What is sufficient data for stable CT reconstruction?

Bill Lionheart, School of Mathematics, University of Manchester

School of Mathematics, Univ. of Manchester and Otto Mønsted Visiting Professor at DTU

November 4, 2013

Thanks to Otto Mønsted

I am Otto Mønsted Visiting Professor here.

Thanks to Otto Mønsted

I am Otto Mønsted Visiting Professor here.

Thanks to Otto Mønsted

I am Otto Mønsted Visiting Professor here.

Thanks to Otto Mønsted

I am Otto Mønsted Visiting Professor here.

Thanks to Otto Mønsted

I am Otto Mønsted Visiting Professor here.

Stable or unstable?

An x-ray CT inversion problem with a specific data set may be stable or unstable. In the context on *continuum* data

- We will call a problem *stable* if it can be stably inverted in some Sobolev space.
- for example filtered back projection for the 2D Radon transform is a stable inverse of the Radon transform with complete data in the right Sobolev spaces.
- It is not stable in L^2 of course.
- So what is unstable?
- The *Fourier slice theorem* connects the FT of a 2D image and its Radon Transform.
- *Paley-Weiner says FT of a compactly supported function is analytic*
- Some limited data problems for the 2D Radon transform can be interpreted as analytic continuation....
- but of course this is unstable!
- Typically when a continuum problem is discretized a stable problem will have singular values that decay like a *negative power* while unstable problem *faster than any negative power*.

Under, correctly or over determined

We will say a problem is correctly determined if there is just enough data to guarantee a unique solution *even if it is unstable*. Over determined if a subset of data gives a correctly determined problem and under determined if the solution is not unique.

Examples

Under, correctly or over determined

We will say a problem is correctly determined if there is just enough data to guarantee a unique solution *even if it is unstable*. Over determined if a subset of data gives a correctly determined problem and under determined if the solution is not unique.

Examples

- For 2D the $[0, \pi)$ sinogram data parallel beam or for fan beam the $[0, \pi + \theta_{\text{fan}})$ data are *correctly determined*

Under, correctly or over determined

We will say a problem is correctly determined if there is just enough data to guarantee a unique solution *even if it is unstable*. Over determined if a subset of data gives a correctly determined problem and under determined if the solution is not unique.

Examples

- For 2D the $[0, \pi)$ sinogram data parallel beam or for fan beam the $[0, \pi + \theta_{\text{fan}})$ data are *correctly determined*
- Some choices of 4D data in 3D x-ray CT, for example laminography with an open set of projection angles, are overdetermined even though highly unstable (Payley-Weiner again)

Under, correctly or over determined

We will say a problem is correctly determined if there is just enough data to guarantee a unique solution *even if it is unstable*. Over determined if a subset of data gives a correctly determined problem and under determined if the solution is not unique.

Examples

- For 2D the $[0, \pi)$ sinogram data parallel beam or for fan beam the $[0, \pi + \theta_{\text{fan}})$ data are *correctly determined*
- Some choices of 4D data in 3D x-ray CT, for example laminography with an open set of projection angles, are overdetermined even though highly unstable (Payley-Weiner again)
- For a source trajectory that does not satisfy Tuy's conditions Grangeat's formula, cone beam data may still give a unique reconstruction, so *correctly determined or over determined* but unstable

Under, correctly or over determined

We will say a problem is correctly determined if there is just enough data to guarantee a unique solution *even if it is unstable*. Over determined if a subset of data gives a correctly determined problem and under determined if the solution is not unique.

Examples

- For 2D the $[0, \pi)$ sinogram data parallel beam or for fan beam the $[0, \pi + \theta_{\text{fan}})$ data are *correctly determined*
- Some choices of 4D data in 3D x-ray CT, for example laminography with an open set of projection angles, are overdetermined even though highly unstable (Payley-Weiner again)
- For a source trajectory that does not satisfy Tuy's conditions Grangeat's formula, cone beam data may still give a unique reconstruction, so *correctly determined or over determined* but unstable
- For a helical source trajectory with the Tam-Danielson [Tam] window of cone beam data, there is a stable reconstruction. So his is stable and *correctly determined*.
- Circular scan cone beam data does not satisfy [Tuy]'s condition so there is no stable reconstruction.

Under, correctly or over determined

We will say a problem is correctly determined if there is just enough data to guarantee a unique solution *even if it is unstable*. Over determined if a subset of data gives a correctly determined problem and under determined if the solution is not unique.

Examples

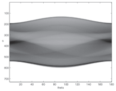
- For 2D the $[0, \pi)$ sinogram data parallel beam or for fan beam the $[0, \pi + \theta_{\text{fan}})$ data are *correctly determined*
- Some choices of 4D data in 3D x-ray CT, for example laminography with an open set of projection angles, are overdetermined even though highly unstable (Payley-Weiner again)
- For a source trajectory that does not satisfy Tuy's conditions Grangeat's formula, cone beam data may still give a unique reconstruction, so *correctly determined or over determined* but unstable
- For a helical source trajectory with the Tam-Danielson [Tam] window of cone beam data, there is a stable reconstruction. So his is stable and *correctly determined*.
- Circular scan cone beam data does not satisfy [Tuy]'s condition so there is no stable reconstruction.
- ....note that means that [FDK] cannot be exact as it *is* stable.

Tam-Danielson window

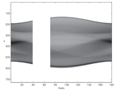
The Tam-Danielson window for a helical source trajectory is a subset of the detector plane.

Types of limited data tomography

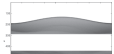
There are stable exact methods for some 2D problems with missing data. Others are unstable or underdetermined. See [Quinto] for an introduction. Lines in the plane given by the angle θ and distance from origin s. Density f, Radon transform $Rf(\theta, s)$



- In *limited angle* tomography, the data is only known for certain values of θ. This might occur in medical applications when a patient is attached to some apparatus, or in non-destructive testing when some views are obscured by some device used for live loading.

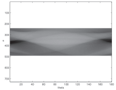


- *The exterior problem* concerns the case where the data $Rf(\theta, s)$ is known only for $|s| > a$ for some $a > 0$. Eg object near centre too dense for the x-rays to penetrate.



Types of limited data tomography: interior

- *The interior problem*,data of the form $Rf(\theta, s)$ for $|s| < a$, and some $a > 0$. Eg the detector width too small to cover the entire object being scanned.



We will consider examples of two algorithms

- Λ-tomography which produces an image of a large area of the domain but only gets jump changes right.
- Two step Hilbert Transform method which gives an exact reconstruction on part of the domain.

Reconstruction formulae

R is the Radon transform and $R^{\#}$ its adjoint the *backprokection* operator.
The Reisz potential is the operator $I^{\alpha} f(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix\cdot\omega} |\omega|^{n-1-\alpha} \hat{f}(\omega) \mathrm{d}\omega$ where the hat denotes Fourier transform and n is the dimension.
The Radon Transform can be inverted using the formula

$$f = \frac{1}{2}(2\pi)^{-1} I^{-\alpha} R^{\#} I^{\alpha-1} Rf$$

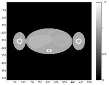
for any $\alpha < 2$. For $\alpha = 0$ this is Filtered Back Projection

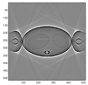
$$f = \frac{1}{2}(2\pi)^{-1} R^{\#} I^{-1} Rf$$

here I^{-1} is the *ramp filter* applied to the s variable.
The basis of *Lambda tomography* [Smith] is to take $\alpha = -1$.

$$\sqrt{-\Delta} f = \mathrm{R}^{\#}\, \mathrm{I}^{-2}\, \mathrm{R}\, f.$$

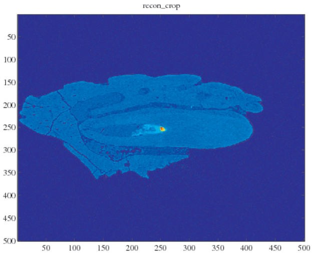
where Δ is the Laplacian. The name is derived from $\Lambda = \sqrt{-\Delta}$, which is a *pseudo differential operator* of order one, and so preserves discontinuities of f.
On the right $\mathrm{I}^{-2} = -d^2/ds^2$ this filter is a local operation, and only the rays on each side of a given ray is needed to calculate it.

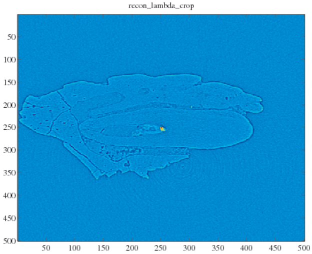




Crocodile fossil Lambda tomography

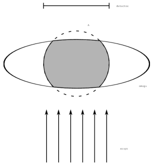
To illustrate Λ-tomography we applied the method to X-ray data of part of a fossilised crocodile, data courtesy of Paul Tafforeau, ESRF. On the left is the filtered backprojection reconstruction and on the right the Λ-tomography reconstruction. Actually with complete data. Note the discontinuities show up well in Λ-tomography but the values are wrong and homogeneous regions become smoothly varying.





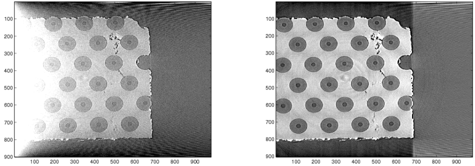
Two step Hilbert transform method

In the interior problem we call *Region A* the part of the domain in which we measure all the lines through each point



The two step Hilbert transform method [Noo] is an exact reconstruction formula for part of region A.
If parts of region *A* lie outside the object then the method gives and exact reconstruction along any straight line inside *A* that intersects both edges of the object. The method involves differentiation of the data, back projection and then inversion of a truncated Hilbert transform in one variable.

Two step Hilbert on composite



A composite sample with a simulated dense block on one end. On the left reconstructed with Filtered Backprojection, and on the right with the two step Hilbert transform method. Reconstruction by David Szotten, experiment in collaboration with Jim Bennett Yu-Chen Hung and Francisco Garcia-Pastor (Materials Science, Manchester)

Source trajectories

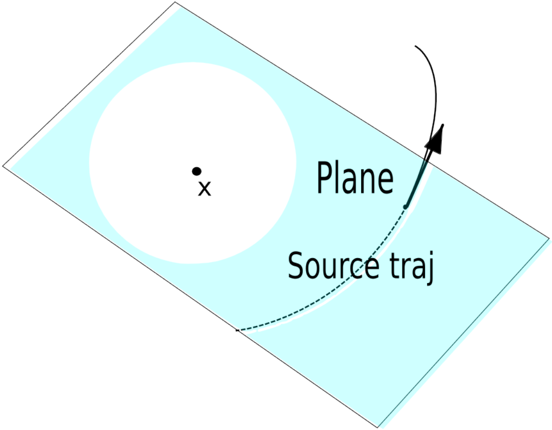
For parallel beam tomography, we can consider the problem as 2D slices (apart from regularisation between slices). So in three dimensions the most interesting case is *cone beam tomography*, where we measure lines going through some curve the *source trajectory*. Two important question considering teh sufficiency of data and stability of reconstruction

- What are good and bad source trajectories?
- Which of the rays intersecting the object do we measure? (detector size)

as a general rule **If you don't measure rays making glancing blows to singularities you can't stably find that singularity** (singularity is usually a surface of discontinuity).

Tuy's condition

More specifically a curve external to the object satisfies the Tuy condition if **any plane through the object intersects the curve cleanly**.



For a example a helix, or circle and line, satisfy Tuy's condition whereas a circle does not.

..source trajectories cont.

[Finch] now tells us

All the rays intersecting a closed curve external to the object are sufficient for reconstruction, but the reconstruction is stable if and only iff Tuy's condition is satisfied

Eg circular scan trajectory unstable, helix stable.

More specifically the microlocal methods of eg [Finch1] show that the reconstruction is unstable "in the direction" in which you do not have Tuy data (ie no glancing blows).

Clearly to make a stable reconstruction we have to impose constraints (a priori knowledge) on the smoothness in these directions.

Cone beam reconstruction algorithms

Here are some examples of algorithms.
$g(a, \theta) := Xf(a, \theta) = \int f(a + t\theta)\mathrm{d}t$

- Feldkamp-Davis-Kress [FDK] is the most widely used algorithm. It is an extension of fan beam filtered back projection to approximately correct for out of plane effects. Also used for helical scan data.

Cone beam reconstruction algorithms

Here are some examples of algorithms.
$g(a, \theta) := Xf(a, \theta) = \int f(a + t\theta)\mathrm{d}t$

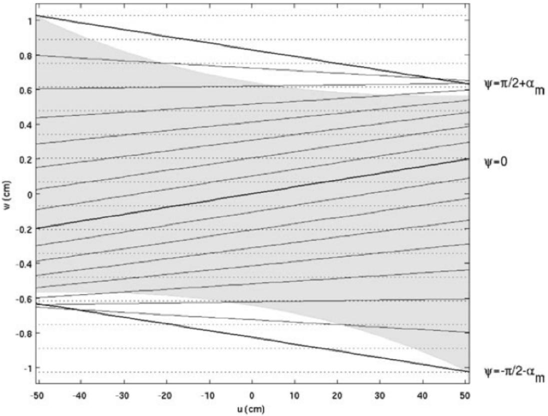
- Feldkamp-Davis-Kress [FDK] is the most widely used algorithm. It is an extension of fan beam filtered back projection to approximately correct for out of plane effects. Also used for helical scan data.
- Tuy's formula [Natterer p174]. $a(\lambda)$ source trajectory.

$$f(x) = \frac{1}{\mathrm{i}(2\pi)^{3/2}} \int_{S^2} \frac{1}{a'(\lambda) \cdot \theta} \frac{\mathrm{d}}{\mathrm{d}\lambda} g(a(\lambda), \cdot\hat{)}(\phi) \, \mathrm{d}\phi$$

Note this needs every ray intersecting the support of f and the source curve.

Cone beam reconstruction algorithms cont

Katsevich's exact reconstruction algorithm [NPH] uses slightly more data than the Tam-Danielson window.



The shaded area shows the TD window the larger area including the "κ-lines" is needed, but not the whole sensor.

Cone beam reconstruction algorithms cont

Katsevich's reconstruction algorithm still involves backprojecting filtered data. But for each point x and source point $a(\lambda)$ the filter is taken along the κ-lines $\xi(\lambda, x, \gamma)$.

$$f(x) = \frac{1}{2\pi} \int \frac{1}{||x - a(\lambda)||} \mathrm{PV} \int_0^{2\pi} \frac{\partial}{\partial q} g(a(q), \xi(\lambda, x, \gamma)) \Big|_{q=\lambda} \frac{\mathrm{d}\gamma \mathrm{d}\lambda}{\sin \gamma}$$

There are other forms more suitable for computation eg [NPH].

Why do lab cone beam CT systems still use FDK?

and indeed only circular trajectories?

Why do lab cone beam CT systems still use FDK? cont

One reason is that Katsevich has a reputation for being unstable, although it is stable in the Sobolev sense (one derivative)

Why do lab cone beam CT systems still use FDK? cont

One reason is that Katsevich has a reputation for being unstable, although it is stable in the Sobolev sense (one derivative)
Actually the problem is more likely that it uses *closer to* an exactly determined data set.

Why do lab cone beam CT systems still use FDK? cont

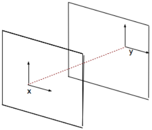
One reason is that Katsevich has a reputation for being unstable, although it is stable in the Sobolev sense (one derivative)
Actually the problem is more likely that it uses *closer to* an exactly determined data set.
This means if the geometry is slightly wrong it produces plausible sharp but wrong images, where as FDK blurs images with inconsistent data.

Why do lab cone beam CT systems still use FDK? cont

One reason is that Katsevich has a reputation for being unstable, although it is stable in the Sobolev sense (one derivative)
Actually the problem is more likely that it uses *closer to* an exactly determined data set.
This means if the geometry is slightly wrong it produces plausible sharp but wrong images, where as FDK blurs images with inconsistent data.
One solution [Varslot] is to correct the geometry by ‘focusing’ inexact reconstructions. Another would be to use the data outside the TD-κ window as a consistencey check.

4D spatial data

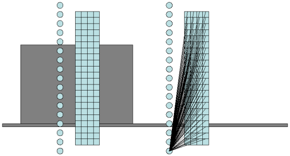
- **Image volume is a function of 3 variables**
- **the set of lines in space is 4 dimensional**
- ..so when sources are used on a sample of a surface rather than curve we have 4D data

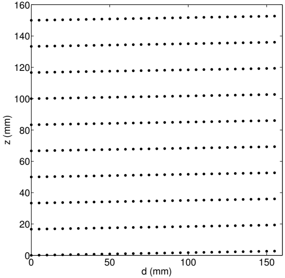


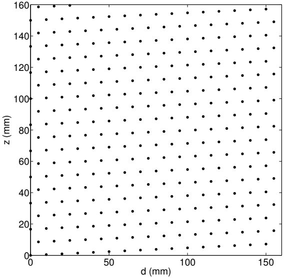
- In this case we have John's equation
- $g(x, y) = \int f(x + t(y - x))\, \mathrm{d}t \qquad \frac{\partial^2 g}{\partial x_1 \partial y_2} - \frac{\partial^2 g}{\partial x_2 \partial y_1} = 0$
- explicit consistency relation on data allows us to check for and reduce errors.
- Examples include laminography and electron CT with a two axis tilt stage, and systems with multiple switched sources .. such as RTT.

The Rapiscan RTT

RTT System Geometry







Standard approach to linear inverse problems

In any linear inverse problem we can reduce to solving the ill-conditioned equation

$$\mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m, \mathbf{A} \in \mathbb{R}^{m\times n},$$

Typically $m > n$.
As $\mathbf{A}$ is *ill-conditioned* we solve instead a regularized problem

$$\mathbf{x}_{\text{Tik}} = \text{arg min}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \|\mathbf{L}\mathbf{x}\|^2$$

This is equivalent to finding the least squares solution of the augmented problem

$$\mathbf{x}_{\text{Tik}} = \text{arg min}_{\mathbf{x}} \left\|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}}\right\|^2$$

where

$$\tilde{\mathbf{A}} = \begin{bmatrix}\mathbf{A}\\ \mathbf{L}\end{bmatrix}, \qquad \tilde{\mathbf{b}} = \begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}$$

Iterative solution

- For matrices that are very large, and usually very sparse, iterative methods are preferred to direct solution of the least squares problem.
- Each iteration typically involves the application of the matrix and its transpose to vectors and these operations are implemented without storing the matrix. We refer to these operations as *forward projection* and *back projection* when $\mathbf{A}$ is the matrix of an x-ray transform.
- Slowly convergent methods, such as Landweber (SIRT) or Kaczmarz (ART) are often employed but the iteration stopped before the residual error falls below the data error.
- ..but in this case choice of algorithm determines solution!!
- Systematic methods choose the regularisation penalty term and then use a quickly converging iterative algorithm such as CGLS
- The Tomography Toolbox [ASTRA] has a CGLS implementation using GPUs you can use.
- Algebraic iterative methods can handle any source or measurement configuration.

CGLS

CGLS is a version of Conjugate gradient for non-square matrices conversing to the least squares solution $x^{(k)} \to A^{\dagger} b$.

Given initial guess $x^{(0)}$, set

$$
\begin{aligned}
& r^{(0)} = b - Ax^{(0)} \\
& p^{(0)} = s^{(0)} + A^T r^{(0)} \\
& \gamma_0 = \|s^{(0)}\|_2^2 \\
& k = 0 \\
& \text{while } \|r^{(k)}\| > \text{acceptable error} \\
& \quad q^{(k)} = Ap^{(k)} \\
& \quad \alpha_k = \gamma_k / \|q^{(k)}\|^2 \\
& \quad x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)} \\
& \quad r^{(k+1)} = r^{(k)} - \alpha_k q^{(k)} \\
& \quad s^{(k+1)} = A^T r^{(k+1)} \\
& \quad \gamma_{k+1} = \|s^{(k+1)}\|_2^2 \\
& \quad \beta_k = \gamma_{k+1} / \gamma_k \\
& \quad p^{(k+1)} = s^{(k+1)} + \beta_k p^{(k)} \\
& \quad k = k + 1
\end{aligned}
$$

Note that in each iteration only one multiplication of a vector by A and one by A^T is needed.

Choice of regularization

- There is a systematic Bayesian approach in which regularization is interpreted as a*prior* distribution [Kaipio].
- In this context we need a strong enough prior so that the instability is removed in the augmented system. Severely illposed problems, or more noisy data need stronger priors.
- Incorporating the distribution of errors, including correlated errors, on your data means that you don't try so hard to fit more erroneous data!
- with systematic regularization at least you know what you assumed. You can then ask how robust your conclusions are with respect to choice of prior.
- Even though parallel beam CT decouples planes, smoothing priors couple them [Hahn].

Over and under determined

In algebraic iterative methods the first thing we often do with the data is back-project.

$$\ker \mathbf{A}^T = (\mathrm{range}\mathbf{A})^\perp$$

so we project on to consistent data.
To a large extent this handles inconsistency due to small errors well.
In limited data problems where **A** has a large null space, the component in the null space will be determined by the prior.
Algebraic iterative methods can handle

- irregularly spaced projection angles
- missing data
- (several levels) of "zoom" data
- arbitrary trajectories

Spatio-temporal regularization

When a system is changing quickly relative to the scan speed but in a way that time steps are highly correlated one can use algebraic iterative methods where the regularization couples space and time.
The time series of images is assembled as a a single vector, each projection is taken as a block of rows in the augmented matrix and a regularization matrix **L** couples space and time.
As a 2D+Time example letting n be the number of image pixels in the x and y directions, and letting p be the number of frames, **L** might have the following Kronecker product decomposition:

$$\mathbf{L} = \alpha_s \mathbf{I}_p \otimes \mathbf{I}_n \otimes \mathbf{D}_n + \alpha_s \mathbf{I}_p \otimes \mathbf{D}_n \otimes \mathbf{D}_n + \alpha_t \mathbf{D}_p \otimes \mathbf{I}_n \otimes \mathbf{D}_n$$

where $\mathbf{D}_n$ is a difference operator in one dimension and $\mathbf{I}_n$ is the $n \times n$ identity while α_s and α_t are special and temporal regularization parameters.

Oil-water fast tomography

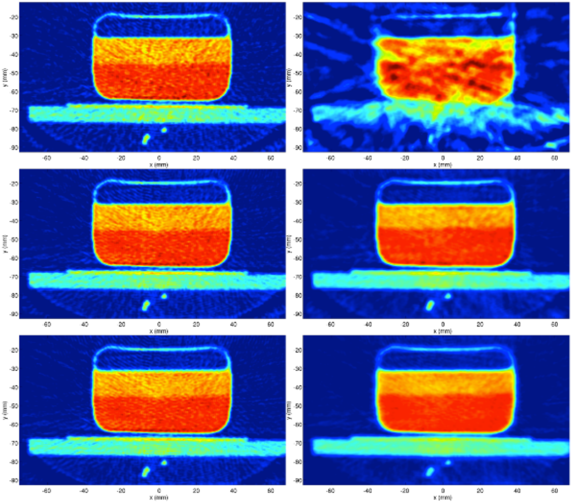


Figure 1: Reconstructed images of a single frame from the oil and water data (left, 245 projections; right, 49 projections; top – bottom, $\alpha_t = 0$, $\alpha_t = 5$, $\alpha_t = 25$)

Rapiscan RTT20 data from [Thompson]

GPU implementation

Recent GPUs are making algebraic iterative methods viable as far as computational time Typical timings for our CGLS code with the GPU based Siddon type forward/back projector, 720 cone beam projections of size 512×512 into a reconstruction volume of $512 \times 512 \times 512$ voxels. Around 20 iterations are typical (with no regularization).

- Forward projection: 5.95 seconds
- Matched back projection: 7.4 seconds
- Total time per iteration: 14.65 seconds (so time for all other CGLS vector operations = 1.3s)
- Specifications of system
  - CPU: 2x 8-core Intel Xeon E5-2687W @ 3.1GHz
  - RAM: 64GB
  - GPU: nVidia Quadro 6000 (448 cores, 6GB RAM, Fermi architecture)
  - OS: Windows 7 professional 64 bit

Data from Will Thompson.

References

[Noo] F Noo, R Clackdoyle, and JD Pack. A two-step Hilbert transform method for 2D image reconstruction. Phys. In Medicine Biol., 49(17):3903– 3923, 2004.

[Smith] KT Smith and F Keinert. Mathematical foundations of computed tomography. Applied Optics, 24, 3950–3957, 1985.

[Szotten] D Szotten, Limited data problems in x-ray and polarized light tomography, PhD thesis, University of Manchester, 2011

[Finch1]DV Finch, Cone Beam Reconstruction with Sources on a Curve, SIAM Journal on Applied Mathematics, 45, 665–673.

[John] F John, The Ultrahyperbolic Differential Equation with Four Independent Variables. Duke Math J, 1938, 300–322.

[Finch2] D Finch, I Lan, and G Uhlmann, Microlocal Analysis of the Restricted X-ray Transform with Sources on a Curve, Inside Out, Inverse Problems and Applications (Gunther Uhlmann, ed.), MSRI Publications, vol. 47, Cambridge University Press, 2003, pp. 193218.

[Quinto] ET Quinto, An Introduction to X-ray tomography and Radon Transforms, Proceedings of Symposia in Applied Mathematics,

[Thompson] WM Thompson, WRB Lionheart, EJ Morton, Real-Time Imaging with a High Speed X-Ray CT System, 6th International Symposium on Process Tomography, Cape Town, 2012.

References cont

[Tam] KC. Tam, S Samarasekera, and F Sauer. Exact cone beam CT with a spiral scan. Physics in Medicine and Biology, 43, 1015, 1998.

[FDK] LA Feldkamp, LC Davis, and JW Kress Practical cone-beam algorithm JOSA A, Vol. 1, 612-619, 1984

[Natterer]]F Natterer, The Mathematics of Computerized Tomography, SIAM Classics, 2001.

[NPH] F Noo, J Pack, and D Heuscher, Exact helical reconstruction using native cone-beam geometries. Physics in Medicine and Biology, 48, 3787, 2003.

[Tuy] H. Tuy. An inversion formula for cone-beam reconstruction. SIAM Journal on Applied Mathematics, 43, 546552, 1983.

[Varslot] T Varslot,A Kingston,G Myers,A Sheppard, High-resolution helical cone-beam micro-CT with theoretically-exact reconstruction from experimental data, Med Phys, Oct;38, 5459, 2011

[Kaipio] J Kaipio, E Somersalo, Statistical and Computational Inverse Problems, Series: Applied Mathematical Sciences, Vol. 160, Springer, 2005

[Hahn] B Hahn and AK Louis Reconstruction in the three-dimensional parallel scanning geometry with application in synchrotron-based x-ray tomography, Inverse Problems 28 045013 2012

[ASTRA] WJ Palenstijn, KJ Batenburg, and J Sijbers, The ASTRA Tomography Toolbox, 13th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2013, 2013.