PILOT-ASSISTED CHANNEL ESTIMATION FOR MIMO OFDM SYSTEMS USING THEORY OF SPARSE SIGNAL RECOVERY

Mohammad A. (Amir) Khojastepour, Krishna Gomadam, Xiaodong Wang

NEC Labs America, Princeton, New Jersey, Email: {amir, kgomadam, wangx}@nec-labs.com

ABSTRACT

In this work, a new framework for channel estimation in MIMO OFDM systems is provided. Sparse channel estimation refers to estimating the time domain channel impulse response by exploiting the fact that the channel has a very few nonzero taps. We formalize the problem and drive necessary and sufficient condition on the number of pilots for perfect channel recovery which leads to a L0 norm optimization problem. A practical suboptimal solution is proposed that is a modified orthogonal matching pursuit (OMP) which exploits the sparsity structure of the MIMO channel. The investigations reveal that the training overhead can be drastically reduced while maintaining the same accuracy as the current state of the art techniques.

***Index Terms*—** Channel estimation, MIMO, OFDM

1. INTRODUCTION

Channel estimation for OFDM systems is traditionally approached in the frequency domain by estimating the frequency response for a few selected subcarriers and using those observations to interpolate the rest of the subcarriers.With this approach, the required number of pilots depends on the coherence bandwidth of the channel. The higher the bandwidth, the lesser the number of pilots. However, this approach takes into account only the length of the impulse response and ignores the sparsity of the wireless channel. Wireless channels are typically sparse, i.e., the time domain impulse response of the wireless channels typically has a very few nonzero taps.

Channel sparsity is attractive from a system design perspective since it can be exploited to design more efficient channel estimation strategies. A recent result on compressed sensing [1] shows that a discrete time signal of length M with only T nonzero coefficients can be exactly reconstructed from just observing any $2T$ samples of its discrete Fourier transform (DFT) if M is prime. This result has a direct application in high data rate OFDM systems where the number of subcarriers is large. While the result in [1] is remarkable, the proposed optimal signal recovery principle is combinatorial in nature. Thus, the optimization problem is relaxed to L1 norm optimization [1]. In [2], the authors employ this algorithm for channel recovery in SISO OFDM systems where the observations are corrupted with noise. The results indicate the scope for potential improvement in wireless channel estimation problems by using the theory of sparse signal recovery. Matching pursuit (MP) is a well-known algorithm used for sparse signal recovery and has many variants [3]. The algorithm iteratively identifies a small subset of the nonzero positions, that contribute to most of the energy in the observations. Although the algorithm is suboptimal and greedy in nature, it is efficient in terms of performance and complexity.

In this work, we extend the theory of sparse signal recovery for application in MIMO OFDM channel estimation problems which exploits the following properties of the multipath MIMO channel. (1) The channel between any pair of transmit and receive antennas has at most T taps, and $\frac{T}{L}$ is relatively small. (2) The positions of the nonzero taps are identical for all the channels associated with the point to point MIMO system. This property follows from the fact that the propagation delay is roughly the same for all transmit-receive antenna pairs. Practical channel models such as Spatial Channel Model (SCM) used in 3GPP incorporate this property for generating multipath MIMO channels [4].

The rest of paper is organized as follows. In Section 2, we discuss the channel and system model. The theoretical limits, i.e., the necessary and sufficient conditions on the minimum number of pilots required for perfect channel recovery in an ideal system where there is no noise at the receivers, are derived in Section 3. In Section 4, we propose a new weighted OMP algorithm that exploits the properties of the MIMO channel. Section 5 contains the numerical results and we conclude in Section 6.

2. CHANNEL AND SYSTEM MODEL

We consider a multipath environment with T clusters or scatterers. The impulse response between the ith transmitter and jth receiver is modeled as

$$h_{ji}(\tau,t) = \sum_{p=1}^{T} \alpha_p^{ji}(t)\delta(\tau - \tau_p(t)T_s) \tag{1}$$

where $\alpha_p(t) \in \mathbb{C}$ and $\tau_p(t) \in \mathbb{R}^+$ are the magnitude and the delay for path p, respectively, and T_s is the sampling in-

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

terval of the system. With block-fading channel assumption where the channel parameters are constant over a block and assuming perfect symbol level synchronization, the equivalent discrete time channel between transmit antenna i and receive antenna j can be modeled as

$$h_{ji}(n) = \sum_{p=1}^{T} \alpha_p^{ji} g((n - \tau_p)T_s) \tag{2}$$

where $g(t)$ represents the effect of the pulse shaping filter and the RF front-ends at both the transmitter and the receiver. It can be noticed that in high data rate communication systems where T_s is very small compared to the maximum delay spread, (2) results in a channel with a very few nonzero taps. For a raised-cosine filter with excess bandwidth of 0.5 or greater, the above channel will have approximately T nonnegligible taps.

We consider a cyclic prefix based OFDM system with pilot aided channel estimation and FFT size M. The channel estimation procedure consists of the following protocol. The training phase spans n_t OFDM symbols in which we assume the channel remains constant. In the ith slot, the ith transmit antenna sends pilots in Q_i subcarriers and the remaining $M - Q_i$ subcarriers are used for data transmission. The set of pilots chosen for different transmit antennas need not be the same or disjoint. The received pilots at receive-antenna j during the mth training slot is given by

$$\begin{aligned} \mathbf{b}_{jm} &= \sqrt{\mathbf{P}_m}\mathbf{S}_m\mathbf{F}\mathbf{h}_{jm} + \mathbf{n}_j,\ m = 1, ..., n_t,\ j = 1, ..., n_r \\ &= \mathbf{A}_m\mathbf{h}_{jm} + \mathbf{n}_j \end{aligned} \tag{3}$$

where $\mathbf{h}_{jm}$ is the $M \times 1$ vector representing the channel between mth transmit antenna and jth receive antenna, $\mathbf{F}$ is the $M \times M$ DFT matrix whose entry corresponding to the pth row and qth column is given by $\exp(-j\frac{2\pi}{M}pq)$, $\mathbf{S}_m$ is the $Q_m \times M$ selection matrix that chooses the Q_m rows of the DFT matrix according to the pilots chosen in the mth slot and the diagonal matrix $\mathbf{P}_m$ is the power loading matrix that determines the power allocated to the pilot subcarriers. $\mathbf{n}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the additive white Gaussian noise for the selected pilot tones.

It can be noticed that the training stage entails a total of $\sum_{m=1}^{n_t} Q_m$ pilots to estimate $n_t n_r$ channels. We seek to establish a theoretical limit on the number of pilots required for perfect channel recovery in an ideal noiseless system and approach the limit with efficient practical algorithms.

3. THEORY FOR PERFECT RECONSTRUCTION

The channel estimation problem considered in this work falls under the category of under-determined system since we wish to estimate the response for a large number of sub-carriers using a limited number of observations. Here, we extend the result of [1] that is applicable to SISO channels to the case that covers MIMO channels. The MIMO channel reconstruction problem is posed as follows: Consider the following noiseless observations of the vectors $\mathbf{h}_{jm}$ which all have common sparsity structure[1] of size T (e.g., $\mathbf{h}_{jm}$ represents the channel in the frequency domain).

$$\mathbf{b}_{jm} = \mathbf{A}_m\mathbf{h}_{jm},\ \ m = 1, 2, ..., n_t,\ \ j = 1, 2, ..., n_r \tag{4}$$

What are the conditions to uniquely determine the channels? Also, what is the minimum number of pilots required to reconstruct all the channels?

Theorem 1 *For the linear inverse problem in (4) with M being prime, perfect reconstruction of all the vectors is guaranteed if and only if*

$$\max\{Q_1, Q_2, \cdots, Q_{n_t}\} \geq 2T\ \&\ \min\{Q_1, Q_2, \cdots Q_{n_t}\} \geq T.$$

The proof is through finding two distinct solutions which satisfy (4) if the conditions of the theorem are not satisfied. We skip the proof due to space limitation.

The theorem suggests that only $(n_t + 1)T$ pilots are enough to completely reconstruct the set of $n_t n_r$ channels. And, the best pilot allocation policy is to allocate $2T$ pilots for one transmit antenna and T pilots to the rest of the antennas. From Theorem 1, one can show that the following optimization problem successfully identifies the set of channels if the conditions in Theorem 1 are satisfied.

$$\min \max\{\| \mathbf{h}_{j1} \|_0, \| \mathbf{h}_{j2} \|_0, ..., \| \mathbf{h}_{jn_t} \|_0\} \tag{5}$$

$$\text{s.t.}\quad \mathbf{b}_{jm} = \mathbf{S}_m\mathbf{F}\mathbf{h}_{jm},\ \ m = 1, 2, ..., n_t \tag{6}$$

Nonetheless, this is a discrete optimization problem and is hard to solve similar to the result for SISO case, i.e., optimization problem for optimal sparse signal recovery in [1]. In [1], the L0 norm optimization problem is relaxed to L1 norm to make it tractable. We have considered several suboptimal algorithm including L1 norm optimization [1], modified FOCUSS [3], and modified MP. The MP algorithms are attractive as they enable low complexity implementations. Due to the lower complexity and its better performance, we discuss the proposed MIMO MP in more detail in the following.

4. MIMO MATCHING PURSUIT (MP)

The MP algorithms work by sequentially selecting a small subset of the tap positions that contribute to the most of the energy in the receive observations. In our case, although the tap positions are the same, we have different observations for different transmit-receive antenna pairs. Further the length of the observation (the number of pilots) is not the same for all antennas. Therefore, we need a robust algorithm that can handle these scenarios.

[1] A set of vectors $\{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_L\}$ is said to have a *common sparsity structure of size* T if for any i and k $\sum_{j=1}^{M} \delta_{\mathbf{f}_i - \mathbf{f}_k}(j) \leq T$ and $\sum_{j=1}^{M} \delta_{\mathbf{f}_k}(j) \leq T$ where $\forall \mathbf{f} : \delta_{\mathbf{f}}(j) = 1$ iff the jth element of $\mathbf{f}$ is nonzero.

To begin with, we provide a simple algorithm to convey the essence of the algorithm. To avoid multiple subscripts, we only consider the case of 2×1. The algorithm can be readily extended to the general $n_t \times n_r$ case. At each iteration of the algorithm, we choose a tap position that minimizes the sum of the norm-squared residue. The residues for the first iteration are set as $\mathbf{b}_1^0 = \mathbf{b}_1$ and $\mathbf{b}_2^0 = \mathbf{b}_2$. At the first iteration, the tap position that maximizes the following is chosen:

$$k_1 = \arg\max_j |\mathbf{A}_{1j}^\dagger \mathbf{b}_1^0|^2 + |\mathbf{A}_{2j}^\dagger \mathbf{b}_2^0|^2 \quad j \in \{1, 2, .., M\} \quad (7)$$

where $\mathbf{A}_{1j}$ is the jth column of $\mathbf{A}_1$. The residues are then updated as $\mathbf{b}_1^1 = \mathbf{P}_{\mathbf{A}_{1j}}^\perp \mathbf{b}_1^0$ and $\mathbf{b}_2^1 = \mathbf{P}_{\mathbf{A}_{2j}}^\perp \mathbf{b}_2^0$ where $\mathbf{P_W}$ and $\mathbf{P_W^\perp}$ are the standard projection operations given by $\mathbf{P_W} = \mathbf{W}(\mathbf{W}^\dagger\mathbf{W})^{-1}\mathbf{W}^\dagger$, $\mathbf{P_W^\perp} = \mathbf{I} - \mathbf{W}(\mathbf{W}^\dagger\mathbf{W})^{-1}\mathbf{W}^\dagger$. At every iteration, the above procedure is followed to add a column to the set of already selected columns. The residues are calculated such that they are orthogonal to the set of selected columns.

We can notice that the above algorithm gives equal weight to both the observations although Q_1 may not be equal to Q_2. Further, for the case of perfect reconstruction of the channels with the minimum number of pilots, only one observation vector is used to identify the nonzero tap positions which are common to both the vectors. Therefore, giving equal weight to the observation vectors during the column selection stage need not yield the best performance. We use the following metric to select a column during each iteration: At the pth iteration, choose a column that minimizes the weighted sum of q-norm of the residues where $q > 0$. That is, the index of the selected column during pth iteration is

$$\begin{aligned}
k_p &= \arg\min_j \min_{\mathbf{x}_1, \mathbf{x}_2} w_1 \parallel \mathbf{b}_1 - \left[\mathbf{C}_1^{p-1}\mathbf{A}_{1j}\right] \mathbf{x}_1 \parallel^q \\
&+ w_2 \parallel \mathbf{b}_2 - \left[\mathbf{C}_2^{p-1}\mathbf{A}_{2j}\right] \mathbf{x}_2 \parallel^q \\
&= \arg\min_j w_1 \left(\parallel \mathbf{b}_1^{p-1} \parallel^2 - \left| \mathbf{A}_{1j}{}^{(p-1)\dagger} \mathbf{b}_1^{p-1} \right|^2 \right)^{\frac{q}{2}} \\
&+ w_2 \left(\parallel \mathbf{b}_2^{p-1} \parallel^2 - \left| \mathbf{A}_{2j}{}^{(p-1)\dagger} \mathbf{b}_2^{p-1} \right|^2 \right)^{\frac{q}{2}} \quad (8)
\end{aligned}$$

where $\mathbf{C}_i^{p-1}$ is the matrix whose columns contain the selected columns of $\mathbf{A}_i$ till (p-1)th iteration. In the optimization problem, the size of x_i is $p \times 1$. Let $\mathbf{C}_1^p(j) = \left[\mathbf{C}_1^{p-1} \; \mathbf{A}_{1j}\right]$, $\mathbf{C}_2^p(j) = \left[\mathbf{C}_2^{p-1} \; \mathbf{A}_{2j}\right]$, $\mathbf{A}_{1j}^{p-1} = \mathbf{P}_{\mathbf{C}_1^{p-1}}^\perp \mathbf{A}_{1j}$, and $\mathbf{A}_{2j}^{p-1} = \mathbf{P}_{\mathbf{C}_2^{p-1}}^\perp \mathbf{A}_{2j}$. In following, we discuss the key steps of the proposed algorithm.

Initialization: The residues are initialized as $\mathbf{b}_1^0 = \mathbf{b}_1$ and $\mathbf{b}_2^0 = \mathbf{b}_2$. Similarly the generating matrices are initialized as $\mathbf{A}_1^0 = \mathbf{A}_1$ and $\mathbf{A}_2^0 = \mathbf{A}_2$

Tap Detection: The column selection metric can be obtained by using $q = 2$ in (8). In the pth iteration, the column is selected which maximizes $w_1 \left| \mathbf{A}_{1j}{}^{(p-1)\dagger} \mathbf{b}_1^{p-1} \right|^2 + w_2 \left| \mathbf{A}_{2j}{}^{(p-1)\dagger} \mathbf{b}_2^{p-1} \right|^2$.

Update: The set of selected columns is updated as $\mathbf{C}_i^p(j) = \left[\mathbf{C}_i^{p-1} \; \mathbf{A}_{ik_p}\right]$. The residues $\mathbf{b}_1^p$ and $\mathbf{b}_2^p$ are calculated as

$$\begin{aligned}
\mathbf{b}_1^p &= (\mathbf{I} - \mathbf{P}_{\mathbf{C}_1^p})\mathbf{b}_1^{p-1} \\
\mathbf{b}_2^p &= (\mathbf{I} - \mathbf{P}_{\mathbf{C}_2^p})\mathbf{b}_2^{p-1} \quad (9)
\end{aligned}$$

The generating matrices are updated as

$$\begin{aligned}
\mathbf{A}_1^p &= \text{nrm}\left(\mathbf{P}_{\mathbf{C}_1^p}^\perp \mathbf{A}_1^{p-1}\right) \\
\mathbf{A}_2^p &= \text{nrm}\left(\mathbf{P}_{\mathbf{C}_2^p}^\perp \mathbf{A}_2^{p-1}\right) \quad (10)
\end{aligned}$$

where the nrm function normalizes the columns of the argument.

Stopping condition: The algorithm continues to iterate until the maximum number of iterations is reached or the weighted norm-squared residue goes below a threshold, i.e. $w_1 \parallel \mathbf{b}_1^p \parallel^2 + w_2 \parallel \mathbf{b}_2^p \parallel^2 \leq \epsilon$. In systems with noise, ϵ is determined according to the signal to noise ratio (SNR). In general, the greater the SNR, the lower the value of ϵ. In our numerical analysis, we assume $\epsilon = 0.01$ for the noiseless case.

Convergence: It is straightforward to notice that the algorithm converges since the metric (weighted squared norm residue) decreases with each iteration.

Tap gain regeneration: After determining the set of columns to represent the given observations, the estimate of the channels at the selected tap positions are obtained through $(\mathbf{C}_i^{p^\dagger}\mathbf{C}_i^p)^{-1}\mathbf{C}_i^{p^\dagger}\mathbf{b}_i$. For the case of noisy observations, the pseudo inverse can be replaced by a regularized inverse. That is, the tap values for the selected column indices can be obtained from $(\mathbf{C}_i^{p^\dagger}\mathbf{C}_i^p + \frac{1}{\text{SNR}}\mathbf{I})^{-1}\mathbf{C}_i^{p^\dagger}\mathbf{b}_i$. We refer to this version as MSE based orthogonal MP.

5. NUMERICAL RESULTS

We consider a SIMO channel with two receive antennas and randomly select T out of L tap positions. The tap values are two i.i.d. $\mathcal{CN}(0, 1)$ corresponding to the tap positions for the two channels of the SIMO system. We consider Q randomly generated pilot positions. In Fig. 1, we plot the accuracy performance as a function of the number of pilots. It can be observed that both OMP1 and OMP2 require only 20 pilots to approach perfect accuracy where the theoretical limit is 16. The figure also shows the performance gain obtained through joint channel estimation where the sparsity structure is exploited. About 5 pilots can be saved through joint MP over independent MP where in the joint MP algorithm, we use $w_1 = 0.5$ and $w_2 = 0.5$. The L1 norm optimization algorithm can also approach perfect accuracy with about 30 pilots.

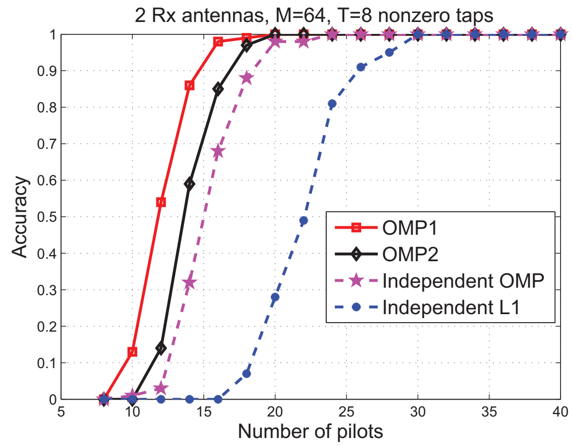

Fig. 1. Accuracy performance in SIMO channel with two receive antennas as a function of the number of transmitted pilots

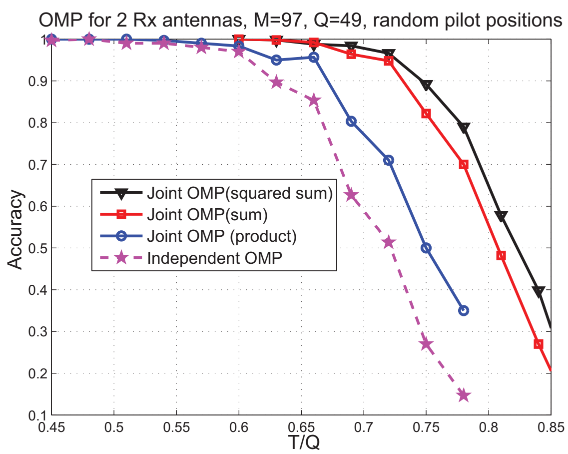

Fig. 2. Correlation combining metrics for OMP $q = 2$.

Intuitively, the MP algorithm, for each channel, correlates the given observation with each columns of the generating matrix. Now, using the results of the above operations, it needs to identify a dominant tap during each iteration. The metric in (8) was arrived at using a weighted least squares problem. Other metric might also be considered such as $k_p = \arg\max_i \quad a_i^2 + b_i^2$ or $k_p = \arg\max_i \quad a_i b_i$ where a_i and b_i are the absolute correlation between the ith column of the generating matrix and the residue for the two channels respectively. In Fig. 2, we compare the performance of different correlation combining metrics. It can be seen that, in terms of accuracy, the squared norm sum of correlation better than the others. This metric is the one obtained in Section 4.

In Fig. 3, we compare the BER performance of the schemes using practical channel model (SCM) [4]. It can be seen that MSE and ideal MP algorithms perform the best.

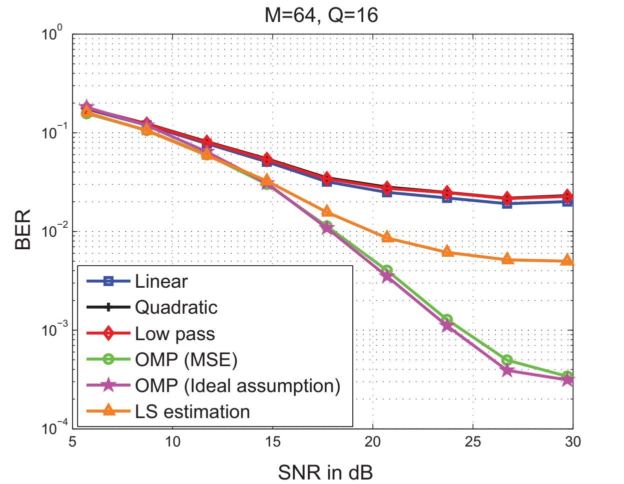

Fig. 3. BER performance of the schemes with BPSK modulation for the SCM channel model for M=64 and Q=16.

The practical state-of-art schemes in frequency domain using linear, quadratic, and Low-pass interpolation as well as direct LS estimation performs much worse than MP and seriously suffer from error floor.

6. CONCLUSION

In this work, we identified two important properties of a MIMO multipath channel in which the individual channels, in addition to being sparse, follow the same sparsity structure. We demonstrated the conditions on the required number of pilots to accurately reconstruct all the channels in a MIMO OFDM system. The modified MP algorithm is presented which has proved very efficient in MIMO channel estimation.

7. REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[2] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," *ICASSP*, pp. 3469–3472, 31 2008-April 4 2008.

[3] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 53, pp. 2477–2488, July 2005.

[4] J. S. P. K. M. M. D. L. J. Salo, G. Del Galdo and C. Schneid, "Matlab implementation of the 3gpp spatial channel model," *3GPP TR 25.996*, Jan 2005.