5.7 Inverse Filtering

The simplest approach to restoration is direct inverse filtering, where we compute an estimate, $\hat{F}(u,v)$, of the transform of the original image by

$$\hat{F}(u,v) = \frac{G(u,v)}{H(u,v)} \qquad (5.7\text{-}1)$$

Substituting the right side of

$$G(u,v) = H(u,v)F(u,v) + N(u,v) \qquad (5.1\text{-}2)$$

in (5.7-1) yields

$$\hat{F}(u,v) = F(u,v) + \frac{N(u,v)}{H(u,v)}. \qquad (5.7\text{-}2)$$

The bad news is that we cannot recover the undegraded image exactly because $N(u,v)$ is not known.

More bad news is that if the degradation function $H(u,v)$ has zero or very small values, so the second term of (5.7-2) could easily dominate the estimate of $\hat{F}(u,v)$.

One approach to get around the zero or small-value problem is to limit the filter frequencies to values near the origin. As discussed earlier, we know that $H(0,0)$ is usually the highest value of $H(u,v)$ in the frequency domain.

Example 5.11: Inverse filtering

The image in Figure 5.25 (b) was inverse filtered with

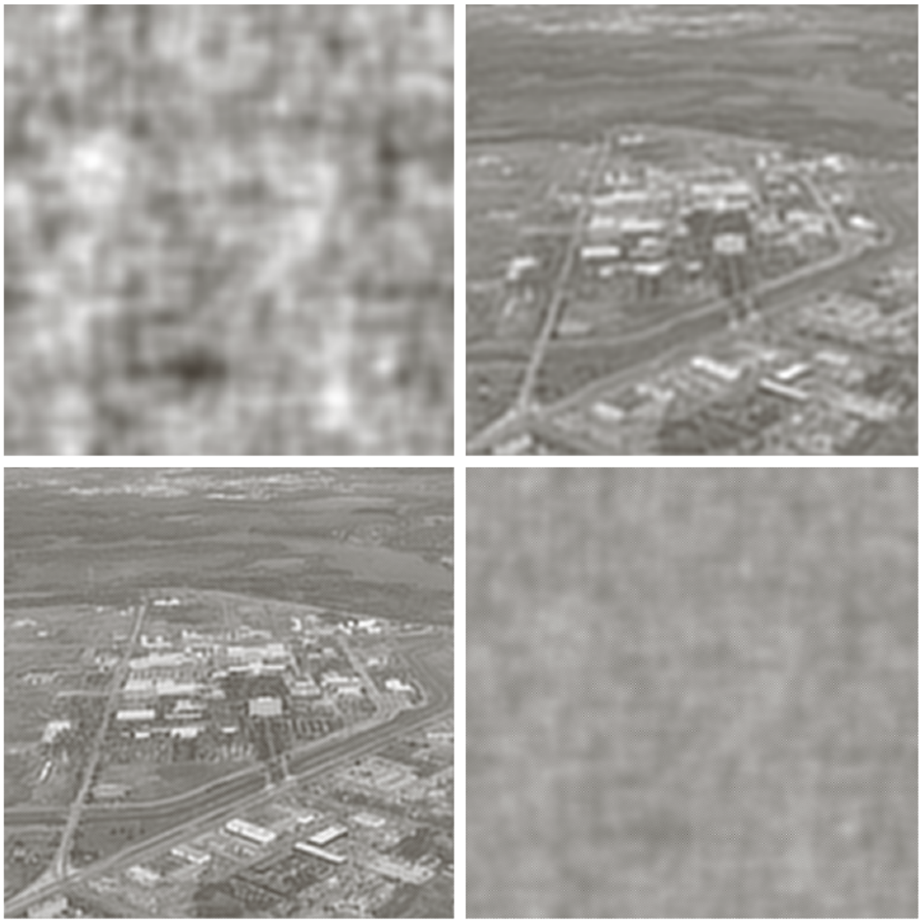
$$\hat{F}(u,v) = \frac{G(u,v)}{H(u,v)} \tag{5.7-1}$$

using the exact inverse of the degradation function that generated that image. That is, the degradation function used was

$$H(u,v) = e^{-k\left[(u-M/2)^2+(v-N/2)^2\right]^{5/6}}$$

with $k = 0.0025$. In this case, $M = N = 480$.

Although a Gaussian-shape function has no zeros and it is not a concern here, the degradation values become so small that the result of full inverse filtering shown in Figure 5.27 (a) is useless.



a b
c d

FIGURE 5.27 Restoring Fig. 5.25(b) with Eq. (5.7-1). (a) Result of using the full filter. (b) Result with H cut off outside a radius of 40; (c) outside a radius of 70; and (d) outside a radius of 85.

Figure 5.27 (b) through (d) show the results of cutting off values of the ratio $G(u,v)/H(u,v)$ outside a radius of 40, 70, and 85, respectively.

Values above 70 started to produce degraded images, and further increases in radius values would produce images that looked more and more like Figure 5.27 (a).

The results in Example 5.11 show the poor performance of direct inverse filtering in general.

5.8 Minimum Mean Square Error (Wiener) Filtering

Here we discuss an approach that incorporates both the degradation function and statistical characteristics of noise into the restoration process.

Considering images and noise as random variables, the objective is to find an estimate $\hat{f}$ of the uncorrupted image f such that the mean square error between them is minimized.

The error measure is given by

$$e^2 = E\left\{(f - \hat{f})^2\right\} \tag{5.8-1}$$

where $E\{\bullet\}$ is the expected value of the argument.

By assuming that

1. the noise and the image are uncorrelated;
2. one or the other has zero mean;
3. the intensity levels in the estimate are a linear function of the levels in the degraded image.

Then, the minimum of the error function in (5.8-1) is given in the frequency domain by the expression

$$\begin{aligned}
\hat{F}(u,v) &= \left[\frac{H^*(u,v)S_f(u,v)}{S_f(u,v)|H(u,v)|^2 + S_\eta(u,v)}\right]G(u,v) \\
&= \left[\frac{H^*(u,v)}{|H(u,v)|^2 + S_\eta(u,v)/S_f(u,v)}\right]G(u,v) \qquad (5.8\text{-}2) \\
&= \left[\frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2 + S_\eta(u,v)/S_f(u,v)}\right]G(u,v)
\end{aligned}$$

The terms in (5.8-2) are as follows:

$\hat{F}(u,v)$ is the frequency domain estimate

$G(u,v)$ is the transform of the degraded image

$H(u,v)$ is the transform of the degradation function

$H^*(u,v)$ is complex conjugate of $H(u,v)$

$|H(u,v)|^2 = H^*(u,v)H(u,v)$

$S_\eta(u,v) = |N(u,v)|^2 =$ power spectrum of the noise

$S_f(u,v) = |F(u,v)|^2 =$ power spectrum of the undegraded image

This result is known as the Wiener filter, which also is commonly referred to as the minimum mean square error filter or the least square error filter.

The Wiener filter does not have the same problem as the inverse filter with zeros in the degradation function, unless the entire denominator is zero for the same value(s) of u and v.

If the noise is zero, then the Wiener filter reduces to the inverse filter.

One of the most important measures is the signal-to-noise ratio, approximated using frequency domain quantities such as

$$SNR = \frac{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}|F(u,v)|^2}{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}|N(u,v)|^2} \qquad (5.8\text{-}3)$$

The mean square error given in statistical form in (5.8-1) can be approximated also in terms a summation involving the original and restored images:

$$MSE = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[f(x,y)-\hat{f}(x,y)\right]^2 \qquad (5.8\text{-}4)$$

If one considers the restored image to be signal and the difference between this image and the original to be noise, we can define a signal-to-noise ratio in the spatial domain as

$$SNR = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\hat{f}(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[f(x,y)-\hat{f}(x,y)\right]^2} \qquad (5.8\text{-}5)$$

The closer f and $\hat{f}$ are, the larger this ratio will be.

If we are dealing with white noise, the spectrum $|N(u,v)|^2$ is a constant, which simplifies things considerably. However, $|F(u,v)|^2$ is usually unknown.

An approach is used frequently when these quantities are not known or cannot be estimated:

$$\hat{F}(u,v) = \left[\frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2+K}\right]G(u,v) \qquad (5.8\text{-}6)$$

where K is a specified constant that is added to all terms of $|H(u,v)|^2$.

Note: White noise is a random signal (or process) with a flat power spectral density. In other words, the signal contains equal power within a fixed bandwidth at any center frequency.

Example 5.12: Comparison of inverse and Wiener filtering

Figure 5.28 shows the advantage of Wiener filtering over direct inverse filtering.

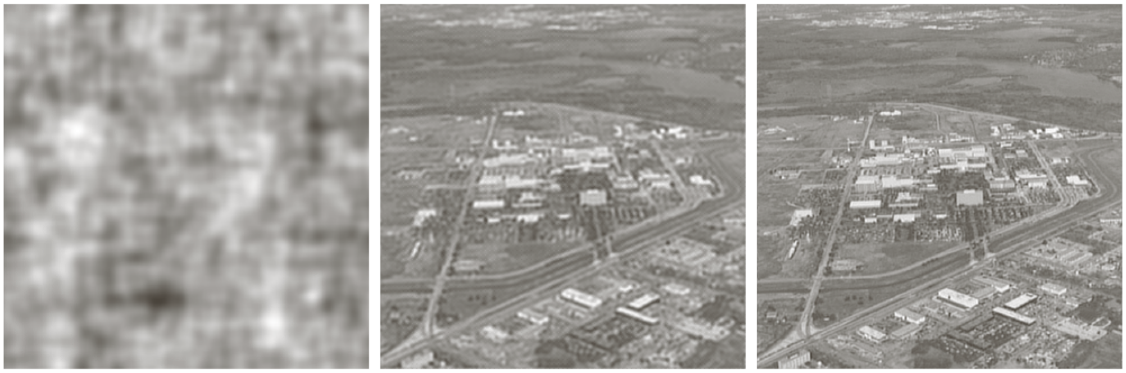


FIGURE 5.28 Comparison of inverse and Wiener filtering. (a) Result of full inverse filtering of Fig. 5.25(b). (b) Radially limited inverse filter result. (c) Wiener filter result.

Figure 5.28 (a) is the full inverse-filtered result from Figure 5.27 (a).

Figure 5.28 (b) is the radially limited inverse result of Figure 5.27 (c).

Figure 5.28 (c) shows the result obtained using

$$\hat{F}(u,v) = \left[\frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + K} \right] G(u,v) \qquad (5.8\text{-}6)$$

with the degradation function

$$H(u,v) = e^{-k\left[(u-M/2)^2 + (v-N/2)^2\right]^{5/6}}$$

used in Example 5.11. The value of K was chosen interactively to yield the best visual result.

By comparing Figure 5.25 (a) and Figure 5.28 (c), we see that the Wiener filter yielded a result very close in appearance to the original image.

Example 5.13: Further comparisons of Wiener filtering

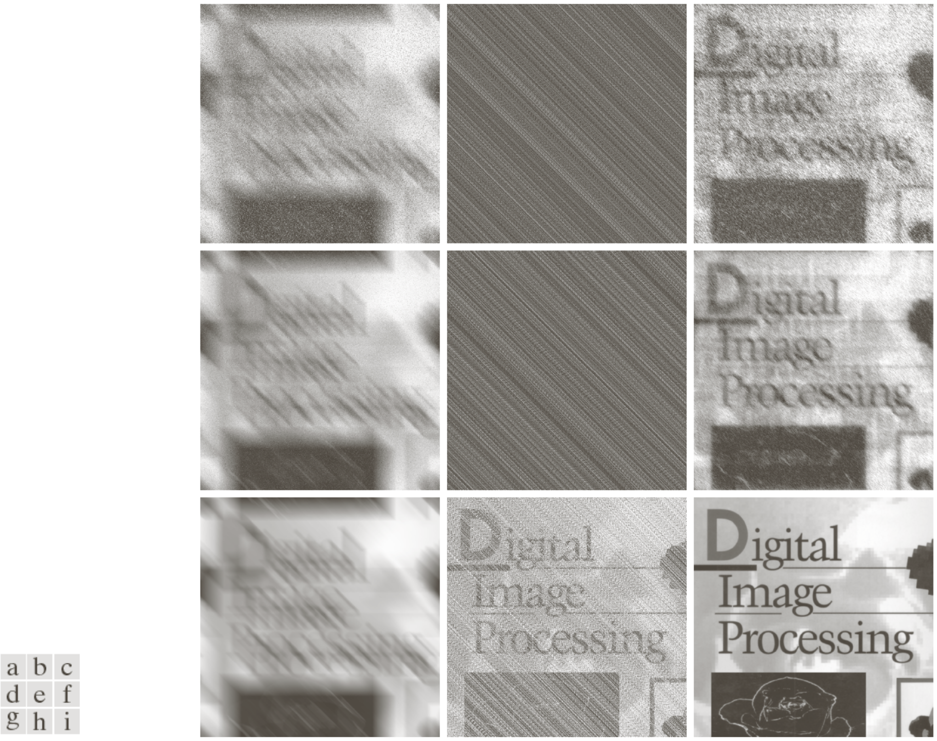


FIGURE 5.29 (a) 8-bit image corrupted by motion blur and additive noise. (b) Result of inverse filtering. (c) Result of Wiener filtering. (d)–(f) Same sequence, but with noise variance one order of magnitude less. (g)–(i) Same sequence, but noise variance reduced by five orders of magnitude from (a). Note in (h) how the deblurred image is quite visible through a “curtain” of noise.

5.10 Geometric Mean Filter

It is possible to generalize the Wiener filter slightly to the so-called geometric mean filter:

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)|^2}\right]^{\alpha} \left[\frac{H^*(u,v)}{|H(u,v)|^2 + \beta\left[\frac{S_\eta(u,v)}{S_f(u,v)}\right]}\right]^{1-\alpha} G(u,v), \qquad (5.10\text{-}1)$$

where α and β are positive real constants.

If $\alpha = 1$, this filter reduces to the inverse filter.

If $\alpha = 0$, the filter becomes the so-called parametric Wiener filter, which reduces to the standard Wiener filter when $\beta = 1$.

If $\alpha = 1/2$, this filter becomes a product of the two quantities raised to the same power, which is the definition of the geometric mean. When $\beta = 1$, the filter is also commonly referred to as the spectrum equalization filter.

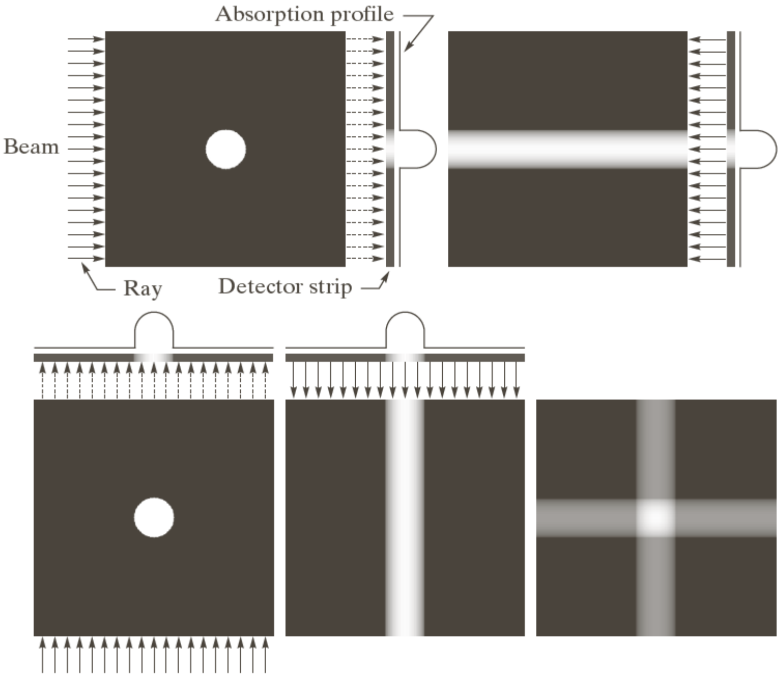
With $\beta = 1$, as α decreases below $1/2$, the filter performance will tend more toward to the inverse filter; as α increases above $1/2$, the filter will behave more like the Wiener filter.

5.11 Image Reconstruction from Projections

In this section, we will examine the problem of reconstructing an image from a series of projections, with a focus on X-ray computed tomography (CT), which is one of the principal applications of digital image processing in medicine.

Introduction

Consider Figure 5.32 (a), which consists of a single object on a uniform background.



a b
c d e

FIGURE 5.32 (a) Flat region showing a simple object, an input parallel beam, and a detector strip. (b) Result of back-projecting the sensed strip data (i.e., the 1-D absorption profile). (c) The beam and detectors rotated by 90°. (d) Back-projection. (e) The sum of (b) and (d). The intensity where the back-projections intersect is twice the intensity of the individual back-projections.

Suppose that we pass a thin, flat beam of X-rays from left to right, and assume that the energy of the beam is absorbed more by the object than by the background. Using a strip of X-ray absorption detectors on the other side will yield the signal, whose amplitude (intensity) is proportional to absorption.

The approach is to project the 1-D signal back across the direction from which the beam came, as Figure 5.32 (b) shows. This approach is called backprojection.

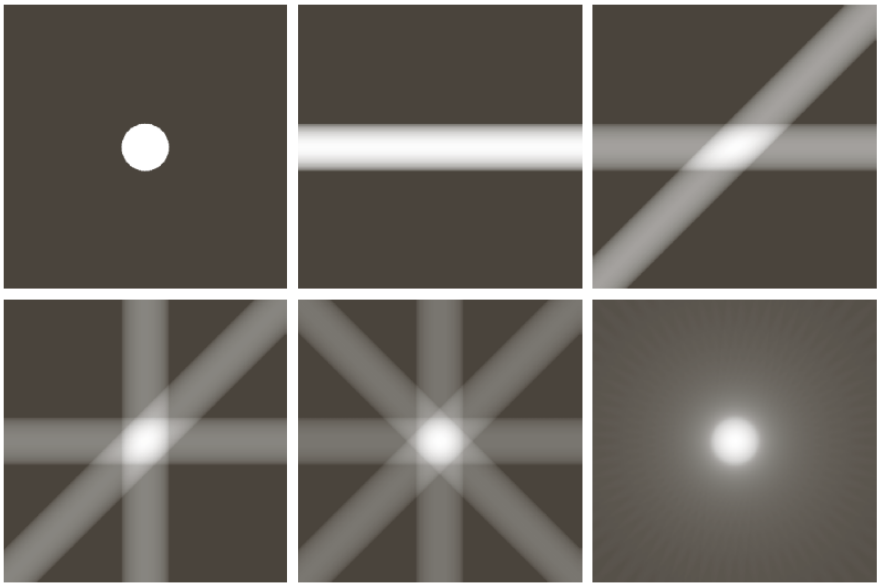
We certainly cannot determine a single object or a multitude of objects along the path of the beam by a single project.

If we rotate the position of the source-detector pair by 90^o and repeat the previous procedure, we will get a backprojection image shown in Figure 5.32 (d). Adding this result to Figure 5.32 (b) will result an image illustrated in Figure 5.32 (e).

We should be able to learn more about the shape of the object in question by taking more views in the same manner, as shown in Figure 5.33.

a b c
d e f

FIGURE 5.33 (a) Same as Fig. 5.32(a). (b)–(e) Reconstruction using 1, 2, 3, and 4 backprojections 45° apart. (f) Reconstruction with 32 backprojections 5.625° apart (note the blurring).



As the number of projections increases, the strength of non-intersecting backprojects deceases relative to the strength of regions in which multiple backprojects intersect.

Figure 5.33 (f) shows the result formed from 32 projections.

The reconstructed image seems to be a reasonably good approximation to the shape of the original object. However, the image is blurred by a “halo” effect, which shows a “star” in Figure 5.33 (e). As the number of views increases, the shape of the “halo” becomes circular, as shown in Figure 5.33 (e).

Blurring in CT reconstruction is an important issue and will be addressed in later discussion.

Since the projections 180^o apart are mirror images of each other, we only need to consider angle increments halfway around a circle in order to generate all the projects required for reconstruction.

Example 5.16: Backprojection of a simple planar region containing two objects

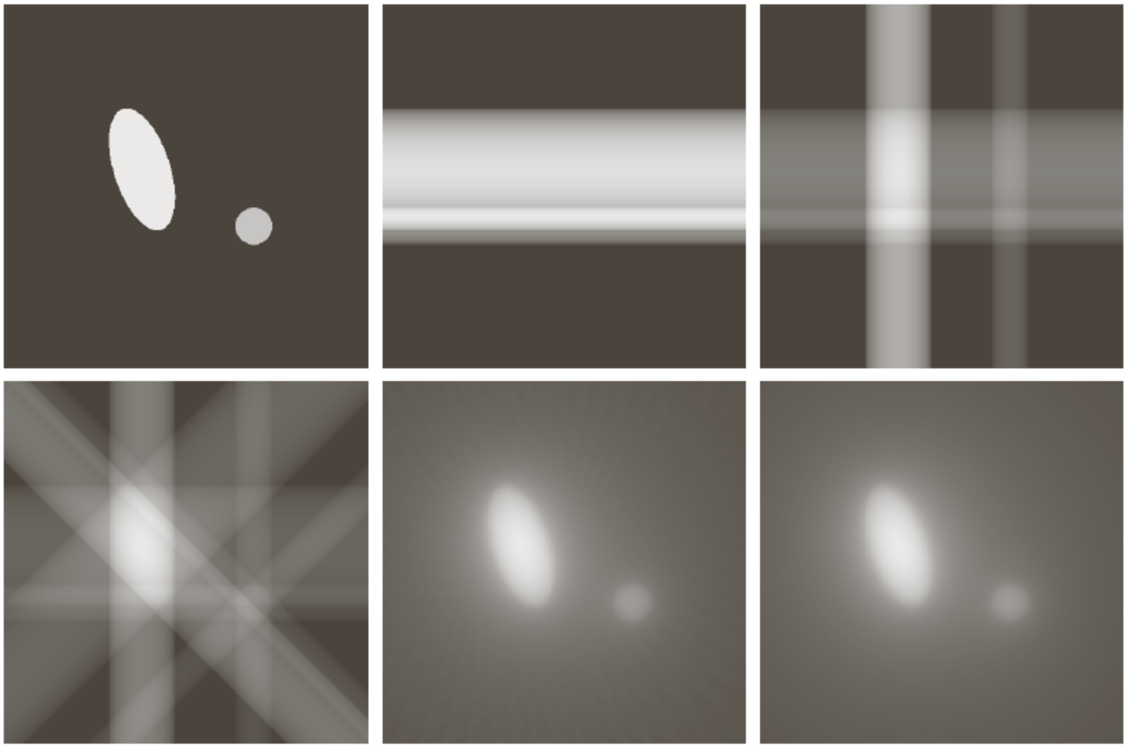


FIGURE 5.34 (a) A region with two objects. (b)–(d) Reconstruction using 1, 2, and 4 backprojections 45° apart. (e) Reconstruction with 32 backprojections 5.625° apart. (f) Reconstruction with 64 backprojections 2.8125° apart.

Principles of Computed Tomography (CT)

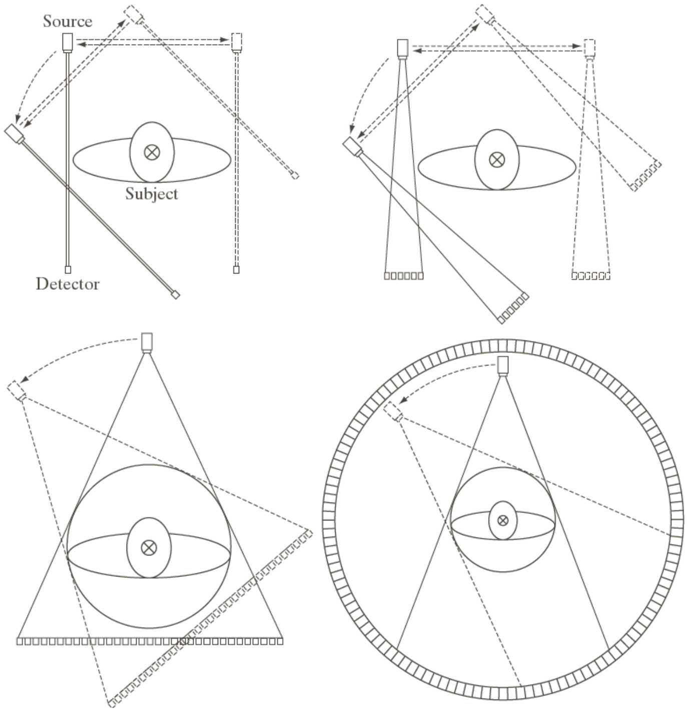
The theoretical foundation of CT dates back to Johann Radon, a mathematician from Vienna who derived a method in 1907 for projecting a 2-D object along parallel rays as part of his work on line integrals. The method is referred as the Radon transform now.

Allan M. Cormack, a physicist at Tufts University, partially “rediscovered” these concepts and applied them to CT. Cormack published his initial findings in 1963 and 1964. He provided the mathematical formulae needed for the reconstruction and built a CT prototype to show his ideas.

Working independently, electrical engineer Godfrey N. Hounsfield and his colleagues at EMI in London formulated a similar solution and built the first medical CT machine.

Cormack and Hounsfield shared the 1979 Nobel Prize in Medicine for their contributions to medical tomography.

Figure 5.35 shows the first four generations of CT scanners.



a b
c d

FIGURE 5.35 Four generations of CT scanners. The dotted arrow lines indicate incremental linear motion. The dotted arrow arcs indicate incremental rotation. The cross-mark on the subject's head indicates linear motion perpendicular to the plane of the paper. The double arrows in (a) and (b) indicate that the source/detector unit is translated and then brought back into its original position.

The fifth-generation (G5) CT scanners eliminate all mechanical motion by employing electron beams controlled electromagnetically.

The sixth-generation (G6) CT scanners rotate the source-detector pair continuously through 360^o, while the patient is moved at a constant speed along the axis perpendicular to the scan.

The seventh-generation (G7) CT scanners (also called multislice CT scanners) use parallel banks of detectors to collect volumetric CT data simultaneously.

Projections and the Radon Transform

A straight line in Cartesian coordinates can be described either by its slope-intercept form

$$y = ax + b\,,$$

or, as in Figure 5.36, by its normal representation

$$x\cos\theta + y\sin\theta = \rho\,. \qquad (5.11\text{-}1)$$

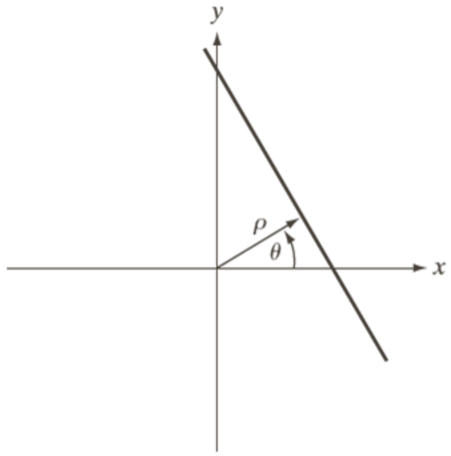


Figure 5.36 Normal representation of a straight line.

The projection of a parallel-ray beam may be modeled by a set of such lines, as shown in Figure 5.37.

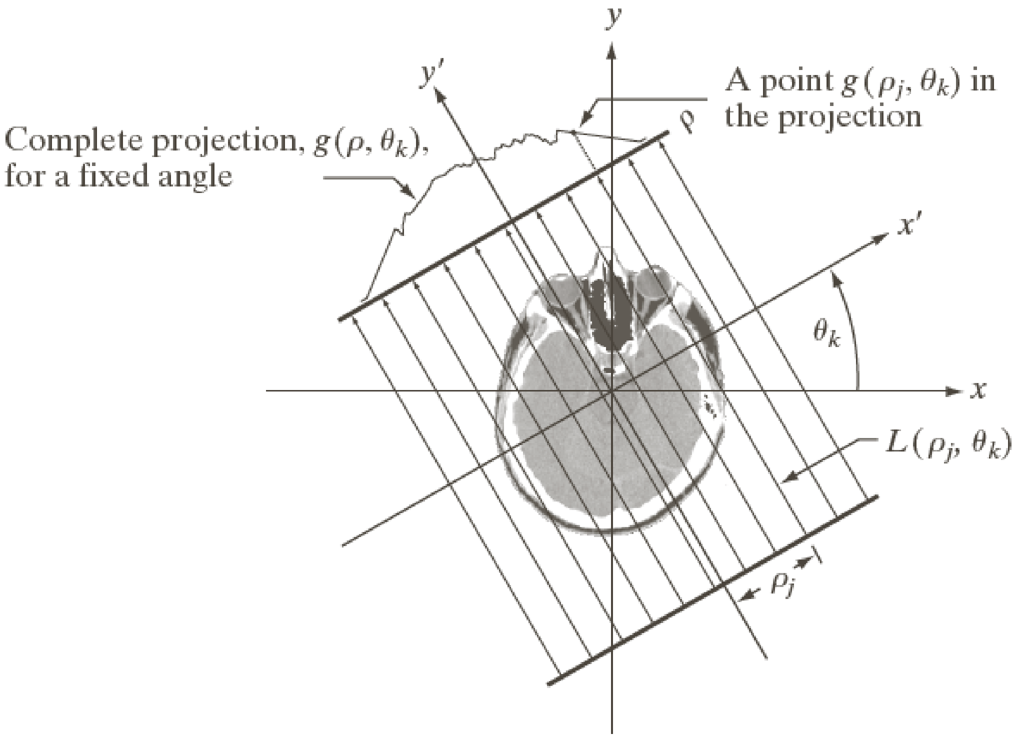

y
y'
A point $g(\rho_j, \theta_k)$ in the projection
Complete projection, $g(\rho, \theta_k)$, for a fixed angle
x'
θ_k
x
$L(\rho_j, \theta_k)$
ρ_j

Figure 5.37 Geometry of a parallel-ray beam.

An arbitrary point in the projection signal is given by the raysum along the line

$$x \cos\theta_k + y \sin\theta_k = \rho_j \,.$$

In the case of continuous, the raysum is a line integral, given by

$$\mathrm{g}(\rho_j, \theta_k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta_k + y\sin\theta_k - \rho_j)dxdy \quad (5.11\text{-}2)$$

Recall the properties of the impulse, δ, the right side of (5.11-2) is zero unless the argument of δ is zero. It indicates that the integral is computed only along the line $x\cos\theta_k + y\sin\theta_k = \rho_j$.

If we consider all values of ρ and θ, (5.11-2) generalizes

$$\mathrm{g}(\rho,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta+y\sin\theta-\rho)dxdy\,. \quad (5.11\text{-}3)$$

The equation (5.11-3) gives the projection of $f(x,y)$ along an arbitrary line in the xy-plane, is called the Radon transform.

The Radon transform is the cornerstone of reconstruction from projections, with CT being its principle application in the field of image processing.

In the discrete case, (5.11-3) becomes

$$\mathrm{g}(\rho,\theta)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\delta\left(x\cos\theta+y\sin\theta-\rho\right), \quad (5.11\text{-}4)$$

where x, y, ρ, and θ are now discrete variables.

If we fix θ and allow ρ to vary, (5.11-4) simply sums the pixels of $f(x,y)$ along the line defined by specified values of these two parameters.

Incrementing through all values of ρ required to span the image (with θ fixed) yields one projection. Changing θ and repeating the same procedure will yield another projection.

Example 5.17: Using the Radon transform to obtain the projection of a circular region.

We want to obtain the Radon transform for the projection of the circular object

$$f(x,y) = \begin{cases} A & x^2 + y^2 \le r^2 \\ 0 & \text{otherwise} \end{cases},$$

where A is a constant and r is the radius of the object. The circular object is shown in Figure 5.38 (a).

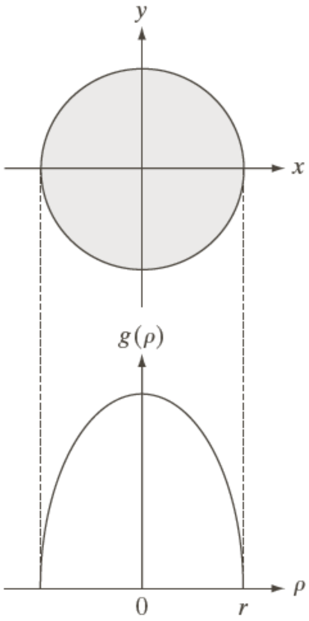


FIGURE 5.38 A disk and a plot of its Radon transform, derived analytically. Here we were able to plot the transform because it depends only on one variable. When g depends on both ρ and θ, the Radon transform becomes an image whose axes are ρ and θ, and the intensity of a pixel is proportional to the value of g at the location of that pixel.

Since the object is circularly symmetric, its projections are the same for all angles, so all we need is to obtain the projection for $\theta = 0^o$. From (5.11-3), we get

$$\begin{aligned} \mathrm{g}(\rho,\theta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x-\rho)dxdy \\ &= \int_{-\infty}^{\infty} f(\rho,y)dy \end{aligned}$$

This is a line integral along the line $L(\rho,0)$.

Note that $\mathrm{g}(\rho,\theta) = 0$ when $|\rho| > r$. When $|\rho| \leq r$, the integral is evaluated from $y = -\sqrt{r^2 - \rho^2}$ to $y = \sqrt{r^2 - \rho^2}$.

Therefore,

$$\mathrm{g}(\rho,\theta) = \int_{-\sqrt{r^2-\rho^2}}^{\sqrt{r^2-\rho^2}} f(\rho,y)dy = \int_{-\sqrt{r^2-\rho^2}}^{\sqrt{r^2-\rho^2}} A\,dy$$

It yields

$$\mathrm{g}(\rho,\theta) = \mathrm{g}(\rho) = \begin{cases} 2A\sqrt{r^2 - \rho^2} & |\rho| \leq r \\ 0 & \text{otherwise} \end{cases}$$

Figure 5.38 (b) shows the result.

$\mathrm{g}(\rho,\theta) = \mathrm{g}(\rho)$ indicates that g is independent of θ because the object is symmetric about the origin.

When the Radon transform, $g(\rho,\theta)$, is displayed as an image with ρ and θ as rectilinear coordinates, the result is called a sinogram, similar in concept to displaying the Fourier spectrum. Like the Fourier spectrum, a sinogram contains the data necessary to reconstruct $f(x,y)$.

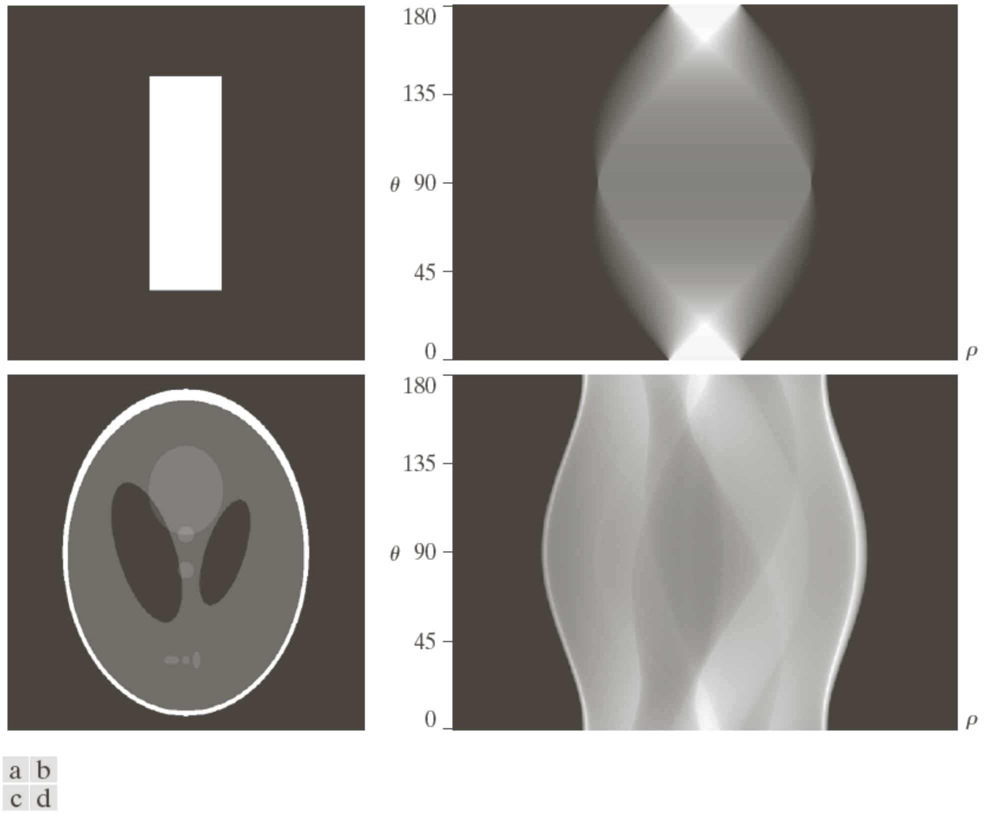


FIGURE 5.39 Two images and their sinograms (Radon transforms). Each row of a sinogram is a projection along the corresponding angle on the vertical axis. Image (c) is called the *Shepp-Logan phantom*. In its original form, the contrast of the phantom is quite low. It is shown enhanced here to facilitate viewing.

Figure 5.39 (b) is the sinogram of the rectangle shown in Figure 5.39 (a).

Figure 5.39 (c) shows an image of the Shepp-Logan phantom, a widely used synthetic image designed to simulate the absorption of major areas of the brain. The sinogram of Figure 5.39 (c) is shown in Figure 5.39 (d).

To obtain a formal expression for a back-projected image from Radon transform, referring to Figure 5.37, we begin with a single point, $g(\rho_j, \theta_k)$, of the complete projection, $g(\rho, \theta_k)$, for a fixed value of rotation, θ_k.

Forming part of an image by back-projecting this single point is simply to copy the line $L(\rho_j, \theta_k)$ onto the image, where the value of each point in that line is $g(\rho_j, \theta_k)$. Repeating this process of all values of ρ_j in the projected signal results

$$f_{\theta_k}(x, y) = g(\rho, \theta_k) = g(x\cos\theta_k + y\sin\theta_k, \theta_k).$$

This equation holds for an arbitrary value of θ_k, therefore, we can write in general that the image formed from a single backprojection obtained at an angle θ is given by

$$f_\theta(x, y) = g(x\cos\theta + y\sin\theta, \theta). \qquad (5.11\text{-}5)$$

We form the final image by integrating over all the back-projected images

$$f(x, y) = \int_0^\pi f_\theta(x, y) d\theta \qquad (5.11\text{-}6)$$

In the discrete case, the integral becomes a sum of all back-projected images:

$$f(x, y) = \sum_{\theta=0}^{\pi} f_\theta(x, y) \qquad (5.11\text{-}7)$$

For example, if 0.5^o increments are being used, the summation is from 0 to 179.5.

A back-projected image formed in this manner is referred to as a laminogram, which is only an approximation to the image from which the projections were generated.

Example 5.18: Obtaining back-projected images from sinograms

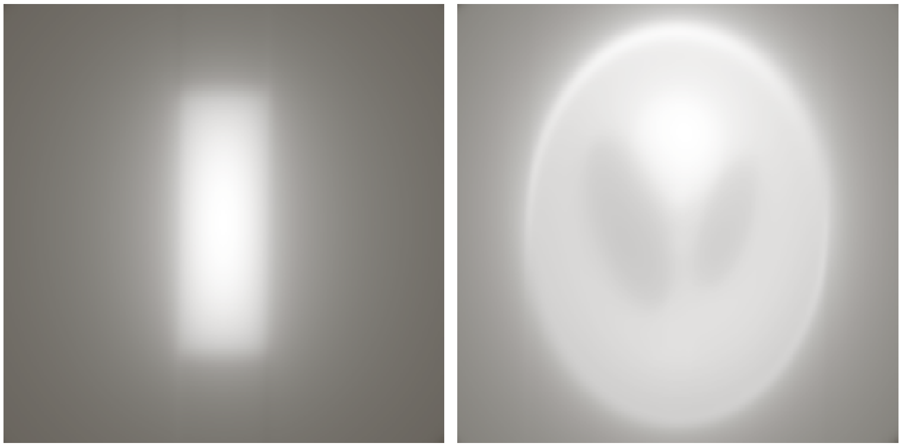
Equation

$$f(x,y) = \sum_{\theta=0}^{\pi} f_\theta\,(x,y) \qquad (5.11\text{-}7)$$

was used to generate the back-projected images in Figure 5.32 through Figure 5.34 from projections obtained with

$$g(\rho,\theta) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\delta\,(x\cos\theta + y\sin\theta - \rho)\,. \qquad (5.11\text{-}4)$$

These equations were also used to generate Figure 5.40 (a) and Figure 5.40 (b), which show the back-projected images corresponding to the sinograms in Figure 5.39 (b) and Figure 5.39 (d).



a b

FIGURE 5.40 Backprojections of the sinograms in Fig. 5.39.

Note that there is a significant amount of blurring shown in Figure 5.40 (a) and (b). It is obvious that a straight use of Equations (5.11-4) and (5.11-7) will not yield acceptable results.