Spatial blind source separation

François Bachoc

Institut de Mathématiques de Toulouse
Université Paul Sabatier

Joint work with Marc Genton (KAUST, Saudi Arabia), Klaus Nordhausen (Jyväskylä, Finland), Anne Ruiz-Gazen (Toulouse, France) and Joni Virta (Turku, Finland)

MASCOT NUM 2021 meeting
April 2021
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

\( Z_1 : \mathbb{R}^d \rightarrow \mathbb{R} \)

\( \vdots \)

\( Z_p : \mathbb{R}^d \rightarrow \mathbb{R} \)

called the sources.

Assume that we observe the mixed random fields

\( X_1 : \mathbb{R}^d \rightarrow \mathbb{R} \)

\( \vdots \)

\( X_p : \mathbb{R}^d \rightarrow \mathbb{R} \)

with

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

where \( \Omega \) 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 $\rightarrow$ we want to recover $p$ speakers ($d = 1$, signal processing).
- $p$ pollutant concentrations measured over a region $\rightarrow$ 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:

Objective

Knowing the **unmixing matrix** $\Omega^{-1}$ 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 $(X_1, \ldots, X_p)$ (complex) $\Rightarrow$ modeling independently the distributions of $Z_1, \ldots, Z_p$ (simpler).

- **Predicting** $X_1, \ldots, X_p$ by multivariate Kriging (cost $O(p^3 n^3)$) $\Rightarrow$ predicting independently $Z_1, \ldots, Z_p$ by univariate Kriging (cost $O(p n^3)$) (Muehlmann, Nordhausen, Yi, 2020).

$\Rightarrow$ We want to estimate $\Omega^{-1}$. 
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 \( \Omega \) multiplied by \( \lambda > 0 \),
  - \( Z_i \) multiplied by \( 1/\lambda \).

\[\Rightarrow\] We assume that
  \[
  \text{Var}(Z_1(s)) = 1, \ldots, \text{Var}(Z_p(s)) = 1
  \]
  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.

\[\Rightarrow\] We want to estimate \( Z_1, \ldots, Z_p \) up to signs and order of the components.

\[\Rightarrow\] We want to estimate \( \Omega^{-1} \) up to signs and order of the rows.
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
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 $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a kernel,
  - let
    \[
    X = \begin{pmatrix}
    X_1 \\
    \vdots \\
    X_p
    \end{pmatrix},
    \]
  - let
    \[
    \hat{M}(f) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} f(s_i - s_j)X(s_i)X(s_j)^\top
    \]
    \[(p \times p)\]
    (assume $X_1, \ldots, X_p$ centered for simplicity).
Different types of kernels

- Let $f_0(s) = 1\{s = 0\}$.
  \[\hat{M}(f_0) = \frac{1}{n} \sum_{i=1}^{n} X(s_i)X(s_i)^\top\]
  (empirical covariance matrix).

- **Ball** kernel:
  \[f(s) = 1\{\|s\| \leq h\}\]

- **Ring** kernel:
  \[f(s) = 1\{h_1 \leq \|s\| \leq h_2\}\]

- **Gaussian** kernel:
  \[f(s) = e^{-\|s\|^2/h^2}\]
Co-diagonalization

Unmixing matrix estimator

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

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

and

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

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

- $\hat{\Gamma}(f)$ estimates $\Omega^{-1}$.
- **Intuition:** Can show that $\hat{\Gamma}(f) = \Omega^{-1}$ would make the above matrices diagonal in expectation.
- Similar method exists for independent observations and time series ($d = 1$) (see e.g. Belouchrani et al, 1997).
Co-diagonalization: pros and cons

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

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

$(p \times p)$.

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

### Unmixing matrix estimator

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

\[
\hat{\Gamma} \in \arg\max_{\Gamma} \sum_{l=1}^k \sum_{j=1}^p \left[ \left( \hat{\Gamma} \hat{M}(f_l) \Gamma^\top \right)_{j,j} \right]^2.
\]  

(1)

- \( \hat{\Gamma}(f) \) estimates \( \Omega^{-1} \).
- **Intuition:** Same principle as before but we want all the matrices
  \[
  \hat{\Gamma} \hat{M}(f_0) \hat{\Gamma}^\top, \hat{\Gamma} \hat{M}(f_1) \hat{\Gamma}^\top, \ldots, \hat{\Gamma} \hat{M}(f_k) \hat{\Gamma}^\top
  \]
  to be approximately diagonal.
- Similar method exists for independent observations and time series
  \( (d = 1) \) (see e.g. Belouchrani et al., 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.
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
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.

$\implies$ 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 $Z_1, \ldots, Z_p$. 
Consider kernels $f_1, \ldots, f_k$ satisfying some technical conditions (allows balls, rings and Gaussian).

Let $d_w$ be a distance between probability distributions such that

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

(Dudley, 2002).

Let $\text{vect}(A)$ 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

\[
\hat{M}(f_0), \hat{M}(f_1), \ldots, \hat{M}(f_k).
\]

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

\[
\tilde{\Gamma}_n = \hat{\Gamma}_n
\]

up to order of the rows and multiplication of the rows by \(\pm 1\).

Furthermore, let \(\mathcal{L}_n\) be the distribution of

\[
\sqrt{n} \ \text{vect} \ (\tilde{\Gamma}_n - \Omega^{-1}).
\]

Then we have

\[
d_w(\mathcal{L}_n, \mathcal{N} [0, V_n(f_1, \ldots, f_k)]) \xrightarrow{n \to \infty} 0.
\]

The sequence of matrices \(V_n(f_1, \ldots, f_k)\) is bounded. See paper.
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
Results on simulated data

- y-axis: mean error criterion.

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

Ring is better than ball. Using both is robust.
Results on simulated data

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

---

François Bachoc
Spatial blind source separation
22 / 27
Results on simulated data

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

⇒ 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 $\hat{M}(f_0)$ and $\hat{M}(f_1)$,
  - $f_1$ 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.

![Map of gold standard and estimators](image-url)
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.
- **Follow-up work:** Dimension reduction ([Muehlmann, Bachoc, Nordhausen, Yi, 2020](#)).
- **Open questions:** Fixed-domain asymptotics? Data driven selection of kernels?

The paper:


**Thank you for your attention!**


