Machine Intelligence 2

2.3 Second Order Blind Source Separation

Prof. Dr. Klaus Obermayer

Fachgebiet Neuronale Informationsverarbeitung (NI)

SS 2017

Second Order Blind Source Separation

Second Order Blind Source Separation

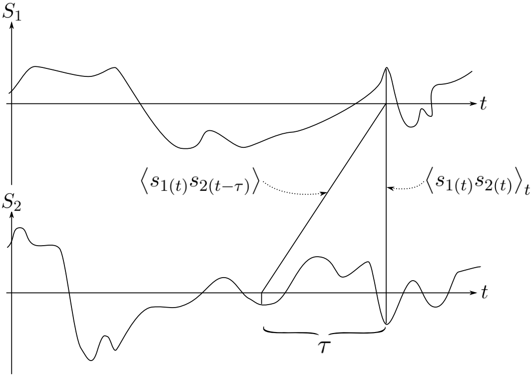
Problem: Noisy observations

- $\rightsquigarrow$ Estimation of higher order moments is unreliable.
- $\rightsquigarrow$ Can we do it with the estimation of second order moments (correlations) only?
- $\rightsquigarrow$ Yes: but only under the assumption of finite length autocorrelations.

Typical examples: Time series, images, videos, etc.

Observations: $\underline{\mathbf{x}}_{(t)}$, recorded at different times (from $\underline{\mathbf{x}} = \underline{\mathbf{A}} \cdot \underline{\mathbf{s}}$)

Example: Two sources & two time shifts



statistical independence $\overset{?}{\underset{\Rightarrow}{\leftarrow}}$ cross-correlation function vanishes

Separation: find unmixing matrix $\underline{\mathbf{W}}$, s.t. both cross-correlations (absolute values) are minimized

Step 1: PCA and sphering

(a) "centering" of the data: $< \underline{\mathbf{x}} >= \underline{\mathbf{0}}$

(b) solve the eigenvalue problem

$$\underline{\mathbf{C}}_{\mathrm{x}}^{(0)} \underline{\mathbf{e}}_k = \lambda_k \underline{\mathbf{e}}_k \quad \text{with} \quad \Big[\overbrace{\underline{\mathbf{C}}_{\mathrm{x}}^{(0)}}^{\tau=0}\Big]_{ij} = \Big\langle \mathrm{x}_{i(t)} \mathrm{x}_{j(t)} \Big\rangle_t$$

(c) transformation into the eigenbasis and sphering

$$\underline{\mathbf{u}} = \underline{\mathbf{M}}_0 \cdot \underline{\mathbf{x}}$$

$$\underline{\mathbf{M}}_0 = \underbrace{\Lambda_0^{-\frac{1}{2}}}_{\substack{\text{diagonal matrix of} \\ \text{inverse eigenvalues}}} \cdot \underbrace{\underline{\mathbf{E}}_0^T}_{\substack{\text{matrix of} \\ \text{eigenvectors}}}$$

Step 2: Rotation of eigenvectors

- $\rightsquigarrow$ ansatz: $\underbrace{\underline{\mathbf{s}}}_{\substack{\text{true sources}\\\text{(independent)}}} = \underline{\mathbf{B}} \cdot \underline{\mathbf{u}}$
- $\rightsquigarrow$ for statistically independent sources we obtain

$$\left\langle s_{i(t)} s_{j(t)} \right\rangle_t = \sum_{k,l=1}^{N} B_{ik} \underbrace{\left\langle u_{k(t)} u_{l(t)} \right\rangle_t}_{\stackrel{!}{=}\delta_{kl}:(\textit{cf. sphering})} B_{lj}^T$$

$$= \sum_{k=1}^{N} B_{ik} B_{kj}^T \stackrel{!}{=} \delta_{ij}$$

$$\underline{\mathbf{B}} \cdot \underline{\mathbf{B}}^T = \underline{\mathbf{1}} \rightsquigarrow \text{orthogonal transformation}$$

Source separation requires an additional orthogonal transformation.

Step 3: Determination of B

Approach: Determination of $\underline{\mathbf{B}}$ through diagonalization of a time-shifted cross-correlation matrix $\underline{\mathbf{C}}_u^{(\tau)}$:

(a) solve the eigenvalue problem

$$\underline{\mathbf{C}}_u^{(\tau)} \underline{\mathbf{e}}_k = \lambda_k \underline{\mathbf{e}}_k \text{ with } \left[\underline{\mathbf{C}}_u^{(\tau)}\right]_{ij} = \left\langle u_{i(t)} u_{j(t-\tau)} \right\rangle_t$$

(b) transformation into the eigenbasis via $\underline{\mathbf{E}}_\tau (= \underline{\mathbf{B}})$

$$\widehat{\underline{\mathbf{s}}} = \underbrace{\underline{\mathbf{E}}_\tau^T}_{\substack{\text{matrix of} \\ \text{eigenvectors}}} \underline{\mathbf{u}} = \underbrace{\underline{\mathbf{E}}_\tau \underline{\mathbf{\Lambda}}_0^{-1/2} \underline{\mathbf{E}}_0^T}_{\underline{\mathbf{W}}} \underline{\mathbf{x}}$$

Molgedey, L., and Schuster, H. G. (1994). Separation of a mixture of independent signals using time delayed correlations. Phys. Rev. Lett. 72:3634 - 3637.

Making it practical

Noisy data: only approximate independence

$\Rightarrow$ approach can be extended to design **noise robust algorithms**

- minimizing sensor noise by omitting $\tau = 0$
- (approximate) joint diagonalization of multiple cross-correlation matrices

$$\left[\underline{\mathbf{C}}_{\mathrm{x}}^{(\tau)}\right]_{ij} = \frac{1}{T}\sum_{t=0}^{T-1} \mathrm{x}_{i(t)}\mathrm{x}_{j(t-\tau)}$$

Two examples: QDIAG & FFDIAG

Algorithm 1: QDIAG (Vollgraf & Obermayer, 2006)

Cost function: squared sum of all non-diagonal elements

$$E^T_{[\underline{\mathbf{W}}]} = \sum_{\tau} \underbrace{\alpha_\tau}_{\text{weighting factors}} \sum_{i \neq j} \left(\underline{\mathbf{W}} \cdot \underline{\mathbf{C}}_{\mathrm{x}}^{(\tau)} \cdot \underline{\mathbf{W}}^T \right)^2_{ij}$$

Optimization problem: constrained minimization of cross-correlations

$$E^T_{[\underline{\mathbf{W}}]} \overset{!}{=} \min$$

$$\left(\underline{\mathbf{W}} \cdot \underline{\mathbf{C}}_{\mathrm{x}}^{(0)} \cdot \underline{\mathbf{W}}^T \right)_{ii} = 1 \text{ for all } i \rightarrow \quad \text{avoid trivial solutions}$$

- two versions: $O(k \cdot N^3)$ or $O(N^5)$, k: no. of shifts, N: no. of sources
- allows for arbitrary (rectangular) matrices $\underline{\mathbf{W}}$
- code and docu: `https://www.ni.tu-berlin.de/menue/software_and_data/approximate_simultaneous_matrix_diagonalization_qdiag/`

Algorithm 2: FFDIAG (Ziehe et al., 2004)

Cost function: equally weighted squared sum

$$E^T_{[\underline{\mathbf{W}}]} = \sum_{\tau} \sum_{i \neq j} \left(\underline{\mathbf{W}} \cdot \underline{\mathbf{C}}_{\mathrm{x}}^{(\tau)} \cdot \underline{\mathbf{W}}^T \right)^2_{ij}$$

Optimization problem: constrained minimization

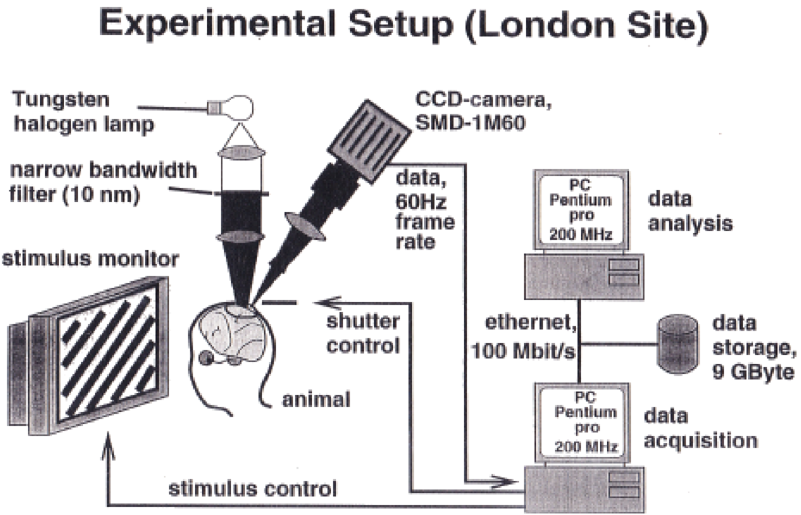
$E^T_{[\underline{\mathbf{W}}]} \overset{!}{=} \min$ minimize cross-correlation

invertability of $\underline{\mathbf{W}}$ constraint to avoid trivial solutions

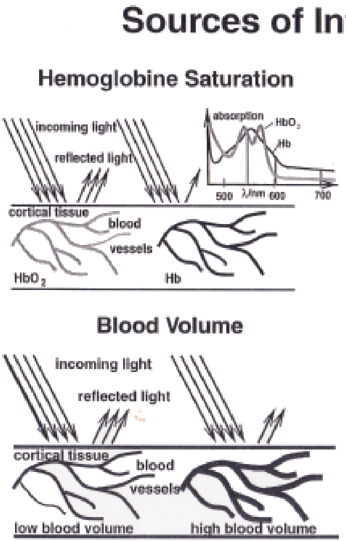
- computational complexity: $O(k \cdot N^2)$, approaching $O(N^3)$ for large N, k: no. of shifts, N: no. of sources
- requires square matrices $\underline{\mathbf{W}}$, no weighting
- code: `http://www.user.tu-berlin.de/aziehe/code/`

Application to optical imaging of brain activity: Stetter et al. (2000)

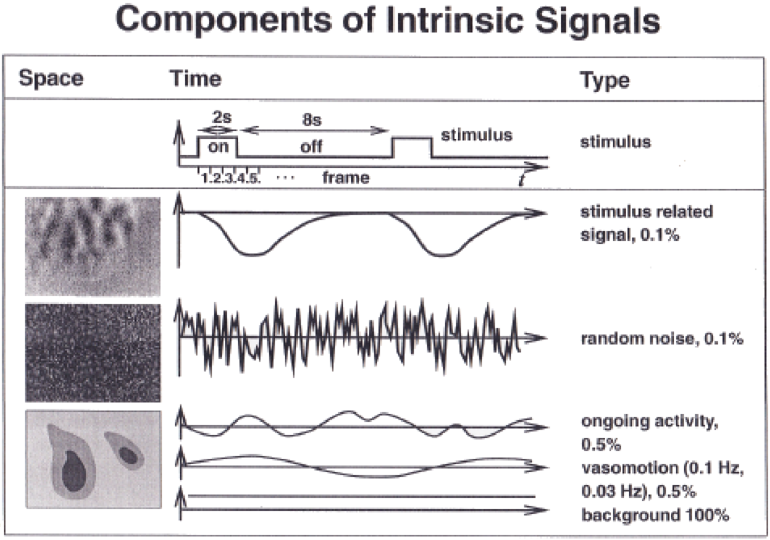
Example: Optical imaging



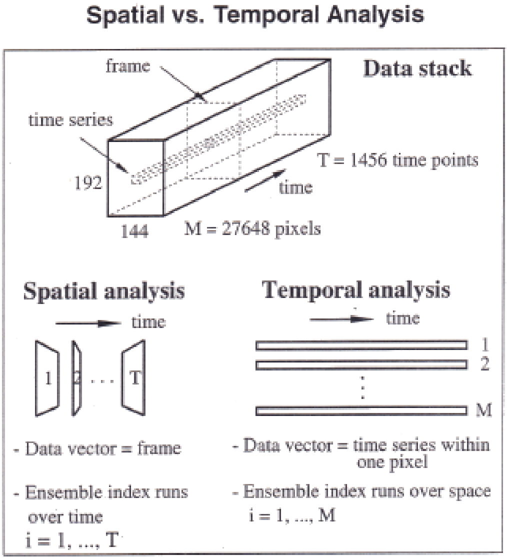
Example: Optical imaging



Example: Optical imaging

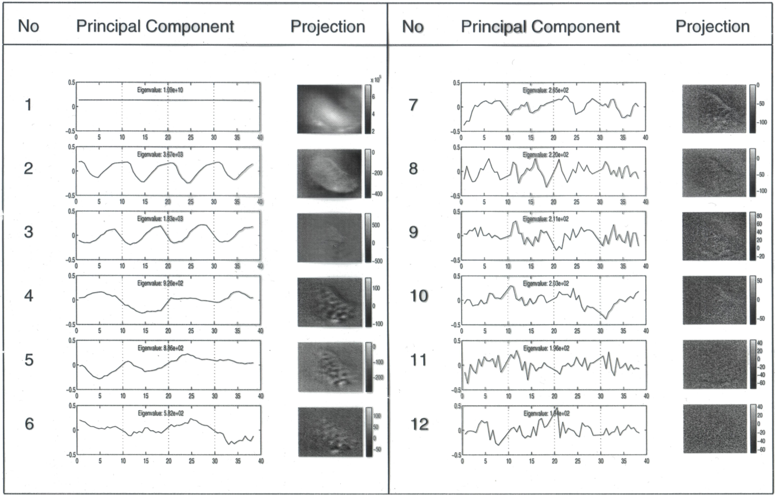


Example: Optical imaging



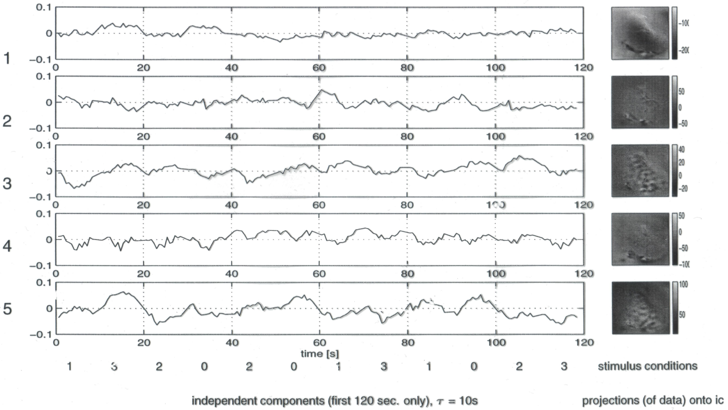
Example: Optical imaging

PCA on Averaged Single Condition Image Series:
Temporal Principal Components and Data Projections



Example: Optical imaging

Source Separation Using the Molgedey-Schuster Algorithm



Source data: Principal components No. 11-15 from long continuous image time series (1456 frames)
processing: Molgedey-Schuster source separation with $\tau = 10$sec