New Sparse Adaptive Algorithms Based on the Natural Gradient and the L_0-Norm

Konstantinos Pelekanakis, *Member, IEEE*, and Mandar Chitre, *Senior Member, IEEE*

***Abstract*—A new algorithmic framework for sparse channel identification is proposed. Although the focus of this paper is on sparse underwater acoustic channels, this framework can be applied in any field where sequential noisy signal samples are obtained from a linear time-varying system. A suit of new algorithms is derived by minimizing a differentiable cost function that utilizes the underlying Riemannian structure of the channel as well as the L_0-norm of the complex-valued channel taps. The sparseness effect of the proposed algorithms is successfully demonstrated by estimating a mobile shallow-water acoustic channel. The clear superiority of the new algorithms over state-of-the-art sparse adaptive algorithms is shown. Moreover, the proposed algorithms are employed by a channel-estimate-based decision-feedback equalizer (CEB DFE). These CEB DFE structures are compared with a direct-adaptation DFE (DA DFE), which is based on sparse and nonsparse adaptation. Our results confirm the improved error-rate performance of the new CEB DFEs when the channel is sparse.**

***Index Terms*—Acoustic echo cancellation, improved-proportionate affine projection algorithm (IPAPA), improved-proportionate normalized least mean square (IPNLMS), L_0-norm, L_1-RRLS, proportionate algorithms, sparse equalization, sparse recursive least squares (RLS), underwater acoustic communications.**

I. INTRODUCTION

ADVANCES in underwater acoustic (UWA) communications over the last two decades have made possible to conceive various high data-rate applications in shallow-water channels [1]. While tremendous progress has been achieved, reliable phase-coherent acoustic communications still remains a challenging goal due to the complex propagation phenomena these channels exhibit. For example, in short and medium ranges, rich sound scattering off the physical boundaries generates a multipath spread on the order of tens of milliseconds. In addition, sea surface motion could induce a Doppler spread on the order of 10 Hz [2]. Consequently, such shallow-water channels could have a spread factor (i.e., the product of delay spread with Doppler spread) close to one, and so reliable channel estimation is often infeasible by employing standard adaptive filters [3]. Moreover, the long delay spread (usually hundreds of symbols) may render a coherent receiver prohibitively complex if special attention on the channel estimation algorithm is not paid.

A remedy for improved channel tracking and reduced receiver complexity is by exploiting the sparse multipath structure of the acoustic link [4]. In particular, when the UWA channel is modeled by the delay-spread function, one notes that a large fraction of its energy is concentrated in a small fraction of its duration. Current research on exploiting the inherent channel sparseness proposes matching pursuit (MP) algorithms [5], [6], a hierarchical Bayesian model [7], an iterative detector/estimator (IDE) algorithm [8], and a mixture of L_2- and L_1-norms [9]. Although the above approaches have shown promising results in various shallow-water channels, some notable issues must be mentioned here.

The proposed algorithms in [5] and [6] are very sensitive to the prior choice of the number of the active channel taps (i.e., the nonzero taps) and the stopping criteria. The algorithms in [7] and [8] require cubic computational complexity (with respect to the number of channel taps), which may become prohibitively costly for long delay-spread channels. Furthermore, the model in [7] is based on the assumption that the channel taps are Gaussian random variables, yet, the results in [2] challenge the applicability of the central limit theorem since intermittent distinct arrivals were identified in the channel response of a surf zone environment. To address the time-varying complexities of shallow-water channels, Li and Preisig [5] and Sen Gupta and Preisig [9] employ the delay-Doppler-spread function. In certain wideband channels, this approach could increase the number of channel parameters to over 1000, which is not desirable since the proposed algorithms in [5] and [9] are not linear in complexity. Last, a common characteristic of all the above algorithms is that they use batch optimization, i.e., algorithm optimization is solved for a fixed set of received channel symbols (typically hundreds of them).

In this work, we follow an online optimization strategy, namely, at every received channel symbol, the algorithm tries to generate a better channel estimate. Online optimization is the *de facto* standard in acoustic echo cancellation (AEC) applications where sparse impulse responses of 1000 taps are often encountered [10, Ch. 4]. Motivated by the need of linear algorithmic complexity, sparse versions of the standard normalized least-mean-square (NLMS) algorithm and the affine projection algorithm (APA) emerged. These algorithms are the so-called

Manuscript received March 08, 2012; revised July 09, 2012; accepted September 26, 2012. Date of publication November 30, 2012; date of current version April 10, 2013. This work was supported by the Ministry of Education (MOE) under Grant R-263-000-521-133.

Associate Editor: M. Stojanovic.

K. Pelekanakis is with the Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore, Singapore 119223, Singapore (e-mail: costas@arl.nus.edu.sg).

M. Chitre is with the Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore, Singapore 119223, Singapore and also with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576, Singapore (e-mail: mandar@arl.nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JOE.2012.2221811

improved-proportionate NLMS (IPNLMS) algorithm [11] and the improved-proportionate APA (IPAPA) [12], respectively. Both IPNLMS and IPAPA have been successfully applied in identifying a shallow-water acoustic link [13]. Although these algorithms exploit sparseness using natural gradient (NG) adaptation, they do not capitalize on the sparseness effect of L_0/L_1-norm.

Recently, L_0/L_1-norm-based sparse adaptive algorithms have been proposed in the AEC literature. A differentiable approximation of the L_0-norm is used to regularize the LMS cost function in [14]. A sparse recursive least squares (RLS) algorithm was proposed in [15], where Angelosante *et al.* solved the L_1-norm regularized RLS cost function by using an efficient least absolute shrinkage and selection operator (Lasso) approach. Eksioglu [16] used a reweighted L_1-norm within the RLS cost function and the resulting algorithm, termed as L_1-RRLS, showed better performance than the sparse RLS (SPARLS) [17]. In the UWA communications arena, the authors proposed an L_0-norm constrained IPNLMS algorithm [18], an enhancement of IPNLMS-ℓ_0 [19].

This paper proposes a framework for generating adaptive algorithms for complex-valued sparse channel estimation. Although the current focus is on UWA channels, this algorithmic framework can be readily applied to any linear time-varying channels. After describing the system model and NG adaptation in Section II, the framework is analyzed in Section III. The derived algorithms utilize NG adaptation (i.e., the channel impulse response lies in a Riemannian space) and an L_0-norm proxy of the channel vector. In Section IV, we compare the new algorithms with L_1-RRLS, RLS, IPNLMS, and IPAPA in a channel identification experiment by using both field and simulation data. In addition, the new algorithms are used by a channel-estimate-based decision-feedback equalizer (CEB DFE). These CEB DFE structures are compared with a direct-adaptation DFE (DA DFE) based on the RLS algorithm and a new sparse DA DFE based on the IPAPA. The paper is concluded in Section V.

Notation and Definitions: Superscripts $^\intercal$, †, and * stand for transpose, Hermitian transpose, and conjugate, respectively. Column vectors (matrices) are denoted by boldface lowercase (uppercase) letters. The $N \times N$ identity matrix is denoted as $\mathbf{I}_N$. The L_1-norm of a complex number x is defined as $|x|_1 \triangleq |\mathrm{Re}\{x\}| + |\mathrm{Im}\{x\}|$. The complex sign function of x is defined as $c\mathrm{sgn}(x) \triangleq \mathrm{sgn}(\mathrm{Re}\{x\}) + j \cdot \mathrm{sgn}(\mathrm{Im}\{x\})$, where $\mathrm{sgn}(\cdot)$ stands for the sign function of a real number. The L_0-norm of a K-tap complex vector $\mathbf{x}$ is denoted as $\|\mathbf{x}\|_0$ and is equal to the number of the nonzero taps of $\mathbf{x}$. The L_1-norm of $\mathbf{x}$ is defined as $\|\mathbf{x}\|_1 \triangleq \sum_{i=0}^{K-1} |x_i|_1$. The gradient of a scalar function $f(\mathbf{x})$ with respect to $\mathbf{x}$ is denoted as $\nabla_{\mathbf{x}}(f(\mathbf{x}))$.

II. System Model and Preliminaries

The baseband (complex) representation of the channel impulse response, input/output signals, and additive noise process will be used throughout this paper. We assume that the output (received) signal is sampled at the Nyquist rate and is given in vector form by

$$y[n] = \mathbf{h}[n]^\dagger \mathbf{u}[n] e^{j\theta[n]} + w[n] \tag{1}$$

where $\mathbf{h}[n] \triangleq [h_0[n]\, h_1[n] \dots h_{K-1}[n]]^\intercal$ denotes the channel impulse response at discrete time n, $\mathbf{u}[n] \triangleq [u[n]\, u[n-1] \dots u[n-K+1]]^\intercal$ contains the K most recent samples of the input signal, $\theta[n]$ is the residual carrier phase occurring after imperfect Doppler compensation and/or mismatch between the transmitter and receiver sampling clocks, and $w[n]$ denotes the additive noise. We implicitly assume that $\theta[n]$ varies much faster than $\mathbf{h}[n]$, and thus, it can be estimated separately. Moreover, we assume that $\mathbf{h}[n]$ is a sparse vector for every n, namely, most of the coefficients are close to zero and only few of them are large.

In this work, we borrow ideas from the AEC paradigm to design our own channel estimation algorithms. Although UWA channels are substantially different from AEC channels in terms of multipath formation and ambient noise, they share a common characteristic: long and sparse impulse responses [10]. Sparse adaptive algorithms for AEC applications have been developed for over a decade. Perhaps, the IPAPA [12] stands out as the most prominent sparse adaptive algorithm with linear computational complexity. For our purposes, we extend the IPAPA to include complex-valued impulse responses. Let $\hat{\mathbf{h}}[n]$ denote the estimated channel response, then the IPAPA channel update equations are given by

$$\mathbf{U}[n] = \left[\mathbf{u}[n]e^{j\theta(n)} \dots \mathbf{u}(n-L+1)e^{j\theta(n-L+1)}\right] \tag{2}$$

$$\mathbf{y}[n] = [y[n]\, y[n-1] \dots y[n-L+1]]^\intercal \tag{3}$$

$$\mathbf{e}[n]^* = \mathbf{y}[n]^* - \mathbf{U}[n]^\dagger \hat{\mathbf{h}}[n-1] \tag{4}$$

$$\delta = (1-\beta)\delta_{\mathrm{NLMS}}/2K, \qquad \beta \in [-1, 1] \tag{5}$$

$$\mathbf{B}[n] = (\mathbf{U}[n]^\dagger \mathbf{G}_{\hat{\mathbf{h}}}[n-1]^{-1}\mathbf{U}[n] + \delta\, \mathbf{I}_L)^{-1} \tag{6}$$

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu \mathbf{G}_{\hat{\mathbf{h}}}[n-1]^{-1}\mathbf{U}[n]\mathbf{B}[n]\mathbf{e}[n]^* \tag{7}$$

where $\mathbf{U}[n]$ is the $K \times L$ matrix of input samples, $\mathbf{y}[n]$ is the vector that contains the last L output samples, $\mathbf{e}[n]$ is the $L \times 1$ *a priori* error vector, $\mu \in [0, 1]$ is a step-size constant, $\delta_{\mathrm{NLMS}} > 0$ is the regularization parameter of the NLMS algorithm, and $\mathbf{G}_{\hat{\mathbf{h}}}[n]^{-1}$ is a $K \times K$ diagonal matrix whose diagonal entries depend on $\hat{\mathbf{h}}[n]$. In particular, the diagonal entries $g_k[n], 0 \le k \le K-1$ of $\mathbf{G}_{\hat{\mathbf{h}}}[n]^{-1}$ are found using the procedure [11]

$$\gamma_k[n] = (1-\beta)\frac{\|\hat{\mathbf{h}}[n]\|_1}{K} + (1+\beta)|\hat{h}_k[n]|_1 \tag{8}$$

$$g_k[n] = \frac{\gamma_k[n]}{\sum_{i=0}^{K-1} \gamma_k[n]}. \tag{9}$$

Substituting (8) in (9), we have that

$$g_k[n] = \frac{1-\beta}{2K} + (1+\beta)\frac{|\hat{h}_k[n]|_1}{2\|\hat{\mathbf{h}}[n]\|_1 + \alpha} \tag{10}$$

where α denotes a small positive constant to avoid division by zero during initialization of the algorithm. The parameter β controls the sparseness of $\hat{\mathbf{h}}[n]$. For sparse channels, β should be chosen close to 1, while for nonsparse channels, $\beta = -1$.

It is worthy to note that if the L_1-norm in (8) is replaced by a different norm, new algorithms will emerge. For instance, the L_1-norm was replaced by the L_0-norm in [19]. In addition, the IPAPA reduces to several known algorithms as follows:

- when $L = 1$, IPAPA reduces to the IPNLMS algorithm [11];
- when $\beta = -1$, IPAPA reduces to APA [20, Ch. 6];
- when $L = 1$ and $\beta = -1$, IPAPA reduces to the NLMS algorithm [20, Ch. 6].

The sparseness effect of the IPAPA is achieved due to the *proportionate* matrix $\mathbf{G}_{\hat{\mathbf{h}}}[n]^{-1}$. The proportionate term is coined to signify that at every iteration, $\mathbf{G}_{\hat{\mathbf{h}}}[n]^{-1}$ assigns a combination of fixed and variable step-size parameters at each filter tap. The variable step-size parameter is a function of the tap's previously estimated magnitude according to (10). Correspondingly, active filter taps converge fast, which makes the overall algorithm to have faster initial convergence than the APA. The fixed step-size parameter makes the algorithm to have robust performance in nonsparse channels.

Although the IPAPA was initially derived without minimizing a cost function, it is essentially an algorithm that leverages on NG adaptation [21]. In particular, if one assumes that the underlying space between $\hat{\mathbf{h}}[n]$ and $\hat{\mathbf{h}}[n-1]$ is warped, i.e., Riemannian, then $\mathbf{G}_{\hat{\mathbf{h}}}[n]$ is a positive–definite matrix that describes the curvature of that space, i.e., $\mathbf{G}_{\hat{\mathbf{h}}}[n]$ is a Riemannian metric tensor. The fact that $\hat{\mathbf{h}}[n]$ lies in a warped space is based on prior knowledge that $\mathbf{h}[n]$ lies close to some axis of $\mathbb{C}^K$ since most of the filter taps must be close to zero. One way to visualize this warped space is as follows: for a region close to the axes, any direction orthogonal to those axes should be larger than the ordinary Euclidean distance [22]. In addition, for Riemannian spaces, the ordinary (Euclidean) gradient does not represent the steepest ascent direction but rather the NG does so. If $J[n]$ denotes a differentiable cost function associated with $\hat{\mathbf{h}}[n]$, then the NG update of $\hat{\mathbf{h}}[n]$ is given by [21]

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu \mathbf{G}_{\hat{\mathbf{h}}}[n-1]^{-1} \cdot \nabla_{\mathbf{h}^*}(J[n-1]). \tag{11}$$

One notes that (7) and (11) have similar form. In Section III, we formalize the connection between IPAPA and NG adaptation based on a new cost function.

An alternative way to interpret $\mathbf{G}_{\hat{\mathbf{h}}}[n]$ is that of Bayesian priors, namely, each tap is assumed to be a random variable with a known prior probability density function (pdf) that captures sparseness [23]. Sparseness is achieved by assuming that with high probability the tap will have a small value, while with low probability the tap will have a large value. For more information about how prior knowledge is encoded in the Riemannian space of $\hat{\mathbf{h}}[n]$, the interested reader is directed to [24]

III. New Algorithmic Framework

We now introduce the proposed framework that combines NG adaptation and L_0-norm regularization. Let the channel update vector be expressed as

$$\mathbf{r}[n] = \hat{\mathbf{h}}[n] - \hat{\mathbf{h}}[n-1]. \tag{12}$$

Also, let the $L \times 1$ *a posteriori* error vector be defined as

$$\bar{\mathbf{e}}[n]^* = \mathbf{y}[n]^* - \mathbf{U}[n]^\dagger \hat{\mathbf{h}}[n] \tag{13}$$
$$= \mathbf{e}[n]^* - \mathbf{U}[n]^\dagger \mathbf{r}[n] \tag{14}$$

where (14) follows from substituting (12) and (4) into (13). According to Kivinen and Warmuth [26], an efficient adaptive algorithm must be conservative (avoid radical changes of $\hat{\mathbf{h}}[n]$ from one iteration to the next) and corrective (ensure better channel estimate if the same input and output were to be observed at two consecutive times). Toward this end, we consider the following cost function:

$$J[n] = \|\bar{\mathbf{e}}[n]\|_2^2 + \delta \mathbf{r}[n]^\dagger \mathbf{G}_{\hat{\mathbf{h}}}[n-1]\mathbf{r}[n] + \gamma \|\hat{\mathbf{h}}[n]\|_0 \tag{15}$$

where δ and γ are positive regularization parameters. The term $\|\bar{\mathbf{e}}[n]\|_2^2$ in (15) ensures better channel estimates if the same input matrix $\mathbf{U}$ and output vector $\mathbf{y}$ were to be observed at two consecutive symbol periods. The regularizing term $\mathbf{r}[n]^\dagger \mathbf{G}_{\hat{\mathbf{h}}}[n-1]\mathbf{r}[n]$ denotes the Riemannian distance between $\hat{\mathbf{h}}[n]$ and $\hat{\mathbf{h}}[n-1]$ and ensures the conservativeness of the algorithm. The regularizing term $\|\hat{\mathbf{h}}[n]\|_0$ is used to further accelerate the convergence of the zero filter taps. Note that if different matrices $\mathbf{G}[n]$ and L_0-norm proxies are employed, different algorithms will be generated. For the remainder of this paper, $\mathbf{G}_{\hat{\mathbf{h}}}[n]^{-1}$ and δ are given by (10) and (5), respectively. In addition, $\|\hat{\mathbf{h}}[n]\|_0$ is approximated by the differentiable function[1]

$$\|\hat{\mathbf{h}}[n]\|_0 \simeq \sum_{k=0}^{K-1} 1 - e^{-\eta |\hat{h}_k[n]|_1}, \qquad \eta > 0 \tag{16}$$

which is a complex extension of the real L_0-norm used in [14].

A few remarks regarding the parameters L, γ, and η are in order. It is known (from APA [20, Ch. 6]) that a longer observation window L increases the converge rate of the algorithm when the input signal correlation matrix has a large eigenvalue spread. On the other hand, a large L degrades the algorithm performance when the UWA channel exhibits rapid fluctuations. In fast-varying channels, a large L captures the average behavior of the channel, which is less sparse than the true snapshot of the channel. The parameter $\epsilon = 1/\eta$ defines the interval $[-\epsilon, \epsilon]$ such that all $\mathrm{Re}\{\hat{h}_k[n]\}$ and $\mathrm{Im}\{\hat{h}_k[n]\}$ that fall within that interval are attracted toward the zero value [14]. A thumbrule suggests to choose ϵ two or three times less than the smallest nonzero tap. Note that for $\gamma \neq 0$, $J[n]$ is not a convex cost function and so the algorithm could theoretically diverge from the global minimum. However, if the regularization parameter γ is chosen small enough (typically close to 10^{-4}), the nonconvexity of $J[n]$ does not become an issue as our results below show.

In Appendix A, the new algorithm, the so-called L_0-IPAPA, is derived by computing $\nabla_{\mathbf{r}[n]^*} J[n] = 0$ and $\nabla_{\hat{\theta}[n]} J[n] = 0$. Since $\mathbf{G}_{\hat{\mathbf{h}}}[n]$ is diagonal and $L \ll K$, the required number of complex multiplications for L_0-IPAPA is $O(K^2(L+1) + 2K(L^2+L+1))$. Moreover, L_0-IPAPA reduces to other new or already known algorithms as follows:

- when $\gamma = 0$, L_0-IPAPA reduces to IPAPA with complexity $O(K(L^2+2L+2))$;

[1]Strictly speaking, the function is not differentiable at zero, but this is not a problem in practice since we allow the channel taps to be arbitrarily close to zero.

TABLE I
LIST OF ALGORITHM ABBREVIATIONS

Abbreviation	Algorithm name	Reference
RLS	Recursive least-squares	[20]
L_1-RRLS	Reweighted recursive least-squares with L_1 norm	[16]
NLMS	Normalized least-mean-square	[20]
L_0-NLMS	Normalized least-mean-square with L_0 norm	
APA	Affine projection algorithm	[20]
L_0-APA	Affine projection algorithm with L_0 norm	
IPNLMS	Improved-proportionate normalized least-mean-square	[11]
L_0-IPNLMS	Improved-proportionate normalized least-mean-square with L_0 norm	
IPAPA	Improved-proportionate affine projection algorithm	[12]
L_0-IPAPA	Improved-proportionate affine projection algorithm with L_0 norm	

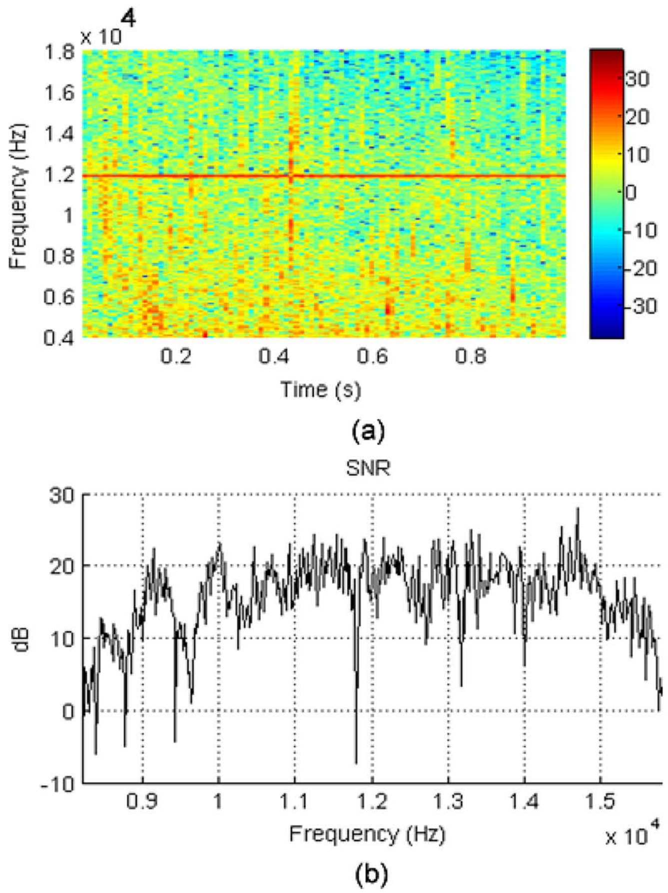

Fig. 1. (a) Spectrogram of ambient plus sensor noise. The colorbar is in a decibel scale. (b) SNR as a function of frequency.

- when $\beta = -1$, L_0-IPAPA reduces to a new algorithm, L_0-APA with complexity $O(K^2(L+1)+K(2L^2+L+1))$;
- when $L = 1$, L_0-IPAPA reduces to L_0-IPNLMS [18] with complexity $O(2K^2 + 6K)$;
- when $L = 1$, $\beta = -1$, L_0-IPAPA reduces to a new algorithm, L_0-NLMS with complexity $O(2K^2 + 4K)$;
- when $L = 1$, $\delta = 0$, L_0-IPAPA reduces to the ℓ_0-LMS algorithm [14].

IV. EXPERIMENTAL AND SIMULATION RESULTS

The goal of this section is twofold. First, the new sparse algorithms derived from our framework are applied in a UWA channel identification experiment. Their estimation accuracy is benchmarked against sparse and nonsparse adaptive algorithms. All employed algorithms of this section are summarized in Table I. Second, the channel estimation accuracy of all algorithms is associated with the error-rate performance of a DFE receiver. We use both experimental and simulated data to support our findings. Unless otherwise stated, the algorithm parameters are chosen as follows:

- $\mu = 0.2$ and $\delta_{\mathrm{NLMS}} = 10\sigma_t^2$ (σ_t^2 being the power of the transmitted signal) for NLMS, APA, IPNLMS, IPAPA, L_0-NLMS, L_0-APA, L_0-IPNLMS, and L_0-IPAPA;
- $\gamma = 4 \times 10^{-4}$ for L_0-NLMS, $\gamma = 10^{-4}$ for L_0-IPNLMS with $\beta = 0$, and $\gamma = 5 \times 10^{-5}$ for L_0-IPNLMS with $\beta = 0.5$;
- $\gamma = 10^{-3}$ for L_0-APA, $\gamma = 4 \times 10^{-4}$ for L_0-IPAPA with $\beta = 0$, and $\gamma = 10^{-4}$ for L_0-IPAPA with $\beta = 0.5$;
- $\gamma = 2$ for L_1-RRLS;
- $L = 3$ for APA, L_0-APA, IPAPA, and L_0-IPAPA;
- $\lambda = 0.999$ and $\delta = 10\sigma_t^2$ for RLS and L_1-RRLS;
- $\epsilon = 0.1$ for L_0-NLMS, L_0-IPNLMS, L_0-APA, L_0-IPAPA, and L_1-RRLS;
- $K_1 = 0.01$ and $K_2 = K_1/100$ for all algorithms.

A. Experimental Results

The data were recorded during the Focused Acoustic Fields (FAF) experiment off the coast of Pianosa Island, Italy, on July 22, 2005. The transmitter was attached on the hull of the research vessel *Leonardo*, 4.5 m below the sea level. The receiver was a 0.75-m-long 16-element horizontal linear array mounted on the bow of moving autonomous undersea vehicle (AUV). The interelement spacing of the linear array was 5 cm. The range of the link was approximately 700 m, the sea depth was 85 m, and the sound-speed profile was downward refracting. The transmitted channel symbol stream was a continuous repetition of a 6250-symbols/s-rate, quadrature phase-shift keying (QPSK)-modulated pseudonoise (PN) sequence. The PN-sequence was pulse shaped by a square-root cosine filter with roll-off factor 0.25 and truncation length ± 4 symbol intervals. The resulting waveform was modulated onto a 12-kHz carrier frequency. The operational bandwidth was 7.812 kHz.

Before we proceed with our findings, it is instructive to report on the signal-to-noise ratio (SNR) the receiver experienced. Fig. 1(a) is generated using data from the outermost sensor of the linear array. Clearly, the noise exhibits nonstationary statistics. This is mainly due to the AUV thruster and a nearby patrolling speedboat. Moreover, a strong tonal appears at 11.8 kHz due to

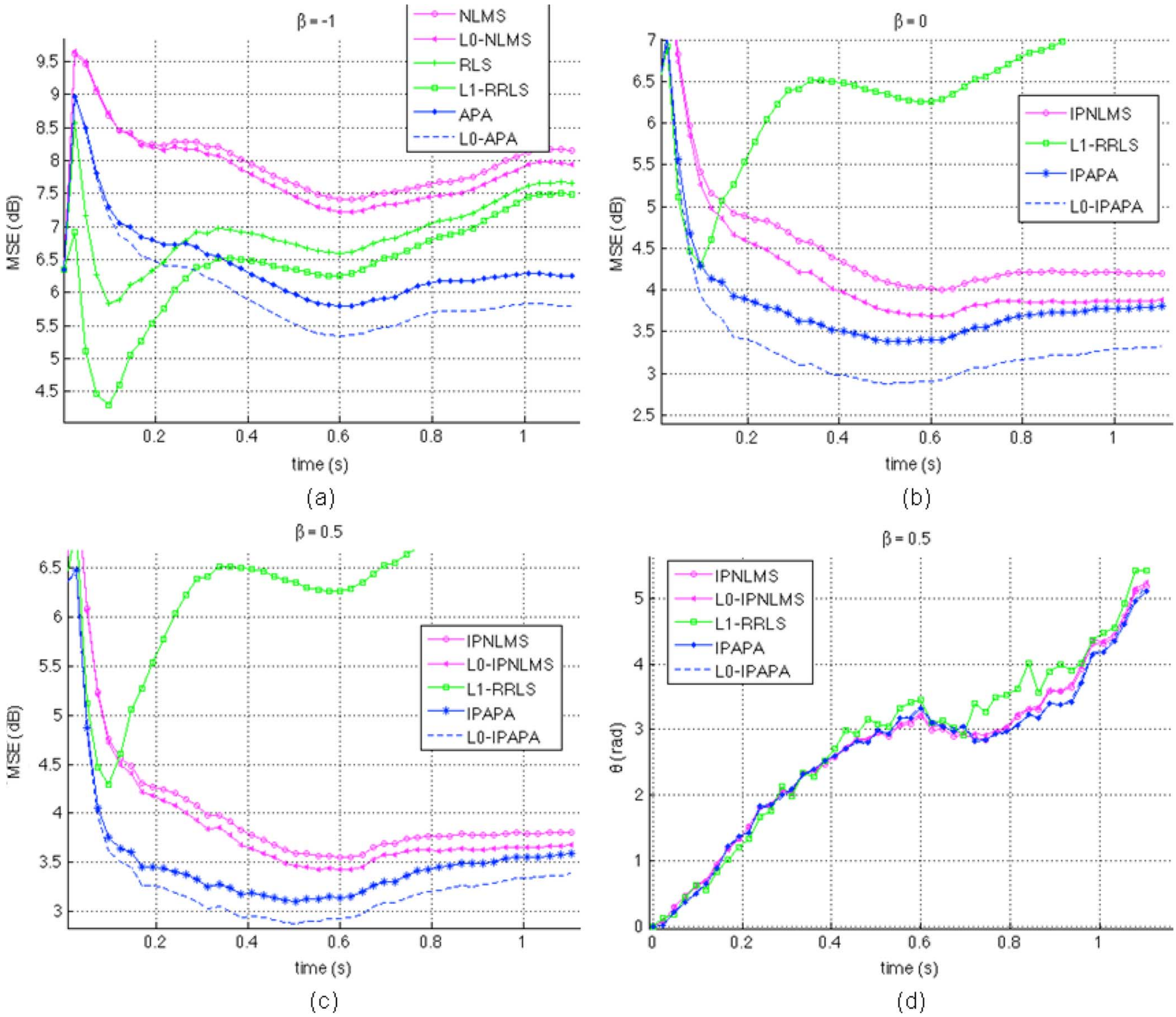

Fig. 2. (a) Learning curves for L_1-RRLS, RLS, APA, L_0-APA, NLMS, and L_0-NLMS. (b)–(c) Learning curves for L_1-RRLS and all NG algorithms with $\beta = 0$ and 0.5. (d) Carrier-phase estimate for L_1-RRLS and all NG algorithms with $\beta = 0.5$.

the array electronics. Fig. 1(b) corresponds to the same sensor and shows the SNR across the entire bandwidth averaged over 1000 symbol intervals. The average SNR is about 15 dB.

Before channel estimation, the received PN sequence is shifted to baseband, lowpass filtered, and downsampled to 2 samples/symbol (s/s). The mean squared *a priori* error (MSE), defined as

$$\frac{1}{N}\sum_{n=1}^{N}\left|y[n]-\hat{\mathbf{h}}[n-1]^{\dagger}\mathbf{u}[n]e^{j\hat{\theta}[n]}\right|^{2} \tag{17}$$

is utilized as a performance metric. The estimated channel $\hat{\mathbf{h}}[n]$ is updated at the symbol rate. Fig. 2(a)–(c) illustrates results of all employed algorithms over a period of 7000 QPSK symbols. These results correspond to the channel seen by the outermost sensor of the linear array.

Fig. 2(a) compares L_1-RRLS with RLS, APA, L_0-APA, NLMS, and L_0-NLMS. Note that L_1-RRLS shows the fastest convergence rate but its performance degrades after 0.2 s. The simulation results below show that this happens due to the ill-conditioning of the correlation matrix of $\mathbf{u}[n]$. The sparseness effect of the L_1-norm causes L_1-RRLS to outperform RLS (recall that L_1-RRLS reduces to RLS when $\gamma = 0$ [16]). L_0-APA exhibits 1.5-dB faster convergence rate than that of L_0-NLMS and shows the best steady-state tracking from the rest of the algorithms. Moreover, the sparseness effect of the L_0-norm renders L_0-NLMS and L_0-APA about 0.2 and 0.5 dB better in steady-state than NLMS and APA, respectively.

Fig. 2(b) and (c) compares L_1-RRLS with all NG algorithms for $\beta = 0$ and $\beta = 0.5$, respectively. Clearly, L_0-IPAPA outperforms all other algorithms. The convergence rate of L_0-IPAPA is about 1 dB faster than that of L_0-IPNLMS. As β increases from 0 to 0.5, the convergence rate of the L_0-IPAPA slightly improves. In contrast, all other NG algorithms show about 0.5-dB improved performance in both convergence rate and steady-state tracking. These results validate that the UWA

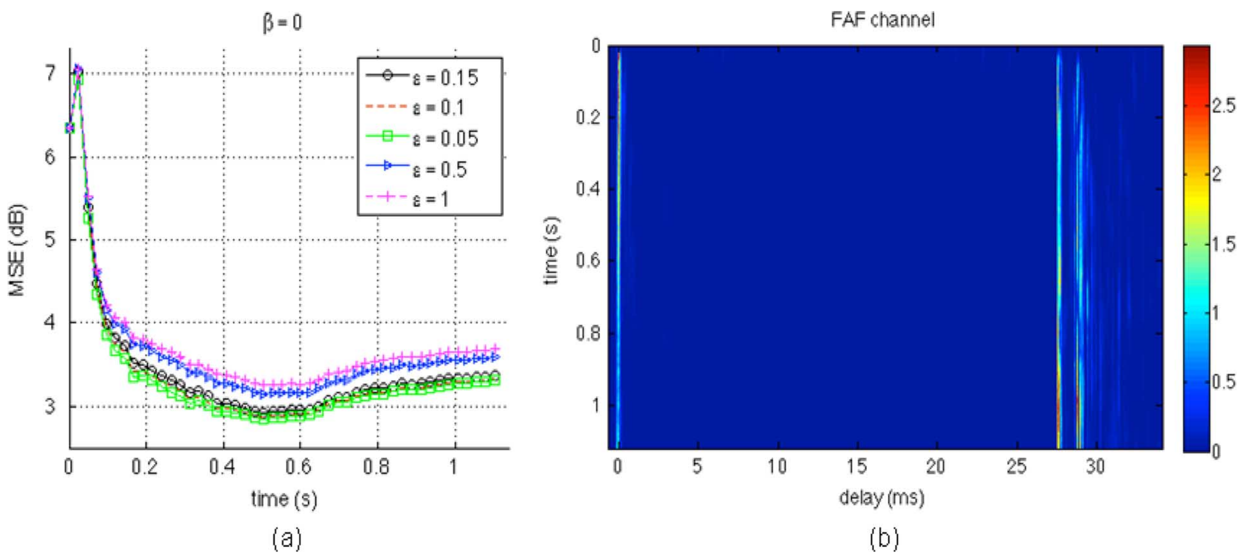

Fig. 3. (a) Learning curves of L_0-IPAPA for $\beta = 0$ and $\epsilon = 0.05, 0.1, 0.15, 0.5$, and 1. (b) Snapshots of the amplitude of the FAF impulse response. The horizontal axis represents multipath delay and the vertical axis represents absolute time. The colorbar is in linear scale. The snapshots are generated at the symbol rate. The figure is generated using L_0-IPAPA with $\beta = 0$ and $\epsilon = 0.1$.

channel has an underlying Riemannian structure that can be exploited. Moreover, the combination of NG adaptation and L_0-norm improves channel estimation performance.

Fig. 2(d) shows the carrier-phase estimate $\hat{\theta}[n]$ versus time for L_1-RRLS and all NG algorithms with $\beta = 0.5$. This estimate corresponds to the mean Doppler the received signal experienced due to the motion of the AUV. All NG algorithms present similar estimates, which is rather plausible, since their corresponding MSE differences are less than 1 dB at steady state. For brevity, the plots of the carrier-phase estimates for other choices of β are omitted since they convey the same information.

Fig. 3(a) tests the MSE of L_0-IPAPA for $\beta = 0$ and different values of ϵ. For $\epsilon \leq 0.15$, the MSE performance is slightly affected. A small decrease in performance is observed when $\epsilon \geq 0.5$, i.e., when ϵ is closer to the nonzero taps. Overall, these results show that L_0-IPAPA is robust on the choice of ϵ.

The time evolution of the estimated channel impulse amplitude can be seen in Fig. 3(b). For a sampling rate of 2 samples/symbol, the delay spread of the adaptive receiver is 434 taps. A 26-ms quiet period between the direct arrival and later multipath arrivals can be clearly seen rendering the channel sparse. It is obvious that the AUV motion renders channel estimation challenging since any adaptive algorithm must cope with the time-varying sparseness of the channel.

We now report on the error rate of a communication receiver that employs the above algorithms to adapt a DFE. The confluence of low SNR and platform motion led us to employ four hydrophones from the linear array. The DFE echoes the structure in [27], namely, the intersymbol interference (ISI) is canceled by combining previous channel estimates and symbol decisions before adaptive feedforward (FF) equalization. We noticed that the DFE Doppler tolerance can be maintained for the first 3000 symbols before performance degrades. This is explained by recalling that the receiver compensates only for the mean Doppler and not for the actual Doppler spread. A 13-tap FF filter is associated with each sensor and is centered around the direct arrival. The four FF filters are jointly adapted via the RLS algorithm

TABLE II
EXPERIMENTAL RESULTS FOR ALL DFEs

	SER / SNR_{out} (dB)	SER / SNR_{out} (dB)
	scenario A	scenario B
L_1-RRLS	0.0228 / 7.9	0.0144 / 8.4
$\beta = -1$		
NLMS	0.0276 / 7.5	0.0096 / 8.5
L_0-NLMS	0.0256 / 7.7	0.0064 / 9.0
APA	0.0232 / 7.7	0.0084 / 8.8
L_0-APA	0.0136 / 8.2	0.0048 / 9.3
$\beta = 0$		
IPNLMS	0.0164 / 8.5	0.0032 / 9.7
L_0-IPNLMS	0.0132 / 8.6	0.0028 / 9.8
IPAPA	0.0184 / 8.4	0.0024 / 9.7
L_0-IPAPA	0.0212 / 8.5	0.0020 / 9.9
$\beta = 0.5$		
IPNLMS	0.0144 / 8.5	0.0032 / 9.8
L_0-IPNLMS	0.0120 / 8.6	0.0024 / 9.8
IPAPA	0.0132 / 8.7	0.0032 / 9.8
L_0-IPAPA	0.0184 / 8.6	0.0032 / 9.9
DA-RLS	0.0388 / 7.2	0.0252 / 7.7
DA-IPAPA	0.0352 / 7.5	0.0152 / 8.4

($\lambda = 0.996$). The DFE performance is computed in terms of the symbol error rate (SER) and the average output SNR (per symbol). The latter is denoted as SNR_{out} and is a measure of how efficiently the DFE removes the ISI [25]. Table II lists the results for each channel estimation algorithm. The middle

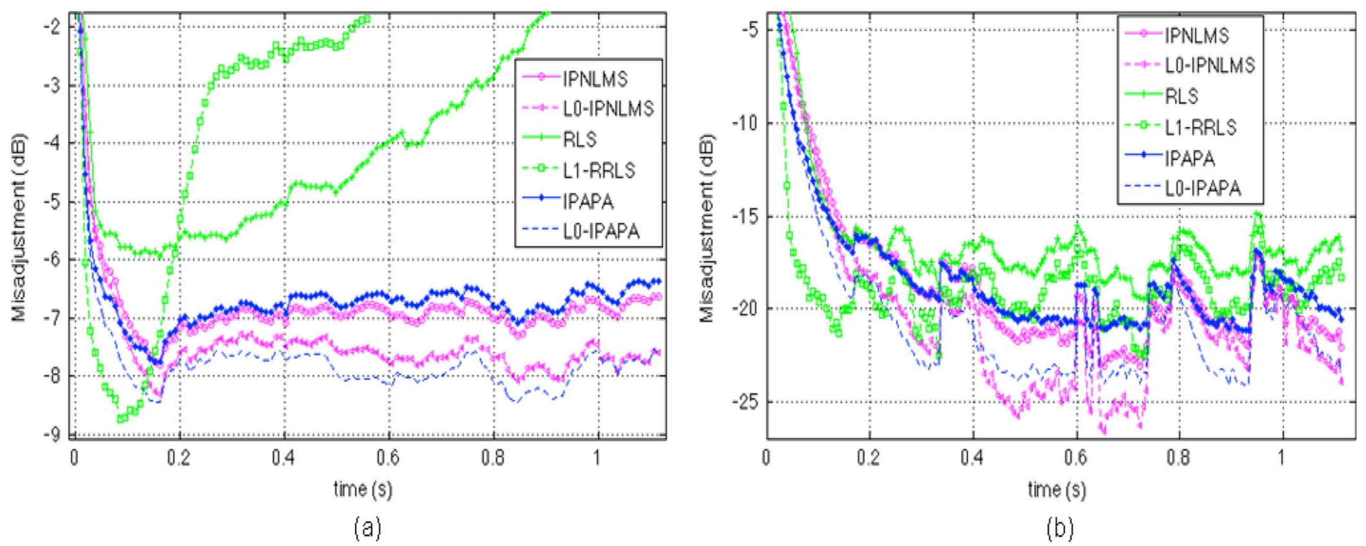

Fig. 4. Normalized misadjustment for L_1-RRLS, RLS, IPAPA, L_0-IPAPA, IPNLMS, and L_0-IPNLMS: (a) input signal is the pulse-shaped PN sequence used in the FAF experiment; (b) input signal is independent white complex Gaussian noise.

column (scenario A) corresponds to the case where the DFE is trained for the first 500 symbols and then is switched to decision-directed mode for the next 2500 symbols. The rightmost column (scenario B) corresponds to the case where the DFE runs only in training mode for 3000 symbols. In scenario A, all NG algorithms outperform L_1-RRLS but the L_0-norm effect for $\beta = 0, 0.5$ is not obvious. Scenario B shows that L_0-IPAPA/L_0-IPNLMS is consistently better than IPAPA/IPNLMS for all choices of β. Comparing the two scenarios, one notes that erroneous symbol decisions make all DFEs to experience roughly 1-dB loss in $\mathrm{SNR_{out}}$.

We conclude the experimental results by computing the error-rate performance of a DFE receiver based on direct adaptation (DA) [28], namely, the DFE does not rely on explicit channel estimates. The DFE structure has four FF filters of 13 taps each and one feedback filter of 213 taps. Two algorithms are employed for DFE adaptation: the standard RLS ($\lambda = 0.996$) and the IPAPA ($\mu = 0.1$, $L = 3$, $\beta = -1$, and 0.5 for the FF and FB filters, respectively). The DA-IPAPA receiver is described in Appendix B. The performance results are summarized in Table II. Note that DA-IPAPA outperforms DA-RLS. Moreover, DA-IPAPA is less costly in terms of computational complexity (recall that IPAPA is $O(K)$ while RLS is $O(K^2)$). The advantage of DA-IPAPA is due to the sparseness of the FB filter. Finally, note that DA DFEs are inferior to their channel-estimate-based counterparts due to their slower tracking of channel fluctuations.

TABLE III
SIMULATION RESULTS FOR ALL DFES

	SER / SNR_{out} (dB)	SER / SNR_{out} (dB)
	scenario A	scenario B
$\beta = -1$		
NLMS	0.3764 / 1.95	0.02464 / 8.74
L_0-NLMS	0.3390 / 2.34	0.01791 / 9.35
APA	0.3894 / 1.75	0.00599 / 10.13
L_0-APA	0.2901 / 2.98	0.00258 / 11.32
$\beta = 0$		
IPNLMS	0.0407 / 9.47	0.00033 / 12.09
L_0-IPNLMS	0.0277 / 10.37	0.00026 / 12.37
IPAPA	0.1322 / 7.18	0.00020 / 11.90
L_0-IPAPA	0.0415 / 10.41	0.00006 / 12.55
$\beta = 0.5$		
IPNLMS	0.0277 / 10.37	0.00017 / 12.36
L_0-IPNLMS	0.0276 / 10.57	0.00015 / 12.44
IPAPA	0.1162 / 7.64	0.00012 / 12.05
L_0-IPAPA	0.0478 / 10.08	0.00006 / 12.44
DA-RLS	0.6228 / -1.90	0.01476 / 10.30
DA-IPAPA	0.5088 / 2.00	0.00062 / 12.78

B. Simulation Results

The experimental results showed that our framework leverages on channel sparseness, however both IPAPA and IPNLMS showed marginal performance difference against their L_0-norm counterparts. This marginal difference is attributed to the low SNR. Thus, the question of how much the L_0-norm improves performance in higher SNR still remains. To address this question, we replicate the FAF experiment by using simulated Gaussian noise.

For the following simulations, the FAF channel shown in Fig. 3(b) is used as our testbed. The duration of the input signal is 7000 symbols. The channel output is generated by using (1) ($\theta[n] = 0$ for all n) at a rate of 2 samples/symbol. In addition, the channel output is corrupted by independent white complex Gaussian noise to achieve an average SNR (per channel symbol) of 20 dB. To ensure convergence of all algorithms, every channel realization is frozen for 15 symbol durations. The performance results for each algorithm are computed by averaging 100 independent trials.

The first simulation test compares L_1-RRLS with RLS and all NG algorithms in the context of channel estimation. The normalized misadjustment, given by

$$20\log_{10}\left(\frac{\|\mathbf{h}[n]-\hat{\mathbf{h}}[n]\|_2}{\|\mathbf{h}[n]\|_2}\right) \tag{18}$$

is the performance metric. All NG algorithms have $\beta = 0$. Fig. 4(a) and (b) illustrates results when the input signal is the pulse-shaped PN sequence (used in the FAF experiment) and independent complex white Gaussian noise, respectively. Clearly, L_0-IPAPA/L_0-IPNLMS show improved performance than IPAPA/IPNLMS. For example, L_0-IPNLMS is about 4 dB better than IPNLMS, as can be seen in Fig. 4(b). Although the L_1-RRLS algorithm exhibits the fastest convergence rate, its performance in Fig. 4(a) rapidly degrades due to the ill-conditioning of the input signal correlation matrix. Also, L_1-RRLS exhibits inferior channel tracking to the rest of the algorithms when the input signal is white.

The second simulation test evaluates the SER and $\mathrm{SNR}_{\mathrm{out}}$ of the DFE structures used in the previous section. Each DFE employs one FF filter with 13 taps. Table III summarizes the results. Scenario A assumes that the DFE is trained for the first 500 symbols and then is switched to decision-directed mode for the next 6500 symbols. Scenario B assumes that the DFE runs in training mode. In both scenarios, all L_0-norm-based algorithms outperform their NG counterparts. In scenario A, for instance, the L_0-norm improves IPAPA and IPNLMS, respectively, by 3.7 and 1 dB, when $\beta = 0$. DA-IPAPA outperforms DA-RLS but is inferior to all CEB DFEs. Also, the simulation results confirm that CEB DFEs based on IPAPA are more sensitive to the reliability of the decisions that are fed back due to the longer observation window L. This observation is consistent with the experimental results above. Finally, note that all DA DFEs and all NG algorithms ($\beta = -1$) fail in decision-directed mode.

V. Conclusion

A new algorithmic framework for sparse system identification was introduced. It utilized a differentiable cost function that leveraged on NG adaptation and L_0-norm regularization. New sparse adaptive algorithms were derived with quadratic computational complexity. The clear superiority of the proposed algorithms over the sparse L_1-RRLS, IPAPA, and IPNLMS was demonstrated based on data recorded from a mobile shallow-water channel. Moreover, the proposed algorithms were employed to adapt a CEB DFE receiver. These DFE receivers were compared with an RLS-based DA DFE and an IPAPA-based DA DFE. The CEB DFEs demonstrated improved error-rate performance due to their faster tracking of the time-varying sparseness of the channel.

Our results were based on algorithms with fixed sparseness parameter β. Since UWA channels are highly dynamic, it would be efficient to dynamically adapt β based on some time-varying sparseness measure. Moreover, reducing the quadratic computational complexity of the proposed algorithms to a linear order is highly desirable for on-chip implementation. We leave these two challenges as a future research direction.

Appendix A
Derivation of the L_0-IPAPA

The algorithm is derived by setting $\nabla_{\mathbf{r}[n]^*}J[n] = 0$ and $\nabla_{\hat{\theta}[n]}J[n]=0$, where $J[n]$ is given by (15). We first compute

$$\begin{aligned}\nabla_{\mathbf{r}[n]^*}J[n] = \nabla_{\mathbf{r}[n]^*}\left(\|\bar{\mathbf{e}}[n]\|_2^2\right) + \nabla_{\mathbf{r}[n]^*}(\delta\mathbf{r}[n]^\dagger\mathbf{G}_{\hat{\mathbf{h}}}[n-1]\mathbf{r}[n]) \\ + \nabla_{\mathbf{r}[n]^*}(\gamma\|\hat{\mathbf{h}}[n]\|_0). \end{aligned} \tag{19}$$

To this end, we have

$$\begin{aligned}\nabla_{\mathbf{r}[n]^*}\left(\|\bar{\mathbf{e}}[n]\|_2^2\right) &= \nabla_{\mathbf{r}[n]^*}(\bar{\mathbf{e}}[n]^{\mathsf{T}}\bar{\mathbf{e}}[n]^*) \\ &= \nabla_{\mathbf{r}[n]^*}(\bar{\mathbf{e}}[n]^{\mathsf{T}})\bar{\mathbf{e}}[n]^* \qquad (20)\\ &= -\mathbf{U}[n]\bar{\mathbf{e}}[n]^* \\ &= -\mathbf{U}[n](\mathbf{e}[n]^* - \mathbf{U}[n]^\dagger\mathbf{r}[n]) \\ &= \mathbf{U}[n]\mathbf{U}[n]^\dagger\mathbf{r}[n] - \mathbf{U}[n]\mathbf{e}[n]^* \qquad (21)\end{aligned}$$

where (20) is obtained by applying the product rule and identifying that $\nabla_{\mathbf{r}[n]^*}(\bar{\mathbf{e}}[n]^*) = 0$. Also, we have

$$\nabla_{\mathbf{r}[n]^*}(\delta\mathbf{r}[n]^\dagger\mathbf{G}_{\hat{\mathbf{h}}}[n-1]\mathbf{r}[n]) = \delta\mathbf{G}_{\hat{\mathbf{h}}}[n-1]\mathbf{r}[n]. \tag{22}$$

Using the chain rule, the gradient of $\|\hat{\mathbf{h}}[n]\|_0$ with respect to $r_k[n]^*, k = 0,\ldots,K-1$, is equal to

$$\nabla_{r_k[n]^*}(\gamma\|\hat{\mathbf{h}}[n]\|_0) = \frac{\gamma\eta}{2}e^{-\eta|\hat{h}_k[n]|_1}\,c\mathrm{sgn}(\hat{h}_k[n]). \tag{23}$$

We now define the vector $\boldsymbol{\nu}[n]$ with entries

$$\nu_k[n] = e^{-\eta|\hat{h}_k[n]|_1}\,c\mathrm{sgn}(\hat{h}_k[n]), \qquad k=0,\ldots,K-1. \tag{24}$$

Combining terms from (21)–(24), we have the following vector equation:

$$(\delta\mathbf{G}_{\hat{\mathbf{h}}}[n-1] + \mathbf{U}[n]\mathbf{U}[n]^\dagger)\mathbf{r}[n] + \frac{\gamma\eta}{2}\boldsymbol{\nu}[n] = \mathbf{U}[n]\mathbf{e}[n]^*. \tag{25}$$

From the above equation, we note that it is tedious to solve for $\mathbf{r}[n]$ since $\boldsymbol{\nu}[n]$ depends on $\hat{\mathbf{h}}[n]$ in a nonlinear fashion. At steady state, however, it is plausible to assume that $\bar{e}[n] \simeq e[n]$, and thus, $\boldsymbol{\nu}[n] \simeq \boldsymbol{\nu}[n-1]$. Using this assumption, we can solve for $\mathbf{r}[n]$ by using the matrix inversion lemma [20]. Thus, we have

$$\mathbf{A}[n] = \mathbf{G}_{\hat{\mathbf{h}}}[n-1]^{-1}\mathbf{U}[n] \tag{26}$$

$$\mathbf{B}[n] = (\mathbf{U}[n]^\dagger\mathbf{A}[n] + \delta\,\mathbf{I}_L)^{-1} \tag{27}$$

$$\mathbf{C}[n] = \mathbf{A}[n]\mathbf{B}[n] \tag{28}$$

$$\mathbf{D}[n] = \mathbf{G}_{\hat{\mathbf{h}}}[n-1]^{-1} - \mathbf{C}[n]\mathbf{A}[n]^\dagger \tag{29}$$

$$\mathbf{r}[n] = \mathbf{C}[n]\mathbf{e}[n]^* - \frac{\gamma\eta}{2\delta}\mathbf{D}[n]\boldsymbol{\nu}[n-1]. \tag{30}$$

Furthermore, to exercise control over the change of the tap values from one iteration to the next, we introduce a step-size parameter $\mu \in (0,1]$. Thus, the channel update equation is deduced as follows:

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu \mathbf{G}_{\hat{\mathbf{h}}}[n-1]^{-1}\mathbf{U}[n]\mathbf{B}[n]\mathbf{e}[n]^* - \frac{\mu\gamma\eta}{2\delta}\mathbf{D}[n]\boldsymbol{\nu}[n-1]. \quad (31)$$

We now compute $\nabla_{\hat{\theta}[n]} J[n]$. Since $\hat{\theta}[n]$ appears only in the first entry of $\bar{\mathbf{e}}[n]$, we have

$$\begin{aligned}\nabla_{\hat{\theta}[n]} J[n] &= \nabla_{\hat{\theta}[n]}(|\bar{e}[n]|^2) \simeq \nabla_{\hat{\theta}[n]}(|e[n]|^2)\\ &= 2\mathrm{Im}\left\{\hat{\mathbf{h}}[n-1]^\dagger \mathbf{u}[n]e^{j\hat{\theta}[n]}e[n]^*\right\}\\ &= 2\mathrm{Im}\left\{\hat{\mathbf{h}}[n-1]^\dagger \mathbf{u}[n]e^{j\hat{\theta}[n]}y[n]^*\right\}. \end{aligned} \quad (32)$$

Although the gradient–descent method could be used to compute $\hat{\theta}[n+1]$, following the suggestion in [25], $\hat{\theta}[n+1]$ is computed by means of a second-order phase-locked loop (PLL) as follows:

$$\hat{\theta}[n+1] = \hat{\theta}[n] + K_1\Phi[n] + K_2\sum_{i=0}^{n-1}\Phi[i] \quad (33)$$

where K_1 and K_2 are positive phase-tracking parameters and $\Phi[n] = -\mathrm{Im}\{\hat{\mathbf{h}}[n-1]^\dagger \mathbf{u}[n]y[n]^*\}$. The algorithm described by (31) and (33) will be called L_0-IPAPA hereafter. The L_0-IPAPA is initialized with $\hat{\mathbf{h}}[0] = \mathbf{0}$ and $\hat{\theta}[0] = 0$.

APPENDIX B
THE DA-IPAPA

Let the vector $\mathbf{h}_{ff}[n]$ of length N_{ff} and the vector $\mathbf{h}_{fb}[n]$ of length N_{fb} denote the FF filter and the feedback filter of the DFE, respectively. A fractionally spaced DFE coupled with carrier-phase tracking is algorithmically described as follows [25]:

$$e[n] = d[n] - \hat{d}[n] \quad (34)$$
$$\hat{d}[n] = p[k] - q[n] \quad (35)$$
$$p[n] = \mathbf{h}_{ff}[n]^\dagger \mathbf{u}[n]e^{-j\theta[n]} \quad (36)$$
$$q[n] = \mathbf{h}_{fb}[n]^\dagger \mathbf{d}[n] \quad (37)$$
$$\theta[n] = \theta[n-1] + K_1\Phi[n] + K_2\sum_{i=0}^{n-1}\Phi[i] \quad (38)$$
$$\Phi[n] = \mathrm{Im}\{p[n](d[n]+q[n])^*\} \quad (39)$$
$$\mathbf{u}[n] = [u[nT+NT_s]\dots u[nT-NT_s]]^\intercal \quad (40)$$
$$\mathbf{d}[n] = [d[n-1]\dots d[n-N_{fb}]]^\intercal \quad (41)$$

where $d[n]$ denotes the transmitted/decided symbol when the DFE operates in training/decision-directed mode, $\hat{d}[n]$ is the symbol estimate, $e[n]$ is the error signal, $p[n]$ is the output of the FF filter, $q[n]$ is the output of the feedback filter, $\theta[n]$ is the carrier-phase estimate, K_1 and K_2 are phase tracking parameters, $\Phi[n]$ is the phase detector output, $\mathbf{u}[n]$ is the received (baseband) signal vector of length $N_{ff} = 2N+1$, $1/T_s$ is the sampling rate of the received signal, $1/T$ is the symbol rate, and $\mathbf{d}[n]$ is a vector containing the N_{fb} previously decided symbols.

In the DFE context, the Lth-order IPAPA is written as

$$\mathbf{D}[n] = [\mathbf{d}[n]\dots\mathbf{d}[n-L+1]] \quad (42)$$
$$\mathbf{U}[n] = \left[\mathbf{u}[n]e^{-j\theta[n]}\dots\mathbf{u}[n-L+1]e^{-j\theta[n-L+1]}\right] \quad (43)$$
$$\bar{\mathbf{d}}[n] = [d[n]\dots d[n-L+1]]^\intercal \quad (44)$$
$$\mathbf{e}[n]^* = \bar{\mathbf{d}}^*[n] - (\mathbf{U}[n]^\dagger \mathbf{h}_{ff}[n] - \mathbf{D}[n]^\dagger \mathbf{h}_{fb}[n]) \quad (45)$$
$$\mathbf{A}_{fb}[n] = \mathbf{G}_{\mathbf{h}_{fb}}[n-1]\mathbf{D}[n] \quad (46)$$
$$\mathbf{A}_{ff}[n] = \mathbf{G}_{\mathbf{h}_{ff}}[n-1]\mathbf{U}[n] \quad (47)$$
$$\mathbf{h}_{fb}[n] = \mathbf{h}_{fb}[n-1] - \mu\mathbf{A}_{fb}[n](\mathbf{D}[n]^\dagger\mathbf{A}_{fb}[n] + \delta_{fb}\mathbf{I}_L)^{-1}\mathbf{e}[n]^* \quad (48)$$
$$\mathbf{h}_{ff}[n] = \mathbf{h}_{ff}[n-1] + \mu\mathbf{A}_{ff}[n](\mathbf{U}[n]^\dagger\mathbf{A}_{ff}[n] + \delta_{ff}\mathbf{I}_L)^{-1}\mathbf{e}[n]^* \quad (49)$$

where $\mathbf{D}[n]$ is the $N_{fb} \times L$ matrix of input signal symbols, $\mathbf{U}[n]$ is the $N_{ff} \times L$ matrix of output signal samples, $\mathbf{e}[n]$ is the $L \times 1$ error signal vector, and δ_{ff} and δ_{fb} are regularization parameters given by (5). The matrices $\mathbf{G}_{\mathbf{h}_{ff}}$ and $\mathbf{G}_{\mathbf{h}_{fb}}$ depend on $\mathbf{h}_{ff}[n]$ and $\mathbf{h}_{fb}[n]$, respectively, and are computed using the method described in (10). Since $\mathbf{h}_{fb}$ is known to be a sparse vector, parameter β should be chosen close to 0.5.

ACKNOWLEDGMENT

The authors would like to thank the personnel of the NATO Undersea Research Centre (NURC) for carrying out the Focused Acoustic Fields (FAF) experiment.

REFERENCES

[1] D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 4–27, Jan. 2000.

[2] J. C. Preisig and G. B. Deane, "Surface wave focusing and acoustic communications in the surf zone," *J. Acoust. Soc. Amer.*, vol. 116, pp. 2067–2080, 2004.

[3] J. G. Proakis, "Adaptive equalization techniques for acoustic telemetry channels," *IEEE J. Ocean. Eng.*, vol. 16, no. 1, pp. 21–31, Jan. 1991.

[4] M. A. Chitre, S. Shahabudeen, and M. Stojanovic, "Underwater acoustic communications and networking: Recent advances and future challenges," *J. Mar. Technol. Soc.*, vol. 42, no. 1, pp. 103–116, 2008.

[5] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 927–939, Oct. 2007.

[6] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pt. 2, pp. 1708–1721, Mar. 2010.

[7] J. Ling, K. Zhao, J. Li, and M. Lundberg Nordenvaad, "Multi-input multi-output underwater communications over sparse and frequency modulated acoustic channels," *J. Acoust. Soc. Amer.*, vol. 130, no. 1, pp. 249–262, 2011.

[8] F. Wan, U. Mitra, and A. Molisch, "The modified iterative detector/estimator algorithm for sparse channel estimation," in *Proc. IEEE OCEANS Conf.*, Seattle, WA, Sep. 2010, DOI: 10.1109/OCEANS.2010.5664395.

[9] A. Sen Gupta and J. Preisig, "A geometric mixed norm approach to shallow water acoustic channel estimation and tracking," *J. Phys. Commun.*, vol. 5, Special Issue on Compressive Sensing in Communications, no. 2, pp. 119–128, Jun. 2012.

[10] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin, Germany: Springer-Verlag, 2006.

[11] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2002, vol. 2, pp. 1881–1884.

[12] O. Hoshuyama, R. A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2004, vol. 4, pp. 161–164.

[13] K. Pelekanakis and M. Chitre, "Comparison of sparse adaptive filters for underwater acoustic channel equalization/estimation," in *Proc. IEEE Int. Conf. Commun. Syst.*, Singapore, Nov. 2010, pp. 395–399.
[14] Y. Gu, J. Jin, and S. Mei, "ℓ_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 774–777, Sep. 2009.
[15] D. Angelosante, J. Bazerque, and G. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1-norm," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3436–3447, Jul. 2010.
[16] E. M. Eksioglu, "Sparsity regularized recursive least squares adaptive filtering," *IET Signal Process.*, vol. 5, pp. 480–487, 2011.
[17] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4013–4025, Aug. 2010.
[18] K. Pelekanakis and M. Chitre, "Natural gradient-based adaptive algorithms for sparse underwater acoustic channel identification," in *Proc. 4th Underwater Acoust. Meas. Conf.*, Kos Island, Greece, Jun. 2011, pp. 1403–1410.
[19] C. Paleologu, J. Benesty, and S. Ciochina, "An improved proportionate NLMS algorithm based on the ℓ_0 norm," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Dallas, TX, Mar. 2010, pp. 309–312.
[20] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[21] S. I. Amari and S. C. Douglas, "Why natural gradient?," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Seattle, WA, May 1998, vol. 2, pp. 1213–1216.
[22] S. L. Gay and S. C. Douglas, "Normalized natural gradient adaptive filtering for sparse and nonsparse systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2002, vol. II, pp. 1405–1408.
[23] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, Jr., "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1883–1894, Aug. 2002.
[24] R. E. Mahony and R. C. Williamson, "Prior knowledge and preferential structures in gradient descent learning algorithms," *J. Mach. Learn. Res.*, vol. 1, pp. 311–355, 2001.
[25] M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Phase-coherent digital communications for underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 19, no. 1, pp. 100–111, Jan. 1994.
[26] J. Kivinen and M. K. Warmuth, "Exponentiated gradient versus gradient descent for linear predictors," *Inf. Comput.*, vol. 132, no. 1, pp. 1–64, Jan. 1997.
[27] M. Stojanovic, "Efficient processing of acoustic signals for high rate information transmission over sparse underwater channels," *J. Phys. Commun.*, pp. 146–161, Jun. 2008.
[28] M. Stojanovic, J. Catipovic, and J. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," *J. Acoust. Soc. Amer.*, vol. 94, no. 3, pt. 1, pp. 1621–1631, 1993.

Konstantinos Pelekanakis (S'06–M'09) received the Diploma from the Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece, in 2001 and the M.Sc. and Ph.D. degrees in mechanical and ocean engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2004 and 2009, respectively.

Currently, he is a Research Fellow with the Acoustic Research Laboratory (ARL), National University of Singapore (NUS), Singapore. His current research lies in the areas of multiple-input–multiple-output (MIMO) underwater acoustic communications and learning algorithms for sparse system identification.

Mandar Chitre (M'03–SM'11) received the B.Eng. (honors) and M.Eng. degrees in electrical engineering from the National University of Singapore (NUS), Singapore, in 1997 and 2000, respectively, the M.Sc. degree in bioinformatics from the Nanyang Technological University (NTU), Singapore, in 2004, and the Ph.D. degree from NUS in 2006.

From 1997 to 1998, he worked with the Acoustic Research Laboratory (ARL), NUS, as a Research Engineer. From 1998 to 2002, he headed the technology division of a regional telecommunications solutions company. In 2003, he rejoined ARL, initially as the Deputy Head (Research) and is now the Head of the laboratory. He also holds a joint appointment with the Department of Electrical and Computer Engineering at NUS as an Assistant Professor. His current research interests are underwater communications, autonomous underwater vehicles, and underwater signal processing.

Dr. Chitre has served on the technical program committees of the IEEE OCEANS Conference, the International Conference on UnderWater Networks and Systems (WUWNet), the Defense Technology Asia (DTA) Conference, and the Water Side Security (WSS) Conference, and has served as reviewer for many international journals. He was the chairman of the student poster committee for the 2006 IEEE OCEANS Conference in Singapore. In the past years, he has served as the Vice Chairman, Secretary, and Treasurer for the IEEE Oceanic Engineering Society (OES; Singapore chapter) and is currently the IEEE Technology Committee Co-Chair of underwater communication, navigation, and positioning. He also serves as a Technical Co-Chair for the 2012 IEEE International Conference on Communications Systems.