Fully 3D PET Image Reconstruction Using
A Fourier Preconditioned Conjugate-Gradient Algorithm

Jeffrey A. Fessler and Edward P. Ficaro

4240 EECS Bldg., University of Michigan, Ann Arbor, MI 48109-2122

ABSTRACT

Since the data sizes in fully 3D PET imaging are very large, iterative image reconstruction algorithms must converge in very few iterations to be useful. One can improve the convergence rate of the conjugate-gradient (CG) algorithm by incorporating preconditioning operators that approximate the inverse of the Hessian of the objective function. If the 3D cylindrical PET geometry were not truncated at the ends, then the Hessian of the penalized least-squares objective function would be approximately shift-invariant, i.e. $\mathbf{G'G}$ would be nearly block-circulant, where $\mathbf{G}$ is the system matrix. We propose a Fourier preconditioner based on this shift-invariant approximation to the Hessian. Results show that this preconditioner significantly accelerates the convergence of the CG algorithm with only a small increase in computation.

I. INTRODUCTION

Statistical methods for tomographic image reconstruction from fully 3D PET scans are particularly promising since the number of unknown parameters (pixels) is the same as in the 2D problem typically, but the number of measurements (rays) is many times larger. However, since the data sizes in fully 3D PET imaging are very large, an iterative image reconstruction algorithm stands to be clinically useful only if it converges in very few iterations.

Due to the large data sizes in 3D PET, it would be particularly inconvenient to have to acquire and store a separate set of "sinograms" or projections corresponding to random-coincidence events. Hence it is likely that in most 3D PET scans the random coincidences will be precorrected in real time using the delay-window method, rather than collected as a separate projection set (on systems that support this mode). This precorrection makes the data non-Poisson, so the usual Poisson-based iterative algorithms are inappropriate. We propose instead to reconstruct the image volume by minimizing a penalized weighted least-squares (PWLS) objective function [1]. (Although see [2] for possible alternatives.)

II. THEORY

Both the 2D and 3D reconstruction problems can be described by the measurement model:

$$y = \mathbf{G}\underline{x} + \text{noise} \qquad (1)$$

where $\underline{y}$ is the measured sinograms, $\mathbf{G}$ is the system model (geometric response), and $\underline{x} = [x_1, \ldots, x_p]'$ is the image or volume, where p is the number of pixels or voxels. We would like to estimate $\underline{x}$ from $\underline{y}$ by minimizing the PWLS objective function:

$$\hat{\underline{x}} = \arg\min_{\underline{x}} \Phi(\underline{x})$$

$$\Phi(\underline{x}) = (\underline{y} - \mathbf{G}\underline{x})'\mathbf{W}(\underline{y} - \mathbf{G}\underline{x}) + \beta R(\underline{x}), \qquad (2)$$

where $\mathbf{W}$ is an estimate of the inverse of the noise covariance [1], $R(\underline{x})$ is a penalty function that discourages rough images, and β controls the tradeoff between resolution and noise [3].

Most of the convex penalty functions proposed for regularization of imaging problems can be expressed in the following very general form, e.g. [4]:

$$R(\underline{x}) = \sum_{k=1}^{K} \psi_k([\mathbf{C}\underline{x} - \underline{c}]_k), \qquad (3)$$

where $\mathbf{C}$ is a $K \times p$ matrix and $\underline{c} \in \mathbb{R}^K$, for some user-defined number K of soft "constraints" of the form $[\mathbf{C}\underline{x}]_k \approx c_k$.

The standard roughness penalty, which penalizes differences between neighboring pixel values, is the special case of (3) where $\underline{c} = \underline{0}$, where K is the number of pairs of neighboring pixels[1], and where each row of $\mathbf{C}$ contains one "+1" and one "-1" entry so that

$$[\mathbf{C}\underline{x}]_k = x_{j_{k,1}} - x_{j_{k,2}}, \quad k = 1, \ldots, K,$$

where $x_{j_{k,1}}$ and $x_{j_{k,2}}$ are two neighboring pixels.

The preconditioned CG algorithm is naturally suited to the PWLS objective, and has been used extensively in tomography, e.g. [5–11]. The preconditioned form of the Polak-Ribiere CG method [12] is

[1] For 3D penalty functions, $K \approx 3p$ and $K \approx 13p$ for first and second-order neighborhoods respectively.

given as follows:

$$g^n = -\nabla'\Phi(\underline{x}^n) \quad \text{(-gradient)}$$

$$p^n = \mathbf{M}g^n \quad \text{(precondition)} \tag{4}$$

$$\gamma_n = \begin{cases} 0, & n = 0 \\ \dfrac{\langle g^n - g^{n-1}, p^n \rangle}{\langle g^{n-1}, p^{n-1} \rangle}, & n > 0 \end{cases}$$

$$\underline{d}^n = \underline{p}^n + \gamma_n \underline{d}^{n-1} \quad \text{(search direction)}$$

$$\alpha_n = \arg\min_{\alpha} \Phi(\underline{x}^n + \alpha \underline{d}^n) \quad \text{(step size)} \tag{5}$$

$$\underline{x}^{n+1} = \underline{x}^n + \alpha_n \underline{d}^n \quad \text{(update)}.$$

Computing the gradient of the objective function requires a forward projection and backprojection:

$$-\nabla'\Phi(\underline{x}) = \mathbf{G}'\mathbf{W}(\underline{y} - \mathbf{G}\underline{x}) - \beta\mathbf{C}'\text{diag}\left\{\dot{\psi}_k([\mathbf{C}\underline{x} - \underline{c}]_k)\right\}$$

$$= \mathbf{G}'\mathbf{W}(\underline{y} - \mathbf{G}\underline{x}) - \beta\mathbf{C}'\mathbf{D}_\omega(\underline{x})(\mathbf{C}\underline{x} - \underline{c})$$

$$= [\mathbf{G}'\mathbf{W}\underline{y} + \mathbf{C}'\mathbf{D}_\omega(\underline{x})\underline{c}] - [\mathbf{G}'\mathbf{W}\mathbf{G} + \beta\mathbf{C}'\mathbf{D}_\omega(\underline{x})\mathbf{C}]\underline{x}$$

$$= \underline{b}(\underline{x}) - \mathbf{H}_\omega(\underline{x}), \tag{6}$$

where

$$\mathbf{H}_\omega(\underline{x}) = \mathbf{G}'\mathbf{W}\mathbf{G} + \beta\mathbf{C}'\mathbf{D}_\omega(\underline{x})\mathbf{C}, \tag{7}$$

$$\mathbf{D}_\omega(\underline{x}) = \text{diag}\{\omega_k([\mathbf{C}\underline{x} - \underline{c}]_k)\},$$

$$\underline{b}(\underline{x}) = \mathbf{G}'\mathbf{W}\underline{y} + \mathbf{C}\mathbf{D}_\omega(\underline{x})\underline{c},$$

$$\omega_k(t) = \frac{\dot{\psi}_k(t)}{t} = \frac{\frac{d}{dt}\psi_k(t)}{t}. \tag{8}$$

The computation of $\mathbf{G}'\mathbf{W}(\underline{y} - \mathbf{G}\underline{x})$ is the most time consuming step in the algorithm, so minimizing the number of iterations is essential.

One can significantly improve the convergence rate of the conjugate-gradient algorithm by incorporating preconditioning operators $\mathbf{M}$ in (4) that approximate $\mathbf{H}^{-1}$, the inverse of the Hessian of the objective function, defined by

$$\mathbf{H}(\underline{x}) = \nabla^2\Phi(\underline{x}) = \mathbf{G}'\mathbf{W}\mathbf{G} + \beta\mathbf{C}'\mathbf{D}_{\ddot{\psi}}(\underline{x})\mathbf{C}$$

where $\ddot{\psi}_k(t) = d^2/dt^2\psi_k(t)$ and

$$\mathbf{D}_{\ddot{\psi}}(\underline{x}) = \text{diag}\left\{\ddot{\psi}_k([\mathbf{C}\underline{x} - \underline{c}]_k)\right\}.$$

But we would also like $\mathbf{M}$ to be easy to compute, unlike $\mathbf{H}^{-1}$.

The standard *diagonal preconditioner* is simply:

$$\mathbf{M}_D = \text{diag}\{\mathbf{H}_{jj}\},$$

which improves the convergence rate somewhat, but is suboptimal since it ignores the off-diagonal structure of the Hessian.

For some tomographic systems, the matrices $\mathbf{G}'\mathbf{G}$ and $\mathbf{C}'\mathbf{C}$ are approximately block-Toeplitz or block-circulant, i.e. they correspond to *shift-invariant* operators. Block-circulant matrices are diagonalized by the Fourier basis. In other words,

$$\mathbf{K}(\eta) = \mathbf{G}'\mathbf{G} + \eta\mathbf{C}'\mathbf{C} \approx \mathbf{Q}'\boldsymbol{\Omega}(\eta)\mathbf{Q},$$

where $\mathbf{Q}$ is the 2D or 3D DFT matrix, and $\boldsymbol{\Omega}(\eta)$ is the diagonal matrix of the DFT coefficients of the kernel of $\mathbf{K}$. For the *unweighted* least-squares problem with $\mathbf{W} = \mathbf{I}/\sigma^2$ and for a *quadratic penalty* (QPULS), we have

$$\mathbf{H} = \frac{1}{\sigma^2}\mathbf{G}'\mathbf{G} + \beta\mathbf{C}'\mathbf{C}$$

so the following *Fourier preconditioner* leads to very rapid convergence:

$$\mathbf{M}_F = \mathbf{Q}'\boldsymbol{\Omega}^{-1}(\sigma^2\beta)\mathbf{Q} \approx \mathbf{K}^{-1}(\sigma^2\beta).$$

Clinthorne *et al.* applied this idea to 2D tomography in [13].

In PET the noise covariance $\mathbf{W}^{-1}$ is highly nonuniform, so the Fisher information matrix $\mathbf{G}'\mathbf{W}\mathbf{G}$ is very *shift variant*. Thus Fourier preconditioning is suboptimal. In [10, 14] we derived the following *combined diagonal/Fourier* preconditioner:

$$\mathbf{M} = \text{diag}\left\{\kappa_j^{-1}\right\}\mathbf{T}'\mathbf{Q}'\boldsymbol{\Omega}^{-1}(\beta)\mathbf{Q}\mathbf{T}\,\text{diag}\left\{\kappa_j^{-1}\right\}$$

which uses the following approximation from [3]:

$$\mathbf{G}'\mathbf{W}\mathbf{G} \approx \text{diag}\{\kappa_j\}\mathbf{G}'\mathbf{G}\,\text{diag}\{\kappa_j\}, \tag{9}$$

where

$$\kappa_j = \sqrt{\frac{\sum_i g_{ij}^2 \mathbf{W}_{ii}}{\sum_i g_{ij}^2}}. \tag{10}$$

The above form is for quadratic penalties with the modification described in [3].

Additional preconditioners for *nonquadratic penalties* are described in [14].

III. 3D PET

If the 3D cylindrical PET geometry were not truncated at the ends, then the Hessian of the penalized least-squares objective function would be approximately shift-invariant. We propose a Fourier preconditioner based on this shift-invariant approximation

to the Hessian. This approach is a 3D generalization of the ideas in [13,10].

To evaluate this approach, we performed a simulation for a small 3D PET geometry consisting of four 64^2 image slices and 3D projections at tilts $\phi = \{-1°, 0°, 1°\}$. Each sinogram had 64 radial bins by 70 angles over 180°. The measurement model is as in (1), where the three terms are defined in Fig. 2. To estimate the image volume $\underline{x}$, we would like to minimize the penalized weighted least-squares objective function in (2), where $R(\underline{x})$ is a 3D roughness penalty.

For simplicity, we initially tried a block-diagonal 2D preconditioner:

$$\mathbf{M} = \begin{bmatrix} \mathbf{Q'\Omega Q} & & & \\ & \mathbf{Q'\Omega Q} & & \\ & & \mathbf{Q'\Omega Q} & \\ & & & \mathbf{Q'\Omega Q} \end{bmatrix}$$

where $\mathbf{Q}$ is the 2D Fourier matrix, and Ω is the frequency response of the kernel of the approximately Toeplitz matrix $\mathbf{G}'_{0,0}\mathbf{G}_{0,0} + \beta\mathbf{R}_{2D}$ and $\mathbf{R}_{2D}$ is the in-plane roughness penalty. In the implementation we use 2D FFTs to perform the multiplication by $\mathbf{Q}$.

The simulated object is shown in Fig. 1. Using noiseless data and initializing with a uniform image, we iterated the conjugate-gradient algorithm with and without the preconditioner described above. Fig. 1 compares the objective function $\Phi(\hat{x})$ at each iteration of the algorithm. Clearly the proposed preconditioner accelerates the convergence of the conjugate-gradient algorithm. The computation of the FFTs is very mild compared to the forward and back-projection operations, so it adds only a small increase in computation to the conjugate-gradient algorithm.

Future work includes more realistic system sizes, noisy data, nonquadratic penalties, non-negativity constraint, a 3D Fourier preconditioner (using 3D FFT), comparisons with the grouped coordinate ascent algorithm [15] etc. See http://www.eecs.umich.edu/~fessler for current preprints.

IV. ACKNOWLEDGEMENT

REFERENCES

[1] 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.

[2] M. Yavuz and J. A. Fessler, "Objective functions for tomographic reconstruction from randoms-precorrected PET scans," in Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf., 1996. To appear.

[3] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," IEEE Tr. Im. Proc., vol. 5, no. 9, pp. 1346–58, Sept. 1996.

[4] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," IEEE Tr. Med. Im., vol. 14, no. 1, pp. 132–137, Mar. 1995.

[5] S. Kawata and O. Nalcioglu, "Constrained iterative reconstruction by the conjugate gradient method," IEEE Tr. Med. Im., vol. 4, no. 2, pp. 65–71, June 1985.

[6] A. R. Formiconi, A. Pupi, and A. Passeri, "Compensation of spatial system response in SPECT with conjugate gradient reconstruction technique," Phys. Med. Biol., vol. 34, pp. 69–84, 1990.

[7] 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.

[8] 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.

[9] D. S. Lalush and B. M. W. Tsui, "A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm," Med. Phys., vol. 22, no. 8, pp. 1273–84, Aug. 1995.

[10] S. D. Booth and J. A. Fessler, "Combined diagonal/Fourier preconditioning methods for image reconstruction in emission tomography," in Proc. IEEE Intl. Conf. on Image Processing, vol. 2, pp. 441–4, 1995.

[11] A. H. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," IEEE Tr. Im. Proc., vol. 5, no. 5, pp. 740–53, May 1996.

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

[13] 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.

[14] J. A. Fessler and S. D. Booth. Conjugate-Gradient Preconditioning Methods for Shift-Variant Image Reconstruction, 1996. In preparation.

[15] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," IEEE Tr. Med. Im., 1996. To appear.
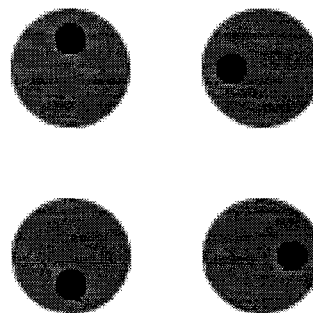
Fig. 1. 4-slice object used in simulation.

$$y = \begin{bmatrix} \underline{y}_{0,1} \\ \underline{y}_{0,2} \\ \underline{y}_{0,3} \\ \underline{y}_{0,4} \\ \cdots \\ \underline{y}_{1,1} \\ \underline{y}_{1,2} \\ \underline{y}_{1,3} \\ \underline{y}_{1,4} \\ \cdots \\ \underline{y}_{-1,1} \\ \underline{y}_{-1,2} \\ \underline{y}_{-1,3} \\ \underline{y}_{-1,4} \end{bmatrix}, \quad G = \begin{bmatrix} \mathbf{G}_{0,0} & & & \\ & \mathbf{G}_{0,0} & & \\ & & \mathbf{G}_{0,0} & \\ & & & \mathbf{G}_{0,0} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \mathbf{G}_{1,0} & \mathbf{G}_{1,1} & & \\ & \mathbf{G}_{1,0} & \mathbf{G}_{1,1} & \\ & & \mathbf{G}_{1,0} & \mathbf{G}_{1,1} \\ & & & \mathbf{G}_{1,0} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \mathbf{G}_{-1,0} & & & \\ \mathbf{G}_{-1,-1} & \mathbf{G}_{-1,0} & & \\ & \mathbf{G}_{-1,-1} & \mathbf{G}_{-1,0} & \\ & & \mathbf{G}_{-1,-1} & \mathbf{G}_{-1,0} \end{bmatrix}, \quad x = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \underline{x}_3 \\ \underline{x}_4 \end{bmatrix}$$

Fig. 2. Model definition for small 3D PET system used in simulations having 4 64^2 slices and 3 tilt angles $\phi = \{-1^\circ, 0^\circ, 1^\circ\}$. $\mathbf{G}_{\phi,z}$ denotes the system response for a given tilt ϕ and axial z-offset, and $\underline{y}_{\phi,z}$ denotes the corresponding sinogram measurement, each of which was 64 radial bins by 70 angles over 180°. The usual 2D PET geometry corresponds to the upper 4×4 block. Note that the additional rows of the system and measurement matrix mean additional counts and information, but of course at the price of superposition of the different slices.

Fig. 3. Minimization of the objective function by conjugate-gradient algorithm with and without the proposed Fourier preconditioner. The objective converges towards its minimum more rapidly with the proposed Fourier preconditioner.