A review of second-order blind identification methods

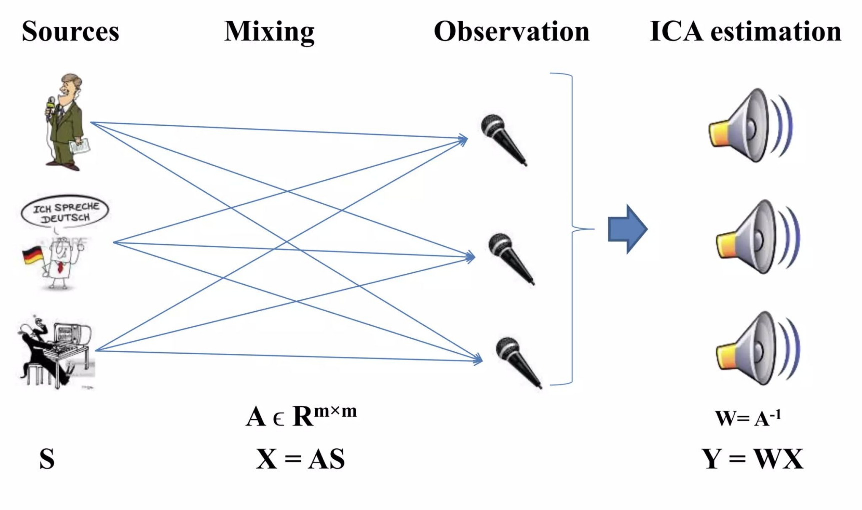
Yan Pan, Markus Matilainen, Sara Taskinen, Klaus Nordhausen

Presenter: Yaorong Xiao

Contents

- BSS, ICA and some popular algorithms
- Go through intuition of SOBI and the algorithm
- Some variant and optimized SOBI methods in recent years
- Application

BSS and ICA



$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

Our goal is to find a transformation matrix $\mathbf{W}$ such that:

$$\mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{A}\mathbf{S}$$

If $\mathbf{W} = \mathbf{A}^{-1}$, we can recover the original sources $\mathbf{S}$.

Essentially, we are seeking a transformation that satisfies:

$$\mathbf{y} = \mathbf{W}\mathbf{X}$$

Where $\mathbf{y}$ is as close to $\mathbf{S}$ as possible.

Some popular algorithms

- Infomax (Bell, A. J., & Sejnowski, T. J. (1995))

$$I(Y,X) = H(Y) - H(Y|X)$$

- FastICA: maximize non-Gaussianity
  - In probability theory, the **central limit theorem** (**CLT**) states that, under appropriate conditions, the distribution of a normalized version of the sample mean converges to a standard normal distribution.

- Joint Diagonalization(JD) based algorithm: **SOBI**, JADE(Joint Approximation Diagonalization of Eigen-matrices), etc..

$$C^*_{ML} = \sum_{l=1}^{L} \mathrm{off}(\hat{R}_l)$$

Order statistics

- First order statistics(Mean):

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Second order statistics(variance):**

$$\text{Variance} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \text{Mean})^2$$

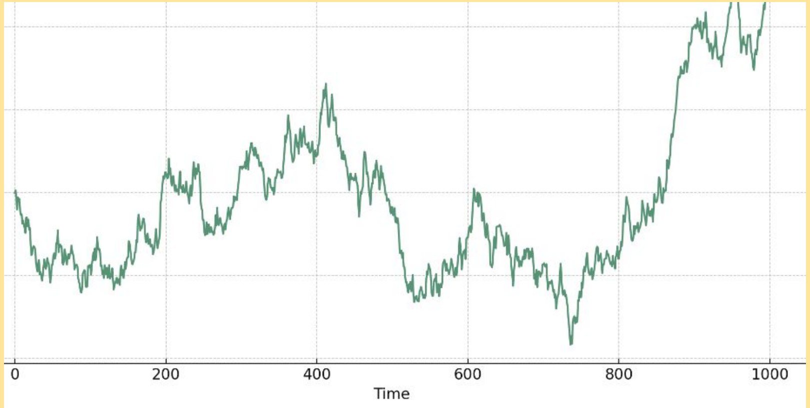
- Higher order statistics:

$$\text{Skew}(X) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3}$$

$$\text{Kurt}(X) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\sigma^4}$$

Intuition

- Assumption of ICA:
    - Source signals are mutually independent.



- Weaker Assumption of SOBI:
    - Sources signals are uncorrelated
    - Unique temporal structure of source signals.
    - a deterministic ergodic sequence

$$E[\mathbf{s}(t+\tau)\mathbf{s}(t)^*] = \mathbf{diag}\left[\rho_1(\tau), \cdots, \rho_n(\tau)\right]$$

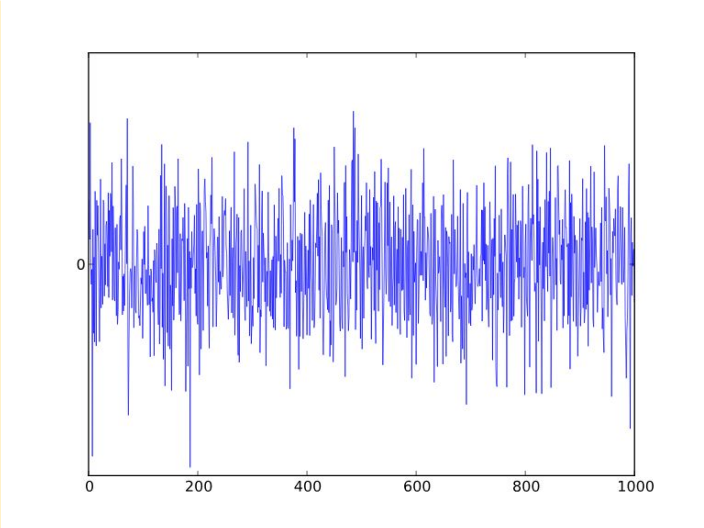
$$\lim_{T\to\infty} T^{-1} \sum_{t=1,T} \mathbf{s}(t+\tau)\mathbf{s}(t)^* \stackrel{\text{def}}{=} E[\mathbf{s}(t+\tau)\mathbf{s}(t)^*] = \mathbf{diag}\left[\rho_1(\tau), \cdots, \rho_n(\tau)\right]$$

SOBI

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

$$\mathbf{R}(0) = E[\mathbf{x}(t)\mathbf{x}^*(t)] = \mathbf{A}\mathbf{R}_s(0)\mathbf{A}^H + \sigma^2\mathbf{I} \qquad (6)$$

$$\mathbf{R}(\tau) = E[\mathbf{x}(t+\tau)\mathbf{x}^*(t)] = \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^H \qquad \tau \neq 0 \quad (7)$$



Whitening

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

$$E[\mathbf{W}\mathbf{y}(t)\mathbf{y}(t)^*\mathbf{W}^H] = \mathbf{W}\mathbf{R}_y(0)\mathbf{W}^H = \mathbf{W}\mathbf{A}\mathbf{A}^H\mathbf{W}^H = \mathbf{I}.$$

Equation (10) shows that if $\mathbf{W}$ is a whitening matrix, then $\mathbf{W}\mathbf{A}$ is a $n \times n$ unitary matrix. It follows that for any whitening matrix $\mathbf{W}$, there exists a $n \times n$ unitary matrix $\mathbf{U}$ such that $\mathbf{W}\mathbf{A} = \mathbf{U}$. As a consequence, matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{W}^{\#}\mathbf{U} \tag{11}$$

$$\mathbf{z}(t) \stackrel{\text{def}}{=} \mathbf{W}\mathbf{x}(t) = \mathbf{W}[\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)] = \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t)$$

Joint Diagonalization(JD)

$$E[\mathbf{s}(t+\tau)\mathbf{s}(t)^*] = \mathbf{diag}\left[\rho_1(\tau), \cdots, \rho_n(\tau)\right]$$

Consider a set $\mathcal{M} = \{\mathbf{M}_I, \cdots, \mathbf{M}_K\}$ of K matrices of size $n \times n$. The "joint diagonality" (JD) criterion is defined, for any $n \times n$ matrix $\mathbf{V}$, as the following nonnegative function of $\mathbf{V}$:

$$\mathcal{C}(\mathcal{M}, \mathbf{V}) \stackrel{\text{def}}{=} \sum_{k=1, K} \mathbf{off}\left(\mathbf{V}^H \mathbf{M}_k \mathbf{V}\right). \tag{20}$$

Ablin, Pierre, Jean-François Cardoso, and Alexandre Gramfort. "Beyond Pham's algorithm for joint diagonalization." *arXiv preprint arXiv:1811.11433* (2018).

Uniqueness of Joint Diagonalization

Theorem 3—Essential Uniqueness of Joint Diagonalization: Let $\mathcal{M} = \{\mathbf{M}_1, \cdots, \mathbf{M}_K\}$ be a set of K matrices where, for $1 \le k \le K$, matrix $\mathbf{M}_k$ is in the form $\mathbf{M}_k = \mathbf{U}\mathbf{D}_k\mathbf{U}^H$ with $\mathbf{U}$ a unitary matrix, and $\mathbf{D}_k = \mathrm{diag}\,[d_1(k), \cdots, d_n(k)]$. Any joint diagonalizer of $\mathcal{M}$ is essentially equal to $\mathbf{U}$ if and only if

$$\forall 1 \le i \ne j \le n \quad \exists k,\ 1 \le k \le K \quad d_i(k) \ne d_j(k). \qquad (21)$$

The sufficiency of (21) is established by proving that any linear combination (with at least two nonzero factors) of the vectors $\mathbf{u}_i$, $i = 1, \cdots, n$ cannot be a common eigenvector of the matrices $\mathbf{M}_k$, $k = 1, \cdots, K$:

Let $\mathbf{v} = \sum_{1 \le i \le n} \alpha_i \mathbf{u}_i$ be a common eigenvector of the matrices $\mathbf{M}_k$, $k = 1, \cdots, K$, and assume, for example, that $\alpha_1 \ne 0$. According to (21), for any index i, $1 < i \le n$, there exists an index k such that $d_1(k) \ne d_i(k)$. For this index k, we have by hypotheses

$$\mathbf{M}_k\mathbf{v} = \lambda_k\mathbf{v} = \sum_{j=1}^{n} \lambda_k \alpha_j \mathbf{u}_j$$

and

$$\mathbf{M}_k\mathbf{v} = \sum_{j=1}^{n} \alpha_j \mathbf{M}_k \mathbf{u}_j = \sum_{j=1}^{n} \alpha_j d_j(k) \mathbf{u}_j.$$

By identification, we have $\alpha_j[d_j(k) - \lambda_k] = 0$ for $1 \le j \le n$. Since $\alpha_1 \ne 0$ and $d_1(k) \ne d_i(k)$, this leads to $\lambda_k = d_1(k)$ and $\alpha_i = 0$. Q.E.D.

Next, we establish the necessity of (21). Assume that there exists a pair (i, j) such that $d_i(k) = d_j(k)$ for $k = 1, \cdots, K$. Then, any linear combination of the vectors $\mathbf{u}_i$ and $\mathbf{u}_j$ is a common eigenvector of the matrices $\mathbf{M}_k$, $k = 1, \cdots, K$. Q.E.D.

Steps of SOBI

1) Estimate the sample covariance $\hat{\mathbf{R}}(0)$ from T data samples. Denote by $\lambda_1, \cdots, \lambda_n$ the n largest eigenvalues and $\mathbf{h}_1, \cdots, \mathbf{h}_n$ the corresponding eigenvectors of $\hat{\mathbf{R}}(0)$.
2) Under the white noise assumption, an estimate $\hat{\sigma}^2$ of the noise variance is the average of the $m - n$ smallest eigenvalues of $\hat{\mathbf{R}}(0)$. The whitened signals are $\mathbf{z}(t) = [z_1(t), \cdots, z_n(t)]^T$, which are computed by $z_i(t) = (\lambda_i - \hat{\sigma}^2)^{-(1/2)}\mathbf{h}_i^*\mathbf{x}(t)$ for $1 \leq i \leq n$. This is equivalent to forming a whitening matrix by

$$\hat{\mathbf{W}} = [(\lambda_1 - \hat{\sigma}^2)^{-(1/2)}\mathbf{h}_1, \cdots, (\lambda_n - \hat{\sigma}^2)^{-(1/2)}\mathbf{h}_n]^H.$$

3) Form sample estimates $\underline{\hat{\mathbf{R}}}(\tau)$ by computing the sample covariance matrices of $\mathbf{z}(t)$ for a fixed set of time lags $\tau \in \{\tau_j | j = 1, \cdots, K\}$.
4) A unitary matrix $\hat{\mathbf{U}}$ is then obtained as joint diagonalizer of the set $\{\underline{\hat{\mathbf{R}}}(\tau_j) | j = 1, \cdots, K\}$.
5) The source signals are estimated as $\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^H\hat{\mathbf{W}}\mathbf{x}(t)$, and/or the mixing matrix $\mathbf{A}$ is estimated as $\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#}\hat{\mathbf{U}}$.

Second Order Blind Identification (SOBI)

- Differences in the spectral density and **autocorrelation function** of the source signals.(1987, Fety)

- Algorithm for multiple unknown signals extraction (AMUSE)(1990)

- SOBI uses multi cross-autocovariance matrix for JD.(1997)

Limitations

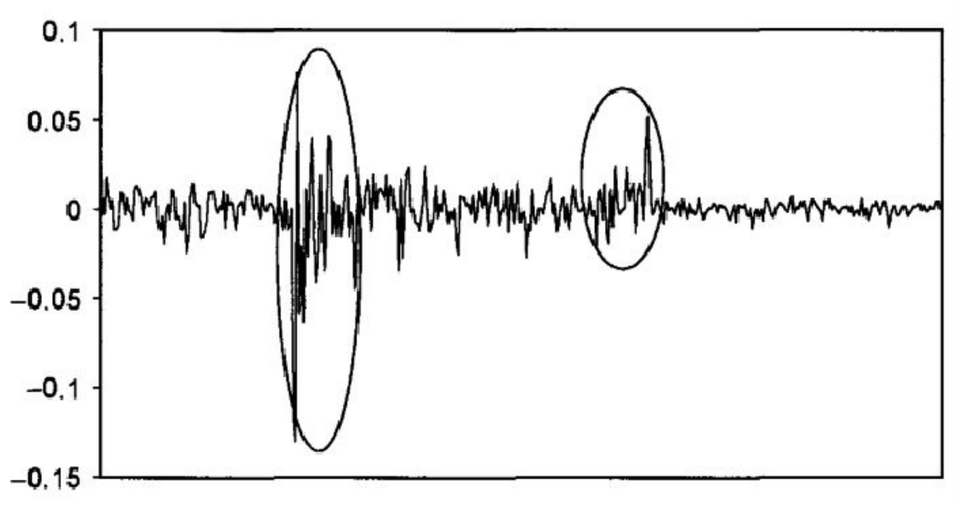
- Selection of time lags.
- Joint diagonalization performance.
- The method does not work well when the volatility cannot be considered as fixed throughout the time series.
- The mixing procedure is assumed to stay constant all the time.
- Multidimensional source signals
- Robustness to outlier

Source signal with volatility clustering, vSOBI

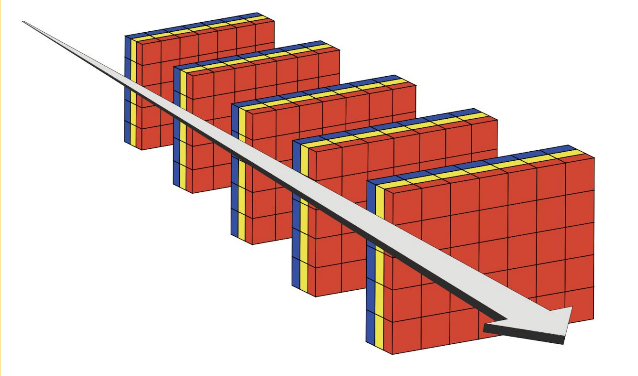
- Stronger assumption(mutual independent)

$$\sum_{\tau \in \mathcal{T}} \sum_{i=1}^{p} \left(\mathrm{E}\left(G\left(\boldsymbol{u}_i' \boldsymbol{x}_{t+\tau}^{st} \right) G\left(\boldsymbol{u}_i' \boldsymbol{x}_t^{st} \right) \right) - \mathrm{E}\left(G\left(\boldsymbol{u}_i' \boldsymbol{x}_t^{st} \right) \right) \mathrm{E}\left(G\left(\boldsymbol{u}_i' \boldsymbol{x}_{t+\tau}^{st} \right) \right) \right)^2,$$

where G is any twice continuously differentiable function.



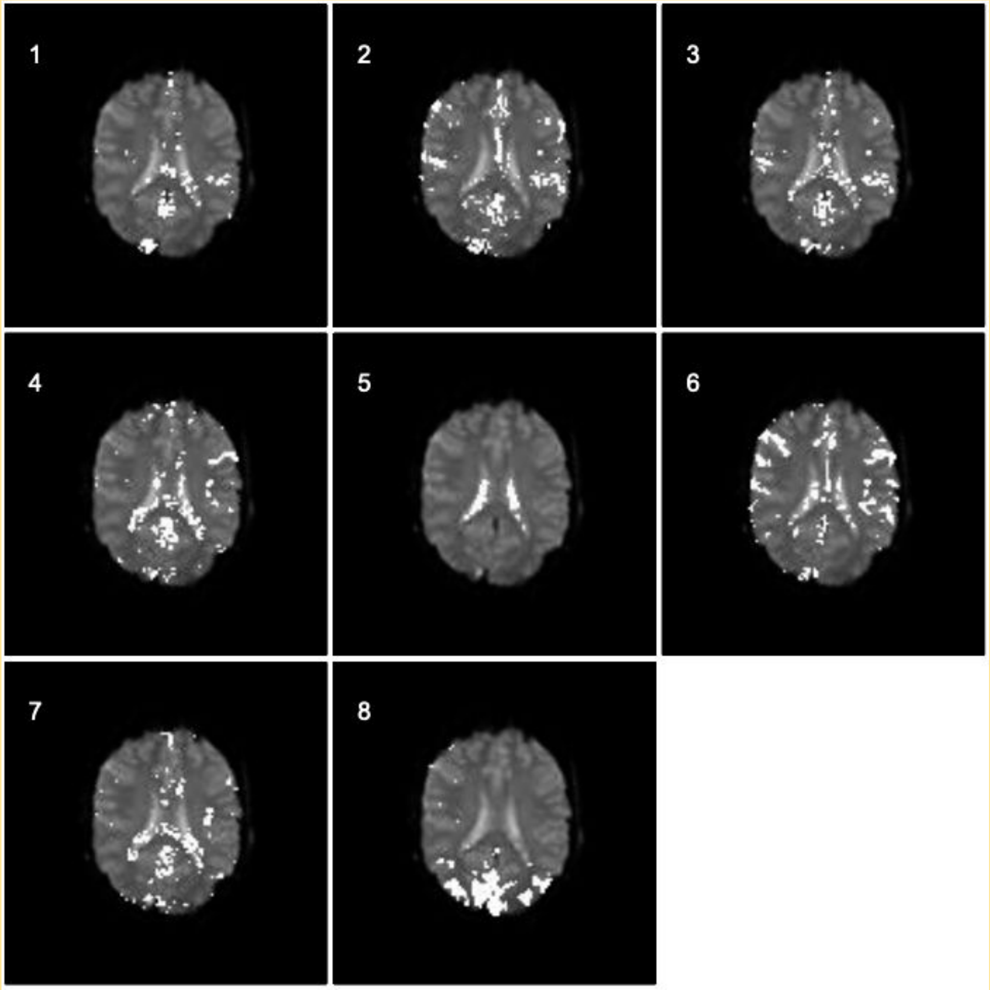
Multidimension SOBI (mdSOBI)



$$\mathrm{Cov}_\tau(\boldsymbol{x_t}) = E\left(\boldsymbol{x}_\mathbf{t}\mathbf{x}'_{\boldsymbol{t}+\tau}\right) = E\left(\boldsymbol{x}_{t_1,\ldots,t_m}\boldsymbol{x}'_{t_1+\tau_1,\ldots,t_m+\tau_m}\right),$$

$$\mathrm{Cov}_\tau^m(\mathbb{X}_t) = \frac{1}{\rho_m}\mathrm{E}\left(\boldsymbol{X}_t^{(m)}\left(\boldsymbol{X}_{t+\tau}^{(m)}\right)'\right)$$

$$\sum_{\tau\in\mathcal{T}^m} \|\,\mathrm{diag}(\boldsymbol{U}\,\mathrm{Cov}_\tau^m\left(\mathbb{X}_t^{st}\right)\boldsymbol{U}'\|^2$$



fMRI data were recorded from six subjects (3 female, 3 male, age 20–37) performing a visual task. In five subjects, five slices with 100 images (TR/TE = 3000/60 msec) were acquired with five periods of rest and five photic simulation periods with rest. Simulation and rest periods comprised 10 repetitions each, i.e. 30s. Resolution was $3 \times 3 \times 4$ mm.

Robust SOBI

Replace autocovariance matrix as a **sign** autocovariance。SAM-SOBI

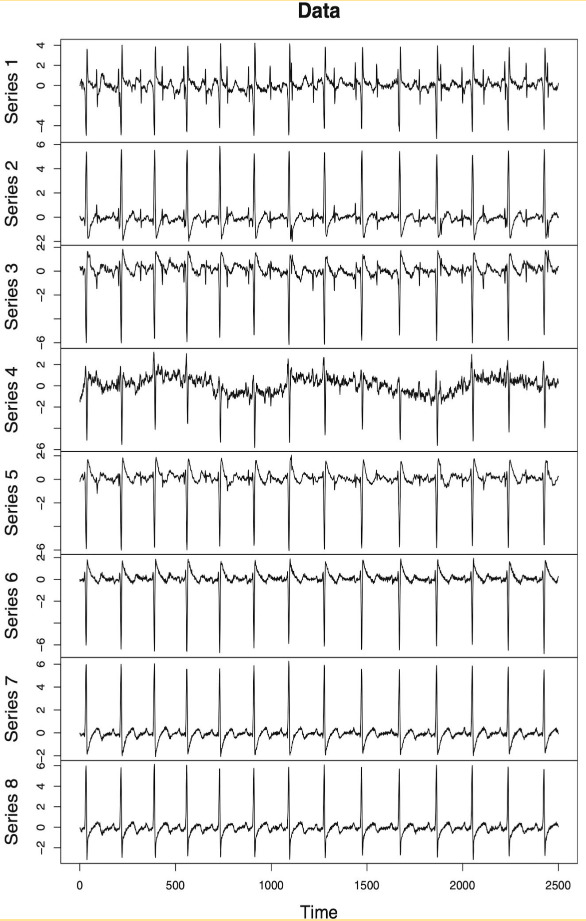
$$\boldsymbol{x}_t^{st} = \mathrm{SCov}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_S), \tag{20}$$

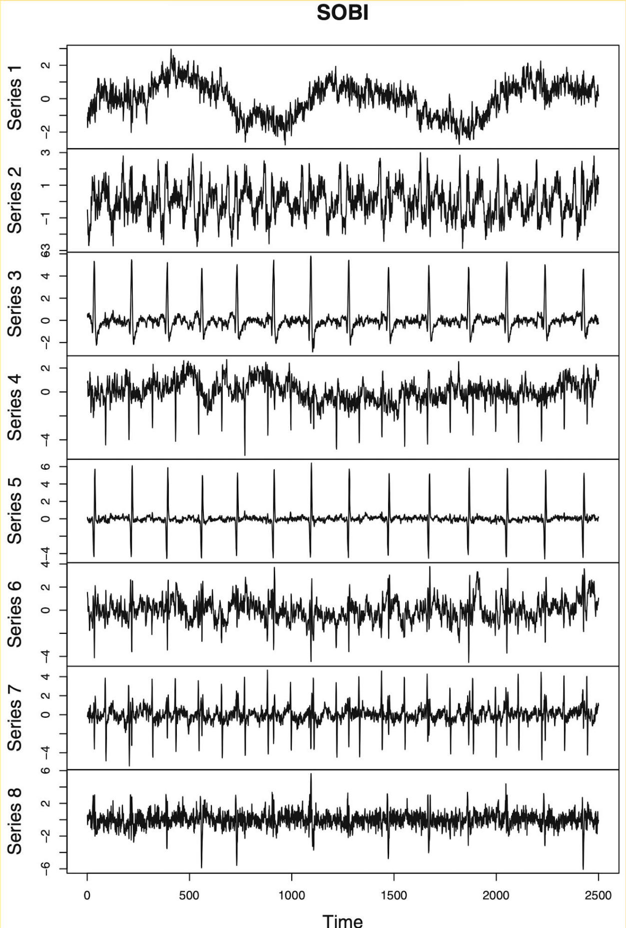
$$\mathrm{SCov}_\tau = E\left(\frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|}\frac{\boldsymbol{x}'_{t+\tau}}{\|\boldsymbol{x}_{t+\tau}\|}\right)$$

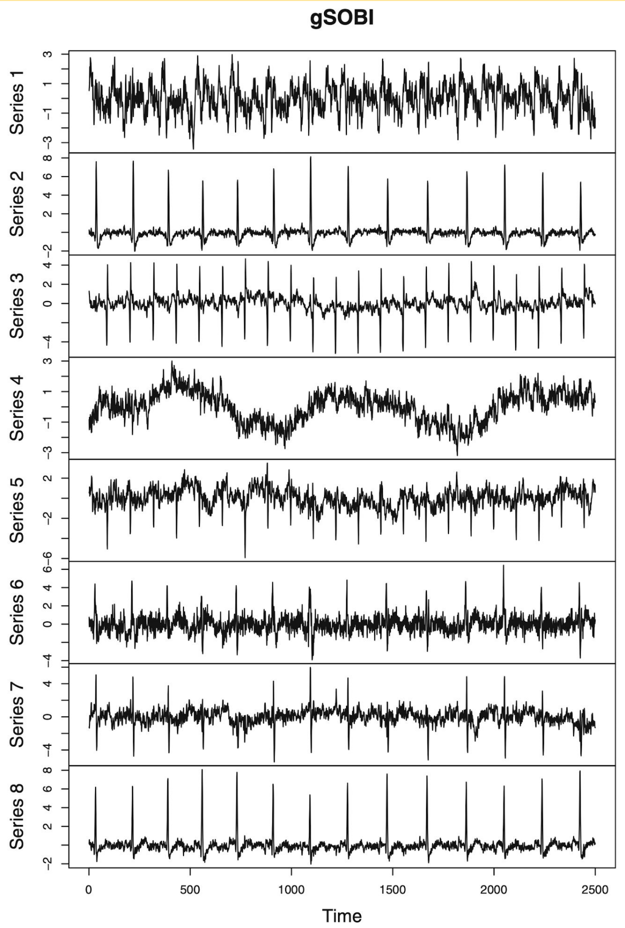
Ilmonen et al. (2015) proposed an eSAM-SOBI method,

the sample mean and the sample covariance matrix in (20) are replaced by the Hettmansperger –Randles estimates of location and scatter (Hettmansperger & Randles, 2002)

Application on ECG








THANK YOU
HAPPY
FRIDAY