A method for the Local Prediction for the ECG Data

6.1 Introduction:

We have a time series $X$ of the ECG signal of a healthy male. This data is collected at the sampling rate of 1024 Hz (courtesy of Dr. Pradhan, NIMHANS). In figure 6.1, we show a graph in which the horizontal axis represents time and the vertical axis represents the value of the ECG signal at different points of time.

There are 649 to 787 data points between any two consecutive peaks. We divide the data set into overlapping sections in such a way that each section contains only
one peak. Then in each section we calculate the highest value of $X$. This generates a list of the peak locations of $X$. Let $E$ denote the peak locations. The vertical green lines indicate the peak locations in figure 6.2.

$$\text{Figure 6.2: Locations of the peaks in the ECG signal}$$

Since the sampling rate of $X$ is 1024 per second, its sampling interval is $1/h = 0.0009765$ second.

### 6.2 Embedding in 2 dimensional complex domain ($C^2$):

We construct a Takens matrix (as described in 2.2) from the data series $X$. Let us denote this matrix by $H$. We choose 99868 rows and 100 columns in this matrix. So we embed the data series in 100 dimensions by creating the time delay vectors $(x_0, x_1, ..., x_{99})$, $(x_1, x_2, ..., x_{100})$ and so on.

$$H = \begin{pmatrix}
    x_0 & x_1 & \ldots & x_{99} \\
    x_1 & x_2 & \ldots & x_{100} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{99768} & x_{99769} & \ldots & x_{99867}
\end{pmatrix}$$

We construct a specific filter which maps each time delay vector into a point in the 2-dimensional complex domain ($C^2$). First, we find the Fourier Transform of the data series $X$. Then, we find that the first two dominating peaks in the transform. 142 and 426 are the bin numbers corresponding to the dominating peaks and $\omega_1 = 9.136$ Hz and $\omega_2 = 27.4$ Hz are the corresponding angular frequencies. Figure 6.3 shows the amplitude of the Fourier components with respect to time. Let us consider the 100 by 2 matrix $A$. An $n$th element in the first column of $A$ is of the form $-e^{i\omega_1 nh}/100$, an $n$th element in the second column of $A$ is of the form $-e^{i\omega_2 nh}/100$, $n$ is the row number, $h$ is the sampling interval, $\omega_1$ and $\omega_2$ are defined above.
6.2. Embedding in 2 dimensional complex domain ($C^2$):

\[ A = \begin{pmatrix}
1 & 1 \\
e^{-i\omega_1 h}/100 & e^{-i\omega_2 h}/100 \\
\vdots & \vdots \\
e^{-99i\omega_1 h}/100 & e^{-99i\omega_2 h}/100
\end{pmatrix} \]

Let us consider the relationship $K = HA$.

\[ K = \begin{pmatrix}
x_0 & x_1 & \ldots & x_{99} \\
x_1 & x_2 & \ldots & x_{100} \\
\vdots & \vdots & \ddots & \vdots \\
x_{99768} & x_{99769} & \ldots & x_{99867}
\end{pmatrix} \begin{pmatrix}
1 & 1 \\
e^{-i\omega_1 h}/100 & e^{-i\omega_2 h}/100 \\
\vdots & \vdots \\
e^{-99i\omega_1 h}/100 & e^{-99i\omega_2 h}/100
\end{pmatrix} \]

$K$ has 99868 rows and 2 columns. The above relationship converts each of the Takens vectors of $H$ into a two dimensional vector in the complex domain. The significance of this transformation is that it filters in the most influential frequencies of the signal. Also it embeds the real data series $X$ in a 2-dimensional complex domain $(C_0, C_1)$, which can be treated as a 4-dimensional real time series $(x, y, z, u)$, where $x$ and $y$ are the real and imaginary parts of $C_0$ respectively, $z$ and $u$ are the real and imaginary parts of $C_1$ respectively. If we plot $x$ against $y$ we get the following state space picture.

Figure 6.3: Fourier transform of the ECG signal
6.3 Creating Equivalence Classes in the data:

In section 6.1 we identified the location of the peaks. Based on the distance from the nearest peak, we assign a value to every point in the time series. We call the distance from the previous peak as the 'AD' value and the distance to the next peak as the 'BC' value. In figure 6.5, we assume that $P_1$ and $P_2$ are two consecutive peaks and $A$ is a point in the data set. The distance between $P_1$ and $A$ is called the AD value of $A$ and the distance between $A$ and $P_2$ is called the BC value of $A$.
6.4. Finding the Local Affine Map: 61

Now we create an equivalence class or fiber in the data series based on the value of AD or BC. Each such class is denoted by the symbol \( N(w) \). For \( w = 0 \) to 249, \( N(w) \) consists of all the data vectors, which are \( (250 - w) \) units apart from the next peak. In other words, it is a collection of all the data points for which the BC value is \( (250 - w) \). For \( w = 250 \) to 700, \( N(w) \) is a collection of all the data vectors which are \( (w - 250) \) unit apart from the previous peak, i.e. the AD value of each point is \( (w - 250) \).

In figure 6.6, we plot the values of the ECG signal \( X \) with respect to time. The vertical red lines indicate the points of \( N(w) \) and the green lines indicate the points of \( N(w + 1) \).

![Figure 6.6: Locations of two consecutive points in the ECG signal](image)

6.4 Finding the Local Affine Map:

In this section we show that there exists an affine transformation between any two consecutive fibers. Therefore if we know the parameters of these affine transformations, \( N(w + 1) \) can be predicted from \( N(w) \) for any \( w \). Both \( N(w) \) and \( N(w + 1) \) have 4 columns, since our filtered data set \( (x, y, z, u) \) is 4-dimensional. Each affine transformation is of the following form:

\[
T(w)N(w) + affine(w) = N(w + 1)
\]

where \( T \) is a time-dependent \( 4 \times 4 \) matrix, \( N \) is the state vector at time step \( n \) (the precise time is \( n/1024 \) seconds) and ”affine” is a \( 4 \times 1 \) vector which depends upon time. Let us describe the method. We subtract the row mean from \( N(w) \) and \( N(w + 1) \) and we denote the new matrices as \( A(w) \) and \( B(w) \). We compute the generalized inverse of \( A(w) \) and compute \( GI(A(w)) \times B(w) \) to find the transformation between \( A(w) \) and \( B(w) \). \( GI(A(w)) \times B(w) \) is denoted as \( T(w) \).

\[
N(w) - meanN(w) = A(w)
\]

\[
N(w + 1) - meanN(w + 1) = B(w)
\]

\[
A(w) \times T(w) = B(w)
\]
So the affine transformation is

\[
\begin{pmatrix}
x_{n+1} \\
y_{n+1} \\
z_{n+1} \\
u_{n+1}
\end{pmatrix} - \text{meanN}(w+1)^T = \begin{pmatrix}
x_n \\
y_n \\
z_n \\
u_n
\end{pmatrix} T(w) + \text{meanN}(w)^T T(w)
\]

i.e.

\[
\begin{pmatrix}
x_{n+1} \\
y_{n+1} \\
z_{n+1} \\
u_{n+1}
\end{pmatrix} = \begin{pmatrix}
x_n \\
y_n \\
z_n \\
u_n
\end{pmatrix} T(w) + \text{affine}(w)
\]

where

\[
(x_{n+1}, y_{n+1}, z_{n+1}, u_{n+1}) \text{ is a point in } N(w+1) \text{ and } (x_n, y_n, z_n, u_n) \text{ is a point in } N(w),
\]

\[
\text{affine}(w) = \text{meanN}(w+1)^T - \text{meanN}(w)^T T(w)
\]

By using this method, for each \( w \), we find a \( T(w) \) and an \( \text{affine}(w) \) by which we can predict \( N(w+1) \) from \( N(w) \).

**6.5 Calculation of Errors and the State Space Reconstruction:**

We calculate the error \( E(w) \) to assess the accuracy of the prediction. \( E(w) \) is the difference between the predicted \( N(w+1) \) and the original \( N(w+1) \). \( E(w) \) has 4 columns as our filtered ECG signal is 4-dimensional. We see that for \( w = 20 \), the error is quite small. Now we start with a specific initial condition and apply the local transformations to predict the data for a longer time. We see that the prediction of ECG is quite good and we can even reconstruct the state space picture. In the graph of figure 6.9, we plot the predicted ECG signal and the original ECG signal with respect to time. In figure 6.10 we show the reconstruction of the state space picture.

**6.6 Conclusion:**

In this chapter, we propose a method of predicting ECG data locally. We divide the whole data set into disjoint classes or fibers. We denote each fiber as \( N(w) \). For each \( w \), we find an affine transformation by which we can predict \( N(w+1) \) from \( N(w) \). If we start from an initial condition and apply these local transformations we can reconstruct the state space picture. The main significance of this local prediction is that it does a lot of data compression.
6.6. Conclusion:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00017805553</td>
<td>-0.0004379083</td>
<td>-0.00067781724</td>
<td>-0.0002713717</td>
</tr>
<tr>
<td>1</td>
<td>0.00018469215</td>
<td>-0.00036758270</td>
<td>-0.00056579349</td>
<td>-0.0002275575</td>
</tr>
<tr>
<td>2</td>
<td>-0.00008795058</td>
<td>0.00061270696</td>
<td>0.00127849371</td>
<td>0.00038692572</td>
</tr>
<tr>
<td>3</td>
<td>0.00028228362</td>
<td>-0.00070400836</td>
<td>0.00164818433</td>
<td>0.00044303232</td>
</tr>
<tr>
<td>4</td>
<td>0.00075267159</td>
<td>0.00056434365</td>
<td>-0.000231301408</td>
<td>-0.00032363163</td>
</tr>
<tr>
<td>5</td>
<td>-0.00019320152</td>
<td>0.00044093414</td>
<td>-0.00103857476</td>
<td>-0.0002809164</td>
</tr>
<tr>
<td>6</td>
<td>0.00033729996</td>
<td>0.00005229307</td>
<td>0.00134961567</td>
<td>0.00033695954</td>
</tr>
<tr>
<td>7</td>
<td>0.00025723894</td>
<td>0.00002274649</td>
<td>0.00069657004</td>
<td>0.001483606</td>
</tr>
<tr>
<td>8</td>
<td>-0.00025442689</td>
<td>0.00046571274</td>
<td>0.00066472962</td>
<td>0.00028973861</td>
</tr>
<tr>
<td>9</td>
<td>-0.00022193414</td>
<td>-0.0004901727</td>
<td>-0.00117219188</td>
<td>-0.00031257499</td>
</tr>
<tr>
<td>10</td>
<td>-0.00013467377</td>
<td>0.00022195373</td>
<td>0.0002971642</td>
<td>0.0003654186</td>
</tr>
<tr>
<td>11</td>
<td>-0.00030805168</td>
<td>-0.00034547233</td>
<td>-0.00078061016</td>
<td>-0.00016074896</td>
</tr>
<tr>
<td>12</td>
<td>-0.0000535325</td>
<td>0.00005239425</td>
<td>0.00039774909</td>
<td>0.00003282733</td>
</tr>
<tr>
<td>13</td>
<td>0.00021713232</td>
<td>-0.00023490313</td>
<td>-0.00024412739</td>
<td>-0.00014298915</td>
</tr>
<tr>
<td>14</td>
<td>-0.00034456586</td>
<td>-0.00018175937</td>
<td>-0.00065249215</td>
<td>-0.0002142528</td>
</tr>
<tr>
<td>15</td>
<td>0.00008294247</td>
<td>0.00056728066</td>
<td>0.00195773439</td>
<td>0.0061295577</td>
</tr>
</tbody>
</table>

Figure 6.7: difference between the actual and the predicted values of $N(21)$

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.1836854721</td>
<td>-0.57764149179</td>
<td>0.20042041672</td>
<td>-0.37527195341</td>
</tr>
<tr>
<td>1</td>
<td>1.0258985019</td>
<td>-0.5027173182</td>
<td>0.16967145737</td>
<td>-0.8465312792</td>
</tr>
<tr>
<td>2</td>
<td>1.1303195988</td>
<td>-0.4584510432</td>
<td>0.1807042515</td>
<td>-0.5344476178</td>
</tr>
<tr>
<td>3</td>
<td>1.1437632697</td>
<td>-0.5627939516</td>
<td>0.16468735934</td>
<td>-0.9493958293</td>
</tr>
<tr>
<td>4</td>
<td>1.1678436096</td>
<td>-0.5827963957</td>
<td>0.17066491013</td>
<td>-0.7571259104</td>
</tr>
<tr>
<td>5</td>
<td>1.1603175119</td>
<td>-0.5686255796</td>
<td>0.16555296715</td>
<td>-0.55573736601</td>
</tr>
<tr>
<td>6</td>
<td>1.1606995012</td>
<td>-0.5652662016</td>
<td>0.18738869167</td>
<td>-0.3516018782</td>
</tr>
<tr>
<td>7</td>
<td>1.16176936124</td>
<td>-0.5727882857</td>
<td>0.16487623296</td>
<td>-0.5693292541</td>
</tr>
<tr>
<td>8</td>
<td>1.0927326865</td>
<td>-0.5364361827</td>
<td>0.17815670131</td>
<td>-0.89930211789</td>
</tr>
<tr>
<td>9</td>
<td>0.91917517621</td>
<td>-0.5295044343</td>
<td>0.19189971097</td>
<td>-0.5399121626</td>
</tr>
<tr>
<td>10</td>
<td>1.1243408768</td>
<td>-0.5476513125</td>
<td>0.19156174934</td>
<td>-0.5241232156</td>
</tr>
<tr>
<td>11</td>
<td>1.09648523111</td>
<td>-0.5541655378</td>
<td>0.14924996949</td>
<td>-0.5029242156</td>
</tr>
<tr>
<td>12</td>
<td>1.12902135767</td>
<td>-0.5988885859</td>
<td>0.20590231487</td>
<td>-0.58149896473</td>
</tr>
<tr>
<td>13</td>
<td>1.11582712823</td>
<td>-0.55136262347</td>
<td>0.17481725204</td>
<td>-0.9250793988</td>
</tr>
<tr>
<td>14</td>
<td>1.08447202724</td>
<td>-0.5312389554</td>
<td>0.17883986367</td>
<td>-0.8816295129</td>
</tr>
<tr>
<td>15</td>
<td>1.11293325204</td>
<td>-0.54134645057</td>
<td>0.1918865843</td>
<td>-0.9115847168</td>
</tr>
</tbody>
</table>

Figure 6.8: values of $N(21)$
Chapter 6. A method for the Local Prediction for the ECG Data

Figure 6.9: plotting the actual and the predicted signals simultaneously

Figure 6.10: Reconstruction of the state space picture