Image reconstruction in limited data X-ray tomography

Ville Kolehmainen[1]

[1]Department of Applied Physics,
University of Eastern Finland (UEF), Kuopio, Finland

SSIP 2016, Szeged, July 2016

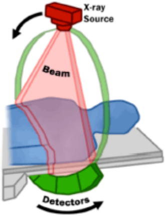
Joint work with :

- Kati Niinimäki
- Samuli Siltanen
- Matti Lassas
- Jari Kaipio
- Maaria Rantala
- Erkki Somersalo
- Antti Vanne
- Martti Kalke

Outline:

Introduction

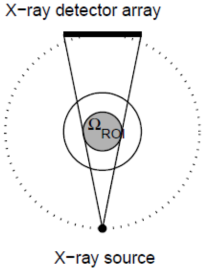
Computerized Tomography (CT)



- Filtered Back Projection:

$$f(x) = \frac{1}{4\pi^2} \int_{S^1} \int_{\mathbb{R}} \frac{\frac{d}{ds}(Rf)(\theta, s)}{x \cdot \theta - s} ds \, d\theta$$

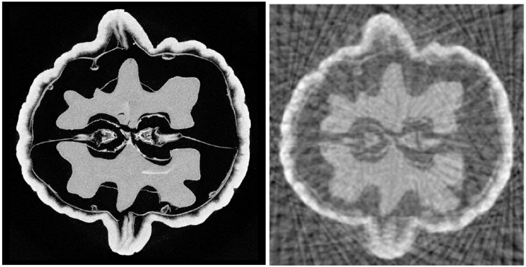
CT vs. Sparse (Low Dose) Data



- Conventional CT; high radiation dose due to the very large number of projections → usage limited by patient safety.
- Dose reduction a central design factor in next generation scanners → Allows completely new applications.
- Dose reduction can be obtained by:
  1. Short exposure time → poor SNR
  2. Using limited field of view (local tomography)
  3. Decreasing the number of projection images

FBP reconstruction using full and sparse data

1200 projections 30 projections



The reconstruction of f from sparse projection data becomes *ill-posed* problem $\Rightarrow$ Advanced reconstruction methods needed.

This lecture: We describe a Bayesian approach for low dose tomography problem and consider two test cases

Computational model of X-ray tomography

Tomography is based on measuring densities of matter using X-ray attenuation data

Detector → X-ray projection image

664 x 872 matrix.
Black = high photon count
White = low photon count

Measured photon count

$$I_j = I_0 \exp\left(- \int_{L_j} f(s)ds \right)$$

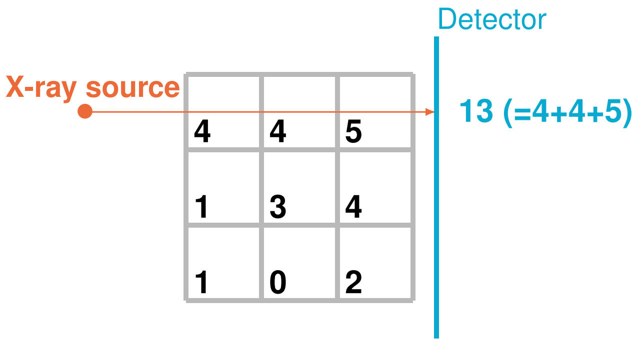
Transform to linear attenuation data by

$$-\log\left(\frac{I_j}{I_0} \right) = \int_{L_j} f(s)ds$$

Discretization of line integrals

$$m_j = \int_{L_j} f(s)ds \approx \sum_{i=1}^{n} a_{ji} f_i$$

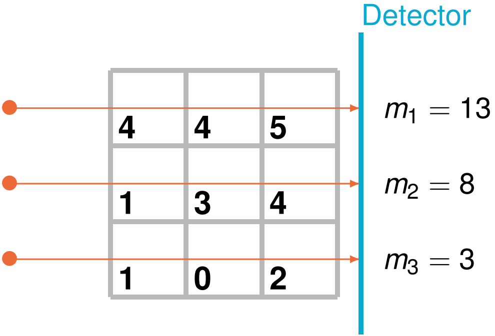
where a_{ji} is the length of ray j in pixel i.



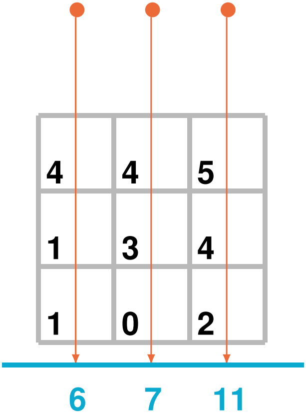
A projection image is produced by parallel X-rays and several detector pixels (here three pixels)

$$m_j = \int_{L_j} f(s)ds \approx \sum_{i=1}^{n} a_{ji} f_i$$

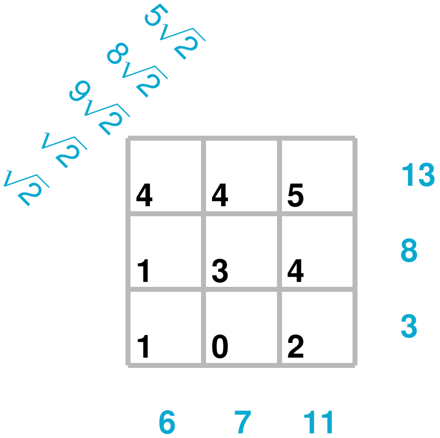
where a_{ji} is the length of ray j in pixel i.



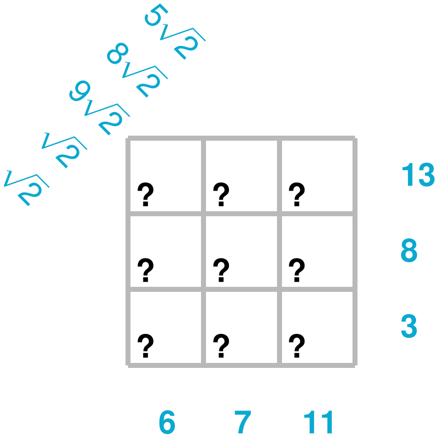
For tomographic imaging it is essential to record projection images from different directions

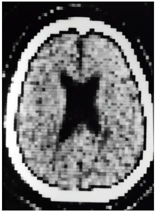


The direct problem of tomography is to find the projection images from known tissue

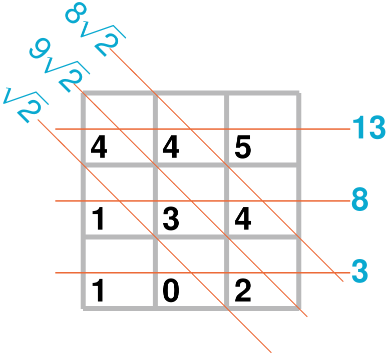


The inverse problem of tomography is to reconstruct the interior from X-ray data





When collecting sparse data, different objects may produce the same data

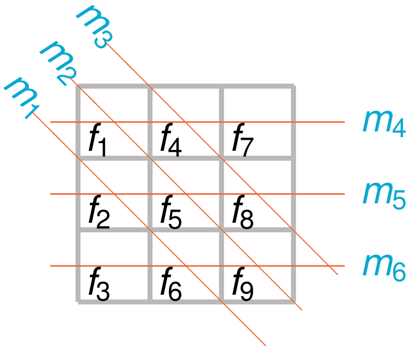


5	6	2
1	5	2
4	0	-1

9	1	3
1	0	7
3	0	0

Reconstruction requires additional
a priori information

We write the computation model in matrix form and assume Gaussian noise



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

$$m = Af$$

Our measurement model is $m = Af + \varepsilon$ with independently distributed Gaussian noise (white noise) with standard deviation $\sigma > 0$.

Bayesian inversion

Bayesian inversion complements measurement data with *a priori* knowledge

Consider the model $m = Af + \varepsilon$, where $m \in \mathbb{R}^k$ and $f \in \mathbb{R}^n$. The inverse problem is to find f when measurement m is given.

We use probability theory to model our lack of information in the inverse problem. The conditional probability

$$\pi(f \,|\, m) = \frac{\pi(f)\pi(m \,|\, f)}{\pi(m)}$$

is called the *posterior distribution*.

In case of white noise $\varepsilon_j \sim \mathcal{N}(0, \sigma^2)$, the *likelihood distribution* is

$$\pi(m \,|\, f) = C \exp\left(-\frac{1}{2\sigma^2}\|Af - m\|_2^2\right).$$

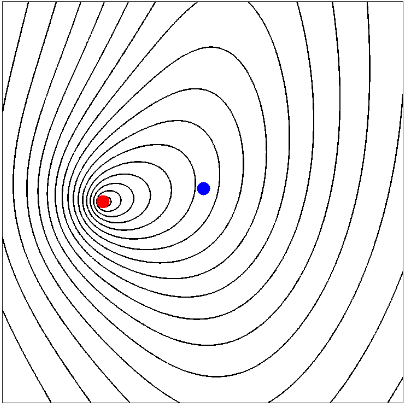
The result of Bayesian inversion is the posterior distribution, but typically one looks at estimates

Maximum a posteriori (MAP) estimate:

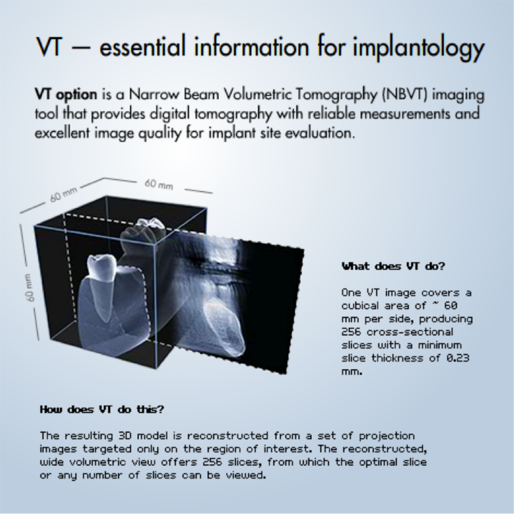
$$\arg\min_{f \in \mathbb{R}^n} \pi(f \mid m)$$

Conditional mean (CM) estimate:

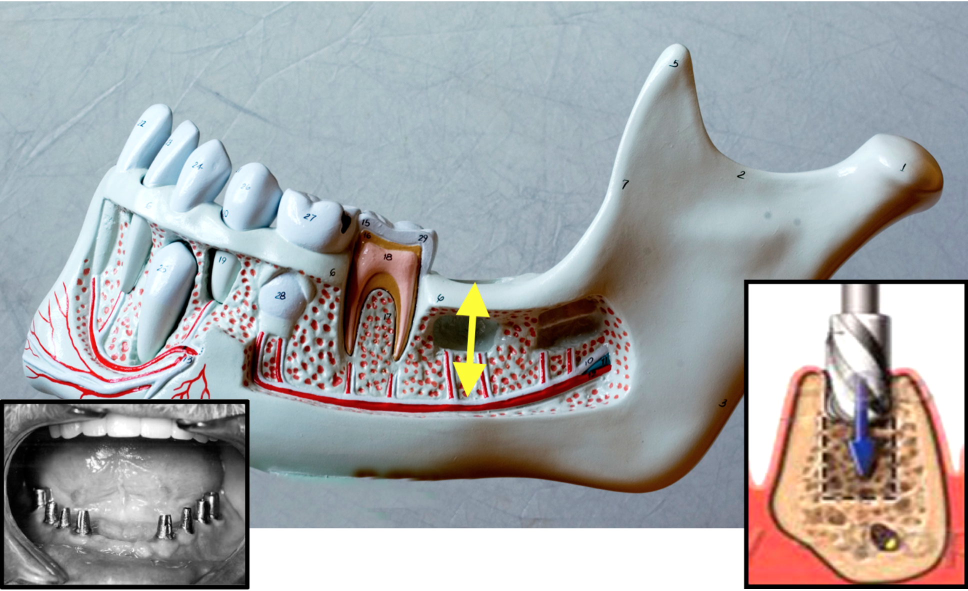
$$\int_{\mathbb{R}^n} f \, \pi(f \mid m) \, df$$



Case 1: Limited angle tomography in dental imaging



Application: dental implant planning, where a missing tooth is replaced with an implant



Nowadays, a digital panoramic imaging device is standard equipment at dental clinics

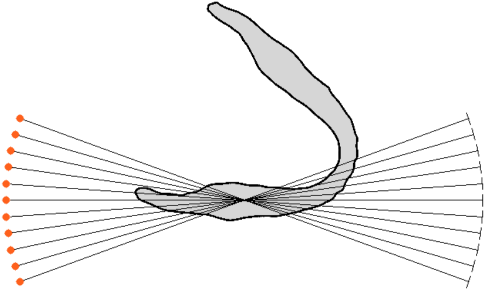
Panoramic images are not good enough for dental implant planning because of geometric distortion.

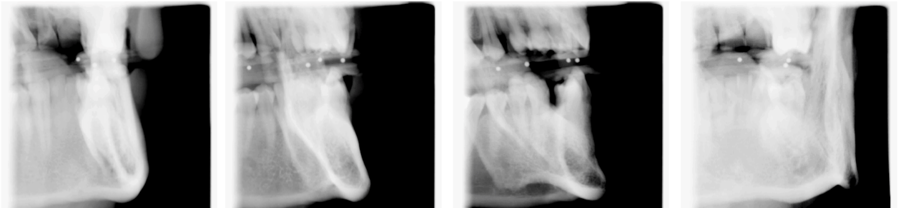
Our solution: Reprogram the panoramic X-ray device so that it collects projection data for 3D reconstruction

11 projection images of the mandibular area

40 degrees angle of view

1000×1000 image size, formed by a scanning movement





Can we obtain sufficient 3D reconstruction from so limited data?

We seek solution by the Bayesian approach. Prior information available in dental imaging:

- Different tissue types (enamel, bone, gum, pulp chamber) and possible artificial materials (fillings, previous implants) are approximately homogeneous
- Attenuation (density) of tissues is non-negative (X-radiation does not intensify inside tissue)
- There are "crisp" boundaries between the different tissues

We model the prior knowledge by the following quantitative and qualitative models:

Non-negativity prior

$$\pi_+(f) = \begin{cases} 0 & \textit{if } f_j < 0 \textit{ for any } j \\ 1 & \textit{otherwise.} \end{cases}$$

Total variation (TV) - prior

$$\pi(f) \propto \exp\left(-\alpha\left\{\|L_H f\|_1 + \|L_V f\|_1\right\}\right)$$

Figure 5. Images having total variations (from left to right) 18, 28 and 40.

Posterior model:

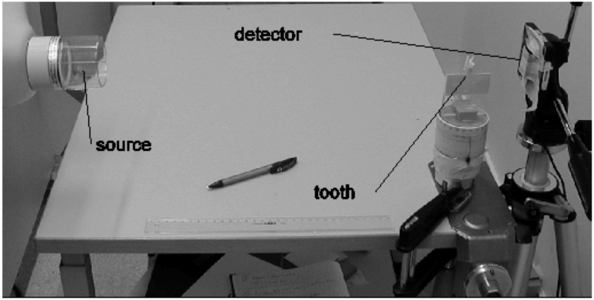
$$\pi(f|m) \propto \exp\left\{-\frac{1}{2\sigma^2}\|m - Af\|^2 - \alpha\left\{\|L_{\text{H}}f\|_1 + \|L_{\text{V}}f\|_1\right\}\right\}$$

Computation of the MAP estimate

$$f_{\mathcal{MAP}} = \arg\min_{f\geq 0}\left\{\frac{1}{2\sigma^2}\|m - Af\|^2 + \alpha\left\{\|L_{\text{H}}f\|_1 + \|L_{\text{V}}f\|_1\right\}\right\}$$

Nowdays there is a large variety of optimization algorithms that can be used for the solution of the MAP estimate (e.g. ADMM, Chambolle-Pock, ...)

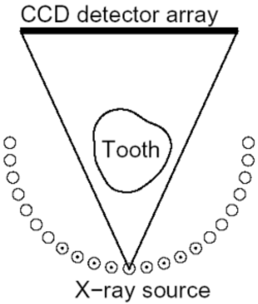
Simple experiment using a 2D model

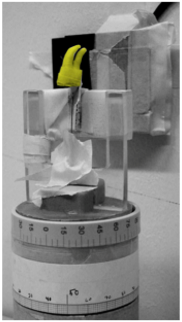


Experimental setup

- Dental x-ray source
- intraoral CCD detector
- A rotating platform
- Target: a tooth specimen

Projection geometry

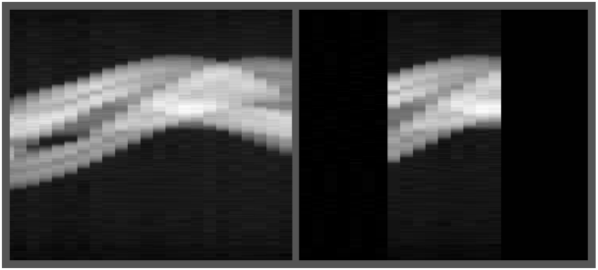




Data from a tooth specimen

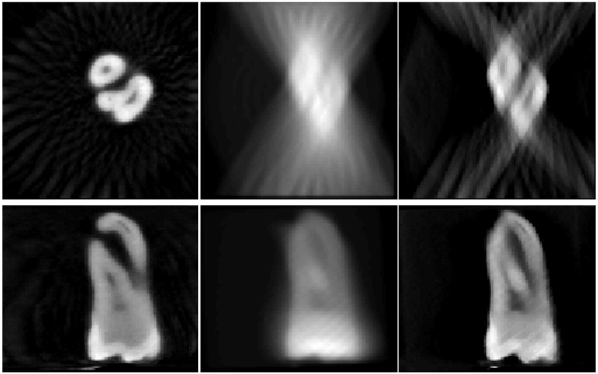
- 23 projections from full angle (187 deg)
- 9 projections from limited angle (68 deg)

Sinograms



Full angle　　　　　　　　　　Limited angle

Comparison of Bayesian inversion (MAP-TV) and tomosynthesis (backprojection)



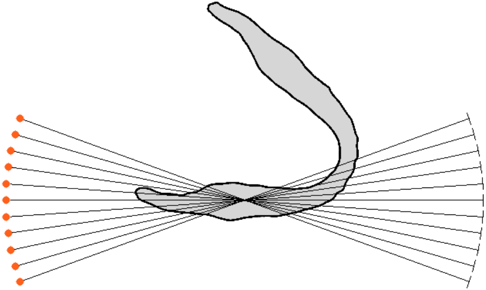
Left: MAP-TV (full angle), Middle: Backprojection (limited angle), Right: MAP-TV (limited angle)

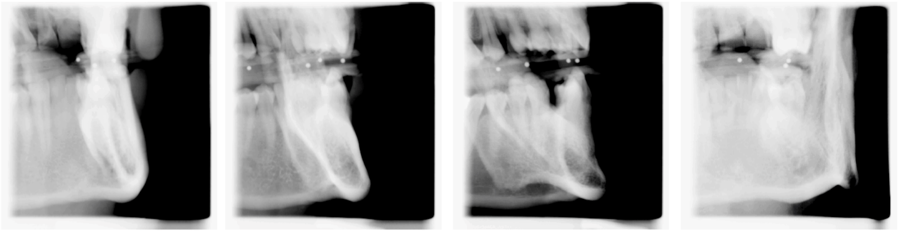
First experiment from a dry skull using panoramic device

11 projection images of the mandibular area

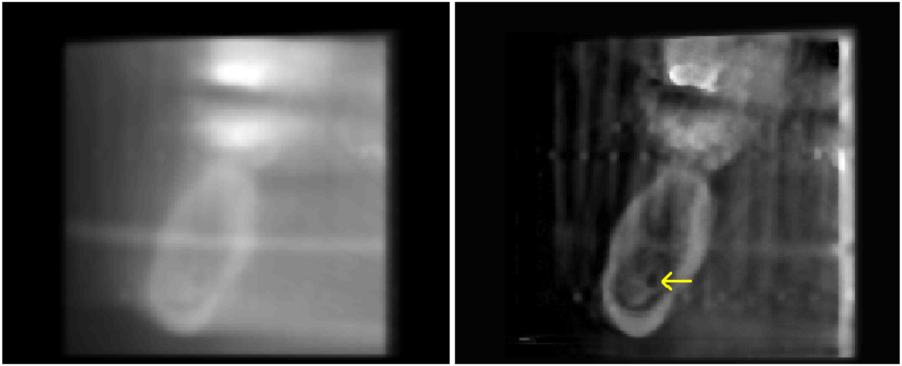
40 degrees angle of view

1000×1000 image size, formed by a scanning movement





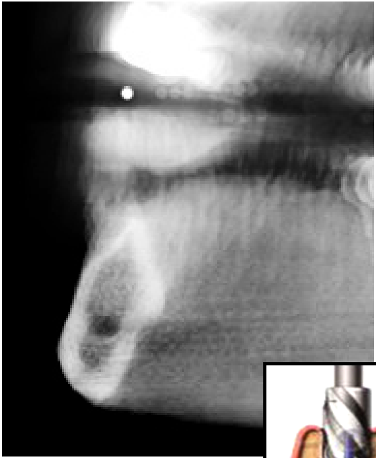
First result from the dry skull (tomosynthesis left, MAP-TV right)

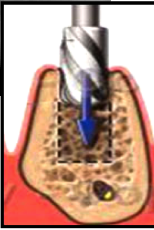


Computational considerations:

- Number of data $\sim$ 1 million, number of unknowns $\sim$ 7 millions.
- 3D implementation
- Clinically acceptable computation times could be obtained by optimized implementation. We investigated parallel CPU computing and GPU based computation.

Finally, here are example images of a patient:





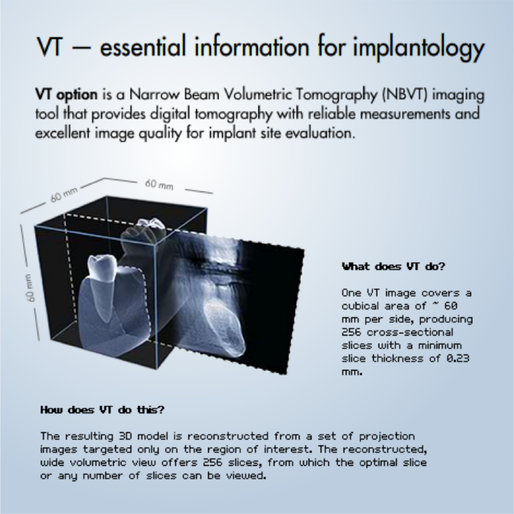
Kolehmainen, Vanne, Siltanen, Järvenpää, Kaipio, Lassas & Kalke **2006**,
Kolehmainen, Lassas & Siltanen **2008**, Cederlund, Kalke & Welander **2009**,

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 that can provide sufficient reconstruction from the *sparse data*.



The radiation dose of the VT device is the lowest among 3D dental imaging modalities

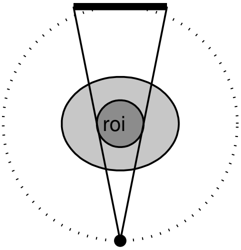
Modality	μ**Sv**
Head CT	2100
CB Mercuray	558
i-Cat	193
NewTom 3G	59
VT device	**13**

Ludlow, Davies-Ludlow, Brooks & Howerton **2006**

Case 2: Local tomography

Prototype of a dental CBCT device

Local tomography problem:



Principle of local tomography.

- Let image domain Ω s.t. $D \subset \Omega$, where D denotes the body.
- Decompose to disjoint subdomains

$$\Omega = \Omega_{roi} \bigcup \Omega_{out}$$

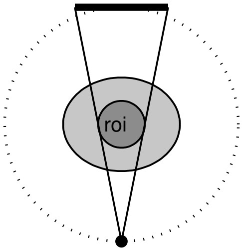
where Ω_{roi} represents the *region of interest* that is present in all the projection images.

- The objective in local tomography is to find accurate reconstruction of

$$f|_{\Omega_{roi}}$$

using (truncated) x-ray projection data.

Image reconstruction in Local tomography



Principle of local tomography.

- Fine scale details cannot be reconstructed in Ω_{out}.
- The contribution of $f|_{\Omega_{out}}$ has to be taken into account in the projection model $\rightarrow$ large number of unknowns.
- We propose a multiresolution approach; Attenuation function f is represented in a reduced wavelet basis with finer scale repsentation available only inside Ω_{roi} and coarse representation in Ω_{out}

$\rightarrow$ *Significant model reduction while accuracy is retained in the ROI.*

Wavelet expansion of f

- The wavelet functions in 2D:
  - Scaling function $\phi_{jk}(x) = 2^j \phi(2^j x - z_{jk})$
  - Wavelet function $\psi_{jk\ell}(x) = 2^j \psi^\ell(2^j x - z_{jk})$,

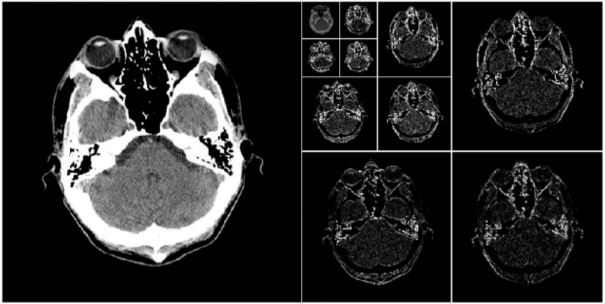
  where $j \in \mathbb{Z}$ refers to the scale, $k \in \mathbb{Z}$ to the location z_{jk} in space and ℓ the wavelet type.
- The wavelet expansion of a function $f : [0,1]^2 \to \mathbb{R}$

$$f = \sum_{k=1}^{K_{J_0}} \underbrace{\langle f, \phi_{J_0 k} \rangle}_{c_{J_0 k}} \phi_{J_0 k} + \sum_{j=J_0}^{J-1} \sum_{k=1}^{K_j} \sum_{\ell=1}^{3} \underbrace{\langle f, \psi_{jk\ell} \rangle}_{w_{jk\ell}} \psi_{jk\ell}.$$

- Formally, we write matrix form:

$$f = Bw = B \begin{bmatrix} (c_{J_0 k}) \\ (w_{jk\ell}) \end{bmatrix}$$

Example of wavelet expansion:



- Original image and the wavelet coefficients using 3 scaling levels.
- Wavelet types ($\ell = 1, 2, 3$):

$$\phi(\boldsymbol{x}) = \phi(x_1)\phi(x_2) \text{ (scaling)} \qquad \psi^2(\boldsymbol{x}) = \psi(x_1)\phi(x_2) \text{ (vertical)}$$
$$\psi^1(\boldsymbol{x}) = \phi(x_1)\psi(x_2) \text{ (horizontal)} \qquad \psi^3(\boldsymbol{x}) = \psi(x_1)\psi(x_2) \text{ (diagonal)}$$

Besov space norm of f:

- The Besov norm of a function f can be expressed with wavelet coefficients

$$\|f\|_{B_p^{sq}} = \|c_{J_0 k}\|_{\ell^p} + \left[\sum_{j=0}^{J-1} \left(2^{jp\left(s+1-\frac{2}{p}\right)} \|w_{jkl}\|_{\ell^p}\right)^q\right]^{\frac{1}{q}}.$$

- Regularity controlled by p,s and q.

Bayesian inversion with wavelets and Besov prior

- Measurement model

$$y = Af + \epsilon = ABw + \epsilon$$

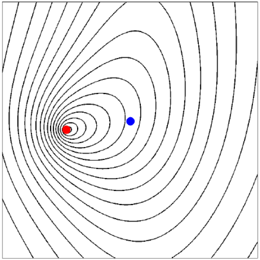
- Besov space prior

$$\pi(w) \propto \exp\left(-\alpha \|Bw\|^p_{B^{sq}_p}\right)$$

$\rightarrow$ Posterior probability distribution model

$$\pi(w|y) \propto \exp\left(-\frac{1}{2}\|ABw - y\|^2_{\Gamma_\epsilon^{-1}} - \alpha \|Bw\|^p_{B^{sq}_p}\right),$$

MAP estimate



MAP estimate (red)

- The *maximum a posterior* (MAP) estimate

$$w_{\mathcal{MAP}} = \arg\min_{w} \left\{ \frac{1}{2} \|ABw - y\|^2_{\Gamma_\epsilon^{-1}} + \alpha \|Bw\|^p_{B^{sq}_p} \right\}$$

- MAP estimate computed using Polak-Ribière conjugate gradient optimization method.

Multiresolution model for local tomography

- Basic idea:
  1) Use all the wavelet coefficients up to the finest scale J in Ω_{roi} and
  2) only a partial number of scaling levels ($J_{out} < J$) in $\Omega_{out} := \Omega \setminus \Omega_{roi}$
- Let $w \in \mathbb{R}^{n_f}$ denote the full wavelet expansion when all the J scales are used everywhere in Ω.
- Denoting by $\mathcal{S} \subset \{1, 2, \ldots, n_f\}$ the set of indices that contain
  i) all the scales up to J in $\Omega_{\mathcal{ROI}}$
  ii) the scales up to J_{out} in Ω_{out}

  we can write reduced wavelet expansion

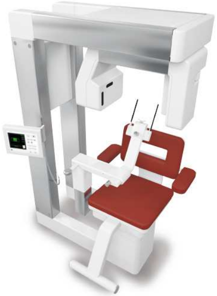
$$f = \tilde{B}\tilde{w}, \quad \tilde{w} = Pw \in \mathbb{R}^n, \quad \tilde{B} = BP^{\mathcal{T}}, \quad n \leq n_f$$

  for the multiresolution representation of f.

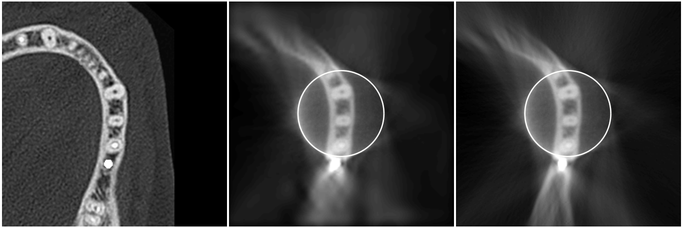
MAP estimate with the multiresolution model

$$\tilde{w}_{\mathcal{MAP}} = \arg\min_{\tilde{w}} \left\{ \frac{1}{2} \|A\tilde{B}\tilde{w} - y\|^2_{\Gamma_\epsilon^{-1}} + \alpha \|\tilde{B}\tilde{w}\|^p_{B^{sq}_p} \right\}$$

Results using simulated local tomography data

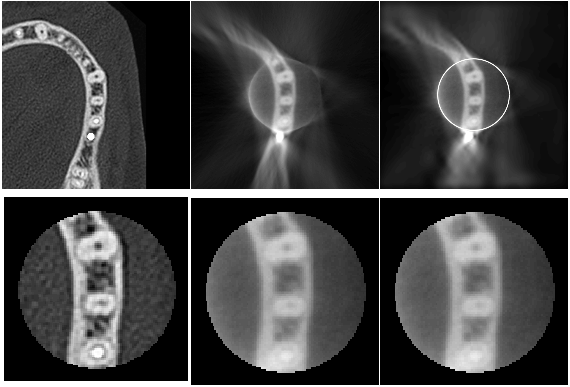


- Simulated (2D) local tomography data of a jawbone phantom (187 projections from total view angle of 187°).
- Besov norm parameters (p, q, s): $p = 1.5$, $q = 1.5$ and $s = 0.5$.

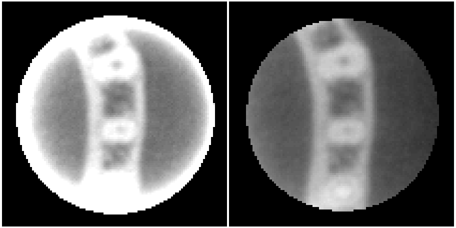


- Number of scaling levels in the ROI (marked with white circle): $J_{roi} = 5$.
- Reconstructions ($\delta_{\Omega_{\mathcal{R}OI}}$ is the relative L_2-reconstruction error in $\Omega_{\mathcal{R}OI}$ w.r.t the original phantom).
  - Middle: $J_{out} = 1$, number of unknowns $n = 14070$, $\delta_{\Omega_{\mathcal{R}OI}} = 19.3\%$.
  - Right: $J_{out} = 5$, number of unknowns $n = 76990$, $\delta_{\Omega_{\mathcal{R}OI}} = 19.1\%$.

Comparision with the TV prior:



- Middle: MAP with the total variation (TV) prior. Number of unknowns $n_{pix} = 65536$, $\delta_{\Omega_{ROI}} = 19.1\%$.
- Right: Multiresolution model with $J_{out} = 1$. Number of unknowns $n = 14070$, $\delta_{\Omega_{ROI}} = 19.3\%$.



- Left: MAP with the TV-prior using truncated measurement model (i.e., $\Omega = \Omega_{\mathcal{ROI}}$)

$$y \approx A_{\mathcal{ROI}} f_{\mathcal{ROI}} + \epsilon.$$

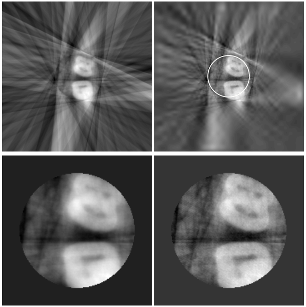
Number of unknowns $n_{pix} = 7484$, $\delta_{\Omega_{\mathcal{ROI}}} = 150.8\%$.

- Right: Multiresolution model with $J_{out} = 1$. Number of unknowns $n = 14070$, $\delta_{\Omega_{\mathcal{ROI}}} = 19.3\%$.

Experimental setup

- Data: 23 projections of a jawbone specimen from total view-angle of 187°.

Results with experimental data



- Data: 23 projections of a jawbone specimen from total view-angle of 187°.
- Left: MAP with the TV prior. Number of unknowns $n_{pix} = 248004$, computation time 12 min 36 s.
- Right: Multiresolution model with $J_{roi} = 6$ and $J_{out} = 1$. Number of unknowns $n = 29130$, computation time 2 min 14 s.

S-curve method

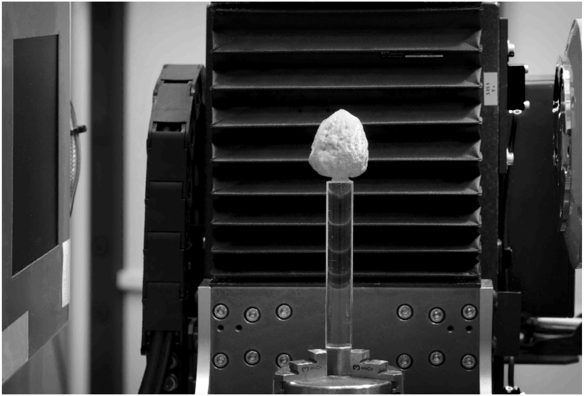
About selection of the prior parameter α

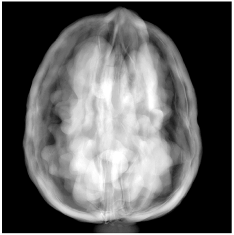
- In many applications, expected level of sparsity (e.g. # non-zero wavelet coefficients, TV norm) ca be extracted from anatomical atlases or previous CT-scans of the patient (e.g. IMGRT). How could we utilize this information in tuning the prior model?
- Let $\hat{S}$ denote the expected level of sparsity and let $S(\alpha)$ denote the sparsity of the reconstruction f_α. The idea in the *S-curve* method is to select α such that

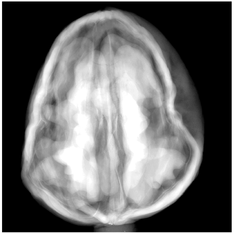
$$S(\alpha) = \hat{S}$$

- We tested the idea using sparse X-ray data from a walnut. Sparsity level $\hat{S}$ was extracted from photographs.

We collected X-ray projection data of a walnut from 1200 directions

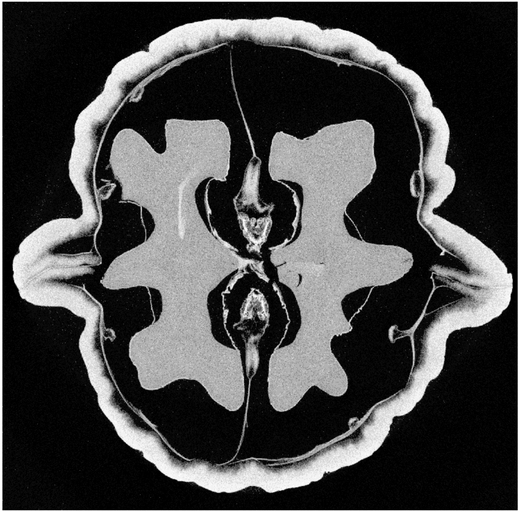




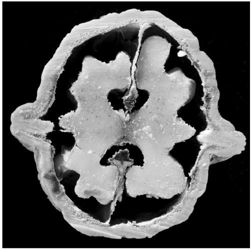


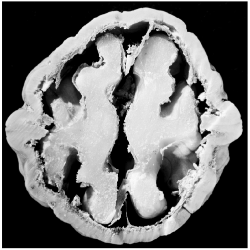
The data was collected by Keijo Hämäläinen and Aki Kallonen at University of Helsinki.

This is the reconstruction using all 1200 projections and filtered back-projection



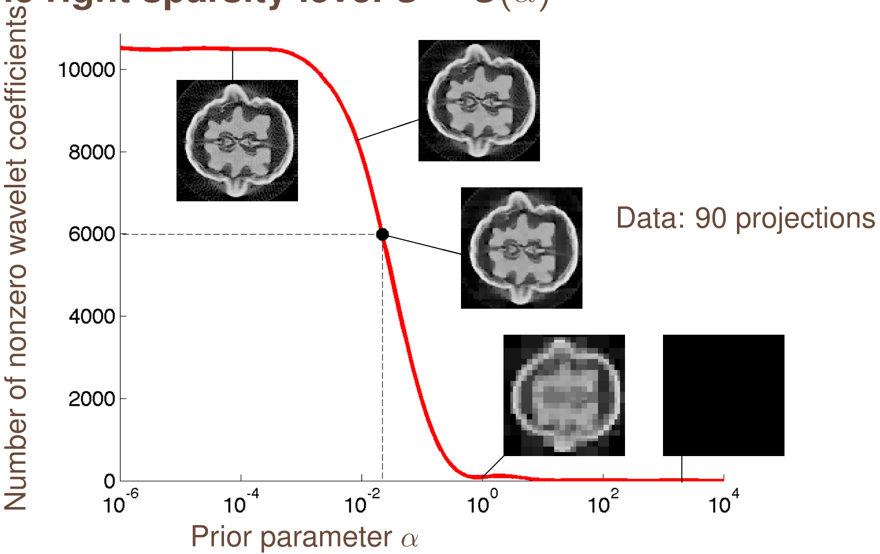
We took photographs of walnuts cut in half





These photos are used for estimating the expected sparsity level $\hat{S}$ in a two-dimensional tomographic reconstruction. Special thanks go to Esa Niemi for his careful job in sawing the walnuts.

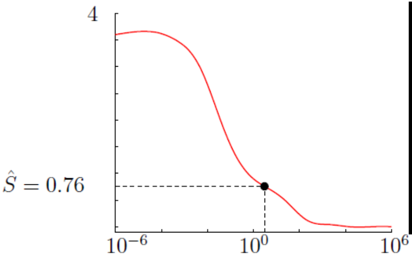
The S-curve method determines value of α giving the right sparsity level $\hat{S} = S(\alpha)$

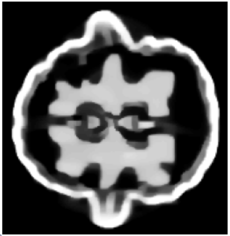


Kolehmainen, Lassas, Niinimäki & Siltanen **2012**

S-curve for TV prior

- Sparsity measured directly in TV norm $S(\alpha) = TV(f_\alpha)$
- The photographs were scaled by $\tilde{f}_p = \frac{\|m\|}{\|Af_p\|} f_p$ to intensity levels that can be expected in the tomographic reconstruction.





$n = 256$
$\alpha = 3$

Filtered back-projection vs TV prior (90 projections)