# ANALYSIS OF LOW RANK TRANSFORM DOMAIN ADAPTIVE FILTERING ALGORITHM

*Balaji Raghothaman, Darel Linebarger*

Program in Electrical Engineering EC33,
University of Texas at Dallas,
Richardson, TX 75083-0688.

*Dinko Begušić*

Faculty of Electrical Engineering,
University of Split,
HR-21000 Split, Croatia.

## ABSTRACT

This paper analyzes an SVD based low rank transform domain adaptive filtering algorithm and proves that it performs better than Normalized LMS. The method extracts an underdetermined solution from an overdetermined least squares problem, using a part of the unitary transformation formed by the right singular vectors of the data matrix. The aim is to get as close to the solution of an overdetermined system as possible, using an underdetermined system. Previous work based on the same framework, but with the DFT as the transformation [1, 2], has shown considerable improvement in performance over conventional time domain methods like NLMS and Affine Projection. The analysis of the SVD-based variant helps us to understand the convergence behavior of the DFT-based low complexity method. We prove that the SVD-based method gives a lower residual than NLMS. Simulations confirm the theoretical results.

## 1. INTRODUCTION

Transform domain methods have been used in adaptive filtering problems for reducing the complexity in various ways. The Frequency Domain Adaptive Filter [3] uses the convolution property of the DFT. The efficiency of subband methods [4] is due to their slower rate of adaptation. Subspace methods can also be considered transform domain methods which use the singular vectors of the data as transforms. These are effective rank-reducing mechanisms [5], useful for low rank problems.

Our transform domain adaptive filtering framework [1, 2], is based on the linear least squares problem. Let

$$\mathbf{X}_{M\times N} = [\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_N}]$$

be the matrix formed from the received signal, $\mathbf{y}_{N\times 1} = [y_1, y_2, \ldots y_N]^T$ be the desired signal vector and $\mathbf{e}_{N\times 1} = [e_1, e_2, \ldots e_N]^T$ be the *a priori* error vector, in an echo cancellation problem as shown in figure 1. Let $\mathbf{h}_{M\times 1}$ and $\mathbf{\Delta h}_{M\times 1}$ be the adaptive filter coefficient vector, and its increment respectively. Time subscripts have not been shown for brevity. We are trying to solve the system

$$\mathbf{X}^H(\mathbf{h}+\Delta\mathbf{h}) = \mathbf{y}, \quad \text{or equivalently,} \quad \mathbf{X}^H\Delta\mathbf{h} = \mathrm{e}. \quad (1)$$

In this system, there are $M$ unknowns (filter taps), and $N$ equations.

The UTD authors would like to acknowledge support from the UTD Telecommunications DSP Consortium, TARP Grant 009741-039 and Texas Instruments.

## 2. LOW RANK TRANSFORM DOMAIN ALGORITHM

We modify the problem defined by (1), by applying $\mathbf{Q_1}$, a subset of a unitary transformation given by $\mathbf{Q}_{N\times N} = [\mathbf{Q_1}_{(N\times P)}\mathbf{Q_2}_{(N\times N-P)}]$ where $P < M$, so that (1) now becomes

$$\mathbf{Q_1}^H\mathbf{X}^H\mathbf{\Delta h} = \mathbf{Q_1}^H\mathrm{e}. \quad (2)$$

This new underdetermined problem has a minimum norm solution given by

$$\mathbf{\Delta h} = \mathbf{X}\mathbf{Q_1}(\mathbf{Q_1}^H\mathbf{X}^H\mathbf{X}\mathbf{Q_1})^{-1}\mathbf{Q_1}^H\mathrm{e}, \quad (3)$$

and the updated filter is given by

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu\mathbf{\Delta h}_k, \quad (4)$$

where $\mathbf{h}_k$ denotes the adaptive filter at time $k$. It is to be noted that NLMS [6] and Affine Projection(**AP**)[7, 8], fit into this framework, using $\mathbf{Q} = \mathbf{I}_{N\times N}$, an identity. For our low complexity algorithm [2], we used the DFT as the transform $\mathbf{Q}$. We used one DFT vector as our $\mathbf{Q}_1$ for each iteration. In [1], the near optimal $\mathbf{Q}_1$ is proven to be obtained by $\max_{\mathbf{Q}_1}||\mathbf{Q}_1^H\mathrm{e}||$. In [2], we described an approximate but efficient alternative which involved generating a set of random numbers whose pdf confirmed to $|\mathrm{DFT}(\mathbf{x})|^2$. These random numbers then represented which column of the DFT matrix to use as the transform for each iteration. This and other modifications resulted in a low complexity high performance algorithm for adaptive filtering, whose performance we demonstrated in an echo canceler setup (figure 2).

## 3. SVD-AP: CONVERGENCE ANALYSIS

In this section, we attempt to understand the convergence characteristics of the algorithm described in section 2, by analyzing a similar algorithm, based on the SVD instead of the DFT.

In SVD-AP, the set of right singular vectors of the data matrix at each iteration forms the full-rank transform for that iteration, and a subset of those singular vectors is selected as the low rank transformation. We derive expressions for the tap weight convergence in the mean square and the residual, and compare them with the corresponding expressions for NLMS given in the literature [9]. We neglect the computational complexity involved with the choice of the SVD as our transformation. We use the rank 1 version of our algorithm, and call it **SVD-AP(1)**.

Let us define $\mathbf{R} = E[\mathbf{x}_k\mathbf{x}_k^H] = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, with $\lambda_j, \quad j = 0, 1, \ldots, M-1$, as the eigenvalues. Now consider $E\left[\mathbf{X}_k\mathbf{X}_k^H\right]$, $\mathbf{X}_k$ being of size $M \times N$, where $N \geq M$. Its EVD is identical to that of $\mathbf{R}$, except that its eigenvalues are greater by a factor of $N$, *i.e.,* $E\left[\mathbf{X}_k\mathbf{X}_k^H\right] = \mathbf{U}(N\mathbf{\Lambda})\mathbf{U}^H$. Using the stationarity of the signal, the instantaneous SVD of the data matrix $\mathbf{X}_k$ can be written as

$$\mathbf{X}_k = \mathbf{U}_{M\times M}\sqrt{N\left[\mathbf{\Lambda}_{M\times M} \quad \mathbf{0}_{M\times(N-M)}\right]}\mathbf{V}^H_{N\times N}. \quad (5)$$

The unitary transform to be used is now

$$\mathbf{Q} = \mathbf{V} = [\mathbf{V}_0 \quad \mathbf{V}_1 \quad \ldots \quad \mathbf{V}_{N-1}].$$

We calculate the transform domain error vector as $\mathbf{V}^H\mathbf{e}$. Our low rank transform to be used at time $k$ will now be given by $\max||\mathbf{V}_i^H\mathbf{e}||$, where $\mathbf{V}_i$ is the $i^{th}$ column of $\mathbf{V}$, *i.e.,* the $i^{th}$ right singular vector. From now on, we will refer to $i$ as the **index** of the low rank transform at time $k$. The system is the same as defined in (2) and the solution as defined in (3), with $\mathbf{Q}_1 = \mathbf{V}_i$.

It is to be noted here that for $N > M$, $\mathbf{X}\mathbf{V}_i = \mathbf{0}$, for $M \leq i < N$, and hence the solution is trivial. In other words, selection of a $\mathbf{V}_i$ that is not in the column space of $\mathbf{X}^H$ is possible, but not desirable. So we have to limit our selection of indices to the first $M$.

The convergence properties of LMS [10] and NLMS [9] have been extensively analyzed in literature. We shall follow a procedure similar to a majority of these works, in that we shall project the tap weight error onto the eigenvectors of the data correlation matrix.

Let $\mathbf{h_o}$ be the optimum filter, *i.e.,* the actual echo path. Let $\mathbf{h}_k$ be the adaptive filter at time instant $k$ and let

$$\zeta_k = \mathbf{h}_o - \mathbf{h}_k \quad (6)$$

be the tap weight error vector at time $k$. The length $N$ residual vector at time $k$ is given by

$$\mathbf{e}_k = \mathbf{y}_k - \mathbf{X}_k^H\mathbf{h}_{k-1}. \quad (7)$$

where $\mathbf{y}_k$ is the desired signal vector at time $k$, given by

$$\mathbf{y}_k = \mathbf{X}_k^H\mathbf{h}_o + \mathbf{\Gamma}_k. \quad (8)$$

The term $\mathbf{\Gamma}_k$ represents an additive zero-mean white Gaussian noise vector, with each element having a variance $\sigma^2$.

In order to derive the total tap weight weight error power, we decompose $\zeta_k$ into its modal components projected onto the eigen space $\mathbf{U}$ of the correlation matrix:

$$\widetilde{\lambda}_{j_k} = \mathbf{U}_j^H E[\zeta_k\zeta_k^H]\mathbf{U}_j \quad (9)$$

where $\mathbf{U}_j$ is the $j^{th}$ column of $\mathbf{U}$. The tap weight error power is now given by $E[\zeta_k^H\zeta_k] = \sum_{j=0}^{M-1}\widetilde{\lambda}_{j_k}$. Using equations (4), (3), (7) and (8) and simplifying, we get the expressions for modal tap weight errors at time $k+1$ as

$$\widetilde{\lambda}_{j_{k+1}} = \begin{cases} (1-\mu)^2\widetilde{\lambda}_{j_k} + \frac{\mu^2\sigma^2}{N\lambda_j}, & j = i \\ \widetilde{\lambda}_{j_k}, & j \neq i. \end{cases} \quad (10)$$

It can be seen from (10) that only the mode corresponding to the current index is affected. The residual at time $k+1$, $e_{k+1}$, is given by the first element of (7). The expression for the MSE can be found from the literature [9] to be

$$E[|e_{k+1}|^2] = \left(\sum_{j=0}^{M-1}\lambda_j\widetilde{\lambda}_{j_k}\right) + \sigma^2. \quad (11)$$

Substituting (10) in (11), we get the MSE as

$$E[|e_{k+1}|^2] = \left(\sum_j \lambda_j\widetilde{\lambda}_{j_{k-1}}\right) \underbrace{-\mu(2-\mu)\lambda_i\widetilde{\lambda}_{i_{k-1}}}_{\mathbf{A}} \overbrace{+\sigma^2[1+\frac{\mu^2}{N}]}^{\mathbf{B}}. \quad (12)$$

### 3.1. Comparison with NLMS

Slock [9] follows essentially the same steps in order to obtain the expressions for the modal tap weight error and MSE for NLMS. With some approximations, we get the MSE learning curve for NLMS as

$$E[|e_{L_{k+1}}|^2] = \left(\sum_j \lambda_j\widetilde{\lambda}_{L,j_{k-1}}\right) \underbrace{-\mu(2-\mu)\frac{\sum_j \lambda_j^2\widetilde{\lambda}_{L,j_{k-1}}}{\sum_j \lambda_j}}_{\mathbf{C}} \overbrace{+\sigma^2\left[1+\frac{\mu^2\sum_j\lambda_j^2}{\left(\sum_j\lambda_j\right)^2}\right]}^{\mathbf{D}}. \quad (13)$$

The subscript L is used to distinguish values for NLMS from the similar values for SVD-AP. Comparing (12) and (13) tells us that the signal term for the residual in SVD-AP(1) does not depend on the signal eigenvalues, while the signal term in NLMS does.

Consider the case when we start with equal modal tap weight errors in both NLMS and SVD-AP(1) at time $k-1$. That is to say, $\widetilde{\lambda}_{j_{k-1}} = \widetilde{\lambda}_{L,j_{k-1}}$, for $j = 0, 1, \ldots, M-1$. Now we shall compare the one-step improvement in the residual in the two algorithms. We compare the signal terms (**A** and **C**) and the noise termw (**B** and **D**) separately. We can also neglect the subscript $L$ in $\widetilde{\lambda}_{L,j_{k-1}}$, since the modal tap weight errors of NLMS and SVD-AP(1) are assumed to be initially equal.

Consider the signal term of $E[|e_{k+1}|^2] - E[|e_{L_{k+1}}|^2]$. We have to prove that

$$\lambda_i \widetilde{\lambda}_{i_{k-1}} \geq \frac{\sum_j \lambda_j^2 \widetilde{\lambda}_{j_{k-1}}}{\sum_j \lambda_j} \tag{14}$$

Since we have selected $i$ at iteration $k$ such that $\lambda_i \widetilde{\lambda}_{i_{k-1}} = \max_j \lambda_j \widetilde{\lambda}_{j_{k-1}}$, we can write the following inequality for the RHS of (14):

$$\frac{\sum_j \lambda_j^2 \widetilde{\lambda}_{j_{k-1}}}{\sum_j \lambda_j} \leq \lambda_i \widetilde{\lambda}_{i_{k-1}} \frac{\sum_j \lambda_j}{\sum_j \lambda_j} = \lambda_i \widetilde{\lambda}_{i_{k-1}}, \tag{15}$$

which is nothing but the condition in (14). Considering the noise term of $E[|e_{k+1}|^2] - E[|e_{L_{k+1}}|^2]$, we have to prove

$$\frac{1}{N} \leq \frac{\sum_j \lambda_j^2}{\left(\sum_j \lambda_j\right)^2} \tag{16}$$

This condition is easily proved using the **Cauchy-Schwartz** inequality. Since we have proved both the signal and noise terms in the residual of SVD-AP(1) are less than equal to their counterparts in NLMS, we can say that the performance of SVD-AP(1) is always better than or equal to that of NLMS.

### 3.2. Relevance to DFT-AP(1)

The analysis of SVD-AP(1) gives us a good insight into the working of the DFT-based low complexity version, M-DFT-AP(2). We extrapolate many of the conclusions derived in the SVD-AP analysis to DFT-AP. The modes of convergence of the tap weight are now the squares of the DFT coefficients of the tap weight error vector. If we select frequency bin $i$ for a particular iteration, the tap weight error is decreased mainly in that bin. Some leakage effects also cause some convergence in other bins, unlike the SVD, because of imperfect diagonalization of the correlation matrix. A justification can also be found for the selection of indices using the DFT of the signal. From (10) and (12), it is clear that the maximum reduction in the tap weight convergence at any given iteration is obtained when **both $\widetilde{\lambda}_{i_k}$ and $\lambda_i$ are high**. On the other hand, it can be proved that

$$E[\|\mathbf{V}_i^H \mathbf{e}_{k+1}\|^2] = N\lambda_i \widetilde{\lambda}_{i_k} + \sigma^2. \tag{17}$$

Hence when we maximize $\|\mathbf{V}_i^H \mathbf{e}_{k+1}\|$, we are actually achieving $\max_i[\lambda_i \widetilde{\lambda}_{i_k}]$ Thus ideally, we have to select indices $i$ which will maximize $\lambda_i \widetilde{\lambda}_{i_k}$. After convergence of the tap weights has occurred, all $\widetilde{\lambda}_{i_k}$ have sufficiently converged so that the choice of indices is determined solely by $\lambda_i$, the data eigenvalues (square of DFT coefficients in the case of DFT-AP). Hence, it is reasonable to select indices that fit the statistics of the DFT coefficients of $\mathbf{x}$, which is what we do in the low complexity DFT-based method. But for initial convergence, this argument is not fully suited, since at that time, the $\widetilde{\lambda}_{i_k}$'s are also significant.

Finally, the performance degradation that occurs in the case of $N < M$ can be explained using the above analysis. For a data matrix $\mathbf{X}_{M\times N}$, there will be $M$ left singular vectors and $N$ right singular vectors in the SVD. When $N < M$, we are able to adapt in only any one of the first $N$ of the possible $M$ modes. The mode containing the largest error could be one of the other $M-N$ modes, hence the adaptation might not be taking place in the best possible mode. A similar logic can be applied in the DFT-based method, *i.e.,* the modes of adaptation are limited, hence a degradation in performance occurs there too.

## 4. SIMULATIONS

Figure 2 compares the residual error obtained in an echo canceler configuration, from SVD-AP(1), optimal DFT-AP(1), affine projection of order 4 (AP(4)) and normalized LMS. The tap weight error curves were similar to the residual curves. For the input signal, an AR process with one pole at $0.2$ is used. The echo path is a filter of length 128. This figure demonstrates that our algorithm performs better than the time domain algorithms.

Figure 3 compares the theoretical and simulated residual and tap weight error for SVD-AP(1). The input used is a lowpass AR(1) process with a pole at 0.5. A 128-tap filter is used as the echo filter and an SNR of 40 db is maintained. The SVD, averaged over $k-1$ and $k$, is used as the transform at $k$, in order to avoid instability. The time-varying singular values are used for the theoretical calculation of tap weight error and residual ($\lambda_j$'s vary with time). It can be seen that the simulation for SVD-AP(1) matches the theoretical curve fairly closely. The degradation in the residual simulation compared to the theoretical values can be attributed to the changing statistics (*i.e.,* SVD) of the data. This introduces some deviation from perfect diagonalization, which is not accounted for in the derivation.

## 5. CONCLUSIONS

In this paper, we analyzed an algorithm based on a new transform domain framework, with the SVD as the transform and showed that it provides better performance than NLMS. The SVD based analysis gives us good insight into the working of the DFT-based low complexity algorithm described in [2]. Apart from acoustic echo cancellation, this algorithm can also be applied to system identification and possibly data echo cancellation problems.

## 6. REFERENCES


[1] D. Linebarger, B. Raghothaman, D. Begušić, R. Degroat, E. Dowling, and S. Oh. Low rank transform domain adaptive filtering. In *Proc. of ASILOMAR*, pages 123–127, 1997.

[2] B. Raghothaman, D. Linebarger, and D. Begušić. Low complexity low rank transform domain adaptive filtering. In *Proc. of DSP Workshop*, Aug. 1998.

[3] J. Shynk. Frequency-Domain and Multirate Adaptive Filtering. *IEEE SP Magazine*, pages 15–36, Jan. 1992.

[4] A. Gilloire and M. Vetterli. Adaptive Filtering in Subbands. In *Proceedings of ICASSP 1988*, pages 1572–1575, 1988.

[5] S. Goldstein and I. Reed. Reduced-Rank Adaptive Filtering. *IEEE Tr. SP.*, 45(2):492–496, Feb. 1997.

[6] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, third edition, 1996.

[7] S. Gay and S. Tavaitha. The fast affine projection algorithm. In *Proceedings of ICASSP, 1995*, pages 3023–3026, 1995.

[8] M. Montazeri and P. Duhamel. A set of algorithms linking NLMS and block RLS algorithms. *IEEE Tr. Signal Proc.*, 43(2):444–453, Feb. 1995.

[9] D. Slock. On the Convergence Behaviour of the LMS and the Normalized LMS Algorithms. *IEEE Tr. SP*, 41(9):2811–2825, Sep. 1993.

[10] B. Widrow, J. M. McCool, M. J. Larimore, and C. R. Johnson Jr. Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter. *Proc. of the IEEE*, 64(8):1151–1162, Aug. 1976.


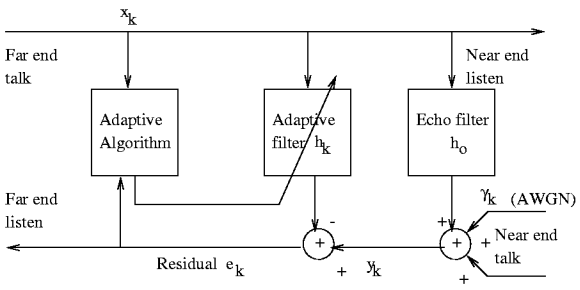


Figure 1: Block Diagram of Adaptive Echo Canceler

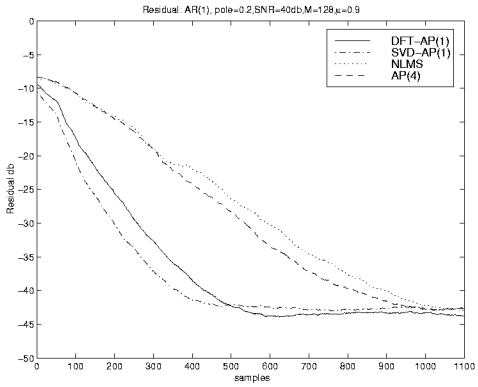


Figure 2: Comparison of Residuals: Ideal SVD-AP(1), DFT-AP(1), AP(4) and NLMS

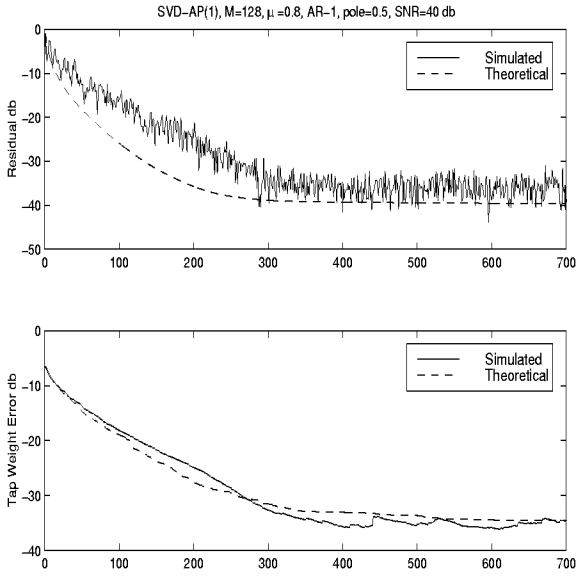


Figure 3: Comparison of theoretical and simulated residual for SVD-AP(1)