COMBINED DIAGONAL/FOURIER PRECONDITIONING METHODS FOR IMAGE RECONSTRUCTION IN EMISSION TOMOGRAPHY

Scott D. Booth and Jeffrey A. Fessler

Bioengineering Program and Dept. of Electrical Engineering and Computer Science
University of Michigan

ABSTRACT

Iterative methods for tomographic image reconstruction often converge slowly. Preconditioning methods can often accelerate gradient-based iterations. Previous preconditioning methods for PET reconstruction have used either diagonal or Fourier-based preconditioners. Fourier-based preconditioners are well suited to problems with near-circulant Hessian matrices. However, due to the nonuniform Poisson noise variance in PET, the circulant approximation to the Hessian is suboptimal. This paper shows that a particular combined diagonal/Fourier preconditioner yields a more accurate approximation to the Hessian and gives significantly faster convergence rates than does either preconditioner used alone.

1. INTRODUCTION

Statistical methods for reconstructing PET and SPECT images require the maximization of some objective function. For realistic image sizes, direct maximization methods are computationally intractable, so iterative methods are required. Several authors have described conjugate gradient methods for PET and SPECT [1–4]. This paper describes a new preconditioning method for accelerating the convergence of the conjugate gradient iteration for penalized weighted least-squares image reconstruction in PET.

Kaufman [1, 5] showed that the EM algorithm for emission tomography is equivalent to a gradient ascent iteration with an estimate-dependent diagonal preconditioner. Mumcuoglu *et al* [2] incorporated this diagonal preconditioner into a conjugate-gradient algorithm. Clinthorne *et al* [6] proposed a Fourier-based preconditioner for iterative reconstruction in PET, which exploits the fact that $\mathbf{A}^T\mathbf{A}$ is approximately circulant-block-circulant (where $\mathbf{A}$ is the system matrix described below). Both diagonal and Fourier preconditioners increase the convergence rate of gradient-based iterations. Unfortunately, the Poisson measurement noise in PET imaging has a covariance matrix $\mathbf{K}$ with a very nonuniform diagonal, due to the effects of attenuation, detector efficiency, etc. [7]. Therefore the term $\mathbf{A}^T\mathbf{K}^{-1}\mathbf{A}$ within the Hessian (see (2) below) is not particularly well approximated by a circulant-block-circulant preconditioner (see Fig. 1). (Clinthorne *et al* used $\mathbf{K} = \mathbf{I}$ in their simulations [6].) Further, $\mathbf{A}^T\mathbf{A}$ is not particularly well approximated by any diagonal matrix, since the $1/r$ characteristics of tomographic systems give significant off-diagonal content. Thus, neither the Fourier or diagonal preconditioner is optimal for PET image reconstruction.

In related work, we have shown that $\mathbf{A}^T\mathbf{K}^{-1}\mathbf{A}$ can be reasonably well approximated by a matrix of the form $\mathbf{\Lambda}\mathbf{A}^T\mathbf{A}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a particular diagonal matrix [8, 9]. This property suggests that a *combined diagonal/Fourier preconditioner* should yield faster convergence than is realized by the use of either preconditioner alone. In the following we summarize the development of the combined diagonal/Fourier preconditioner, and present empirical results demonstrating significantly improved convergence rates.

This work was supported in part by DOE grant DE-FG02-87ER60561 and NIH grants CA-60711 and CA-54362.

2. THEORY

PET projection measurements approximately obey the Poisson statistical model:

$$y \sim \text{Poisson}\{\mathbf{A}x\},$$

where $x = [x_1 \dots x_p]^T$ is a vector of the true emission distribution, y is a vector of the measured projection data, and $\mathbf{A}$ is the projection matrix. One useful statistical criterion for estimating x from y is the penalized weighted least-squares objective:

$$\Phi(x) =$$

$$\frac{1}{2}(y - \mathbf{A}x)^T \mathbf{K}^{-1}(y - \mathbf{A}x) + \beta \sum_{j=1}^{p} \sum_{k \in \mathcal{N}_j} w_{jk}\phi(x_j - x_k),$$

where $\mathbf{K}$ is the covariance matrix of y (or the data-based approximate covariance [7]) and the term on the right is a roughness penalty. The unregularized estimate (with β=0) is poorly conditioned, so some regularization is required to ensure a stable solution. For simplicity, in this paper we use the quadratic penalty $\phi(x) = \frac{1}{2}x^2$ and a first-order neighborhood $\mathcal{N}_j$, although the method is applicable to other convex penalties. The regularization parameter β controls the resolution/noise tradeoff. Methods for choosing β to specify a desired resolution are described in [8, 9].

For a quadratic penalty, the objective simplifies to

$$\Phi(x) = \frac{1}{2}(y - \mathbf{A}x)^T \mathbf{K}^{-1}(y - \mathbf{A}x) + \frac{1}{2}\beta x^T \mathbf{R}x,$$

where the matrix $\mathbf{R}$ is the (symmetric) Hessian of the penalty. The usual “uniform” roughness penalty matrix $\mathbf{R}^u$ has 4’s along its diagonal and -1’s in the off-diagonal positions corresponding to each pixel's neighbors. Surprisingly, this uniform penalty leads to *nonuniform spatial resolution* [8, 9]. Thus, here we follow [8, 9] and use a modified penalty in which we replace w_{jk} above with $w_{jk}\kappa_j\kappa_k$, where

$$\kappa_j = \sqrt{\sum_i A_{ij}^2 K_{ii}^{-1} / \sum_i A_{ij}^2}.$$

In this case one can show that the penalty Hessian satisfies

$$\mathbf{R} \approx \mathbf{\Lambda} \mathbf{R}^u \mathbf{\Lambda}, \tag{1}$$

where $\mathbf{\Lambda} = \text{diag}\{\kappa_j\}$. This modified penalty leads to nearly uniform spatial resolution [8, 9], and also fortuitously leads to a Hessian that is more amenable to preconditioning.

For simplicity, in this paper we ignore the nonnegativity constraint for x, although the methods could be extended to include a barrier function as described in [2]. In the unconstrained case, one can minimize the quadratic objective by setting its derivative to zero, yielding the "ideal" estimate:

$$\hat{x} = \left(\mathbf{A}^T\mathbf{K}^{-1}\mathbf{A} + \beta\mathbf{R}\right)^{-1}\mathbf{A}^T\mathbf{K}^{-1}y.$$

This corresponds to the solution of the matrix equation $\mathbf{H}\hat{x} = b$, where $\mathbf{H}$ is the Hessian of Φ and b is the backprojection of y:

$$\begin{aligned} \mathbf{H} &= \mathbf{A}^T\mathbf{K}^{-1}\mathbf{A} + \beta\mathbf{R} \\ b &= \mathbf{A}^T\mathbf{K}^{-1}y. \end{aligned} \tag{2}$$

Unfortunately, the analytical solution $\hat{x} = \mathbf{H}^{-1}b$ cannot be computed directly because the $\mathbf{A}$ matrix is too large. Thus, iterative methods must be used to compute $\hat{x}$.

The preconditioned conjugate-gradient iteration is a well known method for solving $\mathbf{H}\hat{x} = b$. In the following sections, we describe this algorithm and discuss three choices of preconditioner.

2.1. Conjugate Gradient Function Minimization

Traditional methods of iterative function minimization, including the steepest-descent (SD) method, use the gradient of the objective function $\nabla\Phi(\hat{x})$ to determine a series of direction vectors d^n over which Φ is minimized:

$$\hat{x}^{n+1} = \hat{x}^n + \alpha d^n,$$

where α is chosen to minimize $\Phi(\hat{x}^n + \alpha d^n)$ [10]. The method of conjugate gradients modifies the search directions d^n to ensure that they are mutually orthogonal. This method minimizes quadratic objective functions in a finite number of steps, and converges faster than steepest-descent for both quadratic and nonquadratic objectives [2, 10].

A preconditioned form of the Polak-Ribiere Conjugate Gradient (CG) method [2, 10] is summarized below:

$$\hat{x}^{n+1} = \hat{x}^n + \alpha_n d^n$$

$$d^n = p^n + \gamma_n d^{n-1}$$

$$p^n = -\mathbf{M}\nabla^T\Phi(\hat{x}^n)$$

$$\gamma_n = \frac{\left[\nabla\Phi(\hat{x}^n) - \nabla\Phi\left(\hat{x}^{n-1}\right)\right] p^n}{\nabla\Phi(\hat{x}^{n-1})\, p^{n-1}}$$

$$\nabla^T\Phi(\hat{x}^n) = \mathbf{A}^T\mathbf{K}^{-1}(\mathbf{A}\hat{x}^n - y) + \beta\mathbf{R}\hat{x}^n = \mathbf{H}\hat{x}^n - b.$$

The positive-definite matrix $\mathbf{M}$ determines the preconditioner, and the step size α_n is determined as follows:

$$\begin{aligned} \alpha_n &= \min_\alpha \Phi(\hat{x}^n + \alpha d^n) \\ &= \frac{(d^n)^T\mathbf{A}^T\mathbf{K}^{-1}(y - \mathbf{A}\hat{x}^n) - \beta(d^n)^T\mathbf{R}\hat{x}^n}{(d^n)^T\mathbf{A}^T\mathbf{K}^{-1}\mathbf{A}d^n + \beta(d^n)^T\mathbf{R}d^n}. \end{aligned}$$

Preconditioning the CG algorithm is equivalent to solving the equation $\mathbf{M}\mathbf{H}\hat{x} = \mathbf{M}b$. Thus, the ideal preconditioner would be $\mathbf{M} = \mathbf{H}^{-1}$. Since we cannot directly compute $\mathbf{H}^{-1}$ for realistic imaging scenarios, we seek preconditioners that approximate $\mathbf{H}^{-1}$.

Figure 1a displays a typical Hessian matrix for a small imaging problem (one for which the Hessian could be easily computed). Note that the Hessian is approximately circulant-block-circulant, but the Hessian is nonuniform along the diagonal. The preconditioners described below exploit these two properties.

2.2. Diagonal Preconditioning

The simplest preconditioner uses the inverse of the diagonal elements of $\mathbf{H}$:

$$\mathbf{M} = \mathbf{D}^{-1}, \quad \mathbf{D} = \text{diag}\{H_{jj}\}.$$

Since $\mathbf{A}$ is sparse, the diagonals H_{jj} of the Hessian are easily computed without computing $\mathbf{H}$ itself. This diagonal matrix $\mathbf{D}$ effectively rescales the problem so that $\mathbf{D}^{-1}\mathbf{H}$ typically has a smaller condition number than $\mathbf{H}$. Applying this preconditioner to the CG algorithm is trivial:

$$p^n = \mathbf{D}^{-1}\nabla^T\Phi(\hat{x}^n).$$

Figure 1b displays the diagonally-preconditioned Hessian.

Note that this diagonal preconditioner differs from the EM-based diagonal preconditioner used in [2, 5].

2.3. Fourier Preconditioning

Since $\mathbf{A}^T\mathbf{A}$ and $\mathbf{R}^u$ are approximately block-circulant[2], the Fourier basis is a natural choice for a preconditioner [3]. The Fourier preconditioner is derived from the point-spread function of the $\mathbf{A}^T\mathbf{A}$ (projection-backprojection) operation. A single pixel at the center of the image field (a delta function) is projected and then backprojected. This creates a 2D spatially-invariant blurring filter which models the $\mathbf{A}^T\mathbf{A}$ operation:

$$\mathbf{\Omega} = \text{diag}\left\{\text{fft}_2\left[\left(c\mathbf{A}^T\mathbf{A} + \beta\mathbf{R}^u\right)\delta_{(0,0)}\right]\right\}$$

$$c\mathbf{A}^T\mathbf{A} + \beta\mathbf{R}^u \approx \mathbf{Q}\mathbf{\Omega}\mathbf{Q}^\star,$$

where $\mathbf{Q}$ is the 2D Fourier basis, $\mathbf{Q}^\star$ denotes the conjugate transpose of $\mathbf{Q}$, c is a constant related to the covariance matrix $\mathbf{K}$, and $\mathbf{R}^u$ is the uniform roughness penalty matrix discussed above. Figure 1c displays the Fourier-preconditioned

[2]More precisely, if $\mathbf{A} = \text{diag}\{c_i\}\mathbf{G}$, where c_i are scan-dependent correction factors and $\mathbf{G}$ is the system geometric response, then it is $\mathbf{G}^T\mathbf{G}$ that is approximately circulant-block-circulant, rather than $\mathbf{A}^T\mathbf{A}$. These important details will be treated more precisely in a forthcoming paper.

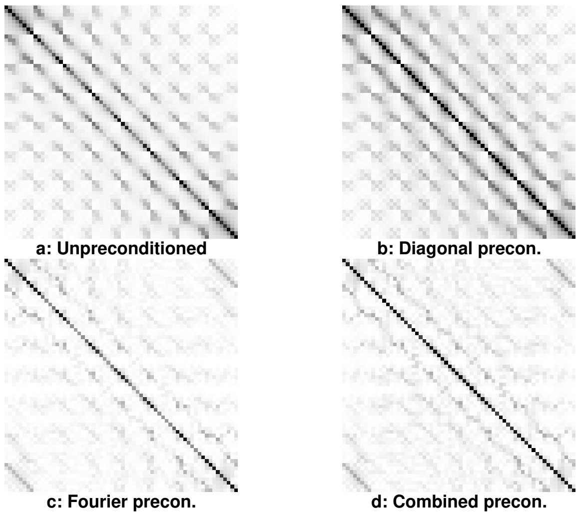

Figure 1: Images of the preconditioned Hessian $\mathbf{MH}$ for an 8-by-8-pixel test image. Preconditioners: (a) none, (b) diagonal, (c) Fourier, (d) combined diagonal/Fourier. The diagonal elements of the Fourier preconditioned Hessian range from .55 to 1.48, whereas the diagonal elements of $\mathbf{MH}$ for the combined diagonal/Fourier preconditioner range from .77 to 1.04. The eigenvalues of $\mathbf{MH}$ similarly have narrower spread for the combined diagonal/Fourier preconditioner, which leads to faster convergence.

Hessian $\mathbf{Q}\mathbf{\Omega}^{-1}\mathbf{Q}^{\star}\mathbf{H}$. This preconditioner can be efficiently incorporated into the CG algorithm using 2D FFTs:

$$p^n = \mathbf{Q}\mathbf{\Omega}^{-1}\mathbf{Q}^{\star}\nabla^T\Phi\left(\hat{x}^n\right) = \mathrm{fft}_2^{-1}\left(\mathbf{\Omega}^{-1}\mathrm{fft}_2\left[\nabla^T\Phi\left(\hat{x}^n\right)\right]\right).$$

Since the image space is discrete (divided into pixels), one must ensure that the back-projected point-spread function $\mathbf{A}^T\mathbf{A}\delta_{(0,0)}$ is symmetric about the "center" of the image (i.e. the pixel that corresponds to zero-frequency in the Fourier domain). In practice, this requires first averaging the $\mathbf{A}^T\mathbf{A}\delta_{(0,0)}$ matrix with flipped and shifted versions of itself before computing $\mathbf{\Omega}$. This ensures that the direction vector p^n is a real vector.

2.4. A Combined Diagonal/Fourier Method

Since the utility of the Fourier preconditioning method hinges on the assumption that the Hessian is approximately circulant-block-circulant, making this assumption more accurate should make the Fourier preconditioner more effective. In [9] the approximation $\mathbf{A}^T\mathbf{K}^{-1}\mathbf{A} \approx \mathbf{\Lambda}\mathbf{A}^T\mathbf{A}\mathbf{\Lambda}$ is discussed in more detail, where $\mathbf{\Lambda}$ was defined above. Using this approximation, note that:

$$\mathbf{H} = \mathbf{A}^T\mathbf{K}^{-1}\mathbf{A} + \beta\mathbf{R} \approx \mathbf{\Lambda}\mathbf{A}^T\mathbf{A}\mathbf{\Lambda} + \beta\mathbf{\Lambda}\mathbf{R}^u\mathbf{\Lambda}$$

$$= \mathbf{\Lambda}\left(\mathbf{A}^T\mathbf{A} + \beta\mathbf{R}^u\right)\mathbf{\Lambda} \approx \mathbf{\Lambda}\mathbf{Q}^{\star}\mathbf{\Omega}\mathbf{Q}\mathbf{\Lambda}.$$

This suggests that the combined diagonal/Fourier preconditioner

$$\mathbf{M} = \mathbf{\Lambda}^{-1}\mathbf{Q}^{\star}\mathbf{\Omega}^{-1}\mathbf{Q}\mathbf{\Lambda}^{-1}$$

should give a much closer approximation to $\mathbf{H}^{-1}$ than either diagonal or Fourier preconditioner alone. (Note that the diagonal $\mathbf{\Lambda}$ used here differs significantly from the conventional diagonal $\mathbf{D}$.) Figure 1d shows the preconditioned Hessian $\mathbf{MH}$ for the combined diagonal/Fourier preconditioner. The corresponding implementation in the CG algorithm is

$$p^n = \mathbf{\Lambda}^{-1}\mathrm{fft}_2^{-1}\left(\mathbf{\Omega}^{-1}\mathrm{fft}_2\left[\mathbf{\Lambda}^{-1}\nabla^T\Phi\left(\hat{x}^n\right)\right]\right).$$

The additional computation involved in the multiplication by the diagonal matrices $\mathbf{\Lambda}^{-1}$ is negligible relative to the FFTs. The resulting acceleration is remarkable, as shown by the results in the next section.

3. METHODS

Using Matlab on a DEC 3000/800 workstation, we performed iterative CG image reconstruction on a set of simulated projection data using the three preconditioning methods described above. For comparison, we also performed two steepest-descent reconstructions—one with no preconditioning, and one with the combined preconditioner. (The steepest-descent algorithm is obtained from the CG method by setting $\gamma = 0$.)

A Radon transform matrix $\mathbf{A}$ was chosen to simulate a tomographic projection of 94 bins by 70 angles. The data set y was created by projecting the intensity distribution x of a 64×64 simulated emission distribution (Hoffman brain phantom) using $\mathbf{A}$ to form a projection vector $\bar{y}$ representing the mean count rate seen by each detector:

$$\bar{y} = \mathbf{A}x.$$

This noiseless projection was then used as the parameter vector for a 6580-dimensional Poisson random variable y representing the number of counts actually recorded by the detectors in a time period t:

$$y = \mathrm{Poisson}\left(t\bar{y}\right).$$

The simulated scan time t was chosen to give a total of approximately 600,000 counts over the duration of the simulated scan. The diagonal covariance matrix $\mathbf{K}$ of the projection data y is estimated using the noisy data based on the Poisson statistics [7]

$$K_{ii} = \max\{10, y_i\}.$$

The iterations were initialized with a zero image. We used the modified roughness penalty (1) with $\beta = 0.001$.

4. RESULTS

The convergence rates for the unpreconditioned and the three preconditioned forms of the CG image reconstructions are illustrated in Fig.2, which plots the norm-error criterion

$$\frac{\|\hat{x}^n - \hat{x}^\infty\|}{\|\hat{x}^\infty\|} = \frac{\sqrt{\sum_{j=1}^p\left(\hat{x}_j^n - \hat{x}_j^\infty\right)^2}}{\sqrt{\sum_{j=1}^p\left(\hat{x}_j^\infty\right)^2}}$$

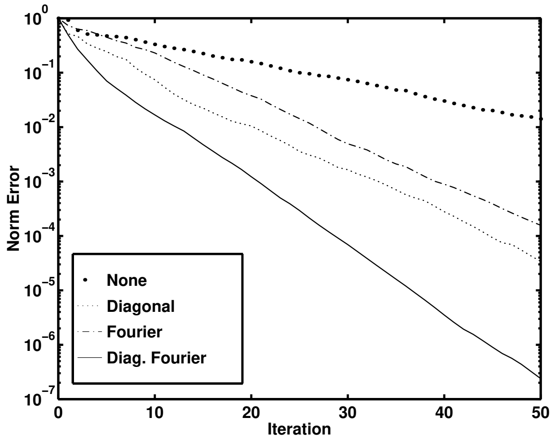

Figure 2: Normalized error $\|\hat{x}^n - \hat{x}^\infty\| / \|\hat{x}^\infty\|$ versus iteration n for unpreconditioned CG and for three preconditioned forms of CG. The initial image was zero.

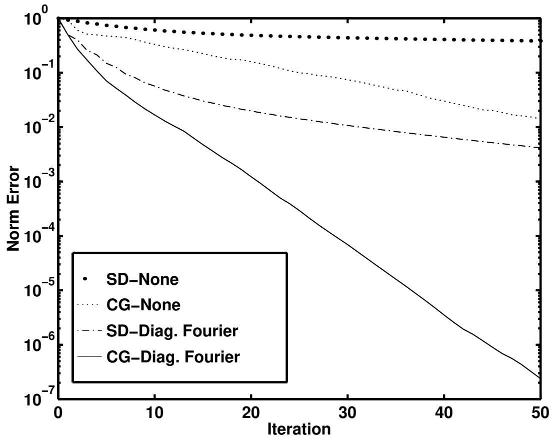

Figure 3: Comparison of the convergence rates of the SD and CG algorithms with combined diagonal/Fourier preconditioning and without preconditioning.

against the iteration number n, where $\hat{x}^\infty$ is a "converged" reconstruction created by iterating the CG algorithm until the overall change in the image from one iteration to the next was negligible. Figure 3 compares the convergence of the CG and SD algorithms for the unpreconditioned case and for the combined diagonal/Fourier preconditioner. Note that SD with combined diagonal/Fourier preconditioning is still significantly slower than CG with the combined preconditioner, suggesting there is still room for further improvement through the use of better preconditioners.

5. DISCUSSION

We have described three preconditioning methods for PET image reconstruction using the conjugate-gradient algorithm. The results clearly demonstrate that all three preconditioners improve the convergence rate, with the combined diagonal/Fourier preconditioner yielding significantly faster convergence than either diagonal or Fourier preconditioning alone. In other simulation results (not shown), the ranking of the diagonal preconditioner versus Fourier preconditioner depended on the initial image, but the combined diagonal/Fourier preconditioner consistently demonstrated the fastest convergence, reflecting the improved approximation to the Hessian illustrated in Fig. 1. In this example, the algorithm converged to within single-precision floating point accuracy (about 10^{-6}) in about 30 iterations. For noisy data, a numerical precision of, say, 1% is probably adequate, and in this case $\|\hat{x}^n - \hat{x}^\infty\| / \|\hat{x}^\infty\|$ falls below 0.01 and about $n = 8$ iterations. Future work will include modifying the preconditioner to accommodate the fact that $\mathbf{R}^u$ is less circulant-block-circulant when one uses nonquadratic penalties.

6. REFERENCES

[1] L Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET", *IEEE Tr. Med. Im.*, vol. 12, no. 2, pp. 200–214, June 1993.

[2] E U Mumcuoglu, R Leahy, S R Cherry, and Z Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images", *IEEE Tr. Med. Im.*, vol. 13, no. 3, pp. 687–701, Dec. 1994.

[3] D S Lalush and B M W Tsui, "A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm", *Med. Phys.*, 1994, Submitted.

[4] A H Delaney and Y Bresler, "A fast iterative tomographic reconstruction algorithm", in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 1995, vol. 4, pp. 2295–2298.

[5] L Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography", *IEEE Tr. Med. Im.*, vol. 6, no. 1, pp. 37–51, Mar. 1987.

[6] N H Clinthorne, T S Pan, P C Chiao, W L Rogers, and J A Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions", *IEEE Tr. Med. Im.*, vol. 12, no. 1, pp. 78–83, Mar. 1993.

[7] J A Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography", *IEEE Tr. Med. Im.*, vol. 13, no. 2, pp. 290–300, June 1994.

[8] J A Fessler and W L Rogers, "Uniform quadratic penalties cause nonuniform image resolution (and sometimes vice versa)", in *Conf. Rec. of the IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1994, vol. 4, pp. 1915–1919.

[9] J A Fessler et al., "Spatial resolution properties of penalized maximum-likelihood image reconstruction methods", 1994, In preparation.

[10] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling, *Numerical recipes in C*, Cambridge Univ. Press, 1988.