Chapter 13: Image Reconstruction

Set of 178 slides based on the chapter authored by
J. Nuyts and S. Matej
of the IAEA publication (ISBN 78–92–0–143810–2):
Nuclear Medicine Physics:
A Handbook for Teachers and Students

Objective:

To familiarize the student with how 2-D or 3-D images of tracer distribution can be reconstructed from a series of projection images acquired with a gamma camera or a positron emission tomography (PET) system.

Slide set prepared in 2015
R. Fulton (Westmead Hospital and University of Sydney)

IAEA
International Atomic Energy Agency

CHAPTER 13 TABLE OF CONTENTS

13.1 INTRODUCTION

The Inverse Problem

- ☐ Reconstructing an image of the tracer distribution within a patient from projection images acquired with a gamma camera or positron emission tomograph is an example of an ‘inverse problem’.
- ☐ The reconstruction is the inverse of the acquisition.
- ☐ Making software to compute the true tracer distribution from the acquired data turns out to be more difficult than the ‘forward’ direction, i.e. making software to simulate the acquisition.

13.1 INTRODUCTION

There are basically two approaches to image reconstruction: analytical reconstruction and iterative reconstruction.

- ☐ The analytical approach is based on mathematical inversion, yielding efficient, non-iterative reconstruction algorithms.
- ☐ In the iterative approach, the reconstruction problem is reduced to computing a finite number of image values from a finite number of measurements. Iterative inversion tends to require more computer power, but can cope with more complex (and hopefully more accurate) models of the acquisition process.

13.2 ANALYTICAL RECONSTRUCTION

- ☐ The (n-dimensional) X ray transform maps an image of dimension n to the set of all possible line integrals.
- ☐ In all PET and in almost all single photon emission computed tomography (SPECT) applications, the measured projections can be well approximated as a subset of the (possibly attenuated) X ray transform, because the mechanical (SPECT) or electronic (PET) collimation is designed to acquire information along lines (the line of response (LOR)).
- ☐ Consequently, reconstruction involves computing the unknown image Λ from (part of) its X ray transform.

IAEA

13.2 ANALYTICAL RECONSTRUCTION

- PET projections are often represented as a set of projections or a set of sinograms.

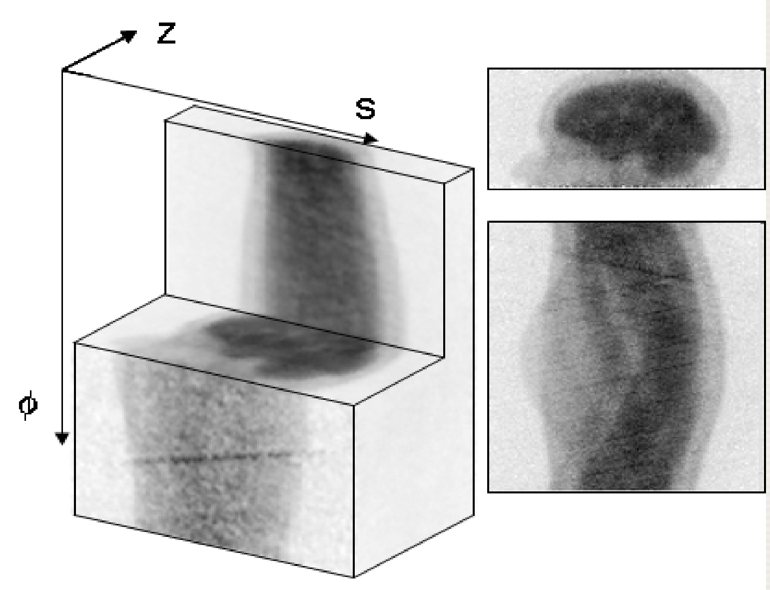


FIG. 13.1. The relation between projections and sinograms in parallel beam projection. The parallel beam (PET) acquisition is shown as a block with dimensions s, ϕ and z. A cross-section at fixed ϕ yields a projection; a cross-section at fixed z yields a sinogram.

13.2 ANALYTICAL RECONSTRUCTION

- The central slice (or central section) theorem gives a relation between the Fourier transform of an image and the Fourier transforms of its parallel projections.
- Using this theorem we can construct the 2D Fourier transform of the image from the 1-D Fourier transforms of the acquired projections.

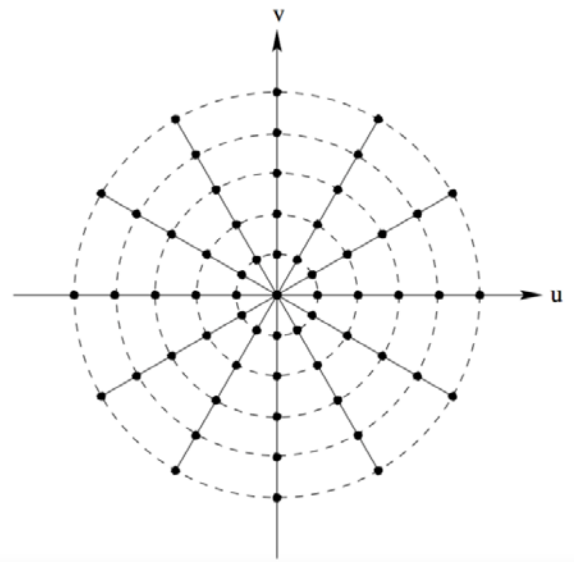


FIG 13.1a. The Fourier transforms of six projections of an object at 30° intervals give estimates of the Fourier transform of the original object along the six onial lines shown. The dots represent the location of estimates of the object's Fourier transform. Note that the Fourier transform is less densely sampled as the distance from the centre increases (i.e. with increasing spatial frequency).

13.2 ANALYTICAL RECONSTRUCTION

- Having so constructed the 2D Fourier transform (FT) of the image, we can obtain the image by simply computing the inverse Fourier transform.
- In practice, this method is rarely used because of the need to convert the radially acquired samples to rectangular by interpolation; the closely related filtered back-projection algorithm is far more popular.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

X ray transform: projection and back-projection

- In 2-D, the Radon transform and X ray transform are identical. Mathematically, the 2-D X ray (or Radon) transform of the image Λ can be written as follows

$$\begin{aligned} \mathrm{Y}(s,\phi) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Lambda(x,y)\,\delta_{s=x\cos\phi+y\sin\phi}\ dx\ dy \\ &= \int_{-\infty}^{\infty} \Lambda(s\cos\phi - t\sin\phi, s\sin\phi + t\cos\phi)\ dt \end{aligned} \tag{13.1}$$

where the δ function is infinite at the points on the LOR (s,ϕ) and zero elsewhere. With the notation used here, $\phi = 0$ corresponds to projection along the y axis.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- Eq. 13.1 describes the acquisition process in 2-D PET and in SPECT with parallel hole collimation, if attenuation can be ignored.
- Assuming that $\Lambda(x, y)$ represents the tracer distribution at transaxial slice *Z* through the patient, then $\Upsilon(s,\phi)$ represents the corresponding sinogram, and contains the z-th row of the projections acquired at angles ϕ.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- The X ray transform has an adjoint operation that appears in both analytical and iterative reconstruction, usually called the back-projection operator

$$
\begin{aligned}
\mathrm{B}(x,y) &= \mathrm{Backproj}\,(\mathrm{Y}(s,\phi)) \\
&= \int_0^{\pi} \mathrm{d}\phi \int_{-\infty}^{\infty} \mathrm{Y}(s,\phi)\,\delta_{s=x\cos\phi+y\sin\phi}\;ds \qquad (13.2) \\
&= \int_0^{\pi} \mathrm{Y}(x\cos\phi + y\sin\phi, \phi)\;d\phi
\end{aligned}
$$

- The back-projection is not the inverse of the projection, $\mathrm{B}(x, y) \neq \Lambda(x, y)$. Intuitively, the back-projection sends the measured activity back into the image by distributing it uniformly along the projection lines.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- ☐ As illustrated in Fig. 13.2, projection followed by back-projection produces a blurred version of the original image. This blurring corresponds to the convolution of the original image with the 2-D convolution kernel $1/\sqrt{x^2+y^2}$.

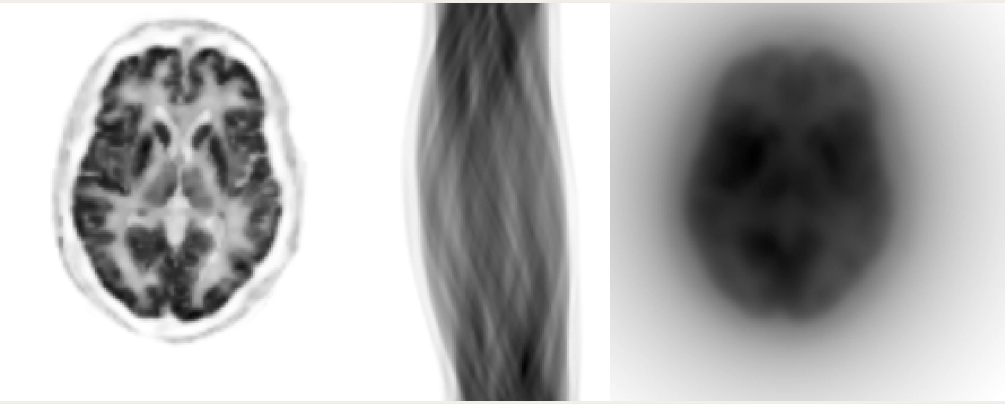


FIG. 13.2. The image (left) is projected to produce a sinogram (centre), which in turn is back-projected, yielding a smoothed version of the original image.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

Central Slice Theorem

- ☐ The central slice theorem gives a very useful relation between the 2-D Fourier transform of the image and the 1-D Fourier transform of its projections (along the detector axis). Consider the projection along the y axis, ϕ = 0, and its 1-D Fourier transform:

$$Y(s,0) = \int_{-\infty}^{\infty} \Lambda(s,t)\; dt \qquad (13.3)$$

$$\begin{aligned}(\mathcal{F}_1 Y)(\nu_s,0) &= \int_{-\infty}^{\infty} Y(s,0)e^{-i2\pi\nu_s s}\; ds \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Lambda(s,t)e^{-i2\pi\nu_s s}\; dt\; ds\end{aligned} \qquad (13.4)$$

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- ☐ If we compare this to the 2-D Fourier transform of the image $\Lambda(x, y)$:

$$(\mathcal{F}_2\Lambda)(\nu_x,\nu_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Lambda(x,y)e^{-i2\pi(\nu_x x+\nu_y y)}\,dx\,dy \qquad (13.5)$$

- ☐ Both expressions are equal if we set $\nu_y = 0$:

$$(\mathcal{F}_1 Y)(\nu_s,0)=(\mathcal{F}_2\Lambda)(\nu_x,0) \qquad (13.6)$$

- ☐ $(\mathcal{F}_1 Y)(\nu_s,0)$ is the 1-D Fourier transform of the projection along the y axis and $(\mathcal{F}_2\Lambda)(\nu_x,0)$ is a ‘central slice’ along the ν_x axis through the 2-D Fourier transform of the image. Equation (13.6) is the central slice theorem for the special case of projection along the y axis.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- ☐ This result would still hold if the object had been rotated or, equivalently, the x and y axes. Consequently, it holds for any angle ϕ:

$$(\mathcal{F}_1 \mathrm{Y})(\nu_s, \phi) = (\mathcal{F}_2 \Lambda)(\nu_s \cos\phi, \nu_s \sin\phi) \qquad (13.7)$$

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

Two dimensional filtered back-projection

- ☐ The central slice theorem (Eq. (13.7)) can be directly applied to reconstruct an unknown image $\Lambda(x, y)$ from its known projections $\mathrm{Y}(s,\phi)$. The 1-D Fourier transform of the projections provides all possible central slices through $(\mathcal{F}_2\Lambda)(v_x,v_y)$ if $\mathrm{Y}(s, \phi)$ is known for all ϕ in an interval with a length of at least π (Tuy's condition). Consequently, $(\mathcal{F}_2\Lambda)(v_x, v_y)$ can be constructed from the 1-D Fourier transform of $\mathrm{Y}(s, \phi)$. Inverse 2-D Fourier transform then provides $\Lambda(x, y)$.
- ☐ However, a basic Fourier method implementation with a simple interpolation in Fourier space does not work well.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- In contrast, in the case of the filtered back-projection (FBP) algorithm derived below, a basic real-space implementation with a simple convolution and a simple interpolation in the back-projection works well. Inverse Fourier transform of Eq. (13.5) yields:

$$\Lambda(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\mathcal{F}_2\Lambda)(\nu_x,\nu_y)e^{i2\pi(\nu_x x+\nu_y y)}\,d\nu_x\,d\nu_y \qquad (13.8)$$

- This can be rewritten with polar coordinates as:

$$\Lambda(x,y)=\int_{-\infty}^{\infty}\mathrm{d}\nu\int_{0}^{\pi}(\mathcal{F}_2\Lambda)(\nu\cos\phi,\nu\sin\phi)e^{i2\pi(x\nu\cos\phi+y\nu\sin\phi)}\left|\nu\right|\,d\phi \qquad (13.9)$$

13.2 ANALYTICAL RECONSTRUCTION
13.2.1 Two dimensional tomography

- ☐ Application of the central slice theorem (Eq. (13.7)) and reversing the order of integration finally results in:

$$\Lambda(x,y)=\int_0^{\pi}\mathrm{d}\phi\int_{-\infty}^{\infty}(\mathcal{F}_1\mathrm{Y})(v,\phi)\left|v\right|e^{i2\pi v(x\cos\phi+y\sin\phi)}\;dv \qquad (13.10)$$

which is the FBP algorithm. It involves four steps:

a) Apply 1-D FT to $\mathrm{Y}(s,\phi)$ to obtain $(\mathcal{F}_1\mathrm{Y})(v,\phi)$;

b) Filter $(\mathcal{F}_1\mathrm{Y})(v,\phi)$ with the so-called ramp filter $|v|$;

c) Apply the 1-D inverse FT to obtain the ramp filtered projections

$$\hat{\mathrm{Y}}(s,\phi)=\int(\mathcal{F}_1\Lambda)(v,\phi)\left|v\right|e^{i2\pi vs}\;dv$$

d) Apply the back-projection operator Eq. (13.2) to $\hat{\mathrm{Y}}(s,\phi)$ to obtain the desired image $\Lambda(x,y)$.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- It should be noted that the ramp filter sets the DC component (i.e. the amplitude of the zero frequency) of the image to zero, while the mean value of the reconstructed image should definitely be positive.
- As a result, straightforward discretization of FBP causes significant negative bias. The problem is reduced with ‘zero padding’ before computing the Fourier transform with fast Fourier transform (FFT). Zero padding involves extending the sinogram rows with zeros at both sides.
- This increases the sampling in the frequency domain and results in a better discrete approximation of the ramp filter. However, a lot of zero padding may be required.

IAEA

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- ☐ Instead of filtering in the Fourier domain, the ramp filtering can also be implemented as a 1-D convolution in the spatial domain. For this, the inverse Fourier transform of $|v|$ is required. This inverse transform actually does not exist, but approximating it as the limit for $\varepsilon \rightarrow 0$ of the well behaved function $|v|e^{-\varepsilon|v|}$ gives:

$$\mathcal{F}^{-1}\left(|v|e^{-\varepsilon|v|}\right) = \frac{\varepsilon^2 - (2\pi s)^2}{(\varepsilon^2 + (2\pi s)^2)^2} \qquad (13.11)$$

$$\approx -\frac{1}{(2\pi s)^2} \quad \text{for} \quad |s| >> \varepsilon \qquad (13.12)$$

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- ❑ In practice, band limited functions are always worked with, implying that the ramp filter has to be truncated at the frequencies $\nu = \pm 1/(2\tau)$, where τ represents the sampling distance. The corresponding convolution kernel h then equals:

$$h(s) = \mathcal{F}^{-1}(|\nu| b(\nu)) = \frac{1}{2\tau^2}\frac{\sin(\pi s/\tau)}{\pi s/\tau} - \frac{1}{4\tau^2}\left(\frac{\sin(\pi s/(2\tau))}{\pi s/(2\tau)}\right)^2 \qquad (13.13)$$

with $b(\nu) = 1$ if $|\nu| \leq 1/(2\tau)$

$= 0$ if $|\nu| > 1/(2\tau)$

- ❑ h is normally only needed for samples $s = n\tau$: $h(n\tau) = 1/(4\tau^2)$ if $n = 0$, $h(n\tau) = 0$ if n is even, and $h(n\tau) = -1/(n\pi\tau)^2$ if n is odd.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- The filter can either be implemented as a convolution or the Fourier transform can be used to obtain a digital version of the ramp filter. Interestingly, this way of computing the ramp filter also reduces the negative bias mentioned above. The reason is that this approach yields a non-zero value for the DC component. When the filtering is done in the frequency domain, some zero padding before FFT is still recommended because of the circular convolution effects, but far less is needed than with straightforward discretization of $|\nu|$.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- Although this is not obvious from the equations above, an algorithm equivalent to FBP is obtained by first back-projecting $\mathrm{Y}(s,\phi)$ and then applying a 2-D ramp filter to the back-projected image $\mathrm{B}(x, y)$:

$$\mathrm{B}(x,y) = \int_0^{\pi} \mathrm{Y}(x\cos\phi + y\sin\phi, \phi)\ d\phi \qquad (13.14)$$

$$(\mathcal{F}_2\Lambda)(\nu_x,\nu_y) = \sqrt{\nu_x^2+\nu_y^2}\,(\mathcal{F}_2\mathrm{B})(\nu_x,\nu_y) \qquad (13.15)$$

- This algorithm is often referred to as the ‘back-project-then-filter’ algorithm.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

- ☐ Filtered back-projection assumes that the projections $Y(s,\phi)$ are line integrals. As discussed in Chapter 11, PET and SPECT data are not line integrals because of attenuation, detector non-uniformities, the contribution of scattered photons and/or random coincidences, etc.
- ☐ It follows that one has to recover (good estimates of) the line integrals by precorrecting the data for these effects. However, a particular problem is posed by the attenuation in SPECT because, different from PET, the attenuation depends on the position along the projection line, precluding straightforward precorrection. Three solutions are described in the next slide.

13.2 ANALYTICAL RECONSTRUCTION

13.2.1 Two dimensional tomography

Solutions:

a) If it can be assumed that the attenuation is constant inside a convex body contour, then filtered back-projection can be modified to correct for the attenuation.

b) If the attenuation is not constant, an approximate correction algorithm proposed by Chang can be applied. It is a post-correction method, applied to the image obtained without any attenuation correction.

c) Use a modified filtered back-projection algorithm, compensating for non-uniform attenuation in SPECT, (Novikov, 2000). An equivalent algorithm was derived by Natterer [1].

[1] NATTERER, F., Inversion of the attenuated Radon transform, Inverse Probl. 17 (2001) 113–119.14

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- ☐ Several very interesting methods in image reconstruction, including Fourier rebinning, are based on the so-called frequency–distance relation [1]. This is an approximate relation between the orthogonal distance to the detector and the direction of the frequency in the sinogram.

[1] XIA, W., LEWITT, R.M., EDHOLM, P.R., Fourier correction for spatially variant collimator blurring in SPECT, IEEE Trans. Med. Imaging 14 (1995) 100–115.

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- Consider the PET acquisition of a point source, as illustrated in Fig 13.3. Usually, the acquisition is described by rotating the projection lines while keeping the object stationary.

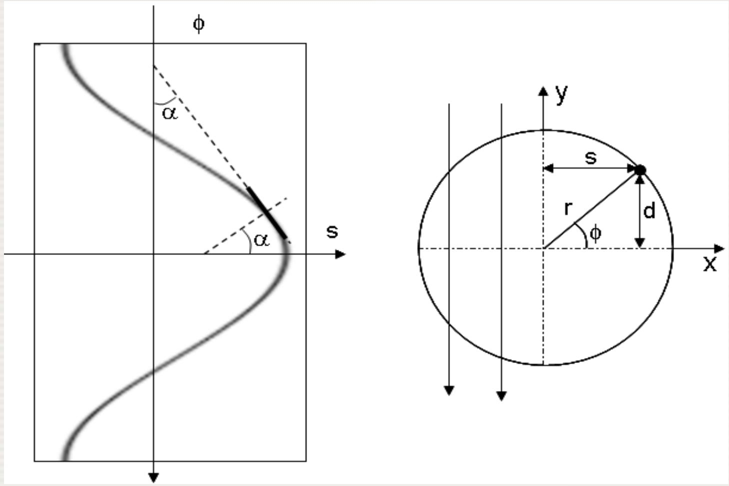


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- ☐ However, here the equivalent description is considered, where projection is always along the y axis, and tomographic acquisition is obtained by rotating the object around the origin.

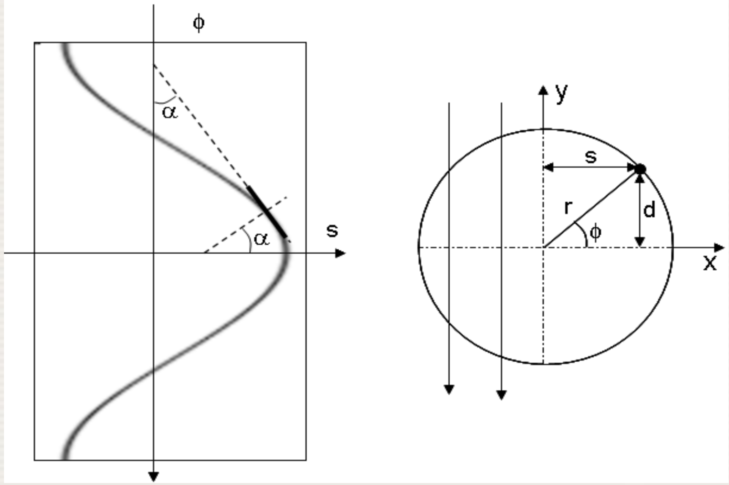


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- ❑ Suppose that the point is located on the *x* axis when $\phi = 0$. When acquiring the parallel projections for angle ϕ, the point has polar coordinates (r, ϕ), with r the distance from the centre of the field of view (FOV) and ϕ the angle with the *x* axis. The distance to the x axis is $d = r \sin \phi$.

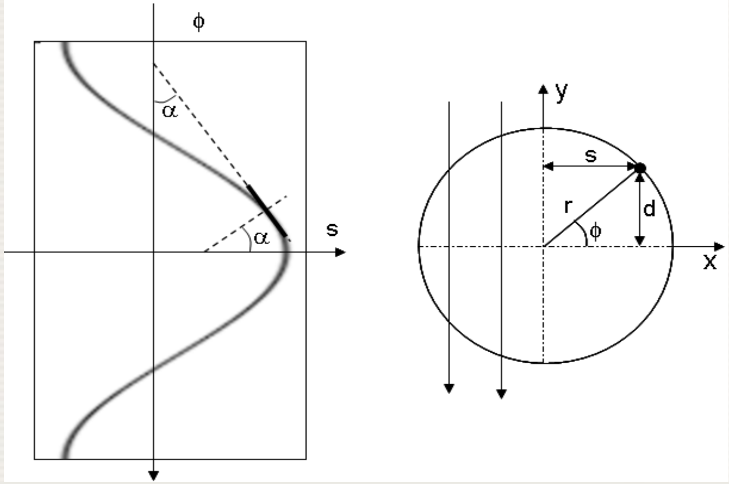


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- The corresponding sinogram $Y(s, \phi)$ is zero everywhere, except on the curve $s = r \cos \phi$. The complete sinogram is obtained by rotating the point over 360° $\phi = -\pi \ldots \pi$.
- Consider a small portion of this curve, which can be well approximated as a tangential line segment near a particular point (s, ϕ).

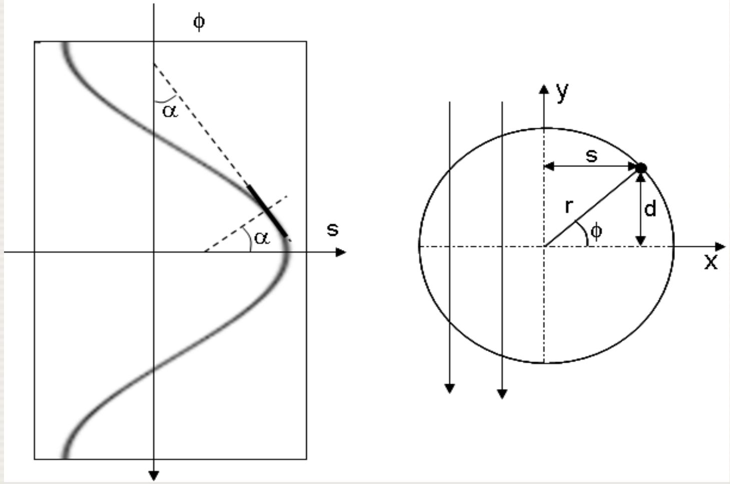


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- ☐ In the 2-D Fourier transform of the sinogram, this line segment contributes mostly frequencies in the direction orthogonal to the line segment, represented by the angle α, given by:

$$\tan\alpha = \frac{\partial}{\partial\phi}(r\cos\phi) = -r\sin\phi = -d. \qquad (13.16)$$

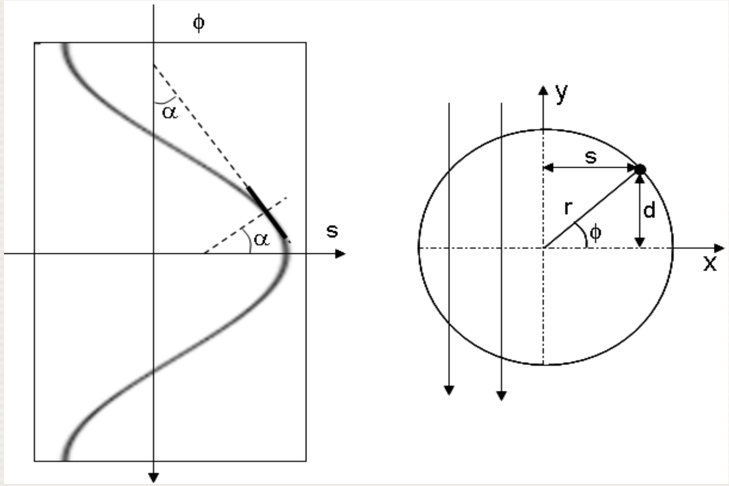


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- ☐ Thus, in the 2-D Fourier transform , the value at a particular point (v_s,v_ϕ) carries mostly information about points located at a distance $d = -\tan\alpha = -v_\phi/v_s$ from the line through the centre, parallel to the detector.

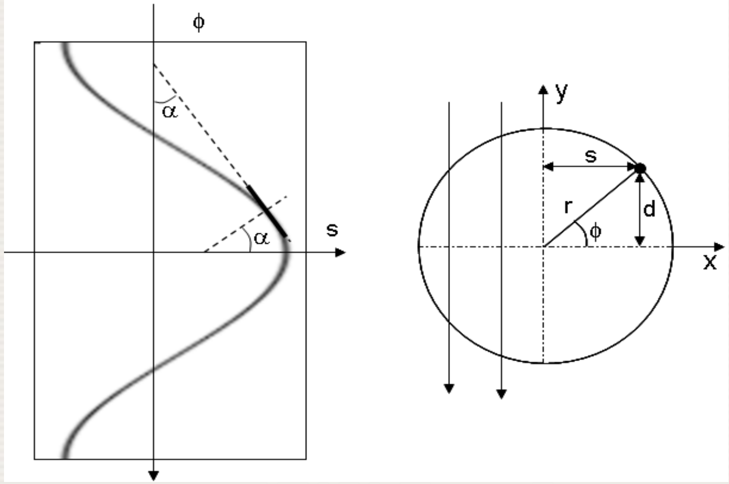


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.2 Frequency-distance relation

- This relation can be exploited to apply distance dependent operations to the sinogram. One example is distance dependent deconvolution, to compensate for the distance dependent blurring in SPECT. Another is Fourier rebinning, where data from oblique sinograms are rebinned into direct sinograms.

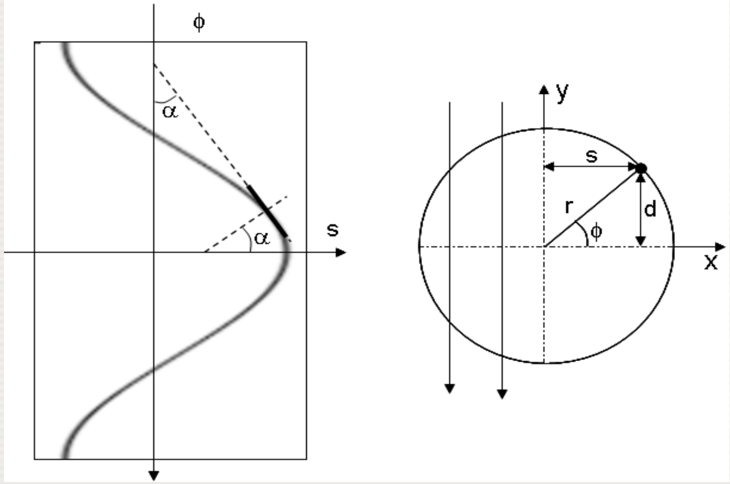


FIG. 13.3. The frequency–distance principle. Left: sinogram; right: vertical projection of a point located at polar coordinates (r,ϕ).

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

Filtered back-projection

- In PET each pair of detectors in coincidence defines a single LOR. In this section, the discrete nature of the detection is ignored, since the analytical approach is more easily described assuming continuous data. Consider the X ray transform in 3-D, which can be written as:

  $$Y(\hat{\boldsymbol{u}},\boldsymbol{s}) = \int_{-\infty}^{\infty} \Lambda(\boldsymbol{s} + t\hat{\boldsymbol{u}})\,dt \qquad (13.17)$$

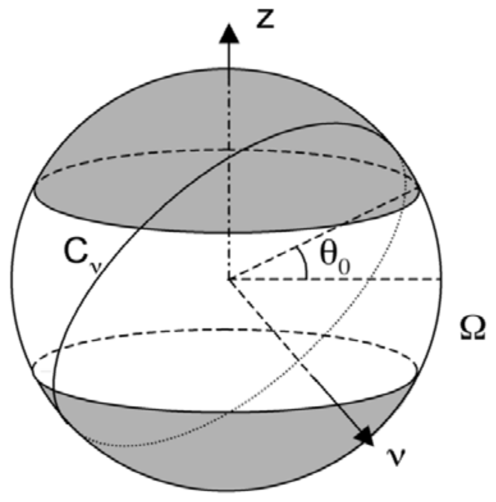
  where the LOR is defined as the line parallel to $\hat{\boldsymbol{u}}$ and through the point $\boldsymbol{s}$. The vector $\hat{\boldsymbol{u}}$ is a unit vector, and the vector $\mathbf{s}$ is restricted to the plane orthogonal to $\hat{\boldsymbol{u}}$, hence $(\hat{\boldsymbol{u}},\boldsymbol{s})$ is 4-D.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- Most PET systems have cylindrical symmetry. For this reason, the inversion of Eq. (13.17) is studied for the case where $\hat{\mathbf{u}}$ is restricted to the band Ω_{θ_0} on the unit sphere, defined by $|u_z| \leq \sin\theta_0$, as illustrated in Fig. 13.4. Only half of the sphere is actually needed because $Y(\hat{\mathbf{u}},\mathbf{s}) = Y(-\hat{\mathbf{u}},\mathbf{s})$, but working with the complete sphere is more convenient.

FIG. 13.4. Each point on the unit sphere corresponds to the direction of a parallel projection. An ideal rotating gamma camera with a parallel hole collimator only travels through the points on the equator. An idealized 3-D PET system would also acquire projections along oblique lines; it collects projections for all points of the set Ω. The set Ω, defined by θ_0, is the non-shaded portion of the unit sphere. To recover a particular frequency **ν** (of the Fourier transform of the object), at least one point on the circle C_ν is required.



13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- ☐ With $\theta_0 = 0$, the problem reduces to 2-D parallel projection (for multiple slices), which was shown to have a unique solution. It follows that with $|\theta_0| > 0$, the problem becomes overdetermined, and there are infinitely many ways to compute the solution. This can be seen as follows.

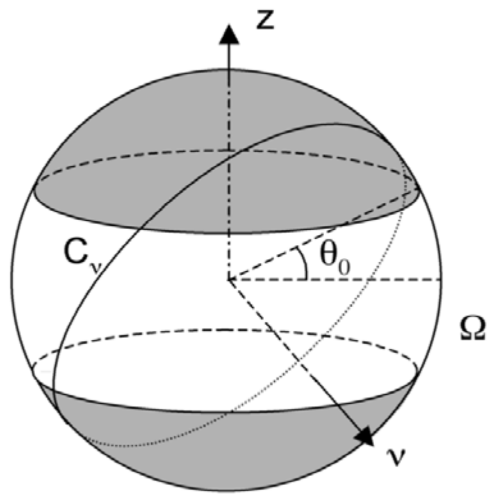


FIG. 13.4. Each point on the unit sphere corresponds to the direction of a parallel projection. An ideal rotating gamma camera with a parallel hole collimator only travels through the points on the equator. An idealized 3-D PET system would also acquire projections along oblique lines; it collects projections for all points of the set Ω. The set Ω, defined by θ_0, is the non-shaded portion of the unit sphere. To recover a particular frequency $\mathbf{\nu}$ (of the Fourier transform of the object), at least one point on the circle C_ν is required.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- Each point of Ω corresponds to a parallel projection. From the central slice theorem, this provides a central plane perpendicular to of the 3-D Fourier transform $\mathcal{L}(\mathbf{v})$ *of* $\Lambda(\mathbf{x})$. The set Ω_0 (i.e. all points on the equator of the unit sphere), provides all planes intersecting the v_z axis, sufficient to recover the entire image $\Lambda(\mathbf{x})$ via inverse FFT.

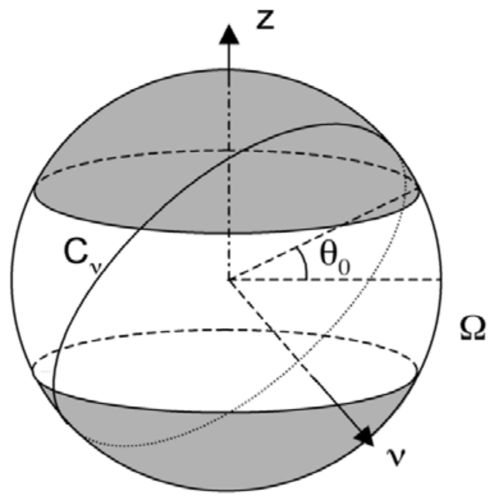


FIG. 13.4. Each point on the unit sphere corresponds to the direction of a parallel projection. An ideal rotating gamma camera with a parallel hole collimator only travels through the points on the equator. An idealized 3-D PET system would also acquire projections along oblique lines; it collects projections for all points of the set Ω. The set Ω, defined by θ_0, is the non-shaded portion of the unit sphere. To recover a particular frequency $\mathbf{v}$ (of the Fourier transform of the object), at least one point on the circle C_v is required.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- The set Ω_0 with $\theta_0 > 0$ provides additional (oblique) planes through $\mathcal{L}(\boldsymbol{v})$, which are obviously redundant. A simple solution would be to select a sufficient subset from the data. However, if the data are noisy, a more stable solution is obtained by using all of the measurements. This is achieved by computing $\mathcal{L}(\boldsymbol{v})$ from a linear combination of all available planes:

$$\mathcal{L}(\boldsymbol{v}) = \int_{\Omega_{\theta_0}} \mathcal{Y}(\hat{\boldsymbol{u}},\boldsymbol{v}) \mathrm{H}(\hat{\boldsymbol{u}},\boldsymbol{v}) \delta(\hat{\boldsymbol{u}},\boldsymbol{v}) \mathrm{d}\hat{\boldsymbol{u}} \qquad (13.18)$$

where $\mathcal{Y}(\hat{\boldsymbol{u}},\boldsymbol{v})$ is the 2-D Fourier transform with respect to $\boldsymbol{s}$ of the projection $\mathrm{Y}(\hat{\boldsymbol{u}},\boldsymbol{s})$. The Dirac function $\delta(\hat{\boldsymbol{u}},\boldsymbol{v})$ selects the parallel projections which are perpendicular to $\boldsymbol{v}$ (i.e. the points on the circle C_v in Fig. 13.4).

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- Finally, the filter $H(\hat{\mathbf{u}},\boldsymbol{v})$ assigns a particular weight to each of the available datasets $y(\hat{\mathbf{u}},\boldsymbol{v})$. The combined weight for each frequency should equal unity, leading to the filter equation:

$$\int_{\Omega_{\theta_0}} \mathrm{H}(\hat{\mathbf{u}},\boldsymbol{v})\delta(\hat{\mathbf{u}},\boldsymbol{v})d\hat{\mathbf{u}} = 1 \qquad (13.19)$$

- A solution equivalent to that of unweighted least squares (LS) is obtained by assigning the same weight to all available data. This results in the Colsher filter:

$$\begin{aligned} \mathrm{H}_{\mathrm{C}}(\hat{\mathbf{u}},\boldsymbol{v}) &= |\boldsymbol{v}|/(2\pi) && \text{if } \sin\psi \le \sin\theta_0 \\ &= |\boldsymbol{v}|/(4 \arcsin(\sin\theta_0/\sin\psi)) && \text{if } \sin\psi > \sin\theta_0 \end{aligned} \qquad (13.20)$$

where ψ is the angle between $\boldsymbol{v}$ and the z axis: $v_z/|\boldsymbol{v}| = \cos\psi$.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- The direct Fourier reconstruction method can also be applied here, by straightforward inverse Fourier transform of Eq. (13.18). However, a filtered back-projection approach is usually preferred, which can be written as:

$$\Lambda(\boldsymbol{x}) = \int_{\Omega_{\theta_0}} d\hat{\boldsymbol{u}}\,\Upsilon^F(\hat{\boldsymbol{u}}, \boldsymbol{x} - (\boldsymbol{x}\cdot\hat{\boldsymbol{u}})\hat{\boldsymbol{u}}) \qquad (13.21)$$

- Here, Υ^F is obtained by filtering Υ with the Colsher filter (or another filter satisfying Eq. (13.19):

$\Upsilon^F(\hat{\boldsymbol{u}},\boldsymbol{s}) = \mathcal{F}^{-1}(H_C(\hat{\boldsymbol{u}},v)\,y(\hat{\boldsymbol{u}},v))$. The coordinate $\boldsymbol{s} = \boldsymbol{x}-(\boldsymbol{x}\cdot\hat{\boldsymbol{u}})\hat{\boldsymbol{u}}$ is the projection of the point $\boldsymbol{x}$ on the plane perpendicular to $\hat{\boldsymbol{u}}$; it selects the LOR through $\boldsymbol{x}$ in the parallel projection $\hat{\boldsymbol{u}}$.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

The reprojection algorithm

- The previous analysis assumed that the acceptance angle θ_0 was a constant, independent of $\mathbf{x}$. As illustrated in Fig. 13.5, this is not the case in practice.

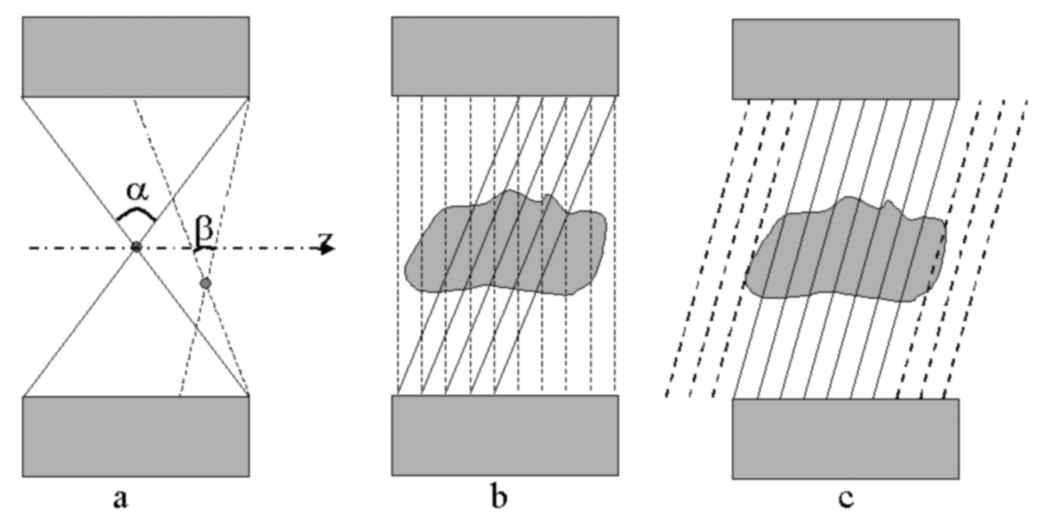


FIG. 13.5. An axial cross-section through a cylindrical PET system, illustrating that the acceptance angle is position dependent (a). The projections are complete for orthogonal to the z axis (these are the 2-D multislice parallel beam projections) and are truncated for the oblique parallel projections (b). The truncation becomes more severe for more oblique projections. In the reprojection algorithm, the missing oblique projections (dashed lines) are computed from a temporary multislice 2-D reconstruction (c)

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- ☐ As the acceptance angle is position dependent, the required filtering is position dependent as well, and cannot be implemented as a shift-invariant convolution (or Fourier filter). Several strategies for dealing with this truncation have been developed.
- ☐ In the reprojection algorithm we start with a first reconstruction, using the smallest acceptable angle over all positions x in the FOV. This usually means that only the parallel projections orthogonal to the z axis are used. The missing oblique projection values are computed from this first reconstruction (Fig. 13.4) and used to complete the measured oblique projections.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

Rebinning techniques

- ❑ The complexity (estimated as the number of LORs) increases linearly with the axial extent for 2-D PET, but quadratically for 3-D PET. To keep the processing time acceptable, researchers have sought ways to reduce the size of the data as much as possible, while minimizing the loss of information induced by this reduction.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- Most PET systems have a cylindrical detector surface: the detectors are located on rings with radius R, and the rings are combined in a cylinder along the z axis. The data are usually organized in sinograms which can be written as:

$$Y_{\mathrm{P}}(s,\phi,z,\Delta_z)=\int_{-\infty}^{\infty}\mathrm{d}t\,\Lambda(s\cos\phi+t\hat{u}_x,s\sin\phi+t\hat{u}_y,z+t\hat{u}_z), \qquad (13.22)$$

where $\hat{\boldsymbol{u}}$ is a unit vector in the direction of the LOR:

$\hat{\boldsymbol{u}}=\boldsymbol{u}/||\boldsymbol{u}||$ with $\boldsymbol{u}=(-\sin\phi,\cos\varphi,\Delta_z/(2\sqrt{R^2-s^2}))$

where s is the distance between the LOR and the z axis. The LOR corresponds to detectors with axial positions $\mathrm{z}-\Delta_z/2$ and $\mathrm{z}+\Delta_z/2$,and ϕ is the angle between the y axis and the projection of the LOR on the xy plane.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- In practice, $s << R$ and, as a result, the direction of the LOR, the vector $\hat{\boldsymbol{u}}$, is virtually independent of s. In other words, a set of LORs with fixed Δ_z can then be treated as a parallel projection with good approximation. LORs with $\Delta_z = 0$ are often called ‘direct’ LORs, while LORs with $\Delta_z \neq 0$ are called ‘oblique’.
- The basic idea of rebinning algorithms is to compute estimates of the direct sinograms from the oblique sinograms. If the rebinning algorithm is good, most of the information from the oblique sinograms will go into these estimates. As a result, the data have been reduced from a complex 3-D geometry into a much simpler 2-D geometry without discarding measured signal.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- The final reconstruction can then be done with 2-D algorithms, which tend to be much faster than fully 3-D algorithms. A popular approach is to use Fourier rebinning, followed by maximum likelihood reconstruction.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

Single slice and multislice rebinning

- The simplest way to rebin the data is to treat oblique LORs as direct LORs, giving the approximation:

$$Y_P(s,\phi,z,\Delta_z) \approx Y_P(s,\phi,z,0) \qquad (13.23)$$

- The approximation is only exact if the object consists of points located on the z axis, and it introduces mis-positioning errors that increase with increasing distance to the z axis and increasing Δ_z. Consequently, *single slice rebinning* (SSRB) is applicable when the object is small and positioned centrally in the scanner or when Δ_z is small. The axial extent of most current PET systems is too large to rebin an entire 3-D dataset with Eq. (13.23).

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- However, SSRB is used on all PET systems to reduce the sampling of the Δ_z dimension in the 3-D data, by combining sinograms with similar Δ_z. This typically reduces the data size with a factor of about ten, when compared to the finest possible sampling.
- Application of Eq. (13.23) obviously causes blurring in the z direction, to a degree proportional to the distance from the z axis. However, it may also cause severe inconsistencies in the sinograms, producing blurring artefacts in the xy planes of the reconstructed images as well.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- ☐ Lewitt et al. [1] proposed distributing the oblique LOR values $Y_P(s,\phi,z,\Delta_z)$ over all LORs with $z \in [z-\Delta_z R_f / (2R), z+\Delta_z R_f / (2R)]$, i.e. over all slices intersected by the LOR, and within an FOV with radius R_f.
- ☐ This so-called *multislice rebinning* (MSRB) reduces the inconsistencies in the sinograms, eliminating most of the xy blurring artefacts in the reconstruction, but the improvement comes at the cost of strong axial blurring. This blurring depends strongly on z, and is approximately independent of x and y. A z-dependent 1-D axial filter is applied to reduce this axial blurring. MSRB is superior to SSRB, but the noise characteristics are not optimal.

[1] LEWITT, R.M., MUEHLLEHNER, G., KARP, J.S., Three-dimensional image reconstruction for PET by multi-slice rebinning and axial image filtering, Phys. Med. Biol. 39 (1994) 321–339.

13.2 ANALYTICAL RECONSTRUCTION
13.2.3 Fully 3-D tomography

Fourier rebinning

- ☐ Fourier rebinning is based on the frequency–distance principle, which was explained above. The Fourier rebinning method is most simply formulated when the projection is written as follows:

$$Y(s,\phi,z,\delta)=\int_{-\infty}^{\infty} dt\Lambda(s\cos\phi-t\sin\phi, s\sin\phi+t\cos\phi, z+t\delta) \qquad (13.24)$$

where $\delta = \tan\theta$, θ is the angle between the LOR and the xy plane, and the integration variable t is the distance between the position on the LOR and the z axis.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- It follows that:

$$Y(s,\phi,z,\delta)=\frac{Y_P\left(s,\phi,z,\Lambda_z=2\delta\sqrt{R^2-s^2}\right)}{\sqrt{1+\delta^2}} \qquad (13.25)$$

$$\approx\frac{Y_P\left(s,\phi,z,\Lambda_z=2\delta R\right)}{\sqrt{1+\delta^2}} \qquad (13.26)$$

where the approximation is valid wherever $s << R$. In this case, no interpolation is needed; it is sufficient to scale the PET data with the weight factor $\sqrt{1+\delta^2}$.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- ☐ Fourier rebinning uses the frequency–distance principle to find the distance d corresponding to a particular portion of the oblique sinogram. As illustrated in Fig. 13.6, distance is used to locate the direct sinogram to which this portion should be assigned.
- ☐ Denoting the 2-D Fourier transform of Y with respect to s and φ as $\mathcal{Y}$:

$$\mathcal{Y}(v_s, v_\phi, z, \delta) \approx \mathcal{Y}(v_s, v_\phi, z - \delta(v_\phi / v_s), 0) \quad (13.27)$$

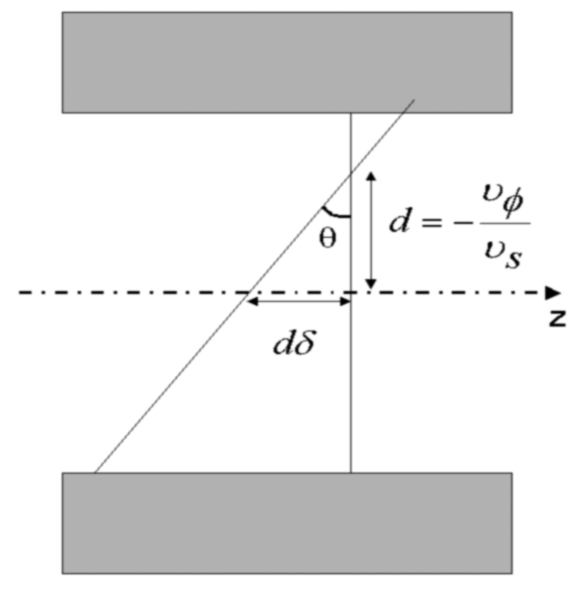


FIG. 13.6. Fourier rebinning: the distance from the rotation axis is obtained via the frequency–distance principle. This distance is used to identify the appropriate direct sinogram.

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- Eq. (13.27) explains how to distribute frequency components from a particular oblique sinogram into different direct sinograms. Frequencies located on the line $v_\phi = v_s$ in the oblique sinogram z can be assigned to that same line in the direct sinogram $z+\delta d$.
- The final rebinning algorithm (often called ‘FORE’) is obtained by averaging all of the available estimates of the direct sinogram:

$$
\begin{aligned}
\mathcal{Y}(\nu_{\mathrm{s}},\nu_\phi,\mathrm{z},0) &\approx \frac{1}{\delta_{\max}}\int_0^{\delta_{\max}} \mathrm{d}\delta\, \mathcal{Y}\!\left(\nu_{\mathrm{s}},\nu_\phi,\mathrm{z}+\delta\frac{\nu_\phi}{\nu_{\mathrm{s}}},\delta\right) \text{ if } |\nu_{\mathrm{s}}| \gg 0 \\
&\approx \mathcal{Y}(0,0,z,0) \ \text{ if } \ \nu_s \approx 0,\ \nu_\phi \approx 0 \qquad (13.28) \\
&\approx 0 \quad \text{ if } \ |\nu_\phi / \nu_s| > R_f
\end{aligned}
$$

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

- ☐ It should be noted that the rebinning expression is only valid for large v_s. In the low frequency range, only the direct sinogram is used. The last line of Eq. (13.28) holds because the image $\Lambda(x,y,z)$ is assumed to be zero outside the FOV $\sqrt{x^2+y^2} > R_f$.
- ☐ A more rigorous mathematical derivation of the frequency–distance relation is given in [1]. Alternative derivations based on exact rebinning expressions are in [2].
- ☐ After Fourier rebinning, the resulting 2-D dataset can be reconstructed with any 2-D reconstruction algorithm. A popular method is the combination of Fourier rebinning with a 2-D statistical reconstruction algorithm.

[1] DEFRISE, M., A factorization method for the 3D X-ray transform, Inverse Probl. 11 (1995) 983–994

[2] DEFRISE, M., et al., Exact and approximate rebinning algorithms for 3D PET data, IEEE Trans. Med. Imaging 16 (1997) 145–158

13.2 ANALYTICAL RECONSTRUCTION

13.2.3 Fully 3-D tomography

Exact rebinning methods

- ☐ Fourier rebinning is an approximate method, but was found to be sufficiently accurate for apertures up to $\theta_0 =$ 25°, and it is, therefore, largely sufficient for most current PET systems. However, there is a tendency towards still larger acceptance angles, and a more exact Fourier rebinning algorithm may be needed in the future. An example of an 'exact' rebinning algorithm is FOREX.
- ☐ FOREX is slower than FORE, but still considerably faster than 3-D filtered back-projection with reprojection.

13.2 ANALYTICAL RECONSTRUCTION

13.2.4 Time of flight positron emission tomography

- In time of flight (TOF) PET, the difference in arrival time of the two detected photons is used to estimate the position of their emission along the LOR. TOF projections correspond to Gaussian convolutions along lines, rather than to line integrals, as illustrated in Fig. 13.7.

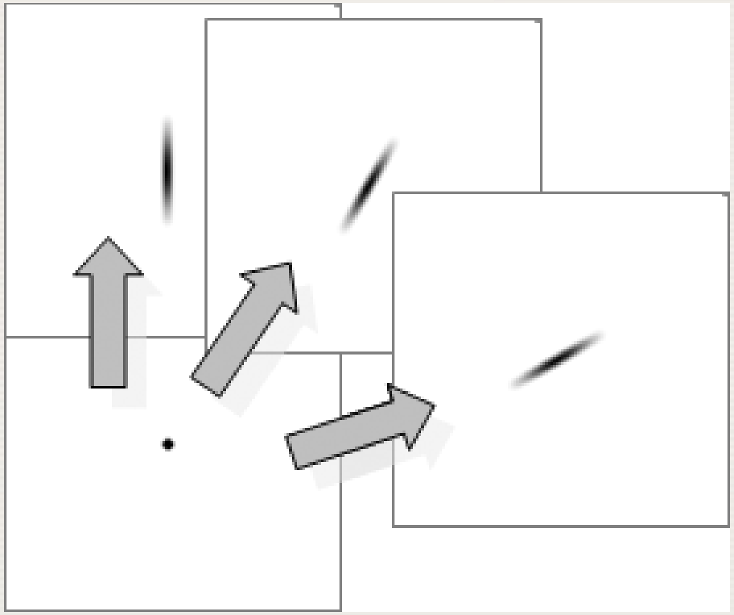


FIG. 13.7. Time of flight projection can be well modelled as a 1-D Gaussian convolution in the direction of the line of response.

13.2 ANALYTICAL RECONSTRUCTION

13.2.4 Time of flight positron emission tomography

- The corresponding TOF back-projection corresponds to convolving the measured data with the same 1-D Gaussians, followed by summation over all angles. If σ_{TOF} is the standard deviation of the TOF-blurring kernel, then TOF projection followed by TOF back-projection corresponds to convolution with the blurring kernel:

$$B_{\mathrm{TOF}}(x,y) = \frac{\mathrm{Gauss}\,(x, y, \sqrt{2}\sigma_{\mathrm{TOF}})}{\sqrt{x^2+y^2}} \qquad (13.34)$$

$$= \frac{1}{\sqrt{x^2+y^2}} \frac{1}{2\sqrt{\pi}\sigma_{\mathrm{TOF}}} \exp\left(-\frac{x^2+y^2}{4\sigma_{\mathrm{TOF}}^2}\right) \qquad (13.35)$$

- Note: Gaussian blurring is present in the projection and in the back-projection, hence s.d. of $\sqrt{2}\sigma_{\mathrm{TOF}}$ in Eq. (13.34).

13.2 ANALYTICAL RECONSTRUCTION

13.2.4 Time of flight positron emission tomography

- The filter required in TOF PET filtered back-projection is derived by inverting the Fourier transform of B_{TOF}, and equals:

$$\mathrm{TOF\_recon\_filter}(\nu) = \frac{1}{\exp(-2\pi^2\sigma^2{}_{\mathrm{TOF}}\nu^2) I_0(2\pi^2\sigma^2{}_{\mathrm{TOF}}\nu^2)} \qquad (13.36)$$

  where I_0 is the zero order modified Bessel function of the first kind.

- This FBP expression is obtained by using the ‘natural’ TOF back-projection, defined as the adjoint of the TOF projection. It also appears in LS approaches, and it has been shown that with this back-projection definition, FBP is optimal in an (unweighted) LS sense.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

Discretization

- In analytical reconstruction, it is initially assumed that the unknown object can be represented as a function, and that the acquired data can also be represented as a function. The reconstruction algorithm is then derived by mathematical inversion, and finally the resulting algorithm is discretized to make it ready for software implementation.
- In iterative reconstruction, one usually starts by discretizing the problem. This reduces the reconstruction problem to finding a finite set of unknown values from a finite set of equations, a problem which can be solved with numerical inversion.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- The advantage of numerical inversion is that only a model for the acquisition process is needed, not for its inverse. That makes it easier (although it may still be non-trivial) to take into account some of the undesired but unavoidable effects that complicate the acquisition, such as photon attenuation, position dependent resolution, gaps between the detectors, patient motion, etc.
- After discretization, the unknown image values and the known measured values can be represented as column vectors $\boldsymbol{\lambda}$ and $\boldsymbol{y}$.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- The PET or SPECT acquisition process is characterized by the system matrix $\boldsymbol{A}$ and an additive contribution $\overline{\boldsymbol{b}}$, and $\boldsymbol{n}$ is the measurement noise:

$$\boldsymbol{y} = \boldsymbol{A\lambda} + \overline{\boldsymbol{b}} + \boldsymbol{n} \quad or \quad y_i = \sum_{j=1}^{J} A_{ij}\lambda_j + \overline{b}_i + n_i, \quad i = 1,..,I \qquad (13.37)$$

where y_i denotes the number of photons measured at LOR i, where the index i runs over all of the sinogram elements. The index j runs over all of the image voxels, and $\boldsymbol{A}_{ij}$ is the probability that a unit of radioactivity in j gives rise to the detection of a photon (SPECT) or photon pair (PET) in LOR i. Image reconstruction now consists of finding $\boldsymbol{\lambda}$, given $\boldsymbol{A}$, $\boldsymbol{y}$ and $\overline{\boldsymbol{b}}$, and a statistical model for $\boldsymbol{n}$.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

Objective functions

- The presence of the noise precludes exact reconstruction. For this reason, the reconstruction is often treated as an optimization task: it is assumed that a useful clinical image can be obtained by maximizing a well chosen objective function. When the statistics of the noise are known, a Bayesian approach can be applied, searching for the image that maximizes the conditional probability on the data.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- Bayes' rule is written:

$$\hat{\boldsymbol{\lambda}} = \underset{\lambda}{\operatorname{argmax}} \frac{p(\boldsymbol{y} \,|\, \boldsymbol{\lambda})\, p(\boldsymbol{\lambda})}{p(\boldsymbol{y})} = \underset{\lambda}{\operatorname{argmax}} (\ln p(\boldsymbol{y} \,|\, \boldsymbol{\lambda}) + \ln p(\boldsymbol{\lambda})) \qquad (13.38/39)$$

- The probability $p(\boldsymbol{y}|\lambda)$ gives the likelihood of measuring a particular sinogram $\boldsymbol{y}$, when the tracer distribution equals λ. This distribution is often simply called the likelihood. The probability $p(\lambda)$ represents the a priori knowledge about the tracer distribution, available before PET or SPECT acquisition. This probability is often called the *prior distribution*. The knowledge available after the measurements equals $p(\boldsymbol{y}|\lambda)p(\lambda)$ and is called the *posterior distribution*.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- To keep things simple, it is often assumed that no prior information is available, i.e. $p(\boldsymbol{\lambda}|\mathbf{y}) \sim p(\mathbf{y}|\boldsymbol{\lambda})$. Finding the solution then reduces to maximizing the likelihood $p(\mathbf{y}|\boldsymbol{\lambda})$ (or its logarithm). In this section, maximum likelihood algorithms are discussed.
- Maximum a posteriori algorithms are discussed in Section 13.3.5, as a strategy to suppress noise propagation.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- A popular approach to solve equations of the form of Eq. (13.37) is LS estimation. This is equivalent to a maximum likelihood approach, if it is assumed that the noise is Gaussian with a zero mean and a fixed, position independent standard deviation σ. The probability to measure the noisy value $\boldsymbol{y}_i$ when the expected value was $\sum_j A_{ij}\lambda_j + \bar{b}_i$ then equals:

$$p_{\mathrm{LS}}\left(y_i \mid \sum_j A_{ij}\lambda_j + \bar{b}_i \right) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\boldsymbol{y}_i - (\sum_j A_{ij}\lambda_j + \bar{b}_i))^2}{2\sigma^2} \right) \qquad (13.40)$$

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- As the noise in the sinogram is not correlated, the likelihood (i.e. the probability of measuring the entire noisy sinogram $\mathbf{y}$) equals:

$$p_{\mathrm{LS}}(\boldsymbol{y}\,|\,\boldsymbol{\lambda}) = p_{\mathrm{LS}}(\boldsymbol{y}\,|\,\boldsymbol{A\lambda}+\bar{\boldsymbol{b}}) = \prod_i p_{\mathrm{LS}}(y_i\,|\sum_j A_{ij}\lambda_j + \bar{b}_i). \qquad (13.41)$$

- It is more convenient to maximize the logarithm of p_{LS}; dropping constants the objective function L_{LS} is:

$$L_{\mathrm{LS}} = -\sum_i \left(y_i - \left(\sum_j A_{ij}\lambda_j + \bar{b}_i\right)\right)^2 = -(\boldsymbol{y}-(\boldsymbol{A\lambda}+\bar{\boldsymbol{b}}))'(\boldsymbol{y}-(\boldsymbol{A\lambda}+\bar{\boldsymbol{b}})) \qquad (13.42)$$

- where the prime denotes matrix transpose. Setting the first derivatives w.r.t. λ_j to zero for all j, one obtains:

$$\boldsymbol{A}'(\boldsymbol{y}-\boldsymbol{A\lambda}-\bar{\boldsymbol{b}}) = 0 \quad \text{and} \quad \boldsymbol{\lambda} = (\mathbf{A'A})^{-1}(\boldsymbol{y}-\bar{\boldsymbol{b}}) \qquad (13.43)$$

provided that $\boldsymbol{A'A}$ is non-singular.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- In Eq. (13.43) the operator $\mathbf{A}$ is the discrete projection; its transpose $\mathbf{A}'$ is the discrete back-projection. Its analytical counterpart was given in Eq. (13.2) and illustrated in Fig. 13.2. Fig. 13.8 shows $\mathbf{A}$ and $\mathbf{A}'\mathbf{A}$ on an image of 3 point sources, using ideal parallel beam projection.

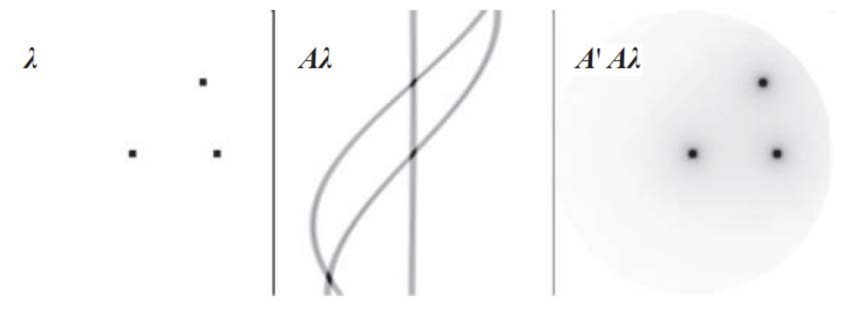


FIG. 13.8. The image of point sources is projected and back-projected again along ideal parallel beams. This yields a shift-invariant blurring.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- The figure shows the resulting point spread functions of $A'A$ for each of the point sources. They are identical: for ideal parallel beam projection, $A'A$ is shift-invariant, equivalent to a convolution. It follows that $(A'A)^{-1}$ is the corresponding shift-invariant deconvolution, which is easily computed via the Fourier transform.

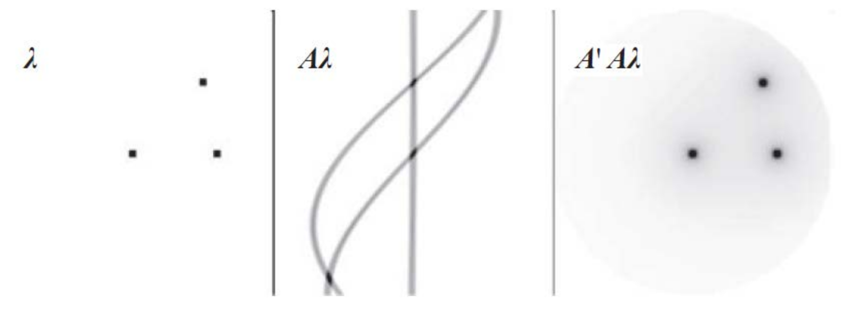


FIG. 13.8. The image of point sources is projected and back-projected again along ideal parallel beams. This yields a shift-invariant blurring.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- In this situation, LS reconstruction (Eq. (13.43)) is the discrete equivalent of the back-project-then-filter algorithm (Eq. (13.15)), applied to the data after precorrection for $\bar{b}$.

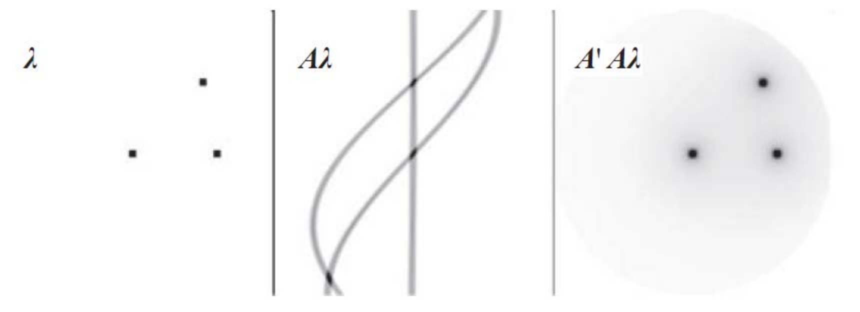


FIG. 13.8. The image of point sources is projected and back-projected again along ideal parallel beams. This yields a shift-invariant blurring.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- Figure 13.9 illustrates $\boldsymbol{A}$ and $\boldsymbol{A'A}$ for a projector that models the position dependent blurring of a typical parallel beam SPECT collimator. The blurring induced by $\boldsymbol{A'A}$ is now shift-variant — it cannot be modelled as a convolution and its inverse cannot be computed with the Fourier transform.
- For real life problems, direct inversion of $\boldsymbol{A'A}$ is not feasible. Instead iterative optimization is applied to find the maximum of Eq. (13.42).

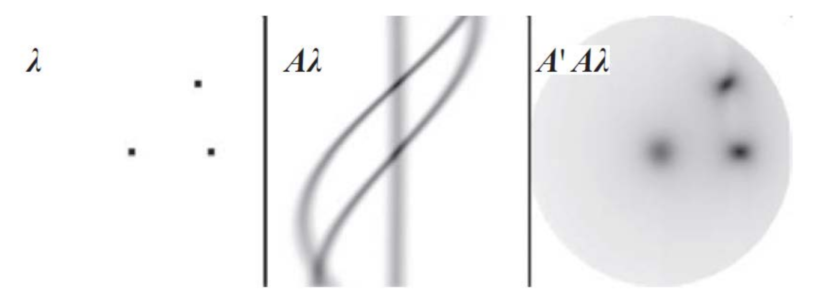


FIG. 13.9. The image of point sources is projected and back-projected again with collimator blurring. This yields a shift-variant blurring.

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- Proceeding as before, and replacing σ with σ_i in Eq. (13.40), the weighted least squares objective function is:

$$L_{\mathrm{WLS}} = -\sum_i \frac{(y_i - (\sum_j A_{ij}\lambda_j + \bar{b}_i))^2}{\sigma_i^{\,2}} \qquad (13.44)$$

$$= -(\boldsymbol{y} - (\boldsymbol{A}\boldsymbol{\lambda} + \bar{\boldsymbol{b}}))'\boldsymbol{C}_y^{\,-1}(\boldsymbol{y} - (\boldsymbol{A}\boldsymbol{\lambda} + \bar{\boldsymbol{b}}))$$

where $\boldsymbol{C}y$ is the covariance matrix of the data. For emission tomography it is a diagonal matrix with elements $\boldsymbol{C}y[\mathrm{i,i}]=\sigma_i^2$. The corresponding WLS reconstruction can be written

$$\boldsymbol{\lambda} = (\boldsymbol{A}'\boldsymbol{C}_y^{\,-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{C}_y^{\,-1}(\boldsymbol{y} - \bar{\boldsymbol{b}}) \qquad (13.45)$$

13.3 ITERATIVE RECONSTRUCTION

13.3.1 Introduction

- The operator $A'C_y^{-1}A$ is always shift-variant, even for parallel beam tomography (Fig. 13.10).

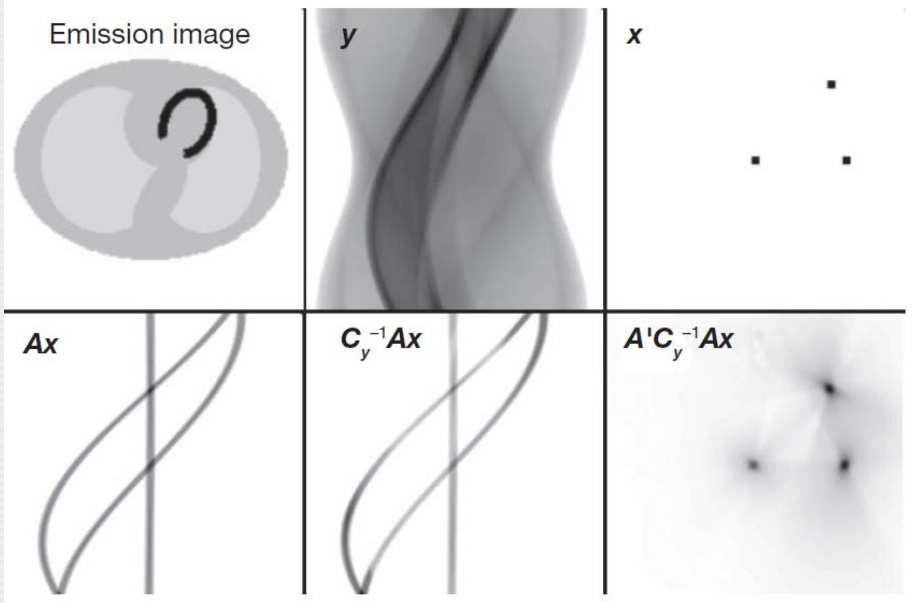


FIG 13.10. The noise-free sinogram $\bar{y}$ is computed for a particular activity distribution. Setting $C_y = \mathrm{diag}(\bar{y})$, the operator $A'C_y^{-1}A$ can be analysed by applying it to the image of a few point sources, x. The image x is projected, the sinogram Ax is divided by $\bar{y}$ on a pixel basis and the result is back-projected. Clearly, position dependent blurring is obtained. Consequently, iterative optimization must be used for WLS reconstruction.

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

- Many iterative reconstruction algorithms have been proposed to minimize the objective functions L_{WLS} and L_{ML}.
- Here, only two approaches are briefly described: preconditioned conjugate gradient methods and optimization transfer, with expectation maximization as a special case of the latter.

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

Preconditioned gradient methods (see Handbook Sec. 13.3.2 for further details)

- The objective function will be optimized when its 1st derivatives are zero:

$$\hat{y}_i = \sum_i A_{ij} \lambda_j + \overline{b_i} \qquad (13.48)$$

$$\frac{\partial L_{\mathrm{WLS}}(\boldsymbol{\lambda})}{\partial \lambda_j} = \sum_i A_{ij} \frac{y_i - \hat{y}_i}{\sigma_i{}^2} \qquad (13.49)$$

$$\frac{\partial L_{\mathrm{ML}}(\boldsymbol{\lambda})}{\partial \lambda_j} = \sum_i A_{ij} \frac{y_i - \hat{y}_i}{\hat{y}_i} \qquad (13.50)$$

- This can be optimized by a steepest ascent method, as

$$\mathbf{d}^k = \nabla L(\boldsymbol{\lambda}^{k-1});\ \boldsymbol{\lambda}^k = \boldsymbol{\lambda}^{k-1} + \alpha_k \mathbf{d}^k;\ \alpha_k = \arg\max_{\alpha} L(\boldsymbol{\lambda}^{k-1} + \alpha \mathbf{d}^k) \quad (13.51)$$

where k and $k\text{-}1$ denote iteration numbers and ∇L is the vector of the 1st derivatives of L w.r.t. λ_j.

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

Conjugate gradient methods

- Figure 13.11 shows the convergence of the steepest gradient ascent algorithm for a nearly quadratic function of two variables. The algorithm starts moving in the direction of the maximum gradient (i.e. perpendicular to the isocontour), and keeps moving along the same line until a maximum is reached.

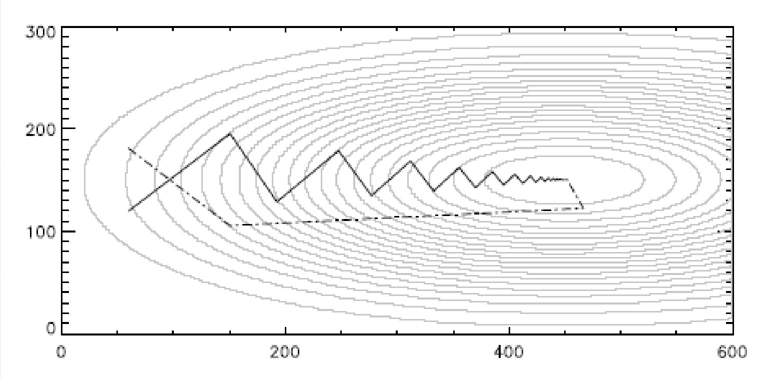


FIGURE 13.11. Dotted line: isocontours of the objective function. Solid line: convergence of the steepest gradient ascent algorithm. Dashed line: convergence of conjugate gradient ascent.

- This often leads to a zigzag line, requiring many iterations for good convergence.

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

- ☐ The conjugate gradient algorithm is designed to avoid these oscillations. The first iteration is identical to that of the steepest gradient ascent.
- ☐ However, in the following iterations, the algorithm attempts to move in a direction for which the gradient along the previous direction(s) remains the same (i.e. equal to zero).
- ☐ The idea is to eliminate the need for a new optimization along these previous directions.

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

- Let $\boldsymbol{d}_{old}$ be the previous direction and $\boldsymbol{H}$ the Hessian matrix (i.e. the second derivatives). It is now required that the new direction $\boldsymbol{d}_{new}$ be such that the gradient along $\boldsymbol{d}_{old}$ does not change. When moving in direction $\boldsymbol{d}_{new}$, the gradient will change (using a quadratic approximation) as $\boldsymbol{H}\boldsymbol{d}_{new}$. Requiring that the resulting change along $\boldsymbol{d}_{old}$ is zero yields the condition:

$$\boldsymbol{d}'_{old}\boldsymbol{H}\boldsymbol{d}_{new} = 0 \qquad (13.57)$$

- This behaviour is illustrated by the dashed line in Fig. 13.11. (For further information see Handbook Sec. 13.3.2.2 and 13.3.2.3).

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

Optimization transfer

- The log-likelihood function (Eq. (13.47)) can be maximized by setting its gradients (Eq. (13.50)) to zero for all $j = 1 \ldots J$. A problem is that each of these derivatives is a function of many voxels of λ, which makes the set of equations very hard to solve. The idea of 'optimization transfer' is to replace the problematic log-likelihood function with another function $\Phi(\lambda)$ that leads to a simpler set of equations, usually one where the derivative with respect to λ_j is only a function of λ_j and not of the other voxels of λ. That makes the problem separable into J 1-D optimizations, which are easily solved.

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

- The key is to design $\Phi(\lambda)$ in such a way that maximization of $\Phi(\lambda)$ is guaranteed to increase $L(\lambda)$. This leads to an iterative algorithm, since new functions Φ will have to be designed and maximized repeatedly to maximize $\boldsymbol{L}$. At iteration k, the surrogate function $\Phi(\lambda)$ needs to satisfy the following conditions (see Fig. 13.12):

$$\Phi(\boldsymbol{\lambda}^{(k)}) = L(\boldsymbol{\lambda}^{(k)}) \qquad (13.60)$$

$$\Phi(\boldsymbol{X}) \le L(\boldsymbol{X}) \qquad (13.61)$$

- It follows that the new reconstruction image $\lambda^{(k+1)}$ which maximizes $\Phi(\lambda)$ has higher likelihood than $\lambda^{(k)}$:

$$L(\boldsymbol{\lambda}^{(k)}) = \Phi(\boldsymbol{\lambda}^{(k)}) \le \Phi(\boldsymbol{\lambda}^{(k+1)}) \le L(\boldsymbol{\lambda}(^{k+1)}) \qquad (13.62)$$

13.3 ITERATIVE RECONSTRUCTION

13.3.2 Optimization algorithms

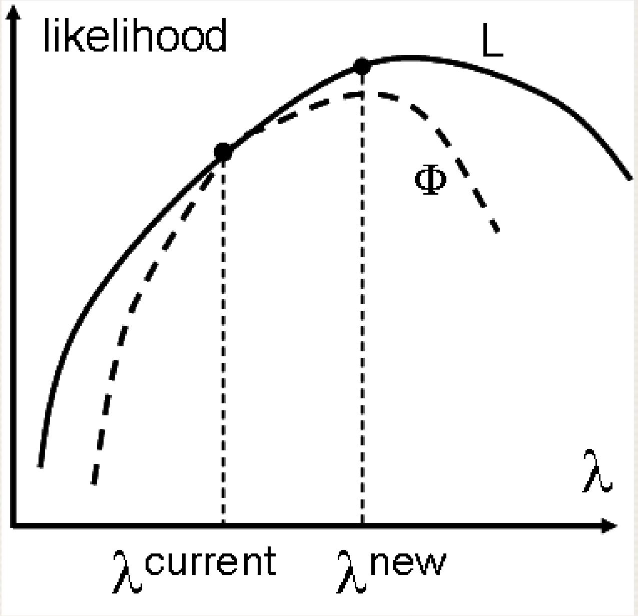


FIG. 13.12. Optimization transfer: a surrogate function is designed, which is equal to the likelihood in the current reconstruction, and less or equal everywhere else.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

Reconstruction from sinogram data

- ☐ Recall that we wish to find the image λ that maximizes the likelihood function L_{ML} of Eq. (13.47). The expectation maximization does this in a remarkable way. Instead of concentrating on L_{ML}, an alternative (different) likelihood function is derived by introducing a set of so-called ‘complete data’ $\boldsymbol{x}_{ij}$, defined as the number of photons that were emitted at voxel j and detected in LOR i during the measurement. These unobserved data are ‘complete’ in the sense that they describe in more detail than the observed data $\boldsymbol{y}_i$ what happened during the measurement. The variables $\boldsymbol{x}_{ij}$ are Poisson distributed.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- Just as for the actual data $\boldsymbol{y}_i$, one can write the log-likelihood function for observing the data $\boldsymbol{x}_{ij}$ while $\bar{\boldsymbol{x}}_{ij} = \boldsymbol{A}_{ij}\boldsymbol{\lambda}_j$ were expected:

$$L_x(\boldsymbol{\lambda}) = \sum_i \sum_j x_{ij} \ln(A_{ij}\lambda_j) - A_{ij}\lambda_j \qquad (13.63)$$

- However this cannot be computed because the $\boldsymbol{x}_{ij}$ are not available. The emission measurement only produces sums of the complete data, since

$$y_i = \sum_j A_{ij} x_{ij} + b_i \qquad (13.64)$$

where $\boldsymbol{b}_i$ is the actual (also unobserved) additive contribution $\boldsymbol{b}_i$ in LOR i.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ Expectation maximization (EM) requires computing the expectation of L_x, based on available data and the current reconstruction $\lambda^{(k)}$. From the latter we could write $\boldsymbol{E}(\boldsymbol{x}_{ij}|\boldsymbol{\lambda}^{(k)}) = \boldsymbol{A}_{ij}\lambda_j^{(k)}$. Since $\boldsymbol{x}_{ij}$ should satisfy Eq. (13.64) we have

$$E(x_{ij} \,|\, \boldsymbol{\lambda}^{(k)}, \boldsymbol{y}) = \frac{y_i}{\sum_j A_{ij}\lambda_j^{(k)} + \overline{b}_i} A_{ij}\lambda_j^{(k)} \qquad (13.65)$$

where $\overline{b}_i$ is the noise-free estimate of b_i, which is assumed to be available.

- ☐ Inserting this in Eq. (13.63) produces the expectation $L_x(\lambda)$ and completes the expectation ($\boldsymbol{E}$) step.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- For the maximization (M) step, the first derivatives are simply set to zero:

$$\frac{\partial L_x(\boldsymbol{\lambda})}{\partial \lambda_j} = \sum_i \left(\frac{y_i}{\sum_j A_{ij} \lambda_j^{(k)} + \bar{b}_i} A_{ij} \lambda_j^{(k)} \frac{1}{\lambda_j} - A_{ij} \right) = 0 \qquad (13.66)$$

- This is easily solved for λ_j, yielding the new reconstruction $\lambda_j^{(k+1)}$:

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_i A_{ij}} \sum_i A_{ij} \frac{y_i}{\sum_j A_{ij} \lambda_j^{(k)} + \bar{b}_i} \qquad (13.67)$$

- This is the well-known MLEM algorithm for emission tomography.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ It can be shown that this recipe has the wonderful feature that each new EM iteration increases the value of the likelihood L_{ML}. It should be noted that the complete data $\boldsymbol{x}_{ij}$ do not appear in Eq. (13.67); they are needed in the derivation but they do not need to be computed explicitly. This is very fortunate as there is a huge number of them.
- ☐ An initial image $\lambda^{(1)}$ is required to start the iterations. As experience (and theoretical analysis) has shown that higher spatial frequencies have slower convergence, and because smooth images are preferred, the initial image is usually chosen to be uniform, by setting $\lambda_j^{(1)} = C$ and $j = 1 \ldots J$, where C is a strictly positive constant.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- The MLEM algorithm is multiplicative, implying that it cannot change the value of a reconstruction voxel, when the current value is zero. For this reason, the voxels in the initial image should only be set to zero if it is known a priori that they are indeed zero.
- The derivation of the MLEM algorithm uses the assumption that all y_i, all x_{ij} and all λ_j are non-negative. Assuming that $y_i \geq 0$ and $i = 1 \ldots I$, and considering that the probabilities A_{ij} are also non-negative, it is clear that when the initial image $\lambda^{(1)}$ is non-negative, all subsequent images $\lambda^{(k)}$ will be non-negative as well.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- However, when, for some reason, a reconstruction value becomes negative (e.g. because one or a few sinogram values y_i are negative), then convergence is no longer guaranteed. In practice, divergence is almost guaranteed in that case. Consequently, if the sinogram is pre-preprocessed with a procedure that may produce negatives (e.g. randoms subtraction in PET), MLEM reconstruction will only work if all negative values are set to a non-negative value.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

Reconstruction from list-mode data

- ☐ The measured data y_i considered in the derivations above represent the number of counts acquired within an individual crystal pair i (LOR i), that is, y_i represents the sum of those acquired events (indexed by m) that were assigned (histogrammed) to the i-th LOR: $y_i = \sum_{m \in i} 1$
- ☐ However, in modern PET systems, the number of possible LORs within the FOV typically greatly exceeds the number of events acquired in a clinical study. Thus, the binned data are very sparse and it is more efficient to store and process each acquired event (with all of its relevant information) separately, in the so-called ‘list-mode’ format.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- Modification of the maximum-likelihood algorithms is straightforward (whether MLEM or accelerated algorithms based on ordered subsets discussed in later), as shown in works by Parra and Barrett [1], and by Reader et al [2].
- It should be noted that the same is not true about other algorithms, for example, algorithms with additive updates.
- The MLEM algorithm for the list-mode data can be obtained by replacing y_i in the MLEM equation (Eq. (13.67)) by the sum over events, skipping the LORs with zero counts (which do not contribute to the MLEM sum).

[1] PARRA, L., BARRETT, H.H., List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET, IEEE Trans. Med. Imaging 17 2 (1998) 228–235.

[2] READER, A.J., ERLANDSSON, K., FLOWER, M.A., OTT, R.J., Fast accurate iterative reconstruction for low-statistics positron volume imaging, Phys. Med. Biol. 43 4 (1998) 835–846.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ Combining the sum over LORs i with the sum over events m:

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum\limits_{i \in \text{LORs}} A_{ij}} \sum_{m \in \text{event-list}} A_{i_m j} \frac{1}{\sum\limits_j A_{i_m j} \lambda_j^{(k)} + \bar{b}_{i_m}} \qquad (13.68)$$

  where i_m represents the LOR index in which the m-th event has been recorded.

- ☐ The main difference is that the MLEM sum is now evaluated (including calculations of the relevant forward and backprojections) only over the list of the available events (in any order).

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ However, it is important to mention here that the normalizing term in front of the sum (sensitivity matrix $\sum_i A_{ij}$) still has to be calculated over all possible LORs, and not only those with non-zero counts.
- ☐ This represents a challenge for the attenuated data (attenuation considered as part of the system matrix $\boldsymbol{A}$), since the sensitivity matrix has to be calculated specifically for each object and, therefore, it cannot be pre-computed. For modern systems with a large number of LORs, calculation of it often takes more time than the list-mode reconstruction itself. For this reason some approximate methods have been proposed [1, 2].

[1] QI, J., Calculation of the sensitivity image in list-mode reconstruction, IEEE Trans. Nucl. Sci. 53 (2006) 2746–2751.

[2] MATEJ, S., et al., Efficient 3-D TOF PET reconstruction using view-grouped histo images: DIRECT — Direct Image Reconstruction for TOF, IEEE Trans. Med. Imaging 28 (2009) 739–751.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

Reconstruction of time of flight (TOF) PET data

- ☐ In TOF PET, the probability of a pair of photons arriving from a particular point along the LOR (as reported based on the difference of their detection times) is given by a Gaussian kernel having a width determined by the timing uncertainty of the detection system. In contrast, in the non-TOF case, the probability of detecting the event is approximately uniform along the LOR.
- ☐ Modification of iterative reconstruction algorithms to account for the TOF is straightforward. Integrations along the LORs (the main component of the system matrix A) just need to be replaced with the TOF kernel weighted integrations along the LORs.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- The forward-projection (or backprojection) in a certain direction can now be viewed, and performed, as a convolution of the image with a proper TOF kernel in the LOR direction (Fig. 13.13).

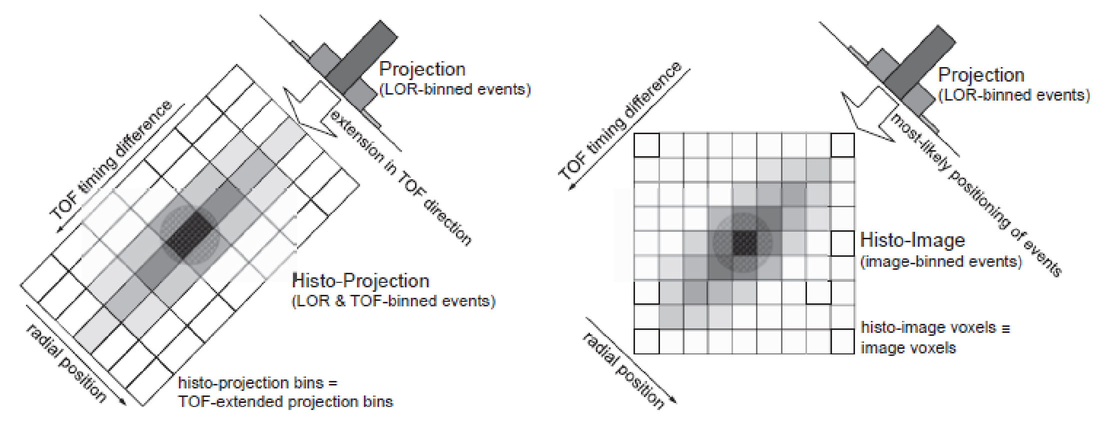


FIG. 13.13. Comparison of the data formats for binned time of flight (TOF) data (left: histo-projection for a 45° view) and for the DIRECT (direct image reconstruction for TOF) approach (right: histo-image for a 45° view). Histo-projections can be viewed as an extension of individual non-TOF projections into TOF directions (time bins), and their sampling intervals relate to the projection geometry and timing resolution. Histo-images are defined by the geometry and desired sampling of the reconstructed image. Acquired events and correction factors are directly placed into the image resolution elements of individual histo-images (one histo-image per view) having a one to one correspondence with the reconstructed image voxels.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- The rest of the algorithm, i.e. formulas derived in the previous subsections, stays exactly the same (only the form of the system matrix $\boldsymbol{A}$ is changed). Additional information provided by the TOF measurements, leading to more localized data, results in faster, and more uniform, convergence, as well as in improved signal to noise ratios in reconstructed images, as widely reported in the literature.
- The TOF mode of operation has some practical consequences (and novel possibilities) on the ways the acquired data are stored and processed.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ The list-mode format is very similar to the non-TOF case. The event structure is just slightly expanded by a few bits (5–8 bits/event) to include the TOF information, and the events are processed event by event as in the non-TOF case.
- ☐ On the other hand, the binned data undergo considerable expansion when accommodating the TOF information. Namely, the projection (X ray transform) structures are expanded by one dimension, that is, each projection bin is expanded in the LOR direction into the set of time bins forming the so-called histo-projections (see Fig. 13.13 (left)).

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ In practice, the effect of this expansion on the data size is not as bad as it appears, because the localized nature of TOF data allows decreased angular sampling (typically about 5–10 times) in both azimuthal and co-polar directions (views), while still satisfying angular sampling requirements.
- ☐ TOF also allows a conceptually different approach of data partitioning, leading to more efficient reconstruction implementations, by using the DIRECT (direct image reconstruction for TOF) approach utilizing so-called histo-images (see Fig. 13.13 (right)).

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

- ☐ In the DIRECT approach, the data are directly histogrammed (deposited), for each view, into image resolution elements (voxels) of desired size.
- ☐ Similarly, all correction arrays and data are estimated or calculated in the same histo-image format. The fact that all data and image structures are now in image arrays (of the same geometry and size) makes possible very efficient computer implementations of the data processing and reconstruction operations.

13.3 ITERATIVE RECONSTRUCTION

13.3.3 Maximum-likelihood expectation-maximization

Reconstruction of dynamic data

- ☐ Data acquired from an object dynamically changing with time in activity distribution, or in morphology (shape), or in both is referred to as dynamic data. An example of the first case would be a study looking at temporal changes in activity uptake in individual organs or tissues, so-called time–activity curves.
- ☐ An example of the second case would be a gated cardiac study providing information about changes of the heart morphology during the heart beat cycle (such as changes of the heart wall thickness or movements of the heart structures). Further details: Handbook Sec. 13.3.3.4.

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

Ordered subsets expectation maximization

- ☐ The MLEM algorithm requires a projection and a back-projection in every iteration, which are operations involving a large number of computations.
- ☐ Typically, MLEM needs several tens to hundreds of iterations for good convergence.
- ☐ Consequently, MLEM reconstruction is slow and many researchers have studied methods to accelerate convergence.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- The method most widely used is ordered-subsets expectation-maximization (OSEM). The MLEM algorithm (Eq. (13.67)) is rewritten here for convenience:

$$\hat{y}_i^{(k)} = \sum_j A_{ij} \lambda_j^{(k)} + \bar{b}_i \qquad (13.69)$$

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_i A_{ij}} \sum_i A_{ij} \frac{y_i}{\hat{y}_i^{(k)}} \qquad (13.70)$$

where k is the iteration number and $\lambda^{(1)}$ is typically set to a uniform, strictly positive image.

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- In OSEM, the set of all projections $\{1 \ldots I\}$ is divided into a series of subsets $\boldsymbol{S}_t$, $t = 1 \ldots T$. Usually, these subsets are exhaustive and non-overlapping, i.e. every projection element i belongs to exactly one subset $\boldsymbol{S}_t$. In SPECT and PET, the data $\boldsymbol{y}$ are usually organized as a set of (parallel or fanbeam) projections, indexed by projection angle ϕ.
- Therefore, the easiest way to produce subsets of $\boldsymbol{y}$ is by assigning all of the data for each projection angle to exactly one of the subsets.
- However, if the data $\boldsymbol{y}$ are stored in list-mode (see Section 13.3.3), the easiest way is to simply cut the list into blocks, assigning each block to a different subset.

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- The OSEM algorithm can then be written as:

```
Initialize $\lambda_j$, $j = 1,\ldots J$
for $k = 1,\ldots K$
    for $t = 1,\ldots T$
        $\hat{y}_i = \sum_j A_{ij}\lambda_j + \bar{b}_i, \quad \mathrm{i} \in \boldsymbol{S}_{\mathrm{t}}$
        for $j = 1,\ldots J$
            $\lambda_j = \frac{\lambda_j}{\sum_{i \in \boldsymbol{S}_t} A_{ij}} \sum_{i \in \boldsymbol{S}_t} A_{ij} \frac{y_i}{\hat{y}_i}$        (13.71)
```

13.3 ITERATIVE RECONSTRUCTION
13.3.4 Acceleration

- If all of the projections are combined into a single subset, the OSEM algorithm is identical to the MLEM algorithm.
- Otherwise, a single OSEM iteration k consists of T sub-iterations, where each sub-iteration is similar to an MLEM iteration, except that the projection and back-projection are only done for the projections of the subset $\boldsymbol{S}_t$. If every sinogram pixel i is in exactly one subset, the computational burden of a single OSEM iteration is similar to that of an MLEM iteration. However, MLEM would update the image only once, while OSEM updates it T times. Experience shows that this improves convergence by a factor of about T, which is very significant.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- ☐ Convergence is only guaranteed for consistent data and provided that there is subset balance, which requires:

$$\sum_{i \in \boldsymbol{S}_t} A_{ij} = \sum_{i \in \boldsymbol{S}_u} A_{ij} \qquad (13.72)$$

  where $\boldsymbol{S}_t$ and $\boldsymbol{S}_u$ are different subsets.
- ☐ In practice, these conditions are never satisfied, and OSEM can be shown to converge to a limit cycle rather than to a unique solution, with the result that the OSEM reconstruction is noisier than the corresponding MLEM reconstruction.
- ☐ However, in many applications, the difference between the two is not clinically relevant.

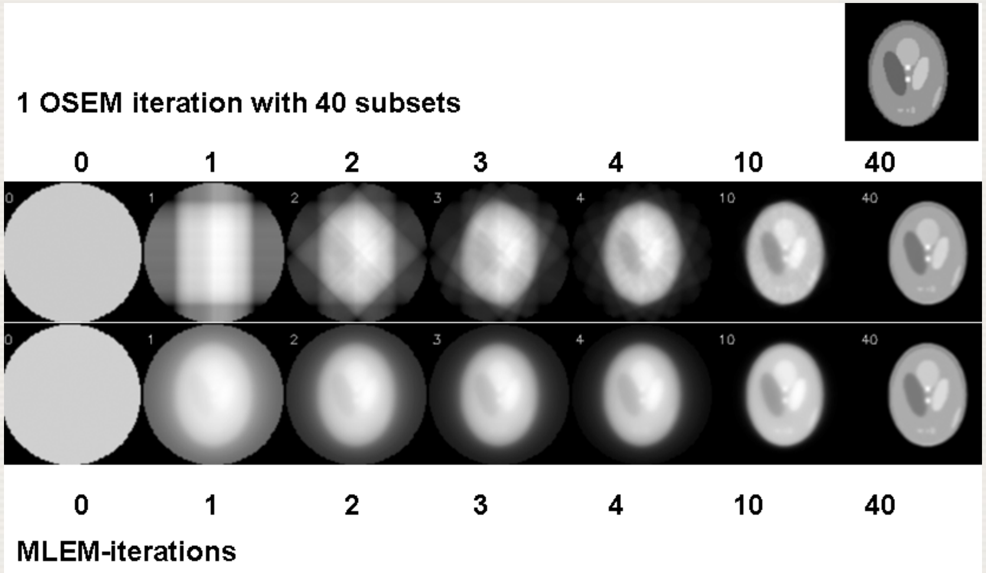
IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- ❑ The procedure is illustrated with a simple simulation in Fig. 13.14. As there was no noise and no attenuation, convergence of OSEM is guaranteed in this example.

FIG. 13.14. A simulation comparing a single ordered-subsets expectation-maximization (OSEM) iteration with 40 subsets, to 40 maximum-likelihood expectation-maximization (MLEM) iterations. The computation time of the MLEM reconstruction is about 40 times longer than that of OSEM. In this example, there were only two (parallel beam) projection angles per subset, which is clearly visible in the first OSEM iteration.



13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

Refinements of the OSEM algorithm

- ☐ As mentioned above, OSEM converges to a limit cycle: after many iterations, it starts cycling through a series of solutions rather than converging to the maximum likelihood solution. When compared to the initial image (usually a uniform image), these series of solutions are ‘relatively close’ to the maximum likelihood solution.
- ☐ Consequently, the convergence of OSEM is initially much faster but otherwise similar to that of MLEM; the better performance of MLEM only becomes noticeable at high iteration numbers. Thus, a simple solution to avoid the limit cycle is to gradually decrease the number of subsets.

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- This approach preserves the initial fast convergence of OSEM, avoiding the limit cycle by returning to MLEM at high iteration numbers. A drawback of this approach is that convergence becomes slower each time the number of subsets is reduced. In addition, there is no theory available that prescribes how many sub-iterations should be used for each OSEM iteration.
- Many algorithms have been proposed that use some form of relaxation to obtain convergence under less restrictive conditions than those of OSEM.

13.3 ITERATIVE RECONSTRUCTION

13.3.4 Acceleration

- As an example, relaxation can be introduced by rewriting the OSEM Eq. (13.71) in an additive way. Then, a relaxation factor α is inserted to scale the update term to obtain RAMLA (row-action maximum likelihood algorithm:

$$\lambda_j^{\text{new}} = \lambda_j^{\text{old}} + \alpha\lambda_j^{\text{old}} \sum_{i \in \boldsymbol{S}_t} A_{ij} \left(\frac{y_i}{\hat{y}_i} - 1 \right) \quad \text{with} \quad \alpha < \frac{1}{\max_t \left(\sum_{i \in \boldsymbol{S}_t} A_{ij} \right)} \qquad (13.73)$$

- The relaxation factor α decreases with increasing iteration number to ensure convergence. It should be noted that setting $\alpha = 1 / \sum_{i \in \boldsymbol{S}_t} A_{ij}$ for all (sub-)iterations yields OSEM.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- MLEM maximizes the likelihood, by making the computed projections (from current reconstruction) as similar as possible to the measured projections, where similarity is measured based on the Poisson distribution.
- An upper limit of the likelihood would be obtained when the measured and calculated projections are identical. However, this is never possible, because Poisson noise introduces inconsistencies.
- Nevertheless, a large part of the noise is consistent, and obtainable as the projection of a (noisy) activity distribution. This part of the noise propagates into the reconstructed image, and is responsible for the so-called ‘deterioriation’ of the MLEM image at high iterations.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

Stopping iterations early

- ☐ An ‘accidental’ feature of the MLEM algorithm is its frequency dependent convergence: low spatial frequencies converge faster than higher frequencies.
- ☐ This is due to the low-pass effect of the back-projection operation. This effect is easily verified for the reconstruction of the activity in a point source, if the MLEM reconstruction is started from a uniform image. The first iteration then yields the back-projection of the point source measurement.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- ☐ As discussed in Section 13.2.1, this yields an image with intensity $\lambda(x,y) \propto 1/\sqrt{x^2+y^2}$, if the point source was located at $(0,0)$.
- ☐ Each iteration multiplies with a similar back-projection, implying that after t iterations, the image intensity at (x,y) is proportional to $1/(x^2+y^2)^{t/2}$, so that the peak at $(0,0)$ becomes a bit sharper with every iteration. For more complicated objects, the evolution is more subtle.

13.3 ITERATIVE RECONSTRUCTION
13.3.5 Regularization

- It follows that reducing the number of iterations has an effect which is similar to reducing the cut-off frequency of a low-pass filter. However, the effect on the resolution is position dependent, as illustrated in Fig. 13.15.

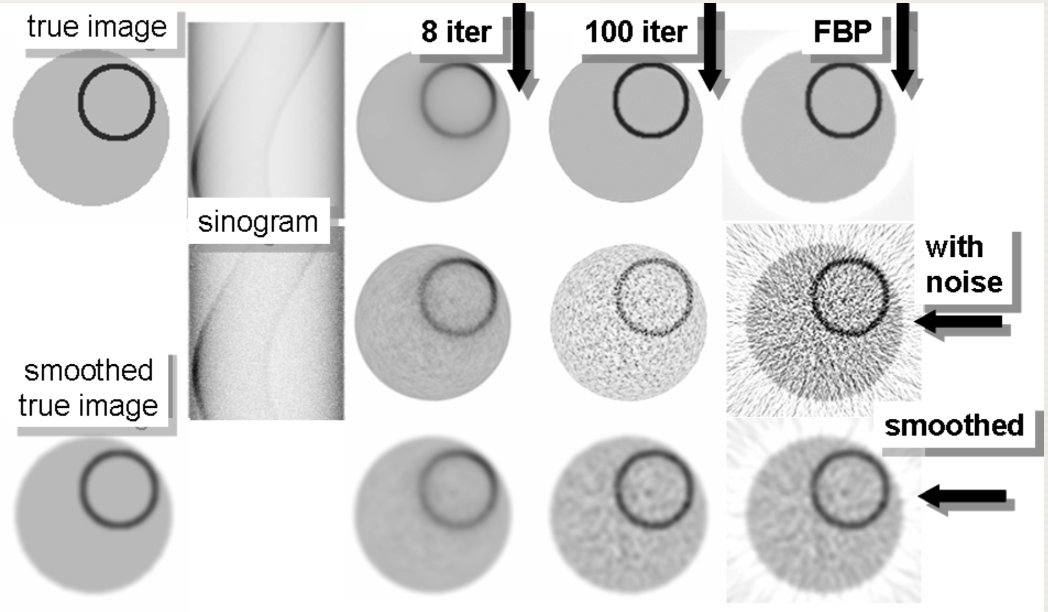


FIG. 13.15. Simulation study illustrating position dependent convergence in PET with attenuation. After 8 iterations, convergence in highly attenuated regions is poor. After 100 iterations, good convergence is obtained, but with strong noise propagation. Post-smoothing yields a fair compromise between noise and nearly position independent resolution.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

Post-smoothed maximum likelihood

- ☐ The noise in the higher MLEM iterations is high frequency noise, and there are strong negative correlations between neighbouring pixels. As a result, a modest amount of smoothing strongly suppresses the noise at the cost of a mild loss of resolution. This is illustrated in the third row of Fig. 13.15.
- ☐ If the MLEM implementation takes into account the (possibly position dependent) spatial resolution effects, then the resolution should improve with every MLEM iteration. After many iterations, spatial resolution should be good, similar or even better than the sinogram resolution, but the noise will have propagated greatly.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- Let us assume that the obtained spatial resolution corresponds to a position dependent point spread function which can be approximated as a Gaussian with a full width at half maximum (FWHM) of $F_{\mathrm{ML}}(x, y)$. Assume further that this image is post-smoothed with a (position independent) Gaussian convolution kernel with an FWHM of F_p. The local point spread function in the smoothed image will then have an FWHM of $\sqrt{(F_{\mathrm{ML}}(x,y))^2 + F_p^2}$
- If enough iterations are applied and if the post-smoothing kernel is sufficiently wide, the following relation holds F_p » $F_{\mathrm{ML}}(x, y)$ and, therefore,

$$\sqrt{(F_{\mathrm{ML}}(x,y))^2 + F^2_p} \approx F_p$$

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- Under these conditions, the post-smoothed MLEM image has a nearly position independent and predictable spatial resolution.
- Thus, if PET or SPECT images are acquired for quantification, it is recommended to use many iterations and post-smoothing, rather than a reduced number of iterations, for noise suppression.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

Smoothing basis functions

- An alternative approach to counter noise propagation is to use an image representation that does not accommodate noisy images. Instead of representing the image with a grid of non-overlapping pixels, a grid of smooth, overlapping basis functions can be used. The two mostly used approaches are the use of spherical basis functions or ‘blobs’ and the use of Gaussian basis functions or sieves.
- In the first approach, the projector and back-projector operators are typically adapted to work directly with line integrals of the basis functions.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- In the sieves approach, the projection of a Gaussian blob is usually modelled as the combination of a Gaussian convolution and projection along lines. The ‘blob’ approach produces a better approximation of the mathematics, while the sieves approach yields a faster implementation.
- The blobs or sieves are probably most effective when their width is very similar to the spatial resolution of the tomographic system. In this setting, the basis function allows accurate representation of the data measured by the tomographic system, and prevents reconstruction of much of the (high frequency) noise.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- If the blob or sieve is wider than the spatial resolution of the tomographic system, then its use during reconstruction produces Gibbs over- and undershoots, also known as ‘ringing’. This effect always arises when steep edges have to be represented with a limited frequency range, and is related to the ringing effects observed with very sharp low-pass filters. For some imaging tasks, these ringing artefacts are a disadvantage.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

Maximum a posteriori or penalized likelihood

- ☐ Smoothing the MLEM image is not a very elegant approach: first, the likelihood is maximized, and then it is decreased again by smoothing the image. It seems more elegant to modify the objective function, such that the image that maximizes it does not need further processing. This can be done with a Bayesian approach, which is equivalent to combining the likelihood with a penalty function.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- It is assumed that a good reconstruction image λ will be obtained if that image maximizes the (logarithm of the) probability $p(\lambda|\mathbf{y})$ given by Eq. (13.39) and repeated here for convenience:

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\lambda} (\ln p(\mathbf{y} \,|\, \boldsymbol{\lambda}) + \ln p(\boldsymbol{\lambda})) \qquad (13.74)$$

- The second term represents the a priori knowledge about the tracer distribution, and it can be used to express our belief that the true tracer distribution is fairly smooth. This is usually done with a Markov prior.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- In a Markov prior, the a priori probability for a particular voxel, given the value of all other voxels, is only a function of the direct neighbours of that voxel:

$$p(\lambda_j \mid \lambda_k, \forall k \neq j) = p(\lambda_j \mid \lambda_k,\ k \in \boldsymbol{N}_j) \qquad (13.75)$$

where $\boldsymbol{N}_j$ denotes the set of neighbour voxels of j. Such priors are usually written in the following form:

$$p(\boldsymbol{\lambda}) = \ln p(\boldsymbol{\lambda}) = \sum_j \ln p(\lambda_j \mid \lambda_k, k \in \boldsymbol{N}_j) = -\beta \sum_j \sum_{k \in \boldsymbol{N}_j} E(\lambda_j \lambda_k) \qquad (13.76)$$

where the ‘energy’ function E is designed to obtain the desired noise suppressing behaviour, and the parameter β is the weight assigned to the prior.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- A higher weight results in smoother images, at the cost of a decreased likelihood, i.e. poorer agreement with the acquired data. In most priors, the expression is further simplified by making E a function of a single variable, the absolute value of the difference $|\lambda_j - \lambda_k|$. Some popular energy functions $E(|\lambda_j - \lambda_k|)$ are shown in Fig. 13.16.

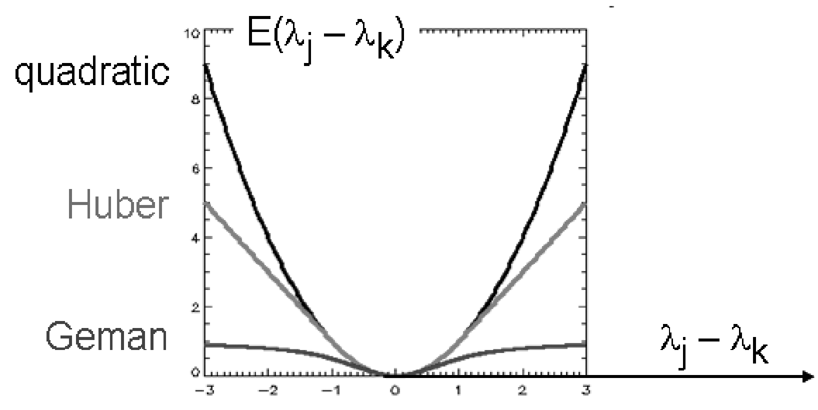


FIG. 13.16. The energy function of the quadratic prior, the Huber prior and the Geman prior.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- A simple and effective one is the quadratic prior ; a maximum a posteriori (MAP) reconstruction with this prior is shown in Fig. 13.17. Better preservation of strong edges is obtained with the Huber prior: it is quadratic for $|\lambda_j - \lambda_k| \leq \delta$ and linear for $|\lambda_j - \lambda_k| > \delta$, with a continuous first derivative at δ. Consequently, it applies less smoothing than the quadratic prior for differences larger than δ, as illustrated in Fig. 13.17.

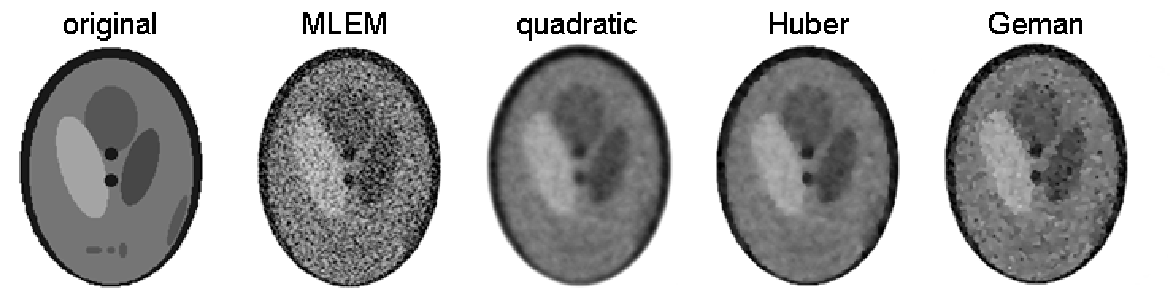


FIG. 13.17. MLEM and MAP reconstructions of the Shepp–Logan phantom. Three different smoothing priors were used: quadratic, Huber and Geman. The latter smooth small differences quadratically, but are more tolerant for large edges.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- It can be shown that the prior (Eq. (13.76)) is a concave function of λ if $E|\lambda_j - \lambda_k|$ is a convex function. Consequently, the quadratic and Huber energy functions yield a concave prior: it has a single maximum. In contrast, the Geman prior is not concave (see Fig. 13.16) and has local maxima. Such concave priors require careful initialization, because the final reconstruction depends on the initial image and on the behaviour of the optimization algorithm.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- Figure 13.18 shows that MAP reconstructions produce position dependent spatial resolution, similar to MLEM with a reduced number of iterations.

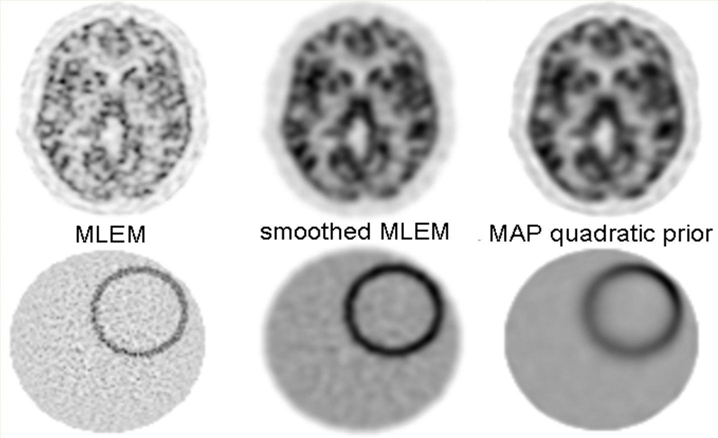


FIG. 13.18. Maximum-likelihood expectation-maximization (MLEM), smoothed MLEM and maximum a posteriori (MAP) (quadratic prior) reconstructions of simulated PET data of a brain and a ring phantom. The ring phantom reveals position dependent smoothing for MAP. The prior produces more smoothing in regions where the likelihood is ‘weaker’, e.g. regions that have contributed only a few photons to the measurement due to high attenuation.

- The reason is that the prior is applied with a uniform weight, whereas the likelihood provides more information about some voxels than about others.

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- The prior can be made position dependent as well, to ensure that the balance between the likelihood and the prior is about the same in the entire image. In that case, MAP with a quadratic prior produces images which are very similar to MLEM images with post-smoothing: if the prior and smoothing are tuned to produce the same spatial resolution, then both algorithms also produce nearly identical noise characteristics.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- Many papers have been devoted to the development of algorithms for MAP reconstruction. A popular algorithm is the so-called ‘one step late’ algorithm. Inserting the derivative of the prior P in Eq. (13.66) yields:

  $$\frac{\partial(L_x(\boldsymbol{\lambda})+P(\boldsymbol{\lambda}))}{\partial\lambda_j}=\sum_i\left(\frac{y_i}{\hat{y}_i^{(k)}}A_{ij}\lambda_j^{(k)}\frac{1}{\lambda_j}-A_{ij}\right)+\frac{\partial P(\boldsymbol{\lambda})}{\partial\lambda_j}=0 \qquad (13.77)$$

  where $\hat{\mathbf{y}}_i^{(k)}$ is the projection of the current reconstruction for detector i.

- A problem with this equation is that $\partial P(\boldsymbol{\lambda})/\partial\lambda_j$ is itself a function of the unknown image $\boldsymbol{\lambda}$. To avoid this problem, the derivative of the prior is simply evaluated in the current reconstruction $\boldsymbol{\lambda}^{(k)}$.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.5 Regularization

- ☐ The equation can then be solved to produce the MAP update expression:

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_i A_{ij} - \left.\frac{\partial P(\boldsymbol{\lambda})}{\partial \lambda_j}\right|_{\lambda^{(k)}}} \sum_i A_{ij} \frac{y_i}{\hat{y}_i^{(k)}} \qquad (13.78)$$

- ☐ Owing to the approximation, convergence is not guaranteed. The algorithm usually works fine, except with very high values for the prior.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ In typical emission data, the true events (having a Poisson character) are distorted and contaminated by a number of physical effects.
- ☐ To make the best use of the acquired data and of our knowledge of the acquisition system, these effects should be included in the reconstruction model.
- ☐ Distortion effects include resolution effects (such as detector resolution, collimator effects, and in PET also non-colinearity and positron range) and motion effects.
- ☐ Contamination effects can be divided, by their character and the way they are treated, into multiplicative and additive terms.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Multiplicative factors:
  - attenuation of the annihilation photons by the object,
  - the probability of the detector elements detecting an event once they are hit by the photon (detector normalization factors),
  - coefficients accounting for the decay time,
  - the geometrical restriction of directions/LORs for which true events are detected (axial acceptance angle, detector gaps).

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ Additive factors:
  - scattered coincidences,
  - random coincidences (in PET).

13.3.6 Corrections

- ☐ The most straightforward approach:
  - pre-correct the data before reconstruction for the contamination effects (multiplying by multiplicative correction coefficients and subtracting the scatter and random estimates), so as to approximate the X ray transform (or attenuated X ray transform in the SPECT case) of the reconstructed object.
- ☐ For analytical reconstruction approaches (derived for the ideal X ray transform data), the data always have to be pre-corrected.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- For the statistical reconstruction methods, derived based on the statistical properties of the data, an attempt is made to preserve the Poisson character of the data as much as possible by including the correction effects inside the reconstruction model.
- Theoretically, the most appropriate way is to include the multiplicative and scatter effects directly into the system matrix.
- The system matrix would have to include not only an accurate model of the direct data (true events) but also of the physical processes of the generation of the contamination scatter data.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- However, inclusion of the scatter model into the system matrix tremendously increases the number of non-zero elements of the system matrix, i.e. the matrix is not sparse anymore, and consequently the system is more ill-posed (the contamination data are typically quite noisy) and computationally exceedingly expensive, and, thus, not feasible for routine clinical use.

IAEA

13.3 ITERATIVE RECONSTRUCTION
13.3.6 Corrections

- ☐ The more practical, and commonly used, approach is to include correction effects as multiplicative factors and additive terms within the forward projection model of the iterative reconstruction approaches:

$$\boldsymbol{y} = \boldsymbol{A\lambda} + \boldsymbol{b} \qquad (13.82)$$

where the effects directly influencing the direct (true) data are included inside the system matrix $\boldsymbol{A}$, and the additive terms are represented by $\boldsymbol{b}$.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

Factors affecting direct events — multiplicative effects

- In PET, the sequence of the physical effects that occur as true coincident events are generated and detected can be described as a factorization of the system matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{A}_{\text{det.sens}} \boldsymbol{A}_{\text{det.blur}} \boldsymbol{A}_{\text{att}} \boldsymbol{A}_{\text{geom}} \boldsymbol{A}_{\text{tof}} \boldsymbol{A}_{\text{positron}} \qquad (13.83)$$

where $\boldsymbol{A}_{\text{positron}}$ models the positron range, $\boldsymbol{A}_{\text{tof}}$ models the timing accuracy for TOF PET systems, $\boldsymbol{A}_{\text{geom}}$ is the geometric projection matrix, a geometrical mapping between the source (voxel j) and data (projection bin i), $\boldsymbol{A}_{\text{att}}$ is a diagonal matrix containing attenuation factors on individual LORs, $\boldsymbol{A}_{\text{det.blur}}$ models the accuracy of reporting the true LOR positions, and $\boldsymbol{A}_{\text{det.sens}}$ models the probability that a photon pair reaching the detectors will be reported.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ $A_{\mathrm{det.sens}}$ is a unique multiplicative factor for each detector crystal pair (LOR) modelled by normalization coefficients, but it can also include the detector axial extent and detector gaps.
- ☐ In practice, the attenuation operation A_{att} is often moved to the left (to be performed after the blurring operation). This is strictly correct only if the attenuation factors change slowly, i.e. they do not change within the range of detector resolution kernels. However, even if this is not the case, a good approximation can be obtained by using blurred (with the detector resolution kernels) attenuation coefficients.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ❑ In this case, the multiplicative factors $\boldsymbol{A}_{\text{det.sens}}$ and $\boldsymbol{A}_{\text{att}}$ can be removed from the system matrix $\boldsymbol{A}$ and applied only after the forward projection operation as a simple multiplication operation (for each projection bin).
- ❑ The rest of the system matrix (except $\boldsymbol{A}_{\text{positron}}$, which is object dependent) can now be pre-computed, whether in a combined or a factorized form, since it is now independent of the reconstructed object.
- ❑ On the other hand, the attenuation factors $\boldsymbol{A}_{\text{att}}$ (and $\boldsymbol{A}_{\text{positron}}$, if considered) have to be calculated for each given object.

IAEA

13.3 ITERATIVE RECONSTRUCTION
13.3.6 Corrections

- In SPECT, the physical effects affecting the true events can be categorized and factorized into the following sequence:

$$A = A_{\text{det.sens}} A_{\text{det.blur}} A_{\text{geom,att}} \qquad (13.84)$$

where $A_{\text{det.sens}}$ includes multiplicative factors (such as detector efficiency and decay time), $A_{\text{det.blur}}$ represents the resolution effects within the gamma camera (the intrinsic resolution of the system), and $A_{\text{geom,att}}$ is the geometric projection matrix, also including the collimator effects (such as the depth dependent resolution) and the depth and view dependent attenuation factors.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- For gamma cameras, the energy and linearity corrections are usually performed in real time, and the remaining (detector efficiency) normalization factors are usually very close to one and can be, for all practical purposes, ignored or pre-corrected.
- In theory, the decay correction should be performed during the reconstruction, because it is different for each projection angle. However, for most tracers, the decay during the scan is very modest, and in practice it is usually either ignored or done as a pre-correction. The attenuation component is object dependent and needs to be recalculated for each reconstructed object.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

Additive contributions

- The main additive contaminations are scatter (SPECT and PET) and random events (PET). The simplest possibility of dealing with them is to subtract their estimates ($\bar{s}$ and $\bar{r}$) from the acquired data. While this is a valid (and necessary) pre-correction step for the analytical reconstructions, it is not recommended for statistical approaches since it changes the statistical properties of the data, causing them to lose their Poisson character. As the maximum likelihood algorithm is designed for Poisson distributed data, its performance is suboptimal if the data noise is different from Poisson.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Furthermore, subtraction of the estimated additive terms from the noisy acquired data can introduce negative values into the pre-corrected data, especially for low count studies.
- The negative values have to be truncated before the maximum likelihood reconstruction, since it is not able to correctly handle the negative data.
- This truncation, however, leads to a bias in the reconstruction.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ We could consider the scatter and randoms directly in the (full) system model, i.e. include a complete physical model of the scatter and random components into a Monte Carlo calculation of the forward projection.
- ☐ However, this approach is exceedingly computationally expensive and is not feasible for practical use.
- ☐ A practical and common approach for dealing with the additive contaminations is to add their estimate ($\bar{\boldsymbol{b}} = \bar{\boldsymbol{s}} + \bar{\boldsymbol{r}}$) to the forward projection in the matrix model of the iterative reconstruction, i.e. $\boldsymbol{A\lambda} + \bar{\boldsymbol{b}}$ as considered in the derivation of the MLEM reconstruction (Eq. (13.67)).

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Special treatment has to be considered for clinical scanners in which the random events ($\boldsymbol{r}$, estimated by delayed coincidences) are on-line subtracted from the acquired data ($\boldsymbol{y}$, events in the coincidence window — prompts).
- After subtracting the delays from the prompts (both being Poisson variables), the resulting data ($\boldsymbol{\gamma}$) are not Poisson anymore, since var($\boldsymbol{\gamma}_i$) = var($\boldsymbol{y}_i - \boldsymbol{r}_i$) = var($\boldsymbol{y}_i$) + var($\boldsymbol{r}_i$).
- To regain Poisson characteristics, the shifted Poisson approach can be used, using the fact that adding a constant value to the Poisson variable changes the mean but preserves the variance of the result.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- To modify the mean of the subtracted data $\boldsymbol{\gamma}$ to be equal to their variance (i.e. var($\boldsymbol{y}_i$) + var($\boldsymbol{r}_i$)), we need to add to the subtracted data an estimate (of the mean) of the randoms $\bar{\boldsymbol{r}}$ multiplied by two. This gives

$$\mathrm{mean}(\gamma_i + 2\bar{r}_i) = \mathrm{mean}(y_i - r_i + 2\bar{r}_i) = \mathrm{mean}(y_i) + \mathrm{mean}(r_i)$$

which is equal to $\mathrm{var}(\gamma_i + 2\bar{r}_i) = \mathrm{var}(y_i) + \mathrm{var}(r_i)$. The MLEM algorithm using the shifted Poisson model can then be written as:

$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_i A_{ij}} \sum_i A_{ij} \frac{\gamma_i + 2\bar{r}_i}{\sum_j A_{ij}\lambda_j^{(k)} + \bar{s}_i + 2\bar{r}_i} \qquad (13.85)$$

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Even in the shifted Poisson case, the negative values in the subtracted data and consequent truncation leading to the bias and artefacts cannot be completely avoided.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Examples of reconstructions from data with a subtracted additive term, using the regular MLEM algorithm and using MLEM with the shifted Poisson model, are shown in Fig. 13.19.

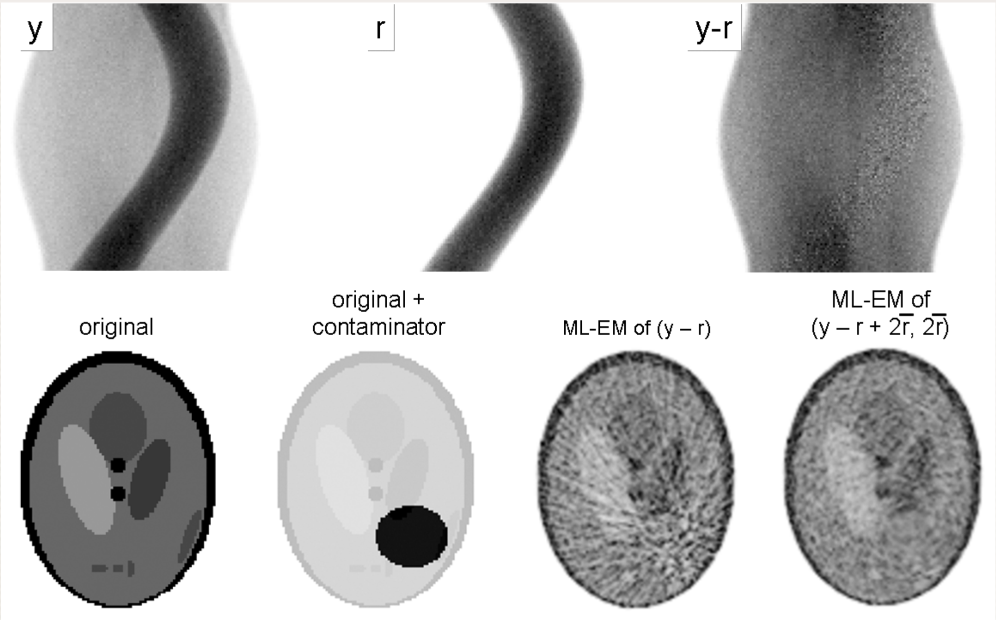


FIG. 13.19. Illustration of (exaggerated case of) reconstructions from contaminated data $\boldsymbol{y}$ from which the additive contamination term $\boldsymbol{r}$ was subtracted (both data and contamination term are Poisson). The top row shows the sinograms. The increased noise level in the contaminated area in the sinogram $(\boldsymbol{y}-\boldsymbol{r})$ should be noted. The bottom row shows the true image without and with the contaminator, the maximum-likelihood expectation-maximization (MLEM) reconstruction from the subtracted data $(\boldsymbol{y}-\boldsymbol{r})$ and the shifted Poisson MLEM reconstruction, in which the estimated (noiseless) additive term $2\bar{\boldsymbol{r}}$ is added to the subtracted data and forward projection as given by Eq. (13.85).

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ As the counts were relatively high in this simulation, the subtraction did not produce negatives. MLEM of $(\boldsymbol{y} - \boldsymbol{r})$ creates streaks because the reliability of the subtracted data is overestimated.
- ☐ It should be noted that in the reconstruction model (as well as in the pre-correction approaches) the estimates of the scatter and randoms have to be treated in the same way as the estimates of the true events in the forward projection, including consideration of the normalized or un-normalized events, attenuation corrected or uncorrected data, gaps in the data, etc.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

Finite spatial resolution

- ☐ A number of physical and geometrical effects and limitations (such as positron range, acollinearity, depth of interaction, size of detector crystal elements, inter-crystal scatter, collimator geometry, etc.) affect PET and SPECT spatial resolution.
- ☐ To get the most out of the acquired data and to correct for the resolution degradation, these effects have to be properly modelled in the system matrix of statistical reconstruction, as considered in the components ($\boldsymbol{A}_{\mathrm{det.blur}}$, $\boldsymbol{A}_{\mathrm{geom}}$, $\boldsymbol{A}_{\mathrm{positron}}$) of the factorized system matrix as outlined previously.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ This step does not influence the mathematical definition of the reconstruction algorithm; only the form of its system matrix is changed.
- ☐ However, it has very practical consequences for the complexity of the algorithm implementation, for computational demands and most importantly for the quality of the reconstructed images.
- ☐ By including resolution effects into the reconstruction model, more of the data is being used for the reconstruction of each point in the space, with the true signal component becoming more consistent, and noise components becoming less consistent with the model.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Thus, the resolution modelling helps twice, by improving the image resolution while at the same time reducing the image noise, as illustrated in Fig. 13.20 for simulated SPECT data.

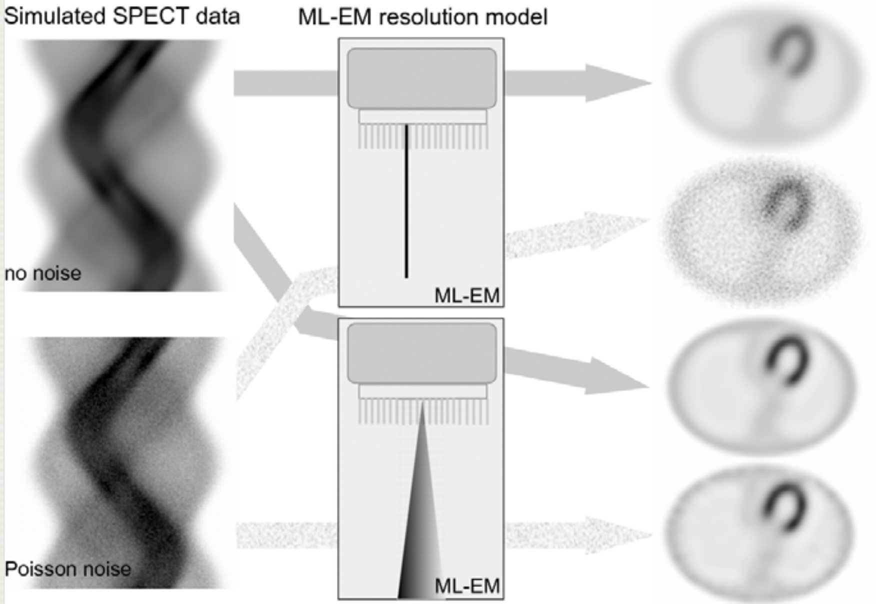


FIG. 13.20. Examples of the effects of resolution modelling within statistical iterative reconstruction. Data were simulated for a SPECT system with depth dependent resolution. It is clearly seen that using the proper resolution model within statistical reconstruction (lower two images on the right) not only improves resolution of the images, but also helps to efficiently suppress the noise component.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- This is quite different from the filtering case, where the noise suppression is always accompanied by resolution deterioration.
- On the other hand, the resolution modelling has a price in terms of a considerable increase in the computational load (both in space/memory and time) since the system matrix is much less sparse, that is, it contains a larger proportion of non-zero elements.
- This not only leads to more computational load per iteration, but also to a slower convergence of the iterative reconstruction and, consequently, to the need for more iterations.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Resolution effects can be subdivided into the effects dependent on the particular object, such as the positron range, and the effects influenced by the scanner geometry, design and materials (which can be determined beforehand for the given scanner).
- The positron range depends on the particular attenuation structures in which the the positrons annihilate, and also varies from isotope to isotope.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Furthermore, the shape of the probability function (kernel) of the positron annihilation abruptly changes at the boundaries of two tissues, such as at the boundary of the lungs and surrounding soft tissues, and, thus, it strongly depends on the particular object's morphology and is quite challenging to model accurately.
- In general, the positron range has a small effect (compared to the other effects) for clinical scanners, particularly for studies using ^{18}F-labelled tracers, and can often be ignored. However, for small animal imaging and for other tracers (such as ^{82}Rb), the positron range becomes an important effect to be considered.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ There is a whole spectrum of approaches to determine and implement the scanner dependent resolution models.
- ☐ Approach 1. (Simplest, but least accurate):
  - approximate the system resolution model by a spatially invariant resolution kernel, usually a spherically symmetric Gaussian, with the shape (FWHM) estimated from point source measurements at one or more representative locations within the given scanner.
  - This typically provides satisfactory results within the central FOV of large, whole body PET scanners. However, for PET systems with smaller ring diameters (relative to the reconstruction FOV), (e.g. animal systems and SPECT systems with depth dependent resolution (in particular with non-circular orbits), it is desirable to use more accurate spatially variant resolution models.

13.3.6 Corrections

☐ Approach 2:

- Use analytically calculated resolution functions (usually spatially variant anisotropic kernels) for each location (LOR) as determined based on analytical models of physical effects affecting the resolution.
- This approach is usually limited to simple analytical models representing (or approximating) only basic physical characteristics of the system. The resolution kernels are usually calculated in real time during the reconstruction process when they are needed within the forward and back-projection calculations.
- In SPECT, distance dependent collimator blurring requires convolution kernels that become wider and, therefore, need more computation, with increasing distance to the collimator.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

☐ Approach 3 (More accurate but computationally very demanding):

- Use Monte Carlo simulations of the resolution functions based on a set of point sources at various (ideally all) image locations.
- Setting up an accurate mathematical model (transport equations tracing the photon paths through the detector system/crystals) is relatively easy within the Monte Carlo simulations, compared to the analytical approach of determining the resolution function.
- However, to obtain sufficient statistics to get the desired accuracy of the shape of the resolution functions is extremely time consuming. Consequently, simplifications often have to be made in practice, such as determining the resolution kernels only at a set of representative locations and interpolating/extrapolating from them the resolution kernels at other locations.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ Approach 4 (Most accurate but most involved):
  - Use experimental measurements of the system response by measuring physical point sources at a set of image locations within the scanner.
  - This is a tedious and very time consuming process, involving point sources with long half-life isotopes and usually requiring the use of accurate robotic stages to move the point source.
  - Among the biggest challenges is to accumulate a sufficient number of counts to obtain an accurate point spread function, even at a limited number of locations.
  - Consequently, the actual resolution kernels used in the reconstruction model are often estimated by fitting analytical functions (kernels) to the measured data, rather than directly using the measured point spread functions.

13.3.6 Corrections

- ☐ Although it has been shown in the literature that proper system models lead to improved reconstructed image quality, they can never fully recover information that has been lost through resolution effects and other instrumentation limitations.
- ☐ Due to the increased level of modelling, the system matrix becomes more dense, and the reconstruction becomes more ill-posed, making it impossible to attain perfect recovery for the realistic data.
- ☐ Improved instrumentation as well as novel and more accurate reconstruction models will play an important role in improving image quality and quantitative accuracy.

13.3 ITERATIVE RECONSTRUCTION
13.3.6 Corrections

Motion corrections

- ☐ Owing to the relatively long acquisition times, motion effects, caused by patient movement and organ motion and deformation, cannot be avoided in emission tomography.
- ☐ With the continuous improvements of PET and SPECT technology, leading to improved spatial resolution, signal to noise ratio, image quality and accuracy of quantitative studies, corrections for motion effects become more important.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ In fact, artefacts caused by motion are becoming the single most important factor for image degradation, especially in PET or PET/computed tomography (CT) imaging of the upper torso region.
- ☐ For example, motion effects can lead to the loss of small lesions by blurring them out completely, or to their misplacement into the wrong anatomical region (e.g. into the liver from the lungs, or vice versa).
- ☐ The two main sources of motion related artefacts in emission studies are the motion during the emission scan and the discrepancy (caused by the motion) between the attenuation and emission data.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- The motion during the emission scan means that the emission paths (LORs) through the object (as considered in the system matrix) change during the scan time. If this time dependent change is not accounted for, the system model becomes inconsistent with the data, which results in artefacts and motion blurring in the reconstructed images.
- On the other hand, the transmission scan (CT) is relatively short and can usually be done in a breath-hold mode. Consequently, the attenuation image is usually motion-free and captures only one particular patient position and organ configuration (time frame).

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- If the attenuation factors obtained from this fixed-time position attenuation image are applied to the emission data acquired at different time frames (or averaged over many time frames), this leads to artefacts in the reconstructed images, which tend to be far more severe in PET than in SPECT.

IAEA

13.3.6 Corrections

☐ Motion correction of emission data:

- The first step is subdividing the data (in PET, typically list-mode data) into a sufficient number of time frames to ensure that the motion within each frame is small.
- For the organ movement, the frames can be distributed over a period of the organ motion (e.g. breathing cycle).
- For the patient motion, the frames would be typically longer and distributed throughout the scan time.
- Knowledge about the motion can be obtained using external devices, such as cameras with fiducial markers, expansion belts or breathing sensors for respiratory motion, the electrocardiogram signal for cardiac motion, etc.
- There are also a limited number of approaches for estimating the motion directly from the data.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- Once the data are subdivided into the set of the frames, the most straightforward approach is to reconstruct data independently in each frame.
- The problem with this approach is that the resulting images have a poor signal to noise ratio because the acquired counts have been distributed into a number of individual (now low count) frames.
- To improve the signal to noise ratio, the reconstructed images for individual frames can be combined (averaged) after they are registered (and properly deformed) to the reference time frame image.

IAEA

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- However, for statistical non-linear iterative reconstruction algorithms, this is not equivalent to (and typically of a lower quality than) the more elaborate motion correction approaches, taking into account all of the acquired counts in a single reconstruction.
- For rigid motion (e.g. in brain imaging), the events on LORs (LOR_i) from each time frame, or time position, can be corrected for motion by translation (using affine transformations) into the new LORs (LOR_i) in the reference frame (see next slide), in which the events would be detected if there were no motion.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

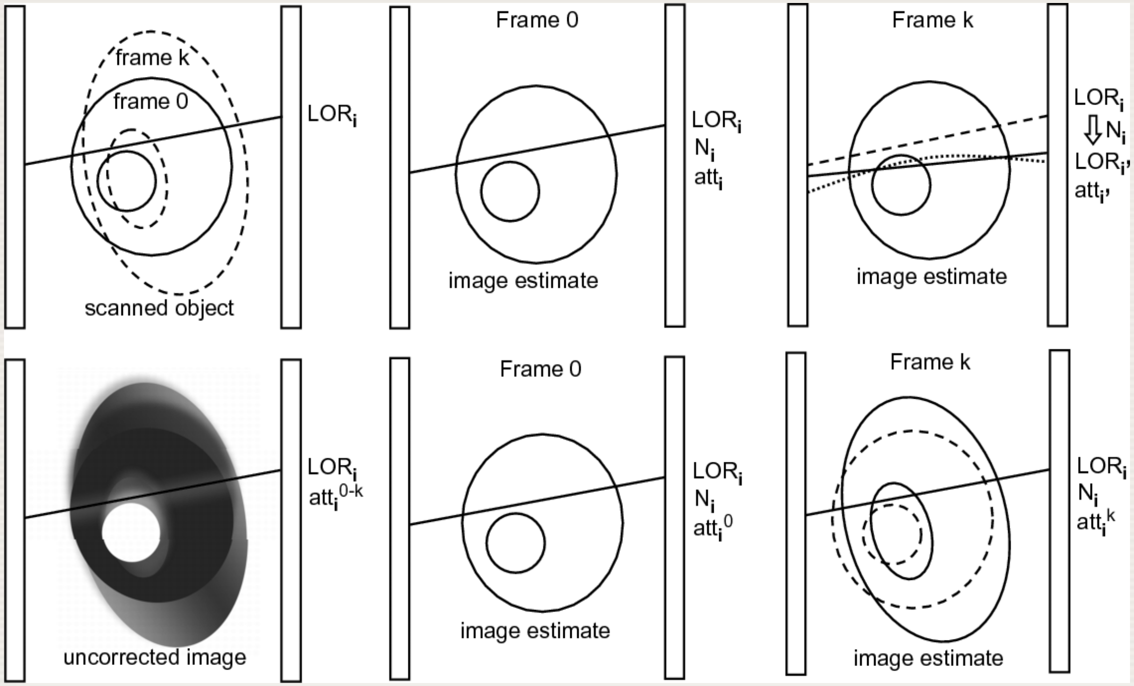


FIG. 13.21. Illustration of motion corrections for events acquired within line of response LOR_i with corresponding normalization N_i and attenuation att_i factors. Left top: positions and shapes of the object in the reference time frame 0 and frame k. Left bottom: illustration of blurring in the reconstruction combining events from all frames without motion correction (attenuation factors are also averaged over the whole range of the frames att_i^{0-k}). Middle column: processing within the reference time frame. Right top: LOR based motion correction for frame k — the LOR_i (dashed line) has to be transformed to the LOR_i (solid line for rigid motion, dotted line for non-rigid motion) which represents the paths that the photons would travel through the reference object if there were no motion. It should be noted that although the LORs are transformed, the normalization factors are used for the crystal pairs (LORs) in which the events were detected (N_i), while the used attenuation factors are for the transformed paths (att_i). Right bottom: image based motion correction, including image morphing of the estimated image from the reference frame (dashed lines) into the given frame (solid line).

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ Reconstruction is then done in a single reference frame using all acquired counts, leading to a better signal to noise ratio in the reconstructed images. Care has to be taken with the detector normalization factors so that the events are normalized using the proper factors (N_i) for the LORs on which they were actually detected (and not into which they were translated).
- ☐ Attenuation factors are obtained on the transformed lines (att_i) through the attenuation image in the reference frame. Care also has to be given to the proper treatment of data LORs with events being translated into, or out of, the detector gaps or detector ends.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- For non-rigid (elastic) motion, which is the case for most of the practical applications, the motion correction procedures become quite involved. There are two basic possibilities.
- The first approach is to derive the transformations of individual paths of events (LORs) from each frame into the reference frame (see Fig. 13.21 (top right, dotted line)). For the non-rigid motion, the transformed paths through the reference object frame are not straight lines anymore, thus leading to very large computational demands for the calculations of the forward and backprojection operations.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ The second, more efficient, approach involves morphing the image estimate (of the reference image) into the frame for which current events (LORs) are being processed (see Fig. 13.21 (bottom right, solid line)).
- ☐ It should be noted that some pre-sorting of the data is considered, so that events from each frame are processed together (using a common image morphing operation).
- ☐ Here, the acquired LORs (LOR_i) and their normalization coefficients (N_i) are directly used without modification.

13.3 ITERATIVE RECONSTRUCTION

13.3.6 Corrections

- ☐ However, the sensitivity matrix still needs to be carefully calculated, taking into consideration update and subset strategy, e.g. including the morphing operation if subset data involve several frames.
- ☐ This is, however, a simpler operation than in the LOR based case since the morphing is done in the image domain.
- ☐ This image based approach is not only more efficient, but also better reflects/models the actual data acquisition process during which the acquired object is being changed (morphed).

IAEA

13.4 NOISE ESTIMATION

13.4.1 Noise propagation in filtered back-projection

- ☐ The pixel variance in an image reconstructed with FBP can be estimated analytically, by propagating the uncorrelated Poisson noise in the data through the reconstruction operation. The FBP algorithm can be written as:

$$\Lambda(x,y)=\int_0^{x} d\phi \int_{-\infty}^{\infty} \mathrm{Y}(x\cos\phi + y\sin\phi - s)\mathrm{h}(s)\mathrm{ds}, \qquad (13.86)$$

where $\mathrm{h}(s)$ is the convolution kernel, combining the inverse Fourier transform of the ramp filter and a possible low-pass filter to suppress the noise.

13.4 NOISE ESTIMATION

13.4.1 Noise propagation in filtered back-projection

- ☐ The variance on the measured sinogram $\mathrm{Y}(s, \phi)$ data equals its expectation $\bar{\mathrm{Y}}(s, \phi)$; the covariance between two different sinogram values $\mathrm{Y}(s, \phi)$ and $\mathrm{Y}(s', \phi')$ is zero. Consequently, the covariance between two reconstructed pixel values $\Lambda(x, y)$ and $\Lambda(x', y')$ equals:

$$\mathrm{covar}(\Lambda(x,y),\Lambda(x',y')) = \int_0^x d\phi \int_{-\infty}^{\infty} \bar{\mathrm{Y}}(x\cos\phi + y\sin\phi - s)ds$$
$$\mathrm{h}(s)\mathrm{h}(s+(x'-x)\cos\phi+(y'-y)\sin\phi) \quad (13.87)$$

- ☐ This integral is non-zero for almost all pairs of pixels.
  - As $\mathrm{h}(s)$ is a high-pass filter, neighbouring reconstruction pixels tend to have fairly strong negative correlations. The correlation decreases with increasing distance between (x, y) and (x', y').

IAEA

13.4 NOISE ESTIMATION

13.4.1 Noise propagation in filtered back-projection

- ❑ The variance is obtained by setting $x = x'$ and $y = y'$, which produces:

$$\mathrm{var}(\Lambda(x, y)) = \int_0^x d\phi \int_{-\infty}^{\infty} \bar{\mathrm{Y}}(x\cos\phi + y\sin\phi - s)\left|\mathrm{h}(s)\right|^2 ds \qquad (13.88)$$

13.4 NOISE ESTIMATION

13.4.1 Noise propagation in filtered back-projection

- Figure 13.22 shows the variance image of the FBP reconstruction of a simulated PET sinogram of a heart phantom.

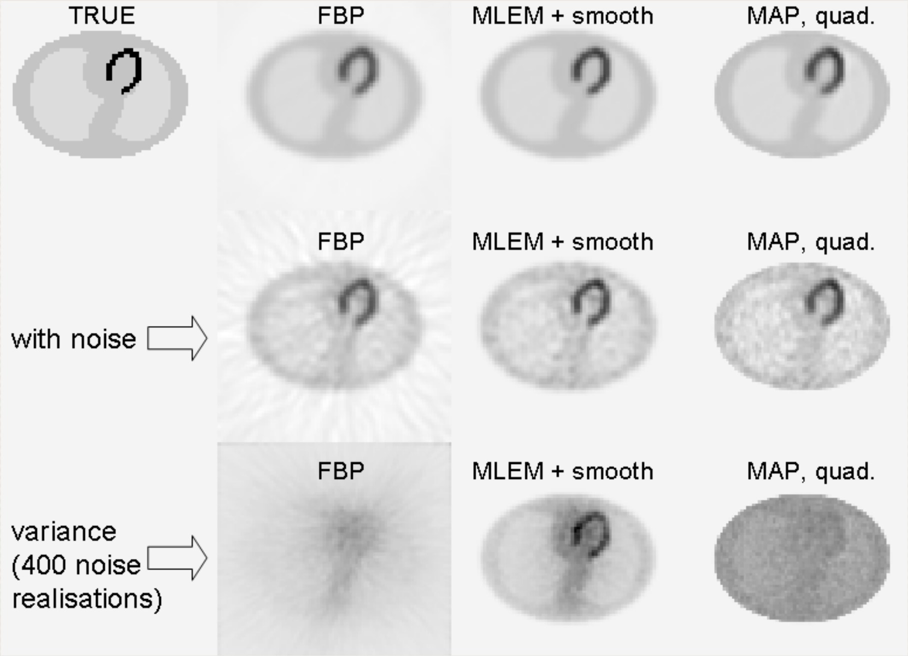


FIG. 13.22. Simulated PET reconstructions of a heart phantom. Reconstructions were done with filtered back-projection (FBP), maximum-likelihood expectation-maximization (MLEM) with Gaussian post-smoothing and with maximum a posteriori (MAP) using a quadratic prior. For each algorithm, a noise-free and a noisy reconstruction are shown, and also the pixel variance obtained from 400 independent Poisson noise realizations on the simulated PET data. All reconstructions (first two rows) are shown on the same grey value scale. A second scale was used to display the three variance images. The noisy filtered back-projection (FBP) image contains negative pixels (displayed in white with this scale). The noise creates streaks that extend to the edge of the image. As a result, the variance is non-zero in the entire image.

13.4 NOISE ESTIMATION

13.4.2 Noise propagation in ML-EM

- The noise analysis of MLEM (and MAP) reconstruction is more complicated than that for FBP because these algorithms are non-linear. However, the MLEM algorithm has some similarity with the WLS algorithm, which can be described with matrix operations. The WLS reconstruction was described previously; Eq. (13.45) is repeated here for convenience (the additive term was assumed to be zero for simplicity):

$$\boldsymbol{\lambda} = (\boldsymbol{A}'\boldsymbol{C}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{C}_y^{-1}\boldsymbol{y} \qquad (13.89)$$

- $\boldsymbol{C}_y$ is the covariance of the data, defined as $\boldsymbol{C}_y = E(\boldsymbol{y} - \overline{\boldsymbol{y}})(\boldsymbol{y} - \overline{\boldsymbol{y}})'$, where E denotes the expectation, and $\overline{\boldsymbol{y}}$ is the expectation of $\boldsymbol{y}$.

13.4 NOISE ESTIMATION

13.4.2 Noise propagation in ML-EM

- The covariance of the reconstruction is then:

$$\begin{aligned} \boldsymbol{C}_\lambda &= E(\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}})(\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}})' \\ &= (\boldsymbol{A}'\boldsymbol{C}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{C}_y^{-1}E(\boldsymbol{y} - \bar{\boldsymbol{y}})(\boldsymbol{y} - \bar{\boldsymbol{y}})'\boldsymbol{C}_y^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{C}_y^{-1}\boldsymbol{A})^{-1} \qquad (13.90) \\ &= (\boldsymbol{A}'\boldsymbol{C}_y^{-1}\boldsymbol{A})^{-1} \end{aligned}$$

- This matrix gives the covariances between all possible pixel pairs in the image produced by WLS reconstruction. The projection $\boldsymbol{A}$ and back-projection $\boldsymbol{A}'$ have a low pass characteristic. Consequently, the inverse $(\boldsymbol{A}'\boldsymbol{C}_y^{-1}\boldsymbol{A})^{-1}$ acts as a high-pass filter. It follows that neighbouring pixels of WLS reconstructions tend to have strong negative correlations, as is the case with FBP. Owing to this, the MLEM variance decreases rapidly with smoothing.

13.4 NOISE ESTIMATION

13.4.2 Noise propagation in ML-EM

- Figure 13.22 shows mean and noisy reconstructions and variance images of MLEM with Gaussian post-smoothing and MAP with a quadratic prior. For these reconstructions, 16 iterations with 8 subsets were applied. MAP with a quadratic prior produces fairly uniform variance, but with a position dependent resolution. In contrast, post-smoothed MLEM produces fairly uniform spatial resolution, in combination with a non-uniform variance.