Adaptive Sparse Channel Estimation under Symmetric alpha-Stable Noise

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

Abstract-We tackle the problem of channel estimation in environments that exhibit both sparse, time-varying impulse responses and impulsive noise with Symmetric alpha-Stable (S α S) statistics. Two novel frameworks are proposed for designing online adaptive algorithms that exploit channel sparseness and achieve robust performance against impulses. The first framework generates recursive least-squares (RLS)-type algorithms based on a differentiable cost function that combines robust nonlinear methods with sparse-promoting L₀ norm regularization. The second framework employs the natural gradient (NG) and incorporates non-linear methods for the channel prediction error as well as the L₀ norm of the channel taps. From these frameworks, we derive linear and quadratic complexity algorithms. The improved performance of the proposed RLS-type and NGtype algorithms relative to conventional robust algorithms, such as the recursive least M-estimate (RLM) algorithm and the recursive least p-norm (RLP) algorithm, is validated by using extensive computer simulations as well as signal analysis from an underwater acoustic communications experiment. In addition, we discovered that RLM is not robust under specific $S\alpha S$ noise conditions, contrary to the claim in [34]. Finally, our results also demonstrate the clear superiority of the NG-type algorithms over their RLS-type counterparts.

Index Terms—Robust system identification, outlier rejection, robust statistics, M-estimate algorithm.

I. INTRODUCTION

S PARSE channels are typically encountered in wireless links such as digital TV [1], acoustic echo [2], and underwater acoustic [3]. They are called *sparse* since most of the energy of the impulse response is concentrated in a small fraction of its duration. Exploiting sparseness in channel estimation by using adaptive filters has gained considerable interest since the late 1990s [4-6]. It is well known that improved estimation performance in terms of steady-state misadjustment and channel tracking can be achieved by using sparse prior information. In addition, a receiver that explicitly adapts to a sparse channel can attain reduced complexity if only the significant channel coefficients are retained.

Proportionate-type algorithms, i.e., algorithms that update each channel coefficient in proportion to its estimated magnitude were among the first paradigms to use sparse prior

Manuscript received August 5, 2013; revised December 18, 2013 and February 10, 2014; accepted March 6, 2014. The associate editor coordinating the review of this paper and approving it for publication was Z. Wang.

This work was partially supported by MOE Tier-1 grant R-263-000-521-133.

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

Digital Object Identifier 10.1109/TWC.2014.042314.131432

information [2]. Popular examples are the improved proportionate NLMS (IPNLMS) [7] algorithm and the improved proportionate affine projection algorithm (IPAPA) [8]. The first algorithm that used a sparseness-promoting L_p norm ($p \in$ (0,1]) within its cost function was the pNLMS [9]. Based on the expectation-maximization (EM) algorithm, a sparse RLS, termed as sparse RLS (SPARLS), was studied in [10]. An L₁ norm-regularized RLS based on the least-absolute shrinkage and selection operator (Lasso) approach was proposed in [11]. On a different track, Slavakis et al [12] proposed a sparse online algorithm using projections on closed convex sets and Murakami et al [13] introduced an Adaptive Proximal Forward-Backward Splitting (APFBS) scheme. Recently, the authors introduced an algorithmic framework that leveraged on natural gradient (NG) adaptation combined with L_0 norm regularization [14].

The above papers, among many others, assume that the observation noise is Gaussian and so using the L_2 norm of the channel prediction error (i.e., the difference between the observed signal and the filter output) in the cost function is optimal. However, a number of man-made and physical noise processes depart from the Gaussian assumption due to their impulsive nature. Examples are: multiple access interference in radio channels [15], double talk in acoustic echo cancellation [16], biological noise [17] or ice cracking [18] in various underwater acoustic channels. Such environments require the use of robust adaptive filters since L_2 norm-based algorithms suffer severe performance degradation.

Studies that propose sparse channel estimation in the presence of impulsive noise are scarce. Vega et al proposed a variable step-size IPNLMS algorithm [19]. Subsequently, an improved proportionate affine projection sign algorithm (RIP-APSA) based on the L₁ norm of the error signal was introduced by [20]. Yamamoto et al [21] robustified the APFBS scheme by employing a Huber loss function [22]. A notable issue with the aforementioned algorithms is that they were tested in impulsive noise that obeys a Gaussian-mixture density function. Gaussian-mixture models and the related Middleton class A model are often used to model impulsive noise environments (e.g., man-made impulse noise against a Gaussian noise background) [23]. Although these models lend themselves to computer simulations, they are not suitable for modeling many natural noise sources (e.g., snapping shrimp noise in warm shallow waters [17]). Another drawback of the Gaussian-mixture distribution is that its tail decays exponentially while empirical evidence manifests that algebraic decay of heavy-tailed noise processes often occurs in communications as well as in various fields of engineering, physics, and economics [24].

The family of alpha-stable distributions provides an accu-

rate model for heavy-tailed noise [25]. The significance and validity of alpha-stable models is justified by the stability property and the Generalized Central Limit Theorem (GCLT). In this work, we deal with the family of Symmetric alpha-Stable (S α S) distributions due to its ability to model many impulsive noise processes in communications channels, and, in fact, includes the Gaussian density as a special case.

For signal processing applications, the pivotal property of $S\alpha S$ random processes is their infinite second and higher order moments. As shown in [25], the minimum dispersion criterion is an appropriate measure of optimality since minimizing the error signal dispersion, the average estimation error is simultaneously minimized. In addition, minimizing the error dispersion is equivalent to minimizing the fractional lower order moment (FLOM) of the estimation error. Variants of LMS and RLS algorithms based on FLOMs were proposed in the stable signal processing community. Characteristic examples are the least mean p-norm (LMP) [25], the recursive least p-norm (RLP) [26], and recursive least mean p-norm (RLMP) [27]. It is worth stressing that both LMP and RLP are tailored to real-valued (passband) channels. Recently, the authors introduced an algorithmic framework for complexvalued (baseband) channels [28]. That paper serves as a brief precursor to this work.

Our main contribution is the development of two new algorithmic frameworks that systematically generate sparse adaptive filters robust against $S\alpha S$ noise. The first framework generates RLS-type algorithms. This framework uses an objective function with two parts: a data fidelity component that is robust against outliers and a L₀ norm regularization component. The second framework leverages on natural gradient (NG) adaptation by using a sparse-aware Riemannian distance. In addition, this framework incorporates robust non-linear methods for the channel prediction error and the L₀ norm of the channel taps. Better flexibility in terms of computational complexity than the RLS-type framework is shown. New linear and quadratic complexity algorithms are derived. Finally, all proposed algorithms are tested by identifying experimental sparse underwater acoustic channels in simulated and reallife noise with $S\alpha S$ properties. Their improved performance relative to conventional robust algorithms is demonstrated. Contrary to Chan and Zou [34], we show that the recursive least M-estimate (RLM) algorithm is not robust under specific noise regimes. Furthermore, our analysis manifests the superiority of the NG-type framework over its RLS-type counterpart.

The remainder of this paper is organized as follows: Section II describes the system model. The RLS-type and the NG-type frameworks are presented, respectively, in Sections III and IV. Simulation and experimental results are reported in Section V. Finally, the paper is concluded in Section VI.

II. NOTATION AND SYSTEM MODEL

A. Notation

Superscripts ^T, [†], 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 . Let $z \in \mathbb{C}$ and $p \geq 1$. The \mathbf{L}_p norm of z is defined as $|z|_p \triangleq (|\operatorname{Re}\{z\}|^p + |\operatorname{Im}\{z\}|^p)^{1/p}$. The sign function of z is defined as csgn $(z) \triangleq sgn(\operatorname{Re}\{z\}) + j \cdot sgn(\operatorname{Im}\{z\})$, where

 $sgn(\cdot)$ stands for the sign function of a real scalar. Let $\mathbf{z} \in \mathbb{C}^N$. The sign function of \mathbf{z} is given by the column vector $csgn(\mathbf{z})$ with elements $csgn(z_i)$, $i=0, \ldots, N-1$. The L_p norm of \mathbf{z} is defined as $\|\mathbf{z}\|_p \triangleq (\sum_{i=0}^{N-1} |z_i|_p^p)^{1/p}$. The L_0 norm of \mathbf{z} , denoted as $\|\mathbf{z}\|_0$, equals the number of the non-zero entries of \mathbf{z} . The complex gradient of a scalar function $f(\mathbf{z})$ with respect to \mathbf{z} is denoted as $\nabla_{\mathbf{z}} f(\mathbf{z})$ and is defined in [29].

B. System model

We employ the baseband representation of the channel, transmitted/received signals, and additive noise process. Let us consider an impulse response, which is described by the unknown K-tap vector $\mathbf{h}[n] = [h_0[n] h_1[n] \dots h_{K-1}[n]]^T$ at discrete time n. In addition, we assume that $\mathbf{h}[n]$ is slowly time-varying and sparse, namely, most of the coefficients are close to zero and only few of them are large. The received signal is expressed as

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

where $\mathbf{u}[n] = [u[n] \ u[n-1] \dots u[n-K+1]]^{\mathsf{T}}$ contains the K most recent samples of the transmitted signal, and w[n]denotes the noise. The passband noise samples are assumed independent and identically distributed (i.i.d.) following the Symmetric alpha-Stable (S α S) distribution with characteristic function $\varphi(\omega) = e^{-\gamma |\omega|^{\alpha}}$. The characteristic exponent $\alpha \in (0, 2]$ describes the impulsiveness of the noise (smaller α leads to more frequent occurrence of impulses) and the dispersion $\gamma > 0$ controls the spread of the distribution around its location parameter (which is zero for our purposes). When $\alpha=2$, the S α S probability density function (pdf) boils down to the Gaussian pdf and γ is equal to half the variance. For mathematical and practical reasons (it is rare to find S α S noise with $\alpha < 1$ in practical systems), we restrict our work to the class of S α S distributions where $\alpha \in (1, 2]$ [25]. The objective of this paper is to perform recursive estimation of h[n] with limited complexity and memory given sequential observations $\{y[i], \mathbf{u}[i]\}_{i=1}^{n}$.

In digital communications, performance analysis is often reported against the signal-to-noise ratio (SNR) per information symbol, denoted as E_s/N_0 , where E_s is the energy of the information symbol and N_0 is the power spectral density (PSD) of the noise. Since the concept of the PSD cannot be applied in S α S noise, we must resort to a different SNR measure. An SNR measure for baseband signals is rather involved because the in-phase and quadrature components of the S α S noise are generally dependent [30]. In passband, however, the parameters α and γ can be easily estimated and so an SNR definition can be readily defined as

$$E_s/N_0 (\mathbf{dB}) \triangleq 10 \log_{10} \frac{N_s P_s}{2\gamma^{2/\alpha}},$$
 (2)

where N_s is the ratio of the symbol interval over the sample interval, P_s is the received signal power, and $\gamma^{2/\alpha}$ plays the same role as the variance. When α =2, equation (2) becomes the usual E_s/N_0 definition in Gaussian noise.

III. RLS-TYPE FRAMEWORK

The RLS algorithm is one of the most important adaptive filter algorithms due to its fast convergence rate in nonstationary environments, insensitivity to the eigenvalue spread of the input correlation matrix, and modular structure that offers fast implementations (e.g., lattice structure). Hence, it is desirable to have RLS-type algorithms that are robust in $S\alpha S$ noise and exploit channel sparseness for improved performance.

Let us express the a posterior and prior error as $\bar{e}[i]=y[i] - \hat{h}[n]^{\dagger}\mathbf{u}[i]$ and $e[i]=y[i] - \hat{h}[n-1]^{\dagger}\mathbf{u}[i]$, respectively, where $i \leq n$ and $\hat{h}[n]$ stands for the estimate of h[n]. Elaborating on the robust filtering approach [31], we consider a real, non-negative valued *loss function*, denoted as $f(\bar{e})$, whose purpose is to down weight large errors due to impulses. We also define the complex *score function* $\psi(\bar{e})=\partial f(\bar{e})/\partial \bar{e}$ and the complex *weight function* $q(\bar{e})=\psi(\bar{e})/\bar{e}^*$. The proposed cost function is written as:

$$J[n] = \sum_{i=0}^{n} \lambda^{n-i} f\left(\bar{e}[i]\right) + \zeta \left\| \hat{\mathbf{h}}[n] \right\|_{0},$$
(3)

where $\lambda \in (0, 1)$ is the forgetting factor and $\zeta \geq 0$ is a regularization parameter. The regularizing term $\|\hat{\mathbf{h}}[n]\|_0$ helps to further accelerate the convergence of the inactive (close to zero) filter taps. It is obvious that if different functions $f(\bar{e})$ and L_0 norm proxies are employed, different algorithms will be generated. Specific examples of $f(\bar{e})$ are given below. For the remainder of this paper, the L_0 norm is approximated by the differentiable function¹

$$\left\| \hat{\mathbf{h}}[n] \right\|_{0} \simeq \sum_{k=0}^{K-1} 1 - e^{-\eta \left| \hat{h}_{k}[n] \right|_{1}}, \eta > 0, \tag{4}$$

which is a complex extension of the real L₀ norm used in [32]. The parameter $\epsilon = 1/\eta$ defines the interval $[-\epsilon, \epsilon]$ such that all Re $\{\hat{h}_k[n]\}$ and Im $\{\hat{h}_k[n]\}$ that fall within that interval are attracted towards the zero value. Note that equation (3) is not a convex cost function, however, if η is close to 10 and ζ is chosen sufficiently small the algorithm converges to meaningful solutions, as indicated in [32].

The RLS-type algorithm is derived in Appendix A by setting $\nabla_{\hat{\mathbf{h}}[n]*} J[n]=0$. The following equations summarize the result:

$$e[n] = y[n] - \hat{\mathbf{h}}[n-1]^{\dagger} \mathbf{u}[n], \qquad (5)$$

$$\mathbf{k}[n] = \frac{q(e[n])\mathbf{\Psi}[n-1] \quad \mathbf{u}[n]}{\lambda + q(e[n])\mathbf{u}[n]^{\dagger}\mathbf{\Phi}[n-1]^{-1}\mathbf{u}[n]},\tag{6}$$

$$\Phi[n]^{-1} = \lambda^{-1} \left(\Phi[n-1]^{-1} - \mathbf{k}[n] \mathbf{u}[n]^{\dagger} \Phi[n-1]^{-1} \right), \quad (7)$$

$$\nu_{k}[n] = e^{-\eta |h_{k}[n]|_{1}} csgn\left(h_{k}[n]\right)$$

, k = 0, ..., K - 1, (8)

$$\mathbf{\hat{h}}[n] = \mathbf{\hat{h}}[n-1] + \mathbf{k}[n]e[n]^* +$$

$$\frac{\lambda - 1}{\lambda} \frac{\zeta \eta}{2} \left(\mathbf{I}_K - \mathbf{k}[n] \mathbf{u}[n]^{\dagger} \right) \mathbf{\Phi}[n-1]^{-1} \nu[n-1].$$
(9)

The algorithm is initialized for $\hat{\mathbf{h}}[0]=\mathbf{0}$ and $\boldsymbol{\Phi}[0]^{-1}=\kappa^{-1}\mathbf{I}_{K}$, κ being a small positive real number. It is worthy to note that if $f(\bar{e}[i])=\bar{e}[i]\bar{e}[i]^*$ is employed in (3), then q(e[n])=1 and so (5)-(9) will be called the L₀-RLS algorithm hereafter. The L₀-RLS algorithm requires $O(K^2)$ algebraic operations per datum. If in addition $\zeta=0$, then the resulting algorithm becomes the standard RLS [33].

A. The L_0 -RLM algorithm

The work in [34] introduces the recursive least M-estimate (RLM) algorithm, a real-valued adaptive algorithm based on Hampel's three-part redescending M-estimate cost function. The algorithm is designed to cope with contaminated Gaussian noise, namely, the observed noise consists of two components: a Gaussian component and an impulsive interference component. In addition, the authors claim that RLM is robust under $S\alpha S$ noise, however, no performance results were reported towards that front. We check the validity of this claim in Section V.

Our aim is to improve RLM to yield lower misadjustment in sparse channels. To this end, we modify Hampel's threepart redescending M-estimate function so that it conforms with the chosen complex gradient operator [29]. Dropping the time index for notational convenience, the loss function has the form

$$f(\bar{e}) = \begin{cases} \bar{e}\bar{e}^* & , 0 \le |\bar{e}|_2 < \xi \\ 2\xi |\bar{e}|_2 - \xi^2 & , \xi \le |\bar{e}|_2 < \Delta \\ \xi (T + \Delta) - \xi^2 + \xi \frac{(|\bar{e}|_2 - T)^2}{\Delta - T} & , \Delta < |\bar{e}|_2 < T \\ \xi (T + \Delta) - \xi^2 & , T < |\bar{e}|_2 \end{cases}$$
(10)

where the threshold parameters ξ , Δ , and T are used for outlier suppression. The score and weight function are computed as

$$\psi(\bar{e}) = \begin{cases}
\bar{e}^{*} & , 0 \leq |\bar{e}|_{2} < \xi \\
\frac{\xi}{|\bar{e}|_{2}} \bar{e}^{*} & , \xi \leq |\bar{e}|_{2} < \Delta \\
\xi \frac{|\bar{e}|_{2} - T}{\Delta - T} \frac{\bar{e}^{*}}{|\bar{e}|_{2}} & , \Delta < |\bar{e}|_{2} < T \\
0 & , T < |\bar{e}|_{2}
\end{cases}$$
(11)

and

$$q(\bar{e}) = \begin{cases} 1 & ,0 \le |\bar{e}|_2 < \xi \\ \frac{\xi}{|\bar{e}|_2} & ,\xi \le |\bar{e}|_2 < \Delta \\ \frac{\xi \frac{|\bar{e}|_2 - T}{\Delta - T} \frac{1}{|\bar{e}|_2}}{0} & ,\Delta < |\bar{e}|_2 < T \\ 0 & ,T < |\bar{e}|_2 \end{cases}$$
(12)

respectively.

We now describe a way to continuously estimate the threshold parameters ξ , Δ , and T under the assumption of contaminated Gaussian noise. We stress that our system model in (1) has no background Gaussian noise. Based on [34], a robust estimation of the variance of the real part of the baseband noise is computed by using the median operator as follows:

$$\sigma_r^2[n] = \lambda_\sigma \sigma_r^2[n-1] + c(1-\lambda_\sigma) \operatorname{med}(\mathbf{a}[n]), \quad (13)$$

where λ_{σ} is a forgetting factor, $\mathbf{a}[n] = [e_r^2[n] \dots e_r^2[n - N_w + 1]]^{\mathrm{T}}$ is the real part of the prior error signal, and $c = 1.483(1 + 5/(N_w - 1))$ is a finite sample correction factor that ensures consistent estimates. The variance of the imaginary part of the baseband noise, $\sigma_i^2[n]$, is computed similarly. Using the Rayleigh distribution for $|e[n]|_2$ with parameter $\sigma^2[n] = 0.5(\sigma_r^2[n] + 10^{-10})$

¹Strictly speaking the function is not differentiable along the real and imaginary axis but this is not a problem in practice since we allow the channel taps to be arbitrarily close to these axes.

 $\sigma_i^2[n])$, the threshold parameters are chosen by the following expressions: $\xi{=}2.45\sigma[n]$ (i.e., $\Pr\{|e[n]|_2<\xi\}{=}0.95)$, $\Delta{=}2.72\sigma[n]$ (i.e., $\Pr\{|e[n]|_2<\Delta\}{=}0.975)$, and $T{=}3.03\sigma[n]$ (i.e., $\Pr\{|e[n]|_2<T\}{=}0.99$). The algorithm described by (12) and (5)-(9) will be called L_0-RLM hereafter. The channel update of the L_0-RLM requires $O(N_w \log_2(N_w))$ additional operations for the computation of $\sigma^2[n]$ as compared to L_0-RLS. Finally, note that if $\zeta{=}0$, then L_0-RLM reduces to the complex form of RLM.

B. The L_0 -RLSA

As discussed above, the parameters ξ , Δ , T of the L₀-RLM are based on the steady-state error signal. Consequently, the algorithm performance may be compromised when large channel fluctuations or large impulses occur during its convergence period. Note that the performance results in [34] are based on the condition that no impulses occur during the convergence period of RLM. Clearly, that is not the case for the channels considered here.

Here, we propose a loss function $f(\bar{e})$ that does not depend on any threshold parameters and is resilient to large impulses through the merits of the L₁ norm, that is, $f(\bar{e}[i])=|\bar{e}[i]|_1$. The score function is computed as

$$\psi(\bar{e}[i]) = 0.5(csgn(\bar{e}[i]))^* \tag{14}$$

and the weight function in (6) becomes

$$q(e[n]) = 0.5(csgn(e[n])/e[n])^*.$$
(15)

The algorithm described by (15) and (5)-(9) will be called the L₀-RLSA (L₀ norm-recursive least sign algorithm) hereafter.

C. The L_0 -RLP algorithm

The work in [26] introduces the RLP algorithm, a realvalued adaptive algorithm that exhibits robust performance in S α S noise. The RLP is established on the important observation that the mean square error (MSE) is not a valid optimality criterion since S α S distributions lack moments of order $p \ge \alpha$. However, all moments of order $p < \alpha$ do exist and so the minimum dispersion error is mathematically meaningful as an optimality criterion. This fact motivates the usage of the L_p norm ($p \in [1, \alpha)$) of the a posterior error in the cost function.

Here, we enhance RLP so that it yields lower misadjustment in sparse channels. Inspired by the structure of the RLM loss function, we use a mixture of L_2 and L_p norms as follows:

$$f(\bar{e}) \triangleq \begin{cases} \bar{e}\bar{e}^* & , 0 \le |\bar{e}|_2 < \xi \\ |\bar{e}|_p^p & , \xi \le |\bar{e}|_2 < \Delta \\ |\Delta|^p & , \Delta \le |\bar{e}|_2 \end{cases}$$
(16)

where the threshold parameters ξ , Δ are proportional to the dispersion γ of the observed passband S α S noise. The score and weight function are computed as

$$\psi(\bar{e}) = \begin{cases} \bar{e}^* & , 0 \le |\bar{e}|_2 < \xi \\ \frac{p}{2} [|\operatorname{Re}\{\bar{e}\}|^{p-1} \operatorname{sgn}(\operatorname{Re}\{\bar{e}\}) & , \xi \le |\bar{e}|_2 < \Delta \\ -j \ |\operatorname{Im}\{\bar{e}\}|^{p-1} \operatorname{sgn}(\operatorname{Im}\{\bar{e}\})] & , \Delta \le |\bar{e}|_2 \end{cases}$$
(17)

and

$$q(\bar{e}) = \begin{cases} 1 & , 0 \leq |\bar{e}|_2 < \xi \\ \frac{p}{2\bar{e}^*} [|\operatorname{Re}\{\bar{e}\}|^{p-1} sgn(\operatorname{Re}\{\bar{e}\}) & , \xi \leq |\bar{e}|_2 < \Delta \\ -j \ |\operatorname{Im}\{\bar{e}\}|^{p-1} sgn(\operatorname{Im}\{\bar{e}\})] & , \Delta \leq |\bar{e}|_2 \end{cases}$$

$$(18)$$

respectively. The algorithm described by (5)-(9) and (18) will be called L₀-RLP hereafter. The L₀-RLP requires $O(K^2)$ algebraic operations per datum. If in addition ζ =0, then the L₀-RLP reduces to the complex form of RLP.

IV. FRAMEWORK BASED ON THE NATURAL GRADIENT

Below, we propose two algorithmic frameworks that leverage on natural gradient (NG) adaptation. Before we embark on our discussion, it is instructive to define the a priori error vector,

$$\mathbf{e}[n]^* = [e[n]^* e[n-1]^* \dots e[n-L+1]^*]^{\mathsf{T}} = \mathbf{y}[n]^* - \mathbf{U}[n]^{\dagger} \hat{\mathbf{h}}[n-1],$$
(19)

and a posteriori error vector,

$$\mathbf{\bar{e}}[n]^* = [\bar{e}[n]^* \, \bar{e}[n-1]^* \dots \bar{e}[n-L+1]^*]^{\mathrm{T}}$$
$$= \mathbf{y}[n]^* - \mathbf{U}[n]^{\dagger} \mathbf{\hat{h}}[n], \qquad (20)$$

where $\mathbf{U}[n] = [\mathbf{u}[n] \mathbf{u}[n-1] \dots \mathbf{u}[n-L+1]]$ is the $K \times L$ matrix of input samples and $\mathbf{y}[n] = [y[n] y[n-1] \dots y[n-L+1]]^T$ contains the L most recent output samples. Let us also denote the channel update vector as

$$\mathbf{r}[n] = \mathbf{\hat{h}}[n] - \mathbf{\hat{h}}[n-1].$$
(21)

Then, the a-posteriori error vector be also be written as:

$$\bar{\mathbf{e}}[n]^* = \mathbf{e}[n]^* - \mathbf{U}[n]^{\dagger}\mathbf{r}[n], \qquad (22)$$

where (22) follows from substituting (21) into (20).

A. Background

The general framework for online linear prediction algorithms proposed by Kivinen and Warmuth [4] is the stepping stone for the cost functions proposed in this section. This framework suggests that an efficient online algorithm must exhibit a balance between its requirements to be *conservative* (i.e., a radical change from $\hat{\mathbf{h}}[n-1]$ to $\hat{\mathbf{h}}[n]$ should be avoided) and *corrective* (i.e., better channel prediction must occur if the same input and output were to be observed at two consecutive times). Usually, the correctiveness and conservativeness are on opposite ends, thus an efficient cost function (to be minimized with respect to $\hat{\mathbf{h}}[n]$) could be written as

$$J[n] = f(\bar{\mathbf{e}}[n]) + \delta D(\hat{\mathbf{h}}[n], \hat{\mathbf{h}}[n-1]), \qquad (23)$$

where $f(\bar{\mathbf{e}}[n])$ is a scalar *loss function* and $D(\hat{\mathbf{h}}[n], \hat{\mathbf{h}}[n-1])$ denotes the scalar *distance function* between $\hat{\mathbf{h}}[n]$ and $\hat{\mathbf{h}}[n-1]$. The distance function may not be a metric. For instance, the Kullback-Leibler (KL) divergence is used as D in [4]. The magnitude of the positive parameter δ keeps the relative balance between correctiveness, induced by the loss function,

and conservativeness, induced by the distance function. In addition, this framework lends itself to incorporating additional properties for $\hat{\mathbf{h}}[n]$. For example, if the filter energy is constrained to be equal to X, then this constraint is introduced via a new Lagrangian multiplier ζ , as follows:

$$J[n] = f(\mathbf{\bar{e}}[n]) + \delta D(\mathbf{\hat{h}}[n], \mathbf{\hat{h}}[n-1]) + \zeta \left(\left\| \mathbf{\hat{h}}[n] \right\|_{2}^{2} - X \right).$$
(24)

This framework is the basis for many adaptive algorithms; cf [19], [37], [28].

B. Quadratic complexity framework

Using the above algorithmic framework, the merits of both robust methods and sparseness constraints can be integrated into a single objective function as:

$$J[n] = \sum_{i=n-L+1}^{n} f(\bar{e}[i]) + \delta \mathbf{r}[n]^{\dagger} \mathbf{P}[n-1]\mathbf{r}[n] + \zeta \left\| \hat{\mathbf{h}}[n] \right\|_{0}.$$
(25)

The loss function $f(\bar{e}[i])$ ensures robustness against outliers, $\delta, \zeta \geq 0$ are regularization parameters and L is the length of the observation window. The choice of L is upper-limited (usually $L \leq 10$) by the time-variation of the channel as well as the hardware memory/complexity requirements. The matrix $\mathbf{P}[n]$ is Hermitian and positive definite whose entries depend on $\hat{\mathbf{h}}[n]$, i.e., $\mathbf{P}[n]$ is a $K \times K$ Riemannian metric tensor. Thus, the term $\mathbf{r}[n]^{\dagger}\mathbf{P}[n-1]\mathbf{r}[n]$ denotes the Riemannian distance between $\hat{\mathbf{h}}[n]$ and $\hat{\mathbf{h}}[n-1]$. The fact that $\hat{\mathbf{h}}[n]$ lies in a non-isotropic (Riemannian) space is based on the prior knowledge that $\mathbf{h}[n]$ must be close to some axis of \mathbb{C}^{K} since most of the filter taps are near zero. In addition, in non-isotropic (Riemannian) spaces, it is well known that the ordinary Euclidean gradient does not represent the steepest ascent direction of a cost function [36]. Thus, regularizing J[n]with a Riemannian distance is well motivated. In addition, the regularizing term $\|\hat{\mathbf{h}}[n]\|_0$ (approximated by (4)) accelerates the convergence of the zero-valued coefficients. The parameter ζ should be tuned close to 10^{-5} for moderate SNR, as suggested in [32].

A plausible question is: "how could one find **P** for any sparse channel?". The study in [35] suggests that the parameter space of a sparse channel may be visualized as a space having the following warping: for regions close to the coordinate axes, distances in any direction orthogonal to those axes should be larger than the Euclidean distances. A good choice is $\mathbf{P}[n]^{-1}=\mathbf{G}[n]$, where $\mathbf{G}[n]$ is the *proportionate matrix* of the PNLMS algorithm [2]. Given the superiority of IPNLMS [7] over PNLMS, we choose $\mathbf{G}[n]$ to be a diagonal matrix with elements $\{g_k[n]\}_{k=0}^{K-1}$, which are computed as follows:

$$\ell_k[n] = (1 - \beta) \frac{\left\| \hat{\mathbf{h}}[n] \right\|_1}{K} + (1 + \beta) \left\| \hat{h}_k[n] \right\|_1, \quad (26)$$

$$g_k[n] = \frac{\ell_k[n]}{\sum_{k=0}^{K-1} \ell_k[n]},$$
(27)

where $\beta \in [-1, 1]$. Substituting (26) in (27), we have

...

$$g_k[n] = \frac{1-\beta}{2K} + (1+\beta) \frac{\left|\hat{h}_k[n]\right|_1}{2\left\|\hat{\mathbf{h}}[n]\right\|_1 + \varepsilon},$$
(28)

where ε denotes a small positive constant to avoid division by zero during initialization of the algorithm. Note that if β =-1, **P**[n] becomes proportional to the identity matrix, i.e., the channel space is Euclidean (the channel is assumed nonsparse). For very sparse channels, β should be chosen between 0 and 0.5. Furthermore, δ is chosen as [7]

$$\delta = (1 - \beta)\delta'/2K,\tag{29}$$

where δ' is the regularization parameter of the normalized least-mean-square (NLMS) algorithm.

The general type of the algorithm is derived by computing $\nabla_{\mathbf{r}[n]^*} J[n]=0$. The algebra is presented in the Appendix B. The channel update equations are summarized below:

$$\mathbf{e}[n]^* = \mathbf{y}[n]^* - \mathbf{U}[n]^{\dagger} \hat{\mathbf{h}}[n-1], \tag{30}$$

$$\mathbf{A}[n] = \mathbf{G}[n-1]\mathbf{U}[n], \tag{31}$$

$$\mathbf{B}[n] = \left(\mathbf{U}[n]^{\dagger}\mathbf{A}[n] + \delta \mathbf{Q}[n]^{-1}\right)^{-1}, \qquad (32)$$

$$\mathbf{C}[n] = \mathbf{A}[n]\mathbf{B}[n], \tag{33}$$

$$\mathbf{D}[n] = \mathbf{G}[n-1] - \mathbf{C}[n]\mathbf{A}[n]^{\dagger}, \qquad (34)$$

$$\nu_k[n] = e^{-\eta |h_k[n]|_1} \operatorname{csgn}\left(\hat{h}_k[n]\right), k = 0, \dots, K-1, \quad (35)$$

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu \mathbf{C}[n]\mathbf{e}[n]^* - \frac{\mu\zeta\eta}{2\delta}\mathbf{D}[n]\nu[n-1], \quad (36)$$

where $\mu \in (0,1]$ is the step-size parameter. Since $\mathbf{G}[n]$ is diagonal and $L \ll K$, the required number of algebraic operations is $O(K^2)$. Initialization of the algorithm starts with $\hat{\mathbf{h}}[0]=\mathbf{0}$.

In light of the above framework, new and existing algorithms are derived as follows:

- if $f(\bar{e})=\bar{e}\bar{e}^*$, then $\mathbf{Q}[n]=\mathbf{I}_L$ and (30)-(36) describe the L_0 -IPAPA [14]. If in addition β =-1 and ζ =0, then the L_0 -IPAPA reduces to the affine projection algorithm (APA) [33],
- if $q(\bar{e})$ is given by (12), (30)-(36) will be called L₀-IPMAPA (L₀ norm-improved-proportionate M-estimate affine projection algorithm) hereafter. If in addition ζ =0, then the L₀-IPMAPA reduces to IPMAPA (note O(K)computational complexity),
- if $q(\bar{e})$ is given by (18), (30)-(36) will be called L₀-IPpNAPA (L₀ norm-improved-proportionate p-norm affine projection algorithm) hereafter. If in addition ζ =0, then the L₀-IPpNAPA reduces to IPpNAPA (note O(K) computational complexity).

C. Linear complexity framework

The $O(K^2)$ complexity of L₀-IPMAPA and L₀-IPpNAPA may become objectionable from a hardware perspective when long filters are required. Such is the case in acoustic echo or broadband underwater acoustic channels. Towards reducing the computational complexity, we propose the following cost function:

$$J[n] = \sum_{i=n-L+1}^{n} f(\bar{e}[i])$$
(37)

subject to
$$\mathbf{r}[n]^{\dagger}\mathbf{P}[n-1]\mathbf{r}[n] \le \mu^2$$
. (38)

Using Lagrange multipliers, the modified cost function becomes

$$J[n] = \sum_{i=n-L+1}^{n} f(\bar{e}[i]) + \delta\left(\mathbf{r}[n]^{\dagger} \mathbf{G}^{-1}[n-1]\mathbf{r}[n] - \mu^{2}\right)$$
(30)

where δ here is the Lagrange multiplier.

Setting $\nabla_{\mathbf{r}[n]^*} J[n]=0$, we have

$$-\sum_{i=n-L+1}^{n}\psi(\bar{e}[i])\mathbf{u}[i] + \delta \mathbf{G}^{-1}[n-1]\mathbf{r}[n] = 0.$$
(40)

Note that it is tedious to solve for $\mathbf{r}[n]$ using (40) since $\{\psi(\bar{e}[i])\}_{i=n-L+1}^{n}$ depends on $\hat{\mathbf{h}}[n]$. We circumvent this issue by assuming $\bar{e}[i] \simeq e[i], i=n-L+1, \ldots, n$, at steady-state. Then, $\mathbf{r}[n]$ is expressed as:

$$\mathbf{r}[n] = \frac{1}{\delta} \mathbf{G}[n-1] \mathbf{U}[n] \psi[n], \qquad (41)$$

where $\psi[n] = [\psi(e[n]) \dots \psi(e[n - L + 1])]^{T}$. To obtain the Lagrange multiplier δ , we substitute (41) into (38). Hence, we have

$$\delta = \frac{1}{\mu} \sqrt{\left\|\overline{\mathbf{x}}[n]\right\|_{2}^{2}},\tag{42}$$

$$\overline{\overline{\mathbf{x}}}[n] = \mathbf{G}[n-1]^{1/2} \mathbf{U}[n] \psi[n], \qquad (43)$$

where $\mathbf{G}[n]^{1/2}$ denotes the Cholesky decomposition of $\mathbf{G}[n]$. Recall that $\mathbf{G}[n]$ is diagonal and so $\mathbf{G}[n]^{1/2}$ is equal to the square root of the entries of $\mathbf{G}[n]$. Hence, the channel update equation is given by the formula

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu \frac{\mathbf{G}[n-1]^{1/2}\overline{\mathbf{x}}[n]}{\sqrt{\kappa + \|\overline{\mathbf{x}}[n]\|_2^2}},$$
(44)

where $\mu > 0$ and κ is a small positive constant used to avoid possible division by zero during initialization of the algorithm. Since $L \ll K$, it is straightforward to see that (44) requires O(K) operations per datum. Also note the following:

- if $\psi(e[i])$ is given by (11), then (44) will be called NGMAPA (natural gradient-based M-estimate affine projection algorithm) hereafter,
- if ψ(e[i]) is given by (17), then (44) will be called NGpNAPA (natural gradient-based p-norm affine projection algorithm) hereafter,
- if $\psi(e[i])$ is given by (14), then (44) will be called NGAPSA (natural gradient-based affine projection sign algorithm) hereafter. If β =-1, then the NGAPSA reduces to the complex form of the APSA [37]. If *L*=1, then the NGAPSA reduces to the NGSA (natural gradient sign algorithm).

V. Performance under $S\alpha S$ Noise

A theoretical analysis of the proposed frameworks is exceptionally difficult because $S\alpha S$ distributions do not possess second order moments. In this section, the effectiveness of all algorithms derived by the two frameworks is tested by running numerical simulations as well as analyzing experimental data. Impulse responses obtained from broadband underwater acoustic communications channels are used to support our findings.



Fig. 1. The time-varying channel used in simulations. The x-axis shows multipath delay, the y-axis shows absolute time and the z-axis shows the channel amplitude in linear scale.

A. Simulation results

The time-varying channel to be estimated is shown in Figure 1. This channel corresponds to a wideband underwater acoustic link that was measured during the Focused Acoustic Fields (FAF) experiment. Each channel snapshot was estimated at rate of 6250 Hz. Figure 1 is generated by keeping each channel snapshot fixed for 8 ms. For further details about the FAF experimental setup, the interested reader is directed to [14]. The channel to be identified is clearly sparse and the filter length required to capture the entire impulse response is 371 taps. The simulated input signal is independent white complex Gaussian noise with unit power and the output is generated according to (1). The simulated $S\alpha S$ noise is generated in passband using [38] and then is shifted to baseband using the lowpass filter of the FAF experiment. Four types of noise series are considered, i.e., 1) high rate of impulses in low SNR ($\alpha = 1.2, E_s/N_0 = 15 \text{ dB}$), 2) high rate of impulses in high SNR ($\alpha = 1.2, E_s/N_0 = 25 \text{ dB}$), 3) low rate of impulses in low SNR ($\alpha = 1.65, E_s/N_0=15 \text{ dB}$), and 4) low rate of impulses in high SNR ($\alpha = 1.65, E_s/N_0=25 \text{ dB}$). The performance measure is the normalized misadjustment (in dB), $20 \log_{10}(||\mathbf{h}[n] - \hat{\mathbf{h}}[n]||_2/||\mathbf{h}[n]||_2)$, and is computed after averaging 100 independent runs.

The parameters of all algorithms are chosen as follows:

- λ =0.995 for all RLS-type filters,
- λ_{σ} =0.99 for all M-estimate filters,
- $\eta=10$ for all filters with L₀ norm,
- β =0.5, L=4 for all NG-type filters,
- μ =0.1, δ' =10, ζ =5 · 10⁻⁴ for L₀-IPpNAPA, IPpNAPA, L₀-IPMAPA and IPMAPA,
- ζ =0.5 for all RLS-type filters with L₀ norm;
- μ =0.25, ξ =2 γ , Δ =100 γ for NGpNAPA and NGMAPA;
- $p=\alpha$ -0.15², $\xi=2\gamma$ for L₀-RLP, RLP, L₀-IPpNAPA, and NGpNAPA.

In Figures 2(a)-(d), we have plotted the curves of all RLStype algorithms for different values of α and E_s/N_0 . The following observations are in order:

 ^{2}p must be as close to α as possible but not equal to α . In practice, α is estimated so one should be conservative on the choice of p.



Fig. 2. Learning curves of RLS-type algorithms for different α and Es/N₀; RLM: (a) N_w=12, (b) N_w=12, (c) N_w=40, (d) N_w=40. RLP/L₀-RLP: (a) Δ =10 γ , (b) Δ =40 γ , (c) Δ =6 γ , (d) Δ =100 γ .



Fig. 3. Learning curves for different α and Es/N₀. L₀-IPMAPA: (a) N_w=12, (b) N_w=12, (c) N_w=12, (d) N_w=40. L₀-IPpNAPA: (a) Δ =10 γ , (b) Δ =40 γ , (c) Δ =10 γ , (d) Δ =100 γ .



Fig. 4. Learning curves of APSA, NGSA, NGAPSA, NGMAPA, and NGpNAPA for different α and Es/N₀



Fig. 5. Learning curves of L₀-RLP, NGpNAPA, and L₀-IPpNAPA for different α and Es/N₀.

- L₀-RLP is consistently robust and exhibits the lowest misadjustment for α =1.2 regardless the E_s/N_0 .
- In spite of the SNR, L₀-RLM cannot cope with impulsive noise for α =1.2. RLM [34] shows similar behavior (not shown for brevity). On the other hand, L₀-RLM shows similar misadjustment with L₀-RLP and L₀-RLSA for α =1.65. These results are contrary to the general claim that RLM is robust in S α S noise.
- The L₀-RLSA is consistently robust but exhibits slower convergence rate than L₀-RLP for α=1.2. Note though that L₀-RLSA is easier to use than L₀-RLM or L₀-RLP since it depends on fewer parameters.
- The sparseness effect of the L_0 norm improves RLP [26] for all pairs of α and E_s/N_0 . Similar results hold for RLM and RLSA and are omitted for brevity.
- As expected, L₀-RLS is not robust against impulsive noise (its misadjustment is greater than 0 dB for α =1.2 and is not shown for visualization purposes).

Figures 3(a) through (d) present a comparison of the misadjustment of all algorithms derived from (30)-(36). We make the following observations:

- L₀-IPpNAPA is consistently robust and exhibits the lowest misadjustment for α =1.2 regardless the E_s/N_0 .
- In spite of the SNR, L₀-IPMAPA cannot cope with impulsive noise for α=1.2. However, it shows similar performance with L₀-IPpNAPA for α=1.65.
- The sparseness effect of the L₀ norm is validated since L₀-IPpNAPA is consistently superior over both IPpNAPA and IPMAPA.
- L₀-IPAPA [14] fails for α =1.2, it has poor performance for α =1.65 and E_s/N_0 =15 dB, but it shows similar performance with L₀-IPpNAPA when α =1.65 and E_s/N_0 =25 dB.

All linear complexity algorithms stemming from (44) are compared in Figures 4(a)-(d). Note the following:

- The NGpNAPA consistently achieves the best convergence rate all pairs of α and E_s/N_0 .
- NGpNAPA, NGMAPA and NGAPSA consistently demonstrate similar channel tracking.
- Regarding the effect of the observation window L, the NGSA shows consistently slower convergence speed among all other sparse algorithms, however, the algorithm exhibits the best tracking for $E_s/N_0=15$ dB regardless the choice of α . This result is justified by noting that the error term in the cost function becomes smaller (recall NGSA uses L=1) in steady state and so NGSA becomes more robust (conservative) against impulses.
- The sparseness of all proposed algorithms is confirmed since the non-sparse (but robust) APSA [37] exhibits poor performance for all pairs of α and E_s/N_0 .

In Figures 5(a)-(d), we compare the best algorithms from each framework. In particular, L_0 -RLP is compared with L_0 -IPpNAPA and NGpNAPA. In light of the results presented, we observe that:

- The NG-type algorithms outperform L₀-RLP.
- Although L₀-IPpNAPA uses a smaller step-size μ than NGpNAPA, it shows faster convergence for α =1.2. Furthermore, L₀-IPpNAPA exhibits up to 2dB better tracking than NGpNAPA when α =1.6 despite the SNR.

B. Experimental results

The dataset analyzed here was recorded during the ROMA-NIS 2010 experiment in the area of Selat Pauh in Singapore waters. The transmitter was mounted on a rigid tripod 2.5 m above the seabed. The receiver was a 1.3 m diameter, 2dimensional circular array consisted of 508 acoustic sensors [40]. Here, we analyze data from one sensor of the array. The average water depth was about 15 m and the transmitterreceiver range was about 80 m. The transmitted signal was a 10⁴ bits/s-rate, BPSK-modulated, pseudo-random data. The PN-sequence was pulse-shaped by a square-root cosine filter with roll-off factor 0.25 and truncation length \pm 5 symbol intervals. The resulting waveform was modulated onto a 30 kHz carrier frequency.

A notable feature of this channel is its very high SNR due to the very short range. To test our algorithms in a realistic scenario, on-site recorded ambient noise is appropriately scaled and added to the passband received signal so that $E_s/N_0=15$ dB. Prior to adding extra noise to the received signal, the noise series is bandpass filtered for estimating the S α S parameters α (based on [41]) and γ (based on [42]). In particular, we find that $\alpha=1.44$ and $\gamma=715.28$. Figure 6(a) illustrates the received passband signal after noise addition and Figure 6(b) validates that the ambient noise is S α S distributed.

Prior to channel estimation, the received signal is shifted to baseband, low-pass filtered, and downsampled to 1 sample/bit. Since the channel is unknown, the misadjustment cannot be applied here. Consequently, the mean absolute error (MAE), defined as

$$MAE[n] = \frac{1}{n} \sum_{i=1}^{n} |y[i] - \hat{\mathbf{h}}[i-1]^{\dagger} \mathbf{u}[i]|_{1}$$
$$= \frac{1}{n} \sum_{i=1}^{n} \left| \left(\mathbf{h}[i] - \hat{\mathbf{h}}[i-1] \right)^{\dagger} \mathbf{u}[i] + w[n] \right|_{1}$$
(45)

is utilized as a performance metric. Note that when the channel estimate is very close to the true channel, the MAE converges to the first order moment $E\{|w[n]|_1\}$. The dB scale of the MAE is defined as $20 \log_{10}(MAE)$. It is important to stress that we do not change the parameter values used in simulations except for Δ =10 γ (L₀-RLP and L₀-IPpNAPA) and N_w =12 (L₀-RLM and L₀-IPMAPA).

The time evolution of the estimated channel amplitude is shown in Figure 7. For a sampling rate of 1 sample/bit, the required length of the adaptive filter is 206 taps. Clearly, the acoustic channel has a sparse multipath structure and any amplitude fluctuations are attributed to environmental changes since both the transmitter and the receiver are stationary.

In light of the results shown in Figure 8(a), we observe that L_0 -RLP shows the best convergence rate but similar channel tracking with L_0 -RLSA. In addition, L_0 -RLP outperforms RLP[26] validating the sparseness effect of the L_0 norm. L_0 -RLM achieves inferior performance against L_0 -RLP and L_0 -RLSA. These results are in close agreement with the simulation results shown in Figure 2(a).

Figure 8(b) shows the MAE performances of all algorithms derived by (30)-(36). L_0 -IPpNAPA presents better convergence rate but same tracking as L_0 -IPMAPA. As expected, L_0 -IPAPA shows higher MAE than its robust counterparts. It is reassuring to note that these results are in agreement with those shown in Figure 3(c).



Fig. 6. (a) received passband signal for Es/N₀=15dB; (b) goodness of fit of passband ambient noise series to $S\alpha S$ distribution.



Fig. 7. Snapshots of the ROMANIS channel. The x-axis shows multipath delay, the y-axis shows absolute time and the z-axis shows the channel amplitude in linear scale. The snapshots are generated at the bit rate.

Figure 8(c) compares all linear complexity algorithms generated by (44). Clearly, the NGpNAPA exhibits the best performance. The NGSA shows the largest MAE among all sparse algorithms. As expected, the sparse-indifferent APSA [37] shows the worst behavior. Observe that these experimental results are in close agreement with the simulation results shown in Figure 4(c).

Figure 8(d) combines the results from Figures 8(a)-(c) and compares L_0 -RLP with L_0 -IPpNAPA and NGpNAPA. The NG-type algorithms outperform L_0 -RLP validating the superiority of the NG-type framework. The L_0 -IPpNAPA achieves faster convergence than NGpNAPA. On the other hand, NGpNAPA demonstrates slightly better channel tracking. Again, note that these results are in close agreement with the simulation results shown in Figure 5(a).

VI. DISCUSSION AND CONCLUSIONS

We investigated two frameworks for developing complexvalued sparse robust adaptive filters. Although the focus of this paper was on sparse channel estimation in the presence of $S\alpha S$ noise, the proposed algorithms can be applied in any field where noisy samples are obtained from a sparse linear time-varying system.

The first framework was inspired by the RLS algorithm and proposed a cost function that coupled robust methods for outlier suppression with a L₀ norm penalty. Three new $O(K^2)$ algorithms were generated: the L₀-RLM, the L₀-RLP, and the L₀-RLSA. Both computer simulations and experimental data analysis verified that these algorithms improve the traditional robust algorithms RLM [34] and RLP [26]. We also found that L₀-RLP and L₀-RLSA were consistently robust regardless of noise parameter α while L₀-RLM/RLM lost robustness when α was close to one. Our results contradict the claim in [34] that RLM is generally robust in S α S noise.

The second framework took advantage of the non-isotropic (Riemannian) space of the channel and generated robust algorithms based on NG adaptation. Two $O(K^2)$ algorithms, i.e., the L₀-IPMAPA and the L₀-IPpNAPA were introduced. Our data analysis revealed that the L₀-IPpNAPA was robust despite the choice of α but the L₀-IPMAPA (like L₀-RLM) lost robustness when α was close to one. In addition, three O(K) algorithms were introduced, i.e., the NGMAPA, the NGpNAPA and the NGAPSA. These algorithms demonstrated firmly robust performance for all S α S noise regimes and exhibited comparable channel tracking with their $O(K^2)$ counterparts. This result is very promising from a hardware implementation point of view since one could start with an $O(K^2)$ algorithm to achieve fast convergence followed by an O(K) algorithm for fast tracking.

Elaborating on the loss function $f(\bar{e})$, our results demonstrated that a mixture of L₂ and L_p norms is more efficient in terms of convergence rate than the L₁ norm or the Hampel's M-estimate function.

Finally, our results demonstrated the clear superiority of the NG-framework over the RLS-framework in sparse channels. This is due to the fact that the NG-type filters use the Riemannian distance to modify the gradient search direction for faster adaptation. It would be intriguing to incorporate a Riemannian distance in the RLS-type framework to test its algorithmic performance. We leave this challenge as a future research direction.

APPENDIX A DERIVATION OF (5)-(9)

Computing $\nabla_{\hat{\mathbf{h}}[n]^*} J[n]$, where J[n] is given by (3), we have:

$$\nabla_{\hat{\mathbf{h}}[n]^*} J[n] = \nabla_{\hat{\mathbf{h}}[n]^*} \left(\sum_{i=0}^n \lambda^{n-i} f(\bar{e}[i]) \right) + \nabla_{\mathbf{h}[n]^*} \left(\zeta \left\| \hat{\mathbf{h}}[n] \right\|_0 \right).$$
(46)



Fig. 8. Learning curves of proposed algorithms based on experimental data.

Computing each of the above terms individually, we have

$$\nabla_{\hat{\mathbf{h}}[n]^*} \left(\sum_{i=0}^n \lambda^{n-i} f(\bar{e}[i]) \right)$$
$$= \sum_{i=0}^n \lambda^{n-i} \frac{\partial f(\bar{e}[i])}{\partial \bar{e}[i]} \frac{\partial \bar{e}[i]}{\partial \hat{\mathbf{h}}[n]^*}$$
(47)

$$= -\sum_{i=0}^{n} \lambda^{n-i} \psi(\bar{e}[i]) \mathbf{u}[i] = -\sum_{i=0}^{n} \lambda^{n-i} q(\bar{e}[i]) \bar{e}[i]^* \mathbf{u}[i] \quad (48)$$

$$= -\sum_{i=0}^{n} \lambda^{n-i} q(\bar{e}[i]) \left(y[i]^* - \mathbf{u}[i]^{\dagger} \hat{\mathbf{h}}[n] \right) \mathbf{u}[i] \qquad (49)$$

$$= -\sum_{i=0}^{n} \lambda^{n-i} q(\bar{e}[i]) y[i]^* \mathbf{u}[i]$$
$$+ \sum_{i=0}^{n} \lambda^{n-i} q(\bar{e}[i]) \left(\mathbf{u}[i] \mathbf{u}[i]^{\dagger} \right) \hat{\mathbf{h}}[n]$$
(50)

where (47) holds due to the chosen gradient operator [29]. The gradient of $\|\hat{\mathbf{h}}[n]\|_0$ with respect to $h_k[n]^*, k = 0, \dots, K-1$, is equal to

$$\nabla_{h_k[n]^*} \left(\zeta \left\| \hat{\mathbf{h}}[n] \right\|_0 \right) = \frac{\zeta \eta}{2} e^{-\eta \left| \hat{h}_k[n] \right|_1} \operatorname{csgn} \left(\hat{h}_k[n] \right).$$
(51)

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

$$\nu_k[n] = e^{-\eta \left| \hat{h}_k[n] \right|_1} csgn\left(\hat{h}_k[n] \right), k = 0, \dots, K - 1.$$
 (52)



Computing $\nabla_{\hat{\mathbf{h}}[n]^*}J[n]{=}0$ and after some algebra we have

$$\sum_{i=0}^{n} \lambda^{n-i} q(\bar{e}[i]) \left(\mathbf{u}[i] \mathbf{u}[i]^{\dagger} \right) \mathbf{\hat{h}}[n]$$
$$= \sum_{i=0}^{n} \lambda^{n-i} q(\bar{e}[i]) y[i]^* \mathbf{u}[i] - \frac{\zeta \eta}{2} \nu[n].$$
(53)

From the above equation, it is not straightforward to find $\hat{\mathbf{h}}[n]$ since knowledge of $\bar{e}[i]$ implies knowledge of $\hat{\mathbf{h}}[n]$. However, at steady-state, it is plausible to assume that $\bar{e}[n] \simeq e[n]$. Following [34], [26], we define the weighted complex auto-correlation matrix, $\boldsymbol{\Phi}[n]$, and the weighted complex cross-correlation vector, $\mathbf{p}[n]$, as:

$$\mathbf{\Phi}[n] = \sum_{i=0}^{n} \lambda^{n-i} q(e[i]) \left(\mathbf{u}[i] \mathbf{u}[i]^{\dagger} \right)$$
$$= \lambda \mathbf{\Phi}[n-1] + q(e[n]) \left(\mathbf{u}[n] \mathbf{u}[n]^{\dagger} \right), \qquad (54)$$

$$\mathbf{p}[n] = \sum_{i=0} \lambda^{n-i} q(e[i]) y[i]^* \mathbf{u}[i]$$
$$= \lambda \mathbf{p}[n-1] + q(e[n]) y[n]^* \mathbf{u}[n], \tag{55}$$

respectively. Hence, (53) can be written as

ć

$$\mathbf{\Phi}[n]\hat{\mathbf{h}}[n] = \mathbf{p}[n] - \frac{\zeta \eta}{2}\nu[n]$$
(56)

and so

$$\hat{\mathbf{h}}[n] = \mathbf{\Phi}[n]^{-1} \left(\mathbf{p}[n] - \frac{\zeta \eta}{2} \nu[n] \right).$$
(57)

$$\mathbf{t}[n] = \mathbf{p}[n] - \frac{\zeta \eta}{2} \nu[n] = \mathbf{p}[n] - \tilde{\nu}[n]$$
(58)

before invoking the matrix inversion lemma to compute $\Phi[n]^{-1}$. Elaborating as in [39], we note that

$$\mathbf{t}[n] = \lambda \mathbf{p}[n-1] + q(e[n])y[n]^* \mathbf{u}[n] - \tilde{\nu}[n] + \lambda \tilde{\nu}[n-1] - \lambda \tilde{\nu}[n-1]$$
(59)

$$\simeq \lambda \left(\mathbf{p}[n-1] - \tilde{\nu}[n-1]\right) + q(e[n])y[n]^* \mathbf{u}[n] + (\lambda - 1)\tilde{\nu}[n-1]$$
(60)

$$= \lambda \mathbf{t}[n-1]$$
(60)

+
$$q(e[n]) y[n]^* \mathbf{u}[n] + (\lambda - 1)\tilde{\nu}[n-1]$$
 (61)

where (61) is obtained by assuming $\tilde{\nu}[n-1] \simeq \tilde{\nu}[n]$. Applying the matrix inversion lemma to compute $\Phi[n]^{-1}$, we have that

$$\mathbf{\Phi}[n]^{-1} = \lambda^{-1} \left(\mathbf{\Phi}[n-1]^{-1} - \mathbf{k}[n]\mathbf{u}[n]^{\dagger} \mathbf{\Phi}[n-1]^{-1} \right),$$
(62)

where $\mathbf{k}[n]$ is given by

$$\mathbf{k}[n] = \frac{q(e[n])\mathbf{\Phi}[n-1]^{-1}\mathbf{u}[n]}{\lambda + q(e[n])\mathbf{u}[n]^{\dagger}\mathbf{\Phi}[n-1]^{-1}\mathbf{u}[n]}.$$
 (63)

Hence, the update for $\hat{\mathbf{h}}[n]$ becomes

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mathbf{k}[n]e[n]^* + \frac{\lambda - 1}{\lambda} \frac{\zeta \eta}{2} \left(\mathbf{I}_K - \mathbf{k}[n]\mathbf{u}[n]^\dagger \right) \mathbf{\Phi}[n-1]^{-1}\nu[n-1].$$
(64)

APPENDIX B DERIVATION OF (30)-(36)

Computing $\nabla_{\mathbf{r}[n]^*} J[n]$, where J[n] is given by (25), we have:

$$\nabla_{\mathbf{r}[n]^*} J[n] = \nabla_{\mathbf{r}[n]^*} \left(\sum_{i=n-L+1}^n f(\bar{e}[i]) \right) + \nabla_{\mathbf{r}[n]^*} \left(\delta \mathbf{r}[n]^{\dagger} \mathbf{P}[n-1] \mathbf{r}[n] \right) + \nabla_{\mathbf{r}[n]^*} \left(\zeta \left\| \mathbf{\hat{h}}[n] \right\|_0 \right).$$
(65)

Computing each of the above terms individually, we have

$$\nabla_{\mathbf{r}[n]^*} \left(\sum_{i=n-L+1}^n f(\bar{e}[i]) \right)$$
$$= -\sum_{i=n-L+1}^n \psi(\bar{e}[i]) \mathbf{u}[i] = -\sum_{i=n-L+1}^n q(\bar{e}[i]) \bar{e}[i]^* \mathbf{u}[i] \quad (66)$$

$$= -\sum_{i=n-L+1}^{n} q(\bar{e}[i])(e[i]^* - \mathbf{u}[i]^{\dagger}\mathbf{r}[n])\mathbf{u}[i]$$
(67)

$$= -\sum_{i=n-L+1}^{n} q(\bar{e}[i])e[i]^*\mathbf{u}[i] + \sum_{i=n-L+1}^{n} q(\bar{e}[i])\left(\mathbf{u}[i]\mathbf{u}[i]^{\dagger}\right)\mathbf{r}[n]$$
(68)

$$= -\mathbf{U}[n]\mathbf{Q}[n]\mathbf{e}[n]^* + \mathbf{U}[n]\mathbf{Q}[n]\mathbf{U}[n]^{\dagger}\mathbf{r}[n], \qquad (69)$$

where $\mathbf{Q}[n]$ is an $L \times L$ diagonal matrix with elements $q(\bar{e}[n]), ..., q(\bar{e}[n-L+1])$. We also have

$$\nabla_{\mathbf{r}[n]^*} \left(\delta \mathbf{r}[n]^{\dagger} \mathbf{P}[n-1] \mathbf{r}[n] \right) = \delta \mathbf{P}[n-1] \mathbf{r}[n].$$
(70)

The gradient of $\|\hat{\mathbf{h}}[n]\|_0$ with respect to $r_k[n]^*, k = 0, \ldots, K-1$, is equal to $\nu[n]$ from (52).

Setting $\nabla_{\mathbf{r}[n]^*} J[n]=0$ and combining terms we have the following vector equation:

$$\left(\delta \mathbf{P}[n-1] + \mathbf{U}[n]\mathbf{Q}[n]\mathbf{U}[n]^{\dagger}\right)\mathbf{r}[n] + \frac{\zeta\eta}{2}\nu[n] = \mathbf{U}[n]\mathbf{Q}[n]\mathbf{e}[n]^{*}.$$
(71)

From the above equation, we note that it is tedious to solve for $\mathbf{r}[n]$ since $\nu[n]$ depends on $\hat{\mathbf{h}}[n]$ in a non-linear fashion. At steady-state, however, it is plausible to assume that $\bar{e}[n] \simeq e[n]$ and thus, $\nu[n] \simeq \nu[n-1]$. Using this assumption, we can solve for $\mathbf{r}[n]$ by using the matrix inversion lemma [33]. Thus, we have:

$$\mathbf{r}[n] = \mathbf{C}[n]\mathbf{e}[n]^* - \frac{\zeta \eta}{2\delta} \mathbf{D}[n]\nu[n-1], \qquad (72)$$

where

$$\mathbf{A}[n] = \mathbf{P}[n-1]^{-1}\mathbf{U}[n] = \mathbf{G}[n-1]\mathbf{U}[n]$$
 (73)

$$\mathbf{B}[n] = \left(\mathbf{U}[n]^{\dagger}\mathbf{A}[n] + \delta \mathbf{Q}[n]^{-1}\right)^{-1}$$
(74)

$$\mathbf{C}[n] = \mathbf{A}[n]\mathbf{B}[n] \tag{75}$$

$$\mathbf{D}[n] = \mathbf{G}[n-1] - \mathbf{C}[n]\mathbf{A}[n]^{\dagger}$$
(76)

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:

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu \mathbf{C}[n]\mathbf{e}[n]^* - \frac{\mu\zeta\eta}{2\delta}\mathbf{D}[n]\nu[n-1].$$
(77)

ACKNOWLEDGMENT

The authors are indebted to the personnel of the NATO Undersea Research Centre (NURC) for carrying out the FAF experiment. Also, the authors would like to thank Dr. Venugopalan Pallayil and Mr. Unnikrishnan K. C. for their leadership and help during the ROMANIS'10 experiment.

REFERENCES

- W. F. Schreiber, "Advanced television systems for terrestrial broadcasting: some problems and some proposed solutions," *Proc. IEEE*, vol. 83, no. 6, pp. 958–981, June 1995.
- [2] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Process.*, vol. 8, no. 5, pp. 508–518, 2000.
- [3] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for realtime digital underwater acoustic communications," in *Proc. 1995 MTS/IEEE OCEANS*, vol. 3, pp. 1417–1422.
- [4] J. Kivinen and M. K.Warmuth, "Exponentiated gradient versus gradient descent for linear predictors," *Inform. Comput.*, vol. 132, no. 1, pp. 1–64, 1997.
- [5] R. E. Mahony and R. C.Williamson, "Prior knowledge and preferential structures in gradient descent learning algorithms," *J. Machine Learn. Res.*, vol. 1, pp. 311–355, 2001.
- [6] R. K. Martin *et al*, "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal. Process.*, vol. 50, no. 8, pp. 1883–1894, 2002.
- [7] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc.* 2002 IEEE ICASSP, vol. 2, pp. 1881–1884.
- [8] O. Hoshuyama, R. A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proc. 2004 IEEE ICASSP*, vol. 4, pp. 161–164.

- [9] B. D. Rao and B. Song, "Adaptive filtering algorithms for promoting sparsity," in *Proc. 2003 IEEE ICASSP*, vol. 6, pp. VI361–VI364.
- [10] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: the sparse RLS algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4013–4025, 2010.
- [11] 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, 2010.
- [12] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted balls," *IEEE Trans. Signal Process.*, vol. 59, pp. 936–952, 2011.
- [13] Y. Murakami, et al., "A sparse adaptive filtering using time-varying softthresholding techniques," in Proc. 2010 IEEE ICASSP, pp. 3734–3737.
- [14] K. Pelekanakis and M. Chitre, "New sparse adaptive algorithms based on the natural gradient and the L₀-norm," *IEEE J. Ocean. Eng.*, pp. 323–332, vol. 38, 2013.
- [15] B. Aazhang and H. V. Poor, "Peformance of DS/SSMA communications in impulsive channels—part I: linear correlation receivers," *IEEE Trans. Commun.*, vol. COMM-35, pp. 1179–1187, 1987.
- [16] T. Gänsler *et al.*, "Double-talk robust fast converging algorithms for network echo cancelation," *IEEE Trans. Speech, Audio Process.*, vol. 8, no. 6, pp. 656–663, 2000.
- [17] M. A. Chitre, J. R. Potter, and S. H. Ong, "Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise," *IEEE J. Ocean. Eng.*, vol. 31, no. 2, pp. 497–503, 2006.
- [18] M. Bouvet and S. C. Schwartz, "Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, pp. 621–626, 1989.
- [19] L. R. Vega *et al.*, "A family of robust algorithms exploiting sparsity in adaptive filters," *IEEE Trans. Speech, Audio Process.*, vol. 17, no. 4, pp. 572–581, 2009.
- [20] Z. Yang, Y. R. Zheng, and S. L. Grant, "Proportionate affine projection sign algorithms for network echo cancellation," *IEEE Trans. Speech, Audio Process.*, vol. 19, no. 8, pp. 2273–2284, 2011.
- [21] T. Yamamoto, M. Yamagishi, and L. Yamada, "Adaptive proximal forward-backward splitting for sparse system identification under impulsive noise," in *Proc. 2012 EUSIPCO*, pp. 2620–2624.
- [22] P. J. Huber, Robust Statistics. John, John Wiley & Sons, 1981.
- [23] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: new methods and results for class A and class B noise models," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1129–1149, 1999.
- [24] R. Adler, R. Feldman, and M. S. Taqqu, Ed., A Practical Guide to Heavy Tails Statistical Techniques for Analyzing Heavy-Tailed Distributions. Birkhauser, 1997.
- [25] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proc. IEEE*, vol. 81, pp. 986–1009, 1993.
- [26] A. T. Georgiadis and B. Mulgrew, "Adaptive Bayesian decision feedback equaliser for alpha-stable noise environments," *Signal Process.*, vol. 81, no. 8, pp. 1603–1623, 2001.
- [27] M. Belge and E. L. Miller, "A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise," *IEEE Signal Process. Lett.*, vol. 7, no. 4, pp. 86–89, 2000.
- [28] K. Pelekanakis and M. Chitre, "A class of affine projection filters that exploit sparseness under symmetric alpha-stable noise," in *Proc. 2013 IEEE/MTS Oceans.*
- [29] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Microwaves, Optics and Antennas, IEE Proceedings H*, vol. 130, no. 1, pp. 11–16, 1983.
- [30] A. Mahmood, M. Chitre, and M. Armand, "PSK communication with passband additive symmetric α-stable noise," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2990–3000, 2012.
- [31] P. Petrus, "Robust Huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, 1999.

- [32] G. Su *et al.*, "Performance analysis of L₀-norm constraint least mean square algorithm," *IEEE Trans. Signal Process.* vol. 60, pp. 2223–2235, 2012.
- [33] S. Haykin, Adaptive Filter Theory, 4th ed. Prentice-Hall, 2002.
- [34] S. C. Chan and Y. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.
- [35] S. L. Gay and S. C. Douglas, "Normalized natural gradient adaptive filtering for sparse and nonsparse systems," in *Proc. 2002 IEEE ICASSP*, vol. 2, pp. 1405–1408.
- [36] S. I. Amari and S. C. Douglas, "Why natural gradient?" in Proc. 1998 IEEE ICASSP, vol. 2, pp. 1213–1216.
- [37] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 327–330, Apr. 2010.
- [38] J. P. Nolan, STABLE program for Windows. Available: http://academic2.american.edu/jpnolan/stable/stable.html.
- [39] E. M. Eksioglu, "Sparsity regularized recursive least squares adaptive filtering," *IET Signal Process.*, vol. 5, pp. 480–487, 2011.
 [40] M. Chitre, S. Kuselan, and V. Pallayil, "Ambient noise imaging in warm
- [40] M. Chitre, S. Kuselan, and V. Pallayil, "Ambient noise imaging in warm shallow waters; robust statistical algorithms and range estimation," J. of the Acoustical Society of America, vol. 132, no. 2, pp. 838–847, 2012.
- [41] J. H. McCulloch, "Simple consistent estimators of stable distribution parameters," *Commun. Statist. — Simula.*, vol. 15, no. 4, pp. 1109–1136, 1986.
- [42] E. F. Fama and R. Roll, "Parameter estimates for symmetric stable distributions," J. Amer. Statist. Assoc., vol. 66, pp. 331–338, June 1971.



Konstantinos Pelekanakis (S'06–M'09) received his Diploma from the Department of Electronic and Computer Engineering, Technical University of Crete, Greece, in 2001 and his 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) at the National University of Singapore (NUS). His current research lies in the areas of acoustic communications and learning

algorithms for sparse system identification.



Mandar Chitre (M'03–SM'11) received B.Eng. (Hons) and M.Eng. degrees in electrical engineering from the National University of Singapore (NUS), a M.Sc. degree in bioinformatics from the Nanyang Technological University (NTU), Singapore, and a Ph.D. degree from NUS. From 1997 to 1998, he worked with the Acoustic Research Laboratory (ARL), NUS in Singapore 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 a

Deputy Head and is now the Head of the laboratory. He also holds a joint appointment with the Department of Electrical & 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, Workshop on Underwater Networking (WUWNet), Defence Technology Asia (DTA), IEEE International Conference on Communication Systems (ICCS) and Waterside Security (WSS) conferences and has served as reviewer for numerous international journals. He was the chairman of the student poster committee for IEEE OCEANS'06 in Singapore, and the founding chairman of the IEEE OES Singapore Autonomous Underwater Vehicle Challenge (SAUVC). He is also the IEEE technology committee co-chair of underwater communication, navigation & positioning.