Tensor-based Blind fMRI Source Separation Without the Gaussian Noise Assumption — A β-Divergence Approach

Christos Chatzichristos[1,*], Michiel Vandecapelle[1,2,*], Eleftherios Kofidis[3], Sergios Theodoridis[4,5], Lieven De Lathauwer[1,2] and Sabine Van Huffel[1]
[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS, Leuven, Belgium
christos.chatzichristos@esat.kuleuven.be, * Equal contribution
[2] KU Leuven Kulak, Group Science, Engineering and Technology, Kortrijk, Belgium
[3] Dept. of Statistics and Insurance Science, University of Piraeus, Greece
[4] Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece
[5] Chinese University of Hong Kong, Shenzhen, China

***Abstract*—The advantages of tensor- over matrix-based methods have been recently demonstrated in the context of functional magnetic resonance imaging (fMRI) blind source unmixing. However, these methods rely on the assumption of a Gaussian distribution for the noise, which suggests a least squares criterion for the tensor decomposition. One can instead argue that a Rician model for the fMRI noise is much more accurate and hence alternative cost functions should also be investigated. In this paper, β-divergences are used to parametrize the Canonical Polyadic Decomposition (CPD) fitting to fMRI data and the effect of β on the source separation performance is evaluated, for different values of signal to noise ratio (SNR). Our results confirm that the commonly used squared error is not the best choice, particularly at low SNRs.**

I. Introduction

Functional Magnetic Resonance Imaging (fMRI) is a noninvasive neuro-imaging technique, which indirectly studies brain activity, by measuring fluctuations of the Blood Oxygenation Level Dependent (BOLD) signal [1]. The fMRI concept builds upon the Magnetic Resonance Imaging (MRI) technology and the properties of oxygen-rich blood. The signal detected from the MR scanner is the sum of the signals from dipoles (nuclei) all over the region under examination. In order to form an image, the dipoles must somehow be spatially resolved. To this end, MRI uses strong magnetic fields (varying in phase and frequency) in order to define the different voxels-nuclei from where the signal is transmitted [2].

The signal is measured through a quadrature detector that produces real and imaginary signals (based on the different phases and frequencies of the transmitted magnetic gradients). The collected raw complex-valued data, which lie in the frequency domain, form the so-called k-space. Each point in that space contains spatial frequency and phase information about all the voxels of the image. The real and imaginary parts of the k-space image are Gaussian distributed and the noise that corrupts the data is also Gaussian [3].

The image space can thus be acquired through an Inverse Fourier Transform (IFT) of the k-space and since the IFT is a linear transformation, the noise distribution in the image space is also Gaussian. Furthermore, the variance of the noise will be uniform over the whole field of view and, due to the IFT, the noises in the real and imaginary voxels can be assumed uncorrelated [4].

MR scanners typically provide magnitude images, mainly due to memory limitations and also for avoiding the problem of phase artifacts (by deliberately discarding the phase information) [5], [6]. By only taking the magnitude of the complex data, the noise on the obtained data follows a Rician instead of a Gaussian distribution [4].

The localization of the activated brain areas from the fMRI data is a challenging Blind Source Separation (BSS) problem [7], in which the sources consist of a combination of spatial maps (areas activated) and time courses (timings of activation). fMRI data involve multiple modes, such as trial, session and subject, in addition to the intrinsic modes of time and space [8]. Up to recently, multivariate bi-linear (i.e., matrix-based) methods, based on the concatenation of different modes, have been the-state-of-the-art in fMRI BSS [9], [10]. However, by definition, such methods fall short in exploiting the inherently multi-way nature of fMRI data.

In contrast, this is preserved in multi-linear (tensor) models, which, in general, a) produce unique (modulo scaling and permutation ambiguities) representations under mild conditions [11], b) can improve the ability of extracting spatiotemporal modes of interest [8], [12], [13], and c) facilitate neurophysiologically meaningful interpretations [8]. The state-of-the-art in tensorial methods for analyzing multi-subject fMRI data include (among others [14], [15], [16]) the Canonical Polyadic Decomposition (CPD)-based analysis [8], which views the fMRI data as a 3rd-order tensor, with modes space $\times$ time $\times$ subjects.

However, in all of those tensor-based blind fMRI source unmixing methods, a Gaussian approximation of the noise is used, which suggests the adoption of the least squares (LS) criterion for the estimation of the tensor decomposition. In

this paper, we model the Rician nature of the fMRI noise via β-divergence cost functions and investigate whether, for certain types of data and specific squared noise distributions, these can lead to more meaningful low-rank approximations than those obtained with the LS cost function. This is done by simulating fMRI data, adding noise with different Signal-to-Noise Ratios (SNRs) and decomposing the resulting tensor with a β-divergence CPD for different values of β. It is shown that, indeed, the LS criterion can be outperformed by choosing a suitable β-divergence cost function. For overlapping sources, higher values of β work well, while for sources that have different noise variances in their real and imaginary parts, lower β values offer the best results.

A. Notation

Vectors, matrices and higher-order tensors are denoted by bold lower-case, upper-case and calligraphic upper-case letters, respectively. The outer product of two or more vectors is denoted by $\circ$.

II. Methods

The magnitude images are formed by calculating the magnitude, pixel by pixel, from the real and the imaginary images. This is a nonlinear mapping and therefore the noise distribution is no longer Gaussian. Assuming that $A(x)$ is the image pixel density of the MR image in the absence of noise (with $A_r(x)$ and $A_i(x)$ its real and imaginary parts, respectively) at voxel location x, the magnitude signal is:

$$M(x) = \sqrt{(A_r(x) + N_r(x))^2 + (A_i(x) + N_i(x))^2},$$

with N_r, $N_i \sim \mathcal{N}(0, \sigma^2)$, where σ denotes the standard deviation of the Gaussian noise in the real and the imaginary images. Hence, in the presence of noise, the probability density function of the magnitude image is given by [6], [4]:

$$p(M) = \frac{M}{\sigma^2} e^{-(M^2+A^2)/2\sigma^2} I_0\left(\frac{AM}{\sigma^2}\right),$$

with M and A being the mean magnitude signal intensity and the mean measured signal intensity, respectively, and $I_0(\cdot)$ is the modified zeroth–order Bessel function of the first kind. This is known as the Rice density and for low SNRs it is quite different from the Gaussian distribution.

While the stationarity of the Rician noise (in the sense of having the same variance in its real and imaginary parts) has been the keystone of the statistical signal processing in MRI for years, it cannot be applied when parallel imaging reconstruction is considered [17]. Parallel imaging methods are used to increase the acquisition rate via subsampled acquisitions of the k-space. In those methods, the known placement and sensitivities of the receiver coils are used to assist spatial localization of the MR signal, and hence a reduction in the number of phase-encoding steps during image acquisition is achieved. As a result, the main assumption of a single value of σ in order to characterize the whole data set is no longer valid. When parallel imaging techniques are used, due to the reconstruction process and the use of fewer phase gradient steps, the variance of noise becomes space-dependent, and with a different value of σ for real and imaginary parts [17].

III. CPD with β-divergence

CPD (or PARAFAC) [18], [8], [19], approximates the 3rd-order tensor of fMRI data, $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_{xyz} \times I_t \times I_s}$, by a sum of R (= estimated number of sources) rank-1 tensors, namely

$$\boldsymbol{\mathcal{T}} \approx \sum_{r=1}^{R} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r.$$

Letting $\boldsymbol{A} = \left[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_R\right]$ be the matrix that contains the R spatial components (I_{xyz} voxels per component) and defining $\boldsymbol{B} = \left[\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_R\right]$ and $\boldsymbol{C} = \left[\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_R\right]$ similarly, corresponding to the associated time courses (I_t time points) and the subject activation levels (I_s subjects), respectively, the CPD of $\boldsymbol{\mathcal{T}}$, denoted by $[\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$, allows us to recover the space-time sources in a completely blind manner.

Its main advantage, besides its simplicity, is the fact that it is unique (up to permutation and scaling) under mild conditions [20], [21]. Uniqueness of CPD is crucial to its application in fMRI. In fact, it was demonstrated [12], [13], [22] that CPD with fMRI data is robust to overlaps (spatial and/or temporal) of the sources. On the other hand, the result of CPD is largely dependent on the correct estimation of the tensor rank, R [23], [24].

Numerous algebraic and optimization-based algorithms have been designed to estimate the CPD of a tensor [19]. Most of these methods fit a CPD to the tensor in a LS sense, i.e. minimizing the Frobenius norm of the residual tensor $\boldsymbol{\mathcal{T}} - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$. Assuming that the tensor entries are perturbed by independent and identically distributed (i.i.d.) Gaussian noise, this gives the maximum likelihood estimate. However, magnitude images are instead expected to have a Rician noise distribution, as explained above. As a result, fitting a CPD with a different criterion may lead to better results for this type of data and investigating this possibility is the goal of this paper.

For nonnegative data, such as magnitude images, a possible alternative to the squared error is the class of β-divergence cost functions, a special class of Bregman divergences [25], defined for $\beta \in \mathbb{R}$. This class of cost functions encompasses not only the squared error itself ($\beta = 2$), but also the Itakura—Saito (IS) divergence ($\beta = 0$) and the Kullback—Leibler (KL) divergence ($\beta = 1$) and interpolates between them continuously. An important property of β-divergence cost functions is that values of $\beta < 2$ penalize errors in small entries more heavily compared to the squared distance. On the other hand, for values of $\beta > 2$, errors in large entries are penalized more heavily. β-divergences can therefore be a good choice when the tensor has entries of different magnitudes and one is primarily interested in fitting either the smaller or larger entries better compared to the LS approach. These cost functions have been applied successfully to the decomposition of tensors containing count data or spectral data and gave more meaningful low-rank approximations than those obtained with

the squared cost [26]. For example, for fMRI data, the KL divergence has been demonstrated to outperform the squared distance as a criterion for different model choices [27], [28], [29].

For each tensor entry x, the β-divergence $d_\beta(x, y)$ between x and its estimate y from the CPD is defined as follows:

$$d_\beta(x,y) = \begin{cases} \frac{x^\beta + (\beta-1)y^\beta - \beta(xy^{(\beta-1)})}{\beta(\beta-1)} & \beta \in \mathbb{R}\backslash\{0,1\} \\ x\ln(\frac{x}{y}) - x + y & \beta = 1 \\ \frac{x}{y} - \ln(\frac{x}{y}) - 1 & \beta = 0 \end{cases}$$

Note that the squared distance depends only on the absolute difference between x and y, while β-divergences also consider the values x and y themselves. A number of methods have been proposed to compute a nonnegative CPD of a tensor with a β-divergence cost function [30], [31], [32]. In this paper, we will use the algorithm from [33], which uses second-order information to compute a nonnegative CPD of a tensor for general cost functions and thus also for the β-divergence. Following a generalized Gauss–Newton approach, the algorithm uses curvature information from the Hessian matrix of the cost function and thus offers fast local convergence, while it exploits any available structure in the tensor.

IV. Simulations

In order to test the performance of the separation algorithm for Rician noise, complex-valued data were generated, as the Rician noise is obtained from the modulus of complex data that are perturbed by Gaussian noise. A simulated complex fMRI-like set of 8 spatial components was generated and mixed with a set of time courses to obtain a simulated fMRI-like dataset [34]. The sources are generated based on basic knowledge of the statistical characteristics of the underlying neuronal sources.

Eight complex-valued spatial maps (60×60 voxels each) were simulated along with their corresponding time courses (100 samples each). The magnitude spatial maps and the time courses of each source are as shown in Fig. 1. In an fMRI experiment, the phase difference induced by the task activation is typically less than $\pi/9$ [35]. Therefore, we keep the phase of each pixel uniformly distributed in the range $[-\pi/18, \pi/18]$. The phase of each complex-valued time point is generated proportional to its magnitude, but is again restricted to the range mentioned. The spatial sources are unfolded into one-dimensional vectors and mixed by the corresponding time courses. The same procedure is followed for S different subjects, and for each subject every source occurs with different random (real) amplitude. Hence a tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{3600\times100\times S}$ with complex-valued entries is generated. Gaussian complex noise is added to the tensor and the magnitude of the noisy tensor yields the magnitude (image) tensor $\tilde{\boldsymbol{\mathcal{T}}}$, whose CPD provides the source estimates.

Three different simulation setups are tested. The first with the initial eight sources used in [34], the second with the same sources but with noise with different standard deviation in the real and imaginary parts (simulated parallel imaging methods) and the last one with an extra source (Source 9, Fig. 1), which has high spatial and temporal correlation with Source 1. Four different SNRs, ranging from low to high values, are tested in each case, namely: $95.4, 47.6, 22.4$ and 11.2 dB with noise variance equal to 0.5. The rank, R, was set equal to the number of sources for all the simulation setups. The Pearson correlation of the obtained sources with the ground truth is presented in Figures 2, 3 and 4 for the three different scenarios, respectively.

As we can note from Fig. 2, where only sources with low overlap are considered, the best β value is equal to 2.5 and not the commonly used $\beta = 2$ (LS) case. In the high SNR case, for different β values, the results are approximately the same while as the SNR decreases the difference between the performance with $\beta = 2.5$ and $\beta = 2$ increases. In Fig. 5 (left), a histogram is shown of observed values for one spatial

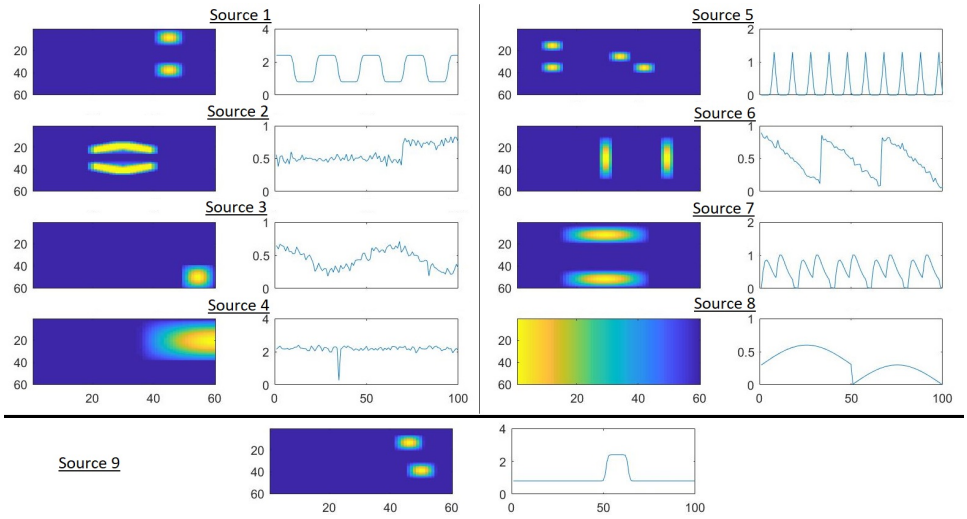

Fig. 1. Magnitude spatial maps and time courses of the complex-valued sources used.

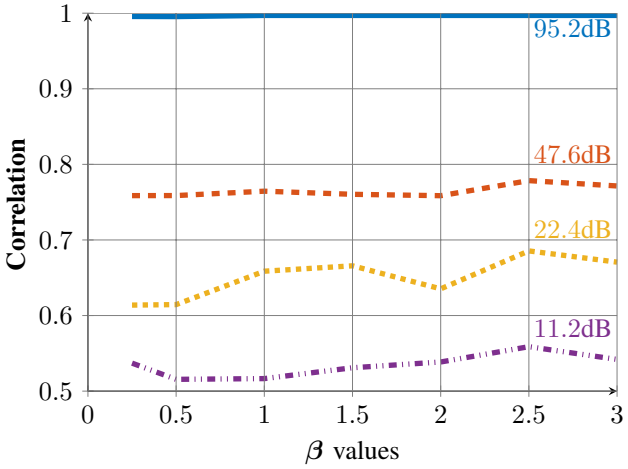

Fig. 2. First simulation with the 8 sources used in [34].

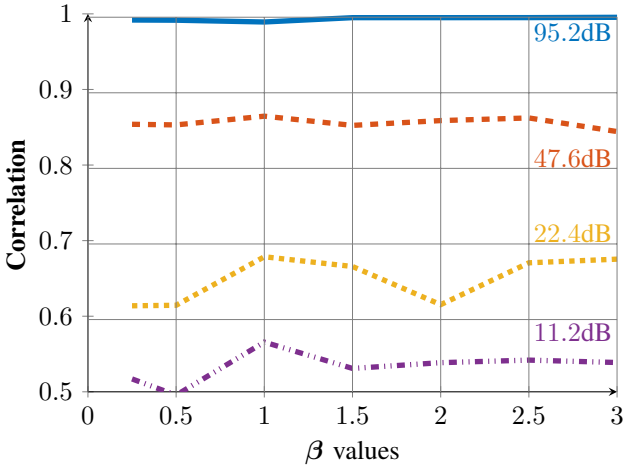

Fig. 3. Second simulation with the 8 sources used in [34] and different noise variance in the real and imaginary parts of the complex-valued source signals.

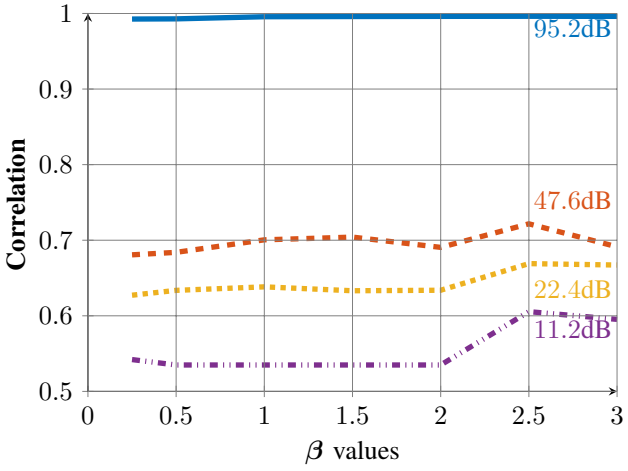

Fig. 4. Third simulation with 9 sources, where Source 9 with high overlap is added to the 8 sources used in [34].

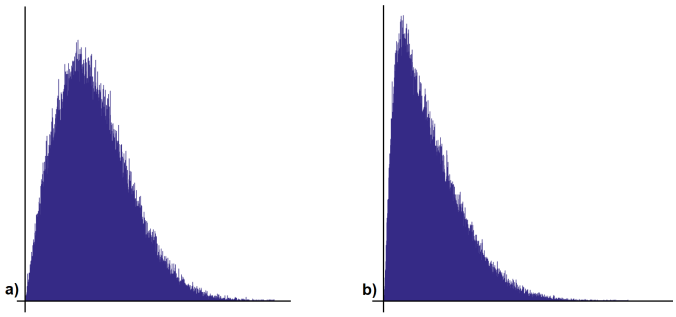

Fig. 5. Histograms for the observed values in one spatial point with real and complex noise variances that a) are equal and b) with the noise variance of the complex part five times that of the real part.

point with a constant time course, where both the real and imaginary parts have been perturbed by Gaussian distributed noise with equal variance. The Rician nature of the data is clearly visible.

In the second case (Fig. 3), with different noise variance in the real and imaginary parts, we note that $\beta = 1$ (KL divergence) gives the best performance, while also $\beta = 2.5$ still has better performance than $\beta = 2$. In Fig. 5 (right), a histogram is shown for one spatial point where the complex part has been perturbed by Gaussian distributed noise with five times the variance of the noise of the real part. Most entries are very small, but the right tail of the distribution is long, pointing to a large number of outliers. It seems that the KL divergence is more robust to these outliers and the non-circularity of the complex data that the different noise variances can introduce.

Last, when an extra source with high spatial and temporal overlap is added (Source 9 in Fig. 1), the gain in performance with $\beta = 2.5$ (and $\beta = 3$) compared to $\beta = 2$ increases, as can be noted in Fig. 4. The overlapping sources result in an increase of the signal magnitude in the specific area, hence these values are fitted relatively better than the non-overlapping values. As mentioned before, for $\beta < 2$, errors in smaller entries are penalized more heavily than for the squared distance, while for $\beta > 2$ the converse is true. Hence in the case of overlapping sources, relatively more importance is given to fitting these entries well compared to the squared distance.

V. Conclusions

To the best of the authors' knowledge, this is the first time that the Gaussian noise assumption and its influence on the fMRI BSS performance are tested in a tensorial framework. Given that the use of a LS criterion (implied by the Gaussian assumption) for fitting a CPD to fMRI data is not optimal in the presence of Rician noise, we have considered alternative cost functions in this paper, based on β-divergences. Different values of β have been tested and it has been shown that $\beta = 1$ (KL divergence) performs best in cases where different noise variance affects the real and imaginary data. Furthermore, $\beta = 2.5$ outperforms the LS criterion ($\beta = 2$) by producing better separation results in all other cases.

As future work real data will be considered, while also additional constraints that are meaningful for fMRI BSS will be tested. For example, the use of Block Term Decomposition [36], [22] (BTD) and the use of regularizers that force independence or sparsity in the spatial maps, will be examined.

ACKNOWLEDGMENTS

Michiel Vandecappelle is supported by an SB Grant from the Research Foundation–Vlaanderen (FWO). Research furthermore supported by: (1) Flemish Government: This work was supported by the Fonds de la Recherche Scientifique–FNRS and the Fonds Wetenschappelijk Onderzoek–Vlaanderen under EOS Project n° 30468160 (SeLMA); (2) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n° 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information; (3) KU Leuven Internal Funds C16/15/059.

REFERENCES

[1] M. A. Lindquist, "The statistical analysis of fMRI data," *Statistical Science*, vol. 23, no. 4, pp. 439–464, Jun. 2008.

[2] P. Wright, "Magnetic Resonance Imaging," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 56–66, Jan. 1997.

[3] D. B. Rowe and B. R. Logan, "An fMRI activation method using complex data," in *IEEE Int. Symposium Biomed. Imag. Nano to Macro*. Arlington, VA, USA: IEEE, Apr. 2004.

[4] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Res. Med.*, vol. 34, no. 6, pp. 910–914, Feb. 1995.

[5] J. Sijbers, A. den Dekker, J. V. Audekerke, M. Verhoye, and D. V. Dyck, "Estimation of the noise in magnitude MR images," *Magn. Reson. Imag.*, vol. 16, no. 1, pp. 87 – 90, Jan. 1998.

[6] D. Varadarajan and J. P. Haldar, "A majorize-minimize framework for Rician and non-central chi MR images," *IEEE Trans. Medical Imag.*, vol. 34, no. 10, pp. 2191–2202, Oct. 2015.

[7] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press, 2015.

[8] A. H. Andersen and W. S. Rayens, "Structure-seeking multilinear methods for the analysis of fMRI data," *NeuroImage*, vol. 22, no. 2, pp. 728–739, Jun. 2004.

[9] V. D. Calhoun and T. Adalı, "Unmixing fMRI with independent component analysis," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 2, pp. 79–90, Mar. 2006.

[10] L. Kangjoo, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1076–1089, Dec. 2011.

[11] N. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometrics*, vol. 14, no. 3, pp. 229–239, May 2000.

[12] A. Stegeman, "Comparing Independent Component Analysis and the PARAFAC model for artificial multi-subject fMRI data," Unpublished Technical Report, Univ. of Groeningen, Feb. 2007.

[13] N. E. Helwig and S. Hong, "A critique of Tensor Probabilistic Independent Component Analysis: implications and recommendations for multi-subject fMRI data analysis," *J. Neuroscience Methods*, vol. 213, no. 2, pp. 263–273, Mar. 2013.

[14] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject fMRI analysis," *NeuroImage*, vol. 25, no. 1, pp. 294–311, Mar. 2005.

[16] C. Chatzichristos, E. Kofidis, Y. Kopsinis, M. M. Moreno, and S. Theodoridis, "Higher-order block term decomposition for spatially folded fMRI data," in *Latent Variable Analysis and Signal Separation Conf. (LVA/ICA)*, Grenoble, France, Feb. 2017.

[15] K. Madsen, N. W. Churchill, and M. Mørup, "Quantifying functional connectivity in multi-subject fMRI data using component models," *Human Brain Mapping*, vol. 38, no. 2, pp. 882–899, Oct. 2016.

[17] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, and A. Tristán-Vega, "Noise estimation in parallel MRI: GRAPPA and SENSE," *Magn. Reson. Med.*, vol. 32, no. 3, pp. 281–290, Apr. 2014.

[18] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Work. Papers in Phonetics*, pp. 1–84, 1970.

[19] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[20] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and Its Applications*, vol. 18, pp. 95–138, 1977.

[21] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Parts I–II," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 855–903, Mar. 2013.

[22] C. Chatzichristos, E. Kofidis, M. M. Moreno, and S. Theodoridis, "Blind fMRI source unmixing via higher-order tensor decompositions," *J. Neuroscience Methods*, vol. 315, pp. 17–47, Mar. 2019.

[23] R. Bro and H. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *J. Chemometrics*, vol. 17, no. 5, pp. 274–286, May 2003.

[24] J. L. Castellanos, S. Gómez, and V. Guerra, "The triangle method for finding the corner of the L-curve," *Appl. Numer. Math.*, vol. 43, no. 4, pp. 359–373, Dec. 2002.

[25] R. Hennequin, B. David, and R. Badeau, "Beta-divergence as a subclass of Bregman divergence," *IEEE Signal Process. Letters*, vol. 18, no. 2, pp. 83–86, Feb. 2011.

[26] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura–Saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[27] S. Ferdowsi, V. Abolghasemi, and S. Sanei, "A comparative study of α-divergence based NMF techniques for fMRI analysis," in *Eur. Signal Process. Conf. (EUSIPCO)*, Barcelona, Spain, Sep. 2011.

[28] A. M. Wink and J. B. Roerdink, "BOLD noise assumptions in fMRI," *Int. J. of Biomed. Imag., Article ID 12014*, 2006.

[29] L. Lauwers and K. Barbé, "General linear models under Rician noise for fMRI data," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015)*, Brisbane, QLD, Australia, April 2015.

[30] D. Hong, T. G. Kolda, and J. A. Duersch, "Generalized canonical polyadic tensor decomposition," *arXiv:1808.07452*, Jan. 2019.

[31] C. Févotte and A. Ozerov, "Notes on nonnegative tensor factorization of the spectrogram for audio source separation: statistical insights and towards self-clustering of the spatial cues," in *Int. Symposium Computer Music Modeling Retrieval (CMMR 2010)*, Málaga, Spain, Jun. 2010.

[32] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Trans. Fund. Electr., Commun. Comp. Sciences*, vol. 92, no. 3, pp. 708–721, Mar. 2009.

[33] M. Vandecappelle, N. Vervliet, and L. De Lathauwer, "Canonical polyadic decomposition with general cost functions using generalized Gauss–Newton," Internal Report 19-05, ESAT-STADIUS, KU Leuven (Leuven, Belgium), May 2019.

[34] V. D. Calhoun, T. Adalı, G. D. Pearlson, P. C. M. Van Zijl, and J. Pekar, "Independent component analysis of fMRI data in the complex domain," *Magn. Reson. Med.*, vol. 48, no. 1, pp. 180–192, Jul. 2002.

[35] F. G. Hoogenrad, J. R. Reichenbach, E. M. Haacke, S. Lai, K. Kuppusamy, and M. Sprenger, "In vivo measurement of changes in venous blood-oxygenation with high resolution functional MRI at .95 Tesla by measuring changes in susceptibility and velocity," *Magn. Reson. Med.*, vol. 39, no. 1, pp. 97–107, Jul. 1998.

[36] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms–Part I-III," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 1022–1083, Sep. 2008.