Mining the Bilinear Structure of Data
with Approximate Joint Diagonalization

L. Korczowski*, F. Bouchard, C. Jutten, M. Congedo
GIPSA-lab, Univ. Grenoble Alpes, CNRS, France.
*contact : louis.korczowski@gmail.com

Blind Source Separation by AJD

Approximate Joint Diagonalization (AJD) of a matrix set can solve the linear Blind Source Separation problem [1]. We show how the linear and bilinear joint diagonalization can be applied for extracting sources according to a composite model where some of the sources have a linear structure and others a bilinear structure. This is the case of Event Related Potentials (ERPs). To do so, we developed a Jacobi-like method named Gauss Planar Transformation. It simultaneously diagonalizes the observation matrices X_k and several spatial statistics R_x of the observation (e.g. cospectra and covariance matrices).

The proposed model achieves higher performance in term of shape and robustness for the estimation of ERP sources in a Brain Computer Interface experiment [2].

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t)$$

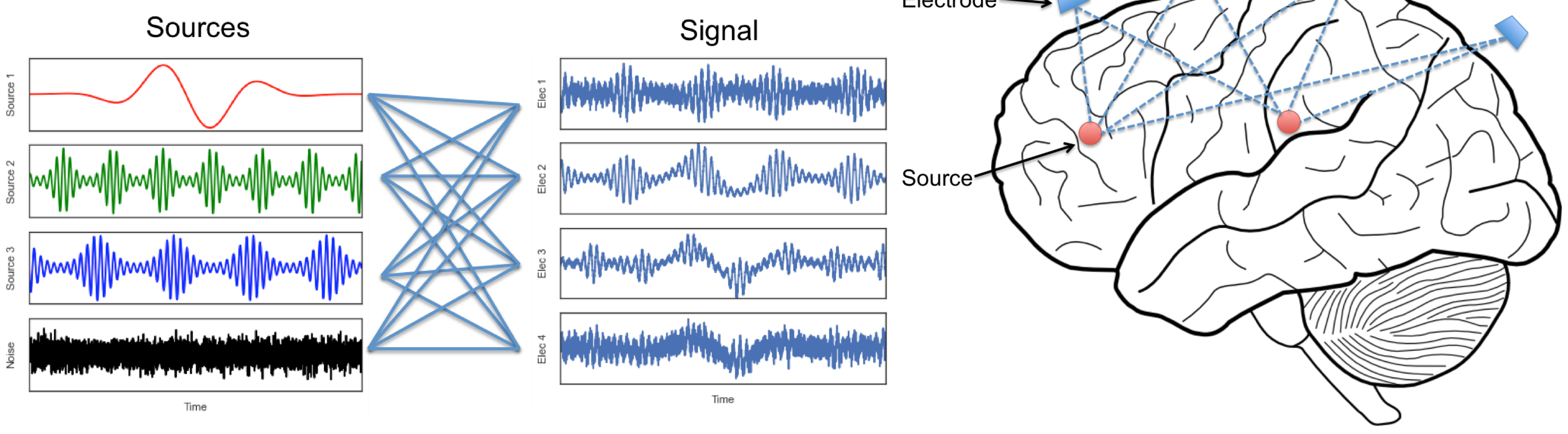

Fig.1 In a linear case , we want to retrieve the independent sources s(t). However using only the spatial statistics can fails to discriminate the transient sources such as Source 1, therefore we propose to use their temporal structure in a flexible manner.

Composite Model

With AJD, data such as EEG evoked potentials X_k can be modeled as additive contribution of *evoked* sources and *continuous* (ongoing) sources (Fig 2) that follow respectively:

- **A Bilinear Model (BAJD)**

$$\underbrace{\boldsymbol{X}_k}_{N\times T} = \underbrace{\boldsymbol{A}}_{N\times N}\underbrace{\boldsymbol{S}_k}_{N\times N}\underbrace{\boldsymbol{E}^T}_{T\times N} + \underbrace{\boldsymbol{N}_k}_{N\times T}$$

can be diagonalized by finding the spatial and temporal unmixing matrix B and D that minimize

$$\mathrm{f}_b(\boldsymbol{B},\boldsymbol{D}) = \sum_{k=1}^{K} \left\| \mathrm{off}(\boldsymbol{B}^T\boldsymbol{X}_k\boldsymbol{D}) \right\|_F^2 \quad (1)$$

- **A Linear Model (AJD)**

$$\underbrace{\boldsymbol{R}_x(l)}_{N\times N} = \boldsymbol{A}\underbrace{\boldsymbol{R}_s(l)}_{N\times N}\boldsymbol{A}^T$$

can be diagonalized by finding the spatial unmixing matrix B that minimizes

$$\mathrm{f}(\boldsymbol{B}) = \sum_{l=1}^{L} \left\| \mathrm{off}(\boldsymbol{B}^T\boldsymbol{R}_x(l)\boldsymbol{B}) \right\|_F^2 \quad (2)$$

Thus, for the Composite (CAJD) model, we jointly minimize the cost functions (1) and (2)

$$\mathrm{f}_c(\boldsymbol{B},\boldsymbol{D}) = \alpha\,\mathrm{f}(\boldsymbol{B}) + (1-\alpha)\mathrm{f}_b(\boldsymbol{B},\boldsymbol{D}) \quad (3)$$

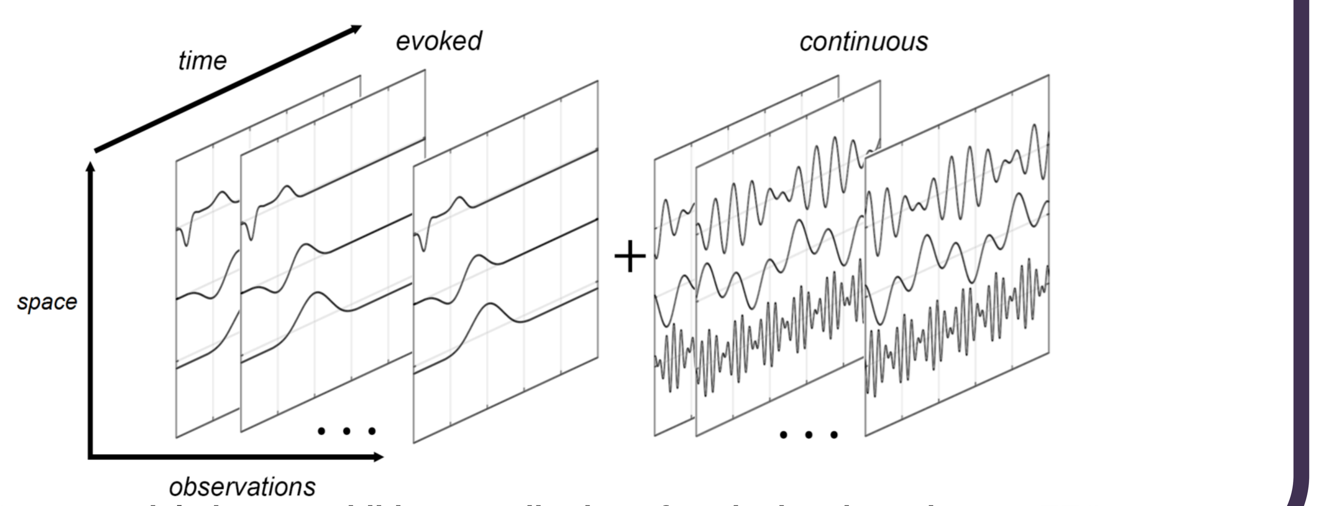

Fig. 2 ERPs are modeled as an additive contribution of evoked and continuous EEG sources.

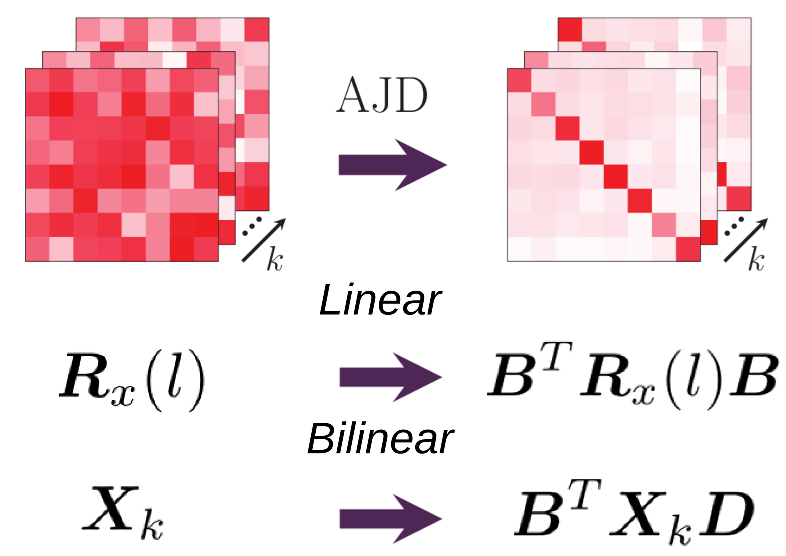

Fig. 3 By minimizing the cost function (4)/(5) we are jointly diagonalizing a set of matrices symmetric or asymmetric for the linear or bilinear model respectively.

TABLE 1 NOTATIONS

A / E	*Spatial/temporal mixing matrix*
B / D	*Spatial/temporal unmixing matrix*
$\boldsymbol{x}(t)$ / X_k	*Observations*
$S(t)$ / S_k	*Sources*
$\boldsymbol{R}_x(l)$	*Symmetric Definite Matrices containing the spatial statistics (covariances and/or cospectra)*
$\boldsymbol{n}(t)$ / N_k	*Noise supposed zero-mean Gaussian*
α	*linear/bilinear ponderation weight*

Gauss Planar Transformation

In order to minimize fc, we use a novel Jacobi-like algorithm inspired from eigenvector Jacobi method and from the Gauss planar elimination method that we call Gauss Planar Transformations (GPT). The transformation is defined such as we want to minimize iteratively for all *k*, the *(ij)*-element of B / D in (3) similarly to [3]

$$\begin{cases} \mathrm{f}_c^{ij}(\beta) = (1-\alpha)\sum_{k=1}^{K}((\boldsymbol{b}_i^T+\beta\boldsymbol{b}_j^T)\boldsymbol{X}_k\boldsymbol{d}_j)^2 \\ \qquad\qquad + 2\alpha\sum_{f=1}^{F}((\boldsymbol{b}_i^T+\beta\boldsymbol{b}_j^T)\boldsymbol{R}_x(l)\boldsymbol{b}_j)^2 \\ \mathrm{f}_c^{ij}(\gamma) = (1-\alpha)\sum_{k=1}^{K}(\boldsymbol{b}_j^T\boldsymbol{X}_k(\boldsymbol{d}_i+\gamma\boldsymbol{d}_j))^2 \end{cases}$$

which has a closed form minimizers:

$$\beta = \frac{(1-\alpha)\sum_{k=1}^{K}(\boldsymbol{b}_i^T\boldsymbol{X}_k\boldsymbol{d}_j)(\boldsymbol{b}_j^T\boldsymbol{X}_k\boldsymbol{d}_j) + 2\alpha\sum_{f=1}^{F}(\boldsymbol{b}_i^T\boldsymbol{R}_x(l)\boldsymbol{b}_j)(\boldsymbol{b}_j^T\boldsymbol{R}_x(l)\boldsymbol{b}_j)}{(1-\alpha)\sum_{k=1}^{K}(\boldsymbol{b}_j^T\boldsymbol{X}_k\boldsymbol{d}_j)^2 + 2\alpha\sum_{f=1}^{F}(\boldsymbol{b}_j^T\boldsymbol{R}_x(l)\boldsymbol{b}_j)^2}$$

$$\gamma = \frac{\sum_{k=1}^{K}(\boldsymbol{b}_j^T\boldsymbol{X}_k\boldsymbol{d}_i)(\boldsymbol{b}_j^T\boldsymbol{X}_k\boldsymbol{d}_j)}{\sum_{k=1}^{K}(\boldsymbol{b}_j^T\boldsymbol{X}_k\boldsymbol{d}_j)^2}.$$

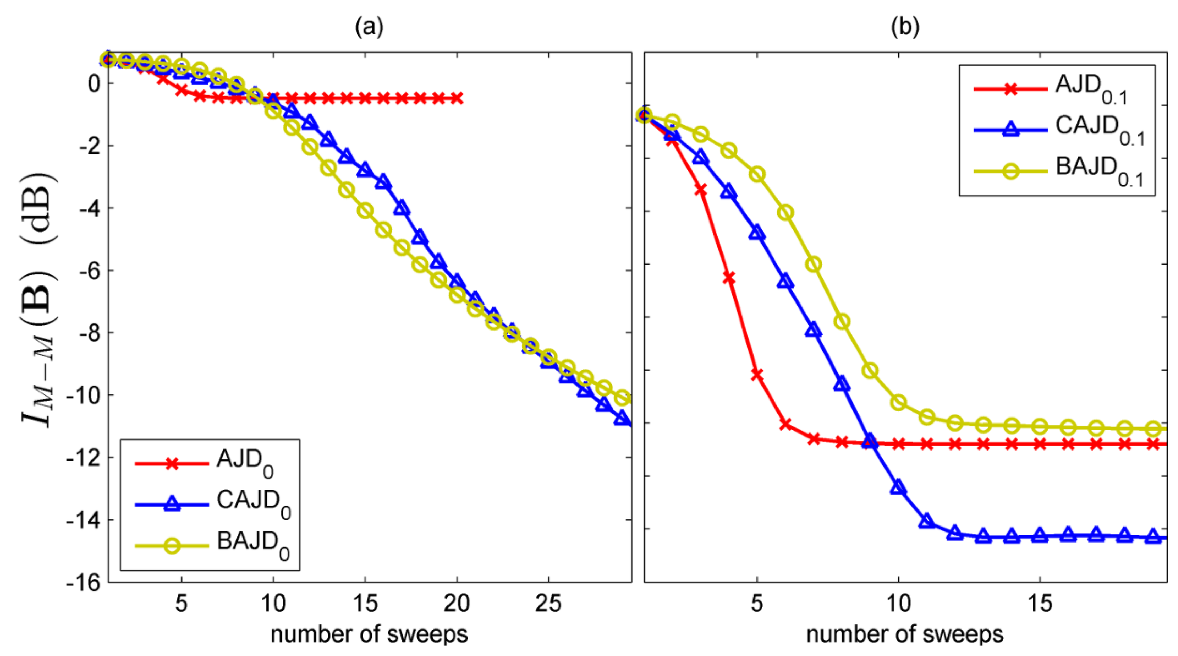

Fig. 4 Moreau-Macchi criterion [4] on the matrix B. Average performance of GPT algorithms on 100 independent realizations (a) without noise (σ = 0) and (b) with noise (σ = 0.1)

TABLE 2 MOREAU-MACCHI CRITERION [4] AFTER CONVERGENCE. 100 INDEPENDENT DRAWS OF MATRICES A AND E WITH σ = 0.1.

		median	10% quantile	90% quantile
B	CAJD	$1.06e{-2}$	$4.79e{-2}$	$2.69e{-2}$
	BAJD	$3.12e{-1}$	$1.21e{-1}$	$7.09e{-1}$
	AJD	$3.26e{-1}$	$2.87e{-1}$	$3.78e{-1}$
D	CAJD	$4.26e{-1}$	$1.24e{-1}$	$8.74e{-1}$
	BAJD	$4.17e{-1}$	$1.28e{-1}$	$8.81e{-1}$

Results

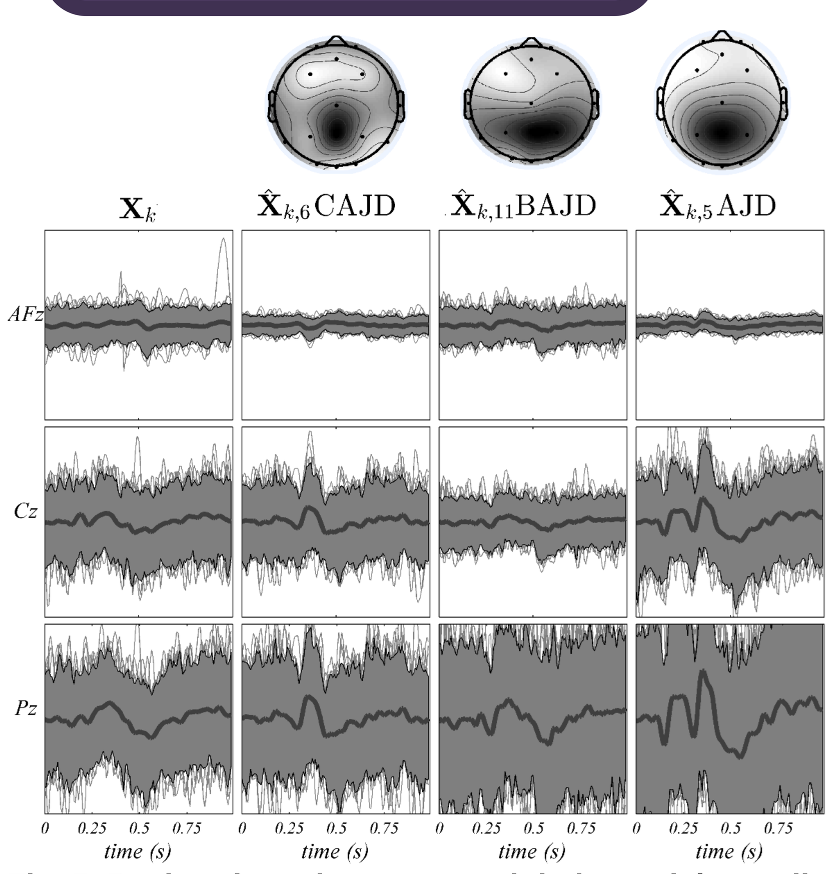

Fig. 5 Backprojected sources and their spatial contribution. Black line: ensemble average. Grey area: 10-90% quantiles. CAJD shows a smaller variability at the single trial level compared to the linear and bilinear models alone.

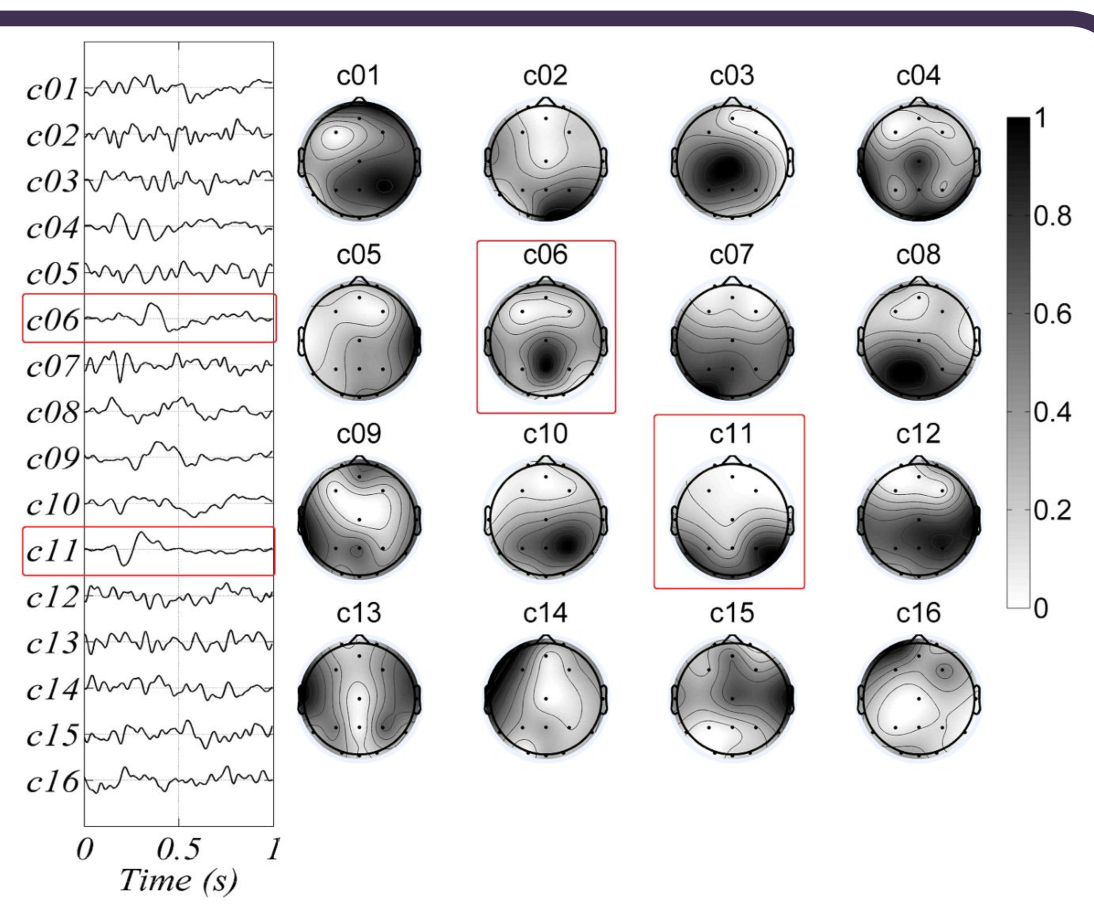

Fig. 6 Left: columns of the estimated temporal mixing matrix. Right: columns of the estimated spatial mixing matrix projected on a topographic scalp map. Highlight : Source c06 corresponds to P3b, source c11 corresponds to early visual complex [5].

Conclusion

Composite AJD

-the use of both models improve the extractions of sources by exploiting the temporal structure while being flexible on the common spatial unmixing matrix.

-compatible with the use of several diversity (non-stationarity/spectral coloration) that can help the separation.

Future work

-apply for BCI classification

-assess the quality of the estimated EEG sources compared to others BSS methods (such as ICA).

References

[1] P. Comon and C. Jutten. Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press, February 2010.
[2] L. Korczowski, A. Barachant, A. Andreev, C. Jutten, and M. Congedo. Brain Invaders 2 : an open source Plug & Play multi-user BCI videogame. In 6th International Brain-Computer Interface Meeting, pages 224, Pacific Grove, CA, United States, May 2016. BCI Society.
[3] V. Maurandi and E. Moreau. A Decoupled Jacobi-Like Algorithm for Non-Unitary Joint Diagonalization of Complex-Valued Matrices. IEEE Signal Processing Letters, 21(12):1453–1456, December 2014.
[4] E. Moreau and O. Macchi. New self-adaptative algorithms for source separation based on contrast functions. In , IEEE Signal Processing Workshop on Higher-Order Statistics, 1993, pages 215–219, 1993.
[5] J. Polich. Updating P300: An Integrative Theory of P3a and P3b. Clin Neurophysiol, 118(10):2128–2148, October 2007.

This research is partly funded by the European Research Council project CHESS 2012-ERC-AdG-320684

Grenoble Images Parole Signal Automatique
UMR CNRS 5216 – Grenoble Campus
38400 Saint Martin d'Hères - FRANCE