# Noise Cancellation in Sub-Bands Using the Conjugate Gradient Method

X. Fan

DSC Communications Corporation
Transmission Division
Plano, TX 75075

N. H. Younan and C. D. Taylor

Dept. of Electrical & Computer Engineering
Mississippi State University
Mississippi State, MS 39762

## Abstract

*In this paper, an adaptive filter bank is considered for noise cancellation with the conjugate gradient (CG) adaptive algorithm being used as the optimization tool. It uses the exponentially windowed pass instantaneous gradients and it has convergence properties superior to the LMS and lower computational complexity than the RLS. With the appropriately designed analysis and synthesis filter banks and the decimation factor such that aliasing can be ignored, the proposed technique provides satisfactory performance with faster convergence and high computational efficiency. Therefore, it is an attractive alternative for real time noise cancellation.*

## 1: Introduction

More than three decades, the need for real time application provides the enormous impetus for the development of adaptive signal processing. In many cases, noise cancellation, echo cancellation, and channel equalization require very long impulse response. It is quite common that hundreds or even thousands of the FIR filter coefficients may be needed so that the desired level of the filter performance can be achieved. Accordingly, the computation is intensive and restricts the long FIR adaptive filter from some real time application scenario which calls for efficiency and fast updating.

One solution to reduce the computational complexity is using adaptive IIR filters, which can have infinite long impulse response with relatively few coefficients. However, because of the potential problems of instability, slow convergence, and local minima, adaptive IIR filters have not been widely used. Another disadvantage of both adaptive FIR and IIR filters is the increased misadjustment and the degraded performance when they are used in the non-stationary environment [1]. In considering these problems, it has been suggested using filter banks to divide the problem into several sub-band problems [2,3]. So, instead of updating a long FIR filter, several short FIR filters are updated parallelly in each sub-band, where the signal is near stationary. If each sub-band of the signal considered has little spectrum spillover from the adjacent bands, adaptive filter banks can be designed to suppress the noise band-by-band.

The sub-band noise cancellation using filter banks offers several advantages. First, it provides the possibility to choose the adaptive algorithm independently for each sub-band so that it can fit best for the characteristics of the signal and the noise in that band and the best noise canceling results can be obtained. Second, the order and the convergence constant of each adaptive filter can be independently determined. Thus, a satisfactory tracking capability can be achieved at lower misadjustment. Third, since updating the coefficients of several short FIR filter can be performed in parallel, it offers higher computational efficiency than updating a long FIR filter sequentially. Therefore, the combined use of filter banks and efficient adaptive algorithms provide an attractive alternative for real time signal processing.

## 2: The Filter Bank Configuration

The sub-band adaptive noise cancellation scheme considered here consists of three filter banks working cooperatively to achieve these advantages. These are: The analysis filter bank for anti-aliasing and dividing the input signal into several sub-bands; the synthesis filter bank for image elimination; and the adaptive filter bank for noise cancellation. Since the analysis filter bank may not be able to eliminate the spillover from adjacent bands in real time applications and, in addition, critical decimation may cause undesirable aliasing in the output,



1058-6393/95 $4.00 © 1995 IEEE

the sub-critical decimation is often used. However, the computational complexity increases accordingly compared with the critical decimation. Note that the design of analysis and synthesis filter banks is independent of the adaptive filter banks design. The basic structure of an n channel adaptive filter bank is illustrated in Figure 1. Several other structural representations are also available [4].

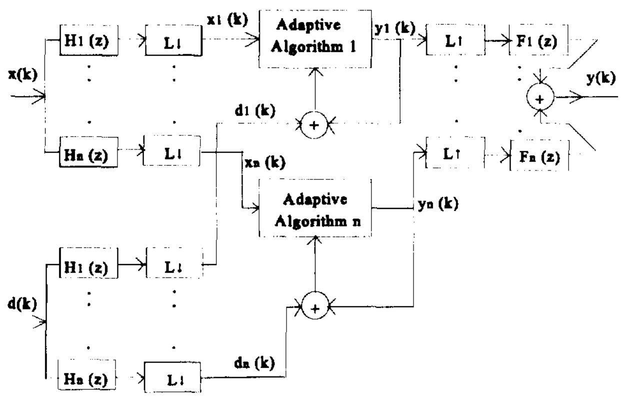


**Figure 1: Basic n Channel Adaptive Filter Bank**

Here, x(k) is the input signal, d(k) is the desired response, y(k) is the output, $H_i(z)$ is the analysis filter bank response, $F_i(z)$ is the synthesis filter bank response, and L is the decimation factor, i.e., $L \leq n$. To reduce the aliasing effects, L is often chosen smaller than n. If d(k) is a delayed version of x(k), then it becomes a bank of adaptive line enhancers. Note that for all r, $1 \leq r \leq n$, the following holds

$$X_r(z) = \frac{1}{n}\sum_{l=0}^{n-1} H_r(z^{1/L} W^l)\, X(z^{1/L} W^l) \tag{1}$$

and

$$Y(z) = \sum_{r=1}^{n} F_r(z)\, Y_r(z^L) \tag{2}$$

where $X_r(z)$, $Y_r(z)$, and Y(z) are the z transforms of $x_r(k)$, $y_r(k)$, and y(k) respectively and $W = e^{-j2\pi/n}$. Let $W_r(z,k)$ be the adaptive filter transfer function, then

$$Y_r(z) = W_r(z,k)\, X_r(z) \tag{3}$$

The relationship between X(z) and Y(z) can be established as follows

$$\begin{aligned} Y(z) &= \sum_{r=1}^{n} F_r(z)\, Y_r(z^L) \\ &= \sum_{r=1}^{n} F_r(z)\, W_r(z^L,k)\, X_r(z^L) \end{aligned} \tag{4}$$

Using equation (1) in equation (4) yields the following

$$\begin{aligned} Y(z) &= \frac{1}{n}\sum_{r=1}^{n}\sum_{l=0}^{n-1} F_r(z) W_r(z^L,k) H_r(zW^l) X(zW^l) \\ &= \left[\frac{1}{n}\sum_{r=1}^{n} F_r(z)\, W_r(z^L,k)\, H_r(z)\right] X(z) \\ &+ \frac{1}{n}\sum_{r=1}^{n}\sum_{l=1}^{n-1} F_r(z)\, W_r(z^L,k)\, H_r(zW^l)\, X(zW^l) \end{aligned} \tag{5}$$

Obviously, the first term of equation (5) represents the aliasing and image-free input and output relation with a transfer function

$$T(z,k) = \frac{1}{n}\sum_{r=1}^{n} F_r(z)\, W_r(z^L,k)\, H_r(z) \tag{6}$$

and the second term of equation (5) is the undesired images due to decimation and interpolation.

In general, T(z,k) is time dependent. If x(k) is both noise free and stationary, and other noise sources such as quantization and channel noise are not considered, it is possible to design $H_r(z)$ and $F_r(z)$ such that the analysis and synthesis banks can satisfy the perfect reconstruction condition (PR), that is: 1) $T(z,k)=z^{-p}$, where p is a positive integer (simply set $W_r(z,k)=1$, $1 \leq r \leq n$), 2) the second term of equation (5) is reduced to zero. In this case, the output y(k) will be a delayed version of x(k) with no aliasing, i.e., y(k)=x(k-p). Since x(k) considered here can be non-stationary and noise contaminated, therefore, adaptive filters are used for noise suppression. It can be argued that because of the time dependent nature of the adaptive filter weights $W_r(z,k)$ and the misadjustment of the adaptation process, it is almost impossible to design $F_r(z)$ and $H_r(r)$ ($1 \leq r \leq n$) such that $T(z,k)=z^{-p}$, for all time index k. On the other hand, perfect reconstruction does not serve the noise cancellation purpose. Otherwise, noisy signal x(k) is perfectly reconstructed at the other end of the filter banks. This yields a delayed version x(k-p) leaving the noise intact, which is not desired. In spite of this, $H_r(z)$ and $F_r(z)$ can be designed off-line independently of $W_r(z,k)$ according to the PR condition so that x(k-p) can be obtained when the contaminating noise in x(k) has no effects so that it can be ignored. The adaptive filter bank can be designed independently of $H_r(z)$ and $F_r(z)$ to suppress the noise while reconstruct the signal with as little distortion as possible.

## 3: The Conjugate Gradient Algorithm

The conjugate gradient method (CG) is a well known method for optimization. Its performance in terms of computational complexity and convergence properties is somewhat between the steepest descent and Newton's methods [6]. Originally, it is used for optimizing quadratic problems. And it is modified to optimize problems with general nonlinear objective functions. Here, a CG algorithm requiring no line search is considered. It is known that adaptive conjugate gradient FIR algorithm has convergence properties superior to those of the LMS and lower computational complexity than the RLS methods. In addition, the implementation of the CG method only involves vector multiplications, thus it requires less computer storage than the RLS method which requires matrix multiplication. Advances in the hardware technology also made it possible for real time implementation

### 3.1: Algorithm

The coefficient vector $\mathbf{W}_r(k)$, the input vector $\mathbf{X}_r(k)$, the output $y_r(k)$ , and the error $e_r(k)$ of the "rth" channel adaptive filter at time k are defined as follows

$$\begin{aligned} \boldsymbol{W}_r(k) &= [w_r^0(k), w_r^1(k), \ldots, w_r^{M_r-1}(k)]^T \\ \boldsymbol{X}_r(k) &= [x_r(k), x_r(k-1), \ldots, x_r(k-M_r+1)]^T \\ y_r(k) &= \boldsymbol{W}_r(k) * \boldsymbol{X}_r(k) \\ e_r(k) &= d_r(k) - y_r(k) \end{aligned} \quad (7)$$

where $M_r$ is the filter order and * stands for convolution. Then, the function to be minimized is

$$\begin{aligned} F[\boldsymbol{W}_r(k)] &= E\left\{e_r^2(k)\right\} \\ &= E\left\{\left(d_r(k) - \boldsymbol{W}_r^T(k) * \boldsymbol{X}_r(k)\right)^2\right\} \end{aligned} \quad (8)$$

with respect to $\mathbf{W}_r(k)$. Denote the gradient of $F[\mathbf{W}_r(k)]$ as $\nabla F[\mathbf{W}_r(k)]$, then, the algorithm is summarized in the following steps:

*Step 1*: starting with $\mathbf{W}_r(0)$, compute

$$\begin{aligned} \boldsymbol{g}_r(0) &= \nabla F[\boldsymbol{W}_r(0)] \\ \boldsymbol{d}_r(0) &= -\boldsymbol{g}_r(0) \\ \boldsymbol{q}_r(0) &= \boldsymbol{W}_r(0) - \boldsymbol{g}_r(0) \\ \boldsymbol{p}_r(0) &= \nabla F[\boldsymbol{q}_r(0)] \end{aligned} \quad (9)$$

*Step 2*: For k = 0, 1, ..., $M_r$ - 1, compute

$$\begin{aligned} \alpha_r(k) &= \frac{\boldsymbol{g}_r^T(k)\, \boldsymbol{d}_r(k)}{\boldsymbol{d}_r^T(k)\,[\boldsymbol{g}_r(k) - \boldsymbol{p}_r(k)]} \\ \boldsymbol{W}_r(k+1) &= \boldsymbol{W}_r(k) + \alpha_r(k)\, \boldsymbol{d}_r(k) \\ \boldsymbol{g}_r(k+1) &= \nabla F[\boldsymbol{W}_r(k+1)] \\ \boldsymbol{q}_r(k+1) &= \boldsymbol{W}_r(k+1) - \boldsymbol{g}_r(k+1) \\ \boldsymbol{p}_r(k+1) &= \nabla F[\boldsymbol{q}_r(k+1)] \\ \beta_r(k) &= \frac{\boldsymbol{g}_r^T(k+1)\, \boldsymbol{g}_r(k+1)}{\boldsymbol{g}_r^T(k)\, \boldsymbol{g}_r(k)} \\ \boldsymbol{d}_r(k+1) &= -\boldsymbol{g}_r(k+1) + \beta_r(k)\, \boldsymbol{d}_r(k) \end{aligned} \quad (10)$$

*Step 3*: Replace $\mathbf{W}_r(0)$ by $\mathbf{W}_r(M_r)$ and go to step 1. since the gradient involves taking the expectation value, i.e.,

$$\nabla F[\boldsymbol{W}_r(k)] = -2E\left\{e_r(k)\, \boldsymbol{x}_r(k)\right\} \quad (11)$$

then, its estimate is used. Accordingly, equation (11) can be written as

$$\nabla \hat{F}[\boldsymbol{W}_r(k)] = -2 \sum_{i=k-M_r+1}^{k} \lambda_r^{k-i+1}\, e_r(i)\, \boldsymbol{x}_r(i) \quad (12)$$

where $0 < \lambda_r < 1$. Note, the number of instantaneous gradients averaged can be larger or smaller depending on the signal and noise characteristics in the sub-band. The algorithm can be further simplified using the fixed step size $\alpha_r(k)=\alpha_r$. Thus, it is reduced to the well known

Fletcher-Reeves algorithm and the computational cost is reduced about 50% accordingly. However, it may meet with some instability problem when $\beta_r(k)$ is larger than unity. One way of solving this problem is to set $\beta_r(k)=1$ whenever it exceeds unity. Note that $M_r$, $\lambda_r$, and $\alpha_r$ ($1 \le r \le n$) can be independently determined for each channel.

### 3.2: Misadjustment

The misadjustment is a major measure of the performance of adaptive algorithms. In the stationary environment, at time k, the misadjustment is defined as [6]

$$M = \frac{MSE_{ave.excess}}{\xi_{min}} \tag{13}$$

where $\xi_{min}$ is the minimum mean square error (MSE) when the adaptive weights have converged to the Wiener solution $\mathbf{W}_r^*$, that is

$$\xi_{min} = E\left\{d_r^2(k)\right\} - \boldsymbol{P}_r^T(k)\,\boldsymbol{W}_r^* \tag{14}$$

with

$$\boldsymbol{P}_r(k) = E\left\{d_r^2(k)\,x_r(k)\right\} \tag{15}$$

and the average excess MSE is given by

$$MSE_{ave.excess} = E\left\{\boldsymbol{V}_r^T(k)\,\boldsymbol{R}\,\boldsymbol{V}_r(k)\right\} \tag{16}$$

Here, $\mathbf{V}_r(k) = \mathbf{W}_r(k) - \mathbf{W}_r^*$ and $\mathbf{R}$ is the covariance matrix of the input x(k) which is assumed to be stationary.

From equation (8), it is apparent that the error function $F[\mathbf{W}_r(k)]$ is quadratic with respect to $\mathbf{W}_r(k)$. If the true gradient is available, then $\mathbf{W}_r(k)$ can reach $\mathbf{W}_r^*$ at no more than $M_r$ steps. However, because of the noise presence, equation (12) is used instead of equation (11) in the execution of the algorithm and $\mathbf{V}_r(k)$ fluctuates at the bottom of the error surface even when the steady state has been reached. Based on a very rough assumption that $\mathbf{W}_r(k) \simeq \mathbf{W}_r^*$ after $M_r$ iteration, then

$$E\left\{\boldsymbol{V}_r^T(k)\boldsymbol{R}\,\boldsymbol{V}_r(k)\right\} \approx E\left\{\alpha_r^2(k)\,\boldsymbol{d}_r^T(k)\boldsymbol{R}\,\boldsymbol{d}_r(k)\right\} \tag{17}$$

This expression indicates that misadjustment is an increasing function of $\alpha^r(k)$. A large value of $\alpha_r(k)$ gives a good tracking capability but the misadjustment is increased accordingly, while a small $\alpha_r(k)$ does the opposite. In addition, it can be seen from equation (17) that the noise in the gradient estimates will eventually propagate and cause the noise in the weight vector. The analysis of convergence and misadjustment of the CG algorithm needs further investigation.

## 4: Numerical Example

To evaluate the foregoing technique, a simulated data is considered. The noisy input considered in this simulation has the form

$$x(k) = e^{-0.01kT}\left[\sin(2\pi f_1 kT) + \sin(2\pi f_2 kT)\right] + n(k) \tag{18}$$

where $f_1$ = 35.3 KHz, $f_2$ = 1.2 KHz, and n(k) is white Gaussian with zero mean and variance of 0.2, which corresponds to about SNR=20 dB. A two channels filter bank is used. Based on [5], the anti-aliasing filters are: $H_1(z) = a_0(z^2) + z^{-1}a_1(z^2)$ and $H_2(z) = a_0(z^2) - z^{-1}a_1(z^2)$, where $a_0(z)=(\beta_0+z^{-1})/(1+\beta_0 z^{-1})$ and $a_1(z)=(\beta_1+z^{-1})/(1+\beta_1 z^{-1})$ with $\beta_0$ = 0.211056 and $\beta_1$ = 0.686504. $H_1(z)$ has a stopband attenuation of 40 dB and the stopband edge is about $0.62\pi$. The synthesis filters are $F_1(z) = H_2(-z)$ and $F_2(z) = -H_1(-z)$. The frequency responses of $H_1(z)$ and $H_2(z)$ are illustrated in Figure 2. The orders of the two adaptive filters are 16 and 9 respectively. The fixed step sizes considered here are $\alpha_1$ = 0.0025 and $\alpha_2$ = 0.001, whereas $\lambda_1$ and $\lambda_2$ are chosen to be 0.6 and 0.8 respectively.

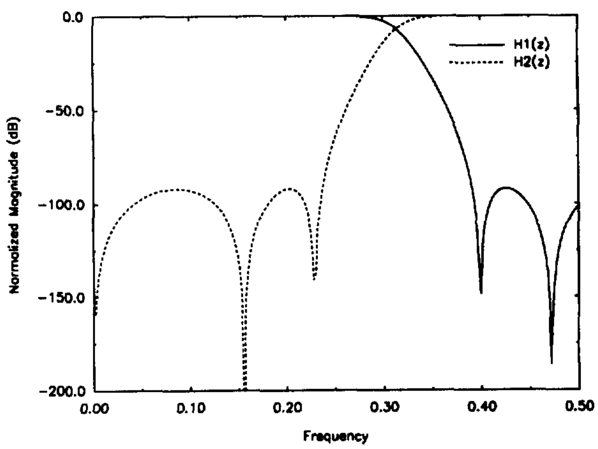


**Figure 2: Frequency Response of Hi(z)**

The adaptation errors in the low-pass and high-pass channels (channel 1 and 2) are illustrated in Figures 3 and 4. For comparison, an order of 25 FIR filter using the LMS algorithm is considered. The adaptation constant in this case is $\mu = 0.0025$. Its adaptation error is illustrated in Figure 5. Compared to the error of figure 5 obtained from the LMS algorithm, Figures 3 and 4 illustrate remarkable improvements. Accordingly, the presented scheme is a useful tool for real time noise cancellation.

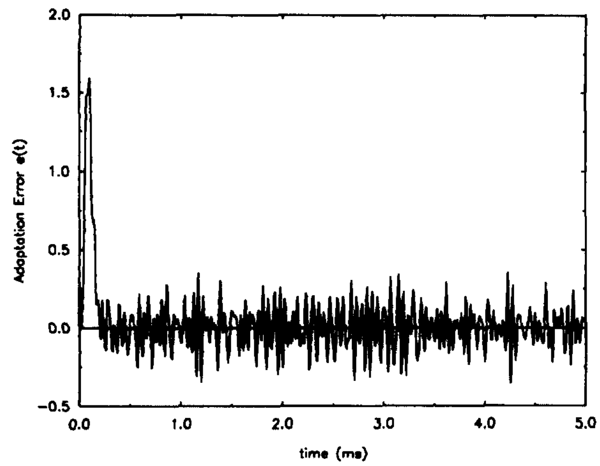


Figure 3: CG Channel 1 Error

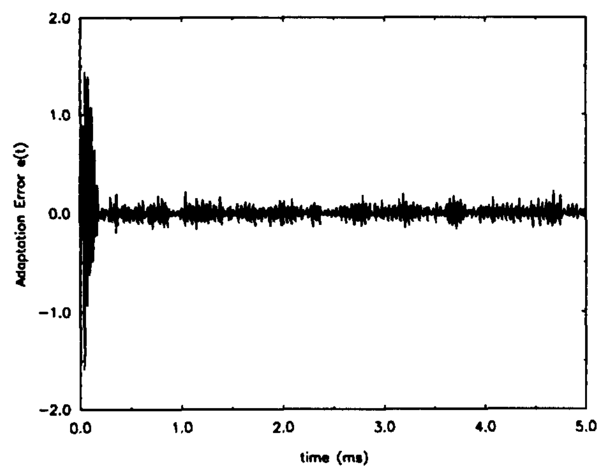


Figure 4: CG Channel 2 Error

## 5: Conclusion

Adaptive filter banks have been effectively used to suppress the noise in sub-bands. With the appropriate design of the analysis and synthesis filters and choice of a decimation factor to avoid aliasing, it is shown that this technique provides satisfactory performance with high computational efficiency and faster convergence. Results for simulated data are presented to demonstrate the applicability of this techniques.

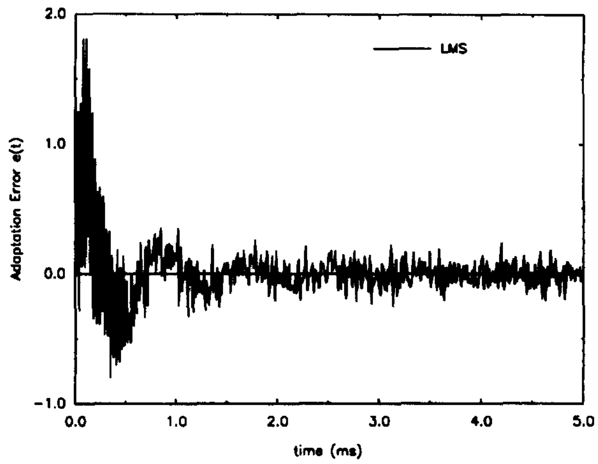


Figure 5: LMS Error

## References


[1] J.R. Zeidler, "Performance Analysis of LMS Adaptive Prediction Filters," Proc. of The IEEE, Vol. 78, No. 12, December 1990.

[2] A. Gilloire and M. Vetterli, "Adaptive Filtering in Subbands," Proc. IEEE Intl Conf. Acoust., Speech, Sig. Proc., New York, 1988, pp.1572-1575.

[3] M.R. Petraglia and S.K. Mitra, "On the use of the filter banks in adaptive filtering," Proc. 22nd Asilomar Conf. Signal, Systems, and Computers, Pacific Grove, CA, October 1988.

[4] J. Shynk, "Frequency Domain and Multirate Adaptive Filtering," IEEE Signal Processing Magazine, January 1992.

[5] P.P. Vaidyanathan, "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial," Proceedings of The IEEE, Vol. 78, No. 1, January 1990.

[6] P.M. Clarkson, Optimal and Adaptive Signal Processing, CRC Press, 1993.