31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Fast PCA via UTV Decomposition and Application on EEG Analysis

Yodchanan Wongsawat

***Abstract*— In the mean square error sense, principal component analysis (PCA) or Karhunen-Loeve transform (KLT) can optimally summarize the high dimensional data into only a few meaningful ones. However, for the biomedical signal analysis, e.g. electroencephalogram (EEG), the data need to be updated or downdated very often. This fact makes the PCA impractical to be employed, especially in real-time signal analysis. In this paper, we propose the fast computational method for approximating the PCA such that the new transform, called fast PCA (fastPCA), can easily be updated and downdated. The fastPCA is calculated via the UTV decomposition which is the method normally used to approximate the rank-revealing property of the singular value decomposition (SVD). The merit of the fastPCA is also illustrated via the application on EEG analysis.**

I. Introduction

Many applications in biomedical signal processing normally employ the principal component analysis (PCA) or Kahunen-Loeve transform (KLT) since it can optimally summarize the high dimensional data into only a few important coefficients. The applications include EEG or ECG compression, feature extraction via eigen analysis, data reduction, etc. However, for the online analysis of biomedical signals, e.g. electroencephalogram (EEG), electrocardiogram (ECG), electromyogram(EMG), the data need to be updated or downdated very often. This fact makes the PCA impractical to use in the real-time analysis.

There are many methods used for calculating the PCA [1]. One of the widely used methods is calculating the PCA via the singular value decomposition (SVD). The SVD is the classical matrix decomposition normally used for obtaining rank of a matrix [2]. However, in real time applications, new data have to be updated/downdated. Since the SVD requires high computational complexity on updating/downdating its eigenvectors and eigenvalues, other decompositions which can eliminate this disadvantages need to be investigated. In [3] and [4], a more efficient rank-revealing decomposition when data need to be updated/downdated called UTV decomposition is introduced. The UTV decomposition decomposes a matrix into the product of orthogonal matrices and the upper or lower triangular matrix. If the upper triangular matrix is used, the decomposition is called URV. In addition, the UTV decomposition is called ULV when the lower triangular matrix is used. Since the UTV decomposition needs only the upper or lower triangular matrix instead of the diagonal matrix in the SVD, updating/downdating algorithms consume only $O(n^2)$ operations instead of $O(n^3)$ operations. Even though the UTV decomposition has been introduced for many years, nobody fully exploits the properties of this decomposition to calculate the PCA.

In this paper, we estimate the PCA matrix by employing the UTV decomposition. Since the UTV decomposition can efficiently reveal the rank of the matrix as well as the SVD, we can similarly derive the PCA matrix based on this decomposition rendering our proposed transform called fast PCA (fastPCA). The merit of the fastPCA is evaluated via the EEG analysis, e.g. summarizing the principal frequencies of the multichannel EEG [5].

This work is supported in part by the young researcher funding of Mahidol University

Y. Wongsawat is with Department of Biomedical Engineering, Mahidol University, 25/25 Phuttamonthon Sai4, Salaya, Nakornpathom 73170, Thailand. egyws@mahidol.ac.th

II. Calculation of the PCA

Principal component analysis (PCA) or Karhunen-Loeve transform (KLT) is the transform that maps a real random vector $\mathbf{x} = (x_0, x_1, \ldots, x_{m-1})^T$ to a random vector $\mathbf{y} = (y_0, y_1, \ldots, y_{m-1})^T$ such that $\mathbf{y}$ is completely decorrelated. In particular, if $\mathbf{\Phi}^T$ is the matrix representation of the PCA, then $\mathbf{y} = \mathbf{\Phi}^T \mathbf{x}$ and

$$\begin{aligned} E\left\{\mathbf{y}\mathbf{y}^T\right\} &= E\left\{\mathbf{\Phi}^T\mathbf{x}\mathbf{x}^T\mathbf{\Phi}\right\} \\ &= \mathbf{\Phi}^T E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{\Phi} = \mathbf{\Phi}^T\mathbf{R}\mathbf{\Phi} = \mathbf{\Lambda}, \end{aligned}$$

where the columns of $\mathbf{\Phi}$ are the normalized eigenvectors of $\mathbf{R} = E\left\{\mathbf{x}\mathbf{x}^T\right\} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T$, $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues with respect to $\mathbf{\Phi}$ and T denotes the transpose operation. It can also be proved that the PCA yields optimal energy compaction, i.e. minimizing the entropy [6]. Therefore, the transformed vector $\mathbf{y}$ is widely used in data compression.

III. UTV decomposition

Suppose that $\mathbf{A}_{m\times n}$ has rank close to k. That is the singular values of $\mathbf{A}_{m\times n}$ satisfy

$$\sigma_1 \geq \ldots \geq \sigma_k \gg \sigma_{k+1} \ldots \geq \sigma_n. \quad (1)$$

There exist orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{V}^T, \quad (2)$$

where

1) $\mathbf{T} = \begin{bmatrix} \mathbf{S} & \mathbf{C} \\ \mathbf{Z} & \mathbf{E} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$,
2) $\mathbf{S}$ is the upper/lower triangular matrix of the size $k \times k$,
3) $\mathbf{E}$ is the upper/lower triangular matrix of the size $(n-k) \times (n-k)$,
4) if $\mathbf{S}$ and $\mathbf{E}$ are upper triangular matrices, $\mathbf{Z}$ is zeros matrix, and $\sqrt{\|C\|^2 + \|E\|^2} \cong \sqrt{\sigma_{k+1}^2 + \ldots + \sigma_n^2}$, UTV is called URV,
5) if $\mathbf{S}$ and $\mathbf{E}$ are lower triangular matrices, $\mathbf{C}$ is zeros matrix, and $\sqrt{\|Z\|^2 + \|E\|^2} \cong \sqrt{\sigma_{k+1}^2 + \ldots + \sigma_n^2}$, UTV is called ULV.

By peeling of the large singular values of $\mathbf{A}$ once at a time, we can obtain the rank-revealing triangular matrix $\mathbf{T}$. Consequently, matrices $\mathbf{U}$ and $\mathbf{V}$ can also be obtained via the givens transformations resulting from each step that we estimate the singular vector [4].

According to the standard perturbation theory [2], the rank-revealing performance of the UTV decomposition will approach the SVD if the norm of the off-diagonal elements of $\mathbf{T}$ approach zero. Specifically, since $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, Equation (2) is actually the SVD when $\mathbf{S}$ and $\mathbf{E}$ are diagonal matrices, and $\mathbf{C}$ and $\mathbf{Z}$ are zero matrices (matrices that all of their elements are zeros). This is a very important property of the UTV decomposition because, as the UTV decomposition preserves the similar property as the SVD, the computational load is reduced. The flop (floating-point operations, i.e. additions and multiplications) count for computing the UTV decomposition is $4m^2k + 4mnkp + 8mnk$ while the flop count for computing the SVD (using Golub-Reinsch SVD algorithm [2]) is $4m^2n + 8mn^2 + 9n^3$. It should be noted that the UTV decomposition is faster than the SVD when $n/k \geq (p+4)/3$, where the integer p is the average number of either power (or Lanczos iterations per deflation step) [7]. More importantly, the updated/downdated algorithms of the UTV decomposition can be done only with $O(n^2)$ operators instead of $O(n^3)$ of the SVD [7], [8].

IV. FASTPCA

A. Calculation of the PCA via UTV decomposition

Similar to the SVD, suppose that $\mathbf{A}$ is the matrix of the size $m \times n$, where $\mathbf{a}_i$ denotes its corresponding i-th column. The autocorrelation matrix, $\mathbf{R}$, of $\mathbf{A}$ can be approximated as $\mathbf{R} = \frac{1}{n}\mathbf{A}\mathbf{A}^T$. According to Sections II and III, $\mathbf{A}$ can be decomposed as $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{V}^T$. Since $\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T = \mathbf{R} = \frac{1}{n}\mathbf{A}\mathbf{A}^T$ and $\mathbf{V}$ is orthonormal, i.e. $\mathbf{V}^T\mathbf{V} = \mathbf{I}$,

$$\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T = \frac{1}{n}\mathbf{U}\mathbf{T}\mathbf{V}^T\mathbf{V}\mathbf{T}^T\mathbf{U}^T = \frac{1}{n}\mathbf{U}\mathbf{T}\mathbf{T}^T\mathbf{U}^T.$$

If $\mathbf{T}\mathbf{T}^T$ is close to a diagonal matrix, we can approximate $\mathbf{\Lambda} \approx \frac{1}{n}\mathbf{T}\mathbf{T}^T$ and hence $\mathbf{\Phi} \approx \mathbf{U}$. In other words, the PCA matrix of the fastPCA can be approximately calculated by employing the matrix $\mathbf{U}$ of the UTV decomposition. Hence, the fastPCA-transformed vector $\mathbf{y}$ of a random vector $\mathbf{x}$ can be defined as $\mathbf{y} = \mathbf{U}^T\mathbf{x}$.

B. Performance Analysis of fastPCA

The mathematical performance of the proposed fastPCA can be analyzed as follows:

Theorem 1: If $\mathbf{T}\mathbf{T}^T$ is diagonal, then $\mathbf{\Phi}$ spans the same space as $\mathbf{U}$.

Proof: According to the SVD, $\mathbf{A} = \mathbf{U_s}\mathbf{\Sigma}\mathbf{V_s}^T$, we can easily show that $\mathbf{\Lambda} = \frac{1}{N}\mathbf{\Sigma}^2$ and $\mathbf{\Phi} = \mathbf{U}_s$. Now, let $\mathbf{T}$ be the upper triangular matrix. Suppose that $\mathbf{T}\mathbf{T}^T$ is diagonal, according to (2), the off-diagonal elements $\mathbf{C}$ and $\mathbf{E}$ are equal to zeros. According to [4], bound of the subspace distance (dist) between the SVD and UTV is given as

$$\text{dist}(\Re(\mathbf{U}), \Re(\mathbf{U}_s)) \leq \frac{\|\mathbf{C}\|_2 \|\mathbf{E}\|_2}{\sigma_{\min}(\mathbf{S})^2 - \|E\|_2^2}, \quad (3)$$

where $\Re(\mathbf{U})$ and $\Re(\mathbf{U}_s)$ denote the column spaces (ranges) of the matrices $\mathbf{U}$ and $\mathbf{U}_s$, respectively. $\sigma_{\min}(\mathbf{S})$ denotes the minimum diagonal element of $\mathbf{S}$ in (2). That is the right hand side of (3) is zero, hence $\mathbf{U}$ and $\mathbf{U}_s$ span the same space. Similarly, the same conclusion can also be proved when $\mathbf{T}$ is the lower triangular matrix.

V. APPLICATION ON EEG ANALYSIS

The fastPCA could be efficiently used in applications such as real-time EEG (or ECG) compression and analysis, feature extraction via eigen analysis, data reduction, etc. In this paper, we demonstrate the usefulness of the fastPCA by analyzing the frequency of the bases derived from the ensemble of each EEG channel.

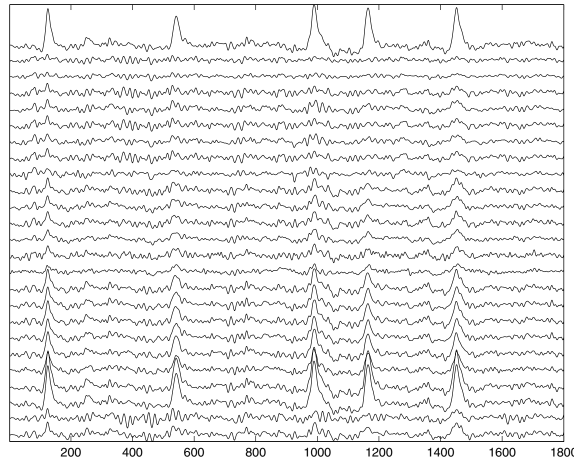

Fig. 1. Original 25-channel EEG contaminated by eyeblink artifacts (channels 1 to 25 arrange from bottom to top, respectively); y-axis denotes the sample numbers

A. Data Acquisition

The fastPCA is applied to real EEG measurements. The database of EEGs contaminated by the eyeblink artifacts is provided by the School of Psychology, Cardiff University, UK. The scalp 25-channel EEG was obtained using Silver/Silver-Chloride electrodes placed at locations defined by the 10-20 system [9]. The 8-second of 25-channel EEG was sampled at 200 Hz (1,600 samples/channel), and bandpass filtered with cut-off frequencies of 2 Hz and 30 Hz.

B. Simulation Results

According to [5], we can simultaneously summarize the frequency information of the 25-channel EEG by calculating its fast Fourier transform (FFT) or power spectral density (PSD) of their bases (each column of the PCA (or KLT) matrix $\mathbf{\Phi}^T$ calculated as in II). In this section, we select two sets of data in Fig.1 to evaluate the merit of the fastPCA. The first dataset (called Data1) is selected from sample numbers 301 to 500 and the second dataset (called Data2) is selected from sample numbers 401 to 600. That is the matrix $\mathbf{A}$ to be decomposed by the SVD and UTV decomposition is of the size 200×25. It is clear that Data1 is selected to avoid the eyeblink artifact while there is one eyeblink artifact appeared in Data2. Fig. 2 illustrates the estimated singular values and singular values of the UTV decomposition and SVD, respectively, of both Data1 (Fig. 2(a)) and Data2 Fig. 2(b). This implies that both Data can be optimally represented (in the mean square error sense) by employing approximately 4-5 bases. In order to summarize the frequency of the 25-channel EEG, we take the FFT of each column vector of matrix $\mathbf{U}$ of the UTV decomposition compare with each corresponding column vector of matrix $\mathbf{U}_s$ of the SVD. Figs.3 and 4 show that the fastPCA can efficiently summarize the contribution of each frequency band in both multichannel EEGs from Data1 and Data2 as well as the traditional PCA. In Figs. 3(a) and (d), both the fastPCA and PCA result in basis which can represent the trend of the multichannel EEG at the frequency around 2 Hz. Figs. 3(b) and (e) show that both the fastPCA and PCA result in the basis which can represent alpha activities around 9-13 Hz. Figs. 3(c) and (f) demonstrate the similar results of both the fastPCA and PCA which can reveal the beta activities around 12-18 Hz. Similarly, the same result can be obtained in Fig. 4 except that Figs. 4(a) and (d) reveal the eyeblink artifact together with the trend of the multichannel EEG at the frequency around 2-4 Hz.

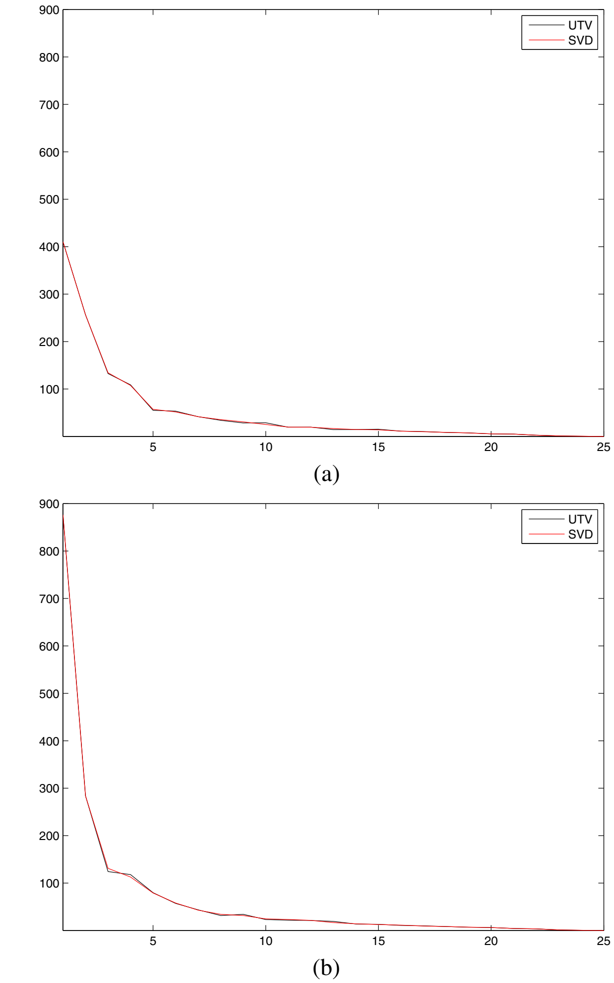

Fig. 2. Estimated singular values and singular values of the UTV decomposition and SVD of (a) Data1 and (b) Data2.

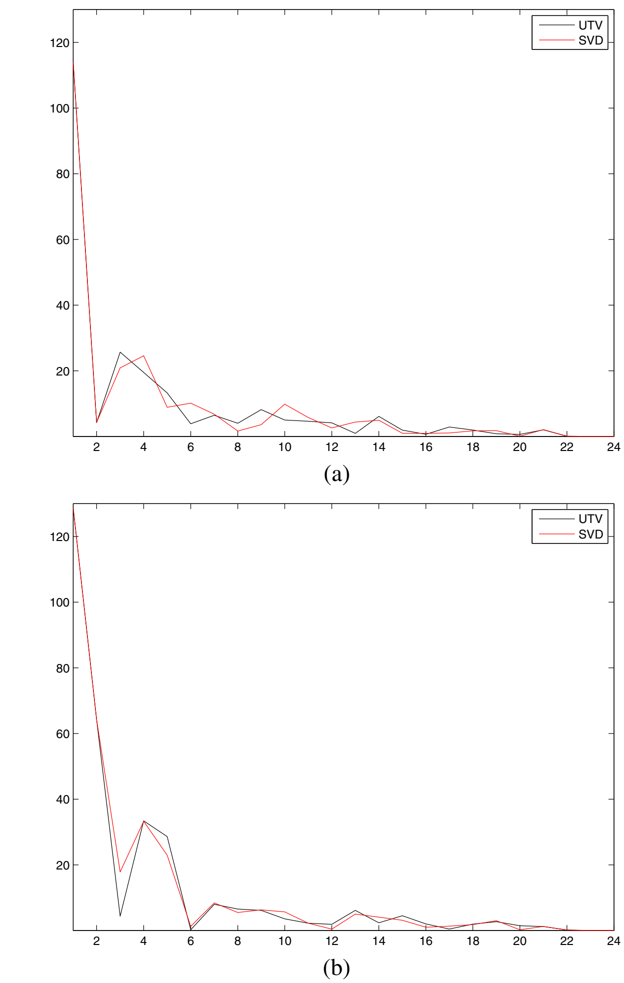

Fig. 5. The absolute values of the PCA-transformed coefficients calculated by the UTV decompostion and SVD of (a) Data1 and (b) Data2.

Fig.5 demonstrates that the absolute PCA-transformed coefficients calculated via the fastPCA can efficiently approximate the traditional PCA calculated using the SVD. For better visualization, 24 out of 200 samples of the PCA-transformed coefficients of the first channel are depicted in Fig.5. Both Data1 (Fig.5(a)) and Data2 (Fig.5(b)) result in the same trend of the energy compaction except that the EEG from Data1 seems to be smoother than Data2, hence, the energies decay faster.

Since we only use 200 samples for each channel of the EEG, computational load on calculating the SVD seems to be acceptable. However, if we would like to continuously include or exclude some data (e.g. continuously update the data in real time application, or in the case that we would like to update some more data from 200 samples to 300 samples and so on to 1,800 samples, or we would like to downdate some old data and update some new data), fastPCA can obviously outperform the traditional PCA in the sense of computational complexity (see Section III)

VI. Conclusion

This paper has presented the new method to compute the PCA-transform matrix via the UTV decomposition. The proposed method called fastPCA can efficiently update and downdate the PCA-transform matrix with lower complexity than the traditional PCA calculated via the SVD. The mathematical performance analysis of the fastPCA has also been illustrated. The fastPCA has been applied to simultaneously analyze the frequency information of the multichannel EEG.

VII. Acknowledgments

The author would like to thank Dr. Kianoush Nazarpour and Prof. Saeid Sanei, the school of engineering, Prof. Edward Wilding, the school psychology, Cardiff University, for providing the dataset.

References

[1] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1986.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1996.

[3] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, pp. 1535–1541, Jun. 1992.

[4] R. D. Fierro, P. C. Hansen, and P. S. K. Hansen, "UTV Tools: MATLAB templates for rank-revealing UTV decompositions," *Numerical Algorithms*, vol. 20, pp. 165–194, 1999.

[5] C.Tenke and J.Kayser, "Reference-free quantification of EEG spectra: Combining current source density (CSD) and frequency principal components analysis (fPCA)," *Clinical Neurophysiology*, vol. 116, pp. 2826–2846, 2005.

[6] K. R. Rao and P. C. Yip, *The Transform and Data Compression Handbook*, CRC Press, Boca Raton, FL, 2001.

[7] J. L. Barlow, P. A. Yoon, and H. Zha, "An algorithm and a stability theory for downdating the ulv decomposition," *BIT Numerical Mathematics*, vol. 36, pp. 14–40, Mar. 1996.

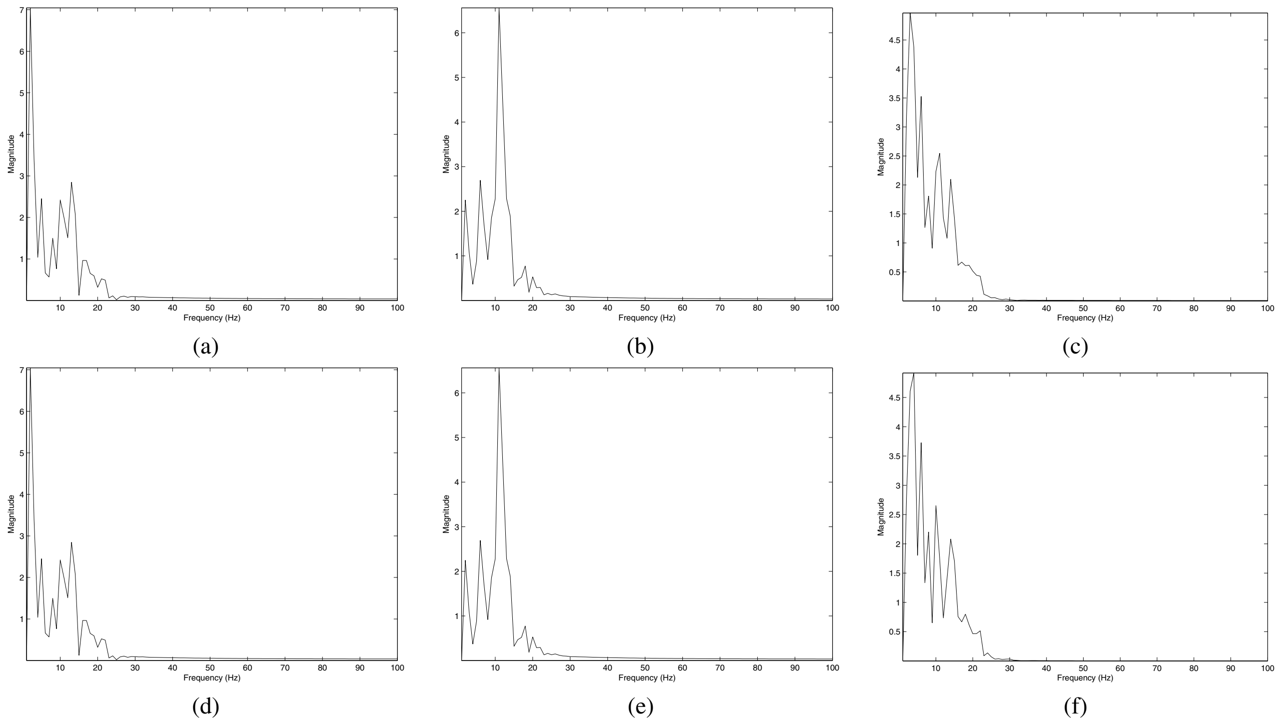

Fig. 3. FFT of (a) the 1-st column, (b) the 2-nd column and (c) the 3-rd column of $\mathbf{U}$ calculated from the UTV decomposition of Data1; FFT of (c) the 1-st column, (d) the 2-nd column and (e) the 3-rd column of $\mathbf{U}s$ calculated from the SVD of Data1.

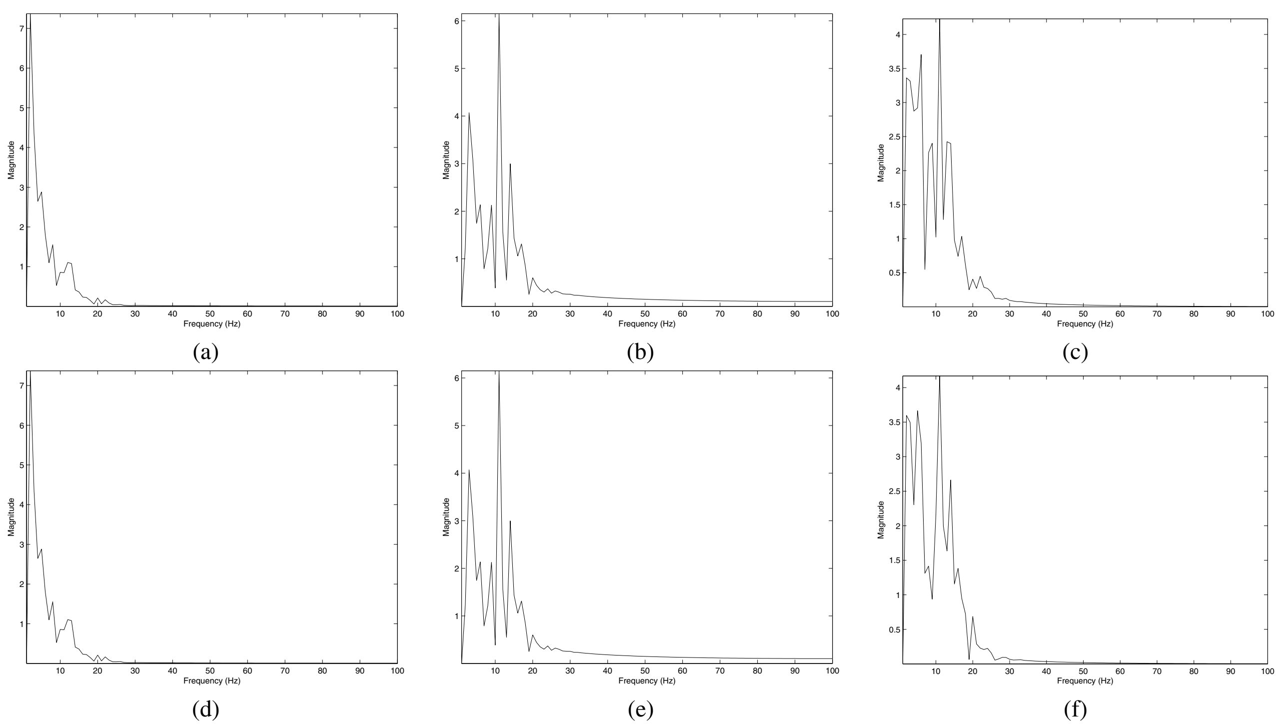

Fig. 4. FFT of (a) the 1-st column, (b) the 2-nd column and (c) the 3-rd column of $\mathbf{U}$ calculated from the UTV decomposition of Data2; FFT of (c) the 1-st column, (d) the 2-nd column and (e) the 3-rd column of $\mathbf{U}s$ calculated from the SVD of Data2.

[8] R. D. Fierro and P. C. Hansen, "Lower-rank revealing UTV decomposition," *Numerical Algorithms*, vol. 15, pp. 37–55, 1997.
[9] S. Sanei and J. A. Chambers, *EEG Signal Processing*, John Wiley and Sons, West Sussex, England, 2007.