LETTER

Sparsity Regularized Affine Projection Adaptive Filtering for System Identification

Young-Seok CHOI[†a)], ***Member***

SUMMARY A new type of the affine projection (AP) algorithms which incorporates the sparsity condition of a system is presented. To exploit the sparsity of the system, a weighted l_1-norm regularization is imposed on the cost function of the AP algorithm. Minimizing the cost function with a subgradient calculus and choosing two distinct weightings for l_1-norm, two stochastic gradient based sparsity regularized AP (SR-AP) algorithms are developed. Experimental results show that the SR-AP algorithms outperform the typical AP counterparts for identifying sparse systems.

key words: *system identification, adaptive filter, affine projection, sparsity, sparse system*

1. Introduction

Adaptive filtering algorithms have gained popularity and proven to be efficient in various applications such as system identification, channel equalization, echo cancellation, and so on. Among various adaptive filtering methods, the relative simplicity and ease of implementation of the normalized least mean square (NLMS) algorithm have made it a popular choice for adaptive filtering applications. However, its convergence rate is significantly deteriorated for correlated input signals [1], [2]. To overcome this issue, the affine projection (AP) algorithm was introduced [3]. The AP algorithm makes use of multiple input vectors in updating the filter weights, leading to a faster convergence over the LMS-type filters which update the filter weights based only on the current input vector [2], [3]. In spite of the impressive features of the AP algorithm, its use is limited when identifying sparse systems, which is common in practice. Examples include echo paths [4] and multipath wireless communication channels [5]. To address this issue, variants of the AP algorithm which employ the variable gain parameters in accordance with the magnitude of the filter weights have been presented [6], [7]. However, these proportionate AP algorithms do not exploit the sparsity condition of an underlying system to be identified.

More recently, motivated by compressive sensing (CS) framework, a new type of adaptive filtering, which makes use of the sparsity condition of the system directly, has been presented [8], [9]. The core idea behind this approach is to incorporate prior knowledge for the sparse system of interest by imposing an l_1-norm based sparsity regularization. Adding the sparsity constraint (l_1-norm regularization) to the cost function leads to the shrinkage of the least relevant weights of the filter to zeros. However, most preceding works have focused on the LMS and the recursive least square (RLS) algorithms, thus an AP algorithm which exploits the sparsity has been lacking. Along this line, this work presents a new family of sparse AP algorithms in a manner of incorporating a weighted l_1-norm regularization into the cost function of the classical AP algorithm. Through a subgradient calculus and the distinct choice of the weighted l_1-norm regularization, two stochastic gradient based sparsity regularized AP (SR-AP) algorithms are derived: First, a simple l_1-norm sparse AP algorithm is presented. Second, a weighted l_1-norm sparse AP algorithm based on an estimate of the actual sparseness of the system is obtained.

Numerical experiments show that by inheriting the merits of the AP algorithm, the resulting SR-AP algorithms possess superior convergence properties over conventional AP ones, especially when the system is sparse. The remainder of this letter is organized as follows: Sect. 2 briefly reviews the AP algorithm in the context of system identification. In Sect. 3, the proposed SR-AP algorithms are developed. In Sect. 4, the simulation results are presented. Section 5 concludes this study.

2. Affine Projection Algorithm for System Identification

Consider a desired signal $d(i)$ that arise from the system identification model

$$d(i) = \mathbf{u}_i\mathbf{h}^\circ + v(i), \tag{1}$$

where i is the time index, $\mathbf{h}^\circ$ is a column vector for the impulse response of an unknown system that we wish to estimate, $v(i)$ accounts for measurement noise with zero mean and variance σ_v^2, and $\mathbf{u}_i = [u(i)\ \ u(i-1)\ \ \cdots u(i-M+1)]$ is an $1 \times M$ row input vector. In [10], the cost function of the AP algorithms is given by

$$J_{\mathrm{AP}}(i) = E[\mathbf{e}_i^*(\mathbf{U}_i\mathbf{U}_i^*)^{-1}\mathbf{e}_i]/2, \tag{2}$$

where

$$\mathbf{U}_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{u}_{i-1} \\ \vdots \\ \mathbf{u}_{i-K+1} \end{bmatrix}, \quad \mathbf{d}_i = \begin{bmatrix} d(i) \\ d(i-1) \\ \vdots \\ d(i-K+1) \end{bmatrix},$$

Manuscript received September 9, 2013.
Manuscript revised December 11, 2013.
[†]The author is with the Department of Electronic Engineering, Gangneung-Wonju National University, Gangneung, 210–702, Republic of Korea.
a) E-mail: yschoi@gwnu.ac.kr
DOI: 10.1587/transinf.E97.D.964

$\mathbf{e}_i = \mathbf{d}_i - \mathbf{U}_i\mathbf{h}_{i-1}$, and $\mathbf{h}_{i-1}$ is an estimate for $\mathbf{h}^\circ$ at $(i-1)$th iteration. By minimizing the cost function (2), the update recursion of the AP algorithm is represented as

$$\mathbf{h}_i = \mathbf{h}_{i-1} + \mu\mathbf{U}_i^*(\mathbf{U}_i\mathbf{U}_i^* + \delta\mathbf{I}_K)^{-1}\mathbf{e}_i, \tag{3}$$

where μ is the step-size, ρ is the regularization parameter, and $\mathbf{I}_K$ denotes a $K \times K$ identity matrix.

3. Sparsity Regularized Affine Projection Algorithm

In order to take into account the sparsity characteristic, an augmented cost function which incorporates a weighted l_1-norm into (2) is newly formulated as

$$J_{\text{SR-AP}}(i) = \mathrm{E}\left[\mathbf{e}_i^*(\mathbf{U}_i\mathbf{U}_i^*)^{-1}\mathbf{e}_i\right]/2 + \gamma\|\mathbf{W}\mathbf{h}_{i-1}\|_1, \tag{4}$$

where $\|\mathbf{W}\mathbf{h}_{i-1}\|_1 = \sum_{k=0}^{k=M-1} w_k|h_{i-1,k}|$ accounts for a weighted l_1-norm of the estimated filter weight, $\mathbf{W}$ is the $M \times M$ weighting identity matrix whose diagonal elements are w_k, and $h_{i,k}$, $k = 0, 1, \ldots, M-1$ denote the kth weight of $\mathbf{h}_i$. In addition, γ is a positive valued parameter which provides a trade-off between the error related term and the sparsity. Then, a stochastic gradient update recursion with the aim of minimizing (4) is derived as follows:

$$\begin{aligned}\mathbf{h}_i &= \mathbf{h}_{i-1} - \mu\nabla_{\mathbf{h}}J_{\text{SR-AP}}(i)\\ &= \mathbf{h}_{i-1} + \mu\mathbf{U}_i^*(\mathbf{U}_i\mathbf{U}_i^* + \delta\mathbf{I}_K)^{-1}\mathbf{e}_i - \mu\gamma\nabla_{\mathbf{h}}^s\|\mathbf{W}\mathbf{h}_{i-1}\|_1,\end{aligned} \tag{5}$$

where $\nabla_{\mathbf{h}}^s f(\cdot)$ denotes a subgradient vector of the function $f(\cdot)$ with respect to $\mathbf{h}$. Since the weighted l_1-norm is not differentiable with respect to $\mathbf{h}_{i-1}$ when $\mathbf{h}_{i-1}$ equals zero, here, the subgradient calculus is employed [12]. The subgradient vector $\nabla_{\mathbf{h}}^s\|\mathbf{W}\mathbf{h}_{i-1}\|_1$ can be obtained as $\mathbf{W}^T\mathrm{sgn}(\mathbf{W}\mathbf{h}_{i-1}) = \mathbf{W}\mathrm{sgn}(\mathbf{h}_{i-1})$, since $\mathbf{W}$ is assumed as a diagonal matrix with positive-valued elements. Then, a framework of the AP algorithms with sparsity can be written as follows as:

$$\mathbf{h}_i = \mathbf{h}_{i-1} + \mu\mathbf{U}_i^*(\mathbf{U}_i\mathbf{U}_i^* + \delta\mathbf{I}_K)^{-1}\mathbf{e}_i - \mu\gamma\mathbf{W}\mathrm{sgn}(\mathbf{h}_{i-1}). \tag{6}$$

Note that an update recursion (6) reduces to the typical AP algorithm if $\gamma = 0$. Here, by choosing the weighting matrix $\mathbf{W}$, two versions of the AP algorithm with the sparsity constraint are developed: First, the use of the identity matrix as the weighting matrix, i.e., $\mathbf{W} = \mathbf{I}_M$, leads to the following update recursion

$$\mathbf{h}_i = \mathbf{h}_{i-1} + \mu\mathbf{U}_i^*(\mathbf{U}_i\mathbf{U}_i^* + \delta\mathbf{I}_K)^{-1}\mathbf{e}_i - \mu\gamma\mathrm{sgn}(\mathbf{h}_{i-1}), \tag{7}$$

which is referred to as the sparsity regularized AP-1 (SR-AP-1) algorithm.

Second, the choice of the weights inversely proportional to the magnitude of the system weights results in an approximation of the actual sparseness of the underlying system, i.e., the l_0-norm of the system [8]. Due to unavailability of the system weights, here, the magnitude of the current filter weights are used as an alternative as follows [13]:

$$w_j = \frac{1}{|h_{i-1,j}| + \epsilon}, \tag{8}$$

where $h_{i-1,j}$ is the j-th tap of $\mathbf{h}_{i-1}$ and ϵ is a small positive value to avoid singularity in the case when $|h_{i-1,j}| = 0$. Then, the weighting matrix $\mathbf{W}$ consists of the values of w_j as the j-th diagonal elements. Finally, the second AP algorithm with sparsity constraint is given by

$$\mathbf{h}_i = \mathbf{h}_{i-1} + \mu\mathbf{U}_i^*(\mathbf{U}_i\mathbf{U}_i^* + \delta\mathbf{I}_K)^{-1}\mathbf{e}_i - \mu\gamma\frac{\mathrm{sgn}(\mathbf{h}_{i-1})}{|\mathbf{h}_{i-1}| + \epsilon}, \tag{9}$$

where the vector division operation accounts for a component-wise division. This update recursion is referred to as the sparsity regularized AP-2 (SR-AP-2) algorithm.

4. Experimental Results

To assess the performance of the proposed SR-APs, the system identification simulations were carried out. A system to be identified has 64 taps and a few taps of them, i.e., L taps, have non-zero values, indicating the sparse characteristic. Then, the degree of sparsity is represented as $S = L/64$. The adaptive filter is assumed to have identical length of $M = 64$. Figure 1 shows an example of a sparse system $\mathbf{h}^\circ$ of $L = 8$. The input signal is obtained by filtering a white, zero-mean, Gaussian random sequence through a first-order system $F(z) = 1/(1 - 0.9z^{-1})$. The signal-to-noise ratio (SNR) is computed by $10\log_{10}(E[y(i)^2]/E[v(i)^2])$, where $y(i) = \mathbf{u}_i\mathbf{h}^\circ$.The measurement noise $v(i)$ is added to $y(i)$.The mean square deviation (MSD), $E\|\mathbf{h}^\circ - \mathbf{h}_i\|^2$, is taken and averaged over 50 independent trials. For the conventional APs and the SR-APs, the projection order $k = 4$, and the step-size parameter $\mu = 1$ (and $\mu = 0.15$ for the AP in Figs. 2 and 6) are chosen in the following system identifications. In addition, the number of non-zero taps is set to $L = 4$ except Fig. 5 where various values of L are considered.

Figure 2 illustrates the MSD curves of the classical APs and two SR-APs, i.e., SR-AP-1 and SR-AP-2, in the case of SNR = 30dB. For comparison purpose, the improved proportionate AP (IPAP) [6], [7] and variable step-size AP (VS-AP) [11] are considered. The parameters, $\alpha = -0.5$ for the IPAP and $C = 0.003$ for the VS-AP, are chosen, respectively. For both the SR-AP-1 and SR-AP-2, $\gamma = 3 \times 10^{-4}$ is used.

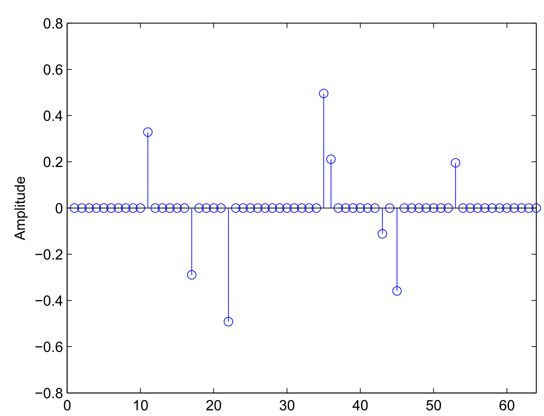

Fig. 1 Example of sparse system $\mathbf{h}^\circ$.

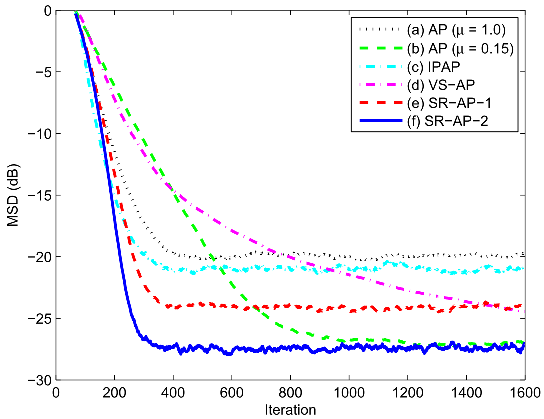

Fig. 2 MSD curves of the AP, IPAP, VS-AP, and SR-AP algorithms [$K = 4$, $L = 4$, SNR = 30dB].

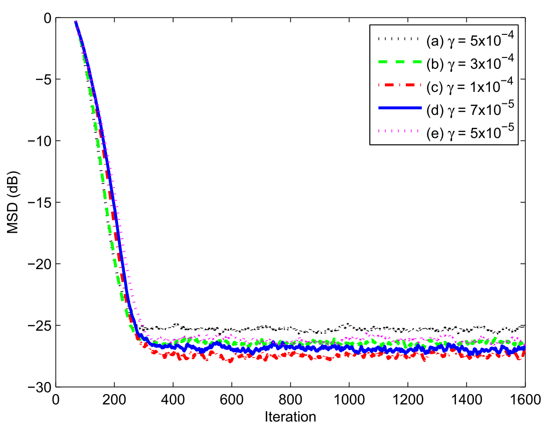

Fig. 3 MSD curves of the SR-AP-2 algorithm with various values of γ [$K = 4$, $L = 4$, SNR = 30dB].

In addition, the parameter $\epsilon = 0.1$ is chosen for the SR-AP-2. In the figure, it is clear that the SR-AP-2 indicates the best performance as well as the SR-AP-1 outperforms the classical APs, i.e., the AP, IPAP, and VS-AP in terms of the convergence rate and the steady-state misalignment.

Figure 3 shows the MSD curves of the SR-AP-2 when various values of γ are chosen. In the figure, the convergence performance is not highly sensitive to γ.

Then, in order to validate the convergence performance of the classical APs and the SR-APs under various SNRs (from −5 to 30 dB), the steady-state MSD values are compared in Fig. 4. As can be seen, the SR-AP-2 outperforms other APs under various SNRs. In addition, the SR-AP-1 is better than the conventional APs and comparable with the VS-AP for more than 5 dB in terms of the steady-state misalignment.

Second, the convergence properties of the classical APs and SR-AP are compared under various sparsity conditions. The same number of tap of system with the first simulation is used ($M = 64$) and the different sparsity conditions (S=2/64, 8/64, and 16/64) are considered. Figure 5 shows the MSD curves of the AP, IPAP and SR-AP-2 in the case of SNR = 30dB. It clearly shows that the more severe the spar-

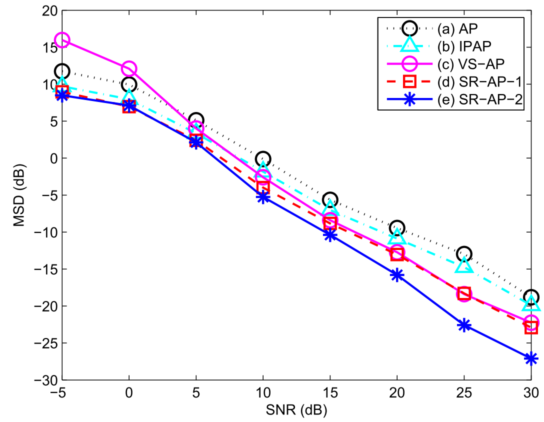

Fig. 4 Steady-state MSD values of the AP, IPAP, VS-AP and SR-AP algorithms under various SNRs [$K = 4$, $L = 4$, SNR $= -5, 0, 5, 10, 15, 20, 25, 30$dB].

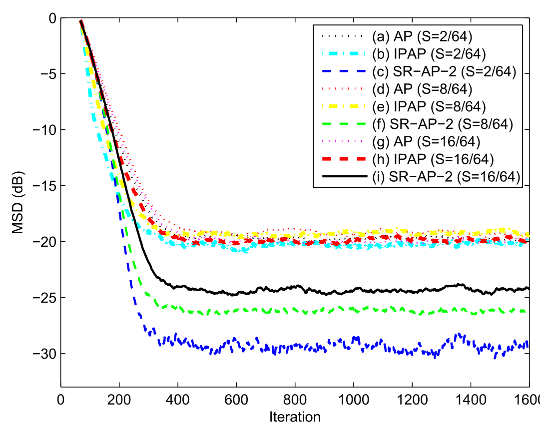

Fig. 5 MSD curves of the AP, IPAP, and SR-AP-2 algorithms for various sparsity conditions [$K = 4$, SNR = 30dB, S = 2/64, 8/64, 16/64].

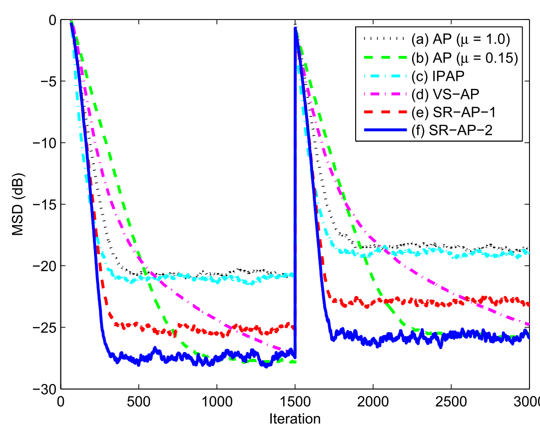

Fig. 6 MSD curves of the AP, IPAP, VS-AP, and SR-AP algorithms for time-varying system identification setup [$K = 4$, $L = 4$, SNR = 30dB].

sity of the system, the better the convergence performance of the SR-AP over the classical APs.

Finally, Fig. 6 illustrates the tracking performance of the SR-AP algorithms with regard to a sudden change in the unknown system. At the 1500*th* iteration, the sparsity condition of the unknown system is altered in that the positions of non-zero taps ($L = 4$) are changed. The figure clearly shows that the proposed SR-AP algorithms keep track of sudden weight change without degrading the convergence rate and the steady-state error, outperforming the AP counterparts.

5. Conclusion

This work presented a novel family of the AP algorithms which employs the sparsity constraint in identifying sparse systems. The proposed AP algorithms take into account the sparsity property by incorporating the variants of the system's l_1-norm into the cost function. Employing the subgradient calculus and choosing the weighting matrix, two stochastic gradient AP algorithms with the sparsity constraint were developed. The resulting SR-AP algorithms have proven their superiority over the conventional AP counterparts, especially in cases when systems are severely sparse.

References

[1] S. Haykin, Adaptive filter theory, Prentice Hall, Upper Saddle River, NJ, 2002.

[2] A.H. Sayed, Fundamentals of adaptive filtering, Wiley, New York, 2003.

[3] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," Electro. Commun. Jpn., vol.67-A, no.5, pp.19–27, 1984.

[4] D.L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancellers," IEEE Trans Signal Speech Audio Process., vol.8, no.5, pp.508–518, Sept. 2000.

[5] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," SIAM Review, vol.51, no.1, pp.34–81, 2009.

[6] O. Hoshuyama, R.A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," Proc. Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP 2004), pp.IV-161–IV-164, 2004

[7] C. Paleologu, S. Ciochina, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," IEEE Signal Process. Lett., vol.17, no.29, pp.165–168, Feb. 2010.

[8] Y. Chen, Y. Gu, and A.O. Hero, "Sparse LMS for system identification," Proc. Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP 2009), pp.3125–3128, 2009.

[9] Y. Gu, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system identification," IEEE Signal Process. Lett., vol.16, no.9, pp.774–777, Sept. 2009.

[10] Y.-S. Choi and W.-J. Song, "Noise constrained data-reusing adaptive filtering algorithms for system identification," IEICE Trans. Fundamentals., vol.E95-A, no.6, pp.1084–1087, June 2012.

[11] H.-C. Shin, Ali H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," IEEE Signal Process. Lett., vol.11, no.2, pp.132–135, Feb. 2004.

[12] D. Bertsekas, A. Nedic, and A. Ozdaglar, Convex analysis and optimization, Athena Scientific, Cambridge, MA USA, 2003.

[13] E.J. Candes, M.B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighting l_0 Minimization," J. Fourier Anal. Appl., vol.14, pp.877–905, 2008.