

Sparse Channel Estimation with Zero Tap Detection

C. Carbonelli

Department of Information Engineering
University of Pisa
cecilia.carbonelli@iet.unipi.it

S. Vedantam, U. Mitra

Communications Science Institute
University of Southern California
{vedantam,ubli}@usc.edu

Abstract

Algorithms for the estimation of a channel whose impulse response is characterized by a large number of zero tap coefficients are developed and compared. Exploiting the sparsity of the channel, the estimation problem is transformed into an equivalent on-off keying (OOK) detection problem, whose solution gives an indication on the position of the zero taps. The proposed schemes are compared to the standard least squares estimate (LSE) via simulations in terms of mean square error (MSE) and bit error rate (BER). A scheme based on sphere decoding appears to give the best performance while maintaining moderate complexity.

1 Introduction

Channel estimation for the purposes of equalization is a long standing problem in communications and signal processing. In the current work, we focus on channel estimation, when the channel has structure. Specifically, we consider the estimation of sparse channels, that is, channels whose time domain response consists of a large number of zero taps. In the conventional least-squares approach all the taps of the impulse response can be nonzero and minimal *a priori* information about the unknown channel is required. On the other hand, when the impulse response is “long” the accuracy of this estimate is no longer satisfactory and the computational burden increases; thus we consider methods that exploit the fact that the channel is sparse.

Our investigation of this problem is primarily motivated by ultrawideband communications where the sparse channel model has utility from two perspectives. First, a widely accepted channel model [3] is based on a statistical model based on clusters which yields relatively sparse channel realizations. Second, for complexity reduction purposes, it is of interest to

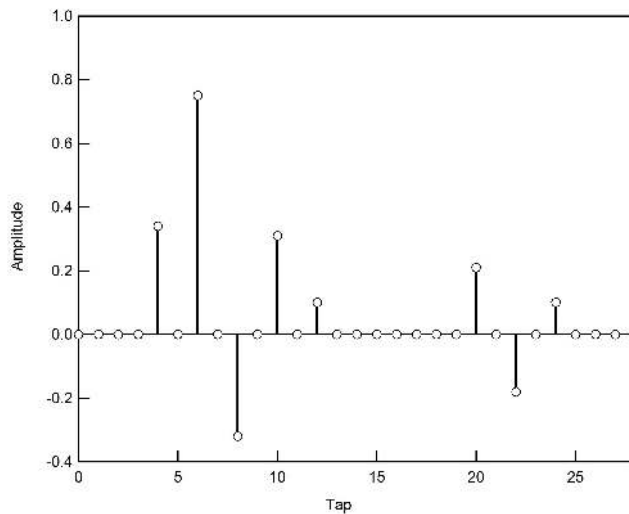


Figure 1: Time domain impulse response of a sparse channel.

estimate only dominant multipath fingers and thus impose a very sparse structure. Sparse channel estimation has gained some interest of late [1, 2] due to the fact that certain cellular [4, 5] and underwater channel models are best modeled as sparse. In [2], the properties of least-squares estimators for sparse channels are investigated. In particular, it is assumed in [2] that the number of desired non-zero taps is known *a priori*; for unknown model order, a penalty function is added to dissuade a large number of non-zero taps. In [1], the matching pursuit algorithm is used to estimate channel taps one by one by maximizing the correlation of a column of the mixture matrix with a residual signal. For scenarios where there might be large inter-path interference due to pulse shapes of long duration or the existence of dense clusters within a sparse channel, the matching pursuit algorithm will result in performance losses.

Our approach shares some features with these prior works, but does not presuppose a known channel order. We consider the channel estimation problem in two stages. First, with an initial estimate, we pose a detection problem by casting the determination of non-zero taps into an on-off-keying (OOK) communication problem. In the second stage, the channel estimate is refined once the locations of the non-zero taps have been determined. To this end, we consider three different schemes: the Viterbi algorithm, the sphere-decoding algorithm proposed in [6] and the multistage detector illustrated in [7] respectively.

The remainder of this paper is organized as follows: Section 2 describes the system model and the optimal OOK detector. Three strategies for the zero-tap detection problem are introduced in Section 3. A second estimate of the channel is then formed exploiting these results in Section 4. We compare the proposed schemes via simulation in Section 5 and draw conclusions in Section 6.

2 Signal Model

We assume a single-user system, where a pulse shaped symbol is transmitted through a channel modeled by a finite-impulse response filter. The received signal is sampled; the resulting signal in a symbol duration can be written in a matrix form as:

$$\mathbf{x} = \mathbf{\Theta}\mathbf{b} + \mathbf{n}, \quad (1)$$

where,

$$\mathbf{\Theta} = \mathbf{U}\text{diag}(\mathbf{h}), \quad (2)$$

and

$$\mathbf{U} = \begin{pmatrix} u[0] & 0 & 0 & \dots & 0 \\ u[1] & u[0] & 0 & \dots & 0 \\ u[2] & u[1] & u[0] & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & u[M-1] & \dots & u[0] \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & \dots & 0 & u[M-1] \end{pmatrix}, \quad (3)$$

$u[i]$, $i = 0, \dots, M-1$, being the sampled pulse shape at sampling instant i (M is the number of samples per transmitted pulse). Moreover, $\mathbf{h} = [h[0] \ h[1] \ \dots \ h[L-1]]^T$ are the taps of the discrete-time FIR filter (of length L), which models the channel impulse response (CIR): see Figure 1. As we consider sparse channels, the vector \mathbf{h} has non-zero values in the set of positions $\mathbf{P} = (p_1 \ p_2 \ \dots \ p_m)$, and \mathbf{b} is a sparse vector whose entries are obtained as:

$$b_i = \begin{cases} 1, & i \in \mathbf{P} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Notice that the weight of \mathbf{b} is m , the number of non-zero taps in the CIR. Finally \mathbf{n} is the vector of the noise samples, which is assumed to be additive, white Gaussian and thus $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Rather than estimate \mathbf{h} directly, we consider the detection of the vector \mathbf{b} which can be posed as a classic On Off Keying (OOK) detection problem (see *e.g.* [8]). To this end consider the following modified OOK detection problem with received signal:

$$x[i] = Ab_i + n[i], \quad b_i = \begin{cases} 0, & P(b_i = 0) = p \\ 1, & P(b_i = 1) = q, \end{cases} \quad (5)$$

where $n[i]$ is white Gaussian noise with standard deviation σ for each i . In this case, the optimum receiver is a matched filter followed by a threshold device, where the threshold is chosen in order to minimize the probability of error at the output of the correlator. More precisely, the total probability of error is:

$$p_\epsilon = pQ\left(\frac{\lambda}{\sigma}\right) + q\left(1 - Q\left(\frac{\lambda - A}{\sigma}\right)\right), \quad (6)$$

where the Q function is defined as:

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-\rho^2/2} d\rho, \quad (7)$$

and λ is the parameter we have to optimize. Deriving (6) with respect to λ and exploiting the properties of the Q function we get:

$$\begin{aligned} \frac{dp_\epsilon}{d\lambda} &= \frac{1}{\sqrt{2\pi}\sigma^2} \left[qe^{-\frac{(\lambda-A)^2}{2\sigma^2}} - pe^{-\frac{\lambda^2}{2\sigma^2}} \right] = 0 \\ \frac{p}{q} &= e^{\frac{A^2 - 2\lambda A}{2\sigma^2}}, \end{aligned}$$

and finally the optimum threshold value is:

$$\lambda_{opt} = \frac{A}{2} + \frac{\sigma^2}{A} \ln\left(\frac{p}{q}\right). \quad (8)$$

Notice that when $p = q = 1/2$, $\lambda_{opt} = A/2$.

3 Zero-Tap Detection

We now present three methods for detecting zero-taps in the channel impulse response. These schemes are typically used in detection problems and in the general case, their computational cost is much higher than that of the regular least squares (LS) approach. However, at least for the third approach, we expect better performance in the mean squared error sense.

3.1 Approximate Maximum Likelihood Detector

Let us reconsider our estimation/detection problem in light of the previous results. Assume that a first LS estimate of the CIR ($\hat{\mathbf{h}}$) has been obtained disregarding any structure on the nature of the channel as:

$$\hat{\mathbf{h}} = [\mathbf{U}^T \mathbf{U}]^{-1} \mathbf{U}^T \mathbf{x}. \quad (9)$$

The result in (9) will be referred to as the Least Squared Estimator (LSE).

When the probabilities are not known *a-priori* (they depend on the statistical nature of the channel) a possible approximation is to fix $p = q = 1/2$ and resort to an approximate maximum likelihood sequence detector in order to detect the vector \mathbf{b} , *i.e.*:

$$\hat{\mathbf{b}} = \arg \max_{\tilde{\mathbf{b}}} \{ \mathbf{p}(\mathbf{x} | \tilde{\mathbf{b}}, \hat{\mathbf{h}}) \} \quad (10)$$

where,

$$p(\mathbf{x} | \tilde{\mathbf{b}}, \hat{\mathbf{h}}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(\mathbf{x}-\mu)^T (\mathbf{x}-\mu)} \quad (11)$$

and

$$\begin{aligned} \mu &= \hat{\Theta} \tilde{\mathbf{b}} \\ \hat{\Theta} &= \mathbf{U} \text{diag}(\hat{\mathbf{h}}). \end{aligned}$$

After standard manipulations and neglecting all the terms that do not depend on $\tilde{\mathbf{b}}$, (10) can be expressed as:

$$\hat{\mathbf{b}} = \arg \max_{\tilde{\mathbf{b}}} \left\{ \sum_{i=0}^{L-1} \lambda_{\tilde{\mathbf{b}}}^{(i)} \right\}, \quad (12)$$

where,

$$\begin{aligned} \lambda_{\tilde{\mathbf{b}}}^{(i)} &= 2\tilde{b}_i \hat{h}[i] z(i) - |\tilde{b}_i|^2 \hat{h}[i]^2 \phi(i, i) - 2 \sum_{j=0}^{i-1} \tilde{b}_i \tilde{b}_j \hat{h}[i] \hat{h}[j] \phi(i, j) \\ z(i) &= \sum_{k=0}^{L-1} u(k-i) x(k) \\ \phi(i, j) &= \sum_{k=0}^{L-1} u(k-i) u(k-j). \end{aligned}$$

It can be shown that, due to the finite duration of the transmitted pulse, the problem in (12) can be solved using the Viterbi algorithm [8] with number of states 2^M . This result is referred to as (Approximate) Maximum Likelihood detector-Estimator (AMLE).

3.2 Iterative Multistage Detector

A second strategy is derived from the multistage detector proposed in [7] and referred in the following as Multistage Detector Estimator (MDE). The basic idea is to select a vector $\hat{\mathbf{b}}$ which minimizes

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b} \in \{0,1\}^L} \|\mathbf{x} - \hat{\Theta}\mathbf{b}\|^2. \quad (13)$$

Notice that the problem in (13) is equivalent to:

$$\begin{cases} \text{minimize} & f(\tilde{\mathbf{b}}) = \frac{1}{2}\tilde{\mathbf{b}}^T\mathbf{Q}\tilde{\mathbf{b}} + \mathbf{c}^T\tilde{\mathbf{b}}, \\ \text{subject to} & \tilde{\mathbf{b}} \in \{0,1\}^L. \end{cases} \quad (14)$$

where,

$$\begin{aligned} \mathbf{Q} &= \hat{\Theta}^T\hat{\Theta}, \\ \mathbf{c}^T &= -\mathbf{x}^T\hat{\Theta}. \end{aligned} \quad (15)$$

Observe that in the original application of the Multistage Detector [7], the columns of Θ corresponded to the spreading code of the different users in a multiuser direct-sequence code-division multiple-access system and $\mathbf{b} \in \{-1, 1\}^L$ contained the information bits of the users. As a consequence, a first transformation into an equivalent 0-1 problem has to be performed. Instead, for the signal model in (1), Θ is as described and \mathbf{b} is already a 0-1 vector. Nevertheless, this particular structure of Θ causes some conditioning issues as will be explained later, in the simulation results.

Once we have found the unconstrained solution $\hat{\mathbf{x}}$ to (14), we iterate towards the constrained solution $\hat{\mathbf{x}}^I$ as described in [7] observing that the continuous function minimum is upper bounded by the value taken by the function for any zero-one argument. In other terms, if the continuous minimum $f(\mathbf{x})$ gets closer to the zero-one minimum $f(\hat{\mathbf{x}}^I)$, the unconstrained solution $\hat{\mathbf{x}}$ will also approach to $\hat{\mathbf{x}}^I$. Note that, to guarantee convergence of this algorithm, we need \mathbf{Q} to be positive definite at every step of the iteration. In [9] it is shown that this result can be easily achieved searching for the index p -th of the current minimum which is furthest away from a zero-one value and then decreasing the p -th diagonal element of \mathbf{Q} by a quantity ϕ_p , which has been properly chosen so that the new function is still positive definite. The continuous minimum will also change by a quantity Δf that is always positive. The computational complexity of this scheme is linear with the number of taps to

be detected, apart from the computation of the quadratic function in (14) at each iteration stage.

Denoting \mathbf{g}_p the p -th column of \mathbf{Q}^{-1} , g_{pp} the p -th element along the diagonal of \mathbf{Q}^{-1} and q_{pp} the p -th element along the diagonal of \mathbf{Q} , the algorithm can be summarized as follows, if we start from the original guess \mathbf{x}_m and corresponding function value $f_m = f(\mathbf{x}_m)$:

1. For $i = 1, \dots, L$ let

$$t_i = \begin{cases} \hat{x}_i, & \text{if } \hat{x}_i \leq 0.5 \\ 1 - \hat{x}_i, & \text{if } \hat{x}_i > 0.5 \end{cases}$$

2. Find the index p such that $|t_p| = \max_{i=1, \dots, L} |t_i|$, *i.e.* \hat{x}_p is the component of the current minimum which is furthest away from a 0-1 value.

3. Compute

$$\phi_p = 2t_p/g_{pp}$$

$$q_{pp} = q_{pp} - \phi_p; c_p = c_p + \frac{1}{2}\phi_p$$

$$\tau = \phi_p/(\phi_p g_{pp} - 1)$$

$$\hat{\mathbf{x}} = \hat{\mathbf{x}} + \tau \mathbf{g}_p (0.5 - \hat{x}_p)$$

$$\hat{\mathbf{x}}^I = \frac{\text{sign}(\hat{\mathbf{x}} - 0.5) + 1}{2}$$

$$\mathbf{Q}^{-1} = \mathbf{Q}^{-1} - \tau \mathbf{g}_p \mathbf{g}_p^T$$

$$f(\hat{\mathbf{x}}^I) = \frac{1}{2}(\hat{\mathbf{x}}^I)^T \mathbf{Q} \hat{\mathbf{x}}^I + \mathbf{c}^T \hat{\mathbf{x}}^I$$

4. If $f(\hat{\mathbf{x}}^I) < f_m$, $f_m = f(\hat{\mathbf{x}}^I)$ and $\mathbf{x}_m = \hat{\mathbf{x}}^I$. Repeat from 1 until K -th stage, where K is independent of the number of taps in the channel.

3.3 Sphere Detection

We next consider the sphere decoding algorithm - referred in the following as Sphere Decoding Estimator (SDE). It was originally proposed in [6] and its complexity is discussed in [10]. A flowchart of the algorithm is available in [11]. SDE solves the problem:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b} \in \mathbf{D}_{\mathbf{CZ}^L}} \|\mathbf{x} - \hat{\mathbf{\Theta}} \mathbf{b}\|^2. \quad (16)$$

and decodes the received signal \mathbf{x} in (1) to the closest point on a lattice, generated by transforming the lattice \mathbf{D} through $\hat{\Theta}$. This is achieved via searching in a hypersphere of given radius. Computing this radius (ideally the covering radius of the lattice), unfortunately, is mostly experimental since determining the covering radius of an arbitrary lattice is NP-hard. As discussed in [10], the complexity of the SDE is polynomial for small enough noise variances and this is the major motivation for its consideration as a candidate scheme for our problem.

We now provide a description of the sphere decoding algorithm as applied to the problem considered. We consider *Hard Output* decoding here – that is the algorithm itself provides estimates of the vector \mathbf{b} . We first set \mathbf{D} to be $\{0,1\}^L$. Now, let r be the radius of the hypersphere to search over, $\hat{\mathbf{b}}$ be the original guess for the vector to be estimated, and $\mathbf{R} = [r_{i,j}]_{i,j=1}^L$ denote the Cholesky decomposition of $\hat{\Theta}$. Note that this decomposition always exists since the LS estimates can never be perfect. The following steps then provide an outline for the SDE. (This description follows along the lines of that in [10].)

1. Set $k = L$, $r'_L = \sqrt{r^2 - \|\mathbf{x}\|^2 + \|\hat{\Theta}\hat{\mathbf{b}}\|^2}$, $\hat{b}_{L|L+1} = \hat{b}_L$
2. Set $z = \frac{r'_k}{r_{k,k}}$, $U(b_k) = \lfloor z + \hat{b}_{k|k+1} \rfloor$, $b_k = -1$
3. $b_k = b_k + 1$. If $b_k \leq U(b_k)$ and $b_k \leq 1$ go to 5, else go to 4
4. $k = k + 1$, go to 3
5. If $k = 1$ go to 6. Else
 $k = k - 1$, $\hat{b}_{k|k+1} = \hat{b}_k + \text{sum}_{j=k+1}^L \frac{r_{k,j}}{r_{k,k}}(x_j - \hat{b}_j)$, $r'_k = \sqrt{r'^2_{k+1} - r^2_{k+1,k+1}(x_{k+1} - \hat{b}_{k+1|k+2})^2}$
6. Save b_k and go to 3

4 “Structured” Channel Estimates

Once the position of the zero taps has been detected, the next step is estimating the ‘active’ taps in the channel impulse response. To this end observe that a new estimate ($\hat{\mathbf{h}}^*$) of the channel response can be found solving the following LS problem:

$$\hat{\mathbf{h}}^* = \arg \min_{\tilde{\mathbf{h}}^*} \{ \|\mathbf{x} - \mathbf{U} \text{diag}(\mathbf{b}) \tilde{\mathbf{h}}^* \|^2 \}. \quad (17)$$

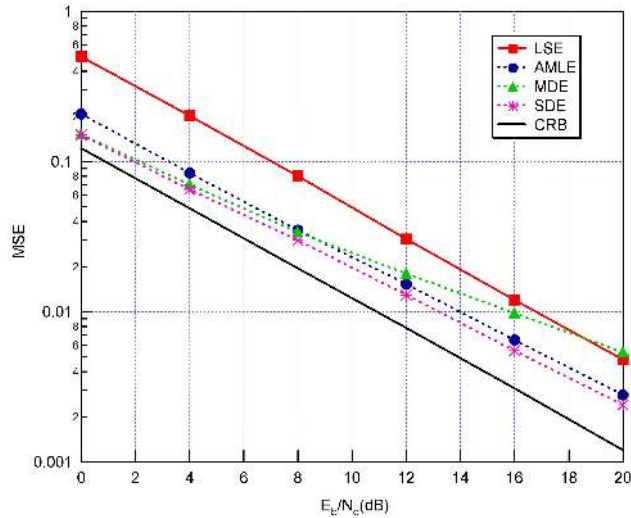


Figure 2: Residual MSE for the schemes proposed.

Calculating the gradient of the above expression with respect to $\tilde{\mathbf{h}}^*$ and setting it to zero, we get the following estimator for $\hat{\mathbf{h}}^*$:

$$\hat{\mathbf{h}}^* = \mathbf{Q}_b^\dagger \text{diag}(\hat{\mathbf{b}}) \mathbf{U}^T \mathbf{x}, \quad (18)$$

where \mathbf{Q}_b^\dagger denotes the pseudoinverse of

$$\mathbf{Q}_b = \text{diag}(\hat{\mathbf{b}}) \mathbf{U}^T \mathbf{U} \text{diag}(\hat{\mathbf{b}}) \quad (19)$$

and $\hat{\mathbf{b}}$ is the vector detected through (12), (13) or (16).

5 Results

In our simulation results, we consider a channel with $L = 30$ and $m = 8$. For the pulse shape we have chosen a root raised cosine function with roll-off factor 0.5 and M has been set to 6. The proposed estimation schemes are compared in terms of MSE and BER.

Figure 2 shows a plot of the residual Mean Squared Error (MSE) for the schemes as a function of the SNR of the transmitted signal and the corresponding Cramer-Rao bound (CRB) [12]. The MSE is determined as:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^N \|\mathbf{h} - \hat{\mathbf{h}}_n\|^2. \quad (20)$$

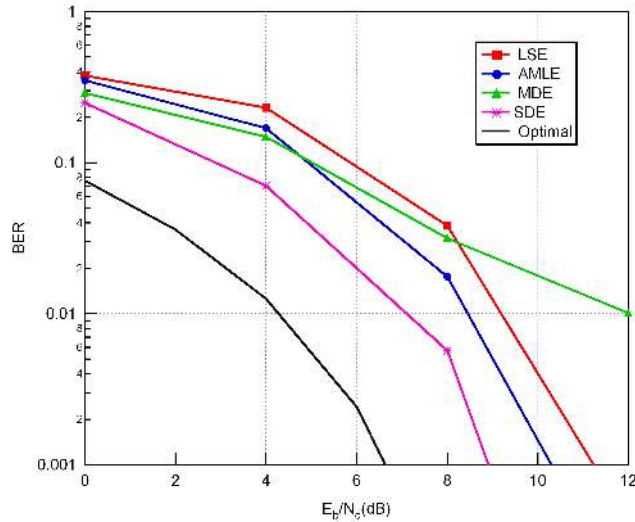


Figure 3: BER Performance of the schemes proposed.

where $N = 100$ is the number of Monte Carlo runs in the simulation. The true channel and the channel estimate for run n are denoted by \mathbf{h} and $\hat{\mathbf{h}}_n$ respectively. The superiority of the estimation strategies which exploit the sparsity of the channel is patent here at least for the SDE and the AMLE while the MDE outperforms the LSE only at low SNR. The following observation can explain this behavior. As the SNR increases, the estimation accuracy of the LS algorithm improves and several taps will be very small. As a consequence, the Hessian \mathbf{Q} is almost positive semi-definite and the MDE cannot converge to an integer solution, *i.e.* some taps of the continuous minimum from each iteration will start oscillating among a few points. In our work, the values provided in the figures for the MSE and the BER have been obtained by averaging over those realizations where the MDE converged to a 0-1 solution. A modified MDE is also proposed in [7] to improve the stability of the scheme.

Figure 3 compares the BER at the output of the correlation receiver for the LSE, AMLE, MDE, SDE and the optimal value provided by the single-user bound. It is seen that at high SNR's the SDE outperforms the LSE by around 2dB and the AMLE by approximately 1dB. The MDE is close to the SDE at low SNR's but suffers from the problem described above at high SNR's.

As seen from Figures 2 and 3, the sphere decoding algorithm outperforms all other schemes for the detection of the bit vector. This robustness stems from an efficient implementation

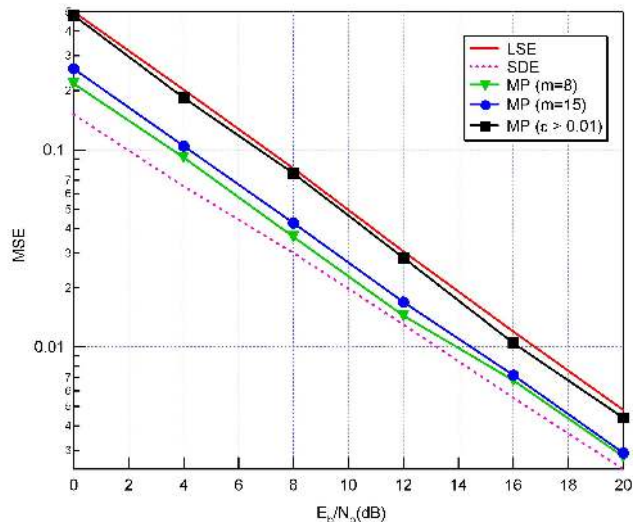


Figure 4: Comparison between SDE, MP and LSE.

of an algorithm to search for the global solution of a restricted minimization problem as described in the sections above. Thus, the SDE would provide the “best-possible” solution to the problem as long as the noise in the problem is spherically symmetric. This performance is at the cost of a comparatively larger number of computations (polynomial or exponential in the worst case) that need to be carried out to form the estimate with SDE as opposed to the MDE which provides its estimates in a linear number of computations.

One would expect the AMLE to provide the optimal solution to the problem considered. However, this algorithm is found to be sensitive to the probability of a tap being 0 as described by the derivation in Section (3.1). Moreover, the exponential complexity of this algorithm is a further deterrent to its utility as compared to the SDE.

Finally, in Figure 4 we contrast the SDE and the LSE with the Matching Pursuit (MP) proposed in [1]. Despite its simplicity, this iterative algorithm is extremely sensitive to the choice of the stopping rule. In fact, we can terminate the process when m^* non-zero taps have been detected, or, as an alternative, when the norm ϵ of the residual signal is smaller than a certain threshold T_h . For the simulation results in Figure 4 we have considered three different criteria: $m^* = m = 8$, $m^* = 15$ and $T_h = 0.01$. It is seen that even in the most optimistic situation, i.e. when the order of the channel is known, the SDE still outperforms the MP. On the other hand, stopping the iteration when the residual error meets a given

tolerance results in a large performance loss (note that the MP and the LSE are almost indistinguishable).

6 Conclusions

In this paper, we investigate the possibility of improving the channel estimates by formulating the estimation problem as an OOK detection problem. As future work, some other issues need to be taken care of. Using estimates of the *a-priori* probabilities of a tap being zero, the approximate maximum likelihood estimator also improves. This information could be obtained from the “soft” estimates provided by sphere decoding. Alternatively, it can be derived from the statistical properties of the channel model.

While the methods considered here assume that the channel estimates formed are “noise-free”, in reality, this is seldom the case. In such a scenario, we need to model the problem as a composite hypothesis testing problem of (potential) size 2^L . The estimation error in the channel is now seen in the form of a coloring of the noise term. Moreover, this colored noise has a variance which would change from hypothesis to hypothesis. A complete solution to this problem would yield the best possible estimates of the vector \mathbf{b} under “real-world” conditions.

References

- [1] S. F. Cotter and B. D. Rao, “Sparse channel estimation via matching pursuit with application to equalization”, *IEEE Trans. Commun.*, Vol.50, No.3, March 2002.
- [2] J. Homer, I. Mareels, R. R. Bitmead, B. Wahlberg and F. Gustafsson, “LMS estimation via structural detection”, *IEEE Trans. Signal. Processing*, Vol.46, No.10, October 1998.
- [3] R. J. -M Cramer, R. A. Scholtz and M. Z. Win, “Evaluation of an ultra-wide-band propagation Channel”, *IEEE Trans. Antennas Propagat.*, Vol.50, No.5, May 2002.
- [4] Cost 207, “Digital land mobile radio communications: final report”, *Office for Official Publications of the European Communities*, Luxembourg, 1989.
- [5] W. F. Schreiber, “Advanced television system for terrestrial broadcasting: some problems and some proposed solutions”, *Proc.IEEE*, vol.83, pp.958-981, June 1995.
- [6] U. Finke and M. Pohst, “Improved methods for calculating vectors of short length in a lattice including a complexity analysis”, *Math. Comput.* Vol 44, pp.463-471, Apr. 1985.

- [7] Z-L. Shi, W. Du, and P. F. Driessen, "A new multistage detector for synchronous communications", *IEEE Trans. Signal. Processing*, Vol.44, No.5, May 1996.
- [8] S.Biglieri and E.Benedetto, *Principles of Digital Transmission : With Wireless Applications*, Kluwer Academic / Plenum Publishers, New York, 1999.
- [9] M. W. Carter , "The indefinite zero-one quadratic problem", *Discrete Appl. Math.*, Vol. t, pp.23-44, 1984.
- [10] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding", in *Proc. Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol 2 , pp 1051-1055, 2001.
- [11] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels", *IEEE Trans. on Inf. Theory*, Vol. 45, pp.1639-1642, July 1999.
- [12] S. Kay, *Fundamental of Statistical Processing: Estimation Theory*, Prentice Hall PTR, 1st edition, January 27, 1998.