

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

Tao Dai, Andy Adler, and Behnam Shahrrava

**Abstract**—This paper presents a forgetting factor scheme for variable step-size affine projection algorithms (*APA*). The proposed scheme uses a forgetting processed input matrix as the projection matrix of pseudo-inverse to estimate system deviation. This method introduces temporal weights into the projection matrix, which is typically a better model of the real error's behavior than homogeneous temporal weights. The regularization overcomes the ill-conditioning introduced by both the forgetting process and the increasing size of the input matrix. This algorithm is tested by independent trials with coloured input signals and various parameter combinations. Results show that the proposed algorithm is superior in terms of convergence rate and misadjustment compared to existing algorithms. As a special case, a variable step size *NLMS* with forgetting factor is also presented in this paper.

**Keywords**—Adaptive signal processing, affine projection algorithms, variable step-size adaptive algorithms, regularization.

## I. INTRODUCTION

Adaptive signal processing algorithms have been widely used in numerous applications, such as noise cancellation, system identification and data feature extraction. These algorithms are designed to minimize a performance cost function. The Least Mean Square (*LMS*) algorithm [1][2], based on minimizing Mean Squared Error (MSE), is a common algorithm of this type. The Normalized Least Mean Square (*NLMS*) [3] algorithm is one of the most widely used adaptive algorithms because of its computational simplicity and robustness. However, highly correlated/colored input signals can deteriorate the convergence rate of *LMS* type algorithms [1]. Recursive Least Squares *RLS* algorithm [4] is known for fast convergence even with colored inputs, however, with cost of higher computational complexity. The Affine Projection Algorithm (*APA*) was proposed by Ozeki and Umeda [5] using affine subspace projections. As the generalization of the *NLMS*, the algorithm applies de-correlation of the input process to speed up the convergence, without evident increase of computations and instability problems met by *RLS*[6]. A Fast Affine Projection (*FAP*) was proposed by Gay and Tavathia [7] for acoustic echo cancellation. Shin and Sayed [8] provided a unified treatment of the transient performance of the *APA* family. Sankaran and Beex [9] analyzed convergence behaviors of the *APA* class.

Tao Dai is with the Department of System and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. email: tdai@sce.carleton.ca

Andy Adler is with the Department of System and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. email: adler@sce.carleton.ca

Behnam Shahrrava is with the Department of Electrical and Computer Engineering, University of Windsor, Ottawa, Ontario, Canada. email: shahrrav@uwindsor.ca

In conventional *LMS*, *NLMS*, and *APA* algorithms, a fixed step size  $\mu$  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. Harris *et al.* [10] used a feedback coefficient based on the sign of the gradient of the squared error; Mader *et al.* [11] proposed an optimum step size for *NLMS*; Shin *et al.* [12] proposed a criterion to measure the adaptation states and developed a variable step-size *APA* based on this criterion.

In this paper, a forgetting factor method for the variable step size affine projection algorithm is presented. This approach uses a forgetting factor processed input matrix as the projection matrix of pseudo-inverse to estimate weights deviation. Compared with existing algorithms, it significantly improves convergence performance in terms of speed and misadjustment. However, as the input matrix size increases, especially when forgetting process is introduced, the matrix singularity deteriorates and makes the projected error inaccurate. The regularization method resolves this problem and it gives further improvement over the previous method.

## II. METHODS

### A. Optimal Variable Step-Size APA

The Affine Projection Algorithm (*APA*)[5] updates the weight vector  $\mathbf{w}$  via

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

where

$$U_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_{i-1} \\ \dots \\ \mathbf{x}_{i-K+1} \end{bmatrix} \quad \mathbf{d}_i = \begin{bmatrix} d_i \\ d_{i-1} \\ \dots \\ d_{i-K+1} \end{bmatrix} \quad \mathbf{w}_i = \begin{bmatrix} w_{0,i} \\ w_{1,i} \\ \dots \\ w_{L-1,i} \end{bmatrix}$$

The subscript  $i$  is the time index corresponding to the  $i^{\text{th}}$  sampling instant;  $K$  is the *APA* order or signal window width;  $L$  is filter order;  $\mathbf{w}_i \in \mathbb{R}^{L \times 1}$  is the filter weights vector;  $\mathbf{x}_i \in \mathbb{R}^{1 \times L}$  is the input vector;  $d_i$  is the desired signal;  $\mu$  is the step size; The error signal is  $\mathbf{e}_i = \mathbf{d}_i - U_i \mathbf{w}_{i-1}$ ; the superscript asterisk denotes complex conjugation.

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

$$d_i = \mathbf{x}_i \mathbf{w}^\circ + v_i \quad (2)$$

where  $\mathbf{w}^\circ \in \mathbb{R}^{L \times 1}$  is an unknown vector to be estimated.  $v$  is a zero mean Gaussian noise sequence.  $\mathbf{x}$  and  $v$  are independent.

Shin *et al.* [12] proposed the optimal variable step-size APA (VS-APA) in which (1) can be written as

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i \quad (3)$$

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

$$\mathbf{p}_i \triangleq U_i^* (U_i U_i^*)^{-1} U_i \tilde{\mathbf{w}}_{i-1} \quad (4)$$

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

$$E[\mathbf{p}_i] = E[U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i] \quad (5)$$

where  $E(\cdot)$  is the expectation operator. Shin *et al.* [12] selected the optimal adaptive filter as the minimizer of  $\|\mathbf{p}_i\|$ . For this case,

$$\mathbf{p}_i = U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i \quad (6)$$

and can be estimated as follows:

$$\hat{\mathbf{p}}_i = \hat{\mathbf{p}}_{i-1} + (1 - \alpha) \mathbf{p}_i \quad (7)$$

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

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu_i U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i \quad (8)$$

where

$$\mu_i = \mu_{max} \frac{\|\hat{\mathbf{p}}_i\|^2}{\|\hat{\mathbf{p}}_i\|^2 + C} \quad (9)$$

and  $\|\cdot\|$  denotes the Euclidean norm of a vector. For a positive constant,  $C$  is related to  $\frac{1}{\alpha} Tr\{E[(U_i U_i^*)^{-1}]\}$ , which can be approximated as  $K/SNR$ . Here the  $Tr(\cdot)$  is the trace of a matrix. When  $\|\hat{\mathbf{p}}_i\|^2$  is large,  $\mathbf{w}_i$  is far from  $\mathbf{w}^\circ$  and  $\mu_i$  is close to  $\mu_{max}$ ; when  $\|\hat{\mathbf{p}}_i\|^2$  is small,  $\mathbf{w}_i$  approaches  $\mathbf{w}^\circ$  and  $\mu_i$  is close to zero.

### B. Optimal Variable Step Size APA with Forgetting Factor

In this section, a variable step size APA with a forgetting factor  $\alpha$  is proposed. The motivation is similar as that of the RLS algorithm. Choosing a value for  $\alpha$  that is less than one introduces memory into the the structure of matrix  $U_i$ , because such a weighting scheme would assign relatively larger weights to recent input samples and smaller weights to input samples in the remote past.

One entry of the input matrix  $U_i$  is generally denoted as  $x_{i-k-l}$  that is in the  $(k+1)^{th}$  row and  $(l+1)^{th}$  column of  $U_i$  for  $k=0, \dots, K-1$  and  $l=0, \dots, L-1$ .

By introducing a forgetting factor  $\alpha$ , where  $0 < \alpha \leq 1$ , and multiplying every element in  $U_i$  by  $\alpha^{k+l}$ , a new observation matrix  $Q_i$  is introduced where its elements are given by

$$q_{i-k-l} = \alpha^{k+l} x_{i-k-l}, \quad (10)$$

where  $q_{i-k-l}$  denotes the element in row  $k+1$  and column  $l+1$  of  $Q_i$  for  $k=0, \dots, K-1$  and  $l=0, \dots, L-1$ .

For simplicity, by decomposing the factor  $\alpha^{k+l}$  into two terms, (10) can be rewritten as follows:

$$q_{i-k-l} = \alpha^k x_{i-k-l} \alpha^l. \quad (11)$$

Then (11) can be written in matrix form as

$$Q_i = \Lambda^{(K)} U_i \Lambda^{(L)}, \quad (12)$$

where  $\Lambda^{(K)} = \text{diag}(\alpha^0, \alpha^1, \dots, \alpha^{K-1})$  and  $\Lambda^{(L)} = \text{diag}(\alpha^0, \alpha^1, \dots, \alpha^{L-1})$ . Note that  $\Lambda^{(K)}$  can be considered as a column operation on matrix  $Q_i$  that assigns larger weight to recent regressors. Whereas,  $\Lambda^{(L)}$  is for row operation on  $Q_i$  which assigns larger weights to recent element of each regressor.

In this way, the filter has a tracking mechanism that can track slow statistical variations of the measured data. Since the weighting scheme gives recent data more significance, the algorithm is sensitive to system dynamics [2].

Two algorithms are preliminarily proposed by replacing  $U_i$  by  $Q_i$  in (8) and (6) as follows:

#### Algorithm 1:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu_i Q_i^* (Q_i Q_i^*)^{-1} \mathbf{e}_i \quad (13a)$$

$$\mu_i = \mu_{max} \frac{\|\hat{\mathbf{h}}_i\|^2}{\|\hat{\mathbf{h}}_i\|^2 + C} \quad (13b)$$

$$\hat{\mathbf{h}}_i = \alpha \hat{\mathbf{h}}_{i-1} + (1 - \alpha) \mathbf{h}_i \quad 0 \leq \alpha < 1 \quad (13c)$$

#### Algorithm 2:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu_i U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i \quad (14a)$$

$$\mu_i = \mu_{max} \frac{\|\hat{\mathbf{h}}_i\|^2}{\|\hat{\mathbf{h}}_i\|^2 + C} \quad (14b)$$

$$\hat{\mathbf{h}}_i = \alpha \hat{\mathbf{h}}_{i-1} + (1 - \alpha) \mathbf{h}_i \quad 0 \leq \alpha < 1 \quad (14c)$$

where in both algorithms

$$\mathbf{h}_i \triangleq Q_i^* (Q_i Q_i^*)^{-1} \mathbf{e}_i \quad (15)$$

As seen, the only difference between these two algorithms is that, in the Algorithm 2,  $U_i$  is only replaced by  $Q_i$  during the error evaluation phase (6), not during the weights updating phase (14a). The Algorithm 1 is not recommended because: replacing  $U_i$  by  $Q_i$  during the both phases introduces possible convergence instability. This is discussed in details in the section III.

The algorithm 2 is the proposed variable step size affine projection algorithm with a forgetting factor (VS-APA-FF). A special case is the variable step size NLMS with forgetting factor (VS-NLMS-FF). This can be obtained by setting  $K=1$  when the input matrix  $U_i$  is a row vector and the forgetting factor processing is implemented only in the row direction as follows:

$$Q_i = U_i \Lambda^{(L)}. \quad (16)$$

### III. ALGORITHM STABILITY ANALYSIS

The convergence behavior of the general form of APA class had been investigated by Sankaran and Beex [9]. In [13], Ikeda *et al.* derived the convergence rates of the APA family for both white and the coloured signals. As proved in [9],  $\mu \in (0, 2)$  is a necessary and sufficient condition for the APA class to be stable; when  $\mu = 1$ , the algorithm conducts the fastest convergence. Therefore, the proposed variable step size APA with forgetting factor (VS-APA-FF), (Algorithm 2), guarantees stability as long as the initial step size is within  $(0, 2)$  [9].

To investigate the effect of introducing the forgetting scheme on the weight update, the convergence behavior of the traditional fixed step size APA [14] is firstly investigated:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i \quad (17)$$

Subtracting both sides of (17) from  $\mathbf{w}^\circ$  leads to the following weight-error recursion:

$$\tilde{\mathbf{w}}_i = [I - \mu U_i^* (U_i U_i^*)^{-1} U_i] \tilde{\mathbf{w}}_{i-1} \quad (18)$$

where  $I$  is an  $L \times L$  identity matrix.

The singular value decomposition (SVD) of the input matrix  $U$  is given by

$$U = R \Sigma V^* \quad (19)$$

where  $R$  and  $V$  are  $K \times K$  and  $L \times L$  unitary matrices respectively, and the  $K \times L$  matrix  $\Sigma$  is defined as

$$\Sigma = \begin{bmatrix} \mathbf{S} & \mathbf{0} \end{bmatrix} \quad (20)$$

where  $\mathbf{S}$  is a diagonal matrix,

$$\mathbf{S} = \text{diag}(s_1, s_2, \dots, s_K). \quad (21)$$

and the  $s_k$ 's which are called the eigenvalues of  $U$  are positive square roots of the eigenvalues of  $U_i U_i^*$  and are usually ordered in a decreasing manner,  $s_1 \geq s_2 \geq \dots \geq s_K > 0$ .

Substituting (19) into (18) gives

$$\tilde{\mathbf{w}}_i = V_i [I - \mu \Sigma_i^* (\Sigma_i \Sigma_i^*)^{-1} \Sigma_i] V_i^* \tilde{\mathbf{w}}_{i-1}. \quad (22)$$

Multiplying on the left by  $V_i^*$  then gives

$$V_i^* \tilde{\mathbf{w}}_i = [I - \mu \Sigma_i^* (\Sigma_i \Sigma_i^*)^{-1} \Sigma_i] V_i^* \tilde{\mathbf{w}}_{i-1}. \quad (23)$$

In this result, let define the rotational coefficient error vector as  $\check{\mathbf{w}}_i = V_i^* \tilde{\mathbf{w}}_i$  and write

$$\check{\mathbf{w}}_i = [I - \mu \Sigma_i^* (\Sigma_i \Sigma_i^*)^{-1} \Sigma_i] \check{\mathbf{w}}_{i-1} \quad (24)$$

Now each element of  $\check{\mathbf{w}}_i$  converges at its own rate, which is determined by the corresponding eigenvalue of the transformation matrix  $T_i = [I - \mu \Sigma_i^* (\Sigma_i \Sigma_i^*)^{-1} \Sigma_i]$ . Suppose the  $k^{th}$  eigenvalue of  $T_i$  is  $\lambda_k$ ,  $k = 1, \dots, K$ , then the  $k^{th}$  element of  $\check{\mathbf{w}}_i$  has converge rate as

$$\lambda_k = 1 - \mu \frac{s_k}{s_k} = 1 - \mu \quad (25)$$

Conclusions: 1.  $\mu \in (0, 2)$  is the sufficient and necessary condition for APA class algorithm convergence; 2. if  $\mu = 1$ , the algorithm converges at the highest rate. These are consistent with the conclusions drawn by Sankaran and Beex [9].

Next, let investigate the convergence behavior of the APA algorithm if a forgetting weighted process is introduced into the weight updating phase (17). After being weighted, the singular value decomposition of the input matrix  $Q$  is (here the time index is ignored for simplicity, without loss of generality):

$$\begin{aligned} Q &= \Lambda^{(K)} U \Lambda^{(L)} = \Lambda^{(K)} [R \Sigma V^T] \Lambda^{(L)} \\ &= R [\Lambda^{(K)} \Sigma \Lambda^{(L)}] V^T = R \Sigma' V^T \end{aligned} \quad (26)$$

where

$$\Sigma' = \Lambda^{(K)} \Sigma \Lambda^{(L)} \quad (27)$$

$\Sigma' \in \mathfrak{R}^{K \times L}$  and

$$[\Sigma']_{k,l} = \begin{cases} \lambda_k & k = l \\ 0 & k \neq l \end{cases} \quad (28)$$

for  $k = 1, 2, \dots, K$ ;  $l = 1, 2, \dots, L$ .

The (17) becomes

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu Q_i^* (Q_i Q_i^*)^{-1} \mathbf{e}_i \quad (29)$$

Subtracting both sides of (29) from  $\mathbf{w}^\circ$

$$\tilde{\mathbf{w}}_i = (I - P_i') \tilde{\mathbf{w}}_{i-1} \quad (30)$$

where  $P_i' = \mu Q_i^* (Q_i Q_i^*)^{-1} U_i$ . The equation (24) changes to

$$\check{\mathbf{w}}_i = [I - \mu \Sigma_i^* (\Sigma_i' \Sigma_i'^*)^{-1} \Sigma_i'] \check{\mathbf{w}}_{i-1} \quad (31)$$

the weighted transformation matrix becomes  $T_i' = [I - \mu \Sigma_i^* (\Sigma_i' \Sigma_i'^*)^{-1} \Sigma_i']$ . Suppose the  $k^{th}$  eigenvalue of  $T_i'$  is  $\lambda_k'$ ,  $k = 1, \dots, K$ , then the  $k^{th}$  element of  $\check{\mathbf{w}}_i$  has converge rate as

$$\lambda_k' = 1 - \mu \frac{s_k}{s_k} = 1 - \mu \frac{s_k}{s_k} = 1 - \mu^{2(1-k)} \quad (32)$$

when  $\lambda_k'$  and  $k$  are chosen such that  $\frac{2}{\mu} < 2(1-k)$ , certain converge rate  $|\lambda_k'| > 1$  which means the algorithm is unstable.

Conclusion: The APA algorithm may be unstable if the forget-weighting process is introduced into the weight updating phase (Algorithm 1). Although parameters can be chosen carefully to assure  $|\lambda_k'| \leq 1$ , applying the proposed forgetting process on the weight updating phase is not recommended.

In the Section IV, the proposed algorithm VS-APA-FF is further upgraded.

#### IV. REGULARIZATION OF THE WEIGHTED PROJECTION

According to equation 6 and 15, the accuracy of the error evaluation depends on the condition number (define by the ratio of the maximum eigenvalue and the minimum eigenvalue) of the projection matrix ( $U$  or  $Q$ ). For a projection matrix with large condition number, even a small amplitude of noise in the error signal will be amplified which makes the  $\mathbf{h}$  or  $\mathbf{p}$  relatively noisy. Thus the VS-APA and VS-APA-FF algorithms adopt a smoothing function, in the form of (7), to alleviate this problem, however, with the cost loss of error signal fidelity, which sacrifices convergence speed and/or misadjustment.

In the previously proposed algorithm (Algorithm 2, denoted as VS-APA-FF), the weighted projection matrix  $Q$  has a larger condition number compared with the non-weighted  $U$ . This can be illustrated as follows:

The condition number of  $U$  is:

$$\text{cond} U = \max \lambda / \min \lambda = s_1 / s_K \quad (33)$$

from (12), the SVD of the weighted input matrix  $Q$  is:

$$\begin{aligned} Q &= \Lambda^{(K)} U \Lambda^{(L)} = \Lambda^{(K)} [R \Sigma V^*] \Lambda^{(L)} \\ &= R [\Lambda^{(K)} \Sigma \Lambda^{(L)}] V^* \\ &= R \Sigma' V^* \end{aligned} \quad (34)$$

where  $\Sigma'$  is a  $K \times L$  matrix with all zero entities except  $[\Sigma']_{j,j} = \lambda_j$ ,  $j = 1, 2, \dots, K$ . The condition number of  $Q$  is:

$$\text{cond} Q = \lambda_1 / \lambda_K = \lambda_1 / \lambda_K = \lambda_1^{2(1-K)} \text{cond} U \quad (35)$$

which illustrates the increased condition number by a factor of  $2^{(1-K)}$ . According to equation (15), this means that the estimated  $\mathbf{h}$  is more sensitive to the noise in error signal.

In stead of solving this problem using a smoothing function as the VS-APA and VS-APA-FF, this problem is addressed to use a Tikhonov regularization approach, under which (15) becomes:

$$\mathbf{h}_i = Q_i^*(Q_i Q_i^* + \lambda^2 I)^{-1} \mathbf{e}_i \quad (36)$$

where  $I$  is the identity matrix on the assumption that each element in  $\mathbf{h}_i$  has identical variance;  $\lambda$  is a hyperparameter to control the amount of regularization. The modified algorithm becomes (denoted as VS-APA-FF-REGU):

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \mu_i U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i \\ \mu_i &= \mu_{max} \frac{\|\hat{\mathbf{h}}_i\|^2}{\|\hat{\mathbf{h}}_i\|^2 + C} \end{aligned} \quad (37)$$

Note that the smoothing functions (such as (7), (14c)) is no longer needed (effectively  $\lambda = 0$ ), since the regularization process accomplishes this function.

## V. SIMULATION RESULTS

The proposed algorithm is validated by simulations of a system identification model (Fig. 1).

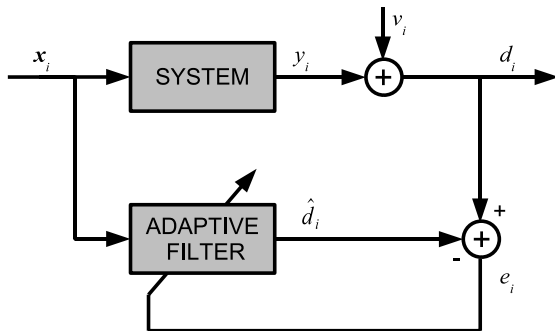


Fig. 1. System identification model

Following the test framework of Shin *et al.* [12], 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 coloured 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 first order filter,  $G_1(z) = 1/(1 - 0.9z^{-1})$  or a 4<sup>th</sup> 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  $y_i$  ( $y_i = \mathbf{x}_i \mathbf{w}^o$ ) and the SNR of the measurement signal is calculated by

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

The simulation results are obtained from the ensemble average of 100 independent trials (ten thousand iterations each), with  $\mu_{max} = 1$  and a smoothing factor  $\lambda = 0.99$  for VS-APA and VS-APA-FF. The convergence is evaluated by Mean Square Deviation (MSD) which is calculated by

$$E(\|\tilde{\mathbf{w}}_i\|^2) = E(\|\mathbf{w}^o - \mathbf{w}_i\|^2)$$

Figs.4-5 illustrate effects of different forgetting factors of VS-APA-FF through G1 and G2 colorizations, respectively. The special case of VS-APA-FF is VS-NLMS-FF (when  $K=1$ ) which is illustrated in fig. 4(a)4(c) and 5(a). (Note that when  $\lambda = 1$ , the VS-APA-FF becomes the original VS-APA. Therefore, VS-APA can be regarded as a special case of VS-APA-FF).

From figs.4-5, following observations can be made:

- VS-APA-FF outperforms VS-APA by applying a temporal weighted projection matrix  $Q$ .
- in figs.4-5, some adaptation curves are noisy (e.g. figs.4(b)and figs.3(d), when  $K$  is large and  $\lambda$  is small), which also illustrated the increased condition number of the projection matrix.
- noise color affects adaptation performance. (This applies for all APA class)

Using experimental conditions described previously, and  $\lambda = 0.5$ ,  $\mu = 1$ , simulation comparisons between VS-APA, VS-APA-FF and the regularized version VS-APA-FF-REGU are illustrated by figure 4 (noise G2) and figure 5 (noise G1). For some cases, when the update matrix of VS-APA-FF becomes severely ill-conditioned (Figure 4(a)5(a)5(b)5(c)) and experiences unsatisfactory convergence, the VS-APA-FF-REGU can still converge quickly with low misadjustment. Therefore, the conclusion is that VS-APA-FF-REGU is a good complement for VS-APA-FF, when the forgetting processed input matrix is close to singular.

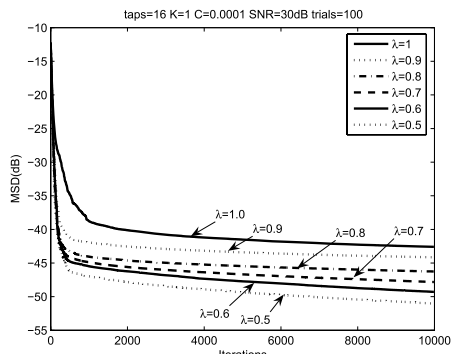
On the other hand, by using a larger forgetting factor (e.g.,  $\lambda = 0.9$ ) when the VS-APA-FF has less singularity problem, the VS-APA-FF-REGU is still advantageous over the VS-APA and the VS-APA-FF, illustrated in Figure (6,7).

## VI. CONCLUSIONS

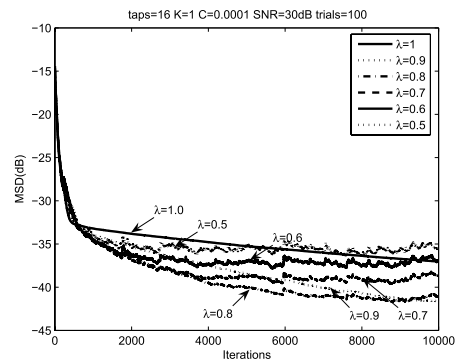
This paper presents an upgraded variable step size APA algorithm, VS-APA-FF, and its regularized version VS-APA-FF-REGU. With a projection matrix processed with a forgetting factor, VS-APA-FF obtains upgraded convergence performance in terms of higher convergence rate and lower misadjustment. However, due to the singularity becomes worse after been weighted, VS-APA-FF is incapable of some cases, such as the APA window size  $K$  is large and a small forgetting factor is used. The Tikhonov regularization is used to overcome the deteriorated singularity problem of the processed input matrix. The regularized algorithm is more stable and converges better than previous algorithms.

## REFERENCES

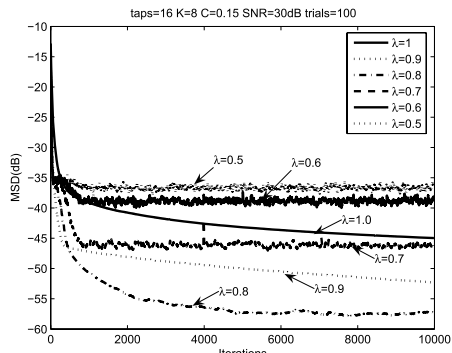
- [1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [2] S. Haykin, *Adaptive Filter Theory*. 4th ed. Upper Saddle River, NJ: Prentice Hall, 2002.



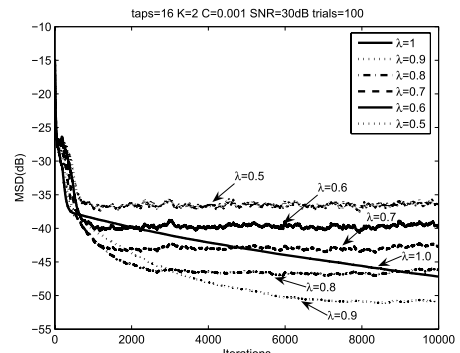
(a)



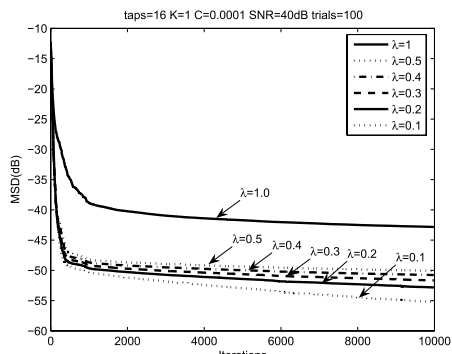
(a)



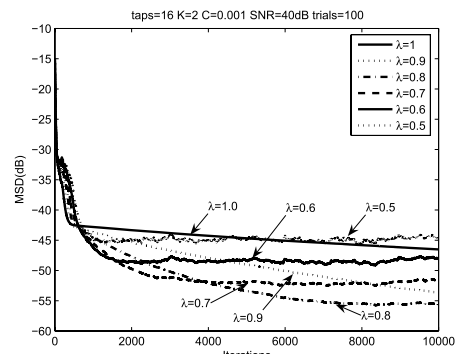
(b)



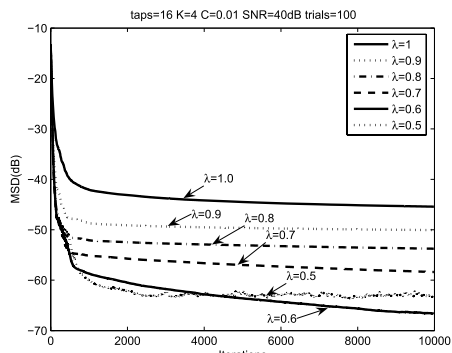
(b)



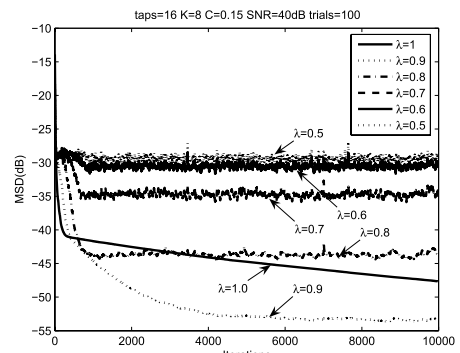
(c)



(c)



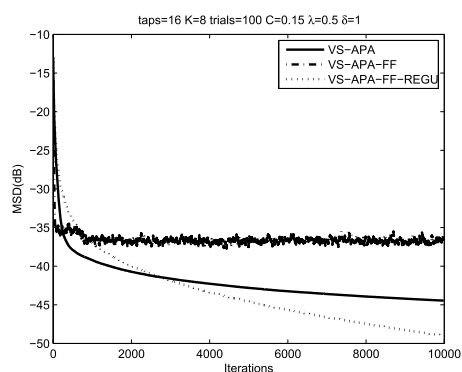
(d)



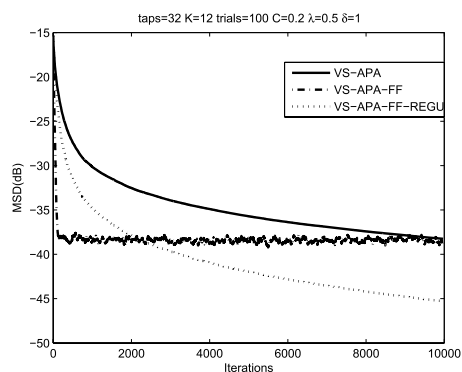
(d)

Fig. 2. Effects of  $\lambda$  in VS-APA-FF (VS-APA when  $\lambda = 1.0$ ), G2 colorization. (a) VS-NLMS-FF K=1, taps=16, C=0.0001, SNR=30dB; (b) K=8, taps=16, C=0.15, SNR=30dB; (c) VS-NLMS-FF K=1, taps=16, C=0.0001, SNR=40dB; (d) K=4, taps=16, C=0.01, SNR=40dB;

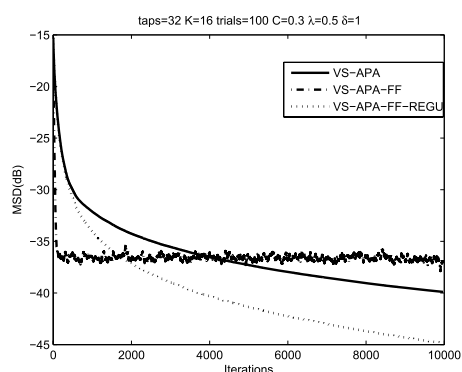
Fig. 3. Illustration of different forgetting factors for VS-APA-FF (VS-APA when  $\lambda = 1.0$ ), G1 colorization. (a) VS-NLMS-FF K=1, taps=16, C=0.0001, SNR=30dB; (b) K=2, taps=16, C=0.001, SNR=30dB; (c) K=2, taps=16, C=0.001, SNR=40dB; (d) K=8, taps=16, C=0.15, SNR=40dB;



(a)

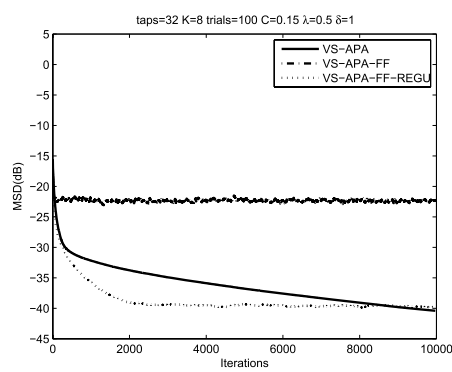


(b)

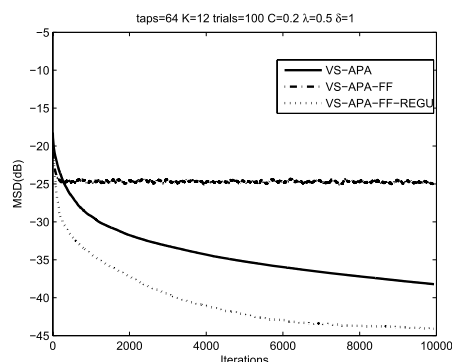


(c)

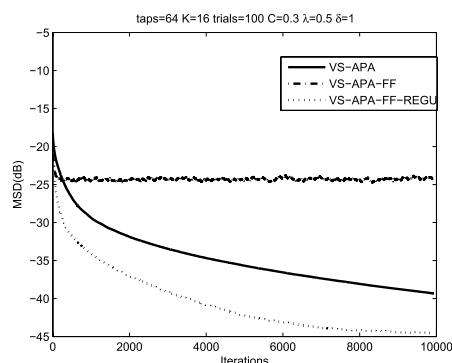
Fig. 4. Comparisons among VS-APA, VS-APA-FF, and VS-APA-FF-REGU, G2 colorization.  $\lambda = 0.5$ . (a)  $K=8$ , taps=16,  $C=0.15$ ; (b)  $K=12$ , taps=32,  $C=0.2$ ; (c)  $K=16$ , taps=32,  $C=0.3$



(a)



(b)

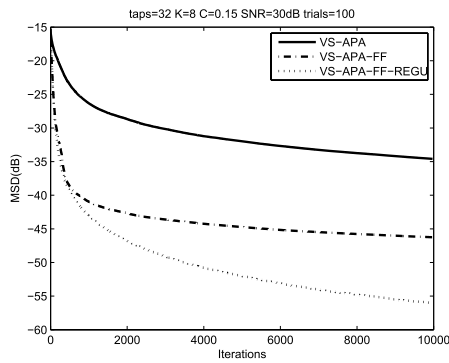


(c)

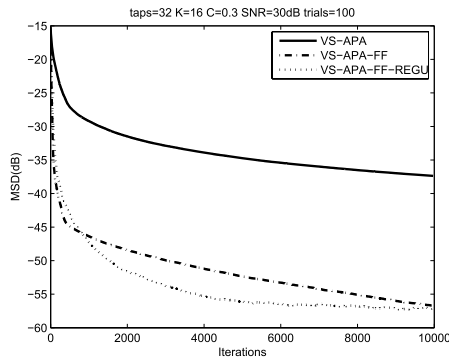
Fig. 5. Comparisons among VS-APA, VS-APA-FF, and VS-APA-FF-REGU, G1 colorization.  $\lambda = 0.5$ . (a)  $K=8$ , taps=32,  $C=0.15$ ; (b)  $K=12$ , taps=64,  $C=0.2$ ; (c)  $K=16$ , taps=64,  $C=0.3$

[3] J. Nagumo and A. Noda, "A learning method of system identification", *IEEE Trans. Automat. Control*, 12 (1967) 282-287.  
 [4] S.J. Orfanidis, *Optimum Signal Processing, An Introduction*. MacMillan, New York, 1985.  
 [5] 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.  
 [6] F. Yu and M. Bouchard, "Recursive least-squares algorithms with good numerical stability for multichannel active noise control" vol. 5, pp. 3221-3224. ICASSP 2001  
 [7] S.L. Gay and S. Tavathia, "The fast affine projection algorithm", in Proc. ICASSP, pp. 3023-3026, May 1995.  
 [8] H.C. Shin and A.H. Sayed, "Transient behavior of affine projection algorithms", *ICASSP*, VI 353-356, 2003.  
 [9] S.G. Sankaran and A.A. Beex, "Convergence behavior of affine pro-

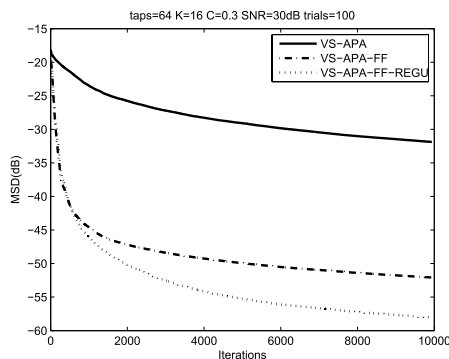
jection algorithms," *IEEE Trans. Signal Processing*, vol. 48, no. 4, pp. 1086-1096, 2000  
 [10] R.W. Harris, D. Chabries, and F. Bishop, "A variable step (VS) adaptive filter algorithm". *IEEE Trans. Acoust. Speech Signal Process.* vol.34, i2. 309-316, 1986  
 [11] A. Mader, H. Puder, and G.U. Schmidt, "Step-size control for acoustic echo cancellation filters-An overview", *Signal Process.*, vol. 80, pp. 1697-1719, Sept 2000.  
 [12] H.C. Shin, A.H. Sayed and W.J. Song, "Variable step-size NLMS and affine projection algorithms", *IEEE Signal Proc. Letters*, Vol.11, No.2, pp. 132-135, Feb 2004.  
 [13] K. Ikeda, "Convergence analysis of block orthogonal projection and affine projection algorithms", *Signal Processing*, 82, 491-496, 2002.  
 [14] S.L. Gay, "Affine projection algorithm", in *Least-Mean-Square Adaptive Filters*, ISBN 0-471-21570-8, Ed. S. Haykin and B. Widrow, Wiley interscience, 2003.



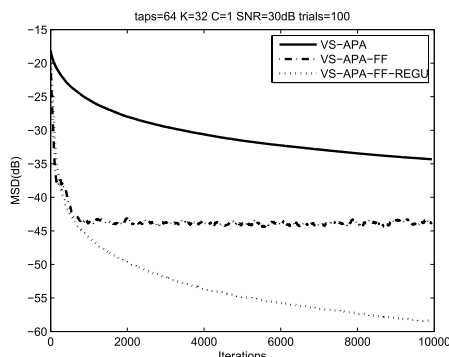
(a)



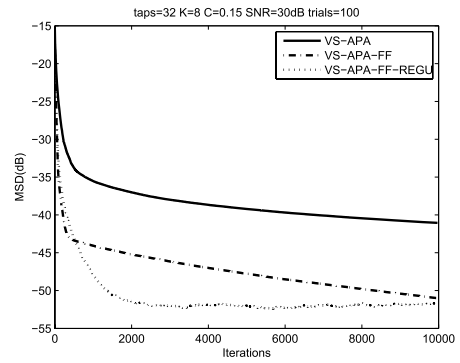
(b)



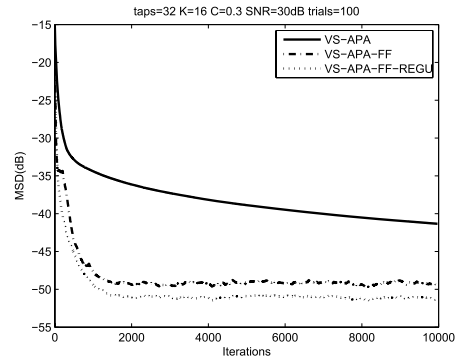
(c)



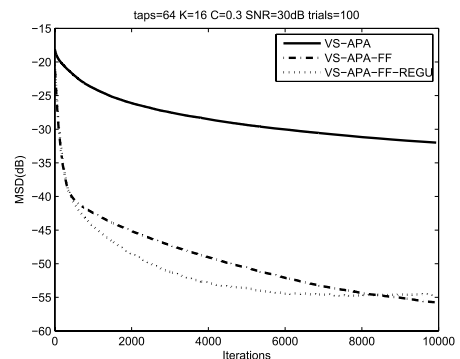
(d)



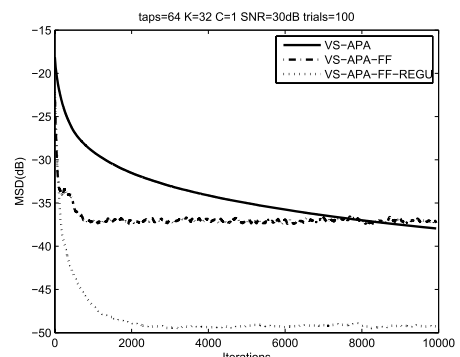
(a)



(b)



(c)



(d)

Fig. 6. Comparisons between algorithms: *VS-APA*, *VS-APA-FF* and *VS-APA-FF-REGU*. G2 colorization,  $\lambda = 0.9$ ,  $\delta = 1$ , SNR=30dB. (a) K=8, taps=32, C=0.15; (b) K=16, taps=32, C=0.3; (c) K=16, taps=64, C=0.3; (d) K=32, taps=64, C=1;

Fig. 7. Comparisons between algorithms: *VS-APA*, *VS-APA-FF* and *VS-APA-FF-REGU*. G1 colorization,  $\lambda = 0.9$ ,  $\delta = 1$ , SNR=30dB. (a) K=8, taps=32, C=0.15; (b) K=16, taps=32, C=0.3; (c) K=16, taps=64, C=0.3; (d) K=32, taps=64, C=1;

**Tao Dai** received the B.A.Sc. degree in electrical engineering in 1995 from Zhejiang University, Hangzhou, P.R.China, and the M.A.Sc. degrees in electrical and computer engineering in 2003, from University of Windsor, Windsor, Canada.

At this time, he is a Ph.D. candidate in the Carleton University, Ottawa, Canada. His research interests are primarily in the area of medical imaging, image processing, adaptive signal processing and medical instrumentation.

**Andy Adler** is associate professor and Canada Research Chair (Tier II) in biomedical engineering in Systems and Computer Engineering at Carleton University in Ottawa, Canada. His research interests are in biometrics imaging and security systems, and the associated algorithms, measurement devices, and privacy and security aspects; and development of non-invasive biomedical measurement technologies and sensors, including the medical image and signal processing algorithms. Previously, he taught and researched at the University of Ottawa, and worked in senior technology positions in biometric security applications.

**Behnam Shahrava** received the Ph.D. degree in Electrical Engineering from the University of Waterloo, Waterloo, ON, Canada, in 1998. He is currently an assistant professor with the Department of Electrical and Computer Engineering, University of Windsor, Canada. His research interests are in the areas of advanced statistical and adaptive signal processing and their applications, primarily in wireless communications.