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 radiographs \( g_\phi(r, z) \)

Forward model:
Line integrals or Radon Transform

\[
g_\phi(r) = \int_{L_{r,\phi}} f(x, y) \, dl + \epsilon_\phi(r) \\
= \int \int f(x, y) \, \delta(r - x \cos \phi - y \sin \phi) \, dx \, dy + \epsilon_\phi(r)
\]

Inverse problem: Image reconstruction

Given the forward model \( \mathcal{H} \) (Radon Transform) and a set of data \( g_{\phi_i}(r), i = 1, \ldots, M \) find \( f(x, y) \)
2D and 3D Computed Tomography

\[ g_\phi(r_1, r_2) = \int_{\mathcal{L}_{r_1, r_2, \phi}} f(x, y, z) \, dl \quad g_\phi(r) = \int_{\mathcal{L}_{r, \phi}} f(x, y) \, dl \]

Forward problem: \( f(x, y) \) or \( f(x, y, z) \) \( \rightarrow \) \( g_\phi(r) \) or \( g_\phi(r_1, r_2) \)

Inverse problem: \( g_\phi(r) \) or \( g_\phi(r_1, r_2) \) \( \rightarrow \) \( f(x, y) \) or \( f(x, y, z) \)
CT as a linear inverse problem

\[ g(s_i) = \int_{L_i} f(r) \, dl_i + \epsilon(s_i) \longrightarrow \text{Discretization} \longrightarrow g = Hf + \epsilon \]

- \text{\( g, f \) and \( H \) are huge dimensional}
Inversion : Regularization theory

Inverse problems = Ill posed problems

→ Need for prior information

Functional space (Tikhonov) :

\[ g = \mathcal{H}(f) + \epsilon \rightarrow J(f) = \| g - \mathcal{H}(f) \|_2^2 + \lambda \| Df \|_2^2 \]

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

• Minimum norme LS (MNLS) : \( J(f) = \| g - H(f) \|_2^2 + \lambda \| f \|_2^2 \)

• Classical regularization : \( J(f) = \| g - H(f) \|_2^2 + \lambda \| Df \|_2^2 \)

• More general regularization :

\[ J(f) = Q(g - H(f)) + \lambda \Omega(Df) \]

or

\[ J(f) = \Delta_1(g, H(f)) + \lambda \Delta_2(f, f\infty) \]

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

\[ M : \quad g = H f + \epsilon \]

- Observation model \( M \) + Hypothesis on the noise \( \epsilon \)
  \[ p(g \mid f; M) = p_\epsilon(g - H f) \]
- A priori information \( p(f \mid M) \)
- Bayes :
  \[ p(f \mid g; M) = \frac{p(g \mid f; M) p(f \mid M)}{p(g \mid M)} \]

Link with regularization:

Maximum A Posteriori (MAP) :

\[ \hat{f} = \arg \max_f \{ p(f \mid g) \} = \arg \max_f \{ p(g \mid f) \ p(f) \} \]
\[ = \arg \min_f \{ - \ln p(g \mid f) - \ln p(f) \} \]

with \( Q(g, H f) = - \ln p(g \mid f) \) and \( \lambda \Omega(f) = - \ln p(f) \)

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

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

Piecewise 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?

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

\[
p(f(r)) = \sum_k P(z(r) = k) \mathcal{N}(m_k, v_k)
\]

Mixture of Gaussians

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

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

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

\[
\begin{align*}
f(r) & | z(r) \\
f(r) & | z(r') \\
d(f(r^2), z(r), z(r')) & z(r) \\
d(f(r^2), z(r), z(r')) & z(r) | z(r')
\end{align*}
\]
Case 1: \( f \mid z \) Gaussian iid, \( z \) iid

Independent Mixture of Independent Gaussians (IMIG):

\[
p(f(r) \mid z(r) = k) = \mathcal{N}(m_k, v_k), \quad \forall r \in \mathcal{R}
\]

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

\[
p(z) = \prod_r p(z(r) = k) = \prod_r \alpha_k = \prod_k \alpha_k^{n_k}
\]

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

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

we have:

\[
p(f \mid z) = \prod_{r \in \mathcal{R}} \mathcal{N}(m_z(r), v_z(r))
\]

\[
p(z) = \prod_r \alpha_z(r) = \prod_k \alpha_k^{\sum_{r \in \mathcal{R}} \delta(z(r) - k)} = \prod_k \alpha_k^{n_k}
\]
Case 2: \( f \mid \sim \text{Gauss-Markov}, \quad \sim \text{iid} \)

Independent Mixture of Gauss-Markov (IMGM):

\[
p(f(r) \mid z(r), z(r'), f(r'), r' \in \mathcal{V}(r))
= \mathcal{N}(\mu_z(r), \nu_z(r)), \forall r \in \mathcal{R}
\]
\[
\mu_z(r) = \frac{1}{|\mathcal{V}(r)|} \sum_{r' \in \mathcal{V}(r)} \mu_z^*(r')
\]
\[
\mu_z^*(r') = \delta(z(r') - z(r)) f(r') + (1 - \delta(z(r') - z(r)) m_z(r')
\]
\[
\rho(f \mid z) \propto \prod_r \mathcal{N}(\mu_z(r), \nu_z(r)) \propto \prod_k \alpha_k \mathcal{N}(m_k 1, \Sigma_k)
\]
\[
\rho(z) = \prod_r \nu_z(r) = \prod_k \alpha_k^{n_k}
\]

with \( 1_k = 1, \forall r \in \mathcal{R}_k \) and \( \Sigma_k \) a covariance matrix \((n_k \times n_k)\).
Case 3 : \( f | z \) Gauss iid, \( z \) Potts

Gauss iid as in Case 1:

\[
p(f | z) = \prod_{r \in R} \mathcal{N}(m_z(r), v_z(r)) = \prod_k \prod_{r \in R_k} \mathcal{N}(m_k, v_k)
\]

Potts-Markov:

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

\[
p(z) \propto \exp \left\{ \gamma \sum_{r \in R} \sum_{r' \in V(r)} \delta(z(r) - z(r')) \right\}
\]
Case 4: $f \mid z$ Gauss-Markov, $z$ Potts

Gauss-Markov as in Case 2:

$$p(f(r) \mid z(r), z(r'), f(r'), r' \in \mathcal{V}(r)) = \mathcal{N}(\mu_z(r), \nu_z(r)), \forall r \in \mathcal{R}$$

$$\mu_z(r) = \frac{1}{|\mathcal{V}(r)|} \sum_{r' \in \mathcal{V}(r)} \mu_z^*(r')$$

$$\mu_z^*(r') = \delta(z(r') - z(r)) f(r') + (1 - \delta(z(r') - z(r))) m_z(r')$$

$$p(f \mid z) \propto \prod_r \mathcal{N}(\mu_z(r), \nu_z(r)) \propto \prod_k \alpha_k \mathcal{N}(m_k 1, \Sigma_k)$$

Potts-Markov as in Case 3:

$$p(z) \propto \exp \left\{ \gamma \sum_{r \in \mathcal{R}} \sum_{r' \in \mathcal{V}(r)} \delta(z(r) - z(r')) \right\}$$
Summary of the two proposed models

- \( f | \sim \) Gaussian iid \\
  \( z \sim \) Potts-Markov

  (MIG with Hidden Potts)

- \( f | \sim \) Markov \\
  \( z \sim \) Potts-Markov

  (MGM with hidden Potts)
Bayesian Computation

\[ p(f, z, \theta | g) \propto p(g | f, z, \nu_\varepsilon) p(f | z, m, v) p(z | \gamma, \alpha) p(\theta) \]

\[ \theta = \{ \nu_\varepsilon, (\alpha_k, m_k, \nu_k), k = 1, \ldots, K \} \]

- Direct computation and use of \( p(f, z, \theta | g; \mathcal{M}) \) 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
  \[ p(f, z, \theta | g; \mathcal{M}) \]
  by
  \[ q(f, z, \theta) = q_1(f) q_2(z) q_3(\theta) \]

  - Choice of approximation criterion: \( KL(q : p) \)
  - Choice of appropriate families of probability laws
    for \( q_1(f) \), \( q_2(z) \) and \( q_3(\theta) \)
MCMC based algorithm

\[ p(f, z, \theta | g) \propto p(g | f, z, \theta) p(f | z, \theta) p(z) p(\theta) \]

General scheme :

\[ \hat{f} \sim p(f | \hat{z}, \hat{\theta}, g) \longrightarrow \hat{z} \sim p(z | \hat{f}, \hat{\theta}, g) \longrightarrow \hat{\theta} \sim (\theta | \hat{f}, \hat{z}, g) \]

- Sample \( f \) from \( p(f | \hat{z}, \hat{\theta}, g) \propto p(g | f, \theta) p(f | \hat{z}, \hat{\theta}) \)
  Needs optimisation of a quadratic criterion.

- Sample \( z \) from \( p(z | \hat{f}, \hat{\theta}, g) \propto p(g | \hat{f}, \hat{z}, \hat{\theta}) p(z) \)
  Needs sampling of a Potts Markov field.

- Sample \( \theta \) from
  \[ p(\theta | \hat{f}, \hat{z}, g) \propto p(g | \hat{f}, \sigma^2 \epsilon I) p(f | \hat{z}, (m_k, v_k)) p(\theta) \]
  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

\[
g|f = Hf + \epsilon \\
g|f \sim \mathcal{N}(Hf, \sigma^2 \epsilon I)
\]

Gaussian

\[
f|z \quad \text{iid Gaussian}
\]

Gauss-Markov

\[
z \quad \text{iid}
\]

Potts

\[
q(r) \in \{0, 1\} \quad \text{binary}
\]

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

Unsupervised Bayesian estimation:

\[
p(f, z, \theta|g) \propto p(g|f, z, \theta) p(f|z, \theta) p(\theta)
\]
Results: 2D case

Original

Backprojection

Filtered BP

LS

Gauss-Markov+pos

GM+Line process

GM+Label process
Some results in 3D case

M. Defrise

Phantom

FeldKamp

Proposed method
Some results in 3D case

FeldKamp

Proposed method
Some results in 3D case

A photography of metalique esponge

Reconstruction by proposed method

Experimental setup

Génrateur de Rayon X

Détecteur Medipix2

Feldkamp

EM 2D

Notre méthode
Application: liquid evaporation in metallic sponge

Time 0

Time 1

Time 2
Conclusions

- Gauss-Markov-Potts are useful prior models for images incorporating regions and contours
- Bayesian computation needs often approximations (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)