PAPER

Zero norm based analysis model for image smoothing and reconstruction

To cite this article: Jiebo Song *et al* 2020 *Inverse Problems* **36** 115009

IOP Publishing Inverse Problems
Inverse Problems **36** (2020) 115009 (22pp) https://doi.org/10.1088/1361-6420/abb87e

Zero norm based analysis model for image smoothing and reconstruction

Jiebo Song[1], Jia Li[1], Zhengan Yao[1], Kaisheng Ma[2] and Chenglong Bao[3,*]

[1] School of Mathematics, Sun Yat-Sen University, Guangzhou, People's Republic of China
[2] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, People's Republic of China
[3] Yau Mathematical Sciences Center, Tsinghua University, Beijing, People's Republic of China

E-mail: clbao@mail.tsinghua.edu.cn

Received 1 January 2020, revised 29 July 2020
Accepted for publication 15 September 2020
Published 15 October 2020

Abstract

The sparsity-based approaches have demonstrated promising performance in image processing. In this paper, for better preservation of the salient edge structures of images, we propose an $\ell_0 + \ell_2$-norm based analysis model, which requires solving a challenging non-separable ℓ_0-norm related minimization problem, and we also propose an inexact augmented Lagrangian method with proven convergence to a local minimum. Extensive experiments in image smoothing, including texture removal and context smoothing, show that our method achieves better visual results over various sparsity-based models and the CNN method. Also, experiments on sparse view CT reconstruction further validate the advantage of the proposed method.

Keywords: L0 optimization, image smoothing and restoration, sparse view CT imaging

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

1. Introduction

Image smoothing and reconstruction are two critical and classical problems in computational photography and image processing. In general, image smoothing is to enhance salient structures such that the image is further used on many other applications, including edge extraction, image abstraction, pencil drawing and detail magnification [1–3]. Image reconstruction is to

*Author to whom any correspondence should be addressed.

recover all the details of the ground truth image from the corrupted observations, such as denoising, deblurring and medical computed tomography (CT) image reconstruction. In both applications, the main challenge is to develop an effective approach that preserves the salient edges while suppressing unrelated structures. To address the above problems, many models have been proposed and have witnessed the success in some applications. Among these works, regularization based models are representative ones, which can be formulated as solving the following minimization:

$$\min_u f(u, u_0) + \lambda R(u), \tag{1}$$

where u_0 and u are the input and output images, respectively, f denotes the data fidelity, R denotes the regularization term and λ is a positive constant that balances the above two terms. One commonly used regularization is the so-called sparsity prior, e.g. ℓ_1-norm, ℓ_0-norm, and their variants. Compared to the Tikhonov regularization which always results in a blur image, the sparsity based methods better preserve the image edges and improves the visual quality of images. One typical sparse regularization is the so-called Rudin–Osher–Fatemi model [4], which promotes the sparsity of image gradients using ℓ_1-norm or $\ell_{1,2}$-norm. Following the spirit of this model, the ℓ_1-norm and its related convex sparse regularizations have been extensively explored for improving the reconstruction results. Despite their numerical success in many applications, the ℓ_1-norm based models cause bias on large coefficients which noticeably lower the visual quality [5, 6]. Besides, Wu *et al* [7] have shown that the convex sparse prior has difficulty in preserving the image contrast.

In order to overcome the above weaknesses of convex regularizations, many non-convex models have been proposed such as ℓ_p-norm [8], smoothly clipped absolute deviations [5], and minimax concave penalty [6]. Among these models, the ℓ_0-norm is the most straightforward one, which is non-convex and non-smooth. In recent years, it has been successfully applied to many tasks in image processing, e.g. image smoothing [1], segmentation [9], super-resolution [10], deblurring [11, 12] and enhancement [13]. Specifically, the ℓ_0-norm of image gradients that globally controls the sparsity of images has been applied in image smoothing [1], text deblurring [11] and medical image reconstruction [14–16]. These ℓ_0-norm based models have satisfactory performance for keeping the sparsity of images and preserving the image edges. However, due to the discontinuity of ℓ_0-norm, it may generate undesired artifacts in the smooth region, e.g. negligible edges, outliers, or sharp spots. To balance the sparsity and the smoothness, we propose to impose an additional squared ℓ_2-norm in the original ℓ_0-norm based model. Thus, our proposed model is to solve the following minimization:

$$\min_u f(u, u_0) + \lambda \|Wu\|_0 + \frac{\beta}{2}\|Wu\|_2^2 \tag{2}$$

where $\lambda, \beta > 0$, $W : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a predefined transform. It is worth noting that the proposed regularization is a non-convex relaxation of the elastic net [17] that encourages the group sparsity. In figure 1, the results demonstrate the advantages of the proposed model compared to other regularization based models in image smoothing. The ℓ_2-norm based model blurs the whole image, the elastic net and the ℓ_0-norm based models can not remove the noise as small edges or spots still exist. However, our model simultaneously keeps the salient edges while suppressing the noise and obtains better visual results.

Due to the existence of ℓ_0-norm, the minimization problem (2) is non-convex and non-smooth. Thus, designing an efficient and convergent numerical algorithm for solving (2) is a challenging problem. According to the choices of W, there are mainly two different

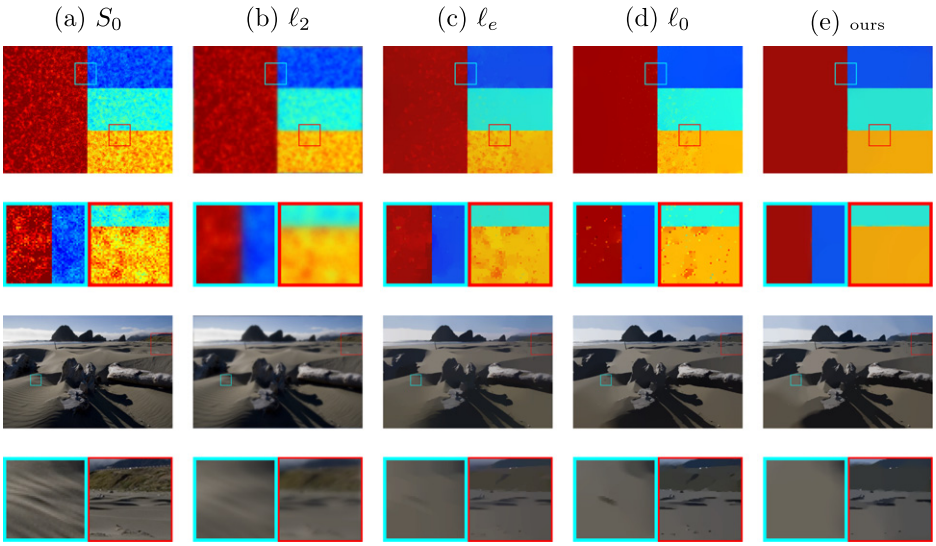

Figure 1. Image smoothing. From left to right: original images, results of ℓ_2-norm based, ℓ_e-norm based [17], ℓ_0-norm based [1] models and our method. The ℓ_2-norm based model blurs images, the ℓ_e and ℓ_0 norm based models obtain higher sparsity, but are still struggling to deal with the situation of existing large scale of noise or some outliers and sharp spots. On contrast, our method yields superior results. Our method has less noise in stripes and sandbeach, especially in the zoom-in areas.

formulations. When W is a bijection, the problem (2) is equivalent to

$$\min_v f(W^{-1}v, u_0) + \lambda\|v\|_0 + \frac{\beta}{2}\|v\|_2^2, \tag{3}$$

where $v = Wu$. In this case, the ℓ_0-norm is separable, and iterative hard thresholding method can be applied [18, 19]. This method achieves good performance in image reconstruction [14, 16] and the global convergence to a local minimizer is established. However, when W is not a bijection, two minimizations (2) and (3) are not equivalent, and the non-separable structure in ℓ_0-norm makes the problem more difficult. One solution is to solve a relaxed version of (2) by adding a penalty term that penalizes the difference between the range space of W and its sparsity. This relaxation reduces the non-separable structure to a separable form such that the previous iterative hard thresholding methods can be applied. Besides, the alternating direction method of multipliers and its variants [12, 15], and the mean doubly augmented Lagrangian method [20] are proposed for solving the non-separable ℓ_0-norm related minimization. However, their analysis does not apply to the problem (2), and thus designing a practical numerical algorithm for solving (2) with theoretical convergence analysis is a challenging problem. Recently, Chen *et al* [21] proposed an augmented Lagrangian method for non-Lipschitz non-convex programming and proved the convergence under certain assumptions, but our model (2) does not satisfy these assumptions. Although the analysis in [21] is not directly applicable to the model (2), their method provides some insights for solving (2). Motivated by the provided general algorithm in [21], we propose an inexact augmented Lagrangian method for solving (2), in which the alternating scheme [16] is applied to approximately solve the subproblem. Moreover, we establish the convergence to a local minimum of

(2) for the sequence generated by the proposed algorithm. In summary, the main contributions of this paper are listed as follows.

- A new sparsity-based approach is proposed with a hybrid ℓ_0-norm and squared ℓ_2-norm with successful applications to image smoothing and reconstruction. This model keeps the salient edges by the sparsity regularization while not introducing unwanted artifacts.
- An inexact augmented Lagrangian method is proposed for solving the resulting non-separable ℓ_0-norm related minimization. Moreover, we analyze the convergence of the proposed algorithm and show the sub-convergence to a local minimizer when the generated sequence is bounded. To the best of our knowledge, this could be the first practical algorithm that contains the convergence proof for solving the non-separable ℓ_0-norm related minimization problem.
- Experiments on texture removal and context smoothing show that our model outperforms other sparsity-based methods, including a convolutional neural network (CNN) based approach. Besides, the proposed method outperforms other regularization based methods in terms of PSNR in sparse-view CT reconstruction.

The rest of the paper is organized as follows: in section 2, the proposed algorithm for solving (2) is proposed with application to the image smoothing and CT imaging reconstruction; the convergence analysis of the proposed algorithm is given in section 3; numerical experiments on both image smoothing and CT reconstruction are presented in section 4 and finally, we conclude in section 5.

2. Proposed approach

Before presenting our numerical method, we first introduce useful definitions for our analysis. Given a point x and a set $\mathcal{X}$, the distance from x to $\mathcal{X}$ is defined as

$$\operatorname{dist}(x, \mathcal{X}) = \inf\{\|x - z\| | z \in \mathcal{X}\}. \tag{4}$$

Definition 2.1. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function.

- The domain of f is $\operatorname{dom} f = \{x | f(x) < +\infty\}$.
- f is lower semicontinuous if $\liminf_{x \to x_0} f(x) \geqslant f(x_0)$.
- f is proper if $f(x) > -\infty$ for all x and $\operatorname{dom} f \neq \emptyset$.

Definition 2.2. Let f be a proper and lower semi-continuous function.

- The Fréchet subdifferential of f is defined as

$$\hat{\partial} f(x) = \left\{ u : \liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geqslant 0 \right\}, \tag{5}$$

  if $x \in \operatorname{dom} f$, and $\hat{\partial} f(x) = \emptyset$ otherwise.
- The limiting subdifferential of f is defined as

$$\partial f(x) = \{u : \exists x^k \to x, f(x^k) \to f(x) \quad \text{and} \quad u^k \in \hat{\partial} f(x^k) \to u\}. \tag{6}$$

- x is a **critical point** of f if $0 \in \partial f(x)$.

Algorithm 1. Inexact augmented Lagrangian method for solving (7).

Initialization: $u^0, y^0 = Wu^0, z^0, \rho_0 > 0, \eta, \tau \in (0,1), \gamma > 1, \Gamma > L_{\rho_0}(u^0, y^0, z^0)$ and $\{\epsilon_k\} \subset \mathbb{R}_+$
for $k = 1$ to N **do**
  find (u^k, y^k), such that

$$(u^k, y^k) \approx \arg\min_{u,y} L_{\rho_k}(u, y, z^{k-1}) \tag{11}$$

  which satisfies (10).
  $z^{k+1} \leftarrow z^k + \rho_k(Wu^k - y^k)$
  if $\eta\|Wu^{k-1} - y^{k-1}\|_2 \geqslant \|Wu^k - y^k\|_2$
  $\rho_{k+1} = \rho_k$.
  else
  $\rho_{k+1} \leftarrow \max(\gamma\rho_k, \|z^{k+1}\|_2^{1+\tau})$.
end
Return: u^N

2.1. The inexact augmented Lagrangian method

By introducing an auxiliary variable y, the minimization (2) is equivalent to

$$\min_{u,y} f(u, u_0) + \lambda\|y\|_0 + \frac{\beta}{2}\|y\|_2^2, \quad \text{s.t. } y = Wu. \tag{7}$$

Throughout this paper, we use $f(u)$ as $f(u, u_0)$ for simplicity and assume that $f(u)$ is lower bounded, convex, smooth and its gradient $\nabla f(u)$ satisfies:

$$\|\nabla f(u) - \nabla f(v)\| \leqslant L\|u - v\|, \quad \forall\, u, v, \tag{8}$$

where $L > 0$. Given $\rho > 0$, the augmented Lagrangian function of (7) is

$$L_\rho(u, y, z) = f(u, u_0) + \frac{\beta}{2}\|y\|_2^2 + \lambda\|y\|_0 + \langle Wu - y, z\rangle + \frac{\rho}{2}\|Wu - y\|_2^2. \tag{9}$$

The inexact augmented Lagrangian method updates the current estimate (u^{k-1}, y^{k-1}, z^k) by the following steps:

(Step 1). Update (u^k, y^k) by solving $\min_{u,y} L_{\rho_k}(u, y, z^k)$ approximately;
(Step 2). Update the multiplier $z^{k+1} = z^k + \rho_k(Wu^k - y^k)$;
(Step 3). Update the penalty parameter ρ_{k+1} by certain criterion.

The above inexact augmented Lagrangian method is a useful framework and can be applied for solving the non-smooth and non-Lipschitz minimization problems [21]. To ensure the convergence of the algorithm, we have to specify the accuracy required in (step 1) and the rule for updating the penalty parameter ρ_k. Following from the recent proposed method [21], we obtain the new estimation (u^{k+1}, y^{k+1}) in (step 1) whenever the following condition is met:

$$\text{dist}(0, \partial_{u,y} L_{\rho_k}(u^k, y^k, z^k)) \leqslant \epsilon_k, \quad \text{and} \quad L_{\rho_k}(u^k, y^k, z^k) \leqslant \Gamma \tag{10}$$

where $\epsilon_k > 0$ is the tolerance and $\Gamma > 0$ is an upper bound for the augmented Lagrangian function $L_{\rho_k}(u^k, y^k, z^k)$. We summarize our numerical method in algorithm 1.

2.2. Proximal alternative shecme for solving subproblem (11) in algorithm 1

In algorithm 1, there is a subproblem (11) which has to be solved approximately, i.e. the conditions (10) hold. In this subsection, we aim at designing a numerical algorithm for solving

the subproblem (11) that satisfies the termination condition within finite iterations, and thus algorithm (1) is well defined. Let

$$P(u,y) = f(u) + \frac{\beta}{2}\|y\|_2^2 + \frac{\rho_k}{2}\|y - Wu + \hat{z}\|_2^2 \quad \text{and} \quad r(y) = \lambda\|y\|_0, \tag{12}$$

where $\hat{z} = z^k/\rho_k$. Then, the subproblem (11) at kth iteration is equivalent to

$$\min_{u,y} H(u,y) = P(u,y) + r(y). \tag{13}$$

The problem (13) is a typical two-block non-convex problem in which the non-convex term is separable. In recent years, the proximal alternating schemes [18, 22, 23] have been proposed and show the success in image processing and machine learning. Thus, we adopt this method for solving (13). More specifically, at kth step, we choose the initialization $(u^{k,0}, y^{k,0})$ as

$$(u^{k,0}, y^{k,0}) = \begin{cases} (u^0, y^0), & \text{if } L_{\rho_k}(u^{k-1}, y^{k-1}, z^k) \geqslant \Gamma, \\ (u^{k-1}, y^{k-1}), & \text{if } L_{\rho_k}(u^{k-1}, y^{k-1}, z^k) < \Gamma. \end{cases} \tag{14}$$

Since $Wu^0 = y^0$ and $\Gamma > L_{\rho_0}(u^0, y^0, z^0)$, we know

$$L_{\rho_k}(u^{k,0}, y^{k,0}, z^k) < \Gamma. \tag{15}$$

Once the initial point is chosen, the iterative scheme is given as follows.

- Update u by fixing y. Since f is convex and smooth with respect to u, we know $H(u,y)$ is convex and smooth with respect to u. To update $u^{k,j+1}$, we choose a surrogate function $\tilde{H}$ as

$$\tilde{H}(u, y^{k,j}) = \begin{cases} H(u, y^{k,j}), & \text{if } \nabla^2 f + \rho_k W^\top W \succ 0, \\ H(u, y^{k,j}) + \dfrac{\alpha_j}{2}\|u - u^{k,j}\|_2^2, & \text{otherwise}, \end{cases} \tag{16}$$

  where $\alpha_j > 0$. It is easy to know $\tilde{H}(\cdot, y^{k,j})$ is strongly convex and smooth, and thus admits a unique minimizer. We update $u^{k,j+1}$ by minimizing $\tilde{H}$ approximately, i.e. given a tolerance $\eta_j > 0$, we accept $u^{k,j+1}$ as long as

$$\|\nabla_u \tilde{H}(u^{k,j+1}, y^{k,j})\|_2 \leqslant \eta_j. \tag{17}$$

  As $\tilde{H}$ is smooth and strongly convex, the termination condition (17) holds in finite iterations when minimizing $\tilde{H}$.
- Update y by fixing u. Let $\mu_j > 0$, we adopt the proximal method to update $y^{k,j+1}$. Formally, we need to solve the minimization:

$$\min_y H(u^{k,j+1}, y) + \frac{\mu_j}{2}\|y - y^{k,j}\|_2^2. \tag{18}$$

  It is easy to know the above problem is equivalent to

$$\min_y \lambda\|y\|_0 + \left(\frac{\beta + \rho_k + \mu_j}{2}\right)\|y - \frac{\rho_k(Wu^{k,j+1} - \hat{z}) + \mu_j y^{k,j}}{\beta + \rho_k + \mu_j}\|_2^2, \tag{19}$$

  which has an analytical solution:

$$y^{k,j+1} = \mathcal{H}_{\sqrt{2\lambda/(\beta+\rho_k+\mu_j)}}\left(\frac{\rho_k(Wu^{k,j+1} - \hat{z}) + \mu_j y^{k,j}}{\beta + \rho_k + \mu_j}\right), \tag{20}$$

Algorithm 2. Proximal alternating scheme for solving (11).

Inputs: $\lambda, \beta, 0 < a < b$ and $\{\alpha_j, \mu_j\} \subset (a, b)$ and $\{\eta_j\} \subset \mathbb{R}_+$.
Initialization: Choose $(u^{k,0}, y^{k,0})$ according to (14).
while (10) *is not satisfied* **do**
  subproblem-u:
    Update $u^{k,j+1} \approx \arg\min_u \tilde{H}(u, y^{k,j})$ with $\|\nabla_u \tilde{H}(u^{k,j+1}, y^{k,j})\| \leqslant \eta_j$
  Subproblem-y:
    Update $y^{k,j+1}$ according to (20).
end
Return: $u^{k+1} = u^{k+1,0} = u^{k,N}, y^{k+1} = y^{k+1,0} = y^{k,N}$.

where $\mathcal{H}$ is the so-called hard thresholding operator, i.e. $\mathcal{H}_\lambda(x) = x$ if $|x| > \lambda$ and $\mathcal{H}_\lambda(x) = 0$ if $|x| \leqslant \lambda$.

In summary, the detailed proximal alternating scheme for solving subproblem (13) is present in algorithm 2.

2.3. Applications in image smoothing and CT imaging reconstruction

In this subsection, we apply algorithm 1 to image smoothing and CT imaging reconstruction. It is known that the biggest computational bottleneck is solving the subproblem-u in algorithm 2 as $f(u)$ is chosen from applications. We will show that in both image smoothing and CT reconstruction the subproblem-u can be solved efficiently.

Image smoothing. Image smoothing aims at capturing or highlighting the important structures while smoothing the unimportant contents. Two typical kinds of image smoothing tasks include texture removal and context smoothing. In this case, $u \in R^{M\times N\times 3}$, M and N present the size of image, the data fidelity f is chosen as $f(u) = \frac{1}{2}\|u - u_0\|_2^2$. The subproblem (16) is reduced to

$$\min_u \tilde{H}(u, y^{k,j}) = \frac{1}{2}\|u - u_0\|_2^2 + \frac{\rho_k}{2}\|Wu - y^{k,j} - \hat{z}\|_2^2, \tag{21}$$

as $\nabla^2 f(u) + \rho_k W^\top W = I + \rho_k W^\top W \succ 0$. The first order optimality of (21) implies $u^{k,j+1}$ satisfies

$$(I + \rho_k W^\top W)u^{k,j+1} = u_0 + \rho_k W^\top (y^{k,j} + \hat{z}). \tag{22}$$

Using the fast Fourier transform, the analytical form of $u^{k,j+1}$ is

$$u^{k,j+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(u_0) + \rho_k \mathcal{F}(W)^*(\mathcal{F}(\hat{z}) + \mathcal{F}(y^{k,j}))}{\mathcal{F}(\mathbf{1}) + \rho_k \mathcal{F}(W)^*\mathcal{F}(W)}\right). \tag{23}$$

CT imaging reconstruction. Medical CT images derive from x-ray radiation, which is harmful to absorb excess capacity. For better medical diagnosis, we have to develop algorithms that can reconstruct more accurate and clear images from sparse-view projection CT images. Once the projection operator A is fixed, we choose $f(u) = \frac{1}{2}\|Au - u_0\|_2^2$, $u \in R^{M\times N}$ and the subproblem (16) is

$$\min_u \tilde{H}(u, y^{k,j}) = \frac{1}{2}\|Au - u_0\|_2^2 + \frac{\rho_k}{2}\|Wu - y^{k,j} - \hat{z}\|_2^2 + \frac{\alpha_j}{2}\|u - u^{k,j}\|_2^2. \tag{24}$$

By the optimal condition of (24), we know $u^{k,j+1}$ satisfies

$$(A^\top A + \rho_k W^\top W + \alpha_j I)u^{k,j+1} = A^\top u_0 + \rho_k W^\top (y^{k,j} + \hat{z}) + \alpha_j u^{k,j}. \tag{25}$$

It is known that the conjugate gradient method can be applied for updating $u^{k,j+1}$ approximately such that the condition (17) holds.

3. Convergence analysis

This section firstly proves that the algorithm 1 is well defined, and then the generated sequence converges to a KKT point of (7). Finally, we show that the KKT point of (7) is a local minimum of (2) when W is a wavelet tight frame. Let $r_0 = \Gamma - L_{\rho_k}(u^{k,0}, y^{k,0}, z^k)$ and $\lambda_{\min}$ be the minimum eigenvalue of $\nabla^2 f(u) + \rho_k W^\top W$ for all u, we have $r_0 > 0$ from (14) and $\lambda_{\min} \geqslant 0$ from the convexity of f. Moreover, define the increments of the consecutive steps as

$$d_u^{k,j} = u^{k,j} - u^{k,j+1} \quad \text{and} \quad d_y^{k,j} = y^{k,j} - y^{k,j+1}. \tag{26}$$

Lemma 3.1. *Let $\{(u^{k,j}, y^{k,j})\}$ be the sequence generated by algorithm 2. Then, there exists some $0 < \tilde{c}_1 \leqslant c = \max(\lambda_{\min}, a)$ such that*

$$H(u^{k,j}, y^{k,j}) - H(u^{k,j+1}, y^{k,j+1}) \geqslant \frac{\tilde{c}_1}{2}\left(\|d_u^{k,j}\|_2 - \frac{\eta_j}{\tilde{c}_1}\right)^2 - \frac{\eta_j^2}{\tilde{c}_1} + \frac{a}{2}\|d_y^{k,j}\|_2^2, \tag{27}$$

where $0 < a \leqslant \inf\{\alpha_j, \mu_j | j \in \mathbb{N}\}$ as defined in algorithm 2.

Proof. By the mean value theorem, we have

$$H(v, y) \geqslant H(u, y) + \langle \nabla_u H(u, y), v - u\rangle + \frac{\lambda_{\min}}{2}\|v - u\|_2^2, \quad \forall\ u, v. \tag{28}$$

If $\lambda_{\min} > 0$, we know $u^{k,j+1}$ satisfies $\|\nabla_u H(u^{k,j+1}, y^{k,j})\|_2 \leqslant \eta_j$ as $\tilde{H}(u, y^{k,j}) = H(u, y^{k,j})$. Applying (28), we have

$$\begin{aligned} H(u^{k,j}, y^{k,j}) &\geqslant H(u^{k,j+1}, y^{k,j}) + \langle \nabla_u H(u^{k,j+1}), d_u^{k,j}\rangle + \frac{\lambda_{\min}}{2}\|d_u^{k,j}\|_2^2 \\ &\geqslant H(u^{k,j+1}, y^{k,j}) - \eta_j\|d_u^{k,j}\|_2 + \frac{\lambda_{\min}}{2}\|d_u^{k,j}\|_2^2 \\ &= H(u^{k,j+1}, y^{k,j}) + \frac{\lambda_{\min}}{2}\left(\|d_u^{k,j}\|_2 - \frac{\eta_j}{\lambda_{\min}}\right)^2 - \frac{\eta_j^2}{\lambda_{\min}}. \end{aligned}$$

If $\lambda_{\min} = 0$, we know $u^{k,j+1}$ satisfies $\|\nabla_u \tilde{H}(u^{k,j+1}, y^{k,j})\|_2 \leqslant \eta_j$ and $\nabla^2 \tilde{H} \succeq \alpha_j I$. Similarly, we have

$$\begin{aligned} H(u^{k,j}, y^{k,j}) = \tilde{H}(u^{k,j}, y^{k,j}) &\geqslant \tilde{H}(u^{k,j+1}, y^{k,j}) - \eta_j\|d_u^{k,j}\|_2 + \frac{\alpha_j}{2}\|d_u^{k,j}\|_2^2 \\ &\geqslant H(u^{k,j+1}, y^{k,j}) + \frac{\alpha_j}{2}\left(\|d_u^{k,j}\|_2 - \frac{\eta_j}{\alpha_j}\right)^2 - \frac{\eta_j^2}{\alpha_j}. \end{aligned}$$

Since $\alpha_j \geqslant a$, we choose $\tilde{c}_1 = a$ if $\lambda_{\min} = 0$ and $\tilde{c}_1 = \lambda_{\min}$ if $\lambda_{\min} > 0$. Then, $0 < \tilde{c}_1 \leqslant c$ and the above two inequalities imply

$$H(u^{k,j}, y^{k,j}) - H(u^{k,j+1}, y^{k,j}) \geqslant \frac{\tilde{c}_1}{2}\left(\|u^{k,j} - u^{k,j+1}\|_2 - \frac{\eta_j}{\tilde{c}_1}\right)^2 - \frac{\eta_j^2}{\tilde{c}_1}. \tag{29}$$

By the optimality condition of $y^{k,j+1}$, we have

$$H(u^{k,j+1}, y^{k,j}) - H(u^{k,j+1}, y^{k,j+1}) \geqslant \frac{\mu_j}{2}\|d_y^{k,j}\|_2^2 \geqslant \frac{a}{2}\|d_y^{k,j}\|_2^2, \tag{30}$$

as $\mu_j \geqslant a$. Sum (29) and (30), we obtain (27). □

Telescoping the inequality (27), it has

$$H(u^{k,0}, y^{k,0}) - H(u^{k,j}, y^{k,j}) \geqslant \sum_{i=0}^{j-1} \frac{\tilde{c}_1}{2}\left(\|d_u^{k,i}\|_2 - \frac{\eta_i}{\tilde{c}_1}\right)^2 - \frac{\eta_i^2}{\tilde{c}_1} + \frac{a}{2}\|d_y^{k,i}\|_2^2, \tag{31}$$

for all $j \in \mathbb{N}$. Since f is lower bounded, we know H is lower bounded by $\bar{H}$. If $\sum_{i=0}^{\infty}\eta_i^2 < \infty$, it has

$$\infty > H(u^{k,0}, y^{k,0}) - \bar{H} \geqslant \sum_{i=0}^{\infty} \frac{\tilde{c}_1}{2}\left(\|d_u^{k,i}\|_2 - \frac{\eta_i}{\tilde{c}_1}\right)^2 + \frac{a}{2}\|d_y^{k,i}\|_2^2 - \sum_{i=1}^{\infty}\frac{\eta_i^2}{\tilde{c}_1}. \tag{32}$$

Thus, we have

$$\|d_u^{k,i}\|_2 - \eta_i/\tilde{c}_1 \to 0 \quad \text{and} \quad \|d_y^{k,i}\|_2 \to 0, \quad \text{as } i \to \infty. \tag{33}$$

Lemma 3.2. *Let $\{(u^{k,j}, y^{k,j})\}$ be the sequence generated by algorithm 2. Then, there exist $\tilde{c}_2 > 0$ such that*

$$\operatorname{dist}(0, \partial H(u^{k,j+1}, y^{k,j+1})) \leqslant \tilde{c}_2(\|d_u^{k,j}\|_2 + \|d_y^{k,j}\|_2) + \eta_j \tag{34}$$

Proof. By elementary calculus, the subdifferential in (13) is

$$\partial H(u, y) = \begin{pmatrix} \nabla_u H(u, y) \\ \partial_y H(u, y) \end{pmatrix} = \begin{pmatrix} \nabla f(u) + \rho_k W^\top(Wu - y + \hat{z}) \\ \rho_k(y - Wu - \hat{z}) + \beta y + \lambda\partial\|y\|_0 \end{pmatrix}. \tag{35}$$

Applying (35), we know

$$\begin{aligned} &\|\nabla_u H(u^{k,j+1}, y^{k,j+1})\|_2 \\ &= \|(\nabla_u H(u^{k,j+1}, y^{k,j+1}) - \nabla_u H(u^{k,j+1}, y^{k,j})) + \nabla_u H(u^{k,j+1}, y^{k,j})\|_2 \\ &\leqslant \|\rho_k W^\top(y^{k,j} - y^{k,j+1})\|_2 + \|\nabla_u H(u^{k,j+1}, y^{k,j})\|_2. \end{aligned} \tag{36}$$

If $\lambda_{\min} > 0$, it has $\|\nabla_u H(u^{k,j+1}, y^{k,j})\|_2 \leqslant \eta_j$ from (17). If $\lambda_{\min} = 0$, it has

$$\begin{aligned} \eta_j \geqslant \|\nabla_u \tilde{H}(u^{k,j+1}, y^{k,j})\|_2 &= \|\nabla_u H(u^{k,j+1}, y^{k,j}) + \alpha_j(u^{k,j+1} - u^{k,j})\|_2 \\ &\geqslant \|\nabla_u H(u^{k,j+1}, y^{k,j})\|_2 - \alpha_j\|d_u^{k,j}\|_2. \end{aligned} \tag{37}$$

Let $\tilde{\rho}_k = \rho_k\lambda_W$ and λ_W be the maximum singular value of W, (36) and (37) imply

$$\|\nabla_u H(u^{k,j+1}, y^{k,j+1})\|_2 \leqslant \tilde{\rho}_k\|d_y^{k,j}\|_2 + b\|d_u^{k,j}\|_2 + \eta_j, \tag{38}$$

as $\alpha_j \leqslant b$. Furthermore, from the optimality condition of (18), we know

$$0 \in \rho_k(y^{k,j+1} - Wu^{k,j+1} - \hat{z}) + \beta y^{k,j+1} + \lambda\partial\|y^{k,j+1}\|_0 + \mu_j(y^{k,j+1} - y^{k,j}). \tag{39}$$

Combing (39) with (35), we know $\mu_j d_y^{k,j} \in \partial_y H(u^{k,j+1}, y^{k,j+1})$, and thus

$$\begin{aligned}\operatorname{dist}(0, \partial H(u^{k,j+1}, y^{k,j+1})) &\leqslant \|\nabla_u H(u^{k,j+1}, y^{k,j+1})\|_2 + \mu_j \|d_y^{k,j}\|_2 \\ &\leqslant (\tilde{\rho}_k + b)\|d_y^{k,j}\|_2 + b\|d_u^{k,j}\|_2 + \eta_j,\end{aligned}$$

as $\mu_j \leqslant b$. By choosing $\tilde{c}_2 = \tilde{\rho}_k + b$, the inequality (34) holds. □

Theorem 3.3. *Let $\{(u^{k,j}, y^{k,j})\}$ be the sequence generated by algorithm 2. Suppose the tolerance $\{\eta_j\}$ satisfies $\sum_{j=1}^{\infty} \eta_j^2 < cr_0$ where $c = \max(\lambda_{\min}, a)$. Then, for any $\epsilon_k > 0$, there exists a finite $N_k \in \mathbb{N}$ such that*

$$\operatorname{dist}(0, \partial H(u^{k,N_k}, y^{k,N_k})) \leqslant \epsilon_k \quad \text{and} \quad H(u^{k,N_k}, y^{k,N_k}) \leqslant \Gamma. \tag{40}$$

Proof. From (31), it has

$$H(u^{k,j}, y^{k,j}) \leqslant H(u^{k,0}, y^{k,0}) + \sum_{i=0}^{j-1} \frac{\eta_i^2}{\tilde{c}_1} \leqslant H(u^{k,0}, y^{k,0}) + r_0 \leqslant \Gamma \tag{41}$$

as $H(u^{k,0}, y^{k,0}) = L_{\rho_k}(u^{k,0}, y^{k,0}, z^k)$. Combining (33) with (34), it has

$$\begin{aligned}&\limsup_{j\to\infty} \operatorname{dist}(0, \partial H(u^{k,j+1}, y^{k,j+1})) \\ &\leqslant \lim_{j\to\infty} \tilde{c}_2(\|d_u^{k,j}\|_2 + \|d_y^{k,j}\|_2) + \eta_j = 0,\end{aligned} \tag{42}$$

which implies (40) holds. □

Since $H(u^{k,j}, y^{k,j}) = L_{\rho_k}(u^{k,j}, y^{k,j}, z^k)$, the termination condition (17) is met in finite iterations, and thus the algorithm 1 is well defined.

In the next, we will prove the convergence property of algorithm 1. We firstly write the KKT point of (7). That is, a point (u, y) is called a KKT point of (7) if there exists z such that

$$0 = \nabla f(u) + W^\top z, \quad 0 \in \beta y - z + \lambda \partial \|y\|_0, \quad Wu - y = 0. \tag{43}$$

Define

$$\varphi_\rho(u, y, z) = f(u) + \frac{\beta}{2}\|y\|_2^2 + \frac{\|\rho(Wu - y) + z\|_2^2 - \|z\|_2^2}{2\rho}. \tag{44}$$

Then, we know $L_\rho(u, y, z) = \varphi_\rho(u, y, z) + r(y)$ where $r(y)$ is defined in (12). Motivated by the analysis in [21], the convergence results of algorithm 1 are established as follows.

Theorem 3.4. *Let $\{(u^k, y^k, z^k)\}$ be the sequence generated by algorithm 1 and assume $\lim_{k\to\infty} \epsilon_k \to 0$. Then, the following statements hold.*

(*a*) $\lim_{k\to\infty} Wu^k - y^k = 0$.

(*b*) *If $\{(u^k, y^k, z^k)\}$ is bounded, there exists an accumulation point (u^*, y^*) of $\{(u^k, y^k)\}$ such that (u^*, y^*) is a KKT point of (7).*

Before proving theorem 3.4, we first calculate the limiting subdifferential of $\lambda\|y\|_0$.

Lemma 3.5. *Let $r(y) = \lambda\|y\|_0$. Then, we have*

$$\partial r(y) = (\partial r_1(y_1), \ldots, \partial r_n(y_n))^\top, \tag{45}$$

where $\partial r_i(y_i) = \{0\}$ if $y_i \neq 0$ and $\partial r_i(y_i) = \mathbb{R}$ if $y_i = 0$.

Proof. As $r(y) = \sum_{i=1}^n r_i(y_i) = \sum_{i=1}^n |y_i|_0$, thus $\partial r(y) = \partial r_1(y_1) \times \partial r_2(y_2) \times \cdots \times \partial r_n(y_n)$. If $y_i \neq 0$, since $r_i(y_i) = 1$ for all $y_i \neq 0$, we know $\partial r_i(y_i) = \{0\}$. If $y_i = 0$, for any $u \in \mathbb{R}$, we have

$$\liminf_{x\to 0, x\neq 0} \frac{r_i(x) - r_i(y_i) - u(x - y_i)}{|x|} = \liminf_{x\to 0, x\neq 0} \frac{1 - ux}{|x|} \geqslant 0. \tag{46}$$

Thus, $\partial r_i(0) = \hat{\partial} r_i(0) = \mathbb{R}$. □

Now, we are ready to prove theorem 3.4.

Proof. Let $s^k = Wu^k - y^k$. We prove (i) by considering two cases of $\{\rho_k\}$.

(a) $\{\rho_k\}$ is bounded. By the update scheme of $\{\rho_k\}$ in algorithm 1, there exists some k_0 such that for all $k \geqslant k_0$, it has $\|s^k\|_2 \leqslant \eta\|s^{k-1}\|_2$. Since $\eta \in (0, 1)$, we know $s^k \to 0$ as $k \to \infty$.

(b) $\{\rho_k\}$ is unbounded. By the update scheme of $\{\rho_k\}$, $\{\rho_k\}$ is updated by infinite times. Let $\{\rho_{j_1}, \rho_{j_2}, \ldots\}$ denote all elements in $\{\rho_k\}$ and $\mathcal{J} = \{j_1, j_2, \ldots, \}$ be arranged in an increasing order. Then, it has $\{\rho_{j_\ell}\} \to \infty$ and

$$\rho_i = \rho_{j_\ell}, \quad \forall j_\ell \leqslant i < j_{\ell+1}, \quad \ell \geqslant 1, \tag{47}$$

$$\rho_{j_\ell} = \max(\gamma\rho_{j_\ell - 1}, \|z^{j_\ell}\|_2^{1+\tau}). \tag{48}$$

Define $\underline{j}(k) = \max\{j \in \mathcal{J} | k \geqslant j\}$ for all $k \geqslant j_1$. By the update scheme of dual variable $\{z^k\}$, we know

$$z^k = z^{k-1} + \rho_{k-1}s^{k-1} = z^{\underline{j}(k)} + \rho_{\underline{j}(k)} \sum_{i=0}^{k-\underline{j}(k)-1} s^{\underline{j}(k)+i}. \tag{49}$$

Moreover, we have $\|s^{\underline{j}(k)+i+1}\|_2 \leqslant \eta\|s^{\underline{j}(k)+i}\|_2$ for all $0 \leqslant i < k - \underline{j}(k) - 1$. Together with (49) and the fact that $\{\rho_k\}$ is increasing, it has

$$\frac{\|z^k\|_2}{\rho_k} \leqslant \frac{\|z^{\underline{j}(k)}\|_2}{\rho_{\underline{j}(k)}} + \sum_{i=0}^{k-\underline{j}(k)-1} \eta^i \|s^{\underline{j}(k)}\|_2 \leqslant \frac{\|z^{\underline{j}(k)}\|_2}{\rho_{\underline{j}(k)}} + \frac{1}{1-\eta}\|s^{\underline{j}(k)}\|_2. \tag{50}$$

Since $\rho_{\underline{j}(k)} = \max(\gamma\rho_{\underline{j}(k)-1}, \|z^{\underline{j}(k)}\|_2^{1+\tau})$, it is easy to know $\|z^{\underline{j}(k)}\|_2/\rho_{\underline{j}(k)} \leqslant \rho_{\underline{j}(k)}^{-\tau/(1+\tau)}$ and it implies

$$\frac{\|z^{\underline{j}(k)}\|_2}{\rho_{\underline{j}(k)}} \to 0, \quad \text{as } k \to \infty. \tag{51}$$

Meanwhile, by the termination condition (10), it has $\varphi_{\rho_{\underline{j}(k)}}(u^{\underline{j}(k)}, y^{\underline{j}(k)}, z^{\underline{j}(k)}) + r(y^{\underline{j}(k)}) < \Gamma$ which implies

$$\left\| s^{\underline{j}(k)} + \frac{z^{\underline{j}(k)}}{\rho_{\underline{j}(k)}} \right\|_2^2 \leqslant \frac{2}{\rho_{\underline{j}(k)}} \left(\Gamma - f(u^{\underline{j}(k)}) - \frac{\beta}{2}\|y^{\underline{j}(k)}\|_2^2 - r(y^{\underline{j}(k)}) \right) + \frac{\|z^{\underline{j}(k)}\|_2^2}{\rho_{\underline{j}(k)}^2}. \tag{52}$$

Let $k \to \infty$ in (52), we know $\|s^{j(k)}\|_2 \to 0$ as $f(u)$ is bounded below, $\{\rho_{j(k)}\} \to \infty$ and (51). Thus, let $k \to \infty$ in (51), it has

$$\frac{\|z^k\|_2}{\rho_k} \to 0, \quad \text{as } k \to \infty. \tag{53}$$

Similarly, as $\varphi_{\rho_k}(u^k, y^k, z^k) + r(y^k) < \Gamma$, it has

$$\left\| s^k + \frac{z^k}{\rho_k} \right\|_2^2 \leqslant \frac{2}{\rho_k} \left(\Gamma - f(u^k) - \frac{\beta}{2}\|y^k\|_2^2 - r(y^k) \right) + \frac{\|z^k\|_2^2}{\rho_k^2}. \tag{54}$$

As f is bounded below and $\{\rho_k\} \to \infty$, (54) implies $s^k \to 0$ as $k \to \infty$.

Let $\{(u^k, y^k)\}_{k\in\mathcal{K}}$ be a convergent subsequence of $\{(u^k, y^k)\}$ and assume that $\{(u^k, y^k)\}_{k\in\mathcal{K}} \to \{(u^*, y^*)\}$. By statement (i), we have $s^* = Wu^* - y^* = 0$. Moreover, from the termination condition (10), we know there exists $\xi^k = (\xi_u^k, \xi_y^k)^\top$ where

$$\xi_u^k = \nabla f(u^k) + W^\top z^k + \rho_k W^\top (Wu^k - y^k) = \nabla f(u^k) + W^\top z^{k+1}, \tag{55}$$

$$\xi_y^k \in \beta y^k - z^k + \lambda \partial \|y^k\|_0 + \rho_k (y^k - Wu^k) = \beta y^k + \lambda \partial \|y^k\|_0 - z^{k+1}, \tag{56}$$

such that $\|\xi^k\|_2 \leqslant \epsilon_k$. Since $\{z^k\}$ is bounded, there exists a subsequence $\mathcal{K}_1 \subset \mathcal{K}$ such that $\{z^{k+1}\}_{\mathcal{K}_1}$ converges to some z^*. Denote $\mathcal{I} = \{i | y_i^* \neq 0\}$ and $\mathcal{I}^c = \{i | y_i^* = 0\}$. As $\{y^k\}_{\mathcal{K}_1} \to y^*$, there exists some k_0 such that $y_i^k \neq 0$ for all $k_0 \leqslant k \in \mathcal{K}_1$ and $i \in \mathcal{I}$. Let $y_\mathcal{I} = \{y_i | i \in \mathcal{I}\}$, as $\|y\|_0 = \sum_{i=1}^n |y_i|_0$, then ξ_y^k can be written as $\xi_y^k = (\xi_{y_\mathcal{I}}^k, \xi_{y_{\mathcal{I}^c}}^k)$

$$\xi_{y_\mathcal{I}}^k = \beta y_\mathcal{I}^k - z_\mathcal{I}^k + \lambda \partial \|y_\mathcal{I}^k\|_0 = \beta y_\mathcal{I}^k - z_\mathcal{I}^k, \tag{57}$$

$$\xi_{y_{\mathcal{I}^c}}^k \in \beta y_{\mathcal{I}^c}^k - z_{\mathcal{I}^c}^k + \lambda \partial \|y_{\mathcal{I}^c}^k\|_0. \tag{58}$$

Since $\{\epsilon_k\} \to 0$, it has $\{\xi_u^k, \xi_{y_\mathcal{I}}^k\}_{\mathcal{K}_1} \to 0$. Then, let $k \to \infty$ and $k \in \mathcal{K}_1$ in (23) and (57), we know

$$\nabla f(u^*) + W^\top z^* = 0 \quad \text{and} \quad \beta y_\mathcal{I}^* - z_\mathcal{I}^* = 0. \tag{59}$$

For all $i \in \mathcal{I}^c$, by lemma 3.5, we know $z_i^* \in \lambda \partial \|y_i^*\|_0$ which implies $z_{\mathcal{I}^c}^* \in \lambda \partial \|y_{\mathcal{I}^c}^*\|_0$. Together with (59) and the fact $s^* = Wu^* - y^* = 0$, we conclude that (u^*, y^*) satisfies (43) and thus is a KKT point of (7). □

Remark 1. We make the following remarks for theorem 3.4.

- The boundedness requirement of $\{z^k\}$ can be relaxed to $\{z_{\mathcal{I}^c}^k\}$ as $\|\beta y_\mathcal{I}^k - z_\mathcal{I}^k\|_2 \to 0$ and $y_\mathcal{I}^k \to y_\mathcal{I}^*$ as $k \to \infty$ and $k \in \mathcal{K}_1$.
- Let $W = (W_\mathcal{I}^\top, W_{\mathcal{I}^c}^\top)^\top$, the system in (59) can be reduced to

$$W_{\mathcal{I}^c}^\top z_{\mathcal{I}^c}^* = -\nabla f(u^*) - \beta y_\mathcal{I}^*. \tag{60}$$

  Since $y_{\mathcal{I}^c} = 0$, $z_{\mathcal{I}^c}^*$ can be arbitrary chosen as long as the system (60) is satisfied. It is known that the system (60) has at least one solution if $\nabla f(u^*) + \beta y_\mathcal{I}^*$ lies in the range space of $W_{\mathcal{I}^c}^\top$ which implies (u^*, y^*) satisfies the KKT equation (43). One sufficient condition is $W_{\mathcal{I}^c}^\top$ has full of column rank.

In the next, we will show that the KKT point (u^*, y^*) is a local minimizer of (2).

Theorem 3.6. *Let $\{(u^k, y^k, z^k)\}$ be the sequence generated by algorithm 1 and W be a wavelet tight frame and assume $\lim_{k\to\infty} \epsilon_k \to 0$. If $\{u^k, y^k, z^k\}$ is bounded, then there exists an accumulation point u^* of $\{u^k\}$ such that u^* is a local minimizer of (2).*

Proof. Let (u^*, y^*, z^*) be an accumulation point of $\{(u^k, y^k, z^k)\}$. By theorem 3.4, we know (u^*, y^*, z^*) satisfies the KKT system (43). Moreover, as W is a wavelet tight frame, we have $W^\top W = I$, then the KKT system implies

$$0 \in \beta u^* + \nabla f(u^*) + \lambda W^\top \partial \|y^*\|_0 \quad \text{and} \quad Wu^* - y^* = 0 \tag{61}$$

Define $\mathcal{I} = \{i | y_i^* \neq 0\}$ and $W = (W_{\mathcal{I}}^\top, W_{\mathcal{I}^c}^\top)^\top$. Lemma 3.5 implies

$$W^\top \partial \|y^*\|_0 = \{W_{\mathcal{I}^c}^\top \nu | \nu \in \mathbb{R}^{|\mathcal{I}_u^c|}\}, \tag{62}$$

and $W_{\mathcal{I}} u^* \neq 0$ and $W_{\mathcal{I}^c} u^* = 0$. We will prove that $W^\top \partial \|y^*\|_0 \subseteq \hat{\partial}(\|Wu^*\|_0)$. Since $W_{\mathcal{I}} u^* \neq 0$, there exists a neighborhood $\mathcal{N}$ of u^* such that $W_{\mathcal{I}} u \neq 0$ for all $u \in \mathcal{N}$. Define $\mathcal{S} = \{u | W_{\mathcal{I}^c} u = 0\}$, we have $\|Wu\|_0 = \|Wu^*\|_0$ if $u \in \mathcal{N} \cap \mathcal{S}$ and $\|Wu\|_0 \geqslant \|Wu^*\|_0 + 1$ if $u \in \mathcal{N} \backslash \mathcal{S}$. Therefore, for all $v \in W^\top \partial \|y^*\|_0$, we have

$$\begin{aligned} &\liminf_{y \to u^*, y \neq u^*} \frac{\|Wy\|_0 - \|Wu^*\|_0 - \langle v, y - u^* \rangle}{\|y - u^*\|_2} \\ &= \liminf_{y \to u^*, y \neq u^*, y \in \mathcal{S}} \frac{-\langle W_{\mathcal{I}^c}^\top \nu, y - u^* \rangle}{\|y - u^*\|_2} = 0, \end{aligned} \tag{63}$$

as $u^* \in \mathcal{S}$. This implies $W^\top \partial \|y^*\|_0 \subseteq \hat{\partial}(\|Wu^*\|_0) \subseteq \partial(\|Wu^*\|_0)$. By (61), we know

$$0 \in \beta u^* + \nabla f(u^*) + \lambda \partial(\|Wu^*\|_0), \tag{64}$$

which means u^* is a critical point of (2). Using theorem 3.1 in [24] extended from theorem 3.6 in [25], we have u^* is a local minimizer of (2). □

4. Numerical experiments

In this section, we evaluate the performance of our algorithm on image smoothing and CT imaging reconstruction. We compare our method with classical ℓ_2-norm, ℓ_e-norm [17], ℓ_0-norm [1] based models and the deep CNN [26]. Our results reveal that the proposed method can achieve higher sparsity and reduce more unnecessary details while keeping the main structure. All the experiments are implemented in MATLAB using a laptop equipped with a 2.6 GHz 2-core processor and 16GB memory.

4.1. Image smoothing

Image smoothing aims at smoothing the unwanted regions while keeping the salient structures, which is the basic process for salient object detection, image segmentation, image restoration, etc. In this part, we set W as the gradient operator ∇ and we test two kinds of image smoothing tasks: texture removal and context smoothing. All the test images come from [1, 26, 27]. The regularization term with ℓ_2-norm, ℓ_e-norm, and ℓ_0-norm are chosen as benchmarks. More precisely, we implement ℓ_2-norm and ℓ_0-norm model by setting $\lambda = 0$ and $\beta = 0$ respectively. For the ℓ_e-norm model, we also apply our algorithm to solve the convex model by replacing the ℓ_0-norm with ℓ_1-norm. Furthermore, the deep CNN based method [26] is also compared.

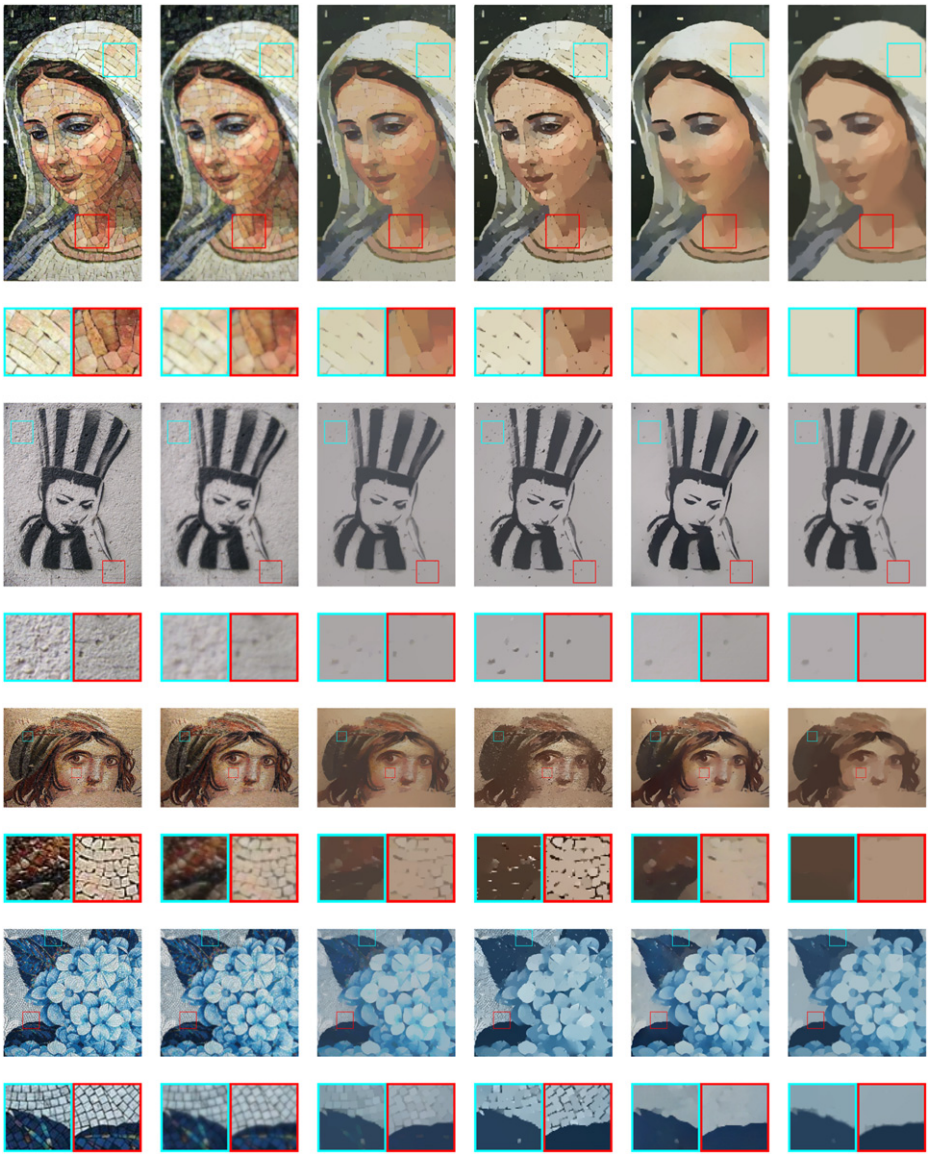

Figure 2. Performance on texture removal. From left to right: original images, results of ℓ_2-norm, ℓ_e-norm [17], ℓ_0-norm [1], CNN [26], and ours, respectively. Our method can remove unnecessary edges and is outstanding in texture removal.

In our model, parameters λ and β in our proposed model are used for leveraging the effect of ℓ_0 and ℓ_2-norm, where λ is critical to control the level of structure coarseness, β is a balanced parameter. Generally, we set $\lambda = 0.01$ and $\beta = \lambda$ in smoothing problem. Especially, we set $\beta = 10\lambda$ for the case of existing large scale of noise, and we set $\beta = 100\lambda$ for texture removal. Additionally, we use the KKT conditions or the following stopping criteria as termination $\|y^k - Wu^k\|_2 < 10^{-10}$.

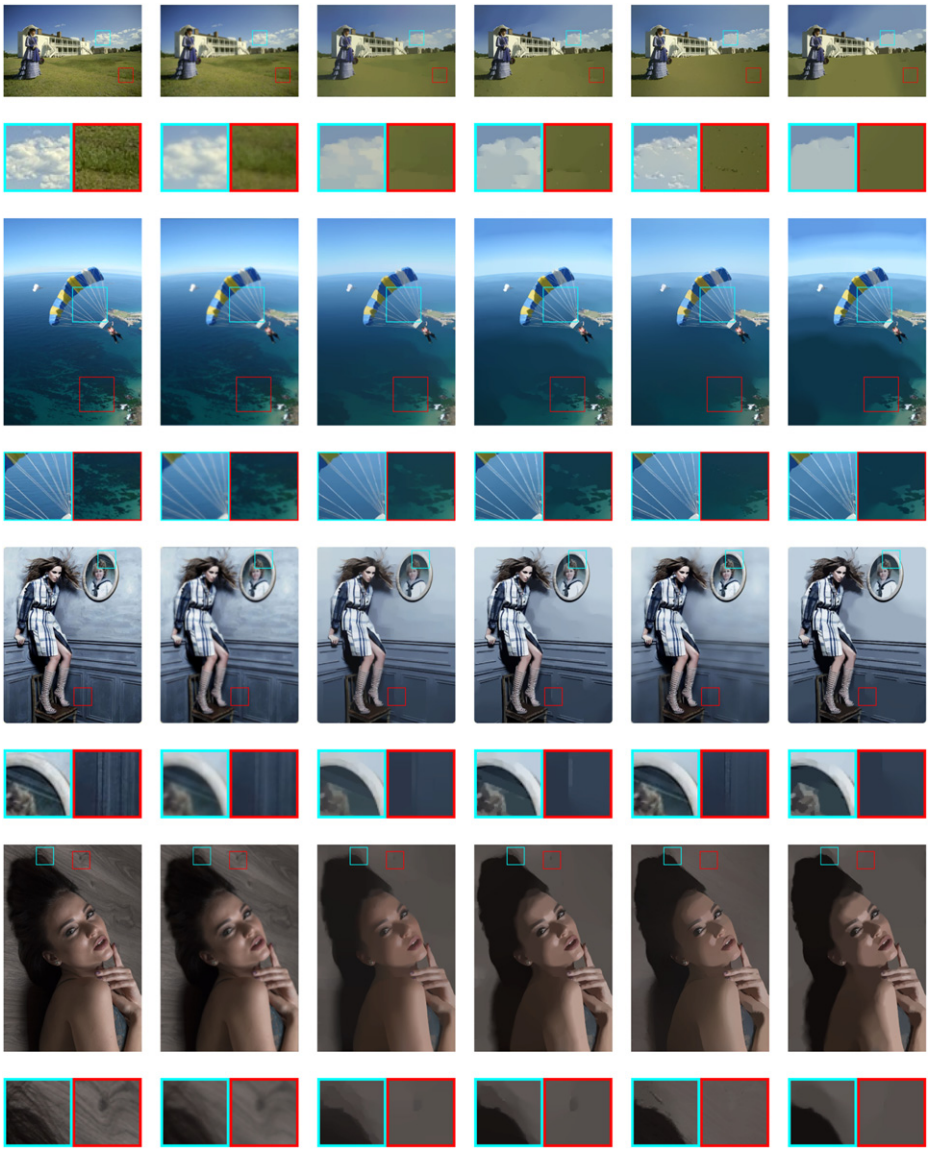

Figure 3. Performance on context smoothing. From left to right: original images, results of ℓ_2-norm, ℓ_e-norm [17], ℓ_0-norm [1], CNN [26] and ours, repectively. Our method is good at sparsity approximation while preserving the salient edges.

Figure 2 shows the results of texture removal. The ℓ_2-norm based model blurs the texture; the ℓ_e-norm based model dims the texture, but still looks hazy; the ℓ_0-norm based model removes most texture, but remains lots of unnecessary edges, and plenty of sharp edges and spots have remained. It is noticed that our proposed model is better than many other sparsity-based benchmarks at the situation of texture removal. In our model, the ℓ_2-norm is helpful to remove sharp gradients, and the ℓ_0-norm maintains sparsity. Moreover, our model is comparable to the deep CNN model, and even slightly better in some cases.

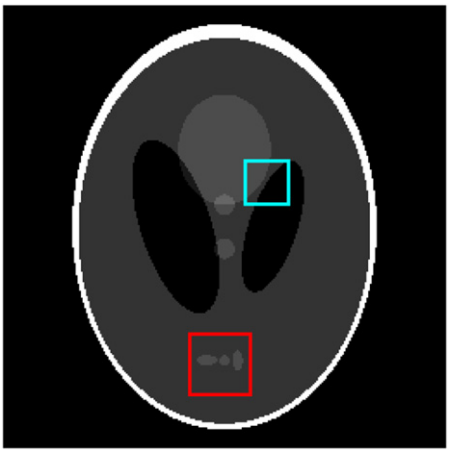

Figure 4. Original modified Shepp–Logan phantom with the marked ROIs.

Figure 3 shows the results of context smoothing. The ℓ_e-norm and the ℓ_0-norm based method, our method, and the CNN based methods all present higher sparsity than the ℓ_2-norm based images, yet differences remain. In the case of background smoothing, our results of the female model images demonstrate the advantages over the ℓ_e-norm based model. Compared to the ℓ_0-norm based model, our model removes more trivial outliers and sharp spots, especially the outliers and amount of noise in the zoom-in areas, e.g., the background of the last two images. In the case of edge-preserving, as the second image shows, our method obtains higher sparsity in the ocean while keeps the edge-preserving of parachute. Besides, our model achieves better visual results than CNNs. In summary, compared with the existing sparsity promising models, our proposed model can not only achieve higher sparsity, but also be able to remove large scale of noise, trivial outliers, spots and unnecessary edges. Moreover, our method is competitive with deep learning based method.

4.2. CT imaging from lower dose—sparse view reconstruction

In clinical practice of CT imaging, it is necessary to reduce the x-ray dose which is harmful to patients. One typical approach is to reconstruct the CT images from only a small set of projections, which is known as the sparse view reconstruction. In this section, both simulation and real data are tested. We first choose the modified Shepp–Logan phantom [28] as the simulated data, and then we choose a CT slice of a patient's chest as the clinical one, which comes from 'Low Dose CT Grand Challenge'[4]. In this section, we choose W as the Haar wavelet tight frame and compare it with canonical filtered back-projection (FBP) method and regularization based methods including ℓ_2-norm, elastic model(denoted as the ℓ_e-norm) and ℓ_0-norm. Similar as the case in image smoothing, the ℓ_2-norm and ℓ_e-norm based models are solved by algorithm 1, the ℓ_0-norm model is solved by the algorithm in [1]. From the reconstruction results, typical regions-of-interest (ROIs) are chosen for qualitative comparison. Also, the quantitative comparison for the reconstruction is measured by the peak signal-to-noise ratios (PSNR) value

[4] Which is provided by Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine, and supported by Grants EB017095 and EB017185 from the National Institute of Biomedical Imaging and Bioengineering.

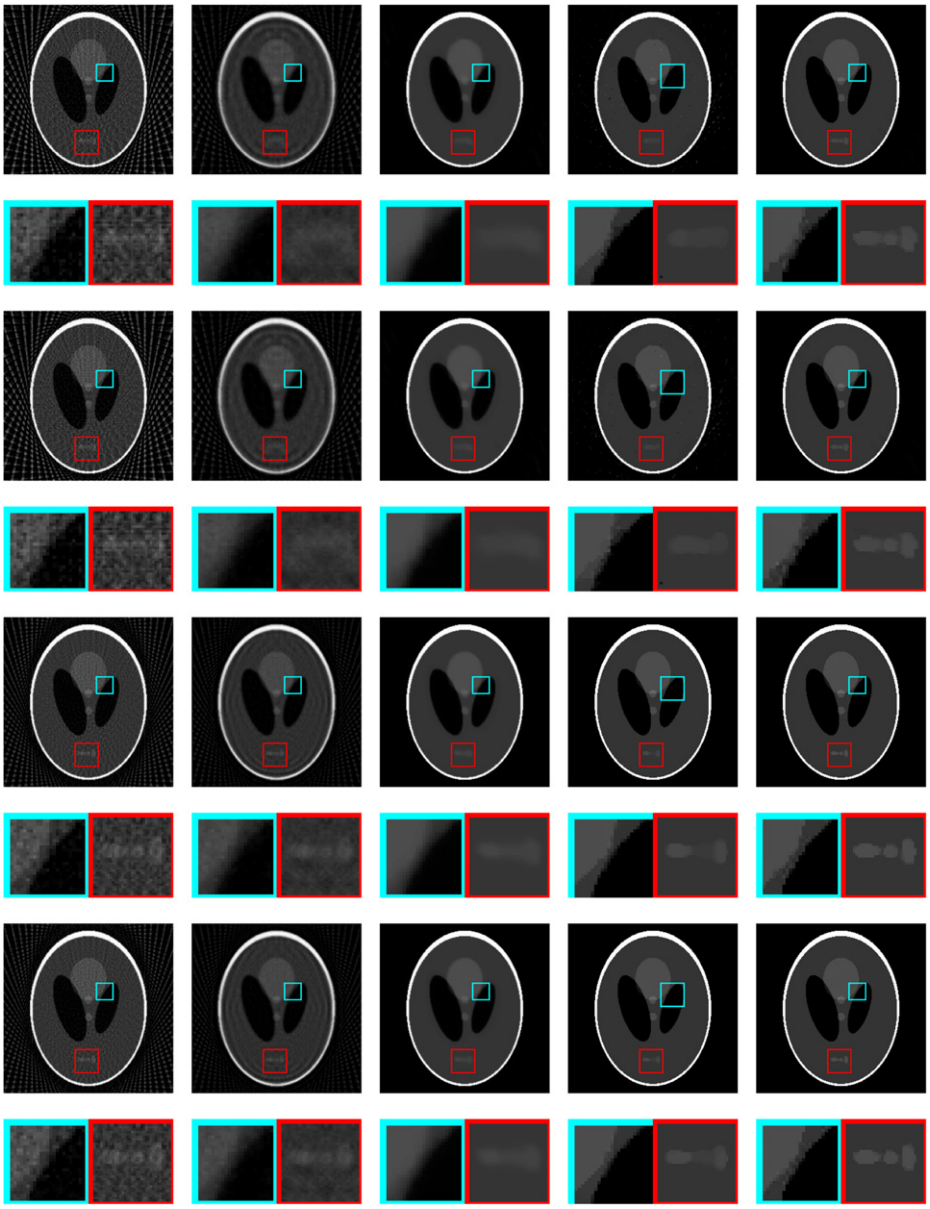

Figure 5. Performance on 'Shepp–Logan' images restoration. From top to bottom: 30 projections, 30 projections + noise, 45 projections and 45 projections + noise. From left to right: results of FBP, ℓ_2 norm, ℓ_e norm [17], ℓ_0 norm [1], and ours, respectively.

which is defined as:

$$\mathrm{PSNR}(u,\tilde{u}) = 10\,\log_{10}\left(\frac{MN\|u_{\max}-u_{\min}\|_2^2}{\|u-\tilde{u}\|_2^2}\right), \tag{65}$$

where $\tilde{u}$ denotes the truth image, $u_{\max}, u_{\min}$ denote the maximal and minimal pixel value of u, M and N are the sizes of the image.

Table 1. Reconstruction results on PSNR: our method obtains the highest PSNR values.

Images	FBP	ℓ_2	ℓ_e	ℓ_0	**Ours**
Proj.30	17.30	20.35	28.15	28.08	**33.42**
Proj.30n	17.29	20.35	28.11	28.00	**33.18**
Proj.45	20.62	21.48	30.92	36.38	**44.69**
Proj.45n	20.60	22.19	30.88	36.22	**44.07**

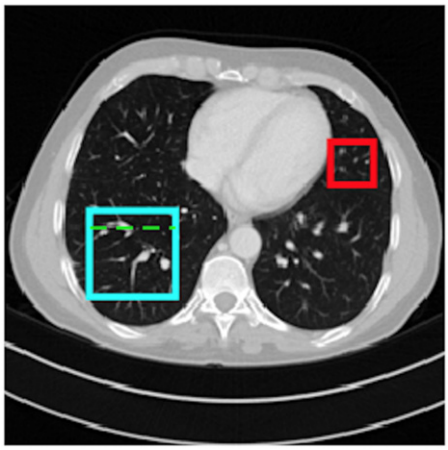

Figure 6. Original medical image with the marked ROI.

4.2.1. Simulated data. The transformation A in simulated data is the discretized version of radon transformation [29]. We considered 30 and 45 sparse projection views and set $\lambda = 0.0001, \beta = \lambda$ for the modified Shepp–Logan phantom. Additionally, we also consider the noisy measurement case for 30 projection and 45 projection views with 20% Gaussian noise. As shown in figure 4, two ROIs are selected for visual comparison.

The reconstruction results are shown in figure 5 and table 1. It is obvious that our method obtain better results than those from the ℓ_2-norm, ℓ_0-norm and ℓ_e-norm based models. More precisely, the FBP method leads to a bunch of trivial noise or spots, the ℓ_2 and ℓ_e-methods obtain blurring reconstructions, the ℓ_0-method generates obvious noisy artifacts from 30 projections, and over smooth features appear in 45 projections (e.g. the tiny dots in the red square). Moreover, for CT imaging from the noisy sparse projection views, our method also has the the best performance among all the methods in these numerical simulations.

4.2.2. Clinical data. We evaluate our method on CT images with the x-ray projection of sparsely sampled 15, 30, 45 and 60 views and compare them with the same baselines. We set $\lambda = 0.01, \beta = 20\lambda$ in our CT image restoration model. Except for comparing the visual results between original CT image and the reconstructed images, two ROIs (the big blue squareR_{left} and the small right red square R_{right}) and a profile line (green line) to compare the PSNR are selected for further demonstration as shown in figure 6.

Figure 7 shows the reconstruction results from ℓ_2, ℓ_1, ℓ_0-norm and our $\ell_0 + \ell_2$-norm. Models with ℓ_0-norm have higher sparsity than ℓ_2 and ℓ_e-norm. When comparing with the ℓ_0 method,

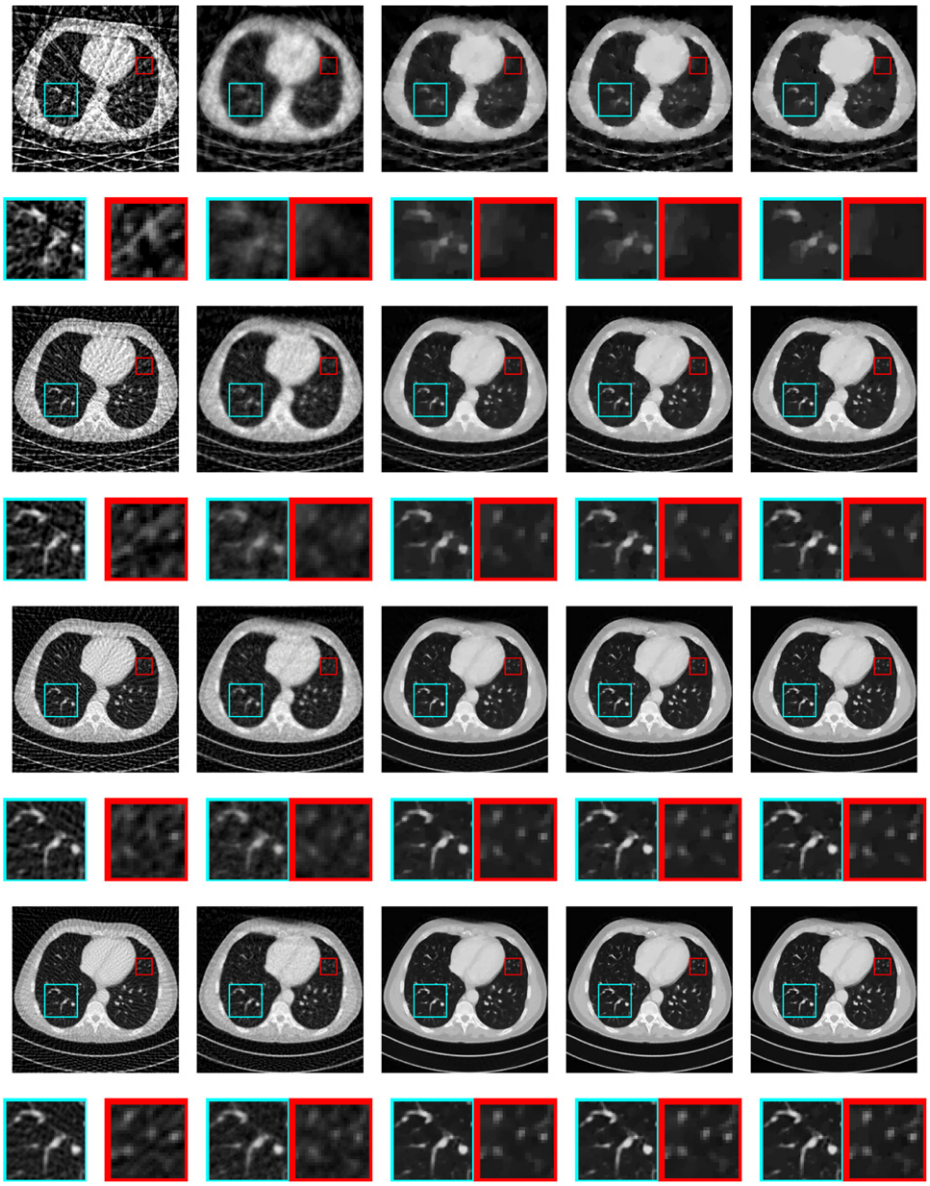

Figure 7. Performance on medical CT images restoration. From left to right: results of FBP method, ℓ_2 norm, ℓ_e norm [17], ℓ_0 norm [1], and ours, respectively.

the visual superiority of our method is not obvious. The zoom-in ROI in figure 7 shows that our method slightly outperforms the ℓ_0 regularized model. In order to see the particular priority, we compare the profile line to show the specific pixel values on each reconstructed images and ROIs. Figure 8 shows the original profile line and the reconstructed profile line of each method in each projection. It is obvious that the red lines (ours) are closer to the grey lines(truth) and are more flatten and sparser than others, especially in projection of 15 and 30.

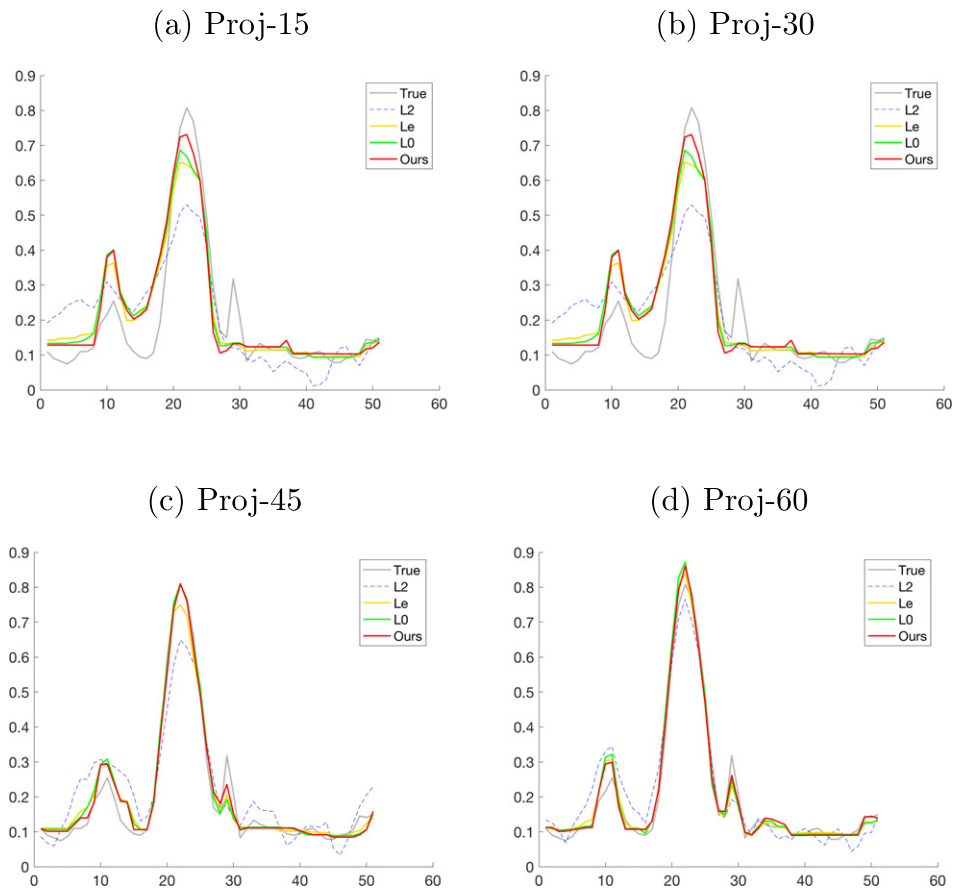

Figure 8. Performance on reconstruction of scanlines in projection of 15, 30, 45 and 60 views: the red lines (ours) obtain higher sparsity and are more close to the gray lines (groundtruth), especially in projection of 15 and 30 views.

Additionally, table 2 illustrates that our method obtains the highest PSNR value in every example. In particular, the PSNR values of our results are 0.1–1.0 dB larger than the results from the ℓ_0 regularized model. As a whole, figure 8 and table 2 verify that our model is applicable to CT reconstruction from lower dose and superior to other sparse regularization based methods.

5. Conclusion

In this paper, we propose a hybrid sparsity-based analysis model for image smoothing and CT imaging reconstruction. Thanks to the combination of ℓ_0-norm and the ℓ_2-norm, our model balances the sparsity and smoothness when representing the ground truth image. To solve the resulted non-separable ℓ_0-norm related minimization, an inexact augmented Lagrangian method is proposed with proved convergence to a local minimum. Moreover, our experiments

Table 2. Reconstruction results on PSNR: our method obtains the highest PSNR values in both global (full image) and details (region R_{left} and R_{right}).

Images	FBP	ℓ_2	ℓ_e	ℓ_0	**Ours**
Proj.15-full	13.05	18.52	20.85	20.97	**21.19**
Proj.15-R_{left}	16.15	18.58	19.94	20.17	**20.70**
Proj.15-R_{right}	16.87	20.66	23.26	23.24	**23.36**
Proj.30-full	17.79	21.34	25.50	25.83	**25.93**
Proj.30-R_{left}	21.34	20.60	24.73	25.30	**26.32**
Proj.30-R_{right}	21.90	24.23	27.17	27.31	**27.62**
Proj.45-full	21.17	23.74	29.80	30.22	**30.43**
Proj.45-R_{left}	24.86	22.96	29.01	29.91	**30.71**
Proj.45-R_{right}	24.53	26.07	30.51	30.99	**31.40**
Proj.60-full	23.78	25.43	32.28	32.45	**32.92**
Proj.60-R_{left}	27.06	25.59	33.00	33.43	**33.96**
Proj.60-R_{right}	27.67	27.76	33.75	34.66	**34.71**

validate the advantages of our proposed model when comparing with other sparsity-based models.

Funding

This work is partially supported by National Natural Sciences Foundation of China (No. 11901338 and No. 11801594), Tsinghua University Initiative Scientific Research Program and Guangdong-Hong Kong-Macau Applied Math Center grant 2020B1515310011.

ORCID iDs

Jiebo Song https://orcid.org/0000-0002-6124-2850
Chenglong Bao https://orcid.org/0000-0002-1201-1212

References

[1] Xu L, Lu C, Xu Y and Jia J 2011 Image smoothing via l_0 gradient minimization *ACM Trans. Graph.* **30** 174
[2] Basu M 2002 Gaussian-based edge-detection methods-a survey *IEEE Trans. Syst. Man Cybern.* C **32** 252–60
[3] DeCarlo D and Santella A 2002 Stylization and abstraction of photographs *ACM Trans. Graph.* **21** 769–76
[4] Rudin L I, Stanley O and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[5] Fan J and Li R 2001 Variable selection via nonconcave penalized likelihood and its oracle properties *J. Am. Stat. Assoc.* **96** 1348–60
[6] Zhang C-H 2010 Nearly unbiased variable selection under minimax concave penalty *Ann. Stat.* **38** 894–942
[7] Wu C, Liu Z and Wen S 2018 A general truncated regularization framework for contrast-preserving variational signal and image restoration: motivation and implementation *Sci. China Math.* **61** 1711–32

[8] Prakash J, Shaw C B, Manjappa R, Rajan K and Yalavarthy P K 2013 Sparse recovery methods hold promise for diffuse optical tomographic image reconstruction *IEEE J. Sel. Top. Quantum Electron.* **20** 74–82
[9] Han Y, Feng X-C and George B 2013 Variational and PCA based natural image segmentation *Pattern Recognit.* **46** 1971–84
[10] Faramarzi E, Rajan D and Christensen M P 2013 Unified blind method for multi-image super-resolution and single/multi-image blur deconvolution *IEEE Trans. Image Process.* **22** 2101–14
[11] Pan J, Hu Z, Su Z and Yang M-H 2016 l_0-regularized intensity and gradient prior for deblurring text images and beyond *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 342–55
[12] Xu L, Zheng S and Jia J 2013 Unnatural l_0 sparse representation for natural image deblurring *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1107–14
[13] Soo-Chang Pei S-C, Chih-Tsung Shen C-T and Tzu-Yen Lee T-Y 2012 Visual enhancement using constrained L0 gradient image decomposition for low backlight displays *IEEE Signal Process. Lett.* **19** 813–6
[14] Zhang Y, Dong B and Lu Z 2013 l_0 minimization for wavelet frame based image restoration *Math. Comput.* **82** 995–1015
[15] Yu W and Zeng L 2015 ℓ_0 gradient minimization based image reconstruction for limited-angle computed tomography *PloS One* **10** e0130793
[16] Bao C, Dong B, Hou L, Shen Z, Zhang X and Zhang X 2016 Image restoration by minimizing zero norm of wavelet frame coefficients *Inverse Problems* **32** 115004
[17] Zou H and Hastie T 2005 Regularization and variable selection via the elastic net *J. R. Stat. Soc.* B **67** 301–20
[18] Bolte J, Sabach S and Teboulle M 2014 Proximal alternating linearized minimization for nonconvex and nonsmooth problems *Math. Program.* **146** 459–94
[19] Lu Z 2014 Iterative hard thresholding methods for l_0 regularized convex cone programming *Math. Program.* **147** 125–54
[20] Dong B and Zhang Y 2013 An efficient algorithm for l_0 minimization in wavelet frame based image restoration *J. Sci. Comput.* **54** 350–68
[21] Chen X, Guo L, Lu Z and Ye J J 2017 An augmented Lagrangian method for non-Lipschitz nonconvex programming *SIAM J. Numer. Anal.* **55** 168–93
[22] Xu Y and Yin W 2013 A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion *SIAM J. Imaging Sci.* **6** 1758–89
[23] Bao C, Ji H, Quan Y and Shen Z 2016 Dictionary learning for sparse coding: algorithms and convergence analysis *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 1356–69
[24] Xue F and Wu C 2019 Every critical point of an l_0 composite minimization problem is a local minimizer (arXiv:1912.04498)
[25] Feng X, Wu C and Zeng C 2018 On the local and global minimizers of l_0 gradient regularized model with box constraints for image restoration *Inverse Problems* **34** 095007
[26] Fan Q, Yang J, Wipf D, Chen B and Tong X 2018 Image smoothing via unsupervised learning *ACM Trans. Graph.* **37** 2018
[27] Farbman Z, Fattal R, Lischinski D and Szeliski R 2008 Edge-preserving decompositions for multi-scale tone and detail manipulation *ACM Trans. Graph.* **27** 67
[28] Gach H M, Tanase C and Boada F 2008 2D and 3D Shepp–Logan phantom standards for MRI *19th International Conference on Systems Engineering* (IEEE) 521–6
[29] Lim J S 1990 *Two-Dimensional Signal and Image Processing* (Englewood Cliffs, NJ: Prentice-Hall) pp 42–5