Sparse tomography

Samuli Siltanen

Department of Mathematics and Statistics
University of Helsinki, Finland

Seminar talk at Shanghai Jiao Tong University
September 13, 2012
Finnish Centre of Excellence in Inverse Problems Research

https://wiki.helsinki.fi/display/inverse/Home
This is a joint work with

Keijo Hämäläinen, University of Helsinki

Aki Kallonen, University of Helsinki

Ville Kolehmainen, University of Eastern Finland

Matti Lassas, University of Helsinki

Kati Niinimäki, University of Eastern Finland

Eero Saksman, University of Helsinki
Outline

Principle of X-ray tomography

Sparse-data tomography

Continuous and discrete Bayesian inversion

Promoting sparsity using Besov space priors

Applications to medical X-ray imaging
Outline

Principle of X-ray tomography

Sparse-data tomography

Continuous and discrete Bayesian inversion

Promoting sparsity using Besov space priors

Applications to medical X-ray imaging
Godfrey Hounsfield and Allan McLeod Cormack were the first to develop X-ray tomography.

Nobel prize was awarded to Hounsfield (top) and Cormack in 1979.
Reconstruction of a function from its line integrals was first invented by

Johann Radon (1887-1956).

This is the famous inversion formula from 1917 for the Radon transform $Rf$ of a function $f$:

$$f(x) = \frac{1}{4\pi^2} \int_{S^1} \int_{\mathcal{R}} \frac{d}{ds}(Rf)(\theta, s) \frac{x \cdot \theta - s}{d\theta} ds \, d\theta$$
Filtered back-projection (FBP) is mathematical technology used on a daily basis in hospitals around the world. The quality of 3D reconstruction using FBP is excellent.

However, a comprehensive data set is mandatory for FBP.
Mathematical interpretation of X-ray measurements

X-ray source

1000  1000  1000

Aluminium  Aluminium

1000  500  250

Detector
Mathematical interpretation of X-ray measurements

X-ray source

1000

1000

1000

Aluminium

Aluminium

Detector

1000

500

250

Logarithm

6.9

6.2

5.5
Mathematical interpretation of X-ray measurements

X-ray source

1000

1000

1000

Aluminium

Aluminium

Logarithm  6.9  6.2  5.5
Density  0.0  0.7  1.4

Detector
Tomography is based on measuring densities of matter using attenuation data.

![Diagram showing X-ray source and detector with a grid of numbers. The sum of the numbers along the line is 13.]
A projection image is produced by a set of (roughly) parallel X-rays

![Diagram showing a 3x3 matrix and calculations for each row resulting in 13, 8, and 3.](image-url)
Direct problem of tomography is to find the radiographs from given tissue.
Inverse problem of tomography is to find the tissue from radiographs.

9 unknowns, 11 linear equations
Inverse problem of tomography is to find the tissue from radiographs.
The limited angle problem is harder than the full angle problem.

9 unknowns, 11 linear equations

9 unknowns, 6 linear equations
In limited angle 3D imaging there are many tissues matching the radiographs.

\[
\begin{array}{ccc}
8\sqrt{2} & \quad & \\
9\sqrt{2} & \quad & \\
1\sqrt{2} & \quad & \\
4 & 4 & 5 & 13 \\
1 & 3 & 4 & \\
1 & 0 & 2 & 3 \\
\end{array}
\]

\[
\begin{array}{ccc}
& 5 & 6 & 2 \\
1 & 5 & 2 & \\
4 & 0 & -1 & \\
9 & 1 & 3 & \\
1 & 0 & 7 & \\
3 & 0 & 0 & \\
\end{array}
\]

*a priori* information is needed!
We write the reconstruction problem in matrix form and assume Gaussian noise

\[
x = [x_1, x_2, \ldots, x_9]^T
\]

\[
m = [m_1, m_2, \ldots, m_6]^T
\]

Construct system matrix \( A \) so that

\[
Ax = m
\]

Our measurement is \( m = Ax + \epsilon \)

with Gaussian noise \( \epsilon \) of standard deviation \( \sigma > 0 \).
Outline

Principle of X-ray tomography

Sparse-data tomography

Continuous and discrete Bayesian inversion

Promoting sparsity using Besov space priors

Applications to medical X-ray imaging
Due to the radiation dose, a CT scan is only appropriate for seriously ill patients

In filtered backprojection, the mathematical reconstruction formula assumes dense angular sampling of full-angle data. Then the inverse problem is only mildly ill-posed. We can say that **The chosen mathematics requires high radiation dose.**

Think the opposite: take as few X-ray images as possible. Then the inverse problem becomes very ill-posed. Use advanced mathematics to form a reconstruction that is good enough for the clinical task. **Now the low level of radiation dose requires new mathematics and more computational power.**
A series of projects started in Finland in 2001, aiming for a new type of low-dose 3D imaging. The goal was a mathematical algorithm with

**Input:** small number of digital X-ray images from any X-ray device.
**Output:** Three-dimensional reconstruction with high enough quality for the clinical task at hand.

Products of Instrumentarium Imaging in 2001:
Experimental setup for chairside 3D imaging models the clinical situation
Details of this limited angle experiment

Opening angle 60 degrees

Seven digital intraoral radiographs (664 x 872 pixels each)

There are 42,496,000 unknowns and 4,053,056 linear equations. Computation is divided into 400 approximately 2D problems.
Tomo-synthesis

Bayes-MAP

S, Kolehmainen, Järvenpää, Kaipio, Koistinen, Lassas, Pirttilä and Somersalo 2003
We reconstructed the 3D volume using total variation regularization and non-negativity.

We wanted to calculate the minimizer of the TV functional

\[ \|Ax - m\|_2^2 + \alpha \|Lx\|_1, \]

where \( L \) is a first-order difference matrix.

Due to the large scale of the problem we used the approximate absolute value function \( |t|_\beta := \sqrt{t^2 + \beta} \) in the total variation norm. Also, we implemented the non-negativity constraint by a smooth, iteratively steeper barrier.

The above leads to a differentiable object functional, and we used the gradient-based Barzilai-Borwein method to find the minimizer.
Application: dental implant planning, where a missing tooth is replaced by an implant
Three-dimensional information is crucial for dental implant planning

The hole must be drilled deep enough for sturdy attachment but not so deep that the mandibular nerve is damaged.

Two-dimensional X-ray projection images are not suitable for assessing the proper depth because of geometric distortion.

Three-dimensional reconstruction of the tissue is needed, but a traditional CT scan is not practical due to high cost, too low resolution and too high radiation dose.

Low-dose 3D X-ray imaging is an ideal solution: it is based on a small number of X-ray projection images recorded with a cost-effective device.
Panoramic dental imaging shows all the teeth simultaneously.

Panoramic imaging was invented by Yrjö Veli Paatero in the 1950’s.
Nowadays, a digital panoramic imaging device is standard equipment at dental clinics.

Panoramic images are not good enough for dental implant planning because of geometric distortion.
We reprogram the panoramic X-ray device so that it collects projection data by scanning

11 projection images of the mandibular area

40 degrees angle of view

1000×1000 image size, formed by a scanning movement
Here are example images of a patient

Kolehmainen, Vanne, S, Järvenpää, Kaipio, Lassas & Kalke 2006,
Kolehmainen, Lassas & S 2008, Cederlund, Kalke & Welander 2009,
Hyvönen, Kalke, Lassas, Setälä & S 2010,
United States patent 7269241
This low-dose 3D imaging technique has been commercialized by Palodex Group.

The VT device has been in the market from year 2007.

Remarkably, an existing 2D panoramic imaging device becomes a 3D imaging product just by a software update.

The core of that update is an inversion algorithm.
The radiation dose of the VT device is the lowest among 3D dental imaging modalities

<table>
<thead>
<tr>
<th>Modality</th>
<th>$\mu$Sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head CT</td>
<td>2100</td>
</tr>
<tr>
<td>CB Mercuray</td>
<td>558</td>
</tr>
<tr>
<td>i-Cat</td>
<td>193</td>
</tr>
<tr>
<td>NewTom 3G</td>
<td>59</td>
</tr>
<tr>
<td>VT device</td>
<td>13</td>
</tr>
</tbody>
</table>

Ludlow, Davies-Ludlow, Brooks & Howerton 2006
We needed a new computational approach for removing the approximate absolute value function

Using $|t|_\beta := \sqrt{t^2 + \beta}$ in the TV norm leads to smoothing. We want crisp, non-negative, large-scale TV-regularizations.

There are several candidates for the computational algorithm:

- **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.
We removed the approximate absolute value function using quadratic programming

The minimizer of the functional

\[
\arg\min_{x \in \mathbb{R}^n_+} \left\{ \|Ax - m\|^2_2 + \alpha \|Lx\|_1 \right\}
\]

can be transformed into the standard form

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

where \( Q \) is symmetric and \( E \) implements equality constraints.

A large-scale QP method was developed in Kolehmainen, Lassas, Niinimäki & S 2012 and Hämäläinen, Kallonen, Kolehmainen, Lassas, Niinimäki & S (submitted).
Reduction to quadratic programming

Denote horizontal and vertical differences by \( L_H x = u_H^+ - u_H^- \) and \( L_V x = u_V^+ - u_V^- \), where \( u_H^\pm, u_V^\pm \geq 0 \). TV minimization is now

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

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

\[
z = \begin{bmatrix} x \\ u_H^+ \\ u_H^- \\ u_V^+ \\ u_V^- \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{1}{\sigma^2} A^T A & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad c = \begin{bmatrix} -2 A^T m \\ \alpha \mathbf{1} \\ \alpha \mathbf{1} \\ \alpha \mathbf{1} \\ \alpha \mathbf{1} \end{bmatrix},
\]

Now the TV minimization becomes

\[
\arg\min_{z \in \mathbb{R}^{5n}} \left\{ \frac{1}{2} z^T Q z + c^T z \right\} \quad \text{with} \quad Ez = b \quad \text{and} \quad z \geq 0.
\]
We collected X-ray projection data of a walnut from 1200 directions

The data was collected by Keijo Hämäläinen and Aki Kallonen at University of Helsinki.
This is the reconstruction using all 1200 projections and filtered back-projection
When only few projection angles are available, TV regularization performs better than FBP.

These images were computed by Kati Niinimäki.
Outline

Principle of X-ray tomography

Sparse-data tomography

Continuous and discrete Bayesian inversion

Promoting sparsity using Besov space priors

Applications to medical X-ray imaging
Our starting point is the *continuum model* for an indirect measurement.

Consider a quantity $U$ that can be observed indirectly by

$$M = AU + \mathcal{E},$$

where $A$ is a smoothing linear operator, $\mathcal{E}$ is white noise, and $U$ and $M$ are functions.

In the Bayesian framework, $U = U(x, \omega)$, $M = M(z, \omega)$ and $\mathcal{E} = \mathcal{E}(z, \omega)$ are taken to be random functions.
X-ray tomography: Continuum model

\[ M = AU + \mathcal{E} \]
A measurement device gives only a finite number of data, leading to practical measurement model

The measurement is a realization of the random variable

\[ M_k = P_k M = A_k U + \mathcal{E}_k, \]

where \( A_k = P_k A \) and \( \mathcal{E}_k = P_k \mathcal{E} \). Here \( P_k \) is a linear orthogonal projection with \( k \)-dimensional range.

The given data is a realization \( \hat{m}_k = M_k(z, \omega_0) \) of the measurement \( M_k(z, \omega) \), where \( \omega_0 \in \Omega \) is a specific element in the probability space.
X-ray tomography: Practical measurement model

\[ M_k = A_k U + \mathcal{E}_k \]
The inverse problem

This study concentrates on the inverse problem

given a realization \( \hat{m}_k \), estimate \( U \),

using estimates and confidence intervals related to a Bayesian posterior probability distribution.
Numerical solution of the inverse problem is based on the discrete \textit{computational model}

We need to discretize the unknown function $U$. Assume that $U$ is \textit{a priori} known to take values in a function space $Y$.

Choose a linear projection $T_n: Y \rightarrow Y$ with $n$-dimensional range $Y_n$, and define a random variable $U_n := T_n U$ taking values in $Y_n$. This leads to the \textit{computational model}

$$M_{kn} = A_k U_n + \mathcal{E}_k.$$ 

Note that realizations of $M_{kn}$ can be simulated by computer but cannot be measured in reality.
X-ray tomography: Computational model

\[ M_{kn} = A_k U_n + \mathcal{E}_k \]
The computational model $M_{kn} = A_k U_n + \mathcal{E}_k$ involves two independent discretizations:

$P_k$ is related to the measurement device, and $T_n$ is related to the finite representation of $U$. 
The numbers $k$ and $n$ are independent
The numbers $k$ and $n$ are independent

\[ k = 8 \]
The numbers $k$ and $n$ are independent

$k = 8$

$n = 48$
The numbers $k$ and $n$ are independent

$k = 8$

$n = 156$
The numbers $k$ and $n$ are independent

$k = 8$

$n = 440$
The numbers $k$ and $n$ are independent

\[ k = 16 \]
\[ n = 440 \]
The numbers $k$ and $n$ are independent

\[ k = 24 \]
\[ n = 440 \]
Bayesian estimates are drawn from the posterior distribution related to the computational model

The posterior distribution is defined by

$$
\pi_{kn}(u_n \mid m_{kn}) = C \pi(u_n) \pi(m_{kn} \mid u_n).
$$

Data is given as a realization of the practical measurement model

$$
M_k = A_k U + \mathcal{E}_k: \hat{m}_k = M_k(\omega_0).
$$

The conditional mean estimate \( u_{kn}^{CM} \) is defined by

$$
u_{kn}^{CM} := \int_{\mathbb{R}^n} u_n \pi_{kn}(u_n \mid \hat{m}_k) \, d\mu(u_n),
$$

and the maximum a posteriori estimate \( u_{kn}^{MAP} \) is defined by

$$
\pi_{kn}(u_{kn}^{MAP} \mid \hat{m}_k) = \max_{u_n} \{ \pi_{kn}(u_n \mid \hat{m}_k) \}. 
$$
Bayesian MAP estimation with total variation prior is equivalent to TV regularization

The likelihood is \( C \exp\left(-\frac{1}{2\sigma^2}\|A_k u_n - \hat{m}_k\|_2^2\right) \) and the TV prior is \( C \exp(-\delta\|Lu_n\|_1) \) with \( \delta > 0 \). The MAP estimate is defined by

\[
\pi_{kn}(u_{kn}^{\text{MAP}} \mid \hat{m}_k) = \max_{u_n}\{ C \exp\left(-\frac{1}{2\sigma^2}\|A_k u_n - \hat{m}_k\|_2^2 - \delta\|Lu_n\|_1\right) \}.
\]

Taking logarithm gives

\[
u_{kn}^{\text{MAP}} = \arg\min_{u_n}\left\{ \frac{1}{2\sigma^2}\|A_k u_n - \hat{m}_k\|_2^2 + \delta\|Lu_n\|_1 \right\}.
\]

Now we see that \( u_{kn}^{\text{MAP}} \) is the minimizer of the TV functional

\[
\|A_k u_n - \hat{m}_k\|_2^2 + \alpha\|Lu_n\|_1
\]

with \( \alpha = 2\sigma^2\delta \).
Bayesian inversion using total variation prior does not have a well-defined continuous limit

Theorem (Lassas and S 2004)

Total variation prior is not discretization-invariant.

Sketch of proof: Apply a variant of the central limit theorem to the independent, identically distributed random consecutive differences.

New numerical experiments are reported in Kolehmainen, Lassas, Niinimäki and S (2012).
Discretization dilemma with total variation prior

Parameter fixed

Parameter $\sim \sqrt{n}$

CM

MAP
Discretization dilemma with total variation prior

Parameter fixed

Parameter $\sim \sqrt{n}$

CM

Diverges

Becomes smooth

MAP

Becomes zero
Outline

Principle of X-ray tomography

Sparse-data tomography

Continuous and discrete Bayesian inversion

Promoting sparsity using Besov space priors

Applications to medical X-ray imaging
Bayesian inversion using Besov space priors

Theorem (Lassas, Saksman and S 2009)
Besov space priors are discretization-invariant.

Sketch of proof: Construction of well-defined Bayesian inversion theory in infinite-dimensional Besov spaces that allow wavelet bases. Discretizations are achieved by truncating the wavelet expansion.

Numerical experiments are reported in Kolehmainen, Lassas, Niinimäki and S (2012).

Deterministic Besov space regularization was first introduced in Daubechies, Defrise and De Mol 2004.
The wavelet transform divides an image into three types of details at different scales.
This is how the image decomposition is defined mathematically

We can represent discrete images by wavelet expansion

\[
f(x, y) = \sum_{k_1=0}^{2J_0-1} \sum_{k_2=0}^{2J_0-1} c_{J_0 \vec{k}} \phi_{J_0, \vec{k}}(x, y) + \sum_{j=J_0}^{J-1} \sum_{\ell=1}^{3} \sum_{k_1=0}^{2j-1} \sum_{k_2=0}^{2j-1} w_{j k \ell} \psi_{j, \vec{k}}^\ell(x, y),
\]

where \( \vec{k} = (k_1, k_2) \) and the coefficients \( c_{J_0 \vec{k}} \) and \( w_{j k \ell} \) are defined by

\[
c_{J_0 \vec{k}} = \langle f, \phi_{J_0 \vec{k}} \rangle = \int_{\mathbb{T}^2} f(x, y) \phi_{j k}(x, y) dxdy,
\]

\[
w_{j k \ell} = \langle f, \psi_{j k}^\ell \rangle = \int_{\mathbb{T}^2} f(x, y) \psi_{j k}^\ell(x, y) dxdy.
\]
Besov space norm can be defined in terms of the wavelet coefficients

A function $f$ belongs to $B_{pq}^s(\mathbb{T}^2)$ if and only if the following expression is finite:

$$\left( \sum_{k_1=0}^{2^{J_0}-1} \sum_{k_2=0}^{2^{J_0}-1} |c_{J_0 \vec{k}}|^p \right)^{\frac{1}{p}} + \left( \sum_{j=J_0}^{\infty} 2^{jq(s+1-\frac{2}{p})} \left( \sum_{\ell=1}^{3} \sum_{k_1=0}^{2j-1} \sum_{k_2=0}^{2j-1} |w_{j\vec{k}\ell}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}$$

We focus on the case $p = 1$ and $q = 1$ and $s = 1$:

$$\|f\|_{B_{11}^1(\mathbb{T}^2)} = \sum_{k_1=0}^{2^{J_0}-1} \sum_{k_2=0}^{2^{J_0}-1} |c_{J_0 \vec{k}}| + \sum_{j=J_0}^{\infty} \sum_{\ell=1}^{3} \sum_{k_1=0}^{2j-1} \sum_{k_2=0}^{2j-1} |w_{j\vec{k}\ell}|.$$
We promote sparsity by minimizing a sum of $\ell^2$ and $\ell^1$ norms.

In the case of Besov space penalty we minimize

$$ f^\text{MAP} = \arg\min_{f \in \mathbb{R}^n} \left\{ \frac{1}{2\sigma^2} \|Kf - m\|_{\ell^2}^2 + \alpha \sum_j W_j |\langle f, \psi_j \rangle| \right\}, $$

where $\psi_j$ are the wavelet basis functions.
We took photographs of walnuts cut in half

These photos are used for estimating the expected number of nonzero wavelet coefficients in a two-dimensional tomographic reconstruction. Special thanks go to Esa Niemi for his careful job in sawing the walnuts.
The S-curve method determines a regularization parameter value giving the right sparsity level.
Let’s see how the S-curve method works for tomographic problems with different numbers of projection data.
Reconstruction from 90 projections

\[ \hat{S} = 5936 \]

\[ \alpha = 0.024 \]
Reconstruction from 45 projections

\[ \hat{S} = 5936 \]

\[ \alpha = 0.011 \]
Reconstruction from 30 projections

\[ \hat{S} = 5936 \]

\[ \alpha = 0.019 \]
Reconstruction from 15 projections

\[ \hat{S} = 5936 \]

\[ \alpha = 0.015 \]
All reconstructions with comparison to filtered back-projection

90 angles

45 angles

30 angles

15 angles

$\alpha = 0.024$

$\alpha = 0.011$

$\alpha = 0.019$

$\alpha = 0.015$
Let’s check that the MAP estimates based on Besov $111$-prior are discretization-invariant.
Outline

Principle of X-ray tomography

Sparse-data tomography

Continuous and discrete Bayesian inversion

Promoting sparsity using Besov space priors

Applications to medical X-ray imaging
Limited angle tomography results for X-ray mammography

[Tomosynthesis]

[Rantala et al. 2006]
Thanks to GE Healthcare

MAP estimate, Besov prior, p=1.5=q and s=0.5
Local tomography results for dental X-ray imaging; data measured from specimen
Besov space method compared to \( \Lambda \)-tomography in the region of interest

Lambda-tomography   MAP using Besov prior with \( p=q=1.5 \) and \( s=0.5 \)

Thanks to Palodex Group
Thank you for your attention!

Preprints available at www.siltanen-research.net.