Estimation of Sparse Multipath Channels

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

Anna Scaglione
Dept. of Electrical and Computer Engineering
University of California, Davis, CA 95616
Email: ascaglione@ucdavis.edu

Abstract—In many communication systems the channel impulse response can be characterized with a parametric form, though the channel estimation is often performed using an equivalent discrete-time linear time-invariant system (usually modeled as a Moving Average (MA) system). When the number of parameters which describe the channel is less than the number of unknowns in the MA model, the ML estimate of the parameters describing the channel may lead to a better estimate of the channel response. However, this ML estimation procedure is highly complex.

The objectives of this paper are 1) to cast the parameter estimation problem as a sparse estimation problem, 2) to compare the performance of this estimate with the CRB of the parameter estimation problem and the least squares estimate, and 3) to present novel guidelines on the amount of resources which one must devote to training for identification of the channel.

I. INTRODUCTION

In wireless communication systems, the channel impulse response must typically be acquired before decoding the data. For this purpose, it is common to insert training or probe sequences periodically into the data stream either in frequency or time. In this paper we are concerned with the estimation of multi-path channels where the transmitted signal arrives at the receiver over several propagation paths, each with a corresponding delay and path gain. A pertinent question to ask is how many resources must be devoted to training for accurate estimation of this channel.

Under the assumption we have a parametric model for the channel impulse response, the ML estimation procedure is highly complex due to the unknown number of paths as well as the continuous valued delays associated with these paths. For this reason it is common to model the channel as a Moving Average (MA) system, whose order is known a priori. In the presence of Gaussian noise, the corresponding ML estimate is equivalent to a linear least squares estimate (LSE) and is efficient, meaning it achieves the Cramer-Rao bound (CRB).

If the number of parameters which describe the channel is much less than the number of unknown coefficients in the MA system, for large enough observations, one expects a performance loss in the LSE in comparison to the ML estimate of the parameters describing the channel. However, the ML procedure is too complex, searching over the order of the paths, and for each order trying to estimate the associated path gains and continuous valued delays.

Under the assumption that the MA model is over-parametrizing the channel, a different class of approaches has been proposed in [1], [2], [3], [4]. The idea is to cast the original estimation problem as one which is sparse in the number of parameters. The goal is then to find the sparsest solution to a system of linear equations given some constraint on the reconstruction error. This methodology mediates between the strategy of estimating the MA discrete-time model and that of estimating the continuous valued parameters capturing the multi-path propagation.

In this paper we examine the multi-path channel estimation problem, and similar to [1] and [4] provide a model capitalizing on sparsity. We compare the performance of the proposed estimator with the CRB of the corresponding continuous valued parameter estimation problem which is derived in Section II-A. We will provide insight into the design of the training sequences, specifying the amount of resources as well as the type of training, which guarantee identifiability of the channel in the absence of noise in Section IV-C.

II. SYSTEM MODEL

We consider a single-input single-output (SISO) system affected by a linear time-invariant channel. The corresponding complex-baseband received signal in the presence of additive white Gaussian noise is given as:

\[ y(t) = \sum_{p=1}^{P} a_p s(t - \tau_p) + w(t), \]

where the coefficients \( a_p \in \mathbb{C} \) and \( \tau_p \) are the attenuation and delay of the \( p \)-th propagation path. We make the assumption that the multi-path is causal (i.e. the delays \( \tau_p \geq 0 \forall p \)). The complex envelope \( s(t) \), of the digitally modulated transmit signal is:

\[ s(t) = \sum_{n=0}^{N-1} x[n] p(t - nT_s) \]

where \( 1/T_s \) is the symbol rate, \( p(t) \) is the convolution of the pulse shaping filter at the transmitter and the matched receive filter, which we assume is a Nyquist pulse, and \( x[n] \) is a sequence of \( N \) pilot symbols. After sampling the received
signal at a rate $W = 1/T_s$, the discrete-time signal is:

$$y[k] = \sum_{n=0}^{N-1} x[n] \sum_{p=1}^{P} a_p p ((k - n)T_s - \tau_p) + w[k]$$

(3)

where $w[k] \sim \mathcal{CN}(0, \sigma_w^2)$ are samples of additive white Gaussian noise. As common in wireless settings, we assume that the channel has finite memory $L - 1$. Hence, we may express our received signal as:

$$y[k] = \sum_{l=0}^{L-1} h[l] x[k - l] + w[k]$$

(4)

where now

$$h[l] = \sum_{p=1}^{P} a_p p (lT_s - \tau_p)$$

(5)

In vector notation we have:

$$y = X h + w$$

(6)

where $h = (h[0], \ldots, h[L-1])^T$ and denoting $h^*$ as the initial sample of our observations, the Toeplitz matrix $X$ is:

$$\{X\}_{kl} = x[k + k^* - l]; \quad k \in [0, K - 1], l \in [0, L - 1].$$

(7)

Commonly, the coefficients of the MA model of the channel are estimated directly, i.e. the vector $h$. Under the assumption that $X^H X$ is invertible, this is a simple linear estimator in the observations, where the LSE is given by:

$$\hat{h}_{LS} = (X^H X)^{-1} X^H y$$

(8)

Clearly, in order to uniquely identify the channel in the absence of noise $K \geq L$ and rank($X$) $= L$. In the presence of noise, one can increase the length of the training sequence $x$ at the expense of the data rate, in order to improve the accuracy of the channel estimate.

Since the CRB is a lower bound for any unbiased estimate of the channel, it is a natural choice for a performance comparison. Further, by examining the conditions required for the ML estimate, we can provide insight into the design of accurate channel estimators.

A. The Multipath Parametric Model and its CRB

As opposed to the LSE (8) of the coefficients of the channel impulse response, in the multi-path parametric model (5), instead, we wish to estimate the unknown parameters $\theta = \{P, a, \tau\}$ which consist of two sets of continuous parameters, namely the complex path gains $a = (a_1, \ldots, a_P)$ and the real delays $\tau = (\tau_1, \ldots, \tau_P)$, as well as the discrete-parameter $P$ indicating the order. If the number of propagation paths $P$ is such that $2P < L$, the estimation of the MA model of the channel is over-parameterizing the observations, and it may be advantageous to estimate the parameters $\theta$ of the channel instead.

In the derivation of the CRB and the ML estimator, we make the assumption that the total number of paths $P$ is known a priori, and therefore the parameter vector $\xi = \{a^*, a, \tau\}$. The observations may then cast into the following vector model:

$$y = XG(\tau)a + w = S(\tau)a + w$$

(9)

where the matrices $G(\tau)$ and $S(\tau)$ for $p \in [1, P]$:

$$\{G\}_{lp} = p (lT_s - \tau_p)$$

$$\{S(\tau)\}_{kp} = \sum_{l=0}^{L-1} x[k^* + k - l]p (lT_s - \tau_p)$$

(10)

An estimate of the channel impulse response $h$ can now be obtained from an estimate of the path gains $a$ and delays $\tau$ as:

$$\hat{h}_{PM} = G(\hat{\tau}) \hat{a}$$

(11)

In the following, for simplification of notation, we omit the dependence on $\tau$ and use $S$ and $G$ in lieu of $S(\tau)$ and $G(\tau)$. The log-likelihood function for $\xi = \{a^*, a, \tau\}$ given the observation $y$ is:

$$\ln f(y; \xi) = C - \frac{1}{\sigma_w^2} \|y - Sa\|^2_2$$

(12)

where $C$ is a constant. In order to maximize the likelihood function, we take the complex gradient of the likelihood function with respect to $a^*$ and set it equal to zero. This results in the following condition on the estimate of the path attenuations:

$$\hat{a} = (S^H S)^{-1} S^H y$$

(13)

Likewise, taking the gradient of the likelihood function with respect to $\tau$, and substituting (13) in this expression results in the condition:

$$\mathbb{R}\{\text{diag}(y^H S (S^H S)^{-1}) D^H (I - S (S^H S)^{-1} S^H ) y\} = 0$$

(14)

where $D$ is a $K \times P$ matrix with columns:

$$\{D\}_{kp} = \frac{\partial}{\partial \tau_p} \sum_{l=0}^{L-1} x[k^* + k - l]p (lT_s - \tau_p)$$

This structure is intuitively pleasing, since the matrix $(I - S (S^H S)^{-1} S^H)$ spans the space orthogonal to the column span of $S$ and in the noiseless case $(I - S (S^H S)^{-1} S^H) y = 0$.

The complex Fisher information matrix for the parameter vector as $\xi = \{a^*, a, \tau\}$, is:

$$I(\xi) = \frac{1}{\sigma_w^2} \begin{bmatrix}
S^H S & 0 & S^H D A \\
0 & S^H D^H S & S^H D A^* \\
A^H D^H S & A^T D^H S^* & 2 \mathbb{R} \{A^H D^H DA\}
\end{bmatrix}$$

(15)

where $A = \text{diag}(a)$ is a $P \times P$ diagonal matrix. Thus the Cramer-Rao lower bound for the propagation delays is [5]:

$$\text{CRB}^{-1}(\tau) = \frac{2}{\sigma_w^2} \mathbb{R} \{A^H D^H (I - S (S^H S)^{-1} S^H ) DA\}.$$ 

(16)

It is now obvious that the estimation of the path delays is affected by the corresponding path amplitudes through the

\[^1\text{We assume that } S^H S \text{ is full rank.}\]
matrix $A$. Note that if $P > K$ in general $S^H S$ is not invertible and $(S^H S)^{-1}$ will have to be replaced by $(S^H S)^1$. In this case $S(S^H S)^1 S^H = I_{K \times K}$ for any $S$ with full row rank, implying the Fisher Information matrix is rank deficient and the delays are no longer identifiable. The CRB of the path amplitudes obtained from (15) is:

$$CRB(a) = \sigma_a^2 (S^H S)^{-1} + \Psi$$ (17)

where

$$\Psi = (S^H S)^{-1} S^H DACRB(\tau) A^H D S (S^H S)^{-1}.$$ (18)

Equation (17) clearly shows that the amplitude estimation is penalized when the delays are not known and is coupled with the delay error via the additional contribution represented by $\Psi$.

In order to compare the performance of the estimators we propose in the next section, we must map the performance of the estimator for $\xi$ to the estimate of the impulse response $h$. Since the impulse response $h_{PM}$ can be expressed as a function of the parameters via (11), the MSE of $h_{PM}$ is lower bounded by, [6]:

$$MSE_{PM} \geq tr(J^{-1}(\xi) J^H).$$ (19)

where $J = [G, 0, G'A]$ is the Jacobian of the transformation (11). Using the formula for the inverse of a block matrix as well as the matrix inversion lemma, with some algebra we obtain:

$$JI^{-1}(\xi) J^H = GCRB(a) G^H + G' ACRB(\tau) A^H G^H - \Phi,$$ (20)

where $G'$ is the gradient of $G$ given by:

$$\{G'_l\}_P = \frac{\partial}{\partial \tau_p} p(lT_s - \tau_p), \quad l \in [0, L - 1], \quad p \in [1, P]$$ (21)

and

$$\Phi = G' ACRB(\tau) A^H D S (S^H S)^{-1} G^H + G(S^H S)^{-1} S^H DACRB(\tau)(G'A)^H$$ (22)

Now that CRB has been derived, we proceed to map our problem into a sparse estimation problem and compare its performance to that of the CRB.

### III. Sparse Channel Model

In certain situations, the channels may be characterized by large delay spreads with relatively limited multipath. In these situations the number of parameters describing the channel impulse response may be much smaller than the channel length $L$. Thus the impulse response can be considered *sparse*, namely, there are only a few nonzero taps in $h$. Before describing this sparse channel model, we provide a brief review of sparse signal reconstruction.

### IV. Sparse Signal Recovery

Frequently, signal processing problems are mapped into a linear observation model consisting of some linear transformation of sparse vector, such as:

$$u = \Phi v_0$$ (23)

where $\Phi \in \mathbb{R}^{K \times L}$, while $v_0$ is $S$-sparse, no more than $S$ of its elements are nonzero. An important question to answer is whether the vector $v_0$ is identifiable. Obviously, if $K \leq L$ or rank($\Phi$) $< L$, the system of equations is underdetermined and the solution is not unique. However, we have the additional knowledge that $v_0$ is sparse. We thus search over the solution space for the sparsest solution, as in the following optimization problem:

$$\hat{v} = \arg \min_\varepsilon \| \varepsilon \|_0 \quad \text{s.t.} \quad u = \Phi \hat{v}.$$ (24)

However, this optimization problem is generally combinatorial in nature. A number of authors [7], [8] have proposed a more tractable relaxation of (24), namely:

$$\hat{v} = \arg \min_\varepsilon \| \varepsilon \|_1 \quad \text{s.t.} \quad u = \Phi \hat{v}$$ (25)

which is a convex optimization problem which can be solved using standard algorithms. There are still two fundamental questions to answer: 1) when is a sparse solution to the system of equations necessarily the sparsest, implying (24) results in a unique solution, and 2) under what conditions is a unique solution to (24) also a unique solution of (25), or alternatively, when are the two problems equivalent. This is currently the subject of a large amount of research, including [9], [8], [10].

In the literature, these two questions are often addressed using the following two metrics [11]:

**Restricted Isometry Constant:** Assume that the columns of $\Phi$ have unit Euclidean norm. The $S$ restricted isometry constant $\delta_S$ is defined as the smallest constant such that for every subset of $T \leq S$ columns of $\Phi$, the matrix $\Phi_T$ which contains them is such that:

$$(1 - \delta_S) ||v||_2^2 \leq ||\Phi_T h||_2^2 \leq (1 + \delta_S) ||v||_2^2$$

**Restricted Orthogonality Constant:** The $S, S'$ restricted orthogonality constant $\theta_{S,S'}$ with $S + S' \leq L$ is the smallest quantity such that:

$$|\langle \Phi_T v, \Phi_T v' \rangle| \leq \theta_{S,S'} \| v \|_2 \| v' \|_2$$

holds for all disjoint sets $T, T' \subset \{1, ..., L\}$ of cardinality $|T| \leq S$ and $|T'| \leq S'$. In regard to first question, that of a unique sparsest solution, every subset of $2S$ columns of $\Phi$ must be linearly independent, which is implied by $\delta_{2S} < 1$. Sufficient conditions for equivalence between the solutions of the two problems based on these metrics are rather pessimistic and difficult to check.

This discussion has been limited to noiseless observations, however, in the presence of noise our model is now:

$$u = \Phi v_0 + w$$ (26)
where we assume $w \sim \mathcal{CN}(0, \sigma_w^2 I)$. The problem (25) can be modified by constraining the reconstruction error. An example, and the one we will use in this paper, is the following convex optimization problem [11] (known as the Dantzig Selector):

$$\hat{v} = \arg\min_{v} \|v\|_1 \text{ s.t. } \|\Phi H (u - \Phi v)\|_\infty \leq \beta \quad (27)$$

The following theorem, adapted from [11], states that the error of the Dantzig selector is bounded:

**Theorem 4.1:** Suppose $\|v_0\|_0 \leq S$ and the matrix $\Phi$ satisfies $\delta_{2S} + \theta_{2S} < 1$, the squared error obtained using the Dantzig selector is such that [11]:

$$\|v_0 - \hat{v}\|_2^2 \leq C_1^2 S \beta^2$$

with probability exceeding $1 - L \exp(-\beta^2/\sigma_w^2)$ where $C_1 = 4/(1 - \delta_S - \theta_{2S})$.

Since the corresponding ML estimate of $v_0$ must satisfy $\Phi H (\Phi \hat{v} - u) = 0$, the Dantzig Selector appears a natural choice for estimating $v_0$. In order to justify the use of sparse estimators, we must map the channel model into a sparse model, which is the subject of the next section.

### A. Channel Estimation as a Sparse Estimation Problem

The multipath nature is the result of scattering from several reflectors. The delays corresponding to a particular scatterer may have very small difference in comparison to the symbol period $T_s$, while the difference in delays from different scatterers may be several symbol periods or larger. In this case, the delays corresponding to a particular scatter coalesce into a single propagation path with delay $\tau_p$ and attenuation $a_p$. Thus if there are only a few scatters with a large delay spread, the channel can be considered sparse. This idea is exploited in [1] and [2], where the authors make the assumption that the impulse response $h$ is sparse. Similar to [4], instead, the parameter space of $\tau$ can be discretized assuming that each $\tau_q = \tau_q^* \in T_s$ where $\tau_q^* \in \mathbb{Z}$ and the signal is bandlimited, resulting in the following equivalent model:

$$h[l] = \sum_{r \in \mathcal{R}} a_r \text{sinc}(\pi W (l T_s - r \in T_s)). \quad (28)$$

Note that this model does not neglect the sidelobes of each pulse replica as in [1] and [2]. In this case the number of non zero coefficients $a_r$ is approximately equal to the number of scatterers which may be much less than $|\mathcal{R}|$, where $|\mathcal{R}|$ is the cardinality of the set $\mathcal{R}$. We adopt a similar parametrization where in general the sinc pulse is replaced by a Nyquist pulse, i.e.:

$$h[l] \approx \sum_{r \in \mathcal{R}} a_r \text{sinc}(\pi W (l T_s - r \in T_s)). \quad (29)$$

Adopting (29) as the channel model, the channel and the observation are given by the following vectors:

$$h = G \alpha \approx P \alpha \quad \Rightarrow \quad y \approx X P \alpha + w, \quad (30)$$

where:

$$\{P\}_{lr} = p(l T_s - r \in T_s), \quad l = 0, \ldots, L - 1; \quad r \in \mathcal{R}, \quad (31)$$

and $\alpha$ is a sparse vector which is non zero on the support:

$$S = \{ \forall \in \mathcal{R} : \min |r\tau_s - \tau_p|, p \in [1, P]\}.$$ 

The $\alpha_r, \forall \in S$ represent the corresponding path amplitudes. In essence, the idea is to approximate the actual $P$ columns of $G$, which are smooth functions of the parameters $\tau$, as an unknown subset of columns from the fat matrix $P$, forming an over-complete basis for $h$ which is likely to lead to a sparse vector of coefficients $\alpha$. The Dantzig selector for this model is (see also [4]):

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 \text{ s.t. } \|X P \alpha \|_\infty \leq \beta \quad (32)$$

In the following, we denote the support of the unknown sparse vector $\alpha$ (or $h$ when assuming the impulse response is sparse) as $S$, and its cardinality as $\mathcal{S}$, i.e. the number of nonzero components in the vector.

### B. Performance

In the following, we assume that $h$ can be exactly modeled as $h = P \alpha_0$ where $\alpha_0$ has support $S$. Let $\hat{h}_{SP}$ denote the estimate obtained from the sparse mapping (29) discussed above, i.e.:

$$\hat{h}_{SP} = P \hat{\alpha}, \quad (33)$$

and let:

$$MSE_{SP} = E\{\|\hat{h}_{SP} - h\|^2\}. \quad (34)$$

Let $P_S$ be the $K \times S$ matrix containing the columns of $P$ which correspond to the support $S$ of $\alpha$. A simple lower bound for $MSE_{SP}$ is obtained by assuming the support $S$ is known, which we refer to as the genie-aided bound, namely:

$$MSE_{SP} \geq \sigma_w^2 \text{tr}(P_S (P_S^H X^H X P_S)^{-1} P_S^H) \sim O(\sigma_w^2 S). \quad (35)$$

Neglecting the possible residual bias, the analysis of the CRB in Section II-A can be used to provide a tighter lower bound for the MSE since the model used in deriving the CRB assumes only knowledge of the cardinality of the support but not the support itself, i.e.

$$MSE_{SP} \gtrsim \text{tr}(J^{-1}(\xi) J^H). \quad (36)$$

The sparse estimator (32) is in general biased, as the estimator tends to a sparse solution, meaning its estimate is biased towards the all zero vector. This is clear in that if the constraint on the reconstruction error $\beta$ is relaxed enough, the zero solution will minimize the L1 norm. It is still useful to compare the sparse estimator performance with the CRB, as it can be used as a reference for comparison with other unbiased estimates of the channel. Further, this bound naturally can be used as a criterion for deciding whether or not it is beneficial to estimate the parameters $\xi$ of the channel in order to improve upon the performance of the LS estimate (8). The criterion is that of comparing the bound obtained from the CRB with MSE of the LS estimator:

$$MSE_{LS} = \sigma_w^2 \text{tr}(X H X^{-1}) \sim O(\sigma_w^2 L) \quad (37)$$
Clearly, if
\[ MSE_{LS} < tr(JI^{-1}(\xi)J^H), \]
then an unbiased estimator of the \( \tau \) and \( \alpha \) cannot outperform the LS estimator in terms of mean-squared error. It may also be possible to predict the trend of the error of the sparse estimator using the following lemma:

**Lemma 4.2:** Assume \( \|\alpha_0\| \leq S \), \( \beta = \sqrt{\sigma_w^2(1+a) \log R} \) with \( a \geq 0 \) and the matrix \( XP \) is such that \( \delta_{2S} + \theta_{S,2S} < 1 \). With probability exceeding \( 1 - 1/R^a \), the squared error

\[ \|P(\alpha_0 - \hat{\alpha})\|^2 \leq \lambda_{\max}(P^H P)C_1^2 \sigma_w^2 S(1 + a) \log R \] (39)

**Proof:** As a consequence of Theorem 4.1, with probability exceeding \( 1 - 1/R^a \), the norm of the error \( \|\hat{\alpha} - \alpha_0\|^2 \leq C_1^2 \sigma_w^2 (1 + a) S \log R \). The proof is completed by noting:

\[ \|P(\hat{\alpha} - \alpha_0)\|^2 \leq \lambda_{\max}(P^H P)\|\hat{\alpha} - \alpha_0\|^2. \]

An interesting observation is that under the conditions in which Lemma 4.2 is valid, the squared error of the Dantzig selector scales as \( O(\sigma_w^2 s \log R) \) and for large enough \( R \), will tend to have the same trend as the genie-aided estimate, which is \( O(\sigma_w^2 s) \). Further, as a result of the structure of (20) it is not difficult to observe that the bound (36) is \( O(\sigma_w^2 P) \), and since \( S \approx P \) this bound has a similar trend as well.

**C. Training Design for Identifiability**

In this section we provide new guidelines on the design of the training sequence in order to guarantee the identifiability (i.e. a unique sparsest solution) of the channel in the absence of noise via a sparse mapping. We focus on sufficient conditions such that the following problem has a unique solution:

\[ \min \|\hat{\alpha}\|_0, \quad \text{s.t.} \quad y = XP\hat{\alpha}. \] (40)

Unlike other random projection models common to the sparse signal recovery literature, the matrix \( X \) or \( XP \) has a specific Toeplitz structure. As a consequence of this structure, rather than trying to provide conditions on the number of observations \( K \) needed using a pseudo-noise training sequence, we focus on the structure of the column span of \( X \) that will guarantee identifiability. This is advantageous as it allows us to clarify the number of dimensions which must be devoted to training, while the orthogonal subspace of \( X \) can still be used to transmit unknown information which can be eliminated by right multiplication \( X^H y \) when estimating the channel.

Thus we map the problem to an equivalent formulation where the observations are a linear combination of columns from a Vandermonde matrix. Under the conditions in the following Lemma, modified from [12], each row of this Vandermonde matrix has distinct elements.

**Lemma 4.3:** Let \( K \) be a prime number, then the matrix:

\[ \{F\}_{q,t} = |\gamma[q]| e^{-j2\pi q j d_k}, \quad t = 1,...,T, \quad q = 1,...,T \] (41)

has nonzero determinant (and is therefore invertible), for any \( T \leq K \), if \( p_1,...,p_T \) and \( d_1,...,d_T \) are distinct elements of the integers modulo \( K \) and \( \gamma[q] \neq 0 \) \( \forall q \).

This lemma implies that any \( Q \times T \) submatrix of \( F \) with \( Q \geq T \) has full column rank, and is an injective mapping. In order to make use of this lemma and make our analysis tractable, we make several assumptions on the observation size \( K \) relative to \( L \) and \( N \), the initial sample \( k^* \), and the structure of the input sequence \( x[n] \). However, for large \( K \) and \( L \) our analysis sheds insight on more general cases. More specifically, we assume that:

\[ k^* = L - 1; N = K + L - 1; R = [0, R - 1]; \]

\[ x[n] = x[K + n], \quad \text{for} \quad n = 0,...,L - 1. \] (42)

These assumptions imply the matrix \( X \) is circulant. In the following we define \( \gamma[m] \) as the following Fourier coefficient:

\[ \gamma[m] = \sum_{k=0}^{K-1} x[k^* + k] e^{-j2\pi mk}, \quad m \in [0, K - 1]. \] (43)

The coefficients \( \gamma[m] \) characterize the spectral support of the training sequence, specifically, let \( Q = |\gamma[0]| \), then \( Q \) specifies the amount of spectral support. Restating identifiability conditions which we provided in [3] for the case of a sparse impulse response \( \mathbf{h} \).

**Lemma 4.4:** Let \( \mathbf{h} \) be such that \( \|\mathbf{h}\|_0 \leq S \), then the solution to the following problem is unique:

\[ \min \|\hat{\mathbf{h}}\|_0, \quad \text{s.t.} \quad y = X\hat{\mathbf{h}}. \] (44)

if \( Q \geq 2S \) and \( K \) is prime.

**Proof:** See Appendix A.

For the sparse mapping (30), we replace \( h[l] \) with (29). We thus arrive with a similar lemma for the identifiability of this mapping.

**Lemma 4.5:** For the sparse mapping (30), with \( L = K \), \( K/\epsilon \) prime, and \( p(t) = \text{sinc}(\pi Wt) \), the coefficients \( \alpha \) and \( \mathbf{h} \) are identifiably if the \( Q \geq 2S \) where \( \|\alpha\|_0 \leq S \).

**Proof:** See Appendix B.

In the case where our training consists of a sum of pilot tones, such as OFDM, this analysis implies that the number of pilot tones needed to uniquely identify the channel impulse response \( \mathbf{h} \) in the absence of noise is only twice the number of nonzero components in the impulse response. This is pleasing as we can be much more efficient in the use of our bandwidth. For example, in the case of the least squares estimator, we require \( L \) out of \( K \) possible tones to be used for training. However, under the assumption that the channel impulse response \( \|\mathbf{h}\|_0 \leq S \), we require only \( 2S \) tones and thus our efficiency:

\[ \rho_{SP} = \frac{K - 2S}{K + L - 1} \leq \frac{K - L}{K + L - 1}. \]

We note that these results do not imply uniqueness of the solution to (25). However, they can be used as guidelines for the design of training sequences for identification of the channel.
V. Numerical Results

We simulate the performance of the sparse channel estimator based on our mapping and compare it with the genie-aided and CRB bounds derived. The training sequence is constructed such that the matrix \( X \) is circulant and further each of the \( Q \) nonzero coefficients \( \gamma[m] = 1 \) for all \( m \in Q \). We set the number of observations \( K = 11 \) and the number of channel taps \( L = 11 \), and \( p(t) = \text{sinc}(\pi W t) \) for all simulations.

In Fig. 1 and Fig. 2, the channels are directly generated using the model (29), where \( S \) denotes the support of \( \alpha \). The support \( S \) is randomly generated uniformly and each of the \( S \) nonzero elements \( \alpha_r \sim \mathcal{CN}(0, 1/S) \) \( \forall r \in S \) where we have normalized the path gains by \( S \) in order to eliminate any performance gain as \( S \) increases. The \( S \) randomly selected indices \( r \in S \) each correspond to a delay \( \tau_r = r \varepsilon T_s \), where we have selected \( R = 27 \) and \( \varepsilon = K/29 \) such that \( K/\varepsilon \) is prime. In all the simulations we define the \( SNR = Q/(K \sigma_w^2) \).

We denote Sparse 1 as the estimate under the assumption the impulse response \( h \) is sparse, i.e.

\[
\hat{h} = \arg \min_{||h||_1} ||X^H(y - Xh)||_\infty \leq \beta, \quad (45)
\]

and Sparse 2 as the estimate obtained using (32), with \( \beta = \sqrt{2 \sigma_w^2 \log(L) } \) and \( \beta = \sqrt{2 \sigma_w^2 \log(R) } \), respectively. For both estimators, the Dantzig Selector is implemented using a log-barrier interior-point method.

In Fig. 1, we set \( Q = L \) and thus all the \( \gamma[m] = 1 \) for all \( m \). The MSE of the channel estimate is compared to the genie-aided LS bound (35), the LS MSE, as well as the bound (19) from the CRB. In Fig. 2 we now examine the case where \( Q = 6 \) where we fix \( Q = \{1, 3, 5, 7, 9, 11\} \) implying the system of equation is underdetermined. We also plot the MSE of the channel impulse response estimate obtained using (45), though we note that in this case, the impulse response \( h \) may not truly be sparse due to the sidelobes of the pulse shaping filter \( p(t) \) for different delays.

In Fig. 3, we generate the channel according to the model (5), where the delays \( \tau_p \sim \mathcal{U}(0, (L - 1)T_s) \) and \( a \sim \mathcal{CN}(0, S^{-1}) \). In this case we again set \( Q = 11 \). The MSE of the estimates obtained using (45) and (32) is compared with the CRB.

In Fig. 1, we observe that when the channel is generated according to the model (29), the performance of the sparse channel estimator is actually better than the performance of the LS estimator and has the same error trend as the CRB and genie-aided bounds.

In Fig. 2, we see that even when the spectral support \( Q < L \), the sparse estimator can indeed recover the channel impulse response. For the case of a single path, it even has the same trend as the genie-aided bound. The estimate obtained from (45) cannot however recover the channel impulse response as the impulse response is not necessarily sparse. We note that for the case \( P = 3 \), the sparse estimator is not able to reliably estimate the channel. However, the Fisher Information matrix \( I(\xi) \) is singular for most delays in this case as well.

Fig. 3 demonstrates that there is a slight decrease in the performance of the sparse estimator, though it still is able to outperform the LS estimator even in the case when the delays are truly continuous. We have observed that the performance can be improved by increasing the resolution of the discrete set of delays, i.e. by increasing \( R \) while appropriately decreasing \( \varepsilon \).

VI. Conclusions

In this paper we have proposed a mapping of multipath fading into a sparse observation model which allows us to use sparse signal recovery methods to estimate the channel impulse response. The performance approaches that of an ML estimator for a parametric multipath model. We also provided a criterion based on the CRB that can be used to assess when this approach may be superior to estimating the overall equivalent discrete time response, which is applicable to other similar approaches. Future directions include the extension to the case of multipath channels in the presence of doppler shifts as well as MIMO channels.
and therefore $\Gamma U$, must have full column rank as a result of Lemma (4.3). The matrix $h U = g$ enough. We make the following approximation (which is valid for large enough $W$).

$$\{WXP\}_{m,r} = \gamma[m] \left( \frac{1}{K} \sum_{l=0}^{L-1} e^{-j2\pi \frac{ml}{K}} \sum_{d=-K/2}^{K/2-1} e^{j2\pi d(l-cr)} \right)$$

If we make the assumption $L = K$, we observe:

$$\{WXP\}_{m,r} = \begin{cases} \gamma[m] e^{-j2\pi \frac{mr}{K}} & 0 \leq m \leq \frac{K}{2} - 1 \\ \gamma[m] e^{-j2\pi \frac{r(m-K)}{K}} & \frac{K}{2} \leq m \leq K - 1 \end{cases}$$

Hence, for the sparse mapping (30) with $K/\varepsilon$ prime and with $L = K$, the matrix $WXP$ satisfies the conditions of Lemma (4.3). The proof follows under the same arguments given in Appendix (A).

**APPENDIX A**

**PROOF OF LEMMA 4.4**

Let us introduce the DFT matrix:

$$\{W\}_{mk} = e^{-j2\pi mk/K} \quad k, m \in [0, K-1]$$

Under the conditions (42) the matrix $X$ is circulant and thus for $m \in [0, K-1], l \in [0, L-1]$:

$$\{WX\}_{ml} = \sum_{k=0}^{K-1} x[l-1+k-l] e^{-j2\pi kl/K} = \gamma[m] e^{-j2\pi ml/K}. \tag{46}$$

Hence, the FFT of the observation vector is such that:

$$Wy = WXh = \Gamma Wh. \tag{47}$$

where $\Gamma = \text{diag}(\gamma)$ and $W$ consists of the first $L$ columns of the matrix $W$. Let $Q = \|\gamma\|_0$ and $K$ prime. Suppose there exist vectors $g$ and $h$, such that $\Gamma W(g - h) = 0$. Let the support of $g$ and $h$ be denoted $T'$ and $T$, respectively, with $|T'|, |T| \leq S$ and $|T| \leq |T'| \leq Q/2$. Defining $e = g - h$ with support $U = T \cup T'$, then $|U| \leq Q$. The matrix $\Gamma W U$, where the matrix $W_U$ contains only the columns of $W$ corresponding to $U$, must have full column rank as a result of Lemma (4.3). The linear mapping associated with this matrix must be injective, and therefore $\Gamma W e = 0$ implies $g = h$.

**APPENDIX B**

**PROOF OF LEMMA 4.5**

The DFT of the observations is now given as:

$$Wy = WXP \alpha$$

We make the following approximation (which is valid for large enough $K$)

$$p((l - cr)T_s) \approx \frac{1}{KT_s} \sum_{d=-K/2}^{K/2-1} P\left( \frac{d}{KT_s} \right) e^{j2\pi d(l-cr)/K} \tag{48}$$

where $P(f)$ is the Fourier transform of $p(t)$, and for the case $p(t) = \text{sinc}(\pi W t)$, $P(d/(KTS)) = 1/W$. Thus

$$\{WXP\}_{m,r} = \gamma[m] \left( \frac{1}{K} \sum_{l=0}^{L-1} e^{-j2\pi \frac{ml}{K}} \sum_{d=-K/2}^{K/2-1} e^{j2\pi d(l-cr)} \right) \tag{49}$$

**REFERENCES**


