Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Special Issue: LVA
Special Issue :Latent Variable Analysis and Signal Separation 00 (2012) 1–15

ISI Sparse Channel Estimation Based on SL0 and its Application in ML Sequence-by-Sequence Equalization☆

Rad Niazadeh[a], Sina Hamidi Ghalehjegh[b], Massoud Babaie-Zadeh[a], Christian Jutten[c]

[a]*Department of Electrical Engineering, Sharif University of Technology, Iran*
[b]*Department of Electrical and Computer Engineering, McGill University, Canada*
[c]*GIPSA-Lab, Grenoble, and Institut Universitaire de France, France*

Abstract

In this paper, which is an extended version of our work at LVA/ICA 2010 [1], the problem of Inter Symbol Interface (ISI) Sparse channel estimation and equalization will be investigated. We firstly propose an adaptive method based on the idea of Least Mean Square (LMS) algorithm and the concept of smoothed l_0 (SL0) norm presented in [2] for estimation of sparse ISI channels. Afterwards, a new non-adaptive fast channel estimation method based on SL0 sparse signal representation is proposed. ISI channel estimation will have a direct effect on the performance of the ISI equalizer at the receiver. So, in this paper we investigate this effect in the case of optimal Maximum Likelihood Sequence-by-sequence Equalizer (MLSE) [3]. In order to implement this equalizer, we first introduce an equivalent F-model for sparse channels, and then using this model we propose a new method called pre-filtered Parallel Viterbi Algorithm (or pre-filtered PVA) for *general ISI sparse channels* which has much less complexity than ordinary Viterbi Algorithm (VA) and also with no considerable loss of optimality, which we have examined by doing some experiments in Matlab/Simulink. Indeed, Simulation results clearly show that the proposed concatenated estimation-equalization methods have much better performance than the usual equalization methods such as Linear Mean Square Equalization (LMSE) for ISI sparse channels, while preserving simplicity at the receiver.

Keywords: Channel Estimation, Channel Equalization, Adaptive Filters, Sparse Recovery, Viterbi Algorithm

1. Introduction

In the framework of digital communication over band-limited channels, there are special scenarios in which one can model the overall channel as a Finite Impulse Response (FIR) sparse filter which will produce interferences with previous samples, i.e. Inter Symbol Interference (ISI). By the constraint of sparsity, we mean that only a few taps of the overall ISI channel are non-zero. As a result, the vectorized model of the mentioned FIR Channel Impulse Response (CIR) could be considered as a sparse vector. Such channels may be encountered, for example, in wireless multipath fading channels, acoustic underwater channels, or even in simplified models of ultra wideband communications [4]. Other applications of sparse channels, such as modelling the overall underground channel for propagation of seismic waves (which appears in the problem of finding the location of oil wells), have also been reported in literature of geological signal processing [5]. In most of these scenarios, the communication between transmitter and receiver should be as fast and reliable as it can be. Among the major factors that affects the speed and reliability of the

☆This work has been partially funded by Iran Telecom Research Center (ITRC) and Iran National Science Foundation (INSF). Part of this work was done when the third author was in sabbatical at the University of Minnesota.

Email addresses: `radniazadeh@gmail.com` (Rad Niazadeh), `shghamidi@gmail.com` (Sina Hamidi Ghalehjegh), `mbzadeh@yahoo.com` (Massoud Babaie-Zadeh), `Christian.Jutten@inpg.fr` (Christian Jutten)

overall data communication, one can mention the performance and convergence speed of channel estimation and equalization [3]. In the process of channel estimation, usually the transmitter sends a known-to-receiver training data, and the receiver estimates the CIR using the received observations. This estimated channel will then be used by other elements of the communication systems such as channel equalizer or detectors at the receiver side [3]. In the process of the channel equalization, the receiver tries to compromise the effect of produced ISI, so that the overall detection error of the communication system will be decreased. Accordingly, having a high accurate and fast channel estimator, as well as a high performance equalizer is at the point of interest in the literature of communications through band-limited channels [3].

Considering the problem of estimating the sparse channel, many efforts have been done to design batch estimation algorithms [4, 6, 7]. Most of these algorithms try to exploit the sparsity of the channel or detecting the locations of non-zero taps of the CIR. In addition, adaptive algorithms for sparse channel estimation are also proposed which are based on iterative estimation of the taps of a sparse channel [8], or even adapting the well-known methods of sparse recovery, such as CoSaMP [9], to the problem of sparse channel estimation [10]. At the other side, many efforts have been done to design high-performance equalizers for sparse channels [11, 12]. Among all of the choices for the structure of the equalizer, one may try to use the optimum equalizer in the sense of Mean Square Error (MSE). This equalizer is Maximum Likelihood Sequence-by-Sequence equalizer (MLSE) which is implemented using Viterbi Algorithm [3]. For the case of sparse channels, a simplified structure known as Parallel Viterbi Algorithm (PVA) has been proposed by Mcginty et al. [13], which exploits the sparsity of the channel for decreasing the computation complexity of the MLSE equalizer. Although using PVA reduces the computational complexity of the receiver, it just works for a special type of sparse channels, known as zero-pad channel, in which non-zero taps are equally spaced [13].

In this work, which is an extended version of our work presented at LVA/ICA 2010 [1], we investigate these two problems of channel estimation and equalization in the case of ISI sparse channels with a novel approach. According to our best knowledge, all of the works around the problem of channel estimation and equalization have considered the ISI channel as an "FIR causal filter" with no constraint on the taps. The first contribution of our work is to solve these two problems under the assumption of using a matched filter structure at the receiver, which is a well-known receiver structure when communicating data through a band-limited channel (ISI channel) [3, 4]. In this situation, the resulting sparse FIR channel can be assumed to be minimum phase [3] (the reason of this assumption is described in Sect. 2, in which we shown that by sampling the output of the channel and using a whitening discrete-time filter the end, we may model the whole CIR as a minimum-phase FIR filter). This assumption can make our equalization problem much easier to solve. More accurately, as will be shown in Sect. 3 of this work, it will help us to adapt PVA using a pre-filter at the receiver in order to implement MLSE with much less complexity. All in all, using this assumption we develop our algorithms in order to find the solutions to both the estimation and equalization problems. Finally, we experimentally examine the efficiency of the proposed algorithms in a concatenation of estimation and equalization stages.

This paper is divided in three main parts. In the first part of this paper, motivated by the Smoothed l_0-norm (SL0) algorithm [2] which is known as a fast sparse representation technique, we propose two new approaches for sparse channel estimation problem. Firstly, by modifying LMS algorithm introduced by Widrow and Hoff [14], we propose a new adaptive algorithm for channel estimation, called SL0-LMS. This algorithm is similar to the algorithms proposed in [8] (which are named as Zero-Attracting LMS (ZA-LMS) and Re-weighted Zero-Attracting LMS (RZA-LMS) in [8]): it just differs by the regularization term and the possibility of varying this regularization term and the weight of this term adaptively. We mathematically show that by exploiting the sparsity information and using the concept of Smoothed l_0-norm we can improve the estimation performance in the sense of MSE, comparing to standard LMS method. we also provide a local convergence analysis of our proposed algorithm. We show that by choosing some parameters of our algorithm wisely at each iteration, we can guarantee both stability and performance of our algorithm. Furthermore, we will experimentally investigate the effect of adding a smoothed l_0 norm penalty term to the cost function on the learning curve of LMS, and will investigate the improved tracking behaviour and steady state error of our proposed algorithm. Accordingly, we will experimentally show that it has better steady state and tracking behaviour in comparison to ZA-LMS and RZA-LMS. The second proposed algorithm is a non-adaptive algorithm which uses SL0 in a direct manner to estimate channel coefficients by finding the sparsest solution[1] of an under-determined system of linear equations [2]. Although this algorithm is non-adaptive, it uses the speed of SL0 while preserving enough accuracy for the equalization stage. It is important to mention that none of the proposed algorithms in this part depend on the assumption of modelling the channel as a minimum phase filter and can be used if we do not use the matched filter structure at the receiver.

[1]Solution with minimum l_0-norm, i.e. minimum number of non-zero coefficients.

In the next part, high-performance equalization based on the MLSE (also known as the Viterbi equalization [3]) will be described. In conventional ISI sparse channels, implementation of the MLSE using the ordinary Viterbi algorithm is almost impossible (because the computational complexity of the receiver grows exponentially with the channel memory [3]). However, as mentioned earlier, in some special cases known as zero-pad ISI channels the parallel Viterbi algorithm is used instead of the ordinary Viterbi algorithm [15, 13]. In PVA a number of trellises are used after symbol deinterlacing and placing symbols into different groups. In fact, symbols that produce ISI on each other will be placed in the same group and Viterbi equalization will be performed separately on each group. The performance of this structure is the same as of a conventional Viterbi equalizer (as this structure is just an implementation of the Viterbi equalizer), while having much lesser complexity. In the case of *general* ISI sparse channels, according to our best knowledge, no such solution is yet presented in the literature. So, we propose a new method for using PVA in *general* ISI sparse channels based on the modelling of the ISI channel with a minimum-phase FIR filter. Our idea is based on the usage of a pre-filter at the receiver. In fact, this filter will re-shape the channel structure to a zero-pad channel which is presented in [15] and so, applying the PVA method will be possible afterwards. Although this method is not optimal, as we will see in experimental results, we do *not* have considerable amount of loss in optimality in most practical cases.

In the last part, computer simulations with Matlab/Simulink will be used to obtain the performance of our proposed algorithms. The results of these mentioned simulations will verify the performance of the presented algorithms on the both estimation of the channel and equalization of the produced ISI. It is important to note that the overall performance of the receiver system (as a concatenation of channel estimation and ISI equalization) depends on both of these parts which will be evaluated in these simulations and is considered as the overall measure of efficiency of the proposed algorithms for both estimation and equalization tasks[2].

2. ISI Sparse Channel Estimation

In this section, we would like to estimate the ISI sparse channel coefficients while transmitting digital bits through this channel using Binary Phase Shift Keying (BPSK) modulation [3][3]. We assume that the BPSK modulator maps the incoming digital bits to the symbols $\{\pm 1\}$. Accordingly, assuming a signaling interval of T, the transmitter sends $s(t) = \sum_{n=-\infty}^{+\infty} u(n)g(t - nT)$ through the channel, in which $u(n)$ is the transmitted $\{\pm 1\}$ symbol at n-th signaling interval (the transmitted symbol corresponding to the transmitted bit at n-th signaling interval) and $g(t)$ is the pulse shape of BPSK modulator. Including the BPSK modulator and de-modulator with the channel and assuming that we are using matched filter structure at the receiver, we can model the overall waveform channel with a corresponding discrete time channel. More accurately, assume that the analog channel impulse response is $c(t)$, and it follows by an Additive White Gaussian Noise (AWGN) process $n(t)$ with spectral density equal to $\frac{N_0}{2}$. Lets define $h(t) \triangleq g(t) * c(t)$. When the transmitted signal $s(t)$ passes through the channel, the following signal will be received at the receiver side:

$$r(t) = \sum_{n=-\infty}^{+\infty} u(n)h(t - nT) + n(t) \ . \tag{1}$$

By use of a matched filter $h(T - t)$ at the receiver, the following signal will be extracted at the output of matched filter:

$$y(t) = \sum_{n=-\infty}^{+\infty} u(n)x(t - nT) + z(t) \ , \tag{2}$$

in which $x(t) \triangleq h(t) * h(-t)$ and $z(t)$ is the filtered Gaussian noise. By sampling the output of the matched filter at $t = (n + 1)T$, it is convenient to model the overall channel as a symmetric non-causal FIR filter with taps equal to $\{x(n) : -M \leq n \leq M\}$ as following:

$$y(n) = u(n) * x(n) + z(n) \ , \tag{3}$$

[2]Parts of this work have been presented in LVA/ICA 2010 [1]. In the current paper we provide further explanation and details of our algorithm, a local convergence and performance analysis for our proposed SL0-LMS algorithm, and new experiments for measuring the tracking behaviour of this algorithm.

[3]In this section, we will use the notations in [3, Chapter 9] for indicating all continuous time and discrete time signals that appears in data communications through a band-limited channel.

in which $x(n)$, $y(n)$, and $z(n)$ are defined as:

$$\begin{aligned} x(n) &= x(t)|_{t=nT}, \quad -M \leq n \leq M \\ y(n) &= y(t)|_{t=nT}, \quad -\infty \leq n \leq +\infty \\ z(n) &= z(t)|_{t=nT}, \quad -\infty \leq n \leq +\infty \ . \end{aligned} \tag{4}$$

It is easy to see that $z(n)$ is a colored noise with spectral density of $\mathscr{Z}\{z(n)\} = \frac{N_0}{2}\mathscr{Z}\{x(n)\} = \frac{N_0}{2}X(z)$, in which $\mathscr{Z}\{.\}$ denotes the Z-transform operator [3].

This model of describing an ISI channel is called the "X" model in the area of band-limited communications [3]. Additionally, we force the sparsity constraint on the taps of $x(n)$ which is a valid constraint in the scenarios that were mentioned in the previous section. Although this model is very common in the context of ISI channels and communications through band-limited channels [3], we did not use this model in this paper. Instead, we use a simpler model called "F" model [3] which can be extracted by decomposing the Z-transform of channel impulse response into a minimum phase filter and its conjugate reciprocal[4] ($X(z) = F(z)F^*(\frac{1}{z^*})$) and then adding a whitening filter according to this decomposition to the end of sampler [3]. More accurately, by concatenating a whitening filter $G(z) = \frac{1}{F^*(\frac{1}{z^*})}$ to the end of sampler, the equivalent discrete channel would be a minimum phase FIR filter with taps equal to $\{f(n) : 0 \leq n \leq M\}$, in which $f(n) = \mathscr{Z}^{-1}\{X(z)G(z)\} = \mathscr{Z}^{-1}\{F(z)\}$, followed by an additive white Gaussian noise $v(n)$ which is independent of the input and has a variance of $\frac{N_0}{2}$ [3]. This model of describing an ISI channel is shown in Fig. 1, and is described in the following equation:

$$d(n) = u(n) * f(n) + v(n) \ , \tag{5}$$

in which $d(n)$ is the output of the whitening filter $G(z)$.

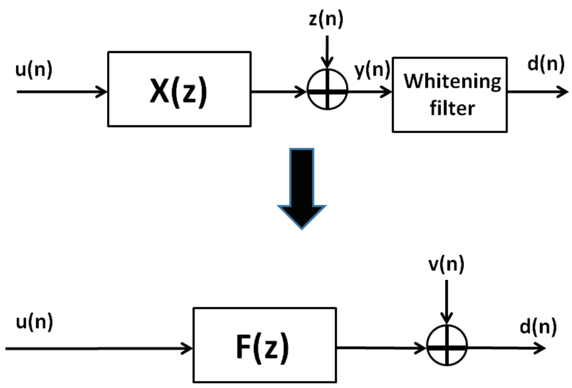


Figure 1: Discrete time models of the resulting ISI channel

To involve the sparsity constraint, the total ISI channel can be modelled as an FIR minimum-phase filter $\mathbf{f} = [f_0, \cdots f_{M-1}]^T$, which has only a few non-zero coefficients. To show that this assumption is reasonable, consider the relation between $x(n)$ and $f(n)$ in the frequency domain. Hence, we have $x(n) = f(n) * f^*(-n)$ in which $*$ denotes the convolution sum. So according to this relation, when $x(n)$ is very sparse it is very unlikely that $f(n)$ would not be sparse. Although we have currently no mathematical proof for this, it can heuristically be seen from the convolution sum, in which it is very unlikely that the convolution of two non sparse signal produces a sparse signal. It can also be experimentally verified. Therefore, we use the above mentioned minimum phase sparse FIR model for the ISI channel which will make our problem much easier to solve.

Now, let $d(n)$ be the last observed sample of the noisy output signal of the channel, as in (5), and let $\mathbf{u}(n)$ be a vector that contains the last M samples of the input signal of the ISI sparse channel, that is:

$$\mathbf{u}(n) = [u(n), u(n-1), \cdots, u(n-M+1)]^T \ . \tag{6}$$

[4]The possibility of this decomposition is due to the symmetry of $x(n)$.

Consequently, we have the following input-output relation for the channel:

$$d(n) = \mathbf{f}^T \mathbf{u}(n) + v(n) \ . \tag{7}$$

In addition to the channel model in (7), we assume that the receiver knows an upper-bound for the sparsity degree of the channel as a prior information, i.e. we have $||\mathbf{f}||_0 \leq \alpha$ and $\alpha \in \mathbb{Z}_+$ is known at the side of receiver. Note that having a prior information about an upper-bound for the degree of sparsity of the channel is one of prerequisites of our algorithms in this paper, while it is also one of well-known assumptions in the context of sparse channel estimation [4, 10]. Knowing this prior information and considering the model in (7), receiver uses a semi-random training sequence (which is known to the receiver) in order to estimate the channel. This sequence is generated by producing a random block of binary digits with length equal to M and transmitting it periodically (note that the BPSK modulator will map these bits to ± 1 symbols).

2.1. Adaptive Sparse Channel Estimation

As was mentioned earlier, in many scenarios often prior information about the unknown channel is available. One important example is when the impulse response of the unknown channel is known to be sparse, containing only a few large coefficients interspersed among many small ones. Exploiting such prior information can improve the filtering performance, i.e. the performance of channel estimation in the sense of mean square error. In this section, the adaptation of ordinary Least Mean Square (LMS) method will be investigated. We try to use the idea of the smoothed l_0 norm introduced in [2] and change the LMS algorithm by adding a regularization term to its cost function, so that the final algorithm can extract the sparse nature of the channel.

2.1.1. Review of ZA-LMS and RZA-LMS.

During the learning steps, i.e. when transmitting the training sequence, the standard LMS which is based on iteratively minimizing the cost function $J(\mathbf{w}(n)) = E\{e^2(n)\}$ adaptively estimates $\mathbf{f}$ by the following recursion:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{u}(n) \ . \tag{8}$$

where $\mathbf{w}(n) = [w_0(n), w_1(n), \ldots, w_{M-1}(n)]^T$ is the estimated adaptive filter at the nth iteration (which is an estimation for$\mathbf{f}$), μ is the step size parameter and $e(n) = d(n) - \mathbf{w}(n)^T\mathbf{u}(n)$ [14]. Note that (8) is just a recursion step of the simple steepest decent algorithm when the gradient vector of $J(\mathbf{w}(n))$, i.e. $E\{e(n)\mathbf{u}(n)\}$, is replaced by the stochastic gradient, i.e. $e(n)\mathbf{u}(n)$ [14]. For considering the spars structure of the channel impulse response, ZA-LMS and RZA-LMS have been developed by Chen et al. [8], in which a different regularization term has been added to the cost function of LMS in each of these algorithm. In ZA-LMS, the cost function is modified by adding a penalty term based on the l_1 norm[5] to enforce some sparsity on $\mathbf{w}(n)$. More accurately we have the following cost function:

$$J_1(\mathbf{w}(n)) = \frac{1}{2}E\{e^2(n)\} + \gamma||\mathbf{w}(n)||_1 \ , \tag{9}$$

in which γ is the regularization factor. Using steepest descent and by replacing the gradient vector of $J_1(\mathbf{w}(n))$ with its stochastic gradient, the channel coefficients update equation will then be [8]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{u}(n) - \mu\gamma \ \mathrm{sgn}(\mathbf{w}(n)) \ . \tag{10}$$

In RZA-LMS the cost function is defined as [8]:

$$J_2(\mathbf{w}(n)) = \frac{1}{2}E\{e^2(n)\} + \gamma\sum_{i=1}^{M}\log(1+\varepsilon|w_i|) \ , \tag{11}$$

in which γ is again the regularization factor. In (11), the log-sum term has been used because it behaves more similarly to the l_0 norm than $||\mathbf{w}(n)||_1$. So, the update equation will be

$$w_i(n+1) = w_i(n) + \mu e(n)u(n-i) - \mu\gamma\varepsilon\frac{\mathrm{sgn}(w_i(n))}{1+\varepsilon|w_i(n)|} \ , \tag{12}$$

for every $0 \leq i \leq M-1$.

[5]In Maximum A Posterior estimation theory, this cost function is associated to a Laplacian prior on the coefficients of $\mathbf{w}(n)$ [16].

2.1.2. The New Smoothed l_0-LMS Algorithm (SL0-LMS).

Inspired from the idea of the SL0 algorithm [2] and for exploiting the sparse nature of the estimated channel in a more accurate way, we propose replacing the above mentioned cost function with:

$$J_3(\mathbf{w}(n)) = \frac{1}{2}E\{e^2(n)\} + \gamma_n g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ , \tag{13}$$

in which γ_n is the variable regularization factor and $g_{\text{reg}}^{(n)}(.)$ is a multi-variable smooth function. In fact, $g_{\text{reg}}^{(n)}(\mathbf{w}(n))$ is a regularization term at n-th iteration and is based on the smooth approximation of $||\mathbf{w}(n)||_0$ defined in [2]. More accurately we use the following approximation:

$$||\mathbf{w}(n)||_0 \approx g_{\text{reg}}^{(n)}(\mathbf{w}(n)) = M - \sum_{i=1}^{M} e^{-w_i^2(n)/2\sigma_n^2} \ , \tag{14}$$

in which σ_n is the smoothness factor at n-th iteration and is updated at each iteration. So, following the same procedure as in LMS, i.e. by taking the gradient vector of $J_3(\mathbf{w}(n))$ with respect to $\mathbf{w}(n)$ and replacing the gradient with stochastic gradient, the update equation of channel coefficients will then be

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{u}(\mathbf{n}) - \mu\gamma_n \nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ , \tag{15}$$

or equivalently

$$w_i(n+1) = w_i(n) + \mu e(n)u(n-i) - \rho_n \frac{w_i(n)}{\sigma_n^2} e^{-w_i^2(n)/2\sigma_n{}^2} \ , \tag{16}$$

in which $\rho_n \triangleq \mu\gamma_n$. It is important to mention that our proposed adaptive algorithm in (15) has two main advantages in comparison to other similar methods such as LMS, ZA-LMS or RZA-LMS. Firstly, we use a smooth approximation of $||\mathbf{w}(n)||_0$, which proved to be an appropriate and near-accurate approximation[6] as stated in [2], and it is also easy to work with, as one can find a continuous gradient for this smoothed norm function. Secondly, we have provided some degrees of freedom by varying the regularization factor (γ_n) and the smoothness factor (σ_n) at each step of iteration. The adaptation of these parameters will be discussed in Sect. 2.1.3, in which we show mathematically that by this adaptations the performance of our algorithm will be better than ordinary LMS in the sense of MSE.

Since $g_{\text{reg}}^{(n)}(\mathbf{w}(n))$ is a better approximation of $||\mathbf{w}(n)||_0$ for small enough σ_n than the approximations used in the ZA-LMS and RZA-LMS [2], one intuitively may expect that the SL0-LMS will have a better MSE performance than the ZA-LMS and RZA-LMS, as the smoothed l_0. As will be seen in Sect. 4, experimental results show that our algorithm has a better MSE performance as well as better learning and tracking behaviour than those mentioned algorithms. However, we do not have any rigorous mathematical proof for this at this point. Moreover, our algorithm requires adjusting the parameters γ_n and σ_n at each iteration, and thus it has more computational complexity in comparison to the LMS, ZA-LMS and RZA-LMS. This computational complexity is one of the main drawbacks of our proposed SL0-LMS.

2.1.3. Convergence and Performance Analysis of SL0-LMS

In this section, we provide an analysis on the convergence and MSE performance of SL0-LMS. Our analysis is based on "*independence theory*" [17] (called also "*independence assumption*" [18]), which is a well-known simplifying assumption used in adaptive filter's literature[7] to analyze the convergence of the LMS [18, 17, 14, 19]. More accurately, considering the channel model in (7) we have the following theorem that compares the MSE performance of SL0-LMS to that of standard LMS:

Theorem 2.1. *Assume that $\{v(n)\}$ is a Gaussian i.i.d random process with variance of σ_v^2 and $\{u(n)\}$ is a symmetric Bernoulli i.i.d random process ($u(n) \in \{+1, -1\}$ and $\mathbb{P}\{u(n) = +1\} = \mathbb{P}\{u(n) = -1\} = 0.5$). Let $\mathbf{w}(n)$ and $\mathbf{w}'(n)$ be coefficients updated by (15) and (8) by the same step size μ. Furthermore, suppose that the independence assumption is satisfied, i.e. $v(n)$, $\mathbf{u}(n)$, $\mathbf{w}(n)$ and $\mathbf{w}'(n)$ are mutually independent. Also assume that there exists a known parameter $\alpha > 0$ such that $||\mathbf{f}||_0 \le \alpha$. So, if $\mathbf{w}(0) = \mathbf{w}'(0)$, and at each iteration of (15) parameters σ_n and ρ_n are selected as*

$$\sigma_n \ge \max\left\{ \max_{0\le i\le M-1} |w_i(n)|, \max_{0\le i\le M-1} |f_i| \right\} \ , \tag{17}$$

[6] As shown in [2], (14) tends to equality when $\sigma_n \to 0$.

[7] Mathematically, the independence assumption is questionable. However, it can be justified to be a good approximation for small μ's [18, Section 6.2].

and

$$0 \le \rho_n \le \max\left\{0, 2(1-\mu)\frac{g_{reg}^{(n)}(\mathbf{w}(n)) - \alpha}{\|\nabla g_{reg}^{(n)}(\mathbf{w}(n))\|_2^2}\right\} \ , \tag{18}$$

then for every $n > 1$ we have:

$$E\{\|\mathbf{w}(n) - \mathbf{f}\|_2^2\} \le E\{\|\mathbf{w}'(n) - \mathbf{f}\|_2^2\} \ . \tag{19}$$

The proof is left to the Appendix.

Remark 1. Let $\mathbf{R_u}$ denotes the covariance matrix of $\mathbf{u}(n)$. As it has been shown in [18], if $0 < \mu < \frac{2}{3\text{tr}(\mathbf{R_u})} = \frac{2}{3M}$, then the standard LMS algorithm will converge under the independence assumption [8]. So, following the result of Theorem 2.1, one can say that if $0 < \mu < \frac{2}{3M}$ and if σ_n and ρ_n are updated using (17) and (18) respectively, then the SL0-LMS algorithm will converge.

Remark 2. As it can be seen from (18), whenever the condition $g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \le \alpha$ is satisfied, then ρ_n will be equal to zero. This can be interpreted intuitively as following: Whenever the sparsity constraint is satisfied, we will leave the regularization term and will focus on lowering the MSE. Once the condition is not satisfied, then a penalty term for decreasing sparsity will be added to the cost function.

Remark 3. As it can be seen from Theorem 2.1, while updating σ_n based on (17) will guarantee that our algorithm performs better than the ordinary LMS in the sense of MSE, it also requires knowing $\mathbf{f}$ which is completely impractical as our algorithm is going to estimate $\mathbf{f}$ at the first place. However, with a little change in the update of σ_n, one may interpret the Theorem 2.1 as the *local convergence* [18] analysis of our algorithm. More accurately, assume that instead of using (17), one updates σ_n based on the following equation:

$$\sigma_n > \max_{0 \le i \le M-1} |w_i(n)| + \epsilon \ , \tag{20}$$

in which $\epsilon > 0$ is an arbitrary small enough real number. This update equation is practical, as it does not depend on knowing $\mathbf{f}$. Moreover, local convergence can be obtained by this choice of σ_n. More accurately, assume that we know for $n = n_0$, $\mathbf{w}(n_0)$ is close enough to $\mathbf{f}$, so that $|\max_{0 \le i \le M-1} |f_i| - \max_{0 \le i \le M-1} |w_i(n_0)|| < \epsilon$. Now, it is easy to check that $\sigma_n > \max_{0 \le i \le M-1} |f_i|$, and so the condition in (17) will be satisfied. Following the same steps as in the proof of Theorem 2.1, one may conclude that our algorithm will have a better MSE than LMS for $n > n_0$, which will result in the local convergence of our algorithm for suitable small enough μ (Remark 1).

Remark 4. It is seen that choosing σ_n according to (17) guarantees the convergence of our algorithm (under independence assumption and the choice of μ according to Remark 1). However, this choice is not small enough for having a good approximation of $\|\mathbf{w}(n)\|_0$ by $g_{\text{reg}}^{(n)}(\mathbf{w}(n))$ (note however that even this choice results in a better estimation in the sense of MSE than LMS, as stated in Theorem 2.1). One may also propose to use a fixed and small value, such as $\sigma \ll 1$, for σ_n to have a better approximation $g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \simeq \|\mathbf{w}(n)\|_0$. This choice results in a non-convex regularization term, which makes theoretical convergence analysis very tricky. So, we will study this variant of the SL0-LMS algorithm only experimentally. We call this variant of the SL0-LMS as the Fixed Smoothness SL0-LMS (FS-SL0-LMS), and we name the other variant with variable σ_n (updated using (20)) as the Variable Smoothness SL0-LMS (VS-SL0-LMS). While we do not have any rigorous mathematical proof for convergence or performance improvement of the FS-SL0-LMS in comparison to the VS-SL0-LMS, we numerically show that (in Sect. 4) it has better MSE performance than the VS-SL0-LMS. It can also be experimentally seen that the FS-SL0-LMS converges slightly slower than VS-SL0-LMS.

2.2. Non-Adaptive SL0 Based Sparse Channel Estimation

In this section, we propose a non-adaptive channel estimation algorithm which will estimate the channel coefficients *after* observing $m \ll M$ successive samples of the output signal of the channel. The relation of these observations to the input sequence can be described in the matrix form as follows :

$$\mathbf{d}_{m\times 1} = \mathbf{A}_{m\times M}.\mathbf{f}_{M\times 1} + \mathbf{v} \ . \tag{21}$$

[8] As stated in [18], this range is more reliable for $0 < \mu \ll \frac{2}{3M}$, as the independence assumption is is a more accurate approximation for small μ.

in which, $\mathbf{v}$ is a vector including samples of the Gaussian noise, $\mathbf{d}$ is the vector of observations, $\mathbf{f}$ is the vector of channel taps and $\mathbf{A}$ is a Toeplitz random matrix known to the receiver whose rows are circular shift of $\mathbf{q} = [q_0, \cdots, q_{M-1}]^T$ and $\mathbf{q}$ is a M-length random block of ±1s. Consequently, if we start observing the channel output after l transmissions, then $\mathbf{A}$ can be expressed as the following matrix:

$$\mathbf{A} = \begin{bmatrix} q_l & q_{l-1} & q_{l-2} & \cdots & q_0 & q_{M-1} & \cdots & q_{l+1} \\ q_{l+1} & q_l & q_{l-1} & \cdots & q_1 & q_0 & \cdots & q_{l+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{l+m-1} & q_{l+m-2} & q_{l+m-3} & \cdots & q_{m+M-1} & q_{m+M-2} & \cdots & q_{m+l} \end{bmatrix} . \tag{22}$$

Note that in (22), $q_i = q_{i+kM}$ for every $k \in \mathbb{Z}$, as q_i is periodically repeated with a period of M. Now, (21) is a noisy under-determined system of linear equations which has to be solved under the sparsity constraints. Now, we want to find the sparsest solution of (21) while having a constraint on the square error, i.e. we are trying to solve the following optimization problem:

$$\text{argmin } \|\hat{\mathbf{f}}\|_0 \quad s.t \quad \|\mathbf{A}\hat{\mathbf{f}} - \mathbf{d}\|_2^2 \le \epsilon \ , \tag{23}$$

in which we assume that in (21), $\|\mathbf{v}\|_2^2 \le \epsilon$. There are many methods for finding the sparsest solution of a noisy under-determined system of linear equations such as the Basis Pursuit DeNoising (BPDN) [20], the Least-Absolute Shrinkage and Selection Operator (LASSO) [21] and the Robust SL0 [2, 22]. Numerically, we have seen that there is not much difference in the accuracy of the Robust SL0, LASSO and BPDN in our application and so, we use the Robust-SL0 according to its speed in comparison to the other two algorithms. So, in our experimental results we just mentioned the results of the Robust-SL0 algorithm.

3. Pre-Filtered PVA Equalization in General ISI Sparse Channels

After estimating channel coefficients using one of the methods of the previous section, high-performance equalization in *general* ISI sparse channels will be investigated in this section. One of the properties of such channels is that they have a large amount of memory while having a few number of significant coefficients. Hence, implementation of the optimum ISI equalizer (MLSE) using Viterbi algorithm is almost impossible for these channels (because computational complexity of the receiver grows exponentially with respect to channel memory [3]). The PVA, i.e. an algorithm that uses parallel trellises, can be used in a special case of such channels known as *the zero-pad channel* [15], in which the sparse channel has equally spaced coefficients. Assume that $\mathbf{w}$ is a zero-pad channel with the length $M = K.L + 1$, then it has a structure as follows:

$$\mathbf{w} = [\underbrace{w_0\, 0\, 0\, \ldots}_{K \text{ coefficients}}\ \underbrace{w_K\, 0\, 0\, \ldots}_{K \text{ coefficients}}\ \underbrace{w_{2K}\, 0\, 0\, \ldots}_{K \text{ coefficients}} \ldots\, 0\, 0\, w_{L.K}] \ . \tag{24}$$

As in [15], it is evident that received symbols $\{u_0, u_K, u_{2K} \ldots\}$ will have interference between themselves while having no interference with any other symbols, and consequently, they will generate output symbols $\{d_0, d_K, d_{2K} \ldots\}$. Similarly, symbols $\{u_1, u_{K+1}, u_{2K+1} \ldots\}$ have ISI between themselves and have no interference with any other symbols and so on. So, we can equalize this channel using K parallel trellises, each one using Viterbi algorithm for a channel with coefficients as:

$$\mathbf{w}' = [w_0\ \ w_K\ \ w_{2K}\ \ \ldots\ \ w_{L.K}]^T \ . \tag{25}$$

and the input of the jth trellis are symbols $\{d_{i.k+j} : \forall i \in \mathrm{N}\}$. So, this PVA structure will have less overall complexity than a single trellis in these channels. In fact, the complexity of a single trellis is of $O(2^{(M-1)})$ and the complexity of PVA is of $O(K.2^L)$ which is much less than the single trellis case. For example, the parameters have been chosen as $K = L = 8$ and $M - 1 = KL = 64$ in our experiment, and hence implementation of normal Viterbi algorithm which requires 2^{64} states is impossible. But having 8 trellises each with $2^8 = 64$ states is practical and possible. In the case of the general ISI sparse channels, reducing complexity of the Viterbi algorithm is much more sophisticated. For this case, according to the authors' best knowledge, no exact solution has been presented in the literature of data communications and signal processing [15]. To find a solution for the case of general ISI sparse channel, firstly we use the matched filter structure at the receiver and so, as was mentioned before, we can use the "F" model of the channel in which the channel impulse response can be assumed to be an FIR minimum-phase sparse filter. In this way, we can change the channel to be similar to the *comb* shape channel described above, i.e. the zero-pad channel. Precisely, our idea is based on a pre-filtered equalization. This pre-filter will *reshape* the channel impulse response to a zero-pad channel which has coefficients similar to the original channel, but they are moved so that the coefficients of the resulting channel will be equally spaced. In other words, assume that $\mathbf{f}$ has at most $\alpha = L + 1$ non-zero taps. We will generate a channel impulse response, i.e. $\tilde{\mathbf{f}}$, so that it has the same $L + 1$ non-zero taps, but they are equally spaced among $M = KL + 1$ channel taps. We use a pre-filter that reshape $\mathbf{f}$ to $\tilde{\mathbf{f}}$. In order to build such a filter, we use an IIR filter structure which has estimated channel ($\mathbf{f}$) as its denominator coefficients and re-shaped channel ($\tilde{\mathbf{f}}$) as its numerator coefficients. In other words we use the following filter:

$$W_{\text{pre}}(z) = \frac{\tilde{F}(z)}{F(z)} \quad . \tag{26}$$

in which $F(z)$ and $\tilde{F}(z)$ are the Z-transforms of the channel impulse response and its reshaped version respectively. According to the fact that the channel impulse response is minimum-phase, this IIR filter will be stable and causal. So, PVA can be implemented afterwards using a few number of trellises (as described in Fig. 2), which will cause a

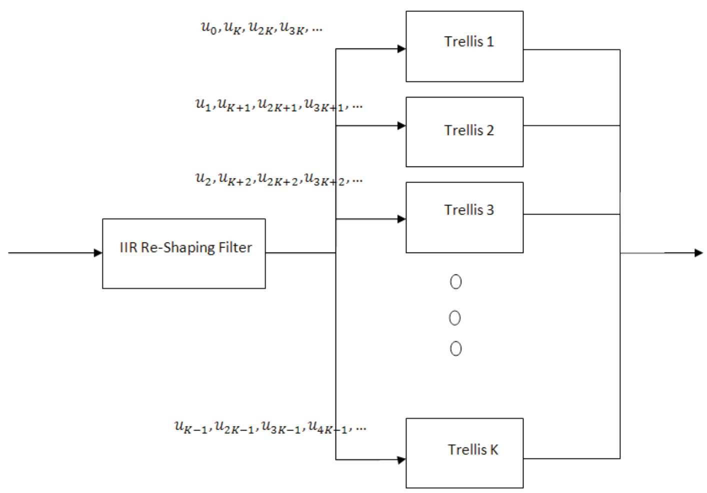


Figure 2: The proposed Pre-filtered Parallel Viterbi algorithm

great amount of computational complexity reduction; nevertheless, the equalization structure will depend on the CIR, which is one of the drawbacks of our proposed method.

4. Experimental Results

In this section, we first compare the performance of channel estimators mentioned in Sect. 2.1 which are the SL0-LMS (Eq. 15), ZA-LMS (Eq. 10), RZA-LMS (Eq. 12) and the standard LMS. In the first experiment, the input signal is an equiprobable random vector of ± 1 with length 5000 and the additive noise is white Gaussian random sequence of length 5000 and variance 10^{-3}. The channel has 256 taps where only 28 of them (randomly selected) are non zero. The value of each non-zero tap is also independently and randomly selected according to a Gaussian distribution with a mean of 0 and a variance of 1. The parameters are set as $\mu_{\text{LMS}} = \mu_{\text{ZA-LMS}} = \mu_{\text{RZA-LMS}} = \mu_{\text{SL0-LMS}} = 0.01$, $\gamma_{\text{ZA-LMS}} = \gamma_{\text{RZA-LMS}} = 0.001$, $\gamma_{\text{ZA-LMS}} = \gamma_{\text{RZA-LMS}} = 0.001$, $\varepsilon = 10$, $\sigma_0 = 0.007$ and $\alpha = \|\mathbf{f}\|_0 = 28$. We simulate the behaviour of both variants of SL0-LMS with fixed and variable smoothness factor. For the VS-SL0-LMS, σ_n is updated using (20) with $\epsilon = 0.01$. For FS-SL0-LMS, we have $\sigma_n = \sigma = 0.01$. ρ_n is updated using (18) for both of the variants. The learning curves of various adaptive estimation methods ($\mathbb{E}\{\|\mathbf{w}_n - \mathbf{f}\|^2\}$) with equal step-sizes is shown in Fig. 3.

It is clear that for this long sparse channel, the FS-SL0-LMS is drastically better than LMS,ZA-LMS and RZA-LMS, both in steady-state and convergence rate behaviour although it requires more computational complexity, as ρ_n is updated adaptively at each iteration. Furthermore, VS-SLO-LMS converges faster than LMS, ZA-LMS and RZA-LMS, while it has better steady-state behaviour that ZA-LMS and LMS. When comparing FS-SL0-LMS with VS-SL0-LMS, one can see that FS-SL0-LMS has much better MSE performance than VS-SL0-LMS, while it just has slight slower convergence behaviour.

In the second experiment, the input signal is an equiprobable random sequence of ± 1 and additive noise is a white Gaussian random sequence with variance of 10^{-3}, both with a length of 1500. The channel is an unknown time varying system, whose impulse response switches between 3 modes, each having 16 taps. In the first mode only the $5th$ tap has a value equal to 1 and the others are set to zero, so the channel is sparse. After 500 iterations, the channel switches to mode 2 that all the odd taps have a value equal to 1, while the even taps remain to be zero. In mode 3 that occurs after 1000 iterations, all the even taps are set to be -1, while all the odd taps remain 1. We run the simulation 4000 times. The parameters are set as $\mu_{\text{LMS}} = \mu_{\text{ZA-LMS}} = \mu_{\text{RZA-LMS}} = \mu_{\text{SL0-LMS,ZA}} = 0.01$, $\gamma_{\text{ZA-LMS}} = \gamma_{\text{RZA-LMS}} = 0.01$, $\gamma_{\text{ZA-LMS}} = \gamma_{\text{RZA-LMS}} = 0.001$ and $\varepsilon = 10$. The SL0-LMS algorithm in this experiment works with a fixed smoothness factor $\sigma_n = 0.02$. The upper-bound α is chosen at each mode adaptively, corresponding to the sparsity degree of the channel in that mode. The result is shown in Fig. 4. As one can see, in the first mode, when the channel is sparse, all 3 algorithm surpass standard LMS, while in mode 2 and 3, ZA-LMS does not have a good performance. Nevertheless, in all 3 modes the SL0-LMS algorithm (with fixed smoothness factor) has the best performance either in convergence rate or steady-state behaviour, even though in mode 2 and 3 the channel is completely non-sparse.

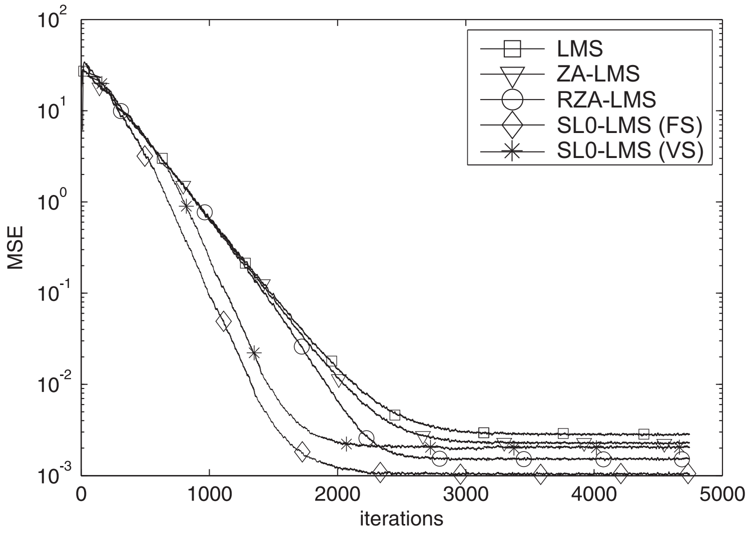


Figure 3: Learning curves of different methods.

In the third experiment the input signal is a correlated signal, generated by $u(n) = \mathcal{Q}\{0.8u(n-1)+v(n)\}$, where $v(n)$ is a white Gaussian noise and $\mathcal{Q}\{.\}$ is a hard-limiter quantizer, i.e. if $x \geq 0$, then $\mathcal{Q}\{x\} = 1$ and if $x < 0$, then $\mathcal{Q}\{x\} = -1$. The channel is the same as the second experiment, except the switch times between modes are set to 5000 and 10000. We run the simulation 2000 times. The parameters are set as $\mu_{\text{LMS}} = \mu_{\text{ZA-LMS}} = \mu_{\text{RZA-LMS}} = \mu_{\text{SL0-LMS,ZA}} = 0.01$, $\gamma_{\text{ZA-LMS}} = \gamma_{\text{RZA-LMS}} = 0.01$, $\gamma_{\text{ZA-LMS}} = \gamma_{\text{RZA-LMS}} = 0.001$ and $\varepsilon = 10$. The SL0-LMS algorithm in this experiment works with a fixed smoothness factor $\sigma_n = 0.02$. The upper-bound α is chosen at each mode adaptively, corresponding to the sparsity degree of the channel in that mode. The result is shown in Fig. 5. We can see that the SL0-LMS algorithm (with a fixed smoothness factor) has a better performance in steady-state behaviour but it is the worst in convergence rate.

After that, using simulation with a Matlab/Simulink model for concatenation of proposed channel estimators and pre-filtered PVA equalizer, we will test the efficiency of our proposed methods. We design two experiments. In these experiments, our equalization methods will be compared to the adaptive LMS Equalizer (LMSE) and an approximate Bit Error Rate (BER) bound for the MLSE introduced in [3]. To calculate this bound, suppose that the estimated path in the trellis diagram diverges from the correct path at time k and remerges again with the correct path at time

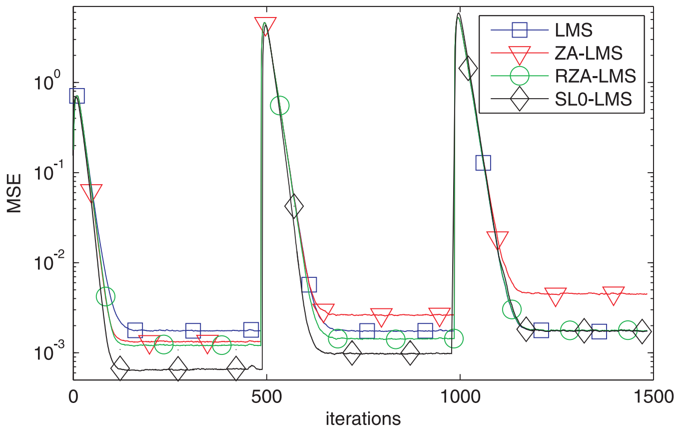


Figure 4: Tracking and steady-state behaviors of 16-order adaptive filters, driven by white input signal.

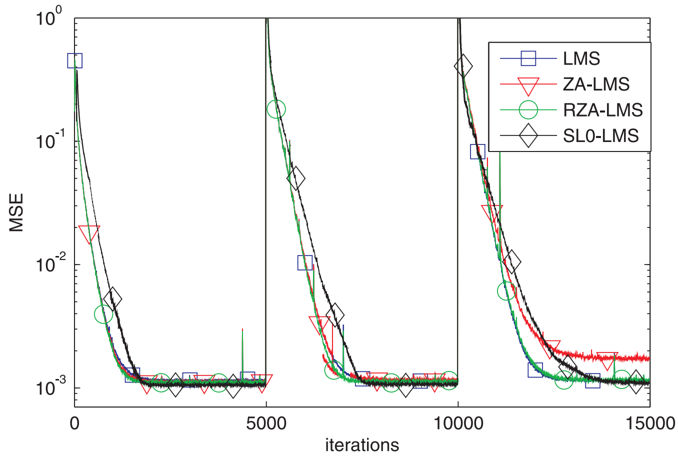


Figure 5: Tracking and steady-state behaviors of 16-order adaptive filters, driven by correlated input signal.

$k + l$. If we define $\{u(j)\}$ as original transmitted binary symbols and $\{\tilde{u}(j)\}$ as estimated binary symbols from the trellis diagram, for such an error event we have $\tilde{u}(k) \neq u(k)$ and $\tilde{u}(k + l - M) \neq u(k + l - M)$, but $\tilde{u}(j) = u(j)$ for $k - M + 1 \leq j \leq k - 1$ and $k + l - M + 1 \leq j \leq k + l - 1$. So, it is convenient to define an error vector, ϵ, corresponding to this error event as:

$$\epsilon = [\epsilon_k, \epsilon_{k+1}, \ldots, \epsilon_{k+l-M}], \tag{27}$$

where the components of ϵ are defined as:

$$\epsilon_j \triangleq \frac{1}{2}(u(j) - \tilde{u}(j)) , \tag{28}$$

For such an error event, the probability of occurrence is well approximated and upper-bounded by [3]:

$$\mathbb{P}(\epsilon) \leq Q\left(\sqrt{2\gamma_{\text{av}}\delta^2(\epsilon)}\right) \prod_{i=k}^{k+l-M} \frac{2 - |\epsilon_i|}{2} , \tag{29}$$

in which $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{t^2}{2}} d_t$, $\gamma_{\text{av}} = \frac{E_b}{N_0} = \frac{1}{N_0}$, and $\delta^2(\epsilon)$ is defined as:

$$\delta^2(\epsilon) \triangleq \mathbf{f}^T \mathbf{\Theta} \mathbf{f}, \tag{30}$$

where $\mathbf{\Theta}$ is defined as:

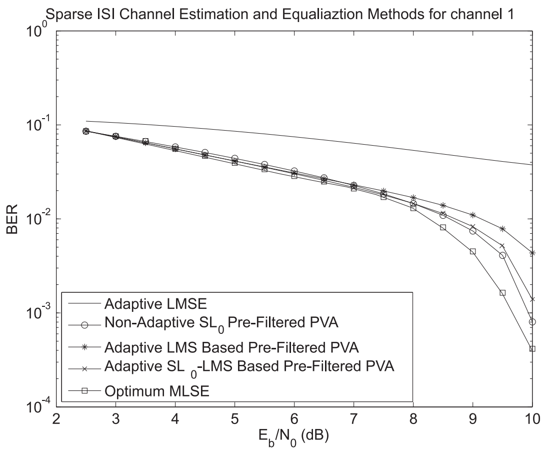
$$\mathbf{\Theta} \triangleq \begin{bmatrix} \epsilon_k & 0 & 0 & \dots & 0 & \dots & 0 \\ \epsilon_{k+1} & \epsilon_k & 0 & \dots & 0 & \dots & 0 \\ \epsilon_{k+2} & \epsilon_{k+1} & \epsilon_k & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \epsilon_{k+l-1} & \dots & \dots & \dots & \dots & \dots & \epsilon_{k+l-M} \end{bmatrix}. \tag{31}$$

Now, let E be the set of all error events ϵ starting at time k and let $w(\epsilon)$ be the corresponding number of non-zero components in each error event. Then the probability of a symbol error (in our case bit error) is upper-bounded by the union bound as:

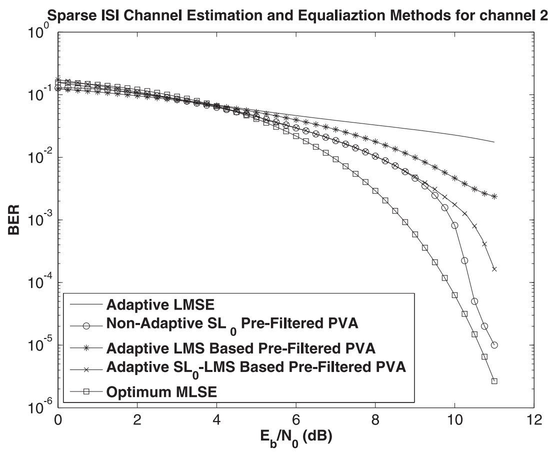
$$\begin{aligned} \mathbb{P}(e) &\leq \sum_{\epsilon \in E} w(\epsilon)\mathbb{P}(\epsilon) \\ &\leq \sum_{\epsilon \in E} w(\epsilon) Q\left(\sqrt{2\gamma_{\text{av}}\delta^2(\epsilon)}\right) \prod_{i=k}^{k+l-M} \frac{2-|\epsilon_i|}{2}. \end{aligned} \tag{32}$$

As it can be seen from (32), the MLSE upper bound for the probability of error can be calculated by computing the sum in (32). This computation requires a combinatorial search over all possible choices for ϵ. So, instead of doing a combinatorial search, we confine our search to error events with length of $l = 2, 3, 4$, and will escape from the combinatorial search by this approximation.

Moreover, channel taps in our experiments are chosen at random, with the constraint that the channel is minimum-phase according to our model. In order to generate such channels, we have generated random channels, and then have selected two of those that are minimum phase. The channels are also chosen so that channel for experiment 1 is close to the comb shape channel (taps are nearly equally spaced), while the channel for experiment 2 is not. It is also important to mention that between two proposed variants of SL0-LMS, we have chosen a fixed smoothness SL0-LMS with a fixed $\sigma_n = 0.01$ for these experiments, as it may have better estimation performance than variable smoothness SL0-LMS. The resulting BER vs Signal to Noise Ratio (SNR) per Bit is shown in Fig. 6(a) and Fig. 6(b).



(a) experiment 1.



(b) experiment 2.

Figure 6: Comparison of the BER-SNR curves for our estimation/equalization methods.

It is important to note that the derivation of the optimum MLSE's BER curves is impossible during the simulation (according to its exponential complexity with respect to channel memory, i.e. $M = 64$), and so we use approximate bounds for "Optimum MLSE" curves in these figures (as described in (32)) instead of using the exact curves. Advantages of our estimation methods and the proposed pre-filtered PVA equalizer could be seen in these results. In fact, our adaptive estimator works much better than an ordinary LMS channel estimator, while our pre-filtered PVA equalizer also performs more accurate. Indeed, the concatenation of our proposed channel estimators (both adaptive SL0-LMS and non-adaptive SL0 based channel estimators) with our proposed pre-filtered PVA equalizer has a BER near the well-approximated bound in (32), especially at low and middle SNR. Hence, our proposed communication system is near optimal at those SNRs, as MLSE is the optimum equalizer in the sense of error probability.

5. Conclusion

In this paper, we extended our work presented at LVA/ICA 2010 by providing further theoretical details and simulations. First, we proposed two sparse channel estimators, one adaptive and one non-adaptive, based on the concept of smoothed l_0 norm. Our adaptive SL0-LMS algorithm exploits the sparse structure of the channel by adding a regularization term (the smoothed l_0 norm of the channel) to the LMS cost function. Them we provided a mathematical analysis for showing the local convergence of this proposed algorithm and its performance improvement over standard LMS. We also experimentally showed that the SL0-LMS has both better tracking behaviour and steady state error, in comparison to similar algorithms like the ZA-LMS or RZA-LMS. Additionally, we proposed a simple non-adaptive channel estimator that provides an accurate estimate of the channel by observing a small number of output samples. Second, we provided a method to apply the Viterbi equalizer to the case of general sparse ISI channels, using the equivalent F-model of the channel. In fact, we introduced a pre-filtered parallel Viterbi equalization method for such channels. We showed experimentally that with the use of the proposed estimation/equalization methods, we benefit from the speed of adaptive filtering, the optimality of ML equalizer and the complexity reduction of PVA. According to that, we have no significant loss of performance at the receiver while having a large reduction in complexity. In fact, pre-filtering may increase the noise power, but as we have shown experimentally, the performance of our method is not much less than the MLSE bound and so is appropriate in the sense of error performance.

As for further research on this topic, one may find more practical solutions to the problem of equalization. Indeed, implementation of our method requires prior information about the estimated channel, and mostly to know the sparsity degree of the channel. Finding a structure which is free from such constraints is desirable in many scenarios, such as where the receiver has no prior knowledge about the degree of sparsity. Furthermore, one may try to find the necessity conditions for the convergence and performance improvement of SL0-LMS in comparison to either LMS, ZA-LMS or RZA-LMS instead of relying on sufficient conditions proposed in this paper.

Appendix

Proof of Theorem 2.1 According to (15) and (7) we have:

$$\mathbf{w}(n+1) - \mathbf{f} = \left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)(\mathbf{w}(n) - \mathbf{f}) + \mu v(n)\mathbf{u}(n) - \rho_n \nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ . \tag{33}$$

By doing some simple manipulations we have:

$$\begin{aligned} \|\mathbf{w}(n+1) - \mathbf{f}\|_2^2 = {} & (\mathbf{w}(n) - \mathbf{f})^T \left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)^2 (\mathbf{w}(n) - \mathbf{f}) + \mu^2 v^2(n)\|\mathbf{u}(n)\|_2^2 + \rho_n^2 \|\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n))\|_2^2 \\ & + 2\mu v(n)(\mathbf{w}(n) - \mathbf{f})^T \left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)\mathbf{u}(n) - 2\rho_n(\mathbf{w}(n) - \mathbf{f})^T \left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \\ & - 2\mu\rho_n v(n)\mathbf{u}^T(n)\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ . \end{aligned} \tag{34}$$

So, noting that from the independence assumption $v(n)$,$\mathbf{u}(n)$ and $\mathbf{w}(n)$ are mutually independent, we will have:

$$\begin{aligned} E\left\{\|\mathbf{w}(n+1) - \mathbf{f}\|_2^2 | \mathbf{w}(n)\right\} = {} & (\mathbf{w}(n) - \mathbf{f})^T E\left\{\left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)^2\right\}(\mathbf{w}(n) - \mathbf{f}) \\ & + \mu^2 E\left\{v^2(n)\right\} E\left\{\|\mathbf{u}(n)\|_2^2\right\} + \rho_n^2 \|\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n))\|_2^2 \\ & + 2\mu E\left\{v(n)\right\}(\mathbf{w}(n) - \mathbf{f})^T E\left\{\left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)\mathbf{u}(n)\right\} \\ & - 2\rho_n(\mathbf{w}(n) - \mathbf{f})^T E\left\{\left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}(n)^T\right)\right\}\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \\ & - 2\mu\rho_n E\{v(n)\} E\left\{\mathbf{u}^T(n)\right\}\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ . \end{aligned} \tag{35}$$

Note that $\|\mathbf{u}(n)\|_2^2 = M$ as $u(n) \in \{\pm 1\}$. Furthermore, $E\{v(n)\} = 0$ and $E\{v^2(n)\} = \sigma_v^2$, and also $E\{\mathbf{u}(n)\mathbf{u}^T(n)\} = \mathbf{I}$ as $u(n)$ is an i.i.d process. Accordingly, we also have:

$$E\left\{\left(\mathbf{I} - \mu\mathbf{u}(n)\mathbf{u}^T(n)\right)^2\right\} = \mathbf{I} - 2\mu E\{\mathbf{u}(n)\mathbf{u}^T(n)\} + \mu^2 E\{\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{u}(n)\mathbf{u}^T(n)\} = (1 - 2\mu + M\mu^2)\mathbf{I} \ . \tag{36}$$

By substituting (36) in (35) we have:

$$E\left\{\|\mathbf{w}(n+1)-\mathbf{f}\|_2^2|\mathbf{w}(n)\right\} = (1-2\mu+M\mu^2)\|\mathbf{w}(n)-\mathbf{f}\|_2^2 + 2\rho_n(1-\mu)(\mathbf{f}-\mathbf{w}(n))^T\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) + M\mu^2\sigma_v^2 + \rho_n^2\|\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n))\|_2^2 \ . \tag{37}$$

Lets define $\mathscr{S} \triangleq \left\{\mathbf{x}\in\mathbb{R}^M : \max_{0\le i\le M-1}|x_i| \le \max\left\{\max_{0\le i\le M-1}|w_i(n)|, \max_{0\le i\le M-1}|f_i|\right\}\right\}$. It is easy to check that the mentioned set of points is a convex hull. Now, if $\sigma_n \ge \max\left\{\max_{0\le i\le M-1}|w_i(n)|, \max_{0\le i\le M-1}|f_i|\right\}$, then for every $\mathbf{x}\in\mathscr{S}$, $\nabla^2 g_{\text{reg}}^{(n)}(\mathbf{x})$ is positive semi-definite, and so $g_{\text{reg}}^{(n)}$ is a convex function at $\mathbf{x}$. Moreover, $\mathbf{w}(n)\in\mathscr{S}$ and $\mathbf{f}\in\mathscr{S}$. Accordingly we have:

$$(\mathbf{f}-\mathbf{w}(n))^T\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \le g_{\text{reg}}^{(n)}(\mathbf{f}) - g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ . \tag{38}$$

As stated in [2], $g_{\text{reg}}^{(n)}(\mathbf{f}) = M - \sum_{i=1}^M e^{-w_i^2(n)/2\sigma_n^2}$ tends to $\|\mathbf{f}\|_0$ as $\sigma_n \to 0$. It is also easy to check that $g_{\text{reg}}^{(n)}$ is a decreasing function with respect to σ_n, and so we can make an upper-bound for $g_{\text{reg}}^{(n)}(\mathbf{f})$ as:

$$g_{\text{reg}}^{(n)}(\mathbf{f}) \le \|\mathbf{f}\|_0 \le \alpha \ , \tag{39}$$

and so we have:

$$(\mathbf{f}-\mathbf{w}(n))^T\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \le \alpha - g_{\text{reg}}^{(n)}(\mathbf{w}(n)) \ . \tag{40}$$

Therefore :

$$E\left\{\|\mathbf{w}(n+1)-\mathbf{f}\|_2^2|\mathbf{w}(n)\right\} \le (1-2\mu+M\mu^2)\|\mathbf{w}(n)-\mathbf{f}\|_2^2 + M\mu^2\sigma_v^2 + 2\rho_n(1-\mu)\left(\alpha - g_{\text{reg}}^{(n)}(\mathbf{w}(n))\right) + \rho_n^2\|\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n))\|_2^2 \ . \tag{41}$$

Lets define

$$h(\rho_n) \triangleq 2\rho_n(1-\mu)\left(\alpha - g_{\text{reg}}^{(n)}(\mathbf{w}(n))\right) + \rho_n^2\|\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n))\|_2^2 \ . \tag{42}$$

Now, it is easy to see that if $0 \le \rho_n \le \max\left\{0, 2(1-\mu)\frac{g_{\text{reg}}^{(n)}(\mathbf{w}(n))-\alpha}{\|\nabla g_{\text{reg}}^{(n)}(\mathbf{w}(n))\|_2^2}\right\}$ then $h(\rho_n)\le 0$ and so we have:

$$E\left\{\|\mathbf{w}(n+1)-\mathbf{f}\|_2^2|\mathbf{w}(n)\right\} \le (1-2\mu+M\mu^2)\|\mathbf{w}(n)-\mathbf{f}\|_2^2 + M\mu^2\sigma_v^2 \ . \tag{43}$$

By taking expectation with respect to $\mathbf{w}(n)$ we have:

$$E\left\{\|\mathbf{w}(n+1)-\mathbf{f}\|_2^2\right\} \le (1-2\mu+M\mu^2)E\left\{\|\mathbf{w}(n)-\mathbf{f}\|_2^2\right\} + M\mu^2\sigma_v^2 \ . \tag{44}$$

For standard LMS, it is straightforward to check that (by putting $\rho_n = 0$ in (37) and taking expectations from both sides with respect to $\mathbf{w}(n)$):

$$E\left\{\|\mathbf{w}'(n+1)-\mathbf{f}\|_2^2\right\} = (1-2\mu+M\mu^2)E\left\{\|\mathbf{w}'(n)-\mathbf{f}\|_2^2\right\} + M\mu^2\sigma_v^2 \ . \tag{45}$$

Hence, if $E\left\{\|\mathbf{w}'(0)-\mathbf{f}\|_2^2\right\} = E\left\{\|\mathbf{w}(0)-\mathbf{f}\|_2^2\right\}$, (19) can be obtained by induction on n, which completes the proof. ∎

References

[1] R. Niazadeh, S. Hamidi Ghalehjegh, M. Babaie-Zadeh, C. Jutten, Adaptive and non-adaptive ISI sparse channel estimation based on SL0 and its application in ML sequence-by-sequence equalization, in: Ninth International Conference on Latent Variable Analysis and Signal Seperation (LVA-ICA, formerly known as ICA), Springer-Verlag, 2010, pp. 579–587.

[2] H. Mohimani, M. Babaie-Zadeh, C. Jutten, A fast approach for overcomplete sparse decomposition based on smoothed l_0-norm, IEEE Trans. Signal Processing 57 (1) (2009) 289–301.

[3] J. Proakis, Digital Communication, 4th edition, Mc-Graw Hill, New york, 2001.

[4] S. V. C. Carbonelli, U. Mitra, Sparse channel estimation with zero tap detection, IEEE Trans. Communication 6 (5) (2007) 1743–11754.

[5] N. Chapman, I. Barrodale, Deconvolution of marine seismic data using the l1 norm, Geophysical Journal of the Royal Astronomical Society 72 (1) (1983) 93–100.

[6] S. Cotter, B. Rao, Sparse channel estimation via matching pursuit with application to equalization, IEEE Trans. Communication 50 (3) (2002) 374–378.

[7] S. V. W. Li, J. Preisig, Estimation of rapidly time-varying sparse channels, IEEE Journal of Oceanic Engineering 32 (4) (2007) 927–940.

[8] Y. Chen, Y. Gu, A. H. III, Sparse LMS for system identification, ICASSP, 2009, pp. 3125–3128.
[9] D. Needell, J. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Applied and Computational Harmonic Analysis 26 (3) (2009) 301–321.
[10] N. Kalouptsidis, G. Mileounis, B. Babadi, V. Tarokh, Adaptive algorithms for sparse non-linear channel estimation, Relation 10 (1) (2010) 5320–5325.
[11] M. Kocic, D. Brady, M. Stojanovic, Sparse equalization for real-time digital underwater acoustic communications, in: OCEANS'95. MTS/IEEE. Challenges of Our Changing Global Environment. Conference Proceedings., Vol. 3, IEEE, 1995, pp. 1417–1422.
[12] M. Stojanovic, L. Freitag, M. Johnson, Channel-estimation-based adaptive equalization of underwater acoustic signals, in: OCEANS'99 MTS/IEEE. Riding the Crest into the 21st Century, Vol. 2, IEEE, 1999, pp. 590–595.
[13] N. C. Mcginty, R. Kennedy, P. Hoeher, Parallel trellis viterbi algorithm for sparse channels, IEEE Communication Letters 2 (5) (1998) 143–145.
[14] B. Widrow, S. Streams, Adaptive Signal Processing, Prentice Hall, New Jersey, 1985.
[15] J. Mietzner, S. Hoeher, I. Land, P. Hoeher, Trellis-based equalization for sparse ISI channels revisited, ISIT, 2005, pp. 229–233.
[16] H. Poor, An introduction to signal detection and estimation, Springer, 1994.
[17] S. Haykin, Adaptive filter theo˜, Prentice Hall, New Jersey.
[18] B. Farhang-Boroujeny, Adaptive filters: theory and applications, Wiley, 1999.
[19] A. Feuer, E. Weinstein, Convergence analysis of lms filters with uncorrelated gaussian data, Acoustics, Speech and Signal Processing, IEEE Transactions on 33 (1) (1985) 222–230.
[20] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM review 43 (1) (2001) 129–159.
[21] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.
[22] A. Eftekhari, M. Babaie-Zadeh, C. Jutten, H. Moghaddam, Robust-SL0 for stable sparse representation in noisy settings, ICASSP, 2009, pp. 3433–3436.