Tomographic image reconstruction: theory and applications to photon and proton tomography

Eleonora Vanzi

Medical Physics Careggi University Hospital Firenze, Italia Tel. 055 7946970 eleonora.vanzi@unifi.it

Summary

- Introduction
- The mathematics of tomographic reconstruction
 - Inverse problems
 - The Radon Transform
 - Reconstruction algorithms
- Examples of photon tomography
- Proton Computed Tomography (pCT)
 - What's different?
 - The PRIMA (PRoton IMAging) experiment: preliminary results

Just to start...

• Two examples of medical imaging techniques using photons

Medical Imaging: Radiology

Radiography



Radiograph of the hand of Albert von Kolliker, made at the conclusion of Roentgen's lecture and demonstration at the Wurzburg Physical-Medical Society on 23 January 1896

Computed Tomography (CT)



(SPECT)

Medical Imaging: Nuclear Medicine



Comparing...

In both examples, two kinds of imaging techniques have been shown

- In the following we will focus on tomographic imaging and describe the methods for their production.
- Observe that tomographic imaging was considered a revolution in Medicine: the inventors of the X-ray CT, Sir Hounsfield and Cormack were awarded Nobel Prize for Physiology or Medicine in 1979
- Tomographic imaging is based on the mathematical procedure called tomographic reconstruction, which requires a series of planar images acquired at different angles



TOMOGRAPHIC RECONSTRUCTION

The Mathematics

Inverse Problems



Mathematically speaking, the reconstruction of an unknown object from its angular projections is a typical "Inverse Problem"

Unknown distribution

- Physically, the *direct problem* is the one formulated in the causeeffect direction: object \longrightarrow projections
- Real measurement systems have a limited pass-band: the direct problem is associated to a loss of information
- In the inverse problem, where object and projections exchange their roles, we deal with a lack of information.

Inverse problems are generally **ill-posed**.

Inverse Problems

- A problem is <u>well posed</u>, according to Hadamard (1865-1963) definition, if the following conditions are satisfied:
 - 1) A solution exists for all possible data sets
 - 2) The solution is unique
 - 3) The solution's behavior hardly changes when there's a slight change in the initial condition (continuous dependence)
- Problems that are not well-posed in the sense of Hadamard are termed **ill-posed**.
- Direct problems are generally well-posed.
- Inverse problems are often ill-posed.



- 1. EXISTENCE: there may be data for which the solution does not exist (for ex. noisy data)
- UNIQUENESS: the inverse problem solution may not be unique (for ex. For objects differing in frequency out of the pass-band)
- 3. CONTINUOUS DEPENDENCE: similar data can be produced by very different objects

Inverse Problems

- Due to the ill-position of inverse-problem and due to the noise on data, it makes no sense searching for "the solution" of the problem. We can search for "a solution" compatible with experimental data.
- The class of possible solutions can be very large
- Techniques aiming to limit the class of possible solutions are said "regularization methods":
 - constraints on the solution (positive, limited amplitude, smoothness degree,....)
 - ➤ use of a-priori information on the transformation kernel
 - ➤ use of noise filters
 - number of iterations in iterative methods

Radon Transform

The Radon Transform is the most simple model for the operating principle of a lot of tomographic acquisition devices (CT, SPECT, PET...)



$$p(s,\theta) = \iint_{-\infty}^{\infty} X(x,y) \quad \delta(s+x \, sen\theta - y \, \cos\theta) dx \, dy$$

Inversion Formula (Radon, 1917):

$$X(r,\varphi) = \frac{1}{2\pi^2} \int_{0}^{\pi} \int_{-\infty}^{+\infty} \frac{\partial p(s,\theta)}{\partial s} \frac{1}{r\sin(\varphi-\theta)-s} ds d\theta$$

The FBP algorithm directly derives from the RT inversion formula

$$X(r,\varphi) = \frac{1}{2\pi^2} \int_{0}^{\pi} \int_{-\infty}^{+\infty} \frac{\partial p(s,\theta)}{\partial s} \frac{1}{r\sin(\varphi-\theta)-s} ds d\theta =$$

= $\frac{1}{2\pi^2} \int_{0}^{\pi} \left[\frac{\partial p(s,\theta)}{\partial s} * \frac{1}{s} \right]_{r\sin(\varphi-\theta)} d\theta$ where the integration over θ constitutes the **back**-projection operator

Let us call *S* the spatial frequency in the projection space and $\tilde{P}(S,\theta)$ the Fourier transform of projections with respect to the spatial variable *s*.

Let's now remember that

$$F\left[\frac{\partial p(s,\theta)}{\partial s} * \frac{1}{s}\right] = F\left[\frac{\partial p(s,\theta)}{\partial s}\right] \cdot F\left[\frac{1}{s}\right] \quad \text{and} \quad \begin{bmatrix}F\left[\frac{\partial p(s,\theta)}{\partial s}\right] = 2\pi \ i \ S \ \widetilde{P}(S,\theta) \\ F\left[\frac{1}{s}\right] = -\pi \ i \ \text{sgn}(S)$$

We can rewrite the previous equation using the FT of the integrand and backprojecting its inverse FT

Filtered Back Projection (FBP)
$$X(r,\varphi) = \frac{1}{2\pi^2} \int_{0}^{\pi} \int_{-\infty}^{+\infty} \frac{\partial p(s,\theta)}{\partial s} \frac{1}{r\sin(\varphi-\theta)-s} ds d\theta =$$

$$= \dots = \int_{0}^{\pi} \int_{-\infty}^{+\infty} |S| \widetilde{P}(S,\theta) e^{2\pi i Sr \sin(\varphi-\theta)} dS d\theta$$

Since in the practice projections are discrete, their FT is band limited. Therefore we can replace the ramp function |S| with a function with limited support:

$\tilde{c}(S) = |S| w(S)$

The window function w(S) is a low-pass filter that can have different shapes (Rectangular, Hann, Butterworth,...)



CT

$$X(r,\varphi) = \int_{0}^{\pi} \int_{-\infty}^{+\infty} \widetilde{c}(S) \widetilde{P}(S,\theta) e^{2\pi i Sr \sin(\varphi-\theta)} dS d\theta \qquad \widetilde{c}(S) = |S| w(S)$$

The FBP algorithm is based on this equation

FBP 1:
$$X = B\{F^{-1}[\tilde{c} F(p)]\}$$

A more efficient implementation can be written using the convolution theorem

FBP 2:
$$X = B \{ c * p \}$$



FIGURE 25-16

Backprojection. Backprojection reconstructs an image by taking each view and *smearing* it along the path it was originally acquired. The resulting image is a blurry version of the correct image.



Lakshminarayanan weights

FIGURE 25-17

Filtered backprojection. Filtered backprojection reconstructs an image by filtering each view before backprojection. This removes the blurring seen in simple backprojection, and results in a mathematically exact reconstruction of the image. Filtered backprojection is the most commonly used algorithm for computed tomography systems.

Modified Radon Transform

In general, the imaging systems PSF may differ from a Dirac Delta function due to many physical factors:

- detector responses
- particles interactions

• ...

In order to obtain a reliable solution the projection equation must be modified

$$p(s,\theta) = \iint_{-\infty}^{\infty} X(x,y) (F(s+x \, sen \, \theta - y \, \cos \theta) \, dx \, d$$

Analytical methods can not be used anymore to find the solution.

The projection equation is discretized and numerical methods are used to solve the system

$$p_{jn} = \sum_{k} F_{jn}^{k} X^{k}$$

 p_{jn} = value of the projection bin *n* acquired at angle *j* X^k = value of the *k*-th object voxel F^k_{jn} = contribution of the *k*-th object voxel to the *n*-th acquired at angle *j*

 $p(s, \theta)$

(x,y) Object

projection

Reconstruction algorithms

$$p_{jn} = \sum_{k} F_{jn}^{k} X^{k}$$

Making a tomographic reconstruction is solving a linear system of equations.

The choice of the algorithm depends on the computational burden, especially when used in the clinical routine (this explains why FBP survived until now!).

Some examples:

• Algebraic methods: ART

OSEM (Ordered Subsets Expectation Maximization)

- **Maximum Likelihood** statistical methods: the function to be maximized is defined on the base of the projection statistical properties. For ex. EM algorithm for Poissonian data
- **Statistical Bayesian** methods: generalization of the previous class, including a priori information about the solution characteristics.
- Least Squares methods: a quadratic function is minimized. No information about data statistical properties.. Ex: *Conjugate Gradients*

Examples: SPECT Let us suppose... this is the patient!



The patient can be opened and his striata filled with radioactive material.





Examples: SPECT

The patient is scanned...





For example: 40 frames (120 angles) 60 sec/frame (40 min acquisition)

...and the image is reconstructed! (OSEM)



Example: PET

• A **positron emitting** radio-agent is injected.

• The two back-to-back 511KeV photons produced by positron annihilation are detected in coincidence by a **ring of detectors** surrounding the patient.



Coincidences along the same angular directions define the projections.





...and the image is reconstructed (OSEM)!



...and the map of X-ray attenuation coefficient *inside* the body can be obtained through tomographic reconstruction of the acquired *projections*.

pCT: PROTON COMPUTED TOMOGRAPHY

Proton Radiotherapy

- Proton radiotherapy exploits the fact that proton energy loss shows a maximum at the end of the particle path (Bragg Peak), as described by the Bethe-Bloch equation.
- Conveniently choosing the proton energy, the peak can be positioned on the therapy target, with reduced dose to healthy organs positioned in front and behind the target volume.



Fig. 3. Proton energy loss in water as a function of depth for two incident proton energies (without energy-range straggling). The open symbols indicate the energy of the protons, and closed symbols the energy deposited in 1 mm water.

> Typical energies and ranges in «A150 tissue equivalent palstic»: 130 MeV \rightarrow 12 cm 200MeV \rightarrow 25.8 cm 250 MeV \rightarrow 37.7 cm

Proton Radiotherapy treatment planning

• In proton radiotherapy, to accurately know and plan the local energy deposition, the knowledge of the stopping power distribution at the proton energy is required.

• The stopping power is given by the Bethe-Bloch equation



and mass numbers

Proton Radiotherapy treatment planning

• Attempts have been made to obtain the Stopping Power map through conversion of CT numbers, but the method is inaccurate and produces an unacceptable uncertainty on proton range calculations

•The most accurate way to determine this information is through tomographic imaging performed directly with protons.



Other advantages:

•PCT can also be used for pretreatment patient positioning verification

• The dose delivered to the patient in proton imaging is lower than that delivered with X-ray imaging

PCT imaging equations

Many equations have been proposed in literature. We will consider that proposed by **Wang**, *Med. Phys.* 37(8), 2010, p. 4138



Multiple Coulomb Scattering

The "true" path of protons in a thickness *x* of material is determined by a sequence of elastic collisions with the atomic nuclei.



Due to the random nature of interactions it is not possible to know the true path. However, the distribution of the exit displacement (y) and exit angle (θ) is described by two Gaussian functions centered in the initial position/direction.

$$\overline{y} = 0 \qquad \sigma_y \propto \sigma_\theta x / \sqrt{3}$$

$$\overline{\theta} = 0 \qquad \sigma_\theta \propto \sqrt{x / X_0}$$
Ex: 250 MeV
protons in 20 cm
H_2O:
 $\sigma_{\theta} \sim 1^\circ, \sigma_{y} \sim 3$ mm

X_o radiation length

Most Likely Path (MLP)

If the entry and exit positions and directions of a proton that traverses a certain thickness of material are known, then it is possible to estimate the Most Likely Path of the proton within the medium.



PCT scanner design



• Two entrance and two exit Position Sensitive Detectors

• A calorimeter

- Indeed, for each protons it is necessary to measure:
 - entrance position and direction
 - exit position and direction
 - residual energy: E_{res}
 - (entry energy is known from the accelerator)

$$\int_{Path} S(x, y, E_0) dl = \int_{E_{res}}^{E_0} \left[\frac{S}{\rho} (H_2 O, E_0) \middle/ \frac{S}{\rho} (H_2 O, E) \right] dE$$



Iterative vs. FBP

• **ITERATIVE**:

- recovery of spatial resolution
- long modeling and computational time

• <u>FBP</u>

approximate solution



 very fast and easy (the reconstruction of a 256x256x256 image volume required 22 seconds on a standard personal computer, Intel Core i3-380M CPU, 4GB RAM, 64 bit Linux Operating System)

(a)
 (a)
 (b)
 (c)
 (c)

•

If FBP is not able to produce a pCT image with sufficient accuracy for treatment planning, FBP images can however have sufficient resolution for a first assessment of the object. Therefore FBP can be the algorithm of choice when a pCT image has to be produced in a short time, such as for:

- patient positioning verification in proton treatment facilities
- producing an image that can be used as the starting point for iterative methods

The PRIMA collaboration

Preliminary results in FBP reconstruction of pCT data

Presented at RESMDD 2012, 9-12 Oct. 2012, Florence, Italy (in press in NIMA Proceedings)

The PRIMA collaboration

PRIMA (Proton IMAging) is an Italian collaboration, involving researchers from Istituto Nazionale di Fisica Nucleare (Catania, Cagliari and Firenze Sections and Laboratori Nazionali del Sud) and from the University of Firenze, Catania and Sassari. The collaboration is devoted to the design and manufacture of a proton imaging prototype system.



<u>The PRIMA detector</u> Silicon microstrip tracker, followed by a YAG:Ce calorimeter. The traker measures the (x,y) coordinates on four planes (P1-P4). The calorimeter measures the proton residual energy.



The beam & the phantom

Results we are going to show have been obtained on experimental data acquired at LNS (Laboratori Nazionali del Sud, Catania, Italy)



Data analysis: the tomographic equation

• Definition of the tomographic equation (Wang, Med.Phys. 37(8), 2010: 4138)

Unknown stopping
power distribution
(at E₀)
$$\int_{Path} S(x, y, E_0) dl = \int_{E_{res}}^{E_0} \left[\frac{S}{\rho} (H_2 O, E_0) / \frac{S}{\rho} (H_2 O, E) \right] dE$$
 (projection)

• Evaluation of the "projection" term (through numerical integration starting from NIST tables and using the measured E_{res})



Mean measured residual energy plotted at P3 plane Corresponding Wang projections

Data analysis: definining protons trajectories

- 1. The proton path in the phantom is curved due to MCS, but FBP can not handle this information, so:
 - We have not to use the MLP formalism but only to define *a straight line* for the event
- 2. The line could be that connecting the entry and exit points on the phantom...

...But FBP should produce a-priori information about the object for successive iterative algorithms, without needing a-priori information itself! So:

• We don't want to use the phantom boundary to define the line

We used the line connecting the impact points on P₂ and P₃, ignoring the information from P₁ and P₄. This means that results we will show could have been obtained with a simpler tracker made with two planes only





Now we have to approssimate these lines with lines perpendicular to the detectors

Data analysis: data rebinning

We defined a plane parallel to detector's planes and passing through the phantom axis.

The plane was sampled in a 256×256 matrix, 200μ m pixel size.

For each event, the associated projection bin was determined by the intersection of the line connecting P2 and P3 impact points with the plane.



Data analysis: data selection

In order to fulfill the FBP requirement of rectilinear trajectories perpendicular to the detector, only events with small deviations from the projection direction *should* be selected.

We can define acceptance intervals (Δx , Δy) and use in FBP reconstruction only events with $|\mathbf{x}_3 - \mathbf{x}_2| < \Delta x$ and $|\mathbf{y}_3 - \mathbf{y}_2| < \Delta y$.



Small acceptance interval means:

- «more Radon» protons (resolution should increase)
- more rejected events

Δx (mm)	Δy (mm)	Stat. res.
0.2	0.2	1%
0.4	0.4	4%
1	1	22.4%

Is it really necessary?

Data analysis: data selection

Let's consider the measured deviations $|x_3-x_2|$ and the experimental acquisition geometry:

- Most of the events have deviations < 3mm
- 3 mm deviations on the P₃-P₂ distance produce at most
 260 μm deviations within the phantom



This effect is comparable to the σ of the lateral displacement of proton trajectories from the line connecting entry and exit points due to MCS (~250 µm), therefore the inclusion of **all** events in FBP reconstruction will not worsen resolution too much. Moreover, the increase of the noise level will partially mask the resolution recovery obtainable with cuts on directions





We used a Butterworth filter as window function. It is described by two parameters: order (n) and cut-off frequency (fc).

The filter is "harder" for:

- lower order
- higher cut-off frequency

"Hard" filter means:higher resolution

2n

• higher noise

Results





Images: 256×256×256, 200 µm pixel

Butterworth filter: order 2, cut-off 20/128 of the Nyquist freq.

Good quality images without cuts: possibility of reducing acquisition times and dose in patient procedures



Results

pCT images have been coregistered to phantom images obtained with a microCT scanner (GE Flex-XO) used for small-animal imaging in preclinical studies. No distortions are highlighted by the image fusion.



Resolution evaluation

The edge of the phantom was fitted with an *erf* function on 20 slices in the homogeneous region.

The derivative of the *erf* function is a Gaussian function that describes the Line Spread Function (LSF) of the imaging system.



The mean FWHM of the LSF over the 20 slices was evaluated and used to quantify resolution.



FBP filter choice

There is a trade-off between resolution and noise.



Conclusions on FBP in PCT

- Even using the information on two planes only and without cuts on events to be used in reconstruction, good quality images were obtained, sufficient for patient positioning verification
- Resolution could be improved further:
 - During acquisition: refining the angular sampling and increasing the number of events
 - During data analysis: with cuts on angles and energy
- These results are valid for the considered experimental set-up. The good performances of the pCT scanner encourages working on the development of a similar pCT equipment with an enlarged field of view...
-and on iterative algorithms implementing the MLP formalism.

Thank you for your kind attention

The PRIMA collaboration



Silicon microstrip tracker, followed by a YAG:Ce calorimeter. The traker measures the (x,y) coordinates on four planes (P1-P4). The calorimeter measures the proton residual energy.

Microstrip sensor:

- Thickness: 200±15 μm
- Strip pitch: 200 µm
- Number of strips per sensor: 256
- Active area: 51×50.66 mm²

Calorimeter:

- Four YAG:Ce (Yttrium Aluminum Garnet activated by Cerium) scintillating crystals assembled in an optically decoupled 22 matrix.
- Crystal dimensions: 3×3cm²cross-section and 10 cm depth (stop up to 230 MeV kinetic energy protons.

$$\begin{split} X(r,\varphi) &= \int_{0}^{\pi+\infty} \widetilde{C}(S) \ \widetilde{P}(S,\theta) \ e^{2\pi i Sr \sin(\varphi-\theta)} dS \ d\theta \\ F\left[\frac{\partial p(s,\theta)}{\partial s}\right] = 2\pi \ i \ S \ \widetilde{P}(S,\theta) \\ &= \frac{1}{2\pi^2} \int_{0}^{\pi} \left[\frac{\partial p(s,\theta)}{\partial s} * \frac{1}{s}\right] = F\left[\frac{\partial p(s,\theta)}{\partial \theta} \right] \cdot F\left[\frac{1}{s}\right] = -\pi \ i \ \text{sgn}(S) \\ X(r,\varphi) &= \frac{s}{2\pi} \int_{0}^{\pi+\infty} \frac{\partial p(s,\theta)}{\partial \theta} \cdot \frac{1}{s} = \frac{1}{2\pi^2} \int_{0}^{\pi+\infty} \frac{\partial p(s,\theta)}{\partial \theta} \cdot \frac{1}{s} \left[\frac{S}{\rho}(H_2O,E_0) / \frac{S}{\rho}(H_2O,E)\right] dE \\ X(r,\varphi) &= \frac{1}{2\pi^2} \int_{0}^{\pi+\infty} \frac{\partial p(s,\theta)}{\partial s} \frac{1}{r\sin(\varphi-\theta)-s} \ ds \ d\theta \\ &= \dots = \int_{0}^{\pi+\infty} |S| \ \widetilde{P}(S,\theta) \ e^{2\pi i Sr \sin(\varphi-\theta)} dS \ d\theta \end{split}$$