Bayesian Discrete Tomography from a few number of projections

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

Invited talk at
Workshop on Discrete Tomography, Polytechnico de Milan, Italy
Contents

1. Limited angle Tomography
2. Basic Bayesian approach
3. Two main steps:
   - Choosing appropriate Prior model
   - Do the computational efficiently
4. Hierarchical prior modelling
   - Sparsity enforcing models through Student-t and IGSM
   - Gauss-Markov-Potts models
5. Computational tools: JMAP, Gibbs Sampling MCMC, VBA
6. Case study: Image Reconstruction with only two projections
7. Implementation issues
   - Main GPU implementation steps: Forward and Back Projections
   - Multi-Resolution implementation
8. Conclusions
Limited angle Tomography: Limitations of analytical methods

Original Data Backprojection Filtered Backprojection

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 3/30
Algebraic methods: Discretization

\[ f(x, y) = \sum_j f_j b_j(x, y) \]

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

\[ g(r, \phi) = \int_L f(x, y) \, dl \]

\[ g_i = \sum_{j=1}^N H_{ij} f_j + \epsilon_i \]

\[ g_k = H_k f + \epsilon_k, \quad k = 1, \ldots, K \quad \rightarrow \quad g = H f + \epsilon \]

\( g_k \) projection at angle \( \phi_k \), \( g \) all the projections.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy.
Algebraic methods

\[ \mathbf{g} = \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_K \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \rightarrow \mathbf{g}_k = \mathbf{H}_k \mathbf{f} + \epsilon_k \rightarrow \mathbf{g} = \sum_k \mathbf{H}_k \mathbf{f} + \epsilon_k = \mathbf{H} \mathbf{f} + \epsilon \]

- \( \mathbf{H} \) is huge dimensional: 2D: \( 10^6 \times 10^6 \), 3D: \( 10^9 \times 10^9 \).
- \( \mathbf{H} \mathbf{f} \) corresponds to forward projection
- \( \mathbf{H}^t \mathbf{g} \) corresponds to Back projection (BP)
- \( \mathbf{H} \) may not be invertible and even not square
- \( \mathbf{H} \) is, in general, ill-conditioned
- In limited angle tomography \( \mathbf{H} \) is under determined, so the problem has infinite number of solutions
- Minimum Norm Solution

\[ \hat{\mathbf{f}} = \mathbf{H}^t (\mathbf{H} \mathbf{H}^t)^{-1} \mathbf{g} = \sum_k \mathbf{H}^t_k (\mathbf{H}_k \mathbf{H}_k^t)^{-1} \mathbf{g}_k \]

can be interpreted as the Filtered Back Projection solution.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 5/30
Algebraic methods

- **Minimum Norm Solution:**
  
  \[
  \min \| f \|_2^2 \text{ s.t. } H f = g \quad \rightarrow \quad \hat{f} = H^t (H H^t)^{-1} g
  \]

- **Least square Solution:**
  \[
  \hat{f} = \arg \min_f \{ J(f) = \| g - H f \|_2^2 \} \quad \rightarrow \quad \hat{f} = (H^t H)^{-1} H^t g
  \]

- **Quadratic Regularization:**
  \[
  J(f) = \| g - H f \|_2^2 + \lambda \| f \|_2^2 \quad \rightarrow \quad \hat{f} = (H^t H + \lambda I)^{-1} H^t g
  \]

- **L1 Regularization:** \( J(f) = \| g - H f \|_2^2 + \lambda \| f \|_1 \)

- **Lpq Regularization:** \( J(f) = \| g - H f \|_p^p + \lambda \| D f \|_q^q \)

- **More general Regularization:**
  \[
  J(f) = \sum_i \phi(g_i - [H f]_i) + \lambda \sum_j \psi((D f)_j)
  \]

  \[
  J(f) = \Delta_1(g, H f) + \lambda \Delta_2(f, f_0)
  \]

  with \( \Delta_1 \) and \( \Delta_2 \) any distances (L2, L2, ..) or divergence (KL)

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 6
Computed Tomography with only two projections

\[ g(r, \phi) = \int_{L} f(x, y) \, dl \]

\[ f(x, y) = \sum_{j} b_{j}(x, y) f_{j} \]

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

\[ g_{i} = \sum_{j=1}^{N} H_{ij} f_{j} + \epsilon_{i} \]

\[ g = Hf + \epsilon \]

Case study: Reconstruction from 2 projections

\[ g_{1}(x) = \int f(x, y) \, dy, \]

\[ g_{2}(y) = \int f(x, y) \, dx \]

Very ill-posed inverse problem

\[ f(x, y) = g_{1}(x) g_{2}(y) \Omega(x, y) \]

\[ \Omega(x, y) \text{ is a Copula:} \]

\[ \int \Omega(x, y) \, dx = 1 \]

\[ \int \Omega(x, y) \, dy = 1 \]
Simple example

\[
\begin{bmatrix}
1 & 3 & 4 \\
2 & 4 & 6 \\
3 & 7 & 6
\end{bmatrix}
\begin{bmatrix}
? & ? & 4 \\
? & ? & 6 \\
? & ? & ?
\end{bmatrix}
\begin{bmatrix}
f_1 \\
f_2 \\
f_3 \\
f_4 \\
f_5 \\
f_6 \\
f_7 \\
f_8 \\
f_9 \\
g_1 \\
g_2 \\
g_3 \\
g_4 \\
g_5 \\
g_6
\end{bmatrix}
= \begin{bmatrix}
1 & -1 & 0 \\
-1 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
g_1 \\
g_2 \\
g_3
\end{bmatrix}
\]

▷ $Hf = g \rightarrow \hat{f} = H^{-1}g$ if $H$ invertible.

▷ $H$ is rank deficient: $\text{rank}(H) = 3$

▷ 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.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 8/30
Prior information or constraints

- Positivity: \( f_j > 0 \) or \( f_j \in \mathbb{R}^+ \)
- Boundedness: \( 1 > f_j > 0 \) or \( f_j \in [0, 1] \)
- Smoothness: \( f_j \) depends on the neighborhoods.
- Sparsity: many \( f_j \) are zeros.
- Sparsity in a transform domain: \( f = Dz \) and many \( z_j \) are zeros.
- Discrete valued (DV): \( f_j \in \{0, 1, ..., K\} \)
- Binary valued (BV): \( f_j \in \{0, 1\} \)
- Compactness: \( f(r) \) is non zero in one or few non-overlapping compact regions
- Combination of the above mentioned constraints
- Main mathematical questions:
  - Which combination results to unique solution?
  - How to apply them?
Deterministic approaches

- Iterative methods: SIRT, ART, Quadratic or L1 regularization, Bloc Coordinate Descent, Multiplicative ART, ...
- Criteria: \( J(f) = \| g - Hf \|^2 + \lambda \| Df \|^2 \)
- Gradient based algorithms: \( \nabla J(f) = -2H^t(g - Hf) + 2\lambda D^tDf \)
- Simplest algorithm:
  \[
  \hat{f}^{(k+1)} = \hat{f}^{(k)} + \alpha^{(k)} \left[ H^t(g - H\hat{f}^{(k)}) + 2\lambda D^tD\hat{f}^{(k)} \right]
  \]
- More criteria:
  \[
  J(f) = \sum_i \phi(g_i - [Hf]_i) + \lambda \sum_j \psi((Df)_j)
  \]
  with \( \phi(t) \) and \( \psi(t) = \{ t^2, |t|, |t|^p, ... \} \) or
  \[
  J(f) = \Delta_1(g, Hf) + \lambda \Delta_2(f, f_0)
  \]
- Imposing constraints in each iteration (example: DART)
- Mathematical studies of uniqueness and convergence of these algorithms are necessary
Bayesian estimation approach

\[ M : \quad g = Hf + \epsilon \]

- **Observation model** \( M \) + Hypothesis on the noise \( \epsilon \):
  \[ p(g|f; M) = p\epsilon(g - Hf) \]

- **A priori information** \( p(f|M) \)

- **Bayes**:
  \[ p(f|g; M) = \frac{p(g|f; M) p(f|M)}{p(g|M)} \]

- **Maximum A Posteriori (MAP)**:
  \[ \hat{f} = \arg\max_f \{ p(f|g) \} = \arg\max_f \{ p(g|f) p(f) \} \]
  \[ = \arg\min_f \{ J(f) = -\ln p(g|f) - \ln p(f) \} \]

- **Link with Regularization**:
  \[ \hat{f} = \arg\min_f \{ J(f) = \Delta_1(g, Hf) + \lambda R(f) \} \]

with \( \Delta_1(g, Hf) = -\ln p(g|f) \) and \( \lambda R(f) = -\ln p(f) \)
Case of linear models and Gaussian priors

\[ \mathbf{g} = \mathbf{H} \mathbf{f} + \epsilon \]

- Prior knowledge on the noise:
  \[ \epsilon \sim \mathcal{N}(0, \nu_\epsilon^2 \mathbf{I}) \rightarrow p(\epsilon | \mathbf{f}) \propto \exp \left[ -\frac{1}{2\nu_\epsilon^2} \| \mathbf{g} - \mathbf{H} \mathbf{f} \|^2 \right] \]

- Prior knowledge on \( \mathbf{f} \):
  \[ \mathbf{f} \sim \mathcal{N}(0, \nu_f^2 (\mathbf{D}' \mathbf{D})^{-1}) \rightarrow p(\mathbf{f}) \propto \exp \left[ -\frac{1}{2\nu_f^2} \| \mathbf{D} \mathbf{f} \|^2 \right] \]

- A posteriori:
  \[ p(\mathbf{f} | \mathbf{g}) \propto \exp \left[ -\frac{1}{2\nu_\epsilon^2} \| \mathbf{g} - \mathbf{H} \mathbf{f} \|^2 - \frac{1}{2\nu_f^2} \| \mathbf{D} \mathbf{f} \|^2 \right] \]

- MAP: \[ \hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \{ p(\mathbf{f} | \mathbf{g}) \} = \arg \min_{\mathbf{f}} \{ J(\mathbf{f}) \} \]
  with \[ J(\mathbf{f}) = \| \mathbf{g} - \mathbf{H} \mathbf{f} \|^2 + \lambda \| \mathbf{D} \mathbf{f} \|^2, \quad \lambda = \frac{\nu_\epsilon^2}{\nu_f^2} \]

- Advantage: characterization of the solution
  \[ p(\mathbf{f} | \mathbf{g}) = \mathcal{N}(\hat{\mathbf{f}}, \hat{\Sigma}) \text{ with} \]
  \[ \hat{\mathbf{f}} = \left( \mathbf{H}' \mathbf{H} + \lambda \mathbf{D}' \mathbf{D} \right)^{-1} \mathbf{H}' \mathbf{g}, \quad \hat{\Sigma} = \nu_\epsilon \left( \mathbf{H}' \mathbf{H} + \lambda \mathbf{D}' \mathbf{D} \right)^{-1} \]
MAP estimation with other priors:

\[ \hat{f} = \arg \min_f \{ J(f) \} \quad \text{with} \quad J(f) = \frac{1}{\nu_\epsilon} \| g - Hf \|^2 + \Omega(f) \]

Separable priors:

- **Gaussian:**
  \[ p(f_j) \propto \exp[-\alpha |f_j|^2] \quad \rightarrow \quad \Omega(f) = \alpha \sum_j |f_j|^2 = \|f\|^2 \]

- **Generalized Gaussian:**
  \[ p(f_j) \propto \exp[-\alpha |f_j|^p], \quad 1 < p < 2 \quad \rightarrow \quad \Omega(f) = \alpha \sum_j |f_j|^p = \|f\|^p \]

- **Gamma:**
  \[ p(f_j) \propto f_j^\alpha \exp[-\beta f_j] \quad \rightarrow \quad \Omega(f) = \alpha \sum_j \ln f_j + \beta f_j \]

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

Markovian models:

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

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy.
Sparsity enforcing models

- 3 classes of models: 1- Generalized Gaussian, 2- Mixture models and 3- Heavy tailed (Cauchy and Student-t)
- Student-t model

\[
St(f|\nu) \propto \exp \left[ -\frac{\nu + 1}{2} \log \left(1 + \frac{f^2}{\nu}\right) \right]
\]

- Infinite Gaussian Scaled Mixture (IGSM) equivalence

\[
St(f|\nu) \propto \int_{0}^{\infty} \mathcal{N}(f|0,1/z) \mathcal{G}(z|\alpha,\beta) \, dz, \quad \text{with } \alpha = \beta = \nu/2
\]

\[
\begin{align*}
    p(f|z) &= \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0,1/z_j) \propto \exp \left[ -\frac{1}{2} \sum_j z_j f_j^2 \right] \\
    p(z|\alpha,\beta) &= \prod_j \mathcal{G}(z_j|\alpha,\beta) \propto \prod_j z_j^{\alpha - 1} \exp \left[ -\beta z_j \right] \\
    &\propto \exp \left[ \sum_j (\alpha - 1) \ln z_j - \beta z_j \right] \\
    p(f,z|\alpha,\beta) &\propto \exp \left[ -\frac{1}{2} \sum_j z_j f_j^2 + (\alpha - 1) \ln z_j - \beta z_j \right]
\end{align*}
\]
Non stationary noise and sparsity enforcing model

- Non stationary noise:
  \[ g = Hf + \epsilon, \quad \epsilon_i \sim \mathcal{N}(\epsilon_i|0, \nu_{\epsilon_i}) \rightarrow \epsilon \sim \mathcal{N}(\epsilon|0, V_\epsilon = \text{diag} [\nu_{\epsilon_1}, \cdots, \nu_{\epsilon_M}]) \]

- Student-t prior model and its equivalent IGSM:
  \[ f_j | \nu_{f_j} \sim \mathcal{N}(f_j|0, \nu_{f_j}) \text{ and } \nu_{f_j} \sim \mathcal{IG}(\nu_{f_j}|\alpha_{f_0}, \beta_{f_0}) \rightarrow f_j \sim \text{St}(f_j|\alpha_{f_0}, \beta_{f_0}) \]

\[
\begin{align*}
\alpha_{f_0}, \beta_{f_0} & \quad \alpha_{\epsilon_0}, \beta_{\epsilon_0} \\
\nu_{f} & \quad \nu_{\epsilon} \\
f & \quad \epsilon \\
H & \quad g
\end{align*}
\]

\[
\begin{align*}
\{ p(g|f, \nu_{\epsilon}) & = \mathcal{N}(g|Hf, V_\epsilon), \quad V_\epsilon = \text{diag} [\nu_{\epsilon}] \\
p(f|\nu_f) & = \mathcal{N}(g|0, V_f), \quad V_f = \text{diag} [\nu_f] \\
p(\nu_{\epsilon}) & = \prod_i \mathcal{IG}(\nu_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\
p(\nu_f) & = \prod_i \mathcal{IG}(\nu_{f_j}|\alpha_{f_0}, \beta_{f_0})
\end{align*}
\]

\[ p(f, \nu_{\epsilon}, \nu_f|g) \propto p(g|f, \nu_{\epsilon}) p(f|\nu_f) p(\nu_{\epsilon}) p(\nu_f) \]

- JMAP: \( (\hat{f}, \hat{\nu}_{\epsilon}, \hat{\nu}_f) = \arg \max_{(f, \nu_{\epsilon}, \nu_f)} \{ p(f, \nu_{\epsilon}, \nu_f|g) \} \)

- VBA: Approximate \( p(f, \nu_{\epsilon}, \nu_f|g) \) by \( q_1(f) q_2(\nu_{\epsilon}) q_3(\nu_f) \)

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy.
Sparse model in a Transform domain

\[ g = Hf + \epsilon, \quad f = Dz, \quad z \text{ sparse} \]

\[
\begin{align*}
\{ p(g|z, v_\epsilon) &= \mathcal{N}(g|HDf, v_\epsilon I) \\
p(z|v_z) &= \mathcal{N}(z|0, V_z), \quad V_z = \text{diag}[v_z] \\
p(v_\epsilon) &= \mathcal{IG}(v_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) \\
p(v_z) &= \prod_i \mathcal{IG}(v_{z j}|\alpha_{z 0}, \beta_{z 0}) 
\end{align*}
\]

\[ p(z, v_\epsilon, v_z, v_\xi|g) \propto p(g|z, v_\epsilon) p(z|v_z) p(v_\epsilon) p(v_z) p(v_\xi) \]

- JMAP:
  \[ (\hat{z}, \hat{v}_\epsilon, \hat{v}_z) = \arg \max \{ p(z, v_\epsilon, v_z|g) \} \]

Alternate optimization:

\[ \hat{z} = \arg \min_z \{ J(z) \} \quad \text{with:} \]

\[
\begin{align*}
J(z) &= \frac{1}{2\hat{v}_\epsilon} \| g - HDz \|^2 + \| V_z^{-1/2}z \|^2 \\
\hat{v}_z &= \frac{\beta_{z 0} + \hat{z}_j^2}{\alpha_{z 0} + 1/2} \\
\hat{v}_\epsilon &= \frac{\beta_{\epsilon 0} + \| g - HD\hat{z} \|^2}{\alpha_{\epsilon 0} + M/2}
\end{align*}
\]

- VBA: Approximate
  \[ p(z, v_\epsilon, v_z, v_\xi|g) \] by \( q_1(z) q_2(v_\epsilon) q_3(v_z) \)

Alternate optimization.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy.
Sparse model in a Transform domain 2

\[ \mathbf{g} = \mathbf{Hf} + \epsilon, \quad \mathbf{f} = \mathbf{Dz} + \xi, \quad \mathbf{z} \text{ sparse} \]

\[
\begin{cases}
    p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{Hf}, \mathbf{v}_\epsilon I) \\
p(\mathbf{f}|\mathbf{z}) = \mathcal{N}(\mathbf{f}|\mathbf{Dz}, \mathbf{v}_\xi I), \\
p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|0, \mathbf{V}_z), \quad \mathbf{V}_z = \text{diag}[\mathbf{v}_z] \\
p(\mathbf{v}_\epsilon) = \mathcal{IG}(\mathbf{v}_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) \\
p(\mathbf{v}_z) = \prod_i \mathcal{IG}(\mathbf{v}_{zj}|\alpha_{z0}, \beta_{z0}) \\
p(\mathbf{v}_\xi) = \mathcal{IG}(\mathbf{v}_\xi|\alpha_{\xi 0}, \beta_{\xi 0})
\end{cases}
\]

\[
p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, \mathbf{v}_\xi|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) p(\mathbf{f}|\mathbf{z}_f) p(\mathbf{z}|\mathbf{v}_z) p(\mathbf{v}_\epsilon) p(\mathbf{v}_z) p(\mathbf{v}_\xi)
\]

- JMAP:
  \[
  (\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}_\epsilon, \hat{\mathbf{v}}_z, \hat{\mathbf{v}}_\xi) = \text{arg max } \{ p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, \mathbf{v}_\xi|\mathbf{g}) \} \\
  (\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, \mathbf{v}_\xi)
  \]

Alternate optimization.

- VBA: Approximate
  \[
p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, \mathbf{v}_\xi|\mathbf{g}) \text{ by } q_1(\mathbf{f}) q_2(\mathbf{z}) q_3(\mathbf{v}_\epsilon) q_4(\mathbf{v}_z) q_5(\mathbf{v}_\xi)
  \]

Alternate optimization.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy.
Gauss-Markov-Potts prior models for images

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

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

- Separable iid hidden variables: \( p(z) = \prod_r p(z(r)) \)
- Markovian hidden variables: \( p(z) \) Potts-Markov:

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

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

\( c(r) = 1 - \delta(z(r) - z(r')) \)

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy.
Four different cases

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

- $f | z \text{ Gaussian iid, } z \text{ iid :}$
  Mixture of Gaussians

- $f | z \text{ Gauss-Markov, } z \text{ iid :}$
  Mixture of Gauss-Markov

- $f | z \text{ Gaussian iid, } z \text{ Potts-Markov :}$
  Mixture of Independent Gaussians
  \text{(MIG with Hidden Potts)}

- $f | z \text{ Markov, } z \text{ Potts-Markov :}$
  Mixture of Gauss-Markov
  \text{(MGM with hidden Potts)}
Gauss-Markov-Potts prior models for images

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

\[
\begin{align*}
g &= Hf + \epsilon \\
p(g|f, v_\epsilon) &= \mathcal{N}(g|Hf, v_\epsilon I) \\
p(v_\epsilon) &= \mathcal{IG}(v_\epsilon|\alpha_{\epsilon 0}, \beta_{\epsilon 0}) \\
p(f|z, \theta) &= \sum_k \prod_{r \in R_k} a_k \mathcal{N}(f(r)|m_k, v_k), \\
\theta &= \{(a_k, m_k, v_k), k = 1, \ldots, K\} \\
p(\theta) &= \mathcal{D}(a|a_0)\mathcal{N}(a|m_0, v_0)\mathcal{IG}(v|\alpha_0, \beta_0) \\
p(z|\gamma) &\propto \exp \left[ \gamma \sum_r \sum_{r' \in N(r)} \delta(z(r) - z(r')) \right] \text{ Potts MRF} \\
p(f, z, \theta|g) &\propto p(g|f, v_\epsilon) p(f|z, \theta) p(z|\gamma) \\
\end{align*}
\]

MCMC: Gibbs Sampling

VBA: Alternate optimization.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 20/30
Bayesian Computation and Algorithms

- Joint posterior probability law of all the unknowns \( f, z, \theta \)

\[
p(f, z, \theta | g) \propto p(g | f, \theta_1) \ p(f | z, \theta_2) \ p(z | \theta_3) \ p(\theta)
\]

- Often, the expression of \( p(f, z, \theta | g) \) is complex.

- Its optimization (for Joint MAP) or its marginalization or integration (for Marginal MAP or PM) is not easy.

- Two main techniques:
  - MCMC: Needs the expressions of the conditionals \( p(f | z, \theta, g) \), \( p(z | f, \theta, g) \), and \( p(\theta | f, z, g) \)
  - VBA: Approximate \( p(f, z, \theta | g) \) by a separable one

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

and do any computations with these separable ones.
MCMC based algorithm

\[ p(f, z, \theta | g) \propto p(g | f, z, \theta_1) p(f | z, \theta_2) p(z | \theta_3) p(\theta) \]

General Gibbs sampling scheme:

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

- Generate samples \( f \) using \( p(f | \hat{z}, \hat{\theta}, g) \propto p(g | f, \theta) p(f | \hat{z}, \hat{\theta}) \)
  When Gaussian, can be done via optimization of a quadratic criterion.

- Generate samples \( z \) using \( p(z | \hat{f}, \hat{\theta}, g) \propto p(g | \hat{f}, \hat{z}, \hat{\theta}) p(z) \)
  Often needs sampling (hidden discrete variable)

- Generate samples \( \theta \) using
  \[ p(\theta | \hat{f}, \hat{z}, g) \propto p(g | \hat{f}, \sigma^2) p(\hat{f}| \hat{z}, (m_k, v_k)) p(\theta) \]
  Use of Conjugate priors \( \rightarrow \) analytical expressions.

- After convergence use samples to compute means and variances.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 22/30
Application in CT: Reconstruction from 2 projections

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

\[ f | z \]
iid Gaussian
or
Gauss-Markov

\[ z \]
 iid
or
Potts

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

\[ p(f, z, \theta | g) \propto p(g | f, \theta_1) p(f | z, \theta_2) p(z | \theta_3) p(\theta) \]
Proposed algorithms

\[ p(f, z, \theta | g) \propto p(g | f, \theta_1) p(f | z, \theta_2) p(z | \theta_3) p(\theta) \]

- MCMC based general scheme:

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

Iterative algorithm:

- Estimate \( f \) using \( p(f | \hat{z}, \hat{\theta}, g) \propto p(g | f, \theta) p(f | \hat{z}, \hat{\theta}) \)
  Needs optimization 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

\[ p(\theta | \hat{f}, \hat{z}, g) \propto p(g | \hat{f}, \sigma^2) p(\hat{f} | \hat{z}, (m_k, v_k)) p(\theta) \]
Conjugate priors \( \rightarrow \) analytical expressions.

- Variational Bayesian Approximation

\[ \text{Approximate } p(f, z, \theta | g) \text{ by } q_1(f) q_2(z) q_3(\theta) \]
Results with two projections

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

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 25/30
Implementation issues

- In almost all the algorithms, the step of computation of $\hat{f}$ needs an optimization algorithm.
- The criterion to optimize is often in the form of

$$J(f) = \|g - Hf\|^2 + \lambda\|Df\|^2$$

- Very often, we use the gradient based algorithms which need to compute

$$\nabla J(f) = -2H^t(g - Hf) + 2\lambda D^tDf$$

- So, for the simplest case, in each step, we have

$$\hat{f}^{(k+1)} = \hat{f}^{(k)} + \alpha^{(k)} \left[ H^t(g - H\hat{f}^{(k)}) + 2\lambda D^tD\hat{f}^{(k)} \right]$$
Gradient based algorithms

\[ \hat{f}^{(k+1)} = \hat{f}^{(k)} + \alpha \left[ H' \left( g - H\hat{f}^{(k)} \right) - \lambda D'D\hat{f}^{(k)} \right] \]

1. Compute \( \hat{g} = H\hat{f} \) (Forward projection)
2. Compute \( \delta g = g - \hat{g} \) (Error or residual)
3. Compute \( \delta f_1 = H'\delta g \) (Backprojection of error)
4. Compute \( \delta f_2 = -D'D\hat{f} \) (Correction due to regularization)
5. Update \( \hat{f}^{(k+1)} = \hat{f}^{(k)} + [\delta f_1 + \delta f_2] \)

Steps 1 and 3 need great computational cost and have been implemented on GPU.

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 27/30
Multi-Resolution Implementation

Scale 1: black \( g^{(1)} = H^{(1)} f^{(1)} \) \((N \times N)\)

Scale 2: green \( g^{(2)} = H^{(2)} f^{(2)} \) \((N/2 \times N/2)\)

Scale 3: red \( g^{(3)} = H^{(3)} f^{(3)} \) \((N/4 \times N/4)\)
Results with 4 projection

Original

Projections

Initialization

Final result

A. Mohammad-Djafari, Bayesian Discrete Tomography from a few number of projections, Mars 21-23, 2016, Polytechnico de Milan, Italy. 29/30
Conclusions

- Limited angle Computed Tomography is a very ill-posed Inverse problem
- Analytical methods have many limitations
- Algebraic methods push further these limitations
- Deterministic Regularization methods push still further the limitations of ill-conditioning.
- Probabilistic and in particular the Bayesian approach has many potentials
- 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