APPLICATION OF SPARSE SIGNAL RECOVERY TO PILOT-ASSISTED CHANNEL ESTIMATION

Matthew Sharp and Anna Scaglione

Cornell University
School of Electrical and
Computer Engineering
Ithaca, NY 14853
Email:mds72@cornell.edu, anna@ece.cornell.edu

ABSTRACT

We examine the application of current research in sparse signal recovery to the problem of channel estimation. Specifically, using an Orthogonal Frequency Division Multiplexed (OFDM) transmission scheme with Pilot Symbol Assisted Modulation (PSAM), we consider the problem of identifying a frequency selective channel from a limited number Q out of a possible M tones of an OFDM symbol. The main observation is that if M is chosen as prime, one can identify the channel uniquely if $Q \geq 2T$, where T is the number of nonzero taps in the frequency-selective channel. The identifiability result requires the minimization of the l_0 norm, leading to an intractable combinatorial search problem. Several methods have been proposed to deal with these issues, and the one we examine involves l_1 norm regularization known as basis pursuit [1]. We apply these methods specifically to the problem of estimating a frequency selective channel with PSAM. As a result, the bandwidth efficiency of the system is increased due to the sparsity of the channel.

***Index Terms*—** Sparse signal recovery, Channel estimation, OFDM, PSAM

1. INTRODUCTION

In broadband transmission, the data is corrupted by inter-symbol-interference (ISI). To overcome this, the majority of communication standards use pilot symbols to estimate the coefficients of the equivalent discrete-time baseband channel impulse response. Three common training methods are (*i*) preamble based training, where the pilot sequence is included at the beginning of a data burst; (*ii*) PSAM, where the training sequence is inserted in the data stream either in frequency or time, (*iii*) and superimposed training, where the training sequence is added to the data sequence. Few methods for estimation of the channel with PSAM training make assumptions on the *sparsity* of the channel. Several exceptions are, e.g. [2],[3],[4].

A sparse channel can be considered as a channel with a large delay spread but with relatively few nonzero coefficients. Applications with sparse channels include high definition television (HDTV) signals, [5][6], and underwater acoustic channels [7]. The sparsity of the channel allows identification of the channel from a limited number of measurements, smaller than the length of the channel. Suppose, for example, we have the model

$$\mathbf{y} = \mathbf{X}\mathbf{h}$$

where $\mathbf{y}$ is of length N and the sparse signal $\mathbf{h}$ is of length $L >> N$. There are known results and restrictions which guarantee the channel is uniquely identifiable in a noise-free scenario, but most methods to recover the signal are combinatorial in nature as they consider minimization of the l_0 norm. Several methods for sparse reconstruction of $\mathbf{h}$ with a more tractable solution include the matching pursuit algorithm (MP) [8], and l_1 norm regularization such as basis pursuit [1]. To the best of our knowledge, the matching pursuit algorithm does not guarantee unique identifiability of the channel, and the sufficient conditions on $\mathbf{X}$ for unique identification of the channel using l_1 norm regularization due not readily allow one to find deterministic matrices satisfying these conditions. It has been shown in [9] that certain types of random measurement matrices can uniquely recover the channel with high probability, such as when the elements of $\mathbf{X}$ are drawn independently from a Gaussian distribution.

When estimating a frequency selective channel using PSAM with the pilot symbols separated from the data symbols in time (time-division multiplexed training) the matrix $\mathbf{X}$ above will be a Toeplitz matrix. Recently in [10], it was shown that Toeplitz matrices with entries drawn independently from probability distributions satisfy the conditions in [9], thus allowing unique identification of a sufficiently sparse channel with high-probability.

Considering PSAM for a single carrier transmission system, to have data which is not affected by ISI from the information symbols, we need to transmit a number of pilot symbols Q greater than or equal to the channel length L of the channel. Each additional equation requires an extra pilot symbol. The main advantage of formulating the problem in the Fourier domain is that, for a number of tones which is prime, it is possible to give necessary and sufficient conditions on the identifiability of the channel that cannot be given in the time domain counterpart of the problem. The sufficient conditions available in [10] would lead to worse spectral efficiency compared to what is forecasted for the OFDM case. In Section 2 we present the model of our system, and in Sections 3 and 4 we discuss the application of known uncertainty principles and methods for estimating the sparse channel.

2. SYSTEM MODEL

Let $x[n]$ be the discrete-time complex baseband equivalent signal representing the pilot sequence and let us assume that $x[n]$ has finite duration P. Further, let us model the discrete-time complex

This material is based upon work supported by the Office of Naval Research (ONR) under Contract No. N00014-05-C-0070. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of ONR.

baseband channel $\mathbf{h} = (h[0], ..., h[L-1])$ of finite length L as a deterministic *sparse* impulse response, i.e.

$$h[l] = \sum_{t=1}^{T} h_t \delta[l - d_t] \quad (1)$$

where, without loss of generalization, $0 \leq d_1 \leq d_2 \leq ... \leq d_T$ are unknown parameters. We refer to the set $\mathcal{T} = \{d_1, ..., d_T\}$ as the support of $h[l]$ and indicate its cardinality by $|\mathcal{T}|$ or T. The output of the system due to the pilot signal only is:

$$y[n] \;=\; \sum_{l=0}^{L-1} h[l] x[n-l] + \tilde{w}[n] \quad (2)$$

where $\tilde{w}[n] \sim \mathcal{CN}(0, N_0)$. Assuming there are no guard intervals between the pilot and data sequence, we define the vector $\mathbf{y} = (y[L-1], ..., [L+N-1]$ which contains the received signals which contain contributions only from the pilot sequence. We may then express (2) in vector form as:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \tilde{\mathbf{w}} \quad (3)$$

where $\mathbf{X}$ is an $N \times L$ Toeplitz convolution matrix. In general, if no sparsity constraint is made on the channel, we require $N \geq L$ meaning we need $P \geq 2L - 1$ pilot symbols in order to recover the channel. However, if we assume the channel is sparse, the number of pilot symbols P can be reduced. This particular problem is considered in [2], where the authors proposed the use of i.i.d random pilot symbols and used l_1 norm regularization to estimate the channel. However, no sufficient conditions were specified for *perfect* recovery of the channel.

As stated previously, in [10] the authors investigated the conditions under which a Toeplitz convolution matrix with more columns than rows allows the exact recovery of a sparse signal in a noise-free system with high probability. The sufficient conditions proposed in [10] apply clearly to the problem at hand as we can estimate the channel coefficients with a limited number of randomly generated pilot symbols.

Our intention here is to resort to the work of [11] to give strict identifiability constraints and show how they can result in more efficient transmission of information through the channel, by instead decreasing the amount of bandwidth spent on the pilot sequence.

In particular let us assume OFDM transmission, known for its high spectral efficiency. The use of OFDM allows us to formulate the problem in the Fourier domain. Let $X[k]$ represent the comb of pilots in the OFDM symbol,

$$X_p[k] = \sum_{q=1}^{Q} \delta[k - p_q] \quad (4)$$

with $\Omega = \{p_1, ..., p_Q\}$ and $p_q \in \{0, 1, ..., M-1\}$. The OFDM modulated pilot sequence $x[n]$ is generated from the M-point DFT:

$$\begin{aligned} x_p[n] &= \frac{1}{\sqrt{P}} \sum_{k=0}^{M-1} X_p[k] e^{2\pi jkn/M}, \;\; 1 - L \leq n \leq M - 1 \\ x[n] &= \frac{1}{\sqrt{P}} \sum_{q=1}^{Q} e^{2\pi j p_q n/M} + x_s[n] \end{aligned} \quad (5)$$

We note that the sequence $x[n]$ incorporates a cyclic prefix of length $L-1$ and the signal $x_s[n]$ corresponds to the data tones. Thus the $P = M + L - 1$ length pilot sequence given by the vector $\mathbf{x}_p$ will consists of Q tones out of a possible M tones. Removing the cyclic prefix at the receiver, let $\mathbf{y} = (y[L-1], ..., y[L+M-1])^T$, Thus after taking the M point DFT of $\mathbf{y}$, we have

$$\begin{aligned} z[d] &= \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} y[L-1+m] e^{-\frac{j2\pi dm}{M}} \\ z[p_q] &= \sqrt{\frac{M}{P}} H(e^{-jw_q}) + w[p_q] \end{aligned} \quad (6)$$

where $\omega_q = 2\pi p_q/M$. We can therefore obtain an estimate of the partial frequency response of the channel given by $\hat{H}(e^{-jw_q})$ using for example an zero-forcing or minimum mean squared error estimator. In the following, let $\hat{H}(e^{-j\Omega})$ denote the vector corresponding to the partial frequency response of $\mathbf{h}$ at frequencies $p_q \in \Omega$. In vector notation we may express this as:

$$\mathbf{z} = \sqrt{\frac{M}{P}} \boldsymbol{\mathcal{S}} \mathbf{F} \mathbf{h} + \mathbf{w} \quad (7)$$

where $\mathbf{F}$ is the DFT matrix and $\boldsymbol{\mathcal{S}}$ is a $Q \times L$ selection matrix selecting only the rows of the DFT matrix with frequencies in Ω.

3. UNCERTAINTY PRINCIPLE

We now examine strict identifiability conditions for recovering the channel. Consider the noise-free system where our estimate of the partial frequency response is exact, i.e. $\hat{H}(e^{-j\Omega}) = H(e^{-j\Omega})$. We want to know if we can recover the sparse impulse response $\mathbf{h}$ from its partial frequency response. The question can be formulated as follows: What is the minimum number Q of pilot tones necessary to uniquely identify the sparse impulse response $\mathbf{h}$? This question is equivalent to finding conditions such that the solution to the following combinatorial optimization problem is unique.

$$\min \|\mathbf{g}\|_{l_0} \;\text{ such that }\; H(e^{-j\Omega}) = G(e^{-j\Omega}) \quad (8)$$

where $\|\mathbf{g}\|_{l_0}$ is simply the number of non-zero elements of $\mathbf{g}$.

The answer is provided in the following Lemma, first proved by Chebotarev in 1926 (see [12]). We repeat the lemma from ([11], Lemma 1.3):

Lemma 3.1 *Let M be a prime number, the matrix*

$$\{\mathbf{A}\}_{q,t} = e^{-\frac{j2\pi p_q d_t}{M}}, \quad q = 1, ..., T \;\; t = 1, ...T$$

has nonzero determinant (and is therefore invertible) for any $1 \leq T \leq M$ if $p_1, ..., p_Q$ are distinct elements of $\mathbf{Z}/M\mathbf{Z}$ and $d_1, ..., d_T$ are also distinct elements of $\mathbf{Z}/M\mathbf{Z}$, where $\mathbf{Z}/M\mathbf{Z}$ denotes the integers modulo M.

A simple corollary of Lemma 3.1 is for $Q \geq T$ the matrix $\mathbf{A}$ is full column rank. This lemma leads to the following theorem, stated in ([13], Theorem 1.1):

Theorem 3.2 *The sparse channel with $|\mathcal{T}| = T$ is uniquely identifiable from Q out of a possible M pilot tones if M is prime and $Q \geq 2T$*

The proof of this theorem is given in the appendix.

The consequence of Theorem 3.2 is that if $L > 2T$ we can reduce the number of pilot tones that are strictly required to identify the channel. If no sparsity constraint is made on the channel in general we require L pilot tones to estimate the channel. However, if we assume the channel is at most T sparse, the bandwidth efficiency of the PSAM multi-carrier system is increased:

$$\rho_{MC} = \frac{M-Q}{M+L-1} \leq \frac{M-2T}{M+L-1} \tag{9}$$

Unfortunately, the combinatorial optimization problem (8) is complex to solve since one must look at all subsets of $\mathbf{h}$ of size T for the set which gives $H(e^{-j\Omega})$. Furthermore, the presence of noise in the system resulting in noisy estimate $\hat{H}(e^{j\Omega})$ requires relaxing the strict equality $G(e^{j\Omega}) = \hat{H}(e^{j\Omega})$. These problem are discussed in the following sections.

4. CHANNEL ESTIMATION VIA L_1 NORM MINIMIZATION

The pioneering work in [13] presented a tractable solution to the problem (8) by demonstrating that under a stricter set of conditions, the l_1 norm relaxation of (8) given by the following convex optimization problem,

$$\min \|\mathbf{g}\|_{l_1} \text{ such that } H(e^{-j\Omega}) = G(e^{-j\Omega}) \tag{10}$$

results in unique identification of the signal. Specifically, they demonstrated that if one selects $|\Omega|$ rows from the DFT matrix uniformly at random, and the vector $\mathbf{h}$ has support of size T, such that

$$T \leq C_\gamma (\log M)^{-1} |\Omega| \tag{11}$$

for some constant $C_\gamma > 0$, then with probability at least $1 - O(M^{-\gamma})$ (10) will uniquely recover the signal $\mathbf{h}$. The bound in (11) results from letting $\mathbf{h}$ be a picket fence signal with exactly $\sqrt{M}$ spikes separated every $\sqrt{M}$ samples. The recovery of this signal requires that the intersection $\Omega \cap \mathcal{T}$ cannot be empty.

In light of this work, with respect to the problem of recovering a sparse channel with $|\mathcal{T}| = T$ from Q out of a possible M pilot tones, we can still uniquely recover with high probability the channel impulse response from $Q \leq L$ pilot tones chosen uniformly at random. Though the theory suggests that the set of frequencies Ω must be chosen at random, one wonders if we can choose a deterministic set of Q pilot tones and still exactly recover $\mathbf{h}$. This question is addressed through simulation in section 6.

5. SPARSE CHANNEL ESTIMATION WITH NOISE

The convex optimization problem (10) can be modified for noisy measurements as:

$$\min \|\mathbf{g}\|_{l_1} \text{ such that } \|\hat{H}(e^{-jw}) - G(e^{-jw})\|_{l_2}^2 \leq \nu \tag{12}$$

where $\nu \geq \|w\|_{l_2}^2$ and $\mathbf{w}$ is the noise in the system. In [14] the solution to the problem was shown to be stable in the presence of noise, i.e. $\|\hat{\mathbf{h}} - \mathbf{h}\|_{l_2}^2 \leq C\nu$ for some $C > 0$. The constraint relies on the random value ν. If we assume that $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I})$, then we know that it has a χ^2 distribution with $2Q$ degrees of freedom and thus we can choose α such that $\alpha \geq \|\mathbf{w}\|_2^2$ with high probability.

6. SIMULATIONS

In the first simulation, for various values of M, we fix the number of tones to transmit, Q, and vary T as a percentage of Q. For each value of T we run 100 iterations of the noise-free system randomly generating at each iteration Ω and $\mathcal{T}$. The taps of the channel supported on $\mathcal{T}$ are generated as i.i.d. complex Gaussian circular symmetric random variables with variance $\sigma_h^2 = 1$. Fig. 1 shows the percentage for each support size that the channel is correctly identified. We also compare the randomly generated pilot tones with a fixed pilot tone placement scheme where we select T tones, starting with the first discrete frequency, and increase the gap between tones by one until we have exceeded $M/2$. The remaining tones are all placed at the end of the OFDM symbol. This design is done to intentionally avoid a symmetric placement of while generating analogous patterns for different M. In the the case of randomly selected tones, we observe the performance remains unchanged regardless of whether or not M is prime. However, for our fixed selection of pilot tones, the performance is worse than in the case of randomly selected tones.

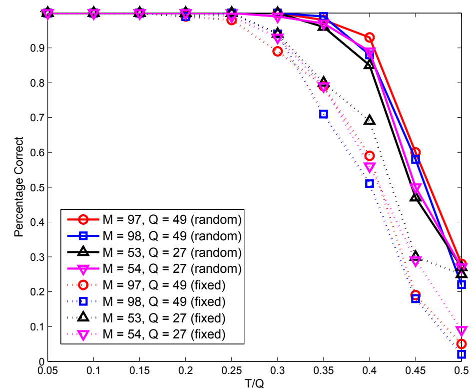

Fig. 1. Success Rate of (10) in correctly estimating the complex sparse channel $\mathbf{h}$ in a noise-free system with Q out of $M = L$ randomly selected (solid) and fixed (dotted) tones for different percentages T/Q.

We now wish to examine whether there exists a fixed placement of pilot tones which performs better than the randomly selected placement of pilot tones. In Fig. 2, we perform a similar simulation as in Fig. 1, with $M = L = 10$. In this case, we examine the percentage of correctly identified channels over 1000 iterations for each value of T. For our fixed selection of pilot tones, $\Omega = \{1, 2, 5, 7, 10\}$ and $\Omega = \{1, 2, 5, 7, 9, 10\}$ for $Q = 5$, and $Q = 6$ respectively. We see that for our fixed assignment of pilot ones, the performance is better for $T = 1$ and $T = 2$ than the randomly selected tones. Thus, there do exist fixed assignments of pilot tones which perform better than randomly selected tones.

In Fig. 3, we plot the mean square error (MSE) of the same fixed pilot tone assignment for $Q = 6$ as in Fig. 2 for a noisy system using (12). The noise $\mathbf{w} \sim \mathcal{CN}(0, N_o \mathbf{I})$ and we define our $SNR = Q/N_o$ since each pilot has unit energy at the receiver (5). The T taps of the channel are again uniformly selected and each tap on the channel support $\mathcal{T}$ is generated as i.i.d. complex Gaussian circular symmetric with variance $\sigma_h^2 = 1$. We see that the MSE tends to increase as the support size T is increased. Further, for $T = 3$, as observed in Fig. 2, the identifiability of the channel is lost, and therefore the MSE performance has an error floor as the SNR is increased.

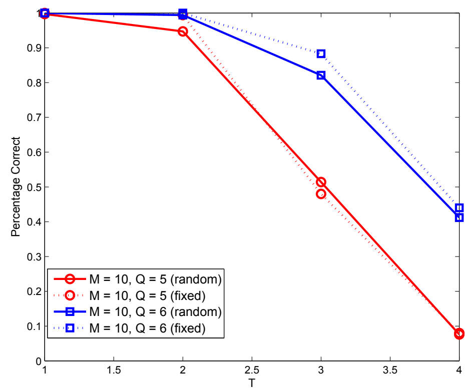

Fig. 2. Success Rate of (10) in correctly estimating the complex sparse channel $\mathbf{h}$ in a noise-free system with Q out of $M = 10$ randomly selected (solid) and fixed (dotted) tones for different T.

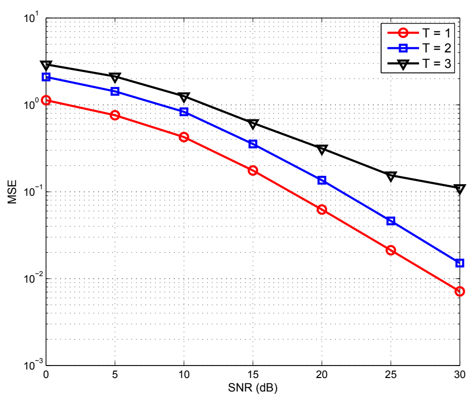

Fig. 3. Mean square error $\|\hat{\mathbf{h}} - \mathbf{h}\|_{l_2}^2$ for $M = L = 10$, $Q = 6$, with a fixed set of pilot tones $\Omega = \{1, 2, 5, 7, 9, 10\}$.

7. CONCLUSION

We have applied the theory of sparse signal recovery to the problem of channel estimation. Using a multi-carrier transmission scheme we have shown that the number of pilot tones needed to estimate a sufficiently sparse channel can be greatly reduced, thus increasing the spectral efficiency. In particular,in a noise-free system, we can perfectly estimate the channel if the number of pilot tones $Q \geq 2T$. In the noisy case, there are alternative methods for estimating the channel, and the l_1 norm regularization problem has been shown to be stable in the presence of noise. We have shown there are choices of pilot tone assignment which perform better than the case of randomly selected pilot tones. Further, in our experience, we have observed that with a uniform assignment of pilot tones, identifiability of the channel tends to be lost when M is not prime.

8. APPENDIX

8.1. Proof of Theorem 3.2

A proof of this theorem is given in [13], and we state a similar proof here.

Suppose there exists $\mathbf{g}$ and $\mathbf{h}$ such that $G(e^{j\omega_q}) = H(e^{j\omega_q})$. Let the support of $\mathbf{g}$ be denoted by $\tilde{\mathcal{T}}$ and the support of $\mathbf{h}$ by $\mathcal{T}$, and $\tilde{T} = |\tilde{\mathcal{T}}|$, $T = \mathcal{T}|$. Assuming $\tilde{T} \leq T \leq \frac{1/2}{Q}$, and defining $\mathbf{e} = \mathbf{g} - \mathbf{h}$ with support $\epsilon = \tilde{\mathcal{T}} \bigcup \mathcal{T}$, then we must have $|\epsilon| \leq Q$. Let $\epsilon = \{\xi_1, \xi_2, ..., \xi_{|\epsilon|}\}$, then due to Lemma 3.1, the matrix $\mathbf{A}$ given by:

$$\{\mathbf{A}\}_{q,t} = e^{-\frac{j2\pi p_q \xi_t}{M}}, \quad t = 1, ..., |\epsilon|$$

must have full column rank and the linear map $\mathcal{A}$ associated with the matrix $\mathbf{A}$ must be injective. Therefore it is impossible for $E(e^{j\omega_q}) = 0$. Thus by contradiction, $\mathbf{g} = \mathbf{h}$. ■

9. REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[2] M. Cetin and B. Sadler, “Semi-blind sparse channel estimation with constant modulus symbols,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’05)*, vol. 3, Nagoya, Japan, Mar. 2005, pp. 561–564.

[3] S. Cotter and B. Rao, “Sparse channel estimation via matching pursuit with application to equalization,” *IEEE Trans. on Comm.*, vol. 50, pp. 374–377, Mar. 2002.

[4] O. Rabaste and T. Chonavel, “Estimation of multipath channels with long impulse response at low SNR via an MCMC method,” *IEEE Trans. Sig. Proc.*, vol. 55, pp. 1312–1325, Apr. 2007.

[5] W. F. Schreiber, “Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions,” in *Proc. IEEE*, vol. 83.

[6] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, “Reduced complexity decision feedback equalization for multipath channels with large delay spreads,” *IEEE Trans. Comm.*, vol. 47, pp. 927–937, June 1999.

[7] M. Kocic *et al.*, “Sparse equalization for real-time digital underwater acoustic communications,” in *Proc. OCEANS’95*, San Diega, CA, 1995, pp. 1417–1422.

[8] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Sig. Proc.*, vol. 41, pp. 3397–3415, Dec. 1993.

[9] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Info. Theory*, vol. 51, pp. 4203–4215, 2005.

[10] W. U. Bajwa, J. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, “Toeplitz-stuctured compressed sensing matrices,” in *Proc. IEEE Workshop on Stat. Signal. Proc. (SSP’07)*, Madison, WI, Aug. 2007, pp. 294–298.

[11] T. Tao, “An uncertainty principle for cyclic groups of prime order,” *Math Res. Letter*, vol. 12, pp. 121–127, 2005.

[12] P. Stevenhagen and J. H.W. Lenstra, “Chebotarev and his density theorem,” *Math Intelligencer*, vol. 18, pp. 26–37, 1996.

[13] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. on Info. Theory*, vol. 52, pp. 489–509, 2006.

[14] ——, “Stable signal recovery from incomplete and innacurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2005.