Two main steps in Bayesian approach: Prior modeling and Bayesian computation

Ali Mohammad-Djafari

Groupe Problèmes Inverses
Laboratoire des Signaux et Systèmes
UMR 8506 CNRS - SUPELEC - Univ Paris Sud 11
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, FRANCE.

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

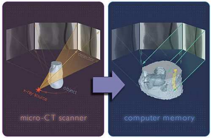
MaxEnt2009, 5-10 July 2009, Oxford, Mississippi, USA

Content

- Image reconstruction in Computed Tomography :
  An ill posed invers problem
- Two main steps in Bayesian approach :
  Prior modeling and Bayesian computation
- Prior models for images :
  - Separable Gaussian, GG, ...
  - Gauss-Markov, General one layer Markovian models
  - Hierarchical Markovian models with hidden variables (contours and regions)
  - Gauss-Markov-Potts
- Bayesian computation
  - MCMC
  - Variational and Mean Field approximations (VBA, MFA)
- Application : Computed Tomography in NDT
- Conclusions and Work in Progress
- Questions and Discussion

Computed Tomography :
Making an image of the interior of a body

- $f(x,y)$ a section of a real 3D body $f(x,y,z)$
- $g_\phi(r)$ a line of observed radiographe $g_\phi(r,z)$
- Forward model :
  Line integrals or Radon Transform



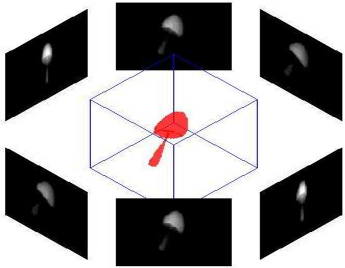
$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x,y)\,\mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem : Image reconstruction

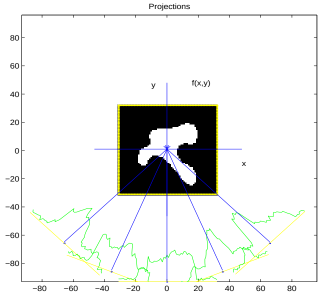
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x,y)$

2D and 3D Computed Tomography

3D



2D

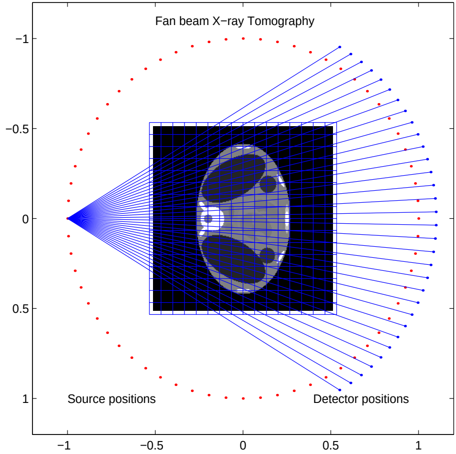


$$g_\phi(r_1,r_2) = \int_{\mathcal{L}_{r_1,r_2,\phi}} f(x,y,z)\, \mathrm{d}l \qquad g_\phi(r) = \int_{\mathcal{L}_{r,\phi}} f(x,y)\, \mathrm{d}l$$

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 $f(x,y,z)$

CT as a linear inverse problem



$$g(s_i) = \int_{L_i} f(\boldsymbol{r}) \, \mathrm{d}l_i + \epsilon(s_i) \longrightarrow \text{Discretization} \longrightarrow \boldsymbol{g} = \boldsymbol{H} \boldsymbol{f} + \boldsymbol{\epsilon}$$

- $\boldsymbol{g}$, $\boldsymbol{f}$ and $\boldsymbol{H}$ are huge dimensional

Inversion : Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

Functional space (Tikhonov) :

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

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

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

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

or

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

Limitations :

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

Bayesian estimation approach

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

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

Link with regularization :

Maximum A Posteriori (MAP) :

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

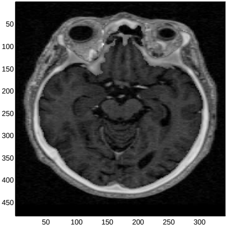
with $\quad Q(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) = -\ln p(\boldsymbol{g}|\boldsymbol{f}) \quad$ and $\quad \lambda\Omega(\boldsymbol{f}) = -\ln p(\boldsymbol{f})$

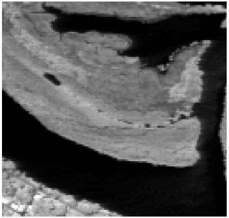
But, Bayesian inference is not only limited to MAP

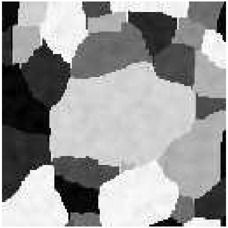
Two main steps in the Bayesian approach

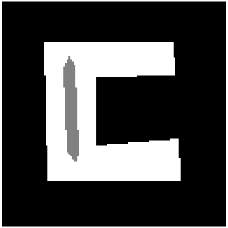
- Prior modeling
  - Separable :
    Gaussian, Generalized Gaussian, Gamma,
    mixture of Gaussians, mixture of Gammas, ...
  - Markovian : Gauss-Markov, GGM, ...
  - Separable or Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects
  - MAP,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

Which images I am looking for ?

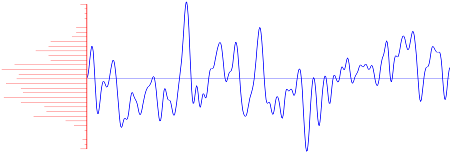






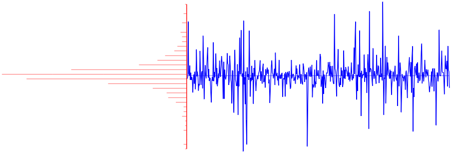


Which image I am looking for ?



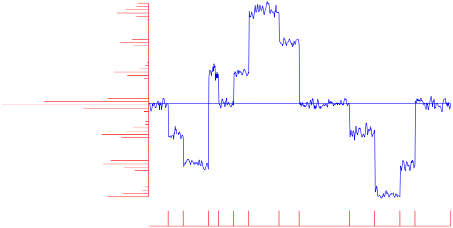
Gaussian

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



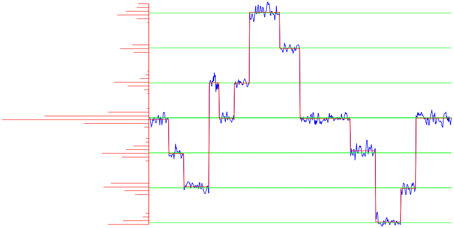
Generalized Gaussian

$p(f_j|f_{j-1}) \propto \exp\left\{-\alpha|f_j - f_{j-1}|^p\right\}$



Piecewize Gaussian

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



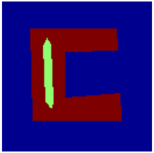
Mixture of GM

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

Gauss-Markov-Potts prior models for images

"In NDT applications of CT, the objects are, in general, composed of a finite number of materials, and the voxels corresponding to each materials are grouped in compact regions"

How to model this prior information ?



$f(\boldsymbol{r})$ $\qquad$ $z(\boldsymbol{r}) \in \{1, ..., K\}$

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

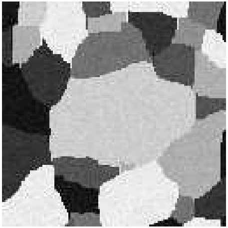
$$p(f(\boldsymbol{r})) = \sum_k P(z(\boldsymbol{r}) = k) \mathcal{N}(m_k, v_k) \quad \text{Mixture of Gaussians}$$

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

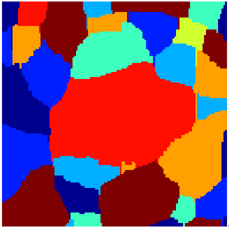
Four different cases

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)

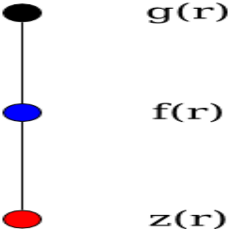


$f(\boldsymbol{r})$

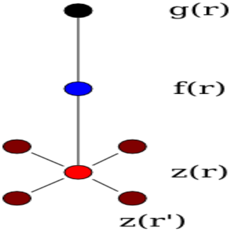


$z(\boldsymbol{r})$

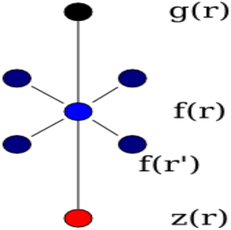
Four different cases



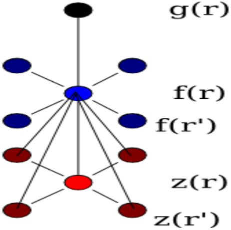
Case 1 : Mixture of Gaussians



Case 2 : Mixture of Gauss-Markov

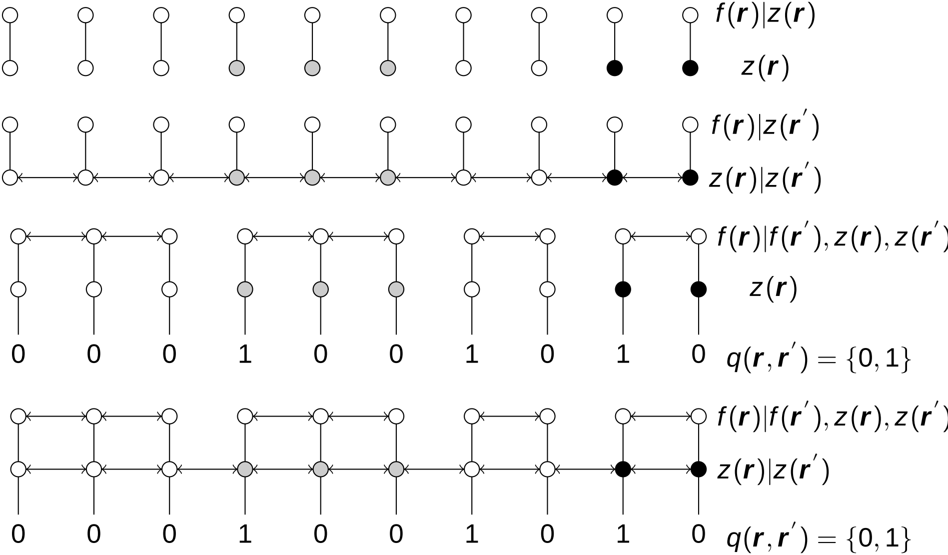


Case 3 : MIG with Hidden Potts



Case 4 : MGM with hidden Potts

Four different cases



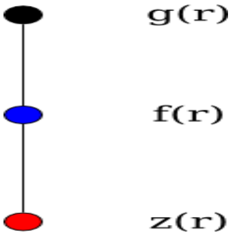
Case 1 : $f|z$ Gaussian iid, z iid

Independent Mixture of Independent Gaussiens (IMIG) :

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

$$p(f(\boldsymbol{r}))=\sum_{k=1}^{K}\alpha_k\mathcal{N}(m_k,v_k), \text{with } \sum_k\alpha_k=1.$$

$$p(\boldsymbol{z})=\prod_{\boldsymbol{r}}p(z(\boldsymbol{r})=k)=\prod_{\boldsymbol{r}}\alpha_k=\prod_k\alpha_k^{n_k}$$



Noting $\quad \mathcal{R}_k=\{\boldsymbol{r}:z(\boldsymbol{r})=k\}, \quad \mathcal{R}=\cup_k\mathcal{R}_k,$

$$m_z(\boldsymbol{r})=m_k, v_z(\boldsymbol{r})=v_k, \alpha_z(\boldsymbol{r})=\alpha_k, \forall\boldsymbol{r}\in\mathcal{R}_k$$

we have :

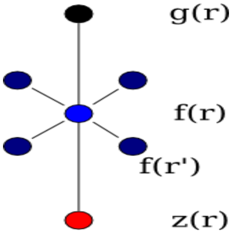
$$p(\boldsymbol{f}|\boldsymbol{z}) = \prod_{\boldsymbol{r}\in\mathcal{R}}\mathcal{N}(m_z(\boldsymbol{r}),v_z(\boldsymbol{r}))$$

$$p(\boldsymbol{z}) = \prod_{\boldsymbol{r}}\alpha_z(\boldsymbol{r})=\prod_k\alpha_k^{\sum_{\boldsymbol{r}\in\mathcal{R}}\delta(z(\boldsymbol{r})-k)}=\prod_k\alpha_k^{n_k}$$

Case 2 : $f|z$ Gauss-Markov, z iid

Independent Mixture
of Gauss-Markov (IMGM) :

$$p(f(\boldsymbol{r})|z(\boldsymbol{r}), z(\boldsymbol{r}'), f(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r}))$$

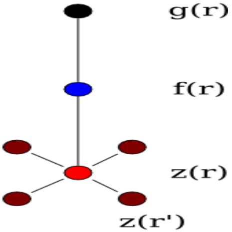


$$\begin{aligned} &= \mathcal{N}(\mu_z(\boldsymbol{r}), v_z(\boldsymbol{r})), \forall \boldsymbol{r} \in \mathcal{R} \\ \mu_z(\boldsymbol{r}) &= \frac{1}{|\mathcal{V}(\boldsymbol{r})|} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \mu_z^*(\boldsymbol{r}') \\ \mu_z^*(\boldsymbol{r}') &= \delta(z(\boldsymbol{r}') - z(\boldsymbol{r}))\, f(\boldsymbol{r}') + (1 - \delta(z(\boldsymbol{r}') - z(\boldsymbol{r}))\, m_z(\boldsymbol{r}') \\ &= \quad (1 - c(\boldsymbol{r}')) \quad f(\boldsymbol{r}') \; + \qquad c(\boldsymbol{r}') \qquad m_z(\boldsymbol{r}') \end{aligned}$$

$$\begin{aligned} p(\boldsymbol{f}|\boldsymbol{z}) &\propto \prod_{\boldsymbol{r}} \mathcal{N}(\mu_z(\boldsymbol{r}), v_z(\boldsymbol{r})) &&\propto \prod_k \alpha_k \mathcal{N}(m_k 1, \boldsymbol{\Sigma}_k) \\ p(\boldsymbol{z}) &= \prod_{\boldsymbol{r}} v_z(\boldsymbol{r}) &&= \prod_k \alpha_k^{n_k} \end{aligned}$$

with $1_k = 1, \forall \boldsymbol{r} \in \mathcal{R}_k$ and $\boldsymbol{\Sigma}_k$ a covariance matrix $(n_k \times n_k)$.

Case 3 : $f|z$ Gauss iid, z Potts



Gauss iid as in Case 1 :

$$p(\boldsymbol{f}|\boldsymbol{z}) = \prod_{\boldsymbol{r}\in\mathcal{R}} \mathcal{N}(m_z(\boldsymbol{r}), v_z(\boldsymbol{r})) = \prod_k \prod_{\boldsymbol{r}\in\mathcal{R}_k} \mathcal{N}(m_k, v_k)$$

Potts-Markov :

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

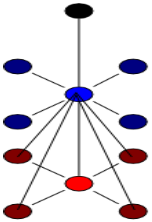
$$p(\boldsymbol{z}) \propto \exp\left\{\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

Case 4 : $f|z$ Gauss-Markov, z Potts

Gauss-Markov as in Case 2 :

$p(f(\boldsymbol{r})|z(\boldsymbol{r}), z(\boldsymbol{r}'), f(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) =$

$\mathcal{N}(\mu_z(\boldsymbol{r}), v_z(\boldsymbol{r})), \forall \boldsymbol{r} \in \mathcal{R}$



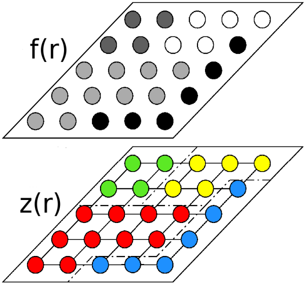
$$\begin{aligned} \mu_z(\boldsymbol{r}) &= \frac{1}{|\mathcal{V}(\boldsymbol{r})|} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \mu_z^*(\boldsymbol{r}') \\ \mu_z^*(\boldsymbol{r}') &= \delta(z(\boldsymbol{r}') - z(\boldsymbol{r}))\, f(\boldsymbol{r}') + (1 - \delta(z(\boldsymbol{r}') - z(\boldsymbol{r}))\, m_z(\boldsymbol{r}') \end{aligned}$$

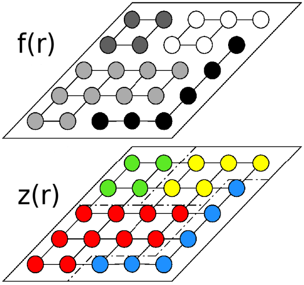
$$p(\boldsymbol{f}|\boldsymbol{z}) \quad \propto \prod_{\boldsymbol{r}} \mathcal{N}(\mu_z(\boldsymbol{r}), v_z(\boldsymbol{r})) \propto \prod_k \alpha_k \mathcal{N}(m_k 1, \boldsymbol{\Sigma}_k)$$

Potts-Markov as in Case 3 :

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

Summary of the two proposed models





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

(MIG with Hidden Potts)

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

(MGM with hidden Potts)

Bayesian Computation

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

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

- Direct computation and use of $p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g};\mathcal{M})$ is too complex
- Possible approximations :
  - Gauss-Laplace (Gaussian approximation)
  - Exploration (Sampling) using MCMC methods
  - Separable approximation (Variational techniques)
- Main idea in Variational Bayesian methods :
  Approximate
  $p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g};\mathcal{M})$ by $q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$
  - Choice of approximation criterion : $KL(q : p)$
  - Choice of appropriate families of probability laws for $q_1(\boldsymbol{f})$, $q_2(\boldsymbol{z})$ and $q_3(\boldsymbol{\theta})$

MCMC based algorithm

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

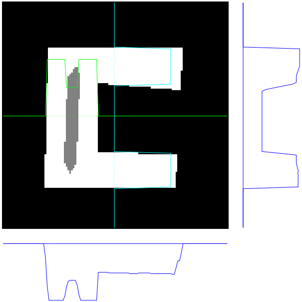
General scheme :

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

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

Application of CT in NDT

Reconstruction from only 2 projections



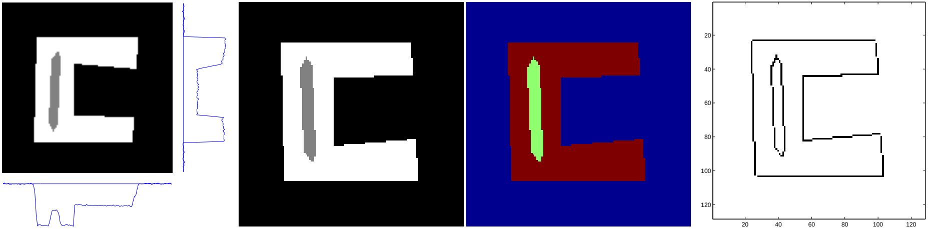
$$g_1(x) = \int f(x,y)\, dy$$

$$g_2(y) = \int f(x,y)\, dx$$

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

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

Application in CT



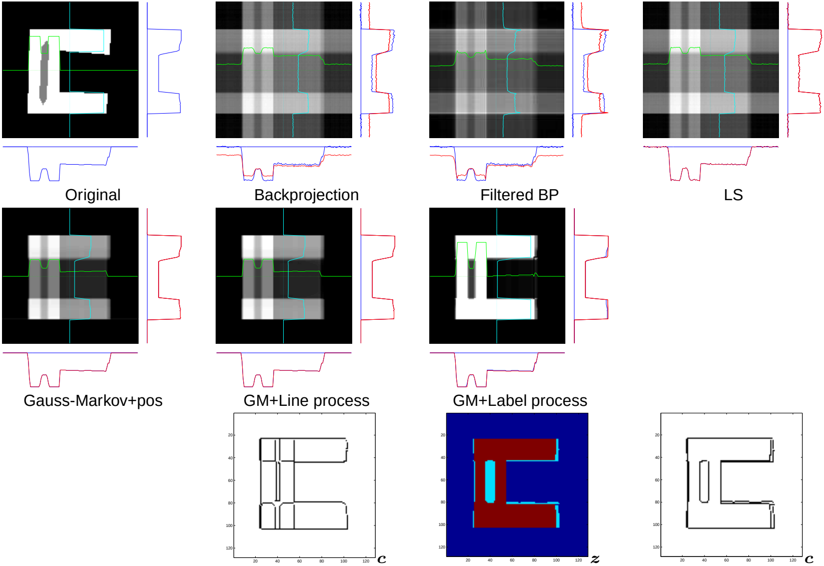
$g \mid f$	$f \mid z$	z	q
$g = Hf + \epsilon$	iid Gaussian	iid	$q(r) \in \{0, 1\}$
$g \mid f \sim \mathcal{N}(Hf, \sigma_\epsilon^2 I)$	or	or	$1 - \delta(z(r) - z(r'))$
Gaussian	Gauss-Markov	Potts	binary

Forward model | Gauss-Markov-Potts Prior Model | Auxilary

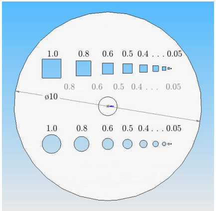
Unsupervised Bayesian estimation :

$$p(f, z, \theta \mid g) \propto p(g \mid f, z, \theta)\, p(f \mid z, \theta)\, p(\theta)$$

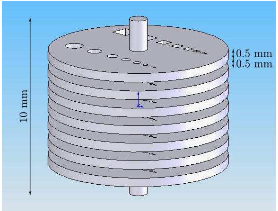
Results : 2D case



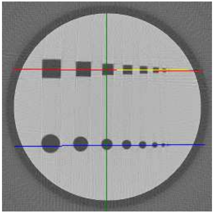
Some results in 3D case



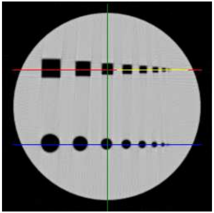
M. Defrise



Phantom

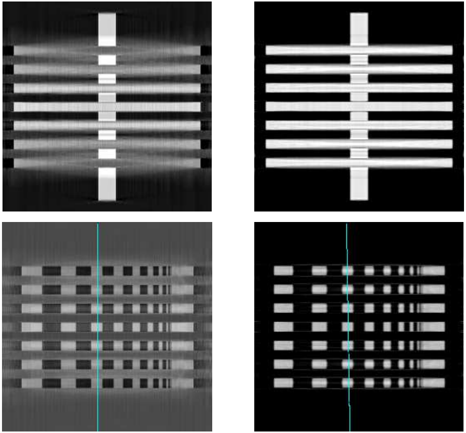


FeldKamp



Proposed method

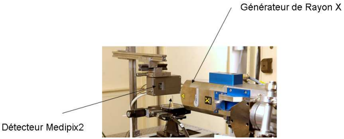
Some results in 3D case



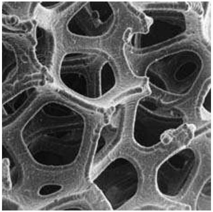
FeldKamp Proposed method

Some results in 3D case

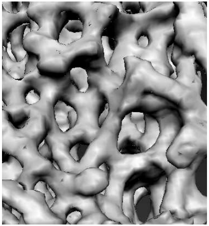
Experimental setup

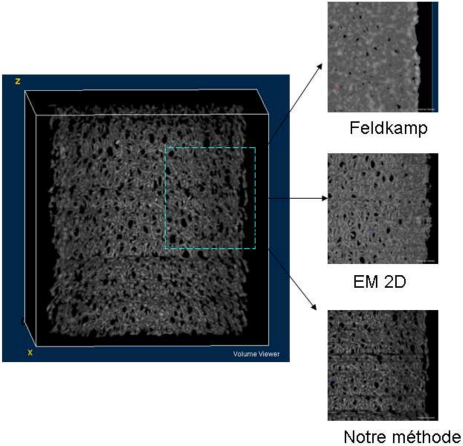


A photograpy of metalique esponge

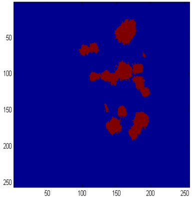


Reconstruction by proposed metho

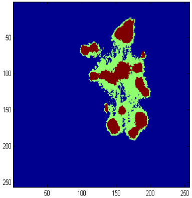




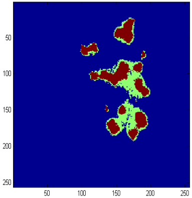
Application : liquid evaporation in metalic esponge



Time 0



Time 1



Time 2

Conclusions

- Gauss-Markov-Potts are useful prior models for images incorporating regions and contours
- Bayesian computation needs often pproximations (Laplace, MCMC, Variational Bayes)
- Application in different CT systems (X ray, Ultrasound, Microwave, PET, SPECT) as well as other inverse problems

Work in Progress and Perspectives :

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