Greedy Algorithms for Sparse Adaptive Decision Feedback Equalization

Aris S. Lalos[‡], Evangelos Vlachos[‡], Kostas Berberidis[‡], Athanasios A. Rontogiannis[*]
[‡] Department of Computer Engineering and Informatics
University of Patras & C.T.I RU-8, 26500, Rio - Patra, Greece
[*] Institute for Space Applications and Remote Sensing, National Observatory of Athens, 15236, Palea Penteli, Athens, Greece
Emails: {lalos,vlaxose,berberid}@ceid.upatras.gr, tronto@space.noa.gr

***Abstract*—In this paper we propose two new adaptive decision feedback equalization (DFE) schemes for channels with long and sparse impulse responses. It has been shown that for a class of channels, and under reasonable assumptions concerning the DFE filter sizes, the feedforward (FF) and feedback (FB) filters possess also a sparse form. The sparsity form of both the channel impulse response (CIR) and the equalizer filters is properly exploited and two novel adaptive greedy schemes are derived. The first scheme is a channel estimation based one. In this scheme, the non-negligible taps of the involved CIR are first estimated via a new greedy algorithm, and then the FF and FB filters are adaptively computed by exploiting a useful relation between these filters and the CIR. The channel estimation part of this new technique is based on the steepest descent (SD) method and offers considerably improved performance as compared to other adaptive greedy algorithms that have been proposed. The second scheme is a direct adaptive sparse equalizer based on a SD-based greedy algorithm. Compared to non sparsity aware DFE, both of our schemes exhibit faster convergence, improved tracking capabilities and reduced complexity.**

I. Introduction

In several high-speed wireless communication systems, the involved multipath channels are characterized by long CIR consisting of a few dominant components some of which may have quite large time delays with respect to the main signal. Some typical applications of the kind are Digital Video Broadcasting, Broadband Wireless Networks and Underwater Digital Communications, to name but a few. Adaptive decision feedback equalization (DFE) has been widely used in single-carrier transmission systems as an effective technique for reducing the introduced Intersymbol Interference (ISI). Moreover, it has been shown that for the particular class of channels we consider here, and under reasonable assumptions concerning the filter sizes, both the DFE filters (i.e., feedforward and feedback) exhibit a sparse structure [1].

The exploitation of sparsity has been attracting recently an interest of exponential growth. Two major algorithmic approaches to sparse system identification are ℓ_1-minimization (basis pursuit), and greedy algorithms (matching pursuit). Basis pursuit methods solve a convex minimization problem, in which the ℓ_0 quasi-norm is replaced by the ℓ_1 norm. Greedy algorithms, on the other hand, compute iteratively the support set of the signal and construct an approximation of it until a halting condition is met [2], [3]. The ℓ_1-minimization methods provide theoretical performance guarantees but they lack the speed of greedy techniques. Recently developed greedy algorithms, such as those developed in [2], [3] provide some performance guarantees, as the ℓ_1-minimization approaches, offering less computational cost and storage.

Adaptive implementations of the aforementioned approaches have been proposed in the relevant literature. Chen et al. [4] incorporated two different sparsity constraints based on ℓ_1 relaxation in order to improve the performance of LMS-type adaptive methods. In [5], Angelosante et al. developed a recursive sub-gradient based approach for solving the batch Lasso estimator. An ℓ_1-regularized RLS type algorithm based on a low complexity Expectation-Maximization scheme is derived in [6] by Babadi et al. Mileounis et al. in [7] propose a conversion procedure that turns greedy algorithms into adaptive schemes for sparse system identification. Due to the fact that the update rule of RLS cannot be directly restricted to the support set (set of non zero coefficients), the authors propose an LMS update of this set.

To the best of our knowledge, little work has been done towards developing sparse equalization schemes. In this paper, we propose two different approaches for designing a DFE. The first one is channel estimation based, i.e., first, the non-negligible taps of the involved CIR are estimated via a new adaptive greedy algorithm, and then the FF and FB filters are approximated by sparse vectors by exploiting a useful relation between these filters and the CIR. The channel estimation part of this new technique is based on the steepest descent method and offers a significantly improved performance as compared to the algorithm proposed in [7]. In the second approach, a direct adaptive sparse equalization scheme based on a SD-based greedy algorithm is presented. It has been verified through simulations that the sparse channel estimation based equalization scheme outperforms the direct sparse equalization one. This can be justified by the fact that the sparsity order of the first approach is smaller. Furthermore the convergence properties of the adaptive schemes are studied theoretically.

The rest of the paper is organized as follows: In Section 2, the problem is formulated and some preliminaries concerning greedy methods are provided. In Section 3, the new adaptive schemes are derived and analyzed theoretically. Simulation results are presented in Section 4. Finally, Section 5 concludes

978-1-4673-0753-6/11/$26.00 ©2011 IEEE

the paper.

II. Problem Formulation and Preliminaries

The sampled output of a sparse multipath channel can be expressed as follows:

$$x\left(k\right)=\sum_{l=1}^{S}h_{k_l}^{*}u\left(k-k_l\right)+\eta\left(k\right) \tag{1}$$

where S is the number of the dominant CIR components appearing at the symbol-spaced time instants, h_{k_l} is the complex amplitude of the l-th component, and k_l its respective delay. Delay k_1 corresponds to the main signal ($k_1=0$), while the remaining ones correspond either to causal ($k_l>0$) or to anticausal ($k_l<0$) components. The input $\{u(k)\}$ is an independent identically distributed (i.i.d.) symbol sequence with variance σ_u^2 and the noise samples $\eta(k)$ are assumed zero mean and uncorrelated across time, each with known variance σ_η^2. The symbol-spaced CIR spans n_1 precursor and n_2 postcursor symbols, respectively and can be written in vector form as $\mathbf{h}=\left[h_{-n_1}\cdots h_0\cdots h_{n_2}\right]^T$. From the total $N=n_1+n_2+1$ CIR coefficients, only S are assumed to be nonnegligible, located at the positions $k_1,\cdots,k_S$.

The intersymbol interference involved in the system described by (1) can be mitigated through a DFE. The DFE is designed to extract the stream $u\left(k\right)$ and it consists of a FF and a FB filter of temporal span K_f and K_b taps, respectively. The input to the FF filter $\mathbf{f}$ can be described by the $K_f\times 1$ vector

$$\mathbf{x}(k)=\left[\begin{array}{ccc} x(k) & \cdots & x(k-K_f+1) \end{array}\right]^T \tag{2}$$

Similarly, if $\tilde{d}(k)$ denotes the output of the DFE and $d(k)=f\{\tilde{d}(k)\}$ is the corresponding decision device output, then the input to the FB filter $\mathbf{b}$ can be expressed by the $K_b\times 1$ vector

$$\mathbf{d}(k)=\left[\begin{array}{ccc} d(k-1) & \cdots & d(k-K_b) \end{array}\right]^T \tag{3}$$

By using the above definitions, the output of the DFE can be compactly expressed as

$$\tilde{d}(k)=\mathbf{w}^H\mathbf{y}(k),\quad \mathbf{w}=\left[\mathbf{f}^T\ \ \mathbf{b}^T\ \ \right]^T, \tag{4}$$

$$\mathbf{y}(k)=\left[\mathbf{x}^T(k)\ \mathbf{d}^T(k)\ \right]^T \tag{5}$$

In the minimum mean-squared error (MMSE) DFE, the equalizer coefficients $\mathbf{w}$ are obtained by minimizing the MMSE cost function $\hat{\mathbf{w}}=arg\min_{\mathbf{w}}E\{|u(k)-\tilde{d}(k)|^2\}$. Under the assumption that previously detected symbols are correct, the FF and FB filters $\mathbf{f},\mathbf{b}$ can be expressed in terms of the CIR coefficients. More specifically, by assuming that the maximum sum of any K_f consecutive CIR coefficients, h_i with $i\in[-n_1,K_f],i\neq 0$, is less than $|h_0|$, and that the FF length is several times larger than the anticausal part of the channel ($K_f\geq 3n_1$), we can estimate $\mathbf{f}$ by using the following approximation originally proposed in [1]:

$$\hat{\mathbf{f}}\approx h_0^{*-1}\mathbf{e}_{K_f}-h_0^{*-2}\sum_{k_i<0}h_{k_i}^{*}\mathbf{e}_{K_f+k_i}$$
$$+h_0^{*-3}\sum_{k_i+k_j<0}\sum h_{k_i}^{*}h_{k_j}^{*}\mathbf{e}_{K_f+k_i+k_j} \tag{6}$$

where $\mathbf{e}_K$ is the unit $K_f\times 1$ vector with 1 as its K_f-th entry.

Thus we easily derive the following results concerning $\hat{\mathbf{f}}$: i)There exists one coefficient equal to h_0^{*-1} as the last element of $\hat{\mathbf{f}}$, ii)For each $k_i<0$ there exists a coefficient equal to $-h_0^{*-2}h_{k_i}^{*}$ located at the position K_f+k_i, iii) For each combination of k_i,k_j with $k_i+k_j<0$ there exists a coefficient equal to $-h_0^{*-3}h_{k_i}^{*}h_{k_j}$ located at the position $K_f+k_i+k_j$. The FB filter $\mathbf{b}$ can be obtained as :

$$\hat{\mathbf{b}}=-\begin{bmatrix} h_{K_f} & h_{K_f+1}\cdots h_{n_2} & 0 & \cdots & 0\\ h_{K_f-1} & h_{K_f}\ \ \cdots\cdots h_{n_2} & \cdots & 0 & \\ \vdots & \vdots\quad\vdots & \vdots & \vdots & \\ h_1 & h_2\ \ \cdots\cdots\cdots\cdots & h_{n_2} & & \end{bmatrix}^H\hat{\mathbf{f}}. \tag{7}$$

Thus for the class of channels we consider here and under reasonable assumptions concerning the filter sizes, the filters $\mathbf{f},\mathbf{b}$ can be approximated by sparse vectors.

A. The CoSaMP Algorithm

The $N\times 1$ vector $\mathbf{h}$ is called sparse since only a few of its entries $\{h_{k_1},\ldots,h_{k_S}\}$ are nonzero. Upon defining the nonzero support set of $\mathbf{h}$ as $supp\left(\mathbf{h}\right)=\{i\in\{1,\ldots,N\}:h_i\neq 0\}$, sparsity amounts to having $|supp(\mathbf{h})|=S<<N$, where $|\cdot|$ denotes set cardinality. Suppose that such a sparse signal vector $\mathbf{h}$ is to be estimated from the n observations $\{x(k)\}$, $k=1,\ldots,n$. Each observation (channel output) can be written as:

$$x(k)=\mathbf{h}^H\mathbf{u}(k)+\eta(k),\quad k=1,\ldots,n \tag{8}$$

where $\mathbf{u}(k)=[u(k+n_2),u(k+n_2-1),\ldots,u(k-n_1)]^T$ are known $N\times 1$ transmitted vectors.

Given $\{x(k),\mathbf{u}(k)\}$, for $k=1,\ldots,n$ and σ^2, a batch approach can be applied to the concatenated model

$$\mathbf{x}(n)=\mathbf{U}(n)\mathbf{h}^{*}+\boldsymbol{\eta}(n) \tag{9}$$

$$\mathbf{x}(n)=[x(n)\ldots x(1)]^T,\ \mathbf{U}(n)=[\mathbf{u}(n),\ldots,\mathbf{u}(1)], \tag{10}$$

The $n\times 1$ vector $\boldsymbol{\eta}(n)$ consists of noise samples. Recovery of $\mathbf{h}$ can be pursued by finding the sparsest estimate of $\mathbf{h}$ that satisfies the ℓ_2 norm error tolerance δ, i.e.,

$$\min_{\mathbf{h}}\|\mathbf{h}\|_{\ell_0}\quad s.t.\quad \|\mathbf{x}(n)-\mathbf{U}(n)\mathbf{h}^{*}\|_{\ell_2}\leq\delta \tag{11}$$

Convex relaxation methods cope with the intractability of the above formulation by approximating the ℓ_0 quasi-norm by convex ℓ_1-minimization. The ℓ_2 constraint can be interpreted as a noise removal mechanism when $\delta\geq\|\boldsymbol{\eta}(n)\|_{\ell_2}^2$.

Another approach to the combinatorial problem (11) is based on dynamic programming. In this type of approach the combinatorial problem is circumvented by heuristically choosing which values of $\mathbf{h}$ are non-zero (support set), setting the others to zero and solving the resulting least-squares problem $\hat{\mathbf{h}}=arg\min_{\mathbf{h}}\|\mathbf{x}(n)-\mathbf{U}(n)\mathbf{h}^{*}\|_{\ell_2}^2$.

The most popular algorithms of this type are greedy algorithms, like Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP). Although OMP is quite fast, it is unknown whether it succeeds on noisy measurements. A more sophisticated algorithm, called Compressed Sampling Matching

TABLE I
BASIC STEPS OF THE COSAMP ALGORITHM

Repeat until a stoping criterion is met:
1) Identification of the $2S$ largest components of the proxy signal.
2) Support merging, which forms the union of the set of newly identified indexes of the most correlated columns with the set of indices corresponding to the S largest components of the least squares estimate obtained in the previous iteration
3) Estimation via LS on the merged set of components
4) Pruning, which restricts the LS estimate to its S largest components
5) Sample update, which updates the error residual.

Pursuit algorithm (CoSaMP) and developed by Needell and Tropp [2], is known to provide nearly optimal performance guarantees. CoSaMp takes advantage of the measurement matrix $\mathbf{U}(n)$ which is approximately orthonormal. Thus, the largest components of the so-called proxy signal vector $\mathbf{p}(n) = \mathbf{U}^H(n)\mathbf{U}(n)\mathbf{h}$ most likely correspond to the non-zero entries of $\mathbf{h}$ in case the measurement matrix satisfies the RIP property [8]. The basic steps of the CoSaMP algorithm are presented in table I.

III. SPARSE ADAPTIVE DFE TECHNIQUES

When the CIR changes within a burst (a case arising for relatively long bursts and/or under fast varying conditions), the batch approaches fail to equalize the channel and efficient adaptive methods are required. Two different equalization schemes that are based on the conversion of greedy algorithms into adaptive ones are proposed. The first scheme exploits the sparsity of the CIR by applying a new adaptive greedy algorithm to estimate the non-zero taps $h_{k_1}, \dots, h_{k_S}$. Then, the FF and FB filters are approximated by sparse vectors expressed in terms of $h_{k_1}, \dots, h_{k_S}$, as shown in eqs. (6),(7). The second scheme exploits the sparsity of the DFE equalizer by applying directly the new adaptive greedy algorithm to estimate the non negligible taps of the DFE . The sparsity orders of the equalizer filters are assumed to be fixed and in both schemes are adjusted to be the same.

A. Scheme I: Sparse Channel Estimation Based Adaptive DFE

The sampled output of the sparse multipath channel at time n may be written as :

$$x(n) = \mathbf{h}^H(n)\mathbf{u}(n) + \eta(n) \tag{12}$$

If a sparsity constraint is not imposed, the CIR can be obtained in an adaptive way as the minimizer of the following cost function:

$$\hat{\mathbf{h}}(n) = arg \min_{\mathbf{h}(n)} \sum_{k=1}^{n} \lambda^{n-k} \left| x(k) - \mathbf{h}^H(n)\mathbf{u}(k) \right|^2 \tag{13}$$

where $\lambda \in (0, 1]$ is a forgetting factor. Thus, at time n the above cost function can be alternatively written as:

$$J\left(\mathbf{h}(n), \mathbf{\Phi}_{uu}(n), \mathbf{z}_{xu}(n)\right) = \frac{\mathbf{h}^H(n)\mathbf{\Phi}_{uu}(n)\mathbf{h}(n)}{2} - Re\{\mathbf{h}^H(n)\mathbf{z}_{xu}(n)\} \tag{14}$$

where matrix $\mathbf{\Phi}_{uu}(n)$ stands for the $N \times N$ exponentially time-averaged input autocorrelation matrix, and $\mathbf{z}_{xu}(n)$ for the crosscorrelation vector, defined in matrix-vector form as:

$$\mathbf{\Phi}_{uu}(n) = \mathbf{U}(n)\mathbf{W}(n)\mathbf{U}^H(n), \tag{15}$$

$$\mathbf{z}_{xu}(n) = \mathbf{U}(n)\mathbf{W}(n)\mathbf{x}^*(n) \tag{16}$$

where $\mathbf{W}(n) = diag(1, \lambda, \dots, \lambda^{n-1})$. The $n \times N$ matrix $\mathbf{U}(n)$ and the $n \times 1$ vector $\mathbf{x}(n)$ have been already defined in (10).The proposed algorithm relies on the CoSaMP structure. More specifically, the error residual at time n is given by

$$\mathbf{e}(n) = \mathbf{x}(n) - \mathbf{U}(n)\hat{\mathbf{h}}^*(n) \tag{17}$$

where $\hat{\mathbf{h}}(n)$ is a sparse vector with $N - S$ zeros obtained in the previous iteration (time slot). At each iteration, the current approximation induces an error residual that corresponds to the part of the CIR that has not been approximated. As the algorithm progresses, the samples are updated so that they reflect the current error residual. These samples are used to construct a proxy signal, which point toward the largest S entries of the error residual. Thus the new proxy signal may be computed similarly to that of the CoSaMP algorithm [2] in order to incorporate the current correlation and forget the data, capturing the variations on the support set of the channel:

$$\mathbf{p}(n) = \mathbf{U}(n)\mathbf{W}(n)\mathbf{e}^*(n) \tag{18}$$

By combining eqs. (17) and (18) it can be easily shown that the proxy signal correpsonds to the gradient of the cost function defined in (14). The gradient can be evaluated as

$$\mathbf{g}(n) = \mathbf{z}_{xu}(n) - \mathbf{\Phi}_{uu}(n)\hat{\mathbf{h}}(n) \tag{19}$$

By inspecting eqs. (15) and (16) it can be easily seen that $\mathbf{g}(n) = \mathbf{p}(n)$. Thus in order to identify the new support set we select the largest $2S$ coefficients of the estimated gradient, i.e.

$$\Omega = max\left(\left|\mathbf{g}(n)\right|, 2S\right). \tag{20}$$

To proceed, we form a union of Ω with the set of indices corresponding to the S non-zero components of the LS estimate $\mathbf{h}(n)$ obtained in the previous iteration:

$$\Lambda = \Omega \cup supp\left(\hat{\mathbf{h}}(n)\right) \tag{21}$$

Then, we focus on estimating the coefficients of the channel that correspond to the index set Λ. These coefficients denoted as $\mathbf{h}_{|\Lambda}(n)$ can be computed as the minimizer of the cost function $J\left(\mathbf{h}_{|\Lambda}(n), \mathbf{\Phi}_{uu,|\Lambda}(n), \mathbf{z}_{xu,|\Lambda}(n)\right)$ with respect to $\mathbf{h}_{|\Lambda}(n)$. Note that $\mathbf{h}_{|\Lambda}(n)$ and $\mathbf{z}_{zu,|\Lambda}(n)$ denote the $Q \times 1$ coefficient and crosscorelation sub-vectors corresponding to the index set Λ, while $\mathbf{\Phi}_{uu,|\Lambda}(n)$ is a $Q \times Q$ matrix formed by the corresponding to Λ rows and columns of $\Phi_{uu}(n)$, with $Q \in [2S, 3S]$.

The coefficient sub-vector estimate $\hat{\mathbf{h}}_{|\Lambda}(n)$, can be updated by a standard adaptive algorithm. The least mean squares (LMS) is one of the most widely used algorithms in adaptive filtering due to its simplicity, robustness and low complexity. However, his poor convergence properties usually render it

978-1-4673-0753-6/11/$26.00 ©2011 IEEE

impractical in cases where fast convergence is required. On the other hand, the recursive least squares (RLS) algorithm is one order of magnitude costlier but significantly improves the convergence speed of LMS. However, the update rule for RLS cannot be directly restricted to the index support set. To overcome these drawbacks and take advantage of the fact that we already possess an estimate of the gradient of the cost function defined in (14), we apply a steepest descent coefficient update.

Following the SD method, the $Q \times 1$ channel coefficient subvector $\hat{\mathbf{h}}_{|\Lambda}(n)$ can be updated as:

$$\hat{\mathbf{h}}_{|\Lambda}(n+1) = \hat{\mathbf{h}}_{|\Lambda}(n) + \alpha(n)\mathbf{g}_{|\Lambda}(n) \tag{22}$$

Note that $\hat{\mathbf{h}}_{|\Lambda}(n)$ and $\mathbf{g}_{|\Lambda}(n)$ denote the coefficient and the gradient subvectors corresponding to the index set Λ. In the proposed algorithm, the step size $\alpha(n)$ is selected as the minimizing argument of $J\left(\mathbf{h}_{|\Lambda}(n), \mathbf{\Phi}_{uu,|\Lambda}(n), \mathbf{z}_{xu|\Lambda}(n)\right)$ with respect to $\alpha(n)$, i.e.,

$$\alpha(n) = \frac{\mathbf{g}_{|\Lambda}(n)^H \mathbf{g}_{|\Lambda}(n)}{\mathbf{g}_{|\Lambda}(n)^H \mathbf{\Phi}_{uu,|\Lambda}(n)\mathbf{g}_{|\Lambda}(n)} \tag{23}$$

As verified by simulations, the proposed greedy scheme significantly improves the performance of the scheme presented in [7] at the cost of a slight increase in complexity. This increase is due to the computation of the gradient $\mathbf{g}(n)$ in (19) and the extra matrix-by-vector product $\mathbf{\Phi}_{uu,|\Lambda}(n)\mathbf{g}_{|\Lambda}(n)$ in (23), being of the order of $\mathcal{O}(NS)$ and $\mathcal{O}\left(S^2\right)$ respectively.

In the next step of the algorithm, the CIR estimate $\hat{\mathbf{h}}(n+1)$ results from the S largest coefficients of $\hat{\mathbf{h}}_{|\Lambda}(n+1)$ located at the positions determined by the index set:

$$\Lambda_S = max\left(\left|\hat{\mathbf{h}}_{|\Lambda}(n+1)\right|, S\right) \tag{24}$$

while the remaining $N - S$ coefficients are set to zero, i.e.

$$\hat{\mathbf{h}}_{|\Lambda_S^c}(n+1) = \mathbf{0}_{(N-S)\times 1} \tag{25}$$

where Λ_S^c represents the complement of set Λ_S.

Finally, the equalizer filters (FF and FB) are computed from the estimated nonzero channel taps $\hat{\mathbf{h}}_{|\Lambda_S}(n+1)$ as described in (6), (7). The proposed scheme is summarized in the upper part of Table 1.

B. Scheme II: Sparse Adaptive DFE

The output of the equlizer at ech time n is defined as in (4).The equalizer coefficients $\mathbf{w}(n)$ can be obtained as the solution of the following minimization problem:

$$\hat{\mathbf{w}}(n) = arg \min_{\mathbf{w}(n)} \sum_{k=1}^{n} \lambda^{n-k} \left|d(k) - \mathbf{w}^H(n)\mathbf{y}(k)\right|^2 \tag{26}$$

where $d(k)$, $\mathbf{w}(n)$ and $\mathbf{y}(k)$ correspond to the DFE output, coefficients and input defined in eqs. (4),(5).

As mentioned previously, for a sparse multipath channel the FF and FB filters of the corresponding MMSE DFE are also sparse [1]. Thus, based on (26), the greedy algorithm proposed in the previous subsection can also be applied for direct sparse channel equalization. Similarly to the proposed SD scheme,

TABLE II
SUMMARY OF PROPOSED SCHEMES

	Algorithm description of **Scheme I** *at n-th iteration*
Step 1:	*Update the autocorelation matrix* $\mathbf{\Phi}_{uu}(n)$ *and the crosscorrelation vector* $\mathbf{z}_{xu}(n)$ $\mathbf{\Phi}_{uu}(n) = \lambda\mathbf{\Phi}_{uu}(n-1) + \mathbf{u}(n)\mathbf{u}^H(n)$ $\mathbf{z}_{xu}(n) = \lambda\mathbf{z}_{xu}(n-1) + \mathbf{u}(n)x^*(n)$
Step 2:	*Estimate gradient vector* $\mathbf{g}(n)$ *via eq.* (19)
Step 3:	*Identify the* $2S$ *large components and Merge support sets as in eqs.* (20), (21)
Step 4:	*Perform an SD update by using eqs.*(22),(23)
Step 5:	*Prune support set and impose zeros* (24), (25)
Step 6:	*Estimate non-zero FF, FB taps from the estimated sparse CIR as in eqs.* (6),(7)
Step 7:	*Estimate the decision feedback equalizer output from eq.* (4)

	Algorithm description of **Scheme II** *at n-th iteration*
Step 1:	*Update the autocorelation matrix* $\mathbf{\Phi}_{yy}(n)$ *and the crosscorrelation vector* $\mathbf{z}_{dy}(n)$ *as in Scheme I* $\mathbf{\Phi}_{yy}(n) = \lambda\mathbf{\Phi}_{yy}(n-1) + \mathbf{y}(n)\mathbf{y}^H(n)$, $\mathbf{z}_{dy}(n) = \lambda\mathbf{z}_{dy}(n-1) + \mathbf{y}(n)d^*(n)$
Step 2:	*Estimate gradient vector* $\mathbf{g}(n)$ *as in eqs.* (19)
Step 3:	*Identify the* $2S'$ *large components of* $\mathbf{w}(n)$ *and Merge support sets as in eqs.* (20), (21)
Step 4:	*Perform an SD Equalizer Update* $\alpha(n) = \frac{\mathbf{g}_{\vert\Lambda'}(n)^H \mathbf{g}_{\vert\Lambda'}(n)}{\mathbf{g}_{\vert\Lambda'}(n)^H \mathbf{\Phi}_{yy,\vert\Lambda'}(n)\mathbf{g}_{\vert\Lambda'}(n)}$ $\mathbf{w}_{\vert\Lambda'}(n) = \mathbf{w}_{\vert\Lambda'}(n-1) + \alpha(n)\mathbf{g}_{\vert\Lambda'}(n)$
Step 5:	*Prune sparse equalizer by imposing zeros*
Step 6:	*Estimate the decision feedback equalizer output from eq.* (4)

we initially select a subset of the equalizer coefficients $\mathbf{w}(n)$ (by forming the proxy signal and selecting its $2S'$ largest components*) and solve the resulting constrained LS problem. Provided that all previous decisions are correct, the non-zero taps of the equalizer $\mathbf{w}_{|\Lambda'}(n)$ can be computed as the minimizer of the cost function $J\left(\mathbf{w}_{|\Lambda'}(n), \mathbf{\Phi}_{yy,|\Lambda'}(n), \mathbf{z}_{dy,|\Lambda'}(n)\right)$ with respect to $\mathbf{w}_{|\Lambda'}(n)$.† $\mathbf{\Phi}_{yy,|\Lambda'}(n)$ is a $Q' \times Q'$ matrix and $\mathbf{z}_{dy,|\Lambda'}(n)$ a $Q' \times 1$ vector derived from the exponentially time-averaged input data autocorrelation matrix $\mathbf{\Phi}_{yy}(n)$ and the crosscorrelation vector $\mathbf{z}_{dy}(n)$ respectively, with $2S' \leq Q' \leq 3S'$. Finally, the minimizers are approximated by performing an SD step (see step 4 of Scheme II, in Table 1). The new adaptive equalization algorithm is summarized in the lower part of Table II.

C. Performance of schemes I,II

The convergence properties of Schemes I and II depend on the correct recovery of the support set of the sparse signal. According the CoSaMP algorithm, the proxy signal acts as an indicator for the non-zero elements of the sparse signal,

*With S' we denote the number of the non negligible taps of the FF and FB filter.

†Note the cost function in (26) is quadratic with respect to $\mathbf{w}$ for every n even under error propagation conditions.

provided that the measurement matrix satisfies the Restricted Isometry Property (RIP) [8].

The RIP property is directly related to the singular values of the measurement matrix. More specifically, a measurement matrix $\mathbf{U}(n)$ satisfies the RIP if the extremal eigenvalues of the Gram matrix $\mathbf{\Phi}(n) = \mathbf{U}(n)\mathbf{U}^H(n)$ lie in the range $(1-\delta_S, 1+\delta_S)$, where $\delta_S \in (0,1)$ and S the sparsity order of the estimated sparse signal. As it will be shown, under some reasonable assumptions the eigenvalues of the input auto-correlation matrices for both of our schemes lie in such a range. So, a natural selection of the proxy signal for both cases will be the cross-correlation vector for the initial step, and the gradient vector for the successive steps of the algorithms.

In case of scheme I, the $N \times 1$ gradient vector (19) assuming high SNR conditions ($\sigma_\eta^2 \to 0$) can be written as :

$$\mathbf{g}_h(n)=\mathbf{U}(n)\mathbf{W}(n)\mathbf{U}(n)^H\big(\mathbf{h}(n)-\hat{\mathbf{h}}(n)\big) \tag{27}$$

where $\mathbf{\Phi}_{uu}(n)$ is the $N \times N$ autocorrelation matrix that can be written as the Gram matrix of $\mathbf{W}(n)^{1/2}\mathbf{U}(n)^H$. In order to correctly recover the support set, the matrix $\mathbf{U}(n)^H$ must satisfy the Exponentially-weighted RIP [7], which is the case when assuming that $\mathbf{U}(n)$ is constructed from i.i.d. Gaussian entries with zero mean and variance $1/N$. Furthermore, in the minimum mean square (MMSE) case the input autocorrelation matrix $\mathbf{\Phi}_{uu}$ is a scaled identity matrix $\frac{1}{1-\lambda}\sigma_u^2\mathbf{I}_N$, with eigenvalues in the desired range.

In case of scheme II, the $(K_f+K_b) \times 1$ gradient vector can be written as :

$$\mathbf{g}_w(n)=\mathbf{\Phi}_{yy}(n)\big(\mathbf{w}(n)-\mathbf{w}(n-1)\big) \tag{28}$$

where $\mathbf{\Phi}_{yy}(n)$ is the $(K_f+K_b)\times(K_f+K_b)$ autocorrelation matrix of the equalizer input. In order to correctly recover the support set, the normalized matrix $\alpha\mathbf{\Phi}_{yy}(n)$ should have eigenvalues bounded in the range $(1-\delta_{S'}, 1+\delta_{S'})$, where $\delta_{S'} \in (0,1)$ and S' is the sparsity order of the equalizer coefficients $\mathbf{w}(n)$.

Lemma 1: The eigenvalues of the input autocorrelation submatrix of the MMSE-DFE are bounded in the range $(c-\rho, c+\rho)$, where $c=\mathcal{O}(1)$, $\rho = S'\big(\mathcal{O}(\epsilon)+S\mathcal{O}(\epsilon^2)\big)+\sigma_\eta^2$ and S' is the sparsity order of the equalizer coefficients $\mathbf{w}(n)$, under the following assumptions :

1) the sparse channel with S non zero coefficients, has amplitude of order $\mathcal{O}(1)$ at the specular path h_0, and $\mathcal{O}(\epsilon)$ at each of the scattered paths, where $\epsilon \in (0,1)$, such that $\sum_{k=1}^{S}|h_k| < |h_0|$
2) the delay of the FF filter is ΔT_s symbol periods, with $\Delta \geq K_f - 1 + n_1$.

Proof: The input autocorrelation matrix of the MMSE DFE can be written as

$$\mathbf{\Phi}_{yy}=\begin{bmatrix}\mathbf{H}\mathbf{H}^H+\sigma_\eta^2\mathbf{I}_{K_f} & \mathbf{H}\mathbf{J}_\Delta\\ \mathbf{J}_\Delta^T\mathbf{H}^H & \mathbf{I}_{K_b}\end{bmatrix} \tag{29}$$

where $\mathbf{I}_{K_b}$ is the $K_b \times K_b$ identity matrix. For the shake of simplicity we will assume that the transmitted symbol energy is equal to 1. Matrix $\mathbf{H}$ is a $K_f \times (K_f+N)$ Toeplitz matrix consisting of the channel coefficients. With $\mathbf{J}_\Delta$ we denote a $(K_f+N) \times K_b$ matrix of 0's and 1's, which has the upper $\Delta+1$ rows zeroed and an identity matrix of dimension $\min(K_b, K_f+N-\Delta-1)$ with zeros to the right (when $K_f+N-\Delta-1 < K_b$), zeros below (when $K_f+N-\Delta-1 > K_b$), or no zeros to the right or below exactly fitting in the bottom of $\mathbf{J}_\Delta$ (when $K_f+N-\Delta-1 = K_b$).

According to the Gershgorin theorem [9], each eigenvalue of $\mathbf{\Phi}_{yy}$ belongs in a circle with center the value of the diagonal entry of the matrix and radius the sum of its off-diagonal entries. To exploit the special structure of matrix $\mathbf{\Phi}_{yy}$, we will examine the off-diagonal entries of the first K_f and the last K_b rows separately.

The upper Toeplitz block of $\mathbf{\Phi}_{yy}$ can be written as $\mathbf{\Phi}_{xx} = \mathbf{H}\mathbf{H}^H + \sigma_\eta^2\mathbf{I}_{K_f} = |h_0|^2\mathbf{I}_{K_f} + (\sigma_\eta^2\mathbf{I}_{K_f}+\mathbf{F}_1)$, where the Hermitian matrix $\mathbf{F}_1$ results from $\mathbf{H}\mathbf{H}^H$ after subtracting the diagonal matrix $|h_0|^2\mathbf{I}_{K_f}$. If with $\mathbf{F}_0$ we denote the $K_f \times (K_f+N)$ Toeplitz matrix which results from $\mathbf{H}$ after subtracting the quantity h_0 from its n_1+1 diagonal, defined as $\mathbf{F}_0 = \mathbf{H} - \mathbf{D}_{n_1+1}$ then it is straightforward to show that the matrix $\mathbf{F}_1 = \mathbf{D}_{n_1+1}\mathbf{F}_0^H + \mathbf{F}_0\mathbf{D}_{n_1+1}^H + \mathbf{F}_0\mathbf{F}_0^H$ consists of elements with amplitude of order $\kappa_1 = \mathcal{O}(\epsilon)+S\mathcal{O}(\epsilon^2)$. Provided that $|\Pi_1|$ is the sparsity order of the FF part of the equalizer, the maximum row sum of the Π_1 columns will be $\|\mathbf{F}_{1,|\Pi_1|}\|_\infty = |\Pi_1|\kappa_1$. By taking into account the second assumption, that $\Delta \geq K_f-1+n_1$, it can be easily verified that the submatrix $\mathbf{\Phi}_{xd} = \mathbf{H}\mathbf{J}_\Delta$, does not contain the channel coefficient h_0. If with Π_2 we denote the set of the non zero coefficients of the FB part of the equalizer, then $\|\mathbf{\Phi}_{xd,\Pi_2}\|_\infty = |\Pi_2|\mathcal{O}(\epsilon)$. Furthermore, considering the off-diagonal entries of the last K_b rows of Φ_{yy}, we have that $\|\mathbf{\Phi}_{xd,\Pi_1}\|_\infty = |\Pi_1|\mathcal{O}(\epsilon)$.

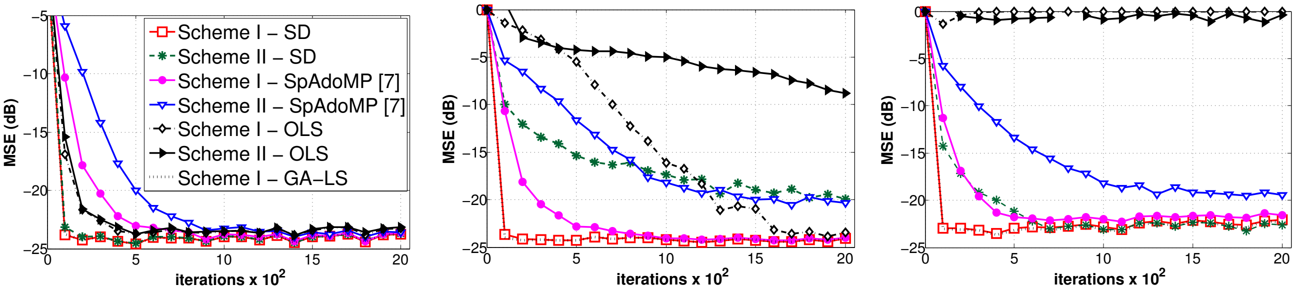
Thus, if $S' = \Pi_1 + \Pi_2$ denotes the set with the non zero coefficients of the equalizer, then the eigenvalues of the $S' \times S'$ submatrix of Φ_{yy} will be bounded in discs with center $c = \mathcal{O}(1)$ and radius $\rho = S'\big(\mathcal{O}(\epsilon)+S\mathcal{O}(\epsilon^2)\big)+\sigma_\eta^2$. ■

When the channel and the equalizer filter are sparse enough, so as $c > \rho$, the eigenvalues of the normalized matrix $\alpha\mathbf{\Phi}_{yy}(n)$ with $\alpha = 1/c$ are bounded the range $(1-\frac{\rho}{c}, 1+\frac{\rho}{c})$ with $\frac{\rho}{c} \in (0,1)$, as posed by RIP statement.

IV. Simulation Results

In this section we present some indicative simulation results of the new equalization schemes. We consider as a test channel a typical terrestrial HDTV CIR[‡] containing 6 multipath components with amplitudes -20, 0, -20, -18, -14, -10 dB, while the corresponding time delays with respect to the main peak are $-20T_s$,$0T_s$, $5T_s$, $20T_s$, $50T_s$, $120T_s$, where T_s is the symbol period. The multipath component phases were chosen randomly. The input sequence consisted of QPSK symbols, while complex white Gaussian noise was added to the channel output resulting in an SNR of 25 dB. The FF and FB filters had a temporal span of $K_f = 30$, and $K_b = 128$ taps, respectively.

[‡]The relevant documents can be found at the Web site of ATSC (Advanced Television Standards Committee), http://www.atsc.org



(a) All Trainning (λ=0.998) (b) Training period: $20T_s$ (λ=0.998) (c) Slow fading channel ($\lambda = 0.98$)

Fig. 1. *Performance of adaptive equalizers at SNR=25dB.*

Initially, to study the convergence of the equalizers, the Doppler effect was ignored and the channel was kept static for an interval of $4096T_s$. Seven different DFE algorithms were tested: (1,2) Schemes I,II based on the novel greedy SD algorithm, (3,4) Schemes I, II based on the greedy algorithm proposed in [7], (5,6) Schemes I,II based on the Ordinary LS (OLS) algorithm, (7) the genie-aided (GA) LS which is the support aware LS estimator applied to a reduced equalization model after removing the regressors corresponding to the zero entries of the channel. Fig. 1 (a) shows that Schemes I have superior performance than Schemes II. This is so because in the former case, fewer non-zero coefficients are adaptively estimated resulting in faster convergence rate. The convergence of the novel greedy SD algorithm is very close to that of the optimal GA-LS for both schemes and is significantly faster than that of the OLS and SpAdOMP algorithms.

Error propagation effects in decision-directed mode were studied by simulating a system that operates over a static channel and employing a training period of only $20T_s$. As observed from Fig. 1 (b) channel estimation based equalizers (Scheme I) are more robust to error propagation effects, while the convergence speed of direct equalization schemes (Scheme II) is seriously affected. Moreover, the non-sparsity aware algorithm OLS converges slowly, especially for Scheme II.

The tracking performance of the algorithms was tested by simulating a system that operates over a modified HDTV channel, where the first and second multipath casual components vary according to the autoregressive model $h_i(n) = \alpha h_i(n-1) + \sqrt{1-|\alpha|^2} v(n)$ [10] where $\alpha = J_o(2\pi f_D T_s)$, $J_o(\cdot)$ is the zero-th order Bessel function and $v(n)$ denotes a white noise process with unit variance and a normalized Doppler frequency $f_D T_s$=1.1 10^{-5} was simulated. The support set of the sparse channel also undergoes gradual changes. The second multipath component in the time delay $20T_s$, is gradually reduced and is nullified after 1000 symbol periods while a new component with increasing amplitude gradually appears in the time delay $5T_s$ with respect to the main peak of the CIR. After 1000 symbol periods the amplitude of the new multipath component is equal to -5dB. Fig. 1 (c) shows the performance of the DFE algorithms when the parameters changes dynamically in time. Our schemes exhibit improved tracking capabilities as compared to [7]. On the other hand, OLS could not track the changes of the channel, given that the training period is only $20T_s$. Finally, the SER performance of the proposed equalizers operating over slow fading frequency selective channels at different SNR conditions, showed that the proposed schemes have superior performance as compared to the algorithm proposed in [7] and the OLS schemes.

V. Conclusion

In this paper two new adaptive DFE schemes for channels with long and sparse impulse responses have been developed. These schemes improve significantly the performance of a conventional DFE using the principles of underlying greedy algorithms. The first scheme that performs channel estimation based equalization exhibits a superior performance as compared to the other one which performs direct equalization. The noticeable fast convergence of this scheme allows the use of shorter training sequence in the applications of interest. Future research is focused on adaptive greedy algorithms without prior knowledge on the sparsity order.

References

[1] A. Rontogiannis and K. Berberidis, "Efficient decision feedback equalization for sparse wireless channels," *IEEE Transactions on Wireless Communications*, vol. 2, no. 3, pp. 570–581, May 2003.

[2] D. Needell and J. A. Tropp, "Cosamp: Iteratives signal recovery from incomplete and inaccurate samples," California Institute of Technology, Pasadena, Tech. Rep., 2008.

[3] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[4] C. Yilun, Y. Gu, and A. O. Hero, "Sparse lms for system identification," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3125–3128.

[5] D. Angelosante and G. B. Giannakis, "Rls-weighted lasso for adaptive estimation of sparse signals," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3245–3248.

[6] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Sparls: the sparse rls algorithm," *IEEE Transactions on Singal Processing*, vol. 58, pp. 4013–4025, August 2010.

[7] G. Mileounis, B. Babadi, N. Kalouptsidis, and V. Tarokh, "An adaptive greedy algorithm with application to nonlinear communications," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 2998–3007, Jun 2010.

[8] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[9] R. S. Varga, *Gershgorin and His Circles*. ser. Springer Series in Computational Mathematics. Berlin, Germany: Springer-Verlag, 2004, no. 36.

[10] A. H. Sayed, *Fundamentals of Adaptive Filtering*. NY: Wiley, 2003.