



# Optimization of the Constrained Algebraic Reconstruction Technique for the Performance of a Variety of Visual Tasks

K. M. Hanson\*  
Los Alamos National Laboratory, MS P940  
Los Alamos, NM 87545 USA

This work demonstrates the application of a method to optimize image reconstruction algorithms on the basis of the performance of specific visual tasks that are to be accomplished using the reconstructed images. The evaluation of task performance is numerically realized by a Monte Carlo simulation of the complete imaging chain, including the final inference based on the reconstructions. Fundamental to this evaluation is that it yields an average response by consideration of many initial scenes. It is shown that the use of the nonnegativity constraint in the Algebraic Reconstruction Technique can significantly improve performance in situations where there is a severe lack of measurements when the relaxation factor is optimized. There is no indication in any of the cases studied hitherto that the nonnegativity constraint can improve performance in situations where the data are complete, but noisy.

## INTRODUCTION

The value of an image is ultimately related to how well it enables one to perform a desired visual interpretation. This dictum is evident in the clinical setting, where imaging systems are to be judged on the basis of their diagnostic capabilities. To meet the need for a valid evaluation of image processing algorithms, I explored [Hanson, 1988a] a method to numerically evaluate image-recovery algorithms on the basis of how well the resulting reconstructions allow one to perform specified visual tasks. The capabilities of this approach were demonstrated by studying the effect of the nonnegativity constraint on the performance of the Algebraic Reconstruction Technique

---

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

(ART) [Gordon et al., 1970], an iterative tomographic reconstruction algorithm, in a specific imaging problem. In the first study [Hanson, 1988a] the task was simple detection of small discs and the performance index was the detectability index  $d'$ . The nonnegativity constraint was found to improve detection for an insufficiency of data by about a factor of three. Optimization of constrained ART by a proper choice of the relaxation factor resulted in even better detectability [Hanson, 1988b], gaining about another factor of ten. A number of other high-order visual tasks are interesting. The value of considering such high-order visual tasks is that they rely on the higher spatial frequencies in the reconstruction, which are believed to dominate the performance of other high-order tasks such as medical diagnosis [Hanson, 1983]. Optimization of ART with respect to one of these, estimation of object location, has already been reported [Hanson, 1989].

In this work I extend the range of visual tasks to include another high-order task, the discrimination of binary pointlike objects from a single object. I also investigate the effect on object detection when the background is not known *a priori*. This task is interesting because it is conjectured that the human observer cannot incorporate the known background level into his decision in an absolute way and so must estimate the background locally. In this presentation I summarize the results of optimization of the constrained ART algorithm with respect to the performance of the above visual tasks and compare them with standard measures of the fidelity of reconstructed images.

## METHOD

The overall method for evaluating a reconstruction algorithm used here has been described before [Hanson, 1988a]. In this method one numerically evaluates a task performance index for a specified imaging situation. This technique consists of a Monte Carlo simulation of the entire imaging process including random scene generation, data taking, reconstruction, and performance of the specified task. The accuracy of the task performance is determined by comparison of the results with the known original scene using an appropriate figure of merit. Repetition of this process for many randomly generated scenes provides a statistically significant estimate of the performance index. When the ability to perform a task is marginal, task performance is inherently a statistical question. This is clearly true when data are significantly degraded by random noise. But, it also is often true in measurement geometries in which the data are limited. The artifacts produced by the ambiguities arising from the lack of data depend on the scene. As the scenes vary in an uncontrollable manner, so do the artifacts. Interpretation of the image can thus vary from one scene to the next in a random fashion.

An advantage of the Monte Carlo method is that the reconstruction al-

gorithm may be optimized for any fixed number of iterations. Alternatively, the number of iterations may be varied to achieve the optimum performance for algorithms that tend to diverge after many iterations. Such behavior is observed [Veklerov and Llacer, 1987] in the EM (Estimation Maximization) algorithm and, when confronted by inconsistent data, some implementations of ART [Gordon et al., 1970].

Most of the tasks addressed in this paper are based on the estimation of parameters associated with an assumed object. The parameter values are estimated for a given set of data by using a minimum chi-squared ( $\chi^2$ ) fitting procedure, which essentially finds the parameter set that best match the data. This procedure is equivalent to maximum likelihood estimation [Hanson, 1984] when the measurement noise is gaussian distributed and uncorrelated. For measurements of constant noise variance, this procedure amounts to minimizing the mean square residuals, known simply as least-squares estimation. The fitting algorithm I use is essentially identical to the CHIFIT program presented by Bevington [Bevington, 1969] for fitting a nonlinear function of the parameters. Nonlinear estimation is necessary because the function describing an object is nonlinearly related to its position and its width.

The Algebraic Reconstruction Technique (ART) [Gordon et al., 1970] is an iterative algorithm that reconstructs a function from its projections. To define the relaxation parameters to be discussed shortly, a brief description of the algorithm follows. Assume that  $N$  projections of the unknown function  $f$  are measured. Considering these measurements to be a vector, the  $i$ th measurement is written as

$$g_i = H_i f, \quad i = 1, \dots, N, \quad (1)$$

where  $H_i$  is the corresponding row of the measurement matrix. An initial guess is made; for example,  $f^0 = 0$ . In the ART algorithm the estimate is updated by iterating on the individual measurements one at a time:

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

where  $f^k$  is the  $k$ th estimate of the image vector  $f$ ,  $i = k \bmod(N) + 1$ , and  $\lambda^k$  is a relaxation factor for the  $k$ th update. In constrained ART, a non-negativity constraint is enforced by setting any component of  $f^{k+1}$  to zero that has been made negative by the above updating procedure. The index  $K$  indicates the iteration number ( $K = \text{int}(k/N)$ ), which, in the standard nomenclature, corresponds to one pass through all  $N$  measurements. Ten iterations were used to obtain all the results reported here. The relaxation factor is expressed as

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

The proper choice of the relaxation parameters  $\lambda_0$  and  $r_\lambda$  is the crux of optimization of the ART algorithm. The literature offers meager guidance as to the choice of the relaxation factor [Hanson, 1988b].

## RESULTS

In previous publications I dealt with scenes composed of twenty randomly placed nonoverlapping discs, ten with an amplitude of 1.0 and ten with 0.1, superimposed on a background of zero. The discs were all 8 pixels in diameter and the size of the reconstruction was 128 x 128. Each evaluation of the reconstruction algorithm involved averaging the response to ten different, but similar scenes. In the first study [Hanson, 1988b] the focus was on the simple detection task in which it is assumed that the observer knows the position and shape of the object as well as the background. The decision variable was taken to be simply the average value of the reconstruction calculated over the area of the disc. The relaxation factor parameters were optimized with respect to the detectability index  $d'_{\text{DET}}$ , which is based on the two histograms of the decision variable for regions where the objects are known to be present and separately for regions where there is no object. The optimization of constrained ART resulted in a major increase in  $d'_{\text{DET}}$  as well as much improved apparent image quality.

In another study [Hanson, 1989] I looked at a different type of task, namely object localization. In that study, the position of the discs was estimated using the minimum  $\chi^2$  fitting procedure. The reconstruction values in circular regions surrounding each known disc were used as the input data. The diameter of the circular fit regions was chosen to be 14 pixels. The function used to fit these values was a disc with slightly tapered edges. The amplitude and the position of the disc were allowed to vary in the fit while the disc diameter and edge taper were fixed. The background was assumed to be zero. The rms error in the position of the discs,  $\sigma_\Delta$ , was used as the optimization function. It is seen in Fig. 1 that this function is very similar to  $1/d'_{\text{DET}}$  for both the low-contrast and high-contrast discs. It is noted however that the position uncertainty for the high-contrast discs does not deteriorate as quickly as one approaches the upper righthand corner as does that for the low-contrast discs.

I have recently addressed the task of detection of the discs when the background is unknown. This task is easily handled within the framework of the minimum  $\chi^2$  procedure by varying the amplitude of the disc and the value of an assumed flat background in the fitting procedure. The optimum choice for the diameter of the fit region is found to be 14 pixels for the type of scenes used. The resulting estimated amplitude of the disc is taken to be the proper decision variable. It is seen in Fig. 1 that the detectability index derived from this analysis,  $d'_{\text{AMPL}}$ , has a dependence on the relaxation parameters similar to that of  $d'_{\text{DET}}$ , but that it is offset by a substantial

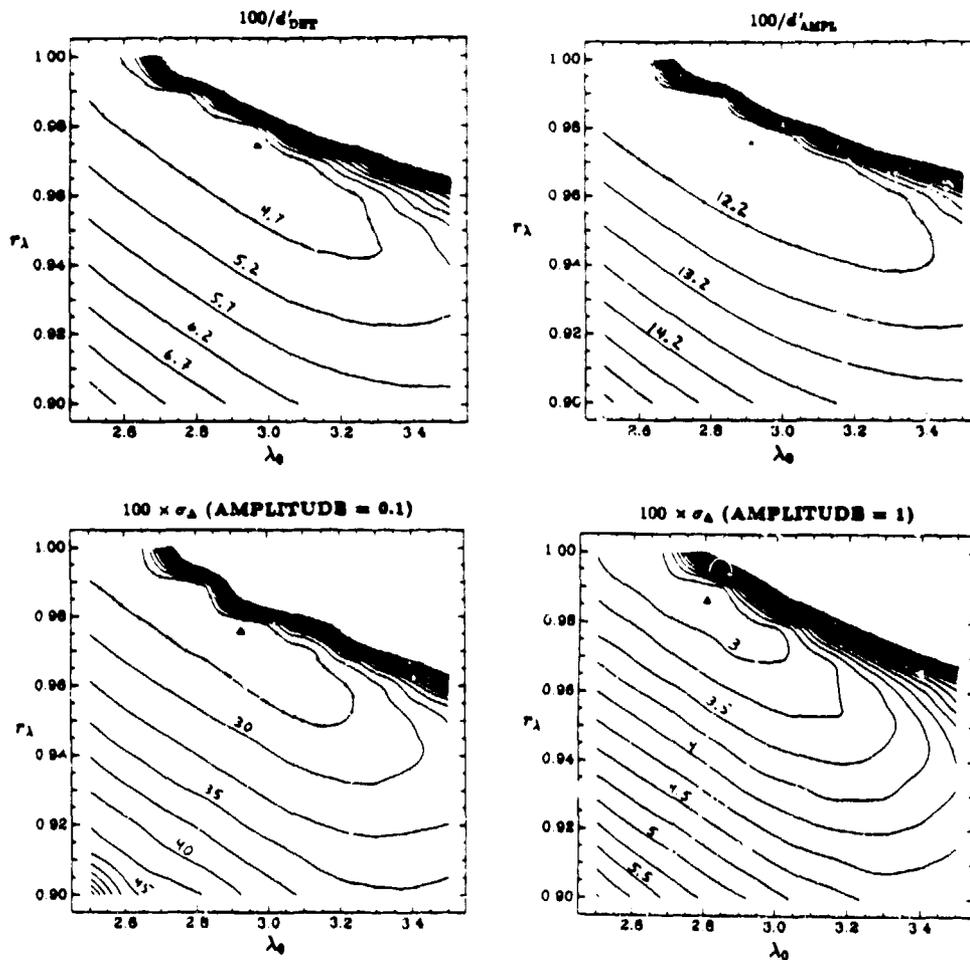


Figure 1: Contour plots of four optimization functions based on task performance plotted as a function of the relaxation parameters  $\lambda_0$  and  $r_\lambda$  used in 10 iterations of the constrained ART reconstruction algorithm. The measurement data consist of 12 noiseless, parallel projections spanning  $180^\circ$  and the source images consist of 10 high-contrast and 10 low-contrast discs, all randomly placed. The functions are  $d'_{\text{DET}}$ , the detectability index for the low-contrast discs;  $d'_{\text{AMPL}}$ , the detectability index based on fitting the amplitude of the low-contrast discs together with a flat background; and  $\sigma_\Delta$ , the rms uncertainty in the position of the discs (shown for both low- and high-contrast discs). The coarse sampling ( $10 \times 10$  points) of these functions, necessitated by the lengthy computation time required for each function evaluation, accounts for the slight scalloping effects.

amount. The lack of *a priori* knowledge of the background decreases detectability at the minimum by about a factor of 2.6. One might expect a worsening of the detectability in this case, but not by such a large amount. On the basis of stationary uncorrelated noise, one expects about a 20% decrease in detectability caused by the necessity to estimate the background. A large part of the observed degradation comes from the contribution to  $d'_{\text{AMPL}}$  that arises from sampling the background of the reconstruction. The size of the suggested values of the relaxation factors is perhaps surprising since values near unity are most often used. However, this is explained by the observation that the nonnegativity constraint thwarts any negative updates made in regions where the reconstruction is already zero. Thus more emphasis must be given to the updates to make them effective in matching the measurements.

Figure 2 shows contour plots for several standard measures of the quality of reconstructions. Two of these functions are measures of the vector-space distance between the reconstruction and the original image; the rms error is based on the L2 norm of the difference (the square root of the mean squared value of the pixel differences) and the L1 error is based on the L1 norm (the mean value of the absolute differences). The rms error gives more emphasis to large deviations than the L1 error. The third function measures how closely the projections of the reconstruction match the data. It is called here the rms residual (the rms value of the difference between the data and the projections of the reconstruction), which is not directly related to the reconstruction per se.

All the optimization functions displayed in Figs. 1 and 2 indicate the upper righthand corner of the  $\lambda_0 - r_\lambda$  space is undesirable. However, careful inspection of the plots reveals that both the rms error and L1 error permit operation much closer to the corner than do any of the task performance indices. It was shown that an operating point chosen to minimize the rms error resulted in more artifacts and considerably poorer detectability  $d'_{\text{DET}}$  than optimization based on  $d'_{\text{DET}}$  itself [Hanson, 1988b]. The additional artifacts were not very large, but they had a significant effect on the detectability of the low-contrast discs. From the plots, it is observed that optimization with respect to the L1 error would not degrade the detection of the low-contrast discs very much. In fact, it has been suggested [Yeung and Hermar, 1989] on the basis of a study very similar to this one that the L1 error might be better than the rms error for judging the quality of reconstructions. However, such a conclusion cannot be drawn on the basis of a single test. There are a number of reasons to be suspicious of the L1 error, as will be pointed out below.

The point of using task performance as a basis for the evaluation of reconstructed images is that it focuses on what is important to accomplish the desired goal. Summary measures of image reconstruction quality, such as rms error and L1 error, are not specific enough to be good indicators of how

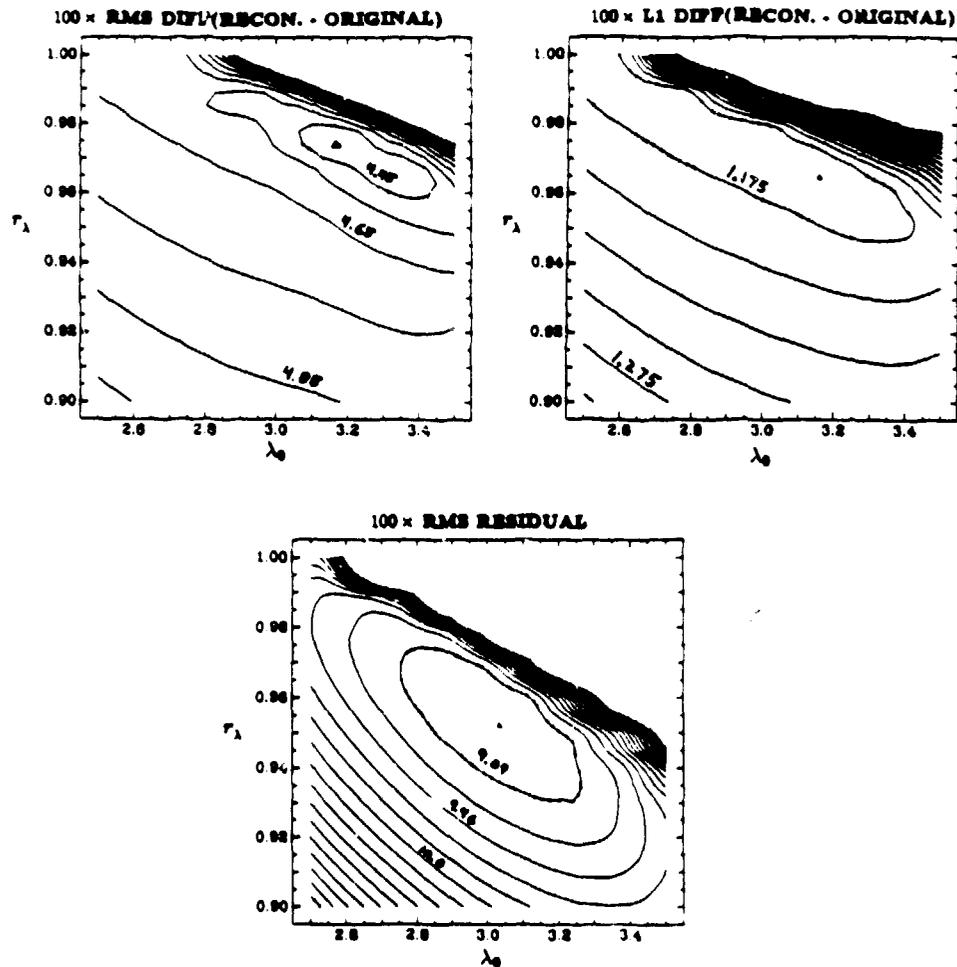


Figure 2: Contour plots of three functions often used to measure the quality of reconstructions. These functions were generated for the same reconstructions employed in Fig. 1. The measurement data consist of 12 noiseless, parallel projections spanning  $180^\circ$ . Although these functions generally have characteristics similar to those based on task performance, they differ in essential details.

well one can perform specific tasks. For example, in the present case dealing with objects of different amplitudes, detection of the low-contrast discs is very sensitive to slight artifacts in the background regions. On the other hand, the rms error and L1 error can easily be overpowered by the rather large errors associated with and surrounding the high-contrast objects. Since these regions are not of interest for detection of the low-contrast discs, the irrelevance of these measures to this task is obvious. In their favor, it may be said that the rms error and L1 error are easily calculated. Also when reconstructions are dominated by random noise and not by artifacts, in some situations it is possible to relate the accuracy of parameter estimation to the rms error using the standard method to propagate errors.

I next consider the task of binary discrimination, which is a revised version of the famous Raleigh criterion. The task is to determine whether a detected pointlike object is a single star or a binary pair. Figure 3 shows the first test image in the sequence used to evaluate the performance of this task. Each scene consists of eight binary objects and eight single objects. The binary objects consist of two 2D gaussian functions, each with a full-width at half maximum (fwhm) of 4 pixels and an amplitude of 0.5. The separation parameter, defined as the distance between the two gaussians divided by their fwhm width in the direction of their separation, is 1.5. The singlets are asymmetric gaussians with a width of 10 pixels (fwhm) along the principal axis and a width of 4 pixels (fwhm) in the orthogonal direction. Their amplitude is 0.42. These parameters provide the best match to the doublets specified above as determined by fitting the doublets with a single gaussian. The principal axes and the positions of these objects are randomly chosen, with the only restriction being that the objects not be too close to each other. The task is to decide whether the objects are singlets or doublets without regard for asymmetries in object shape. This criterion requires that there be ample evidence of two peaks to call it a binary. The method used to perform this task is based on the minimum  $\chi^2$  fit to circular regions, 20 pixels in diameter, centered on the original objects. In the fit, the amplitude, the position, the angle of the principal axis, the width along and perpendicular to the principal axis, the value of the flat background and the separation parameter are all varied. Since there may be multiple minima in this eight dimensional space, especially in the direction of the tilt angle, each data set is fit six different times with randomly chosen initial angles. In half of these trials, the separation parameter is set to zero and in the other half, to a value of 1.5 to adequately sample both possibilities. The separation parameter is used as the decision variable to calculate the binary discrimination index  $d'_{\text{BIN}}$  in the same way as is done for the detectability index [Hanson, 1988a].

The reconstructions obtained with 10 iterations of unconstrained and constrained ART for 12 parallel, noiseless projections covering  $180^\circ$  are shown in Fig. 3. The nonnegativity constraint brings about a dramatic reduction in the background clutter caused by the artifacts emanating from

Figure 3: The first image (top) in the sequence of 10 trials used to test binary discrimination. The reconstruction (bottom left) obtained with 10 iterations of the unconstrained ART reconstruction algorithm for relaxation parameters  $\lambda_0 = 1.0$  and  $r_\lambda = 0.80$  and that obtained with constrained ART for  $\lambda_0 = 3.17$  and  $r_\lambda = 0.97$  (bottom right) both for measurement data consisting of 12 noiseless, parallel projections spanning  $180^\circ$ . In the latter case, the relaxation parameters represent those that yield the largest binary discriminability,  $d'_{\text{BIN}} = 5.04$ . For the unconstrained reconstructions,  $d'_{\text{BIN}} = 0.89$ , which is relatively insensitive to the relaxation parameters.

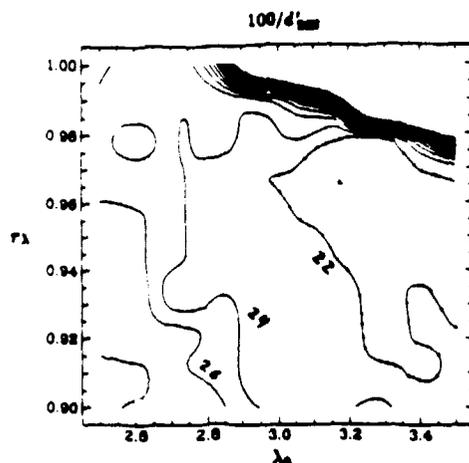


Figure 4: Contour plot of the optimization function based on binary discrimination. The measurement data consist of 12 noiseless, parallel projections spanning  $180^\circ$ . The irregular behavior is caused by the random search used in the minimum  $\chi^2$  procedure to find the global minimum, as explained in the text.

the objects themselves. It also cleans up the objects themselves, enhancing their definition as binary or single objects. The binary discrimination index shows an improvement of more than a factor of five. The nonnegativity constraint leads to a decrease in the rms error of from 0.040 to 0.014 and likewise in the L1 error of from 0.023 to 0.005. The rms residual, however, increases from 0.026 to 0.108. This increase is not unreasonable as the ART algorithm, under certain circumstances, provides the least square residual solution with minimum norm [Censor et al., 1983]. If this is the case, the addition of a side constraint can only increase the rms residual.

Figure 4 shows the contour plot for the binary discrimination index. Because the scenes are chosen to have roughly the same fraction of its area nonzero as in the disc images used in the earlier studies, it is expected that the optimization map for the binary discrimination problem will be similar. Indeed this expectation is more or less fulfilled. The lack of smoothness in this function is attributed to the random starting values of the parameters used in the fitting procedure.

Figure 5 shows reconstructions obtained from plentiful but noisy data. Gaussian-distributed random numbers are added to the projections to simulate the effects of noise. The rms value of the noise is 4. For comparison, the peak amplitude of the projections of one of the individual gaussians is about 1.8. At first, the constrained reconstruction appears to be much better than the unconstrained one. However, closer examination reveals that although the noise in the background region of the reconstruction is drasti-

Figure 5: The reconstruction (left) obtained with 10 iterations of the unconstrained ART reconstruction algorithm and the unconstrained reconstruction (right) both for relaxation parameters  $\lambda_0 = 0.2$  and  $r_\lambda = 0.80$  and for measurement data consisting of 100 parallel projections spanning  $180^\circ$  with additive gaussian noise (rms value = 4). For the 10 unconstrained reconstructions,  $d'_{\text{BIN}} = 2.59$  and for the constrained reconstructions,  $d'_{\text{BIN}} = 2.56$ .

cally reduced by the nonnegativity constraint, there is little change in the shape and amplitude of the objects. The Monte Carlo evaluation (based on ten scenes) finds the binary discriminability for the constrained reconstructions is essentially identical to that for the unconstrained ones. This conclusion mirrors what was observed for the other visual tasks under the same data-taking conditions. For example, detection of low-contrast discs was not improved despite the apparent decrease in the background noise. The lack of improved performance is contrary to the observed decrease in the rms error from 0.048 to 0.026, and the even larger relative decrease in the L1 error from 0.032 to 0.010 that attends the invocation of the nonnegativity constraint. Thus, to use either of these as a measure of reconstruction quality would be grossly misleading. The reason why task performance does not follow the rms error in this case is undoubtedly related to the introduction of the nonlinear constraint, which disrupts the properties of the noise. The nonnegativity constraint has the effect of increasing the rms residual slightly from 1.77 to 1.91.

## SUMMARY

I have shown that in imaging situations in which the data are incomplete, the optimization of constrained ART realized through a judicious selection of

the relaxation factor can significantly improve the performance of a variety of visual tasks based on the reconstructions. The nonnegativity constraint yields no improvement in task performance when the data are complete, but noisy. For unconstrained ART, little improvement can be achieved through optimization.

The optimization functions based on the performance of all the visual tasks studied to date show similar trends as a function of the two relaxation parameters  $\lambda_0$  and  $\tau_\lambda$ . The optimum operating points in terms of these parameters vary with the data-taking situation, but they are nearly the same for both of these tasks. From the contour plots presented here, it is clear that an optimization function that consisted of a weighted sum of all the tasks presented here would have a broad valley, and the optimum operating point obtained would not much affect the performance of any individual task.

The use of either the rms error or the L1 error as a basis for the evaluation of reconstruction algorithms seems unwise. Optimization of these, particularly the rms error, leads to reconstructions that are undesirable for the purpose of performing some kinds of tasks. The nonnegativity constraint decreases the magnitudes of both of these errors. However, in situations in which confusion in the reconstruction is caused by measurement noise and not artifacts, the performance of visual tasks is not improved. Hence, these measures of image quality may be misleading.

## ACKNOWLEDGEMENTS

I am indebted to the following colleagues for their continued interest in this work and for their many astute comments: H. H. Barrett, K. J. Myers, and R. F. Wagner

## References

- [Bevington, 1969] R. R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York.
- [Censor et al., 1983] Y. Censor, P. P. B. Eggermont, and D. Gordon. Strong underrelaxation in Kaczmarz's method for inconsistent systems. *Numer. Math.*, 41:83-92.
- [Gordon et al., 1970] 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.
- [Hanson, 1983] K. M. Hanson. Variations in task and the ideal observer. *Proc. SPIE*, 419:60-67.

- [Hanson, 1984] K. M. Hanson. Optimal object and edge localization in the presence of correlated noise. *Proc. SPIE*, 454:9-17.
- [Hanson, 1988a] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *Proc. SPIE*, 914:336-343.
- [Hanson, 1988b] K. M. Hanson. POPART - Performance OPTimized Algebraic Reconstruction Technique. *Proc. SPIE*, 1001:318-325.
- [Hanson, 1989] K. M. Hanson. Optimization for object localization of the constrained algebraic reconstruction technique. *Proc. SPIE*, 1090.
- [Veklerov and Llacer, 1987] E. Veklerov and J. Llacer. Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Trans. Med. Imaging*, MI-6:313-319.
- [Yeung and Herman, 1989] K. T. D. Yeung and G. T. Herman. Objective measures to evaluate the performance of reconstruction algorithms. *Proc. SPIE*, 1092:326-335.

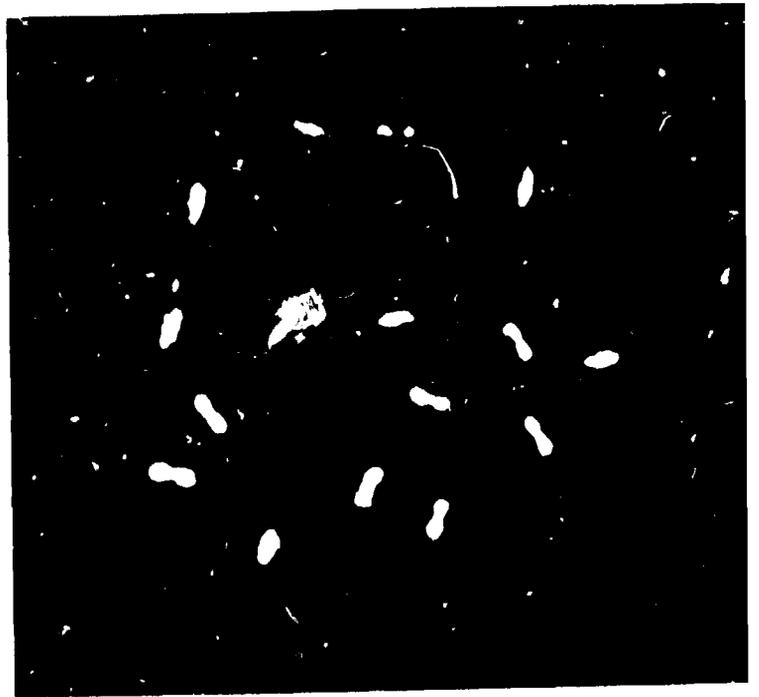


Fig. 3

REPRODUCED FROM BEST  
AVAILABLE COPY

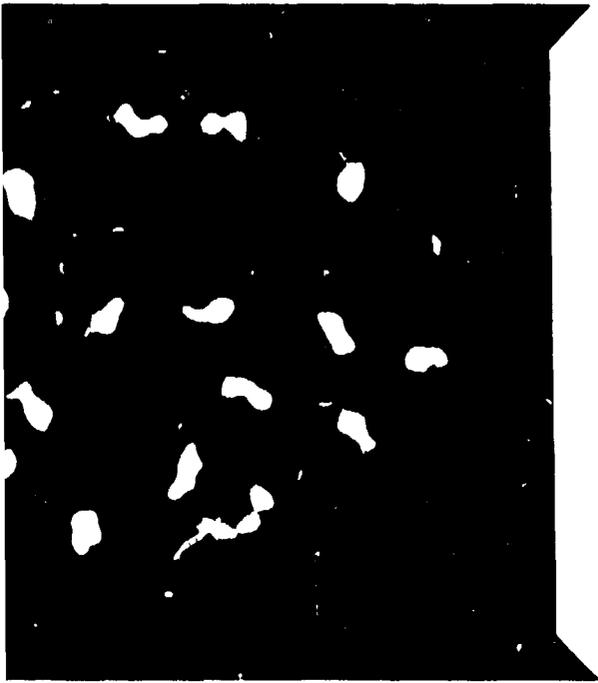


Fig 5

REPRODUCED FROM BEST  
AVAILABLE COPY