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

Phys. Med. Biol. **60** (2015) 1965–1987 doi:10.1088/0031-9155/60/5/1965

Iterative CBCT reconstruction using Hessian penalty

Tao Sun[1], Nanbo Sun[1], Jing Wang[2] and Shan Tan[1]

[1] Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
[2] Department of Radiation Oncology, The University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390, USA

E-mail: shantan@mail.hust.edu.cn and Jing.Wang@UTSouthwestern.edu

Received 13 July 2014, revised 12 November 2014
Accepted for publication 3 December 2014
Published 12 February 2015

Abstract

Statistical iterative reconstruction algorithms have shown potential to improve cone-beam CT (CBCT) image quality. Most iterative reconstruction algorithms utilize prior knowledge as a penalty term in the objective function. The penalty term greatly affects the performance of a reconstruction algorithm. The total variation (TV) penalty has demonstrated great ability in suppressing noise and improving image quality. However, calculated from the first-order derivatives, the TV penalty leads to the well-known staircase effect, which sometimes makes the reconstructed images oversharpen and unnatural. In this study, we proposed to use a second-order derivative penalty that involves the Frobenius norm of the Hessian matrix of an image for CBCT reconstruction. The second-order penalty retains some of the most favorable properties of the TV penalty like convexity, homogeneity, and rotation and translation invariance, and has a better ability in preserving the structures of gradual transition in the reconstructed images. An effective algorithm was developed to minimize the objective function with the majorization-minimization (MM) approach. The experiments on a digital phantom and two physical phantoms demonstrated the priority of the proposed penalty, particularly in suppressing the staircase effect of the TV penalty.

Keywords: cone-beam CT, iterative reconstruction, Hessian regularization, total variation penalty, majorization–minimization approach

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

 © 2015 Institute of Physics and Engineering in Medicine Printed in the UK

1. Introduction

There is growing interest in using the on-board cone-beam computed tomography (CBCT) for image-guided radiation therapy (Jaffray *et al* 2002). A great concern is that the excessive radiation dose due to its repeated use during a treatment course poses potential risk to the health of patients (Islam *et al* 2006, Brenner and Hall 2007). One way to reduce radiation dose is to lower mAs level in CT projection data acquisition. However, excessive quantum noise leads to degraded reconstructed CBCT images (Wang *et al* 2008a), which renders low-mAs CBCT a less attractive option for the therapeutic guidance.

Statistical iterative reconstruction (SIR) algorithms have shown superior performance in improving image quality (Ouyang *et al* 2011). The choice of the penalty term in the objective function greatly affects the quality of the reconstructed images. In the general maximum *a posterior* (MAP) estimate in inverse problems like CBCT reconstruction, the penalty term imposes prior knowledge on the desired results. The Markov random field (MRF) (Geman and Geman 2009) has been widely used and can derive a variety of distinct penalties by choosing different potential functions. One common model is the Gaussian Markov random field (GMRF) (Hanson and Wecksung 1983). However, the Gaussian prior penalty results in excessive noisy or blurred images. Some non-Gaussian distribution function have been used such as the Huber function (Chlewicki *et al* 2004), the log-cosh function (Green 1990) and the Gibbs functions (Lange 1990). Thibault *et al* proposed the q-generalized Gaussian Markov random field (q-GGMRF) penalty (Thibault *et al* 2007) which generalizes many existing penalties. Other commonly used penalties include the isotropic penalty (Fessler 1994), the anisotropic penalty (Wang *et al* 2008b) and the total variation (TV) penalty (Sidky *et al* 2006, Jia *et al* 2010), etc.

The TV penalty has shown the state-of-the-art performance in suppressing noises and preserving edges (Sidky *et al* 2006, Jia *et al* 2010). Besides, it has been used as the sparsifying transform function in many compress sensing-based reconstruction algorithms (Choi *et al* 2010, Lee *et al* 2012, Park *et al* 2012). The success of the TV penalty lies in its good properties such as simplicity, convexity, and rotation and shift invariance (Boyd and Vandenberghe 2004). These properties make it possible to derive various efficient algorithms. Despite its success, the TV penalty often leads to the well-known staircase effect (Chan *et al* 2000, Yuan *et al* 2009). The reason is that the TV penalty favors minimizing the first-order derivatives and thus tends to have piecewise constant results (Yuan *et al* 2009). This tendency can be highly undesirable in biomedical imaging where the intensity of neighboring tissues usually has gradual transitioning. The staircase effect in the biomedical image may falsely reflect the true structure of the tissues and leads to incorrect interpretation of the images.

In this study, we proposed to use the Hessian penalty for CBCT reconstruction that involves the Frobenius norm of the Hessian matrix of an image to suppress the staircase effect observed in the TV penalty. The Hessian penalty is a second-order derivative penalty. The motivation behind is that the second-order derivatives have a weaker penalty on the absolute difference between neighboring voxels and allow for piecewise smooth reconstructed results, thus better reflecting the natural intensity transition in CBCT images. The Hessian penalty retains most favorable properties of TV such as convexity, homogeneity, rotation and translation invariance (Boyd and Vandenberghe 2004), and have shown improved performance in suppressing the staircase effect and enhancing the image quality for image denoising (Chan *et al* 2000) and microscopy image debluring (Lefkimmiatis *et al* 2012). In previous CBCT reconstruction methods (Fessler 1994, Wang *et al* 2008b), most of the penalty terms are in quadratic form by themselves, and thus the optimization of the associated objective functions can be directly performed by algorithms such as Gauss-Seidel update and conjugate gradient descent. Unfortunately, the Hessian penalty is not

quadratic, which makes the minimization of the objective function challenging. We developed an effective algorithm to minimize the objective function with the majorization-minimization (MM) approach. The objective function was upper-bounded by a sequence of quadratic majorizers, which can be minimized directly using the Gauss-Seidel update strategy. In this way, the optimization problem of the original objective function was converted to a sequence of simple optimization problems. We conducted a systematic comparison on a simulated phantom and two physical phantoms. The results demonstrated that the proposed penalty outperforms the TV penalty, particularly in suppressing the staircase effect. To the best of our knowledge, this is the first time that the Hessian penalty was used in CBCT reconstruction.

2. Mathematical models

2.1. PWLS image reconstruction

The model of x-ray CT data acquisition is based on the line integration along the x-ray path of the tissue attenuation, which can be calculated through the logarithm transform of the ratio of incident and detected photon numbers

$$\widehat{p}_l = \ln\frac{I_{l0}}{I_l} = \int_l \mu(x, y, z)\,\mathrm{d}l, \tag{1}$$

where l is the x-ray path, $\widehat{p}_l$ is the projection datum corresponding to the path and μ is the linear attenuation coefficient image. The reconstruction is aimed at estimating μ from a series of line integral values, i.e. the observed projection data $\widehat{p}$. The statistical noise property of the x-ray CT projection data after logarithm transform approximately follows a Gaussian distribution (Wang *et al* 2008). The variance of the noise at the detector bin i can be estimated as follows:

$$\sigma_i^2 = \exp(\overline{p}_i)/N_{i0}, \tag{2}$$

where N_{i0} is the incident photon number, and $\overline{p}_i$ and σ_i^2 are the mean and variance of projection datum, all at the detector bin i. The reconstruction can be formulated as a minimization problem with the objective function (Wang *et al* 2008b)

$$\Phi(\mu) = \frac{1}{2}(\widehat{p} - A\mu)^T \Sigma^{-1}(\widehat{p} - A\mu) + \beta R(\mu), \tag{3}$$

where Σ is a diagonal matrix with diagonal element σ_i^2, and A is the projection matrix, in which the element A_{ij} is the intersection length of the ith projection with the image voxel μ_j. The first term in equation (3) is the weighted least-squares (WLS) criterion or a data-fidelity term. The elements of the diagonal matrix Σ play the role of weighting in the WLS cost function and determine the contribution of each measurement. The second term is a prior constraint or a penalty term, where β controls the trade-off between the data-fidelity term and the penalty term. The image reconstruction task is to find an attenuation coefficient image μ by minimizing the objective function with a positive constraint. The method for CBCT reconstruction using equation (3) is referred to as the penalized weighted least-square (PWLS) criterion (Wang *et al* 2008b).

2.2. Hessian penalty

The prior constraint in equation (3) enforces a smoothness penalty on the solution. Minimization of the total variation (TV) of an image has shown good properties in noise suppressing and

edge preserving for CBCT reconstruction (Wang *et al* 2008). Let μ be a continuously differentiable 3D image. The TV penalty is defined as

$$R_{\mathrm{TV}}(\mu) = \int_{\Omega} \| \nabla \mu \|_2 \, \mathrm{d}\mathbf{x}, \tag{4}$$

where $\Omega \subset \mathbb{R}^3$, $\nabla \mu = (\mu_x, \mu_y, \mu_z)^T$.

The TV penalty usually leads to the annoying staircase effect. In this study, we proposed to use a new penalty that has good properties in noise suppressing and edge preserving while being able to overcome the staircase effect of the TV penalty. Our motivation was based on the consideration that the first-order derivative in the TV penalty tends to minimize the difference between neighboring voxels, thus leading to the piecewise constant images. Instead of using the first-order derivatives, we chose to utilize the second-order derivatives in the penalty term. The proposed Hessian penalty for CBCT reconstruction can be defined as (Lefkimmiatis *et al* 2012)

$$R_{\mathrm{Hessian}}(\mu) = \int_{\Omega} \| H_{\mu}(\mathbf{x}) \|_F \, \mathrm{d}\mathbf{x}, \tag{5}$$

where

$$H_{\mu}(\mathbf{x}) = \begin{pmatrix} \mu_{xx}(\mathbf{x}) & \mu_{xy}(\mathbf{x}) & \mu_{xz}(\mathbf{x}) \\ \mu_{yx}(\mathbf{x}) & \mu_{yy}(\mathbf{x}) & \mu_{yz}(\mathbf{x}) \\ \mu_{zx}(\mathbf{x}) & \mu_{zy}(\mathbf{x}) & \mu_{zz}(\mathbf{x}) \end{pmatrix} \tag{6}$$

and the symbol $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, which is a natural extension of the l_2 norm in the TV penalty. This Hessian penalty satisfies good properties such as convexity, homogeneity, and rotation and translation invariance (Boyd and Vandenberghe 2004). The convexity arises from the integration of the norm of a linear operator, and the rotation invariance arises from the fact that Hessian matrix norm is fully determined by its eigenvalues (Boyd and Vandenberghe 2004). In addition, the following equation holds for the Hessian penalty (Lefkimmiatis *et al* 2012)

$$\| H_{\mu}(\mathbf{x}) \|_F = \frac{3}{4\pi} \| D^2_{\mathbf{u},\mathbf{v}} \mu(\mathbf{x}) \|_{L_{2(S)}}, \tag{7}$$

where $\mathbf{u}$ and $\mathbf{v}$ are unit-norm vectors represented in spherical coordinate as

$$\mathbf{n} = (\sin\theta \, \cos\varphi, \ \sin\theta \, \sin\varphi, \ \cos\theta), \ \theta \in [0, \pi], \ \varphi \in [0, 2\pi), \tag{8}$$

$S = [0, \pi] \times [0, 2\pi]$. $\mathbf{n}$ is $\mathbf{u}$ or $\mathbf{v}$. $\mathrm{D}^2_{\mathbf{u},\mathbf{v}}$ is the second directional derivative in the directions specified by the vectors, and defined as $\mathrm{D}^2_{\mathbf{u},\mathbf{v}}\mu(\mathbf{x}) = \mathrm{D}^1_{\mathbf{u}}(D^1_{\mathbf{v}}\mu)(\mathbf{x}) = \mathbf{u}^T H_{\mu}(\mathbf{x})\,\mathbf{v}$, where $\mathrm{D}^1_{\mathbf{u}}$ and $\mathrm{D}^1_{\mathbf{v}}$ are the first-order derivative in the direction specified by $\mathbf{u}$ and $\mathbf{v}$, respectively. This equation indicates the translation invariance property. The excellent properties of the Hessian penalty enable us to derive efficient optimization algorithms for CBCT reconstruction.

2.3. Separable footprints forward projection

In the previous section, the element A_{ij} of the projection matrix A is defined as the length of the intersection of the projection ray from the x-ray source to the ith detector bin with the image voxel μ_j. The element in the projection matrix can be calculated according to the Siddon's algorithm (Siddon 1985). This algorithm is easy to be understood, and has been widely used

for SIR CBCT reconstruction in the literature (Jia *et al* 2011, Ouyang *et al* 2011). However, this line intersection model is imprecise and computationally complicated for our pixel-wise update strategy. In this study, we adopted a more precise and more efficient forward projection algorithm instead. The main idea of the separable footprints (SF) projection algorithm is that the true voxel footprint functions can be approximated by 2D separable functions (Long *et al* 2010). Specifically we chose SF-TR projector which uses trapezoid functions in the transaxial direction and rectangular functions in the axial direction. The element $A(s,t,\gamma;x,y,z)$ of the projection matrix A can be mathematically represented as

$$A(s,t,\gamma;x,y,z) = v(s,t,\gamma)\cdot u(\gamma;x,y)\cdot F_1(s,\gamma;x,y)\cdot F_2(t,\gamma;x,y,z)\,, \tag{9}$$

where s and t are the transaxial and axial coordinate in the detector plane, respectively, γ is the polar angle of the projection, (x,y,z) is the coordinate of the pixel, F_1 is a trapezoid function in the transaxial direction, and F_2 is a rectangle function in the axial direction. The footprint functions F_1 and F_2 approximate the shape of the true footprint. The factors $v(s,t,\gamma)$ and $u(\gamma;x,y)$ are simple amplitude functions. Readers who are interested in the details can refer to Long *et al* (2010).

3. Miniminaztion of objective functions

3.1. Derivative filters

In discrete case, the TV penalty in equation (4) can be written as

$$R_{\mathrm{TV}}(\mu) = \sum_{x,y,z}\sqrt{\mu_x^2(x,y,z)+\mu_y^2(x,y,z)+\mu_z^2(x,y,z)}\ . \tag{10}$$

The proposed Hessian penalty in equation (5) can be written as

$$R_{\mathrm{Hessian}}(\mu) = \sum_{x,y,z}\sqrt{\begin{array}{c}\mu_{xx}^2(x,y,z)+\mu_{yy}^2(x,y,z)+\mu_{zz}^2(x,y,z)+2\mu_{xy}^2(x,y,z)\\ +2\mu_{xz}^2(x,y,z)+2\mu_{yz}^2(x,y,z)\end{array}} \tag{11}$$

which can be implemented using the discrete filtering (Arigovindan *et al* 2013). The content under the root symbol in equation (11) consists of the sum of six different quadratic terms. Let us denote the index of a voxel by the triplet (x,y,z) and all possible filters yielding second-order derivatives as $L_i(x,y,z)\,;\ i=1,2,\cdots,6$. Their exact forms can be found in the appendix A. Equation (11) can be equally stated as

$$R_{\mathrm{Hessian}}(\mu) = \sum_{x,y,z}\sqrt{\sum_{i=1}^{6}[L_i(x,y,z)*\mu(x,y,z)\ ]^2}\ . \tag{12}$$

This form of penalty is equivalent to equation (11) but more concise and consistent with that of the TV penalty in equation (10). It also makes it easier for the derivation and description of the following minimization approach.

3.2. Majorization–minimization approach

To minimize the objective function, we adopted the majorization-minimization (MM) approach (Hunter and Lange 2004). The main idea of the approach is to obtain a sequence of quadratic functions that upper-bound the original objective function and converge to its optimum.

For the minimization problem

$$\mu^* = \arg \min_{\mu} \Phi(\mu), \tag{13}$$

with a sequence of independent variable $\{\mu^{(0)}, \mu^{(1)}, \cdots, \mu^{(t)}, \mu^{(t+1)}, \cdots\}$, we take its majorizer $Q(\mu; \mu^{(t)})$ at a fixed point $\mu^{(t)}$, which satisfies the following two conditions

$$\begin{aligned} Q(\mu; \mu^{(t)}) &\geq \Phi(\mu), \forall \mu \\ Q(\mu^{(t)}; \mu^{(t)}) &= \Phi(\mu^{(t)}) \end{aligned}. \tag{14}$$

Supposing that $\mu^{(0)}$ is a given initial point and $\mu^{(t+1)}$ is obtained by

$$\mu^{(t+1)} = \arg \min_{\mu} Q(\mu; \mu^{(t)}), \tag{15}$$

we then have

$$\Phi(\mu^{(t+1)}) \leq Q(\mu^{(t+1)}; \mu^{(t)}) \leq Q(\mu^{(t)}; \mu^{(t)}) = \Phi(\mu^{(t)}). \tag{16}$$

This inequality forms the foundation of the majorization-minimization (MM) approach (Hunter and Lange 2004).

Instead of minimizing the original objective function $\Phi(\mu)$, we first upper-bounded it by a sequence of majorizers $Q(\mu; \mu^{(t)})$, where $\mu^{(0)}$ is a given initial point and $\mu^{(t+1)}$ is obtained by solving the optimizing problem in equation (15). Based on the inequality in equation (14), $\Phi(\mu)$ is monotonically non-increasing on the set $\{\mu^{(0)}, \mu^{(1)}, \cdots, \mu^{(t)}, \mu^{(t+1)}, \cdots\}$. $\Phi(\mu^{(t)})$ converges monotonically to $\Phi = \Phi(\mu^*)$ for a stationary point μ^* of $\Phi(\mu)$ under the condition that $Q(\mu; \mu')$ is continuous in both μ and μ'. In addition, $\mu^{(t)}$ converges to the global minimum of $\Phi(\mu)$ if $\Phi(\mu)$ is strictly convex. The proof of these converging properties of MM approach can be found in Wu (1983). Since in this study these majorizers can be simply chosen as quadratic forms, the optimization problem of the original function can be converted to a sequence of simple optimization problems.

3.3. Quadratic majorizers of penalty functions

We first developed a quadratic majorizer of the TV penalty (10) and the proposed Hessian penalty (12). We employed the inequality holding for a general function

$$\sqrt{g(x)} \leq \frac{\sqrt{g(y)}}{2} + \frac{\sqrt{g(x)}}{2\sqrt{g(y)}}, \forall x, \forall y : g(y) \neq 0, \tag{17}$$

with equality if and only if $g(x) = g(y)$. Using the fact that the inequality relation is closed under the formation of sums and non-negative products, we then obtained that the function

$$\begin{aligned} Q_{R_{\mathrm{TV}}}(\mu; \mu^{(t)}) = \sum_{x,y,z} \Bigg\{ & \frac{1}{2}\sqrt{\left[\mu_x^{(t)}(x,y,z)\right]^2 + \left[\mu_y^{(t)}(x,y,z)\right]^2 + \left[\mu_z^{(t)}(x,y,z)\right]^2} \\ & + \frac{\mu_x^2(x,y,z) + \mu_y^2(x,y,z) + \mu_z^2(x,y,z)}{2\sqrt{\left[\mu_x^{(t)}(x,y,z)\right]^2 + \left[\mu_y^{(t)}(x,y,z)\right]^2 + \left[\mu_z^{(t)}(x,y,z)\right]^2}} \Bigg\} = \frac{1}{2}\mu^T R_{TV}^{(t)} \mu + \text{const} \end{aligned} \tag{18}$$

and

$$Q_{R_{\text{Hessian}}}(\mu;\mu^{(t)}) = \sum_{x,y,z}\left\{\frac{1}{2}\sqrt{\sum_{i=1}^{6}[L_i(x,y,z)*\mu^{(t)}(x,y,z)]^2} + \frac{\sum_{i=1}^{6}[L_i(x,y,z)*\mu(x,y,z)]^2}{2\sqrt{\sum_{i=1}^{6}[L_i(x,y,z)*\mu^{(t)}(x,y,z)]^2}}\right\} = \frac{1}{2}\mu^T R^{(t)}_{Hessian}\mu + \text{const} \tag{19}$$

are valid majorizers of $R_{\text{TV}}(\mu)$ and $R_{\text{Hessian}}(\mu)$ at the fixed point $\mu^{(t)}$, where $R^{(t)}_{TV}$ and $R^{(t)}_{\text{Hessian}}$ are asymmetric square matrixes, and the constants are independent of μ, but depend only upon $\mu^{(t)}$. Since the WLS term $\Phi_{\text{WLS}} = \frac{1}{2}(\widehat{p} - A\mu)^T\Sigma^{-1}(\widehat{p} - A\mu)$ in equation (3) is quadratic and satisfies the two conditions in equation (14), function

$$Q_{\Phi}(\mu;\mu^{(t)}) = \frac{1}{2}(\widehat{p} - A\mu)^T\Sigma^{-1}(\widehat{p} - A\mu) + \beta Q_R(\mu;\mu^{(t)}) \tag{20}$$

is a quadratic majorizer of the original objective function $\Phi(\mu)$, where $Q_R(\mu;\mu^{(t)})$ is $Q_{R_{\text{TV}}}(\mu;\mu^{(t)})$ or $Q_{R_{\text{Hessian}}}(\mu;\mu^{(t)})$, specifically. Let $\frac{\partial}{\partial\mu}Q_{\Phi}(\mu;\mu^{(t)}) = 0$, the minimization of the majorizers is equivalent to solving the following sequence of linear equations

$$(A^T\Sigma^{-1}A + \beta R^{(t)})\mu = A^T\Sigma^{-1}\widehat{p}\,. \tag{21}$$

Equation (21) can be efficiently solved by the Gauss–Seidel update strategy (Wang *et al* 2008b). Note that for a fixed voxel μ_j, the cross term $\mu_j\mu_i$ arises from the derivative operation and only exists when μ_i is no more than one voxel away from μ_j in equation (18) and when no more than two voxels away in equation (19). As a result, the majorizers in equations (18) and (19) can be equivalently stated as

$$Q_R(\mu;\mu^{(t)}) = \frac{1}{2}\sum_j\sum_{i\in N_j} w'^{(t)}_{ji}(\mu_j - \mu_i)^2 + \text{const}, \tag{22}$$

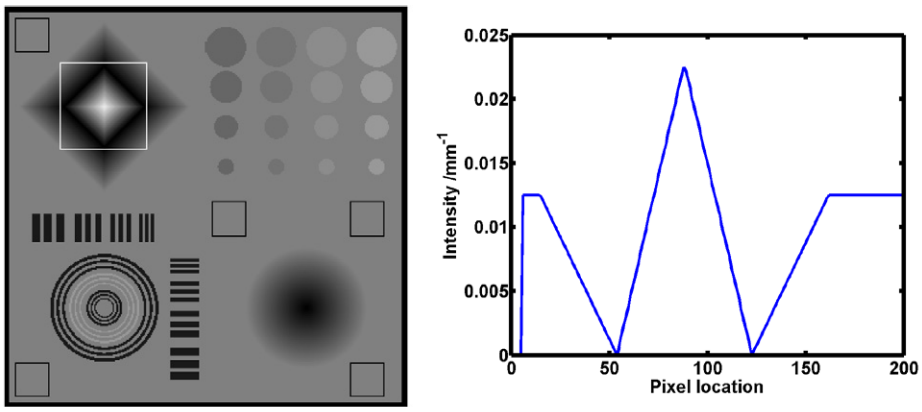
where $Q_R(\mu;\mu^{(t)})$ is $Q_{R_{\text{TV}}}(\mu;\mu^{(t)})$ or $Q_{R_{\text{Hessian}}}(\mu;\mu^{(t)})$ specifically, and N_j is the set of neighbor voxels of voxel j. Details about the algorithm are described in the appendix A.

4. Materials and evaluation

4.1. Simulated phantom

We first performed a simulation study using a Compressed Sensing (CS) phantom (Smith 2010). Figure 1(*a*) shows one typical slice of the CS phantom. The phantom has a uniform background with attenuation coefficient of 0.0125 mm^{-1}. It contains an octahedron with linearly gradual transitioning intensity. Figure 1(*b*) plots the horizontal profile through the center of the octahedron along the shown slice. It also contains a set of line objects for studying the image resolution, a set of circular cylinders with different diameters and intensity for studying the contrast-to-noise ratio (CNR) and a ball with linearly gradual transitioning intensity similar to the octahedron.

The digital phantom was forward projected mathematically using the SF forward projection algorithm. The total angle samples over 360° was 360. The dimensions of each projection



(a) (b)

Figure 1. (*a*) One representative slice of the CS phantom, (*b*) horizontal profile through the center of the octahedron at the shown slice.

image were 800 × 200 pixels with the pixel size of 0.776 × 0.776 mm^2. The distance of source-to-axial distance is 100 cm and the source-to-detector distance is 150 cm, which are the same as those in the physical phantom experiments described in the following section. Randomly generated Gaussian noises were added to the projections according to the noise model of equation (2). The incident photon number in equation (2) was set to be 5 × 10^3, 1 × 10^4 and 5 × 10^4 in order to simulate low-dose, medium-dose and high-dose protocol, respectively.

The simulated projection data at different noise level was used for reconstruction. The reconstructed images had dimensions of 350 × 350 × 16 voxels, where the voxel size is 0.776 × 0.776 × 0.776 mm^3.

4.2. Physical phantoms

Two physical phantoms were used to evaluate the performance of the proposed algorithm. The first is a commercial calibration phantom CatPhan® 600 (The Phantom Laboratory, Inc., Salem, NY). The second is an anthropomorphic head phantom. The projection data were acquired by ExactArms (kV source/detector arms) of a Trilogy(tm) treatment system (Varian Medical Systems, Palo Alto, CA). The number of projections for a full 360° rotation is around 634. The dimension of each acquired projection image is 397 mm × 298 mm, containing 1024 × 768 pixels. The system has a FOV of 25 cm × 25 cm (full-fan mode) in the transverse plane and 17 cm in the longitudinal direction, which can be increased to 45 cm × 45 cm in the transverse plane by shifting the detector laterally (half-fan mode).

For each phantom, the x-ray tube current was set at 10 mA (low dose) and 80 mA (high dose) during the acquisition of CBCT projection images. In both phantom studies, the tube voltage was set to 125 kVp and the duration of the x-ray pulse at each projection view was 10 ms. The projection date were acquired in full-fan mode with a bowtie filter. The distance of source-to-axial distance is 100 cm and source-to-detector distance of 150 cm. The size of the reconstructed image is 350 × 350 × 16 and 550 × 550 × 32 for the CatPhan® 600 and the anthropomorphic head phantom, respectively. The voxel size in the reconstructed image is 0.776 × 0.776 × 0.776 mm^3 and 0.388 × 0.388 × 0.388 mm^3, respectively. Here, we chose smaller voxel size in order to obtain more details for the anthropomorphic head phantom.

4.3. Comparison algorithms

Besides of the TV penalty, we also compared the proposed Hessian penalty with the Huber penalty, which has been frequently used in CT reconstruction (Elbakri and Fessler 2002, Thibault *et al* 2007, Wang *et al* 2014). The Huber penalty has the following form

$$H(t) = \begin{cases} t^2/2 & |t| < \delta \\ \delta(|t| - \delta) + \delta^2/2 & |t| \geq \delta \end{cases}$$

It applies a linear penalty to the difference between neighboring voxels if the absolute difference voxel value $|t|$ is greater than the threshold δ and a quadratic penalty to the small differences which usually occur at gradual transitioning regions. With a proper choice of the threshold, the Huber penalty has an excellent performance for CT reconstruction. For a fair comparison, the Huber penalty was integrated into our minimization framework using the MM approach described in section 3. In this study, the threshold parameter δ was set to 0.001 mm^{-1} in accord with the value used in (Elbakri and Fessler 2002, Wang *et al* 2008).

4.4. Evaluation indexes

To quantify the quality of the reconstructed images, we selected the peak signal-to-noise ratio (PSNR), increase in the signal-to-noise ratio (ISNR), contrast-to-noise ratio (CNR) and structural similarity (SSIM). In our experiments, the FDK reconstructed image is used as an initial guess of the iterative reconstruction algorithms.

PSNR is defined as

$$\text{PSNR} = 10\log_{10}\frac{\mu_{\max}^2}{\text{MSE}} \tag{24}$$

where MSE is the mean-squared error between the reconstructed image and the reference image and $\mu_{\max}$ is the maximum possible value of the image. Higher PSNR means less error between the reconstructed image and the reference image.

ISNR is defined as

$$\text{ISNR} = 10\log_{10}\frac{\text{MSE}_{\text{in}}}{\text{MSE}_{\text{out}}} \tag{25}$$

where MSE_{in} and MSE_{out} are the mean-squared errors between the FDK reconstructed image and the reference image, and the iterative reconstructed image and the reference image, respectively. Higher ISNR means better improvement in the quality of the iterative reconstructed image over the FDK reconstructed image.

CNR is defined as

$$\text{CNR} = \left|\mu_{\text{ROI}} - \mu_{\text{Ref}}\right| / \sqrt{\sigma_{\text{ROI}}^2 + \sigma_{\text{Ref}}^2} \tag{26}$$

where μ_{ROI} and μ_{Ref} are the mean intensity of a selected ROI and the reference uniform background region, σ_{ROI} and σ_{Ref} are the standard deviation of the ROI and the reference region. Higher CNR means better preservation of the signal from noises.

The SSIM metric is defined as (Wang *et al* 2004)

$$\text{SSIM}(a,b) = \frac{\left(2\mu_a\mu_b + C_1\right)\left(2\sigma_{ab} + C_2\right)}{\left(\mu_a^2 + \mu_b^2 + C_1\right)\left(\sigma_a^2 + \sigma_b^2 + C_2\right)} \tag{27}$$

where a and b are two local windows of size 8 × 8 pixels in two images. The two windows have the same position, and μ_a and σ_a, $\mu_b$and σ_b are their mean and standard deviation, respectively. σ_{ab} is the covariance between the two windows. C_1 and C_2 are two constants to avoid instability. In this study, C_1 and C_2 were chosen as $C_1 = (0.01\mu_{\max})^2$ and $C_2 = (0.03\mu_{\max})^2$. SSIM is used to measure the similarity in the structure between the two windows where higher value means higher similarity. As the two windows move pixel-by-pixel over the reconstructed image and the reference image, we obtain a SSIM map. In practice, we use a single Mean-SSIM (MSSIM) value to evaluate the overall image quality by simply averaging the SSIM values.

5. Experimental results

5.1. CS Phantom

5.1.1. Convergence analysis. Convergence of the PWLS may depend on the choice of penalties. In this work, the l_2-norm of the difference of the attenuation coefficients between successive iterations (residual) was calculated as the indicator of convergence of the iterative algorithms. Figure 2 shows the convergence curves of the fully sampled CS phantom reconstructed using the PWLS algorithm with the TV, Huber and Hessian penalties. After 20 iterations, good convergence was achieved for all three penalties. In this study, 30 iterations were used for all PWLS algorithms to reconstruct images for the following analysis.

5.1.2. Evaluation. We selected a ROI of size 76 × 76 pixels (indicated by a white square in figure 1(*a*)) at the octahedron that we were particularly interested in. The noise level of the image was characterized by the mean standard deviations of five uniform regions with size of 30 × 30 pixels (indicate by the black squares in figure 1(*a*)). For each group of experiments with different incident photon numbers, the trade-off parameter β was chosen so that the iterative reconstructed images the different penalties were at the comparable noise level.

For a better illustration of the staircase effect, the ROI of the reconstructed images are enlarged and shown in figure 3. It is clearly shown that the images reconstructed with the TV penalty have several piecewise constant regions while those reconstructed with the Hessian penalty and the Huber penalty are more natural.

We further plot the horizontal profile through the center of the octahedron. As is shown in figure 4, several stairs can be seen on the profiles in the images reconstructed using the TV penalty, particularly in the low-dose simulation. While in those profiles reconstructed with the Hessian penalty and the Huber penalty, such effect has been obviously suppressed. This is consistent with the observation in figure 3.

For a quantitative comparison, table 1 listed the evaluation index values of the selected ROI (the white square in figure 1(*a*)) in the reconstructed images, where the noise-free mathematical phantom is used as the reference image. The proposed Hessian penalty has the best performance in all evaluation indexes at all noise levels.

5.2. CatPhan® 600 phantom

Representative slices of the reconstructed images from the CatPhan® 600 phantom are shown in figure 5. The noise level of the reconstructed images was characterized by the standard deviation of the reference region indicated by a white square in figure 5(*a*). The noise level of the low-dose FDK and high-dose FDK reconstructed results are 2.40×10^{-3} and 8.62×10^{-4},

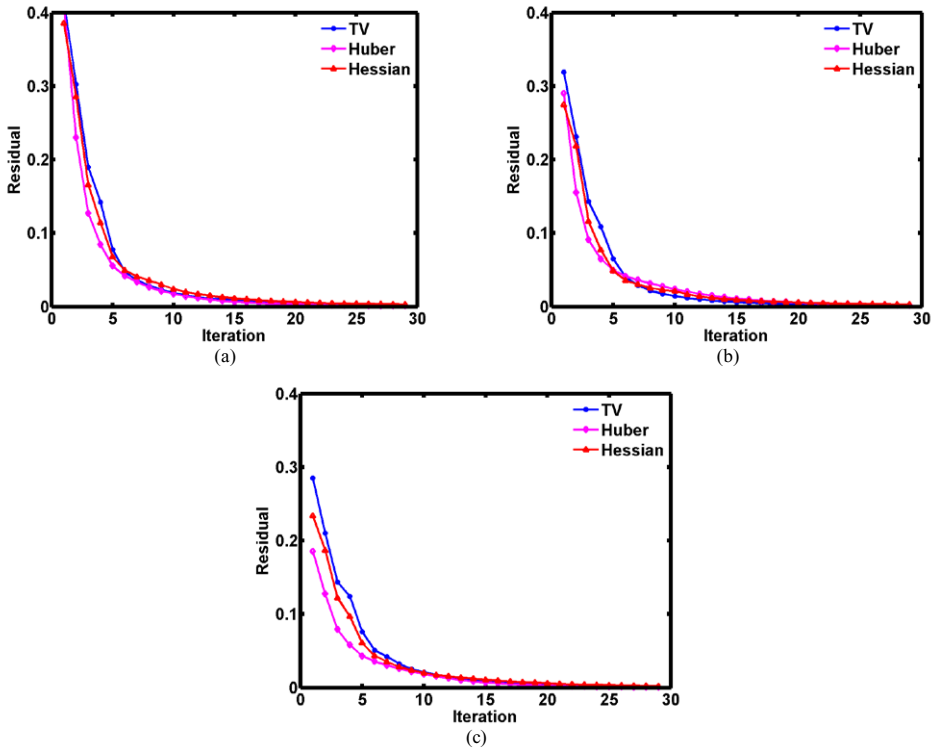


Figure 2. Convergence curves of the iterative reconstruction of the CS phantom using TV, Huber and Hessian penalties with different incident photon numbers: (*a*) 5×10^3, (*b*) 1×10^4, and (*c*) 5×10^4.

respectively. The noise level of the iterative reconstructed results using the TV penalty, the Huber penalty and the Hessian penalty are 8.33×10^{-4}, 8.40×10^{-4} and 8.40×10^{-4}, respectively. In all cases, the noises of the initial FDK reconstructed image from low-dose protocol are greatly removed.

The parameter β in equation (3) trades off between the data-fidelity term and the penalty term. As can be seen in figure 6, the noise level of the reconstructed images decreases monotonously as β increases for all three penalties. In this figure, the upper dashed line indicates the noise level of the FDK reconstructed low-dose image, and the lower dashed line indicates that of the FDK reconstructed high-dose image.

We chose six groups of reconstructed results based on the relation curves between the noise level and trade-off parameter displayed in figure 6. For each group, β was delicately chosen so that the noise levels of three iterative reconstructed results are comparable. The noise levels of the six groups of reconstructed results are listed in the upper part of table 2. For a quantitative comparison, we selected four low-contrast ROIs with size of 17×17 pixels. Their exact locations are indicated by arrows in figure 5(*a*). The CNRs of the four ROIs from the three PWLS algorithms at different noise levels (achieved by using different smoothing parameter β) are listed in the lower part of table 2. As can be observed from the table, most of the CNR values from the Hessian results are equivalent to those from the TV and Huber results at each noise level, indicating similar preservation of the contrast of the objects.

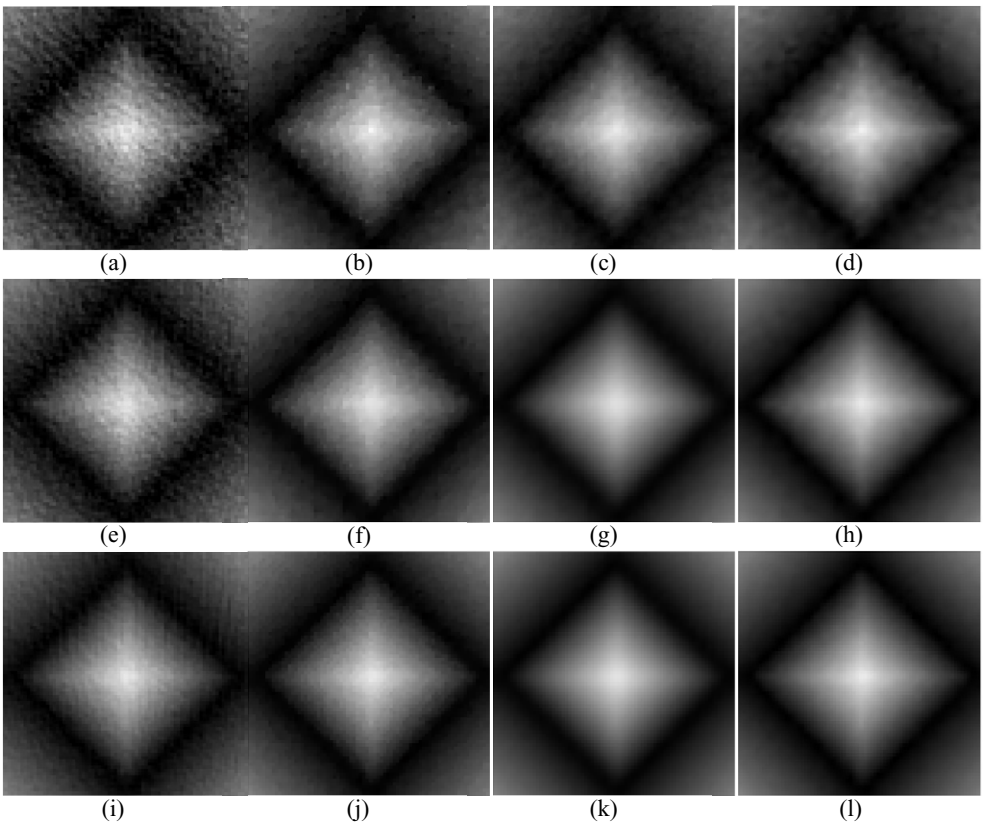


Figure 3. Enlarged ROI of the CS phantom reconstructed by FDK and iterative reconstruction algorithms at different noise levels. High noise level: (*a*) FDK, (*b*) TV penalty, (*c*) Huber penalty, and (*d*) Hessian penalty. Medium noise level: (*e*) FDK, (*f*) TV penalty, (*g*) Huber penalty, and (*h*) Hessian penalty. Low noise level: (*i*) FDK, (*j*) TV penalty, (*k*) Huber penalty, and (*l*) Hessian penalty. Display window: [0, 0.025] mm^{-1}.

5.3. Anthropomorphic head phantom

Representative slices of the reconstructed results from the anthropomorphic head phantom are shown in figure 7. The parameter β was delicately chosen so that the iterative reconstruction results are at the same noise level. We focused on a selected ROI (indicated by a white rectangle in figure 7(*a*)). The standard deviation in the uniform region (indicated by a black square in figure 7(*a*)) equals to 5.32×10^{-4} for the TV penalty (figure 7(*c*)), 5.35×10^{-4} for the Huber penalty (figure 7(*d*)), and 5.34×10^{-4} for the Hessian penalty (figure 7(*f*)). Since the noise-free image is not available, we adopted the high-dose FDK result as a reference to see if the iterative reconstructions from low-dose projection are comparable in quality to the FDK reconstructed results from high-dose projection. The MSSIM due to the TV penalty, the Huber penalty, and the Hessian penalty is 0.9856, 0.9893, and 0.9915, respectively. Figure 8 shows the SSIM map images of the selected ROI between the iterative reconstructed images and the FDK image of high-dose protocol. We can observe that the value of SSIM in figure 8(*c*) is higher than those in figures 8(*a*) and (*b*), indicating that the Hessian penalty has a better preservation of the gradual transform of intensity in image than both the TV penalty and the Huber penalty.

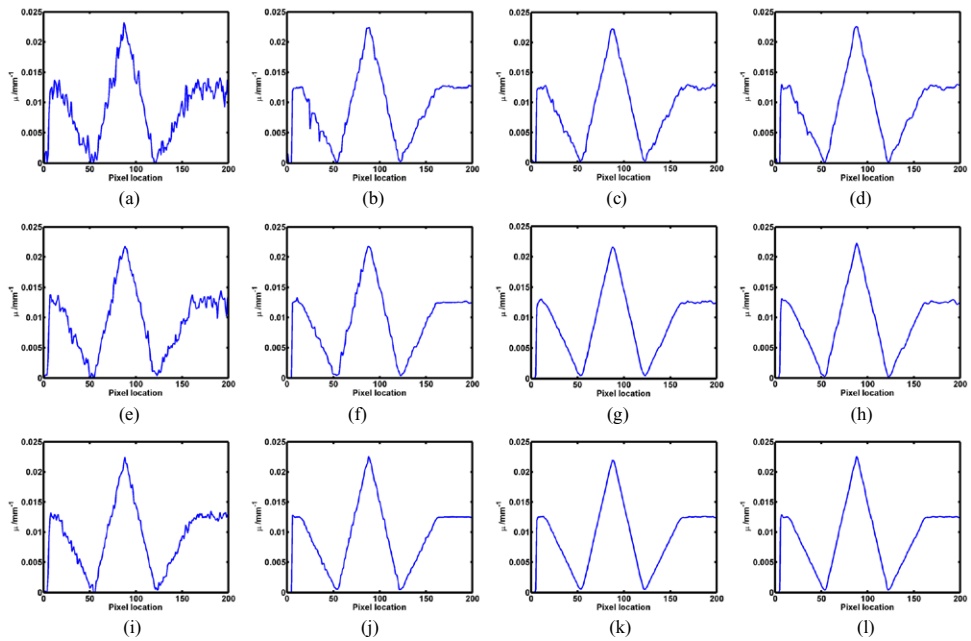


Figure 4. Horizontal profiles through the selected ROI of the CS phantom reconstructed by FDK and iterative reconstruction algorithms at different noise levels. High noise level: (*a*) FDK, (*b*) TV penalty, (*c*) Huber penalty, and (*d*) Hessian penalty. Medium noise level: (*e*) FDK, (*f*) TV penalty, (*g*) Huber penalty, and (*h*) Hessian penalty. Low noise level: (*i*) FDK, (*j*) TV penalty, (*k*) Huber penalty, and (*l*) Hessian penalty.

Table 1. A summary of the evalutation indexes of the iterative reconstruction of the CS phantom at different noise levels.

Incident photon number	Penalty	Noise level ($\times 10^{-4}$)	PSNR (dB)	ISNR(dB)	MSSIM
5×10^3	FDK	12.0	26.90	0.00[a]	0.6587
	TV	3.27	33.77	7.13	0.9075
	Huber	3.44	36.84	10.20	0.9500
	Hessian	3.43	**37.14**	**10.50**	**0.9545**
1×10^4	FDK	9.01	29.82	0.00	0.7895
	TV	1.29	36.84	7.18	0.9492
	Huber	1.30	39.83	10.17	0.9761
	Hessian	1.34	**42.10**	**12.44**	**0.9854**
5×10^4	FDK	5.27	34.84	0.00	0.9258
	TV	0.83	39.34	4.54	0.9701
	Huber	0.86	42.30	7.51	0.9846
	Hessian	0.81	**44.74**	**9.95**	**0.9914**

[a] Note that FDK results are used as reference in ISNR.

6. Discussion

The weighted least-square criterion reflects that the measured data with a lower SNR contributes less to the estimation of the attenuation map. The PWLS objective function is equivalent

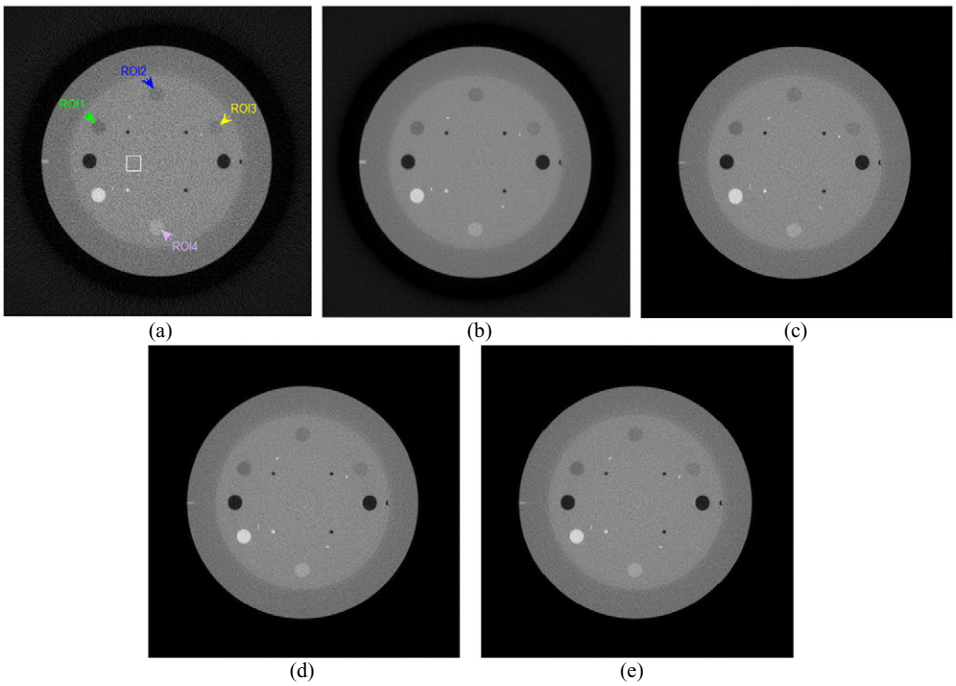


Figure 5. Representative slices of CatPhan® 600 phantom. FDK reconstructed images: (*a*) low-dose protocol (10 mA/10 ms), and (*b*) high-dose protocol (80 mA/10 ms). PWLS reconstructed images with a low-dose protocol: (*c*) TV penalty, (d) Huber penalty, and (*e*) Hessian penalty. Display window: [0, 0.04] mm^{-1}.

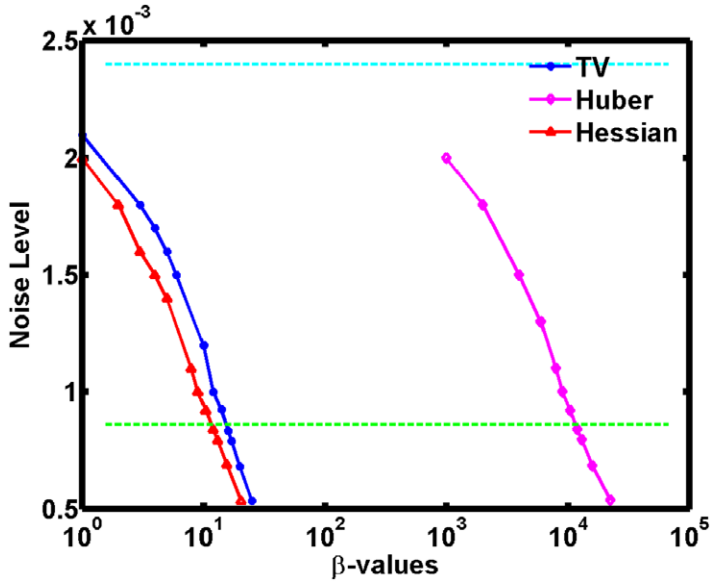


Figure 6. Curves describing the relation between noise level and trade-off parameter for TV, Huber and Hessian penalties, where β values are log-plotted.

Table 2. CNRs of four selected ROIs for the CatPhan® 600 phantom at different noise levels.

Group no.			I	II	III	IV	V	VI
Noise level ($\times 10^{-3}$)		FDK (80 mA)	8.62					
		FDK (10 mA)	24.0					
		TV	5.35	6.87	8.42	9.26	10	21
		Huber	5.38	6.86	8.40	9.19	10	20
		Hessian	5.35	6.89	8.40	9.21	10	20
CNR	ROI1	FDK (80 mA)	3.58					
		FDK (10 mA)	1.47					
		TV	8.58	6.71	5.50	5.01	4.51	2.20
		Huber	8.54	6.74	5.52	5.05	4.56	2.31
		Hessian	8.56	6.69	5.51	5.03	4.56	2.30
	ROI2	FDK (80 mA)	2.58					
		FDK (10 mA)	0.92					
		TV	5.98	4.68	3.83	3.49	3.14	1.53
		Huber	5.96	4.68	3.82	3.50	3.15	1.60
		Hessian	5.99	4.67	3.84	3.51	3.18	1.61
	ROI3	FDK (80 mA)	2.02					
		FDK (10 mA)	0.78					
		TV	4.78	3.74	3.06	2.79	2.50	1.20
		Huber	4.77	3.74	3.05	2.79	2.51	1.25
		Hessian	4.79	3.73	3.07	2.80	2.54	1.26
	ROI4	FDK (80 mA)	2.51					
		FDK (10 mA)	0.90					
		TV	5.83	4.51	3.67	3.35	2.98	1.45
		Huber	5.78	4.53	3.69	3.37	3.03	1.51
		Hessian	5.81	4.49	3.67	3.35	3.03	1.51

to the penalized maximum likelihood or maximum *a posteriori* criterion for Gaussian distributed noise. In this work, we assume that the projection datum is statistically independent with the weighted matrix Σ being diagonal. This may be improved with further consideration of correlation between neighboring projection datum. Besides, the performance of iterative reconstruction algorithms for x-ray CT may be improved by more accurate noise modeling.

The penalty term reflects the prior information of the CT images. The TV penalty encourages smoothness among neighboring pixels thus has a strong ability in suppressing noise. However, this leads to the staircase effect. As was demonstrated in the simulation experiments, clear piecewise constant regions can be observed in the octahedron in figures 3(*b*), (*f*) and (*j*). Such over-smoothness may suppress the gradual transition of intensity in images. Instead of using the first-order derivation, the proposed Hessian penalty uses the second-order derivation, thus keeping a better balance between suppressing noise and preserving intensity transition. In the simulation experiment and the anthropomorphic head phantom experiment, the iterative reconstructed images using the Hessian penalty are more similar to the original images visually. This was further validated by the SSIM maps and MSSIM values.

The Huber penalty applies a quadratic penalty to the small differences and a linear penalty to the large difference. It should be noted that the performance of the Huber penalty depends on the choice of the threshold δ. An improper choice of this parameter can degrade the performance. As an example, we display in figure 9 the enlarged ROI and in figure 10 the representative slices of the CS phantom reconstructed using Huber penalty with the threshold 0.0005 mm^{-1}, 0.001 mm^{-1} and 0.002 mm^{-1}, respectively, and all other settings were kept

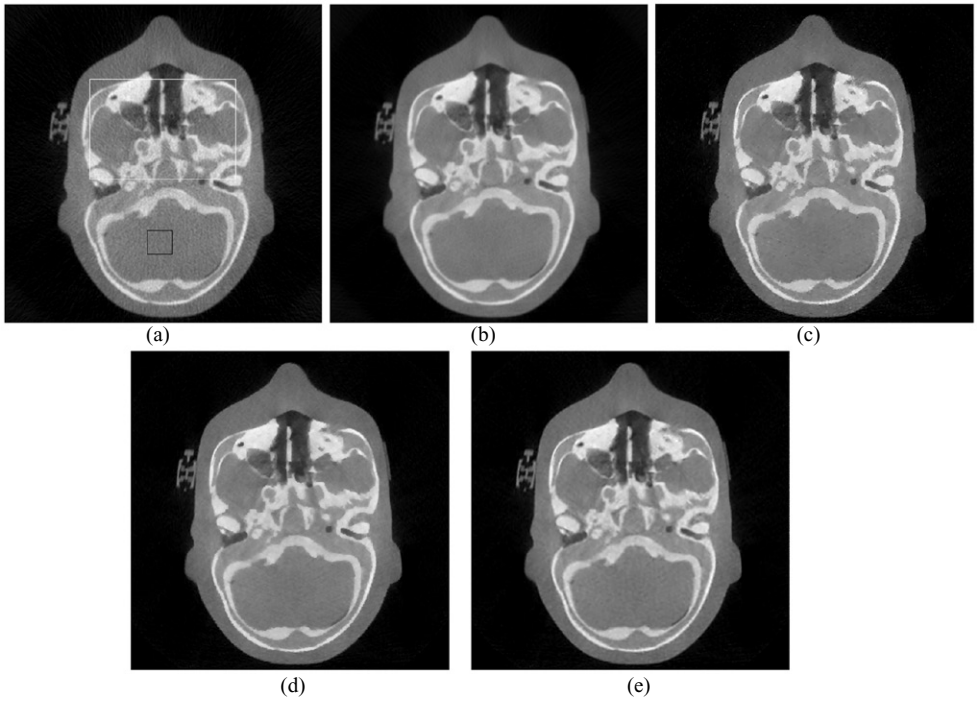


Figure 7. Representative slices of anthropomorphic head phantom. FDK reconstructed images: (*a*) low-dose protocol (10 mA/10 ms), and (*b*) high-dose protocol (80 mA/10 ms), (*c*) TV penalty with low-dose protocol, (*d*) Huber penalty with low-dose protocol, and (*e*) Hessian penalty with low-dose protocol. Display window: [0, 0.045] mm^{-1}.

exactly the same as those in figure 3. A small value of 0.0005 mm^{-1} leaded to an obvious staircase effect and several piecewise constant regions can be observed in figures 9(*a*)–(*c*), similar to those in figures 3(*b*), (*f*) and (*j*) due to the TV penalty. The reason might be related to the fact that when the threshold is too small the Huber penalty generally applies a linear penalty and approximates to the TV penalty. On the contrary, a large value of 0.002 mm^{-1} tends to blur the edges as can be observed in the line and circle objects in figures 10(*a*)–(*f*). In fact, when the threshold is large, the Huber penalty generally applies a quadratic penalty and thus approximates to the Gaussian penalty. Instead of using a hard threshold, the proposed Hessian penalty uses the second-order derivation and utilizes the gradient information by its nature, which can be considered as one of its potential advantages over the Huber penalty. The Hessian penalty has a better preservation of object edges than the Huber penalty with threshold of both 0.002 mm^{-1} and 0.001 mm^{-1}, as shown in figure 10.

The staircase effect due to the TV penalty has been thoroughly studied in the field of image denoising and image restoration. One way to overcome the effect in the variational framework is to add higher-order operations into the corresponding PDE equation. Chen *et al* suggested to add a nonlinear fourth order diffusive term in the Euler–Lagrange equations (Chan *et al* 2000). Lysaker *et al* proposed a fourth-order partial differential equation (Lysaker *et al* 2003). Lixin *et al* attempted to alleviate staircase effect by coupling the gradient fidelity term (Lixin and Deshen 2008). Another way is to use a sparsifying transform, such as wavelet transform (Durand and Froment 2003) and DCT (Gothandaraman *et al* 2001). For implementation, the higher-order operations can be involved in various norm-based penalties. In equation (5),

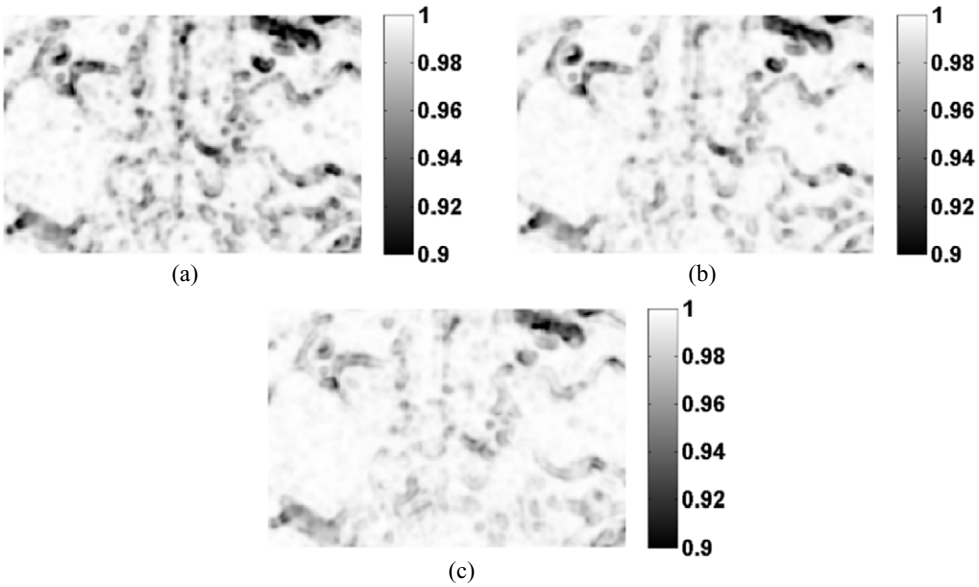


Figure 8. SSIM maps of anthropomorphic head phantom slices reconstructed using iterative reconstruction algorithms with low-dose protocol and FDK with high-dose protocol: (*a*) TV penalty, (*b*) Huber penalty, and (*c*) Hessian penalty.

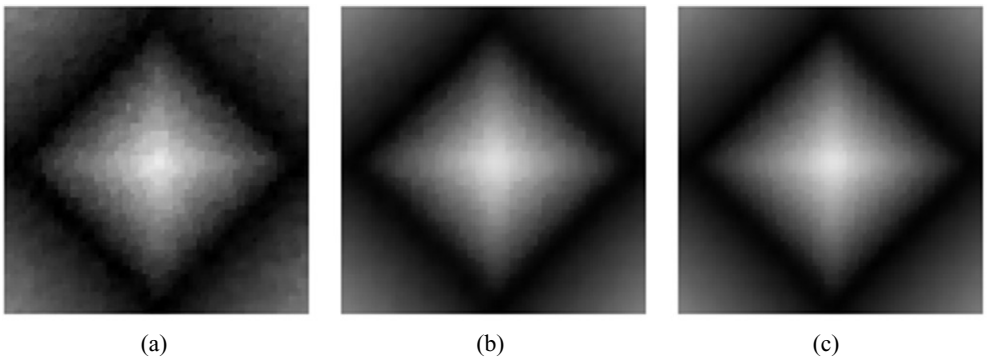


Figure 9. Enlarged ROI of the CS phantom reconstructed by PWLS using Huber penalty with threshold set to be $0.0005\,\mathrm{mm}^{-1}$ (*a*) high noise level, (*b*) medium noise level, and (*c*) low noise level. Display window: $[0, 0.025]\,\mathrm{mm}^{-1}$.

we simply adopted the Frobenius norm of the Hessian matrix as a natural extension of the l_2 norm. The choice of Frobenius norm enables us to develop an efficient algorithm using the majorization-minimization strategy. Experiments in image restoration indicate that a potential improvement may be obtained by using a general Schatten norm (Lefkimmiatis *et al* 2012).

Recently, some novel patch-based regulation frameworks have been proposed for image denoising such as non-local means (NL-means) (Buades *et al* 2005), Block-Matching and 3D filtering (BM3D) (Katkovnik *et al* 2010) and spatial adapted patch-based denoising (Kervrann and Boulanger 2006). These patch based methods employ extra structural information through similarity analysis of small image patches localized along the whole image, and are able to better preserve weak edges and texture details. Further performance improvement may be

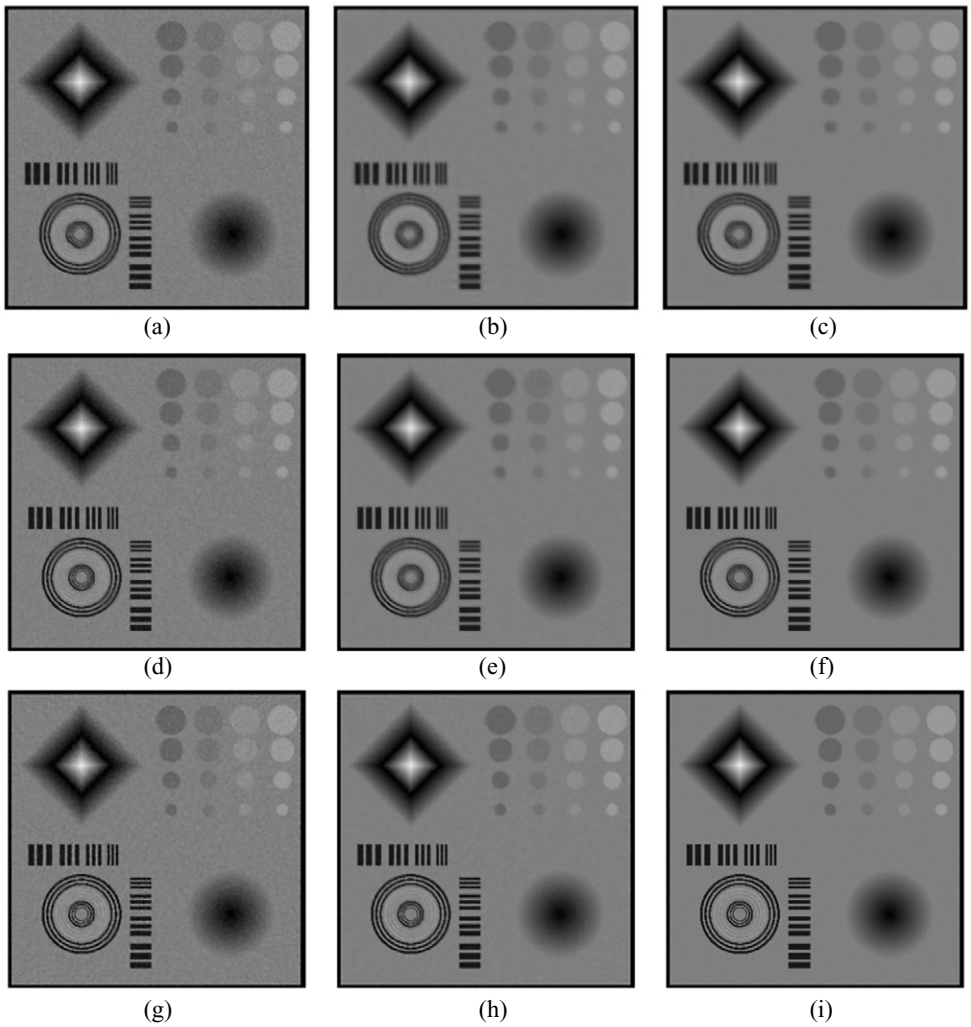


Figure 10. Representative slices of the CS phantom reconstructed by PWLS using Huber penalty with threshold 0.002 mm^{-1}: (*a*) High noise level, (*b*) Medium noise level, and (*c*) Low noise level. Huber penalty with threshold 0.001 mm^{-1}: (*d*) High noise level, (*e*) Medium noise level, and (*f*) Low noise level. Hessian penalty: (*g*) High noise level, (*h*) Medium noise level, and (*i*) Low noise level. Display window: [0, 0.025] mm^{-1}.

obtained by integrating the non-local patch similarity information into the proposed iterative reconstruction algorithm with Hessian penalty.

The TV penalty has a good ability in preserving object edges. However, this sometimes leads to oversharpness. We can observe that even though the image reconstructed using the TV penalty in figure 7(*c*) has more sharpen edges than that reconstructed using the Hessian penalty in figure 7(*e*), it is less similar in structure to the reconstructed image with high-dose protocol. To further combine the advantage of preserving gradual transition of the Hessian penalty and the advantage of preserving resolution of the TV penalty, one possible approach would be to combine the second-order derivation and the first-order derivation to get a mixed-order penalty.

Iterative reconstruction algorithms have shown advantages for CBCT imaging in terms of noise suppressing and structure preservation, however, computational time could be a challenge for its practical use. Recently, many efficient and theoretically fast converging algorithms have been proposed such as the FISTA (Beck and Teboulle 2009) and the NESTA (Choi *et al* 2010, Becker *et al* 2011). Besides, combining graphics processing unit (GPU) with NVIDIA's parallel Compute Unified Device Architecture (CUDA) for CBCT reconstruction has shown great shrink in computation time (Jia *et al* 2010, Noël *et al* 2010). In our implementation, we do not need to precompute and save the projection matrix A. The corresponding column of A was calculated using separable footprints (SF) algorithm described in section 2.3 during the updating of a specific voxel. It takes about 2 min to finish one iteration to reconstruct the CBCT images of a size of 350 × 350 × 16 using a PC with 3.3GHz CPU. Nevertheless, the reconstruction can be sped up by GPU since the SF algorithm can be efficiently adapted for parallel accelerating (Wu and Fessler 2011).

7. Conclusion

In this study, we proposed a Hessian penalty for CBCT reconstruction. We developed an effective algorithm to minimize the objective function using the majorization-minimization strategy. Comparison studies have shown better ability of the proposed penalty in preserving intensity transitions of images than the TV penalty and its potential to overcome the staircase-effect.

Acknowledgement

This work was supported in part by National Natural Science Foundation of China (NNSFC), under Grant Nos. 60971112 and 61375018, and Fundamental Research Funds for the Central Universities, under Grant No. 2012QN086. J Wang was supported in part by grants from the Cancer Prevention and Research Institute of Texas (RP130109 and RP110562-P2) and a grant from the American Cancer Society (RSG-13-326-01-CCE).

Appendix A

A.1. Derivative filters

For a 3D image, there are six commonly used second-order derivatives. Let us represent the voxel index by the triplet (x, y, z). For a discrete image μ, the filtering operations can be defined as

$$L_1(x,y,z) * \mu(x,y,z) = \mu(x+1,y,z) - 2\mu(x,y,z) + \mu(x-1,y,z)$$
$$L_2(x,y,z) * \mu(x,y,z) = \mu(x,y+1,z) - 2\mu(x,y,z) + \mu(x,y-1,z)$$
$$L_3(x,y,z) * \mu(x,y,z) = \mu(x,y,z+1) - 2\mu(x,y,z) + \mu(x,y,z-1)$$
$$L_4(x,y,z) * \mu(x,y,z) = \sqrt{2}\,[\mu(x,y,z) - \mu(x-1,y,z) - \mu(x,y-1,z) + \mu(x-1,y-1,z)\,]$$
$$L_5(x,y,z) * \mu(x,y,z) = \sqrt{2}\,[\mu(x,y,z) - \mu(x-1,y,z) - \mu(x,y,z-1) + \mu(x-1,y,z-1)\,]$$
$$L_6(x,y,z) * \mu(x,y,z) = \sqrt{2}\,[\mu(x,y,z) - \mu(x,y-1,z) - \mu(x,y,z-1) + \mu(x,y-1,z-1)\,]$$

A.2. Update of $w_{ji}'^{(t)}$

In section 3.3, we introduced the majorization-minimization approach to upper-bound the original penalty function with a quadratic form. The difference of the TV penalty and the proposed Hessian penalty lies in the matrix $R^{(t)}$, where $R^{(t)}$ is $R_{\mathrm{TV}}^{(t)}$ or $R_{\mathrm{Hessian}}^{(t)}$ in equations (18) and (19), specifically. Obviously, $R_{\mathrm{Hessian}}^{(t)}$ has more non-zero elements than $\mathrm{R}_{\mathrm{TV}}^{(t)}$ alongside the main diagonal. For voxel μ_j, N_j is the set of neighbor voxels which are at most 2-voxel far from it. With the definition in equation (22), we can easily obtain that $w_{ji}'^{(t)} := -\mathrm{R}_{ji}^{(t)}/2$, $\mu_i \in N_j$, $i \neq j$. $w_{ji}'^{(t)}$ is updated in each iteration. For the TV penalty, the detailed update formula of $w_{ji}'^{(t)}$ can be referred in Wang *et al* (2008b). For the proposed Hessian penalty, we first defined $W(x, y, z)$ as

$$W(x, y, z) := \frac{1}{\sqrt{\sum_{i=1}^{6} [L_i(x, y, z) * \mu^{(t)}(x, y, z)]^2}}$$

where L_i is the derivative filter defined in the previous section. The update formula of $w_{ji}'^{(t)}$ can be expressed as:

For each neighbor voxel $\mu_i \in N_j$ of voxel μ_j

if $i = (x+1, y, z)$ then $w_{ji}'^{(t)} = W(x, y, z) + 3W(x+1, y, z) + W(x+1, y+1, z) + W(x+1, y, z+1)$

if $i = (x, y+1, z)$ then $w_{ji}'^{(t)} = W(x, y, z) + 3W(x, y+1, z) + W(x+1, y+1, z) + W(x, y+1, z+1)$

if $i = (x, y, z+1)$ then $w_{ji}'^{(t)} = W(x, y, z) + 3W(x, y, z+1) + W(x+1, y, z+1) + W(x, y+1, z+1)$

if $i = (x-1, y, z)$ then $w_{ji}'^{(t)} = W(x-1, y, z) + 3W(x, y, z) + W(x, y+1, z) + W(x, y, z+1)$

if $i = (x, y-1, z)$ then $w_{ji}'^{(t)} = W(x, y-1, z) + 3W(x, y, z) + W(x, y, z+1) + W(x+1, y, z)$

if $i = (x, y, z-1)$ then $w_{ji}'^{(t)} = W(x, y, z-1) + 3W(x, y, z) + W(x+1, y, z) + W(x, y+1, z)$

if $i = (x, y+1, z+1)$ then $w_{ji}'^{(t)} = -W(x, y+1, z+1)$

if $i = (x, y+1, z-1)$ then $w_{ji}'^{(t)} = -W(x, y+1, z)$

if $i = (x, y-1, z-1)$ then $w_{ji}'^{(t)} = -W(x, y, z)$

if $i = (x, y-1, z+1)$ then $w_{ji}'^{(t)} = -W(x, y, z+1)$

if $i = (x+1, y, z+1)$ then $w_{ji}'^{(t)} = -W(x+1, y, z+1)$

if $i = (x+1, y, z-1)$ then $w_{ji}'^{(t)} = -W(x+1, y, z)$

if $i = (x-1, y, z-1)$ then $w_{ji}'^{(t)} = -W(x, y, z)$

if $i = (x-1, y, z+1)$ then $w_{ji}'^{(t)} = -W(x, y, z+1)$

if $i = (x+1, y+1, z)$ then $w_{ji}'^{(t)} = -W(x+1, y+1, z)$

if $i = (x+1, y-1, z)$ then $w_{ji}'^{(t)} = -W(x+1, y, z)$

if $i = (x-1, y-1, z)$ then $w_{ji}'^{(t)} = -W(x, y, z)$

if $i = (x-1, y+1, z)$ then $w_{ji}'^{(t)} = -W(x, y+1, z)$

if $i = (x+2, y, z)$ then $w_{ji}'^{(t)} = -W(x+1, y, z)/2$

if $i = (x, y+2, z)$ then $w_{ji}'^{(t)} = -W(x, y+1, z)/2$

if $i = (x, y, z+2)$ then $w_{ji}'^{(t)} = -W(x, y, z+1)/2$

if $i = (x-2, y, z)$ then $w_{ji}'^{(t)} = -W(x-1, y, z)/2$

if $i = (x, y-2, z)$ then $w_{ji}'^{(t)} = -W(x, y-1, z)/2$

if $i = (x, y, z-2)$ then $w_{ji}'^{(t)} = -W(x, y, z-1)/2$

A.3. Gauss–Seidel update

The task for image reconstruction is to estimate the attenuation coefficient distribution map μ from the projection data $\widehat{p}$ by minimizing equation (3). In this study, the Gauss-Seidel update algorithm was adopted for the minimization problem.

Initialization:

$$\mu = \mathrm{FDK}\,\{\widehat{p}\,\}$$

$$r = \widehat{p} - \mathrm{A}\mu$$

$$s_j = A_j{}^T \Sigma^{-1} A_j, \ \ \forall j$$

For each iteration
Begin
For each pixel μ_j
Begin

$$\text{update}\left\{w_{ji}^{'(t)}\right\}$$

$$\mu_j^{\mathrm{old}} := \mu_j$$

$$\mu_j^{\mathrm{new}} := \frac{A_j{}^T \sum^{-1} r + s_j\,\mu_j^{\mathrm{old}} + \beta \sum_{i\in N_j} w_{ji}^{'(t)}\,\mu_i}{s_j + \beta \sum_{i\in N_j} w_{ji}^{'(t)}}$$

$$\mu_j := \max\left\{\mu_j^{\mathrm{new}},\ 0\right\}$$

$$r := r + A_j\left(\mu_j^{\mathrm{old}} - \mu_j\right)$$

End
End

References

Arigovindan M *et al* 2013 High-resolution restoration of 3D structures from widefield images with extreme low signal-to-noise-ratio *Proc. Natl Acad. Sci. USA* **110** 17344–9

Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imag. Sci.* **2** 183–202

Becker S, Bobin J and Candès E J 2011 NESTA: a fast and accurate first-order method for sparse recovery *SIAM J. Imag. Sci.* **4** 1–39

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

Brenner D J and Hall E J 2007 Computed tomography: an increasing source of radiation exposure *New Engl. J. Med.* **357** 2277–84

Buades A, Coll B and Morel J-M 2005 A non-local algorithm for image denoising *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (San Diego, CA, USA, 20–26 June 2005)* pp 60–5

Chan T, Marquina A and Mulet P 2000 High-order total variation-based image restoration *SIAM J. Sci. Comput.* **22** 503–16

Chlewicki W, Hermansen F and Hansen S 2004 Noise reduction and convergence of Bayesian algorithms with blobs based on the Huber function and median root prior *Phys. Med. Biol.* **49** 4717

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
Durand S and Froment J 2003 Reconstruction of wavelet coefficients using total variation minimization *SIAM J. Sci. Comput.* **24** 1754–67
Elbakri I A and Fessler J A 2002 Statistical image reconstruction for polyenergetic x-ray computed tomography *IEEE Trans. Med. Imaging* **21** 89–99
Fessler J A 1994 Penalized weighted least-squares image reconstruction for positron emission tomography *IEEE Trans. Med. Imaging* **13** 290–300
Geman S and Geman D 2009 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41
Gothandaraman A, Whitaker R and Gregor J 2001 Total variation for the removal of blocking effects in DCT based encoding *Proc. Int. Conf. on Image Processing (Thessaloniki, Greece, October 2001)* pp 455–8
Green P J 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imaging* **9** 84–93
Hanson K M and Wecksung G W 1983 Bayesian approach to limited-angle reconstruction in computed tomography *JOSA* **73** 1501–9
Hunter D R and Lange K 2004 A tutorial on MM algorithms *Am. Stat.* **58** 30–7
Islam M K *et al* 2006 Patient dose from kilovoltage cone beam computed tomography imaging in radiation therapy *Med. Phys.* **33** 1573–82
Jaffray D A, Siewerdsen J H, Wong J W and Martinez A A 2002 Flat-panel cone-beam computed tomography for image-guided radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 1337–49
Jia X, Dong B, Lou Y and Jiang S B 2011 GPU-based iterative cone beam CT reconstruction using tight frame regularization *Phys. Med. Biol.* **56** 3787
Jia X, Lou Y, Li R, 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
Katkovnik V, Foi A, Egiazarian K and Astola J 2010 From local kernel to nonlocal multiple-model image denoising *Int. J. Comput. Vis.* **86** 1–32
Kervrann C and Boulanger J 2006 Optimal spatial adaptation for patch-based image denoising *IEEE Trans. Image Process.* **15** 2866–78
Lange K 1990 Convergence of EM image reconstruction algorithms with Gibbs smoothing *IEEE Trans. Med. Imaging* **9** 439–46
Lee H, Xing L, Davidi R, Li R, Qian J and Lee R 2012 Improved compressed sensing-based cone-beam CT reconstruction using adaptive prior image constraints *Phys. Med. Biol.* **57** 2287
Lefkimmiatis S, Bourquard A and Unser M 2012 Hessian-based regularization for 3D microscopy image restoration *9th IEEE Int. Symp. on Biomedical Imaging (Barcelona, Spain, 2–5 May 2012)* pp 1731–4
Lefkimmiatis S, Ward J P and Unser M 2013 Hessian Schatten-norm regularization for linear inverse problems *IEEE Trans. Image Process.* **22** 1873–88
Lixin Z and Deshen X 2008 Staircase effect alleviation by coupling gradient fidelity term *Image Vis. Comput.* **26** 1163–70
Long Y, Fessler J A and Balter J M 2010 3D forward and back-projection for x-ray CT using separable footprints *IEEE Trans. Med. Imaging* **29** 1839–50
Lysaker M, Lundervold A and Tai X-C 2003 Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time *IEEE Trans. Image Process.* **12** 1579–90
Noël P B, Walczak A M, Xu J, Corso J J, Hoffmann K R and Schafer S 2010 GPU-based cone beam computed tomography *Comput. Methods Prog. Biomed.* **98** 271–7
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
Park J C *et al* 2012 Fast compressed sensing-based CBCT reconstruction using Barzilai–Borwein formulation for application to on-line IGRT *Med. Phys.* **39** 1207–17
Siddon R L 1985 Fast calculation of the exact radiological path for a 3D CT array *Med. Phys.* **12** 252–5
Sidky E Y, Kao C-M and Pan X 2006 Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT *J. X-ray Sci. Technol.* **14** 119–39
Smith D 2010 www.mathworks.cn/matlabcentral/fileexchange/29364-compressed-sensing-mri-phantom--v1-1-/content/csphantom.filex.zip

Thibault J-B, Sauer K D, Bouman C A and Hsieh J 2007 A 3D statistical approach to improved image quality for multislice helical CT *Med. Phys.* **34** 4526–44

Yuan J, Schnörr C and Steidl G 2009 Total-variation based piecewise affine regularization *Scale Space and Variational Methods in Computer Vision* (New York: Springer) pp 552–64

Wang J *et al* 2008 An experimental study on the noise properties of x-ray CT sinogram data in Radon space *Phys. Med. Biol.* **53** 3327

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12

Wang J, Li T, Liang Z and Xing L 2008a Dose reduction for kilovotage cone-beam computed tomography in radiation therapy *Phys. Med. Biol.* **53** 2897

Wang J, Li T and Xing L 2008b Iterative image reconstruction for CBCT using edge-preserving prior *Med. Phys.* **36** 252–60

Wang A S *et al* 2014 Soft-tissue imaging with C-arm cone-beam CT using statistical reconstruction *Phys. Med. Biol.* **59** 1005

Wu C J 1983 On the convergence properties of the EM algorithm *Ann. Stat.* **11** 95–103

Wu M and Fessler J A 2011 GPU acceleration of 3D forward and backward projection using separable footprints for x-ray CT image reconstruction *Proc. Int. Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (Potsdam, Germany, 11–15 July 2011)* pp 56–9