SIAM J. IMAGING SCIENCES
Vol. 14, No. 2, pp. 749–777

Limited-Angle CT Reconstruction via the L_1/L_2 Minimization*

Chao Wang†, Min Tao‡, James G. Nagy§, and Yifei Lou†

Abstract. In this paper, we consider minimizing the L_1/L_2 term on the gradient for a limited-angle scanning problem in computed tomography (CT) reconstruction. We design a specific splitting framework for an unconstrained optimization model so that the alternating direction method of multipliers (ADMM) has guaranteed convergence under certain conditions. In addition, we incorporate a box constraint that is reasonable for imaging applications, and the convergence for the additional box constraint can also be established. Numerical results on both synthetic and experimental datasets demonstrate the effectiveness and efficiency of our proposed approach, showing significant improvements over the state-of-the-art methods in the limited-angle CT reconstruction.

Key words. L_1/L_2 minimization, limited-angle computed tomography, alternating direction method of multipliers, nonconvex optimization

AMS subject classifications. 65K10, 68U10, 49N45, 49M20, 92C55

DOI. 10.1137/20M1341490

1. Introduction.

Recent developments in science and technology have led to a revolution in data processing, as large datasets are becoming increasingly available and useful. In medical imaging, a series of imaging modalities, such as x-ray computed tomography (CT) [31, 5, 13, 14], magnetic resonance imaging (MRI) [42], and electroencephalography (EEG) [34, 35], offer different perspectives to facilitate diagnostics. On the other hand, however, one often faces "small data," e.g., only a small number of CT scans are allowed for the sake of radiation dose. In this paper, we are particularly interested in a limited-angle CT reconstruction problem, which often occurs in many medical imaging applications. In breast imaging, a technique gaining wide interest is tomosynthesis (sometimes referred to as 3D mammography) [65, 73], which is a limited angle tomography approach designed to produce pseudo three-dimensional images while keeping the radiation exposure to approximately that of traditional two-dimensional mammograms.

*Received by the editors June 1, 2020; accepted for publication (in revised form) March 19, 2021; published electronically June 8, 2021.
https://doi.org/10.1137/20M1341490
Funding: The work of the first author was partially supported by HKRGC grant CityU11301120 and National Science Foundation CCF HDR TRIPODS grant 1934568. The work of the second author was partially supported by National Key Research and Development Program of China grant 2018AAA0101100, Natural Science Foundation of China grant 11971228, and Jiangsu Provincial National Natural Science Foundation of China grant BK20181257. The work of the third author was partially supported by NSF grant DMS-1819042 and NIH grant 5R01CA181171-04. The work of the fourth author was partially supported by NSF CAREER grant 1846690.

†Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX 75080 USA (chaowang.hk@gmail.com, yifei.lou@utdallas.edu).
‡Department of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (taom@nju.edu.cn).
§Mathematics and Computer Science, Emory University, Atlanta, GA 30322 USA (jnagy@emory.edu).

The CT data collection is a nonlinear process due to the polychromatic nature [31, 19] of the x-ray source. A common practice in CT adopts some linearization and discretization schemes that express the formation model as $f = Au$, where f denotes the measurement data, u is the attenuation coefficients to be recovered, and A is a projection matrix. Specifically for this paper, we consider two types of projection geometries: parallel beam and fan beam, which are popular in the CT reconstruction literature. For parallel beam, the complete scanning angle is 180°, while it is 360° for fan beam. If we restrict the maximum scanning angle, it becomes the so-called limited-angle scanning, which is much more challenging than the CT reconstruction from the complete scanning. Some conventional methods in the CT reconstruction include filtered back projection (FBP) [15, 55], simultaneous iterative reconstruction technique (SIRT) [31], and simultaneous algebraic reconstruction technique (SART) [1, 30]. These approaches do not involve any regularization and perform poorly in the case of limited-angle and/or noisy data, resulting in severe streaking artifacts [16, 46].

When data is insufficient, one often requires reasonable assumptions to be imposed as a regularization term in order to reconstruct a desired solution. As such, the CT reconstruction can be formulated as minimizing an objective function that consists of a data fidelity term and a regularization term. There are two commonly studied data fitting terms for CT reconstruction: least squares (LS) [22, 29, 54] and weighted least squares (WLS) [57]. In this paper, we focus on the LS data fitting with a discussion of WLS in section 5.2. As for regularizations, the celebrated total variation (TV) [12, 29, 53, 54, 58, 69] prefers piecewise constant images. However, two noticeable drawbacks for TV are loss of contrast and staircasing artifacts. To resolve its limitations, Jia et al. [28] utilized a tight frame regularization and implemented the algorithm on graphics processing units (GPUs) to achieve fast computation. Recently, a combination of TV and wavelet tight frame was discussed in [41]. The extension of TV in a nonlocal fashion by exploiting patch similarity was examined in [40] for the regular CT reconstruction and in [43] for the limited-angle case.

The TV seminorm is equivalent to the L_1 norm on the gradient. It is well known that L_1 is the tightest convex approximation to the L_0 *norm*,[1] which is used to enforce sparsity for signals of interest. There are several alternatives to approximating the L_0 norm, such as L_p with $0 < p < 1$ [11, 66], transformed L_1 [44, 59, 71, 72], and L_1-L_2 [36, 37, 38, 45, 67]. Algorithmically, Candés, Wakin, and Boyd [6] proposed an iteratively reweighted L_1 (IRL1) algorithm to solve for the L_0 minimization. This idea was reformulated as a scale-space algorithm in [27].

Motivated by recent works using L_1/L_2 [50, 56, 61] for sparse signal recovery, we apply the L_1/L_2 form on the gradient, leading to a new regularization term. This regularization is rather generic in image processing, and we find it works particularly well for piecewise constant images, owing to its scale-invariant property when approximating L_0. In addition, the proposed regularization can mitigate the staircasing artifacts produced by TV, as the L_2 norm of the gradient in the denominator should be away from zero. Extensive experiments demonstrate that our method outperforms the state of the art in CT reconstruction, and significant improvements are achieved for the limited-angle case. L_1/L_2 on the gradient was originally proposed in [50], which included the MRI reconstruction as a proof-of-concept example under

[1]Note that L_0 is not a norm but is often referred to as one.

the noiseless setting, but it lacks practicality and convergence analysis of the algorithm. The contributions of this work are threefold:

1. We propose a novel regularization together with a box constraint for the limited-angle CT reconstruction.
2. We design a specific splitting scheme for solving several related models so that the convergence of ADMM can be established under certain conditions.
3. We present extensive CT reconstruction results (using phantoms/experimental data under parallel/fan beam) to demonstrate the practicality of the proposed approach.

The rest of the paper is organized as follows. In section 2, we present some preliminary materials, such as notation, TV definition, and a previous work of L_1/L_2 [50]. We discuss the proposed models and algorithms in section 3, followed by convergence analysis in section 4. Experimental studies are conducted in section 5 using the projection data of two phantoms subject to two types of noise (Gaussian noise and Poisson noise) as well as two experimental datasets. Finally, conclusions and future work are given in section 6.

2. Preliminaries.

Suppose an underlying image is defined on an $m \times n$ Cartesian grid and denote the Euclidean space $\mathbb{R}^{m\times n}$ as X. We adopt a linear index for the 2D image, i.e., for $u \in X$, $u_{ij} \in \mathbb{R}$ is the $((i-1)m+j)$th component of u. We define a discrete gradient operator,

$$\nabla u := (\nabla_x u, \nabla_y u), \tag{2.1}$$

with ∇_z being the forward difference operator in the z-direction for $z \in \{x, y\}$. Denote $Y = X \times X$. Then $\nabla u \in Y$, and for any $\mathbf{p} \in Y$, its $((i-1)m+j)$th component is $p_{ij} = (p_{ij,1}, p_{ij,2})$. We use a bold letter $\mathbf{p}$ to indicate that it contains two elements in each component. With these notations, we define the inner products by

$$\langle x, y\rangle_X = \sum_{i,j=1}^{m,n} x_{ij}y_{ij} \quad \text{and} \quad \langle \mathbf{p}, \mathbf{q}\rangle_Y = \sum_{i,j=1}^{m,n}\sum_{k=1}^{2} p_{ij,k}q_{ij,k}, \tag{2.2}$$

as well as the corresponding norms

$$\|x\|_2 = \sqrt{\langle x, x\rangle_X} \quad \text{and} \quad \|\mathbf{p}\|_2 = \sqrt{\langle \mathbf{p}, \mathbf{p}\rangle_Y}. \tag{2.3}$$

2.1. Total variation.

By incorporating the TV regularization [51] into the data fitting terms, we can obtain the two models

$$\min_u \|\nabla u\|_1 \quad \text{s.t.} \quad Au = f, \tag{2.4}$$

$$\min_u \|\nabla u\|_1 + \frac{\lambda}{2}\|Au - f\|_2^2, \tag{2.5}$$

where ∇ is defined in (2.1). We refer to (2.4) as a constrained formulation, while (2.5) is an unconstrained one. The latter is often used when the noise is present and the parameter $\lambda > 0$ in (2.5) shall be tuned according to the noise level. Note that the TV term, $\|\nabla u\|_1$, is equivalent to the L_1 norm of the gradient, which can be formulated as the *anisotropic TV*,

$$\|\nabla u\|_1 = \|\nabla_x u\|_1 + \|\nabla_y u\|_1, \tag{2.6}$$

or the *isotropic TV*, defined by $\sum_{i,j=1}^{m,n}\sqrt{(\nabla_x u)_{ij}^2+(\nabla_y u)_{ij}^2}$. The anisotropic TV was shown to be superior to the isotropic one for CT reconstruction [12]. Here, we also adopt the anisotropic TV to define the L_1 norm on the gradient. In addition, the difference of anisotropic and isotropic TV was proposed in [39] for general imaging applications. There are many efficient algorithms to minimize (2.4) or (2.5), including dual projection [7], primal-dual [8], split Bregman [18], and the alternating direction method of multipliers (ADMM) [3].

2.2. L_1/L_2 on the gradient. We review a model of L_1/L_2 on the gradient in a constrained formulation [50],

$$\min_u \frac{\|\nabla u\|_1}{\|\nabla u\|_2} \quad \text{s.t.} \quad Au=f, \tag{2.7}$$

which is referred to as L_1/L_2-con. Here $\|\cdot\|_1$ and $\|\cdot\|_2$ are defined by (2.6) and (2.3), respectively. We apply the ADMM framework [3] to minimize (2.7) by rewriting it into the equivalent form

$$\min_{u,\mathbf{d},\mathbf{h}} \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}\|_2} \quad \text{s.t.} \quad Au=f,\ \mathbf{d}=\nabla u,\ \mathbf{h}=\nabla u, \tag{2.8}$$

with two auxiliary variables $\mathbf{d}$ and $\mathbf{h}$. Note that we denote $\mathbf{d}$ and $\mathbf{h}$ in bold to indicate that they have two components corresponding to x and y derivatives. The augmented Lagrangian for (2.8) is given by

$$\begin{aligned}\mathcal{L}(u,\mathbf{d},\mathbf{h};w,\mathbf{b}_1,\mathbf{b}_2) = \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}\|_2} &+ \langle \lambda w, f-Au\rangle + \frac{\lambda}{2}\|Au-f\|_2^2 \\ &+ \langle \rho_1\mathbf{b}_1, \nabla u-\mathbf{d}\rangle + \frac{\rho_1}{2}\|\mathbf{d}-\nabla u\|_2^2 \\ &+ \langle \rho_2\mathbf{b}_2, \nabla u-\mathbf{h}\rangle + \frac{\rho_2}{2}\|\mathbf{h}-\nabla u\|_2^2,\end{aligned} \tag{2.9}$$

where $w,\mathbf{b}_1,\mathbf{b}_2$ are Lagrange multipliers (or dual variables) and λ,ρ_1,ρ_2 are positive parameters. The ADMM iterations proceed as follows:

$$\begin{cases} u^{(k+1)} = \arg\min_u \mathcal{L}(u,\mathbf{d}^{(k)},\mathbf{h}^{(k)};w^{(k)},\mathbf{b}_1^{(k)},\mathbf{b}_2^{(k)}),\\ \mathbf{d}^{(k+1)} = \arg\min_{\mathbf{d}} \mathcal{L}(u^{(k+1)},\mathbf{d},\mathbf{h}^{(k)};w^{(k)},\mathbf{b}_1^{(k)},\mathbf{b}_2^{(k)}),\\ \mathbf{h}^{(k+1)} = \arg\min_{\mathbf{h}} \mathcal{L}(u^{(k+1)},\mathbf{d}^{(k+1)},\mathbf{h};w^{(k)},\mathbf{b}_1^{(k)},\mathbf{b}_2^{(k)}),\\ w^{(k+1)} = w^{(k)} + f - Au^{(k+1)},\\ \mathbf{b}_1^{(k+1)} = \mathbf{b}_1^{(k)} + \nabla u^{(k+1)} - \mathbf{d}^{(k+1)},\\ \mathbf{b}_2^{(k+1)} = \mathbf{b}_2^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}. \end{cases} \tag{2.10}$$

For more details, please refer to [50] that presented a proof-of-concept example when A^TA and $\nabla^T\nabla$ can be simultaneously diagonalizable by the fast Fourier transform (FFT). In this paper, the matrix A corresponds to a projection matrix, where the inverse of $\lambda A^T A + (\rho_1 + \rho_2)\nabla^T\nabla$ cannot be computed via FFT.

As the splitting scheme (2.8) involves two-block variables of u and $(\mathbf{d}, \mathbf{h})$, it is hard to establish the convergence of (2.10). To prove for the convergence of ADMM, the existing literature [20, 49, 63] requires some associated function (e.g., objective function, merit function, and augmented Lagrangian function) to be coercive, separable, or Lipschitz differentiable (on a certain domain), neither of which holds for the L_1/L_2 functional.

3. The proposed models.

Here we consider an unconstrained formulation of L_1/L_2 in order to deal with noisy data. As opposed to (2.8), we propose a different splitting scheme, under which we can establish the ADMM convergence. We then discuss a variant in section 3.2 to incorporate a box constraint, which is reasonable for the CT reconstruction problems.

3.1. Unconstrained formulation.

The unconstrained L_1/L_2 formulation is given by

$$\min_u \frac{\|\nabla u\|_1}{\|\nabla u\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2, \tag{3.1}$$

which is referred to as L_1/L_2-uncon.

We design a specific splitting scheme that reformulates (3.1) into

$$\min_{u,\mathbf{h}} \frac{\|\nabla u\|_1}{\|\mathbf{h}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 \quad \text{s.t.} \quad \mathbf{h} = \nabla u. \tag{3.2}$$

The corresponding augmented Lagrangian function is expressed as

$$\mathcal{L}_{\text{uncon}}(u, \mathbf{h}; \mathbf{b}_2) = \frac{\|\nabla u\|_1}{\|\mathbf{h}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \langle \rho_2 \mathbf{b}_2, \nabla u - \mathbf{h}\rangle + \frac{\rho_2}{2}\|\mathbf{h} - \nabla u\|_2^2, \tag{3.3}$$

with a dual variable $\mathbf{b}_2$ and a positive parameter ρ_2. The ADMM framework involves the following iterations:

$$\begin{cases} \mathbf{u}^{(k+1)} = \arg\min_u \mathcal{L}_{\text{uncon}}(u, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}), \\ \mathbf{h}^{(k+1)} = \arg\min_{\mathbf{h}} \mathcal{L}_{\text{uncon}}(u^{(k+1)}, \mathbf{h}; \mathbf{b}_2^{(k)}), \\ \mathbf{b}_2^{(k+1)} = \mathbf{b}_2^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}. \end{cases} \tag{3.4}$$

Same as in [50], the $\mathbf{h}$-update has a closed-form solution given by

$$\mathbf{h}^{(k+1)} = \begin{cases} \tau^{(k)}\mathbf{g}^{(k)} & \text{if } \mathbf{g}^{(k)} \neq \mathbf{0}, \\ \mathbf{e}^{(k)} & \text{otherwise}, \end{cases} \tag{3.5}$$

where $\mathbf{g}^{(k)} = \nabla u^{(k+1)} + \mathbf{b}_2^{(k)}$, $\mathbf{e}^{(k)}$ is a random vector with its L_2 norm being $\sqrt[3]{\frac{\|\nabla u^{(k+1)}\|_1}{\rho_2}}$, and $\tau^{(k)} = \frac{1}{3} + \frac{1}{3}(C^{(k)} + \frac{1}{C^{(k)}})$ with

$$C^{(k)} = \sqrt[3]{\frac{27D^{(k)} + 2 + \sqrt{(27D^{(k)} + 2)^2 - 4}}{2}} \quad \text{and} \quad D^{(k)} = \frac{\|\nabla u^{(k+1)}\|_1}{\rho_2\|\mathbf{g}^{(k)}\|_2^3}.$$

The u-subproblem can be expressed as

$$\min_u \frac{\|\nabla u\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \nabla u - \mathbf{b}_2^{(k)}\|_2^2. \tag{3.6}$$

With $\mathbf{h}^{(k)}$ and $\mathbf{b}_2^{(k)}$ fixed, we can apply ADMM to find the optimal solution of (3.6). Specifically by introducing an auxiliary variable $\mathbf{d}$, we rewrite (3.6) as

$$\min_{u,\mathbf{d}} \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \nabla u - \mathbf{b}_2^{(k)}\|_2^2 \quad \text{s.t.} \quad \mathbf{d} = \nabla u. \tag{3.7}$$

The augmented Lagrangian corresponding to (3.7) is given by

$$\begin{aligned}\mathcal{L}_{\text{uncon}}^{(k)}(u, \mathbf{d}; \mathbf{b}_1) = {}& \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \nabla u - \mathbf{b}_2^{(k)}\|_2^2 \\ & + \langle \rho_1 \mathbf{b}_1, \nabla u - \mathbf{d}\rangle + \frac{\rho_1}{2}\|\mathbf{d} - \nabla u\|_2^2,\end{aligned}$$

where $\mathbf{b}_1$ is a dual variable and λ, ρ_1 are positive parameters. Here we have k in the superscript of $\mathcal{L}_{\text{uncon}}$ to indicate that it is the Lagrangian for the u-subproblem in (3.4) at the kth iteration. The ADMM framework to minimize (3.7) leads to the iterations

$$\begin{cases} u_{j+1} = \arg\min_u \mathcal{L}_{\text{uncon}}^{(k)}(u, \mathbf{d}_j; (\mathbf{b}_1)_j), \\ \mathbf{d}_{j+1} = \arg\min_{\mathbf{d}} \mathcal{L}_{\text{uncon}}^{(k)}(u_{j+1}, \mathbf{d}; (\mathbf{b}_1)_j), \\ (\mathbf{b}_1)_{j+1} = (\mathbf{b}_1)_j + \nabla u_{j+1} - \mathbf{d}_{j+1}, \end{cases} \tag{3.8}$$

where the subscript j represents the inner loop index, as opposed to the superscript k for outer iterations in (3.4). Note that $\mathcal{L}_{\text{uncon}}^{(k)}(u, \mathbf{d}; \mathbf{b}_1)$ resembles the augmented Lagrangian $\mathcal{L}(u, \mathbf{d}, \mathbf{h}^{(k)}; w, \mathbf{b}_1, \mathbf{b}_2^{(k)})$ with $w = 0$ defined in (2.9), and hence (3.4) with one iteration of (3.8) for the u-subproblem is equivalent to the previous approach [50]. If we can reach the optimal solution of the u-subproblem, the convergence can be guaranteed; see section 4.

We then elaborate on how to solve the two subproblems in (3.8). By taking the derivative of $\mathcal{L}_{\text{uncon}}^{(k)}$ with respect to u, we obtain a closed-form solution,

$$u_{j+1} = \Big(\lambda A^T A - (\rho_1 + \rho_2)\triangle\Big)^{-1}\Big(\lambda A^T f + \rho_1 \nabla^T(\mathbf{d}_j - (\mathbf{b}_1)_j) + \rho_2 \nabla^T(\mathbf{h}^{(k)} - \mathbf{b}_2^{(k)})\Big), \tag{3.9}$$

where $\triangle = -\nabla^T\nabla$ denotes the Laplacian operator. For a general system matrix A that cannot be diagonalized by the Fourier transform, we adopt the conjugate gradient (CG) descent iterations [48] to solve for (3.9). The $\mathbf{d}$-subproblem in (3.8) has a closed-form solution, i.e.,

$$\mathbf{d}_{j+1} = \mathbf{shrink}\left(\nabla u_{j+1} + (\mathbf{b}_1)_j, \frac{1}{\rho_1\|\mathbf{h}^{(k)}\|_2}\right), \tag{3.10}$$

where $\mathbf{shrink}(\mathbf{v}, \mu) = \text{sign}(\mathbf{v}) \max\{|\mathbf{v}| - \mu, 0\}$.

We summarize in Algorithm 1 for minimizing the L_1/L_2-uncon model (3.1). Admittedly, Algorithm 1 involves 3 levels of iterations: outer/inner ADMM and CG for solving (3.9), which is not computationally appealing. An alternative is the linearized ADMM [47] so as to avoid the CG iterations, which will be explored in the future.

Algorithm 1 The L_1/L_2 unconstrained minimization (L_1/L_2-uncon).

Input: projection matrix A and observed data f
Parameters: $\rho_1, \rho_2, \lambda, \bar{\epsilon} \in \mathbb{R}^+$, and kMax, jMax$\in \mathbb{Z}^+$
Initialize: $\mathbf{h}, \mathbf{b}_1, \mathbf{b}_2, \mathbf{d}$, and $k, j = 0$
while $k <$ kMax or $|u^{(k)} - u^{(k-1)}|/|u^{(k)}| > \bar{\epsilon}$ **do**
  while $j <$ jMax or $|u_j - u_{j-1}|/|u_j| > \bar{\epsilon}$ **do**
    $u_{j+1} = (\lambda A^T A - (\rho_1 + \rho_2)\triangle)^{-1}(\lambda A^T f + \rho_1 \nabla^T(\mathbf{d}_j - (\mathbf{b}_1)_j) + \rho_2 \nabla^T(\mathbf{h}^{(k)} - \mathbf{b}_2^{(k)}))$
    $\mathbf{d}_{j+1} = \mathbf{shrink}\left(\nabla u_{j+1} + (\mathbf{b}_1)_j, \frac{1}{\rho_1 \|\mathbf{h}^{(k)}\|_2}\right)$
    $(\mathbf{b}_1)_{j+1} = (\mathbf{b}_1)_j + \nabla u_{j+1} - \mathbf{d}_{j+1}$
    $j = j + 1$
  end while
  return $u^{(k+1)} = u_j$
  $\mathbf{h}^{(k+1)} = \begin{cases} \tau^{(k)}\left(\nabla u^{(k+1)} + \mathbf{b}_2^{(k)}\right), & \nabla u^{(k+1)} + \mathbf{b}_2^{(k)} \neq 0, \\ \mathbf{e}^{(k)}, & \nabla u^{(k+1)} + \mathbf{b}_2^{(k)} = 0 \end{cases}$
  $\mathbf{b}_2^{(k+1)} = \mathbf{b}_2^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}$
  $k = k + 1$ and $j = 0$
end while
return $\mathbf{u}^* = \mathbf{u}^{(k)}$

3.2. Box constraint.

It is reasonable to incorporate a box constraint for image processing applications [9, 32], since pixel values are usually bounded by $[0, 1]$ or $[0, 255]$. Specifically for CT, the pixel value has physical meanings, and hence the bound can often be estimated in advance [31, 5]. The box constraint is particularly helpful for the L_1/L_2 model to prevent its divergence [61]. We add a general box constraint $u \in [c, d]$ to (3.1), thus leading to

$$\min_u \frac{\|\nabla u\|_1}{\|\nabla u\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 \quad \text{s.t.} \quad u \in [c, d], \tag{3.11}$$

referred to as L_1/L_2-box. To derive an algorithm for solving the L_1/L_2-box model, we rewrite (3.11) equivalently as

$$\min_{u,\mathbf{h}} \frac{\|\nabla u\|_1}{\|\mathbf{h}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \Pi_{[c,d]}(u) \quad \text{s.t.} \quad \mathbf{h} = \nabla u, \tag{3.12}$$

where $\Pi_S(t)$ is an indicator function enforcing t into the feasible set S, i.e.,

$$\Pi_S(t) = \begin{cases} 0 & \text{if } t \in S, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.13}$$

The augmented Lagrangian function for (3.12) can be expressed as

$$\mathcal{L}_{\text{box}}(u, \mathbf{h}; \mathbf{b}_2) = \mathcal{L}_{\text{uncon}}(u, \mathbf{h}; \mathbf{b}_2) + \Pi_{[c,d]}(u). \tag{3.14}$$

By using ADMM, we have the same update rules for $\mathbf{h}$ and $\mathbf{b}_2$ as in (3.4), while the u-subproblem is given by

$$u^{(k+1)} = \arg\min_u \frac{\|\nabla u\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \nabla u - \mathbf{b}_2^{(k)}\|_2^2 + \Pi_{[c,d]}(u). \tag{3.15}$$

We introduce two variables, $\mathbf{d}$ for the gradient and v for the box constraint, thus getting

$$\min_{u,\mathbf{d},v} \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \nabla u - \mathbf{b}_2^{(k)}\|_2^2 + \Pi_{[c,d]}(v) \quad \text{s.t.} \quad \mathbf{d} = \nabla u, u = v. \tag{3.16}$$

The augmented Lagrangian corresponding to (3.16) becomes

$$\begin{aligned} \mathcal{L}_{\text{box}}^{(k)}(u, \mathbf{d}, v; \mathbf{b}_1, e) = & \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\rho_2}{2}\|\nabla u - \mathbf{h}^{(k)} + \mathbf{b}_2^{(k)}\|_2^2 + \Pi_{[c,d]}(v) + \frac{\lambda}{2}\|Au - f\|_2^2 \\ & + \langle \rho_1 \mathbf{b}_1, \nabla u - \mathbf{d}\rangle + \frac{\rho_1}{2}\|\mathbf{d} - \nabla u\|_2^2 + \langle \beta e, u - v\rangle + \frac{\beta}{2}\|v - u\|_2^2, \end{aligned} \tag{3.17}$$

where $\mathbf{b}_1, e$ are dual variables and λ, ρ_1, β are positive parameters. Similar to (3.8), there is a closed-form solution of the u-subproblem,

$$\begin{aligned} u_{j+1} = & \left(\lambda A^T A + (\rho_1 + \rho_2)\triangle + \beta I\right)^{-1} \Big(\lambda A^T f + \rho_1 \nabla^T(\mathbf{d}_j - (\mathbf{b}_1)_j) \\ & + \rho_2 \nabla^T(\mathbf{h}^{(k)} - \mathbf{b}_2^{(k)}) + \beta(v^{(k)} - e^{(k)})\Big). \end{aligned} \tag{3.18}$$

The update for $\mathbf{d}$ is the same as (3.10), and we update v by projecting it onto $[c, d]$, i.e., $v_{j+1} = \min\{\max\{u_{j+1} + e_j, c\}, d\}$. The pseudocode with the additional box constraint is summarized in Algorithm 2.

4. Convergence analysis.

We intend to establish the convergence of Algorithms 1–2. We observe that the ADMM framework for both models shares the same structure,

$$\begin{cases} u^{(k+1)} = \arg\min_u \mathcal{L}(u, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}), \\ \mathbf{h}^{(k+1)} = \arg\min_{\mathbf{h}} \mathcal{L}(u^{(k+1)}, \mathbf{h}; \mathbf{b}_2^{(k)}), \\ \mathbf{b}_2^{(k+1)} = \mathbf{b}_2^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}, \end{cases} \tag{4.1}$$

where $\mathcal{L}$ is either $\mathcal{L}_{\text{uncon}}$ or $\mathcal{L}_{\text{box}}$. We show that the sequence generated by ADMM for L_1/L_2-uncon either diverges due to unboundedness or has a convergent subsequence, while the sequence for L_1/L_2-box always has a convergent subsequence. For this purpose, we introduce Lemma 4.2 for an upper bound of $\|\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}\|_2$ in terms of $\|u^{(k+1)} - u^{(k)}\|_2$ and $\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2$. Lemmas 4.3 and 4.4 are standard in convergence analysis [26, 33, 62, 63] to guarantee that the augmented Lagrangian decreases sufficiently and the subgradient at each iteration is bounded by successive errors, respectively. The lemmas require the following three assumptions:

A1 : $\mathcal{N}(\nabla) \bigcap \mathcal{N}(A) = \{0\}$, where $\mathcal{N}$ denotes the null space and ∇ is defined in (2.1).

Algorithm 2 The L_1/L_2 minimization with a box constraint (L_1/L_2-box).

Input: projection matrix A, observed data f, and a bound $[c, d]$ for the original image
Parameters: $\rho_1, \rho_2, \lambda, \beta, \bar{\epsilon} \in \mathbb{R}^+$, and kMax, jMax $\in \mathbb{Z}^+$
Initialize: $\mathbf{h}, \mathbf{b}_1, \mathbf{b}_2, \mathbf{d}, w = 0, e$, and $k, j = 0$
while $k <$ kMax or $|u^{(k)} - u^{(k-1)}|/|u^{(k)}| > \bar{\epsilon}$ **do**
  while $j <$ jMax or $|u_j - u_{j-1}|/|u_j| > \bar{\epsilon}$ **do**
    $u_{j+1} = (\lambda A^T A - (\rho_1 + \rho_2)\triangle + \beta I)^{-1}(\lambda A^T f + \rho_1 \nabla^T(\mathbf{d}_j - (\mathbf{b}_1)_j) + \rho_2 \nabla^T(\mathbf{h}^{(k)} - \mathbf{b}_2^{(k)}) + \beta(v^{(k)} - e^{(k)}))$
    $\mathbf{d}_{j+1} = \mathbf{shrink}\left(\nabla u_{j+1} + (\mathbf{b}_1)_j, \frac{1}{\rho_1 \|\mathbf{h}^{(k)}\|_2}\right)$
    $v_{j+1} = \min\{\max\{u_{j+1} + e_j, c\}, d\}$
    $(\mathbf{b}_1)_{j+1} = (\mathbf{b}_1)_j + \nabla u_{j+1} - \mathbf{d}_{j+1}$
    $e_{j+1} = e_j + u_{j+1} - v_{j+1}$
    $j = j + 1$
  end while
  return $u^{(k+1)} = u_j$
  $\mathbf{h}^{(k+1)} = \begin{cases} \tau^{(k)}\left(\nabla u^{(k+1)} + \mathbf{b}_2^{(k)}\right), & \nabla u^{(k+1)} + \mathbf{b}_2^{(k)} \neq 0, \\ \mathbf{e}^{(k)}, & \nabla u^{(k+1)} + \mathbf{b}_2^{(k)} = 0 \end{cases}$
  $\mathbf{b}_2^{(k+1)} = \mathbf{b}_2^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}$
  $k = k + 1$ and $j = 0$
end while
return $\mathbf{u}^* = \mathbf{u}^{(k)}$

A2 : The sequence $\{u^{(k)}\}$ generated by (4.1) is bounded, and then so is $\{\nabla u^{(k)}\}$ and we denote $M = \sup_k\{\|\nabla u^{(k)}\|_1\}$.

A3 : The norm of $\{\mathbf{h}^{(k)}\}$ generated by (4.1) has a lower bound, i.e., there exists a positive constant ϵ such that $\|\mathbf{h}^{(k)}\|_2 \geq \epsilon \ \forall k$.

Remark 4.1. Assumption A1 is standard in image processing [10, 39]. Assumption A2 requires the boundedness of $\{u^{(k)}\}$, and hence the convergence results can be interpreted as the sequence either diverges (due to unboundedness) or converges to a critical point. To make the L_1/L_2 regularization well-defined, we shall have $\|\mathbf{h}\|_2 > 0$. Certainly, $\|\mathbf{h}\|_2 > 0$ does not imply a uniform lower bound of ϵ, but we can redefine the divergence of an algorithm by including the case of $\|\mathbf{h}^{(k)}\|_2 < \epsilon$, which can be checked numerically with a preset value of ϵ.

Please refer to the appendix for the proofs of these lemmas, based on which we can establish the convergence in Theorems 4.5 and 4.6 for Algorithms 1 and 2, respectively. Furthermore, Theorems 4.7 and 4.8 extend the convergence analysis to the case when the u-subproblem in (4.1) can be solved inexactly.

Lemma 4.2. *Under assumptions* A1 *and* A2*, the sequence* $\{u^{(k)}, \mathbf{h}^{(k)}, \mathbf{b}_2^{(k)}\}$ *generated by* (4.1) *satisfies*

$$\left\|\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}\right\|_2^2 \leq \left(\frac{32mn}{\rho_2^2 \epsilon^4}\right) \left\|u^{(k+1)} - u^{(k)}\right\|_2^2 + \left(\frac{8M^2}{\rho_2^2 \epsilon^6}\right) \left\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\right\|_2^2. \tag{4.2}$$

Lemma 4.3 (sufficient descent). *Under assumptions* A1–A3 *and a sufficiently large* ρ_2*, the sequence* $\{u^{(k)}, \mathbf{h}^{(k)}, \mathbf{b}_2^{(k)}\}$ *generated by* (4.1) *satisfies*

$$\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}; \mathbf{b}_2^{(k+1)}) \le \mathcal{L}(u^{(k)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) - c_1\|u^{(k+1)} - u^{(k)}\|_2^2 - c_2\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2, \tag{4.3}$$

where c_1 *and* c_2 *are two positive constants.*

Lemma 4.4 (subgradient bound). *Under assumptions* A1–A3 *and a sufficiently large* ρ_2*, there exists a vector* $\boldsymbol{\eta}^{(k+1)} \in \partial\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}; \mathbf{b}_2^{(k+1)})$ *and a constant* $\gamma > 0$ *such that*

$$\|\boldsymbol{\eta}^{(k+1)}\|_2^2 \le \gamma\left(\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2 + \|\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}\|_2^2\right). \tag{4.4}$$

Theorem 4.5 (convergence of L_1/L_2-uncon). *Under assumptions* A1–A3 *and a sufficiently large* ρ_2*, the sequence* $\{u^{(k)}, \mathbf{h}^{(k)}\}$ *generated by* (3.4) *has a subsequence convergent to a critical point of* (3.2)*.*

Proof. We first show that if $\{u^{(k)}\}$ is bounded, then $\{\mathbf{h}^{(k)}, \mathbf{b}_2^{(k)}\}$ is also bounded. As $\|u^{(k)}\|_2$ is bounded, so is $\|\nabla u^{(k)}\|_1$. It follows from assumption A2 and the optimality condition for $\mathbf{b}_2$ in (A.3) that we have

$$\|\mathbf{b}_2^{(k)}\|_2 = \left\| \frac{\|\nabla u^{(k)}\|_1}{\rho_2} \frac{\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|^3} \right\|_2 \le \frac{\|\nabla u^{(k)}\|_1}{\rho_2 \epsilon^2}.$$

Therefore, $\{\mathbf{b}_2^{(k)}\}$ is bounded and hence $\{\mathbf{h}^{(k)}\}$ is also bounded due to the $\mathbf{h}$-update (3.5) and boundedness of ∇u. Then it follows from the Bolzano–Weierstrass theorem that the sequence $\{u^{(k)}, \mathbf{h}^{(k)}, \mathbf{b}_2^{(k)}\}$ has a convergent subsequence, denoted by $(u^{(k_j)}, \mathbf{h}^{(k_j)}, \mathbf{b}_2^{(k_j)}) \to (u^*, \mathbf{h}^*, \mathbf{b}_2^*)$, as $k_j \to \infty$. In addition, we can estimate that

$$\begin{aligned} &\mathcal{L}_{\text{uncon}}(u^{(k)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) \\ &= \frac{\|\nabla u^{(k)}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \nabla u^{(k)} - \mathbf{b}_2\|_2^2 - \frac{\rho_2}{2}\|\mathbf{b}_2^{(k)}\|_2^2 \\ &\ge \frac{\|\nabla u^{(k)}\|_1}{\|\mathbf{h}^{(k)}\|_2} - \frac{\|\nabla u^{(k)}\|_1^2}{\rho_2 \epsilon^4}, \end{aligned}$$

which gives a lower bound of $\mathcal{L}_{\text{uncon}}$ owing to the boundedness of $u^{(k)}$. Therefore, we have that $\mathcal{L}_{\text{uncon}}(u^{(k)}, \mathbf{h}^{(k)}, \mathbf{b}_2^{(k)})$ converges due to its monotonic decreasing by Lemma 4.3.

We then sum inequality (4.3) from $k = 0$ to K, thus getting

$$\begin{aligned} &\mathcal{L}_{\text{uncon}}(u^{(K+1)}, \mathbf{h}^{(K+1)}; \mathbf{b}_2^{(K+1)}) \\ &\le \mathcal{L}_{\text{uncon}}(u^{(0)}, \mathbf{h}^{(0)}; \mathbf{b}_2^{(0)}) - c_1 \sum_{k=0}^{K} \|u^{(k+1)} - u^{(k)}\|_2^2 - c_2 \sum_{k=0}^{K} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2. \end{aligned}$$

Letting $K \to \infty$, we have that both $\sum_{k=0}^{\infty} \|u^{(k+1)} - u^{(k)}\|_2^2$ and $\sum_{k=0}^{\infty} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2$ are finite, indicating that $u^{(k)} - u^{(k+1)} \to 0$, $\mathbf{h}^{(k)} - \mathbf{h}^{(k+1)} \to 0$. Then by Lemma 4.2, we get

$\mathbf{b}_2^{(k)} - \mathbf{b}_2^{(k+1)} \to 0$. By $(u^{(k_j)}, \mathbf{h}^{(k_j)}, \mathbf{b}_2^{(k_j)}) \to (u^*, \mathbf{h}^*, \mathbf{b}_2^*)$, we have $(u^{(k_j+1)}, \mathbf{h}^{(k_j+1)}, \mathbf{b}_2^{(k_j+1)}) \to (u^*, \mathbf{h}^*, \mathbf{b}_2^*)$, and $\nabla u^* = \mathbf{h}^*$ (by the update of $\mathbf{b}_2$). Here, by Lemma 4.4, we have $\mathbf{0} \in \partial \mathcal{L}_{\text{uncon}}(u^*, \mathbf{h}^*, \mathbf{b}_2^*)$, and hence $(u^*, \mathbf{h}^*)$ is a critical point of (3.2). ∎

For the box model (3.11) with an explicit bounded assumption on u, we can prove that the ADMM framework has the same convergence results as in Theorem 4.5 without assumption A2. The proof is thus omitted.

Theorem 4.6 (convergence of L_1/L_2-box). *Under assumptions* A1 *and* A3 *and a sufficiently large ρ_2, the sequence $\{u^{(k)}, \mathbf{h}^{(k)}\}$ generated by Algorithm 2 always has a subsequence convergent to a critical point of* (3.11).

Theorem 4.7 (convergence of inexact scheme in L_1/L_2-uncon). *Under assumption* A1 *and a sufficiently large ρ_2, one can solve the u-subproblem in* (3.6) *within an error tolerance ε_{k+1}, i.e.,*

$$\|\tilde{u}^{(k+1)} - u^{(k+1)}\|_2^2 \le \varepsilon_{k+1}, \tag{4.5}$$

and the sequence $\{\tilde{\mathbf{h}}^{(k+1)}, \tilde{\mathbf{b}}_2^{(k+1)}\}$ is generated by an inexact ADMM scheme, i.e.,

$$\begin{cases} \tilde{\mathbf{h}}^{(k+1)} = \arg\min_{\mathbf{h}} \mathcal{L}_{\text{uncon}}(\tilde{u}^{(k+1)}, \mathbf{h}; \tilde{\mathbf{b}}_2^{(k)}), \\ \tilde{\mathbf{b}}_2^{(k+1)} = \tilde{\mathbf{b}}_2^{(k)} + \nabla \tilde{u}^{(k+1)} - \tilde{\mathbf{h}}^{(k+1)}. \end{cases} \tag{4.6}$$

If $\{\tilde{u}^{(k)}, \tilde{\mathbf{h}}^{(k)}\}$ satisfies assumptions A2–A3 *and $\sum_k \varepsilon_k < +\infty$, then this sequence has a subsequence convergent to a critical point of* (3.2).

Proof. As $\tilde{u}^{(k)}$ is bounded, we denote $\tilde{M} := \sup_k \|\nabla \tilde{u}^k\|_1$ and $\hat{M} := \sup_k\{\|\tilde{u}^{(k)}\|_2\}$. Following the proof of Lemma 4.3, we have

$$\mathcal{L}_{\text{uncon}}(u^{(k+1)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) \le \mathcal{L}_{\text{uncon}}(\tilde{u}^{(k)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) - \frac{\sigma\lambda}{2}\|u^{(k+1)} - \tilde{u}^{(k)}\|_2^2,$$

$$\mathcal{L}_{\text{uncon}}(\tilde{u}^{(k+1)}, \tilde{\mathbf{h}}^{(k+1)}; \tilde{\mathbf{b}}_2^{(k)}) \le \mathcal{L}_{\text{uncon}}(\tilde{u}^{(k+1)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) - \frac{\rho_2 - 3\tilde{L}}{2}\|\tilde{\mathbf{h}}^{(k+1)} - \tilde{\mathbf{h}}^{(k)}\|_2^2,$$

where $\tilde{L} = \frac{2\tilde{M}}{\epsilon^3}$. Simple calculations lead to

$$\begin{aligned} \|u^{(k+1)} - \tilde{u}^{(k)}\|_2^2 &= \|(u^{(k+1)} - \tilde{u}^{(k+1)}) + (\tilde{u}^{(k+1)} - \tilde{u}^{(k)})\|_2^2 \\ &\ge \|u^{(k+1)} - \tilde{u}^{(k+1)}\|^2 + \|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\|^2 - 2\|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\|\|u^{(k+1)} - \tilde{u}^{(k+1)}\| \\ &\ge \|u^{(k+1)} - \tilde{u}^{(k+1)}\|^2 + \|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\|^2 - 4\hat{M}\varepsilon_{k+1} \\ &\ge \|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\|^2 - 4\hat{M}\varepsilon_{k+1}. \end{aligned}$$

It follows from Lemma A.2 that

$$\begin{aligned} \mathcal{L}_{\text{uncon}}(\tilde{u}^{(k+1)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) &\le \tilde{\mathcal{L}}_{\text{uncon}}(u^{(k+1)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) + \frac{\tilde{L}}{2}\|\tilde{u}^{(k+1)} - u^{(k+1)}\|_2^2 \\ &\le \mathcal{L}_{\text{uncon}}(u^{(k+1)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) + \frac{\tilde{L}\varepsilon_{k+1}}{2}. \end{aligned}$$

Analogous to Lemma 4.2, it holds that

$$\left\|\tilde{\mathbf{b}}_2^{(k+1)} - \tilde{\mathbf{b}}_2^{(k)}\right\|_2^2 \le \kappa_1 \left\|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\right\|_2^2 + \kappa_2 \left\|\tilde{\mathbf{h}}^{(k+1)} - \tilde{\mathbf{h}}^{(k)}\right\|_2^2,$$

where $\kappa_1 := \frac{32mn}{\rho_2^2\epsilon^4}$ and $\kappa_2 := \frac{8\tilde{M}^2}{\rho_2^2\epsilon^6}$. By combining all the inequalities, we get

$$\begin{aligned} &\mathcal{L}_{\text{uncon}}(\tilde{u}^{(k+1)}, \tilde{\mathbf{h}}^{(k+1)}; \tilde{\mathbf{b}}_2^{(k+1)}) \\ &\le \mathcal{L}_{\text{uncon}}(\tilde{u}^{(k)}, \tilde{\mathbf{h}}^{(k)}; \tilde{\mathbf{b}}_2^{(k)}) - \tilde{c}_1\|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\|_2^2 - \tilde{c}_2\|\tilde{\mathbf{h}}^{(k+1)} - \tilde{\mathbf{h}}^{(k)}\|_2^2 + \tilde{c}_3\varepsilon_{k+1}, \end{aligned} \tag{4.7}$$

where $\tilde{c}_1 := \frac{\sigma\lambda}{2} - \frac{16mn}{\rho_2\epsilon^4}$, $\tilde{c}_2 := \frac{\rho_2 - 3L}{2} - \frac{4\tilde{M}^2}{\rho_2\epsilon^6}$, and $\tilde{c}_3 := 2\hat{M}\sigma\lambda + \frac{\tilde{L}}{2}$. We can choose a sufficiently large ρ_2 such that $\tilde{c}_1, \tilde{c}_2 > 0$. Summing inequality (4.7) with k from 0 to K and letting $K \to \infty$, we obtain that $\sum_{k=0}^{\infty} \|\tilde{u}^{(k+1)} - \tilde{u}^{(k)}\|_2^2$ and $\sum_{k=0}^{\infty} \|\tilde{\mathbf{h}}^{(k+1)} - \tilde{\mathbf{h}}^{(k)}\|_2^2$ are finite, since $\tilde{c}_3 \sum_k \varepsilon_k < +\infty$ by assumption. The rest of the proof follows along the same lines as that of Theorem 4.5 and is thus omitted. ∎

Similarly, we have the convergence of inexact scheme in L_1/L_2-box under a restriction that $\tilde{u}^{(k+1)}$ should belong to the feasible set, i.e., $\tilde{u}^{(k+1)} \in [c, d]$.

Theorem 4.8 (convergence of inexact scheme in L_1/L_2-box). *Under assumption* A1 *and a sufficiently large* ρ_2*, one can solve the* u*-subproblem in* (3.15) *within an error tolerance* ε_k *and feasible set, i.e.,*

$$\|\tilde{u}^{(k+1)} - u^{(k+1)}\|_2^2 \le \varepsilon_{k+1} \quad \text{and} \quad \tilde{u}^{(k+1)} \in [c, d]. \tag{4.8}$$

If $\sum_k \varepsilon_k < +\infty$ *and* $\tilde{\mathbf{h}}^{(k)}$ *satisfies assumption* A3*, then the sequence* $\{\tilde{u}^{(k)}, \tilde{\mathbf{h}}^{(k)}\}$ *has a subsequence convergent to a critical point of* (3.11).

Remark 4.9. Theorems 4.5–4.8 are about subsequential convergence, which is weaker than global convergence, i.e., the entire sequence converges. If the augmented Lagrangian $\mathcal{L}$ has the Kurdyka–Łojasiewicz (KL) property [2], global convergence can be shown in a similar way as [20, Theorem 3.1]. Unfortunately, the KL property is an open problem for the L_1/L_2 functional. On the other hand, it is true that Theorems 4.7 and 4.8 relax the accuracy of solving the u-subproblem within the tolerance ε_k at the kth iteration, but in practice we solve for a fixed number of iterations, under which the convergence remains open in the optimization literature.

5. Experimental results. We carry out extensive experiments to demonstrate the performance of the proposed approaches in comparison to the state of the art. We test on two phantoms of Shepp–Logan (SL) by the MATLAB command `phantom` and FORBILD (FB) [70] as well as two experimental datasets of a walnut [21] and a lotus [4], all shown in Figure 1. Note that the MATLAB SL phantom has a higher contrast than the original one presented in [52], as illustrated in Figure 2 by the horizontal and vertical profiles. The reference images for the experimental data are reconstructed from a complete scanning via the Tikhonov regularization (MATLAB function is provided in [4, 21]). As the FB phantom has a very low image contrast, we display it with the grayscale window of [1.03, 1.10] in order to reveal its structures. Using the SL phantom, we discuss some computational aspects of the proposed

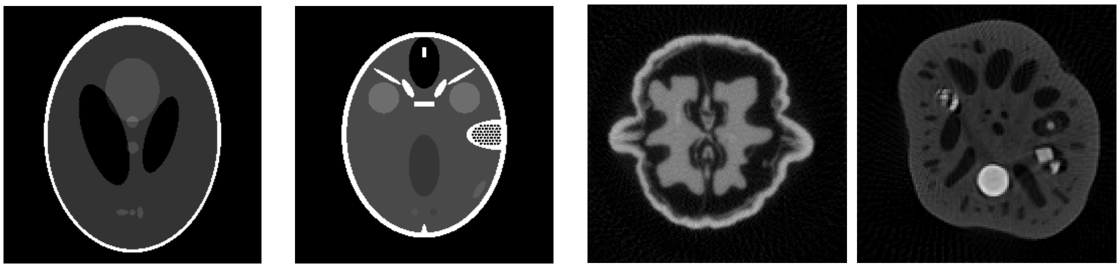

Figure 1. *Ground truth of Shepp–Logan (SL) phantom and FORBILD (FB) head phantom with the grayscale window of* $[0,\ 1]$ *and* $[1.03,\ 1.10]$*, respectively. The last two are reference images of a walnut and a lotus reconstructed by using the complete projection data with the grayscale window of* $[0,\ 0.6]$*.*

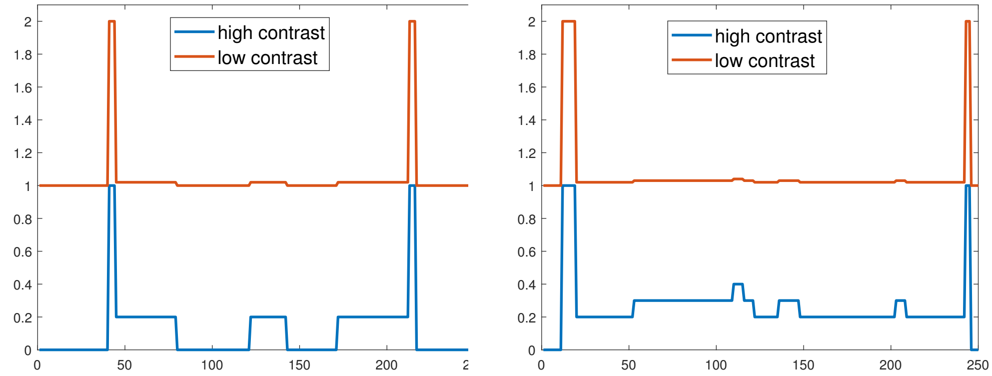

Figure 2. *Horizontal (left) and vertical (right) profiles of two SL phantoms with low contrast and high contrast, the latter of which is used in this paper.*

algorithms in subsection 5.1. We then present numerical results on synthetic data (SL and FB) in subsection 5.2 and experimental data (walnut and lotus) in subsection 5.3. All the numerical experiments are conducted on a desktop with CPU (Intel i7-5930K, 3.50 GHz) and MATLAB 9.7 (R2019b).

To synthesize the limited-angle CT projection data, we discretize both SL and FB phantoms at a resolution of 256×256. The forward operator A is generated as the discrete Radon transform with the same resolution as the digital phantoms. We use the IR and AIR toolbox [17, 23] to simulate *parallel beam* and *fan beam* for the CT scanning. Both settings are sampled at $\theta_{\text{Max}}/30$ over a range of θ_{Max}, resulting in a subsampled data of size 362×31. We use the same number of projections when we vary ranges of projection angles. Note that complete scanning ranges for parallel beam and fan beam are $180°$ and $360°$, respectively. Therefore, fan beam is more challenging to reconstruct than parallel beam with the same value of θ_{Max}. We then add either Gaussian noise or Poisson noise to the projected data. The Gaussian noise follows the zero mean Gaussian distribution with a standard deviation set by a noise level multiplying the maximum intensity of the projected data. We consider two Gaussian noise levels: 0.5% and 0.1%. A more realistic noise distribution for CT data is that the data has the Poisson distribution with the mean $I_0 \exp(-f)$, where I_0 denotes the number of incident x-ray photons and f is the noise-free sinogram. We consider two Poisson noise levels: $I_0 = 10^4$ and 10^5. The larger the value of I_0 is, the higher the signal-to-noise ratio is for the measured data.

We evaluate the performance in terms of the root mean squared error (RMSE) and the overall structural similarity index (SSIM) [64]. RMSE is defined as

$$\mathrm{RMSE}(u^*, \tilde{u}) := \frac{\|u^* - \tilde{u}\|_2}{N_{\text{pixel}}},$$

where u^* is the restored image, $\tilde{u}$ is the ground truth, and N_{pixel} is the total number of pixels. SSIM is the mean of local similarity indices,

$$\mathrm{SSIM}(u^*, \tilde{u}) := \frac{1}{N}\sum_{i=1}^{N} \mathrm{ssim}(x_i, y_i),$$

where x_i, y_i correspond to the ith 8×8 windows for u^* and $\tilde{u}$, respectively, and N is the number of such windows. Note that $N \neq N_{pixel}$ if we do not consider zero-padded pixels along the edges. The local similarity index is defined as

$$\mathrm{ssim}(x, y) := \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

where the averages/variances of x, y are denoted as μ_x/σ_x^2 and μ_y/σ_y^2, respectively. Here, c_1 and c_2 are two fixed constants to stabilize the division with weak denominator, which are set to be $c_1 = c_2 = 0.05$.

We compare the proposed L_1/L_2 model with a clinical standard approach of SART [1], the TV model (2.4), referred to as L_1, as well as two nonconvex regularizations: L_p for $p = 0.5$ and L_1-L_2 [39] on the gradient. To solve for the L_p model, we replace the soft shrinkage by the proximal operator corresponding to L_p, derived in [66], and apply the same ADMM framework as the L_1 minimization. As for L_1-L_2, we modify the MATLAB package provided by the authors [39] for the CT reconstruction problem. All these regularization methods are solved in a constrained formulation with the same box constraint to make a fair comparison. We pose the box constraints: $[0, 1]$ for SL and $[0, 1.8]$ for FB, since we know the upper/lower bounds of the ground-truth images. As for the experimental images, we set the box constraint as $[0, 0.5]$, which is estimated from the reference images. The initial condition of u is chosen to be a zero vector for all the methods. We set the maximum iteration in the inner loop and outer loop for both L_1/L_2 and L_1-L_2 as 5 and 300, respectively, while the maximum iteration of L_1 and L_p is 500. The (outer) stopping criterion is $\frac{\|u^{(k)} - u^{(k-1)}\|_2}{\|u^{(k)}\|_2} \leq 10^{-5}$. As for the other parameters in L_1/L_2, we set $\rho_1 = \rho_2 = \rho$ and find the optimal combination among the candidate set of $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$ and $\rho, \beta \in \{0.1, 1, 10\}$ that gives the lowest RMSE. We tune parameters at each noise level for every testing dataset. In a similar way, we tune the parameters individually for L_1, L_p, and L_1-L_2.

5.1. Algorithm behavior. In this section, we discuss computational aspects of the proposed algorithms. We first analyze the influence of the box constraint on the reconstruction results. The analysis is based on the SL phantom from parallel beam CT projection data

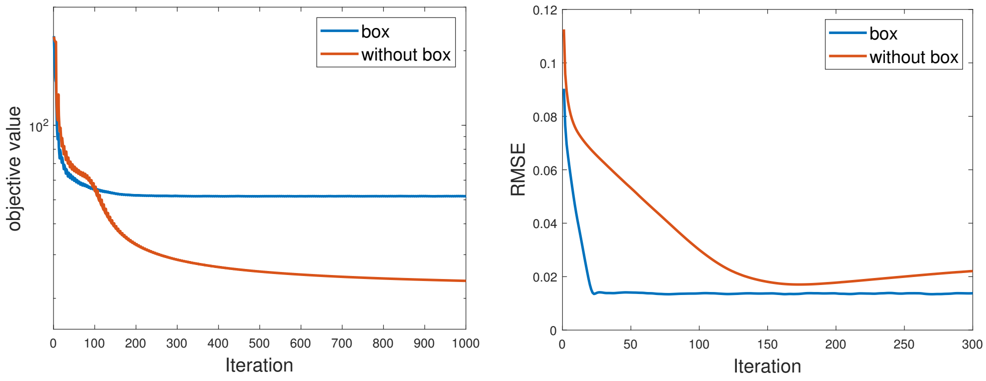

Figure 3. *The effects of the box constraint in terms of the objective value (left) and RMSE (right).*

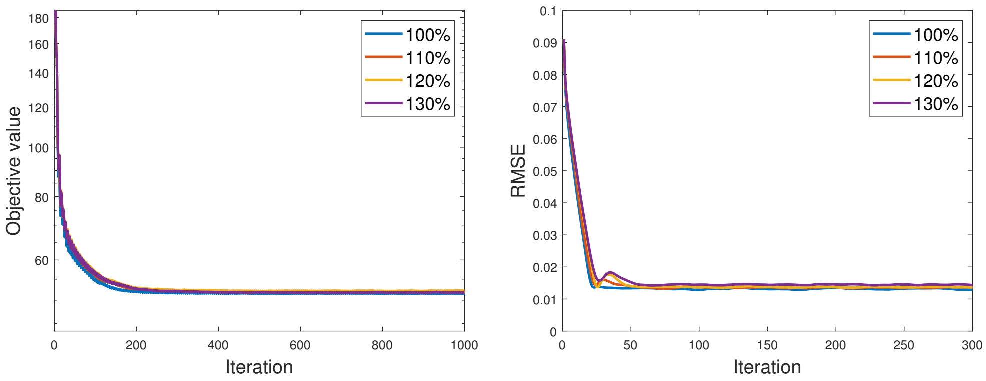

Figure 4. *The effects of the upper bound of the box constraint in terms of the objective value (left) and RMSE (right).*

with the scanning range of 135° subject to Gaussian noise of 0.5%. The fidelity of the CT reconstruction and the convergence are assessed in terms of objective values and $\mathrm{RMSE}(u^{(k)}, \tilde{u})$ versus outer iteration counter k. In Figure 3, we present algorithmic behaviors of the box constraint on the unconstrained model. Here we set jMax to be 5 (we will discuss the effects of inner iteration number shortly.) We plot both inner and outer iterations in Figure 3, showing that the proposed algorithms with and without the box constraint are convergent, as the objective functions decrease. On the other hand, the box constraint yields smaller RMSE compared to the one without the box. Moreover, the box constraint helps avoid local minimizers, as the RMSE of the algorithm without the box increases and the objective function keeps going down. Therefore, the box constraint plays an important role in the success of our approach for the CT reconstruction.

We then discuss the effect of upper bound of the box constraint, i.e., d, on the CT reconstruction performance. Again, we consider the SL phantom from paralleled beam CT projection with the scanning range of 135° subject to a noise level of 0.5%. We compare the oracle upper bound (100%) with relaxed bounds (110%, 120%, and 130%). In Figure 4, we plot the objective values and RMSE with respect to iteration numbers. All the curves of

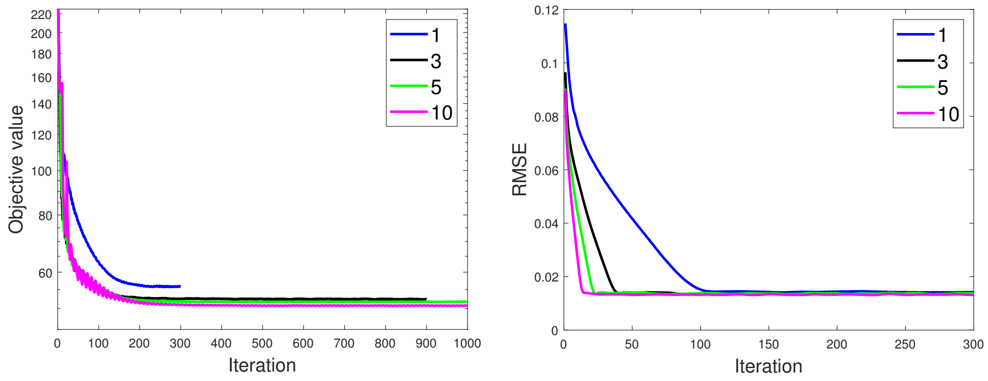

Figure 5. *The effects of the maximum number of the inner loops in terms of the objective value (left) and RMSE (right).*

objective values in different d are almost the same, while the RMSE shows the accuracy is slightly different. Figure 4 demonstrates that the proposed method is insensitive to the upper bound of the box constraint.

Finally, we discuss the influence of jMax on the sparse recovery performance. Fixing the maximum outer iterations at 300, we examine the results of jMax$= 1, 3, 5,$ and 10. In Figure 5, we plot the objective values and RMSE with respect to iterations (counting both inner and outer loops). The objective function with only one inner iteration does not decrease as much as the ones with more inner iterations. RMSE reaches a lower value by fewer outer iterations when using larger jMax. Following Figure 5, we set jMax to be 5 throughout the experiments.

5.2. Synthetic dataset. We start with the parallel beam CT reconstruction of the SL phantom from 90° and 150° projection ranges, labeled SL-90°/SL-150°, with 0.5% Gaussian noise. The quantitative results in terms of SSIM and RMSE are reported in Table 1. As illustrated in Figure 6, SART fails to recover the ellipse shape of the skull with such small ranges of projection angles. Both L_1 and L_1-L_2 models are unable to restore the bottom skull and preserve details of some ellipses in the middle. The L_p model leads to a nearly perfect reconstruction of the skull, but containing a lot of salt-and-pepper artifacts inside the brain. The proposed L_1/L_2 method yields a reasonable recovery in a balanced manner. In the case of SL-150°, L_p is superior over the other approaches, while the proposed method is the second best. This outcome is consistent with the compressed sensing literature [68] that L_p performs quite well for the incoherent problem, which corresponds to a larger scanning angle in the CT reconstruction. The performance of L_p decays for narrow scanning ranges, as reported in Table 1.

We present the visual results of FB-90° and FB-150° with 0.1% Gaussian noise in Figure 7. None of the methods can get satisfactory recovery results under the grayscale window of $[1.03, 1.10]$. Large fluctuations inside of the skull are produced by the competing methods, among which L_1/L_2 can restore the most details of the image. Furthermore, we plot the horizontal and vertical profiles in Figure 8, which illustrates that L_1/L_2 leads to the smallest fluctuations compared to others. In contrast to the simple SL phantom, L_p does not work well for FB. We also observe a well-known artifact of the L_1 method, i.e., loss of contrast, as

Table 1
Parallel beam CT reconstruction of the SL phantom by SART, L_1, L_p, L_1-L_2, and L_1/L_2.

Noise	Range	SART		L_1		L_p		L_1-L_2		L_1/L_2	
		SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE
0.5%	90°	0.56	0.138	0.88	0.075	0.91	0.029	0.78	0.087	**0.96**	**0.017**
	150°	0.58	0.106	0.98	0.038	**0.99**	**0.008**	0.88	0.034	0.98	0.011
0.1%	90°	0.58	0.137	0.96	0.041	**1.00**	0.006	0.88	0.072	**1.00**	**0.003**
	150°	0.60	0.104	0.98	0.035	**1.00**	0.005	0.99	0.076	**1.00**	**0.001**

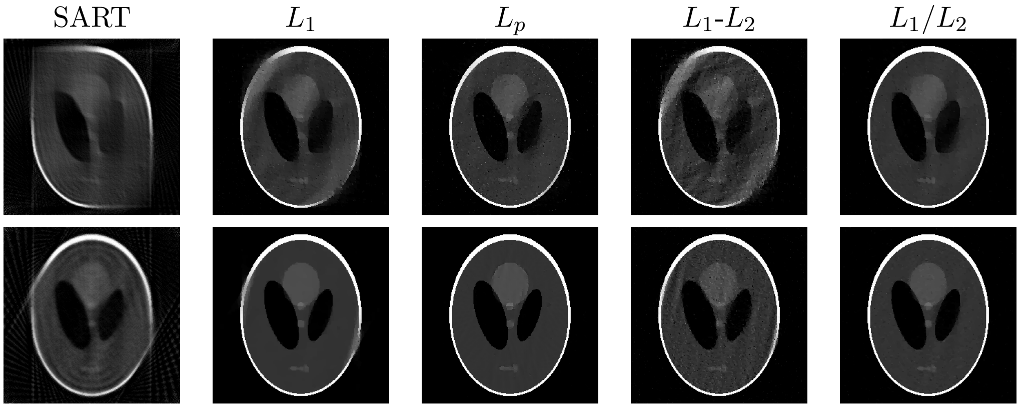

Figure 6. *CT reconstruction from* 90° *(top) and* 150° *(bottom) parallel beam projection for the SL phantom with* 0.5*% noise. The grayscale window is* $[0, 1]$.

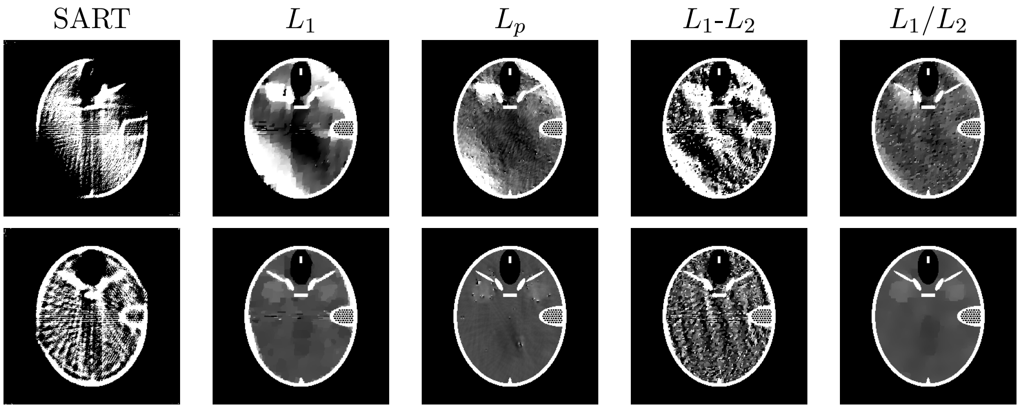

Figure 7. *CT reconstruction from* 90° *(top) and* 150° *(bottom) parallel beam projection for the FB phantom with* 0.1*% Gaussian noise. The grayscale window is* $[1.03, 1.10]$.

its profile fails to reach the height of jump on intervals such as $[160, 180]$ in the left plot and $[220, 230]$ in the right plot of Figure 8, while L_1/L_2 makes a good recovery in these regions. As shown in Figure 7, errors in these low contrast regions are magnified when we display the restored image in a narrow grayscale window. Furthermore, the profile plots in Figure 8 confirm that our approach performs very well for high contrast details [60]. We report the

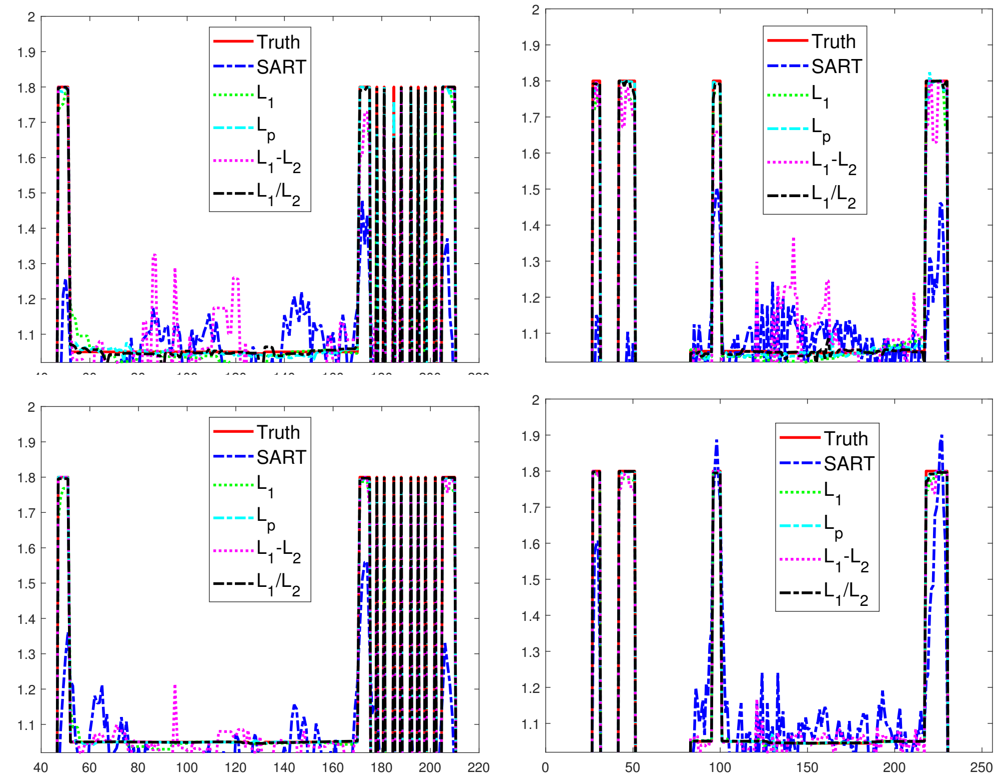

Figure 8. *Horizontal and vertical profiles generated via SART, L_1, L_p, L_1-L_2, and L_1/L_2 in the range of projection 90° (top) and 150° (bottom) for the FB phantom.*

Table 2

Parallel beam CT reconstruction of the FB phantom by SART, L_1, L_p, L_1-L_2, and L_1/L_2.

Noise	Range	SART		L_1		L_p		L_1-L_2		L_1/L_2	
		SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE
0.5%	90°	0.26	0.275	0.82	0.135	0.77	0.101	0.65	0.169	**0.91**	**0.080**
	150°	0.28	0.206	0.90	0.059	0.70	0.078	0.70	0.107	**0.95**	**0.028**
0.1%	90°	0.30	0.266	0.93	0.101	0.97	0.049	0.78	0.123	**0.99**	**0.012**
	150°	0.32	0.192	0.99	0.026	**1.00**	0.003	0.94	0.026	**1.00**	**0.002**

quantitative results of FB in Table 2. Comparing Tables 1 and 2 shows that all the methods yield better performance for smaller noise levels and a larger range of scanning angles. In addition, the recovery results of FB are much worse than those of SL, which are largely due to low contrast structures in FB.

We then test fan beam CT reconstruction using the SL phantom with 0.5% Gaussian noise. Note that fan beam with the same scanning angle is more ill-posed than in the cases of parallel beam. Figure 9 illustrates that the ellipse shape of skull cannot be completely recovered except for the proposed method. In the case of SL-150°, L_1/L_2 recovers the image with RMSE of 0.014, while RMSEs of other approaches all exceed 0.020. Overall, the proposed L_1/L_2 approach achieves significant improvements over SART, L_1, and L_1-L_2. Here L_p is

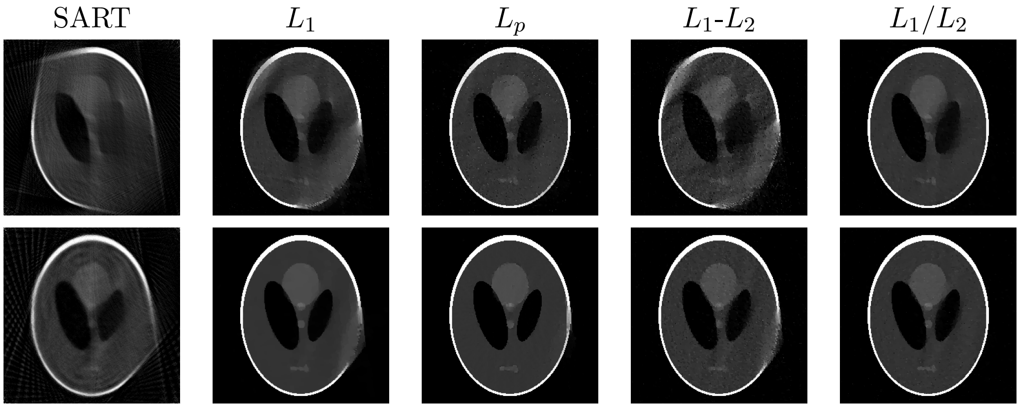

Figure 9. *CT reconstruction from* 90° *(top) and* 150° *(bottom) fan beam projection for the SL phantom with* 0.5% *Gaussian noise. The grayscale window is* $[0, 1]$.

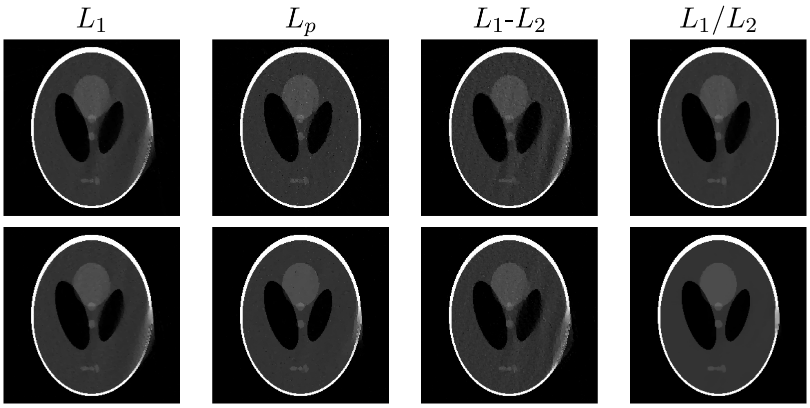

Figure 10. *CT reconstruction from the* 150° *fan beam projection for the SL phantom with Poisson noise* $I_0 = 10^5$ *using LS (top) and WLS (bottom) data-fidelity terms. The grayscale window is* $[0, 1]$.

comparable to L_1/L_2 only in the case of wider scanning ranges and ground-truth images with simple geometries.

Lastly, we consider more realistic noise statistics, i.e., Poisson noise, for the CT problem. Under such a noise model, we also examine a popular data fitting term, called weighted least squares (WLS) [57], to measure the data misfit. In fact, WLS replaces the LS term in (3.1) by $\frac{\lambda}{2}\|Au-f\|_W^2 := \frac{\lambda}{2}(Au-f)^T W(Au-f)$, where $W = \text{diag}(\exp(-f))$. As a result, we can simply modify the LS implementations to fit in WLS. We present one example of reconstructing the SL phantom from 150° fan beam projection with noise level $I_0 = 10^5$. Figure 10 shows similar results of LS and WLS. Specifically, LS gives a better recovery of the skull, while WLS has fewer fluctuations inside the brain. We further compare the two data terms under different noise levels in Table 3, reporting minor improvements of WLS over LS for all the regularization methods.

Table 3

CT reconstruction from 150° fan beam projection for the SL phantom with Poisson noise by L_1, L_p, L_1-L_2, and L_1/L_2.

I_0	Data-fitting	L_1		L_p		L_1-L_2		L_1/L_2	
		SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE
10^4	LS	0.93	0.057	0.90	0.054	0.79	0.068	0.85	0.053
	WLS	0.91	0.056	0.88	0.051	0.78	0.066	**0.95**	**0.051**
10^5	LS	0.96	0.042	0.95	0.017	0.91	0.046	0.99	0.009
	WLS	0.97	0.041	0.98	0.028	0.93	0.045	**0.99**	**0.007**

5.3. Experimental dataset. We set up a limited-angle CT problem from two experimental datasets [4, 21]. The reference image of the walnut is of size 164×164, while that of the lotus is 128×128. The sinogram for walnut is $f \in \mathbb{R}^{164\times 120}$ (3° per projection), and the projection matrix $A \in \mathbb{R}^{19680\times 26896}$. In the lotus case, $f \in \mathbb{R}^{429\times 120}$ and $A \in \mathbb{R}^{51480\times 16384}$. When we perform the limited-angle CT reconstruction, we take partial data from f. Specifically, we consider a 150° scanning angle by selecting the first 50 projections, i.e., extracting the corresponding rows of A and the columns of sinogram to generate the projection matrix and sinogram, respectively. Since the real data contains noise generated by the CT machine, we do not add additional noise in the sinogram. The reference images shown in Figure 1 are reconstructed from the complete scanning data by using the Tikhonov regularization. We further impose a region of interest (ROI) when computing the quantitative evaluation metrics. The ROI is a circle with radius of 62 for walnut and 72 for lotus.

We consider a $[0, 0.5]$ box constraint on all the regularization methods (L_1, L_p, L_1-L_2, and L_1/L_2), which is estimated from the reference images. We do not assume any noise type (or noise level), and we only consider LS as the data fitting term. The optimal parameters are selected based on the "eyeball" norm of the restored image, focusing on textures and details such as the walnut's shell and its inner structure. The reconstruction results are presented in Figures 11 and 12 for walnut and lotus, respectively, within the corresponding ROIs and under a grayscale window of $[0, 0.6]$. In Figure 11, SART produces a lot of artifacts. The L_1 model makes a good recovery, but loses some details in the bottom-left corner with blurring inner texture. All these nonconvex regularization models have sharper images than L_1, while L_1/L_2 can have a higher contrast, especially for the internal region of the walnut. The lotus is more difficult to reconstruct, as its root is filled with attenuating objects that cause severe metal artifacts. In Figure 12, the restored image via our proposed model has fewer streaking artifacts than the ones by other approaches. Lastly, we provide some quantitative analysis in Table 4. All these regularization methods have similar performance in terms of SSIM and RMSE, while L_1 has the best results. As reference images have some obvious streaking artifacts, the method with the best quantitative measures does not grant the optimal performance.

6. Conclusions and future works. Following a preliminary work [50], we considered the use of L_1/L_2 on the gradient as a regularization for imaging applications. We formulated an unconstrained model, which is novel and suitable when noise is present. We also incorporated a box constraint that is reasonable and yet helpful for the CT reconstruction problem. We provided convergence guarantees for the proposed algorithms under mild conditions. We

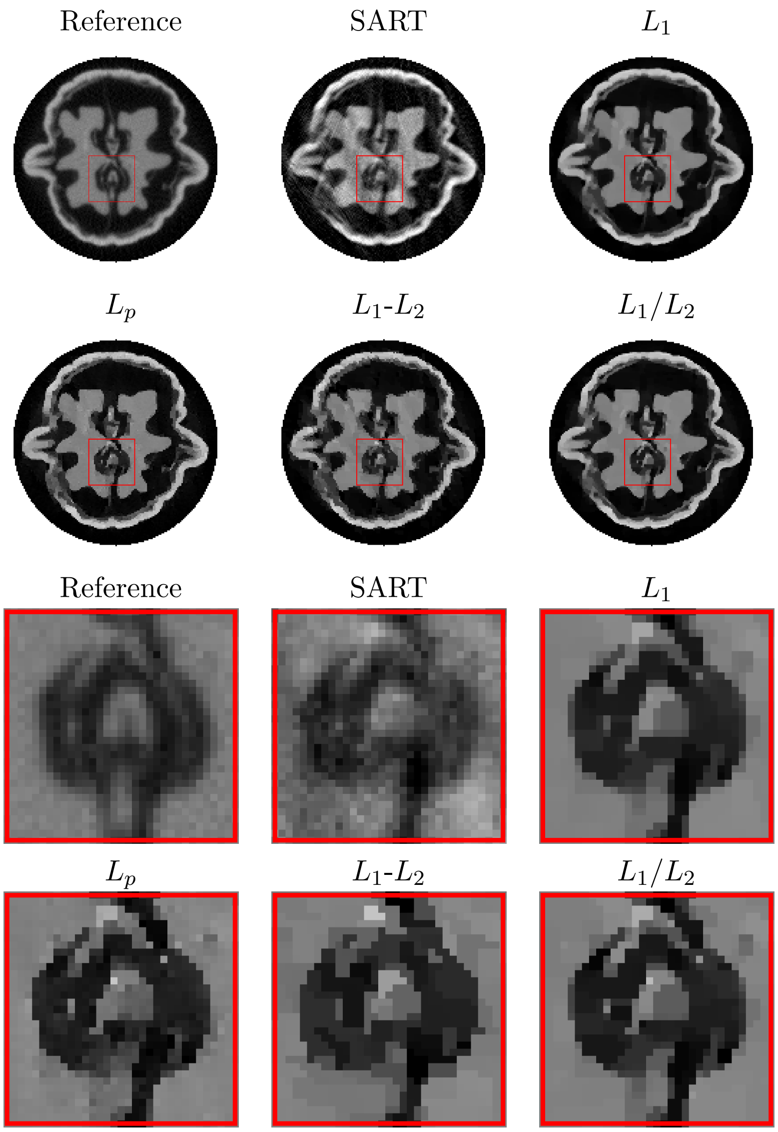

Figure 11. *CT reconstruction of a walnut in the* 150° *projection range. The internal region of the walnut is zoomed-in and highlighted in red. The display window is* $[0, 0.6]$.

conducted extensive experiments to demonstrate that our approaches outperform the state of the art in the limited-angle CT reconstruction subject to either Gaussian noise or Poisson noise. Specifically, we validated the effectiveness and efficiency of our approach with two experimental datasets.

As both L_1 and L_1/L_2 models take about 10 minutes to run on MATLAB, we will implement the algorithms on the GPU for fast computation. Extensions to a higher dimension as well as to other medical and biological applications with real data, e.g., MRI, cone-beam CT, positron emission tomography (PET), and transmission electron microscopy (TEM), are worth exploring in the future.

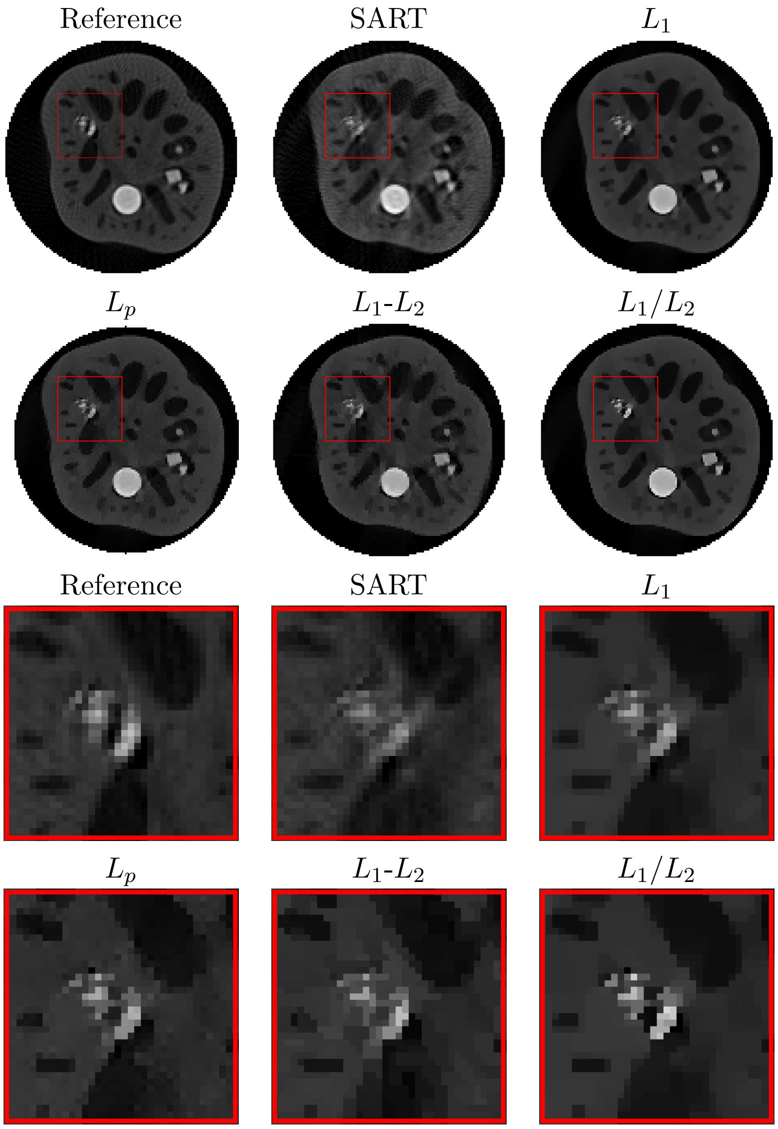

Figure 12. *CT reconstruction of a lotus in the* 150° *projection range. One hole filled with attenuating objects is zoomed-in and highlighted in red. The display window is* $[0, 0.6]$.

Table 4
CT reconstruction of experimental data.

Dataset	SART		L_1		L_p		L_1-L_2		L_1/L_2	
	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE
walnut	0.87	0.049	**0.92**	**0.036**	0.91	0.040	0.90	0.041	0.91	0.041
lotus	0.93	0.022	**0.96**	**0.015**	0.96	0.016	0.95	0.017	0.96	0.017

Appendix A. Proofs. To prepare for convergence analysis, we summarize some equivalent conditions for strong convexity and Lipschitz smooth functions in Lemmas A.1 and A.2, respectively.

Lemma A.1. *A function $f(x)$ is called strongly convex with parameter μ if and only if one of the following conditions holds:*
(a) $g(x) = f(x) - \frac{\mu}{2}\|x\|_2^2$ *is convex;*
(b) $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x-y\|_2^2 \ \forall x, y;$
(c) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y-x\|_2^2 \ \forall x, y.$

Lemma A.2. *The gradient of $f(x)$ is Lipschitz continuous with parameter $L > 0$ if and only if one of the following conditions holds:*
(a) $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2 \ \forall x, y;$
(b) $g(x) = \frac{L}{2}\|x\|_2^2 - f(x)$ *is convex;*
(c) $f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y-x\|_2^2 \ \forall x, y.$

We show in Lemma A.3 that the gradient of the function $f(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|_2}$ is Lipschitz continuous on a set with a lower bound.

Lemma A.3. *Given a function $f(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|_2}$ and a set $\mathcal{M}_\epsilon := \{\mathbf{x} | \|\mathbf{x}\|_2 \geq \epsilon\}$ for a positive constant $\epsilon > 0$, we have*

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq \frac{2}{\epsilon^3}\|\mathbf{x}-\mathbf{y}\|_2 \ \forall \ \mathbf{x}, \mathbf{y} \in \mathcal{M}_\epsilon.$$

Proof. Some calculations lead to $\nabla f(\mathbf{x}) = -\frac{\mathbf{x}}{\|\mathbf{x}\|_2^3}$ and $\nabla^2 f(\mathbf{x}) = -\frac{1}{\|\mathbf{x}\|_2^3}I + 3\mathbf{x}\mathbf{x}^T\frac{1}{\|\mathbf{x}\|_2^5}$ with the identify matrix I. Then $\forall \mathbf{y}$, one has

$$\mathbf{y}^T\nabla^2 f(\mathbf{x})\mathbf{y} = -\frac{\mathbf{y}^T\mathbf{y}}{\|\mathbf{x}\|_2^3} + 3\frac{\mathbf{y}^T\mathbf{x}\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2^5} \leq 2\frac{\mathbf{y}^T\mathbf{y}}{\|\mathbf{x}\|_2^3} \leq \frac{2}{\epsilon^3}\mathbf{y}^T\mathbf{y},$$

which implies that the maximum spectral radius of Hessian of f is less than $\frac{2}{\epsilon^3}$. ■

A.1. Proof of Lemma 4.2.

Proof. It follows from the optimality condition of the $\mathbf{h}$-subproblem in (4.1) that

$$-\frac{a^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|^3}\mathbf{h}^{(k+1)} + \rho_2\left(\mathbf{h}^{(k+1)} - \nabla u^{(k+1)} - \mathbf{b}_2^{(k)}\right) = 0, \tag{A.1}$$

where $a^{(k)} := \|\nabla u^{(k)}\|_1$. Using the dual update $-\mathbf{b}_2^{(k+1)} = \mathbf{h}^{(k+1)} - \nabla u^{(k+1)} - \mathbf{b}_2^{(k)}$, we have

$$\mathbf{b}_2^{(k+1)} = -\frac{a^{(k+1)}}{\rho_2}\frac{\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2^3}, \tag{A.2}$$

and similarly

$$\mathbf{b}_2^{(k)} = -\frac{a^{(k)}}{\rho_2}\frac{\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|_2^3}. \tag{A.3}$$

We can estimate

$$\begin{aligned} \|\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}\|_2 &= \frac{1}{\rho_2} \left\| a^{(k+1)} \frac{\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2^3} - a^{(k)} \frac{\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|^3} \right\|_2 \\ &\leq \frac{1}{\rho_2} \left(\frac{1}{\|\mathbf{h}^{(k+1)}\|_2^2} |a^{(k+1)} - a^{(k)}| + a^{(k)} \left\| \frac{\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2^3} - \frac{\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|_2^3} \right\|_2 \right). \end{aligned} \tag{A.4}$$

For the first term in (A.4), we use the facts that $\|\mathbf{x}\|_1 \leq \sqrt{l}\|\mathbf{x}\|_2$ for a vector $\mathbf{x}$ of length l and $\|\nabla\|_2^2 \leq 8$, thus leading to

$$\begin{aligned} |a^{(k+1)} - a^{(k)}| &\leq \|\nabla(u^{(k+1)} - u^{(k)})\|_1 \leq \sqrt{2mn}\|\nabla(u^{(k+1)} - u^{(k)})\|_2 \\ &\leq \sqrt{2mn} \cdot \|\nabla\|_2 \cdot \|u^{(k+1)} - u^{(k)}\|_2 \leq 4\sqrt{mn}\|u^{(k+1)} - u^{(k)}\|_2. \end{aligned} \tag{A.5}$$

Note that $u \in \mathbb{R}^{m\times n}$ and $\nabla u \in \mathbb{R}^{m\times n\times 2}$ (thus of length $2mn$.) Invoking Lemma A.3, we get

$$a^{(k)} \left\| \frac{\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2^3} - \frac{\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|_2^3} \right\|_2 \leq \frac{2M}{\epsilon^3} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2. \tag{A.6}$$

By putting together (A.4)–(A.6) and using the Cauchy–Schwarz inequality, we get (4.2). ∎

A.2. Proof of Lemma 4.3. In order to prove Lemma 4.3, we show in Lemma A.4 that the augmented Lagrangian decreases sufficiently with respect to $u^{(k)}$.

Lemma A.4. *Under the same assumptions as in Lemma 4.3, there exists a constant $\bar{c}_1 > 0$ such that*

$$\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) - \mathcal{L}(u^{(k)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) \leq -\frac{\bar{c}_1}{2}\|u^{(k+1)} - u^{(k)}\|_2^2 \tag{A.7}$$

holds for the augmented Lagrangian corresponding to L_1/L_2-uncon and L_1/L_2-box.

Proof. Denote by σ the smallest eigenvalue of the matrix $A^TA + \nabla^T\nabla$. We show that σ is strictly positive. If $\sigma = 0$, there exists a vector x such that $x^T(A^TA + \nabla^T\nabla)x = 0$. It is straightforward that $x^TA^TAx \geq 0$ and $x^T\nabla^T\nabla x \geq 0$. Therefore, one shall have $x^TA^TAx = 0$ and $x^T\nabla^T\nabla x = 0$, which contradicts assumption A1 that $\mathcal{N}(\nabla) \bigcap \mathcal{N}(A) = 0$. Therefore, we have that

$$v^T(A^TA + \nabla^T\nabla)v \geq \sigma\|v\|_2^2 \quad \forall v,$$

which implies that $\mathcal{L}_{\text{uncon}}(u, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)})$ with fixed $\mathbf{h}^{(k)}$ and with $\mathbf{b}_2^{(k)}$ strongly convex with parameter $\bar{c}_1 = \sigma\lambda$ (we can choose $\rho_2 \geq \lambda$ as it is sufficiently large). It follows from (3.14) that the only difference between $\mathcal{L}_{\text{uncon}}$ and $\mathcal{L}_{\text{box}}$ is the indicator function $\Pi_{[c,d]}(u)$. Since the indicator function is convex, then $\mathcal{L}_{\text{box}}$ is strongly convex with the same parameter c_1. We can unify $\mathcal{L}_{\text{uncon}}$ and $\mathcal{L}_{\text{box}}$ to be $\mathcal{L}$. Then Lemma A.1 leads to

$$\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) \leq \mathcal{L}(u^{(k)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) - \frac{\sigma\lambda}{2}\|u^{(k+1)} - u^{(k)}\|_2^2.$$

Therefore, we can choose $\bar{c}_1 = \sigma\lambda$ such that inequality (A.7) holds. ∎

Now we are ready to prove Lemma 4.3.

Proof. Denote $a = \|u^{(k+1)}\|_1$ and $L = \frac{2M}{\epsilon^3}$. Lemmas A.2 and A.3 lead to

$$\frac{a}{\|\mathbf{h}^{(k+1)}\|_2} \leq \frac{a}{\|\mathbf{h}^{(k)}\|_2} - \left\langle \frac{a\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|_2^3}, \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)} \right\rangle + \frac{L}{2}\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2. \tag{A.8}$$

Denoting $\mathbf{z} = \nabla u^{(k+1)} + \mathbf{b}_2^{(k)}$ and using the optimality condition of $\mathbf{h}^{(k+1)}$ (A.1), we get

$$\begin{aligned} &\frac{\rho_2}{2}\|\mathbf{h}^{(k+1)} - \mathbf{z}\|_2^2 - \frac{\rho_2}{2}\|\mathbf{h}^{(k)} - \mathbf{z}\|_2^2 \\ &= \frac{\rho_2}{2}\|\mathbf{h}^{(k+1)}\|_2^2 - \frac{\rho_2}{2}\|\mathbf{h}^{(k)}\|_2^2 - \left\langle -\frac{a\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|^3} + \rho_2\mathbf{h}^{(k+1)}, \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)} \right\rangle \\ &= \left\langle \frac{a\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|^3}, \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)} \right\rangle - \frac{\rho_2}{2}\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2. \end{aligned} \tag{A.9}$$

Combining (A.8) and (A.9), we obtain

$$\begin{aligned} &\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}; \mathbf{b}_2^{(k)}) - \mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) \\ &\leq \left\langle \frac{a\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2^3} - \frac{a\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|_2^3}, \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)} \right\rangle - \frac{\rho_2 - L}{2}\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2 \\ &\leq \left\| \frac{a\mathbf{h}^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2^3} - \frac{a\mathbf{h}^{(k)}}{\|\mathbf{h}^{(k)}\|_2^3} \right\|_2 \left\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\right\|_2 - \frac{\rho_2 - L}{2}\left\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\right\|_2^2 \\ &\leq -\frac{\rho_2 - 3L}{2}\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2. \end{aligned} \tag{A.10}$$

Lastly, from the update of $\mathbf{b}_2$, we compute

$$\begin{aligned} &\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}; \mathbf{b}_2^{(k+1)}) - \mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}; \mathbf{b}_2^{(k)}) \\ &= \frac{\rho_2}{2}\left(\|\mathbf{b}_2^{(k)}\|_2^2 - \|\mathbf{b}_2^{(k+1)} - 2\mathbf{b}_2^{(k)}\|_2^2\right) \leq \frac{\rho_2}{2}\|\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}\|_2^2. \end{aligned} \tag{A.11}$$

By putting inequalities (A.7), (A.10), and (A.11) together with Lemma 4.2, we have

$$\mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}; \mathbf{b}_2^{(k+1)}) \leq \mathcal{L}(u^{(k)}, \mathbf{h}^{(k)}; \mathbf{b}_2^{(k)}) - c_1\|u^{(k+1)} - u^{(k)}\|_2^2 - c_2\|\mathbf{h}^{(k)} - \mathbf{h}^{(k+1)}\|_2^2,$$

where $c_1 = \frac{\bar{c}_1}{2} - \frac{16mn}{\rho_2\epsilon^4}$ and $c_2 = \frac{\rho_2\epsilon^3 - 6M}{2\epsilon^3} - \frac{16M^2}{\rho_2\epsilon^6}$. For sufficiently large ρ_2, we can have $c_1, c_2 > 0$. ■

Remark A.5. It seems that we need a very large value of ρ_2 to guarantee $c_1, c_2 > 0$ in Lemma 4.3. Fortunately, it is just a sufficient condition for convergence, and we can choose a reasonable value of ρ_2 in practice; please refer to section 5 for parameter tuning.

A.3. Proof of Lemma 4.4.

Proof. To accommodate the models (with and without box), we express the optimality condition of (4.1) as follows:

$$\text{(A.12)} \qquad \begin{cases} \frac{p^{(k+1)}}{\|\mathbf{h}^{(k)}\|_2} + q^{(k+1)} + r^{(k+1)} + \rho_2 \nabla^T(\nabla u^{(k+1)} - \mathbf{h}^{(k)} + \mathbf{b}_2^{(k)}) = 0, \\ -\frac{\|\nabla u^{(k+1)}\|_1}{\|\mathbf{h}^{(k+1)}\|_2^3}\mathbf{h}^{(k+1)} + \rho_2(\mathbf{h}^{(k+1)} - \nabla u^{(k+1)} - \mathbf{b}_2^{(k)}) = 0, \\ \mathbf{b}_2^{(k+1)} = \mathbf{b}_2^{(k)} + \nabla u^{(k+1)} - \mathbf{h}^{(k+1)}, \end{cases}$$

where $p^{(k+1)} \in \partial\|\nabla u^{(k+1)}\|_1$, $q^{(k+1)} := \lambda A^T(Au^{(k+1)} - f)$, and $r^{(k+1)}$ either belongs to $\partial(\Pi_{[c,d]}(u^{(k+1)}))$ with the box constraint or is zero otherwise. Let $\eta_1^{(k+1)}, \eta_2^{(k+1)}, \eta_2^{(k+1)}$ be

$$\text{(A.13)} \qquad \begin{cases} \eta_1^{(k+1)} := \frac{p^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2} + q^{(k+1)} + r^{(k+1)} + \rho_2 \nabla^T(\nabla u^{(k+1)} - \mathbf{h}^{(k+1)} + \mathbf{b}_2^{(k+1)}), \\ \eta_2^{(k+1)} := -\frac{\|\nabla u^{(k+1)}\|_1}{\|\mathbf{h}^{(k+1)}\|_2^3}\mathbf{h}^{(k+1)} + \rho_2(\mathbf{h}^{(k+1)} - \nabla u^{(k+1)} - \mathbf{b}_2^{(k+1)}), \\ \eta_3^{(k+1)} := \rho_2(\nabla u^{(k+1)} - \mathbf{h}^{(k+1)}). \end{cases}$$

Clearly, we have

$$\begin{aligned} \eta_1^{(k+1)} &\in \partial_u \mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}, \mathbf{b}_2^{(k+1)}), \\ \eta_2^{(k+1)} &\in \partial_{\mathbf{h}} \mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}, \mathbf{b}_2^{(k+1)}), \\ \eta_3^{(k+1)} &\in \partial_{\mathbf{b}_2} \mathcal{L}(u^{(k+1)}, \mathbf{h}^{(k+1)}, \mathbf{b}_2^{(k+1)}) \end{aligned}$$

for $\mathcal{L} = \mathcal{L}_{\text{uncon}}$ or $\mathcal{L}_{\text{box}}$. Combining (A.12) and (A.13) leads to

$$\begin{cases} \eta_1^{(k+1)} = -\frac{p^{(k+1)}}{\|\mathbf{h}^{(k)}\|_2} + \frac{p^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2} + \rho_2\nabla^T(\mathbf{h}^{(k)} - \mathbf{h}^{(k+1)}) + \rho_2 \nabla^T(\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}), \\ \eta_2^{(k+1)} = \rho_2(\mathbf{b}_2^{(k)} - \mathbf{b}_2^{(k+1)}), \\ \eta_3^{(k+1)} = \rho_2(\mathbf{b}_2^{(k+1)} - \mathbf{b}_2^{(k)}). \end{cases}$$

The chain rule of subgradient [24, 25] suggests that $\partial\|\nabla u\|_1 = \nabla^T \mathbf{q}$, where

$$\mathbf{q} = \{\mathbf{q} \mid \langle \mathbf{q}, \nabla u\rangle_Y = \|\nabla u\|_1, \ |q_{ijk}| \le 1 \, \forall i,j,k\}.$$

Therefore, we have an upper bound for $\|p^{(k+1)}\|_2 \le \|\nabla^T\|_2 \|\mathbf{q}^{(k+1)}\|_2 \le 2\sqrt{2mn}$. Simple calculations show that

$$\begin{aligned} &\left\| \frac{p^{(k+1)}}{\|\mathbf{h}^{(k)}\|_2} - \frac{p^{(k+1)}}{\|\mathbf{h}^{(k+1)}\|_2} \right\|_2 = \left| \frac{1}{\|\mathbf{h}^{(k)}\|_2} - \frac{1}{\|\mathbf{h}^{(k+1)}\|_2} \right| \left\| p^{(k+1)} \right\|_2 \\ &\le \frac{1}{\epsilon^2} \left\| \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)} \right\|_2 \left\| p^{(k+1)} \right\|_2 \le \frac{2\sqrt{2mn}}{\epsilon^2} \left\| \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)} \right\|_2. \end{aligned}$$

Finally, by setting $\gamma = \max\{26\rho^2, \ 24\rho^2 + \frac{24mn}{\epsilon^4}\}$, (4.4) follows immediately. ∎

REFERENCES

[1] A. H. Andersen and A. C. Kak, *Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm*, Ultrason. Imaging, 6 (1984), pp. 81–94.

[2] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2007), pp. 1205–1223, https://doi.org/10.1137/050644641.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.

[4] T. A. Bubba, A. Hauptmann, S. Huotari, J. Rimpeläinen, and S. Siltanen, *Tomographic X-ray Data of a Lotus Root Filled with Attenuating Objects*, preprint, https://arxiv.org/abs/1609.07299, 2016.

[5] J. T. Bushberg and J. M. Boone, *The Essential Physics of Medical Imaging*, Lippincott Williams & Wilkins, 2011.

[6] E. J. Candés, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, J Fourier Anal Appl., 14 (2008), pp. 877–905.

[7] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 20 (2004), pp. 89–97.

[8] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011).

[9] R. H. Chan and J. Ma, *A multiplicative iterative algorithm for box-constrained penalized likelihood image restoration*, IEEE Trans. Image Process., 21 (2012), pp. 3168–3181.

[10] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005, https://doi.org/10.1137/1.9780898717877.

[11] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 10 (2007), pp. 707–710.

[12] Z. Chen, X. Jin, L. Li, and G. Wang, *A limited-angle CT reconstruction method based on anisotropic TV minimization*, Phys. Med. Biol., 58 (2013), pp. 2119–2141.

[13] B. De Man and J. A. Fessler, *Statistical iterative reconstruction for x-ray computed tomography*, Bio. Math. Promising Dir. Imaging Therapy Plan. Inverse Prob., (2010), pp. 113–40.

[14] I. A. Elbakri and J. A. Fessler, *Statistical image reconstruction for polyenergetic X-ray computed tomography*, IEEE Trans. Med. Imaging, 21 (2002), pp. 89–99.

[15] L. A. Feldkamp, L. C. Davis, and J. W. Kress, *Practical cone-beam algorithm*, J. Opt. Soc. Am. A, 1 (1984), pp. 612–619.

[16] J. Frikel and E. T. Quinto, *Characterization and reduction of artifacts in limited angle tomography*, Inverse Probl., 29 (2013), 125007.

[17] S. Gazzola, P. C. Hansen, and J. G. Nagy, *IR tools: A MATLAB package of iterative regularization methods and large-scale test problems*, Numer. Algorithms, 81 (2019), pp. 773–811.

[18] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343, https://doi.org/10.1137/080725891.

[19] R. Gu and A. Dogandžić, *Polychromatic X-ray CT Image Reconstruction and Mass-Attenuation Spectrum Estimation*, preprint, https://arxiv.org/abs/1509.02193, 2015.

[20] K. Guo, D. Han, and T. Wu, *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints*, Int. J. Comput. Math., 94 (2017), pp. 1653–1669.

[21] K. Hämäläinen, L. Harhanen, A. Kallonen, A. Kujanpää, E. Niemi, and S. Siltanen, *Tomographic X-ray Data of a Walnut*, preprint, https://arxiv.org/abs/1502.04064, 2015.

[22] X. Han, J. Bian, D. R. Eaker, T. L. Kline, E. Y. Sidky, E. L. Ritman, and X. Pan, *Algorithm-enabled low-dose micro-CT imaging*, IEEE Trans. Med. Imag., 30 (2010), pp. 606–620.

[23] P. C. Hansen and M. Saxild-Hansen, *AIR tools—MATLAB package of algebraic iterative reconstruction methods*, J. Comput. Appl. Math., 236 (2012), pp. 2167–2178.

[24] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Vol. I, Springer-Verlag, Heidelberg, 1996.

[25] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Vol. II, Springer-Verlag, Heidelberg, 1996.

[26] M. Hong, Z.-Q. Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optim., 26 (2016), pp. 337–364, https://doi.org/10.1137/140990309.

[27] Y. Huang, O. Taubmann, X. Huang, V. Haase, G. Lauritsch, and A. Maier, *Scale-space anisotropic total variation for limited angle tomography*, IEEE Trans. Radiat. Plasma Med. Sci., 2 (2018), pp. 307–314.

[28] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, *GPU-based iterative cone-beam CT reconstruction using tight frame regularization*, Phys. Med. Biol., 56 (2011), pp. 3787–3807.

[29] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, *GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation*, Med. Phys., 37 (2010), pp. 1757–1760.

[30] M. Jiang and G. Wang, *Convergence of the simultaneous algebraic reconstruction technique (SART)*, IEEE Trans. Image Process., 12 (2003), pp. 957–961.

[31] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898719277.

[32] K. Kan, S. W. Fung, and L. Ruthotto, *PNKH-B: A Projected Newton-Krylov Method for Large-Scale Bound-Constrained Optimization*, preprint, https://arxiv.org/abs/2005.13639, 2020.

[33] G. Li and T. K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM J. Optim., 25 (2015), pp. 2434–2460, https://doi.org/10.1137/140998135.

[34] F. Liu, J. Rosenberger, Y. Lou, R. Hosseini, J. Su, and S. Wang, *Graph regularized EEG source imaging with in-class consistency and out-class discrimination*, IEEE Trans. Big Data, 3 (2017), pp. 378–391.

[35] F. Liu, S. Wang, J. Qin, Y. Lou, and J. Rosenberger, *Estimating latent brain sources with low-rank representation and graph regularization*, in Brain Informatics, Springer, 2018, pp. 304–316.

[36] Y. Lou and M. Yan, *Fast L_1-L_2 minimization via a proximal operator*, J. Sci. Comput., 74 (2018), pp. 767–785.

[37] Y. Lou, P. Yin, Q. He, and J. Xin, *Computing sparse representation in a highly coherent dictionary based on difference of L_1 and L_2*, J. Sci. Comput., 64 (2015), pp. 178–196.

[38] Y. Lou, P. Yin, and J. Xin, *Point source super-resolution via non-convex L_1 based methods*, J. Sci. Comput., 68 (2016), pp. 1082–1100.

[39] Y. Lou, T. Zeng, S. Osher, and J. Xin, *A weighted difference of anisotropic and isotropic total variation model for image processing*, SIAM J. Imaging Sci., 8 (2015), pp. 1798–1823, https://doi.org/10.1137/14098435X.

[40] Y. Lou, X. Zhang, S. J. Osher, and A. L. Bertozzi, *Image recovery via nonlocal operators*, J. Sci. Comput., 42 (2010), pp. 185–197.

[41] X. Luo, W. Yu, and C. Wang, *An image reconstruction method based on total variation and wavelet tight frame for limited-angle CT*, IEEE Access, 6 (2017), pp. 1461–1470.

[42] M. Lustig, D. L. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnet. Reson. Med., 58 (2007), pp. 1182–1195.

[43] D. Lv, Q. Zhou, J. K. Choi, J. Li, and X. Zhang, *Nonlocal TV-Gaussian prior for Bayesian inverse problems with applications to limited CT reconstruction*, Inverse Probl. Imag., 14 (2020), pp. 117–132.

[44] J. Lv and Y. Fan, *A unified approach to model selection and sparse recovery using regularized least squares*, Ann. Appl. Stat., 37 (2009), pp. 3498–3528.

[45] T.-H. Ma, Y. Lou, and T.-Z. Huang, *Truncated l_{1-2} models for sparse recovery and rank minimization*, SIAM J. Imaging Sci., 10 (2017), pp. 1346–1380, https://doi.org/10.1137/16M1098929.

[46] L. V. Nguyen, *How strong are streak artifacts in limited angle computed tomography?*, Inverse Probl., 31 (2015), 055003.

[47] H. Nien and J. A. Fessler, *Fast X-ray CT image reconstruction using a linearized augmented Lagrangian method with ordered subsets*, IEEE Trans. Med. Imaging, 34 (2014), pp. 388–399.

[48] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2006.

[49] J.-S. Pang and M. Tao, *Decomposition methods for computing directional stationary solutions of a class of nonsmooth nonconvex optimization problems*, SIAM J. Optim., 28 (2018), pp. 1640–1669, https://doi.org/10.1137/17M1110249.

[50] Y. Rahimi, C. Wang, H. Dong, and Y. Lou, *A scale-invariant approach for sparse signal recovery*, SIAM J. Sci. Comput., 41 (2019), pp. A3649–A3672, https://doi.org/10.1137/18M123147X.

[51] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[52] L. A. Shepp and B. F. Logan, *The Fourier reconstruction of a head section*, IEEE Trans. Nucl. Sci., 21 (1974), pp. 21–43.
[53] E. Y. Sidky, C. Kao, and X. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*, J. X-Ray Sci. Technol., 14 (2006), pp. 1–21.
[54] E. Y. Sidky and X. Pan, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Phys. Med. Biol., 53 (2008), pp. 4777–4807.
[55] X. Tang, J. Hsieh, R. A. Nilsen, S. Dutta, D. Samsonov, and A. Hagiwara, *A three-dimensional-weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric CT-helical scanning*, Phys. Med. Biol., 51 (2006), pp. 855–874.
[56] M. Tao and Y. Lou, *Minimization of L_1 over L_2 for Sparse Signal Recovery with Convergence Guarantee*, http://www.optimization-online.org/DB_HTML/2020/10/8064.html, 2020.
[57] 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., 34 (2007), pp. 4526–4544.
[58] R. Tovey, M. Benning, C. Brune, M. J. Lagerwerf, S. M. Collins, R. K. Leary, P. A. Midgley, and C.-B. Schönlieb, *Directional sinogram inpainting for limited angle tomography*, Inverse Probl., 35 (2019), 024004.
[59] C. Wang, R. Chan, M. Nikolova, R. Plemmons, and S. Prasad, *Nonconvex optimization for 3-dimensional point source localization using a rotating point spread function*, SIAM J. Imaging Sci., 12 (2019), pp. 259–286, https://doi.org/10.1137/18M1178566.
[60] C. Wang, M. Tao, C.-N. Chuah, J. Nagy, and Y. Lou, *Minimizing L1 over L2 Norms on the Gradient*, preprint, https://arxiv.org/abs/2101.00809, 2021.
[61] C. Wang, M. Yan, Y. Rahimi, and Y. Lou, *Accelerated schemes for the L_1/L_2 minimization*, IEEE Trans. Signal Process., 68 (2020), pp. 2660–2669.
[62] F. Wang, W. Cao, and Z. Xu, *Convergence of multi-block Bregman ADMM for nonconvex composite problems*, Sci. China Info. Sci., 61 (2018), 122101.
[63] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 29–63.
[64] 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., 13 (2004), pp. 600–612.
[65] Z. Wang, Z. Huang, Z. Chen, L. Zhang, X. Jiang, K. Kang, H. Yin, Z. Wang, and M. Stampanoni, *Low-dose multiple-information retrieval algorithm for x-ray grating-based imaging*, Nucl. Instrum. Methods Phys., 635 (2011), pp. 103–107.
[66] Z. Xu, X. Chang, F. Xu, and H. Zhang, *$L_{1/2}$ regularization: A thresholding representation theory and a fast solver*, IEEE Trans. Neural Netw. Learn. Syst., 23 (2012), pp. 1013–1027.
[67] P. Yin, E. Esser, and J. Xin, *Ratio and difference of l_1 and l_2 norms and sparse representation with coherent dictionaries*, Comm. Info. Systems, 14 (2014), pp. 87–109.
[68] P. Yin, Y. Lou, Q. He, and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*, SIAM J. Sci. Comput., 37 (2015), pp. A536–A563, https://doi.org/10.1137/140952363.
[69] H. Yu and G. Wang, *A soft-threshold filtering approach for reconstruction from a limited number of projections*, Phys. Med. Biol., 55 (2010), pp. 3905–3916.
[70] Z. Yu, F. Noo, F. Dennerlein, A. Wunderlich, G. Lauritsch, and J. Hornegger, *Simulation tools for two-dimensional experiments in x-ray computed tomography using the FORBILD head phantom*, Phys. Med. Biol., 57 (2012), pp. N237–N252.
[71] S. Zhang and J. Xin, *Minimization of transformed L_1 penalty: Closed form representation and iterative thresholding algorithms*, Comm. Math. Sci., 15 (2017), pp. 511–537.
[72] S. Zhang and J. Xin, *Minimization of transformed L_1 penalty: Theory, difference of convex function algorithm, and robust application in compressed sensing*, Math. Program., 169 (2018), pp. 307–336.
[73] Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M. M. Goodsitt, L. M. Hadjiiski, J. Ge, and C. Zhou, *A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis*, Med. Phys., 33 (2006), pp. 3781–3795.