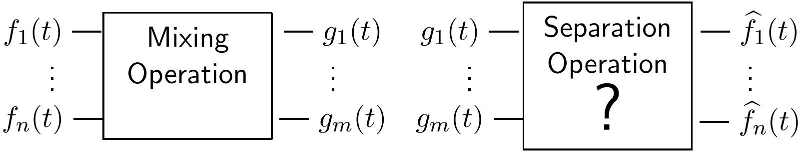
Sources separation methods: An overview

Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes,
UMR8506 CNRS-SUPELEC-UNIV PARIS SUD 11
SUPELEC, 91192 Gif-sur-Yvette, France
http://lss.supelec.free.fr

Email: djafari@lss.supelec.fr
http://djafari.free.fr

Introduction

Mixing models and separation:



- General Linear Mixing Model :

$$\boldsymbol{g}(t) = \int \boldsymbol{A}(t, t')\boldsymbol{f}(t') \, \mathrm{d}t'$$

- Convolutional Mixing Model:

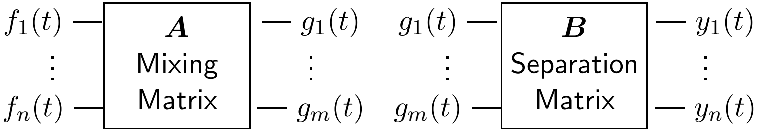
$$\boldsymbol{g}(t) = \int \boldsymbol{A}(t - t')\boldsymbol{f}(t') \, \mathrm{d}t'$$

- Instantaneous Mixing Model:

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t)$$

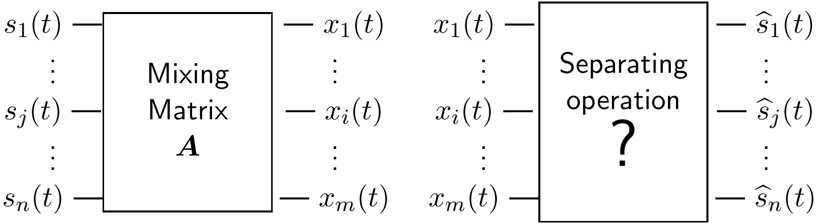
Introduction

Convolutional Mixing $\longrightarrow$ **Multi Chanel Deconvolution**
Instantaneous Mixing $\longrightarrow$ **Source Separation**



- Undeterminations: $\longrightarrow \boldsymbol{B} = \boldsymbol{P}\,\boldsymbol{\Lambda}\,\boldsymbol{A}^{-1}$
  where $\boldsymbol{P}$ is a Permutation matrix and $\boldsymbol{\Lambda}$ a scale (diagonal) matrix.
- Main Hypothesis:: $f_1(t), \ldots, f_n(t)$ are:
  non correlated (PCA) or
  independents (ICA).
- Classical methods : Infomax, Contrast function based, Higher Ordre Sup., Maximum Likelihood, Bayesian Approach

Introduction



Signals:

$$x_i(t) = \sum_{j=1}^{n} A_{i,j} s_j(t) + \epsilon_i(t), \quad t \in \mathcal{T}, \qquad i = 1, \cdots, m$$

Images:

$$x_i(\boldsymbol{r}) = \sum_{j=1}^{n} A_{i,j} s_j(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \ \boldsymbol{r} \in \mathcal{R}, \qquad i = 1, \cdots, m$$

General source separation problem

$$x_i(t) = \sum_{j=1}^{N} A_{ij} s_j(t) + \epsilon_i(t), \quad i = 1, \cdots, M$$

$$\boldsymbol{x}(t) = \boldsymbol{A}\,\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t), \quad t = t_1, \cdots, t_T$$

$$\boldsymbol{X} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} \quad \boldsymbol{E} = \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \vdots \\ \epsilon_M(t) \end{bmatrix} \quad \boldsymbol{S} = \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{bmatrix}$$

$$\boldsymbol{X} = \boldsymbol{A}\,\boldsymbol{S} + \boldsymbol{E}$$

Extension for images

$$x_i(\boldsymbol{r}) = \sum_{j=1}^{N} A_{ij} s_j(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \quad \boldsymbol{r} = (x, y) \in R^2$$

- $\boldsymbol{A}$: Mixing matrix, Loading matrix
- $\boldsymbol{s}(t)$: sources, factors (principales, independent), codebook, ...
- $\boldsymbol{x}(t)$: observations, mixtures, data

General source separation problem

$$g_i(t) = \sum_{j=1}^{N} A_{ij} f_j(t) + \epsilon_i(t), \quad i = 1, \cdots, M$$

$$\boldsymbol{g}(t) = \boldsymbol{A}\, \boldsymbol{f}(t) + \boldsymbol{\epsilon}(t), \quad t = t_1, \cdots, t_T$$

$$\boldsymbol{G} = \begin{bmatrix} g_1(t) \\ g_2(t) \\ \vdots \\ g_M(t) \end{bmatrix} \quad \boldsymbol{E} = \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \vdots \\ \epsilon_M(t) \end{bmatrix} \quad \boldsymbol{F} = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_N(t) \end{bmatrix}$$

$$\boldsymbol{G} = \boldsymbol{A}\, \boldsymbol{F} + \boldsymbol{E}$$

Extension for images

$$g_i(\boldsymbol{r}) = \sum_{j=1}^{N} A_{ij} f_j(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \quad \boldsymbol{r} = (x, y) \in R^2$$

- $\boldsymbol{A}$: Mixing matrix, Loading matrix
- $\boldsymbol{f}(t)$: sources, factors (principales, independent), codebook, ...
- $\boldsymbol{g}(t)$: observations, mixtures, data

General source separation problem

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t), \quad t \in [1, \cdots, T]$$

$$\boldsymbol{g}(\boldsymbol{r}) = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}) + \boldsymbol{\epsilon}(\boldsymbol{r}), \quad \boldsymbol{r} = (x, y) \in R^2$$

- $\boldsymbol{f}$ unknown sources
- $\boldsymbol{A}$ mixing matrix, $\boldsymbol{a}_{*j}$ steering vectors
- $\boldsymbol{g}$ observed signals
- $\boldsymbol{\epsilon}$ represents the errors of modeling and measurement

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} \longrightarrow g_i = \sum_j a_{ij} f_j \longrightarrow \boldsymbol{g} = \sum_j \boldsymbol{a}_{*j} f_j$$

$$\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} f_1 & 0 & f_2 & 0 \\ 0 & f_1 & 0 & f_2 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix}$$

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} = \boldsymbol{F}\boldsymbol{a} \quad \text{with} \quad \boldsymbol{F} = \boldsymbol{f} \odot \boldsymbol{I}, \quad \boldsymbol{a} = \text{vec}(\boldsymbol{A})$$

- $\boldsymbol{A}$ known, estimation of $\boldsymbol{f}$: $\quad \boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$
- $\boldsymbol{f}$ known, estimation of $\boldsymbol{A}$: $\quad \boldsymbol{g} = \boldsymbol{F}\boldsymbol{a} + \boldsymbol{\epsilon}$
- Joint estimation of $\boldsymbol{f}$ and $\boldsymbol{A}$: $\quad \boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{F}\boldsymbol{a} + \boldsymbol{\epsilon}$

Deterministic methods

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} \quad \text{or} \quad \boldsymbol{G} = \boldsymbol{A}\boldsymbol{F}$$

Matrix factorization, Sources separation, Compressed sensing

- $\boldsymbol{A}$ known, find $\boldsymbol{f}$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{f}\|^2 \right\} = (\boldsymbol{A}'\boldsymbol{A} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}'\boldsymbol{g}$$

- $\boldsymbol{f}$ known, find $\boldsymbol{A}$

$$\widehat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{A}\|^2 \right\} = \boldsymbol{g}\boldsymbol{f}'(\boldsymbol{f}\boldsymbol{f}' + \lambda \boldsymbol{I})^{-1}$$

- Both $\boldsymbol{A}$ and $\boldsymbol{f}$ are unknown

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{A}}) = \arg\min_{(\boldsymbol{f}, \boldsymbol{A})} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda_1 \|\boldsymbol{f}\|^2 + \lambda_2 \|\boldsymbol{A}\|^2 \right\}$$

Alternate optimisation

Deterministic methods

- Both $\boldsymbol{A}$ and $\boldsymbol{f}$ are unknown:

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{A}}) = \arg \min_{(\boldsymbol{f},\boldsymbol{A})} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda_1 \|\boldsymbol{f}\|^2 + \lambda_2 \|\boldsymbol{A}\|^2 \right\}$$

- Undeterminations:
  - Permutation: $\boldsymbol{A}\boldsymbol{P}, \boldsymbol{P}'\boldsymbol{f}$
  - Scale: $k\boldsymbol{A}, \frac{1}{k}\boldsymbol{f}$
- Alternate optimisation

$$\begin{cases} \widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda_1 \|\boldsymbol{f}\|^2 \right\} = (\boldsymbol{A}'\boldsymbol{A} + \lambda_1 \boldsymbol{I})^{-1} \boldsymbol{A}'\boldsymbol{g} \\ \widehat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda_2 \|\boldsymbol{A}\|^2 \right\} = \boldsymbol{g}\boldsymbol{f}'(\boldsymbol{f}\boldsymbol{f}' + \lambda_2 \boldsymbol{I}) \end{cases}$$

- Importance of initialization and other constraintes such as positivity
  - Non-negative Matrix decomposition

Source separation: PCA or ICA approach

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} \longrightarrow \widehat{\boldsymbol{s}} = \boldsymbol{B}\boldsymbol{x}$$

- $\boldsymbol{A}$ mixing matrix
- $\boldsymbol{B}$ separation (demixing) matrix: $\boldsymbol{B} = \boldsymbol{A}^{-1} \longrightarrow \widehat{\boldsymbol{s}} = \boldsymbol{s}$
- Find $\boldsymbol{B}$ such that the components (sources) $\widehat{\boldsymbol{s}}$ be
  - Uncorrelated: Principal Components Analysis (PCA)
  - Independent: Independent Components Analysis (ICA)

- PCA: $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} \longrightarrow \text{cov}[\boldsymbol{x}] = \boldsymbol{A}\text{cov}[\boldsymbol{s}]\boldsymbol{A}'$
  - Estimate $\text{cov}[\boldsymbol{x}] = \frac{1}{T}\sum_t(\boldsymbol{x}(t) - \bar{\boldsymbol{x}})(\boldsymbol{x}'(t) - \bar{\boldsymbol{x}})$
  - SVD decomposition: $\text{cov}[\boldsymbol{x}] = \boldsymbol{U}\Lambda\boldsymbol{U}'$
  - Identify: $\widehat{\boldsymbol{A}} = \boldsymbol{U}$, $\text{cov}[\boldsymbol{s}] = \boldsymbol{\Lambda} \longrightarrow \widehat{\boldsymbol{s}} = \Lambda^{1/2}\widehat{\boldsymbol{A}}'\boldsymbol{x}$
  - Uniqueness ?
    $\widehat{\boldsymbol{A}} = \boldsymbol{R}\boldsymbol{B}$ is also a solution for any rotational matrix $\boldsymbol{R}$.

Source separation: ICA approach

- ICA: Sources are supposed to be independent
- $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} \longrightarrow p_x(\boldsymbol{x}) = |\boldsymbol{A}|-1\, p_s(\boldsymbol{A}\boldsymbol{s})$
- $\boldsymbol{B} = \boldsymbol{A}^{-1} \longrightarrow \widehat{\boldsymbol{s}} = \boldsymbol{B}\boldsymbol{x} = \boldsymbol{s} \longrightarrow p_s(\boldsymbol{B}\boldsymbol{x}) = |\boldsymbol{B}|^{-1}\, p_s(\boldsymbol{x})$

- Independence criteria:
  - Entropy: maximize the entropy $H = -\int p(\boldsymbol{s}) \ln p(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}$
  - Infomax: $KL\left(\prod_j p_j(\hat{s}_j) : p(\widehat{\boldsymbol{s}})\right)$
- $KL\left(\prod_j p_j([\boldsymbol{B}\boldsymbol{x}]_j) : p(\boldsymbol{B}\boldsymbol{x})\right)$ is a function of $\boldsymbol{B}$
- Minimization with respect to $\boldsymbol{B}$ gives ICA algorithms.

General Bayesian source separation problem

$$p(\boldsymbol{f}, \boldsymbol{A} | \boldsymbol{g}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) = \frac{p(\boldsymbol{g} | \boldsymbol{f}, \boldsymbol{A}, \boldsymbol{\theta}_1) \, p(\boldsymbol{f} | \boldsymbol{\theta}_2) \, p(\boldsymbol{A} | \boldsymbol{\theta}_3)}{p(\boldsymbol{g} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)}$$

- $p(\boldsymbol{g} | \boldsymbol{f}, \boldsymbol{A}, \boldsymbol{\theta}_1)$ likelihood
- $p(\boldsymbol{f} | \boldsymbol{\theta}_2)$ and $p(\boldsymbol{A} | \boldsymbol{\theta}_3)$ priors
- $p(\boldsymbol{f}, \boldsymbol{A} | \boldsymbol{g}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ joint posterior
- $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ hyper-parameters

Two approaches:

- Estimate first $\boldsymbol{A}$ and then use it for estimating $\boldsymbol{f}$
- Joint estimation

In real application, we also have to estimate $\boldsymbol{\theta}$:

$$p(\boldsymbol{f}, \boldsymbol{A}, \boldsymbol{\theta} | \boldsymbol{g}) = \frac{p(\boldsymbol{g} | \boldsymbol{f}, \boldsymbol{A}, \boldsymbol{\theta}_1) \, p(\boldsymbol{f} | \boldsymbol{\theta}_2) \, p(\boldsymbol{A} | \boldsymbol{\theta}_3) \, p(\boldsymbol{\theta})}{p(\boldsymbol{g})}$$

Bayesian inference for sources $\boldsymbol{f}$ when $\boldsymbol{A}$ is known

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$$

- Prior knowledge on $\boldsymbol{\epsilon}$:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}|0, v_\epsilon \boldsymbol{I}) \longrightarrow p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{A}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{A}\boldsymbol{f}, v_\epsilon \boldsymbol{I}) \propto \exp\left\{\frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2\right\}$$

- Simple prior models for $\boldsymbol{f}$: $\quad p(\boldsymbol{f}|\alpha) \propto \exp\left\{-\alpha\|\boldsymbol{f}\|^2\right\}$
- Expression of the posterior law:

$$p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{A}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{A})\, p(\boldsymbol{f}) \propto \exp\left\{-\frac{1}{2v_\epsilon} J(\boldsymbol{f})\right\}$$

with $\qquad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2, \quad \lambda = v_\epsilon \alpha$

- Link between MAP estimation and regularization

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{A})\} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$$

- Solution: $\widehat{\boldsymbol{f}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}'\boldsymbol{g}$

Bayesian inference for sources $\boldsymbol{f}$ when $\boldsymbol{A}$ is known

- More general prior model $p(\boldsymbol{f}) \propto \exp\{-\alpha\Omega(\boldsymbol{f})\}$
- MAP:

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\Omega(\boldsymbol{f}), \quad \lambda = v_\epsilon \alpha$$

$$\boxed{p(\boldsymbol{f}|\boldsymbol{\theta}, \boldsymbol{g})} \longrightarrow \boxed{\begin{array}{c}\text{Optimization of} \\ J(\boldsymbol{f}) = \frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \alpha\Omega(\boldsymbol{f})\end{array}} \longrightarrow \widehat{\boldsymbol{f}}$$

- Different priors=Different expressions for $\Omega(\boldsymbol{f})$
- Solution can be obtained using appropriate optimisation algorithm.

MAP estimation with sparsity enforcing priors

- Gaussian: $\Omega(\boldsymbol{f}) = \|\boldsymbol{f}\|^2 = \sum_j |f_j|^2$

$$J(\boldsymbol{f}) = \frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \alpha\|\boldsymbol{f}\|^2 \longrightarrow \widehat{\boldsymbol{f}} = [\boldsymbol{A}'\boldsymbol{A} + \lambda\boldsymbol{I}]^{-1}\boldsymbol{A}'\boldsymbol{g}$$

- Generalized Gaussian:

$$\Omega(\boldsymbol{f}) = \gamma \sum_j |f_j|^\beta)$$

- Student-t model:

$$\Omega(\boldsymbol{f}) = \frac{\nu+1}{2} \sum_j \log\left(1 + f_j^2/\nu\right)$$

- Elastic Net model:

$$\Omega(\boldsymbol{f}) = \sum_j \left[\gamma_1 |f_j| + \gamma_2 f_j^2\right]$$

For an extended list of such sparsity enforcing priors see:
A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," EURASIP Journal on Advances in Signal Processing, vol. Special issue on Sparse Signal Processing, 2012.

Estimation of $\boldsymbol{A}$ when the sources $\boldsymbol{f}$ are known

Source separation is a bilinear model:

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} = \boldsymbol{F}\boldsymbol{a} = \boldsymbol{A}\boldsymbol{f}$$

$$\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} f_1 & 0 & f_2 & 0 \\ 0 & f_1 & 0 & f_2 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix}$$

$$\boldsymbol{F} = \boldsymbol{f} \odot \boldsymbol{I}, \quad \boldsymbol{a} = \text{vec}(\boldsymbol{A})$$

- Problem is more ill-posed (underdetermined).
- We need absolutely to impose constraintes on elements or the structure of $\boldsymbol{A}$, for example:
  - Positivity of the elements
  - Toeplitz or TBT structure
  - Symmetry
  - Sparsity
- The same Bayesian approach then can be applied.

Estimation of $\boldsymbol{A}$ when the sources $\boldsymbol{f}$ are known

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{F}\boldsymbol{a} + \boldsymbol{\epsilon}$$

- Prior on noise:

$$p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{A}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{A}\boldsymbol{f}, v_\epsilon \boldsymbol{I}) \quad \begin{aligned} &\propto \exp\left\{\tfrac{1}{2v_\epsilon}\|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|^2\right\} \\ &\propto \exp\left\{\tfrac{1}{2v_\epsilon}\|\boldsymbol{g}-\boldsymbol{F}\boldsymbol{a}\|^2\right\}\end{aligned}$$

- Simple prior models for $\boldsymbol{a}$:

$$p(\boldsymbol{A}|\alpha) \propto \exp\left\{-\alpha\|\boldsymbol{a}\|^2\right\} \propto \exp\left\{-\alpha\|\boldsymbol{A}\|^2\right\}$$

- Expression of the posterior law:

$$p(\boldsymbol{A}|\boldsymbol{g},\boldsymbol{f}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{A})\, p(\boldsymbol{A}) \propto \exp\left\{-J(\boldsymbol{A})\right\}$$

$$\text{with} \qquad J(\boldsymbol{A}) = \frac{1}{2v_\epsilon}\|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|^2 + \alpha\|\boldsymbol{A}\|^2$$

- MAP estimation:

$$\widehat{\boldsymbol{a}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda\boldsymbol{I})^{-1}\boldsymbol{F}'\boldsymbol{g} \leftrightarrow \widehat{\boldsymbol{A}} = \boldsymbol{g}\boldsymbol{f}'(\boldsymbol{f}\boldsymbol{f}' + \lambda\boldsymbol{I})^{-1}$$

Bayesian source separation: both $\boldsymbol{A}$ and $\boldsymbol{f}$ unknown

$$p(\boldsymbol{f},\boldsymbol{A}|\boldsymbol{g},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3)=\frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{A},\boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2)\,p(\boldsymbol{A}|\boldsymbol{\theta}_3)}{p(\boldsymbol{g}|\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3)}$$

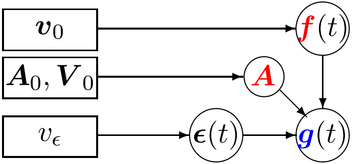
Two approaches:

- Joint estimation
- Estimate first $\boldsymbol{A}$ and then use it for estimating $\boldsymbol{f}$
- Joint estimation (JMAP):
  $(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{A}})=\arg\max_{(\boldsymbol{f},\boldsymbol{A})}\{p(p(\boldsymbol{f},\boldsymbol{A}|\boldsymbol{g},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3)\}$

  $$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{A}})=\arg\min_{(\boldsymbol{f},\boldsymbol{A})}\left\{\|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|^2+\lambda_1\|\boldsymbol{f}\|^2+\lambda_2\|\boldsymbol{A}\|^2\right\}$$
- Permutation and scale ideterminations: needs good choices for priors
- Alternate optimisation

  $$\begin{cases}\widehat{\boldsymbol{f}}=\arg\min_{\boldsymbol{f}}\left\{\|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|^2+\lambda_1\|\boldsymbol{f}\|^2\right\}=(\boldsymbol{A}'\boldsymbol{A}+\lambda_1\boldsymbol{I})^{-1}\boldsymbol{A}'\boldsymbol{g}\\ \widehat{\boldsymbol{A}}=\arg\min_{\boldsymbol{A}}\left\{\|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|^2+\lambda_2\|\boldsymbol{A}\|^2\right\}=\boldsymbol{g}\boldsymbol{f}'(\boldsymbol{f}\boldsymbol{f}'+\lambda_2\boldsymbol{I})\end{cases}$$
- Importance of initialization and other constraintes such as positivity

General case: Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$



$$p(f_j(t)|v_{0j}) = \mathcal{N}(0, v_{0j})$$
$$p(\boldsymbol{f}(t)|\boldsymbol{v}_0) \propto \exp\left\{-\frac{1}{2}\sum_j f_j^2(t)/v_{0j}\right\}$$
$$p(\boldsymbol{A}_{ij}|\boldsymbol{A}_{0ij}, \boldsymbol{V}_{0ij}) = \mathcal{N}(\boldsymbol{A}_{0ij}, \boldsymbol{V}_{0ij})$$
$$p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{V}_0)$$

$$p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I})$$

$$\begin{aligned} p(\boldsymbol{f}_{1..T}, \boldsymbol{A}|\boldsymbol{g}_{1..T}) &\propto p(\boldsymbol{g}_{1..T}|\boldsymbol{A}, \boldsymbol{f}_{1..T}, v_\epsilon)\, p(\boldsymbol{f}_{1..T})\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) \\ &\propto \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{v}_0)\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) \end{aligned}$$

$$p(\boldsymbol{f}(t)|\boldsymbol{g}_{1..T}, \boldsymbol{A}, v_\epsilon, \boldsymbol{v}_0) = \mathcal{N}(\widehat{\boldsymbol{f}}(t), \widehat{\boldsymbol{\Sigma}})$$

$$p(\boldsymbol{A}|\boldsymbol{g}_{1..T}, \boldsymbol{f}_{1..T}, v_\epsilon, \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{V}})$$

Two approaches:

- Alternate joint MAP (JMAP) estimation
- Bayesian Variational Approximation

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$: Alternate JMAP

Let do some simplification:

$$\begin{array}{ll} \boldsymbol{v}_0 = [v_f, .., v_f]', & \text{All sources a priori same variance } v_f \\ \boldsymbol{v}_\epsilon = [v_\epsilon, .., v_\epsilon]', & \text{All noise terms a priori same variance } v_\epsilon \\ \boldsymbol{A}_0 = 0, \quad \boldsymbol{V}_0 = v_a \boldsymbol{I} & \end{array}$$

$$p(\boldsymbol{f}(t)|\boldsymbol{g}(t), \boldsymbol{A}, v_\epsilon, \boldsymbol{v}_0) = \mathcal{N}(\widehat{\boldsymbol{f}}(t), \widehat{\boldsymbol{\Sigma}})$$
$$\left\{ \begin{array}{l} \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \boldsymbol{I})^{-1} \\ \widehat{\boldsymbol{f}}(t) = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \boldsymbol{I})^{-1} \boldsymbol{A}'\boldsymbol{g}(t), \quad \lambda_f = v_\epsilon / v_f \end{array} \right.$$

$$p(\boldsymbol{A}|\boldsymbol{g}(t), \boldsymbol{f}(t), v_\epsilon, \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{V}})$$
$$\left\{ \begin{array}{l} \widehat{\boldsymbol{V}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_f \boldsymbol{I})^{-1} \\ \widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t) \boldsymbol{f}'(t) \left(\sum_t \boldsymbol{f}(t) \boldsymbol{f}'(t) + \lambda_a \boldsymbol{I}\right)^{-1}, \quad \lambda_a = v_\epsilon / v_a \end{array} \right.$$

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$: Alternate JMAP

$$
\begin{aligned}
p(\boldsymbol{f}_{1..T}, \boldsymbol{A}|\boldsymbol{g}_{1..T}) &\propto p(\boldsymbol{g}_{1..T}|\boldsymbol{A}, \boldsymbol{f}_{1..T}, v_\epsilon)\, p(\boldsymbol{f}_{1..T})\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) \\
&\propto \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{z}(t))\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)
\end{aligned}
$$

Joint MAP: Alternate optimization

$$
\left\{
\begin{array}{ll}
\widehat{\boldsymbol{f}}(t) = (\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}} + \lambda_f \boldsymbol{I})^{-1} \widehat{\boldsymbol{A}}'\boldsymbol{g}(t), & \lambda_f = v_\epsilon / v_f \\
\widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t)\widehat{\boldsymbol{f}}'(t) \left(\sum_t \widehat{\boldsymbol{f}}(t)\widehat{\boldsymbol{f}}'(t) + \lambda_a \boldsymbol{I}\right)^{-1} & \lambda_a = v_\epsilon / v_a
\end{array}
\right.
$$

Alternate optimization Algorithm:

$$
\begin{array}{ccccc}
\boldsymbol{A}^{(0)} \longrightarrow \widehat{\boldsymbol{A}} \longrightarrow & \boxed{\left(\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}} + \lambda_f \boldsymbol{I}\right)^{-1} \widehat{\boldsymbol{A}}'\boldsymbol{g}} & \longrightarrow \widehat{\boldsymbol{f}}(t) \\
\uparrow & & \downarrow \\
\widehat{\boldsymbol{A}} \longleftarrow & \boxed{\sum_t \boldsymbol{g}(t)\widehat{\boldsymbol{f}}'(t) \left(\sum_t \widehat{\boldsymbol{f}}(t)\widehat{\boldsymbol{f}}'(t) + \lambda_a \boldsymbol{I}\right)^{-1}} & \longleftarrow \widehat{\boldsymbol{f}}(t)
\end{array}
$$

Variational Bayesian Approximation

Can we do better? Yes, VBA is a good solution.

- Main idea: Approximate a joint pdf $p(\boldsymbol{x})$ difficult to handle by a simpler one (for example a separable one $q(\boldsymbol{x}) = \prod_j q_j(x_j)$)
- Criterion: minimize

$$\mathsf{KL}(q|p) = \int q \ln \frac{q}{p} = \left\langle \ln \frac{q}{p} \right\rangle_q = \sum_j H(q_j) - < \ln p(\boldsymbol{x}) >_q$$

- Solution: $q_j(x_j) \propto \exp\left\{- < \ln p(\boldsymbol{x}) >_{q_{-j}}\right\}$
- In our case: Approximate $p(\boldsymbol{f}, \boldsymbol{A}|\boldsymbol{g})$ by a separable one $q(\boldsymbol{f}, \boldsymbol{A}) = q_1(\boldsymbol{f})q_2(\boldsymbol{A})$
- Solution obtained by alternate optimization:

$$\left\{ \begin{array}{l} q_1(\boldsymbol{f}) \propto \exp\left\{- < \ln p(\boldsymbol{f}, \boldsymbol{A}|\boldsymbol{g}) >_{q_2(\boldsymbol{A})}\right\} \\ q_2(\boldsymbol{A}) \propto \exp\left\{- < \ln p(\boldsymbol{f}, \boldsymbol{A}|\boldsymbol{g}) >_{q_1(\boldsymbol{f})}\right\} \end{array} \right.$$

Joint Estimation: Variational Bayesian Approximation

$$p(\boldsymbol{f}_{1..T}, \boldsymbol{A}|\boldsymbol{g}_{1..T}) \longrightarrow q_1(\boldsymbol{f}_{1..T}|\widetilde{\boldsymbol{A}}, \boldsymbol{g}_{1..T})\; q_2(\boldsymbol{A}|\widetilde{\boldsymbol{f}}_{1..T}, \boldsymbol{g}_{1..T})$$

$$q_1(\boldsymbol{f}(t)|\boldsymbol{g}(t), \boldsymbol{A}, v_\epsilon, \boldsymbol{v}_0) = \mathcal{N}(\widehat{\boldsymbol{f}}(t), \widehat{\boldsymbol{\Sigma}})$$
$$\left\{\begin{array}{l} \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \widehat{\boldsymbol{V}})^{-1} \\ \widehat{\boldsymbol{f}}(t) = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \widehat{\boldsymbol{V}})^{-1} \boldsymbol{A}'\boldsymbol{g}(t), \quad \lambda_f = v_\epsilon/v_f \end{array}\right.$$

$$q_2(\boldsymbol{A}|\boldsymbol{g}(t), \boldsymbol{f}(t), v_\epsilon, \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{V}})$$
$$\left\{\begin{array}{l} \widehat{\boldsymbol{V}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_f \widehat{\boldsymbol{\Sigma}})^{-1} \\ \widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t)\boldsymbol{f}'(t) \left(\sum_t \boldsymbol{f}(t)\boldsymbol{f}'(t) + \lambda_a \widehat{\boldsymbol{\Sigma}}\right)^{-1}, \quad \lambda_a = v_\epsilon/v_a \end{array}\right.$$

$\boldsymbol{A}^{(0)} \longrightarrow \widehat{\boldsymbol{A}} \longrightarrow$, $\boldsymbol{V}^{(0)} \longrightarrow \widehat{\boldsymbol{V}} \longrightarrow$

$$\boxed{\begin{array}{l} \widehat{\boldsymbol{f}}(t) = \left(\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}} + \lambda_f \widehat{\boldsymbol{V}}\right)^{-1} \widehat{\boldsymbol{A}}'\boldsymbol{g}(t) \\ \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \widehat{\boldsymbol{V}})^{-1} \end{array}}$$

$\longrightarrow \widehat{\boldsymbol{f}}(t)$, $\longrightarrow \widehat{\boldsymbol{\Sigma}}$

$\Uparrow$ $\qquad\qquad$ $\Downarrow$

$\widehat{\boldsymbol{A}} \longleftarrow$, $\widehat{\boldsymbol{V}} \longleftarrow$

$$\boxed{\begin{array}{l} \widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t)\widehat{\boldsymbol{f}}'(t) \left(\sum_t \widehat{\boldsymbol{f}}(t)\widehat{\boldsymbol{f}}'(t) + \lambda_a \widehat{\boldsymbol{\Sigma}}\right)^{-1} \\ \widehat{\boldsymbol{V}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_f \widehat{\boldsymbol{\Sigma}})^{-1} \end{array}}$$

$\longleftarrow \widehat{\boldsymbol{f}}(t)$, $\longleftarrow \widehat{\boldsymbol{\Sigma}}$

Bayesian Sparse Sources Separation

Three main steps:

- Assigning priors (sparsity enforcing):
  - Simple priors: $p(\boldsymbol{f})$ and $p(\boldsymbol{A})$
  - Hierarchical priors: $p(\boldsymbol{f}|\boldsymbol{z})\,p(\boldsymbol{z})$ and $p(\boldsymbol{A}|\boldsymbol{q})\,p(\boldsymbol{q})$
- Obtaining the expressions of $p(\boldsymbol{f},\boldsymbol{A},\boldsymbol{\theta}|\boldsymbol{g})$ or $p(\boldsymbol{f},\boldsymbol{A},\boldsymbol{z},\boldsymbol{q},\boldsymbol{\theta}|\boldsymbol{g})$
- Doing the computations:
  - Joint optimization of $p(\boldsymbol{f},\boldsymbol{A},\boldsymbol{\theta}|\boldsymbol{g})$;
  - MCMC Gibbs sampling methods which need generation of samples from the conditionals $p(\boldsymbol{f}|\boldsymbol{A},\boldsymbol{\theta},\boldsymbol{g})$, $p(\boldsymbol{A}|\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{g})$ and $p(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{A},\boldsymbol{g})$;
  - Bayesian Variational Approximation (BVA) methods which approximate $p(\boldsymbol{f},\boldsymbol{A},\boldsymbol{\theta}|\boldsymbol{g})$ by a separable one

$$q(\boldsymbol{f},\boldsymbol{A},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\widetilde{\boldsymbol{A}},\widetilde{\boldsymbol{\theta}},\boldsymbol{g})\; q_2(\boldsymbol{A}|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}},\boldsymbol{g})\; q_3(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{A}},\boldsymbol{g})$$

and then using them for the estimation.

Conclusions

- General source separation problem
  - Estimation of $\boldsymbol{f}$ when $\boldsymbol{A}$ is known
  - Estimation of $\boldsymbol{A}$ when the sources $\boldsymbol{f}$ are known
  - Joint estimation of the sources $\boldsymbol{f}$ and the mixing matrix $\boldsymbol{A}$
- Priors which enforce sparsity:
  - Generalized Gaussian, Student-t, Elastic nets, ...
  - Scaled Gaussian Mixture, Mixture of Gaussians or Gammas, Bernoulli-Gaussian
- Computational tools:
  - Alternate optimization of JMAP criterion
  - MCMC
  - Variational Bayesian Approximation
- Advanced Bayesian methods: Non-Gaussian, Dependent and nonstationnary signals and images.
- Some domaines of applications
  - Acoustic Source localization, Radar and SAR imaging, Spectrometry, Cosmic Microwave Background, Sattelite Image separation, Hyperspectral image processing

References

A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," EURASIP Journal on Advances in Signal Processing, vol. Special issue on Sparse Signal Processing, (2012).

N. Bali and A. Mohammad-Djafari, "Bayesian Approach With Hidden Markov Modeling and Mean Field Approximation for Hyperspectral Data Analysis," IEEE Trans. on Image Processing 17: 2. 217-225 Feb. (2008).

F. Su and A. Mohammad-Djafari, "An Hierarchical Markov Random Field Model for Bayesian Blind Image Separation," 27-30 May 2008, Sanya, Hainan, China: International Congress on Image and Signal Processing (CISP 2008).

N. Bali, A. Mohammad-Djafari, "Bayesian Approach With Hidden Markov Modeling and Mean Field Approximation for Hyperspectral Data Analysis," IEEE Trans. on Image Processing 17: 2. 217-225 Feb. (2008).

H. Snoussi and A. Mohammad-Djafari, " Estimation of Structured Gaussian Mixtures: The Inverse EM Algorithm," IEEE Trans. on Signal Processing 55: 7. 3185-3191 July (2007).

N. Bali and A. Mohammad-Djafari, "A variational Bayesian Algorithm for BSS Problem with Hidden Gauss-Markov Models for the Sources," in: Independent Component Analysis and Signal Separation (ICA 2007) Edited by:M.E. Davies, Ch.J. James, S.A. Abdallah, M.D. Plumbley. 137-144 Springer (LNCS 4666) (2007).

N. Bali and A. Mohammad-Djafari, "Hierarchical Markovian Models for Joint Classification, Segmentation and Data Reduction of Hyperspectral Images" ESANN 2006, September 4-8, Belgium. (2006)

M. Ichir and A. Mohammad-Djafari, "Hidden Markov models for wavelet-based blind source separation," IEEE Trans. on Image Processing 15: 7. 1887-1899 July (2005)

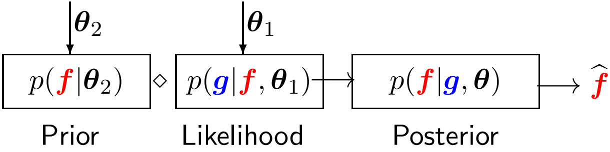
S. Moussaoui, C. Carteret, D. Brie and A Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods sampling," Chemometrics and Intelligent Laboratory Systems 81: 2. 137-148 (2005).

H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images" Journal of Electronic Imaging 13: 2. 349-361 April (2004)

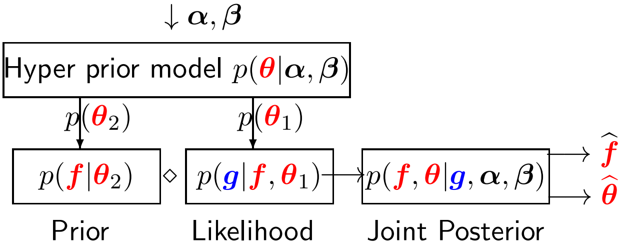
H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," Journal of VLSI Signal Processing Systems 37: 2/3. 263-279 June/July (2004)

Summary of Bayesian estimation with different levels

- Simple Bayesian Model and Estimation



- Full Bayesian Model and Hyperparameter Estimation scheme



Summary of Bayesian estimation with different levels

- Marginalization for Hyperparameter Estimation

$$\boxed{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \boxed{p(\boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \widehat{\boldsymbol{\theta}} \longrightarrow \boxed{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}},\boldsymbol{g})} \longrightarrow \widehat{\boldsymbol{f}}$$

Joint Posterior　　Marginalize over $\boldsymbol{f}$

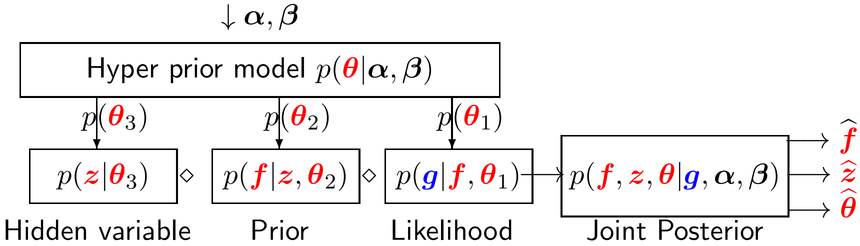
- Full Bayesian Model with a Hierarchical Prior Model

$$\begin{array}{ccccccc} \boldsymbol{\theta}_3 & & \boldsymbol{\theta}_2 & & \boldsymbol{\theta}_1 & & \\ \downarrow & & \downarrow & & \downarrow & & \\ \boxed{p(\boldsymbol{z}|\boldsymbol{\theta}_3)} & \diamond & \boxed{p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)} & \diamond & \boxed{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)} & \longrightarrow & \boxed{p(\boldsymbol{f},\boldsymbol{z}|\boldsymbol{g},\boldsymbol{\theta})} \longrightarrow \widehat{\boldsymbol{f}},\ \widehat{\boldsymbol{z}} \end{array}$$

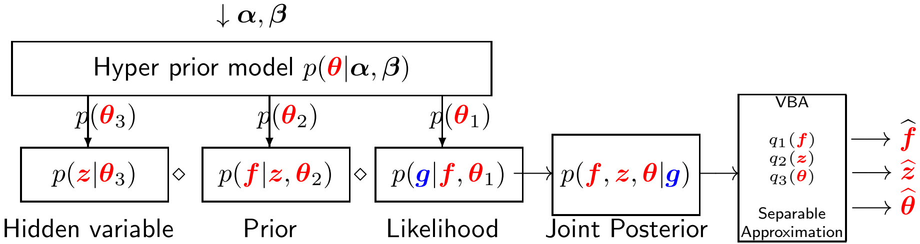
Hidden variable　　Prior　　Likelihood　　Joint Posterior

Summary of Bayesian estimation with different levels

- Full Bayesian Hierarchical Model with Hyperparameter Estimation



- Full Bayesian Hierarchical Model and Variational Approximation



Prior models with hidden variables

- Example 1: MoG model:

$$p(f_j|\lambda, v_1, v_0) = \lambda \mathcal{N}(f_j|0, v_1) + (1-\lambda)\mathcal{N}(f_j|0, v_0)$$

$$\left\{ \begin{array}{l} p(f_j|z_j = 0, v_0) = \mathcal{N}(f_j|0, v_0), \\ p(f_j|z_j = 1, v_1) = \mathcal{N}(f_j|0, v_1), \end{array} \right. \text{ and } \left\{ \begin{array}{l} P(z_j = 0) = \lambda, \\ P(z_j = 1) = 1 - \lambda \end{array} \right.$$

$$\left\{ \begin{array}{l} p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}\left(f_j|0, v_{z_j}\right) \propto \exp\left\{-\frac{1}{2}\sum_j \frac{f_j^2}{v_{z_j}}\right\} \\ P(z_j = 1) = \lambda, \qquad P(z_j = 0) = 1 - \lambda \end{array} \right.$$

- Example 2: Student-t model

$$\mathcal{S}t(f|\nu) = \int_0^\infty \mathcal{N}(f|, 0, 1/z)\, \mathcal{G}(z|\alpha, \beta)\; \mathrm{d}z, \quad \text{with } \alpha = \beta = \nu/2$$

$$\left\{ \begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{z}) & = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0, 1/z_j) \propto \exp\left\{-\frac{1}{2}\sum_j z_j f_j^2\right\} \\ p(\boldsymbol{z}|\alpha, \beta) & = \prod_j \mathcal{G}(z_j|\alpha, \beta) \propto \prod_j z_j^{(\alpha-1)} \exp\{-\beta z_j\} \\ & \propto \exp\left\{\sum_j (\alpha-1)\ln z_j - \beta z_j\right\} \\ p(\boldsymbol{f}, \boldsymbol{z}|\alpha, \beta) & \propto \exp\left\{-\frac{1}{2}\sum_j z_j f_j^2 + (\alpha-1)\ln z_j - \beta z_j\right\} \end{array} \right.$$

- With these models we have:

Bayesian Computation and Algorithms

- Often, the expression of $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ is complex.
- Its optimization (for Joint MAP) or
  its marginalization or integration (for Marginal MAP or PM)
  is not easy
- Two main techniques:
  MCMC and Variational Bayesian Approximation (VBA)
- MCMC:
  Needs the expressions of the conditionals
  $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})$, $p(\boldsymbol{z}|\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g})$, and $p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{g})$
- VBA: Approximate $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable one

$$q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$$

  and do any computations with these separable ones.

General case

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t), \quad t \in [1, \cdots, T]$$

$$p(\boldsymbol{g}(t)|\boldsymbol{f}(t), \boldsymbol{A}, v_\epsilon, \boldsymbol{v}_0) = \mathcal{N}(\boldsymbol{g}(t) - \boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I}) p(\boldsymbol{f}(t)|v_f) = \mathcal{N}(0, v_0 \boldsymbol{I})$$

$$p(\boldsymbol{f}(t)|\boldsymbol{g}(t), \boldsymbol{A}, v_\epsilon, \boldsymbol{v}_0) = \mathcal{N}(\widehat{\boldsymbol{f}}(t), \widehat{\boldsymbol{\Sigma}})$$

Full Bayesian and Variational Bayesian Approximation

- Full Bayesian: $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)\, p(\boldsymbol{\theta})$
- Approximate $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\boldsymbol{g})\, q_2(\boldsymbol{\theta}|\boldsymbol{g})$ and then continue computations.
- Criterion $\text{KL}(q(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) : p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}))$
- $\text{KL}(q : p) = \int\int q \ln q/p = \int\int q_1 q_2 \ln \frac{q_1 q_2}{p}$
- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

$$\left\{ \begin{array}{ll} \widehat{q}_1(\boldsymbol{f}) & \propto \exp\left\{ \langle \ln p(\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}; \mathcal{M}) \rangle_{\widehat{q}_2(\boldsymbol{\theta})} \right\} \\ \widehat{q}_2(\boldsymbol{\theta}) & \propto \exp\left\{ \langle \ln p(\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}; \mathcal{M}) \rangle_{\widehat{q}_1(\boldsymbol{f})} \right\} \end{array} \right.$$

$\boxed{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow$ **[Variational Bayesian Approximation]** $\longrightarrow q_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}}$, $\longrightarrow q_2(\widehat{\boldsymbol{\theta}}) \longrightarrow \widehat{\boldsymbol{\theta}}$

Bayesian estimation approach

$$\boldsymbol{x}(t) = \boldsymbol{A}\,\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t) \quad \text{ou} \quad \boldsymbol{x}_{1..T} = \boldsymbol{A}\,\boldsymbol{s}_{1..T} + \boldsymbol{\epsilon}_{1..T} \quad \text{ou} \quad \boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{E}$$

$$p(\boldsymbol{A}, \boldsymbol{s}_{1..T} | \boldsymbol{x}_{1..T}) \propto p(\boldsymbol{x}_{1..T} | \boldsymbol{A}, \boldsymbol{s}_{1..T})\, p(\boldsymbol{A})\, p(\boldsymbol{s}_{1..T})$$

$$\boldsymbol{x}(\boldsymbol{r}) = \boldsymbol{A}\,\boldsymbol{s}(\boldsymbol{r}) + \boldsymbol{\epsilon}(\boldsymbol{r}) \quad \text{ou} \quad \boldsymbol{x}(\boldsymbol{r}) = \boldsymbol{A}\,\boldsymbol{s}_{\boldsymbol{r}\in\mathcal{R}} + \boldsymbol{\epsilon}(\boldsymbol{r}) \quad \text{ou} \quad \boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{E}$$

$$p(\boldsymbol{A}, \boldsymbol{s}_{\boldsymbol{r}\in\mathcal{R}} | \boldsymbol{x}(\boldsymbol{r})) \propto p(\boldsymbol{x}(\boldsymbol{r}) | \boldsymbol{A}, \boldsymbol{s}_{\boldsymbol{r}\in\mathcal{R}})\, p(\boldsymbol{A})\, p(\boldsymbol{s}_{\boldsymbol{r}\in\mathcal{R}})$$

$$p(\boldsymbol{A}, \boldsymbol{S} | \boldsymbol{X}) \propto p(\boldsymbol{X} | \boldsymbol{A}, \boldsymbol{S})\, p(\boldsymbol{A})\, p(\boldsymbol{S})$$

$$p(\boldsymbol{X} | \boldsymbol{A}, \boldsymbol{S}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{S}, \boldsymbol{\Sigma}_\epsilon), \quad p(\boldsymbol{A}) = \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{\Sigma}_0) \text{ or uniform}$$

◊ Important step: Choice of $p(\boldsymbol{S})$

Bayesian estimation approach

$$\boldsymbol{x}(t) = \boldsymbol{A}\,\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t)$$

$$p(\boldsymbol{A}, \boldsymbol{s}_{1..T}|\boldsymbol{x}_{1..T}) \propto p(\boldsymbol{x}_{1..T}|\boldsymbol{A}, \boldsymbol{s}_{1..T})\, p(\boldsymbol{A})\, p(\boldsymbol{s}_{1..T})$$

◊ 3 directions:

1. Joint estimation: $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{s}}_{1..T})$ using $p(\boldsymbol{A}, \boldsymbol{s}_{1..T}|\boldsymbol{x}_{1..T})$. For example JMAP:

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{s}}_{1..T}) = \arg \max_{(\boldsymbol{A}, \boldsymbol{s}_{1..T})} \{J(\boldsymbol{A}, \boldsymbol{s}_{1..T}) = \ln\, p(\boldsymbol{A}, \boldsymbol{s}_{1..T}|\boldsymbol{x}_{1..T})\}$$

2. $\boldsymbol{A}$ estimation: $\widehat{\boldsymbol{A}}$ using $p(\boldsymbol{A}|\boldsymbol{x}_{1..T})$. For example:

$$\widehat{\boldsymbol{A}} = \arg\max_{\boldsymbol{A}} \{J(\boldsymbol{A}) = \ln\, p(\boldsymbol{A}|\boldsymbol{x}_{1..T})\}$$

3. $\boldsymbol{s}$ estimation: $\widehat{\boldsymbol{s}}$ using $p(\boldsymbol{s}_{1..T}|\boldsymbol{x}_{1..T})$. For example:

$$\widehat{\boldsymbol{s}}_{1..T} = \arg\max_{\boldsymbol{s}} \{J(\boldsymbol{s}_{1..T}) = \ln\, p(\boldsymbol{s}_{1..T}|\boldsymbol{x}_{1..T})\}$$

Gaussian white case: PCA, MNF, PMF, NMF and SOBI

- White and Gaussian signals $\boldsymbol{s}(t), \boldsymbol{\epsilon}(t) \longrightarrow \boldsymbol{x}(t)$:

$$\boldsymbol{x}(t) = \boldsymbol{A}\,\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t) \longrightarrow \boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} + \boldsymbol{\epsilon}$$

$$p(\boldsymbol{x}, \boldsymbol{s}|\boldsymbol{A}) = p(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{s})\,p(\boldsymbol{s})$$

$$p(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{s}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{s}, \boldsymbol{\Sigma}_\epsilon), \quad p(\boldsymbol{s}) = \mathcal{N}(0, \boldsymbol{\Sigma}_s) \longrightarrow p(\boldsymbol{x}|\boldsymbol{A}) = \mathcal{N}(0, \boldsymbol{A}\boldsymbol{\Sigma}$$

- PCA : Estimate $\boldsymbol{\Sigma}_x$ by $\frac{1}{T}\sum_t \boldsymbol{x}(t)\boldsymbol{x}'(t)$,
  svd and keep all the non-zero svd: $\boldsymbol{\Sigma}_x = \boldsymbol{A}\boldsymbol{\Sigma}_s\boldsymbol{A}'$
- Minimum Norm Factorization (MNF) :
  Estimate $\boldsymbol{\Sigma}_x$, svd and keep all svd $\geq \sigma_\epsilon$: $\boldsymbol{\Sigma}_x = \boldsymbol{A}\boldsymbol{\Sigma}_s\boldsymbol{A}' + \boldsymbol{\Sigma}_\epsilon$
- Positive Matrix Factorization (MNF) :
  Decompose $\boldsymbol{\Sigma}_x$ in positive definite matrices [Paatero & Tapper, 94]
- Non-negative Matrix Factorization (NMF) :
  Decompose $\boldsymbol{\Sigma}_x$ in Non-negative definite matrices [Lee& Seung,99]

Gaussian white case: PCA, MNF, PMF, NMF and SOBI (2)

Accounting for non stationnarity

◊ SOBI : $1..T = 1..T_1..T_2..T_k..T_{k+1}..T$

$$\boldsymbol{\Sigma}_x(k) = \frac{1}{T_{k+1} - T_k} \sum_{t=T_k}^{T_{k+1}-1} \boldsymbol{x}(t)\boldsymbol{x}'(t)$$

Joint Diagonalization of $\boldsymbol{\Sigma}_x(k)$

Non Gaussian white case: ICA, JADE, NNICA and Tappered ICA

◊ White Non Gaussian signals and Exact model (no noise):

$$\boldsymbol{s}(t) \longrightarrow \boldsymbol{x}(t) \longrightarrow \boldsymbol{y}(t) = \boldsymbol{A}^{-1}\boldsymbol{x}(t) \longrightarrow \boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{x}(t)$$

ICA: Find $\boldsymbol{B}$ in such a way that the components of $\boldsymbol{y}$ be the most independent

Different measures of independencies:

$$S(\boldsymbol{y}) = -\int p(y_i) \ln p(y_i) \, \mathrm{d}y_i$$

$$KL(p(\boldsymbol{y}) : \prod_i p(y_i)) = \int p(y_i) \ln \frac{\prod_i p(y_i)}{p(\boldsymbol{y}} \, \mathrm{d}y_i$$

Different choices and approximations for $p(y_i) \longrightarrow$ contrast functions, cumulants basis criteria

Non Gaussian white case: ICA, JADE, NNICA and Tappered ICA (2)

◊ White Non Gaussian signals (Accounting for noise)

$$\boldsymbol{x}(t) = \boldsymbol{A}\,\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t) \longrightarrow \boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} + \boldsymbol{\epsilon}$$

$$p(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{\Sigma}_\epsilon) = \int p(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{s}, \boldsymbol{\Sigma}_\epsilon)\, p(\boldsymbol{s})\, \mathsf{d}\boldsymbol{s}$$

◊ ICA (Maximum Likelihood) :

$$\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Sigma}_\epsilon}) = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{\theta})\}$$

◊ EM iterative algorithm :

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \mathsf{E}\left\{\ln p(\boldsymbol{x}, \boldsymbol{s}|\boldsymbol{\theta})|\boldsymbol{\theta}'\right\} \\ \boldsymbol{\theta}' &= \arg\max_{\boldsymbol{\theta}} \left\{Q(\boldsymbol{\theta}, \boldsymbol{\theta}')\right\} \end{aligned}$$

Choice of a priori and JMAP algorithm

◊ Sources $\boldsymbol{s} \longrightarrow$ Mixture of Gaussians [Moulines97]:

$$p(s_j) = \sum_{i=1}^{q_j} \alpha_{ji} \mathcal{N}\left(m_{ji}, \sigma_{ji}^2\right), \qquad j = 1..n$$

◊ Mixing matrix $\boldsymbol{A} \longrightarrow$ a priori Gaussian for its elements:

$$p(\boldsymbol{A}_{ij}) = \mathcal{N}\left(\mu_{ij}, \sigma_{a,ij}^2\right)$$

◊ Scalar iteratif algorithms:

$$\left\{ \begin{array}{rcl} \widehat{\boldsymbol{s}}_j(t)^{(k+1)} & = & \arg\max_{\boldsymbol{s}_j} \left\{ \ln p\left(\boldsymbol{s}_j | \boldsymbol{x}(t), \widehat{\boldsymbol{A}}^{(k)}, \widehat{\boldsymbol{s}}_{l \neq j}(t)^{(k)}\right) \right\} \\ \widehat{\boldsymbol{A}}^{(k+1)} & = & \arg\max_{\boldsymbol{A}} \left\{ \ln p\left(\boldsymbol{A} | \widehat{\boldsymbol{s}}_{1..T}^{(k+1)}, \boldsymbol{x}_{1..T}\right) \right\} \end{array} \right.$$