Bayesian inference methods for sources separation

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

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

General Bayesian source separation problem

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

- $p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{A}, \boldsymbol{\theta}_1)$ likelihood
- $p(\boldsymbol{f}|\boldsymbol{\theta}_2)$ and $p(\boldsymbol{A}|\boldsymbol{\theta}_3)$ priors
- $p(\boldsymbol{f}, \boldsymbol{A}|\boldsymbol{g}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_3)$ 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 $\boldsymbol{f}$: $p(\boldsymbol{f}|\alpha) \propto \exp\{-\alpha\Omega(\boldsymbol{f})\}$
- Expression of the posterior law:

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

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

- Link between MAP estimation and regularization

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

MAP and link with regularization

- 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:

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

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 $\mathrm{KL}(q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) : p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}))$
- $\mathrm{KL}(q:p) = \int\int q \ln q/p = \int\int q_1 q_2 \ln \frac{q_1 q_2}{p}$
- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

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

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

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

Source separation is a bilinear model:

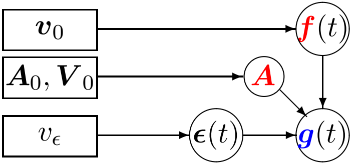
$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} = \boldsymbol{F}\boldsymbol{a} = \boldsymbol{A}\boldsymbol{f}$$

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

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

- Problem is more ill-posed.
- We need absolutely to impose constraintes on elements or the structure of $\boldsymbol{A}$, for example:
  - Positivity of the elements
  - Toeplitz or TBBT structure,
  - Symmetry $\quad p(\boldsymbol{A}) \propto \exp\left\{-\alpha \|\boldsymbol{I} - \boldsymbol{A}'\boldsymbol{A}|^2\right\}$
  - Sparsity $\quad p(\boldsymbol{A}) \propto \exp\left\{-\alpha \sum_{i,j} |\boldsymbol{A}_{ij}|\right\}$
- The same Bayesian approach then can be applied

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



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

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

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

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$ ..

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

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

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

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

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

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$ ..

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

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:

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

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$ with a Gaussian prior model..

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

$$q_1(\boldsymbol{f}(t)|\boldsymbol{g}(t), \boldsymbol{A}, v_\epsilon, \boldsymbol{v}_0) = \mathcal{N}(\widehat{\boldsymbol{f}}(t), \widehat{\boldsymbol{\Sigma}})$$

$$\begin{cases} \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \widehat{\boldsymbol{V}})^{-1} \\ \widehat{\boldsymbol{f}}(t) = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \widehat{\boldsymbol{V}})^{-1} \boldsymbol{A}'\boldsymbol{g}(t), \quad \lambda_f = v_\epsilon / v_f \end{cases}$$

$$q_2(\boldsymbol{A}|\boldsymbol{g}(t), \boldsymbol{f}(t), v_\epsilon, \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{V}})$$

$$\begin{cases} \widehat{\boldsymbol{V}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_f \widehat{\boldsymbol{\Sigma}})^{-1} \\ \widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t)\boldsymbol{f}'(t) \left(\sum_t \boldsymbol{f}(t)\boldsymbol{f}'(t) + \lambda_a \widehat{\boldsymbol{\Sigma}}\right)^{-1} \quad \lambda_a = v_\epsilon / v_a \end{cases}$$

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

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

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

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

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

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

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

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 $\boldsymbol{A}$
- General Bayesian inference for source separation
- Full Bayesian with hyperparameter estimation
- Priors which enforce sparsity
  - Generalized Gaussian, Student-t
  - Mixture of Gaussians or Gammas, Bernoulli-Gaussian
- Computational tools: Laplace approximation, MCMC and Variational Bayesian Approximation
- Advanced Bayesian methods: Non-Gaussian, Dependent and nonstationnary signals and images.
- Some domaines of applications
  - Source localization, Spectrometry, CMB, Sattelite Image separation, Hyperspectral image processing