# A Class of Affine Projection Filters that Exploit Sparseness under Symmetric alpha-Stable noise

Konstantinos Pelekanakis and Mandar Chitre
Acoustic Research Laboratory
Tropical Marine Science Institute
National University of Singapore
Singapore, 119223
E-mail: {costas, mandar}@arl.nus.edu.sg

***Abstract*—A new algorithmic framework for sparse underwater acoustic channel estimation under the presence of Symmetric $\alpha$-Stable (S$\alpha$S) noise is proposed. The framework is inspired by the low-complexity affine projection algorithm (APA) and employs a novel cost function that includes robust methods for impulsive noise suppression and a sparseness constraint. The constraint exploits the non-isotropic (Riemannian) structure of the channel, which leads to natural gradient adaptation. Based on this framework, new algorithms are derived with linear algorithmic complexity. The effectiveness of the new algorithms is confirmed by computer simulations.**

## I. Introduction

Symmetric $\alpha$-Stable (S$\alpha$S) distributions model many random phenomena, such as fluctuations in gravitational fields, stock prices, low-frequency atmospheric noise, and man-made noises [1]. In shallow water acoustics, Chitre *et al* [2] showed that the impulsive ambient noise due to snapping shrimp can be characterized as S$\alpha$S. It is well known that, second and higher order moments of S$\alpha$S random variables do not exist. Hence, any adaptive filter for channel identification that is based on $\mathrm{L}_2$ norm minimization will suffer poor performance.

Underwater acoustic (UWA) communication channels often exhibit sparse time-varying impulse responses [4]. Using sparse prior information, improved performance in terms of channel tracking and computational complexity is possible. Unfortunately, classical adaptive algorithms [3] (e.g., normalized least-mean-square (NLMS), affine projection algorithm (APA), and recursive least–squares (RLS)) cannot exploit such a prior information.

Studies that propose sparse-promoting adaptive filters in the presence of impulsive noise are scarce. Vega *et al* proposed a variable step-size improved proportionate NLMS (IPNLMS) algorithm [5]. Subsequently, a real-coefficient improved proportionate affine projection sign algorithm (RIP-APSA) based on the $\mathrm{L}_1$ norm of the error signal was introduced by [6]. Yamamoto *et al* [7] robustified the Adaptive Proximal Forward-Backward Splitting (APFBS) scheme by employing a Huber loss function [8]. It is worthy to note that all aforementioned algorithms deal with real-valued channels. Recently, the authors introduced a complex-valued, sparse, robust algorithm that employs the $\mathrm{L}_p$ norm ($p \in [1, 2)$) of the error signal [9].

This paper is an incremental work of [9] and proposes a new algorithmic framework that generates a class of sparse robust APAs. The proposed framework employs a cost function that incorporates robust non-linear methods for impulsive noise attenuation and a sparse-promoting constraint based on the Riemannian distance between the current and previous channel estimate. Three new natural gradient-based algorithms are derived. These algorithms also enjoy linear computational complexity, which makes them attractive for on-chip implementation. The effectiveness of the new algorithms is demonstrated by identifying an experimental sparse UWA channel in simulated S$\alpha$S noise.

### A. Notation and definitions

Superscripts $^\intercal$, $^\dagger$, and $^*$ stand for transpose, Hermitian transpose, and conjugate, respectively. Column vectors (matrices) are denoted by boldface lowercase (uppercase) letters. Let $z \in \mathbb{C}$ and $p \geq 1$. The $\mathrm{L}_p$ 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. 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 defined in [10].

## II. System model

We employ the baseband representation of the channel impulse response, transmitted/received signals, and additive noise process. Let us consider an UWA channel, which is modeled by the unknown $K$-tap vector $\mathbf{h}[n]{=}[h_0[n]\, h_1[n] \ldots h_{K-1}[n]]^\intercal$ 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 channel output signal is expressed as

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

where $\mathbf{u}[n]{=}[u[n]\ u[n-1] \ldots u[n-K+1]]^\intercal$ is the vector of the $K$ most recent input samples and $w[n]$ is a complex S$\alpha$S random variable. We assume that the passband noise stems from the family of Symmetric alpha-Stable (S$\alpha$S) distribution with characteristic function $\varphi(\omega){=}e^{-\gamma|\omega|^\alpha}$, where the characteristic exponent $\alpha \in (0, 2]$ describes the impulsiveness of the noise (smaller $\alpha$ leads to more impulsive noise) 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$\alpha$S probability density function (pdf) boils down to the Gaussian pdf and $\gamma$ is equal to half the variance. For mathematical and practical reasons, we restrict our work to the class of S$\alpha$S distributions where $\alpha \in (1, 2]$ [1]. The objective of this work is to perform recursive estimation of $\mathbf{h}[n]$ given sequential observations $\{y[i], \mathbf{u}[i]\}_{i=1}^{n}$.

## III. Linear complexity algorithmic framework

The APA stands out as an alternative to NLMS and RLS due to its ability to compromise between low complexity and memory requirements and fast convergence rate [3]. We capitalize on the APA property and propose sparse robust affine projection-type of algorithms with linear complexity.

Let $\hat{\mathbf{h}}[n]$ denote the estimate of $\mathbf{h}[n]$ and let $\mathbf{r}[n]$=$\hat{\mathbf{h}}[n] - \hat{\mathbf{h}}[n-1]$ be the channel update vector. We also express the a posterior error and prior error as $\bar{e}[i]$=$y[i] - \hat{\mathbf{h}}[n]^{\dagger}\mathbf{u}[i]$ and $e[i]$=$y[i] - \hat{\mathbf{h}}[n-1]^{\dagger}\mathbf{u}[i]$, respectively, for $i \leq n$. The new class of algorithms is derived by minimizing the following cost function:

$$J[n] = \sum_{i=n-L+1}^{n} f\left(\bar{e}[i]\right) \tag{2}$$

$$\text{subject to } \mathbf{r}[n]^{\dagger}\mathbf{P}[n-1]\mathbf{r}[n] \leq \mu^2, \tag{3}$$

where $f\left(\bar{e}\right)$ denotes a real, non-negative valued, *loss function*, whose purpose is to down weight large errors due to impulses. Specific examples of $f\left(\bar{e}\right)$ are given below. The parameter $L$ is the length of the observation window and also the projection order of the filter. There is a tradeoff on the choice of $L$; increasing $L$ makes the algorithm to converge faster provided that the impulse response remains static over that window length. $\mathbf{P}[n]$ is a $K \times K$ Hermitian, positive definite matrix whose entries depend on $\hat{\mathbf{h}}[n]$, i.e., $\mathbf{P}[n]$ is a Riemannian metric tensor. The term $\mathbf{r}[n]^{\dagger}\mathbf{P}[n-1]\mathbf{r}[n]$ denotes the Riemannian distance between $\hat{\mathbf{h}}[n]$ and $\hat{\mathbf{h}}[n-1]$. Note that the constraint $\mathbf{r}[n]^{\dagger}\mathbf{P}[n-1]\mathbf{r}[n] \leq \mu^2$ promotes sparseness since $\hat{\mathbf{h}}[n]$ is only allowed to move close to some axis of $\mathbb{C}^K$.

Finding a suitable $\mathbf{P}[n]$ is not an easy task. Fortunately, the study in [11] suggests that $\mathbf{G}[n]$=$\mathbf{P}[n]^{-1}$, where $\mathbf{G}[n]$ is the *proportionate matrix* of the PNLMS algorithm that leverages sparseness [12]. Since the IPNLMS algorithm [13] is superior over PNLMS, we choose $\mathbf{G}[n]$ to be a diagonal matrix, whose diagonal elements, $\{g_k[n]\}_{k=0}^{K-1}$, are computed as follows:

$$g_k[n] = \frac{1-\beta}{2K} + (1+\beta)\frac{\left|\hat{h}_k[n]\right|_1}{2\left\|\hat{\mathbf{h}}[n]\right\|_1 + \varepsilon}, \tag{4}$$

where $\varepsilon$ denotes a small positive constant to avoid division by zero during initialization of the algorithm. The parameter $\beta \in [-1, 1]$ controls the sparseness of the solution; sparseness is enhanced as $\beta$ approaches one while highly dispersive channels should tune $\beta$ close to -1.

Using Lagrange multipliers, the modified cost function becomes

$$J[n] = \sum_{i=n-L+1}^{n} f\left(\bar{e}[i]\right) + \delta\left(\mathbf{r}[n]^{\dagger}\mathbf{G}^{-1}[n-1]\mathbf{r}[n] - \mu^2\right) \tag{5}$$

where $\delta$ here is the Lagrange multiplier. Computing $\nabla_{\mathbf{r}[n]^*}J[n]$, we have

$$-\sum_{i=n-L+1}^{n} \psi(\bar{e}[i])\mathbf{u}[i] + \delta\mathbf{G}^{-1}[n-1]\mathbf{r}[n], \tag{6}$$

where $\psi(\bar{e})$=$\partial f(\bar{e})/\partial\bar{e}$[1] denotes the complex *score function*. The general type of the algorithm is derived by computing $\nabla_{\mathbf{r}[n]^*}J[n]$=0, which leads to natural gradient adaptation [11]. Note that it is tedious to solve for $\mathbf{r}[n]$ using (6) 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:

$$\mathbf{r}[n] = \frac{1}{\delta}\mathbf{G}[n-1]\mathbf{U}[n]\boldsymbol{\psi}[n], \tag{7}$$

where $\boldsymbol{\psi}[n]$=$[\psi(e[n]) \ldots \psi(e[n\text{-}L\text{+}1])]^{\intercal}$ and $\mathbf{U}[n]$=$[\mathbf{u}[n]\ \mathbf{u}[n-1] \ldots \mathbf{u}[n\text{-}L\text{+}1]]$ is the $K \times L$ matrix of input samples. To obtain the Lagrange multiplier $\delta$, we substitute (7) into (3). Hence, we have

$$\delta = \frac{1}{\mu}\sqrt{\|\bar{\bar{\mathbf{x}}}[n]\|_2^2}, \tag{8}$$

$$\bar{\bar{\mathbf{x}}}[n] = \mathbf{G}[n-1]^{1/2}\mathbf{U}[n]\boldsymbol{\psi}[n], \tag{9}$$

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

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mu\frac{\mathbf{G}[n-1]^{1/2}\bar{\bar{\mathbf{x}}}[n]}{\sqrt{\kappa + \|\bar{\bar{\mathbf{x}}}[n]\|_2^2}}, \tag{10}$$

where $\mu$>0 and $\kappa$ is a small positive constant used to avoid possible division by zero during initialization of the algorithm. Since $\mathbf{G}[n]$ is diagonal and $L \ll K$, it is straightforward to see that (10) requires $O(K)$ operations per datum.

### A. The NGMAPA

Inspired by the recursive least M-estimate (RLM) algorithm [14], we employ a loss function that is designed to cope with contaminated Gaussian noise, namely, the observed noise consists of two components: a Gaussian component and an impulsive interference component. To this end, we modify Hampel's three-part redescending M-estimate function so that it conforms with the chosen complex gradient operator. Dropping the time index for notational convenience, the loss function is defined as:

$$f(\bar{e}) = \begin{cases} \bar{e}\bar{e}^* & ,0 \leq |\bar{e}|_2 < \xi \\ 2\xi\,|\bar{e}|_2 - \xi^2 & ,\xi \leq |\bar{e}|_2 < \Delta \\ \xi(T+\Delta) - \xi^2 + \xi\dfrac{(|\bar{e}|_2 - T)^2}{\Delta - T} & ,\Delta < |\bar{e}|_2 < T \\ \xi(T+\Delta) - \xi^2 & ,T < |\bar{e}|_2 \end{cases}, \tag{11}$$

[1] $f(z)$ may not be analytic [10]

where the threshold parameters $\xi, \Delta, T$ are continuously estimated based on a desired confidence degree for outlier suppression. The score function is computed as

$$\psi(\bar{e}) = \begin{cases} \bar{e}^* & ,0 \leq |\bar{e}|_2 < \xi \\ \frac{\xi}{|\bar{e}|_2}\bar{e}^* & ,\xi \leq |\bar{e}|_2 < \Delta \\ \xi\frac{|\bar{e}|_2 - T}{\Delta - T}\frac{\bar{e}^*}{|\bar{e}|_2} & ,\Delta < |\bar{e}|_2 < T \\ 0 & ,T < |\bar{e}|_2 \end{cases} . \quad (12)$$

The confidence degree for outlier suppression is computed based on the power of the background Gaussian noise as follows: the in-phase and quadrature power components of the baseband Gaussian noise, denoted as $\sigma_r^2(n)$ and $\sigma_i^2(n)$, respectively, are estimated by using a median operator of the $N_w \times 1$ prior error signal $[e[n] \ldots e[n-N_w+1]]^\intercal$ [14]. We assume the Rayleigh distribution for $|e[n]|_2$ with parameter $\sigma^2(n)$=$0.5(\sigma_r^2(n) + \sigma_i^2(n))$. 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 given by equations (12) and (10) will be called NGMAPA (natural gradient-based M-estimate affine projection algorithm) hereafter.

### *B. The NGAPSA*

As discussed above, the parameters $\xi, \Delta, T$ of the NGMAPA are based on the steady-state error signal. Hence, the algorithm performance may be hampered during convergence if rapid channel fluctuations occur. Here, we propose a new algorithm that does not depend on any threshold parameters. This is achieved by employing $f(\bar{e}[i])$=$|\bar{e}[i]|_1$ into (2). Then, the score function is found as $\psi(\bar{e}[i])$=$0.5(csgn(\bar{e}[i]))^*$ and the algorithm described by equation (10) will be called the natural gradient-based affine projection sign algorithm (NGAPSA) hereafter. Note that

- if $L$=1, then the NGAPSA reduces to the NGSA (natural gradient sign algorithm).
- if $\beta$=-1, then the NGAPSA reduces to the complex form of the APSA [15].

### *C. The NGpNAPA*

For signal processing in S$\alpha$S noise environments, the mean square error (MSE) is not a valid optimality criterion since S$\alpha$S distributions lack moments of order $p \geq \alpha$. However, all moments of order $p<\alpha$ do exist and so the minimum dispersion error is mathematically meaningful as an optimality criterion. Inspired by the structure of (11), we propose a loss function that behaves like least-squares in the absence of impulses and uses the $L_p$ norm, $p \in [1, \alpha)$, in the presence of impulses. This loss function takes the form

$$f(\bar{e}) \triangleq \begin{cases} \bar{e}\bar{e}^* & ,0 \leq |\bar{e}|_2 < \xi \\ |\bar{e}|_p^p & ,\xi \leq |\bar{e}|_2 < \Delta \\ |\Delta|^p & ,\Delta \leq |\bar{e}|_2 \end{cases} , \quad (13)$$

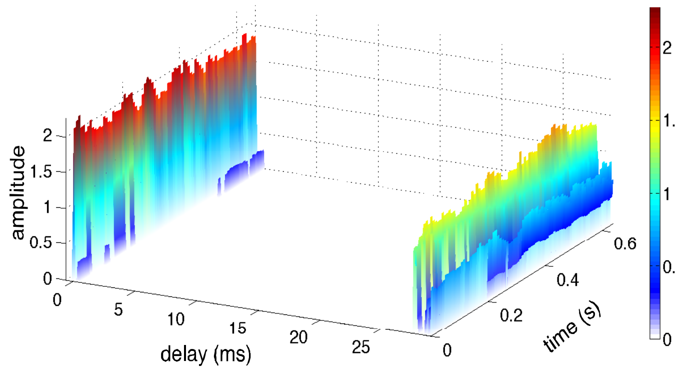


Fig. 1. The time-varying channel used in simulations. The x-axis shows multipath delay, the y-axis shows absolute time and the z-axis shows the channel amplitude in linear scale. The channel length is $K$=371 taps and each channel snapshot is kept fixed for 50 symbol intervals (8 ms).

where the threshold parameters $\xi$, $\Delta$ are proportional to the dispersion $\gamma$ of the observed passband S$\alpha$S noise. The score function is computed as

$$\psi(\bar{e}) = \begin{cases} \bar{e}^* & ,0 \leq |\bar{e}|_2 < \xi \\ \frac{p}{2}[|\text{Re}\{\bar{e}\}|^{p-1} sgn(\text{Re}\{\bar{e}\}) \\ -j\ |\text{Im}\{\bar{e}\}|^{p-1} sgn(\text{Im}\{\bar{e}\})] & ,\xi \leq |\bar{e}|_2 < \Delta \\ 0 & ,\Delta \leq |\bar{e}|_2 \end{cases} \quad (14)$$

and the algorithm described by equations (14) and (10) will be called NGpNAPA (natural gradient p-norm affine projection algorithm) hereafter.

## IV. Simulation Results

The effectiveness of all proposed algorithms, namely, NGMAPA, NGpNAPA, NGAPSA and NGSA is tested by comparing them with APSA [15] and RLS [3]. The sparse, time-varying channel to be identified was measured during the Focused Acoustic Fields (FAF) experiment off the coast of Pianosa Island, Italy in 2005. Fig. 1 illustrates the amplitude evolution of the FAF channel. The simulated input signal is a 6250 symbols/ s-rate Q-PSK modulated PN-sequence sampled at 1 sample/symbol. S$\alpha$S noise is generated in passband using [16], then shifted to baseband using the lowpass filter of the FAF experiment and added to the channel output. The following passband SNR definition is employed

$$E_s/N_0 \text{ (dB)} \triangleq 10\log_{10}\frac{N_s P_s}{2\gamma^{2/\alpha}}, \quad (15)$$

where $N_s$ is the ratio of 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 $\alpha$=2, (15) becomes the usual $E_s/N_0$ definition in Gaussian noise. Four types of noise signals are considered, i.e., 1) high rate of impulses in low

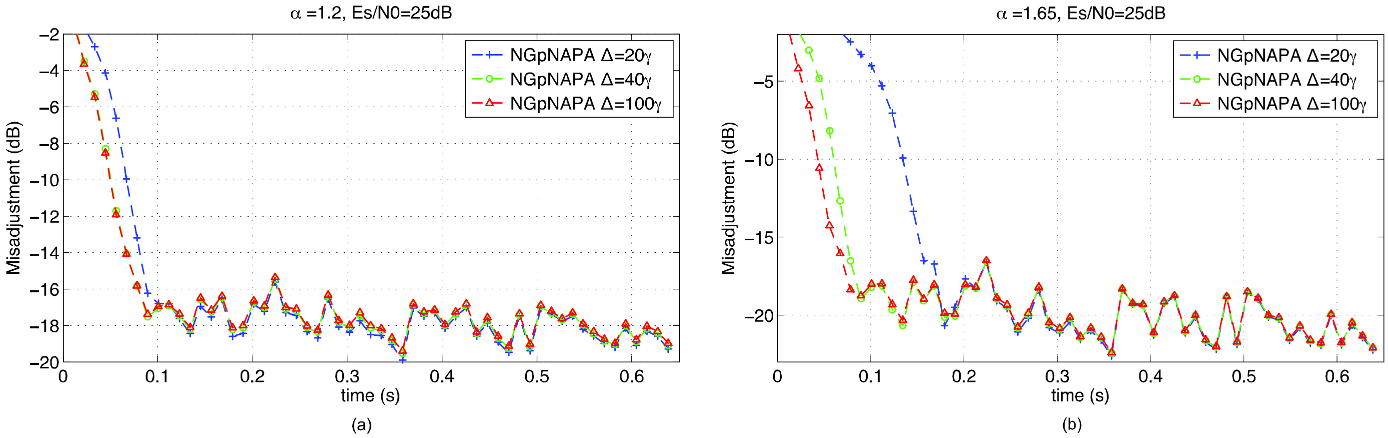


Fig. 2. Learning curves of NGpNAPA for different $\Delta$, $\alpha$, and Es/N$_0$

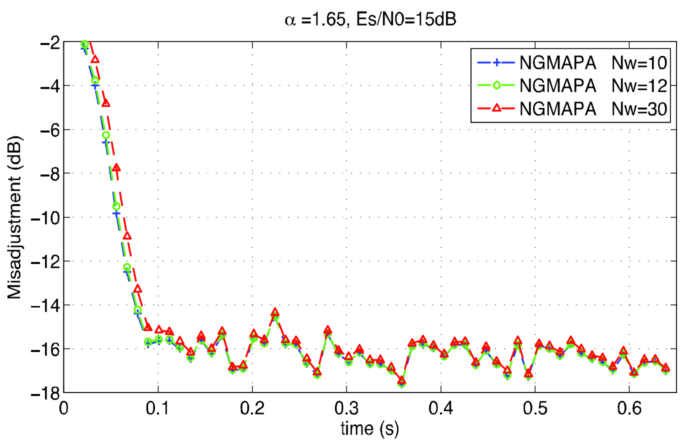


Fig. 3. Learning curves of NGMAPA for different N$_w$.

SNR ($\alpha$ =1.2, $E_s/N_0$=$15dB$), 2) high rate of impulses in high SNR ($\alpha$ =1.2, $E_s/N_0$=$25dB$), 3) low rate of impulses in low SNR ($\alpha$ =1.65, $E_s/N_0$=$15dB$), and 4) low rate of impulses in high SNR ($\alpha$ =1.65, $E_s/N_0$=$25dB$). 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. Unless otherwise specified, the algorithm parameters are chosen as:

- $\mu$=0.25, $\beta$=0.5 for NGMAPA, NGpNAPA, NGAPSA, and NGSA
- $L$=4 for NGMAPA, NGpNAPA, NGAPSA, and APSA,
- $p$=$\alpha$-0.15, $\xi$=$\gamma$ for NGpNAPA,
- $\lambda$=0.995 for RLS.

Fig. 2 tests the misadjustment sensitivity to parameter $\Delta$ for the NGpNAPA. Clearly, when $\Delta$=100$\gamma$ the algorithm shows the fastest convergence rate. This is plausible because during convergence the error signal is dominated by large channel estimation errors while measurement noise is less significant. Note that the steady-state performance is insensitive to our choices of $\Delta$. Insensitivity to $\Delta$ is also found in both convergence rate and steady-state for the other combinations of $\alpha$ and $E_s/N_0$ (results are omitted for brevity). This behavior is valid since the error signal is dominated by measurement noise and so $\Delta$=20$\gamma$ is a good choice.

Fig. 3 tests the misadjustment sensitivity to parameter $N_w$ for the NGMAPA. We see that $N_w$=10 makes algorithm convergence slightly faster. The same result occurs for the other pairs of $\alpha$ and $E_s/N_0$ (results are omitted for brevity). This happens because the method that computes the background Gaussian noise power assumes a steady-state error signal. A large $N_w$ increases the memory of the algorithm to past error signal values, which is not efficient during the convergence period since these values are affected by large channel estimation errors.

Figures 4(a) through 4(d) present a comparison of the misadjustment of all algorithms (note $\Delta$=100$\gamma$ for NGpNAPA and $N_w$=12 for NGMAPA) for different values of $\alpha$ and $E_s/N_0$. The results show that the NGpNAPA exhibits the fastest convergence rate. Among the sparse algorithms, the NGSA shows the slowest convergence speed, however, the algorithm exhibits the best tracking for $E_s/N_0$=$15dB$ regardless the choice of $\alpha$. This result is justified by noting that the error term in the cost function is smaller (recall $L$=1) and so NGSA becomes more robust against impulses. The sparseness effect of all proposed algorithms is confirmed since the non-sparse (but robust) APSA exhibits consistently poor performance. The robustness of the proposed algorithms is further validated by observing that performance of the non-robust RLS. Clearly, the RLS algorithm consistently fails except when $\alpha$=1.65 and $E_s/N_0$=$25dB$.

## V. Conclusion

A novel APA-type framework for sparse UWA channel identification in the presence of S$\alpha$S noise was proposed. The framework was based on a cost function that employed robust methods for outlier rejection and a sparse-aware Riemannian distance that modified the gradient search direction leading to natural gradient (NG) adaptation. Three $O(K)$ algorithms were generated, namely, the NGMAPA, the NGpNAPA and the

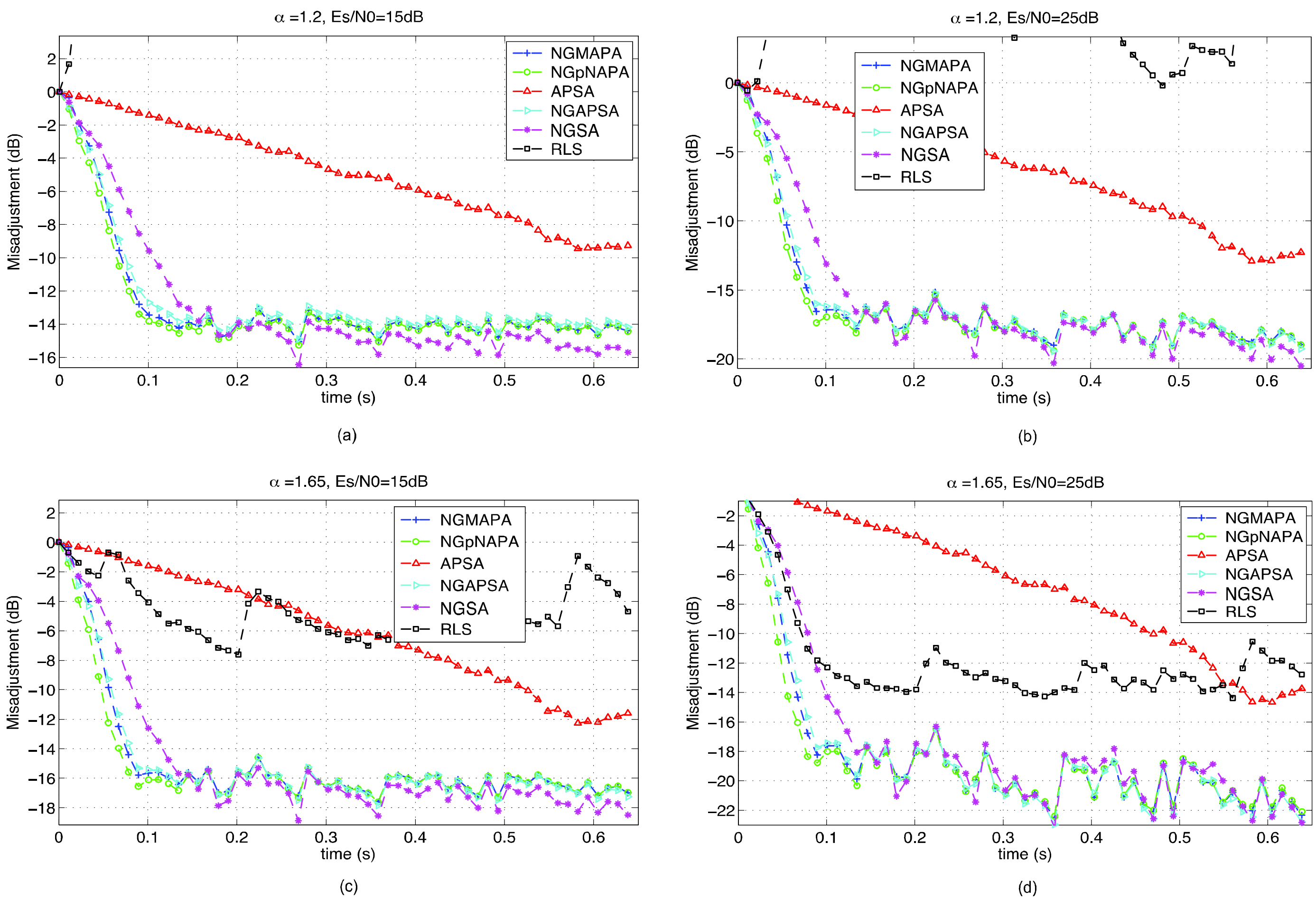


Fig. 4. Learning curves of all algorithms for different $\alpha$ and Es/$N_0$

NGAPSA. The NGMAPA was based on the Hampel's three-part redescending M-estimate function, the NGpNAPA was based on a mixture of $L_2$ and $L_p$ norms and NGAPSA was based on the $L_1$ norm. The effectiveness of these algorithms was confirmed by estimating an experimental sparse acoustic link in low/high SNR and slow/fast rate of impulses. In addition, the NGpNAPA showed the fastest convergence rate corroborating the efficiency of the $L_p$ norm.

## References

[1] 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.

[2] 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.

[3] S. Haykin, *Adaptive Filter Theory*, 4th Ed., Prentice-Hall, Englewood Cliffs, NJ, 2002.

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

[5] 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, May 2009.

[6] 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, Nov. 2011.

[7] T. Yamamoto, M. Yamagishi, and L. Yamada, "Adaptive proximal forward-backward splitting for sparse system identification under impulsive noise," *Proc. EUSIPCO*, pp.2620-2624, Aug. 2012.

[8] P. J. Huber, *Robust Statistics* John, John Wiley & Sons, New York, 1981.

[9] K. Pelekanakis, H. Liu, and M. Chitre, "An algorithm for sparse underwater acoustic channel identification under Symmetric $\alpha$-Stable noise," *in Proc. IEEE OCEANS/MTS*, pp.1-6, Santander, Spain, June 2011.

[10] 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.

[11] S. L. Gay and S. C. Douglas, "Normalized natural gradient adaptive filtering for sparse and nonsparse systems," *Proc. IEEE ICASSP*, vol. II, pp. 1405–1408, 2002.

[12] 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.

[13] J. Benesty and S.L. Gay, "An improved PNLMS algorithm", *IEEE ICASSP*, vol.2, pp. 1881-1884, May 2002.

[14] 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. Sig. Proc.*, vol. 52, no. 4, pp. 975-991, April 2004.

[15] 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.

[16] J.P. Nolan. STABLE program for Windows [Online]. Available: http://academic2.american.edu/~jpnolan/stable/stable.html.