Variable Step-Size Affine Projection Algorithm with a Weighted and Regularized Projection Matrix

Tao Dai
School of Information Technology and Engineering (SITE), University of Ottawa, Canada
e-mail: tdai@site.uottawa.ca

Andy Adler
School of Information Technology and Engineering (SITE), University of Ottawa, Canada
e-mail: adler@site.uottawa.ca

Behnam Shahrrava
Department of Electrical and Computer Engineering, University of Windsor, Canada
e-mail: shahrrav@uwindsor.ca

Abstract

*This paper presents a regularized modification to the weighted variable step-size affine projection algorithm (*APA*). The regularization overcomes the ill-conditioning introduced by both the forgetting process and the increasing size of the input matrix. The algorithm was tested by trials with colorized input signals and different parameter combinations. Simulations illustrate that the proposed algorithm is superior in terms of convergence speed and misadjustment compared with existing algorithms.*

Keywords *— Adaptive Signal Processing; Affine Projection Algorithms; Variable Step-size Adaptive Algorithms; Matrix Singularity; Regularization.*

1 Introduction

Adaptive signal processing algorithms have been widely used in numerous applications, such as noise cancelation, system identification and data feature extraction. These algorithms are designed to minimize a performance cost function. The Least Mean Squares (*LMS*) algorithm [1], based on minimizing Mean Squared Error (MSE), is a common algorithm of this type. The Normalized Least Mean Square (*NLMS*) algorithm is one of the most widely used adaptive filters because of its computational simplicity. However, colored input signals can deteriorate the convergence rate of *LMS* type algorithms [1]. To address this problem, the Affine Projection Algorithm (*APA*), a generalized form of *NMLS*, was proposed by Ozeki et al. [2] using affine subspace projections. Shin et al.[3] provided a unified treatment of the transient performance of the *APA* family. Sankaran et al.[4] analyzed convergence behaviors of *APA* class.

In conventional *LMS*, *NLMS*, and *APA* algorithms, a fixed step size μ governs the tradeoff between the convergence rate and the misadjustment. To realize both fast convergence and low steady-state deviation, a variable step (*VS*) is necessary[5][6][7]. Harris et al.[5] used a feedback coefficient based on the sign of the gradient of the squared error; Mader et al.[6] proposed an optimum step size for NLMS. Shin et al.[7] proposed a criterion to measure the adaptation states and developed a variable step-size *APA* based on this criterion. In [8], Dai et al. proposed a weighted method for the variable step size affine projection algorithm, which processes the projection matrix with a forgetting factor to better estimate weights deviation.

The method in [8] improves convergence performance compared with existing schemes. However, as the input matrix size is increased, especially when the forgetting process is introduced, the matrix becomes ill-conditioned, and the projected error estimate becomes worse. In this paper, we address this ill-conditioning by introducing a regularization term to the weighted projection matrix of [8]. This modification gives further improvement and robustness to the previous method.

2 Methods

2.1 Optimal Variable Step-Size APA

The input vector, $\mathbf{x}_i$, and the desired scalar output, d_i, are related by

$$d_i = \mathbf{x}_i \mathbf{w}^\circ + v_i$$

The subscript i is the time index corresponding to the i^{th} sampling instant; $\mathbf{w}^\circ$ is an unknown $L \times 1$ column vector to be estimated; $\mathbf{x}$ is a $1 \times L$ row vector; v is a zero mean Gaussian independent noise sequence, such that $\mathbf{x}$ and v are independent.

The Affine Projection Algorithm (*APA*)[2] updates weights via

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \tag{1}$$

where $U_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_{i-1} \\ \cdots \\ \mathbf{x}_{i-K+1} \end{bmatrix}$ $\mathbf{d}_i = \begin{bmatrix} d_i \\ d_{i-1} \\ \cdots \\ d_{i-K+1} \end{bmatrix}$ $\mathbf{w}_i = \begin{bmatrix} w_{0,i} \\ w_{i,i} \\ \cdots \\ w_{L-1,i} \end{bmatrix}$ and the error signal is $\mathbf{e}_i = \mathbf{d}_i - U_i \mathbf{w}_{i-1}$. $\mathbf{x}_i$ is the input vector at the i^{th} sampling instant. $\mathbf{d}$ is the desired signal; μ is the step size; K is the *APA* order or signal window width, L is filter order, and * is the conjugate transpose operator.

Shin et al.[7] proposed the optimal variable step-size *APA* (*VS-APA*) in which (1) can be written as

$$\widetilde{\mathbf{w}}_i = \widetilde{\mathbf{w}}_{i-1} - \mu U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \tag{2}$$

where $\widetilde{\mathbf{w}}_i = \mathbf{w}^\circ - \mathbf{w}_i$.

$$\mathbf{p}_i \triangleq U_i^* \left(U_i U_i^*\right)^{-1} U_i \widetilde{\mathbf{w}}_{i-1} \tag{3}$$

which is the projection of $\widetilde{\mathbf{w}}_{i-1}$ onto $\Re(U_i^*)$, the range space of U_i^*. Based on the definition of $\mathbf{p}$,

$$E\left[\mathbf{p}_i\right] = E\left[U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i\right] \tag{4}$$

Shin et al.[7] select the optium adaptive filter as the minimizer of $\|\mathbf{p}_i\|$. For this case, $\mathbf{p}_i$ is estimated as follows:

$$\hat{\mathbf{p}}_i = \alpha\hat{\mathbf{p}}_{i-1} + (1-\alpha)\mathbf{p}_i \tag{5}$$

for a smoothing factor α, $0 \leq \alpha < 1$. Then the variable step-size APA becomes

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu_i U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \tag{6}$$

where

$$\mu_i = \mu_{max}\frac{\|\hat{\mathbf{p}}_i\|^2}{\|\hat{\mathbf{p}}_i\|^2 + C} \tag{7}$$

for a positive constant, C is related to $\sigma_v^2 Tr\{E[(U_iU_i^*)^{-1}]\}$, which can be approximated as K/SNR. When $\|\hat{\mathbf{p}}_i\|^2$ is large, $\mathbf{w}_i$ is far from $\mathbf{w}^\circ$ and μ_i is close to μ_{max}; when $\|\hat{\mathbf{p}}_i\|^2$ is small, $\mathbf{w}_i$ approaches $\mathbf{w}^\circ$ and μ_i is close to zero.

2.2 Optimal Variable Step Size APA with Forgetting Factor

Previously, we introduced a forgetting factor into the pseudo-inverse projection matrix, resulting in a marked convergence enhancement [8]. The input matrix at time i can be described as:

$$\begin{aligned} &U_i[k+1, l+1] = x_{i-k-l} \\ &k = 0, 1, \ldots, K-1; \quad l = 0, 1, \ldots, L-1; \end{aligned} \tag{8}$$

By introducing a forgetting factor λ, $0 < \lambda \leq 1$,

$$U_i'[k+1, l+1] = x_{i-k-l}\lambda^{k+l} = \lambda^k x_{i-k-l}\lambda^l. \tag{9}$$

In matrix notation, we represent this as

$$U_i' = \Lambda^{(K)} U_i \Lambda^{(L)} \tag{10}$$

where $\Lambda^{(m)}$ is an $m \times m$ diagonal matrix with

$$\left[\Lambda^{(m)}\right]_{j,j} = \lambda^{j-1} \quad j = 1, 2, \ldots, m$$

then (4) becomes

$$\mathbf{p}_i' = U_i'^* \left(U_i' U_i'^*\right)^{-1} \mathbf{e}_i \tag{11}$$

The newly generated projection matrix in (10) is time-dependent; the latest data are more significant in the pseudo-inverse matrix by which the error vector is projected.

The proposed variable step size APA with forgetting factor (*VS-APA-FF*) is:

$$\begin{aligned} &\mathbf{w}_i = \mathbf{w}_{i-1} + \mu_i U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \\ &\mu_i = \mu_{max}\frac{\|\hat{\mathbf{p}}_i'\|^2}{\|\hat{\mathbf{p}}_i'\|^2 + C} \\ &\hat{\mathbf{p}}_i' = \alpha\hat{\mathbf{p}}_{i-1}' + (1-\alpha)\mathbf{p}_i' \qquad 0 \leq \alpha < 1 \end{aligned} \tag{12}$$

Note that U_i is only replaced by U_i' during the error evaluation phase (11), not during the weights updating phase because of instability which has been observed in some simulations of replacing U_i by U_i' for both. This phenomenon is most possibly due to the ill-conditioning of the input matrix U_i caused by forgetting process.

A special case of this algorithm is the variable step size NLMS with forgetting factor (VS-NLMS-FF) obtained by setting $K = 1$. For this case, the input matrix U_i is a row vector and the forgetting factor processing is implemented only in the row direction.

$$U_i' = U_i \Lambda^{(L)} \tag{13}$$

2.3 Regularization of the Ill-Conditioned Projection Matrix

In (11) of the previously proposed algorithm, $(U_i'U_i'^*)$ is potentially ill-conditioned with small singular values. Using the *singular value decomposition (SVD)*, U' can be decomposed as:

$$U' = R\Sigma V^* \tag{14}$$

where R and V are $K\times K$ and $L\times L$ unitary matrices, respectively. Σ is a $K\times L$ matrix with nonnegative diagonal elements of singular values σ_i, The ill-conditionness of U is characterized by its condition number,

$$cond\,U = \sigma_{max}/\sigma_{min} = \sigma_1/\sigma_K \tag{15}$$

from (10), the SVD of the weighted input matrix U' is:

$$\begin{aligned} U' &= \Lambda^{(K)} U \Lambda^{(L)} = \Lambda^{(K)}[R\Sigma V^*]\Lambda^{(L)} \\ &= R(\Lambda^{(K)}\Sigma\Lambda^{(L)})V^* \\ &= R\Sigma' V^* \end{aligned} \tag{16}$$

where Σ' is a $K \times L$ matrix with all zero entities except $[\Sigma']_{j,j} = \lambda^{2(j-1)}\sigma_j$, $j = 1, 2, \ldots, K$. The condition number of the weighted input matrix U' is:

$$cond\,U' = \sigma_1/[\lambda^{2(K-1)}\sigma_K] = \lambda^{2(1-K)}\,cond\,U$$

which illustrates the increasing condition number due to decrease in λ and increase in K. Because of this ill-conditioning, the estimated $\mathbf{p}'$ may not be a true evaluation of the error signal. Even if the error signal is stable, the projected $\mathbf{p}'$ could be unstable. Thus the VS-APA and VS-APA-FF algorithms adopt a smoothing function, in the form of (5), to alleviate this problem with the cost loss of error signal fidelity, which sacrifices convergence speed and/or misadjustment.

We propose to address this problem using a Tikhonov regularization approach, under which (11) becomes:

$$\mathbf{p}_i' = U_i'^*(U_i'U_i'^* + \delta^2 I)^{-1}\mathbf{e}_i. \tag{17}$$

where I is the identity matrix, and δ is a hyperparameter to control the amount of regularization. The modified algorithm becomes:

$$\begin{aligned} &\mathbf{w}_i = \mathbf{w}_{i-1} + \mu_i U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \\ &\mu_i = \mu_{max}\frac{\|\hat{\mathbf{p}}_i'\|^2}{\|\hat{\mathbf{p}}_i'\|^2 + C} \end{aligned} \tag{18}$$

Note that the smoothing function is no longer needed since the regularization process accomplishes this function.

3 Simulation Results

The performance of the proposed algorithm is illustrated by simulations of a system identification model[4]. The system to be simulated is represented by a moving average model with L taps. The adaptive filter has the same number of taps. The goal of the adaptive processing is to estimate system parameters by optimizing the adaptive filter parameters iteratively using the proposed algorithm. Two colorized Gaussian noises are used as input signals. The input signal colorizations are obtained by filtering a white Gaussian random noise(zero mean, unit variance) through a 1^{st} order filter, $G_1(z) = 1/(1 - 0.9z^{-1})$ or a 4^{th} order filter

$$G_2(z) = \frac{1 + 0.9z^{-1} + 0.6z^{-2} + 0.81z^{-3} - 0.329z^{-4}}{1 + z^{-1} + 0.21z^{-2}}$$

The measurement noise v_i is added to y_i ($y_i = \mathbf{x}_i\mathbf{w}^\circ$) and the SNR of the measurement signal is calculated by

$$SNR = 10log_{10}(\frac{E[y_i^2]}{E[v_i^2]})$$

The simulation results are obtained by averaging 100 independent trials, with a smoothing factor $\alpha = 0.99$ for *VS-APA* and *VS-APA-FF*. The convergence is evaluated by Mean Square Deviation (MSD) which is calculated by $E(\|\tilde{\mathbf{w}}_i\|^2) = E(\|\mathbf{w}^\circ - \mathbf{w}_i\|^2)$. Figure 1 gives a *VS-APA-FF* example, illustrating effects of different forgetting factors on optimization performance. For this case, $\lambda = 0.7$ is the optimal value for the best convergence. Empirically, we obtained recommended forgetting factors for *VS-APA-FF* (Table I) on various cases. (Note that when $\lambda = 1$, the *VS-APA-FF* becomes the original *VS-APA*. Therefore, *VS-APA* is a special case of *VS-APA-FF*).

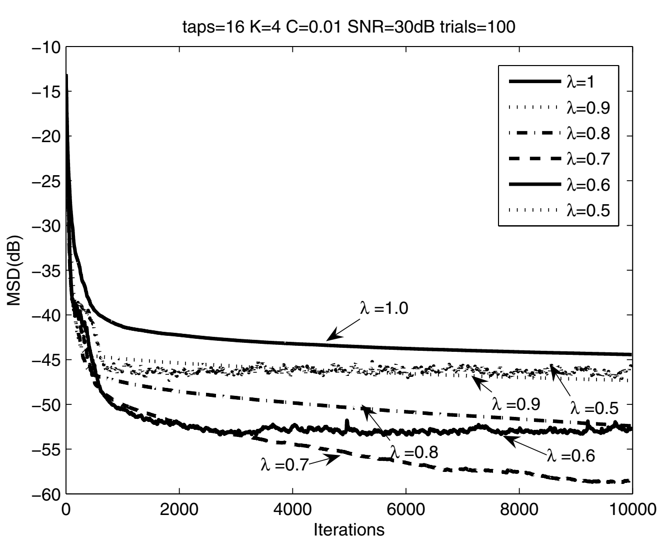


Figure 1. MSD vs. iteration number for *VS-APA-FF* for effect of different forgetting factors λ. (L=16, K=4, SNR=30dB, G2 colorization)

From Table I we conclude:

- the optimal value of forgetting factor increases with the increment of the APA order K. *VS-APA-FF* outperforms *VS-APA* for small K and is gradually beaten by *VS-APA* when K increases ($e.g.K = 4, 8$, G1 colorization, $SNR = 30dB$) due to increasing ill-conditioning.
- *VS-APA-FF* is good at low noise conditions compared to *VS-APA*. In other words, the advantage of *VS-APA-FF* over *VS-APA* becomes less significant with increased noise level.
- noise color affects adaptation performance. (This applies for all APA class)

TABLE I

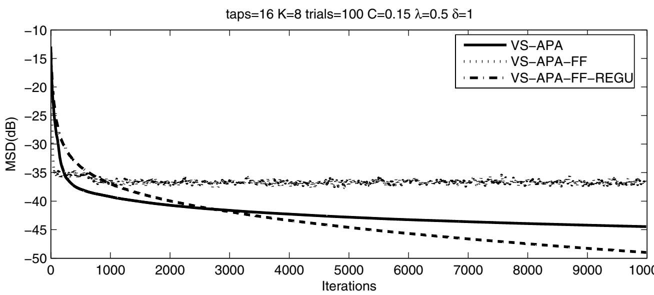
Recommended values of forgetting factor λ for VS-APA-FF. *(L=16)*

K	C	λ			
		G1		G2	
		SNR =30dB	SNR =40dB	SNR =30dB	SNR =40dB
1	0.0001	**0.8**	**0.4**	**0.5**	**0.1**
2	0.001	**0.9**	**0.8**	**0.5**	**0.3**
4	0.01	**1**	**0.9**	**0.7**	**0.6**
8	0.15	**1**	**0.9**	**0.8**	**0.8**

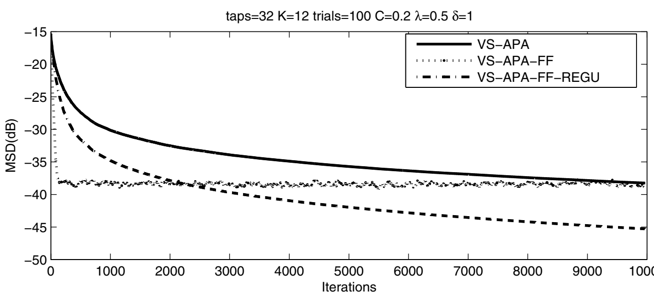
Using experimental conditions described previously, and $\lambda = 0.5$, $\delta = 1$, simulation comparisons between *VS-APA*, *VS-APA-FF*, and the regularized version *VS-APA-FF-REGU* porposed here, are illustrated by figure 2 (noise G2) and figure 3 (noise G1). For some cases, *VS-APA-FF* converges quickly but with high misadjustment (Figure 2(b),2(c)), while *VS-APA-FF-REGU* converges more slowly but with much lower misadjustment. When the update matrix of *VS-APA-FF* becomes severly ill-conditioned (Figure 2(a)3(a)3(b)3(c)) and behaves even worse than *VS-APA*, the *VS-APA-FF-REGU* can still converge quickly and with low misadjustment. Therefore we conclude that *VS-APA-FF-REGU* is a good complement for *VS-APA-FF*, when the forgetting processed input matrix is close to singular.

4 Conclusions

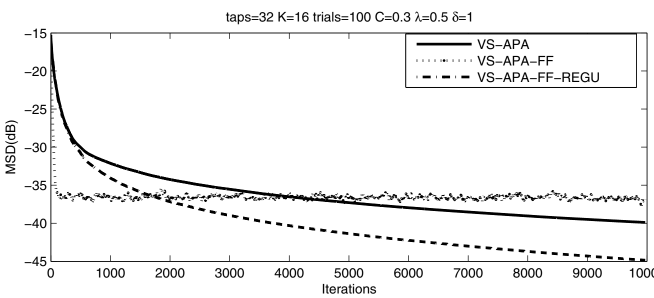
This paper presents a new variable step size APA algorithm, *VS-APA-FF-REGU*, with a projection matrix processed with a forgetting factor and using a regularization term. The ill-conditioning of the projection matrix becomes significant when the size of input matrix is large, especially when the forgetting process is introduced. The Tikhonov regularization is used to overcome the ill-conditionness of the forgetting processed input matrix. The proposed algorithm is more stable and converges better than previous algoriths.



(a)

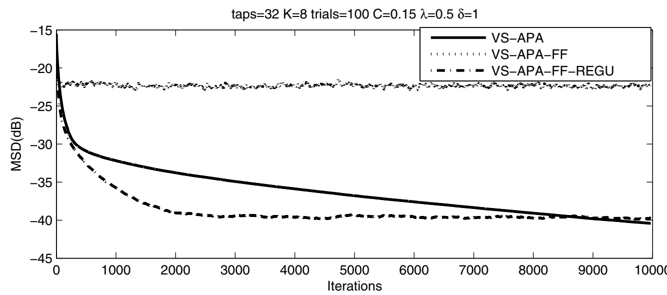


(b)

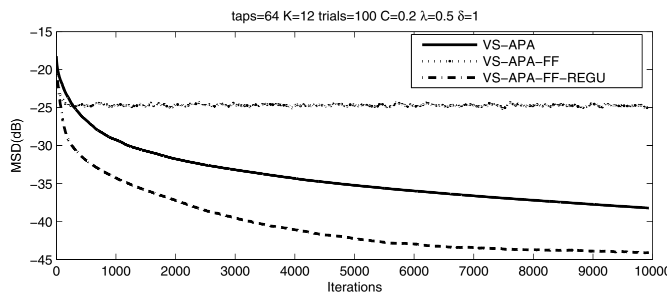


(c)

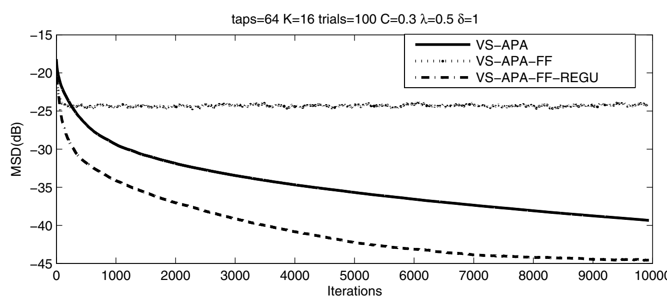
Figure 2. Comparisons among VS-APA, VS-APA-FF, and VS-APA-FF-REGU, G2 colorization. $\lambda = 0.5$. (a) K=8, taps=16, C=0.15; (b) K=12, taps=32, C=0.2; (c)K=16, taps=32, C=0.3



(a)



(b)



(c)

Figure 3. Comparisons among VS-APA, VS-APA-FF, and VS-APA-FF-REGU, G1 colorization. $\lambda = 0.5$. (a) K=8, taps=32, C=0.15; (b) K=12, taps=64, C=0.2; (c)K=16, taps=64, C=0.3

In the weighted and regularized variable step size *APA*, choosing a proper regularization parameter δ is essential. An empirical δ is adopted in this paper. The strategy of deciding the optimal value of δ for various situations will be of importance for further algorithm updating.

References

[1] B.Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

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

[3] H-C Shin, A.H.Sayed, "Transient behavior of affine projection algorithms", *ICASSP*, VI 353-356, 2003.

[4] Sundar G. Sankaran,A. A.Beex, "Convergence behaviour of affine projection algorithms", *IEEE Trans. Signal Processing*, vol 48, no 4, Apr 2000

[5] Richard W. Harris, Douglas M.Chabries, F.Avery Bishop, "A Variable Step (VS) Adaptive Filter Algorithm", *Trans on Acoustics, Speech, and Signal Proc.*, Vol ASSP-34, no 2, Apr 1986

[6] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters-An overview", *Signal Process.*, vol. 80, pp. 1697-1719, Sept 2000.

[7] H-C Shin, A.H. Sayed, W-J Song, "Variable step-Size NLMS and affine projection algorithms", *IEEE Signal Proc. Letters*, Vol.11,No.2, Feb 2004.

[8] T.Dai, B.Shahrrava, X.Chen, "A variable step-size affine projection algorithm with a weighted projection matrix", *IEEE Canadian Conference (CCECE)*,320-323,May 2005.