# Whitening-Rotation-Based Semi-Blind MIMO Channel Estimation

Aditya K. Jagannatham, *Student Member, IEEE,* and Bhaskar D. Rao, *Fellow, IEEE*

***Abstract*—This paper proposes a whitening-rotation (WR)-based algorithm for semi-blind estimation of a complex flat-fading multi-input multi-output (MIMO) channel matrix $H$. The proposed algorithm is based on decomposition of $H$ as the matrix product $H = WQ^H$, where $W$ is a whitening matrix and $Q$ is unitary rotation matrix. The whitening matrix $W$ can be estimated blind using only received data while $Q$ is estimated exclusively from pilot symbols. Employing the results for the complex-constrained Cramer–Rao Bound (CC-CRB), it is shown that the lower bound on the mean-square error (MSE) in the estimate of $H$ is directly proportional to its number of unconstrained parameters. Utilizing the bounds, the semi-blind scheme is shown to be very efficient when the number of receive antennas is greater than or equal to the number of transmit antennas. Closed-form expressions for the CRB of the semi-blind technique are presented. Algorithms for channel estimation based on the decomposition are also developed and analyzed. In particular, the properties of the constrained maximum-likelihood (ML) estimator of $Q$ for an orthogonal pilot sequence is examined, and the constrained estimator for a general pilot sequence is derived. In addition, a Gaussian likelihood function is considered for the joint optimization of $W$ and $Q$, and its performance is studied. Simulation results are presented to support the algorithms and analysis, and they demonstrate improved performance compared to exclusively training-based estimation.**

***Index Terms*—Channel estimation, constrained Cramer–Rao bound, constrained estimation, Cramer-Rao bound (CRB), iterative general maximum likelihood (IGML), multi-input multi-output (MIMO), orthogonal pilot maximum-likelihood (OPML), unitary.**

## I. Introduction

MULTI–INPUT multi-output (MIMO) and smart antenna systems are widely being studied for employment in current and upcoming wireless communication systems. Smart antenna systems, which are built with multiple antennas on receive and/or transmit side, offer a variety of gains such as improved signal-to-noise ratio (SNR) due to diversity of reception or transmission and also enhanced signal quality from interference suppression. In addition to these, MIMO systems also give the additional advantage of increased data communication rates for the same SNR by using the multiple spatial multiplexing modes available for communication.

Manuscript received April 8, 2004; revised April 8, 2005. This work was supported by CoRe research Grant Cor00-10074 and $\mathrm{Cal(IT)^2}$. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Behrouz Farhang-Boroujeny.

The authors are with the Center for Wireless Communication, University of California San Diego, La Jolla, CA 92093 USA (e-mail: ajaganna@ucsd.edu; brao@ucsd.edu).

Digital Object Identifier 10.1109/TSP.2005.862908

Channel parameters provide key information for the operation of wireless systems and hence need to be estimated accurately. As the number of data channels increases in MIMO systems, the number of associated training streams for the estimation of these channel coefficients increases proportionately. This increase in the pilot data overload results in reduced spectral efficiency. Moreover, such pilot-based techniques tend not to use the statistical information available in unknown data symbols to improve channel estimates. Semi-blind techniques can potentially enhance the quality of such estimates by making a more complete use of available data. Overhead costs can be reduced by achieving pilot-based estimation quality for smaller training symbol pay loads. With a few known training symbols along with blind statistical information, such techniques can avoid the convergence problems associated with blind techniques. The MIMO channel estimation problem is further complicated because, as the diversity of the MIMO system increases, the SNR (per bit) required to achieve the same system performance [in terms of bit error rate (BER)] decreases. The SNR at each antenna is even lower. For instance, employing binary orthogonal frequency-shift-keying (FSK) modulation and at an operation BER of $2 \times 10^{-3}$, while an SNR of 25 dB is required with a single receive antenna, an SNR of 12 dB suffices with four antennas [1]. The SNR at each antenna is even lower. Such low-SNR environments call for more training symbols thus compromising the effective data rate. Hence, more robust channel estimation techniques which use both training and blind data completely are attractive.

We utilize the fact that the MIMO channel matrix $H$ can be decomposed as the product $H = WQ^H$, where $W$ is a whitening matrix and $Q$ is a unitary matrix, i.e., $QQ^H = Q^HQ = \mathbf{I}$. It is well known that $W$ can be computed blind from the second-order statistical information in received output data. Training data can then be utilized to estimate only the unitary matrix $Q$. Significant estimation gains can then be achieved by estimation of such orthogonal matrices which are parameterized a much fewer number of parameters. A more rigorous justification of this statement is given in subsequent sections. Such a whitening-rotation (WR) factorization-based estimation procedure naturally arises in the independent component analysis (ICA)-based framework for source separation, where it has been noted that when the sources are uncorrelated Gaussian, the channel matrix can be estimated blind up to a rotation matrix. A more complete discussion of ICA can be found in [2] and [3]. A totally blind higher order statistics algorithm based on such a decomposition is elaborated in [4], for any source distribution.

Extensive work has been done by Slock *et al.* in [5] and [6] where several semi-blind techniques have been reported. More relevant literature to our semi-blind estimation scheme can be found in Pal's work [7], [8]. However, it does not consider the problem of a constrained estimator for $Q$. Our research is novel in the following aspects. First, we use the theory of complex-constrained Cramer–Rao bound (CC-CRB) reported in [9] to quantify exactly how much improvement in performance can be achieved over a traditional training-based technique. Also, since $Q$ is a unitary constrained matrix, optimal estimation of $Q$ necessitates the construction of constrained estimators. Such an estimator can be found in [10] and [11] for an orthogonal pilot sequence. We refer to this as the orthogonal pilot maximum-likelihood (OPML) estimator and examine its properties. Another salient feature of this work is the development of a novel iterative general maximum-likelihood (IGML) algorithm for the constrained estimation of $Q$ employing any (not necessarily orthogonal) pilot sequence. We then present the rotation optimized maximum-likelihood (ROML) algorithm as a low-complexity alternative to the IGML estimator.

The paper is organized as follows. The next section describes the problem setup. An analysis of the constrained CRBs is given in Section III and estimation algorithms are presented in Section IV. Finally, simulation results are given in Section V and we conclude with Section VI.

## II. Problem Formulation

Consider a flat-fading MIMO channel matrix $H \in \mathbb{C}^{r\times t}$ where $t$ is the number of transmit antennas and $r$ is the number of receive antennas in the system, and each $h_{ij}$ represents the flat-fading channel coefficient between the $i$th receiver and $j$th transmitter. Denoting the complex received data by $\mathbf{y} \in \mathbb{C}^{r\times 1}$, the equivalent base-band system can be modeled as

$$\mathbf{y}(k) = H\mathbf{x}(k) + \eta(k) \tag{1}$$

where $k$ represents the time instant, $\mathbf{x} \in \mathbb{C}^{t\times 1}$ is the complex transmitted symbol vector, and $\eta$ is spatio-temporally white additive Gaussian noise such that $\mathrm{E}\{\eta(k)\eta(l)^H\} = \delta(k,l)\sigma_n^2\mathbf{I}$ where $\delta(k,l) = 1$ if $k = l$ and 0 otherwise. Also, the sources are assumed to be spatially and temporally independent with identical source power $\sigma_s^2$ i.e., $\mathrm{E}\{\mathbf{x}(k)\mathbf{x}(l)^H\} = \delta(k,l)\sigma_s^2\mathbf{I}$. The SNR of operation is defined as $\mathrm{SNR} \triangleq (\sigma_s^2)/(\sigma_n^2)$. Now assume that the channel has been used for a total of $N$ symbol transmissions. Out of these $N$ transmissions, the initial $L$ symbols are known training symbols and the observed outputs are thus training outputs. Stacking the training symbols as a matrix we have $X_p = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(L)]$ where $X_p \in \mathbb{C}^{t\times L}$. $Y_p \in \mathbb{C}^{r\times L}$ is given by similarly stacking the received training outputs. The remaining $N-L$ information symbols transmitted are termed as "blind symbols," and their corresponding outputs as "blind outputs." $X_b \in \mathbb{C}^{t\times N-L}, Y_b \in \mathbb{C}^{r\times N-L}$ can be defined analogously for the blind symbols. $\{[X_p, Y_p], Y_b\}$ is the complete available data.

Consider two possible estimation strategies. $H$ can be estimated exclusively using the pilot $X_p$ given as

$$\hat{H}_{\mathrm{TS}} = Y_p X_p^\dagger \tag{2}$$

where $X_p^\dagger$ denotes the Moore–Penrose pseudo-inverse of $X_p$. This qualifies as training-based estimation and is simple to implement. However, it results in poor usage of available bandwidth since the pilot itself conveys no source information. Alternatively, $H$ may be estimated from blind data without the aid of any pilot. Thus, in effect this reduces to the case $L = 0$ and only blind data $Y_b$ is available. This is very efficient in usage of bandwidth since it totally eliminates the need for a pilot. However, most second-order statistics-based blind techniques are limited to estimating the channel matrix up to a scaling and permutation indeterminacy as detailed in [3] and [12]. Blind methods that employ higher order statistics typically require a large number of data symbols. Moreover, such techniques are often computationally complex and result in ill-convergence. Based on the above observations, one is motivated to find a technique which performs reasonably well in terms of bandwidth efficiency and computational complexity. Moreover, pilot symbols are usually feasible in communication scenarios. Hence, the focus of our work has been to develop a semi-blind estimation procedure which uses a small number of pilot symbols along with blind data. Such a procedure serves the dual purpose of reducing the required pilot overhead at the same time achieving a greater estimation accuracy for a given number of pilot symbols.

Consider a MIMO channel $H \in \mathbb{C}^{r\times t}$ which has at least as many receive antennas as transmit antennas i.e., $r \geq t$. Then, the channel matrix $H$ can be decomposed as $H = WQ^H$, where $W \in \mathbb{C}^{r\times t}$ and $Q \in \mathbb{C}^{t\times t}$ is unitary i.e., $Q^HQ = QQ^H = \mathbf{I}$. The matrix $W$ is popularly termed as the whitening matrix.[1] $Q$ induces a rotation on the space $\mathbb{C}^{t\times 1}$ and is therefore known as the rotation matrix. For instance, consider the singular value decomposition (SVD) of $H$ given as $H = P\Sigma V^H$. A possible choice for $W, Q$, which is employed in subsequent portions of this work, is given by

$$W = P\Sigma, \quad \text{and} \quad Q = V. \tag{3}$$

It then becomes clearly evident that all such matrices $W$ satisfy the property $WW^H = HH^H$ and it is well known that $W$ can be determined from the blind data $Y_b$. $Q$ can then be exclusively determined from $X_b$. This semi-blind estimation procedure is termed as a WR scheme. Such a technique potentially improves estimation accuracy because the matrix $Q$ by virtue of its unitary constraint is parameterized by a fewer number of parameters and hence can be determined with greater accuracy from the limited pilot data $X_p$. The precise improvement in quantitative terms is presented in the next section.

To avoid repetition, we present here a list of assumptions which may be potentially employed in our work. The exact subset of assumptions used will be stated specifically in the result.

A1) $W \in \mathbb{C}^{r\times t}$ is perfectly known at the output.

A2) $X_p \in \mathbb{C}^{t\times L}$ is orthogonal i.e., $X_pX_p^H = \sigma_s^2 L\,\mathbf{I}_{t\times t}$.

A1) is reasonable if we assume the transmission of a long data stream from which $W$ can be estimated with considerable accuracy, and A2) can be easily achieved for signal constellations such as the binary-phase-shift keying (BPSK), quadra-

[1]If $\mathbf{a} \in \mathbb{C}^{t\times 1}$ is a random vector such that $\mathrm{E}\{\mathbf{a}\mathbf{a}^H\} = \mathbf{I}$ and $\mathbf{b} \in \mathbb{C}^{r\times 1}$ is obtained by transforming $\mathbf{a}$ as $\mathbf{b} = H\mathbf{a}$, then $W$ can be employed to decorrelate or whiten $\mathbf{b}$ as $\mathbf{c} = W^\dagger\mathbf{b}$, i.e., $\mathrm{E}\{\mathbf{c}\mathbf{c}^H\} = \mathbf{I}$.

ture-phase-shift keying (QPSK), etc., by using an integer orthogonal structure such as the Hadamard matrix.

## III. Estimation Accuracy for Semi-Blind Approaches

We now present a general result to quantify the improvement in estimation accuracy of semi-blind schemes over training-based channel estimators. The CRB is frequently used as a framework to study the estimation efficiency. However, semi-blind approaches involve estimation of constrained complex parameter vectors. Therefore, in our analysis, we use the CC-CRB framework developed in [9], inspired by the result in [13], which provides an ideal setting to study the performance of such schemes. However, from the CRB matrices, which describe a lower bound on the estimation covariance, it is harder to interpret the achievable estimation accuracy in quantitative terms. This necessitates the development of a positive scalar measure to evaluate and contrast the performance of different estimators. Frequently, the trace of the covariance or the MSE in estimation is used to quantify the performance of an estimator. We next present a result which justifies the use of such a positive scalar measure.

*Lemma 1:* Let $A, B \in \mathbb{C}^{n \times n}$ be positive definite matrices and let $A \geq B$ i.e., $\mathbf{u}^H A \mathbf{u} \geq \mathbf{u}^H B \mathbf{u}, \forall \mathbf{u} \in \mathbb{C}^{n \times 1}$. Then $\mathrm{tr}(A) = \mathrm{tr}(B) \Leftrightarrow A = B$.

*Proof:* It is easy to see that $A = B \Rightarrow \mathrm{tr}(A) = \mathrm{tr}(B)$. To prove the converse, observe that $A \geq B \Rightarrow G = A - B \geq 0$ and hence $G$ is positive semi-definite (PSD). Further $\mathrm{tr}(A) = \mathrm{tr}(B) \Rightarrow \mathrm{tr}(G) = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i = 0$ where $\lambda_i$ are the eigenvalues of $G$. However, $G$ is PSD, and hence $\lambda_i \geq 0, \quad \forall i$. Therefore, $\lambda_i = 0, \quad \forall i \Rightarrow G = 0 \Rightarrow A = B$. □

Setting $A, B$ to be the error covariance and the covariance lower bound (obtained from the CRB analysis), respectively, it is easy to see that if the trace of the covariance approaches the trace of the bound, then the covariance itself approaches the bound. Thus, given the estimation error matrix $E \triangleq \hat{H} - H$, it is reasonable to consider the mean of the squared Frobenius norm of $E$ given by $\mathrm{E}\{\|E\|_F^2\} = \mathrm{E}\{\mathrm{tr}(EE^H)\}$, as a performance measure. We now present a central result which relates the MSE of estimation to the number of unconstrianed parameters in $H$.

*Lemma 2:* Under A2), the minimum estimation error in $H$ is directly proportional to $\mathcal{N}_\theta$ the number of unconstrained real parameters required to describe $H$ and in fact

$$\mathrm{E}\left\{\|\hat{H} - H\|_F^2\right\} \geq \frac{\sigma_n^2}{2\sigma_s^2 L} \mathcal{N}_\theta. \tag{4}$$

*Proof:* $H$ is an $r \times t$ dimensional matrix and therefore has $2rt$ real parameters. Let parameter vector $\bar{\gamma}$ be defined as $\bar{\gamma} \triangleq [\mathrm{vec}(H^T)^T, \mathrm{vec}(H^H)^T]$ where $\mathrm{vec}(H)$ denotes a stacking of of the columns of $H$ as $\mathrm{vec}(H) = [h_1^T, h_2^T, \ldots, h_t^T]^T$ and $h_i$ denotes the $i$th column of $H$ for $1 \leq i \leq t$. Since we are concerned with a constrained parameter estimation problem, we wish to employ the CC-CRB. For this purpose, we will need to redefine the following notation. Let the extended set of constraints on $\bar{\gamma}$ be given as $\mathbf{f}(\bar{\gamma}) = 0$ such that $\mathbf{f}(\bar{\gamma}) \in \mathbb{F}^{k \times 1}$, where $\mathbb{F}$ is the space of functions $f$ such that $f : \mathcal{C}^{2rt} \rightarrow \mathcal{R}$. Let $F(\bar{\theta}) \in \mathbb{C}^{k \times 2rt}$ be defined as $F(\bar{\theta}) \triangleq (\partial \mathbf{f}(\bar{\gamma}))/(\partial \bar{\gamma})$. Thus, there exists a matrix $U$ such that the columns of $U$ form an orthonormal basis of the null space of $F(\bar{\theta})$.

Since the number of unconstrained parameters in $H$ is $\mathcal{N}_\theta$, the number of constraints on the system is given as $2rt - \mathcal{N}_\theta$. This can be seen as follows. Let the elements of $H$ be stacked as $\bar{\delta} \triangleq [\mathrm{vec}(\mathrm{Re}(H))^T, \mathrm{vec}(\mathrm{Im}(H))^T]^T \in \mathbb{R}^{2rt \times 1}$. Define $\bar{\zeta} \triangleq [\zeta_1, \zeta_2, \ldots, \zeta_{\mathcal{N}_\theta}]^T$ as the vector of the unconstrained parameters $\zeta_i, 1 \leq i \leq \mathcal{N}_\theta$. Let the parametric representation of the elements of $\bar{\delta}$ be given as $\delta_j \triangleq \chi_j(\bar{\zeta}), 1 \leq j \leq 2rt$, and $\chi_j : \mathbb{R}^{\mathcal{N}_\theta \times 1} \rightarrow \mathbb{R}$. Let $\tilde{\delta} \triangleq [\delta_1, \delta_2, \ldots, \delta_{\mathcal{N}_\theta}]^T$. Define the vector function $\bar{\chi}$ as $\bar{\chi} \triangleq [\chi_1, \chi_2, \ldots, \chi_{\mathcal{N}_\theta}]^T$. Therefore, $\bar{\chi} : \mathbb{R}^{\mathcal{N}_\theta \times 1} \rightarrow \mathbb{R}^{\mathcal{N}_\theta \times 1}$ as $\bar{\chi}(\bar{\zeta}) = \tilde{\delta}$. Now, by the inverse function theorem [14], under mild conditions[2] on $\bar{\chi}$, there exists an inverse function $\bar{\bar{\chi}} : \mathbb{R}^{\mathcal{N}_\theta \times 1} \rightarrow \mathbb{R}^{\mathcal{N}_\theta \times 1}$ such that $\bar{\bar{\chi}}(\tilde{\delta}) = \bar{\zeta}$. The $2rt - \mathcal{N}_\theta$ constraints on the parameter vector $\bar{\delta}$ and in turn on the elements of $H$ are then obtained by the constraint equations

$$\chi_j(\bar{\bar{\chi}}(\tilde{\delta})) - \delta_j = 0, \quad \mathcal{N}_\theta + 1 \leq j \leq 2rt. \tag{5}$$

Therefore, $\mathrm{rank}(F(\bar{\theta})) = 2rt - \mathcal{N}_\theta$, the number of nonredundant constraints. It follows that $U \in \mathbb{C}^{2rt \times \mathcal{N}_\theta}$. From [9], the CC-CRB for the estimation of $\bar{\gamma}$ is given as

$$\mathrm{E}\{(\hat{\bar{\gamma}} - \bar{\gamma})(\hat{\bar{\gamma}} - \bar{\gamma})^H\} \geq U(U^H J U)^{-1} U^H \tag{6}$$

where $J$ is the unconstrained complex Fisher information matrix (FIM). $J$ for the above scenario is then given as $J = (\sigma_s^2 L)/(\sigma_n^2) \mathbf{I}_{2rt \times 2rt}$ [15]. Substituting this expression for $J$ in (6) and considering the trace of resulting matrices on both sides as justified by Lemma 1, we have

$$\begin{aligned} \mathrm{tr}(\mathrm{E}\left\{(\hat{\bar{\gamma}} - \bar{\gamma})(\hat{\bar{\gamma}} - \bar{\gamma})^H\right\}) &\geq \mathrm{tr}((U^H J U)^{-1}) \\ \mathrm{E}\left\{2\|\hat{H} - H\|_F^2\right\} &\geq \mathrm{tr}\left(\frac{\sigma_n^2}{\sigma_s^2 L} \mathbf{I}_{\mathcal{N}_\theta \times \mathcal{N}_\theta}\right) \qquad (7) \\ &= \frac{\sigma_n^2}{\sigma_s^2 L} \mathcal{N}_\theta \\ \mathrm{E}\left\{\|\hat{H} - H\|_F^2\right\} &\geq \frac{\sigma_n^2}{2\sigma_s^2 L} \mathcal{N}_\theta. \qquad (8) \end{aligned}$$

□

Thus, the above result validates the claim that the estimation of a matrix with fewer unconstrained parameters, i.e., a constrained matrix, can result in a significant improvement in estimation accuracy. We next examine the significance of the result in Lemma 2 as applied to the WR-based semi-blind algorithm.

### A. Estimation Accuracy of the WR Scheme

The following result, which compares the lower bounds of estimation errors of the training-based and WR schemes, gives critical insight into the estimation accuracy of the proposed semi-blind scheme.

*Lemma 3.1:* Under assumptions A1) and A2), the potential gain of the semi-blind algorithm (in decibels) in terms of MSE of estimation is $10 \log_{10}((2r/t))$.

*Proof:* Under A1), since $W$ is perfectly known, it suffices to estimate the unitary matrix $\hat{Q}$ to estimate the channel matrix as $\hat{H} = W\hat{Q}$. From [16], the number of real parameters required to parameterize $Q$ which under A1) equals the number of unconstrained parameters in $H$ is given as $\mathcal{N}_Q = t^2$. However, the general matrix $H$ has $\mathcal{N}_H = 2rt$ un-constrained real

[2]Existence of inverse function requires the derivative $\bar{\chi}$ be continuous and the linear operator $\bar{\chi}'$ be invertible. A rigorous formulation can be found in [14].

parameters. Hence, from the result in Lemma 2 the estimation gain in decibels of the semi-blind scheme, which estimates the constrained unitary matrix rather than the complex matrix $H$ is given by

$$\mathcal{G} = 10\log_{10}\left(\frac{\mathcal{N}_H}{\mathcal{N}_Q}\right)\, dB = 10\log_{10}\left(\frac{2r}{t}\right) \text{ dB} \quad (9)$$

which completes the proof. □

Two advantages of the WR scheme can be seen from the above result.

1) In the case when the number of receivers equals the number of transmitters i.e., $r = t$, the algorithm can potentially perform 3 dB more efficiently than estimating $H$ directly.
2) The estimation gain progressively increases as $r$, the number of receive antennas, increases. This can be expected since as $r$ increases, the complexity of estimating $H$ (size $r \times t$) increases while that of $Q$ (size $t \times t$) remains constant.

Thus for a size $8 \times 4$ complex channel matrix H, i.e., $H \in \mathbb{C}^{8\times 4}$, the estimation gain of the semi-blind technique is 6 dB which represents a significant improvement over the conventional technique described in (2).

### B. Constrained CRB of the WR Scheme

An exact expression is now derived for the variance bound in each element of $H$. To begin with, we assume that only A1) holds. Let the channel matrix be factorized using its singular value decomposition (SVD) as $H = P\Sigma Q^H$ where $P \in \mathbb{C}^{r\times t}, Q \in \mathbb{C}^{t\times t}$ are orthogonal matrices such that $P^HP = \mathbf{I}_{r\times r}, Q^HQ = \mathbf{I}_{t\times t}$, $\Sigma = \text{diag}(\sigma_1,\sigma_2,\ldots,\sigma_t)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_t > 0$. As seen earlier in (3), $W$ can be given as $W = P\Sigma$. Let $\mathbf{q}_i$ for $1 \leq i \leq t$ be the columns of $Q$. Define the desired parameter vector to be estimated $\bar{\rho} \triangleq [\text{vec}(Q)^T, \text{vec}(Q^*)^T]^T = [\mathbf{q}_1^T,\mathbf{q}_2^T,\ldots,\mathbf{q}_t^T,\mathbf{q}_1^H,\mathbf{q}_2^H,\ldots,\mathbf{q}_t^H]^T$. It can then be seen that $\bar{\rho}$ is a constrained parameter vector and the constraints are given as

$$\mathbf{q}_i^H\mathbf{q}_i = 1, \quad 1 \leq i \leq t \quad (10)$$
$$\mathbf{q}_i^H\mathbf{q}_j = 0, \quad 1 \leq i < j \leq t. \quad (11)$$

Let $U_f \in \mathbb{C}^{2t^2\times t^2}$ be defined as

$$U_f = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{q}_1 & 0 & \mathbf{q}_2 & 0 & \mathbf{q}_3 & \cdots \\ 0 & \mathbf{q}_1 & 0 & \mathbf{q}_2 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ -\mathbf{q}_1^* & -\mathbf{q}_2^* & 0 & 0 & 0 & \cdots \\ 0 & 0 & -\mathbf{q}_1^* & \mathbf{q}_2^* & 0 & \cdots \\ 0 & 0 & 0 & 0 & -\mathbf{q}_1^* & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (12)$$

From [9], $C_Q$, the CC-CRB for the estimation error of $\bar{\rho}$, can be obtained as

$$C_Q = U_f\left(U_f^HJU_f\right)^{-1}U_f^H \quad (13)$$

and the Fisher information matrix $J \in \mathbb{C}^{2t^2\times 2t^2}$ for the unconstrained case is given by the block diagonal matrix $J = (1)/(\sigma_n^2)(\mathbf{I}_{2\times 2}\otimes\Sigma^2\otimes X_pX_p^H)$. Block partitioning $C_Q$ as

$$C_Q = \begin{bmatrix} C_{Q_{11}} & C_{Q_{12}} \\ C_{Q_{21}} & C_{Q_{22}} \end{bmatrix} \quad (14)$$

the CRB for the estimation of $\bar{\omega} = \text{vec}(Q)$ is given by $C_{Q_{11}}$. Let $\bar{\theta} = \text{vec}(H^T)$ and $\Gamma = W\otimes\mathbf{I}_{t\times t}$. We then have $\bar{\theta} = \Gamma\bar{\omega}$. Hence, from the property of the CRB under transforms [15] the error covariance of estimation of the channel matrix $H$ is then given as

$$\text{E}\{(\hat{\bar{\theta}}-\bar{\theta})(\hat{\bar{\theta}}-\bar{\theta})^H\} \geq \Gamma C_{Q_{11}}\Gamma^H. \quad (15)$$

Equation (15) gives the bound for a general pilot $X_p$. In addition, if A2) holds, then from [15] it follows that $J = (\sigma_s^2L)/(\sigma_n^2)(\mathbf{I}_{2\times 2}\otimes\Sigma^2\otimes\mathbf{I})$ and is therefore diagonal. Further it can be verified that $U^HJU$ is also diagonal and is given as $U^HJU = (\sigma_s^2L)/(\sigma_n^2)(1/2)\tilde{\Sigma}$ where $\tilde{\Sigma} \in \mathbb{C}^{t^2\times t^2}$ is given as $\tilde{\Sigma} = \text{diag}([2\sigma_1^2, \sigma_1^2+\sigma_2^2, \sigma_2^2+\sigma_1^2, 2\sigma_2^2, \sigma_1^2+\sigma_3^2,\ldots])$. Hence, $C_{Q_{11}} = U_1(\sigma_n^2)/(\sigma_s^2L)((1/2)\tilde{\Sigma})^{-1}U_1^H$. Substituting these quantities in (15), the CRB for the estimation of $\bar{\theta}$ is obtained as

$$\begin{aligned} &\text{E}\{(\hat{\bar{\theta}}-\bar{\theta})(\hat{\bar{\theta}}-\bar{\theta})^H\} \\ &\quad \geq (P\Sigma\otimes\mathbf{I}_{t\times t})U_1\frac{\sigma_n^2}{\sigma_s^2L}\left(\frac{1}{2}\tilde{\Sigma}\right)^{-1}U_1^H(P\Sigma\otimes\mathbf{I}_{t\times t})^H \\ &\quad = C_H. \end{aligned} \quad (16)$$

The variance of the $(k,l)$ element of $H$ is obtained as $C_H((k-1)t+l,(k-1)t+l)$

$$\begin{aligned} &\text{E}\{|\hat{H}(k,l)-H(k,l)|^2\} \\ &\quad \geq C_H((k-1)t+l,(k-1)t+l) \\ &\quad = \frac{\sigma_n^2}{\sigma_s^2L}\sum_{i=1}^{t}\sum_{j=1}^{t}\frac{\sigma_i^2}{\sigma_j^2+\sigma_i^2}|P_{k,i}|^2|Q_{l,j}|^2 \end{aligned} \quad (17)$$

where $p_{k,i}, q_{j,l}$ represent the $(k,i)$ element of $P$ and $(j,l)$ element of $Q$, respectively. Thus, (18) give the variance for the estimation of each element of $H$. The weighing factor $(\sigma_i^2)/(\sigma_j^2+\sigma_i^2)$ in each term of the above summation results in the net reduction of estimation error over the training-based scheme as given in Lemma 3.

## IV. Algorithms

### A. Orthogonal Pilot ML Estimator

Under A1) and A2), $\hat{Q}$ the constrained OPML estimator of $Q$ such that $\hat{Q} : \mathbb{C}^{r\times L} \rightarrow \mathcal{S}$, where $\mathcal{S}$ is the manifold of $t\times t$ unitary matrices, is then obtained by minimizing the likelihood

$$||Y_p - WQ^HX_p||^2 \quad \text{such that } QQ^H = \mathbf{I}. \quad (18)$$

It is shown in [17] and [18] that $\hat{Q}$ under the above conditions is given by

$$\hat{Q} = V_QU_Q^H \quad \text{where } U_Q\Sigma_QV_Q^H = \text{SVD}\left(W^HY_pX_p^H\right). \quad (19)$$

The above equation thus yields a closed form expression for the computation of $\hat{Q}$, the ML estimate of $Q$. The channel matrix

$H$ is then estimated as $\hat{H} = W\hat{Q}^H$. We next present properties of the above estimator.

*1) Properties of the OPML Estimator:* In this section we discuss properties of the OPML estimator. We show that the estimator is biased and hence does not achieve the CRB for finite sample length. However, from the properties of ML estimators, it achieves the CRB asymptotically as the sample length increases. Further, it is also shown that the bound is achieved for all sample lengths at high SNR.

P1) There does not exist a finite-length constrained unbiased estimator of the rotation matrix $Q$ and hence $\hat{Q}$, the OPML estimator of $Q$, is biased.

*Proof*: Let there exist $\hat{Q}$ such that $\hat{Q} : \mathbb{C}^{r\times L} \rightarrow \mathcal{S}$ is an constrained unbiased estimator of $Q$. $\mathbb{C}^{r\times L}$ is the observation space $(Y_b)$ and $\mathcal{S}$ is the manifold of orthogonal matrices. Then $\hat{Q} = Q + \bar{E}$ where $\bar{E}$ is such that $\mathrm{E}\{\bar{E}\} = \mathbf{0}$. Now since $\hat{Q}$ is a constrained estimator we have $\hat{Q}\hat{Q}^H = \mathbf{I}$ and therefore

$$(Q + \bar{E})^H(Q + \bar{E}) = \mathbf{I}$$

which when simplified using the fact that $QQ^H = \mathbf{I}$ yields

$$Q^H\bar{E} + \bar{E}^HQ + \bar{E}\bar{E}^H = \mathbf{0}.$$

Rearranging terms in the above expression and taking the expectation of quantities on both sides (where the expectation is with respect to the distribution of $E$ conditioned on $Q$) yields

$$\mathrm{tr}(Q^H\mathrm{E}\{\bar{E}\} + \mathrm{E}\{\bar{E}\}^HQ) = -\mathrm{tr}(\mathrm{E}\{\|\bar{E}\|^2\}). \quad (20)$$

It can immediately be observed that the right-hand side is strictly less than 0 while the left-hand side is equal to zero (by virtue of $\mathrm{E}\{\bar{E}\} = \mathbf{0}$) and hence the contradiction. □

The above result then implies that the CRB cannot be achieved in a general scenario as there does not exist an unbiased estimator necessary for the achievement of the CRB. However, the properties presented next guarantee the asymptotic achievability of the CRB both in sample length and SNR.

P2) The OPML estimator achieves the CRB given in (18) as the pilot sequence length $L \rightarrow \infty$.

*Proof:* It follows from the asymptotic property of ML estimators, reviewed in [15]. □

P3) The OPML estimator of $Q$ achieves the CRB given in (18) at high SNR, i.e., as $(\sigma_s^2)/(\sigma_n^2) \rightarrow \infty$.

*Proof:* The above result can be proved using the theory of matrix eigenspace perturbation analysis detailed in [19]. The detailed proof can be found in http://dsp.ucsd.edu/aditya/pertproof.pdf □

### B. Iterative ML Procedure for General Pilot

The ML estimate of $Q$ for an orthogonal pilot $X_p$ is given by (19). In this section we present the IGML algorithm to compute the estimate for any given pilot sequence $X_p$, i.e., when A2) does not necessarily hold. As it is shown later, the proposed IGML scheme reduces to the OPML under A2). The ML cost-function to be minimized is given as in (18). Let A1) hold true and $\tilde{Y}_p \triangleq P^HY_p$. With constraints given by (10) and (11), the Lagrange cost $f(Q, \bar{\lambda}, \bar{\mu})$ to be minimized can then be formulated as

$$f(Q, \bar{\lambda}, \bar{\mu}) = \sum_{i=1}^{t} \left\| \tilde{Y}_p(i) - \sigma_i \mathbf{q}_i^H X_p \right\|^2 + \sum_{i=1}^{t} \mathrm{Re}\left\{\lambda_i \left(\mathbf{q}_i^H\mathbf{q}_i - 1\right)\right\} + \sum_{i=1}^{t}\sum_{j=i+1}^{t} \mathrm{Re}\left\{\mu_{ij}\mathbf{q}_i^H\mathbf{q}_j\right\}$$

where $\lambda_i \in \mathbb{R}, \mu_{ij} \in \mathbb{C}$ are the Lagrange multipliers, $\tilde{Y}_p(i) \in \mathbb{C}^{1\times L}$ is the i-th row (output at the i-th receiver) and $\mathbf{q}_i$ is the i-th column of $Q$ for $1 \leq i, j \leq t$. Define the matrix of Lagrange multipliers $S \in \mathbb{C}^{t\times t}$ as $S_{ii} \triangleq \lambda_i, S_{ij} \triangleq \mu_{ij}$ if $i > j$ and $S_{ij} \triangleq \mu_{ji}^*$ if $i < j$. Observe that $S$ is a Hermitian symmetric matrix i.e., $S = S^H$. The above cost function can now be differentiated with respect to $\mathrm{Re}\{\mathbf{q}_i\}, \mathrm{Im}\{\mathbf{q}_i\}$ for $1 \leq i \leq t$. These quantities can then be equated to 0 for extrema and after some manipulation, the resulting equations can be represented in terms of complex matrices as

$$X_p\tilde{Y}_p^H\Sigma - X_pX_p^HQ\Sigma^2 = QS \quad (21)$$

where $Q$ is unitary. We avoid repeated mention of this constraint in the foregoing analysis and it is implicitly assumed to hold. Let $\mathcal{A} \triangleq X_p\tilde{Y}_p^H\Sigma = X_pY_p^HW$.

$$Q^H\mathcal{A} - Q^HX_pX_p^HQ\Sigma^2 = S.$$

As noted, $S = S^H$ and therefore the Lagrange multiplier matrix $S$ can be eliminated as

$$Q^H\mathcal{A} - \mathcal{A}^HQ = Q^HX_pX_p^HQ\Sigma^2 - \Sigma^2Q^HX_pX_p^HQ. \quad (22)$$

Adding and subtracting $L\sigma_s^2\Sigma^2$ in (22) and rearranging terms yields

$$Q^H\left(\mathcal{A} + \left(L\sigma_s^2\mathbf{I}_{t\times t} - X_pX_p^H\right)Q\Sigma^2\right) = \left(\mathcal{A}^H + \Sigma^2Q^H\left(L\sigma_s^2\mathbf{I}_{t\times t} - X_pX_p^H\right)\right)Q.$$

Let $\mathcal{T} \triangleq \mathcal{A} + (L\sigma_s^2\mathbf{I}_{t\times t} - X_pX_p^H)Q\Sigma^2$. Thus, from the above equation, $Q^H\mathcal{T}$ is Hermitian symmetric or in other words $Q^H\mathcal{T} = \mathcal{T}^HQ$. Also, if $U_\mathcal{T}\Lambda_\mathcal{T}V_\mathcal{T}^H = \mathrm{SVD}(\mathcal{T})$ then, $\tilde{Q}^HU_\mathcal{T}\Lambda_\mathcal{T}V_\mathcal{T}^H = \mathrm{SVD}(Q^H\mathcal{T})$. We have then from the symmetry of $Q^H\mathcal{T}$,

$$Q^HU_\mathcal{T} = V_\mathcal{T} \Rightarrow Q = U_\mathcal{T}V_\mathcal{T}^H. \quad (23)$$

Expression (23) gives the critical step in the IGML algorithm which is succinctly presented below (some of the definitions above are repeated for the sake of completeness).

IGML Algorithm: Let A1) hold, i.e., $\hat{W} = W = P\Sigma$. $X_p$ is the transmitted pilot symbol sequence and not necessarily orthogonal. We then compute the constrained ML estimate of $\hat{Q}$ as follows.

S.1 Compute $\mathcal{A} = X_pY_p^HW$, where $Y_p$ is the received output data.

S.2 Let $\hat{Q}_0$ denote the initial estimate of the unitary matrix $Q$. Compute $\hat{Q}_0$ by employing $X_p, W$

```
and Y_p in (19).
S.3 Repeat for 𝒩 iterations. At the kth iteration i.e., 1 ≤ k ≤ 𝒩,
    S.3.1 Let T_k = A + (Lσ_s^2 I_{t×t} − X_p X_p^H) Q̂_{k−1} Σ^2.
    S.3.2 Compute refined estimate of Q_k from T_k by
          employing (23).
S.4 Finally estimate H as Ĥ = W Q̂_𝒩^H.
```

$\mathcal{N}$, the number of iterations is small and typically $\mathcal{N} \leq 5$ as found in our simulations. It can now also be noticed that if A2) holds, $X_p X_p^H = L\sigma_s^2 \mathbf{I}$. Therefore, $\mathcal{T} = \mathcal{A} = X_p Y_p^H W$. The SVD of $\mathcal{T}$ is then given by $U_{\mathcal{T}} \Lambda_{\mathcal{T}} V_{\mathcal{T}}^H = V_Q \Sigma_Q U_Q^H$. It follows that the IGML solution given as

$$\hat{Q} = U_{\mathcal{T}} V_{\mathcal{T}}^H = V_Q U_Q^H, \tag{24}$$

is similar to the solution given in (19). Thus, when $X_p$ is orthogonal, the IGML algorithm converges in a single iteration to the OPML solution.

Finally, we wish to compare the CRB of estimation of $H$ for the IGML and OPML schemes. Let $X_p$, the pilot for the IGML scheme be a random sequence such that $\mathrm{E}\{X_p X_p^H\} = L\sigma_s^2 \mathbf{I}$ or in other words it is statistically white. Denoting by $\bar{J}$ the unconstrained FIM for IGML, we have from Section III, $\bar{J} = \mathbf{I}_{2\times 2} \otimes (1)/(\sigma_n^2) X_p X_p^H$. Therefore, $\mathrm{E}\{\bar{J}\} = J$. The average CRB of IGML, where the averaging is over the distribution of $X_p$ is then given as $\mathrm{CRB}_{\mathrm{IGML}} = \mathrm{E}\{\bar{J}^{-1}\}$. Employing Jensen's inequality for matrices from [20] we have

$$\mathrm{CRB}_{\mathrm{IGML}} = \mathrm{E}\{\bar{J}^{-1}\} \geq (\mathrm{E}\{\bar{J}\})^{-1} = J^{-1} = \mathrm{CRB}_{\mathrm{OPML}}. \tag{25}$$

Thus, the error in the estimation of $H$ is minimum for an orthogonal pilot $X_p$. Similar optimality properties of orthogonal pilots have been previously reported in [21] and [22].

*1) "Rotation-Optimization" ML:* The above suggested IGML scheme to compute $\hat{Q}$ for a general pilot sequence $X_p$ might be computationally complex owing to the multiple SVD computations involved. Thus, to avoid the complexity involved in the full computation of the optimal ML solution, we propose a simplistic ROML procedure for the suboptimal estimation of $Q$, thus trading complexity for optimality. The first step of ROML involves construction of a modified cost function as

$$\min_{Q} ||\tilde{W} Y_p - Q^H X_p||^2 \quad \text{where} \;\; QQ^{\mathrm{H}} = \mathbf{I}. \tag{26}$$

$\bar{Y}_p = \tilde{W} Y_p$ is the whitening pre-equalized data. The closed form solution $\hat{Q}$ for the modified cost in (26) is given as

$$\hat{Q} = V_h U_h^H \quad \text{where} \;\; U_h S_h V_h^H = \mathrm{SVD}\left(\tilde{W} Y_p X_p^H\right) \tag{27}$$

which can be implemented with low complexity. This result for problem (26) follows by noting its similarity to problem (18). Several choices can then be considered for the pre-equalization filter $\tilde{W}$. The standard zero-forcing (ZF) equalizer is given by $\tilde{W}_{\mathrm{ZF}} = W^{\dagger}$ (where $\dagger$ denotes the Moore–Penrose pseudo-inverse) and is usually referred to as "data whitening" in literature. However, ZF is susceptible to noise enhancement as frequently cited in literature. Alternatively, a robust MMSE prefilter is given as $\tilde{W}_{\mathrm{MMSE}} = \sigma_s^2 W^H \left(\sigma_s^2 W W^H + \sigma_n^2 \mathbf{I}\right)^{-1}$.

$\hat{Q}$ given by (27) is a reasonably accurate closed form estimate of $Q$. However, the resulting estimate does not have any statistical optimality properties as it does not compute the solution to the true cost function given in (18). This estimate of $Q$ can now be employed to initialize the IGML procedure to minimize the true cost. However, to avoid the complexity associated with an SVD computation, a constrained minimization procedure (e.g., "`fmincon`" in MATLAB) can now be employed to converge to the solution with the $t^2$ nonlinear constraints given by the unit norm and mutual orthogonality of the rows of $Q$. This procedure then yields $\hat{Q}$ which is close to the optimal ML estimate and the low computational cost of the proposed solution makes it attractive to implement in practical systems.

### C. Total Optimization

This procedure builds on the above described schemes. The ML schemes (OPML and IGML) for estimating the unitary matrix are optimal given perfect knowledge of $W$. However, in finite total symbol run situations where this assumption is not valid (for example in fast fading mobile environments where the data symbols available in the channel coherence time are limited and hence the estimated whitening matrix may not be exact as assumed earlier), the disjoint estimation of the whitening matrix from blind symbols and rotation matrix from pilot symbols is not optimal. We present a scheme for such a system to iteratively compute the joint solution for $W$ and $Q$ based on minimizing a Gaussian likelihood cost function.

*1) Initialization of $W$ and $Q$:* $W$ can be estimated from the output correlation matrix $R_y$ which is given as

$$R_y = \sigma_s^2 W W^H + \sigma_n^2 \mathbf{I}. \tag{28}$$

The ML estimate of $R_y$ can be computed blindly from the entire received data $\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(N)$ as $\hat{R}_y = (1/N) \sum_{i=1}^{N} \mathbf{y}(i)\mathbf{y}(i)^H$. Using relation (28) and assuming that $\sigma_s^2$ and $\sigma_n^2$ are known at the receiver, $\hat{W}\hat{W}^H$ may be estimated as

$$\hat{H}\hat{H}^H = \frac{1}{\sigma_s^2}\left(\hat{R}_y - \sigma_n^2 \mathbf{I}\right) = \hat{W}_{\mathrm{ML}} \hat{W}_{\mathrm{ML}}^H. \tag{29}$$

$\hat{W}_{\mathrm{ML}}$ can then be computed from a Cholesky factorization of $\hat{R}_y$. $\hat{Q}$, the initial estimate of $Q$ is then computed by employing $\hat{W}$ in the OPML or IGML algorithms outlined in Sections IV-A and IV-B, respectively.

*2) Likelihood for Total Optimization:* In order to arrive at a reasonably tractable likelihood function, we now assume that the transmitted data $\mathbf{x}(k), k = L+1, \ldots, N$ is Gaussian, i.e., $\mathbf{x} \sim \mathcal{N}(0, \sigma_s^2 \mathbf{I})$. The likelihood of the complete received data, conditioned on the pilot symbols $X_p$ is given as

$$\mathcal{L}(W, Q) = \underbrace{\frac{1}{2}(N-L)\ln|\mathcal{R}(W)| - \sum_{i=L+1}^{N} \mathbf{y}(i)^H \mathcal{R}(W)^{-1} \mathbf{y}(i)}_{\mathcal{L}_1(W)} - \underbrace{\frac{1}{\sigma_n^2} \sum_{j=1}^{L} ||\mathbf{y}(j) - WQ\mathbf{x}(j)||^2}_{\mathcal{L}_2(W,Q)} \tag{30}$$

where $\mathcal{R}(W) \triangleq \sigma_s^2 WW^H + \sigma_n^2 \mathbf{I}$. $\mathcal{L}_1$ is a function entirely of blind data, and $\mathcal{L}_2$ depends only on training data. This cost function can be minimized for $W$ to compute $\hat{W}$ as given below.

Total Optimization: Let $X_p$ be the transmitted pilot symbol sequence, not necessarily orthogonal. We then compute estimates of $W$ and $Q$ matrices as follows.

T.1 Compute $\hat{W}_0$, the initial estimate of $W$ from (29).

T.2 Compute $\hat{Q}_0$ by employing $X_p, \hat{W}_0$ in the IGML algorithm in Section IV-B.

T.3 Repeat for $\mathcal{N}_T$ iterations. At the $k$th iteration, i.e., $1 \leq k \leq \mathcal{N}_T$

T.3.1 Using $\hat{W}_{k-1}$ as an initial estimate, compute $\hat{W}_k$ by minimizing $\mathcal{L}(W, \hat{Q}_k)$ ("fminunc" in MATLAB).

T.3.2 Compute the IGML estimate of $\hat{Q}_k$ from $\hat{W}_k$.

T.4 Finally estimate $H$ as $\hat{H} = \hat{W}_{\mathcal{N}_T} \hat{Q}_{\mathcal{N}_T}^H$.

It is seen from the simulation results that minimization of the above likelihood yields an improved estimate of the channel matrix $H$ even when elements of the transmitted symbol vectors $\mathbf{x}(k)$ are drawn from a discrete signal constellation. This solution however involves a computational overhead. Nevertheless it provides a useful benchmark for the estimation of the flat-fading channel matrix $H$. Practical implementation of this algorithm would require a recipe for efficient numerical computation.

As the data length $N$ increases with pilot length $L$ kept constant, the effect of $\mathcal{L}_2$ on the above expression diminishes for the estimation of $W$. Hence, for large blind data lengths $N$, maximizing the likelihood expression $\mathcal{L}$ with respect to $W$, reduces to maximization of $\mathcal{L}_1$. The solution $W$ is then given by the ML estimate in (29). The second step maximizes $\mathcal{L}_2$, which is the cost function optimized by the OPML and IGML algorithms. Thus, as $N \to \infty$, the total optimization scheme reduces to a one iteration algorithm involving the ML estimation of $\hat{W}$ followed by the constrained ML estimation of $\hat{Q}$.

## V. Simulation Results

Our simulation setup consists of an $8 \times 4$ MIMO channel $H$ (i.e., $r = 8, t = 4$). $H$ was generated as a matrix of zero-mean circularly symmetric complex Gaussian random entries such that the sum variance of the real and imaginary parts was unity. For an orthogonal pilot, the source symbol vectors $\mathbf{x} \in \mathbb{C}^{4\times 1}$ are assumed to be drawn from a BPSK constellation and the orthonormality condition is achieved by using the Hadamard structure. But otherwise, for general pilot sequences and data vectors, symbols were drawn from a 16-QAM signal constellation. Further, the transmitted for the transmitted training symbol vectors $X_p X_p^H = L\sigma_s^2 \mathbf{I}$ and for the data vectors $\mathrm{E}\{x(k)x(l)^H\} = \delta(k,l)\sigma_s^2 \mathbf{I}$, thus maintaining the source power $\sigma_s^2$ constant. Noise vectors $\eta(k)$ were generated as spatio-temporally uncorrelated complex Gaussian random vectors and with variance of each element equal to $\sigma_n^2$. The SNR of operation was measured as $\mathrm{SNR} = 10\log_{10}((\sigma_s^2)/(\sigma_n^2))$. Simulations described below investigate the performance of the proposed semi-blind algorithm under different conditions.

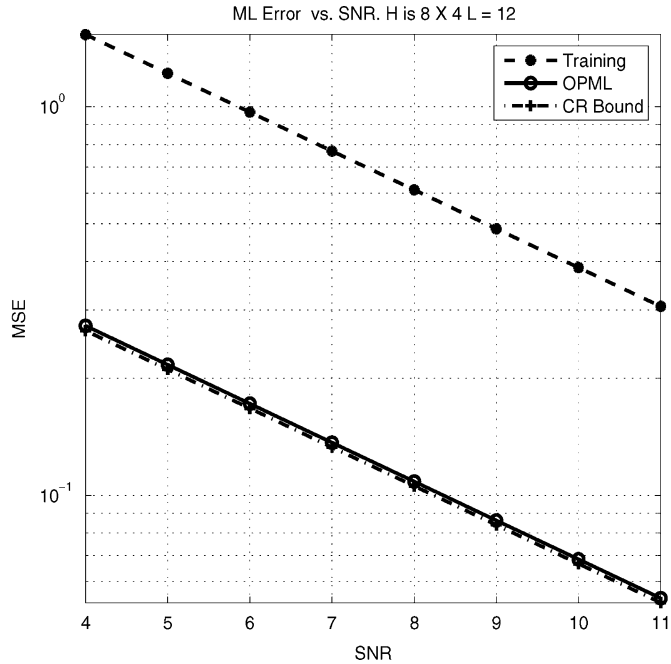


Fig. 1. MSE versus SNR of OPML semi-blind channel estimation and the semi-blind CRB with perfect knowledge of $W$. Also shown for reference is MSE of the exclusively training-based channel estimate. $H$ is an $8 \times 4$ complex flat-fading channel matrix and pilot length $L = 12$.

*Experiment 1:* In this experiment, we demonstrate the enhancement in estimation accuracy that can be achieved by the use of statistical side information (white data) as qunatified by Lemma 3. For this purpose, we evaluate the MSE performance of the different constrained ML estimators of $Q$ under A1) and compare it to the training-based estimate given by (2) which neglects white data. The MSE of estimation of the channel matrix $H$ has been averaged over 1000 instantiations of the channel noise $\eta$. In Fig. 1, this MSE has been plotted versus SNR in the range 4 dB $\leq$ SNR $\leq$ 11 dB for the OPML semi-blind scheme. As noted in Section III-A, the MSE of semi-blind scheme is 6 dB lower than that of exclusively training-based channel estimation. The CRB of the semi-blind scheme is also plotted for reference.

Next we compute the MSE for different pilot lengths $L$ in the range $20 \leq L \leq 100$. A statistically white pilot ($\mathrm{E}\{X_p X_p^H\} = L\sigma_s^2 \mathbf{I}$) was employed for the IGML, ROML and training-based schemes while an orthogonal pilot $X_p$ was used for the OPML scheme with $X_p X_p^H = L\sigma_s^2 \mathbf{I}$, thus maintaining constant source power. The left-hand side of Fig. 2 shows the error for these different schemes and also that for the exclusive training-based scheme. It can be seen that the semi-blind schemes are 6 dB more efficient than the training scheme as suggested by Lemma 3. OPML performs very close to the CRB while the IGML progressively improves toward the CRB as the pilot length increases. On the right-hand side of Fig. 2, which is a blown-up version of the same plot, it is seen that the ROML because of its suboptimality loses slightly (0.5 dB) in terms of estimation gain when compared to the other constrained estimators.

*Experiment 2:* We now consider the effect of estimation inaccuracies in $W$ arising from the availability of finite blind data.

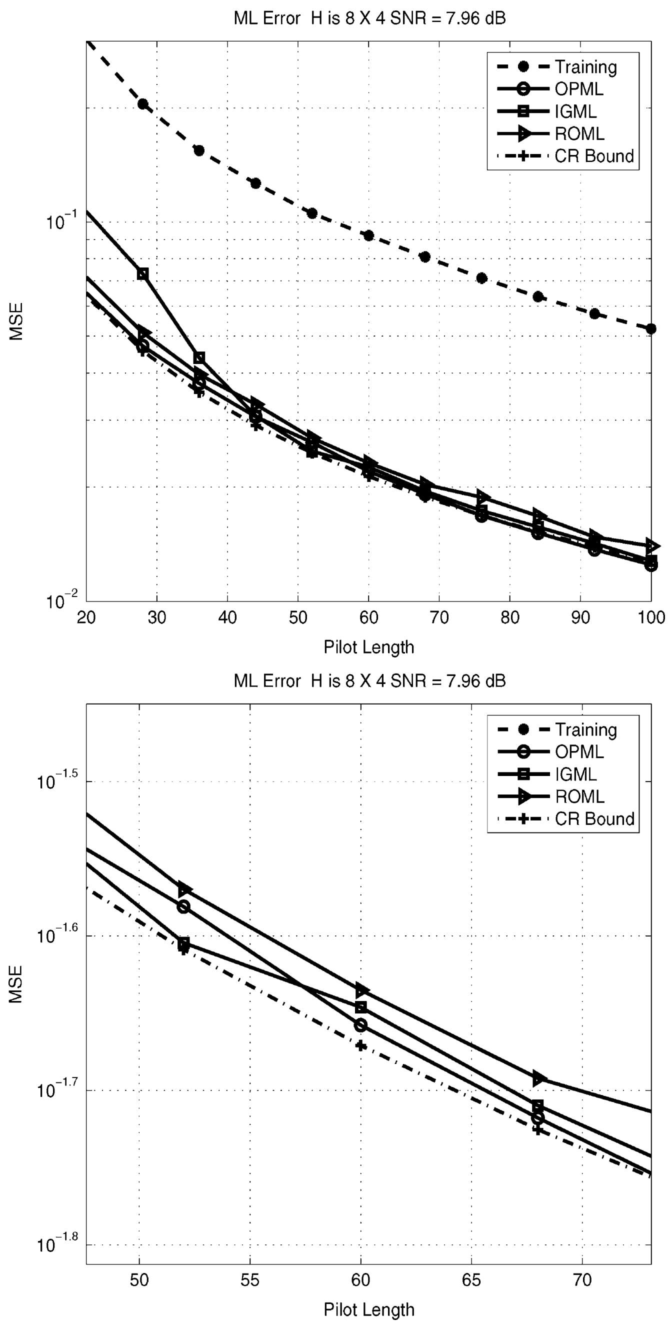


Fig. 2. Computed MSE versus pilot length ($L$) for the OPML, IGML, ROML, and exclusive training-based channel estimation. $H$ is an $8 \times 4$ complex flat-fading channel matrix and SNR $= 8$ dB.

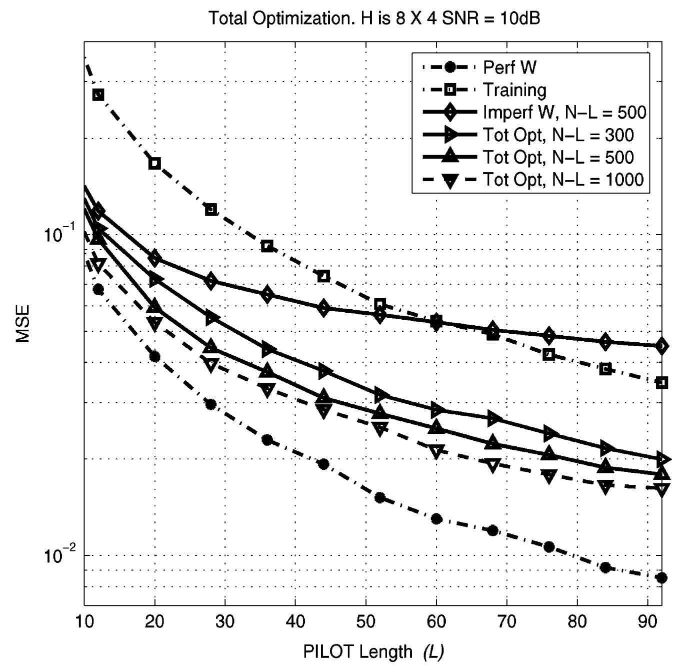


Fig. 3. Comparison of OPML with perfect $W$, OPML with imperfect or estimated $W$, total optimization, and training-based estimation of $H$.

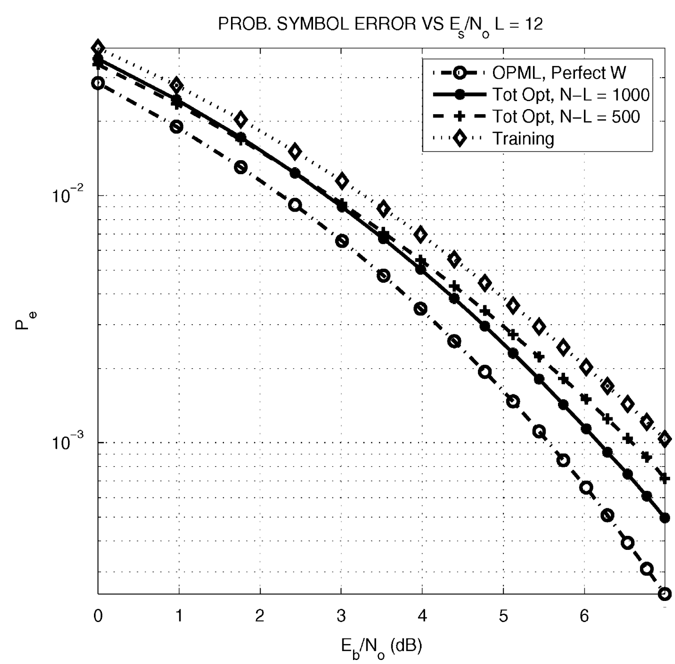


Fig. 4. Probability of bit error versus SNR for $8 \times 4$ MIMO system employing OPML; total optimization ($N = 1000, 500$). The performance of the exclusively training-based channel estimate is also given for comparison.

We demonstrate the performance of the total optimization (TotOpt) procedure for the joint optimization of $W$ and $Q$ and contrast it with the MSE of the IGML estimate with imperfect $W$. We consider estimation of $W$ from $N - L = 300$, 500, and 1000 blind data symbols with the source symbols drawn from a 16-QAM constellation and employing (29). The pilot sequence $X_p$ was orthogonal. As in the previous experiment, we consider the MSE in estimation for different pilot lengths $20 \leq L \leq 100$. It can then be seen from Fig. 3 that while the OPML with imperfect $W$ for $N - L = 500$ performs marginally better than the training sequence-based technique ($L \leq 60$ training symbols), the TotOpt scheme which optimizes the likelihood in (30) performs consistently better than the training sequence-based scheme in all the cases. Their performance is also compared to the situation of availability of perfect knowledge of $W$ (perf W), which can be seen to achieve the best performance. As noted in Section IV-C, the performance of TotOpt approaches that of the OPML with perfect $W$ as $N \rightarrow \infty$.

*Experiment 3:* Finally, we consider $P_e$ of detection of the transmitted symbol vectors employing $\hat{H}$ estimated from different schemes. We illustrate the performance of OPML with perfect knowledge of $W$ at the receiver and total optimization with $N - L = 1000$ and 500 blind symbols. The performance of the exclusively training-based estimate of $H$ is also plotted for $L = 12$. Fig. 4 shows the probability of error detection vs SNR for a linear MMSE receiver at the output for an $8 \times 4$ system $H$. It can be seen that at an SNR of 6 dB the semi-blind scheme achieves about a 1-dB improvement in probability of bit error detection performance and thus improves over the exclusively training-based estimate.

## VI. Conclusion

A semi-blind scheme based on a whitening-rotation decomposition of the channel matrix $H$ has been proposed for MIMO flat-fading channel estimation. The algorithm computes the whitening matrix $W$ blind from received data and the unitary matrix $Q$ exclusively from the pilot data. Closed-form expressions for the CRB of the proposed scheme have been derived employing the CC-CRB framework. Using the bounds, it is shown that the lower bound for the MSE in channel matrix estimation is directly proportional to the number of unconstrained parameters leading to the conclusion that the semi-blind scheme can be very efficient when the number of receive antennas is greater than or equal to the number of transmit antennas. We also develop and analyze algorithms for channel estimation based on the decomposition. Properties of the constrained ML estimator of $Q$ have been studied and an iterative constrained $Q$-estimator has been detailed for nonorthogonal pilot sequences. In the absence of perfect knowledge of $W$, a Gaussian likelihood function has been presented for the joint estimation of $W$ and $Q$. Simulation results have been presented to support the algorithms and analysis and they demonstrate improved performance compared to exclusively training-based estimation.

## References


[1] J. G. Proakis, *Digital Communications*, International ed. New York: McGraw-Hill Higher Education, 2001.

[2] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[3] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.

[4] V. Zarzoso and A. K. Nandi, "Adaptive blind source separation for virtually any source probability density function," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 477–488, Feb. 2000.

[5] E. De Carvalho and D. T. M. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Proc. 31st Asilomar Conf.*, vol. 2, 1998, pp. 1624–1628.

[6] A. Medles and D. T. M. Slock, "Semiblind channel estimation for MIMO spatial multiplexing systems," in *Proc. Vehicular Technology Conf. Fall 2001*, vol. 2, Oct. 2001, pp. 1240–1244.

[7] D. Pal, "Fractionally spaced semi-blind equalization of wireless channels," in *Proc. 26th Asilomar Conf.*, vol. 2, 1992, pp. 642–645.

[8] ——, "Fractionally spaced equalization of multipath channels: A semi-blind approach," in *Proc. 1993 Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, 1993, pp. 9–12.

[9] A. K. Jagannatham and B. D. Rao, "Cramer–Rao lower bound for constrained complex parameters," *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 875–878, Nov. 2004.

[10] A. Medles, D. T. M. Slock, and E. D. Carvalho, "Linear prediction based semi-blind estimation of MIMO FIR channels," in *Proc. 3rd IEEE Workshop Signal Processing in Wireless Communications (SPAWC)*, Taiwan, R.O.C., Mar. 2001, pp. 58–61.

[11] Y. Sung *et al.*, "Semiblind channel estimation for space-time coded WCDMA," in *Proc. 36th Asilomar Conf. Signals Systems*, 2002, pp. 1637–1641.

[12] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[13] P. Stoica and B. C. Ng, "On the Cramer–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[14] A. E. Taylor and W. R. Mann, *Advanced Calculus*, 3rd ed. New York: Wiley .

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*, 1st ed. Englewood Cliffs, NJ: Prentice-Hall PTR, 1993.

[16] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1984.

[18] A. Medles, D. T. M. Slock, and E. De Carvalho, "Linear prediction based semi-blind estimation of MIMO FIR channels," in *Proc. 3rd IEEE Workshop Signal Processing in Wireless Communications (SPAWC)*, Taiwan, R.O.C., pp. 58–61.

[19] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, 1st ed. London, U.K.: Oxford Univ. Press, 1965.

[20] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. Boston, MA: Academic, 1967.

[21] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[22] T. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Sep. 1999, pp. 22–24.



**Aditya K. Jagannatham** (S'03) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 2001 and the M.S. degree from the Department of Electrical and Computer Engineering at the University of California, San Diego, in 2003. He is currently working toward the Ph.D. degree at the University of California.

From June 2003 to September 2003, he was an intern at Zyray wireless (now Broadcom) San Diego, where he contributed to the WCDMA effort and the IEEE 802.11n high-throughput WLAN standard. His research interests are in the area of digital signal processing, statistical estimation theory and its applications in multiple-input multiple-output, OFDM, and code-division multiple-access wireless communications.

Mr. Jagannatham was awarded the Bhavesh Gandhi Memorial award from the Indian Institute of Technology for best seminar.



**Bhaskar D. Rao** (S'80–M'83–SM'91–F'00) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 1979 and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1981 and 1983, respectively.

Since 1983, he has been with the University of California at San Diego, La Jolla, where he is currently a Professor with the Electrical and Computer Engineering Department. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human–computer interactions.

Dr. Rao has been a Member of the Statistical Signal and Array Processing Technical Committee of the IEEE Signal Processing Society. He is currently a member of the Signal Processing Theory and Methods Technical Committee.