Sparse X-ray tomography for medical imaging

Samuli Siltanen

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

Seminar talk
University of Bologna, March 9, 2017
Finnish Centre of Excellence in Inverse Problems Research

http://wiki.helsinki.fi/display/inverse/Home
Outline

**X-ray imaging**

Mathematical model of X-ray attenuation

Tomographic imaging with sparse data

A multiresolution parameter choice method for TV

Industrial case study: low-dose 3D dental X-ray imaging

Hospital case study: diagnosing osteoarthritis
Wilhelm Conrad Röntgen invented X-rays and was awarded the first Nobel Prize in Physics in 1901.
Godfrey Hounsfield and Allan McLeod Cormack developed X-ray tomography.

Hounsfield (top) and Cormack received Nobel prizes in 1979.
Reconstruction of a function from its line integrals was first invented by Johann Radon in 1917

\[ f(P) = -\frac{1}{\pi} \int_0^\infty \frac{dF_p(q)}{q} \]

Johann Radon (1887-1956)
Contrast-enhanced CT of abdomen, showing liver metastases
Head CT can be used for detecting and monitoring brain hemorrhage.

Source: LearningRadiology.com
Unusual variant of the Nutcracker Fracture of the calcaneus and tarsal navicular

[Image: Axial slice of the right foot, Another axial slice, Sagittal slice, 3D render]

[Gajendran, Yoo & Hunter, Radiology Case Reports 3 (2008)]]
Show lotus root video!

www.youtube.com/watch?v=eWwD_EZuzBl&t=7s
Outline

X-ray imaging

Mathematical model of X-ray attenuation

Tomographic imaging with sparse data

A multiresolution parameter choice method for TV

Industrial case study: low-dose 3D dental X-ray imaging

Hospital case study: diagnosing osteoarthritis
X-ray intensity attenuates inside matter, here shown with two homogeneous blocks

https://www.youtube.com/watch?v=Z_IBFQcn0l8
A digital X-ray detector counts how many photons arrive at each pixel.

X-ray source

1000

Detector

1000

photon count
Adding material between the source and detector reveals the exponential X-ray attenuation law.
We take logarithm of the photon counts to compensate for the exponential attenuation law.
Final calibration step is to subtract the logarithms from the empty space value (here 6.9)

<table>
<thead>
<tr>
<th>photon count</th>
<th>log</th>
<th>line integral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>6.9</td>
<td>0.0</td>
</tr>
<tr>
<td>500</td>
<td>6.2</td>
<td>0.7</td>
</tr>
<tr>
<td>250</td>
<td>5.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>
After calibration we are observing how much attenuating matter the X-ray encounters

https://www.youtube.com/watch?v=TKqcrDGPsAI
This sweeping movement is the data collection mode of first-generation CT scanners

https://www.youtube.com/watch?v=TbLaQo3rgEE
Rotating around the object allows us to form the so-called *sinogram*

https://www.youtube.com/watch?v=5Vyc1TzmNl8
This is an illustration of the standard reconstruction by filtered back-projection

https://www.youtube.com/watch?v=ddZeLNh9aac
Outline

X-ray imaging

Mathematical model of X-ray attenuation

Tomographic imaging with sparse data

A multiresolution parameter choice method for TV

Industrial case study: low-dose 3D dental X-ray imaging

Hospital case study: diagnosing osteoarthritis
Let us study a simple two-dimensional example of tomographic imaging.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

X-ray source → 13 (4+4+5) → Detector
A projection image is produced by parallel X-rays and several detector pixels (here three pixels)

\[
\begin{array}{ccc}
4 & 4 & 5 \\
1 & 3 & 4 \\
1 & 0 & 2 \\
\end{array}
\]

- 13 \((=4+4+5)\)
- 8 \((=1+3+4)\)
- 3 \((=1+0+2)\)
For tomographic imaging it is essential to record projection images from different directions.
The length of X-rays traveling inside each pixel is important, thus here the square roots
The direct problem of tomography is to find the projection images from known tissue
The inverse problem of tomography is to reconstruct the interior from X-ray data.
The limited-angle problem is harder than the full-angle problem.

9 unknowns, 6 equations

9 unknowns, 11 equations
In limited-angle imaging, different objects may produce the same data.

Mathematically this means that the matrix $A$ has nontrivial kernel.
We write the reconstruction problem in matrix form

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>$f_4$</td>
<td>$f_7$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_2$</td>
<td>$f_5$</td>
<td>$f_8$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_3$</td>
<td>$f_6$</td>
<td>$f_9$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\begin{bmatrix}
  m_1 \\
m_2 \\
m_3 \\
m_4 \\
m_5 \\
m_6
\end{bmatrix} = \begin{bmatrix}
  f_1 \\
f_2 \\
f_3 \\
f_4 \\
f_5 \\
f_6 \\
f_7 \\
f_8 \\
f_9
\end{bmatrix}$,

Measurement model: $m = Af$
This is the matrix equation related to the above measurement

\[
\begin{bmatrix}
  m_1 \\
  m_2 \\
  m_3 \\
  m_4 \\
  m_5 \\
  m_6
\end{bmatrix} =
\begin{bmatrix}
  0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 \\
  \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} \\
  0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 \\
  1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
  0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
  0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3 \\
  f_4 \\
  f_5 \\
  f_6 \\
  f_7 \\
  f_8 \\
  f_9
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_1 \\
  \varepsilon_2 \\
  \varepsilon_3 \\
  \varepsilon_4 \\
  \varepsilon_5 \\
  \varepsilon_6
\end{bmatrix}
\]
Construction of the sinogram

Angle of X-rays: 3.0 degrees
Construction of the sinogram

Angle of X-rays: 12.2 degrees
Construction of the sinogram

Angle of X-rays: 21.5 degrees
Construction of the sinogram

Angle of X-rays: 30.7 degrees
Construction of the sinogram

Angle of X-rays: 39.9 degrees
Construction of the sinogram

Angle of X-rays: 49.2 degrees
Construction of the sinogram

Angle of X-rays: 58.4 degrees
Construction of the sinogram

Angle of X-rays: 67.6 degrees
Construction of the sinogram

Angle of X-rays: 76.8 degrees
Construction of the sinogram

Angle of X-rays: 86.1 degrees
Construction of the sinogram

Angle of X-rays: 95.3 degrees
Construction of the sinogram

Angle of X-rays: 104.5 degrees
Construction of the sinogram

Angle of X-rays: 113.8 degrees
Construction of the sinogram

Angle of X-rays: 123.0 degrees
Construction of the sinogram

Angle of X-rays: 132.2 degrees
Construction of the sinogram

Angle of X-rays: 141.5 degrees
Construction of the sinogram

Angle of X-rays: 150.7 degrees
Construction of the sinogram

Angle of X-rays: 159.9 degrees
Construction of the sinogram

Angle of X-rays: 169.2 degrees
Construction of the sinogram

Angle of X-rays: 178.4 degrees
Construction of the sinogram

Angle of X-rays: 187.6 degrees
Construction of the sinogram

Angle of X-rays: 196.8 degrees
Construction of the sinogram

Angle of X-rays: 206.1 degrees
Construction of the sinogram

Angle of X-rays: 215.3 degrees
Construction of the sinogram

Angle of X-rays: 224.5 degrees
Construction of the sinogram

Angle of X-rays: 233.8 degrees
Construction of the sinogram

Angle of X-rays: 243.0 degrees
Construction of the sinogram

Angle of X-rays: 252.2 degrees
Construction of the sinogram

Angle of X-rays: 261.5 degrees
Construction of the sinogram

Angle of X-rays: 270.7 degrees
Construction of the sinogram

Angle of X-rays: 279.9 degrees
Construction of the sinogram

Angle of X-rays: 289.2 degrees
Construction of the sinogram

Angle of X-rays: 298.4 degrees
Construction of the sinogram

Angle of X-rays: 307.6 degrees
Construction of the sinogram

Angle of X-rays: 316.8 degrees
Construction of the sinogram

Angle of X-rays: 326.1 degrees
Construction of the sinogram

Angle of X-rays: 335.3 degrees
Construction of the sinogram

Angle of X-rays: 344.5 degrees
Construction of the sinogram

Angle of X-rays: 353.8 degrees
We have object and data for the inverse problem

\[ f \in \mathbb{R}^{32 \times 32} \]

\[ A \]

\[ A f \in \mathbb{R}^{49 \times 39} \]
Illustration of the ill-posedness of tomography

Difference 0.00254
Illustration of the ill-posedness of tomography
Illustration of the ill-posedness of tomography

Difference 0.00004
Singular Value Decomposition (SVD) of the above tomographic measurement matrix $A$

Nonzero elements of matrix $A$:

Singular values of matrix $A$: diagonal of $D$ in $A = UDV^T$
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 $18.5$
Standard Tikhonov regularization

\[ \text{arg min}_{f \in \mathbb{R}^n} \left\{ \| Af - m \|_2^2 + \alpha \| f \|_2^2 \right\} \]

Original phantom

Reconstruction

Relative square norm error 35\%
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 13%
Constrained total variation (TV) regularization

$$\arg \min_{f \in \mathbb{R}^n_+} \left\{ \| Af - m \|_2^2 + \alpha \left( \| L_h f \|_1 + \| L_v f \|_1 \right) \right\}$$

Original phantom

TV regularized reconstruction
Relative square norm error 3%
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

\[ \text{Filtered back-projection} \]

\[ \text{Constrained Tikhonov regularization} \]

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

$$\text{arg min}_{f \in \mathbb{R}^n_+} \{ \| Af - m \|_2^2 + \alpha \| \nabla f \|_1 \}$$
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
Sparse-data reconstruction of the walnut using Total Generalized Variation (TGV)

Filtered back-projection

TGV: thanks to Kristian Bredies!
Daubechies, Defrise and de Mol introduced a revolutionary method in 2004

The sparsity-promoting iteration works like this:

\[ f_n = S_\mu(f_{n-1} + A^T(m - Af_{n-1})) , \]

where the soft-thresholding operator \( S_\mu \) is defined by

\[ S_\mu(g) = \sum_{j \in J} S_\mu(\langle g, \psi_j \rangle)\psi_j(x). \]

Here \( \psi_j \) are for example wavelets or shearlets, forming a frame, and

\[ S_\mu(x) = \begin{cases} 
 x + \frac{\mu}{2} & \text{if } x \leq -\frac{\mu}{2} \\
 0 & \text{if } |x| < \frac{\mu}{2} \\
 x - \frac{\mu}{2} & \text{if } x \geq \frac{\mu}{2}. 
\end{cases} \]
Illustration of the Haar wavelet transform
Sparse-data reconstruction of the walnut using Haar wavelet sparsity

Filtered back-projection

Constrained Besov regularization

\[
\arg \min_{f \in \mathbb{R}^n} \left\{ \| Af - m \|_2^2 + \alpha \| f \|_{B_{11}} \right\}
\]
Sparse-data reconstruction of the walnut using Daubechies 2 wavelet sparsity

Filtered back-projection

Constrained Besov regularization

$$\arg\min_{f \in \mathbb{R}_+^n} \left\{ \|Af - m\|_2^2 + \alpha \|f\|_{B_{11}} \right\}$$
Sparse-data reconstruction of the walnut using shearlet sparsity

Filtered back-projection

Thanks to Gitta Kutyniok!
http://www.shearlab.org/
Sparse-data reconstruction of the lotus using TV regularization and shearlet sparsity
Outline

X-ray imaging

Mathematical model of X-ray attenuation

Tomographic imaging with sparse data

A multiresolution parameter choice method for TV

Industrial case study: low-dose 3D dental X-ray imaging

Hospital case study: diagnosing osteoarthritis
This part of the talk is a joint work with

Kati Niinimäki, University Paris-Sud, France

Matti Lassas, University of Helsinki, Finland

Keijo Hämäläinen, University of Helsinki, Finland

Aki Kallonen, University of Helsinki, Finland

Ville Kolehmainen, University of Eastern Finland

Esa Niemi, University of Helsinki, Finland
How to choose the regularization parameter in the total variation (TV) functional?

**Heuristics:** Rullgård 2008

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

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

**Discrepancy principle:** Wen & Chan 2012

**S-curve method:** Kolehmainen, Lassas, Niinimäki & S 2012

**Dantzig estimation:** Frick, Marниц & Munk 2012

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

**KKT system:** Chen, Loli Piccolomini & Zama 2014

**Discrepancy principle:** Toma, Sixou & Peyrin 2015

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

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:**

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

**Discrete model:**
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:**
We define the total variation (TV) norm consistently for continuous and discrete cases

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

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

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:

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

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 \( \alpha \)

Computations by Kati Niinimäki using a primal-dual interior point method.
Low-noise TV reconstructions of a walnut using several regularization parameters

$\alpha = 0.001$  $\alpha = 1$  $\alpha = 1000$

Too small $\alpha$  Just right $\alpha$  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 $\alpha = 1$.

$128 \times 128$  $192 \times 192$  $256 \times 256$
Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 1000$. 

128 × 128  
192 × 192  
256 × 256
TV norms of low-noise reconstructions with various resolutions and parameters $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Resolution</th>
<th>128 x 128</th>
<th>192 x 192</th>
<th>256 x 256</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td></td>
<td>1.51</td>
<td>2.29</td>
<td>3.64</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td></td>
<td>1.51</td>
<td>2.29</td>
<td>3.46</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td></td>
<td>1.50</td>
<td>2.23</td>
<td>2.97</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td></td>
<td>1.43</td>
<td>1.85</td>
<td>1.93</td>
</tr>
<tr>
<td>$10^{0}$</td>
<td></td>
<td>1.08</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td>$10^{1}$</td>
<td></td>
<td>0.78</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>$10^{2}$</td>
<td></td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>$10^{3}$</td>
<td></td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>$10^{4}$</td>
<td></td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>$10^{5}$</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$10^{6}$</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### TV norms of low-noise reconstructions with various resolutions and parameters $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Resolution\n128 × 128</th>
<th>192 × 192</th>
<th>256 × 256</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td>1.51</td>
<td>2.29</td>
<td>3.64</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>1.51</td>
<td>2.29</td>
<td>3.46</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>1.50</td>
<td>2.23</td>
<td>2.97</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>1.43</td>
<td>1.85</td>
<td>1.93</td>
</tr>
<tr>
<td>$10^{0}$</td>
<td>1.08</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td>$10^{1}$</td>
<td>0.78</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>$10^{2}$</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>$10^{3}$</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>$10^{4}$</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>$10^{5}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$10^{6}$</td>
<td>0</td>
<td>0</td>
<td>0</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 $\alpha = 10$
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>5% noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128$^2$</td>
<td>192$^2$</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>1.51</td>
<td>2.29</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>1.51</td>
<td>2.29</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>1.50</td>
<td>2.23</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>1.43</td>
<td>1.85</td>
</tr>
<tr>
<td>$10^0$</td>
<td>1.08</td>
<td>1.11</td>
</tr>
<tr>
<td>$10^1$</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>$10^2$</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>$10^3$</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>$10^5$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$10^6$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

[Niinimäki, Lassas, Hämäläinen, Kallonen, Kolehmainen, Niemi & S, SIAM Journal on Imaging Sciences 2016]
How about theory?
There are some related results in the literature

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

We use geometric arguments similar to those here:

These works consider TV functionals and $\Gamma$-convergence when discretization is refined, but without a measurement operator:
2009 Chambolle, Giacomini & Lussardi
2012 Gennip & Bertozzi
2013 Bellettini, Chambolle & Goldman
2013 Trillos & Slepčev

This paper achieves a result analogous with ours, using wavelet frames in the finite-dimensional functionals:
2012 Cai, Dong, Osher & Shen
Assumptions on the linear forward map $A$

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

(A) $A : L^2(D) \to L^2(\Omega)$ is compact and $A : L^1(D) \to D'(\Omega)$ is continuous with some open and bounded set $\Omega \subset \mathbb{R}^2$. This covers the case of classical Radon transform with image domain $D$ and sinogram domain $\Omega$. We denote the set of distributions by $D'(\Omega)$.

(B) $A : L^1(D) \to \mathbb{R}^M$ is bounded. This covers the practically important discrete pencil beam model of tomographic measurements.
Definition of discrete and continuous regularization functionals

Let $D$ be the square $[0, 1]^2 \subseteq \mathbb{R}^2$. Use anisotropic $BV(D)$ norm

$$
\|u\|_{BV} = \|u\|_{L^1} + V(u) = \|u\|_{L^1} + \int_D \left( \left| \frac{\partial u(x)}{\partial x_1} \right| + \left| \frac{\partial u(x)}{\partial x_2} \right| \right) dx.
$$

Define $S_\infty : BV(D) \to \mathbb{R}$ and $S_j : BV(D) \to \mathbb{R} \cup \{\infty\}$ by

$$
S_\infty(u) = \|A u - m\|_{L^2(\Omega)}^2 + \alpha_1 \|u\|_{L^1(D)} + \alpha V(u)
$$

with positive regularization parameters $\alpha_1 > 0$ and $\alpha > 0$, and

$$
S_j(u) = \begin{cases} 
S_\infty(u), & \text{for } u \in \text{Range}(T_j), \\
\infty, & \text{for } u \notin \text{Range}(T_j).
\end{cases}
$$

Linear operator $T_j$ projects to functions that are piecewise constant on a regular $2^j \times 2^j$ square 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, SIAM Journal on Imaging Sciences 2016]
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 about the noise amplitude.

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

Downside: 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

X-ray imaging

Mathematical model of X-ray attenuation

Tomographic imaging with sparse data

A multiresolution parameter choice method for TV

Industrial case study: low-dose 3D dental X-ray imaging

Hospital case study: diagnosing osteoarthritis
The VT device was developed in 2001–2012 by

Nuutti Hyvönen
Seppo Järvenpää
Jari Kaipio
Martti Kalke
Petri Koistinen
Ville Kolehmainen
Matti Lassas
Jan Moberg
Kati Niinimäki
Juha Pirttilä
Maaria Rantala
Eero Saksman
Henri Setälä
Erkki Somersalo
Antti Vanne
Simopekka Vänskä
Richard L. Webber
Application: dental implant planning, where a missing tooth is replaced with an implant.
Panoramic dental imaging shows all the teeth simultaneously.

Panoramic imaging was invented by Yrjö Veli Paatero in the 1950's.
This is the classical imaging procedure of the panoramic X-ray device

https://www.youtube.com/watch?v=QFTXegPxC4U
The resulting image shows a sharp layer positioned inside the dental arc.
Nowadays, a digital panoramic imaging device is standard equipment at dental clinics. A panoramic dental image offers a general overview showing all teeth and other structures simultaneously. Panoramic images are not suitable for dental implant planning because of unavoidable geometric distortion.
We reprogram the panoramic X-ray device so that it collects projection data by scanning https://www.youtube.com/watch?v=motthjiP8ZQ
We reprogram the panoramic X-ray device so that it collects projection data by scanning

Number of projection images: 11
Angle of view: 40 degrees
Image size: $1000 \times 1000$ pixels

The unknown vector $f$ has 7,000,000 elements.
CBCT imaging gives 100 times more radiation than VT reconstruction


Outline

X-ray imaging

Mathematical model of X-ray attenuation

Tomographic imaging with sparse data

A multiresolution parameter choice method for TV

Industrial case study: low-dose 3D dental X-ray imaging

Hospital case study: diagnosing osteoarthritis
This part of the talk is a joint work with

Tatiana Bubba, University of Helsinki, Finland

Sakari Karhula, Oulu University Hospital, Finland

Juuso Ketola, Oulu University Hospital, Finland

Miika T. Nieminen, University of Oulu, Finland

Zenith Purisha, University of Helsinki, Finland

Simo Saarakkala, Oulu University Hospital, Finland
Normal Knee

Osteoarthritis

Image by Bruce Blaus, CC BY-SA 4.0
https://commons.wikimedia.org/w/index.php?curid=44968165
We consider small specimens of human bone imaged using microtomography.

Slice of 3D reconstruction by FDK based on 596 angles. Three-dimensional structure.
We pick out a smaller region of interest for osteoarthritis analysis.

Slice of 3D reconstruction by FDK based on 596 angles

Slice of 3D region of interest after binary thresholding
We use several numerical quality measures applied to segmented three-dimensional bone structure:

- Trabecular thickness (Tb.Th)
- Trabecular separation (Tb.Sp)

[Source: Bouxsein, Boyd, Christiansen, Guldberg, Jepsen, & Müller 2010]
We use several numerical quality measures applied to segmented three-dimensional bone structure.

\[ \text{BV/TV} = 17.7\% \]
\[ \text{BV/TV} = 6.5\% \]
\[ \text{BV/TV} = 4.2\% \]

[Bouxsein, Boyd, Christiansen, Guldberg, Jepsen, & Müller 2010]
The goal is to reduce measurement time by recording fewer radiographs.

3D FDK reconstruction based on 40 angles

3D shearlet-sparsity reconstruction based on 40 angles
# Results from FDK reconstructions

<table>
<thead>
<tr>
<th></th>
<th>Healthy bone</th>
<th>Osteoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>596</td>
<td>41.0%</td>
<td>8.1</td>
</tr>
<tr>
<td>120</td>
<td>41.3%</td>
<td>6.5</td>
</tr>
<tr>
<td>60</td>
<td>41.7%</td>
<td>4.0</td>
</tr>
<tr>
<td>40</td>
<td>44.2%</td>
<td>3.4</td>
</tr>
</tbody>
</table>
## Results from 3D shearlet-sparsity reconstructions

<table>
<thead>
<tr>
<th></th>
<th>BV/TV</th>
<th>Tb.Th</th>
<th>Tb.N</th>
<th>Tb.Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy bone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>596</td>
<td>41.0%</td>
<td>8.1</td>
<td>0.05</td>
<td>15.3</td>
</tr>
<tr>
<td>120</td>
<td>41.4%</td>
<td>8.9</td>
<td>0.05</td>
<td>16.0</td>
</tr>
<tr>
<td>60</td>
<td>41.2%</td>
<td>8.8</td>
<td>0.05</td>
<td>16.0</td>
</tr>
<tr>
<td>40</td>
<td>40.9%</td>
<td>9.0</td>
<td>0.05</td>
<td>15.8</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>596</td>
<td>19.0%</td>
<td>5.9</td>
<td>0.03</td>
<td>21.0</td>
</tr>
<tr>
<td>120</td>
<td>20.3%</td>
<td>6.6</td>
<td>0.03</td>
<td>21.7</td>
</tr>
<tr>
<td>60</td>
<td>20.8%</td>
<td>6.8</td>
<td>0.03</td>
<td>21.5</td>
</tr>
<tr>
<td>40</td>
<td>21.2%</td>
<td>7.1</td>
<td>0.03</td>
<td>21.8</td>
</tr>
</tbody>
</table>
Thank you for your attention!
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
Another great resource is Per Christian Hansen’s 3D tomography toolbox TVreg

**TVreg**: Software for 3D Total Variation Regularization (for Matlab Version 7.5 or later), developed by Tobias Lindstrøm Jensen, Jakob Heide Jørgensen, Per Christian Hansen, and Søren Holdt Jensen.

Website: http://www2.imm.dtu.dk/ pcha/TVReg/
These books are recommended for learning the mathematics of practical X-ray tomography

1983 Deans: The Radon Transform and Some of Its Applications
1986 Natterer: The mathematics of computerized tomography
1988 Kak & Slaney: Principles of computerized tomographic imaging
1996 Engl, Hanke & Neubauer: Regularization of inverse problems
1998 Hansen: Rank-deficient and discrete ill-posed problems
2001 Natterer & Wübbeling: Mathematical Methods in Image Reconstruction
2008 Buzug: Computed Tomography: From Photon Statistics to Modern Cone-Beam CT
2008 Epstein: Introduction to the mathematics of medical imaging
2010 Hansen: Discrete inverse problems
2012 Mueller & S: Linear and Nonlinear Inverse Problems with Practical Applications
2014 Kuchment: The Radon Transform and Medical Imaging
There are three main approaches for noise-robust solution of inverse problems

1. **Variational regularization.** Regularized solution is the minimizer of $\|A(f) - m\|_Y^2 + \alpha R(f)$.
   - The same code applies to many problems.
   - Repeated solution of direct problem needed.
   - Can get stuck in local minima.

2. **Problem-specific regularization.** Derive an analytical formula for recovering the unknown.
   - Can deal efficiently with a specific nonlinearity.
   - Each formula applies to only one inverse problem.

3. **Bayesian inversion.** In the measurement equation $m = A(x) + \varepsilon$, model $x, m, \varepsilon$ as random vectors. The solution is the *posterior distribution* $\pi(x|m) = \pi(x) \pi(m|x)/\pi(m)$.
   - Very flexible framework, includes uncertainty quantification.
   - Computationally heavy.
There are many computational approaches for computing the minimum

**Primal-dual algorithms** Chambolle, Chan, Chen, Esser, Golub, Mulet, Nesterov, Zhang

**Thresholding** Candès, Chambolle, Chaux, Combettes, Daubechies, Defrise, DeMol, Donoho, Pesquet, Starck, Teschke, Vese, Wajs

**Bregman iteration** Cai, Burger, Darbon, Dong, Goldfarb, Mao, Osher, Shen, Xu, Yin, Zhang

**Splitting approaches** Chan, Esser, Fornasier, Goldstein, Langer, Osher, Schönlieb, Setzer, Wajs

**Nonlocal TV** Bertozzi, Bresson, Burger, Chan, Lou, Osher, Zhang

We found that **quadratic programming** works well for us.
Quadratic programming (QP) for TV regularization

The minimizer of the functional

$$\arg\min_{f \in \mathbb{R}^n_+} \left\{ \|Af - m\|_2^2 + \alpha \|L_H f\|_1 + \alpha \|L_V f\|_1 \right\}$$

can be transformed into the standard form

$$\arg\min_{z \in \mathbb{R}^{5n}} \left\{ \frac{1}{2} z^T Q z + c^T z \right\}, \quad z \geq 0, \quad Ez = b,$$

where $Q$ is symmetric and $E$ implements equality constraints.

Large-scale primal-dual interior point QP method was developed in Kolehmainen, Lassas, Niinimäki & S (2012) and Hämäläinen, Kallonen, Kolehmainen, Lassas, Niinimäki & S (2013).
Reduction to $\arg \min_{z \in \mathbb{R}^{5n}} \left\{ \frac{1}{2} z^T Q z + c^T z \right\}$

Denote horizontal and vertical differences by

$$L_H f = u_H^+ - u_H^- \quad \text{and} \quad L_V f = u_V^+ - u_V^-,$$

where $u_H^+, u_V^+ \geq 0$. TV minimization is now

$$\arg \min_{f \in \mathbb{R}^n_+} \left\{ f^T A^T A f - 2 f^T A^T m + \alpha 1^T (u_H^+ + u_H^- + u_V^+ + u_V^-) \right\},$$

where $1 \in \mathbb{R}^n$ is vector of all ones. Further, we denote

$$z = \begin{bmatrix} f \\ u_H^+ \\ u_H^- \\ u_V^+ \\ u_V^- \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{1}{\sigma^2} A^T A & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad c = \begin{bmatrix} -2 A^T m \\ \alpha 1 \\ \alpha 1 \\ \alpha 1 \end{bmatrix}. $$
Projected Barzilai-Borwein minimization

For $f \in \mathbb{R}^n$, denote $\|f\|_\beta := \sum_{i=1}^{n} \sqrt{(f_i)^2 + \beta}$, with a small parameter $\beta > 0$. We minimize

$$G_\beta(f) := \frac{1}{2} \|Af - \tilde{g}\|_2^2 + \alpha (\|L_Hf\|_\beta + \|L_Vf\|_\beta),$$

with a non-negativity constraint:

$$f^{k+1} = P \left( f^k - \lambda_k \nabla G_\beta(f^k) \right), \quad k = 0, \ldots, k_{\text{max}} - 1.$$

The step size is

$$\lambda_k = \frac{(f^k - f^{k-1})^T(f^k - f^{k-1})}{(f^k - f^{k-1})^T(\nabla G_\beta(f^k) - \nabla G_\beta(f^{k-1}))}$$

and $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the projection

$$(P(f))_i = \begin{cases} f_i & \text{if } f_i \geq 0 \\ 0 & \text{if } f_i < 0 \end{cases}, \quad i = 1, \ldots, n.$$
We present another real-data example involving an arrangement of 10 sugarcubes.

We use 120 projections. Shown above is a reconstruction using FBP.
**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>256 $\times$ 256</th>
<th>512 $\times$ 512</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td>0.85</td>
<td>2.30</td>
<td>7.40</td>
<td></td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>0.84</td>
<td>2.30</td>
<td>6.80</td>
<td></td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>0.93</td>
<td>2.20</td>
<td>7.00</td>
<td></td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>0.91</td>
<td>1.80</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>$10^{0}$</td>
<td>0.76</td>
<td>0.91</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$10^{1}$</td>
<td>0.53</td>
<td>0.54</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>$10^{2}$</td>
<td>0.37</td>
<td>0.37</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>$10^{3}$</td>
<td>0.27</td>
<td>0.30</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>$10^{4}$</td>
<td>0.17</td>
<td>0.14</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>$10^{5}$</td>
<td>0.11</td>
<td>0.33</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>$10^{6}$</td>
<td>0.14</td>
<td>0.28</td>
<td>1.90</td>
<td></td>
</tr>
</tbody>
</table>
TV reconstruction using rounded absolute-value function and projected Barzilai-Borwein

We used 120 projections, parameter $\alpha = 10$, and resolution $512 \times 512$. Data collected by Aki Kallonen, computations by Esa Niemi.
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.

The approximation lemma on the right serves as the foundation of the proof. Generalizations of the result to higher dimension or to other TV norms would require modifying this key lemma.

**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

\[ v_1 = a \]
\[ v_1 = b \]

\[ v_2 = a \]
\[ v_2 = b \]

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.
Total Generalized Variation Regularization

The usual Total Variation (TV) seminorm is

\[ TV(u) = \int_{\Omega} d|\nabla u| = \sup \left\{ \int_{\Omega} u \text{ div } v \, dx \mid v \in C^1_c(\Omega, \mathbb{R}^d), \|v\|_{\infty} \leq 1 \right\}. \]

One can also consider higher-order derivatives:

\[ TV^2(u) = \sup \left\{ \int_{\Omega} u \text{ div}^2 v \, dx \mid v \in C^2_c(\Omega, S^{d \times d}), \|v\|_{\infty} \leq 1 \right\}, \]

where \( S^{d \times d} \) denotes the set of symmetric matrices.

Total generalized variation [Bredies, Kunisch & Pock 2011] is defined as

\[ TGV^2_{\beta}(u) = \sup \left\{ \int_{\Omega} u \text{ div}^2 v \, dx \mid v \in C^2_c(\Omega, S^{d \times d}), \right. \]

\[ \left. \|v\|_{\infty} \leq \beta, \|\text{div } v\|_{\infty} \leq 1 \right\}. \]