A SEMI-BLIND TECHNIQUE FOR MIMO CHANNEL MATRIX ESTIMATION

AdityaKiran Jagannatham and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093-0407
E-mail: ajaganna@ucsd.edu, brao@ece.ucsd.edu

ABSTRACT

In this work we present a semi-blind algorithm for the estimation of a flat-fading MIMO channel matrix H. The algorithm is based on a decomposition of the channel matrix H as the product of a *Whitening* matrix W and a *Unitary* matrix Q. The whitening matrix can be estimated blindly from all received data. Several techniques are then suggested to estimate the optimum rotation matrix from training samples. Since it uses both blind and training data, the algorithm is semi-blind in nature. Theoretical results show that estimation of the channel matrix based on estimating only the Q matrix from pilot data can perform more efficiently than estimating H directly from the pilot data. However, performance of the technique depends on the accuracy with which W is estimated and is found to typically perform well in low SNR and fading environments.

I. INTRODUCTION

MIMO and *smart antenna* systems are now widely being employed to combat the problems of multi-user interference, fading in wireless channels, and to achieve high data rates. As the number of input data streams increases on MIMO channels, employing entirely pilot data to learn the channel parameters would result in poorer spectral efficiency. Moreover, such techniques tend not to use the information in the 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 reduced training symbol pay loads. With few known training symbols, such techniques can avoid convergence problems associated with blind techniques.

The channel estimation problem is further complicated in multi-antenna systems because, as the diversity of the MIMO system increases, the SNR(per bit) required to achieve the same system performance(in BER terms) decreases. The SNR at each antenna is even lower. For example, an m = 4 orthogonal carrier system operating at bit error probability $P_e = 2 \times 10^{-3}$ and diversity = 1, requires an SNR of 25 dB, while at diversity = 4, the working SNR is 12 dB [1] and the SNR at each antenna could be as low as 6 db.

Such low SNR environments call for more training symbols, compromising the effective data rate. Hence, more robust channel estimation techniques which use training and blind data completely are attractive.

Work on semi-blind techniques has been reported earlier in [2] and [3] for the design of fractional semi-blind equalizers for multi-path channels. In [4] and [5] error bounds and asymptotic properties of blind and semi-blind techniques are analyzed.

In this paper, we utilize the fact that the channel matrix $\mathbf{H}$ can be decomposed as $H = WQ$, where W is a whitening matrix and Q is unitary. W can be estimated employing exclusively all the output data. Training data has to be utilized to estimate only the Q matrix. This sort of an estimation has been used in an ICA based framework for source separation, where when the sources are uncorrelated gaussian, the channel matrix can be estimated blind up to a rotation matrix. More details on the source separation problem can be found in [6].

Such a two-step estimation procedure can potentially be superior because,

- The orthonormal matrix $\mathbf{Q}$ is parameterized by a significantly lesser number of parameters than the complete channel matrix H, and hence can be estimated with greater accuracy from the limited pilot data.

- As the number of receive antennas r grows, number of parameters $(2rt)$ needed to be estimated for

H increases, while that (t^2) for Q remains constant.

II. ALGORITHM

Consider a MIMO channel with inputs drawn from t spatially and temporally independent sources represented by $\mathbf{x} = [x_1, x_2, ..., x_t]^T$ such that $E(\mathbf{x}(k)\mathbf{x}^H(l)) = \delta_{kl}\sigma_s^2 I$. The channel outputs are $\mathbf{y}$ are given as,

$$\mathbf{y}(k) = H\mathbf{x}(k) + v(k) \qquad (1)$$

where H is an $r \times t$ channel matrix and v is spatio-temporally uncorrelated noise. $E(v(k)v(l)^T) = \delta_{kl}\sigma_n^2 \mathbf{I}$. A Rayleigh fading channel is considered and the channel is assumed constant over the transmission period. The output correlation matrix is

$$R_y = \sigma_s^2 HH^H + \sigma_n^2 I = \sigma_s^2 WW^H + \sigma_n^2 I \qquad (2)$$

Out of a total of N data transmissions for which the channel outputs $\{\mathbf{y}(1), \mathbf{y}(2), ..., \mathbf{y}(N)\}$ are observed, the pilot data $\{\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(L)\}$ is known for the initial L transmissions. Let $Y = [\mathbf{y}(1)\mathbf{y}(2)....\mathbf{y}(L)]$ and $X = [\mathbf{x}(1)\mathbf{x}(2).....\mathbf{x}(L)]$. It is desired to compute the best estimate of H from the complete available data.

We describe the conventional pilot based technique in the next section followed by the semi-blind algorithm.

II.1. Training Sequence Based Estimation (TS):

The least squares estimate of the complete channel matrix H using only pilot data is given as

$$\hat{H}_{TS} = YX^\dagger \qquad (3)$$

where $X^\dagger$ is the *Moore-Penrose pseudo-inverse* of X. This is referred to as the *training sequence* (TS) technique.

II.2. Estimating the Q matrix, W known

The techniques described in this section address the problem of estimating only the optimal rotation matrix Q using pilot data. The next section focuses on the problem of joint estimation of W and Q. Under the Gaussian noise assumption, the ML estimate of Q is obtained by minimizing

$$\|Y - WQX\|_F^2 \quad \text{where} \quad QQ^H = I \qquad (4)$$

II.2.1. **SVD$_1$** *Technique:*

When the whitening matrix W is known (i.e. when the estimate of W obtained from blind data is reasonably

accurate), a simple algorithm for estimating Q can be found by modifying the cost function in (4) as,

$$\min_Q \|W^\dagger \bar{Y} - Q\bar{X}\|_F^2 \quad \text{where} \quad QQ^H = I \qquad (5)$$

The solution to this problem is addressed below.

Lemma 1. The cost minimizing $\tilde{Q}$ for the cost function in (5) is given as:

$$\tilde{Q} = V_h U_h^H \quad \text{where} \quad U_h S_h V_h^H = \text{SVD}[\bar{X}Y^H W^{\dagger H}] \qquad (6)$$

where SVD denotes a *Singular Value Decomposition*.

Proof. Given in [7]. ☐

II.2.2. Rotation Optimization - 'RotOpt'

Though the least squares cost function (5) results in a simple algorithm, the resulting estimate does not have any statistically optimal properties. The true 2-norm error function to be minimized is (4). The *RotOpt* procedure involves finding the optimal rotation matrix Q to this exact cost function. It can be carried out using any standard numerical optimization routine such as the gradient descent, lagrange multiplier optimization routines. The $\tilde{Q}$ matrix computed in (6) is used to initialize all such optimization routines. Since it is a close approximation to Q, any standard procedure is sure to yield good results. In the work described in the paper, we specifically employed MATLAB based optimization routines.

II.2.3. SVD_2 Technique

If the system allows a certain degree of freedom in design, then the training symbols $X = [\mathbf{x}_1\mathbf{x}_2...\mathbf{x}_L]$ may appropriately be chosen to simplify the expression in (4) and thus arrive at a solution of reduced complexity. The result stated below, describes one such design choice.

Lemma 2. If $XX^H = \zeta I$ then the cost minimizing Q in (4) is given as

$$Q = VU^H \quad \text{where} \quad USV^H = \text{SVD}[XY^H W] \qquad (7)$$

Matrices X which satisfy $XX^H = \zeta I$ exist if modulation format is BPSK or more generally, if the signal subspace contains an *anti-podal* subset and $L = 2^n$, n is a positive number. This choice of pilot signal alleviates the need for a numerical procedure to optimize the *RotOpt* cost function and reduces it to an SVD computation without compromising performance.

II.3. Joint Estimation of W and Q

This section addresses the problem of estimating W blindly, and also jointly optimal estimates of W and Q.

II.3.1. Blind Estimation of W

The ML estimate of R_y is given as

$$\hat{R}_y = \frac{1}{N} \sum_{i=1}^{N} y_i y_i^H \tag{8}$$

Using (2),

$$\hat{W} = U_w \sqrt{S_w}, \tag{9}$$

where

$$U_w S_w V_w^H = \text{SVD} \left[\frac{1}{\sigma_s^2} \left\{ \hat{R}_y - \sigma_n^2 \mathbf{I} \right\} \right]. \tag{10}$$

σ_s^2 and σ_n^2 are assumed known.

II.3.2. Total Optimization - 'TotOpt'

This procedure builds on the above described procedures. After computing Q from the above procedures, the estimates of both W and Q can be refined using cost functions derived from the entire data set. Assuming data is *Gaussian*, the likelihood function for the total data comprising of the training data $\{(X_1, Y_1), (X_2, Y_2)..., (X_L, Y_L)\}$ and blind data $\{Y_{L+1}, ..., Y_N\}$ is given as

$$\mathcal{L}(W, Q) = -\mathcal{L}_1(W) - \mathcal{L}_2(W, Q) \tag{11}$$

where $\mathcal{L}_1(W) = (N - L)\frac{1}{2} \ln[\det(WW^H + \sigma_n^2 I)] - \sum_{i=L+1}^{N} Y_i^H (WW^H + \sigma_n^2 I)^{-1} Y_i,$

$$\mathcal{L}_2(W, Q) = \frac{1}{\sigma_n^2} \sum_{j=1}^{L} (Y_j - WQX_j)^H (Y_j - WQX_j) \tag{12}$$

$\mathcal{L}_1$ is a function of the blind data and $\mathcal{L}_2$ is a function of only the training data. This cost function can then be minimized for W with the earlier computed W as an initial estimate. Successive iterations of *RotOpt* and *TotOpt* can be performed to progressively improves estimates of W and Q.

As the data length N increases with pilot length L constant, the effect of $\mathcal{L}_2$ on the above expression weakens for the estimation of W. Hence, for large data lengths, the likelihood expression increasingly looks like $\mathcal{L}_1$, and maximizing it expression wrto W reduces to the blind estimation described by (9). Computation of Q then reduces to a minimization of $\mathcal{L}_2$ which is the cost function minimized by the *RotOpt* algorithm.

III. BOUNDS AND PERFORMANCE ANALYSIS:

In this section, the two competing techniques mentioned above are analyzed and compared through an analytical computation of performance bounds. The CR bound for the *RotOpt* technique, under the assumption of perfect knowledge of W is presented in the lemma below.

III.1. CR Bound for estimation of Q:

Theorem 1. If W is exactly known and $X \sim \mathcal{N}(0, \sigma_s^2 I)$, the CR bound for the error of estimation of $H = WQ$ from the complete data set $\{x(1), x(2), ..., x(L), y(1) y(2), ..., y(N)\}$ is

$$E(\|H - W\hat{Q}\|_F^2) \geq \frac{t^2}{2} \frac{\sigma_n^2}{\sigma_s^2 L} \tag{13}$$

$\hat{Q}$ is any unbiased estimate of Q. Note that t^2 is the number of parameters required to describe a complex $t \times t$ unitary matrix.

III.2. Bound for estimation of M:

Theorem 2. The error bound for estimation of the matrix H from the reference data $\{\mathbf{x_1}, ..., \mathbf{x_L}\}$ is given as

$$E(\|H - \hat{H}\|_F^2) \geq rt \frac{\sigma_n^2}{\sigma_s^2 L} \tag{14}$$

where $2rt$, is the number of parameters required to describe the complex $r \times t$ channel matrix H and $\hat{H}$ is any estimate of H.

From theorem (2) and theorem (1) it is evident that

1. Under conditions of perfect knowledge of W, and $r = t$, rotation estimation is **3 dB** more efficient than estimating H.

2. As the number of receive antennas r increases, the error of estimation of H increases, while that of Q remains constant.

III.3. Bound for estimation of W and Q:

Relaxing the previous assumption of the availability of a precise estimate of W, a lower bound, based on a first order analysis, for the *RotOpt* procedure is given as:

$$E(\|H - \hat{W}\hat{Q}\|_F^2) \geq$$

$$\frac{t^2}{2L} \frac{\sigma_n^2}{\sigma_s^2} + \frac{1}{N} \frac{\sum_{i=1}^{r} r_{ii}^2}{4\|W\|_F^2 \sigma_s^2} + 2\sqrt{\frac{t^2}{2L} \frac{\sigma_n^2}{\sigma_s^2} \frac{1}{N} \frac{\sum_{i=1}^{r} r_{ii}^2}{\|W\|_F^2 \sigma_s^2}} \tag{15}$$

$\hat{W}$ is estimated using (9). $r_{ii} = [R_y]_{ii}$. It can be seen from the above expression that the performance of the semi-blind algorithm is sensitive to $\|W\|_F^2$ ($= \|H\|_F^2$) and decreases with SNR for high SNR.
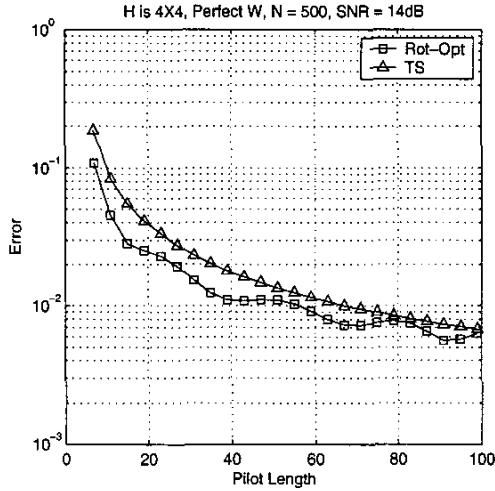
Fig. 1. Expt[1]-Error Variance $\|H - \hat{H}\|_F^2$ Vs Pilot Length L for the two competing techniques- TS and RotOpt, assuming the whitening matrix W is exactly known, N = 500. MPL = ∞
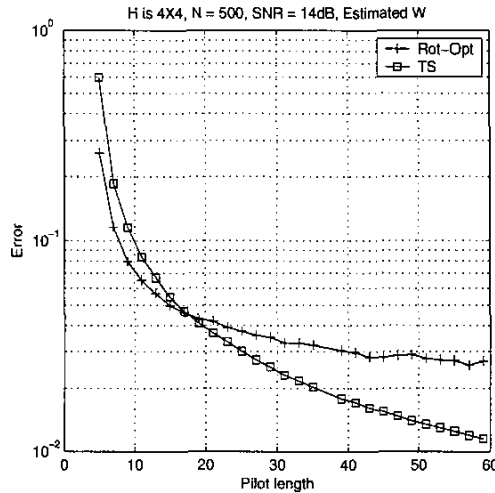
Fig. 2. Expt[2]-Error Variance $\|H - \hat{H}\|_F^2$ Vs Pilot Length L for TS and *RotOpt* in the case when W is estimated blind from output data $y_1, ..., y_N$. N = 400. MPL = 20.
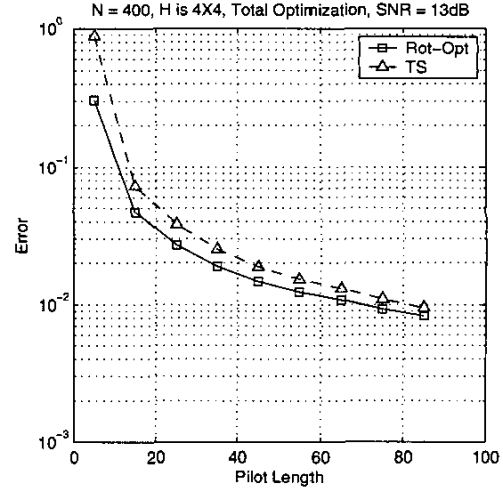
Fig. 3. Expt[2]-Error Variance $\|H - \hat{H}\|_F^2$ Vs Pilot Length L for TotOpt and TS techniques. N = 400. MPL = ∞.

IV. SIMULATIONS

In the simulations carried out, the elements of the channel matrix were generated as independent unit variance circular Gaussian random variables. The input vectors x were drawn from $t = 4$ independent (temporally and spatially) 16-QAM sources, and for different values of r ($= 4, 8, 12$), the number of receive antennas . Input SNR ($= \frac{\sigma_s^2}{\sigma_n^2}$) at each source was $\approx$ 14dB.

The final error is calculated as $\|M - \tilde{M}\|_F^2$. The error was averaged over multiple realizations ($\approx$30) in each experiment. The experiments are described below. Data length N was assumed equal to 400 samples unless specified otherwise. The variance of the estimate of different techniques is computed and plotted as a function of the pilot length L. To evaluate the effectiveness of the methods developed, we define the following parameter.

Definition 1. Maximum Pilot-Length (MPL)for a total number of data transmissions N is defined as the maximum number of training symbols (L) for which the *RotOpt* technique outperforms the TS technique.

The MPL is a compact representation of the range of pilot sequence lengths for which the developed techniques achieve better accuracy in the estimation of H compared to (3), training sequence based estimation. A high value of MPL indicates a wider range for which the developed methods are superior.

Experiment 1. Perfect knowledge of W is assumed. Q is then estimated using exclusively the training samples

$r \times t$	FADE (dB)	SVD	Rot-Opt
4×4	0	0	20
4×4	-6	0	55
4×4	-12	0	∞
8×4	0	0	30
8×4	-6	60	80
8×4	-12	∞	∞
12×4	-6	80	100

Table 1. Table showing MPL Vs Fade, for several r, t. MPL $= \infty$ implies *RotOpt* always performs better.

and *RotOpt* technique. Fig. 1 shows the error variance of the *RotOpt* technique, contrasted to the pilot based one. As is seen, the rotation optimization technique performs approximately 3dB better than estimating the whole matrix H from pilot data (MPL $= \infty$).

Experiment 2. No prior assumption was made on W and it was estimated directly from data. Fig. 2 shows the performance curves of both the estimation procedures as a function of the pilot length L. The *RotOpt* technique performs better for short pilot lengths (MPL ≈ 20).

Followed by the above procedure the *TotOpt* procedure is employed and optimum solution is found for the total likelihood cost function. MPL $= \infty$, and the technique outperforms the TS method (Fig 3).

Experiment 3. The performance of the above described techniques is investigated in a more severe fading environment. The matrix H is scaled with a factor P_o where $P_o \leq 1.00$ effectively reducing the SNR by $10 \log[P_o^2]$ dB (attenuation of the fade). Fig. 4 shows the plot for the case when H is 8×4 and the fade is -6 dB.

Table 1 gives the MPL for *RotOpt* and the SVD_1 technique for several values of r and P_o. It is evident that the MPL increases progressively as

- Increasing signal attenuation (i.e. Deeper Fade)

- Increasing number of receive antennas (r)

V. CONCLUSIONS

We have presented a semi-blind algorithm for the estimation of flat fading MIMO channel matrices, based on decomposing the matrix as the product of a whitening matrix and an unitary matrix. Theoretical results have been given to demonstrate that this algorithm can result in up to 3dB improvement in performance over the conventional estimation algorithm. Simulation results have been presented and the procedure has been seen to perform well in low SNR environments.

Po = 0.5(−6dB), H is 8X4, SNR = 14dB
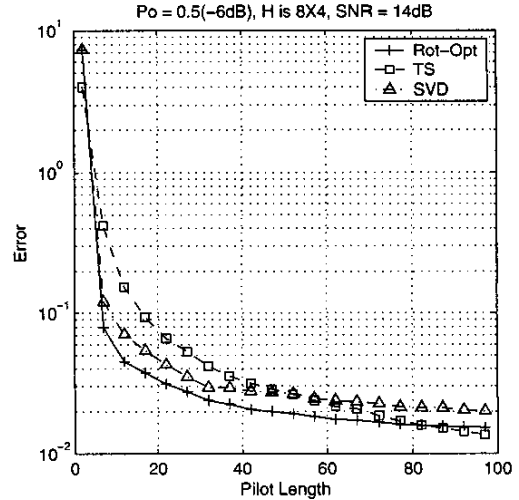
Fig. 4. Expt[3]-Error Variance $\|H - \hat{H}\|_F^2$ Vs Pilot Length L, for Fading Channels channel matrix H is 8×4, Fade = -6dB, MPL = 80

VI. REFERENCES

[1] John G. Proakis, *Digital Communications*, McGraw-Hill Higher Education, NewYork,NY-10020, international edition, 2001.

[2] Debajyoti Pal, "Fractionally spaced semi-blind equalization of wireless channels," *The Twenty-Sixth Asilomar Conference*, vol. 2, pp. 642–645, 1992.

[3] Debajyoti Pal, "Fractionally spaced equalization of multipath channels: a semi-blind approach," *1993 International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 9–12, 1993.

[4] Dirk T.M. Slock Elisabeth De Carvalho, "Cramer-rao bounds for semi-blind and training sequence based channel estimation," *First IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp. 129–32, 1997.

[5] Dirk T.M.Slock Elisabeth De Carvalho, "Asymptotic performance of ml methods for semi-blind channel estimation," *Thirty-First Asilomar Conference*, vol. 2, pp. 1624–8, 1998.

[6] Jean Francois Cardoso, "Blind signal separation : Statistical principles," *Proceedings of the IEEE*, vol. 86, pp. 2009–25, Oct 1998.

[7] Charles F. Van Loan Gene H. Golub, *Matrix Computations*, Johns Hopkins Univ Pr, second edition, 1984.