A Preliminary Investigation of 3D Preconditioned Conjugate Gradient Reconstruction for Cone-Beam CT

Lin Fu[a], Bruno De Man[a], Kai Zeng[a], Thomas M. Benson[a],
Zhou Yu[b], Guangzhi Cao[b], Jean-Baptiste Thibault[b]
[a]GE Global Research, One Research Circle, Niskayuna, NY, USA 12309;
[b]GE Healthcare Technologies, 3000 N Grandview Blvd, Waukesha, WI, USA 53188

ABSTRACT

Model-based iterative reconstruction (MBIR) methods based on maximum *a posteriori* (MAP) estimation have been recently introduced to multi-slice CT scanners. The model-based approach has shown promising image quality improvement with reduced radiation dose compared to conventional FBP methods, but the associated high computation cost limits its widespread use in clinical environments. Among the various choices of numerical algorithms to optimize the MAP cost function, simultaneous update methods such as the conjugate gradient (CG) method have a relatively high level of parallelism to take full advantage of a new generation of many-core computing hardware. With proper preconditioning techniques, fast convergence speeds of CG algorithms have been demonstrated in 3D emission and 2D transmission reconstruction. However, 3D transmission reconstruction using preconditioned conjugate gradient (PCG) has not been reported. Additional challenges in applying PCG in 3D CT reconstruction include the large size of clinical CT data, shift-variant and incomplete sampling, and complex regularization schemes to meet the diagnostic standard of image quality. In this paper, we present a ramp-filter based PCG algorithm for 3D CT MBIR. Convergence speeds of algorithms with and without using the preconditioner are compared.

Keywords: 3D model-based iterative CT reconstruction, preconditioned conjugate gradient

1. INTRODUCTION

Iterative CT reconstruction algorithms based on the maximum *a posteriori* (MAP) estimation were developed over a decade ago [1–4] but have only very recently been introduced commercially on multi-slice clinical CT scanners [5,6]. The model-based approach is based on incorporation of accurate modeling of system optics, noise statistics in the transmission data, and *a priori* knowledge of the object being imaged. However, the associated long computation time is a major challenge to their widespread use in clinical environments.

Many types of algorithms have been explored for minimizing the MAP cost function in CT reconstruction. Sequential algorithms such as iterative coordinate descent (ICD) [4,5,7] has fast convergence rate if given a good initial estimate, but requires column access to the system matrix and has relatively large computation cost per iteration. Simultaneous algorithms, such as gradient-based methods with various surrogate functions [2,3], perform forward- and back-projection operations over the whole image volume, and thus have a higher level of parallelism to take full advantage of multi- and many-core computing hardware. In between ICD and gradient-based updates are grouped-coordinate descent algorithms that use all the data to update a subset of the pixels each (sub)iteration [8–10]. These algorithms offer a compromise that combines efficient iterations with faster convergence than gradient-based methods, albeit with somewhat more complicated implementation.

In his paper we focus on simultaneous algorithms based on conjugate gradient (CG). Similar to many standard gradient-based algorithms, CG converges very slowly for tomographic inversion problems due to their ill-conditioned nature. However, significant enhancement in convergence speeds has been demonstrated when combining CG with a positive definite preconditioning matrix tailored to approximate the inverse of the Hessian matrix of the MAP cost function [11–14]. Various forms of preconditioner have been studied in the context of iterative tomographic reconstruction. Diagonal preconditioners [12–15] can be effective but suboptimal because they do not account for the off-diagonal structure of the Hessian. Fourier preconditioners [16,17] can address the off-diagonal structure (such as the $1/r$ correlation between voxels) for shift-invariant systems, but they are not effective to address shift-variance caused by statistical noise

Supported in part by NIH grant 1-R01-HL-098686.

Medical Imaging 2012: Physics of Medical Imaging, edited by Norbert J. Pelc, Robert M. Nishikawa, Bruce R. Whiting,
Proc. of SPIE Vol. 8313, 83133O · © 2012 SPIE · CCC code: 1605-7422/12/$18 · doi: 10.1117/12.910903

modeling. More effective preconditioning approaches use a combination of diagonal and circulant matrices [11,18], multiple FFTs [11], or operator splitting methods [19,20] to address additional shift-variance.

Although preconditioned conjugate gradient (PCG) approaches have demonstrated remarkable convergence speeds in the context of 3D emission reconstruction [13] and 2D parallel-beam transmission reconstruction [11], their effectiveness in 3D cone-beam settings as in modern multi-slice CT scanners remains to be evaluated. In this study, we design a PCG algorithm for 3D cone-beam CT reconstruction with statistical noise modeling and edge-preserving regularization. A quadratic surrogate function is introduced to avoid the inconveniences of the 'line search' procedure to determine the step size parameter along the conjugate gradient direction. A combined diagonal/circulant preconditioner is designed to speed up reconstruction. The convergence speeds of algorithms with and without using the preconditioner will be compared.

2. METHODS

2.1 Cost function

One approach to statistical image reconstruction in x-ray CT uses a regularized weighted least-squares cost function of the form

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \Phi(\boldsymbol{x}) \triangleq \left\{\frac{1}{2}(\boldsymbol{y} - \mathbf{A}\boldsymbol{x})^T \boldsymbol{W}(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}) + \beta U(\boldsymbol{x})\right\},$$

where $\boldsymbol{x} = \{x_1, \dots, x_N\}$ denotes the vector of unknown 3D image space, $\boldsymbol{y} = \{y_1, \dots, y_M\}$ is the vector of sinogram measurements, $\boldsymbol{W} = \text{diag}\{w_1, \dots, w_M\}$ denotes the statistical weighting, $\boldsymbol{A}$ is a $M \times N$ system matrix, $U(\boldsymbol{x})$ is a regularizer, and β is a scalar that controls the regularization strength [4].

2.2 Non-quadratic edge-preserving prior

We employ a convex q-generalized Gaussian Markov random field (q-GGMRF) [5] to model the prior distribution of the image. The prior energy function takes the form

$$U(\boldsymbol{x}) = \sum_{k>j} \omega_{jk} \rho(x_j - x_k),$$

where the ω_{jk} are directional weighting coefficients, which we choose as the inverse of the distance between voxels j and k, and $\rho(\cdot)$ is a potential function given by

$$\rho(\Delta) = \frac{|\Delta|^p}{1 + \left|\frac{\Delta}{c}\right|^{p-q}},$$

where $1 \le q \le p \le 2$ to ensure convexity. This prior features edge-preserving properties with a threshold c that determines the transition from low to high contrast regions so that the edge-preserving behavior can be tuned to specific use cases.

2.3 Preconditioner

We use a preconditioner which we call a "ramp-based" one because an analytical ramp filter is used to approximate the inverse of the geometric response of $\mathbf{A}^T\mathbf{A}$. Ideally, a preconditioner should invert of the Hessian of the MAP objective function, which takes the form

$$\nabla^2 \Phi(\boldsymbol{x}) = \mathbf{A}^T \mathbf{W} \mathbf{A} + \beta \nabla^2 U(\boldsymbol{x}).$$

Since we cannot directly compute this matrix inverse, simplifications have to be introduced. With approximations, statistical weighting in sinogram-domain may be translated into image-domain [11]

$$\mathbf{A}^T\mathbf{W}\mathbf{A} \approx \text{diag}(\boldsymbol{\kappa})\ \mathbf{A}^T\mathbf{A}\ \text{diag}(\boldsymbol{\kappa}), \text{ where } \kappa_j = \sqrt{\frac{\sum_i a_{ij}^2 W_i}{\sum_i a_{ij}^2}}.$$

In CT scanners with relatively small cone angle, $\mathbf{A}^T\mathbf{A}$ is well modeled by a shift-invariant filter

$$\mathbf{A}^T\mathbf{A} \approx \mathbf{I}_{N_z} \otimes \left[\boldsymbol{Q}^T \boldsymbol{\Lambda} \mathbf{Q}\right],$$

where $\boldsymbol{Q}$ is the orthonormal version of the 2-D DFT operator, $\boldsymbol{\Lambda}$ is diagonal matrix corresponding to the $1/r$ filter in 2D, "⊗" denotes the Kronecker product operator that replicates the 2D filter by the number of slices in the Z direction.

As an initial evaluation, we assume that the regularization parameter β is relatively small so that the Hessian is dominated by the data fit term. Ignoring the regularization term, we obtain the following proposed preconditioner for 3D reconstruction

$$\mathbf{M} = \text{diag}(\boldsymbol{\kappa})^{-1}\left[\mathbf{I}_{N_z} \otimes (\mathbf{Q}^T \boldsymbol{\Lambda}^{-1} \mathbf{Q})\right]\text{diag}(\boldsymbol{\kappa})^{-1}.$$

The above preconditioner takes a similar form as the combined diagonal/circulant matrix proposed by Fessler and Booth [11], who constructed the circulant kernel by taking a column of $\mathbf{A}^T\mathbf{A} + \beta\boldsymbol{R}$, with $\boldsymbol{R}$ being the Hessian of a quadratic prior term, and making it symmetric and block-circulant by averaging flipped and shifted versions of itself. In this work the circulant kernel is constructed directly as an analytical ramp filter in the Fourier space. The analytical design is more convenient and automatically guarantees positive-definiteness. To balance between accuracy and computational overhead of the preconditioner, we ignore the shift-variance caused by edge-preserving regularization and the cone-beam effects. These approximations will only affect convergence speeds but will not change the final solution defined by cost function. More advanced methods based on multiple FFTs [11] or operator splitting methods [19,20] may be used to address more challenging shift-variant factors.

2.4 Conjugate gradient

We use Polka-Ribiere form of PCG algorithm to compute an image update

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} + \alpha^{(n)}\boldsymbol{d}^{(n)}$$

$$\boldsymbol{d}^{(n)} = \boldsymbol{p}^{(n)} + \gamma^{(n)}\boldsymbol{d}^{(n-1)}$$

$$\gamma^{(n)} = \frac{\boldsymbol{p}^{(n)^T}(\boldsymbol{g}^{(n)} - \boldsymbol{g}^{(n-1)})}{\boldsymbol{p}^{(n-1)^T}\boldsymbol{g}^{(n-1)}}$$

$$\boldsymbol{p}^{(n)} = \mathbf{M}\boldsymbol{g}^{(n)}$$

where $\boldsymbol{g}^{(n)}$ is the gradient vector of the objective function evaluated at $\boldsymbol{x} = \boldsymbol{x}^{(n)}$, $\mathbf{M}$ is the preconditioning matrix, and $\alpha^{(n)}$ is a step size parameter. For non-quadratic functions, it is necessary to check if $\boldsymbol{d}^{(n)}$ is a descent direction at each iteration, i.e., $\boldsymbol{d}^{(n)^T}\boldsymbol{g}^{(n)} > 0$. If the condition is not satisfied, we reset $\gamma^{(n)}$ to zero and $\boldsymbol{d}^{(n)} = \boldsymbol{p}^{(n)}$ will used as the step direction.

For non-quadratic cost functions, standard CG algorithms usually employs a "line search" procedure to determine the step size $\alpha^{(n)}$ along the current conjugate gradient direction [11–13]. While the "line search" procedure generally does not require extra forward/back projection operations, it does require evaluations of the first and second gradients of the prior term, which can be of high computation cost in 3D due to the large number of neighborhood pairs. To ease computation, we employ a quadratic surrogate functional substitution that Yu et al. proposed for the q-GGMRF prior [7] which allows us obtain a closed form solution for $\alpha^{(n)}$ without using "line search". A quadratic function $\rho_s(\cdot;\cdot)$ is constructed at an expansion point Δ_0 to majorize the original potential function $\rho(\cdot)$

$$\rho_s(\Delta;\Delta_0) \triangleq \frac{\rho'(\Delta_0)}{2\Delta_0}\Delta^2 \geq \rho(\Delta);$$

Then a surrogate prior function $U_s\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right)$ is formed by replacing each $\rho(\cdot)$ in $U(\boldsymbol{x})$ by $\rho_s(\cdot;\cdot)$

$$U_s\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right) \triangleq \sum_{k>j}\omega_{jk}\rho_s\left(x_j - x_k; x_j^{(n)} - x_k^{(n)}\right) \geq U(\boldsymbol{x}),$$
$$U_s\left(\boldsymbol{x}^{(n)};\mathbf{x}^{(n)}\right) = U\left(\boldsymbol{x}^{(n)}\right),$$

which further gives us the following quadratic surrogate substitution of the overall MAP cost function

$$\Phi_s\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right) \triangleq \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \beta U_s\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right).$$

Then $\alpha^{(n)}$ can be computed in closed form by minimizing the following 1D quadratic problem. Monotonic decrease of original MAP cost function is still guaranteed.

$$\begin{aligned}\alpha^{(n)} &= \arg\min_{\alpha} \Phi_s\left(\boldsymbol{x}^{(n)} + \alpha \boldsymbol{d}^{(n)}; \boldsymbol{x}^{(n)}\right) \\ &= -\frac{\boldsymbol{d}^{(n)^T}\boldsymbol{g}^{(n)}}{\boldsymbol{d}^{(n)^T}\left[\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A} + \beta\nabla^2 U_s(\boldsymbol{x}; \boldsymbol{x}^{(n)})|_{\boldsymbol{x}=\boldsymbol{x}^{(n)}}\right]\boldsymbol{d}^{(n)}} \\ &= -\frac{\boldsymbol{d}^{(n)^T}\boldsymbol{g}^{(n)}}{\boldsymbol{d}^{(n)^T}\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A}\boldsymbol{d}^{(n)} + \beta\sum_{j=1}^{N}\sum_{k>j}\omega_{jk}\rho_s''\left(x_j^{(n)} - x_k^{(n)}\right)\left(d_j^{(n)} - d_k^{(n)}\right)^2}\ .\end{aligned}$$

3. COMPUTER SIMULATION

We performed a numerical phantom simulation using CatSim [21]. The simulation includes a 64 row axial scan of an XCAT phantom [22] with 1000 views evenly distributed over one full rotation for a 1.0 second gantry period. We used cone-beam geometry with a source-to-iso distance of 540 mm, a source-to-detector distance of 960 mm and 64 x 888 detector elements of size 1.1mm x 1.0mm. The source model included a mono-energetic spectrum at 60keV and 100mA tube current; we did not include focal spot deflection or electronic noise. Iterative reconstruction was performed over a 512x512x64 voxel grid with a 50cm field of view and 0.625mm slice thickness. We compared the convergence rates of the reconstruction algorithms with and without the ramp-based preconditioner described above. We implemented both steepest descent (SD) and conjugate-gradient (CG) methods. All algorithms are initialized with standard FBP reconstructions.

Figure 1 compares reconstructions of center slices. Clearly, after 30 iterations, the ramp-CG reconstruction is already much closer to the reference solution after 200 CG iterations than the CG reconstruction is. Figure 2 compares reconstructions of end slices in the field of view. Again, the ramp-PCG reconstruction is much closer to the reference solution than the CG reconstruction. Some loss of resolution is observed in the edge slice due to the change of geometric response and statistical weighting. Modified regularization design has been proposed to achieve nearly uniform resolution [23] and is straightforward to be incorporated into the PCG framework described here. The modified regularization will also make the Hessian matrix less shift-variant and more amenable to preconditioning.

Figure 3 shows the image-domain l2 distance to the converged solution as a function of iteration number. The l2 distance for each reconstruction slice is computed separately to study the effect of cone-angle dependence of convergence speeds. The curves show that in all slices, the ramp-based preconditioner brings drastic improvement in convergence rate for both SD and CG algorithms. Comparing the curves of different slices, the acceleration caused by the preconditioner is the most significant near the center slices, which is expected due to the more shift-invariant system response in this regions. More advanced methods such as multiple FFTs [11] or operator splitting methods [19,20] may further improve the effectiveness of the preconditioner by addressing more shift-variant factors in the edge slices.

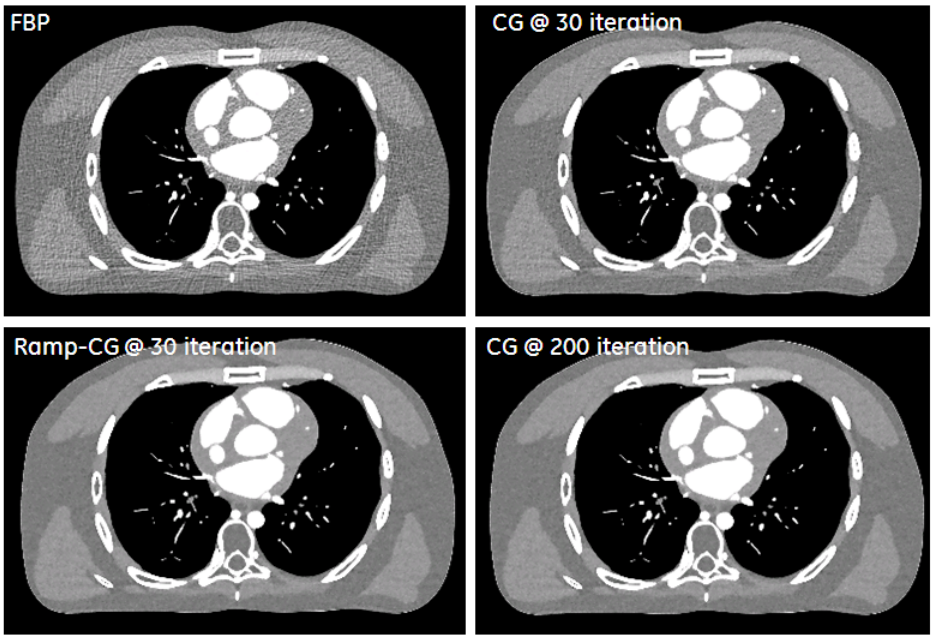

Figure 1. Center slice images computed by different algorithms: FBP reconstruction (top left), CG reconstruction at 30 iterations (top right), Ramp-PCG reconstruction at 30 iterations (bottom left), and CG reconstruction at 200 iterations (bottom right). Display window width = 400 HU.

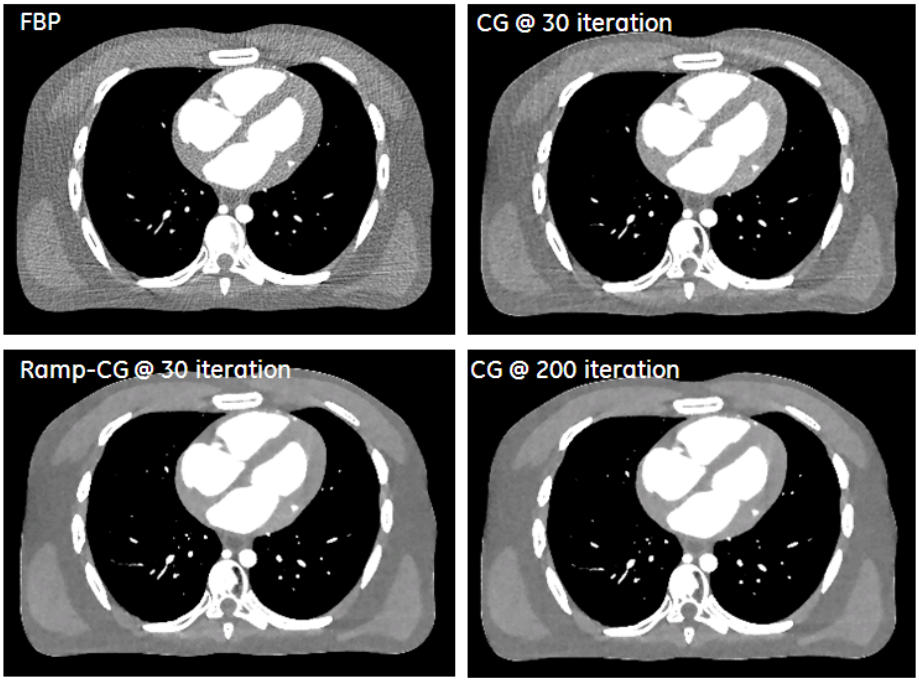

Figure 2. Edge slice images computed by different algorithms: FBP reconstruction (top left), CG reconstruction at 30 iterations (top right), Ramp-PCG reconstruction at 30 iterations (bottom left), and CG reconstruction at 200 iterations (bottom right). Display window width = 400 HU.

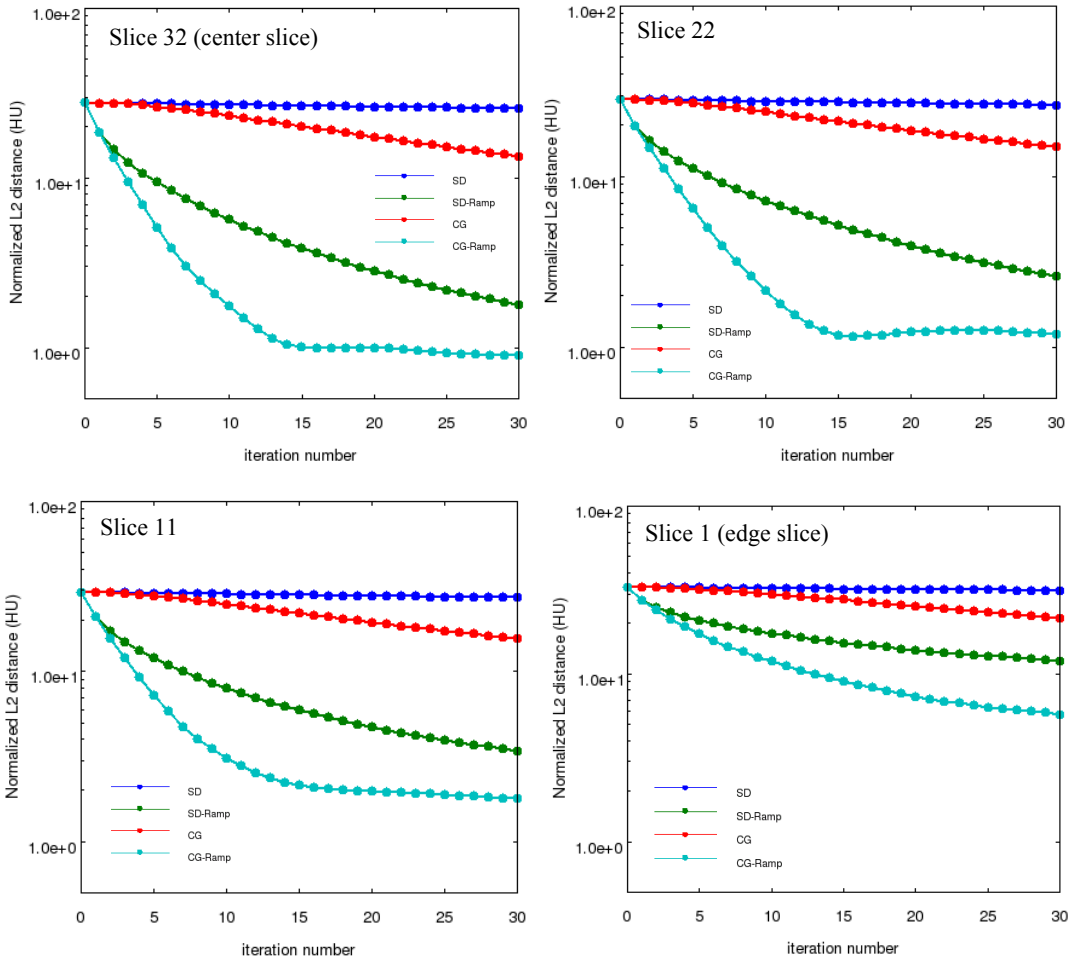

Figure 3. Normalized $l2$distance to the image computed by CG after 200 iterations versus iteration number. "-Ramp" denotes algorithms with preconditioning. All algorithms optimize the same cost function and start from the same FBP estimate. In all slices, the ramp-based methods bring significant improvements in convergence rate for both the SD and CG algorithms.

4. PHYSICAL PHANTOM DATA

We also tested different MBIR algorithms using a GE performance phantom. The phantom was scanned axially on a 64-slice GE HD750 scanner at 120 kV and 70mAs. All axial reconstructed images are of size 512 x 512 with each slice having thickness of 0.625 mm. The reconstruction field of view is 350 mm in diameter. Standard FBP images were used as initial estimation. Figure 4 shows that with only 10 iterations, the PCG reconstruction already produces images very similar to the reference image produced by 200 standard CG iterations, with much better noise reduction compared to the standard CG reconstruction at 10 iterations.

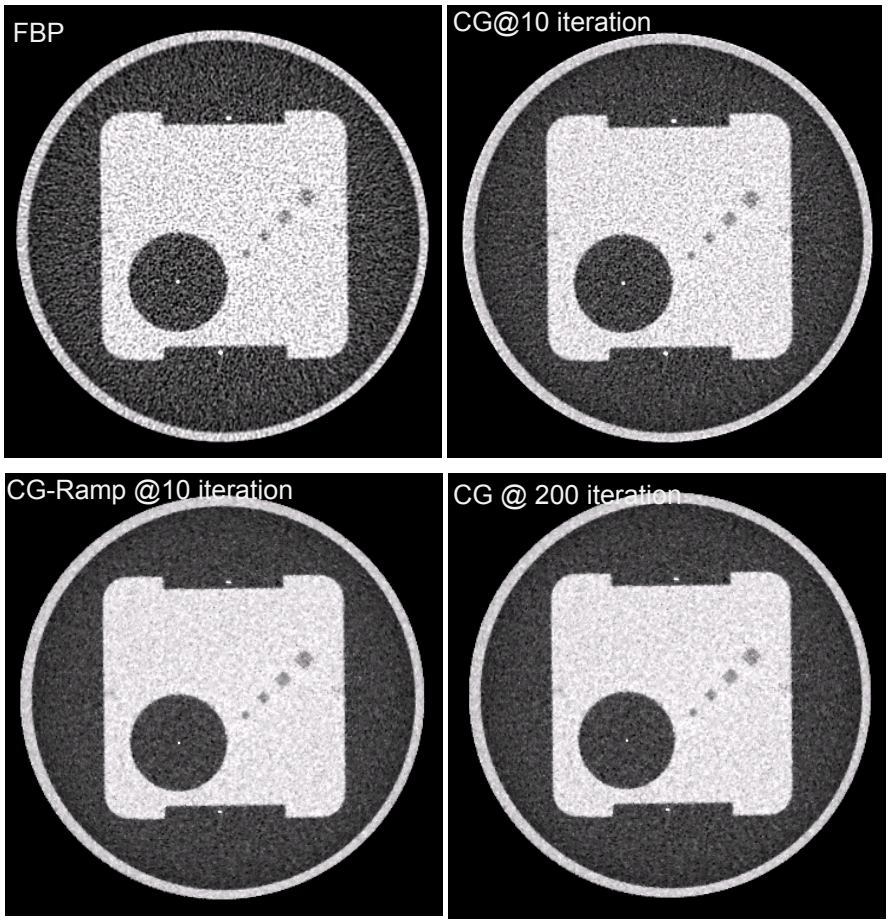

Figure 4. GE performance phantom images computed by different algorithms: FBP reconstruction (top left), CG reconstruction at 10 iterations (top right), PCG reconstruction at 10 iterations (bottom left), and CG reconstruction at 200 iterations (bottom right). Display window width = 200 HU.

5. CONCLUSION

We performed an initial evaluation of a ramp-based PCG algorithm for cone-beam CT MBIR. Modeling of non-uniform noise statistics and edge-preserving image prior were included in the cost function. Computer simulation and real phantom data consistently showed that the preconditioning approach achieves significant acceleration in convergence rate of SD or CG algorithms. The acceleration was the most dramatic in center slices where the cone-beam effect is relatively small. However, even at the edge slices, the preconditioner still brought significant speed up to both SD and CG algorithms. The computational cost of applying the ramp-based preconditioner is similar to that of a Fourier-domain filtering operation, which is much less compared to forward and back projection operations. As an initial evaluation, the preconditioner used in this study did not account for the shift-variance caused by object-dependent and edge-preserving regularization. The incorporation of statistical weighting is also based on empirical approximations. The performance of the preconditioner may be further improved by incorporation of more sophisticated space-variant design [11] or combining with operator splitting methods [19,20].

REFERENCES

[1] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, "Iterative reconstruction for helical CT: a simulation study," *Phys. Med. Biol.* **43**, 729–737 (1998).

[2] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Reduction of metal streak artifacts in X-ray computed tomography using a transmission maximum a posteriori algorithm," *IEEE Trans. Nucl. Sci.* **47**, 977–981 (2000).

[3] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. Imag. Proc.* **4**, 1430–1450 (1995).

[4] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Proc.* **41**, 534–548 (1993).

[5] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.* **34**, 4526–4544 (2007).

[6] A. Ziegler, T. Nielsen, and M. Grass, "Iterative reconstruction of a region of interest for transmission tomography," *Med. Phys.* **35**, 1317–1327 (2008).

[7] Z. Yu, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. Imag. Proc.* **20**, 161–175 (2011).

[8] B. De Man, S. Basu, J.-B. Thibault, J. Hsieh, J. A. Fessler, C. Bouman, and K. Sauer, "A study of different minimization approaches for iterative reconstruction in x-ray CT," in *proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, pp. 2708–2710 (2005).

[9] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange., "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. Med. Imag.* **16**, 166–175 (1997).

[10] T. Benson, B. De Man, L. Fu, and J.-B. Thibault, "Block-based iterative coordinate descent," in *proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, pp. 2856–2859 (2010).

[11] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Imag. Proc.* **8**, 688–699 (1999).

[12] 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 Trans. Med. Imag.* **13**, 687–701 (1994).

[13] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High resolution 3D Bayesian image reconstruction using the microPET small animal scanner," *Phys. Med. Biol.* **43**, 1001–1013 (1998).

[14] L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET.," *IEEE Trans. Med. Imag.* **12**, 200–214 (1993).

[15] D. S. Lalush and B. M. Tsui, "A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm," *Med. Phys.* **22**, 1273–1284 (1995).

[16] 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 Trans. Med. Imag.* **12**, 78–83 (1993).

[17] J. Nuyts, P. Suetens, and L. Mortelmans, "Acceleration of maximum likelihood reconstruction, using frequency amplification and attenuation compensation.," *IEEE Trans. Med. Imag.* **12**, 643–652 (1993).

[18] G. Chinn and S. C. Huang, "A general class of preconditioners for statistical iterative reconstruction of emission computed tomography.," *IEEE Trans. Med. Imag.* **16**, 1–10 (1997).

[19] S. Ramani and J. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction," *IEEE Trans. Med. Imag.* **to appear** (2012) [doi:10.1109/TMI.2011.2175233].

[20] G. McGaffin, S. Ramani, and A. J. Fessler, "Reduced memory augmented Lagrangian algorithm for 3D iterative x-ray CT image reconstruction," in *proc. SPIE*, p. to appear (2012).

[21] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, et al., "CatSim: a new computer assisted tomography simulation environment," in *Proc. SPIE*, p. 65102G (2007) [doi:10.1117/12.710713].

[22] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.* **37**, 4902 (2010) [doi:10.1118/1.3480985].

[23] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs.," *IEEE Trans. Imag. Proc.* **5**, 1346–1358 (1996).