Estimation of Rapidly Time-Varying Sparse Channels

Weichang Li, *Member, IEEE*, and James C. Preisig, *Member, IEEE*

***Abstract*—The estimation of sparse shallow-water acoustic communication channels and the impact of estimation performance on the equalization of phase coherent communication signals are investigated. Given sufficiently wide transmission bandwidth, the impulse response of the shallow-water acoustic channel is often sparse as the multipath arrivals become resolvable. In the presence of significant surface waves, the multipath arrivals associated with surface scattering fluctuate rapidly over time, in the sense that the complex gain, the arrival time, and the Dopplers of each arrival all change dynamically. A sparse channel estimation technique is developed based on the *delay-Doppler-spread function* representation of the channel. The *delay-Doppler-spread function* may be considered as a first-order approximation to the rapidly time-varying channel in which each channel component is associated with Doppler shifts that are assumed constant over an averaging interval. The sparse structure of the *delay-Doppler-spread function* is then exploited by sequentially choosing the dominant components that minimize a least squares error. The advantage of this approach is that it captures both the channel structure as well as its dynamics without the need of explicit dynamic channel modeling. As the symbols are populated with the sample Dopplers, the increase in complexity depends on the channel Doppler spread and can be significant for a severely Doppler-spread channel. Comparison is made between nonsparse recursive least squares (RLS) channel estimation, sparse channel impulse response estimation, and estimation using the proposed approach. The results are demonstrated using experimental data. In training mode, the proposed approach shows a 3-dB reduction in signal prediction error. In decision-directed mode, it improves significantly the robustness of the performance of the channel-estimate-based equalizer against rapid channel fluctuations.**

***Index Terms*—Channel-estimate-based equalization, delay-Doppler-spread function, matching pursuit, sparse estimation, time-varying channels.**

I. Introduction

PHASE coherent demodulation of communication signals relies, explicitly or implicitly, on the accurate estimation of the impulse response of the communication channels. For broadband transmissions in shallow-water surface scattered acoustic channels, the principal challenges for channel estimation are the rapid channel fluctuations and the sparse structure of the channel impulse response. At sufficiently wide bandwidth, many shallow-water surface scattered channels can have sparse structures, formed by multipath arrivals that are resolved in delay [1]. This is reflected in that both the channel impulse response and the scattering function have an extended overall delay spread yet most of the channel energy is localized around several small regions (for example, see Figs. 1 and 2). Due to signal interactions with surface gravity waves, this sparse picture is also a highly dynamic one, with the delay, Doppler and complex gain associated with the surface arrivals changing rapidly over time [1], [2].

The combination of sparse structure and rapid fluctuations, together with the extended overall delay spread, creates a challenging channel estimation problem. However, the sparse structure of the channel impulse response can also be exploited to improve algorithm performance, due to the potential of reducing the number of taps to be tracked. As a result, the tracking algorithms can have a reduced computational complexity and memory requirement, and more importantly, smaller noise errors and an increased rate of channel fluctuations that can be tracked [3], [4].

Existing techniques for sparse system estimation in adaptive filtering and channel estimation can be categorized roughly into two types. Algorithms of the first type are approximation schemes that solve the nonlinear optimization problem of minimizing the squared residual prediction error as a function of the gain and the delay location of all the dominant taps. Among them are the sparse decision feedback equalizer (DFE) [5], the adaptive delay filter [6], the adaptive echo canceller [7], and the thresholding recursive least squares (RLS) algorithm [3], [4], [8], [9]. The common strategy of these algorithms is to break the original optimization problem over the whole gain/delay space into a sequence of optimization problems with smaller parameter spaces. In [5], a sparse DFE algorithm is derived by optimizing over the gains first and then finding the optimal delays. The adaptive delay filter [6] approximates the original problem by sequentially optimizing over the gain/delay of each individual tap. The adaptive echo canceller [7] and the threshold RLS are similar in that a full-tap adaptive filter is used first as an auxiliary filter to provide tap location and then the detected delay locations are transferred to a set of lower order filters to adapt the coefficients of those identified taps. The adaptive echo canceller uses a combination of various criteria to pick the dominant taps while the threshold RLS uses a simple energy criterion.

Algorithms of the second type, in the context of channel estimation, choose the most important *taps* of the sampled channel impulse response. This corresponds to finding the lowest dimensional linear combination of delayed versions of the transmitted symbol sequences to represent the received signal. Among the explicit sparse estimation techniques are the L_p-norm regularized method [10], [11], [12] and greedy algorithms such as the

Manuscript received September 1, 2006; revised April 27, 2007; accepted July 20, 2007. This work was supported by the U.S. Office of Naval Research under Grants N00014-05-10085 and N00014-06-1-0196.

Associate Editor: D. A. Abraham.

W. Li was with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: wli@whoi.edu).

J. C. Preisig is with the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA.

Digital Object Identifier 10.1109/JOE.2007.906409

0364-9059/$25.00 © 2007 IEEE

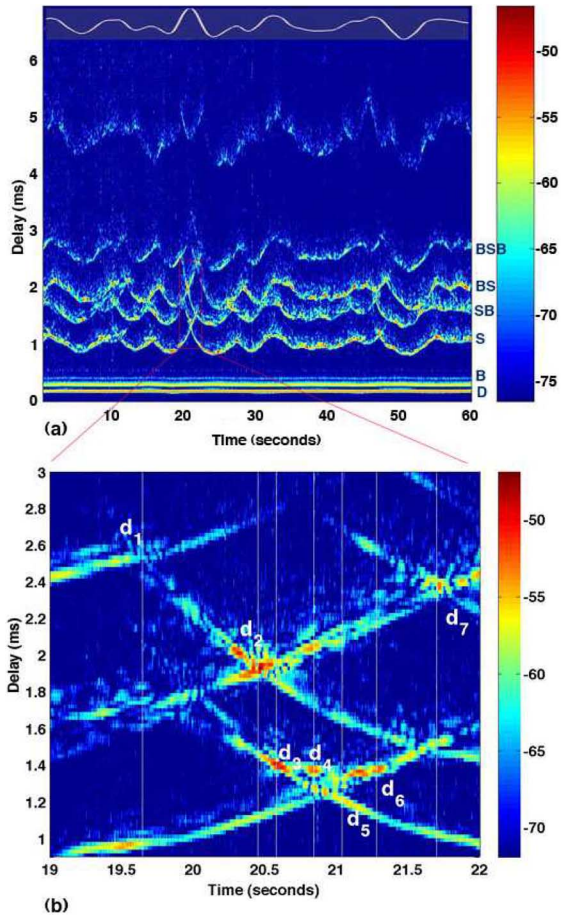


Fig. 1. Channel impulse response estimate. (a) Horizontal axis is time in second and the vertical axis is delay in milliseconds. On the very top of (a) is the smoothed surface waveheight time series measured at the specular point. The arrivals with one surface interaction are labeled as D (direct arrival), B (bottom scattered arrival), S (surface scattered arrival), SB (surface and then bottom), BS (bottom and then surface), and BSB (bottom-surface-bottom). (b) Enlarged plot of the delay-time region (0.9 ms, 3 ms, 19 s, 22 s) of (a). $d_1 \sim d_7$ labels dynamic events.

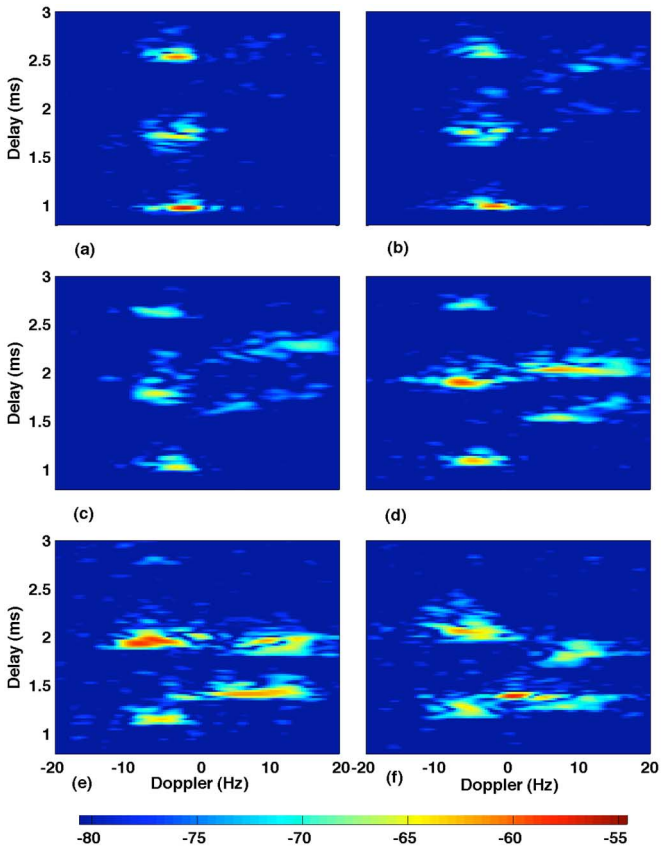


Fig. 2. Time-varying scattering function sequence I. Each plot shows the Doppler-delay region (−20 Hz, 20 Hz, 0.8 ms, 3 ms). Adjacent plots are approximately 273 ms (or 6552 symbols) apart. (a) 19.5617 s. (b) 19.8347 s. (c) 20.1077 s. (d) 20.3807 s. (e) 20.6537 s. (f) 20.9267 s.

matching pursuit (MP) algorithm [13] and its orthogonal versions (OMP) [14]. The MP algorithm and its orthogonal variants are computationally more efficient but remain analytically less tractable.[1] These methods originated from the signal representation literature where a dictionary subset is chosen iteratively to seek the most compact representation of the signal. Sparse channel estimation and equalization using the MP algorithm or its orthogonalized variant have recently increased noticeably [16], [17], [18].

The main limitation of these sparsing methods is that they require the sparse structure of the channel impulse response to remain stable over a certain time scale, which could be easily violated for the channels considered here.

A sparse channel estimation approach is developed in this paper, based on the authors' previously reported development [19]. It utilizes the channel *delay-Doppler-spread function* representation to accommodate time variations of the channel impulse response. The dominant components of the *delay-Doppler-spread function* are then identified sequentially to minimize the least squares (LS) signal prediction error. The advantage of the proposed approach is its capability of tracking the first-order channel dynamics. In the event of significant waves causing a highly dynamic channel, the resulting reduction in the signal prediction error is critical as otherwise the channel can be untrackable. An important point to note is that the *delay-Doppler-spread function* representation allows both the time variations and the sparse structure of the channel impulse response to be addressed simultaneously in a single approach.

Channel estimation and equalization approaches based on the *delay-Doppler-spread function* have previously been studied. In [20] and [21], the dominant delay-Doppler components are detected during initialization and periodic reinitializations via binary hypothesis test over each delay-Doppler location based on scattering function estimate. The channel tracker then runs on those detected components assuming that their locations stay fixed. In comparison, the approaches taken in this paper simultaneously track the locations of the dominant taps as well as their coefficients.

The remainder of this paper is organized as follows. Section II formulates the problem of channel estimation in terms of the channel impulse response as well as the *delay-Doppler-spread function*. Sparse estimation algorithms, specifically the basic MP algorithms, the OMP algorithm, and the order-recursive LS-MP algorithm, are presented in Section III. Section III

[1]The recent work by Tropp [15] has established some optimal conditions for the OMP algorithm, but the results are for exact-sparse cases only.

also discusses in detail the sparse estimation of the channel *delay-Doppler-spread function*. Section IV shows experimental results of channel estimation and equalization. Equalization results are obtained by using a residual prediction error decision feedback equalizer (RPE-DFE) [22]. Concluding remarks follow in Section V.

Throughout this paper, the following notations are used: superscripts *, ^{t}, and ^{h} denote complex conjugate, transpose, and Hermitian, respectively; $\mathbf{I}$ and $\mathbf{0}$ are identity matrices and zero matrices (or vectors) of appropriate size; $||\mathbf{a}||$ is the L_2-norm of the vector $\mathbf{a}$; and $||\mathbf{A}||$ denotes the induced norm of the matrix $\mathbf{A}$, i.e., the maximum singular value of $\mathbf{A}$.

II. Problem Formulation

A. Channel Impulse Response Estimation

In this paper, the sampled complex baseband representation will be used for all signals and channel responses. Here, the time-varying channel is represented by the *input delay-spread function*, denoted by $g(t,\tau)$, for which the sampled data input–output relationship is [23]

$$y[i] = \sum_{k=1}^{K} g^*[i,k]x[i-k+1] + v[i] \tag{1}$$

$$= \mathbf{g}^h[i]\mathbf{x}[i] + v[i] = \mathbf{x}^t[i]\mathbf{g}^*[i] + v[i] \tag{2}$$

where $y[i] \triangleq y(i\Delta t)$ is the received signal, $v[i]$ is the ambient noise, and $\mathbf{g}[i]$ is a column vector containing the sampled *input delay-spread function*, with its kth element given by $g[i,k] \triangleq g(i\Delta t, \tau_k)$. Here, $\tau_k \triangleq \tau_0 + (k-1)\Delta\tau$ for $k = 1,\ldots,K$ are the sampled delays. $\mathbf{x}[i]$ is the transmitted symbol vector whose kth element is the $(i-k+1)$th symbol. τ_0 is the reference delay. Δt and $\Delta\tau$ are the sample intervals in time and delay, respectively. K is the number of uniformly sampled delay taps, i.e., the channel dimension.

Stacking M successively received samples together and assuming that $\mathbf{g}[i]$ remains constant over these M samples yields the following:

$$\mathbf{y}[i] = \mathbf{C}[i]\mathbf{g}^*[i] + \mathbf{v}[i] \tag{3}$$

where

$$\mathbf{y}[i] \triangleq [\, y[i+M-1] \quad \ldots \quad y[i] \,]^t \tag{4}$$

$$\mathbf{v}[i] \triangleq [\, v[i+M-1] \quad \ldots \quad v[i] \,]^t \tag{5}$$

$$\mathbf{C}[i] \triangleq [\, \mathbf{x}[i+M-1] \quad \ldots \quad \mathbf{x}[i] \,]^t \triangleq [\, \mathbf{c}_1[i] \quad \mathbf{c}_2[i] \quad \ldots \quad \mathbf{c}_K[i] \,] \tag{6}$$

are of dimension $M \times 1$, $M \times 1$, and $M \times K$, respectively. Here, the columns of $\mathbf{C}[i]$ and $\mathbf{c}_k[i]$ are $M \times 1$ vectors; M is the length of the averaging window. For a sparse channel, let P denote the number of significant components in $\mathbf{g}_i$, that is, $K-P$ components are equal to zero or are negligible.

The basic assumption held by (3) is that

$$\mathbf{g}[i+m] = \mathbf{g}[i], \qquad \text{for } m = 1,\ldots,M-1. \tag{7}$$

The validity of this assumption depends on the length of the averaging window M as well as the rate of fluctuation of the channel taps.

For the purpose of channel estimation, it is assumed here that the symbol sequence matrix $\mathbf{C}[i]$ in (3) is known. For communication applications, this is the case either in the training mode where the transmitted symbols are known to the receiver, or in the decision-directed mode in which the previously detected symbols are considered known.

$\mathbf{g}[i]$ can be solved from (3) using a least squares (LS) method. The exponentially weighted recursive LS (EWRLS) algorithm is often applied to track slow variations of the channel

$$\widehat{\mathbf{g}}_{\text{ewrls}}[i] = \arg\min_{\mathbf{g}} \sum_{m=0}^{i} \lambda^{i-m} \left| y(m) - \mathbf{g}^h\mathbf{x}[m] \right|^2 \tag{8}$$

where λ is the forgetting factor. The effective averaging window length of the EWRLS algorithms is $M_\lambda = 1/(1-\lambda)$ [24] during which the channel is assumed constant. In general, the averaging window length should be proportional to the channel dimension (a rule of thumb value is 2–3 times the channel length) to maintain the stability of the algorithms. Consequently, a long averaging window is necessary for the channels of large dimension. Another effect of the averaging window length is that a longer averaging window results in a smaller noise contribution. However, with a long averaging window, the assumption (7) becomes rather restrictive or even unrealistic if the channel variation is significant during that period of time. Therefore, the choice of averaging window length is a tradeoff among the noise error contribution, the algorithm stability, and the tracking capability. Unfortunately, an acceptable choice is not always available in cases where the channel is overspread, i.e., has large delay-spread and also fluctuates rapidly.

B. Estimation of the Channel Delay-Doppler-Spread Function

Denote the sampled *delay-Doppler-spread function* by $u_{l,k} \triangleq u(\nu_l, \tau_k)$. Here, the sampled delays τ_k were defined earlier, and $\nu_l \triangleq \nu_0 + (l-1)\Delta\nu$ for $l = 1,\ldots,L$ are the sampled Dopplers with ν_0 and $\Delta\nu$ as the reference Doppler and the Doppler sample interval, respectively. $u_{l,k}$ is the Fourier transform of the sampled *input delay-spread function* [23], such that

$$g[i,k] = \sum_{l=1}^{L} u_{l,k} e^{j2\pi\nu_l i\Delta t}. \tag{9}$$

Substituting (9) into (1) yields

$$y[i] = \sum_{k=1}^{K}\sum_{l=1}^{L} u_{l,k}^* e^{-j2\pi\nu_l i\Delta t} x[i-k+1] + v[i] \tag{10}$$

$$= \mathbf{u}^h \left(\boldsymbol{\phi}[i] \otimes \mathbf{x}[i]\right) + v[i] \tag{11}$$

$$= \left(\boldsymbol{\phi}[i] \otimes \mathbf{x}[i]\right)^t \mathbf{u}^* + v[i] \tag{12}$$

where

$$\mathbf{u} \triangleq [\, u_{1,1} \quad u_{1,2} \quad \ldots \quad u_{1,K} \ldots \quad u_{L,K} \,]^t \tag{13}$$

$$\boldsymbol{\phi}[i] \triangleq [\, e^{-j2\pi\nu_1 i\Delta t} \quad e^{-j2\pi(\nu_1) i\Delta t} \quad \ldots \quad e^{-j2\pi(\nu_L) i\Delta t} \,]^t \tag{14}$$

are $KL \times 1$ and $L \times 1$ column vectors, respectively. $\otimes$ denotes the Kronecker product. Equation (10) essentially represents the received signal as the weighted sum of a sequence of delayed and Doppler shifted transmitted signals.

The *delay-Doppler-spread function* explicitly models the channel fluctuations in terms of the Fourier components of the time-variations of each tap and may be considered as a first-order approximation to the rapid channel fluctuations.

Again, stacking M continuously received signals and assuming that $\mathbf{u}$ remains constant over these M samples yields the following:

$$\mathbf{y}[i] = \mathcal{C}[i]\mathbf{u}^* + \mathbf{v}[i] \tag{15}$$

where $\mathbf{y}[i]$ and $\mathbf{v}[i]$ are defined in (4) and(5), respectively, and

$$\mathcal{C}[i] \triangleq [\,\mathbf{C}_{\nu_1}[i] \quad \ldots \quad \mathbf{C}_{\nu_L}[i]\,] \triangleq [\,\boldsymbol{c}_1[i] \quad \boldsymbol{c}_2[i] \quad \ldots \quad \boldsymbol{c}_{KL}[i]\,]. \tag{16}$$

Here, $\boldsymbol{c}_n[i]$ for $n = 1, \ldots, KL$, denoting the columns of $\mathcal{C}[i]$, are $M \times 1$ vectors. $\mathbf{C}_{\nu_l}[i]$ are the Doppler populated symbol matrices defined as follows. Denote the mth rows of $\mathbf{C}_{\nu_l}[i]$ as $\tilde{\mathbf{c}}_{\nu_l,m}[i]$, then

$$\tilde{\mathbf{c}}_{\nu_l,m}[i] \triangleq \mathbf{x}^t[i + M - m]e^{-j2\pi\nu_l(i+M-m)\Delta t} \tag{17}$$

for $l = 1, \ldots, L$ and $m = 1, \ldots, M$. Here again, M is the length of the averaging window, K is the channel delay dimension, and L is the Doppler dimension. For a sparse channel, let P denote the number of significant components in $\mathbf{u}$.

The basic assumption held by (15) is that $\mathbf{u}$ remains constant throughout the M sample periods. The validity of this assumption depends on the length of the averaging window M as well as the rate of fluctuation of the channel delay-Doppler components. Note that for channels with large yet slowly varying Dopplers, the assumption that the *delay-Doppler-spread function* remains constant is far less restrictive than assuming that the channel impulse response is constant over an equivalent period of time.

During channel estimation, the Doppler populated symbol sequence matrix $\mathcal{C}_i$ is assumed known. One might attempt to solve for $\mathbf{u}[i]$ directly from (15) using the LS method as one would solve for $\mathbf{g}[i]$ from (3). However, the resulting $\mathbf{u}[i]$ would have a large number of degrees of freedom. For the type of channels considered here, which are very sparse in the delay-Doppler domain, doing so will result in an overparameterization of the problem as the majority of the delay-Doppler components have little energy. As a result, the algorithm would become unstable and very sensitive to noise. While regularized LS algorithms such as ridge regression can be used to get less noise-sensitive solutions, they will inevitably introduce bias and artificial smoothing across channel coefficients. Also, the choice of the regularization parameter can be difficult to make. Furthermore, an extremely long averaging window, as required by the large dimensionality of $\mathbf{u}[i]$, would in turn restrict the maximum rate of fluctuation of $\mathbf{u}$ that can be tracked by the algorithm.

In Section III, sparse estimation techniques are presented which avoid these issues and improve the channel estimation performance. They reduce the number of channel components to be estimated, either in delay domain alone or on the delay-Doppler plane. As shown in [15], the matching pursuit (MP) algorithm can lead to the solutions close to those of the basis pursuit which essentially solves the L_p-norm regularized problem. The LS-MP algorithm developed in Section III can be viewed as trying to minimize the LS residual error, hence control the bias introduced by the regularization.

Once the channel *delay-Doppler-spread function* is estimated, an estimate of the channel impulse response can be obtained via (9).

C. Performance Metrics

The purpose of channel estimation is to provide an accurate estimate of the channel impulse response to the subsequent equalizer. The equalizer then detects the newly received symbols based on that channel estimate. An implicit assumption here is that the channel estimate computed based on symbols received previously can predict reasonably well the channel through which the new symbols to be detected are transmitted. Thus, it is essentially a prediction problem. For one-step prediction, the following performance metrics can be used.

1) Channel prediction error ϵ_g^2

$$\epsilon_g^2[i+1] \triangleq ||\mathbf{g}[i+1] - \widehat{\mathbf{g}}[i+1|i]||^2 \tag{18}$$

where $\widehat{\mathbf{g}}[i+1|i]$ is the prediction of the channel impulse response at time $i+1$ based on signals received up to time i. In general, $\widehat{\mathbf{g}}[i+1|i] = \widehat{\mathbf{g}}[i|i] \triangleq \widehat{\mathbf{g}}[i]$ for algorithms not based on any dynamic channel models, such as the LS type of algorithms. Since in reality the true channel impulse response $\mathbf{g}[i+1]$ is unknown, ϵ_g^2 is not a very useful metric in practice. Instead, the following *signal residual prediction error* is often used as a good surrogate for ϵ_g^2.

2) Signal residual prediction error ϵ_y^2

$$\begin{aligned}\epsilon_y^2[i+1] &\triangleq |y[i+1] - \widehat{y}[i+1|i]|^2 \\ &= \left|y[i+1] - \mathbf{x}^t[i+1]\widehat{\mathbf{g}}^*[i+1|i]\right|^2 \\ &= \left|\mathbf{x}^t[i+1]\left(\mathbf{g}^*[i+1] - \widehat{\mathbf{g}}^*[i+1|i]\right) + v[i+1]\right|^2\end{aligned} \tag{19}$$

where $\widehat{y}[i+1|i] = \mathbf{x}[i+1]^t\widehat{\mathbf{g}}^*[i+1|i]$ is the prediction of the signal received at time $i+1$ based on signals received up to time i. ϵ_y^2 is directly measurable and is used in the remainder of this paper to evaluate various channel estimation algorithms.

III. Sparse Channel Estimation

In this section, the basic MP algorithm, the orthogonal MP (OMP) algorithm, and an order-recursive LS-MP algorithm are introduced and applied to estimate the channel impulse response and the channel *delay-Doppler-spread function*. All three of these algorithms sequentially select "dominant" taps of the impulse response or the *delay-Doppler-spread function* and then calculate tap coefficients according to some criteria. Both the MP and OMP algorithms select the dominant channel tap that maximizes the projection of the residual vector onto the corresponding symbol vector. Then, the MP algorithm obtains the tap coefficient directly from that projection while OMP calculates a joint LS solution for the coefficients of all the selected taps. In contrast, the order-recursive LS-MP algorithm both selects

the dominant channel taps and calculates the tap coefficients to minimize the residual error achieved by the LS solution.

Consider the problem of solving for $\mathbf{x}$ from the following:

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{20}$$

where $\mathbf{C}$ is an $M \times K$ matrix and $\mathbf{y}$ and $\mathbf{x}$ are $M \times 1$ and $K \times 1$ vectors, respectively. Both $\mathbf{C}$ and $\mathbf{y}$ are assumed known. A solution $\widehat{\mathbf{x}}$ may be viewed as the coefficient vector associated with the representation of $\mathbf{y}$ in terms of the columns of $\mathbf{C}$. The sparsest solution $\widehat{\mathbf{x}}$ leads to the most compact representation of $\mathbf{y}$, in the sense that it has the least number of nonzero coefficients. The columns of $\mathbf{C}$ and $\mathbf{c}_j$, for $j = 1, \ldots, K$, are often called the dictionaries.

This general formulation (20) contains as special cases both the problem of channel impulse response estimation as specified by (3) and the channel *delay-Doppler-spread function* estimation as specified by (15), with the noise terms dropped.

A. MP Algorithm

The basic MP algorithm is an iterative procedure which can sequentially identify the dominant channel taps and estimate the associated tap coefficients. At each iteration, it selects one column of $\mathbf{C}$ that correlates best with the approximation residual from the previous iteration [13]. Here, the residual vector $\mathbf{r}_{p-1}$ is defined as $\mathbf{y}$ minus the contributions of all the columns identified in the previous $p-1$ iterations, with $\mathbf{r}_0 = \mathbf{y}$. More specifically, at the pth iteration, the column of $\mathbf{C}$ onto which the residual vector $\mathbf{r}_{p-1}$ has the maximal rank-one projection, denoted as $\mathbf{c}_{s_p}$, is found as

$$s_p = \arg \max_{j \notin I_{p-1}} \frac{\left|\mathbf{c}_j^h \mathbf{r}_{p-1}\right|^2}{\|\mathbf{c}_j\|^2} \tag{21}$$

where $I_{p-1} \triangleq \{s_1, \ldots, s_{p-1}\}$ is the index set of all previously selected columns. Then, $\widehat{x}_p$, the element of $\widehat{\mathbf{x}}$ associated with $\mathbf{c}_{s_p}$, is found as follows:

$$\widehat{x}_p = \frac{\mathbf{c}_{s_p}^h \mathbf{r}_{p-1}}{\|\mathbf{c}_{s_p}\|^2}. \tag{22}$$

The residual vector is computed as

$$\mathbf{r}_p = \mathbf{r}_{p-1} - \widehat{x}_p \mathbf{c}_{s_p} \tag{23}$$

with $\mathbf{r}_0 = \mathbf{y}$. Both (21) and (22) directly involve $b_{p,j} \triangleq \mathbf{c}_j^h \mathbf{r}_p$ for $j = 1, \ldots, K$ instead of $\mathbf{r}_p$ alone. Thus, there is no need to compute the residual vector $\mathbf{r}_p$ if $b_{p,j}$ are available. According to (22) and (23), $b_{p,j}$ can be recursively computed as

$$b_{p,j} = b_{p-1,j} - \frac{\mathbf{c}_j^h \mathbf{c}_{s_p}}{\|\mathbf{c}_{s_p}\|^2} b_{p-1,s_p} = b_{p-1,j} - \widehat{x}_p \mathbf{c}_j^h \mathbf{c}_{s_p}. \tag{24}$$

The MP algorithm is summarized in Table I.

B. Orthogonal Matching Pursuit

When the set of columns chosen according to (21) are not orthogonal, the set of coefficients $\widehat{x}_i$ obtained in (22) may not give the minimal residual error. This is because although $\mathbf{r}_p$ is orthogonal to $\mathbf{c}_{s_p}$, it is not necessarily orthogonal to $\mathbf{c}_{s_j}$ for $j < p$, which means that with $\widehat{x}_p$ calculated according to (22), the residual vector $\mathbf{r}_p$ is not orthogonal to the space spanned by $\mathbf{c}_{s_j}$ for $j \leq p$ and it is not of minimal length. The OMP algorithm, which addresses this issue, was proposed previously in [14]. At each step after a new column has been chosen according to (21), the OMP algorithm recomputes the coefficients as follows:

$$\widehat{\mathbf{x}}_{p,\text{omp}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{C}_{s,p}\mathbf{x}\|^2 \tag{25}$$

$$= \left[\mathbf{C}_{s,p}^h \mathbf{C}_{s,p}\right]^{-1} \mathbf{C}_{s,p}^h \mathbf{y} \tag{26}$$

where $\mathbf{C}_{s,p} \triangleq [\,\mathbf{c}_{s_1} \quad \ldots \quad \mathbf{c}_{s_p}\,]$. Equation (26) essentially carries out the LS minimization based on the set of columns chosen at each iteration. However, the rule of selecting the new column is the same as (21).

TABLE I
BASIC MP ALGORITHM

initialization	(27)
$\mathbf{r}_0 = \mathbf{y}$	(28)
$b_{0,j} = \mathbf{c}_j^h \mathbf{r}_0, \quad \text{for } j = 1, \cdots, K$	(29)
$s_1 = \arg \max_{j=1,\cdots,K} \frac{\lvert b_{0,j}\rvert^2}{\|\mathbf{c}_j\|^2}$	(30)
$I_1 = \{s_1\}$	(31)
$\widehat{x}_1 = \frac{b_{0,s_1}}{\|\mathbf{c}_{s_1}\|^2}$	(32)
$b_{1,j} = b_{0,j} - \widehat{x}_1 \mathbf{c}_j^h \mathbf{c}_{s_1}, \quad \text{for } j = 1, \cdots, K, j \notin I_1$	(33)
the pth iteration, $p > 1$	(34)
$s_p = \arg \max_{j=1,\cdots,K, j \notin I_{p-1}} \frac{\lvert b_{p-1,j}\rvert^2}{\|\mathbf{c}_j\|^2}$	(35)
$I_p = \{I_{p-1}\ ,\ s_p\}$	(36)
$\widehat{x}_p = \frac{b_{p-1,s_p}}{\|\mathbf{c}_{s_p}\|^2}$	(37)
$b_{p,j} = b_{p-1,j} - \widehat{x}_p \mathbf{c}_j^h \mathbf{c}_{s_p}, \quad \text{for } j = 1, \cdots, K, j \notin I_p$	(38)

C. Order-Recursive LS-MP Algorithm

In general, the set of columns selected by the MP algorithm does not necessarily lead to a representation of $\mathbf{y}$ that has the smallest LS error. The order-recursive LS-MP algorithm developed in [1] uses an LS error criterion as the rule of column selection at each iteration. Rather than picking the new column onto which the rank-one projection of the residual vector is maximized as both the MP and the OMP algorithm do, it seeks the new column which, together with all the previously selected $(p-1)$ columns, represents the received signal with the minimum squared residual error. As a result, the selected p columns span a subspace onto which the projection of the received signal is maximized. Here, p is the index of the current iteration.

More specifically, assume that after the $p-1$st iteration, the set of identified columns is $\mathbf{C}_{s,p-1} \triangleq [\,\mathbf{c}_{s_1} \quad \mathbf{c}_{s_2} \quad \ldots \quad \mathbf{c}_{s_{p-1}}]$; the associated coefficient vector $\widehat{\mathbf{x}}_{p-1}$ is obtained using the LS method, i.e.,

$$\widehat{\mathbf{x}}_{p-1} = \mathbf{R}_{p-1}^{-1} \mathbf{z}_{p-1} \tag{39}$$

where

$$\mathbf{R}_{p-1} = \mathbf{C}_{s,p-1}^{h}\mathbf{C}_{s,p-1} \tag{40}$$

$$\mathbf{z}_{p-1} = \mathbf{C}_{s,p-1}^{h}\mathbf{y} \tag{41}$$

and $\mathbf{R}_{p-1}$ is assumed invertible.

The associated residual vector $\mathbf{r}_{p-1}$ and the squared residual error $\|\mathbf{r}_{p-1}\|^2$ can be calculated as

$$\mathbf{r}_{p-1} = \mathbf{y} - \mathbf{C}_{s,p-1}\widehat{\mathbf{x}}_{p-1} \tag{42}$$

$$\|\mathbf{r}_{p-1}\|^2 = \|\mathbf{y}\|^2 - \mathbf{z}_{p-1}^{h}\mathbf{R}_{p-1}^{-1}\mathbf{z}_{p-1}. \tag{43}$$

The second equality follows from the orthogonality principle.

At the pth iteration, the algorithm finds a new column out of the set of remaining columns, denoted by $\mathbf{c}_{s_p}$, which gives the minimum squared residual error, that is

$$s_p = \arg\min_{j\notin I_{p-1}} \|\mathbf{r}_{p,j}\|^2 \tag{44}$$

where

$$\mathbf{r}_{p,j} = \mathbf{y} - [\,\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\,]\,\mathbf{x}_{p,j} \tag{45}$$

and $\mathbf{x}_{p,j}$ contains the associated LS coefficients given by

$$\mathbf{x}_{p,j} = \mathbf{R}_{p,j}^{-1}\mathbf{z}_{p,j} \tag{46}$$

where

$$\mathbf{R}_{p,j} = [\,\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\,]^{h}\,[\,\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\,] \tag{47}$$

$$\mathbf{z}_{p,j} = [\,\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\,]^{h}\,\mathbf{y}. \tag{48}$$

It is shown in the Appendix that the column selected according to (44), denoted as $\mathbf{c}_{s_p}$, is the solution to the following optimization problem:

$$s_p = \arg\max_{j\notin I_{p-1}} \frac{\left|\mathbf{c}_j^{h}\mathbf{r}_{p-1}\right|^2}{\delta_{p-1,j}} \tag{49}$$

where $\delta_{p-1,j} \triangleq \mathbf{c}_j^{h}\left[\mathbf{I} - \mathbf{C}_{s,p-1}\mathbf{R}_{p-1}^{-1}\mathbf{C}_{s,p-1}^{h}\right]\mathbf{c}_j$ [see (78) in the Appendix] has the geometrical meaning as the LS distance between the column $\mathbf{c}_j$ and the subspace spanned by $\mathbf{C}_{s,p-1}$. $\mathbf{r}_{p,j}$ in (45), $\mathbf{x}_{p,j}$ in (46), and $\delta_{p-1,j}$ in (49) can all be computed in an order-recursive manner, as shown in the Appendix.

Equations (49) and (21) differ only in the denominator. They become identical if all columns are orthogonal such that $\mathbf{C}_{s,p-1}^{h}\mathbf{c}_j = \mathbf{0}$ and $\delta_{p-1,j} = \|\mathbf{c}_j\|^2$. In (21), the cross correlation between the residual vector and a column is normalized by the L_2 norm of the column while in (49) the cross correlation is normalized by the squared distance between the new column and the previous column set $\mathbf{C}_{s,p-1}$. This change of normalization in (49) may be viewed as an orthogonalization step at each iteration. That is, if a new column is associated with some dominant tap but is closely affine to some columns in $\mathbf{C}_{s,p-1}$, it may still be selected by the new algorithm due to its smaller distance to the space spanned by $\mathbf{C}_{s,p-1}$ (yet not equal to zero), even though it may not be selected by the MP or OMP algorithm.

TABLE II
ORDER-RECURSIVE LS-MP ALGORITHM

initialization	(50)
$\mathbf{r}_0 = \mathbf{y}$	(51)
$b_{0,j} = \mathbf{c}_j^{h}\mathbf{r}_0, \quad \text{for } j = 1,\cdots,K$	(52)
$\delta_{0,j} = \|\mathbf{c}_j\|^2, \quad \text{for } j = 1,\cdots,K$	(53)
$s_1 = \arg\max_{j=1,\cdots,K} \frac{\lvert b_{0,j}\rvert^2}{\delta_{0,j}}$	(54)
$\mathbf{C}_{s,1} = \mathbf{c}_{s_1}, \quad I_1 = \{s_1\}$	(55)
$R_1^{-1} = \|\mathbf{c}_{s_1}\|^{-2}$	(56)
$\widehat{x}_1 = R_1^{-1} b_{0,s_1}$	(57)
$b_{1,j} = b_{0,j} - \mathbf{c}_j^{h}\mathbf{c}_{s_1}\widehat{x}_1, \quad \text{for } j = 1,\cdots,K,\ j\notin I_1$	(58)
$\mathbf{Q}_1 = \mathbf{I} - \mathbf{C}_{s,1}R_1^{-1}\mathbf{C}_{s,1}^{h}$	(59)
$\delta_{1,j} = \mathbf{c}_j^{h}\mathbf{Q}_1\mathbf{c}_j, \quad \text{for } j = 1,\cdots,K,\ j\notin I_1$	(60)
the pth iteration, $p > 1$*:*	(61)
$s_p = \arg\max_{j=1,\cdots,K,\ j\notin I_{p-1}} \frac{\lvert b_{p-1,j}\rvert^2}{\delta_{p-1,j}}$	(62)
$\mathbf{C}_{s,p} = \left[\mathbf{C}_{s,p-1}, \mathbf{c}_{s_p}\right], \quad I_p = \{I_{p-1}, s_p\}$	(63)
$\mathbf{d}_{p-1} = \begin{bmatrix} -\mathbf{R}_{p-1}^{-1}\mathbf{C}_{s,p-1}^{h}\mathbf{c}_{s_p} \\ 1 \end{bmatrix}$	(64)
$\mathbf{R}_p^{-1} = \begin{bmatrix} \mathbf{R}_{p-1}^{-1} & \mathbf{0} \\ \mathbf{0}^{t} & 0 \end{bmatrix} + \delta_{p-1,s_p}^{-1}\mathbf{d}_{p-1}\mathbf{d}_{p-1}^{h}$	(65)
$\widehat{x}_p = \delta_{p-1,s_p}^{-1} b_{p-1,s_p}$	(66)
$\widehat{\mathbf{x}}_p = \begin{bmatrix} \widehat{\mathbf{x}}_{p-1} \\ 0 \end{bmatrix} + \mathbf{d}_{p-1}\widehat{x}_p$	(67)
$b_{p,j} = b_{0,j} - \mathbf{c}_j^{h}\mathbf{C}_{s,p}\widehat{x}_p, \quad \text{for } j = 1,\cdots,K,\ j\notin I_p$	(68)
$\delta_{p,j} = \delta_{p-1,j} - \frac{\left\lvert\mathbf{c}_j^{h}\mathbf{C}_{s,p}\mathbf{d}_{p-1}\right\rvert^2}{\delta_{p-1,s_p}}, \quad \text{for } j = 1,\cdots,K,\ j\notin I_p$	(69)

Geometrically, the order-recursive LS-MP algorithm approximates the original NP-hard[2] optimization problem by a set of successively higher dimension LS minimization problems starting from a 1-D problem. In the same light, the MP algorithm is essentially an approximation to the original problem via a set of successive 1-D projections. The differences between the two, reflected not only in the coefficient calculation, but also in column selection, depend largely on the orthogonality of the columns of the symbol sequence matrix. However, because a globally optimal p dimension subspace is not necessarily a simple augmentation of the optimal subspace of dimension $p-1$ by a newly selected column vector, the order-recursive LS-MP algorithm is, in general, not globally optimal.

The order-recursive LS-MP algorithm is summarized in Table II.

D. Sparse Channel Impulse Response Estimation

The sparse estimate of the channel impulse response at time i is obtained by applying either the basic MP algorithm, the OMP algorithm, or the order-recursive LS-MP algorithm to the system of equations specified by (3), which hereafter will be referred to as MP-IR, OMP-IR, and LSMP-IR, respectively. In general, the sampling intervals in both time and delay are

[2]NP is a complexity term and stands for nondeterministic polynomial time. NP-hard generally refers to problems for which there is no polynomial-time algorithms.

known. The set of parameters for the algorithms that need to be specified in this case include the length of channel impulse response K, the averaging window length M, and the number of dominant taps P.

The channel length K is generally chosen such that the most significant part of the delay spread is covered. It affects the size of the candidate column set from which those associated with energetic taps will be identified. A channel length K corresponds to columnwise inner-products of the order K^2M. For a given P, the effect of K becomes mainly computational as long as it covers the overall channel delay spread.

Several factors are relevant in choosing the averaging window length M, including the fluctuating rate of the channel impulse response, the channel dimension, the number of dominant taps, the signal-to-noise ratio (SNR), and the coherence structure of the resulting symbol sequence matrix. First, the assumption (7) requires that the channel does not change significantly over M samples. For instance, considering a single tap with Doppler $\nu = 5$ Hz and assuming the sample frequency $f_s = 24\,000$ Hz. The phase of this tap will change by $\pi/2$ over an averaging window of length $M = 1200$. Thus, estimation based on a longer averaging window length would introduce significant error. Second, the choice of M should be proportional to P, the number of dominant taps. In the order-recursive LS-MP, it is necessary to have $M > P$, otherwise the set of selected columns becomes linearly dependent and the problem is underdetermined with nonunique solutions. In the presence of moderate-to-large noise level, a longer averaging window is able to suppress the noise effects assuming that the noises are independent. Finally, a larger M, in general, will lead to a symbol matrix with better orthogonality among its columns, a property desirable for sparse processing algorithms [15].

The choice of the number of dominant taps P is discussed in Section III-F.

Note that when the symbol sequences are nearly orthogonal, as in the case where M-sequence is used with a sufficiently long averaging window, then the basic MP algorithm becomes nearly as good as the order-recursive LS-MP algorithm.

E. Sparse Estimation of Channel Delay-Doppler-Spread Function

Applying the MP, the OMP, or the order-recursive LS-MP algorithms to (15) generates sparse estimates of the channel *delay-Doppler-spread function*, which hereafter will be referred to as MP-SF, OMP-SF, and LSMP-SF, respectively. As was the case with the channel impulse response estimation, the averaging window length M and the length of the channel impulse response K need to be specified. Additionally, the Doppler range and the number of Doppler samples L also need to be specified. The dimension of the *delay-Doppler-spread function* is now KL, a large number in general.

The Doppler range is chosen to cover the most significant part of the channel Doppler spread. Once the Doppler range has been chosen, the number of Doppler samples L should be selected such that the chosen averaging time window length can resolve the Doppler sample interval.

To be more specific, consider $\boldsymbol{c}_m[i]$ and $\boldsymbol{c}_n[i]$, two columns of the Doppler-populated symbol matrix $\mathcal{C}[i]$ obtained by populating the column $\mathbf{c}_l[i]$ of the symbol matrix with two different Dopplers ν_m and ν_n, respectively

$$\boldsymbol{c}_{m,k}[i] = \mathbf{c}_{l,k}[i] e^{j2\pi\nu_m(i+M-k)\Delta t} \tag{70}$$

$$\boldsymbol{c}_{n,k}[i] = \mathbf{c}_{l,k}[i] e^{j2\pi\nu_n(i+M-k)\Delta t} \tag{71}$$

for $k = 1, \ldots, M$. Here, k is the element index in $\boldsymbol{c}_m[i]$, $\boldsymbol{c}_n[i]$, and $\mathbf{c}_l[i]$. For simplicity, assume $|\mathbf{c}_{l,k}[i]| = 1$. It then follows that

$$\alpha_{m,n} \triangleq \frac{\left|\boldsymbol{c}_m^h[i]\boldsymbol{c}_n[i]\right|}{||\boldsymbol{c}_m[i]||\,||\boldsymbol{c}_n[i]||} = \left|\frac{\sin(M\pi\Delta\nu_{m,n}\Delta t)}{M\sin(\pi\Delta\nu_{m,n}\Delta t)}\right| \tag{72}$$

and $0 \le \alpha_{m,n} \le 1$. $\alpha_{m,n} \approx 1$ indicates that $\boldsymbol{c}_m[i]$ and $\boldsymbol{c}_n[i]$ are nearly affine and cannot be easily distinguished from the MP point of view. Thus, to resolve $\boldsymbol{c}_m[i]$ from $\boldsymbol{c}_n[i]$, it is necessary to keep $\alpha_{m,n}$ small. For large M, it can be shown [25, p. 47] that $\alpha_{m,n} \le 1/\sqrt{2}$ when

$$\Delta\nu_{m,n} > \frac{1.4}{\pi M\Delta t} \tag{73}$$

which can be considered as a lower bound of $\Delta\nu$. For $M = 800$ and a sample frequency $f_s = 24\,000$ Hz, this corresponds to $\Delta\nu > 13.37$ Hz.

F. Stopping Criteria

Another parameter critical to the performance of these sparse algorithms is the number of dominant taps P, which determines when to stop the iterations in these algorithms. Qualitatively speaking, the choice of P is a tradeoff between the variance of the channel estimate and the errors caused by insufficient parameterization. It also has an effect on the tracking capability of the algorithm. A small P leads to lower variance and fast tracking, but would introduce errors in the channel estimate by neglecting certain taps. There are several theoretical criteria that one can resort to in order to find various optimal choices of P, such as the minimum description length (MDL) and the Akaike information criterion (AIC). Without delving into these optimal criteria and their applicability to this case, the emphasis here is to show that the response of various residual errors to the choice of P can be quite different and the key is to seek P that minimizes the prediction error defined in (19).

In Fig. 5(a), the dependency of two types of errors on P are shown, including the residual vector $\mathbf{r}[n] \triangleq \mathbf{y}[n] - \mathcal{C}[n]\widehat{\mathbf{u}}^*[n]$ and the signal prediction error $\epsilon_y[n] \triangleq y[n+1] - \tilde{\mathbf{c}}_1[n+1]\widehat{\mathbf{u}}^*[n]$. These curves are obtained for *delay-Doppler-spread function* estimation using the order-recursive LS-MP algorithm. As shown, the magnitude squares of $\mathbf{r}[n]$ decrease monotonically as P increases, although the amount of reduction diminishes as P increases further. The signal prediction error, however, decreases initially as P increases, but then rises as P increases passing 200. This makes intuitive sense. Initially, the channel is undermodeled when P is small. Therefore, increasing P can reduce the associated undermodeling error. As P increases and the channel becomes overparameterized, noise contribution as

TABLE III
OPERATION COUNTS IN (35)–(38)

(35)	$2[K-(p-1)]$
(37)	1
(38)	$2(K-p)$

TABLE IV
OPERATION COUNTS IN (62)–(69)

(62)	$2[K-(p-1)]$
(64)	$(p-1)^2$
(65)	p^2+p
(66)	1
(67)	p
(68)	$(K-p)(p+2)$
(69)	$(K-p)p$

well as tracking error rise with P. As discussed previously, channel estimation for the purpose of equalization is really a prediction problem, hence the choice of P should be made to minimize the signal prediction error. This is an important difference from the problem of sparse representation.

However, establishing the direct connection between the prediction error to the value of P needs statistical modeling of the channel dynamics and is still an ongoing work. An *ad hoc* approach is to terminate the MP process when either the residual error decreases below certain level which may depend on various system parameters such as input SNR levels, or when the amount of reduction in the residual error between consecutive iterations falls below certain threshold.

G. Complexity and Performance Tradeoff

Another important issue is the complexity of the channel algorithms. In both the basic MP and the order-recursive LS-MP algorithms, the computation mainly consists of the following two parts: 1) the calculation of the inner products between column pairs of $\mathbf{C}$ and those between columns of $\mathbf{C}$ and $\mathbf{y}$ and 2) the iterations in which the algorithm identifies the dominant channel components and estimates their coefficients.

These inner products are precalculated before the iterations. In both algorithms, the computation of this part is of the order of $\mathcal{O}(MK+MK^2)$ based on a block of received signal of length M. When data arrive sequentially and the transmitted sequence is shifted symbol by symbol, the computation is reduced to $\mathcal{O}(2K+2K^2)$ since both $\mathbf{C}^h\mathbf{C}$ and $\mathbf{C}^h\mathbf{y}$ can be updated recursively.

The amount of computations involved in the iterations can be significantly different. In the basic MP algorithm, the operation counts for (35)–(38) at the pth iteration are listed in Table III. For P iterations, the total operation counts are of the order of $\mathcal{O}\left(4KP-2P^2+P\right)$. For the order-recursive LS-MP algorithm, the operation counts associated with (62)–(69) at the pth iteration are listed in Table IV. Using series summation over $p=1,\ldots,P$, it follows that for P iterations the total operation counts are of the order of $\mathcal{O}\left(4KP-2P^2+2P+KP^2+KP\right)$, with an increase over that of the MP algorithm of the order of $\mathcal{O}\left(KP^2+KP+P\right)$.

The overall computation for sequential data processing, including the calculation of inner products and the iterations, is of the order of $\mathcal{O}(2K^2+2K+4KP-2P^2+P)$ for the basic MP algorithm and $\mathcal{O}(2K^2+2K+4KP-2P^2+2P+KP^2+KP)$ for the order-recursive LS-MP algorithm. For example, with 116 channel delay taps, nine Doppler samples (so that $K=116\times 9=1044$), and the number of dominant components $P=100$, the LS-MP *delay-Doppler-spread function* estimation increases the computation approximately by a factor of four over the MP *delay-Doppler-spread function* estimation and approximately 306 over the basic MP channel impulse response estimation that uses $K=116$ and a reduced $P=40$.

Note that during sequential data processing, it is possible to reduce the overall complexity of the *delay-Doppler-spread function* estimation by reducing the frequency at which the dominant delay-Doppler component locations are updated, assuming that the locations change slower than the coefficients. For example, instead of running the LS-MP symbol-by-symbol, one could run LS-MP block-by-block while the coefficients are updated using standard RLS algorithm between blocks.

IV. EXPERIMENTAL RESULTS

To demonstrate the performance of various channel estimation algorithms presented in Section III as well as their effects on equalization, data from a surf-zone experiment has been analyzed and the results based on two receiver operating scenarios are presented in this section: one based on pure training mode and the other on decision-directed mode with periodic training. In both scenarios, channel impulse response estimates are obtained using both the EWRLS and the MP-IR algorithms. Channel *delay-Doppler-spread function* estimates are obtained via the MP-SF and the LSMP-SF algorithms. Equalization performance is evaluated using the RPE-DFE [22], based on the channel impulse response estimates or the impulse response converted from the *delay-Doppler-spread function* estimates. In the training mode, the transmitted symbols are assumed perfectly known to both the channel estimator and the DFE feedback filter. In the decision-directed mode, only the symbol estimates are available. A periodic training strategy is adopted in the decision-directed mode so that every block of 2000 symbols contains 600 known ones.

The WavefrontsII experiment during which the data used here was gathered used binary phase-shift keying (BPSK) modulation with a symbol rate of 24 000 symbols/s at a carrier frequency of 18 kHz. The experiment geometry included a source in a water depth of 6 m and a receiver 40 m directly inshore. The details of the experiment setup can be found in [2]. The received signals were downsampled to a fractional sample rate of 2 samples per symbol. The coefficients of the RPE-DFE feedforward and feedback filters were calculated based on the impulse response estimates. There was no error correcting code used on the transmitted sequence during the experiment.

A. Channel Characteristics

Fig. 1(a) shows a 1-min history of the time-varying channel impulse response estimated using an EWRLS algorithm with an exponential forgetting factor $\lambda=0.98$. As shown, the channel has a delay spread about 7 ms or 168 symbols. The multipath arrivals are well resolved in delay as identified in the figure. They form a fairly sparse multipath channel structure which also

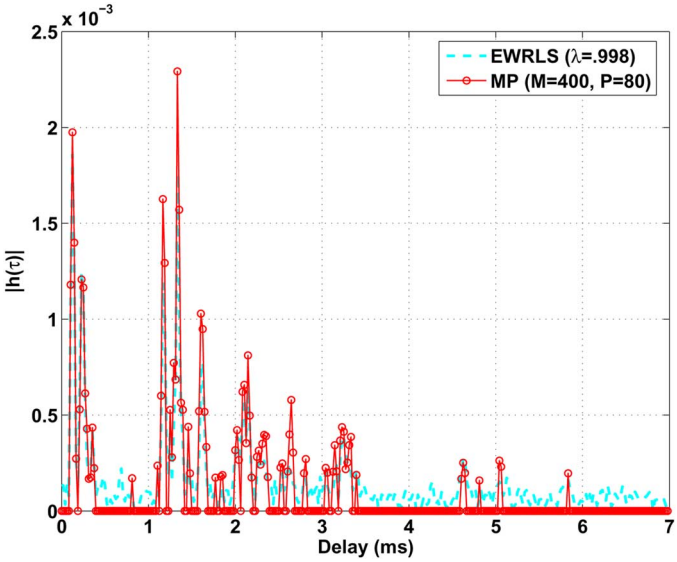


Fig. 3. Sparse and nonsparse channel impulse response estimation. The sparse estimate (dash line with dots) is obtained via the basic MP with averaging window length $M = 400$ and the number dominant taps is 80; nonsparse estimation (solid line) uses an EWRLS with $\lambda = 0.998$.

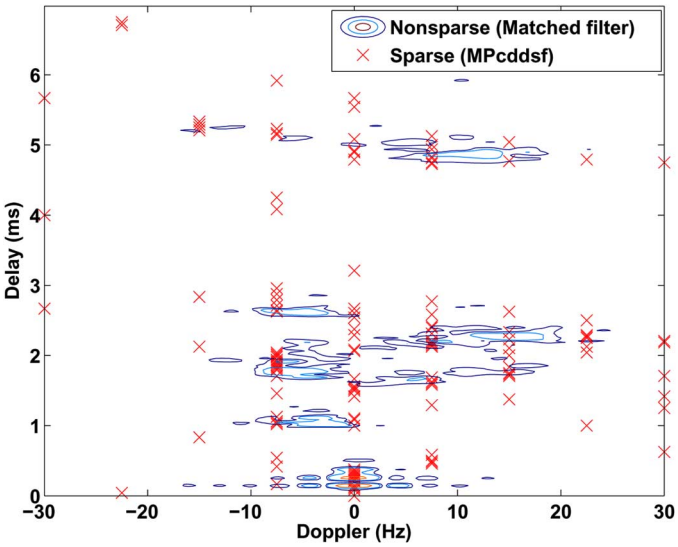


Fig. 4. Sparse and nonsparse estimation of the channel delay-Doppler-spread function. The sparse estimate (the cross marks) is obtained via the MP algorithm with averaging window length $M = 800$ and number of dominant components $P = 200$; the nonsparse estimate (the contour plot) is obtained via the matched filter.

varies over time. Plotted on the very top is the time series of the surface waveheight inferred from the pressure measurements obtained near the specular reflection point for the first surface bounce path. The correlation between the delay migration pattern of those surface arrivals and the surface waveheight time series is evident. Around 20 s a wave passage can be observed and the channel becomes noticeably dynamic, as shown in Fig. 1(b) where the bifurcation of single surface arrival into two oppositely moving arrivals forms butterfly patterns in the multipath structure. This multipath bifurcation is a result of a surface wave focusing [2]. The delay migration of the advancing branch of the first surface arrival has a corresponding path-centric Doppler shift around 7–8 Hz which together with the negative Doppler shift of the retreating branch gives rise to a total path-centric Doppler spread about 15–16 Hz. The Doppler spread is larger in later surface arrivals. Note that the path-centric Dopplers are closely related to but generally not the same as the delay-centric Dopplers [1]. Still, the path-centric Dopplers are good indications of the channel fluctuations. Several dynamic events have been labeled in Fig. 1(b) $(d_1 \sim d_7)$ for the time duration between 19.5 and 22 s. Among them, d_1, d_2, d_4, and d_7 see path crossing and others are associated with significant amplitude fluctuation and Doppler. It is shown in Figs. 6–9 that these dynamic events correspond to large errors in the channel estimation and equalization.

Fig. 2 shows a sequence of six snapshots of the time-varying scattering function estimated using a matched filter between 19.562 and 20.927 s when the surface gravity wave passes through the specular region. The time interval between adjacent estimates is 273 ms or 6552 symbols. The matched filter has an averaging window of 341 ms which can resolve approximately 2.93-Hz Doppler. Only the first three surface arrivals are covered in each snapshot. The sequence shows a nearly circular rotation of the second surface arrival which in plate 4 has a Doppler spread about 30 Hz. In addition, most of the channel energy is localized around some small areas on the delay-Doppler plane that correspond to those strong arrivals.

Fig. 3 plots a sparse channel impulse response estimate on top of a full channel impulse response estimate, obtained using the basic MP algorithm and the EWRLS algorithm, respectively. The basic MP algorithm uses an averaging window of $M = 400$ symbols and chooses 40 taps per subchannel totaling $P = 80$ taps. The EWRLS uses a forgetting factor $\lambda = 0.98$, corresponding to an equivalent averaging window length 500 symbols. The two show a good match at those energetic tap locations while the quiescent tap locations are not picked up by the basic MP algorithm. In Fig. 4, the location of the dominant components of the *delay-Doppler-spread function* (marked via crosses) obtained using the MP algorithm are plotted on the top of the contours of the nonsparse estimates of the *delay-Doppler-spread function*, at $t = 20.1077$ s. The sparse estimate is obtained using the basic MP algorithm with an averaging window of $M = 800$ symbols and the number of dominant components $P = 200$. The nonsparse estimate is obtained using the same matched filter used to generate the snapshots in Fig. 2. The contours of 0, -6, and -25 dB level below the peak value are shown. The *delay-Doppler-spread function* as shown has a delay span of 7 ms and a Doppler span of [-30, 30] Hz. Due to the Doppler sampling step used in the sparse estimation, 6.7 Hz in this case, the contour peaks of the matched filter estimate are not picked up by the MP algorithm. However, within the range of that Doppler sampling step, the association of those dominant components picked up by the MP algorithm with the contour peaks is evident.

B. Performance in Training Mode

In training mode, all channel estimation algorithms assume perfect knowledge of the transmitted sequence. Fig. 6 shows in training mode the signal prediction residual error of various channel estimation algorithms, over the time period from 18.4

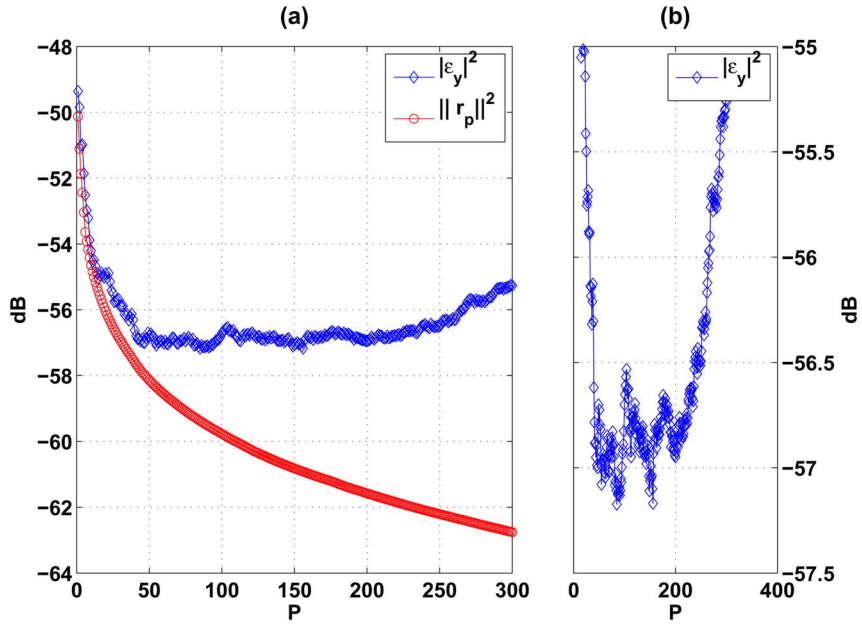


Fig. 5. *Delay-Doppler-spread function* estimation using the LSMP algorithm. (a) Magnitude squares of the residual error $\mathbf{r}_p[n] \triangleq \mathbf{y}[n] - \mathcal{C}[n]\hat{\mathbf{u}}_n^*$ and the prediction error $\epsilon_y[n+1] \triangleq |y[n+1] - \bar{\mathbf{c}}_1[n+1]\hat{\mathbf{u}}_n^*|^2$, as functions of P. Note that $\|\mathbf{r}_p[n]\|^2$ decreases monotonically as P increases while $|\epsilon_y[n+1]|^2$ decreases initially and then increases with P after $P > 200$. (b) Closeup plot of $|\epsilon_y[n+1]|^2$.

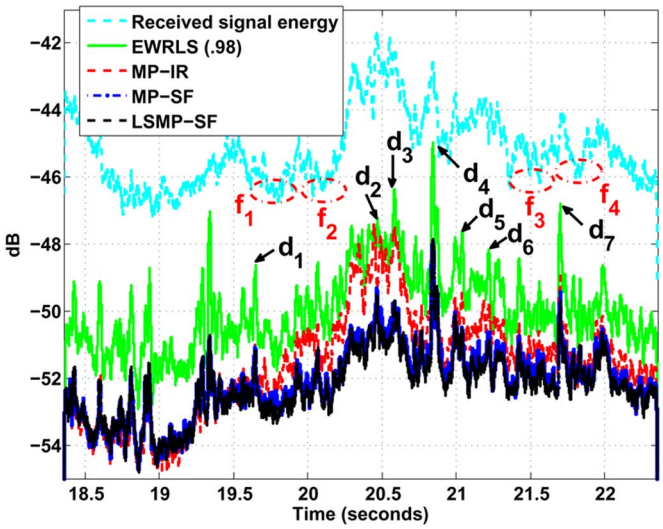


Fig. 6. Signal prediction residual errors in the training mode, with channel impulse response estimated using the EWRLS and the MP-IR algorithms and channel delay-Doppler-spread function estimation using the MP-SF and the LSMP-SF algorithms. The top curve is the total received signal energy.

to 22.4 s. For comparison, the total energy of the signal received during that period is also shown in Fig. 6. The signal prediction errors as shown have been averaged over 400 symbols.

The labels $d_1 \sim d_7$ signal the correspondence between the residual error peaks and the dynamic events previously marked in Fig. 1. Regions marked by $f_1 \sim f_4$ have low received signal level and, consequently, cause degraded equalizer performance.

The sparse estimations of the channel *delay-Doppler-spread function*, obtained using either the basic MP or the order-recursive LS-MP algorithm, achieve a uniform 2–3-dB reduction in the signal prediction error compared with the full channel impulse response estimation using EWRLS. While the sparse channel impulse response estimate reaps a similar gain before 19.5 s, it starts to deteriorate afterwards and hits the EWRLS curve around 20.5 s. Notice that the EWRLS shows some worst case residual error about 1 dB below the total received energy. This indicates that it is not effectively tracking the channel. The uniform gain of the *delay-Doppler-spread-function*-based sparse estimates can be attributed to the explicit incorporation of the channel variations via Doppler population and also the improvement of tracking capability by focusing on the dominant components. While the initial gain of the sparse channel impulse response estimate can be attributed to the similar tracking capability improvement due to the reduced number of taps, its failure to keep up with the variations when the channel fluctuates more rapidly after 19.5 s is caused by the lack of channel variation modeling.

Equalization based on 100% training has also been carried out for each case although for brevity the results are not shown. In the full training mode, the input to the feedback filter of the RPE-DFE consists of previous symbols that are assumed perfectly known, hence there is no error propagation. The DFE consists of a feedforward filter of 200 fractionally spaced taps and a feedbackward filter of 70 taps. The coefficients of both filters are computed from the channel estimates [22]. The results show similar comparative trends as in the residual errors. In addition, the soft decision error and the bit error deteriorate as the received signal level drops. This characteristic is also found in the decision-directed results shown in Section IV-C.

Note that although the absolute level of performance gain of the *delay-Doppler-spread-function*-based sparse channel estimation is not large, the results from the decision-directed mode in Section IV-C show that the realized improvement in the channel estimation performance makes a substantial improvement in the subsequent equalization performance.

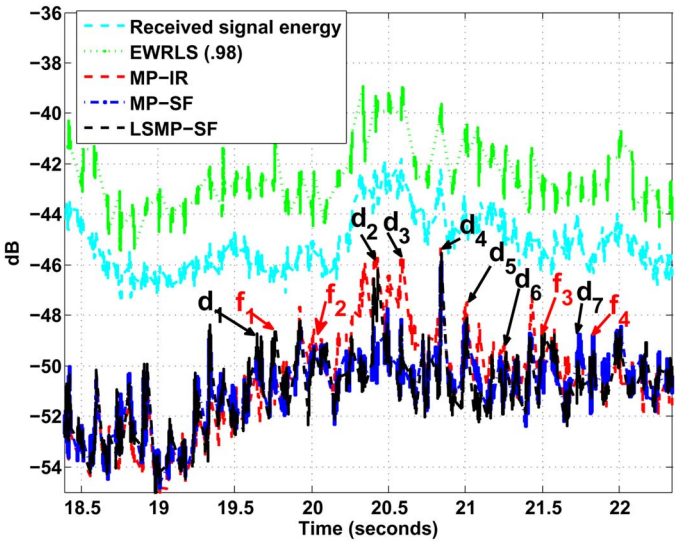


Fig. 7. Signal prediction residual errors in the decision-directed mode with periodic training, with channel impulse response estimated using the EWRLS and the MP-IR algorithms and channel delay-Doppler-spread function estimation using the MP-SF and the LSMP-SF algorithms. The top curve is the total received signal energy.

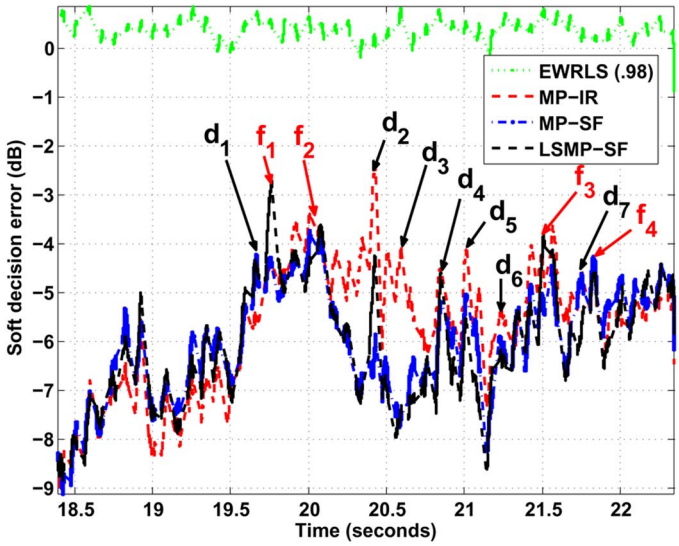


Fig. 8. Soft decision errors in decision-directed mode with periodic training, obtained using the RPE-DFE. The coefficients of RPE-DFE are computed from the channel impulse response estimates obtained directly using the EWRLS algorithm, the MP-IR algorithm, and converted from delay-Doppler-spread function estimates obtained from the MP-SF and the LSMP-SF algorithms.

C. Performance in Decision-Directed Mode

In the decision-directed mode, only the previously detected symbols are available to both the channel estimator and the feedback filter of the RPE-DFE. The DFE filter lengths are the same as in the training mode and the errors are evaluated similarly as in the training mode. Periodic training has been adopted to maintain robustness against channel fluctuations throughout the 1-min transmission data. Specifically, the first 600 symbols in every block of 2000 symbols are assumed known corresponding to 0.7 b/s/Hz spectral efficiency. While this is by no means an economic usage of the bandwidth, it is a reasonable choice considering the severe channel fluctuations and the fact that the transmitted sequence used in the experiment was not protected by error-correcting code. In fact, as shown in Figs. 7–9, even at such frequent training, the combination of a regular EWRLS channel estimation and the RPE-DFE still fails.

Although the data are processed in the decision-directed mode with periodic training, only the decision-directed portion of the error results are plotted. More specifically, for each block of 2000 symbols, only the data for symbols 601 to 2000 are included in the figures. Those in the training and the transition period are not shown.

Figs. 7–9 show in the decision-directed mode the signal prediction residual error of various channel estimation algorithms and the soft decision error and the bit error rate (BER) of the RFE-DFE based on those channel estimates, over the same time period as in Fig. 6.

First, as shown, the EWRLS channel impulse response estimate yields large symbol estimation errors (see Fig. 8 for soft decision error and Fig. 9 for BER) which in turn degrade the channel estimation performance (Fig. 7). In fact, the BER of the EWRLS algorithm in the decision-directed mode is between 40% and 50%. All the sparse-channel-estimation-based approaches as shown are fairly robust to channel fluctuations.

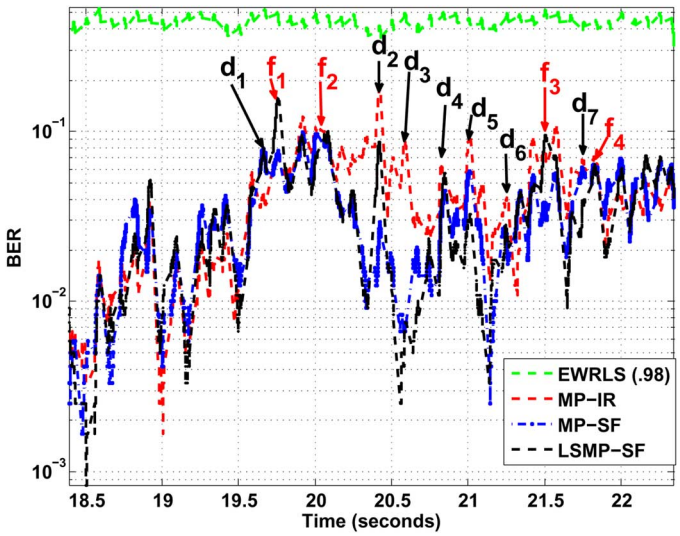


Fig. 9. BERs in the decision-directed mode with periodic training, obtained using the RPE-DFE. The coefficients of the RPE-DFE are computed from the channel impulse response estimates obtained directly using the EWRLS and the MP-IR algorithms, and converted from the delay-Doppler-spread function estimates obtained from the MP-SF and the LSMP-SF algorithms.

Notably, the two *delay-Doppler-spread-function*-based sparse estimation methods achieve significant error reduction over the MP channel IR estimation between 20 and 20.75 s when the channel varies rapidly.

Second, compared with those in the training mode (Fig. 6), the residual curves associated with sparse algorithms in the decision-directed mode as shown in Fig. 7 appear to have more peaky errors. In addition to peaks corresponding to the dynamic events $d_1 \sim d_7$ as labeled, there are also error peaks (identified by $f_1 \sim f_4$) at times when the received signal level is low. These residual error peaks may be attributed to the errors caused by an inaccurate symbol detection which has higher occurrences at lower SNR.

Similar labeling of both the dynamic events and the low SNR regions is used in Figs. 7 and 8 where, in low SNR regions $f_1 \sim f_4$, the *delay-Doppler-spread-function*-based approaches do not outperform sparse channel impulse response estimation. In fact, at f_1, both MP-SF and LSMP-SF perform worse than MP-IR, which may be attributed to the increased noise contribution and the fact that with fewer unknowns MP-IR is less sensitive to noise. However, between 20 and 20.75 s and especially at d_2 and d_3 both figures show significant error reduction from the MP-IR to both MP-SF and LSMP-SF. This is because the channel fluctuations dominate during these time periods. Both MP-SF and LSMP-SF are able to track these fluctuations but MP-IR cannot due to its lack of dynamic modeling. Also notice that performance gain is marginal at d_1, d_4, and d_7 when there are path crossings. This may be attributed to the fact that both MP-SF and LSMP-SF still average over a significantly long-time window. Inherently, they assume that the dynamics of the channel variations, e.g., delay-Doppler locations, are constant throughout the window which is violated when there are crossing paths.

V. Conclusion

A *delay-Doppler-spread-function*-based channel sparse estimation approach is proposed in this paper. The *delay-Doppler-spread function* representation accommodates the channel time variations by explicitly populating the symbol sequence in Doppler domain. Then, the dominant components on the delay-Doppler plane are identified using sparse estimation techniques. The main advantage of the proposed method is that both channel variations and sparse structure are addressed in a single approach. Parameter selection for the sparse channel estimation algorithms is also discussed. The performance gain of the proposed approach is demonstrated through analyzing a set of surf-zone experimental data. In the training mode, the results show uniform gain of the proposed approach over the EWRLS channel estimate in the signal prediction error. In the decision-directed mode, the proposed approaches are shown to be robust against severe channel fluctuations, while the traditional EWRLS channel estimator fails.

Appendix

Following (47) and (48)

$$\mathbf{R}_{p,j} = \left[\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\right]^h \left[\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\right] = \begin{bmatrix} \mathbf{R}_{p-1} & \mathbf{C}_{s,p-1}^h \mathbf{c}_j \\ \mathbf{c}_j^h \mathbf{C}_{s,p-1} & \|\mathbf{c}_j\|^2 \end{bmatrix} \tag{74}$$

$$\mathbf{z}_{p,j} = \left[\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\right]^h \mathbf{y} = \begin{bmatrix} \mathbf{z}_{p-1} \\ \mathbf{c}_j^h \mathbf{y} \end{bmatrix}. \tag{75}$$

Using the block matrix inversion formula [26] yields

$$\mathbf{R}_{p,j}^{-1} = \begin{bmatrix} \mathbf{R}_{p-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \delta_{p-1,j}^{-1} \mathbf{d}_{p-1} \mathbf{d}_{p-1}^h \tag{76}$$

where

$$\mathbf{d}_{p-1} \triangleq \begin{bmatrix} -\mathbf{R}_{p-1}^{-1} \mathbf{C}_{s,p-1}^h \mathbf{c}_j \\ 1 \end{bmatrix} \tag{77}$$

$$\begin{aligned} \delta_{p-1,j} &\triangleq \|\mathbf{c}_j\|^2 - \left[\mathbf{C}_{s,p-1}^h \mathbf{c}_j\right]^h \mathbf{R}_{p-1}^{-1} \left[\mathbf{C}_{s,p-1}^h \mathbf{c}_j\right] \\ &= \mathbf{c}_j^h \left[\mathbf{I} - \mathbf{C}_{s,p-1} \mathbf{R}_{p-1}^{-1} \mathbf{C}_{s,p-1}^h\right] \mathbf{c}_j \\ &\triangleq \mathbf{c}_j^h \mathbf{Q}_{p-1} \mathbf{c}_j \end{aligned} \tag{78}$$

where $\mathbf{Q}_{p-1} \triangleq \mathbf{I} - \mathbf{C}_{s,p-1} \mathbf{R}_{p-1}^{-1} \mathbf{C}_{s,p-1}^h$ is the projection matrix associated with the nullspace of $\mathbf{C}_{s,p-1}$. $\delta_{p-1,j}$ is the Shur complement of $\mathbf{R}_{p-1}$ and has the geometric meaning as the LS distance between the vector $\mathbf{c}_j$ and the subspace spanned by $\mathbf{C}_{s,p-1}$. $\delta_{p-1,j} = \|\mathbf{c}_j\|^2$ if $\mathbf{c}_j$ is orthogonal to $\mathbf{C}_{s,p-1}$.

Combining (45), (46), and (74)–(78) yields the following recursion form for the residual vector:

$$\mathbf{r}_{p,j} = \mathbf{r}_{p-1} - \frac{\mathbf{Q}_{p-1} \mathbf{c}_j^h \mathbf{r}_{p-1}}{\delta_{p-1,j}} \mathbf{c}_j. \tag{79}$$

Suppose the s_pth column is selected at the pth iteration, then (79) is evaluated using $j = s_p$.

It is then analog to (23) and becomes identical if $\mathbf{Q}_{p-1}$ is an identity matrix. It can be shown that $\mathbf{r}_p$ is orthogonal to $\mathbf{c}_{s_i}$ for all $i < p$.

Denoting $b_{p-1,j} \triangleq \mathbf{c}_j^h \mathbf{r}_{p-1}$, it follows that

$$\begin{aligned} b_{p,j} &= b_{p-1,j} - \left[\mathbf{c}_j^h \mathbf{Q}_{p-1} \mathbf{c}_{s_p}\right] \delta_{p-1,s_p}^{-1} \mathbf{c}_{s_p}^h \mathbf{r}_{p-1} \\ &= b_{p-1,j} - \left[\mathbf{c}_j^h \mathbf{Q}_{p-1} \mathbf{c}_{s_p}\right] \widehat{x}_p \end{aligned} \tag{80}$$

where

$$\widehat{x}_p \triangleq \delta_{p-1,s_p}^{-1} \mathbf{c}_{s_p}^h \mathbf{r}_{p-1} \tag{81}$$

is the coefficient associated with $\mathbf{c}_{s_p}$. Combining (46), (76), and (81) yields the new coefficient vector

$$\widehat{\mathbf{x}}_p = \begin{bmatrix} \widehat{\mathbf{x}}_{p-1} \\ 0 \end{bmatrix} + \mathbf{d}_{p-1} \widehat{x}_p. \tag{82}$$

Using the orthogonality principle and the idempotent property of $\mathbf{Q}_{p-1}$, i.e., $\mathbf{Q}_{p-1}\mathbf{Q}_{p-1} = \mathbf{Q}_{p-1}$, the squared residual error associated with the set of columns $\left[\mathbf{C}_{s,p-1} \quad \mathbf{c}_j\right]$ is given by

$$\begin{aligned} \|\mathbf{r}_{p,j}\|^2 &= \|\mathbf{y}\|^2 - \mathbf{z}_{p,j}^h \mathbf{R}_{p,j}^{-1} \mathbf{z}_{p,j} \\ &= \|\mathbf{y}\|^2 - \mathbf{z}_{p-1}^h \mathbf{R}_{p-1}^{-1} \mathbf{z}_{p-1} - \delta_{p-1,j}^{-1} \left|\mathbf{y}^h \mathbf{Q}_{p-1} \mathbf{c}_j\right|^2 \\ &= \|\mathbf{r}_{p-1}\|^2 - \delta_{p-1,j}^{-1} \left|\mathbf{r}_{p-1}^h \mathbf{c}_j\right|^2. \end{aligned} \tag{83}$$

Consequently, the newly chosen column $\mathbf{c}_{s_p}$ should maximize the second term in (83)

$$s_p = \arg \max_{j \notin I_{p-1}} \frac{\left|\mathbf{c}_j^h \mathbf{r}_{p-1}\right|^2}{\delta_{p-1,j}} = \arg \max_{j \notin I_{p-1}} \frac{|b_{p-1,j}|^2}{\delta_{p-1,j}}. \tag{84}$$

Also, using $\mathbf{Q}_{p-1}$ and (76)–(78), it can be shown that

$$\delta_{p,j} = \delta_{p-1,j} - \frac{\left|\mathbf{c}_j^h \mathbf{C}_{s,p}\mathbf{d}_{p-1}\right|^2}{\delta_{p-1,s_p}}. \tag{85}$$

Furthermore, using (46), it follows that

$$b_{p,j} = b_{0,j} - \mathbf{c}_j^h \mathbf{C}_{s,p}\widehat{\mathbf{x}}_p. \tag{86}$$

Equations (85) and (86) are listed in Table II as (68) and (69), respectively.

REFERENCES

[1] W. Li, "Estimation and tracking of rapidly time-varying broadband acoustic communication channels," Ph.D. dissertation, MIT/WHOI Joint Program, Massachusetts Inst. Technol., Cambridge, MA, 2005.

[2] J. C. Preisig and G. B. Deane, "Surface wave focusing and acoustic communications in the surf zone," *J. Acoust. Soc. Amer.*, vol. 116, no. 3, pp. 2067–2080, Sep. 2004.

[3] M. Stojanovic, L. Freitag, and M. Johnson, "Channel-estimation-based adaptive equalization of underwateracoustic signals," in *Proc. MTS/IEEE OCEANS*, 1999, vol. 2, pp. 985–990.

[4] M. Stojanovic, "Efficient acoustic signal processing based on channel estimation for high rate underwater information," *J. Acoust. Soc. Amer.*, submitted for publication.

[5] A. Rontogiannis and K. Berberidis, "Bandwidth efficient transmission through sparse channels using a parametric channel-estimation-based DFE," *Proc. Inst. Electr. Eng.—Commun.*, vol. 152, no. 2, pp. 251–256, 2005.

[6] Y.-F. Cheng and D. Etter, "Analysis of an adaptive technique for modeling sparse systems," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 2, pp. 254–264, Feb. 1989.

[7] P.-W. Yip and D. Etter, "An adaptive multiple echo canceller for slowly time-varying echo," *IEEE Trans. Commun.*, vol. 38, no. 10, pp. 1693–1698, Oct. 1990.

[8] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. MTS/IEEE OCEANS*, 1995, vol. 3, pp. 1417–1422.

[9] S. Ozen, W. Hillery, M. Zoltowski, S. M. Nereyanuru, and M. Fimoff, "Structured channel estimation based decision feedback equalizers for sparse multipath channels with applications to digital TV receivers," in *Proc. 36th Asilomar Conf.*, 2002, vol. 1, pp. 558–564.

[10] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[11] D. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," M.S. thesis, Electr. Eng. Comp. Sci., Massachusetts Inst. Technol., Cambridge, MA, 2003.

[12] J. J. Fuchs and B. Delyon, "Minimal L1-norm reconstruction function for oversampled signals: Applications to time-delay estimation," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1666–1673, Jul. 2000.

[13] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[14] B. K. Natarajan, "Sparse approximation solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.

[15] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[16] S. Cotter and B. Rao, "The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels," in *Proc. 34th Asilomar Conf. Signals Syst. Comput.*, 2000, vol. 2, pp. 1772–1776.

[17] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[18] M. Cetin and B. Sadler, "Semi-blind sparse channel estimation with constant modulus symbols," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Mar. 2005, vol. III, pp. 561–564.

[19] W. Li and J. C. Preisig, "Sparse estimation and equalization of rapidly varying wideband acoustic communication channels," in *Proc. IEEE/MTS OCEANS*, Boston, MA, 2006, pp. 1–6.

[20] T. H. Eggen, A. B. Baggeroer, and J. C. Preisig, "Communication over Doppler spread channels, Part I: Channel and receiver presentation," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 62–72, Jan. 2000.

[21] T. H. Eggen, J. C. Preisig, and A. B. Baggeroer, "Communication over Doppler spread channels, Part II: Receiver characterization and practical results," *IEEE J. Ocean. Eng.*, vol. 26, no. 4, pp. 612–622, Oct. 2001.

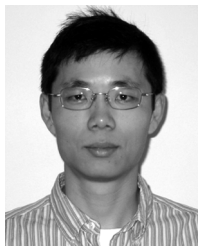
[22] J. C. Preisig, "Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment," *J. Acoust. Soc. Amer.*, vol. 118, no. 1, pp. 263–278, Jul. 2005.

[23] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 4, pp. 360–393, Dec. 1963.

[24] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002, p. 437.

[25] H. L. Van-Tree, *Optimum Array Processing*. New York: Wiley, 2002, p. 47.

[26] D. Harville, *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag, 1997, p. 99.



Weichang Li (S'04–M'05) received the B.S. and S.M. degrees in acoustics and electronic engineering from Harbin Engineering University, Harbin, China, in 1993 and 1996, the S.M. degree in electrical engineering and computer science and ocean engineering from Massachusetts Institute of Technology, Cambridge, in 2002, and the Ph.D. degree in electrical and oceanographic engineering from the Massachusetts Institute of Technology/Woods Hole Oceanographic Institute Joint Program in Oceanography in February 2006.

From 1996 to 1999, he was a Research Engineer and then Assistant Scientist, working on acoustical signal processing and DSP, at the Institute of Acoustics, Chinese Academy of Sciences, China. Currently, he is a Postdoctoral Fellow at Woods Hole Oceanic Institute (WHOI), Woods Hole, MA. His research interests are in the areas of signal processing, underwater acoustic communications, and digital holographic imaging of biological particles.

Dr. Li is the recipient of the 2006 Office of Naval Research (ONR) Ocean Acoustic Postdoctoral Fellowship Award and is a member of the Acoustical Society of America.

James C. Preisig (S'79–M'80) received the B.S. degree in electrical engineering from the United States Coast Guard Academy in 1980, the S.M. and E.E. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1988, and the Ph.D. degree in electrical and ocean engineering from the Massachusetts Institute of Technology/Woods Hole Oceanographic Institution (WHOI) Joint Program in Oceanography and Oceanographic Engineering in 1992.

He was a Postdoctoral Investigator at WHOI from 1992 to 1994 and a Visiting Assistant Professor at Northeastern University from 1994 to 1997. Since July 1997, he has been on the scientific staff of the Department of Applied Ocean Physics and Engineering at WHOI and is currently an Associate Scientist with Tenure. His research interests are in the areas of adaptive signal processing, system identification, underwater acoustic propagation modeling, underwater acoustic communications, and numerical optimization.

Dr. Preisig is the recipient of the 1999 U.S. Office of Naval Research Ocean Acoustics Young Faculty Award and is a member of the Acoustical Society of America's Underwater Acoustics and Signal Processing Technical Committees. He is also an Associate Editor of the IEEE JOURNAL OF OCEANIC ENGINEERING and served as a member of the IEEE Sensor Array and Multichannel Signal Processing Technical Committee from 1998 to 2004.