Clinical Entropy based PCA for Multi-lead Electrocardiogram Signals

L. N. Sharma, *Member, IACSIT*, S. Dandapat, *Member, IEEE*, and A. Mahanta, *Member, IEEE*

Dept. of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
Guwahati, India
lns@iitg.ernet.in

Abstract— In this work, an information theoretic approach is proposed for principal component analysis (PCA) of multi-lead electrocardiogram signals. Clinical information is evaluated from the inverse of the diagonal eigenvalue matrix. It is termed as Clinical Entropy (Centropy). Clinical entropy (Centropy) based PCA method shows improved performance compared to the conventional PCA. The proposed method exhibits superior signal quality with higher cross correlation (CC), lower percentage root mean square difference (PRD) and lower root mean square error (RMSE) values.

Keywords—ECG, PRD, PCA, Entropy

I. Introduction

Conventional multivariate signal analysis tool such as principal component analysis (PCA) [1] can be employed to process multichannel electrocardiogram signals (MECG). PCA transforms a set of correlated variables into a new set of uncorrelated variables which are called principal components (PC). By selecting suitable number of PCs, the dimensions can be reduced with minimal loss. PCA has been effectively used for noise elimination, signal separation, data reduction, feature extraction [2],[3] and separation of respiratory and the non-respiratory segments in an ECG signal [4]. In an ECG signal, the clinical components such as P-wave, QRS-complex and T-wave appear at different frequency bands [5], [6]. These clinical information are captured by the eigenvalues. The quantification and subsequent preservation of clinical information in multichannel ECG signals are challenging tasks. This paper proposes a novel clinical entropy (Centropy) based PCA method for multichannel ECG signal processing.

II. Clinical Entropy based PCA

Conventional PCA can be applied to multi-lead or multichannel ECG signals by forming appropriate multivariate matrix. The multivariate N×M , signal matrix, **S** , is constructed with M number of channels as column and N number of samples in each channel. Each element, s_{ij} , represents the ith sample of the jth channel. The square symmetric covariance matrix, $\mathbf{C_s}$, is evaluated as

$\mathbf{C_s} = \dfrac{1}{(M-1)}([\mathbf{S}][\mathbf{S}]^T)$. It exploits all the possible signal correlations between the channels. The main objective is to minimize redundancy measured by covariance and to maximize the signal measured by variance. Orthogonal eigenvector matrix, $\mathbf{V_s}$, diagonalizes the covariance matrix, $\mathbf{C_s}$, as

$$\mathbf{V_s}\mathbf{C_s}\mathbf{V_s^{-1}} = \Lambda_s \qquad (1)$$

Λ_s is the diagonal matrix where the diagonal elements, λ_i, $(i = 1, 2, L, M)$, are the eigenvalues. The eigenvectors with higher eigenvalues are the principal components. These orthogonal eigenvectors represent the signals in the direction of maximum variances. In conventional PCA [1], the threshold, T_s , for the selection of PC is defined as

$$T_s = \dfrac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{M}\lambda_i} \times 100 \qquad (2)$$

where λ_i is the ith eigenvalue, **M** is the total number of eigenvalues, **p**, is the number of eigenvalues selected. Generally, to select the subset of PC, cumulative percentage of total variation of variances [1] is taken. Thus, with lower value of T_s more dimension reduction is possible. Careful implementation of PCA is required for retention of clinical information in MECG signals. Higher eigenvalues are resulted from significant energy contributed by P-wave, QRS-complex and T-wave. The noise and other signals components are captured by lower eigenvalues. Proper selection of eigenvalues can ensure retention of clinical information. In this work, clinical entropy (Centropy) is proposed for selection of appropriate number of eigenvalues. For evaluation of Centropy, first, the diagonal eigenvalue matrix (Λ_s) is inverted. The probability of each diagonal element (λ_i') of this inverse matrix is defined as

$$P_i = \frac{\lambda_i'}{\sum_{i=1}^{M} \lambda_i'} \qquad (3)$$

where λ_i' is the reciprocal of ith eigenvalue. Higher eigenvalues will result in lower probability values. The self-information contributed by the ith eigenvalue can be given as

$$I_i = -\log(P_i) \qquad (4)$$

Hence, the entropy information [7] contributed by i^{th} eigenvalue is given as $H_i = -P_i \log(P_i)$ and it is termed as clinical entropy (Centropy). It is expected that the eigenvalue

of all eigenvalues. Thus, the self-information and the entropy of ith eigenvalue for conventional PCA can be estimated using this probability value.

III. Results and Discussion

Multichannel signals from CSE multilead measurement library [8] datasets M01-040 and M01-004 are used to form multivariate data matrix, **S**. In this matrix there are 12 columns which correspond to 12 channels of data. Each column consists of 4096 samples. First, mean removal, amplitude normalization and base-line wander removal for each channel data are carried-out. The matrix, **S**, is subjected to covariance analysis. For the proposed Centropy based PCA, self-information and entropy information of eigenvalues are evaluated using the inverse of the diagonal eigenvalue matrix.

**Fig.1 (a) ECG signal of
lead-I, recontructed signals (b) without PC1 with the highest
eigenvalue, (c) without PC1 and PC2, (d) without PC1, PC2 and PC5.**
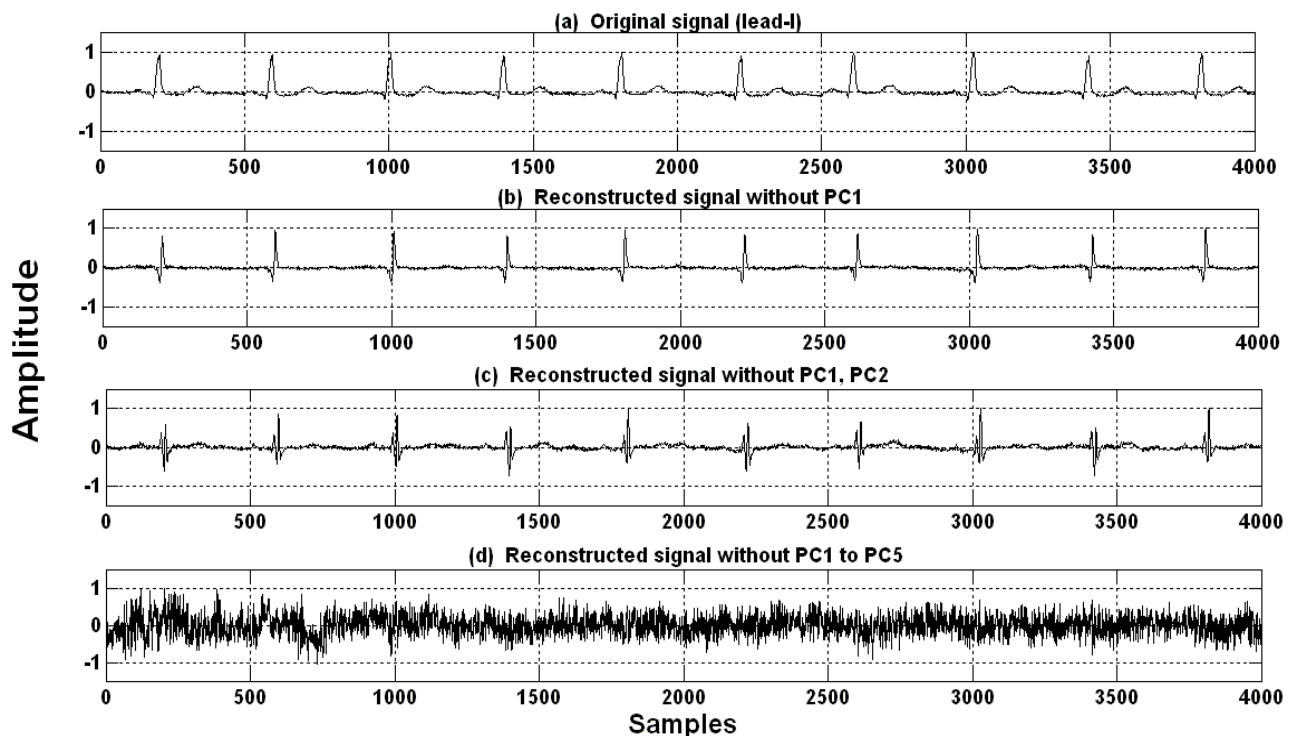
with lower probability will have higher self-information and higher entropy. In this work, the selection of number of PCs is based on total cumulative entropy. The Centropy based threshold, (T_H), is defined as

$$T_H = \frac{\sum_{i=1}^{p'} H_i}{\sum_{i=1}^{M} H_i} \times 100 \qquad (5)$$

For conventional PCA, probability of ith eigenvalue can be defined as the ratio between the ith eigenvalue and the sum

Fig.1 shows lead-I ECG signal and corresponding reconstructed signal after elimination of selected principal components. There are 12 PCs. Based upon the magnitudes, the eigenvalues are arranged in descending order and corresponding PCs are named as PC1 to PC12. Fig.1 (b) shows the reconstructed signal without PC1. It is observed that the low frequency content of the signal with clinical components, P-wave, T-wave and some part of QRS-complex, are distorted. Fig.1. (c) shows the reconstructed signal without PC1 and PC2. It is observed that, in addition to P-wave and T-wave, the QRS-complex is significantly affected. After removal of five PCs (PC1 to PC5), the ECG signal is

completely distorted as shown in Fig.1. (d). This result shows

**Fig.3 RMSE and cross correlation
plots; (a) RMSE with reduction of PCs and (b) CC with reduction of
PCs for signals of lead-II, lead-aVF and lead-V2 respectively. The
dataset-M01-004 is taken from CSE multilead measurement library.**
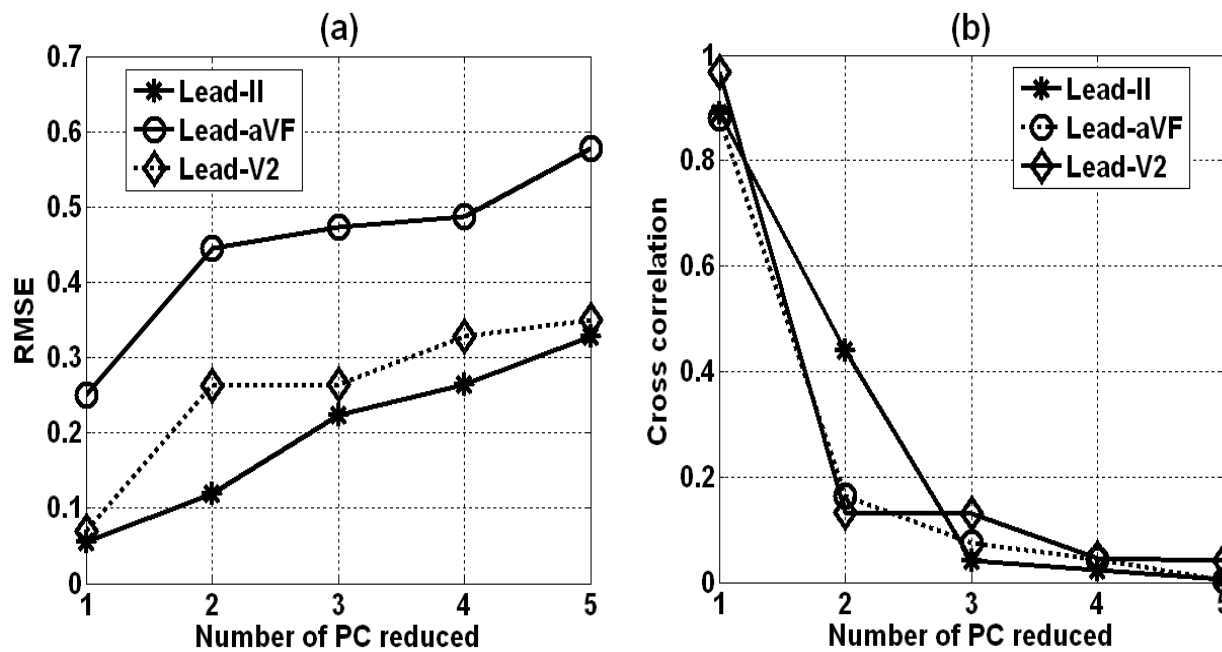
the clinical significance of PCs with higher eigenvalues.

**Fig. 2 Scree and
Centropy plots for Conventional and Centropy PCA. CSE multilead
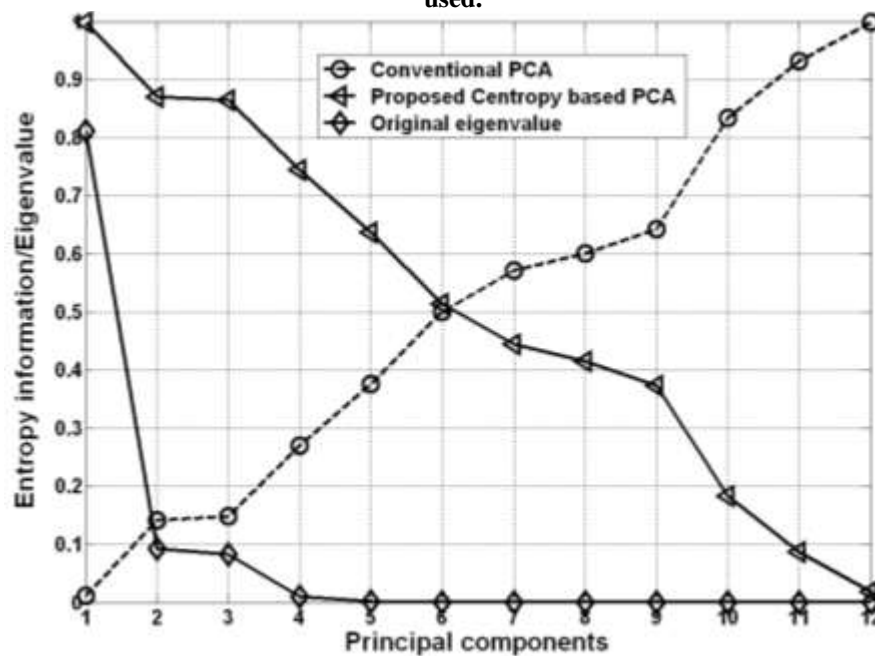measurement library, dataset-M01-040 is
used.**

Fig.2, shows the RMSE and CC values for reconstructed signals after removal of PCs. The results for Lead-II, lead-aVF and lead-V2 signals are shown in the figure. From Fig.2.(a), it

is observed that the RMSE values increase with removal of more number of PCs. Similarly, it is observed that the CC values decrease with removal of more number of PCs as shown in Fig.2. (b). The above results (Fig.1 and Fig.2) prove that the clinical significance of PCs is higher for those with higher eigenvalues.

eigenvalues are associated with the significant clinical information such as P-wave, QRS-complex and T-wave in an ECG signal. From the results shown in Fig.3, it is observed that the Centropy values are higher for higher eigenvalues. Hence, higher values of Centropy represent significant clinical information in an ECG signal. This justifies the assumption that the Centropy can quantify the clinical information. On

Fig. 4 Original signals and recontructed signals using proposed Centropy method and conventional PCA. In panels (a) and (d) original signals of lead-V4 and V5; (b) and (e) reconstructed signals using proposed Centropy PCA and, (c) and (f) reconstructed signals using conventional PCA. The dataset-M01-040 is taken from CSE multilead measurement library.
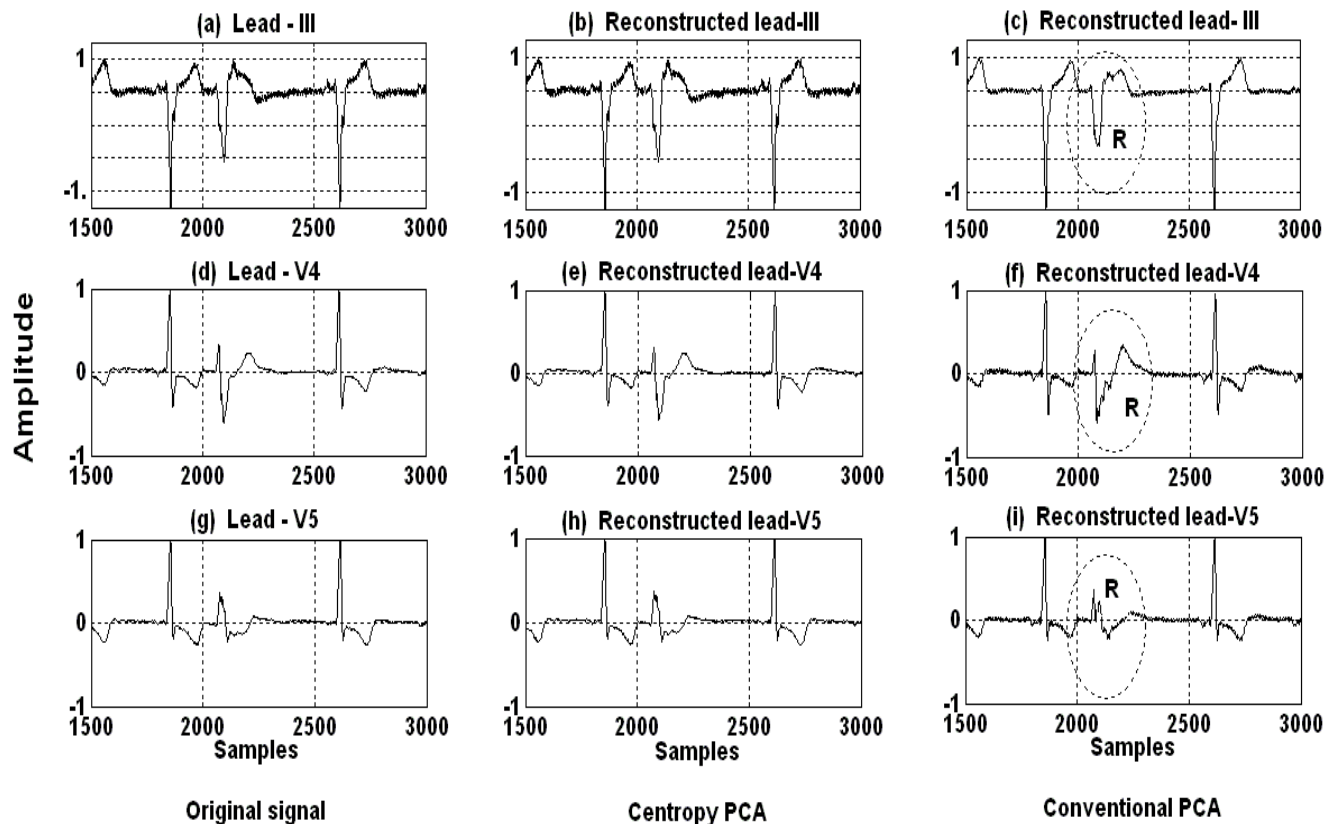

Fig.3 shows, the scree (eigenvalue), the conventional PCA based entropy and Centropy values. The x-axis represents the principal component number (eigenvalue number). Higher

contrary, the conventional PCA based entropy values are lower for higher eigenvalues as shown in Fig.3.

Table 1 Distortion measures: PRD (in %), RMSE (in %) and CC for proposed method and conventional PCA

Parameters	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6
PRD-Proposed	0.56	0.35	0.06	0.47	0.49	0.75	1.62	1.34	1.43	3.67	2.21	3.72
PRD-Conventional	11.79	4.33	14.38	6.12	3.13	8.93	11.43	15.29	18.79	21.62	19.86	11.70
RMSE-Proposed	0.07	0.09	0.04	0.10	0.09	0.48	0.70	0.29	0.32	0.61	0.38	0.67
RMSE-Conventional	1.58	1.08	8.96	1.31	0.59	5.70	4.96	3.33	4.16	3.62	3.39	2.12
CC-Proposed	1.00	1.00	1.00	1.00	1.00	1.00	0.999	0.999	0.999	0.999	0.999	0.999
CC-Conventional	0.996	0.999	0.990	0.998	0.999	0.996	0.995	0.990	0.987	0.977	0.980	0.993

Fig.4 shows segments of ECG signals from three channels and corresponding results for Centropy based PCA and conventional PCA. The original signals from lead-III, V4 and V5 are shown in Fig.4.(a), (d) and (g) respectively.

The reconstructed signals for Centropy based PCA and from conventional PCA are shown in Fig.4. (b), (e), (h) and Fig.4.(c), (f) and (i) respectively. For retaining 99% of the total variance, the number of PC required is 6 in case of Centropy based PCA whereas for conventional PCA, the number of PCs is 3. It is observed that the segment of reconstructed signal in Fig.4. (c) which is marked as 'R' is distorted compared to the same segment in the original signal shown in Fig.4.(a). The reconstructed signal in Fig.4. (b) shows no perceivable distortion compared to the original signal. The clinical information like P-wave, QRS-complex and T-wave are well preserved with sufficient clinical fidelity. Similar results are observed with the other two signals. These results show that the Centropy based PCA method performs better compared to the conventional PCA based method from the point of view of preservation of clinical information.

The performances of Centropy based PCA and conventional PCA are quantitatively evaluated using distortion metrics, PRD [5], [6] RMSE and CC for all 12 channel data. The results are shown in Table-1. It is observed that the PRD and the RMSE values are lower for all the channels in case of the proposed Centropy based PCA method compared to the conventional PCA based method. The CC values are higher in case of Centropy based PCA method for all the channels. Lower PRD and RMSE values and higher CC values prove that the reconstructed signal quality is better in case of the Centropy based PCA method. These results prove that the proposed Centropy based PCA method is superior to the conventional PCA based method.

IV. Conclusions

In this work, a novel clinical entropy based PCA is proposed for multichannel electrocardiogram signals. It is shown that the Centropy quantifies the clinical information. For preservation of signal fidelity in terms of clinical information, threshold based on Centropy can be successfully used for selection of principal components. The evaluation of measures such as PRD, RMSE and CC show that Centropy based PCA perform superior to conventional PCA.

References

[1] G. I. T. Joliffe, "Principal Component Analysis," second edition, *Springer, New York, NY, USA*, 2002.

[2] Francisco Castells, Pablo Laguna, Leif Sornmo, Andreas Bollmann, and JoseMillet Roig, "Principal Component Analysis in ECG Signal Processing," *EURASIP Journal on Advances in Signal Processing, Hindawi Publishing Corporation*, Volume 2007, Article ID 74580, 21 pages.

[3] M. P. S. Chawla, "A comparative analysis of principal component and independent component techniques for electrocardiograms," Neural Comput and Applic, Springer-Verlag London Limited, 18:pp.539 - 556, 2009.

[4] Philip Langley, Emma J. Bowers, Alan Murray, "Principal Component Analysis as a Tool for Analyzing Beat-to-Beat Changes in ECG Features: Application to ECG-Derived Respiration," *IEEE Transactions on biomedical engineering*, vol. 57, no. 4, pp.821 - 829, 2010.

[5] L.N. Sharma, S. Dandapat, A. Mahanta, "ECG signal denoising using higher order statistics in Wavelet subbands," *Signal Process. Control, Elsevier*, vol 5, pp.214 - 222, 2010.

[6] L.N. Sharma, S. Dandapat, A. Mahanta, "Kurtosis-based noise estimation and multiscale energy to denoise ECG signal," *Signal, Image and Video Processing, Springer*, DOI: 10.1007/s11760-011-0227-7, Springer-Verlag London Limited 2011.

[7] M.S.Manikandan, and S.Dandapat, "Multiscale Entropy-Based Weighted Distortion Measure for ECG Coding," *IEEE Signal Processing Letters*, vol.15, 2008.

[8] Jos L. Willems (CSE Project Leader), "CSE Multilead Atlas, Measurement Results - Data Set 3," *Common Standards for Quantitative Electrocardiography, Commission of the European Communities, Medical and Public Health Research*, Ref. Nr. CSE 88- 04-15, Leuven, 1988