

Z-Index Parameterization for Volumetric CT Image Reconstruction via 3-D Dictionary Learning

Ti Bai¹, Hao Yan, Xun Jia¹, Steve Jiang, Ge Wang², *Fellow, IEEE*, and Xuanqin Mou¹

Abstract—Despite the rapid developments of X-ray cone-beam CT (CBCT), image noise still remains a major issue for the low dose CBCT. To suppress the noise effectively while retain the structures well for low dose CBCT image, in this paper, a sparse constraint based on the 3-D dictionary is incorporated into a regularized iterative reconstruction framework, defining the 3-D dictionary learning (3-DDL) method. In addition, by analyzing the sparsity level curve associated with different regularization parameters, a new adaptive parameter selection strategy is proposed to facilitate our 3-DDL method. To justify the proposed method, we first analyze the distributions of the representation coefficients associated with the 3-D dictionary and the conventional 2-D dictionary to compare their efficiencies in representing volumetric images. Then, multiple real data experiments are conducted for performance validation. Based on these results, we found: 1) the 3-D dictionary-based sparse coefficients have three orders narrower Laplacian distribution compared with the 2-D dictionary, suggesting the higher representation efficiencies of the 3-D dictionary; 2) the sparsity level curve demonstrates a clear Z-shape, and hence referred to as Z-curve, in this paper; 3) the parameter associated with the maximum curvature point of the Z-curve suggests a nice parameter choice, which could be adaptively located with the proposed Z-index parameterization (ZIP) method; 4) the proposed

3-DDL algorithm equipped with the ZIP method could deliver reconstructions with the lowest root mean squared errors and the highest structural similarity index compared with the competing methods; 5) similar noise performance as the regular dose FDK reconstruction regarding the standard deviation metric could be achieved with the proposed method using (1/2)/(1/4)/(1/8) dose level projections. The contrast-noise ratio is improved by $\sim 2.5/3.5$ times with respect to two different cases under the (1/8) dose level compared with the low dose FDK reconstruction. The proposed method is expected to reduce the radiation dose by a factor of 8 for CBCT, considering the voted strongly discriminated low contrast tissues.

Index Terms—Dictionary learning, sparse representation, cone-beam CT, noise suppression, regularization parameter.

I. INTRODUCTION

AS A powerful tool to visualize internal structures of an object in a non-invasive fashion, cone-beam CT (CBCT) [1] has been applied in many scenarios, such as patient setup in radiation therapy [2], intraoperative imaging [3], and maxillofacial visualization [4].

Despite the rapid developments, image noise still remains a major issue in low dose CBCT. One of the reasons comes from the demands for radiation dose reduction as low as reasonably achievable (ALARA) due to the potential radiation damage to the human body. Basically, low dose CBCT can be achieved by either collecting fewer projections (few-view protocol) or reducing the exposure level (low-exposure protocol). In this work, we will focus on the low-exposure protocol, as it can be simply implemented by reducing the tube current and is advantageous sampling-wise. Low-exposure protocol, however, would inevitably result in noisy projection data. And data noise would be propagated into reconstructed images, possibly rendering the images less useful or useless.

A great effort has been devoted to image noise reduction. Specifically, by accommodating measurement statistics, modeling data acquisition geometry, and enforcing physical constraints, regularized iterative reconstruction algorithms often produce superior image quality with highly noisy measurements, and hence having become increasingly popular. In the context of iterative reconstruction, an appropriate physical constraint about the underlying image, i.e., the regularizer, is regarded as being of primary importance (e.g., references [5]–[8]). Thanks to the rapid development of

Manuscript received July 31, 2017; revised September 26, 2017; accepted September 27, 2017. Date of publication October 5, 2017; date of current version November 29, 2017. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFA0202003, in part by the National Natural Science Foundation of China under Grant 61571359, in part by China National Thirteen-Five Major Projects of Digital Medical Equipment under Grant 2016YFC0105202, in part by NIH under Grant R01EB016977, Grant 1R01CA154747-01, Grant 1R21CA178787-01A1, Grant 1R21EB017978-01A1, and Grant U01EB017140, and in part by China Scholarship Council. (*Corresponding author: Xuanqin Mou.*)

T. Bai and X. Mou are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing 100048, China (e-mail: tibaiw@163.com; xqmou@mail.xjtu.edu.cn).

H. Yan is with Cyber Medical Corporation, Shaanxi 710049, China (e-mail: hao.yan@cybermedical.com).

X. Jia and S. Jiang are with the Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX 75390 USA (e-mail: xun.jia@utsouthwestern.edu; steve.jiang@utsouthwestern.edu).

G. Wang is with the Biomedical Imaging Center/Cluster, CBIS, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: wanggg@rpi.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2017.2759819

compressive sensing theory [9], the sparsity-promotion regularizers have been successful, most of which could be applied on both the few-view and low-exposure protocols. For example, Sidky *et al.* [10] and Yu *et al.* [11] proposed an iterative reconstruction algorithm by minimizing the total variation (TV) of the image. Tang *et al.* [12] compared the TV based reconstruction method with the well-accepted penalized weighted least squares method [13] and the q-Generalized Gaussian Markov Random Field (g-GGMRF) method [14]. Yan *et al.* [15] proposed an adaptive-weighted TV regularizer for better edge reservation performance. Provided a high-quality image which resembles the image under reconstruction, Chen *et al.* [6] developed a method referred to as prior image constrained compressive sensing (PICCS) for accurate reconstruction of dynamic CT images. In low dose CBCT domain, many iterative reconstruction algorithms were published, with an emphasis on the design of the regularizer [16], [17]. Sidky and Pan [5] developed a 3D-TV minimization method for volumetric image reconstruction from a circular CBCT scan, which is referred to as adaptive-steepest-descent-projection-onto-convex sets (ASD-POCS) algorithm. Jia *et al.* [16] constructed an iterative CBCT reconstruction framework regularized by the tight frame (TF) based sparse representation, attaining competitive performance as the TV minimization method. Considering the correlations of the images in different energy channels, Gao *et al.* [18] enforced a low rank constraint among the images so as to enable high-quality reconstructions.

Recently, learning based image processing techniques gained significant interest, with a primary example known as dictionary learning [7], [19], [20]. The basic idea is a well-accepted assumption that in natural scene images, there exist abundant structures which could be sparsely represented with a redundant dictionary. Xu *et al.* [7] incorporated a dictionary learning based sparse constraint into the statistical x-ray CT iterative reconstruction framework. Li *et al.* [21] combined dictionary learning and TV minimization based sparse constraints together to facilitate dual-energy CT reconstruction.

Currently, most of the dictionary learning based sparse representation techniques are for 2D cases. Intuitively, 3D structures in volumetric images should be directly targeted by training a 3D dictionary, which consists of 3D atoms. It is expected that the trained 3D dictionary could represent the structures more efficiently by exploiting the spatial correlations in all the three dimensions simultaneously. And consequently, this higher representation efficiency could also facilitate the later denoising task better. Motivated by this, in 2014, we reported a 3D dictionary learning (3DDL) based reconstruction framework for low dose volumetric CT in the 56th AAPM Annual Meeting & Exhibition [22]. In this paper, we will expand our previous idea. Besides, we will also further analyze the representation efficiency differences between the 2D/3D dictionaries. This could be used to justify the potential denoising performance of the proposed 3DDL method.

Regularization parameter is another critical issue in the regularized iterative reconstruction framework. It is well-accepted that the noise and resolution properties highly depend on the selection of the regularization parameter. Numerous

selection strategies have been proposed, such as the discrepancy principle [23], generalized cross-validation [24], Stein's unbiased risk estimate [25], and L-curve [26]. Specifically, under the spirit of the discrepancy principle, by enforcing the fidelity error to be compatible with a predetermined tolerance, an accelerated barrier optimization compressed sensing (ABOCS) reconstruction method was proposed to facilitate CBCT reconstruction with adaptive parameter selection [8], [17]. In fact, the above reviewed ASD-POCS algorithm also employed similar idea to adaptively select the step size of the steepest descent. On the other hand, in intensity-modulated radiation therapy, Zhu and Xing [27] used the Pareto frontier to make a balanced trade-off between the dose performance and the segment number, where the Pareto frontier is constructed based on the solutions of two different objective functions. The idea of the L-curve rule is similar as the above Pareto frontier strategy. Basically, in the iterative reconstruction context, the L-curve is shaped by the fidelity terms and the regularization terms associated with different regularization parameter choices. And the desired parameter is considered as the one corresponding to the "corner" of the L-curve. However, there are two disadvantages in the L-curve method. First, in order to visualize and locate the "corner" better, one often needs to plot the L-curve in a different scale, such as log-log/sqrt-sqrt scales, and the scale function is also case-dependent [28]. Second, the resultant L-curve is *implicitly* correlated with the regularization parameter, making it inconvenient to develop an adaptive selection strategy. Realizing that our trained 3D dictionary may be an efficient feature descriptor, it could be used to construct an informative regularizer which can indicate the potential image qualities independently. Therefore, in this work, based on our constructed 3DDL based sparse regularizer, we will also propose an adaptive regularization parameter selection rule, which is *explicitly* correlated with the regularization parameter.

In summary, in this paper, in order to justify the proposed 3DDL method, we will first evaluate the representation efficiencies of the 2D/3D dictionaries for spatial structures. Then, a new adaptive regularization parameter selection method would be described. Based on these two results, we will qualitatively and quantitatively compare the performance of the proposed 3DDL method to other competing algorithms, such as the previous 2D dictionary learning method and the TF method. And we will also explore the dose reduction potential of our 3DDL method. Moreover, to make the proposed method clinically practical, we parallelize the whole program on the graphic processing units (GPU) using several algorithmic tricks.

II. METHODS AND MATERIALS

A. Formulation

For presentation, we will first formulate the iterative reconstruction and introduce related notations. Basically, the objective of an image reconstruction is to find the unknown true image $\hat{\mathbf{x}} \in \mathbb{R}^{N \times 1}$ from observed measurements $\mathbf{y} \in \mathbb{R}^{M \times 1}$ (the transmission data through log transform) defined by $\mathbf{y} = \mathbf{A}\hat{\mathbf{x}} + \boldsymbol{\zeta}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the system matrix, $\boldsymbol{\zeta} \in \mathbb{R}^{M \times 1}$

denotes the noise which can be modeled as a zero-mean Gaussian distribution with ray-dependent variances [29], M and N are the amount of the rays and pixels, respectively. By incorporating certain physical constraints, the regularized statistical iterative reconstruction is formulated as

$$\min_{\mathbf{x}} \phi(\mathbf{x}) + \beta \mathbf{R}(\mathbf{x}), \quad (1)$$

where β denotes the regularization parameter controlling the relative weight between the fidelity term $\phi(\mathbf{x})$ and the regularization term $\mathbf{R}(\mathbf{x})$. A common choice about the fidelity term is $\phi(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\mathbf{W}}^2$, where $\|\mathbf{u}\|_{\mathbf{W}}^2 \triangleq \mathbf{u}^T \mathbf{W} \mathbf{u}$, \mathbf{T} represents the transpose operator, $\mathbf{W} = \text{diag}\{w_{ii}\} \in \mathbb{R}^{M \times M}$ is a diagonal matrix consisting of the statistical weights that are inversely proportional to the measurements' variances [29].

Specifically, the dictionary learning based statistical iterative reconstruction [7] can be written as

$$\begin{aligned} \min_{\mathbf{x}, \alpha_s} & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\mathbf{W}}^2 + \beta \sum_s \|\alpha_s\|_0, \\ \text{s.t.}, & \|\mathbf{E}_s \mathbf{x} - \mathbf{D}\alpha_s\|_2^2 \leq \epsilon, \quad \forall s, \end{aligned} \quad (2)$$

where \mathbf{E}_s denotes the extraction operator for the s^{th} data block which can be sparsely represented by a learned dictionary \mathbf{D} , and the associated coefficients are α_s , ϵ denotes the tolerance. Note that each column of the dictionary \mathbf{D} is a vector rearranging of an atom.

B. Optimization Algorithm

The solution of problem (2) could be obtained by alternatively iteratively solving the following two sub-problems:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\mathbf{W}}^2 + \beta \sum_s \|\mathbf{E}_s \mathbf{x} - \mathbf{D}\alpha_s\|_2^2, \quad (3)$$

$$\min_{\alpha_s} \|\alpha_s\|_0, \quad \text{s.t.}, \|\mathbf{E}_s \mathbf{x} - \mathbf{D}\alpha_s\|_2^2 \leq \epsilon, \quad \forall s. \quad (4)$$

Sub-problem (3) is of simple quadratic form, which can be optimized by the order subsets based separable quadratic surrogate (OS-SQS) method [30]:

$$\mathbf{x}^{j+1} = \mathbf{x}^j - \frac{V \mathbf{A}_C^T \mathbf{W}_C (\mathbf{A}_C \mathbf{x}^j - \mathbf{y}_C) + \beta \sum_s \mathbf{E}_s^T (\mathbf{E}_s \mathbf{x}^j - \mathbf{D}\alpha_s)}{\mathbf{A}^T \mathbf{W} \mathbf{A} + \beta \sum_s \mathbf{E}_s^T \mathbf{E}_s \mathbf{I}}, \quad (5)$$

where V is the number of the subsets, subscript C denotes one subset of the projections, superscript j denotes the j^{th} iteration, \mathbf{I} is the unity vector. In this work, to accelerate the above algorithm, two additional algorithmic tricks are employed, i.e., Nesterovs weighting strategy [31] and double surrogates strategy [32].

Sub-problem (4) is a typical sparse coding task. Here, the Cholesky decomposition based orthogonal matching pursuit (OMP) algorithm is considered as the solver [33]. For the OMP algorithm, in this work, the maximum amount of atoms used for sparse coding is set to be 8, which is a good and robust choice according to our experience. Specifically, to address the bottleneck that the DL based sparse coding stage is much slower than the conventional regularizers such as TV minimization, an optimized GPU computing strategy is implemented by assigning each thread block to be in charge

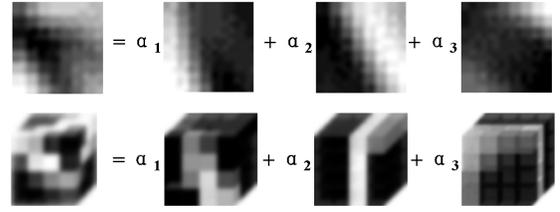


Fig. 1. Illustration of 2D (first row) and 3D (second row) dictionary learning based sparse representations.

of the sparse coding for the corresponding data block. In our implementation, the data block and the dictionary are stored in the shared memory and global memory, respectively. This could dramatically decrease computational cost associated with data access. Detailed GPU implementations could be found in the supplemental material.¹

C. 3D Dictionary Learning

Basically, even for the 3D case, the involved sparse regularizer for the volumetric image still could be constructed based on a 2D dictionary $\mathbf{D} \in \mathbb{R}^{K \times L}$ in a slicewise method, where L denotes the number of the 2D atoms of size $P \times Q$, and $K = P \times Q$. However, intuitively, the 3D structures in the volumetric images should be directly targeted by training a 3D dictionary $\mathbf{D} \in \mathbb{R}^{K \times L}$ comprising of L 3D atoms of $P \times Q \times R$, and $K = P \times Q \times R$.

In fact, the motivation we construct a 3D dictionary learning based regularizer for the volumetric image reconstruction is that, we expect our 3D dictionary could represent the 3D structures more efficiently by capturing the spatial correlations in all the three dimensions simultaneously. On the other hand, in the field of neuroscience, it has been found that the more efficiently the trained dictionary could represent the structures, the sparser the representation coefficients are, and also the more the atoms like the characteristics of the simple cells' receptive fields in human primary visual cortex [34]. Therefore, we also expect our 3D dictionary could result in sparser coefficients, and hence facilitate the later denoising stages better. This assumption would be validated in the following results section.

In this work, based on a number of existing CBCT images, a 3D dictionary $\mathbf{D} \in \mathbb{R}^{64 \times 256}$ containing 256 $4 \times 4 \times 4$ atoms is trained with the open source optimization toolbox sparse modeling software (SPAMS) [35]. This dictionary has the same size as the commonly used 2D dictionary $\mathbf{D} \in \mathbb{R}^{64 \times 256}$ containing 256 8×8 atoms [7]. A simple illustration about the 2D/3D dictionary learning based sparse representation is given in Fig. 1.

D. Regularization Parameter Selection

We first give a brief description about the L-curve method, and then present our selection strategy. Mathematically, an L-curve [26] in g -scale could be expressed as

$$L = \{g(\phi(\mathbf{x}_\beta)), g(\mathbf{R}(\mathbf{x}_\beta))\}, \beta \in \{0, \dots, +\infty\}, \quad (6)$$

¹Please see the supplemental material for the detailed GPU implementation.

where \mathbf{x}_β is the solution of problem (1) equipped with a regularization parameter β . In the L-curve method, the L-shape “corner” location of the above plot is used to determine the desired regularization parameter [26].

As shown in (6), in practice, in order to visualize and locate the “corner” better, one always needs to scale the original L-curve with a positive, monotone increasing function, such as the log-log/sqrt-sqrt scales, and this choice is also case-dependent [28]. Moreover, from (6), it is found that L-curve is *implicitly* correlated with β , and hence is not convenient to develop an adaptive selection strategy.

In this work, we aim to construct a new curve which is *explicitly* correlated with β by devising a new scale function, so as to adaptively guide its selection. Firstly, it has been proven that $\phi(\mathbf{x}_\beta) \triangleq f(\beta)$ is monotone increasing with respect to β [36]. Therefore, we could inverse $\phi(\mathbf{x}_\beta)$ as $f^{-1}(\phi(\mathbf{x}_\beta)) = \beta$. Secondly, we assume that the trained dictionary is an excellent feature descriptor for the spatial structures. As a consequence, it is expected that the constructed dictionary learning based sparse regularizer $\mathbf{R}(\mathbf{x}_\beta) = \sum_s \|\alpha_s\|_0$ itself could be very informative about the image quality of the reconstructions without any scale function. Note that $\mathbf{R}(\mathbf{x}_\beta)$ represents the total non-zero elements in the sparse coefficient matrix. In order to further highlight its physical meaning, one could normalize it as $\psi(\beta) = \frac{\mathbf{R}(\mathbf{x}_\beta)}{\mathbf{B}}$, where $\psi(\beta)$ means the averaged sparsity level among all the data blocks under sparse coding, whose total amount is \mathbf{B} . Therefore, unlike the conventional log-log/sqrt-sqrt scales, our scale function is selected to be $f^{-1} \cdot \mathbf{B}^{-1}$, which is adaptive to the data. Applying our scale function on (6), one could formulate the new curve as:

$$Z = \{\beta, \psi(\beta)\}, \beta \in \{0, \dots, +\infty\}. \quad (7)$$

Because our newly constructed curve could be divided into the following three stages, exhibiting a clear Z-shape, therefore it is referred to as “Z-curve” hereafter in this work. Basically, a small β would result in a highly noisy reconstruction with low dose projections, for whom, one needs to use almost the maximum allowed amount of atoms to well represent the extracted noisy data blocks.

1) Steady-Descent Stage: Considering one gradually increases β from a small value, due to the existing high noise, the required amount of atoms may be still around the maximum amount or slightly less. This process would shape a level-up or steady-descent stage in the Z-curve.

2) Steep-Descent Stage: When the noise of the data block is reduced to a certain level under a relative large β , such as slightly less than the tolerance ϵ , if one continues to increase β , the required amount of atoms would be less than the maximum amount and decreased rapidly, this would shape a steep-descent stage.

3) Steady-Descent Stage: Until most of the noise is removed effectively, due to the high representation efficiency of our 3D dictionary, only 1 ~ 2 atoms are required for sparse coding according to our experience. From now, if one continues to increase β , the structures begin to be blurred, but the required amount of atoms would keep steady because the minimum amount of atoms is 1 even for the oversmoothed

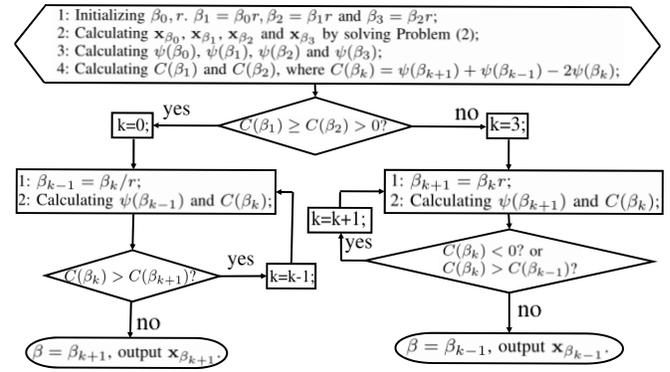


Fig. 2. The workflow for the ZIP method.

structures. This would shape another level-up or steady-descent stage.

As described above, we could consider the desired β as the value associated with the second “corner” of the Z-curve which is the transition point between the second and the third stages, where the noise has been effectively-removed while the structures are still well-retained.

Similar as the L-curve method, this “corner” could be located as the point having the maximum curvature [26], or the maximum second derivative. Considering the shape of our Z-curve (as will be demonstrated in Fig. 4, 11 and 14), from the first stage (steady-descent) to the second stage (steep-descent), the Z-curve is concave with negative second derivative, while from the second stage (steep-descent) to the third stage (steady-descent), the Z-curve is convex with positive second derivative. Regarding this property, we further develop an adaptive strategy to locate the second “corner” of the Z-curve, and hence adaptively select the regularization parameter β . This strategy is referred to as Z-index parameterization (ZIP) method in this work. Workflow could be found in Fig. 2.

Basically, both initial parameter β_0 and ratio r only affect the converge speed of the program. And we only need r to be slightly larger than 1 for a fine tuning of the parameter. Starting from four β , we iteratively estimate the associated curvatures based on the resultant $\psi(\beta) = \frac{\sum_s \|\alpha_s\|_0}{\mathbf{B}}$. In this work, the curvature $C(\beta_k)$ is approximated with the second order difference, as in Line 4 of the initialization step in Fig. 2, where k is the iterator. Basically, if the condition $C(\beta_1) \geq C(\beta_2) > 0$ is satisfied, it indicates that the initial parameter is too large, and hence, one needs to decrease the parameter gradually. If the new curvature $C(\beta_k)$ is larger than the old curvature $C(\beta_{k+1})$, it suggests that one may decrease the parameter further for a larger curvature; otherwise, β_{k+1} could be regarded as the desired parameter having the maximum curvature, and output the corresponding reconstruction $\mathbf{x}_{\beta_{k+1}}$. In contrast, if the condition $C(\beta_1) \geq C(\beta_2) > 0$ cannot be satisfied, it indicates that the initial parameter is too small, and hence, one needs to increase the parameter gradually. The same logic as the described above for the other branch can be used for updating the parameter in this branch. When $C(\beta_k) < 0$, the current β_k belongs to the concave part of the Z-curve, we should increase β further.

In this work, we empirically set $\beta_0 = 0.1$ and $r = 1.2$, which works well among all the involved cases in this work.

E. Experiments

1) *Experimental Data*: Two realistic datasets were collected, including one head-neck (HN) full-fan scan and one prostate half-fan scan. Both datasets were collected from an on board imager integrated in a TrueBeam medical accelerator (Varian Medical System, Palo Alto, CA), where the source-to-axis and source-to-detector distances are 1000mm and 1500mm, respectively. The collected projection data were deidentified in the Radon space. They were rebinned using a 2×2 mode, resulting an imager of 512×384 with a detector size of $0.776\text{mm} \times 0.776\text{mm}$. To be specific, the HN patient was scanned in a full-fan mode to acquire 363 projections in a 200 degrees arc, the tube voltage was 100kVp, and the exposure levels were 0.4 mAs per projection, respectively. The prostate patient (referred to as prostate patient 1 hereafter) was scanned in a half-fan mode with a 160mm lateral shift, acquiring 656 projections in 360 degrees with an exposure level of 1.25 mAs per projection, and the tube voltage was 125kVp. To demonstrate the performance of the proposed 3DDL method in the low dose configurations, Poisson noise for 1×10^4 and 3×10^4 photon incidents per ray was superimposed into the above HN and prostate patient 1 raw datasets (before log transform), respectively. To further consider the electronic noise which cannot be ignored in the low dose configurations, extra zero-mean Gaussian noise was also added to the above generated noisy projections. For both datasets, the standard deviation of the Gaussian noise was set to be 10. Then, the above noisy datasets were truncated by 1×10^4 (for the HN patient case) or 3×10^4 (for the prostate patient 1 case) from above, and by 1 from below, to ensure the non-negative projection value and avoid the numeric overflow from the log transform. Finally, the log transform was performed to simulate projection data. The reconstructed images were of $512 \times 512 \times 512$ and $512 \times 512 \times 256$ with voxel sizes of $0.6\text{mm} \times 0.6\text{mm} \times 0.6\text{mm}$ and $1.0\text{mm} \times 1.0\text{mm} \times 1.0\text{mm}$, corresponding to the HN and prostate patient 1 cases, respectively.

Moreover, to explore the dose reduction potential of the proposed 3DDL method, two groups of realistic datasets were collected based on two different patients receiving image guided radiation therapy (IGRT). One of the them is a female patient suffering from gynecological disease (referred to as GYN patient hereafter), and the other is a male patient suffering from prostate cancer (referred to as prostate patient 2 hereafter). Each group of datasets contains 6 different datasets. Specifically, in each group, these 6 realistic datasets were collected in three consecutive fractions (denote as FX 1, FX 2 and FX 3 hereafter) which were conducted every week. And in each fraction, during image guiding, a low dose protocol scan was followed by a regular dose protocol scan. For all the three different regular dose protocol scans, the exposure level settings were $80\text{mA} \times 13\text{ms}$, while for the three low dose protocol scans, the exposure level settings were $40\text{mA} \times 13\text{ms}$, $20\text{mA} \times 13\text{ms}$ and $10\text{mA} \times 13\text{ms}$, corresponding to FX 1, FX 2

and FX 3, respectively. For both cases, 618 projections were acquired in 360 degrees with the same geometry parameters as the above prostate patient 1 case, except that the lateral half-fan shift was 148mm. And also, the reconstructed images were of $512 \times 512 \times 256$ with voxel sizes of $1.0\text{mm} \times 1.0\text{mm} \times 1.0\text{mm}$ for both cases. Therefore, based on these datasets, we could evaluate the performance of the proposed 3DDL method under $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose level situations.

2) *Analysis for the Representation Efficiencies of the 2D/3D Dictionaries*: As described in Section II-C, it is important to explore the representation efficiencies of the 2D/3D dictionaries for the spatial structures. The result can be used to justify the 3DDL method in a natural scene statistical viewpoint. Specifically, it is well-accepted that the representation coefficients of the natural scene images shall follow a zero-mean Laplacian distribution that is highly peaked around zero with heavy tails [37]. Therefore, we conduct the representation efficiency analysis by investigating the resultant Laplacian distributions. It is expected that the narrower the Laplacian distribution is, the more efficient the trained dictionary can represent the structures, and hence facilitate the later processing stages better, such as the denoising task in our context.

In this study, the regular dose FDK reconstructions [38] of the HN and prostate patient 1 are employed to explore the representation efficiencies. In details, for the HN patient case, a group of 1×10^5 3D data samples of $4 \times 4 \times 4$ are randomly extracted from the volumetric images. Then, three groups of 1×10^5 2D data samples of $8 \times 8 \times 1$, $8 \times 1 \times 8$ and $1 \times 8 \times 8$ are randomly extracted from the transversal, coronal and sagittal views, respectively. The same rules are also applied on the prostate patient 1 case. The above data samples are fed into the following equation to obtain the corresponding representation coefficients

$$\min_{\alpha} \|\mathbf{X} - \mathbf{D}\alpha\|_2^2 + \gamma \|\alpha\|_1, \quad (8)$$

where each column of \mathbf{X} is a vector stacking representation of the 2D/3D data sample, \mathbf{D} is the 2D/3D dictionary, α denotes the resultant representation coefficients, and γ is a Lagrangian multiplier. Equation (8) is a typical least absolute shrinkage and selection operator (LASSO) problem [39], which could be solved equivalently with the following constraint problem:

$$\min_{\alpha_{:,i}} \|\alpha_{:,i}\|_1, \quad \text{s.t.}, \|\mathbf{X}_{:,i} - \mathbf{D}\alpha_{:,i}\|_2^2 \leq \tau, \quad \forall i \quad (9)$$

where $\mathbf{X}_{:,i}$ denotes the i^{th} sample (column) of \mathbf{X} whose sparse coefficients are $\alpha_{:,i}$, i.e., the i^{th} column of α . In the practical implementation, we employ the SPAMS software package as the solver, and set the tolerance $\tau = 0$ for all the 2D/3D cases to fully explore the sparseness of the resultant coefficients. We will analyze the distributions of the resultant coefficients and calculate the associated variance of each distribution to quantify the representation efficiencies.

3) *Regularization Parameter Selection Strategy*: All the data cases are used to illustrate the introduced Z-curve in Section II-D. Specifically, 21 different parameters β are fed into problem (2). For the prostate patient 1 case, β ranges from 2.5×1.2^{-10} to 2.5×1.2^{10} . For all the other cases, β ranges from 0.5×1.2^{-10} to 0.5×1.2^{10} . Ratio between different β

is 1.2. During reconstruction, the OS setting for the HN patient case is 11 subsets with 33 projections per subset, denoted as 11(\times 33) OS protocol hereafter. They are 8(\times 82) and 6(\times 103) OS protocols for the prostate patient 1 and the GYN/prostate patient 2 cases, respectively. The whole optimization program is terminated after 10 iterations for all the cases. Consequently, 21 different images ranging from over-noisy to over-smooth are reconstructed. Based on these reconstructions, our Z-curve $\{\beta, \psi(\beta)\}$ as well as the associated curvatures are plotted.

Moreover, to compare with our 3DDL based sparse regularizer, we also perform the TV minimization based CBCT reconstruction for the HN and prostate patient 1 cases with 21 different β . They range from 6.25×1.2^{-10} to 6.25×1.2^{10} and from 6.0×1.2^{-10} to 6.0×1.2^{10} , corresponding to the HN and prostate patient 1 cases, respectively, ratio between different β is 1.2. Based on these reconstructions, curve $\{\beta, \mathbf{R}(\mathbf{x}_\beta)\}$ is plotted, where $\mathbf{R}(\mathbf{x}_\beta)$ is calculated to be the value of the total variation of the image.

4) Comparison Studies Among Different Regularizers: In this study, based on the HN and prostate patient 1 datasets, we qualitatively and quantitatively compare the proposed 3DDL method with two existing methods, namely, 2D dictionary learning based method [7] and TF method [16]. Specifically, for the 2D dictionary learning based method, three 2D dictionary learning based sparse constraints are consecutively applied on the transversal, coronal and sagittal views for a fair comparison, though this strategy suffers from heavy computation burden. Furthermore, to illustrate the inherent 2D property of the 2D dictionary learning based method for volumetric image reconstruction, three additional strategies are employed to utilize the 2D dictionary learning based regularizer, i.e., in a slice-wise fashion for each of the three different views. The employed 2D dictionary contains 256 atoms of 8×8 . For fair comparison, and also taking the isotropic voxel resolution of volumetric image into account, in the 3DDL method, the employed 3D dictionary also contains 256 3D atoms of $4 \times 4 \times 4$. Besides, both the 2D and 3D dictionary training processes employ the same high-quality CBCT images as the sample source. For brevity, we will use the following abbreviations to represent the different methods:

3DDL: the proposed 3DDL method.

2DDL: three 2D dictionary learning based sparse constraints are consecutively enforced on all the three views.

TF: the TF method.

881DL: the 2D dictionary learning based sparse constraint is only enforced on the transversal view in a slice-wise method.

818DL: the 2D dictionary learning based sparse constraint is only enforced on the coronal view in a slice-wise method.

188DL: the 2D dictionary learning based sparse constraint is only enforced on the sagittal view in a slice-wise method.

For all the iterative methods, the OS settings and the iteration number are the same as Section II-E.3. For all the dictionary learning based methods, the sparsity level (maximum amount of atoms allowed while sparse coding) is selected to 8, the tolerance parameter ϵ is selected to be 1×10^{-3} , the stride is 1 during image patch/block extraction. It is noted that a same pretrained 2D dictionary is used among all the 2DDL/881DL/818DL/188DL methods for all the experiments

in this paper. And also, a same pretrained 3D dictionary is used for the 3DDL method in all the experiments in this paper. Both datasets are also reconstructed by the FDK algorithm to benchmark the regularized iterative reconstructions.

In this study, for the proposed 3DDL method, β is selected according to our ZIP method as presented in Fig. 2. For the other methods, β is manually selected for the best visual quality.

For both cases, because the regular dose reconstructions are available, the results are quantified with the root mean squared errors (RMSE) and the structure similarity (SSIM) index (the closer to 1, the better the image is) [40].

The RMSE is calculated based on the whole 3D volume as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{x,y,z} (\mathbf{x}(x,y,z) - \hat{\mathbf{x}}(x,y,z))^2}{N_x N_y N_z}}, \quad (10)$$

where N_t denotes that number of voxels along the t dimension, $\mathbf{x}(x,y,z)$ and $\hat{\mathbf{x}}(x,y,z)$ represent the voxel values at spatial location (x,y,z) of the low dose reconstruction and the regular dose FDK reconstruction, respectively.

Since the SSIM index is devised for the 2D images, in this work, the SSIM index is calculated based on a single presented transversal slice, as follows:

$$\text{SSIM} = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_{x,\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}, \quad (11)$$

where \mathbf{x} and $\hat{\mathbf{x}}$ represent the transversal slices for the low dose reconstruction and the regular dose FDK reconstruction, whose mean values/standard deviations are μ_x/σ_x and $\mu_{\hat{x}}/\sigma_{\hat{x}}$, respectively. $\sigma_{x,\hat{x}}$ denotes the covariance between \mathbf{x} and $\hat{\mathbf{x}}$. $c_1 = 3 \times 10^{-5}$ and $c_2 = 3 \times 10^{-4}$ are two small positive constants to stabilize the division. The transversal images used for SSIM calculation could be found in the full image of Fig. 5 and 7.

To further demonstrate the noise textures of different methods, we also calculate the 3D noise power spectrums (NPS) of the reconstructions with respect to the above 6 regularized iterative reconstruction methods. To be specific, same as the above low dose protocol simulations both for the HN and prostate patient 1 cases, 20 independent low dose projection data simulations for each case are firstly conducted. Then, the above 6 regularized iterative reconstruction methods are used to perform reconstruction for each of the simulated noisy data. In each reconstruction, the parameter β is absolutely the same as the above selected β used in the comparison studies. A cubic volume of interest (VOI) around the center of the field of view is used to measure the 3D NPS. The physical dimensions of the VOI are 60mm \times 60mm \times 60mm (HN patient case) and 100mm \times 100mm \times 100mm (prostate patient 1 case). The 3D NPS is calculated using

$$\begin{aligned} \text{NPS}(f_x, f_y, f_z) &= \frac{\Delta x \Delta y \Delta z}{N_x N_y N_z} \frac{\sum_{i=1}^{20} |\text{DFT}_{3D}[\mathbf{x}^i(x,y,z) - \bar{\mathbf{x}}(x,y,z)]|^2}{20}, \quad (12) \end{aligned}$$

where Δ_t and N_t denote the voxel size and the number of voxels along the t dimension. DFT_{3D} represents the 3D

discrete Fourier transform. And $\bar{\mathbf{x}}(x, y, z)$ is calculated as the averaged image of all the 20 individual reconstructions $\mathbf{x}^i(x, y, z)$, defined as

$$\bar{\mathbf{x}}(x, y, z) = \frac{1}{20} \sum_{i=1}^{20} \mathbf{x}^i(x, y, z). \quad (13)$$

In both data cases, the NPS will be calculated and displayed in the frequency ranges within \pm Nyquist frequency, which are 0.83mm^{-1} (HN patient case) and 0.5mm^{-1} (prostate patient 1 case).

5) Low Dose Potential Exploration: The GYN patient and the prostate patient 2 cases are employed to explore the low dose potential of the proposed 3DDL method. In details, for both cases, each of the 6 datasets collected in different fractions and different dose levels is firstly FDK reconstructed. And then for the three low dose projection datasets, the proposed 3DDL method is also employed for the volumetric image reconstruction, where $6(\times 10^3)$ OS protocol and 10 iterations are used. The sparsity level and the tolerance while sparse coding are still set to be 8 and 1×10^{-3} in both cases. It is noted that the reconstructed images are not absolutely the same despite that they are from the same patient, because the associated projections are collected in different time, even when they belong to a same fraction. Specifically, based on a selected region-of-interest (ROI), we measure the contrast-to-noise ratio (CNR) for different reconstructions. The CNR is calculated as $\text{CNR} = \frac{2|S - S_b|}{\sigma + \sigma_b}$, where S and S_b represent the mean intensities of the ROI and the background, σ and σ_b are the associated standard deviations, respectively. Besides, we also evaluate the noise performance based on a selected flat area by calculating the standard deviation (STD). In addition, in order to assess the low contrast tissue discrimination ability of different reconstructions, three radiologists are asked to vote the reconstructions in three different levels: strongly discriminated, weakly discriminated and unclear. The level with the most votes will be regarded as the level of the reconstruction. For example, if one reconstruction has two votes for strongly discriminated and one vote for weakly discriminated, then the low contrast tissue of the reconstruction could be regarded as strongly discriminated. Specifically, if one reconstruction has one vote for each level, then it will be regarded as unclear. In this subject experiments, we employ the presented coronal views for the radiologists' evaluation. In this study, β is also selected based on our ZIP method as presented in Fig. 2.

6) Computational Overhead Comparison: Without loss of generality, we will take the HN and prostate patient 1 cases as examples to conduct the computational overhead comparison. It is believed that similar observations can be also achieved from the other cases. Specifically, for all the involved methods in this work, we separately calculate the computational consumptions from the data fidelity term update and the regularization term update. The computational overhead for these two parts are calculated as the averaged time consumptions among all the 10 iterations. All the algorithms are implemented in the CUDA 7.0 programming environment on a NVIDIA GeForce GTX 980 video card which is installed on a personal computer (Intel i5-4460 CPU and 8GB RAM).

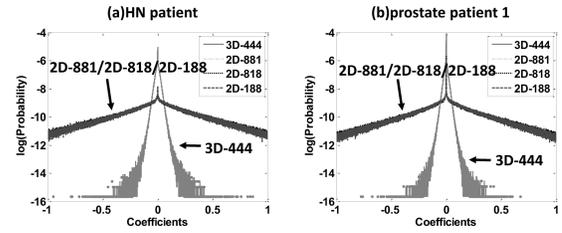


Fig. 3. Distributions of the representation coefficients among different dictionaries for the HN patient (a) and prostate patient 1 (b). The x-axis is the values of the coefficients, the y-axis is the logarithmic probabilities. 3D-444 denotes the distributions of the coefficients for the 3D data samples represented by the 3D dictionary of dimension $4 \times 4 \times 4$. 2D-881/2D-818/2D-188 denote the distributions of the coefficients with the 2D dictionary for the 2D data samples extracted from the transversal/coronal/sagittal views, respectively.

III. RESULTS

A. Representation Efficiency Results

Fig. 3 plots the distributions of the representation coefficients among different dictionaries. Note that the y-axis is set to be the logarithmic probabilities to show the Laplacian nature of these probability distributions more clearly [37]. It is obvious that for both cases, the 3D dictionary based sparse coefficients consistently have three orders narrower Laplacian distributions compared to the 2D dictionary based sparse coefficients, suggesting the higher representation efficiencies of the 3D dictionary. The variances of these distributions are summarized in Table S2² in the supplemental material. It is shown that the variances associated with the 3D dictionary are much smaller compared to those with respect to the 2D dictionaries. Additional two experiments involving a dental patient case and a small animal case are conducted for a further validation, see Fig. S3³ in the supplemental material.

B. Regularization Parameter Selection Results

Fig. 4 presents the Z-curve results of the HN and prostate patient 1 cases. From Fig. 4 (a) and (b), it could be found that the trends of both our curves $\{\beta, \psi(\beta)\}$ could be divided into three stages, i.e., firstly steady-descent stage, then steep-descent stage, finally steady-descent stage, demonstrating a clear Z-shape, matching well with our analysis in Section II-D. Regarding the curvature curve $\{\beta, C(\beta)\}$, it shows that the point with maximum curvature corresponds to the second corner of the Z-curve, as indicated by the arrows. Moreover, the curvature of the left-most part is negative, indicating the concave part of the Z-curve. As for the TV based regularizer, the “potential corner” is less apparent compared to our 3DDL based sparse regularizer, especially in the prostate patient 1 case, as illustrated in Fig. 4 (c) and (d). It is noted that to fully illustrate the Z-shape of the curves, the starting parameters in Fig. 4 (a) and (b) are different from the initial parameter $\beta_0 = 0.1$ used in our ZIP method.

²Please see the supplemental material for the results of the variances of the distributions.

³Please see the supplemental material for the distributions of the other two cases.

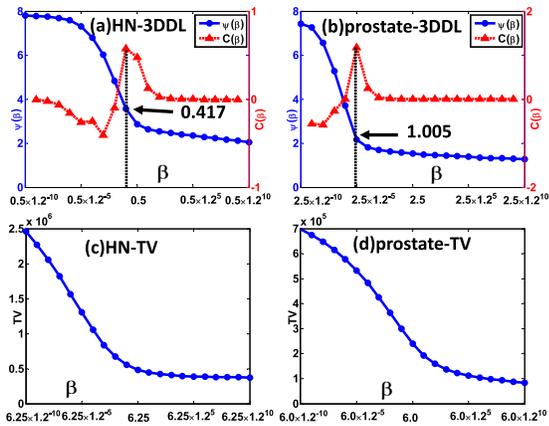


Fig. 4. Z-curve results for the HN and prostate patient 1 cases. The first and the second rows correspond to the 3DDL and TV based sparse regularizers, respectively. The first and the second columns correspond to the HN and prostate patient 1 cases, respectively. The arrows indicate the corresponding β having maximum curvature.

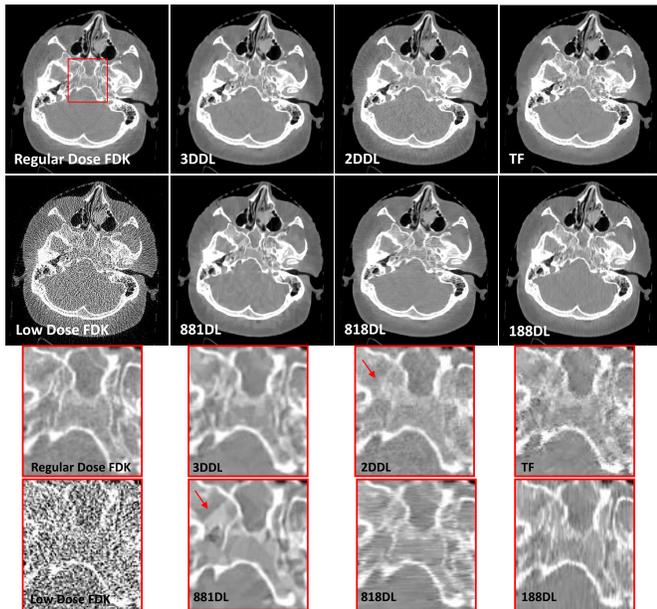


Fig. 5. Transversal views of the HN patient images reconstructed by different methods. From left to right in the first row, the images are regular dose FDK reconstruction, reconstructions from the 3DDL, 2DDL and TF methods, respectively. From left to right in the second row, the images are reconstructed by the FDK, 881DL, 818DL and 188DL methods, respectively. The last two rows show the corresponding zoomed-in ROIs of the box in the first two rows. The display window is $[-750\ 750]$ HU.

C. Performance Comparison Among Different Regularizers

Fig. 5 are the transversal views of the reconstructed images for the HN patient case. It is observed that the low dose FDK reconstruction is overwhelmed by the noise, while the noise is substantially suppressed by the regularized iterative reconstruction algorithms. As indicated by the zoomed-in ROIs in Fig. 5, if the 2D dictionary learning based sparse constraint is enforced on only one view, the unprocessed views exhibit directional streak artifacts, such as the horizontal and vertical streak artifacts corresponding to the 818DL and

188DL sub-figures, respectively. Similar directional streak artifacts could be also observed from the coronal and sagittal views, see Fig. S4 and S5⁴ in the supplemental material, respectively. Regarding the processed views, the structures are distorted, and there also induce some fake structures that should not exist. For example, as indicated by the arrows in the 881DL zoomed-in ROI of Fig. 5, part of the soft bone is missing and replaced with a flat region compared to the regular dose FDK reconstruction. The fake structures could be further observed from the soft tissue part of the other two views, such as the 818DL sub-figure in Fig. 4S and the 188DL sub-figure in Fig. S5 in the supplemental material.⁵ The directional streak artifacts can be alleviated if the 2D dictionary learning based sparse constraint is enforced on all the three views consecutively, i.e., the 2DDL method. However, compared to the 3DDL method, the 2DDL method exhibits lower spatial resolution and higher image noise, as indicated by 2DDL images in Fig. 5. Another disadvantage of the 2DDL method is the inherent high computational cost, considering that three individual sparse coding stages are required for regularization. Regarding the TF method, one cannot well distinguish the subtle structures which are blurred due to the reduced resolution and the remained noise in the high contrast region, as indicated by the TF sub-figure in Fig. 5. On the other hand, from these results, it can be seen that the 3DDL method achieves promising results in enhancing the anatomical structures and in removing the noise effectively, and hence validates its efficacy. Quantitatively, with the regular dose FDK reconstruction as the reference, the RMSE and SSIM values are listed in Table S3⁶ in the supplemental material. The lowest RMSE and the highest SSIM further verify that the 3DDL method outperforms other algorithms.

Fig. 6 presents the NPS of the HN patient images reconstructed with different methods. It is noted that the NPS slightly loses its radial symmetry due to the different x-ray paths length. Considering this radial symmetry losses, to better visualize the NPS, the representative cuts of the $f_x - f_z$ and $f_y - f_z$ plans are selected to be the $\frac{1}{10}$ Nyquist frequency, i.e., 0.08mm^{-1} . From the first column, one could find that the noise powers of the 2DDL/TF reconstructions are stronger than the 3DDL reconstructions by comparing their NPS. On the other hand, the NPS of the 881DL/818DL/188DL reconstructions clearly suggest the above directional streak artifacts. For example, if the sparse constraint is only enforced on the transversal view, i.e., the 881DL method, there will induce many directional streak artifacts in the other two views. These will result in large frequency magnitudes along the Z direction of the NPS, as shown in the 881DL sub-figures in Fig. 6. Similar phenomena could also be observed from the 818DL/188DL NPS.

Fig. 7 illustrates the transversal views of the prostate patient 1 case with different methods. It is shown that the

⁴Please see the supplemental material for the results of the coronal and sagittal views of the HN patient case.

⁵Please see the supplemental material for the results of the coronal and sagittal views of the HN patient case.

⁶Please see the supplemental material for the RMSE and SSIM results of the HN patient case.

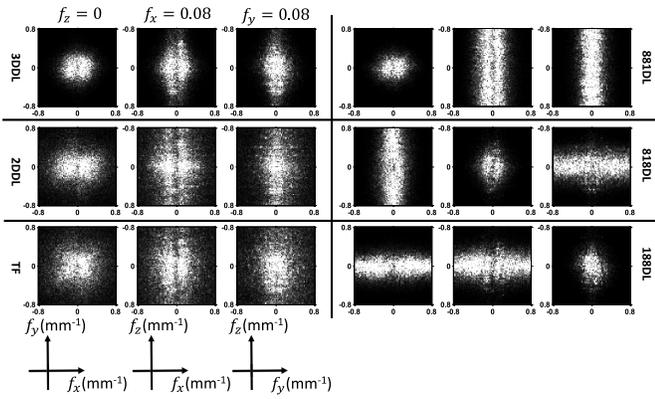


Fig. 6. The NPS of the HN patient images reconstructed by different methods. From top to bottom in the first column, the images are the NPS with respect to the 3DDL/2DDL/TF reconstructions. From top to bottom in the second column, the images are the NPS with respect to the 881DL/818DL/188DL reconstructions. The display range is the same for all the cases ($[0\ 3000]$ HU²mm³).

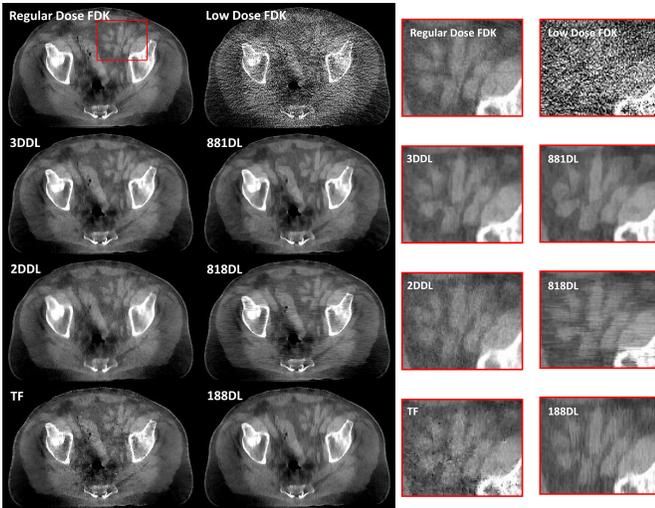


Fig. 7. Transversal views of the prostate patient 1 images reconstructed by different methods. From top to bottom in the first column, the images are regular dose FDK reconstruction, reconstructions from the 3DDL, 2DDL and TF methods, respectively. From top to bottom in the second column, the images are reconstructed by the FDK, 881DL, 818DL and 188DL methods, respectively. The last two columns show the corresponding zoomed-in ROIs of the box in the first two rows. The display window is $[-300\ 150]$ HU.

3DDL method efficiently suppresses noise and well retains anatomical structures both for the low contrast and high contrast regions. Regarding the TF method, the structures are contaminated by the remained pepper-like noise, as shown by the TF sub-figure in Fig. 7. It can be seen that the 2DDL method exhibits stronger noise with comparable resolution, if not inferior, compared to the 3DDL method. On the other hand, directional streak artifacts are observed from the reconstructions, as demonstrated by the 818DL/188DL sub-figures in Fig. 7, while the structures are distorted for the processed view. These distorted structures could be further observed from the other two views, as indicated by the 818DL sub-figure in Fig. S6 and the 188DL sub-figure in Fig. S7 in the supplemental material.⁷ The calculated RMSE/SSIM values

⁷Please see the supplemental material for the results of the coronal and sagittal views of the prostate patient 1 case.

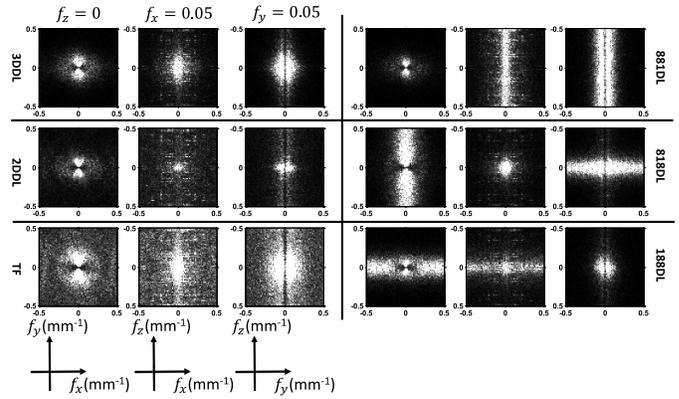


Fig. 8. The NPS of the prostate patient 1 images reconstructed by different methods. The sub-figures are in the same arrangement as those in Fig. 6. The display range is the same for all the cases ($[0\ 3000]$ HU²mm³).

are in Table S4⁸ in the supplemental material. As expected, the 3DDL method quantitatively outperforms the other competitors in terms of the lowest RMSE and the highest SSIM measures, which are consistent with the visual observations that the 3DDL method leads to more naturally and visually pleasant denoising results by better preserving the image texture areas.

Fig. 8 presents the NPS of the prostate patient 1 images reconstructed with different methods. In this case, due to the longer paths along the lateral direction, the NPS shows severe radial symmetry loss. Besides, the NPS with respect to the 2DDL method appears to be "redder" compared to the 3DDL method, considering the stronger noise appearance in the spatial domain of the 2DDL reconstruction, as indicated by sub-figures in Fig. 7. The NPS associated with the TF method indicates the noisier reconstruction compared to the 3DDL reconstruction. Once again, the NPS associated with the 881DL/818DL/188DL methods suggest the directional streak artifacts for the unprocessed views.

It is noted that for both cases, the regularization parameter β in the 3DDL method is adaptively selected according to our ZIP method in Fig. 2. It turns out to be 0.430 and 1.070, matching well with the suggested β based on the maximum curvature of our Z-curve, which are 0.417 and 1.005 (see Fig. 4 (a) and (b)), corresponding to the HN and prostate patient 1 cases, respectively. Regarding the nice image qualities delivered by our 3DDL method, the efficacy of our ZIP method is validated.

D. Low-Dose Potential Exploration

Fig. 9 presents the transversal view of the reconstructed images associated with the GYN patient case in three different fractions. From the second column, it is observed that as the dose decreases from $\frac{1}{2}$ dose level (row 1 in the second column) to $\frac{1}{8}$ dose level (row 3 in the second column), the FDK reconstructions become noisier and noisier. This phenomena could be observed more clearly from the coronal views in Fig. 10. The associated zoomed-in ROI and

⁸Please see the supplemental material for the RMSE and SSIM results of the prostate patient 1 case.

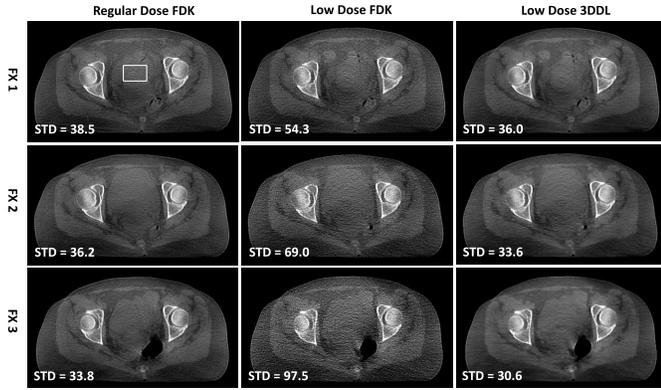


Fig. 9. Transversal view of the reconstructed images of the GYN patient case in different fractions and different dose protocols. From top to bottom, the images correspond to FX 1, FX 2 and FX 3, respectively. From left to right, the images are regular dose FDK reconstructions, low dose FDK reconstructions and low dose 3DDL reconstructions, respectively. In the second column, from top to bottom, the dose levels are $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ compared to the regular dose levels in the first column. The region in the box is selected for the STD calculation, whose values are displayed in the lower-left corner of each reconstruction, the unit is HU. The display window is $[-400\ 350]$ HU.

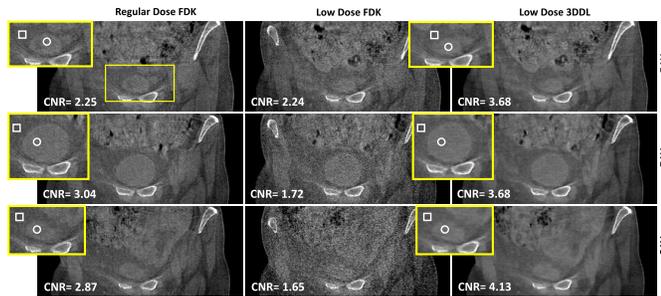


Fig. 10. Coronal view of the reconstructed images of the GYN patient case in different fractions and different dose protocols. The full image are in the same arrangement as those in Fig. 9. The insets are the associated zoomed-in ROIs of the box, indicating the region used for CNR calculation, where the contents inside the circle and the box of the insets are regarded as the foreground and the background, respectively. The calculated CNR value is shown in the lower-left corner. The third column shares the same ROIs as the second column for the CNR calculation. The display windows is $[-400\ 350]$ HU.

the sagittal view are presented in Fig. S10 and S8 in the supplemental material,⁹ respectively. In contrast, for all the dose levels, the proposed 3DDL method always could produce comparable image qualities, and one could consistently tell the ROIs apart from the background due to the efficiently suppressed noise and fine reserved structures, as indicated by the zoomed-in ROIs. Quantitative evaluations in terms of STD and CNR also validate the above observations, as indicated by the corresponding STD and CNR values in Fig. 9 and 10, respectively. Specifically, regarding the noise performance, the STDs of the FDK reconstructions increase from ~ 35 HU to ~ 100 HU as the dose levels decrease from the regular dose to the $\frac{1}{8}$ dose level. With the proposed method, the noise is suppressed effectively in terms of that the STD with respect

⁹Please see the supplemental material for the results of the zoomed-in ROI and sagittal views of the GYN patient case.

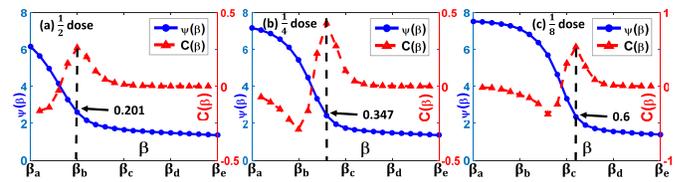


Fig. 11. Z-curve results for the GYN patient cases. (a) \sim (c) correspond to the $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose level cases, respectively. The values of $\beta_a \sim \beta_e$ are 0.5×1.2^{-10} , 0.5×1.2^{-5} , 0.5 , 0.5×1.2^5 and 0.5×1.2^{10} , respectively. The arrows indicate the suggested β of the maximum curvature points.

to the 3DDL method is comparable to that of the regular dose FDK reconstruction. As for the CNR metric, the proposed method could improve the CNR values by $\sim 1.6/2.1/2.5$ times associated with $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose levels, comparing to the low dose FDK reconstructions, even higher than the regular dose cases. Moreover, in our method, one could also find that the CNRs are comparable among all the three low dose protocols. These quantitative results further validate the efficacy of the proposed 3DDL method. It is noted that during the CNR calculation, the ROI with slightly higher intensities is chosen as the the foreground (marked as circle), and the region with lower intensities near the ROI is randomly chosen as the background (marked as box) despite manually. In addition, because the images are slightly different across different fractions and different dose protocols, slightly different ROIs used for the STD and CNR calculations are employed for the regular dose protocols and the low dose protocols, while in the same fraction, the low dose FDK reconstruction and the 3DDL reconstruction share the same ROI for their calculations. As for the clinical subjective evaluations of the low contrast tissue discrimination ability based on the presented coronal view (Fig. 10), from Table I, it is found that the reconstructions associated with our method are voted as strongly discriminated among different dose levels, while the $\frac{1}{4}/\frac{1}{8}$ dose level FDK reconstructions are voted as unclear, which further validates our method.

Fig. 11 presents the Z-curve results of the three GYN patient cases in three different low dose levels. Again, in Fig. 11 (b) and (c), the $\{\beta, \psi(\beta)\}$ curves exhibit obvious Z-shapes. If one uses more smaller β , a clear Z-shape for the $\frac{1}{2}$ dose level case in Fig. 11 (a) is also expected. As indicated by the arrows, the β with respect to the maximum curvature points are 0.201/0.347/0.6, corresponding to the $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose level cases, respectively.

On the other hand, the low dose 3DDL images in Fig. 9 and 10 are reconstructed based on the adaptively selected β with our ZIP method, which turns out to be 0.207/0.358/0.619, corresponding to $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose level cases, respectively. These matches well with the indicated β based on the maximum curvature points in Fig. 11.

Fig. 12 and 13 demonstrate the transversal and coronal views of the reconstructed images associated with the prostate patient 2 cases in three different fractions. Similarly, it is found that as the dose level decreases from regular dose to $\frac{1}{8}$ dose level, the FDK reconstructions become noisier and noisier. Quantitatively, the STDs increase from ~ 40 HU (regular dose) to ~ 120 HU ($\frac{1}{8}$ dose). With the proposed method, the noise

TABLE I

VOTES OF THE LOW CONTRAST TISSUE DISCRIMINATION ABILITY OF DIFFERENT METHODS FOR THE GYN AND PROSTATE PATIENT 2 CASES. SD, WD AND U DENOTES STRONGLY DISCRIMINATED, WEAKLY DISCRIMINATED AND UNCLEAR, RESPECTIVELY. LD AND RD REPRESENTS THE LOW DOSE AND REGULAR DOSE CASES, RESPECTIVELY

Dose levels		$\frac{1}{2}$ dose			$\frac{1}{4}$ dose			$\frac{1}{8}$ dose		
Methods		LD-3DDL	LD-FDK	RD-FDK	LD-3DDL	LD-FDK	RD-FDK	LD-3DDL	LD-FDK	RD-FDK
GYN	SD	2			3		1	2		
	WD	1	2	3		1	1	1		3
	U		1			2	1		3	
Prostate 2	SD	3						3		
	WD		1	2	2		3			2
	U		2	1	1	3			3	1

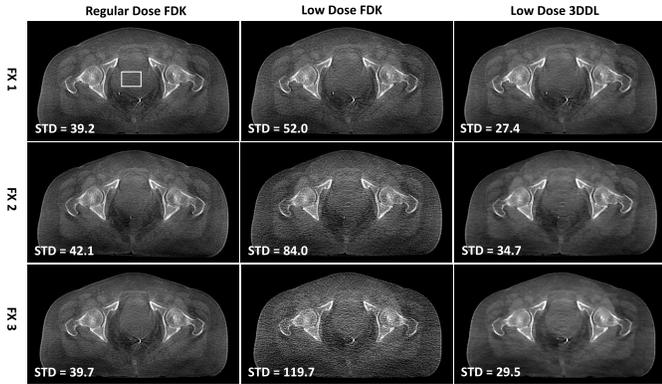


Fig. 12. Transversal view of the reconstructed images of the prostate patient 2 case in different fractions and different dose protocols. The sub-figures are in the same arrangement as those in Fig. 9. The display window is $[-400\ 350]$ HU.

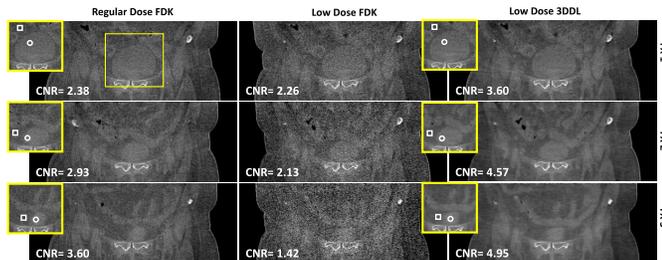


Fig. 13. Coronal view of the reconstructed images of the prostate patient 2 case in different fractions and different dose protocols. The sub-figures are in the same arrangement as those in Fig. 10. The display window is $[-400\ 350]$ HU.

are removed effectively and even weaker than the regular dose FDK reconstructions regarding the smaller STDs. It should be noted that despite of the reduced noise, the resolution of the 3DDL reconstructions is still kept well and comparable with the low dose FDK reconstruction. This could be observed from Fig. 12, considering that both the low dose FDK reconstruction and the 3DDL reconstruction give clear metal markers with comparable visual resolutions which is relevant for the patient positioning during the clinical IGRT. From Fig. 13, one could see that with the proposed method, the CNRs are improved by $\sim 1.5/2.5/3.5$ times compared to the low dose FDK reconstructions with respect to the $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose levels. Due to the effectively-suppressed noise with the proposed method, the soft tissues could be clearly distinguished for all the low dose cases, as indicated by the zoomed-in ROIs in Fig. S11 in the supplemental material.¹⁰ The sagittal views

¹⁰Please see the supplemental material for the results of the zoomed-in ROI of the prostate patient 2 case.

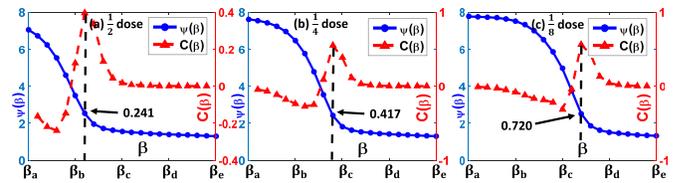


Fig. 14. Z-curve results for the prostate patient 2 case. The sub-figures are in the same arrangements as those in Fig. 11.

TABLE II

COMPUTATIONAL OVERHEADS (UNIT: SECONDS) OF DIFFERENT METHODS FOR THE HN AND PROSTATE PATIENT 1 CASES. F AND R INSIDE THE PARENTHESES DENOTE THE FIDELITY AND REGULARIZATION TERMS, RESPECTIVELY

Cases	3DDL	2DDL	TF	881DL	818DL	188DL
HN (F)	13.04	13.99	13.02	13.91	13.02	13.17
HN (R)	20.49	61.62	5.46	21.29	20.86	20.81
Prostate (F)	9.33	9.76	9.28	9.53	9.52	9.95
Prostate (R)	14.93	50.16	2.25	20.6	16.84	17.94

are also presented in Fig. S9 in the supplemental material.¹¹ Regarding the clinical evaluation results (Table I) based on the coronal view in Fig. 13, one could find that the reconstructions associated with our method are voted as strongly discriminated and weakly discriminated in the $\frac{1}{2}/\frac{1}{8}$ and $\frac{1}{4}$ dose levels, respectively, while the low dose FDK reconstructions in all the three different dose levels are voted as unclear. It is noted that both in Fig. 9 and 12, there exist slight ring artifacts in the center of the transversal image, which is supposed to be caused by the inconsistent detector response.

Fig. 14 presents the z-curve results of the three prostate patient 2 cases in three different low dose levels. Clear Z-curves as well as the corners could be observed. The parameters associated with the point having maximum curvature and selected by our ZIP method are 0.241/0.417/0.720 and 0.249/0.430/0.743, corresponding to the $\frac{1}{2}/\frac{1}{4}/\frac{1}{8}$ dose level cases, respectively, matching well with each other.

E. Computational Overhead

Table II lists the computational overheads for all the methods considered in this work. It is observed that the time consumption for the fidelity term update is quite stable among different methods, however, for the regularization term update, the time consumption is highly correlated with the choice of the regularizer. Specifically, similar computational overheads are required for the 3DDL method and

¹¹Please see the supplemental material for the results of the sagittal views of the prostate patient 2 case.

the 881DL/818DL/188DL methods, while the 2DDL method suffers from significantly higher computational overhead. Besides, one of the biggest advantages of the TF method is its low computation complexity.

IV. DISCUSSIONS AND CONCLUSIONS

In this study, a 3D dictionary learning based sparse regularizer has been constructed for low dose CBCT reconstruction, being validated in multiple realistic data experiments. A statistical analysis on the representation efficiencies of the 2D/3D dictionary were carried out, suggesting the higher representation efficiency of the 3D dictionary. This observation could be used to justify the performance of the proposed 3DDL method. Realizing that the constructed dictionary is an excellent feature descriptor, a new adaptive regularization parameter selection strategy, i.e., ZIP method, was proposed. Based on the suggested β with the ZIP method, our 3DDL method can deliver superior image quality in terms of well-preserved structures and effectively-suppressed noise, and outperforms the competing methods, such as 2DDL/881DL/818DL/188DL/TF methods. Moreover, the whole program was well parallelized by employing several algorithmic tricks, attaining a high computational efficiency.

The higher representation efficiency of the 3D dictionary over the 2D dictionary may be explained by the fact that the 3D dictionary could sufficiently capture spatial correlations in all the three dimensions simultaneously, while the 2D dictionary could only make use of the planar spatial correlations. As mentioned in Section II-C, a more efficient representation could facilitate the later denoising stage. Indeed, this has been experimentally validated to a certain extent in our comparison studies as described in Section III-C.

Considering that the dictionary is an excellent feature descriptor, we realize that the associated averaged sparsity level $\psi(\beta)$ while sparse coding may be a good image quality indicator. Motivated by this, in this work, an adaptive regularization parameter selection strategy based on the constructed Z-curve $\{\beta, \psi(\beta)\}$, i.e., ZIP method, was proposed to facilitate our 3DDL method, as described in Section II-D. It is believed that same strategy also applies on the 2D dictionary learning based slice CT reconstruction. However, we did not use this strategy for the 2DDL method in our comparison study. This is because one could not determine a concrete value for the sparsity level $\psi(\beta)$ in the 2DDL method, considering the fact that the sparse constraint was applied on all the three views consecutively. The reason why the TV value curve with respect to β doesn't work may be that the TV value is sensitive to the noise power. In other words, the noisier the image is, the larger the TV value is, as indicated in Fig. 4. By contrast, our 3DDL based averaged sparsity level counts the averaged amount of the used atoms while sparse coding, and hence, is truncated from above and below, which is 8 and 1 in this work. This property could help produce a well Z-curve shape. Despite the promising performance of our ZIP method, one of the main drawbacks is that multiple iterations are required so as to find the desired β . A fast ZIP method is our ongoing research, quick estimation based on a subset of image and then extension to the original data may be a direction. Moreover, the proposed

ZIP method is heuristic, a more comprehensive analysis about the image qualities of the associated reconstruction is required for a better understanding.

The cause of the directional streak artifacts is that the noise in the processed view is smoothed out when the 2D dictionary based processing is applied on only one view. As a result, the intersections of the unprocessed views with the processed view would exhibit directional streak artifacts, such as the horizontal/vertical streak artifacts in the transversal views if the coronal/ sagittal views are processed, as illustrated by the 818DL/188DL sub-figures in Fig. 5. The reason of the distorted structures in the processed view may be explained by the fact that the directional streak artifacts are spread out through the cone beam forward projection. If the sparse constraint are applied on all the three views consecutively, i.e., the 2DDL method, the directional streak artifacts could be alleviated. However, the 2DDL method may incorrectly interpret the directional streak artifacts from the previous steps as the potential structures. To avoid this side effect, a large tolerance maybe required. Moreover, one may need to carefully select the suitable tolerance for each of the three sparse coding steps in the 2DDL method. In this work, the tolerances were set to be the same for all the three steps and also same as the 881DL/818DL/188DL methods. Another disadvantage of the 2DDL method is the high computational cost, as indicated by Table II. This is mainly because that three individual sparse coding stages are required in each iteration for the 2DDL method. The TF method employs a group of piecewise linear TF basis functions consisting of low pass filters for low frequency components, as well band pass and high pass filters for edges. As a consequence, in processing high contrast regions, the high pass filters are required to represent the structures, and may result in the pepper-like noise artifacts.

It should be noted that in this work, a same 3D dictionary is employed for all the cases, i.e., containing 256 3D atoms of size $4 \times 4 \times 4$. From the results presented in Section III, it is found that the proposed method could consistently deliver high-quality reconstructions despite the different voxel size settings in the HN patient case ($0.5\text{mm} \times 0.5\text{mm} \times 0.5\text{mm}$) and the prostate/GYN patient cases ($1.0\text{mm} \times 1.0\text{mm} \times 1.0\text{mm}$). Therefore, it is expected that under typical isotropic voxel size settings for CBCT reconstruction, the performance of the proposed method will not heavily depend on the voxel sizes. However, if one prefers anisotropic resolution, which often occurs with helical single/multi-slice CT scans at large pitches, a 3D dictionary with anisotropic 3D atoms and different atom sizes should be considered in reference to the thickness in the longitudinal direction.

In this work, 12 realistic datasets, divided into two groups based on a GYN patient and a prostate patient and collected from different fractions and different dose levels, were employed to investigate the dose reduction potential of the proposed 3DDL method. And the experimental results indicated that when the dose level was decreased to one-eighth of the regular dose, our method still could achieve comparable or even higher image qualities as the clinical-used regular dose FDK reconstructions. Despite the insightful results, it still should be noted that the practical dose reduction

performance is site-specific, patient-specific and task-oriented, which still needs a thorough investigation. In the future, we would conduct more detailed studies to explore the extent to which the proposed 3DDL method can be applied in various clinical data.

In summary, the constructed 3D dictionary has exhibited a higher representation efficiency over the 2D dictionary, demonstrating a potential of enhancing the image quality of CBCT reconstruction. A new adaptive regularization parameter selection method, i.e., ZIP method, has been proposed to facilitate our 3DDL method. Based on two real data experiments, it is expected that by using the proposed method, the dose could be reduced to $\frac{1}{8}$ regular dose levels without obvious image quality degradations, considering that comparable STD and CNR values as the regular dose FDK reconstructions could be achieved.

ACKNOWLEDGMENT

The authors would like to thank Dr. Chenxia Li, Dr. Yonghao Du and Dr. Hui Zhang from the first affiliated hospital of Xi'an Jiaotong University for their professional evaluations on the reconstructed images. The authors would also like to thank all the anonymous reviews for their constructive suggestions. These have greatly improved the qualities of this work.

REFERENCES

- [1] D. A. Jaffray and J. H. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: Initial performance characterization," *Med. Phys.*, vol. 27, no. 6, pp. 1311–1323, 2000.
- [2] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, "Flat-panel cone-beam computed tomography for image-guided RT," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 48, no. 3, p. 194, 2000.
- [3] S. Lee *et al.*, "Intraoperative C-arm cone-beam CT: Quantitative analysis of surgical performance in skull base surgery," *Laryngoscope*, vol. 122, no. 9, p. 1925–1932, 2012.
- [4] T. Kiljunen, T. Kaasalainen, A. Suomalainen, and M. Kortensniemi, "Dental cone beam CT: A review," *Phys. Med.*, vol. 31, no. 8, pp. 844–860, 2015.
- [5] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.
- [6] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.
- [7] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.
- [8] T. Niu and L. Zhu, "Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: Phantom studies," *Med. Phys.*, vol. 39, no. 7, pp. 4588–4598, 2012.
- [9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [10] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. Technol.*, vol. 14, no. 2, pp. 119–139, 2006.
- [11] G. Yu, L. Li, J. Gu, and L. Zhang, "Total variation based iterative image reconstruction," in *Proc. 1st Int. Workshop Comput. Vis. Biomed. Image Appl.*, 2005, pp. 526–534.
- [12] J. Tang, B. E. Nett, and G.-H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Phys. Med. Biol.*, vol. 54, no. 19, pp. 5781–5804, 2009.
- [13] W. Jing, T. Li, and X. Lei, "Iterative image reconstruction for CBCT using edge-preserving prior," *Med. Phys.*, vol. 36, no. 1, pp. 252–260, 2009.
- [14] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.
- [15] Y. Liu, J. Ma, H. Zhang, J. Wang, and Z. Liang, "Low-mAs X-ray CT image reconstruction by adaptive-weighted tv-constrained penalized re-weighted least-squares," *J. X-Ray Sci. Technol.*, vol. 22, no. 4, pp. 437–457, 2014.
- [16] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," *Phys. Med. Biol.*, vol. 56, no. 13, pp. 3787–3807, 2011.
- [17] T. Niu, X. Ye, Q. Fruhauf, M. Petrongolo, and L. Zhu, "Accelerated barrier optimization compressed sensing (ABOCS) for CT reconstruction with improved convergence," *Phys. Med. Biol.*, vol. 59, no. 7, pp. 1801–1814, 2014.
- [18] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse Problems*, vol. 27, no. 11, pp. 115012–115033, 2011.
- [19] Y. Chen *et al.*, "Discriminative feature representation: An effective postprocessing solution to low dose CT imaging," *Phys. Med. Biol.*, vol. 62, no. 6, pp. 2103–2131, 2017.
- [20] Y. Chen *et al.*, "Artifact suppressed dictionary learning for low-dose CT image processing," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2271–2292, Dec. 2014.
- [21] L. Li, Z. Chen, and P. Jiao, "Dual-energy CT reconstruction based on dictionary learning and total variation constraint," in *Proc. Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2012, pp. 2358–2361.
- [22] T. Bai *et al.*, "WE-G-18A-04: 3D dictionary learning based statistical iterative reconstruction for low-dose cone beam CT imaging," *Med. Phys.*, vol. 41, no. 6, p. 527, 2014.
- [23] B. Jin, Y. Zhao, and J. Zou, "Iterative parameter choice by discrepancy principle," *IMA J. Numer. Anal.*, vol. 32, no. 4, pp. 1714–1732, 2012.
- [24] G. H. Golub and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [25] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and sure-based methods," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3659–3672, Aug. 2012.
- [26] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, p. 561–580, Dec. 1992.
- [27] L. Zhu and L. Xing, "Search for IMRT inverse plans with piecewise constant fluence maps using compressed sensing techniques," *Med. Phys.*, vol. 36, no. 5, pp. 1895–1905, 2009.
- [28] T. Reginska, "A regularization parameter in discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 17, no. 3, pp. 740–749, 1996.
- [29] J. Wang *et al.*, "An experimental study on the noise properties of X-ray CT sinogram data in radon space," *Phys. Med. Biol.*, vol. 53, no. 12, pp. 3327–3341, 2008.
- [30] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, Feb. 2002.
- [31] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$," *Doklady an SSSR*, vol. 269, no. 3, pp. 543–547, 1983.
- [32] J. H. Cho and J. A. Fessler, "Accelerating ordered-subsets image reconstruction for X-ray CT using double surrogates," in *Proc. SPIE*, 2012, pp. 145–150.
- [33] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 220–224.
- [34] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [35] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
- [36] K. Ito, B. Jin, and T. Takeuchi, "A regularization parameter for non-smooth Tikhonov regularization," *SIAM J. Sci. Comput.*, vol. 33, no. 3, pp. 1415–1438, 2011.
- [37] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. CVPR*, 1999, pp. 541–547.
- [38] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 1, no. 6, pp. 612–619, 1984.
- [39] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
- [40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.