Optimization of reconstruction algorithms using Monte Carlo simulation

*K. M. Hanson**
Los Alamos National Laboratory, MS P940
Los Alamos, New Mexico 87545
USA

ABSTRACT: A method for optimizing reconstruction algorithms is presented that is based on how well a specified task can be performed using the reconstructed images. Task performance is numerically assessed by a Monte Carlo simulation of the complete imaging process including the generation of scenes appropriate to the desired application, subsequent data taking, reconstruction, and performance of the stated task based on the final image. The use of this method is demonstrated through the optimization of the Algebraic Reconstruction Technique (ART), which reconstructs images from their projections by an iterative procedure. The optimization is accomplished by varying the relaxation factor employed in the updating procedure. In some of the imaging situations studied, it is found that the optimization of constrained ART, in which a nonnegativity constraint is invoked, can vastly increase the detectability of objects. There is little improvement attained for unconstrained ART. The general method presented may be applied to the problem of designing neutron-diffraction spectrometers.

Introduction

The overall purpose of an imaging system is to provide information about the object or scene being imaged. For mission-oriented imaging systems, the type of scenes expected and the kind of information desired can frequently be specified. In such a case an imaging system should be optimized on the basis of how well the specified tasks can be performed using the resulting images. Here this approach to optimization is applied to only one aspect of the complete imaging system, that of the image reconstruction algorithm. It is shown that such an optimization is distinctly practical and can be extremely beneficial.

Several classes of measures have been employed in the past on which to base the optimization of reconstruction algorithms [1]. Some are based on the fidelity of the reconstructed images, such as the conventional measure of the

*This work was supported by the United States Department of Energy under contract number W-7405-ENG-36.

rms difference between the reconstruction and the original image, simply called the rms error. Experience teaches us that this does not always seem to be correlated with the usefulness of images. There are alternative measures based on how closely the estimated reconstruction reproduces the measurement data, for example, the mean-square residual. Unfortunately, without further constraints reconstruction based on minimizing the mean-square residual is known to be ill-conditioned or even worse, ill-posed [1].

In the approach to algorithm optimization presented here, an algorithm is rated on the basis of how well one can perform stated tasks using the reconstructed images. As shown in Ref. [2], task performance in a well specified imaging situation is readily assessed numerically through a Monte Carlo technique that is used to simulate the complete imaging process. The optimization procedure involves maximizing task performance by varying whatever free parameters exist in the reconstruction algorithm.

This article closely follows one that appeared in conjunction with an SPIE conference [3]. The main thrust of the present article is the solution of the tomographic reconstruction problem in which a two-dimensional image is to be determined from a set of projections (line integrals) taken through it. However, the same kind of difficulties that exist in tomographic reconstruction are present in other image-recovery problems. That goes for the deblurring of blurred data in either one or two dimensions. The technique presented here for evaluation and optimizaton of a reconstruction algorithm has obvious applications to many of the questions that have been posed during this workshop regarding the best design of neutron-diffraction spectrometers. It is well to remember that the data-collection system includes both the spectrometer design and the subsequent data analysis, which includes any deblurring that might be deemed necessary. Optimization of the quality of the final data should also include the effects of the data processing that may be required for the proper interpretation of the data.

Method to Calculate Task Performance

For linear imaging systems the effects of image noise on task performance can be predicted for a variety of simple tasks [4]. The same cannot be said of the effects of artifacts. The masking effects of measurement noise are truly random in nature. The random noise process results in each set of measurements being different, even when the scene being imaged does not change. However, reconstruction from limited data typically produces artifacts in the reconstructed images that behave differently than the fluctuations arising from random noise. They manifest themselves as seemingly unpredictable irregularities that look like noise, but in a strict sense, they are not. They are deterministic since they can be predicted from the combined knowledge of the measurement geometry, the scene, and the reconstruction algorithm. Since these artifacts depend on the scene, a single realization of a simple scene is plainly inadequate to judge a reconstruction algorithm. It is necessary to obtain a statistically meaningful average of the response of an algorithm to many realizations of the ensemble of scenes with which it must cope.

A Monte Carlo technique, one that employs pseudo-random numbers to generate its results, is used to simulate the entire imaging process from scene generation to the final task performance, because it can readily provide the above variations within the ensemble.

The method requires first a complete specification of the entire problem in the following manner:

a) Define the class of scenes to be imaged with as much complexity as exists in the intended application.

b) Define the geometry of the measurements. The deficiencies in the measurements such as blur, uncertainties in the geometry, and uncertainties in the measurements (noise) should be specified.

c) Define clearly the task to be performed. Details concerning what is known about the signal and the background must be stated explicitly.

d) Define the method of task performance. This method should be consistent with the intended application and the *a priori* known information.

The simulation procedure is then performed by doing the following:

e) Create a representative scene and the corresponding measurement data by means of a Monte Carlo simulation technique.

f) Reconstruct the scene with the algorithm being tested.

g) Perform the specified task using the reconstructed image.

h) Repeat steps e) through g) a sufficient number of times to obtain the necessary statistics on the accuracy of the task performance.

Finally, determine how well the task has been performed:

i) Evaluate the task performance using the relevant measure of performance.

The advantage of this numerical approach is that it readily handles complex imaging situations, nonstationary imaging characteristics, and nonlinear reconstruction algorithms. Its major disadvantage is that it provides an evaluation that is valid only for the specific imaging situation investigated.

ART

The Algebraic Reconstruction Technique (ART) [5] is an iterative algorithm that reconstructs a function from its projections. It has proven to be a very successful algorithm in tomographic reconstruction, particularly for estimating a function when there is a limited amount of data available. It is identical to the Kaczmarz algorithm [6], which provides a pseudoinverse solution to a singular system of linear equations [7] and works particularly well when the matrix is sparse. Assume that N projection measurements are made of the unknown function f, which will be considered a vector. As these measurements are linearly related to f, they may be written as

$$g_i = H_i f, \quad i = 1, ..., N, \tag{1}$$

where g_i is the ith measurement and H_i is the corresponding row of the measurement matrix. The ART algorithm proceeds as follows. An initial guess is made, for example, $f^0 = 0$. Then the estimate is updated by iterating on the individual measurements taken in turn:

$$f^{k+1} = f^k + \lambda^k H_i^{\mathrm{T}} \left[\frac{g_i - H_i f^k}{H_i^{\mathrm{T}} H_i} \right], \tag{2}$$

where $i = k$ mod(N)+1 and λ^k is a relaxation factor for the kth update. Any applicable constraints are invoked after each update. For example, for constrained ART in which a nonnegativity constraint is enforced, when $f^{k+1} < 0$, set $f^{k+1} = 0$. In the absence of constraints, the normalization of (2) is such that when $\lambda^k = 1$, f^{k+1} is guaranteed to satisfy the measurement equation (1). In the standard nomenclature one iteration is completed after the full set of N measurements has been processed. We use the index K to indicate the iteration number ($K = \mathrm{int}(k/N)$). Variable relaxation (or damping) factors are used here to attenuate successive updates during the reconstruction. We will express the relaxation factor as

$$\lambda^K = \lambda_0 (r_\lambda)^{K-1}. \tag{3}$$

The proper choice of the relaxation factor is the issue at hand. There is very little guidance on this choice in the literature. It is known [8] that if a solution to the measurement equations exists, the ART algorithm will converge to it in the limit of an infinite number of iterations provided that $2 > \lambda_k > 0$. A value of unity is often suggested. Censor *et al.* [9] have shown that unconstrained ART ultimately converges to a minimum-norm least-squares solution if the relaxation factor approaches zero slowly enough. However, λ^K will asymptotically approach zero for any value of $r_\lambda < 1$. The value appropriate to a finite number of iterations remains uncertain. In previous work the author has assumed for λ_0 and r_λ the nominal values of 1.0 and 0.8 for problems involving a limited number of projections, and 0.2 and 0.8 for problems involving many (~100) views [2]. This choice for r_λ makes the final λ^K at ten iterations about seven times smaller than the initial one λ_0. In our experience unconstrained ART converges reasonably well in ten iterations. Next we discuss a way to find the best choice for the relaxation parameters for a given problem.

Optimization of ART

The use of numerically calculated task performance will be demonstrated by searching for the optimum choice of λ_0 and r_λ for the ART algorithm. For the present purpose, the class of scenes is assumed to consist of a number of non-overlapping discs placed on a zero background. For this example, each scene contains 10 high-contrast discs of amplitude 1.0 and 10 low-contrast discs with amplitude 0.1. The discs are randomly placed within a circle of reconstruction, which has a diameter of 128 pixels in the reconstructed image. The diameter

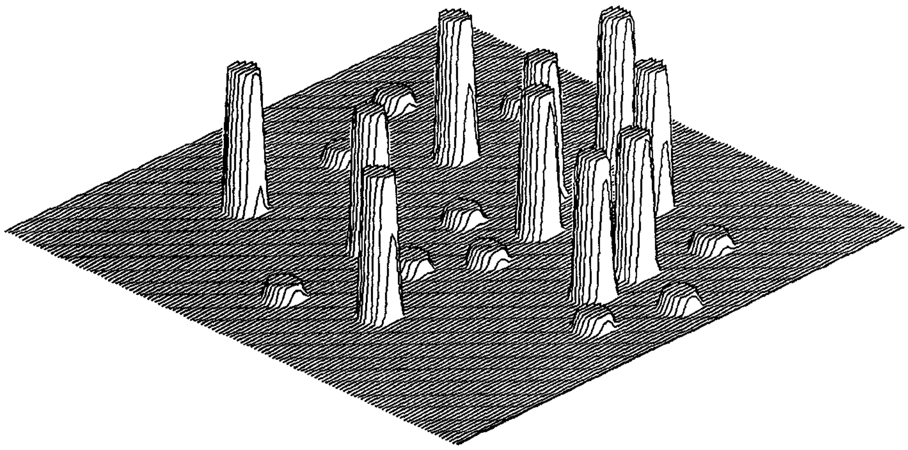

Figure 1: The first randomly generated scene consisting of 10 high-contrast and 10 low-contrast discs. The evaluation of task performance is based on an average over ten similar scenes.

of each disc is 8 pixels. The first of the series of images generated for these tests is shown in Fig. 1. In this computed tomographic (CT) problem, the measurements are assumed to consist of a specified number of parallel projections, each containing 128 samples. Ten iterations of ART are used in all of the present examples. It is assumed that the task to be performed is the detection of the low-contrast discs. To produce noisy data, random noise is added to the projection measurements using a Gaussian-distributed random number generator.

The result of reconstructing Fig. 1 from 12 noiseless views spanning 180° is shown in Fig. 2. The seemingly random fluctuations in the background are actually artifacts produced by the limited number of projections and arise mainly from the high-contrast discs. As the artifacts depend on the positions of the discs, it is important to allow for random placement of the discs to allow for the full range of artifacts. It appears that the nonnegativity constraint improves the reconstruction considerably in that it has reduced the confusion caused by the fluctuations in the background. However, upon careful examination, one finds that some of the low-contrast discs have not been reproduced. Also, there still remain many fluctuations in the background that may mislead one to suspect the presence of discs in places where none exist in reality. Thus, on the basis of this single example, one cannot say with certainty whether or not the detection of the low-contrast discs is improved by the nonnegativity constraint. A statistically significant comparison between reconstructions with and without the constraint must be made to assess its value.

The task to be performed is assumed to be the simple detection of the low-contrast discs. It is assumed that the position of a possible disc is known beforehand as is the background. To perform the stated task of detection, it is

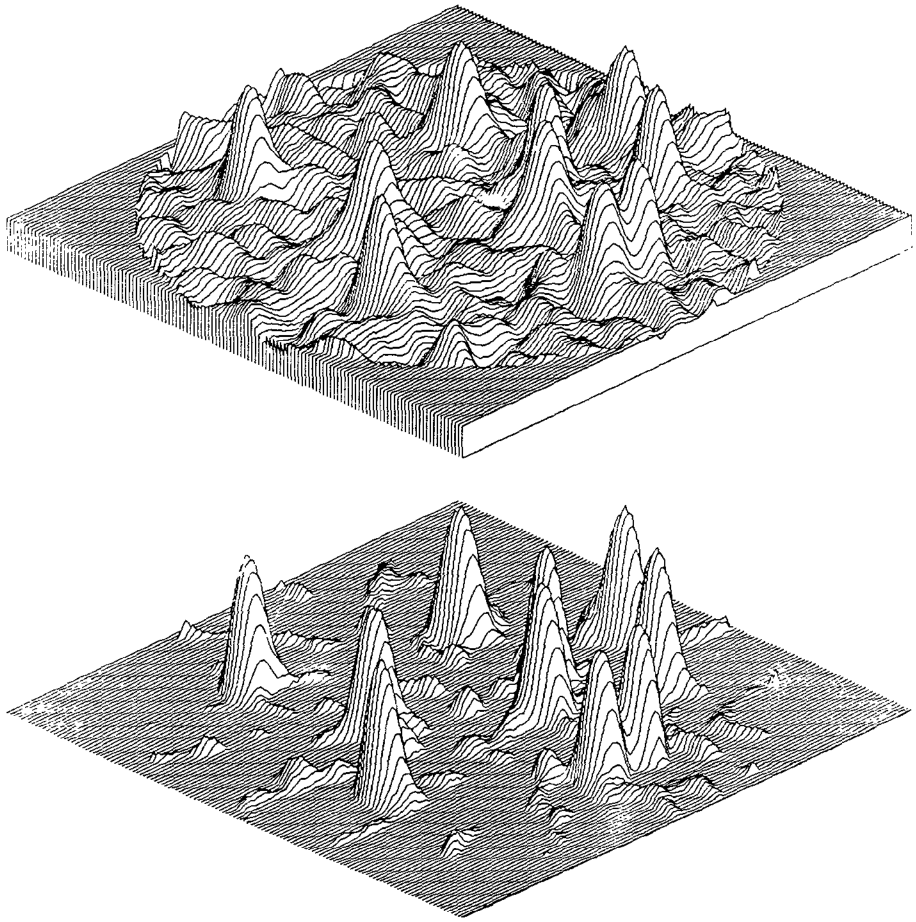

Figure 2: Reconstructions of Fig. 1 from 12 noiseless parallel projections subtending 180° obtained with 10 iterations of the ART algorithm (top) without and (bottom) with the nonnegativity constraint. These reconstructions were obtained with $\lambda_0 = 1.0$ and $r_\lambda = 0.8$.

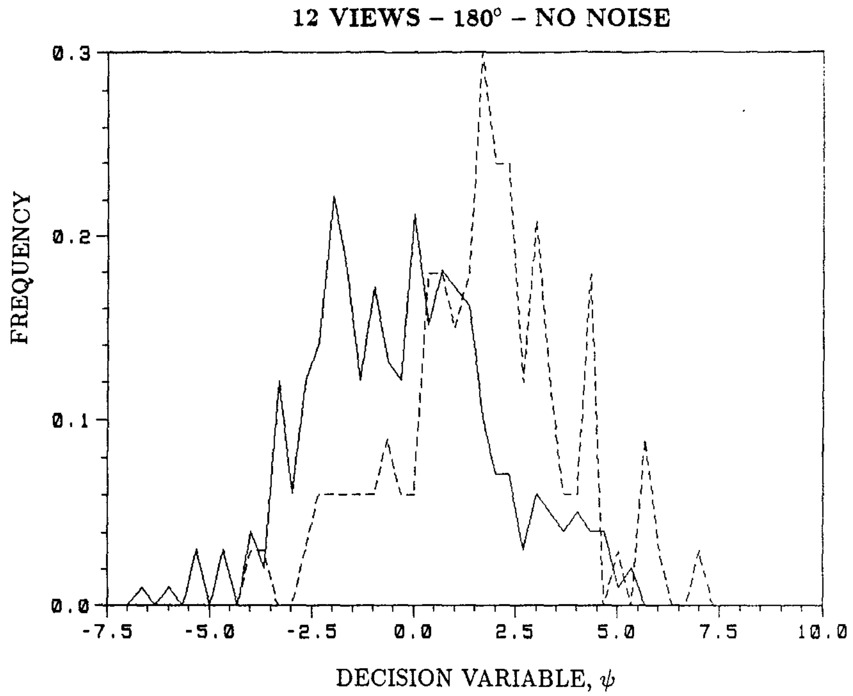

Figure 3: The frequency distributions of the decision variable (the sum over a circular region) evaluated where a low-contrast disc is known to exist (dashed line) and where none exists (solid line) for ART reconstructions without the nonnegativity constraint. These results summarize the detection performance obtained from reconstructions from 12 views for 10 randomly-generated scenes.

assumed that the sum over the area of the disc provides an appropriate decision variable ψ. This sum is an approximation to the matched filter, which is known to be the optimum decision variable when the image is corrupted by additive uncorrelated Gaussian noise [10]. Ignored is the blurring effects of the finite resolution of the discretely-sampled reconstruction. Neither is account taken of the known correlation in the noise in CT reconstructions [11] that have been derived from projections containing uncorrelated noise. After reconstruction, the sums over each region where the low-contrast objects are known to exist are calculated, as well as those over each region where none exist. These two data sets may be displayed as histograms in this decision variable as shown in Fig. 3. To perform the detection task, a disc will be said to be present at each location where the value of the decision variable is above a chosen threshold. The degree of separation between these two distributions is often characterized by the detectability index d', given by

$$d' = \frac{\psi_1 - \psi_0}{\sqrt{\frac{\sigma_1^2 + \sigma_0^2}{2}}}, \tag{4}$$

where ψ_1 and σ_1 are the mean and rms deviation of the frequency distribution when the object is present and those with the subscript 0 are when the object is not present. This quantity is sometimes called the signal-to-noise ratio (SNR) for detection. Clearly, larger d' implies better separation of the two distributions and hence better detectability. For the histograms shown in Fig. 3 obtained for unconstrained reconstructions, d' is 0.871. When the same analysis is carried out on reconstructions employing the nonnegativity constraint, a value of 2.054 is obtained. We conclude that the nonnegativity constraint has improved detectability. As noted in Ref. [2], the detectability index based on the area under the receiver operating characteristic curve d_A may be more appropriate for the binary decision task. But d' has better statistical accuracy than d_A and is more likely to be a continuous function of the parameters that can be varied in the reconstruction procedure. Thus d' is the preferred choice for the purpose of optimization.

Fig. 4 shows how two choices for optimization functions depend on λ_0 and r_λ for constrained ART. There is a definite minimum in these functions indicating optimum operating points for these two parameters. However, the minima are at different values of these parameters. Which operating point should we choose? Fig. 5 shows the reconstructions obtained using the relaxation parameters for optimization with respect to $100/d'$ and the rms error in the reconstruction. There is an enormous improvement in the quality of both reconstructions over those shown in Fig. 2. Optimization with respect to $100/d'$ appears to be preferable because it yields a d' that is twice as large as the optimization with respect to rms error. The latter also leads to annoying streak artifacts, which are quite visible in a good display of the reconstruction. The same kind of contour plots for unconstrained ART are relatively flat and uninteresting.

Fig. 6 shows reconstructions obtained from noisy data. Because of the large number of views, the data are complete. For the unconstrained and constrained reconstructions, d' is found to be 1.995 and 1.825, respectively. In this case the nonnegativity constraint has worsened detectability, contrary to what might be concluded from a first glance. The CPU time required to calculate these detectabilities took about one hour on a VAX 8700, which is about four times faster than a VAX 785.

The optimum values for λ_0 and r_λ were found for various conditions of data collection using a function minimizer from the NAG library[1] called E04JB. This routine finds the parameters for the minimum of a function after many evaluations of the function. From 20 to 100 function evaluations are required for the cases studied here. Table 1 tabulates the results obtained with unconstrained

[1]Numerical Algorithm Group, 7 Banbury Road, Oxford OX2 6NN, UK

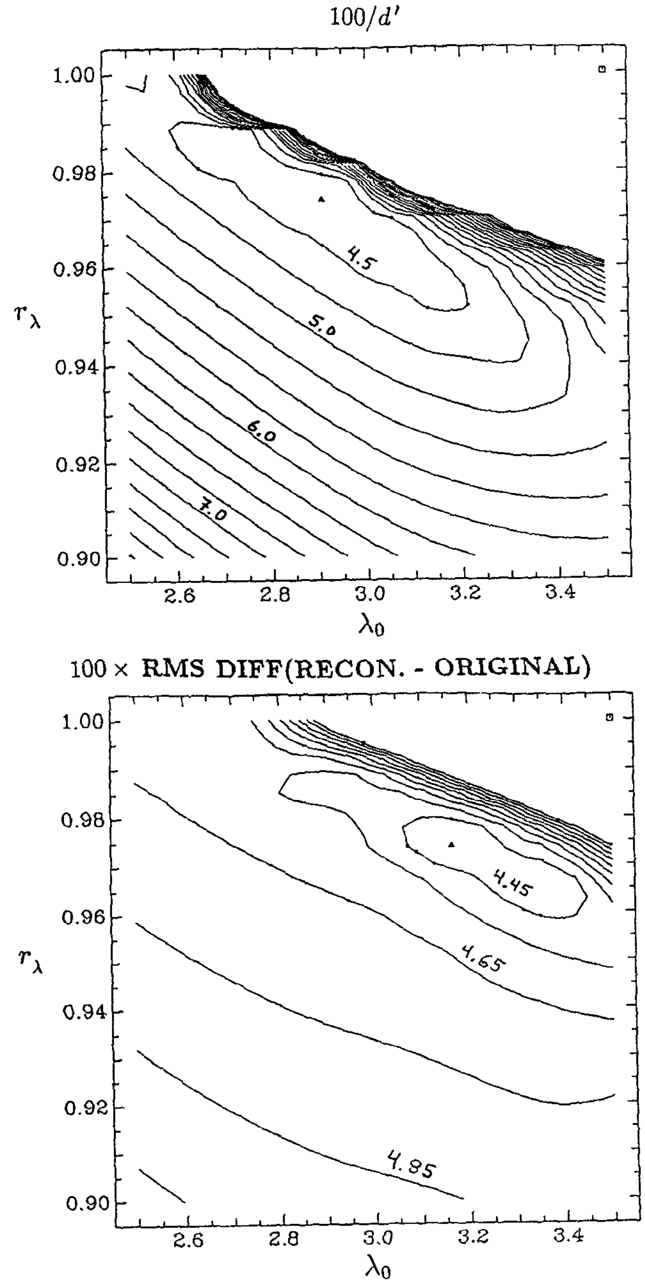

Figure 4: Contour plots of two optimization functions plotted as a function of the relaxation parameters λ_0 and r_λ used in the constrained ART reconstruction algorithm. The measurements consist of 12 noiseless, parallel projections spanning 180°. The coarse sampling (10 x 10 points) of these functions, necessitated by the lengthy computation time required for each function evaluation, accounts for the scalloping effects.

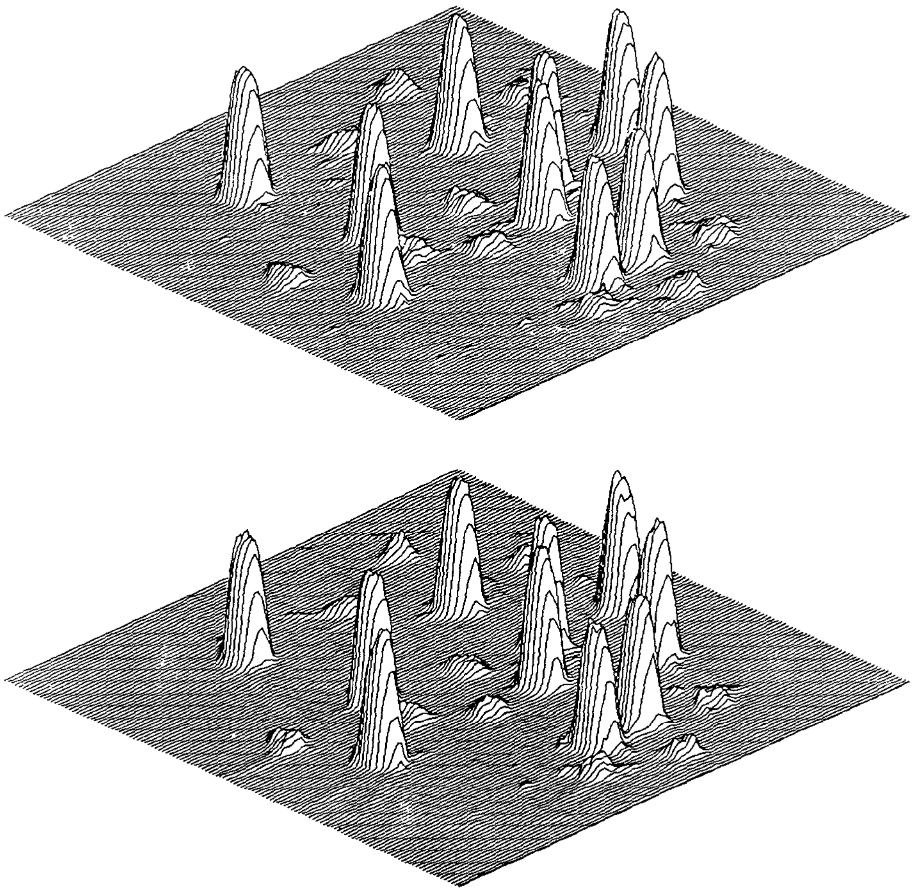

Figure 5: Optimized reconstructions of Fig. 1 from 12 noiseless parallel projections subtending 180° obtained with constrained ART. The reconstruction on the top is obtained with $\lambda_0 = 2.96$ and $r_\lambda = 0.975$, which is the optimum for detectability. The reconstruction on the bottom is obtained with $\lambda_0 = 3.25$ and $r_\lambda = 0.975$, which produces the smallest rms difference between the reconstruction and the original image. Although the rms error in the reconstruction is a common measure for the quality of reconstruction, it yields more visible artifacts and reduces d' from its optimum of 23.5 to 12.6.

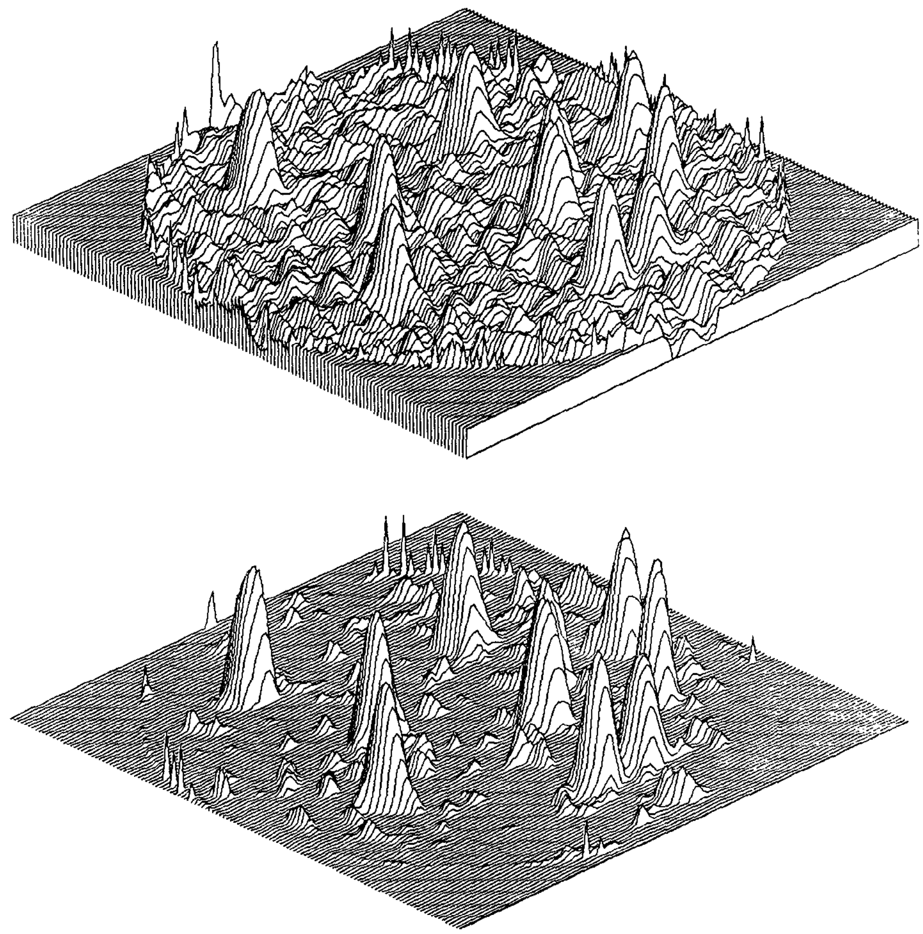

Figure 6: Reconstructions of Fig. 1 from 100 noisy parallel projections subtending 180° obtained with the ART algorithm (top) without and (bottom) with the nonnegativity constraint. The noise added to the projection measurements has an rms amplitude of 8, which is ten times the peak projection value for one of the low-contrast discs. These reconstructions were obtained with $\lambda_0 = 0.2$ and $r_\lambda = 0.8$.

Table 1: Summary of the effect of optimization with respect to the detectability index d' on reconstructions obtained using 10 iterations of unconstrained ART. The optimum operating point was found by varying the parameters that control the relaxation factor used in the ART algorithm, λ_0 and r_λ, as discussed in the text. There is generally little improvement in detectability.

number	$\Delta\theta$	rms	nominal			optimized		
proj.	(deg.)	noise	λ_0	r_λ	d'	λ_0	r_λ	d'
100	180	8	0.2	0.8	1.995	0.107	0.820	2.013
8	180	0	1.0	0.8	0.464	0.915	0.463	0.485
12	180	0	1.0	0.8	0.871	0.427	0.729	0.932
16	180	0	1.0	0.8	1.960	1.047	0.998	1.969
16	90	0	1.0	0.8	1.122	1.714	0.993	1.202
16	180	2	1.0	0.8	1.653	2.247	0.635	1.662

Table 2: Summary of the effect of optimization with respect to the detectability index d' on ART reconstructions incorporating the nonnegativity constraint. When the measurement geometry limits the reconstruction rather than noise in the data, dramatic improvement in detectability is seen to be possible.

number	$\Delta\theta$	rms	nominal			optimized		
proj.	(deg.)	noise	λ_0	r_λ	d'	λ_0	r_λ	d'
100	180	8	0.2	0.8	1.825	0.052	0.859	1.908
8	180	0	1.0	0.8	0.653	3.450	0.959	4.91
12	180	0	1.0	0.8	2.054	2.959	0.975	23.46
16	180	0	1.0	0.8	4.782	2.794	0.951	40.13
16	90	0	1.0	0.8	2.050	2.782	0.967	6.30
16	180	2	1.0	0.8	2.372	3.012	0.712	2.747

ART. In most cases relatively little improvement in detectability is achieved by optimization compared to that obtained with the nominal relaxation factors. In the noiseless cases, a value of unity for λ^K yields essentially the same results as the optimized values, a choice that is in agreement with common practice. However, for noisy data it seems desirable for r_λ to be less than unity and, when there are many views, λ_0 should be small. These choices are reasonable as they promote significant averaging over all the views. As a rule of thumb, for noisy but complete data, the relaxation factor should be approximately equal to the reciprocal of the number of views for the last few iterations.

The results of optimizing constrained ART are presented in Table 2. The nonnegativity constraint is seen to be generally useful with the nominal relaxation factors, particularly when the data are limited by the measurement geometry. But with optimization, huge improvements in detectability are obtained in these cases. Very large relaxation factors are preferred, in fact much larger than might be expected. However, when it is realized that the nonnegativity constraint has the effect of undoing the agreement with each measurement that should result from an update, it seems reasonable that overrelaxation is needed. Neither the use of nonnegativity nor the optimization has much benefit when the data are complete but noisy. It is possible that this conclusion depends heavily on the type of task posed and the decision variable adopted for the performance of the detection task. It seems that the task of identification of the discs as separate entities might yield a different conculsion about the value of the nonnegativity constraint when the data are noisy.

Discussion

In some of the imaging situations studied, the use of the nonnegativity constraint in ART significantly increases the detectability of objects, especially when the data consist of a limited number of noiseless projections. Optimization is accomplished by varying the relaxation factor, both in terms of its initial value and the rate of its decline with iteration number. The detectability in the reconstructions obtained with constrained ART is dramatically enhanced by the optimization procedure in some cases. It is found that optimization of ART with respect to conventional measures of reconstruction quality, such as rms difference from the original image, results in reconstructions with more artifacts and lower detectability. For unconstrained ART, little improvement was achieved through optimization.

It is concluded that it is important to optimize image-reconstruction algorithms on the basis of what is most important, which can often be defined in terms of a task that is to be performed using the final image. The approach taken here is based on a Monte Carlo simulation of the complete imaging process from the composition of the original scene to the final interpretation of the reconstructed image. This method is consistent with the assertion that an algorithm can only be properly evaluated by testing it on a statistically meaningful sample of trials in which all the uncontrollable variables in the problem are varied. This numerical simulation technique has several great advantages. It can be used to evaluate the net effect of complex scenes on the reconstructed images. It is particularly useful in situations that do not lend themselves to analytic analysis, as in nonlinear algorithms like constrained ART. It can optimize the performance of iterative algorithms for an arbitrary number of iterations. These issues cannot be addressed directly by theoretical approaches to optimization. The major disadvantage of relying on the Monte Carlo numerical technique is that each result applies only to the specific imaging situation tested and generalizations are seldom possible.

The method for optimizing tomographic reconstruction presented here suggests a way to evaluate and optimize the design of neutron-diffraction spectrometers together with the performance of the required data-unfolding schemes. For a postulated mix of broad and narrow peaks that occur on a variable background, one could determine how well the presence of each peak is detected. To push the technique further, it would be possible to ascertain how well one could estimate the various parameters associated with each peak (amplitude, width, position) from the final reconstructed data. This approach to data evaluation can provide a firm basis upon which to make decisions about spectrometer design.

References

[1] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice-Hall, Englewood Cliffs, New Jersey, 1977.

[2] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *Proc. SPIE* **914**, 336–343, 1988.

[3] K. M. Hanson. POPART - Performance OPtimized Algebraic Reconstruction Technique. *Proc. SPIE* **1001**, 318–325, 1988.

[4] K. M. Hanson. Variations in task and the ideal observer. *Proc. SPIE* **419**, 60–67, 1983.

[5] R. Gordon, R. Bender, and G. Herman. Algebraic reconstruction techniques for three-dimensional electron microscopy and x-ray photography. *J. Theor. Biol.* **29**, 471–481, 1970.

[6] S. Kaczmarz. Angenährte Auflösung von Systemen linearer Gleichungen. *Bull. Acad. Polon. Sci. Lett.* **A35**, 355–357, 1937.

[7] K. Tanabe. Projection method for solving a singular system of linear equations and its applications. *Numer. Math.* **17**, 203–214, 1971.

[8] G. T. Herman, A. Lent, and P. H. Lutz. Relaxation methods for image reconstruction. *Commun. Assoc. Comput. Mach.* **21**, 152–158, 1978.

[9] Y. Censor, P. P. B. Eggermont, and D. Gordon. Strong underrelaxation in Kaczmarz's method for inconsistent systems. *Numer. Math.* **41**, 83–92, 1983.

[10] A. D. Whalen. *Detection of Signals.* Academic, New York, 1971.

[11] K. M. Hanson. Detectability in computed tomographic images. *Med. Phys.* **6**, 441–451, 1979.