APPLYING CSISZÁR’S I-DIVERGENCE TO BLIND SPARSE CHANNEL ESTIMATION

Feng Wan and Urbashi Mitra

Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, USA

ABSTRACT

Compressed sensing (CS) has renewed interest in sparse channel estimation. Herein, a semi-blind, iterative, sparse channel estimation method is proposed. The new method is based on minimizing Csiszár’s I-divergence using Schulz & Snyder’s iterative deautocorrelation algorithm. First, it is shown that the desired methods can be adapted to the problem of interest. The proposed semi-blind method accurately estimates the significant tap locations of a sparse channel, and their corresponding magnitudes. A method for determining the channel coefficients up to a phase ambiguity is presented. The simulation results show that although limited pilots are used, the proposed semi-blind iterative algorithm achieves performance comparable to that of training-based compressed sensing methods.

Index Terms— sparse channel estimation, compressed sensing, Csiszár’s I-divergence, semi-blind, OFDM.

1. INTRODUCTION

A number of naturally occurring channels in wireless communications are best modeled as sparse channels, that is, channels whose time domain impulse response consists of a large number of zero taps. Such channels include those for more conventional radio-wireless communications, ultrawideband communications and underwater acoustic channels. A variety of algorithms have been used in the past to estimate such channels, exploiting sparsity to improve estimation from unstructured initial estimates followed by thresholding [1] to modeling as an on-off keying problem [2] to greedy based algorithms such as orthogonalized matching pursuit (see e.g. [3]). Many of these methods are sensitive to knowledge of the channel order. The introduction of compressed sensing (CS) [4, 5] has provided renewed interest in the sparse model estimation problem. By showing that classical training sequences satisfy the restricted isometry property [6] for convolutive channels, the application of CS with pilot sequences has proven quite successful for a variety of channel estimation problems: SISO wireless channels [6], and UWA channels [7]. These works have used a variety of iterative methods to solve l1 norm minimization problems or model order penalized least squares problems.

In order to satisfy the RIP, a large number of pilots are needed to achieve good performance for most of these CS methods. To increase spectral efficiency, a cyclic-prefix (CP) based blind method has been proposed for non-zero tap detection of OFDM systems in [8]. However, this detection scheme requires a large number of OFDM symbols as well as a large CP length in order to obtain accurate detection of non-zero tap locations. In our previous work [9], based on an analysis of the second-order statistics of the received signal passing through a sparse channel, an efficient semi-blind non-zero tap detection algorithm was developed for OFDM channel estimation. However, our method still required the use of pilots.

In this paper, we provide preliminary results on a promising approach for semi-blind iterative estimation of a sparse channel. As the noise-free correlation function of the received signal is equal to the correlation function of the channel, we propose to formulate the problem as a blind deautocorrelation problem. We propose to employ an iterative deautocorrelation technique [10] based on the minimization of Csiszár’s I-divergence [11]. The motivation for the use of these methods is the fact that with the iterative de-autocorrelation technique, it has been proven that the difference of two consecutive estimates converges to zero in the l1 norm [12]. Essential to the use of these methods is showing a key approximation of the magnitude of the noiseless received signal autocorrelation function which is done by exploiting properties of sparse channels. Pilots are employed to initialize the iterations and to resolve ambiguities. Our proposed method, which estimates the non-zero tap locations and the associated channel magnitudes, is compared to a variety of CS based methods: Orthogonal Matching Pursuit (OMP) [3], iterative detector/estimator (IDE) [13], Haupt-Nowak (HN) [14] and iterative hard thresholding (IHT) [15], and shown to offer comparable (and sometimes better) performance without the use of pilot sequences. Throughout the paper, we adopt the following notation: δ(·) Delta function, ∗ conjugate, ⊛ circular convolution.

2. SIGNAL MODEL AND PROPERTIES

In this section, we discuss two key properties: the white nature of a single OFDM symbol and an approximation of the magnitude of the noiseless output correlation function. It is this latter property which enables us to employ Csiszár’s I-divergence as a metric for the blind channel estimation.

We assume a sparse, deterministic, unknown channel, denoted by H ∈ C^{L×1}. The non-zero integer tap locations are described by l_d where d = 0, · · · , D − 1, where 0 = l_0 < l_1 < · · · < l_{D−1}. Thus, the non-zero tap coefficients are given by

\[ z(d) = h(l_d). \]  

(1)

The effective channel is given by z(d). For OFDM, if the length of the cyclic prefix is greater than or equal to L, the time-domain signal model for the frequency-selective fading channel is given by

\[ y(n) = h(n) \ast x(n) + v(n), \]  

(2)

where v(n) ∈ C is a spatio-temporally uncorrelated noise with zero-mean and variance \( \sigma_v^2 \).

The correlation function of the received signal y(n) is defined as,

\[ r_y(m) \triangleq E\{y(n)y^*(n-m)\}, \ (m = 0, 1, \cdots, P). \]  

(3)

Using Equations (2), (3) and v(n) = 0, one can easily obtain that, the estimated version of \( r_y(m) \) by using one OFDM symbol can be written as

\[ \hat{r}_y(m) = \sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1} h(l_1) h^*(l_2) \hat{r}_x(m+l_2-l_1), \]  

(4)

This work is supported in part by ONR N00014-09-1-0700, NSF CNS-0832186, NSF CNS-0821750 (MRI), and NSF OCE-0520324.
where $\hat{r}_y (m) \triangleq \hat{E} \{ x (n) x^* (n - m) \}$. One can easily verify that, for a constant modulus signal with the power of $\sigma_x^2 = 1$, $\hat{r}_x (m)$ is a Kronecker delta function. Thus, we have

$$\hat{r}_y (m) = \sum_{l=0}^{L-1} h (l) h^* (l - m).$$

(5)

It is seen from Equation (5) that the correlation function of the received signal estimated by using one OFDM symbol is the same as the correlation function of $h (l)$ in the noise-free case.

Due to the sparse channel assumption, we now show an approximation to the magnitude of the output correlation function, in the noiseless case, Equation (6). The I-divergence is only applicable to the deautocorrelation problem of a real-valued, non-negative signal. Thus, as previously noted, we shall examine estimation of the location of the non-zero taps of the sparse channel and the magnitudes of the channel coefficients. However, we underscore that our methods will be validated on complex channels of arbitrary phase in Section 4.

We wish to show that the following approximation holds for highly sparse channels,

$$|\hat{r}_y (m)| \approx \sum_{l=0}^{L-1} |h (l) | |h (l - m)|.$$

(6)

First, it is trivial to show that,

$$|\hat{r}_y (0)| = \sum_{l=0}^{L-1} |h (l) | |h (l)|.$$

(7)

Next we consider other lag values, i.e., $m > 0$. Due to channel sparsity, the correlation function of the received signal, $\hat{r} (m)$, is also sparse. In particular, we have shown [9] that $\hat{r} (m)$ can be expressed in terms of the effective channel $z (d)$ in the absence of noise as follows,

$$\hat{r}_y (m) = \sum_{i=0}^{D-1} \sum_{j=0}^{D-1} \delta (m - i_l + l_j) \sum_{l=0}^{D-1} z (i) z^* (j).$$

(8)

Our argument for the approximation is as follows. Due to sparsity, the likelihood that more than one pair of non-zero taps coinciding for a particular value of $m > 0$ is low. Thus the only relevant values of $m$ are such that $m \in \{ l_i - l_j \}_{i \neq j}$: further, it is only the contribution of the pair of channel taps $z (l_i)$ and $z (l_j)$ that constitutes the correlation value, that is,

$$\hat{r}_y (m = l_i - l_j) = \sum_{j=0}^{D-1} \sum_{i=0}^{D-1} \delta ((l_i - l_j) - l_i + l_j) z (i) z^* (j) = z (i) z^* (j), \ (i \neq j).$$

(9)

Thus, if our conjecture holds, we can re-write the magnitude of the noiseless received signal correlation function as

$$|\hat{r}_y (m)| \approx \sum_{i=0}^{D-1} \sum_{j=0}^{D-1} \delta (m - i_l + l_j) \sum_{z=i}^{D-1} z (i) \sum_{z=j}^{D-1} z^* (j).$$

(10)

It is straightforward to show that Equation (8) is equivalent to Equation (6) above.

Now, we consider the probability of no overlap, i.e., if we assume a length $L$ Bernoulli sequence with the probability of a non-zero tap being $\frac{D}{L}$, it is trivial to show that $\mathbb{E} [r_y (m)] = (L - m) (\frac{D}{L})^2$ which clearly converges to zero as $\frac{D}{L} \to 0$. Similarly, one can show that for four distinct tap locations $l_i, l_j, l_k, l_m$ we have the following $P [l_i - l_j = l_k - l_m] = 2 \left( \frac{D}{L} \right)^2 \sum_{l=0}^{L-1} (1 - \frac{l}{L})^2$ which at worst case is of the order $O \left( \frac{L^4}{L^2} \right)$. Thus, it is clear that for significant sparsity, our conjecture holds; however, this will not be true in practice. In the case of more than one pair of taps contributing to a term in the autocorrelation function, we can show that a related channel will be estimated, wherein the tap locations will be properly estimated, but a function of the desired coefficients will result. Given that the estimation of MSTs is the most challenging aspect of sparse channel estimation, the algorithm retains some utility.

3. PROPOSED SEMI-BLIND ITERATIVE ALGORITHM

Equation (6) describes an autocorrelation of a real-valued nonnegative signal, thus, blind estimation of the sparse channel without phase information can be obtained by using the deautocorrelation technique on the minimization of Csizár’s I-divergence, which is given by,

$$I (\hat{r}_y; \hat{r}_y') = \sum_{m=-P}^{P} \left\{ |\hat{r}_y (m)| \ln |\hat{r}_y (m)| + |\hat{r}_y' (m)| - |\hat{r}_y (m)| \right\}$$

(11)

where

$$|\hat{r}_y' (m)| = \sum_{l=0}^{L-1} |\hat{h} (l) | |\hat{h} (l - m)|,$$

(12)

and $\hat{r}_y$, $\hat{r}_y'$ are the vectors whose components are comprised of $r_y (m)$ and $r_y' (m)$, respectively. Our semi-blind channel estimation problem can be formulated as: given $\hat{r}_y$, the absolute value of the estimated second order statistics (SOS) of the received signal, find a $\hat{h}$ such that

$$\hat{h} = \arg \min \frac{I (\hat{r}_y; \hat{r}_y')}{|\hat{h}]}$$

(13)

where $\hat{h}$ is comprised of $|\hat{h} (l)|, \ (l = 0, 1, \ldots, L - 1)$.

We employ the iterative algorithm of [10] tailored to the sparse channel case. The necessary conditions for $\hat{h}$ to satisfy Equation (13) are the following stationary point conditions:

$$\frac{\partial I (\hat{r}_y; \hat{r}_y')}{\partial |\hat{h} (l)|} = 0, \ (l = 0, \ldots, L - 1).$$

(14)

Equation (11), we have

$$\frac{\partial I (\hat{r}_y; \hat{r}_y')}{\partial |\hat{h} (l)|} = \sum_{m=-P}^{P} \left\{ 1 - |\hat{r}_y (m)|^2 - |\hat{r}_y' (m)|^2 \right\}$$

(15)

Using Equation (12), one can obtain

$$\frac{\partial |\hat{r}_y' (m)|}{\partial |\hat{h} (l)|} = |\hat{h} (l - m)| + |\hat{h} (l + m)|.$$
Substituting Equations (16) - (18) into (15) yields,
\[
\frac{\partial}{\partial \hat{h}(l)} \left( \frac{\|Y_f\|}{\|Y_f^*\|} \right) = 2\sqrt{r_0} \sum_{m=-P}^{P} \left( \frac{\hat{h}(l-m) + \hat{h}(l+m)}{\hat{r}_y(m)} \right). \tag{19}
\]

Based on the previous results and [10], a blind iterative algorithm can be derived which is summarized in the sequel. The non-increasing behavior of the I-divergence with each iteration was shown in [10]. The difference of two consecutive estimates converges to zero in the \(l_1\) norm [12]; these properties suggest that the proposed algorithm will result in high performance sparse channel estimation despite being blind.

Due to correlation function sparsity, we perform the channel update only for the relevant lags as shown in Equation (22) thus reducing complexity and the impact of noise. Considering a channel with a normalized unit power, one can easily verify the noise-dulling complexity and the impact of noise. The difference of two consecutive estimates converges to zero in the \(l_1\) norm [12]; these properties suggest that the proposed algorithm will result in high performance sparse channel estimation despite being blind.

The first step in our algorithm is to detect the most significant lags as quantified by the non-increasing behavior of the I-divergence with each iteration. The difference of two consecutive estimates converges to zero in the \(l_1\) norm [12]; these properties suggest that the proposed algorithm will result in high performance sparse channel estimation despite being blind.

1. Initialization:
   (1.1) Determine the SOS from the received data:
   \[
   \hat{r}_y(m) = \frac{1}{K} \sum_{n=0}^{N-1} y(n) y^*(n-m). \tag{20}
   \]
   (1.2) Compare \(\hat{r}(l)\) with \(\eta\), detect the MSLs \(q_i\), \(i = 0, 1, \ldots, W - 1\).
   (1.3) Give an initial channel estimate \(\hat{h}((0))(l) = \frac{1}{\sqrt{L}}\), \(\forall l\).

2. Loop: at the \(t\)-th iteration \(t \geq 0\),
   (2.1) Determine estimated channel SOS:
   \[
   \hat{r}_y(m) = \sum_{l=0}^{L-1} \hat{h}^{(t)}(l) \hat{h}^{(t)}(l-m). \tag{21}
   \]
   (2.2) Update channel:
   \[
   \hat{h}^{(t+1)}(l) = \hat{h}^{(t)}(l) \frac{2\sqrt{r_0}}{K} \sum_{i=-W+1}^{W-1} \left\{ \hat{h}^{(t)}(l-q_i) \hat{r}_y(q_i) \right\} + \hat{h}^{(t)}(l+q_i) \hat{r}_y(q_i). \tag{22}
   \]
   (2.3) Check stopping rule (e.g., \(I(\hat{r}_y) < \zeta\)); otherwise, increase \(t\) and go to (2.1).

The algorithm is inherently sparse as we only examine the significant lags as quantified by the \(q_i\); after the iteration ceases, the final estimate is further thresholded to limit the impact of small noisy estimates for the remaining lags. To ensure that taps are not missed, a loose thresholding is employed in order to determine the \(q_i\). It should be pointed out that a cyclic shift of the channel is an ambiguity that may result even in the noise-free case as cyclically shifted channels have the same correlation function.

**Blind algorithm versus training-based CS methods**

We compare the proposed purely blind part of the algorithm with the training-based compressed sensing methods, in which 128 pilot subcarriers are used. Herein, we correct for the cyclic shift ambiguity. The CS methods exploit the pilot information in order to determine the projection matrices. We consider a complex Rayleigh channel, modelled by a 6-tap FIR filter with a delay spread \(L = 100\) and the fixed MSTS 0, 8, 25, 58, 78, 99. The estimation performance is evaluated in terms of the MSE defined as

\[
\text{MSE} = \frac{1}{N_{MC}} \sum_{n=1}^{N_{MC}} \left\| \hat{h}_n - h_n \right\|^2, \tag{23}
\]

where \(N_{MC}\) is the number of Monte Carlo iterations. Note that in the proposed blind method, \(\hat{h}(l)\) is estimated directly, while in the CS methods, \(\hat{h}(l)\) is first estimated and then used to obtain \(\hat{h}_n(l)\). Figure 1 shows the MSE plots from 1000 Monte Carlo iterations for the proposed algorithm, the unconstrained least squares (LS) estimate, the OMP, the IDE, the HN and the IHT algorithms. The proposed method achieves a gain of 8.5 \(\sim\) 13 dB over the LS method. It is also seen that the performance of the proposed method is better than IDE and HN when the SNR is less than 18 dB, while it is slightly worse than the IHT method. Our numerical results indicate that the proposed blind algorithm can achieve a performance comparable to CS methods without exploiting the use of pilot information.

**MST detection for Rayleigh channel with random MST positions**

In this experiment, the MST detection for Rayleigh channel with ran-
dom MST positions is investigated. The number of MSTs is set to 6 for a delay spread \( L = 100 \). We take a semi-blind approach, thus we initialize with a LS channel estimate by using \( K_p = 64 \) pilot sub-carriers in the blind deautocorrelation and therefore the resulting MSTs are not subject to any position ambiguity. The MST detection performance for the LS with hard thresholding, the OMP and the IHT using \( K_p = 64 \) and \( K_p = 128 \) pilot subcarriers is also provided. Here, a normalized probability error of the MST detection is defined as \( P_e = \frac{N_{\text{dif}}}{D} \) where \( N_{\text{dif}} \) is the number of the misdetected significant taps in the estimated channel. As \( N_{\text{dif}} \) may be larger than the true sparsity \( D \), \( P_e \) might be larger than 1. Figure 3 show the results from 1500 Monte Carlo iterations. It is seen that the MST detection performance of the semi-blind MST detection algorithm significantly outperforms that of the LS, the OMP and the IHT using \( K_p = 64 \) pilot subcarriers. Also the semi-blind algorithm is superior to the LS and the OMP methods at a low or moderate SNR level.

The semi-blind channel estimation algorithm has been extended to a blind MST detection algorithm for a general overlap case. Simulation results have confirmed that the proposed semi-blind iterative algorithm as well as the MST detection algorithm can achieve very good performance by utilizing much less pilot overhead.

6. REFERENCES