Spatial blind source separation

François Bachoc

Institut de Mathématiques de Toulouse
Université Paul Sabatier

Joint work with **Marc Genton** (KAUST), **Klaus Nordhausen** (Vienna), **Anne Ruiz-Gazen** (Toulouse) and **Joni Virta** (Turku)

CMStatistics 2019
London
December 2019

Outline

1. The spatial blind source separation problem
2. A solution by co-diagonalization of two local covariance matrices
3. An improved solution by approximate diagonalization of several local covariance matrices
4. Asymptotic results
5. Numerical results

Mixing of independent sources

Consider p unobserved independent stationary random fields

- $Z_1 : \mathbb{R}^d \to \mathbb{R}$
- $\vdots$
- $Z_p : \mathbb{R}^d \to \mathbb{R}$

called the sources.

Assume that we observe the mixed random fields

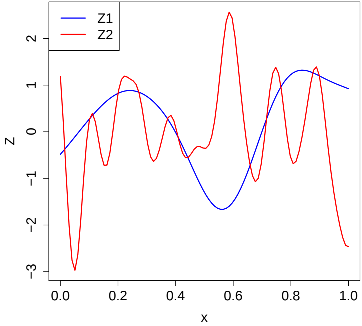
- $X_1 : \mathbb{R}^d \to \mathbb{R}$
- $\vdots$
- $X_p : \mathbb{R}^d \to \mathbb{R}$

with

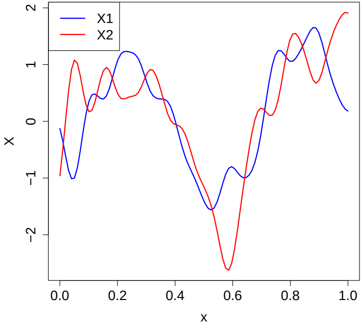
$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \Omega \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix}$$

where Ω is the $p \times p$ unknown mixing matrix.

Illustration (d=1)



Unobserved source fields Z_1, Z_2.



Observed mixed fields X_1, X_2.

Here

$$\Omega = \begin{pmatrix} 1 & 0.3 \\ 1 & -0.4 \end{pmatrix}.$$

Application examples

- Sound signal registered at p sensors $\longrightarrow$ we want to recover p speakers ($d = 1$, signal processing).
- p pollutant concentrations measured over a region $\longrightarrow$ we want to recover p main independent sources of pollution ($d = 2$, spatial statistics).
- Determining main drivers for time series ($d = 1$, finance).
- Recovering neuron sources in EEGs ($d = 1$, neurosciences).

A reference:

Comin, P. & Jutten, C., Handbook of Blind Source Separation: Independent component analysis and applications, *Academic press*, 2010.

Objective

$\Longrightarrow$ Knowing the unmixing matrix Ω^{-1} would be useful.

- Recovery of the independent sources with

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix} = \Omega^{-1} \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}.$$

- Interpretation of the independent sources by subject experts.
- Modeling the distribution of $(X_1, \dots, X_p)$ (complex) $\Longrightarrow$ modeling independently the distributions of $Z_1, \dots, Z_p$ (simpler).
- Predicting $X_1, \dots, X_p$ by multivariate Kriging (cost $O(p^3 n^3)$) $\Longrightarrow$ predicting independently $Z_1, \dots, Z_p$ by univariate Kriging (cost $O(pn^3)$).

$\Longrightarrow$ We want to estimate Ω^{-1}.

Identifiability aspects

- In

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \Omega \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix},$$

  the observed $X_1, \ldots, X_p$ are unchanged if
  - column i of Ω multiplied by $\lambda > 0$,
  - Z_i multiplied by $1/\lambda$.

  $\Longrightarrow$ We assume that

$$\mathrm{Var}(Z_1(s)) = 1, \ldots, \mathrm{Var}(Z_p(s)) = 1$$

  for $s \in \mathbb{R}^d$.

- Still now
  - Z_i can not be distinguished from $-Z_i$,
  - the order of $Z_1, \ldots, Z_p$ can not be estimated.

  $\Longrightarrow$ We want to estimate $Z_1, \ldots, Z_p$ up to signs and order of the components.

  $\Longrightarrow$ We want to estimate Ω^{-1} up to signs and order of the rows.

Observations and local covariance matrices

- **Observations:** We observe $X_1, \ldots, X_p$ at the observation points

$$s_1, \ldots, s_n \in \mathbb{R}^d.$$

  Our observations are thus
  - $X_1(s_1), \ldots, X_1(s_n)$
  - $\vdots$
  - $X_p(s_1), \ldots, X_p(s_n)$.
- **Local covariance matrices:**
  - let $f : \mathbb{R}^d \to \mathbb{R}$ be a *kernel*,
  - let

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix},$$

  - let

$$\widehat{M}(f) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} f(s_i - s_j) X(s_i) X(s_j)^\top$$

  $(p \times p)$
  (assume $X_1, \ldots, X_p$ centered for simplicity).

Different types of kernels

- Let $f_0(s) = \mathbf{1}\{s = 0\}$.
  $\Longrightarrow$ We have

$$\widehat{M}(f_0) = \frac{1}{n} \sum_{i=1}^{n} X(s_i) X(s_i)^\top$$

  (empirical covariance matrix).
- Ball kernel:

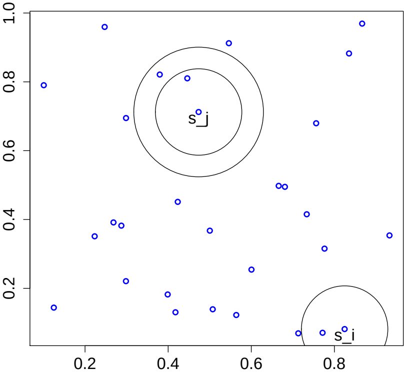
$$f(s) = \mathbf{1}\{||s|| \leq h\}.$$

- Ring kernel:

$$f(s) = \mathbf{1}\{h_1 \leq ||s|| \leq h_2\}.$$

- Gaussian kernel:

$$f(s) = e^{-||s||^2/h^2}.$$



Co-diagonalization

Unmixing matrix estimator

Estimator $\widehat{\Gamma}(f)$ by co-diagonalization of $\widehat{M}(f_0)$ and $\widehat{M}(f)$:

$$\widehat{\Gamma}(f)\widehat{M}(f_0)\widehat{\Gamma}(f)^\top = I_p$$

and

$$\widehat{\Gamma}(f)\widehat{M}(f)\widehat{\Gamma}(f)^\top = \widehat{\Lambda}(f),$$

where $\widehat{\Lambda}(f)$ is a diagonal matrix.

- $\widehat{\Gamma}(f)$ estimates Ω^{-1}.
- **Intuition:** Can show that $\widehat{\Gamma}(f) = \Omega^{-1}$ would make the above matrices diagonal in expectation.
- Similar method exists for independent observations and time series ($d = 1$) (see e.g. Belouchrani et a. 1997).
- Method suggested in the spatial setting ($d \geq 2$) in Nordhausen et al (2015).

Co-diagonalization: pros and cons

+ $\widehat{\Gamma}(f)$ can be computed explicitly by diagonalization of

$$\widehat{M}(f_0)^{-1/2}\widehat{M}(f)\widehat{M}(f_0)^{-1/2}$$

$(p \times p)$.

+ No need to model the random fields $X_1, \ldots, X_p$ (the estimator is semi-parametric).

- The estimation quality strongly depends on the choice of f.

Approximate diagonalization

Consider k kernels $f_1, \ldots, f_k : \mathbb{R}^d \to \mathbb{R}$.

Unmixing matrix estimator

Estimator $\widehat{\Gamma}(f_1, \ldots, f_k) = \widehat{\Gamma}$ satisfies

$$\widehat{\Gamma} \in \underset{\substack{\Gamma: \\ \Gamma \widehat{M}(f_0) \Gamma^\top = I_p}}{\operatorname{argmax}} \sum_{l=1}^{k} \sum_{j=1}^{p} \left[\left(\Gamma \widehat{M}(f_l) \Gamma^\top \right)_{j,j} \right]^2 . \tag{1}$$

- $\widehat{\Gamma}(f)$ estimates Ω^{-1}.
- **Intuition:** Same principle as before but we want all the matrices

  $$\hat{\Gamma} \widehat{M}(f_0) \hat{\Gamma}^\top, \hat{\Gamma} \widehat{M}(f_1) \hat{\Gamma}^\top, \ldots, \hat{\Gamma} \widehat{M}(f_k) \hat{\Gamma}^\top$$

  to be approximately diagonal.
- Similar method exists for independent observations and time series ($d = 1$) (see e.g. Belouchrani et a. 1997).
- Here we extend to the spatial setting.

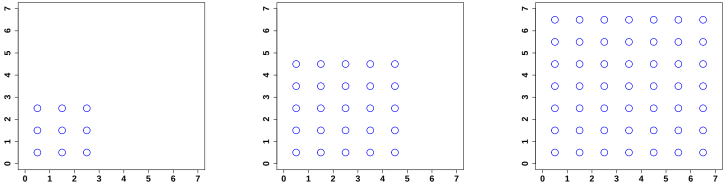
Approximate diagonalization: comments

- No explicit solution of the optimization problem.
- The cost function has complexity $O(kp^3)$.
- Efficient algorithms exist, e.g. Given's rotations (Clarkson, 1988).
- We have more flexibility to choose $f_1, \ldots, f_k$ for a better estimation.
- Typically, a mix of different types of kernels is recommended.

Asymptotic framework

- We let $n \to \infty$ and p be fixed.

Increasing-domain asymptotics: Infinite sequence $(s_i)_{i \in \mathbb{N}}$ of observation locations covering an infinite domain.



$\Longrightarrow$ Asymptotic weak dependence between observations.

Gaussianity: We assume that $Z_1, \ldots, Z_p$ are Gaussian random fields.

- Technical conditions on the covariance functions of $Z_1, \ldots, Z_p$.

Some notation

- Consider kernels $f_1, \dots, f_k$ satisfying some technical conditions (allows balls, rings and Gaussian).
- Let d_w be a distance between probability distributions such that

$$\mathcal{L}_n \xrightarrow[n\to\infty]{d} \mathcal{L}_\infty \iff d_w(\mathcal{L}_n, \mathcal{L}_\infty) \xrightarrow[n\to\infty]{} 0$$

  (Dudley, 2002).
- Let $\mathrm{vect}(A)$ be the column vector obtained by row vectorization of a matrix A.

Central limit theorem

We show: Theorem

- Let $(\hat{\Gamma}_n)$ be any sequence of matrices that approximately diagonalizes

$$\widehat{M}(f_0), \widehat{M}(f_1), \ldots, \widehat{M}(f_k).$$

- Then there exists a sequence $(\check{\Gamma}_n)$ such that for all $n \in \mathbb{N}$

$$\check{\Gamma}_n = \hat{\Gamma}_n$$

  up to order of the rows and multiplication of the rows by ± 1.

- Furthermore, let $\mathcal{L}_n$ be the distribution of

$$\sqrt{n} \, \mathrm{vect} \left(\check{\Gamma}_n - \Omega^{-1} \right).$$

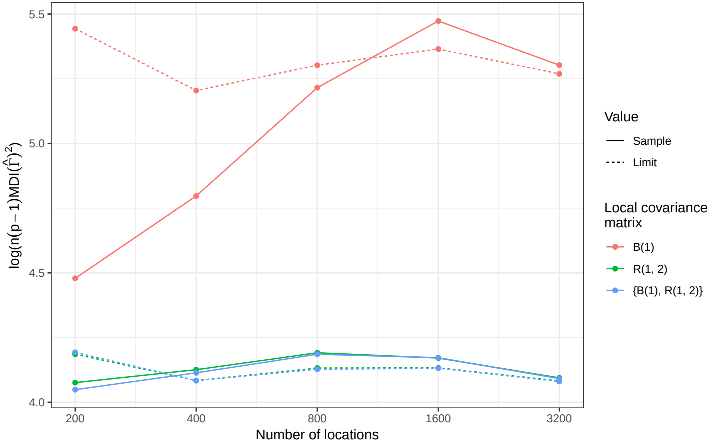
- Then we have

$$d_w \Big(\mathcal{L}_n, \mathcal{N} \left[0, V_n(f_1, \ldots, f_k) \right] \Big) \underset{n \to \infty}{\longrightarrow} 0.$$

- The sequence of matrices $V_n(f_1, \ldots, f_k)$ is bounded. See paper.

Results on simulated data

- y-axis: mean error criterion.

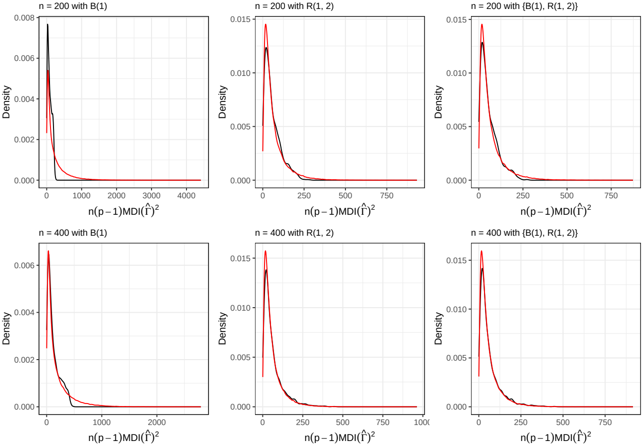


$\Longrightarrow$ As n increases, asymptotic and empirical error criteria get closer.

$\Longrightarrow$ Ring is better than ball. Using both is robust.

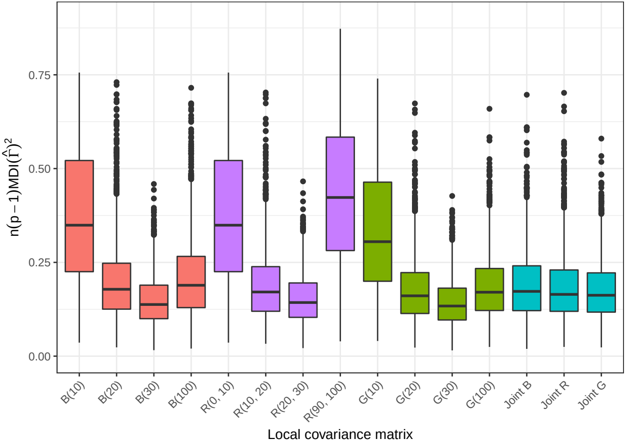
Results on simulated data

- Empirical (**black**) and asymptotic (red) distributions of error criterion.



Results on simulated data

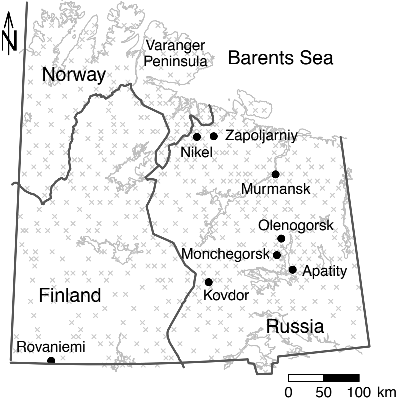
- *x*-axis: Ball (B), ring (R), Gaussian (G) and joint kernels.
- *y*-axis: mean error criterion.



$\Longrightarrow$ Using combinations of kernels is robust.

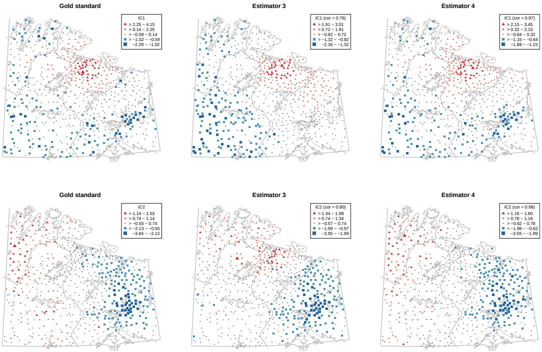
Real data example

- $n = 594$ samples of terrestrial moss in Finland, Norway, Russia.
- $p = 31$ concentrations of chemical elements.
- (Nordhausen et al, 2015).



Real data example

- **Left**, gold standard: 2 most important estimated sources in Z by
  - co-diagonalization of $\widehat{M}(f_0)$ and $\widehat{M}(f_1)$,
  - f_1 is the ball kernel with radius $50km$,
  - chosen carefully by hand with a subject expert.
- **Middle:** f_0 and f_1; ball kernel with radius $100km$.
- **Right:** f_0 and f_1, f_2, f_3; ring kernels with varying radii.



Conclusion

- Unmixing the random fields for easier modeling, easier prediction, interpretation.
- Algorithms are semi-parametric and scale well with dataset size.
- Approximate joint diagonalization with multiple kernels is more robust.
- We have extended procedures and asymptotic results from time series to random fields.
- **Multiple open questions:** Fixed-domain asymptotics? Data driven selection of kernels? Dimension reduction?

The paper:

F. Bachoc, M. G. Genton, K. Nordhausen, A. Ruiz-Gazen and J. Virta, Spatial blind source separation, *Biometrika*, forthcoming, 2019. arxiv.org/abs/1812.09187.

Thank you for your attention!

Bibliography

- Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F. & Moulines, E., A blind source separation technique using second-order statistics, *IEEE Transactions on Signal Processing*, 45, 434-444, 1997.
- Nordhausen, K., Oja, H., Filzmoser, P. & Reimann, C., Blind source separation for spatially correlated compositional data, *Mathematical Geosciences*, 47, 753-770, 2015.
- Clarkson, D. B., Algorithm AS 211: The F-G Diagonalization Algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37, 147-151, 1988.
- Dudley, D. B., Real Analysis and Probability, *Cambridge University Press.*, 2002.