Variational Bayesian Approximation for Linear Inverse Problems with a hierarchical prior models

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

Seminar 1 given at
Tsinghua University, Beijing, China, December 9, 2013

General linear inverse problem

Signal processing : $g(t) = \mathcal{H} f(t) + \epsilon(t), \quad t \in [1, \cdots, T]$
Image processing : $g(\boldsymbol{r}) = \mathcal{H} f(\boldsymbol{r}) + \epsilon(\boldsymbol{r}), \ \boldsymbol{r} = (x, y) \in R^2$
More general : $g(\boldsymbol{s}) = [\mathcal{H} f(\boldsymbol{r})](\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \ \boldsymbol{r} = (x, y), \boldsymbol{s} = (u, v)$

- f unknown quantity (input)
- $\mathcal{H}$ Forward operator:
  (Convolution, Radon, Fourier or any Linear operator)
- g observed quantity (output)
- ϵ represents the errors of modeling and measurement

Discretization:

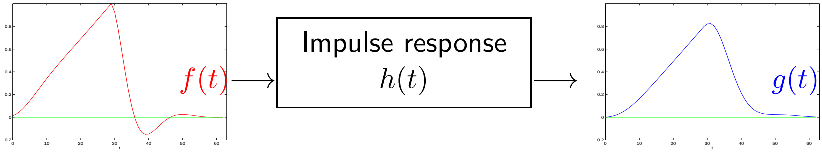
$$\boldsymbol{g} = \boldsymbol{H} \boldsymbol{f} + \boldsymbol{\epsilon}$$

- Forward operation $\boldsymbol{H} \boldsymbol{f}$
- Adjoint operation $\boldsymbol{H}' \boldsymbol{g}$: $\quad < \boldsymbol{H}' \boldsymbol{g}, \boldsymbol{f} > = < \boldsymbol{H} \boldsymbol{f}, \boldsymbol{g} >$
- Inverse operation (if exists) $\boldsymbol{H}^{-1} \boldsymbol{g}$, but this is never the case.

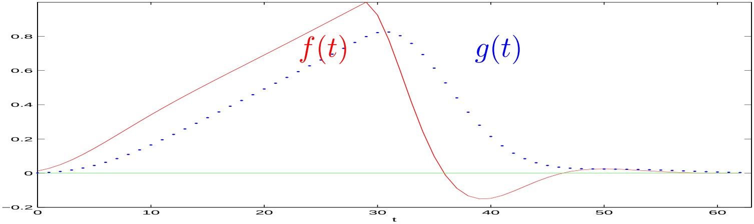
Example 1: Signal Deconvolution

Forward model: Convolution

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



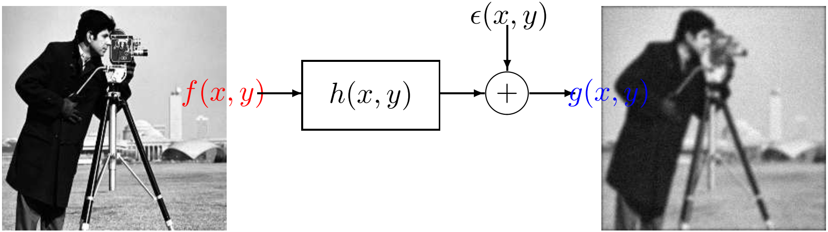
Inversion: Deconvolution



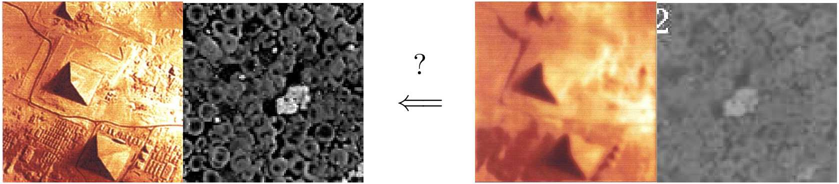
Example 2: Image restoration

Forward model: 2D Convolution

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

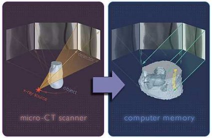


Inversion: Image Deconvolution or Restoration



Example 3: Image Reconstruction in Computed Tomography

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



- Forward model:
  Line integrals or Radon Transform

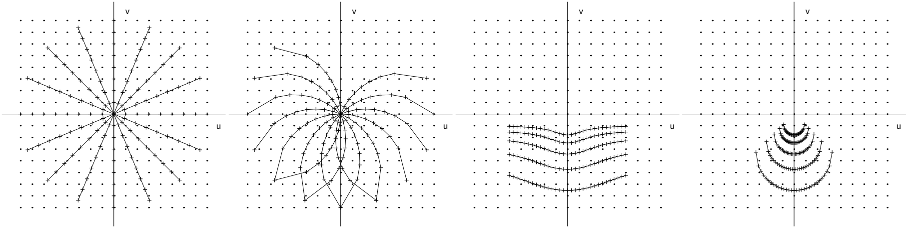
$$
\begin{aligned}
g(r, \phi) &= \int_{L_{r,\phi}} f(x, y) \, \mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x, y) \, \delta(r - x \cos\phi - y \sin\phi) \, \mathrm{d}x \, \mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem: Image reconstruction

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

Example 4: Fourier Synthesis in different imaging systems

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



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 $f(x, y)$

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

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

Maximum A Posteriori (MAP):

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

Posterior Mean (PM):

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

Supervised and Unsupervised Bayesian Inference

- 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) or Posterior Mean (PM) $\longrightarrow \widehat{\boldsymbol{f}}$

- Unsupervised Bayesian inference:
  Joint estimation of $\boldsymbol{f}$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$

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

Full Bayesian approach

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

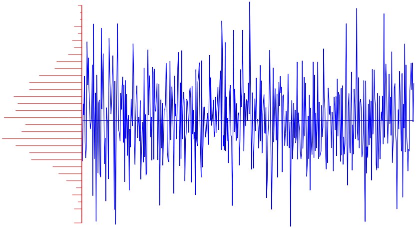
- Forward & errors model: $\longrightarrow p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1; \mathcal{M})$
- Prior models $\longrightarrow p(\boldsymbol{f}|\boldsymbol{\theta}_2; \mathcal{M})$
- Hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \longrightarrow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

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

Two main steps in the Bayesian approach

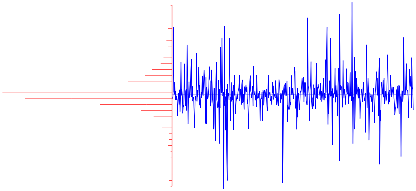
- Prior modeling: $p(\boldsymbol{f}|\boldsymbol{\theta}_1)$
  - Separable:
    Gaussian, Generalized Gaussian, Gamma,
    mixture of Gaussians, mixture of Gammas, ...
  - Markovian: Gauss-Markov, GGM, ...
  - Separable or Markovian with hidden variables
    (contours, region labels)
- Bayesian computational aspects
  - 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 Approximation (VBA)

Prior models: Separable $p(\boldsymbol{f}) = \prod_j p(f_j)$



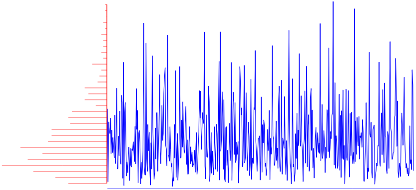
Gaussian

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



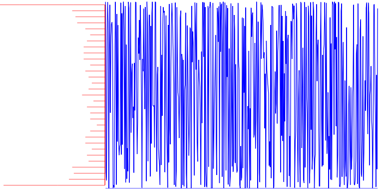
Generalized Gaussian

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



Gamma

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



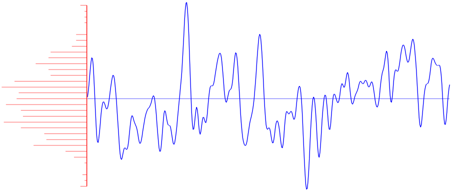
Beta

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

Prior models: Markovian models

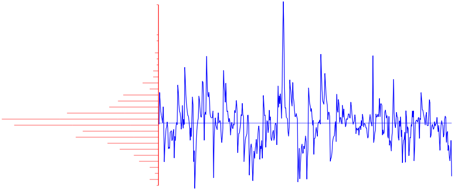
Gauss-Markov (G-M)

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_j |f_j - f_{j-1}|^2\right]$$



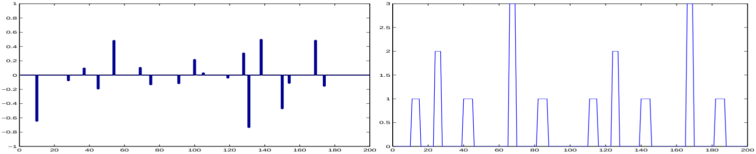
Generalized G-M

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_j |f_j - f_{j-1}|^\beta\right]$$

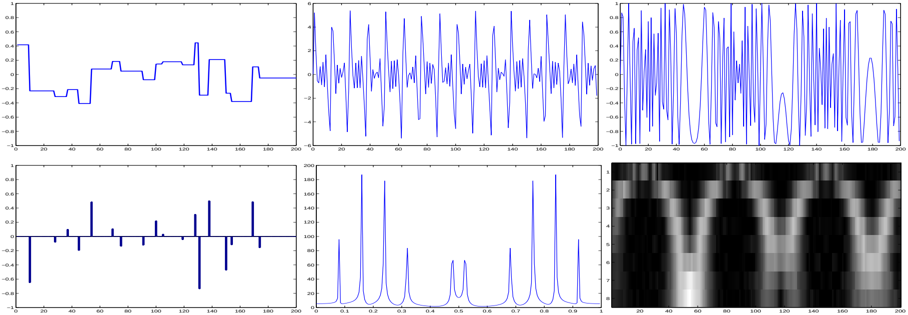


Sparsity enforcing prior models

- Sparse signals: Direct sparsity



- Sparse signals: Sparsity in a Transform domaine



Prior models: Sparsity enforcing 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 $\nu = 1$.

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

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

- Examples:

  Gaussian noise, Gaussian prior:

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

  Gaussian noise, Double Exponential prior:

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

- 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:

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

Marginal MAP and PM estimates

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

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

  and then:
  - MAP: $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})\right\}$
  - Posterior Mean: $\widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \,\mathrm{d}\boldsymbol{f}$
- EM and BEM Algorithms

- Variational Bayesian Approximation:
  Main idea: Approximate $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by a simpler, for example a separable one:

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

  and then continue computations with $q_1(\boldsymbol{f})$ and $q_2(\boldsymbol{\theta})$.

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 $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

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

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

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

VBA: 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

$$\left\{\begin{array}{lll}\widehat{q}_1(\boldsymbol{f}) & \text{in the same family than} & p(\boldsymbol{f}|\boldsymbol{\theta},\boldsymbol{g};\mathcal{M}) \\ \widehat{q}_2(\boldsymbol{\theta}) & \text{in the same family than} & p(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{g};\mathcal{M})\end{array}\right.$$

Variational Bayesian Approximation and model selection

- Joint posterior law:

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

where

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

- Approximate $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g},\mathcal{M})$ by a simpler $q(\boldsymbol{f},\boldsymbol{\theta})$ by minimizing

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

- Free energy:

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

- Link between the evidence of the model $\ln p(\boldsymbol{g}|\mathcal{M})$, $\mathsf{KL}(q:p)$ and $\mathcal{F}(q)$:

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

Variational Bayesian Approximation

- Alternate optimization of $\mathcal{F}(q)$ or $\text{KL}(q:p)$ with respect to q_1 and q_2 results in

$$\begin{cases} q_1(\boldsymbol{f}) & \propto \exp\left[-\langle \ln p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{g})\rangle_{q_2(\boldsymbol{\theta})}\right] \\ q_2(\boldsymbol{\theta}) & \propto \exp\left[-\langle \ln p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{g})\rangle_{q_1(\boldsymbol{f})}\right] \end{cases}$$

- $q_1 = \delta(\boldsymbol{f}-\widetilde{\boldsymbol{f}}), \quad q_2 = \delta(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}) \rightarrow$ JMAP:

$$\begin{cases} \widetilde{\boldsymbol{\theta}} & = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})\} \\ \widetilde{\boldsymbol{f}} & = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})\} \end{cases}$$

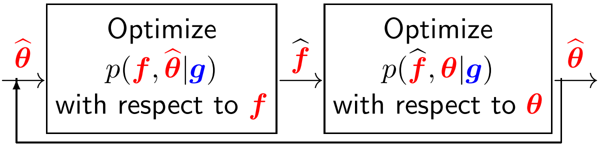
- q_1 free but $q_2 = \delta(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}) \leftarrow$ Expectation-Maximization (EM):

$$\begin{cases} \text{E stap:} & Q(\boldsymbol{\theta}) = \langle \ln p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{g})\rangle_{p(\boldsymbol{f}|\boldsymbol{g},\widetilde{\boldsymbol{\theta}})} \\ \text{M stap:} & \widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{Q(\boldsymbol{\theta})\} \end{cases}$$

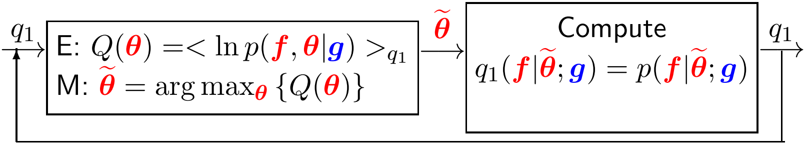
- If p is in the exponential family, then q will be too.

Comparison between JMAP, BEM and VBA

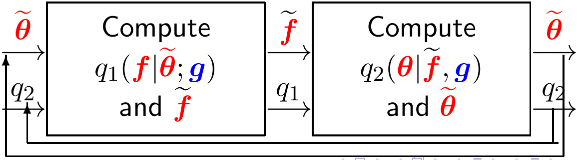
- JMAP:



- BEM:



- VBA:



Hierarchical models and hidden variables

- All the mixture models and some of simple models can be modeled via hidden variables $\boldsymbol{z}$.
- Example 1: Student-t model

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

- Example 2: MoG model:

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

- With these models we have:

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

Bayesian Variational Approximation

- Alternate optimization of $\mathcal{F}(q)$ with respect to q_1, q_2 and q_3 results in

$$
\begin{array}{rcl}
q(\boldsymbol{f}) & \propto & \exp\left[-\left\langle \ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{g})\right\rangle_{q(\boldsymbol{z})q(\boldsymbol{\theta})}\right] \\
q(\boldsymbol{z}) & \propto & \exp\left[-\left\langle \ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{g})\right\rangle_{q(\boldsymbol{f})q(\boldsymbol{\theta})}\right] \\
q(\boldsymbol{\theta}) & \propto & \exp\left[-\left\langle \ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{g})\right\rangle_{q(\boldsymbol{f})q(\boldsymbol{z})}\right]
\end{array}
$$

- If p is in the exponential family, then q will be too.
- Other separable decompositions:

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

$$q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) = \prod_j q_{1j}(f_j|z_j)\, q_{2j}(z_j) \prod_k q_{3k}(\boldsymbol{\theta}_k)$$

BVA with Student-t priors

Scale Mixture Model of Student-t:

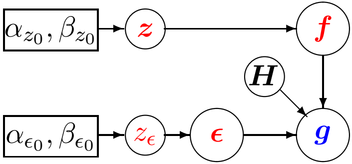
$$\mathcal{S}t(f_j|\nu) = \int_0^\infty \mathcal{N}(f_j|0, \frac{1}{z_j})\, \mathcal{G}(z_j|\nu/2, \nu/2)\, \mathrm{d}z_j$$

Hidden variables z_j:

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

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

- Graphical model:



BVA with Student-t priors Algorithm

Likelihood:

$$p(\boldsymbol{g}|\boldsymbol{f}, z_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \tfrac{1}{z_\epsilon}\boldsymbol{I})$$
$$p(z_\epsilon|\alpha_{z0}, \beta_{z0}) = \mathcal{G}(z_\epsilon|\alpha_{z0}, \beta_{z0})$$

Prior laws:

$$p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j \mathcal{N}\left(f_j|0, \tfrac{1}{z_j}\right)$$
$$p(\boldsymbol{z}|\alpha_0, \beta_0) = \prod_j \mathcal{G}(z_j|\alpha_0, \beta_0)$$

Joint posterior law:

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

Approximation by:

$$q(\boldsymbol{f}, z_\epsilon) = q_1(\boldsymbol{f}|z_\epsilon)\prod_j q_2(z_j)q_3(z_\epsilon)$$

$$q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}}) = \mathcal{N}(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$$
$$\widetilde{\boldsymbol{\mu}} = \langle\lambda\rangle_q \widetilde{\boldsymbol{\Sigma}}\boldsymbol{H}'\boldsymbol{g}$$
$$\widetilde{\boldsymbol{\Sigma}} = (\langle\lambda\rangle_q \boldsymbol{H}'\boldsymbol{H} + \widetilde{\boldsymbol{Z}})^{-1} \text{with} \quad \widetilde{\boldsymbol{Z}} = \operatorname{diag}[\widetilde{\boldsymbol{z}}]$$
$$q_{2j}(z_j) = \mathcal{G}(z_j|\widetilde{\alpha}_j, \widetilde{\beta}_j)$$
$$\widetilde{\alpha}_j = \alpha_{00} + \tfrac{1}{2}$$
$$\widetilde{\beta}_j = \beta_{00} + < f_j^2 > /2$$

$$q_3(z_\epsilon) = \mathcal{G}(z_\epsilon|\widetilde{\alpha}_{z_\epsilon}, \widetilde{\beta}_{z_\epsilon}),$$
$$\widetilde{\alpha}_{z_\epsilon} = \alpha_{z_0} + (n+1)/2$$
$$\widetilde{\beta}_{z_\epsilon} = \beta_{z_0} + \tfrac{1}{2}[\|\boldsymbol{g}\|^2$$
$$-2\langle\boldsymbol{f}\rangle_q' \boldsymbol{H}'\boldsymbol{g} + \boldsymbol{H}'\langle\boldsymbol{f}\boldsymbol{f}'\rangle_q \boldsymbol{H}]$$

$$< \boldsymbol{f} > = \widetilde{\boldsymbol{\mu}}, \quad < \boldsymbol{f}\boldsymbol{f}' > = \widetilde{\boldsymbol{\Sigma}} + \widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}}', \quad < f_j^2 > = [\widetilde{\boldsymbol{\Sigma}}]_{jj} + \widetilde{\mu}_j^2, \quad \widetilde{\lambda} = \frac{\widetilde{\alpha}_z}{\widetilde{\beta}_z}, \quad \tilde{z}_j = \frac{\widetilde{\alpha}_j}{\widetilde{\beta}_j}$$

$\widetilde{\lambda} \rightarrow$, $\widetilde{\boldsymbol{z}} \rightarrow$

$$q_1(\boldsymbol{f}|\widetilde{\boldsymbol{z}}, \widetilde{\lambda}) = \mathcal{N}(\widetilde{\boldsymbol{f}}, \widetilde{\boldsymbol{\Sigma}})$$
$$\widetilde{\boldsymbol{f}} = \widetilde{\lambda}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{H}'\boldsymbol{g}$$
$$\widetilde{\boldsymbol{\Sigma}} = (\widetilde{\lambda}\boldsymbol{H}'\boldsymbol{H} + \widetilde{\boldsymbol{z}}^{-1})^{-1}$$

$\widetilde{\boldsymbol{f}} \rightarrow$, $\widetilde{\boldsymbol{\Sigma}} \rightarrow$

$$q_{2j}(z_j|\widetilde{\boldsymbol{f}}) = \mathcal{G}(z_j|\widetilde{\alpha}_j, \widetilde{\beta}_j)$$
$$\widetilde{\alpha}_j = \alpha_{00} + \frac{n+1}{2}$$
$$\widetilde{\beta}_j = \beta_{00} + \frac{1}{2}\left\langle f_j^2\right\rangle_q$$
$$\tilde{z}_j = \widetilde{\alpha}_j/\widetilde{\beta}_j$$

$\widetilde{\boldsymbol{f}} \rightarrow$, $\widetilde{\boldsymbol{\Sigma}} \rightarrow$, $\tilde{z}_j \rightarrow$

$$q_3(z|\widetilde{\boldsymbol{f}}) = \mathcal{G}(z|\widetilde{\alpha}_z, \widetilde{\beta}_z)$$
$$\widetilde{\alpha}_z = \alpha_{z_0} + \frac{n+1}{2}$$
$$\widetilde{\beta}_z = \beta_{z_0} + \frac{1}{2}[\|\boldsymbol{g}\|^2$$
$$-2 < \boldsymbol{f} >' \boldsymbol{H}'\boldsymbol{g} + \boldsymbol{H}' < \boldsymbol{f}\boldsymbol{f}' > \boldsymbol{H}]$$
$$\widetilde{\lambda} = \widetilde{\alpha}_z/\widetilde{\beta}_z$$

$\widetilde{\lambda} \rightarrow$, $\widetilde{\boldsymbol{z}} \rightarrow$

Implementation issues

- To compute $\boldsymbol{f}$, at each iteration, a gradient based optimization algorithm is used. This step is a common step with all the methods. The criterion to optimize is often quadratic:

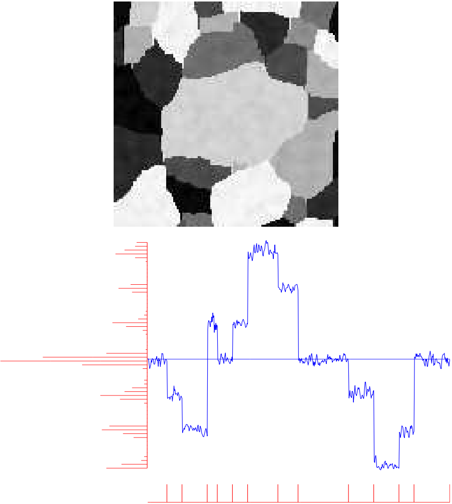
$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2$$

  and its gradient is

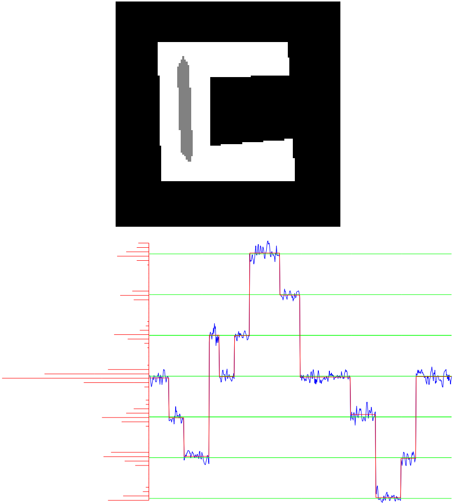
$$\nabla J(\boldsymbol{f}) = 2\boldsymbol{H}'(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + 2\lambda\boldsymbol{D}'\boldsymbol{D}\boldsymbol{f}$$

- 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=forward(f,...)`
  - Adjoint operator : $\boldsymbol{H}'\boldsymbol{g} \longrightarrow \boldsymbol{f}$ `f=adjoint(g,...)`
- For any particular application, we have to write specific programs (`forward & adjoint`).
- Often the main computational cost is in these two programs. being Bayesian does not cost much more
- The main computational cost for BEM and VBA is the computation and inversion of the covariances.
- Complete factorization may reduce this to the computation of

Two other Hierarchical models

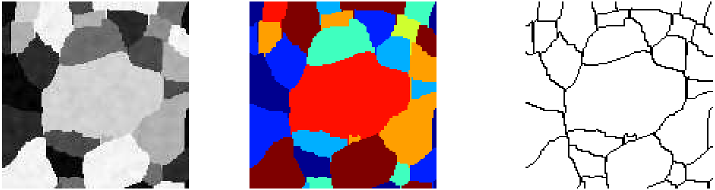


Contours: $p(f_j|(1-q_j)f_{j-1})$



Regions: $p(f_j|z_j = k) = \mathcal{N}(m_k, v_k)$

Gauss-Markov-Potts prior models for images



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

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

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

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

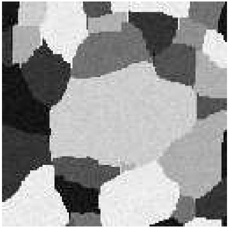
$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left[\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]$$

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

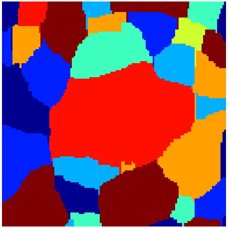
Four different cases

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)



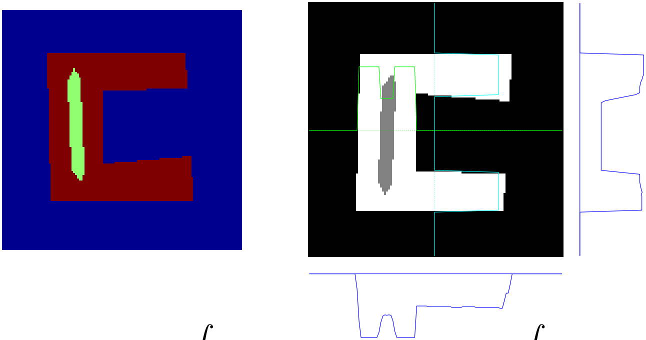
$f(\boldsymbol{r})$



$z(\boldsymbol{r})$

Application of CT in NDT

Reconstruction from only 2 projections

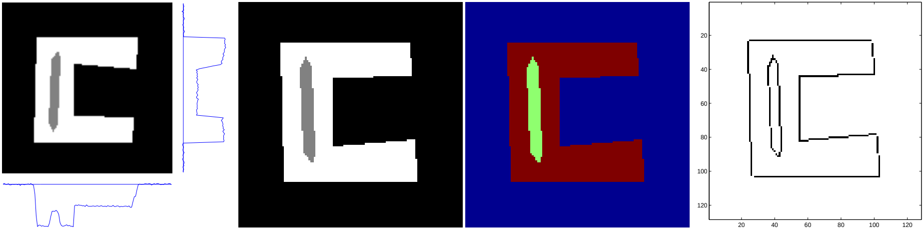


$$g_1(x) = \int f(x,y) \, \mathrm{d}y, \qquad g_2(y) = \int f(x,y) \, \mathrm{d}x$$

- 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:

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

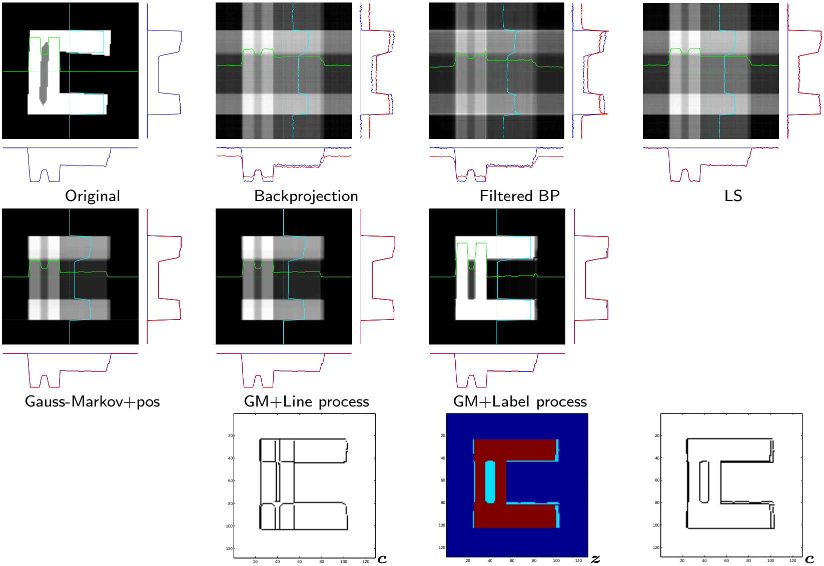
Application in CT



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

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

Results



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

Conclusions

- Inverse problems arise in many signal and image processing applications.
- Two main steps in the Bayesian approach: Prior modeling and Bayesian computation
- Different prior models: Separable, Markovian and hierarchical models
- Different Bayesian computational tools: Gaussian approximation, MCMC and VBA
- We use these models for inverse problems in different signal and image processing applications such as:
  - Spectral and periodical components estimation in biological time series
  - X ray Computed Tomography,
  - Signal deconvolution in Proteomic and molecular imaging
  - Diffraction Optical Tomography
  - Microwave Imaging, Acoustic imaging and sources localization
  - Synthetic Aperture Radar (SAR) Imaging

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).

2. S. Zhu, A. Mohammad-Djafari, H. Wang, B. Deng, X. Li and J. Mao J, “Parameter estimation for SAR micromotion target based on sparse signal representation,” EURASIP Journal on Advances in Signal Processing, vol. Special issue on Sparse Signal Processing, (2012).

3. 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

4. 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: