

COMPUTED TOMOGRAPHY FROM X-RAYS: OLD 2-D RESULTS, NEW 3-D PROBLEMS

KEITH MILLER

Department of Mathematics
University of California
Berkeley, CA 94720-3840
kmiller@math.berkeley.edu

Abstract: We consider old results on 2-D computerized tomography methods and their relevance to new fully 3-D problems. We examine the 2-D filtered back-projection method (FBP) from several perspectives to better understand how it works. Based on that understanding, we question whether stable, reliable reconstruction algorithms can be found for *wide-detector* cone-beam 3-D machines with their resulting large-slant beams. We use a numerical method of “point response function fitting” to compute convolution kernels $H(u, v)$ (for use with back-projection) for a model slant-beam problem. These kernels exhibit disturbing growing and spreading oscillations which would greatly amplify errors in the projection data.

Keywords computerized tomography; fully 3-D reconstruction; filtered back-projection; cone-beam CT

1. INTRODUCTION

This paper is dedicated to the memory of Carlo Pucci—co-advisor on my Ph.D. thesis research, my mentor in several new fields of mathematics, my lifelong friend.

We consider old results in 2-D CT methods and their relevance to new fully 3-D CT problems. In particular, an understanding of how the reliable 2-D filtered back-projection (FBP) method manages to succeed leads the author to question whether stable, reliable reconstruction algorithms can be found for *wide-detector* cone-beam helical-scan 3-D machines with their resulting large-slant beams.

In Sections 2, 3, 5 we examine the 2-D FBP method from several perspectives in order to gain a more intuitive understanding of how it works. In Section 4 we introduce 3-D planar-fan-beam and 3-D cone-beam machines and discuss, from an intuitive point of view, the problem of slant-beams with the latter. In Section 6 we use a numerical method of “point response

function fitting” (PRFF) to investigate a 3-D slant-beam model problem (with constant slant $\beta > 0$) which models certain aspects of a wide-detector cone-beam machine. With it we compute a convolution kernel $H(u, v)$ which, under convolution with the projection data and back-projection, yields a “ δ -blurred” reconstruction having our desired point response function (a Gaussian of radius δ). The computed kernels $H(u, v)$, for $\beta = .10$ or $.25$, show strongly growing and widely spreading oscillations (absent in the zero slant $\beta = 0$ case) which would *greatly* amplify errors in the projection data.

I am greatly indebted to the gracious hospitality of my hosts at Oxford University Computing Laboratory, where this work was done and presented in colloquium during a sabbatical in February–March 2002.

2. 2-D PARALLEL BEAM TOMOGRAPHY

One has an unknown “radiographic density function” $f(x_1, x_2)$ with support in the unit disk D_1 , as shown in Figure 1; for example, a cross-section of a human head. On the line at angle θ from the positive x_2 axis and a directed distance t from the origin $(0, 0)$ one has an X-ray source and a detector and one measures the beam attenuations $I(t, \theta)/I_0$, where I_0 is the unobstructed intensity and $I(t, \theta)$ is the intensity after passing through the object. Thus, one measures the beam projections

$$(A_\theta f)(t) = \int_{\text{beam}} f ds = g(z, \theta) = -\log(I(t, \theta)/I_0). \quad (2.1)$$

These are the line integrals of f along the beams, for $|t| \leq 1$, $|\theta| \leq \pi/2$.

First generation machines, in the early 1970’s, used this geometry and recorded the beam densities $g(t_i, \theta_j)$ for approximately $300 \times 360 = 108,000$ equally spaced beams. One attempts to reconstruct the gray levels $f(x_1, x_2)$ at the centers of a perhaps 256×256 matrix of pixels, for TV display.

Some deceptive nonuniqueness results by Keenan Smith led me into working on this problem. Smith’s result (from his AMS talk in Monterey, California, in 1975, and from [11] in 1977) was the following.

THEOREM. *“A finite number of projections tells you absolutely nothing”. Given any function $g(x_1, x_2)$ on a compact subset of the disk D_1 (say $D_{1/2}$) one can extend it to a function $f(x_1, x_2)$ on D_1 whose projections are all zero.*

Thus, one can add a tumor on, or leave a tumor out, and still fit exactly the same projection data! Smith had some very convincing pictures.

This therefore *seemed* to be an *extremely* ill-posed problem. It was Carlo Pucci who pioneered the use of a prescribed global bound to stabilize ill-posed problems, and it was Carlo who gave me the initial insight for the

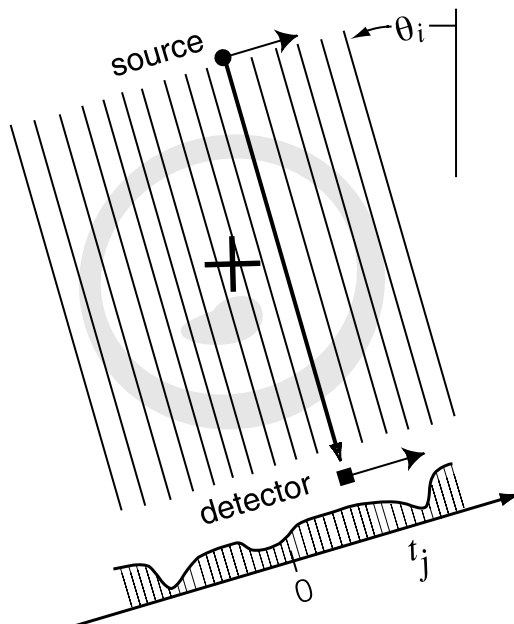


FIGURE 1. The 2-D parallel beam geometry. Projection data are $g(t, \theta) = -\log$ of the measured beam attenuations $I(t, \theta)/I_0$.

Method of Partial Eigenfunction Expansion of my thesis, [4] and [5], a general method which is “optimal but for a factor of at most two”. In 1977 I applied the method to this CT problem, assuming a prescribed bound of E on f and a bound of ϵ on the data error, i.e.,

$$\|f\| \leq E, \quad \|Af - g\| \leq \epsilon. \quad (2.2)$$

The first norm is weighted by $\exp((x_1^2 + x_2^2)/2)$ and the second norm, for the composite projection operator $A = (A_{\theta_1}, \dots, A_{\theta_N})$, is weighted by $\exp(t^2/2)$. The orthogonal eigenfunctions of $A^T A$, of course, turn out to involve Hermite polynomials in x_1 and x_2 . Because this problem is ill-posed, we attempt to reconstruct not f itself, but instead $\rho * f$, the “mollification” or “blurring” of f with a “point response function” ρ of radius δ (for example, the Gaussian kernel of (3.4)). See the AMS abstract [7].

Thinking that there might be real financial gain in this high-powered medical field, I privately hired a talented graduate student, Alan Robbins, to program the method. Initial results showed that it was taking many many eigenfunctions to get any resolution at all. Then friends in the radiology department at U. di Firenze in 1977 showed me their new “color” fan-beam CAT scanner. Such resolution! The problem was not very ill-posed at all! I would need $\approx 40,000$ of my eigenfunctions to get such

resolution. The engineers seemed to have solved the problem empirically—*despite* Smith’s doomsday theorem.

I turned therefore to analysis of the engineers’ “Convolution” or “Filtered Back Projection” (FBP) method. We review now how that method works for the 2-D parallel beam problem.

3. FOURIER TRANSFORMS AND THE FBP METHOD

For a function f of the 2-D variable $x = (x_1, x_2)$ on the plane R^2 we have its Fourier transform \hat{f} , a function of the 2-D transform variable $\xi = (\xi_1, \xi_2)$

$$\hat{f}(\xi) = (2\pi)^{-1} \iint e^{-ix \cdot \xi} f(x) dx. \quad (3.1)$$

We have the inversion formula for $f(x)$, and that the transform of the convolution $\rho * f$ is the product of the convolutions

$$f(x) = (2\pi)^{-1} \iint e^{ix \cdot \xi} \hat{f}(\xi) d\xi, \quad (3.2)$$

$$(\widehat{\rho * f})(\xi) = (2\pi) \hat{\rho}(\xi) \hat{f}(\xi). \quad (3.3)$$

In particular, for the Gaussian kernel $\rho(x)$ of radius δ and total integral 1 we have that its transform $\hat{\rho}(\xi)$ is a Gaussian of radius $1/\delta$

$$\rho(x) = \frac{1}{2\pi\delta^2} \exp(-|x|^2/2\delta^2), \quad (3.4)$$

$$\hat{\rho}(\xi) = \frac{1}{2\pi} \exp(-\delta^2|\xi|^2/2). \quad (3.5)$$

Let $g(x_1)$ be the projection of $f(x_1, x_2)$ in the $\theta = 0$ direction, i.e.,

$$g(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2. \quad (3.6)$$

Henceforth we follow the common convention of neglecting the pesky factors of 2π , $(2\pi)^{-1/2}$, etc., which pop up in work with the Fourier transform. We take the 1-D transform \hat{g} of this projection g to get

$$\begin{aligned} \hat{g}(\xi_1) &= \int g(x_1) e^{-i\xi_1 x_1} dx_1 = \int \left(\int f(x_1, x_2) dx_2 \right) e^{-i\xi_1 x_1} dx_1 \\ &= \iint f(x_1, x_2) e^{-i(\xi_1 x_1 + 0x_2)} dx_1 dx_2 = \hat{f}(\xi_1, 0). \end{aligned} \quad (3.7)$$

Similarly, letting the ξ_1, ξ_2 transform space have polar coordinates (r, θ) , with $-\infty < r < \infty$ and $|\theta| \leq \pi/2$, we get that $\hat{f}(\xi)$ on the line at angle θ

is given by the 1-D transform of the projection $g(t, \theta)$ of f at angle θ , i.e.,

$$\hat{f}(\xi_1, \xi_2) = \hat{g}(r, \theta). \quad (3.8)$$

It suffices to reconstruct the desired $\rho * f$ at the origin (one can shift the origin later)

$$(\rho * f)(0) = \iint (\widehat{\rho * f})(\xi) d\xi = \iint \hat{\rho}(\xi) \hat{f}(\xi) d\xi. \quad (3.9)$$

We convert this integral to polar coordinates, using (3.8)

$$\begin{aligned} (\rho * f)(0) &= \int_{-\pi/2}^{\pi/2} \left(\int_{-\infty}^{\infty} \hat{\rho}(r) \hat{g}(r, \theta) |r| dr \right) d\theta \\ &= \int_{-\pi/2}^{\pi/2} \left(\int_{-\infty}^{\infty} \hat{\rho}(r) \left(\int_{-\infty}^{\infty} g(t, \theta) e^{-irt} dt \right) |r| dr \right) d\theta \\ &= \int_{-\pi/2}^{\pi/2} \left(\int_{-\infty}^{\infty} H(t) g(t, \theta) dt \right) d\theta = \int_{-\pi/2}^{\pi/2} (H * g_\theta)(0) d\theta, \end{aligned} \quad (3.10)$$

where the convolution *kernel function* H is given by

$$H(t) = \int_{-\infty}^{\infty} |r| \hat{\rho}(r) e^{-irt} dr. \quad (3.11)$$

This kernel function H is shown in Figure 3. It has a positive central lobe of height $\approx \delta^{-2}$ and width $\approx \delta$, two negative side lobes of height $O(\delta^{-2})$ and width $O(\delta)$, and negative side tails which are asymptotic to $-t^{-2}$.

Numerically, one replaces the $\int(\cdot) d\theta$ integral in (3.10) by the trapezoid rule over the equally spaced projection angles θ_k . For the general pixel center (x_1, x_2) one just shifts the origin and *back-projects* from the convolution functions $(H * g_{\theta_k})(t)$. Since f is L^2 bounded, with its support in the unit disk, \hat{f} in (3.9) is very smooth. Also, $\hat{\rho}(r)$ is essentially zero for $r \gg 1/\delta$. But the trapezoid rule is amazingly accurate for smooth 2π periodic functions of θ . In fact, if

$$F(\theta) = \sum_{n=0}^{\infty} (a_n \cos n\theta + b_n \sin n\theta) \quad (3.12)$$

then the error of the trapezoid rule to approximate the integral on $[-\pi, \pi]$, with $\Delta\theta = 2\pi/N$ and nodes at $0, \pm\Delta\theta, \pm2\Delta\theta, \dots$, is

$$\text{error} = 2\pi(a_N + a_{2N} + \dots) \approx 2\pi a_N. \quad (3.13)$$

This is because the error is zero for the sine terms (by antisymmetry) and for the cosine terms with $n = 0$ or n not a multiple of N . In 1978 I derived error bounds, again under the assumptions (2.2), showing that this FBP

method, with midpoint quadrature, is amazingly accurate once one gets $\Delta\theta$ several times smaller than the blurring radius δ .

For small t , say $|t| \leq 3$, the kernel $H(t)$ can be well approximated by a few terms of its Taylor series. For larger t , by successive integrations by parts, one can obtain a highly accurate approximation with a few terms of an asymptotic expansion in the negative even powers t^{-2}, t^{-4}, \dots . In fact, the coefficient for t^{-2} is $-\hat{\rho}(0)$. Since $\hat{\rho}(0)$ is proportional to the total integral of ρ (which is always 1) this means that the *difference* between two such kernels is $O(t^{-4})$. Thus for example, the kernel $H(t)$ for $\rho = \rho_\delta - \rho_{4\delta}$, where these are the Gaussian kernels of radius δ and 4δ , decays so rapidly that it is essentially zero for $|t| \geq 16$. This has implications for *limited region illumination*. One can't obtain the δ -blur $\rho_\delta * f$ by illuminating only a limited region of the object, but one *can* obtain the difference between a sharp δ -blur and a much broader 4δ -blur. The small support of $H_\delta - H_{4\delta}$ also suggests the possibility of *multigrid* methods for computing the convolution $\rho_\delta * f = (\rho_\delta - \rho_{4\delta}) * f + \rho_{4\delta} * f$ since the latter convolution can probably be computed with a coarser resolution.

For a stable reconstruction of $\rho_\delta * f$ one needs a full 180° of projections. One cannot depend on *limited angle illumination*, for example with finely spaced projections from only 90° of θ 's. This is because one can easily discover nontrivial objects $f(x, y)$, with features considerably larger than δ , whose projections from 90° of θ 's (or 135° , or \dots) are nearly zero. Consider, for example,

$$f(x, y) = (a \cos ny + b \sin ny) \exp(-(x^2 + y^2)/2), \quad (3.14)$$

with $\sqrt{a^2 + b^2} = 1$. Its projections from the $\theta = 0$ direction are

$$g(t, 0) = \int_{-\infty}^{\infty} f(t, s) ds = a \cdot \sqrt{2\pi} \exp(-n^2/2) \cdot \exp(-t^2/2). \quad (3.15)$$

On beams with angle $\theta \neq 0$, the line integrals are similar except the wavelength of the sinusoid oscillation is lengthened by the factor $1/\cos\theta$. Thus the $\exp(-n^2/2)$ above gets replaced by $\exp(-(n \cos\theta)^2/2)$, and this is exceedingly small provided $n \cos\theta$ is ≥ 6 , for example. In order to detect such "ghost" oscillations one must view them "*edge-on*", that is, from projection angles of $\approx 90^\circ$.

Second generation 2-D CT machines do not use the parallel-beam (t_j, θ_i) geometry. Instead they use a fan-beam (β_j, α_i) geometry with a single source at rotated angles α_i and a bank of detectors at positions β_j opposite it. Let the source angle α go all the way around, i.e., $-\pi \leq \alpha < \pi$; then we still get all the same line integrals (repeated twice). We want to compute the δ -blur $(\rho * f)$ at a point (x_0, y_0) in the object. Let the fan-beam (α, β) beam have parallel-beam coordinates (θ, t) with respect to (x_0, y_0) as center

(a simple geometrical change of coordinates). We have equispaced (β_j, α_i) with line integral data $\bar{g}(\beta, \alpha) = g(t, \theta)$.

A *first* approach is interpolation or “rebinning” on the equispaced (β, α) data \bar{g} to try to get data g with equispacing in the (t, θ) coordinates. Then one applies the parallel-beam method.

A *second* approach is to use a *change of variables* in the exact $\iint \dots dt d\theta$ formula (3.10) to get an exact $\iint \dots d\beta d\alpha$ formula (with a somewhat altered but still smooth kernel function). Then one applies equispacing quadrature on that new formula with the $\bar{g}(\beta_j, \alpha_i)$ data. Since our integrand has compact support in the β coordinate and 2π periodicity in the α direction this quadrature should be highly accurate once $\Delta\beta$ and $\Delta\alpha$ get sufficiently small (for the given blurring radius δ).

Lakshminarayanan [3] did something like this in 1975, as reported by Shepp [10]. But, it was based on the *exact* Radon inverse transform (not blurred) to get a *singular* integral with respect to $d\beta d\alpha$, which he then smooths out with respect to β by Fourier filtering (uniformly, independent of x_0, y_0, α). This gives an efficient and reliable method, but with a distorted point response function for points far from the center. This is what most fan-beam machines use today, as the “FBP method”.

I continued developing a “Change of Variables” fan-beam code, but with some nonlinear filtering techniques to attempt to limit “streaking” from hard objects. But then I lost my programmer, who returned to graduate school in another department. Other imperatives intervened, and all work stopped. I never published my results (on the Eigenfunction Expansion method, on the error bounds for the parallel-beam FBP method, on the fan-beam Change of Variables approach) although I talked about them extensively in colloquia and conference talks in Firenze, Berkeley, Argentina, ... in '78–'80.

We turn now from old 2-D results to new fully 3-D CT problems.

4. 3-D PLANAR-FAN-BEAM AND CONE-BEAM MACHINES

For fully 3-D imaging, helical *planar-fan-beam* machines operate much as a 2-D fan-beam machine. The patient, with unknown density function $f(x, y, z)$, lies on the z -axis of the machine with a *single* 2π ring of detectors and a single source opposite the detectors, all in a plane perpendicular to the z -axis. Now, however, as the source rotates the patient is moved slowly in the z direction; thus the source appears to trace a helical path around and around the unknown $f(x, y, z)$ object. Provided that the “pitch” h of the helix (the z distance that the helix advances under each 2π rotation) is somewhat less than 2δ (where δ is the blurring radius of the desired 3-D point response function $\rho(x, y, z)$) rather straightforward modifications of

the basic 2-D FBP algorithm should yield reliable 3-D CT image reconstructions. Note that a 3-D object f of the form $p(z)$ times the “ghost” oscillation $f(x, y)$ of (3.14) is essentially “invisible” from 90° (or 135° , or \dots) of the projection angles θ . If h were to be considerably larger than δ , then by choosing $p(z)$ to be a positive “hump” of height 1 with support in $|z| \leq h/2$ we would have a 3-D object f (with $(\rho * f)(0) \approx 1$) which would be essentially invisible in all the projection data as the source spirals around in its helix and illuminates this f only from the “wrong” (i.e., non “edge-on”) projection angles θ .

The problems arise with modern *cone-beam* machines. In an attempt to speed up the imaging, and to limit the X-ray exposure to the patient, manufacturers have developed machines with multiple rings of detectors around the z -axis. There are machines with 8 rings, 16 rings; and even 256 rings are in the works. A schematic is shown in Figure 2; there is a single source and opposite it a 2-dimensional array of detectors (parametrized by u in the z -direction and v in the angular direction wrapping around the z -axis, with $|u| \leq Z$ and $|v| \leq \pi$). Again, the source moves around the unknown $f(x, y, z)$ object in a helix, but because of the multiple rings one attempts to use a “pitch” h which is much greater than δ ; in fact, one attempts h ’s which are a goodly fraction of the axial width $2Z$ of the detector bank.

Now, however, it is *not* the case that the beams which pass through a given point (x_0, y_0, z_0) all lie in a single plane. There are recent algorithms, see, for example, that of Katsevich [2], which do fairly well. But there are *no* algorithms comparable to the 2-D FBP method in reliability. Algorithms found in [8] and [9], for example, tend to be *rebinning* approaches (trying to find planar data sets in the 3-D helical data), or *iterative* approaches (trying to find pixelated $\tilde{f}(x, y, z)$ whose projections approximately fit the data, or trying to find formulae in terms of Radon derivatives and filtering of these singular integrals. All these methods have lots of artifacts and don’t achieve good point response functions.

It seems that the greatest problem probably arises from *slant-beams*, such as that (near the edge of the detectors, with source angle $\theta = 0$ and detector coordinates $v = 0$ and $u \approx Z$) passing through the indicated point in Figure 2. As the source rotates through nearby angles ($\theta \approx 0$) *all* the beams through this point have approximately the same large slant and lie approximately in a cone which is *far* from planar.

Based upon an understanding of how the 2-D reconstruction method manages to succeed, I have strong doubts that highly stable 3-D methods can be devised using data sets from such highly slanted beams. Moreover, it seems to me that for these complicated 3-D geometries we can’t *guess* a solution; we have to go more numerical. I propose a numerical method of

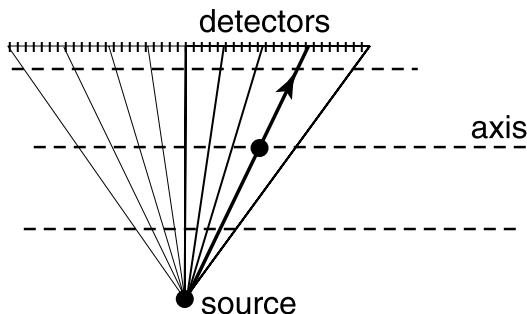


FIGURE 2. Schematic of a helical-scan wide cone-beam machine for 3-D reconstruction. A single source and opposite it a 2-dimensional array of detectors (in rings wrapping around the z -axis). As source rotates, patient is moved slowly in the z direction.

“*point response function fitting*” (PRFF), and use it (in Section 5) to illustrate the working of the 2-D parallel beam FBP method and (in Section 6) to investigate a 3-D *slant-beam model problem* which perhaps throws some light on the perils and possibilities of wide-detector cone-beam machines.

5. PRFF FOR 2-D PARALLEL BEAMS

We want to find a convolution kernel $H(t)$ whose *back-projections*, when θ averaged, fit the desired point response function $\rho(r)$ (for example, the Gaussian of radius δ of (3.4)). This approach is similar to that in the 1979 thesis of Mark Davison [1].

Let (\cdot, \cdot) denote the L^2 inner product of two functions on the unit disk D_1 . Recall that in (3.10) we found the exact formula

$$\begin{aligned} (\rho, f) &= \int_{-\pi/2}^{\pi/2} \left(\int_{-1}^1 H(t)g(t, \theta)dt \right) d\theta \\ &= \int_{-\pi/2}^{\pi/2} (BP_\theta(H), f)d\theta = \left(\int_{-\pi/2}^{\pi/2} BP_\theta(H)d\theta, f \right) \end{aligned} \quad (5.1)$$

where $BP_\theta(H)$ is the “*back-projection*” of $H(t)$ at angle θ (that is, the function whose value is the constant $H(t)$ on each of the (t, θ) beams shown in Figure 1). Thus, we see that this θ -average of these back-projections exactly fits the desired $\rho(r)$.

When we used only N equispaced projection angles $\{\theta_k\}$ we got instead

$$(\rho, f) \approx \left(\sum_{k=1}^N BP_{\theta_k}(H)\Delta\theta, f \right) \equiv (\tilde{\rho}, f) \quad (5.2)$$

where $\tilde{\rho}$ is our approximate point response function. Assuming that our unknown f has the L_2 bound $\|f\|_{D_1} \leq E$, the error (with exact data $g(t, \theta_k)$) satisfies

$$\text{error} = (\rho - \tilde{\rho}, f) \leq \|\rho - \tilde{\rho}\|_{D_1} E. \quad (5.3)$$

Suppose instead that the convolution kernel $H(t)$ is unknown. We can find a good approximate $H(t)$ by finding one whose θ -averaged back-projections $\tilde{\rho}$ well fits the desired ρ . Let's normalize the geometry here by letting $\rho(r) = \exp(-r^2/2)$, the Gaussian of radius 1, and let our object $f(x, y)$ have support in D_R , the disk of radius R , with $R \gg 1$. By rotational symmetry of ρ and $\tilde{\rho}$ it suffices to consider the fit of $\tilde{\rho}(r)$ to $\rho(r)$ on the single ray $(0, r)$, $0 \leq r \leq R$, of the positive x -axis. When the detector is at angle θ from the x -axis the point r on this object ray projects onto the point $t = r \cos \theta$ on the detector. Thus, we want to find a function $H(t)$ (symmetric in t) such that its θ -averaged back-projections

$$\tilde{\rho}(r) \equiv \frac{1}{\pi/2} \int_0^{\pi/2} H(r \cos \theta) d\theta \quad (5.4)$$

is $\approx \rho(r)$ for $0 \leq r \leq R$.

We expand $H(t)$ in cosines

$$H(t) = \sum_{n=0}^N a_n \cos(n\pi t/R); \quad (5.5)$$

hence

$$\tilde{\rho}(r) = \sum_{n=0}^N a_n C_n(r) \approx \rho(r), \quad (5.6)$$

where

$$C_n(r) \equiv \frac{1}{\pi/2} \int_0^{\pi/2} \cos(n\pi(r \cos \theta)/R) d\theta \quad (5.7)$$

is the n -th θ -averaged back-projected cosine basis function. We then take a fine grid of θ and r values, approximate all integrals by the trapezoid rule, then compute the coefficients a_n so as to minimize the L_2 fit in (5.6) (with the 1-D L_2 norm on $[0, R]$ weighted by r so as to give the 2-D L_2 norm on D_R).

Shown in Figure 3 for $0 \leq r, t \leq R = 12$ are our desired $\rho(r) = \exp(-r^2/2)$, the N -th of the θ -averaged basis functions $C_n(r)$, and the convolution kernel $H(t)$ found (which matches the exact kernel of (3.11) extremely closely).

We now are prepared to understand how the 2-D FBP method manages to succeed. Note that $H(t)$ has a positive central peak of height $\approx \delta^{-2}$ and width $\approx \delta$, two negative side lobes of height $O(\delta^{-2})$ and widths $O(\delta)$, and

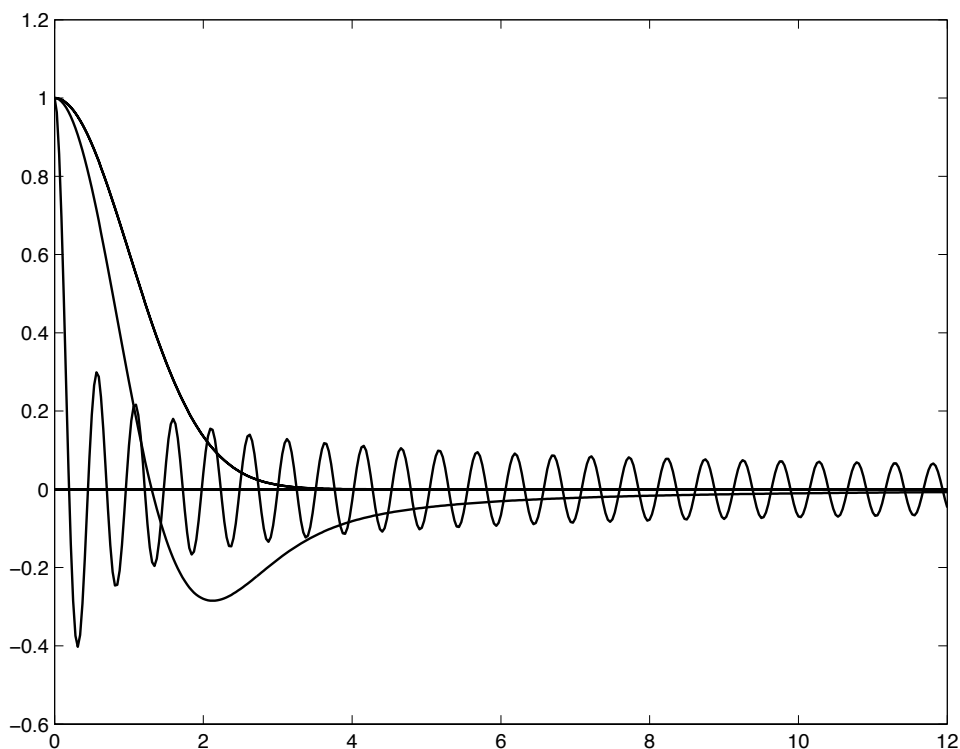


FIGURE 3. The numerical PRFF for 2-D parallel beams: the desired $\rho(r) = \exp(-r^2/2)$, the θ -averaged cosine basis function $C_N(r)$, the convolution kernel $H(t)$ found with its positive central peak and negative side lobes.

negative side tails which are asymptotic to $-t^{-2}$. Figure 4 is a schematic showing the contours, on the unit disk D_1 , of only 4 of the back-projections of this kernel $H(t)$. The positive lobe is indicated by solid lines and the negative side lobes by dotted lines. We take the average of these over all $|\theta| \leq \pi/2$. Near the center the positive lobes add up to fit the positive peak of our desired ρ . Away from the center, where ρ is ≈ 0 , the positive and negative lobes of these neighboring back-projections manage to cancel each other out to \approx zero, provided that one samples with a sufficiently fine $\Delta\theta$. Moreover, it is clearly the lobes of *nearby* θ 's which accomplish most of the cancellation.

For fully 3-D reconstruction this *planar beam* method (with the beams through a point all in a plane perpendicular to the z -axis) would succeed for the same reasons. The desired point response function $\rho(x_1, x_2, z)$ would just be the 2-D $\rho(x_1, x_2)$ of (3.4) multiplied by the Gaussian $(\delta\sqrt{2\pi})^{-1} \exp(-z^2/2\delta^2)$, and the desired convolution kernel $H(t, z)$ for back-projection would just be the 2-D kernel $H(t)$ times that same Gaussian.

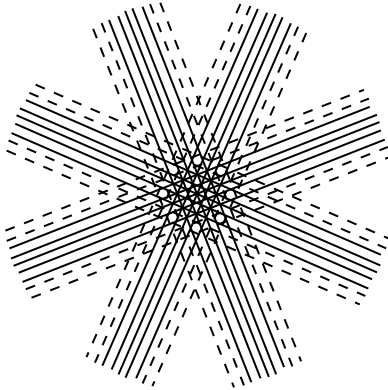


FIGURE 4. Contours of 4 back-projections of the kernel $H(t)$. Taking the average of these (over all θ) the positive peaks (solid contours) and negative side lobes (dotted contours) manage to cancel each other when away from the center.

Now, as the projection angle θ is rotated, the positive and negative lobes of neighboring back-projections of $H(t, z)$ would pass through the same regions of (x_1, x_2, z) space and would manage to cancel each other out to zero.

The case would be different, however, for 3-D reconstruction from *slant-beams* such as those in Figure 2 discussed previously. Now the lobes of neighboring back-projections of a similarly structured $H(t, z)$ kernel would *not* pass through the same regions and would probably not manage to fully cancel each other out.

6. PRFF FOR A 3-D SLANT-BEAM MODEL PROBLEM

We consider a 3-D model problem which may shed some light on the slant-beam difficulties of wide cone-beam machines. This model problem has sufficient symmetries that it is amenable to our numerical PRFF method. Essentially we let the source in Figure 2 go out to infinity along the darkened slant-beam shown.

Consider the geometry shown in Figure 5. The unknown object $f(x_1, x_2, z)$ has its support in the cylinder $-\infty \leq z \leq \infty$, $0 \leq r \leq R$, $r \equiv (x_1^2 + x_2^2)^{1/2}$. The X-ray beams come in with slant $dz/dr = \beta$ from a source at ∞ , pass through the object, and impinge on a rectangular array of detectors with coordinates (u, v) in $[-Z, Z] \times [-R, R]$. Here u is in the axial (z) direction and v is in the radial direction (perpendicular to the page). This source-detector array is then rotated around the z -axis through angles θ , $-\pi/2 \leq \theta < \pi/2$. Thus, we have the beams parametrized by (u, v, θ) and measure the line integrals $g(u, v, \theta)$ of f along these beams. We desire to

reconstruct the “ δ -blurring” of f with the Gaussian 3-D point response function ρ of radius δ . By translation of the origin it suffices to reconstruct this value at the origin, that is, to reconstruct the L_2 inner product (ρ, f) .

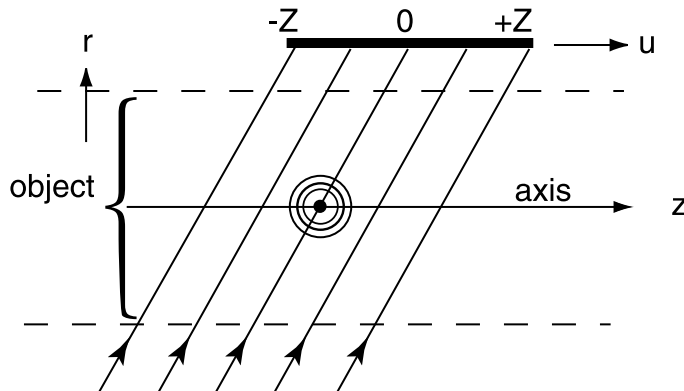


FIGURE 5. 3-D slant-beam model problem. Beams with slant β , from source at ∞ , impinge on rectangular detector with coordinates (u, v) in $[-Z, Z] \times [-R, R]$. Source and detector then rotate around z -axis.

As with the 2-D problem in Section 5, it suffices here to find a *convolution kernel* $H(u, v)$ whose *back-projections* along the beams, when θ averaged, yield a $\tilde{\rho}$ which approximately fits the desired point response function ρ . We want $H(u, v)$ to have its support well within the detector rectangle $[-Z, Z] \times [-R, R]$. Again, we normalize our geometry by setting $\delta = 1$ and letting R be $\gg 1$. Thus, our desired ρ is

$$\begin{aligned} \rho(x_1, x_2, z) &= \exp(-(x_1^2 + x_2^2 + z^2)/2) \\ &= \exp(-(r^2 + z^2)/2) = \rho(z, r). \end{aligned} \quad (6.1)$$

By rotational symmetry of ρ and $\tilde{\rho}$ it suffices to consider the fit to ρ of our θ -averaged $\tilde{\rho}$ on a single object plane within the object cylinder; for example, the plane with 3-D coordinates $(r, 0, z)$ for $-R \leq r \leq R$, $-\infty < z < \infty$. For a given θ , $-\pi/2 \leq \theta < \pi/2$, a point (r, z) in the object plane projects along the beams to the point in the detector plane with coordinates

$$\begin{aligned} u &= u(z, r, \theta) = z - \beta r \sin \theta \\ v &= v(z, r, \theta) = r \cos \theta. \end{aligned} \quad (6.2)$$

This projection and back-projection between (z, r) and (u, v) coordinates, for a given θ , are shown in Figure 6. At angle $\theta \neq 0$ the detector rectangle back-projects onto a slanted elongated quadrilateral on the object plane, as shown. As θ rotates between $-\pi/2$ and $\pi/2$ these back-projections range in the object plane over the hour glass-shaped domain \mathcal{D} shown (we are

interested only in points (z, r) with $|r| \leq R$. However, some points (z, r) in \mathcal{D} will project under (6.2) into points (u, v) with $|u| > Z$, outside the detector rectangle; we must therefore be sure that our approximate $H(u, v)$ decays (smoothly, for the sake of numerical integration) to be essentially zero for $|u| \geq Z$.

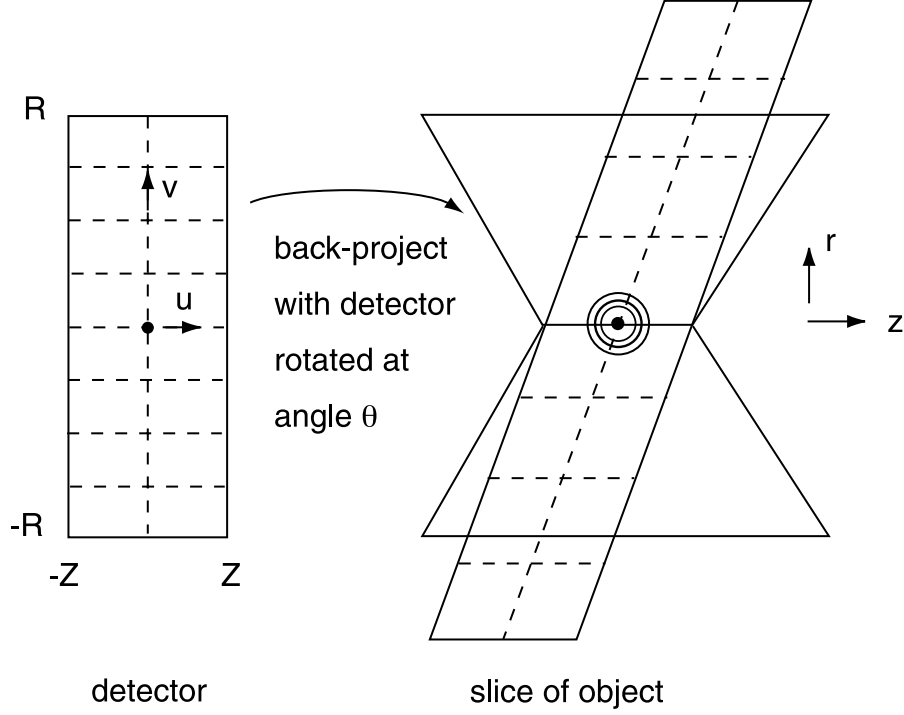


FIGURE 6. Projection and back-projection between detector and slice of object when detector rotated at angle θ .

We therefore want to find a kernel function $H(u, v)$ whose θ -averaged back-projections

$$\tilde{\rho}(z, r) \equiv \frac{1}{\pi} \int_{-\pi/2}^{+\pi/2} H(u(z, r, \theta), v(z, r, \theta)) d\theta \quad (6.3)$$

are $\approx \rho(z, r)$ for all (r, z) in \mathcal{D} . Because $H(u, v)$ will be symmetric in both u and v , it suffices to consider only (z, r) in the positive quadrant \mathcal{D}_+ of \mathcal{D} .

We expand $H(u, v)$ in cosines, with a smooth decay factor of

$$\eta(u) = \exp(-(1.4u/z)^8), \quad (6.4)$$

to ensure that H is essentially zero for $|u| \geq Z$;

$$H(u, v) = \sum_{m,n=0,0}^{M,N} a_{mn} \cos(m\pi u/Z) \cdot \cos(n\pi v/R) \cdot \eta(u). \quad (6.5)$$

Thus,

$$\tilde{\rho}(z, r) = \sum_{m,n=0,0}^{M,N} a_{mn} C_{mn}(z, r) \quad (6.6)$$

where

$$C_{mn}(z, r) \equiv \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos(m\pi u(z, r, \theta)\pi/Z) \cos(n\pi v(z, r, \theta)/R) \eta(u(z, r, \theta)) d\theta \quad (6.7)$$

is the m, n -th θ -averaged back-projected basis function.

We operate in Matlab. We put a fine deformed Cartesian grid on \mathcal{D}_+ . Then, for each grid point $(z, r)_{ij}$ in \mathcal{D}_+ , and for each m, n , we compute the θ -averaged values $C_{mn}(z, r)_{ij}$ using a fine grid of angles θ_k . All integrals, with respect to θ or r or z , are discretized using the trapezoid rule with equispacing (which is extremely accurate once the grid is sufficiently fine).

We then compute the coefficients a_{mn} in (6.6) so as to minimize

$$\psi \equiv \|\rho - \tilde{\rho}\|_{\mathcal{D}_+}^2 + \gamma^2 \left(\sum_{m,n} a_{mn}^2 \right), \quad (6.8)$$

where γ is a small regularization coefficient. The 2-D L_2 norm on \mathcal{D}_+ in (6.8) is, of course, weighted by r so as to convert it into a 3-D norm on the object cylinder. This minimization problem (6.8) converts into a matrix linear least squares problem and we solve the resulting normal equations with their large nonsparse matrix $A^T A + \gamma^2 I$.

The regularization in (6.8) serves two purposes. In the first place, because of the $\eta(u)$ decay factor the $\cos(mu) \cos(nu) \eta(u)$ basis functions in (6.5) are nowhere near orthogonal on the detector rectangle; in fact, they are undoubtedly nearly linearly dependent, and the matrix of the normal equations (without the regularization) would be nearly singular. In the second place, the regularization with a given γ ensures that the resulting $\sum a_{mn}^2$ is as small as possible, for the resulting fit $\|\rho - \tilde{\rho}\|$ to the data. See [6] for a discussion. But $\sum a_{mn}^2$ is proportional to the L^2 norm squared of $H(u, v)/\eta(u)$ on the detector rectangle. By keeping this bounded we ensure both that $H(u, v)$ is decently L_2 bounded on the rectangle and that our $H(u, v)$ decays smoothly to be essentially zero for $|u| \geq Z$ outside the detector rectangle.

Moreover, it is not sufficient that $\|\rho - \tilde{\rho}\|$ be small. It is also important, because of data error, that the norm of H on the data set be decently bounded. Suppose that the projection data $g(u, v, \theta)$ has a root mean square error of ϵ , and that the unknown f has the L^2 bound $\|f\| \leq E$. Then the total error of our approximation to (ρ, f) is bounded by

$$\text{error} \leq \|\rho - \tilde{\rho}\|_{\text{object}} \cdot E + \|H\|_{\text{dataset}} \cdot \epsilon. \quad (6.9)$$

Sometimes, for truly ill-posed problems, one can make $\|\rho - \tilde{\rho}\|$ *small* only by making H *huge* and oscillatory. But then we would have to seek a balance and try to artificially ensure that $H(u, v)$ does not oscillate too wildly with a huge norm.

Our resulting $H(u, v)$ with a *modest* slant, $\beta = .10$, is shown in Figure 7. Here we have used $Z, R = 10, 24$. Shown are 37 “ u -slices”, offset downward, ranging from $v = 0$ at the top to $v = 24$ at the bottom. Note that as v increases $H(u, v)$ has developed strong positive growth along the v -axis and spreading oscillations in the u direction. Clearly we have chosen Z sufficiently large to not interfere with these oscillations. The fit of $\tilde{\rho}$ to ρ is very good, with $\rho - \tilde{\rho}$ on the \mathcal{D}^+ grid having ∞ norm $= 6.7 \times 10^{-5}$, rms norm $= 2.9 \times 10^{-6}$, and the r -weighted rms norm $\|\rho - \tilde{\rho}\|/\|1\| = 1.4 \times 10^{-6}$.

Our resulting $H(u, v)$ with *zero* slant, $\beta = 0$, $Z, R = 10, 24$ is not shown because, as expected, this H is just of the form $H(u, v) = e^{-u^2/2}H(v)$, where the $H(v)$ is the 2-D method’s kernel function $H(t)$ of Figure 3. Its graph thus totally lacks the oscillations seen in the bottom half of Figure 7. The fit of $\tilde{\rho}$ to ρ is very good, with $\rho - \tilde{\rho}$ on the \mathcal{D}_+ grid having ∞ norm $= 8.9 \times 10^{-5}$, rms norm $= 3.6 \times 10^{-6}$, and the r -weighted rms norm $\|\rho - \tilde{\rho}\|/\|1\| = 6.9 \times 10^{-7}$.

Our resulting $H(u, v)$ with a *large* slant, $\beta = .25$, is shown in Figure 8. Here $Z, R = 16, 16$. Shown are 25 “ u -slices”, offset downward, ranging from $v = 0$ at the top to $v = 24$ at the bottom slice. Note that as v increases $H(u, v)$ has developed a huge positive growth along the v -axis and huge widely spreading oscillations in the u direction. It seems that we have chosen Z sufficiently large to not interfere with these spreading oscillations. For this run we have changed the decay factor $\eta(u)$ to $\exp(-(1.2u/Z)^{16})$ instead of (6.4) in order to move the decay cutoff further out toward $u = Z$. The fit of $\tilde{\rho}$ to ρ is good, with $\rho - \tilde{\rho}$ on the \mathcal{D}_+ grid having ∞ norm $= 1.9 \times 10^{-3}$, rms norm $= 2.1 \times 10^{-4}$ and the r -weighted rms norm $\|\rho - \tilde{\rho}\|/\|1\| = 1.5 \times 10^{-4}$.

It is important to know that these oscillations are not an artifact of the numerics or the γ regularization. For *all* our computations we found it sufficient (since the humps and oscillations of $\rho(r, z)$ and $H(u, v)$ turn out to have dimensions ≈ 1) to use $M = 3Z + 1$ and $N = 3R + 1$. The number of lines in the z, r directions for the grid on \mathcal{D}_+ are $nZ, nR = 4M, 2N$

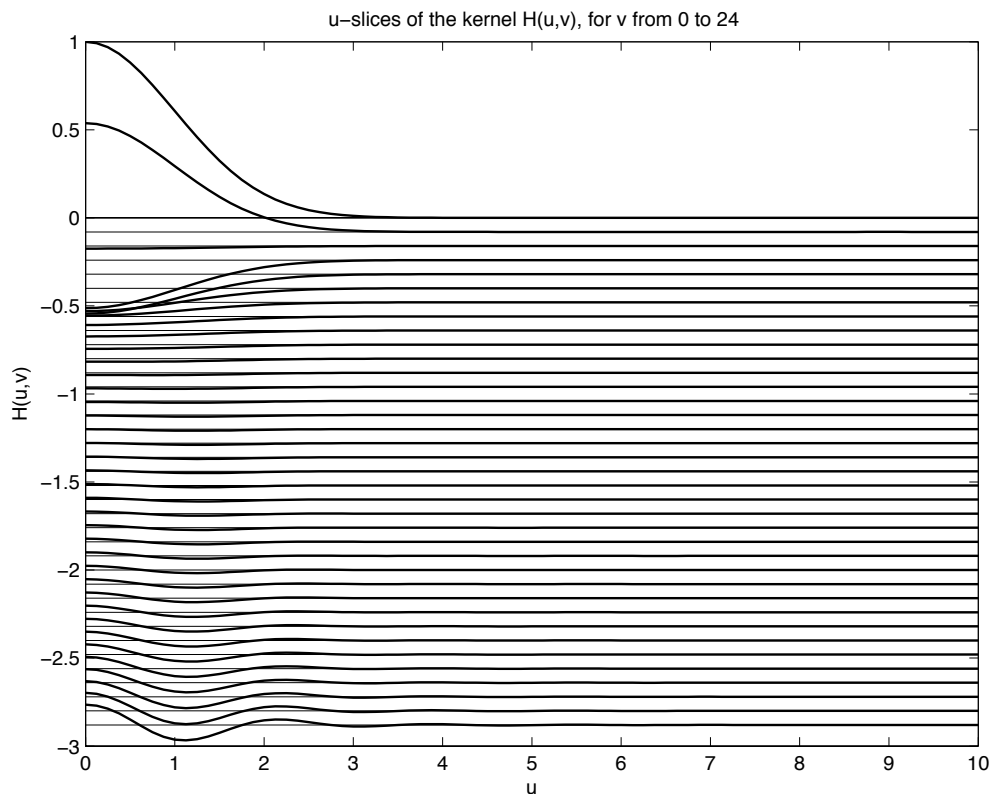


FIGURE 7. $H(u, v)$ for a *modest* slant, $\beta = .10$, with detector dimensions $Z, R = 10, 24$. Shown are “ u -slices” of H , offset downward, from $v = 0$ to $v = 24$. Note the growing and spreading oscillations.

and the number of angles θ_k for the θ -averaging is $nT = 3N$. It is clear that M, N have been chosen sufficiently large to give adequate resolution of $H(u, v)$, because in all 3 cases the cosine coefficients c_{mn} are ≈ 0 for larger m, n .

All three cases (with $\beta = 0, .10$, or, $.25$) were run with the regularization coefficient $\gamma = .10$, but the resulting $H(u, v)$ was essentially unchanged with $\gamma = .01$ or 1.0 . Only with a quite large $\gamma = 10.0$ did one see a substantial change; this large γ did damp the large oscillations and huge norm of H a bit, but only at the expense of largely destroying the desired fit of ρ to $\tilde{\rho}$.

7. DISCUSSION OF THE SLANT-BEAM RESULTS

The strongly growing and widely spreading oscillations of the *slant-beam* kernel $H(u, v)$ found in Figures 7 and 8 with $\beta = .10$ and $.25$ are troubling.

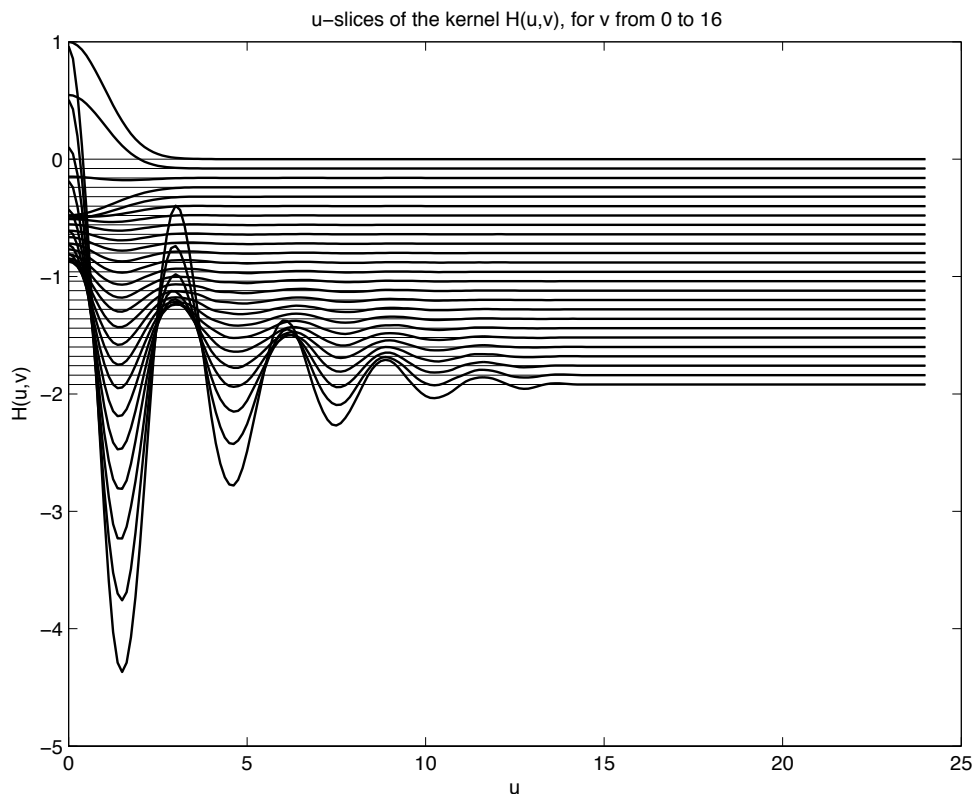


FIGURE 8. $H(u, v)$ for a *large* slant, $\beta = .25$, with detector dimensions $Z, R = 16, 16$. Shown are “ u -slices” of H , offset downward from $v = 0$ to $v = 16$. Note the huge growing and spreading oscillations.

Clearly one wants to operate with a much larger ratio of object size R to blurring resolution δ than the $24/1$ or $16/1$ that we have considered here (but these R 's have filled the 512 Mb RAM of our computer). With a larger R the oscillations for the $\beta = .10$ case would undoubtedly grow to yield a kernel which (like that for $\beta = .25$) is much *much* larger than that for the $\beta = 0$ planar-beam case. But convolution with such a huge kernel would *greatly* amplify errors in the $g(u, v, \theta)$ data, as noted in (6.9).

For much smaller slant-angles β one could perhaps augment the present pure numerics with asymptotics to discover a good approximate formula for this $H(u, v)$ with its then *slowly* growing and spreading oscillations. And *perhaps* one could even discover asymptotic corrections (to the planar-beam helical method discussed in the first paragraph of Section 4) which would be adequate for an actual helical-scan cone-beam machine, provided that the detector is sufficiently narrow in the z direction. But the author has strong doubts that a stable, reliable reconstruction algorithm can be

found for a *single-source wide-cone-beam* machine whose helical pitch h is $\gg \delta$.

The author would instead suggest a system of *multiple*-width sources to go with the multiple-width detectors, leading to 4-D overspecified data for the reconstruction of the 3-D object. Of course, one would have to analyze whether the strength of these multiple sources could be kept sufficiently small so as to not increase the overall X-ray dosage.

REFERENCES

- [1] M.E. Davison, Filtered back-projection methods for reconstructing a function from an arbitrary finite set of its projections, Ph.D. Thesis in Mathematics, University of California, Berkeley, June 1979, 125 pages.
- [2] A. Katsevich, A general scheme for constructing inversion algorithms for cone-beam CT, *International Journal of Mathematics and Mathematical Sciences*, 21 (2003), 1305–1321.
- [3] A.V. Lakshminarayanan, Reconstruction from divergent ray data, SUNY Technical Report Number 92 (1975), Computer Science Department, Buffalo, New York.
- [4] K. Miller, Three circle theorems in partial differential equations and applications to improperly posed problems, *Archive for Rational Mechanics and Analysis* 16 (1964), 126–154.
- [5] K. Miller, An eigenfunction expansion method for problems with over-specified data, *Annali della Scuola Normale Superiore di Pisa*, Series 3, 19 (1965), 397–405.
- [6] K. Miller, Least squares methods for ill-posed problems with a prescribed bound, *SIAM Journal of Numerical Analysis*, 1 (1970), 52–73.
- [7] K. Miller, An optimal method for the X-ray reconstruction problem, preliminary report, *AMS Notices*, January 1978, A-161: paper delivered at AMS meeting, Atlanta, Georgia, January 1978.
- [8] Proceedings of the Sixth International Meeting on Fully Three-Dimensional Reconstruction in Radiology and Nuclear Medicine, Asilomar Conference Grounds, Pacific Grove, CA, October 30–November 2, 2001, R. H. Heusman, Editor. The papers for this meeting were published in *Physics in Medicine and Biology*, Vol. 47, No. 15, 2002.
- [9] Proceedings of the 2003 International Meeting on Fully Three-Dimensional Reconstruction in Radiology and Nuclear Medicine, Saint Malo, France, June 29–July 4, 2003, Yves Bizais, Editor. The papers for this meeting were published in *Physics in Medicine and Biology*, Vol. 49, No. 11, 2004.
- [10] L.A. Shepp and J.B. Kruskal, Computerized tomography: the new medical X-ray technology, *American Mathematical Monthly*, June–July (1978), 420–439.
- [11] K.T. Smith, D.C. Solmon and S.L. Wagner, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, *Bull. AMS*, 83 (1977), 1227–1270.