Preconditioning Methods for Improved Convergence Rates in Iterative Reconstructions

Neal H. Clinthorne, *Member, IEEE,* Tin-Su Pan, *Member, IEEE,* Ping-Chun Chiao, *Member, IEEE,* W. Leslie Rogers, *Member, IEEE,* and John A. Stamos

Abstract— **Because of the characteristics of the tomographic inversion problem, iterative reconstruction techniques often suffer from poor convergence rates—especially at high spatial frequencies. By using preconditioning methods, the convergence properties of most iterative methods can be greatly enhanced without changing their ultimate solution. To increase reconstruction speed, we have applied spatially-invariant preconditioning filters that can be designed using the tomographic system response and implemented using 2-D frequency-domain filtering techniques.**

In a sample application, we performed reconstructions from noiseless, simulated projection data, using preconditioned and conventional steepest-descent algorithms. The preconditioned methods demonstrated residuals that were up to a factor of 30 lower than the unassisted algorithms at the same iteration. Applications of these methods to regularized reconstructions from projection data containing Poisson noise showed similar, although not as dramatic, behavior.

I. Introduction

In SPECT and PET, accurately modeling the effects of imaging system resolution, attenuation, and scatter, in addition to employing a regularized solution criterion for noise suppression, will improve quantitative accuracy. Because the set of equations resulting from this modeling process can be difficult to solve directly, however, iterative methods are usually employed. While iterative reconstruction techniques allow a straightforward implementation, they have a major disadvantage in that their progress toward the desired solution can often be painfully slow. Since this behavior is primarily a consequence of the ill-posed nature of the tomographic inversion problem, regularized solution objectives—necessary to reduce the noise-sensitivity of the reconstruction—have the side-benefit of improving converging rates. Once the desired, regularized solution objective has been specified, however, the convergence rate of most iterative methods can be further enhanced—without altering the chosen solution criterion—by preconditioning [1].

Preconditioning techniques have been applied in many areas of numerical analysis including tomographic reconstruction [2]–[4]. These methods hinge on being able to find a system of equations "close" in some sense to the true system, but more easily invertible. In this work, we have exploited the similarity of the tomographic reconstruction problem to that of deconvolving a 2-D spatially invariant blur and have developed preconditioners that can be represented as frequency-domain filtering operations. These preconditioning filters typically have a high-pass characteristic and can be tailored to individual tomographs by employing the system response in their design.

The next section presents the relevant background to least-squares reconstruction methods. A general framework for iterative solution techniques and convergence rate expressions are given in Section III. Section IV discusses necessary qualities for preconditioners and Section V presents reconstructions of simulated SPECT data using a preconditioned steepest-descent iteration with both least-squares and regularized least-squares objective functionals.

Manuscript received November 12, 1991; revised May 27, 1992. This work was supported by the National Cancer Institute, DHHS, under PHS Grant RO1-CA54362.

N. H. Clinthorne, P.-C. Chiao, and W. L. Rogers are with the Division of Nuclear Medicine. The University of Michigan, Ann Arbor, MI 48109.

T.-S. Pan is with the Department of Nuclear Medicine, University of Massachusetts, Worcester, MA 01655.

J. A. Stamos is with the Department of Nuclear Engineering, The University of Michigan, Ann Arbor, MI 48109.

IEEE Log Number 9206091.

II. Background

We assume that the reconstruction problem can be formulated as the solution to an over-determined set of imaging equations

$$\boldsymbol{y} = A\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{x}$ is an m-vector representing the emission density of a "pixelated" object, $\boldsymbol{y}$ is an n-vector of projection data and A is the $n \times m$ system transfer matrix which can include, in addition to an accurate model of the tomograph resolution, the attenuation map of the object and a model for γ-ray scattering.

The projection data are usually corrupted by measurement noise and since the imaging system is assumed to be overdetermined ($n > m$ and A^T is of full column-rank), the above equation may have no formal solution. It is common practice in this case to solve (1) in a least-squares sense, that is, we choose $\boldsymbol{x}^*$ to minimize $\phi(\hat{\boldsymbol{x}})$ where

$$\phi^2(\hat{\boldsymbol{x}}) = \|\boldsymbol{y} - A\hat{\boldsymbol{x}}\|_2^2 \equiv (\boldsymbol{y} - A\hat{\boldsymbol{x}})^T(\boldsymbol{y} - A\hat{\boldsymbol{x}}). \tag{2}$$

If there is no measurement error in the projection vector, choosing $\boldsymbol{x}^*$ to minimize (2) recovers the original emission density $\boldsymbol{x}$ otherwise $\boldsymbol{x}^*$ is the unique solution such that the Euclidian distance of the projection data to $A\boldsymbol{x}^*$ is minimized.

To account for the effects of nonstationary or covarying noise in the projection data, "weighted" least-squares solution objectives are often employed. The weighting is typically

accomplished by choosing the solution $\boldsymbol{x}^*$ that minimizes the squared "K-norm," defined as

$$\phi^2(\hat{\boldsymbol{x}}) = \|\boldsymbol{y} - A\hat{\boldsymbol{x}}\|^2_{K^{-1}} \equiv (\boldsymbol{y} - A\hat{\boldsymbol{x}})^T K^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{x}}) \quad (3)$$

where K is usually taken to be the noise covariance matrix if it is known or an estimate thereof.

The solution to the above minimization problem is given by the well-known least-squares normal equations [1].

$$\boldsymbol{x}^* = (A^T K^{-1} A)^{-1} A^T K^{-1} \boldsymbol{y}. \quad (4)$$

It is easy to show (Appendix I) that if the projection data is corrupted by Poisson counting noise, the unconstrained maximum-likelihood estimate of the object $\boldsymbol{x}_{ML}$ also satisfies (4) providing that the noise covariance matrix $K(\boldsymbol{x}) = \text{diag}(A\boldsymbol{x})$ is simultaneously estimated as a function of the object estimates,

$$\hat{\boldsymbol{x}}_{ML} = (A^T K^{-1}(\hat{\boldsymbol{x}}_{ML}) A)^{-1} A^T K^{-1}(\hat{\boldsymbol{x}}_{ML}) \boldsymbol{y}. \quad (5)$$

The normal equations are therefore useful in describing solutions with respect to a variety of statistical criteria.

III. Solution Methods

A. Generalized Landweber Iteration

For a quadratic objective function, inspection of (3) reveals that it can solved directly for the desired solution; however, even a 64×64 pixel estimate of the object requires inversion of the $4096 \times 4096 A^T K^{-1} A$ matrix—a computationally intensive procedure. Although the required matrix inversion is feasible on many computer systems, it is nevertheless impractical when the matrix A changes from slice-to-slice or patient-to-patient, which occurs for example, as the attenuation distribution changes in SPECT.

Just as the normal equations are useful for describing the solution with respect to a number of statistical criteria, the generalized Landweber iteration [5] can be used to describe a wide class of simultaneous-update iterative methods. The iteration, a variant of the successive approximation method, can be written as

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_k + M_k^{-1} A^T K_k^{-1} (\boldsymbol{y} - A\hat{\boldsymbol{x}}_k) \quad (6)$$

where $\hat{\boldsymbol{x}}_k$ is the estimate of the object distribution at iteration k, A, and K are defined above, and M_k is a positive-definite, symmetric *preconditioning matrix* which can dramatically alter the convergence properties of (6) without changing the fixed-point of the iteration.

The above iteration can describe a wide variety of iterative methods. As examples: the EM algorithm is obtained by setting $K_k = \text{diag}(A\hat{\boldsymbol{x}}_k)$ and $M_k = \text{diag}(\hat{\boldsymbol{x}}_k)$; the SIRT method results if $M_k = I$ and $K_k = I$; and with proper choice of M_k, the conjugate-gradient and steepest-descent methods can similarly be obtained.

B. Convergence Rates

Since the preconditioner M_k does not alter the fixed-point of (6), by properly choosing M_k the convergence rate of most currently used iterative methods can be enhanced without altering their ultimate solution. In order to more fully characterize the necessary properties for good preconditioners, we first present conditions for convergence and an expression for the convergence rate.

Define the residual vector at each iteration by

$$\varepsilon_k = A^T K_k^{-1} (\boldsymbol{y} - A\hat{\boldsymbol{x}}_k), \quad (7)$$

or the backprojection of the weighted projection-space error vector. Using this formulation, a recurrence relation for the residual vectors can be derived for the case in which K is constant with iteration (Appendix II) and can be used to obtain the following inequality:

$$\|\varepsilon_{k+1}\| \leq \|I - A^T K^{-1} A M_k^{-1}\| \, \|\varepsilon_k\|. \quad (8)$$

A sufficient condition for the convergence of (6) to a unique solution is provided by an appeal to the contraction mapping principle which states that if the norm of the operator $(I - A^T K^{-1} A M_k^{-1})$ is less than unity (i.e., a contraction), convergence is assured [6], [7]. This condition is satisfied if the eigenvalues of $A^T K^{-1} A M_k^{-1}$ lie in the open interval (0, 2).

Inspection of (8) reveals that the smaller $\|I - A^T K^{-1} A M_k^{-1}\|$, the faster convergence will be. If the eigenvalues of $A^T K^{-1} A M_k^{-1}$ lie in the interval (0, 1), the matrix norm in (8) reduces to

$$\|I - A^T K^{-1} A M_k^{-1}\| = 1 - \lambda_{\min}(A^T K^{-1} A M_k^{-1}) \quad (9)$$

where the notation $\lambda_{\min}(B)$ indicates the minimum eigenvalue of B. Convergence rates in unpreconditioned iterative methods suffer because while the maximum eigenvalue of the composite matrix $A^T K^{-1} A$ may be close to one, the minimum eigenvalue is usually very close to zero (i.e., the condition number of the system matrix is large). To be useful in improving the convergence rate, the preconditioner, M_k must therefore improve the condition of the system thereby decreasing the norm (9).

C. Regularized Solutions

In practice the poorly conditioned nature of the system response matrix causes other difficulties—most notably, amplification of the errors in the projection data resulting from Poisson counting noise. In order to reduce this sensitivity to noise, a stabilizing term is often added to the objective functional to penalize solutions inconsistent with our prior notions of how they should behave:

$$\phi^2_\alpha(\hat{\boldsymbol{x}}) = \|\boldsymbol{y} - A\hat{\boldsymbol{x}}\|^2_{K^{-1}} + \alpha \|Q\hat{\boldsymbol{x}}\|^2_W. \quad (10)$$

The first term, as before, is a measure of the distance of the solution from the data, but now the additional term introduces a penalty for the deviation of certain features of the object (e.g., smoothness) from those deemed acceptable *a priori*. The regularization parameter, α, controls the balance between

solutions agreeing entirely with the data and those agreeing with our prior knowledge. The matrix Q in (10) is often chosen to approximate a two-dimensional, second derivative operator so that the "roughness" of the reconstructed objects is penalized [8], [9]. The weighting matrix W can be employed to apply the smoothing in a spatially variant manner; for example, it can be chosen to prevent smoothing across known boundaries in the reconstructed image [10].

The direct solution for the $\hat{\boldsymbol{x}}$ that minimizes (10) can be written immediately as

$$\hat{\boldsymbol{x}} = (A^T K^{-1} A + \alpha Q^T W Q)^{-1} A^T K^{-1} \boldsymbol{y}. \tag{11}$$

But, as before, the size of the problem often precludes the use of the direct approach. An iterative method, converging to the regularized solution, is given by

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_k + M_k^{-1}(A^T K_k^{-1} \boldsymbol{y} - (A^T K_k^{-1} A + \alpha Q^T W Q)\hat{\boldsymbol{x}}_k). \tag{12}$$

Paralleling the development of (9) for the unregularized iteration, an analogous expression for the preconditioned, regularized iteration can be derived as

$$\|\varepsilon_{k+1}\| \leq \|I - (A^T K^{-1} A + \alpha Q^T W Q) M_k^{-1}\| \, \|\varepsilon_k\| \tag{13}$$

where now $\varepsilon_k = A^T K^{-1} \boldsymbol{y} - (A^T K^{-1} A + \alpha Q^T W Q)\hat{\boldsymbol{x}}_k$.

IV. Preconditioner Selection

One immediately sees from (8) that from the standpoint of convergence rate, the optimum choice of preconditioner would be $M_k = A^T K^{-1} A$ (or $A^T K^{-1} A + \alpha Q^T W Q$ for the regularized case). With this selection, the iteration (6) becomes Newton's method and, for a quadratic objective, converges in one iteration. Nevertheless, this choice is impractical since the preconditioner is just as difficult to invert as the matrix $A^T K^{-1} A$. We seek preconditioners that both are close to the true system *and* are efficiently invertible. If the condition number of $A^T K^{-1} A M_k^{-1}$ is less than that of $A^T K^{-1} A$ alone, the convergence rate of the iteration (6) will improve.

Numerous types of preconditioners have been applied to iterative problems ranging from the diagonal preconditioner arising naturally in the EM algorithm [11] to incomplete Cholesky factorizations of the system matrix, polynomials in the system matrix, and Fourier methods [1]–[4]. Since tomographic inversion is similar to deconvolving a two-dimensional spatially invariant blur (after an unfiltered backprojection of the projection data), we have chosen the Fourier method. For its implementation, we exploit the well-known relationship between spatially invariant point spread functions (psf) and block-circulant matrices, which can be efficiently inverted using the 2-D FFT [8].

V. Results

To illustrate the advantage of preconditioning, we simulated a parallel-collimation rotating gamma-camera system having 128 equally spaced angular views of a 64×64 pixelated object having no attenuation. The number of samples per projection was 64, the length of each projection element was the same as that of one side of a square object pixel, and the simulated resolution was two samples fwhm, which was accounted for in the system response matrix, A. The test object used in this study was the "complex phantom" we have previously employed for analyzing algorithm performance [12].

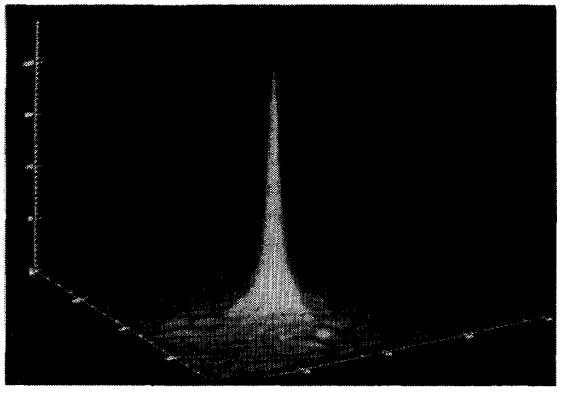

Fig. 1. Point response function of the $A^T A$ operation for a point at the center of the simulated system.

A. Preconditioning Filters

We chose, as preconditioning operators, high-pass filters that attempt to undo the blurring operations inherent in the forward projection and backprojection operations. To derive the filters, we generated a point spread function of the $A^T A$ operation by forward projecting and backprojecting a one-pixel point source located at the center of the field-of-view. The resulting psf is shown in Fig. 1. This response was transformed to the frequency domain and 2-D "inverse-filters" were generated using the formula

$$H(u,v) = \frac{1}{F(u,v) + \Gamma} \tag{14}$$

where $H(u,v)$ is the preconditioning filter and $F(u,v)$ is the Fourier transform of the response in Fig. 1. Since the prf of this system is spatially varying, it cannot expected that a filter derived from the response at a single image location will accurately deconvolve blurring at all locations. Because of this, we added the gain-limiting term Γ in (14) to limit the effect of the filter where $F(u,v)$ is small. Profiles through the center of the two-dimensional filters for four values of Γ chosen as 0.5%, 1.0%, 5.0%, and 20.0% of the maximum of $F(u,v)$(Γ_1–Γ_4, respectively) are shown in Fig. 2.

B. Noiseless Reconstructions

We used a steepest-descent variant of (6) to perform the reconstruction because it adaptively adjusts the magnitude of the correction vector such that convergence is assured for quadratic $\phi(x)$. The following iteration was used:

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_k + r_k M^{-1} A^T (\boldsymbol{y} - A\hat{\boldsymbol{x}}_k), \tag{15}$$

$$r_k = \frac{\varepsilon_k^T M^{-1} \varepsilon_k}{\varepsilon_k^T M^{-1} (A^T A) M^{-1} \varepsilon_k}, \tag{16}$$

$$\varepsilon_k = A^T (\boldsymbol{y} - A\hat{\boldsymbol{x}}_k). \tag{17}$$

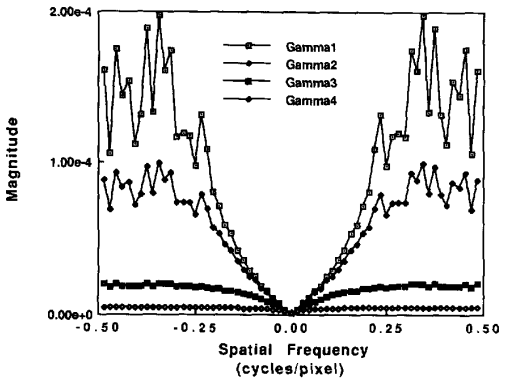

Fig. 2. Preconditioning filters generated from psf in Fig. 1 for four values of Γ. $\Gamma_1 = 0.005F_{\max}$, $\Gamma_2 = 0.01F_{\max}$, $\Gamma_3 = 0.05F_{\max}$, $\Gamma_4 = 0.20F_{\max}$. $F_{\max} = \max\{F(u, v)\}$.

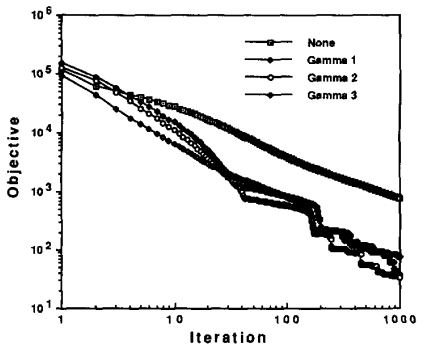

Fig. 3. Magnitude of objective function plotted against iteration for no preconditioning and preconditioning filters using Γ_1, Γ_2, and Γ_3.

At each iteration, preconditioning was accomplished by: (1) transforming the correction vector $A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}}_k)$ to the frequency domain, (2) multiplying this by the preconditioning filter, and (3) transforming the preconditioned update vector back to the spatial domain. The iteration was implemented in C under VMS on a DEC MicroVax III computer system. Reconstruction times for the 64×64 object were approximately 40 s/iteration with no preconditioning and 55 s/iteration with preconditioning.

Reconstructions of the simulated data were performed using no preconditioning and using preconditioning filters Γ_1, Γ_2, and Γ_3 shown in Fig. 2. The norm $\|\boldsymbol{y} - A\hat{\boldsymbol{x}}_k\|$ is plotted in Fig. 3 for reconstructions using each of the filters to 1000 iterations. Since there is no noise in the projection data, this norm should reach zero when convergence is attained. It is evident that the preconditioned methods converge more quickly than the unaided steepest-descent iteration.

The reconstructed images at 1, 10, 100, and 1000 iterations are shown in Fig. 4 with the slowest converging algorithm in the top row (no preconditioning), to the fastest converging at 1000 iterations in the bottom row (preconditioned with Γ_2). Note that, as expected, the preconditioned reconstructions recover the high object frequencies more quickly.

C. Regularized Reconstructions with Noisy Data

As mentioned above, to reduce the variance in the reconstruction due to projection counting noise, regularized

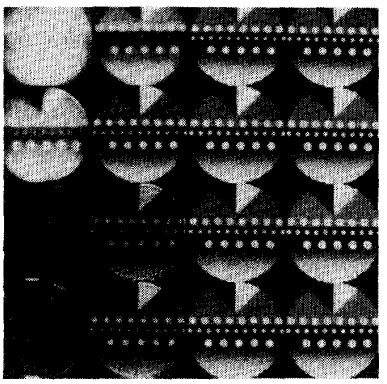

Fig. 4. Reconstructed images for each filter at 1, 10, 100, and 1000 iteration. Row-order from top: no preconditioning, Γ_3, Γ_1, Γ_2.

solutions are usually employed. To evaluate the effects of preconditioning on regularized iterations, we used the same phantom, this time with projection data containing Poisson counting noise equivalent to one million events, along with a zeroth-order Tikhonov stabilizer [10]. The resulting objective requiring minimization takes the form

$$\phi_\alpha^2(\boldsymbol{x}) = (\boldsymbol{y} - A\boldsymbol{x})^T(\boldsymbol{y} - A\boldsymbol{x}) + \alpha \boldsymbol{x}^T\boldsymbol{x} \tag{18}$$

where α is the regularization parameter controlling the degree of noise smoothing. Note that this stabilizer has the action of adding the small constant α to each of the singular values of A^TA before inversion, thereby significantly limiting the maximum singular value of the inverse.

A steepest-descent iterative method to minimize (18) can be written as

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_k + r_k M^{-1}(A^T\boldsymbol{y} - (A^TA + \alpha I)\hat{\boldsymbol{x}}_k), \tag{19}$$

$$r_k = \frac{\varepsilon_k^T M^{-1}\varepsilon_k}{\varepsilon_k^T M^{-1}(A^TA + \alpha I)M^{-1}\varepsilon_k}, \tag{20}$$

$$\varepsilon_k = A^T\boldsymbol{y} - (A^TA + \alpha I)\hat{\boldsymbol{x}}_k. \tag{21}$$

This algorithm was again implemented on the MicroVax III and reconstruction times showed little difference from those for the unregularized case above.

Using a value for the regularization parameter of 0.004 $\|A^TA\|_1$, we evaluated the convergence of the unpreconditioned reconstruction and the preconditioned reconstructions using the filters Γ_2, Γ_3, and Γ_4. The value of the objective function is plotted in Fig. 5 for each reconstruction to 50 iterations. The behavior is similar to that of the unregularized case with a notable exception: the convergence in all cases is much faster due to the regularized solution objective.

Reconstructed images are shown in Fig. 6 at 5, 10, 20, and 50 iterations for the unaided steepest-descent method in the top row, filter Γ_4 in the second row, filter Γ_3 in third row, and finally, Γ_2 at the bottom. Notice that as far as visual quality, there is little change in the preconditioned reconstructions after 20 iterations. All preconditioned iterations have essentially reached the same numerical solution at 50 iterations. In contrast, the unassisted steepest-descent method is still converging at 50 iterations.

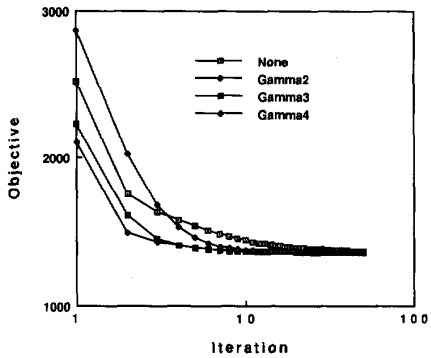

Fig. 5. Magnitude of objective function for the regularized iteration (19) to 50 iterations.

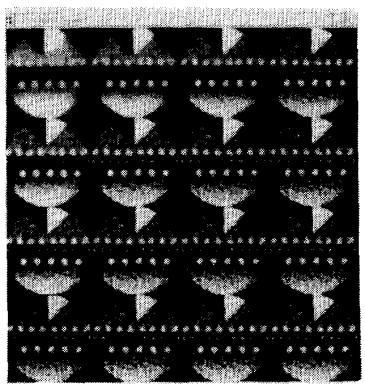

Fig. 6. Images reconstructed using regularized objective functional and preconditioned steepest-descent iteration at 5, 10, 20, and 50 iterations. Top row: no preconditioning. Second through fourth rows: preconditioned with filters Γ_4, Γ_3, and Γ_2, respectively.

VI. Discussion

Close inspection of the images in Fig. 4 reveals that since preconditioning improves the high-frequency convergence rate, the iterates for some filters tend to appear "noisier" than if no preconditioning is used. As the algorithm nears convergence, however, this behavior tends to subside. Additionally, the reconstructed images in the top row of Fig. 4 demonstrate that the appearance of these high frequencies is not peculiar to the preconditioned methods and may be a natural consequence of the convergence properties of this simulated system. Another result of faster high frequency convergence rates is that preconditioning will have a tendency to amplify the Poisson counting noise faster than the unaided algorithm. This is due strictly to the fact that preconditioning causes the iteration to take a different—and faster—path to convergence.

Even though the residuals in Fig. 3 are quite small for the preconditioned methods, there are still significant discrepancies between all reconstructions at 1000 iterations and the true object—especially at intensity discontinuities. This is due to the very small singular values of the matrix A^TAM^{-1} that are difficult to recover even with preconditioning. In practice, a regularized solution criterion would nearly always be employed and would not only reduce the effects of noise amplification but also improve convergence properties by increasing the magnitude very small singular values of A^TA.

In this paper, we applied preconditioning to the steepest-descent algorithm and solved for an unweighted least-squares estimate. But, since (6) can describe a variety of currently employed reconstruction methods, preconditioning should be useful in enhancing the convergence rates of all these methods—including the maximum likelihood estimate. The steepest-descent iteration was chosen in this work because it is simple to implement and guaranteed to converge for quadratic objective functions. The preconditioned conjugate-gradient algorithm [1] would be more appropriate in practice.

Although preconditioning can improve convergence speed, these methods generally involve a time penalty. For the FFT methods used here, the penalty was only about 25%. This was more than offset by the increase in convergence rate for both the unregularized and regularized iterations. In the noiseless, unregularized example, the convergence rate using the preconditioned methods was approximately four times that of the unpreconditioned rate in the early iterations. This became as much as a factor of 30 in the late iterations. Results for the regularized example were not as dramatic, but yet, we still observed improvements in the convergence properties that outweighted the additional reconstruction time. Furthermore, as the dimensions of the reconstruction problem grow, the preconditioning overhead, which increases at the FFT rate $O(n \log n)$ will decrease relative to the projection and backprojection times, which tend to increase as $O(n^2)$.

The preconditioning filters chosen for this study were simple apodized inverse-filters designed from spatially invariant approximations to the spatially varying system response. In this study, choosing the gain limiting term as Γ_2 seemed to give the best overall convergence performance in both examples. Nevertheless, the design of optimal Fourier preconditioning filters for tomographic systems in an open question in which the singular-value decompositions of the matrices $A^TK^{-1}A$ and $A^TK^{-1}AM^{-1}$ will play a major role in answering.

VII. Conclusion

Preconditioning is useful for enhancing convergence rates of iterative reconstruction methods when it is desirable not to alter a preselected solution criterion. In this study we developed the idea of applying block-circulant preconditioning methods to ECT image reconstruction problems. We demonstrated that convergence rates for a test object and a simulated rotating gamma camera tomograph could be significantly improved for both least-squares and regularized least-squares examples. While the Fourier methods described in this paper are quite simple to implement and can be applied to a variety of iterative methods, there are several routes for further investigation as well—including the determination of optimal Fourier preconditioning filters and the investigation of alternative preconditioners. In particular, polynomial preconditioners are attractive not only because they are simple to implement but also because their effect on the spectrum of $A^TK^{-1}A$ can be readily determined [3].

VIII. Appendix I
Maximum Likelihood Estimation

If we assume that the projection data, $\boldsymbol{y} = (y_1, \cdots, y_n)^T$, is Poisson-distributed with mean $A\boldsymbol{x}$ where $\boldsymbol{x} = (x_1, \cdots, x_m)^T$, then maximizing the likelihood of observing the projection data given the object is the same as maximizing,

$$\phi(\boldsymbol{x}) = \sum_{i=1}^{n} \left[y_i \ln \sum_k a_{ik} x_k - \sum_k a_{ik} x_k \right] \tag{22}$$

where a_{ik} is the corresponding element of the system response matrix A.

To solve for the most-likely estimate, we set all derivatives to zero and solve for $\boldsymbol{x}$:

$$\frac{\partial \phi(\boldsymbol{x})}{\partial x_j} = \sum_{i=1}^{n} \left[\frac{a_{ij} y_i}{\sum_k a_{ik} x_k} - a_{ij} \right] = 0, \tag{23}$$

or equivalently,

$$\sum_{i=1}^{n} \frac{a_{ij}}{\sum_k a_{ik} x_k} \left[y_i - \sum_k a_{ik} x_k \right] = 0. \tag{24}$$

If we restack (24) into vector notation and take the elements of K to be $k_{ii} = \sum_p a_{ip} x_p$, $k_{ij} = 0, i \neq j$ we obtain

$$A^T K^{-1}(\boldsymbol{x}) \boldsymbol{y} - A^T K^{-1}(\boldsymbol{x}) A \boldsymbol{x} = 0, \tag{25}$$

from which (4) can be obtained.

Note that although equation (4) represents a maximum likelihood estimate of the object parameters it is *not* generally the same estimate that would be produced by the EM algorithm. The difference lies in that (4) allows some negative pixel values in $\boldsymbol{x}$ while estimates of $\boldsymbol{x}$ produced via the EM algorithm are naturally constrained to lie in the positive orthant of $\mathbb{R}^m$.

IX. Appendix II
Convergence Expressions

Define the projection space error vector at iteration $k+1$ as

$$\zeta_{k+1} = (\boldsymbol{y} - A\boldsymbol{x}_{k+1}). \tag{26}$$

Substituting (6) into (26) yields the recurrence relation for the projection space error

$$\zeta_{k+1} = (I - AM_k^{-1} A^T K_k^{-1}) \zeta_k. \tag{27}$$

From this, if K^{-1} does not vary with iteration, we can derive the following recurrence for the correction vectors, ε_k defined in (7), by premultiplying both sides of (27) by $A^T K^{-1}$:

$$\varepsilon_{k+1} = (I - A^T K^{-1} A M_k^{-1}) \varepsilon_k. \tag{28}$$

Taking norms gives,

$$\|\varepsilon_{k+1}\| = \|(I - A^T K^{-1} A M_k^{-1}) \varepsilon_k\| \tag{29}$$

$$\leq \|I - A^T K^{-1} A M_k^{-1}\| \, \|\varepsilon_k\| \tag{30}$$

where the inequality follows from the submultiplicative property of the matrix and vector 2-norms [1].

Acknowledgment

We would like to express our thanks to the anonymous reviewers for their excellent suggestions.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations, Second Edition*. Baltimore, MD: The Johns Hopkins University Press, 1989.
[2] P. P. B. Eggermont, G. T. Herman, and A. Lent, "Iterative algorithms for large partitioned linear systems with applications to image reconstruction," *Linear Algebra Appl.*, vol. 40, 37–67, 1981.
[3] T.-S. Pan, "The generalized Landweber iteration in positron emission tomography," Ph.D. dissertation, Univ. Michigan, Dep. EECS, Ann Arbor, MI, 1991.
[4] C.-C. J. Kuo and B. C. Levy, "Discretization and solution of elliptic partial differential equations—A digital signal processing approach," *IEEE Proc.*, vol. 78, pp. 1808–1842, 1990.
[5] O. N. Strand, "Theory and methods related to the singular-function expansion and Landweber's iteration for integral equations of the first kind," *SIAM J. Numer. Anal.*, vol. 11, no. 4, pp. 798–825, 1974.
[6] R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained iterative restoration algorithms," *IEEE Proc.*, vol. 69, pp. 432–450, 1981.
[7] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York: Springer-Verlag, 1982, pp. 125–134.
[8] B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Comput.*, vol. C-22, pp. 805–812, 1973.
[9] G. Demoment, "Image reconstruction and restoration: overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2024–2036, 1989.
[10] X. Wang, "Weighted regularization and its applications," Ph.D. dissertation, Univ. Michigan, Dep. EECS, Ann Arbor, MI, 1991.
[11] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 37–51, 1987.
[12] J. A. Stamos, W. L. Rogers, N. H. Clinthorne, and K. F. Koral, "Object-dependent performance comparison of two iterative reconstruction algorithms," *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 611–614, 1988.