Total Variation Regularization and Large Scale Volume Reconstructions in Tomography

Sami S. Brandt
Linear Volumetric Imaging

(a) Tomography
(b) Confocal z-stack imaging
(c) Microrotation confocal imaging
Contents

• Bayesian reconstruction problem in volumetric imaging
• Sparsity regularisation by total variation and spatial derivative priors
• High Dimensional Optimization – EM – Skilling-Bryan
• Application example – Microrotation reconstruction
Micro-rotation fluorescence imaging

Micro-rotation imaging aims at
1) To image living cells in natural environment
2) To improve image resolution in 3D

http://www.pfid.org/AUTOMATION/
Examples of micro-rotation images

A human living cell, expressing fluorescence at nuclear envelope

Laksameethanasan et al. 2008
Step 0: Image Registration

• Imaging geometry need to be solved prior to the reconstruction problem
• Beyond the scope of this talk, but we have studied that too
  – TEM: Brandt et al. (2001a, 2001b), Brandt and Ziese (2005), Brandt (2006);
  – X-ray tomography: Brandt and Kolehmainen (2007);
Projection Model

• Assume a linear projection model

\[ m_i(x, y) = A_i f(x, y, z), \]

where \( A_i: C_3 \rightarrow C_2 \)

• Assuming a linear and shift invariant system

\[ m_i(x, y) = h(x, y, z) * f_i(x, y, z) \bigg|_{z=d}, \]
Bayesian Reconstruction

- Consider the discretised model
  \[ \hat{m} = Af \]
- The complete solution is the posterior
  \[ p(f|m) = \frac{p(f)p(m|f)}{p(m)} \propto p(f)p(m|f), \]
  where \( p(m|f) \) is the likelihood and \( p(f) \) is the prior.
Likelihood

- Obtained from the noise model
- Gaussian noise
  \[ p(m|f) \propto \exp \left( -\frac{1}{2\sigma} ||m - Af||^2 \right) \]
- Poisson noise (photon counting)
  \[ p(m|f) = \prod_{j=1}^{KM} \left( \frac{1}{m_j!} \right) \exp(m^T \log(Af) - 1^T Af) \]
Sparsity Prior

- We may take the desired sparsity of the solution into account in the prior $p(f)$.
- What choices do we have?

For instance:
- Pseudo norm
- One norm (Lasso)
- Total variation
- Spatial derivative priors
Total variation Prior

- In the continuous case
  \[ p_f \propto \exp\left(-\lambda \int |\nabla f| \, dV\right) \]
- If \( f \) is the characteristic function of the set \( B \)
  \[ \text{TV}(f) = \int |\nabla f| \, dV = \text{length}(\partial B) \]
- Discrete definition (four neighbourhood)
  \[ p(f) = \exp\left(-\frac{\lambda}{2} \sum_{i \in N_j} |f_i - f_j| \right) \]
Total Variation Example

Images with the same energy but increasing total variation

TV = 18  TV = 26  TV = 42
Spatial Derivative Priors

- We may use a more general class of priors

\[ p_f \propto \exp\left(-\lambda \int |G \cdot f| \, dV \right) \]

where \( G: C_3 \rightarrow C_3 \) is a linear operator.

- We have used the Laplacian instead of gradient

\[ p_f \propto \exp\left(-\lambda \int |\Delta f| \, dV \right) \]
Discrete Laplacian Example

Images with the same energy but increasing total absolute (discrete) Laplacian

SD = 32

SD = 42

SD = 52
How does this relate to sparsity of the solution?

• Consider negative log posterior

\[ E(f) = -\log p(m|f) - \lambda \log p(f) \]

• Computing the MAP estimate is multiobjective optimization

• The regularization parameter is chosen so that the fit (likelihood) is at desired level (Morozov discrepancy principle)

• From that subset the prior is maximized
How does this relate to sparsity of the solution?

• Consider first the one norm prior

\[ E(f) = -\log p(m|f) + \lambda \|f\|_1 \]

\[ \|f\|_1 = C_\lambda^2 \]

\[ -\log p(m|f) = C_\lambda^1 \]
How does this relate to sparsity of the solution?

- For the spatial derivative priors

\[ E(f) = -\log p(m|f) + \lambda \|Gf\|_1 \]

\[ \|g\|_1 \equiv \|Gf\|_1 = C_\lambda^2 \]
How does this relate to sparsity of the solution?

• For the spatial derivative priors

\[ E(f) = -\log p(m|f) + \lambda \|Gf\|_1 \]
How does this relate to sparsity of the solution?

The linearity preserves the vertex

\[-\log p(m|f) = C^1_{\lambda}\]

\[f : \|Gf\|_1 = C^2_{\lambda}\]
How does this relate to sparsity of the solution?

- Total variation favours sparse solutions in the first derivative (edge preservation)
- Our Spatial derivative prior favours sparsity in the second derivative (edge and smoothness preservation)
- The 1-norm imposes the sparsity for a large class of linear operators
Implementation of the Prior

- The laplacian computed by convolution with the LoG filter (Gaussian interpolation)

\[
TV(f) = \int |Gf| dV
\]

\[
\approx \sum_l \left| \sum_k f_l \Delta g(r_l - r_k) \right| = \|Gf\|_1
\]

where G is the Toepliz matrix corresponding to the 3D convolution with the LoG kernel.
Implementation of the Prior

• To make the energy function differentiable at zero, we approximate

\[ t \approx \beta^{-1} \cosh(\beta t) \]

in \( \|Gf\|_1 \approx \|Gf\|_1 \)

• The prior finally takes the form

\[ p(f) \propto \exp(-\lambda \|Gf\|_1) \]
Computation of the MAP Estimate

- Poisson noise and the spatial derivative prior yields the optimization problem

\[
\min_{f} \left\{ 1^T (Af) - m^T \log(Af) + \lambda \|Gf\|_1 \right\}
\]

with subject to \( f_i \geq 0 \forall i \)

- Here we consider two algorithms:
  - Expectation Maximization (EM)
  - Non-linear optimization by Skilling-Bryan
EM algorithm

- Solution by the iteration (Green 1990, Dey 2006, Laksameethanasan et al. 2008)

\[
f_{k+1} = \frac{f_k}{A^T 1 + \lambda \nabla \|Gf_k\|_1}(A^T m - A f_k)
\]

where

\[
\nabla \|Gf_k\|_1 = G^T \tanh(\beta Gf_k)
\]

- Note: the matrices for forward projection \(A\) and its adjoint \(A^T\) are not computed
Toy Reconstruction Example

ML

MAP-TV

MAP-SD

(intensity profile)
Generalized Skilling-Bryan method

• A 2\textsuperscript{nd} order, non-linear optimization algorithm to compute the MAP estimate

\[
\arg \max_f p(f|m), \quad f_i \geq 0 \forall i
\]

which is equivalent to maximizing

\[
q(f) = s(f) - \lambda c(f), \quad f_i \geq 0 \forall i
\]

where \( s(f) = \log p(f), \quad c(f) = -\log p(m|f) \)

and \( \lambda \) is the regularisation parameter
<table>
<thead>
<tr>
<th></th>
<th>Functions</th>
<th>Gradient $g$</th>
<th>Hessian $H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian noise $c_G$</td>
<td>$\frac{1}{2} | m - Af |^2$</td>
<td>$-A^T(m - Af)$</td>
<td>$A^T A$</td>
</tr>
<tr>
<td>Poisson noise $c_P$</td>
<td>$A^T(Af) - m^T \log(Af)$</td>
<td>$-A^T(\frac{m}{Af} - 1)$</td>
<td>$A^T \text{diag}\left(\frac{m}{(Af)(Af)}\right) A$</td>
</tr>
<tr>
<td>Gaussian prior $s_G$</td>
<td>$-\frac{1}{2} | f |^2$</td>
<td>$-f$</td>
<td>$-I$</td>
</tr>
<tr>
<td>Entropy prior $s_E$</td>
<td>$A^T f - f^T \log(f/f_0)$</td>
<td>$-\log(f/f_0)$</td>
<td>$-\text{diag}(1/f)$</td>
</tr>
<tr>
<td>TV prior $s_T$</td>
<td>$-A^T \beta^{-1} \log(\cosh(\beta Gf))$</td>
<td>$-G^T \tanh(\beta Gf)$</td>
<td>$-G^T \text{diag}(\beta \text{sech}^2(\beta Gf)) G$</td>
</tr>
</tbody>
</table>
Skilling-Bryan Algorithm

- The objective function is approximated by the 2nd order Taylor expansion

\[ q(f + p) \approx q(f) + g_q^T p + \frac{1}{2} p^T H_q p \]

- To allow high-dimensional search, the maximization is done in a subspace.
Subspace Selection

• The step is solved up to the 2nd order:

\[ \mathbf{p} = - (\mathbf{H}_q + \gamma \mathbf{I})^{-1} \mathbf{g}_q, \]
\[ \approx - (\mathbf{I} + \gamma^{-1} \mathbf{H}_q) \mathbf{g}_q, \]
\[ \approx - \mathbf{g}_s + \lambda \mathbf{g}_c + \gamma^{-1} [(\mathbf{H}_s - \lambda \mathbf{H}_c)(\mathbf{g}_s - \lambda \mathbf{g}_c)] \]

• The step lies in the subspace spanned by

\[ \mathbf{g}_c, \mathbf{g}_s, \mathbf{H}_c \mathbf{g}_c, \mathbf{H}_s \mathbf{g}_s, \mathbf{H}_s \mathbf{g}_c \text{ and } \mathbf{H}_c \mathbf{g}_s \]
Subspace Selection

- The basis is summarised by

  \[ e_1 = f g_s, \]
  \[ e_2 = f g_c, \]
  \[ e_3 = f H_c \left( \frac{e_1}{\|g_s\|} - \frac{e_2}{\|g_c\|} \right) + f H_s \left( \frac{e_1}{\|g_s\|} - \frac{e_2}{\|g_c\|} \right) \]

- The gradient directions are weighted by \( f \) to increase the weight for high values (positivity constraint)
Some final details

- The step is solved in the subspace
  \[ p = Ex = x_1 e_1 + x_2 e_2 + x_3 e_3, \]
  under the constraint \( C < C_{\text{aim}} \)
- The new iterate is obtained as
  \[ f_{\text{new}} = f + Ex, \]
  whilst it needs to be protected against negative values (optional rescaling)
Properties of the Skilling-Bryan Algorithm

- Resembles Levenberg-Marguardt Method
- Inherent positivity constraint – thus natural for tomographic reconstruction
- Converges faster than 1st order methods based on line-search
- Each iteration evaluates six projection-backprojection pairs (computational bottleneck)
Conclusions

• We have studied the Bayesian approach for volumetric reconstruction problems
• TV and Spatial Derivative Priors impose sparse solutions in the derivatives
• Extended Skilling-Bryan Optimization
  – Especially for convex objective functions with the positivity constraint
  – Suitable for very high dimensional problems
  – Fast convergence (≈ 20 iterations)
Acknowledgements

• Danai Laksameethanasan, Olivier Renaud, Spencer Shorte

• AUTOMATION project (EU sixth framework programme)

http://www.pfid.org/AUTOMATION/
References

Brandt, S.S. (2006), in Frank J. (editor), Electron Tomography (2nd ed.)