LMS Based Affine Projection Algorithms for Noise Reduction

Jackson Cornell
Department of Electrical and Computer Engineering
University of Florida
jcornell@ufl.edu

Michael Kim
Department of Electrical and Computer Engineering
University of Florida
michaelkim@ufl.edu

***Abstract*—** In this paper we present a speech enhancement method based on Affine Projection Algorithms (APA). Tests of two APA members are performed for the purpose of interference canceling on a noisy speech signal. The performances of the normalized LMS filter and the non-regularized APA by gradient descent are first evaluated using a 2nd order filter. After obtaining metrics such as the Mean Squared Error, frequency response, and SNR improvement, the optimal filter order is calculated, and the results are compared.

***Keywords*—** Adaptive Signal Processing, Noise Reduction, Affine Projection Algorithms, LMS, NLMS, SNR, Mean Squared Error, Speech Enhancement

I. Introduction

Adaptive signal processing is a field of study wherein a filter's weights are iteratively adapted to minimize an error measurement [1]. One common method in signal processing is the Least Mean Squares (LMS) filter, which uses stochastic gradient descent to minimize the least-squares cost function. The least-squares cost function finds a set of parameters that minimizes the sum of the squared deviations between the observed response and a function of the parameters and input response [2]. These parameters are found iteratively using stochastic gradient descent, a process where the algorithm descends the cost function and finds its minimum.

The normalized least mean squares filter is related to the LMS adaptive filter and solves the problem of selecting a fixed learning rate while avoiding divergence by normalizing the step size by the input power. It is therefore able to achieve better convergence characteristics than that of LMS. Figure 1 shows a general block diagram of the NLMS filter. In this case, the desired signal d[n] represents the speech with vacuum cleaner noise added and the input signal x[n] contains only the vacuum cleaner noise. The adaptive filters find the optimal weights that transforms the signal x[n] to d[n] and uses an error signal e[n] to find these weights. This problem is unique in that the filtered signal is not the transformed signal y[n], but the error signal itself that returns the audio with reduced noise.

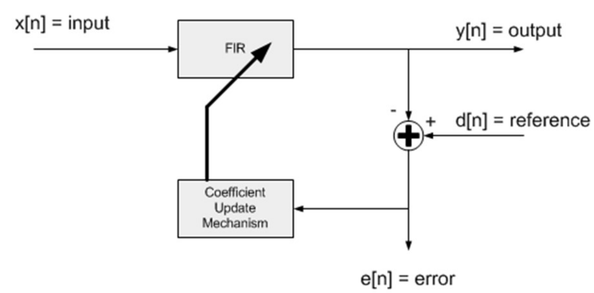

Figure 1: Normalized LMS Filter

II. Methods

A. Data

The given dataset consists of a .mat file with two channels labeled as desired and input. The data itself is two speech signals collected by microphones in a noisy environment: one contains the speech added with vacuum cleaner noise and the other just the noise. Each signal was sampled at 21 kHz and is about 3 seconds in length. The objective is to design a machine learning algorithm using the Affine Projection Algorithm (APA) family to denoise the desired input from the machine noise to understand the speech.

B. Algorithms

The APA family of algorithms generalize the mean squared error filters to multiple input vectors, allowing for better tracking of the optimal filter weights [3]. Compared to other methods such as the RLS solution, APA has better performance and a lower complexity than that of other adaptive algorithms because it reuses old data, resulting in fast convergence speeds when the input signal is highly correlated. Variants within the family can differ by their search method or regularization in the cost function. For this project, two algorithms were used to extract the speech signal from the noise: APA 1 and APA 2. APA 1 generalizes the LMS filter and was chosen so that it could follow the signal power in a quicker manner than APA 2. It updates its weights using gradient descent:

$$w(i) = w(i-1) + \eta U(i)[d(i) - U(i)^T w(i-1)]$$

APA 2 generalizes the NLMS filter when the kernel size is and was chosen due to its ease of implementation. It updates its weights by using Newton's Recursion:

$$w(i) = w(i-1) + \eta U(i)[U(i)^T U(i) + \varepsilon I]^{-1}[d(i) - U(i)^T w(i-1)]$$

The difference between the two algorithms is that, as the signal power changes, APA 2 is less prone to divergence given equal step sizes [4]. As the step size for the NLMS is normalized, NLMS converges faster and the estimated error value between the desired signal and filter output is smaller than that of LMS. The two algorithms were compared with equal hyperparameters: filter length of 2, input size of 1, and step size of 0.001. A very small value of 1e-6 was chosen for epsilon for APA 2 to ensure that the matrix inverse can be performed without regularizing the resulting value. These parameters were chosen as a baseline for further investigation on optimal hyperparameter tuning, as well as chosen for ease of visualization.

C. Hyperparameters

The signal-to-noise ratio is defined as the ratio of the signal power to the noise power corrupting the signal. All hyperparameters were chosen to maximize the SNR measurement of Echo Return Loss Enhancement (ERLE) given by the following:

$$ERLE = 10\log\left(\frac{E[d^2]}{E[e^2]}\right)$$

For each hyperparameter, an array of values was passed into the respective APA function, and the value with the highest ERLE SNR value was chosen. This was performed with input size, filter size, epsilon, and step size in that order. Except for epsilon—which was only computed for APA 2—this method was used individually on both APA 1 and APA 2.

III. Results

A. 2-Tap Filter Order

An analysis of the two algorithms with a filter size of 2, step size of 0.001, and input size of 1 was first performed as a baseline for improving the hyperparameters. The figure below shows a three-dimensional plot of the performance surface contours at time 1.5 seconds. Note the quadratic form and convex shape showcases a minimum for which the weights will converge to. The optimum weight values are seen at the projection of the bottom of the performance surface.

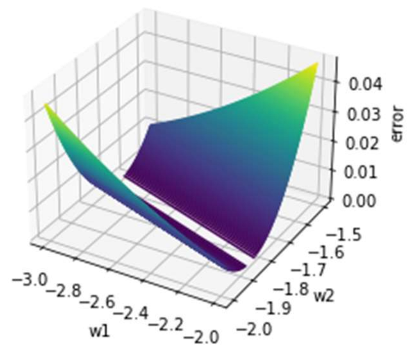

Figure 2: 2-Tap Performance Surface Contour

The figure below shows how the weights for the filters change over time. The weight tracks assist in obtaining the best solution of the adaptive filter. APA 2 seems to show a tendency to overfit the data, while APA 1 has a better estimate, as demonstrated by the fact that the weights hardly change even when the power of the noise changes at about 2 seconds in.

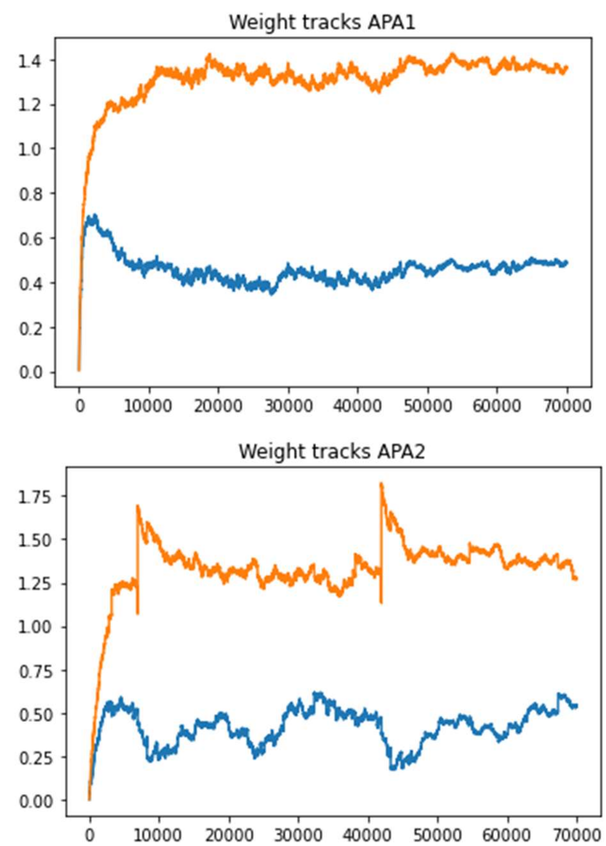

Figure 3: 2-Tap Weight Tracks

At first glance, the learning curves may seem to be diverging, but it is important to remember that the desired data is the error measurement, rather than the output y[n], as is common with adaptive filtering problems. Nonetheless, there is a clear reduction in error with increasing epochs.

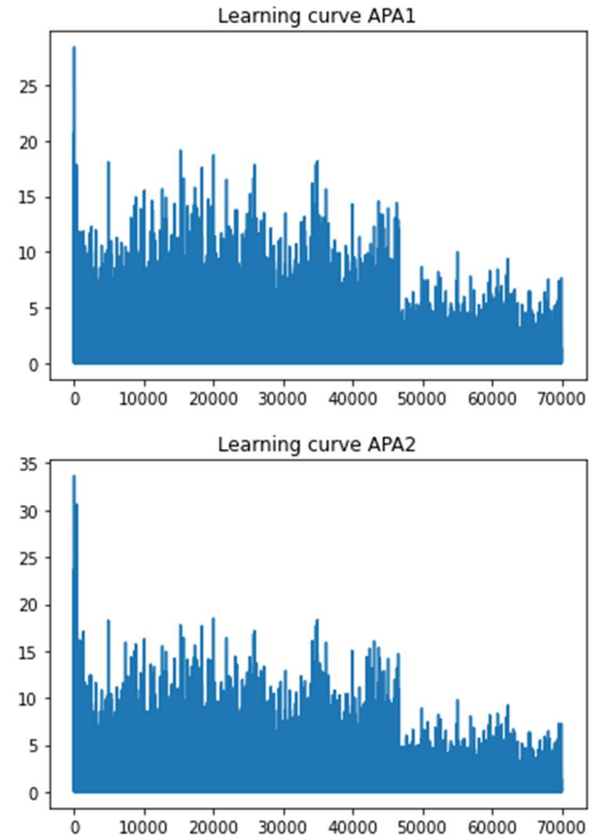

Figure 4: 2-Tap Learning Curves

The frequency plots show the system response from the desired signal to the noise-filtered signal (or, the error). In other words, the frequency plots demonstrate a model that filters out the noise.

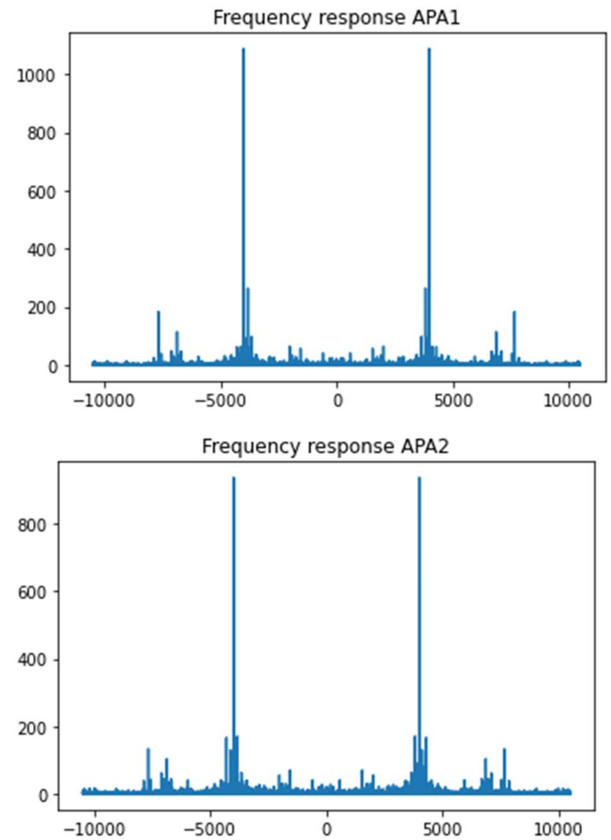

Figure 5: 2-Tap Frequency Responses

The SNR improvement describes how much the noise that appears in the desired signal is attenuated when observing the error signal. From our results, we can see that the normalized LMS/APA 2 returns a greater ERLE SNR value than the APA 1.

```
APA1 ERLE SNR Improvement (dB): 8.760316800887626
APA2 ERLE SNR Improvement (dB): 30.409528346257268
```

Figure 6: 2-Tap SNR Improvements

B. Optimal Hyperparameters

The next stage in our analysis was choosing the hyperparameters (filter order, input size, step size, and epsilon) that best filtered the noise from the desired signal. To do this, we iteratively ran through a series of common values for each filter and chose the value that maximized the ERLE SNR. The following is the resulting optimal hyperparameters:

Table 1: APA 1

Hyperparameter	**Value**
Input Size	10
Filter Order	19
Step Size	0.001

Table 2: APA 2

Hyperparameter	Value
Input Size	1
Filter Order	2
Step Size	0.001
Epsilon	1e-6

Below shows the weight tracks of the two filters to demonstrate how they converge. APA 2's optimal filter size was 2 taps, so the weight tracks are the same as above. However, APA 1's optimal filter size was 19, and the below figures demonstrates how the weights converge with very little spikes.

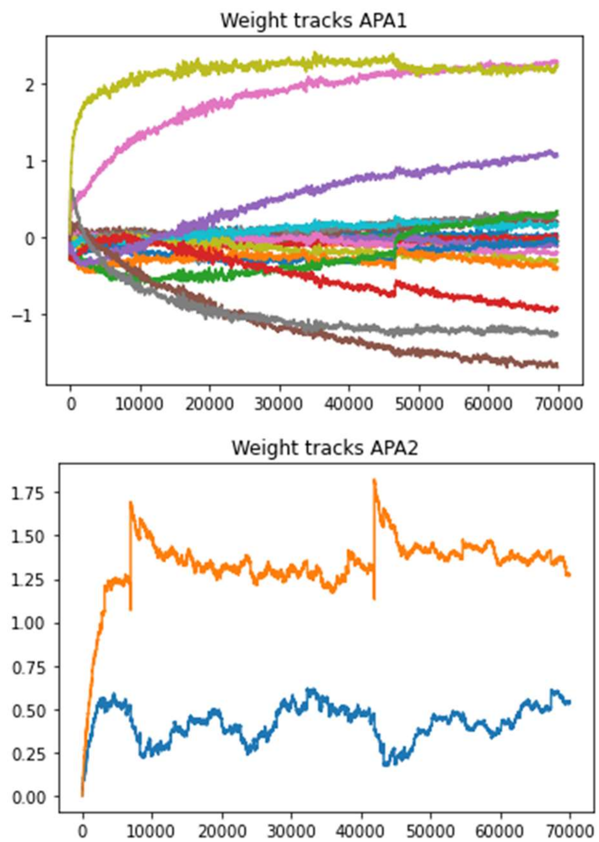

Figure 7: Optimal Filter Order Weight Tracks

The learning curves show that APA 1 has a significant decrease in noise compared to APA 2. Although APA 1 still has many spikes, it is again crucial to recall that the error is the extracted speech signal, and not a true error measurement.

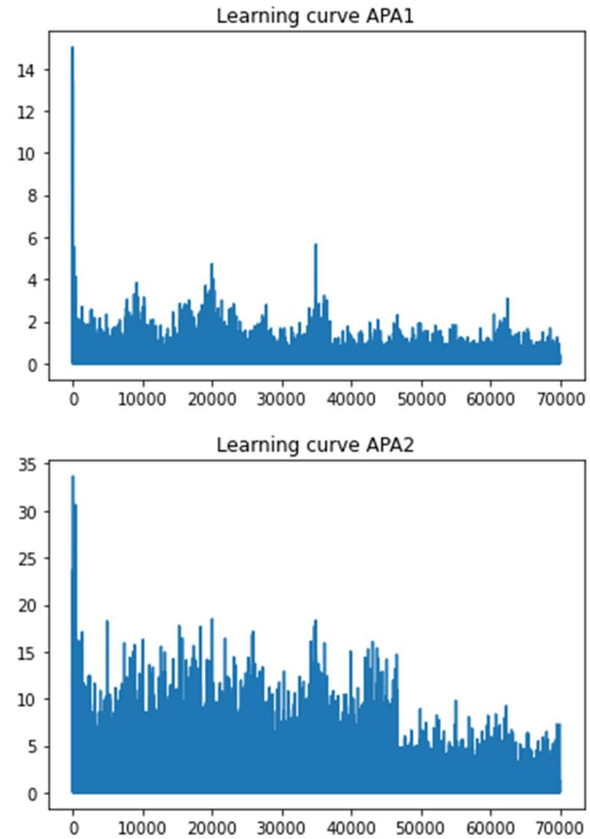

Figure 8: Optimal Filter Order Learning Curves

The frequency responses show a difference in complexity of the two models. APA 1, with a filter size of 19, can model the required filter to remove the noise much better than that of APA 2, which has a very simple frequency response. Additionally, it is more straightforward to analyze the both the noise and speech with APA 1's frequency response. A higher magnitude in the lower frequencies indicates that the speech is of a lower frequency than that of the noise. The highly attenuated higher frequencies further support this claim that the noise is mostly of higher frequencies.

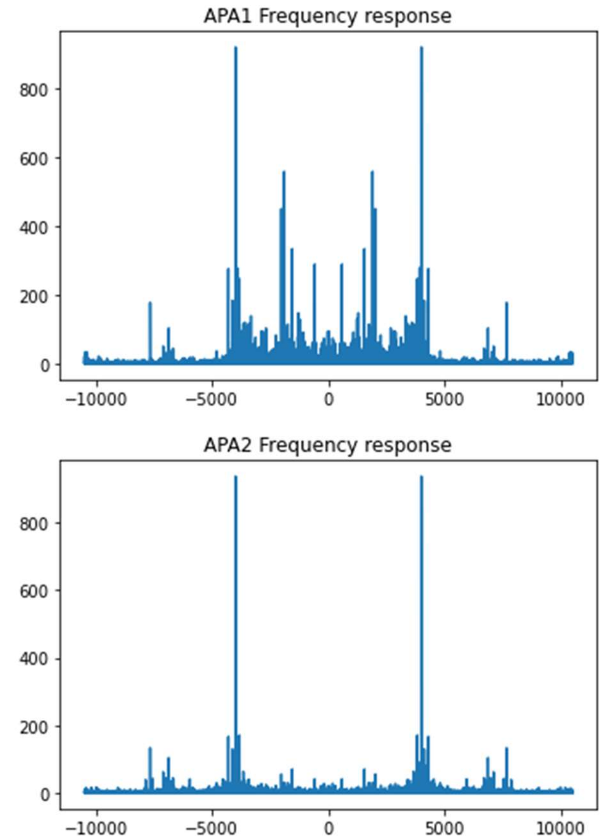

Figure 9: Optimal Filter Order Frequency Responses

The filters' input lengths were found by testing several commonly used values between 1 and 25. Interestingly, APA 2's best performing length was by far L=1, which resulted in a 30 dB SNR. Both filters performed well with an input length of L=10 and was APA 1's optimal input length.

```
APA1 ERLE SNR Improvement (dB), L = 1 : 8.760316800887626
APA2 ERLE SNR Improvement (dB), L = 1 : 30.409528346257268
APA1 ERLE SNR Improvement (dB), L = 5 : 8.135032872997197
APA2 ERLE SNR Improvement (dB), L = 5 : 4.778747377372045
APA1 ERLE SNR Improvement (dB), L = 10 : 15.13345257269631
APA2 ERLE SNR Improvement (dB), L = 10 : 14.796987250264015
APA1 ERLE SNR Improvement (dB), L = 15 : 4.282019342190497
APA2 ERLE SNR Improvement (dB), L = 15 : 5.50192051383113
APA1 ERLE SNR Improvement (dB), L = 25 : 10.211686716961824
APA2 ERLE SNR Improvement (dB), L = 25 : 11.456056114903978
```

Optimal filter length was found using the same process. When two values (such as 15 and 20) have a high SNR, the process is run again with values between those two, thus the testing of sizes 18 through 22 for APA 1. APA 1's optimal filter length was K=19, with an SNR of 30 dB. Interestingly, APA 2's optimal filter length was K=2, meaning that both the starting input length and filter length were optimal.

```
APA1 ERLE SNR Improvement (dB), K = 18 : 25.37318605087771
APA1 ERLE SNR Improvement (dB), K = 19 : 32.75842773929312
APA1 ERLE SNR Improvement (dB), K = 20 : 24.162449371750018
APA1 ERLE SNR Improvement (dB), K = 21 : 18.154681589367947
APA1 ERLE SNR Improvement (dB), K = 22 : 14.890173018855267
APA2 ERLE SNR Improvement (dB), K = 2 : 30.409528346257268
APA2 ERLE SNR Improvement (dB), K = 5 : 18.130099758713822
APA2 ERLE SNR Improvement (dB), K = 10 : 11.246480417584714
APA2 ERLE SNR Improvement (dB), K = 15 : 0.12708971702342511
APA2 ERLE SNR Improvement (dB), K = 20 : -0.8676858520671897
APA2 ERLE SNR Improvement (dB), K = 30 : 8.412332273744799
```

Epsilon was tested with a series of very small values. Because epsilon is a variable whose function is to allow the input matrix squared to be inverted, it should be very small so as not to affect the resulting value in any non-negligible manner. Thus, it was to be expected that lower values of epsilon will result in better performances.

```
APA2 ERLE SNR Improvement (dB), eps = 1e-06 : 30.409528346257268
APA2 ERLE SNR Improvement (dB), eps = 1e-05 : 17.47916642683491
APA2 ERLE SNR Improvement (dB), eps = 0.0001 : 11.020942957837622
APA2 ERLE SNR Improvement (dB), eps = 0.001 : 9.097515427519385
APA2 ERLE SNR Improvement (dB), eps = 0.01 : 8.398487771118257
APA2 ERLE SNR Improvement (dB), eps = 0.1 : 8.286379485138381
```

The filters experience an optimal SNR at a step size of about 0.001 and then experiences a steep decline as the step size increases. This is to be expected as for step sizes of about 0.01, the solution begins to diverge.

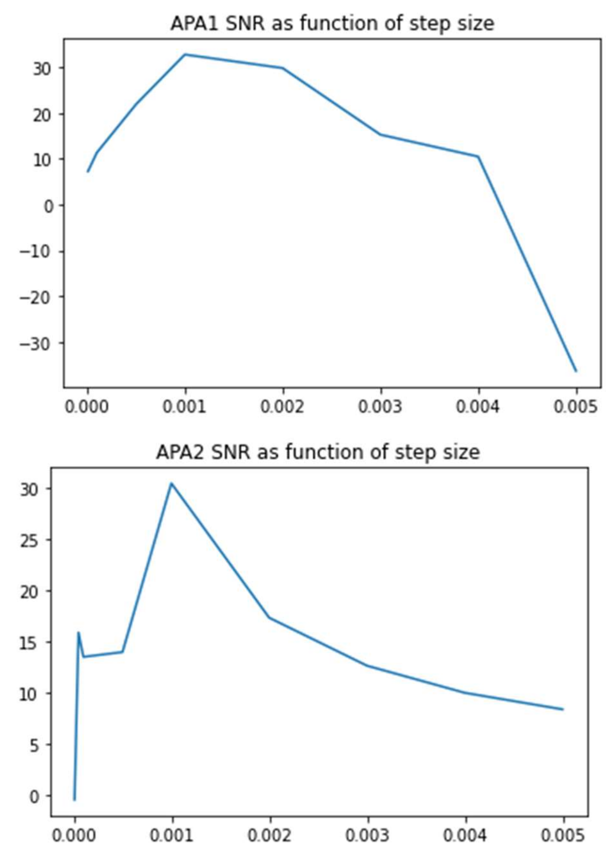

Figure 10: SNR as Function of Step Size

IV. Discussion

A. Performance and Computation Tradeoff Comparisons

Misadjustment is defined as the normalized excess MSE divided by the minimum MSE. For an FIR adaptive filter, the average excess MSE is equal to the input signal power multiplied by the sum of the variances of the weight noises [5]. In theory, the misadjustment of the APA class is independent of the number of input signal vectors used [6].

We experienced very high misadjustment values, which we attribute to the error being the filtered signal, rather than a true measurement of error. Because of these high values, we computed them on a logarithmic scale for better readability. In the future, an error measurement other than MSE should be taken for these types of problems to better reflect results. Nonetheless, the better performing filter had a smaller misadjustment, indicating a better performance.

APA1 output with optimal parameters
```
Misadjustment: 8.540982646538387
```

APA2 output with optimal parameters
```
Misadjustment: 9.777791002740045
```

Non-stationarity may arise in cases where the desired signal is time varying such as in system identification. This is because the autocorrelation function is time invariant, but the cross-correlation function is time varying. This may also occur when the input is non-stationary, for example, equalization or noise cancellation problems.

Generally, APA has advantages when the input signals are highly correlated. As the speech signal is non-stationary, we should not expect the learning curve to converge to a solution. We can compare the prediction error with major changes in a spectrogram i.e., abrupt changes in frequency during spoken words. As you proceed to the subsequent stages of the time series, the adaptive filters will quickly adapt to that region of the signal. Based on the results obtained, we can conclude that the rate of convergence increases as you increase the filter order.

Using the ERLE SNR and by playing the audio, we can reliably assess how well the four tested models did at filtering the audio. The APA 1 filter with optimal hyperparameters did the best at reducing the noise—though there is still significant noise—the speech "I will not condone a course of action that will lead us to war" is clearly heard. The next best performing were the APA 2 optimal and 2-tap filters, with the speech being heard, but difficult to discern what exactly is being said. The APA 1 2-tap filter performed the worse, with the speech barely being audible above the noise. These results correspond to the computed SNR for each filter.

Comparing the two algorithms with equal input and filter sizes (as was done in the 2-tap filter analysis), APA 2 is the more computationally expensive filter. The biggest contribution to this is the matrix inversion required for the online update, which for large input and filter sizes, can be quite computationally demanding. APA 1, on the other hand, only requires matrix multiplication.

However, with the optimal hyperparameters, APA 2 only had an input size of L=1 and a filter size of K=2, making it less computationally expensive. APA 1, on the other hand, required an input size of L=10 and a filter size of K=19. Regardless, APA 1 is the preferred method for filtering audio, due to the fact that the computational trade-off is not that significant (it does not require matrix inversion) and the fact that the speech was much clearer.

V. Conclusion

In this paper, we presented a solution to a noise reduction problem for speech signals using LMS based Affine Projection Algorithms. A comparative analysis was conducted to assess the performance of two adaptive filters within the APA family. We observed that for a given speech signal corrupted by noise, as the SNR increases and misadjustment decreases, the audio of the output signal is of higher quality. Based on the results we obtained, we can conclude that APA is effective in the task of interference canceling.

VI. References

[1] Samuel D. Steams, "Fundamentals of Adaptive Signal Processing" https://course.ece.cmu.edu/~ece792/handouts/LO_Chap_AdaptiveFilters.pdf

[2] Engineering Statistics Handbook https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd431.htm

[3] "Affine Projection Algorithms: Evolution to Smart and Fast Algorithm and Applications" http://www.eurasip.org/Proceedings/Eusipco/Eusipco2012/Conference/papers/1569579637.pdf

[4] Jose Principe, LMS and NLMS algorithms.pdf pg.10

[5] H. J. Butterweck, "A new interpretation of the misadjustment in adaptive filtering," 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, 1996, pp. 1641-1643 vol. 3, doi: 10.1109/ICASSP.1996.544119.

[6] "On Ways to Improve Adaptive Filter Performance" https://vtechworks.lib.vt.edu/handle/10919/30198