Bayesian inference with hierarchical prior models for inverse problems in imaging systems

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

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

Content

1. What is probabilty law ? Basics of probability theory
2. How to assign a probability law to a quantity?
   Maximum Entropy (ME), Maximum Likelihood (ML),
   Parametric and Non Parametric Bayesian.
3. Inverse problems:
   Examples and Deterministic regularization methods
4. Bayesian approach for inverse problems
5. Prior modeling
   - Gaussian, Generalized Gaussian (GG), Gamma, Beta,
   - Gauss-Markov, GG-Marvov
   - Sparsity enforcing priors (Bernouilli-Gaussian, B-Gamma,
   Cauchy, Student-t, Laplace)
6. Full Bayesian approach (Estimation of hyperparameters)
7. Hierarchical prior models
8. Bayesian Computation and Algorithms for Hierarchical models
9. Gauss-Markov-Potts family of priors
10. Applications and case studies

1. What we mean by a probability law?
Basics of probability theory

- We may re-consider the classical definitions of *random variable* and *probability*.
- *Hazard* does not exists.
- When we say that a quantity is *random*, it means that we do not have enough information about it.
- A probability measures a degree of *rational* belief in the truth of a proposition (Bernoulli 1713 and Laplace 1812)
- A *probability law* is not inherent to physics or real world.
- We *assign* a probability law to a quantity to translate what we know about it.
- A probability law is a *mathematical model*.
- A *probability law* is always conditional to what we know.

Direct and undirect observation?

- *Direct observation* of a few quantities are possible:
  length, time, electrical charge, number of particles
- For many others, we only can measure them by transforming them. Example:
  Thermometer transforms variation of temepratture to variation of length.
- When measuring (observing) a quantity, the *errors* are always present.
- Even for direct observation of a quantity we may define a probability law

Discrete and continuous variables

- A quantity can be discrete or continuous
- For discrete value quantities we define a probability distribution

$$P(X = k) = \pi_k, \; k = 1, \cdots, K \quad \text{with} \sum_{k=1}^{K} \pi_k = 1$$

- For continuous value quantities we define a probability density.

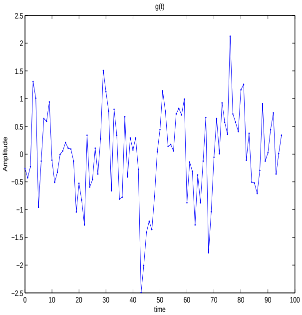
$$P(a < X \leq b) = \int_a^b p(x) \, \mathrm{d}x \quad \text{with} \int_{-\infty}^{+\infty} p(x) \, \mathrm{d}x = 1$$

- For both cases, we may define:
  - Most probable
  - Expected value
  - Variance
  - Higher order moments
  - Entropy

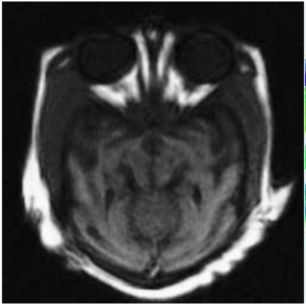
Representation of signals and images

- Signal: $f(t), f(x), f(\nu)$
  - $f(t)$ Variation of temperature in a given position as a function of time t
  - $f(x)$ Variation of temperature as a function of the position x on a line
  - $f(\nu)$ Variation of temperature as a function of the frequency ν
- Image: $f(x,y), f(x,t), f(\nu,t), f(\nu_1,\nu_2)$
  - $f(x,y)$ Distribution of temperature as a function of the position (x,y)
  - $f(x,t)$ Variation of temperature as a function of x and t
  - ...
- 3D, 3D+t, 3D+ν, ... signals: $f(x,y,z), f(x,y,t), f(x,y,z,t)$
  - $f(x,y,z)$ Distribution of temperature as a function of the position (x,y,z)
  - $f(x,y,z,t)$ Variation of temperature as a function of (x,y,z) and t
  - ...

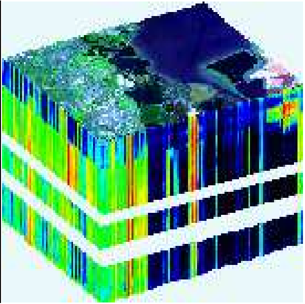
Representation of signals



1D signal



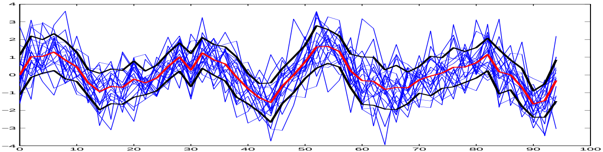
2D signal=image



3D signal

Signals and images

- A signal $f(t)$ can be represented by $p(f(t), t = 0, \cdots, T-1)$



- An image $f(x,y)$ can be represented by $p(f(x,y), (x,y) \in \mathcal{R})$
- Finite domaine observations $\boldsymbol{f} = \{f(t), t = 0, \cdots, T-1\}$
- Image $\boldsymbol{F} = \{f(x,y)\}$ a 2D table or a 1D table $\boldsymbol{f} = \{f(x,y), (x,y) \in \mathcal{R}\}$
- For a vector $\boldsymbol{f}$ we define $p(\boldsymbol{f})$. Then, we can define
  - Most probable value: $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f})\}$
  - Expected value : $\boldsymbol{m} = \mathrm{E}\{\boldsymbol{f}\} = \int \boldsymbol{f} p(\boldsymbol{f}) \, \mathrm{d}\boldsymbol{f}$
  - CoVariance matrix: $\boldsymbol{\Sigma} = \mathrm{E}\{(\boldsymbol{f}-\boldsymbol{m})(\boldsymbol{f}-\boldsymbol{m})'\}$
  - Entropy $H = \mathrm{E}\{-\ln p(\boldsymbol{f})\} = -\int p(\boldsymbol{f}) \ln p(\boldsymbol{f}) \, \mathrm{d}\boldsymbol{f}$

2. How to assign a probability law to a quantity?

- A scalar quantity f is directly observed N times: $\boldsymbol{f} = \{f_1, \cdots, f_N\}$. We want to assign a probability law $p(f)$ to it to be able to compute its most probable value, its mean, its variance, its entropy, ...

- This is an ill-posed problem: Many possible solutions
- Needs prior knowledge

- Main Mathematical methods:
  - Maximum Entropy
  - Maximum Likelihood approach
  - Parametric Bayesian approach
  - Non Parametric Bayesian approach

Maximum Entropy

- First select a finite set of $\phi_k(.)$. For example arithmetic moments $\phi_k(x) = x^k$ or harmonic moments $\phi_k(x) = e^{j\omega_k x}$ and then compute:

$$\mathsf{E}\left\{\phi_k(f)\right\} = \frac{1}{N}\sum_{j=1}^{N} \phi_k(f_j) = d_k, \quad k = 1, \cdots, K$$

- Next, find $p(f)$ which has its entropy

$$H = -\int p(f) \ln p(f) \, \mathrm{d}f$$

  maximum subject to the constraints

$$\mathsf{E}\left\{\phi_k(f)\right\} = \int \phi_k(f)\, p(f) \, \mathrm{d}f = d_k, \quad k = 1, \cdots, K.$$

- Lagrangian technic

Maximum Entropy

- Solution:

$$p(f) = \frac{1}{Z} \exp\left[\sum_{k=1}^{K} \lambda_k \phi_k(f)\right] = \exp\left[\sum_{k=0}^{K} \lambda_k \phi_k(f)\right]$$

with $\phi_0 = 1$ and $\lambda_0 = -\ln Z$ and where

$$Z = \int \exp\left[\sum_{k=1}^{K} \lambda_k \phi_k(f)\right] \, \mathrm{d}f$$

and where $\lambda_k, k = 1, \cdots, K$ are obtained from the K constraints and Z from the normailty $\int p(f) \, \mathrm{d}f = 1$.

Maximum Likelihood

- First select a parametric family $p(f_j|\boldsymbol{\theta})$ (Prior knowledge)
- Then, assuming that the data are observed independently from each other, the likelihood is defined

$$p(\boldsymbol{f}|\boldsymbol{\theta}) = \prod_{j=1}^{N} p(f_j|\boldsymbol{\theta})$$

- Maximum Likelihood estimte of $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{f}|\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \left\{ -\sum_{j=1}^{N} \ln p(f_j|\boldsymbol{\theta}) \right\}$$

- For generalized exponential families, there is a direct link between ME and ML methods.

Parametric Bayesian

- Select a parametric family $p(f_j|\boldsymbol{\theta})$ (Prior knowledge)
- Define the likelihood: $p(\boldsymbol{f}|\boldsymbol{\theta}) = \prod_{j=1}^{N} p(f_j|\boldsymbol{\theta})$
- Assign a prior probabilty law $p(\boldsymbol{\theta})$ to $\boldsymbol{\theta}$
  (Jeffrey's priors, Conjugate priors, Reference priors, Invariance principles, Fischer Information, ...)
- Use the Bayes rule:

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

- Estimate $\boldsymbol{\theta}$, for example:

Maximum A Posteriori (MAP) : $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{f})\}$
Posterior Mean (PM) : $\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\boldsymbol{f})\, \mathrm{d}\boldsymbol{\theta}$

- Use it $p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}})$

Non Parametric Bayesian

- How to define a probability law to a probabilty law ?
- Infinite dimentional: Dirichlet Process, Pitman-Yor Process. ...
- Pitman-Yor Infinite Mixture of Gaussians:

$$p(f_j|\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \alpha_k \mathcal{N}(f_j|\mu_k, v_k)$$

- In practice, the number of components K^* is obtained from the data

$$p(f_j|\boldsymbol{\theta}) = \sum_{k=1}^{K^*} \alpha_k \mathcal{N}(f_j|\mu_k, v_k) \text{ with } \sum_{k=1}^{K^*} \alpha_k = 1$$

- Needs priors on α_k, μ_k, v_k:

$$\begin{array}{l} p(\boldsymbol{\alpha}) = \mathcal{D}(\boldsymbol{\alpha}|\alpha_0) \\ p(\mu_k|v_k) = \mathcal{N}(\mu_k|m_0, v_k/\rho_0) \\ p(v_k) = \mathcal{IG}(v_k|\alpha_0, \beta_0) \end{array}$$

3. Inverse problems : 3 main examples

- Example 1:
  Measuring variation of temperature with a therometer
  - $f(t)$ variation of temperature over time
  - $g(t)$ variation of length of the liquid in thermometer
- Example 2: Seeing outside of a body: Making an image using a camera, a microscope or a telescope
  - $f(x, y)$ real scene
  - $g(x, y)$ observed image
- Example 3: Seeing inside of a body: Computed Tomography usng X rays, US, Microwave, etc.
  - $f(x, y)$ a section of a real 3D body $f(x, y, z)$
  - $g_\phi(r)$ a line of observed radiographe $g_\phi(r, z)$

- Example 1: Deconvolution
- Example 2: Image restoration
- Example 3: Image reconstruction

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: Convolution

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

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

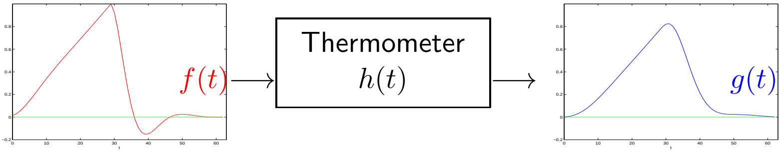
- Inverse problem: Deconvolution

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

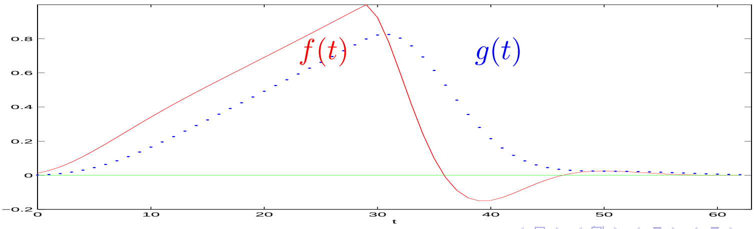
Measuring variation of temperature with a therometer

Forward model: Convolution

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

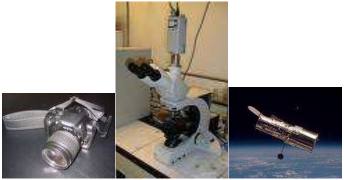


Inversion: Deconvolution



Seeing outside of a body: Making an image with a camera, a microscope or a telescope

- $f(x, y)$ real scene
- $g(x, y)$ observed image



- Forward model: Convolution

$$g(x, y) = \iint f(x', y')\, h(x - x', y - y')\, \mathrm{d}x'\, \mathrm{d}y' + \epsilon(x, y)$$

$h(x, y)$: Point Spread Function (PSF) of the imaging system

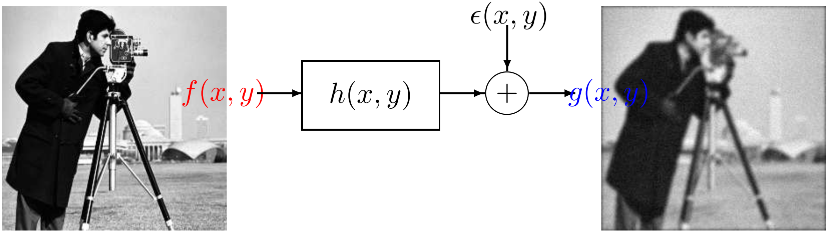
- Inverse problem: Image restoration

Given the forward model $\mathcal{H}$ (PSF $h(x, y)$))
and a set of data $g(x_i, y_i), i = 1, \cdots, M$
find $f(x, y)$

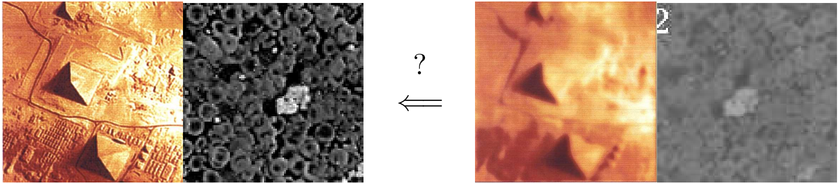
Making an image with an unfocused camera

Forward model: 2D Convolution

$$g(x,y) = \iint f(x',y')\, h(x-x',y-y')\, \mathrm{d}x'\, \mathrm{d}y' + \epsilon(x,y)$$

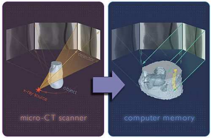


Inversion: Image Deconvolution or Restoration



Seeing inside of a body: Computed Tomography

- $f(x,y)$ a section of a real 3D body $f(x,y,z)$
- $g_\phi(r)$ a line of observed radiographe $g_\phi(r,z)$



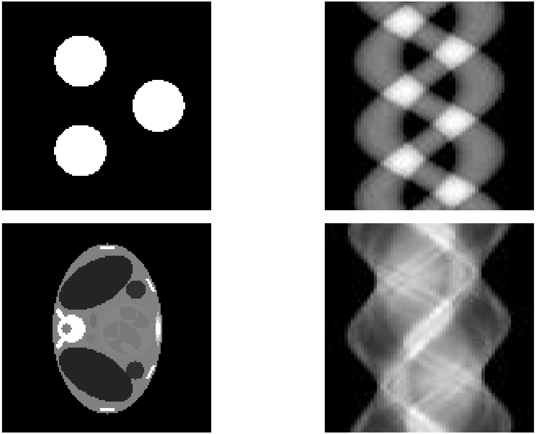
- Forward model:
  Line integrals or Radon Transform

$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x,y) \, \mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x,y) \, \delta(r - x\cos\phi - y\sin\phi) \, \mathrm{d}x \, \mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem: Image reconstruction

  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x,y)$

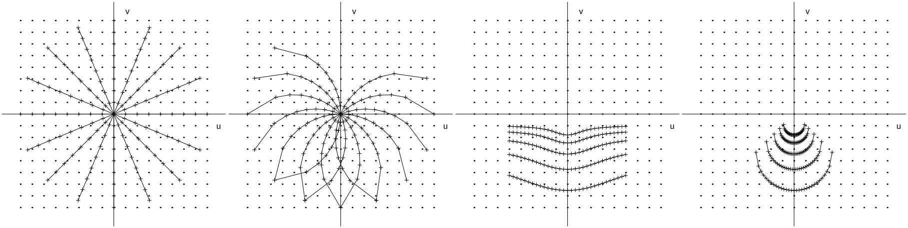
Computed Tomography: Radon Transform



Forward:	$f(x,y)$	$\longrightarrow$	$g(r,\phi)$
Inverse:	$f(x,y)$	$\longleftarrow$	$g(r,\phi)$

Fourier Synthesis in different imaging systems

$$G(\omega_x, \omega_y) = \iint f(x,y) \exp\left[-j\left(\omega_x x + \omega_y y\right)\right] \, \mathrm{d}x \, \mathrm{d}y$$



X ray Tomography   Diffraction   Eddy current   SAR & Radar

Forward problem: Given $f(x,y)$ compute $G(\omega_x, \omega_y)$
Inverse problem : Given $G(\omega_x, \omega_y)$ on those algebraic lines, cercles or curves, estimate $f(x,y)$

General formulation of inverse problems

- General non linear inverse problems:

$$g(\boldsymbol{s}) = [\mathcal{H} f(\boldsymbol{r})](\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \quad \boldsymbol{r} \in \mathcal{R}, \quad \boldsymbol{s} \in \mathcal{S}$$

- Linear models:

$$g(\boldsymbol{s}) = \int f(\boldsymbol{r})\, h(\boldsymbol{r}, \boldsymbol{s})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s})$$

  If $h(\boldsymbol{r}, \boldsymbol{s}) = h(\boldsymbol{r} - \boldsymbol{s}) \longrightarrow$ Convolution.

- Discrete data:

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, m$$

- Inversion: Given the forward model $\mathcal{H}$ and the data $\boldsymbol{g} = \{g(\boldsymbol{s}_i), i = 1, \cdots, m)\}$ estimate $f(\boldsymbol{r})$
- Well-posed and Ill-posed problems (Hadamard): existance, uniqueness and stability
- Need for prior information

Inverse problems: Discretization

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, M$$

- $f(\boldsymbol{r})$ is assumed to be well approximated by

$$f(\boldsymbol{r}) \simeq \sum_{j=1}^{N} f_j\, b_j(\boldsymbol{r})$$

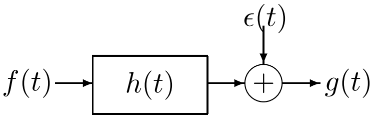
with $\{b_j(\boldsymbol{r})\}$ a basis or any other set of known functions

$$g(\boldsymbol{s}_i) = g_i \simeq \sum_{j=1}^{N} f_j \int h(\boldsymbol{s}_i, \boldsymbol{r})\, b_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}, \quad i = 1, \cdots, M$$

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \;\text{ with }\; H_{ij} = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, b_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$$

- $\boldsymbol{H}$ is huge dimensional
- LS solution : $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{Q(\boldsymbol{f})\}$ with
  $Q(\boldsymbol{f}) = \sum_i |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^2 = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$
  does not give satisfactory result.

Convolution: Discretization



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

- The signals $f(t)$, $g(t)$, $h(t)$ are discretized with the same sampling period $\Delta T = 1$,
- The impulse response is finite (FIR) : $h(t) = 0$, for t such that $t < -q\Delta T$ or $\forall t > p\Delta T$.

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

Convolution: Discretized matrix vector forms

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

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

- $\boldsymbol{g}$ is a $(M+1)$-dimensional vector,
- $\boldsymbol{f}$ has dimension $M+p+q+1$,
- $\boldsymbol{h} = [h(p), \cdots, h(0), \cdots, h(-q)]$ has dimension $(p+q+1)$
- $\boldsymbol{H}$ has dimensions $(M+1) \times (M+p+q+1)$.

Convolution: Discretized matrix vector form

- If system is causal ($q = 0$) we obtain

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

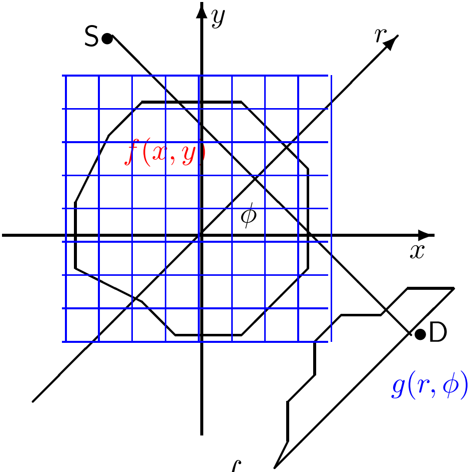
- $\boldsymbol{g}$ is a $(M+1)$-dimensional vector,
- $\boldsymbol{f}$ has dimension $M+p+1$,
- $\boldsymbol{h} = [h(p), \cdots, h(0)]$ has dimension $(p+1)$
- $\boldsymbol{H}$ has dimensions $(M+1) \times (M+p+1)$.

Convolution: Causal systems and causal input

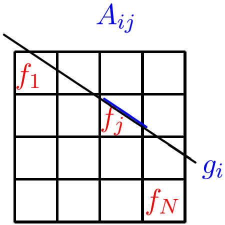
$$
\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix}
=
\begin{bmatrix}
h(0) & & & & & \\
h(1) & \ddots & & & & \\
\vdots & & & & & \\
h(p) & \cdots & & h(0) & & \\
0 & \ddots & & & \ddots & \\
\vdots & & & & & \\
0 & \cdots & 0 & h(p) & \cdots & h(0)
\end{bmatrix}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

- $\boldsymbol{g}$ is a $(M+1)$-dimensional vector,
- $\boldsymbol{f}$ has dimension $M+1$,
- $\boldsymbol{h} = [h(p), \cdots, h(0)]$ has dimension $(p+1)$
- $\boldsymbol{H}$ has dimensions $(M+1) \times (M+1)$.

Discretization of Radon Transfrom in CT



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



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

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

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

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

Inverse problems: Deterministic methods

Data matching

- Observation model

  $g_i = h_i(\boldsymbol{f}) + \epsilon_i, \quad i = 1, \ldots, M \longrightarrow \boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$
- Misatch between data and output of the model $\Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f}))$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{\Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f}))\}$$

- Examples:

– LS $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^2 = \sum_i |g_i - h_i(\boldsymbol{f})|^2$

– L_p $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^p = \sum_i |g_i - h_i(\boldsymbol{f})|^p, \quad 1 < p < 2$

– KL $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \sum_i g_i \ln \frac{g_i}{h_i(\boldsymbol{f})}$

- In general, does not give satisfactory results for inverse problems.

Inverse problems: Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

Functional space (Tikhonov):

$$g = \mathcal{H}(f) + \epsilon \longrightarrow J(f) = ||g - \mathcal{H}(f)||_2^2 + \lambda||\mathcal{D}f||_2^2$$

Finite dimensional space (Philips & Towmey): $\boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$

- Minimum norme LS (MNLS): $J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{f}||^2$
- Classical regularization: $J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{D}\boldsymbol{f}||^2$
- More general regularization:

$$J(\boldsymbol{f}) = \mathcal{Q}(\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})) + \lambda\Omega(\boldsymbol{D}\boldsymbol{f})$$

or

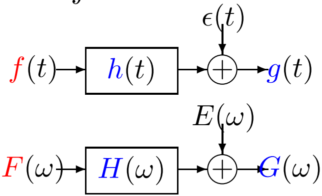
$$J(\boldsymbol{f}) = \Delta_1(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) + \lambda\Delta_2(\boldsymbol{f}, \boldsymbol{f}_\infty)$$

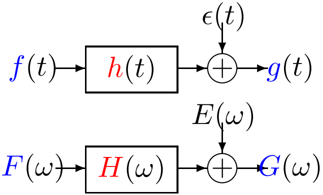
Limitations:

- Errors are considered implicitly white and Gaussian
- Limited prior information on the solution
- Lack of tools for the determination of the hyperparameters

Convolution, Identification, Deconvolution and Blind deconvolution problems

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





$$G(\omega) = H(\omega)\,F(\omega) + E(\omega) \qquad G(\omega) = H(\omega)\,F(\omega) + E(\omega)$$

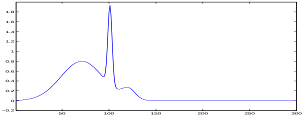
$$F(\omega) = \frac{G(\omega)}{H(\omega)} + \frac{E(\omega)}{H(\omega)} \qquad H(\omega) = \frac{G(\omega)}{F(\omega)} + \frac{E(\omega)}{F(\omega)}$$

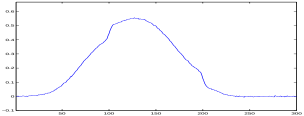
- Convolution: Given h and f compute g
- Identification: Given f and g estimate h
- Simple Deconvolution: Given h and g estimate f
- Blind Deconvolution: Given g estimate h and f

Identification and Deconvolution: discretized formulation

Deconvolution	Identification
$\boldsymbol{g} = \boldsymbol{H}\,\boldsymbol{f} + \boldsymbol{\epsilon}$	$\boldsymbol{g} = \boldsymbol{F}\,\boldsymbol{h} + \boldsymbol{\epsilon}$
$J(\boldsymbol{f}) = \Vert\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\Vert^2 + \lambda_f \Vert\boldsymbol{D}_f\boldsymbol{f}\Vert^2$	$J(\boldsymbol{h}) = \Vert\boldsymbol{g} - \boldsymbol{F}\boldsymbol{h}\Vert^2 + \lambda_h \Vert\boldsymbol{D}_h\boldsymbol{h}\Vert^2$
$\nabla J = -2\boldsymbol{H}'(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + 2\lambda_f \boldsymbol{D}_f'\boldsymbol{D}_f\boldsymbol{f}$	$\nabla J = -2\boldsymbol{F}'(\boldsymbol{g} - \boldsymbol{F}\boldsymbol{h}) + 2\lambda_h \boldsymbol{D}_h'\boldsymbol{D}_h\boldsymbol{h}$
$\widehat{\boldsymbol{f}} = [\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{D}_f'\boldsymbol{D}_f]^{-1}\boldsymbol{H}'\boldsymbol{g}$	$\widehat{\boldsymbol{h}} = [\boldsymbol{F}'\boldsymbol{F} + \lambda_h\boldsymbol{D}_h'\boldsymbol{D}_h]^{-1}\boldsymbol{F}'\boldsymbol{g}$
Circulante approximation: $\widehat{f}(\omega) = \frac{H^*(\omega)}{\vert H(\omega)\vert^2 + \lambda_f \vert D_f(\omega)\vert^2}\, g(\omega)$	$\widehat{f}(\omega) = \frac{\vert F^*(\omega)}{\vert F(\omega)\vert^2 + \lambda_h \vert D_h(\omega)\vert^2}\, g(\omega)$
Wiener Filtering: $\widehat{f}(\omega) = \frac{H^*(\omega)}{\vert H(\omega)\vert^2 + \frac{S_{\epsilon\epsilon}(\omega)}{S_{ff}(\omega)}}\, g(\omega)$	$\widehat{f}(\omega) = \frac{F^*(\omega)}{\vert F(\omega)\vert^2 + \frac{S_{\epsilon\epsilon}(\omega)}{S_{hh}(\omega)}}\, g(\omega)$

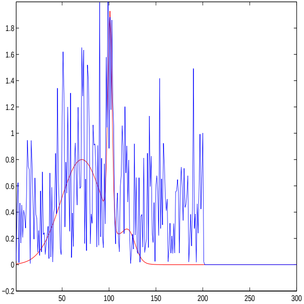
Deconvolution example



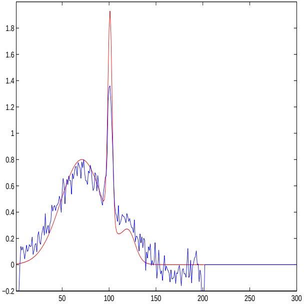


Forward: $f(t) \longrightarrow g(t) = h(t) * f(t) + \epsilon(t)$

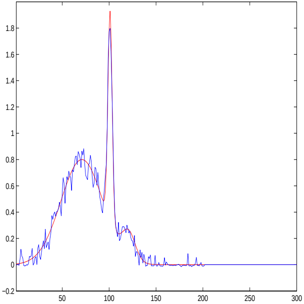
Inverse: $\widehat{f}(t) \longleftarrow g(t)$



Inverse Filtering



Wiener



Regularization

Inversion: Probabilistic methods

Taking account of errors and uncertainties $\longrightarrow$ Probability theory

- Maximum Likelihood (ML)
- Minimum Inaccuracy (MI)
- Probability Distribution Matching (PDM)
- Maximum Entropy (ME) and Information Theory (IT)
- Bayesian Inference (Bayes)

Advantages:

- Explicit account of the errors and noise
- A large class of priors via explicit or implicit modeling
- A coherent approach to combine information content of the data and priors

Limitations:

- Practical implementation and cost of calculation

4. Bayesian inference for inverse problems

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

- Observation model $\mathcal{M}$ + Hypothesis on the noise $\boldsymbol{\epsilon} \longrightarrow$
  $$p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M}) = p_{\boldsymbol{\epsilon}}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})$$
- A priori information $\quad p(\boldsymbol{f}|\mathcal{M})$
- Bayes : $$p(\boldsymbol{f}|\boldsymbol{g};\mathcal{M}) = \frac{p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M})\, p(\boldsymbol{f}|\mathcal{M})}{p(\boldsymbol{g}|\mathcal{M})}$$

Link with regularization :

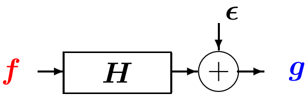
Maximum A Posteriori (MAP) :

$$\begin{aligned} \widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{g}|\boldsymbol{f})\; p(\boldsymbol{f})\} \\ &= \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f}) - \ln p(\boldsymbol{f})\} \end{aligned}$$

with $\quad Q(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) = -\ln p(\boldsymbol{g}|\boldsymbol{f}) \quad$ and $\quad \lambda\Omega(\boldsymbol{f}) = -\ln p(\boldsymbol{f})$

Bayesian inference for inverse problems

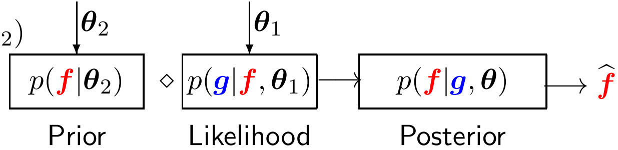
- Linear Inverse problems: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$



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

Case of linear models and Gaussian priors

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

- Hypothesis on the noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I})$

$$p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right]$$

- Hypothesis on $\boldsymbol{f}$: $\boldsymbol{f} \sim \mathcal{N}(0, \sigma_f^2 \boldsymbol{I})$

$$p(\boldsymbol{f}) \propto \exp\left[-\frac{1}{2\sigma_f^2}\|\boldsymbol{f}\|^2\right]$$

- A posteriori:

$$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \frac{\sigma_\epsilon^2}{\sigma_f^2}\|\boldsymbol{f}\|^2\right]\right]$$

- MAP : $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$

  with $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2, \qquad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$

- Advantage : characterization of the solution

$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{P}})$ with $\widehat{\boldsymbol{f}} = \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g}$, $\widehat{\boldsymbol{P}} = \sigma_\epsilon^2\left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{I}\right)^{-1}$

MAP estimation with other priors:

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

Separable priors:

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

Markovian models:

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

Main advantages of the Bayesian approach

- MAP $=$ Regularization
- Posterior mean ? Marginal MAP ?
- More information in the posterior law than only its mode or its mean
- Meaning and tools for estimating hyper parameters
- Meaning and tools for model selection
- More specific and specialized priors, particularly through the hidden variables
- More computational tools:
  - Expectation-Maximization for computing the maximum likelihood parameters
  - MCMC for posterior exploration
  - Variational Bayes for analytical computation of the posterior marginals
  - ...

MAP estimation and Compressed Sensing

$$\begin{cases} \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \\ \boldsymbol{f} = \boldsymbol{W}\boldsymbol{z} \end{cases}$$

- $\boldsymbol{W}$ a code book matrix, $\boldsymbol{z}$ coefficients
- Gaussian:

$$p(\boldsymbol{z}) = \mathcal{N}(0, \sigma_z^2 \boldsymbol{I}) \propto \exp\left[-\frac{1}{2\sigma_z^2}\sum_j |z_j|^2\right]$$
$$J(\boldsymbol{z}) = -\ln p(\boldsymbol{z}|\boldsymbol{g}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{W}\boldsymbol{z}\|^2 + \lambda \sum_j |z_j|^2$$

- Generalized Gaussian (sparsity, $\beta = 1$):

$$p(\boldsymbol{z}) \propto \exp\left[-\lambda \sum_j |z_j|^\beta\right]$$
$$J(\boldsymbol{z}) = -\ln p(\boldsymbol{z}|\boldsymbol{g}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{W}\boldsymbol{z}\|^2 + \lambda \sum_j |z_j|^\beta$$

- $\boldsymbol{z} = \arg\min_{\boldsymbol{z}} \{J(\boldsymbol{z})\} \longrightarrow \widehat{\boldsymbol{f}} = \boldsymbol{W}\widehat{\boldsymbol{z}}$

Bayesian Estimation: Two simple priors

- Example 1: Linear Gaussian case:

$$\left\{\begin{array}{l} p(\boldsymbol{g}|\boldsymbol{f},\theta_1)=\mathcal{N}(\boldsymbol{H}\boldsymbol{f},\theta_1\boldsymbol{I}) \\ p(\boldsymbol{f}|\theta_2)=\mathcal{N}(0,\theta_2\boldsymbol{I}) \end{array}\right. \longrightarrow p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})=\mathcal{N}(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{P}})$$

with

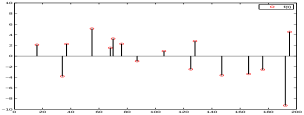
$$\left\{\begin{array}{l} \widehat{\boldsymbol{f}}=(\boldsymbol{H}'\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \widehat{\boldsymbol{P}}=\theta_1(\boldsymbol{H}'\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}, \quad \lambda=\frac{\theta_1}{\theta_2} \end{array}\right.$$

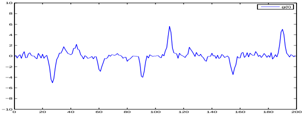
$$\widehat{\boldsymbol{f}}=\arg\min_{\boldsymbol{f}}\{J(\boldsymbol{f})\} \text{ with } J(\boldsymbol{f})=\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2+\lambda\|\boldsymbol{f}\|_2^2$$

- Example 2: Double Exponential prior & MAP:

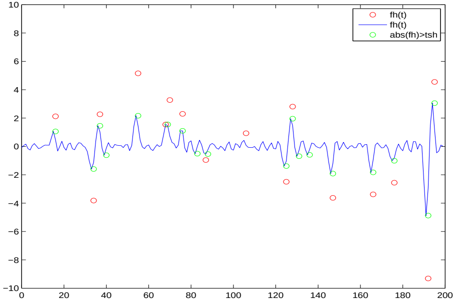
$$\widehat{\boldsymbol{f}}=\arg\min_{\boldsymbol{f}}\{J(\boldsymbol{f})\} \text{ with } J(\boldsymbol{f})=\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2+\lambda\|\boldsymbol{f}\|_1$$

Deconvolution example

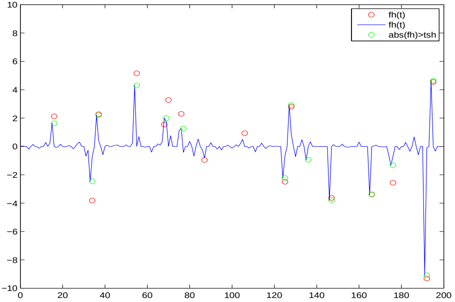




Forward:	$f(t)$	$\longrightarrow$	$g(t) = h(t) * f(t) + \epsilon(t)$
Inverse:	$\widehat{f}(t)$	$\longleftarrow$	$g(t)$



Quadratic Reg. (Gaussian)

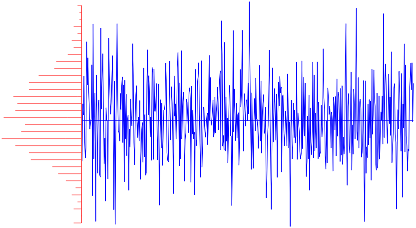


L_1 Reg. (Laplace)

Two main steps in the Bayesian approach

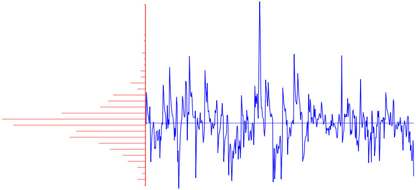
- Prior modeling
  - Separable:
    Gaussian, Gamma,
    Sparsity enforcing: Generalized Gaussian, mixture of Gaussians, mixture of Gammas, ...
  - Markovian:
    Gauss-Markov, GGM, ...
  - Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects
  - MAP, Posterior mean, Marginal MAP
  - MAP needs optimization algorithms
  - Posterior mean needs integration methods
  - Marginal MAP and Hyperparameter estimation need integration and optimization
  - Approximations:
    - Gaussian approximation (Laplace)
    - Numerical exploration MCMC
    - Variational Bayes (Separable approximation)

5. Prior modeling of signals



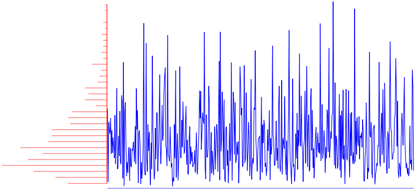
Gaussian

$$p(f_j) \propto \exp\left[-\alpha |f_j|^2\right]$$



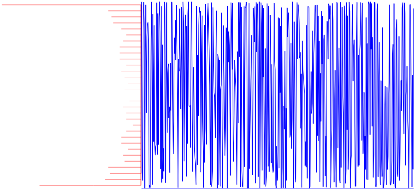
Generalized Gaussian

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



Gamma

$$p(f_j) \propto f_j^\alpha \exp\left[-\beta f_j\right]$$

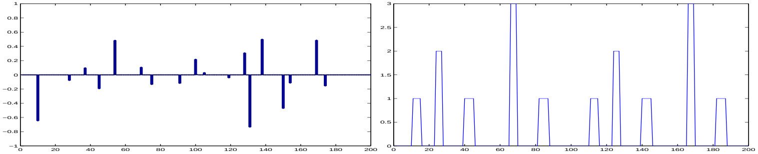


Beta

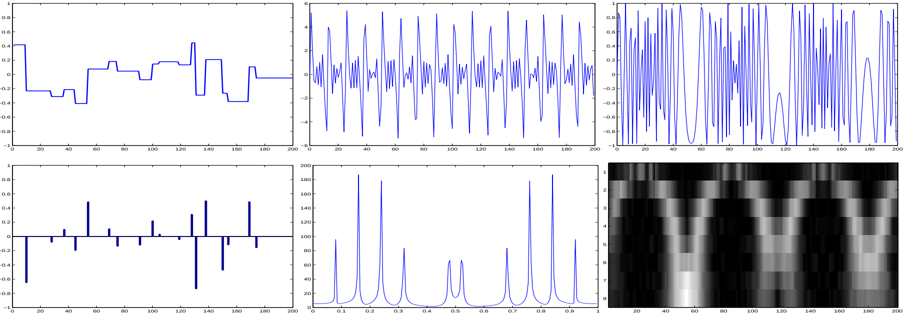
$$p(f_j) \propto f_j^\alpha (1 - f_j)^\beta$$

Sparsity enforcing prior models

- Sparse signals: Direct sparsity



- Sparse signals: Sparsity in a Transform domaine



Sparsity enforcing prior models

- Simple heavy tailed models:
  - Generalized Gaussian, Double Exponential
  - Student-t, Cauchy
  - Elastic net

  - Symmetric Weibull, Symmetric Rayleigh
  - Generalized hyperbolic

- 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.

- 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 $\nu = 1$.

- Elastic net prior model

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

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 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

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

MAP, Joint MAP

- Inverse problems: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$
- Posterior law:

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

- Examples:

  Gaussian noise, Gaussian prior and MAP:

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

  Gaussian noise, Double Exponential prior and MAP:

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

- 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:

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

6. Full Bayesian approach

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

- 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: $\qquad (\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\}$
- Marginalization: $\begin{cases} p(\boldsymbol{f}|\boldsymbol{g}; \mathcal{M}) & = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{\theta} \\ p(\boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M}) & = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f} \end{cases}$
- Posterior means: $\begin{cases} \widehat{\boldsymbol{f}} & = \int\int \boldsymbol{f}\, p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{f} \\ \widehat{\boldsymbol{\theta}} & = \int\int \boldsymbol{\theta}\, p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta} \end{cases}$
- 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: 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

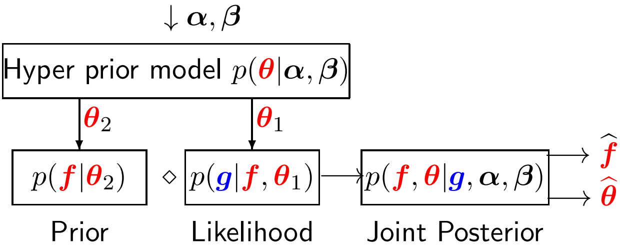
  Posterior Mean: $\widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \,\mathrm{d}\boldsymbol{f}$

- Needs the expression of the Likelihood:

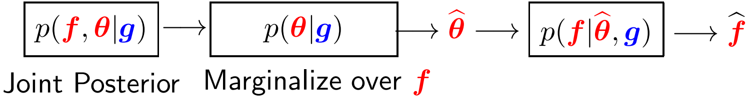
$$p(\boldsymbol{g}|\boldsymbol{\theta}) = \int p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2) \,\mathrm{d}\boldsymbol{f}$$

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

Full Bayesian Model and Hyperparameter Estimation



Full Bayesian Model and Hyperparameter Estimation scheme



Marginalization for Hyperparameter Estimation

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:

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

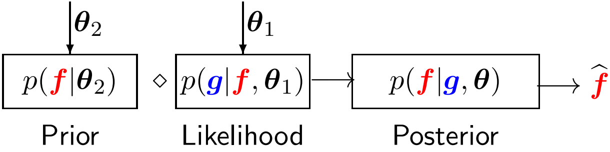
- GEM (Bayesian) algorithm:

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

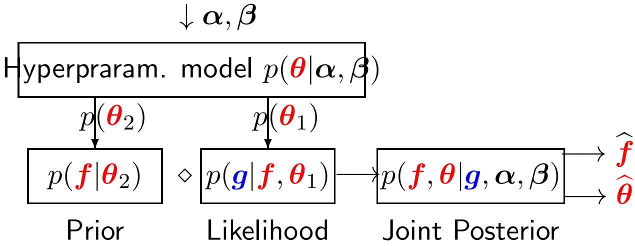
$$\boxed{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \boxed{\text{EM, GEM}} \longrightarrow \widehat{\boldsymbol{\theta}} \longrightarrow \boxed{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})} \longrightarrow \widehat{\boldsymbol{f}}$$

Summary of Bayesian estimation 1

- Simple Bayesian Model and Estimation



- Full Bayesian Model and Hyperparameter Estimation



Summary of Bayesian estimation 2

- Marginalization for Hyperparameter Estimation

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

Joint Posterior — Marginalize over $\boldsymbol{f}$

- Variational Bayesian Approximation

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

Variational Bayesian Approximation

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

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

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

Variational Bayesian Approximation

$$\mathcal{M} : \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \longrightarrow p(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g}|\mathcal{M}) = p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}, \mathcal{M})\, p(\boldsymbol{f}|\boldsymbol{\theta}, \mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M})$$

$$p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}, \mathcal{M}) = p(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g}|, \mathcal{M})/p(\boldsymbol{g}|\mathcal{M})$$

$$\mathsf{KL}(q : p) = \iint q(\boldsymbol{f}, \boldsymbol{\theta}) \ln \frac{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})}{q(\boldsymbol{f}, \boldsymbol{\theta})}\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta}$$

$$\begin{aligned} p(\boldsymbol{g}|\mathcal{M}) &= \iint q(\boldsymbol{f}, \boldsymbol{\theta}) \frac{p(\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}|\mathcal{M})}{q(\boldsymbol{f}, \boldsymbol{\theta})}\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta} \\ &\geq \iint q(\boldsymbol{f}, \boldsymbol{\theta}) \ln \frac{p(\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}|\mathcal{M})}{q(\boldsymbol{f}, \boldsymbol{\theta})}\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

Free energy:

$$\mathcal{F}(q) = \iint q(\boldsymbol{f}, \boldsymbol{\theta}) \ln \frac{p(\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}|\mathcal{M})}{q(\boldsymbol{f}, \boldsymbol{\theta})}\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta}$$

Evidence of the model $\mathcal{M}$:

$$p(\boldsymbol{g}|\mathcal{M}) = \mathcal{F}(q) + \mathsf{KL}(q : p)$$

BVA: Separable Approximation

$$p(\boldsymbol{g}|\mathcal{M}) = \mathcal{F}(q) + \mathsf{KL}(q:p)$$

$$q(\boldsymbol{f},\boldsymbol{\theta}) = q_1(\boldsymbol{f})\; q_2(\boldsymbol{\theta})$$

Minimizing $\mathsf{KL}(q:p)$ = Maximizing $\mathcal{F}(q)$

$$(\widehat{q}_1,\widehat{q}_2) = \arg\min_{(q_1,q_2)} \{\mathsf{KL}(q_1 q_2 : p)\} = \arg\max_{(q_1,q_2)} \{\mathcal{F}(q_1 q_2)\}$$

$\mathsf{KL}(q_1 q_2 : p)$ is convexe wrt q_1 when q_2 is fixed and vise versa:

$$\left\{\begin{array}{ll} \widehat{q}_1 & = \arg\min_{q_1} \{\mathsf{KL}(q_1 \widehat{q}_2 : p)\} = \arg\max_{q_1} \{\mathcal{F}(q_1 \widehat{q}_2)\} \\ \widehat{q}_2 & = \arg\min_{q_2} \{\mathsf{KL}(\widehat{q}_1 q_2 : p)\} = \arg\max_{q_2} \{\mathcal{F}(\widehat{q}_1 q_2)\} \end{array}\right.$$

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

BVA: Choice of family of laws q_1 and q_2

- Case 1 : $\longrightarrow$ Joint MAP

$$\left\{\begin{array}{ll}\widehat{q}_1(\boldsymbol{f}|\widetilde{\boldsymbol{f}}) & = \delta(\boldsymbol{f}-\widetilde{\boldsymbol{f}}) \\ \widehat{q}_2(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}) & = \delta(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}})\end{array}\right. \longrightarrow \left\{\begin{array}{l}\widetilde{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f},\widetilde{\boldsymbol{\theta}}|\boldsymbol{g};\mathcal{M})\right\} \\ \widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{p(\widetilde{\boldsymbol{f}},\boldsymbol{\theta}|\boldsymbol{g};\mathcal{M})\right\}\end{array}\right.$$

- Case 2 : $\longrightarrow$ EM

$$\left\{\begin{array}{ll}\widehat{q}_1(\boldsymbol{f}) & \propto p(\boldsymbol{f}|\boldsymbol{\theta},\boldsymbol{g}) \\ \widehat{q}_2(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}) & = \delta(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}})\end{array}\right. \longrightarrow \left\{\begin{array}{ll}Q(\boldsymbol{\theta},\widetilde{\boldsymbol{\theta}}) = & \left\langle \ln p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g};\mathcal{M})\right\rangle_{q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}})} \\ \widetilde{\boldsymbol{\theta}} & = \arg\max_{\boldsymbol{\theta}} \left\{Q(\boldsymbol{\theta},\widetilde{\boldsymbol{\theta}})\right\}\end{array}\right.$$

- Appropriate choice for inverse problems

$$\left\{\begin{array}{ll}\widehat{q}_1(\boldsymbol{f}) & \propto p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}},\boldsymbol{g};\mathcal{M}) \\ \widehat{q}_2(\boldsymbol{\theta}) & \propto p(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{g};\mathcal{M})\end{array}\right. \longrightarrow \left\{\begin{array}{l}\text{Prise en compte des incertitudes} \\ \text{de } \widehat{\boldsymbol{\theta}} \text{ pour } \widehat{\boldsymbol{f}} \text{ et vise versa.}\end{array}\right.$$

Blind Deconvolution: Bayesian approach

Deconvolution	Identification
$\boldsymbol{g} = \boldsymbol{H}\,\boldsymbol{f} + \boldsymbol{\epsilon}$	$\boldsymbol{g} = \boldsymbol{F}\,\boldsymbol{h} + \boldsymbol{\epsilon}$
$p(\boldsymbol{g}\vert\boldsymbol{f}) = \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, \boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I})$	$p(\boldsymbol{g}\vert\boldsymbol{h}) = \mathcal{N}(\boldsymbol{F}\boldsymbol{h}, \boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I})$
$p(\boldsymbol{f}) = \mathcal{N}(0, \boldsymbol{\Sigma}_f = [\boldsymbol{D}_f' \boldsymbol{D}_f]^{-1})$	$p(\boldsymbol{h}) = \mathcal{N}(0, \boldsymbol{\Sigma}_h = [\boldsymbol{D}_h' \boldsymbol{D}_h]^{-1})$
$p(\boldsymbol{f}\vert\boldsymbol{g}) = \mathcal{N}(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\Sigma}}_f)$	$p(\boldsymbol{h}\vert\boldsymbol{g}) = \mathcal{N}(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{\Sigma}}_h)$
$\widehat{\boldsymbol{\Sigma}}_f = \sigma_\epsilon^2 [\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{D}_f' \boldsymbol{D}_f]^{-1}$	$\widehat{\boldsymbol{\Sigma}}_h = \sigma_\epsilon^2 [\boldsymbol{F}'\boldsymbol{F} + \lambda_h \boldsymbol{D}_h' \boldsymbol{D}_h]^{-1}$
$\widehat{\boldsymbol{f}} = [\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{D}_f' \boldsymbol{D}_f]^{-1} \boldsymbol{H}'\boldsymbol{g}$	$\widehat{\boldsymbol{h}} = [\boldsymbol{F}'\boldsymbol{F} + \lambda_h \boldsymbol{D}_h' \boldsymbol{D}_h]^{-1} \boldsymbol{F}'\boldsymbol{g}$

- Blind Deconvolution: Joint posterior law:

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

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

with $J(\boldsymbol{f}, \boldsymbol{h}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda_f \|\boldsymbol{D}_f \boldsymbol{f}\|^2 + \lambda_h \|\boldsymbol{D}_h \boldsymbol{h}\|^2$

- iterative algorithm

Blind Deconvolution: Variational Bayesian Approximation algorithm

- Joint posterior law:

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

- Approximation: $p(\boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g})$ by $q(\boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{h})$
- Criterion of approximation: Kullback-Leiler

$$\mathsf{KL}(q|p) = \int q \,\ln \frac{q}{p} = \int q_1\, q_2 \,\ln \frac{q_1\, q_2}{p}$$

$$\begin{aligned} \mathsf{KL}(q_1\, q_2|p) &= \int q_1 \,\ln q_1 + \int q_2 \,\ln q_2 - \int q \ln p \\ &= -\mathcal{H}(q_1) - \mathcal{H}(q_2) + \langle - \ln p((\boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g})\rangle_q \end{aligned}$$

- When the expression of q_1 and q_2 are obtained, use them.

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

- Joint MAP:

$$\boldsymbol{h}^{(0)} \longrightarrow \boldsymbol{H} \longrightarrow \boxed{\widehat{\boldsymbol{f}} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{\Sigma}_f)^{-1} \boldsymbol{H}'\boldsymbol{g}} \longrightarrow \widehat{\boldsymbol{f}}$$

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

$$\widehat{\boldsymbol{h}} \longleftarrow \boxed{\widehat{\boldsymbol{h}} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_h \boldsymbol{\Sigma}_h)^{-1} \boldsymbol{F}'\boldsymbol{g}} \longleftarrow \boldsymbol{F}$$

- VBA:

$$\begin{array}{l} \boldsymbol{h}^{(0)} \longrightarrow \boldsymbol{H} \longrightarrow \\ \boldsymbol{\Sigma}_h^{(0)} \longrightarrow \boldsymbol{\Sigma}_h \longrightarrow \end{array} \boxed{\begin{array}{c} \widehat{\boldsymbol{f}} = \left(\boldsymbol{H}'\widehat{\boldsymbol{\Sigma}}_h\boldsymbol{H} + \lambda_f \boldsymbol{\Sigma}_f\right)^{-1} \boldsymbol{H}'\boldsymbol{g} \\ \widehat{\boldsymbol{\Sigma}}_f = \sigma_\epsilon^2 (\boldsymbol{H}'\widehat{\boldsymbol{\Sigma}}_h\boldsymbol{H} + \lambda_f \boldsymbol{\Sigma}_f)^{-1} \end{array}} \begin{array}{l} \longrightarrow \widehat{\boldsymbol{f}} \\ \longrightarrow \widehat{\boldsymbol{\Sigma}}_f \end{array}$$

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

$$\begin{array}{l} \widehat{\boldsymbol{h}} \longleftarrow \\ \widehat{\boldsymbol{\Sigma}}_h \longleftarrow \end{array} \boxed{\begin{array}{c} \widehat{\boldsymbol{h}} = \left(\boldsymbol{F}'\widehat{\boldsymbol{\Sigma}}_f\boldsymbol{F} + \lambda_h \boldsymbol{\Sigma}_h\right)^{-1} \boldsymbol{F}'\boldsymbol{g} \\ \widehat{\boldsymbol{\Sigma}}_h = \sigma_\epsilon^2 (\boldsymbol{F}'\widehat{\boldsymbol{\Sigma}}_f\boldsymbol{F} + \lambda_h \boldsymbol{\Sigma}_h)^{-1} \end{array}} \begin{array}{l} \longleftarrow \boldsymbol{F} \\ \longleftarrow \boldsymbol{\Sigma}_f \end{array}$$

- Link with Message Passing and Belief Propagation methods

7. Hierarchical models and hidden variables

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

$$p(f) = \sum_{k=1}^{K} \alpha_k p_k(f) \longrightarrow \left\{ \begin{array}{l} p(f|z=k) = p_k(f), \\ P(z=k) = \alpha_k, \quad \sum_k \alpha_k = 1 \end{array} \right.$$

- Example 1: MoG model: $p_k(f) = \mathcal{N}(f|m_k, v_k)$
  2 Gaussians: $p_0 = \mathcal{N}(0, v_0), p_1 = \mathcal{N}(0, v_1), \alpha_0 = \lambda, \alpha_1 = 1 - \lambda$

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

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

$$\left\{ \begin{array}{l} p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}\left(f_j|0, v_{z_j}\right) \propto \exp\left[-\frac{1}{2}\sum_j \frac{f_j^2}{v_{z_j}}\right] \\ p(\boldsymbol{z}) = \lambda^{n_1}(1-\lambda)^{n_0}, \qquad n_0 = \sum_j \delta(z_j), \quad n_1 = \sum_j \delta(z_j - 1) \end{array} \right.$$

Hierarchical models and hidden variables

- Example 2: Student-t model

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

- Infinite mixture

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

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

Hierarchical models and hidden variables

- Example 3: Laplace (Double Exponential) model

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

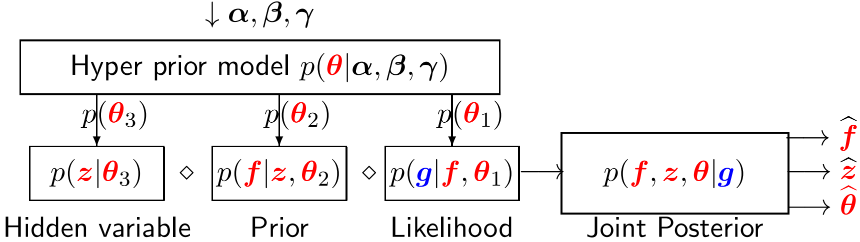
$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{z}) & = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0,z_j) \propto \exp\left[-\frac{1}{2}\sum_j f_j^2/z_j\right] \\ p(\boldsymbol{z}|\frac{a^2}{2}) & = \prod_j \mathcal{E}(z_j|\frac{a^2}{2}) \propto \exp\left[\sum_j \frac{a^2}{2} z_j\right] \\ p(\boldsymbol{f},\boldsymbol{z}|\frac{a^2}{2})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

- With these models we have:

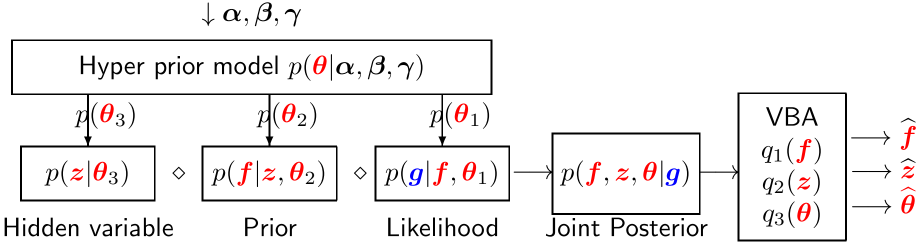
$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)\,p(\boldsymbol{z}|\boldsymbol{\theta}_3)\,p(\boldsymbol{\theta})$$

Summary of Bayesian estimation 3

- Full Bayesian Hierarchical Model with Hyperparameter Estimation



- Full Bayesian Hierarchical Model and Variational Approximation



8. Bayesian Computation and Algorithms for Hierarchical models

- 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.

Bayesian Variational Approximation

- Objective: Approximate $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable one $q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{f})\, q_3(\boldsymbol{\theta})$
- Criterion:

$$\mathsf{KL}(q:p) = \int q \ln \frac{q}{p} = \left\langle \ln \frac{q}{p} \right\rangle_q$$

- Free energy: $\mathsf{KL}(q:p) = \ln p(\boldsymbol{g}|\mathcal{M}) - \mathcal{F}(q)$ where:

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

  with $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g}|\mathcal{M}) = p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta})\, p(\boldsymbol{z}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$ and $\mathcal{F}(q)$ is the free energy associated to q defined as

$$\mathcal{F}(q) = \left\langle \ln \frac{p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g}|\mathcal{M})}{q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})} \right\rangle_q$$

- For a given model $\mathcal{M}$, minimizing $\mathsf{KL}(q:p)$ is equivalent to maximizing $\mathcal{F}(q)$ and when optimized, $\mathcal{F}(q^*)$ gives a lower bound for $\ln p(\boldsymbol{g}|\mathcal{M})$.

BVA with Scale Mixture Model of Student-t priors

Scale Mixture Model of Student-t:

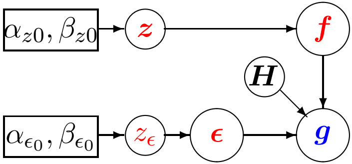
$$\mathcal{S}t(f_j|\nu) = \int_0^\infty \mathcal{N}(f_j|0, 1/z_j)\,\mathcal{G}(z_j|\nu/2, \nu/2)\ \mathrm{d}z_j$$

Hidden variables z_j:

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

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

- Graphical model:

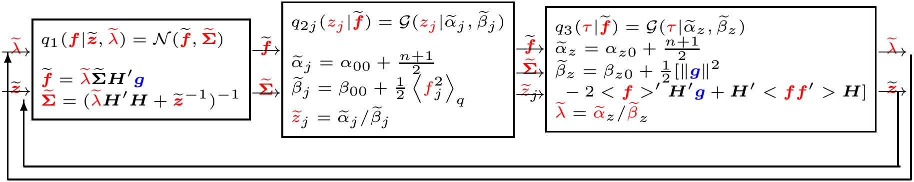


12. BVA with Student-t priors Algorithm

$$
\begin{cases}
p(\boldsymbol{g}|\boldsymbol{f}, z_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, (1/z_\epsilon)\boldsymbol{I}) \\
p(z_\epsilon|\alpha_{z0}, \beta_{z0}) = \mathcal{G}(z_\epsilon|\alpha_{z0}, \beta_{z0}) \\
p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j \mathcal{N}(f_j|0, 1/z_j) \\
p(\boldsymbol{z}|\alpha_0, \beta_0) = \prod_j \mathcal{G}(z_j|\alpha_0, \beta_0) \\
\\
q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}}) = \mathcal{N}(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}}) \\
\widetilde{\boldsymbol{\mu}} = \langle\lambda\rangle_q \widetilde{\boldsymbol{\Sigma}}\boldsymbol{H}'\boldsymbol{g} \\
\widetilde{\boldsymbol{\Sigma}} = (\langle\lambda\rangle_q \boldsymbol{H}'\boldsymbol{H} + \widetilde{\boldsymbol{Z}})^{-1}, \\
\text{with } \widetilde{\boldsymbol{Z}} = \widetilde{\boldsymbol{z}}^{-1} = \text{diag}\,[\widetilde{\boldsymbol{z}}]
\end{cases}
$$

$$
\begin{cases}
q_{2j}(z_j) = \mathcal{G}(z_j|\widetilde{\alpha}_j, \widetilde{\beta}_j) \\
\widetilde{\alpha}_j = \alpha_{00} + 1/2 \\
\widetilde{\beta}_j = \beta_{00} + < f_j^2 > /2 \\
\\
q_3(z_\epsilon) = \mathcal{G}(z_\epsilon|\widetilde{\alpha}_{z\epsilon}, \widetilde{\beta}_{z\epsilon}), \\
\widetilde{\alpha}_{z\epsilon} = \alpha_{z0} + (n+1)/2 \\
\widetilde{\beta}_{z\epsilon} = \beta_{z0} + 1/2[\|\boldsymbol{g}\|^2 \\
-2\,\langle \boldsymbol{f}\rangle_q' \boldsymbol{H}'\boldsymbol{g} + \boldsymbol{H}'\,\langle \boldsymbol{f}\boldsymbol{f}'\rangle_q \boldsymbol{H}]
\end{cases}
$$

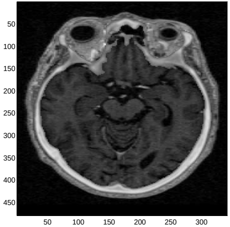
$$
\begin{cases}
< \boldsymbol{f} >= \widetilde{\boldsymbol{\mu}} \\
< \boldsymbol{f}\boldsymbol{f}' >= \widetilde{\boldsymbol{\Sigma}} + \widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}} \\
< f_j^2 >= [\widetilde{\boldsymbol{\Sigma}}]_{jj} + \widetilde{\mu} \\
\widetilde{\lambda} = \widetilde{\alpha}_z/\widetilde{\beta}_z \\
\widetilde{z}_j = \widetilde{\alpha}_j/\widetilde{\beta}_j
\end{cases}
$$

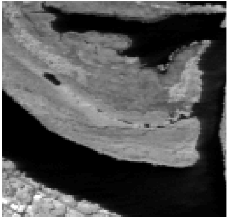


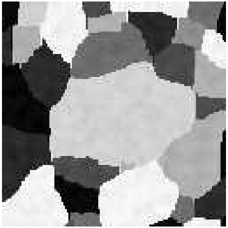
13. Implementation issues

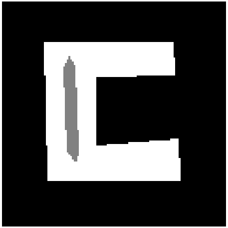
- In inverse problems, often we do not have access directly to the matrix $\boldsymbol{H}$. But, we can compute:
  - Forward operator : $\boldsymbol{H}\boldsymbol{f} \longrightarrow \boldsymbol{g}$ `g=direct(f,...)`
  - Adjoint operator : $\boldsymbol{H}'\boldsymbol{g} \longrightarrow \boldsymbol{f}$ `f=transp(g,...)`
- For any particular application, we can always write two programs (`direct` & `transp`) corresponding to the application of these two operators.
- To compute $\widetilde{\boldsymbol{f}}$, we use a gradient based optimization algorithm which will use these operators.
- We may also need to compute the diagonal elements of $[\boldsymbol{H}'\boldsymbol{H}]$.. We also developped algorithms which computes these diagonal elements with the same programs (`direct` & `transp`)

Which images I am looking for?

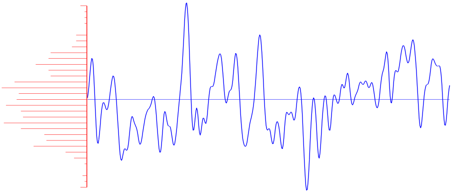




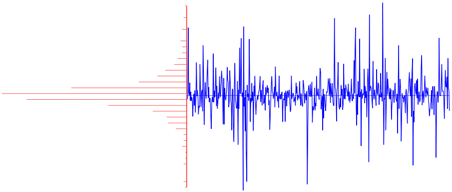




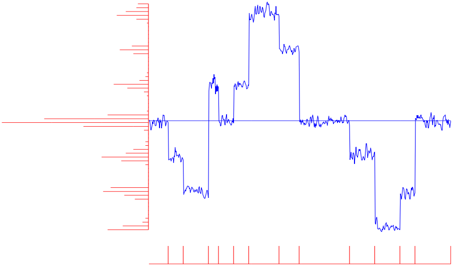
Which image I am looking for?



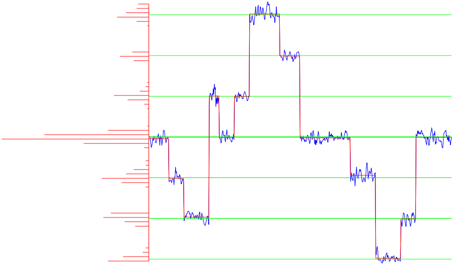
Gauss-Markov



Generalized GM

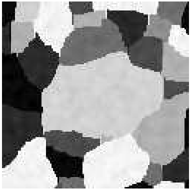


Piecewize Gaussian

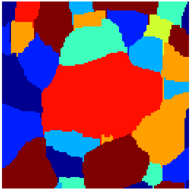


Mixture of GM

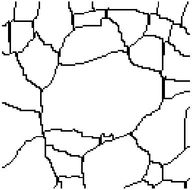
9. Gauss-Markov-Potts prior models for images



$f(\boldsymbol{r})$



$z(\boldsymbol{r})$



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

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

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

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

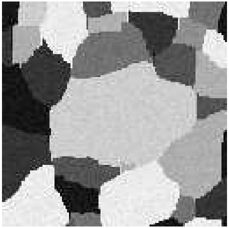
$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left[\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]$$

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

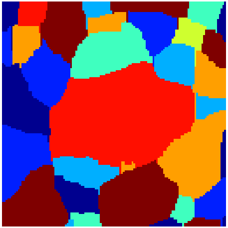
Four different cases

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)



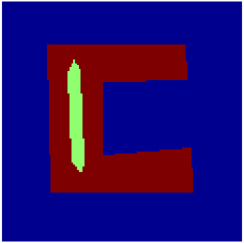
$f(\boldsymbol{r})$

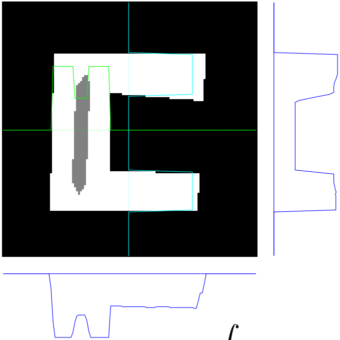


$z(\boldsymbol{r})$

Application of CT in NDT

Reconstruction from only 2 projections



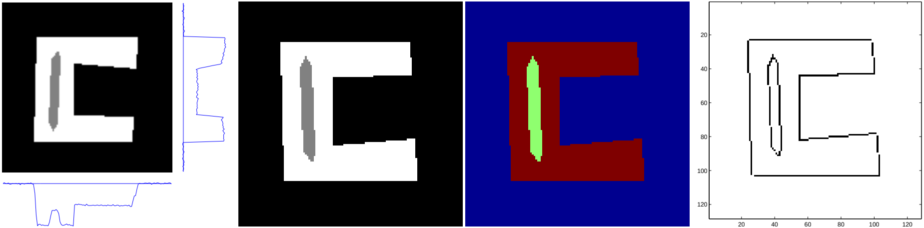


$$g_1(x) = \int f(x,y) \, \mathrm{d}y, \qquad g_2(y) = \int f(x,y) \, \mathrm{d}x$$

- 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



$\boldsymbol{g}|\boldsymbol{f}$

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

$\boldsymbol{g}|\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, \sigma_\epsilon^2 \boldsymbol{I})$

Gaussian

$\boldsymbol{f}|\boldsymbol{z}$

iid Gaussian

or

Gauss-Markov

$\boldsymbol{z}$

iid

or

Potts

$\boldsymbol{c}$

$c(\boldsymbol{r}) \in \{0, 1\}$

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

binary

Proposed algorithm

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

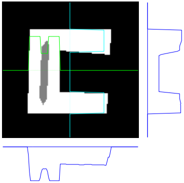
General scheme:

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

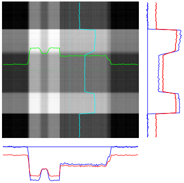
Iterative algorithme:

- 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.

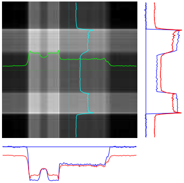
Results



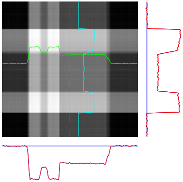
Original



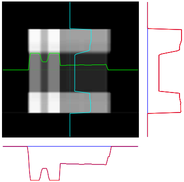
Backprojection



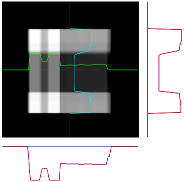
Filtered BP



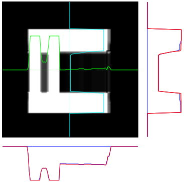
LS



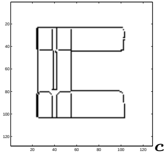
Gauss-Markov+pos

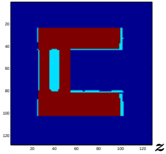


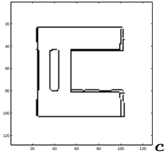
GM+Line process



GM+Label process





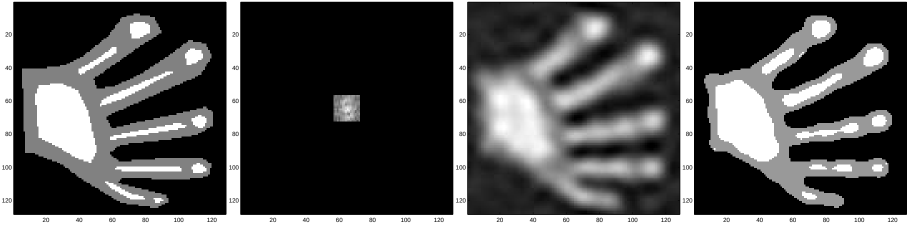


Application in Microwave imaging

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

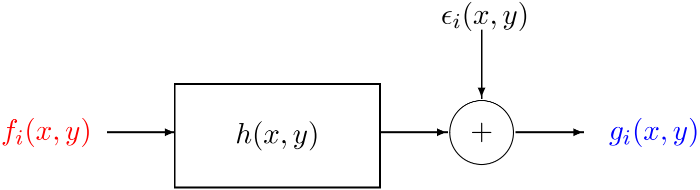
$$g(u,v) = \iint f(x,y) \exp\left[-j(ux+vy)\right] \,\mathrm{d}x \,\mathrm{d}y + \epsilon(u,v)$$

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

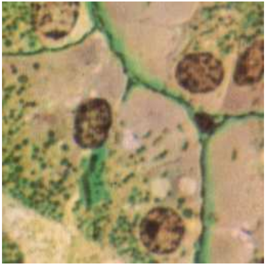


$f(x,y)$ $\quad g(u,v)$ $\quad \widehat{\boldsymbol{f}}$ IFT $\quad \widehat{\boldsymbol{f}}$ Proposed method

Color (Multi-spectral) image deconvolution

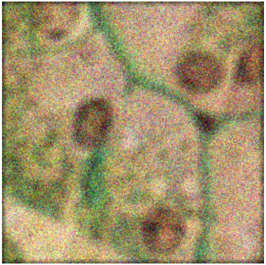


Observation model : $\boldsymbol{g}_i = \boldsymbol{H}\boldsymbol{f}_i + \boldsymbol{\epsilon}_i, \quad i = 1,2,3$



$?$

$\Longleftarrow$

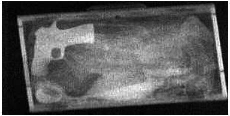


Images fusion and joint segmentation

(with O. Féron)

$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2) \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \\ p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \end{cases}$$

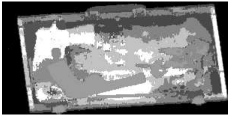
$\boldsymbol{g}_1$



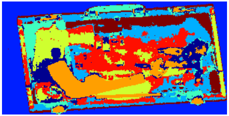
$\boldsymbol{g}_2$

$\longrightarrow$

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



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

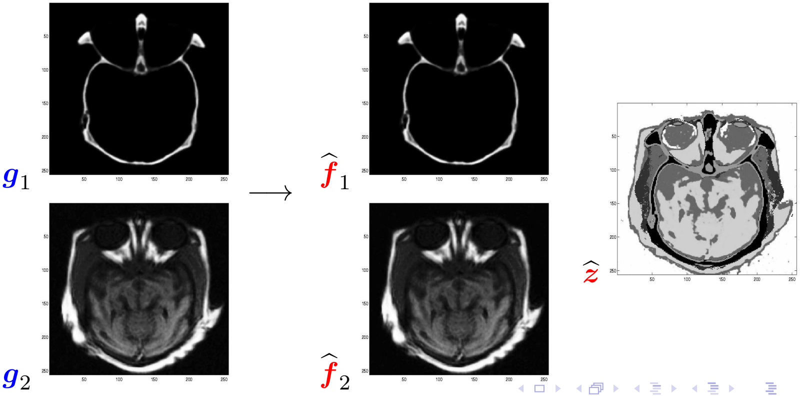


$\widehat{\boldsymbol{z}}$

Data fusion in medical imaging

(with O. Féron)

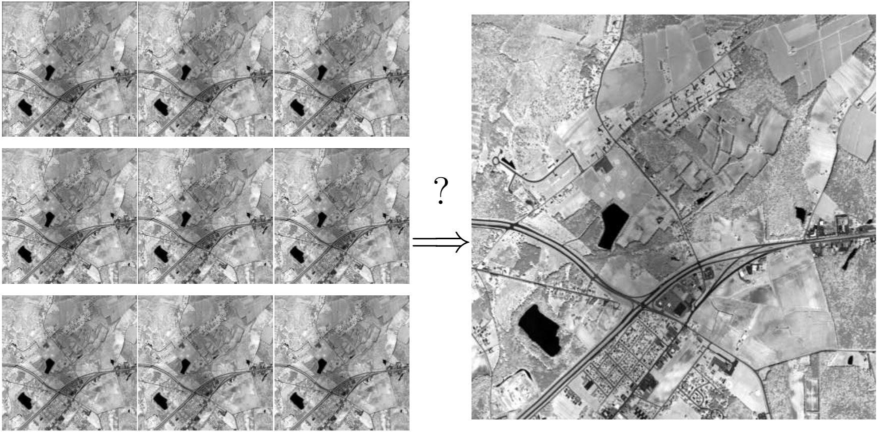
$$
\begin{cases}
g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\
p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2) \\
p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \\
p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]
\end{cases}
$$



Super-Resolution

(with F. Humblot)

$$\left\{\begin{array}{l} g_i(\boldsymbol{r}) = [\mathcal{D}\mathcal{M}\mathcal{B} f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma^2_{i\,k}) \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \\ p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \end{array}\right.$$

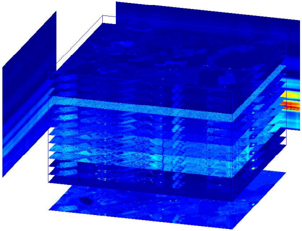


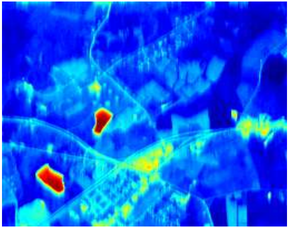
Low Resolution images High Resolution image

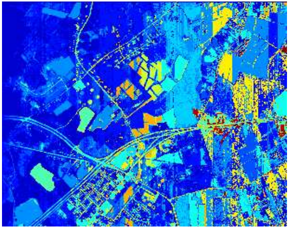
Joint segmentation of hyper-spectral images

(with N. Bali & A. Mohammadpour)

$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma_{i\,k}^2), \quad k = 1, \cdots, K \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \\ p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \\ m_{ik} \quad \text{follow a Markovian model along the index} \quad i \end{cases}$$



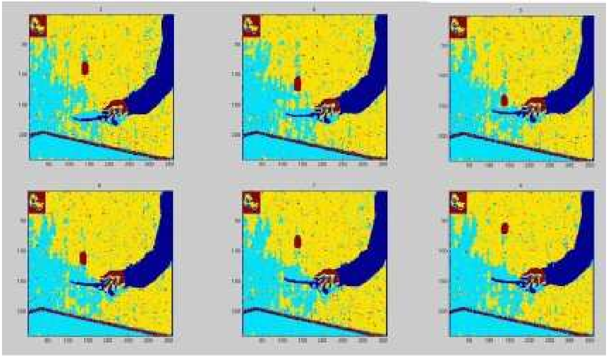




Segmentation of a video sequence of images

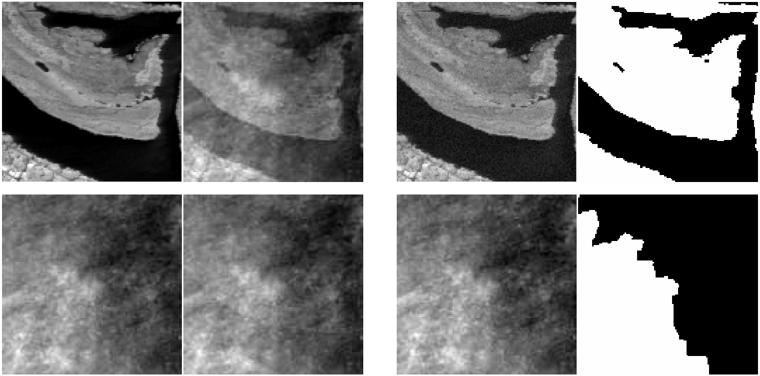
(with P. Brault)

$$\left\{\begin{array}{l} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z_i(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2), \quad k = 1, \cdots, K \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}_i) \\ p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \\ z_i(\boldsymbol{r}) \;\; \text{follow a Markovian model along the index} \;\; i \end{array}\right.$$



Source separation: (with H. Snoussi & M. Ichir)

$$
\left\{
\begin{array}{l}
g_i(\boldsymbol{r}) = \sum_{j=1}^{N} A_{ij} f_j(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\
p(f_j(\boldsymbol{r}) | z_j(\boldsymbol{r}) = k) = \mathcal{N}(m_{j\,k}, \sigma^2_{j\,k}) \\
p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \\
p(A_{ij}) = \mathcal{N}(A_{0\,ij}, \sigma^2_{0\,ij})
\end{array}
\right.
$$



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

Conclusions

- Bayesian Inference for inverse problems
- Different prior modeling for signals and images:
  Separable, Markovian, without and with hidden variables
- Sprasity enforcing priors
- Gauss-Markov-Potts models for images incorporating hidden regions and contours
- Two main Bayesian computation tools: MCMC and VBA
- Application in different CT (X ray, Microwaves, PET, SPECT)

Current Projects and Perspectives :

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

Current Applications and Perspectives

We use these models for inverse problems in different signal and image processing applications such as:

- Period estimation in biological time series
- Signal deconvolution in Proteomic and molecular imaging

- X ray Computed Tomography
- Diffraction Optical Tomography
- Microwave Imaging, Acoustic imaging and sources localization
- Synthetic Aperture Radar (SAR) Imaging

Thanks to:

Graduated PhD students:

- Sh. Zhu (2012: SAR Imaging)
- D. Fall (2012: Emission Positon Tomography, Non Parametric Bayesian)
- D. Pougaza (2011: Copula and Tomography)
- H. Ayasso (2010: Optical Tomography, Variational Bayes)
- S. Fékih-Salem (2009: 3D X ray Tomography)
- N. Bali (2007: Hyperspectral imaging)
- O. Féron (2006: Microwave imaging)
- F. Humblot (2005: Super-resolution)
- M. Ichir (2005: Image separation in Wavelet domain)
- P. Brault (2005: Video segmentation using Wavelet domain)
- H. Snoussi (2003: Sources separation)
- Ch. Soussen (2000: Geometrical Tomography)
- G. Montémont (2000: Detectors, Filtering)
- H. Carfantan (1998: Microwave imaging)
- S. Gautier (1996: Gamma ray imaging for NDT)
- M. Nikolova (1994: Piecewise Gaussian models and GNC)
- D. Prémel (1992: Eddy current imaging)

Thanks to:

Present PhD students:

- C. Cai (Multispectral X ray Tomography)
- N. Chu (Acoustic sources localization)
- Th. Boulay (Non Cooperative Radar Target Recognition)
- R. Prenon (Proteomic and Masse Spectrometry)
- L. Gharsali (Microwave imaging for Cancer detection)
- M. Dumitru (Multivariate time series analysis for biological signals)

Present Master students:

- A. Cai (Non-circular X ray Tomography)
- F. Fuc (Multi component signal analysis for biology applications)

Post-Docs:

- J. Lapuyade (2011: Dimentionality Reduction and multivariate analysis)
- S. Su (2006: Color image separation)
- A. Mohammadpour (2004: HyperSpectral image segmentation)

Thanks my colleagues and collaborators

- B. Duchêne & A. Joisel (L2S) (Inverse scattering and Microwave Imaging)
- N. Gac (L2S) (GPU Implementation)
- Th. Rodet (L2S) (Computed 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)
- J. Picheral (SSE, Supélec) (Acoustic sources localization)
- D. Blacodon (ONERA) (Acoustic sources separation)
- J. Lagoutte (Thales Air Systems) (Non Cooperative Radar Target Recognition)
- P. Grangeat (LETI, CEA, Grenoble) (Proteomic and Masse Spectrometry)
- F. Lévi (CNRS-INSERM, Hopital Paul Brousse) (Biological rythms and Chronotherapy of Cancer)

References 1

- A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," EURASIP Journal on Advances in Signal Processing, vol. Special issue on Sparse Signal Processing, (2012).
- 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.
- 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.
- H. Ayasso and Ali Mohammad-Djafari Joint NDT Image Restoration and Segmentation using Gauss-Markov-Potts Prior Models and Variational Bayesian Computation, 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.
- N. Bali and A. Mohammad-Djafari, "Bayesian Approach With Hidden Markov Modeling and Mean Field Approximation for Hyperspectral Data Analysis," IEEE Trans. on Image Processing 17: 2. 217-225 Feb. (2008).
- H. Snoussi and J. Idier., "Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures," *IEEE Trans. on Signal Processing*, 2006.

References 2

- 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.
- H. Snoussi and A. Mohammad-Djafari, "Estimation of Structured Gaussian Mixtures: The Inverse EM Algorithm," IEEE Trans. on Signal Processing 55: 7. 3185-3191 July (2007).
- N. Bali and A. Mohammad-Djafari, "A variational Bayesian Algorithm for BSS Problem with Hidden Gauss-Markov Models for the Sources," in: Independent Component Analysis and Signal Separation (ICA 2007) Edited by:M.E. Davies, Ch.J. James, S.A. Abdallah, M.D. Plumbley. 137-144 Springer (LNCS 4666) (2007).

References 3

- N. Bali and A. Mohammad-Djafari, "Hierarchical Markovian Models for Joint Classification, Segmentation and Data Reduction of Hyperspectral Images" ESANN 2006, September 4-8, Belgium. (2006)
- M. Ichir and A. Mohammad-Djafari, "Hidden Markov models for wavelet-based blind source separation," IEEE Trans. on Image Processing 15: 7. 1887-1899 July (2005)
- S. Moussaoui, C. Carteret, D. Brie and A Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods sampling," Chemometrics and Intelligent Laboratory Systems 81: 2. 137-148 (2005).
- H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images" Journal of Electronic Imaging 13: 2. 349-361 April (2004)
- H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," Journal of VLSI Signal Processing Systems 37: 2/3. 263-279 June/July (2004)
- F. Su and A. Mohammad-Djafari, "An Hierarchical Markov Random Field Model for Bayesian Blind Image Separation," 27-30 May 2008, Sanya, Hainan, China: International Congress on Image and Signal Processing (CISP 2008).
- N. Chu, J. Picheral and A. Mohammad-Djafari, "A robust super-resolution approach with sparsity constraint for near-field wideband acoustic imaging," *IEEE International Symposium on Signal Processing and Information Technology* pp 286–289, Bilbao, Spain, Dec14-17,2011