Nested Loop Algorithm for Parallel Model Based Iterative Reconstruction

Zhou Yu, Lin Fu, Debashish Pal, Jean-Baptiste Thibault, Charles A. Bouman and Ken D. Sauer

***Abstract*— Model based iterative reconstruction (MBIR) algorithms have been used in clinical studies to allow significant dose reduction in CT scans while maintaining the diagnostic image quality. Simultaneous-update algorithms, which can take advantage of massively parallel computer architectures, are promising to significantly improve the speed of MBIR. To achieve this goal, we also need to improve the convergence speed of these algorithms. In this paper, we propose a fast converging simultaneous-update algorithm using a nested loop structure. Preliminary experimental results show that the proposed algorithm has faster convergence speed compared to algorithms such as conjugate gradient and preconditioned conjugate gradient methods.**

***Index Terms*— Computed tomography, iterative reconstruction, nested loop, preconditioner.**

I. Introduction

Recent applications of model based iterative reconstruction (MBIR) algorithms to medical CT have demonstrated significant improvement in image quality by increasing resolution as well as reducing noise and artifacts [1], [2]. Clinical studies also show that MBIR algorithms can be used as a tool to allow significant dose reduction in CT scans while maintaining diagnostic image quality [3]. With ever advancing computing technologies, massively parallel architectures have emerged, such as the newest multi-core CPUs and GPUs. These new hardware technologies bring promise to significantly speed up the MBIR algorithms [4]. Taking advantage of these new technologies requires developing algorithms that are highly parallel and yet have fast convergence properties. Simultaneous-update algorithms, such as variations of expectation maximization (EM) [5], conjugate gradients (CG) [6], and ordered subsets (OS) [7], are attractive since they are easier to map on to highly parallel computer architectures to reduce per iteration computational cost. However, compared to sequential-update algorithms such as iterative coordinate descent (ICD) [8], [9], simultaneous-update algorithms tend to require many more iterations to converge. Therefore, it is critical to speed up the convergence of simultaneous-update algorithms.

In this paper, we propose a nested loop framework to accelerate the convergence of simultaneous-update algorithms. Our method is composed of inner and outer loop iterations. In each outer loop iteration, we create a local approximation to the cost function. The approximate problem is then solved by inner loop iterations with relatively low computational cost. The inner loop solution is used to compute an update direction for the outer loop. The outer loop then computes an optimal step size so that it guarantees the cost function will decrease monotonically. Similar nested loop algorithms have been explored in PET reconstruction problems [10]–[12]. In the CT reconstruction problem, we propose to construct the inner loop problem using an image space approximation to the Hessian matrix of the original cost function.

The project described was partly supported by Grant No. 1-R01-HL-098686 from NIH and its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH.

Zhou Yu, Debashish Pal and Jean-Baptiste Thibault are with GE Healthcare Technologies, 3000 N Grandview Bvd, W-1180, Waukesha, WI 53188.

Lin Fu is with GE Global Research Center

Charles Bouman is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-0501.

Ken Sauer is with the Department of Electrical Engineering, 275 Fitzpatrick, University of Notre Dame, Notre Dame, IN 46556-5637.

II. Method

A. Objective Function

MBIR algorithms typically work by first forming an objective function which incorporates an accurate system model [13], statistical noise model [1] and prior model [14]. The image is then reconstructed by computing an estimate which minimizes the resulting objective function.

Let x denote the image and y denote the measurement data. We consider both x and y as random vectors, and our goal is to reconstruct the image by computing the maximum *a posteriori* (MAP) estimate given by

$$x^* = \arg\min f(x) \tag{1}$$

$$f(x) = \{\frac{1}{2}J(x,y) + \Phi(x)\} \tag{2}$$

where $J(x,y)$ is the log likelihood term that penalizes the inconsistency between the image and the measurement, $\Phi(x)$ is the negative log of the prior distribution that penalizes the noises in the image. One example of $J(x,y)$ is in quadratic form

$$J(x,y) = ||Ax - y||^2_W \tag{3}$$

where A is the system matrix, W is a diagonal weighting matrix. The i^{th} diagonal entry of the matrix W, denoted by w_i, is typically chosen to be approximately inversely proportional to the estimate of the variance in the measurement y_i [1], [9]. We will consider the data mismatch term in (3) to illustrate the algorithm framework in this paper. However, the proposed algorithm can also be applied to other forms of data mismatch terms, such as the Poisson log likelihood function, as long as $f(x)$ remains strictly convex.

B. Nested Loop Algorithm

The idea in this paper is to create a sequence of sub-problems that optimize simpler approximate cost functions, while still guaranteeing that the solution of the sub-problems will converge to the solution of the original cost function. To achieve this, we propose a nested loop framework. Let n be

the outer loop iteration index, and $x^{(n)}$ be the image estimate after the n^{th} iteration. In each outer loop iteration, we first create a local approximate cost function, $h^{(n)}(x)$, which must satisfy,

$$\nabla h^{(n)}(x^{(n-1)}) = \nabla f^{(n)}(x^{(n-1)}) \quad (4)$$

We then minimize $h^{(n)}$ using inner loop iterations. If we update the image directly using the inner loop solution, this does not necessarily guarantee convergence. Instead, we use the solution of the sub-problem to compute an update direction, and then solve a 1D optimization problem to determine the update step size. Since the cost function is minimized along the search direction, it is guaranteed to decrease monotonically with every outer loop iteration.

In the following, we propose a method to apply the nested loop framework to CT iterative reconstruction problem. First, we need to derive the approximate cost function used in each outer loop. We can rewrite the cost function in (2) and (3) as

$$f(x) = ||x - x^{(n-1)}||^2_{A^tWA} + x^t\Theta^{(n-1)} + \Phi(x) + c^{(n-1)}, \quad (5)$$

where $\Theta^{(n-1)} = A^tW(Ax^{(n-1)} - y)$ and $c^{(n-1)}$ is a constant. In CT iterative reconstruction, the most expensive computation components in each iteration are typically the forward projection and the back projection, i.e. A and A^t. Therefore, we create the approximate cost function by replacing A^tWA in (5) with a simpler operator M, i.e.

$$h^{(n)}(x) = ||x - x^{(n-1)}||^2_M + x^t\Theta^{(n-1)} + \Phi(x) + c^{(n-1)} \quad (6)$$

Written in this form, it is easy to verify that the condition in (4) holds for the proposed $h^{(n)}(x)$. Notice that, the regularization term in $h^{(n)}(x)$ is also calculated exactly. The only approximation in $h^{(n)}(x)$ is in the second derivative of the cost function by replacing A^tWA with M.

Designing M appropriately requires balancing between two objectives. First, M needs to be a close approximation of A^tWA. Second, M must be easy to pre-compute so the matrix vector multiplication can be computed at low cost. In this paper, we use the approximation to A^tWA operator proposed by Fessler and Booth in [15], that is,

$$A^tWA \approx DKD, \quad (7)$$

where D is a diagonal matrix, with the j^{th} diagonal element to be $d_i = \sqrt{\frac{\sum_j a_{ij}^2 w_j}{\sum_j a_{ij}^2}}$, and K is a circulant matrix approximation to the A^tA operator. Notice that the proposed operator is a pure image space operator. It is very easy to compute since it only requires image scaling and filtering, which can be efficiently implemented with fast Fourier transforms (FFT).

Second, we solve the minimization problem of $h^{(n)}$ iteratively using inner loops. The inner loop problem described in (6) is similar to an image space denoising problem. One can solve this problem with a simple gradient based method, such as gradient descent. Here, we can also accelerate the inner loop convergence using an image space preconditioner, such as Fourier based preconditioners proposed in [15] and [16].

Finally, we update the image based on the inner loop solution. Let $\tilde{x}^{(n)}$ denote the solution of the inner loop. We compute the update direction using $u^{(n)} = \tilde{x}^{(n)} - x^{(n-1)}$, and then compute the step size $\beta^{(n)}$. The step size is computed to minimize the cost along the update direction, that is,

$$\beta^{(n)} = \arg\min_{\beta} f\left(x^{(n-1)} + \beta u^{(n)}\right) \quad (8)$$

This way, we can guarantee $f(x^{(n)}) < f(x^{(n-1)})$. We can compute an approximate solution to (8) using a closed form formula similar to the one used in [15]. Since the step size formula is derived using a surrogate function, the monotonicity of sequence $\{f(x^{(n)})\}$ is still guaranteed.

Fig. 1 shows the pseudo code of the proposed algorithm. In each outer loop, we first formulate an approximate cost function, $h^{(n)}(x)$, which is then optimized by the inner loop iterations in line 5. By eliminating A and A^t operation in $h^{(n)}$, each inner loop has very low computational cost. In line 6, the result of the inner loop is used to compute the update direction, followed by the step size calculation in line 7.

The outer loop algorithm is similar to a gradient descent algorithm except the search direction is computed from the inner loop solution rather than the gradient direction. We can easily generalize this algorithm by using other gradient based method in the outer loop such as conjugate gradient, etc.

1: $x \leftarrow$ FBP reconstruction
2: $\Theta \leftarrow A'W(Ax - y)$
3: **repeat**
4: $\quad x_{prev} \leftarrow x$
5: $\quad x \leftarrow \arg\min_v ||v - x_{prev}||^2_M + v^t\Theta + \Phi(v)$
6: $\quad u \leftarrow x - x_{prev}$
7: $\quad \beta \leftarrow \arg\min_\beta f\left(x_{prev} + \beta u\right)$
8: $\quad x \leftarrow x_{prev} + \beta u$
9: $\quad \Theta \leftarrow \Theta + \beta A'WAu$
10: **until** x is converged

Fig. 1. Pseudo code of one example of nested loop algorithm. In each outer loop, we first formulate an approximate cost function. Line 5 is solved iteratively using the inner loop iterations. The result of the inner loop is used to compute an update direction in line 6. Finally, the line search step in line 7 guarantees the monotonicity of the cost function.

III. Experimental Results

In this section, we compare the convergence speed of Nested Loop algorithm to CG and two variations of Preconditioned CG algorithms. Here, the nested loop (NL) algorithm and PCG algorithms use the same approximation to A^tWA with image space operator DKD, in which K is a shift invariant filter with frequency response of the form $K(f) = \frac{1}{|f|+c}$, where $f \in [-0.5, 0.5]$. The difference among the three methods lies in the modeling of the regularization function. In the method labeled as PCG-P1, the Hessian of the regularization term is ignored. Therefore, its preconditioner is given by

$$P_1 = D^{-1}K^{-1}D^{-1}. \quad (9)$$

In the method labeled as PCG-P2, we model the Hessian of the regularization term as DRD, where R has the frequency response

$$R(f) = 1 - \cos(2\pi f). \quad (10)$$

In this case, the preconditioner is given by

$$P_2 = D^{-1}(K + \alpha R)^{-1} D^{-1}. \quad (11)$$

Finally, in the nested loop algorithm, we compute the regularization term exactly in the inner loop problem. The inner loop is solved using 10 iterations of PCG, where we use P_1 as the preconditioner.

Fig. 2 shows the Fourier transform of the preconditioner kernel used in the experiment. Notice that, by modeling the regularization term in the cost function, the high frequency gain in the preconditioner P_2 is suppressed.

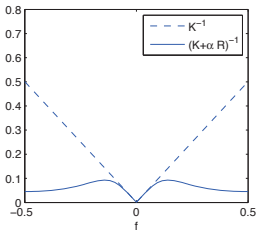


Fig. 2. This figure shows the Fourier transform of the preconditioner kernel used in the experiment, K^{-1} and $(K + \alpha R)^{-1}$, where $\alpha = 10$

The data we use to test the algorithm is a low dose axial scan of the GE performance phantom shown in Fig. 3. In this phantom, wires and resolution bars are used to measure the spatial resolution of the reconstruction, and a uniform region provides the noise measurements.

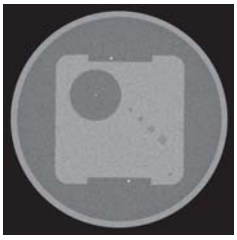


Fig. 3. This figure shows GE performance phantom used in the experiment. We use the wire in the phantom to measure in-plane resolution and the uniform region to measure the noise standard deviation

In Fig. 4, the results of various algorithms are compared against a reference image computed from a sufficiently converged reference NH-ICD [9] algorithm. In (a), we compute the root mean squared difference (RMSD) in the ROI volume between the current image and the reference image. Fig. 4 (b) and (c) characterize the convergence speed of high frequency components for each algorithm. In (b), we use the wire in the image to measure the 50 percent MTF after each iteration, and then plot the MTF convergence curve. The horizontal dashed line shows the MTF achieved by reference NH-ICD algorithm sufficiently converged after 20 iterations. In (c), we measure the noise standard deviation in a uniform ROI after each iteration. The horizontal line shows the noise level in a the reference NH-ICD reconstruction. Comparing the convergence plots, we find a consistent trend: As expected, preconditioning significantly accelerates the convergence speed of CG. By including the regularization term in the preconditioner design (P_2), we can further speed up the convergence. Finally, the NL method has the fastest convergence speed, which can be attributed to its ability to compute the regularization function exactly in the inner loops.

As shown in (c), in the uniform area, the PCG-P2 algorithm has a similar convergence speed to the NL algorithm. However, in (b), the NL method appears to be much faster than the PCG P_2 method. This is probably because, in the PCG-P2 method, we choose the parameter α to match the regularization strength in the uniform area. Since we use an edge-preserving regularization function, the same parameter can be too strong in the area around the wire causing slower resolution recovery. On the other hand, since the NL algorithm computes regularization term exactly, it shows consistently fast convergence speed in both the uniform area and around the wire without the need to choose parameters to optimize the convergence behavior.

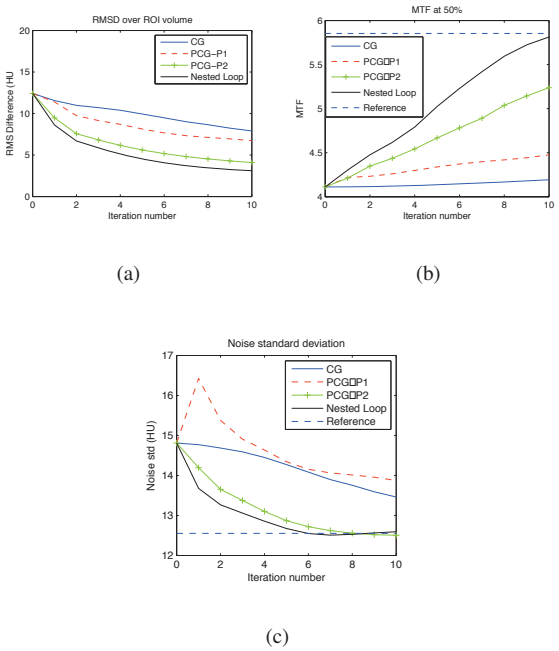
Fig. 5 shows the reconstructed image from a body scan data, in which (a) shows the denoised FBP image used as initial condition for the iterative reconstruction, (b) and (c) shows the image after 3 and 10 iterations of the nested loop algorithm, and (d) shows the fully converged image generated using 20 iterations of NH-ICD algorithm. The figure shows the image resolution improves very quickly using the proposed algorithm, and reaches the same solution as the NH-ICD algorithm.

Finally, let us comment on the computational cost of each algorithm. Typically, forward and back projection are the most computationally expensive components of each iteration. The cost of applying preconditioners (additional FFTs) is relatively low. Therefore, the per iteration cost of the PCG algorithm and the CG algorithm are very similar. The nested loop algorithm has slightly higher computational cost mainly due to multiple FFTs per inner loop iteration. However, in general, the inner loop cost may still be low compared to the projector costs. In practice, one can adjust the number of inner loop iterations to balance convergence speed and per iteration computation cost.

IV. Discussion and Future Work

The idea of using an approximate cost function is also used in the optimization transfer techniques. In these methods, one typically designs a surrogate function that satisfies equation (4), and upper bounds $f(x)$ for all x. However, to satisfy the upper bound condition, $h^{(n)}(x)$ tends to have very large curvature, and therefore, the update steps tend to be very small. In our approach, we replace the upper bound condition with a line search step, giving us more flexibility to explore different choices of $h^{(n)}(x)$.

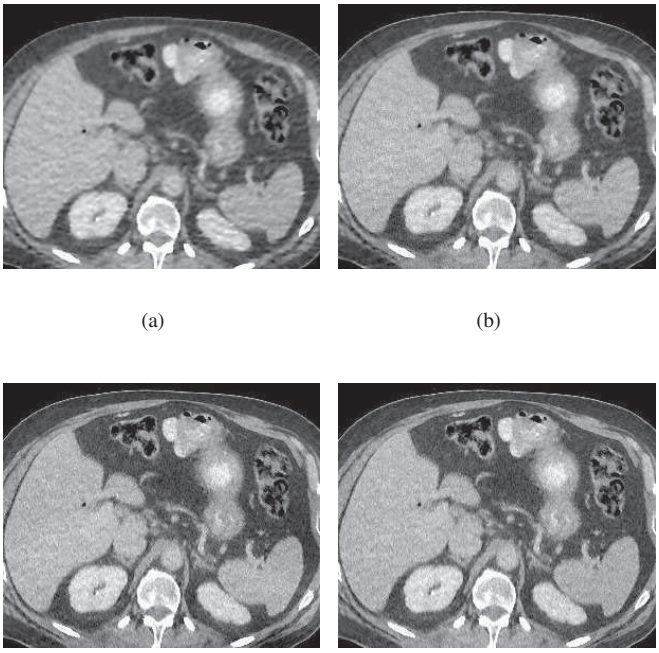
The idea to approximate $A^t W A$ has been well explored in preconditioner methods. In these methods, the idea is to design a matrix operator P directly to approximate the inverse of the Hessian $H^{-1}(x)$. However, in our approach, we only attempt to approximate the forward Hessian matrix. In general, approximating $H(x)$ instead of $H^{-1}(x)$ has two benefits. First, we can compute part of the cost function, in our case, the regularization term, exactly. In the typical preconditioner methods, the Hessian matrix of the regularization function is either completely ignored, or approximated. Using the same approximation to $A^t W A$, the proposed method outperforms



(a) (b)

(c)

Fig. 4. This figure shows convergence curves comparing the results of various algorithms. In (a), we show the root mean squared difference (RMSD) between the image and the fully converged reference image. In (b), we use the wire in the image to measure the 50 percent MTF after each iteration, and then plot the MTF convergence curve. The horizontal dashed line shows the MTF achieved by ICD algorithm after 20 iterations. In (c), we measure the noise standard deviation in a uniform ROI after each iteration. The horizontal line shows the noise level in a fully converged ICD reconstruction. All convergence curves show that the nested loop algorithm has the fastest convergence speed of all methods considered here.



(a) (b)

(c) (d)

Fig. 5. This figure shows the reconstructed image of a body scan, in which (a) shows the FBP initial image, (b) and (c) shows the image after 3 and 10 iterations, and (d) shows the fully converged image generated using NH-ICD. The figure shows that the image resolution improves very quickly using the proposed algorithm.

the preconditioner based methods mostly because it models the regularization function exactly. Second, it provides us the flexibility to use more sophisticated and more accurate models of the Hessian. In this paper, we construct the inner-loop algorithm using well-known approximations. In future work, we will further explore the flexibility of the algorithm to design a more sophisticated approximation of the $A^t W A$ operator.

REFERENCES

[1] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[2] A.Ziegler, Th.Köhler, and R.Proksa, "Noise and resolution in images reconstructed with FBP and OSC algorithms for CT," *Med. Phys.*, vol. 34, no. 2, pp. 585–598, 2007.

[3] G. Yadav, S. Kulkarni, Z. R. Colon, J. Thibault, and J. Hsieh, "Dose reduction and image quality benefits using model based iterative reconstruction technique for computed tomography," in *Fifty-second annual meeting of the American Association of Physicists in Medicine*, 2010.

[4] K. Mueller and R. Yagel, "Fast implementations of algebraic methods for three-dimensional reconstruction from cone-beam data," *IEEE Trans. on Medical Imaging*, vol. 18, no. 6, pp. 538–548, 1999.

[5] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-1, no. 2, pp. 113–122, October 1982.

[6] E. Ü. Mumcuoğlu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 687–701, December 1994.

[7] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 601–609, December 1994.

[8] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[9] Z. Yu, J.-B. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Fast model-based X-Ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. on Image Processing*, vol. 20, no. 1, pp. 161–175, 2011.

[10] L. Fu, "Residual correction algorithms for statistical image reconstructionin positron emission tomography," Ph.D. Thesis, University of California, Davis, Davis, Ca, Feb 2010.

[11] G. Wang and J. Qi, "Acceleration of the direct reconstruction of linear parametric images using nested algorithms," *Physics in Medicine and Biology*, vol. 55, no. 5, 2010.

[12] I. Hong, Z. Burbar, C. Michel, and R. Leahy, "Ultrafast preconditioned conjugate gradient algorithm for fully 3D PET reconstruction," in *IEEE Medical Imaging Conference*, Orlando, FL, Oct. 2010.

[13] B. DeMan and S. Basu, "Distance-driven projection and backprojection in three-dimensions," *Physics in Medicine and Biology*, vol. 49, pp. 2463–2475, 2004.

[14] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[15] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. on Image Processing*, vol. 8, no. 5, pp. 688–99, 1999.

[16] L. Fu, B. DeMan, K. Zeng, T. M. Benson, Z. Yu, G. Cao, and J.-B. Thibault, "A preliminary investigation of 3D preconditioned conjugate gradient reconstruction for cone-beam CT," in *SPIE Medical Imaging Conference*, 2012, pp. 8313–134.