Bayesian Approach with prior laws enforcing sparsity for inverse problems and sources separation

Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes,
UMR8506 CNRS-SUPELEC-UNIV PARIS SUD 11
SUPELEC, 91192 Gif-sur-Yvette, France

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

1. Preliminaries on Bayesian approach
2. Bayesian inference for inverse problems
   - Full Bayesian with hyperparameter estimation and model selection
   - Priors which enforce sparsity
     - Generalized Gaussian, Mixture of Gaussians, Student-t, ...
     - Gauss-Markov-Pottz
   - Computational tools: Laplace approximation, MCMC and Variational Bayesian Approximation
3. Sources separation as an inverse problem
   - Estimation of sources with known mixing matrix
   - Estimation of the mixing matrix with known sources
   - Joint estimation or marginalization
4. Links with classical methods: PCA, FA and ICA
5. Advanced Bayesian methods: Non-Gaussian, Dependent and nonstationnary signals and images.
6. Some results and applications
   - X ray Computed Tomography, Spectrometry, CMB, Sattelite Image separation, Hyperspectral image processing
1. Preliminaries on Bayesian inference

Bayesian Inference:
- Make inference about hypotheses using all available information
- Needs to identify all hypotheses explicitly
- Needs to model the link between hypotheses and observations
- No needs for complicated mathematics
- The most difficult part is to learn "Thinking Bayesian"
- People who have already learned "Orthodox statistics" may have some difficulties
- Whent you "get it", you will find it much easier to understand
- Conceptually simple, Logically consistent, Uniform (always the same), Powerful, Elegant
Bayesian inference for Scientists

- How do Hypotheses and Forward model predict the potential data?
- How do the Observed data support those Hypotheses and Model?
Probability

- What is the probability?
  Degree of belief always conditionned on what we know

- What is the probability of a fair coin comes up head?
  - Before tossing the coin?
  - After tossing, but before looking?
  - After tossing and looking by me but before telling to you?
  - After tossing and looking by yourself?
  - After tossing but looking at it through a low cost camera?

- Will it rain tomorrow?

- Is the millionth digit of $\pi$ the digit 3?

- Probability Axioms
  - $0 \leq P(A|I) \leq 1$
  - $P(A|A, I) = 1$
  - $P(A|I) + P(A^c|I) = 1$
  - $P(A, B|I) = P(A|B, I) P(B|I)$
  - $P(A|B, I) = P(A, B|I)/P(B|I)$ if $P(B|I) \neq 0$
Bayes’ rule

Bayes’ rule:

\[
P(A|D) = \frac{P(A, D)}{P(D)} = \frac{P(D|A) P(A)}{P(D)} \propto P(D|A) P(A)
\]

\[
P(D) = \sum_i P(D, A_i) = \sum_i P(D|A_i) P(A_i)
\]

Particular case of 2 states: \( A_2 = A_1^c \) (Mutually exclusive states)

Prior odds = \( \frac{P(A_1)}{P(A_2)} \)

Posterior odds = \( \frac{P(A_1|D)}{P(A_2|D)} = \frac{P(D|A_1)}{P(D|A_2)} \frac{P(A_1)}{P(A_2)} \)

Odds = \( \frac{Probability}{1 - Probability} \) → Probability = \( \frac{Odds}{1 - Odds} \)
Bayes’ rule for continuous case

Continuous case:

\[ p(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{p(x|\theta) p(\theta)}{p(x)} \propto p(x|\theta) p(\theta) \]

\[ p(x) = \int p(x, \theta) \, d\theta = \int p(x|\theta) p(\theta) \, d\theta \]

posterior \( \propto \) likelihood \( \times \) prior

- The posterior law summarizes all we know after we have considered our prior knowledge and the data.

- In practice, we need to summarize it:
  - Mode: \( \hat{\theta}_{Mod} = \arg \max_\theta \{ p(\theta|x) \} \)
  - Mean: \( \hat{\theta}_{Mean} = \int \theta p(\theta|x) \, d\theta \)
  - Median: \( \theta_{Med} : P(\theta > \theta_{Med}) = P(\theta < \theta_{Med}) \)
Credible intervals

- We can also compute Credible intervals:

\[ P(\theta \in [a, b]|x) = \int_a^b p(\theta \in [a, b]|x) \, d\theta \]

Credible interval is \([a, b]\) such that \(P(\theta \in [a, b]|x) = 0.95\).

- This is different from "Orthodox Confidence Intervals":
  - Define an estimator \(\hat{\theta}\) which is a function of data \(x\) (which are random!)
  - Compute the sampling distribution \(g(\theta|\theta)\)
  - Compute \(\alpha = 1 - \int_a^b g(u|\theta) \, du\)
  - Compute \(a\) and \(b\) such that \(\alpha = .05\)

- Even if, in some cases, the results are numerically the same, the interpretations are not the same:
The Bayesian way is much more understandable!
A very simple example: Gaussian law

\[ x_i \sim \mathcal{N}(\mu, \nu), \ i = 1, \ldots, n \rightarrow p(x_i | \mu, \nu) = \frac{1}{\sqrt{2\pi\nu}} \exp \left\{ -\frac{1}{2} \frac{(x_i - \mu)^2}{\nu} \right\} \]

\[ x_i = \mu + \epsilon_i \text{ with } \epsilon_i \sim \mathcal{N}(0, \nu), \ i = 1, \ldots, n \]

Likelihood:

\[ p(x | \mu, \nu) = (2\pi\nu)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_i \frac{(x_i - \mu)^2}{\nu} \right\} \]

\[ = (2\pi\nu)^{-n/2} \exp \left\{ -\frac{1}{2\nu} (S + n(\mu - \bar{x})) \right\} \]

with

\[ \bar{x} = \frac{1}{n} \sum_i x_i \text{ and } S = \sum_i (x_i - \bar{x})^2 \]

- Maximum Likelihood: \( \hat{\mu} = \bar{x}, \hat{\nu} = \frac{1}{n} S \)

- Bayesian:
  - Known \( \nu \), flat prior for \( \mu \) \( \rightarrow \hat{\mu} = \bar{x} \)
  - Known \( \nu \), Gaussian prior for \( \mu \)
  - Known \( \mu \), unknown \( \nu \) with Jeffreys' prior
  - Inverse Gamma prior for \( \nu \) and Gaussian prior for \( \mu \)
    (Conjugate priors)
Example of the Gaussian law

Bayesian:

- Known $v$, flat prior for $\mu$
  \[
  \rightarrow p(\mu|\mathbf{x}, v) = \mathcal{N}(\bar{x}, v/n) \rightarrow \hat{\mu} = \bar{x}
  \]

- Known $v$, Gaussian prior for $\mu$:
  \[
  p(\mu|\mu_0, v_0) = \mathcal{N}(\mu_0, v_0) \rightarrow p(\mu|\mathbf{x}, v) = \mathcal{N}(\hat{\mu}, \hat{v})
  \]
  with $\frac{1}{\hat{v}} = \frac{n}{v} + \frac{1}{v_0}$, $\hat{\mu} = (1 - w)\bar{x} + w\mu_0$, with $w = \hat{v}/v_0$

When $v_0 \rightarrow \infty$ (flat prior) $\rightarrow \hat{\mu} = \bar{x}$

- Known $\mu$, unknown $v$ with Jeffreys’ prior $p(v) = 1/v$:
  \[
  p(v|\mathbf{x}, \mu) \propto v^{-n/2} \exp \left\{ -\frac{1}{2v} (S + n(\mu - \bar{x})) \right\} v^{-1}
  \]
  $\propto v^{-(n+1)/2} \exp \left\{ -\frac{1}{2v} (S + n(\mu - \bar{x})) \right\}$
  \[
  = IG(\hat{\alpha}, \hat{\beta})
  \]
  with $\begin{cases}
  \hat{\alpha} = \frac{n+1}{2} \\
  \hat{\beta} = (S + n(\mu - \bar{x})) / 2
  \end{cases}$
  $\rightarrow \hat{v} = \frac{\hat{\beta}}{\hat{\alpha}} = \frac{1}{n - 1} \sum_i (x_i - \mu)^2$
Example of the Gaussian law

Bayesian:

- Unknown $\mu$ with flat prior $p(\mu) = \text{cte}$,
- Unknown $v$ with Jeffreys’ prior $p(v) = 1/v$:

$$
p(\mu, v|\mathbf{x}) \propto v^{-n/2} \exp \left\{ -\frac{1}{2v}(S + n(\mu - \bar{x}) \right\} v^{-1}$$

$$
\propto v^{-(n+1)/2} \exp \left\{ -\frac{1}{2v}(S + n(\mu - \bar{x}) \right\}
$$

$$
= \mathcal{N}(\bar{x}, v/n) \mathcal{IG}(\hat{\alpha}, \hat{\beta})
$$

with \[
\begin{align*}
\hat{\alpha} &= \frac{n+1}{2} \\
\hat{\beta} &= (S + n(\mu - \bar{x})) / 2
\end{align*}
\]

$$\hat{\nu} = \frac{\hat{\beta}}{\hat{\alpha}} = \frac{1}{n-1} \sum_{i}(x_i-\mu)^2$$

We may integrate out $\mu$:

$$p(v|\mathbf{x}) = \int p(\mu, v|\mathbf{x}) \, d\mu$$

$$\propto \int v^{-(n+1)/2} \exp \left\{ -\frac{1}{2v}(S + n(\mu - \bar{x}) \right\}$$

$$\propto \left(1 + \frac{n(\mu-\bar{x})^2}{S} \right)^{-n/2}$$

$$S_t \left( \nu = n - 1, t = \frac{\mu-\bar{x}}{v\sqrt{n}}, v = S/v \right) \propto \left(1 + \frac{t^2}{v} \right)^{-(\nu+1)/2}$$
Example of the Gaussian law with Conjugate priors

- Unknown $\mu$, Unknown $v$ with Conjugate priors:

$$p(\mu, v | x) \propto v^{-n/2} \exp \left\{ -\frac{1}{2v} (S + n(\mu - \bar{x}) \right\} p(v) p(\mu)$$

$$\left\{ \begin{array}{l}
p(\mu | \mu_0, v_0) = N(\mu_0, v_0) \\
p(v | \alpha_0, \beta_0) = IG(\alpha_0, \beta_0) \end{array} \right. \longrightarrow p(\mu, v | x) = N(\hat{\mu}, \hat{v}) IG(\hat{\alpha}, \hat{\beta})$$

$$\left\{ \begin{array}{l}
\hat{\mu} = (1 - w) \bar{x} + w \mu_0, \text{ with } \frac{\hat{v}}{v_0} \\
\hat{v} = (\frac{1}{v} + \frac{1}{v_0})^{-1} \end{array} \right. $$

$$\left\{ \begin{array}{l}
\hat{\alpha} = \frac{n+1}{2} \\
\hat{\beta} = (S + n(\mu - \bar{x})^2 / 2 \end{array} \right. $$
Example of Normal Linear Regression

- A line through a set of points \((x_i, y_i)\) with \(y_i = \beta_0 + \beta_1 x_i\)

  \[ f(x, \beta_1, \beta_2) = \beta_0 + \beta_1 x \]

- Change of variables \(\alpha_0 = \beta_0 + \beta_1 \bar{x}, \alpha_1 = \beta_1\)

  \[ f(x, \alpha_1, \alpha_2) = \alpha_0 + \alpha_1 (x - \bar{x}) \]

- Likelihood: \(y_i = \alpha_0 + \alpha_1 (x_i - \bar{x}) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \nu)\)

  \[ p(x, y|\alpha_1, \alpha_2, \nu) \propto \nu^{-n/2} \exp \left\{ -\frac{1}{2\nu} \sum_i (y_i - \alpha_0 + \alpha_1 (x - \bar{x})^2 \right\} \]
Example of Normal Linear Regression

- Defining

\[ \hat{\alpha}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \]

\[ S_e = \sum_i (y_i - \bar{y})^2 - \hat{\alpha}_1 \sum_i (x_i - \bar{x})(y_i - \bar{y}) \]

\[ S_x = \sum_i (x_i - \bar{x})^2 \]

- Likelihood:

\[ p(x, y|\alpha_1, \alpha_2, v) \propto v^{-n} \exp \left\{ -\frac{1}{2v} S_e + n(\alpha_0 - \bar{y})^2 + S_x(\alpha_1 - \hat{\alpha}_1)^2 \right\} \]

- Flat priors on \( \alpha_0 \) and \( \alpha_1 \) and Jeffrey’s on \( v \):

\[ p(\alpha_0, \alpha_1, v|x, y) \propto v^{-(n+1)} \exp \left\{ -\frac{1}{2v} S_e + n(\alpha_0 - \bar{y})^2 + S_x(\alpha_1 - \hat{\alpha}_1)^2 \right\} \]
Example of Normal Linear Regression

- Marginalizing over $\alpha_1$:

$$ p(\alpha_0, v|x, y) \propto v^{-n} \exp \left\{ -\frac{1}{2v} S_e + n(\alpha_0 - \bar{y})^2 \right\} $$

- Marginalizing over $v$:

$$ p(\alpha_0|x, y) \propto v^{-n} \left( 1 + \frac{n}{S_e} (\alpha_0 - \bar{y})^2 \right)^{-(n-1)/2} $$

- Marginalizing over $\alpha_0$ and then over $v$:

$$ p(\alpha_1|x, y) \propto \left( 1 + \frac{S_x}{S_e} (\alpha_1 - \hat{\alpha}_1)^2 \right)^{-(n-1)/2} $$

- Marginalizing over $\alpha_0$ and then over $\alpha_1$:

$$ p(v|x, y) \propto v^{-(n+1)} \exp \left\{ -\frac{1}{2v} S_e \right\} $$
Example of Normal Linear Regression

- Marginalizing over $v$

\[ p(\alpha_0, \alpha_1 | x, y) \propto (S_e + n(\alpha_0 - \hat{\alpha}_0)^2 + S_x(\alpha_1 - \hat{\alpha}_1)^2)^{-n/2} \]

- Note that $p(\alpha_0, \alpha_1 | x, y) \neq p(\alpha_0 | x, y) p(\alpha_1 | x, y)$, so a posteriori $\alpha_0$ and $\alpha_1$ are not independent, but they are uncorrelated because $\text{cov}[\alpha_0, \alpha_1 | x, y] = 0$.

- If we come back to $\beta_0$ and $\beta_1$:

\[ p(\beta_0, \beta_1 | x, y) \propto \left( S_e + n(\beta_0 - \hat{\beta}_0)^2 + 2n\bar{x}(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) + S_x(\beta_1 - \hat{\beta}_1)^2 \right)^{-n/2} \]

\[ \hat{\beta}_0 = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{x}, \quad \hat{\beta}_1 = \hat{\alpha}_1 \]

- Extensions:
  - Errors on $x_i$
  - Errors on both $y_i$ and $x_i$
  - More general regressions (Linear and Non-Linear)
General Linear Model

- General Linear Model: \( y = A\theta + \epsilon \)
- Likelihood

\[
p(y|\theta) \propto v^{-n} \exp \left\{ -\frac{1}{2v}(y - A\theta)'(y - A\theta) \right\} \\
\propto v^{-n} \exp \left\{ -\frac{S^2}{2v} \right\}
\]

- Maximum Likelihood:

\[
\hat{\theta} = (A' A)^{-1} A' y
\]

- A simple calculation:

\[
S = (y - A\theta)'(y - A\theta) \\
= (\hat{\theta} - \theta)'(\hat{\theta} - \theta) + (y - A\hat{\theta})'(y - A\hat{\theta}) \\
= S_l(\theta) + S_e
\]

- Flat prior on \( \theta \) and Jeffrey's Prior on \( v \)

\[
p(\theta, v|y) \propto v^{-(n+1)} \exp \left\{ -\frac{1}{2v}S_l(\theta) + S_e \right\} \\
\propto v^{-n} \exp \left\{ -\frac{S}{2v} \right\}
\]
General Linear Model

- Marginalizing over all $\theta_i$:

$$
p(\theta_{-i}, v|y) \propto v^{-(n)} \exp \left\{ -\frac{1}{2v} S_l(\theta_{-i}) + S_e \right\}
\propto v^{-n} \exp \left\{ -\frac{S}{2v} \right\}
$$

- Each marginalization over each $\theta_i$ loses one power of $v$.
- Marginalization over all $k - 1$ variables $\theta_{-i}$ and then over $v$, we get a $S_t$ distribution with $n - k$ degrees of freedom where

$$
t = \frac{\theta_k - \hat{\theta}_k}{s \sqrt{m_{kk}}}, \quad s^2 = \frac{S_e}{\nu}
$$

and $m_{kk}$ is the $(kk)$ element of the design matrix $M = (A' A)^{-1}$

- The posterior marginal of $v$ is Inverse Gamma
General Linear Model

- Informative prior: \( p(\theta|\theta_0, V_0) \)

\[
p(\theta, v|y) \propto v^{-(n+1)} \exp \left\{ -\frac{1}{2v} S_t(\theta) + S_e \right\} p(\theta|\theta_0, V_0) \\
\propto v^{-n} \exp \left\{ -\frac{S}{2v} - (\theta - \theta_0)'V_0^{-1}(\theta - \theta_0) \right\}
\]

- When \( v \) is known:

\[
\hat{\theta} = \left( \frac{1}{v} A' A + V_0 \right)^{-1} \left( \frac{1}{v} A' y + V_0 \theta_0 \right)
\]

- One can continue other calculations

- More extentions:
  - Integrate out \( v \rightarrow S_t \) distribution
  - Laplace (Gaussian) approximation of \( t \) distribution around its maximum \( \hat{\theta} \)
  - Prediction, ...
Preliminaries on Bayesian inference

- Probabilistic model: $\mathcal{M} : g \sim p(g|\theta; \mathcal{M})$
- Frequentist view: $\theta$ unknown "fixed" parameters
- Maximum Likelihood:

  $$\hat{\theta} = \arg \max_{\theta} \{ p(g|\theta; \mathcal{M}) \}$$

- Bayesian approach: Probabilistic Prior Information $\theta \sim p(\theta)$

  $$p(\theta|g; \mathcal{M}) = \frac{p(g|\theta; \mathcal{M}) p(\theta|\mathcal{M})}{p(g|\mathcal{M})}$$

  $$p(g|\mathcal{M}) = \int p(g|\theta; \mathcal{M}) p(\theta|\mathcal{M}) \, d\theta$$

- Infer on $\theta$ using the posterior $p(\theta|g; \mathcal{M})$
Model selection

- **Frequentist view:** Likelihood ratio:

\[
\frac{p(g|\theta; M_1)}{p(g|\theta; M_2)} \quad \text{or} \quad \frac{p(g|\hat{\theta}; M_1)}{p(g|\hat{\theta}; M_2)} \quad \text{or} \quad \frac{p(g|M_1)}{p(g|M_2)}
\]

- **Bayesian view:**

\[
p(M_1|\theta, g) \propto p(g|\theta; M_1) \, p(\theta|M_1) \, P(M_1)
\]

\[
p(M_1|g) \propto p(g|M_1) \, P(M_1)
\]

\[
p(M_1|g, \theta) \quad \text{or} \quad \frac{p(M_1|g, \hat{\theta})}{p(M_2|g, \hat{\theta})} \quad \text{or} \quad \frac{p(M_1|g)}{p(M_2|g)}
\]
2. Bayesian inference for inverse problems

Example: Measuring variation of temperature with a thermometer

- $f(t)$ variation of temperature over time
- $g(t)$ variation of length of the liquid in thermometer

Forward model $\mathcal{M}$: Convolution

$$g(t) = \int f(t') h(t - t') \, dt' + \epsilon(t)$$

$h(t)$: impulse response of the measurement system

Inverse problem: Deconvolution

Given the forward model $\mathcal{M}$ (impulse response $h(t)$) and a set of data $g(t_i), i = 1, \ldots, M$ find $f(t)$
Measuring variation of temperature with a thermometer

Forward model: Convolution

\[ g(t) = \int f(t') h(t - t') \, dt' + \epsilon(t) \]

Inversion: Deconvolution
Convolution/Deconvolution: Discrete form

\[ g(t) = h(t) \ast f(t) + \epsilon(t) = \int h(t') f(t - t') \, dt' + \epsilon(t) \]

\[ g(m) = \sum_{k=-q}^{p} h(k) f(m - k) + \epsilon(m), \quad m = 0, \cdots, M \]

Matrix-Vector form: \( \mathbf{g} = A \mathbf{f} + \mathbf{\epsilon} \)
\[
\begin{bmatrix}
  g(0) \\
g(1) \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  g(M)
\end{bmatrix}
= \begin{bmatrix}
  h(0) \\
  \vdots \\
  h(1) \\
  \vdots \\
  h(p) \\
  \vdots \\
  0 \\
  \vdots \\
  \vdots \\
  0 \\
  \vdots \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\begin{bmatrix}
  f(0) \\
  \vdots \\
  f(1) \\
  \vdots \\
  f(p) \\
  \vdots \\
  0 \\
  \vdots \\
  \vdots \\
  0 \\
  \vdots \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\]
Computed tomography (CT)

A Multislice CT Scanner

\[ g(s_i) = \int_{L_i} f(r) \, dl_i + \epsilon(s_i) \]

Discretization

\[ g = Af + \epsilon \]
Computed Tomography

\[
g(r, \phi) = \int_L f(x, y) \, dl
\]

\[
f(x, y) = \sum_j f_j b_j(x, y)
\]

\[
b_j(x, y) = \begin{cases} 
1 & \text{if } (x, y) \in \text{pixel } j \\
0 & \text{else}
\end{cases}
\]

\[
g_i = \sum_{j=1}^N A_{ij} f_j + \epsilon_i
\]

\[
g = Af + \epsilon
\]
Computed Tomography with two projections

\[ f_{ij} \]

\[ \sum_{j=1}^{n} f_{ij} = g_{1i}, \quad i = 1, \ldots, m \]

\[ \sum_{i=1}^{m} f_{ij} = g_{2j}, \quad j = 1, \ldots, n \]
Application of CT in NDT

Reconstruction from only 2 projections

\[ g_1(x) = \int f(x, y) \, dy, \quad g_2(y) = \int f(x, y) \, dx \]

- Given the marginals \( g_1(x) \) and \( g_2(y) \) find the joint distribution \( f(x, y) \).
- Infinite number of solutions: \( f(x, y) = g_1(x) g_2(y) \Omega(x, y) \)

\( \Omega(x, y) \) is a Copula:

\[ \int \Omega(x, y) \, dx = 1 \quad \text{and} \quad \int \Omega(x, y) \, dy = 1 \]
\[
\begin{bmatrix}
g_1 \\
g_2 \\
\vdots \\
g_8
\end{bmatrix}
= 
\begin{bmatrix}
1000100010001000 \\
0100100010001000 \\
0010001000100010 \\
0001000100010001 \\
0000000000001111 \\
0000000111100000 \\
0001111000000000 \\
1111000000000000
\end{bmatrix}
\begin{bmatrix}
f_1 \\
f_2 \\
\vdots \\
f_{16}
\end{bmatrix}
\]

\[ g = Af \]

- **Forward problem**: Given \( f \) compute \( g \)
- **Inverse problem**: Given \( g \) estimate \( f \)
  - Existance
  - Uniqueness
  - Stability
### Existance and Uniqueness:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>5</td>
<td>15</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>15</td>
<td>5</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>15</td>
<td>5</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>5</td>
<td>15</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stability:

\[
A = \begin{bmatrix}
10 & 7 & 8 & 7 \\
7 & 5 & 6 & 5 \\
8 & 6 & 10 & 9 \\
7 & 5 & 9 & 10 \\
\end{bmatrix}
\]

\[
A^{-1} = \begin{bmatrix}
25 & -41 & 10 & -6 \\
-41 & 68 & -17 & 10 \\
10 & -17 & 5 & -3 \\
-6 & 10 & -3 & 2 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
10 & 7 & 8 & 7 \\
7 & 5 & 6 & 5 \\
8 & 6 & 10 & 9 \\
7 & 5 & 9 & 10 \\
\end{bmatrix}
\begin{bmatrix}
f_1 \\
f_2 \\
f_3 \\
f_4 \\
\end{bmatrix} = \begin{bmatrix}
32 \\
23 \\
33 \\
31 \\
\end{bmatrix} \quad \rightarrow \quad f = \begin{bmatrix}
1 \\
1 \\
1 \\
1 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
10 & 7 & 8 & 7 \\
7 & 5 & 6 & 5 \\
8 & 6 & 10 & 9 \\
7 & 5 & 9 & 10 \\
\end{bmatrix}
\begin{bmatrix}
f_1 + \delta f_1 \\
f_2 + \delta f_2 \\
f_3 + \delta f_3 \\
f_4 + \delta f_4 \\
\end{bmatrix} = \begin{bmatrix}
32.1 \\
22.9 \\
33.1 \\
30.9 \\
\end{bmatrix} \quad \rightarrow \quad f + \delta f = \begin{bmatrix}
9.2 \\
-12.6 \\
4.5 \\
-1.1 \\
\end{bmatrix}
\]

\[
\frac{||\delta g||}{||g||} = \frac{1}{300} \quad \rightarrow \quad \frac{||\delta f||}{||f||} = \frac{10}{1}
\]

\[
\frac{||\delta f||}{||f||} = \text{cond}(A) \frac{||\delta g||}{||g||}
\]
Bayesian inference for inverse problems

- Linear Inverse problems: \( g = A f + \epsilon \)

- Bayesian inference:

\[
p(f | g, \theta) = \frac{p(g | f, \theta_1) p(f | \theta_2)}{p(g | \theta)}
\]

with \( \theta = (\theta_1, \theta_2) \)

- Point estimators:
  
  - Maximum A Posteriori (MAP)
    
    \[
    \hat{f} = \text{arg max}_f \{p(f | g, \theta)\}
    \]
  
  - Posterior Mean (PM)
    
    \[
    \hat{f} = E_{p(f | g, \theta)} \{f\} = \int f \ p(f | g, \theta) \, df
    \]
Simple Bayesian Inference: Known hyperparameters $\theta$

\[ f \xrightarrow{A} g \]

Forward model

\[ p(f|\theta_2) \quad \diamond \quad p(g|f, \theta_1) \quad \rightarrow \quad p(f|g, \theta) \]

Prior \quad \text{Likelihood} \quad \text{Posterior}

\[ p(f|g, \theta) = \frac{p(g|f, \theta_1) p(f|\theta_2)}{p(g|\theta)} \]
Case of linear models and Gaussian priors

\[ g = Af + \epsilon \]

- Hypothesis on the noise: \( \epsilon \sim \mathcal{N}(0, \sigma^2_\epsilon I) \)
  \[ p(g|f) \propto \exp \left\{ -\frac{1}{2\sigma^2_\epsilon} \| g - Af \|^2 \right\} \]

- Hypothesis on \( f \): \( f \sim \mathcal{N}(0, \sigma^2_f (D^t D)^{-1}) \)
  \[ p(f) \propto \exp \left\{ -\frac{1}{2\sigma^2_f} \| Df \|^2 \right\} \]

- A posteriori:
  \[ p(f|g) \propto \exp \left\{ -\frac{1}{2\sigma^2_\epsilon} \| g - Af \|^2 - \frac{1}{2\sigma^2_f} \| Df \|^2 \right\} \]

- MAP:
  \[ \hat{f} = \arg \max_f \{ p(f|g) \} = \arg \min_f \{ J(f) \} \]
  with \( J(f) = \| g - Af \|^2 + \lambda \| Df \|^2 \), \( \lambda = \frac{\sigma^2_\epsilon}{\sigma^2_f} \)

- Advantage: characterization of the solution
  \[ f|g \sim \mathcal{N}(\hat{f}, \hat{P}) \text{ with } \hat{f} = \hat{P} A' g, \quad \hat{P} = \left( A'A + \lambda D^t D \right)^{-1} \]
Simple Bayesian Model and Estimation

- Example 1: Linear Gaussian case

\[
\begin{align*}
p(g|f, \theta_1) &= \mathcal{N}(Af, \theta_1 I) \\
p(f|\theta_2) &= \mathcal{N}(0, \theta_2 I)
\end{align*}
\rightarrow p(f|g, \theta) = \mathcal{N}(\hat{f}, \hat{P})
\]

with

\[
\begin{align*}
\hat{P} &= (A'A + \lambda I)^{-1}, \quad \lambda = \frac{\theta_1}{\theta_2} \\
\hat{f} &= \hat{P} A'g
\end{align*}
\]

\[
\hat{f} = \arg \min_f \{ J(f) \} \quad \text{with} \quad J(f) = \|g - Af\|_2^2 + \lambda \|f\|_2^2
\]

- Example 2: Gaussian noise, Double Exponential prior & MAP:

\[
\hat{f} = \arg \min_f \{ J(f) \} \quad \text{with} \quad J(f) = \|g - Af\|_2^2 + \lambda \|f\|_1
\]
MAP estimation with other priors:

\[
\hat{f} = \arg \min_f \{ J(f) \} \quad \text{with} \quad J(f) = \| g - Af \|^2 + \lambda \Omega(f)
\]

Separable priors:

- **Gaussian:** 
  
  \[
p(f_j) \propto \exp \left\{ -\alpha |f_j|^2 \right\} \quad \rightarrow \quad \Omega(f) = \alpha \sum_j |f_j|^2
  \]

- **Gamma:** 
  
  \[
p(f_j) \propto f_j^\alpha \exp \left\{ -\beta f_j \right\} \quad \rightarrow \quad \Omega(f) = \alpha \sum_j \ln f_j + \beta f_j
  \]

- **Beta:** 
  
  \[
p(f_j) \propto f_j^\alpha (1 - f_j)^\beta \quad \rightarrow \quad \Omega(f) = \alpha \sum_j \ln f_j + \beta \sum_j \ln(1 - f_j)
  \]

- **Generalized Gaussian:** 
  
  \[
p(f_j) \propto \exp \left\{ -\alpha |f_j|^\beta \right\}, \quad 1 < \beta < 2 \quad \rightarrow \quad \Omega(f) = \alpha \sum_j |f_j|^\beta,
  \]

Markovian models:

\[
p(f_j | f) \propto \exp \left\{ -\alpha \sum_{i \in N_j} \phi(f_j, f_i) \right\} \quad \rightarrow \quad \Omega(f) = \alpha \sum_j \sum_{i \in N_j} \phi(f_j, f_i),
\]
Different prior models for signals and images: Separable

Gaussian

\[ p(f_j) \propto \exp\{-\alpha |f_j|^2\} \]

Generalized Gaussian

\[ p(f_j) \propto \exp\{-\alpha |f_j|^p\}, \quad 1 \leq p \leq 2 \]

Gamma

\[ p(f_j) \propto f_j^\alpha \exp\{-\beta f_j\} \]

Beta

\[ p(f_j) \propto f_j^\alpha (1 - f_j)^\beta \]
Different prior models: Simple Markovian

\[ p(f_j | f) \propto \exp \left\{ -\alpha \sum_{i \in v_j} \phi(f_j, f_i) \right\} \rightarrow \Phi(f) = \alpha \sum_j \sum_{i \in V_j} \phi(f_j, f_i) \]

- 1D case and one neighbor \( V_j = j - 1 \):
  \[ \Phi(f) = \alpha \sum_j \phi(f_j - f_{j-1}) \]

- 1D Case and two neighbors \( V_j = \{j - 1, j + 1\} \):
  \[ \Phi(f) = \alpha \sum_j \phi(f_j - \beta(f_{j-1} + f_{j-1})) \]

- 2D case with 4 neighbors:
  \[ \Phi(f) = \alpha \sum_{\mathbf{r} \in \mathcal{R}} \phi \left( f(\mathbf{r}) - \beta \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} f(\mathbf{r}') \right) \]

- \( \phi(t) = |t|^\gamma \): Generalized Gaussian
Different prior models: Markovian with hidden variables

Piecewise Gaussians
(contours hidden variables)

\[ p(f_j|q_j, f_{j-1}) = \mathcal{N} \left( (1 - q_j)f_{j-1}, \sigma_f^2 \right) \]

Mixture of Gaussians (MoG)
(regions labels hidden variables)

\[ p(f_j|z_j = k) = \mathcal{N} \left( m_k, \sigma_k^2 \right) \]

\[ p(f|q) \propto \exp \left\{ -\alpha \sum_j |f_j - (1 - q_j)f_{j-1}|^2 \right\} \]

\[ p(f|z) \propto \exp \left\{ -\alpha \sum_k \sum_{j \in R_k} \left( \frac{f_j - m_k}{\sigma_k} \right)^2 \right\} \]
Gauss-Markov-Potts prior models for images

\[
f(r)
\]

\[
z(r)
\]

\[
c(r) = 1 - \delta(z(r) - z(r'))
\]

\[
p(f(r)|z(r) = k, m_k, v_k) = \mathcal{N}(m_k, v_k)
\]

\[
p(f(r)) = \sum_k P(z(r) = k) \mathcal{N}(m_k, v_k)
\]

Mixture of Gaussians

- Separable iid hidden variables: \( p(z) = \prod_r p(z(r)) \)
- Markovian hidden variables: \( p(z) \) Potts-Markov:

\[
p(z(r)|z(r'), r' \in \mathcal{V}(r)) \propto \exp \left\{ \gamma \sum_{r' \in \mathcal{V}(r)} \delta(z(r) - z(r')) \right\}
\]

\[
p(z) \propto \exp \left\{ \gamma \sum_{r \in \mathcal{R}} \sum_{r' \in \mathcal{V}(r)} \delta(z(r) - z(r')) \right\}
\]
Four different cases

To each pixel of the image is associated 2 variables $f(r)$ and $z(r)$

- $f|z$ Gaussian iid, $z$ iid : Mixture of Gaussians
- $f|z$ Gauss-Markov, $z$ iid : Mixture of Gauss-Markov
- $f|z$ Gaussian iid, $z$ Potts-Markov : Mixture of Independent Gaussians (MIG with Hidden Potts)
- $f|z$ Markov, $z$ Potts-Markov : Mixture of Gauss-Markov (MGM with hidden Potts)
Full Bayesian, Joint MAP, Marginalization

- Unknown hyperparameters $\theta$
- Full Bayesian: Joint Posterior:

$$ p(f, \theta | g) \propto p(g | f, \theta_1) p(f | \theta_2) p(\theta) $$

- Joint MAP:

$$ (\hat{f}, \hat{\theta}) = \arg \max_{(f, \theta)} \{ p(f, \theta | g) \} $$

- Iterative algorithm:

$$ \begin{cases} 
\hat{f}^{(k)} = \arg \max_f \left\{ p(f | \hat{\theta}^{(k-1)}, g) \right\} \\
\hat{\theta}^{(k)} = \arg \max_{\theta} \left\{ p(\theta | \hat{f}^{(k-1)}, g) \right\} 
\end{cases} $$
Full Bayesian Model and Hyperparameter Estimation

↓ $\alpha, \beta$

Hyper prior model $p(\theta | \alpha, \beta)$

$\theta_2$ $\theta_1$

$p(f | \theta_2)$ $\diamond$ $p(g | f, \theta_1)$ $\rightarrow$ $p(f, \theta | g, \alpha, \beta)$

Prior Likelihood Joint Posterior

Full Bayesian Model and Hyperparameter Estimation scheme

$p(f, \theta | g)$ $\rightarrow$ $p(\theta | g)$ $\rightarrow$ $\hat{\theta}$ $\rightarrow$ $p(f | \hat{\theta}, g)$ $\rightarrow$ $\hat{f}$

Joint Posterior Marginalize over $f$

Marginalization for Hyperparameter Estimation
Full Bayesian: Marginal MAP and PM estimates

- Marginal MAP: $\hat{\theta} = \arg \max_{\theta} \{p(\theta | g)\}$ where

$$p(\theta | g) = \int p(f, \theta | g) \, df \propto p(g | \theta) p(\theta)$$

and then $\hat{f} = \arg \max_{f} \{p(f | \hat{\theta}, g)\}$ or

Posterior Mean: $\hat{f} = \int f \, p(f | \hat{\theta}, g) \, df$

- Needs the expression of the Likelihood:

$$p(g | \theta) = \int p(g | f, \theta_1) p(f | \theta_2) \, df$$

Not always analytically available $\rightarrow$ EM, SEM and GEM algorithms
Full Bayesian approach

\[ g = A f + \epsilon \]

- Forward & errors model: \( \rightarrow p(g|f, \theta_1; \mathcal{M}) \)
- Prior models \( \rightarrow p(f|\theta_2; \mathcal{M}) \)
- Hyperparameters \( \theta = (\theta_1, \theta_2) \rightarrow p(\theta|\mathcal{M}) \)
- Bayes: \( \rightarrow p(f, \theta|g; \mathcal{M}) = \frac{p(g|f, \theta; \mathcal{M}) p(f|\theta; \mathcal{M}) p(\theta|\mathcal{M})}{p(g|\mathcal{M})} \)
- Joint MAP: \( (\hat{f}, \hat{\theta}) = \arg\max_{(f, \theta)} \{p(f, \theta|g; \mathcal{M})\} \)
- Marginalization:
  \[
  \begin{align*}
  p(f|g; \mathcal{M}) &= \int p(f, \theta|g; \mathcal{M}) \, df \\
  p(\theta|g; \mathcal{M}) &= \int p(f, \theta|g; \mathcal{M}) \, d\theta
  \end{align*}
  \]
- Posterior means:
  \[
  \begin{align*}
  \hat{f} &= \int \int f \, p(f, \theta|g; \mathcal{M}) \, df \, d\theta \\
  \hat{\theta} &= \int \int \theta \, p(f, \theta|g; \mathcal{M}) \, df \, d\theta
  \end{align*}
  \]
- Evidence of the model:

\[
p(g|\mathcal{M}) = \int\int p(g|f, \theta; \mathcal{M}) p(f|\theta; \mathcal{M}) p(\theta|\mathcal{M}) \, df \, d\theta
\]
Full Bayesian: EM and GEM algorithms

- EM and GEM Algorithms: \( f \) as hidden variable, \( g \) as incomplete data, \((g, f)\) as complete data
  - \( \ln p(g|\theta) \) incomplete data log-likelihood
  - \( \ln p(g, f|\theta) \) complete data log-likelihood

- Iterative algorithm:
  
  \[
  \begin{align*}
  \text{E-step:} & \quad Q(\theta, \hat{\theta}^{(k)}) = \mathbb{E}_{p(f|g, \hat{\theta}^{(k)})} \left\{ \ln p(g, f|\theta) \right\} \\
  \text{M-step:} & \quad \hat{\theta}^{(k)} = \arg \max_{\theta} \left\{ Q(\theta, \hat{\theta}^{(k-1)}) \right\}
  \end{align*}
  \]

- GEM (Bayesian) algorithm:
  
  \[
  \begin{align*}
  \text{E-step:} & \quad Q(\theta, \hat{\theta}^{(k)}) = \mathbb{E}_{p(f|g, \hat{\theta}^{(k)})} \left\{ \ln p(g, f|\theta) + \ln p(\theta) \right\} \\
  \text{M-step:} & \quad \hat{\theta}^{(k)} = \arg \max_{\theta} \left\{ Q(\theta, \hat{\theta}^{(k-1)}) \right\}
  \end{align*}
  \]
Variational Bayesian Approximation

- Approximate \( p(f, \theta | g) \) by \( q(f, \theta | g) = q_1(f | g) q_2(\theta | g) \) and then continue computations.
- Criterion \( \text{KL}(q(f, \theta | g) : p(f, \theta | g)) \)
- Iterative algorithm \( q_1 \rightarrow q_2 \rightarrow q_1 \rightarrow q_2, \cdots \)

\[
\begin{cases}
\hat{q}_1(f) \propto \exp \left\{ \langle \ln p(g, f, \theta; M) \rangle_{\hat{q}_2(\theta)} \right\} \\
\hat{q}_2(\theta) \propto \exp \left\{ \langle \ln p(g, f, \theta; M) \rangle_{\hat{q}_1(f)} \right\}
\end{cases}
\]

\[ p(f, \theta | g) \rightarrow \text{Variational Bayesian Approximation} \rightarrow q_1(f) \rightarrow \hat{f} \]

\[ \rightarrow q_2(\theta) \rightarrow \hat{\theta} \]
Two main steps in the Bayesian approach

- Prior modeling
  - Separable:
    Gaussian, Generalized Gaussian, Gamma, mixture of Gaussians, mixture of Gammas, ...
  - Markovian: Gauss-Markov, GGM, ...
  - Separable or Markovian with hidden variables (contours, region labels)

- Choice of the estimator and computational aspects
  - MAP, Posterior mean, Marginal MAP
  - MAP needs optimization algorithms
  - Posterior mean needs integration methods
  - Marginal MAP needs integration and optimization
  - Approximations:
    - Gaussian approximation (Laplace)
    - Numerical exploration MCMC
    - Variational Bayes (Separable approximation)
Different prior models for signals and images: Separable

**Gaussian**
\[ p(f_j) \propto \exp\left\{-\alpha|f_j|^2\right\} \]

**Generalized Gaussian**
\[ p(f_j) \propto \exp\left\{-\alpha|f_j|^p\right\}, \quad 1 \leq p \leq 2 \]

**Gamma**
\[ p(f_j) \propto f_j^\alpha \exp\left\{-\beta f_j\right\} \]

**Beta**
\[ p(f_j) \propto f_j^\alpha (1 - f_j)^\beta \]
2. Sparsity enforcing prior models

- **Simple heavy tailed models:**
  - Generalized Gaussian, Double Exponential
  - Symmetric Weibull, Symmetric Rayleigh
  - Student-t, Cauchy
  - Generalized hyperbolic
  - Elastic net

- **Hierarchical mixture models:**
  - Mixture of Gaussians
  - Bernoulli-Gaussian
  - Mixture of Gammas
  - Bernoulli-Gamma
  - Mixture of Dirichlet
  - Bernoulli-Multinomial
Simple heavy tailed models

- Generalized Gaussian, Double Exponential

\[ p(f | \gamma, \beta) = \prod_j g_G(f_j | \gamma, \beta) \propto \exp \left\{ -\gamma \sum_j |f_j|^\beta \right\} \]

\( \beta = 1 \) Double exponential or Laplace.
\( 0 < \beta \leq 1 \) are of great interest for sparsity enforcing.
Simple heavy tailed models

- Symmetric Weibull

\[ p(f|\gamma, \beta) = \prod_j \mathcal{W}(f_j|\gamma, \beta) \propto \exp \left\{ -\gamma \sum_j |f_j|^\beta + (\beta - 1) \log |f_j| \right\} \]

\( \beta = 2 \) is the Symmetric Rayleigh distribution.

\( \beta = 1 \) is the Double exponential and

\( 0 < \beta \leq 1 \) are of great interest for sparsity enforcing.
Simple heavy tailed models

- Student-t and Cauchy models

\[ p(f|\nu) = \prod_j St(f_j|\nu) \propto \exp \left\{ -\frac{\nu + 1}{2} \sum_j \log \left( 1 + \frac{f_j^2}{\nu} \right) \right\} \]

Cauchy model is obtained when \( \nu = 1 \).
Simple heavy tailed models

• Elastic net prior model

\[
p(f|\nu) = \prod_j \mathcal{E}\mathcal{N}(f_j|\nu) \propto \exp \left\{ - \sum_j (\gamma_1 |f_j| + \gamma_2 f_j^2) \right\}
\]
Simple heavy tailed models

- Generalized hyperbolic (GH) models

\[ p(f|\delta, \nu, \beta) = \prod_j \left( \delta^2 + f_j^2 \right)^{\nu-1/2}/2 \exp \{ \beta x \} \cdot K_{\nu-1/2}(\alpha \sqrt{\delta^2 + f_j^2}) \]
Mixture models

- Mixture of two Gaussians (MoG2) model

\[
p(f|\lambda, v_1, v_0) = \prod_j \left[ \lambda \mathcal{N}(f_j|0, v_1) + (1 - \lambda) \mathcal{N}(f_j|0, v_0) \right]
\]

- Bernoulli-Gaussian (BG) model

\[
p(f|\lambda, v) = \prod_j p(f_j) = \prod_j \left[ \lambda \mathcal{N}(f_j|0, v) + (1 - \lambda) \delta(f_j) \right]
\]
- Mixture of Gammas

\[ p(f | \lambda, v_1, v_0) = \prod_j \left[ \lambda G(f_j | \alpha_1, \beta_1) + (1 - \lambda) G(f_j | \alpha_2, \beta_2) \right] \]

- Bernoulli-Gamma model

\[ p(f | \lambda, \alpha, \beta) = \prod_j \left[ \lambda G(f_j | \alpha, \beta) + (1 - \lambda) \delta(f_j) \right] \]
• Mixture of Dirichlets model

\[ p(f|\lambda, a_1, \alpha_1, a_2, \alpha_2) = \prod_j [\lambda D(f_j|a_1, \alpha_1) + (1 - \lambda) D(f_j|a_2, \alpha_2)] \]

where

\[ D(f_j|a, \alpha) = \prod_{k=1}^{K} \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)\Gamma(\alpha_K)} a_k^{\alpha_k - 1}, \quad \alpha_k \geq 0, \quad a_k \geq 0 \]

where \( a = \{a_1, \ldots, a_K\} \) and \( \alpha = \{\alpha_1, \ldots, \alpha_K\} \)

with \( \sum_k \alpha_k = \alpha \) and \( \sum_k a_k = 1 \).

• Bernoulli-Multinomial (BMultinomial) model

\[ p(f|\lambda, a, \alpha) = \prod_j [\lambda \delta(f_j) + (1 - \lambda) Mult(f_j|a, \alpha)] \]
Hierarchical models and hidden variables

- All the mixture models and some of simple models can be modeled via hidden variables $z$.

- Example 1: MoG model:

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

$$\begin{cases} 
p(f_j | z_j = 0, v_0) = \mathcal{N}(f_j | 0, v_0), \\
p(f_j | z_j = 1, v_1) = \mathcal{N}(f_j | 0, v_1),
\end{cases} \quad \text{and} \quad \begin{cases} 
P(z_j = 0) = \lambda, \\
P(z_j = 1) = 1 - \lambda
\end{cases}$$

$$\begin{align*}
p(f | z) &= \prod_j p(f_j | z_j) = \prod_j \mathcal{N}(f_j | 0, v_{z_j}) \\&\propto \exp \left\{-\frac{1}{2} \sum_j \frac{f_j^2}{v_{z_j}}\right\}
\end{align*}$$

$$P(z_j = 1) = \lambda, \quad P(z_j = 0) = 1 - \lambda$$
Hierarchical models and hidden variables

Example 2: Student-t model

\[ St(f|\nu) \propto \exp \left\{ -\frac{\nu + 1}{2} \log (1 + f^2/\nu) \right\} \]

\[ = \int_0^\infty N(f|0, 1/z) \mathcal{G}(z|\alpha, \beta) \, dz, \quad \text{with } \alpha = \beta = \nu/2 \]

\[
\begin{align*}
p(f|z) &= \prod_j p(f_j|z_j) = \prod_j N(f_j|0, 1/z_j) \propto \exp \left\{ -\frac{1}{2} \sum_j z_j f_j^2 \right\} \\
p(z|\alpha, \beta) &= \prod_j \mathcal{G}(z_j|\alpha, \beta) \propto \prod_j z_j^{(\alpha-1)} \exp \left\{ -\beta z_j \right\} \\
&\quad \propto \exp \left\{ \sum_j (\alpha - 1) \ln z_j - \beta z_j \right\} \\
p(f, z|\alpha, \beta) &\propto \exp \left\{ -\frac{1}{2} \sum_j z_j f_j^2 + (\alpha - 1) \ln z_j - \beta z_j \right\}
\end{align*}
\]
Hierarchical models and hidden variables

Example 3: Laplace (Double Exponential) model

\[ D \mathcal{E}(f|a) = \frac{a}{2} \exp \{-a|f|\} = \int_0^\infty \mathcal{N}(f|0, z) \mathcal{E}(z|a^2/2) \, dz, \quad a > 0 \]

\[
\begin{align*}
p(f|z) &= \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0, z_j) \propto \exp \left\{ -\frac{1}{2} \sum_j f_j^2 / z_j \right\} \\
p(z|a^2/2) &= \prod_j \mathcal{E}(z_j|a^2/2) \propto \exp \left\{ \sum_j \frac{a^2}{2} z_j \right\} \\
p(f, z|a^2/2) &\propto \exp \left\{ -\frac{1}{2} \sum_j f_j^2 / z_j + \frac{a^2}{2} z_j \right\}
\end{align*}
\]

With these models we have:

\[ p(f, z, \theta|g) \propto p(g|f, \theta_1) \, p(f|z, \theta_2) \, p(z|\theta_3) \, p(\theta) \]
Bayesian Computation and Algorithms

- Often, the expression of $p(f, z, \theta|g)$ 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 $p(f|z, \theta, g)$, $p(z|f, \theta, g)$, and $p(\theta|f, z, g)$.

**VBA:** Approximate $p(f, z, \theta|g)$ by a separable one

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

and do any computations with these separable ones.
Summary of Bayesian estimation with different levels

\[ p(f|\theta_2) \quad \diamondsuit \quad p(g|f, \theta_1) \quad \rightarrow \quad p(f|g, \theta) \rightarrow \hat{f} \]

\[ \text{Prior} \quad \cdot \quad \text{Likelihood} \quad \rightarrow \quad \text{Posterior} \]

Simple Bayesian Model and Estimation

\[ \downarrow \alpha, \beta \]

\[ \text{Hyper prior model } p(\theta|\alpha, \beta) \]

\[ \theta_2 \quad \downarrow \quad \theta_1 \]

\[ p(f|\theta_2) \quad \diamondsuit \quad p(g|f, \theta_1) \quad \rightarrow \quad p(f, \theta|g, \alpha, \beta) \rightarrow \hat{f}, \hat{\theta} \]

\[ \text{Prior} \quad \cdot \quad \text{Likelihood} \quad \rightarrow \quad \text{Joint Posterior} \]

Full Bayesian Model and Hyperparameter Estimation scheme
Summary of Bayesian estimation with different levels

$$p(f, \theta|g) \rightarrow p(\theta|g) \rightarrow \hat{\theta} \rightarrow p(f|\hat{\theta}, g) \rightarrow \hat{f}$$

Joint Posterior  Marginalize over $f$

Marginalization for Hyperparameter Estimation

$$\theta_3 \downarrow p(z|\theta_3) \diamond p(f|z, \theta_2) \diamond p(g|f, \theta_1) \rightarrow p(f, z|g, \theta) \Rightarrow \hat{f}$$

Hidden variable  Prior  Likelihood  Joint Posterior

Full Bayesian Model with a Hierarchical Prior Model
Summary of Bayesian estimation with different levels

↓ \( \alpha, \beta \)

Hyper prior model \( p(\theta|\alpha, \beta) \)

\( \theta_3 \) \ □  \( \theta_2 \) \ □  \( \theta_1 \)  \ □

\( p(z|\theta_3) \)  \ □  \( p(f|z, \theta_2) \)  \ □  \( p(g|f, \theta_1) \)  \ □

Hidden variable  Prior  Likelihood  Joint Posterior

Full Bayesian Hierarchical Model with Hyperparameter Estimation

↓ \( \alpha, \beta \)

Hyper prior model \( p(\theta|\alpha, \beta) \)

\( \theta_3 \) \ □  \( \theta_2 \) \ □  \( \theta_1 \)  \ □

\( p(z|\theta_3) \)  \ □  \( p(f|z, \theta_2) \)  \ □  \( p(g|f, \theta_1) \)  \ □

Hidden variable  Prior  Likelihood  Joint Posterior

Full Bayesian Hierarchical Model and Variational Approximation

VBA
\( q_1(f) \)  \ □  \( q_2(z) \)  \ □  \( q_3(\theta) \)  \ □

Separable Approximation

\( \hat{f} \)  \ □  \( \hat{z} \)  \ □  \( \hat{\theta} \)
3. Bayesian Inference for Sources Separation

- Source separation problem
  \[ f(t) \xrightarrow{\epsilon(t)} A f(t) + g(t) \quad t = 1, \ldots, T \]

- Stationary case
  \[ f \xrightarrow{\epsilon} A f + g \]

- Estimation of sources \( f \) when the mixing matrix \( A \) is known
- Estimation of the mixing matrix \( A \) when sources are known \( f \)
- Joint Estimation of the mixing matrix \( A \) and sources \( f \)

- Nonstationary case
Estimation of sources $f$ with known mixing matrix $A$

\[ g = Af \]

Exact Model without errors

\[ f = A^+ g = A'(AA')^{-1} g \]

Realistic Model with errors

\[ g = Af + \epsilon \]

Maximum Likelihood

\[ p(g|f, \epsilon) \]

Bayesian Estimation

\[ p(f|g, \epsilon) \propto p(g|f, \epsilon) p(f|\nu_f) \]
Estimation of sources $f$ with known mixing matrix $A$

\[ v_f \xrightarrow{A} f \]

Uncorrelated Gaussian
\[
p(f | v_f) = \mathcal{N}(0, v_f I) \\
\propto \exp \left\{ -\frac{1}{2v_f} \| f \|^2 \right\}
\]

\[ \hat{f} = \arg \min_f \{ J(f) \} \]
\[ J(f) = \| g - Af \|^2 + \lambda \| f \|^2 \]
\[ \hat{f} = (A'A + \lambda I)^{-1} A'g \]

Correlated Gaussian
\[
p(f | v_f \Sigma_f) = \mathcal{N}(0, v_f \Sigma_f) \\
\propto \exp \left\{ -\frac{1}{2v_f} \| Df \|^2 \right\}
\]
with $\Sigma_f = (D^t D)^{-1}$

\[ \hat{f} = \arg \min_f \{ J(f) \} \]
\[ J(f) = \| g - Af \|^2 + \lambda \| Df \|^2 \]
\[ \hat{f} = (A'A + \lambda D^t D)^{-1} A'g \]
Estimation of the mixing matrix $A$ with known sources $f$

- Bilinear problems:

$$g = Af = Fa$$

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

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

- Estimation of $f$ with known $A$: $g = Af$

- Estimation of $A$ with known $f$: $g = Af = Fa$

Underdetermination:
Needs constraints or prior information on $A$

- Joint Estimation of $f$ and $A$: $g = Af = Fa$

Underdetermination:
Needs constraints or prior information on $A$ and on $f$
Estimation of the mixing matrix $A$ with known sources $f$

$$g = Af = Fa$$

Exact Model without Errors

$$a = F^+ g = F^t (FF^t)^{-1} g$$

Realistic Model with errors

$$g = Af + \epsilon = Fa + \epsilon$$

Maximul Likelihood

$$p(g|a, \epsilon)$$

Bayesian Estimation

$$p(a|g, \epsilon) \propto p(g|a, \epsilon) p(a|v_a)$$
Estimation of the mixing matrix $A$ with known sources $f$

**Gaussian**

$p(a|v_f) = \mathcal{N}(0, v_a I)$

\[ \propto \exp \left\{ -\frac{1}{2v_a} \sum_j |a_j|^2 \right\} \]

\[ \hat{a} = \arg\min_a \{ J(a) \} \]

\[ J(a) = \| g - F a \|^2 + \lambda \sum_j |a_j|^2 \]

\[ \hat{a} = (F^t F + \lambda I)^{-1} F^t g \]

\[ \hat{A} = g f^t (f f^t + \lambda I)^{-1} \]

**Generalized Gaussian**

$p(a|\lambda, \beta) = \mathcal{GG}(\lambda, \beta)$

\[ \propto \exp \left\{ -\lambda \sum_j |a_j|^\beta \right\} \]

\[ \hat{a} = \arg\min_a \{ J(a) \} \]

\[ J(a) = \| g - F a \|^2 + \lambda \sum_j |a_j|^\beta \]

No analytic expression
Joint Estimation of $A$ and $f$

$g = Af = Fa$

Exact Model without errors

$g = Af + \epsilon = Fa + \epsilon$

Realistic Model with errors

- Indeterminations (scale and permutation)
- Needs constraints or prior information
- Example: Positivity $\rightarrow$ NNMF

$$(\hat{A}, \hat{f}) = \arg \min_{(A > 0, f > 0)} \{ \|g - Af\|^2 \}$$

$$\left\{ \begin{array}{l}
\hat{f}(t) = (\hat{A}'\hat{A})^{-1}\hat{A}'g(t) \\
\hat{A} = \sum_t g(t)\hat{f}'(t) \left( \sum_t \hat{f}(t)\hat{f}'(t) \right)^{-1}
\end{array} \right.$$  

Apply positivity $\hat{f} > 0$

Apply positivity $\hat{A} > 0$
Joint Estimation of $A$ and $f$ 

Maximul Likelihood

$$p(g(t)|A, f(t), v_\epsilon)$$

Bayesian Estimation

$$p(A, f(t)|g(t), v_\epsilon) \propto p(g(t)|A, f(t), v_\epsilon) \\ p(A|v_a) p(f(t)|v_f)$$

$$\begin{cases} 
\hat{f}(t) = (\hat{A}'\hat{A} + \lambda_f I)^{-1}\hat{A}'g(t), \\
\hat{A} = \sum_t g(t)\hat{f}'(t) \left(\sum_t \hat{f}(t)\hat{f}'(t) + \lambda_a I\right)^{-1} 
\end{cases}$$

$$\lambda_f = v_\epsilon/v_f$$

$$\lambda_a = v_\epsilon/v_a$$
Joint Estimation of $A$ and $f$

Joint Posterior:

$\mathcal{p}(A, f \mid g, \theta) \propto \mathcal{p}(g \mid A, f, v_\epsilon) \mathcal{p}(A \mid v_a) \mathcal{p}(f \mid v_f)$

Integration over $f$ can be done easily.
4. Links with classical methods: PCA, ICA, Neural Networks, ...

Inversion ou Decomposition

\[ g(t) = Af(t) + \epsilon(t) \]

Given that \( g(t) = Af(t) + \epsilon(t) \), estimate \( f(t) \) and \( A \)

- Maximum Likelihood
- Bayesian estimation

Given \( g(t) \), find \( \phi(.) \) and the separating matrix \( B \) such that \( y(t) \) has:

- Uncorrelated components (PCA)
- Independent components (ICA)

Link between \( y \) and \( \hat{f} \) and between \( B \) and \( \hat{A} \)?

Prior modeling
Links with classical methods: PCA

Classical PCA:

\[ g(t) = Af(t) \rightarrow \text{cov}[g] = A\text{cov}[f]A' \]

- Estimate \( \text{cov}[g] \) from the data: \( \text{cov}[g] = \frac{1}{T} \sum_t g(t)g'(t) \)
- Singular Value Decomposition (SVD): \( \text{cov}[g] = U\Lambda U' \)
- Identify \( A = U \) and \( \text{cov}[f] = \Lambda \)
  (Assumes sources to be uncorrelated)
- Now, given \( A \), estimate sources by \( f(t) = (A'A)^{-1}A'g(t) \)

- Indetermination:
  Note that any \( A = RU \) with \( R \) any orthogonal matrix
  is also a solution
Link with PCA

Probabilistic PCA:

\[ g(t) = Af(t) + \epsilon(t) \rightarrow \text{cov}[g] = A\text{cov}[f]A' + \text{cov}[\epsilon] \]

- \( \epsilon(t) \sim \mathcal{N}(0, \sigma^2_{\epsilon} I) \rightarrow p(g(t)|A, f(t), \sigma^2_{\epsilon}) = \mathcal{N}(Af(t), \sigma^2_{\epsilon} I) \)
- \( f(t) \sim \mathcal{N}(0, \Lambda) \)
- \( p(g(t)|A, \sigma^2_{\epsilon}, \Lambda) = \mathcal{N}(0, A\Lambda A' + \sigma^2_{\epsilon} I) \)
- Estimation of \( \text{cov}[g] \) by \( \frac{1}{T} \sum_t g(t)g'(t) \)
- SVD: \( \text{cov}[g] = U\Lambda U' \) and its identification with \( A\Lambda A' + \sigma^2_{\epsilon} I \)
- If \( \sigma^2_{\epsilon} = 0 \rightarrow \) Classical PCA
- The identification is not unique and needs constraints or prior information
- Maximum Likelihood is not unique
Link with PCA

Bayesian Probabilistic PCA:
\[ g(t) = Af(t) + \epsilon(t) \]

- \[ p(g(t)|A, f(t), \sigma_\epsilon^2) = \mathcal{N}(Af(t), \sigma_\epsilon^2 I) \]
- \[ p(f(t)|\Lambda) = \mathcal{N}(0, \Lambda) \]
- \[ p(g(t)|A, \sigma_\epsilon^2, \Lambda) = \mathcal{N}(0, A\Lambda A' + \sigma_\epsilon^2 I) \]
- \[ P(A|A_0, V_0) = \mathcal{N}(A_0, V_0) \]
- \[ p(A, f(t)|g(t)) \propto p(g(t)|A, f(t), \sigma_\epsilon^2) p(f(t)|\Lambda) p(A|A_0, V_0) \]
- Integration over \( f(t) : \)
\[
p(A|g(t), \sigma_\epsilon^2, \Lambda) \propto p(g(t)|A, \sigma_\epsilon^2, \Lambda) p(A|A_0, V_0) \]
\[
\propto |A\Lambda A' + \sigma_\epsilon^2 I|^{-1/2} \exp \left\{ g' [A\Lambda A' + \sigma_\epsilon^2 I]^{-1} g \right\} \]
\[
\exp \left\{ -\frac{1}{2} \sum_i \sum_j [A - A_0]_{ij}^2 / [V_0]_{ij} \right\} \]

- MAP solution can be computed iteratively

- Full Bayesian:
\[
p(A, \sigma_\epsilon^2, \Lambda|g(t)) \propto p(g(t)|A, \sigma_\epsilon^2, \Lambda) p(A|A_0, V_0) p(\Lambda|\Lambda_0, \alpha_0) p(\sigma_\epsilon^2|\alpha_0, \beta_0) \]
Link with ICA

Classical ICA:

\[ g(t) = Af(t) \longrightarrow p_g(g(t)|A) = \text{det}(A) p_f(A^{-1}g(t)) \]

- Noting by \( B = A^{-1} \) and by \( y = Bg \) and assuming sources to be independent \( p_f(f) = \prod_i p_j(f_j) \):

\[
\ln p(g_{1..T}|B) = \frac{T}{2} \ln \text{det}(B) + \sum_t \sum_j p_j(y_j) + c
\]

- Maximum Likelihood:

\[
\hat{B} = \arg \max_B \{ J(B) = \ln p(g_{1..T}|B) \}
\]

- Iterative Gradient based algorithm

\[
B^{(k+1)} = B^{(k)} - \gamma H(y) \quad \text{with} \quad H(y) = \phi(y) y^t - I
\]

\[
\phi(y) = [\phi_1(y_1), \ldots, \phi_n(y_n)]^t \quad \text{and} \quad \phi_j(z) = -\frac{p_j'(z)}{p_j(z)}.
\]
<table>
<thead>
<tr>
<th>Distribution</th>
<th>Density Function</th>
<th>Mean Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss</td>
<td>( p(z) \propto \exp {-\alpha z^2} )</td>
<td>( \phi(z) = 2\alpha z )</td>
</tr>
<tr>
<td>Laplace</td>
<td>( p(z) \propto \exp {-\alpha</td>
<td>z</td>
</tr>
<tr>
<td>Cauchy</td>
<td>( p(z) \propto \frac{1}{1 + (z/\alpha)^2} )</td>
<td>( \phi(z) = \frac{2z/\alpha^2}{1+(z/\alpha)^2} )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( p(z) \propto z^\alpha \exp {-\beta z} )</td>
<td>( \phi(z) = -\alpha/z + \beta )</td>
</tr>
<tr>
<td>Sub-Gaussian</td>
<td>( p(z) \propto \exp \left{-\frac{1}{2} z^2 \right} \text{sech}^2(z) )</td>
<td>( \phi(z) = z + \tanh(z) )</td>
</tr>
<tr>
<td>Mixture of Gauss.</td>
<td>( p(z) \propto \exp \left{-\frac{1}{2} (z - \alpha)^2 \right} ) + ( \exp \left{-\frac{1}{2} (z + \alpha)^2 \right} )</td>
<td>( \phi(z) = \alpha z - \alpha \tanh(\alpha z) )</td>
</tr>
</tbody>
</table>
Link with Learning in Neural Network

\[ H(y) = \frac{\partial}{\partial B} \left[ \sum_i \ln p_i(y_i) - \ln |\text{det}(B)| \right]. \]

Optimization by a Gradient based algorithm:

\[ \Delta B \propto H(y) = [I - \phi(y)y^t]B \]
5. Advanced Bayesian methods

Non Gaussian, Dependent, Colored sources

\[ g(t) = A f(t) + \epsilon(t) \quad \longrightarrow \quad g_{1..T} = Af_{1..T} + \epsilon_{1..T} \]

\[
\ln p(g_{1..T}|f_{1..T}) = \sum_t \sum_i q_i(g_i(t) - [Af]_i(t))
\]

\[
\ln p(f_{1..T}) = \sum \sum r_j(f_j(t)), \quad p(A) \propto \exp \left\{ -\frac{1}{2\sigma_a^2} \sum \sum a_{ij}^2 \right\}
\]

\[
\ln p(A, f_{1..T}|g_{1..T}) = \sum \sum q_i(g_i(t) - [Af]_i(t)) + \sum \sum r_j(f_j(t)) + \frac{1}{2\sigma_a^2} \sum \sum a_{ij}^2 + cte.
\]

\[
\hat{f}^{(k)} = \arg \max_f \left\{ \sum_i q_i(g_i - [Af]_i) + \sum_j r_j(f_j) \right\}
\]

\[
\hat{A}^{(k)} = \arg \max_A \left\{ \sum_i q_i(g_i - [Af]_i) + \frac{1}{2\sigma_a^2} \sum \sum a_{ij}^2 \right\}
\]
Gaussian / Non Gaussian

- **Gaussian laws** for the noise and sources:
  \[
  \begin{align*}
  f(t) &= (A' A + \lambda I)^{-1} A' g(t) \\
  A &= \sum_t g(t) f'(t)(\sum_t f(t) f'(t) + \mu I)^{-1}
  \end{align*}
  \]

  with \( \lambda = \sigma^2_{\epsilon}/\sigma^2_s \) and \( \mu = \sigma^2_{\epsilon}/\sigma^2_a \).

- **Non Gaussian sources** \( f(t) \):
  \[
  \begin{align*}
  y(t) &= (A' A + \lambda I)^{-1} A' g(t) \\
  f(t) &= \phi(y(t)) \\
  A &= \sum_t g(t) f'(t)(\sum_t f(t) f'(t) + \mu I)^{-1}
  \end{align*}
  \]

  NN Learning
  \[
  A = g f'(f f' + \mu I)^{-1}
  \]
Dependent sources

- **Gaussian laws** for the noise and sources:

\[
\begin{align*}
    f(t) &= (A' A + \lambda D' D)^{-1} A' g(t) \\
    A &= \sum_t g(t) f'(t) (\sum_t f(t) f'(t) + \mu I)^{-1}
\end{align*}
\]

with \( \lambda = \sigma^2_{\epsilon}/\sigma^2_s \) and \( \mu = \sigma^2_{\epsilon}/\sigma^2_a \).

- **Non Gaussian sources** \( f \):

\[
\begin{align*}
    y(t) &= (A' A + \lambda D' D)^{-1} A' g \\
    f(t) &= \phi(y(t)) \\
    A &= \sum_t g(t) f'(t) (\sum_t f(t) f'(t) + \mu I)^{-1}
\end{align*}
\]

\[
\[
A = g f'(f f' + \mu I)^{-1}
\]
Independent but temporally colored sources

IID case:

\[ p(f(t)) = \sum_j p_j(f_j(t)), \forall t \rightarrow p(f_{1..T}) = \sum_j \sum_t p_j(f_j(t)) \]

Spatially independent but temporally colored sources

\[ p(f_{1..T}) = \sum_j p_j(f_j(1), \ldots, f_j(T)) \]

Main difficulty: Modelization of \( p(f_j(1), \ldots, f_j(T)) \)

- Markovian Models:

\[ p(f_j(1), \ldots, f_j(T)) = p(f_j(1)) \prod_{t=2}^{T} p(f_j(t)|f_j(t-1)) \]

- Gauss-Markov (First order AR Model):

\[ p(f_j(1), \ldots, f_j(T)) \propto \exp \left\{ -\frac{1}{2} \sum_t ((f_j(t) - \alpha f_j(t-1))^2 \right\} \]
Joint Estimation of $A$ and $f$ with a Gaussian prior model

\[
p(f_j(t)|v_{0j}) = \mathcal{N}(0, v_{0j})
\]
\[
p(f(t)|v_0) \propto \exp \left\{ -\frac{1}{2} \sum_j f_j^2(t)/v_{0j} \right\}
\]
\[
p(A_{ij}|A_{0ij}, V_{0ij}) = \mathcal{N}(A_{0ij}, V_{0ij})
\]
\[
p(A|A_0, V_0) = \mathcal{N}(A_0, V_0)
\]
\[
p(g(t)|A, f(t), v_\epsilon) = \mathcal{N}(Af(t), v_\epsilon I)
\]

\[
p(f_{1..T}, A|g_{1..T}) \propto p(g_{1..T}|A, f_{1..T}, v_\epsilon) p(f_{1..T}) p(A|A_0, V_0)
\]
\[
\propto \prod_t p(g(t)|A, f(t), v_\epsilon) p(f(t)|z(t)) p(A|A_0, V_0)
\]

\[
p(f(t)|g_{1..T}, A, v_\epsilon, v_0) = \mathcal{N}(\hat{f}(t), \hat{\Sigma})
\]
\[
p(A|g_{1..T}, f_{1..T}, v_\epsilon, A_0, V_0) = \mathcal{N}(\hat{A}, \hat{V})
\]

\[
p(f(t)|g_{1..T}, v_\epsilon, v_0) = \mathcal{N}(\hat{f}(t), \hat{\Sigma})
\]
\[
p(A|g_{1..T}, v_\epsilon, A_0, V_0) = \mathcal{N}(\hat{A}, \hat{V})
\]
Joint Estimation of $A$ and $f$ with a Gaussian prior model.

$$\mathbf{v}_0 = [v_f, \ldots, v_f]' , \quad \text{All sources a priori same variance } v_f$$

$$\mathbf{v}_\epsilon = [v_\epsilon, \ldots, v_\epsilon]' , \quad \text{All noise terms a priori same variance } v_\epsilon$$

$$A_0 = 0, \quad V_0 = v_a I$$

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

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

$$p(A|g(t), f(t), v_\epsilon, A_0, V_0) = \mathcal{N}(\hat{A}, \hat{V})$$

$$\begin{cases} 
\hat{V} = (F' F + \lambda_f I)^{-1} \\
\hat{A} = \sum_t g(t) f'(t) (\sum_t f(t) f'(t) + \lambda_a I)^{-1} \quad \lambda_a = v_\epsilon/v_a 
\end{cases}$$
Joint Estimation of $A$ and $f$ with a Gaussian prior model.

$$p(f_{1..T}, A | g_{1..T}) \propto p(g_{1..T} | A, f_{1..T}, v_\epsilon) \ p(f_{1..T}) \ p(A | A_0, V_0)$$
$$\propto \prod_t p(g(t) | A, f(t), v_\epsilon) \ p(f(t) | z(t)) \ p(A | A_0, V_0)$$

Joint MAP:

$$\begin{cases} 
\hat{f}(t) = (\hat{A}' \hat{A} + \lambda_f I)^{-1} \hat{A}' g(t), \\
\hat{A} = \sum_t g(t) \hat{f}'(t) \left( \sum_t \hat{f}(t) \hat{f}'(t) + \lambda_a I \right)^{-1} 
\end{cases}$$

$$\lambda_f = v_\epsilon / v_f$$
$$\lambda_a = v_\epsilon / v_a$$

Algorithm:

$$A^{(0)} \rightarrow \hat{A} \rightarrow $$

$$\quad \left( \hat{A}' \hat{A} + \lambda_f I \right)^{-1} \hat{A}' g$$

$$\rightarrow \hat{f}(t)$$

$$\uparrow$$

$$\hat{A} \leftarrow \sum_t g(t) \hat{f}'(t) \left( \sum_t \hat{f}(t) \hat{f}'(t) + \lambda_a I \right)^{-1} \leftarrow \hat{f}(t)$$
Joint Estimation of $A$ and $f$ with a Gaussian prior model.

VBA: $p(f_{1..T}, A|g_{1..T}) \longrightarrow q_1(f_{1..T}|A, g_{1..T}) q_2(A|f_{1..T}, g_{1..T})$

$$q_1(f(t)|g(t), A, v_\epsilon, v_0) = \mathcal{N}(\hat{f}(t), \hat{\Sigma})$$

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

$$q_2(A|g(t), f(t), v_\epsilon, A_0, V_0) = \mathcal{N}(\hat{A}, \hat{V})$$

$$\begin{cases} 
\hat{V} = (F' F + \lambda_f \hat{\Sigma})^{-1} \\
\hat{A} = \sum_t g(t) f'(t) \left(\sum_t f(t) f'(t) + \lambda_a \hat{\Sigma}\right)^{-1} \quad \lambda_a = v_\epsilon/v_a 
\end{cases}$$
Joint Estimation of $A$ and $f$ with a Gaussian prior model.

Algorithm:

\[ A^{(0)} \rightarrow \hat{A} \rightarrow \hat{f}(t) = \left( \hat{A}'\hat{A} + \lambda_f \hat{V} \right)^{-1} \hat{A}'g(t) \]
\[ \hat{\Sigma} = (A'\hat{A} + \lambda_f \hat{V})^{-1} \rightarrow \hat{f}(t) \]

\[ \hat{A} = \sum_t g(t)\hat{f}'(t) \left( \sum_t \hat{f}(t)\hat{f}'(t) + \lambda_a \hat{\Sigma} \right)^{-1} \rightarrow \hat{f}(t) \]
\[ \hat{V} = (F'F + \lambda_f \hat{\Sigma})^{-1} \rightarrow \hat{\Sigma} \]
Joint Estimation of $A$ and $f$, $v_\epsilon$ and $v$

\[
p(f_j(t)|v_0j) = \mathcal{N}(0, v_j)
\]
\[
p(f(t)|v_0) \propto \exp \left\{ -\frac{1}{2} \sum_j f_j^2(t) / v_0j \right\}
\]
\[
p(A_{m,n}|A_{0m,n}, V_{0m,n}) = \mathcal{N}(A_{0m,n}, V_{0m,n})
\]
\[
p(A|A_0, V_0) = \mathcal{N}(A_0, V_0)
\]
\[
p(g(t)|A, f(t), v_\epsilon) = \mathcal{N}(Af(t), v_\epsilon I)
\]
\[
p(v_\epsilon|\alpha_\epsilon, \beta_\epsilon) = \mathcal{G}(\alpha_\epsilon, \beta_\epsilon)
\]
\[
p(v_j|\alpha_0, \beta_0) = \mathcal{G}(\alpha_0, \beta_0)
\]

\[
p(f_{1..T}, A, v_\epsilon, v|g_{1..T}) \propto p(g_{1..T}|A, f_{1..T}, v_\epsilon) \ p(f_{1..T}|v) \ p(A|A_0, V_0) \ p(v_\epsilon|\alpha_\epsilon, \beta_\epsilon) \ \prod_j p(v_j|\alpha_0, \beta_0)
\]
\[
\propto \ \prod_t p(g(t)|A, f(t), v_\epsilon) \ p(f(t)|v(t)) \ p(A|A_0, V_0) \ p(v_\epsilon|\alpha_\epsilon, \beta_\epsilon) \ \prod_j p(v_j|\alpha_0, \beta_0)
\]
Joint Estimation of $A$ and $f$, $\nu_\epsilon$ and $\nu$ with dependent and colored sources.

$p(f_1..T, A, \nu_\epsilon, \nu | g_{1..T}) \propto p(g_{1..T}|A, f_{1..T}, \nu_\epsilon) \ p(f_{1..T}|\nu) \ p(A|A_0, V_0) p(\nu_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \ \prod_j p(\nu_j|\alpha_0, \beta_0)$

$\propto \ \prod_t p(g(t)|A, f(t), \nu_\epsilon) \ p(f(t)|z(t)) \ p(A|A_0, V_0) p(\nu_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \ \prod_j p(\nu_j|\alpha_0, \beta_0)$
Joint Estimation of $A$ and $f$ with a variance modulated prior model inducing sparsity

$p(v_j(t)|\alpha_{j0}, \beta_{j0}) = IG(\alpha_{j0}, \beta_{j0})$
$p(f_j(t)|v_j(t)) = N(0, v_j(t))$
$p(f(t)|v(t)) \propto \exp \left\{ - \sum_j f^2_j(t)/v_j(t) \right\}$

$p(A|A_0, V_0) = N(A_0, V_0)$
$p(g(t)|A, f(t), v_\epsilon) = N(Af(t), v_\epsilon I)$
$p(v_\epsilon|\alpha_{\epsilon0}, \beta_{\epsilon0}) = G(\alpha_{\epsilon0}, \beta_{\epsilon0})$

$p(f_{1..T}, A, v_{1..T}, v_\epsilon|g_{1..T}) \propto p(g_{1..T}|A, f_{1..T}, v_\epsilon) p(f_{1..T}|v_{1..T}) \prod_j p(v_{1..T}|\alpha_{j0}, \beta_{j0}) p(A|A_0, V_0)$

$\propto \prod_t p(g(t)|A, f(t), v_\epsilon) p(f(t)|v(t)) \prod_t \prod_j p(v_j(t)|\alpha_{j0}, \beta_{j0}) p(A|A_0, V_0)$
Joint Estimation of $A$ and $f$ with a variance modulated prior model..

$p(v_j(t)|\alpha_{j0}, \beta_{j0}) = IG(\alpha_{j0}, \beta_{j0})$

$p(f_j(t)|v_j(t)) = N(0, v_j(t))$

$p(f(t)|v(t)) \propto \exp\left\{-\sum_j f_j^2(t)/v_j(t)\right\}$

$p(A|A_0, V_0) = N(A_0, V_0)$

$p(g(t)|A, f(t), v_\epsilon) = N(Af(t), v_\epsilon I)$

$p(v_\epsilon|\alpha_{\epsilon0}, \beta_{\epsilon0}) = G(\alpha_{\epsilon0}, \beta_{\epsilon0})$

$p(f(t), A, v(t), v_\epsilon|g(t)) \propto p(g(t)|A, f(t), v_\epsilon) p(f(t)|v(t)) \prod_j p(v_j(t)|\alpha_{j0}, \beta_{j0}) p(A|A_0, V_0)$

$p(f(t)|g(t), A, v(t), v_\epsilon, \alpha_{j0}, \beta_{j0}) = N(\hat{f}(t), \hat{\Sigma})$

$p(A|g(t), f(t), v_\epsilon, A_0, V_0) = N(\hat{A}, \hat{V})$

$p(v_j(t)|f_j(t), \alpha_{j0}, \beta_{j0}) = G(\hat{\alpha}_j, \hat{\beta}_j)$

$p(v_\epsilon|g(t), f(t), \alpha_{\epsilon0}, \beta_{\epsilon0}) = G(\hat{\alpha}_{\epsilon0}, \hat{\beta}_{\epsilon0})$
Joint Estimation of $A$ and $f(t)$ with a Mixture of Gaussians prior model

$$f(t) \rightarrow A \rightarrow g(t) \quad t = 1, \ldots, T$$

For $j$, let

$$z_{j}(t) = k|\alpha_{j_k} = \alpha_{j_k}, \sum_k \alpha_{j_k} = 1$$

$$p(f_{j}(t)|z_{j}(t) = k) = \mathcal{N}(m_{jk}, v_{jk})$$

$$p(f(t)|z(t), m, v) = \sum_j \mathcal{N}(m_{z_j(t)}, v_{z_j(t)})$$

$$p(A|A_0, V_0) = \mathcal{N}(A_0, V_0)$$

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

$$p(v_\epsilon|\alpha_\epsilon, \beta_\epsilon) = \mathcal{IG}(\alpha_\epsilon, \beta_\epsilon)$$

$$p(f_{1..T}, A, z_{1..T}, v_\epsilon|g_{1..T}) \propto p(g_{1..T}|A, f_{1..T}, v_\epsilon) p(f_{1..T}|z_{1..T}) \prod_j p(z_{j1..T}|\alpha_{j0}, \beta_{j0}) p(A|A_0, V_0) p(v_\epsilon|\alpha_\epsilon, \beta_\epsilon)$$
Joint Estimation of $A$, $f(t)$ and $\theta$ with mixture model

\[ p(z_j(t) = k|\alpha_{jk}) = \alpha_{jk}, \quad \sum_k \alpha_{jk} = 1 \]
\[ p(f_j(t)|z_j(t) = k) = \mathcal{N}(m_{jk}, v_{jk}) \]
\[ p(f(t)|z(t), m, v) = \sum_j \mathcal{N}(m_{z_j(t)}, v_{z_j(t)}) \]
\[ p(A|A_0, V_0) = \mathcal{N}(A_0, V_0) \]
\[ p(g(t)|f(t), \nu_\epsilon) = \mathcal{N}(Af(t), \nu_\epsilon I) \]
\[ p(\nu_\epsilon|\alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) = \mathcal{IG}(\alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) \]
\[ p(m|m_0, v_0) = \mathcal{G}(\alpha_{m_0}, \beta_{m_0}) \]
\[ p(v|\alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) = \mathcal{IG}(\alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) \]

\[ p(f_{1..T}, A, z_{1..T}, \nu_\epsilon, m, v|g_{1..T}) \propto \prod_t p(g(t)|A, f(t), \nu_\epsilon) p(f(t)|z(t)) \]
\[ \prod_t \prod_j p(z_j(t)|\alpha_{j0}, \beta_{j0}) \]
\[ p(A|A_0, V_0) p(\nu_\epsilon|\alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) \]
\[ p(m|m_0, v_0) p(v|\alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) \]

\[ p(f(t)|g(t), A, z(t), \nu_\epsilon, \alpha_{j0}, \beta_{j0}) = \mathcal{N}(\hat{f}(t), \hat{\Sigma}) \]
\[ p(A|g(t), f(t), \nu_\epsilon, A_0, V_0) = \mathcal{N}(\hat{A}, \hat{V}) \]
\[ p(z_j(t) = k|f_j(t), \alpha_{j0}, \beta_{j0}) = \hat{\alpha}_{j_k} \]
\[ p(\nu_\epsilon|g(t), f(t), \alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) = \mathcal{G}(\hat{\alpha}_{\nu_\epsilon}, \hat{\beta}_{\nu_\epsilon}) \]
\[ p(m|f(t), m_0, v_0) = \mathcal{N}(\hat{m}, \hat{v}) \]
\[ p(v_j|f_j(t), \alpha_{\nu_\epsilon}, \beta_{\nu_\epsilon}) = \mathcal{IG}(\hat{\alpha}_j, \hat{\beta}_j) \]
Joint Estimation of $A$ and $f(t)$ with mixture model (common $z$)


cardinality

\[ p(z(t) = k|\alpha_k) = \alpha_k, \quad \sum_k \alpha_k = 1 \]

\[ p(f_j(t)|z(t) = k) = \mathcal{N}(m_{jk}, v_{jk}) \]

\[ p(f(t)|z(t), m, v) = \sum_j \mathcal{N}(m_{z_j}(t), v_{z_j}(t)) \]

\[ p(A|A_0, V_0) = \mathcal{N}(A_0, V_0) \]

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

\[ p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{G}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(m|m_0, v_0) = \mathcal{G}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(v|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{G}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(m|m_0, v_0) = \mathcal{G}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{G}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(m|f(t), m_0, v_0) = \mathcal{N}(\hat{m}, \hat{v}) \quad p(v_j|f_j(t), \alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\hat{\alpha}_j, \hat{\beta}_j) \]

\[ p(f_1..T, A, z_1..T, v_\epsilon|g_1..T) \propto \prod_t p(g(t)|A, f(t), v_\epsilon) \quad p(f(t)|z(t)) \]

\[ \prod_t p(z(t)|\alpha_k, \beta_k) \]

\[ p(A|A_0, V_0) \quad p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(m|m_0, v_0) \quad p(v|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \]

\[ p(m|f(t), m_0, v_0) = \mathcal{N}(\hat{m}, \hat{v}) \quad p(v_j|f_j(t), \alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\hat{\alpha}_j, \hat{\beta}_j) \]
6. A few applications

- X ray Computed Tomography for Non Destructive Testing
- Spectrometry (with Said Mousaoui)
- Cosmic Microwave Background in Radio Astronomy (with Hichem Snoussi)
- Satellite image separation (with Mahiedding Ichir)
- Hyperspectral imaging (with Nadia Bali and Adel Mohammadpour)
Which images I am looking for?
Which image I am looking for?

- Gauss-Markov
- Generalized GM
- Piecewise Gaussian
- Mixture of GM
Different prior models for signals and images

- **Separable**
  \[
  p(f) = \prod_j p_j(f_j) \propto \exp\left\{ -\beta \sum_j \phi(f_j) \right\}
  \]
  \[
  p(f) \propto \exp\left\{ -\beta \sum_{r \in R} \phi(f(r)) \right\}
  \]

- **Markoviens (simple)**
  \[
  p(f_j | f_{j-1}) \propto \exp\left\{ -\beta \phi(f_j - f_{j-1}) \right\}
  \]
  \[
  p(f) \propto \exp\left\{ -\beta \sum_{r \in R} \sum_{r' \in V(r)} \phi(f(r), f(r')) \right\}
  \]

- **Markovien with hidden variables**
  \[
  z(r) \text{ (lines, contours, regions)}
  \]
  \[
  p(f | z) \propto \exp\left\{ -\beta \sum_{r \in R} \sum_{r' \in V(r)} \phi(f(r), f(r'), z(r), z(r')) \right\}
  \]
Different prior models for images: Separable

- **Gaussian**
  \[ p(f_j) \propto \exp\left\{-\alpha |f_j|^2\right\} \]

- **Generalized Gaussian**
  \[ p(f_j) \propto \exp\left\{-\alpha |f_j|^p\right\}, \quad 1 \leq p \leq 2 \]

- **Gamma**
  \[ p(f_j) \propto f_j^\alpha \exp\left\{-\beta f_j\right\} \]

- **Beta**
  \[ p(f_j) \propto f_j^\alpha (1 - f_j)^\beta \]
Different prior models: Simple Markovian

\[ p(f_j | f) \propto \exp \left\{ -\alpha \sum_{i \in v_j} \phi(f_j, f_i) \right\} \rightarrow \Phi(f) = \alpha \sum_j \sum_{i \in V_j} \phi(f_j, f_i) \]

- 1D case and one neighbor \( V_j = j - 1 \):
  \[ \Phi(f) = \alpha \sum_j \phi(f_j - f_{j-1}) \]

- 1D Case and two neighbors \( V_j = \{j - 1, j + 1\} \):
  \[ \Phi(f) = \alpha \sum_j \phi(f_j - \beta(f_{j-1} + f_{j-1})) \]

- 2D case with 4 neighbors:
  \[ \Phi(f) = \alpha \sum_{r \in R} \phi \left( f(r) - \beta \sum_{r' \in \mathcal{V}(r)} f(r') \right) \]

- \( \phi(t) = |t|^\gamma\): Generalized Gaussian
Different prior models: Markovian with hidden variables

Piecewise Gaussians
(contours hidden variables)

\[ p(f_j | q_j, f_{j-1}) = \mathcal{N} \left( (1 - q_j) f_{j-1}, \sigma_f^2 \right) \]

Mixture of Gaussians (MoG)
(regions labels hidden variables)

\[ p(f_j | z_j = k) = \mathcal{N}(m_k, \sigma_k^2) \]

\[ p(f | q) \propto \exp \left\{ -\alpha \sum_j \left| f_j - (1 - q_j) f_{j-1} \right|^2 \right\} \]
\[ p(f | z) \propto \exp \left\{ -\alpha \sum_k \sum_{j \in R_k} \left( \frac{f_j - m_k}{\sigma_k} \right)^2 \right\} \]
Gauss-Markov-Potts prior models for images

\[ \begin{align*}
  f(r) & \quad z(r) & \quad c(r) = 1 - \delta(z(r) - z(r')) \\
  p(f(r) | z(r) = k, m_k, v_k) &= \mathcal{N}(m_k, v_k) \\
  p(f(r)) &= \sum_k P(z(r) = k) \mathcal{N}(m_k, v_k) \quad \text{Mixture of Gaussians} \\
  p(z(r) | z(r'), r' \in \mathcal{V}(r)) & \propto \exp \left\{ \gamma \sum_{r' \in \mathcal{V}(r)} \delta(z(r) - z(r')) \right\} \\
  p(z) & \propto \exp \left\{ \gamma \sum_{r \in \mathcal{R}} \sum_{r' \in \mathcal{V}(r)} \delta(z(r) - z(r')) \right\}
\end{align*} \]

- Separable iid hidden variables: \( p(z) = \prod_r p(z(r)) \)
- Markovian hidden variables: \( p(z) \) Potts-Markov:
**Four different cases**

To each pixel of the image is associated 2 variables $f(r)$ and $z(r)$

- $f|z$ Gaussian iid, $z$ iid : Mixture of Gaussians

- $f|z$ Gauss-Markov, $z$ iid : Mixture of Gauss-Markov

- $f|z$ Gaussian iid, $z$ Potts-Markov : Mixture of Independent Gaussians (MIG with Hidden Potts)

- $f|z$ Markov, $z$ Potts-Markov : Mixture of Gauss-Markov (MGM with hidden Potts)
Summary of the two proposed models

\[ f \mid z \sim \text{Gaussian iid} \]
\[ z \sim \text{Potts-Markov} \]
(MIG with Hidden Potts)

\[ f \mid z \sim \text{Markov} \]
\[ z \sim \text{Potts-Markov} \]
(MGM with hidden Potts)
Bayesian Computation

\[ p(f, z, \theta | g) \propto p(g | f, z, v_\epsilon) p(f | z, m, v) p(z | \gamma, \alpha) p(\theta) \]

\[ \theta = \{v_\epsilon, (\alpha_k, m_k, v_k), k = 1, \cdots, K\} \quad p(\theta) \quad \text{Conjugate priors} \]

- Direct computation and use of \( p(f, z, \theta | g; \mathcal{M}) \) is too complex
- Possible approximations:
  - Gauss-Laplace (Gaussian approximation)
  - Exploration (Sampling) using MCMC methods
  - Separable approximation (Variational techniques)

- Main idea in Variational Bayesian methods:
  Approximate
  \[ p(f, z, \theta | g; \mathcal{M}) \quad \text{by} \quad q(f, z, \theta) = q_1(f) q_2(z) q_3(\theta) \]

  - Choice of approximation criterion : \( KL(q : p) \)
  - Choice of appropriate families of probability laws
  for \( q_1(f), q_2(z) \) and \( q_3(\theta) \)
MCMC based algorithm

\[ p(f, z, \theta | g) \propto p(g | f, z, \theta) p(f | z, \theta) p(z) p(\theta) \]

General scheme:

\[ \hat{f} \sim p(f | \hat{z}, \hat{\theta}, g) \longrightarrow \hat{z} \sim p(z | \hat{f}, \hat{\theta}, g) \longrightarrow \hat{\theta} \sim (\theta | \hat{f}, \hat{z}, g) \]

- Estimate \( f \) using \( p(f | \hat{z}, \hat{\theta}, g) \propto p(g | f, \theta) p(f | \hat{z}, \hat{\theta}) \)
  Needs optimisation of a quadratic criterion.

- Estimate \( z \) using \( p(z | \hat{f}, \hat{\theta}, g) \propto p(g | \hat{f}, \hat{z}, \hat{\theta}) p(z) \)
  Needs sampling of a Potts Markov field.

- Estimate \( \theta \) using
  \[
  p(\theta | \hat{f}, \hat{z}, g) \propto p(g | \hat{f}, \sigma^2 \epsilon I) p(\hat{f} | \hat{z}, (m_k, v_k)) p(\theta)
  \]
  Conjugate priors \( \longrightarrow \) analytical expressions.
Application of CT in NDT

Reconstruction from only 2 projections

\[ g_1(x) = \int f(x, y) \, dy, \quad g_2(y) = \int f(x, y) \, dx \]

- Given the marginals \( g_1(x) \) and \( g_2(y) \) find the joint distribution \( f(x, y) \).
- Infinite number of solutions: \( f(x, y) = g_1(x) \, g_2(y) \, \Omega(x, y) \)
  \( \Omega(x, y) \) is a Copula:
  \[ \int \Omega(x, y) \, dx = 1 \quad \text{and} \quad \int \Omega(x, y) \, dy = 1 \]
Application in CT

\[ g|f = Hf + \epsilon \]
\[ g|f \sim \mathcal{N}(Hf, \sigma^2 I) \]

Gaussian

\[ f|z \]

iid Gaussian

or

Gauss-Markov

\[ c \]

\[ c(r) \in \{0, 1\} \]

or

Potts

or

\[ 1 - \delta(z(r) - z(r')) \]

binary

A. Mohammad-Djafari, 6 ème École d’été de Peyresq, Séparation de sources, Juillet 25-30, 2011.
Results

Original

Backprojection

Filtered BP

LS

Gauss-Markov+pos

GM+Line process

GM+Label process

A. Mohammad-Djafari, 6 ème École d’été de Peyresq, Séparation de sources, Juillet 25-30, 2011.
Application in Microwave imaging

\[ g(\omega) = \int f(r) \exp \{-j(\omega \cdot r)\} \, dr + \epsilon(\omega) \]

\[ g(u, v) = \iint f(x, y) \exp \{-j(ux + vy)\} \, dx \, dy + \epsilon(u, v) \]

\[ g = Hf + \epsilon \]

\( f(x, y) \) \hspace{1cm} \( g(u, v) \) \hspace{1cm} \( \hat{f} \) \hspace{1cm} \text{IFT} \hspace{1cm} \hat{f} \) Proposed method
Application in Microwave imaging
Conclusions

- Bayesian Inference for inverse problems
- Approximations (Laplace, MCMC, Variational)
- Gauss-Markov-Potts are useful prior models for images incorporating regions and contours
- Separable approximations for Joint posterior with Gauss-Markov-Potts priors
- Application in different CT (X ray, US, Microwaves, PET, SPECT)

Perspectives:

- Efficient implementation in 2D and 3D cases
- Evaluation of performances and comparison with MCMC methods
- Application to other linear and non linear inverse problems: (PET, SPECT or ultrasound and microwave imaging)
Images fusion and joint segmentation

(with O. Féron)

\[
\begin{align*}
    g_i(r) &= f_i(r) + \epsilon_i(r) \\
    p(f_i(r) | z(r) = k) &= \mathcal{N}(m_{ik}, \sigma^2_{ik}) \\
    p(f | z) &= \prod_i p(f_i | z)
\end{align*}
\]
Data fusion in medical imaging

(with O. Féron)

\[
\begin{align*}
g_i(r) &= f_i(r) + \epsilon_i(r) \\
p(f_i(r)|z(r) = k) &= \mathcal{N}(m_{ik}, \sigma_{ik}^2) \\
p(f|z) &= \prod_i p(f_i|z)
\end{align*}
\]
Joint segmentation of hyper-spectral images

(with N. Bali & A. Mohammadpour)

\[
\begin{align*}
  g_i(r) &= f_i(r) + \epsilon_i(r) \\
  p(f_i(r) | z(r) = k) &= \mathcal{N}(m_{ik}, \sigma_{ik}^2), \quad k = 1, \cdots, K \\
  p(f | z) &= \prod_i p(f_i | z) \\
  m_{ik} & \text{ follow a Markovian model along the index } i
\end{align*}
\]
Segmentation of a video sequence of images

(with P. Brault)

\[
\begin{align*}
    g_i(r) = f_i(r) + \epsilon_i(r) \\
    p(f_i(r)|z_i(r) = k) = \mathcal{N}(m_{ik}, \sigma^2_i k), \quad k = 1, \ldots, K \\
    p(f|z) = \prod_i p(f_i|z_i) \\
    z_i(r) \text{ follow a Markovian model along the index } i
\end{align*}
\]
Source separation
(with H. Snoussi & M. Ichir)

\[
g_i(r) = \sum_{j=1}^{N} A_{ij} f_j(r) + \epsilon_i(r)
\]

\[
p(f_j(r) | z_j(r) = k) = \mathcal{N}(m_{jk}, \sigma_{jk}^2)
\]

\[
p(A_{ij}) = \mathcal{N}(A_{0ij}, \sigma_{0ij}^2)
\]
Some references

Thanks, Questions and Discussions

**Thanks to:**
**My graduated PhD students:**
- H. Snoussi, M. Ichir, (Sources separation)
- F. Humblot (Super-resolution)
- H. Carfantan, O. Féron (Microwave Tomography)
- S. Fékih-Salem (3D X ray Tomography)

**My present PhD students:**
- H. Ayasso (Optical Tomography, Variational Bayes)
- D. Pougaza (Tomography and Copula)
- Sh. Zhu (SAR Imaging)
- D. Fall (Emission Positron Tomography, Non Parametric Bayesian)

**My colleagues in GPI (L2S) & collaborators in other institutes:**
- B. Duchêne & A. Joisel (Inverse scattering and Microwave Imaging)
- N. Gac & A. Rabanal (GPU Implementation)
- Th. Rodet (Tomography)
- A. Vabre & S. Legoupil (CEA-LIST), (3D X ray Tomography)
- E. Barat (CEA-LIST) (Positron Emission Tomography, Non Parametric Bayesian)
- C. Comtat (SHFJ, CEA) (PET, Spatio-Temporal Brain activity)

**Questions and Discussions**
Conclusions and Perspectives

- Bayesian approach is powerful.
- We proposed a list of different probabilistic prior models which can be used for sparsity enforcing.
- We classified these models in two categories: simple heavy tails and hierarchical mixture models.
- We showed how to use these models for inverse problems where the desired solutions are sparse.
- Different algorithms have been developed and their relative performances are compared.
- We use these models for inverse problems in different signal and image processing applications such as:
  - Synthetic Aperture Radar (SAR) Imaging
  - Signal deconvolution in Proteomic and molecular imaging
  - X-ray Computed Tomography, Diffraction Optical Tomography, Microwave Imaging, ...
Main references on sparsity


