

RESEARCH

Open Access



# Low-dose CT reconstruction via L1 dictionary learning regularization using iteratively reweighted least-squares

Cheng Zhang<sup>1,2,3</sup>, Tao Zhang<sup>2</sup>, Ming Li<sup>1</sup>, Chengtao Peng<sup>4</sup>, Zhaobang Liu<sup>1</sup> and Jian Zheng<sup>1\*</sup>

\*Correspondence:  
zhengj@sibet.ac.cn

<sup>1</sup> Suzhou Institute of Biomedical Engineering and Technology of Chinese Academy of Sciences, Suzhou 215163, China  
Full list of author information is available at the end of the article

## Abstract

**Background:** In order to reduce the radiation dose of CT (computed tomography), compressed sensing theory has been a hot topic since it provides the possibility of a high quality recovery from the sparse sampling data. Recently, the algorithm based on DL (dictionary learning) was developed to deal with the sparse CT reconstruction problem. However, the existing DL algorithm focuses on the minimization problem with the  $L_2$ -norm regularization term, which leads to reconstruction quality deteriorating while the sampling rate declines further. Therefore, it is essential to improve the DL method to meet the demand of more dose reduction.

**Methods:** In this paper, we replaced the  $L_2$ -norm regularization term with the  $L_1$ -norm one. It is expected that the proposed  $L_1$ -DL method could alleviate the over-smoothing effect of the  $L_2$ -minimization and reserve more image details. The proposed algorithm solves the  $L_1$ -minimization problem by a weighting strategy, solving the new weighted  $L_2$ -minimization problem based on IRLS (iteratively reweighted least squares).

**Results:** Through the numerical simulation, the proposed algorithm is compared with the existing DL method (adaptive dictionary based statistical iterative reconstruction, ADSIR) and other two typical compressed sensing algorithms. It is revealed that the proposed algorithm is more accurate than the other algorithms especially when further reducing the sampling rate or increasing the noise.

**Conclusion:** The proposed  $L_1$ -DL algorithm can utilize more prior information of image sparsity than ADSIR. By transforming the  $L_2$ -norm regularization term of ADSIR with the  $L_1$ -norm one and solving the  $L_1$ -minimization problem by IRLS strategy,  $L_1$ -DL could reconstruct the image more exactly.

**Keywords:** Dictionary learning, Image reconstruction,  $L_1$ -norm, Iteratively reweighted least squares

## Background

Nowadays, X-ray CT (computed tomography) is still one of the most important medical imaging technologies. Compared to other imaging methods, like ultrasonic imaging or magnetic resonance imaging, CT imaging has its own advantages to provide patients' anatomical structure. CT images are of higher quality than the ultrasonic images. Compared to the magnetic resonance imaging, CT imaging is of faster imaging speed and CT images have a bit higher spatial resolution.

However, high quality CT images are now based on a noticeable X-ray radiation dose to the patient, which may result in a non-negligible lifetime risk of genetic or cancerous diseases [1]. This fact has become a major concern for clinical applications of CT scans. Therefore, this article focuses on reducing the radiation dose in CT and generating the clinically qualified image.

In order to reduce the radiation dose, one direct way is to lower mAs levels in CT data acquisition protocols. However, this approach will result in insufficient numbers of X-ray photons received by the detectors and hence increase the quantum noise level. This is a great challenge for these advanced methods which have taken the noise models into account. For example, PWLS (penalized weighted least-squares) based methods [2] can only deal with the noise-contaminated sinogram data to some extent. As a consequence, the radiation dose cannot be reduced evidently by this approach if the reconstructed images need to be qualified for clinical diagnosis. Another way to reduce imaging dose is to decrease the number of X-ray projections operated by fewer sampling angles. Yet, this will lead to serious streaking artifacts in the image reconstructed by the analytic-based algorithms like FBP (filtered backprojection algorithm) [3], since the analytic-based algorithms require that the number of projections should follow the Shannon/Nyquist sampling theorem [4].

To solve the under-sampled reconstruction problem, algebraic algorithms transform the problem to a series of linear equations and the reconstructed image is acquired by the iterative method. SART (simultaneous algebraic reconstruction technique) [5] is one of the typical algebraic iterative methods. But the images reconstructed by traditional algebraic algorithms do not satisfy the clinical image quality demand. Researchers are trying to improve the performance of the iterative reconstruction algorithms by introducing prior information of the reconstructed images to the reconstruction process.

Recently, the compressed sensing theory [6, 7] has been applied to the CT reconstruction problem. The reconstruction problem of the compressed sensing algorithm can be written as a constrained optimization

$$\arg \min_{\boldsymbol{\mu}} (\mathbf{A}\boldsymbol{\mu} - \hat{\mathbf{g}})^2 + \lambda R(\boldsymbol{\mu}) \quad (1)$$

where  $(\mathbf{A}\boldsymbol{\mu} - \hat{\mathbf{g}})^2$  is the data fidelity term,  $\mathbf{A}$  is the projection matrix modeling the forward projection,  $\boldsymbol{\mu}$  is the image vector to be reconstructed,  $\hat{\mathbf{g}}$  is the projection vector,  $R(\boldsymbol{\mu})$  is the regularization term including the prior information,  $\lambda$  is the regularization parameter adjusting the relative penalties on the regularization term and the data fidelity term. The regularization term is modeled by the prior information. One commonly used regularization term is the TV (total variation) norm, which is the sum of the absolute coefficients of the DGT (discrete gradient transform) of the reconstructed image. TV-based algorithm is usually used to solve the CT reconstruction problem since that most CT images are piecewise constant. So the TV norm is small enough to reflect the image sparsity. Algorithms like ASD-POCS (adaptive steepest descent projection onto convex sets) [8] and GPBB (gradient projection Barzilai Borwein) [9] are typical TV-based reconstruction algorithms. Besides the TV norm, the recent coming DL (dictionary learning) methods can generate the regularization term, which divide the CT image into many overlapped patches and calculate the sparse representations of the patches

under the basis of an over-complete dictionary. While the TV-based methods take the image as a whole to measure the image sparsity, the DL methods have advantages over TV since they extract the sparse prior information from each overlapped image patch, which utilizes more sparse information than TV. Reference [10] combines the SIR (statistical iterative reconstruction) model with DL regularization term to deal with the low-dose reconstruction problem.

Although the compressed sensing algorithms behave well in the low-dose CT reconstruction problem, some low-contrast details of the reconstructed image are lost, especially when the sampling rate decreases further. In order to reserve more image information and satisfy the need of further radiation reduction, the TV-based algorithm developed a weighted TV regularization to help preserve the edge of the image [11]. When it comes to the DL-based algorithm, it is common that the regularization term is  $L_2$ -norm error. Usually, algorithms based on  $L_2$ -norm error minimization may lead to over-smoothing of the image, causing loss of detail. One way to alleviate this problem is to develop algorithms to minimize the  $L_1$ -norm error. In this work, we develop a DL reconstruction algorithm with the  $L_1$ -norm regularization term while the  $L_1$ -minimization problem is approximated by iteratively solving the weighted  $L_2$ -minimization problem, known as IRLS (iteratively reweighted least squares) [12, 13]. The proposed  $L_1$ -DL ( $L_1$  dictionary learning) algorithm is compared with ADSIR (adaptive dictionary based statistical iterative reconstruction), SART and GPBB to demonstrate the improvement of image quality based on the  $L_1$ -norm regularization term.

The rest of the paper is organized as follows. In “Methods” section, firstly, the backgrounds of ADSIR and IRLS are reviewed. Then the  $L_1$ -DL algorithm and its corresponding optimizing methods are described. After that, the workflow of the algorithm is provided. In section Simulation, a series of experiments are performed to demonstrate the proposed algorithm’s superiority. Finally, the section Conclusion with corresponding discussions and further analysis is provided.

## Methods

### Review of ADSIR

The previous work developed a DL based approach for low-dose X-ray CT with the statistical reconstruction model [10]. The related algorithm is reviewed in details as follows.

### SIR model

Let  $I$  and  $N$  be integers and  $\mathbb{R}$  be the real space. By assuming a monochromatic source, measured data follow the Poisson distribution phantoms with full display

$$y_i \sim \text{Poisson}\{b_i e^{-g_i} + \gamma_i\}, \quad i = 1, \dots, I \quad (2)$$

where  $\mathbf{b} = (b_1, b_2, \dots, b_I)^T \in \mathbb{R}^{I \times 1}$  is the entrance X-ray intensity,  $\mathbf{y} = (y_1, y_2, \dots, y_I)^T \in \mathbb{R}^{I \times 1}$  is the exit X-ray intensity,  $\mathbf{g} = (g_1, g_2, \dots, g_I)^T \in \mathbb{R}^{I \times 1}$  is the integral of the linear attenuation coefficient with  $g_i = [\mathbf{A}\boldsymbol{\mu}]_i = \sum_{j=1}^{N^2} a_{ij}\mu_j$ ,  $\mathbf{A} = \{a_{ij}\} \in \mathbb{R}^{I \times N^2}$  is the system matrix, the reconstructed image  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_{N^2})^T$  is a linear attenuation coefficient distribution, which transforms the initial image of  $N \times N$  pixels to a vector  $\boldsymbol{\mu} \in \mathbb{R}^{N^2 \times 1}$ ,  $\gamma_i$  represents the read-out noise.

The objective function of the SIR model is as

$$\arg \min_{\boldsymbol{\mu}} \sum_{i=1}^I \frac{\omega_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - \hat{g}_i)^2 \tag{3}$$

where  $\sum_{i=1}^I (\omega_i/2)([\mathbf{A}\boldsymbol{\mu}]_i - \hat{g}_i)^2$  is the data fidelity term,  $\hat{\mathbf{g}} = (\hat{g}_1, \hat{g}_2, \dots, \hat{g}_I)^T \in \mathbb{R}^{I \times 1}$  is the measured data of  $\mathbf{g}$  calculated by  $\hat{g}_i = \ln(b_i/(y_i - \gamma_i))$ ,  $\omega_i = (y_i - \gamma_i)^2/y_i$  is the statistical weight.

In the SIR model, the statistical weight reflects the confidence of the projection measurement along each path. The projection data through denser paths would have lower SNR (signal to noise ratios). Compared to the SART which minimizes a least square function, SIR deals with a statistically weighted least square function. However, this development is insufficient for the low-dose CT reconstruction that it is essential to introduce the regularization constraint.

**DL model**

Let  $N_0$  and  $K$  be integers. The DL regularization term is represented as

$$R(\boldsymbol{\mu}) = \sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s\|_2^2 + \sum_{s=1}^S \nu_s \|\boldsymbol{\alpha}_s\|_0 \tag{4}$$

where  $\mathbf{E}_s = \{e_{nj}^s\} \in \mathbb{R}^{N_0^2 \times N^2}$  is an operator to extract patches with  $N_0 \times N_0$  pixels from the image, the image patches are overlapping. With a sliding distance of one pixel, the total number of the patches is  $S = (N - N_0 + 1) \times (N - N_0 + 1)$ .  $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K) \in \mathbb{R}^{N_0^2 \times K}$  is the training dictionary, whose column  $\mathbf{d}_k \in \mathbb{R}^{N_0^2 \times 1}$  is called an atom with the same size of a patch. Usually, the dictionary is redundant or over-complete ( $N_0^2 \ll K$ ).  $\boldsymbol{\alpha}_s \in \mathbb{R}^{K \times 1}$  has few nonzero components as a sparse representation of the patch by the dictionary basis  $\mathbf{D}$ .  $\nu_s$  is the regularization parameter different from  $\lambda$ .

**ADSIR**

By introducing the DL regularization term to the SIR model, the objective function of ADSIR is as

$$\min_{\boldsymbol{\mu}, \boldsymbol{\alpha}, \mathbf{D}} \sum_{i=1}^I \frac{\omega_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - \hat{g}_i)^2 + \lambda \left( \sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s\|_2^2 + \sum_{s=1}^S \nu_s \|\boldsymbol{\alpha}_s\|_0 \right) \tag{5}$$

Since  $\boldsymbol{\mu}$ ,  $\boldsymbol{\alpha}_s$  and  $\mathbf{D}$  are all unknown, the algorithm is iterated by the alternating minimization scheme, which divides the primary problem into two recursive steps—update of the dictionary model and update of the image. During each iteration process, keep the image  $\boldsymbol{\mu}$  unchanged firstly when the dictionary model is updated. And the objective function (5) becomes

$$\min_{\boldsymbol{\alpha}, \mathbf{D}} \sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s\|_2^2 + \sum_{s=1}^S \nu_s \|\boldsymbol{\alpha}_s\|_0 \tag{6}$$

which is the dictionary learning and sparse representation problem. The objective function (5) can be transformed to

$$\begin{aligned} & \min_{\alpha, \mathbf{D}} \sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \alpha_s\|_2^2 \\ & \|\alpha_s\|_0 \leq L_0^S \end{aligned}$$

where the sparse level  $L_0^S$  is set as a fixed number, usually from 5 to 10. Then the L0-norm problem as (5) is transformed to the OMP question, which has no need to determine the value of  $v_s$ . The dictionary  $\mathbf{D}$  is updated by the classic K-SVD (K Singular Value Decomposition) algorithm [14]. Then, the sparse representation  $\alpha_s$  is updated by using the OMP (orthogonal matching pursuit) [15] algorithm based on recent dictionary. Once the dictionary model has been updated in the current iteration process, the image  $\boldsymbol{\mu}$  should be updated with  $\alpha_s$  and  $\mathbf{D}$  invariable. In other words, the problem transforms to the form as

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^I \frac{\omega_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - \hat{g}_i)^2 + \lambda \sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \alpha_s\|_2^2 \tag{7}$$

which consists of the data fidelity term  $\sum_{i=1}^I \omega_i ([\mathbf{A}\boldsymbol{\mu}]_i - \hat{g}_i)^2 / 2$  and the regularization term  $\sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \alpha_s\|_2^2$ . Since the regularization term is already a separable quadratic function. By replacing the data fidelity term with a separable paraboloid surrogate [16], the optimization can be iteratively solved by

$$\begin{aligned} \mu_j^{t+1} = & \left[ \mu_j^t - \frac{\sum_{i=1}^I (a_{ij} \omega_i ([\mathbf{A}\boldsymbol{\mu}^t]_i - \hat{g}_i)) + 2\lambda \sum_{s=1}^S \sum_{n=1}^{N_0^2} e_{nj}^s ([\mathbf{E}_s \boldsymbol{\mu}^t]_n - [\mathbf{D}\alpha_s]_n)}{\sum_{i=1}^I (a_{ij} \omega_i \sum_{k=1}^{N^2} a_{ik}) + 2\lambda \sum_{s=1}^S \sum_{n=1}^{N_0^2} e_{nj}^s \sum_{k=1}^{N^2} e_{nk}^s} \right]_+ \\ & j = 1, 2, \dots, N^2 \end{aligned} \tag{8}$$

**IRLS**

Consider a sparse signal  $x$  with length  $N$  (sparse means the signal has few nonzero components, that is  $\|x\|_0 \ll N$ ) is encoded by an  $M \times N$  measurement matrix  $\Phi$  with  $M < N$ , and the encoded signal is  $y = \Phi x$  with length  $M$ . Referred to [12, 13], the objective function with  $L_p$ -norm minimization to solve the sparse signal is as

$$\min_x \|x\|_p^p, \text{ subject to } \Phi x = y \quad (0 < p \leq 1) \tag{9}$$

IRLS can be used for solving (9) by replacing the  $L_p$ -norm with a weighted  $L_2$ -norm

$$\min_x \sum_{i=1}^N w_i x_i^2, \text{ subject to } \Phi x = y \tag{10}$$

where the weights are computed from previous iteration result  $x^{(t-1)}$ . To make the  $L_2$ -norm approximate to the  $L_p$ -norm, the weights are calculated by

$$w_i = \left| (x_i^{(t-1)})^2 + \varepsilon \right|^{\frac{p}{2}-1} \tag{11}$$

where a small  $\varepsilon > 0$  is provided to ensure stability. Then the signal is iterated by

$$x_i^{(t)} = Q_t \Phi^T (\Phi Q_t \Phi^T)^{-1} y \tag{12}$$

where  $Q_i = \text{diag}(1/w_1, 1/w_2, \dots, 1/w_n)$  is the diagonal matrix with entries  $1/w_i = |(x_i^{(t-1)})^2 + \varepsilon|^{1-\frac{p}{2}}$

**$L_1$ -DL**

In the ADSIR model, the regularization term  $\sum_{s=1}^S ||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||_2^2$  is the sum of the  $L_2$ -norm of the difference between the image patch and its sparse representation. The  $L_2$ -norm constraint tends to distribute the energy of  $\sum_{s=1}^S ||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||_2^2$  to each image patch  $\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s$  uniformly. However, most CT images are piecewise constant, so that most image patches have small values of  $||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||$  (equal or close to zero). A small part of the image patches have large values of  $||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||$  because they contain edge details and image information. In other word, the distribution of  $||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||$  is sparse. So we propose the  $L_1$ -DL method to make  $||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||$  converge to the sparse distribution by the  $L_1$ -norm regularization term. The  $L_1$ -DL utilizes more prior information of image sparsity (the sparse distribution of  $||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||$ ) than ADSIR.

Derived from the ADSIR algorithm, the objective function of the  $L_1$ -DL method is generated by replacing the  $L_2$ -norm of the regularization term with the  $L_1$ -norm, which is

$$\min_{\boldsymbol{\mu}, \boldsymbol{\alpha}, \mathbf{D}} \sum_{i=1}^I \frac{\omega_i}{2} ( [\mathbf{A} \boldsymbol{\mu}]_i - \hat{g}_i )^2 + \lambda \left( \sum_{s=1}^S ||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||_1 + \sum_{s=1}^S \nu_s ||\boldsymbol{\alpha}_s||_0 \right) \tag{13}$$

To make this optimization problem solvable, the patch-based weighted  $L_2$ -norm similar to IRLS is introduced. Since the DL theory takes the sparse representation of the image patch as the sparse constraint, the weights are calculated by the information of each image patch, which is

$$w_s = 1 / \left[ \left( \frac{1}{N_0^2} \sum_{i=1}^{N_0^2} | [\mathbf{E}_s \boldsymbol{\mu}^{(t-1)}]_i - [\mathbf{D}^{(t-1)} \boldsymbol{\alpha}_s^{(t-1)}]_i | \right) + \varepsilon \right] \tag{14}$$

where  $N_0^2$  is the dimension of  $\mathbf{E}_s \boldsymbol{\mu}^{(t-1)}$  and  $\mathbf{D}^{(t-1)} \boldsymbol{\alpha}_s^{(t-1)}$ , the script  $(t - 1)$  means the corresponding terms are previous iteration results, a small  $\varepsilon > 0$  is to ensure stability. The weights is not computed from each component of the vector  $\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s$ . We firstly calculate the average absolute value of the components of the vector  $\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s$ , and then take the reciprocal value of it as the weight of this patch. This strategy is different from the IRLS algorithm, which is in order to fit the nature of DL method that the image patch is treated as the basic unit.

By introducing the weights, the optimization problem becomes

$$\min_{\boldsymbol{\mu}, \boldsymbol{\alpha}, \mathbf{D}} \sum_{i=1}^I \frac{\omega_i}{2} ( [\mathbf{A} \boldsymbol{\mu}]_i - \hat{g}_i )^2 + \lambda \left( \sum_{s=1}^S w_i ||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||_2^2 + \sum_{s=1}^S \nu_s ||\boldsymbol{\alpha}_s||_0 \right) \tag{15}$$

The iteration is also operated by the alternating minimization scheme. When updating the dictionary model, the objective function (15) becomes

$$\sum_{s=1}^S w_i ||\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s||_2^2 + \sum_{s=1}^S \nu_s ||\boldsymbol{\alpha}_s||_0. \tag{16}$$

By the transformation  $\mathbf{E}_s \boldsymbol{\mu}' = \sqrt{w_s} \mathbf{E}_s \boldsymbol{\mu}, \boldsymbol{\alpha}'_s = \sqrt{w_s} \boldsymbol{\alpha}_s$  and the equation  $\|\boldsymbol{\alpha}_s\|_0 = \|\sqrt{w_s} \boldsymbol{\alpha}_s\|_0$ , (16) can be transformed to

$$\sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu}' - \mathbf{D} \boldsymbol{\alpha}'_s\|_2^2 + \sum_{s=1}^S v_s \|\boldsymbol{\alpha}'_s\|_0 \tag{17}$$

which is as the same form as (6), so that the dictionary model can be updated by K-SVD and OMP.

When updating the image, the objective function (15) becomes

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^I \frac{\omega_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - \hat{g}_i)^2 + \lambda \sum_{s=1}^S w_s \|\mathbf{E}_s \boldsymbol{\mu} - \mathbf{D} \boldsymbol{\alpha}_s\|_2^2. \tag{18}$$

Similar to (8), the formula to update the image is

$$\mu_j^{t+1} = \left[ \mu_j^t - \frac{\sum_{i=1}^I (a_{ij} \omega_i ([\mathbf{A}\boldsymbol{\mu}^t]_i - \hat{g}_i)) + 2\lambda \sum_{s=1}^S w_s \sum_{n=1}^{N_0^2} e_{nj}^s ([\mathbf{E}_s \boldsymbol{\mu}^t]_n - [\mathbf{D} \boldsymbol{\alpha}_s]_n)}{\sum_{i=1}^I (a_{ij} \omega_i \sum_{k=1}^{N^2} a_{ik}) + 2\lambda \sum_{s=1}^S w_s \sum_{n=1}^{N_0^2} e_{nj}^s \sum_{k=1}^{N^2} e_{nk}^s} \right]_+ \tag{19}$$

$j = 1, 2, \dots, N^2$

The convergence of L1-DL is much more difficult to prove, and is considered beyond the scope of this paper. However, our experimental results to be reported below seem suggesting the convergence of our proposed algorithms.

Above all, the workflow of the developed algorithm is exhibited in Algorithm I. In addition, the ordered subsets convex (OSC) algorithm [17] is utilized to accelerate the convergence.

**ALGORITHM I. Workflow of the developed algorithm**

- Initialize  $\boldsymbol{\mu}^{(0)}, \mathbf{D}^{(0)}, \boldsymbol{\alpha}_s^{(0)}$ , and  $t = 1$ , set  $w_s = 1$
- While the stopping criterion is not satisfied, do
1. Implement the OSC algorithm for acceleration;
  2. Extract patches from the intermediate image  $\boldsymbol{\mu}^{(t-1)}$ ;
  3. Transform the dictionary optimizing model (16) to the regular one as (17);
  4. Update the dictionary  $\mathbf{D}^{(t)}$  by K-SVD algorithm;
  5. Update the sparse coding  $\boldsymbol{\alpha}_s^{(t)}$  by OMP algorithm;
  6. Update the image  $\boldsymbol{\mu}^{(t)}$  by (19),  $t = t + 1$ ;
  7. Update the weight  $w_s$  by (14);
- Output the final image.

The exiting condition is as follows:

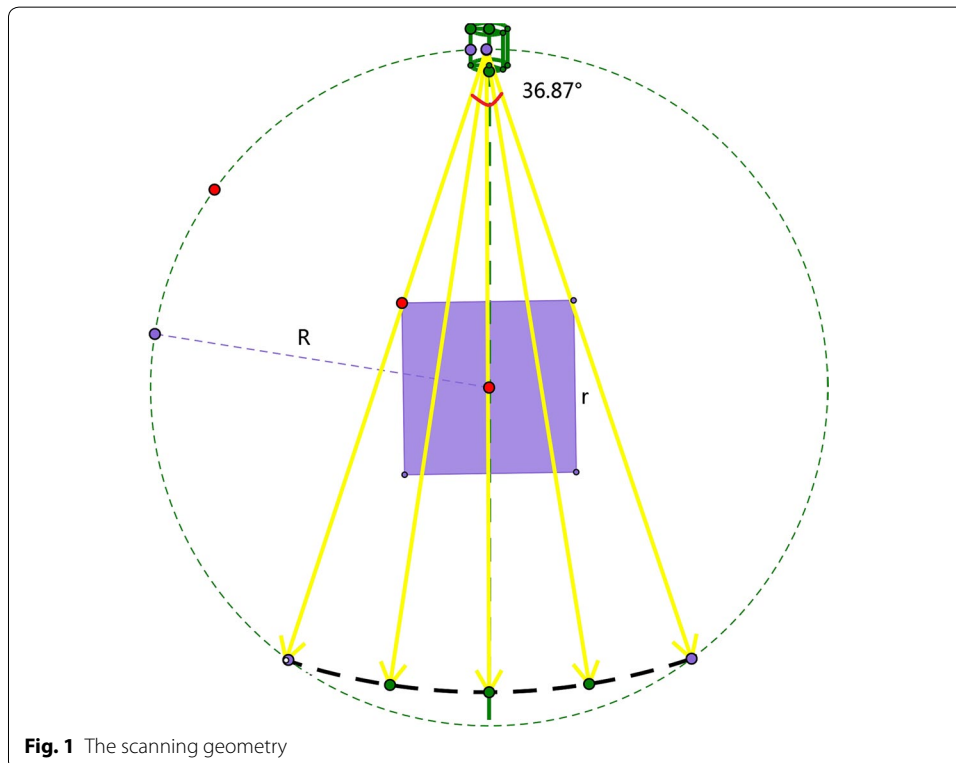
$$\left\{ \begin{array}{l} \frac{|\delta^{(t)} - \delta^{(t-1)}|}{\delta^{(t)}} < \varepsilon_1, \quad \delta^{(t)} = \sum_{i=1}^I \frac{\omega_i}{2} ([\mathbf{A}\boldsymbol{\mu}^{(t)}]_i - \hat{g}_i)^2 \\ \frac{|\eta^{(t)} - \eta^{(t-1)}|}{\eta^{(t)}} < \varepsilon_2, \quad \eta^{(t)} = \sum_{s=1}^S w_s \|\mathbf{E}_s \boldsymbol{\mu}^{(t-1)} - \mathbf{D}^{(t)} \boldsymbol{\alpha}_s^{(t)}\|_2^2 \end{array} \right.$$

Both conditions must be met at the same time, then the iteration process stops.

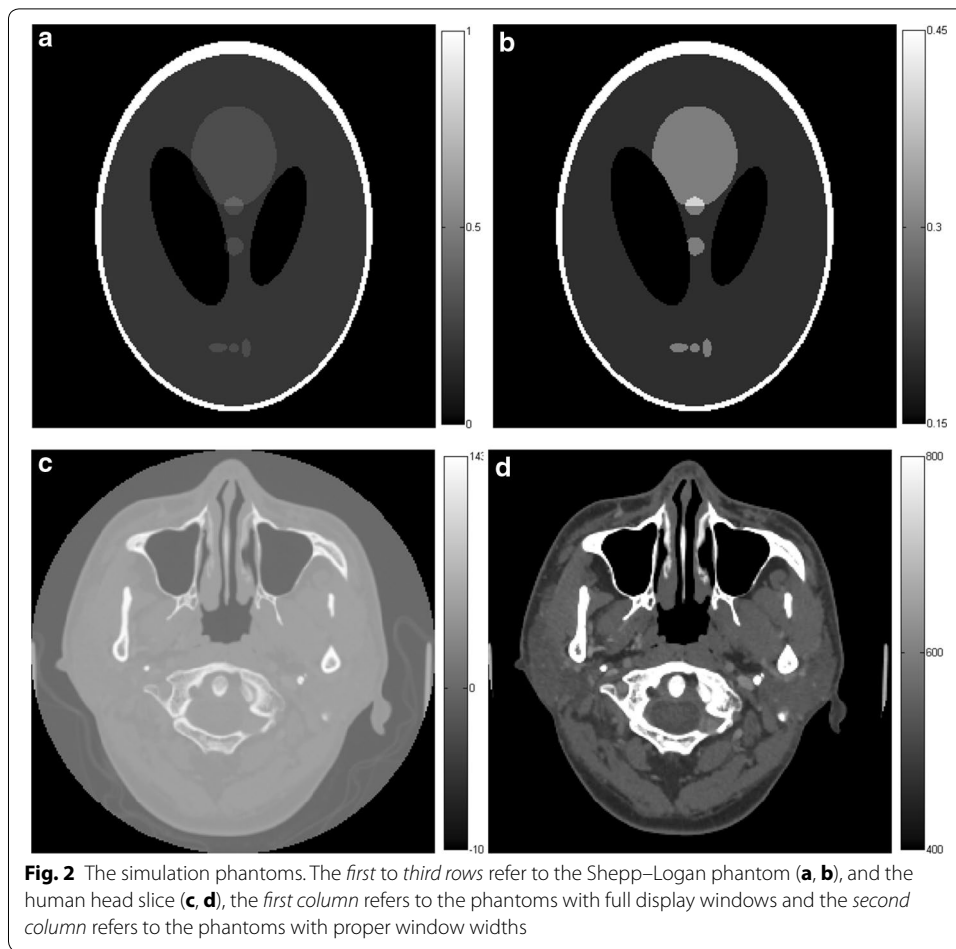
### Simulation results

To verify the effectiveness of the proposed  $L_1$ -DL algorithm on low-dose CT reconstruction, several simulation experiments are designed. All the simulations are performed in MATLAB on a dual-core PC with 3.10 GHz Intel Core i5-2400. The proposed algorithm is compared with SART, GPBB and ADSIR. The scanning geometry is the fan-beam geometry shown in Fig. 1. The size of the phantom is  $r = 20$  cm, the radius of the scanning circle (or the distance from the radiation source to the central point of the phantom) is  $R = 40$  cm. For each projection view, 512 detector elements were equi-angularly distributed with the field angle being  $36.87^\circ$ . The distance from the radiation source to the detector elements is 75.895 cm. During the simulation, the scanning circle covers  $360^\circ$  around the imaging phantom, and the size of the reconstructed image is  $256 \times 256$  pixels. The phantoms under simulations are respectively the Shepp–Logan phantom, and the human head slice from clinic, which are shown in Fig. 2 with full and part display window. The biomedical images are often observed by a proper window width to find more details. The number of different densities in the Shepp–Logan phantom is 6, and we set the density of water as 0.2.

Some details about the proposed algorithm (like the setting of some variables) are explained as follows. The image patches are overlapping with a sliding distance of one pixel, and the patches are with  $N_0 \times N_0$  pixels. The dictionary is consisted of  $K$  atoms. When updating the dictionary by K-SVD algorithm, the sparsity level is  $L_0^D$ . The sparsity level of the sparse representation  $\alpha_s$  is set as  $L_0^S$ . The iterative stopping criterions are  $\varepsilon_1$  and  $\varepsilon_2$ . The values of all the corresponding variables are provided in Table 1. In the image processing field,  $K = 4 \times N_0 \times N_0$  is a conventional choice to ensure the redundancy of the dictionary. The patch size also influences the quality of the image. If the patch size is too small, it could not effectively catch features in an image. On the other hand, a larger





**Table 1** Summary of the parameter selections

No.	Variable	Meaning	Value
1	$N_0$	Scale of the patch	8
2	$K$	Number of atoms in the dictionary	256
3	$L_0^D$	Sparsity level for dictionary learning	5
4	$L_0^S$	Sparsity level for sparse representation	5
5	$\epsilon_1$	Stopping criterion about the projection	0.001
6	$\epsilon_2$	Stopping criterion about the sparse representation	0.001

patch size may lead to an over-smoothed image and corresponds to a larger number of atoms in a dictionary, which would increase the computational cost. For image processing,  $N_0 = 8$  is a proper value with considering both image quality and computational cost [18]. The sparsity level  $L_0^D$  and  $L_0^S$  are empirically determined according to the complexity of the image to be reconstructed.

In addition, all the experiments are simulated by photon number counting method. To perform the OSC algorithm, the transmission data can be calculated with  $y_i = b_i e^{-\hat{g}_i} = e^{-\hat{g}_i}$  by setting the entrance X-ray intensity  $b_i$  as the number of photons. L1-DL and ADSIR set the initial guess of the image as a random matrix while GPBB and SART set all the values of the elements of the initial matrix as 1. OSC is utilized to

accelerate the convergence of L1-DL and ADSIR, with 10 subsets. The number of iterations of OSC is 30. GPBB and SART are not accelerated by OSC. GPBB and SART stop the iteration when the number of iterations reaches 1000. ADSIR stops when the conditions below are met at the same time:

$$\begin{cases} \frac{|\delta^{(t)} - \delta^{(t-1)}|}{\delta^{(t)}} < \varepsilon_1, & \delta^{(t)} = \sum_{i=1}^I \frac{\omega_i}{2} ([\mathbf{A}\boldsymbol{\mu}^{(t)}]_i - \hat{g}_i)^2 \\ \frac{|\eta^{(t)} - \eta^{(t-1)}|}{\eta^{(t)}} < \varepsilon_2, & \eta^{(t)} = \sum_{s=1}^S \|\mathbf{E}_s \boldsymbol{\mu}^{(t-1)} - \mathbf{D}^{(t)} \boldsymbol{\alpha}_s^{(t)}\|_2^2 \end{cases}$$

**Reconstruction of different sparse levels**

In this simulation, the forward simulation and inverse reconstruction are all performed in 2D. During the simulation, the scanning circle covers 360° range around the phantom. The scanning step of tomographic angels of the Shepp–Logan phantom is set as 3°(120 views) and 6°(60 views) respectively. The scanning step of tomographic angels of the human head slice slice is set to 2°(180 views) and 4°(90 views) respectively.

We choose the SART, GPBB, and ADSIR algorithms to be the comparisons besides our proposed L<sub>1</sub>-DL algorithm. The regularization parameter of GPBB is chosen by tests (from 0.1 to 2, the length of step is 0.1). We choose the best one to perform GPBB. When it comes to the DL methods, a proper selection of the regularization parameter λ is a vital problem. A bigger λ weakens the effect of the data fidelity term, generating a loss of some fine details in the image while a smaller λ weakens the effect of the regularization term as the sparse constraint, resulting in more noise and streak artifacts in the reconstructed image. In this article, the regularization parameter of ADSIR is determined by a similar model as Ref. [19] creates. The model firstly reconstructs the image by setting λ as infinite, then calculates the difference between the forward projection and the scanning data. After that, the proper value of λ can be calculated by a fitting function based on the difference. In the proposed method (L<sub>1</sub>-DL), the weights introduced to the algorithm would influence the regularization parameter. To eliminate this effect, we multiply the weights with a constant which is the average value of  $(1/N_0^2) \sum_{i=1}^{N_0^2} |[\mathbf{E}_s \boldsymbol{\mu}^{(t-1)}]_i - [\mathbf{D}^{(t-1)} \boldsymbol{\alpha}_s^{(t-1)}]_i|$ . The weights are calculated by

$$w_s = C / \left[ \left( \frac{1}{N_0^2} \sum_{i=1}^{N_0^2} |[\mathbf{E}_s \boldsymbol{\mu}^{(t-1)}]_i - [\mathbf{D}^{(t-1)} \boldsymbol{\alpha}_s^{(t-1)}]_i| \right) + \varepsilon \right] \tag{20}$$

$$C = \frac{1}{S} \sum_{s=1}^S \frac{1}{N_0^2} \sum_{i=1}^{N_0^2} |[\mathbf{E}_s \boldsymbol{\mu}^{(t-1)}]_i - [\mathbf{D}^{(t-1)} \boldsymbol{\alpha}_s^{(t-1)}]_i|$$

With this modification, we select the regularization parameter of L<sub>1</sub>-DL as the same as the one of ADSIR. The regularization parameters of different phantom simulations are shown in Table 2. The rationality and influence of the regularization parameter selection will be discussed later.

**Table 2 Selections of the regularization parameters with different sampling rates**

	Shepp–Logan		Head slice	
	120 veivs	60 views	180 views	90 views
DL methods	$3.831 \times 10^3$	$3.823 \times 10^3$	380.7	380.1
GPBB	0.2	0.4	0.2	0.5

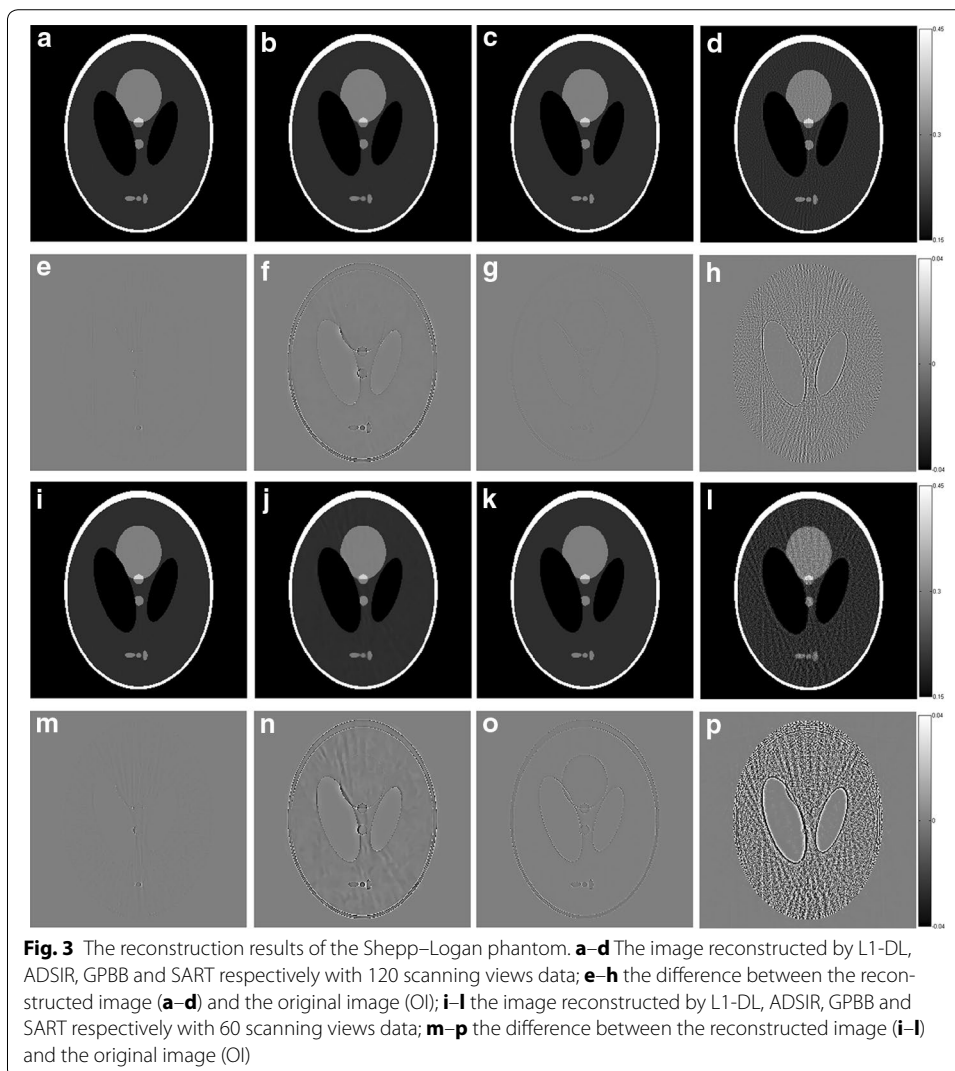
The reconstruction results of the simulations by using these four algorithms are shown in Figs. 3, 4 (Shepp–Logan phantom) and Figs. 5, 6 (human head slice). The negative values in reconstructed image during the iterative process are set to be zero. The results indicate the density of the phantoms. The images of human head slice are displayed by transforming the density to the CT attenuation value. The CT value is calculated by

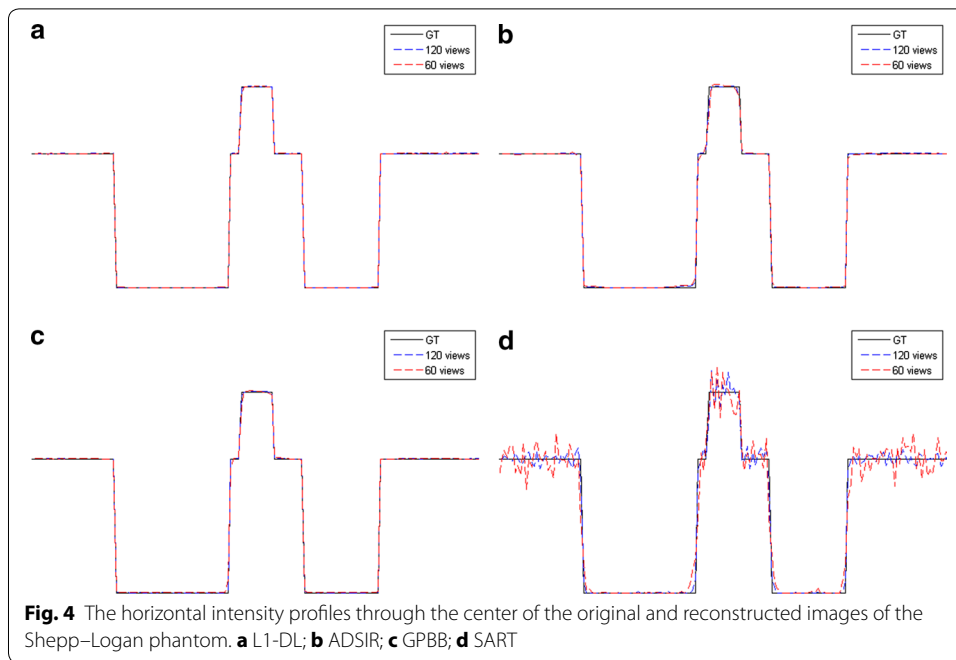
$$\text{CT value} = \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}}} \times 1000 \quad (21)$$

where  $\mu_{\text{water}}$  is the linear attenuation coefficient of water.

By taking the original phantom as a gold standard to provide a numeric quantification of the results, the RMSE (root mean square error) of the reconstructed images is introduced to measure the difference between the reconstructed image and the original image in  $L_2$ -norm. This criterion is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j} (\mu_{ij} - \mu_{ij}^{\text{truth}})^2}{N^2}} \quad (22)$$

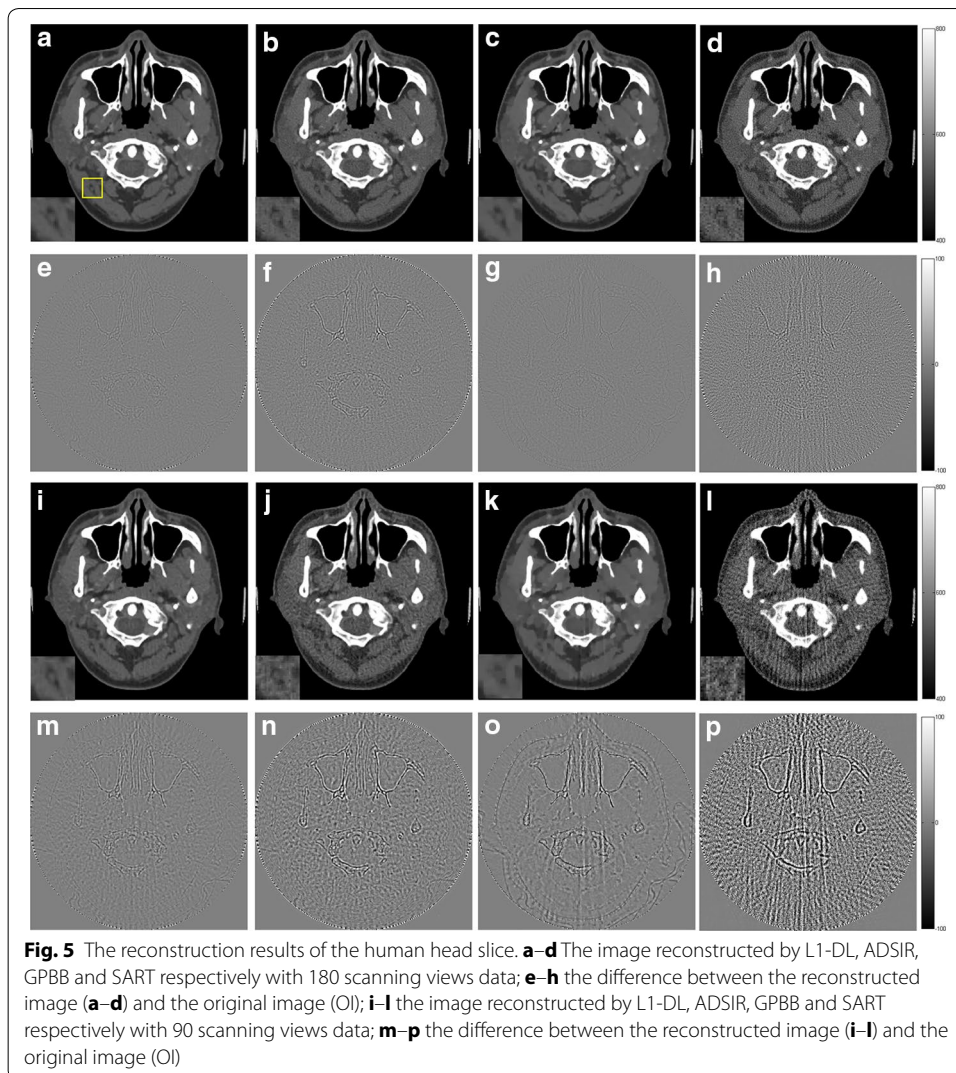




where  $\mu_{ij}^{truth}$  means the gray-value of the original image. The unit of RMSE can be HU by transform the linear attenuation coefficient to CT value according to Eq. (21). It is easy to demonstrate that the smaller the value of RMSE is, the better quality of the image is. The quantitative results are shown in Table 3.

In Figs. 3 and 5, all the reconstructed images of the same phantom are shown in the same proper display window. In Fig. 3, the results of SART are the worst among the four algorithms. Even with 120 sampling views, the result is ruined by the streak artifacts, which deteriorates seriously when the sampling rate decline to half. The reason for the bad results of SART is that this algorithm is a simple iterative algorithm that has no regularization term to utilize the prior information about the reconstructed image. Compared to SART, GPBB and ADSIR perform better with help of the regularization term utilizing the sparse constraint. Figure 3f, n contain some edge structures, which proves the over-smoothing effect of ADSIR. When projection views reduce to 60, the image reconstructed by ADSIR is influenced by some artifacts. L<sub>1</sub>-DL and GPBB perform well in the 120 views situation. The RMSE of these two methods are 1.647HU and 2.658HU respectively, which are both tiny. And Fig. 3e, g reveal that the difference is hard to recognize. When the sampling views reduce, tiny edge structures emerge in Fig. 3o and the RMSE of GPBB increases to 11.04. Figure 4 display the horizontal intensity profiles through the center of reconstructed images compared to the original image. It is shown that the effects of the reduction of the sampling views are arranged as: SART > ADSIR > GPBB > L<sub>1</sub>-DL. In Table 3, the RMSE of L<sub>1</sub>-DL under 60 views situation is much bigger than the RMSEs of ADSIR and SART with 120 views situation, which certifies that the proposed algorithm can adapt to further radiation dose reduction Table 4.

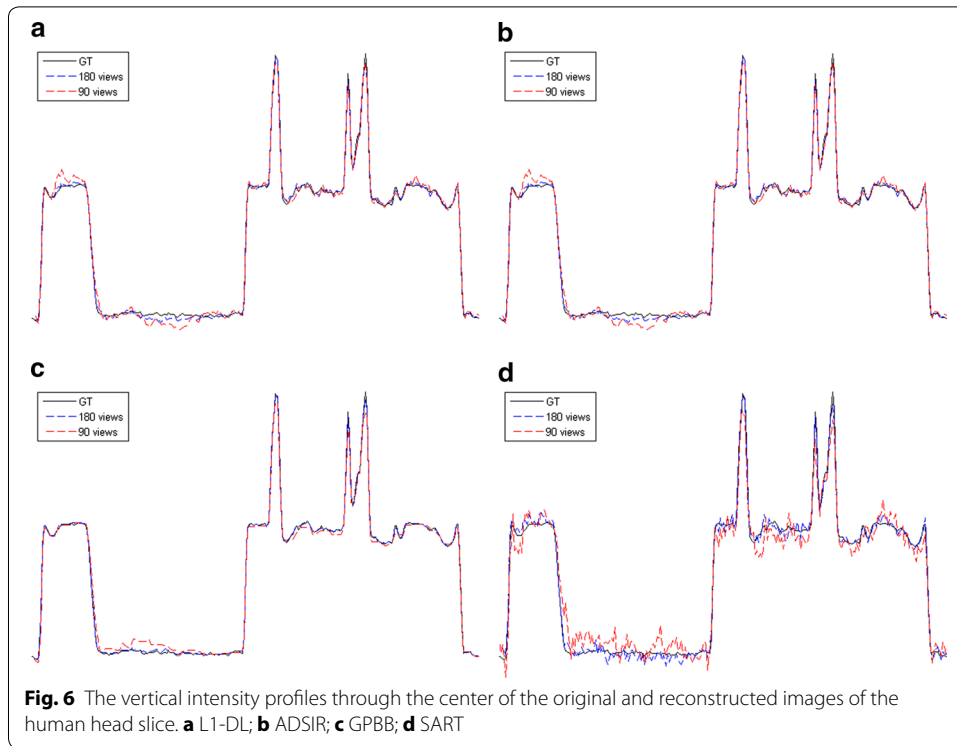
When it comes to the human head slice in Figs. 5 and 6, the results are similar to the Shepp-Logan phantom. But the original image of the human head slice is more complex



than the Shepp–Logan phantom, and is close to the real reconstruction process of the biomedical images. Although the RMSE of the proposed algorithm is a little bigger than GPBB in 180 views situation,  $L_1$ -DL is better than GPBB when the number of views reduces to 90. In 90 views situation, the image of GPBB, displayed as Fig. 5k, has some artifacts and the spatial resolution is worse than  $L_1$ -DL by checking the enlarged region of the images. In addition, the RMSEs of the proposed algorithm is much better than the ones of ADSIR. Considering the proposed algorithm is modified based on the reconstruction model of ADSIR, the improvement of image quality is obvious.

#### Robustness to the noise

In the practical applications, the measurements of the radiation projections are usually polluted by noise, which demands that the algorithm should be robust to the noise. To evaluate the tolerance to noise of the proposed algorithm and other three ones, we choose to add some Poisson noise to the projection data for test. The scanning step of tomographic angles of the Shepp–Logan phantom is set to  $6^\circ$  (60 views) and



**Table 3** RMSEs (HU) of reconstructed images of different sampling rates

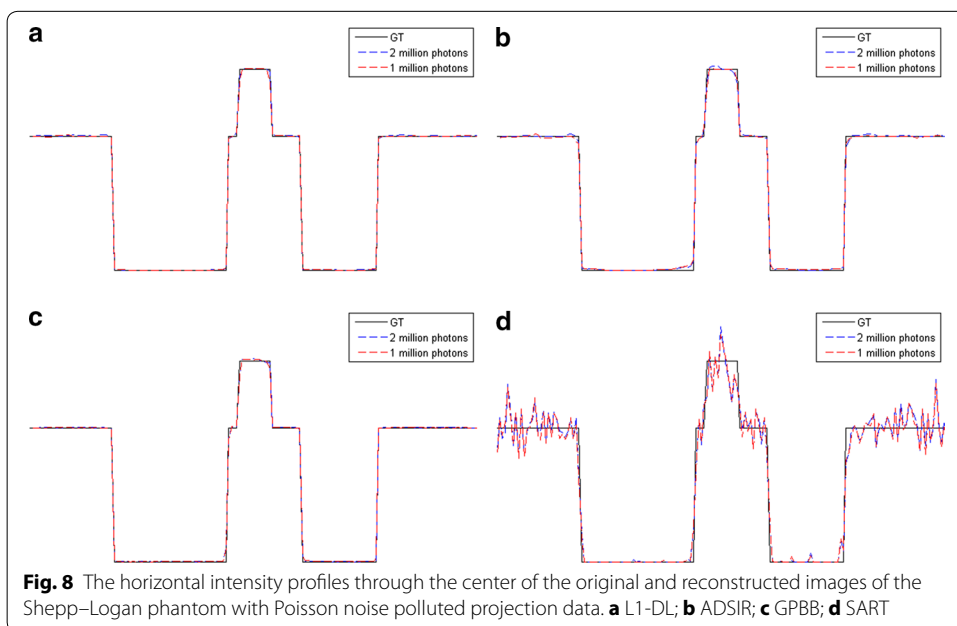
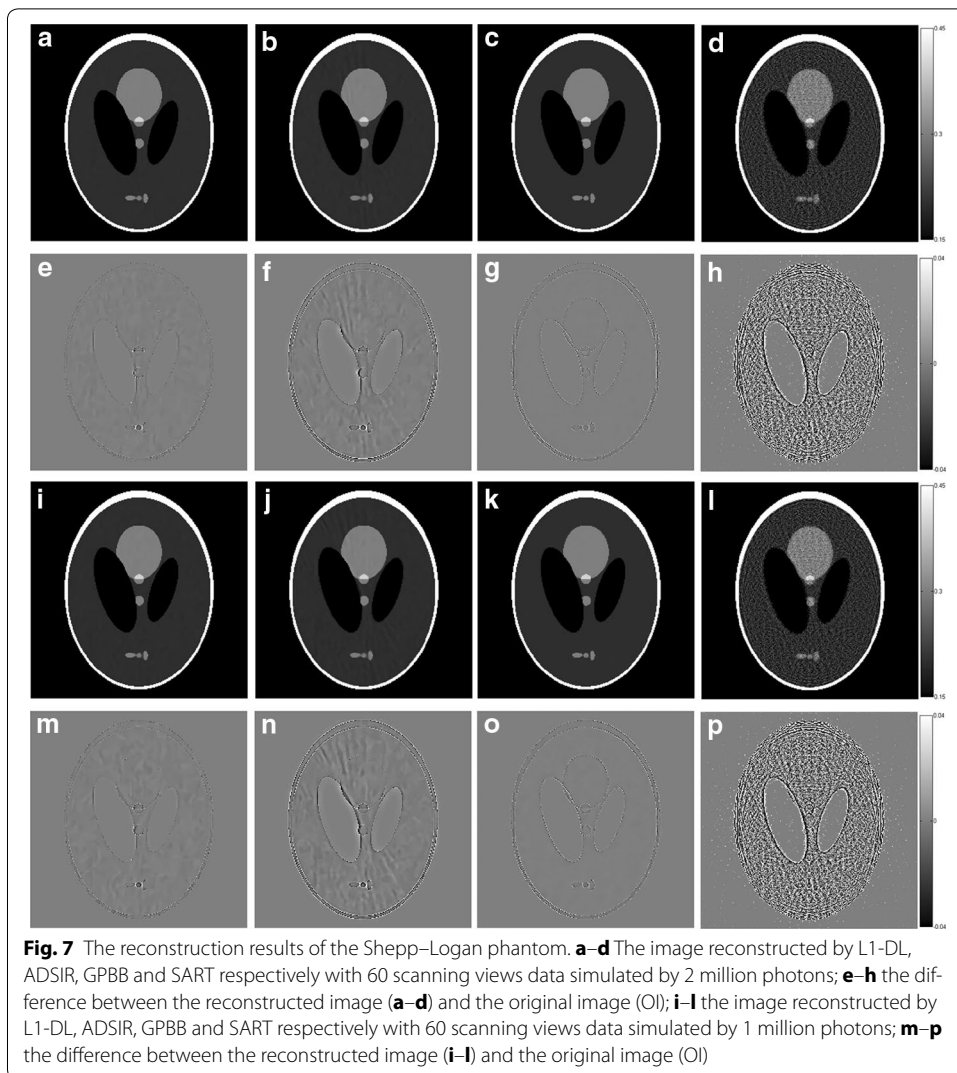
Phantom	L <sub>1</sub> -DL	ADSIR	GPBB	SART
Shepp–Logan				
120 views	1.647	22.62	2.658	34.73
60 views	2.867	31.72	11.04	94.62
Head slice				
180 views	6.228	15.50	4.282	17.87
90 views	10.28	25.55	14.78	41.42

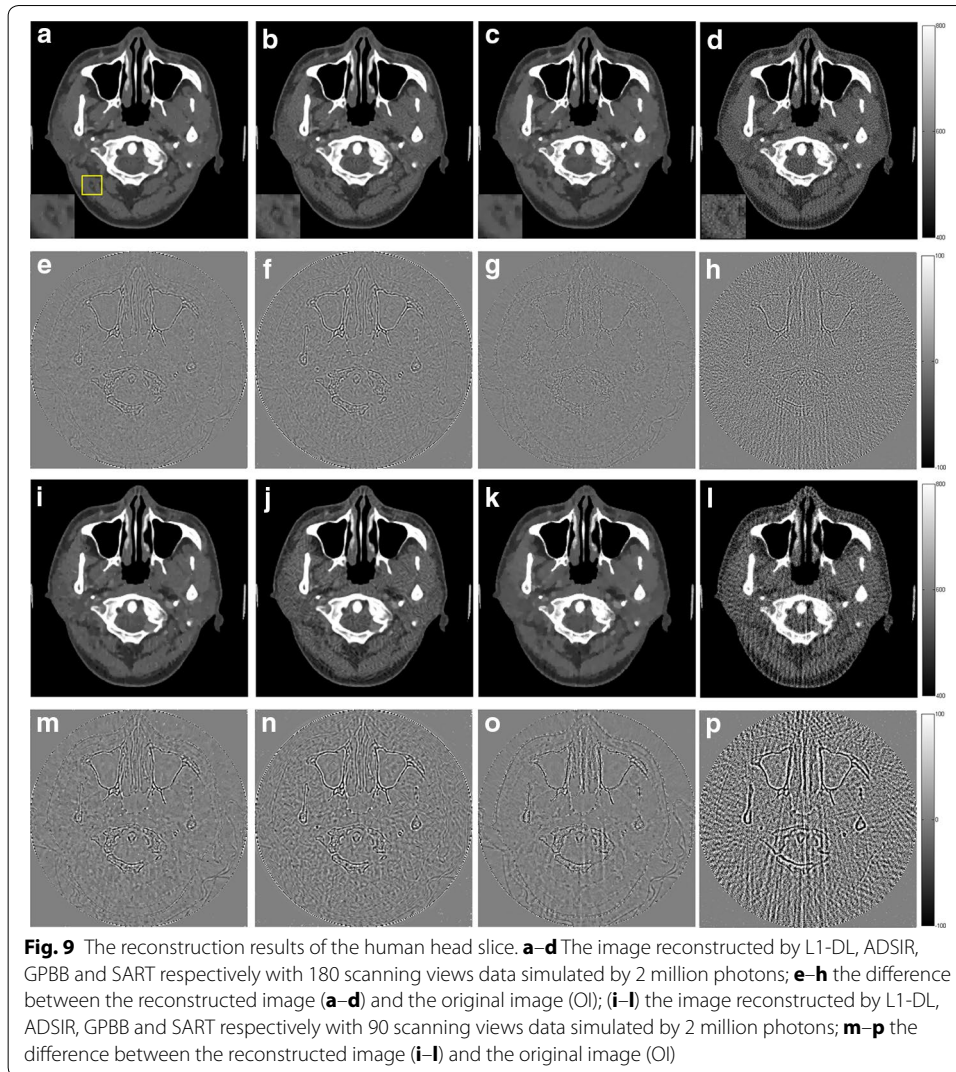
**Table 4** Selections of the regularization parameters with different noise levels

	Shepp–Logan		Head slice	
	60 views 2 million photons	60 views 1 million photons	180 views 2 million photons	90 views 2 million photons
DL methods	$4.841 \times 10^3$	$8.141 \times 10^3$	$1.451 \times 10^3$	$1.725 \times 10^3$
GPBB	0.4	0.5	0.5	0.6

the simulation numbers of photons emitting from the X-ray source to each detector are 2 million and 1 million. The detected numbers of photons are polluted by Poisson noise. The scanning step of the human head slice is set to 2°(180 views) and 4°(90 views) respectively. The simulation number of photons is 2 million.

The reconstructed images of the simulations are shown in Figs. 7, 8, 9, 10 and the RMSEs are provided in Table 5. Compared to the results with no noise polluted in Figs. 3, 4, 5 and 6, the image quality of SART degrades fast with the noise level increasing



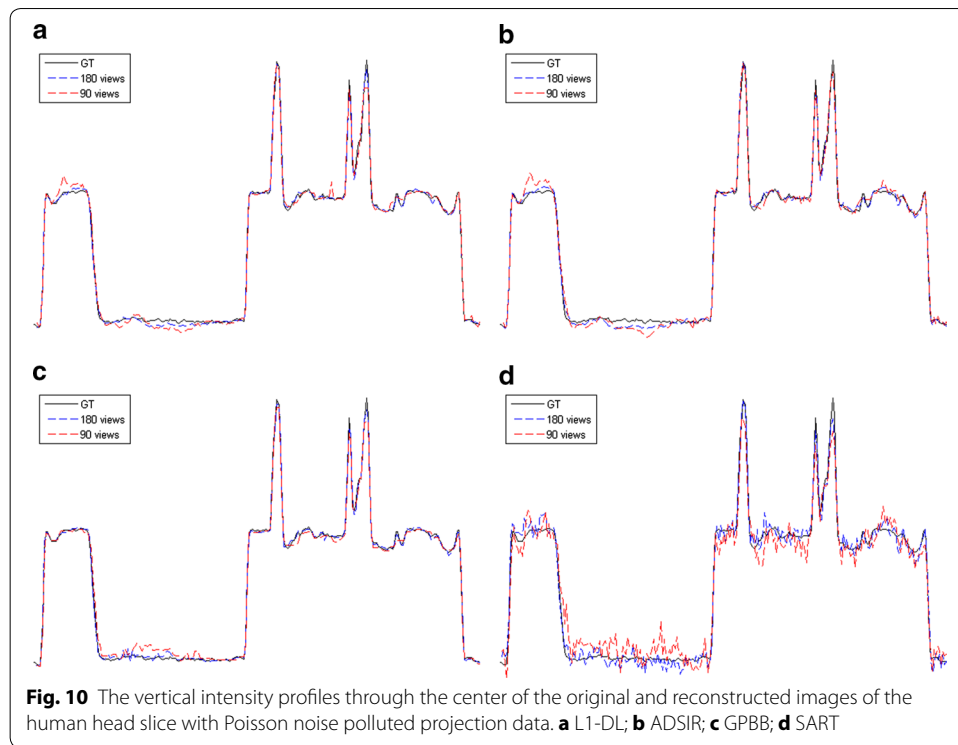


while GPBB,  $L_1$ -DL and ADSIR are more robust to the noise. In the Shepp–Logan experiments,  $L_1$ -DL is better than GPBB and ADSIR comparing the RMSEs. The difference between the reconstructed image and original image shown in Fig. 7 indicates the image quality is arranged as:  $L_1$ -DL > GPBB > ADSIR > SART. In the head slice simulations, the RMSE of SART is a bit smaller than ADSIR, but the image reconstructed by SART are ruined by noise and artifacts. When it comes to  $L_1$ -DL and GPBB, the RMSE of GPBB is better than  $L_1$ -DL in 180 views situation, and the RMSE of these two methods are same in 90 views situation. The results (Fig. 9i–l) of the 90 views simulation with noise indicate that all the four algorithms lose the structure in the yellow rectangle region.  $L_1$ -DL is still better than ADSIR and has a bit higher spatial resolution than GPBB.

#### Convergence rate

To explain the convergence rate of  $L_1$ -DL compared to ADSIR, the RMSEs of the images reconstructed by these two algorithms are shown as functions of the number of

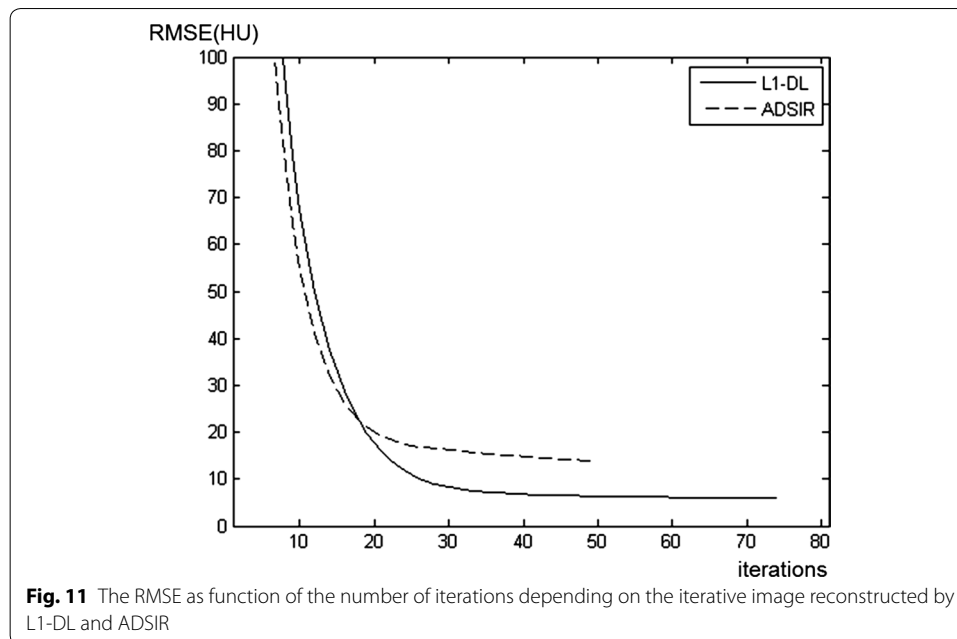




**Table 5 RMSEs (HU) of reconstructed images of different noise levels**

Phantom	L <sub>1</sub> -DL	ADSIR	GPBB	SART
Shepp–Logan				
60 views, 2 million photons	10.87	31.3715	14.78	139.4
60 views, 1 million photons	11.68	34.84	15.43	139.9
Head slice				
180 views, 2 million photons	13.46	24.81	9.582	22.86
90 views, 2 million photons	16.83	30.54	16.83	44.49

iterations in Fig. 11. The projection data are simulated on the human head slice with 180 scanning views, which is not polluted by noise. L<sub>1</sub>-DL stops at the 72th iteration while ADSIR stops at the 49th iteration. Since the iterative process of L<sub>1</sub>-DL only has one additional step, which is updating the weight function, it takes almost the same time as ADSIR to perform one iteration. The times consumed by one iteration of ADSIR and L<sub>1</sub>-DL are 89 and 92 s respectively. By checking the first 20 iterations, it can be found that the RMSE decreasing rate of ADSIR is only a bit faster than L<sub>1</sub>-DL at first. So it is claimed that L<sub>1</sub>-DL needs more iterations than ADSIR for convergence because of the better reconstructed result, but the convergence rate of the two algorithms are almost the same. However, it takes several tens of minutes for both ADSIR and L<sub>1</sub>-DL to reach the stopping points, so that some accelerating methods and high computation efficiency are the expectations for the real time imaging.



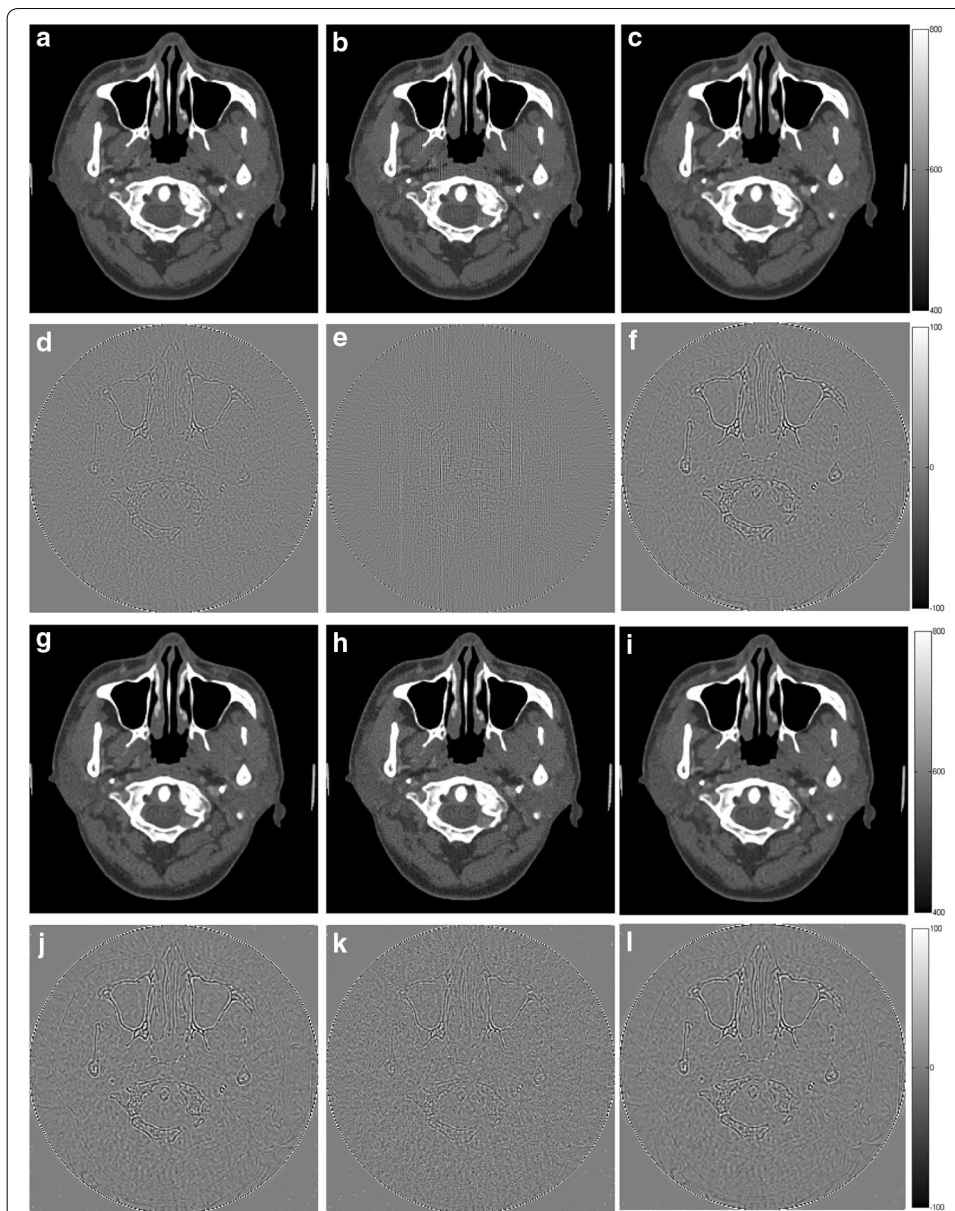
### Regularization parameter investigation

In this part, the rationality and influence of the regularization parameter are discussed in detail. To verify that the selections of this parameter in former text are rational and explore the influence to the results by other parameter values, the images are reconstructed by different regularization parameters. The phantom under simulation is the human head slice in Fig. 2. The projection data are 180 views with no noise, 180 views with noise simulated by 2 million photons. To each projection model, the regularization parameters are chosen as the one calculated by the fitting function (cited as  $\lambda$ ), the one multiplied by 0.2 (cited as  $0.2\lambda$ ) and the one multiplied by 5 (cited as  $5\lambda$ ). The results of ADSIR are shown in Figs. 11 and 12. The RMSEs of the results are shown in Table 6.

From the data in Table 6, we can see the images reconstructed by a bigger regularization parameter ( $5\lambda$ ) will have larger RMSEs, which means the reconstructed images are influenced by the smoothing effect and fewer image details are reserved. When this parameter is set smaller ( $0.2\lambda$ ), the RMSEs might be a little smaller than the ones reconstructed by the proper regularization parameter, such as the data as italics in Table 6 shows. However, the images in Figs. 11 and 12 (the second column) are ruined by noise or streak artifacts in different extents. Therefore, it is important to select the proper value of the regularization parameter for the DL-based methods and the simulations in this section prove that the selections in this article are rational.

### Discussion

The simulation results indicate that the proposed  $L_1$ -DL algorithm is a useful and robust method for the sparse CT reconstruction.  $L_1$ -DL utilizes more prior information of image sparsity than ADSIR benefited by the  $L_1$ -norm DL regularization term.  $L_1$ -DL is an improved method of ADSIR, and the simulation results demonstrate the image quality improvement of  $L_1$ -DL than ADSIR.



**Fig. 12** The reconstruction results of the human head slice by ADSIR with different regularization parameter. **a–c** The image reconstructed by 180 scanning views data with no noise; **d–f** the difference between the reconstructed image (**a–c**) and the original image (OI); **g–i** the image reconstructed by 180 scanning views data simulated by 2 million photons; **j–l** the difference between the reconstructed image (**g–i**) and the original image (OI). The first to third columns refer to the results with  $\lambda$ ,  $0.2\lambda$  and  $5\lambda$  respectively

**Table 6** RMSEs (HU) of reconstructed images of different regularization parameters

Noise level	RMSE (HU)		
	$\lambda$	$0.2\lambda$	$5\lambda$
None	16.18	10.25	25.18
2 million photons	24.81	17.85	27.11

By comparing the simulation results, another comparison algorithm, GPBB, is a little better than  $L_1$ -DL in some situations when comparing the RMSEs. However, GPBB is not good at reserving the edge details and structures, especially when the sampling rate reduces further. It is claimed that  $L_1$ -DL has a higher spatial resolution than GPBB.

In the simulation to explain the convergence rate of  $L_1$ -DL,  $L_1$ -DL stops at the 72th iteration while ADSIR stops at the 49th iteration and the time of one iteration is about 90 s. So it consumes about 108 and 74 min for the reconstruction of  $L_1$ -DL and ADSIR. When it comes to the reconstruction processes of GPBB and SART with the same projection data, the time of one iteration of these two algorithm is about 1 s. Since GPBB and SART stops when the iteration number reaches 1000, the reconstruction time is about 17 min. In addition, the iteration process of GPBB and SART can be accelerated by GPU, so that the time can be reduced to less than a minute. Accelerating  $L_1$ -DL, ADSIR and other DL-based methods is an important factor for the practical application of the DL-based methods.

## Conclusion

In this work, we propose to replace the  $L_2$ -norm regularization term with the  $L_1$ -norm one to improve image quality reconstructed by the DL-based method. The new objective function is optimized by the adaptive weighted  $L_2$ -norm strategy, which is similar to the IRLS algorithm. By involving this modification, the proposed  $L_1$ -DL algorithm behaves better than the existing DL-based method (ADSIR), and other two comparing algorithms. Experimental results show that the proposed algorithm can satisfy the demand of further radiation reduction in CT scanning since it needs fewer scanning data for high-quality recovery. In addition, the proposed algorithm retains the robust characteristic to the projection noise as a DL-based algorithm. Our future work will focus on two aspects. One of them is accelerating the DL-based methods to make the real time imaging with low-dose radiation possible. The other is looking for some possible strategies to utilize more prior information and further improve the image reconstruction result. For example, by utilizing a proper way to distinguish structural information and noise in the image, the DL regularization term can be designed based on the distinguishing results, which is a promising method to preserve more structural information and improve the image quality.

## Abbreviations

CT: computed tomography; DL: dictionary learning; IRLS: iteratively reweighted least squares; ADSIR: adaptive dictionary based statistical iterative reconstruction; PLWS: penalized weighted least-squares; FBP: filtered backprojection algorithm; SART: simultaneous algebraic reconstruction technique; TV: total variation; DGT: discrete gradient transform; ASD-POCS: adaptive steepest descent projection onto convex sets; GPBB: gradient projection Barzilai Borwein; IRLS: iteratively reweighted least squares; SIR: statistical iterative reconstruction; ADSIR: adaptive dictionary based statistical iterative reconstruction; SNR: signal to noise ratios; K-SVD: k singular value decomposition; OMP: orthogonal matching pursuit; RMSE: root mean square error.

## Authors' contributions

Study concept and design (CZ and ML); drafting of the manuscript (CZ); critical revision of the manuscript for important intellectual content (CZ and JZ); obtained funding (JZ); administrative, technical, and material support (CZ, JZ and TZ); study supervision (CP and ZL). All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Suzhou Institute of Biomedical Engineering and Technology of Chinese Academy of Sciences, Suzhou 215163, China.

<sup>2</sup> Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China.

<sup>3</sup> University of Chinese Academy of Sciences, Beijing 100049, China. <sup>4</sup> Department of Electronic Science and technology, University of Science and Technology of China, Hefei 230061, China.

### Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.61301042 and No.61201117), the Natural Science Foundation of Jiangsu Province (No. BK20151232), the Science and Technology Program of Suzhou (No. ZXY2013001), and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (No.2014281).

### Competing interests

The authors declare that they have no competing interests.

Received: 7 December 2015 Accepted: 17 May 2016

Published online: 18 June 2016

### References

1. Berrington de González A, Mahesh M, Kim K, Bhargavan M, Lewis R, Mettler F, Land C. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Arch Intern Med*. 2009;169(22):2071–7. doi:10.1001/archinternmed.2009.440.
2. Wang J, Li T-F, Lu H-B, Liang Z-R. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. *IEEE Trans Med Imaging*. 2006;25(10):1272–83.
3. Liao H-Y, Sapiro G. Sparse representations for limited data tomography, in biomedical imaging: from nano to macro, 2008. ISBI 2008. 5th IEEE international symposium on 2008; pp. 1375–78. doi:10.1109/ISBI.2008.4541261.
4. Jerri AJ. The Shannon sampling theorem—its various extensions and applications: a tutorial review. *Proc IEEE*. 1977;65(11):1565–96.
5. Song B, Park J, Song W. A novel, fast, variable step size gradient method for solving simultaneous algebraic reconstruction technique (SART)-type reconstructions: an example application to CBCT. *Med Phys*. 2011. doi:10.1118/1.3611782.
6. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory*. 2006;52(2):489–509.
7. Candes EJ, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans Inf Theory*. 2006;52(12):5406–25.
8. Sidky EY, Pan X-C. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol*. 2008;53(17):4777–807.
9. Justin CP, Song B-Y, Kim JS, Park SH, Kim HK, Liu Z-W, Suh TS, William YS. Fast compressed sensing-based CBCT reconstruction using Barzilai–Borwein formulation for application to on-line IGRT. *Med Phys*. 2012;39(3):1207–17.
10. Xu Q, Yu H-Y, Mou X-Q, Zhang L, Hsieh J, Wang G. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans Med Imaging*. 2012;31(9):1682–97.
11. Tian Z, Jia X, Yuan K, Pan T, Jiang S-B. Low-dose CT reconstruction via edge-preserving total variation regularization. *Phys Med Biol*. 2011;56(18):5949–67.
12. Endra O, Gunawan D. Comparison of l1-minimization and iteratively reweighted least squares-l p-minimization for image reconstruction from compressive sensing. In *advances in computing, control and telecommunication technologies (ACT)*, 2010 second international conference on 2010. doi:10.1109/ACT.2010.31.
13. Miosso CJ, Borries RV, Arguez M, Velazquez L, Quintero C, Potes CM. Compressive sensing reconstruction with prior information by iteratively reweighted least-squares. *IEEE Trans Signal Process*. 2009;57(6):2424–31.
14. Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process*. 2006;54(11):4311–22.
15. Tropp JA, Gilbert AC. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans Inf Theory*. 2007;53(12):4655–66.
16. Elbakri IA, Fessler JA. Statistical image reconstruction for polyenergetic X-ray computed tomography. *Medical Imaging, IEEE Transactions on*. 2002;21(2):89–99.
17. Kamphuis C, Beekman FJ. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. *IEEE Trans Med Imaging*. 1998;17(6):1101–5.
18. Julien M, Sapiro G, Elad M. Learning multiscale sparse representations for image and video restoration. *Multiscale Model Simul*. 2008;7(1):214–41.
19. Zhang C, Zhang T, Zheng J, Li M, Lu Y-F, You J-L, Guan Y-H. A model of regularization parameter determination in low-dose X-ray CT reconstruction based on dictionary learning. *Computat Math Methods Med*. 2015. doi:10.1155/2015/831790.