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 inference

Bayesian Inference:

- Make inference about hypotheses using all available information
- Needs to identify all hypotheses explicitely
- 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, Elegent

Bayesian inference for Scientists

Theory Hypotheses Model	Creativity ⟵ Science	Inference, Verification Falsification
Deduction Forward modeling ↓		Induction Inversion ↑
Predictions	Observation ⟶ (Errrors)	Data

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

Probability

- What is the probabilty ?
  Degree of beleif always conditionned on what we know
- What is the probabilty 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 milionth digit of π the digit 3 ?
- Probability Axioms
  - $0 \leq P(A|I) \leq 1$
  - $P(A|A, I) = 1$
  - $P(A|I) + P(A^c|I) = 1$
  - $P(A, B|I) = P(A|B, I)\, P(B|I)$
  - $P(A|B, I) = P(A, B|I)/P(B|I) \quad \text{if} \quad P(B|I) \neq 0$

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)

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

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

$$\text{Odds} = \frac{\text{Probablity}}{1-\text{Probablity}} \longrightarrow \text{Probablity} = \frac{\text{Odds}}{1-\text{Odds}}$$

Bayes' rule for continuous case

Continuous case:

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

$$p(x) = \int p(x,\theta)\,\mathrm{d}\theta = \int p(x|\theta)\,p(\theta)\,\mathrm{d}\theta$$

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: $\widehat{\theta}_{Mod} = \arg\max_\theta \{p(\theta|x)\}$
  - Mean: $\widehat{\theta}_{Mean} = \int \theta\, p(\theta|x)\,\mathrm{d}\theta$
  - Median: $\theta_{Med} : \; P(\theta > \theta_{Med}) = P(\theta < \theta_{Med})$

Credible intervals

- We can also compute Credible intervals:

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

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

- This is different from "Orthodox Confidence Intervals":
  - Define an estimator $\widehat{\theta}$ which is a function of data x (which are random!)
  - Compute the sampling distribution $g(\boldsymbol{\theta}|\theta)$
  - Compute $\alpha = 1 - \int_a^b g(u|\theta) \, \mathrm{d}u$
  - 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

$$x_i \sim \mathcal{N}(\mu, v), i = 1, .., n \longrightarrow p(x_i|\mu, v) = \frac{1}{\sqrt{2\pi v}} \exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{v}\right\}$$

$$x_i = \mu + \epsilon_i \quad \text{with} \quad \epsilon_i \sim \mathcal{N}(0, v), i = 1, .., n$$

Likelihood:

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

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

with $\bar{x} = \frac{1}{n}\sum_i x_i$ and $S = \sum_i (x_i - \bar{x})^2$

- Maximum Likelihood: $\widehat{\mu} = \bar{x}$, $\widehat{v} = \frac{1}{n}S$
- Bayesian:
  - Known v, flat prior for $\mu \longrightarrow \widehat{\mu} = \bar{x}$
  - Known v, Gaussian prior for μ
  - Known μ, unknown v with Jeffreys' prior
  - Inverse Gamma prior for v and Gaussian prior for μ (Conjugate priors)

Example of the Gaussian law

Bayesian:

- Known v, flat prior for μ
  $\longrightarrow p(\mu|\boldsymbol{x}, v) = \mathcal{N}(\bar{x}, v/n) \longrightarrow \widehat{\mu} = \bar{x}$
- Known v, Gaussian prior for μ:

  $$p(\mu|\mu_0, v_0) = \mathcal{N}(\mu_0, v_0) \longrightarrow p(\mu|\boldsymbol{x}, v) = \mathcal{N}(\widehat{\mu}, \widehat{v})$$
  with $\frac{1}{\widehat{v}} = \frac{n}{v} + \frac{1}{v_0}$, $\quad \widehat{\mu} = (1-w)\bar{x} + w\mu_0$, with $w = \widehat{v}/v_0$

  When $v_0 \longrightarrow \infty$ (flat prior) $\longrightarrow \widehat{\mu} = \bar{x}$
- Known μ, unknown v with Jeffreys' prior $p(v) = 1/v$:

$$\begin{aligned} p(v|\boldsymbol{x}, \mu) &\propto v^{-n/2} \exp\left\{-\tfrac{1}{2v}(S + n(\mu - \bar{x})\right\} \; v^{-1} \\ &\propto v^{-(n+1)/2} \exp\left\{-\tfrac{1}{2v}(S + n(\mu - \bar{x})\right\} \\ &= \mathcal{IG}(\widehat{\alpha}, \widehat{\beta}) \end{aligned}$$

with $\left\{ \begin{array}{l} \widehat{\alpha} = \frac{n+1}{2} \\ \widehat{\beta} = (S + n(\mu - \bar{x}))/2 \end{array} \right. \longrightarrow \widehat{v} = \dfrac{\widehat{\beta}}{\widehat{\alpha}} = \dfrac{1}{n-1} \sum_i (x_i - \mu)^2$

Example of the Gaussian law

Bayesian:

- Unknown μ with flat prior $p(\mu) = cte$,
  Unknown v with Jeffreys' prior $p(v) = 1/v$:

$$
\begin{aligned}
p(\mu, v|\boldsymbol{x}) & \propto v^{-n/2} \exp\left\{-\frac{1}{2v}(S + n(\mu - \bar{x})\right\}\; v^{-1} \\
& \propto v^{-(n+1)/2} \exp\left\{-\frac{1}{2v}(S + n(\mu - \bar{x})\right\} \\
& = \mathcal{N}(\bar{x}, v/n)\; \mathcal{IG}(\widehat{\alpha}, \widehat{\beta})
\end{aligned}
$$

$$
\text{with} \left\{ \begin{array}{l} \widehat{\alpha} = \frac{n+1}{2} \\ \widehat{\beta} = (S + n(\mu - \bar{x})^)/2 \end{array} \right. \longrightarrow \widehat{v} = \frac{\widehat{\beta}}{\widehat{\alpha}} = \frac{1}{n-1} \sum_i (x_i - \mu)^2
$$

We may integrate out μ:

$$
\begin{aligned}
p(v|\boldsymbol{x}) & = \int p(\mu, v|\boldsymbol{x})\; \mathrm{d}\mu \\
& \propto \int v^{-(n+1)/2} \exp\left\{-\frac{1}{2v}(S + n(\mu - \bar{x})\right\} \\
& \propto \left(1 + \frac{n(\mu-\bar{x})^2}{S}\right)^{-n/2} \\
& \mathcal{S}_t\left(\nu = n-1, t = \frac{\mu - \bar{x})}{v\sqrt{n}}, v = S/\nu\right) \propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}
\end{aligned}
$$

Example of the Gaussian law with Conjugate priors

- Unknown μ, Unknown v with Conjugate priors:

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

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

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

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

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

$$p(\boldsymbol{x}, \boldsymbol{y} | \alpha_1, \alpha_2, v) \propto v^{-n/2} \exp\left\{ -\frac{1}{2v} \sum_i (y_i - \alpha_0 + \alpha_1 (x - \bar{x})^2 \right\}$$

Example of Normal Linear Regression

- Defining

$$\begin{array}{ll} \widehat{\alpha}_1 & = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \\ S_e & = \sum_i (y_i - \bar{y})^2 - \widehat{\alpha}_1 \sum_i (x_i - \bar{x})(y_i - \bar{y}) \\ S_x & = \sum_i (x_i - \bar{x})^2 \end{array}$$

- Likelihood:

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

- Flat priors on α_0 and α_1 and Jeffrey's on v:

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

Example of Normal Linear Regression

- Marginalizing over α_1:

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

- Marginalizing over v:

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

- Marginalizing over α_0 and then over v:

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

- Marginalizing over α_0 and then over α_1:

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

Example of Normal Linear Regression

- Marginalizing over v

$$p(\alpha_0, \alpha_1 | \boldsymbol{x}, \boldsymbol{y}) \propto \left(S_e + n(\alpha_0 - \widehat{\alpha}_0)^2 + S_x(\alpha_1 - \widehat{\alpha}_1)^2\right)^{-n/2}$$

- Note that $p(\alpha_0, \alpha_1 | \boldsymbol{x}, \boldsymbol{y}) \neq p(\alpha_0 | \boldsymbol{x}, \boldsymbol{y})\, p(\alpha_1 | \boldsymbol{x}, \boldsymbol{y})$, so a posteriori α_0 and α_1 are not independent, but they are uncorrelated because $\text{cov}[\alpha_0, \alpha_1 | \boldsymbol{x}, \boldsymbol{y}] = 0$.
- If we come back to β_0 and β_1:

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

$$\widehat{\beta}_0 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \bar{x}, \quad \widehat{\beta}_1 = \widehat{\alpha}_1$$

- 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: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$
- Likelihood

$$\begin{array}{ll} p(\boldsymbol{y}|\boldsymbol{\theta}) & \propto v^{-n} \exp\left\{-\frac{1}{2v}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\theta})'(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\theta})\right\} \\ & \propto v^{-n} \exp\left\{-\frac{S}{2v}\right\} \end{array}$$

- Maximum Likelihood:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{y}$$

- A simple calculation:

$$\begin{array}{ll} S & = (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\theta})'(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\theta}) \\ & = (\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}) + (\boldsymbol{y}-\boldsymbol{A}\widehat{\boldsymbol{\theta}})'(\boldsymbol{y}-\boldsymbol{A}\widehat{\boldsymbol{\theta}}) \\ & = S_l(\boldsymbol{\theta}) + S_e \end{array}$$

- Flat prior on $\boldsymbol{\theta}$ and Jeffrey's Prior on v

$$\begin{array}{ll} p(\boldsymbol{\theta}, v|\boldsymbol{y}) & \propto v^{-(n+1)} \exp\left\{-\frac{1}{2v} S_l(\boldsymbol{\theta}) + S_e\right\} \\ & \propto v^{-n} \exp\left\{-\frac{S}{2v}\right\} \end{array}$$

General Linear Model

- Marginalizing over all θ_i:

$$\begin{array}{rl} p(\boldsymbol{\theta}_{-i}, v|\boldsymbol{y}) & \propto v^{-(n)} \exp\left\{-\frac{1}{2v} S_l(\boldsymbol{\theta}_{-i}) + S_e\right\} \\ & \propto v^{-n} \exp\left\{-\frac{S}{2v}\right\} \end{array}$$

- Each marginalization over each θ_i loses one power of v.
- Marginalization over all $k-1$ variables $\boldsymbol{\theta}_{-i}$ and then over v, we get a $\mathcal{S}_t$ distribution with $n-k$ degrees of freedom where

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

  and m_{kk} is the (kk) element of the design matrix $\boldsymbol{M} = (\boldsymbol{A}'\boldsymbol{A})^{-1}$
- The posterior marginal of v is Inverse Gamma

General Linear Model

- Informative prior: $p(\theta|\boldsymbol{\theta}_0, \boldsymbol{V}_0)$

$$p(\boldsymbol{\theta}, v|\boldsymbol{y}) \begin{array}{l} \propto v^{-(n+1)} \exp\left\{-\frac{1}{2v} S_l(\boldsymbol{\theta}) + S_e\right\} p(\theta|\boldsymbol{\theta}_0, \boldsymbol{V}_0) \\ \propto v^{-n} \exp\left\{-\frac{S}{2v} - (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \boldsymbol{V}_0^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\} \end{array}$$

- When v is known:

$$\widehat{\boldsymbol{\theta}} = (\frac{1}{v} \boldsymbol{A}' \boldsymbol{A} + \boldsymbol{V}_0)^{-1} (\frac{1}{v} \boldsymbol{A}' \boldsymbol{y} + \boldsymbol{V}_0 \boldsymbol{\theta}_0)$$

- One can continue other calculations
- More extentions:
  - Integrate out $v \longrightarrow \mathcal{S}_t$ distribution
  - Laplace (Gaussian) approximation of $_t$ distribution around its maximum $\widehat{\boldsymbol{\theta}}$
  - Prediction, ...

Preliminaries on Bayesian inference

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

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

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

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

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

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

Model selection

- Frequentist view: Likelihood ratio:

$$\frac{p(\boldsymbol{g}|\boldsymbol{\theta};\mathcal{M}_1)}{p(\boldsymbol{g}|\boldsymbol{\theta};\mathcal{M}_2)} \quad \text{or} \quad \frac{p(\boldsymbol{g}|\widehat{\boldsymbol{\theta}};\mathcal{M}_1)}{p(\boldsymbol{g}|\widehat{\boldsymbol{\theta}};\mathcal{M}_2)} \quad \text{or} \quad \frac{p(\boldsymbol{g}|\mathcal{M}_1)}{p(\boldsymbol{g}|\mathcal{M}_2)}$$

- Bayesian view:

$$p(\mathcal{M}_1|\boldsymbol{\theta},\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{\theta};\mathcal{M}_1)\; p(\boldsymbol{\theta}|\mathcal{M}_1)\; P(\mathcal{M}_1)$$

$$p(\mathcal{M}_1|\boldsymbol{g}) \propto p(\boldsymbol{g}|\mathcal{M}_1)\; P(\mathcal{M}_1)$$

$$\frac{p(\mathcal{M}_1|\boldsymbol{g},\boldsymbol{\theta})}{p(\mathcal{M}_2|\boldsymbol{\theta},\boldsymbol{g})} \quad \text{or} \quad \frac{p(\mathcal{M}_1|\boldsymbol{g},\widehat{\boldsymbol{\theta}})}{p(\mathcal{M}_2|\boldsymbol{g},\widehat{\boldsymbol{\theta}})} \quad \text{or} \quad \frac{p(\mathcal{M}_1|\boldsymbol{g})}{p(\mathcal{M}_2|\boldsymbol{g})}$$

2. Bayesian inference for inverse problems

Example: Measuring variation of temperature with a therometer

- $f(t)$ variation of temperature over time
- $g(t)$ variation of length of the liquid in thermometer
- Forward model $\mathcal{M}$: Convolution

$$g(t) = \int f(t')\, h(t-t')\, \mathrm{d}t' + \epsilon(t)$$

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

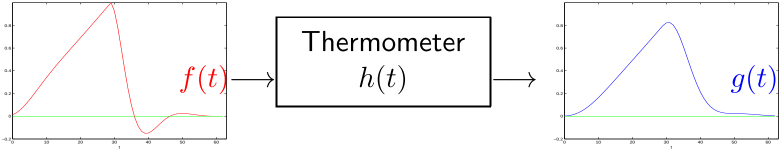
- Inverse problem: Deconvolution

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

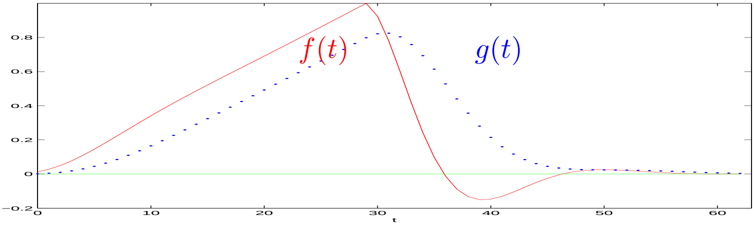
Measuring variation of temperature with a therometer

Forward model: Convolution

$$g(t) = \int f(t')\, h(t-t')\, \mathrm{d}t' + \epsilon(t)$$



Inversion: Deconvolution



Convolution/Deconvolution: Discrete forme

$$g(t) = h(t) * f(t) + \epsilon(t) = \int h(t') f(t - t') \, \mathrm{d}t' + \epsilon(t)$$

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

Matrix-Vector form: $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$

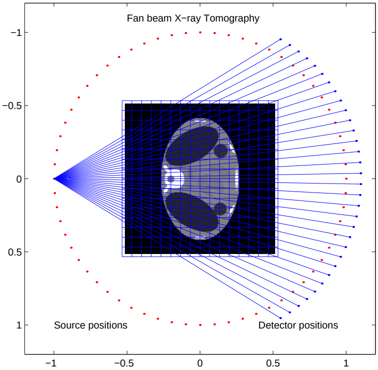
$$\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix} = \begin{bmatrix} h(p) & \cdots & h(0) & \cdots & h(-q) & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & \ddots & & \ddots & & & & \vdots \\ \vdots & & & & & & & & & \vdots \\ \vdots & & & h(p) & \cdots & h(0) & \cdots & h(-q) & & \vdots \\ \vdots & & & & & & & & & \vdots \\ \vdots & & & & & \ddots & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & h(p) & \cdots & h(0) & \cdots & h(-q) \end{bmatrix} \begin{bmatrix} f(-p) \\ \vdots \\ f(0) \\ f(1) \\ \vdots \\ f(M) \\ f(M+1) \\ \vdots \\ f(M+q) \end{bmatrix} + \begin{bmatrix} \epsilon(0) \\ \epsilon(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \epsilon(M) \end{bmatrix}$$

$$
\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix}
=
\begin{bmatrix}
h(0) & & & & & \\
h(1) & \ddots & & & & \\
\vdots & & & & & \\
h(p) & \cdots & & h(0) & & \\
0 & \ddots & & & \ddots & \\
\vdots & & & & & \\
0 & \cdots & 0 & h(p) & \cdots & h(0)
\end{bmatrix}
\begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ f(M) \end{bmatrix}
$$

$$
\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix}
=
\begin{bmatrix}
h(0) & & & & & & h(p) & \cdots & h(1) \\
h(1) & \ddots & & & & & & \ddots & \vdots \\
\vdots & & & & & & & & h(p) \\
h(p) & \cdots & & h(0) & & & & & 0 \\
0 & \ddots & & & \ddots & & & & \vdots \\
\vdots & & & & & & & & \\
0 & \cdots & 0 & h(p) & \cdots & h(0) & 0 & \cdots & 0
\end{bmatrix}
\begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ f(M) \\ 0 \\ \vdots \\ 0 \end{bmatrix}
$$

Computed tomography (CT)

A Multislice CT Scanner

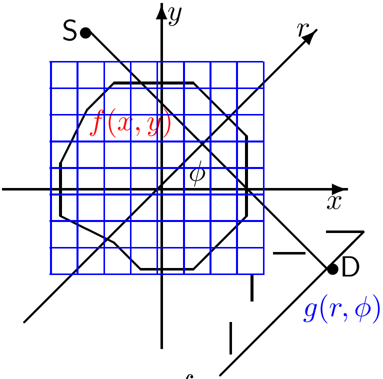


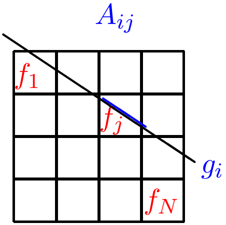
$$g(s_i) = \int_{L_i} f(\boldsymbol{r}) \; \mathrm{d}l_i + \epsilon(s_i)$$

Discretization

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

Computed Tomography





$$f(x,y) = \sum_j f_j \, b_j(x,y)$$

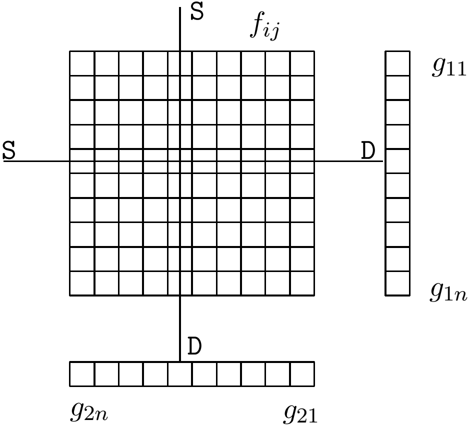
$$b_j(x,y) = \begin{cases} 1 & \text{if } (x,y) \in \text{ pixel } j \\ 0 & \text{else} \end{cases}$$

$$g(r,\phi) = \int_L f(x,y) \, \mathrm{d}l$$

$$g_i = \sum_{j=1}^{N} A_{ij} \, f_j + \epsilon_i$$

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

Computed Tomography with two projections



g_{1i} vertical projection

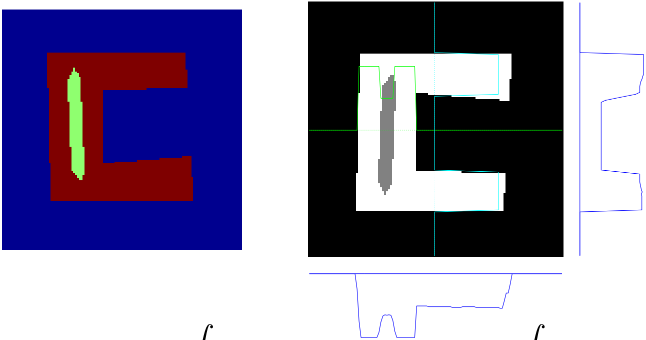
g_{2j} horizontal projection

$$\sum_{j=1}^{n} f_{ij} = g_{1i}, \quad i = 1, \cdots m$$

$$\sum_{i=1}^{m} f_{ij} = g_{2j}, \quad j = 1, \cdots n$$

Application of CT in NDT

Reconstruction from only 2 projections

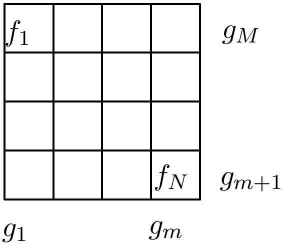


$$g_1(x) = \int f(x,y)\,\mathrm{d}y, \qquad g_2(y) = \int f(x,y)\,\mathrm{d}x$$

- 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:

$$\int \Omega(x,y)\,\mathrm{d}x = 1 \quad \text{and} \quad \int \Omega(x,y)\,\mathrm{d}y = 1$$

$$
\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ \vdots \\ g_8 \end{bmatrix}
=
\begin{bmatrix}
1000100010001000 \\
0100010001000100 \\
0010001000100010 \\
0001000100010001 \\
0000000000001111 \\
0000000011110000 \\
0000111100000000 \\
1111000000000000
\end{bmatrix}
\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ \vdots \\ \vdots \\ f_{16} \end{bmatrix}
$$



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

- Forward problem: Given $\boldsymbol{f}$ compute $\boldsymbol{g}$
- Inverse problem: Given $\boldsymbol{g}$ estimate $\boldsymbol{f}$

– Existance

– Uniqueness

– Stability

Existance and Uniqueness:

				40
	?			40
				40
				40
40	40	40	40	

10	10	10	10	40
10	10	10	10	40
10	10	10	10	40
10	10	10	10	40
40	40	40	40	

40	0	0	0	40
0	40	0	0	40
0	0	40	0	40
0	0	0	40	40
40	40	40	40	

15	5	5	15	40
5	15	15	5	40
5	15	15	5	40
15	5	5	15	40
40	40	40	40	

20	0	0	20	40
0	20	20	0	40
0	20	20	0	40
20	0	0	20	40
40	40	40	40	

20	10	5	5	40
10	20	5	5	40
5	5	20	10	40
5	5	10	20	40
40	40	40	40	

Stability:

$$\boldsymbol{A} = \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \quad \boldsymbol{A}^{-1} = \begin{bmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix} \longrightarrow \boldsymbol{f} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

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

$$\frac{||\delta \boldsymbol{g}||}{||\boldsymbol{g}||} = \frac{1}{300} \longrightarrow \frac{||\delta \boldsymbol{f}||}{||\boldsymbol{f}||} = \frac{10}{1}$$

$$\frac{||\delta \boldsymbol{f}||}{||\boldsymbol{f}||} = \text{cond}(\boldsymbol{A}) \frac{||\delta \boldsymbol{g}||}{||\boldsymbol{g}||}$$

Bayesian inference for inverse problems

- Linear Inverse problems: $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$
- Bayesian inference:

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

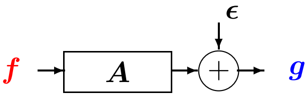
  with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$
- Point estimators:
  - Maximum A Posteriori (MAP)

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

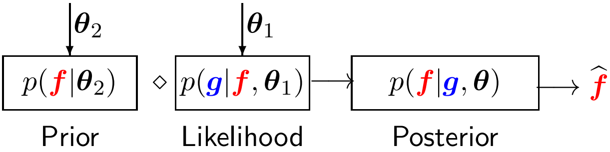
  - Posterior Mean (PM)

$$\widehat{\boldsymbol{f}} = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})}\{\boldsymbol{f}\} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{f}$$

Simple Bayesian Inference: Known hyperparameters $\boldsymbol{\theta}$



Forward model



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

Case of linear models and Gaussian priors

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

- Hypothesis on the noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I}) \longrightarrow$
  $p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2\right\}$
- Hypothesis on $\boldsymbol{f}$: $\boldsymbol{f} \sim \mathcal{N}(0, \sigma_f^2 (\boldsymbol{D}^t\boldsymbol{D})^{-1}) \longrightarrow$
  $p(\boldsymbol{f}) \propto \exp\left\{-\frac{1}{2\sigma_f^2}\|\boldsymbol{D}\boldsymbol{f}\|^2\right\}$
- A posteriori:
  $$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 - \frac{1}{2\sigma_f^2}\|\boldsymbol{D}\boldsymbol{f}\|^2\right\}$$
- MAP : $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$
  with $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2, \qquad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$
- Advantage : characterization of the solution

$$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{P}}) \text{ with } \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{P}}\boldsymbol{A}'\boldsymbol{g}, \quad \widehat{\boldsymbol{P}} = \left(\boldsymbol{A}'\boldsymbol{A} + \lambda \boldsymbol{D}^t\boldsymbol{D}\right)^{-1}$$

Simple Bayesian Model and Estimation

- Example 1: Linear Gaussian case

$$\begin{cases} p(\boldsymbol{g}|\boldsymbol{f},\theta_1) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f},\theta_1\boldsymbol{I}) \\ p(\boldsymbol{f}|\boldsymbol{\theta}_2) = \mathcal{N}(0,\theta_2\boldsymbol{I}) \end{cases} \longrightarrow p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta}) = \mathcal{N}(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{P}})$$

with

$$\begin{cases} \widehat{\boldsymbol{P}} = (\boldsymbol{A}'\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}, \quad \lambda = \frac{\theta_1}{\theta_2} \\ \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{P}}\boldsymbol{A}'\boldsymbol{g} \end{cases}$$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \text{ with } J(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|_2^2 + \lambda\|\boldsymbol{f}\|_2^2$$

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

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \text{ with } J(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}\|_2^2 + \lambda\|\boldsymbol{f}\|_1$$

MAP estimation with other priors:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{with} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\Omega(\boldsymbol{f})$$

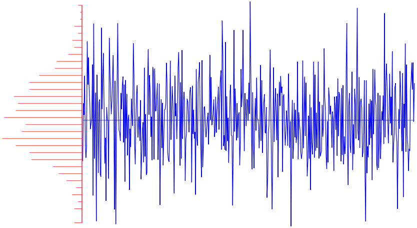
Separable priors:

- Gaussian: $p(f_j) \propto \exp\left\{-\alpha|f_j|^2\right\} \longrightarrow \Omega(\boldsymbol{f}) = \alpha \sum_j |f_j|^2$
- Gamma:
  $p(f_j) \propto f_j^\alpha \exp\{-\beta f_j\} \longrightarrow \Omega(\boldsymbol{f}) = \alpha \sum_j \ln f_j + \beta f_j$
- Beta:
  $p(f_j) \propto f_j^\alpha (1-f_j)^\beta \longrightarrow \Omega(\boldsymbol{f}) = \alpha \sum_j \ln f_j + \beta \sum_j \ln(1-f_j)$
- Generalized Gaussian:
  $p(f_j) \propto \exp\left\{-\alpha|f_j|^\beta\right\}, \; 1<\beta<2 \longrightarrow \Omega(\boldsymbol{f}) = \alpha \sum_j |f_j|^\beta,$

Markovian models:

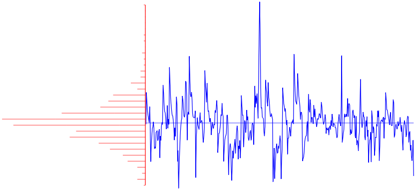
$$p(f_j|\boldsymbol{f}) \propto \exp\left\{-\alpha \sum_{i\in N_j} \phi(f_j, f_i)\right\} \longrightarrow \quad \Omega(\boldsymbol{f}) = \alpha \sum_j \sum_{i\in N_j} \phi(f_j, f_i),$$

Different prior models for signals and images: Separable



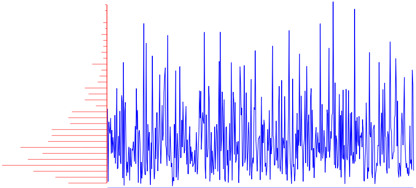
Gaussian

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



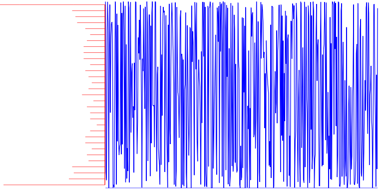
Generalized Gaussian

$$p(f_j) \propto \exp\left\{-\alpha|f_j|^p\right\}, \quad 1 \leq p \leq 2$$



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

$$p(f_j|\boldsymbol{f}) \propto \exp\left\{-\alpha \sum_{i\in v_j} \phi(f_j, f_i)\right\} \longrightarrow \quad \Phi(\boldsymbol{f}) = \alpha \sum_j \sum_{i\in V_j} \phi(f_j, f_i)$$

- 1D case and one neigbor $V_j = j-1$:

$$\Phi(\boldsymbol{f}) = \alpha \sum_j \phi(f_j - f_{j-1})$$

- 1D Case and two neighbors $V_j = \{j-1, j+1\}$:

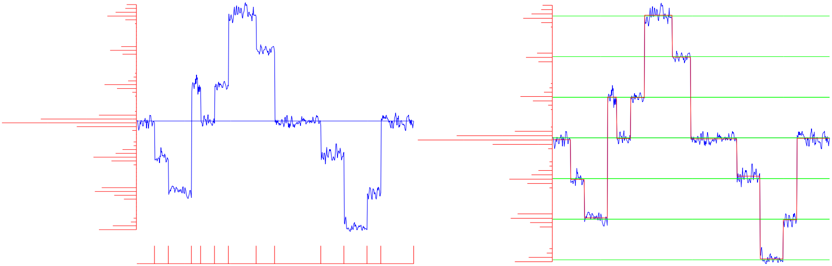
$$\Phi(\boldsymbol{f}) = \alpha \sum_j \phi\left(f_j - \beta(f_{j-1} + f_{j-1})\right)$$

- 2D case with 4 neighbors:

$$\Phi(\boldsymbol{f}) = \alpha \sum_{\boldsymbol{r}\in\mathcal{R}} \phi\left(f(\boldsymbol{r}) - \beta \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} f(\boldsymbol{r}')\right)$$

- $\phi(t) = |t|^\gamma$: Generalized Gaussian

Different prior models: Markovian with hidden variables



Piecewise Gaussians

(contours hidden variables)

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

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

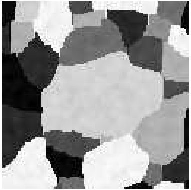
Mixture of Gaussians (MoG)

(regions labels hidden variables)

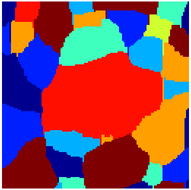
$p(f_j|z_j = k) = \mathcal{N}\left(m_k, \sigma_k^2\right)$ & z_j markovi

$$p(\boldsymbol{f}|\boldsymbol{z}) \propto \exp\left\{-\alpha \sum_k \sum_{j\in\mathcal{R}_k} \left(\frac{f_j - m_k}{\sigma_k}\right)^2\right\}$$

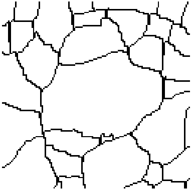
Gauss-Markov-Potts prior models for images



$f(\boldsymbol{r})$



$z(\boldsymbol{r})$



$c(\boldsymbol{r}) = 1 - \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))$

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

$$p(f(\boldsymbol{r})) = \sum_k P(z(\boldsymbol{r}) = k)\, \mathcal{N}(m_k, v_k) \quad \text{Mixture of Gaussians}$$

- Separable iid hidden variables: $p(\boldsymbol{z}) = \prod_{\boldsymbol{r}} p(z(\boldsymbol{r}))$
- Markovian hidden variables: $p(\boldsymbol{z})$ Potts-Markov:

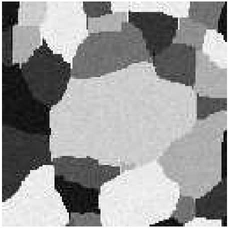
$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left\{\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

$$p(\boldsymbol{z}) \propto \exp\left\{\gamma \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

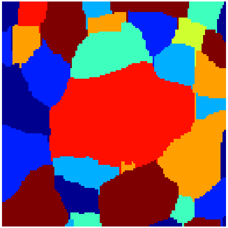
Four different cases

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

- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ iid :
  Mixture of Gaussians
- $\boldsymbol{f}|\boldsymbol{z}$ Gauss-Markov, $\boldsymbol{z}$ iid :
  Mixture of Gauss-Markov
- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Independent Gaussians
  (MIG with Hidden Potts)
- $\boldsymbol{f}|\boldsymbol{z}$ Markov, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Gauss-Markov
  (MGM with hidden Potts)



$f(\boldsymbol{r})$



$z(\boldsymbol{r})$

Full Bayesian, Joint MAP, Marginalization

- Unknown hyperparameters $\boldsymbol{\theta}$
- Full Bayesian: Joint Posterior:

$$p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)\, p(\boldsymbol{\theta})$$

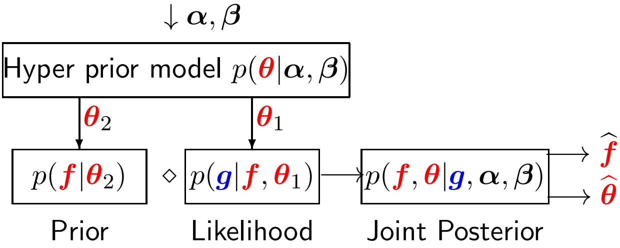
- Joint MAP:

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

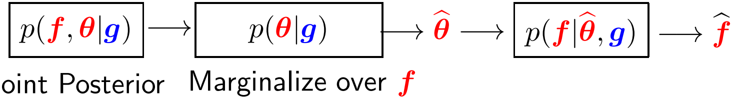
- Iterative algorithm:

$$\left\{\begin{array}{ll} \widehat{\boldsymbol{f}}^{(k)} & = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}^{(k-1)},\boldsymbol{g})\right\} \\ \widehat{\boldsymbol{\theta}}^{(k)} & = \arg\max_{\boldsymbol{\theta}} \left\{p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}}^{(k-1)},\boldsymbol{g})\right\} \end{array}\right.$$

Full Bayesian Model and Hyperparameter Estimation



Full Bayesian Model and Hyperparameter Estimation scheme



Marginalization for Hyperparameter Estimation

Full Bayesian: Marginal MAP and PM estimates

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

$$p(\boldsymbol{\theta}|\boldsymbol{g}) = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f} \propto p(\boldsymbol{g}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

  and then $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})\right\}$ or

  Posterior Mean: $\widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})\, \mathrm{d}\boldsymbol{f}$

- Needs the expression of the Likelihood:

$$p(\boldsymbol{g}|\boldsymbol{\theta}) = \int p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)\, \mathrm{d}\boldsymbol{f}$$

  Not always analytically available $\longrightarrow$ EM, SEM and GEM algorithms

Full Bayesian approach

$$\mathcal{M}: \qquad \boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$$

- Forward & errors model: $\longrightarrow p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1; \mathcal{M})$
- Prior models $\longrightarrow p(\boldsymbol{f}|\boldsymbol{\theta}_2; \mathcal{M})$
- Hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \longrightarrow p(\boldsymbol{\theta}|\mathcal{M})$
- Bayes: $\longrightarrow p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M}) = \frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\, p(\boldsymbol{f}|\boldsymbol{\theta};\mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M})}{p(\boldsymbol{g}|\mathcal{M})}$
- Joint MAP: $(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\}$
- Marginalization: $\begin{cases} p(\boldsymbol{f}|\boldsymbol{g};\mathcal{M}) & = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f} \\ p(\boldsymbol{\theta}|\boldsymbol{g};\mathcal{M}) & = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{\theta} \end{cases}$
- Posterior means: $\begin{cases} \widehat{\boldsymbol{f}} & = \int\int \boldsymbol{f}\, p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta} \\ \widehat{\boldsymbol{\theta}} & = \int\int \boldsymbol{\theta}\, p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta} \end{cases}$
- Evidence of the model:

$$p(\boldsymbol{g}|\mathcal{M}) = \iint p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}; \mathcal{M}) p(\boldsymbol{f}|\boldsymbol{\theta}; \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta}$$

Full Bayesian: EM and GEM algorithms

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

$$\left\{\begin{array}{ll} \text{E-step:} & Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(k)}) = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{\theta}}^{(k)})}\{\ln p(\boldsymbol{g},\boldsymbol{f}|\boldsymbol{\theta})\} \\ \text{M-step:} & \widehat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}}\left\{Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(k-1)})\right\} \end{array}\right.$$

- GEM (Bayesian) algorithm:

$$\left\{\begin{array}{ll} \text{E-step:} & Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(k)}) = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{\theta}}^{(k)})}\{\ln p(\boldsymbol{g},\boldsymbol{f}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})\} \\ \text{M-step:} & \widehat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}}\left\{Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(k-1)})\right\} \end{array}\right.$$

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

Variational Bayesian Approximation

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

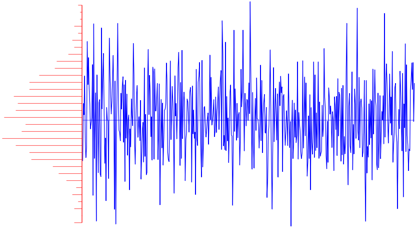
$$\left\{ \begin{array}{ll} \widehat{q}_1(\boldsymbol{f}) & \propto \exp\left\{ \langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{\widehat{q}_2(\boldsymbol{\theta})} \right\} \\ \widehat{q}_2(\boldsymbol{\theta}) & \propto \exp\left\{ \langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{\widehat{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

$$\boxed{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \boxed{\text{Variational Bayesian Approximation}} \begin{array}{l} \longrightarrow q_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}} \\ \longrightarrow q_2(\widehat{\boldsymbol{\theta}}) \longrightarrow \widehat{\boldsymbol{\theta}} \end{array}$$

Two main steps in the Bayesian approach

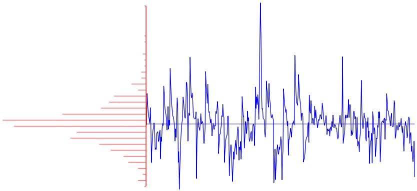
- Prior modeling
  - Separable:
    Gaussian, Generalized Gaussian, Gamma,
    mixture of Gaussians, mixture of Gammas, ...
  - Markovian: Gauss-Markov, GGM, ...
  - Separable or Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects
  - MAP, 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



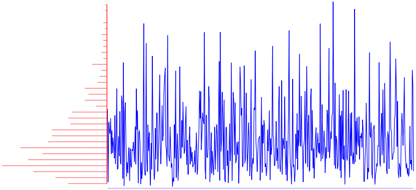
Gaussian

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



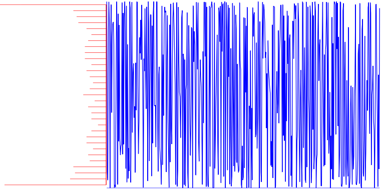
Generalized Gaussian

$$p(f_j) \propto \exp\left\{-\alpha |f_j|^p\right\}, \quad 1 \leq p \leq 2$$



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

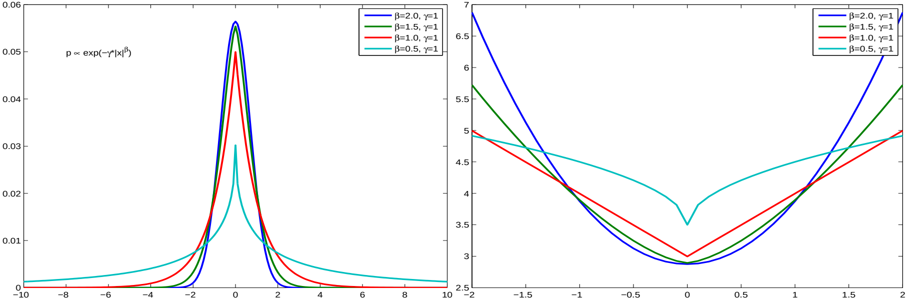
- Simple heavy tailed models:
  - Generalized Gaussian, Double Exponential
  - Symmetric Weibull, Symmetric Rayleigh
  - Student-t, Cauchy
  - Generalized hyperbolic
  - Elastic net
- Hierarchical mixture models:
  - Mixture of Gaussians
  - Bernoulli-Gaussian
  - Mixture of Gammas
  - Bernoulli-Gamma
  - Mixture of Dirichlet
  - Bernoulli-Multinomial

Simple heavy tailed models

- Generalized Gaussian, Double Exponential

$$p(\boldsymbol{f}|\gamma,\beta) = \prod_j \mathcal{GG}(f_j|\gamma,\beta) \propto \exp\left\{-\gamma \sum_j |f_j|^\beta\right\}$$

$\beta = 1$ Double exponential or Laplace.
$0 < \beta \leq 1$ are of great interest for sparsity enforcing.



Generalized Gaussian family

Simple heavy tailed models

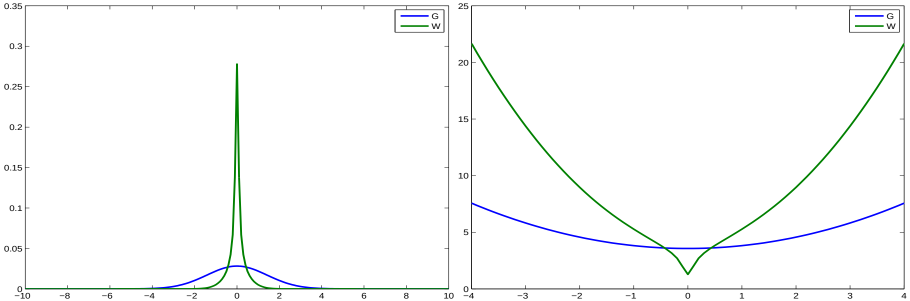
- Symmetric Weibull

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

$\beta = 2$ is the Symmetric Rayleigh distribution.

$\beta = 1$ is the Double exponential and

$0 < \beta \leq 1$ are of great interest for sparsity enforcing.



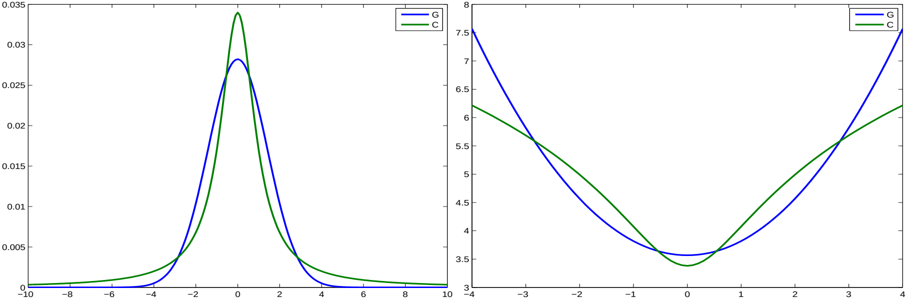
Symmetric Weibull family

Simple heavy tailed models

- Student-t and Cauchy models

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{S}t(f_j|\nu) \propto \exp\left\{-\frac{\nu+1}{2}\sum_j \log\left(1+f_j^2/\nu\right)\right\}$$

Cauchy model is obtained when $\nu = 1$.

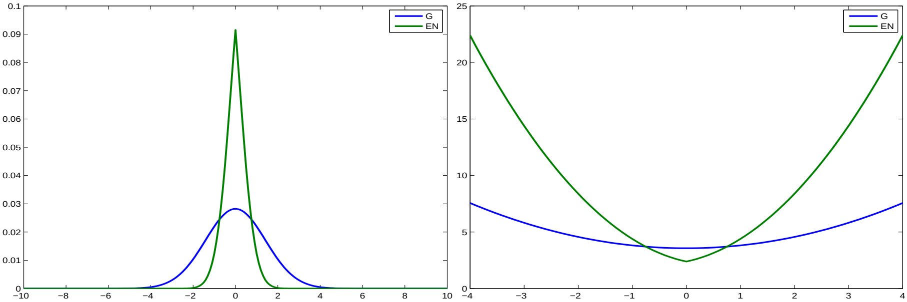


Student-t and Cauchy families

Simple heavy tailed models

- Elastic net prior model

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{EN}(f_j|\nu) \propto \exp\left\{-\sum_j (\gamma_1|f_j| + \gamma_2 f_j^2)\right\}$$

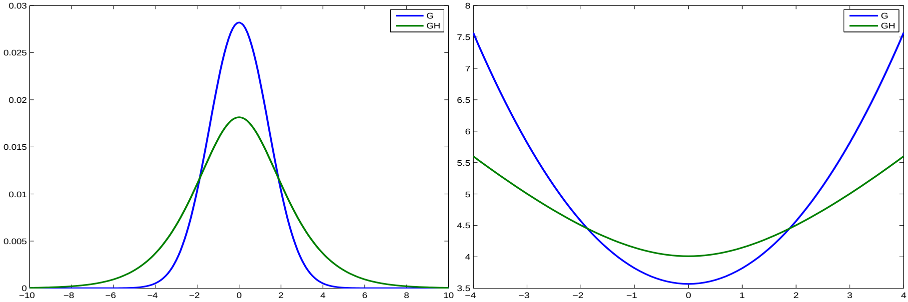


Elastic Net family

Simple heavy tailed models

- Generalized hyperbolic (GH) models

$$p(\boldsymbol{f}|\delta,\nu,\beta) = \prod_j (\delta^2 + f_j^2)^{(\nu-1/2)/2} \exp\{\beta x)\} K_{\nu-1/2}(\alpha\sqrt{\delta^2 + f_j^2})$$



Generalized hyperbolic family

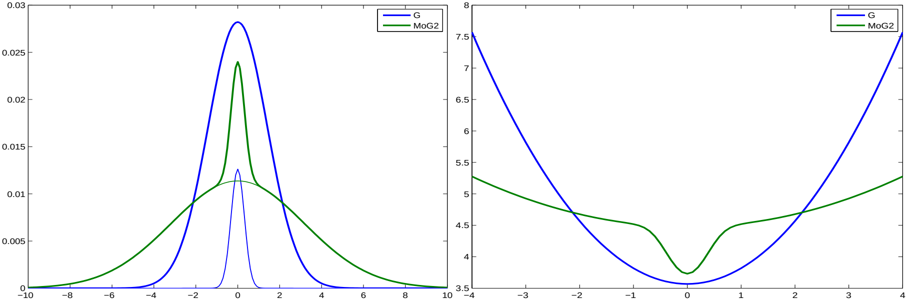
Mixture models

- Mixture of two Gaussians (MoG2) model

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

- Bernoulli-Gaussian (BG) model

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



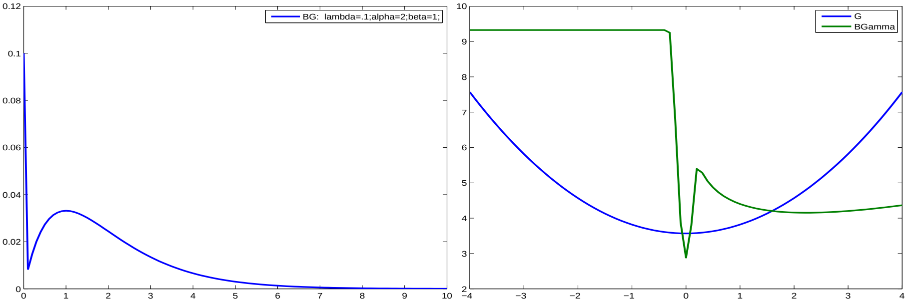
Mixture of 2 Gaussians families

- Mixture of Gammas

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

- Bernoulli-Gamma model

$$p(\boldsymbol{f}|\lambda, \alpha, \beta) = \prod_j \left[\lambda \mathcal{G}(f_j|\alpha, \beta) + (1-\lambda)\delta(f_j)\right]$$



Bernoulli-Gamma families

- Mixture of Dirichlets model

$$p(\boldsymbol{f}|\lambda, \boldsymbol{a}_1, \boldsymbol{\alpha}_1, \boldsymbol{a}_2, \boldsymbol{\alpha}_2) = \prod_j \left[\lambda \mathcal{D}(f_j|\boldsymbol{a}_1, \boldsymbol{\alpha}_1) + (1-\lambda)\mathcal{D}(f_j|\boldsymbol{a}_2, \boldsymbol{\alpha}_2)\right]$$

where

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

where $\boldsymbol{a} = \{a_1, \cdots, a_K\}$ and $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_K\}$
with $\sum_k \alpha_k = \alpha$ and $\sum_k a_k = 1$.

- Bernoulli-Multinomial (BMultinomial) model

$$p(\boldsymbol{f}|\lambda, \boldsymbol{a}, \boldsymbol{\alpha}) = \prod_j \left[\lambda \delta(f_j) + (1-\lambda)\mathcal{M}ult(f_j|\boldsymbol{a}, \boldsymbol{\alpha})\right]$$

Hierarchical models and hidden variables

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

- Example 1: MoG model:

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

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

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

Hierarchical models and hidden variables

- Example 2: Student-t model

$$\begin{aligned}\mathcal{S}t(f|\nu) &\propto \exp\left\{-\frac{\nu+1}{2}\log\left(1+f^2/\nu\right)\right\} \\ &= \int_0^\infty \mathcal{N}(f|,0,1/z)\,\mathcal{G}(z|\alpha,\beta)\;\mathrm{d}z, \quad \text{with } \alpha=\beta=\nu/2\end{aligned}$$

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

Hierarchical models and hidden variables

- Example 3: Laplace (Double Exponential) model

$$\mathcal{DE}(f|a) = \frac{a}{2} \exp\{-a|f|\} = \int_0^\infty \mathcal{N}(f|,0,z)\,\mathcal{E}(z|a^2/2)\,\mathrm{d}z, \quad a > 0$$

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

- With these models we have:

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\; p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)\; p(\boldsymbol{z}|\boldsymbol{\theta}_3)\; p(\boldsymbol{\theta})$$

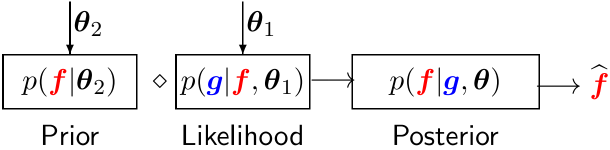
Bayesian Computation and Algorithms

- 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

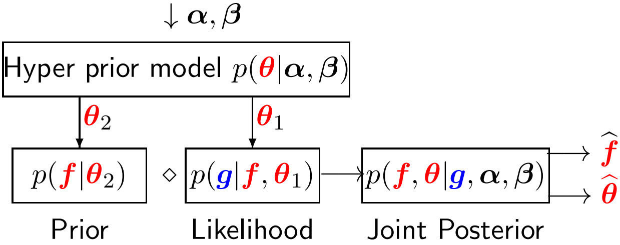
$$q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$$

  and do any computations with these separable ones.

Summary of Bayesian estimation with different levels

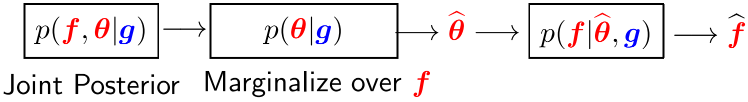


Simple Bayesian Model and Estimation

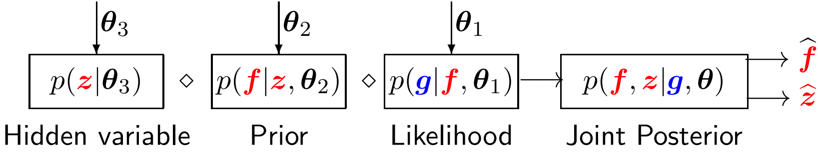


Full Bayesian Model and Hyperparameter Estimation scheme

Summary of Bayesian estimation with different levels

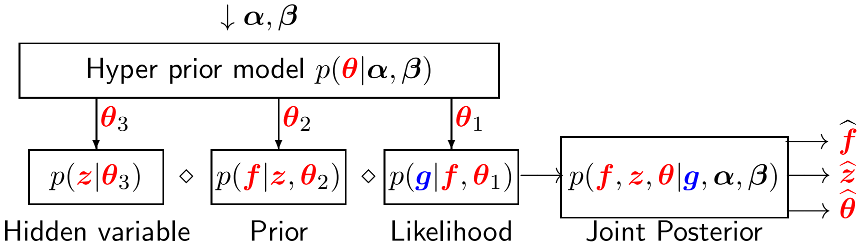


Marginalization for Hyperparameter Estimation

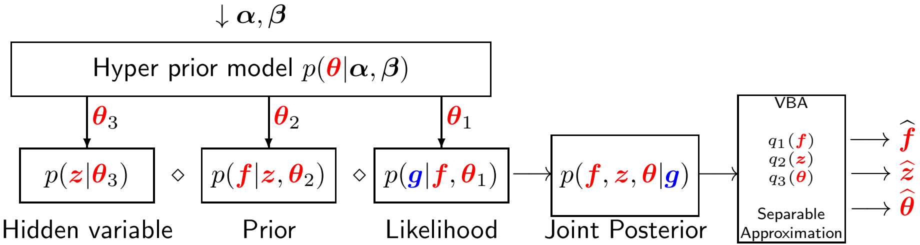


Full Bayesian Model with a Hierarchical Prior Model

Summary of Bayesian estimation with different levels



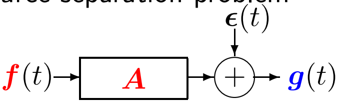
Full Bayesian Hierarchical Model with Hyperparameter Estimation



Full Bayesian Hierarchical Model and Variational Approximation

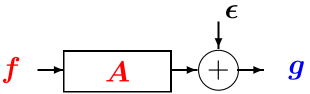
3. Bayesian Inference for Sources Separation

- Source separation problem



$t = 1, \cdots, T$

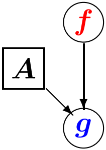
- Stationary case



- 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

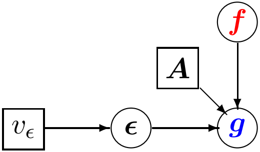
Estimation of sources $\boldsymbol{f}$ with known mixing matrix $\boldsymbol{A}$



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

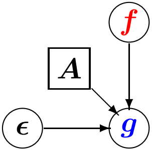
Exact Model without errors

$$\boldsymbol{f} = \boldsymbol{A}^{+}\boldsymbol{g} = \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}\boldsymbol{g}$$



Maximum Likelihood

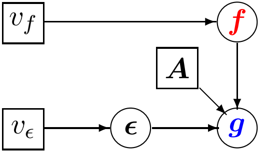
$$p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon)$$



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

Realistic Model with errors

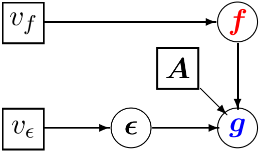
$$\boldsymbol{f} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2\right\}$$



Bayesian Estimation

$$p(\boldsymbol{f}|\boldsymbol{g}, v_\epsilon) \propto p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon)\, p(\boldsymbol{f}|v_f)$$

Estimation of sources $\boldsymbol{f}$ with known mixing matrix $\boldsymbol{A}$

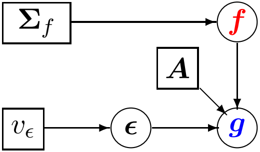


Uncorrelated Gaussian

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

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$$
$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$$

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



Correlated Gaussian

$$p(\boldsymbol{f}|v_f\boldsymbol{\Sigma}_f) = \mathcal{N}(0, v_f \boldsymbol{\Sigma}_f)$$
$$\propto \exp\left\{-\frac{1}{2v_f}\|\boldsymbol{D}\boldsymbol{f}\|^2\right\},$$

with $\boldsymbol{\Sigma}_f = (\boldsymbol{D}^t\boldsymbol{D})^{-1}$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$$
$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2$$

$$\widehat{\boldsymbol{f}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda\boldsymbol{D}^t\boldsymbol{D})^{-1}\boldsymbol{A}'\boldsymbol{g}$$

Estimation of the mixing matrix $\boldsymbol{A}$ with known sources $\boldsymbol{f}$

- Bilinear problems:

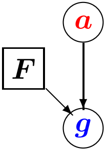
$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} = \boldsymbol{F}\boldsymbol{a}$$

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

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

- Estimation of $\boldsymbol{f}$ with known $\boldsymbol{A}$: $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$
- Estimation of $\boldsymbol{A}$ with known $\boldsymbol{f}$: $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} = \boldsymbol{F}\boldsymbol{a}$
  Underdetermination:
  Needs constraints or prior information on $\boldsymbol{A}$
- Joint Estimation of $\boldsymbol{f}$ and $\boldsymbol{A}$: $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} = \boldsymbol{F}\boldsymbol{a}$
  Underdetermination:
  Needs constraints or prior information on $\boldsymbol{A}$ and on $\boldsymbol{f}$

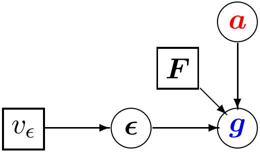
Estimation of the mixing matrix $\boldsymbol{A}$ with known sources $\boldsymbol{f}$



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

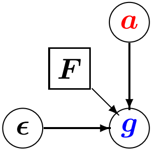
Exact Model without Errors

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



Maximul Likelihood

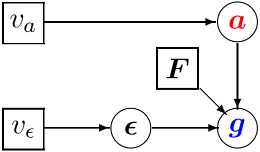
$$p(\boldsymbol{g}|\boldsymbol{a}, v_\epsilon)$$



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

Realistic Model with errors

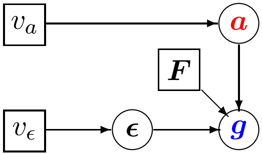
$$\boldsymbol{a} = \arg\min_{\boldsymbol{a}} \left\{ \|\boldsymbol{g} - \boldsymbol{F}\boldsymbol{a}\|^2 \right\}$$



Bayesian Estimation

$$p(\boldsymbol{a}|\boldsymbol{g}, v_\epsilon) \propto p(\boldsymbol{g}|\boldsymbol{a}, v_\epsilon)\, p(\boldsymbol{a}|v_a)$$

Estimation of the mixing matrix $\boldsymbol{A}$ with known sources $\boldsymbol{f}$

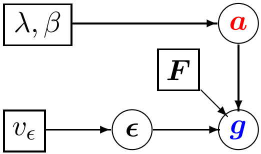


Gaussian

$$p(\boldsymbol{a}|v_f) = \mathcal{N}(0, v_a\boldsymbol{I})$$
$$\propto \exp\left\{-\frac{1}{2v_a}\sum_j |a_j|^2\right\}$$

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \{J(\boldsymbol{a})\}$$
$$J(\boldsymbol{a}) = \|\boldsymbol{g} - \boldsymbol{F}\boldsymbol{a}\|^2 + \lambda \sum_j |a_j|^2$$

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



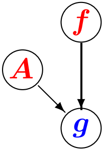
Generalized Gaussian

$$p(\boldsymbol{a}|\lambda, \beta) = \mathcal{GG}(\lambda, \beta)$$
$$\propto \exp\left\{-\lambda \sum_j |a_j|^\beta\right\}$$

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \{J(\boldsymbol{a})\}$$
$$J(\boldsymbol{a}) = \|\boldsymbol{g} - \boldsymbol{F}\boldsymbol{a}\|^2 + \lambda \sum_j |a_j|^\beta$$

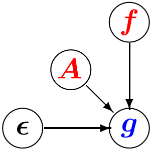
No analytic expression

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



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

Exact Model withour errors



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

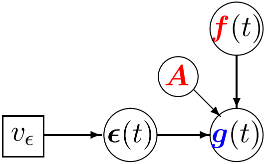
Realistic Model with errors

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

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{f}}) = \arg \min_{(\boldsymbol{A}>0, \boldsymbol{f}>0)} \left\{ \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 \right\}$$

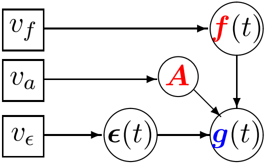
$$\begin{cases} \widehat{\boldsymbol{f}}(t) = (\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}})^{-1}\widehat{\boldsymbol{A}}'\boldsymbol{g}(t) & \text{Apply positiviy } \widehat{\boldsymbol{f}} > 0 \\ \widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t)\widehat{\boldsymbol{f}}'(t) \left(\sum_t \widehat{\boldsymbol{f}}(t)\widehat{\boldsymbol{f}}'(t) \right)^{-1} & \text{Apply positiviy } \widehat{\boldsymbol{A}} > 0 \end{cases}$$

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



Maximul Likelihood

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

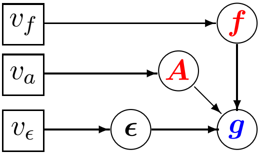


Bayesian Estimation

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

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

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

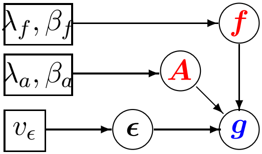


Gaussian

$$p(\boldsymbol{f}|v_f) = \mathcal{N}(0, v_f \boldsymbol{I})$$
$$\propto \exp\left\{-\frac{1}{2v_f}\sum_j |f_j|^2\right\}$$

Gaussian

$$p(\boldsymbol{a}|v_f) = \mathcal{N}(0, v_a \boldsymbol{I})$$
$$\propto \exp\left\{-\frac{1}{2v_a}\sum_j |a_j|^2\right\}$$



Gaussian

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

Generalized Gaussian

$$p(\boldsymbol{a}|\lambda_a, \beta_a) = \mathcal{GG}(\lambda_a, \beta_a)$$
$$\propto \exp\left\{-\lambda_a \sum_j |a_j|^\beta\right\}$$

- Joint Posterior:

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

- Integration over $\boldsymbol{f}$ can be down easily

4. Links with classical methods: PCA, ICA, Neural Networks, ...

Inversion ou Decomposition

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

$\boldsymbol{g}(t) \longrightarrow$ Inversion $\longrightarrow \widehat{\boldsymbol{f}}(t)$, $\longrightarrow \widehat{\boldsymbol{A}}$

Given that $\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t)$, estimate $\boldsymbol{f}(t)$ and $\boldsymbol{A}$

- Maximum Likelihood
- Bayesian estimation

$\boldsymbol{g}(t) \longrightarrow$ Decomposition $\longrightarrow$ $\boldsymbol{y}$(t)=$\phi(\boldsymbol{B}\boldsymbol{g}$(t))

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

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

Link between $\boldsymbol{y}$ and $\widehat{\boldsymbol{f}}$ and between $\boldsymbol{B}$ and $\widehat{\boldsymbol{A}}$?

Prior modeling

Links with classical methods: PCA

Classical PCA:

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) \longrightarrow \text{cov}[\boldsymbol{g}] = \boldsymbol{A}\text{cov}[\boldsymbol{f}]\boldsymbol{A}'$$

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

- Indetermination:
  Note that any $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{U}$ with $\boldsymbol{R}$ any orthogonal matrix is also a solution

Link with PCA

Probabilistic PCA:

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t) \longrightarrow \text{cov}[\boldsymbol{g}] = \boldsymbol{A}\text{cov}[\boldsymbol{f}]\boldsymbol{A}' + \text{cov}[\boldsymbol{\epsilon}]$$

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

Link with PCA

Bayesian Probabilistic PCA:

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

- $p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), \sigma_\epsilon^2) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), \sigma_\epsilon^2 \boldsymbol{I})$
- $p(\boldsymbol{f}(t)|\boldsymbol{\Lambda}) = \mathcal{N}(0, \boldsymbol{\Lambda})$
- $p(\boldsymbol{g}(t)|\boldsymbol{A}, \sigma_\epsilon^2, \boldsymbol{\Lambda}) = \mathcal{N}(0, \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}' + \sigma_\epsilon^2 \boldsymbol{I})$
- $P(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{V}_0)$
- $p(\boldsymbol{A}, \boldsymbol{f}(t)|\boldsymbol{g}(t)) \propto p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), \sigma_\epsilon^2)\, p(\boldsymbol{f}(t)|\boldsymbol{\Lambda})\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)$
- Integration over $\boldsymbol{f}(t)$:

$$\begin{aligned} p(\boldsymbol{A}|\boldsymbol{g}(t), \sigma_\epsilon^2, \boldsymbol{\Lambda}) &\propto p(\boldsymbol{g}(t)|\boldsymbol{A}, \sigma_\epsilon^2, \boldsymbol{\Lambda})\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) \\ &\propto |\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}' + \sigma_\epsilon^2 \boldsymbol{I}|^{-1/2} \exp\left\{\boldsymbol{g}'[\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}' + \sigma_\epsilon^2 \boldsymbol{I}]^{-1}\boldsymbol{g}\right\} \\ &\quad \exp\left\{-\tfrac{1}{2}\sum_i \sum_j [\boldsymbol{A} - \boldsymbol{A}_0]_{ij}^2 / [\boldsymbol{V}_0]_{ij} \|\right\} \end{aligned}$$

- MAP solution can be computed iteratively
- Full Bayesian:

$$p(\boldsymbol{A}, \sigma_\epsilon^2, \boldsymbol{\Lambda}|\boldsymbol{g}(t)) \propto p(\boldsymbol{g}(t)|\boldsymbol{A}, \sigma_\epsilon^2, \boldsymbol{\Lambda})\, p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)\, p(\boldsymbol{\Lambda}|\Lambda_0, \alpha_0)\, p(\sigma_\epsilon^2|\alpha_0, \beta_0)$$

Link with ICA

Classical ICA:

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) \longrightarrow p_g(\boldsymbol{g}(t)|\boldsymbol{A}) = \det(\boldsymbol{A})\, p_f(\boldsymbol{A}^{-1}\boldsymbol{g}(t))$$

- Noting by $\boldsymbol{B} = \boldsymbol{A}^{-1}$ and by $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{g}$ and assuming sources to be independent $p_f(\boldsymbol{f}) = \prod_n p_j(\boldsymbol{f}_j)$:

$$\ln p(\boldsymbol{g}_{1..T}|\boldsymbol{B}) = \frac{T}{2} \ln \det(\boldsymbol{B}) + \sum_t \sum_j p_j(y_j) + c$$

- Maximum Likelihood:

$$\widehat{\boldsymbol{B}} = \arg\max_{\boldsymbol{B}} \{J(\boldsymbol{B}) = \ln p(\boldsymbol{g}_{1..T}|\boldsymbol{B})\}$$

- Iterative Gradient based algorithm

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

$$\boldsymbol{\phi}(\boldsymbol{y}) = [\phi_1(y_1), \ldots, \phi_n(y_n)]^t \quad \text{and} \quad \phi_j(z) = -\frac{p_j'(z)}{p_j(z)}.$$

.

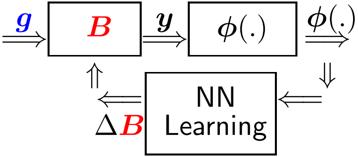
Gauss	$p(z) \propto \exp\left\{-\alpha z^2\right\}$	$\phi(z) = 2\alpha z$
Laplace	$p(z) \propto \exp\left\{-\alpha \vert z \vert\right\}$	$\phi(z) = \alpha \text{sign}(z)$
Cauchy	$p(z) \propto \dfrac{1}{1+(z/\alpha)^2}$	$\phi(z) = \frac{2z/\alpha^2}{1+(z/\alpha)^2}$
Gamma	$p(z) \propto z^\alpha \exp\left\{-\beta z\right\}$	$\phi(z) = -\alpha/z + \beta$
Sub-Gaussian	$p(z) \propto \exp\left\{-\dfrac{1}{2} z^2\right\} \text{sech}^2(z)$	$\phi(z) = z + \tanh(z)$
Mixture of Gauss.	$p(z) \propto \exp\left\{-\frac{1}{2}(z-\alpha)^2\right\} + \exp\left\{-\frac{1}{2}(z+\alpha)^2\right\}$	$\phi(z) = \alpha z - \alpha \tanh(\alpha z)$

Link with Learning in Neural Network

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

Optimization by a Gradient based algorithm:

$$\Delta \boldsymbol{B} \propto H(\boldsymbol{y}) = [\boldsymbol{I} - \boldsymbol{\phi}(\boldsymbol{y})\boldsymbol{y}^t]\boldsymbol{B}$$



5. Advanced Bayesian methods

Non Gaussian, Dependent, Colored sources

$$\boldsymbol{g}(t) = \boldsymbol{A}\,\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t) \longrightarrow \boldsymbol{g}_{1..T} = \boldsymbol{A}\boldsymbol{f}_{1..T} + \boldsymbol{\epsilon}_{1..T}$$

$$\ln p(\boldsymbol{g}_{1..T}|\boldsymbol{f}_{1..T}) = \sum_t \sum_i q_i(g_i(t) - [\boldsymbol{A}\boldsymbol{f}]_i(t))$$

$$\ln p(\boldsymbol{f}_{1..T}) = \sum_t \sum_j r_j(f_j(t)), \quad p(\boldsymbol{A}) \propto \exp\left\{-\frac{1}{2\sigma_a^2}\sum_i \sum_j a_{ij}^2\right\},$$

$$\ln p\left(\boldsymbol{A}, \boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T}\right) = \sum_t \sum_i q_i\left(g_i(t) - [\boldsymbol{A}\boldsymbol{f}]_i(t)\right) + \sum_t \sum_j r_j(f_j(t)) + \frac{1}{2\sigma_a^2}\sum_i \sum_j a_{ij}^2 + cte.$$

$$\left\{\begin{array}{lcl} \widehat{\boldsymbol{f}}^{(k)} & = & \arg\max_{\boldsymbol{f}} \left\{\sum_i q_i\left(g_i - [\boldsymbol{A}\boldsymbol{f}]_i\right) + \sum_j r_j(f_j)\right\} \\ \widehat{\boldsymbol{A}}^{(k)} & = & \arg\max_{\boldsymbol{A}} \left\{\sum_i q_i\left(g_i - [\boldsymbol{A}\boldsymbol{f}]_i\right) + \frac{1}{2\sigma_a^2}\sum_i \sum_j a_{ij}^2\right\} \end{array}\right.$$

Gaussian / Non Gaussian

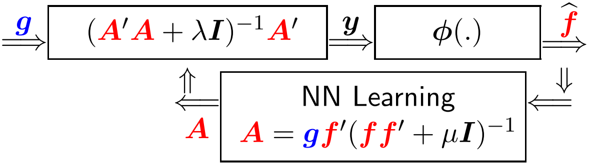
- **Gaussian laws** for the noise and sources :

$$\left\{\begin{array}{lcl} \boldsymbol{f}(t) & = & (\boldsymbol{A}'\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}\boldsymbol{A}'\boldsymbol{g}(t) \\ \boldsymbol{A} & = & \sum_t \boldsymbol{g}(t)\boldsymbol{f}'(t)(\sum_t \boldsymbol{f}(t)\boldsymbol{f}'(t)+\mu\boldsymbol{I})^{-1} \end{array}\right.$$

with $\lambda=\sigma_\epsilon^2/\sigma_s^2$ and $\mu=\sigma_\epsilon^2/\sigma_a^2$.

- **Non Gaussian sources** $\boldsymbol{f}(t)$**:**

$$\left\{\begin{array}{lcl} \boldsymbol{y}(t) & = & (\boldsymbol{A}'\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}\boldsymbol{A}'\boldsymbol{g}(t) \\ \boldsymbol{f}(t) & = & \phi(\boldsymbol{y}(t)) \\ \boldsymbol{A} & = & \sum_t \boldsymbol{g}(t)\boldsymbol{f}'(t)(\sum_t \boldsymbol{f}(t)\boldsymbol{f}'(t)+\mu\boldsymbol{I})^{-1} \end{array}\right.$$



Dependent sources

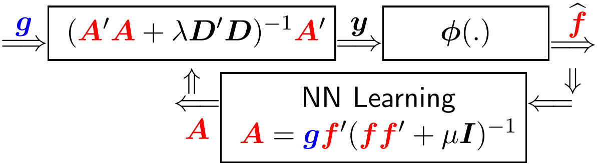
- **Gaussian laws** for the noise and sources :

$$\begin{cases} \boldsymbol{f}(t) & = & (\boldsymbol{A}'\boldsymbol{A} + \lambda \boldsymbol{D}'\boldsymbol{D})^{-1} \boldsymbol{A}'\boldsymbol{g}(t) \\ \boldsymbol{A} & = & \sum_t \boldsymbol{g}(t)\boldsymbol{f}'(t) (\sum_t \boldsymbol{f}(t)\boldsymbol{f}'(t) + \mu \boldsymbol{I})^{-1} \end{cases}$$

with $\lambda = \sigma_\epsilon^2/\sigma_s^2$ and $\mu = \sigma_\epsilon^2/\sigma_a^2$.

- **Non Gaussian sources $\boldsymbol{f}$:**

$$\begin{cases} \boldsymbol{y}(t) & = & (\boldsymbol{A}'\boldsymbol{A} + \lambda \boldsymbol{D}'\boldsymbol{D})^{-1} \boldsymbol{A}'\boldsymbol{g} \\ \boldsymbol{f}(t) & = & \phi(\boldsymbol{y}(t)) \\ \boldsymbol{A} & = & \sum_t \boldsymbol{g}(t)\boldsymbol{f}'(t) (\sum_t \boldsymbol{f}(t)\boldsymbol{f}'(t) + \mu \boldsymbol{I})^{-1} \end{cases}$$



Indpendent but temporally colored sources

IID case:

$$p(\boldsymbol{f}(t)) = \sum_j p_j(f_j(t)), \forall t \longrightarrow p(\boldsymbol{f}_{1..T}) = \sum_j \sum_t p_j(f_j(t))$$

Spatially indpendent but temporally colored sources

$$p(\boldsymbol{f}_{1..T}) = \sum_j p_j(f_j(1), ..f_j(T))$$

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

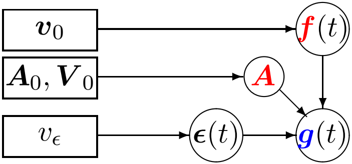
- Markovian Models:

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

- 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 $\boldsymbol{A}$ and $\boldsymbol{f}$ with a Gaussian prior model



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

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

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

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

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

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

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

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

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

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

Joint MAP:

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

Algorithm:

$$
\boldsymbol{A}^{(0)} \longrightarrow \widehat{\boldsymbol{A}} \longrightarrow \boxed{\left(\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}} + \lambda_f \boldsymbol{I}\right)^{-1}\widehat{\boldsymbol{A}}'\boldsymbol{g}} \longrightarrow \widehat{\boldsymbol{f}}(t)
$$

$$
\uparrow \qquad\qquad\qquad\qquad \downarrow
$$

$$
\widehat{\boldsymbol{A}} \longleftarrow \boxed{\sum_t \boldsymbol{g}(t)\widehat{\boldsymbol{f}}'(t)\left(\sum_t \widehat{\boldsymbol{f}}(t)\widehat{\boldsymbol{f}}'(t) + \lambda_a \boldsymbol{I}\right)^{-1}} \longleftarrow \widehat{\boldsymbol{f}}(t)
$$

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

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

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

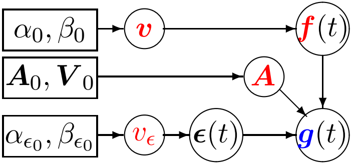
$$q_2(\boldsymbol{A}|\boldsymbol{g}(t), \boldsymbol{f}(t), v_\epsilon, \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{V}})$$
$$\left\{ \begin{array}{l} \widehat{\boldsymbol{V}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_f \widehat{\boldsymbol{\Sigma}})^{-1}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

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

Algorithm:

$$
\begin{array}{ccc}
\begin{array}{l} \boldsymbol{A}^{(0)} \longrightarrow \widehat{\boldsymbol{A}} \longrightarrow \\ \boldsymbol{V}^{(0)} \longrightarrow \widehat{\boldsymbol{V}} \longrightarrow \end{array} &
\boxed{\begin{array}{l} \widehat{\boldsymbol{f}}(t) = \left(\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}} + \lambda_f \widehat{\boldsymbol{V}}\right)^{-1} \widehat{\boldsymbol{A}}'\boldsymbol{g}(t) \\ \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{A}'\boldsymbol{A} + \lambda_f \widehat{\boldsymbol{V}})^{-1} \end{array}} &
\begin{array}{l} \longrightarrow \widehat{\boldsymbol{f}}(t) \\ \longrightarrow \widehat{\boldsymbol{\Sigma}} \end{array} \\
\Uparrow & & \Downarrow \\
\begin{array}{l} \widehat{\boldsymbol{A}} \longleftarrow \\ \widehat{\boldsymbol{V}} \longleftarrow \end{array} &
\boxed{\begin{array}{l} \widehat{\boldsymbol{A}} = \sum_t \boldsymbol{g}(t)\widehat{\boldsymbol{f}}'(t) \left(\sum_t \widehat{\boldsymbol{f}}(t)\widehat{\boldsymbol{f}}'(t) + \lambda_a \widehat{\boldsymbol{\Sigma}}\right)^{-1} \\ \widehat{\boldsymbol{V}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_f \widehat{\boldsymbol{\Sigma}})^{-1} \end{array}} &
\begin{array}{l} \longleftarrow \widehat{\boldsymbol{f}}(t) \\ \longleftarrow \widehat{\boldsymbol{\Sigma}} \end{array}
\end{array}
$$

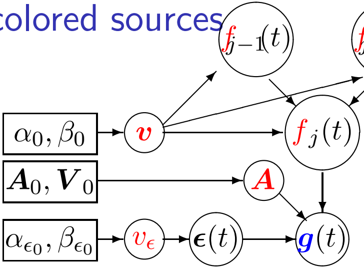
Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$, v_ϵ and $\boldsymbol{v}$



$$
\begin{aligned}
p(f_j(t)|v_{0j}) &= \mathcal{N}(0, v_j) \\
p(\boldsymbol{f}(t)|\boldsymbol{v}_0) &\propto \exp\left\{-\tfrac{1}{2}\sum_j f_j^2(t)/v_{0j}\right\} \\
p(\boldsymbol{A}_{m,n}|\boldsymbol{A}_{0m,n}, \boldsymbol{V}_{0m,n}) &= \mathcal{N}(\boldsymbol{A}_{0m,n}, \boldsymbol{V}_{0m,n}) \\
p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) &= \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{V}_0) \\
p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon) &= \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I}) \\
p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) &= \mathcal{G}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\
p(v_j|\alpha_0, \beta_0) &= \mathcal{G}(\alpha_0, \beta_0)
\end{aligned}
$$

$$
\begin{aligned}
p(\boldsymbol{f}_{1..T}, \boldsymbol{A}, v_\epsilon, \boldsymbol{v}|\boldsymbol{g}_{1..T}) \propto\ & p(\boldsymbol{g}_{1..T}|\boldsymbol{A}, \boldsymbol{f}_{1..T}, v_\epsilon)\, p(\boldsymbol{f}_{1..T}|\boldsymbol{v}) \\
& p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)\, p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \prod_j p(v_j|\alpha_0, \beta_0) \\
\propto\ & \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{v}(t)) \\
& p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)\, p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \prod_j p(v_j|\alpha_0, \beta_0)
\end{aligned}
$$

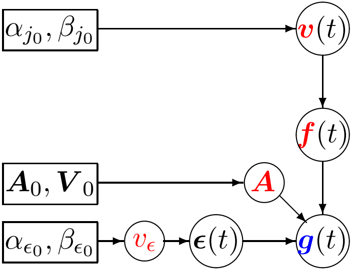
Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$, v_ϵ and $\boldsymbol{v}$ with dependent and colored sources



$p(f_j(t)|v_j) = \mathcal{N}((f_{j-1}(t) + f_{j+1}(t))/2,$
$p(\boldsymbol{f}(t)|\boldsymbol{v}) = \mathcal{N}(0, \boldsymbol{V})$ with $\boldsymbol{V} = \boldsymbol{D}'\text{diag}$
$p(\boldsymbol{A}_{m,n}|\boldsymbol{A}_{0m,n}, \boldsymbol{V}_{0m,n}) = \mathcal{N}(\boldsymbol{A}_{0m,n}, \boldsymbol{V}$
$p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{V}_0)$
$p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I})$
$p(v_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) = \mathcal{G}(\alpha_{\epsilon 0}, \beta_{\epsilon 0})$
$p(v_j|\alpha_0, \beta_0) = \mathcal{G}(\alpha_0, \beta_0)$

$$\begin{aligned} p(\boldsymbol{f}_{1..T}, \boldsymbol{A}, v_\epsilon, \boldsymbol{v}|\boldsymbol{g}_{1..T}) \propto \; & p(\boldsymbol{g}_{1..T}|\boldsymbol{A}, \boldsymbol{f}_{1..T}, v_\epsilon)\, p(\boldsymbol{f}_{1..T}|\boldsymbol{v}) \\ & p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)\, p(v_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) \prod_j p(v_j|\alpha_0, \beta_0) \\ \propto \; & \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{z}(t)) \\ & p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)\, p(v_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) \prod_j p(v_j|\alpha_0, \beta_0) \end{aligned}$$

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}$ with a variance modulated prior model inducing sparsity



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

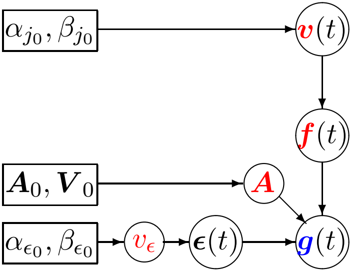
$$p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0) = \mathcal{N}(\boldsymbol{A}_0,\boldsymbol{V}_0)$$

$$p(\boldsymbol{g}(t)|\boldsymbol{A},\boldsymbol{f}(t),v_\epsilon) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I})$$
$$p(v_\epsilon|\alpha_{\epsilon 0},\beta_{\epsilon 0}) = \mathcal{G}(\alpha_{\epsilon 0},\beta_{\epsilon 0})$$

$$p(\boldsymbol{f}_{1..T},\boldsymbol{A},\boldsymbol{v}_{1..T},v_\epsilon|\boldsymbol{g}_{1..T}) \propto p(\boldsymbol{g}_{1..T}|\boldsymbol{A},\boldsymbol{f}_{1..T},v_\epsilon)\, p(\boldsymbol{f}_{1..T}|\boldsymbol{v}_{1..T}) \prod_j p(v_{j\,1..T}|\alpha_{j0},\beta_{j0})\; p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0)$$

$$\propto \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A},\boldsymbol{f}(t),v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{v}(t)) \prod_t\prod_j p(v_j(t)|\alpha_{j0},\beta_{j0})\; p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0)$$

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



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

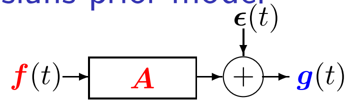
$$p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{V}_0)$$

$$p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I})$$
$$p(v_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) = \mathcal{G}(\alpha_{\epsilon 0}, \beta_{\epsilon 0})$$

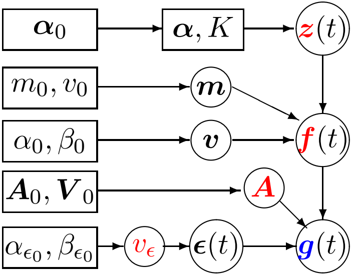
$$p(\boldsymbol{f}(t), \boldsymbol{A}, \boldsymbol{v}(t), v_\epsilon|\boldsymbol{g}(t)) \propto \quad p(\boldsymbol{g}(t)|\boldsymbol{A}, \boldsymbol{f}(t), v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{v}(t)) \prod_j p(v_j(t)|\alpha_{j0}, \beta_{j0})\; p(\boldsymbol{A}|\boldsymbol{A}_0, \boldsymbol{V}_0)$$

$$p(\boldsymbol{f}(t)|\boldsymbol{g}(t), \boldsymbol{A}, \boldsymbol{v}(t), v_\epsilon, \alpha_{j0}, \beta_{j0}) = \mathcal{N}(\widehat{\boldsymbol{f}}(t), \widehat{\boldsymbol{\Sigma}})$$
$$p(\boldsymbol{A}|\boldsymbol{g}(t), \boldsymbol{f}(t), v_\epsilon, \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{V}})$$
$$p(v_j(t)|f_j(t), \alpha_{j0}, \beta_{j0}) = \mathcal{G}(\widehat{\alpha}_j, \widehat{\beta}_j)$$
$$p(v_\epsilon|\boldsymbol{g}(t), \boldsymbol{f}(t), \alpha_{\epsilon 0}, \beta_{\epsilon 0}) = \mathcal{G}(\widehat{\alpha}_{\epsilon 0}, \widehat{\beta}_{\epsilon 0})$$

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}(t)$ with a Mixture of Gaussians prior model



$t = 1, \cdots, T$



$p(z_j(t) = k | \alpha_{j_k}) = \alpha_{j_k}, \quad \sum_k \alpha_{j_k} = 1$

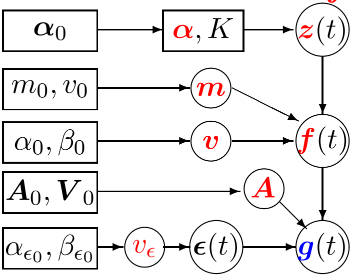
$p(f_j(t) | z_j(t) = k) = \mathcal{N}(m_{j_k}, v_{j_k})$
$p(\boldsymbol{f}(t) | \boldsymbol{z}(t), \boldsymbol{m}, \boldsymbol{v}) = \sum_j \mathcal{N}(m_{z_j(t)}, v_{z_j(t)})$

$p(\boldsymbol{A} | \boldsymbol{A}_0, \boldsymbol{V}_0) = \mathcal{N}(\boldsymbol{A}_0, \boldsymbol{V}_0)$

$p(\boldsymbol{g}(t) | \boldsymbol{f}(t), v_\epsilon) = \mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t), v_\epsilon \boldsymbol{I})$
$p(v_\epsilon | \alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\alpha_{\epsilon_0}, \beta_{\epsilon_0})$

$$p(\boldsymbol{f}_{1..T}, \boldsymbol{A}, \boldsymbol{z}_{1..T}, v_\epsilon | \boldsymbol{g}_{1..T}) \propto p(\boldsymbol{g}_{1..T} | \boldsymbol{A}, \boldsymbol{f}_{1..T}, v_\epsilon)\, p(\boldsymbol{f}_{1..T} | \boldsymbol{z}_{1..T}) \prod_j p(z_{j\,1..T} | \alpha_{j_0}, \beta_{j_0})\, p(\boldsymbol{A} | \boldsymbol{A}_0, \boldsymbol{V}_0)\, p(v_\epsilon | \alpha_{\epsilon_0}, \beta_{\epsilon_0})$$

Joint Estimation of $\boldsymbol{A}$, $\boldsymbol{f}(t)$ and $\boldsymbol{\theta}$ with mixture model

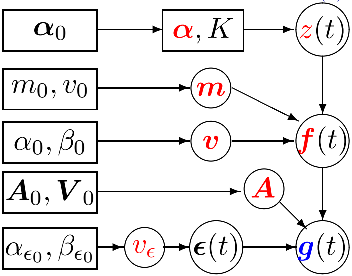


$$p(z_j(t)=k|\alpha_{j_k})=\alpha_{j_k}, \quad \sum_k \alpha_{j_k}=1$$
$$p(f_j(t)|z_j(t)=k)=\mathcal{N}(m_{j_k},v_{j_k})$$
$$p(\boldsymbol{f}(t)|\boldsymbol{z}(t),\boldsymbol{m},\boldsymbol{v})=\sum_j \mathcal{N}(m_{z_j(t)},v_{z_j(t)})$$
$$p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0)=\mathcal{N}(\boldsymbol{A}_0,\boldsymbol{V}_0)$$
$$p(\boldsymbol{g}(t)|\boldsymbol{f}(t),v_\epsilon)=\mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t),v_\epsilon\boldsymbol{I})$$
$$p(v_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{IG}(\alpha_{\epsilon_0},\beta_{\epsilon_0})$$
$$p(\boldsymbol{m}|m_0,v_0)=\mathcal{G}(\alpha_{\epsilon_0},\beta_{\epsilon_0})$$
$$p(\boldsymbol{v}|\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{IG}(\alpha_{\epsilon_0},\beta_{\epsilon_0})$$

$$\begin{aligned} p(\boldsymbol{f}_{1..T},\boldsymbol{A},\boldsymbol{z}_{1..T},v_\epsilon,\boldsymbol{m},\boldsymbol{v}|\boldsymbol{g}_{1..T}) \propto \; & \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A},\boldsymbol{f}(t),v_\epsilon)\, p(\boldsymbol{f}(t)|\boldsymbol{z}(t)) \\ & \prod_t \prod_j p(z_j(t)|\alpha_{j_0},\beta_{j_0}) \\ & p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0)\, p(v_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}) \\ & p(\boldsymbol{m}|m_0,v_0)\, p(\boldsymbol{v}|\alpha_{\epsilon_0},\beta_{\epsilon_0}) \end{aligned}$$

$$p(\boldsymbol{f}(t)|\boldsymbol{g}(t),\boldsymbol{A},\boldsymbol{z}(t),v_\epsilon,\alpha_{j_0},\beta_{j_0})=\mathcal{N}(\widehat{\boldsymbol{f}}(t),\widehat{\boldsymbol{\Sigma}})$$
$$p(\boldsymbol{A}|\boldsymbol{g}(t),\boldsymbol{f}(t),v_\epsilon,\boldsymbol{A}_0,\boldsymbol{V}_0)=\mathcal{N}(\widehat{\boldsymbol{A}},\widehat{\boldsymbol{V}})$$
$$p(z_j(t)=k|f_j(t),\alpha_{j_0},\beta_{j_0})=\widehat{\alpha}_{jk}$$
$$p(v_\epsilon|\boldsymbol{g}(t),\boldsymbol{f}(t),\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{G}(\widehat{\alpha}_{\epsilon_0},\widehat{\beta}_{\epsilon_0})$$
$$p(\boldsymbol{m}|\boldsymbol{f}(t),m_0,v_0)=\mathcal{N}(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{v}}) \quad p(v_j|f_j(t),\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{IG}(\widehat{\alpha}_j,\widehat{\beta}_j)$$

Joint Estimation of $\boldsymbol{A}$ and $\boldsymbol{f}(t)$ with mixture model (common z)



$$p(z(t)=k|\alpha_k)=\alpha_k, \quad \sum_k \alpha_k = 1$$
$$p(f_j(t)|z(t)=k)=\mathcal{N}(m_{j_k}, v_{j_k})$$
$$p(\boldsymbol{f}(t)|z(t),\boldsymbol{m},\boldsymbol{v})=\sum_j \mathcal{N}(m_{z_j(t)}, v_{z_j(t)})$$
$$p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0)=\mathcal{N}(\boldsymbol{A}_0,\boldsymbol{V}_0)$$
$$p(\boldsymbol{g}(t)|\boldsymbol{f}(t),v_\epsilon)=\mathcal{N}(\boldsymbol{A}\boldsymbol{f}(t),v_\epsilon\boldsymbol{I})$$
$$p(v_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{G}(\alpha_{\epsilon_0},\beta_{\epsilon_0})$$
$$p(\boldsymbol{m}|m_0,v_0)=\mathcal{G}(\alpha_{\epsilon_0},\beta_{\epsilon_0})$$
$$p(\boldsymbol{v}|\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{IG}(\alpha_{\epsilon_0},\beta_{\epsilon_0})$$

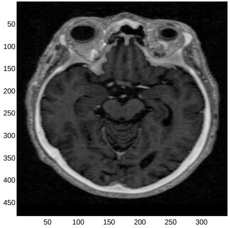
$$\begin{aligned} p(\boldsymbol{f}_{1..T},\boldsymbol{A},z_{1..T},v_\epsilon|\boldsymbol{g}_{1..T}) \propto\ & \prod_t p(\boldsymbol{g}(t)|\boldsymbol{A},\boldsymbol{f}(t),v_\epsilon)\, p(\boldsymbol{f}(t)|z(t)) \\ & \prod_t p(z(t)|\alpha_k,\beta_k) \\ & p(\boldsymbol{A}|\boldsymbol{A}_0,\boldsymbol{V}_0)\, p(v_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}) \\ & p(\boldsymbol{m}|m_0,v_0)\, p(\boldsymbol{v}|\alpha_{\epsilon_0},\beta_{\epsilon_0}) \end{aligned}$$

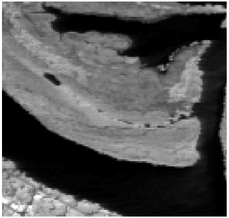
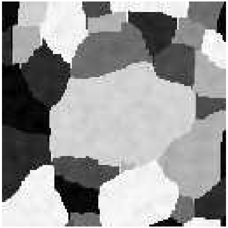
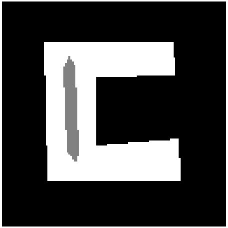
$$p(\boldsymbol{f}(t)|\boldsymbol{g}(t),\boldsymbol{A},z(t),v_\epsilon,\alpha_k,\beta_k)=\mathcal{N}(\widehat{\boldsymbol{f}}(t),\widehat{\boldsymbol{\Sigma}})$$
$$p(\boldsymbol{A}|\boldsymbol{g}(t),\boldsymbol{f}(t),v_\epsilon,\boldsymbol{A}_0,\boldsymbol{V}_0)=\mathcal{N}(\widehat{\boldsymbol{A}},\widehat{\boldsymbol{V}})$$
$$p(z(t)=k|\boldsymbol{f}(t),\alpha_{j_0},\beta_{j_0})=\widehat{\alpha}_k$$
$$p(v_\epsilon|\boldsymbol{g}(t),\boldsymbol{f}(t),\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{G}(\widehat{\alpha}_{\epsilon_0},\widehat{\beta}_{\epsilon_0})$$
$$p(\boldsymbol{m}|\boldsymbol{f}(t),m_0,v_0)=\mathcal{N}(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{v}}) \quad p(v_j|f_j(t),\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{IG}(\widehat{\alpha}_j,\widehat{\beta}_j)$$

6. A few applications

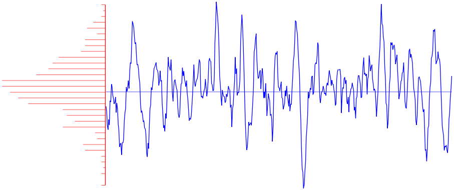
- X ray Computed Tomography for Non Destructive Testing
- Spectrometry (with Said Mousaoui)
- Cosmic Microwave Bacground in Radio Astronomy (with Hichem Snoussi)
- Sattelite image separation (with Mahiedding Ichir)
- Hyperspectral imaging (with Nadia Bali and Adel Mohammadpour)

Which images I am looking for?

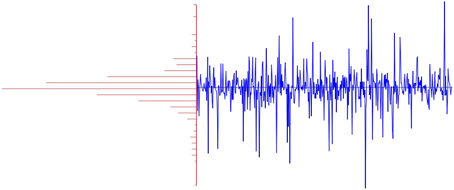



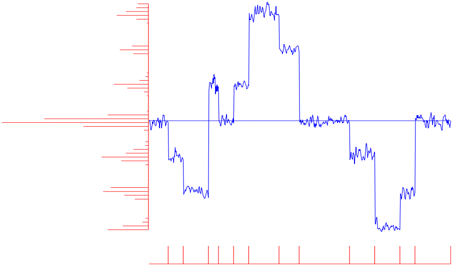
Which image I am looking for?



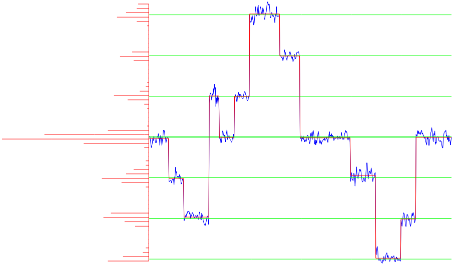
Gauss-Markov



Generalized GM



Piecewize Gaussian



Mixture of GM

Different prior models for signals and images

- Separable $p(\boldsymbol{f}) = \prod_j p_j(f_j) \propto \exp\left\{-\beta \sum_j \phi(f_j)\right\}$

$$p(\boldsymbol{f}) \propto \exp\left\{-\beta \sum_{\boldsymbol{r} \in \mathcal{R}} \phi(f(\boldsymbol{r}))\right\}$$

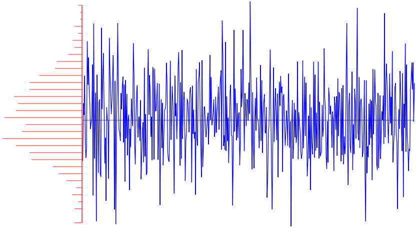
- Markoviens (simple) $p(f_j | f_{j-1}) \propto \exp\left\{-\beta \phi(f_j - f_{j-1})\right\}$

$$p(\boldsymbol{f}) \propto \exp\left\{-\beta \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \phi(f(\boldsymbol{r}), f(\boldsymbol{r}'))\right\}$$

- Markovien with hidden variables
  $z(\boldsymbol{r})$ (lines, contours, regions)

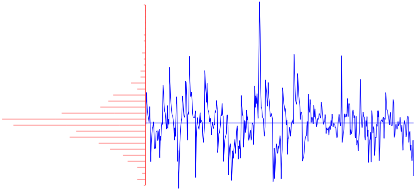
$$p(\boldsymbol{f} | \boldsymbol{z}) \propto \exp\left\{-\beta \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \phi(f(\boldsymbol{r}), f(\boldsymbol{r}'), z(\boldsymbol{r}), z(\boldsymbol{r}'))\right\}$$

Different prior models for images: Separable



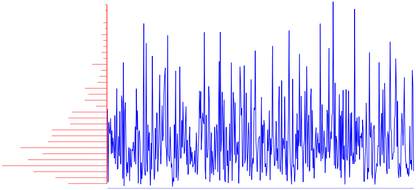
Gaussian

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



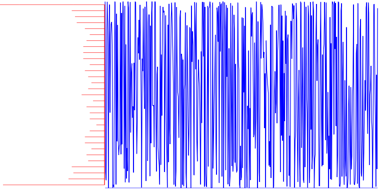
Generalized Gaussian

$$p(f_j) \propto \exp\left\{-\alpha |f_j|^p\right\}, \quad 1 \leq p \leq 2$$



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

$$p(f_j|\boldsymbol{f}) \propto \exp\left\{-\alpha \sum_{i\in v_j} \phi(f_j, f_i)\right\} \longrightarrow \quad \Phi(\boldsymbol{f}) = \alpha \sum_j \sum_{i\in V_j} \phi(f_j, f_i)$$

- 1D case and one neigbor $V_j = j-1$:

$$\Phi(\boldsymbol{f}) = \alpha \sum_j \phi(f_j - f_{j-1})$$

- 1D Case and two neighbors $V_j = \{j-1, j+1\}$:

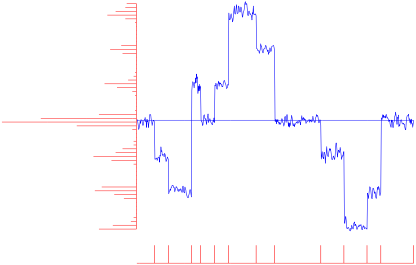
$$\Phi(\boldsymbol{f}) = \alpha \sum_j \phi\left(f_j - \beta(f_{j-1} + f_{j-1})\right)$$

- 2D case with 4 neighbors:

$$\Phi(\boldsymbol{f}) = \alpha \sum_{\boldsymbol{r}\in\mathcal{R}} \phi\left(f(\boldsymbol{r}) - \beta \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} f(\boldsymbol{r}')\right)$$

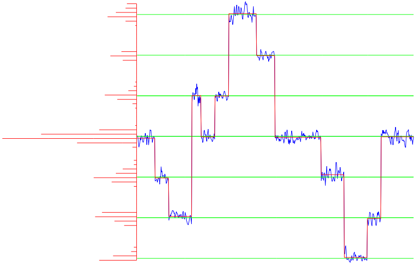
- $\phi(t) = |t|^\gamma$: Generalized Gaussian

Different prior models: Markovian with hidden variables



Piecewise Gaussians

(contours hidden variables)



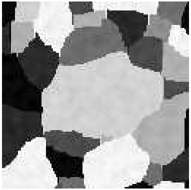
Mixture of Gaussians (MoG)

(regions labels hidden variables)

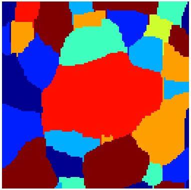
$p(f_j|q_j, f_{j-1}) = \mathcal{N}\left((1-q_j)f_{j-1}, \sigma_f^2\right)$ $p(f_j|z_j = k) = \mathcal{N}\left(m_k, \sigma_k^2\right)$ & z_j markovi

$p(\boldsymbol{f}|\boldsymbol{q}) \propto \exp\left\{-\alpha \sum_j \left|f_j - (1-q_j)f_{j-1}\right|^2\right\}$ $p(\boldsymbol{f}|\boldsymbol{z}) \propto \exp\left\{-\alpha \sum_k \sum_{j\in\mathcal{R}_k} \left(\frac{f_j - m_k}{\sigma_k}\right)^2\right\}$

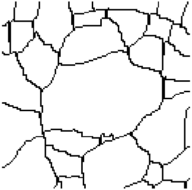
Gauss-Markov-Potts prior models for images



$f(\boldsymbol{r})$



$z(\boldsymbol{r})$



$c(\boldsymbol{r}) = 1 - \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))$

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

$$p(f(\boldsymbol{r})) = \sum_k P(z(\boldsymbol{r}) = k)\, \mathcal{N}(m_k, v_k) \quad \text{Mixture of Gaussians}$$

- Separable iid hidden variables: $p(\boldsymbol{z}) = \prod_{\boldsymbol{r}} p(z(\boldsymbol{r}))$
- Markovian hidden variables: $p(\boldsymbol{z})$ Potts-Markov:

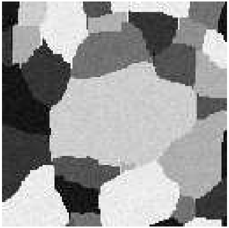
$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left\{\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

$$p(\boldsymbol{z}) \propto \exp\left\{\gamma \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

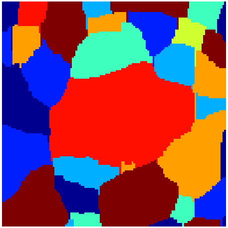
Four different cases

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

- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ iid :
  Mixture of Gaussians
- $\boldsymbol{f}|\boldsymbol{z}$ Gauss-Markov, $\boldsymbol{z}$ iid :
  Mixture of Gauss-Markov
- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Independent Gaussians
  (MIG with Hidden Potts)
- $\boldsymbol{f}|\boldsymbol{z}$ Markov, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Gauss-Markov
  (MGM with hidden Potts)

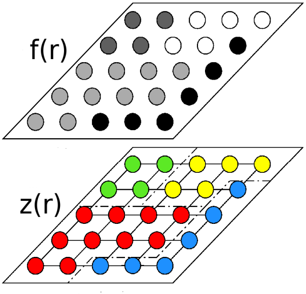


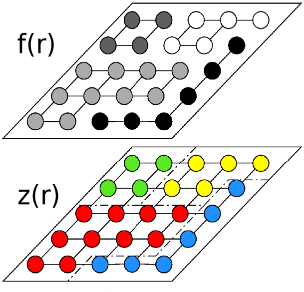
$f(\boldsymbol{r})$



$z(\boldsymbol{r})$

Summary of the two proposed models





$\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid
$\boldsymbol{z}$ Potts-Markov

(MIG with Hidden Potts)

$\boldsymbol{f}|\boldsymbol{z}$ Markov
$\boldsymbol{z}$ Potts-Markov

(MGM with hidden Potts)

Bayesian Computation

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{z}, v_\epsilon)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{v})\, p(\boldsymbol{z}|\gamma, \boldsymbol{\alpha})\, p(\boldsymbol{\theta})$$

$\boldsymbol{\theta} = \{v_\epsilon, (\alpha_k, m_k, v_k), k = 1, \cdot, K\}$ $\quad p(\boldsymbol{\theta})$ Conjugate priors

- Direct computation and use of $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})$ 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

  $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})$ by $q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$
  - Choice of approximation criterion : $KL(q : p)$
  - Choice of appropriate families of probability laws for $q_1(\boldsymbol{f})$, $q_2(\boldsymbol{z})$ and $q_3(\boldsymbol{\theta})$

MCMC based algorithm

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})\, p(\boldsymbol{z})\, p(\boldsymbol{\theta})$$

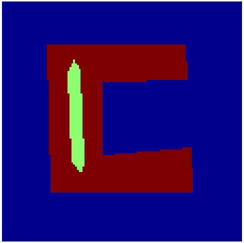
General scheme:

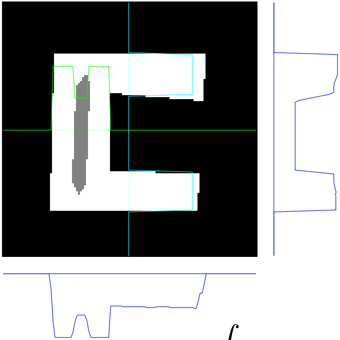
$$\widehat{\boldsymbol{f}} \sim p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{z}} \sim p(\boldsymbol{z}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{\theta}} \sim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

- Estimate $\boldsymbol{f}$ using $p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})\, p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}})$
  Needs optimisation of a quadratic criterion.
- Estimate $\boldsymbol{z}$ using $p(\boldsymbol{z}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}})\, p(\boldsymbol{z})$
  Needs sampling of a Potts Markov field.
- Estimate $\boldsymbol{\theta}$ using
  $p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}},\sigma_\epsilon^2\boldsymbol{I})\, p(\widehat{\boldsymbol{f}}|\widehat{\boldsymbol{z}},(m_k,v_k))\, p(\boldsymbol{\theta})$
  Conjugate priors $\longrightarrow$ analytical expressions.

Application of CT in NDT

Reconstruction from only 2 projections



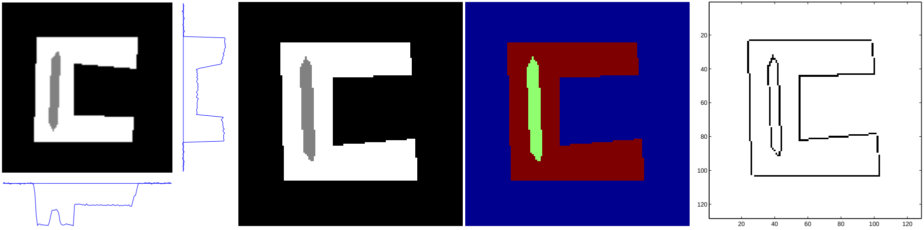


$$g_1(x) = \int f(x,y) \, \mathrm{d}y, \qquad g_2(y) = \int f(x,y) \, \mathrm{d}x$$

- 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:

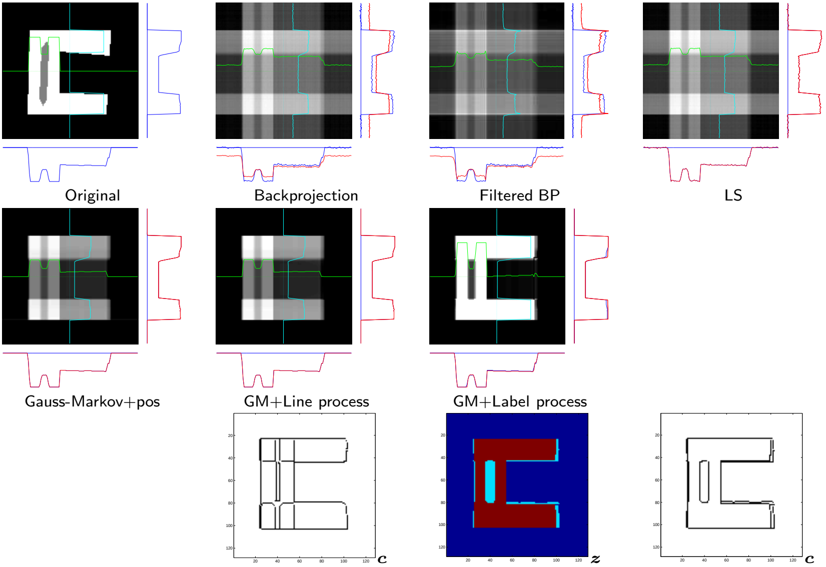
$$\int \Omega(x,y) \, \mathrm{d}x = 1 \quad \text{and} \quad \int \Omega(x,y) \, \mathrm{d}y = 1$$

Application in CT



$\boldsymbol{g}\vert\boldsymbol{f}$	$\boldsymbol{f}\vert\boldsymbol{z}$	$\boldsymbol{z}$	$\boldsymbol{c}$
$\boldsymbol{g} = \boldsymbol{H}\,\boldsymbol{f} + \boldsymbol{\epsilon}$	iid Gaussian	iid	$c(\boldsymbol{r}) \in \{0,1\}$
$\boldsymbol{g}\vert\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{H}\,\boldsymbol{f}, \sigma_\epsilon^2 \boldsymbol{I})$	or	or	$1 - \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))$
Gaussian	Gauss-Markov	Potts	binary

Results

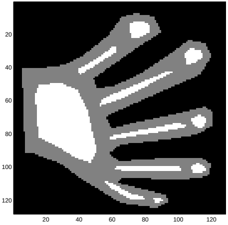

Original
Backprojection
Filtered BP
LS
Gauss-Markov+pos
GM+Line process
GM+Label process
c
z
c

Application in Microwave imaging

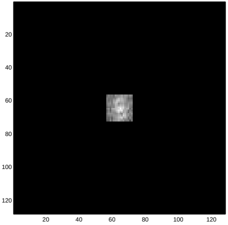
$$g(\boldsymbol{\omega}) = \int f(\boldsymbol{r}) \exp\left\{-j(\boldsymbol{\omega}.\boldsymbol{r})\right\} \, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{\omega})$$

$$g(u,v) = \iint f(x,y) \exp\left\{-j(ux+vy)\right\} \, \mathrm{d}x \, \mathrm{d}y + \epsilon(u,v)$$

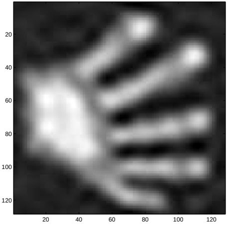
$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$



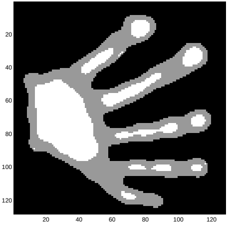
$f(x,y)$



$g(u,v)$

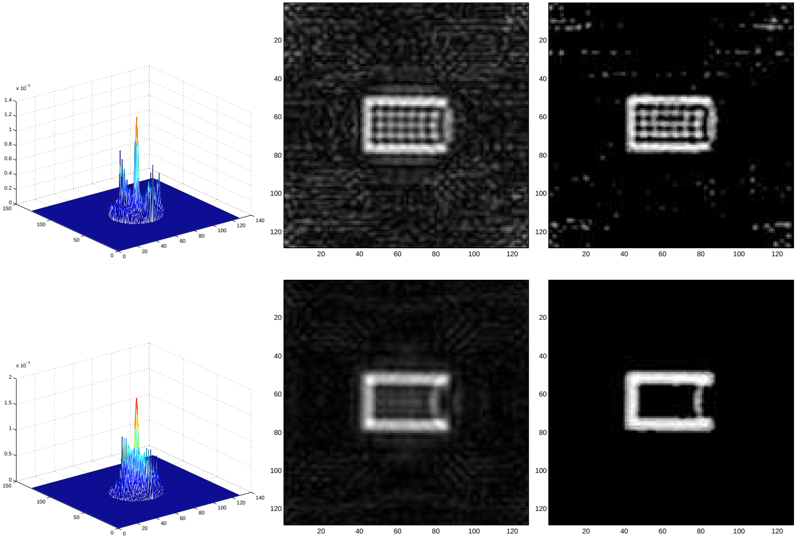


$\widehat{\boldsymbol{f}}$ IFT



$\widehat{\boldsymbol{f}}$ 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 :

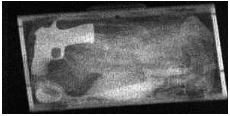
- Efficient implementation in 2D and 3D cases
- Evaluation of performances and comparison with MCMC methods
- Application to other linear and non linear inverse problems: (PET, SPECT or ultrasound and microwave imaging)

Images fusion and joint segmentation

(with O. Féron)

$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma^2_{i\,k}) \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \end{cases}$$

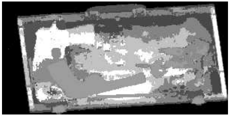
$\boldsymbol{g}_1$



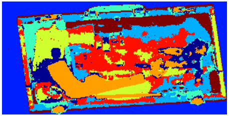
$\boldsymbol{g}_2$

$\longrightarrow$

$\widehat{\boldsymbol{f}}_1$



$\widehat{\boldsymbol{f}}_2$

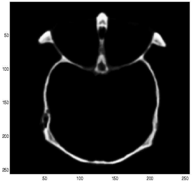


$\widehat{\boldsymbol{z}}$

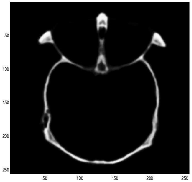
Data fusion in medical imaging

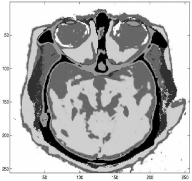
(with O. Féron)

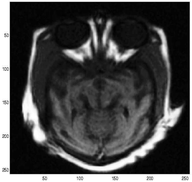
$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2) \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \end{cases}$$

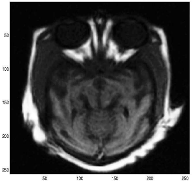
$\boldsymbol{g}_1$ 

$\longrightarrow$

$\widehat{\boldsymbol{f}}_1$ 

$\widehat{\boldsymbol{z}}$ 

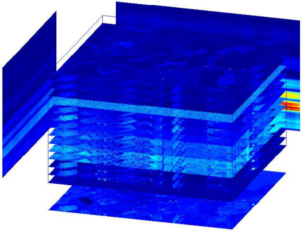
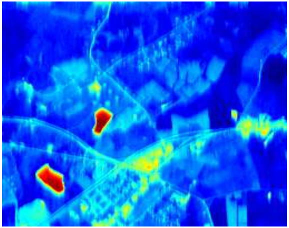
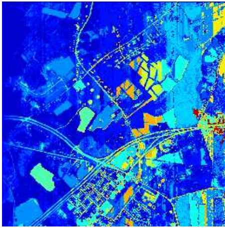
$\boldsymbol{g}_2$ 

$\widehat{\boldsymbol{f}}_2$ 

Joint segmentation of hyper-spectral images

(with N. Bali & A. Mohammadpour)

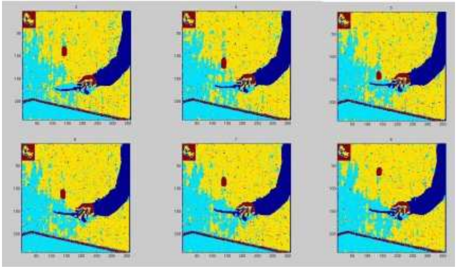
$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma^2_{i\,k}), \quad k = 1, \cdots, K \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \\ m_{ik} \text{ follow a Markovian model along the index } i \end{cases}$$

Segmentation of a video sequence of images

(with P. Brault)

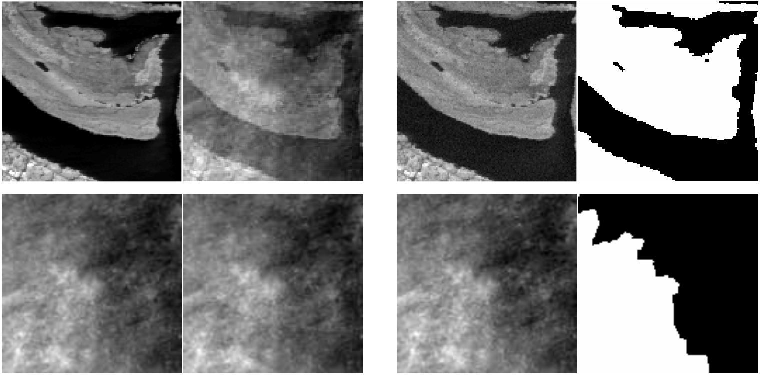
$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z_i(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma_{i\,k}^2), \quad k = 1, \cdots, K \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}_i) \\ z_i(\boldsymbol{r}) \ \text{ follow a Markovian model along the index } \ i \end{cases}$$



Source separation

(with H. Snoussi & M. Ichir)

$$
\begin{cases}
g_i(\boldsymbol{r}) = \sum_{j=1}^{N} A_{ij} f_j(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\
p(f_j(\boldsymbol{r}) | z_j(\boldsymbol{r}) = k) = \mathcal{N}(m_{j\,k}, \sigma^2_{j\,k}) \\
p(A_{ij}) = \mathcal{N}(A_{0ij}, \sigma^2_{0ij})
\end{cases}
$$



$\boldsymbol{f}$ $\boldsymbol{g}$ $\widehat{\boldsymbol{f}}$ $\widehat{\boldsymbol{z}}$

Some references

- A. Mohammad-Djafari (Ed.) Problèmes inverses en imagerie et en vision (Vol. 1 et 2), Hermes-Lavoisier, Traité Signal et Image, IC2, 2009,
- A. Mohammad-Djafari (Ed.) Inverse Problems in Vision and 3D Tomography, ISTE, Wiley and sons, ISBN: 9781848211728, December 2009, Hardback, 480 pp.
- H. Ayasso and Ali Mohammad-Djafari Joint NDT Image Restoration and Segmentation using Gauss-Markov-Potts Prior Models and Variational Bayesian Computation, To appear in IEEE Trans. on Image Processing, TIP-04815-2009.R2, 2010.
- H. Ayasso, B. Duchene and A. Mohammad-Djafari, Bayesian Inversion for Optical Diffraction Tomography Journal of Modern Optics, 2008.
- A. Mohammad-Djafari, Gauss-Markov-Potts Priors for Images in Computer Tomography Resulting to Joint Optimal Reconstruction and segmentation, International Journal of Tomography & Statistics 11: W09. 76-92, 2008.
- A Mohammad-Djafari, Super-Resolution : A short review, a new method based on hidden Markov modeling of HR image and future challenges, The Computer Journal doi:10,1093/comjnl/bxn005:, 2008.
- O. Féron, B. Duchène and A. Mohammad-Djafari, Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data, Inverse Problems, 21(6):95-115, Dec 2005.
- M. Ichir and A. Mohammad-Djafari, Hidden markov models for blind source separation, IEEE Trans. on Signal Processing, 15(7):1887-1899, Jul 2006.
- F. Humblot and A. Mohammad-Djafari, Super-Resolution using Hidden Markov Model and Bayesian Detection Estimation Framework, EURASIP Journal on Applied Signal Processing, Special number on Super-Resolution Imaging: Analysis, Algorithms, and Applications:ID 36971, 16 pages, 2006.
- O. Féron and A. Mohammad-Djafari, Image fusion and joint segmentation using an MCMC algorithm, Journal of Electronic Imaging, 14(2):paper no. 023014, Apr 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.
- A. Mohammad-Djafari, J.F. Giovannelli, G. Demoment and J. Idier, Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems, Int. Journal of Mass Spectrometry, 215(1-3):175-193, April 2002.

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 Positon Tomography, Non Parametric Bayesian)

My colleages in GPI (L2S) & collaborators in other instituts:

- 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) (Positon Emission Tomography, Non Parametric Bayesian)
- C. Comtat (SHFJ, CEA)(PET, Spatio-Temporal Brain activity)

Questions and Discussions

Conclusions and Perspectives

- Bayesian approach is powerfull
- 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

- A. Doucet and P. Duvaut, “Bayesian estimation of state-space models applied to deconvolution of Bernoulli-Gaussian processes,” *Signal Processing*, vol. 57, no. 2, 1997.
- T. Park and G. Casella., “The Bayesian Lasso,” *Journal of the American Statistical Association*, 2008.
- M. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, 2001.
- C. Févotte and S. Godsill, “A Bayesian aproach for blind separation of sparse source,” *IEEE Transactions on Audio, Speech, and Language processing*, 2006.
- J. Griffin and P. Brown, “Inference with normal-gamma prior distributions in regression problems,” *Bayesian Analysis*, 2010.
- N. Polson and J. Scott., “Shrink globally, act locally: sparse Bayesian regularization and prediction,” *Bayesian Statistics 9*, 2010.
- H. Snoussi and J. Idier., “Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures,” *IEEE Trans. on Signal Processing*, 2006.
- P. Williams, “Bayesian regularization and pruning using a Laplace prior,” *Neural Computation*, 1995.
- T. Mitchell and J. Beauchamp, “Bayesian variable selection in linear regression,” *Journal of the American Statistical Association*, 1988.
- J. R. H. Ishwaran, “Spike and slab variable selection: Frequentist and bayesian strategies,” *Annals of Statistics*, 2005.
- J. J. Kormylo and J. M. Mendel, “Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes,” vol. 28, pp. 482–488, 1982.
- M. Lavielle, “Bayesian deconvolution of Bernoulli-Gaussian processes,” *Signal Processing*, vol. 33, pp. 67–79, 1993.