Parameter choice method for total variation regularization

Samuli Siltanen

Department of Mathematics and Statistics
University of Helsinki, Finland
samuli.siltanen@helsinki.fi
http://www.siltanen-research.net

Inverse Days
Lappeenranta University of Technology
December 10, 2015
This is a joint work with

Keijo Hämäläinen, University of Helsinki, Finland

Aki Kallonen, University of Helsinki, Finland

Ville Kolehmainen, University of Eastern Finland

Matti Lassas, University of Helsinki, Finland

Esa Niemi, University of Helsinki, Finland

Kati Niinimäki, University of Eastern Finland
Outline

Total variation regularized tomography

A multiresolution parameter choice method for TV

How about theory?

Bonus material: my new lab
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram
Construction of the sinogram

Unknown: \( f \in \mathbb{R}^{32 \times 32} \)

Data: \( Af \in \mathbb{R}^{49 \times 32} \)
The Singular Value Decomposition $A = UDV^T$ allows analysis of any linear inverse problem.
These phantoms have almost the same sinogram
Naive reconstruction using the Moore-Penrose pseudoinverse; data has 0.1% relative noise

Original phantom, values between zero (black) and one (white)

Naive reconstruction with minimum $-14.9$ and maximum $18.5$
Inverse problem of X-ray tomography: given noisy sinogram, find a stable approximation to $f$

Model space $X = \mathbb{R}^{32 \times 32}$

Data space $Y = \mathbb{R}^{32 \times 49}$

$D(A)$

$A(f) = m$

$A(D(A))$
Constrained Tikhonov regularization
\[ \arg \min_{f \in \mathbb{R}^n_+} \left\{ \|Af - m\|_2^2 + \alpha \|f\|_2^2 \right\} \]

Original phantom

Reconstruction
Relative square norm error 35%
Rudin, Osher and Fatemi (1992): total variation regularization
\[ \arg \min_{f \in \mathbb{R}_+^n} \{ \| Af - m \|_2^2 + \alpha \| \nabla f \|_1 \} \]

Original phantom

TV regularized reconstruction
Relative square norm error 32%
Naive reconstruction using the Moore-Penrose pseudoinverse

Original phantom

Naive reconstruction
Relative square norm error 1246%
This is Professor Keijo Hämäläinen’s X-ray lab
We collected X-ray projection data of a walnut from 1200 directions

Laboratory and data collection by Keijo Hämäläinen and Aki Kallonen, University of Helsinki.

The data is openly available at http://fips.fi/dataset.php, thanks to Esa Niemi and Antti Kujanpää
Reconstructions of a 2D slice through the walnut using filtered back-projection (FBP)

FBP with comprehensive data (1200 projections)

FBP with sparse data (20 projections)
Sparse-data reconstruction of the walnut using non-negative Tikhonov regularization

Filtered back-projection

Constrained Tikhonov regularization

\[
\arg \min_{f \in \mathbb{R}^n_+} \left\{ \|Af - m\|_2^2 + \alpha \|f\|_2^2 \right\}
\]
Sparse-data reconstruction of the walnut using non-negative total variation regularization

\[
\text{arg min}_{f \in \mathbb{R}^n_+} \left\{ \|Af - m\|_2^2 + \alpha \|\nabla f\|_1 \right\}
\]
TV tomography: \[ \arg \min_{f \in \mathbb{R}^n} \{ \| Af - m \|_2^2 + \alpha \| \nabla f \|_1 \} \]

1992 Rudin, Osher & Fatemi: denoise images by taking \( A = I \)
1998 Delaney & Bresler
2001 Persson, Bone & Elmqvist
2003 Kolehmainen, S, Järvenpää, Kaipio, Koistinen, Lassas, Pirttilä & Somersalo (first TV work with measured X-ray data)
2006 Kolehmainen, Vanne, S, Järvenpää, Kaipio, Lassas & Kalke
2006 Sidky, Kao & Pan
2008 Liao & Sapiro
2008 Sidky & Pan
2008 Herman & Davidi
2009 Tang, Nett & Chen
2009 Duan, Zhang, Xing, Chen & Cheng
2010 Bian, Han, Sidky, Cao, Lu, Zhou & Pan
2011 Jensen, Jørgensen, Hansen & Jensen
2011 Tian, Jia, Yuan, Pan & Jiang
2012–present: dozens of articles indicated by Google Scholar
Outline

Total variation regularized tomography

A multiresolution parameter choice method for TV

How about theory?

Bonus material: my new lab
How to choose the regularization parameter in the total variation (TV) functional?

Balancing $\ell^1$ and TV: Clason, Jin & Kunisch 2010

Local variance estimation: Dong, Hintermüller & Rincon-Camacho 2011

Discrepancy principle: Wen & Chan 2012

S-curve method: Kolehmainen, Lassas, Niinemäki & S 2012

Quasi-optimality principle and Hanke-Raus rules: Kindermann, Mutimbu & Resmerita 2013

Including TV parameter in a Karush-Kuhn-Tucker system: Chen, Loli Piccolomini & Zama 2014

Cross validation, Stein’s unbiased risk estimates, L-curve method, …

Practical experience suggests that no single choice rule works perfectly for all applications.

Therefore, it is good to have a collection of rules.
The continuous tomographic model needs to be approximated using a discrete model.

Continuous model:

Discrete model:

In this schematic setup we have 5 projection directions and a 10-pixel detector. Therefore, the number of data points is 50.
The resolution of the discrete model can be freely chosen according to computational resources.

**Continuous model:**

In this schematic setup we have 5 projection directions and a 10-pixel detector. Therefore, the number of data points is 50.

The number of degrees of freedom in the three discrete models below are 16, 64 and 256, respectively.

**Discrete models:**
Intuition: discrete TV reconstructions at different resolutions should converge to a continuous limit 32 × 32
**Intuition:** discrete TV reconstructions at different resolutions should converge to a continuous limit.
**Intuition:** discrete TV reconstructions at different resolutions should converge to a continuous limit
We define the total variation (TV) norm consistently for continuous and discrete cases

Continuous anisotropic TV norm for attenuation coefficient $f : \Omega \to \mathbb{R}$:

$$\int_{\Omega} \left( \left| \frac{\partial f}{\partial x_1} \right| + \left| \frac{\partial f}{\partial x_2} \right| \right) dx.$$ 

Discrete anisotropic TV norm for an image matrix of size $n \times n$:

$$\frac{1}{n} \sum \left| f_{\kappa} - f_{\kappa'} \right|,$$

where the sum is over horizontally and vertically neighboring pixel values $f_{\kappa}$ and $f_{\kappa'}$.

The above is based on this approximate two-dimensional computation:

$$\int_{\Omega} \left| \frac{f(x_1 + \frac{1}{n}, x_2) - f(x_1, x_2)}{1/n} \right| dx \approx (1/n)^2 \sum \left| \frac{f_{\kappa} - f_{\kappa'}}{1/n} \right|,$$

where the sum is over horizontally neighboring pixel values $f_{\kappa}$ and $f_{\kappa'}$. 
Low-noise TV reconstructions of a walnut using several regularization parameters

\( \alpha = 0.001 \) \hspace{1cm} \( \alpha = 1 \) \hspace{1cm} \( \alpha = 1000 \)

Too small \( \alpha \) \hspace{1cm} Just right \( \alpha \) \hspace{1cm} Too large \( \alpha \)
What happens when we compare reconstructions at different resolutions?
Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 0.001$
Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 1$.
Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 1000$.
TV norms of low-noise reconstructions with various resolutions and parameters $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Resolution</th>
<th>$128 \times 128$</th>
<th>$192 \times 192$</th>
<th>$256 \times 256$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td>1.51</td>
<td>2.29</td>
<td>3.64</td>
<td></td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>1.51</td>
<td>2.29</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>1.50</td>
<td>2.23</td>
<td>2.97</td>
<td></td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>1.43</td>
<td>1.85</td>
<td>1.93</td>
<td></td>
</tr>
<tr>
<td>$10^{0}$</td>
<td>1.08</td>
<td>1.11</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td>$10^{1}$</td>
<td>0.78</td>
<td>0.78</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>$10^{2}$</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>$10^{3}$</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>$10^{4}$</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>$10^{5}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$10^{6}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
What happens when we add noise to the data?
5% noise TV reconstructions of a walnut at many resolutions using $\alpha = 0.001$.
5% noise TV reconstructions of a walnut at many resolutions using $\alpha = 10$

128 × 128  192 × 192  256 × 256
5% noise TV reconstructions of a walnut at many resolutions using $\alpha = 10000$
TV norms of reconstructions using various noise levels, resolutions and parameters $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Low noise</th>
<th></th>
<th>5% noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td>1.51 $128^2$</td>
<td>2.29 $192^2$</td>
<td>3.64 $256^2$</td>
<td>2.42 $128^2$</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>1.51 $128^2$</td>
<td>2.29 $192^2$</td>
<td>3.46 $256^2$</td>
<td>2.43 $128^2$</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>1.50 $128^2$</td>
<td>2.23 $192^2$</td>
<td>2.97 $256^2$</td>
<td>2.42 $128^2$</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>1.43 $128^2$</td>
<td>1.85 $192^2$</td>
<td>1.93 $256^2$</td>
<td>2.37 $128^2$</td>
</tr>
<tr>
<td>$10^0$</td>
<td>1.08 $128^2$</td>
<td>1.11 $192^2$</td>
<td>1.11 $256^2$</td>
<td>1.99 $128^2$</td>
</tr>
<tr>
<td>$10^1$</td>
<td>0.78 $128^2$</td>
<td>0.78 $192^2$</td>
<td>0.77 $256^2$</td>
<td>0.86 $128^2$</td>
</tr>
<tr>
<td>$10^2$</td>
<td>0.48 $128^2$</td>
<td>0.48 $192^2$</td>
<td>0.48 $256^2$</td>
<td>0.48 $128^2$</td>
</tr>
<tr>
<td>$10^3$</td>
<td>0.12 $128^2$</td>
<td>0.12 $192^2$</td>
<td>0.12 $256^2$</td>
<td>0.12 $128^2$</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.04 $128^2$</td>
<td>0.04 $192^2$</td>
<td>0.04 $256^2$</td>
<td>0.04 $128^2$</td>
</tr>
<tr>
<td>$10^5$</td>
<td>0 $128^2$</td>
<td>0 $192^2$</td>
<td>0 $256^2$</td>
<td>0 $128^2$</td>
</tr>
<tr>
<td>$10^6$</td>
<td>0 $128^2$</td>
<td>0 $192^2$</td>
<td>0 $256^2$</td>
<td>0 $128^2$</td>
</tr>
</tbody>
</table>

[Niinimäki, Lassas, Hämäläinen, Kallonen, Kolehmainen, Niemi & S, submitted]
Outline

Total variation regularized tomography

A multiresolution parameter choice method for TV

How about theory?

Bonus material: my new lab
Let $D$ be the square $[0, 1]^2 \subset \mathbb{R}^2$. Use anisotropic $BV(D)$ norm
\[
\|u\|_{BV} = \|u\|_{L^1} + V(u) = \|u\|_{L^1} + \int_D \left( \left| \frac{\partial u(x)}{\partial x_1} \right| + \left| \frac{\partial u(x)}{\partial x_2} \right| \right) dx.
\]

Assume either (A) or (B) about the linear operator $A$:

(A) $A : L^2(D) \to L^2(\Omega)$ is compact and $A : L^1(D) \to \mathcal{D}'(\Omega)$ is continuous with some open and bounded set $\Omega \subset \mathbb{R}^2$,

(B) $A : L^1(D) \to \mathbb{R}^M$ is bounded.

Define $S_{\infty} : BV(D) \to \mathbb{R}$ and $S_j : BV(D) \to \mathbb{R} \cup \{\infty\}$ by
\[
S_{\infty}(u) = \|Au - m\|^2_{L^2(\Omega)} + \alpha_1 \|u\|_{L^1(D)} + \alpha V(u)
\]
with positive regularization parameters $\alpha_1 > 0$ and $\alpha > 0$, and
\[
S_j(u) = \begin{cases} 
S_{\infty}(u), & \text{for } u \in \text{Range}(T_j), \\
\infty, & \text{for } u \notin \text{Range}(T_j).
\end{cases}
\]

$T_j$ projects to functions piecewise constant on $2^j \times 2^j$ pixel grid.
Our main theorem ensures the convergence of regularized solutions as resolution grows

- There exists a minimizer $\tilde{u}_j \in \arg\min (S_j)$ for all $j = 1, 2, 3, \ldots$
- There exists a minimizer $\tilde{u}_\infty \in \arg\min (S_\infty)$.
- Any sequence $\tilde{u}_j \in \arg\min (S_j)$ of minimizers has a subsequence $\tilde{u}_{j(\ell)}$ that converges weakly in $BV(D)$ to some limit $w \in BV(D)$. Furthermore, $\lim_{\ell \to \infty} V(\tilde{u}_{j(\ell)}) = V(w)$.
- The limit $w$ is a minimizer: $w \in \arg\min (S_\infty)$.

[Niinimäki, Lassas, Hämäläinen, Kallonen, Kolehmainen, Niemi & S, submitted]
There are some related results in the literature

1992 Vainikko: *On the discretization and regularization of ill-posed problems with noncompact operators*


These works consider TV functionals and $\Gamma$-convergence when discretization is refined:

2009 Chambolle, Giacomini & Lussardi

2012 Cai, Dong, Osher & Shen

2012 Gennip & Bertozzi

2013 Bellettini, Chambolle & Goldman

2013 Trillos & Slepčev

However, they do not consider measurement operators, which are crucial in inverse problems.
How to prove the main theorem?

The proof is an analysis of $\Gamma$-convergence of functionals $S_j$ to $S_\infty$. However, the choice of topologies is very delicate. For details, see http://arxiv.org/abs/0902.2313v1.

Note: related $\Gamma$-convergence results of $TV$ functionals are given in [Chambolle, Giacomini and Lussardi 2009], but they do not consider measurement operators.

This approximation lemma serves as the foundation of the proof:

**Lemma.** For all $u \in BV(D)$ and $\varepsilon > 0$ there exists $j > 0$ and a function $u'$, piecewise constant in the dyadic $2^j \times 2^j$ grid, such that

$$\|u - u'\|_{L^1(D)} + |V(u) - V(u')| < \varepsilon.$$

Recall that

$$V(u) = \int_D \left( \left| \frac{\partial u(x)}{\partial x_1} \right| + \left| \frac{\partial u(x)}{\partial x_2} \right| \right) dx.$$
We need to move from triangulation-based to pixel-based approximation

[Bělík and Luskin 2003]: the desired inequality holds with PW constant functions in a fine triangularization.

However, we need to work with dyadic $2^j \times 2^j$ pixel grids.
We surround any triangle vertex (blue dot) with a “pixel cluster” neighborhood (gray box).
Refine the grid outside clusters so that pixel-wise polygonal chains (on pink) connect the clusters.
Using the anisotropic BV norm reduces the approximation to estimating small intervals

The difference between the BV norms of the piecewise constant functions $v_1$ and $v_2$ comes entirely from jumps over the two red vertical intervals below.
What can we say about the proposed method?

Benefits of our multiresolution TV parameter choice method:
- simple definition,
- easy implementation, and
- no need of a priori information.

Also, it seems to perform well for real tomographic data.

As a downside we mention that several reconstructions need to be computed. Also, it is still unclear why the method works so nicely: if there is convergence for any $\alpha$ in theory, what is the instability we are observing?

The method can be tried out with 3D tomography (it works!) and with other inverse problems and regularizers.
Outline

Total variation regularized tomography

A multiresolution parameter choice method for TV

How about theory?

Bonus material: my new lab
This is my new X-ray laboratory at University of Helsinki
Lego robot under imaging
Lego robot under imaging
This is simply a SIRT reconstruction from 360 views using the ASTRA toolbox

Computation and visualization by Topias Rusanen
Thank you for your attention!

http://www.siltanen-research.net
All Matlab codes freely available at this site!

Part I: Linear Inverse Problems
1 Introduction
2 Naïve reconstructions and inverse crimes
3 Ill-Posedness in Inverse Problems
4 Truncated singular value decomposition
5 Tikhonov regularization
6 Total variation regularization
7 Besov space regularization using wavelets
8 Discretization-invariance
9 Practical X-ray tomography with limited data
10 Projects

Part II: Nonlinear Inverse Problems
11 Nonlinear inversion
12 Electrical impedance tomography
13 Simulation of noisy EIT data
14 Complex geometrical optics solutions
15 A regularized D-bar method for direct EIT
16 Other direct solution methods for EIT
17 Projects