Analysis of Real Time ECG Signal using Principal Component Analysis

Nishant Saxena  
Department of EE&I Engg.,  
Moradabad Institute of Technology,  
Moradabad U.P., India  
email:nishant_el@rediffmail.com

R.S. Anand  
Electrical Engineering Department  
IIT, Roorkee  
Uttarakhand, India  
email:anandfee@iitroorkee.ernet.in

Abstract—Principal Component Analysis (PCA) is one of the most valuable results oriented techniques of applied linear algebra. The minimum effort of PCA provides a roadmap for reducing a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it. Bioelectrical signals express the electrical functionality of different organs in the human body. The Electrocardiogram, also called ECG signal, is one important signal among all bioelectrical signals. The ECG reflects the performance and the properties of the human heart and conveys very important hidden information in its structure. Principal component analysis (PCA) is a statistical technique whose purpose is to condense the information of a large set of correlated variables into a few variables (“principal components”), while not throwing overboard the variability present in the data set. The principal components are derived as a linear combination of the variables of the data set, with weights chosen so that the principal components become mutually uncorrelated. Each component contains new information about the data set, and is ordered so that the first few components account for most of the variability. Noise reduction is closely related to data compression as reconstruction of the original signal usually involves a set of eigenvectors whose noise level is low, and thus the reconstructed signal becomes low noise; such reduction is, however, mostly effective for noise with muscular origin. The purpose of the paper is to provide an overview of PCA in ECG signal compression.

Keywords—ECG, Principal component analysis.

I. INTRODUCTION

Principal Component Analysis (PCA) is a statistical technique whose purpose is to condense the information of a large set of correlated variables into a few variables (“principal components”), while not throwing overboard the variability present in the data set [1]. The principal components are derived as a linear combination of the variables of the data set, with weights chosen so that the principal components become mutually uncorrelated. Each component contains new information about the data set, and is ordered so that the first few components account for most of the variability. In signal processing applications, PCA is performed on a set of time samples rather than on a data set of variables. When the signal is recurrent in nature, like the ECG signal, the analysis is often based on samples extracted from the same segment location of different periods of the signal. Signal processing is today found in virtually any system for ECG analysis, and has clearly demonstrated its importance for achieving improved diagnosis of a wide variety of cardiac pathologies. Signal processing is employed to deal with diverse issues in ECG analysis such as data compression, beat detection and classification, noise reduction, signal separation, and feature extraction. Principal component analysis has become an important tool for successfully addressing many of these issues, and was first considered for the purpose of efficient storage retrieval of ECGs. Over the years, this issue has remained central as a research topic, although the driving force has gradually changed from having been tiny hard disks to become slow transmission links. Noise reduction may be closely related to data compression as reconstruction of the original signal usually involves a set of eigenvectors whose noise level is low, and thus the reconstructed signal becomes low noise; such reduction is, however, mostly effective for noise with muscular origin. Classification of waveform morphologies in arrhythmia monitoring is another early application of PCA, in which a subset of the principal components serves as features which are used to distinguish between normal sinus beats and abnormal waveforms such as premature ventricular beats. A recent application of PCA in ECG signal processing is robust feature extraction of various waveform properties for the purpose of tracking temporal changes due to myocardial ischemia. Historically, such tracking has been based on local measurements derived from the ST-T segment; however, such measurements are unreliable when the analyzed signal is noisy. With correlation as the fundamental signal processing operation, it has become clear that the use of principal components over a more robust and global approach to the characterization of the ST-T segment. Signal separation during atrial fibrillation is another recent application of PCA, the specific challenge being to extract the atrial activity so that the characteristics of this common arrhythmia can be studied without interference from ventricular activity. Such separation is based on the fact that the two activities originate from different bioelectrical sources; separation may exploit temporal redundancy among successive heartbeats as well as spatial redundancy when multilead recordings are analyzed. Principal component analysis can be performed on raw data matrices or genetic similarity or distance matrices estimated from raw data. ECG analysis
systems are usually designed to process ECG signals measured under particular conditions, like resting ECG interpretation, stress test analysis, ambulatory ECG monitoring, intensive care monitoring.

II. THE ANATOMY OF THE HEART

The electrical functionality of different organs in the human body is expressed by bioelectrical signals. The Electrocardiogram, also called ECG signal, is one important signal among all bioelectrical signals. The ECG reflects the performance and the properties of the human heart and conveys very important hidden information in its structure. This information has to be extracted and analyzed before any useful and meaningful interpretations can be started. Extracting or decoding this information or feature from ECG signal has been found very helpful in explaining and identifying various pathological conditions. The feature extraction procedure can be accomplished straightforward by analyzing the ECG visually on paper or screen. However, the complexity and the duration of ECG signals are often quite considerable making the manual analysis a very time-consuming and limited solution. In addition, manual feature extraction is always prone to error. Therefore, ECG signal processing has become an indispensable and effective tool for extracting clinically significant information from ECG signals, for reducing the subjectivity of manual ECG analysis and for developing advanced aid to the physician in making well-founded decisions. Over the past few years automatic analysis of electrocardiograms (ECG) has gained more and more significance in the field of clinical ECG diagnosis.

The human heart is located in the chest between the lungs, behind the sternum and above the diaphragm. It weighs between 200 to 425 grams and is a little larger than the size of a fist [2, 3, 4].

The rhythmic beating of the heart is due to the triggering pulses that originate in an area of specialized tissue in the right atrium of the heart. This area is known as sino-atrial node.

The heart has four chambers: two atria (singular: atrium) and two ventricles. Figure 2.1 shows all the components of the heart. The SA node sends electrical impulses at a certain rate, but heart rate may still change depending on physical demands, stress, or hormonal factors [3]. The rhythmic sequence of contractions is coordinated by the sinoatrial (SA) and atrioventricular (AV) nodes located at the upper and lower walls, respectively, of the right atrium.

The contractions of the heart are controlled by electrical impulses. These electrical impulses fire at a rate which controls the beat of the heart. The cells that create these rhythmical impulses are called pacemaker cells, and they directly control the heart rate.

![Fig. 1. The Human Heart.](image1)

![Fig. 2. ECG Waveform](image2)

**P wave:** Identify the wave of depolarization that spreads from the SA node throughout the atria. It typically has 80-100 ms in duration.

**P - R interval:** It normally lasts from 120 to 200 ms. It represents the time between the onset of atrial depolarization and the onset of ventricular depolarization. If the P-R interval is >0.2 sec, there is an AV conduction block. It is also termed as a first-degree heart block if the impulse is still able to be conducted into the ventricles.

**QRS Complex:** The QRS complex represents the ventricular depolarization. It is the most prominent amplitude of the ECG. It can be used to diagnose bundle branch blocks or abnormal pacemaker site located in the ventricles. This can be detected when the QRS complex is prolonged above 100ms.

**ST Segment:** The ST segment is measured from the onset of the S wave to the onset of the T wave. The T wave represents the repolarization of the ventricles. The ST segment is the time at which the entire ventricle is depolarized. The ST segment is important in the diagnosis of ventricular ischemia or hypoxia because under those conditions, the ST segment can become either depressed or elevated.

**QT interval:** The Q-T interval represents the time for both ventricular depolarization and repolarization to occur. Therefore, it roughly estimates the duration of an average ventricular action potential. This interval can range from 200 to 400 ms. In practice, the Q-T interval is expressed as a “corrected Q-T (QTe)” by taking the Q-T interval and dividing it by the square root of the R-R is the time between two consecutive R waves.

III. PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first
coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the ‘most important’ aspects of the data, but this is not necessarily the case, depending on the application. PCA is also called the (discrete) Karhunen-Loève transform (or KLT, named after Kari Karhunen and Michel Loève) or the Hotelling transform (in honor of Harold Hotelling). PCA has the distinction of being the optimal linear transformation for keeping the subspace that has largest variance. This advantage, however, comes at the price of greater computational requirement if compared, for example, to the discrete cosine transform. Unlike other linear transforms, the PCA does not have a fixed set of basis vectors. Its basis vectors depend on the data set.[15].

A. The Covariance Matrix Method

The goal is to transform a given data set X of dimension M to an alternative data set Y of smaller dimension L. Equivalently, we are seeking to find the matrix Y, where Y is the Karhunen-Loève transform (KLT) of matrix X:

\[ Y = KLT\{X\} \]

1. Organizing the data set: Suppose a training set X with N samples and each sample Xi can be expressed by a row vector with the size of M as follows:

\[ X_i = [X_{i1}, X_{i2}, ..., X_{iM}] \]

The training set is placed into a single matrix X of dimensions N × M, so that N are the number of observations and M is the dimension of the observation vector.

\[ X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1M} \\ \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NM} \end{pmatrix} \]

2. Calculate the empirical mean row vector: The empirical mean along each dimension m = 1…M is calculated. Afterward, all computed mean values are placed into an empirical mean row vector u of dimension M.

\[ u(m) = \frac{1}{N} \sum_{n=1}^{N} X(n,m), m=1,2,...,M \]

3. Find the covariance matrix: As illustrated before, the M × M empirical covariance matrix C is calculated from the outer product of the zero-centered matrix B with itself:

\[ C = E[B \otimes B] = E[B.B^t] = \frac{1}{N-1} B.B^t \], \;

where E is the expected value operator, is the outer product operator, and \( \sim \) is the conjugate transpose operator.

4. Find the eigenvectors and eigenvalues of the covariance matrix: This step will typically require the use of a computer-based algorithm for computing the eigenvalue matrix D and the eigenvector matrix V of the covariance matrix C:

\[ C.V = V.D \], \;

5. The matrix D will take the form of an M × M diagonal matrix, where \( D[p, q] = \lambda_m \) for \( p = q = m \) is the m\textsuperscript{th} eigenvalue of the covariance matrix C, and \( D[p, q] = 0 \) for \( p \neq q \).

6. Rearrange the eigenvectors and eigenvalues: The columns of the eigenvector matrix V and eigenvalue matrix D are sorted out in order of decreasing eigenvalues thereby maintaining the correct pairings between the columns in each matrix.

7. Compute the cumulative energy content for each eigenvector: The eigenvalues represent the distribution of the source data’s energy among each of the eigenvectors, where the eigenvectors form a basis for the data. The cumulative energy content \( g[m] \) for the m\textsuperscript{th} eigenvector is the sum of the energy content across all of the eigenvectors from 1 through m:

\[ g[m] = \sum_{q=1}^{m} D[p, q] \]

for \( p = q \) and \( m = 1,...,M \).

8. Select a subset of the eigenvectors as basis vectors: Save the first L columns of V as the M × L matrix W:

\[ W[p, q] = V[p, q] \]

For \( p = 1,...,Mq = 1,...,L \), the matrix Z is calculated by multiplying the eigenvector matrix with the zero-mean data matrix from the left as follows:

\[ Z = B.V = KLT\{x\} = \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1M} \\ \vdots & \ddots & \vdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{NM} \end{pmatrix} \]
The rows of Z correspond to the observations, whereas the columns refer to the components or dimensions. In fact, the projected PCA-scores or vectors represent the Karhunen-Loève transform (KLT) of the data vectors in the columns of matrix X.[8]

IV. IMPLEMENTATION

The method that we are using is principal component analysis. All we have done is that we have taken normal ECG signal and perform the principal component analysis. To perform the experiment we need some hardware as well as some software. Those are described here.

A. Software/Hardware Requirement

The ECG waveforms have been measured with the help of Real time ECG Software developed by Scientech Instruments. For the implementation of principal component analysis of ECG signal we have used MATLAB software which is used as the simulation package for this paper. All simulations were run on a 1.8 GHz Pentium 4 processor with 512 MB of RAM and ECG measurement setup make Scientech technologies (ST-8251). The kit has been interfaced with the PC for the proper visualization the ECG waveform. At different conditions ECG measurement was done using the ECG measurement setup.

B. Design Procedure

The complete experimental setup contains three accompanying sections:

- Measurement of the ECG waveform
- Compression of the ECG waveform
- Reconstruction of ECG Waveform

C. Analysis and Results

The waveform which has to be analysed has been observed and shown in Fig. 3.

Transform-based compression requires that the ECG first be partitioned into a series of successive blocks, where each block is subjected to data compression. The signal may be partitioned so that each block contains one beat. Each block is positioned around the QRS complex, starting at a fixed distance before the QRS, including the P wave and extending beyond the end of the T wave to the beginning of the next beat. Since the heart rate varies, the distance by which the block extends after the QRS complex is adapted to the prevailing heart rate. Hence, the resulting blocks vary in length, introducing a potential problem in transform-based compression where a fixed block length is assumed. This problem may be solved by padding too short blocks with a suitable sample value, whereas too long blocks can be truncated to the desired length. The use of variable block lengths has been studied in detail in [13,14]; the results show that variable block lengths produce better compression performance than fixed blocks. It should be noted that partitioning of the ECG is bound to fail when certain chaotic arrhythmias are encountered such as ventricular fibrillation during which no QRS complexes are present.

The corresponding Eigen values of this waveform analysis with the MATLAB software has been shown here:

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>7.7696</td>
</tr>
<tr>
<td>2.</td>
<td>2.0530</td>
</tr>
<tr>
<td>3.</td>
<td>1.0886</td>
</tr>
<tr>
<td>4.</td>
<td>0.2225</td>
</tr>
<tr>
<td>5.</td>
<td>0.1388</td>
</tr>
<tr>
<td>6.</td>
<td>0.0407</td>
</tr>
<tr>
<td>7.</td>
<td>0.0213</td>
</tr>
<tr>
<td>8.</td>
<td>0.0177</td>
</tr>
<tr>
<td>9.</td>
<td>0.0099</td>
</tr>
<tr>
<td>10.</td>
<td>0.0098</td>
</tr>
<tr>
<td>11.</td>
<td>0.0082</td>
</tr>
<tr>
<td>12.</td>
<td>0.0072</td>
</tr>
<tr>
<td>13.</td>
<td>0.0057</td>
</tr>
<tr>
<td>14.</td>
<td>0.0049</td>
</tr>
<tr>
<td>15.</td>
<td>0.0037</td>
</tr>
<tr>
<td>16.</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

Fig. 3. Normal ECG measurement

Fig. 4. one cycle of ECG waveform

The one cycle of the ECG waveform has been extracted from this waveform and shown in Fig. 4.
Only three eigenvectors that corresponds to the three dominant eigenvalue are computed and are shown in Figs. 5, 6 and 7.

The reconstructed waveform corresponding to first cycle and given by the equation:

\[ X_{m'} = y_1\phi_1 + y_2\phi_2 + y_3\phi_3 \]

is shown in Figure 8, only three ECG Eigenvectors \( \Phi_1, \Phi_2, \Phi_3 \) and three features \( y_1, y_2, y_3 \) we are able to reconstruct the original cycle of the waveform with fairly good accuracy. Fig. 8 is showing the original and reconstructed waveform in the same window.

With the help of which the parameters of ECG wave are calculated and shown in Table II.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Parameters</th>
<th>Original wave</th>
<th>Reconstructed wave</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>P-Wave</td>
<td>0.25mv</td>
<td>0.23mv</td>
<td>8%</td>
</tr>
<tr>
<td>2.</td>
<td>QRS Complex</td>
<td>1.60mv</td>
<td>1.55mv</td>
<td>3.12%</td>
</tr>
<tr>
<td>3.</td>
<td>T-Wave</td>
<td>0.40mv</td>
<td>0.35mv</td>
<td>12.5%</td>
</tr>
<tr>
<td>4.</td>
<td>P-R Interval</td>
<td>0.20s</td>
<td>0.19s</td>
<td>5%</td>
</tr>
<tr>
<td>5.</td>
<td>Q-T Interval</td>
<td>0.40s</td>
<td>0.35s</td>
<td>5%</td>
</tr>
</tbody>
</table>

With the help of graph drawn below we can analyze the percentage error. On the X-axis the ECG parameters have been shown and on the Y-axis % Error has been shown.

V. CONCLUSION

The principal component analysis of ECG signal is very helpful in the compression of the signal. Although memory is not a very big problem now-a-days but in the transmission of signal it is very useful as the distortion in the signal will be very less in the case of compressed signal. Several PCA-based strategies are available which exploit the fact that the ECG signal exhibits intrabeat and interlead redundancy. Although the underlying principle is the same in all ECG applications, the results are obtained and interpreted in quite diverse ways. In some applications, the goal is to find a more compact representation of the signal, while in others it is to search for specific patterns or to extract a certain physiologic activity. In other applications, PCA may serve as a powerful, intermediate step when addressing problems related to noise reduction and beat classification. To date, PCA has been used to solve signal processing issues, most
notably ECG data compression, as well as clinically oriented issues related to the characterization and diagnosis of myocardial ischemia, ventricular repolarization. In the case of simulation experiment we have come to conclusion that error in the T-wave reconstruction is more but it is not going to affect the diagnosis as it is within the tolerance limit. The reconstruction can be performed with the help on only three eigenvectors.

REFERENCES

[14] M. Mudrov’a, A. Proch’azka “Principal component analysis In image processing”.