Contents

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

Multi Resolution Analysis using Approximation Basis derived from Orthogonal Polynomials

Nilotpal Pathak
Pratyush Garg

1 INTRODUCTION

Signal processing deals with methods to represent signals and extracting meaningful information from them. Transient, non-stationary signals have time varying frequencies and conventional tools like Fourier Transform prove inadequate to capture time-frequency information in these signals. A special time-frequency localized approximation basis is needed to capture information in both time and frequency domain and methods like Short Time Fourier Transform (STFT) and wavelet transform have been used for these applications. However, these methods are limited by the amount of computation they require and subsequently the time they take.

One advantage of the wavelet transform is that it offers the option of approximating the signal at various successive resolutions and hence view features that may not be trivially visible. This multiresolution analysis was first introduced in the context of signal processing by Stephane Mallat and Yves Meyer.

They showed strong interconnection between wavelet representation and multiresolution analysis. Ingrid Daubechies further explored the use of wavelets in signal processing. In the present work, we try to show that similar interconnection exists between orthogonal polynomial approximation and multiresolution analysis. R. Kumar and P. Sircar explored this interconnection and presented the multiresolution of ECG signals using orthogonal polynomials derived on discrete interval. Pradip Sircar, Ram Bilas Pachori, and Rupendra Kumar also separated the different rhythms of EEG signals using orthogonal polynomials and compared the results with wavelet packet transform. In both of these approaches, they did not provide mathematical formulation for the link between orthogonal polynomial approximation, multiresolution analysis and the wavelet transform. We try to bridge that gap in this work and use mathematical formulation to look at EEG signals at different resolutions.

In the subsequent sections, we first look at the approximation process and the mathematical preliminaries. Then we develop the multiresolution analysis framework with orthogonal polynomials as basis and finally analyze the results from simulation on an EEG dataset.

2 Approximation Theory and Chebyshev Polynomials

2.1 Linear Approximation

Real world signals are functions of continuous-time. But any physical device can only measure the signal on a discrete set of time. Approximation theory is concerned with how functions can be best approximated with simpler functions using known values of the the given function on some discrete-time points with quantitatively characterizing the errors introduced thereby.

Let $f(x)$ be a function which we wish to approximate using the class of functions. $\{g_j(x) : j = 1,2,3...M\}$. Let us denote this approximation by $\hat{f}(x)$, then the approximation is given by the equation:

$$f(x) \approx \hat{f}(x) = a_0 g_0(x) + a_1 g_1(x) + ... + a_M g_M(x) \tag{2.1}$$

such that $\{a_j(x) : j = 1,2,3...M\}$ are constants. This is the approximation of linear type. Let us denote the observed values of the function f on points $\{x_n : n = 0,1,2,...N-1\}$ as $\{f_n : f_n = f(x_n);\ n = 0,1,2,...N-1\}$ and approximated values of function as $\{\hat{f}_n : f_n \approx \hat{f}_n;\ n = 0,1,2,...N-1\}$. This gives us the discrete form of eqn 2.1

2.2 Orthogonal Basis Approximation and Linear Transforms

The above linear approximation equation can be written in a compact form as

$$\hat{\mathbf{f}} \approx \mathbf{f} = \mathbf{G}\,\mathbf{a} \tag{2.2}$$

where $\mathbf{f} = [f_0, f_1, ...f_{N-1}]^T$, $\hat{\mathbf{f}} = [\hat{f}_0, \hat{f}_1, ...\hat{f}_{N-1}]^T$, $\mathbf{a} = [a_0, a_1, ...a_M]^T$ and

$$\mathbf{G} = \begin{bmatrix} g_0(x_0) & g_1(x_0) & \dots & g_M(x_0) \\ g_0(x_1) & g_1(x_1) & \dots & g_M(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ g_0(x_{N-1}) & g_1(x_{N-1}) & \dots & g_M(x_{N-1}) \end{bmatrix}$$

Coefficient vector a can be calculated by least square solution to above equation. Solving we get:

$$\mathbf{a} = (\mathbf{G^T G})^{-1}\mathbf{G^T f} \tag{2.3}$$

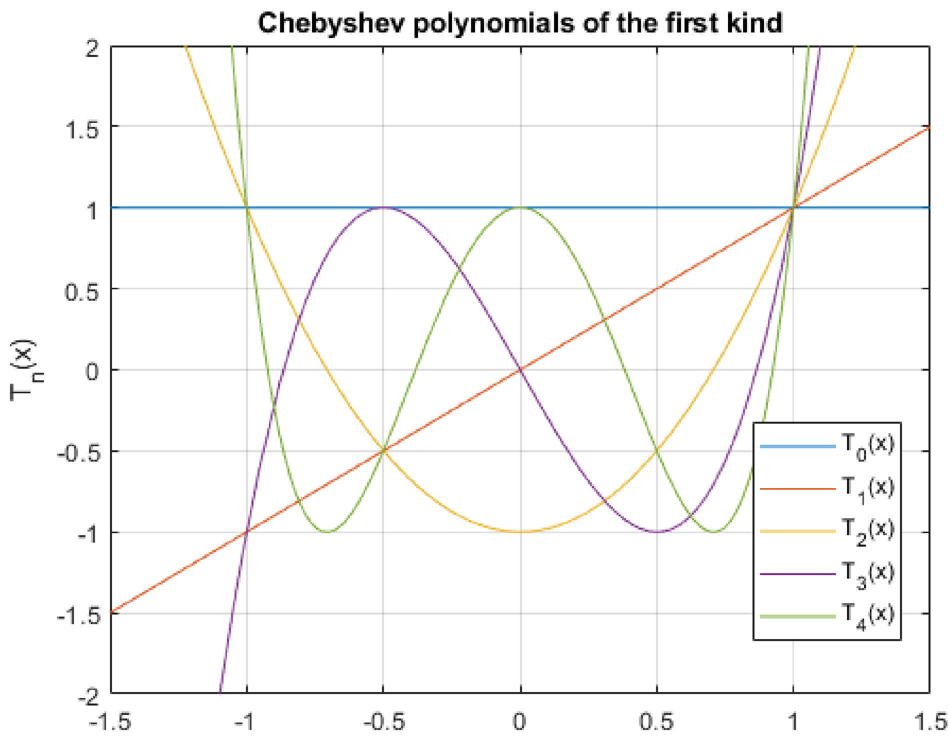
If the basis set/class of functions $\{g_j(x) : j = 1,2,3...M\}$ is orthogonal, then we have $\mathbf{G^T G} = \mathbf{Z}$, where $\mathbf{Z}$ is a diagonal matrix. If we include orthonormality as well, $\mathbf{Z} = \mathbf{I}$ and then the calculation of coefficients from eqn 2.3 reduces to:

$$\mathbf{a} = \mathbf{G^T f} \implies \hat{\mathbf{f}} = \mathbf{G G^T f} \tag{2.4}$$

2.3 Chebyshev Polynomials

Chebyshev Polynomials constitute a system of orthogonal polynomials on the interval $[a,b] = [-1,1]$. Chebyshev polynomials of the first kind $T_k(x)$ are orthogonal with respect to the weight function $w(x) = \frac{1}{\sqrt{1-x^2}}$ and Chebyshev polynomials of the second kind $U_k(x)$ are orthogonal with respect to weight function $w(x) = \sqrt{1-x^2}$. $T_k(x)$ and $U_k(x)$ are solutions to different differential equations and hence have certain properties associated with them. We will limit our analysis to polynomials of the first kind.

The first few polynomials are shown in the attached figure and can be defined using trigonometric formula as:



$$T_k(x) = cos(kcos^{-1}x) \tag{2.5}$$

This graph indicates that their absolute value in the interval $[-1, 1]$ is bounded by 1. We can also confirm that (as defined before) Chebyshev Polynomials are orthogonal with the given weight function. However, according to section 2.1 and 2.2, we would like them to be orthonormalized and discretized.

It turns out, Chebyshev Polynomials of the first kind satisfy discrete orthogonality if x_n's are zeros of $T_N(x)$ or (as they are commonly known) nodes of $T_N(x)$ given by the following equation:

$$x_n = cos\left(\frac{2n+1}{2N}\pi\right) : n = 0, 1, 2....N-1 \tag{2.6}$$

The corresponding orthogonality may be expressed by eqn 2.7 and with x_n's as defined in 2.6, we can get both, the desired discretization and orthonormalization.

$$(T_k(x), T_j(x))_d = \sum_{n=0}^{N-1} T_k(x_n) T_j(x_n) = \begin{cases} 0, & k \neq j \\ N, & k = j = 0 \\ N/2, & k = j \neq 0 \end{cases} \tag{2.7}$$

$$\implies ||T_k(x)||_d = (T_k(x), T_k(x))_d^{1/2} = \begin{cases} \sqrt{N}, & k = 0 \\ \sqrt{N/2}, & k \neq 0 \end{cases} \tag{2.8}$$

3 Multiresolution Analysis

3.1 Definition and Properties

Multiresolution Analysis is defined as a sequence of closed linear subspaces $\{\mathbf{V_j} \subset \mathbf{L^2}(\mathbf{R})\}$ and a sequence of approximating operators O_j. O_j is the approximation operator which approximates $f(x) \in \mathbf{L^2}(\mathbf{R})$ at resolution 2^{-j} . We would like this sequence of subspaces and operators to have some desired properties in order to process information in signals in a orderly and meaningful way. These properties as mentioned in Mallat are:

1. $O_j f(x)$ is the projection of f(x) on vector space $\mathbf{V_j} \subset \mathbf{L^2}(\mathbf{R})$. It is a linear operator and $O_j f(x)$ is not modified if we approximate it again at the same resolution 2^{-j}, i.e. $O_j \circ O_j = O_j$
2. Among all possible approximations of $f(x)$ at 2^{-j}, $O_j f(x)$ is the most similar to $f(x)$. Hence O_j is the orthogonal projection of $f(x)$ on the vector space $\mathbf{V_j}$.
3. The approximation of signal $O_j f(x)$ contains all the necessary information to compute the same signal at a smaller resolution $2^{-(j+1)}$. This is the causality property. Since O_j is the projector operator on $\mathbf{V_j}$, this principle is equivalent to $\forall j \in \mathbf{Z}, \mathbf{V_{j+1}} \subset \mathbf{V_j}$
4. The subspaces of approximated functions can be derived from one another by scaling each approximated function by the ratio of their resolution values. i.e. $\forall j \in \mathbf{Z}, f(x) \in \mathbf{V_{j+1}} \Leftrightarrow f(2x) \in \mathbf{V_j}$
5. When $f(x)$ is translated by some length proportional to 2_j , $O_j f(x)$ is translated by same amount and it is characterized by the same samples which have been translated.

6. As the resolution increases to 1, the approximated signal should converge to original signal. Conversely, as resolution decrease to 0, the approximated signal contains less and less information and converges to a constant value.

We call such a set of subspaces $\{\mathbf{V_j}\}_{j\in\mathbf{Z}}$ a multiresolution of $\mathbf{L^2}(\mathbf{R})$. This is very useful in many applications in signal and image processing.

3.2 Scaling Function and Basis of $\mathbf{V_j}$

In multiresolution analysis, approximation operator O_j is the orthogonal projection on the vector space $\mathbf{V_j}$. In order to numerically characterize this operator, we need to find an orthonormal basis of space $\mathbf{V_j}$. Mallat proved that such an orthonormal basis can be defined by translating and dilating a unique function called scaling function $\phi(x)$. Thus, the basis for $\mathbf{V_j}$ is given by the set:

$$\{\phi_j(x-2^j i_j) = \sqrt{2^{-j}}\phi(2^{-j}x - i_j) : \ i_j \in \mathbf{Z}\} \tag{3.1}$$

The orthogonal projection of $f(x)$ on space $\mathbf{V_j}$ can now be computed using above define orthonormal basis as: $\forall f(x) \in \mathbf{L^2}(\mathbf{R})$

$$O_j f(x) = \sum_{i_j=-\infty}^{\infty} \underbrace{\langle \phi_j(x-2^j i_j), f(x)\rangle}_{\text{Inner Product } a_{j,i_j}} \phi_j(x-2^j i_j) \tag{3.2}$$

Hence the inner product gives the coefficients $\{a_{j,i_j} : i_j \in \mathbf{Z}\}$ of the basis functions for approximation at resolution 2^{-j}.

4 MRA using Chebyshev Polynomials

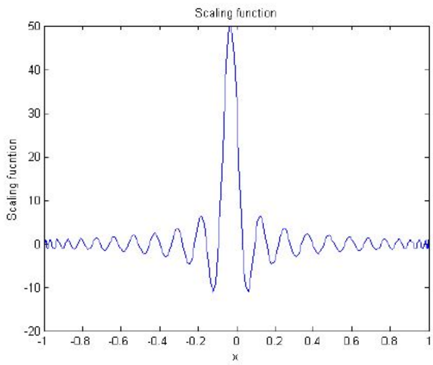
4.1 Christoffel Darboux Kernel and the Polynomial Scaling Function

Christoffel Darboux (CD) kernel polynomials and the polynomial scaling functions are defined as follows:

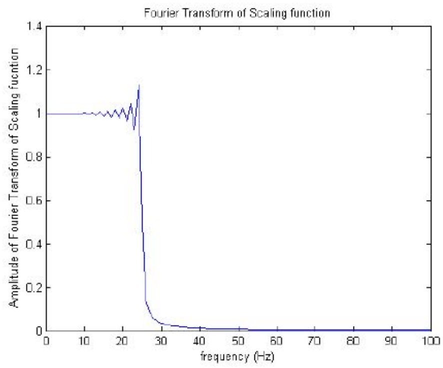
$$K_M(x,y) = \sum_{k=0}^{M} \frac{T_k(x)T_k(y)}{||T_k(x)||^2} = \sum_{k=0}^{M} P_k(x)P_k(y) \tag{4.1}$$

$$\phi_j(x, y) = K_{M_j}(x, y) \tag{4.2}$$

where $P_k(.)$ is the k^{th} order normalized Chebyshev polynomial of the first kind. Let us look at some important properties of the scaling function:



(a) Scaling function



(b) Fourier Transform of Scaling function

1. The inner product of $\phi_j(x, y_i)$ with $f(x)$ reproduces the value at $f(x = y_i)$. i.e. $\langle f(x), \phi_j(x, y_i)\rangle = f(y_i)$. This implies that the scaling function $\phi_j(x, y_i)$ **is localized around** $x = y_i$**.** (as shown in the attached figure)

2. The fourier transform of the scaling function looks like **a bandpass filter**. (as shown in the attached figure) This is consistent with the idea that to get different resolutions, we have to limit the frequency. We talk more about this in subsequent sections.

3. Let $\{y_{j,i_j} : i_j = 0, 1, 2, ..., M_j\}$ be an **arbitrary** parameter set, then the set $\{\phi_j(x, y_i) : i_j = 0, 1, 2, ..., M_j\}$ in itself is **not orthogonal**. The y_{j,i_j}'s need to be the **nodes of** $P_N(.)$ (as explained in section 2.3, eqn 2.6).

4. Chebyshev Polynomials are **orthogonal only in the domain** $[-1, 1]$. Often times we want to approximate functions $f(x)$ on a general interval $[a, b]$, so we require **a change in variable** and shift of domain.

With these observations, we can move on to the algorithm and matrix formulation for the approximation.

4.2 MATRIX FORMULATION

As given in section 2.2, the approximated $\hat{f}$ can be calculated easily from the matrix G as $\hat{\mathbf{f}} = \mathbf{G}\mathbf{G}^{\mathbf{T}}\mathbf{f}$. Here the matrix G is the discretized form of the basis set at N points $\{x_n : n = 0, 1, 2, ..., N-1\}$. The basis set for $\mathbf{V_j}$ is $\{\phi_j(x, y_i) : i_j = 0, 1, 2, ..., M_j\}$, i.e.

$$\mathbf{V_j} = \text{span}\ \{\phi_j(x, y_0),\ \phi_j(x, y_1), ...\phi_j(x, y_{M_j})\} \tag{4.3}$$

Therefore, the corresponding G matrix here will be formed by the above defined set and consequently called $\Phi_{\mathbf{j}}$. The definition follows:

$$\Phi_{\mathbf{j}} = \mathbf{P_j}\mathbf{A_j^T} \tag{4.4}$$

$$\text{where, } \mathbf{P_j} = \begin{bmatrix} P_0(x_0) & P_1(x_0) & \dots & P_{M_j}(x_0) \\ P_0(x_1) & P_1(x_1) & \dots & P_{M_j}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_{N-1}) & P_1(x_{N-1}) & \dots & P_{M_j}(x_{N-1}) \end{bmatrix}$$

$$\mathbf{A_j} = \begin{bmatrix} P_0(y_0) & P_1(y_0) & \dots & P_{M_j}(y_0) \\ P_0(y_1) & P_1(y_1) & \dots & P_{M_j}(y_1) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(y_{M_j}) & P_1(y_{M_j}) & \dots & P_{M_j}(y_{M_j}) \end{bmatrix} \tag{4.5}$$

$$\Phi_{\mathbf{j}} = \begin{bmatrix} \phi_j(x_0, y_0) & \phi_j(x_0, y_1) & \dots & \phi_j(x_0, y_{M_j}) \\ \phi_j(x_1, y_0) & \phi_j(x_1, y_1) & \dots & \phi_j(x_1, y_{M_j}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_j(x_{N-1}, y_0) & \phi_j(x_{N-1}, y_1) & \dots & \phi_j(x_{N-1}, y_{M_j}) \end{bmatrix}$$

Let us also define the space $\mathbf{W_j} = \mathbf{V_{j-1}} \ominus \mathbf{V_j} = \text{span}\ \{P_{M_j+1}, P_{M_j+2}, ...P_{M_{j-1}}\}$ The associated basis will have a wavelet function corresponding to the scaling function before; as given by: $\psi_j(x, z) = \sum_{M_j+1}^{M_{j-1}} P_k(x) P_k(z_{j,i})$ In matrix form:

$$\Psi_{\mathbf{j}} = \mathbf{Q_j}\mathbf{B_j^T} \tag{4.6}$$

$$\text{where, } \mathbf{Q_j} = \begin{bmatrix} P_{M_j+1}(x_0) & P_{M_j+2}(x_0) & \dots & P_{M_{j-1}}(x_0) \\ P_{M_j+1}(x_1) & P_{M_j+2}(x_1) & \dots & P_{M_{j-1}}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ P_{M_j+1}(x_{N-1}) & P_{M_j+2}(x_{N-1}) & \dots & P_{M_{j-1}}(x_{N-1}) \end{bmatrix}$$

$$\mathbf{B_j} = \begin{bmatrix} P_{M_j+1}(z_0) & P_{M_j+2}(z_0) & \dots & P_{M_{j-1}}(z_0) \\ P_{M_j+1}(z_1) & P_{M_j+2}(z_1) & \dots & P_{M_{j-1}}(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ P_{M_j+1}(z_{N-1}) & P_{M_j+2}(z_{N-1}) & \dots & P_{M_{j-1}}(z_{N-1}) \end{bmatrix} \tag{4.7}$$

$$\mathbf{\Psi_j} = \begin{bmatrix} \psi_j(x_0,z_0) & \psi_j(x_0,z_1) & \dots & \psi_j(x_0,z_{M_j-1}) \\ \psi_j(x_1,z_0) & \psi_j(x_1,z_1) & \dots & \psi_j(x_1,z_{M_j-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_j(x_{N-1},z_0) & \psi_j(x_{N-1},z_1) & \dots & \psi_j(x_{N-1},z_{M_j-1}) \end{bmatrix}$$

Now let us revisit section 3.2 and what the approximation and detail signals actually are using the polynomial technique.

$$\hat{f}_j(x) = O_j f(x) = \sum_{i_j=0}^{M_j} \underbrace{\langle \phi_j(x, y_{i_j}), f(x) \rangle}_{a_{j,i_j}} \phi_j(x, y_{i_j}) \tag{4.8}$$

$$\hat{r}_j(x) = D_j f(x) = \sum_{i_j=0}^{M_j-1} \underbrace{\langle \psi_j(x, z_{i_j}), f(x) \rangle}_{d_{j,i_J}} \psi_j(x, z_{i_j}) \tag{4.9}$$

Finally, let us define the mirror filters for the two scale refinement equations:

$$\mathbf{H_{j-1}} = \mathbf{\Phi_{j-1}^T \Phi_j} = \mathbf{A_{j-1} P_{j-1}^T P_j A_j^T} \tag{4.10}$$

$$\mathbf{G_{j-1}} = \mathbf{\Phi_{j-1}^T \Psi_j} = \mathbf{A_{j-1} P_{j-1}^T Q_j B_j^T} \tag{4.11}$$

Now we can go from a lower resolution to a higher one and vice-versa with the following formulas. Equations 4.12, 4.14 go to a lower (less points) resolution whereas eqns 4.13, 4.15 go to a higher resolution and hence require details added to the lower approximation.

$$\mathbf{\Phi_j = \Phi_{j-1} H_{j-1}} \text{ and } \mathbf{\Psi_j = \Phi_{j-1} G_{j-1}} \tag{4.12}$$

$$\mathbf{\Phi_{j-1} = \Phi_j H_j^* + \Psi_j G_j^*} \tag{4.13}$$

$$\mathbf{a_j = \Phi_j^T f = H_{j-1}^T a_{j-1}} \text{ and } \mathbf{d_j = \Psi_j^T f = G_{j-1}^T a_{j-1}} \tag{4.14}$$

$$\mathbf{a_{j-1} = H_{j-1} a_j + G_{j-1} d_j} \tag{4.15}$$

$$\mathbf{\hat{f}_j = O_j f = \Phi_j\, a_j = \Phi_j\, \Phi_j^T f} \tag{4.16}$$

In our simulations, we focused on approximation at a given resolution. Our approach involved calculating the matrix Φ_j using eqns 4.4, 4.5 and finally eqn 4.16 to get the signal at the desired resolution.

4.3 Nyquist Criterion and Aliasing

Let function $f(t)$ is a bandlimited signal with bandwidth B. Sampling Theorem states that to be able to perfectly reconstruct $f(t)$ from its N samples sampled at sampling frequency ν_s, we need $B \leq \nu_s/2$. The error of approximation goes to zero when we choose an M_j for the polynomial such that:

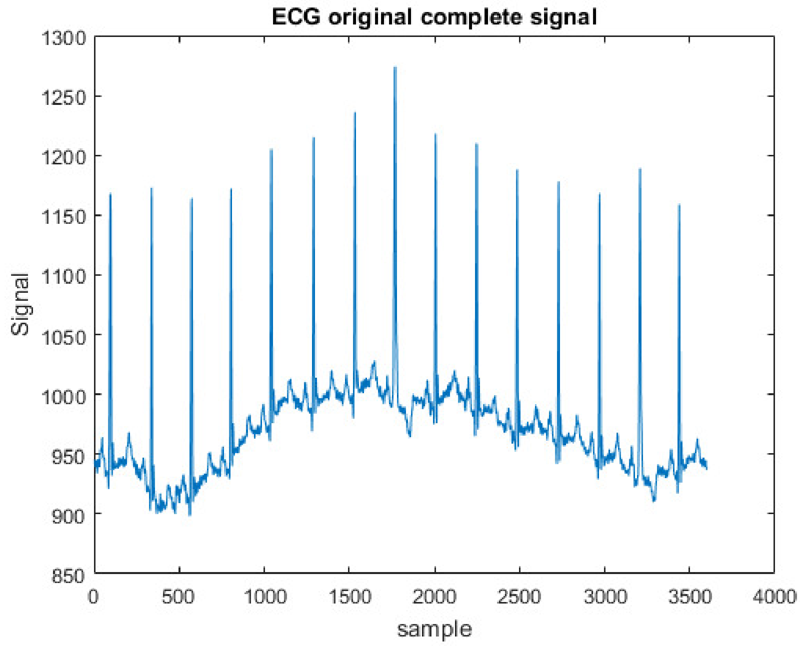
$$B = \frac{M_j \nu_s}{2N} \implies \frac{M_j \nu_s}{2N} \leq \frac{\nu_s}{2} \implies M_j \leq N \tag{4.17}$$

Hence, we set $M_0 = N - 1$ to get the best results within the Nyquist limit.

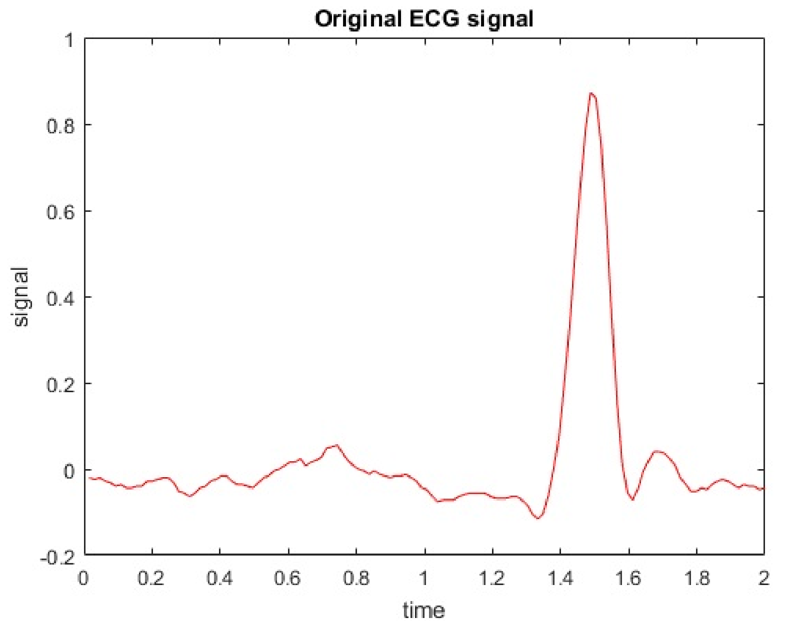
5 Results

5.1 ECG Signals

For the purpose of testing our hypotheses, we have chosen the MIT-BIH Arrhythmia database of ECG signals. The sampling frequency is 360 Hz. The original samples are 10 seconds long (3600 samples). For simplicity, we will run our analysis on 129 points only.



After reducing to 129 points, the signal reduces to the following:



The approximations from the algorithm at various resolutions are given.

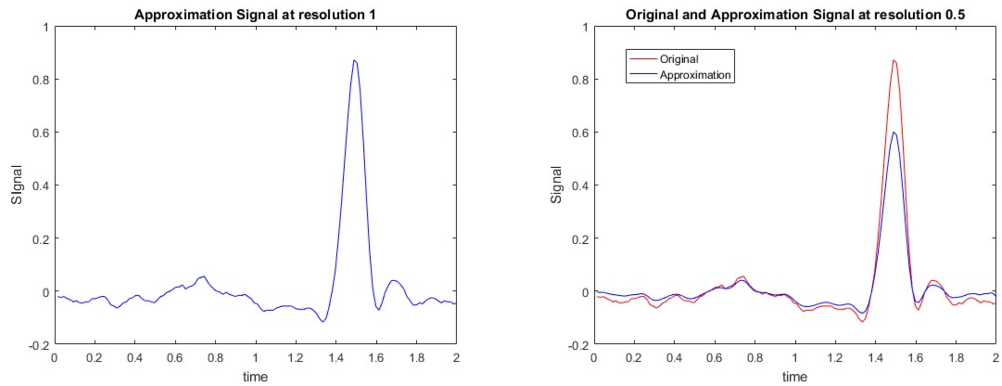
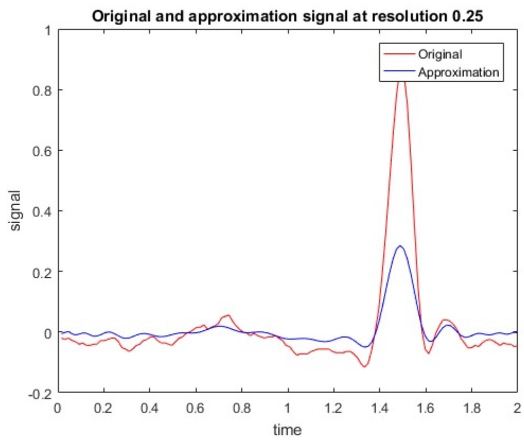


Figure 5.1: Approximations of ECG signal at various resolutions



5.2 EEG SIGNALS

Here, we have used a dataset made public by Bonn university. The rate of sampling is 173.61 Hz and we are only looking at 129 samples as in the earlier case. We will look at two EEG signals in our analysis.

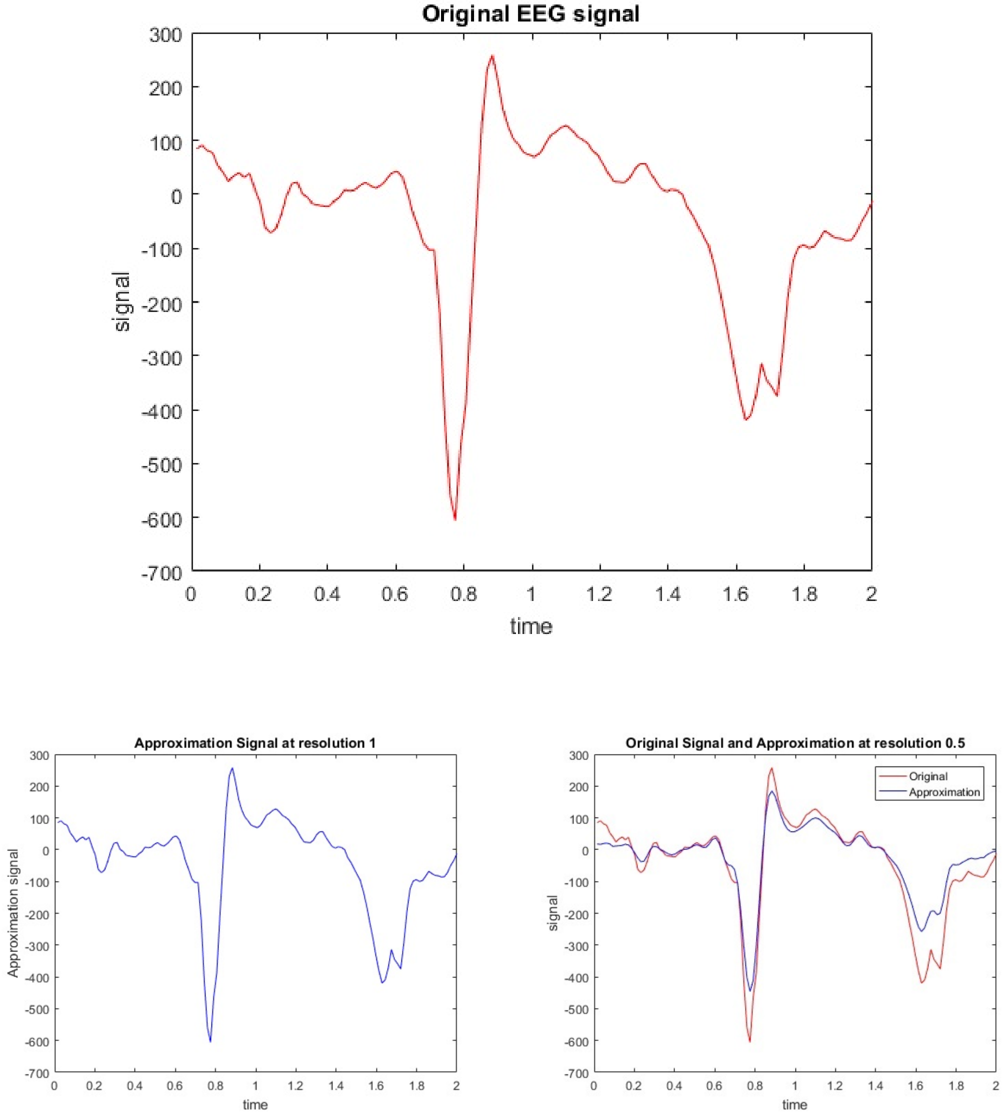
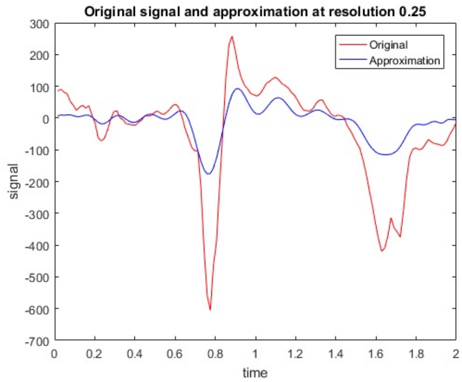
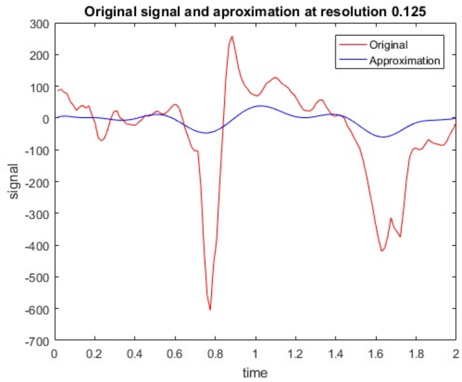
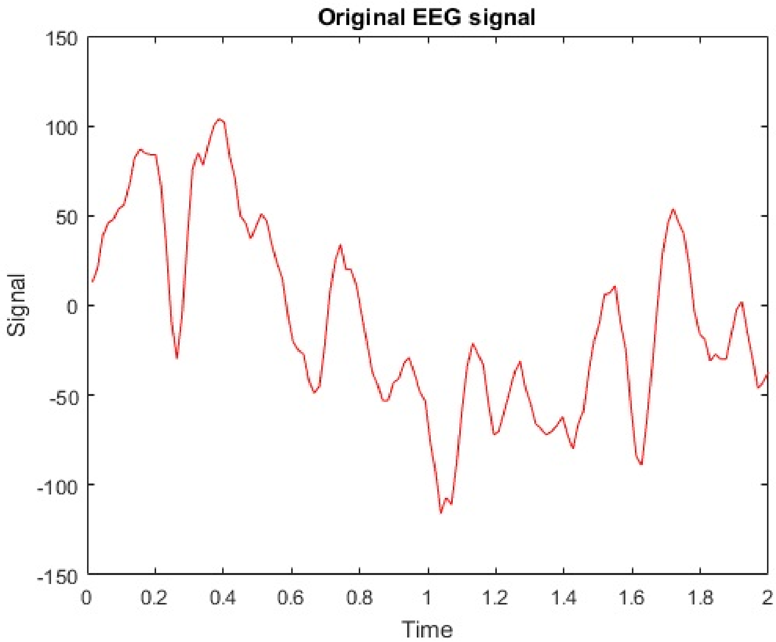


Figure 5.2: Approximations of given EEG signal at various resolutions





We will analyze another EEG signal with this technique. The original function and its approximations at various resolutions follow.



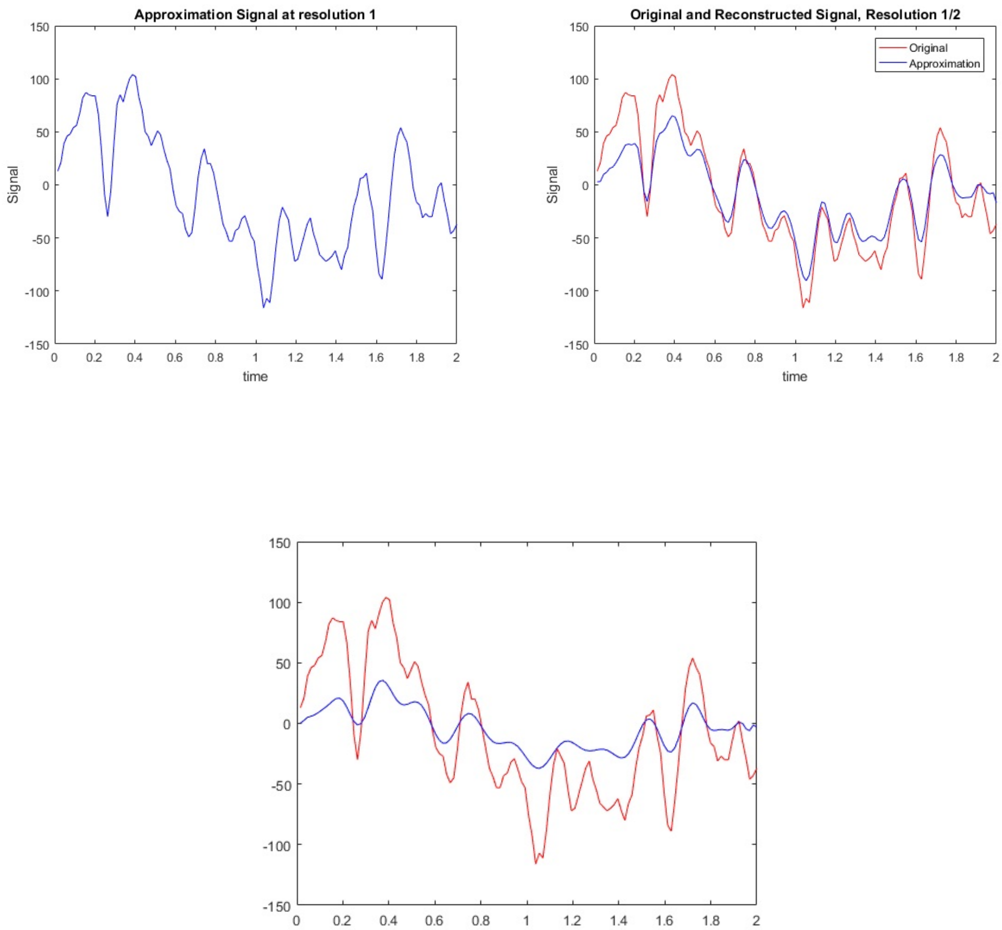


Figure 5.3: Approximation Signal at Resolution 1/4

6 Acknowledgement

We take this opportunity to express our sincere gratitude to Dr. Pradip Sircar for giving us a chance to work in this exciting field. His constant guidance and foresight was the key to the successful completion of this work. Also, his course on Wavelet Transforms helped establish the basic concepts of the UG project and gave us a good foundation to pursue other ideas.

We firmly believe that this project has given us insight into the world of research and inspired us to continue striving even in the face of adversity.

Nilotpal Pathak
Pratyush Garg

7 References

[1] S. G. Mallat. "A theory for multiresolution signal decomposition: the wavelet representation." in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693, Jul 1989. doi: 10.1109/34.192463.

[2] P. Sircar, K. Prasad and B. Harshavardhan. "Analysis of multicomponent speech-like signals by continuous wavelet transform-based technique." Signal Processing Conference, 2006 14th European, Florence, 2006, pp. 1-5.

[3] R. Kumar and P. Sircar. "Multiresolution analysis using orthogonal polynomial approxima- tion." European Signal Processing Conference, 1996. EUSIPCO 1996. 8th, Trieste, Italy, 1996, pp. 1-3.

[4] Pradip Sircar, Ram Bilas Pachori, and Rupendra Kumar. "Analysis of rhythms of EEG signals using orthogonal polynomial approximation." in Proceedings of the 2009 International Conference on Hybrid Information Technology (ICHIT '09). ACM, New York, NY, USA, 176-180

[5] Stephane Mallat. "A Wavelet Tour of Signal Processing". ISBN: 978-0-12-374370-1 .