Missouri University of Science and Technology
Scholars' Mine

Electrical and Computer Engineering Faculty Research & Creative Works

Electrical and Computer Engineering

01 Apr 2007

Variable Regularized Fast Affine Projections

Steven L. Grant
Missouri University of Science and Technology, sgrant@mst.edu

Asif Iqbal Mohammad

Deepak Challa

Follow this and additional works at: https://scholarsmine.mst.edu/ele_comeng_facwork

Part of the Electrical and Computer Engineering Commons

Recommended Citation

S. L. Grant et al., "Variable Regularized Fast Affine Projections," *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, Institute of Electrical and Electronics Engineers (IEEE), Apr 2007.
The definitive version is available at https://doi.org/10.1109/ICASSP.2007.366623

This Article - Conference proceedings is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

VARIABLE REGULARIZED FAST AFFINE PROJECTIONS

Deepak Challa, Steven L. Grant, Asif Mohammad*

University of Missouri-Rolla
1870 Miner Circle, Rolla, MO 65409
[dkcy92, sgrant, ammq2]@umr.edu

ABSTRACT

This paper introduces a variable regularization method for the fast affine projection algorithm (VR-FAP). It is inspired by a recently introduced technique for variable regularization of the classical, affine projection algorithm (VR-APA). In both algorithms, the regularization parameter varies as a function of the excitation, measurement noise, and residual error energies. Because of the dependence on the last parameter, VR-APA and VR-FAP demonstrate the desirable property of fast convergence (via a small regularization value) when the convergence is poor and deep convergence/immunity to measurement noise (via a large regularization value) when the convergence is good. While the regularization parameter of APA is explicitly available for on-line modification, FAP's regularization is only set at initialization. To overcome this problem we use noise-injection with the noise-power proportional to the variable regularization parameter. As with their fixed regularization versions, VR-FAP is considerably less complex than VR-APA and simulations verify that they have the very similar convergence properties.

Index Terms— FAP, APA, regularization, adaptive filter, affine projections

1. INTRODUCTION

A wide variety of adaptive filters are now available in the signal processing community. Each has their advantages and disadvantages. The affine projection algorithm (APA) [1,2] has received considerable attention over the past 15 years or so because it's attributes provide a nice compromise between normalized least mean squares (NLMS) [3] and fast recursive least squares (FRLS) [4]. NLMS is computationally quite efficient and numerically stable, but converges rather slowly when a colored excitation signal is used. FRLS, is less computationally efficient and somewhat difficult to stabilize numerically, but has fast convergence for colored excitation.

The NLMS coefficient update method may be viewed as a one-dimensional affine projection in the parameter space. Under this view, APA is a generalization of NLMS in that it performs an N-dimensional affine projection each sample period [2,5]. When N is greater than or equal to the order of the source model that creates the excitation signal, APA's convergence properties are roughly the same as FRLS's [2]. Depending on the exact implementation, APA generally enjoys a much greater degree of numerical stability than FRLS algorithms – even the so-called stabilized ones. However, depending on N, APA's computational complexity can be higher than FRLS. To address this defect, the fast affine projection (FAP) [2,5,6,7] was introduced in the early 1990's. FAP reduced the computational complexity to roughly that of NLMS.

As the affine projection order, N, increases from one, a simple scalar inversion of the excitation vector's norm in NLMS becomes an N-by-N excitation sample covariance matrix inversion in APA. Often, with highly colored noise excitation, this sample covariance is ill-conditioned and to prevent undue noise amplification, a regularization parameter, δ is added to the matrix diagonal prior to inversion. In [8] a method for dynamically estimating an optimal regularization parameter for APA was described and the subsequent improvement in convergence for stationary excitation signals was demonstrated. In this paper we re-derive the optimal regularization for FAP. Traditionally, FAP's regularization is implemented by way of an initialization parameter that remains fixed thereafter; we overcome this static-regularization problem by using only a very small initial regularization and then using the noise-injection technique to vary the regularization as determined by the VR method.

This paper is arranged as follows: section 2 is a brief review of APA and FAP, section 3 presents the derivation of the variable regularization parameter for FAP and the use of the noise injection method. Finally, simulation results are presented in section 4 and conclusions in section 5.

2. REVIEW OF APA AND FAP

This section presents a brief review of APA and a review of FAP to the extent that the VR algorithm for it may be derived. For a complete derivation of FAP, see [2,5,6,7].

2.1 The Affine Projection Algorithm

The affine projection algorithm, in a relaxed and regularized form, is defined as

$$\mathbf{e}(n) = \mathbf{s}(n) - \mathbf{X}^T(n)\mathbf{h}(n-1) \tag{1}$$

$$\boldsymbol{\varepsilon}(n) = [\mathbf{X}^T(n)\mathbf{X}(n) + \delta\mathbf{I}]^{-1}\mathbf{e}(n) \tag{2}$$

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{X}^T(n)\boldsymbol{\varepsilon}(n) \tag{3}$$

* formerly known as Steven L. Gay

1-4244-0728-1/07/$20.00 ©2007 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on January 23, 2009 at 15:20 from IEEE Xplore. Restrictions apply.

The excitation signal matrix, $\mathbf{X}(n)$, is L by N and has the structure,

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \cdots, \mathbf{x}(n-N+1)] \tag{4}$$

where the $\mathbf{x}(n) = [x(n), \cdots, x(n-L+1)]^T$. The adaptive tap weight vector is $\mathbf{h}(n) = [h_0(n), \cdots, h_{L-1}(n)]^t$, where $h_i(n)$ is the i^{th} coefficient at sample period n. The N-length vector, $\mathbf{e}(n)$, consists of background noise and residual echo left uncancelled by the echo canceller's L-length adaptive tap weight vector, $\mathbf{h}(n)$. The N-length vector, $s(n)$, is the system output consisting of the response of the echo path impulse response, $\mathbf{h}_{ep}$ to the excitation and the additive system noise, $\mathbf{y}(n)$,

$$\mathbf{s}(n) = \mathbf{X}^T(n)\mathbf{h}_{ep} + \mathbf{y}(n) \tag{5}$$

The scalar δ is the regularization parameter for the sample autocorrelation matrix inverse used in (2), the calculation of the N-length normalized residual echo vector, $\boldsymbol{\varepsilon}(n)$. Where $\mathbf{X}^T(n)\mathbf{X}(n)$ may have eigenvalues close to zero, creating problems for the inverse, $\mathbf{X}^T(n)\mathbf{X}(n) + \delta\mathbf{I}$ has δ as its smallest eigenvalue which, if large enough, yields a well behaved inverse. The step-size parameter μ is the relaxation factor. As in NLMS, the algorithm is stable for $0 \le \mu < 2$.

If we define the coefficient error vector as $\Delta\mathbf{h}(n) = \mathbf{h}_{ep} - \mathbf{h}(n)$ then the error vector, $\mathbf{e}(n)$ may be written

$$\mathbf{e}(n) = \mathbf{X}^T(n)\Delta\mathbf{h}(n) + \mathbf{y}(n) \tag{6}$$

Note that if N is set to one, relations(1), (2), and (3) reduce to the familiar NLMS algorithm. Thus, APA is a generalization of NLMS.

2.2 The Fast Affine Projection Algorithm

The complexity of APA is $2LN + K_{inv}N^2$ multiplies per sample period, where K_{inv} is a constant associated with the complexity of the inverse required in (2). FAP performs a complete N-dimensional APA update each sample period with $2L + O(N)$ multiplies per sample [2,5,6,7]. The development of FAP involves reducing the computational complexity of each of the steps in equations (1), (2), and (3). For the variable regularization derivation in section 3 we only need to review FAP's computational reduction of equation (3).

2.3 Fast Adaptive Coefficient Vector Calculation

The "trick" used in FAP to reduce the computational complexity of the coefficient update equation for $\mathbf{h}(n)$ is to introduce an alternate coefficient vector, $\hat{\mathbf{h}}(n)$, whose update each sample period consists only of adding a weighted version of the last column of $\mathbf{X}(n)$. This requires only L multiplications as opposed to NL for the update of equation (3). FAP also provides a method for calculating $\mathbf{e}(n)$ from $\hat{\mathbf{h}}(n)$ which is not shown in this paper.

From (3) the APA tap update is,

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mu\mathbf{X}_n\boldsymbol{\varepsilon}(n) \tag{7}$$

One can also express the current echo path estimate, $\mathbf{h}(n)$, in terms of the original echo path estimate, $\mathbf{h}(0)$, and the subsequent $\mathbf{X}(i)$'s and $\boldsymbol{\varepsilon}(i)$'s,

$$\mathbf{h}(n) = \mathbf{h}(0) + \mu\sum_{i=0}^{n-1}\mathbf{X}(n-i)\boldsymbol{\varepsilon}(n-i) \tag{8}$$

Now, expanding the vector/matrix multiplication,

$$\mathbf{h}(n) = \mathbf{h}(0) + \mu\sum_{i=0}^{n-1}\sum_{j=0}^{N-1}\mathbf{x}(n-j-i)\boldsymbol{\varepsilon}_j(n-i) \tag{9}$$

Assuming that $\mathbf{x}(n) = 0$ for $n \le 0$, (9) can be rewritten as,

$$\mathbf{h}(n) = \mathbf{h}(0) + \mu\sum_{k=0}^{N-1}\mathbf{x}(n-k)\sum_{j=0}^{k}\boldsymbol{\varepsilon}_j(n-k+j)$$
$$+\mu\sum_{k=N}^{n-1}\mathbf{x}(n-k)\sum_{j=0}^{N-1}\boldsymbol{\varepsilon}_j(n-k+j). \tag{10}$$

If the first term and the second pair of summations on the right side of (10) are defined as

$$\hat{\mathbf{h}}(n-1) = \mathbf{h}(0) + \mu\sum_{k=N}^{n-1}\mathbf{x}(n-k)\sum_{j=0}^{N-1}\boldsymbol{\varepsilon}_j(n-k+j) \tag{11}$$

and the first pair of the summations in (10) are recognized as a vector-matrix multiplication,

$$\mathbf{X}(n)\mathbf{E}(n) = \mu\sum_{k=0}^{N-1}\mathbf{x}(n-k)\sum_{j=0}^{k}\boldsymbol{\varepsilon}_j(n-k+j) \tag{12}$$

where,

$$\mathbf{E}(n) = \begin{bmatrix} \varepsilon_0(n) \\ \varepsilon_1(n) + \varepsilon_0(n-1) \\ \vdots \\ \varepsilon_{N-1}(n) + \varepsilon_{N-2}(n-1) \cdots + \varepsilon_0(n-(N-1)) \end{bmatrix} \tag{13}$$

then, (10) can be expressed as

$$\mathbf{h}(n) = \hat{\mathbf{h}}(n-1) + \mu\mathbf{X}(n)\mathbf{E}(n) \tag{14}$$

It is easily seen from (11) that

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) +$$
$$+\mu\mathbf{x}(n-(N-1))\sum_{j=0}^{N-1}\boldsymbol{\varepsilon}_j(n-N+1+j). \tag{15}$$

Or,

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu\mathbf{x}(n-(N-1))E_{N-1}(n) \tag{16}$$

Where $E_{N-1}(n)$ is the last element of $\mathbf{E}(n)$ (note we number the elements from 0 to N-1).

Authorized licensed use limited to: IEEE Xplore. Downloaded on January 23, 2009 at 15:20 from IEEE Xplore. Restrictions apply.

3. VARIABLE REGULARIZATION FOR FAP

We now make the regularization parameter variable with time, explicitly denoting it as $\delta(n)$. In the following derivation we will assume that the relaxation parameter, $\mu = 1$. Similar to the approach in [6] we chose as our optimization criterion that $\delta(n)$ which minimizes the cost function,

$$J = E\left(\left\|\Delta\hat{\mathbf{h}}(n)\right\|^2\right) - E\left(\left\|\Delta\hat{\mathbf{h}}(n-1)\right\|^2\right) \tag{17}$$

where

$$\Delta\hat{\mathbf{h}}(n) = \mathbf{h}_{ep} - \hat{\mathbf{h}}(n) \tag{18}$$

is the FAP coefficient error vector. Applying (18) to (16) we have

$$\Delta\hat{\mathbf{h}}(n) = \Delta\hat{\mathbf{h}}(n-1) - \mathbf{x}(n-N+1)E_N(n-1) \tag{19}$$

From (13), we see that

$$E_{N-1}(n) = \varepsilon_{N-1}(n) + \varepsilon_{N-2}(n-1)\cdots + \varepsilon_0(n-(N-1)) \tag{20}$$

Defining,

$$\mathbf{S}(n) = [\mathbf{X}^T(n)\mathbf{X}(n) + \delta\mathbf{I}]^{-1}, \tag{21}$$

we now make the simplifying assumption that the sample covariance matrix, $\mathbf{X}^T(n)\mathbf{X}(n)$ is fixed and equal to $L\mathbf{R}_x$ where $\mathbf{R}_x$ is the correlation matrix of $\mathbf{x}(n)$. This is a reasonable assumption when N<<L. This implies that

$$\mathbf{S} \approx [L\mathbf{R}_x + \delta\mathbf{I}]^{-1} \tag{22}$$

where we have removed the time index to emphasize the assumption that $\mathbf{S}$ is now non-time-varying. Under these assumptions (2) becomes

$$\boldsymbol{\varepsilon}(n) \approx \mathbf{S}\mathbf{e}(n) \tag{23}$$

Since we have also assumed that the step-size, $\mu = 1$, from [5] we know that FAP sets

$$\mathbf{e}(n) = \left[e(n), \mathbf{0}_{N-1}^T\right]^T \tag{24}$$

where $\mathbf{0}_k$ is an k-length all zero vector. Thus,

$$\boldsymbol{\varepsilon}(n) \approx \mathbf{S}\mathbf{p}_0 e(n) \tag{25}$$

where $\mathbf{p}_k = \left[\mathbf{0}_k^T,\ 1,\ \mathbf{0}_{N-k-1}^T\right]^T \quad 0 \le k \le N-1$. By careful inspection we observe that

$$\varepsilon_{N-1-k}(n-k) \approx \mathbf{p}_{N-1-k}^T \mathbf{S}\mathbf{p}_0 e(n-k). \tag{26}$$

Thus, with some manipulation we can rewrite (20) as

$$E_N(n-1) = \mathbf{p}_0^T \mathbf{S}\breve{\mathbf{e}}^R(n) \tag{27}$$

where the R in the superscript of $\breve{\mathbf{e}}(n)$ denotes the "reverse" of the vector and

$$\breve{\mathbf{e}}(n) = \left[e(n) \quad e(n-1) \quad \cdots \quad e(n-N+1)\right]^T, \tag{28}$$

is a history of the N most recent FAP residual errors – not to be confused with $\mathbf{e}(n)$ of (1) and certainly not that of (24). We can now express (19) as

$$\Delta\hat{\mathbf{h}}(n) = \Delta\hat{\mathbf{h}}(n-1) - \mathbf{x}(n-N+1)\mathbf{p}_0^T\mathbf{S}\breve{\mathbf{e}}^R(n). \tag{29}$$

Using (29) in (17) we write the cost function as,

$$\begin{aligned} J &= E\left(\left\|\mathbf{x}(n-N+1)\right\|^2\left(\mathbf{p}_0^T\mathbf{S}\breve{\mathbf{e}}^R(n)\right)^2\right) \\ &\quad - E\left(2\Delta\hat{\mathbf{h}}^T(n-1)\mathbf{x}(n-N+1)\mathbf{p}_0^T\mathbf{S}\breve{\mathbf{e}}^R(n)\right) \end{aligned} \tag{30}$$

Assuming the excitation signal is white, we have

$$\mathbf{S} \approx \frac{1}{\delta(n) + L\sigma_x^2}\mathbf{I} \tag{31}$$

Then

$$\mathbf{p}_0^T\mathbf{S}\breve{\mathbf{e}}^R(n) = \frac{\breve{e}_0^R(n)}{\delta(n) + L\sigma_x^2} = \frac{e(n-N+1)}{\delta(n) + L\sigma_x^2} \tag{32}$$

Assuming that $\Delta\hat{\mathbf{h}}^T(n-1)$ changes slowly enough that $\Delta\hat{\mathbf{h}}^T(n-1) \approx \Delta\hat{\mathbf{h}}^T(n-N)$ and we are close enough to convergence that $\Delta\hat{\mathbf{h}}^T(n-N) \approx \Delta\mathbf{h}^T(n-N)$, then

$$\Delta\hat{\mathbf{h}}^T(n-1)\mathbf{x}(n-N+1) \approx \Delta\mathbf{h}^T(n-N)\mathbf{x}(n-N+1) \tag{33}$$

Using this in (6) at the sample period $n-N+1$ we have

$$\begin{aligned} &\Delta\hat{\mathbf{h}}^T(n-1)\mathbf{x}(n-N+1) \\ &\qquad \approx y(n-N+1) + e(n-N+1) \end{aligned} \tag{34}$$

Applying (31) and (34) to (30), and assuming

$$\left\|\mathbf{x}(n-N+1)\right\|^2 \approx L\sigma_x^2,$$

The performance index is approximately

$$J \approx \frac{2\sigma_y^2}{\delta(n) + L\sigma_x^2} - \frac{L\sigma_x^2 + 2\delta(n)}{\left(\delta(n) + L\sigma_x^2\right)^2}E\left(e^2(n-N+1)\right) \tag{35}$$

Minimizing J with respect to $\delta(n)$, we find

$$\delta(n) = \frac{L\sigma_y^2\sigma_x^2}{E\left(e^2(n-N+1)\right) - \sigma_y^2}. \tag{36}$$

As in [8] we estimate $E\left(e^2(n-N+1)\right)$ with a time average. In our simulations we find that a time average of length L is sufficient.

The regularization of (36) is applied to FAP via the technique of noise injection. Noise injection is a method of regularizing a signal's covariance by adding Gaussian white noise. The standard deviation of the noise is set to the square root of the desired regularization value. In VR-FAP, this noise is only added in the excitation signal input to the sliding window FRLS part where the calculation of the forward and backward linear predictors and their estimation error energies are calculated [2,5,6,7].

4. SIMULATIONS

Figure 1 shows a comparison of the convergence of FAP, VR-APA and VR-FAP. The excitation signal was white Gaussian noise. The length of the echo path was set to L=512. The projection order, N=2 and the additive noise, $\mathbf{y}(n)$, was set 30dB down from the echo signal. The initial δ for VR-FAP was set

Authorized licensed use limited to: IEEE Xplore. Downloaded on January 23, 2009 at 15:20 from IEEE Xplore. Restrictions apply.

to $(.001)\sigma_x$. For FAP the fixed regularization was $\delta = (5)\sigma_X$. For VR-APA the parameter γ (referred to as δ in [8]) was set to 0.05. VR-FAP performs very well when compared to FAP; the VR-FAP coefficient error goes below -50dB while FAP bottomed-out at -29dB. VR-APA and VR-FAP have similar initial convergence rate but VR-APA only reaches -43dB.

Figure 2 shows the convergence curves of the FAP, VR-APA and VR-FAP for colored noise input. The colored noise was generated using and auto regressive model with one pole at z-0.95, AR1(0.95). The echo path considered was of length, L=512 and the additive noise, $y(n)$, was set to 30dB lower than the echo. The order of the projection, N was 2. δ for VR-FAP and FAP are the same as above. γ for VR-APA is set to 0.08. The performance of the proposed algorithm is significantly improved over FAP. VR-FAP has fast initial convergence rate and finally converges to a value around -32dB where as FAP settles down at around -17dB. For all the simulations the sampling frequency was considered to be 8000Hz.

5. CONCLUSIONS

This paper introduced a variable regularization method for the fast affine projection algorithm (VR-FAP). The regularization parameter varies as a function of the excitation, measurement noise, and residual error energies. Because of the dependence on the last parameter, VR-FAP demonstrates the desirable property of fast initial convergence and deep convergence/immunity to measurement noise. Conventionally, FAP's regularization is only set at initialization. To overcome this we used noise-injection with the noise-power proportional to the variable regularization parameter. As with their fixed regularization versions, VR-FAP is considerably less complex than VR-APA and simulations verify that they both have similar convergence properties.

REFERENCES

[1] K. Ozeki, T. Umeda, "An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Subspace and Its Properties," *Electron. Commun. Japan.*, vol.67-A, no.5, pp.19-27, May 1984.

[2] S.L. Gay, "Affine Projection Algorithms," in *Least-Mean-Square Adaptive Filters*, S. Haykin, B. Widrow, Eds., John Wiley & sons Inc., Chapter 7, pp. 241-291, 2003.

[3] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 2002.

[4] J.M. Cioffi, T. Kailath, "Windowed Fast Transversal Filters Adaptive Algorithms with Normalization," *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-33, No. 3, June 1985.

[5] S.L. Gay, "Fast Projection Algorithms with Application to Voice Excited Echo Cancellers," *Ph.D. Dissertation,* Rutgers University, Piscataway, New Jersey, October, 1994.

[6] S.L. Gay, S. Tavathia, "The Fast Affine Projection Algorithm," in *Proceedings of the IEEE Trans. Acoust., Speech, Signal Process.*, vol.5, pp. 3023-3026, May 1995.

[7] M. Tanaka, Y. Kaneda, S. Makino, J. Kojima, "Fast Projection Algorithm and Its Step Size Control," in *Proceedings of the IEEE Trans. Acoust., Speech, Signal Process.*, vol.2, pp. 945-948, May 1995.

[8] H. Rey, L.R. Vega, S. Tressens, J. Benesty, "Optimum Variable Explicit Regularized Affine Projection Algorithm," in *Proceedings of the IEEE Trans. Acoust., Speech, Signal Process.*, vol.3, pp. 197-200, May 2006.

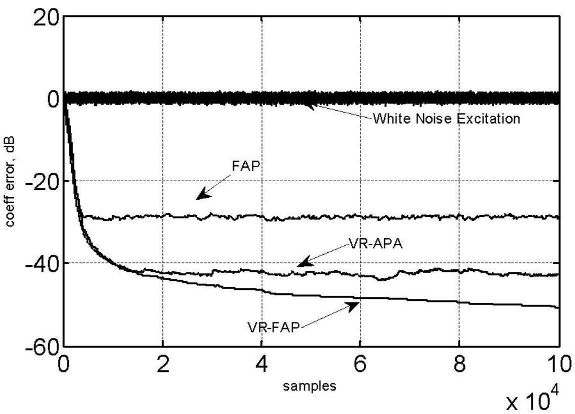

Fig. 1. Coefficient Error (in dB) for white excitation noise, L=512, N=2, SNR = 30dB

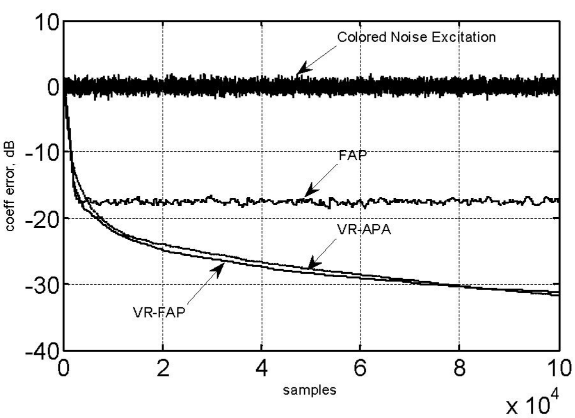

Fig 2. Coefficient Error (in dB) for Colored noise AR1(0.95), L=512, N=2, SNR = 30dB

Authorized licensed use limited to: IEEE Xplore. Downloaded on January 23, 2009 at 15:20 from IEEE Xplore. Restrictions apply.