{S}_e = diag(s_{e,i}, \ldots, s_{e,N_{rx}})$). We note that by construction for $N_{tx} < N_{rx}$:

$$s_{R,i} = \sigma_v^2 \qquad , 1 \leq i \leq N_{rx} - N_{tx}$$
$$s_{R,i+N_{rx}-\min\{N_{tx},N_{rx}\}} = \sigma_v^2 + \sigma_a^2 {d_i}^2\ , 1 \leq i \leq \min\{N_{tx}, N_{rx}\}$$

Let $\mathbf{O} = \mathbf{U}_R^H\,\mathbf{U}_e$, $\mathbf{O}$ is a unitary $N_{tx}\times N_{tx}$ matrix. Let also $\mu_i = (\mathbf{U}_R^H\hat{\mathbf{R}}\mathbf{U}_R)_{ii} = (\mathbf{O}\mathbf{S}_e\mathbf{O}^H)_{ii}$, $i = 1, \ldots, N_{rx}$. Then (up to a constant):

$$\begin{aligned} LL_B(\mathbf{H}) &= -N_B \mathrm{tr}\{\mathbf{S}_R^{-1}\mathbf{O}\mathbf{S}_e\mathbf{O}^H\} - N_B \ln\det(\mathbf{S}_R) \\ &= -N_B \textstyle\sum_{i=1}^{N_{rx}} (\mu_i/s_{R,i} + \ln(s_{R,i}))\ . \end{aligned}$$

It can be shown [3] that $(\mu_i)_{1\leq i\leq N_{rx}}$ majorizes $(s_{e,i})_{1\leq i\leq N_{rx}}$, i.e. $\sum_{i=1}^{N_{rx}} s_{e,i} = \sum_{i=1}^{N_{rx}} \mu_i$ and $\sum_{i=1}^{k} s_{e,i} \leq \sum_{i=1}^{k} \mu_i$, $1 \leq k \leq N_{rx}$. Then the following result is proven in [4]:

$$\textstyle\sum_{i=1}^{N_{rx}} \mu_i/s_{R,i} \quad \geq \quad \sum_{i=1}^{N_{rx}} s_{e,i}/s_{R,i}$$

or equivently

$$-\mathrm{tr}\{\mathbf{O}\mathbf{S}_e\mathbf{O}^H\mathbf{S}_R^{-1}\} \quad \leq \quad -\mathrm{tr}\{\mathbf{S}_e\mathbf{S}_R^{-1}\}.$$

This shows that $LL_B(\mathbf{H})$ is maximized for $\mathbf{O} = \mathbf{I}_{N_{rx}}$, i.e. $\mathbf{U}_R = \mathbf{U}_e$ or equivalently that $\hat{\mathbf{U}}$ corresponds to the $\min\{N_{tx}, N_{rx}\}$ dominant eigenvectors in $\mathbf{U}_e$. Let us now evaluate the optimal $\hat{\mathbf{D}} = diag(\hat{d}_1, \ldots, \hat{d}_{\min\{N_{tx},N_{rx}\}})$. $LL_B(\mathbf{H})$ is separable in the $\hat{d}_i$, is a monotonically increasing function for $0 \leq \hat{d}_i \leq \sqrt{\lfloor s_{R,i+N_{rx}-\min\{N_{tx},N_{rx}\}} - \sigma_v^2 \rfloor_+}/\sigma_a$, $1 \leq i \leq \min\{N_{tx}, N_{rx}\}$, and is monotonically decreasing for $\hat{d}_i \geq \sqrt{\lfloor s_{R,i+N_{rx}-\min\{N_{tx},N_{rx}\}} - \sigma_v^2 \rfloor_+}/\sigma_a$, $1 \leq i \leq \min\{N_{tx}, N_{rx}\}$. $LL_B(\mathbf{H})$ is hence maximized for $\hat{d}_i = \sqrt{\lfloor s_{R,i+N_{rx}-\min\{N_{tx},N_{rx}\}} - \sigma_v^2 \rfloor_+}/\sigma_a$, $1 \leq i \leq \min\{N_{tx}, N_{rx}\}$, which are the dominant $\min\{N_{tx}, N_{rx}\}$ values of $\frac{1}{\sigma_a}\left(\lfloor \mathbf{S}_e - \sigma_v^2 \mathbf{I}_{N_{rx}} \rfloor_+\right)^{1/2}$. This ends the proof. □

Solution Training Part. Given the $\hat{\mathbf{W}}$ estimate, the TS LL part (up to a constant) becomes :

$$\begin{aligned} &\sigma_v^2\ LL_{TS}(\mathbf{Q}/\mathbf{W} = \hat{\mathbf{W}}) \\ &= -\mathrm{tr}\{(\mathbf{Y}^{TS} - \hat{\mathbf{W}}\mathbf{Q}\mathbf{A}^{TS})^H\,(\mathbf{Y}^{TS} - \hat{\mathbf{W}}\mathbf{Q}\mathbf{A}^{TS}) \\ &= 2\Re\,\mathrm{tr}\{\mathbf{A}^{TS}\mathbf{Y}^{TS^H}\hat{\mathbf{W}}\mathbf{Q}\} - \mathrm{tr}(\hat{\mathbf{W}}^H\hat{\mathbf{W}}\mathbf{Q}\mathbf{A}^{TS}\mathbf{A}^{TS^H}\mathbf{Q}^H) \end{aligned}$$

Due to the quadratic constraint ($\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$), the solution is non trivial in general. However, for optimal training $\mathbf{A}^{TS}\,\mathbf{A}^{TS^H} = \beta^{TS}\,\mathbf{I}$, $\beta^{TS} > 0$, $\hat{\mathbf{W}}^H\hat{\mathbf{W}}\mathbf{Q}\mathbf{A}^{TS}\mathbf{A}^{TS^H}\mathbf{Q}^H = \beta^{TS}\hat{\mathbf{W}}^H\hat{\mathbf{W}}$, and the TS LL part is then (up to a constant):

$$LL_{TS}(\mathbf{Q}/\mathbf{W} = 2\,\sigma_v^{-2}\,\Re\,\mathrm{tr}\{\mathbf{A}^{TS}\mathbf{Y}^{TS^H}\hat{\mathbf{W}}\mathbf{Q}\}\ . \qquad (2)$$

Theorem 3 : For white training sequence, $\mathbf{A}^{TS}\,\mathbf{A}^{TS^H} \propto \mathbf{I}$, the solution for the Training Part is:

$$\hat{\mathbf{Q}} = \mathbf{V}\,\mathbf{S}^H\ ,$$

where $\mathbf{S}$ and $\mathbf{V}$ denote here the unitary factors of the Singular Value Decomposition of $\mathbf{A}^{TS}\mathbf{Y}^{TS^H}\hat{\mathbf{W}} = \mathbf{S}\,\Sigma\,\mathbf{V}^H$.

Proof: The maximization of (2) corresponds to a subspace rotation problem and is solved in [2], [5].

IV. Deterministic Semi-Blind (DSB) Approach

In this section we don't exploit known correlations of the inputs, leading to only the exploitation of the subspace. The use of this approach is restricted to the case when a noise subspace of the spatial covariance matrix exists, i.e. $N_{rx} > N_{tx}$. The blind information expresses then the orthogonality of the channel to the noise subspace:

$\mathbf{U}^{\perp,H}\mathbf{H} = 0$. Using a weighted least squares approach we combine the blind and training parts in a quadratic criterion:

$$\min_{\mathbf{H}} \left(\sigma_v^{-2}||\mathbf{Y}^{TS} - \mathbf{H}\mathbf{A}^{TS}||_F^2 \; + \; N_B||\hat{\mathbf{U}}^{\perp H}\mathbf{H}^H||_F^2\right)$$

where $||\mathbf{M}||_F^2 = \text{tr}\,\mathbf{M}^H\mathbf{M}$ is the Frobenius norm of $\mathbf{M}$. Further details about this approach can be found in [2].

V. MIMO Frequency Selective Channel

For a channel length of L symbol periods, the sampled received signal can be written as follows:

$$\mathbf{y}_k = \sum_{l=0}^{L-1} \mathbf{H}_l\,\mathbf{a}_{k-l} + \mathbf{v}_k\,.$$

In this case the TS and Blind parts interfere, hence the training and blind LL parts are no longer separable. To continue to express the LL separately, we assume the use of a cyclic prefix and neglect the effect of the interference with the training signal when evaluating the blind LL. Asymptotically in the length of the blind part N_B and for $N_B >> max(N_{TS}, L)$ this is correct, and leads (up to a constant) to:

$$\begin{aligned} &LL_B(\mathbf{H}) \\ &= -\sum_{k=0}^{N_B-1}[\mathbf{y}^H(f_k)(\sigma_v^2\mathbf{I}_{N_{rx}} + \sigma_a^2\,\mathbf{H}(f_k)\,\mathbf{H}^H(f_k))^{-1}\mathbf{y}(f_k) \\ &\qquad + \ln\det(\sigma_v^2\mathbf{I}_{N_{rx}} + \sigma_a^2\,\mathbf{H}(f_k)\,\mathbf{H}^H(f_k))] \\ &= -\sum_{k=0}^{N_B-1}[\text{tr}(\mathbf{R}^{-1}(f_k)\mathbf{y}(f_k)\mathbf{y}^H(f_k)) + \ln\det\mathbf{R}^{-1}(f_k)] \\ &= -\sum_{k=0}^{N_B-1}[\text{tr}(\mathbf{R}^{-1}(f_k)\hat{\mathbf{R}}(f_k)) + \ln\det\mathbf{R}^{-1}(f_k)] \end{aligned}$$

where $\mathbf{y}(f_k)$, $\mathbf{H}(f_k)$ is the DFT of the sequence $\mathbf{y}_i$, $\mathbf{H}_i$ at the normalized frequency $f_k = \frac{k}{N_B}$, and $\hat{\mathbf{R}}(f_k) = \mathbf{y}(f_k)\mathbf{y}^H(f_k)$ is a highly noisy estimate of $\mathbf{R}(f_k)$.
The maximization of $LL_B(\mathbf{H})$ leads to covariance matching. The problem is then how to do covariance matching of $\hat{\mathbf{R}}(f_k)$ with acceptable complexity. First, to take advantage of the a priori knowledge of the finite channel length (only $\mathbf{R}(k)$, $k = -L+1, -L+2, \ldots, L-2, L-1$ are nonzero, the $\mathbf{R}(k)$ sequence is the inverse DFT of $\mathbf{R}(f_k)$), we do covariance matching in the time domain, this should allow to reduce the complexity.
Second, to be able to do the covariance matching, we have to use an appropriate parameterization of the channel to characterize the channel from the correlation sequence $\hat{\mathbf{R}}(k)$. $\mathbf{P}_K(z) = \mathbf{I} + \sum_{i=1}^{K-1}\mathbf{P}_i z^{-i}$ is a predictor for the channel $\mathbf{H}(z)$ if $\mathbf{P}_K(z)\mathbf{H}(z) = \mathbf{H}_0$. For $N_{rx} \geq N_{tx}$ the channel predictor generically exists and is FIR for $N_{rx} > N_{tx}$, this constitutes an appropriate parameterization of the channel:

$$\mathbf{H}(z) = \mathbf{P}^{-1}(z)\mathbf{U}\mathbf{D}\mathbf{Q} = \mathbf{P}^{-1}(z)\mathbf{W}\mathbf{Q}\,.$$

For $N_{rx} > N_{tx}$ and $K \geq \left\lceil\frac{L-N_{tx}}{N_{rx}-N_{tx}}\right\rceil$ the predictor can be evaluated from $\hat{\mathbf{R}}(k)$, $k = 0, \ldots, K-1$, this fixes the channel up to a unitary matrix: $\hat{\mathbf{H}}(z) = \hat{\mathbf{P}}_K^{-1}(z)\hat{\mathbf{W}}\mathbf{Q}$ $(\mathbf{R}(z) - \sigma_v^2\,\mathbf{I} = \mathbf{P}_K^{-1}(z)\mathbf{W}\mathbf{W}^H\mathbf{P}_K^{-\dagger}(z))$.
However unlike in the flat channel case, there is no trivial method to estimate $\mathbf{Q}$ by ML. If we reduce the exploitation of $\mathbf{P}(z)\,\mathbf{H}(z) = \mathbf{H}_0$ or $\mathbf{P}(q)\,\mathbf{H}_k = \mathbf{H}_0\,\delta_{k0}$ to $\mathbf{W}^{\perp,H}\,\mathbf{H}_0 = 0$ and $\mathbf{P}(q)\,\mathbf{H}_k = 0$, $k > 0$; and combining it with the training part in a weighted least squares approach, then the result is a simple quadratic criterion. Further results on this approach can be found in [2].

VI. Simulations

Fig. 1 to 6 consider the flat channel case with a blind part of length $N_B = 400$, the channel is 4×2 ($N_{rx} \times N_{tx}$) for Fig. 1 and 2, 4×4 for Fig. 3 and 4, and 2×4 for Fig. 5 and 6. We compare the classical TS approach with the GSB/DSB approaches and with the Cramer-Rao Bound (CRB) in terms of normalized mean square estimation error (NMSE). In Fig. 1, 3 and 5 the performances are given for different TS lengths with a fixed SNR $\rho = 10db$, when for Fig. 2, 4 and 6 the performances are given for different SNR with $N_{TS} = 4$. The results show that wherever the condition $N_B >> \rho\,N_{TS}$ is fulfilled, the proposed GSB approach achieves the CRB. The GSB performances show a linear behavior w.r.t the SNR, and outperform the TS approach by a gap corresponding approximatively to the relative reduction of the number of real parameters to be estimated; for 4×2 : $\ln\frac{16}{4} = 6db$, 4×4 : $\ln\frac{32}{16} = 3db$ and 2×4 : $\ln\frac{16}{12} \approx 1db$. The GSB saturates for very high SNR due to the lack of consistency (in SNR) of the channel part estimated blindly. This doesn't appear in the case of the DSB in which all information exploited is consistent in SNR. However, the DSB is suboptimal when $N_{rx} > N_{tx}$ (4×2) and its performance reduces to that of the TS approach for $N_{rx} \leq N_{tx}$.

Acknowledgments

Eurécom's research is partially supported by its industrial partners: Ascom, Swisscom, Thales Communications, ST Microelectronics, CEGETEL, France Télécom, Bouygues Telecom, Hitachi Europe Ltd. and Texas Instruments. The work reported herein was also partially supported by the French RNRT project ERMITAGES.

References

[1] E. de Carvalho, L. Deneire, and D.T.M. Slock. Blind and semi-blind maximum-likelihood techniques for multi-user multichannel identification. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Rhodes, Greece, Sept. 1998.

[2] A. Medles, E. de Carvalho, and D.T.M. Slock. Linear prediction based semi-blind estimation of mimo fir channels. In *Proc. 3rd Workshop on Signal Processing Advances in Wireless Communications(SPAWC 2001)*, Taoyuan, Taiwan, March 2001.

[3] R. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1991.

[4] S.A. Jafar, S. Vishwanath, and A. Goldsmith. Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback. In *Proc. ICC 2001*, Helsinki, Finland, Jun. 2001.

[5] G.H. Golub and C.F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, third edition edition, 1996.

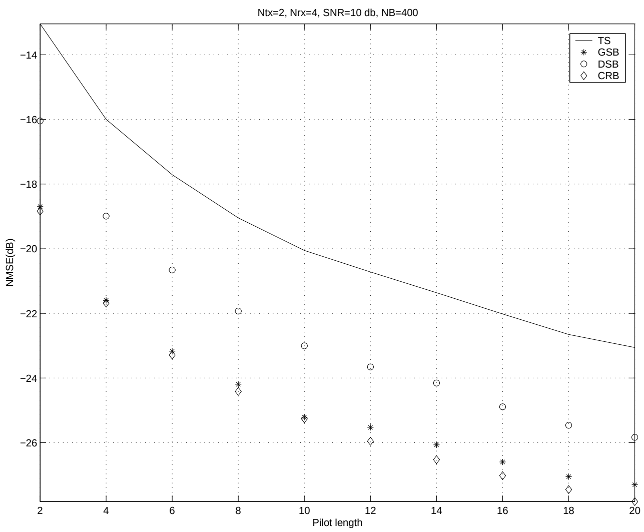

Fig. 1. Normalized MSE: flat channel, $N_{tx} = 2, N_{rx} = 4,\ N_B = 400$, SNR= 10 db.

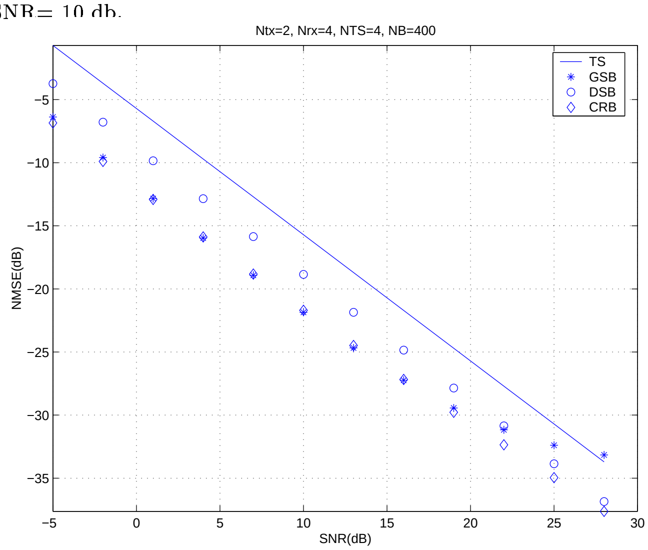

Fig. 2. Normalized MSE: flat channel, $N_{tx} = 2, N_{rx} = 4,\ N_B = 400$, $N_{TS} = 4$.

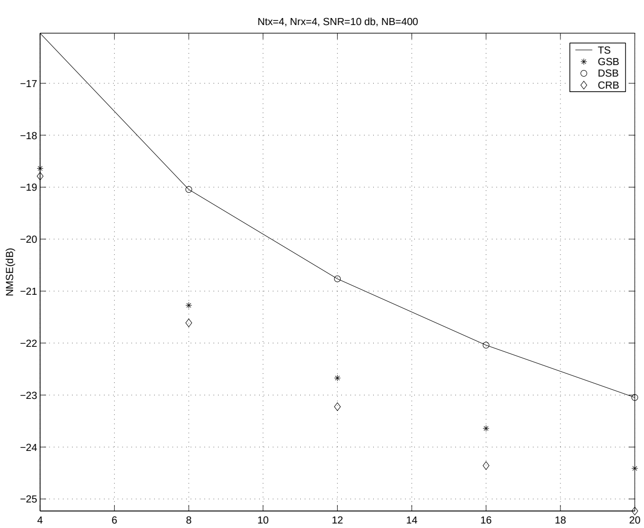

Fig. 3. Normalized MSE: flat channel, $N_{tx} = 4, N_{rx} = 4,\ N_B = 400$, SNR= 10 db.

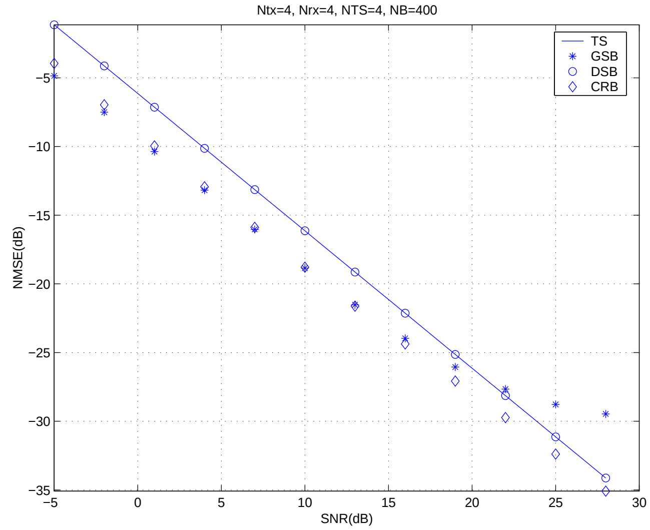

Fig. 4. Normalized MSE: flat channel, $N_{tx} = 4, N_{rx} = 4,\ N_B = 400$, $N_{TS} = 4$.

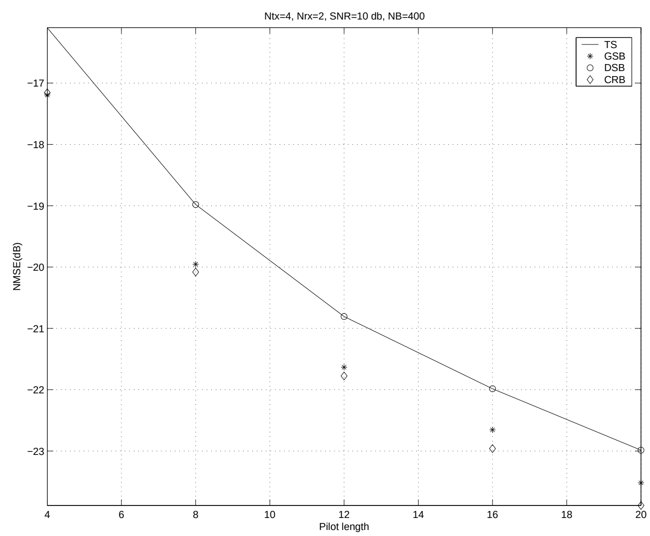

Fig. 5. Normalized MSE: flat channel, $N_{tx} = 4, N_{rx} = 2,\ N_B = 400$, SNR= 10 db.

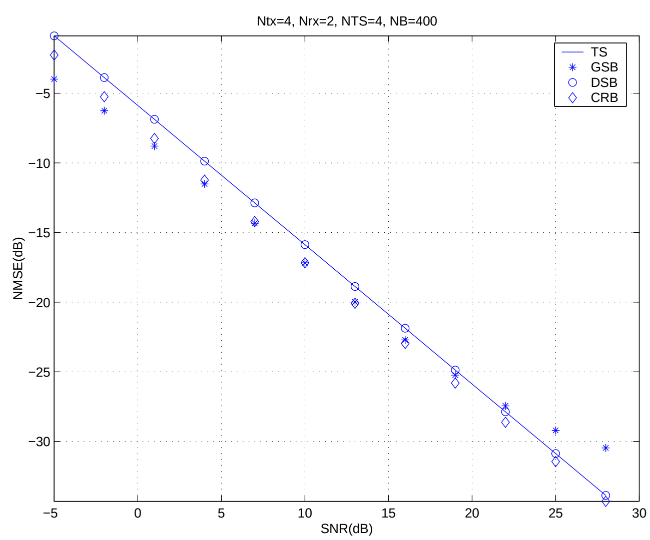

Fig. 6. Normalized MSE: flat channel, $N_{tx} = 4, N_{rx} = 2,\ N_B = 400$, $N_{TS} = 4$.