Gauss-Markov-Potts Priors for Inverse Problems in Imaging Systems

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

Journée Problèmes Inverses et Optimisation de Forme,
13/12/2007
Laboratoire de mathématiques Jean Leray, Nantes, france
Content

- Computed Tomography (CT) as an Invers Problem example
- Classical methods: analytical and algebraic method
- Probabilistic methods
- Bayesian inference approach
- Gauss-Markov-Potts prior models for images
- Bayesian computation
- VB with Gauss-Markov-Potts prior models
- Application in Computed Tomography
- Conclusions
- Questions and Discussion
2D and 3D Computed Tomography

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

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 \( f(x, y, z) \)
X ray Tomography and Radon Transform

\[ g(r, \phi) = -\ln \left( \frac{l}{l_0} \right) = \int_{L_{r,\phi}} f(x, y) \, dl \]

\[ g(r, \phi) = \iint_{D} f(x, y) \delta(r - x \cos \phi - y \sin \phi) \, dx \, dy \]

\[ f(x, y) \xrightarrow{RT} g(r, \phi) \]

IRT
Analytical Inversion methods

\[ g(r, \phi) = \int_D f(x, y) \delta(r - x \cos \phi - y \sin \phi) \, dx \, dy \]

\[ f(x, y) = \left( -\frac{1}{2\pi^2} \right) \int_0^\pi \int_{-\infty}^{+\infty} \frac{\partial}{\partial r} g(r, \phi) \frac{dr}{(r - x \cos \phi - y \sin \phi)} \, dr \, d\phi \]
Filtered Backprojection method

\[ f(x, y) = \left( -\frac{1}{2\pi^2} \right) \int_0^\pi \int_{-\infty}^{+\infty} \frac{\partial g(r, \phi)}{dr} \frac{g(r, \phi)}{(r - x \cos \phi - y \sin \phi)} \, dr \, d\phi \]

Derivation \( \mathcal{D} \):
\[ \overline{g}(r, \phi) = \frac{\partial g(r, \phi)}{\partial r} \]

Hilbert Transform \( \mathcal{H} \):
\[ g_1(r', \phi) = \frac{1}{\pi} \int_0^\infty \overline{g}(r, \phi) \frac{g(r, \phi)}{(r - r')} \, dr \]

Backprojection \( \mathcal{B} \):
\[ f(x, y) = \frac{1}{2\pi} \int_0^\pi g_1(r' = x \cos \phi + y \sin \phi, \phi) \, d\phi \]

\[ f(x, y) = \mathcal{B} \mathcal{H} \mathcal{D} g(r, \phi) = \mathcal{B} \mathcal{F}_1^{-1} |\Omega| \mathcal{F}_1 g(r, \phi) \]

- Backprojection of filtered projections:

\[ \begin{array}{c|c|c|c|c}
   g(r, \phi) & \text{FT} & \text{Filter} & \text{IFT} & f(x, y) \\
   \mathcal{F}_1 & |\Omega| & \mathcal{F}_1^{-1} & g_1(r, \phi) & \mathcal{B} \\
\end{array} \]
Limitations: Limited angle or noisy data

- Limited angle or noisy data
- Accounting for detector size
- Other measurement geometries: fan beam, ...
Limitations: Limited angle or noisy data

Original Data  Backprojection  Filtered Backprojection
CT as a linear inverse problem

\[ g(s_i) = \int_{L_i} f(r) \, dl_i \rightarrow \text{Discretization} \rightarrow g = Hf + \epsilon \]
Classical methods in CT

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

- \( H \) is a huge dimensional matrix of line integrals
- \( Hf \) is the forward or projection operation
- \( H^t g \) is the backward or backprojection operation
- \((H^tH)^{-1}H^t g\) is the filtered backprojection minimizing \( \|g - Hf\|^2 \)
- Iterative methods:
  \[
  \hat{f}^{(k+1)} = \hat{f}^{(k)} + \alpha^{(k)} H^t \left( g - H \hat{f}^{(k)} \right)
  \]
  is the Least squares iterative reconstruction method
- Regularization:
  \[
  J(f) = \|g - Hf\|^2 + \lambda \|Df\|^2.
  \]
Inversion: Deterministic methods

Data matching

- Observation model
  \[ g_i = h_i(f) + \epsilon_i, \quad i = 1, \ldots, M \rightarrow g = H(f) + \epsilon \]

- Misatch between data and output of the model \( \Delta(g, H(f)) \)
  \[ \hat{f} = \arg \min_f \{ \Delta(g, H(f)) \} \]

- Examples:
  - LS
    \[ \Delta(g, H(f)) = \| g - H(f) \|^2 = \sum_i |g_i - h_i(f)|^2 \]
  - \( L_p \)
    \[ \Delta(g, H(f)) = \| g - H(f) \|^p = \sum_i |g_i - h_i(f)|^p, \quad 1 < p < 2 \]
  - KL
    \[ \Delta(g, H(f)) = \sum_i g_i \ln \frac{g_i}{h_i(f)} \]

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

Inverse problems = Ill posed problems
\[ \text{Need of prior information} \]

- Minimum norme LS (MNLS): \[ J(f) = \|g - H(f)\|^2 + \lambda\|f\|^2 \]
- Classical regularization: \[ J(f) = \|g - H(f)\|^2 + \lambda\|Df\|^2 \]
- More general regularization:
  \[ J(f) = Q(g - H(f)) + \lambda\Phi(Df) \]
  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
Inversion : Probabilistic methods

Taking account of errors and uncertainties $\rightarrow$ 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

$$g = Hf + \epsilon$$

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

**Link with regularisation:**
Maximum A Posteriori (MAP):

$$\hat{f} = \arg\max_f \{p(f|g)\} = \arg\max_f \{p(g|f) p(f)\}$$

$$\hat{f} = \arg\min_f \{-\ln p(g|f) - \ln p(f)\}$$

with

$$Q(g, Hf) = -\ln p(g|f) \quad \text{and} \quad \lambda \Phi(f) = -\ln p(f)$$
Case of linear models and Gaussian priors

\[ g = Hf + \epsilon \]

- **Hypothesis on the noise**: \( \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I) \)
  \[ p(g | f) \propto \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \|g - Hf\|^2 \right\} \]

- **Hypothesis on \( f \)**: \( f \sim \mathcal{N}(0, \sigma_f^2 (D^t D)^{-1}) \)
  \[ p(f) \propto \exp \left\{ -\frac{1}{2\sigma_f^2} \|Df\|^2 \right\} \]

- **A posteriori**:
  \[ p(f | g) \propto \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \|g - Hf\|^2 \frac{1}{2\sigma_f^2} \|Df\|^2 \right\} \]

- **MAP**:
  \[ \hat{f} = \arg \max_f \{ p(f | g) \} = \arg \min_f \{ J(f) \} \]
  with
  \[ J(f) = \|g - Hf\|^2 + \lambda \|Df\|^2, \quad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2} \]

- **Advantage**: characterization of the solution
  \[ f | g \sim \mathcal{N}(\hat{f}, \hat{P}) \quad \text{with} \quad \hat{f} = \hat{P}H^t g, \quad \hat{P} = (H^t H + \lambda D^t D)^{-1} \]
MAP estimation with other priors:

\[ \hat{f} = \arg \min_f \{ J(f) \} \quad \text{avec} \quad J(f) = \| g - H f \|^2 + \lambda \Omega(f) \]

Separable priors:

- Gaussian prior:
  \[ p(f_j) \propto \exp \left\{ -\alpha (f_j - m_j)^2 \right\} \quad \rightarrow \quad \Omega(f) = \alpha \sum_j (f_j - m_j)^2 \]

- Gamma prior:
  \[ p(f_j) \propto (f_j/m_j)^\alpha \exp \{-f_j/m_j\} \quad \rightarrow \quad \Omega(f) = \alpha \sum_j \ln \frac{f_j}{m_j} + \frac{f_j}{m_j}, \]

- Beta prior:
  \[ p(f_j) \propto f_j^\alpha (1 - f_j)^\beta \quad \rightarrow \quad \Omega(f) = \alpha \sum_j \ln f_j + \beta \sum_j \ln(1 - f_j), \]

- Generalized Gaussian:
  \[ p(f_j) \propto \exp \left\{ -\alpha |f_j - m_j|^p \right\}, \quad 1 < p < 2 \quad \rightarrow \quad \Phi(f) = \alpha \sum_j |f_j - m_j|^p, \]

Markovian models:

\[ p(f_j | f) \propto \exp \left\{ -\alpha \sum_{i \in N_j} \phi(f_j, f_i) \right\} \quad \rightarrow \quad \Phi(f) = \alpha \sum_j \sum_{i \in N_j} \phi(f_j, f_i), \]
Which images I am looking for?
Which image I am looking for?

Gauss-Markov

Piecewise Gaussian

Generalized GM

Mixture of GM
Gauss-Markov-Potts prior models for images

\[ f(r) \quad z(r) \quad c(r) = 1 - \delta(z(r) - z(r')) \]

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

\[
p(f) = \sum P(z = k) \mathcal{N}(m_k, v_k) \quad \text{Mixture of Gaussians}
\]

Separable iid hidden variables:

\[
p(z) = \prod_r p(z(r))
\]

Markovian hidden variables:

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

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

- \( f \mid z \text{ iid}, \ z \text{ iid} : \) 
  Classical case of 
  Mixture of Gaussians

- \( f \mid z \text{ Markov}, \ z \text{ iid} : \) 
  (Markov composite) 
  Mixture of Gauss-Markov

- \( f \mid z \text{ iid}, \ z \text{ Markov} : \) 
  (Hidden Potts-Markov) 
  Gauss-Potts

- \( f \mid z \text{ Markov}, \ z \text{ Markov} : \) 
  (Gauss-Markov-Potts)
Case 1: \( f \mid z \) iid, \( z \) iid

Independent Mixture of Independent Gaussians (IMIG):

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

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

\[
p(z) = \prod_r p(z(r) = k) = \prod_r \alpha_k = \prod_k \alpha_k^n
\]

Noting

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

we have:

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

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

Independent Mixture of Gauss-Markov (IMGM):

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

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

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

\[
= (1 - c(r')) f(r') + c(r') m_z(r')
\]

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

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

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

Gauss iid as in Case 1:

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

Potts-Markov

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

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

Gauss-Markov as in Case 2 :

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

\[
\begin{align*}
\mu_z(r) &= \frac{1}{|\mathcal{V}(r)|} \sum_{r' \in \mathcal{V}(r)} \mu_z^*(r') \\
\mu_z^*(r') &= \delta(z(r') - z(r)) f(r') + (1 - \delta(z(r') - z(r)) m_z(r')
\end{align*}
\]

$$p(f \mid z) \propto \prod_r \mathcal{N}(\mu_z(r), \nu_z(r)) \propto \prod_k \alpha_k \mathcal{N}(m_k1, \Sigma_k)$$

Potts-Markov as in Case 3 :

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

\[ g = H f + \epsilon \]

- **Forward & errors model**: \( p(g|f, \theta; \mathcal{M}) \)
- **Prior models**: \( p(f|\theta; \mathcal{M}) \) and \( p(\theta|\mathcal{M}) \)
- **Bayes**: \( p(f, \theta|g; \mathcal{M}) = \frac{p(g|f, \theta; \mathcal{M}) p(f|\theta; \mathcal{M}) p(\theta|\mathcal{M})}{p(g|\mathcal{M})} \)
- **Joint MAP**: 
  \[ (\hat{f}, \hat{\theta}) = \arg \max_{(f, \theta)} \{ p(f, \theta|g; \mathcal{M}) \} \]

- **Posterior means**: 
  \[
  \begin{align*}
  \hat{f} &= \int f \ p(f, \theta|g; \mathcal{M}) \ df \ d\theta \\
  \hat{\theta} &= \int \theta \ p(f, \theta|g; \mathcal{M}) \ df \ d\theta
  \end{align*}
  \]

- **Evidence of the model**: 
  \[ p(g|\mathcal{M}) = \int \int p(g|f, \theta; \mathcal{M}) p(f|\theta; \mathcal{M}) p(\theta|\mathcal{M}) \ df \ d\theta \]
Bayesian Computation

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

\[ p(f, z, \theta | g) = \frac{p(g|f, \theta) \ p(f|z, \theta) \ p(z)}{p(g|\theta)} \]

Approximations:

- \( f | z \) iid, \( z \) iid:
  \[ q(f, z, \theta | g) = q_1(f | z) \ q_2(z) \ q_3(\theta). \]

- \( f | z \) iid, \( z \) Markov:
  \[ q(f, z, \theta | g) = q_1(f | z) \ q_{2w}(z_w) \ q_{2b}(z_b) \ q_3(\theta). \]

- \( f \) Markov, \( z \) iid:
  \[ q(f, z, \theta | g) = q_{1w}(f_w | z) \ q_{1b}(f_b | z) \ q_2(z) \ q_3(\theta). \]

- \( f \) Markov, \( z \) iid:
  \[ q(f, z, \theta | g) = q_{1w}(f_w | z) \ q_{1b}(f_b | z) \ q_{2w}(z_w) \ q_{2b}(z_b) \ q_3(\theta) \]
Application of CT in NDT

Reconstruction from only 2 projections
Application in CT

\[ g \mid f = H f + \epsilon \]
\[ g \mid f \sim \mathcal{N}(H f, \nu \epsilon I) \]
Gaussian

\[ f \mid z \]
\[ f \mid z \sim \text{iid Gaussian} \]
\[ f \mid z \sim \text{Gauss-Markov} \]
\[ f \mid z \sim \text{Gauss-Markov} \]

\[ z \mid c \]
\[ z \mid c \sim \text{iid} \]
\[ z \mid c \sim \text{Potts} \]

\[ c(r) \in \{0, 1\} \]
\[ 1 - \delta(z(r) - z(r')) \]
binary
Proposed algorithm

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

General scheme:

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

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

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

- Estimate \( \theta \) using
  \[ p(\theta | \hat{f}, \hat{z}, g) \propto p(g | \hat{f}, \sigma^2_{\epsilon} I) p(\hat{f} | \hat{z}, (m_k, \nu_k)) p(\theta) \]
  Conjugate priors \( \rightarrow \) analytical expressions.
Results

Original

Backprojection

Filtered BP

LS

Gauss-Markov+pos

GM+Line process

GM+Label process
Application in Microwave imaging

\[ g(\omega) = \int f(r) \exp \{-j(\omega \cdot r)\} \, dr + \epsilon(\omega) \]

\[ g(u, v) = \int f(x, y) \exp \{-j(ux + vy)\} \, dx \, dy + \epsilon(u, v) \]

\[ g = Hf + \epsilon \]
Conclusions

- Bayesian Inference for inverse problems
- Approximations (Laplace, MCMC, Variational)
- Gauss-Markov-Potts are useful prior models for images incorporating regions and contours
- Separable approximations for Joint posterior with Gauss-Markov-Potts priors
- Application in different CT (X ray, US, Microwaves, PET, SPECT)

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


- A. Mohammad-Djafari, J.F. Giovannelli, G. Demoment and J. Idier,
Questions and Discussions

- Thanks for your attentions
- ...
- ...
- Questions?
- Discussions?
- ...
- ...
CT from two projections = Joint distribution from its marginals

\[ g_1(x) = \int f(x, y) \, dy \]
\[ g_2(y) = \int f(x, y) \, dx \]

Given the marginals \( g_1(x) \) and \( g_2(y) \) find the joint distribution \( f(x, y) \)

Infinite number of solutions

\[ f(x, y) = g_1(x) g_2(y) \Omega(G_1(x), G_2(y)) \]
\[ \Omega(u, v) \] is a Copula:
\[ \Omega(u, 0) = 0, \Omega(u, 1) = 1, \]
\[ \Omega(0, u) = 0, \Omega(1, u) = 1 \]

\( (x, y) \in [0, 1]^2, G_1(x) \) and \( G_2(y) \) are CDFs of \( g_1(x) \) and \( g_2(y) \)

Example: \( \Omega(u, v) = uv \)

Any link between geometrical structure of \( f \) and Copula functions?