2

PREPRINT UCRL- 82924

CONF-7906100--1

Lawrence Livermore Laboratory

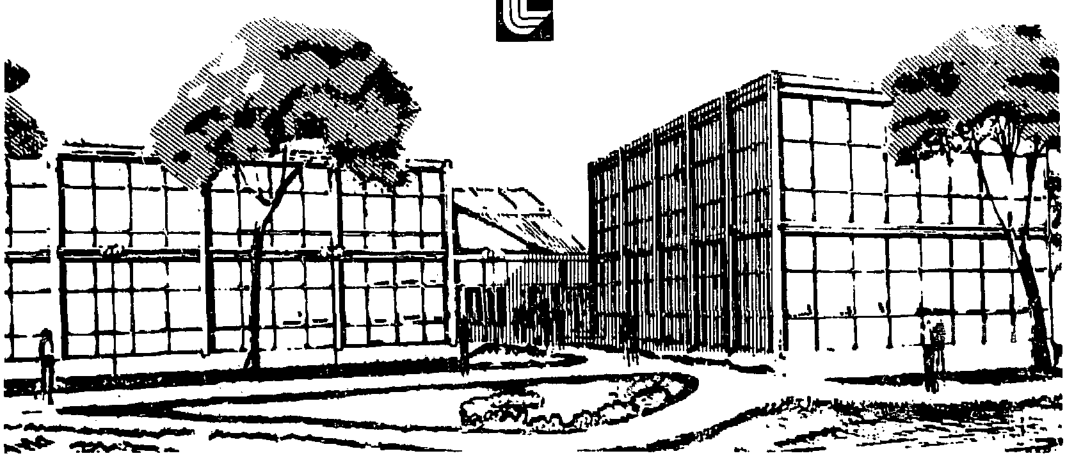
ON OPTIMAL RECONSTRUCTION ANGLES

GRANT O. COOK, JR.
LARRY KNIGHT

JULY 1979

MASTER

This paper was prepared for submission to;
IEEE COMPUTER SOCIETY PROCEEDINGS;
JUNE 20-21, 1979, NEWPORT BEACH, CA.



On Optimal Reconstruction Angles*

Grant Cook, Jr.† and Larry Knight‡

June, 1979

Abstract

The question of optimal projection angles has recently become of interest in the field of reconstruction from projections. Here, studies are concentrated on the $n \times n$ "pixel" space, where iterative algorithms such as ART and direct matrix techniques due to Katz are considered. We attempt to show which angles are best in a Gauss-Markov statistical sense as well as with respect to a function-theoretical error bound. The possibility of making photon intensity a function of angle is also examined. Finally, we study which angles are best to use in an ART-like algorithm. A certain set of unequally spaced angles was found to be preferred in several contexts.

Introduction

The reconstruction space used herein is the pixel space composed of n^2 square regions, denoted by $Z(n)$. Though in the studies of ART-like algorithms other spaces come into play, $Z(n)$ is always the reconstruction space of choice. Our study depends sharply on this decision; in fact, in our search for optimal projection angles, subsets of the angles which correspond to the underlying symmetries of $Z(n)$ are frequently better than any other choice of angles. On the other hand, for one who would take into account the *a priori* known symmetry of an object to be reconstructed, the projection angles chosen should reflect that information. But in general, no such knowledge is available.

Many crucial factors motivate a study of projection angles. Recognizing that the choice of projection angles affects the required detector resolution and the performance of reconstruction algorithms for $Z(n)$, our studies consider stability, numerical sensitivity, accuracy, and, in the iterative schemes, speed of convergence. The set of projection angles is optimized in two related contexts. We consider algorithms which use $Z(n)$ exclusively and are analyzed totally in that space. This involves direct matrix reconstruction techniques, in particular those developed in [14]. Since direct methods may not be feasible on existing computers, ART-like algorithms are studied using projection angle dependent results like those in [10].

I. Some direct matrix methods

Background

The approach to this study is empirical, being founded on theoretical developments due to Katz.[14] Of particular importance in his work is the answer he gives to the question of uniqueness of $Z(n)$ reconstructions. We summarize some of his results as follows. Let p_i and q_i be integers such that

$$\{\theta_i = \tan^{-1}\left(\frac{p_i}{q_i}\right), i=1,m\}$$

†Student guest at the National Magnetic Fusion Energy Computer Center, Lawrence Livermore Laboratory, Livermore, California 94550.

‡Department of Physics, Brigham Young University, Provo, Utah 84602.

*Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore Laboratory under contract number W-7405-ENG-48.

is the set of angles we are using to reconstruct an object in $Z(n)$; m is the number of angles. The detector spacing is intimately related to the chosen angles by the natural choice the $Z(n)$ geometry affords us. It is an integral multiple of (size of a pixel)$/(p_i^2+q_i^2)^{.5}$ With these facts in hand, Katz then proves that a reconstruction in $Z(n)$ is uniquely determined by such projection data if

$$n < 1 + \min\left(\sum_{i=1}^{m} |p_i|, \sum_{i=1}^{m} |q_i| \right)$$

where n is the dimension of $Z(n)$. When this condition is violated, uniqueness ceases to hold and the null space of the projecton transformation becomes non-zero. We will discuss the character of the null space and the role it can play at a later point. But first it is important to clarify these important points. To do so we classify projection angles and detector spacings in a natural way.

A first obvious classification of angles is:
a) Those having rational tangents.
b) Those having irrational tangents.

Class (a) can be further subdivided into two sets. First, there are the angles generated by the Farey series of order n which will henceforth be called "Farey angles of order n". To define them, we must define the term "Farey series":[17] given that p and q are relatively prime integers, the Farey series of order n is that increasing ordered sequence of fractions p/q with $q \leq n$ such that

$$\frac{0}{1} \leq \frac{p}{q} \leq \frac{1}{1}.$$

For example, the Farey series of order 6 is:

$$\frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1}.$$

The angles a Farey series generates are given by

$$0, \frac{\pi}{2}, \frac{\pi}{4}, -\frac{\pi}{4}$$

and

$$\tan^{-1}\left(\frac{p}{q}\right), \tan^{-1}\left(\frac{q}{p}\right), \tan^{-1}\left(\frac{-p}{q}\right), \tan^{-1}\left(\frac{-q}{p}\right)$$

where $\tan^{-1}(1/0)=\pi/2$, and $0 \leq p \leq q$. The other subdivision of the class of angles which have rational tangents consists of all such angles which are not Farey angles. This distinction is important when we come to consider naturally-induced projection data.

The most commonly used projection data in reconstruction algorithms comes from angles of class (b). Equally-spaced detectors are normally employed, with some researchers varying the spacing as a function of angle.[9] A second class of projection data is that generated by the projected images of the vertices of the pixels, with the projected points being the detector boundaries. This is graphically illustrated in Figure 1 for projection information at an angle of 36°. Clearly, the detectors are not equally-spaced, although there is symmetry about the central detector. In contrast, there are angles which do generate equally-spaced projection data. They are in fact the 8 angles generated by the Farey series of order 2. So, another subclassification has arisen. But our classification of projection data is not complete, for those angles generated by Farey series order $n+1$ have projection data which is equally-spaced on the interior detectors. That is, there are outside detectors which have widths which are integral multiples of the minimum detector width at that angle. However, the interior detectors do have equal widths as opposed to the situation for non-Farey angles, so this case is only slightly different from that of completely equally-spaced projection data. Figure 2 illustrates the completely equally-spaced phenomenon for $\tan^{-1}(1/2)$ and Figure 3 shows the "edge" effect for $\tan^{-1}(2/5)$. The reason for larger detector spacings on the edge is that the corner pixels lack the nearby pixels which would otherwise provide vertices to

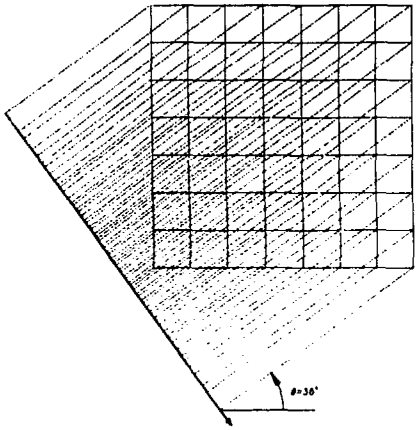


Figure 1

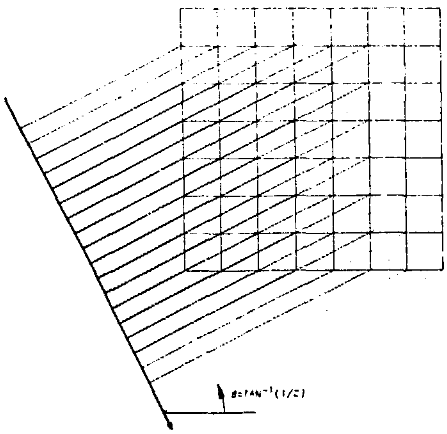


Figure 2

subdivide the larger detectors. Because this kind of projection data is so closely linked to the projection angle in the $Z(n)$ structure, we call it "naturally-induced" projection data. On the basis of these classifications, we can now define the angles of minimum required resolution. They are the m angles in the Farey angles of order n which have the largest detector spacings, or equivalently, those angles which require the least resolution for naturally-induced projection data.[14]

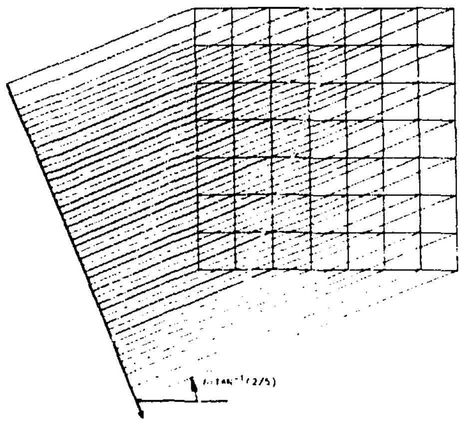


Figure 3

One direct method

We proceed now to sketch Katz's method of implementing this information into a reconstruction algorithm. First, he casts the $Z(n)$ reconstruction problem in the form UA=W. Here A is an $n^2 \cdot 1$ vector which is an element of $Z(n)$; i.e., the reconstruction space we have chosen. W is an $l \cdot 1$ vector containing integrals of the projection data. In particular, 1) its elements are really strip integrals, 2) imply the use of detectors of finite extent, and as such 3) represent some sort of an average. U is an $l \cdot n^2$ matrix which describes the projection data, W, for any element of $Z(n)$ (via matrix multiplication); U depends on a given choice of angles. It is actually a stacked matrix formed from the m different U_{θ_i}, one for each projection angle, θ_i, where the row dimension (the number of detectors at that angle) depends implicitly on θ_i and equals

$$n^2+2n-(n-|p_i|+1)\cdot(n-|q_i|+1)$$

This formula is easily obtained by noting that the number of projected images of vertices in $Z(n)$ is $(n+1)^2$ and that the number of projected images which are duplicates of already existing projections is $(n-|p_i|+1)\cdot(n-|q_i|+1)$ for angles in the Farey series of order n. With a little algebra, the formula can be rewritten in the form

$$(n+1)\cdot(|p_i|+|q_i|)-|p_i|\cdot|q_i|-1$$

For Farey angles of order 2 and angles with rational tangents having $|p_i|$ or $|q_i| = 1$, this number is $n\cdot(|p_i|+|q_i|)$. For all angles outside the Farey angles of order n, the number of duplicated projected images of vertices is zero, so the row dimension is n^2+2n.

The computation of the geometric matrix U is a very difficult matter. Because of this fact, precautions were taken to assure numerical accuracy in its computation. For example, when an angle with a rational tangent occurs, the elements of U_{θ_i} are known to be rational, so advantage is taken of this fact by doing the computations in exact integer arithmetic. However, when angles with irrational tangents are encountered, approximations are necessarily introduced at this step because of the available numerical procedures. Either we can compute the $n^2\cdot(n^2+2n)$ matrix via direct area intersection calculation of the strip integrals, or we can attempt to make the magnitude of p and q satisfying $\tan\theta \cong p/q$ large. The first choice is very difficult to program and has many difficult numerical traps to overcome, such as occur when only a small corner of a pixel lies in the integral's path. If properly done, this is a good but approximate way to compute U_{θ_i}. In the second method, increasing the magnitude of p and q to very large integers would be necessary to come closer and closer to equality in this relation (equality already holds for θ rational). However, considerations of time required for this last computation also restrict the largest magnitude of p and q we can accept, so our subroutines now use the first method for computing U_{θ_i} when $\tan\theta_i$ is irrational.

Since U is seldom square, to solve $UA=W$ some sort of a least squares problem must be solved. Therefore, a form of generalized inverse can be written: $A=(U^tU)^{-1}U^tW$. This expression is truly formidable when it is realized that the matrix to be inverted, U^tU, is of dimension $n^2 \cdot n^2$. Good resolution in a pixel space would require n to be of order 64 or larger. Thus, a matrix inversion of order at least $4096 \cdot 4096$ is essential. At first, it was not known that U^tU was positive definite, although every case we checked was mildly diagonally dominant (diagonal dominance does not imply positive definiteness). In an attempt to exploit the fact that the elements of U^tU are rational when the chosen angles have rational tangents, we explored the possibility of doing exact inversions; i.e., using integer inversion techniques. This implied that U^tU be transformed to an *integer matrix* by multiplying each element by the least common denominator of all elements in U. Because of the complexity of exact inversion techniques in large dimensions, only a limited size matrix was inverted by this technique (see Appendix A for a discussion of this and other details). Fortunately, it was found that in most cases U^tU already has a small determinant and condition number, so it was properly conditioned for standard inversion techniques from the start. However, due to the finite precision of computer representations of numbers, numerical instability will again arise when the dimension of the matrix becomes large enough. At that point, integer inversion techniques could again come into use if the implementation were fast enough and the computer system could provide adequate, readily available storage.

Optimizing the direct method

What can be done to optimize the reconstruction methods given above? Katz has proposed several ideas. First, he derives another family of least square techniques for the problem based on Gauss-Markov statistics. This permits a comparison between projections angles, although not between individual detectors within a detector array. Second, his analysis led to the first rigorous error bounds in $Z(n)$. A factor in this error bound involves a function-theoretic norm which is a function of the projection angles via U. Yet another interesting idea is that of modifying the photon dose (or beam intensity) as a function of angle. Such modifications change both the Gauss-Markov statistics of a given reconstruction geometry, and the norm in the error bound. Each of these ideas provides a separate point of view in this study of optimal angles, and they will be examined in turn.

Gauss-Markov statistics

In attempting to favor those equations containing the "best" information for a particular reconstruction, the least squares problem was recast in the framework of Gauss-Markov statistics. The equation obtained is: $U^tG^{-1}UA=U^tG^{-1}W$. Or, multiplying on the left by the inverse: $A=(U^tG^{-1}U)^{-1}U^tG^{-1}W$. Here, G is the covariance matrix of W in the Gauss-Markov statistic. It should also be noted that the original variance structure used in [14] has been replaced by a more appropriate one.[16] Next, the Gauss-Markov theory says that the covariance of A is $(U^tG^{-1}U)^{-1}$. That is, $(U^tG^{-1}U)^{-1}$ gives information as to how the pixels in a reconstruction are statistically correlated with one another. In addition to minimizing these correlations in some sense, the "loss function", $\det((U^tG^{-1}U)^{-1})$, also provides a measure of the reconstruction errors due to inconsistencies between projections. Hence, minimizing this quantity as a function of the fixed number of projection angles involved provides an indicator as to which angles are statistically optimal.

Before presenting the numerical results, the tools employed in this computer study and subsequently throughout the paper should be presented. To optimize (or minimize) the functions given here, a powerful technique which is a modification of the "Bremermann method"[4] was used. The reasons for this particular choice of optimization algorithm were numerous, and are discussed in Appendix B. Because of the complexity of the functions to be minimized, we also plotted segments of these functions to aid in adapting the optimizer to the non-smooth nature of many of these functions as well as to increase our understanding of the meaning of the functions themselves.

In studying these graphs, it should be noted that it is the first angle from each indicated table of angles which is being varied. Also, as G is a factor in every function plotted in these figures, it should be pointed out that its elements are only computed up to a factor of $(PN)^{-1}$, where P is an attentuation constant characteristic of the object being reconstucted, and N is the total number of photons involved in obtaining one set of projection data.[16] Thus, the value of the loss function is only accurate to a factor of $(PN)^{-n^2}$. When this factor is small enough to render unimportant the 30 orders of magnitude variation in the loss function to be discussed shortly, then the results given here are perhaps not that meaningful. However, when all possible mileage is needed from a limited amount of data, then the overall results can depend crucially on the choice of angles. The function studied in the next section, M_2, also depends on G, but in a different manner. Using this functional dependence, it can be shown that the numbers given in the plots of M_2 are accurate to a factor of $(PN)^{-.5}$.

The results of optimizing the loss function, $\det((U^tG^{-1}U)^{-1})$, was that angles

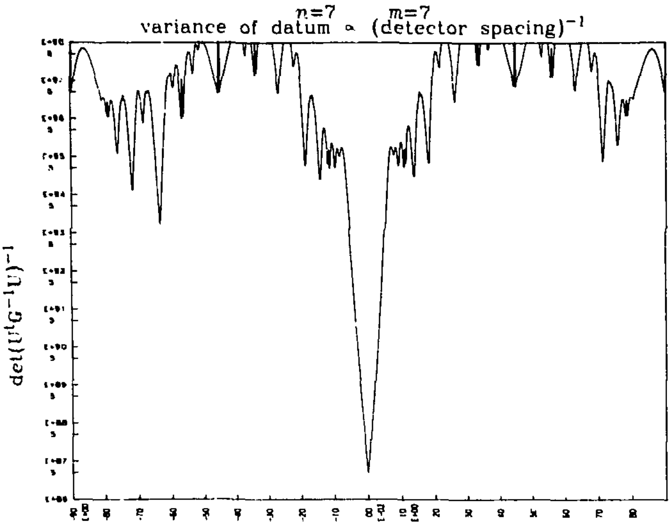
Angles of Minimum Required Resolution

Angle (in radians)	Required Resolution
0.0	1.0000
$\pi/2$	1.0000
$\pi/4$	0.7071
$-\pi/4$	0.7071
$\tan^{-1}(1/2)$	0.4472
$\tan^{-1}(2/1)$	0.4472
$\tan^{-1}(-1/2)$	0.4472

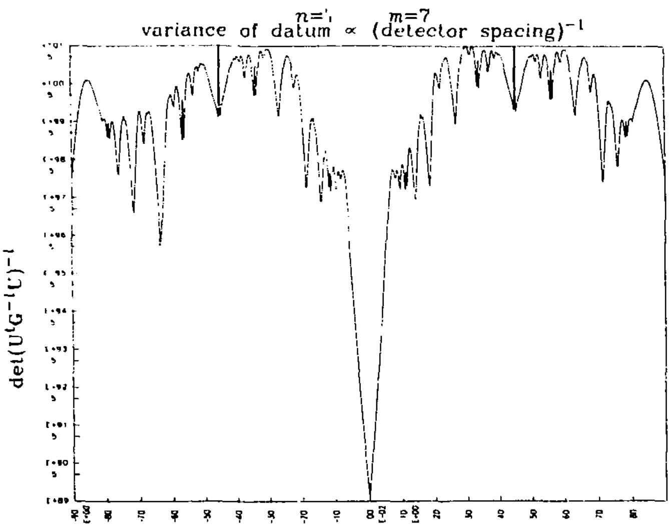
Table 1

from the Farey angles of order n were dramatically better than any other choice. In particular, the 7 angles of minimum required resolution given in Table 1 were preferred for n=7. A plot illustrating the situation is shown in Figure 4, where the set of angles in Table 1 occurs at the center of the plot. The dramatic rise by nearly 10 orders of magnitude in moving the first angle slightly away from 0.0 radians clearly illustrates just how strong this minimum is. Similar plots are shown in Figures 5, 6, and 7. Figure 5 differs from Figure 4 only in the second angle (not varied in the plot). The difference is about .04 radians (see Table 2). Since equally-spaced angles are of such interest in reconstruction work, the case of 7 equally-spaced angles is plotted in Figure 6 at 90° and −90° (these functions are periodic with period π, so each of these graphs show a period of the loss function in a different section of the space). The precise 7 angles used are given in Table 3. Finally, to get an idea of what a randomly selected region of the space looks like, Figure 7 was plotted. It used the last 6 angles in Table 4 as fixed parameters.

One should first note that the loss function shows unpredictable trends (exhibiting frequent, very rapid variations). Hence, optimizing it by any quadratic minimization technique and even most probabilistic ones would be hopeless. It is also of great interest to observe that a great many of the sharp drops in the function occur at Farey angles. For example the −90°, 0°, and 90° ordinates in



θ in degrees
Generated from last 6 angles in Table 1
Figure 4



θ in degrees
Generated from last 6 angles in Table 2
Figure 5

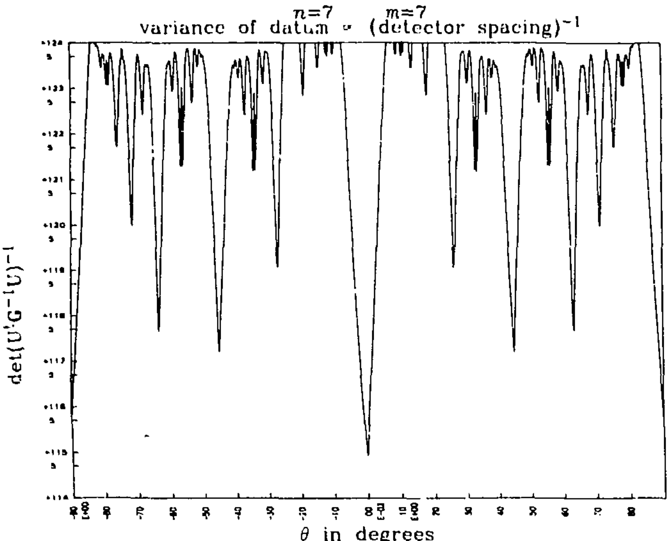
Figure 6 are Farey angles of order 1. Interestingly, the −90° and 90° ordinates correspond to equally-spaced angles, demonstrating that occasionally the Farey angles of order n and a specific set of equally-spaced angles will possess elements in common.

The overall structure observed in these plots was an aid in interpreting the results which were obtained in attempting to optimize the loss function. The modified Bremermann optimizer was started at the 7 equally-spaced angles and rapidly

Angles Near Set of Minimum Required Resolution

Angle (in radians)	Required Resolution
0.0	1.0000
$\tan^{-1}(25/1)$	0.0400
$\pi/4$	0.7071
$-\pi/4$	0.7071
$\tan^{-1}(1/2)$	0.4472
$\tan^{-1}(2/1)$	0.4472
$\tan^{-1}(-1/2)$	0.4472

Table 2



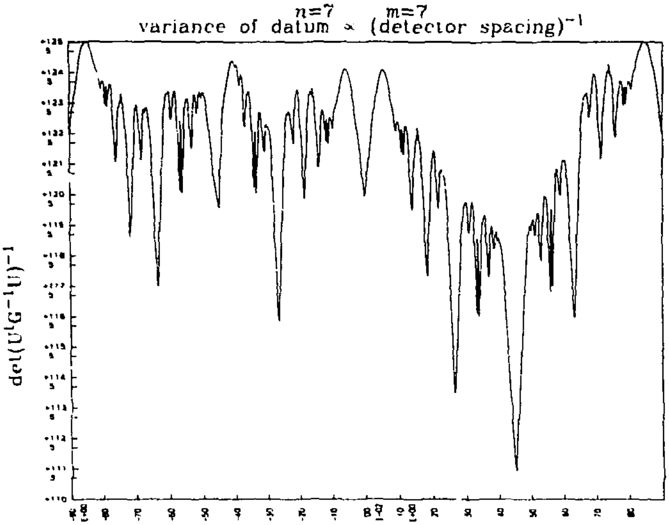
Generated from last 6 angles in Table 3
Figure 6

Angles are Equally-Spaced

Angle (in degrees)	Required Resolution
90.0000	1.0000
64.2857	0.0332
38.5714	0.0099
12.8571	0.0848
−12.8571	0.0848
−38.5714	0.0099
−64.2857	0.0332

Table 3

moved to points with function values several orders of magnitude better. Further improvement came only very slowly, but it is significant to note that invariably the local minima which cause the "hang-up" to occur in the progress of the optimizer involved one or more Farey angles. Indeed, in an optimization which ran 1000 iterations, 4 of the 7 angles of the local minimum were Farey angles to 4 significant figures. Another similar optimization found a better value (by 4 orders of magnitude) with three of the angles being Farey angles. The final function values in



θ in degrees
Generated from last 6 angles in Table 4
Figure 7

A Set of Randomly Selected Angles

Angle (in radians)	Required Resolution
$\tan^{-1}(1/1)$	0.7071
$\tan^{-1}(3/1)$	0.3162
0.0713	0.0712
$\tan^{-1}(-1/3)$	0.3162
-0.7509	0.0487
$\tan^{-1}(-7/4)$	0.1240
-1.5292	0.0416

Table 4

these two cases were $1.077 \cdot 10^{100}$ and 1.104×10^{96} respectively. Yet a third optimization was run from the "random" set of angles listed in Table 4. After 40 iterations, a function value of 7.054×10^{96} was reached at a point with six of the seven angles being nearly Farey angles. These angles are (with the corresponding Farey angle in parenthesis): 1.197(1.107), .791(.785), $-.001$(.000), $-.776$($-.785$), -1.10(-1.107), and -1.53(-1.57). To increase our confidence that the angles of minimum required resolution (see Table 1) do indeed minimize the loss function, we ran the optimizer starting at that point. No better function value was ever found.

These results lead one to the conclusion that Farey angles are very strongly preferred in this statistical framework. Indeed, if a range of angles were chosen for each of the m projection angles to be used, then if that range included a Farey angle of order n, it would be strongly preferred over any other in the chosen range. Though it was disappointing that the computer time could not be afforded to let the optimizer run several thousand iterations and find the global minimum at the angles given in Table 1, it was nonetheless gratifying to find that Farey angles of order n are to preferred, even locally.

All of the computer runs discussed thus far involved Z(7). To justify the use of this example as indicative of the answers to be found in other dimensions, a plot

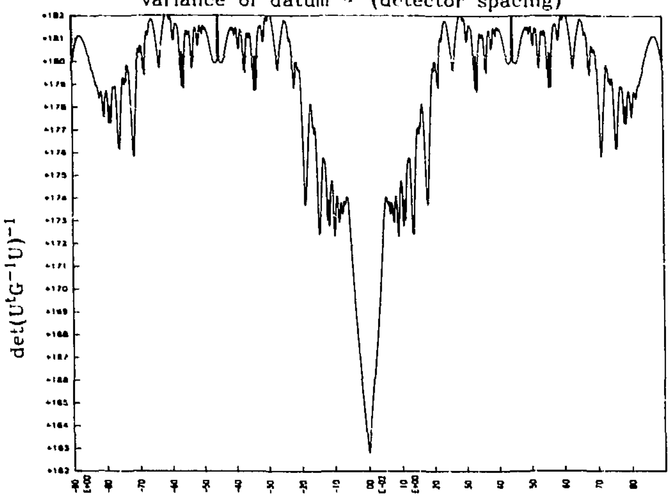
Angles of Minimum Required Resolution

Angle (in radians)	Required Resolution
0.0	1.0000
$\pi/2$	1.0000
$\pi/4$	0.7071
$-\pi/4$	0.7071
$\tan^{-1}(1/2)$	0.4472
$\tan^{-1}(2/1)$	0.4472
$\tan^{-1}(-1/2)$	0.4472
$\tan^{-1}(-2/1)$	0.4472

Table 5

of the loss function in $Z(9)$ about the eight Farey angles of order 2 is shown in

n=9 m=8
variance of datum ∝ (detector spacing)$^{-1}$



θ in degrees
Generated from last 7 angles in Table 5
Figure 8

Figure 8. Its very similar structure to the corresponding plot in $Z(7)$, Figure 4, strengthens this contention.

<u>The operator norm in the error bound</u>

The idea of providing a bound on the error in a $Z(n)$ reconstruction as an approximation to the unknown attentuation distribution is also fairly new. And since the consideration of new variance structures modifies the operator, the former operator, U^tU, must be replaced by the more general operator, $(U^tG^{-1}U)^{-1}$, in the norm. The new factor in the error bound involving the operator norm becomes[16]

$$M_2=\frac{\sigma}{(\text{minimum eigenvalue of } U^tD^{-1}U)^{.5}}$$

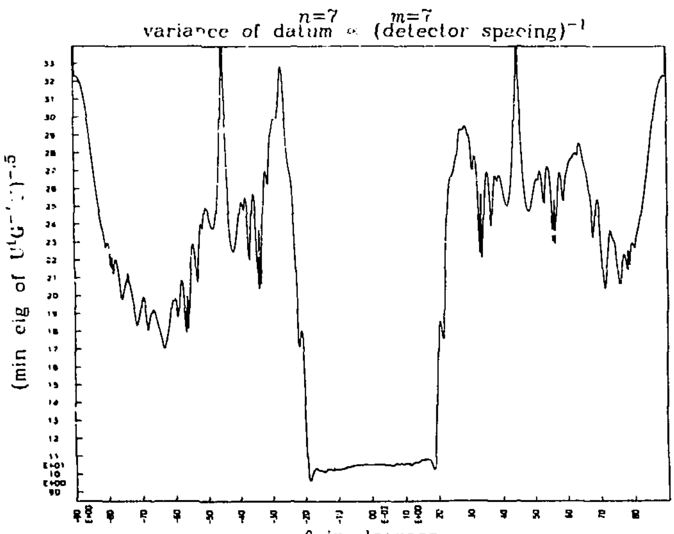
where $G=\sigma^2D$. And the error bound itself changes to

$$E=\epsilon_1(n)+M_2(M^{.5}\epsilon_1(n)+\epsilon_0)$$

$\epsilon_1(n)$ is a measure of how close the original object is to a $Z(n)$ restriction of that object, and ϵ_0 is the norm of the difference between the projection data of the continuous object and the projection data obtained from the $Z(n)$ reconstruction which the algorithm selects. Also, for our purposes here, $M=m$. If there is too much structure in the original object or if that object has very large gradients, $\epsilon_1(n)$ will be large, and consequently our reconstruction will be poor. However, it is of interest to note that if our function exists in $Z(n)$ to start with, and there is no noise in the data, we can reconstruct that function exactly. That we can obtain such an exact reconstruction has been verified by experiment up to $Z(27)$ through an actual computer implementation. The significance of noise in the data is described in the variance structure of the data and thereby in M_2; as M_2 is the crucial parameter of the error estimate which is angle dependent, the minimization of M_2 is this criterion for optimal projection angles.

The result of the optimization of M_2 is that the preferred angles are neither all Farey angles nor equally-spaced. This is again very curious because of local minima occurring throughout the plotted portion of the space which involve Farey angles of order n. Indeed, as an optimization proceeded, the local minima found involved one or more angles which were close to the Farey angles. The best value obtained via 40 iterations of the modified Bremermann optimizer was 46.6 at angles (in radians) of −.960, −.685, −.329, .093, .674, .111, and 1.43. In this set, two angles, .093 and .111, may be considered close to the two Farey angles, 0.0 and .1107.

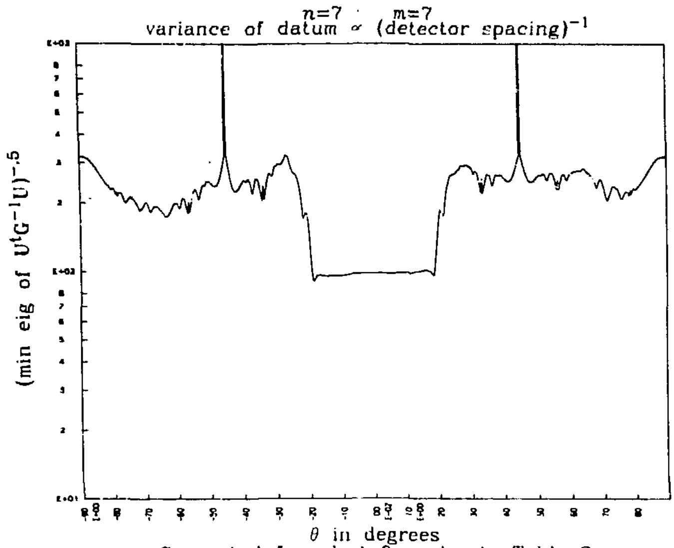
One-dimensional plots of this function, M_2, in the same neighborhoods of the four sets of angles involved in the loss function study are given in Figures 9



Generated from last 6 angles in Table 1

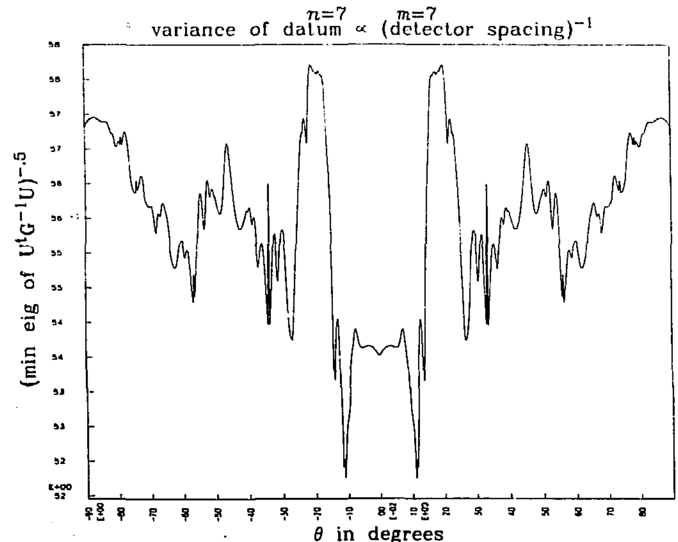
Figure 9

through 12. The first of these graphs, Figure 9, involves the angles of Table 1 at the 0° ordinate. It is striking that the space appears to be so flat in the region of the angles of minimum required resolution (the subset of the Farey angles of order n which have the largest detector widths). This fact is emphasized in Figure 10, a similar graph except that one of the fixed angles has been changed by .04 radians (see Table 2). However, all trace of this flatness at the minimum disappears when a random section of the space is plotted as seen in Figure 12, where the last 6 angles from Table 4 are fixed parameters of the graph. The 7 equally-spaced angles of Table



Generated from last 6 angles in Table 2

Figure 10

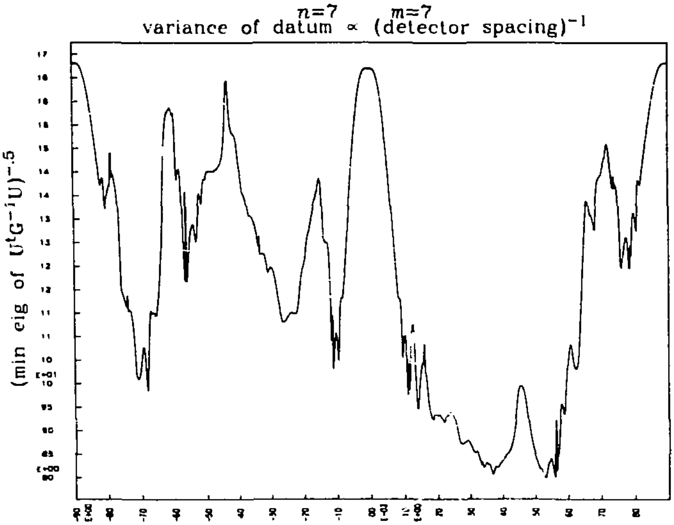


Generated from last 6 angles in Table 3

Figure 11

3 occur at 90° and −90° in Figure 11, where it can be seen that these 7 angles are near a local maximum value of M_2. But, as is also apparent from this figure, there are a large number of points that have better values for M_2 than the angles of minimum required resolution (notice that the maximum of Figure 11 is less that the minimum of Figure 9).

As was pointed out in the discussion of the optimization of M_2, many local minima involve one or more Farey angles. But in Figure 12, some of the most obvious



Generated from last 6 angles in Table 4
Figure 12

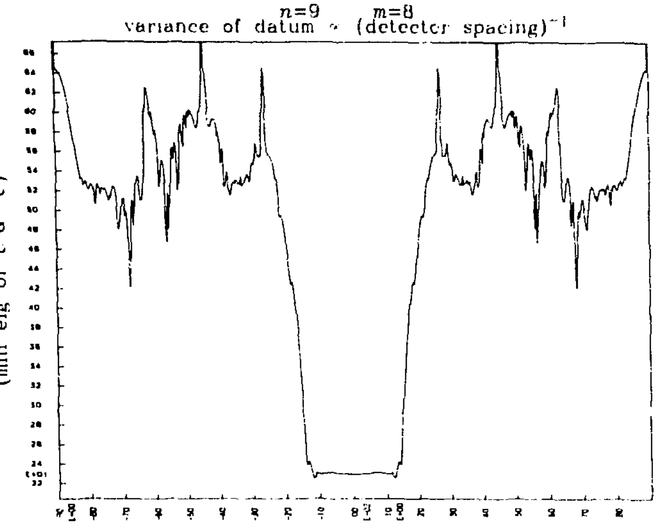
Farey angles (such as 0° and 90°) appear to occur at local maxima of M_2. This may be explained by noting that when two angles get closer and closer together, both M_2 and $\det((U^tG^{-1}U)^{-1})$ get very large. That is, groups of rows of U become pairwise dependent (corresponding rows of U_{θ_i} and U_{θ_k} become less and less linearly independent until they are actually the same), thereby causing U to become deficient in rank. This makes $\det(U^tG^{-1}U)=0$, or $\det((U^tG^{-1}U)^{-1})\to\infty$, a phenomenon dramatically visible in Figure 10 at −45° and 45°. In addition, this phenomenon would probably explain why closely-spaced angles have been found to work poorly in many reconstruction algorithms.[10]

In considering the numbers obtained in these computations, it is evident that many of them are of the same order of magnitude, i.e., 50 to 150. Thus, while the graphs show this function is certainly not smooth, it is definitely far from the wildly varying loss function of the last section. Also, there appear to be no strong local minima anywhere in the plotted domain or in the region sampled by the optimizer. Thus an optimal value of 46.6 at the 7 angles given above would not seem to have a great deal of significance when seen in the context of the error bound. That is, since M_2 multiplies a quantity involving ϵ_1 and ϵ_0 in the error bound, the fact that it may be difficult to make either ϵ_1 or ϵ_0 small in a $Z(n)$ reconstruction implies that an improvement in M_2 by a factor of two (the difference between the value at the optimal set and that at the angles of minimum required resolution) may not change the overall error bound significantly.

By plotting M_2 in the space of Z(9) about the 8 Farey angles of order 2, we once again substantiate the claim that these functions do not possess strong dimensional dependence. This can be seen in Figure 13, where the striking similarities to Figure 9 are amazing. That is, if there existed some non-trivial dimensional dependence in these complicated functions, one would expect an unusual phenomenon, such as the absolute flatness about the angles of minimum required resolution, to change substantially in changing dimensions.

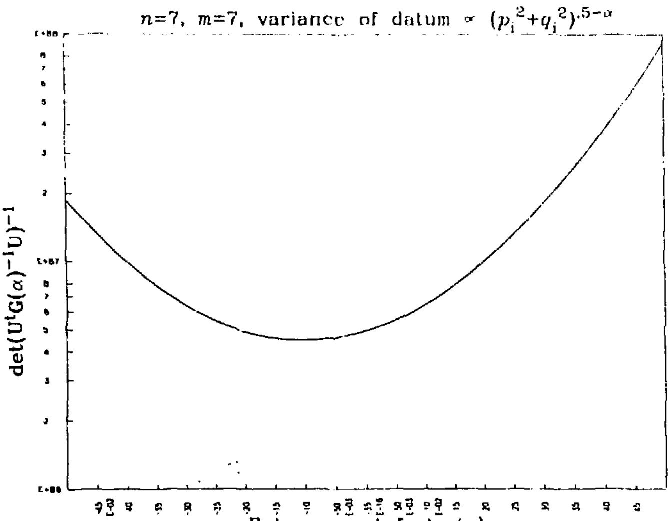
The question of photon enhancement

Since the number of photons which are seen at a detector is proportional to the size of the detector, and also since the detector width is a function of angle when



Generated from last 7 angles in Table 5
Figure 13

the projected images of vertices in $Z(n)$ define the detectors, the idea of enhancing the number of photons seen by detectors of smaller extent seems an interesting possibility. This is particularly the case because the variance of · datum is inversely proportional to the number of photons impinging upon that detector. To maintain comparability, the number of photons is fixed. Then the number of photons used per projection is modified according to an enhancement scheme explained in [15], where the scheme involves the Farey angles exclusively. The effect this procedure has on both the loss function and M_2 can be seen in *Figures 14 and 15, where the*



Uses angles of Table 1
Figure 14

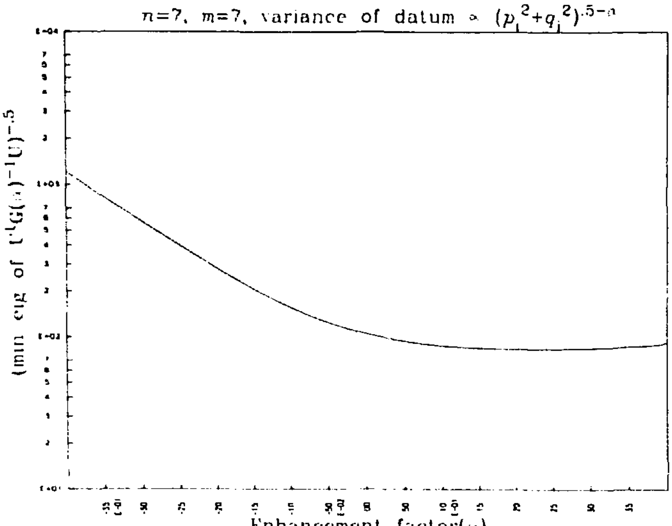


Figure 15

angles of minimum required resolution (Table 1) are the fixed parameters. The abscissa coordinate, α, is a measure of the degree to which enhancement is performed. In the case of the loss function, the preferred amount of enhancement is very small, roughly $-.103$. The graph is nearly flat in the plotted region, leading to the conclusion that no enhancement will yield virtually the same reconstruction characteristics as would the preferred value of $\alpha=-.103$. Strengthening this argument is the fact that $\det((U^tG(\alpha)^{-1}U)^{-1})\to\infty$ as $|\alpha|$ becomes large($\cdot 3$). For the error bound study, M_2 is similarly flat around an enhancement of zero. The preferred enhancement, $\alpha\approx 2.18$, is only marginally better than the value at $\alpha=0.0$. As with the loss function, $M_2(\alpha)\to\infty$ as $|\alpha|$ becomes large, so the conclusion that no enhancement in the photon distribution will work well in practical implementations of direct reconstruction methods is also substantiated in the context of the error bound.

Evaluation of direct methods

As is evidenced by the optimizations of M_2 and the enhancement factor, α, these studies have not been biased toward a particular set of angles. As a result, the strong statistical stability observed at the angles of minimum required resolution demonstrates a clear preference for those angles. That is, the near equivalence of this set of angles and the set found by the optimizer for M_2 means that we can confidently use the Farey angles of order n having the minimum required resolution in the unbiased Gauss-Markov statistics. Also, the photon enhancement study implies that radiation dose need not be modified as a function of angle in order to improve the statistics of the reconstruction. This is a definite plus when practical design considerations are taken into account. When this new information is biased with the fact that of all possible angles, those in the set of Farey angles of order n have the minimum required resolution (largest required detector spacing), a final practical choice is without question that subset of the Farey angles.

There is however, yet another practical consideration in the use of direct matrix methods using this statistical structure and the naturally induced projection data. And that is one of sheer computability. For example, we have been able to compute $(U^tG^{-1}U)^{-1}$ up to dimensions of $n=27$. With the aid of work reported in [15], this can be immediately pushed to $n=37$ (see Appendix C). These facts clearly point to the need for an alternative to direct matrix methods in reconstruction work, all least for the present. Unfortunately, the strength of the unbiased estimator developed via Gauss-Markov is lost when looking for alternative reconstruction algorithms.

II. Optimal angles and ART-like algorithms

ART and the question of uniqueness

One very popular reconstruction algorithm in $Z(n)$ has been ART (Algebraic Reconstruction Technique). Numerous modifications to it have appeared, as well as a whole family of quadratic optimization methods which have been shown[11] to correspond exactly to a number of specific ART-like algorithms currently implemented. All these methods differ from what is done in the direct matrix method discussed above in a very crucial point: the initial approximation. A second difference lies in the fact that the solution obtained from an ART-like algorithm can depend on the first iterate. This is a direct consequence of the lack of guaranteed uniqueness in these algorithms; it will be discussed further below. The initial approximation just referred to consists of making the elements of U exclusively 0 or 1 (see [22] for some less drastic approximations). The specific value of a given entry is determined by testing whether the center of a given pixel lies inside the strip which contributes to a given detector. If it lies within the strip, the corresponding value in U is 1, otherwise it is 0.[9] In some cases, this will be a reasonable approximation, particularly when only a very small corner of a pixel figures in the strip integral. However, in a large number of cases it is a very poor approximation, and the exact influence of this assumption can only be determined by numerical experiments. We call this difficulty the "centroid problem". A further difference between ART-like algorithms and the direct methods of Part I lies in the kind of projection data used. Typically the detectors are equally-spaced, and that spacing does not vary from one projection to another. In examining the effect of these differences upon the ART-like algorithms, possibly the single most important issue is the question of uniqueness. Without the guarantee of a unique object in $Z(n)$ corresponding to the finite amount of detector information in an experiment, a whole new area of problems arises. Thus it is of primary interest to know when there is uniqueness and when there is not.

There are two cases where we can give unequivocal results. First, for angles in the Farey series of order n, whenever the inequality

$$n < 1 + \min\left(\sum_{i=1}^{m} |p_i|, \sum_{i=1}^{m} |q_i| \right)$$

holds, there is but one element of $Z(n)$ corresponding to the naturally-induced projection data.[14] When the inequality ceases to hold, the null space contains non-zero elements of $Z(n)$. These objects have zero projection data at infinite resolution in the directions of each of the m angles used. In [14] they are called *ghosts*, which is a very appropriate term. The second situation where uniqueness holds again involves the naturally-induced detector spacings, but the angles come from the complement of the set of angles allowed in the first situation. In other words, the angles come from the set of all angles not in the set of Farey angles of order n. Once more, the detector spacing guarantees that we have infinite resolution of any element in $Z(n)$ at these angles, and thus the use of any single angle is sufficient to guarantee uniqueness; this is the result of [20]. In spite of the guaranteed uniqueness in this second situation, the results of Part I would seem to argue against the use of a single reconstruction angle, as would the fact that resolution requirements imposed by naturally-induced projection data would be excessive at these angles. In addition, the fact that the projection data is coming from a smooth object and not one in $Z(n)$ argues that it is unphysical to reconstruct two-dimensional structure from a single view.

As noted above, ART-like algorithms do not use the kind of projection data we have just discussed. So what can be said? In general, nothing definitive. However, even if exact "mathematical" *ghosts* are not generated by the combination of the centroid problem and an unnatural detector placement, experience has taught us that "numerical" *ghosts* will effectively be generated in many cases. Computation of any of these *ghosts* is a non-trivial matter, regardless of how close they are to being "mathematical" *ghosts*. Such a computation would depend upon the exact detector placement, and would thus vary from one reconstruction apparatus to another. Instead of undertaking a search for these complicated objects, we propose to modify ART-like algorithms to use projections from the set of Farey angles of order n, where the detectors are placed in the naturally-induced way. Unfortunately, at the deadline for this conference, the first modified ART algorithm was still being developed, so numerical results will appear at a later date.

Choice of angles in a modified ART

Why should Farey angles be preferred here in the context of ART-like algorithms as well as in the exact methods of Part I? Because uniqueness is guaranteed when the naturally-induced projection data is employed. But are there any other criteria which would bias the choice of angle? No statistical structure like Gauss-Markov has been implemented in an ART algorithm to favor one set of projection angles over another (the statistics in [12] involves favoring one reconstruction over all other elements of $Z(n)$), and no rigorous error bound such as in [14] has been given for ART, so yet another criterion needs to be found. As is shown in [10], the convergence factor established in [23] provides just such a biasing mechanism, for not only is that factor dependent upon the projection angles involved, but upon the ordering of their use as well. This convergence factor does not, however, depend upon the detector spacing at a given angle. Therefore, all angles will be allowed in an optimization study.

As is pointed out in [10], experience has shown that the order in which projections are used can greatly affect the performance of ART-like algorithms in that convergence can in some cases be greatly accelerated by a re-ordering of the projections. Two expressions are developed for bounds on this convergence factor.[10] In the case of the first bound, a transformation to another space is first performed so as to permit linear computations. The experience of Hamaker and Solomon was that this first bound was not sharp, so a second bound was derived. However, their theory is not limited to the study of bounds on convergence factors. That is, their theorems constitute a constructive theory for the exact computation of the convergence factor in the case of an arbitrary permutation of equally-spaced angles. Unfortunately, they do not develop a similar construction for unequally spaced angles, so no precise calculations of the convergence factor have been done in those cases.

It was found in our computations that the first bound was not sharp at all, and that the second bound was not significantly better. So, barring new theoretical developments, only numerical experiments can provide an aid in conjecturing what set of angles would be best for ART-like algorithms. Again because of time constraints, we have not yet performed those experiments.

Bounds on the convergence factor were computed in three spaces. The first was the poor bound mentioned above. Those computations gave the preferred orderings for subsets of the Farey angles, equally-spaced angles, and arbitrary sets of angles. There was an ordering of m equally-spaced angles which was superior to any ordering of the first m angles in a sequence of Farey angles, except in the instance when all angles generated by a given order Farey series are used. In particular, for Farey angles of order 2, the ordering -.785, 1.107, .464, -1.107, 0.0, 1.57, -.464, and .785 (in radians) gave a bound of .87 for the convergence factor, while the best ordering of 8 equally-spaced angles, 90., 22.5, -22.5, 67.5, -45., 45., 0.0, and -67.5 (in degrees), gave a bound of .93 . In an optimization using the Bremermann optimizer, yet a third set of 8 angles, .253, .870, .630, 1.435, .497, .999, .345, and .768 (in radians), gave a bound of .81 . The second computed bound gave .70 for both the equally-spaced and Farey angles, and according to a conjecture in [10], the value for an optimal ordering of 8 equally-spaced angles should be below

$.22 \approx 1-2(8/9)^8$. Thus, a final conclusion awaits experiment and possibly further constructive theoretical developments.

The third space used was the finite-dimensional restriction, $Z(n)$, of the infinite-dimensional Hilbert spaces used in deriving the first bound. As noted in the discussion on uniqueness, subsets of the Farey angles of order n can have *ghosts* in $Z(n)$ relative to them if they do not satisfy the inequality given there. These *ghosts* are not unique, and can be shown to be linear combinations of "fundamental" *ghosts* or basis vectors in $Z(n)$.[14] These basic vectors are easily generated by linear translations of the first basis vector,[14] or they can all be computed directly via a formula given in [15]

$$K(z)=h(z)+\sum_{r=1}^{m}(-1)^r\left[\sum_{i_1<i_2<\cdots<i_r=1}^{m} h\left(z+\sum_{j=1}^{r} t_{i_j}\theta_{i_j}\right)\right].$$

Here, h and K are elements of $Z(n)$ and the argument z equals (x,y), a point in the domain of an element of $Z(n)$. Since this basis allows us to generate all possible *ghosts* with respect to any set of Farey angles, it is possible to compute directly the minimum angle between the null spaces corresponding to two sets of projection angles. The angles between the null spaces give the bound on the convergence factor via:

$$c \le 1-\prod_{1\le j\le m-1} \sin^2(\beta_j)$$

where the β_j are the minimum angles between *ghosts* with respect to the one angle in the first subspace and *ghosts* with respect to the remaining m−j angles in the second subspace, with the restriction that these ghosts lie in the orthogonal complement of the intersection of the two null spaces. That is, the angle between null spaces is measured in that portion of each null space which is orthogonal to their intersection. To describe how the direct computation of β_j is done, it is essential to understand that a means exists for computing the "fundamental" *ghosts* described above. No such characterization of the null space exists in infinite dimensions so direct calculation of elements in the null space cannot be done in that case.

There are several steps involved in computing β_j. The first step is to find the "fundamental" *ghosts* with respect to the first angle, the remaining m−j angles, and the combination of the two sets of angles (this gives a basis spanning the intersection of the null spaces). Next the basis of the intersection of the null spaces is orthonormalized via a Gram-Schmidt process. Then, the orthonormalization is completed in each null space by using the basis vectors already spanning each individual null space. The orthonormal basis vectors generated by this completion process then span that part of each corresponding null space which is orthogonal to the intersection of the two null spaces. With these tools in hand, an iterative procedure of alternately projecting an element of the orthogonal portion of one null space onto the other and then reversing the process converges rapidly to give β_j (we are indebted to Myron Katz for the theory which permitted implementation of this algorithm).

The result of computing the β_j for the optimum ordering of the 8 Farey angles found in computing the first bound was a bound on the convergence factor of .994 . How can the bound be yet worse in this case? If the intersection of the null spaces takes up a proportionately larger section of each null space in the infinite-dimensional case, then conceiveably, the angle between two null spaces might be larger in the infinite-dimensional case, thereby leading to smaller observed bounds on the convergence factor. This conjecture is motivated by the observation that in finite-dimensional calculations, the angle between null spaces increases as the non-zero dimension of the intersection of the null spaces increases proportionate to the dimension of those null spaces (see Table 6). However, as noted above, there is no available characterization of the null spaces in infinite dimensions, so the conjecture just given will have to remain just that, a conjecture. Clearly, for the purposes of comparison with other sets of angles, a way to compute the convergence factor itself is still needed.

Null Space Calculations for 8 Farey Angles

j	Dimension of first null space	Dimension of second null space	Dimension of intersection	θ_j (in radians)
1	64	1	0	1.5708
2	56	6	1	0.4243
3	56	16	6	0.5684
4	56	30	16	0.6274
5	72	36	30	0.9249
6	72	42	36	0.9901
7	56	64	42	1.0943

Table 6

A final issue in the Hamaker and Solomon approach is that of working in infinite-dimensional spaces and then applying the results obtained to $Z(n)$. This may indeed be valid, but its rigorous justification would be most welcome. Such justification would definitely establish ART on yet firmer footing.

Summary

It is evident from these studies that there is a great deal of previously unknown structure in $Z(n)$. Many surprises occurred in the discoveries reported here, but possibly the greatest surprise of all was the strong preference of the Farey angles in exact reconstruction methods (direct matrix methods). We believe this is a direct result of the symmetries of $Z(n)$ itself, and therefore, some sort of preference for these angles probably exists for all algorithms using $Z(n)$ as the underlying reconstruction space. And in considering all the possible algorithms, the favorable characteristics of the direct matrix methods studied in Part I should make those methods good candidates for spaces smaller than $Z(38)$. For larger spaces, the results are inconclusive since the experiments comparing a modified ART using Farey angles to other versions of ART (such as those employing optimal orderings of equally-spaced angles) have yet to be performed. While a modified ART using equally-spaced angles in the optimal order might very well be competitive with this proposed algorithm, the large required resolution in the projection data (small detector spacing) would clearly make it less attractive as a practical reconstruction method.

Remark. As has been stated, the strong preference of the Farey angles in the context of Gauss-Markov statistics is remarkable. But one should recall that while the fact of minimizing the loss function, $\det((U^tG^{-1}U)^{-1})$, accomodates inconsistencies between projection angles, it does not accomodate empirically inappropriate data collected by detectors on the same projection rack. That is, measurements on a given projection rack are weighted equally in the statistics discussed here. As there is invariably white noise in the data, inconsistencies between the measurements on a given projection rack necessarily arise and must be dealt with. This problem can be handled by a statistical technique called cross-validation, but we will defer a discussion of it to a subsequent paper.

Appendix A

There are two competitive exact integer inversion techniques currently in use: the two-step integer-preserving Gaussian elimination and the congruential technique. The two-step Gaussian technique is a special case of a more general class of n-step Gaussian elimination processes developed by Bareiss.[1] It requires a fast multiprecision system, as every step of the process involves multiprecision integer additions, subtractions, multiplications and divisions. The higher-order Gaussian elimination methods require a little more computer memory, and are a little more complicated to program, so none have appeared in the literature. The congruential inversion process is built upon modulo arithmetic, and is sometimes called the "modular algorithm" for this reason. Uniqueness of arithmetic processes requires that the modulo bases be primes, and speed of computation and scarcity of computer memory require that the primes be large (just smaller than the size allowed by a word on the computer being used). The inverse is then found by doing a Gaussian elimination modulo a set of primes until a uniqueness condition is determined or a particular bound tells us we have the result if the matrix was not singular. Inverse modulo arithmetic is then done on the series of inverses generated.

Before December, 1977, the only congruential inversion programs of which we were aware[3,8,13] ran roughly twenty times slower than the two-step integer-preserving Gaussian elimination procedure. Since there was a fast multiprecision system at our disposition, we began by implementing the two-step Gaussian elimination system as programmed by Bareiss and his colleagues.[2] As it was of interest to do inversions for much larger dimensions of $Z(n)$, it was necessary that the matrix be kept on disk and that only a portion of it be in core at a given time. Because concurrent disk input and output was not available on that computer system (and would have been very difficult to program), and also because a competitive congruential inversion program appeared at that time,[5] the two-step technique was abandoned. Similar difficulties quickly arose using the congruential method: the number of primes required was becoming enormous since the determinant of the transformed U^tU was growing like e^{c^n}, truly an astronomical number. This meant that an overlapped input-output scheme was essential on the new computer system we were using, and a complicated interface to a double buffering program of Bruce Langdon at Lawrence Livermore Laboratory was then programmed. The optimal reworking of the congruential elimination and back-solution loops in the modulo solver improved the program's speed, but only by a factor of three or so. Thus, current computational resources seem to bound these calculations to roughly $n \leq 15$. It was concluded that the transformed U^tU cannot be inverted for large n on current computers by known methods.

Remark. It should be pointed out that we became aware of the excellent work of McClellan[18,19] only in December of 1977. But it must be admitted that we would still have been prejudiced in favor of the two-step Gaussian elimination method because our multiprecision system was optimally programmed on the IBM 7030 for the architecture of that machine, and it would have taken a large amount of effort and time to take McClellan's "SAC-1" implementation of his "modular algorithm" and make it uniformly competitive with the two-step technique, if indeed it could have become competitive. This was so because the variable-length operations in that multiprecision system allowed special short-cuts to be taken in almost every operation, leading to a huge saving in computer time as well as memory (the exact length of every multiprecision integer in the calculation was known).

Appendix B

The optimization technique used here is a modification of a technique called the "Bremermann method" in the literature, named after Hans Bremermann who discovered it.[4] It has since been improved.[21] We have replaced a numerically unstable procedure it employs (the solution of a cubic equation) by a newly developed stable method.[24] Basically, the method does line searches in Gaussian-distributed random directions in the k-dimensional space in which the function to be minimized is embedded. However, it is distinguishable from other methods which do line searches in that it does a parameterized Lagrangian fit to the five function values obtained at five equally-spaced points along a line of randomly chosen orientation. This spacing is also varied from iteration to iteration. The result of this fit is a quartic whose minima are found by solving the quartic's derivative (a cubic). The function is computed at the smallest of the quartic's minima (if the cubic has three real roots), and the minimization algorithm proceeds at the smallest f the new value and the five values used to compute the Lagrangian fit. The performance of this algorithm is astounding, and has had favorable mention in the literature,[6,7] yet many still doubt its value. There are improvements which still need to be made to the method, such as hybridizing it with a second-order minimization technique so that its convergence to a local or global minimum will be accelerated once it is in the region of attraction of that minimum, and so that the probability that "hang-up" at a local minimum in a deep well will be diminished, but the improvements due to Jaime Milstein already mentioned are already quite adequate for many applications.

The merits of this technique are numerous, and we only mention a few here relevant to our applications. First, no derivatives of the function to be minimized are required. For functions whose derivatives are analytically very large expressions or hard to compute numerically, or, as in our case, not even available due to a lack of analytic representation, this is a tremendous advantage. Second, it is not limited in its effectiveness to the local region of attraction, but "sees" a great deal of the global structure of the function. Finally, and most importantly, it requires only a few function evaluations. When, as is often the case, the function is costly to compute, this criterion alone could determine whether a numerical minimization is even attempted.

Appendix C

Recently (on May 3,1979), Katz discovered that $U^tG^{-1}U$ was not only symmetric, but "centro-symmetric" as well.[16] While searching for an empirically faster way to compute $U^tG^{-1}U$, we had independently discovered (on March 17,1979) that its computation could be decomposed as follows:

$$U^tG^{-1}U = \sum_{i=1}^{m} U^t_{\theta_i} G^{-1}_{\theta_i} U_{\theta_i}$$

This decomposition is essential to the proof of centro-symmetry. We had also been aware empirically that $U^tG^{-1}U$ was centro-symmetric. At the same time as he discovered the centro-symmetry of $U^tG^{-1}U$, Katz also found that U itself is centro-symmetric.[16] These facts permit all calculations to be done on matrices only half the size. Thus, given the computer being used, computations of $(U^tG^{-1}U)^{-1}U^tG^{-1}$ up through n=37 should be possible. Improvement beyond that point will only come with fantastic improvement in computer size and speed. Elaborate overlapped I-O schemes may allow some limited extension of the dimensional reach of the method, but inversion methods employing overlapped I-O are very strictly limited by the response times of the secondary storage. Details of an implementation of such an inversion scheme and a discussion of its drawbacks will appear elsewhere.

Bibliography

(1) Bareiss, E. H.: Sylvester's Identity and Multistep Integer-Preserving Gaussian Elimination. Math. Comp. 22(103),565-578.

(2) Bareiss, E. H., et al.: Two-Step User's Guide. AEC Report No. COO-2280-29, February, 1977.

(3) Borosh, I., and A. S. Fraenkel: Exact Solutions of Linear Equations with Rational Coefficients by Congruence Techniques. Math. Comp. 20(93),107-112.

(4) Bremermann, H.: A Method of Unconstrained Global Optimization. Math. Biosci. 9(1970),1-15.

(5) Cabay, S., and T. P. L. Lam: Algorithm 522, ESOLVE, Congruence Techniques for the Exact Solution of Integer Systems of Linear Equations. ACM Trans. Math. Software 3(4),404-410.

(6) Dixon, L. C. W., and G. P. Szego eds.: Towards Global Optimization. Proceedings of a Workshop at the University of Cagliari, Italy, October, 1974.

(7) ______: Towards Global Optimization 2. North-Holland, 1978.

(8) Fraenkel, A. S., and D. Loewenthal: Exact Solutions of Linear Equations with Rational Coefficients. J. Res. Nat. Bur. Stand. 75B(1 and 2),67-75.

(9) Gordon, R.: A Tutorial on ART. IEEE Trans. Nucl. Sci. NS-21(3),78-93.
(10) Hamaker, C., and D. C. Solomon: The Angles Between the Null Spaces of X-Rays. J. Math. Anal. Appl. 62(1978),1-23.
(11) Herman, G. T. and A. Lent: Quadratic Optimization for Image Reconstruction. I. Computer Graphics Image Processing 5(1976),319-332.
(12) ________: A Computer Implementation of a Bayesian Analysis of Image Reconstruction. Info. Control 31(1976),364-384.
(13) Howell, J.: Algorithm 406: Exact Solution of Linear Equations Using Residue Arithmetic. J. ACM 22(1),38-50.
(14) Katz, M. B.: Questions of Uniqueness and Resolution in Reconstruction from Projections. Lecture Notes in Biomathematics, Vol. 26. Simon Levin ed., Springer Verlag, November, 1978.
(15) ________: Theoretical Design of a Computerized Tomography System: A 4mm Resolution Example. IEEE Trans. Computer Society (to appear).
(16) Personal Communication from Myron Katz of the University of New Orleans.
(17) Knuth, D. E.: The Art of Computer Programming, Vol. 1, Fundamental Algorithms, pp. 157,515, Addison-Wesley, 1969.
(18) McClellan, M. T.: A Comparison of Algorithms for the Exact Solution of Linear Equations. ACM Trans. Math. Software 3(2),147-158.
(19) ________: The Exact Solution of Linear Equations with Rational Coefficients. ACM Trans. Math. Software 3(1),1-25.
(20) Mersereau, R. M. and A. V. Oppenheim: Digital Reconstruction of Multidimensional Signals from Their Projections. Proc. IEEE 62(1974),1319-1339.
(21) Personal Communication from Jaime Milstein of the University of Southern California.
(22) Oppenheim, B. E.: More Accurate Algorithms for Iterative 3-Dimensional Reconstruction. IEEE Trans. Nucl. Sci. NS-21(3),72-77.
(23) Smith, K. T., D. C. Solomon, and S. L. Wagner: Practical and Mathematical Aspects of the Problem of Reconstructing Objects from Radiographs. Bull. Amer. Math. Soc. 83(6),1227-1270.
(24) Vignes, J.: New Methods for Evaluating the Validity of the Results of Mathematic Computations. Math. Comp. Sim. XX(1978),227-249.