Image reconstruction for cone-beam computed tomography using total p-variation plus Kullback–Leibler data divergence*

Ai-Long Cai(蔡爱龙), Lei Li(李磊), Lin-Yuan Wang(王林元), Bin Yan(闫镔),†
Zhi-Zhong Zheng(郑治中), Han-Ming Zhang(张瀚铭), and Guo-En Hu(胡国恩)

National Digital Switching System Engineering & Technological Research Centre, Zhengzhou 450002, *China*

(Received 19 December 2016; revised manuscript received 7 March 2017; published online 27 May 2017)

Accurate reconstruction from a reduced data set is highly essential for computed tomography in fast and/or low dose imaging applications. Conventional total variation (TV)-based algorithms apply the L1 norm-based penalties, which are not as efficient as Lp ($0 < p < 1$) quasi-norm-based penalties. TV with a p-th power-based norm can serve as a feasible alternative of the conventional TV, which is referred to as total p-variation (TpV). This paper proposes a TpV-based reconstruction model and develops an efficient algorithm. The total p-variation and Kullback–Leibler (KL) data divergence, which has better noise suppression capability compared with the often-used quadratic term, are combined to build the reconstruction model. The proposed algorithm is derived by the alternating direction method (ADM) which offers a stable, efficient, and easily coded implementation. We apply the proposed method in the reconstructions from very few views of projections (7 views evenly acquired within 180°). The images reconstructed by the new method show clearer edges and higher numerical accuracy than the conventional TV method. Both the simulations and real CT data experiments indicate that the proposed method may be promising for practical applications.

Keywords: image reconstruction, total p-variation minimization, Kullback–Leibler data divergence, p-shrinkage mapping

PACS: 87.57.nf, 87.59.bd, 87.59.bf, 87.85.Pq **DOI:** 10.1088/1674-1056/26/7/078701

1. Introduction

Computed tomography (CT) configured with circular scanning is widely used in many imaging applications,[1] such as medical imaging and industrial inspection. Accurate image reconstruction from sparse or few-view data is greatly important in areas needing fast imaging and/or low radiation dose.

When observation data is not quite adequate, iterative image reconstruction (IIR) algorithms, especially the optimization-based ones, are often considered. A typical reconstruction method is usually comprised of two aspects, the optimization models and the corresponding algorithms. The optimization models are often the combinations of regularization parts and data fidelity parts. The regularization term is always used to describe some properties of the intended image itself. This may be some sparse representations or de-noising functions on images.

The state-of-the-art regularization methods can be roughly divided into four categories. Firstly, among the earliest ones, the prior information-based ones such as the method proposed by Chen[2–4] have seen some valuable applications. The feature constrained compressed sensing reconstruction[5] is closely related to the above method. Secondly, the methods such as wavelet-, framelet-, or tight frame-based ones have been proposed in recent years, and some[6–8] of them show practical capabilities. These methods are based on harmonic analysis which can help the recovery of the image details for some applications. However, the selection of proper wavelets cannot be very easy, and the most appropriate form of the potential wavelet for practical applications is still an open question. Thirdly, the methods are dictionary learning-based ones. These methods[9,10] are usually training or applying an over-complete dictionary in the reconstruction procedure. However, training or getting such a dictionary possibly needs a large amount of available dataset and it may not work for all applications. Fourthly, much work regarding IIR algorithms has been focused on total variation (TV) minimization,[11–17] which exploits the sparsity of gradient magnitude image (GMI) of images. TV is the L1 norm of GMI and reconstruction by minimizing TV can reduce the amount of data acquisition when GMI is sparse in non-zero entities. Some extensions of TV, such as high-order TV,[18] total generalized variation,[19] and TV-stokes,[20] are proposed for image reconstruction and restoration. These forms in this type incorporate second-order (or higher) information which can help to improve the image quality but meet with sacrifices of efficiency.

The state-of-the-art optimization-based algorithms for reconstruction can also be roughly divided into four categories: 1) the adaptive-steepest-descent-projection-onto-convex-sets (ASD-POCS) framework,[11] 2) the Chambolle–Pock (CP) framework,[12,13,21,22] 3) the split-Bregman (SB) iteration,[23] and 4) the alternating direction minimization

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61372172 and 61601518).
†Corresponding author. E-mail: ybspace@hotmail.com
© 2017 Chinese Physical Society and IOP Publishing Ltd

(ADM) method.[14,15] The ASD-POCS-type methods are well designed to solve the constrained TV minimization by incorporating the traditional POCS and the steepest descent, resulting in relatively good performance for undersampled[24] or low dose dataset.[25–27] The CP framework and its extensions[28] provide a feasible tool for various problems raised in practical applications, and until now some valuable applications[29,30] confirm this property of CP. The SB originates from Bregman iteration and can solve a very broad class of l1-regularized problems. The first attempt of SB for image reconstruction, to the best of our knowledge, is carried out by Vandehinste[31] for sparse-view reconstruction. ADM is a more straightforward iteration scheme and has been widely applied in optimization community. The basic idea of ADM is to decouple the complex problem into several simple ones and deal with these simple ones alternatively and iteratively. Some of the most important applications of ADM are image reconstructions for magnetic resonance imaging (MRI),[32] CT,[33–36] and other x-ray-based imaging applications.[37]

It should be noted that the L1-normed-based TV can yield easy computation using shrinkage mapping[38] (also known as "soft thresholding"), which makes it easy to use in practical CT imaging applications. However, TV is not the most direct method to describe the GMI's sparsity. It is logical to assume that a relaxation closer to the L0 quasi-norm can measure GMI sparsity better than TV, which implies that the p-th power of the Lp norm $\|\boldsymbol{z}\|_p^p \triangleq \sum_i |z_i|^p$ $(0 < p < 1)$ is a feasible alternative. Researchers in both theory and engineering fields have made many efforts to develop practical solvers for Lp minimization.[39,40] Fortunately, recent theoretical progress[40] and applications[16,17,41] show that, although the Lp norm causes non-convex optimization problems, it may be practical for CT imaging with reduced datasets and that efficient solvers can help design reconstruction algorithms[17,42] with greater capabilities than those designed with the conventional TV model.

Besides the regularization term (i.e., the TV term), the forms of the data fidelity terms are also of great importance. The data term should account for the inconsistencies of data observation. It is widely known that the acquisition of projection can be roughly seen as the random Poisson process. Compared to the conventional quadratic term based on least square (or the L2-norm-based term), the Kullback–Leibler (KL) may have better noise suppression capabilities. The Kullback–Leibler data divergence can be a proper choice in building optimization problem. This paper combines the total p-variation (TpV) plus the KL data divergence for few-view CT image reconstruction. The practical algorithms are designed by ADM, and both simulations and real CT data experiments are carried out to verify the performances.

The remainder of this study is organized as follows. Section 2 introduces some necessary theoretical background, and presents the proposed reconstruction model and the design of the proposed algorithm. In Section 3, numerical simulations and real data experiments are used to validate the performance and capability of the proposed algorithm. In Section 4, a brief discussion is given to the results obtained from Section 3, and our conclusions are given in Section 5.

2. Method

A typical CT system mainly consists of an x-ray source, a flat panel detector, a mechanical gantry system, and a computer-based data processing system. When modeling the imaging system, a discrete form of the x-ray transform is usually taken into consideration

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}, \tag{1}$$

where $\boldsymbol{p} \in \mathbb{R}^I$ is the data vector representing the measured projections, $\boldsymbol{x} \in \mathbb{R}^s$ is the object vector, and $\boldsymbol{W} \in \mathbb{R}^{I\times s}$ is the system matrix. With the linear equation expressed above, image reconstruction is to solve Eq. (1). However, even under ideal and consistent conditions, finding the solution is not an easy task. This is mainly for two reasons, one of which is that the projection matrix $\boldsymbol{W}$ is tremendously huge thus computing its (pseudo) inverse is computationally intractable. Another reason is that $\boldsymbol{W}$ is often severely ill-conditioned having huge condition number resulting in numerical computing being unstable.

2.1. General p-shrinkage and its induced penalty function

Conventional TV minimization is often related to the optimization of an L1-norm-based function. With regard to the minimization of an L1-norm-based function, the proximal mappings of L1 functions are defined as

$$\text{prox}_\beta \|\cdot\|_1(\boldsymbol{x}) := \arg\min_{\boldsymbol{y}} \left\{ \|\boldsymbol{y}\|_1 + \frac{\beta}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \right\}, \tag{2}$$

where $\boldsymbol{x}, \boldsymbol{y}$ are vectors of the same dimension, and $\beta > 0$ is a constant scalar. The symbol := in Eq. (2) stands for definition. Actually, the solution $\tilde{\boldsymbol{y}}$ for Eq. (2) is quite concise, and it can be expressed as the soft thresholding or the shrinkage operator

$$\begin{aligned} \tilde{\boldsymbol{y}} &= \text{shrink}(\boldsymbol{x}) \\ &= \max\left(|\boldsymbol{x}| - \frac{1}{\beta}, 0\right) \cdot \text{sgn}(\boldsymbol{x}), \end{aligned} \tag{3}$$

where sgn( ) is the signum function and $|\boldsymbol{x}|$ returns a vector whose entities are absolute values of those in $\boldsymbol{x}$. We denote shrink$(\boldsymbol{x})$ as $S(\boldsymbol{x})$, and $S(\boldsymbol{x})_i = s(|\boldsymbol{x}_i|)\cdot \text{sgn}(\boldsymbol{x}_i)$ for each entity where $s(x) = \max(x - \beta^{-1}, 0)$. However, the above solution is not directly applicable for the situations of Lp $(0 < p \leq 1)$ pseudo-norm. For a class of penalty functions which have similar properties to Lp pseudo-norm-based function (depicted in the theorem in the following context), a p-shrinkage mapping

is proposed as an exact solution. The p-shrinkage mapping[40] $S_p(\boldsymbol{x}) = S_{\beta,p}(\boldsymbol{x})$ for $0 < p \leq 1$ that is considered in this paper is defined by $S_p(\boldsymbol{x})_i = s_p(|\boldsymbol{x}_i|) \cdot \operatorname{sgn}(\boldsymbol{x}_i)$, where the generalized p-shrinkage function $s_p = s_{p,\beta}$ is defined by

$$s_p(x) = \max\left\{x - \beta^{p-2}x^{p-1}, 0\right\}, \quad 0 < p \leq 1. \tag{4}$$

It is straightforward that the p-shrinkage $S_p(\boldsymbol{x})$ and the conventional shrinkage $S(\boldsymbol{x})$ coincide when $p = 1$.

The motivation for using alternative shrinkage mappings in Eq. (4) is to obtain a closed-form formula in the reconstruction algorithm, which means that the alternative shrinkage mappings are required to be proximal mappings of the penalty functions and that their induced penalty functions should be guaranteed to be better measurements of sparsity than L1. The following theorem indeed guarantees these properties and was first proposed and proven by Chartrand.[39]

Theorem 1 Suppose $s : [0,\infty) \to \mathbb{R}$ is continuous, satisfies $x \leq \beta \Rightarrow s(x) = 0$ for some $\beta \geq 0$, is strictly increasing on $[\beta, +\infty)$, and $s(x) \leq x$. Define $S_p(\boldsymbol{x})_i = s_p(|\boldsymbol{x}_i|)\operatorname{sign}(\boldsymbol{x}_i), 0 < p \leq 1$ for each i. Then, S_p is the proximal mapping of a penalty function $G_p(\boldsymbol{w}) = \sum_i g(\boldsymbol{w}_i)$, where g is even, strictly increasing, and continuous on $[0,+\infty)$, differentiable on $(0,+\infty)$, and non-differentiable at 0 if $\beta > 0$ (in which case $\partial g(0) = [-1,1]$). If $x - s(x)$ is also non-increasing on $[\beta,+\infty)$, then g is concave on $[0,+\infty)$ and G_p satisfies the triangle inequality.

The proof of the theorem constructs g using the Legendre–Fenchel transform of an anti-derivative of s. Although g cannot be guaranteed to have a closed-form expression for the penalty function G_p, this is considered an acceptable price to pay for having an explicit proximal mapping, which is much more suitable for implementing a practical, cone-beam CT algorithm than having a closed-form penalty function. In order to have an intuitive impression of the p-shrinkage and its induced penalty functions, we can compute s_p and the corresponding g numerically and some examples are plotted in Fig. 1.

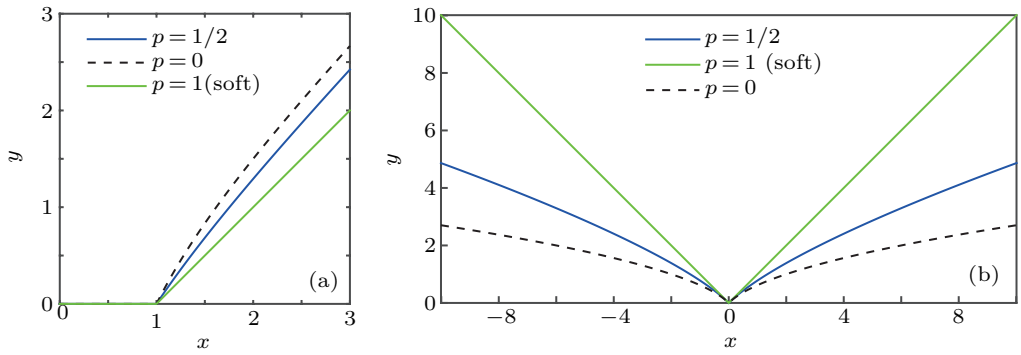


Fig. 1. (color online) Examples of (a) p-shrinkage functions and (b) their induced penalty functions for some typical values of p.

2.2. Minimization of TpV plus Kullback–Leibler term by ADM

In conventional applications, the projection data fitting or fidelity term is always built using the square of L2-norm. However, for applications with different noise properties, L2-norm is not always presenting satisfying outcomes. This generates the needs for seeking an alternative of the projection data fitting term. When the data noise is an important physical factor that must be considered as a multivariate Poisson probability distribution, we then consider the Kullback–Leibler data divergence, which is based on the maximum likelihood expectation maximization method. The KL data error function form is chosen as the fidelity term which can be written as

$$\min_{\boldsymbol{x}} \sum_i \boldsymbol{p}_i \cdot \left(\frac{\boldsymbol{W}\boldsymbol{x}}{\boldsymbol{p}} - \mathbf{1} - \ln\left(\frac{\boldsymbol{W}\boldsymbol{x}}{\boldsymbol{p}}\right)\right)_i, \tag{5}$$

where $\boldsymbol{p}_i$ and $(\cdot)_i$ denote the i-th entity in vectors, and division $\boldsymbol{W}\boldsymbol{x}/\boldsymbol{p}$ in the above equation is a component-wise operation and it returns a vector, and $\mathbf{1} \in \mathbb{R}^l$ is a vector whose entities are all 1. The KL data term is based on the fact that $f(x) = (x - 1 - \ln x) \geq 0$ for $x \in \mathbb{R}_+$. Furthermore, the function $f(x)$ on $x \in \mathbb{R}_+$ is convex which makes the optimization easily solvable.

We incorporate the KL term into the TpV model which can be written as

$$\begin{aligned}\tilde{\boldsymbol{x}} \;=\; & \arg\min_{\boldsymbol{x}} \rho_1 G_p(\boldsymbol{\nabla}\boldsymbol{x}) \\ & + \rho_2 \sum_i \left[\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p} + \boldsymbol{p}\ln\boldsymbol{p} - \boldsymbol{p}\ln\left(\operatorname{pos}(\boldsymbol{W}\boldsymbol{x})\right)\right]_i \\ & + \delta_{\mathbb{R}_+^l}(\boldsymbol{W}\boldsymbol{x}) + \delta_{\mathbb{R}_+^s}(\boldsymbol{x}),\end{aligned} \tag{6}$$

where $\rho_1 > 0$ and $\rho_2 > 0$ are the weights for two terms, and the summation $\sum_i [\boldsymbol{x}]_i$ over i is a component-wise addition for $\boldsymbol{x}$, and $\operatorname{pos}(\boldsymbol{x})$ enforces all negative entities in $\boldsymbol{x}$ to be 0 while keeping other entities unchanged. The operator G_p, the penalty function on image gradient in Eq. (6), is described and defined in the above theorem. Note that when $\boldsymbol{W}\boldsymbol{x}$ exactly equals to $\boldsymbol{p}$ for the ideal and noiseless condition, the KL term vanishes to 0. Once the data is contaminated by noise, the KL becomes non-negative values.

We introduce auxiliary variable $\boldsymbol{z}$ to decouple the TpV

term and the KL term

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{pmatrix}, \tag{7}$$

where $\boldsymbol{z}_1 = \boldsymbol{\nabla}\boldsymbol{x}, \boldsymbol{z}_2 = \boldsymbol{W}\boldsymbol{x}$. Consequently, the KL-T$p$V can be equivalently converted into

$$\begin{aligned}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}_1, \tilde{\boldsymbol{z}}_2) \;=\; & \arg\min_{\boldsymbol{x},\boldsymbol{z}_1,\boldsymbol{z}_2} \rho_1 G_p(\boldsymbol{z}_1) + \delta_{\mathbb{R}_+^s}(\boldsymbol{x}) + \delta_{\mathbb{R}_+^l}(\boldsymbol{z}_2) \\ & + \rho_2 \sum_i [\boldsymbol{z}_2 - \boldsymbol{p} + \boldsymbol{p}\ln\boldsymbol{p} - \boldsymbol{p}\ln(\mathrm{pos}(\boldsymbol{z}_2))]_i, \\ & \text{s.t.} \begin{cases} \boldsymbol{z}_1 = \boldsymbol{\nabla}\boldsymbol{x}, \\ \boldsymbol{z}_2 = \boldsymbol{W}\boldsymbol{x}. \end{cases}\end{aligned} \tag{8}$$

The corresponding augmented Lagrangian (AL) function of Eq. (8) is

$$\begin{aligned}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}_1, \tilde{\boldsymbol{z}}_2) \;=\; & \arg\min_{\boldsymbol{x},\boldsymbol{z}_1,\boldsymbol{z}_2} \rho_1 G_p(\boldsymbol{z}_1) + \delta_{\mathbb{R}_+^s}(\boldsymbol{x}) + \delta_{\mathbb{R}_+^l}(\boldsymbol{z}_2) \\ & + \rho_2 \sum_i [\boldsymbol{z}_2 - \boldsymbol{p} + \boldsymbol{p}\ln\boldsymbol{p} - \boldsymbol{p}\ln(\mathrm{pos}(\boldsymbol{z}_2))]_i \\ & + \frac{\beta_1}{2}\left\|\boldsymbol{\nabla}\boldsymbol{x} - \boldsymbol{z}_1 + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right\|_2^2 \\ & + \frac{\beta_2}{2}\left\|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{z}_2 + \frac{\boldsymbol{\lambda}_2}{\beta_2}\right\|_2^2,\end{aligned} \tag{9}$$

where $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are the multipliers, and $\beta_1, \beta_2 > 0$ are scalars. It should be pointed out that solving the above AL function is actually easily-handled. ADM applies an alternative fashion to tackle with each variable $(\boldsymbol{x}; \boldsymbol{z}_1; \boldsymbol{z}_2)$ independently.

We optimize the above problem (9) with respect to each variable in $(\boldsymbol{x}; \boldsymbol{z}_1; \boldsymbol{z}_2)$ while fixing the others and update each variable in an alternative and iterative fashion. We first obtain the following sub-problem of $\boldsymbol{z}_1$ expressed as

$$\tilde{\boldsymbol{z}}_1 = \arg\min_{\boldsymbol{z}_1} \rho_1 G_p(\boldsymbol{z}_1) + \frac{\beta_1}{2}\left\|\boldsymbol{\nabla}\boldsymbol{x} - \boldsymbol{z}_1 + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right\|_2^2. \tag{10}$$

It is obvious that the solution of Eq. (10) is exactly the p-shrinkage mapping. Applying the p-shrinkage formula directly, we can obtain the optimal solution for $\boldsymbol{z}_1$

$$\begin{aligned}\boldsymbol{z}_1 \;=\; & \max\left\{\left|\boldsymbol{\nabla}\boldsymbol{x} + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right| - \rho_1 \frac{1}{\beta_1^{p-2}}\left|\boldsymbol{\nabla}\boldsymbol{x} + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right|^{p-1}, 0\right\} \\ & \times \mathrm{sgn}\left(\boldsymbol{\nabla}\boldsymbol{x} + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right).\end{aligned}$$

The sub-problem for $\boldsymbol{z}_2$ can be obviously written as

$$\begin{aligned}\tilde{\boldsymbol{z}}_2 \;=\; & \arg\min_{\boldsymbol{z}_2} \delta_{\mathbb{R}_+^l}(\boldsymbol{z}_2) + \rho_2 \langle \boldsymbol{1}, \boldsymbol{z}_2 - \boldsymbol{p}\ln(\mathrm{pos}(\boldsymbol{z}_2))\rangle \\ & + \frac{\beta_2}{2}\left\|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{z}_2 + \frac{\boldsymbol{\lambda}_2}{\beta_2}\right\|^2,\end{aligned} \tag{11}$$

where some constants irrelevant to $\boldsymbol{z}_2$ are omitted and the component-wise summation is replaced by inner product notation. Actually the above problem can be converted into solving a quadratic equation, and we obtain the non-negative solution for $\boldsymbol{z}_2$ as

$$\boldsymbol{z}_2 = \frac{-(\rho_2 - \beta_2 C) + \sqrt{(\rho_2 - \beta_2 C)^2 + 4\beta_2\rho_2\boldsymbol{p}}}{2\beta_2}, \tag{12}$$

where $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{\lambda}_2/\beta_2$. However, the solution for $\boldsymbol{x}$ needs some derivations, and we leave the detailed derivation in Appendix A and write directly here explicitly as

$$\begin{aligned}\boldsymbol{x} \;=\; & \mathrm{pos}\left(\mathbb{F}^{-1}\left\{\boldsymbol{\Lambda}^{-1}\mathbb{F}\left(\beta_1\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{z}_1 - \boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\lambda}_1\right.\right.\right. \\ & \left.\left.\left. + \frac{\beta_2}{\tau}\boldsymbol{x}_k - \beta_2\boldsymbol{g}_k\right)\right\}\right),\end{aligned} \tag{13}$$

where $\tau > 0$, $\boldsymbol{g}_k = \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{x}_k - \boldsymbol{z}_2 + \boldsymbol{\lambda}_2/\beta_2)$, $\boldsymbol{\Lambda} = \beta_1\mathbb{F}\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\nabla}\mathbb{F}^{\mathrm{H}} + \beta_2/\tau\boldsymbol{I}$, and $\mathbb{F}$ is the FFT matrix. We conclude the pseudo-code of the proposed KL-TpV algorithm as Algorithm 1.

Algorithm 1: Pseudocode for K steps ADM iteration scheme for KL-TpV reconstruction.

1: $\xi \leftarrow 1$; $\beta_1, \beta_2 > 0$; $\rho_1 \leftarrow 1, \rho_2 \leftarrow 1$; $k \leftarrow 0$

2: $\boldsymbol{x} \leftarrow \boldsymbol{x}_0$; $\boldsymbol{z}_1 \leftarrow (\boldsymbol{z}_1)_0, \boldsymbol{z}_2 \leftarrow (\boldsymbol{z}_2)_0$; $\boldsymbol{\lambda}_1 \leftarrow \boldsymbol{0}$, $\boldsymbol{\lambda}_2 \leftarrow \boldsymbol{0}$

3: **Do**

4: $\boldsymbol{x}_k \leftarrow \mathrm{pos}\left\{\mathbb{F}^{-1}\left\{\boldsymbol{\Lambda}^{-1}\mathbb{F}\left(\beta_1\boldsymbol{\nabla}^{\mathrm{T}}(\boldsymbol{z}_1)_k - \boldsymbol{\nabla}^{\mathrm{T}}(\boldsymbol{\lambda}_1)_k + \frac{\beta_2}{\tau}\boldsymbol{x}_k - \beta_2\boldsymbol{g}_k\right)\right\}\right\}$,

5: $(\boldsymbol{z}_1)_k \leftarrow \max\left(\left|\boldsymbol{\nabla}\boldsymbol{x}_k + \frac{(\boldsymbol{\lambda}_1)_k}{\beta_1}\right| - \rho_1\frac{1}{\beta_1^{p-2}}\left|\boldsymbol{\nabla}\boldsymbol{x}_k + \frac{(\boldsymbol{\lambda}_1)_k}{\beta_1}\right|^{p-1}, 0\right)\cdot\mathrm{sgn}\left(\boldsymbol{\nabla}\boldsymbol{x}_k + \frac{(\boldsymbol{\lambda}_1)_k}{\beta_1}\right)$,

6: $(\boldsymbol{z}_2)_k \leftarrow \frac{-(\rho_2 - \beta_2\boldsymbol{W}\boldsymbol{x}_k - (\boldsymbol{\lambda}_2)_k) + \sqrt{(\rho_2 - \beta_2\boldsymbol{W}\boldsymbol{x}_k - (\boldsymbol{\lambda}_2)_k)^2 + 4\beta_2\rho_2\boldsymbol{p}}}{2\beta_2}$,

7: $(\boldsymbol{\lambda}_1)_k \leftarrow (\boldsymbol{\lambda}_1)_k + \beta_1\xi(\boldsymbol{\nabla}\boldsymbol{x}_k - (\boldsymbol{z}_1)_k)$,

8: $(\boldsymbol{\lambda}_2)_k \leftarrow (\boldsymbol{\lambda}_2)_k + \beta_2\xi(\boldsymbol{W}\boldsymbol{x}_k - (\boldsymbol{z}_2)_k)$.

9: **until** $k \geq K$, **end do**

10: **return** $\boldsymbol{x}$.

It should be noted that there are two groups of parameters, i.e., 1) the model-parameters p, ρ_1, ρ_2; and 2) the algorithm-parameters ξ, τ, β_1, and β_2. The model-parameters have impacts on the designed solutions and different choices of the model-parameters lead to different reconstructions. Unlike the model parameters that specify feasible solutions, if a reasonable range is chosen, the algorithm parameters have no impact on the solution specification. Instead, they impact the algorithm's convergence path and rate to the feasible solution set. In the selection for these parameters, a direct way is to try and see. For the convenience of implementation for readers, we offer a group of available choice in this paper. Here we choose $\beta_1 = 0.16$, $\beta_2 = 1.0$, $\rho_1 = 0.005$, $\rho_2 = 1.6 \times 10^4$, $\tau = 1.0$, and $\xi = 1.0$.

2.3. Relationship between Kullback–Leibler term and projection data noise

Although the KL data discrepancy applied here is not directly derived from the Poisson noise model, it has a very close relationship with the noise distribution. For the case of a normal dose of exposure, the x-ray CT photon measurements are often modeled as independent Poisson random variables with means of

$$\hat{I}_i(\boldsymbol{x}) = I_0 \exp\left(-\mu_i\right), \tag{14}$$

where $\mu_i = \sum_j \boldsymbol{w}_{ij}\boldsymbol{x}_j$. Therefore, the measured photons I_i are considered as

$$I_i \sim \text{Poisson}\left(\hat{I}_i(\boldsymbol{x})\right), \tag{15}$$

and then we introduce a transmission Poisson likelihood (TPL) data discrepancy term as

$$D\left(I, \hat{I}(\boldsymbol{x})\right) = \sum_i \left[\hat{I}_i(\boldsymbol{x}) - I_i - I_i \ln\left(\hat{I}_i(\boldsymbol{x})/I_i\right)\right]. \tag{16}$$

The derivation for the forming of Eq. (16) is explained in Appendix B. We expand Eq. (14) and plug it into the above equation, resulting in

$$\begin{aligned} & D\left(I, \hat{I}(\boldsymbol{x})\right) \\ = & I_0 \sum_i \left[\exp\left(-\sum_j \boldsymbol{w}_{ij}\boldsymbol{x}_j\right) - (I_i/I_0) - (I_i/I_0)\right. \\ & \left. \times \ln\left(\exp\left(-\sum_j \boldsymbol{w}_{ij}\boldsymbol{x}_j\right)/(I_i/I_0)\right)\right]. \end{aligned} \tag{17}$$

From formulas (5) and (17), it can be found that if we apply the logarithm function on each term in Eq. (17), the two formulas turn out to be the same regardless of notations.

Strictly speaking, formulas (5) and (17) are not equivalent as they cannot be directly induced by each other. However, they share some similar properties in noise suppression because 1) the exponential (or logarithmic) function is monotonic, and 2) both Eqs. (5) and (17) are nonnegative, and 3) reach the minimum value at the same position (i.e., $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}$ or equivalently $\sum_j \boldsymbol{w}_{ij}\boldsymbol{x}_j = -\ln(I_i/I_0)$). From a physics perspective, the KL and TPL data discrepancies are quite different from the often used least square (LS) ones in that KL and TPL give greater weights to higher count measurements while LS treats all measurements equally. This implies that the data discrepancy we apply here can result in better images when noises are presented in the observation. Moreover, it should be pointed out that KL and TPL can be useful even when there are data inconsistencies due to other physical factors besides the stochastic nature of the counts measurement. Furthermore, compared with the TPL data discrepancy, the KL data term is simple and convex, which is very helpful for algorithm development. This is the consideration for the application of the KL term in this work.

3. Results

3.1. Tests with ideal data set

In this experiment, a fan beam geometry is set to simulate the data acquisition of CT imaging. The typical Shepp-Logan phantom with a discretization of 256×256 pixels is utilized as the standard image. This image is generated by MATLAB code phantom ('Shepp-Logan', 256) which returns a low contrast image, widely used by tomography researchers. The distance from the x-ray source to the iso-center is 164.86 mm and the distance from the x-ray source to the center of the line detector is 1052.00 mm. The number of detector bins is 512 with a size of 0.074 mm for each.

In this experiment, 7 projections are evenly acquired within 180° along the circular trajectory. The reconstructions are carried out using alternating direction TV minimization (ADTVM),[35] which is an L2-normed-based fidelity plus TV minimization model (also termed as L2-TV hereafter), KL-TV, and KL-TpV (with different p values), and the results for each algorithm at the iteration number of 2000 are shown in Fig. 2. In addition, the root mean square errors (RMSEs) for these algorithms are listed in Table 1. The RMSE of a reconstructed image $\boldsymbol{x}$ is defined as

$$\text{RMSE} = \sqrt{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2^2 / S}, \tag{18}$$

where $\tilde{\boldsymbol{x}}$ is the background truth, S is the total number of entities in $\boldsymbol{x}$. The numbers in bold face are the relatively better outcomes. Moreover, the curves of RMSE for each algorithm are plotted in Fig. 3. Obviously, the KL-TpV algorithms give relatively clearer images compared to the conventional KL-TV algorithm.

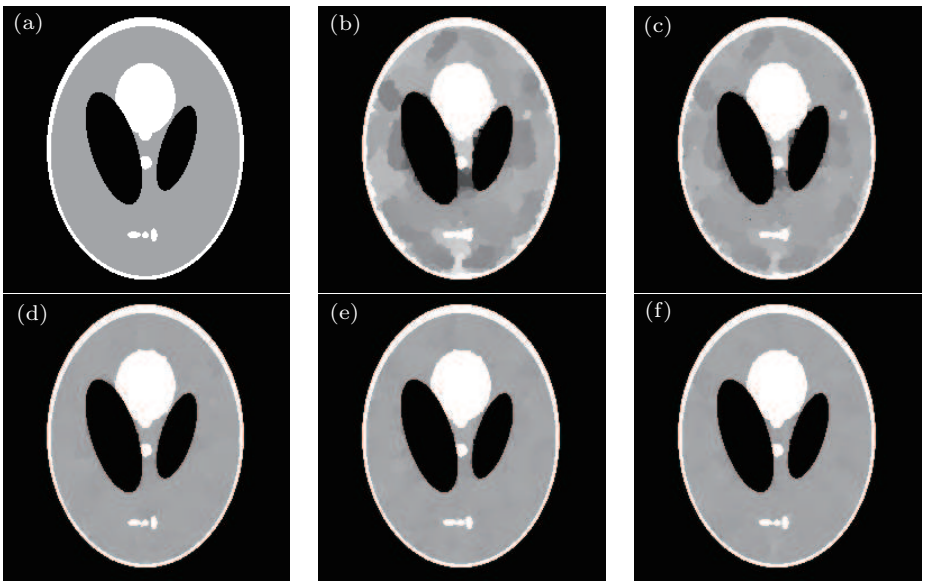


Fig. 2. Background truth and 7-view reconstructions: (a) background truth, (b) ADTVM, (c) KL-TV, (d) KL-TpV ($p = 0.94$), (e) KL-TpV ($p = 0.93$), (f) KL-TpV ($p = 0.92$). The displayed gray window is [0.016 0.022].

Table 1. Numerical comparisons of RMSE at different iterations of each algorithm for 7 views.

Iteration	100	400	600	1000	1600	2000
ADTVM	0.06084	0.01755	0.002028	0.001033	0.0008267	8.276×10^{-4}
KL-TV	0.06347	0.02577	0.009507	0.0009691	0.0007355	5.5764×10^{-4}
KL-TpV p=0.94	0.06081	0.01242	0.002133	0.0007333	**0.0003368**	2.0951×10^{-4}
KL-TpV p=0.93	0.06036	0.01022	0.001907	**0.0006879**	0.0003496	2.0457×10^{-4}
KL-TpV p=0.92	0.05991	0.008157	0.001491	0.0007092	0.0004104	$\mathbf{1.9240}\times10^{-4}$
KL-TpV p=0.90	**0.05902**	**0.004731**	**0.001418**	0.0008034	0.0003554	2.1869×10^{-4}

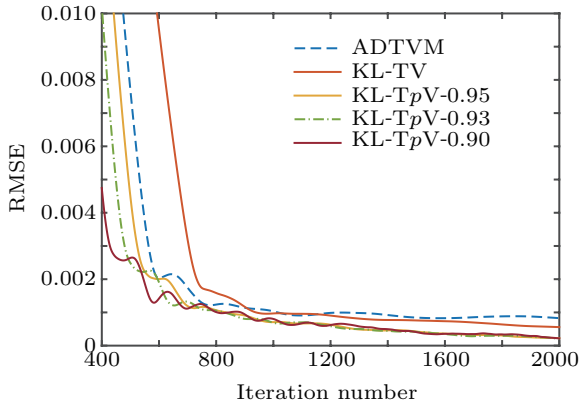


Fig. 3. (color online) Comparison of RMSE curves for ADTVM, KL-TV, and KL-TpV (with different p values) with 7 views of projections.

In the following of this subsection, some more tests are carried out in order to obtain more knowledge of the proposed method. Reconstructions using only 6 views evenly acquired within 180° are performed. Different values of p are chosen within [1.0 0.90]. It should be noted that 6 views are too few to obtain accurate reconstructions, and consequently, we properly increase the number of detector bins to 640 for each view. Moreover, the iteration number for each algorithm is extended to 10000 to obtain meaningful results. Other geometric parameters are the same as those in the first experiment. It must be declared that, for reconstructions with 6 views of projections, the objective of the following investigations is to test the convergence behavior rather than merely getting high-accuracy images since the projections are too few for high-accuracy reconstructions.

Some typical results are shown in Fig. 4 which gives us an intuitive expression of the reconstructions. ADTVM and KL-TV fail to provide clear images, while KL-TpV with $p = 0.90$ and $p = 0.91$ generates clear images. Some more numerical results are listed in Table 2. Similar to those in Table 1, the numbers in bold face are the relatively better outcomes. The behaviors of RMSE for each KL-TpV with some typical values of p are plotted in Fig. 5. A basic knowledge that can be concluded from these two plots is that KL-TpV has a faster error reduction speed (related to the value of p) than KL-TV or ADTVM does.

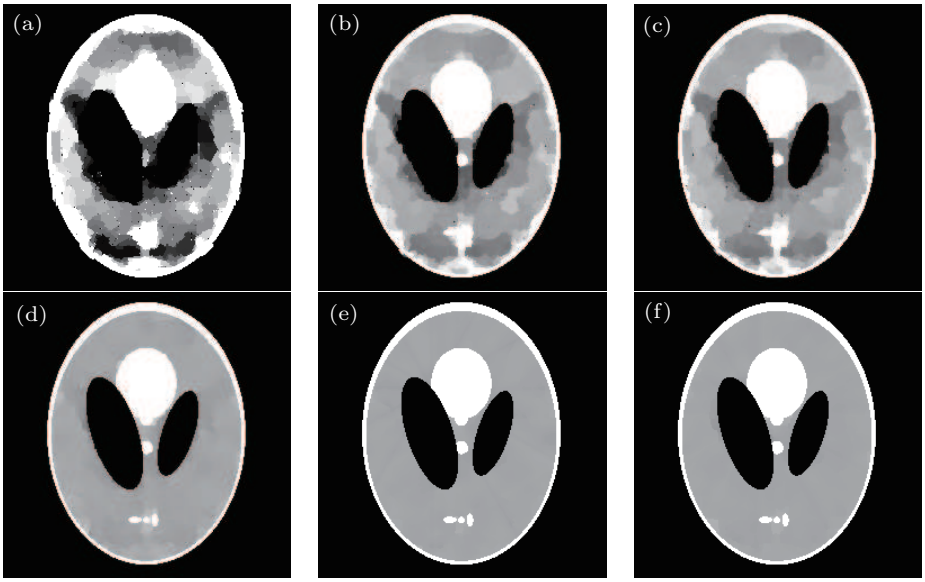


Fig. 4. Reconstructions with 6 views of projections. The results of (a) ADTVM, (b) KL-TV, (c) KL-TpV ($p = 0.95$), (d) KL-TpV ($p = 0.93$), (e) KL-TpV ($p = 0.91$), and (f) KL-TpV ($p = 0.90$) at the iteration number of 10000. The displayed gray window is [0.016 0.022].

Table 2. Numerical comparison of RMSE at different iterations of each algorithm for 6 views.

Iteration	600	1000	2000	4000	8000	10000
ADTVM	0.04936	0.0405	0.0270	0.0182	0.0154	0.0147
KL-TV	0.0429	0.0216	0.0033	0.0038	0.0038	0.0037
KL-TpV $p = 0.96$	0.0322	0.0022	0.0017	0.0016	0.0011	0.0012
KL-TpV $p = 0.95$	0.0292	0.0016	0.0017	0.0014	0.0012	0.0011
KL-TpV $p = 0.93$	0.0225	0.0016	0.0013	9.151×10^{-4}	3.568×10^{-4}	1.697×10^{-4}
KL-TpV $p = 0.91$	0.0156	0.0014	0.0012	7.879×10^{-4}	2.184×10^{-4}	$\mathbf{7.507}\times10^{-5}$
KL-TpV $p = 0.90$	**0.0121**	**0.0014**	**0.0011**	$\mathbf{6.846}\times10^{-4}$	$\mathbf{1.107}\times10^{-4}$	7.932×10^{-5}

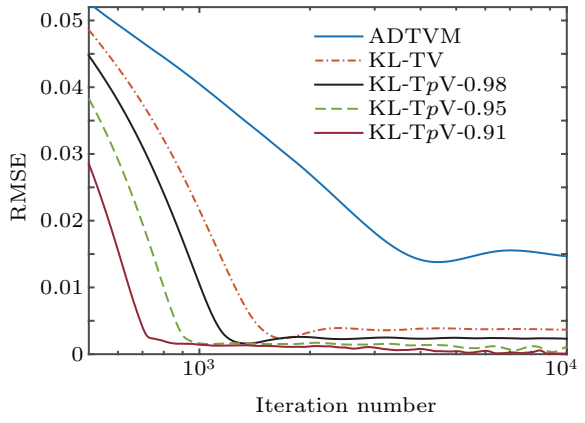


Fig. 5. (color online) Comparison of RMSEs for ADTVM, KL-TV, and KL-TpV (with different p values) with 6 views of projections.

Another important issue remains in the selection of values of p, and here we carry out a comparison between KL-TpV algorithms with different p with 7 and 6 views of projections. In the reconstructions with 7 and 6 views, the best RMSEs of each reconstruction are plotted as the functions of p in Fig. 6. The plots in Fig. 6 indicate that a proper range of p is in (0.91 0.93). Therefore, 0.92 is selected for p in the following tests including simulations with noisy data and real CT projections.

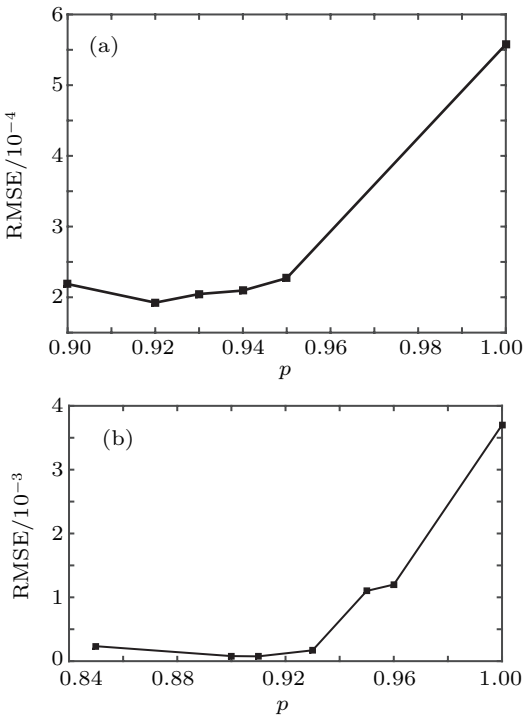


Fig. 6. (color online) Best RMSEs of each reconstruction as the functions of p with 7 views (a) and 6 views (b).

3.2. Simulations with Poisson random noise

Reconstructions with noisy projections are carried out and demonstrated in this subsection. The objective of this simulation is to test the preliminary performance between KL- and L2-norm-based data fidelity in the reconstructions. The Poisson noise is added into the simulated projections. Similar to the previous studies, the phantom we apply here is also the low contrast Shepp–Logan phantom.

In the simulations, the initial photon count is 66000 for the utilized low contrast phantom in the generation of the projection data. In order to get an expression of the noise level, the SART reconstruction with 361 views of projections acquired evenly within 360° is shown in Fig. 7. It should be pointed out that although the noise level here is not very high compared with the background truth image, reconstructions with very few views (e.g. less than 10 projections) are still very difficult tasks. Even very tiny data inconsistencies can result in reconstruction errors for sparse views. The comparisons are carried out using the ASD-POCS (implemented by SART and TV minimization, termed as SART-TV in this paper), ADTVM (L2-TV), KL-TV, and the proposed KL-TpV algorithm.

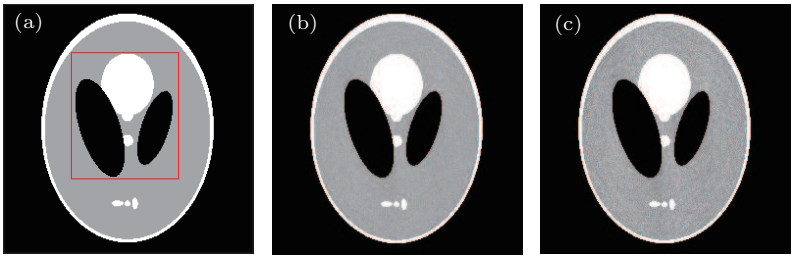


Fig. 7. (color online) Background truth image and SART reconstruction with noise projections: (a) background truth image and (b), (c) SART reconstruction with 361 views under different displayed windows. The displayed gray scale window is [0.016 0.022] in panel (b) and [0.018 0.021] in panel (c).

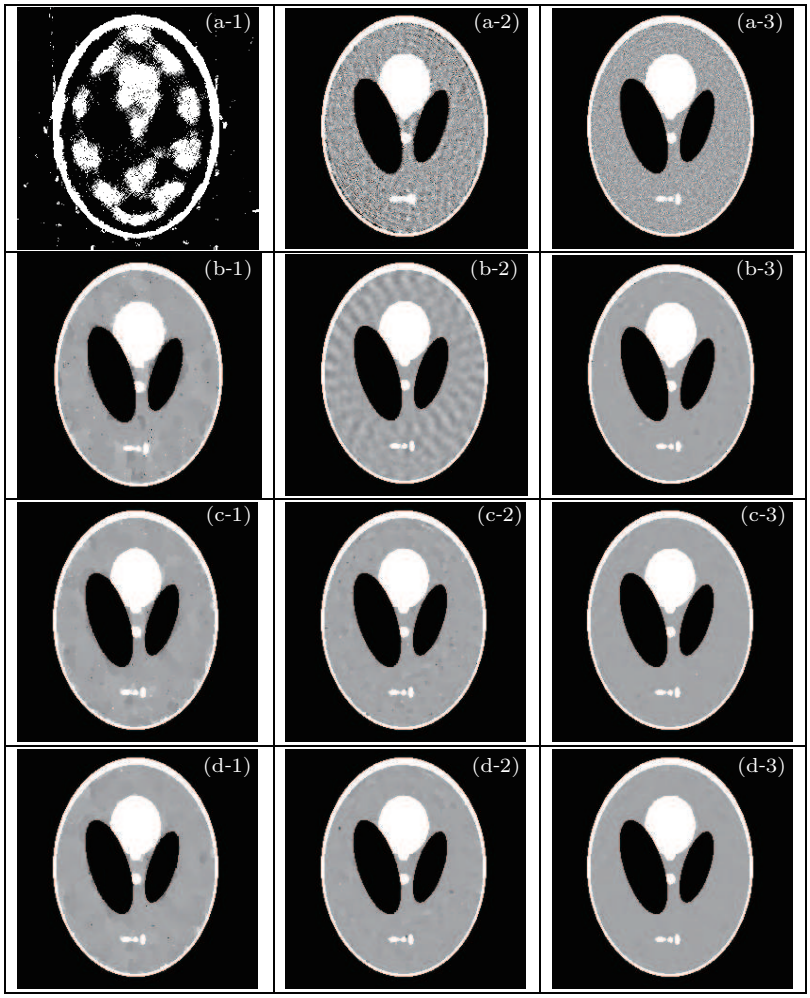


Fig. 8. (color online) Reconstruction comparisons for (a) SART-TV, (b) ADTVM, (c) KL-TV, and (d) KL-TpV with different projection views with Poisson noise. From the left column to the right column, images are reconstructed from 9, 30 and 60 views at the iteration number of 2000. The displayed gray window is [0.018 0.021].

Table 3. Numerical comparisons for each algorithm with noisy data under different views.

View number	9 views		30 views		60 views	
RMSE/SNR	RMSE	SNR in ROI/dB	RMSE	SNR in ROI/dB	RMSE	SNR in ROI/dB
SART-TV	0.06054	10.7533	9.193×10^{-4}	47.1260	2.833×10^{-4}	57.3496
ADTVM (L2-TV)	3.8986×10^{-4}	55.9871	1.084×10^{-4}	65.6909	7.612×10^{-5}	68.7648
KL-TV	$\mathbf{3.208}\times10^{-4}$	**56.2699**	$\mathbf{9.259}\times10^{-5}$	**67.0635**	$\mathbf{5.710}\times10^{-5}$	**71.2620**
KL-TpV (p=0.92)	$\mathbf{3.01}\times10^{-4}$	**56.8234**	$\mathbf{8.410}\times10^{-5}$	**68.9225**	$\mathbf{4.810}\times10^{-5}$	**73.7318**

In this test, 9, 30, and 60 projections evenly acquired within 180° are applied to perform the reconstructions. Typical results are shown in Fig. 8. The signal-to-noise ratio (SNR) of a region of interest (ROI, depicted by a red rectangle in Fig. 7) of each reconstruction is calculated and listed in Table 3. The SNR of $\boldsymbol{x}_{\mathrm{ROI}}$ is computed as

$$\mathrm{SNR} = 20\lg\frac{\|\tilde{\boldsymbol{x}}_{\mathrm{ROI}}\|_2}{\|\tilde{\boldsymbol{x}}_{\mathrm{ROI}} - \boldsymbol{x}_{\mathrm{ROI}}\|_2}, \tag{19}$$

where $\tilde{\boldsymbol{x}}_{\mathrm{ROI}}$ is the background truth in ROI. Also the RMSEs of the entire reconstructed images are listed in Table 3. With visual comparison, the KL-based algorithms (i.e., KL-TV and KL-TpV) give better results than SART-TV or ADTVM (L2-TV) does. These observations are also verified by the SNRs in ROI and the RMSEs of the reconstructions in Table 3.

3.3. Verifications with real CT projections

In this subsection, real CT verifications are carried out to test the proposed algorithm. These projections are acquired from an adult male rabbit *in vivo* on a CBCT imaging system. The distance from the source to the rotation center is 502.8 mm, and the distance from the source to the center of the detector is 1434.7 mm. The flat detector has 1024×768 bins with pixel size of 0.3810 mm×0.3810 mm for each. The reconstructed image is of 512×512×384 with 0.2671 mm×0.2671 mm×0.2671 mm for each voxel. The working voltage and current of the x-ray tube are 100 kV and 200 μA. There are 360 projections acquired evenly within the range of 360°. We first perform the reconstruction with the real CT projections using the standard FDK algorithm. Both full views and sparse views (60 projections evenly chosen with angular step of 3° within 180°) are conducted and the typical results are shown in Fig. 9.

For iterative reconstructions, it is well known that conducting fully three-dimensional reconstructions is very time consuming because the forward- and backward-projection operations are of severely high computational complexity. In this paper, accelerations for the two operations using graphic processing unit (GPU) under NVidia CUDA framework are incorporated in the implementation. Specifically, a fast and parallel algorithm for projection developed by Gao[43] is implemented on the NVidia Tesla K40c card. With the hardware acceleration, the time consumption for each iteration loop is about 40 s.

Reconstruction from full views serves an expression of the references while the result from sparse views indicates the limitation of the analytical algorithm. In Fig. 9, three slices in the transverse direction are selected to demonstrate the results of full and sparse-view reconstruction by FDK. Images reconstructed from sparse views by FDK suffer severely from the streak artifacts and lots of tiny structures are degraded.

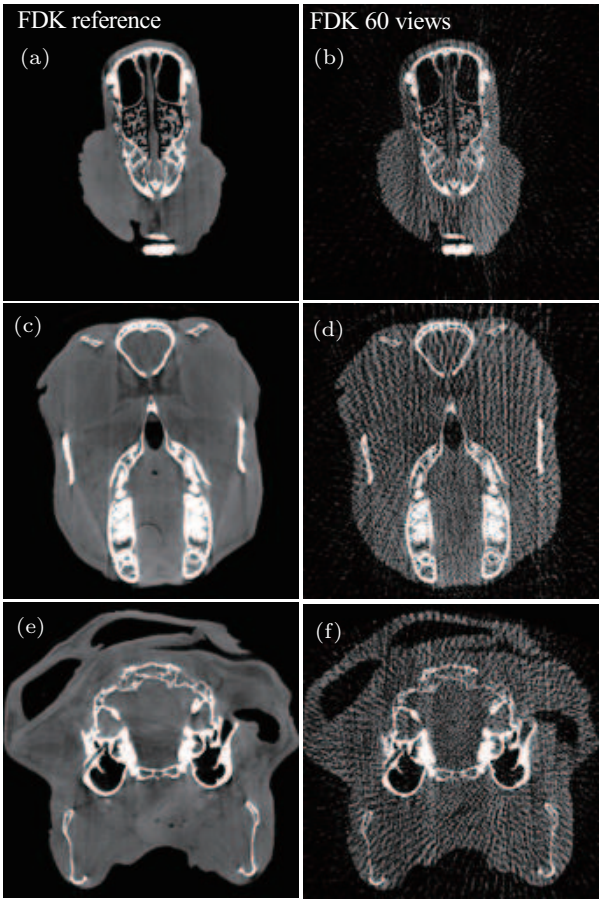


Fig. 9. (color online) FDK reconstructions with real CT projections. (a), (c), (e) FDK reconstructions with full views (i.e., 360 projections) and (b), (d), (f) FDK reconstructions with sparse views (60 projections). From the top to the bottom row, images are at the 72$^{\mathrm{th}}$, 192$^{\mathrm{th}}$ (central transverse section), and 302$^{\mathrm{th}}$ transverse sections. The displayed gray scale window is [0.01 0.06].

We carry out the iterative reconstructions using 60 views of projections acquired the same as sparse-view reconstruction of FDK above. The iterative algorithms applied are the SART-TV, ADTVM (L2-TV), and the proposed KL-TpV ($p = 0.92$) method. For iterative algorithms under real CT projections, it is an essential problem to properly set the stop criterion. We test the KL-TpV algorithms with the real CT data and plot the

relative change (in RMSE) of the current image with the previous adjacent one in Fig. 10. The relative RMSE of the image $\boldsymbol{x}$ at the iteration number k is defined as

$$\text{rel.RMSE}(k) = \sqrt{\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\|_2^2 / S}, k \geq 2, \tag{20}$$

where S is the total number of entities of $\boldsymbol{x}$. It can be easily known that when the iteration number is between 80 and 100 (the red part in Fig. 10), the relative change between two outcomes of adjacent iterations is very tiny and almost no visible change can be observed. Consequently, we set the iteration number to 100 for SART-TV, ADTVM, and KL-TpV algorithms.

Slices on three different directions (sagittal, coronal, and transverse sections) are chosen to demonstrate the reconstruction of each algorithm. Iterative results (with 60 projections) are shown in Figs. 11 and 12. The results of KL-TpV under sparse views are very similar to those of full-view FDK. The comparisons between SART-TV, ADTVM, and KL-TpV indicate that these methods give very close outcomes. However, the KL-TpV results show better qualities in image contrast and artifacts. We take the FDK reconstruction with full views as the reference and calculate the RMSEs of the selected slices in the iterative reconstructions. The RMSEs of the selected slices on the transverse section are listed in Table 4, and those of sagittal and coronal sections are listed in Table 5. Both the images and the numerical statistics indicate that the proposed KL-TpV algorithm can give better reconstructions. In order to compare the details in the structures of the rabbit head of the reconstructions, several ROIs are chosen on slices of sagittal and coronal sections and the zoomed pictures are also shown in Fig. 12.

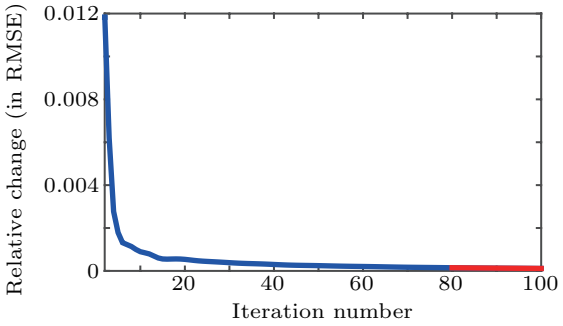


Fig. 10. (color online) The reduction curve of relative changes (in RMSE) of two adjacent outcomes of KL-TpV iterations. The red part indicates that the changes are little on images, and it means that the images are becoming stable.

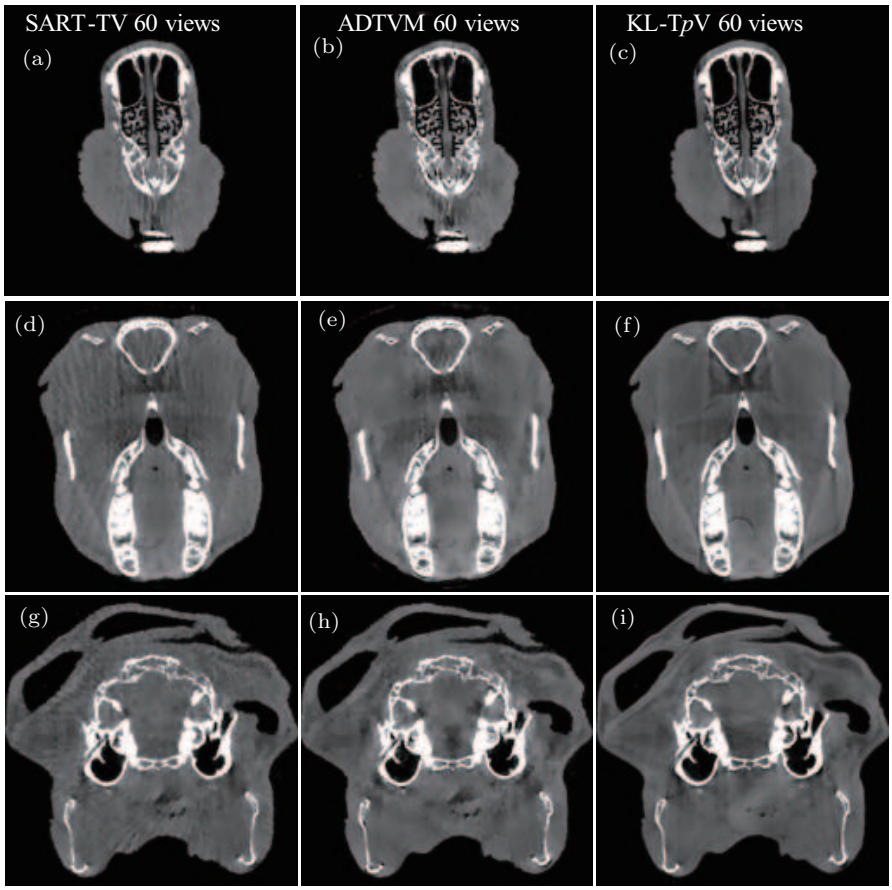


Fig. 11. (color online) Comparisons of reconstructions given by (a), (d), (g) SART-TV, (b), (e), (h) ADTVM, and (c), (f), (i) KL-TpV ($p = 0.92$) on transverse section with 60 views. From the top to the bottom row, the images are results in the 72$^{\text{th}}$, 192$^{\text{th}}$, and 302$^{\text{th}}$ slice of the transverse section. The gray scale of the displayed window is set to [0.01 0.06].

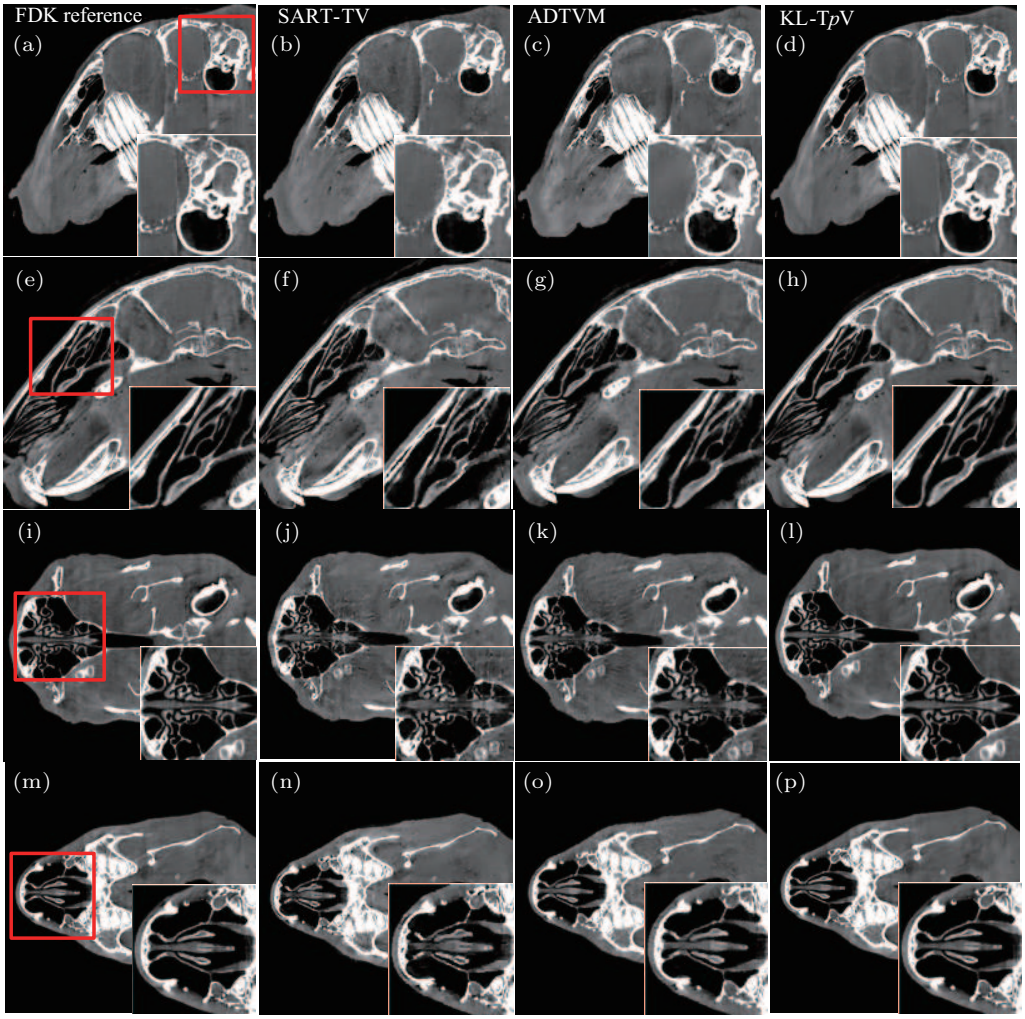


Fig. 12. (color online) Zoomed comparisons of FDK reference, SART-TV, ADTVM, and KL-TpV on sagittal and coronal sections. From the left to the right column the images are results of FDK (full views), SART-TV (sparse views, 60 projections), ADTVM (sparse views, 60 projections), and KL-TpV (sparse views, 60 projections, $p = 0.92$). Panels (a)–(d) and (e)–(h) are results of the 300$^{\text{th}}$ and 275$^{\text{th}}$ slice on the sagittal section. Panels (i)–(l) and (m)–(p) are results of the 160$^{\text{th}}$ and 200$^{\text{th}}$ slice on the coronal section. The gray scale of the displayed window is set to [0.01 0.06].

Table 4. RMSEs of the selected slices on the transverse section.

	SART-TV	ADTVM	KL-TpV ($p = 0.92$)
Slice 72	0.1241	0.0675	0.0153
Slice 192	0.1328	0.0726	0.0173
Slice 302	0.1191	0.0762	0.0171

Table 5. RMSEs of the selected slices on the sagittal and coronal sections.

	SART-TV	ADTVM	KL-TpV ($p = 0.92$)
Slice 300 (sagittal)	0.1635	0.1034	0.0170
Slice 275 (sagittal)	0.1588	0.1102	0.0176
Slice 160 (coronal)	0.1375	0.0922	0.0164
Slice 200 (coronal)	0.1308	0.0759	0.0161

4. Discussion

Although the conventional TV is an often used convex relaxation to the L0 quasi-norm of image gradient, we point out in Section 1 that a better approximation to L0 is the Lp norm-based ones. In this paper, we apply the p-shrinkage induced penalty for forming the reconstruction model. This model is theoretically a better alternative of the TV because the TpV may achieve better images with very few views. In the simulation experiments, the gradient image of the utilized phantom has 2184 non-zero entities, the TV method needs more than $2184 \times 2/512 = 8.5313$ views, and consequently it is impossible for the TV method to reconstruct accurate images with only 7 views according to the sufficient-sampling conditions[44] for sparsity-exploiting reconstruction. However, the reconstructions using TpV are of different situations. The results in Fig. 2 show that the TpV method outperforms the TV method in image quality. Moreover, the numerical comparisons in Table 1 and RMSE curves in Fig. 3 also confirm this judgment.

The optimization built for reconstruction applies the KL data term which shows capabilities in noise suppression over the conventional L2-norm-based methods. The KL term has a very close relationship to the TPL term which is drawn from the maximum likelihood under Poisson random noise assumption. Comparing the results in Figs. 7 and 8, one can eas-

ily find that the images provided by the KL-TpV are clearer than those of SART. Note that the observation data (with 361 views) for SART is relatively sufficient (since the observation number of equations is bigger than that of image pixels), but it still cannot provide satisfying quality. However, KL-TpV algorithms, with only 60 views, generate relatively better outcomes. This can be explained by two aspects: 1) the smoothness function of TpV minimization and 2) noise suppression property of the KL data term. In the last part of Section 3, the real CBCT projections are utilized for tests, which verifies that the proposed method can serve as an alternative of the popular ASD-POCS method with almost no sacrifice on image qualities. The proposed method implemented in this paper is only on the working version stage, and only some necessary optimizations and accelerations are incorporated when conducting three-dimensional reconstructions with big scale of data. For sparse-view situation, the proposed method gives out images with almost the same qualities as the full-view FDK results confirmed by Figs. 11 and 12.

Another essential point is that the KL-TpV reconstruction is designed using ADM rather than CP or SB. This is for the reason that the KL-TpV model is in fact a non-convex function and ADM does not restrict strictly the convexity of the objectives while CP or SB is not directly applicable for this model. Moreover, we apply the ADM for algorithm derivation which can easily incorporate the p-shrinkage operator in the implementation. Even so, we find that, by the experiments, the relatively proper range for p is (0.91 0.93), and certainly some necessary tuning on parameters can improve its performance.

5. Conclusion

This paper proposed an efficient reconstruction algorithm using total p-variation plus Kullback–Leibler data divergence. We introduce the p-shrinkage method in the algorithm which leads to an easy and concise implementation. This new method is derived by ADM and can reach promising results for very few views of projections compared with the conventional method. The experiments demonstrate the effectiveness of this method. However, this is only preliminary research and result. Further optimizations and investigations should be studied in the future.

Appendix A

The $\boldsymbol{x}$ sub-problem can be written as

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \delta_{\mathbb{R}_+^s}(\boldsymbol{x}) + \frac{\beta_1}{2}\left\|\nabla\boldsymbol{x} - \boldsymbol{z}_1 + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right\|^2 + \frac{\beta_2}{2}\left\|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{z}_2 + \frac{\boldsymbol{\lambda}_2}{\beta_2}\right\|^2, \tag{A1}$$

where the variables $\boldsymbol{z}_1, \boldsymbol{z}_2$ are fixed as constants. We use a replacement

$$\left\|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{z}_2 - \frac{\boldsymbol{\lambda}_2}{\beta_2}\right\|_2^2 \approx \left\|\boldsymbol{W}\boldsymbol{x}_k - \boldsymbol{z}_2 - \frac{\boldsymbol{\lambda}_2}{\beta_2}\right\|_2^2 + 2\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{\tau}\|\boldsymbol{x} - \boldsymbol{x}_k\|_2^2, \tag{A2}$$

where $\tau > 0$ is a scalar in $(0, 2)$ and $\boldsymbol{g}_k = \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{x}_k - \boldsymbol{z}_2 - \boldsymbol{\lambda}_2/\beta_2)$. Plug Eq. (A2) into Eq. (A1) and omit the constant part

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \delta_{\mathbb{R}_+^s}(\boldsymbol{x}) + \frac{\beta_1}{2}\left\|\nabla\boldsymbol{x} - \boldsymbol{z}_1 + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right\|^2 + \frac{\beta_2}{2}2\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{\tau}\|\boldsymbol{x} - \boldsymbol{x}_k\|_2^2, \tag{A3}$$

where $\delta_{\mathbb{R}_+^s}(\boldsymbol{x})$ means that all the entities of $\boldsymbol{x}$ should be non-negative. When handling Eq. (A3), we first leave the indicator function and check $\boldsymbol{x}$ last. Thus, computing the derivative of the following function alternatively

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{\beta_1}{2}\left\|\nabla\boldsymbol{x} - \boldsymbol{z}_1 + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right\|^2 + \frac{\beta_2}{2}2\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{\beta_2}{2}\frac{1}{\tau}\|\boldsymbol{x} - \boldsymbol{x}_k\|_2^2, \tag{A4}$$

and setting it to 0, results in

$$\beta_1\nabla^{\mathrm{T}}\left(\nabla\boldsymbol{x} - \boldsymbol{z}_1 + \frac{\boldsymbol{\lambda}_1}{\beta_1}\right) + \beta_2\boldsymbol{g}_k + \frac{\beta_2}{\tau}(\boldsymbol{x} - \boldsymbol{x}_k) = 0. \tag{A5}$$

Consequently, the solution for $\boldsymbol{x}$ is

$$\left(\beta_1\nabla^{\mathrm{T}}\nabla + \frac{\beta_2}{\tau}\right)\boldsymbol{x} = \beta_1\nabla^{\mathrm{T}}\boldsymbol{z}_1 - \nabla^{\mathrm{T}}\boldsymbol{\lambda}_1 + \frac{\beta_2}{\tau}\boldsymbol{x}_k - \beta_2\boldsymbol{g}_k. \tag{A6}$$

It should be noted that on the periodic definition condition, the matrix form of $\nabla^{\mathrm{T}}\nabla$ is block circulant. Therefore, $\nabla^{\mathrm{T}}\nabla$ can be diagonalized using FFT transform as

$$\boldsymbol{\Lambda} = \beta_1\mathbb{F}\nabla^{\mathrm{T}}\nabla\mathrm{F}^{\mathrm{H}} + \frac{\beta_2}{\tau}\boldsymbol{I}, \tag{A7}$$

where $\mathbb{F}^{\mathrm{H}}$ denotes the Hermite form of $\mathbb{F}$. The matrix $\boldsymbol{\Lambda}$ is diagonal and all the entities in the principle diagonal are all positive, which makes it easy to compute its inverse. Thus, the solution can be easily computed as

$$\boldsymbol{x} = \mathbb{F}^{-1}\left\{\boldsymbol{\Lambda}^{-1}\mathrm{F}\left(\beta_1\nabla^{\mathrm{T}}\boldsymbol{z} - \nabla^{\mathrm{T}}\boldsymbol{\lambda}_1 + \frac{\beta_2}{\tau}\boldsymbol{x}_k - \beta_2\boldsymbol{g}_k\right)\right\}. \tag{A8}$$

Appendix B

The transmission Poisson likelihood data term is derived from the negative ln likelihood of Poisson stochastic model of the photon measurements of Eq. (15). This means that minimizing this data term is equivalent to maximizing the likelihood function. The probability distribution of the observed photon measurements can be expressed as

$$P(I = I_i) = \frac{\mathrm{e}^{-\lambda}\lambda^{I_i}}{I_i!}. \tag{B1}$$

If we apply the following likelihood function

$$L(I,\hat{I}(\boldsymbol{x})) = -\ln\left(\prod_i \frac{P(I=I_i)|_{\lambda=I_i}}{P(I=I_i)|_{\lambda=\hat{I}_i(\boldsymbol{x})}}\right), \tag{B2}$$

and use the notation

$$D(I,\hat{I}(\boldsymbol{x})) = \ln\left(\prod_i \frac{P(I=I_i)|_{\lambda=I_i}}{P(I=I_i)|_{\lambda=\hat{I}_i(\boldsymbol{x})}}\right), \tag{B3}$$

with some simple and direct derivations (omitted here), $D(I,\hat{I}(\boldsymbol{x}))$ has a very brief expression as

$$D(I,\hat{I}(\boldsymbol{x})) = \sum_i \left[\hat{I}_i(\boldsymbol{x}) - I_i - I_i \ln\left(\hat{I}_i(\boldsymbol{x})/I_i\right)\right]. \tag{B4}$$

The above expression is exactly the same as Eq. (16).

References

[1] Buzug T M 2008 *Computed Tomography: From Photon Statistics to Modern Cone-beam CT* (1st Edn.) (Berlin Heidelberg: Springer-Verlag) p. 1
[2] Lauzier P T, Tang J and Chen G H 2012 *Phys. Med. Biol.* **57** 2461
[3] Chen G H, Theriault-Lauzier P, Jie T, Nett B, Shuai L, Zambelli J, Zhihua Q, Bevins N, Raval A, Reeder S and Rowley H 2012 *Medical Imaging, IEEE Transactions on* **31** 907
[4] Chen G H, Tang J and Leng S 2008 *Medical Physics* **35** 660
[5] Wu D, Li L and Zhang L 2013 *Phys. Med. Biol.* **58** 4047
[6] Ding H, Gao H, Zhao B, Cho H M and Molloi S 2014 *Phys. Med. Biol.* **59** 6005
[7] Ramani S and Fessler J A 2012 *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on* pp. 1008–1011
[8] Zhao B, Gao H, Ding H and Molloi S 2013 *Medical Physics* **40** 031905
[9] Xu Q, Yu H, Mou X, Zhang L, Hsieh J and Wang G 2012 *IEEE Transactions on Medical Imaging* **31** 1682
[10] Xu Q, Yu H, Wang G and Mou X 2014 *Frontiers of Medical Imaging* pp. 99–109
[11] Sidky E Y and Pan X C 2008 *Phys. Med. Biol.* **53** 4777
[12] Sidky E Y, Jørgensen J H and Pan X C 2012 *Phys. Med. Biol.* **57** 3065
[13] Sidky E Y, Jørgensen J H and Pan X C 2013 *Medical Physics* **40** 031115
[14] Li C 2009 *An Efficient Algorithm For Total Variation Regularization With Applications To The Single Pixel Camera And Compressive Sensing* (PhD Dissertation) Rice University
[15] Wang Y, Yang J, Yin W and Zhang Y 2008 *SIAM Journal on Imaging Sciences* **1** 248
[16] Chartrand R, Sidky E Y and Pan X C 2013 *Proceedings of Asilomar Conference on Signal System and Computers* pp. 665–669
[17] Sidky E Y, Chartrand R, Boone J M and Pan X C 2014 *IEEE Journal of Translational Engineering in Health and Medicine* **2** 1
[18] Yang J, Yu H Y, Jiang M and Wang G 2010 *Inverse Problems* **26** 350131
[19] Bredies K, Kunisch K and Pock T 2010 *SIAM Journal on Imaging Sciences* **3** 349
[20] Liu Y, Liang Z, Ma J, Lu H, Wang K, Zhang H and Moore W 2014 *IEEE Transactions on Medical Imaging* **33** 749
[21] Chambolle A and Pock T 2010 *Journal of Mathematical Imaging and Vision* **40** 1
[22] Pock T and Chambolle A 2011 *IEEE International Conference on Computer Vision (ICCV): IEEE* pp. 1762–1769
[23] Goldstein T and Osher S 2009 *SIAM Journal on Imaging Sciences* **2** 323
[24] Han X, Bian J G, Ritman E L, Sidky E Y and Pan X C 2012 *Phys. Med. Biol.* **57** 5245
[25] Bian J G, Han X, Yang K, Sidky E Y, Boone J M and Pan X C 2011 *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE* pp. 2566–2568
[26] Bian J G, Han X, Sidky E Y, Siewerdsen J H and Pan X C 2010 *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE* pp. 3479–3482
[27] Bian J G, Yang K, Boone J M, Han X, Sidky E Y and Pan X C 2014 *Phys. Med. Biol.* **59** 2659
[28] Barber R F and Sidky E Y 2016 arXiv:1510.08842v08842
[29] Sidky E Y, Kraemer D N, Roth E G, Ullberg C, Reiser I S and Pan X C 2014 *Journal of Medical Imaging* (Bellingham)**1** 031007
[30] Barber R F, Sidky E Y, Schmidt T G and Pan X 2015 arXiv:1511.03384v03381
[31] Vandeghinste B, Goossens B, Beenhouwer J, Pizurica A, Philips W, Vandenberghe S and Staelens S 2011 *11th International Meeting on Fully Three-Dimensional Reconstruction in Radiology and Nuclear Medicine* pp. 431–434
[32] Yang J, Zhang Y and Yin W 2010 *IEEE Journal of Selected Topics in Signal Processing* **4** 288
[33] Wang L Y, Cai A L, Zhang H M, Yan B, Li L, Hu G E and Bao S L 2015 *Journal of X-ray Science and Technology* **23** 83
[34] Cai A L, Wang L Y, Zhang H M, Yan B, Li L, Xi X Q and Li J X 2014 *Journal of X-ray Science and Technology* **22** 335
[35] Zhang H M, Wang L Y, Yan B, Li L, Xi X Q and Lu L Z 2013 *Chin. Phys. B* **22** 078701
[36] Wang L Y, Cai A L, Zhang H M, Yan B, Li L and Hu G E 2013 *Computational and Mathematical Methods in Medicine* **2013**
[37] Li S P, Wang L Y, Yan B, Li L and Liu Y J 2012 *Chin. Phys. B* **21** 108703
[38] Daubechies I, Defrise M and De Mol C 2004 *Communications on Pure and Applied Mathematics* **57** 1413
[39] Chartrand R 2014 *IEEE International Conference on Acoustic, Speech and Signal Processing* pp. 1025–1029
[40] Woodworth J and Chartrand R 2014 *Inverse Problems* **32** 075004
[41] Chartrand R 2009 Biomedical Imaging: From Nano to Macro, 2009 *ISBI '09 IEEE International Symposium on* pp. 262–265
[42] Cai A L, Wang L Y, Yan B, Li L, Zhang H M and Hu G E 2015 *Computerized Medical Imaging and Graphics* **45** 1
[43] Gao H 2012 *Medical Physics* **39** 7110
[44] Jorgensen J S, Sidky E Y and Pan X 2013 *IEEE Transactions on Medical Imaging* **32** 460