Image reconstruction of dense scattering media from CW sources using constrained CGD and a matrix rescaling technique

Jenghwa Chang[1], Harry L. Graber[2], Randall L. Barbour[1,2]

Departments of Pathology[1], and Physiology and Biophysics[2]
SUNY Health Science Center at Brooklyn
450 Clarkson Ave.,
Brooklyn, NY 11203

ABSTRACT

This study reports on results of our efforts to improve the efficiency and stability of previously developed reconstruction algorithms for optical diffusion tomography. The previous studies, which applied regularization and *a priori* constraints to iterative methods - POCS, CGD, and SART algorithms – showed that in most cases, good quality reconstructions of simply structured media were achievable using a perturbation model. The CGD method, which is the most efficient of the three algorithms, was, however, in some instances not able to produce good quality images because of the difficulty in applying range constraints, which can cause divergence. In this study, a scheme is proposed to detect this divergence while imposing range constraints on the previous reconstruction. Once divergence is detected, the conjugate gradient vector is reset and the CGD reconstruction is restarted using the previous reconstruction as the initial value. In addition, a rescaling technique, which rescales the weight matrix to make it more uniform and less ill-conditioned, is also used to suppress numerical errors and accelerate convergence. Two criteria, rescaling the maximum of each column to 1 and rescaling the average of each column to 1, were applied and compared to results without rescaling. The results show that, with properly imposed constraints, good quality images can be obtained using the CGD method. The convergence speed is much slower when constraints are imposed, but still comparable to the POCS and SART algorithms. The rescaling technique produces more stable and more accurate reconstructions, and speeds up the reconstruction significantly for all three algorithms.

1. INTRODUCTION

In the previous studies [1, 2], we have successfully demonstrated the use of perturbation theory for optical diffusion tomography. The inverse problem involves solving a system of linear equations, where the unknowns are the perturbed absorption and scattering cross sections, and the coefficients, called the weight function, are the angular integration of the product of the forward and adjoint solutions of the transport equation [1, 3]. Many methods have been developed to solve systems of linear equations for medical imaging [4-6]. The developed strategies includes direct matrix inversion (Gaussian elimination), singular value decomposition (SVD), Fourier slice theorem and Fourier diffraction theory, and iterative methods. Direct matrix inversion and the SVD method are very efficient for small size problems, but are not suitable for large size matrix inversions and are rarely used in practical image reconstruction. The filtered backprojection algorithm, which uses the Fourier slice theorem and applies the fast FFT algorithm, is currently the most popular reconstruction algorithm used in CT imaging. However, this algorithm require a parallel/fan beam illumination scheme and assumes that the propagating wave front is not significantly perturbed while penetrating through the target medium. This seems impractical in optical imaging because a point source and detector are preferred, and the targets are dense scattering media. The iterative methods, which iteratively update the reconstructed images according to the detected signal and *a priori* information, are well suited for large scale inverse problems because they can readily accommodate any source-detector configuration. Moreover, unlike the direct matrix inversion and SVD methods, useful images can be obtained even after a few iterations.

Previously, we have adopted three iterative algorithms - projection onto convex sets (POCS) [7], conjugate gradient descent (CGD) [8, 9], and simultaneous arithmetic reconstruction technique (SART) [5, 10], for image reconstruction, and have applied positivity constraints. The POCS algorithm is a sequential method, whereas CGD and SART evaluate multiple projections simultaneously. In previous studies we have shown that the CGD method has difficulty incorporating a positivity constraint [2]. The CGD method iteratively updates the reconstruction on the basis of the previous

0-8194-1736-X/95/$6.00

reconstruction and all the preceding conjugate gradient vectors. In some cases, applying constraints to the previous reconstruction can lead to miscalculation of the conjugate gradient vector and result in divergence.

In this study, a scheme was developed to detect this divergence and a rescaling technique was also used to speed up the reconstruction. The convergence behavior and the reconstruction results obtained from the constrained CGD, unconstrained CGD, as well as other algorithms are compared.

2. THEORY

1.1 Perturbation model

The perturbation model [1, 3, 11] adopted for optical diffusion tomography relates differences between the absorption and scattering cross sections of target and reference media, $\Delta\mu_{aj} = \mu_{aj} - \mu_{aj}^{r}$, $\Delta\mu_{sj} = \mu_{sj} - \mu_{sj}^{r}$, to changes in detector readings, $\Delta R_i = R_i - R_i^{r}$, by the following first order approximation:

$$\sum_j w_{aij}\Delta\mu_{aj} + \sum_j w_{sij}\Delta\mu_{sj} = \Delta\phi_i, \quad i = 1,2,\ldots,I. \tag{1}$$

Here, I is the total number of source-detector pairs, and w_{aij} and w_{sij} are called the *weight*, which is proportional to the angular integration of the product of the direct (*i.e.*, photons propagating from the source to voxel *j*) and adjoint solutions (*i.e.*, from the detector to voxel *j*). The above linear equation can be represented in matrix form as:

$$\Delta\mathbf{R} = \mathbf{W}\cdot\Delta\boldsymbol{\mu}, \tag{2}$$

where

$$\Delta\mathbf{R} = \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \vdots \\ \Delta R_I \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_{2J} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1,2J} \\ w_{21} & w_{22} & \cdots & w_{2,2J} \\ \vdots & \vdots & \ddots & \vdots \\ w_{I1} & w_{I2} & \cdots & w_{I,2J} \end{bmatrix}, \quad \Delta\boldsymbol{\mu} = \begin{bmatrix} \Delta\mu_{a1} \\ \vdots \\ \Delta\mu_{aJ} \\ \Delta\mu_{s1} \\ \vdots \\ \Delta\mu_{sJ} \end{bmatrix},$$

where J is the total number of voxels.

1.2 Iterative method

The least squares solution to a system of linear equations is obtained by iteratively modifying the unknowns, $\Delta\boldsymbol{\mu}$, to minimize the mean squared errors:

$$E = \frac{1}{2}\cdot(\mathbf{W}\cdot\Delta\boldsymbol{\mu} - \Delta\boldsymbol{\phi})^T\cdot(\mathbf{W}\cdot\Delta\boldsymbol{\mu} - \Delta\boldsymbol{\phi}) = \frac{1}{2}\cdot\Delta\boldsymbol{\mu}^T\cdot\mathbf{A}\cdot\Delta\boldsymbol{\mu} - \mathbf{b}^T\cdot\Delta\boldsymbol{\mu} + \frac{1}{2}\cdot\Delta\boldsymbol{\phi}^T\cdot\Delta\boldsymbol{\phi}, \tag{3}$$

where $\mathbf{A} = \mathbf{W}^T\cdot\mathbf{W}$, and $\mathbf{b} = \mathbf{W}^T\cdot\Delta\mathbf{R}$. Any vector, $\Delta\boldsymbol{\mu}$, minimizing E is a least squares solution. To minimize E, the derivative of E is calculated and set equal to 0:

$$\mathbf{g}(\Delta\boldsymbol{\mu}) = \frac{\partial}{\partial\Delta\boldsymbol{\mu}}E = \mathbf{A}\cdot\Delta\boldsymbol{\mu} - \mathbf{b} = 0 \tag{4}$$

where $\mathbf{g}(\Delta\boldsymbol{\mu}) = \mathbf{A}\cdot\Delta\boldsymbol{\mu} - \mathbf{b}$ is the gradient of E. Three iterative algorithms - POCS [7], CGD [8, 9], and SART [5, 10], previously described, were used for image reconstruction. Different combinations of regularization constraints were also used: (1) unconstrained, (2) positive range constraint on reconstruction results, (3) positive range constraint on detector readings and reconstruction results. Range constraints on the $\Delta\mathbf{R}$ were imposed prior to reconstruction, with all the negative readings set to zero:

$$\Delta R_t = 0 \text{ if } \Delta R_t < 0. \tag{5}$$

The range constraints on reconstruction results were imposed after each iteration:

$$\Delta \mu_j^n = 0 \text{ if } \Delta \mu_j^n < 0. \tag{6}$$

1.3 Constrained CGD

The CGD method iteratively updates the reconstruction on the basis of the previous reconstruction, and all the preceding gradient and conjugate gradient vectors:

$$\begin{aligned} &\mathbf{g}^0 = \mathbf{A}\cdot\Delta\boldsymbol{\mu}^0 - \mathbf{b}, \quad \beta^1 = 0, \quad \mathbf{d}^1 = -\mathbf{g}^0, \\ &\beta^n = \frac{\left\|\mathbf{g}^{n-1}\right\|^2}{\left\|\mathbf{g}^{n-2}\right\|^2}, \quad \mathbf{d}^n = -\mathbf{g}^{n-1} + \beta^n \mathbf{d}^{n-1}, \\ &\alpha^n = \frac{\mathbf{g}^{n-1^T}\cdot\mathbf{g}^{n-1}}{\mathbf{d}^{n^T}\cdot\mathbf{A}\cdot\mathbf{d}^n}, \quad \Delta\boldsymbol{\mu}^n = \Delta\boldsymbol{\mu}^{n-1} - \alpha^n\mathbf{d}^n, \\ &\mathbf{g}^n = \mathbf{A}\cdot\Delta\boldsymbol{\mu}^n - \mathbf{b} = \mathbf{g}^{n-1} - \alpha^n\cdot\mathbf{A}\cdot\mathbf{d}^n. \end{aligned} \tag{7}$$

where, $\Delta\boldsymbol{\mu}^0$ is the initial guess, $\mathbf{g}^n$ is the gradient vector and $\mathbf{d}^n$ is the conjugate gradient vector. β^n is the conjugate gradient choice which makes $\mathbf{d}^n$ **A**-orthogonal; *i.e.*, the weighted inner-product, $\mathbf{d}^{n^T}\cdot\mathbf{A}\cdot\mathbf{d}^m = 0$ if $n \neq m$. α^n is called the step–size which makes $\mathbf{g}^n$ orthogonal; *i.e.*, $\mathbf{g}^{n^T}\cdot\mathbf{g}^m = 0$ if $n \neq m$. Theoretically, this algorithm should converge after a number of iterations less than or equal to, N, the number of unknowns, since $\{\mathbf{d}^n,\ n = 1, 2, \ldots, N\}$ expands the N-dimension solution space. In practice, round-off error may prevent convergence so that a convergence criteria is needed.

Applying positive range constraints to the previous reconstruction, $\Delta\boldsymbol{\mu}^{n-1}$, leads to miscalculation of the gradient, $\mathbf{g}^n$, and conjugate gradient vectors, $\mathbf{d}^n$, and results in the loss of the **A**-orthogonal properties. Consequently, the conjugate gradient vectors do not expand the N-dimensional solution space so the reconstruction will not converge after N-iterations. In some cases, the reconstruction may even diverge. In general, the mean squared error, $\{E^n\}$, will converge to a minimum, *i.e.*, $E^0 \geq E^1 \geq \ldots \geq E^n \geq \ldots$ From our experience, the divergence happens when the mean squared error, E^n, exceeds the previous one, E^{n-1}. Once this occurs, the mean squared error will continue to increase and exceed the range of the variables after only a few iterations (less than 10). This is not the case when using the POCS algorithm, whose mean squared errors can increase and decrease during the reconstruction and ultimately converge to a minimum. To overcome this difficulty, the subsequent scheme was used to detect this divergence.

Let the ratio of two consecutive mean squared errors be:

$$r^n = \frac{E^n}{E^{n-1}} \tag{8}$$

The reconstruction result becomes divergent when $r^n > 1$. Once this happens, the conjugate gradient vector is reset, *i.e.*, $\mathbf{d}^n = \mathbf{0}$, and the CGD reconstruction is restarted using the previous reconstruction result, $\Delta\boldsymbol{\mu}^{n-1}$, as the initial value.

1.4 Weight Matrix Rescaling

A rescaling technique, which rescales the weight matrix **W** to make it more uniform and less ill-conditioned, was also used to suppress numerical errors and accelerate convergence. Two criteria, rescaling the maximum value of each column to 1, *i.e.*, $w_{ij}' = w_{ij} / \max_{i=1}^{I}\{w_{ij}\}$, and rescaling the average value of each column to 1, *i.e.*, $w_{ij}' = w_{ij} / \sum_{i=1}^{I} w_{ij}$, were applied and compared to the results without rescaling.

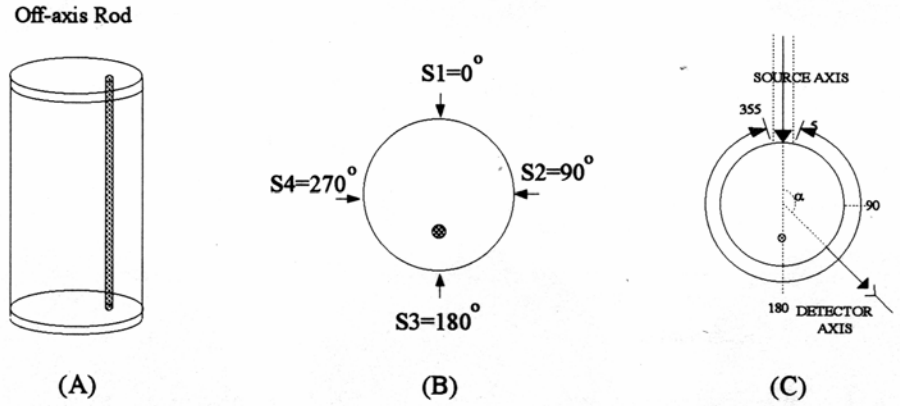


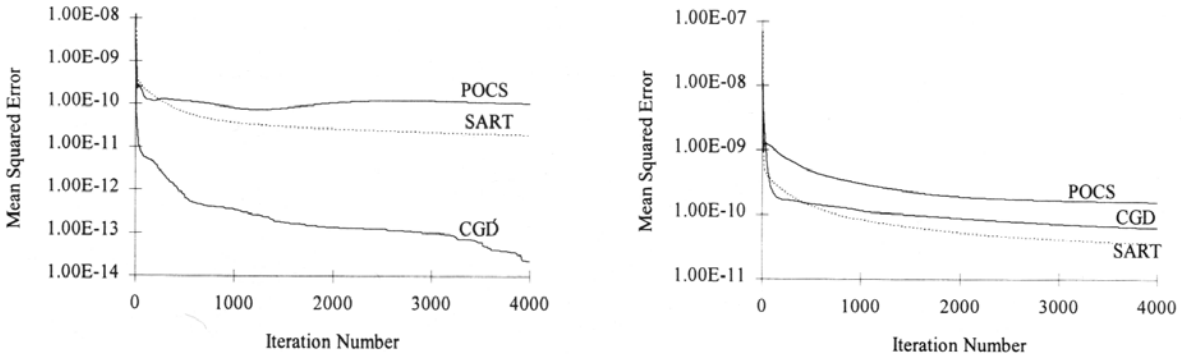
Figure 1. (A) Tissue phantom used for Monte Carlo simulations. (B) Source configuration. (C) Detector configuration for each source. The detectors were oriented normal to the phantom in the plane of the source and located every 10° about the boundary.

3. METHODS

The numerical studies employed Monte Carlo methods to compute the internal light distributions and the flux of light re–emitted across the surface of a 3–D cylindrical, homogeneous, isotropically scattering media. The cylinder axis was infinitely long, and its diameter was 20 mean free pathlengths (mfp; 1 mfp = $1/\mu_T$, where μ_T is the total cross section). The heterogeneity was a single 2–mfp–diameter black rod whose axis was halfway between the cylinder axis and the boundary (Figure 1A). The light source, a pencil beam directed normally to the surface, was sequentially located at 0° , 90°, and 180° about the cylinder (Figure 1B). Detector readings were calculated for each location of the source. From geometrical symmetry, it was assumed that the detector readings corresponding to the source located at 270° are equivalent to those obtained for the 90° source. The detectors were spaced at 10° intervals about the cylinder (Figure 1C). Only the readings of detectors located in the plane containing the source and normal to the cylinder were used by the reconstruction algorithms. 5×10^7 photons were launched into the medium in each of these simulations. Internal light intensities (forward and adjoint) were calculated in separate simulations, in which the cylinder was homogeneous and the average collision density per incident photon was calculated in each of 16,400 voxels. These were arranged in 41, 1–mfp–thick layers perpendicular to the cylinder axis, with 400 voxels in each layer; each voxel's volume was $\pi/4$ mfp^3. The output was reported in units of collisions/incident photon/unit volume. 2×10^8 photons were launched into the medium in each of these simulations. The resultant 3-D weight functions were transformed to 2-D by summing the values of the weight along the direction of the long axis of the cylinder. 2-D image reconstructions were accomplished using all three algorithms and applying the appropriate positivity constraints. For each, a total of 144 source-detector pairs were considered (4 source locations × 36 detectors per source). Thus, in all cases, the reconstruction problems were underdetermined (400 unknowns). Image quality was evaluated for each combination of algorithm and rescaling technique. Reconstructions using constrained CGD were also compared to the results using unconstrained CGD.

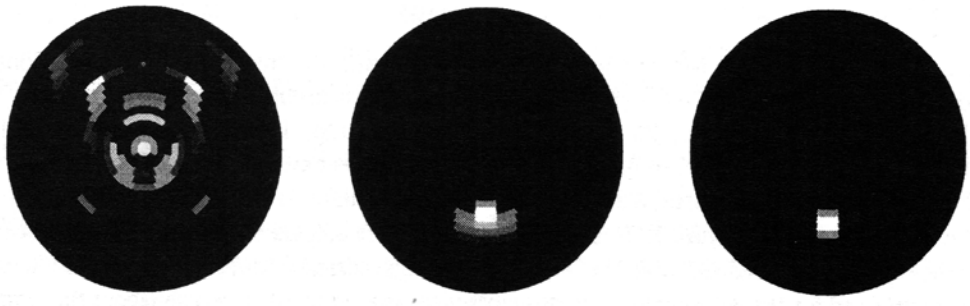
4. RESULTS

Figure 2 shows the mean squared error vs. iteration number for different algorithms. Figure 2A is the results without constraint and Figure 2B with constraint. Figure 3 shows the reconstructed images using the unconstrained and constrained CGD algorithms. The unconstrained and constrained result, (positivity constraints on both detector readings and reconstruction results), is shown in Panels A and B, respectively. The target is shown in Panel C. Results in Figure 4 demonstrate the reconstructed images obtained using the CGD algorithm after 100, 1000, and 10,000 iterations, with different rescaling techniques - no rescaling, rescaling the maximum value of each column to one, and rescaling the average value of each column to one. Figures 5-6 show the corresponding results obtained using the SART and POCS algorithms, respectively.



(A) Unconstrained reconstruction (B) Constrained reconstruction

Figure 2. Mean squared error vs. iteration number for different algorithms: (A) unconstrained reconstruction, (B) constrained reconstruction.



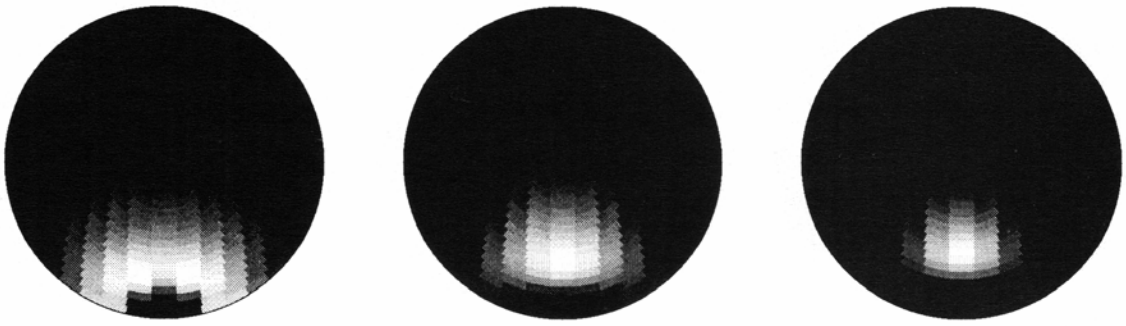
(A) Unconstrained (B) Constrained (C) Target

Figure 3. Reconstruction results obtained using the CGD algorithm with and without range constraints for 10,000 iterations.

5. DISCUSSION

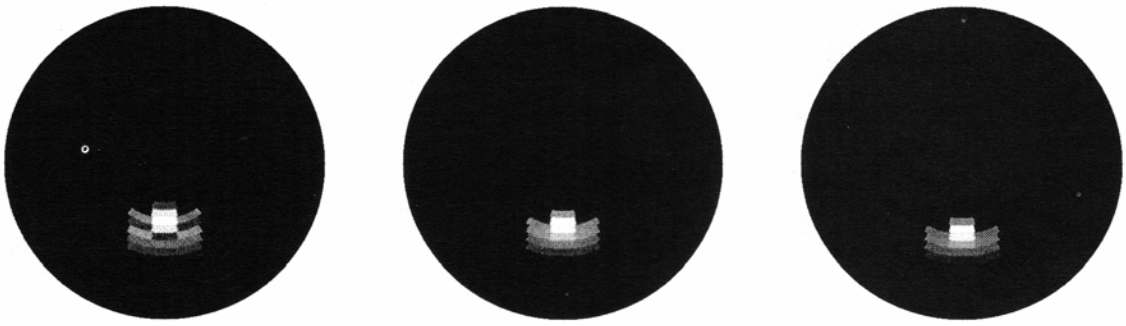
In a previous study, we compared the quality of reconstruction results of simply structured phantoms using various iterative algorithms [2]. With exception of CGD, in nearly all cases good quality reconstructions were obtained. In these studies it was observed that, unlike the other algorithms, efforts to apply a range constraint to the CGD algorithm frequently caused the solution to diverge. This indicates that applying positive range constraints to the previous reconstruction leads to miscalculation of the gradient and conjugate gradient vectors and results in the loss of **A**-orthogonal properties. In this study we have overcome this difficulty by detecting divergence and restarting the

(A) No rescaling



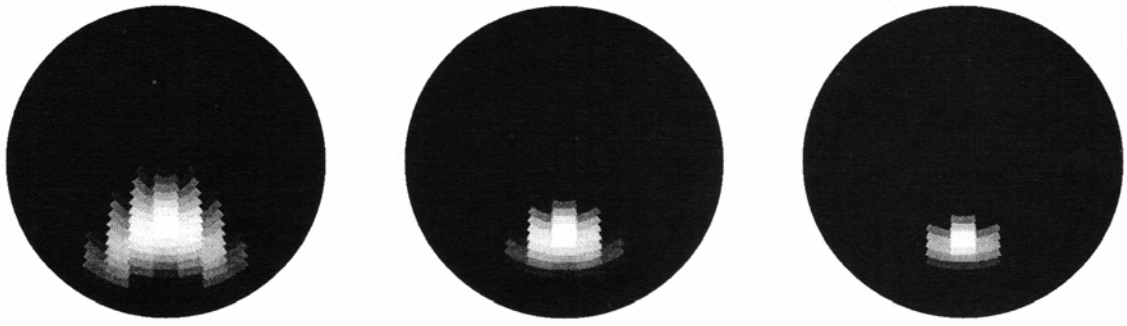
100 iterations 1,000 iterations 10,000 iterations

(B) Rescaling the maximum of each column to 1



100 iterations 1,000 iterations 10,000 iterations

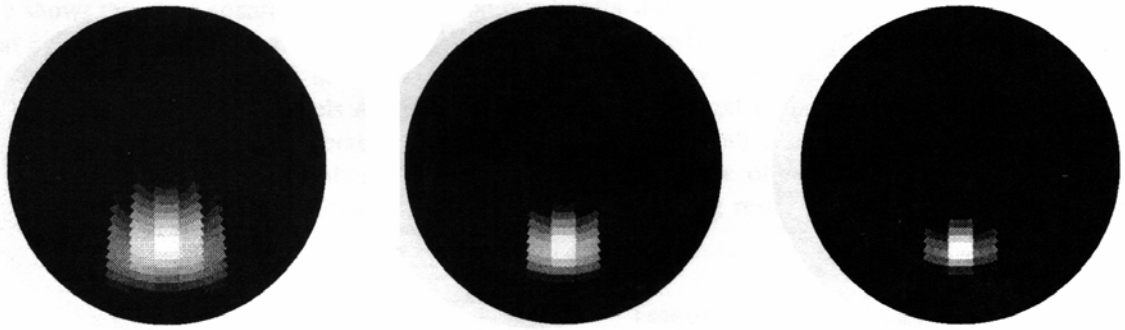
(C) Rescaling the average of each column to 1



100 iterations 1,000 iterations 10,000 iterations

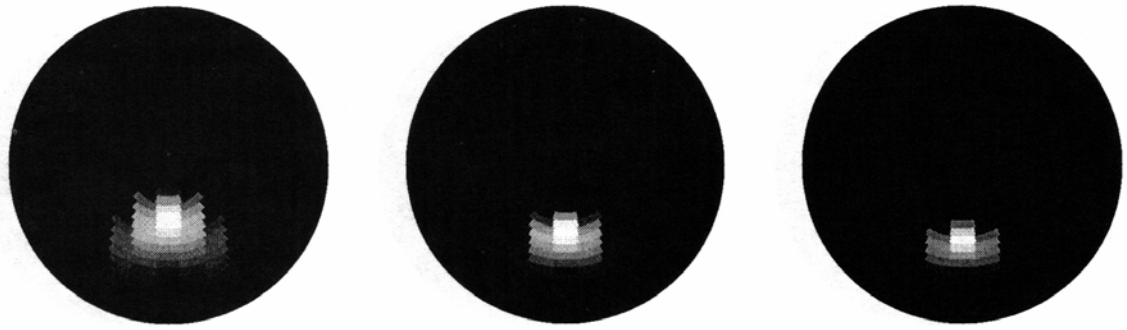
Figure 4. Reconstruction results using the CGD algorithm and (A) no rescaling, (B) rescaling the maximum of each column to 1, or (C) rescaling the average of each column to 1, after 100, 1,000, and 10,000 iterations.

(A) No rescaling



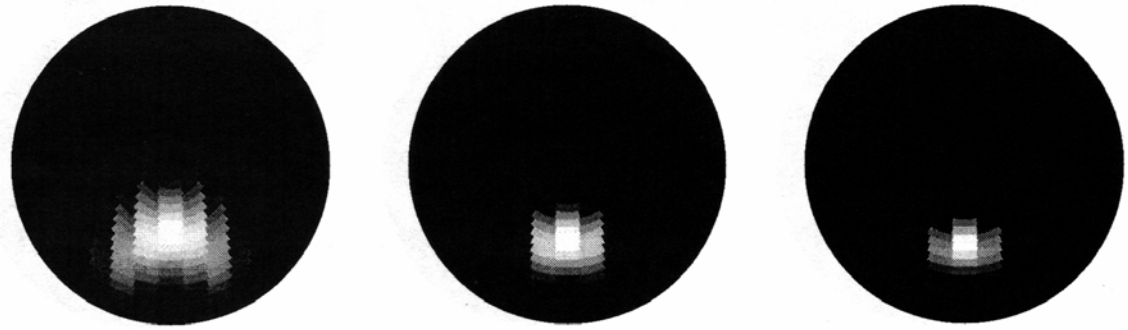
100 iterations 1,000 iterations 10,000 iterations

(B) Rescaling the maximum of each column to 1



100 iterations 1,000 iterations 10,000 iterations

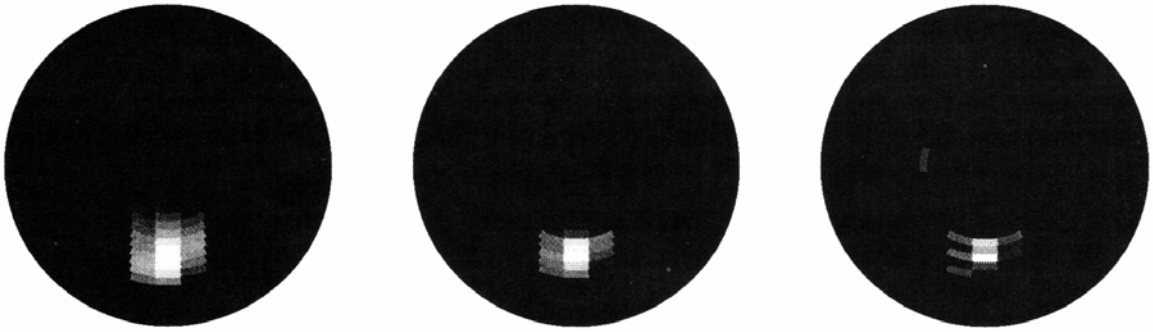
(C) Rescaling the average of each column to 1



100 iterations 1,000 iterations 10,000 iterations

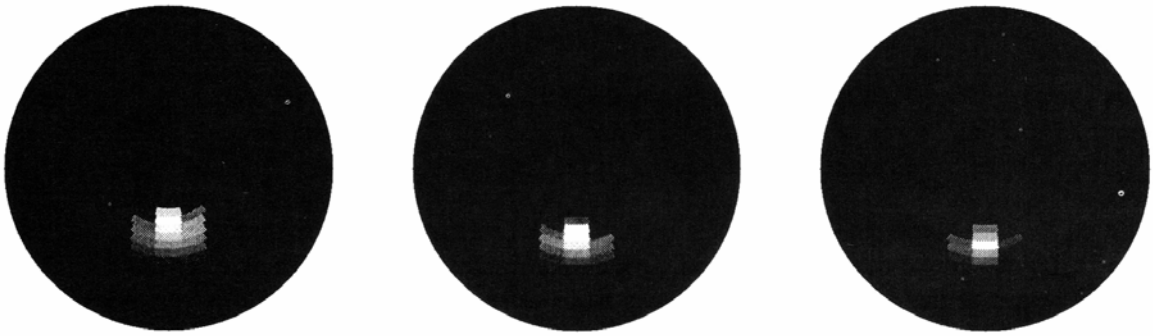
Figure 5. Reconstruction results using the SART algorithm and (A) no rescaling, (B) rescaling the maximum of each column to 1, or (C) rescaling the average of each column to 1, after 100, 1,000, and 10,000 iterations.

(A) No rescaling



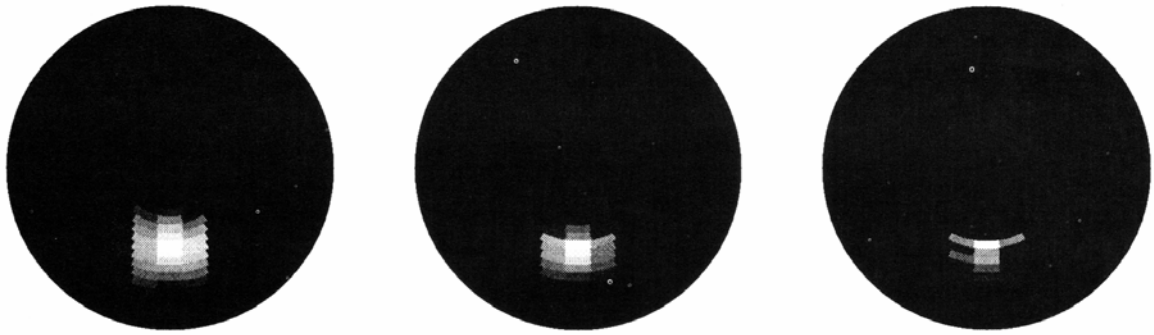
100 iterations 1,000 iterations 10,000 iterations

(B) Rescaling the maximum of each column to 1



100 iterations 1,000 iterations 10,000 iterations

(C) Rescaling the average of each column to 1



100 iterations 1,000 iterations 10,000 iterations

Figure 6. Reconstruction results using the POCS algorithm and (A) no rescaling, (B) rescaling the maximum of each column to 1, or (C) rescaling the average of each column to 1, after 100, 1,000, and 10,000 iterations.

reconstruction using the previous result, $\Delta\mu^{n-1}$, as the initial value. The use of positivity constraints slowed down the convergence of mean squared error significantly, as shown in Figure 2. With constraints, all three algorithms converge at about the same rate. Without constraints, we observe that after 4,000 iterations, the mean squared error for the CGD method is three orders of magnitude less than for SART and POCS. A reduced mean squared error, however, does not ensure a better reconstruction result. This is demonstrated in Figure 3 where the unconstrained result, which has a much smaller mean squared error, produces a significantly poorer quality image than does the constrained result.

A second method we have tested to improve the quality of the reconstructions is the use of a matrix rescaling method. This idea was derived from the recognition that, in dense scattering media, the range of values of weight for a particular voxel is strongly dependent on the source-detector configuration. This results in large variations in the maximum value of the column vectors, $\mathbf{w}_j$, of Eq. 2. Physically, $\mathbf{w}_j$ is an expression of the importance that a particular voxel has on the detector response. Voxels having larger values of $\mathbf{w}_j$ will have a greater impact on a detector than voxels with smaller values. Because of this, it is expected that the CGD method will preferentially update those voxels having the greatest weight. This follows because column vectors that have the largest absolute weight will most strongly influence the computed conjugate gradient. From previous studies we have determined that the largest values of weight typically occur in the vicinity of the source and receiver. As these are located near the surface, at early iterations the resultant image will be predominately localized to these voxels, as shown in Figure 4A. The effect of rescaling is to mathematically minimize differences in importance between the weight vectors thus causing a more uniform update of the reconstructed image.

Results shown in Figure 4B demonstrate that, with rescaling, a more accurate reconstruction is achieved after only 100 iterations. Figure 4C shows that an improved result is also obtained when the average value of the column vector is set equal to one, but this result would appear inferior to the case when the maximum value is scaled to one (Figure 4B). The corresponding results for the SART algorithm, shown in Figure 5, which is also a simultaneous method, reveals that rescaling produces some improvement in image quality at early iterations, but the magnitude of this effect is less than observed for CGD. We note that without rescaling, the SART result is more accurate at early iterations than that obtained by CGD. Inspection of the SART algorithm shows that during the backprojection step, projections to each voxel are divided by the sum of the elements of the column vector for that voxel. This is equivalent to rescaling the average value of each column vector to one. The effect of rescaling column vector on reconstructions obtained using POCS is similar to the results obtained from SART. In POCS, the angles between the constraint sets is a critical determinant of the convergence rate. We believe that the effect of rescaling is to increase these angles, thereby increase the rate of convergence.

In summary, we have demonstrated that improved reconstructions using the CGD method can be obtained by implementing the described divergence detection method. We also show that additional improvements are achievable for all algorithms tested by using matrix rescaling.

6. ACKNOWLEDGMENT

This work was supported in part by NIH grant R01 CA59955, by the New York State Science and Technology Foundation, and by ONR grant N000149510063.

7. REFERENCES

1. J. Chang, H. L. Graber, R. L. Barbour, "Progress toward optical mammography: imaging in dense scattering media using time-independent optical sources," *Proceedings of 1994 IEEE Medical Imaging Conference*, Norfolk, Nov. 1994.

2. J. Chang, H. L. Graber, and R. L. Barbour, "Image reconstruction of targets in random media from continuous wave laser measurements and simulated data," *OSA Proceedings on Advances in Optical Imaging and Photon Migration*, vol. 21, pp. 193-201, Orlando, Mar. 1994.

3. H. L. Graber, J. Chang, R. Aronson, R. L. Barbour, "A perturbation model for imaging in dense scattering media: derivation and evaluation of imaging operators," in *Medical Optical Tomography: Functional Imaging and Monitoring*, SPIE Institutes vol. IS11, pp. 121-143, SPIE Press, 1993.

4. G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, 1980.

5. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging,* IEEE Press, New York, 1988.

6. *Image Recovery: Theory and Application*, H. Stark, ed., Academic Press, New York, 1987

7. D. C. Youla, "Mathematical theory of image reconstruction by the method of convex projections," *Image Recovery : Theory and Application,* Henry Stark, ed., Academic Press, New York, 1987.

8. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization,* Academic Press, New York, 1981.

9. G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, MA, 1986.

10. A. H. Anderson and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm." *Ultrasonic Imaging*, vol. 6, pp. 81-94, 1984.

11. J. Chang, R. Aronson, H. L. Graber, R. L. Barbour "Imaging diffusive media using time-independent and time-harmonic sources: dependence of image quality on imaging algorithms, target volume, weight matrix, and view angles," accompanying paper in this proceedings.