Inverse problems in signal and image processing and Bayesian inference framework: from basic to advanced Bayesian computation

Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes (L2S)
UMR8506 CNRS-CentraleSupélec-UNIV PARIS SUD
SUPELEC, 91192 Gif-sur-Yvette, France
http://lss.centralesupelec.fr
Email: djafari@lss.supelec.fr
http://djafari.free.fr
http://publicationslist.org/djafari

Contents

1. Signal and Image Processing:
   Classical/Inverse problems approaches
2. Inverse problems examples
   - Instrumentation
   - Imaging systems to see outside of a body
   - Imaging systems to see inside of a body
   - Other imaging systems (Acoustics, Radar, SAR,...)
3. Analytical/Algebraic methods
4. Deterministic regularization methods and their limitations
5. Bayesian approach
6. Two main steps: Priors and Computational aspects
7. Case studies:
   Instrumentation, X ray Computed Tomography, Microwave imaging, Acoustic source localisation, Ultrasound imaging, Satellite image restoration, etc.

Signal and Image Processing: Classical/Inverse problems approach

- Classical: You have given a signal or an image, process it. Examples:
  - Signal:
    Detect periodicities, changes, Model it for prediction, ...
    AR, MA, ARMA modeling,... Parameter estimation,...
  - Image: Enhancement, Restoration, Segmentation, Contour detection, Compression, ...
- Model based or Inverse problem approach:
  - What represent the observed signal or image?
  - How they are related to the desired unknowns?
  - Forward modelling / Inversion
  - Examples: Deconvolution, Image restoration, Image reconstruction in Computed Tomography (CT), ...
- PCA, ICA / Blind source Separation,
- Compressed Sensing / L1 Regularization, Bayesian sparsity enforcing

Inverse Problems examples

- Example 1:
  Instrumentation: Measuring the temperature with a thermometer Deconvolution
  - $f(t)$ input of the instrument
  - $g(t)$ output of the instrument
- Example 2: Seeing outside of a body: Making an image using a camera, a microscope or a telescope: Image restoration
  - $f(x, y)$ real scene
  - $g(x, y)$ observed image
- Example 3: Seeing inside of a body: Computed Tomography usng X rays, US, Microwave, etc.: Image reconstruction
  - $f(x, y)$ a section of a real 3D body $f(x, y, z)$
  - $g_\phi(r)$ a line of observed radiographe $g_\phi(r, z)$
- Example 4: Seeing differently: MRI, Radar, SAR, Infrared, etc.: Fourier Synthesis
  - $f(x, y)$ a section of body or a scene
  - $g(u, v)$ partial data in the Fourier domain

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

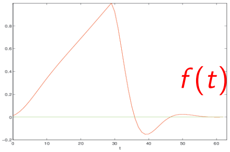
- Inverse problem: Deconvolution

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

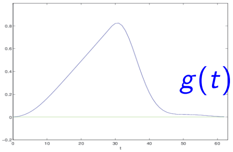
Measuring variation of temperature with a therometer

Forward model: Convolution

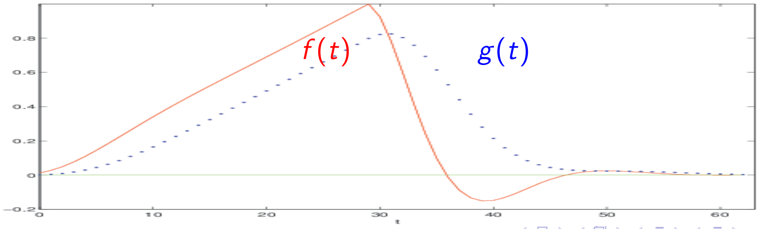
$$g(t) = \int f(t')\, h(t - t')\, \mathrm{d}t' + \epsilon(t)$$



$\longrightarrow$ Thermometer $h(t)$ $\longrightarrow$

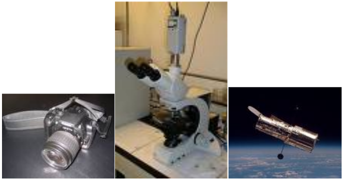


Inversion: Deconvolution



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

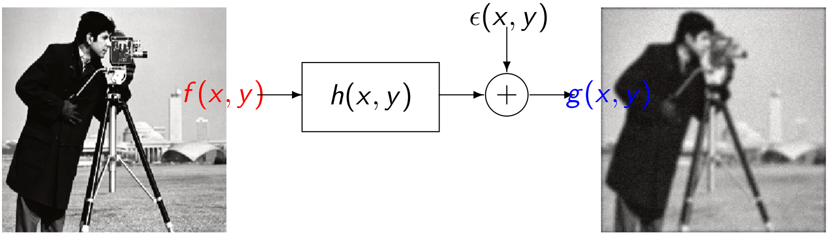
- Inverse problem: Image restoration

  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 $f(x, y)$

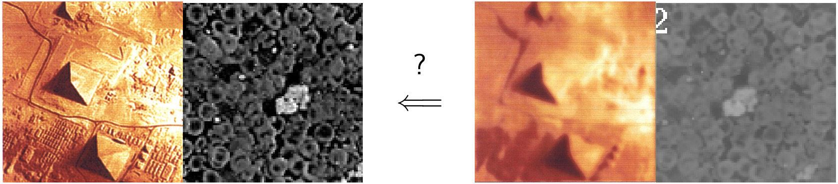
Making an image with an unfocused camera

Forward model: 2D Convolution

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

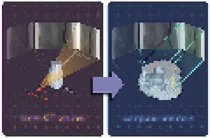


Inversion: Image Deconvolution or Restoration



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 radiography $g_\phi(r,z)$



- Forward model:
  Line integrals or Radon Transform

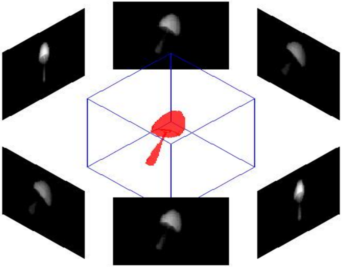
$$g_\phi(r) = \int_{L_{r,\phi}} f(x,y)\,\mathrm{d}l + \epsilon_\phi(r)$$
$$= \iint f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y + \epsilon_\phi(r)$$

- Inverse problem: Image reconstruction

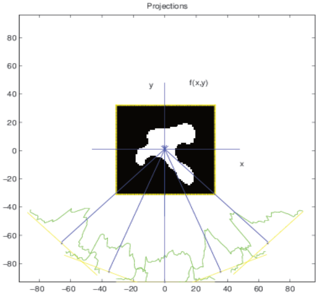
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x,y)$

2D and 3D Computed Tomography

3D



2D

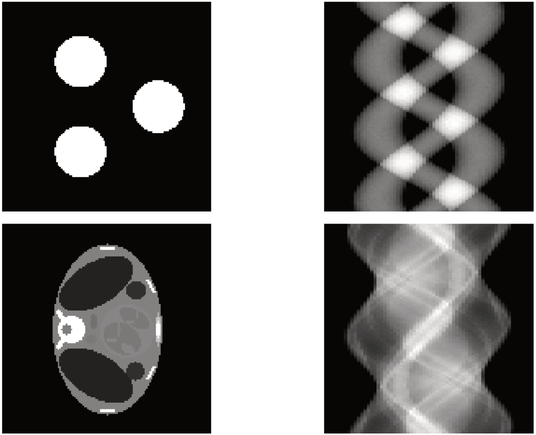


$$g_\phi(r_1, r_2) = \int_{\mathcal{L}_{r_1,r_2,\phi}} f(x,y,z) \, \mathrm{d}l \qquad g_\phi(r) = \int_{\mathcal{L}_{r,\phi}} f(x,y) \, \mathrm{d}l$$

Forward probelm: $f(x,y)$ or $f(x,y,z) \longrightarrow g_\phi(r)$ or $g_\phi(r_1,r_2)$

Inverse problem: $g_\phi(r)$ or $g_\phi(r_1,r_2) \longrightarrow f(x,y)$ or $f(x,y,z)$

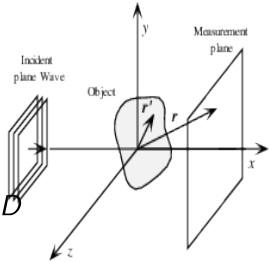
Computed Tomography: Radon Transform



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

Microwave or ultrasound imaging

Measures: diffracted wave by the object $g(\boldsymbol{r}_i)$
Unknown quantity: $f(\boldsymbol{r}) = k_0^2(n^2(\boldsymbol{r}) - 1)$
Intermediate quantity : $\phi(\boldsymbol{r})$

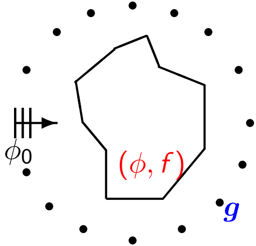


$$g(\boldsymbol{r}_i) = \iint_D G_m(\boldsymbol{r}_i, \boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \ \boldsymbol{r}_i \in S$$

$$\phi(\boldsymbol{r}) = \phi_0(\boldsymbol{r}) + \iint_D G_o(\boldsymbol{r}, \boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \ \boldsymbol{r} \in D$$

Born approximation $(\phi(\boldsymbol{r}') \simeq \phi_0(\boldsymbol{r}'))$ **):**

$$g(\boldsymbol{r}_i) = \iint_D G_m(\boldsymbol{r}_i, \boldsymbol{r}')\phi_0(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \ \boldsymbol{r}_i \in S$$



Discretization:

$$\begin{cases} \boldsymbol{g} = \boldsymbol{G}_m \boldsymbol{F} \boldsymbol{\phi} \\ \boldsymbol{\phi} = \boldsymbol{\phi}_0 + \boldsymbol{G}_o \boldsymbol{F} \boldsymbol{\phi} \end{cases} \longrightarrow \begin{cases} \boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) \\ \text{with} \quad \boldsymbol{F} = \mathrm{diag}(\boldsymbol{f}) \\ \boldsymbol{H}(\boldsymbol{f}) = \boldsymbol{G}_m \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{G}_o \boldsymbol{F})^{-1} \boldsymbol{\phi}_0 \end{cases}$$

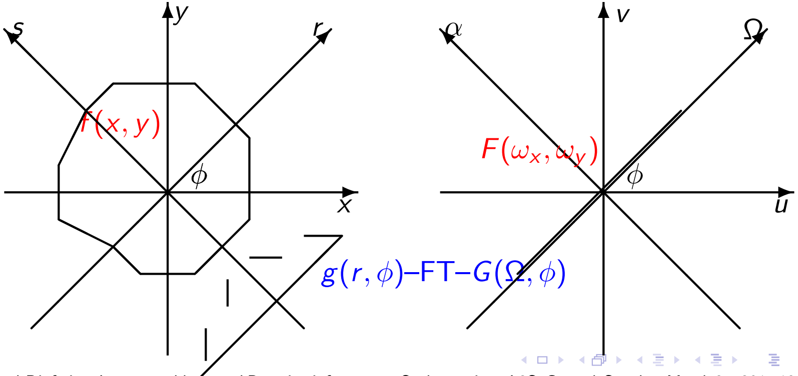
Fourier Synthesis in X ray Tomography

$$g(r,\phi) = \iint f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$$

$$G(\Omega,\phi) = \int g(r,\phi)\exp\left[-j\Omega r\right]\,\mathrm{d}r$$

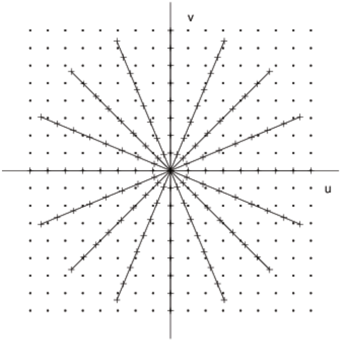
$$F(u,y) = \iint f(x,y)\exp\left[-jvx, yy\right]\,\mathrm{d}x\,\mathrm{d}y$$

$$F(v,y) = G(\Omega,\phi) \qquad \text{for} \quad u = \Omega\cos\phi \quad \text{and} \quad v = \Omega\sin\phi$$

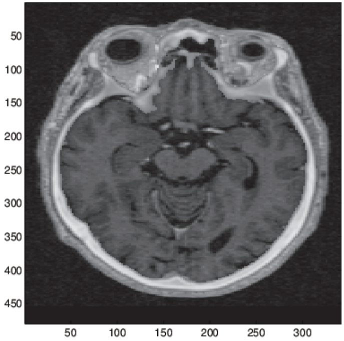


Fourier Synthesis in X ray tomography

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

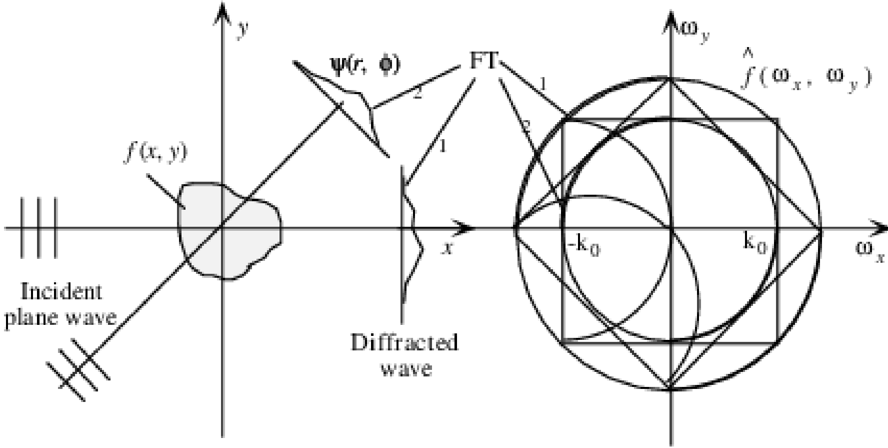


? $\Longrightarrow$



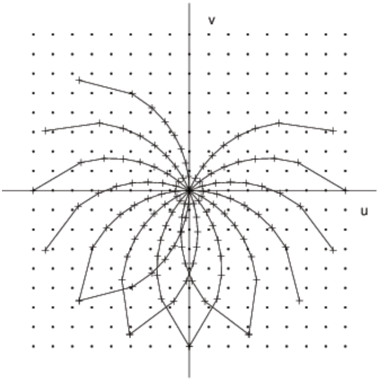
Forward problem: Given $f(x, y)$ compute $G(u, v)$
Inverse problem: Given $G(u, v)$ on those lines estimate $f(x, y)$

Fourier Synthesis in Diffraction tomography

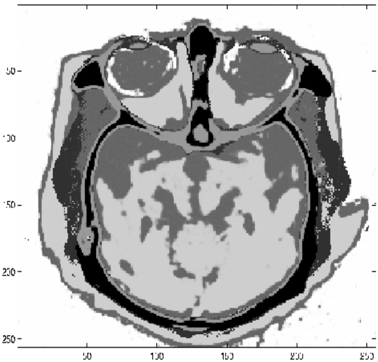


Fourier Synthesis in Diffraction tomography

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



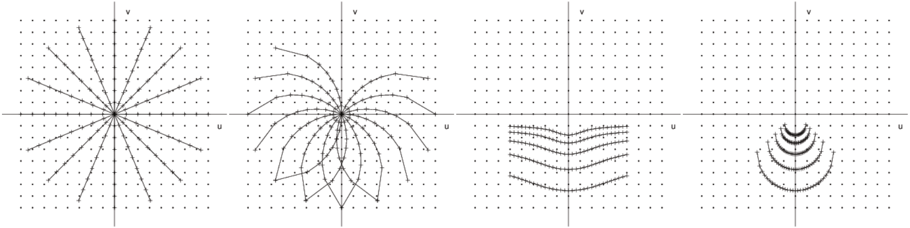
? $\Longrightarrow$



Forward problem: Given $f(x,y)$ compute $G(u,v)$
Inverse problem : Given $G(u,v)$ on those semi cercles estimate $f(x,y)$

Fourier Synthesis in different imaging systems

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



X ray Tomography　　Diffraction　　Eddy current　　SAR & Radar

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

Linear inverse problems

- Deconvolution

$$g(t) = \int f(\tau) h(t-\tau) \, \mathrm{d}\tau$$

- Image restoration

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

- Image reconstruction in X ray CT

$$g(r,\phi) = \int f(x,y) \delta(r - x\cos\phi - y\sin\phi) \, \mathrm{d}x \, \mathrm{d}y$$

- Fourier synthesis

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

- Unified linear relation

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

Linear Inverse Problems

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

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

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

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

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

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

- $\boldsymbol{H}$ is huge dimensional
  - 1D: $10^3 \times 10^3$, 2D: $10^6 \times 10^6$, 3D: $10^9 \times 10^9$
- Due to ill-posedness of the inverse problems, Least squares (LS) methods: $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$ with $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$ do not give satisfactory result. Need for regularization methods: $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$

Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

Functional space (Tikhonov): $g = \mathcal{H}(f) + \epsilon$

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

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

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$$

- 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

$$J(\boldsymbol{f}) = \Delta_1(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) + \lambda\Delta_2(\boldsymbol{D}\boldsymbol{f}, \boldsymbol{f}_0)$$

Limitations:

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

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

Bayesian estimation approach

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

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

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

- Regularization:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f}) = Q(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) + \lambda\Omega(\boldsymbol{f})\}$$

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

Case of linear models and Gaussian priors

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

- Prior knowledge on the noise:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I}) \rightarrow p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right]$$

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

$$\boldsymbol{f} \sim \mathcal{N}(0, \sigma_f^2 (\boldsymbol{D}'\boldsymbol{D})^{-1}) \rightarrow p(\boldsymbol{f}) \propto \exp\left[-\frac{1}{2\sigma_f^2}\|\boldsymbol{D}\boldsymbol{f}\|^2\right]$$

- A posteriori:

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

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

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

- Advantage : characterization of the solution

$$p(\boldsymbol{f}|\boldsymbol{g}) = \mathcal{N}(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{P}}) \text{ with } \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{P}}\boldsymbol{H}'\boldsymbol{g}, \quad \widehat{\boldsymbol{P}} = \left(\boldsymbol{H}'\boldsymbol{H} + \lambda\boldsymbol{D}'\boldsymbol{D}\right)^{-1}$$

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: $p(f_j) \propto \exp\left[-\alpha|f_j|^2\right] \longrightarrow \Omega(\boldsymbol{f}) = \alpha \sum_j |f_j|^2$
- 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:

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

MAP estimation with markovian priors:

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

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

with $\phi(t)$:

Convex functions:

$$|t|^\alpha, \ \sqrt{1+t^2}-1, \ \log(\cosh(t)), \ \begin{cases} t^2 & |t| \leq T \\ 2T|t| - T^2 & |t| > T \end{cases}$$

or Non convex functions:

$$\log(1+t^2), \quad \frac{t^2}{1+t^2}, \quad \arctan(t^2), \quad \begin{cases} t^2 & |t| \leq T \\ T^2 & |t| > T \end{cases}$$

- A great number of methods, optimization algorithms,...

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
- Tools for estimating hyper parameters
- Tools for model selection
- More specific and specialized priors, particularly through the hidden variables and hierarchical models
- More computational tools:
  - Expectation-Maximization for computing the maximum likelihood parameters
  - MCMC for posterior exploration
  - Variational Bayes for analytical computation of the posterior marginals
  - ...

Bayesian Estimation: Simple priors

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

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

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

- Generalized Gaussian prior & MAP:

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

- Double Exponential ($\beta = 1$):

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

Full (Unsupervised) Bayesian approach

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

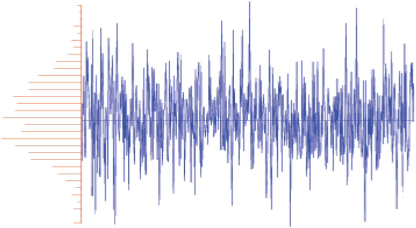
- Forward & errors model: $\longrightarrow p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1; \mathcal{M})$
- Prior models $\longrightarrow p(\boldsymbol{f}|\boldsymbol{\theta}_2; \mathcal{M})$
- Hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \longrightarrow p(\boldsymbol{\theta}|\mathcal{M})$
- 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:

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

Two main steps in the Bayesian approach

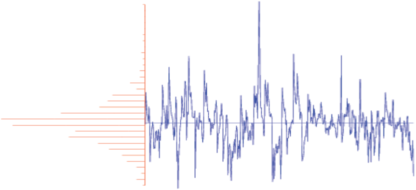
- Prior modeling
  - Separable:
    Gaussian, Gamma,
    Sparsity enforcing: Generalized Gaussian, mixture of Gaussians, mixture of Gammas, ...
  - Markovian:
    Gauss-Markov, GGM, ...
  - Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects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

Different prior models for signals and images: Separable



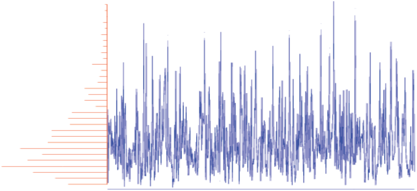
Gaussian

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



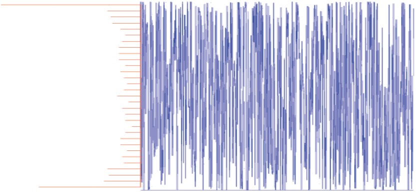
Generalized Gaussian

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



Gamma

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

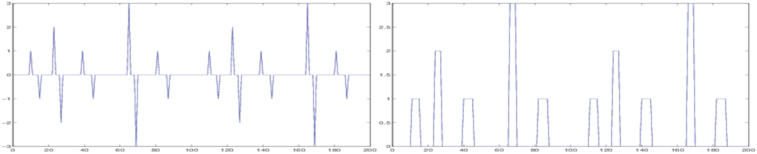


Beta

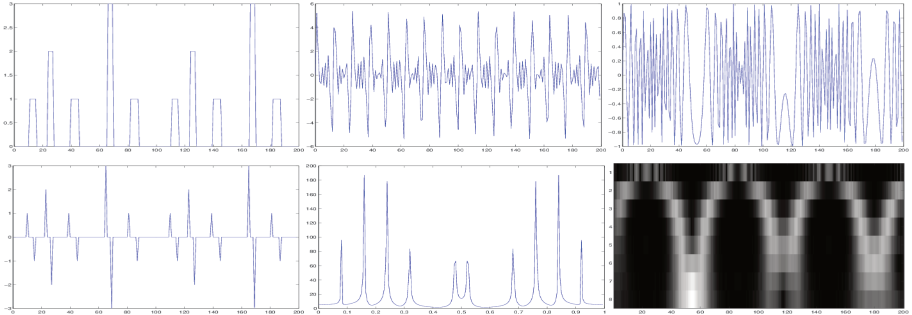
$p(f_j) \propto f_j^{\alpha}(1-f_j)^{\beta}$

Sparsity enforcing prior models

- Sparse signals: Direct sparsity



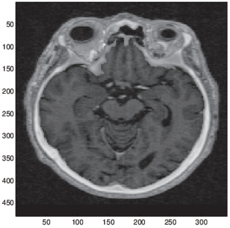
- Sparse signals: Sparsity in a Transform domain

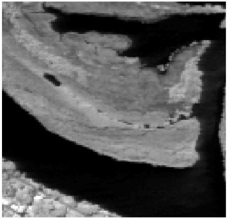


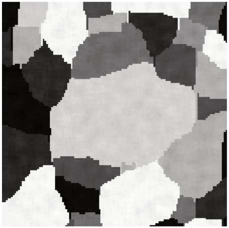
Sparsity enforcing prior models

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

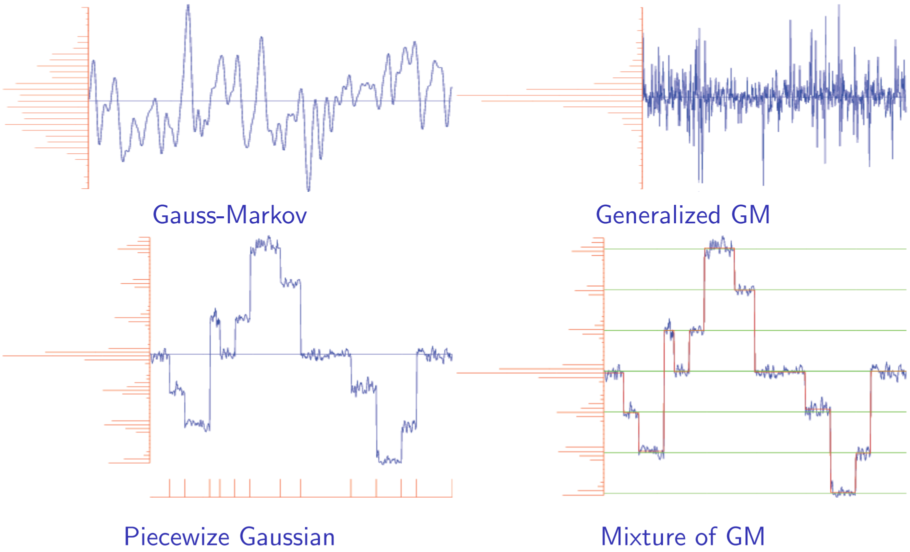
Which images I am looking for?







Which image I am looking for?



Different prior models for signals and images: Separable

- Simple Gaussian, Gamma, Generalized Gaussian

$$p(\boldsymbol{f}) \propto \exp\left[\sum_j \phi(f_j)\right]$$

- Simple Markovian models: Gauss-Markov, Generalized Gauss-Markov

$$p(\boldsymbol{f}) \propto \exp\left[\sum_j \sum_{j \in \mathcal{N}(i)} \phi(f_j - f_i)\right]$$

- Hierarchical models with hidden variables: Bernouilli-Gaussian, Gaussian-Gamma

$$p(\boldsymbol{f}|\boldsymbol{z}) \propto \exp\left[\sum_j p(f_j|z_j)\right] \text{ and } p(\boldsymbol{z}) \propto \exp\left[\sum_j p(z_j)\right]$$

with different choices for $p(f_j|z_j)$ and $p(z_j)$

Hierarchical models and hidden variables

- Student-t model

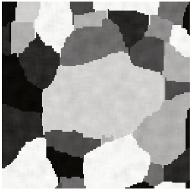
$$\mathcal{S}t(f|\nu) \propto \exp\left[-\frac{\nu+1}{2}\log\left(1+f^2/\nu\right)\right]$$

- Infinite Scaled Gaussian Mixture (ISGM) equivalence

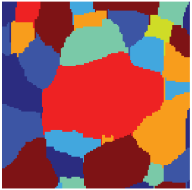
$$\mathcal{S}t(f|\nu) \propto= \int_0^\infty \mathcal{N}(f|,0,1/z)\,\mathcal{G}(z|\alpha,\beta)\,\mathrm{d}z, \quad \text{with } \alpha=\beta=\nu/2$$

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

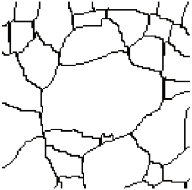
Gauss-Markov-Potts prior models for images



$f(\boldsymbol{r})$



$z(\boldsymbol{r})$



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

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

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

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

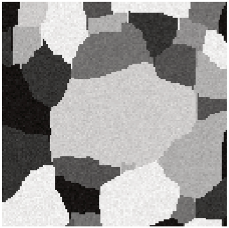
$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left[\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]$$

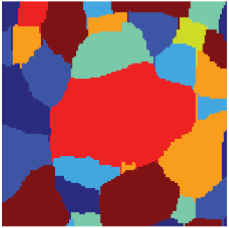
$$p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]$$

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)


$f(\boldsymbol{r})$


$z(\boldsymbol{r})$

Bayesian Computation and Algorithms

- Joint posterior probability law of all the unknowns $\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}$

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

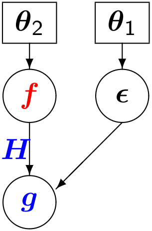
- Often, the expression of $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ is complex.
- Its optimization (for Joint MAP) or
  its marginalization or integration (for Marginal MAP or PM)
  is not easy
- Two main techniques:
  MCMC and Variational Bayesian Approximation (VBA)
-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

Hierarchical models

Simple case (1 layer):



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

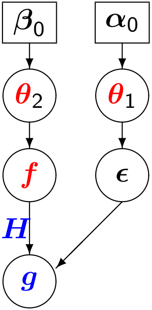
$p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{\theta}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)$

Objective: Infer on $\boldsymbol{f}$

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

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

Unsupervised case (2 layers):



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

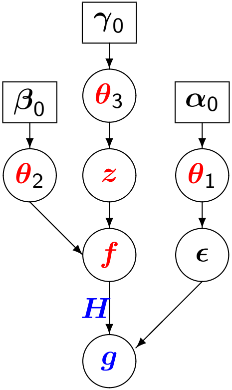
Objective: Infer on $\boldsymbol{f}, \boldsymbol{\theta}$

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

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

VBA: Approximate $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q_1(\boldsymbol{f})\, q_2(\boldsymbol{\theta})$

Hierarchical models (3 layers)



$p(f, z, \theta|g) \propto p(g|f, \theta_1)\, p(f|z, \theta_2)\, p(z|\theta_3)\, p(\theta)$
$p(\theta) = p(\theta_1|\alpha_0)\, p(\theta_2|\beta_0)\, p(\theta_3|\gamma_0)$
Objective: Infer on f, z, θ

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

Marginalization:

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

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

or $p(f|g) = \iint p(f, z, \theta|g)\, \mathrm{d}z\, \mathrm{d}\theta$

VBA:
Approximate $p(f, z, \theta|g)$ by $q_1(f)\, q_2(z)\, q_3(\theta)$

MCMC based algorithm

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

General scheme:

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

- Estimate $\boldsymbol{f}$ using $p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})\, p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}})$
  When Gaussian, can be done via optimisation of a quadratic criterion.

- Estimate $\boldsymbol{z}$ using $p(\boldsymbol{z}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}})\, p(\boldsymbol{z})$
  Often needs sampling (hidden discrete variable)

- Estimate $\boldsymbol{\theta}$ using
  $p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}},\sigma_\epsilon^2\boldsymbol{I})\, p(\widehat{\boldsymbol{f}}|\widehat{\boldsymbol{z}},(m_k,v_k))\, p(\boldsymbol{\theta})$
  Use of Conjugate priors $\longrightarrow$ analytical expressions.

Variational Bayesian Approximation

- Approximate $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\boldsymbol{g})\, q_2(\boldsymbol{\theta}|\boldsymbol{g})$ and then continue computations.
- Criterion $\mathrm{KL}(q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) : p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}))$
- $\mathrm{KL}(q:p) = \int\int q \ln q/p = \int\int q_1 q_2 \ln \frac{q_1 q_2}{p} = \int q_1 \ln q_1 + \int q_2 \ln q_2 - \int\int q \ln p = -H(q_1) - H(q_2) - < \ln p >_q$
- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

$$\left\{\begin{array}{ll} q_1(\boldsymbol{f}) & \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_2(\boldsymbol{\theta})}\right] \\ q_2(\boldsymbol{\theta}) & \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_1(\boldsymbol{f})}\right] \end{array}\right.$$

$p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) \longrightarrow$ [Variational Bayesian Approximation] $\longrightarrow \widehat{q}_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}}$; $\longrightarrow \widehat{q}_2(\boldsymbol{\theta}) \longrightarrow \widehat{\boldsymbol{\theta}}$

JMAP

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

Alternate optimization:

$$\begin{cases} \widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\} \\ \widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\} \end{cases}$$

Main drawbacks:

- Convergence
- Uncertainties in each step are not accounted for

Marginalization

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

- Main drawback: 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

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:

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

- GEM (Bayesian) algorithm:

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

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

JMAP, Marginalization, VBA

- JMAP:

$$\boxed{p(f,\theta|g) \text{ optimization}} \longrightarrow \widehat{f}, \quad \longrightarrow \widehat{\theta}$$

- Marginalization

$$\underbrace{\boxed{p(f,\theta|g)}}_{\text{Joint Posterior}} \longrightarrow \underbrace{\boxed{p(\theta|g)}}_{\text{Marginalize over } f} \longrightarrow \widehat{\theta} \longrightarrow \boxed{p(f|\widehat{\theta},g)} \longrightarrow \widehat{f}$$

- Variational Bayesian Approximation

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

Variational Bayesian Approximation

- Approximate $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f},\boldsymbol{\theta}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{\theta})$ and then use them for any inferences on $\boldsymbol{f}$ and $\boldsymbol{\theta}$ respectively.
- Criterion $\mathrm{KL}(q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) : p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}))$
  $\mathrm{KL}(q : p) = \int\int q \ln \frac{q}{p} = \int\int q_1 q_2 \ln \frac{q_1 q_2}{p}$
- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

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

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

$\longrightarrow q_2(\boldsymbol{\theta}) \longrightarrow \widehat{\boldsymbol{\theta}}$

Variational Bayesian Approximation

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

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

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

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

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 $\mathcal{M}$:

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

Minimizing $\mathrm{KL}(q:p)$ = Maximaizing $\mathcal{F}(q)$

VBA: Separable Approximation

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

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

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

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

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

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

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

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

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

- Appropriate choice for inverse problems

$$\left\{\begin{array}{ll} \widehat{q}_1(\boldsymbol{f}) & \propto p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}},\boldsymbol{g};\mathcal{M}) \\ \widehat{q}_2(\boldsymbol{\theta}) & \propto p(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}},\boldsymbol{g};\mathcal{M}) \end{array}\right. \longrightarrow \left\{\begin{array}{l} \text{Accounts for the uncertainties of} \\ \widehat{\boldsymbol{\theta}} \text{ for } \widehat{\boldsymbol{f}} \text{ and vise versa.} \end{array}\right.$$

Exponential families, Conjugate priors

JMAP, EM and VBA

JMAP Alternate optimization Algorithm:

$$\boldsymbol{\theta}^{(0)} \longrightarrow \widetilde{\boldsymbol{\theta}} \longrightarrow \boxed{\widetilde{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{ p(\boldsymbol{f}, \widetilde{\boldsymbol{\theta}} | \boldsymbol{g}) \right\}} \longrightarrow \widetilde{\boldsymbol{f}} \longrightarrow \widehat{\boldsymbol{f}}$$

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

$$\widehat{\boldsymbol{\theta}} \longleftarrow \widetilde{\boldsymbol{\theta}} \longleftarrow \boxed{\widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ p(\widetilde{\boldsymbol{f}}, \boldsymbol{\theta} | \boldsymbol{g}) \right\}} \longleftarrow \widetilde{\boldsymbol{f}}$$

EM:

$$\boldsymbol{\theta}^{(0)} \longrightarrow \widetilde{\boldsymbol{\theta}} \longrightarrow \boxed{q_1(\boldsymbol{f}) = p(\boldsymbol{f} | \widetilde{\boldsymbol{\theta}}, \boldsymbol{g})} \longrightarrow q_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}}$$

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

$$\widehat{\boldsymbol{\theta}} \longleftarrow \widetilde{\boldsymbol{\theta}} \longleftarrow \boxed{\begin{array}{c} Q(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \langle \ln p(\boldsymbol{f}, \boldsymbol{\theta} | \boldsymbol{g}) \rangle_{q_1(\boldsymbol{f})} \\ \widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ Q(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) \right\} \end{array}} \longleftarrow q_1(\boldsymbol{f})$$

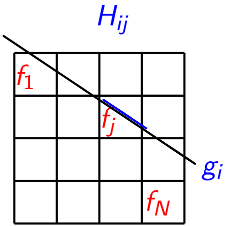
VBA:

$$\boldsymbol{\theta}^{(0)} \longrightarrow q_2(\boldsymbol{\theta}) \longrightarrow \boxed{q_1(\boldsymbol{f}) \propto \exp\left[\langle \ln p(\boldsymbol{f}, \boldsymbol{\theta} | \boldsymbol{g}) \rangle_{q_2(\boldsymbol{\theta})} \right]} \longrightarrow q_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}}$$

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

$$\widehat{\boldsymbol{\theta}} \longleftarrow q_2(\boldsymbol{\theta}) \longleftarrow \boxed{q_2(\boldsymbol{\theta}) \propto \exp\left[\langle \ln p(\boldsymbol{f}, \boldsymbol{\theta} | \boldsymbol{g}) \rangle_{q_1(\boldsymbol{f})} \right]} \longleftarrow q_1(\boldsymbol{f})$$

Computed Tomography: Discretization



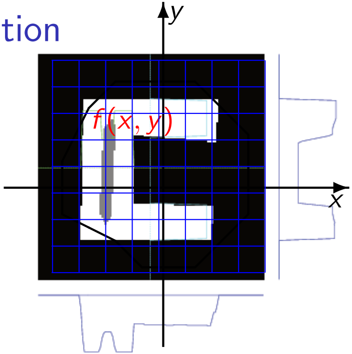
$$g(r,\phi) = \int_L f(x,y)\, \mathrm{d}l$$

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

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

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

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



Case study: Reconstruction from 2 projections

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

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

Very ill-posed inverse problem

$f(x,y) = g_1(x)\, g_2(y)\, \Omega(x,y)$

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

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

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

Simple example

1	3	4
2	4	6
3	7	

?	?	4
?	?	6
3	7	

f_1	f_3	g_3
f_2	f_4	g_4
g_1	g_2	

1	-1	0
-1	1	0
0	0	

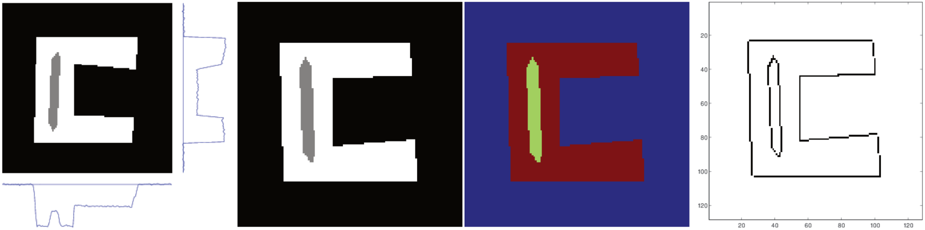
-1	1	0
1	-1	0
0	0	

$$\begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix}$$

f_1	f_4	f_7	g_4
f_2	f_5	f_8	g_5
f_3	f_6	f_9	g_6
g_1	g_2	g_3	

- $\boldsymbol{H}\boldsymbol{f} = \boldsymbol{g} \longrightarrow \widehat{\boldsymbol{f}} = \boldsymbol{H}^{-1}\boldsymbol{g}$ if $\boldsymbol{H}$ invertible.
- $\boldsymbol{H}$ is rank deficient: $\text{rank}(\boldsymbol{H}) = 3$
- Problem has infinite number of solutions.
- How to find all those solutions ?
- Which one is the good one? Needs prior information.
- To find an unique solution, one needs either more data or prior information.

Application in CT: Reconstruction from 2 projections



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

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

Proposed algorithms

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

- MCMC based general scheme:

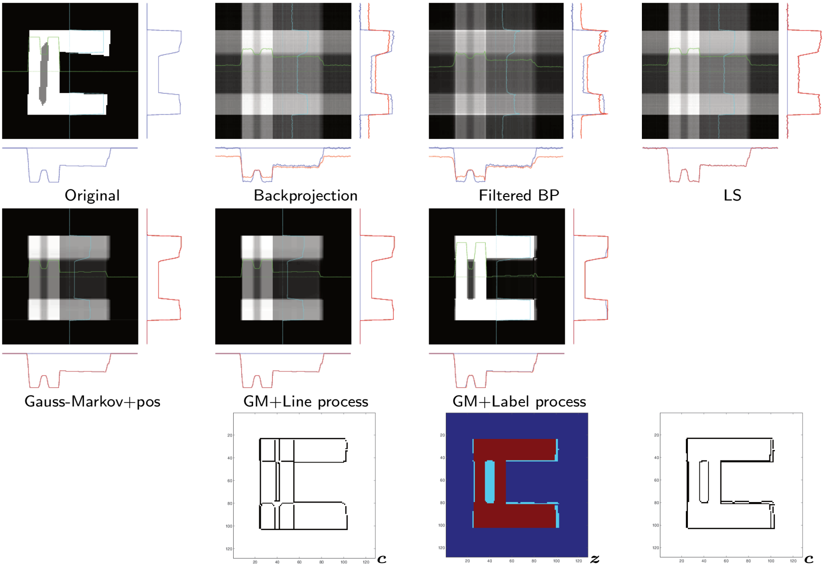
$$\widehat{\boldsymbol{f}} \sim p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{z}} \sim p(\boldsymbol{z}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{\theta}}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

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.

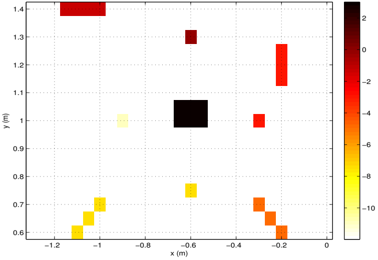
- Variational Bayesian Approximation
  - Approximate $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$

Results

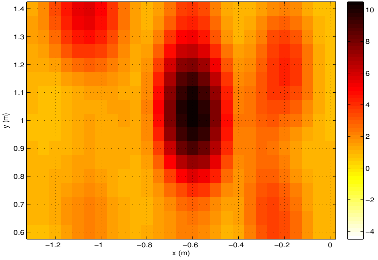

Original
Backprojection
Filtered BP
LS
Gauss-Markov+pos
GM+Line process
GM+Label process
c
z
c

Application in Acoustic source localization

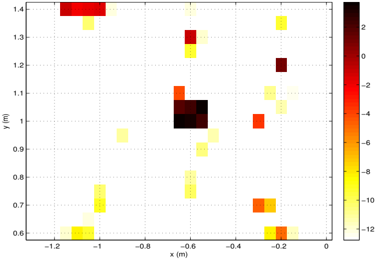
(Ning Chu et al.)



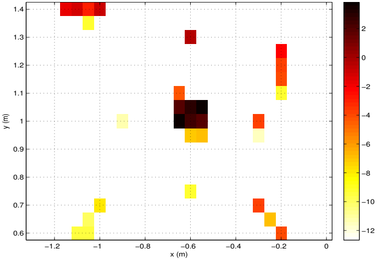
Source powers



Beamforming powers



Bayesian MAP inversion

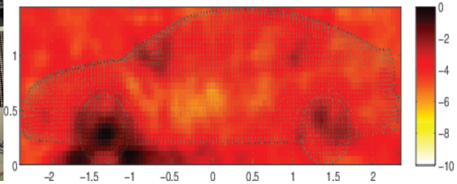


Proposed VBA inversion

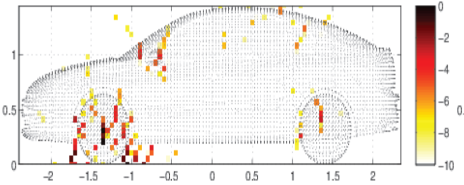
Application in Acoustic source localization

(Ning Chu et al.)

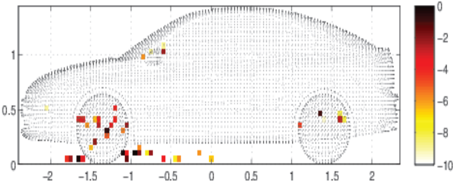
Wind tunnel



Beamforming



DAMAS



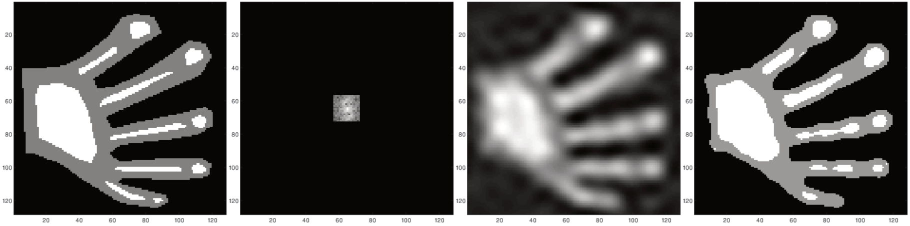
Proposed VBA inference

Application in Microwave imaging

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

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

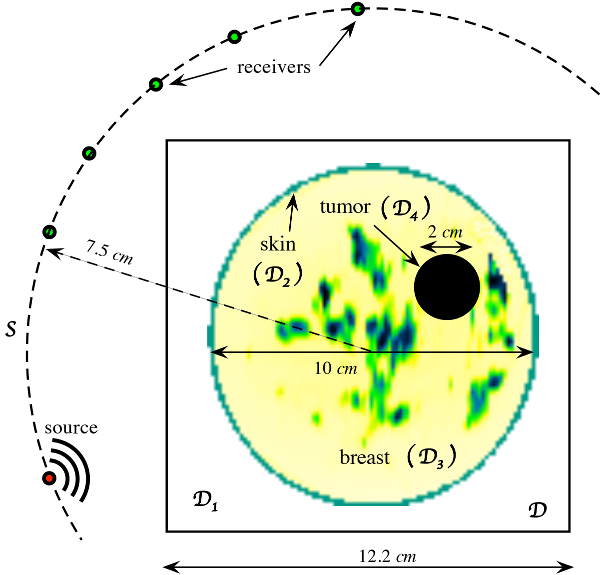
$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$



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

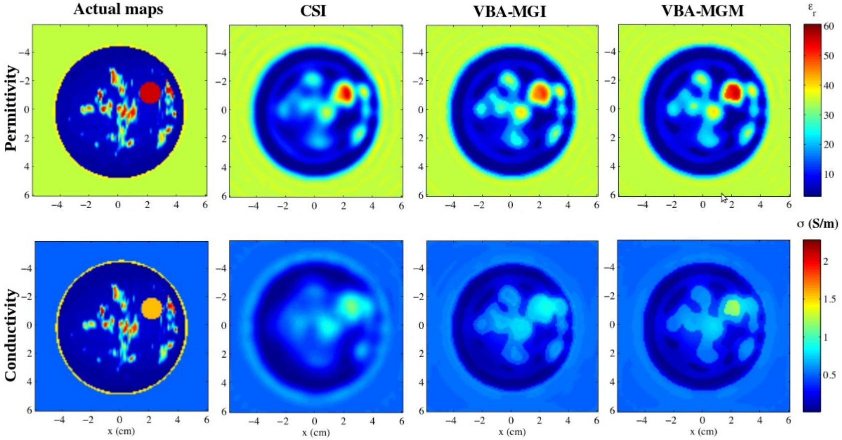
Microwave Imaging for Breast Cancer detection

(L. Gharsalli et al.)



Microwave Imaging for Breast Cancer detection

CSI: Contrast Source Inversion, VBA: Variational Bayesian Approach,
MGI: Independent Gaussian mixture, MGM: Gauss-Markov mixture

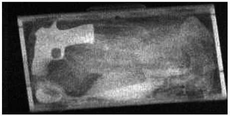


Images fusion and joint segmentation

(with O. Féron)

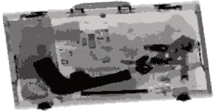
$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma^2_{i\,k}) \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \end{cases}$$

$\boldsymbol{g}_1$

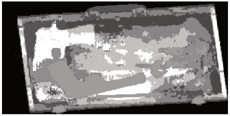


$\boldsymbol{g}_2$

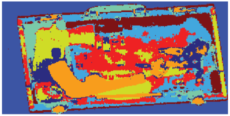
$\longrightarrow$



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



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

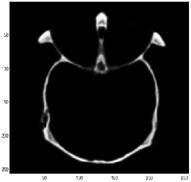


$\widehat{\boldsymbol{z}}$

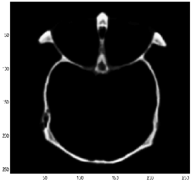
Data fusion in medical imaging

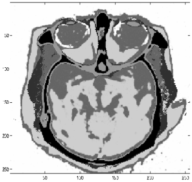
(with O. Féron)

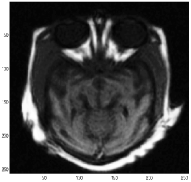
$$\left\{\begin{array}{l} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma^2_{i\,k}) \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \end{array}\right.$$

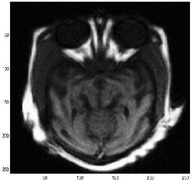
$\boldsymbol{g}_1$ 

$\longrightarrow$

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

$\widehat{\boldsymbol{z}}$ 

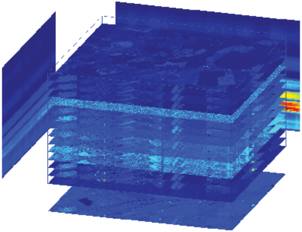
$\boldsymbol{g}_2$ 

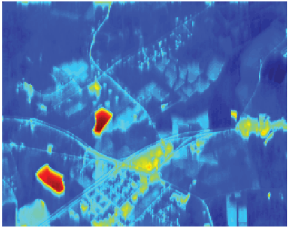
$\widehat{\boldsymbol{f}}_2$ 

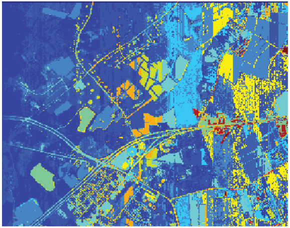
Joint segmentation of hyper-spectral images

(with N. Bali & A. Mohammadpour)

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







Segmentation of a video sequence of images

(with P. Brault)

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

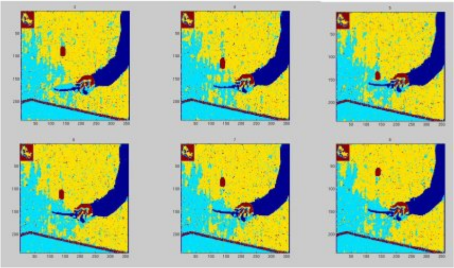
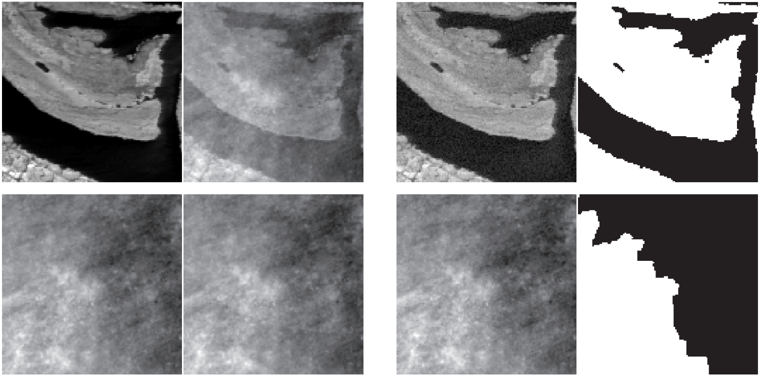


Image separation in Sattelite imaging

(with H. Snoussi & M. Ichir)

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



Conclusions

- Inverse problems arise in many science and engineering applications
- Deterministic Algorithms: Optimization of a two terms criterion, penalty term, regularization term
- Probabilistic: Bayesian approach
- Hierarchical prior model with hidden variables are very powerful tools for Bayesian approach to inverse problems.
- Gauss-Markov-Potts models for images incorporating hidden regions and contours
- Main Bayesian computation tools: JMAP, MCMC and VBA
- Application in different imaging system (X ray CT, Microwaves, PET, Ultrasound, Optical Diffusion Tomography (ODT), Acoustic source localization,...)

Current Projects:

- Efficient implementation in 2D and 3D cases

Thanks to:

Present PhD students:

- L. Gharsali (Microwave imaging for Cancer detection)
- M. Dumitru (Multivariate time series analysis for biological signals)
- S. AlAli (Diffraction imaging for geophysical applications)

Freshly Graduated PhD students:

- C. Cai (2013: Multispectral X ray Tomography)
- N. Chu (2013: Acoustic sources localization)
- Th. Boulay (2013: Non Cooperative Radar Target Recognition)
- R. Prenon (2013: Proteomic and Masse Spectrometry)
- 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)

Older Graduated PhD students:

- 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)

Thanks to:

Older Graduated PhD students:

- 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)

Post-Docs:

- J. Lapuyade (2011: Dimentionality Reduction and multivariate analysis)
- S. Su (2006: Color image separation)
- A. Mohammadpour (2004-2005: HyperSpectral image segmentation)

Colleagues:

- B. Duchêne & A. Joisel (L2S)& G. Perruson (Inverse scattering and Microwave Imaging)
- N. Gac (L2S) (GPU Implementation)
- Th. Rodet (L2S) (Computed Tomography)

Thanks to:

National Collaborators

- 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)

International Collaborators

- K. Sauer (Notre Dame University, IN, USA) (Computed Tomography, Inverse problems)
- F. Marvasti (Sharif University), (Sparse signal processing)
- M. Aminghafari (Amir Kabir University) (Independent Components Analysis)
- A. Mohammadpour (AKU) (Statistical inference)
- Gh. Yari (Tehran Technological University) (Probability and Analysis)

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. Duchêne 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).

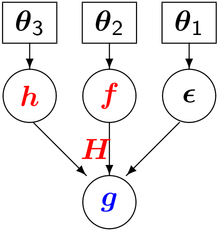
References 3

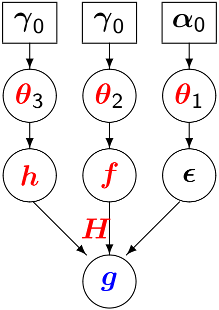
- N. Bali and A. Mohammad-Djafari, “Hierarchical Markovian Models for Joint Classification, Segmentation and Data Reduction of Hyperspectral Images” ESANN 2006, September 4-8, Belgium. (2006)
- M. Ichir and A. Mohammad-Djafari, “Hidden Markov models for wavelet-based blind source separation,” IEEE Trans. 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

Questions, Discussions, Open mathematical problems

- Sparsity representation, low rank matrix decomposition
  - Sparsity and positivity or other constraints
  - Group sparsity
  - Algorithmic and implementation issues for great dimensional applications (Big Data)
  - Joint estimation of Dictionary and coefficients
- Optimization of the KL divergence for Variational Bayesian Approximation
  - Convergency of alternate optimization
  - Other possible algorithms
- Properties of the obtained approximation
  - Does the moments of q's corresponds to the moments of p?
  - How about any other statistics: entropy, ...
- Other divergency or Distance measures?
- Using Sparsity as a prior in Inverse Problems
- Applications in Biological data and signal analysis, Medical imaging, Non Destructive Testing (NDT) Industrial Imaging, Communication, Geophysical imaging, Radio Astronomy, ...

Blind deconvolution (1)





$$\boldsymbol{g} = \boldsymbol{h} * \boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{\epsilon}$$

Simple priors:

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

Objective: Infer on $\boldsymbol{f}, \boldsymbol{h}$

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

VBA: Approximate $p(\boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g})$ by $q_1(\boldsymbol{f})\, q_2(\boldsymbol{h})$

Unsupervised:

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

Objective: Infer on $\boldsymbol{f}, \boldsymbol{h}, \boldsymbol{\theta}$

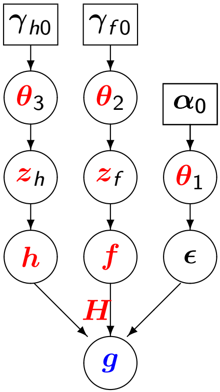
JMAP:

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

VBA:

Approximate $p(\boldsymbol{f}, \boldsymbol{h}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q_1(\boldsymbol{f})\, q_2(\boldsymbol{h})\, q_3(\boldsymbol{\theta})$

Blind deconvolution (2)



$$g = h * f + \epsilon = Hf + \epsilon = Fh + \epsilon$$

Simple priors:

$p(f, h|g, \theta) \propto p(g|f, h, \theta_1)\, p(f|\theta_2)\, p(h|\theta_3)$

Unsupervised:

$p(f, h, \theta|g) \propto p(g|f, h, \theta_1)\, p(f|\theta_2)\, p(h|\theta_3)\, p(\theta)$

Sparsity enforcing prior for f:

$p(f, z_f, h, \theta|g) \propto$
$p(g|f, h, \theta_1)\, p(f|z_f)\, p(z_f|\theta_2)\, p(h|\theta_3)\, p(\theta)$

Sparsity enforcing prior for h:

$p(f, h, z_h, \theta|g) \propto$
$p(g|f, h, \theta_1)\, p(f|\theta_2)\, p(h|z)\, p(z|\theta_3)\, p(\theta)$

Hierarchical models for both f and h:

$p(f, z_f, h, z_h, \theta|g) \propto$
$p(g|f, h, \theta_1)\, p(f|z_f)\, p(z_f|\theta_2)\, p(h|z_h)\, p(z_h|\theta_3)\, p(\theta)$

JMAP:

$(\widehat{f}, \widehat{z}_f, \widehat{h}, \widehat{z}_h, \widehat{\theta}) = \arg\max_{(f, z_f, \widehat{h}, \widehat{z}_h, \theta)} \left\{ p(f, z_f, \widehat{h}, \widehat{z}_h, \theta|g) \right\}$

VBA: Approximate $p(f, z_f, h, z_h, \theta|g)$ by
$q_1(f)\, q_2(z_f)\, q_3(h)\, q_4(z_h)\, q_5(\theta)$