Statistica Sinica **18**(2008): Supplement, S??∼S??

PENALIZED PARAFAC ANALYSIS OF SPONTANEOUS EEG RECORDINGS

Eduardo Martíez-Montes, José M. Sáchez-Bornot and Pedro A. Valdés-Sosa

Cuban Neuroscience Center

Supplementary Material

1. Determination of the Model Order in PARAFAC

Determining the correct number of components in a PARAFAC decomposition is a difficult task, in which experience and prior knowledge on the data to be analyzed is very useful. Although no flawless method exists for this purpose, a few different approaches have been used so far. One is the use of residual analysis (or a Scree-plot), similarly applied to bilinear methods such as Principal Component Analysis. On one hand, if systematic variation is left in the residuals, it is an indication that more components can be extracted. On the other hand one can plot the number of components versus the percentage variance explained by the model or the residuals. Then, only those components explaining large percentage of the variance are considered relevant. Another method is the Split-half analysis, which consists in randomly splitting the data in two halves and fit a PARAFAC model to each of them. Due to the uniqueness, the decomposition will be the same only if the correct number of components is used (Bro (1998)). Another approach is the use of cross-validation (CV) (see e.g., Louwerse, Smilde and Kiers (1999)) and evaluation of information criteria such as Bayesian Information Criteria (BIC) (Morup, Hansen, Herrmann, Parnas and Arnfred (2006)). Finally, other more sophisticated methods have been developed such as the Add-one-Up (Chen, Liu, Cao and Yu (2001)) and the Core Consistency Diagnostic (Corcondia), which apply especially to PARAFAC models (Bro (1998)).

In this paper we use the Corcondia measure, together with the analysis of residual variance. This measure is based on the fact that the PARAFAC model can be seen as a particular case of a more general model known as Tucker 3, whose structural equation is as follows:

$$x_{ijk} = \sum_{d=1}^{N_d} \sum_{e=1}^{N_e} \sum_{f=1}^{N_f} a_{id} b_{je} c_{kf} g_{def} + \varepsilon_{ijk}.$$

The properties and usefulness of this model are extensively described in Bro (1998). For our purposes we only need to show that the main differences of this

model with PARAFAC are i) that the loading matrices can have different number of components (N_d, N_e and N_f respectively) and ii) the existence of the so-called 'core array' $\mathbf{G}_{(N_d \times N_e \times N_f)}$ (with elements g_{def}), which allows for modelling weighted combinations of the loading matrices. Then, this model reduces to PARAFAC when $N_d = N_e = N_f$ and $\mathbf{G}$ is the superidentity three dimensional array (i.e., it has zeroed elements apart from those elements in the superdiagonal which are ones).

After having fitted a PARAFAC model to the data $\mathbf{X}$, verification that the trilinear structure is appropriate can be obtained by calculating the core array of a Tucker 3 model for the same data using the estimated loading matrices by PARAFAC. This estimation is made by least squares regression and if PARAFAC is valid, then this estimated core array $\mathbf{G}$ should resemblance the superidentity array $\mathbf{T}$. A simple way to assess if the model structure is reasonable is therefore to monitor the distribution of superdiagonal and off-superdiagonal elements of $\mathbf{G}$, which can be summarized in a single parameter (Core Consistency) defined as:

$$Core\ Consistency = 100 \left(1 - \frac{\sum_{d=1}^{N_d} \sum_{e=1}^{N_e} \sum_{f=1}^{N_f} \left(t_{def} - g_{def}\right)^2}{\sum_{d=1}^{N_d} \sum_{e=1}^{N_e} \sum_{f=1}^{N_f} g_{def}^2} \right).$$

This is a general measure of the trilinearity of the data, taking the value 100% in the ideal case in which the data conform exactly to the trilinear model. If this is not the case (e.g. if Corcondia is lower than 85%) then either too many components have been extracted, the model is misspecified, or gross outliers disturb the model (Bro (1998)). Among the advantages of the use of this measure is that there is no need for prior assumptions on the distribution of residuals neither to estimate degrees of freedom (as in using information criteria such as CV, BIC). Despite its simplicity, this measure has been shown to be a very powerful tool in assessing the real number of components in several fields of application of the PARAFAC model (Bro and Kiers (2003), Miwakeichi, Martínez-Montes, Valdés-Sosa, Nishiyama, Mizuhara and Yamaguchi (2004), Morup (2005) and Morup, Hansen, Herrmann, Parnas and Arnfred (2006)).

2. Lemma 1 and Its Proof

Lemma 1. *Consider the minimization of the loss function of a multiple penalized linear regression model for a row* $\mathbf{a}^T$*, subject to any constraint:* $\min(\left\|\mathbf{T} - \mathbf{z}\mathbf{a}^T\right\|^2 + \sum \lambda_l P_l(\mathbf{a}))$*. The solution to this problem is equivalent to* $\min(\|\alpha - \mathbf{a}\|^2 + \sum \overline{\lambda}_l P_l(\mathbf{a}))$*, where* α *is the solution of the unconstrained problem:* $\alpha = \arg\min(\left\|\mathbf{T} - \mathbf{z}\alpha^T\right\|^2) = \mathbf{T}^T\mathbf{z}/\mathbf{z}^T\mathbf{z}$*, and* $\overline{\lambda}_l = \lambda_l/\mathbf{z}^T\mathbf{z}$*.*

Proof. Let $\alpha = \mathbf{T}^T\mathbf{z}/\mathbf{z}^T\mathbf{z}$ and $\mathbf{H} = \mathbf{T} - \mathbf{z}\alpha^T$. Then $\min(\left\|\mathbf{T} - \mathbf{z}\mathbf{a}^T\right\|^2 + \sum \lambda_l P_l(\mathbf{a})) = \min(\left\|\mathbf{H} + \mathbf{z}\alpha^T - \mathbf{z}\mathbf{a}^T\right\|^2 + \sum \lambda_l P_l(\mathbf{a})) = \min(\left\|\mathbf{H} + \mathbf{z}(\alpha - \mathbf{a})^T\right\|^2 + \sum \lambda_l P_l(\mathbf{a})) = \min(\|\mathbf{H}\|^2 + 2trace(\mathbf{H}^T\mathbf{z}(\alpha - \mathbf{a})^T) + \mathbf{z}^T\mathbf{z}\,\|\alpha - \mathbf{a}\|^2 + \sum \lambda_l P_l(\mathbf{a}))$. Since $\|\mathbf{H}\|^2$ does not depend on $\mathbf{a}$ (constant term), $\mathbf{H}^T\mathbf{z} = 0$, and $\mathbf{z}^T\mathbf{z}$ is a non-negative scalar, then we can divide the function by $\mathbf{z}^T\mathbf{z}$ and define $\overline{\lambda}_l = \lambda_l/\mathbf{z}^T\mathbf{z}$ to finally get:

$$\min\left(\left\|\mathbf{T} - \mathbf{z}\mathbf{a}^T\right\|^2 + \sum \lambda_l P_l(\mathbf{a})\right) = \min\left(\|\alpha - \mathbf{a}\|^2 + \sum \overline{\lambda}_l P_l(\mathbf{a})\right).$$

3. Data Description and Preproccesing

The data chosen for the study consist of an EEG recording on 16 bipolar derivations (Fp2-F8, F8-T4, T4-T6, T6-O2, O2-P4, P4-C4, C4-F4, F4-Fp2; Fp1-F7, F7-T3, T3-T5, T5-O1, O1-P3, P3-C3, C3-F7, F7-Fp1) for a subject in resting state. The corresponding time-varying spectra are computed with the use of Thomson multitaper method (Thomson (1982)). Then, we end up with a three dimensional data set of 208 320 elements, indexed by 16 derivations, 124 frequencies and 105 time points, that can be subject to PARAFAC analysis as is schematically shown in figure 1 of this supplemental material. Other details about the data can be found in Goldman, Stern, Engel and Cohen (2002).

This data was used previously, together with fMRI data for a concurrent analysis using multi-way Partial Least Squares (Martínez-Montes, Vald és-Sosa, Miwakeichi, Goldman and Cohen (2004)) where only the spectral loading obtained by unconstrained PARAFAC decomposition via ALS was reported. Examination of Corcondia, residual errors and explained variance allowed to determine the appropriate number of components as 3. Here we followed the same preprocessing carried out in that work. Firstly, three outlier time points were identified by the analysis of leverages and removed for subsequent analysis. Secondly, the data was centered along the frequency dimension and scaled through time and space. Finally, the spectral and spatial loadings were normalized while the temporal one kept the scale of the data. In the same way the physical magnitude (microvolts squared) can be assigned to any of the loadings indistinctly, therefore, in this paper we will refer to values of the loadings as energy of the data and omit labels on y-axes of figures.

4. Supplemental figures and Tables

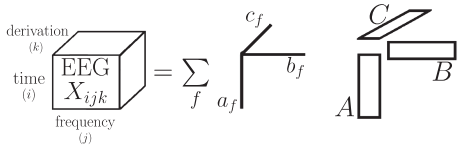

Figure 1. Schematic representation of PARAFAC decomposition. The time-varying EEG spectrum is decomposed as a sum of atoms, each being the outer (tensor) product of signatures corresponding to dimensions. Matrices **A**, **B** and **C** contain atoms as columns.

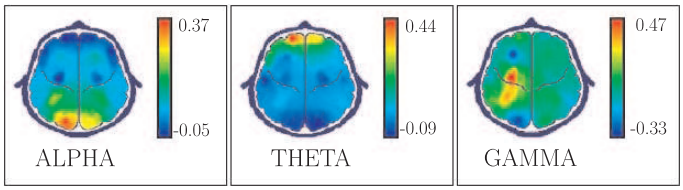

Figure 2. Topographical representation of spatial signatures of the three atoms obtained with unconstrained PARAFAC decomposition. Reproduced from Martínez-Montes, Valdés-Sosa, Miwakeichi, Goldman, and Cohen (2004), (their figure 6). The values in the colorbar represents normalized energy.

Table 1. Parameters of interest for comparing PARAFAC decompositions obtained with different degrees of smoothness for the spectral loadings. Note that among constrained solutions, lambda=1 gives the lowest GCV, RSS and relative distances, while the highest Corcondia.

lambda	0	0.3	0.5	1	5	10	50
logGCV	-3.4581	-3.4138	-3.4493	**-3.4559**	-3.4385	-3.3850	-3.2426
Corcondia	93.43	75.04	93.24	**93.48**	92.59	92.18	90.96
RSS	6512.67	6821.23	6571.55	**6527.91**	6642.44	7008.00	8080.08
df	735.0	549.9	734.9	734.8	734.1	733.2	726.3
Time (s)	0.23	405.36	453.62	140.80	108.11	95.46	90.71
Niter	4	7	14	**6**	6	6	6
RD 1 (%)	0	33.54	9.68	**5.67**	5.75	5.79	8.20
RD 2 (%)	0	81.45	33.21	**28.65**	31.36	32.69	40.67
RD 3 (%)	0	55.26	3.62	**2.43**	5.92	7.91	13.56

Legend: lambda = smoothing parameter; logGCV = logarithm of Generalized Cross Validation function evaluated at lambda; Corcondia = Core Consistency Diagnostic; RSS = residual sum of squares; df = degrees of freedom; Time = time of computation, in seconds; Niter = number of iterations needed for convergence; RD 1 = relative distance between estimated temporal signature and temporal signature of the unconstrained PARAFAC decomposition, in percent. This is defined as the ratio between the norm of the difference between constrained and unconstrained loadings and the norm of the constrained loading. RD 2 and RD 3 = idem to RD 1 but for spectral and spatial loadings respectively.

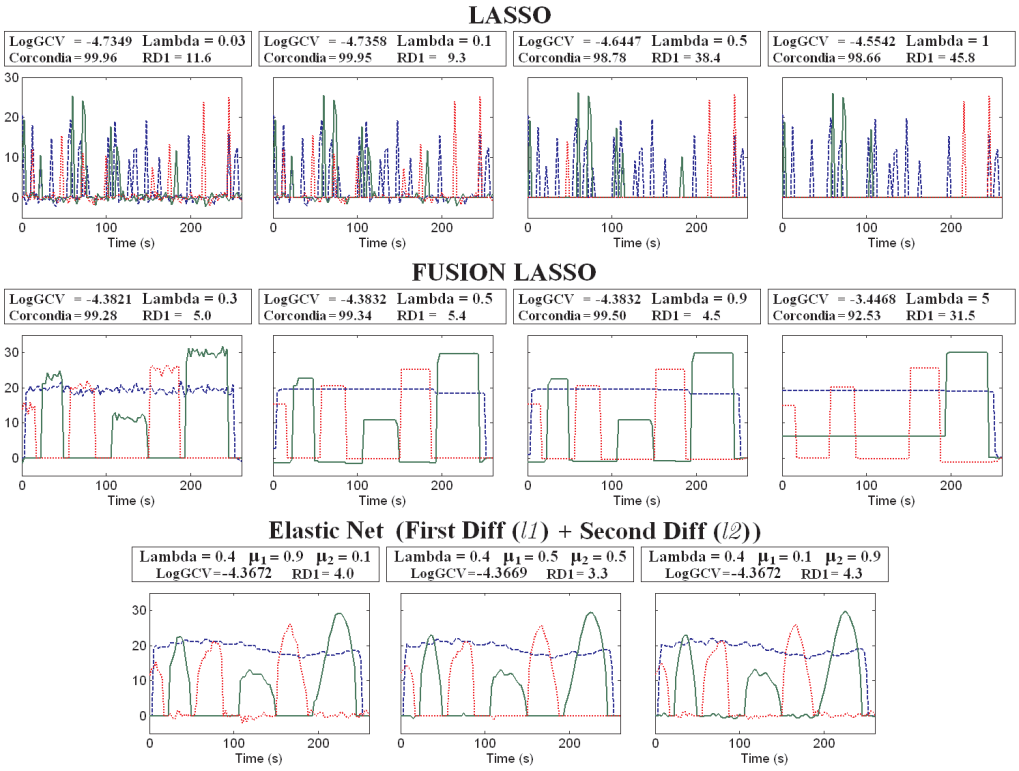

Figure 3. Temporal loadings obtained using different penalizers for the three simulated scenarios presented in figure 3 of the manuscript. In all cases, corresponding values of lambda, logarithm of GCV, Corcondia and relative distances are shown. Green solid line represents the Alpha atom, red dotted line represents the Theta atom and blue dashed line the Gamma atom. Top row: using Lasso penalization for the first simulated scenario (figure 3a of the manuscript), with different weighting parameters. Corcondia was above 99% in all cases, since the data is by construction trilinear, though noisy. However, according to logGCV the optimum value for the weighting parameter is 0.1, which also have the lowest relative distance to the real loading. Middle row: using Fusion Lasso penalization with different values of lambda on the second simulated data (figure 3b of the manuscript). The value $\lambda = 0.9$ seems to be optimal, having the highest Corcondia, the lowest logGCV and the lowest relative distance to the real loading. Bottom row: using Enet penalization on the third simulated scenario (figure 3c of the manuscript). Enet solutions were found using a first order difference operator for the *l1*-norm term and a second order difference operator for the *l2*-norm term. Different values of the parameter λ were explored for three different pairs of weights (μ_1 for *l1*-norm term and μ_2 for the *l2*-norm term) and the solution with the lowest logGCV is presented. The optimal weighting parameter was the same for the three choices of pairs of weights μ_1 and μ_2. Also, values of logGCV and Corcondia (not shown) are almost the same in the three cases and we could select the solution with $\mu_1 = \mu_2 = 0.5$ as the best one based only on the relative distance to the real loading.

References

Bro, R. (1998). Multi-way analysis in the food industry: models, algorithms and applications. Ph.D. Thesis. University of Amsterdam and Royal Veterinary and Agricultural University, Denmark.

Bro, R. and Kiers, H. A. L. (2003). A new efficient method for determining the number of components in PARAFAC models. *J. Chemometr.* **17**, 274-286.

Chen, Z. P., Liu, Z., Cao, Y. C. and Yu, R. Q. (2001). Efficient way to estimate the optimum number of factors for trilinear decomposition. *Anal Chim. Acta* **444**, 295-307.

Goldman, R. I., Stern, J. M., Engel, J. and Cohen, M. S. (2002). Simultaneous EEG and fMRI of the alpha rhythm. *NeuroReport* **13**, 2487-2492.

Louwerse, D. J., Smilde, A. K. and Kiers, H. A. L. (1999). Cross-validation of multiway component models. *J. Chemometr.* **13**, 491-510.

Martínez-Montes, E., Valdés-Sosa, P. A., Miwakeichi, F., Goldman, R. and Cohen, M. (2004). Concurrent EEG/fMRI Analysis by Multi-way Partial Least Squares. Neuroimage **22**, 1023-1034.

Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P. A., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. (2004). Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. *Neuroimage* **22**, 1035-1045.

Morup, M. (2005). Analysis of brain data using multi-way array models on the EEG. Msc Thesis. Technical University of Denmark.

Morup, M., Hansen, L. K., Herrmann, C. S., Parnas, J. and Arnfred, S. M. (2006). Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage* **29**, 938-947.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055-1096.

Neurostatistics Department, Cuban Neuroscience Center.

E-mail: eduardo@cneuro.edu.cu

Neurostatistics Department, Cuban Neuroscience Center.

E-mail: bornot@cneuro.edu.cu

Neurostatistics Department, Cuban Neuroscience Center.

E-mail: peter@cneuro.edu.cu

(Received April 2007; accepted March 2008)