ℓ_0 MINIMIZATION FOR WAVELET FRAME BASED IMAGE RESTORATION

YONG ZHANG, BIN DONG, AND ZHAOSONG LU

ABSTRACT. The theory of (tight) wavelet frames has been extensively studied in the past twenty years and they are currently widely used for image restoration and other image processing and analysis problems. The success of wavelet frame based models, including balanced approach and analysis based approach, is due to their capability of sparsely approximating piecewise smooth functions like images. Motivated by the balanced approach and analysis based approach, we shall propose a wavelet frame based ℓ_0 minimization model, where the ℓ_0 "norm" of the frame coefficients is penalized. We adapt the penalty decomposition (PD) method of Lu and Zhang to solve the proposed optimization problem. Some convergence analysis of the adapted PD method will also be provided. Numerical results showed that the proposed model solved by the PD method can generate images with better quality than those obtained by either analysis based approach or balanced approach in terms of restoring sharp features as well as maintaining smoothness of the recovered images.

1. INTRODUCTION

Mathematics has been playing an important role in the modern developments of image processing and analysis. Image restoration, including image denoising, deblurring, inpainting, tomography, etc., is one of the most important areas in image processing and analysis. Its major purpose is to enhance the quality of a given image that is corrupted in various ways during the process of imaging, acquisition and communication, and it enables us to see crucial but subtle objects residing in the image. Therefore, image restoration is an important step to take towards accurate interpretations of the physical world and making optimal decisions.

1.1. **Image restoration.** Image restoration is often formulated as a linear inverse problem. For the simplicity of notation, we denote the images as vectors in $\mathbb{R}^n$ with n equals to the total number of pixels. A typical image restoration problem is formulated as

$$(1.1) \qquad\qquad f = Au + \eta,$$

where $f \in \mathbb{R}^d$ is the observed image (or measurements), η denotes white Gaussian noise with variance σ^2, and $A \in \mathbb{R}^{d \times n}$ is some linear operator. The objective is to find the unknown true image $u \in \mathbb{R}^n$ from the observed image f. Typically,

the linear operator in (1.1) is a convolution operator for image deconvolution problems, a projection operator for image inpainting and partial Radon transform for computed tomography.

To solve u from (1.1), one of the most natural choices is the following least square problem

$$\min_{u \in \mathbb{R}^n} \|Au - f\|_2^2,$$

where $\| \cdot \|_2$ denotes the ℓ_2-norm. This is, however, not a good idea in general. Taking an image deconvolution problem as an example, since the matrix A is ill-conditioned, the noise η possessed by f will be amplified after solving the above least squares problem. Therefore, in order to suppress the effect of noise and also preserve key features of the image, e.g., edges, various regularization based optimization models were proposed in the literature. Among all regularization based models for image restoration, variational methods and wavelet frame based approaches are widely adopted and have been proven successful.

The trend of variational methods and partial differential equation (PDE) based image processing started with the refined Rudin-Osher-Fatemi (ROF) model [45] which penalizes the total variation (TV) of u. Many of the current PDE based methods for image denoising and decomposition utilize TV regularization for its beneficial edge preserving property (see e.g., [41, 46, 42]). The ROF model is especially effective on restoring images that are piecewise constant, e.g., binary images. Other types of variational models were also proposed after the ROF model. We refer the interested readers to [36, 17, 41, 42, 22, 2, 23, 53] and the references therein for more details.

Wavelet frame based approaches are relatively new and came from a different path. The basic idea for wavelet frame based approaches is that images can be sparsely approximated by properly designed wavelet frames, and hence, the regularization used for wavelet frame based models is the ℓ_1-norm of frame coefficients. Although wavelet frame based approaches take similar forms as variational methods, they were generally considered as different approaches than variational methods because, among many other reasons, wavelet frame based approaches are defined for discrete data, while variational methods assume all variables are functions. Some studies in the literature (see for example [51]) indicated that there was a relation between Haar wavelet and total variation. However, it was not clear if there exists a general relation between wavelet frames and variational models (with general differential operators) in the context of image restorations. In a recent paper [9], the authors established a rigorous connection between one of the wavelet frame based approaches, namely the analysis based approach and variational models. It was shown in [9] that the analysis based approach can be regarded as a finite difference approximation of a certain type of general variational model, and such approximation will be exact when image resolution goes to infinity. Furthermore, through Gamma-convergence, the authors showed that the solutions of the analysis based approach also approximate the solutions of the corresponding variational model. Such connections not only grant geometric interpretation to wavelet frame based approaches, but also lead to even wider applications of them, e.g., image segmentation [27] and 3D surface reconstruction from unorganized point sets [29]. On the other hand, the discretization provided by wavelet frames was shown, e.g., in [18, 20, 10, 11, 9, 28], to be superior to the standard discretizations for some of the variational models, due to the multiresolution structure and redundancy of wavelet

frames which enable wavelet frame based models to adaptively choose a proper differential operator in different regions of a given image according to the order of the singularity of the underlying solutions. For these reasons, as well as the fact that digital images are always discrete, we use wavelet frames as the tool for image restoration in this paper.

1.2. **Wavelet frame based approaches.** We now briefly introduce the concept of tight frames and tight wavelet frames, and then recall some of the frame based image restoration models. Interesting readers should consult [44, 24, 25] for theories of frames and wavelet frames; see [47] for a short survey on theory and applications of frames, and [28] for a more detailed survey.

A countable set $X \subset L_2(\mathbb{R})$ is called a tight frame of $L_2(\mathbb{R})$ if

$$f = \sum_{h \in X} \langle f, h \rangle h \quad \forall f \in L_2(\mathbb{R}),$$

where $\langle \cdot, \cdot \rangle$ is the inner product of $L_2(\mathbb{R})$. The tight frame X is called a tight wavelet frame if the elements of X are generated by dilations and translations of finitely many functions called framelets. The construction of framelets can be obtained by the unitary extension principle (UEP) of [44]. In our implementations, we will mainly use the piecewise linear B-spline framelets constructed by [44]. Given a 1-dimensional framelet system for $L_2(\mathbb{R})$, the s-dimensional tight wavelet frame system for $L_2(\mathbb{R}^s)$ can be easily constructed by using tensor products of 1-dimensional framelets (see, e.g., [24, 28]).

In the discrete setting, we will use $W \in \mathbb{R}^{m \times n}$ with $m \geq n$ to denote fast tensor product framelet decomposition and use $W^\top$ to denote the fast reconstruction. Then by the unitary extension principle [44], we have $W^\top W = I$, i.e., $u = W^\top W u$ for any image u. We will further denote an L-level framelet decomposition of u as

$$Wu = (\ldots, W_{l,j}u, \ldots)^\top \quad \text{for } 0 \leq l \leq L - 1, j \in \mathcal{I},$$

where $\mathcal{I}$ denotes the index set of all framelet bands and $W_{l,j}u \in \mathbb{R}^n$. Under such notation, we have $m = L \times |\mathcal{I}| \times n$. We will also use $\alpha \in \mathbb{R}^m$ to denote the frame coefficients, i.e., $\alpha = Wu$, where

$$\alpha = (\ldots, \alpha_{l,j}, \ldots)^\top, \quad \text{with } \alpha_{l,j} = W_{l,j}u.$$

More details on discrete algorithms of framelet transforms can be found in [28].

Since tight wavelet frame systems are redundant systems (i.e., $m > n$), the representation of u in the frame domain is not unique. Therefore, there are mainly three formulations utilizing the sparseness of the frame coefficients, namely, analysis based approach, synthesis based approach, and balanced approach. Detailed and integrated descriptions of these three methods can be found in [28].

The wavelet frame based image processing started from [18, 19] for high-resolution image reconstructions, where the proposed algorithm was later analyzed in [7]. These works led to the following balanced approach [8]

$$(1.2) \quad \min_{\alpha \in \mathbb{R}^m} \frac{1}{2}\|AW^\top \alpha - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)\alpha\|_2^2 + \left\| \sum_{l=0}^{L-1} \left(\sum_{j \in \mathcal{I}} \lambda_{l,j} |\alpha_{l,j}|^p \right)^{1/p} \right\|_1,$$

where $p = 1$ or 2, $0 \leq \kappa \leq \infty$, $\lambda_{l,j} \geq 0$ is a scalar parameter, and $\|\cdot\|_D$ denotes the weighted ℓ_2-norm with D positive definite. This formulation is referred to as

the balanced approach because it balances the sparsity of the frame coefficient and the smoothness of the image. The balanced approach (1.2) was applied to various applications in [16, 21, 48, 39].

When $\kappa = 0$, only the sparsity of the frame coefficient is penalized. This is called the synthesis based approach, as the image is synthesized by the sparsest coefficient vector (see e.g., [26, 32, 33, 34, 35]). When $\kappa = +\infty$, only the sparsity of canonical wavelet frame coefficients, which corresponds to the smoothness of the underlying image, is penalized. For this case, problem (1.2) can be rewritten as

$$(1.3) \qquad \min_{u \in \mathbb{R}^n} \frac{1}{2} \|Au - f\|_D^2 + \left\| \sum_{l=0}^{L-1} \left(\sum_{j \in \mathcal{I}} \lambda_{l,j} |W_{l,j} u|^p \right)^{1/p} \right\|_1.$$

This is called the analysis based approach, as the coefficient is in range of the analysis operator (see, for example, [11, 31, 50]).

Note that if we take $p = 1$ for the last term of (1.2) and (1.3), it is known as the anisotropic ℓ_1-norm of the frame coefficients, which is the case used for earlier frame based image restoration models. The case $p = 2$, called isotropic ℓ_1-norm of the frame coefficients, was proposed in [9] and was shown to be superior to the anisotropic ℓ_1-norm. Therefore, we will choose $p = 2$ for our simulations.

1.3. **Motivations and contributions.** For most of the variational models and wavelet frame based approaches, the choice of norm for the regularization term is the ℓ_1-norm. Taking wavelet frame based approaches for example, the attempt of minimizing the ℓ_1-norm of the frame coefficients is to increase their sparsity, which is the right thing to do since piecewise smooth functions like images can be sparsely approximated by tight wavelet frames. Although the ℓ_1-norm of a vector does not directly correspond to its cardinality in contrast to the ℓ_0 "norm", it can be regarded as a convex approximation to the ℓ_0 "norm". Such approximation is also an excellent approximation for many cases. It was shown by [12], which generalizes the exciting results of compressed sensing [13, 15, 14, 30], that for a given wavelet frame, if the operator A satisfies certain conditions, and if the unknown true image can be sparsely approximated by the given wavelet frame, one can robustly recover the unknown image by penalizing the ℓ_1-norm of the frame coefficients.

For image restoration, however, the conditions on A as required by [12] are not generally satisfied, which means penalizing the ℓ_0 "norm" and the ℓ_1-norm may produce different solutions. Although both the balanced approach (1.2) and analysis based approach (1.3) can generate restored images with very high quality, one natural question is whether using the ℓ_0 "norm" instead of the ℓ_1-norm can further improve the results.

On the other hand, it was observed, e.g., in [28] (also see Figure 3 and Figure 4), that the balanced approach (1.2) generally generates images with sharper features like edges than the analysis based approach (1.3), because balanced approach emphasizes more on the sparsity of the frame coefficients. However, the recovered images from a balanced approach usually contain more artifacts (e.g., oscillations) than an analysis based approach, because the regularization term of the analysis based approach has a direct link to the regularity of u (as proven by [9]) when compared to a balanced approach. Although such a trade-off can be controlled by the parameter κ in the balanced approach (1.2), it is not very easy to do in practice.

Furthermore, when a large κ is chosen, some of the numerical algorithms solving (1.2) will converge slower than when choosing a smaller κ (see e.g., [48, 28]).

Since penalizing ℓ_1-norm of Wu ensures smoothness while not as much sparsity as the balanced approach, we propose to penalize the ℓ_0 "norm" of Wu instead. Intuitively, this should provide us a balance between sharpness of the features and smoothness for the recovered images. The difficulty here is that ℓ_0 minimization problems are generally hard to solve. Recently, penalty decomposition (PD) methods were proposed by [40] for a general ℓ_0 minimization problem that can be used to solve our proposed model due to its generality. Computational results of [40] demonstrated that their methods generally outperform the existing methods for compressed sensing problems, sparse logistic regression and sparse inverse covariance selection problems in terms of quality of solutions and/or computational efficiency. This motivates us to adapt one of their PD methods to solve our proposed ℓ_0 minimization problem. As proposed in [40], the same block coordinate descent (BCD) method is used to solve each penalty subproblem of the PD method. However, the convergence analysis of the BCD method was missing from [40] when the ℓ_0 "norm" appears in the objective function. Indeed, the convergence of the BCD method generally requires the continuity of the objective function as discussed in [52]. In addition, the BCD method for the optimization problem with the nonconvex objective function has only been proved to converge to a stationary point which is not a local minimizer in general (see [52] for details).

Contributions. The main contributions of this paper are summarized as follows.

1) We propose a new wavelet frame based model for image restoration problems that penalizes the ℓ_0 "norm" of the wavelet frame coefficients. Numerical simulations show that the PD method that solves the proposed model generates recovered images with better quality than those obtained by either balanced approach and analysis based approach.

2) Given the discontinuity and nonconvexity of the ℓ_0 "norm" term in the objective function, we have proved some convergence results for the BCD method which is missing from the literature.

We now leave the details of the model and algorithm to Section 2 and details of simulations to Section 3.

2. Model and Algorithm

We start by introducing some simple notation. The space of symmetric $n \times n$ matrices will be denoted by $\mathcal{S}^n$. If $X \in \mathcal{S}^n$ is positive definite, we write $X \succ 0$. We denote by I the identity matrix, whose dimension should be clear from the context. Given an index set $J \subseteq \{1, \ldots, n\}$, x_J denotes the subvector formed by the entries of x indexed by J. For any real vector, $\|\cdot\|_0$ and $\|\cdot\|_2$, denote the cardinality (i.e., the number of nonzero entries) and the Euclidean norm of the vector, respectively. In addition, $\|x\|_D$ denotes the weighted ℓ_2-norm defined by $\|x\|_D = \sqrt{x^\top D x}$ with $D \succ 0$.

2.1. **Model.** We now propose the following optimization model for image restoration problems,

$$(2.1) \qquad \min_{u \in \mathcal{Y}} \frac{1}{2}\|Au - f\|_D^2 + \sum_i \lambda_i \|(Wu)_i\|_0,$$

where $\mathcal{Y}$ is some convex subset of $\mathbb{R}^n$. Here we are using the multi-index $\boldsymbol{i}$ and denote $(Wu)_{\boldsymbol{i}}$ (similarly for $\lambda_{\boldsymbol{i}}$) the value of Wu at a given pixel location within a certain level and band of wavelet frame transform. Comparing to the analysis based model, we are now penalizing the number of nonzero elements of Wu. As mentioned earlier that if we emphasize too much on the sparsity of the frame coefficients as in the balanced approach or synthesis based approach, the recovered image will contain artifacts, although features like edges will be sharp; if we emphasize too much on the regularity of u like in the analysis based approach, features in the recovered images will be slightly blurred, although artifacts and noise will be nicely suppressed. Therefore, by penalizing the ℓ_0 "norm" of Wu as in (2.1), we can indeed achieve a better balance between sharpness of features and smoothness of the recovered images.

Given that the ℓ_0 "norm" is an integer-valued, discontinuous and nonconvex function, problem (2.3) is generally hard to solve. Some algorithms proposed in the literature, e.g., iterative hard thresholding algorithms [5, 6, 38], cannot be directly applied to the proposed model (2.1) unless $W = I$. Recently, Lu and Zhang [40] proposed a penalty decomposition (PD) method to solve the following general ℓ_0 minimization problem,

$$(2.2) \qquad \min_{x \in \mathcal{X}} f(x) + \nu \|x_J\|_0$$

for some $\nu > 0$ controlling the sparsity of the solution, where $\mathcal{X}$ is a closed convex set in $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, and $\|x_J\|_0$ denotes the cardinality of the subvector formed by the entries of x indexed by J. In view of [40], we reformulate (2.1) as

$$(2.3) \qquad \min_{u \in \mathcal{Y}, \alpha = Wu} \frac{1}{2}\|Au - f\|_D^2 + \sum_{\boldsymbol{i}} \lambda_{\boldsymbol{i}} \|\alpha_{\boldsymbol{i}}\|_0$$

and then we can adapt the PD method of [40] to tackle problem (2.1) directly. As proposed in [40], the same BCD method is used to solve each penalty subproblem of the PD method. In addition, we apply the nonmonotone gradient projection method proposed in [4] to solve one of the subproblems in the BCD method.

2.2. Algorithm for problem (2.3).

In this section, we discuss how the PD method proposed in [40] solving (2.2) can be adapted to solve problem (2.3). Letting $x = (u_1, \ldots, u_n, \alpha_1, \ldots, \alpha_m)$, $J = \{n+1, \ldots, n+m\}$, $\bar{J} = \{1, \ldots, n\}$, $f(x) = \frac{1}{2}\|Ax_{\bar{J}} - f\|_D^2$ and $\mathcal{X} = \{x \in \mathbb{R}^{n+m} : x_J = Wx_{\bar{J}} \text{ and } x_{\bar{J}} \in \mathcal{Y}\}$, we can clearly see that the problem (2.3) takes the same form as (2.2). In addition, there obviously exists a feasible point $(u^{\text{feas}}, \alpha^{\text{feas}})$ for problem (2.3) when $\mathcal{Y} \neq \emptyset$, i.e., there exist $(u^{\text{feas}}, \alpha^{\text{feas}})$ such that $Wu^{\text{feas}} = \alpha^{\text{feas}}$ and $u^{\text{feas}} \in \mathcal{Y}$. In particular, we can choose $(u^{\text{feas}}, \alpha^{\text{feas}}) = (0, 0)$, which is the choice we make for our numerical studies. We now discuss the implementation details of the PD method when solving the proposed wavelet frame based model (2.3).

Given a penalty parameter $\varrho > 0$, the associated quadratic penalty function for (2.3) is defined as

$$(2.4) \qquad p_\varrho(u, \alpha) := \frac{1}{2}\|Au - f\|_D^2 + \sum_{\boldsymbol{i}} \lambda_{\boldsymbol{i}} \|\alpha_{\boldsymbol{i}}\|_0 + \frac{\varrho}{2}\|Wu - \alpha\|_2^2.$$

Then we have the following PD method for problem (2.3) where each penalty subproblem is approximately solved by a BCD method (see [40] for details).

Penalty Decomposition (PD) Method for (2.3)**:**

Let $\varrho_0 > 0$, $\delta > 1$ be given. Choose an arbitrary $\alpha^{0,0} \in \mathbb{R}^m$ and a constant Υ such that $\Upsilon \geq \max\{\frac{1}{2}\|Au^{\text{feas}} - f\|_D^2 + \sum_i \lambda_i \|\alpha_i^{\text{feas}}\|_0, \min_{u \in \mathcal{Y}} p_{\varrho_0}(u, \alpha^{0,0})\}$. Set $k = 0$.

   1) Set $q = 0$ and apply the BCD method to find an approximate solution $(u^k, \alpha^k) \in \mathcal{Y} \times \mathbb{R}^m$ for the penalty subproblem

$$(2.5) \qquad \min\{p_{\varrho_k}(u, \alpha) : \ u \in \mathcal{Y}, \ \alpha \in \mathbb{R}^m\}$$

   by performing steps 1a)–1d):
   1a) Solve $u^{k,q+1} \in \text{Arg} \min_{u \in \mathcal{Y}} p_{\varrho_k}(u, \alpha^{k,q})$.
   1b) Solve $\alpha^{k,q+1} \in \text{Arg} \min_{\alpha \in \mathbb{R}^n} p_{\varrho_k}(u^{k,q+1}, \alpha)$.
   1c) If $(u^{k,q+1}, \alpha^{k,q+1})$ satisfies the stopping criteria of the BCD method, set $(u^k, \alpha^k) := (u^{k,q+1}, \alpha^{k,q+1})$ and go to step 2).
   1d) Otherwise, set $q \leftarrow q + 1$ and go to step 1a).
   2) If (u^k, α^k) satisfies the stopping criteria of the PD method, stop and output u^k. Otherwise, set $\varrho_{k+1} := \delta\varrho_k$.
   3) If $\min_{u \in \mathcal{Y}} p_{\varrho_{k+1}}(u, \alpha^k) > \Upsilon$, set $\alpha^{k+1,0} := \alpha^{\text{feas}}$. Otherwise, set $\alpha^{k+1,0} := \alpha^k$.
   4) Set $k \leftarrow k + 1$ and go to step 1).

end

Remark 2.1. In the practical implementation, we terminate the inner iterations of the BCD method based on the relative progress of $p_{\varrho_k}(u^{k,q}, \alpha^{k,q})$ which can be described as follows:

$$(2.6) \qquad \frac{|p_{\varrho_k}(u^{k,q}, \alpha^{k,q}) - p_{\varrho_k}(u^{k,q+1}, \alpha^{k,q+1})|}{\max(|p_{\varrho_k}(u^{k,q+1}, \alpha^{k,q+1})|, 1)} \ \leq \ \epsilon_I.$$

Moreover, we terminate the outer iterations of the PD method once

$$(2.7) \qquad \frac{\|Wu^k - \alpha^k\|_2}{\max(|p_{\varrho_k}(u^k, \alpha^k)|, 1)} \ \leq \ \epsilon_O.$$

Next we discuss how to solve two subproblems arising in step 1a) and 1b) of the PD method.

2.2.1. *The BCD subproblem in step 1a).* The BCD subproblem in step 1a) is in the form of

$$(2.8) \qquad \min_{u \in \mathcal{Y}} \frac{1}{2}\langle u, Qu \rangle - \langle c, u \rangle$$

for some $Q \succ 0$ and $c \in \mathbb{R}^n$. Obviously, when $\mathcal{Y} = \mathbb{R}^n$, problem (2.8) is an unconstrained quadratic programming problem that can be solved by the conjugate gradient method. Nevertheless, the pixel values of an image are usually bounded. For example, the pixel values of a CT image should be always greater than or equal to zero and the pixel values of a grayscale image is between $[0, 255]$. Then the corresponding $\mathcal{Y}$ of these two examples are $\mathcal{Y} = \{x \in \mathbb{R}^n : x_i \geq lb \ \forall i = 1, \ldots, n\}$ with $lb = 0$ and $\mathcal{Y} = \{x \in \mathbb{R}^n : lb \leq x_i \leq ub \ \forall i = 1, \ldots, n\}$ with $lb = 0$ and $ub = 255$. To solve these types of the constrained quadratic programming problems, we apply the nonmonotone projected gradient method proposed in [4] and terminate it using the duality gap and dual feasibility conditions (if necessary).

For $\mathcal{Y} = \{x \in \mathbb{R}^n : x_i \geq lb \; \forall i = 1, \ldots, n\}$, given a Lagrangian multiplier $\beta \in \mathbb{R}^n$, the associated Lagrangian dual function of (2.8) can be written as:

$$L(u, \beta) = w(u) + \beta^\top (lb - u),$$

where $w(u) = \frac{1}{2}\langle u, Qu \rangle - \langle c, u \rangle$. Based on the Karush-Kuhn-Tucker (KKT) conditions, for an optimal solution u^* of (2.8), there exists a Lagrangian multiplier β^* such that

$$Qu^* - c - \beta^* = 0,$$
$$\beta_i^* \geq 0 \quad \forall i = 1, \ldots, n,$$
$$(lb - u_i^*)\beta_i^* = 0 \quad \forall i = 1, \ldots, n.$$

Then at the sth iteration of the projected gradient method, we let $\beta^s = Qu^s - c$. As $\{u^s\}$ approaches the solution u^* of (2.8), $\{\beta^s\}$ approaches the Lagrangian multiplier β^* and the corresponding duality gap at each iteration is given by $\sum_{i=1}^n \beta_i^s(lb - u_i^s)$. Therefore, we terminate the projected gradient method when

$$\frac{|\sum_{i=1}^n \beta_i^s(lb - u_i^s)|}{\max(|w(u^s)|, 1)} \leq \epsilon_D \quad \text{and} \quad \frac{-\min(\beta^s, 0)}{\max(\|\beta^s\|_2, 1)} \leq \epsilon_F$$

for some tolerances $\epsilon_D, \epsilon_F > 0$.

For $\mathcal{Y} = \{x \in \mathbb{R}^n : lb \leq x_i \leq ub \; \forall i = 1, \ldots, n\}$, given Lagrangian multipliers $\beta, \gamma \in \mathbb{R}^n$, the associated Lagrangian function of (2.8) can be written as:

$$L(u, \beta, \gamma) = w(u) + \beta^\top (lb - u) + \gamma^\top (u - ub),$$

where $w(u)$ is defined as above. Based on the KKT conditions, for an optimal solution u^* of (2.8), there exist Lagrangian multipliers β^* and γ^* such that

$$Qu^* - c - \beta^* + \gamma^* = 0,$$
$$\beta_i^* \geq 0 \quad \forall i = 1, \ldots, n,$$
$$\gamma_i^* \geq 0 \quad \forall i = 1, \ldots, n,$$
$$(lb - u_i^*)\beta_i^* = 0 \quad \forall i = 1, \ldots, n,$$
$$(u_i^* - ub)\gamma_i^* = 0 \quad \forall i = 1, \ldots, n.$$

Then at the sth iteration of the projected gradient method, we let $\beta^s = \max(Qu^s - c, 0)$ and $\gamma^s = -\min(Qu^s - c, 0)$. As $\{u^s\}$ approaches the solution u^* of (2.8), $\{\beta^s\}$ and $\{\gamma^s\}$ approach Lagrangian multipliers β^* and γ^*. In addition, the corresponding duality gap at each iteration is given by $\sum_{i=1}^n (\beta_i^s(lb - u_i^s) + \gamma_i^s(u_i^s - ub))$ and the duality feasibility is automatically satisfied. Therefore, we terminate the projected gradient method when

$$\frac{|\sum_{i=1}^n (\beta_i^s(lb - u_i^s) + \gamma_i^s(u_i^s - ub))|}{\max(|w(u^s)|, 1)} \leq \epsilon_D$$

for some tolerance $\epsilon_D > 0$.

2.2.2. *The BCD subproblem in step 1b).* For $\lambda_{\boldsymbol{i}} \geq 0$, $\varrho > 0$ and $c \in \mathbb{R}^m$, the BCD subproblem in step 1b) is in the form of

$$\min_{\alpha \in \mathbb{R}^m} \sum_{\boldsymbol{i}} \lambda_{\boldsymbol{i}} \|\alpha_{\boldsymbol{i}}\|_0 + \frac{\varrho}{2} \sum_{\boldsymbol{i}} (\alpha_{\boldsymbol{i}} - c_{\boldsymbol{i}})^2.$$

By [40, Proposition 2.2] (see also [1, 5] for example), the solutions of the above subproblem form the following set:

$$(2.9) \qquad \qquad \alpha^* \in H_{\tilde{\lambda}}(c) \quad \text{with } \tilde{\lambda}_{\boldsymbol{i}} := \sqrt{\frac{2\lambda_{\boldsymbol{i}}}{\varrho}} \text{ for all } \boldsymbol{i},$$

where $H_\gamma(\cdot)$ denotes a componentwise hard thresholding operator with threshold γ:

$$(2.10) \qquad [H_\gamma(x)]_{\boldsymbol{i}} = \begin{cases} 0 & \text{if } |x_{\boldsymbol{i}}| < \gamma_{\boldsymbol{i}}, \\ \{0, x_{\boldsymbol{i}}\} & \text{if } |x_{\boldsymbol{i}}| = \gamma_{\boldsymbol{i}}, \\ x_{\boldsymbol{i}} & \text{if } |x_{\boldsymbol{i}}| > \gamma_{\boldsymbol{i}}. \end{cases}$$

Note that H_γ is defined as a set-valued mapping [43, Chapter 5] which is different (only when $|x_{\boldsymbol{i}}| = \gamma_{\boldsymbol{i}}$) from the conventional definition of hard thresholding operator.

2.3. Convergence of the BCD method.
In this subsection, we establish some convergence results regarding the inner iterations, i.e., step 1), of the PD method. In particular, we will show that the fixed point of the BCD method is a local minimizer of (2.5). Moreover, under certain conditions, we prove that the sequence $\{(u^{k,q}, \alpha^{k,q})\}$ generated by the BCD method converges and the limit is a local minimizer of (2.5).

For convenience of presentation, we omit the index k from (2.5) and consider the BCD method for solving the following problem:

$$(2.11) \qquad \min\{p_\varrho(u, \alpha) : \ u \in \mathcal{Y}, \ \alpha \in \mathbb{R}^m\}.$$

Without loss of generality, we assume that $D = I$. We now relabel and simplify the BCD method described in steps 1a)–1c) in the PD method as follows:

$$(2.12) \qquad \begin{cases} u^{q+1} = \arg\min_{u \in \mathcal{Y}} \ \frac{1}{2}\|Au - f\|_2^2 + \frac{\varrho}{2}\|Wu - \alpha^q\|_2^2, \\ \alpha^{q+1} \in \text{Arg}\min_\alpha \ \sum_{\boldsymbol{i}} \lambda_{\boldsymbol{i}}\|\alpha_{\boldsymbol{i}}\|_0 + \frac{\varrho}{2}\|\alpha - Wu^{q+1}\|_2^2. \end{cases}$$

We first show that the fixed point of the above BCD method is a local minimizer of (2.5).

Theorem 2.2. *Given a fixed point of the BCD method (2.12), denoted as (u^*, α^*), then (u^*, α^*) is a local minimizer of $p_\varrho(u, \alpha)$.*

Proof. We first note that the first subproblem of (2.12) gives us

$$(2.13) \qquad \langle A^\top(Au^* - f) + \varrho W^\top(Wu^* - \alpha^*), v - u^* \rangle \geq 0 \quad \text{for all } v \in \mathcal{Y}.$$

By applying (2.9), the second subproblem of (2.12) leads to:

$$(2.14) \qquad \alpha^* \in H_{\tilde{\lambda}}(Wu^*).$$

Define index sets

$$\Gamma_0 := \{\boldsymbol{i} : \alpha_{\boldsymbol{i}}^* = 0\} \quad \text{and} \quad \Gamma_1 := \{\boldsymbol{i} : \alpha_{\boldsymbol{i}}^* \neq 0\}.$$

It then follows from (2.14) and (2.10) that

$$(2.15) \qquad \begin{cases} |(Wu^*)_{\boldsymbol{i}}| \leq \tilde{\lambda}_{\boldsymbol{i}} & \text{for } \boldsymbol{i} \in \Gamma_0, \\ (Wu^*)_{\boldsymbol{i}} = \alpha_{\boldsymbol{i}}^* & \text{for } \boldsymbol{i} \in \Gamma_1, \end{cases}$$

where $(Wu^*)_{\boldsymbol{i}}$ denotes $\boldsymbol{i}$th entry of Wu^*.

Consider a small deformation vector $(\partial h, \partial g)$ such that $u^* + \partial h \in \mathcal{Y}$. Using (2.13), we have

$$
\begin{aligned}
p_\varrho(u^* + \partial h, \alpha^* + \partial g) &= \frac{1}{2}\|Au^* + A\partial h - f\|_2^2 + \sum_i \lambda_i \|(\alpha^* + \partial g)_i\|_0 \\
&\quad + \frac{\varrho}{2}\|\alpha^* + \partial g - W(u^* + \partial h)\|_2^2 \\
&= \frac{1}{2}\|Au^* - f\|_2^2 + \langle A\partial h, Au^* - f\rangle + \frac{1}{2}\|A\partial h\|_2^2 + \sum_i \lambda_i \|(\alpha^* + \partial g)_i\|_0 \\
&\quad + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 + \varrho\langle \alpha^* - Wu^*, \partial g - W\partial h\rangle + \frac{\varrho}{2}\|\partial g - W\partial h\|_2^2 \\
&= \frac{1}{2}\|Au^* - f\|_2^2 + \sum_i \lambda_i \|(\alpha^* + \partial g)_i\|_0 + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 + \frac{1}{2}\|A\partial h\|_2^2 \\
&\quad + \langle \partial h, A^\top(Au^* - f) + \varrho W^\top(Wu^* - \alpha^*)\rangle + \varrho\langle \partial g, \alpha^* - Wu^*\rangle + \frac{\varrho}{2}\|\partial g - W\partial h\|_2^2 \\
&\geq \frac{1}{2}\|Au^* - f\|_2^2 + \sum_i \lambda_i \|(\alpha^* + \partial g)_i\|_0 + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 \\
&\quad + \langle \partial h, A^\top(Au^* - f) + \varrho W^\top(Wu^* - \alpha^*)\rangle + \varrho\langle \partial g, \alpha^* - Wu^*\rangle \\
(\text{By (2.13)}) &\geq \frac{1}{2}\|Au^* - f\|_2^2 + \sum_i \lambda_i \|(\alpha^* + \partial g)_i\|_0 + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 + \varrho\langle \partial g, \alpha^* - Wu^*\rangle \\
&= \frac{1}{2}\|Au^* - f\|_2^2 + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 + \sum_i \left(\lambda_i\|\alpha_i^* + \partial g_i\|_0 + \varrho\partial g_i(\alpha_i^* - (Wu^*)_i)\right).
\end{aligned}
$$

Splitting the summation in the last equation with respect to index sets Γ_0 and Γ_1 and using (2.15), we have

$$
\begin{aligned}
p_\varrho(u^* + \partial h, \alpha^* + \partial g) &\geq \frac{1}{2}\|Au^* - f\|_2^2 + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 \\
&\quad + \sum_{i\in\Gamma_0} \left(\lambda_i\|\partial g_i\|_0 - \varrho\partial g_i(Wu^*)_i\right) + \sum_{i\in\Gamma_1} \lambda_i\|\alpha_i^* + \partial g_i\|_0.
\end{aligned}
$$

Notice that when $|\partial g_i|$ is small enough, we then have

$$
\|\alpha_i^* + \partial g_i\|_0 = \|\alpha_i^*\|_0 \quad \text{for } i \in \Gamma_1.
$$

Therefore, we have

$$
\begin{aligned}
p_\varrho(u^* + \partial h, \alpha^* + \partial g) &\geq \frac{1}{2}\|Au^* - f\|_2^2 + \frac{\varrho}{2}\|\alpha^* - Wu^*\|_2^2 \\
&\quad + \sum_{i\in\Gamma_0} \left(\lambda_i\|\partial g_i\|_0 - \varrho\partial g_i(Wu^*)_i\right) + \sum_{i\in\Gamma_1} \lambda_i\|\alpha_i^*\|_0 \\
&= p_\varrho(u^*, \alpha^*) + \sum_{i\in\Gamma_0} \left(\lambda_i\|\partial g_i\|_0 - \varrho\partial g_i(Wu^*)_i\right).
\end{aligned}
$$

We now show that, for $i \in \Gamma_0$ and $\|\partial g\|$ small enough,

(2.16) $$\lambda_i\|\partial g_i\|_0 - \varrho\partial g_i(Wu^*)_i \geq 0.$$

For the indices i such that $\lambda_i = 0$, first inequality of (2.15) implies that $(Wu^*)_i = 0$ and hence (2.16) holds. Therefore, we only need to consider indices $i \in \Gamma_0$ such that $\lambda_i \neq 0$. Then obviously as long as $|\partial g_i| \leq \frac{\lambda_i}{\varrho|(Wu^*)_i|}$, (2.16) holds. We now conclude that there exists $\varepsilon > 0$ such that for all $(\partial h, \partial g)$ satisfying $\max(\|\partial h\|_\infty, \|\partial g\|_\infty) < \varepsilon$, we have $p_\varrho(u^* + \partial h, \alpha^* + \partial g) \geq p_\varrho(u^*, \alpha^*)$. $\qquad\square$

We next show that under some suitable assumptions, the sequence $\{(u^q, \alpha^q)\}$ generated by (2.12) converges to a fixed point of the BCD method.

Theorem 2.3. *Assume that $\mathcal{Y} = \mathbb{R}^n$ and $A^\top A \succ 0$. Let $\{(u^q, \alpha^q)\}$ be the sequence generated by the BCD method described in (2.12). Then, the sequence $\{(u^q, \alpha^q)\}$ is bounded. Furthermore, any cluster point of the sequence $\{(u^q, \alpha^q)\}$ is a fixed point of (2.12).*

Proof. In view of $\mathcal{Y} = \mathbb{R}^n$ and the optimality condition of the first subproblem of (2.12), one can see that

$$(2.17) \qquad u^{q+1} = (A^\top A + \varrho I)^{-1} A^\top f + \varrho (A^\top A + \varrho I)^{-1} W^\top \alpha^q.$$

Let $x := (A^\top A + \varrho I)^{-1} A^\top f$, $P := \varrho (A^\top A + \varrho I)^{-1}$, equation (2.17) can be rewritten as

$$(2.18) \qquad\qquad u^{q+1} = x + PW^\top \alpha^q.$$

Moreover, by the assumption $A^\top A \succ 0$, we have $0 \prec P \prec I$.

Using (2.18) and (2.10), we observe from the second subproblem of (2.12) that

$$(2.19) \qquad \alpha^{q+1} \in H_{\tilde\lambda}(Wu^{q+1}) = H_{\tilde\lambda}\left(Wx + WPW^\top \alpha^q\right).$$

Let $Q := I - WPW^\top$, then (2.19) can be rewritten as

$$(2.20) \qquad\qquad \alpha^{q+1} \in H_{\tilde\lambda}\left(\alpha^q + Wx - Q\alpha^q\right).$$

In addition, from $W^\top W = I$ we can easily show that $0 \prec Q \preceq I$.

Let $F(\alpha, \beta) := \frac{1}{2}\langle \alpha, Q\alpha \rangle - \langle Wx, \alpha \rangle + \sum_i \bar\lambda_i \|\alpha_i\|_0 - \frac{1}{2}\langle \alpha - \beta, Q(\alpha - \beta)\rangle + \frac{1}{2}\|\alpha - \beta\|_2^2$ where $\bar\lambda = \frac{\lambda}{\rho}$. Then we have

$$(2.21) \quad \text{Argmin}_\alpha F(\alpha, \alpha^q) = \text{Argmin}_\alpha \frac{1}{2}\|\alpha - (\alpha^q + Wx - Q\alpha^q)\|_2^2 + \sum_i \bar\lambda_i \|\alpha_i\|_0.$$

In view of equation (2.20) and (2.21) and the definition of the hard thresholding operator, we can easily observe that $\alpha^{q+1} \in \text{Argmin}_\alpha F(\alpha, \alpha^q)$. By following similar arguments as in [5, Lemma 1, Lemma D.1], we have

$$
\begin{aligned}
F(\alpha^{q+1}, \alpha^{q+1}) &\le F(\alpha^{q+1}, \alpha^{q+1}) + \frac{1}{2}\|\alpha^{q+1} - \alpha^q\|_2^2 - \frac{1}{2}\langle \alpha^{q+1} - \alpha^q, Q(\alpha^{q+1} - \alpha^q)\rangle \\
&= F(\alpha^{q+1}, \alpha^q) \\
&\le F(\alpha^q, \alpha^q),
\end{aligned}
$$

which leads to

$$\|\alpha^{q+1} - \alpha^q\|_2^2 - \langle \alpha^{q+1} - \alpha^q, Q(\alpha^{q+1} - \alpha^q)\rangle \le 2F(\alpha^q, \alpha^q) - 2F(\alpha^{q+1}, \alpha^{q+1}).$$

Since $P \succ 0$, we have

$$
\begin{aligned}
\|W^\top(\alpha^{q+1} - \alpha^q)\|_2^2 &\le \frac{1}{C_1}\langle W^\top(\alpha^{q+1} - \alpha^q), PW^\top(\alpha^{q+1} - \alpha^q)\rangle \\
&= \frac{1}{C_1}\langle \alpha^{q+1} - \alpha^q, (I - Q)(\alpha^{q+1} - \alpha^q)\rangle \\
&= \frac{1}{C_1}\left(\|\alpha^{q+1} - \alpha^q\|_2^2 - \langle \alpha^{q+1} - \alpha^q, Q(\alpha^{q+1} - \alpha^q)\rangle\right) \\
&\le \frac{2}{C_1}F(\alpha^q, \alpha^q) - \frac{2}{C_1}F(\alpha^{q+1}, \alpha^{q+1})
\end{aligned}
$$

for some $C_1 > 0$. Telescoping on the above inequality and using the fact that $\sum_i \lambda_i \|\alpha_i\|_0 \geq 0$, we have

$$
\begin{aligned}
\sum_{q=0}^{N} \|W^\top(\alpha^{q+1} - \alpha^q)\|_2^2 \quad &\leq \quad \frac{2}{C_1} F(\alpha^0, \alpha^0) - \frac{2}{C_1} F(\alpha^{N+1}, \alpha^{N+1}) \\
&\leq \quad \frac{2}{C_1} \left(F(\alpha^0, \alpha^0) - (\frac{1}{2}\langle \alpha^{N+1}, Q\alpha^{N+1} \rangle - \langle Wx, \alpha^{N+1} \rangle) \right) \\
&\leq \quad \frac{2}{C_1} \left(F(\alpha^0, \alpha^0) - K \right),
\end{aligned}
$$

where K is the optimal value of $\min_y\{\frac{1}{2}\langle y, Qy \rangle - \langle Wx, y \rangle\}$. Since $Q \succ 0$, we have $K > -\infty$. Then the last inequality implies that $\lim_{q\to\infty} \|W^\top(\alpha^{q+1} - \alpha^q)\|_2 \to 0$.

By using (2.18) and $P \prec I$, we see that

$$
\begin{aligned}
\|u^{q+1} - W^\top \alpha^{q+1}\|_2 \quad &= \quad \|x + PW^\top \alpha^q - W^\top \alpha^{q+1} + W^\top \alpha^q - W^\top \alpha^q\|_2 \\
&= \quad \|x + (P - I)W^\top \alpha^q - W^\top(\alpha^{q+1} - \alpha^q)\|_2 \\
&\geq \quad \|x + (P - I)W^\top \alpha^q\|_2 - \|W^\top(\alpha^{q+1} - \alpha^q)\|_2 \\
&= \quad \|(I - P)W^\top \alpha^q - x\|_2 - \|W^\top(\alpha^{q+1} - \alpha^q)\|_2 \\
&\geq \quad \|(I - P)W^\top \alpha^q\|_2 - \|x\|_2 - \|W^\top(\alpha^{q+1} - \alpha^q)\|_2 \\
&\geq \quad C_2 \|W^\top \alpha^q\|_2 - \|x\|_2 - \|W^\top(\alpha^{q+1} - \alpha^q)\|_2
\end{aligned}
$$

for some $C_2 > 0$. Then by rearranging the above inequality and using the fact $W^\top W = I$, we have

$$
\begin{aligned}
\|W^\top \alpha^q\|_2 \quad &\leq \quad \frac{1}{C_2}(\|u^{q+1} - W^\top \alpha^{q+1}\|_2 + \|x\|_2 + \|W^\top(\alpha^{q+1} - \alpha^q)\|_2) \\
&= \quad \frac{1}{C_2}(\|W^\top(Wu^{q+1} - \alpha^{q+1})\|_2 + \|x\|_2 + \|W^\top(\alpha^{q+1} - \alpha^q)\|_2) \\
&\leq \quad \frac{1}{C_2}(\|Wu^{q+1} - \alpha^{q+1}\|_2 + \|x\|_2 + \|W^\top(\alpha^{q+1} - \alpha^q)\|_2).
\end{aligned}
$$

By the definition of the hard thresholding operator and (2.19), we can easily see that $\|Wu^{q+1} - \alpha^{q+1}\|_2$ is bounded. In addition, notice that $\|x\|_2$ is a constant and $\lim_{q\to\infty} \|W^\top(\alpha^{q+1} - \alpha^q)\|_2 \to 0$. Thus $\|W^\top \alpha^q\|_2$ is also bounded. By using (2.18) and the definition of the hard thresholding operator again, we can immediately see that both $\{u^{q+1}\}$ and $\{\alpha^{q+1}\}$ are bounded as well.

Suppose that (u^*, α^*) is a cluster point of the sequence $\{(u^q, \alpha^q)\}$. Therefore, there exists a subsequence $\{(u^{q_l}, \alpha^{q_l})\}_l$ converging to (u^*, α^*). Using (2.19) and the definition of the hard thresholding operator, we can observe that

$$
\alpha^* = \lim_{l\to\infty} \alpha^{q_l+1} \in H_{\tilde\lambda}(\lim_{l\to\infty} Wu^{q_l+1}) = H_{\tilde\lambda}(Wu^*).
$$

In addition, it follows from (2.17) that

$$
u^* = (A^\top A + \varrho I)^{-1} A^\top f + \varrho(A^\top A + \varrho I)^{-1} W^\top \alpha^*.
$$

In view of the above two relations, one can immediately conclude that $\{(u^*, \alpha^*)\}$ is a fixed point of (2.12). □

In the view of Theorems 2.2, 2.3 and under some suitable assumptions, we can easily observe the following convergence of the BCD method.

Theorem 2.4. *Assume that $\mathcal{Y} = \mathbb{R}^n$ and $A^\top A \succ 0$. Then, the sequence $\{(u^q, \alpha^q)\}$ generated by the BCD method has at least one cluster point. Furthermore, any cluster point of the sequence $\{(u^q, \alpha^q)\}$ is a local minimizer of (2.11).*

For the PD method itself, similar arguments as in the proof of [40, Theorem 3.2] will lead to that every accumulation point of the sequence $\{(u^k, \alpha^k)\}$ is a feasible point of (2.3). Although it is not clear whether the accumulation point is a local minimizer of (2.3), our numerical results show that the solutions obtained by the PD method are superior to those obtained by the balanced approach and the analysis based approach.

3. Numerical results

In this section, we conduct numerical experiments to test the performance of the PD method for problem (2.3) presented in Section 2 and compare the results with the balanced approach (1.2) and the analysis based approach (1.3). We use the accelerated proximal gradient (APG) algorithm [48] (see also [3]) to solve the balanced approach; and we use the split Bregman algorithm [37, 11] to solve the analysis based approach.

For the APG algorithm that solves balanced approach (1.2), we shall adopt the following stopping criteria:

$$\min\left\{\frac{\|\alpha^k - \alpha^{k-1}\|_2}{\max\{1, \|\alpha^k\|_2\}}, \frac{\|AW^\top \alpha^k - f\|_D}{\|f\|_2}\right\} \leq \epsilon_P.$$

For the split Bregman algorithm that solves the analysis based approach (1.3), we shall use the following stopping criteria:

$$\frac{\|Wu^{k+1} - \alpha^{k+1}\|_2}{\|f\|_2} \leq \epsilon_S.$$

Throughout this section, the codes of all the algorithms are written in MATLAB and all computations below are performed on a workstation with Intel Xeon E5410 CPU (2.33GHz) and 8GB RAM running Red Hat Enterprise Linux (kernel 2.6.18). If not specified, the piecewise linear B-spline framelets constructed by [44] are used in all the numerical experiments. We also take $D = I$ for all three methods for simplicity. For the PD method, we choose $\epsilon_I = 10^{-4}$ and $\epsilon_O = 10^{-3}$ and set $\alpha^{0,0}$, α^{feas} and u^{feas} to be zero vectors. In addition, we choose [4, Algorithm 2.2] and set $M = 20$, $\epsilon_D = 5 \times 10^{-5}$ and $\epsilon_F = 10^{-4}$ (if necessary) for the projected gradient method applied to one of subproblems arising in the BCD method (i.e., step 1a in the PD method).

3.1. Experiments on CT image reconstruction.
In this subsection, we apply the PD method stated in Section 2 to solve problem (2.3) on CT images and compare the results with the balanced approach (1.2) and the analysis based approach (1.3). The matrix A in (1.1) is taken to be a projection matrix based on fan-beam scanning geometry using Siddon's algorithm [49], and η is generated from a zero mean Gaussian distribution with variance $\sigma = 0.01\|f\|_\infty$. In addition, we pick the level of framelet decomposition to be 4 for the best quality of the reconstructed images. For the balanced approach, we set $\kappa = 2$ and take $\epsilon_P = 1.5 \times 10^{-2}$ for the stopping criteria of the APG algorithm. We set $\epsilon_S = 10^{-5}$ for the stopping criteria of the split Bregman algorithm when solving the analysis based approach. Moreover, we take $\mathcal{Y} = \{x \in \mathbb{R}^n : x_i \geq 0 \ \forall i = 1, \ldots, n\}$ for model (2.3), and take

TABLE 1. Comparisons: CT image reconstruction

	Balanced approach	Analysis based approach	PD method
Time	56.0	204.8	147.6
PSNR	56.06	59.90	60.22

$\delta = 10$ and $\varrho_0 = 10$ for the PD method. To measure quality of the restored image, we use the PSNR value defined by

$$\text{PSNR} := -20 \log_{10} \frac{\|u - \tilde{u}\|_2}{n},$$

where u and $\tilde{u}$ are the original and restored images, respectively, and n is the total number of pixels in u.

Table 1 summarizes the results of all three models when applying to the CT image restoration problem and the corresponding images and their zoom-in views are shown in Figure 1 and Figure 2. In Table 1, the CPU time (in seconds) and PSNR values of all three methods are given in the first and second row, respectively. In order to fairly compare the results, we have tuned the parameter λ to achieve the best quality of the restoration images for each individual method. We observe that based on the PSNR values listed in Table 1 the analysis based approach and the PD method obviously achieve better restoration results than the balanced approach. Nevertheless, the APG algorithm for the balanced approach is the fastest algorithm in this experiment. In addition, the PD method is faster and achieves larger PSNR than the split Bergman algorithm for the analysis based approach. Moreover, we can observe from Figure 2 that the edges are recovered better by the PD method and the balanced approach.
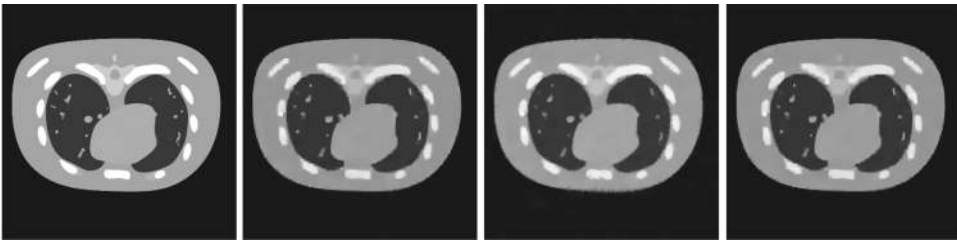

FIGURE 1. CT image reconstruction. Images from left to right are: original CT image, reconstructed image by the balanced approach, reconstructed image by the analysis based approach and reconstructed image by the PD method.

3.2. **Experiments on image deconvolution.** In this subsection, we apply the PD method stated in Section 2 to solve problem (2.3) on image deblurring problems and compare the results with the balanced approach (1.2) and the analysis based approach (1.3). The matrix A in (2.3) is taken to be a convolution matrix with corresponding kernel a Gaussian function (generated in MATLAB by "fspecial('Gaussian', 9, 1.5);") and η is generated from a zero mean Gaussian distribution with variance σ. If not specified, we choose $\sigma = 3$ in our experiments. In addition, we pick the level of framelet decomposition to be 4 for the best quality of the reconstructed images. We set $\kappa = 1$ for the balanced approach and choose both ϵ_P
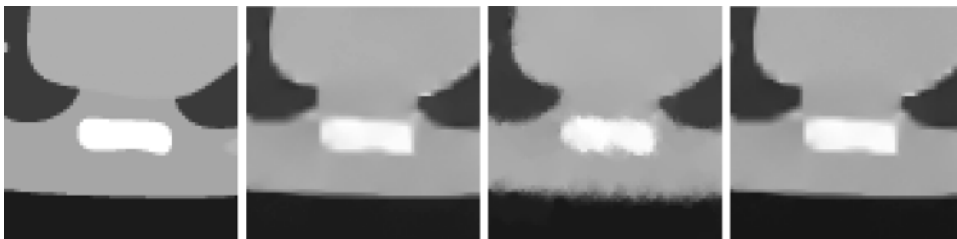
FIGURE 2. Zoom-in views of the CT image reconstruction. Images from left to right are: original CT image, reconstructed image by the balanced approach, reconstructed image by the analysis based approach and reconstructed image by the PD method.

and ϵ_S to be 10^{-4} for the stopping criteria of both APG algorithm and the split Bregman algorithm. Moreover, we set $\mathcal{Y} = \{x \in \mathbb{R}^n : 0 \le x_i \le 255 \; \forall i = 1, \dots, n\}$ for model (2.3), and take $\delta = 10$ and $\varrho_0 = 10^{-3}$ for the PD method. To measure quality of the restored image, we use the PSNR value defined by

$$\mathrm{PSNR} := -20 \log_{10} \frac{\|u - \tilde{u}\|_2}{255n}.$$

We first test all three methods on twelve different images by using piecewise linear wavelet and summarize the results in Table 2. The names and sizes of images are listed in the first two columns. The CPU time (in seconds) and PSNR values of all three methods are given in the last six columns. In addition, the zoom-in views of original images, observed images and recovered images are shown in Figures 3 and 4. In order to fairly compare the results, we have tuned the parameter λ to achieve the best quality of the restoration images for each individual method and each given image.

We first observe that in Table 2, the PSNR values obtained by the PD method are generally better than those obtained by the other two approaches. Although for some of the images (i.e., "Downhill", "Bridge", "Duck" and "Barbara"), the PSNR values obtained by the PD methods are comparable to those of balanced and analysis based approaches, the quality of the restored images cannot only be judged by their PSNR values. Indeed, the zoom-in views of the recovered images in Figure 3 and Figure 4 show that for all tested images, the PD method produces visually superior results than the other two approaches in terms of both sharpness of edges and smoothness of regions away from edges. Takeing the image "Barbara" as an example, the PSNR value of the PD method is only slightly greater than that obtained by the other two approaches. However, the zoom-in views of "Barbara" in Figure 4 show that the face of Barbara and the textures on her scarf are better recovered by the PD method than the other two approaches. This confirms the observation that the penalizing ℓ_0 "norm" of Wu should provide good balance between sharpness of features and smoothness of the reconstructed images. We finally note that the PD method is slower than other two approaches in these experiments but the processing time of the PD method is still acceptable.

We next compare all three methods on the "Portrait I" image by using three different tight wavelet frame systems, i.e., Haar framelets, piecewise linear framelets and piecewise cubic framelets constructed by [44]. We summarize the results in Table 3. The names of three wavelets are listed in the first column. The CPU time

TABLE 2. Comparisons: image deconvolution

Name	Size	Balanced approach		Analysis based approach		PD method	
		Time	PSNR	Time	PSNR	Time	PSNR
Downhill	256	12.5	27.24	6.1	27.36	29.5	27.35
Cameraman	256	18.2	26.65	7.0	26.73	31.1	27.21
Bridge	256	14.5	25.40	5.1	25.46	33.0	25.44
Pepper	256	21.6	26.82	7.5	26.63	32.1	27.29
Clock	256	17.3	29.42	19.9	29.48	22.3	29.86
Portrait I	256	32.7	33.93	19.3	33.98	27.1	35.44
Duck	464	30.6	31.00	16.1	31.11	72.5	31.09
Barbara	512	38.8	24.62	12.3	24.62	77.4	24.69
Aircraft	512	55.9	30.75	35.1	30.81	67.5	31.29
Couple	512	91.4	28.40	41.5	28.14	139.1	29.32
Portrait II	512	45.2	30.23	22.1	30.20	48.9	30.90
Lena	516	89.3	12.91	31.0	12.51	67.0	13.45

TABLE 3. Comparisons among different wavelet representations

Wavelets	Balanced approach		Analysis based approach		PD method	
	Time	PSNR	Time	PSNR	Time	PSNR
Haar	17.9	33.63	20.2	33.80	24.3	34.68
Piecewise linear	32.7	33.93	22.3	33.98	27.1	35.44
Piecewise cubic	61.0	33.95	37.3	34.00	37.8	35.20

TABLE 4. Comparisons among different noise levels for the image "Portrait I"

Variances of noises	Balanced approach		Analysis based approach		PD method	
	Time	PSNR	Time	PSNR	Time	PSNR
$\sigma = 3$	32.7	33.93	22.3	33.98	27.1	35.44
$\sigma = 5$	23.7	32.84	19.4	32.89	27.2	34.48
$\sigma = 7$	19.6	32.11	25.0	32.14	29.7	33.69

(in seconds) and PSNR values of all three methods are given in the last six columns. In Table 3, we can see that the quality of the restored images by using the piecewise linear framelets and the piecewise cubic framelets is better than that by using the Haar framelets. In addition, all three methods are generally faster when using Haar framelets and slower when using piecewise cubic framelets. Overall, all three approaches when using the piecewise linear method have balanced performance in terms of time and quality (i.e., the PSNR value). Finally, we observe that the PD method consistently achieves the best quality of restored images among all of the approaches for all three different tight wavelet frame systems.

Finally, we test how different noise levels affect the restored images obtained from all the three methods. We choose three different noise levels (i.e., $\sigma = 3, 5, 7$) for the image "Portrait I", and test all the three methods by using piecewise linear framelets. We summarize the results in Table 4. The variances of noises are listed in the first column. The CPU time (in seconds) and PSNR values of all three methods

are given in the last six columns. We observe that the qualities of the restored images by all three methods degrade when the noise level increases. Nevertheless, the PD method still outperforms the other two methods.

FIGURE 3. Zoom-in to the texture part of "Downhill", "Cameraman", "Bridge", "Pepper", "Clock", and "Portrait I". Images from left to right are: original image, observed image, results of the balanced approach, results of the analysis based approach and results of the PD method.

FIGURE 4. Zoom-in to the texture part of "Duck", "Barbara", "Aircraft", "Couple", "Portrait II" and "Lena". Images from left to right are: original image, observed image, results of the balanced approach, results of the analysis based approach and results of the PD method.

4. Conclusion

In this paper, we proposed a wavelet frame based ℓ_0 minimization model, which is motivated by the analysis based approach and balanced approach. The penalty decomposition (PD) method of [40] was used to solve the proposed optimization problem. Numerical results showed that the proposed model solved by the PD method can generate images with better quality than those obtained by either analysis based approach or balanced approach in terms of restoring sharp features like edges as well as maintaining smoothness of the recovered images. Convergence analysis of the subiterations in the PD method was also provided.

Acknowledgement

References

[1] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455):939–967, 2001. MR1946364 (2003k:62098)

[2] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations.* Springer, 2006. MR2244145 (2007j:94004)

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. MR2486527 (2010d:35390)

[4] E.G. Birgin, J.M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000. MR1777088 (2001e:90143)

[5] T. Blumensath and M.E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5):629–654, 2008. MR2461601 (2009j:68077)

[6] T. Blumensath and M.E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009. MR2559726 (2010i:94048)

[7] J.F. Cai, R.H. Chan, L. Shen, and Z. Shen. Convergence analysis of tight framelet approach for missing data recovery. *Advances in Computational Mathematics*, 31(1):87–113, 2009. MR2511575 (2010c:42070)

[8] J.F. Cai, R.H. Chan, and Z. Shen. Simultaneous cartoon and texture inpainting. *Inverse Problems and Imaging* 4(3):379–395, 2010. MR2671102 (2011k:94019)

[9] J.F. Cai, B. Dong, S. Osher, and Z. Shen. Image restorations: Total variation, wavelet frames and beyond. *J. Amer. Math. Soc.* 25(4):1033–1089, 2012. MR2947945

[10] J.F. Cai, S. Osher, and Z. Shen. Linearized Bregman iterations for frame-based image deblurring. *SIAM Journal on Imaging Sciences*, 2(1):226–252, 2009. MR2486529 (2010h:94041)

[11] J.F. Cai, S. Osher, and Z. Shen. Split Bregman methods and frame based image restoration. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 8(2):337–369, 2009. MR2581025 (2011a:94016)

[12] E.J. Candes, Y.C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 2010. MR2795875

[13] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006. MR2236170 (2007e:94020)

[14] E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005. MR2243152 (2007b:94313)

[15] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006. MR2300700 (2008c:94009)

[16] A. Chai and Z. Shen. Deconvolution: A wavelet frame approach. *Numerische Mathematik*, 106(4):529–587, 2007. MR2317925 (2008i:42067)

[17] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997. MR1440119 (98c:65099)

[18] R.H. Chan, T.F. Chan, L. Shen, and Z. Shen. Wavelet algorithms for high-resolution image reconstruction. *SIAM Journal on Scientific Computing*, 24(4):1408–1432, 2003. MR1976222 (2004e:94003)

[19] R.H. Chan, S.D. Riemenschneider, L. Shen, and Z. Shen. Tight frame: an efficient way for high-resolution image reconstruction. *Applied and Computational Harmonic Analysis*, 17(1):91–115, 2004. MR2067917 (2005h:94006)

[20] R.H. Chan, L. Shen, and Z. Shen. A framelet-based approach for image inpainting. *technical report*, 4:325, 2005.

[21] R.H. Chan, Z. Shen, and T. Xia. A framelet algorithm for enhancing video stills. *Applied and Computational Harmonic Analysis*, 23(2):153–170, 2007. MR2344608 (2008h:94006)

[22] T. Chan, S. Esedoglu, F. Park, and A. Yip. Total variation image restoration: Overview and recent developments. *Handbook of mathematical models in computer vision*, pages 17–31, 2006. MR2232522

[23] T.F. Chan and J. Shen. *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Society for Industrial Mathematics, 2005. MR2143289 (2007c:94011)

[24] I. Daubechies. *Ten lectures on wavelets*, volume CBMS-NSF Lecture Notes, SIAM, nr. 61. Society for Industrial Mathematics, 1992. MR1162107 (93e:42045)

[25] I. Daubechies, B. Han, A. Ron, and Z. Shen. Framelets: MRA-based constructions of wavelet frames. *Applied and Computational Harmonic Analysis*, 14(1):1–46, 2003. MR1971300 (2004a:42046)

[26] I. Daubechies, G. Teschke, and L. Vese. Iteratively solving linear inverse problems under general convex constraints. *Inverse Problems and Imaging*, 1(1):29–46, 2007. MR2262744 (2008e:35204)

[27] B. Dong, A. Chien, and Z. Shen. Frame based segmentation for medical images. *Communications in Mathematical Sciences*, 9(2):551–559, 2010. MR2815684

[28] B. Dong and Z. Shen. MRA based wavelet frames and applications. *IAS Lecture Notes Series, Summer Program on "The Mathematics of Image Processing", Park City Mathematics Institute*, 2010.

[29] B. Dong and Z. Shen. Frame based surface reconstruction from unorganized points. *J. Comput. Phys.*, 230(22), 8247–8255, 2011. MR2835419 (2012j:65058)

[30] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006. MR2241189 (2007e:94013)

[31] M. Elad, J.L. Starck, P. Querre, and D.L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Applied and Computational Harmonic Analysis*, 19(3):340–358, 2005. MR2186449 (2007h:94007)

[32] M.J. Fadili and J.L. Starck. Sparse representations and bayesian image inpainting. *Proceedings of SPARS*, 2005.

[33] M.J. Fadili, J.L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal*, 52(1):64, 2009.

[34] M.A.T. Figueiredo and R.D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003. MR2008658

[35] M.A.T. Figueiredo and R.D. Nowak. A bound optimization approach to wavelet-based image deconvolution. *IEEE International Conference on Image Processing*, 2005.

[36] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *Image Processing, IEEE Transactions on*, 4(7):932–946, 1995.

[37] T. Goldstein and S. Osher. The split Bregman algorithm for L1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009. MR2496060 (2010e:65087)

[38] K.K. Herrity, A.C. Gilbert, and J.A. Tropp. Sparse approximation via iterative thresholding. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

[39] X. Jia, Y. Lou, B. Dong, and S. Jiang. GPU-based iterative cone beam CT reconstruction using tight frame regularization. *Phys. Med. Biol.* 56(13):3787–3807, 2011.

[40] Z. Lu and Y. Zhang. Penalty decomposition methods for l_0-norm minimization. *preprint*, 2010.

[41] Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, volume 22. Amer Mathematical Society, 2001. MR1852741 (2002j:43001)

[42] S. Osher and R.P. Fedkiw. *Level set methods and dynamic implicit surfaces*. Springer, 2003. MR1939127 (2003j:65002)

[43] R.T. Rockafellar and J.B.W. Roger. *Variational analysis*, volume 317. Springer, 2004. MR1491362 (98m:49001)

[44] A. Ron and Z. Shen. Affine Systems in $L_2(\mathbb{R}^d)$: The Analysis of the Analysis Operator. *Journal of Functional Analysis*, 148(2):408–447, 1997. MR1469348 (99g:42043)

[45] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Journal of Physics D*, 60:259–268, 1992.

[46] G. Sapiro. *Geometric partial differential equations and image analysis*. Cambridge Univ. Press, 2001. MR1813971 (2002a:68142)

[47] Z. Shen. Wavelet frames and image restorations. *Proceedings of the International Congress of Mathematicians, Hyderabad, India*, 2010.

[48] Z. Shen, K. C. Toh, and S. Yun. An accelerated proximal gradient algorithm for frame based image restorations via the balanced approach. *SIAM Journal on Imaging Sciences*, 4(2):573–596, 2011. MR2810898

[49] R.L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics*, 12:252–255, 1985.

[50] J.L. Starck, M. Elad, and D.L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14(10):1570–1582, 2005. MR2483314

[51] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. *SIAM Journal on Numerical Analysis*, pages 686–713, 2005. MR2084232 (2006d:94018)

[52] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001. MR1835069 (2002b:90079)

[53] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008. MR2486032 (2010e:68198)

DEPARTMENT OF MATHEMATICS, SIMON FRASER UNIVERSITY, BURNABY, BC, V5A 1S6, CANADA.
E-mail address: `yza30@sfu.ca`

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF ARIZONA, 617 N. SANTA RITA AVE., TUCSON, ARIZONA, 85721-0089
E-mail address: `dongbin@math.arizona.edu`

DEPARTMENT OF MATHEMATICS, SIMON FRASER UNIVERSITY, BURNABY, BC, V5A 1S6, CANADA.
E-mail address: `zhaosong@sfu.ca`