MULTI-CHANNEL EEG COMPRESSION BASED ON MATRIX AND TENSOR DECOMPOSITIONS

Justin Dauwels*   *Srinivasan K.*†*   *Ramasubba Reddy M.†*   *Andrzej Cichocki‡*

* Nanyang Technological University, Singapore, E-mail: justin@dauwels.com
† Indian Institute of Technology Madras, Chennai, India, E-mail: srinivasan.sivam@gmail.com
‡ RIKEN BSI, Laboratory for Advanced Brain Signal Processing, Wako-Shi, Japan

ABSTRACT

Compression schemes for EEG signals are developed based on matrix and tensor decomposition. Various ways to arrange EEG signals into matrices and tensors are explored, and several matrix and tensor decomposition schemes are applied, including SVD, CUR, PARAFAC, the Tucker decomposition, and recent random fiber selection approaches. Rate-distortion curves for the proposed matrix and tensor-based EEG compression schemes are computed. It shown that PARAFAC has the best compression performance in this context.

Index Terms— tensor decomposition, EEG, Tucker, PARAFAC, SVD

1. INTRODUCTION

Electroencephalograms (EEG) are electrical signals recorded along the scalp or brain surface, generated by firing of neurons within the brain [1]. In various clinical applications, EEG signals are sometimes continuously recorded over extended periods of time (several days, weeks, or potentially even months):

- in neurology intensive care units (ICU), e.g., for stroke patients,

- in telemedicine for neurological patients, where EEG is continuously recorded outside the hospital.

Long-term EEG recordings result in massive EEG data sets. Therefore, it is required to compress the EEG signals before storage or transmission. The main challenges for EEG compression are as follows:

- the number of EEG channels can be large (e.g., 256), especially if accurate inverse modeling is needed,

- high sampling rate may be required (several kHz in the case of cortical EEG; several hundred Hz for scalp EEG), to capture spikes and high-frequency oscillations in the EEG

A variety of techniques have been developed for compressing EEG signals; we refer to [2] for an excellent review on lossless EEG compression. In the following, we will briefly outline the state-of-the-art in EEG compression.

EEG signals are often modeled as auto-regressive (AR) processes, similarly as speech signals. Various AR predictors have been developed: linear AR predictor [2], least squares [2] and adaptive neural network predictors [3]. Further refinements include context-based bias cancelation [4] and detailed prediction residual modeling [3], which improve compression performance at the expense of computational complexity.

Before compressing signals, it may be fruitful to first transform them in an other domain. Various common transformations have been used for EEG compression, including discrete cosine transform [2], subband transform, wavelet transform, wavelet-packet transform, and integer lifting wavelet transform [2, 5].

The emerging field of compressed sensing allows to acquire sparse signals with very few random measurements, well below the Nyquist rate. Compressed sensing of EEG has been explored in a few recent studies [6, 7].

A large number of studies have been devoted to the compression of EEG. However, most of them consider the compression of single EEG signals; multi-channel compression of EEG is less intensively studied. Multi-channel data can naturally be represented as tensors ("multi-way representation"). Multi-way analysis has been applied to EEG signals (especially non-negative decompositions of time-frequency maps), mostly for extracting features (see, e.g., [8, 9]); those studies seem to suggest that multi-way analysis may be effective for EEG compression as well, however, many issues still need to be systematically explored. In this paper, we present results from such systematic study.

In this study, we utilize a variety of matrix and tensor decomposition methods to approximate EEG signals. We explore several ways to arrange EEG signals in the form of matrix and tensor, and we evaluate several matrix and tensor decomposition schemes [10]. Such decomposition schemes include the singular value decomposition (SVD), column-row decomposition for matrices (CUR), PARAFAC, Tucker decomposition, and fiber-sampling tensor decomposition for tensors. We compute rate-distortion curves for our proposed tensor-based schemes for EEG compression.

First we review the matrix and tensor decomposition methods that we consider in this paper (Section 2), and we explain how we arrange EEG data into matrix and tensor form (Section 3). We describe the EEG data analyzed in this study (Section 4), and present results for our matrix and tensor decomposition based compression schemes (Section 5). At the end of the paper, we describe ongoing and future work.

2. MATRIX AND TENSOR DECOMPOSITIONS

Tensors or multi-way arrays provide a natural representation for multi-dimensional data. Tensor decomposition models are important tools for feature extraction and classification since they capture the dependencies in higher-order data-sets. They have found application in many areas, e.g., psychometrics, chemometrics, and signal processing [10, 11]. In this paper, we investigate two matrix factorization methods (SVD and CUR) and four tensor decomposition techniques (PARAFAC, Tucker, Fiber-sampling tensor decomposition, and compact tensor decomposition); in the following, we will briefly review those schemes. (We refer to [10] for an excellent review of tensor decompositions. We use the same notation as in [10]).

2.1. Singular value decomposition (SVD)
A Singular Value Decomposition of a matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ is of the form

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \qquad (1)$$

where $\mathbf{U} \in \mathbb{R}^{M \times M}$, $\mathbf{V} \in \mathbb{R}^{N \times N}$ are the left and right singular vectors respectively. The matrix $\mathbf{\Sigma} \in \mathbb{R}^{M \times N}$ is diagonal: $\Sigma = [\text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_R\}; 0]$, where σ_i are the singular values of $\mathbf{X}$ with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R$ and $R := min(M, N)$. The matrix $\mathbf{X}$ can be approximated by r-singular vectors with $r < R$:

$$\hat{\mathbf{X}} \approx \sum_{i=1}^{r} \mathbf{u}_i \sigma_i \mathbf{v}_i^T. \qquad (2)$$

2.2. Column-Row decomposition (CUR)
CUR decomposition ("pseudo-skeleton decomposition") decomposes a matrix in terms of rows and columns of the original matrix [12]:

$$\mathbf{X} \approx \mathbf{C}\mathbf{U}\mathbf{R} \qquad (3)$$

where $\mathbf{C}$ and $\mathbf{R}$ consists of selected columns and columns respectively of $\mathbf{X}$, and the matrix $\mathbf{U}$ is constructed such that the product $\mathbf{C}\mathbf{U}\mathbf{R}$ is as close as possible to $\mathbf{X}$. This method has been widely used in data analysis tasks, where the interpretation via SVD decomposition is difficult due to the orthogonality of the singular vectors.

2.3. Parallel factor decomposition (PARAFAC)
PARAFAC decomposes a tensor $\mathcal{X}$ into rank-one tensors (vectors):

$$\mathcal{X} \approx \sum_{i=1}^{r} \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i \qquad (4)$$

where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ represent the factors along the three modes, whereas $\circ$ stands for the outer-product along the particular mode. When r equals the rank of the tensor, the PARAFAC decomposition is exact. For the purpose of compression, we set $r < R$.

2.4. Tucker decomposition
The Tucker method decomposes a tensor $\mathcal{X}$ into a core tensor $\mathcal{G} \in \mathbf{R}^{P \times Q \times R}$ and factor matrices ("basis matrices") $\mathbf{A} \in \mathbb{R}^{N \times P}$, $\mathbf{B} \in \mathbb{R}^{N \times Q}$ and $\mathbf{C} \in \mathbb{R}^{M \times R}$:

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \qquad (5)$$

The factor matrices capture the variation along the three modes and the core tensor captures the interaction between them.

2.5. Fiber-sampling tensor decomposition (FSTD)
The FSTD method is an extension of the CUR matrix decomposition to tensors; a tensor is represented by a subset fibers selected along each mode [13]:

$$\mathcal{X} \approx \mathcal{U} \times_1 \mathbf{C}_1 \times_2 \mathbf{C}_2 \times_3 \mathbf{C}_3 \qquad (6)$$

where $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ are matrices formed by fibers along the mode indicated in the subscripts. The core tensor $\mathcal{U}$ is expressed in terms of intersection subtensor $\mathcal{W}$ formed by the intersection of the selected fibers along the three modes:

$$\mathcal{U} = \mathcal{W} \times_1 \mathbf{W}_{(1)}^{\dagger} \times_2 \mathbf{W}_{(2)}^{\dagger} \times_3 \mathbf{W}_{(3)}^{\dagger}, \qquad (7)$$

where $\mathbf{W}_{(i)}^{\dagger}$ represents the pseudo-inverse of the intersection sub-tensor matriced along the mode-i. The tensor is expressed in terms of a subset of elements from the original tensor; this concept is similar to compressed sensing.

2.6. Compact tensor decomposition (TT)
The compact tensor decomposition method ("TT decomposition") approximates any d-dimensional tensor into two matrices along mode-1 and mode-d, and $d - 2$ three-way tensors [14]. In particular, a three-way array $\mathcal{X} \in \mathbb{R}^{N \times N \times M}$ is decomposed as:

$$\mathcal{X} \approx \mathcal{U} \times_1 \mathbf{G}_1 \times_3 \mathbf{G}_3 \qquad (8)$$

where $\mathcal{U} \in \mathbb{R}^{R_1 \times N \times R_2}$, $\mathbf{G}_1 \in \mathbb{R}^{N \times R_1}$ and $\mathbf{G}_3 \in \mathbb{R}^{N \times R_3}$ are the tensor and matrices formed of decomposition. This method is numerically well behaved and considered as an alternative to PARAFAC. Note that the TT decomposition (8) only tries to capture dependencies along two of the three modes (1 and 3), in contrast to PARAFAC, which models relations among all three modes.

3. MATRIX AND TENSOR FORMATION FROM EEG

Multi-channel EEG signals (M signals of length $L = N^2$) can naturally be arranged in the form of a matrix $\mathbf{X} \in \mathbb{R}^{M \times L}$ or tensor $\mathcal{X} \in \mathbb{R}^{N \times N \times M}$. Such arrangement will allow us to explore compact representations of EEG signals. We have considered the following constructions so far:

- **Matrix**: The EEG signals from each of the M channels are arranged as rows of the matrix:

$$\mathbf{X}(k,:) = EEG^{(k)}(1:L) \qquad \forall\, k = 1,\dots,M \quad (9)$$

- **Time tensor** : Single-channel EEG signals are arranged to form a matrix [5]. This matrix is then stacked to form a three-way tensor $\mathcal{X} \in \mathbb{R}^{N \times N \times M}$, whose frontal slices represents signal from a particular channel. It can be expressed by the following relationship:

$$\mathbf{X}_{::k} = \mathrm{matrix}(EEG^{(k)}(1:N^2)) \qquad (10)$$

where $N \times N$ is the dimension of the frontal slice, M the number of channels, $EEG^{(k)}$ the EEG from channel k and matrix() is function that arranges EEG in the form of a matrix; the entries are filled starting at the top left-hand side, from left to right on the odd rows, and from right to left on the even rows.

- **Wavelet tensor** : Wavelet tensor is derived from the time tensor $\mathcal{X}$, by subjecting each frontal slice to 2-D discrete wavelet transform:

$$\mathbf{X}_{::k}^{w} = \text{2D-DWT}(\mathbf{X}_{::k}) \qquad \forall\, k = 1,\dots,M. \quad (11)$$

In the matrix construction, neighboring entries in the rows represent adjacent time-samples from the same EEG channel, whereas neighboring entries in the columns represent samples from adjacent channels at the same time instance. In the time-tensor construction, the mode-1 (column) fibres represents samples from the same channel displaced by N samples. Mode-2 (row) fibres represent adjacent samples from the same channel, whereas entries along the mode-3 fibres (tubes) represent the samples from adjacent channels at the same time instance.

4. EEG DATA SETS

We describe here the two data EEG sets that we have considered in this study.

4.1. Dataset-1 : MCI Vs Control

The first EEG data set comprises two study groups. The first group consists of 22 patients who had complained of memory problems. These subjects were diagnosed as suffering from mild cognitive impairment (MCI) and subsequently developed mild AD (MiAD). The EEG recordings were conducted while all patients were in the MCI stage. The criteria for inclusion into the MCI group were a mini mental state exam (MMSE) score = 24, though the average score in the MCI group was 26 (SD of 1.8). The other group is a control set consisting of 38 age-matched, healthy subjects who had no memory or other cognitive impairments. The average MMSE of this control group is 28.5 (SD of 1.6).

Ag/AgCl electrodes (disks of diameter 8mm) were placed on 21 sites according to 1020 international system, with the reference electrode on the right ear-lobe. EEG was recorded with Biotop 6R12 (NEC San-ei, Tokyo, Japan) at a sampling rate of 200Hz, with analog bandpass filtering in the frequency range 0.5-250Hz and online digital bandpass filtering between 4 and 30Hz, using a third-order Butterworth filter (forward and reverse filtering).

This data set has been analyzed in various studies (e.g., [16]), and we refer to the latter for more detailed information.

4.2. Dataset-2: EEG-MMI database

The second EEG data set consists of three 64-channel recordings chosen from the EEG-Motor Mental Imagery datasets of physiobank database [15]. The EEG is recorded from healthy subjects at 80Hz sampling rate and with 12 bit resolution.

5. RESULTS

We arranged the EEG data described in Section 4 as matrices and tensors, as described in Section 3. Next we applied the matrix/tensor decomposition schemes of Section 2 to the resulting matrices and tensors, resulting in various compression schemes. We assess the performance of the latter by the compression ratio (CR), defined as the ratio of samples in the original EEG data to the number of elements in the matrix/tensor decomposition. The quality of the reconstructed signal is measured as the percent-root mean square distortion (PRD):

$$PRD(\%) = \sqrt{\frac{\sum_i^{N^2} (x(i) - \tilde{x}(i))^2}{\sum_i^{N^2} x(i)^2}} \times 100. \quad (12)$$

The results are summarized in Fig. 1 and Fig. 2. Since the results for time-tensor and wavelet-tensor construction were about the same, we only show results for the time-tensor construction. The results were obtained by varying the dimensions in the matrix/tensor decomposition, including number of singular values (cf.(2)) and rank-one tensors (cf. (4)), columns/rows (cf. (3)) or fibers (cf. (6)). We computed Tucker and TT decomposition for all possible sizes of the core tensor $\mathcal{G}$ and $\mathcal{U}$ respectively. Next, for fixed compression ratios we chose the core tensor size that minimizes the distortion. To avoid overfitting, we performed leave-one-out crossvalidation to assess Tucker and TT decomposition. We considered EEG segments that are $L = 256$ and 1024 samples long, and hence the dimension N in (10) equals 16 and 32 respectively.

As can be seen from Fig. 1 and Fig. 2, the tensor-based compression schemes clearly outperform matrix-based compression schemes, especially at large compression ratios. The results are better for larger EEG segments ($L = 1024$ and $N = 32$). The smallest reconstruction error was obtained with PARAFAC, followed by Tucker decomposition, and then TT decomposition. In this context, CUR and FSTD yield the largest approximation error. However, they do not require the entire matrix $\mathbf{X}$ or tensor $\mathcal{X}$ to be given, in contrast to the other methods; as a result, CUR and FSTD may be useful for compressive sensing, where the EEG is sampled far
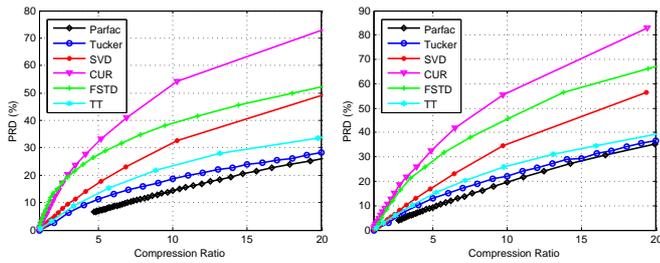
Fig. 1. Percent-root-mean-square distortion variation with compression ratio for the dataset-1 (MCI), for $L = 256$ (right) and $L = 1024$ (left).
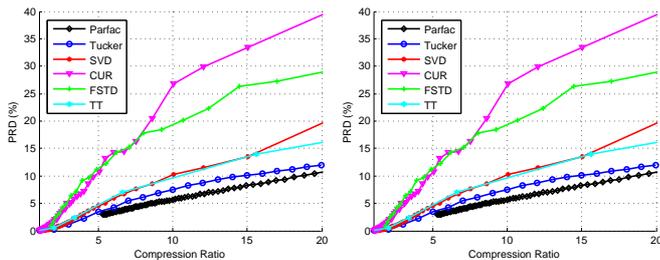
Fig. 2. Percent-root-mean-square distortion variation with compression ratio for the dataset-2 (EEG-MMI database), for $L = 256$ (right) and $L = 1024$ (left).

below Nyquist rate. Most likely, for such methods to be effective though, we will need to explore alternative ways to arrange EEG data in matrix and tensor form, for example, using continuous wavelet transforms. Interestingly, the compression error was significantly smaller for MCI patients than for age-matched control subjects (dataset-1). In future work, we will explore the use of compression schemes in the realm of diagnosing MCI and AD.

6. CONCLUSION

Conventional multi-channel compression schemes usually exploit the intra-channel and inter-channel correlation in separate stages; matrix and tensor-based compression schemes capture intra-channel and inter-channel correlations simultaneously, which is a more elegant approach that seems to yield good compression performance.

We are currently exploring various extensions. For example, the matrix and tensor decompositions may be applied recursively, by applying them to the residual matrix/tensor. Alternatively, the matrices and tensors in the decompositions (e.g., core tensor $\mathcal{G}$ in (5)) may be further compressed, e.g., using arithmetic coding.

We are also considering unequal error protection; it is well-known that the high frequency components in EEG signals are the most contaminated, since they are usually the weakest. Therefore, we can allow higher approximation error for the entries in the wavelet tensor corresponding to high-frequency EEG components. At the same time, we would like to keep the approximation errors for the low-frequency entries as small as possible.

7. REFERENCES

[1] Paul Nunez and Ramesh Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, 2006.

[2] Giuliano Antoniol and Paolo Tonella, "EEG data compression techniques," *IEEE Transactions on Biomedical engineering,*, vol. 44, no. 2, pp. 105–114, Feb. 1997.

[3] N.Sriraam and C.Eswaran, "An adaptive error modeling scheme for the lossless compression of EEG signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 5, pp. 587–594, Sep. 2008.

[4] Nasir Memon, Xuan Kong, and Judit Cinkler, "Context-based lossless and near-lossless compression of EEG signals," *IEEE Transactions on Biomedical Engineering*, vol. 3, no. 3, pp. 231–238, mar. 1999.

[5] K.Srinivasan and M. Ramasubba Reddy, "Efficient preprocessing technique for lossless real-time EEG compression," *Electronics Letters*, vol. 46, no. 1, pp. 26–27, Jan. 2010.

[6] Kok-Kiong Poh and Pina Marziliano, "Compressive sampling of EEG signals with finite rate of innovation," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–12, 2010.

[7] S. Aviyente, "Compressed Sensing Framework for EEG Compression," *Proceedings of IEEE Statistical Signal Processing Workshop,* August 2007.

[8] Morten Mørup, Lars Kai Hansen, Christoph S. Herrmann, Josef Parnas, and Sidse M. Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related eeg," *NeuroImage*, vol. 29, no. 3, pp. 938 – 947, 2006.

[9] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, Anh-Huy Phan, Seungjin Choi, Hyekyoung Lee, Qibin Zhao, Liqing Zhang, and Yuanqing Li, "Noninvasive BCIs: Multiway signal-processing array decompositions," *Computer*, vol. 41, no. 10, pp. 34 –42, oct. 2008.

[10] Tamara G. Kolda and Brett W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, September 2009.

[11] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun ichi Amari N, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, 2009.

[12] Michael W. Mahoney and Petros Drineas, "CUR matrix decompositions for improved data analysis," *Proc. National Acad. of Sciences*, vol. 106, no. 3, pp. 697–702, 2009.

[13] Cesar F. Caiafa and Andrzej Cichocki, "Generalizing the column-row matrix decomposition to milti-way array," *Linear Algebra and its Applications*, vol. 433, pp. 557–573, 2010.

[14] I.V Oseledets, "Compact matrix form of the d-dimensional tensor decomposition," Preprint 2009-01, INM RAS, 2009.

[15] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–220, 2000.

[16] J. Dauwels, F. Vialatte, M. Vialatte, and A. Cichocki, A Comparative Study of Synchrony Measures for the Early Diagnosis of Alzheimers Disease Based on EEG, NeuroImage, 2010, 49:668-693.