GRAPH REGULARISED TENSOR FACTORISATION OF EEG SIGNALS BASED ON NETWORK CONNECTIVITY MEASURES

Loukianos Spyrou and Javier Escudero

School of Engineering, University of Edinburgh

ABSTRACT

Tensor factorisation is a decomposition method for high dimensional data that is used to estimate the prominent factors in some signal. Recently it has been employed with success in the biomedical fields. Regularised tensor factorisation attempts to alleviate overfitting and small sample size estimation errors by constraining the obtained solution to satisfy some metric. In this work, we provide a novel extension to the theory of graph regularisation for regularising multiple graphs and we employ graph regularised tensor factorisation on an electroencephalogram (EEG) dataset. We utilise brain connectivity networks as the basis of our graphs. Subsequently, we perform graph regularised tensor factorisation on the EEG data in order to reduce the noise and interference inherent to the EEG. Furthermore, we employ custom graphs that incorporate prior knowledge of our dataset. We demonstrate the efficacy of the algorithm theoretically and on some real EEG examples. Further applications of the algorithm can be neuroscience applications where there is prior knowledge of the relations between the data and in general in network science for datasets that can be expressed as tensors.

Index Terms— tensor factorisation, graph, regularisation, EEG

1. INTRODUCTION

Tensor factorisation has found many applications in several areas such as antenna array processing, blind source separation, biomedical signal processing, feature extraction, and classification [1]. A tensor is a multi-way representation of data or a multidimensional array. Each dimension in the tensor is called a mode or a way. Using tensor factorisation, the true underlying structure of that data can be preserved. Tensor factorisation methods have been shown to be powerful for describing signals which in general change in time, frequency, and space. Tensor analysis can provide a good way to discover the main features of the data and extract the hidden underlying information especially in the case of having big data size.

Several tensor based methods have been suggested for decomposition and multi-way representation of data. The PARAFAC decomposition [2, 3] is one of the common tensor factorisation methods which is a generalisation of singular value decomposition (SVD) to higher order tensors. Using the Parafac model, the data are decomposed into a sum of rank-1 tensors of lower dimensions than the original data. Therefore, as suggested in [4], it can be employed to compress the high dimensional data and extract their significant features.

The application of tensor decomposition can be significant for biomedical signals, such as EEG, where many transient events and movement related sources and artifacts are involved and most sources are inherently nonstationary. Moreover, the related brain neural process exhibit specific space-time-frequency locations. EEG signals in particular, consist of multichannel recordings with good temporal resolution which subsequently offers good time-frequency resolution. The application of tensor analysis is then logical and the data can be factorised into its space, time and frequency modes [5]. Tensor factorisation has been also applied to multi-subject data where the data can be factorised in the group level, identifying the common components [5, 6].

Tensor factorisation has been employed in brain connectivity studies primarily with the aim of dimensionality reduction or detection of dynamic changes [7, 8]. Graph theory has found applications in many scientific fields in an attempt to analyse interconnections between phenomena, measurements, and systems [9]. A graph consists of a set of nodes and edges describing the connections between the nodes. The edges of weighted graphs describe the strength of the connections. Graphs have been extensively used in a variety of applications in network science such as biological networks, brain networks, and social networks [10, 11, 12, 13]. In brain networks, graphs have been used to describe the brain connectivity [14] which is a measure of the functional and structural integration of the brain.

In this work we combine tensor factorisation and graph theory in an attempt to enhance the capabilities of tensor analyis for EEG data. We use graph regularised tensor factorisation of a four mode tensor (space, frequency, trial, epoch [1]) in order to denoise the EEG and drive the decomposition process based on prior information. Graph regularisers were first theoretically proposed in [15], with applications in matrix

This work was supported by EPSRC, UK, Grant No. EP/N014421/1.

[1]We differentiate between trials, each repetition of the stimuli that produce the brain response, and epoch, the different periods within a trial

factorisation [16, 17] and also tensor factorisation [18]. By using graph regularisation on the frequency mode the tensor components obtained through factorisation will essentially be constrained to satisfy the properties of the underlying brain network. Similarly, in the epoch mode, the relations between the different time periods can be incorporated into the factorisation process. In section 2, we describe the theory behind tensor factorisation. Section 3, introduces the concepts of graphs and graph regularised tensor factorisation. Furthermore, we describe our extension to graph regularisation for multiple graphs. Section 4 shows experimental results and Section 5 concludes the paper.

2. TENSOR FACTORISATION

Tensor factorisation was introduced in [19] and later refined in [20, 21]. It is a generalisation of singular value decomposition (SVD) to higher dimensions. The PARAFAC model [2], factorises a tensor into a sum of rank-1 tensors. As an example, in a four-way Parafac model, each element of a four-way tensor $\underline{\mathbf{X}}$ is factorised into N components in all modes similar to the SVD as:

$$\underline{\mathbf{X}} = \sum_{i}^{I} \mathbf{A}_{(i)} \otimes \mathbf{B}_{(i)} \otimes \mathbf{C}_{(i)} \otimes \mathbf{D}_{(i)} \tag{1}$$

where $\otimes$ refers to the outer product operation. It can also be expressed in mode-i multiplications as:

$$\underline{\mathbf{X}} = \mathbf{J} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D} \tag{2}$$

where $\times_i$ refers to tensor-matrix multiplication in the direction of its ith slab [22]. The tensor $\underline{\mathbf{J}}$ is a tensor with ones at the superdiagonal entries and zeroes everywhere else.

The matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are estimated in an alternating least squares fashion by fixing all matrices except one and solving an ordinary least squares problem:

$$\underset{\mathbf{A}}{\operatorname{argmin}} ||\underline{\mathbf{X}} - \underline{\mathbf{J}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}||_F^2 \tag{3}$$

and similarly for $\mathbf{B}, \mathbf{C}, \mathbf{D}$. For detailed descriprion of the algorithms involved in tensor factorisation refer to [2, 22].

3. GRAPH REGULARISED TENSOR FACTORISATION

3.1. Graph Signals

A weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ is a structure defined by a finite set of nodes (or vertices) $\mathcal{V}$ with $|\mathcal{V}| = n$, a set of edges $\mathcal{E}$ of the form $(v_i, v_j) \in \mathcal{E}$ with $|\mathcal{E}| = n^2 - n$ and a weighted adjacency matrix $\mathbf{W}$ with $w_{ii} = 0 \,\forall\, i$. The entries w_{ij} in the weighted adjacency matrix $\mathbf{W}$ (weight matrix from now on) indicate the strength of connection between nodes. We assume that networks are normalised, i.e. $w_{ij} \in [0, 1]$.

A signal $\mathbf{y} \in \mathcal{R}^n$ can be defined over a graph $\mathcal{G}$ such that each element of the signal corresponds to a node. That way the relation between the signal elements can be described by the elements of the weight matrix $\mathbf{W}$.

3.2. Graph regularised tensor factorisation

A regulariser is a function that penalises deviations according to some metric. There are numerous applications in the signal processing and machine learning fields. As described in [15], a graph regularising function is of the form:

$$R(\mathbf{A}, \mathbf{L}) = tr\{\mathbf{A}^T \mathbf{L} \mathbf{A}\} \tag{4}$$

where $\mathbf{L}$ is the Laplacian of a graph $\mathcal{G}$ i.e. :

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{5}$$

with $\mathbf{D}$ a diagonal matrix containing the degrees of each node of the graph in the diagonal. The matrix $\mathbf{A}$ can be a set of signals $\{\mathbf{y}\}$ defined over the graph and are placed on the columns of $\mathbf{A}$. In this work, we focus on $\mathbf{A}$ being a mode of the tensor $\mathbf{X}$. The regulariser can equivalently be written as:

$$R(\mathbf{A}, \mathbf{W}) = \sum_{i,j} \mathbf{W}_{ij} \sum_{n=1}^{N} (\mathbf{A}_{i,n} - \mathbf{A}_{j,n})^2 \tag{6}$$

This implies that when a connection $\mathbf{W}_{ij}$ between nodes i and j is large, the differences between corresponding entries of $\mathbf{A}$ will be accentuated while for low or zero values of $\mathbf{W}_{ij}$ the converse is true. Therefore, the minimisation of such a function penalises large differences between the i_{th} and j_{th} elements of a mode's components when the network has high values on the corresponding nodes.

The graph regularised tensor factorisation algorithm is similar to the original alternating minimisation algorithm with the addition of the regularisation function [18]:

$$\underset{\mathbf{A}}{\operatorname{argmin}} ||\underline{\mathbf{X}} - \underline{\mathbf{J}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}||_F^2 + \lambda\, R(\mathbf{A}, \mathbf{L}) \tag{7}$$

which results in the following sylvester equation [18, 23]:

$$\mathbf{A}\mathbf{Z}^T\mathbf{Z} + \lambda \mathbf{L}\mathbf{A} = \mathbf{X}\mathbf{Z} \tag{8}$$

where $\mathbf{Z} = (\mathbf{B} \circ \mathbf{C} \circ \mathbf{D})$.

3.3. Multiple graph regularisation

In this work, we provide an extension to the conventional formulation by providing a way to regularising the columns of $\mathbf{A}$ separately with different graphs. That is desirable since we want to have different components of each mode correspond to different brain connectivity networks. This can be accomplished in two ways.

Firstly, by defining a regulariser that operates on single columns as such:

$$R_1(\mathbf{A}_n, \mathbf{L}) = tr\{\mathbf{A}_n^T \mathbf{L}_n \mathbf{A}_n\} \tag{9}$$

That way, each column of $\mathbf{A}$, can be regularised to a different connectivity graph. This type of regulariser is implemented by tensor factorisation that estimates individual components in a deflation procedure [24].

Secondly, we can operate directly on the original regulariser by incorporating the single entry matrix $\mathbf{S}_{kl}$ with a single entry at the k_{th} row and l_{th} column. This is formulated as:

$$R_2(\mathbf{A}, \mathbf{L}) = tr\{\mathbf{S}_{ii}\mathbf{A}^T \mathbf{L} \mathbf{A} \mathbf{S}_{ii}\} \tag{10}$$

where in this case only the i_{th} component of mode $\mathbf{A}$ is regularised to the connectivity network. Since in our work we desire to regularise multiple components this can be accomplished by:

$$R_2(\mathbf{A}, \{\mathbf{L}_1 \ldots \mathbf{L}_n\}) = \sum_{n=1}^{N} tr\{\mathbf{S}_{nn}\mathbf{A}^T \mathbf{L}_n \mathbf{A} \mathbf{S}_{nn}\} \tag{11}$$

In this case R_2 replaces the last term of Eq. (7). This leads to the following generalised sylvester equation:

$$\mathbf{A}\mathbf{Z}^T\mathbf{Z} + \lambda \sum_{n=1}^{N} \mathbf{L}_n \mathbf{A} \mathbf{S}_{nn} = \mathbf{X}\mathbf{Z} \tag{12}$$

This form can be converted to the following [23]:

$$\left(\mathbf{Z}^T\mathbf{Z} \otimes \mathbf{I} + \sum_{n=1}^{N} (\mathbf{S}_{nn} \otimes \lambda \mathbf{L}_n)\right) vec(\mathbf{A}) = vec(\mathbf{X}\mathbf{Z}) \tag{13}$$

where $\otimes$ is the kronecker product and vec is the vectorisation operator. This can be easily solved for $\mathbf{A}$.

3.4. Data tensorisation

EEG signals were recorded for 12 patients with mild cognitive impairment (MCI) and 19 controls while they performed two visual short term memory tasks. A Shape and a Shape-Colour binding task. The EEG data was collected using NeuroScan version 4.3. The EEG was sampled at 250 Hz. A bandpass filter of 0.01-40 Hz was used. 128 EEG channels, corrected for ocular artefacts using ICA, were recorded relying on the 10/20 international system. More information regarding the dataset can be found in [25].

Each subject's data were formed into a 4-mode tensor in spatial (128 electrodes), frequency (15 frequency points of 2Hz bins), epoch (5) and trial (approximately 50 per subject) modes. Epochs indicate the temporal subdivision of trials. We chose 5 overlapping epochs of around $450ms$ to have a small temporal resolution while enabling a good frequency resolution. Tensor factorisation was performed with PARAFAC and with the selection of 6 components per mode ($N = 6$).

4. RESULTS

4.1. Error reduction

In Figure 4.1 we show the group average error reduction of the regularised tensor factorisation algorithm for various penalty parameters λ. We split the data of each patient into train and test sets. The training data of each subject was used to obtain connectivity graphs for each of the 5 epochs. Connectivity was estimated with the imaginary part of coherence [26]. We used those graphs to regularise the spatial mode of the tensor factorisation on the test set. This resulted in a reduction of around 5% in test error. Similarly in Fig 4.1 we show the value of the regularisation function as λ increases.

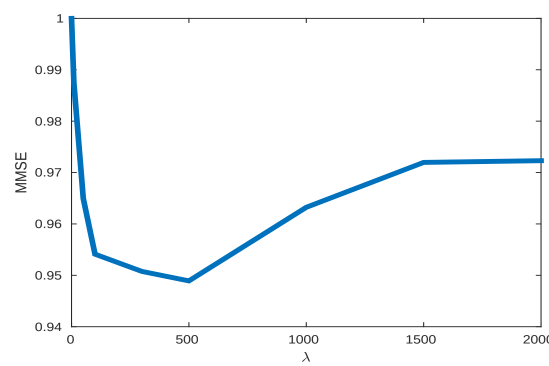

Fig. 1. Group average error ratio between the denoised data and the training average as a function of the regularisation parameter. Each subject's data were half split in a train set where the graphs were learned and a test set where the factorisation was performed.

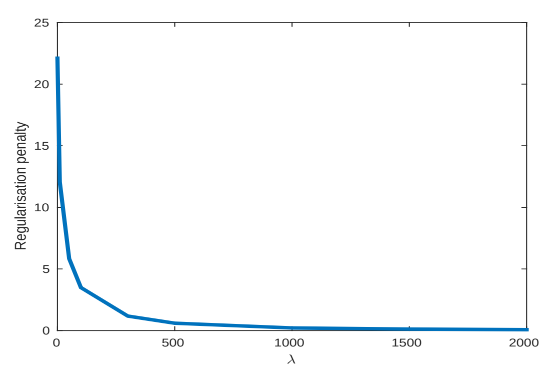

Fig. 2. Group average of the value of the regularisation term as a function of the regularisation parameter.

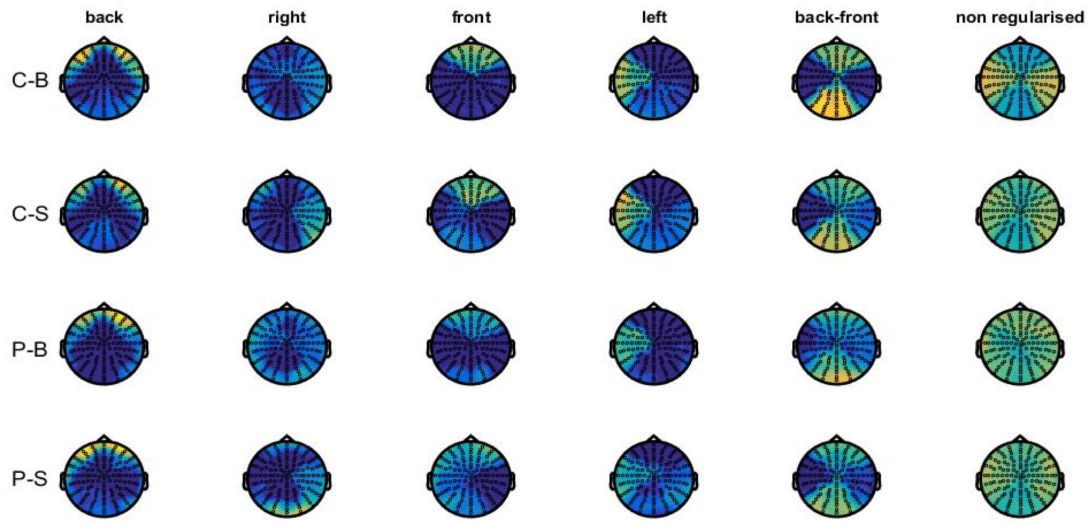

Fig. 3. Group averages of the obtained components of the spatial mode of the 4-mode tensor. The components were regularised separately with graphs that penalised differences between the labeled brain regions and the rest and enhanced the similarity within the labeled brain region. For example, the top left component was regularised to a graph that penalised big differences between the back and the rest of the brain while forcing the elements of that region to be similar to each other. *C* denotes controls, *P* patients while *B* denotes the binding task and *S* the shape task. The colorbar ranges are the same for all figures.

4.2. Custom designed graphs based on electrophysiological a-priori

An important application of graph regularised tensor factorisation, apart from error reduction, can be the driving of the factorisation procedure by a-priori information. Similarly, assumptions about the data can be tested by providing graphs that describe the expected interrelations. In our Altzheimer's dataset the two tasks are expected to produce different brain responses in different brain regions. More specifically, it is expected that there is greater strength in connection between the back and front of the brain [25]. We created 5 custom graphs, the first four concentrating into producing components that are focused on a single region (back-right-front-left) and the fifth one that aims to extract a component that indicates a network between the frontal and rear areas of the brain. The 6th component was not regularised. The results of this experiment are shown in Figure 3.

5. CONCLUSIONS

We developed an algorithm that performs graph regularised tensor factorisation based on brain connectivity graphs. We provided an extension to graph regularisation by allowing different graphs to regularise the various components of a tensor's mode. In general, our methodology can be incorporated in any dataset where we expect or desire that the different components of a mode to exhibit a specific structure.

We based our results on an Altzheimers EEG dataset where we obtained the connectivity graphs either from the data itself or by designing connectivity graphs that test prior clinical assumptions. In the first case, the algorithm resulted in reducing the noise inherent to any EEG dataset in a train-test split of the data. In the second method we designed graphs that express the a-priori connectivity assumptions of the data. The algorithm was able to pick up differences between the two different tasks involved in the dataset. The clinical usefulness of such a method is shown due to the fact that prior knowledge in the design of the graphs led to finding the expected differences between the groups and tasks. Namely, that the shape-colour binding task exhibits larger power in the frontal and back regions of the brain than the shape task.

Further applications of the algorithm can be found in areas that information regarding the relations between the data are known or assumed. Examples can range from functional magnetic resonance imaging studies where networks are more robust and localised. Also, studies in high dimensional social and biological networks can benefit from such a method by regularising the extracted data by known networks.

6. ACKNOWLEDGMENTS

We wish to thank Mario Parra and Agustín Ibáñez for the provision of the dataset [25].

7. REFERENCES

[1] P. Comon, "Tensor decompositions, state of the art and applications," *arXiv preprint arXiv:0905.0454*, 2009.

[2] R. Bro, "PARAFAC. Tutorial and applications," in *Chemometrics and Intelligent Laboratory Systems*, 1997, vol. 38, pp. 149–171.

[3] P. M. Kroonenberg and J. De Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.

[4] A. Cichocki, "Tensor decompositions: a new concept in brain data analysis?," *arXiv preprint arXiv:1305.0395*, 2013.

[5] L. Spyrou, S. Kouchaki, and S. Sanei, "Multiview classification of brain data through tensor factorisation," in *Proceeding of IEEE Machine Learning for Signal Processing (MLSP)*. IEEE, 2015, pp. 1–6.

[6] L. Spyrou, S. Kouchaki, and S. Sanei, "Multiview classification and dimensionality reduction of scalp and intracranial eeg data through tensor factorisation," *Journal of Signal Processing Systems*, pp. 1–12, 2016.

[7] A. G. Mahyari and S. Aviyente, "Identification of dynamic functional brain network states through tensor decomposition," *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 2099–2103, 2014.

[8] B. Pester, C. Ligges, L. Leistritz, H. Witte, and K. Schiecke, "Advanced Insights into Functional Brain Connectivity by Combining Tensor Decomposition and Partial Directed Coherence," *PLoS One*, vol. 10, 2015.

[9] B. Bollobas, *Graph theory: an introductory course*, Springer Science & Business Media, 2012.

[10] N. Deo, *Graph theory with applications to engineering and computer science*, Courier Dover Publications, 2016.

[11] A.-L. Barabási, "Network science: Luck or reason," *Nature*, vol. 489, no. 7417, pp. 507–508, 2012.

[12] N. M Tichy, M. L Tushman, and C. Fombrun, "Social network analysis for organizations," *Academy of management review*, vol. 4, no. 4, pp. 507–519, 1979.

[13] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[14] B. Horwitz, "The elusive concept of brain connectivity," *NeuroImage*, vol. 19, no. 2, pp. 466–470, 2003.

[15] A. A. Smola and R. Kondor, "Kernels and Regularization on Graphs," *Machine Learning*, vol. 2777, pp. 1–15, 2003.

[16] J. J. Y. Wang, H. Bensmail, and X. Gao, "Multiple graph regularized nonnegative matrix factorization," *Pattern Recognition*, vol. 46, no. 10, pp. 2840–2847, 2013.

[17] D. Cai, X. He, and J. Han, "Graph Regularized Nonnegative Matrix Factorization for Data," *Pami*, vol. 33, no. 8, pp. 1–17, 2008.

[18] A. Narita, K. Hayashi, R. Tomioka, and H. Kashima, *Tensor Factorization Using Auxiliary Information*, pp. 501–516, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[19] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, vol. 6, no. 1-4, pp. 164–189, 1927.

[20] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[21] R. A. Harshman, "Foundations of the parafac procedure: models and conditions for an explanatory multimodal factor analysis," 1970.

[22] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2008.

[23] A. Bouhamidi and K. Jbilou, "A note on the numerical approximate solutions for generalized Sylvester matrix equations with applications," *Applied Mathematics and Computation*, vol. 206, no. 2, pp. 687–694, 2008.

[24] C. A. Andersson and R. Bro, "The N-way Toolbox for MATLAB," *Chemometrics and Intelligent Laboratory Systems*, vol. 52, no. 1, pp. 1–4, 2000.

[25] M. Pietto, M. A. Parra, Trujillo N., Flores F., Garca A.M., Bustin J., Richly P., Manes F., Lopera F. Ibez A., and Baez S., "Behavioral and Electrophysiological Correlates of Memory Binding Deficits in Patients at Different Risk Levels for Alzheimers Disease," *Journal of Alzheimer's Disease*, vol. 53, pp. 1325, 2016.

[26] Olaf Sporns and Richard F. Betzel, "Modular Brain Networks," *Annual Review Psychology*, vol. 67, pp. 613–640, 2016.