Stochastic Gradient Pursuit for Adaptive Equalization of Sparse Multipath Channels

Evangelos Vlachos, *Student Member, IEEE*, Aris S. Lalos, and Kostas Berberidis, *Senior Member, IEEE*

***Abstract*—In this paper, a new heuristic algorithm for the sparse adaptive equalization problem, termed as stochastic gradient pursuit, is proposed. A decision-feedback equalization structure is used in order to effectively mitigate the effect of long multipath channels. Diverging from the commonly used approach of sparse channel identification, we exploit the sparsity of the inverse problem under the compressive sensing perspective. Also, an extension to the case where the sparsity order parameter is unknown, is developed. Simulation results verify that the proposed schemes exhibit faster convergence and improved tracking capabilities compared to conventional and other sparse aware equalization schemes, offering at the same time a reduced computational complexity.**

***Index Terms*—Adaptive equalization, compressive sensing, matching pursuit, sparse equalizer, sparse multipath channel.**

I. Introduction

SEVERAL communication systems involve multipath channels which are characterized by long impulse responses, consisting of a few dominant components, some of which may have quite large time delays with respect to the main signal. Some typical examples of this kind are high-definition television (HDTV) [1], broadband mobile network [2], and underwater communication systems [3]. In such cases, conventional equalization methods [4], which are often used in single-carrier transmission systems, should employ very long adaptive equalizers at the receiver's end in order to mitigate effectively the inter-symbol interference (ISI).

However, in applications of the type described above, the implementation of an equalizer becomes a difficult task for two main reasons. Due to the small intersymbol interval, the time available for real time computations is limited. In addition, due to the long span of the introduced ISI, the equalizer must have a large number of taps, which implies heavy computational load per iteration and requirement of longer training sequences. A means to reduce complexity would be to develop sparse equalization methods, where only a small subset of the filter taps is selected to be nonzero, avoiding the computation of the whole equalizer tap vector. An additional motivation to develop sparse equalizers is that in case of sparse multipath channels, where it can be proved that the minimum mean-square error (MMSE) linear or decision-feedback equalizer (DFE) filters are inherently sparse.

During the last decades there have been proposed many different approaches towards developing efficient linear and decision-feedback equalizers, characterized by long impulse responses and few dominant components. A simple approach for the implementation of a sparse equalizer is that of thresholding, [5]–[7]. The nonzero filter taps are estimated based on the strongest taps of the linear or decision-feedback MMSE equalizer. Nevertheless, this approach demands the calculation of the whole equalization vector in each case, which results into schemes with increased computational complexity. On the other hand, two optimum methods for sparse DFEs have been reported in [8] and [9], where the filter tap spacings are jointly optimized with the taps weights. However, both of them are not suitable for real-time execution, due to heavy computational loads. Suboptimum schemes have also been proposed in [10], [11], where the allocation of the nonzero taps is based on the tap positions of infinite length equalizers.

The exploitation of multipath channel sparsity has been recently investigated in the context of system identification for single-carrier [12], [13] and for multi-carrier systems [14]. In [12], a modified least-squares-based structural detection technique is proposed, in order to exploit sparse channel structure and provide improved estimation performance. A conversion procedure that turns the matching pursuit algorithm into an adaptive scheme for sparse system identification, has been proposed in [13], while in [14] the channel is identified in semi-blind fashion using a rough estimation of the nonzero taps in conjunction with the pilot symbols. Furthermore, adaptive equalization schemes that exploit the sparsity of the channel have been reported in [15], where the basic matching pursuit method is used for an adaptive estimation of the channel coefficients. An adaptive DFE algorithm that requires the location of the multipath components of the channel is proposed in [16].

To the best of our knowledge, very little work has been done for the sparse equalization problem [17], [18], viewed as a sparse inverse problem and not based on sparse channel identification. Recently, in parallel to our work, a new design framework for sparse equalizers has been proposed for MMSE linear and decision-feedback equalizers [19]. However, the framework proposed there, was targeting to nonadaptive equalization of a time-invariant channel. In this paper, we investigate the *sparse adaptive equalization* problem viewed from the compressive sensing (CS) perspective [20], [21], where the sensing matrix (i.e., the correlation matrix of the equalizer input

Manuscript received March 02, 2012; revised July 03, 2012; accepted August 02, 2012. Date of publication September 14, 2012; date of current version December 05, 2012. This work was supported in part by THALIS-ENDECON (MIS 376996) research project and by the University of Patras. This paper was recommended by Guest Editor R. Rovatti.

The authors are with the Department of Computer Engineering and Informatics and Computer Technology Institute RU-8, University of Patras, 26500 Patras, Greece (e-mail: vlaxose@ceid.upatras.gr; lalos@ceid.upatras.gr; berberid@ceid.upatras.gr).

Digital Object Identifier 10.1109/JETCAS.2012.2214631

sequence) is updated at each time instant. Within this context, we propose a low-complexity heuristic iterative scheme based on the matching pursuit (MP) algorithm [22] and inexact line search optimization, termed as *stochastic gradient pursuit* (SGP).

The basic MP algorithm [22] offers efficient implementations but suffers from slow convergence. Other variants, such as orthogonal MP (OMP) [23], have superior performance but increased computational complexity due to the orthogonalization step. In order to avoid the cost of OMP but preserve the fast convergence rate, a line search optimization strategy can be adopted [24]. In that case, the estimation of the sparse signal is obtained by making successive steps along a descent direction, which can be computed by various methods, such as gradient descent (GD), conjugate gradient or Newton's method. Due to reasons which will be analyzed in a subsequent section, in our work we present only the GD case, although extensions to the other methods are also possible.

It has been verified through simulations that the proposed scheme exhibits faster convergence and improved tracking capabilities compared to conventional and other sparse aware equalization schemes, offering at the same time a reduced complexity. Furthermore, we develop an extension of SGP where the order of the support set (i.e., the set with the indexes of the nonzero coefficients) of the sparse equalization vector is not known *a priori* but it is recovered in an adaptive manner. Also, the recoverability properties of the proposed greedy scheme are studied theoretically.

The paper is outlined as follows. Section II introduces the signal model, the conventional MMSE linear and decision-feedback equalizer, the concepts and terminology related with CS theory, and comments regarding the sparsity of the inverse filter. In Section III, the sparse adaptive equalization problem is formulated for the cases of known and unknown channels. In Section IV the proposed scheme is presented, with an extension to the varying sparsity order case. In Section V, the performance of the proposed algorithms is investigated, regarding recoverability and complexity. Simulation results are presented in Section VI, followed by conclusions in Section VII.

Notation: Lower-(upper-)case boldface letters are reserved for column vectors (matrixes); $\mathbf{A}_{|\Omega}$ denotes the submatrix with columns of $\mathbf{A}$ based on the index set Ω; $\mathbf{x}_{|\Omega}$ denotes the subvector with elements of $\mathbf{x}$ based on the index set Ω; $(\cdot)^H, (\cdot)^*$ denote the matrix complex conjugate transpose, complex conjugate respectively; $\|\cdot\|_p$ denotes the ℓ_p-norm; $\delta(\cdot)$ denotes the Kronecker delta function.

II. Preliminaries

In this section, first we formulate the signal model and the MMSE linear and decision-feedback equalizers, followed by an overview of concepts and terminology related with CS theory. Finally, we present the conditions under which the inverse filter is sparse.

A. System Model

We consider a communication system operating over a multipath, time-invariant and frequency-selective channel that has more than one physical paths from the transmitter to the receiver. Multipath components tend to be distributed in clusters rather than uniformly over the delay spread, as a physical result of large-scale objects in the scattering environment, while the components within a cluster arise from scattering to small-scale structures [25]. In that case, the channel impulse response (CIR) $h(t)$ can be modeled as

$$h(t) = \sum_{i=0}^{L} h_i \delta(t - \tau_i) \tag{1}$$

where $h_i \in \mathbb{C}$ is the complex gain of the ith path, $\tau_i \in [0, T_m]$ its respective delay with T_m be the maximum delay spread, and L is the number of the resolvable physical paths, which depend on the operating bandwidth W of the communication system. The continuous channel model (1) can be approximated by a discrete counterpart as the sum of gains of all paths

$$h(n) = \sum_{i=0}^{L} h_i \delta(n - iT_s) \tag{2}$$

where T_s is the symbol duration. When the channel has only a few components with nonnegligible gains, compared with the total number of L, then it can be characterized as sparse. We formalize the notion of sparse multipath channel as follows.

Definition 1 (Sparse Multipath Channel): Let a frequency-selective channel $\mathbf{h}$ with CIR expressed in vector form as

$$\mathbf{h} = [h_{-N_1} \ \ldots \ h_0 \ \ldots \ h_{N_2}]^T \tag{3}$$

where samples with positive time indexes are the postcursor taps and samples with negative time indexes are the precursor ones, and $L = N_1 + N_2 + 1$ is the number of the resolvable physical paths within the channel delay spread. We say that channel $\mathbf{h}$ is S-sparse iff $\|\mathbf{h}\|_0 = S \ll L$.

It is worth mentioning here that, even in the best of scenarios, real-world multipath channels can never be exactly S-sparse due to many reasons. Furthermore, if we consider the overall CIR, i.e., the convolved CIR with the transmitter and receiver filters, the resulting vector has some additional nonzero terms. However, in many cases, they can be well approximated as *compressible signals*, i.e., signals with rapidly decaying components when sorted by magnitude.

Next we define the discrete-time system model of a baseband single-carrier communication system, operating at the symbol-rate. Let $s(n)$ be an i.i.d. symbol sequence with variance σ_s^2 and $\eta(n)$ be the complex addictive white Gaussian noise (AWGN) sequence drawn from Gaussian distribution with $\mathcal{N}(0, \sigma_\eta^2)$. The input–output relation for the symbol-spaced case has the following form:

$$x(n) = \sum_{i=0}^{L} h_i s(n - iT_s) + \eta(n). \tag{4}$$

Considering a block of K_f output samples, and assuming that the channel is time-invariant over this block, (4) can be expressed in matrix form as follows:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \boldsymbol{\eta}(n) \tag{5}$$

where the vector with the channel output samples is defined as

$$\mathbf{x}(n) = [x(n)\ x(n-1)\ \ldots\ x(n-K_f+1)]^T \quad (6)$$

the input vector is given by

$$\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-K_f-L+2)]^T \quad (7)$$

the vector with the noise samples is defined as

$$\boldsymbol{\eta}(n) = [\eta(n)\ \eta(n-1)\ \ldots\ \eta(n-K_f+1)]^T \quad (8)$$

and $\mathbf{H}$ is the $K_f \times (K_f+L-1)$ channel matrix with Toeplitz structure.

B. MMSE Decision-Feedback Equalizer

Recall that the MMSE time-domain decision-feedback equalizer consists of two filters, the feedforward filter (FFF) and the feedback filter (FBF). The FFF cascaded with the channel results in a minimum-phase overall system. The FBF cancels the causal ISI using previous decisions.

Next, we describe the MMSE DFE with filter weights' vector denoted by $\mathbf{w} = [\mathbf{w}_{FF}^T\ \mathbf{w}_{FB}^T]^T$, which consists of the FFF and FBF of temporal span K_f and K_b taps, respectively, given by

$$\mathbf{w}_{FF} = \left[w_0\ \ldots\ w_{K_f-1}\right]^T \quad (9)$$

$$\mathbf{w}_{FB} = \left[w_{K_f}\ w_{K_f+1}\ \ldots\ w_{K_f+K_b}\right]^T. \quad (10)$$

Let $\{x(n)\}$ be the input sequence to the equalizer and $\{d(n)\}$ the output of the decision device. The input to the FF filter can be described in vector form by (6) and the input to the FB filter can be expressed as the $K_b \times 1$ vector

$$\mathbf{d}(n-\Delta-1) = [d(n-\Delta-1)\ \ldots\ d(n-\Delta-K_b)]^T \quad (11)$$

where Δ is the causality delay. The MMSE DFE can be obtained by minimizing the following cost function:

$$\mathcal{J} = E\left\{\left|s(n-\Delta) - \mathbf{w}^H\mathbf{y}(n)\right|^2\right\} \quad (12)$$

where $\mathbf{y}(n)$ contains the concatenated FF and FB equalizer input vectors

$$\mathbf{y}(n) = \left[\mathbf{x}(n)^T\ \mathbf{d}(n-\Delta-1)^T\right]^T. \quad (13)$$

Expanding (12), we can write

$$\mathcal{J} = \sigma_s^2 - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r} + (\mathbf{w}-\mathbf{R}^{-1}\mathbf{r})^H\mathbf{R}(\mathbf{w}-\mathbf{R}^{-1}\mathbf{r}) \quad (14)$$

where $\mathbf{R} = E\{\mathbf{y}(n)\mathbf{y}^H(n)\}$ is the $K \times K$ autocorrelation matrix of the equalizer input vector, and $\mathbf{r} = E\{\mathbf{y}(n)s^*(n-\Delta)\}$ is the $K \times 1$ cross-correlation vector of the equalizer input and the desired output, where $K = K_f + K_b$. The optimum mean square error (MSE) solution of (12) can be expressed as the solution of the following system of equations:

$$\mathbf{R}\mathbf{w}_o = \mathbf{r}. \quad (15)$$

The above system can be partitioned as

$$\begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xd} \\ \mathbf{R}_{xd}^H & \sigma_s^2\mathbf{I}_{K_b} \end{bmatrix} \mathbf{w}_o = \begin{bmatrix} \mathbf{r}_{xs} \\ \mathbf{0} \end{bmatrix} \quad (16)$$

where $\mathbf{I}_{K_b}$ is the $K_b \times K_b$ identity matrix, $\mathbf{R}_{xd} = E\{\mathbf{x}(n)\mathbf{d}^H(n-\Delta-1)\}$ is a $K_f \times K_b$ cross-correlation matrix, $\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ is the $K_f \times K_f$ autocorrelation matrix, and $\mathbf{r}_{xs} = E\{\mathbf{x}(n)s^*(n-\Delta)\}$ is the $K_f \times 1$ cross-correlation vector. Furthermore, we can express these correlation quantities based on channel coefficients as follows [4]:

$$\mathbf{R} = \begin{bmatrix} \mathbf{H}\mathbf{H}^H + \sigma_\eta^2\mathbf{I}_{K_f} & \mathbf{H}\mathbf{J}_\Delta \\ \mathbf{J}_\Delta^H\mathbf{H}^H & \sigma_s^2\mathbf{I}_{K_b} \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \mathbf{H}_{|\{\Delta\}} \\ \mathbf{0} \end{bmatrix} \quad (17)$$

where $\mathbf{J}_\Delta = [\mathbf{0}_{K_b\times(\Delta+1)}\ \mathbf{I}_{K_b}\ \mathbf{0}_{K_b\times(K_f+L-K_b-2-\Delta)}]$. Note that, the equations to determine $\mathbf{w}_o$ for the linear equalizer are similar to (16), with $K_b = 0$.

The problem of (15) has a unique solution at $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{r}$, where the cost function of (14) is minimized

$$\mathcal{J}_{\min} = \sigma_s^2 - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}. \quad (18)$$

For any other choice of $\mathbf{w} \neq \mathbf{w}_o$, the cost function is increased by an error, called excess MSE, $\mathcal{J}_e$, so (14) can be written as

$$\mathcal{J} = \mathcal{J}_{\min} + \mathcal{J}_e. \quad (19)$$

C. Compressive Sensing Framework

CS provides a way of obtaining a compressed version of a signal $\mathbf{c} \in \mathbb{C}^K$ using only a small number of linearly combined measurements. The measurements $\mathbf{q} \in \mathbb{C}^m$, are acquired using a sampling matrix $\boldsymbol{\Phi} \in \mathbb{C}^{m\times K}$ as follows:

$$\mathbf{q} = \boldsymbol{\Phi}\mathbf{c}. \quad (20)$$

A sparse approximation $\hat{\mathbf{c}}$ to $\mathbf{c}$ can be obtained by solving the following constrained optimization problem:

$$\min_{\hat{\mathbf{c}}} \|\hat{\mathbf{c}}\|_0 \quad \text{subject to} \quad \|\mathbf{q} - \boldsymbol{\Phi}\hat{\mathbf{c}}\|_2^2 \leq \epsilon \quad (21)$$

where ϵ is a predefined error tolerance. The above optimization problem is computationally intractable, and hence cannot be used for practical applications. Two major classes for the approximation of (21) are convex relaxation methods and heuristic/greedy algorithms.

In convex relaxation methods, the ℓ_0 quasi-norm is replaced by the convex ℓ_1-norm

$$\min_{\hat{\mathbf{c}}} \|\hat{\mathbf{c}}\|_1 \quad \text{subject to} \quad \|\mathbf{q} - \boldsymbol{\Phi}\hat{\mathbf{c}}\|_2^2 \leq \epsilon \quad (22)$$

where the ℓ_1-norm, for the complex norm space $\mathbb{C}^K$, is defined as $\|\mathbf{c}\|_1 = \sum_{i=1}^K(|\Re\{x_i\}| + |\Im\{x_i\}|)$. The main advantage of the ℓ_1-minimization approach is that it is a convex optimization problem that can be solved efficiently by linear programming techniques.

Greedy algorithms, on the other hand, compute iteratively the signal's support set until a halting condition is met. In this approach the combinatorial problem is circumvented by heuristically choosing which values of $\hat{\mathbf{c}}$ are nonzero, setting the others to zero and solving the resulting subproblem.

In the context of CS, the reconstruction performance for both ℓ_1-minimization and greedy methods, can be described in terms of the so-called restricted isometry property (RIP), which is defined as follows.

Definition 2 (Restricted Isometry Property): A matrix $\mathbf{\Phi} \in \mathbb{C}^{m\times K}$ is said to satisfy the RIP, with order S and parameter $\delta_S \in (0,1)$, if

$$(1-\delta_S)\|\mathbf{c}\|_2^2 \leq \|\mathbf{\Phi c}\|_2^2 \leq (1+\delta_S)\|\mathbf{c}\|_2^2 \tag{23}$$

holds for all S-sparse vectors $\mathbf{c} \in \mathbb{C}^K$.

An alternative way to interpret the RIP of the matrix $\mathbf{\Phi}$, with order S and constant δ_S, is by bounding the eigenvalue spread of the Gram matrix of $\mathbf{\Phi}$, namely $\mathbf{\Phi}^H\mathbf{\Phi}$, according to the following definition.

Definition 3 (Restricted Isometry Property and Eigenvalues): A matrix $\mathbf{\Phi} \in \mathbb{C}^{m\times K}$ is said to satisfy the RIP, with order S and parameter $\delta_S \in (0,1)$, when for all $\Omega \subseteq \{1,\ldots,N\}$ such that $|\Omega| \leq S$, it holds that

$$(1-\delta_S) \leq \lambda_{\min}\left(\mathbf{\Phi}_{|\Omega}^H\mathbf{\Phi}_{|\Omega}\right) \leq \lambda_{\max}\left(\mathbf{\Phi}_{|\Omega}^H\mathbf{\Phi}_{|\Omega}\right) \leq (1+\delta_S) \tag{24}$$

where $\lambda_{\min}(\mathbf{\Phi}_{|\Omega}^H\mathbf{\Phi}_{|\Omega})$ and $\lambda_{\max}(\mathbf{\Phi}_{|\Omega}^H\mathbf{\Phi}_{|\Omega})$ denote the minimal and the maximal eigenvalues of $\mathbf{\Phi}_{|\Omega}^H\mathbf{\Phi}_{|\Omega}$ respectively.

D. Sparsity of the Inverse Problem

Sparsity of the channel response does not, in general, imply sparsity of the inverse system, i.e., the equalizer in our case. In this section, we investigate the necessary conditions in order the inverse filter to be inherently or approximately sparse. The overall discrete time CIR, which is the convolution of the CIR, $h(n)$, with the impulse responses of the transmit and receive filters, can be expressed as follows:

$$g(n) = g_0 + \sum_{i=1}^{M} g_i\delta(n-\tau_i) \tag{25}$$

with $M \geq L$. Let the z-transform $\mathcal{Z}\{\cdot\}$ [26] of the overall CIR be written as

$$\mathcal{Z}\{g\} \equiv G(z) = g_0 + \sum_{i=1}^{M} g_i z^{-\tau_i}. \tag{26}$$

In [27], it has been proved that given that the energy of the specular path of the channel is larger than the power spectral density of all scattered paths, i.e.,

$$|g_0|^2 > \left|\sum_{i=1}^{M} g_i z^{-\tau_i}\right|^2 \tag{27}$$

then the CIR is characterized as a minimum-phase sequence [26]. Following the approach in [16] and given that (27) is satisfied, the inverse filter $G^{-1}(z)$ can be expressed as a geometric series

$$G^{-1}(z) = g_0^{-1}\sum_{k=0}^{\infty}(-1)^k\left[\sum_{i=1}^{M}\frac{g_i}{g_0}z^{-\tau_i}\right]^k. \tag{28}$$

Expanding the kth power in (28) we obtain terms proportional to exponentials $z^{-\tau_1-\cdots-\tau_k}$. Hence, taking the inverse z-transform $\mathcal{Z}^{-1}\{\cdot\}$, the infinite-impulse response (IIR) inverse filter contains Kronecker delta functions $\delta(n-\tau_1-\ldots-\tau_k)$, at positions given by the nonnegative-integer-based linear combinations of the positions of the time-shifted components of $g(n)$. A finite-impulse response (FIR) approximation of (28) can be obtained by setting an upper value for the index k. For example, by taking the second-order approximation, with $k \leq 2$, and computing the inverse z-transform we have

$$\mathcal{Z}^{-1}\left\{G^{-1}(z)\right\} \approx g_0^{-1} - g_0^{-2}\sum_{i=-M}^{M} g_i\delta(n-iT_s) + g_0^{-3}\sum_{i=-M}^{M}\sum_{j=-M}^{M} g_i g_j\delta(n-iT_s+jT_s). \tag{29}$$

The above expression indicates the candidate positions of the nonzero terms of the FFF filter. Indeed, by inspecting (29) and given that (27) is satisfied, it can be easily shown that for $k \geq 3$, the additional terms of the kth-order approximation are almost zero.

The FB equalizer filter, in turn, has nonzero components at tap positions that match with the multipath delays of $g(n)$ [10], hence it is characterized as a sparse vector, given that the overall CIR is sparse.

Nevertheless, the CIR of real-world multipath channels is likely to be a mixed-phase sequence, due to phenomena such as Doppler spread and deep fading. Hence, even in case the FBF is a compressible vector, the FFF will probably be a full vector. In such cases, the sparsity order could not be considered as *a priori* known or a static parameter, but it must be adaptively adjusted.

III. Sparse Adaptive Equalization

In this section, we formulate the sparse equalization framework for the design of sparse adaptive decision-feedback equalizers. We describe only the DFE case, as the linear equalizer can be viewed as a special case.

A. Design Based on Known CIR

The conventional MMSE decision-feedback equalizer results from the minimization of the cost function of (12). To obtain an adaptive scheme would demand the minimization of a time-varying cost function, namely

$$\mathcal{J}(\mathbf{w}(n)) = \sigma_s^2 - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r} + \left(\mathbf{w}(n)-\mathbf{R}^{-1}\mathbf{r}\right)^H\mathbf{R}\left(\mathbf{w}(n)-\mathbf{R}^{-1}\mathbf{r}\right) \tag{30}$$

where $\mathbf{R}$, $\mathbf{r}$ are as defined in (17), and $\mathbf{w}(n)$ consists of time-updated FF and FB filters. The third term of the right hand side of (30) represents the excess MSE of the nth adaptation step.

Relaxing the MMSE optimality and favouring sparsity we can formulate the following DFE optimization problem:

$$\min_{\mathbf{w}(n)\in\mathbb{C}^K} \|\mathbf{w}(n)\|_0 \quad \text{subject to} \quad \mathcal{J}(\mathbf{w}(n)) \leq \epsilon. \tag{31}$$

In this work, we consider a heuristic approach in order to approximate the solution of (31). Under this perspective, at time instant n, first we identify the set Ω^n with the nonzero terms of $\mathbf{w}(n)$ (i.e., the support set) and then we seek for the vector $\mathbf{w}_{|\Omega^n}(n)$ which minimizes the cost function $\mathcal{J}(\mathbf{w}_{|\Omega^n}(n))$, namely

$$\min_{\mathbf{w}_{|\Omega^n}(n)\in\mathbb{C}^K} \mathcal{J}\left(\mathbf{w}_{|\Omega^n}(n)\right). \tag{32}$$

Given that the submatrix $\mathbf{R}_{|\Omega^n}$ has a left inverse, there is a unique solution of (32), expressed as the solution of the following linear system of equations:

$$\mathbf{R}_{|\Omega^n}\mathbf{w}_{|\Omega^n}(n) = \mathbf{r}. \tag{33}$$

B. Design Based on Time-Averaged Correlation Sequences

In cases where the CIR is unknown or time-varying, we estimate the autocorrelation matrix $\mathbf{R}(n)$ and cross-correlation vector $\mathbf{r}(n)$ using an exponentially time-averaged window, as follows:

$$\mathbf{R}(n) = \lambda\mathbf{R}(n-1) + \mathbf{y}(n)\mathbf{y}(n)^H \tag{34}$$

$$\mathbf{r}(n) = \lambda\mathbf{r}(n-1) + \mathbf{y}(n)s^*(n) \tag{35}$$

where λ is the forgetting factor with $\lambda \in (0, 1]$. Starting with the conventional adaptive equalizer, the coefficients $\mathbf{w}(n)$ can be obtained at each time instant, as the minimizers of the following cost function:

$$\begin{aligned}\mathcal{J}&(\mathbf{w}(n), \mathbf{R}(n), \mathbf{r}(n)) \\ &= \sum_{k=1}^{n} \lambda^{n-k}\left|s(k) - \mathbf{w}(n)^H\mathbf{y}(k)\right|^2 \\ &= \beta_s(n) - \mathbf{r}(n)^H\mathbf{w}_o(n) \\ &\quad + (\mathbf{w}(n) - \mathbf{w}_o(n))^H\mathbf{R}(n)(\mathbf{w}(n) - \mathbf{w}_o(n))\end{aligned} \tag{36}$$

where $\mathbf{w}_o(n) = \mathbf{R}(n)^{-1}\mathbf{r}(n)$ and $\beta_s = \sum_{k=1}^{n}\lambda^{n-k}|s(k)|^2$. As previously, we can formulate the sparse adaptive DFE problem

$$\min_{\mathbf{w}(n)\in\mathbb{C}^K} \|\mathbf{w}(n)\|_0 \quad \text{subject to} \quad \mathcal{J}(\mathbf{w}(n), \mathbf{R}(n), \mathbf{r}(n)) \leq \epsilon. \tag{37}$$

Using a heuristic algorithm for the approximation of (37), we end up with the following system of equations:

$$\mathbf{R}_{|\Omega^n}(n)\mathbf{w}_{|\Omega^n}(n) = \mathbf{r}(n) \tag{38}$$

where in this case, as opposed to (33), the correlation quantities are time-varying.

IV. Stochastic Gradient Pursuit

In this section, we derive a new algorithm for the solution of the sparse adaptive equalization problem of (37). It is based on a proper incorporation of an inexact line search strategy into an MP-type heuristic optimization technique. The term *stochastic* is intended to point out the fact that the measurement matrix and vector, appearing in (34) and (35), are not deterministic but stochastic quantities.

In the first part, we describe the general philosophy and the basic structure of an MP-type algorithm, and we introduce the inexact line search method which is used to solve the resulting linear system. In the second part, we extend the proposed scheme to the case were the sparsity order parameter is unknown and should be adaptively recovered.

A. Adaptive Equalization Using Matching Pursuit

In basic MP-type algorithms [22], [23], during each iteration, the algorithm selects one or several indexes that represent good partial support set estimates and then adds them to the current support set estimation. Once an index is included in the support set it remains in this set throughout the remainder of the reconstruction process. On the other hand, recently developed MP-type algorithms, such as subspace pursuit (SP) [28] and compressive sampling matching pursuit (CoSaMP) [29], find an estimate of the support set of a predefined order S which is updated during each iteration. An index can be added to or removed from the estimated support set at any stage of the recovery process. However, in order to update the support set, a larger set with $2S$ for SP and $3S$ for CoSaMP must be updated at each iteration, resulting in increased computational complexity. In our work, we develop a greedy strategy which is based on SP/CoSaMP algorithms, but requires less complexity.

Considering the nth time instant and dropping the time index for simplicity, i.e., $\mathbf{w} \equiv \mathbf{w}(n)$, the problem of (37) can be iteratively solved using an MP-type greedy strategy. At the ith iteration the support set Ω^i is updated selecting the new elements based on the so-called proxy signal. In our case, the *proxy signal* is defined as the correlation between the rows of the measurement matrix $\mathbf{R}$ and the error $\mathbf{e}^{i-1} = \mathbf{w}_o - \mathbf{w}^{i-1}$, namely

$$\mathbf{p}^i \equiv \mathbf{R}\mathbf{e}^{i-1} = \mathbf{r} - \mathbf{R}_{|\Omega^{i-1}}\mathbf{w}^{i-1}_{|\Omega^{i-1}} \tag{39}$$

where $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{r}$ is the least-squares (LS) solution. The ith approximation of the proxy $\mathbf{p}^i$, corresponds to the part of the sparse vector that has not been approximated yet. Let Ω^{i-1} be the set of indexes of the S largest components of the previously estimated vector $\mathbf{w}^{i-1}$. Then we update the support set by replacing the index of the weakest component with the index of the strongest component of the proxy vector

$$\Omega^i = arg_\ell \max\left(\left|w_\ell^{i-1}\right|, S-1\right) \cup arg_\ell \max\left(\left|p_\ell^i\right|, 1\right) \tag{40}$$

where $\max(|x_\ell|, S)$ denotes the S largest elements of the vector $\mathbf{x} \in \mathbb{C}^K$ with $\ell = 1, \ldots, K$. Therefore, the support set, with a predefined order equal to S, is refined by only one component, at each stage of the recovery process. This is adequate for convergence to the true support set, since the measurement matrix has linearly independent columns. Once the ith approximation of the support set has been found, the nonzero values of $\mathbf{w}^i$ are easily estimated via LS as

$$\mathbf{w}_{|\Omega^i} = (\mathbf{R}_{|\Omega^i})^\dagger\mathbf{r} \tag{41}$$

Finally, the components of $\mathbf{w}$ which do not belong to the support set, $\mathbf{w}_{\Omega^c_i}$, are set to zero, performing the pruning step.

The steps which are executed within each time instant n, are summarized below.

Step 1) Update the proxy vector via (39).

Step 2) Identify the index with the largest component of the proxy signal.

Step 3) Merge indexes sets, forming the union of the set indexes corresponding to the $S-1$ largest components of the estimate obtained in the previous iteration with the newly identified index of the proxy signal.

Step 4) Estimate the sparse vector via LS on the merged set of components via (41).

Step 5) Prune the elements which do not belong in the support set.

One way to solve in an adaptive fashion the linear system involved in the fourth step is by using the RLS algorithm. However, this solution is not straightforward because the update rule cannot be directly restricted to the varying support set. Hence, we choose an iterative line search method [30], where successive steps are made in a descent direction towards the minimization of the cost function $\mathcal{J}(\mathbf{w}_{|\Omega}, \mathbf{R}, \mathbf{r})$, where for now Ω is assumed to be known and time-invariant. Specifically, the basic steps of the line search method at the kth iteration are the following.

Step 1) Determine the direction of search $\mathbf{d}_{|\Omega}^k$.

Step 2) Find a^k as the a minimizing

$$\min_a \mathcal{J}\left(\mathbf{w}_{|\Omega}^{k-1} + a\mathbf{d}_{|\Omega}^k\right). \tag{42}$$

Step 3) Make a step at the descent direction

$$\mathbf{w}_{|\Omega}^k = \mathbf{w}_{|\Omega}^{k-1} + \alpha^k \mathbf{d}_{|\Omega}^k. \tag{43}$$

As mentioned previously, there are several methods of choosing $\mathbf{d}_{|\Omega}$, such as gradient descent, conjugate gradient (CG) or Newton method (NM). In general, the CG and NM are characterized by faster convergence compared to the GD algorithm, but with increased computational complexity and storage. However, in our case, due to the dimensionality reduction of the problem, the GD algorithm offers similar convergence rate at a smaller cost, as verified by the simulation results in Section VI. Hence, we employ a GD direction rule, where the direction vector corresponds to the opposite of the gradient vector, which is expressed as

$$\mathbf{d}_{|\Omega}^k \equiv \mathbf{g}^k = -\nabla\mathcal{J}\left(\mathbf{w}_{|\Omega}^k, \mathbf{R}, \mathbf{r}\right). \tag{44}$$

The solution of (42) implies that the current residual is orthogonal to the previous, i.e., $\left(\mathbf{g}^k\right)^H \mathbf{g}^{k-1} = 0$, which leads to the next expression for the step size

$$\alpha^k = \frac{\left(\mathbf{g}^{k-1}\right)^H \mathbf{g}^{k-1}}{\left(\mathbf{g}^{k-1}\right)^H \mathbf{R}\mathbf{g}^{k-1}}. \tag{45}$$

We can now proceed with forming the adaptive SGP scheme, incorporating the line search strategy into the MP-type algorithm. At the nth time step of the adaptive algorithm, the proxy signal can be viewed as the residual of the cost function $\mathcal{J}(\mathbf{w}_{|\Omega^{n-1}}(n-1), \mathbf{R}(n), \mathbf{r}(n))$, namely

$$\mathbf{p}(n) \equiv \mathbf{g}(n) = \mathbf{r}(n) - \mathbf{R}_{|\Omega^{n-1}}(n)\mathbf{w}_{|\Omega^{n-1}}(n-1). \tag{46}$$

TABLE I
STOCHASTIC GRADIENT PURSUIT

Algorithm SGP	Complexity Order
`Step 1: gradient update`	
$\mathbf{g}(n) = \mathbf{r}(n) - \mathbf{R}(n)\bar{\mathbf{w}}(n-1)$	$\mathcal{O}(KS)$
`Step 2: support set update`	
$\Lambda = arg\max(\lvert\bar{\mathbf{w}}(n-1)\rvert, S-1)$	$\mathcal{O}(K)$
$\Omega = \Lambda \cup arg\max(\lvert\mathbf{g}_{\lvert\Lambda^c}(n-1)\rvert, 1)$	$\mathcal{O}(K)$
`Step 3: line search optimization`	
$\alpha(n) = \frac{\mathbf{g}_{\lvert\Omega}(n)^H \mathbf{g}_{\lvert\Omega}(n)}{\mathbf{g}_{\lvert\Omega}(n)^H \mathbf{R}_{\lvert\Omega}(n)\mathbf{g}_{\lvert\Omega}(n)}$	$\mathcal{O}(KS)$
$\mathbf{w}_{\lvert\Omega}(n) = \bar{\mathbf{w}}_{\lvert\Omega}(n-1) + \alpha(n)\mathbf{g}_{\lvert\Omega}(n)$	$\mathcal{O}(1)$
`Step 4: pruning`	
$\bar{\mathbf{w}}_{\lvert\Omega}(n) = \mathbf{w}_{\lvert\Omega}(n)$, $\bar{\mathbf{w}}_{\lvert\Omega^c}(n) = \mathbf{0}$	$\mathcal{O}(1)$

Therefore, the computation of the proxy and the direction vector can be combined into one step. Using the time-updated proxy signal, the identification of the nth support set can be obtained by

$$\Omega^n = arg_\ell \max\left(\lvert w_\ell(n-1)\rvert, S-1\right) \cup arg_\ell \max\left(\lvert g_\ell(n)\rvert, 1\right). \tag{47}$$

In order to avoid the cost of running several iterations, we set the GD algorithm to run just one iteration, i.e., for $k=1$. However, in that case, the step size given by (45) does not orthogonalize the current and the previous residuals, resulting in an approximation to the exact line search solution. This is due to the stochastic nature of quantities $\mathbf{R}, \mathbf{r}$ and the time varying support set. For the proposed scheme, we define the step size at the nth time instant as follows:

$$\alpha(n) = \frac{\left(\mathbf{g}_{|\Omega^n}(n)\right)^H \mathbf{g}_{|\Omega^n}(n)}{\left(\mathbf{g}_{|\Omega^n}(n)\right)^H \mathbf{R}_{|\Omega^n}\mathbf{g}_{|\Omega^n}(n)}. \tag{48}$$

Furthermore, instead of using an approximation as the one that follows from (43), we propose the modified approximation

$$\mathbf{w}_{|\Omega^n}(n) = \bar{\mathbf{w}}_{|\Omega^n}(n-1) + \alpha(n)\mathbf{g}_{|\Omega^n}(n) \tag{49}$$

where $\bar{\mathbf{w}}(n-1)$ is the pruned vector, i.e., the estimated equalization vector of the previous adaptation step $n-1$, where the elements not belonging to the previous set, Ω^{n-1}, have been zeroed. The steps of the SGP algorithm are summarized in Table I, for the nth time iteration.

B. Varying Sparsity Order

As it is the case with other greedy pursuit algorithms, the SGP algorithm requires that the sparsity order S of the equalizer vector is *a priori* known. However, in many practical cases such as in time varying channel equalization, the order S is not available or it may change due to channel variations. Underestimation of S will cause the algorithm to diverge from the optimum filter, whereas overestimation will degrade the convergence speed.

Thus, in this section we derive an extension to SGP, termed as ν-SGP, where the sparsity order ν is unknown. The steps of the algorithm are described at the Table II for the nth time iteration. Beginning with $\nu=1$, the ν-SGP algorithm increases the order of the support set by 1 at each time instant n, when the following condition is true:

$$\xi = \lVert\mathbf{g}_\nu\rVert - \lVert\mathbf{g}_{\nu+1}\rVert \notin (0, \kappa\epsilon) \tag{50}$$

TABLE II
SGP WITH VARYING SPARSITY ORDER

Algorithm ν**-SGP**	Complexity Order
Step 1: gradient update	
$\mathbf{g}_\nu = \mathbf{r}(n) - \mathbf{R}(n)\bar{\mathbf{w}}(n-1)$	$\mathcal{O}(K\nu)$
Step 2: support set update	
$\Lambda = arg\max(\lvert\mathbf{w}(n-1)\rvert, \nu-1)$	$\mathcal{O}(K)$
$\Omega = \Lambda \cup arg\max(\lvert\mathbf{g}_{\lvert\Lambda^c}(n-1)\rvert, 1)$	$\mathcal{O}(K)$
$\gamma = arg\max(\lvert\mathbf{w}_{\Omega^c}(n-1)\rvert, 1)$	
$\Omega_e = \Omega \cup \gamma$	
$\mathbf{g}_{\nu+1} = \mathbf{g}_\nu - \mathbf{R}_{\lvert\gamma}(n)\mathbf{w}_{\lvert\gamma}(n-1)$	$\mathcal{O}(K)$
if $\lVert\mathbf{g}_\nu\rVert - \lVert\mathbf{g}_{\nu+1}\rVert \geq \kappa\epsilon$ then	
$\mathbf{g}(n) = \mathbf{g}_{\nu+1}$, $\nu = \nu+1$	
else $\mathbf{g}(n) = \mathbf{g}_\nu$	
Step 3: line search optimization	
$\alpha(n) = \frac{\mathbf{g}_{\lvert\Omega_e}(n-1)^H\mathbf{g}_{\lvert\Omega_e}(n-1)}{\mathbf{g}_{\lvert\Omega_e}(n-1)^H\mathbf{R}_{\lvert\Omega_e}(n)\mathbf{g}_{\lvert\Omega_e}(n-1)}$	$\mathcal{O}(K\nu)$
$\mathbf{w}_{\lvert\Omega_e}(n) = \bar{\mathbf{w}}_{\lvert\Omega_e}(i-1) + \alpha(n)\mathbf{g}_{\lvert\Omega_e}(n-1)$	$\mathcal{O}(1)$
Step 4: pruning	
$\bar{\mathbf{w}}_{\lvert\Omega}(n) = \mathbf{w}_{\lvert\Omega}(n)$, $\bar{\mathbf{w}}_{\lvert\Omega^c}(n) = \mathbf{0}$	$\mathcal{O}(1)$

where κ depends on the eigenvalue spread of the measurement matrix, ϵ is a predefined parameter, $\mathbf{g}_\nu$ is the gradient of the nth iteration following the definition of (46), and $\mathbf{g}_{\nu+1}$ is the gradient based on augmented support set $\Omega_e = \Omega^n \cup arg\max(\lvert\mathbf{w}_{\Omega^c}(n-1)\rvert, 1)$. Thus, given a predefined error tolerance ϵ, the order of the support set is adaptively selected.

At the steady-state, if $\mathbf{w}$ is the ν-sparse solution of (41) with excess MSE $\mathcal{J}_e(\mathbf{w}) \leq \epsilon$, then $\xi \leq \kappa\epsilon$. This follows from

$$\begin{aligned}\xi &\leq \lVert\mathbf{g}_\nu - \mathbf{g}_{\nu+1}\rVert \leq \lVert\mathbf{R}_{\lvert\Omega^c}\mathbf{w}_{\lvert\Omega^c}\rVert \\ &\leq \lVert\mathbf{R}_{\lvert\Omega^c}\rVert\lVert\mathbf{w}_{\lvert\Omega^c}\rVert \leq \lVert\mathbf{R}\rVert\lVert\mathbf{w}_{\lvert\Omega^c}\rVert.\end{aligned} \tag{51}$$

If the matrix $\mathbf{R}$ has bounded eigenvalues in the range $(1-\delta_S, 1+\delta_S)$, where $\delta_S \in (0,1)$, and $\lVert\mathbf{R}\rVert$ denotes the spectral norm of $\mathbf{R}$, then $\lVert\mathbf{R}\rVert = \sqrt{\lambda_{\max}(\mathbf{R}\mathbf{R}^H)} = 1 + \delta_S$. Note that $\mathbf{R}$ is Hermitian symmetric and nonnegative definite, therefore it holds $\lambda(\mathbf{R}\mathbf{R}^H) = \lambda^2(\mathbf{R})$, for any eigenvalue of $\mathbf{R}$. Hence, from (51) it follows that

$$\xi \leq (1+\delta_S)\lVert\mathbf{w}_{\lvert\Omega^c}\rVert. \tag{52}$$

Moreover, based on the Rayleigh–Ritz inequality [4], which for our case the lower bound of the inequality can be expressed as $(1-\delta_S)\lVert\mathbf{w}_{\lvert\Omega^c}\rVert \leq \mathbf{w}^H_{\lvert\Omega^c}\mathbf{R}_{\lvert\Omega^c}\mathbf{w}_{\lvert\Omega^c}$, we have that

$$\xi \leq \frac{1+\delta_S}{1-\delta_S}\mathbf{w}^H_{\lvert\Omega^c}\mathbf{R}_{\lvert\Omega^c}\mathbf{w}_{\lvert\Omega^c} \leq \kappa\epsilon \tag{53}$$

where the excess MSE $\mathcal{J}_e(\mathbf{w}) = \mathbf{w}^H_{\lvert\Omega^c}\mathbf{R}_{\lvert\Omega^c}\mathbf{w}_{\lvert\Omega^c}$ has been bounded by the predefined error tolerance ϵ. The relation of parameter δ_S with the measurement matrix is further analyzed in Section V.

V. PERFORMANCE

In this section we address the issues of performance and recoverability of the adaptive SGP algorithm. In the first part, we prove that the proposed design is asymptotically stable in the mean-squares (MS) sense.

A. Recoverability

The *recoverability* of the proposed algorithm, i.e., the recovery of the correct support set, is guaranteed with high probability for a variety of algorithms, including greedy and ℓ_1-optimization based ones, when the measurement matrix $\mathbf{R}$ satisfies the RIP of order S with constant δ_S. Based on the Definition 2, an alternative way to guarantee recoverability is by bounding the eigenvalue spread of the Gram matrix $\mathbf{R}^H\mathbf{R}$, which appears in the definition of the proxy signal [29]. A particular characteristic of the proposed scheme, is that in the definition of the proxy signal, at (39), appears the measurement matrix $\mathbf{R}$. This implies that matrix $\mathbf{R}$ should preserve the energy of any set with S components of the equalizer filter, and hence, the recoverability depends on the eigenvalue spread of $\mathbf{R}$ and not of $\mathbf{R}^H\mathbf{R}$. The following lemma is concerned with the eigenvalue spread of the MMSE autocorrelation matrix $\mathbf{R}$.

Lemma 1: Let $\mathbf{h}$ be a sparse channel of length L (with N_1 the non-causal length) containing S_c nonzero taps and normalized so that $\lVert\mathbf{h}\rVert = 1$. Then the eigenvalues of any $S \times S$ input autocorrelation submatrix of the MMSE-DFE are bounded in the range $(1-\rho, 1+\rho)$, where $\rho = S(\mathcal{O}(\epsilon) + S_c\mathcal{O}(\epsilon^2)) + \sigma^2_\eta$, under the following assumptions.

1) The specular path h_0 has amplitude of order $\mathcal{O}(1)$, while each of the scattered paths has amplitude of order $\mathcal{O}(\epsilon)$, where $\epsilon \in (0,1)$, such that $\sum_{k=1}^{S_c}\lvert h_k\rvert < \lvert h_0\rvert$.
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 $\mathbf{R}$ in (17). According to the Gershgorin theorem [31], each eigenvalue of $\mathbf{R}$ 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 $\mathbf{R}$, we will examine the off-diagonal entries of the first K_f and the last K_b rows separately. The off-diagonal elements of the first K_f rows of matrix $\mathbf{R}$ are represented by the row elements of the matrix $\mathbf{R}_{xd} = \mathbf{H}\mathbf{J}_\Delta$ and the off-diagonal row elements of matrix $\mathbf{R}_{xx} = \mathbf{H}\mathbf{H}^H + \sigma^2_\eta\mathbf{I}_{K_f}$. By taking into account the second assumption, i.e., $\Delta \geq K_f - 1 + N_1$, it can be easily verified that matrix $\mathbf{R}_{xd}$, does not contain the channel coefficient h_0, and hence the order of the amplitude of its elements is equal to $\mathcal{O}(\epsilon)$. Let us now express the matrix $\mathbf{R}_{xx}$ as the sum of two matrixes, one with the diagonal elements and one with the off-diagonal elements, i.e.,

$$\mathbf{R}_{xx} = \left(\sum_{i=-N_1}^{N_2}\lvert h_i\rvert^2 + \sigma^2_\eta\right)\mathbf{I}_{K_f} + \mathbf{F} \tag{54}$$

where the Hermitian matrix $\mathbf{F}$ results from $\mathbf{H}\mathbf{H}^H$ after subtracting the main diagonal. It is straightforward to show that the matrix $\mathbf{F}$ consists of elements with amplitude of order $\kappa_1 = \mathcal{O}(\epsilon) + S_c\mathcal{O}(\epsilon^2)$. To do so, consider that the channel correlation matrix $\mathbf{H}$ is expressed as $\mathbf{H} = \mathbf{F}_0 + \mathbf{D}_{N_1+1}$, where $\mathbf{F}_0$ is 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. It follows that

$$\mathbf{H}\mathbf{H}^H = \mathbf{D}_{N_1+1}\mathbf{D}^H_{N_1+1} + 2Re\left(\mathbf{D}_{N_1+1}\mathbf{F}^H_0\right) + \mathbf{F}_0\mathbf{F}^H_0 \tag{55}$$

TABLE III
COMPARISON IN TERMS OF COMPLEX MULTIPLICATIONS

Adaptive Algorithm	Complexity Order per Adaptation Step
Conventional RLS	$\mathcal{O}(K^2)$
SpAdOMP	$\mathcal{O}(K)$
CB-DFE	$\mathcal{O}(K)$
SGP DFE	$\mathcal{O}(KS)$
ν-SGP DFE	$\mathcal{O}(K\nu)$ with $\nu \in [1, K]$

where the first term of (55) is a diagonal matrix with $|h_0|^2$ on the main diagonal, the second term contains only off-diagonal elements with amplitude order equal to $\mathcal{O}(\epsilon)$, and the off-diagonal elements of the third term have amplitude order equal to $S_c\mathcal{O}(\epsilon^2)$. Considering the last K_b rows of matrix $\mathbf{R}$, the off-diagonal entries have amplitude of the order equal to $\mathcal{O}(\epsilon)$. Therefore, the $S \times S$ submatrix of $\mathbf{R}$, denoted by $\mathbf{R}_S$, will have a maximum row sum for the off-diagonal elements equal to $S(\mathcal{O}(\epsilon) + S_c\mathcal{O}(\epsilon^2))$. The diagonal elements of matrix $\mathbf{R}$ are different for the first K_f and the last K_b rows, where for the second case are equal to unity. However, the first K_f row elements diverge from unity by σ_η^2. Based on the above analysis, we conclude that each of the eigenvalues of the $S \times S$ submatrix of $\mathbf{R}$ belongs to a circle with center the unity and radius $S(\mathcal{O}(\epsilon) + S_c\mathcal{O}(\epsilon^2)) + \sigma_\eta^2$. ■

B. Complexity

The SGP algorithm has a computational complexity of $\mathcal{O}(KS)$ multiplications per step, which is K/S times faster than the complexity of the nonsparsity aware RLS algorithm [4]. However, there is a slight increase in complexity compared with the greedy LMS scheme presented in [13], due to the computation of the gradient $\mathbf{g}(n)$ and the extra matrix-by-vector product $\mathbf{R}_{|\Omega}(n)\mathbf{g}_{|\Omega}(n)$.

In Table III, the computational complexity of the proposed algorithms is compared with that of 1) the conventional non sparsity aware recursive least squares (RLS), 2) the sparse-channel-based parametric DFE (CB-DFE) which is proposed in [16], and 3) the sparse adaptive OMP (SpAdOMP) which is proposed in [13]. The comparison is carried out in terms of complex multiplications per adaptation step of the algorithm.

VI. SIMULATION RESULTS

In this section, we provide simulation results for the proposed heuristic algorithms in various channel conditions. Their performance has been evaluated in a time invariant, as well as in a time-varying environment. The input sequence consisted of Gray coded 4QAM symbols, while complex white Gaussian noise was added to the channel output.

The performance of the sparsity aware algorithms depends on the predefined sparsity order parameter S, which results in an upper bound for the excess MSE. Furthermore, the number of nonzero equalizer taps needed to approach the nonsparse performance is expected to increase as the input signal-to-noise ratio (SNR) increases, where the input SNR is defined as $\mathrm{SNR} = 10\log_{10}(\sigma_x/\sigma_\eta)$. In the simulations we have used two input SNR regimes: 10 and 30 dB. Note that in 30 dB, due to a lower attainable steady-state error, we may have an increased number of nonzero taps compared to the 10 dB case.

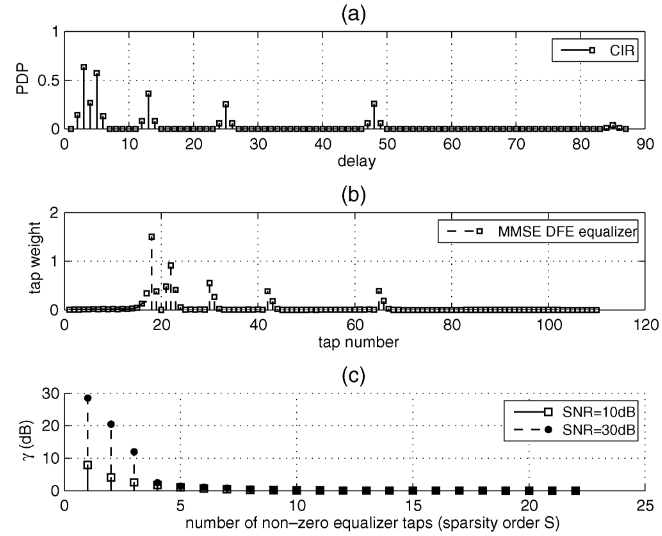

Fig. 1. (a) Power delay profile (PDP) of the sparse W-CDMA PB channel. (b) Equalizer FF and FB tap weights, with $K_f = 20$ and $K_b = 90$. (c) Comparison of misadjustment error γ with respect to sparsity order S.

In Fig. 1, we show the performance of a threshold MMSE-DFE for the "pedestrian B" ("PB") W-CDMA channel for the two SNR regimes, in terms of the misadjustment error γ of the thresholded MMSE DFE

$$\gamma = 10\log_{10}\left(1 + \frac{\mathcal{J}_e}{\mathcal{J}_{\min}}\right). \quad (56)$$

Fig. 1(a) shows the PB CIR convolved with a square-root raised cosine filter (SRRC) with 11.5% rolloff, while Fig. 1(b) shows the full, non sparse aware, MMSE DFE equalizer filter. Comparing the results of two SNR regimes, in Fig. 1(c), we observe the misadjustment error is larger for higher SNR, given the same sparsity order.

A. Time-Invariant Scenario

In this scenario, the channels are assumed to be block fading, that is, they are constant within a block and independent between blocks. In our study, we have used blocks of length $4096T_s$ symbols over several typical multipath channels from W-CDMA and HDTV environments [1], [2]. However, we present the results for the commonly used "PB" W-CDMA channel model, which is worse than the HDTV CIR, due to its strong scattered paths.

First we study the steady-state performance of the proposed scheme. In Fig. 2(a) and (b), the performance of the SGP and SGP with variable order (ν-SGP) is compared, in terms of the MSE learning curves, with the nonsparse aware LS DFE and the threshold LS (thLS). For the latter one, the support set for each iteration is obtained by thresholding the equalizer filter of a non sparse LS DFE. The input SNR was set to 10 and 30 dB for each figure, respectively, while the FF, FB filter lengths were set to $K_f = 20$, $K_b = 90$, and $\lambda = 1$. In each figure, the SGP algorithm, along with the thLS, is presented for three different sparsity order parameters, i.e., $S = \{5, 10, 30\}$ taps. We observe that imposing a smaller sparsity order S, has severe

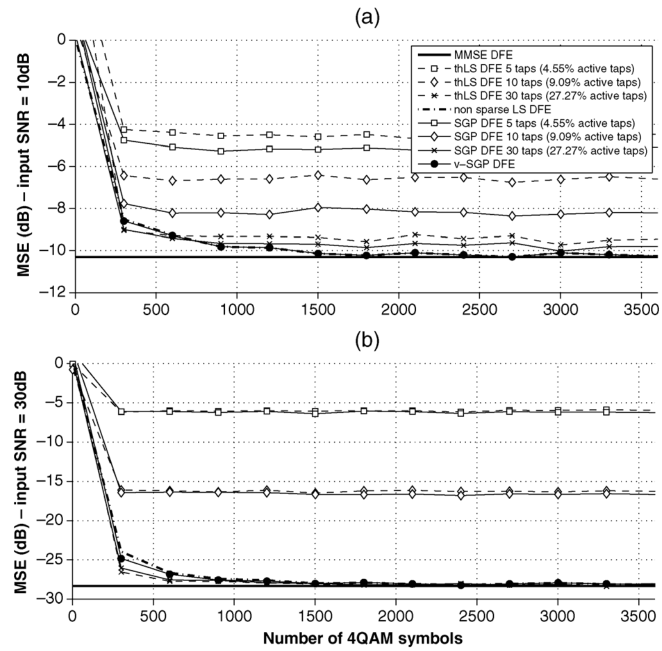

Fig. 2. MSE learning curves for the W-CDMA(PB) channel.

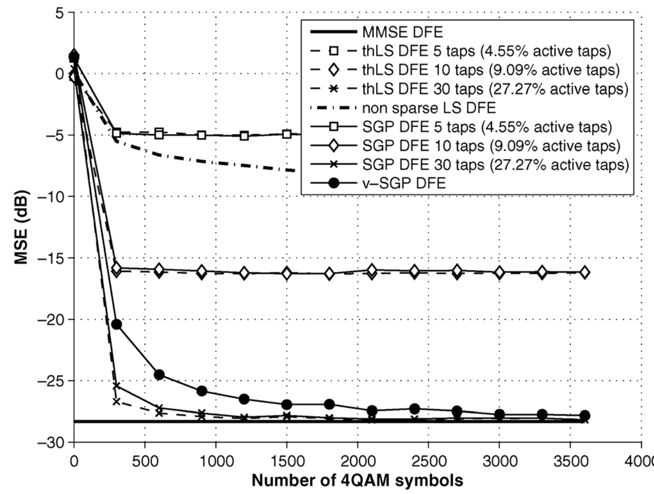

Fig. 3. MSE learning curves for the W-CDMA(PB) channel, input $\mathrm{SNR} = 30\ \mathrm{dB}$ and training period $= 300T_s$.

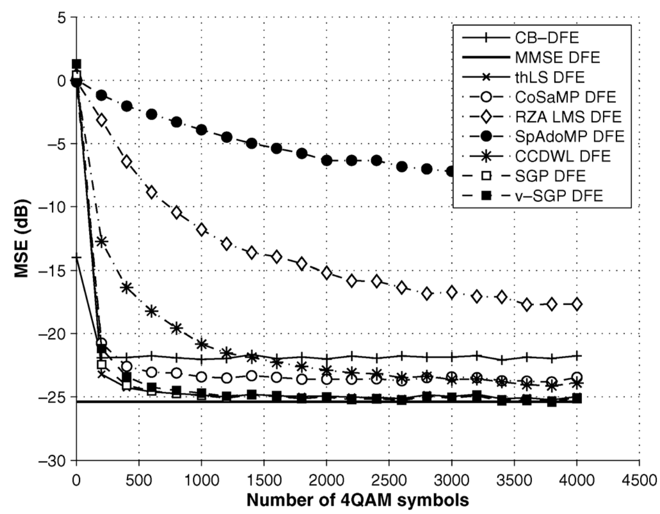

Fig. 4. MSE learning curves for the W-CDMA(PB) channel, input $\mathrm{SNR} = 30\ \mathrm{dB}$.

effects on the MSE in the high SNR case. For $\mathrm{SNR} = 10\ \mathrm{dB}$, using only five nonzero components for the DFE filter, out of the total of 110 taps, results in 10 dB difference from the optimal case, while for $\mathrm{SNR} = 30\ \mathrm{dB}$, wemust increase S up to 30 nonzero components, in order to have the same misadjustment in MSE. We must point out that similar results are obtained by using the typical HDTV CIR and a modified HDTV CIR where (27) is marginally satisfied, however due to space limitations are not presented in this work.

Error propagation effects in decision-directed mode were studied by employing a training period of $300T_s$, which is too sort relatively to the CIR span $L = 83$ and the length of the equalizer filter $K = 110$. As observed in Fig. 3, due to the small length of the training sequence, nonsparse LS algorithm fails to converge to the steady state, whereas for the SGP and ν-SGP algorithms the number of the training symbols is sufficient in order to converge to the steady state. We observe that the ν-SGP, without any prior information, successfully tracks the sparsity order, converging to the optimal steady-state even for very short training period.

In Fig. 4, we compare nine different sparsity aware adaptive DFE schemes, in terms of MSE steady-state error. The tested schemes belong to the greedy algorithms as well as the ℓ_1-based minimization algorithms, and they are: 1), 2) the proposed SGP algorithms (SGP DFE and ν-SGP DFE), 3) the CoSaMP [29] algorithm for adaptive LS DFE where the iterations of the algorithm conducted in the time domain, 4) the reweighted zero-attracting least mean squares (RZA-LMS) [32] algorithm for MMSE DFE with fixed step size equal to 0.008 and parameters $\rho = 10^{-4}$, $\epsilon = 10$, 5) the sparse adaptive OMP (SpAdOMP) [13] algorithm for adaptive LS DFE with step size equal to 10^{-3}, 6) the cyclic coordinate descent weighted lasso (CCDWL) [33] iterative algorithm for MMSE DFE with 100 iterations per adaptation step, 7) the MMSE DFE, 8) the thLS DFE, and 9) the sparse-channel-based parametric DFE (CB-DFE) [16]. For the greedy algorithms the sparsity order S was set to 30 symbols, except for the ν-SGP which adapted the order given the error tolerance $\kappa\epsilon = 0.4$. We observe that only the proposed SGP and the thLS converge to the steady-state with less than 1000 4QAM symbols. CB-DFE scheme cannot converge to the optimum steady state, since the assumption of $\sum_{k=1}^{S} |h_{n_k}| < |h_0|$ is not satisfied under the usage of PB channel with SRRC filter.

Next, we investigate the bit-error-rate (BER) performance of the proposed SGP algorithm, with Gray coded 4QAM modulation. In Fig. 5, we compare the previous nine algorithms in terms of BER with respect to input SNR, and we notice that the proposed schemes outperform the other sparsity aware algorithms, with the ν-SGP performing near the optimum MMSE DFE.

B. Time-Varying Scenario

In this scenario, the tracking performance of the algorithms was tested by simulating a system that operates over a mod-

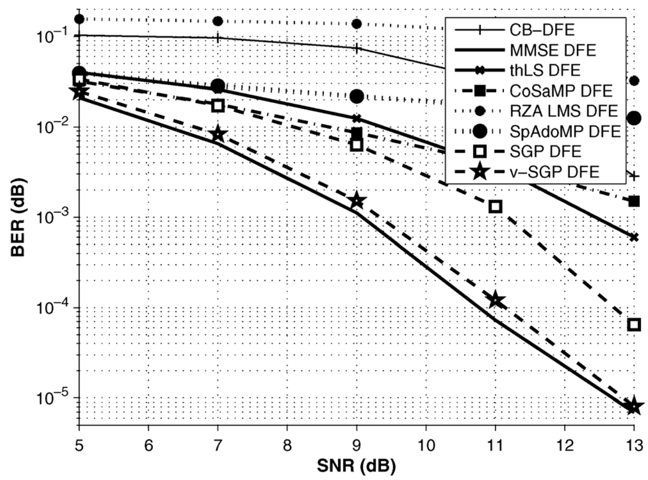

Fig. 5. BER with respect to input SNR for the W-CDMA(PB) channel.

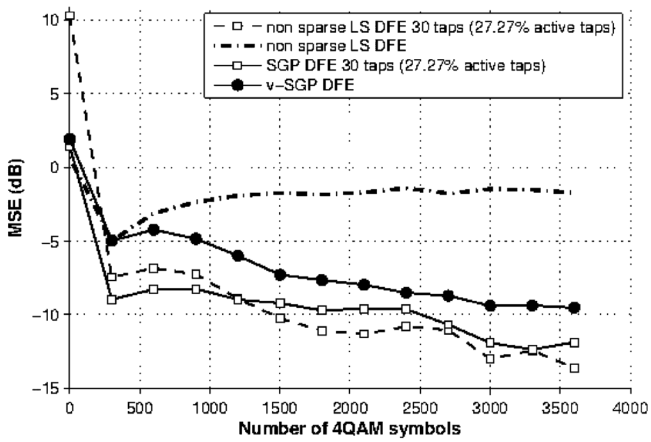

Fig. 6. MSE learning curves for the time varying W-CDMA(PB) channel, input $\mathrm{SNR} = 30$ dB.

ified PB channel, where the first and second multipath components vary according to the autoregressive model $h_i(n) = \alpha h_i(n-1) + \sqrt{1-|\alpha|^2}v(n)$ [4] where $\alpha = J_o(2\pi f_D T_s)$, $J_o(\cdot)$ is the zeroth-order Bessel function with a normalized Doppler frequency $f_D T_s = 1.1\ 10^{-5}$ and $v(n)$ denotes a white noise process with unit variance. The support set of the sparse channel also undergoes gradual changes. The second multipath component is gradually reduced starting at the time delay $23T_s$ and is nullified after 1000 symbol periods while a new component with increasing amplitude gradually appears in the time delay $40T_s$ with respect to the main peak of the CIR. After 1000 symbol periods the amplitude of the new multipath component is equal to -5 dB.

In Fig. 6, we compare the proposed algorithms, the SGP and the ν-SGP, with the conventional LS DFE and the thresholded LS DFE (thLS), in terms of MSE learning curves. The forgetting factor was set to $\lambda = 0.98$, the training period to 400 symbols and the input SNR to 30 dB. We observe that the proposed schemes successfully track the channel changes, along with the thLS, as opposed to the conventional LS.

VII. Conclusion

We described a general framework for the sparse adaptive equalization problem, under the compressive sensing perspective. A new heuristic scheme has been derived, which offers reduced computational complexity compared to existing techniques and is also able to cope with unknown sparsity order. As demonstrated via extensive simulations, the proposed algorithms yield considerable reductions in complexity while maintaining performance comparable to the conventional DFE. Error propagation studies showed a noticeable fast convergence which allows the use of shorter training sequence in the applications of interest.

References

[1] W. F. Schreiber, “Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions,” *Proc. IEEE*, vol. 83, no. 6, pp. 958–981, Jun. 1995.

[2] “3gpp 25.101 user equipment (ue) radio transmission and reception (fdd),” 3rd Generation Partnership Project V.7.4.0 Jun. 2006.

[3] A. Singer, J. Nelson, and S. Kozat, “Signal processing for underwater acoustic communications,” *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 90–96, Jan. 2009.

[4] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[5] M. Melvasalo, P. Janis, and V. Koivunen, Sparse equalization in high data rate wcdma systems pp. 1–5, Jun. 2007.

[6] N. Al-Dhahir and J. M. Cioffi, “Fast computation of channel-estimate based equalizers in packet data transmission,” *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2462–2473, Nov. 1995.

[7] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, “Reduced complexity decision feedback equalization for multipath channels with large delay spreads,” *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 927–937, Jun. 1999.

[8] I. Lee, “Optimization of tap spacings for the tapped delay line decision feedback equalizer,” *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 429–431, Oct. 2001.

[9] S. A. Raghavan, J. K. Wolf, L. B. Milstein, and L. C. Barbosa, “Non-uniformly spaced tapped-delay-line equalizers,” *IEEE Trans. Commun.*, vol. 41, no. 9, pp. 1290–1295, Sep. 1993.

[10] F. K. H. Lee and P. J. McLane, “Design of nonuniformly spaced tapped-delay-line equalizers for sparse multipath channels,” *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 530–535, Apr. 2004.

[11] B. Geller, V. Capellano, J. M. Brossier, A. Essebbar, and G. Jourdain, “Equalizer for video rate transmission in multipath underwater communications,” *IEEE J. Oceanic Eng.*, vol. 21, no. 2, pp. 150–155, Apr. 1996.

[12] J. Homer, I. Mareels, and C. Hoang, “Enhanced detection-guided nlms estimation of sparse fir-modeled signal channels,” *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 53, no. 8, pp. 1783–1791, Aug. 2006.

[13] G. Mileounis, B. Babadi, N. Kalouptsidis, and V. Tarokh, “An adaptive greedy algorithm with application to nonlinear communications,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 2998–3007, Jun. 2010.

[14] F. Wan, W. P. Zhu, and M. N. S. Swamy, “Semi-blind most significant tap detection for sparse channel estimation of ofdm systems,” *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 57, no. 3, pp. 703–713, Mar. 2010.

[15] S. F. Cotter and B. D. Rao, “The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels,” in *Conf. Rec. 34th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2000, vol. 2, pp. 1772–1776.

[16] A. A. Rontogiannis and K. Berberidis, “Efficient decision feedback equalization for sparse wireless channels,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 570–581, May 2003.

[17] A. S. Lalos, E. Vlachos, K. Berberidis, and A. A. Rontogiannis, “Greedy algorithms for sparse adaptive decision feedback equalization,” in *IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2011.

[18] E. Vlachos, A. S. Lalos, G. Lionas, and K. Berberidis, “Compressed sensing techniques for decision feedback equalization of sparse wireless channels,” in *IEEE Veh. Technol. Conf.*, May 2012.

[19] A. Gomaa and N. Al-Dhahir, “A new design framework for sparse fir mimo equalizers,” *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2132–2140, Aug. 2011.

[20] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[21] E. J. Candes, “Compressive sampling,” in *Int. Congress Math.*, 2006, pp. 1433–1452.

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

[23] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition," in *Asilomar Conf. Signals, Syst. Comput.*, 1993.

[24] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370–2382, Jun. 2008.

[25] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[26] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication*, 3rd ed. New York: Kluwer, 2003.

[27] D. Cassioli and A. Mecozzi, "Minimum-phase impulse response channels," *IEEE Trans. Communications*, vol. 57, no. 12, pp. 3529–3532, Dec. 2009.

[28] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[29] D. Needell and J. A. Tropp, "Cosamp: Iteratives signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, Jan. 2010.

[30] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley-Interscience, 1980, vol. 1.

[31] R. S. Varga, *Gershgorin and His Circles*. Berlin, Germany: Springer-Verlag, 2004.

[32] C. Yilun, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3125–3128.

[33] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Stat.*, vol. 1, no. 2, pp. 302–332, Jan. 2007.

Evangelos Vlachos (S'11) was born in Athens, Greece, in 1980. He received the Diploma degree in computer engineering and informatics, and the M.S. degree in signal processing, in 2005 and 2009, respectively, from the University of Patras, Greece, where he is currently pursuing the Ph.D. degree in signal processing and communications.

Since 2009, he has been a Research Assistant with the Computer Technology Institute (CTI), Patras, Greece, and as a Teaching Assistant with the Computer Engineering and Informatics Department, University of Patras. His research interests include equalization and signal processing for communications and sparse signal processing.

Mr. Vlachos is a member of the Technical Chamber of Greece.

Aris S. Lalos was born in Athens, Greece, in 1981. He received the Diploma degree in computer engineering and informatics, the M.A.Sc. degree in signal processing and communications, and the Ph.D. degree in signal processing for wireless communications from the Computer Engineering and Informatics Department (CEID), School of Engineering (SE), University of Patras (UoP), Rio-Patras, Greece, in 2003, 2005, and 2010, respectively.

He has been a research fellow at Signal Processing and Communications Laboratory, CEID, SE, UoP, Rio-Patras, Greece since 2001. He is currently working as a Telecommunication Research Engineer at Analogies S.A., an early stage start up Semiconductor Intellectual property (SIP) design company. His research interests involve MIMO transmission and reception techniques, adaptive algorithms for channel estimation and equalization, cooperative and cognitive networks, compressive sampling, and sparse linear regression.

Dr. A. S. Lalos is a member of the Technical Chamber of Greece, acting also as a technical reviewer for various conferences and journals.

Kostas Berberidis (S'87–M'90–SM'07) received the Diploma degree in electrical engineering the from Democritus University of Thrace, Komotini, Greece, in 1985, and the Ph.D. degree in signal processing and communications from the University of Patras, Patras, Greece, in 1990.

During 1991, he worked at the Signal Processing Laboratory of the National Defense Research Center. From 1992 to 1994 and from 1996 to 1997, he was a researcher at the Computer Technology Institute (CTI), Patras, Greece. In period 1994–1995 he was a Postdoctoral Fellow at CCETT/CNET, Rennes, France. Since December 1997, he has been with the Computer Engineering and Informatics Department (CEID), University of Patras, where he is currently a Professor, and Head of the Signal Processing and Communications Laboratory. Also, since 2008, he has been Director of the Signal Processing and Communications Research Unit of CTI. He has been active in areas as: adaptive filtering, signal processing for communications, wireless sensor networks, biomedical signal processing, etc. He has published more than 120 papers in refereed international technical journals and proceedings of international conferences. He is coauthor of four international patents as well as coauthor of three technical books. He has been the principal investigator of more than 25 national, E.U., and bilateral projects.

Prof. Berberidis has served as a member of scientific and organizing committees of several international conferences, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and as a Guest Editor for the EURASIP JASP. He is currently serving as Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and for the EURASIP Journal on Advances in Signal Processing. Also, since February 2010 he has been serving as Chair of the Greece Chapter of the IEEE Signal Processing Society. He is a Member of the Technical Chamber of Greece and a member of EURASIP.