Fast compressed sensing-based CBCT reconstruction using Barzilai-Borwein formulation for application to on-line IGRT

Justin C. Park
Center for Advanced Radiotherapy Technologies and Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California 92093 and Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, California 92093

Bongyong Song[a)]
Center for Advanced Radiotherapy Technologies and Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California 92093

Jin Sung Kim
Department of Radiation Oncology, Samsung Medical Center, Seoul, South Korea

Sung Ho Park
Department of Radiation Oncology, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, South Korea

Ho Kyung Kim
School of Mechanical Engineering, Pusan National University, Busan, South Korea

Zhaowei Liu
Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, California 92093

Tae Suk Suh
Department of Biomedical Engineering, Catholic University of Korea, Seoul, South Korea

William Y. Song[b)]
Center for Advanced Radiotherapy Technologies and Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California 92093

(Received 28 July 2011; revised 5 January 2012; accepted for publication 10 January 2012; published 15 February 2012)

Purpose: Compressed sensing theory has enabled an accurate, low-dose cone-beam computed tomography (CBCT) reconstruction using a minimal number of noisy projections. However, the reconstruction time remains a significant challenge for practical implementation in the clinic. In this work, we propose a novel gradient projection algorithm, based on the Gradient-Projection-Barzilai-Borwein formulation (GP-BB), that handles the total variation (TV)-norm regularization-based least squares problem for the CBCT reconstruction in a highly efficient manner, with speed acceptable for routine use in the clinic.

Methods: CBCT is reconstructed by minimizing an energy function consisting of a data fidelity term and a TV-norm regularization term. Both terms are simultaneously minimized by calculating the gradient projection of the energy function with the step size determined using an approximate Hessian calculation at each iteration, based on the Barzilai-Borwein formulation. To speed up the process, a multiresolution optimization is used. In addition, the entire algorithm was designed to run with a single graphics processing unit (GPU) card. To evaluate the performance, the Shepp-Logan numerical phantom, the CatPhan 600 physical phantom, and a clinically-treated head-and-neck patient were acquired from the TrueBeam^{TM} system (Varian Medical Systems, Palo Alto, CA). For each scan, in total, 364 projections were acquired in a 200° rotation. The imager has 1024×768 pixels with 0.388×0.388-mm resolution. This was down-sampled to 512×384 pixels with 0.776×0.776-mm resolution for reconstruction. Evenly spaced angles were subsampled and used for varying the number of projections for the image reconstruction. To assess the performance of our GP-BB algorithm, we have implemented and compared with three compressed sensing-type algorithms, the two of which are popular and published (forward–backward splitting techniques), and the other one with a basic line-search technique. In addition, the conventional Feldkamp-Davis-Kress (FDK) reconstruction of the clinical patient data is compared as well.

Results: In comparison with the other compressed sensing-type algorithms, our algorithm showed convergence in $\leq$30 iterations whereas other published algorithms need at least 50 iterations in order to reconstruct the Shepp-Logan phantom image. With the CatPhan phantom, the GP-BB algorithm achieved a clinically-reasonable image with 40 projections in 12 iterations, in less than 12.6 s. This is at least an order of magnitude faster in reconstruction time compared with the most recent reports utilizing GPU technology given the same input projections. For the head-and-neck

 0094-2405/2012/39(3)/1207/11/$30.00

clinical scan, clinically-reasonable images were obtained from 120 projections in 34–78 s converging in 12–30 iterations. In this reconstruction range (i.e., 120 projections) the image quality is visually similar to or better than the conventional FDK reconstructed images using 364 projections. This represents a dose reduction of nearly 67% (120/364 projections) while maintaining a reasonable speed in clinical implementation.

Conclusions: In this paper, we proposed a novel, fast, low-dose CBCT reconstruction algorithm using the Barzilai-Borwein step-size calculation. A clinically viable head-and-neck image can be obtained within ~34–78 s while simultaneously cutting the dose by approximately 67%. This makes our GP-BB algorithm potentially useful in an on-line image-guided radiation therapy (IGRT). © *2012 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3679865]

Key words: CBCT, image reconstruction, low-dose, compressed sensing, on-line IGRT

I. INTRODUCTION

In recent years, the introduction of cone-beam computed tomography (CBCT) in radiation therapy has enabled a precise, on-line positioning (and on-line/off-line replanning) of patients.[1,2] This is possible due to the wealth of information contained in the three-dimensional (3D)-CBCT images including (1) anatomic information,[1,2] (2) geometric information,[3,4] and (3) CT numbers for possible dose calculations for treatment verifications and plan re-optimizations.[5,6]

Because CBCT uses ionizing x-rays to image, however, there is a legitimate concern about hazardous radiation exposure to patients.[7] Due to this, excessive use of imaging should be prohibited and the benefits-vs-harm ratio should be carefully weighed and debated for each treatment, especially for pediatric patients. This concern has now become an issue of central importance in North America, not only in radiation oncology, but in broader radiology community (e.g., Image Wisely™ and Image Gently™ campaigns).

There are rather straightforward ways to reduce the imaging dose for CBCT, either (1) minimize the number of x-ray projections, (2) reduce the current setting in the x-ray tube (mA), and/or (3) reduce the total exposure time (ms). With the current-standard Feldkamp-Davis-Kress (FDK) reconstruction algorithm,[8] however, reducing the projections will cause aliasing artifacts (for example, see Figs. 6 and 7) the severity of which depends inversely on the number of projections, and if the mA and/or ms is reduced, the noise in the image would increase. Both of these properties of the FDK are extremely undesirable, especially if the images are used for guiding precision radiation therapy for cancer eradication.

In recent years, the exciting advances in compressed sensing theory has shown that sparse signals (at least in some known transform domain) can be reconstructed from much smaller number of samples than the Nyquist frequency would mandate.[9–21] In layman's terms, this means that nearly ideal images can be reconstructed even if only a few projections are available. This in turn means that the imaging dose can be safely reduced without compromising the image quality. Past works have shown that for CT-type reconstructions (both fan- and cone-beam), the total variation (TV) formulation has been particularly useful in exploiting prior knowledge of minimal variation in the x-ray attenuation characteristics across the human body.[9,11–13,16–19,21] However, a practical implementation of this method still remains a challenge. The main problem is the iterative nature in solving the TV-based compressed sensing formulation; in general it requires multiple iterations of forward and backward projections of large datasets and cannot be completed in a clinically feasible time frame (e.g., <1 min). Solving this rather cumbersome problem would require multiple innovations encompassing (1) computationally efficient parallel-programming with proper hardware and (2) mathematical formulation of an efficient search algorithm for fast-solution-convergence. The former issue has been resolved successfully with the use of graphics processing units (GPU).[22–25,30] This approach reduced the computational time from several hours to few minutes.[22,23] The motivation of this work is to propose a solution to the latter issue; to achieve a clinically realistic reconstruction time(s) on GPU hardware with an equivalent or better image quality for on-line image-guided radiation therapy (IGRT).

In this work, we propose a gradient projection algorithm that handles the TV-norm regularized least squares problem based on the Barzilai-Borwein (BB) formulation[26,27] in such an efficient manner that a clinically-reasonable patient image is reconstructed in ~12–30 iterations and a total time of ~34–78 s using a single GPU card (NVIDIA GTX 295, Santa Clara, CA). Comparison of our novel approach with the FDK and other published compressed sensing techniques are presented in detail with numerical and physical phantoms, and head-and-neck clinical patient data.

II. METHODS AND MATERIALS

The main problem is to solve the constrained convex optimization problem of the form

$$\min_x f(\boldsymbol{x}) = \left|\left|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\right|\right|_2^2 + \lambda TV(\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{x} \geq 0 \qquad (1)$$

where $\boldsymbol{x}$ = unknown CBCT volume image, $\boldsymbol{A}$ = Radon transform operator, $\boldsymbol{b}$ = measured projections data, λ = regularization constant, and TV = Total Variation (TV) regularization term. In this paper, the matrices are denoted as a boldface-uppercase letters and the vectors are denoted as a boldface-lowercase letters. Note that, in Eq. (1), the three-dimensional (3D) CBCT volume, $\boldsymbol{x}(i,j,k)$, is cast into a vector, $\boldsymbol{x}(l)$. Thus, the two volume representations, $\boldsymbol{x}(i,j,k)$ and $\boldsymbol{x}(l)$, are used interchangeably in this paper.

The TV term we used in this study is defined as

$$TV(\boldsymbol{x}(i,j,k)) = \sum_{i,j}\sqrt{[\boldsymbol{x}(i+1,j,k)-\boldsymbol{x}(i,j,k)]^2+[\boldsymbol{x}(i,j+1,k)-\boldsymbol{x}(i,j,k)]^2+[\boldsymbol{x}(i,j,k+1)-\boldsymbol{x}(i,j,k)]^2} \quad (2)$$

In its form, the first term in Eq. (1) is the fidelity term, it enforces fidelity of $\boldsymbol{x}$ with the measured projections data. The second term (the regularization term) promotes sparsity inherent in the x-ray attenuation characteristics of the human body.

Algorithms of significant acceptance and popularity in solving Eq. (1), so far, have mostly been based on separating the two terms and optimizing them individually in an alternating manner, i.e., the forward–backward splitting technique.[9,11,16,17,21–23] Figure 1(a) illustrates this approach. At iteration n, as commonly used in the gradient descent algorithms, a fixed small step-size α_{const} is chosen to reduce the fidelity term to obtain an intermediate solution $\boldsymbol{x}_{n+1}'$. The $(n+1)$-th solution $\boldsymbol{x}_{n+1}$ which has the minimal total variation is then searched around the $\boldsymbol{x}_{n+1}'$ to complete an iteration. This procedure is repeated until a desired solution $\boldsymbol{x}_{opt}$ is achieved. However, this approach may not result in a fast convergence due to the two-step approach. Intuitively, an algorithm that optimally reaches $\boldsymbol{x}_{n+1}$ in a single step, at each iteration, using a variable step-size (α_{opt}) would be more efficient computationally and faster in convergence [see Fig. 1(b)]. This is the intuition which led to the algorithm proposed herein.

The goal is a gradient projection algorithm that iteratively seeks a solution to Eq. (1) in the direction of the projected gradient while enforcing a non-negativity of the found solution. Let g_n be the gradient of $f(\boldsymbol{x}_n)$ defined as

$$\boldsymbol{g_n} = 2\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}_n - \boldsymbol{b}) + \lambda\nabla TV(\boldsymbol{x}_n) \quad (3)$$

where ^{T} is the transpose operator of the Radon transform matrix $\boldsymbol{A}$, which is physically interpreted as a back projection operation on $\boldsymbol{x}$. We then solve Eq. (1) iteratively using the gradient projection method

$$\boldsymbol{x}_{n+1} = [\boldsymbol{x}_n - \alpha_n\boldsymbol{p}_n]^+ \quad \text{where } [*]^+ = \max[*,0] \quad (4)$$

where

$$p_n(l) = \begin{cases} g_n(l) & \text{if } g_n(l) \le 0 \text{ or } x_n(l) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Here, α_n denotes the step size at iteration n, l denotes the voxel position index, and $\boldsymbol{p}_n$ denotes the projected gradient of the function $f(\boldsymbol{x})$ at $\boldsymbol{x}_n$.

The speed of convergence would be highly dependent on choosing a proper "step-size" α_n in Eq. (4), at each iteration. Remember, the less the number of iterations used to find the optimal solution $\boldsymbol{x}^*$, the less the number of times one needs to calculate the $\boldsymbol{A}$ and $\boldsymbol{A}^T$, which are computationally very expensive.

There are a few approaches in choosing an appropriate α_n including (1) a fixed, small α_n throughout, and (2) a variable α_n obtained through a back-tracking line-search method that satisfies a certain condition for ensuring convergence. For example, the well-known Armijo condition[28] enforces not only a monotonic decrease in the objective function but also a sufficient decrease of the objective function in each

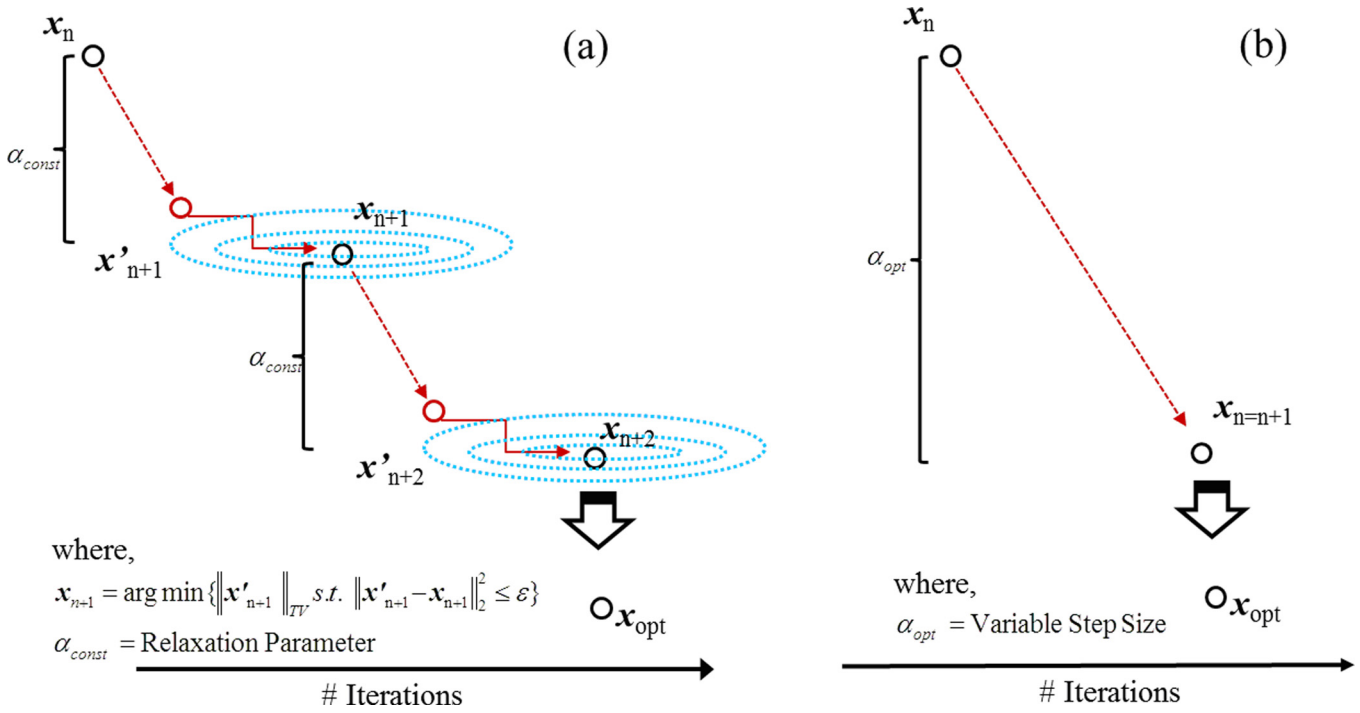

FIG. 1. Illustrations of (a) forward–backward splitting-type optimization, and (b) our one-step proposed approach to solve the TV-based constrained convex optimization problem in Eq. (1).

iteration for guaranteeing the convergence. The first, "fixed step-size" method is simple to implement yet finding an appropriate α_n is not trivial as there's a tradeoff between convergence speed and image quality. The second, "line-search" method is popular and guarantees a monotonic convergence but incurs a relatively high computational burden as the back-tracking line-search is an iterative process in itself, which is logically similar to that illustrated in Fig. 1(a), i.e., an iteration within iteration.

In this paper, a third alternative method is proposed using an approximate second-order solver, proposed by Barzilai and Borwein (BB) (Refs. 26 and 27) where the objective function may not be monotonically decreasing as in the back-tracking "line-search" method, but a much faster convergence is achieved. Unlike most traditional approaches that ensure convergence to an optimal solution by imposing a rather conservative condition of monotonic decrease of the objective function at each and every iteration, the BB method relaxes this constant decrease requirement in order to achieve even faster convergence.[26,32,33] Specifically, the conventional approaches calculate each step-size based on the current gradient of the cost function. As a result, a monotonic convergence is guaranteed throughout the iterative process. However, in the BB approach, the step-size is chosen based on both the current gradient and the previous gradient which could result in a nonmonotonic convergence. Utilization of this additional information (i.e., the past gradient) results in a faster convergence although the monotonic convergence behavior is not guaranteed. Basically, it calculates each step with the formulation [compare with Eq. (4)]

$$\boldsymbol{x}_{n+1} = [\boldsymbol{x}_n - \boldsymbol{H}_n^{-1}\boldsymbol{p}_n]^+ \tag{6}$$

where H_n is an approximation to the true Hessian of $f(\boldsymbol{x})$ at $\boldsymbol{x}_n$ (approximate second-order solver). To calculate $\boldsymbol{H}_n^{-1}$, the BB formulation makes a simple approximation to the Hessian by setting $\boldsymbol{H}_n = \eta^{(n)}\boldsymbol{I}$, where $\boldsymbol{I}$ denotes an identity matrix and $\eta^{(n)}$ is chosen to approximate the true Hessian over the most recent two iteration steps as

$$\boldsymbol{p}_n - \boldsymbol{p}_{n-1} \approx \eta^{(n)}[\boldsymbol{x}_n - \boldsymbol{x}_{n-1}] \tag{7}$$

where $\eta^{(n)}$ is calculated at each iteration that satisfies Eq. (7). In practical implementation, the optimal $\eta^{(n)}$ is solved in the least squares sense by

$$\eta^{(n)} = \frac{[\boldsymbol{x}_n - \boldsymbol{x}_{n-1}]^T[\boldsymbol{p}_n - \boldsymbol{p}_{n-1}]}{\left|\left|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\right|\right|_2^2} \tag{8}$$

Once $\eta^{(n)}$ is calculated, the Eq. (6) is updated by

$$\boldsymbol{x}_{n+1} = [\boldsymbol{x}_n - (\eta^{(n)})^{-1}\boldsymbol{p}_n]^+ \tag{9}$$

For $n=0$, since $\boldsymbol{x}_{n-1}$ and $\boldsymbol{p}_{n-1}$ in Eq. (8) are not known yet, we initialize $(\eta^{(n=0)})^{-1}$ as

$$(\eta^{(0)})^{-1} = \frac{\left|\left|\boldsymbol{g}^{(n)}\right|\right|_2^2}{\left|\left|\boldsymbol{A}\boldsymbol{g}^{(n)}\right|\right|_2^2} \tag{10}$$

which is the closed-form solution of the optimal step size in the quadratic fidelity term in Eq. (1).[28]

The advantage of this technique is at each iteration one needs to carry over only $\boldsymbol{x}_{n-1}$ and $\boldsymbol{p}_{n-1}$ to calculate $\eta^{(n)}$, which must be calculated in the previous step anyway. Thus, is unlike the Gradient-Projection-Barzilai-Linesearch GP-BL method where the step size is calculated via an iterative procedure [step 3 in Fig. 2(c), more on this later], there are no *extra* calculations/iterations that need to be performed to compute α_n which affects the speed of the optimization much favorably. Second, as found in the original BB publication,[26] the convergence of Eq. (6) should be faster than the standard first-order methods such as the back-tracking line-search discussed above, a similar result is shown in this work. Finally, since the entire $f(\boldsymbol{x})$ is minimized simultaneously in Eq. (1) and not iteratively as in the other works

(a) ASD-POCS

$$\begin{aligned}
&Step\,1: \quad \boldsymbol{x}'_{n+1} = \boldsymbol{x}_n - \alpha_{const}\boldsymbol{A}^T(\boldsymbol{A}[\boldsymbol{x}_n] - \boldsymbol{b})\\
&Step\,2: \quad \boldsymbol{x}_{n+1} = \arg\min\{\|\boldsymbol{x}'_{n+1}\|_{TV}\ s.t.\ \|\boldsymbol{x}'_{n+1} - \boldsymbol{x}_{n+1}\|_2^2 \le \varepsilon\}\\
&Step\,3: \quad if\ \ \boldsymbol{x}_{n+1}(i,j,k) \le 0,\ \boldsymbol{x}_{n+1}(i,j,k) = 0
\end{aligned}$$

(b) STF

$$\begin{aligned}
&Step1: \quad \boldsymbol{x}'_{n+1} = \boldsymbol{x}_n - \alpha_{const}\boldsymbol{A}^T(\boldsymbol{A}[\boldsymbol{x}_n] - \boldsymbol{b})\\
&Step2: \quad \boldsymbol{x}_{n+1} = S_{w,1}\{\boldsymbol{x}'_{n+1}\}\\
&Step3: \quad if\ \ \boldsymbol{x}_{n+1}(i,j,k) \le 0,\ \boldsymbol{x}_{n+1}(i,j,k) = 0
\end{aligned}$$

(c) GP-BL

$$\begin{aligned}
&Step\,1: \quad \boldsymbol{g}_n = \boldsymbol{A}^T(\boldsymbol{A}[\boldsymbol{x}_n] - \boldsymbol{b}) + \lambda\nabla_n\|TV(\boldsymbol{x}_n)\|_1^1\\
&Step\,2: \quad \boldsymbol{p}_n(i,j,k) = \boldsymbol{g}_n(i,j,k),\ \ \text{if}\ \boldsymbol{g}_n(i,j,k) \le 0\ \ \text{or}\ \boldsymbol{x}_n(i,j,k) \ge 0\\
&Step\,3: \quad while\ f(\boldsymbol{x}_n - \alpha_n\boldsymbol{p}_n) \ge f(\boldsymbol{x}_n) - \delta\alpha_n\boldsymbol{g}_n^T\boldsymbol{p}_n,\ \ \alpha_n = \beta\alpha_n\\
&Step\,4: \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \alpha_n\boldsymbol{p}_n\\
&Step\,5: \quad if\ \ \boldsymbol{x}_{n+1}(i,j,k) \le 0,\ \boldsymbol{x}_{n+1}(i,j,k) = 0
\end{aligned}$$

(d) GP-BB

$$\begin{aligned}
&Step\,1: \quad \boldsymbol{g}_n = \boldsymbol{A}^T(\boldsymbol{A}[\boldsymbol{x}_n] - \boldsymbol{b}) + \lambda\nabla_n\|TV(\boldsymbol{x}_n)\|_1^1\\
&Step\,2: \quad \boldsymbol{p}_n(i,j,k) = \boldsymbol{g}_n(i,j,k),\ \ \text{if}\ \boldsymbol{g}_n(i,j,k) \le 0\ \ \text{or}\ \boldsymbol{x}_n(i,j,k) \ge 0\\
&Step\,3: \quad \eta^{(k)} = \frac{[\boldsymbol{x}_k - \boldsymbol{x}_{k-1}]^T[\boldsymbol{p}_k - \boldsymbol{p}_{k-1}]}{\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\|_2^2}\\
&Step\,4: \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k - (\eta^{(k)})^{-1}\boldsymbol{p}_k\\
&Step\,5: \quad if\ \ \boldsymbol{x}_{n+1}(i,j,k) \le 0,\ \boldsymbol{x}_{n+1}(i,j,k) = 0
\end{aligned}$$

FIG. 2. Illustration of the computational processes required at each iteration for the four algorithms: (a) ASD-POCS, (b) STF, (c) GP-BL, and (d) GP-BB.

discussed above; and shown in Fig. 1(a), the overall complexity of the implementation is simplified while still guaranteeing an optimal solution.

In our implementation of this Gradient-Projection-Barzilai-Borwein (GP-BB) method, to speed up the algorithm further, the following has been adopted:

(a) For $n=0$, initialize $\boldsymbol{x}_0 = \text{FDK}$. This result in a faster convergence compared with setting $\boldsymbol{x}_0 = 0$.
(b) Two-resolution-level optimization. That is, first set $\boldsymbol{x}$ to $256 \times 256 \times 70$ volume, optimize, then resample to $512 \times 512 \times 70$ volume for a second-level optimization. The resolution at level one and two are $0.97 \times 0.97 \times 2.0$-mm and $0.49 \times 0.49 \times 2.0$-mm, respectively.
(c) The entire code is structured and implemented in C with the CUDA programming environment (NVIDIA, Santa Clara, CA) to utilize the massive parallel computational capability of the GPU hardware. We used a single GTX 295 card ($\sim\$500^{\text{US}}$) that consists of 480 processing cores with 1.24 GHz clock speed and 1792 MB memory. In terms of CPU, an Intel Core™ i7 with 2.68 GHz clock speed, 12.0 GB DDR3 RAM, on a 64-bit Window 7 OS is used.

Three major computational tasks were parallelized in the CUDA environment: (1) the forward projection $\boldsymbol{A}$, (2) the back projection $\boldsymbol{A}^T$, and (3) the vector operations to calculate $\eta^{(n)}$, $\boldsymbol{Ax-b}$, $TV(..)$, etc. For the forward projection operations, each detector pixel is set as a GPU thread and the image voxels which lie along the path from the cone-beam source to the corresponding pixel are summed. Since this summation of voxels, in the ray path, can be independently computed for each detector pixel, this feature has been utilized in the GPU coding as a parallel computations.[30] For the back projection operations, we have instead set each image voxel as a GPU thread. Similar strategies were implemented on the vector operations as well.

To evaluate the performance of our GP-BB algorithm, we have compared it with three other algorithms, two of which are published. First, the adaptive-steepest-descent-projections-onto-convex-set (ASD-POCS) method proposed by Sidky and Pan[17] [described in Fig. 1(a)] was implemented. Second, the soft-threshold filtering approach (STF) proposed by Yu and Wang[21] was implemented. This algorithm is essentially similar to the ASD-POCS except that an approximate solution is proposed over iteratively calculating the second step, shown in Fig. 1(a) (i.e., minimizing the total variation step) in order to reduce the computational burden. Third and finally, we have implemented a first-order Gradient-Projection-Back-tracking-Linesearch (GP-BL) algorithm that attempts to simultaneously minimize both terms in Eq. (1) in a single step as opposed to the forward–backward splitting technique in ASD-POCS and STF. This single step approach is similar to the GP-BB except that an acceptable α_n in Eq. (4) is searched without the approximate second-order Hessian information as explored in the GP-BB method. Essentially, at each iteration in Eq. (4), α_n is found through the back-tracking line-search along the direction of the current gradient.

The computation tasks at each iterative process for each of these algorithms are illustrated in Fig. 2. First, the ASD-POCS iteration is shown in Fig. 2(a). The algorithm starts by finding the intermediate solution $\boldsymbol{x}_{n+1}{}'$ through a SART-type approach with a constant step-size where a promotion of the data fidelity term only is considered (step 1). The TV-norm is then minimized around this $\boldsymbol{x}_{n+1}{}'$ through an iterative, convex optimization process for which we employed a commonly used steepest descent approach in this paper (step 3). Therefore, at each iterative step of the ASD-POCS algorithm, an additional iterative process is required (step 3). Second, the STF approach proposed by Yu and Wang[21] is illustrated in Fig. 2(b). As shown on the figure, the algorithm starts exactly same as the ASD-POCS for step 1. However, instead of using an iterative, convex optimization procedure to minimize the TV around the intermediate solution $\boldsymbol{x}_{n+1}{}'$, the algorithm employs a batch, noniterative, soft-threshold filtering algorithm to remove the extra iteration step and hence reduces the computational time. The soft-threshold filtering procedure is denoted by $S_{w,1}(\boldsymbol{x}_{n+1}{}')$ in step 2 on Fig. 2(b). It is a regularization function that updates the $\boldsymbol{x}_{n+1}{}'$ as a function of TV ($\boldsymbol{x}_{n+1}{}'$) using a closed-form heuristic formula. Interested readers are encouraged to read Yu and Wang[21] for details. Third, the GP-BL algorithm proposed here as an alternative to GP-BB is illustrated in Fig. 2(c). It starts by calculating the gradient of the cost function consisting of the data fidelity and the regularization terms (step 1), followed by a projection of the gradient (step 2). At step 3, a back-tracking line-search is performed by evaluating the objective function with a decreasing step-size α_n until the well-known Armijo condition stated by the inequality in step 3 is met. This not only enables a monotonic decrease in the objective function but also satisfies a sufficient decrease criterion for convergence to the optimal solution. In this study, the constants δ and β are fixed to 0.02 and 0.7, respectively. Once the step-size α_n is obtained, a gradient descent step is conducted while enforcing the non-negative constraint (step 4). Finally, the proposed GP-BB algorithm is illustrated in Fig. 2(d). Referring to the figure, steps 1 and 2 are the same as those in the GP-BL algorithm. However, as we have illustrated our algorithm mathematically, the iterative back-tracking line-search in Fig. 2(c) is replaced by a much simpler, noniterative vector operations for the computation of $\eta^{(n)}$ (step 3). Therefore, favorably compared with the GP-BL approach, the step-size search is performed without the extra iterative calculation of the back-tracking line-search. The gradient descent step is conducted using $\alpha_n = (\eta^{(n)})^{-1}$ (step 4). It should be noted here though that there have been very recent studies reporting the effectiveness of the BB-based approaches for CT reconstruction.[13,35] These efforts could complement our work in developing the most mature form of the BB-based CT/CBCT reconstruction techniques.

The Shepp-Logan numerical phantom, the CatPhan 600 physical phantom (The Phantom Laboratory, Salem, NY), and a clinically-treated head-and-neck patient acquired from the TrueBeam™ system (Varian Medical Systems, Palo Alto, CA) are used for comparison purposes. For the TrueBeam™ scans, a total of 364 projections were acquired in a

200-degree rotation, in a full-fan mode. The imager has 1024×768 pixels with 0.388×0.388-mm resolution. This was down-sampled to 512×384 pixels with 0.776×0.776-mm for the reconstructions. Evenly spaced angles were subsampled and used for varying the number of projections for the image reconstruction.

III. RESULTS

Figure 3 shows the reconstructed 2D images of the Shepp-Logan phantom using the four algorithms described earlier. A total of 40 projections in fan-beam geometry were used for the reconstructions. As can be seen, the Gradient Projection (GP)-type algorithms outperforms the forward–backward splitting-type algorithms in terms of image quality and speed of convergence. At about 50 and 30 iterations, the GP-BL and GP-BB algorithms show convergence, whereas the ASD-POCS and STF algorithms clearly still needs further convergence at 50 iterations. Visually, the GP-BB shows the fastest convergence and this is quantitatively demonstrated in Figs. 4 and 5. In Fig. 4, the line profile comparison is shown after 30 iterations, for example. It is clear from this figure that the level of agreement to the ground truth goes in the order of GP-BB > GP-BL > STF, and ASD-POCS. This finding holds true at all levels of iterations, as illustrated in Fig. 5. Here, the relative error is defined as the mean-squared percent error from the ground truth pixel values:

$$\text{ReletiveError}(\%) = \frac{\sum_{i,j,k} \left(\boldsymbol{x}_{i,j,k} - \boldsymbol{x}_{i,j,k}^{\text{GroundTruth}}\right)^2}{\sum_{i,j,k} \left(\boldsymbol{x}_{i,j,k}^{\text{Ground Truth}}\right)^2} \times 100 \quad (11)$$

where $x_{i,j,k}$ corresponds to the voxel values in the reconstructed volume x and $x^{\text{GroundTruth}}$ refers to the ground truth values of the Shepp-Logan phantom used. As can be seen from the figure, all three algorithms other than the GP-BB continue to decrease at 50 iterations, whereas the GP-BB algorithm reaches saturation at ~25–30 iterations. One thing to note is that, due to the nonmonotonic feature of the BB algorithm discussed in Sec. II, the relative error does not decrease in a smooth manner due to the inherent properties of calculating the step size without conducting a line-search at each iterative step.

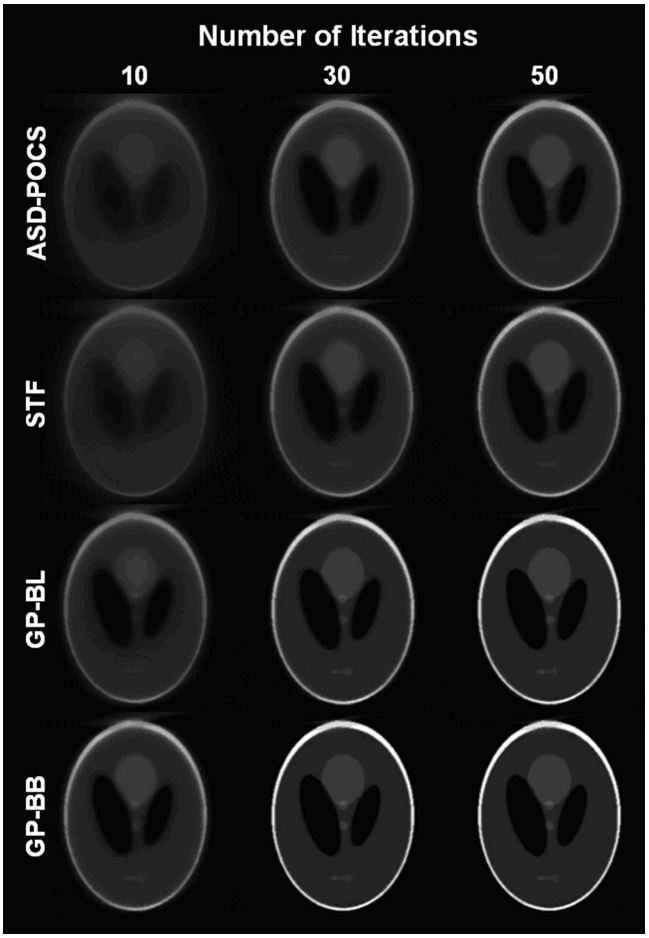

FIG. 3. The reconstructed images of the Shepp-Logan phantom, using the respective four algorithms, as a function of 10, 30, and 50 iterations. A total of 40 projections in fan-beam geometry were used for the reconstructions.

In order to show the computational efficiency of each algorithm, we have measured the computational time performances of the four algorithms implemented on the same GPU card (see Table I). To conduct a fair comparison, we kept all experimental conditions the same for all algorithms. That is, the number of iterations was set to 50, the reconstruction volume was set to $256 \times 256 \times 64$, the number of projections was 42, the detector size was 512×384, the unknown CBCT volume image was all initialized to zero (i.e., $\boldsymbol{x}^{(0)} = \mathbf{0}$), and the multiresolution optimization was not used. It is found that the GP-BL algorithm takes the longest time to compute due to the high computational cost of conducting the back-tracking line-search at each iteration. It can also be observed that those algorithms that have iterations within iterations, i.e., ASD-POCS and GP-BL, the standard deviation is also large compared to the other two algorithms that do not have a second iteration loop. The GP-BB and STF algorithms have a consistent computational time and, moreover, their convergence is noticeably faster than the other two algorithms. It should be noted, though, that the implementations of the ASD-POCS and STF may not have been exactly reproduced as the ones originally proposed and implemented, as all of these algorithms were written in-house. Although we attempted to make the fairest comparison by best implementing the original ideas of the algorithms using the published information, it is difficult to reproduce the same exact performance due to a difference in the experimental setup and data used. As a result, we acknowledge that our evaluations of these algorithms may not represent their best possible implementation and thus performance and image quality.

In addition to the comparisons with the compressed sensing-type algorithms, we've also compared the GP-BB against the conventional and commercially-used, filtered backprojection-type algorithm proposed by Feldkamp, Davis, and Kress (FDK, 1984).[8] Figure 6 show the 3D volumes reconstructed with the two algorithms, using some or all of the 364 projections acquired from the TrueBeam™ CBCT system. Reconstruction times are labeled on the

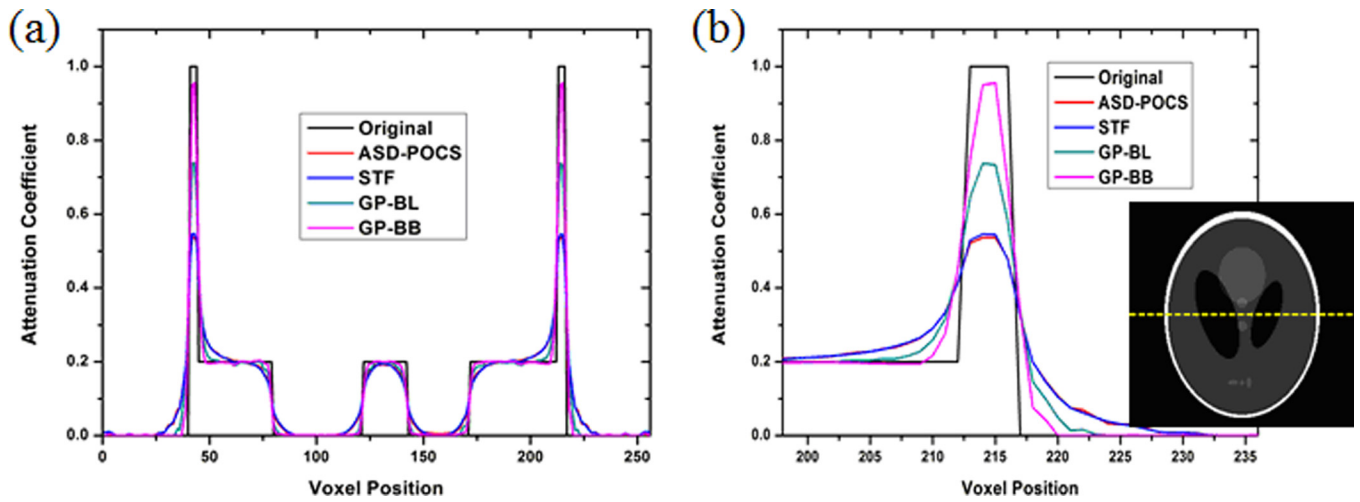

FIG. 4. Line profiles of the respective four algorithms with the (a) full line across the Shepp-Logan phantom, and (b) magnified view of the right one-third. The figure inset shows where the line profiles were generated.

figure. As can be seen, even with the dose reduction to $1/9^{th}$ (=40/364 projections), Fig. 6(b) shows a reasonable image quality achieved by the GP-BB algorithm comparable to that of the FDK reconstructed image using all of 364 projections [Fig. 6(c)] with less noise, while completing this in under 12.6 s. Needless to say, the image quality of Fig. 6(b) is better with minimal aliasing artifacts compared with that of Fig. 6(a), which is FDK reconstructed volume using the same 40 projections. To note, this achievement of 12.6 s is about an order of magnitude faster than that of the latest reports on GPU-accelerated forward–backward splitting-type algorithms, discussed in Fig. 1(a).[22,23]

Figure 7 shows a matrix view of the various image qualities achieved using the GP-BB algorithm as functions of both the number of projections and the number of iterations, for the head-and-neck example patient. The window and level were kept the same for all images. The first row show the images reconstructed with the FDK for comparison. It is observed that as the number of projections increases, the image quality increases in both the FDK and GP-BB algorithms, and as the number of iterations increases in GP-BB, the image quality increases too. It is also observed that, in GP-BB, the qualitative increase in image quality is relatively more significant from 90 to 120 projections than from 120 to 180 projections. This also means that the FDK-initialized GP-BB algorithm always does better than just the FDK alone, for any number of iterations per given number of projections.

Table II shows a comprehensive list of the reconstruction times recorded for various combinations of input condition, encompassing that of the examples shown in Fig. 7. From this list and Fig. 7, we can deduce that, although subjective, visually a "reasonable" image quality for clinical use can be obtained in the range of 12–30 iterations with 120–180 projections. The range of reconstruction times would be between 34 and 117 s, that is, all are within ~2 min or less. In terms of dose reduction, this would be on the order of ~1/3–1/2 depending on the projections used (i.e., 120/364 or 180/364 projections, respectively).

Figure 8 displays, side-by-side, the GP-BB-reconstructed images using 120 projections [Figs. 8(b)–8(e)] in closer comparison with the FDK-reconstructed image using 364 projections [Fig. 8(a)]; currently in-use in clinic). The images using only the 120 projections are displayed here, as opposed to images using 180 projections, since our interest is in generating a reasonable quality images with a minimally necessary radiation dose. Also, as mentioned, there's a

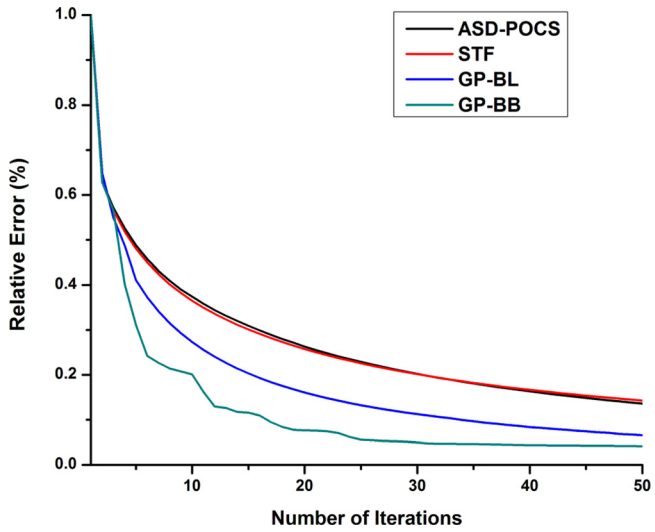

FIG. 5. Mean-squared relative percent error as a function of the number of iterations, for the respective four algorithms. The Shepp-Logan numerical phantom was used as the gold standard.

TABLE I. Computational time recorded to run 50 iterations.

(s)	ASD-POCS	STF	GP-BL	GP-BB
Total computational time	60.47	22.77	66.37	24.56
Average computational time/iteration	1.20	0.45	1.32	0.49
SD	0.36	0.03	0.18	0.04

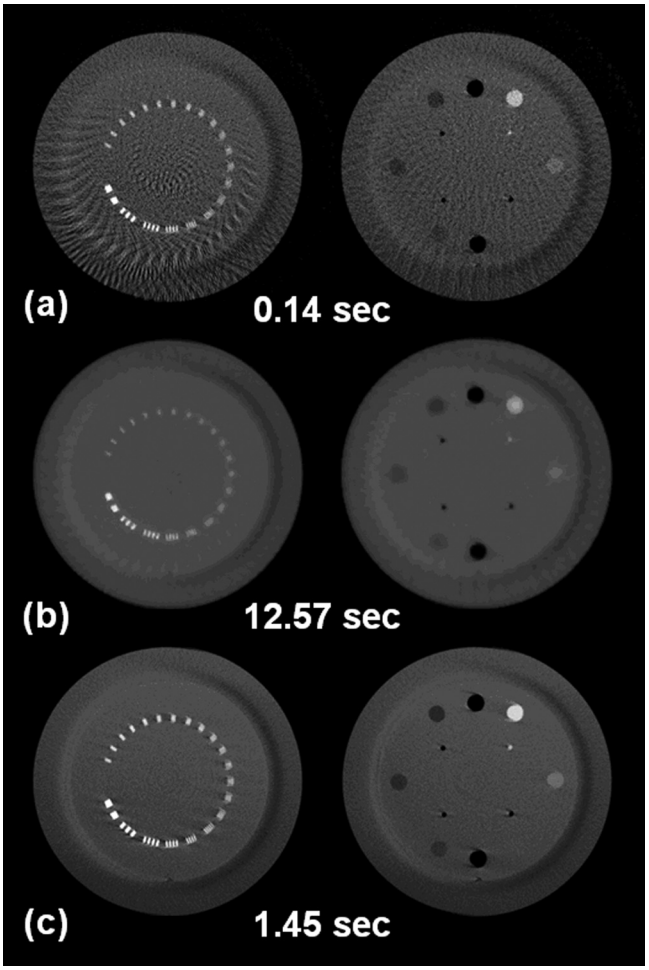

FIG. 6. Spatial and contrast resolution slices of the reconstructed CatPhan 600 phantom using (a) FDK with 40 projections, (b) GP-BB with 40 projections in 12 iterations, and (c) FDK with 364 projections. The reconstruction times are listed on the figure.

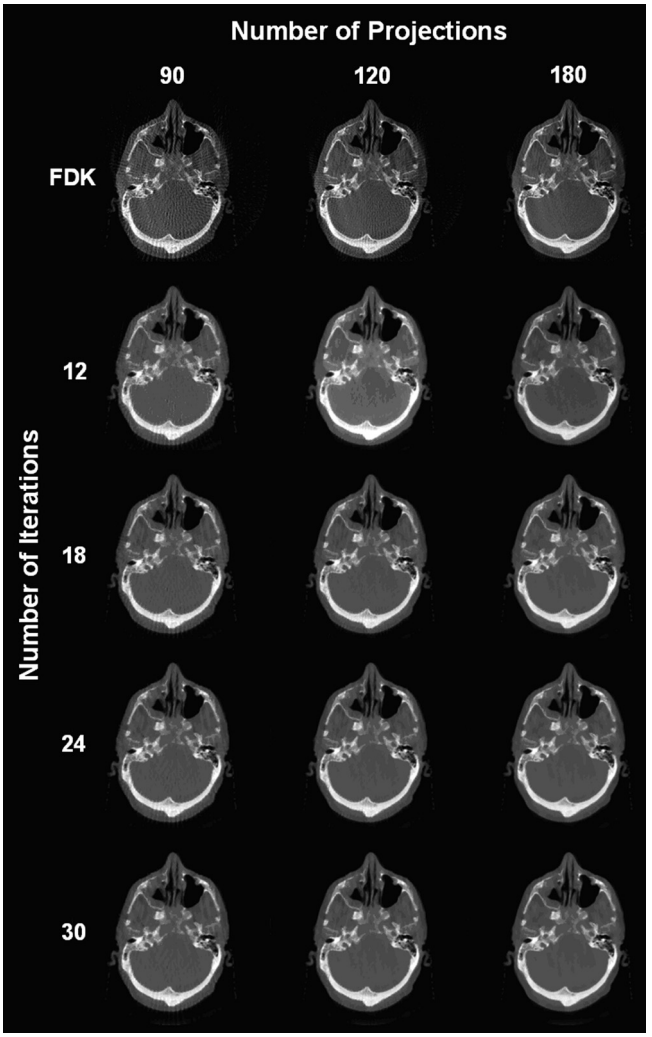

FIG. 7. A matrix view of the various image qualities achieved, using the GP-BB algorithm, as functions of number of projections and number of iterations, for the head-and-neck example patient. The window and level were kept the same for all images.

relatively more benefit in terms of image quality going from 90 to 120 than from 120 to 180 projections (Fig. 7), i.e., diminishing increase in image quality for a given increase in dose. The GP-BB-reconstructed image using 364 projections is also displayed [Fig. 8(f)] to show the limiting image quality that can be achieved with the GP-BB algorithm. First of all, the upper-limit GP-BB image using 364 projections is a visually better quality image than that of the FDK (i.e., less noise, streaking artifacts around bones, etc.), which reaffirms the result shown in Fig. 7; given equal dose the GP-BB always does better. But, more importantly, the image qualities of the 120-projection-images are comparable to the FDK image and are reconstructed within a reasonable 34–78 s. Of course, the necessary image quality for clinical use is quite subjective and requires further (clinical) testing to determine for each site, which is of our future research, it is still encouraging that a "visually" similar quality images can be obtained in one-third the dose in a "reasonable" time frame. To the best of our knowledge, this computational speed achieved using the GP-BB algorithm is the fastest compressed sensing-type optimization that have been proposed for the CBCT reconstruction to date.[12,16,17,20–23]

IV. DISCUSSION

IV.A. Algorithm performance

The translation of compressed sensing-based CBCT reconstruction algorithms onto radiation therapy clinical settings has been difficult due to the fact that the mathematical formulation is solved numerically and not analytically, thus requiring a prohibitive time to solve. A single, complete iteration involves at least one forward and one backward projection calculations, which are computationally expensive. Although significant amount of computational time can be spared by parallelizing the task with GPU programming,[22–25,30] still, the majority of time (e.g., $>80\%$) spent in the reconstruction is on calculating the forward and backward projections. Therefore, for an algorithm/technique to be efficient, it must (1) require a minimal number of forward and backward projection calculations per iteration, and (2) converge in a minimal number of total iterations.

Besides the well-known forward–backward splitting technique that we compared in this work, there are other

TABLE II. List of the reconstruction times recorded for various projections and iterations tested.

Time (s)	No. of projections (views) used				
No. of iterations	60	90	120	180	364
FDK	0.27	0.34	0.50	0.75	1.45
6	9.891	14.48	19.25	28.35	58.11
12	17.58	25.57	33.77	49.94	101.83
18	24.88	36.55	48.21	71.92	146.51
24	32.92	47.62	62.86	92.86	188.37
30	39.91	58.87	77.99	116.55	234.51

compressed sensing-based methods with a focus on achieving faster convergence than previously reported[12,13] (such as based on Nestrov's first-order method). However, on reviewing their works, it was observed that although the convergence rate (determining the number of iterations needed to reach a desired solution) outperforms the comparing counterparts, the algorithms require multiple forward and backward projection calculations at each iteration[12] or an extra iterative procedure to calculate the additional unknown parameters[13] leading to an increased reconstruction times. The nonconvex prior image constrained compressed sensing (NCPICCS) algorithm reported by Ramirez-Giraldo, *et al.*[29] also suffers from the similar complexities where an extra-calculation of forward and backward projections is required to calculate each step size. The GP-BB algorithm, on the other hand, requires (1) only one forward and one backward projection calculations per iteration, which is the least number required for solving any iterative reconstruction techniques, and (2) a simple gradient step-size calculation [i.e., Eq. (8)] that needs only the prior and current values of the gradient and the image volume which occupy <300 MB of memory, thus facilitating easy incorporation onto a single GPU card memory (1.7 GB storage). In our implementation, calculating the step size takes a negligible time, so the great majority of the time is spent on the forward and backward projection calculations (e.g., >98% of time). This demonstrates that the GP-BB algorithm requires only a minimal computational load needed to reach a solution.

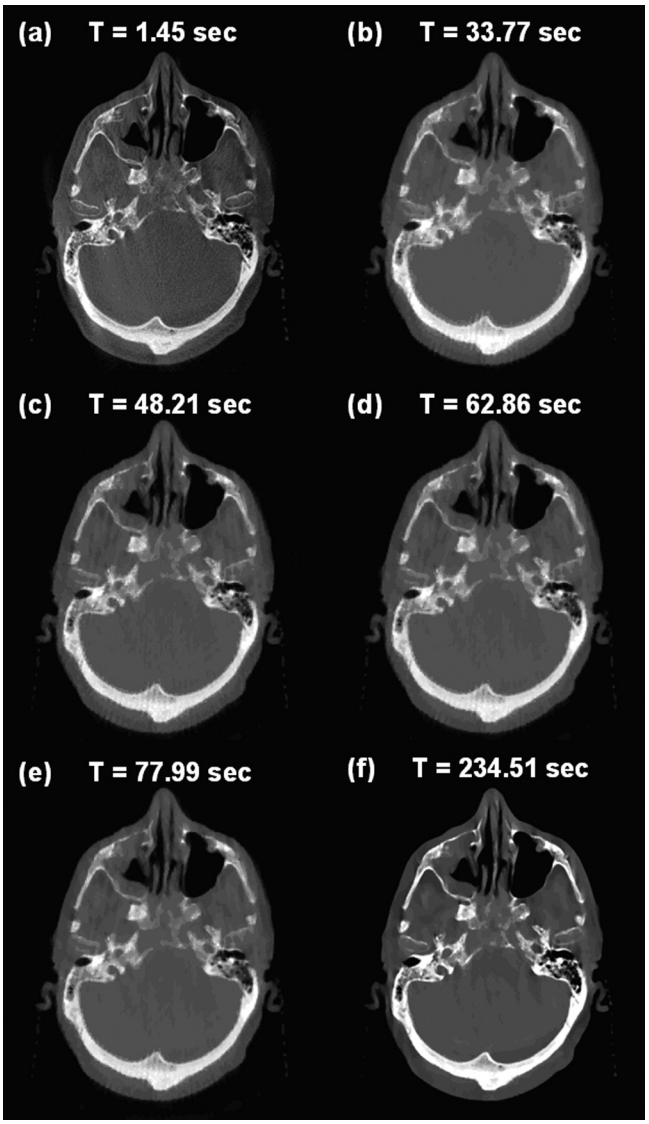

FIG. 8. Selected images from Fig. 6; (a) FDK using 364 projections, (b) GP-BB with 12 iterations using 120 projections, (c) GP-BB with 18 iterations using 120 projections, (d) GP-BB with 24 iterations using 120 projections, and (f) GP-BB with 30 iterations using 364 projections. The reconstruction times are listed on the figure.

IV.B. Dose reduction

It needs to be stated that, if reconstruction time is of no issue and thus enough iterations are allowed, all of the compressed sensing-type algorithms evaluated in this study will eventually reach an optimal solution as anticipated from Fig. 5. That means, if an equal number of projections are used with each algorithm, then the achieved image quality at the end will be identical, and hence, no benefit in terms of dose and/or image quality will be observed for any one algorithm. However, we do not have an infinite time to spare, especially in an on-line IGRT environment, and hence, an algorithm that can produce the most optimal image under a reasonable time and with the least amount of projections (i.e., dose), is favored. Our proposed GP-BB algorithm fits relatively well in this respect. In the Catphan phantom experiment, the GP-BB produced a reasonable image with a highly undersampled projections (40/364 ≅ 89% dose reduction; see Fig. 6) in ~12.6 s. However, for a clinical patient case, about 120 projections or more were needed to generate a reasonable quality images, taking a respectable ~34–78 s. This achievement still represents a significant dose reduction of ≅67%, but any further dose reduction (i.e., less number of projections) is generally not recommended due to a fast degradation of the image quality, although the reconstruction times will further decrease as the number of projections do. This has also been the observation of earlier works.[19,22,23] The possible reason for needing more projections in patients than in phantoms is that the internal anatomy of humans are relatively less sparse, and thus require more data to properly represent it. In addition, since the sparseness is organ-patient specific, much research is needed to determine the appropriate number of projections

needed, and hence, the achievable dose reduction, for each organ and patient. The appropriate mAs setting per projection is another parameter that needs to be examined. Utilizing prior information such as the planning CT would be a one good way to decide what dose reduction is possible-appropriate for each case.

IV.C. Regularization parameter λ

The regularization parameter λ [Eq. (1)] is one of the most influential parameters affecting the image quality. It was our experience that the higher this value, the blurrier and smoother the images, and the smaller it is, the sharper and noisier the images. This is due to the fact that λ is the weighting factor for the TV-norm regularization term in Eq. (1). Thus, if λ is high, then more weight is given in the GP-BB optimization to minimize the variation across the image, and hence, the blurrier but smoother the look. Oppositely, if λ is low, then more weight is given to the fidelity term in Eq. (1), and hence, the high frequency information will survive, thus preserving the noise and streaks. For example, we observed some irregular ripples in a uniform phantom region, in Fig. 4, when a small λ is applied, demonstrating the importance of a proper/optimal λ selection.

Recently, there have been considerable interests in optimizing the λ value in a regularization-type optimization problems.[33,34] Although the purpose of these works is for different applications, we anticipate that similar strategies can also be applied to the TV-based CBCT reconstruction problem as well. But, for now, since there is no global standard in λ value(s) (Ref. 19) for CBCT reconstructions, the selection of λ was subjectively picked by painstakingly repeating a large range of values. From this experience, we learned that for fewer projections, a relatively high λ is needed to suppress the overwhelming noise and streaks, while for more projections, a relatively low λ is sufficient. Specifically, with 100 or less projections, we set $\lambda = 0.0075$, and for >100 projections, we set $\lambda = 0.0025$, for the head-and-neck patient case. Obviously, more research is needed in finding an optimal λ values for various situations, and that this value is likely not only number of projections dependent, but will also be patient and site dependent as well. For best clinical practice, an automated selection of λ based on a prior knowledge, whatever that may be (including a planning CT), will help facilitate the clinical translation of this technology into a busy on-line radiation therapy environment.

IV.E. Future works

In this study, we limited our investigation of the newly proposed GP-BB algorithm to feasibility tests using digital and physical phantoms and a patient case. We have learned that the new algorithm performs relatively well with other published algorithms, and that a reasonable clinical head-and-neck CBCT image can be obtained in about a minute or less. Therefore, the next step is to test this algorithm on various clinical sites, including the half-fan scans (thorax, abdomen, pelvis, etc.), optimize the λ value for the various sites, check the stability of CT numbers for dose calculation purposes, and perform a rigorous image registration testing with the planning CT to calculate the motion deltas for the on-line IGRT use. Determining an appropriate λ value for each clinical situation may be quite difficult to do. An important next step, therefore, is to determine a standard method-index for evaluating the image quality appropriate for the IGRT use, to aid in the search for the best-fit λ values.

Nonetheless, our proposed GP-BB algorithm brings us one step closer to reducing the radiation dose without compromising the image quality, in an on-line IGRT environment, making the iterative CBCT reconstruction approach more practical.

V. CONCLUSIONS

In this paper, we proposed a novel, fast, low-dose CBCT reconstruction algorithm using the Barzilai-Borwein step-size calculation. A clinically viable head-and-neck image can be obtained within ~34–78 s while simultaneously cutting the dose by approximately 67%. This makes our GP-BB algorithm potentially useful in an on-line IGRT.

ACKNOWLEDGMENTS

This project is partially supported by the Clinical & Translational Research Institute 2011 Pilot Innovative Technology Grant provided through the University of California San Diego, and the National Research Foundation of Korea (NRF) Grant Nos. 2011-0022021, 2011-19192, and 2011-0001851 funded by the Ministry of Education, Science & Technology (MEST). There is no conflict of interest.

[a]Electronic mail: bosong@ucsd.edu; Telephone: +1-858-720-1085; Fax: +1-858-822-5568.
[b]Electronic mail: wysong@ucsd.edu; Telephone: +1-858-246-0886; Fax: +1-858-822-5568.

[1]D. A. Jaffray, "Emergent technologies for 3-dimensional image-guided radiation delivery," Semin. Radiat. Oncol. **15**(3), 208–216 (2005).
[2]D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, "Flat-panel cone-beam computed tomography for image-guided radiation therapy," Int. J. Radiat. Oncol., Biol., Phys. **53**(5), 1337–1349 (2002).
[3]J. C. Park, S. H. Park, J. H. Kim, S. M. Yoon, S. S. Kim, J. S. Kim, Z. Liu, T. Watkins, and W. Y. Song, "Four-dimensional cone-beam computed tomography and digital tomosynthesis reconstructions using respiratory signals extracted from transcutaneously inserted metal markers for liver SBRT," Med. Phys. **38**(2), 1028–1036 (2011).
[4]W. Song, B. Schaly, G. Bauman, J. Battista, and J. Van Dyk, "Image-guided adaptive radiation therapy (IGART): Radiobiological and dose escalation considerations for localized carcinoma of the prostate," Med. Phys. **32**(7), 2193–2203 (2005).
[5]J. Hatton, B. McCurdy, and P. B. Greer, "Cone beam computerized tomography: The effect of calibration of the Hounsfield unit number to electron density on dose calculation accuracy for adaptive radiation therapy," Phys. Med. Biol. **54**(15), N329–N346 (2009).
[6]S. Yoo and F. F. Yin, "Dosimetric feasibility of cone-beam CT-based treatment planning compared to CT-based treatment planning," Int. J. Radiat. Oncol., Biol., Phys. **66**(5), 1553–1561 (2006).
[7]W. Y. Song, S. Kamath, S. Ozawa, S. A. Ani, A. Chvetsov, N. Bhandare, J. R. Palta, C. Liu, and J. G. Li, "A dose comparison study between XVI and OBI CBCT systems," Med. Phys. **35**(2), 480–486 (2008).
[8]L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," J. Opt. Soc. Am. A **1**(6), 612–619 (1984).
[9]J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," Phys. Med. Biol. **55**(22), 6575–6599 (2010).

[10]E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**(2), 489–509 (2006).
[11]G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," Med. Phys. **35**(2), 660–663 (2008).
[12]K. Choi, J. Wang, L. Zhu, T. S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," Med. Phys. **37**(9), 5113–5125 (2010).
[13]J. H. Jørgensen, T. L. Jensen, P. C. Hansen, S. H. Jensen, E. Y. Sidky, and X. Pan, "Accelerated gradient methods for total variation based CT image reconstruction," the 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Potsdam, Germany, 2011).
[14]D. L. Donoho, "Compressed sensing," IEEE Trans. Inform Theory **52**(4), 1289–1306 (2006).
[15]D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inf. Theory **52**(1), 6–18 (2006).
[16]L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelriess, "Improved total variation-based CT image reconstruction applied to clinical data," Phys. Med. Biol. **56**(6), 1545–1561 (2011).
[17]E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**(17), 4777–4807 (2008).
[18]T. P. Szczykutowicz and G. H. Chen, "Dual energy CT using slow kVp switching acquisition and prior image constrained compressed sensing," Phys. Med. Biol. **55**(21), 6411–6429 (2010).
[19]J. Tang, B. E. Nett, and G. H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," Phys. Med. Biol. **54**(19), 5781–5804 (2009).
[20]J. Wang, T. F. Li, and L. Xing, "Iterative image reconstruction for CBCT using edge-preserving prior," Med. Phys. **36**(1), 252–260 (2009).
[21]H. Yu and G. Wang, "A soft-threshold filtering approach for reconstruction from a limited number of projections," Phys. Med. Biol. **55**(13), 3905–3916 (2010).
[22]X. Jia, Y. Lou, J. Lewis, R. Li, X. Gu, C. Men, W. Y. Song, and S. B. Jiang, "GPU-based fast low-dose cone beam CT reconstruction via total variation," J. X-Ray Sci. Technol. **19**(2), 139–154 (2011).
[23]X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation," Med. Phys. **37**(4), 1757–1760 (2010).
[24]F. Xu and K. Mueller, "Accelerating popular tomographic reconstruction algorithms on commodity PC graphics hardware," IEEE Trans. Nucl. Sci. **52**(3), 654–663 (2005).
[25]F. Xu and K. Mueller, "Real-time 3D computed tomographic reconstruction using commodity graphics hardware," Phys. Med. Biol. **52**(12), 3405–3419 (2007).
[26]J. Barzilai and J. M. Borwein, "2-Point step size gradient methods," IMA J. Numer. Anal. **8**(1), 141–148 (1988).
[27]M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," IEEE J. Sel. Top. Signal Process. **1**(4), 586–597 (2007).
[28]D. P. Bertsekas, *Nonlinear programming*, 2nd ed. (Athena Scientific, Belmont, Mass., 1999).
[29]J. C. Ramirez-Giraldo, J. Trzasko, S. Leng, L. Yu, A. Manduca, and C. H. McCollough, "Nonconvex prior image constrained compressed sensing (NCPICCS): Theory and simulations on perfusion CT," Med. Phys. **38**(4), 2157–2167 (2011).
[30]J. C. Park, S. H. Park, J. S. Kim, Y. Kim, M. K. Cho, H. K. Kim, Z. Liu, S. B. Jiang, B. Song, and W. Y. Song, "Ultra-fast digital tomosynthesis reconstruction using general-purpose GPU programming for image-guided radiation therapy," Technol. Cancer Res. Treat. **10**(4), 295–306 (2011).
[31]M. Raydan, "The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem," SIAM J. Optim. **7**(1), 26–33 (1997).
[32]L. Grippo and M. Sciandrone, "Nonmonotone globalization techniques for the Barzilai-Borwein gradient method," Comput. Optim. Appl. **23**(2), 143–169 (2002).
[33]J. Wang, H. Guan and T. Solberg, "Inverse determination of the penalty parameter in penalized weighted least-squares algorithm for noise reduction of low-dose CBCT," Med. Phys. **38**(7), 4066–4072 (2011).
[34]L. Zhu and L. Xing, "Search for IMRT inverse plans with piecewise constant fluence maps using compressed sensing techniques," Med. Phys. **36**(5), 1895–1905 (2009).
[35]J. C. Park, B. Y. Song, S. H. Park, J. S. Kim, H. K. Kim, T. S. Suh, Z. Liu, and W. Y. Song, "Fast, iterative, low-dose, cone-beam computed tomography reconstruction using a gradient projection algorithm (Oral presentation) *Joint AAPM/COMP Meeting*, *Vancouver, Canada* (2011).