

Optimum Selection of Savitzky-Golay Filters

B Vrushabh

Summer Intern

Spectrum Lab, Department of Electrical Engineering

Indian Institute of Science, Bangalore

May 7, 2018-June 30, 2018

ACKNOWLEDGEMENTS

I would like to thank my guide, Dr.Chandra Sekhar Seelamantula for believing me and providing an opportunity to work in the Spectrum Lab, Indian Institute of Science, Bangalore.I would like to extend my gratitude towards my mentor Jishnu S for the help and encouragement he provided.

1 Introduction

Real-time signals are often corrupted with noise. Noise can be moving or stationary. In this paper we aim to smoothen the Noisy Signals to a Smooth signal. We try to employ filters to smoothen the data points and compute the approximate clean signal. The filter is based local least-square polynomial approximation and was formulated by the duo, Savitzky and Golay, hence is the name of these filters. Signal denoising is vital for further signal processing.

These Savitzky-Golay Filters (SG Filters) are finite impulse response, low pass, Digital filters. The algorithms expounded in [2] are verified for ECG signals taken from physio-net database. The duo further proved the fact that local least squares polynomial regression is equal to a discrete convolution with a fixed impulse response assuming that data samples are symmetric about the origin. Savitzky and Golay have shown that the shape of the signal is preserved and significant noise reduction is observed. SG filters are characterized by the Order(N) and Length of the filter(M). By choosing appropriate Length and Order, noise can be eliminated to a large extent. In this context we assume the clean signal $s[n]$, noise $w[n]$ which is Additive White Gaussian of zero mean and standard deviation of σ .

2 Background Theory

From [1], Consider an input signal $x[n]$. Considering $2M+1$ samples centred at $n=0$ at an instance we determine the polynomial coefficients, $p(n) = \sum_{k=0}^N a_k n^k$, where N is the order of the filter.

In the works of Savitzky and Golay they showed that the smoothed output value obtained by local least squares polynomial regression is identical to a fixed linear combination of the local set of impulse input samples, all the input impulse samples within the window length interval that is considered for fitting the polynomial.

Hence the output samples can be computed by discrete convolution,

$$y[n] = \sum_{m=-M}^M h[m]x[n-m]$$
$$y[n] = \sum_{m=n-M}^{n+M} h[n-m]x[m]$$

To prove that local ls polynomial regression is equal to discrete convolution of the noisy signal with finite impulse response, we need to minimize the cost function in order to determine optimal polynomial coefficients. The corresponding cost function is, $\sum_{n=-M}^M (p(n) - x[n])^2 = \epsilon_N$

Hence,

$$\frac{\partial \epsilon_N}{\partial a_i} = \sum_{n=-M}^M 2n^i \left(\sum_{k=0}^N a_k n^k - x[n] \right) = 0 \quad (1)$$

By rewriting the equation(1) we get,

$$\sum_{k=0}^N \sum_{n=-M}^M n^{i+k} a_k = \sum_{n=-M}^M n^i x[n] \quad (2)$$

These equations are known as normal equations for least squares approximation problem. We solve the problem by analyzing the equations in the form of matrix. We define a $(2M+1) \times (N+1)$ matrix A, given by

$$A = \{\alpha_{n,i}\}$$

$$\alpha_{n,i} = n^i, -M \leq n \leq M, i=0,1,\dots,N$$

The matrix A is called design matrix (It is popularly known as Vandermonde Matrix). We compute $A^T A$ which is a $(N+1) \times (N+1)$ symmetric matrix represented by B whose elements are,

$$\beta_{i,k} = \sum_{n=-M}^M n^{i+k} = \beta_{k,i}$$

If the input sequence is written as a vector $x = [x[-M], \dots, x[0], \dots, x[M]]^T$. The polynomial coefficients are written as a vector, $a = [a_0, a_1, \dots, a_N]^T$.

From equation 4, we can write,

$$Ba = A^T A a = A^T x$$

The polynomial coefficients are given by, $a = (A^T A)^{-1} A^T x = Hx$

We need to determine a_0 , which is the smoothed output value.

As $y[0] = a_0 = \sum_{m=-M}^M h_{0,m} x[m]$ where $h_{i,n}$ denotes the elements of the $(N+1) \times (2M+1)$ H matrix and $h_{0,m}$ is an element of the zeroth row.

B.GUE-MSE:

It is an estimation model which yields an unbiased estimate of the MSE for any given noise distribution. Unlike its counterparts, it requires the computation of variance of the additive noise. This model is applicable to noise of any category viz., Gaussian, Laplacian etc.

The estimate of the clean signal, $s[n]$ is s' , which is given by,

$$s' = AHx. \quad (3)$$

GUE-MSE for an SG Filter is given by ,

$$R' = 1/(2M+1) (AHX^2 - 2x^T H^T A^T x + 2\sigma^2 \sum_{l=-M}^M (H^T A^T)_{l,l} + x^2) - \sigma^2 \quad (4)$$

Regularized Risk is found by,

$$R_\lambda = R + \lambda/(2M+1)\epsilon\left\{\sum_{n=-M}^M L(s'_n(x))^2\right\} R_\lambda = R + \lambda/(2M+1)\epsilon\left\{\sum_{n=-M}^M \right\} \quad (5)$$

,where λ is the regularization parameter.

C.Selecting the Optimal Length SG Filter using GUE-MSE and Regularized GUE-MSE:

We need to construct an approximate of clean signal from the noisy signal. We need to determine the optimal length of the filter. As mentioned in[2],we follow the algorithm ,but compute the risk using GUE-MSE and Regularized GUE-MSE criterion.This involves employing a LS-fit of a polynomial locally and evaluating the same at the sampling instants within the observation interval which is the filtering action of SG filter.

We define a vector x_n such that every value of x is a centred value while analyzing by padding requisite number of zeros in the vector in x_n .A portion of this vector x_n is considered as one interval for employing LS fit and and evaluation of the Risk. M is the half filter length.'L' is the total length of the signal.We define a vector $M_{vec}=M_{min}:M_{max}$.The algorithm as mentioned in [2],is optimized with the idea of SURE algorithm.We employ i)GUE-MSE ii)Regularized GUE-MSE method to optimize and observe the optimum filter length for every interval of the input signal under consideration for a given order of the filter. $L_{max} = 2M_{max}+1$,given by the user.The steps of the algorithm are repeated for every interval,

Algorithm a:To calculate the Optimal Length L_{opt} for an interval:

Require: $M_{min} = \text{ceil}(0.5*(N+1))$

Ensure: Optimally with respect to GUE-MSE objective

for $i=1:L$

for $M=M_{min}:M_{max}$

$x_n = [\text{zeros}(1,M),x,\text{zeros}(1,M)]$

$WIND = x_n(1,i:2M+i)$

Employ LS fit over the vector WIND.

Evaluate the risk and compute the index at which minimum risk is found.

end for

$M_{opt} \leftarrow M_{vec}(\text{index})$

end for

We need a minimum of $N+1$ samples,else we will have an ill conditioned

system of equations in solving the LS problem. In [2], it is mentioned that for a higher polynomial order, M_{min} has to be increased otherwise the LS fit will not be accurate. The new signal can be constructed by using the Optimum Length which is found for which the risk is minimum, for a given order.

D. Selecting the Optimal Order of SG Filter using GUE:

We need to determine the optimal order of the filter for a given filter length. As mentioned in [2], we follow the algorithm, but compute the risk using GUE-MSE criterion. This is basically considering a window and finding the best polynomial order which yields the minimum MSE and that order is used to reconstruct the signal at the centre value of the window for a given value of M .

The signal is divided into windows of suitable length. The maximum order of the filter is given by the user. The minimum order is set to 1. The filter length should be set at least $Ord_{Max}+1$. We define a vector xn such that every value of x is a centred value while analyzing by padding requisite number of zeros in the vector in xn . A portion of this vector xn , is considered as one interval for employing LS fit and evaluation of the Risk. M is the half filter length. 'L' is the total length of the signal. We define a vector $Ordvec = Ord_{min}:Ord_{max}$. The algorithm as mentioned in [2], is optimized with the idea of SURE algorithm. We employ i) GUE-MSE ii) Regularized GUE-MSE method to optimize and observe the optimum filter length for every interval of the input signal under consideration for a given order of the filter. The steps of the algorithm are given below,

Algorithm b: To calculate the Optimal Order Ord_{opt} for an interval:

Require: $M = \text{ceil}(0.5*(N+1))$

Ensure: Optimally with respect to Regularized GUE-MSE objective

for $i=1:L$

$xn = [\text{zeros}(1,M), x, \text{zeros}(1,M)]$

for $Ord = Ord_{min}:Ord_{max}$

$WIND = xn(1,i:2M+i)$

Employ LS fit over the vector $WIND$.

Evaluate the risk and compute the index of the $Ordvec$ at which risk is minimum.

end for

Compute the index of the $Ordvec$ at which minimum risk is found.

$Ord_{opt} \leftarrow Ordvec(\text{index})$

end for

3 Experimental Results

The performance of the above algorithms is provided for the ECG signal sampled at 720Hz, taken from Physio Bank Database. The clean signal is corrupted by the addition of additive white Gaussian noise of mean 0 and standard deviation of σ .

A) Basic SG filter:

The noisy signal is passed through SG filter and following observations are made. Additive white Gaussian noise of 3dB is added to the clean speech signal. The value of M is set to 30 samples and order N is set equal to 4. The MSE, SNR of the output signal are computed and the results are shown below.

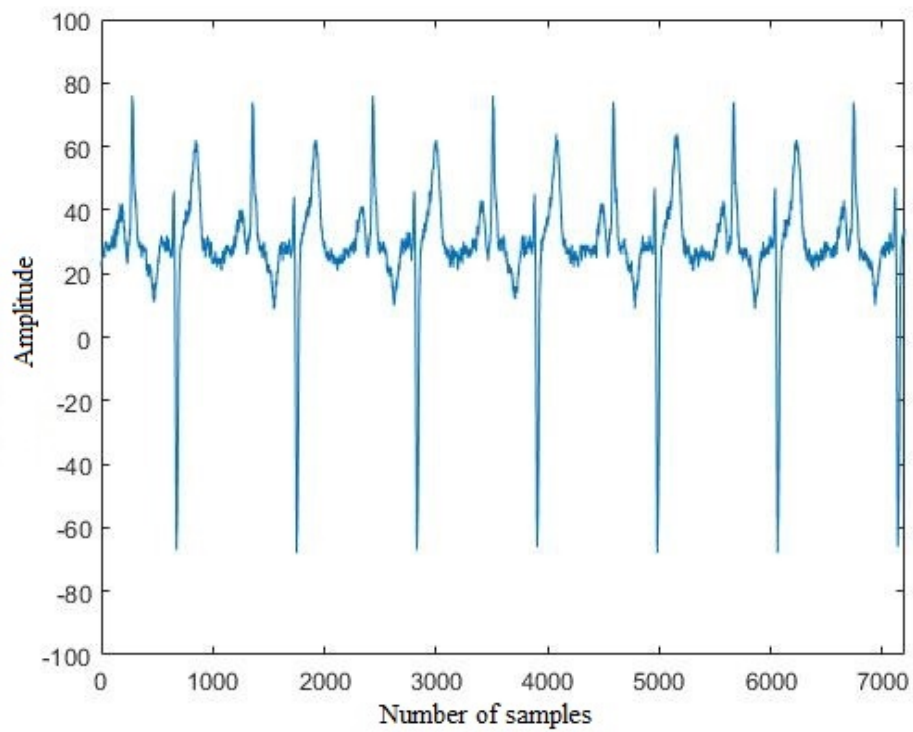


Figure 1: Clean Signal

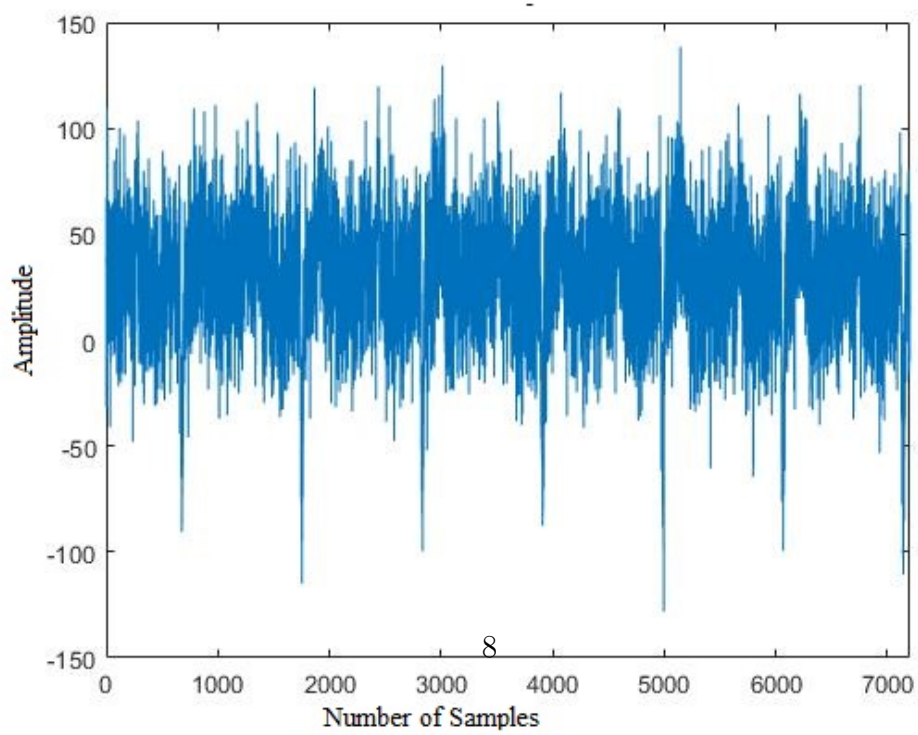


Figure 2: Noisy Signal

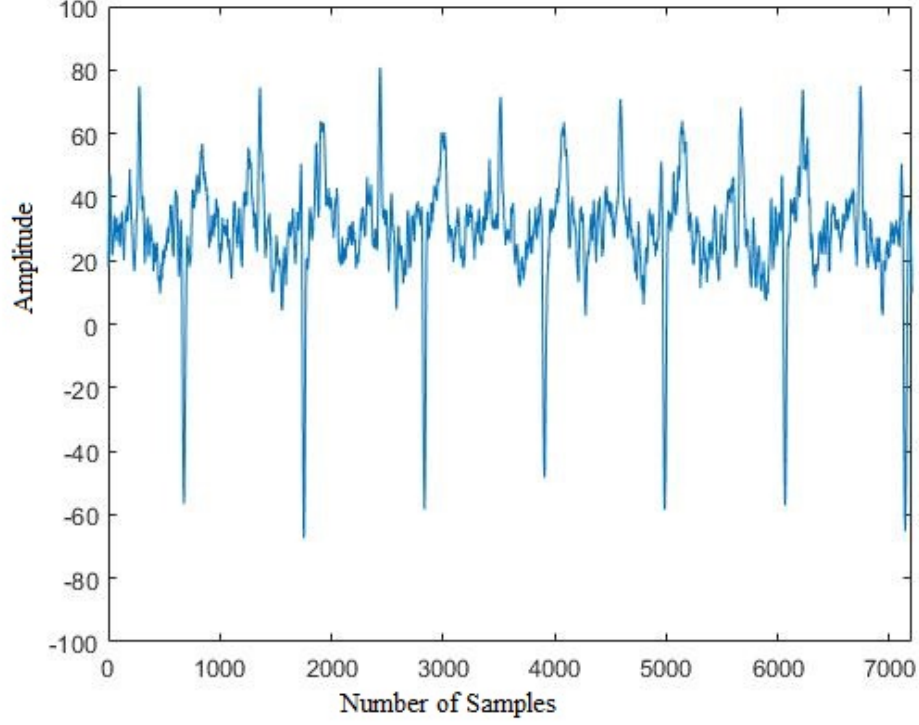


Figure 3: Recovered Signal

We compute the SNR of the recovered signal which gives a value of 16.97dB which is a tremendous amount of improvement against 3dB of the noisy signal. Mean Square Error which is the measure of how close the recovered signal is to that of the clean signal is measured in dB scale and the value is 14.5dB. Further, we plot the variation of MSE with the order, N for a given M and SNR of noise. For this plot M is set at 30 and order N , is varied from 1-6, with an additive Gaussian noise of 5dB. $B = \sum_{n=-M}^{n=M} \{\alpha_{i,n}\} \{\alpha_{n,i}\} = \alpha_{i,n}$

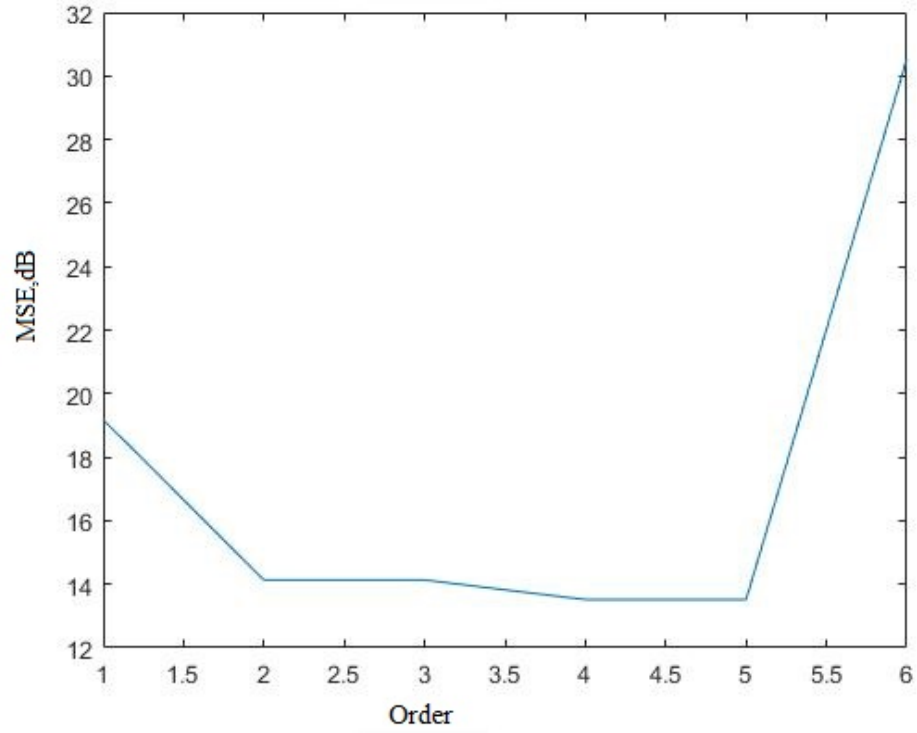


Figure 4: MSE vs Order

Next we plot variation of Output SNR with Input SNR, by setting the value of $M=3$ and $N=4$. Input SNR is varied from 3dB to 15dB

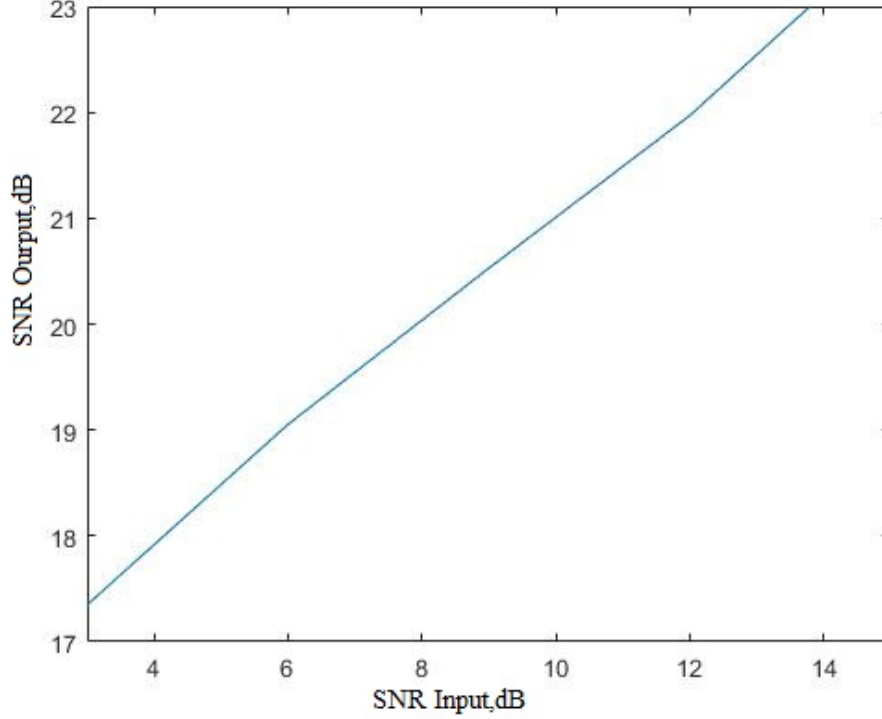


Figure 5: SNR OP vs SNR IN

B) Results with Optimum Filter Length Algorithm-a:

We vary the value of M from 3 to 30 and then reconstruct our signal. We set the order $N = 4$, and SNR of the input as 3dB. The noise has mean 0 and variance σ^2 . We plot the optimum length over the signal variation plot ie., we allow the algorithm to find the best fit of M for which the risk is minimum for that interval with the aid of Regularized GUE-MSE concept.

We find that SNR at the output is when Risk is evaluated by GUE-MSE criterion to be 12.65dB and that by Regularized GUE-MSE criterion is 13.68dB for regularization parameter, $\lambda = 1.2\sigma^2$. Hence there is an increase in SNR of the recovered signal with regularized GUE-MSE. This algorithm gives the optimum half window length M at each point of reconstruction. The plots for different estimation models are as follows,

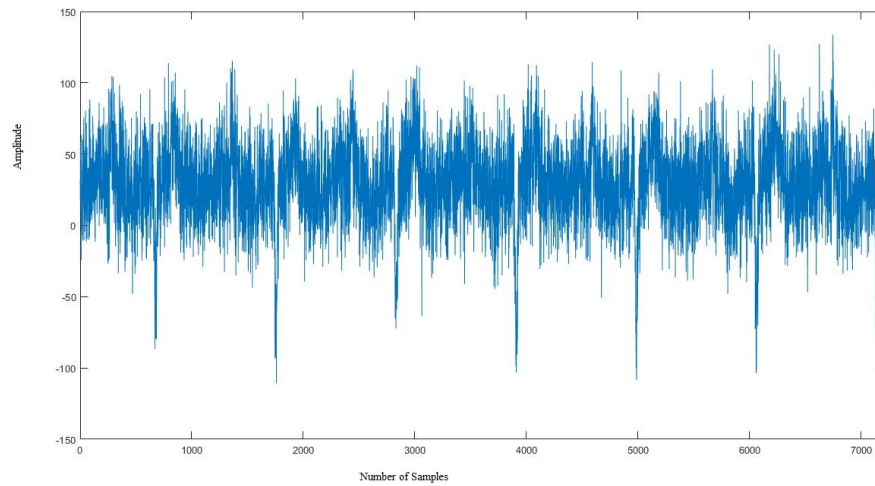


Figure 6: Noisy Signal

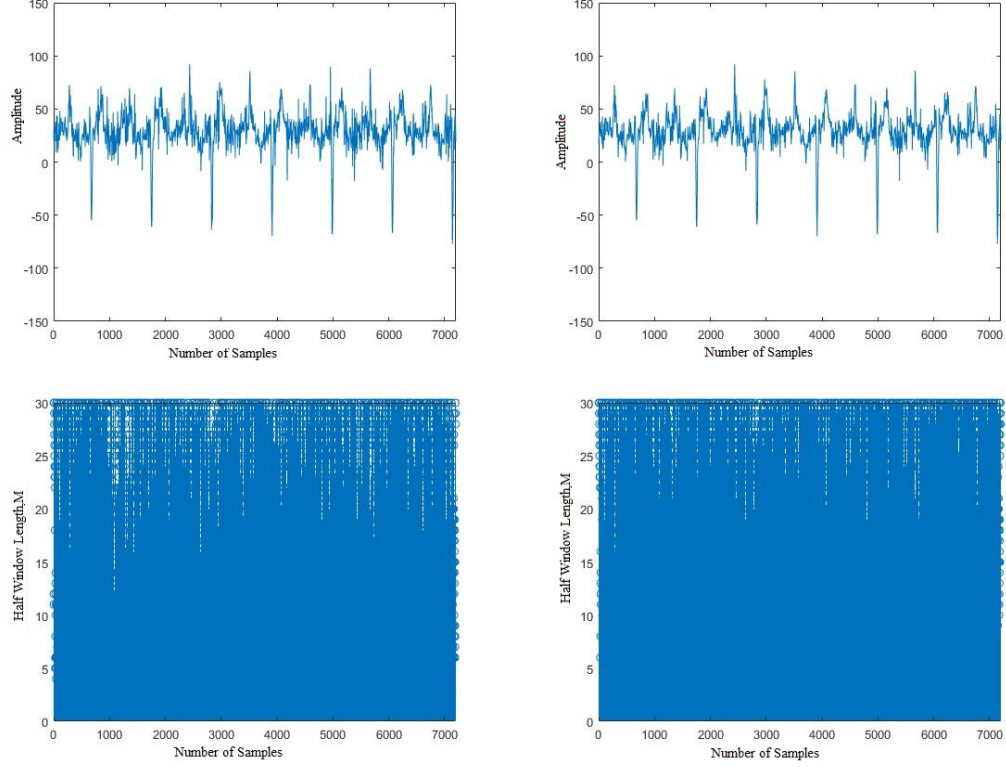


Figure 7: 1.Recovered Signal using GUE-MSE 2.Recovered Signal using Regularized GUE-MSE 3.Optimum M,GUE-MSE 4.Optimum M,Regularized GUE-MSE

We observe that the sites of clean signal at sharp transitions lower values of half window length ,M are attained and when the signal is smooth over a interval higher values of M are attained. B) Results with Optimum Filter Order Algorithm-b:

We vary the value of N from 1 to 6 and then reconstruct our signal point by point. We set the value of $M = 30$, and SNR of the input as 3dB.The noise has mean 0 and variance σ^2 .We plot the optimum length over the signal variation plot ie., with the value of M constant,the value of N is varied from Ord_{Min} to Ord_{Max} and the corresponding risk is computed by GUE-MSE and Regularized GUE-MSE criterion.

We find that SNR at the output is when Risk is evaluated by GUE-MSE criterion to be 15.94dB and that by Regularized GUE-MSE criterion

is 15.91dB, for regularization parameter, $\lambda = 1.2\sigma^2$. This algorithm successfully suppresses noise and maintains signal variations present in the clean signal. High orders viz., 5 and 6 where the variations are sharp. Hence the algorithm restores these sharp transitions. When the signal is relatively smooth the order is minimum, which ensures a high decrease in the noise components.

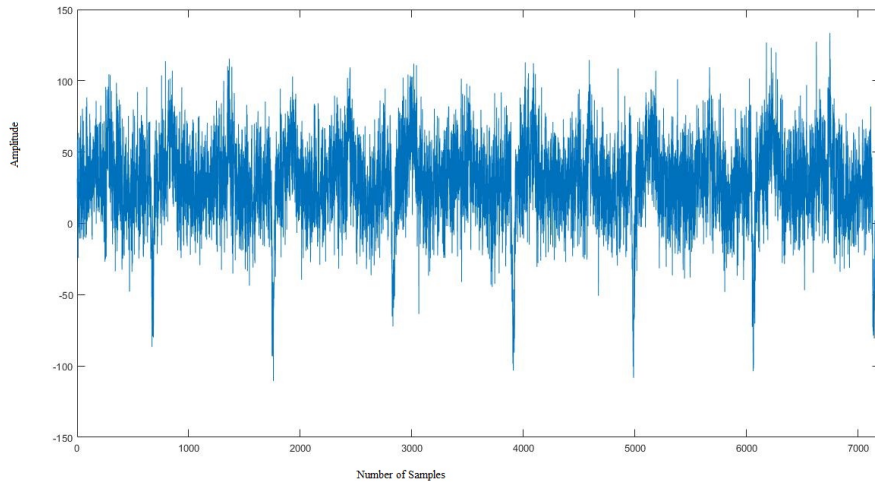


Figure 8: Noisy Signal

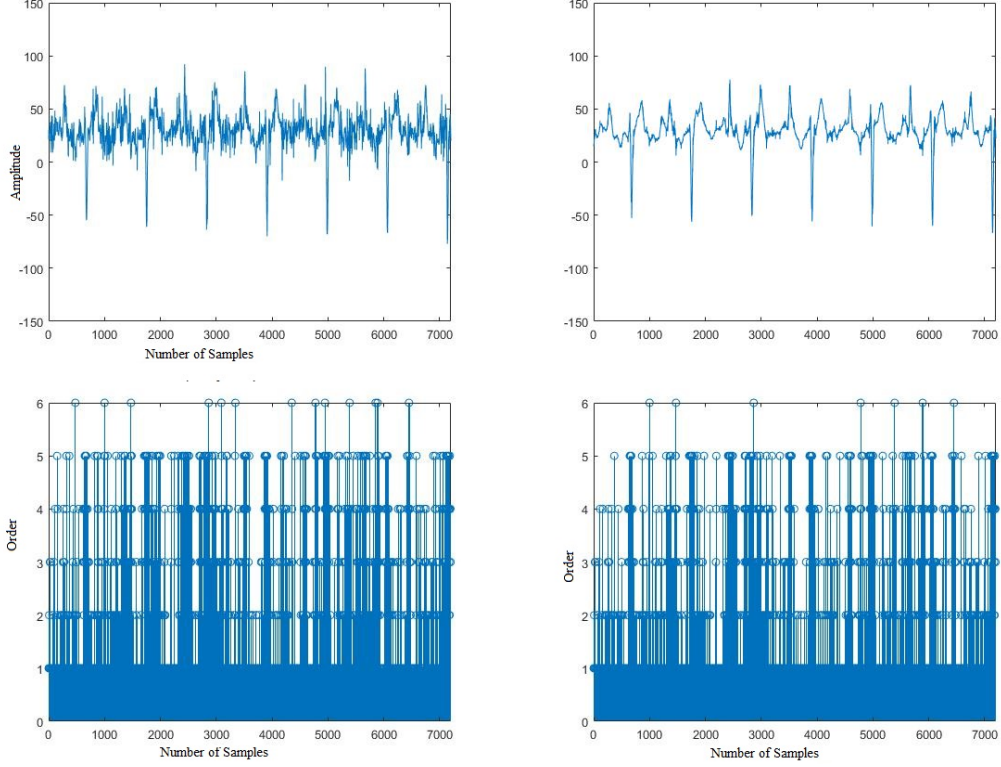


Figure 9: 1.Recovered Signal using GUE-MSE 2.Recovered Signal using Regularized GUE-MSE 3.Optimum Order,GUE-MSE 4.Optimum Order,Regularized GUE-MSE

4 Conclusion

From the above experimental observations, we can conclude that optimizing the order has a better performance than optimizing the window length for ECG signals which is evident from . We then extend the concept of SG filters and the Order ,Window length optimization to speech signals and then plot their corresponding spectrograms. The clean speech signal is corrupted with additive white Gaussian noise.The results and plots are as below for $M=150$ samples and $N=6$.The speech signal is taken from TIMIT Database.

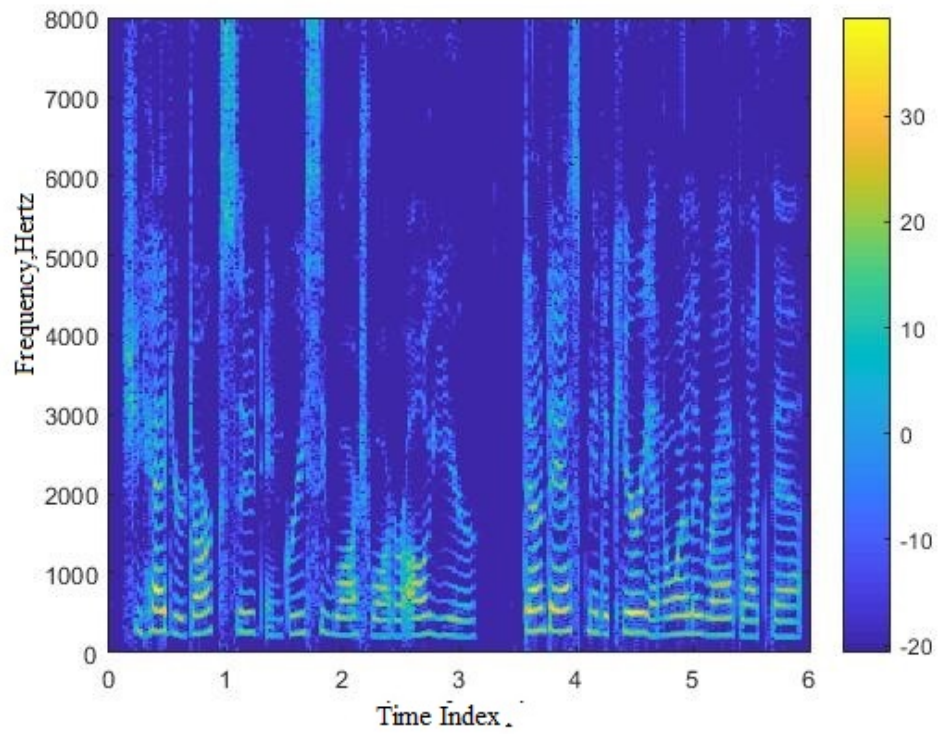


Figure 10: Spectrogram of clean speech signal

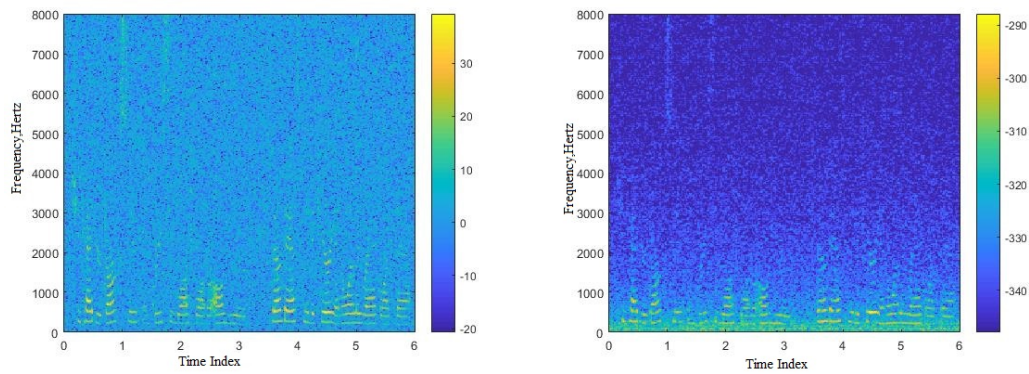


Figure 11: 1.Spectrogram of Noisy Signal 2.Spectrogram of Recovered Signal

From the spectrograms it is evident that the strength of the signal has

decreased significantly. The lower harmonics are dominant (The Fundamental Frequencies of words and alphabets). There is unclear partition between pauses and speech. Hence local least squares polynomial regression is not an ideal method for denoising speech signals. Spectrum subtraction followed by Local ls Polynomial fitting is known to give better results.

References

- [1] S. Hargittai. Savitzky-golay least-squares polynomial filters in ECG signal processing. In *Computers in Cardiology, 2005*. IEEE, 2005.
- [2] Sunder Ram Krishnan and Chandra Sekhar Seelamantula. On the selection of optimum savitzky-golay filters. *IEEE Transactions on Signal Processing*, 61(2):380–391, 2013.
- [3] Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, jul 1964.
- [4] Ronald Schafer. What is a savitzky-golay filter? [lecture notes]. *IEEE Signal Processing Magazine*, 28(4):111–117, 2011.