Accelerated barrier optimization compressed sensing (ABOCS) for CT reconstruction with improved convergence

2014 Phys. Med. Biol. 59 1801

(http://iopscience.iop.org/0031-9155/59/7/1801)

IOP Publishing | Institute of Physics and Engineering in Medicine — Physics in Medicine and Biology

Phys. Med. Biol. **59** (2014) 1801–1814 — doi:10.1088/0031-9155/59/7/1801

Accelerated barrier optimization compressed sensing (ABOCS) for CT reconstruction with improved convergence

Tianye Niu[1], Xiaojing Ye[2], Quentin Fruhauf[1], Michael Petrongolo[1] and Lei Zhu[1]

[1] Nuclear and Radiological Engineering and Medical Physics Programs, The George W Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

E-mail: leizhu@gatech.edu

Received 7 July 2013, revised 11 November 2013
Accepted for publication 9 January 2014
Published 14 March 2014

Abstract

Recently, we proposed a new algorithm of accelerated barrier optimization compressed sensing (ABOCS) for iterative CT reconstruction. The previous implementation of ABOCS uses gradient projection (GP) with a Barzilai–Borwein (BB) step-size selection scheme (GP-BB) to search for the optimal solution. The algorithm does not converge stably due to its non-monotonic behavior. In this paper, we further improve the convergence of ABOCS using the unknown-parameter Nesterov (UPN) method and investigate the ABOCS reconstruction performance on clinical patient data. Comparison studies are carried out on reconstructions of computer simulation, a physical phantom and a head-and-neck patient. In all of these studies, the ABOCS results using UPN show more stable and faster convergence than those of the GP-BB method and a state-of-the-art Bregman-type method. As shown in the simulation study of the Shepp–Logan phantom, UPN achieves the same image quality as those of GP-BB and the Bregman-type methods, but reduces the iteration numbers by up to 50% and 90%, respectively. In the Catphan©600 phantom study, a high-quality image with relative reconstruction error (RRE) less than 3% compared to the full-view result is obtained using UPN with 17% projections (60 views). In the conventional filtered-backprojection reconstruction, the corresponding RRE is more than 15% on the same projection data. The superior performance of ABOCS with the UPN implementation is further demonstrated on the head-and-neck patient. Using 25% projections (91 views), the proposed method reduces the RRE from 21% as in the filtered backprojection (FBP) results to 7.3%. In conclusion, we propose UPN for ABOCS implementation. As compared to GP-BB and the Bregman-type methods, the new method significantly improves the convergence with higher stability and fewer iterations.

Keywords: iterative reconstruction, compressed sensing, total variation, first-order optimization

(Some figures may appear in colour only in the online journal)

1. Introduction

Iterative reconstruction algorithms have been increasingly used in different CT imaging applications (Xia *et al* 2011, Han *et al* 2011, Nett *et al* 2010, Sidky *et al* 2011, Ouyang *et al* 2011, Choi *et al* 2010, Ye *et al* 2011). As compared to the conventional filtered-backprojection (FBP) algorithms, iterative reconstruction shows advantages on image artifact reduction when projection data are heavily undersampled or highly noisy (Han *et al* 2012, Bian *et al* 2010). In this paper, we propose an improved implementation of an iterative CT reconstruction algorithm that was recently developed in our group (Niu and Zhu 2012).

With a linear model, x-ray projections can be approximated as the production of the object linear attenuation coefficient distribution with the system projection matrix. As the measurement data become fewer, the image solution to the linear equation is underdetermined. Recent studies have shown that artifacts from reduced data measurements (such as view aliasing in few-view reconstruction) are effectively suppressed by searching for the solution that has the minimum total variation (TV) and complies with the data fidelity constraints (Tian *et al* 2011, Bian *et al* 2010, Sidky and Pan 2008, Chen *et al* 2010). The success of these TV minimization approaches can be justified by the compressed sensing (CS) theory (Candes *et al* 2006). CS indicates that sparse signals can be well recovered from reduced data measurements by L-1 norm minimization. As the gradient operation sparsifies CT images (Sidky *et al* 2006, Chen *et al* 2010), one embodiment of CS in CT reconstruction is to minimize the L-1 norm of the image gradient, i.e. the TV of the image.

Despite the superior reconstruction performance, TV minimization with data fidelity constraints requires intense computation. A more efficient approach is to convert the problem into a TV-regularization based optimization and to solve using a standard non-negative programming method (Park *et al* 2012). In TV minimization, the image quality is controlled by the data fidelity tolerance ε, while in TV regularization, the main algorithm parameter is the penalty weight λ on the TV term. With proper settings of ε and λ values, these two approaches are equivalent to each other and obtain the same mathematically optimal solution (Pan *et al* 2009). A main disadvantage of TV regularization is that the parameter λ needs to be carefully tuned for different scans, as the optimal λ value is dependent on both the projection noise level and the TV of the true image. On the contrary, the optimal ε value in TV minimization can be accurately estimated directly from the projection data, if noise is dominant in projection errors and the statistics is known (Niu and Zhu 2012, Lauzier and Chen 2012).

To achieve a high computational efficiency while retaining the feature of consistent algorithm parameter settings, we recently proposed a new optimization framework of accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT (CBCT) (Niu and Zhu 2012). In ABOCS, the TV minimization formulation is first converted into a TV regularization based framework with an automatically tuned penalty weight on the data fidelity term. The problem is then efficiently solved using a gradient method. In the published ABOCS implementation, we used gradient projection (GP) with an adaptive Barzilai–Borwein (BB) step-size selection scheme (GP-BB) (Niu and Zhu 2012). Although a high computational efficiency has been demonstrated on preliminary studies, the GP-BB algorithm is empirical and its convergence is not guaranteed (Barzilai and Borwein 1988). In this work, we further improve the ABOCS implementation using a modified Nesterov method (Nesterov 1983). The monotone convergence behavior of Nesterov method, which has been proven in the literature (Jensen *et al* 2011), leads to reduced iterative numbers as well as more reliable stopping criteria. A faster computational speed is also achieved due to the more stable convergence. The method is evaluated on both simulation and phantom studies. We finally present the first results of patient studies using ABOCS reconstruction.

2. Method

2.1. The ABOCS framework

The derivation of ABOCS starts from the TV minimization framework with a data fidelity constraint (Sidky and Pan 2008, Chen *et al* 2008), as shown below:

$$\vec{f}^{*} = \arg\min \|\vec{f}\|_{\mathrm{TV}}, \qquad \text{s.t.,} \quad 0.5\cdot\|M\vec{f}-\vec{b}\|_2^2 \leqslant \varepsilon, \quad f_i \geqslant 0, \tag{1}$$

where the vector $\vec{b}$ with a length of N_d (i.e., number of detector pixels) represents the line integral measurements (i.e., after the logarithmic operation on the raw projections), M is the system matrix modeling the forward projection, $\vec{f}$ is the vectorized patient image to be reconstructed with a length of N_i (i.e., number of image voxels), $\|\bullet\|_2$ calculates the L2-norm in the projection space and $\|\bullet\|_{\mathrm{TV}}$ is the TV term defined as the L1-norm of the spatial gradient image. The user-defined parameter ε is an estimate of total measurement errors.

When statistical noise is dominant in the projection after data correction for deterministic errors, it is possible to estimate ε based on the properties of a known statistical distribution. For example, ε can be accurately estimated from the measured projections, if the measured raw projection obeys Poisson statistics (Zhu *et al* 2009a, Choi *et al* 2010, Wu *et al* 2003, Lauzier and Chen 2012). Effective scatter correction changes the noise variance of the projections, but a precise estimation of ε is still feasible based on the raw data and the scatter-to-primary ratio (SPR) (Zhu *et al* 2009a). As such, the algorithm parameters of equation (1) are consistent for different datasets. Nevertheless, the practical use of TV minimization is hindered by its intense computation due to the fidelity constraint. To accelerate the reconstruction, in ABOCS, we convert equation (1) into a form similar to that of the conventional TV-regularization based algorithm using a logarithmic barrier method: (Lobo *et al* 1998)

$$\vec{f}^{*} = \arg\min\left[\|\vec{f}\|_{\mathrm{TV}} - \log\left(\varepsilon - 0.5\cdot\|M\vec{f}-\vec{b}\|_2^2\right)\right], \quad \text{s.t. } f_i \geqslant 0. \tag{2}$$

Equation (2) is convex (Boyd and Vandenberghe 2004) and therefore can be minimized efficiently using gradient-based algorithms. Note that, equation (2) is equivalent to an optimization problem with the combined TV and data fidelity objective and an automatically adjusted penalty weight on the data fidelity term. Though requiring a slightly longer reconstruction time than the conventional least-square form with a fixed penalty weight, ABOCS uses consistent algorithm parameter values and is therefore more convenient for practical use. More details of the method derivation and its features can be found in Niu and Zhu (2012).

2.2. The unknown-parameter Nesterov algorithm

In (Niu and Zhu 2012), we solve equation (2) using the GP with an adaptive BB step-size selection scheme (Barzilai and Borwein 1988, Zhou *et al* 2006, Park *et al* 2012). In spite of

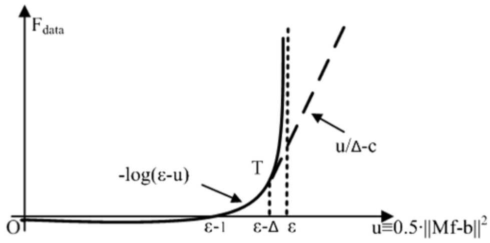

Figure 1. Illustration of the modified logarithmic term (equation (3)).

the high computational efficiency at each iteration, the empirical GP-BB method has a non-monotone convergence behavior, resulting in a sub-optimal performance on reconstruction time (Jensen *et al* 2011). In search of an updating scheme with an improved convergence rate, the Nesterov algorithm becomes a viable option (Nesterov 1983). Among all the first-order approaches, the Nesterov algorithm has been proven to monotonically converge at the optimal rate if the objective function has an Lipschitz continuous gradient (which implies a continuously differentiable function) and is strongly convex (Nesterov 1983).

The Nesterov's method is guaranteed to converge for a continuous and strongly convex objective function (Nesterov 1983). A direct use of the Nesterov algorithm to solve equation (2), however, is hindered by two issues. First, the classic Nesterov algorithm explicitly requires the Lipschitz constant (L) and the strong convex constant (σ), which are used to quantify the continuity and convexity of the objective function, as the input parameters. In iterative CT reconstruction, these two parameter values are difficult to compute due to the large-size and ill-posed system matrix, as well as the singularities in the TV calculation. To circumvent the difficulty in finding the two unknown parameters, we implement the unknown-parameter Nesterov (UPN) method as recently proposed by Jensen *et al* (2011). At each Nesterov iteration, we estimate the L value using a backtracking line search scheme and adjust the σ value using a heuristic formula (Jensen *et al* 2011). The backtracking strategy approximates the Lipschitz constant by iteratively increasing the estimate until the continuity of the objective function is satisfied. The heuristic formula chooses the decremental σ value that satisfies the convexity condition of the objective function in each iteration. Details of the UPN algorithm derivation can be found in Jensen *et al* (2011).

Secondly, the logarithmic term in the objective in equation (2), F_{data}, is not differentiable at the location where $0.5 \cdot \|M\vec{f} - \vec{b}\|_2^2 = \varepsilon$. Therefore, we need to modify the objective function around and outside the singular point of F_{data} without affecting the solution of the optimization framework. In this paper, we use a linear function with a large and finite positive slope to replace the segment of the logarithmic function that has infinite values. Figure 1 shows the linear function (dashed line) and the logarithmic function (solid line). To make the objective continuously differentiable, we design the linear function to be tangent to the log-barrier function at the connection (i.e., the same slope and function value at the tangent point). The modified logarithmic term is thus written as:

$$F_{\text{data}}(u) = \begin{cases} -\log(\varepsilon - u) & \text{if} \quad u \leqslant \varepsilon - \Delta \\ \dfrac{u}{\Delta} - \log(\Delta) - \dfrac{\varepsilon - \Delta}{\Delta} & \text{o.w.,} \end{cases} \tag{3}$$

where $u = 0.5 \cdot \|M\vec{f} - \vec{b}\|_2^2$ and Δ is a user-defined parameter that controls the slope of the linear function.

The resultant objective function becomes:

$$F(\overrightarrow{f}) = \|\overrightarrow{f}\|_{\mathrm{TV}} + \begin{cases} -\log\left(\varepsilon - 0.5\cdot\|M\overrightarrow{f} - \overrightarrow{b}\|_2^2\right), & \text{if} \quad 0.5\cdot\|M\overrightarrow{f} - \overrightarrow{b}\|_2^2 \leqslant \varepsilon - \Delta \\ \dfrac{0.5\cdot\|M\overrightarrow{f} - \overrightarrow{b}\|_2^2}{\Delta} - \log(\Delta) - \dfrac{\varepsilon - \Delta}{\Delta} & \text{o.w.,} \end{cases} \tag{4}$$

and the optimization problem (equation (2)) is re-written as:

$$\overrightarrow{f^*} = \arg\min F(\overrightarrow{f}), \quad \text{s.t. } f_i \geqslant 0. \tag{5}$$

Equation (5) is then solved using the UPN algorithm.

Note that, as seen in figure 1, the linear function is used for solutions outside of the solution pool (i.e. when $0.5\cdot\|M\overrightarrow{f} - \overrightarrow{b}\|_2^2 > \varepsilon - \Delta$). The purpose of the modification is to ensure the objective is differentiable such that the algorithm can rapidly converge to the solution pool. Once the estimated solution satisfies the data fidelity condition (i.e. $0.5\cdot\|M\overrightarrow{f} - \overrightarrow{b}\|_2^2 \leqslant \varepsilon - \Delta$), the linear function does not affect the value of the objective. The objective function essentially becomes the original log-barrier function. As such, the proposed modification improves the convergence without affecting the result of ABOCS reconstruction using the original log-barrier objective.

2.3. Pseudo code for ABOCS with the UPN implementation

In summary, we present the pseudo-code of the UPN approach for ABOCS as below. The following list summarizes the variables in our code:

$N_{\max}$: maximum number of iterations;
L and σ: estimated Lipschitz constant and strong convex constant;
L_0 and σ_0: initial estimates of the Lipschitz constant and strong convex constant;
s_L: magnification factor on L in each step of backtracking;
θ: square root of the ratio between σ and L;
β: intermediate variable in the Nesterov method;
i: index of the detector pixel;
I_0(i): intensity of the flat field at pixel i;
stop: stopping criteria, a constant designated by the user;
μ: user-defined parameter to account for the data error other than Poisson noise;
ε: estimated total noise variance of the line integrals;
Δ : user-defined parameter that controls the slope of the linear function (see figure 1).

The symbol := means assignment. Both image and data space variables are denoted by a vector sign, e.g., $\overrightarrow{f}$ and $\overrightarrow{b}$.

1: $\varepsilon := \mu\cdot\sum_{i=1}^{N_d}\frac{0.5}{I_0(i)\times\exp(-b_i)}$; $N_{\max} := 1000$; $L_0 = 10^3$, $\sigma_0 = 20$, stop $:= -0.999$,
$\theta := \sqrt{\sigma_0/L_0}$, $s_L := 1.3$, $\Delta := 0.02\varepsilon$ *control parameters*
2: $\overrightarrow{h}_0 := \overrightarrow{f}_0 := FBP \quad or \quad ART$ *initial image*
3: for $k = 1, N_{\max}$ do *main loop*
4: $\overrightarrow{d_{\mathrm{TV}}} := \nabla\|\overrightarrow{h}\|_{\mathrm{TV}}$, $\overrightarrow{d_{\mathrm{data}}} := M^T(M\overrightarrow{h} - \overrightarrow{b})$;
5: $\overrightarrow{h_{\mathrm{ind}}} := \begin{cases}1, h_i \neq 0\\ 0, h_i = 0\end{cases}$, $\cos\alpha := \frac{(\overrightarrow{h_{\mathrm{ind}}}\bullet\overrightarrow{d_{\mathrm{TV}}})}{|\overrightarrow{h_{\mathrm{ind}}}\bullet\overrightarrow{d_{\mathrm{TV}}}|}\cdot\frac{(\overrightarrow{h_{\mathrm{ind}}}\bullet\overrightarrow{d_{\mathrm{data}}})}{|\overrightarrow{h_{\mathrm{ind}}}\bullet\overrightarrow{d_{\mathrm{data}}}|}$; (• *denotes element-wise multiplication between two vectors*)
6: $\overrightarrow{f}_{\mathrm{old}} := \overrightarrow{f}$, $\delta h := 0.5\cdot\|M\overrightarrow{h} - \overrightarrow{b}\|_2^2$
7: if $\delta h > \varepsilon - \Delta$, $\lambda := 1/\Delta$; else $\lambda := 1/(\varepsilon - \delta h)$; end if; *calculate the regularization weight outside and inside the feasible solution set, respectively.*

8: $\nabla F(\overrightarrow{f}) := \overrightarrow{d_{\mathrm{TV}}} + \lambda \cdot \overrightarrow{d_{\mathrm{data}}}$; *gradient of the objective as a function of* $\overrightarrow{f}$

9: $\overrightarrow{f} := (\overrightarrow{h} - L^{-1}\nabla F)_+$; *update image and set non-negativity*

10: while $F(\overrightarrow{f}) > F(\overrightarrow{h}) + \nabla F(\overrightarrow{h})^T(\overrightarrow{f} - \overrightarrow{h}) + 0.5L \cdot \|\overrightarrow{f} - \overrightarrow{h}\|_2^2$

11: $L = s_L \cdot L$;

12: $\overrightarrow{f} := (\overrightarrow{h} - L^{-1}\nabla F)_+$;

13: end; *backtracking to find the Lipschitz constant L*

14: $\sigma := \min\left[\sigma_{\mathrm{old}}, \frac{F(\overrightarrow{f_{\mathrm{old}}}) - F(\overrightarrow{h}) - \nabla F(\overrightarrow{h})^T(\overrightarrow{f_{\mathrm{old}}} - \overrightarrow{h})}{0.5 \cdot \|\overrightarrow{f_{\mathrm{old}}} - \overrightarrow{h}\|_2^2}\right]$

15: $\theta_{\mathrm{new}} := 0.5 \cdot \left(\frac{\sigma}{L} - \theta^2 + \sqrt{\left(\frac{\sigma}{L} - \theta^2\right)^2 + 4\theta^2}\right)$

16: $\beta := \theta(1 - \theta)/(\theta^2 + \theta_{\mathrm{new}})$

17: $\overrightarrow{h} := \overrightarrow{f} + \beta\left(\overrightarrow{f} - \overrightarrow{f_{\mathrm{old}}}\right), \theta := \theta_{\mathrm{new}}, \sigma_{\mathrm{old}} := \sigma$;

18: if $\cos\alpha <$ stop && $0.5\|M\overrightarrow{f} - \overrightarrow{b}\|_2^2 \leqslant \varepsilon$ *stopping criteria*

19: return $\overrightarrow{f}$;

20: break;

21: end if;

22: end for;

The parameters in line 1 control the whole algorithm. Their typical values or ranges are shown in the code, which are used to acquire the results in this paper. The initial guess image is generated using the standard FBP reconstruction or the algebraic reconstruction technique (ART), shown in line 2. A zero initial image leads to the same optimal solution but increases the computation time. The main loop, line 3–21, solves the equation (2) using the UPN method with an improved convergence rate. Line 4 and 5 calculate the gradients of TV and data fidelity terms as well as the stopping metric ($\cos(\alpha)$). The formulae of numerical calculations of TV and its gradient can be found in Niu and Zhu (2012). To obtain the gradient of the objective function, we first evaluate the automatic regularization weight based on whether the iteration enters the solution pool or not, shown in line 6 and 7. After computing the gradient of the objective function (line 8), line 9 updates the solution (image) using the reciprocal of L as the step size and sets the non-negativity of the solution (denoted as $(\bullet)_+$). Line 10–13 is a backtracking scheme to find an appropriate L value for the objective function. Line 14 updates the strong convex parameter (σ). Line 15–17 is the standard Nesterov optimal first-order method to update the intermediate variable ($\overrightarrow{f}$). If the stopping criteria are satisfied, the iteration stops and returns an optimal image $\overrightarrow{h}$ (line 19).

Different choices of stopping criteria can be used in our algorithm. In our implementation, we find that the cosine of the angle between the gradient vectors of the data fidelity term and the TV term, $\cos(\alpha)$, accurately indicates the optimality of the solution. According to the Karush Kuhn–Tuchker conditions, at the optimum, the above two vectors must point in exactly opposite directions, i.e., $\cos(\alpha) = -1$ (Sidky and Pan 2008). Specifically, $\cos(\alpha)$ is calculated as:

$$\cos\alpha = \frac{(\overrightarrow{h_{\mathrm{ind}}} \bullet \overrightarrow{d_{\mathrm{TV}}})}{|\overrightarrow{h_{\mathrm{ind}}} \bullet \overrightarrow{d_{\mathrm{TV}}}|} \cdot \frac{(\overrightarrow{h_{\mathrm{ind}}} \bullet \overrightarrow{d_{\mathrm{data}}})}{|\overrightarrow{h_{\mathrm{ind}}} \bullet \overrightarrow{d_{\mathrm{data}}}|}, \tag{6}$$

where $\overrightarrow{d_{\mathrm{TV}}}$ is the gradient of TV term and $\overrightarrow{d_{\mathrm{data}}}$ is the gradient of data fidelity. $\overrightarrow{h_{\mathrm{ind}}}$ is an indicator function with the same dimension of the reconstructed image $\overrightarrow{h}$, defined as: (Sidky and Pan 2008)

$$\overrightarrow{h_{\mathrm{ind}}} = \begin{cases} 1, h_i \neq 0 \\ 0, h_i = 0. \end{cases} \tag{7}$$

In the studies presented in this paper, we used $\cos(\alpha) < -0.999$ as the stopping criterion. Note that, in our previous GP-BB implementation of ABOCS reconstruction, $\cos(\alpha)$ is not used as the stopping criterion due to the instability of the convergence (see the result section for detailed comparisons).

2.4. Evaluation

The performance of the proposed algorithm has been evaluated on a digital Shepp-Logan phantom, the physical Catphan©600 phantom and a head-and-neck patient. The physical phantom data were acquired on our tabletop CBCT system at Georgia Institute of Technology. The geometry of this system exactly matches that of a Varian On-Board Imager CBCT system on the radiation therapy machine. The details of system geometry and hardware parameters can be found in Niu and Zhu (2011). In this study, a fan-beam CT scan, with an illumination width of 1 cm in the vertical direction on the detector, was used to inherently suppress scatter signals. The system was operated in the short-scan mode, with 362 projections covering about 200° of rotation. The same fan-beam geometry was also simulated in the digital phantom study. The reconstructed images have a size of 512 × 512 pixels with 0.5 mm resolution. The projection data on the head-and-neck patient were acquired on the on-board CBCT system installed on the Varian Clinac 23IX radiation therapy machine (Varian Medical System, Palo Alto, CA). The system was operated in the short-scan mode with a bow-tie filter mounted on the outside of the x-ray collimator. 362 projections were acquired in a 195° scan. The reconstructed images have a size of 512 × 512 pixels with 0.684 mm resolution. We limit our investigations on reducing the artifacts from few-view reconstruction. No pre-processing procedures, including scatter and beam-hardening corrections, were applied on the patient projection data. In all the three studies, the detector has 1024 pixels with 0.388 mm resolution and was down-sampled to 512 pixels with 0.776 mm resolution in the reconstruction to avoid large memory consumption.

We compared the performances of UPN, GP-BB and the Bregman operator splitting (BOS) methods (Zhang *et al* 2010, 2011) in the reconstruction of the Shepp-Logan phantom with respect to the image quality and the convergence behavior. BOS is a state-of-the-art optimization algorithm that approximates the solution of a least-square problem and hence eliminates inner iterations required in the classical split Bregman method (Zhang *et al* 2010, 2011). Different noise levels were simulated to test the stability of the proposed algorithm. We then evaluated the ABOCS reconstruction using UPN on the physical phantom and the head-and-neck patient by comparing the results with those of the conventional FBP algorithm.

In the image comparisons, the relative reconstruction error (RRE) of the images compared to the reference, similar to that in (Zhu *et al* 2006), was used as the image quality metric:

$$\mathrm{RRE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - x_{i0})^2}{\sum_{i=1}^{N} x_{i0}^2}} \times 100\%, \tag{8}$$

where x_i is ith pixel the reconstructed image, and x_{i0} is the corresponding pixel in the reference image. The FBP and the iterative reconstruction have different noise levels and artifact patterns even if the projection data are sufficient. Therefore, in the phantom and patient studies, we used the ABOCS reconstruction with full views as the reference, instead of the full-view FBP reconstruction.

3. Result

3.1. Digital phantom study

Besides those used by the classic Nesterov method, the main user-controlled parameter of our algorithm is the Δ which controls the slope of the linear function. Δ depends on the value of data fidelity tolerance ε. A very small Δ value significantly increases the sharpness of the objective function and therefore the number of iterations since the step size is greatly reduced due to the enlarged Lipschitz constant. On the other hand, a large Δ applies a strong weight on the TV regularization term, which makes the algorithm fail to enter the pre-defined solution set. In this paper, we correlate Δ as the percentage of ε and use $\Delta = 0.02\varepsilon$ in the presented studies. We find that, in general, a Δ value in the range of $0.01\varepsilon \sim 0.03\varepsilon$ achieves superior computation performances.

The results on the Shepp-Logan phantom are shown in figure 2. Each iteration takes about 0.3 s using GP-BB and 0.4 s using UPN in MATLAB on a Linux workstation with 6 GB memory and an eight-core Intel Xeon 2.66 GHz CPU. Similar computation time is also found in the physical phantom and patient studies. Two different noise levels (i.e. 5×10^5 and 5×10^4 photons per ray) are simulated to test the algorithm stability. 5×10^5 photons per ray is approximately equivalent to a tube current of 80 mA with a 13 ms pulse width, a standard setting of on-board CBCT systems (Niu and Zhu 2011, 2012). Figures 2(a) and (b) show the ABOCS reconstruction using UPN from 66 projection views and their errors. The algorithm stops at around 200 iterations or less and the RREs for the reconstructed images are 2.0% for figure 2(a1) and 2.3% for figure 2(a2). Note that, the ABOCS reconstruction using GP-BB has a similar image quality at the same objective function value.

Figures 2(c) and (d) show the objective function values and the stopping metric (i.e. $\cos(\alpha)$) at different iterations. The results of UPN are compared with those of GP-BB on the same plot. F^* denotes the optimal objective value obtained by long iterations (i.e. using a more stringent stopping criterion of $\cos(\alpha) < -0.999\,999$). It is seen that the UPN algorithm convergences monotonically towards the optimum, a feature missing in the GP-BB implementation. The convergence of UPN is more stable, especially on the stopping metric, which also results in fewer iterations to reach the same objective function value than GP-BB. After about 130 iterations, the relative difference of objective value of UPN drops below 0.4% and the changes in the reconstructed image are no longer visible at a practical gray scale. In both cases, UPN requires about 50% less iterations than GP-BB to obtain the optimal solution.

To further demonstrate the scientific significance, we also compare the proposed algorithm to a recently developed BOS algorithm (Zhang *et al* 2010, 2011). BOS approximates the solution of a least-square problem and hence eliminates inner iterations required in the classical split Bregman method (Goldstein and Osher 2009). We implement the BOS algorithm for CT reconstruction as a standard least-square problem with a TV regularization term and a carefully-tuned penalty weight. Figure 3 shows the results obtained by the parameter setting that best compromises efficiency and accuracy of the BOS algorithm. Compared to UPN, BOS requires more iterations to converge (see figure 3(c)). On the images reconstructed using the same computation time (figure 3), it is seen that UPN achieves smaller reconstruction errors than BOS. Alternatively, BOS takes more iterations to achieve the same optimal solution of UPN. The higher efficiency of UPN compared to BOS is mainly due to the fact that BOS requires restrictive bound on step sizes according to Lipschitz constant of data fidelity term, i.e. the matrix norm of M^TM (Zhang *et al* 2010). In CT reconstruction, such a step size is too small and causes slow convergence of BOS. On the other hand, UPN adopts Neterov's optimal gradient technique and uses combination of previous two iterations which yields fast

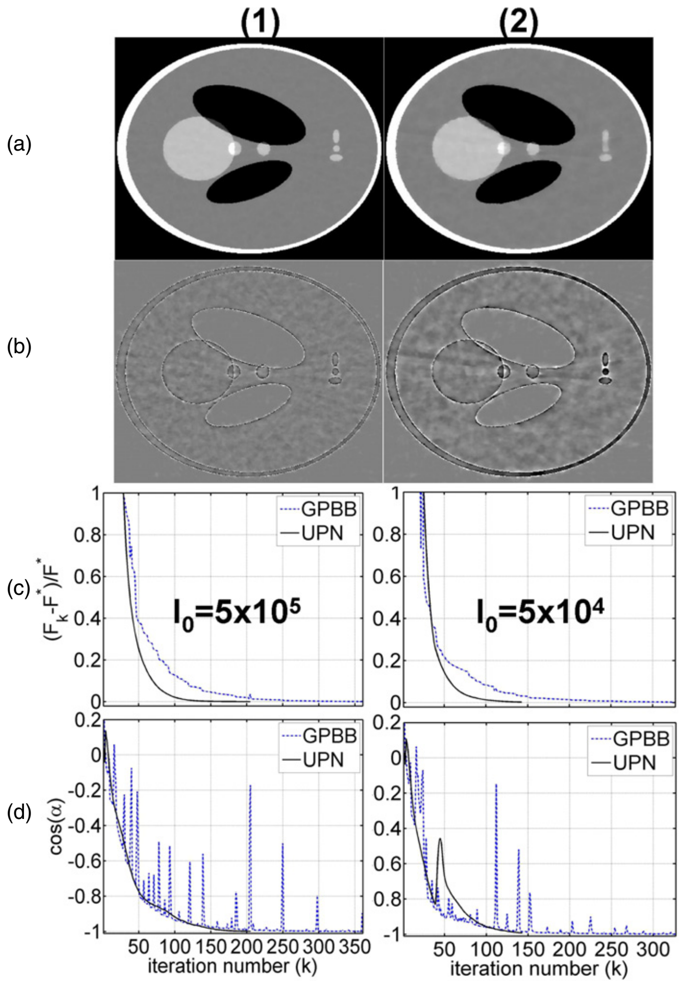

Figure 2. Reconstructed images and plots of the objective function and stopping metric values for GP-BB and UPN on the digital Shepp–Logan phantom. A total of 66 projections are used in the study. Column (1): using a simulated noise level of $I_0 = 5 \times 10^5$ photons per ray; (2): using $I_0 = 5 \times 10^4$. Row (a) reconstructed CT image using ABOCS with the UPN implementation, display window: [−500 500] HU; (b) difference image compared with the ground truth, display window: [−100 100] HU; (c) 1D plot of the objective function values versus the iteration number; (d) 1D plot of the stopping metric versus the iteration number.

convergence (Nesterov 1983). As shown in the simulation study of the Shepp-Logan phantom, UPN achieves the same image quality as those of GP-BB and the Bregman-type method, but reduces the iteration numbers by up to 50% and 90%, respectively.

The reduction of iteration number leads to improvement of computation time. Table 1 compares the computation time for reconstructed images shown in figures 2–5 using UPN, BOS and GP-BB methods. Note that, we terminate the reconstruction of UPN and GP-BB using the same stopping criteria. It is seen that UPN requires less computation time as compared to GP-BB. For a better comparison of UPN and BOS on image qualities, we generate the results of BOS reconstruction using the same computation time as that of UPN. As shown in figure 3,

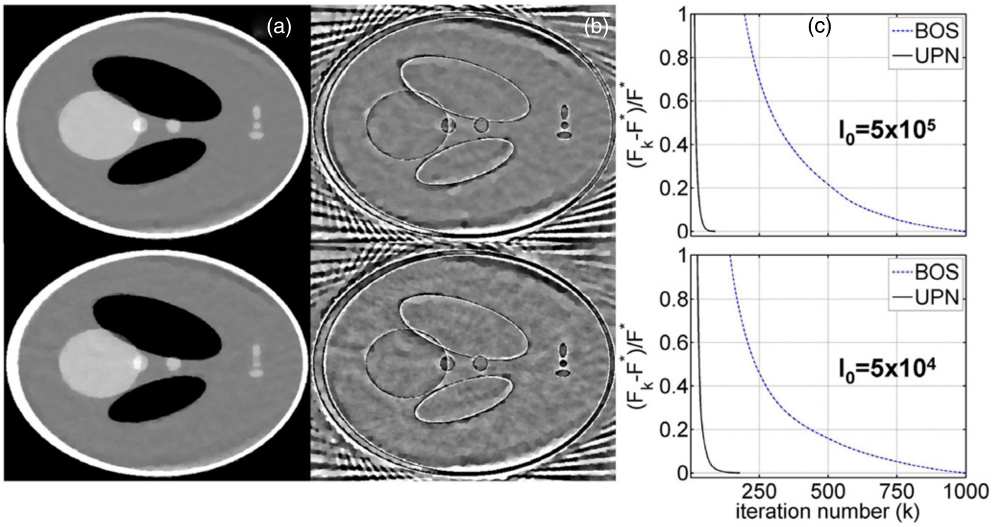

Figure 3. Comparison between BOS and UPN algorithms on the digital Shepp-Logan phantom. A total of 66 projections are used in the study. Row (1): using a simulated noise level of $I_0 = 5 \times 10^5$ photons per ray; (2): using $I_0 = 5 \times 10^4$. Column (a): reconstructed CT image using BOS with the same computation time as required to obtain the UPN results shown in figure 2(a) (100 s for the first row and 57 s for the second row), display window: [−500 500] HU; (b) difference image compared with the ground truth, display window: [−100 100] HU; (c) 1D plot of the objective function values versus the iteration number.

Table 1. Comparison of the computation time (in seconds) for reconstructions shown in figures 2–5 using UPN, BOS and GP-BB algorithms.

	Algorithm	
Data	UPN/BOS	GP-BB
Shepp–Logan $I_0 = 5 \times 10^5$	69	84
Shepp–Logan $I_0 = 5 \times 10^4$	62	110
Catphan© 600	38	47
Head patient	74	125

UPN obtains more accurate reconstruction than BOS with the same computation time. Similar performances of computation time are observed in the following physical phantom and patient studies as well.

3.2. Catphan©600 phantom study

We further evaluate the proposed algorithm on the physical phantom. Figure 4 shows the results on the Catphan©600 phantom. With 60 projection views (about 17% of the total 362 projections from a short scan), the FBP reconstruction has severe view-aliasing artifacts and the RRE compared to the full-view reconstruction is about 15% (see figures 4(a) and (b)). The ABOCS reconstruction using UPN achieves a comparable full-view image quality (figure 4(c)) as that of FBP, and also effectively suppresses the view-aliasing artifacts when the projection views are reduced to 60 (see figure 4(d)). The RRE of the few-view reconstruction is 2.4%.

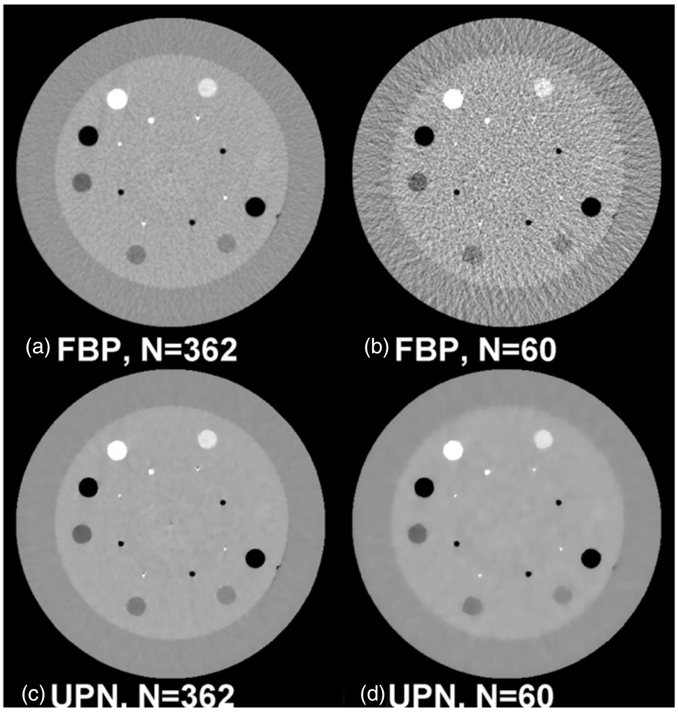

Figure 4. Reconstructed Catphan©600 images from a 200° short-scan mode. (a) FBP: using 362 views; (b) FBP: using 60 views (∼17% of the total 362); (c) ABOCS UPN: using 362 views; (d) ABOCS UPN: using 60 views. Window level: [−600 400] HU.

3.3. Head-and-neck patient study

This head-and-neck patient study well presents the challenges of iterative CT reconstruction in clinical environments. Besides statistical noise, the projections contain considerable errors from non-ideal effects, including photon scatter, beam hardening and patient motion. Nevertheless, the proposed algorithm still significantly improves the FBP reconstruction. The comparison of reconstructed images is shown in figure 5. With 91 views, the FBP reconstruction has an RRE of over 21%. Our algorithm suppresses the streaks from view aliasing and reduces the RRE to 7.3%.

4. Discussion

In this paper, we propose the UPN algorithm for iterative CT reconstruction using the ABOCS optimization framework. As compared to the previous GP-BB algorithm, the new method significantly improves the convergence stability and reduces the iteration number. The fundamental reason for the improvement is that the GP method converges at a rate inversely proportional to the number of iterations ($1/k$, where k is iteration number (Jensen *et al* 2011)), while the Nesterov approach achieves an optimal convergence rate of ($1/k^2$) (Jensen *et al* 2011, Nesterov 1983).

We modify the objective function to make the UPN algorithm implementable on ABOCS. The main parameter of the modified objective is Δ, which specifies the starting point of the linear function to avoid the singularity of the original logarithmic function. A small Δ results

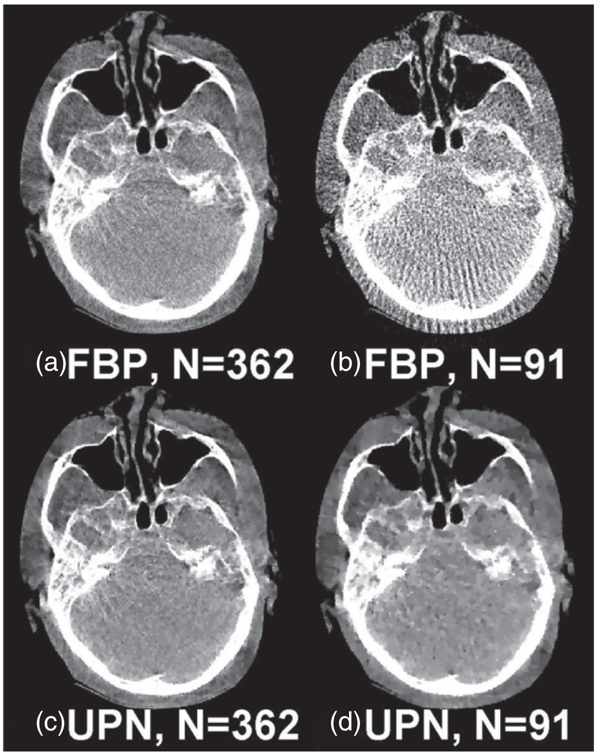

Figure 5. Reconstructed head patient images from a 195° short-scan mode. (a) FBP: using 362 views; (b) FBP: using 91 views (∼25% of the total 362); (c) ABOCS UPN: using 362 views; (d) ABOCS UPN: using 91 views. Window level: [−550 450] HU.

in a large Lypschitz constant and a small step size $(1/L)$, leading to a slow convergence. On the other hand, a large Δ gives larger step size and but may cause instability during the iteration. In our study, we find that the algorithm is not very sensitive to this parameter in a practical range.

Though demonstrated on 2D reconstructions, the proposed method is readily extendable to 3D cone-beam CT reconstruction. To overcome the impediments of large memory use and intensive computation in 3D reconstruction, we will implement the UPN algorithm based on the hardware acceleration techniques including graphics-processing-unit (GPU) (Jia *et al* 2010) or cloud computing (Meng *et al* 2011). The feasibility of the GPU implementation of ABOCS reconstruction has been demonstrated in our previous study (Niu and Zhu 2012). We will further evaluate the performance of the ABOCS reconstruction using UPN on more patient studies. Data correction procedures, such as scatter correction (Niu *et al* 2010, 2012, Zhu *et al* 2009b) and noise suppression (Zhu *et al* 2009a, Wang *et al* 2006, Tang and Tang 2012), will be applied to make our method more valuable in clinical applications. Dose evaluations are important to make studies of iterative reconstruction more complete. We will perform future studies to investigate the relationship between the image quality and imaging dose in a similar way as in Yan *et al* (2012), Leipsic *et al* (2010) and Marin *et al* (2010).

5. Conclusion

We propose UPN for the ABOCS iterative CT reconstruction. The algorithm has been evaluated on both simulated and physical phantoms. As compared to the previous GP-BB and the state-of-the-art BOS algorithm, the new method significantly improves the convergence with higher stability and reduces the total iterations by more than 50% and 90%, respectively. As demonstrated in both phantom and patient studies, the proposed algorithm effectively suppresses view-aliasing artifacts in few-view reconstruction.

Acknowledgments

This work is partially supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R21EB012700 and a Varian MRA grant.

References

Barzilai J and Borwein J M 1988 2-point step size gradient methods *IMA J. Numer. Anal.* **8** 141–8

Bian J G, Siewerdsen J H, Han X A, Sidky E Y, Prince J L, Pelizzari C A and Pan X C 2010 Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT *Phys. Med. Biol.* **55** 6575–99

Boyd S P and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press) (doi:10.1017/CBO9780511804441)

Candes E J, Romberg J and Tao T 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information *IEEE Trans. Inform. Theory* **52** 489–509

Chen G H, Tang J and Leng S H 2008 Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets *Med. Phys.* **35** 660–3

Chen G H, Tang J, Nett B, Qi Z H, Leng S A and Szczykutowicz T 2010 Prior image constrained compressed sensing (PICCS) and applications in x-ray computed tomography *Curr. Med. Imaging Rev.* **6** 119–34

Choi K, Wang J, Zhu L, Suh T S, Boyd S and Xing L 2010 Compressed sensing based cone-beam computed tomography reconstruction with a first-order method *Med. Phys.* **37** 5113–25

Feldkamp L A, Davis L C and Kress J W 1984 Practical cone-beam algorithm *J. Opt. Soc. Am.* A **1** 612–9

Goldstein T and Osher S 2009 The split Bregman method for L1-regularized problems *SIAM J Imaging Sci.* **2** 323–43

Han X A, Bian J G, Eaker D R, Kline T L, Sidky E Y, Ritman E L and Pan X C 2011 Algorithm-enabled low-dose micro-CT imaging *IEEE Trans. Med. Imaging* **30** 606–20

Han X A, Bian J G, Ritman E L, Sidky E Y and Pan X C 2012 Optimization-based reconstruction of sparse images from few-view projections *Phys. Med. Biol.* **57** 5245–73

Jensen T L, Jorgensen J H, Hansen P C and Jensen S H 2011 Implementation of an optimal first-order method for strongly convex total variation regularization *BIT Numer. Methods* **51** 1–28

Jia X, Lou Y F, Li R J, Song W Y and Jiang S B 2010 GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation *Med. Phys.* **37** 1757–60

Lauzier P T and Chen G H 2012 Characterization of statistical prior image constrained compressed sensing: I. Applications to time-resolved contrast-enhanced CT *Med. Phys.* **39** 5930–48

Leipsic J, Nguyen G, Brown J, Sin D and Mayo J R 2010 A prospective evaluation of dose reduction and image quality in chest CT using adaptive statistical iterative reconstruction *Am. J. Roentgenol.* **195** 1095–9

Lobo M S, Vandenberghe L, Boyd S and Lebret H 1998 Applications of second-order cone programming *Linear Algebra Appl.* **284** 193–228

Marin D, Nelson R C, Schindera S T, Richard S, Youngblood R S, Yoshizumi T T and Samei E 2010 Low-tube-voltage, high-tube-current multidetector abdominal CT: improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm-initial clinical experience *Radiology* **254** 145–53

Meng B, Pratx G and Xing L 2011 Ultrafast and scalable cone-beam CT reconstruction using MapReduce in a cloud computing environment *Med. Phys.* **38** 6603–9

Nesterov Y 1983 A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$ *Sov. Math. Dokl.* **27** 372–6

Nett B E, Brauweiler R, Kalender W, Rowley H and Chen G H 2010 Perfusion measurements by micro-CT using prior image constrained compressed sensing (PICCS): initial phantom results *Phys. Med. Biol.* **55** 2333–50

Niu T Y, Al-Basheer A and Zhu L 2012 Quantitative cone-beam CT imaging in radiation therapy using planning CT as a prior: first patient studies *Med. Phys.* **39** 1991–2001

Niu T Y, Sun M S, Star-Lack J, Gao H W, Fan Q Y and Zhu L 2010 Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images *Med. Phys.* **37** 5395–406

Niu T Y and Zhu L 2011 Scatter correction for full-fan volumetric CT using a stationary beam blocker in a single full scan *Med. Phys.* **38** 6027–38

Niu T Y and Zhu L 2012 Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: phantom studies *Med. Phys.* **39** 4588–98

Ouyang L, Solberg T and Wang J 2011 Effects of the penalty on the penalized weighted least-squares image reconstruction for low-dose CBCT *Phys. Med. Biol.* **56** 5535–52

Pan X C, Sidky E Y and Vannier M 2009 Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse Problems* **25** 123009

Park J C, Song B Y, Kim J S, Park S H, Kim H K, Liu Z W, Suh T S and Song W Y 2012 Fast compressed sensing-based CBCT reconstruction using Barzilai–Borwein formulation for application to on-line IGRT *Med. Phys.* **39** 1207–17

Sidky E Y, Duchin Y, Pan X C and Ullberg C 2011 A constrained, total-variation minimization algorithm for low-intensity x-ray CT *Med. Phys.* **38** S117–25

Sidky E Y, Kao C M and Pan X H 2006 Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT *J. X-Ray Sci. Technol.* **14** 119–39 http://iospress.metapress.com/content/1jduv1cll3f9e2br

Sidky E Y and Pan X C 2008 Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization *Phys. Med. Biol.* **53** 4777–807

Tang S J and Tang X Y 2012 Statistical CT noise reduction with multiscale decomposition and penalized weighted least squares in the projection domain *Med. Phys.* **39** 5498–512

Tian Z, Jia X, Yuan K H, Pan T S and Jiang S B 2011 Low-dose CT reconstruction via edge-preserving total variation regularization *Phys. Med. Biol.* **56** 5949–67

Wang J, Li T, Lu H B and Liang Z R 2006 Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography *IEEE Trans. Med. Imaging* **25** 1272–83

Wu T *et al* 2003 Tomographic mammography using a limited number of low-dose cone-beam projection images *Med. Phys.* **30** 365–80

Xia D, Xiao X, Bian J, Han X, Sidky E Y, De Carlo F and Pan X 2011 Image reconstruction from sparse data in synchrotron-radiation-based microtomography *Rev. Sci. Instrum.* **82** 043706

Yan H, Cervino L, Jia X and Jiang S B 2012 A comprehensive study on the relationship between the image quality and imaging dose in low-dose cone beam CT *Phys. Med. Biol.* **57** 2063–80

Ye X J, Chen Y M and Huang F 2011 Computational acceleration for MR image reconstruction in partially parallel imaging *IEEE Trans. Med. Imaging* **30** 1055–63

Zhang X Q, Burger M, Bresson X and Osher S 2010 Bregmanized nonlocal regularization for deconvolution and sparse reconstruction *SIAM J. Imaging Sci.* **3** 253–76

Zhang X Q, Burger M and Osher S 2011 A unified primal-dual algorithm framework based on Bregman iteration *J. Sci. Comput.* **46** 20–46

Zhou B, Gao L and Dai Y H 2006 Gradient methods with adaptive step-sizes *Comput. Optim. Appl.* **35** 69–86

Zhu L, Bennett N R and Fahrig R 2006 Scatter correction method for x-ray CT using primary modulation: theory and preliminary results *IEEE Trans. Med. Imaging* **25** 1573–87

Zhu L, Wang J and Xing L 2009a Noise suppression in scatter correction for cone-beam CT *Med. Phys.* **36** 741–52

Zhu L, Xie Y Q, Wang J and Xing L 2009b Scatter correction for cone-beam CT in radiation therapy *Med. Phys.* **36** 2258–68