Sparsity-Aware Affine Projection Adaptive Algorithms for System Identification

Ran Meng ‡, Rodrigo C. de Lamare ‡ and Vítor H. Nascimento §

‡ *Communications Research Group,*
Department of Electronics, University of York, UK, YO10 5DD
Department of Electronic Systems Engineering,
§ *University of São Paulo, São Paulo, SP, 05508-900 Brazil*
{*rm770,rcdl500*}*@ohm.york.ac.uk, vitor@lps.usp.br*

***Abstract*—In this paper, we propose adaptive algorithms for system identification of sparse systems. We introduce a L_1-norm penalty to improve the performance of affine projection algorithms. This strategy results in two new algorithms, the zero-attracting APA (ZA-APA) and the reweighted zero-attracting AP (RZA-APA). The ZA-APA is derived via the combination of a L_1-norm penalty on the coefficients into a standard APA cost function, which generates a zero attractor in the update function. The zero attractor promotes sparsity in the filter coefficients during the update process, and therefore accelerates convergence when identifying sparse systems. We show that the ZA-APA can achieve a lower mean square error than the standard LMS and AP algorithms. To further improve the performance, the RZA-APA is developed using a reweighted zero attractor. The performance of the RZA-APA is superior to that of the ZA-APA numerically. Simulation results demonstrate the advantages of the proposed adaptive algorithms in both convergence rate and steady-state behavior under sparsity assumptions on the true coefficient vector. The RZA-APA is also shown to be robust when the number of non-zero taps increases.**

I. Introduction

In general, the problem of system identification involves constructing an estimate of an unknown system given only two signals, namely 1) an input signal and 2) a reference signal. Typically, the unknown system is modeled with a finite impulse response (FIR) filter, and adaptive filtering algorithms are employed to compute an estimate of the response of the unknown system being identified. If the system is time-varying, then the problem becomes more involved and includes the task of tracking the unknown system as it changes over time. When there is no available reference signal,we call the problem blind system identification [15], [16]. The system identification problem has numerous applications in control systems, digital communications, and signal processing. [4]

The adaptive filters can be based on various basic algorithms, of which the two most known are the least mean squares (LMS) and the recursive least squares (RLS) [1]. The LMS algorithm is an extremely simple technique from a computational complexity point of view, however, it may have a poor performance with colored signals. The RLS algorithm has often a high performance, however, it is often too complex to implement in real time. This is one of the reasons why designers seek solutions with an improved performance as compared with the LMS and with a significantly lower complexity than the RLS for applications with large filters [6]. As the required adaptive filter lengths grow, the conventional LMS algorithm exhibits a slower convergence rate. For the identification of sparse systems, the poor performance can be explained by observing two aspects: (a) slow convergence of the filter taps to their steady-state values since the convergence rate of the algorithm is proportional to the total filter length; (b) high steady-state misadjustment due to the estimation noise that inevitably occurs during the adaptation of the so-called inactive filter taps (i.e., taps with zero or close to zero values at steady state) [7]. The affine projection algorithm (APA) [1] and its variations [8]–[10], [14] is a popular method in adaptive filtering applications, with complexity and performance intermediary between those of LMS and of RLS. Its applications include echo cancellation, channel equalization, interference cancellation, and so forth.

In many scenarios, impulse responses of unknown systems can be assumed to be sparse, containing only a few large coefficients interspersed among many negligible ones. Using such sparse prior information can improve the filtering/estimation performance. However, standard LMS filters do not exploit such information. In the past years, many algorithms exploiting sparsity were based on applying a subset selection scheme during the filtering process, which was implemented via statistical detection of active taps or sequential partial updating [2]–[4]. Other variants assign proportional step sizes to different taps according to their magnitudes, such as the proportionate normalized LMS (PNLMS) and its variations [5].

Motivated by recent progress in compressive sensing, several authors have considered using the L_1-norm penalty to exploit the sparsity of the sparse systems or signals [2], [3], [11], [12]. The basic idea is to introduce a penalty that favors sparsity in the cost function. In this paper we propose an alternative approach to identifying sparse systems using an APA. We first incorporate an L_1-norm penalty on the coefficients into the quadratic cost function of the standard APA. We incorporate an L_1-norm penalty on the coefficients into the cost function of the standard APA. This results in a modified APA update with a zero attractor for all the taps, which we term ZA-APA. We analytically demonstrate that the ZA-APA achieves better steady-state performance than that of the standard APA for sparse models. To further improve the filtering performance, the reweighted zero-attracting APA (RZA-APA) is also proposed, which employs reweighted step sizes of the zero attractor for different taps, inducing the attractor to selectively promote zero taps rather than uniformly promote zeros on all the taps. Simulation results illustrate that the proposed algorithms outperform the standard APA in both convergence rate and steady-state performance for sparse systems; and the RZA-APA outperforms the ZA-APA numerically. Furthermore,the RZA-APA shows robustness when the number of non-zero taps increases, with little loss in performance with respect to the standard APA in non-sparse situations.

The paper is organized as follows. Section II briefly describes the sparse system identification problem. Section III reviews and develops the ZA-APA and RZA-APA adaptive algorithms for sparse systems. In Section IV, simulation results are provided. Finally, we conclude the paper and discuss possible future directions in Section V.

II. Sparse System Identification Problem Statement

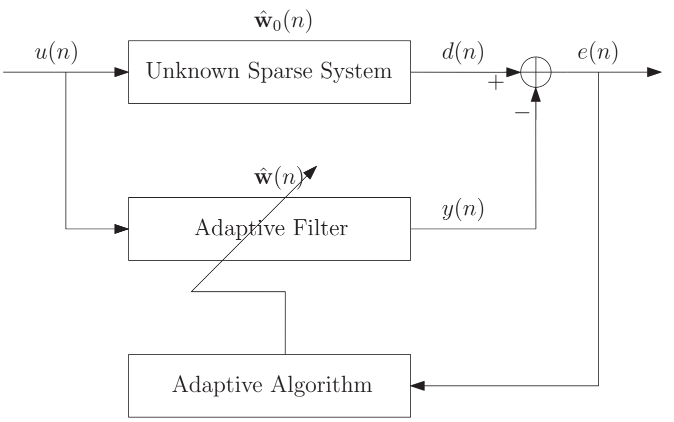

Fig. 1. Block diagram of sparse system identification using an adaptive algorithm.

In a sparse system identification application, the desired signal is the output of an unknown sparse system when excited by an input signal. The input signal is also used as an input for an adaptive filter $\hat{\mathbf{w}}(n)$ with M coefficients to produce an output estimate $y(n)$ which is compared to the reference signal $d(n)$. The error signal $e(n)$ consists of the difference between the desired signal $d(n)$ and the output of the sparse adaptive filter $y(n)$. When the output error $e(n)$ is minimized, the adaptive filter represents a model for the unknown sparse system. The block diagram of the system identification scheme is shown in Fig. 1. Here, $\mathbf{u}(n)$ is the input signal with M samples that is applied to the unknown sparse system, and the response signal $d(n)$ is the reference signal. The problem we are interested in solving is how to identify the unknown sparse system using an adaptive algorithm that is able to identify and exploit the sparse nature of the system.

III. Adaptive Algorithms

In this section, we review the standard APA and derive the proposed ZA-APA and RZA-APA.

A. *Review of the Standard Affine Projection Algorithm*

Let us assume that the last N input signal vectors are organized in a matrix as follows [1]:

$$\mathbf{U}(n) = [\mathbf{u}^T(n), \mathbf{u}^T(n-1), ..., \mathbf{u}^T(n-N+1)], \quad (1)$$

where the $\boldsymbol{u}(n)$ denotes the vector of the input signal at time n, and N denotes the APA order. We can also define some vectors representing the filter output $\mathbf{y}(n)$, the desired signal $\mathbf{d}(n)$ and the error vectors $\mathbf{v}(n)$. These vectors are, respectively given by

$$\mathbf{y}(n) = [y(n), y(n-1), ..., y(n-N+1)]^T, \quad (2)$$

$$\mathbf{d}(n) = [d(n), d(n-1), ..., d(n-N+1)]^T, \quad (3)$$

$$\mathbf{v}(n) = [v(n), v(n-1), ..., v(n-N+1)]^T. \quad (4)$$

From (1)-(4), we can obtain the following equation

$$\mathbf{y}(n) = \hat{\mathbf{w}}^T\mathbf{U}(n) + \mathbf{v}(n) \quad (5)$$

For the APA, the tap-weight vector variation is defined as $\Delta\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n+1) - \hat{\mathbf{w}}(n)$. The objective of the APA is to minimize

$$\begin{aligned} &\|\hat{\mathbf{w}}(n+1) - \hat{\mathbf{w}}(n)\|^2 \\ \text{subject to} \quad &\mathbf{d}(n) - \mathbf{U}(n)\hat{\mathbf{w}}(n+1) = 0 \end{aligned} \quad (6)$$

Here the Lagrange multiplier can be used to find out the solution that minimizes the cost function $J(n)$.

$$\begin{aligned} J(n) = &\|\hat{\mathbf{w}}(n+1) - \hat{\mathbf{w}}(n)\|^2 \\ &+ Re\{[\mathbf{d}(n) - \mathbf{U}(n)\hat{\mathbf{w}}(n+1)]^{\mathrm{H}}\boldsymbol{\lambda}\} \end{aligned} \quad (7)$$

where $\boldsymbol{\lambda} = [\lambda(0), \lambda(1), ..., \lambda(N-1)]^T$ denotes the vector of Lagrange multipliers. By using the method of Lagrange multipliers, the solution of this optimization problem would be the following filter coefficient update equation

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu\,\mathbf{U}^H(n)[\mathbf{U}(n)\mathbf{U}^H(n)]^{-1}\mathbf{e}(n) \quad (8)$$

with $\mu = 1$. In general, a step-size $\mu < 1$ is used to control where μ is the step-size controlling convergence and the steady-state behavior of the APA.

The APA is a generalization of the normalized least mean square (NLMS) adaptive filtering algorithm. When the AP order N is set to one, relations (8) will reduce to the familiar NLMS algorithm.

B. Zero-Attracting Affine Projection Algorithm (ZA-APA)

In the ZA-APA, a new cost function $J_1(n)$ is defined by combining the instantaneous square error with the L_1-norm penalty of the coefficient vector as given by

$$\begin{aligned} J_1(n) &= \|\hat{\mathbf{w}}(n+1) - \hat{\mathbf{w}}(n)\|^2 \\ &+ Re\{[\mathbf{d}(n) - \mathbf{U}(n)\hat{\mathbf{w}}(n+1)]^H \boldsymbol{\lambda}\} \\ &+ \|\hat{\mathbf{w}}(n+1)\|_1 \end{aligned} \tag{9}$$

To minimize the cost function, we can use the method of Lagrange multipliers,

$$\frac{\partial J}{\partial \hat{\mathbf{w}}^*(n+1)} = \hat{\mathbf{w}}(n+1) - \hat{\mathbf{w}}(n) - \mathbf{U}^H(n)\boldsymbol{\lambda} + \alpha\, \mathrm{sgn}[\hat{\mathbf{w}}(n+1)] \tag{10}$$

where sgn[ ] is a function that returns the sign of the arguments. By equating (10) to zero, then we can get

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mathbf{U}^H(n)\boldsymbol{\lambda} - \alpha\, \mathrm{sgn}[\hat{\mathbf{w}}(n+1)] \tag{11}$$

Multiplying both sides by $\mathbf{U}(n)$, we can obtain

$$\begin{aligned} \mathbf{d}(n) &= \mathbf{U}(n)\hat{\mathbf{w}}(n) + \mathbf{U}(n)\mathbf{U}^H(n)\boldsymbol{\lambda} \\ &\quad -\alpha\, \mathbf{U}(n)\mathrm{sgn}[\hat{\mathbf{w}}(n+1)] \end{aligned} \tag{12}$$

Because $\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{U}(n)\hat{\mathbf{w}}(n)$ we can solve for $\boldsymbol{\lambda}$. Assuming that $\mathrm{sgn}[\hat{\mathbf{w}}(n+1)] \approx \mathrm{sgn}[\hat{\mathbf{w}}(n)]$, we can obtain with further manipulations the new filter coefficient update equation

$$\begin{aligned} \hat{\mathbf{w}}(n+1) &= \hat{\mathbf{w}}(n) + \mu\, \mathbf{U}^+(n)\mathbf{e}(n) \\ &+ \alpha\, \mathbf{U}^+(n)\mathbf{U}(n)\mathrm{sgn}[\hat{\mathbf{w}}(n)] \\ &- \alpha \mathrm{sgn}[\hat{\mathbf{w}}(n)] \end{aligned} \tag{13}$$

where $\mathbf{U}^+(n) = \mathbf{U}^H(n)[\mathbf{U}(n)\mathbf{U}^H(n)]^{-1}$.

Comparing the ZA-APA update (13) to the standard APA update (8), the ZA-APA has two additional terms, which attract the tap coefficients to zero whenever the magnitudes of the weight vector are close to zero. We call this the zero attractor feature [2], whose strength is controlled by α. Intuitively, the zero attractor will speed-up convergence when the majority of coefficients of $\hat{\mathbf{w}}$ are zero, i.e., the system is sparse.

In addition, if we set the AP order N to one, (13) could also be considered as the update formula for Zero-Attracting NLMS (ZA-NLMS) algorithm.

C. Reweighted Zero-Attracting Affine Projection Algorithm (RZA-APA)

Unfortunately, the ZA-APA does not distinguish between zero taps and non-zero taps. Since all the taps are forced to zero uniformly, the performance of ZA-APA would be deteriorated when applied to less sparse systems. In order to solve this problem, we adopt a heuristic approach first reported in [13] and employed in [2] to reinforce the zero attractor called the reweighted zero-attracting affine projection algorithm (RZA-APA). For the RZA-APA, we use a new L_1-norm penalty to minimize the cost function,

$$\begin{aligned} \frac{\partial J}{\partial \hat{\mathbf{w}}^*(n+1)} &= \hat{\mathbf{w}}(n+1) - \hat{\mathbf{w}}(n) \\ &\quad - \mathbf{U}^H(n)\boldsymbol{\lambda} + \alpha\, \frac{\mathrm{sgn}(\hat{\mathbf{w}}(n+1))}{1+\varepsilon|\hat{\mathbf{w}}(n+1)|} \end{aligned} \tag{14}$$

Assuming that $\frac{\mathrm{sgn}(\hat{\mathbf{w}}(n+1))}{1+\varepsilon|\hat{\mathbf{w}}(n+1)|} \approx \frac{\mathrm{sgn}[\hat{\mathbf{w}}(n)]}{1+\varepsilon|\hat{\mathbf{w}}(n)|}$, and then we use again the method of Lagrange multipliers to find the solution

$$\begin{aligned} \hat{\mathbf{w}}(n+1) &= \hat{\mathbf{w}}(n) + \mu\, \mathbf{U}^+(n)\mathbf{e}(n) \\ &\quad + \alpha\, \mathbf{U}^+(n)\mathbf{U}(n)S(n) - \alpha S(n) \end{aligned} \tag{15}$$

where

$$\mathbf{U}^+(n) = \mathbf{U}^H(n)[\mathbf{U}(n)\mathbf{U}^H(n)]^{-1}$$
$$S(n) = \frac{\mathrm{sgn}[\hat{\mathbf{w}}(n)]}{1+\varepsilon|\hat{\mathbf{w}}(n)|}$$

where ε is the shrinkage magnitude.

The RZA-APA is more sensitive to taps with small magnitudes. The reweighted zero attractor takes more shrinkage exerted on those taps for which magnitudes are comparable to $1/\varepsilon$; and take less effort on the taps whose $|\hat{\mathbf{w}}(n)| \gg 1/\varepsilon$. In this way, the bias of the RZA-APA can be reduced.

The same as ZA-APA, if we set the AP order $N = 1$, we can also get the update formula of RZA-NLMS algorithm from (15).

D. Computational Complexity Analysis

In this section, we discuss the computational complexity of the existing and proposed algorithms. We assume there are only Q non-zero taps in an M-length sparse system, and the order of the APA is N. For data without time-shifting structure, we detail the computational complexity of the standard APA [1], and the proposed ZA-APA and RZA-APA as shown in Table I. The computational complexity requirements are described in terms of the number of complex arithmetic operations, namely, additions and multiplications. From the table, we note that the complexity of our proposed algorithm is a little higher than that of the standard APA. In addition, we can see that the RZA-APA has only Q more additions and Q more multiplications than that of ZA-APA.

TABLE I
COMPUTE COMPLEXITY

Algorithm	Additions	Multiplications
APA	$N^2M + MN$ $+Q - M + O(N^3)$	$N^2M + N^2 + MN$ $+N + O(N^3)$
ZA-APA	$N^2M + N^2$ $+3MN - 2N - 2M$ $+3Q + O(N^3)$	$N^2M + 2N^2$ $+3MN + 2N$ $+Q + O(N^3)$
RZA-APA	$N^2M + N^2$ $+3MN - 2N - 2M$ $+4Q + O(N^3)$	$N^2M + 2N^2$ $+3MN + N + M$ $+2Q + O(N^3)$

IV. SIMULATION RESULTS

In this section, the performance of the ZA-APA (13) and the RZA-APA (15) are compared with that of the standard NLMS, RZA-NLMS and standard APA. Four experiments have been designed to demonstrate their tracking and steady-state performance. The parameters of the simulation system are shown in Table II

TABLE II
TABLE OF PARAMETERS

Exp.	Algorithm	Step-size μ	ZA α	Reweighted ε
1	AP families	1.3	6×10^{-3}	100
	NLMS families	1.3	2×10^{-3}	100
2	AP families	1.3	6×10^{-3}	100
	NLMS families	1.3	2×10^{-3}	100
3	AP families	1.3	6×10^{-3}	100
	NLMS families	1.3	2×10^{-3}	100
4	AP families	1	2×10^{-4}	100
	NLMS families	1	2×10^{-4}	100

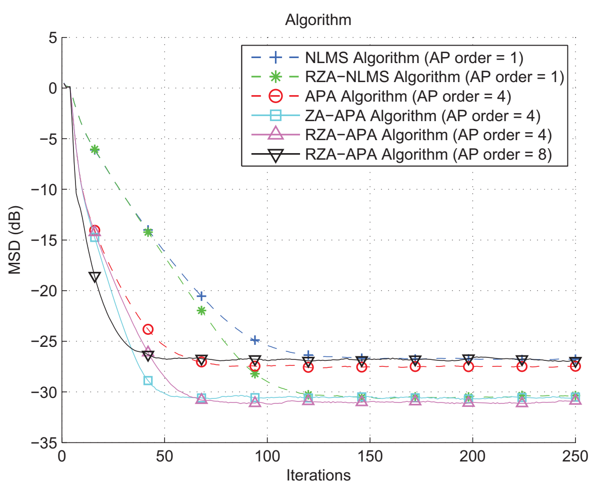

Fig. 2. Simulated MSD for experiment 1

In the first three experiments, there are 16 coefficients in the time-varying system. Note that the number of non-zero taps for each experiment are 1, 8 and 16. In the 1st experiment, we set the 4th tap with value 1 and the others to zero, so that

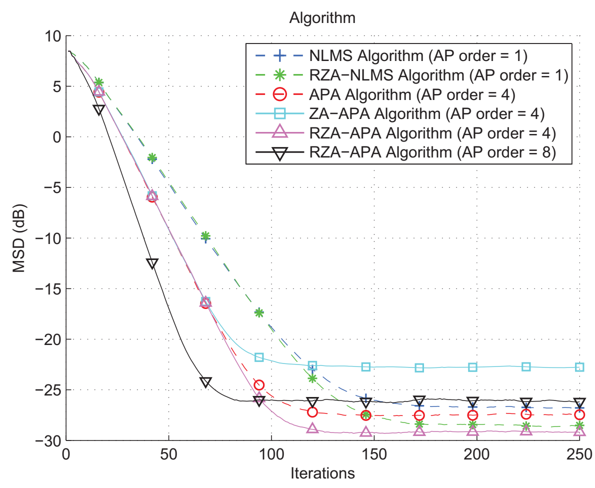

Fig. 3. Simulated MSD for experiment 2

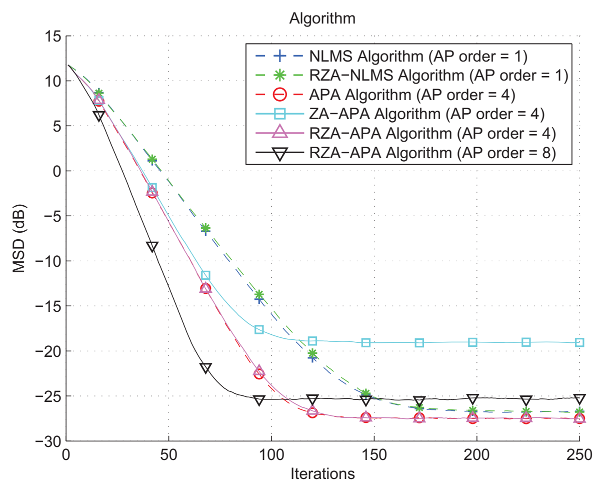

Fig. 4. Simulated MSD for experiment 3

it is a really sparse system with $1/16$ factor of sparsity. In the 2nd experiment, all the odd taps are set to 1, while all the even taps remain equal to zero. For the 3rd experiment, all the taps are set to 1, which means it is a totally non-sparse system. In these three experiments, we choose step-size $\mu = 1.3$ to achieve balance between convergence rate and steady-state performance.

For the first experiment, the average estimate of mean square deviation (MSD) is shown in Fig. 2. As we can see from the MSD results, when the system is very sparse, both the ZA-APA and the RZA-APA achieve faster convergence rate and better steady-state performance than the standard NLMS and APA. And their convergence rate are also much faster than RZA-NLMS. We can also see that if the AP order N increase to 8, the misadjustment become larger, but it can help to get a faster convergence rate. As we can see from Fig. 3, when the number of non-zero taps increases to

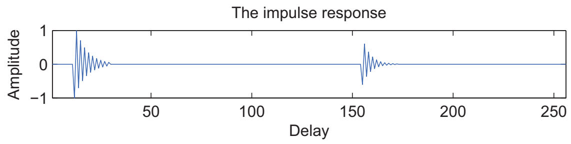

Fig. 5. The impulse response of the system in the fourth experiment.

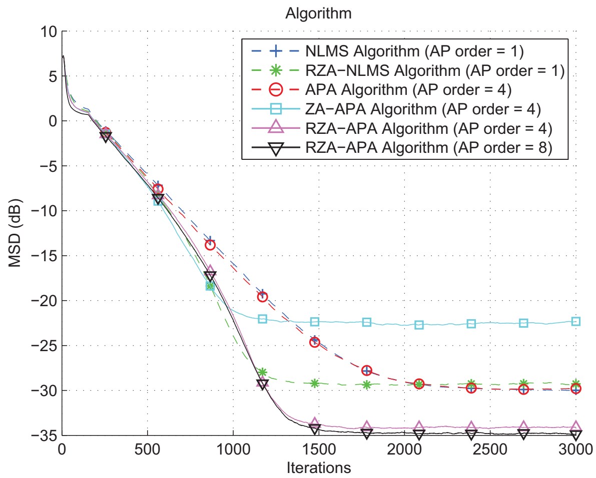

Fig. 6. Simulated MSD for experiment 4

8, the ZA-APA's performance deteriorates, while the RZA-APA still has the best convergence rate. And it also has the best steady-state performance among the three APA-based algorithms. In addtion, the increase of AP order can achieve a better convergence rate, but it will also lead to a larger misadjustment. When the sparsity continues to increase, as what we can see from the Fig. 4, RZA-APA shows a robust performance while the system is non-sparse.

In the fourth experiments, we introduce a 256-tap system with 32 nonzero coefficients. The impulse response of the system is shown in Fig. 5. Because this is a long system, it will be much slower in convergence rate, so we choose step-size $\mu = 1$ to achieve balance between steady-state performance and convergence rate. We can get that for this long sparse system, the RZA-APA significantly outperforms the standard NLMS and APA, as measured by a faster convergence rate and a lower steady-state MSD.

V. Conclusions

In this paper, two affine projection algorithms have been proposed for sparse system identification. The ZA-APA introduces an L_1-norm penalty of the coefficients into its cost function, which can help to reduce the shrinkage in the update formula. It can achieve a higher convergence rate when the majority of coefficients are zero. And the RZA-APA has been proposed to further improve the performance of the ZA-APA algorithm, where a reweighted zero attractor is devised to perform selective coefficient shrinkage. With the help of the reweighted zero attractor, RZA-APA are better both in convergence rate and steady-state performance. The simulations have also shown that the ZA-APA the RZA-APA improve on the standard APA in both tracking and steady-state performance when the system is sparse, and the RZA-APA also shows a robust performance for non-sparse systems. Our future work will include how to choose the parameters of zero-attracting algorithms in a more systematic way and how to obtain analytical expressions to predict the level of MSD.

References

[1] S. Haykin, *Adaptive filter theory*, 4th ed.,Upper Saddle River, N.J: Prentice Hall, 2002.

[2] Y. Chen and Y. Gu and A. Hero, "Sparse LMS for system identification," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*, pp. 3125-3128, Apr. 2009.

[3] Y. Chen and Y. Gu and A. Hero, "Regularized least-mean-square algorithms," 2009[Online]. Available: http://arxiv.org/abs/1012.5066

[4] J.D. Gordy and R.A. Goubran, "Fast system identification using affine projection and a critically sampled subband adaptive filter," *IEEE Trans. on Instrumentation and Measurement Society*, vol. 55, pp. 1242 - 1249, Jul. 2006

[5] Z. Yang and R. Zheng and S. Grant, "Proportionate Affine Projection Sign Algorithms for Network Echo Cancellation," *IEEE Trans. on Audio, Speech, and Language Processing*, 2011[Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5728849

[6] G. Rombouts and M. Moonen, "A sparse block exact affine projection algorithm," *IEEE Trans. on Speech and Audio Processing*, vol. 10, pp. 100-108, Jul. 2002

[7] K. Pelekanakis and M. Chitre, "Comparison of sparse adaptive filters for underwater acoustic channel equalization/Estimation", in *Proc. IEEE Int. Conf. on Communication Systems*, 2010, pp. 395 - 399

[8] S. L. Gay, and S. Tavathia, "The fast affine projection algorithm," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.5, no., pp.3023-3026 vol.5, 9-12 May 1995.

[9] R. C. de Lamare and R. Sampaio-Neto, "Adaptive Reduced-Rank MMSE Filtering with Interpolated FIR Filters and Adaptive Interpolators", *IEEE Sig. Proc. Letters*, vol. 12, no. 3, March, 2005.

[10] R. C. de Lamare and P. S. R. Diniz, "Set-membership adaptive algorithms based on time-varying error bounds for CDMA interference suppression," *IEEE Trans. Vehicular Technology*, vol. 58, pp. 644-654, Feb. 2009.

[11] D. Angelosante and G. B. Giannakis, "RLS-weighted Lasso for adaptive estimation of sparse signals," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*, pp. 3245-3248, 2009.

[12] Z. Yang, R. C. de Lamare and X. Li, "L_1 regularized STAP algorithm with a generalized sidelobe canceler architecture for airborne radar," in *Proc. IEEE Workshop on Statistical Signal Processing,* Nice, France, June 2011.

[13] E.J. Candes, M.B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877-905, 2008.

[14] R. C. de Lamare and R. Sampaio-Neto, "Adaptive Reduced-Rank Processing Based on Joint and Iterative Interpolation, Decimation, and Filtering," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2503-2514, July 2009.

[15] J. Chambers and A. Avlonitis, "A Robust Mixed-Norm Adaptive Filter Algorithm," *IEEE Signal Processing Letters*, vol. 4 , no. 2, pp. 46-48, July 1997.

[16] J. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least Mean Mixed-Norm Adaptive Filtering," *Electronics Letters*, vol. 30, pp. 1574C1575, 1994