

Multimodal approach to estimate the ocular movements during EEG recordings: a coupled tensor factorization method*

Bertrand Rivet¹, Marc Duda¹, Anne Guérin-Dugué¹, Christian Jutten^{1,2} and Pierre Comon¹

Abstract— This paper deals with coupled tensor factorization. A relaxed criterion derived from the advanced coupled matrix-tensor factorization (ACMTF) proposed by Acar *et al.* is described. The proposed relaxed ACMTF (RACMTF) criterion is based on weaker assumptions that are thus more often satisfied when dealing with actual data. Numerical simulations show the benefit of using jointly two data sets when the underlying factors are highly correlated, especially if one of the modality is less noisy than the other one. The proposed method is finally applied on actual Gaze&EEG data to estimate the ocular artifacts into the EEG recordings.

I. INTRODUCTION

Electroencephalography (EEG) is one of the most popular way to explore the brain activity [1] because it is a non-invasive technique. However, the recordings contain ocular artifacts due to the blinks or the eye movements. A classical way to overcome this problem is to ask the user to not move his/her eyes during the experiments by proposing a paradigm which avoids the eye saccades. This leads to very restrictive experiments far from ecological situations. Indeed, this constrain excludes many interesting areas of cognitive sciences related to the visual process such as text reading or visual scene explorations for which the users have to look freely on a screen to perform a given task. Consequently, the EEG recordings can be strongly distorted and the ocular artifacts must be removed from the EEG recordings before their analysis.

In order to attenuate electro-ocular (EOG) artifacts, many algorithms have been proposed in the literature (e.g., [2]). Among which, one can cite independent component analysis (ICA) [3], [4] and regression technique [5], [6]. ICA based methods decompose the EEG signals using spatial filters into components which maximizes an independence criterion and then remove the identified components related to EOG artifacts. Even if ICA is shown to be efficient to remove EOG artifacts, it has two main drawbacks: it needs a large number of sensors and the removed components may still contain some brain activities, leading thus to possibly misinterpretations of the brain waves. On the other hand, regression based methods require additional EOG channels as artifact references and they identify scaling factors between the reference channels and the EEG ones to estimate the ocular artifacts. This alternative approach needs less EEG sensors

than the ICA based ones, but the assumption of equal shape up to a scaling factor between the EOG and the EEG sensors may not be satisfied since EOG recordings also contain some brain activities.

In this paper, we propose to tackle the estimation of EOG artifacts on EEG recordings by a tensor approach using multimodal data which are synchronously recorded. The EEG signals are jointly used with some reference signals on the eye movements recorded by an eye-tracker (ET). The proposed approach lies thus within the joint or coupled tensors factorization framework (e.g., [7], [8], [9], [10]). Compared to [8], [10], which enforce the shared loadings to be the same between modalities, we propose to relax these assumptions by only assuming that some loadings between modalities are linked (i.e. similar or with similar properties) and not necessarily the same. Indeed, the equality assumption is somewhat too restrictive when dealing with real data, especially with brain signals.

The remainder of this paper is organized as follows. In Section II, the proposed method is described. Some numerical experiments and an illustration on real eye-gaze & EEG data are shown in Section III. Finally, conclusions and perspectives are in Section IV.

II. MULTIMODAL TENSOR FACTORIZATION

The advanced coupled matrix-tensor factorization (ACMTF) [10] is recalled in Subsection II-A and the proposed relaxed model is described in Subsection II-B.

We adopt some standardized notations: $\llbracket \boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ denotes a tensor defined by the sum of rank-one tensors: $\llbracket \boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket = \sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$, where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_R]'$, with \cdot' the transposition operator, \circ the tensor product and \mathbf{x}_r the r th column of matrix \mathbf{X} .

A. Coupled tensors factorization

Let $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$ denote two third-order tensors to be jointly factorized. Let's say, without loss of generality, that these two tensors share some components in the second and third factors (\mathbf{B} and \mathbf{C}) while the first one \mathbf{A} is unconstrained. The ACMTF model [10] can then be written as the following optimization problem

$$\min_{\theta} \sum_{i=1}^2 \left\| \mathcal{X}^{(i)} - \llbracket \boldsymbol{\lambda}^{(i)}; \mathbf{A}^{(i)}, \mathbf{B}, \mathbf{C} \rrbracket \right\|_F^2 + \alpha \left\| \boldsymbol{\lambda}^{(i)} \right\|_1 \quad (1)$$

$$\text{s.t. } \forall i, \forall r, \left\| \mathbf{a}_r^{(i)} \right\|_2 = 1, \left\| \mathbf{b}_r \right\|_2 = 1, \left\| \mathbf{c}_r \right\|_2 = 1$$

where $\|\cdot\|_F$, $\|\cdot\|_2$ and $\|\cdot\|_1$ are the Frobenius, ℓ_2 and ℓ_1 norms, respectively, $\alpha \geq 0$ is a penalty parameter and

*This work was funded by ERC-AdG-2012-320864 CHES project and ERC-AdG-2013-320594 DECODA project.

¹ Authors are with Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble France. CNRS, GIPSA-Lab, F-38000 Grenoble France. firstname.lastname@gipsa-lab.grenoble-inp.fr

² Christian Jutten is with Institut Universitaire de France (IUF).

$\theta = \{\{\boldsymbol{\lambda}^{(i)}, \mathbf{A}^{(i)}\}_i, \mathbf{B}, \mathbf{C}\}$ is the set of parameters to be optimized. In this formulation the ℓ_1 norm terms are used to sparsify the weights $\boldsymbol{\lambda}^{(i)}$ so that the unshared components will have weights equal to 0 in one of the data sets while the shared components will have values different from 0 in both data sets. Consequently, the shared components \mathbf{b}_r and \mathbf{c}_r are assumed to be strictly equal in both data sets. We observed on Gaze&EEG data that this assumption leads to overestimate the number of unshared components: some related components are estimated as unshared (different) ones while two components \mathbf{b}_r (and/or \mathbf{c}_r) are closed. Indeed, the relationship between the shared components can be more general than a strict equality: the shared components can only be similar (in some sense) or with similar properties (e.g., variations with the same order of magnitude).

B. Relaxed coupled tensors factorization

To overcome the equality constraint of shared components highlighted in the previous section, we modify the ACMTF model by relaxing criterion (1) such that the proposed optimization problem (RACMTF) is now

$$\begin{aligned} \min_{\theta} & \sum_{i=1}^2 \|\mathcal{X}^{(i)} - \llbracket \boldsymbol{\lambda}^{(i)}; \mathbf{A}^{(i)}, \mathbf{B}^{(i)}, \mathbf{C}^{(i)} \rrbracket\|_F^2 + \alpha \|\boldsymbol{\lambda}^{(i)}\|_1 \\ & + \frac{1}{2} \beta \sum_{r=1}^R \left\| |\mathbf{Q} \mathbf{b}_r^{(1)}| - |\mathbf{Q} \mathbf{b}_r^{(2)}| \right\|_2^2 \\ & + \gamma \sum_{r=1}^R \|\mathbf{c}_r^{(1)} - \mathbf{c}_r^{(2)}\|_1 \\ & - \delta \|\boldsymbol{\lambda}^{(1)} * \boldsymbol{\lambda}^{(2)}\|_{\tilde{0}} \\ \text{s.t. } & \forall i, \forall r, \|\mathbf{a}_r^{(i)}\|_2 = 1, \|\mathbf{b}_r^{(i)}\|_2 = 1, \|\mathbf{c}_r^{(i)}\|_2 = 1 \end{aligned} \quad (2)$$

where $\|\mathbf{x}\| = [|x_1|, \dots, |x_I|]'$ stands for the element wise absolute value, $*$ is the Hadamard product, $\alpha, \beta, \gamma, \delta$ are four non-negative penalty parameters and $\|\mathbf{x}\|_{\tilde{0}} = \sum_i (1 - \exp(-x_i^2/\sigma^2))$ is the smoothed ℓ_0 norm [11], with σ a small enough parameter to approximate the ℓ_0 norm, and \mathbf{Q} is a fixed matrix.

The last term in (2) is used to enforce the identification of the shared components to avoid that similar factors are estimated with different components by maximizing the number of common non-zero elements in the weights $\boldsymbol{\lambda}^{(i)}$; it contributes to restore uniqueness of both decompositions, in the case they were not unique if calculated separately.

For the sake of concise notation, the two additional penalty terms (third and fourth ones) are presented in (2) on factors \mathbf{B} and \mathbf{C} respectively, but they obviously may concern any of the three factors \mathbf{A}, \mathbf{B} or \mathbf{C} . The ℓ_1 norm term on $\mathbf{c}_r^{(i)}$ allows the shared components $\mathbf{c}_r^{(1)}$ and $\mathbf{c}_r^{(2)}$ to have few entries that are different. Finally, the ℓ_2 norm term on $\mathbf{b}_r^{(i)}$ allows the shared components $\mathbf{b}_r^{(1)}$ and $\mathbf{b}_r^{(2)}$ to have some similar properties. For instance, if \mathbf{Q} is the first or the second order (numerical) derivative matrix, this term penalizes the shared components if their derivate values are different in order of magnitude whatever are their signs using the absolute value. Thus, with such \mathbf{Q} matrices, the main idea is that the shared

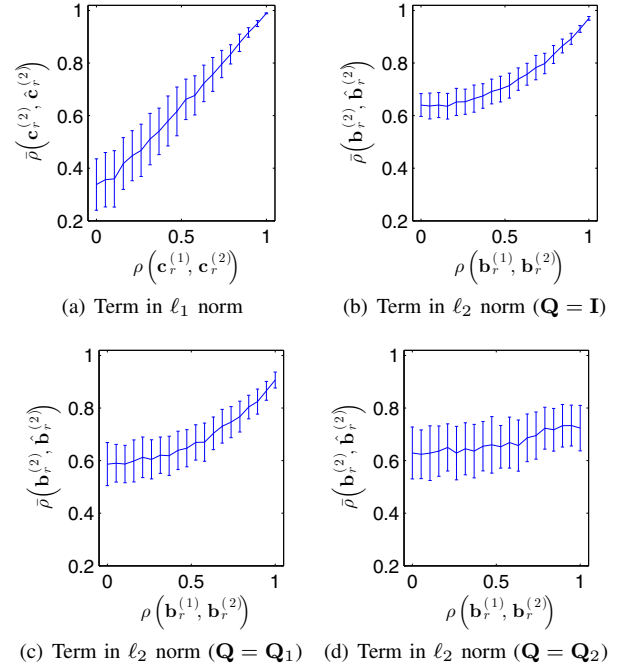


Fig. 1. Influence of penalization terms in the proposed RACMTF (2) criterion on the accuracy of the factorization. The mean plus/minus the standard deviation of average correlation coefficient $\bar{\rho}$ (3) are plotted vs. the correlation coefficient between the two data sets.

components should have the same kind of variations, but for instance one factor in one modality can increase while the other factor in the other modality is decreasing. Finally, one can also express the similarity of the factor using a ℓ_1 norm, instead of the ℓ_2 norm. It is worth noting that such kind of penalizations can also be rewritten in a Bayesian way [12].

To optimize the criterion (2), we used a conjugated gradient descent [13] after converting the constraints to quadratic penalties and substituting the ℓ_1 norm with a differentiable approximation ($\|\mathbf{x}\|_1 \simeq \|\mathbf{x}\|_{\tilde{1}} = \sum_i \sqrt{x_i^2 + \epsilon}$, for small enough ϵ) as done in [10] and the same trick is applied for the absolute value ($|x_i| \simeq \sqrt{x_i^2 + \epsilon}$).

III. EXPERIMENTS

To assess the efficiency of the proposed relaxed ACMTF model (2), first numerical simulations are performed (Section III-A) and then it is applied on actual Gaze&EEG data (Section III-B).

A. Numerical simulations

In each case, we consider the influence of a single relaxation term in the RACMTF criterion (2), i.e. only β or γ is non zero. In other words, we have generated data such that only the second factors $\mathbf{B}^{(i)}$ or the third ones $\mathbf{C}^{(i)}$ are shared, respectively, while the two other ones are unconstrained. The number of components R is fixed and equal to 2. The unconstrained factors are randomly generated using a standard normal distribution. The shared factors are drawn independently from a standard normal distribution for each r , such that the correlation coefficient $\rho(\mathbf{x}_r^{(1)}, \mathbf{x}_r^{(2)}) =$

$\tilde{\mathbf{x}}_r^{(1)'} \tilde{\mathbf{x}}_r^{(2)}$ (with $\tilde{\mathbf{x}}_r^{(1)}$ the factor $\mathbf{x}_r^{(1)}$ divided by its ℓ_2 norm) between component factors $\mathbf{x}_r^{(i)}$ of $\mathbf{X}^{(i)}$ is fixed. All weights $\lambda^{(i)}$ are equal to one. Finally tensors $\mathcal{X}^{(i)}$ are obtained by $\mathcal{X}^{(i)} = \llbracket \lambda^{(i)}; \mathbf{A}^{(i)}, \mathbf{B}^{(i)}, \mathbf{C}^{(i)} \rrbracket + \mathcal{N}^{(i)}$, where entries of the noise tensors $\mathcal{N}^{(i)}$ are drawn from a zero mean normal distribution with standard deviation $\sigma^{(i)}$, which is different for each tensor ($\sigma^{(1)} = .01$ and $\sigma^{(2)} = .1$, i.e. 1% and 10% of noise, respectively). This means that the second data set is more noisy than the first one which often happened when one uses several modalities. For each numerical experiments, 100 configurations are randomly drawn and the results are plotted in Fig. 1. The accuracy of the estimated factors $\mathcal{X}^{(2)}$ (i.e. of the more noisy tensor) is assessed by the average correlation coefficients $\bar{\rho}(\hat{\mathbf{x}}_r^{(2)}, \mathbf{x}_r^{(2)})$ between the estimated factors $\hat{\mathbf{x}}_r^{(2)}$ and the actual ones $\mathbf{x}_r^{(2)}$

$$\bar{\rho}(\hat{\mathbf{x}}_r^{(2)}, \mathbf{x}_r^{(2)}) = \frac{1}{R} \sum_{r=1}^R \frac{\hat{\mathbf{x}}_r^{(2)'} \mathbf{x}_r^{(2)}}{\|\hat{\mathbf{x}}_r^{(2)}\|_2 \|\mathbf{x}_r^{(2)}\|_2}. \quad (3)$$

The closer to one, the better the results.

In the first experiment (Fig. 1(a)), the influence of the ℓ_1 norm term is investigated (i.e. $\beta = 0$). As one can see, the more correlated the two data sets are, the better the factorization is. Without coupled factorizations, the mean accuracy of the factorization $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$ are equal to .99 and .25, respectively. And the accuracy of the coupled factorization of the less noisy tensor $\mathcal{X}^{(1)}$ ($\bar{\rho} = .99$) is almost independent of the correlation between the two tensors. Finally, with highly correlated factors $\mathbf{c}_r^{(1)}$ and $\mathbf{c}_r^{(2)}$, the accuracy of the factorization of the more noisy tensor $\mathcal{X}^{(2)}$ is close to the accuracy of the less noisy one $\mathcal{X}^{(1)}$ showing the advantage of using coupled factorization compared to individual ones.

In the second experiment (Fig. 1(b), 1(c) and 1(d)), the influence of the ℓ_2 norm term is investigated (i.e. $\gamma = 0$) for several matrices \mathbf{Q} , the identity matrix \mathbf{I} , the first order derivative one \mathbf{Q}_1 and the second order derivative one \mathbf{Q}_2 , respectively. As in the previous experiment, one can see the advantage of considering coupled factorization instead of individual ones when dealing with correlated factors: for all three cases, the higher the correlation, the better the accuracy. Although, one can see that the more explicit the coupling represents the signals by themselves, the higher and hence the better the accuracy. Indeed, for a given correlation of the shared factors, the accuracy increases when using the first (Fig. 1(c)) instead of the second derivatives (Fig. 1(d)) and the increase becomes even more pronounced when the signals themselves are coupled (Fig. 1(b)) instead of their first derivatives (Fig. 1(c)).

B. Illustration on actual data

The Gaze&EEG data comes from experiment in visual search where participants had to search a target from a set of distractors [14]. Sixty four active electrodes (BrainProducts GmbH) were mounted on an EEG cap (BrainCapTM) placed on the scalp in compliance with the International 10-20 system. To be compatible with the EEG acquisition, eye

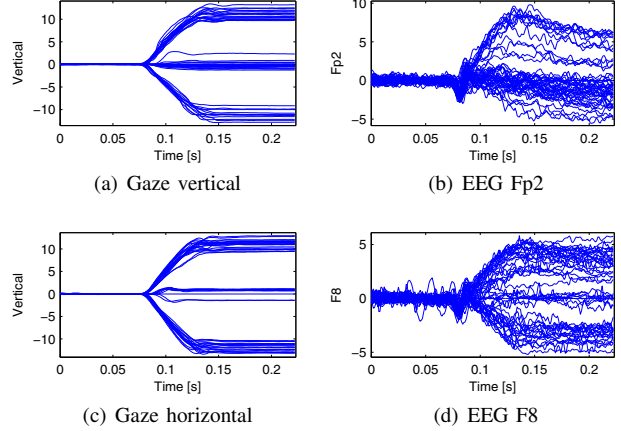


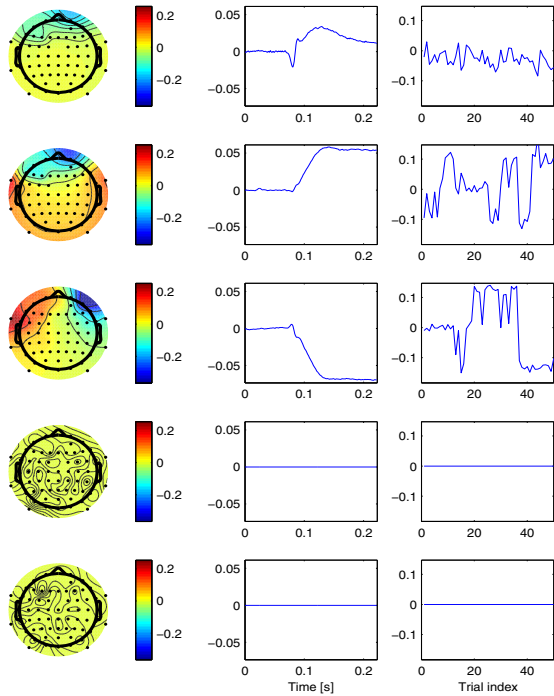
Fig. 2. Gaze (left plots) and EEG (right plots) signals, all trials are stacked on the same plot.

movements were recorded by a remote binocular infrared eye tracker EyeLink 1000 (SR Research) to track the gaze of each eye while the observer was looking at stimuli. The EyeLink system was used in the Pupil-Corneal Reflection tracking mode. For the both acquisition devices, the sampling frequency was 1000Hz. Off-line, EEG signals and gaze samples were synchronized using hardware triggers signals sent in parallel to the EEG recorder and the eye tracker, along the experiment.

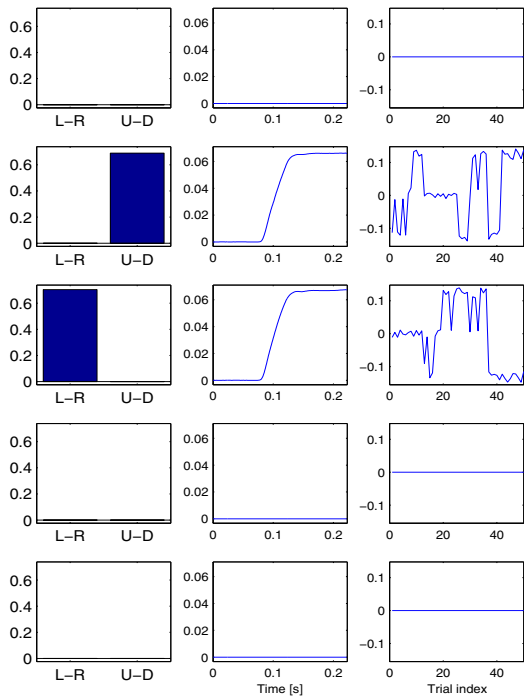
From these data, two tensors (space \times time \times trials) are build synchronized on the beginning of the saccades: each tensor is the concatenation in the third dimension of the matrices which each column is the vector of channel signals at a given time. $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$ are related to the EEG signals and the eye-Gaze signals (Fig. 2), respectively. In addition to the proposed RACMTF criterion (2), we add an extra ℓ_1 norm penalty term on the spatial factors of Gaze signals, $\mathbf{a}_r^{(2)}$, to enforce the identification of horizontal and vertical eye movements separately. For this experiment, the matrix \mathbf{Q} was equal to \mathbf{Q}_1 , so that the first derivative of the temporal shapes of EEG and Gaze signals should be correlated. The result of the coupled factorization is shown in Fig. 3. We have arbitrarily fixed $R = 5$. As one can see three (resp. two) significant EEG (resp. Gaze) factors have been identified, which two of them are coupled with the Gaze ones. And two of the factors (the last two ones) are unnecessary for the factorization. The second factor is clearly related to vertical eye movements as shown by the first order factor $\hat{\mathbf{a}}_2^{(2)}$ and confirmed by the spatial map of $\hat{\mathbf{a}}_2^{(1)}$. On the other hand, the same analysis of the third spatial factors $\hat{\mathbf{a}}_3^{(1)}$ and $\hat{\mathbf{a}}_3^{(2)}$ allows us to conclude that these factors are related to the horizontal eye saccades.

IV. CONCLUSIONS AND PERSPECTIVES

In this article, we proposed a relaxed coupled tensor factorization criterion called RACMTF based on the previous ACMTF criterion proposed by Acar *et al.* [10]. The constraint of equality of the shared factors is replaced by



(a) EEG factors



(b) Gaze factors

Fig. 3. Illustration of the coupled factorization of EEG tensor (Fig. 3(a)) and Gaze one (Fig. 3(b)). Each estimated factor is multiplied by the related $\hat{\lambda}_r^{(i)}$ to show its significance into the tensor. Each column is related to a mode of the tensor ($\mathbf{A}^{(i)}$, $\mathbf{B}^{(i)}$ and $\mathbf{C}^{(i)}$ from left to right) and each row is a factor r . In the first mode of the Gaze tensor, the two spatial dimensions are Left-Right (L-R) and Up-Down (U-D) channels.

a weaker assumption: the shared factors only need to be similar in some sense. Two kinds of similarity have been proposed, either on the factors themselves or either on their first or second order derivatives. The coupled factorization has been shown to improve the accuracy of the estimated factors compared to the one achieved by individual factorization. Numerical experiments shown that the accuracy of the factorization is mainly due to the less noisy input tensor when dealing with highly correlated factors. On actual Gaze&EEG data, the proposed method is shown to recover the horizontal and vertical eye movement, even if the final estimation is sensitive to the initialization of the descent algorithm. Future works will consider other extensions such as synchronization issue (i.e. delays between trials) or time warping.

ACKNOWLEDGEMENTS

The authors want to thank Emmanuelle Kristensen for her help to record the GAZE and EEG data.

REFERENCES

- [1] E. Niedermeyer and F. H. Lopes da Silva, *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
- [2] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: a review," *Neurophysiol Clin*, vol. 30, no. 1, pp. 5–19, 2000.
- [3] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 395 – 404, 1997.
- [4] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation Independent Component Analysis and Applications*. Academic Press, 2010.
- [5] G. Gratton, M. G. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalography and Clinical Neurophysiology*, vol. 55, pp. 468–484, 1983.
- [6] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology*, vol. 118, pp. 1388–2457, 2007.
- [7] A. Banerjee, S. Basu, and S. Merugu, "Multi-way clustering on relation graphs," in *Proc. SIAM Int. Conf. Data Mining*, 2007, pp. 145–156.
- [8] E. Acar, T. Kolda, and D. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," in *Proc. Workshop on Mining and Learning with Graphs*, San Diego, (USA), August 2011.
- [9] H. Becker, P. Comon, and L. Albera, "Tensor-based preprocessing of combined EEG/MEG data," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 275–279.
- [10] E. Acar, E. Papalexakis, G. Gurdeniz, M. Rasmussen, A. Lawaetz, M. Nilsson, and R. Bro, "Structure-revealing data fusion," *BMC Bioinformatics*, vol. 15, no. 1, p. 239, 2014.
- [11] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, January 2009.
- [12] R. Cabral Farias, J. Cohen, C. Jutten, and P. Comon, "Joint decompositions with flexible couplings," in *Proc. Int. Conf. Independent Component Analysis and Blind Source Separation (LVA/ICA)*, 2015, (Submitted to).
- [13] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
- [14] E. Kristensen, A. Guérin-Dugué, and B. Rivet, "Comparison between Adjar and xDAWN algorithms to estimate eye-fixation related potentials distorted by overlapping," in *Proc. International IEEE EMBS Conference on Neural Engineering (NER)*, Montpellier, France, April 2015, pp. 976–979.