This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2009 proceedings

# AN EFFICIENT REGULARIZED SEMI-BLIND ESTIMATOR

*Abla Kammoun*[1] *Karim Abed-Meraim*[1,2] *and Sofiène Affes*[3]

Télécom ParisTech 46, rue Barrault, 75634 Paris Cedex 13, France[1]
ECE Department, College of Engineering, University of Sharjah, P.O. Box 27272, UAE[2]
INRS-EMT, 800, de la Gauchetière, Bureau 6900, Montreal, QC, H5A 1K6 Canada[3]
E-mails: kammoun, abed@tsi.enst.fr, affes@emt.inrs.ca

## 1. ABSTRACT

This paper addresses the issue of the optimization of the regularization constant in semi-blind channel estimation techniques, in which the training sequence-based criterion is combined linearly with the blind subspace criterion. In such semi-blind estimation techniques, the optimization of the regularizing constant with respect to the channel estimation error is mandatory, otherwise, the expected improvement in performance could not be achieved. In this context, recent works proposed numerical methods for the setting of the regularization constant. However, these methods are often sub-optimum and involve high computational complexities. In this paper, we propose to optimize with respect to a regularizing matrix instead of a regularizing scalar. We prove that interestingly in this case, a closed-form expression for the optimum regularizing matrix exists, thereby avoiding iterative algorithms as for the conventional techniques. We also prove that the obtained scheme has slightly better performance in terms of mean square error and bit error rate while ensuring lower complexity.

**Key words:** semi-blind equalization, asymptotic analysis, regularization

## 2. INTRODUCTION

In most current wireless communication systems, equalization techniques are solely based on training. The transmitter sends a known sequence of training symbols before transmitting data sequences. In current mobile telecommunication systems (eg. GSM), the training period can represent up to 30% of the effective data rate, thereby leading to a sub-optimum use of the effective bandwidth. In order to reduce the length of the training period while keeping the same channel estimation quality, methods that take into account the information retrieved from training and data were recently proposed (See [1] and the references therein). They are referred to as semi-blind methods.

However, from a practical point of view, the implementation of these methods implies in general some practical difficulties. The optimal semi-blind estimation technique is implemented by the expectation maximization (EM) algorithm [2, 3], whose complexity grows exponentially with the channel length and the receiving antenna-array size. A sub-optimal alternative which minimizes a weighted sum of the training and the blind cost functions was first proposed in [4]. This method is referred to as regularized semi-blind channel estimation method in reference to the regularizing constant which parametrizes the channel estimate. As opposed to the EM algorithm, this technique exhibits low computational complexity, but its performance is strongly influenced by the tuning of the regularizing constant [5]. Based on an asymptotic analysis, [5] proposed to set the regularizing constant in such a way to minimize the asymptotic channel estimation error. The optimal value for the regularizing constant has no closed-form expression and is determined through the use of iterative algorithms. In [6], it is proposed to evaluate the asymptotic channel estimation error for a finite number of possible values for the regularizing constant and keep thereafter the value that exhibits the least channel estimation error.

In this paper, we propose to use a regularizing matrix instead of regularizing constant. In contrast to preliminary predictions, the optimization of the proposed method is less complicated and yields a closed-form expression for the regularizing matrix. It is shown also to slightly outperform the conventional regularized scheme proposed in [5], while avoiding the need for iterative computations.

The remainder of this paper is as follows: In the next section, we review the blind and the least square channel estimation techniques. After that, we provide the general expression of the semi-blind estimator which is valid for any blind estimation technique that can be represented by a minimization problem of a certain quadratic form. Then, we derive the closed-form expression for the optimal regularizing constant that minimizes the asymptotic channel estimation error. Finally, we present in the last section the simulation results in order to assess the accuracy of the asymptotic results and evaluate the performance of the proposed method.

**Notations:** Subscriptis $^{\mathrm{H}}$ and $^{\#}$ denote respectively hermitian and pseudo-inverse operators. $\mathcal{M}_{m\times n}(\mathbb{C})$ denotes the space of complex $m\times n$ matrices. The $(K\times K)$ identity matrix is denoted by $\mathbf{I}_K$. The $(i,j)$-th entry of a matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. $\|.\|$ denotes the Euclidean norm of a vector.

## 3. BLIND AND LEAST SQAURE CHANNEL ESTIMATION

We consider a Single Input Multiple Output system (SIMO) with $N$ receiving antennas. We assume that the channel is frequency selective and is constant over one single packet. Sampling at the symbol rate $T$ leads to the following discrete time SIMO model:

$$\overrightarrow{\mathbf{y}_k} = \sum_{l=0}^{L} \overrightarrow{\mathbf{h}}_l s_{k-l} + \overrightarrow{\mathbf{v}_k},$$

where :

- $s_k$ are assumed to be iid (independent and identically distributed) complex circular random variables, with $\mathbb{E}(s_k)=0$ and $\mathbb{E}\,|s_k|^2 = 1$
- $\overrightarrow{\mathbf{y}_k}$ denotes the $N\times 1$ received vector at time $kT$,

- $\overrightarrow{\mathbf{v}_k}$ denotes the $N \times 1$ additive noise vector, whose entries are assumed to belong to an iid circular complex random multivariate process with finite fourth-order moment. The noise vector is also assumed to be independent from the input symbols and to be spacially white, i.e, $\mathbb{E}\overrightarrow{\mathbf{v}_k} = 0$ and $\mathbb{E}\overrightarrow{\mathbf{v}_k}\overrightarrow{\mathbf{v}_k}^{\mathrm{H}} = \sigma^2 \mathbf{I}_N$.
- $\overrightarrow{\mathbf{h}_l} = \left[h_l^1, \cdots, h_l^N\right]$ is the channel impulse response vector at the $l$-th tap between the transmitting antenna and the $N$ receiving antennas.

During the training period, the receiver gets $m$ samples $\mathbf{y} \triangleq \left[\overrightarrow{\mathbf{y}}_1^{\mathrm{T}}, \cdots, \overrightarrow{\mathbf{y}}_m^{\mathrm{T}}\right]^{\mathrm{T}}$ that depend on the training sequence. The vector $\mathbf{y}$ can be written as:

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{v},$$

where $\mathbf{v} \triangleq \left[\overrightarrow{\mathbf{v}_1}^{\mathrm{T}}, \cdots, \overrightarrow{\mathbf{v}_m}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\mathbf{S} = [\mathbf{S}_1, \cdots, \mathbf{S}_m]^{\mathrm{T}}$, $\mathbf{S}_k^{\mathrm{T}} = [s_k, \cdots, s_{k-L}] \otimes \mathbf{I}_N$, and $\mathbf{h} = \left[\overrightarrow{\mathbf{h}_0}^{\mathrm{T}}, \cdots, \overrightarrow{\mathbf{h}_L}^{\mathrm{T}}\right]^{\mathrm{T}}$. The least square estimate of $\mathbf{h}$ is therefore given by:

$$\begin{aligned}\widehat{\mathbf{h}} &= \left(\mathbf{S}^{\mathrm{H}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{H}}\mathbf{y}, \\ &= \mathbf{S}^{\#}\mathbf{y},\end{aligned}$$

where it is assumed that the training sequence is properly chosen to avoid the degeneracy of $\mathbf{S}^{\mathrm{H}}\mathbf{S}$.

In the data transmission period, by stacking $M$ observations $\overrightarrow{\mathbf{y}_k}$ in a $N\left(M+1\right)$ vector $\mathbf{Y}_k = \left[\overrightarrow{\mathbf{y}_k}^{\mathrm{T}}, \cdots, \overrightarrow{\mathbf{y}_{k-M}}^{\mathrm{T}}\right]^{\mathrm{T}}$, we will get:

$$\mathbf{Y}_k = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{s}_k + \mathbf{v}_k,$$

where $\boldsymbol{\mathcal{I}}_M(\mathbf{h})$ is the $N(M+1) \times (L+M+1)$ block-Toeplitz matrix given by:

$$\boldsymbol{\mathcal{I}}_M(\mathbf{h}) \triangleq \begin{pmatrix} \boxed{\overrightarrow{\mathbf{h}_0} \cdots \overrightarrow{\mathbf{h}_L}} & & 0 \\ & \boxed{\overrightarrow{\mathbf{h}_0} \cdots \overrightarrow{\mathbf{h}_L}} & \\ & \ddots & \ddots \\ 0 & & \boxed{\overrightarrow{\mathbf{h}_0} \cdots \overrightarrow{\mathbf{h}_L}} \end{pmatrix}.$$

The covariance matrix of the received signal $\mathbf{Y}_k$ can be expressed as:

$$\mathbf{R} = \mathbb{E}\mathbf{Y}_k{\mathbf{Y}_k}^{\mathrm{H}} = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}} + \sigma^2\mathbf{I}.$$

Let $n$ denote the length of the information sequence. The estimated covariance matrix $\widehat{\mathbf{R}}_n$ can be expressed as:

$$\widehat{\mathbf{R}}_n = \frac{1}{n}\sum_{k=1}^{n}\mathbf{Y}_k{\mathbf{Y}_k}^{\mathrm{H}}.$$

Based on the estimation of the covariance matrix, the most common blind estimation techniques evaluate the channel up to a scalar ambiguity by solving the following minimization problem:

$$\min_{\|\mathbf{h}\|=1} \mathbf{h}^{\mathrm{H}}\widehat{\mathbf{A}}_n\mathbf{h},$$

where $\widehat{\mathbf{A}}_n$ is an estimated matrix of $\mathbf{A}$, $\mathbf{A}$ being a matrix that depends on the considered blind estimation technique. For example, this is the case for subspace-based blind techniques as we will see later.

Note that the matrix $\mathbf{A}$ is singular and $\mathbf{h}$ is the unique vector (up to a scalar factor) generating its kernel.

## 4. SEMI-BLIND ESTIMATION

In the conventional regularized semi-blind estimation technique, the blind criterion is combined linearly with the training sequence criterion, thus leading to the following cost function:

$$C(\mathbf{f}, \alpha) = \|\mathbf{y} - \mathbf{S}\mathbf{f}\|^2 + \alpha n \mathbf{f}^{\mathrm{H}}\widehat{\mathbf{A}}_n\mathbf{f}, \quad (1)$$

where $\alpha$ is a regularizing constant. The semi-blind estimator that minimizes (1) is given by:

$$\widehat{\mathbf{h}}_\alpha(\mathbf{A}) = \left(\mathbf{S}^{\mathrm{H}}\mathbf{S} + \alpha n\widehat{\mathbf{A}}_n\right)^{-1}\mathbf{S}^{\mathrm{H}}\mathbf{y}.$$

The optimization of the regularizing constant cannot be done directly and requires in general the use of iterative algorithms. In this paper, we propose to minimize the following cost function that is given by:

$$C(\mathbf{f}, \boldsymbol{\Lambda}) = \|\mathbf{y} - \mathbf{S}\mathbf{f}\|^2 + n\mathbf{f}^{\mathrm{H}}\widehat{\mathbf{P}}_n\boldsymbol{\Lambda}\widehat{\mathbf{P}}_n\mathbf{f}, \quad (2)$$

where $\boldsymbol{\Lambda}$ is a regularizing matrix assumed to be hermitian, and $\widehat{\mathbf{P}}_n$ is an estimate of the orthogonal projector $\mathbf{P}$ onto the space spanned by the columns of $\mathbf{A}$. In this case, the semi-blind estimator that minimizes (2) is given by:

$$\widehat{\mathbf{h}}_{\boldsymbol{\Lambda}}(\mathbf{A}) = \left(\mathbf{S}^{\mathrm{H}}\mathbf{S} + n\widehat{\mathbf{P}}_n\boldsymbol{\Lambda}\widehat{\mathbf{P}}_n\right)^{-1}\mathbf{S}^{\mathrm{H}}\mathbf{y}.$$

Interestingly, we show in this paper that a closed-form expression for the optimum regularizing matrix $\boldsymbol{\Lambda}$ exists, thus avoiding the need for iterative algorithms.

## 5. PERFORMANCE ANALYSIS

Let us first recall the following result:

**Theorem 1** *Let* $\gamma = \lim_{n\to\infty}\frac{n}{m}$. *For any matrix* $\mathbf{A}$ *that verifies:*

- $\mathbf{A}\mathbf{h} = 0$,
- $\delta\mathbf{A} \triangleq \widehat{\mathbf{A}}_n - \mathbf{A} = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$.

*we have*

$$\widehat{\mathbf{h}}_\alpha(\mathbf{A}) - \mathbf{h} = \left(\mathbf{R}_{\mathbf{SS}} + \gamma\alpha\mathbf{A}\right)^{-1}\left(\mathbf{R}_{\mathbf{SV}} - \gamma\alpha\delta\mathbf{A}\mathbf{h}\right) + \mathcal{O}_p(\frac{1}{n}),$$

$$\widehat{\mathbf{h}}_{\boldsymbol{\Lambda}}(\mathbf{A}) - \mathbf{h} = \left(\mathbf{R}_{\mathbf{SS}} + \gamma\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}\right)^{-1}\left(\mathbf{R}_{\mathbf{SV}} - \gamma\delta\left(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}\right)\mathbf{h}\right) + \mathcal{O}_p(\frac{1}{n}),$$

*where* $\mathbf{R}_{\mathbf{SS}} = \frac{1}{m}\mathbf{S}^{\mathrm{H}}\mathbf{S}$ *and* $\mathbf{R}_{\mathbf{SV}} = \frac{1}{m}\mathbf{S}^{\mathrm{H}}\mathbf{v}$.

**Proof** : This theorem could be easily proved along the same lines as in [7].

**Theorem 2** $\sqrt{m}\left(\widehat{\mathbf{h}}_\alpha(\mathbf{A}) - \mathbf{h}\right)$ *is asymptotically normal with covariance matrix* $\Gamma_\alpha(\mathbf{h})$ *given by:*

$$\Gamma_\alpha(\mathbf{h}) = (\mathbf{I} + \alpha\gamma\mathbf{A})^{-1}\left(\sigma^2\mathbf{I} + \alpha^2\gamma^2\lim_{\substack{n\to\infty \\ \frac{n}{m}\to\gamma}} m\mathrm{Cov}(\delta\mathbf{A}\mathbf{h})\right)(\mathbf{I} + \alpha\gamma\mathbf{A})^{-1}. \quad (3)$$

*Also,* $\sqrt{m}\left(\widehat{\mathbf{h}}_{\boldsymbol{\Lambda}}(\mathbf{A}) - \mathbf{h}\right)$ *is asymptotically normal with covariance matrix* $\Gamma_{\boldsymbol{\Lambda}}(\mathbf{h})$ *given by:*

$$\Gamma_{\boldsymbol{\Lambda}}(\mathbf{h}) = (\mathbf{I} + \gamma\mathbf{P}\boldsymbol{\Lambda}\mathbf{P})^{-1}\left(\sigma^2\mathbf{I} + \gamma^2\lim_{\substack{n\to\infty \\ \frac{n}{m}\to\gamma}} m\mathrm{Cov}(\delta(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}\mathbf{h}))\right)(\mathbf{I} + \gamma\mathbf{P}\boldsymbol{\Lambda}\mathbf{P})^{-1}. \quad (4)$$

**Proof** : The asymptotic normality can be proved along the same lines as in [5]. Then, the expression for the covariance matrix is easily deduced from Theorem 1.

**Theorem 3** *The covariance matrix of* $\delta\mathbf{P\Lambda Ph}$ *is given by:*

$$\mathrm{Cov}\delta\left(\mathbf{P\Lambda Ph}\right) = \mathbf{P\Lambda A}^{\#}\mathrm{Cov}(\delta\mathbf{Ah})\mathbf{A}^{\#}\mathbf{\Lambda P} + \mathcal{O}_p(\frac{1}{n^2}).$$

**Proof** Let $\mathbf{P}_\perp$ be the orthogonal projector onto the null space of $\mathbf{A}$. Thus, $\mathbf{P} = \mathbf{I} - \mathbf{P}_\perp$. Using standard perturbation formulae [8], we get:

$$\widehat{\mathbf{P}}_\perp = \mathbf{P}_\perp - \mathbf{P}_\perp\delta\mathbf{A}\mathbf{A}^{\#} - \mathbf{A}^{\#}\delta\mathbf{A}\mathbf{P}_\perp + \mathcal{O}_p(\|\delta\mathbf{A}\|^2).$$

Assuming that $\delta\mathbf{A} = \mathcal{O}_p(\frac{1}{\sqrt{n}})$, which is valid if $\widehat{\mathbf{A}}_n$ is related to $\widehat{\mathbf{R}}_n$ through an infinitely differentiable mapping, we get:

$$\delta\mathbf{Ph} = (\widehat{\mathbf{P}} - \mathbf{P})\mathbf{h} = \mathbf{A}^{\#}\delta\mathbf{Ah} + \mathcal{O}_p(\frac{1}{n}),$$

thus leading to:

$$\begin{aligned}\mathrm{Cov}(\delta(\mathbf{P\Lambda P})\mathbf{h}) &= \mathbf{P\Lambda}\mathrm{Cov}(\delta\mathbf{Ph})\mathbf{\Lambda P} \quad (5)\\ &= \mathbf{P\Lambda A}^{\#}\mathrm{Cov}(\delta\mathbf{Ah})\mathbf{A}^{\#}\mathbf{\Lambda P} + \mathcal{O}_p(\frac{1}{n^2}). \quad (6)\end{aligned}$$

**Corollary 1** *Let* $\mathbf{\Sigma}_\infty = \frac{\gamma}{\sigma^2}\lim_{\substack{n,\to\infty\\ \frac{n}{m}\to\gamma}} m\mathrm{Cov}(\delta\mathbf{Ah})$. *Hence,*

$$\Gamma_{\mathbf{\Lambda}}(\mathrm{h}) = \sigma^2(\mathrm{I}+\gamma\mathrm{P\Lambda P})^{-1}\left(\mathrm{I}+\gamma\mathrm{P\Lambda A}^{\#}\mathbf{\Sigma}_\infty\mathrm{A}^{\#}\mathrm{\Lambda P}\right)(\mathrm{I}+\gamma\mathrm{P\Lambda P})^{-1}.$$

## 6. OPTIMAL REGULARIZING MATRIX Λ

In the following theorem, we give the closed-form expression for the optimal regularizing matrix $\mathbf{\Lambda}$ that minimizes the trace of $\Gamma_{\mathbf{\Lambda}}(\mathbf{h})$.

**Theorem 4** *Assuming that the rank of* $\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}$ *is equal to the rank of* $\mathbf{A}$, *the optimal regularizing matrix* $\mathbf{\Lambda}$ *is given by:*

$$\mathbf{\Lambda} = \left(\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\right)^{\#}.$$

**Proof** : To determine the optimal regularizing matrix $\mathbf{\Lambda}$, we optimize the asymptotic channel estimation error first with respect to the eigenvalues of $\mathbf{P\Lambda P}$ and then with respect to the eigenvectors.

Optimization with respect to the eigenvalues

Consider the eigenvalue decomposition of $\mathbf{P\Lambda P} = \mathbf{U}^{\mathrm{H}}\mathbf{DU}$, where $\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_{N(L+1)})$. Without loss of generality, we assume that $\mathbf{P}$ has a unique zero eigenvalue, i.e, $d_1 = 0$. Then, we have:

$$\mathrm{Tr}(\Gamma_{\mathbf{\Lambda}}(\mathbf{h})) = \sum_{i=2}^{N(L+1)} \frac{1 + \gamma d_i^2\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}{(1+\gamma d_i)^2}.$$

Through simple calculations, we can find that the derivative of $\mathrm{Tr}(\Gamma_{\mathbf{\Lambda}}(\mathbf{h}))$ is given by:

$$\frac{\partial\mathrm{Tr}(\Gamma_{\mathbf{\Lambda}}(\mathbf{h}))}{\partial d_i} = \sum_{i=2}^{N(L+1)} \frac{d_i\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i} - 1}{(1+\gamma d_i)^3}.$$

Hence, the gradient is equal to zero when $d_i = \frac{1}{\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}$. Moreover, one can easily show that in this case, the Hessian matrix is strictly positive.

Consequently, the optimal eigenvalues of $\mathbf{P\Lambda P}$ are given by

$$d_i = \frac{1}{\left[\mathbf{UA}^{\#}\Sigma_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}.$$

One can note that under the assumption that the rank of $\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}$ is equal to that of $\mathbf{A}$, we have

$$\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i} \neq 0 \qquad \forall\, i \geq 2.$$

Optimization with respect to the eigenvector basis

In the following we prove that the optimal basis of eigenvectors is the one that makes $\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}$ diagonal. But before tackling the proof, we shall recall the following results:

**Definition 1** *[9] For two vectors* $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$ *with descending ordered components* $x_1 \geq x_2 \cdots \geq x_n \geq 0$ *and* $y_1 \geq y_2 \cdots \geq y_n \geq 0$, *we say that* $\mathbf{x}$ *is weakly majorized by* $\mathbf{y}$ *and write* $\mathbf{x} \preceq_w \mathbf{y}$ *when:*

$$\sum_{k=1}^{m} x_k \leq \sum_{k=1}^{m} y_k \text{ for all } m = 1, \cdots, n.$$

**Theorem 5** *[9] Let* $\mathbf{A}$ *be a* $n \times n$ *hermitian matrix. Then the descending ordered vector of diagonal entries of* $\mathbf{A}$ *is weakly majorized by the descending ordered vector of eigenvalues.*

**Definition 2** *[9] A real-valued function* $f$ *defined on* $\mathbb{R}^n$ *is said to be Schur-convex (resp. Schur-concave) on* $\mathbb{R}^n$ *if* $x \preceq_w y \Longrightarrow f(x) \leq f(y)$ *(resp. if* $x \preceq_w y \Longrightarrow f(x) \geq f(y)$ *)*

**Proposition 1** *If* $g : \mathbb{R} \to \mathbb{R}$ *is convex (resp. concave) then*

$$f(x_1, \cdots, x_n) = \sum_{k=1}^{n} g(x_k)$$

*is Schur-convex (resp. Schur-concave) on* $\mathbb{R}$.

Substituting $d_i$ by their optimal values, the asymptotic estimation error becomes:

$$\mathrm{Tr}\Gamma_{\mathbf{\Lambda}}(\mathbf{h}) = \sum_{i=2}^{N(L+1)} \frac{1}{1 + \frac{\gamma}{\left[\mathbf{UA}^{\#}\Sigma_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}}. \quad (7)$$

By proposition 1, $\mathbf{x} := (x_1, \cdots, x_{N(L+1)}) :\to \sum_{i=1}^{N(L+1)} \frac{1}{1+\frac{\gamma}{x_i}}$ is Schur-concave since $x \to \frac{1}{1+\frac{\gamma}{x}}$ is concave. Consider the eigenvalue decomposition of $\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#} = \mathbf{V}\Delta\mathbf{V}^{\mathrm{H}}$ then, according to theorem 5, the vector of the diagonal elements of $\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}$ is weakly majorized by the diagonal elements of $\Delta = \mathbf{V}^{\mathrm{H}}\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{V}$. Using the definition of Schur-Concave functions, we conclude that the minimum of the channel estimation error is achieved when $\mathbf{V} = \mathbf{U}^{\mathrm{H}}$. Consequently, the optimal regularizing matrix $\mathbf{\Lambda} = \left(\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\right)^{\#}$.

## 7. APPLICATION: SUBSPACE SEMI-BLIND OPTIMAL REGULARIZED ESTIMATOR

In this section, we consider the case when the blind subspace criterion is chosen. For more details on blind susbspace methods, the reader could refer to [10] and [11].

### 7.1. Determination of the expression of $\mathbf{\Sigma}_\infty$

The blind subspace estimator is defined as:

$$\widehat{\mathbf{h}} = \min_{\|\mathbf{f}\|=1} \mathbf{f}^{\mathrm{H}} \widehat{\mathbf{Q}}_n \mathbf{f},$$

with

$$\widehat{\mathbf{Q}}_n = \boldsymbol{\mathcal{D}}(\widehat{\mathbf{\Pi}}_n)^{\mathrm{H}} \boldsymbol{\mathcal{D}}(\widehat{\mathbf{\Pi}}_n),$$

where $\widehat{\mathbf{\Pi}}_n$ being the estimated noise projector of the autocovariance matrix and $\boldsymbol{\mathcal{D}}$ the operator given by:

$$\begin{array}{rcl} \boldsymbol{\mathcal{D}} : \mathcal{M}_{N(L+1)\times N(L+1)}(\mathbb{C}) & \rightarrow & \mathcal{M}_{N(M+1)(L+M+1)\times N(L+1)}(\mathbb{C}) \\ \mathbf{\Pi} = [\pi_0, \cdots, \pi_M] & \mapsto & \begin{pmatrix} \pi_0 & & 0 \\ \vdots & \ddots & \\ \pi_M & & \pi_0 \\ & \ddots & \vdots \\ 0 & & \pi_M \end{pmatrix}. \end{array}$$

**Theorem 6** *[5] The covariance of $\delta \mathbf{Q}\mathbf{h}$ is given by:*

$$\mathrm{Cov}\,(\delta \mathbf{Q}\mathbf{h}) = \frac{\sigma^2}{n} \boldsymbol{\mathcal{M}}(\mathbf{h}) + \mathcal{O}(\frac{1}{n^2}),$$

*where*

$$\boldsymbol{\mathcal{M}}(\mathbf{h}) = \sum_{|\tau|\leq M} \boldsymbol{\mathcal{D}}^{\mathrm{H}}(\Pi) \left[ \left( \boldsymbol{\mathcal{R}}_{\Pi_M}(\tau) + \sigma^2 \boldsymbol{\mathcal{R}}_{\Phi_M}(\tau) \right) \otimes \boldsymbol{\mathcal{R}}_{\Pi}(\tau) \right] \boldsymbol{\mathcal{D}}(\Pi),$$

*and*

$$\boldsymbol{\mathcal{R}}_{\mathbf{\Pi}}(\tau) \triangleq \mathbf{\Pi} \mathbf{J}_r^{\tau N} \mathbf{\Pi}, \quad (8)$$

$$\boldsymbol{\mathcal{R}}_{\mathbf{\Pi}_M}(\tau) \triangleq \mathbf{\Pi}_M^{\mathrm{T}} \mathbf{J}_p^{\tau} \mathbf{\Pi}_M^*, \quad (9)$$

$$\boldsymbol{\mathcal{R}}_{\mathbf{\Phi}_M}(\tau) \triangleq \mathbf{\Phi}_M^{\mathrm{T}} \mathbf{J}_r^{\tau N} \mathbf{\Phi}_M^*, \quad (10)$$

$$\mathbf{\Pi}_M \triangleq \boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}} \left[ \boldsymbol{\mathcal{I}}_M(\mathbf{h}) \boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}} \right]^{\#} \boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}}, \quad (11)$$

$$\mathbf{\Phi}_M \triangleq \left[ \boldsymbol{\mathcal{I}}_M(\mathbf{h}) \boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}} \right]^{\#} \boldsymbol{\mathcal{I}}_M(\mathbf{h}), \quad (12)$$

*where* $\mathbf{J}_p(\tau) \triangleq \begin{bmatrix} \vdots & \mathbf{I}_{p-\tau} \\ \mathbf{0}_\tau & \cdots \end{bmatrix}$ *for* $\tau \geq 0$ *and* $\mathbf{J}_p^{\tau} \triangleq \left( \mathbf{J}_p^{-\tau} \right)^{\mathrm{T}}$ *if* $\tau < 0$. *Moreover, we have in this case:*

$$\mathbf{\Sigma}_\infty = \boldsymbol{\mathcal{M}}(\mathbf{h}).$$

### 7.2. Practical implementation

From the expression of $\mathbf{\Sigma}_\infty$, we note that the optimal value of $\mathbf{\Lambda}$ depends on the channel statistics which are expressed through $\boldsymbol{\mathcal{R}}_{\mathbf{\Pi}(\tau)}$ and also on the current channel value via the term $\boldsymbol{\mathcal{R}}_{\mathbf{\Phi}_M}$ and $\boldsymbol{\mathcal{R}}_{\mathbf{\Pi}_M}(\tau)$. Assuming that the channel $\mathbf{h}$ has no zeros in common, we can assume that $\mathbf{\Pi}_M$ is equal to the identity matrix thus removing the dependence of the term $\boldsymbol{\mathcal{R}}_{\mathbf{\Pi}_M}(\tau)$ on the channel. To deal with the term $\boldsymbol{\mathcal{R}}_{\mathbf{\Phi}_M}(\tau)$, the work in [5] suggests to substitute the unkown channel vector by an-other estimate, while the work in [7] and [6] proposes to just remove it since it is of order $\sigma^4$ in the expression of the asymptotic channel estimation error and thus could be removed as far as high SNR values are considered.

It has been shown that even for low SNR values, regularization is quite well performed while considering this assumption. Therefore, in this article, we will assume that:

$$\mathbf{\Sigma}_\infty \simeq \boldsymbol{\mathcal{D}}^{\mathrm{H}}(\Pi)\, \mathbf{M}(\mathbf{h}) \boldsymbol{\mathcal{D}}(\Pi), \quad (13)$$

where $\mathbf{M}(\mathbf{h}) = \sum_{|\tau|\leq M} \boldsymbol{\mathcal{R}}_{\Pi_M}(\tau) \otimes \boldsymbol{\mathcal{R}}_{\Pi}(\tau)$. Also, we can prove that in this case, the optimal value of $\mathbf{\Lambda}$ is given by:

$$\mathbf{\Lambda} = \mathbf{Q} \mathbf{\Sigma}_\infty^{\#} \mathbf{Q}.$$

To estimate $\mathbf{\Sigma}_\infty$, we replace $\boldsymbol{\mathcal{D}}(\Pi)$, $\boldsymbol{\mathcal{R}}_{\Pi_M}(\tau)$ and $\boldsymbol{\mathcal{R}}_{\Pi}(\tau)$ by their respective estimates obtained from the estimated covariance matrix. Since the estimate $\widehat{\mathbf{\Sigma}}_\infty$ of $\mathbf{\Sigma}_\infty$ is an ill-conditioned matrix, we compute instead the following estimate given by:

$$\mathbf{\Lambda} = \widehat{\mathbf{Q}}_n \left( \widehat{\Sigma}_\infty + \sigma^2 \mathbf{I}_{N(L+1)} \right)^{-1} \widehat{\mathbf{Q}}_n.$$

## 8. SIMULATIONS

### 8.1. Asymptotic analysis

As it has been already mentioned, the derived asymptotic results hold in the asymptotic regime defined as $n \rightarrow \infty$ and $m \rightarrow \infty$ while $\frac{n}{m} \rightarrow \gamma$. In this section, we assess the accuracy of the derived results for finite data and training periods. Our simulations are conducted in the cases of small and also large training and data periods ($n = 104$, $m = 26$ and $n = 1040$, $m = 260$).

For each case, we set the channel to a fixed value and estimate the Mean Square Error (MSE) given by :

$$\mathrm{MSE} = m \|\widehat{\mathbf{h}} - \mathbf{h}\|^2,$$

where the normalization by $m$ was introduced in order to allow the comparison between the figures pertaining to different choices for $m$ and $n$. More particularly, we estimate the empirical MSE for the least sqaure estimator and for both regularization-based estimators when the regularizing coefficient or matrix is set to its theoretical value or is estimated through the minimization of the channel matrix error. We compare these values to the theoretical MSE which is given by the trace of $\Gamma_\alpha(\mathbf{h})$ or $\Gamma_\Lambda(\mathbf{h})$, depending on the considered estimator. Figs. 1 and 2 display the obtained results. In the legend, 'Emp.MSE.conv.th.reg' and 'Emp.MSE.conv.est.reg' stand for the empirical MSE for the conventional regularized estimator when setting the regularizing coefficient to its optimal value and when estimating numerically the optimum regularizing coefficient, respectively. 'The.MSE.conv.th.reg' stands for the theoretical MSE of the conventional regularized estimator when the regularizing coefficient is set to its theoretical optimum value, and 'Emp.MSE.least.square' stands for the training-based least square estimator. Similar notation is used for the optimum regularized estimator. We note that even for relatively small system dimensions, optimizing the asymptotic results leads approximatively to the expected mean square error.

### 8.2. Bit error rate (BER)

In this section, we compare the BER performance of the least square based receiver with that of the semi-blind regularization-based estimators and that of a genie receiver which knows exactly the channel. We set the number of receiving sensors $N$ to 6 and $L$ to 4. We also assume that the number of training symbols is equal to 26 and that of data symbols is equal to 464. We consider the case when the channel coefficients are Rayleigh distributed with exponential decaying profile (the decay factor is taken to be 0.2). Fig. 3 illustrates the BER obtained over 1000 realizations when a MLSE estimation of the data symbols is performed using the Viterbi algorithm. In this context, we achieve almost the same performance as the conventional receiver and a gain of 0.8 dB over the least square estimation.

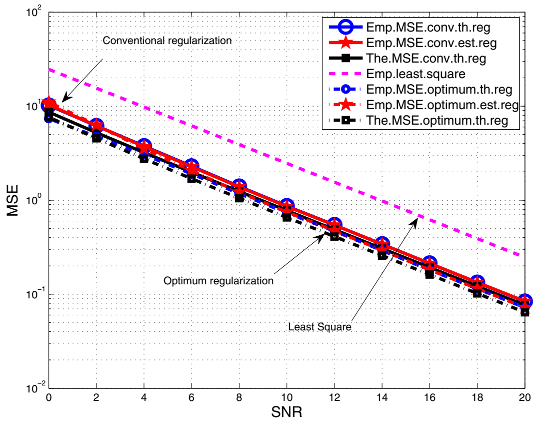


**Fig. 1**. MSE vs SNR for small system dimensions ($N = 4, n = 104, m = 26$).

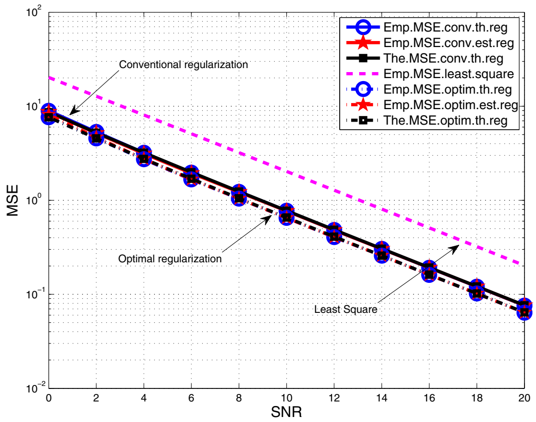


**Fig. 2**. MSE vs SNR for large system dimensions ($N = 4$, $n = 1040$, $m = 260$).

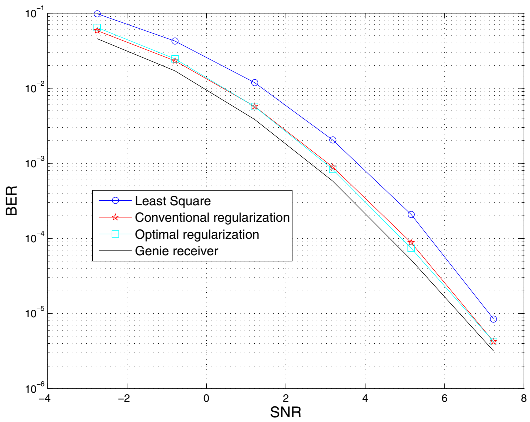


**Fig. 3**. BER vs SNR with semi-blind regularization.

### 8.3. Implementation complexity

It is mentioned in [5] that the localization of the optimal regularizing scalar $\alpha$ requires three iterations, each of which needs the inversion of a matrix of order $N(L + 1)$. However, the proposed method requires only one iteration in which one inversion matrix of order $N(L+1)$ is performed. Hence, we believe that our method exhibits lower complexity since it does not require iterations.

## 9. CONCLUSION

In this paper, we have proposed a new regularization-based method for semi-blind channel estimation. Our technique does not only slightly improve the channel estimation performance, but also exhibits low computational complexity since the setting of the optimal regularizing factor does not involve iterative computations.

## 10. REFERENCES


[1] E. De Carvalho and D. Slock, "Semi-blind Maximum Likelihood Multi-Channel Estimation with Gaussian Prior for the Symbols Using Soft Decision," *Proc. Vehic. Technol. Conf*, May 1998.

[2] H. A. Cirpan and M. K. Tsatsanis, "Stochastic Maximum Likelihood Methods for Semi-Blind Channel Equalization," *IEEE Signal Processing Lett*, vol. 5, pp. 21–24, Jan. 1998.

[3] L. Berriche and K. Abed-Meraim, "Semi-Blind Stochastic Maximum Likelihood for Frequency Selective MIMO Channel Estimation," *PIMRC*, September 2005.

[4] A. Gorokhov and P. Loubaton, "Semi-Blind Second Order Identification of Convolutive Channels," *Proc. ICASSP*, pp. 3905–3908, 1997.

[5] V. Buchoux, O. Cappé, E. Moulines, and A. Gorokhov, "On the Performance of Semi-Blind Subspace-Based Channel Estimation," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1750–1759, June 2000.

[6] S. Lasaulce, P. Loubaton, and E. Moulines, "A Semi-Blind Channel Estimation Technique Based on Second-Order Blind Method for CDMA Systems," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1894–1903, 2003.

[7] S. Lasauce, *Channel Estimation and Multiuser Detection for TD-CDMA Systems*, Ph.D. thesis, ENST Paris, 2001.

[8] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.

[9] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, New York: Academic, 1979.

[10] L. Tong and S. Perreau, "Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

[11] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A Subspace Algorithm for Certain Subspace Blind Identification Problems," *IEEE. Trans. Inform. Theory*, vol. 43, pp. 499–511, March 1997.