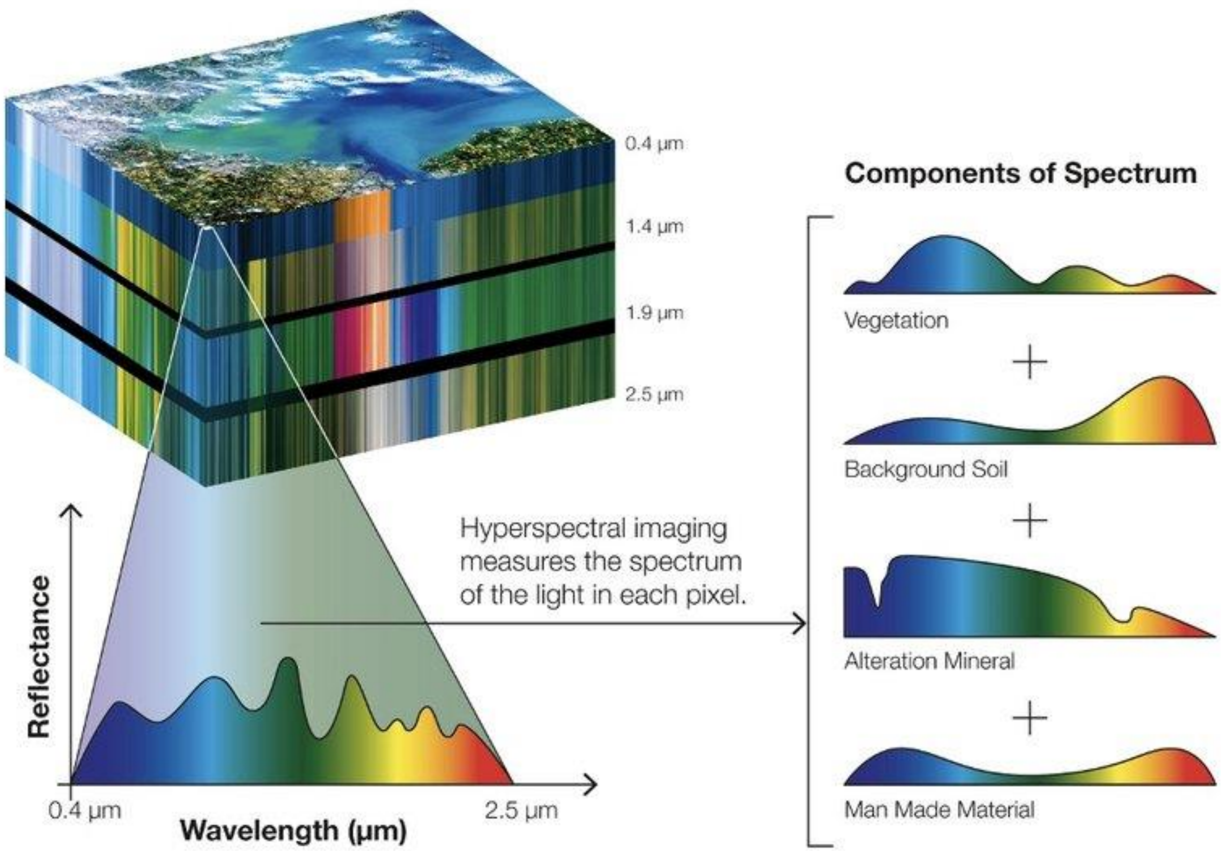
Computational Imaging and Spectroscopy: Blind source separation

Thierry SOREZE
DTU July 2024

DTU Fotonik
Department of Photonics Engineering

Blind source separation (BSS)



Blind source separation (BSS)

Spectral unmixing with Non-Negative Matrix factorization

$$\mathbf{X} = \mathbf{QA} + \mathbf{E}$$

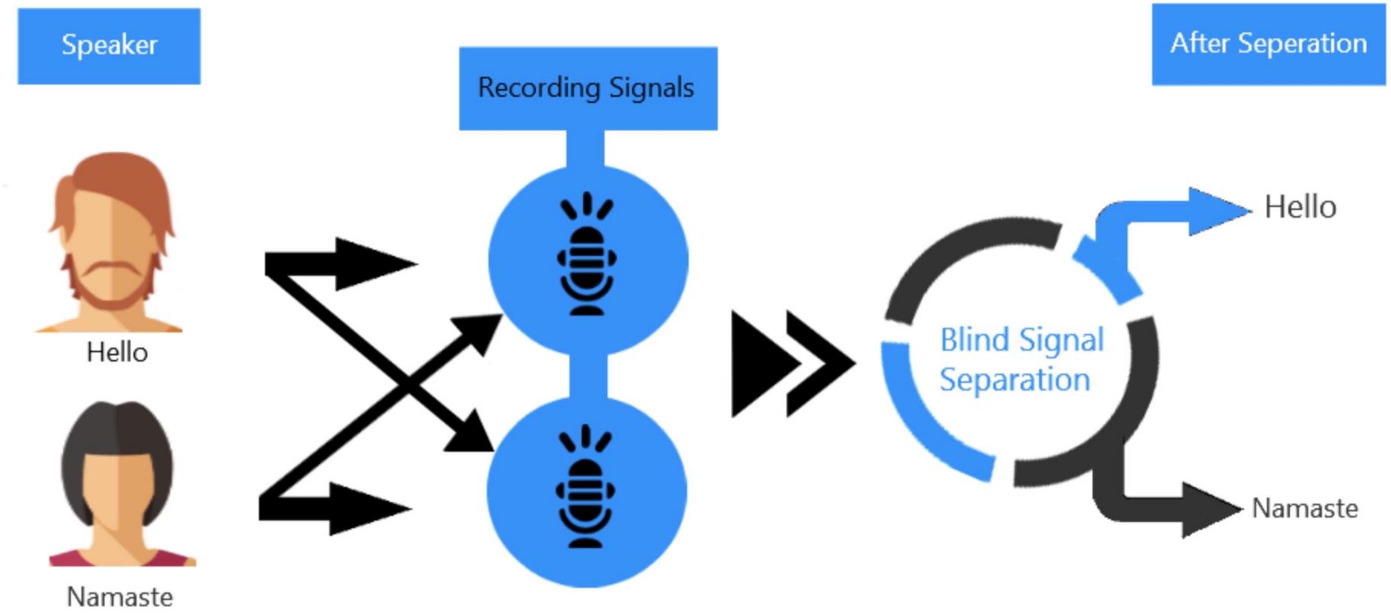
$$\min_{\mathbf{A}<0,\mathbf{Q}<0} \|\mathbf{X} - \mathbf{QA}\|_2^2$$

$$\min_{\mathbf{A}<0,\mathbf{Q}<0} \|\mathbf{X} - \mathbf{QA}\|_2^2 + \boldsymbol{\lambda}\|\mathbf{A}\|_{\boldsymbol{q}} \quad \mathbf{0} < \boldsymbol{q} < \mathbf{1}$$

$$\mathbf{Q} \leftarrow \mathbf{Q}.* \, \mathbf{XA}^{\mathrm{T}}./\mathbf{QAA}^{\mathrm{T}}$$

$$\mathbf{A} \leftarrow \mathbf{A}.* \, \mathbf{Q}^{\mathrm{T}}\mathbf{X}./\left(\mathbf{Q}^{\mathrm{T}}\mathbf{QA} + \frac{\boldsymbol{\lambda}}{\mathbf{2}}\mathbf{Q}^{\boldsymbol{q}-\mathbf{1}}\right)$$

Blind source separation (BSS)



Blind source separation (BSS)

Introduction

In the **BSS** setting, we assume that we are given N_c observations or channels $(y_1, \cdots, y_{N_c})$ where each y_i is a vector of length N. Each measurement is a linear mixture of N_s vectors $(s_1, \cdots, s_{N_s})$, called sources having the same length N

$$y_i[l] = \sum_{j=1}^{N_s} \mathbf{A}[i,j] s_j[l] + \varepsilon_i[l], \qquad \forall i \in [1, \cdots, N_c], \qquad \forall l \in [1, \cdots, N]$$

$\mathbf{A}$ is a $N_c \times N_s$ mixing matrix whose columns will be denoted a_i. In matrix form the problem reads:

$$\mathbf{Y} = \mathbf{AS} + \mathbf{E}$$

Blind source separation (BSS)

Introduction

- $\mathbf{Y}$ is the $N_c \times N$ measurements matrix whose rows are $y_i^T, i = 1, \cdots, N_c$ (observed data)

- $\mathbf{S}$ is the $N_s \times N$ source matrix with rows $s_i^T, i = 1, \cdots, N_s$

- $\mathbf{E}$ is the $N_c \times N$ noise matrix with rows ε_i^T,

Both $\boldsymbol{S}$ and $\boldsymbol{A}$ are unknown

Blind source separation (BSS)

Independent Component Analysis (ICA)

Here we consider the noiseless case $\mathbf{Y} = \boldsymbol{AS}$, which can be written in the following form:

$$\mathbf{Y} = \sum_{i=1}^{N_S} \mathbf{Y}^{(i)} = \sum_{i=1}^{N_S} a_i s_i^{\mathrm{T}}$$

$\mathbf{Y}^{(i)}$ is the contribution of the sources s_i to the data $\mathbf{Y}$

BSS is equivalent to decomposing the matrix $\mathbf{Y}$ of rank N_s into a sum of N_s rank-one matrices $\{\mathbf{Y}^{(i)} = a_i s_i^{\mathrm{T}}\}$ $i = 1, \cdots, N_s$

Blind source separation (BSS)

Independent Component Analysis (ICA)

Let's assume that the sources are random vectors, which are supposed to be known a priori different, and there therefore decorrelated.

We would in this context look to compute their covariance matrix $\mathbf{\Sigma_S}$ expected to be diagonal.

Unfortunately, $\boldsymbol{\Sigma_S}$ is not invariant to orthonormal transformations such as rotation. We then need to go beyond decorrelation.

Blind source separation (BSS)

Independent Component Analysis (ICA)

In the **ICA** framework the sources are assumed to be independent random variables. This holds if their joint pdf is the product of each pdfs

$$\mathrm{pdf}_S = \prod_{i=1}^{N_S} \mathrm{pdf}_{s_i}(s_i)$$

The **ICA** looks for a demixing matrix $\mathbf{B}$ such as the estimated sources, $\tilde{\mathbf{S}} = \mathbf{BAS}$ are independent. We assume the mixing matrix $\mathbf{A}$ to be square and invertible.

Blind source separation (BSS)

Independent Component Analysis (ICA)

Independence

The Kullback-Leibler (**KL**) divergence between two pdf is defined as:

$$\mathbf{KL}\left(\mathrm{pdf}_1 \middle\| \mathrm{pdf}_2\right) = \int_u \mathrm{pdf}_1(\mathbf{u}) \log\left(\frac{\mathrm{pdf}_1(\mathbf{u})}{\mathrm{pdf}_2(\mathbf{u})}\right) d\mathbf{u}$$

The mutual information (**MI**) in the form of the **KL** divergence between the joint density pdf_S and the product of the marginal densities, pdf_{S_i}, is given by

$$\mathbf{MI}(\mathbf{S}) = \mathbf{KL}\left(\mathrm{pdf}_1 \middle\| \prod_{i=1}^{N_S} \mathrm{pdf}_{S_i}\right)$$

Blind source separation (BSS)

Morphospectral Diversity

Let assume that $\mathbf{A} = \left[\varphi_{v,1}, \cdots, \varphi_{v,N_c}\right] \in \mathbb{R}^{N_c \times N_s}$ is a known *spectral* dictionary, and $\boldsymbol{\Phi} = [\varphi_1, \cdots, \varphi_T] \in \mathbb{R}^{N \times T}$ is a *spatial* or *temporal* dictionary

We assume that each source s_i can be represented as a sparse linear decomposition of atoms of $\boldsymbol{\Phi}$; $s_i = \boldsymbol{\Phi}\alpha_i$

Let $\boldsymbol{\alpha}$ be the $N_s \times \mathrm{T}$ matrix whose rows are α_i^T, we can then write the multichannel noiseless data $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\alpha}\boldsymbol{\Phi}^{\mathbf{T}} = \sum_{i=1}^{\boldsymbol{N_s}} \sum_{j=1}^{T} \left(\varphi_{v,i}, \varphi_j^T\right) \alpha_i[j]$$

Each column in $\mathbf{Y}$ reads: $\mathrm{vect}(\mathbf{Y}) = (\mathbf{A} \otimes \boldsymbol{\Phi})\mathrm{vect}(\boldsymbol{\alpha})$

Blind source separation (BSS)

Multichannel sparse decomposition

We suppose $\mathbf{\Psi}$ to be overcomplete. We wish to recover the sparsest solution $\boldsymbol{\alpha}$ from $\mathbf{Y}$ by solving

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{N_s\times T}} \sum_{i=1}^{N_s} \|\alpha_i\|_0 \qquad \text{s.t} \qquad \mathbf{Y} = \mathbf{A}\boldsymbol{\alpha}\mathbf{\Phi}^{\mathbf{T}}$$

Or its relaxed $\ell_1$form

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{N_s\times T}} \sum_{i=1}^{N_s} \|\alpha_i\|_1 \qquad \text{s.t} \qquad \mathbf{Y} = \mathbf{A}\boldsymbol{\alpha}\mathbf{\Phi}^{\mathbf{T}}$$

Blind source separation (BSS)

Generalized MCA

Here we will assume morphological diversity as a source of contrast. We assume that the sources are sparse in a given spatial dictionary $\mathbf{\Phi}$, that is the concatenation of K orthonormal bases $(\mathbf{\Phi}_K)_{k=1,\cdots,K}$ (for formality)

The Generalized **MCA** (**GMCA**) framework assumes a prior that each source is modeled as the linear combination of K morphological components, where each component is sparse in a specific basis

$$\forall i \in \{1, \cdots, N_s\}; \quad s_i = \sum_{k=1}^{K} x_{i.k} = \sum_{k=1}^{K} \mathbf{\Phi}_k \alpha_{i,k}$$

Where $\alpha_i = \left[\alpha_{i,1}^{\mathrm{T}}, \cdots, \alpha_{i,K}^{\mathrm{T}}\right]^{\mathrm{T}}$

Blind source separation (BSS)

Generalized MCA

The **GMCA** seeks an unmixing scheme by estimating $\mathbf{A}$ **,** leading to the sparsest sources $\boldsymbol{S}$ in the dictionary $\mathbf{\Phi}$**.** This is expressed as the following optimization problem

$$\min_{A,\alpha_{1,1},\cdots,\alpha_{N_S,K}} \frac{1}{2}\left\|\mathbf{A}\boldsymbol{\alpha}\mathbf{\Phi}^{\mathbf{T}}\right\|_{\mathrm{F}}^{2} + \lambda \sum_{i=1}^{N_S}\sum_{k=1}^{K}\left\|\alpha_{i,k}\right\|_{p}^{p}$$

Subject to

$$\|\alpha_i\|_2 = 1, \qquad \forall i \in [1,\cdots,N_s]$$

Blind source separation (BSS)

Generalized MCA

Define the $(i,k)th$ multichannel marginal residual by

$$\mathbf{R}_{i,k} = \mathbf{Y} - \sum_{i' \neq i} \sum_{k' \neq k} \alpha_{i'} x_{i',k'}^{\mathrm{T}}$$

Given some assumptions, we can set our optimization problem as a component-wise optimization problem

$$\min \frac{1}{2} \left\| \mathbf{R}_{i,k} - \left(\alpha_i \alpha_{i,k}^{\mathrm{T}} \right) \mathbf{\Phi}^{\mathrm{T}} \right\|_{\mathrm{F}}^{2} + \lambda \left\| \alpha_{i,k} \right\|_{p}^{p}$$

Blind source separation (BSS)

Generalized MCA

The closed form estimation of the morphological components $x_{i,k}$ is

$$\tilde{x}_{i,k} = \Delta_{\mathbf{\Phi}_{k,\lambda'}}\left(\frac{1}{\|\alpha_i\|_2^2}\right)\mathbf{R}_{i,k}^{\mathrm{T}}\alpha_i$$

Note: $\Delta_{D,\lambda}(x)$ is hard thresholding the transformed coefficients in the dictionary D, then reconstructing from the remaining coefficients, with a threshold λ

$$\tilde{\alpha}_i = \frac{1}{\|s_i\|_2^2}\left(\mathbf{Y} - \sum_{i' \neq i} \alpha_i s_{i'}^{\mathrm{T}}\right)s_i$$

Where $s_i = \sum_{k=1}^{K} x_{i,k}$

Blind source separation (BSS)

Generalized MCA algorithm 1/2

Task: Blind source separation

Parameters: The data $\mathbf{Y}$, the dictionary $\mathbf{\Phi}$, number of iterations, N_{iter} number of sources, N_s number of channels N_c, stopping threshold λ_{min}, threshold update schedule

Initialization: $x_{i,k}^{0} = 0$ for all (i, k), $\boldsymbol{A}^{(0)}$ random and threshold λ_0

Blind source separation (BSS)

Generalized MCA algorithm 2/2

For $t = 1\ to\ N_{iter}$ **do**

For $i = 1\ to\ N_s$ **do**

For $k = 1\ to\ K$ **do**

Compute the marginal residuals

$$\mathbf{R}_{i,k}^{(t)} = \mathbf{Y} - \sum_{(i',k')\neq(i,k)} a_{\boldsymbol{i'}}^{(t-1)}\ x_{i',k'}^{(t-1)^{\mathbf{T}}}$$

Estimate the current component $x_{\boldsymbol{i}}^{(t)}$ via thresholding λ_t

$$x_{(\boldsymbol{i}\ \ ,\boldsymbol{k})}^{(t)} = \Delta_{\boldsymbol{\Phi_k},\lambda_t}\left(\mathbf{R}_{i,k}^{t} a_{\boldsymbol{i}}^{(t-1)}\ \right)$$

Update ith source $s_i = \sum_{k=1}^{K} x_{ik}^{(t)}$

Update $a_{\boldsymbol{i}}^{(t)}$ assuming $a_{\boldsymbol{i\neq i'}}^{(t)}$ and the morphological components $x_{i,k}^{(t)}$ are fixed:

$$a_{\boldsymbol{i}}^{(t)} = \frac{\mathbf{1}}{\left\|\boldsymbol{s}_{\boldsymbol{i}}^{(t)}\right\|_2^2}\left(\mathbf{Y} - \sum_{i\neq i'}^{N_s} a_{i'}^{(t-1)} s_{\boldsymbol{i'}}^{(t)^{\mathbf{T}}}\right)\boldsymbol{s}_{\boldsymbol{i}}^{(\boldsymbol{t})}$$

Udpate the threshold λ_t according to the given schedule

If $\lambda_t \leq \lambda_{min}$ **then** **stop**

Morphological component Analysis (MCA)

MCA algorithm

Threshold update strategy:

Prefixed decreasing threshold $\lambda_t = \lambda_0 - t(\lambda_0 - \lambda_{min})/N_{\text{iter}}$, $\lambda_0 = \max_{k \neq k^*} \|\mathbf{T}_k y\|_\infty$

Adaptive strategies:

- ❑ MAD
- ❑ MOM : $\lambda_t = \frac{1}{2}\left(\left\|\mathbf{T}_1 r^{(t)}\right\|_\infty + \left\|\mathbf{T}_2 r^{(t)}\right\|_\infty\right)$

Blind source separation (BSS)

Generalized MCA algorithm

Application 1: inpaint Barbara with GMCA

Apply a mask **M** to Barbara
Inpaint Barbara using GMCA
Ns = 3, Nc = 3

Compute residuals using

$$\mathbf{R}_{i,k}^{(t)} = \left(\mathbf{Y} - \sum_{(i',k') \neq (i,k)} a_{\boldsymbol{i'}}^{(t-1)} \, x_{i',k'}^{(t-1)^{\mathbf{T}}} \right) \mathbf{M}$$