Technical University of Denmark

DTU

Frequency Constrained ShiftCP Modeling of Neuroimaging Data

Mørup, Morten; Hansen, Lars Kai; Madsen, Kristoffer H.

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Frequency Constrained ShiftCP Modeling of Neuroimaging Data

Morten Mørup and Lars K. Hansen
Section for Cognitive Systems,
DTU Informatics
Technical University of Denmark
{mm,lkh}@imm.dtu.dk

Kristoffer H. Madsen
Danish Research Centre for Magnetic Resonance
Copenhagen University Hospital Hvidovre
khm@drcmr.dk

Abstract—The shift invariant multi-linear model based on the CandeComp/PARAFAC (CP) model denoted ShiftCP has proven useful for the modeling of latency changes in trial based neuroimaging data[17]. In order to facilitate component interpretation we presently extend the shiftCP model such that the extracted components can be constrained to pertain to predefined frequency ranges such as alpha, beta and gamma activity. To infer the number of components in the model we propose to apply automatic relevance determination by imposing priors that define the range of variation of each component of the shiftCP model and learning the hyper-parameters of these priors during model estimation.

I. INTRODUCTION

Analysis of neuroimaging data sets such as electro- and magneto-encephalography (EEG and MEG) or functional magnetic resonance imaging (fMRI), is hampered by noise, confounds, and the presence of multiple mixed signal components of interest. To overcome poor signal to noise (SNR) data is typically measured across multiple trials and subsequently averaged. These averaged EEG and fMRI data may be represented by a *space × time* matrix $\boldsymbol{X} \in \mathbb{R}^{I \times J}$ with elements $x_{i,j}$. Bi-linear component analyses are routinely applied to neuroimaging data for exploratory investigations or as a pre-processing step prior to signal detection, see for instance [5], [12], [13]. The bi-linear model reads

$$x_{i,t} = \sum\nolimits_{d=1}^{D} a_{i,d} b_{t,d} + \varepsilon_{i,t},$$

where the data is represented as a sum of components with time profiles $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_D$ and corresponding spatial topographies $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_D$, and where $\varepsilon_{i,j}$ is i.i.d. noise. Since such factor analytic representations are ambiguous, additional constraints may be imposed. For singular value decomposition (SVD) and principal component analysis (PCA) profiles are assumed orthogonal as eigenvectors of the covariance matrix, while for independent component analysis (ICA) an independence assumption is imposed for one of the two modes.

When data are recorded over K repeated trials, we obtain a *space×time×trials* hypermatrix $\mathcal{X} \in \mathbb{R}^{I \times T \times K}$ also commonly denoted a multi-way array or tensor. A natural extension of the bi-linear analysis is to introduce a trial dependent weight $c_{k,d}$ which form the CandeComp/PARAFAC model [7], [4] jointly abbreviated CP given by

$$x_{i,t,k} = \sum\nolimits_{d=1}^{D} a_{i,d} b_{t,d} c_{k,d} + \varepsilon_{i,t,k}.$$

Thus, $\boldsymbol{c}_d$ represents the strength in which the profile time series $\boldsymbol{b}_d$ with spatial topography $\boldsymbol{a}_d$ is expressed throughout the trials. The CP model is unique under mild conditions that in general are satisfied in the presence of noise [11]. Consequently, modeling repeated trials by CP in theory resolves the ambiguities encountered when modeling the data by (bi-linear) factor analysis. The application of CP to EEG was first suggested in [7] and was later reinvented in [14] under the name *topographic component analysis*. In [1] it was further demonstrated how the CP model is useful in the analysis of fMRI. Unfortunately, the CP model is known to suffer from degeneracy issues when analyzing neuroimaging data such as EEG and fMRI [6], [2], [17] which hampers model interpretation as the extracted components become highly correlated. In [17] it was however demonstrated that extending the CP model to include a trial specific delay forming the shiftCP model [9], [10], [17] alleviates issues of degeneracy. The shiftCP model is given by

$$x_{i,t,k} = \sum\nolimits_{d=1,\tau=1}^{D,T} a_{id} b_{t-\tau_{id},d} c_{kd} + \varepsilon_{i,t,k}.$$

Here, the time profile $\boldsymbol{b}_d$ is present with delay τ_{kd} in trial k.

In this contribution we elaborate on the above shiftCP model. We presently extend the shiftCP model such that the model can constrain the components to predefined frequency ranges. As the interpretation of EEG signals is commonly based on the spectral ranges of the signals in terms of activation within frequency bands such as Delta (0-4 Hz), Theta (4-8 Hz), Alpha (8-13 Hz, Beta (13-30 Hz) and Gamma (>30 Hz) constraining components in terms of these types of activation can potentially facilitate component interpretation. We additionally propose the use of automatic relevance determination which has recently been extended to multi-way models [18] in order to infer the number of components in the shiftCP model. The proposed extensions of the shiftCP model is evaluated on synthetic data and real event related EEG. A Matlab implementation of the shiftCP model and the proposed extensions are available for download from www.mortenmorup.dk.

II. METHODS

A. notation

Let $\boldsymbol{A}^{I \times D}$, $\boldsymbol{B}^{T \times D}$ and $\boldsymbol{C}^{K \times D}$ be matrices holding the profiles defined in the shiftCP model in equation (I). Let further $\boldsymbol{a}_d$ denote the dth column of $\boldsymbol{A}$ and $\boldsymbol{a}_{i,:}$ the ith row of $\boldsymbol{A}$. In the following $\boldsymbol{B}$ and $\tilde{\boldsymbol{B}}$ will denote the same matrix in the time and frequency domain respectively, i.e. $\tilde{\mathcal{U}} = \mathfrak{F}(\mathcal{U})$, $\mathcal{U} = \mathfrak{F}^{-1}(\tilde{\mathcal{U}})$ using the discrete fourier transform (DFT) and inverse fourier transform along the modality indexed by $t \in \{1, 2, ..., T\}$. Let further $\circ$ define the direct product, i.e. $(\boldsymbol{P} \circ \boldsymbol{Q})_{lm} = p_{lm} q_{lm}$ and $\odot$ denote the Khatri-Rao product given by $(\boldsymbol{P} \odot \boldsymbol{Q})_{(l-1)M+m,d} = p_{ld} q_{md}$. Finally, the 1-mode, 2-mode and 3-mode matricizing operation are given by $(\mathcal{X}_{(1)})_{i,j+(k-1)J} = x_{i,j,k}$, $(\mathcal{X}_{(2)})_{j,i+(k-1)I} = x_{i,j,k}$ and $(\mathcal{X}_{(3)})_{k,i+(j-1)I} = x_{i,j,k}$ respectively (see also [15] for further details on these operators).

B. The frequency constrained shiftCP model

In the frequency domain the shiftCP model is given by [17]

$$\tilde{x}_{i,f,k} = \sum_{d=1}^{D} a_{i,d} \tilde{b}_{f,d} c_{k,d} exp[-\imath 2\pi \frac{f-1}{T} \tau_{k,d}] + \tilde{e}_{i,f,k}.$$

Thus, the sources $\boldsymbol{b}_d$ are assumed to be periodic such that shifts $\tau_{k,d}$ correspond to the complex multiplication of $\tilde{b}_d$ with the factor $\exp[-\imath 2\pi \frac{f-1}{T} \tau_{k,d}]$. Notice, due to Parseval's identity there is a one-to-one correspondence between the least squares error in the time and frequency domain such that the least squares minimization can be performed arbitrarily between the two domains

$$\sum_{i,t,k} \|\mathcal{E}_{i,t,k}\|^2 = \frac{1}{T} \sum_{i,f,k} \|\tilde{\mathcal{E}}_{i,f,k}\|^2.$$

A simple approach to constrain the shiftCP model to a given frequency range would be to band-pass filter the data $\mathcal{X}$. However, we are presently interested in extracting components pertaining to specific frequency ranges such that the data $\mathcal{X}$ is modelled as a mixture of components pertaining to different frequency ranges. This can be expressed by the following frequency constrained shiftCP model

$$
\begin{aligned}
x_{i,t,k} &= \sum_{d=1}^{D} a_{i,d} b_{t-\tau_{k,d},d} c_{k,d} + \varepsilon_{i,t,k} \\
s.t. \quad &\tilde{b}_{f,d} = 0 \text{ if } w_{fd} = 0.
\end{aligned}
$$

$\boldsymbol{W}$ is a binary matrix where the element $w_{fd} = 1$ indicate that frequency f is included in component d and $w_{fd} = 0$ that the frequency is not included where $\tilde{\boldsymbol{b}}_d$ corresponds to the frequency domain representation of $\boldsymbol{b}_d$.

We will formulate the problem in a Bayesian framework which admit model order estimation through automatic relevance determination, see also [18]. We impose normal distributed priors on $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ with a shared componentwise hyper-parameter λ_d as proposed in [19] for bi-linear models.

Thus,

$$P(\boldsymbol{A}|\boldsymbol{\lambda}) = \prod_d (\frac{\lambda_d}{2\pi})^{I/2} \exp\left(-\frac{\lambda_d \|\boldsymbol{a}_d\|_F^2}{2}\right)$$

$$P(\boldsymbol{B}|\boldsymbol{\lambda}) = \prod_d (\frac{\lambda_d}{2\pi})^{T/2} \exp\left(\frac{\lambda_d \|\boldsymbol{b}_d\|_F^2}{2}\right)$$

$$P(\boldsymbol{C}|\boldsymbol{\lambda}) = \prod_d (\frac{\lambda_d}{2\pi})^{K/2} \exp\left(\frac{\lambda_d \|\boldsymbol{c}_d\|_F^2}{2}\right)$$

We further specify i.i.d. Gamma priors on the elements of $\boldsymbol{\lambda}$, i.e.

$$P(\boldsymbol{\lambda}|\alpha, \beta) = \prod_d \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_d^{\alpha-1} \exp\left(-\beta \lambda_d\right)$$

On τ we use a uniform prior giving equal weight to all of the T potential shifts

$$P(\tau|J) = \begin{cases} \frac{1}{T} & \text{for } -T/2 < \tau \leq T/2 \\ 0 & \text{otherwise} \end{cases}$$

Further assuming Gaussian i.i.d. noise with variance σ^2 we have

$$P(\mathcal{X}|\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{\tau}, \sigma^2) = \prod_{itk} \frac{\exp\left(-\frac{(x_{i,t,k} - \sum_d a_{id} b_{t-\tau_{i,d},d} c_{i,d})^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}},$$

such that the joint log-likelihood is given by

$$
\begin{aligned}
\log P&(\mathcal{X}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{\tau}, \boldsymbol{\lambda}|\sigma^2, \alpha, \beta) = \\
&-\frac{1}{2\sigma^2} \sum_{itk} (x_{i,t,k} - \sum_d a_{id} b_{t-\tau_{k,d},d} c_{k,d})^2 \\
&+\frac{1}{2} \sum_d (I + T + K + \alpha - 1) \log \lambda_d \\
&-\frac{\lambda_d}{2} (\|\boldsymbol{a}_d\|_F^2 + \|\boldsymbol{b}_d\|_F^2 + \|\boldsymbol{c}_d\|_F^2 + \beta) + const.
\end{aligned}
$$

$const$ is a constant independent of the parameters we will infer in the model. We estimate the model parameters by maximizing the above joint posterior applying an alternating strategy where we estimate one set of parameters keeping the remaining set of parameters of the model fixed.

C. Estimating $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$

Let $\tilde{c}_{k,d}^{(f)} = c_{k,d} \exp[-\imath 2\pi \frac{f-1}{J} \tau_{k,d}]$ and define $\tilde{b}_{f,d}^{(k)} = \tilde{b}_{f,d} \exp[-\imath 2\pi \frac{f-1}{J} \tau_{k,d}]$, i.e. $\tilde{\boldsymbol{B}}$ componentwise shifted according to the delays to the kth channel and. Setting $z_{t+T(k-1),d} = c_{k,d} b_{t,d}^{(k)}$, i.e. the Khatri-Rao product between $\boldsymbol{C}$ and the shifted version of $\boldsymbol{B}$ and using n-mode matricizing and the Khatri-Rao product operations we can state the estimation of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ according to

$$\boldsymbol{A} \leftarrow \boldsymbol{X}_{(1)} \boldsymbol{Z} (\sigma^2 \operatorname{diag}(\boldsymbol{\lambda}) + \boldsymbol{Z}^\top \boldsymbol{Z})^{-1},$$

$$\tilde{\boldsymbol{B}}_{f,\mathcal{W}_f} \leftarrow \tilde{\boldsymbol{X}}_{(2)_{f,:}} (\tilde{\boldsymbol{C}}_{\mathcal{W}_f}^{(f)} \odot \boldsymbol{A}_{\mathcal{W}_f}) (\sigma^2 \operatorname{diag}(\boldsymbol{\lambda}_{\mathcal{W}_f}) + \tilde{\boldsymbol{C}}_{\mathcal{W}_f}^{(f)^\top} \tilde{\boldsymbol{C}}_{\mathcal{W}_f}^{(f)} \circ \boldsymbol{A}_{\mathcal{W}_f}^\top \boldsymbol{A}_{\mathcal{W}_f})^{-1},$$

$$\boldsymbol{C}_{k,:} \leftarrow \boldsymbol{X}_{(3)_{k,:}} (\boldsymbol{B}^{(k)} \odot \boldsymbol{A}) (\sigma^2 \operatorname{diag}(\boldsymbol{\lambda}) + \boldsymbol{B}^{(k)^\top} \boldsymbol{B}^{(k)} \circ \boldsymbol{A}^\top \boldsymbol{A})^{-1}.$$

The frequency domain constraints on the components of $\boldsymbol{B}$ are imposed by changing the update for $\boldsymbol{B}$ in the frequency domain to only include the components that are active at the
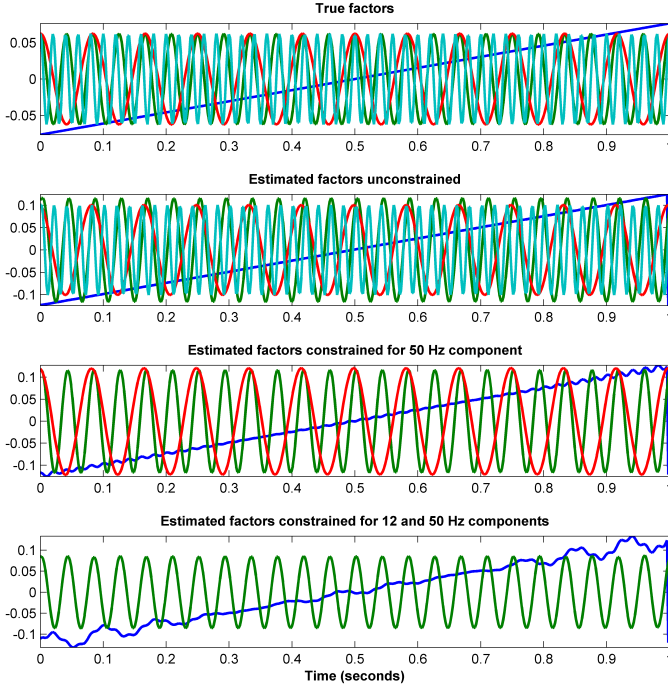
Fig. 1. Synthetic data example. Data is generated according to the shiftCP model and the parameters are inferred without frequency constraints, constraining all the components to disregard the 50 Hz and both 12 Hz and 50 Hz activity respectively. The automatic relevance determination framework was applied in order to prune irrelevant components of the 10 components fitted to the data.

given frequency f given by the set $\mathcal{W}_f = \{d'|w_{f,d'} = 1\}$. Thus, if a component d'' is constrained to not include frequency f we force $\tilde{b}_{f,d''} = 0$. Notice, whereas $\boldsymbol{A}$ and $\boldsymbol{C}$ are updated in the real domain $\boldsymbol{B}$ is updated in the complex domain. However, $\boldsymbol{B}$ is only real valued in the time domain if the following relation holds in the frequency domain [17]

$$\tilde{b}_{T-f+1,d} = \tilde{b}_{f,d}^*,$$

such that $\tilde{\boldsymbol{B}}$ is conjugate symmetric. This constraint is enforced by updating the first $\lfloor T/2 \rfloor + 1$ elements, i.e. up to the Nyquist frequency, while setting the remaining elements according to equation the above. Since the estimation is stated as regular factor analysis problems non-negativity constraints for $\boldsymbol{A}$ and $\boldsymbol{C}$ can be imposed using the active set procedure given in [3], see also [17].

D. Estimating $\boldsymbol{\tau}$

Following [17] we define

$$\boldsymbol{R}_{(3)_{k,:}}^{(d')} = \boldsymbol{X}_{(3)_{k,:}} - \sum_{d \neq d'} c_{k,d}(\boldsymbol{b}_d^{(k)} \odot \boldsymbol{a}_d)^\top,$$

i.e. $\boldsymbol{R}_{(3)_{k,:}}^{(d')}$ is the remaining signal at the kth row when projecting all but the d'^{th} source out of $\boldsymbol{X}_{(3)}$. Let,

$$s_{k,d'}(t) = \sum_i \mathcal{R}_{i,t,k}^{(d')} a_{i,d'},$$
$$\tilde{v}_{k,d'}(f) = \tilde{s}_{k,d'}^*(f)\tilde{b}_{f,d'}.$$

$\tau_{k,d'}$ is then estimated according to [17]

$$\hat{\tau}_{k,d'} = \operatorname*{argmax}_t |v_{k,d'}(t)|$$
$$\tau_{k,d'} = \hat{\tau}_{k,d'} - (T+1).$$

I.e. as the delay corresponding to maximum absolute cross-correlation between $s_{k,d'}(t)$ – the time profile of the residual for the d' component and $\boldsymbol{b}_{d'}$-the component time profile.

E. Estimating $\boldsymbol{\lambda}$ and setting σ^2

Differentiating the joint likelihood with respect to λ_d and setting the gradient to zero we obtain the update

$$\lambda_d = \frac{I + T + K + \alpha - 1}{\|\boldsymbol{a}_d\|_F^2 + \|\boldsymbol{b}_d\|_F^2 + \|\boldsymbol{c}_d\|_F^2 + \beta}.$$

Rather than estimating σ^2 from data we follow [18] and set σ^2 according to a predefined signal to noise ratio (SNR) of 0db. As a result we have

$$\sigma^2 = \frac{\|\mathcal{X}\|_F^2}{(1 + 10^{\text{SNR}/10})ITK} = \frac{\|\mathcal{X}\|_F^2}{2ITK}.$$

III. EXPERIMENTAL RESULTS AND DISCUSSION

In our analysis we set $\alpha = 1$ and $\beta = 10^{-9}\frac{\|\mathcal{X}\|}{ITK}$. The convergence criterion of the algorithm was set to a relative change in fit less than 10^{-6} or when the algorithm had run for 500 iterations. The parameters in the frequency constrained shiftCP model were updated in the sequence; $\boldsymbol{B}, \boldsymbol{\tau}, \boldsymbol{A}, \boldsymbol{C}$ and the updates were accelerated using the approach suggested in [20]. By imposing the componentwise hyper-parameter λ_d defining the range of variation of each component the model order is estimated based on the update for $\boldsymbol{\lambda}$, i.e. the larger λ_d becomes the more the parameters of the dth component will be regularized towards zero. We updated $\boldsymbol{\lambda}$ every 5th iteration. For details on the implementation consult the Matlab script available for download at www.mortenmorup.dk.

A. Simulated data

To investigate that the frequency constrained shiftCP model is able to infer only components within given frequency ranges and that the proposed model order selection approach based on ARD can be used to infer the number of components we considered a synthetical dataset generated according the the shiftCP model based on the following four components; a drift, 12 Hz, 24 Hz and 50 Hz oscillations given by perfect sinusoids. The data was generated with SNR $= 5dB$.

Figure 1 shows the frequency constrained shiftCP analysis of the synthetically generated data without frequency constraints, all components constrained to not include 50 Hz and all components constrained to disregard both 50 Hz and 12 Hz. By updating the hyper-parameters for each of the 10 imposed components (i.e., λ_d) only relevant components prevailed while additional components were pruned from the model. Thus, the model correctly identifies all four components when there is no frequency constraints, three components when constrained for the 50 Hz component and 2 components when

Fig. 2. **Top panel:** 40 component analysis of the real event related EEG data without frequency constraints. **Bottom panel:** 40 component analysis of the real event related EEG data with frequency constraints. The delay and trial dependent strength is plotted in a 2D-plane as a complex vector given by $c_{k,d} \exp(-i2\pi\tau_{k,d}/T)$

constrained both for the 12 Hz and 50 Hz components. Notice that when constraining the components for 12 Hz and both 12 Hz and 50 Hz the drift which also contain 12 Hz and 50 Hz activation is no longer perfectly recovered.

B. Real event related EEG data

We finally analyzed the real event related EEG data considered in [17]. The data is based on a visual object identification task. The data contain a 64 channel recording of a healthy male subject based on the visual stimulus paradigm [8]. The paradigm consists of two types of black and white drawings: (1) objects (Ob), which are easily recognizable everyday type of objects like a chair, a number or a pipe, and (2) non-objects (Nob), which are random re-arrangements of the Ob drawings. Each stimulus category included 313 events and an object was presented up to three times. The data were down sampled to 512 Hz and referenced to digitally linked earlobes. No trials were rejected – instead the data were high-pass filtered > 3Hz to remove the most heavily confounding drift and slow

wave effects. 50 Hz electronic noise was projected out using a multiple linear regression filter in intervals of 2 seconds. The data were cut into trials −500 to 1500 ms forming the data array $I = 64$ *channel* $\times T = 1024$ *time-points* $\times K = 313$ *trials*.

The data was analyzed using both an unconstrained shiftCP model as well as a frequency constrained shiftCP model including 40 components such that sets of 10 components were defined to disregard activation above 12 Hz, frequencies below 12 Hz and above 30 Hz, frequencies below 30 Hz and above 80 Hz and frequencies below 80 Hz respectively. We further constrained C to be non-negative in order to model the trial specific strength such that the sign of topography and time courses for each component are consistent across the trials. In both analysis we used ARD in order to attempt to infer the model order.

In the top panel of figure 2 is shown a 40 component shiftCP analysis of the data. The extracted components are all mixtures of various frequency bands. In the bottom panel of figure 2

is given the analysis imposing various frequency constraints on the components. Clearly the patterns of activation pertains to these specified frequency ranges. In particular, the high frequency components appear to model a strongly localized activation in the left hemisphere. For both models no components were pruned by ARD.

IV. CONCLUSION

We presently extended the shiftCP model of [17] such that the components could be constrained to specific frequency ranges. We further formulated the shiftCP model in a Bayesian framework and used automatic relevance determination (ARD) to automatically estimate from the data the number of components. The success of the approach was demonstrated on synthetic data where the components of the model when constrained for given frequencies disregarded these patterns of activations while the ARD framework correctly pruned all unnecessary components.

In the analysis of real event related EEG it was demonstrated how components could be constrained to capture different consistent frequency domain signatures of the data in separate components. However, in both the shiftCP and the frequency constrained shiftCP analyses no components were pruned by ARD. We attribute this to the EEG data supporting more than the presently considered 40 components. In particular, many of the components extracted resembles patterns of bursts. These bursts may appear multiple times across the duration of each trial of 2 seconds. However, multiple bursts within trials varying in amplitude, shape and relative onsets across trials can only well be accommodated by the shiftCP model by including separate components for each burst. We believe a convolutive CP model as proposed in [16] may be able to address this issue. This will be the focus of future research.

The proposed framework readily generalize to other multi-way models such as the Tucker model, see also [15] for a review of multi-way modeling approaches. By the frequency constrained shiftCP model we have demonstrated how domain knowledge such as frequency information can be imposed in the unsupervised modeling of multi-way neuroimaging data. We note that the proposed frequency constrained modeling approach can also be combined with other types of domain specific constraints such as constraining the estimated temporal delays to given time intervals or imposing orthogonality constraints on the extracted topographic maps.

REFERENCES

[1] Anders H. Andersen and William S. Rayens. Structure-seeking multilinear methods for the analysis of fmri data. *NeuroImage*, 22:728–739, 2004.

[2] C.F. Beckmann and S.M. Smith. Tensorial extensions of independent component analysis for multisubject fmri analysis. *NeuroImage 25*, pages 294–311, 2005.

[3] Rasmus Bro and Sijmen de Jong. A fast non-negativity-constrained least squares algorithm. *J. of Chemometrics*, 11(5):393–401, 1997.

[4] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35:283–319, 1970.

[5] E. Donchin and E. Heffley. Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, pages pp. 555–572, 1978.

[6] Aaron S. Field and Daniel Graupe. Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topography*, 3(4):407–423, 1991.

[7] R. A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[8] C.S. Herrmann, D. Lenz, S. Junge, N.A. Busch, and B. Maess. Memory-matches evoke human gamma-responses. *BMC Neuroscience*, 5(13), 2004.

[9] Sungjin Hong and Richard A. Harshman. Shifted factor analysis part iii: N-way generalization and application. *Journal of Chemometrics*, 17:389–399, 2003.

[10] K.H. Knuth, A.S. Shah, W.A. Truccolo, M. Ding, S.L. Bressler, and C.E. Schroeder. Differentially variable component analysis: Identifying multiple evoked components using trial-to-trial variability. *J. Neurophysiol.*, 95:3257–3276, May 2006.

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

[12] Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. Independent component analysis of electroencephalographic data. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 145–151. The MIT Press, 1996.

[13] M J McKeown, T P Jung, S Makeig, G Brown, S S Kindermann, T W Lee, and T J Sejnowski. Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc Natl Acad Sci U S A*, 95(3):803–810, Feb 1998.

[14] J. Möcks. Topographic components model for event-related potentials and some biophysical considerations. *IEEE Trans. Biomed. Eng.*, 35:482–484, 1988.

[15] M. Mørup. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):24–40, 2011.

[16] M. Mørup, L. K. Hansen, and K. H. Madsen. Modeling latency and shape changes in trial based neuroimaging data. *invited paper, ASILOMAR-SSC session on Tensor Based Array Signal Processing*, 2011.

[17] M Mørup, L.K. Hansen, S.M. Arnfred, L.-K. Lim, and K.M. Madsen. Shift invariant multilinear decomposition of neuroimaging data. *NeuroImage*, 42(4):1439–50, 2008.

[18] Morten Mørup and Lars Kai Hansen. Automatic relevance determination for multi-way models. *Journal of Chemometrics*, 23:352–363, 2009.

[19] V. Y. F. Tan and C. Févotte. Automatic relevance determination in nonnegative matrix factorization. In *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, St-Malo, France, Apr. 2009.

[20] Giorgio Tomasi. *Practical and computational aspects in chemometric data analysis*. PhD thesis, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark, May 2006.