

MODIFIED GAUSS-SEIDEL AFFINE PROJECTION ALGORITHM FOR ACOUSTIC ECHO CANCELLATION

Felix Albu, Constantine Kotropoulos

Department of Informatics, Aristotle University of Thessaloniki,
Box 451, GR-54124 Thessaloniki, Greece
Emails: {felix,costas}@zeus.csd.auth.gr

ABSTRACT

This paper proposes a robust and stable fast affine projection algorithm based on the Gauss-Seidel method, the so called modified Gauss-Seidel fast affine projection algorithm. The proposed algorithm is generalized for simplified Volterra filters as well. The computational complexity of both the modified Gauss-Seidel fast affine projection algorithm and its generalization for Simplified Volterra filters is derived and their performance for acoustic echo cancellation is assessed.

1. INTRODUCTION

In echo cancellation systems, an adaptive filter algorithm is used to reduce the echo. The echo path is usually modeled by a linear filter. The known normalized Least Mean Squares (NLMS) algorithm has been widely used, but it has a slow asymptotic convergence. The affine projection algorithm (APA) can be considered as a generalization of the NLMS algorithm that provides a much improved convergence speed compared to stochastic gradient descent algorithms, although it is sensitive to high level of noise [1]. It has a performance that rivals the more complex Recursive Least Squares (RLS) algorithms in many situations. However, the Fast Affine Projection (FAP) algorithm proposed in [2-3] suffers from numerical instability, when implemented with an embedded fast RLS algorithm. Other difficulties are its memory requirements and the code overhead. In [3] the relation between Row Action Projections (RAP), APA, and Gauss-Seidel is discussed. Other fast affine projection alternatives have been proposed in [4-6]. However, these variants are suitable for values of the step size close to 1 (e.g. $0.7 < \mu \leq 1$). For such values, these algorithms have a fast convergence, but they exhibit a high sensitivity to noisy inputs. They employ some approximations that degrade the performance the original APA can achieve.

Unlike these FAP alternatives the modified Gauss-Seidel FAP (MGSFAP) algorithm, proposed in this paper, can

employ any step size $\mu \in (0,1]$ like the original FAP [2-3].

Therefore, it is more robust to noisy conditions. Moreover, it is numerically less complex than other similar FAPs for typical values of the projection order.

The outline of the paper is as follows. The MGSFAP algorithm is described in Section 2. A generalization of the proposed algorithm for simplified Volterra filters (SVF), abbreviated as MGSFAP-SVF, is also proposed in a similar way with that in [7]. Moreover, a voice activity detector is proposed in order to tackle the sensitivity of SVF structure. In Section 3, the behavior of MGSFAP and MGSFAP-SVF algorithms for echo cancellation in a double-talk scenario is examined. In the same section, the computational complexity of the proposed algorithms is derived and compared to other commonly used FAP algorithms. Section 4 concludes the paper.

2. THE ALGORITHMS

Henceforth, most of the notations in [2] and [5] are used. Let $x(n)$ be the input signal, $y(n)$ be the desired output signal and $e(n)$ be the output error. We denote by L the filter length, and by N the affine projection order. In the following, $\mathbf{R}(n)$ is the auto-correlation matrix of the input signal, \mathbf{I} is the identity matrix, $\delta(n)$ is a regularization factor that prevents the input auto-correlation matrix from becoming ill-conditioned, μ is the step size. Let $\mathbf{h}(n) = [h_1(n), \dots, h_L(n)]$ be the $L \times 1$ filter coefficient vector. $\mathbf{E}(n)$ is an $N \times 1$ vector, while $\bar{\mathbf{E}}(n)$ represents the uppermost $N-1$ elements of $\mathbf{E}(n)$ and $E_{N-1}(n)$ is its last element; Similarly, $\bar{\mathbf{e}}(n)$ contains the upper $N-1$ elements of $\mathbf{e}(n)$. Let $\mathbf{a}(n) = [x(n), \dots, x(n-N+1)]$, and $\tilde{\mathbf{r}}(n)$ be the vector that consists of the $N-1$ lower elements of $\mathbf{r}(n)$. The step size can be chosen within a range from 0 to 1. $\mathbf{R}(n)$ is updated by replacing its first row and first column with

the elements of $r_p(n)$, $p=0, \dots, N-1$, where $r_p(n)$ represents the p -th element of $\mathbf{r}(n)$ and moving down on the diagonal the left top $(N-1) \times (N-1)$ submatrix of $\mathbf{R}(n-1)$. Here are the equations of a general FAP algorithm.

Initialization

$$\mathbf{h}(-1) = \mathbf{0}, \mathbf{E}(-1) = \mathbf{0}, \mathbf{R}(-1) = \delta(-1)\mathbf{I}, \quad (1)$$

$$\boldsymbol{\alpha}(-1) = \mathbf{0}, \mathbf{r}(-1) = [\delta, \mathbf{0}]^T$$

For each $n \geq 0$

$$\mathbf{r}(n) = \mathbf{r}(n-1) + x(n)\boldsymbol{\alpha}(n) - x(n-L)\boldsymbol{\alpha}(n-L) \quad (2)$$

$$e(n) = y(n) - \mathbf{x}^T(n)\mathbf{h}(n-1) - \mu\tilde{\mathbf{r}}^T(n)\bar{\mathbf{E}}(n-1) \quad (3)$$

$$\text{update } \mathbf{R}(n) \text{ using } \mathbf{r}(n) \quad (4)$$

$$\bar{\mathbf{R}}(n) = \mathbf{R}(n) + \delta(n)\mathbf{I} \quad (5)$$

$$\mathbf{e}(n) = \begin{bmatrix} e(n) \\ (1-\mu)\bar{\mathbf{e}}(n-1) \end{bmatrix} \quad (6)$$

$$\text{Solve } \bar{\mathbf{R}}(n)\boldsymbol{\varepsilon}(n) = \mathbf{e}(n) \quad (7a)$$

$$\text{Or } \boldsymbol{\varepsilon}(n) = \bar{\mathbf{R}}^{-1}(n)\mathbf{e}(n) \quad (7b)$$

$$\mathbf{E}(n) = \begin{bmatrix} 0 \\ \bar{\mathbf{E}}(n-1) \end{bmatrix} + \boldsymbol{\varepsilon}(n) \quad (8)$$

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mu\mathbf{x}(n-N+1)E_{N-1}(n) \quad (9)$$

The key element in a FAP algorithm is the approach to solve the linear system in (7a). The choice of the approach (i.e., direct or iterative) determines the stability and robustness of the FAP algorithm. Several proposed FAP algorithms use the following approximation [4-6]: if $\mu = 1$ (non-relaxed case) the error vector reduces to a scalar and only the first column of the inverse of the autocorrelation matrix is needed in order to find $\boldsymbol{\varepsilon}(n)$.

However, this approximation holds for $0.7 < \mu \leq 1$ as well. In [5], the Gauss-Seidel (GS) method is used to solve the linear system $\bar{\mathbf{R}}(n)\boldsymbol{\varepsilon}(n) = \mathbf{e}(n) \cdot \mathbf{b}$, where $\mathbf{b} = [1, 0, \dots, 0]^T$. It is known that if the matrix $\mathbf{R}(n)$ is symmetric and positive definite, the GS method is guaranteed to converge. It was shown in [5] and confirmed in [6] that one GS iteration is enough for a near optimal performance. However, for an arbitrary μ , the solution of the linear system $\bar{\mathbf{R}}(n)\boldsymbol{\varepsilon}(n) = \mathbf{e}(n)$ is preferable to the matrix inversion. Like in the Gauss-Seidel Fast Affine Projection (GSFAP) algorithm, we propose to use the GS method for this purpose. We have called this algorithm the modified Gauss-Seidel fast affine projection (MGSFAP).

It is known that low cost loudspeakers or microphones introduce nonlinear distortions, especially at high

volumes. In such cases, the performance of a linear acoustic canceller degrades. A common approach is to consider nonlinear models [7-8]. Often in the case of the nonlinear distortions the simplified Volterra filters (SVF) can achieve better system identification than a linear one, but at the price of a much higher complexity. We consider memory-less non-linearities that have been proposed as a reliable model for the nonlinearities that occur in amplifiers and loudspeakers [8]. The MGSFAP is generalized for SVFs and the resulted algorithm is abbreviated as MGSFAP-SVF. As shown in [7], only the computation of the input vector and the generalized correlations are affected. We denote by O the order of the simplified Volterra filter. The size of the weight vector becomes OL instead of L . The extended input vector has OL elements and the following form

$$\mathbf{x} = \begin{pmatrix} x(n), \dots, x(n-L+1), x^2(n), \dots, x^2(n-L+1), \dots \\ \dots, x^O(n), \dots, x^O(n-L+1) \end{pmatrix} \quad (10)$$

Let $\mathbf{h}(n) = [h_1(n), \dots, h_{OL}(n)]$ be the filter coefficient vector. The Eq. 11 replaces Eq. 2 as follows for $p = 0, \dots, N-1$

$$r_p(n) = r_p(n-1) + \sum_{o=1}^O x(n)^o x(n-p)^o - \sum_{o=1}^O x(n-N+1)^o x(n-p-N+1)^o \quad (11)$$

As shown in [7], the nonlinear parameters were adapted only during voiced sections. Therefore, a voice activity detection algorithm (VAD) is needed. Its performance requirements are not high. It should determine relatively well the voiced sections, since the stability of the algorithm is greatly improved by the regularization method. Many approaches were proposed, for example there are algorithms based on the short-term energy, zero crossing rate, cepstral features, a periodicity measure, statistical models, etc. (see [9] and the references therein). The energy-based approaches to speech detection suffer from a lack of robustness and fail to work well in a wide range of SNR, unless some care is taken. We used the algorithm proposed in [10] which is exploiting the similarity between the signal and its residual prediction by computing the magnitude of their coherence function. Purely noise frames have a coherence function that admits a value close to 1, while the value of the coherence function on voiced frames is close to 0. The coherence function admits an intermediate value between 0 and 1 for unvoiced frames. A threshold is used in order to decide on the nature of the considered frame. It was found that a threshold value equal to 0.85 is a good compromise for different noise situations [10]. The algorithm achieves a good performance, it has few parameters to adjust and it simple to implement. In integrated systems, some speech

coding function (LP analysis or residual prediction signal) can be re-used in this case.

3. SIMULATIONS

We simulated the algorithm performance using test signals recorded in a car cabin. The convergence of the algorithm was compared by using the squared norm of the difference between the car cabin impulse response and the adaptive filter coefficients in dB. The echo return loss enhancement (ERLE) is computed in windows of 100 ms as follows

$$ERLE = 10 \log_{10} \frac{E[y^2(n)]}{E[e^2(n)]} \quad (12)$$

The Gauss-Seidel procedure requires the execution of a number of iterations in order to obtain an acceptable accuracy in the solution of the linear system.

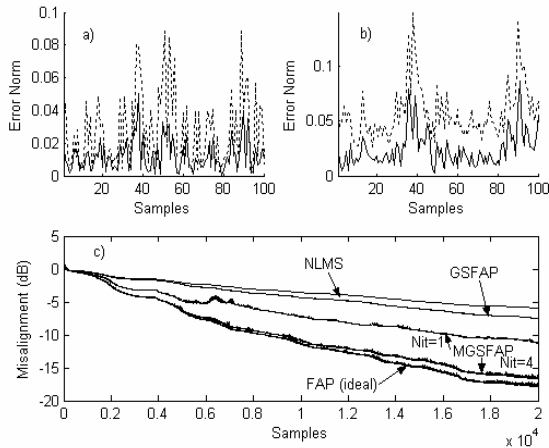


Fig. 1. The error norm between the exact solution and the iterated solution of $\mathbf{\epsilon}(n)$ for two values of the step size a) $\mu=1$ and b) $\mu=1/8$. The result of $N_{it}=4$ GS iterations is plotted with a solid line, while the result of $N_{it}=1$ GS iteration is plotted with a dotted line. c) The learning curves for NLMS, GSFAP, MGSFAP ($N_{it}=1$ and $N_{it}=4$) and ideal FAP algorithms ($SNR=30dB$, $\mu=1/8$, $L=256$, $N=8$).

It can be seen in Figs. 1a and 1b that fewer GS iterations are needed in order to obtain about the same error norm between the exact solution of $\mathbf{\epsilon}(n)$ and the iterative one for $\mu=1$ than for $\mu=1/8$. Fig. 1c shows that the performance of the MGSFAP is improved with the number of GS iterations. It can be seen that MGSFAP performance closely approaches that of the ideal FAP when $N_{it}=4$ for a small step size. Moreover, its performance is better than that of NLMS and GSFAP or other non-relaxed FAPs, (e.g. CGFAP [4]). We found that

two GS iterations are enough for step size values close to 1. Therefore, if a variable step size is used, the number of GS iterations can be varied accordingly. The regularization is a necessary part of the algorithm, especially in noisy conditions and double-talk scenarios. Often when there is near-end speech activity, SNR fall sharply well below 0 dB. We used a simple regularization method based on the approach described in [6]. The regularization parameter is $\delta_n = \rho_x(n)$ if $\rho_x(n) > \gamma \rho_y(n)$ and $\delta_n = 20L \rho_y(n)$ otherwise, where $\gamma=8$ and $\rho_x(n)$ and $\rho_y(n)$ are the time-averaged powers of $x(n)$ and $y(n)$ respectively. It can be seen from Fig. 2 that MGSFAP with 4 GS iterations closely matches the performance of the ideal FAP algorithm in the double-talk scenario. Similar results were obtained for moderate SNR values, different filter lengths, and projection orders.

Next, we examined the performance of the MGSFAP-SVF algorithm. The results are similar with those obtained in [8]. Fig. 3a shows the VAD decision where 1 indicates speech frames, while 0 indicates silence frames. The ERLE performance of the MGSFAP for SVF is lower in high volume level than in the normal volume level (Fig. 3c). Also, it can be seen from Fig. 3b that the difference between the NLMS and MGSFAP for SVF in high volume level is rather small. These benefits were obtained at the cost of much increased complexity than that of the linear algorithms.

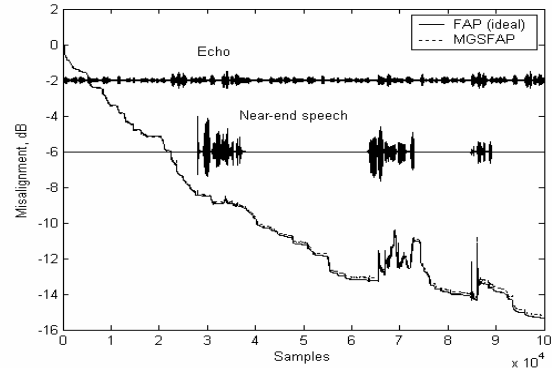


Fig. 2. Comparison of the misalignment (in dB) provided by the ideal FAP and MGSFAP with time-varying regularization in a double-talk scenario for $SNR=30dB$, $\mu=1/4$, $N_{it}=4$, $L=256$, $N=8$.

The linear system can be solved with any direct method (e.g. Cholesky or LDL^T factorization method). If the last method is applied the number of multiply and accumulate (MAC) operations is $N^3/6 + 2N^2 + 17N/6 - 2$ and N divisions in excess of $2L$ that are common to all FAP algorithms. The Cholesky method is more complicated having about the same number of MACs and divisions,

but it has N square root operations (cycles expensive). The MGSFAP algorithm ($N_{it} = 4$) needs $4N^2 + 4N - 2$ MACs and 1 division, while the original FAP [2] needs $20N$ MACs and 5 divisions.

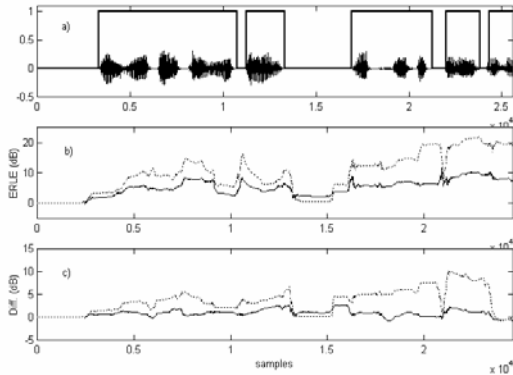


Fig. 3. a) The speech signal, and the VAD decision; b) ERLE of the MGSFAP algorithm ($N = 2, N_{it} = 4$) on high (solid line) and normal (dotted line) volume levels; c) Improvement achieved over NLMS using a 5th order SVF filter, in the same situations.

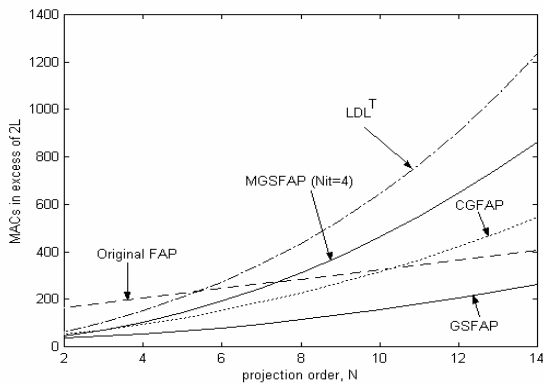


Fig. 4. Computational complexity between the considered FAP algorithms (a division is counted as 25 MACs [4]).

The CGFAP algorithm [4] needs $2N^2 + 9N + 1$ MACs and 1 division, whereas the GSFAP needs $N^2 + 3N + 1$ MACs and 1 division. As can be seen in Fig. 4, the MGSFAP compares favorably with the relaxed methods, and it is more efficient than them for $N < 8$. However, the original FAP [2] is unstable, while MGSFAP was stable in our simulations. If the stabilization procedure proposed in [2] is used then MGSFAP ($N_{it} = 4$) is more efficient for $N < 12$ (typical situation for most voice applications). The complexity of the non-relaxed methods is smaller, but their performance is not very good for lower step sizes (Fig. 1c). The complexity of the MGSFAP-SVF algorithm is much higher. For example, in the reported experiment it needed approximately 2600

MACs, while NLMS needed only 529 MACs, and MGSFAP 559 MACs.

4. CONCLUSIONS

We have proposed the modified Gauss-Seidel fast affine projection algorithm and extended it for simplified Volterra filters. We have demonstrated that the new algorithm has a fast convergence, low complexity and behaves well in a double-talk scenario. Therefore, the proposed algorithms represent interesting options for practical acoustic echo cancellation systems.

Acknowledgement .The work presented was developed within VISNET, a European Network of Excellence (<http://www.visnet-noe.org>), funded under the European Commission IST FP6 programme.

5. REFERENCES

- [1] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol. 67-A, no. 5, 1984.
- [2] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. 1995 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 3023 - 3026, Detroit, Michigan, U.S.A, May 1995.
- [3] S. L. Gay, "Fast projection algorithms with application to voice echo cancellation", Ph.D. Dissertation, Rutgers The State University of New Jersey, New Brunswick, New Jersey, 1994
- [4] H. Ding, "A stable fast affine projection adaptation algorithm suitable for low-cost processors," in *Proc. 2000 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 360 -363, Istanbul, Turkey, June 2000.
- [5] F. Albu, J. Kadlec, N. Coleman, and A. Fagan, "The Gauss-Seidel fast affine projection algorithm," in *Proc. IEEE Workshop Signal Processing Systems (SIPS 2002)*, pp. 109 - 114, San Diego, U.S.A, October 2002.
- [6] H. Sheikhzadeh, R. L. Brennan, and K. R. L. Whyte, "Near-end distortion in over-sampled subband adaptive implementation of affine projection algorithm," in *Proc. 2004 European Signal Processing Conf.*, pp. 413-416, Vienna, Austria, September 2004.
- [7] R. Niemisto, T. Makela, and V. Mllyla, "Robust fast affine projection algorithm for nonlinear acoustic echo cancellation," in *Proc 2002 European Signal Processing Conf.*, pp. 523-526, Toulouse, France, September 2002.
- [8] A. Stenger and W. Kellerman, "Adaptation of a memory-less preprocessor for nonlinear acoustic echo cancelling", *Signal Processing*, vol. 80, pp. 1747-1760, 2000.
- [9] S. G. Tanyer and H. Özer, "Voice activity detection in nonstationary noise," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 478-482, July 2000.
- [10] S. Ben Jebara, "A voice activity detector in noisy environments using linear prediction and coherence method", in *Proc. WSES Multiconference Acoustics Music: Theory and Applications*, pp. 308-311, September 2001.