Preconditioning Methods for Shift-Variant Image Reconstruction

Jeffrey A. Fessler
Dept. of Electrical Engineering and Computer Science
4240 EECS Bldg., University of Michigan, Ann Arbor, MI 48109-2122
Email: fessler@umich.edu, Voice: 313-763-1434, FAX: 313-764-8041

Abstract

Preconditioning methods can accelerate the convergence of gradient-based iterative methods for tomographic image reconstruction and image restoration. Circulant preconditioners have been used extensively for shift-invariant problems. Diagonal preconditioners offer some improvement in convergence rate, but do not incorporate the structure of the Hessian matrices in imaging problems. For inverse problems that are approximately shift-invariant (i.e. approximately block-Toeplitz or block-circulant Hessians), circulant or Fourier-based preconditioners can provide remarkable acceleration. However, in applications with nonuniform noise variance (such as arises from Poisson statistics in emission tomography and in quantum-limited optical imaging), the Hessian of the (penalized) weighted least-squares objective function is quite shift-variant, and the Fourier preconditioner performs poorly. Additional shift-variance is caused by edge-preserving regularization methods based on nonquadratic penalty functions. This paper describes new preconditioners that more accurately approximate the Hessian matrices of shift-variant imaging problems. Compared to diagonal or Fourier preconditioning, the new preconditioners lead to significantly faster convergence rates for the unconstrained conjugate-gradient (CG) iteration. Applications to position emission tomography (PET) illustrate the method.

I. Introduction

Most statistical methods for image reconstruction require minimizing an objective function related to the measurement statistics. For realistic image sizes, direct minimization methods are computationally intractable, so iterative methods are required. For objective functions that are convex, conjugate-gradient (CG) algorithms are appealing for reasons of simplicity and potential for parallelization [1–5], but convergence rates may be undesirably slow. Preconditioning methods can accelerate the convergence of gradient-based iterative methods for tomographic image reconstruction and image restoration.

Circulant or "Fourier" preconditioners have been used extensively for shift-invariant problems [6–8], for which the Hessian of the objective function is approximately block-Toeplitz or block-circulant. However, in applications with nonuniform noise variance (such as arises from Poisson statistics) the Hessian of the natural objective function is shift-variant, and the Fourier preconditioner performs poorly. Significant additional shift-variance is caused by edge-preserving regularization methods based on nonquadratic penalty functions. This paper summarizes new preconditioners that more accurately approximate the Hessian matrices of shift-variant imaging problems. By accommodating nonquadratic penalty functions, this work significantly generalizes our previous method [3]. The new preconditioners lead to significantly faster convergence rates. The method is applied to real position emission tomography (PET) data.

This work was supported in part by NIH grants CA-60711 and CA-54362 and by the Whitaker Foundation.

II. Regularized Reconstruction Problem

Consider the following general inverse problem: find an estimate of object parameters $\underline{x} = [x_1, \ldots, x_p]'$ from a measurement vector $\underline{y}$ related to $\underline{x}$ by

$$\underline{y} = \boldsymbol{G}\underline{x}_{\mathrm{true}} + \text{noise}. \tag{1}$$

One useful statistical criterion for estimating $\underline{x}$ from $\underline{y}$ is the following penalized weighted least-squares objective function [1, 9]:

$$\Phi(\underline{x}) = \frac{1}{2}(\underline{y} - \boldsymbol{G}\underline{x})'\boldsymbol{W}(\underline{y} - \boldsymbol{G}\underline{x}) + \beta R(\underline{x}) \tag{2}$$

where $R(\underline{x})$ is a penalty function that encourages smooth or piecewise-smooth estimates, and β controls the tradeoff between spatial resolution and noise [10]. Usually, $\boldsymbol{W}$ is the inverse of the covariance matrix of $\underline{y}$. Our goal is to compute an estimate $\hat{\underline{x}}$ of $\underline{x}_{\mathrm{true}}$ by minimizing Φ.

Most of the convex penalty functions proposed for regularizing imaging problems can be written as follows:

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

where $\boldsymbol{C} \in \mathbb{R}^{K \times p}$ and $\underline{c} \in \mathbb{R}^K$, for some number K of "soft constraints" of the form $[\boldsymbol{C}\underline{x}]_k \approx c_k$. We consider convex nonquadratic functions ψ_k that are symmetric, twice-differentiable, and have bounded, nonzero 2nd derivatives.

For simplicity we ignore any nonnegativity constraint for $\underline{x}$. One could extend the methods as in [2,11–13].

The Hessian of Φ is given by

$$\boldsymbol{H}(\underline{x}) = \boldsymbol{G}'\boldsymbol{W}\boldsymbol{G} + \beta \boldsymbol{C}'\boldsymbol{D}_{\ddot{\psi}}(\underline{x})\boldsymbol{C}, \quad (4)$$

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

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

The ideal preconditioner for a CG algorithm would be: $\boldsymbol{H}^{-1}(\underline{x})$ which would give superlinear convergence rates. Since $\boldsymbol{H}^{-1}$ is impractical for large p, one must find preconditioners that *approximate* $\boldsymbol{H}^{-1}$. The standard diagonal preconditioner is the inverse of the diagonal elements of $\boldsymbol{H}$:

$$\boldsymbol{M}_{\mathrm{D}}(\underline{x}) = \mathcal{D}[1/\boldsymbol{H}_{jj}(\underline{x})]. \quad (5)$$

It ignores the off-diagonal structure of the Hessian of Φ, the correlation between pixels and the $1/r$ response of tomographic systems, so is suboptimal. We have derived new preconditioners for this problem [14]; the derivation is summarized next.

III. Preconditioners

This section summarizes a sequence of approximations to $\boldsymbol{H}^{-1}$. The final approximations yield a new preconditioner that one can implement efficiently.

The first key approximation is one that we have used for analyzing spatial resolution [10]. Roughly speaking, this approximation brings the diagonal covariance matrix outside of the Fisher information matrix as follows:

$$\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G} \approx \boldsymbol{D}_\kappa \boldsymbol{G}'\boldsymbol{G}\boldsymbol{D}_\kappa, \quad (6)$$

where $\boldsymbol{D}_\kappa = \mathcal{D}[\kappa_j]$ and $\kappa_j = \sqrt{\sum_i g_{ij}^2 \boldsymbol{W}_{ii} / \sum_i g_{ij}^2}$. This exchange is useful because by assumption $\boldsymbol{G}'\boldsymbol{G}$ is approximately block-Toeplitz, i.e. shift-invariant, unlike $\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}$. The matrices on the two sides of approximation (6) are *exactly* equal along their diagonals, and would also be equal off-diagonal if the $\boldsymbol{W}_{ii}$'s were all equal. We also apply an approximation analogous to (6) to the regularization term:

$$\beta \boldsymbol{D}_\kappa^{-1} \boldsymbol{C}'\boldsymbol{D}_{\ddot{\psi}}(\underline{x})\boldsymbol{C}\boldsymbol{D}_\kappa^{-1} \approx \boldsymbol{D}_\eta(\underline{x})\boldsymbol{C}'\boldsymbol{C}\boldsymbol{D}_\eta(\underline{x}), \quad (7)$$

where $\boldsymbol{D}_\eta(\underline{x}) = \mathcal{D}\left[\sqrt{\eta_j(\underline{x})}\right]$ and

$$\eta_j(\underline{x}) = (\beta/\kappa_j^2) \sum_k c_{kj}^2 \ddot{\psi}_k([\boldsymbol{C}\underline{x} - \underline{c}]_k) / \sum_k c_{kj}^2.$$

(The factor $\eta_j(\underline{x})$ is an *effective regularization parameter* for the jth pixel [10].) This leads to the preconditioner[1]:

$$\begin{aligned} \boldsymbol{M}_2(\underline{x}) &= \boldsymbol{D}_\kappa^{-1}\boldsymbol{B}^{-1}(\underline{x})\boldsymbol{D}_\kappa^{-1}, \\ \boldsymbol{B}(\underline{x}) &= \boldsymbol{G}'\boldsymbol{G} + \boldsymbol{D}_\eta(\underline{x})\boldsymbol{C}'\boldsymbol{C}\boldsymbol{D}_\eta(\underline{x}), \end{aligned} \quad (8)$$

which is impractical. The preconditioners that we derive below all have the same form as (8), but with different approximations for the inverse of $\boldsymbol{B}(\underline{x})$.

To proceed, we exploit the heuristic that the effect of the penalty term is predominately *local*. A convenient mathematical expression for this heuristic is as follows:

$$\boldsymbol{B}^{-1/2}(\underline{x})\underline{e}_j \approx \boldsymbol{K}^{-1/2}(\eta_j(\underline{x}))\underline{e}_j \triangleq \underline{v}_j(\underline{x}), \quad (9)$$

$$\boldsymbol{K}(\eta) = \boldsymbol{G}'\boldsymbol{G} + \eta \boldsymbol{C}'\boldsymbol{C}, \quad (10)$$

where $\underline{e}_j$ is the jth unit vector. We build this approximation into a matrix by using the following exact expansion of any positive definite $p \times p$ matrix $\boldsymbol{B}$:

$$\boldsymbol{B}^{-1/2} = \boldsymbol{B}^{-1/2}\boldsymbol{I} = \boldsymbol{B}^{-1/2}\sum_{j=1}^p \underline{e}_j\underline{e}_j' = \sum_{j=1}^p \left(\boldsymbol{B}^{-1/2}\underline{e}_j\right)\underline{e}_j'.$$

Thus by applying (9) we have the approximation:

$$\boldsymbol{B}^{-1/2}(\underline{x}) \approx \sum_{j=1}^p \underline{v}_j(\underline{x})\underline{e}_j' \triangleq \boldsymbol{S}_3(\underline{x}), \quad (11)$$

or equivalently:

$$\boldsymbol{B}^{-1}(\underline{x}) = [\boldsymbol{G}'\boldsymbol{G} + \boldsymbol{D}_\eta(\underline{x})\boldsymbol{C}'\boldsymbol{C}\boldsymbol{D}_\eta(\underline{x})]^{-1} \approx \boldsymbol{S}_3'(\underline{x})\boldsymbol{S}_3(\underline{x}).$$

Thus (11) leads to the (impractical) preconditioner:

$$\boldsymbol{M}_3(\underline{x}) = \boldsymbol{D}_\kappa^{-1}\boldsymbol{S}_3'(\underline{x})\boldsymbol{S}_3(\underline{x})\boldsymbol{D}_\kappa^{-1}. \quad (12)$$

Now we use the fact that $\boldsymbol{K}$ is suitable for circulant approximations. There is a diagonal matrix $\boldsymbol{\Omega}(\eta)$ such that

$$\boldsymbol{K}(\eta) = [\boldsymbol{G}'\boldsymbol{G} + \eta\boldsymbol{C}'\boldsymbol{C}] \approx \boldsymbol{Q}'\boldsymbol{\Omega}^l(\eta)\boldsymbol{Q}, \quad (13)$$

where $\boldsymbol{Q}$ represents the 2D DFT [15]. Thus from (9):

$$\underline{v}_j(\underline{x}) \approx \underline{w}_j(\underline{x}) \triangleq \boldsymbol{Q}'\boldsymbol{\Omega}^{-1/2}(\eta_j(\underline{x}))\boldsymbol{Q}\underline{e}_j.$$

Applying the circulant approximation (13) to $\boldsymbol{M}_3$ gives the (impractical) preconditioner:

$$\begin{aligned} \boldsymbol{M}_5(\underline{x}) &= \boldsymbol{D}_\kappa^{-1}\boldsymbol{S}_5'(\underline{x})\boldsymbol{S}_5(\underline{x})\boldsymbol{D}_\kappa^{-1}, \\ \boldsymbol{S}_5(\underline{x}) &= \sum_{j=1}^p \underline{w}_j(\underline{x})\underline{e}_j'. \end{aligned} \quad (14)$$

[1] For brevity, we have omitted several preconditioners from [14], but have retained the matrix numberings for consistency.

Implementing $\boldsymbol{M}_5$ appears to require $2(p+1)$ 2D FFT's per iteration. Although this is considerably less computation than required for $\boldsymbol{M}_3$, it remains impractical.

To further reduce the number of FFT's, we propose to use interpolation. We choose a small number $m << p$ of values $\{\tilde{\eta}_k\}_{k=1}^m$ that cover (most of) the range of the values of the η_j's, and precompute the 2D DFT coefficients for those values: $\boldsymbol{\Omega}_k \triangleq \boldsymbol{\Omega}(\tilde{\eta}_k), \quad k = 1, \ldots, m$. We then apply interpolation to approximate the 2D DFT's corresponding to the required values η_j:

$$\boldsymbol{\Omega}^{-1/2}(\eta_j(\underline{x})) \approx \sum_{k=1}^{m} \lambda_k(\eta_j(\underline{x}))\boldsymbol{\Omega}_k^{-1/2}, \tag{15}$$

where $\lambda_k \in [0,1]$ are the (linear) interpolation factors with $\sum_{k=1}^m \lambda_k = 1$. Incorporating (15) into (14) yields

$$\begin{aligned} \boldsymbol{S}_5(\underline{x}) &= \sum_{j=1}^{p} \boldsymbol{Q}'\boldsymbol{\Omega}^{-1/2}(\eta_j)\boldsymbol{Q}\underline{e}_j\underline{e}_j' \\ &\approx \sum_{j=1}^{p} \boldsymbol{Q}'\left(\sum_{k=1}^{m} \lambda_k(\eta_j(\underline{x}))\boldsymbol{\Omega}_k^{-1/2}\right)\boldsymbol{Q}\underline{e}_j\underline{e}_j' \\ &= \sum_{k=1}^{m} \boldsymbol{Q}'\boldsymbol{\Omega}_k^{-1/2}\boldsymbol{Q}\left(\sum_{j=1}^{p} \lambda_k(\eta_j(\underline{x}))\underline{e}_j\underline{e}_j'\right) \\ &= \sum_{k=1}^{m} \boldsymbol{Q}'\boldsymbol{\Omega}_k^{-1/2}\boldsymbol{Q}\boldsymbol{D}_{\lambda_k}(\underline{x}) \;\triangleq\; \boldsymbol{Q}'\boldsymbol{S}_9(\underline{x}), \end{aligned}$$

where $\boldsymbol{D}_{\lambda_k}(\underline{x}) = \mathcal{D}[\lambda_k(\eta_j(\underline{x}))]$. Since $\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I}$, this approximation suggests the following preconditioner:

$$\boldsymbol{M}_9(\underline{x}) = \boldsymbol{D}_\kappa^{-1}\boldsymbol{S}_9'(\underline{x})\boldsymbol{S}_9(\underline{x})\boldsymbol{D}_\kappa^{-1} \tag{16}$$

which requires only $2m$ 2D FFT's per iteration. Computation of $2m$ 2D FFT's is $O(mp\log_2 p)$, which for small m is moderate compared with the $O(p^2)$ computation required to compute the gradient. Note that while the preconditioner $\boldsymbol{M}_9$ combines aspects of both the diagonal and Fourier preconditioning methods, the particular diagonal matrices used here are very different from the conventional diagonal preconditioning matrix (5).

The standard Fourier preconditioner is the special of $\boldsymbol{M}_9$ where $\eta_j = \eta_0$ and $\kappa_j = \kappa \; \forall j$. In this case, $m = 1$ suffices, and clearly we will choose $\tilde{\eta}_1 = \eta_0$. Then $\boldsymbol{M}_9$ simplifies to:

$$\boldsymbol{M}_{\mathrm{F}} = \frac{1}{\kappa^2}\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}(\eta_0)\boldsymbol{Q}, \tag{17}$$

which is essentially the standard Fourier preconditioner for shift-invariant problems, e.g. [7].

Implementing $\boldsymbol{M}_9$ using FFTs required about 4-13% more computation time than the CG algorithm with no preconditioning [14, 15], yet the improvement in convergence rate was much greater, as the results below show.

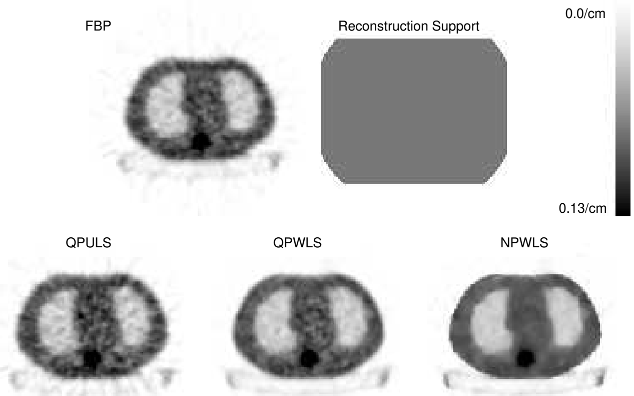

Fig. 1. Grayscale images of the reconstructed FBP image, the reconstruction support (the set of pixels estimated), and the images reconstructed using each of the three objective functions: QPULS, QPWLS, NPWLS.

IV. Numerical Results

We have applied the CG method with various preconditioners to several simulated and real PET scans. We synopsize only representative results from [14, 15].

We acquired a low-count PET transmission scan of a thorax phantom on an CTI ECAT 921 EXACT PET scanner (920653 prompt coincidences, which is quite noisy). Random coincidences were collected separately. We then computed a 2nd-order Taylor expansion of the transmission log-likelihood about the reprojection of an initial reconstruction, to produce a quadratic objective function of the form (2). To compare convergence rates, we examined the normalized l_2 distance between the nth iterate $\underline{x}^n$ and the limiting value $\underline{x}^\infty$: $\|\underline{x}^n - \underline{x}^\infty\|/\|\underline{x}^\infty\|$. To compute $\underline{x}^\infty$, we ran many iterations of a grouped-coordinate ascent algorithm similar to that of [16] using Matlab with double precision. We computed $\underline{x}^n$ in C using single precision for the CG method, applied to a *nonquadratically*-penalized weighted least squares (NPWLS) objective function, with the edge-preserving choice

$$\psi(t) = \delta^2 \left[\, |t/\delta| - \log(1 + |t/\delta|) \,\right], \tag{18}$$

with $\delta = 0.004$ cm. As shown in Fig. 2, due to the shift-variance, the standard Fourier preconditioner is ineffective, but the proposed preconditioner $\boldsymbol{M}_9$ leads to significant improvement in convergence rate, with only a 4-13% increase in computation time per iteration. A more dense matrix $\boldsymbol{G}$, would further enhance the relative benefits of the proposed preconditioners.

We also compared the CG algorithm using the proposed preconditioners to the coordinate-descent algorithms of [1, 9]. As in [1], we found that coordinate-descent often converged faster than *unpreconditioned* CG

[15]. (See Fig. 3.) However, we consistently found that the CG algorithm *with the proposed preconditioners* converged significantly faster than coordinate descent [15]. A nonnegativity constraint could affect these results.

A longer and more complete version of this paper will be available at the author's web site.

`http://www.eecs.umich.edu/~fessler/`

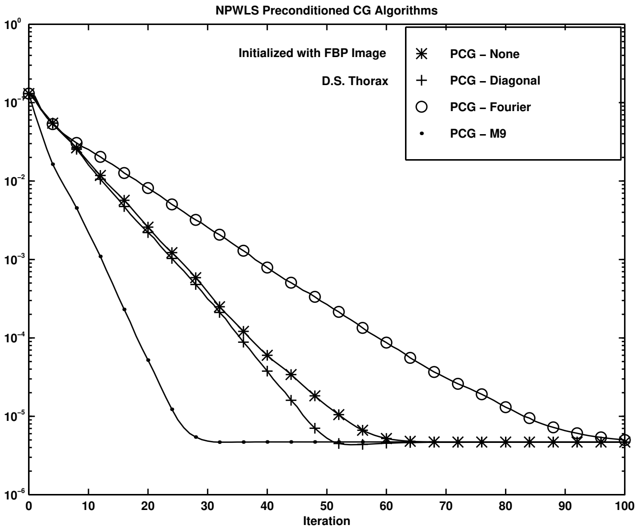

Fig. 2. Normalized l_2 distance to solution $\underline{x}^{\infty}$ versus iteration n for the *nonquadratically*-penalized weighted least squares (NPWLS) objective function. Shown is CG with no preconditioning, with the diagonal preconditioner (5), with the Fourier preconditioner (17), and with the proposed $\boldsymbol{M}_9$ preconditioner (16). The proposed preconditioner provides significant acceleration in convergence rate over conventional preconditioners.

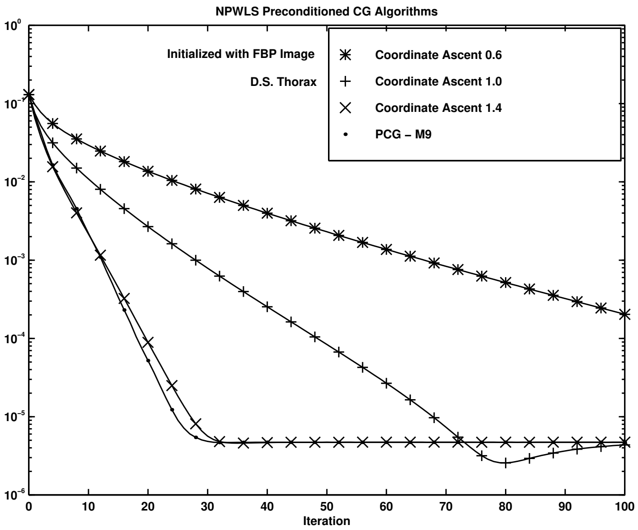

Fig. 3. As in Fig. 2, but comparing the proposed $\boldsymbol{M}_9$ preconditioner to the coordinate descent algorithm described in [9] with three values of its relaxation parameter ω. The preconditioned CG algorithm converges faster than the coordinate descent algorithm, and also requires less computation time per iteration.

References

[1] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Tr. Sig. Proc.*, vol. 41, no. 2, pp. 534–548, February 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, December 1994.

[3] 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*, volume 2, pp. 441–4, 1995.

[4] 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.

[5] J. A. Fessler and E. P. Ficaro, "Fully 3D PET image reconstruction using a Fourier preconditioned conjugate-gradient algorithm," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1599–1602, 1996.

[6] R. H. Chan and M. K. Ng, "Conjugate gradient methods for Toeplitz systems," *SIAM Review*, vol. 38, no. 3, pp. 427–82, September 1996.

[7] 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, March 1993.

[8] J. G. Nagy, R. J. Plemmons, and T. C. Torgersen, "Iterative image restoration using approximate inverse preconditioning," *IEEE Tr. Im. Proc.*, vol. 5, no. 7, pp. 1151–62, July 1996.

[9] 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.

[10] 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, September 1996.

[11] E. U. Mumcuoğlu, R. M. Leahy, and S. R. Cherry, "Bayesian reconstruction of PET images: methodology and performance analysis," *Phys. Med. Biol.*, vol. 41, pp. 1777–1807, 1996.

[12] J. J. Moré and G. Toraldo, "On the solution of large quadratic programming problems with bound constraints," *SIAM J. Optim.*, vol. 1, no. 1, pp. 93–113, February 1991.

[13] M. Bierlaire, P. L. Toint, and D. Tuyttens, "On iterative algorithms for linear least squares problems with bound constraints," *Linear Algebra and its Applications*, vol. 143, pp. 111–43, 1991.

[14] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant image reconstruction," *IEEE Tr. Im. Proc.*, 1997. Submitted.

[15] J. A. Fessler, "Conjugate-gradient preconditioning methods: numerical results," Technical Report 303, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, January 1997. Available on WWW from http://www.eecs.umich.edu/∼fessler.

[16] 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.*, vol. 16, no. 2, pp. 166–75, April 1997.