Image restoration by minimizing zero norm of wavelet frame coefficients

Chenglong Bao[a], Bin Dong[b], Likun Hou[c], Zuowei Shen[a], Xiaoqun Zhang[c,d], Xue Zhang[d]

[a]*Department of Mathematics, National University of Singapore, Singapore 117543*
[b]*Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China*
[c]*Institute of Natural Sciences, Shanghai Jiao Tong Univeristy, Shanghai 200240, China*
[d]*School of Mathematical Sciences, Shanghai Jiao Tong Univeristy, Shanghai 200240, China*

Abstract

In this paper, we propose two algorithms, namely extrapolated proximal iterative hard thresholding (EPIHT) algorithm and EPIHT algorithm with line-search (EPIHT-LS), for solving ℓ_0 norm regularized wavelet frame balanced approach for image restoration. Under the theoretical framework of Kurdyka-Łojasiewicz property, we show that the sequences generated by the two algorithms converge to a local minimizer with linear convergence rate. Moreover, extensive numerical experiments on sparse signal reconstruction and wavelet frame based image restoration problems including CT reconstruction, image deblur, demonstrate the improvement of ℓ_0-norm based regularization models over some prevailing ones, as well as the computational efficiency of the proposed algorithms.

Keywords: Wavelet frame; ℓ_0 regularization; iterative hard threshholding; extrapolation; local minimizer; linear convergence; Kurdyka-Łojasiewicz property.

1. Introduction

The generic image restoration problem is often formulated as an inverse problem

$$b = Au + \epsilon,$$

where $b \in \mathbb{R}^m$ is observed data, A is an ill-posed linear operator, ϵ is a white Gaussian noise or observation error, and u is the true variables to be estimated. To suppress noise and preserve latent image features, many sparse approximations have been proposed, such as total variation, wavelet frame/dictionary based representation. In this paper, we focus on sparse approximation by wavelet frame systems due to its flexibility and promising performance for image restoration problems.

1.1. Image restoration by wavelet frame sparse approximation

In recent years, wavelet frames [31, 34, 49, 69, 68, 70] have been well developed and widely used in image restoration and medical imaging [18, 12, 14, 15, 17, 21, 23, 25, 26, 27, 36, 38, 46, 51, 54, 55]. The basic idea of all the wavelet frame based approaches is that images can be well approximated by few non-zero wavelet frame coefficients. More recently, wavelet frame based approaches have been linked with variational and PDE based approaches in [16, 19, 37] and references therein, where new models

Email addresses: matbc@nus.edu.sg (Chenglong Bao), dongbin@math.pku.edu.cn (Bin Dong), houlk@sjtu.edu.cn (Likun Hou), matzuows@nus.edu.sg (Zuowei Shen), xqzhang@sjtu.edu.cn (Xiaoqun Zhang), zhangxue2100@sjtu.edu.cn (Xue Zhang)

and algorithms for image restoration problems have been introduced as well. In this paper, the wavelet tight frame system we used is the B-spline wavelet tight frame, which was constructed by the Unitary Extension Principle [69].

We start with the balanced approach, originally used in [25, 27] for image super-resolution and further developed for various image restoration tasks in [12, 13, 14, 15]. The balanced approach nowadays can be formulated as the following optimization problem

$$x^* = \arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|AW^\top x - b\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)x\|^2 + \|\boldsymbol{\lambda}\cdot x\|_1, \tag{1}$$

where $\mathbb{R}^n$ is the n-dimensional Euclidean space, W is a multi-level wavelet tight frame transform operator (i.e $W^\top W = I$, others maybe have different definition, see, e.g. [39] for more details), $W^\top$ is the transpose of W, x can be seen as a coefficient vector in the transformed domain of W, D is some weighted positive definite matrix ($\|y\|_D := \sqrt{y^\top Dy}$), $0 < \kappa < \infty$ is the balanced weight, $\boldsymbol{\lambda}$ is a nonnegative sparsity-promoting weight vector that has the same size as x and $\|\boldsymbol{\lambda}\cdot x\|_1 = \sum_i \lambda_i|x_i|$ denotes the weighted ℓ_1 norm.

The balanced approach can be considered as an intermediate between two common sparse approximation models: *analysis* approach which aims to find an image u^* whose representation coefficients are sparsest and *synthesis* approach [43] which seeks for the sparsest representation coefficients x^*. It is well known that these two approaches behave differently when the wavelet tight frame is redundant. However, the balanced model (1) unifies both popular sparse approximation models and provides balanced image quality between sparseness and regularity. When $\kappa = 0$, (1) becomes the synthesis model and when $\kappa \to \infty$, it becomes the analysis model.

The other advantage of the balanced model (1) is that many efficient algorithms can be proposed for this type of convex minimization problem. For instance, the numerical scheme originally used in [25][27] was proved in [15] to be a proximal forward-backward splitting algorithm [28, 5, 32, 33, 56, 64, 73, 30]. More specifically, the proximal forward-backward splitting method for solving (1) is written as

$$x_{k+1} = \mathrm{Prox}_{F_1/\alpha_k}(x_k - \frac{1}{\alpha_k}\nabla F_2(x_k)), \tag{2}$$

where $F_1(x) = \|\boldsymbol{\lambda}\cdot x\|_1$, $F_2(x) = \frac{1}{2}\|AW^\top x - b\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)x\|^2$, $\alpha_k > 0$ is some appropriate step size and $\mathrm{Prox}_{F_1/\alpha_k}(\cdot)$ is the proximal mapping [63]

$$\mathrm{Prox}_{F_1/\alpha_k}(x_0) = \arg\min_{x\in\mathbb{R}^n} F_1(x) + \frac{\alpha_k}{2}\|x - x_0\|^2.$$

It is well-known that when $F_1(x) = \|\boldsymbol{\lambda}\cdot x\|_1$, $\mathrm{Prox}_{F_1/\alpha_k} = \mathcal{S}_{\boldsymbol{\lambda}/\alpha_k}(x)$ is the componentwise soft thresholding operator [35]:

$$\mathcal{S}_{\boldsymbol{\gamma}}(x) = sign(x)\max\{0, |x| - \boldsymbol{\gamma}\}, \text{ with } \boldsymbol{\gamma} = \boldsymbol{\lambda}/\alpha_k. \tag{3}$$

The above iterative soft thresholding based scheme has inspired extensive numerical schemes for solving ℓ_1 based sparse approximation models with applications in diverse signal and image processing tasks, see e.g. [5, 12, 13, 14, 15, 32, 33, 48, 75, 76, 24, 45].

1.2. Motivations

It is well known that under suitable assumptions, the ℓ_1-norm based approaches are capable of obtaining a sparse solution as shown by the comjkgt4rmpressive sensing theory (see e.g. [22]). However, the

non-convex ℓ_0 norm based regularization has its advantages over convex ℓ_1 norm based model in many applications, e.g. image restoration [25, 40, 78], MRI reconstruction [72], bioluminescence tomography [77]. Historically, soft and hard thresholding have been studied by Donoho and Johnstone [42] in the context of nonlinear estimators under orthonormal bases. Although the soft thresholding operator may have less artifacts, it yields loss of contrast and eroded signal peaks as ℓ_1 estimator leads to bias estimation for large coefficients [44].

In this paper, we are interested in minimizing ℓ_0 norm of wavelet frame coefficients for image restoration. More generally, we consider the following ℓ_0-norm minimization problem

$$\min_{x\in\mathbb{R}^n} \; h(x) := \|\boldsymbol{\lambda}\cdot x\|_0 + f(x). \tag{4}$$

where $\|\boldsymbol{\lambda}\cdot x\|_0 = \sum_i \lambda_i|x_i|_0$ denotes the weighted number of nonzero elements in the vector x. It is well-known that the solutions of $\min_{x\in\mathbb{R}^n} \|\boldsymbol{\lambda}\cdot x\|_0 + \frac{1}{2}\|x-c\|^2$ is the following set

$$x^* \in \mathcal{H}_{\sqrt{2\boldsymbol{\lambda}}}(c)$$

where $\mathcal{H}_{\boldsymbol{\gamma}}(\cdot)$ denotes the componentwise hard thresholding operator with thresholding $\boldsymbol{\gamma}\in\mathbb{R}^n$

$$[\mathcal{H}_{\boldsymbol{\gamma}}(c)]_i = \begin{cases} \{[c]_i\}, & \text{if } |[c]_i| > [\boldsymbol{\gamma}]_i, \\ \{0,[c]_i\} & \text{if } |[c]_i| = [\boldsymbol{\gamma}]_i, \\ \{0\}, & \text{if } |[c]_i| < [\boldsymbol{\gamma}]_i \end{cases} \tag{5}$$

and $[\cdot]_i$ denotes the i-th component of a vector.

Due to the non-convexity and discontinuous of ℓ_0 norm, it remains challenging to find an efficient and convergent numerical method to solve (4). A popular method for solving problem (4) is *iterative hard thresholding*. Note that this algorithm has already been considered in [25, 6, 7] for wavelet frame based image super-resolution and compressive sensing reconstruction [41]. The corresponding scheme for solving (4) is as follows

$$x_{k+1} \in \arg\min_{x\in\mathbb{R}^n} \|\boldsymbol{\lambda}\cdot x\|_0 + \frac{L}{2}\|x - x_k + \nabla f(x_k)/L\|^2 + \frac{\mu}{2}\|x-x_k\|^2, \tag{6}$$

where L is the Lipschitz constant of ∇f, which we refer to as *proximal iterative hard thresholding* (PIHT) algorithm.

Recently, ℓ_0-norm minimization problem has been attractive to researchers again. In particular, the proximal forward-backward splitting algorithm has been studied in [3, 10] for solving a general nonconvex and nonsmooth problem of the form

$$\min_x g(x) + f(x), \tag{7}$$

where g is lower semi-continuous and f is smooth. Lu [58] has studied the convergence behavior of PIHT algorithm. In [78], Zhang, Dong and Lu propose an adapted penalty decomposition (PD) method [59] to solve ℓ_0 norm based wavelet frame analysis model and demonstrate significant improvements of their method over some commonly used ℓ_1 minimization models. Dong and Zhang also propose the mean doubly augmented Lagrangian (MDAL) method [40] to solve ℓ_0 analysis based model and show that using ℓ_0-norm can generate higher quality images than ℓ_1-norm based methods by numerical experiments. However, the convergence analysis is not established. Besides, some other types of non-convex problems

and algorithms are also of great interests to researchers, see e.g.[65, 11, 30, 1, 20, 50, 52, 60, 61].

1.3. Our contributions

Motivated by the previous work on balanced approach with ℓ_1 regularization and recent progress on the convergence analysis of ℓ_0 regularization problem, we aim at developing an efficient and convergent numerical solver for the ℓ_0 norm based balanced approach. Due to the non-convexity and discontinuous of ℓ_0 regularization term, it is difficult to accelerate iterative hard thresholding method. With extrapolated step and modification, we propose *extrapolated proximal iterative hard thresholding* (EPIHT) algorithm and *EPIHT algorithm with line search* (EPIHT) algorithm for solving the ℓ_0 norm regularized problem. Under the theoretical framework of Kurdyka-Łojasiewicz property, we show that the sequence generated by EPIHT/EPIHT-LS algorithm converges to a local minimizer of the objective function with linear convergence rate. Finally, we show the performance of EPIHT and EPIHT-LS algorithms by applying them to compressed sensing and image restoration problems. Extensive computational results demonstrate that EPIHT and EPIHT-LS algorithms for ℓ_0 minimization models outperform typical methods for ℓ_1 minimization models (like FISTA, APG, etc) and PIHT algorithm for ℓ_0 minimization models in terms of solution quality and (or) number of iterations.

1.4. Organization

The rest of this paper is organized as follows. In section 2, we introduce the two algorithms EPIHT and EPIHT-LS. In section 3, we establish some convergence results about EPIHT and EPIHT-LS algorithms. In section 4, we propose several ℓ_0-norm based regularization models for some practical problems, and then solve them using EPIHT and EPIHT-LS algorithms. We compare these results with those of the ℓ_1-norm based regularization models solved by some prevailing algorithms for ℓ_1-norm minimization and PIHT algorithm for ℓ_0-norm minimization .

2. Proximal iterative hard thresholding methods

2.1. Model and algorithms

Given that ℓ_0-norm is an integer-valued, discontinuous and non-convex function, to ensure convergence of the proposed iterative scheme, similar as in [58, 71], we consider the following model

$$\min_{x\in\mathbb{R}^n} H(x) := \|\boldsymbol{\lambda}\cdot x\|_0 + F(x) + \frac{t}{2}\|x\|_2^2 = \|\boldsymbol{\lambda}\cdot x\|_0 + G(x), \tag{8}$$

where $t \geq 0$, $F(x)$ is smooth convex function. We emphasize that, the following relaxed ℓ_0 based wavelet frame balanced approach

$$\min_x \|\boldsymbol{\lambda}\cdot x\|_0 + \frac{1}{2}\|AW^\top x - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)x\|^2 + \frac{t}{2}\|x\|^2, \tag{9}$$

which will be used in section 3, is a special case of (8) when $F(x) = \frac{1}{2}\|AW^\top x - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)x\|^2$, $G(x) = F(x) + t\|x\|^2/2$.

Throughout this paper, our common **assumption** on problem (8) is: F is convex differentiable, bounded from below and ∇F is L-Lipschitz continuous.

Before presenting our algorithms, we define the surrogate function $S_\tau(x, y)$ of $H(x)$,

$$S_\tau(x, y) = \|\boldsymbol{\lambda}\cdot x\|_0 + G(y) + \langle \nabla G(y), x - y\rangle + \frac{\tau}{2}\|x - y\|^2, \tag{10}$$

where $\tau > 0$.

We may directly use PIHT to solve the problem (8), namely

PIHT Algorithm [58]

Choose parameters $\boldsymbol{\lambda} > 0, t \geq 0$, $0 < a < b < +\infty$; choose starting point x_0; let $k = 0$.
while k< maximum no. of iterations
Choose $\sigma_k \in (a, b)$ and compute

$$x_{k+1} \in \arg\min_{x \in \mathbb{R}^n} S_{L+t+\sigma_k}(x, x_k) \tag{11}$$

$k = k + 1$
end(while)

The step (11) is given by

$$x_{k+1} \in \mathcal{H}_{\sqrt{\frac{2\boldsymbol{\lambda}}{L+t+\sigma_k}}}(x_k - \frac{1}{L+t+\sigma_k}\nabla G(x_k)),$$

where $\mathcal{H}_{\sqrt{\frac{2\boldsymbol{\lambda}}{L+t+\sigma_k}}}(\cdot)$ is the hard thresholding operator defined in (5).

Note that for a general nonconvex and nonsmooth problem studied in [10]

$$\min_{x,y} s(x) + r(y) + g(x, y), \tag{12}$$

where s, r are lower semi-continuous functions and $g(x, y)$ is a smooth function, a proximal alternating linearized minimization (PALM) algorithm is introduced. When $r(y) \equiv 0$, $s(x) = \|\boldsymbol{\lambda} \cdot x\|_0$ and $g(x, y) = G(x)$, (12) reduces to (8), and PALM algorithm becomes PHIT algorithm. In [58], this algorithm was also studied for ℓ_0 regularized convex cone programming and the convergence to a local minimizer was established. In [3], an inexact forward-backward algorithm was studied for

$$\min_x s(x) + g(x)$$

where s is lower semi-continuous function and $\nabla g(x)$ is Lipschitz continuous. When $s(x) = \|\boldsymbol{\lambda} \cdot x\|_0$, the algorithm in [3] is an inexact version of PIHT.

The first algorithm that we propose is analogous to the extrapolation used in APG algorithm [5, 71] with adaptive restart [66] for convex case. We propose the following *extrapolated* PIHT (EPIHT) algorithm.

EPIHT Algorithm

Choose parameters $\boldsymbol{\lambda} > 0, t > 0$, $0 < a < b < +\infty$ and a sequence of extrapolation weight $0 < \omega_k \leq \omega < 1$; choose starting point $x_{-1} = x_0$; let $k = 0$.
while $k <$ maximum no. of iterations

$$y_{k+1} = x_k + \omega_k(x_k - x_{k-1}).$$

if $H(y_{k+1}) > H(x_k)$

$$y_{k+1} = x_k \tag{13}$$

end(if)

Choose $\sigma_k \in (a, b)$ and compute

$$x_{k+1} \in \arg\min_{x\in\mathbb{R}^n} S_{L+t+\sigma_k}(x, y_{k+1}) \tag{14}$$

$k = k + 1$

end(while)

We note that the inertial algorithm proposed in [65, 11] and a variable metric algorithm proposed in [30] look similar to EPIHT algorithm, but they are different in the following perspectives. Firstly, the algorithms in [11, 30] are proposed for minimizing the sum of a smooth function and a convex function while our EPIHT algorithm is proposed for minimizing the sum of a smooth convex function and ℓ_0-norm regularized term. Secondly, the algorithm in [11] uses a Bregman distance. when we take the Bregman distance function as $\frac{1}{2}\|\cdot\|^2$ and apply the algorithm to ℓ_0 regularization problem, the iterative scheme is

$$x_{k+1} \in \arg\min_{x\in\mathbb{R}^n} \|\boldsymbol{\lambda}\cdot x\|_0 + \frac{1}{4\alpha_k}(\|x - x_k + 2\alpha_k\nabla G(x_k)\|^2 + \|x - (x_k + 2\beta_k(x_k - x_{k-1})\|^2)$$

where $\alpha_k, \beta_k > 0$. The algorithm looks very similar to our EPIHT algorithm, except for the linearization is performed at different point and the setting for parameters α_k, β_k is also different. And for the algorithm proposed in [30], the extrapolation step is taken as $y_{k+1} = y_k + \omega_k(x_k - y_k)$, which is also different with ours.

The second algorithm that we proposed is to apply line search scheme to 'accelerate' the convergence speed. More specifically, we say the step size τ_k is acceptable if the following condition holds:

$$H(x_{k+1}) \le S_{\tau_k}(x_{k+1}, y_{k+1}) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2, \tag{15}$$

where $\sigma_k \in (a, b)$ and a, b are two positive constants. The extrapolated proximal iterative hard thresholding algorithm with line search (EPIHT-LS) is shown in the following.

EPIHT-LS Algorithm

Choose parameters $\boldsymbol{\lambda} > 0, t, \tau > 0, 0 < a < b < +\infty$ and a sequence of extrapolation weight $0 < \omega_k \le \omega < 1$; choose shrinking parameter $0 < \eta < 1$, starting point $x_{-1} = x_0$ and $\tau_{max} \ge \tau_{min} > 0$; let $k = 0$.

while $k <$ maximum no. of iterations

$$y_{k+1} = x_k + \omega_k(x_k - x_{k-1})$$

if $H(y_{k+1}) > H(x_k)$

$$y_{k+1} = x_k \tag{16}$$

end(if)

Choose $\sigma_k \in (a, b)$ and $\tau_0^k \in [\tau_{min}, \tau_{max}]$

Find the smallest integers $i_k \in \{0, 1, 2, \cdots\}$ such that with $\hat{\tau} = \tau_0^k/\eta^{i_k}$

$$H(\hat{x}) \le S_{\hat{\tau}}(\hat{x}, y_{k+1}) - \frac{\sigma_k}{2}\|\hat{x} - y_{k+1}\|^2$$

where $\hat{x} = \arg\min_x S_{\hat{\tau}}(x, y_{k+1})$.

Set $\tau_k = \tau_0^k/\eta^{i_k}$,

$$x_{k+1} = \arg\min_x S_{\tau_k}(x, y_{k+1}) \tag{17}$$

$k = k + 1$

end(while)

The following lemma shows that (15) is a well-defined condition.

Lemma 2.1. *If x_{k+1} is obtained via solving:*

$$x_{k+1} \in \arg\min_x S_{\tau_k}(x, y_{k+1}), \tag{18}$$

the condition (15) *holds whenever $\tau_k \geq L + t + \sigma_k$.*

Proof. Since ∇G is $L + t$ Lipschitz continuous, we have

$$G(x_{k+1}) \leq G(y_{k+1}) + \langle \nabla G(y_{k+1}), x_{k+1} - y_{k+1} \rangle + \frac{L+t}{2}\|x_{k+1} - y_{k+1}\|^2 \tag{19}$$

It implies

$$\begin{aligned} S_{\tau_k}(x_{k+1}, y_{k+1}) &= \|\boldsymbol{\lambda} \cdot x_{k+1}\|_0 + G(y_{k+1}) + \langle \nabla G(y_{k+1}), x_{k+1} - y_{k+1} \rangle + \frac{\tau_k}{2}\|x_{k+1} - y_{k+1}\|^2 \\ &\geq \|\boldsymbol{\lambda} \cdot x_{k+1}\|_0 + G(x_{k+1}) + \frac{\tau_k - L - t}{2}\|x_{k+1} - y_{k+1}\|^2 \\ &\geq H(x_{k+1}) + \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2, \end{aligned} \tag{20}$$

whenever $\tau_k \geq L + t + \sigma_k$. □

Remark 1. *For the choice of τ_0^k, we choose it as the one proposed by [4], namely*

$$\tau_0^k = \min\{\tau_{max}, \max\{\tau_{min}, \frac{\langle \nabla f(x_k) - \nabla f(x_{k-1}), x_k - x_{k-1} \rangle}{\|x_k - x_{k-1}\|^2}\}\}. \tag{21}$$

Remark 2. *It can be seen from the EPIHT-LS algorithm that $\tau_{min} \leq \tau_k \leq \max\{\tau_{max}, L + t + \sigma_k\}/\eta$.*

3. Convergence analysis

3.1. Preliminaries

Given any index set $I \subseteq \{1, 2, \ldots, n\}$, we let

$$C_I := \{x \in \mathbb{R}^n : [x]_i = 0 \text{ for all } i \in I\};$$

conversely, given any $x \in \mathbb{R}^n$, we define the zero element index set of a vector $x \in \mathbb{R}^n$ as

$$I(x) := \{i : [x]_i = 0\}. \tag{22}$$

Definition 3.1 (Subdifferentials [10])**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper and lower semi-continuous function.*

- *For a given $x \in \mathrm{dom} f$, the* Fréchet subdifferential *of f at x, denoted as $\hat{\partial} f(x)$, is the set of all vectors $u \in \mathbb{R}^n$ which satisfy*

$$\liminf_{y \neq x, y \to x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0.$$

When $x \notin \mathrm{dom} f$, we set $\hat{\partial} f(x) = \emptyset$.

- *The* limiting subdifferential *of f at $x \in \mathbb{R}^n$, denoted as $\partial f(x)$, is defined as*

$$\partial f(x) = \{u \in \mathbb{R}^n : \exists x_k \to x, f(x_k) \to f(x), \text{ and } u^k \in \hat{\partial} f(x_k) \to u \text{ as } k \to \infty\}$$

For more details about subdifferential of non-convex function, one can see [67, 62]

Definition 3.2 (Kurdyka-Łojasiewicz property [74])**.** *A function $\psi(x)$ satisfies the Kurdyka-Łojasiewicz (KL) property at point $\bar{x} \in \mathrm{dom}(\partial\psi)$ if there exists $\theta \in [0, 1)$, such that*

$$\frac{|\psi(x) - \psi(\bar{x})|^\theta}{\mathrm{dist}(0, \partial\psi(x))} \tag{23}$$

is bounded under the following conventions: $0^0 = 1$, $\infty/\infty = 0/0 = 0$. In other words, in a certain neighborhood $\mathcal{U}$ of $\bar{x}$, there exists $\phi(s) = cs^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$ such that the KL inequality holds:

$$\phi'(|\psi(x) - \psi(\bar{x})|)\mathrm{dist}(0, \partial\psi(x)) \geq 1, \ \forall x \in \mathcal{U} \cap \mathrm{dom}(\partial\psi) \text{ and } \psi(x) \neq \psi(\bar{x}).$$

where $\mathrm{dom}(\partial\psi) = \{x : \partial\psi(x) \neq \emptyset\}$ and $\mathrm{dist}(0, \partial\psi(x)) = \min\{\|y\| : y \in \partial\psi(x)\}$.

For more about Kurdyka-Łojasiewicz property, one can see [8, 9, 53, 2, 57] and references therein.

Lemma 3.3 (Ex.8.8c in [67])**.** *Assume $f = s + g$, where g is a C^1 function with Lipschitz gradient and s is proper and lower semi-continuous function. For all $x \in \mathrm{dom}(f)$, we have $\partial f(x) = \nabla g(x) + \partial s(x)$.*

In the next, we show that the every stationary point of the ℓ_0 norm related minimization which is defined in (8) is a local minimizer.

Lemma 3.4. *If $0 \in \partial f(x)$, where $f = s + g$, g is a convex and C^1 function, and $s(x) = \|\boldsymbol{\lambda} \cdot x\|_0$, then x is a local minimizer of f.*

Proof. By the Proposition 10.5 in [67], we know

$$0 \in \partial f(x) = \partial_{i \notin I(x)} f(x) \times \partial_{i \in I(x)} f(x).$$

Consequently, $[\nabla g(x)]_{i \notin I(x)} = 0$. Then for

$$\Delta x \in U := \{\Delta x \in \mathbb{R}^n : \|\Delta x\|_\infty < \min_{i \in I(x)} \{[\lambda]_i / |[\nabla g(x)]_i\}; |[\Delta x]_i| < |[x]_i|, \forall i \notin I(x)\},$$

by the convexity of g, we have

$$\begin{aligned} f(x + \Delta x) &= g(x + \Delta x) + \|\boldsymbol{\lambda} \cdot (x + \Delta x)\|_0, \\ &\geq g(x) + \langle \nabla g(x), \Delta x \rangle + \|\boldsymbol{\lambda} \cdot x\|_0 + \sum_{i \in \{i : [\Delta x]_i \neq 0, i \in I(x)\}} [\lambda]_i \\ &= f(x) + \sum_{i \in \{i : [\Delta x]_i \neq 0, i \in I(x)\}} ([\nabla g(x)]_i [\Delta x]_i + [\lambda]_i) \\ &\geq f(x). \end{aligned}$$

The last inequality holds because of $|[g(x)]_i [\Delta x]_i| \leq [\lambda]_i$ for any $\Delta x \in U$.

So x is a local minimizer of f. □

3.2. Global Convergence

This section is devoted to the global convergence property of EPIHT algorithm and EPIHT-LS algorithm. The convergence results of these two algorithms are almost identical, and their proofs are similar as well. Thus, we put these two convergence results together to form the main result of this section. For clarity, we first state this main result.

Theorem 3.5. *Let $H(x)$ be the objective function defined in (8), and $\{x_k\}_{k=0}^{\infty}$ be the sequence generated by EPIHT/EPIHT-LS algorithm, then there exists $\bar{x}$ such that $x_k \to \bar{x}$ as $k \to +\infty$ and $\bar{x}$ is a local minimizer of $H(x)$.*

Before giving the proof of the above theorem, we will first present a few technical lemmas.

Lemma 3.6. *Let $H(x)$ be the objective function defined in (8) and $\{x_k\}$ be the sequence generated by EPIHT/EPIHT-LS algorithm. Then, it has the following descent property:*

$$H(x_k) - H(x_{k+1}) \geq \rho_1 \|x_{k+1} - y_{k+1}\|^2, \tag{24}$$

where $\rho_1 > 0$. Moreover, if $k \to +\infty$, we have

$$\|x_k - y_k\| \to 0, \tag{25}$$

Proof. For EPIHT algorithm, from (10), (14), (13), we have

$$S_{L+t+\sigma_k}(x_{k+1}, y_{k+1}) \leq S_{L+t+\sigma_k}(y_{k+1}, y_{k+1}) = H(y_{k+1}) \leq H(x_k) \tag{26}$$

Combine (26) and Lemma 2.1, we know that

$$H(x_{k+1}) \leq S_{L+t+\sigma_k}(x_{k+1}, y_{k+1}) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2 \leq H(x_k) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2.$$

Similarly, for EPIHT-LS algorithm, from (10), (17) and (16), we have

$$S_{\tau_k}(x_{k+1}, y_{k+1}) \leq S_{\tau_k}(y_{k+1}, y_{k+1}) = H(y_{k+1}) \leq H(x_k). \tag{27}$$

Combine (27) and (15), we know that

$$H(x_{k+1}) \leq S_{\tau_k}(x_{k+1}, y_{k+1}) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2 \leq H(x_k) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2.$$

Thus, for the sequence $\{x_k\}$ generated either by EPIHT algorithm or by EPIHT algorithm, we arrive at the same inequality as follows

$$H(x_{k+1}) \leq H(x_k) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2. \tag{28}$$

Set $\rho_1 = \frac{a}{2} > 0$, the inequality (24) holds.
Sum up k from 0 to j in (24), we have

$$H(x_0) - H(x_{j+1}) \geq \rho_1 \sum_{k=0}^{j} \|x_{k+1} - y_{k+1}\|^2$$

Since $H(x_k)$ is decreasing and bounded below, we have

$$\sum_{j=0}^{+\infty} \|x_k - y_k\|^2 < +\infty,$$

which implies (25). □

In the next, we bound the distance between the limiting subgradient of $H(x_k)$ and the set of stationary points.

Lemma 3.7. *Let $H(x)$ be the objective function defined in (8) and $\{x_k\}$ be the sequence generated by the EPIHT/EPIHT-LS algorithm. Then, $\{x_k\}$ is bounded. Define*

$$s_k = \begin{cases} \nabla G(x_k) - \nabla G(y_k) - (L + t + \sigma_{k-1})(x_k - y_k) & \text{for EPIHT algorithm} \\ \nabla G(x_k) - \nabla G(y_k) - \tau_{k-1}(x_k - y_k) & \text{for EPIHT-LS algorithm} \end{cases},$$

we have $s_k \in \partial H(x_k)$ and

$$\|s_k\| \leq \rho_2 \|x_k - y_k\|, \tag{29}$$

where $\rho_2 = 2(L + t) + b > 0$.

Proof. Since $H(x)$ has coercive property and $H(x_k)$ is decreasing, we know $\{x_k\}$ is a bounded sequence. For EPIHT algorithm, by the optimal condition of (14) and Lemma 3.3, we have

$$0 \in \nabla G(y_k) + (L + t + \sigma_{k-1})(x_k - y_k) + \partial \|\boldsymbol{\lambda} \cdot x_k\|_0.$$

So, $s_k = \nabla G(x_k) - \nabla G(y_k) - (L + t + \sigma_{k-1})(x_k - y_k) \in \partial H(x_k)$. Similarly, for EPIHT-LS algorithm, by the optimal condition of (17) and Lemma 3.3, we have

$$0 \in \nabla G(y_k) + \tau_{k-1}(x_k - y_k) + \partial \|\boldsymbol{\lambda} \cdot x_k\|_0.$$

So, $s_k = \nabla G(x_k) - \nabla G(y_k) - \tau_{k-1}(x_k - y_k) \in \partial H(x_k)$.

Then, as ∇G is $L + t$-Lipschitz continuous, we know the inequality (29) holds. □

In the following, we summarize the properties for the limiting point set started from x^0. Define

$$\Omega(x^0) = \{x : \text{there exists a subsequece } \{x_{k_j}\} \text{ of } \{x_k\} \text{ such that } x_{k_j} \to x\}.$$

Lemma 3.8. *Let $H(x)$ be the objective function defined in (8) and $\{x_k\}$ be the sequence generated by EPIHT/EPIHT-LS algorithm. Then we have:*

- *for any $x \in \Omega(x^0)$, we have $0 \in \partial H(x)$;*
- *$\Omega(x^0)$ is a non-empty, compact and connected set.*
- *H is constant on $\Omega(x^0)$.*

Proof. Assume $\{x_{k_j}\}$ is a sub-sequence of $\{x_k\}$ which converges to $\bar{x}$. From (25), we know $y_{k_j} \to \bar{x}$ as $j \to +\infty$. From (14) and (18), we have

$$S_{\nu_k}(x_{k+1}, y_{k+1}) \leq S_{\nu_k}(\bar{x}, y_{k+1}), \tag{30}$$

where

$$\nu_k = \begin{cases} L + t + \sigma_k, & \text{for EPIHT algorithm,} \\ \tau_k, & \text{for EPIHT-LS algorithm.} \end{cases}.$$

Let $k = k_j - 1$ and $j \to +\infty$ in (30), as $\|x_k - y_k\| \to 0$ and ∇G is Lipschitz continuous, we have

$$\limsup_{j \to +\infty} \|\boldsymbol{\lambda} \cdot x_{k_j}\|_0 \leq \|\boldsymbol{\lambda} \cdot \bar{x}\|_0.$$

Together with the fact that $\|\cdot\|_0$ is lower semi-continuous, we have $\lim_{j\to+\infty}\|\boldsymbol{\lambda}\cdot x_{k_j}\|_0 = \|\boldsymbol{\lambda}\cdot\bar{x}\|_0$ and $\lim_{j\to+\infty} H(x_{k_j}) = H(\bar{x})$. From the definition of limiting subdifferential, Lemma 3.6 and Lemma 3.7, we know $0 \in \partial H(\bar{x})$.
The last two arguments are exactly the same as the proof of lemma 3.5 in [10]. □

In the next, we use Kurdyka-Łojasiewicz property to prove the global convergence for the sequence $\{x_k\}$ generated by EPIHT/EPIHT-LS algorithm.

Lemma 3.9 (Lemma 3.6, [10]). *Let B be a compact set and let σ be a proper and lower semi-continuous function. Assume that σ is constant on B and satisfies the Kurdyka-Łojasiewicz property at each point of B. Then, there exist $\epsilon > 0$, $\eta > 0$ and a concave $\psi : [0,\eta] \to \mathbb{R}_+$ with $\psi(0) = 0$, $\psi^{'}(s) > 0$ for all $s \in (0,\eta)$ and $\psi \in C^1$, continuous at 0, such that for all $\bar{u}$ in B and all u in the following intersection:*

$$\{u : \mathrm{dist}(u,\Omega) \le \epsilon\} \cap \{u : \sigma(\bar{u}) < \sigma(u) \le \sigma(\bar{u}) + \eta\},$$

one has, $\psi^{'}(\sigma(u) - \sigma(\bar{u}))\mathrm{dist}(0, \partial\sigma(u)) \ge 1$.

Now we are ready to give the proof of our main result in this section.
*Proof of **Theorem** 3.5* In the proof of Lemma 3.7, we know the sequence $\{x_k\}$ is bounded. Thus, there exists a subsequence $\{x_{k_j}\}$ such that $x_{k_j} \to \bar{x}$ as $j \to +\infty$. We assume that $H(x_k) > H(\bar{x})$. Otherwise, there exists some K, such that $x_k = x_K$ for all $k > K$ by the descent property of $H(x_k)$ and it is easy to show that $\bar{x}$ is a stationary point. By the fact $\lim_{k\to} H(x_k) = H(\bar{x})$, given $\eta > 0$, there exists K_0 such that $H(x_k) < H(\bar{x}) + \eta$ for all $k > K_0$. And from $\lim \mathrm{dist}(x_k, \Omega(x^0)) = 0$, we have for any $\epsilon > 0$, there exists K_1 such that $\mathrm{dist}(x_k, \Omega(x^0)) < \epsilon$ for all $k > K_1$.

Let $\ell = \max(K_0, K_1)$ and $\Omega(x^0)$ is nonempty and compact and f is constant on $\Omega(x^0)$. We can apply Lemma 3.9 to $B = \Omega(x^0)$, for any $k > \ell$, we have

$$\psi^{'}(H(x_k) - H(\bar{x}))\mathrm{dist}(0, \partial H(x_k)) \ge 1. \tag{31}$$

From Lemma 3.7, we have

$$\psi^{'}(H(x_k) - H(\bar{x})) \ge \frac{1}{\rho_2}\|x_k - y_k\|, \tag{32}$$

where $M > 0$. Meanwhile, as ψ is concave, we have

$$\psi(H(x_k) - H(\bar{x})) - \psi(H(x_{k+1} - H(\bar{x})) \ge \psi^{'}(H(x_k) - H(\bar{x}))(H(x_k) - H(x_{k+1})). \tag{33}$$

Define

$$\triangle_{p,q} := \psi(H(x_p) - H(\bar{x})) - \psi(H(x_q - H(\bar{x})).$$

From lemma 3.6, (32) and (33), there exists $c_0 > 0$, such that for $k > \ell$,

$$\triangle_{k,k+1} \ge \|x_{k+1} - y_{k+1}\|^2 / c_0\|x_k - y_k\|$$

and thus

$$2\|x_{k+1} - y_{k+1}\| \le \|x_k - y_k\| + c_0\ \triangle_{k,k+1} \tag{34}$$

by Cauchy-Schwarz inequality. Summing (34) over i, we have

$$2\|x_{k+1} - y_{k+1}\| + \sum_{i=l+1}^{k}\|x_i - y_i\| \le \|x_\ell - y_\ell\| + C\ \triangle_{\ell+1,k+1},$$

as $\triangle_{p,q} + \triangle_{q,r} = \triangle_{p,r}$. Then, for any $k > \ell$,

$$\sum_{i=\ell+1}^{k} \|x_{i+1} - y_{i+1}\| \leq \|x_\ell - y_\ell\| + C\psi(H(x_{\ell+1}) - H(\bar{x})). \tag{35}$$

Therefore,

$$\begin{aligned} \sum_{i=\ell+1}^{k} \|x_{i+1} - y_{i+1}\| &\geq \sum_{i=\ell+1}^{k} (\|x_{i+1} - x_i\| - \omega_i \|x_i - x_{i-1}\|) \\ &\geq \sum_{i=\ell+1}^{k} (1-\omega)\|x_{i+1} - x_i\| - \omega\|x_\ell - x_{\ell-1}\| \end{aligned} \tag{36}$$

Putting (35) and (36) together, it is easy to see that the sequence $\{x_k\}$ is globally convergent, and thus $x_k \to \bar{x}$. By Lemma 3.4 and Lemma 3.8, we know $\bar{x}$ is a local minimizer of the function $H(x)$. □

In the following, we establish the convergence rate results for the two algorithms.

3.3. Convergence rate of the proposed EPIHT/EPIHT-LS algorithm

Lemma 3.10. *Let $H(x)$ be the objective function defined in* (8) *and $\{x_k\}$ be the sequence generated by the EPIHT/EPIHT-LS algorithm. Then, there exists $k_0 > 0$, such that $I(x_k) = I(x_{k_0})$ for all $k > k_0$.*

Proof. The result holds as $\lim_{k\to+\infty} x_k = \bar{x}$, $\lim_{k\to+\infty} \|x_k\|_0 = \|\bar{x}\|_0$ and $|[x_k]_i| > \rho > 0$ if $i \notin I_k$. □

Lemma 3.11. *Let $H(x) = G(x) + \|\boldsymbol{\lambda} \cdot x\|_0$ be the objective function defined in* (8)*, $\{x_k\}$ be the sequence generated by the EPIHT/EPIHT-LS algorithm and $H(x, I) = G(x) + \mathcal{X}_{C_I}(x)$ where $\mathcal{X}_{C_I}(x)$ is the indicator function of set C_I. There exist $k_0 > 0$, such that for all $k > k_0$ we have $x_k = z_{k-k_0}$ where $\{z_k\}$ is a sequence generated by EPIHT/EPIHT-LS algorithm for minimizing $H(x, I(x_{k_0}))$ which starts from x_{k_0} and $\bar{x}$ is the unique global minimizer of $H(x, I(x_{k_0}))$.*

Proof. From lemma 3.10, there exists k_0 such that $I(x_k) = I(x_{k_0})$ for all $k > k_0$. So, we have

$$x_{k+1} \in \arg\min_x S_{\tau_k}(x, y_{k+1}) = \mathcal{H}_{\tau_k}(y_{k+1} - \nabla G(y_{k+1})/\tau_k) = P_{C_{I(x_{k_0})}}(y_{k+1} - \nabla G(y_{k+1})/\tau_k).$$

We have known that $\{x_k\}, \{y_k\}$ are convergent, $x_k - y_k \to 0$ and $\{\tau_k\}$ have positive lower and upper bound. So, let $k \to +\infty$, we have $\bar{x} = P_{C_{I(x_{k_0})}}(\bar{x} - \nabla G(\bar{x})/\bar{\tau})$ where $\bar{\tau}$ is a accumulation point of $\{\tau_k\}$. Then we can conclude that $z_{k-k_0} = x_k \in P_{C_{I(x_{k_0})}}(y_{k+1} - \nabla G(y_{k+1})/\tau_k)$ and $\bar{x}$ is the unique global minimizer of $H(x, I(x_{k_0}))$ as $H(x, I(x_{k_0}))$ is strongly convex. □

Lemma 3.12 ([74])**.** *Given an index set I. Let $H(x, I) = G(x) + \mathcal{X}_{C_I}(x)$ where $G(x)$ is defined in* (8)*. Then, $H(x, I)$ satisfies KL property with $\psi(s) = \frac{2}{t}\sqrt{s}$.*

Proof. The proof is given in section 2.2 in [74] as $G(x)$ is t-strongly convex. □

Theorem 3.13 (Convergence rate)**.** *Let $H(x)$ be the objective function defined in* (8) *and $\{x_k\}$ be the sequence generated by the EPIHT/EPIHT-LS algorithm. Then, there exist k_0, $C > 0$ and $\tau \in [0, 1)$ such that*

$$\|x_k - \bar{x}\| \leq C\tau^k, \ \forall k > k_0.$$

Proof. By lemma 3.11, there exists k_0, such that for all $k > k_0$, $\{x_k\}$ is a sequence generated by the EPIHT/EPIHT-LS algorithm for minimizing $H(x, I(x_{k_0}))$ and $\bar{x}$ is the unique global minimizer of $H(x, I(x_{k_0}))$. Without loss of generality, we assume that $H(\bar{x}, I(x_{k_0})) = 0$ and $H(x_k, I(x_{k_0})) > 0$. Define $\Delta_k = \sum_{i=k}^{\infty} \|x_i - x_{i+1}\|$. Then, from $H(x_k, I(x_{k_0})) = H(x_k) - n + |I(x_{k_0})|$ for all $k > k_0$, (35) and (36), we have

$$\begin{aligned} \Delta_k &\leq C\psi(H(x_k, I(x_{k_0}))) + \frac{1}{1-w}(\Delta_{k-1} - \Delta_k) + w(\Delta_{k-2} - \Delta_{k-1}) \\ &\leq C\psi(H(x_k, I(x_{k_0}))) + \frac{1}{1-w}(\Delta_{k-2} - \Delta_k) \end{aligned} \tag{37}$$

since $0 < w < 1$ and $\Delta_{k-2} \geq \Delta_{k-1}$. From (32), lemma 3.12 and $\psi(s) = \frac{2}{t}\sqrt{s}$, we have

$$\frac{1}{t}(H(x_k, I(x_{k_0})))^{-1/2} \geq (\rho_2\|x_k - y_k\|)^{-1} \geq (\rho_2(\|x_k - x_{k-1}\| + w\|x_{k-1} - x_{k-2}\|))^{-1}. \tag{38}$$

Then, it implies

$$\psi(H(x_k, I(x_{k_0}))) = \frac{2}{t}(H(x_k, I(x_{k_0})))^{1/2} \leq \frac{2}{t}\frac{1}{t}(\rho_2(\|x_k - x_{k-1}\| + w\|x_{k-1} - x_{k-2}\|)). \tag{39}$$

Let $C_1 = C\frac{2}{t^2}\rho_2$ and $C_2 = \frac{1}{1-w}$, (37) and (39) imply

$$\Delta_k \leq C_1(\Delta_{k-2} - \Delta_k) + C_2(\Delta_{k-1} - \Delta_k) = (C_1 + C_2)(\Delta_{k-2} - \Delta_k). \tag{40}$$

It implies $\Delta_k \leq \frac{C_1+C_2}{1+C_1+C_2}\Delta_{k-2}$. As $\|x_k - \bar{x}\| \leq \Delta_k$, we have $\|x_k - \bar{x}\| \leq C\tau^k$ for $k > k_0$ where $C = \Delta_0$ and $\tau = \sqrt{\frac{C_1+C_2}{1+C_1+C_2}}$. □

4. Numerical implementation

In this section, we will show some numerical results of EPIHT and EPIHT-LS algorithms for solving some ℓ_0 minimization problems of the form (8), and compare them with the results of PIHT algorithm and some previous algorithms designed for ℓ_1-norm minimization in the literature (like FISTA, APG, etc). For set valued operator $\mathcal{H}$ used in EPIHT and EPIHT-LS algorithms, we set $[\mathcal{H}_{\boldsymbol{\gamma}}(x)]_i = \{0\}$ if $|[x]_i| \leq [\gamma]_i$ in all numerical experiments. In section 4.1, we apply EPIHT and EPIHT-LS algorithms to the compressive sensing problem. Our main interest is on the application of EPIHT (as well as EPIHT-LS) algorithm to the wavelet frame based image restoration problems. In section 4.2, we will perform the test on the simulated CT reconstruction and image deblurring. All the experiments are conducted in MATLAB using a desktop computer equipped with a 4.0GHz 8-core AMD processor and 16GB memory.

4.1. Compressive sensing

Compressive sensing problem can be formulated as the following regularization problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2 + g(x)$$

where $A \in R^{m\times n}$ is a data matrix, $b \in R^m$ is an observation vector, and $g(x)$ is the regularization term.

With $\lambda_1, \lambda_2 > 0$, we consider the regularization problem

$$\min_{x\in\mathbb{R}^n} H(x) := \frac{1}{2}\|Ax - b\|^2 + \lambda_1\|x\|_0 + \frac{\lambda_2}{2}\|x\|^2 \tag{41}$$

to be analogue with the Naïve elastic net model [79]

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|^2 + \lambda_1\|x\|_1 + \frac{\lambda_2}{2}\|x\|^2. \tag{42}$$

In numerical experiment, the data matrix $A \in \mathbb{R}^{m\times n}$ is a Gaussian random matrix (mean 0 and standard deviation 1) and the columns of A are normalized to have ℓ_2 norm of 1. We fix the number of rows of A to be $m = 500$ and vary the number of columns n as well as the sparsity parameter s of original signal $\bar{x}$. For each combination of (n, s), we generate the original signal $\bar{x} \in \mathbb{R}^n$ containing s randomly placed ± 1 spikes. The observed data $b \in \mathbb{R}^m$ is generated by

$$b = Ax + \eta,$$

where η is a white Gaussian noise of variance 10^{-4}.

For parameters, we choose $\lambda_2 = 10^{-7}, L = \lambda_{max}(A^\top A + \lambda_2 I)$ and the stopping criteria to be

$$\frac{\|x_k - x_{k-1}\|}{\max\{1, \|x_k\|\}} < tol.$$

with $tol = 10^{-6}$ for all algorithms (namely FISTA, PIHT, EPIHT, EPIHT-LS). The parameter λ_1 is chose according to low relative error $\frac{\|x-\bar{x}\|}{\|\bar{x}\|}$ for the four alsorithms. And for EPIHT algorithm, we choose $\lambda_1 = 0.1$, $\omega_k \equiv 0.8$, $\sigma_k \equiv 0.001$. For EPIHT-LS algorithm, Let $\lambda_1 = 0.1$, $\sigma_k \equiv 0.001$, we choose τ_0^k as (21) where $\tau_{min} = 1e-2L, \tau_{max} = 10L$ and let ω_k be: $\omega_0 = 0.6$; for $k \geq 1$, if $H(x_k) < H(y_{k+1})$, we let $\omega_{k+1} = \max(\omega_k/2, 0.1)$; otherwise, we let $\omega_{k+1} = \min(1.1\omega_k, 0.9)$. We choose $\lambda_1 = 0.1$, $\mu = 0.001$ for PIHT algorithm and $\lambda_1 = 0.001$ for FISTA algorithm. From the convergence analysis in section 3, we know that algorithms for solving ℓ_0 regularization problem can merely guarantee local convergence. Thus, we firstly run FISTA algorithm with $x_0 = A^T b, tol = 10^{-2}$ to get an initial point for PIHT,EPIHT, EPIHT-LS algorithms (the corresponding runtime and iteration number are added in the results of the three algorithms). For each algorithm and each choice of (n, s), we run our experiment 50 times to guarantee that the result is independent of any particular realization of the random matrix and original signal $\bar{x}$. We record the average runtime, the average relative error $\frac{\|x-\bar{x}\|}{\|\bar{x}\|}$ to the original signal $\bar{x}$, the average number of iteration the algorithm needed and their standard variance. The numerical results are listed in Table 1, 2, 3. The bold numbers are the best results.

One may observe from Table 2 that, compared to FISTA, the three ℓ_0-based algorithms, namely PIHT, EPIHT and EPIHT-LS, can always reach solutions of higher precision. The precision of solutions between PIHT, EPIHT and EPIHT-LS is similar. However, we can see from Table 1, 3 that EPIHT enjoys better iteration complexity and less CPU time than PIHT, which shows the advantage of extrapolation. And EPIHT-LS algorithms enjoys better iteration complexity than EPIHT, which shows the advantage of line search. In figure 1 we show the evolution of τ_k in EPIHT-LS algorithm, where we can see that line search leads to a relatively 'larger' stepsize than the uniform parameter used in EPIHT.

4.2. Image restoration

In this subsection, we conduct several numerical experiments on the image restoration problem to demonstrating the improvement of the proposed ℓ_0-norm based regularization algorithm (9) against traditional ℓ_1-norm based regularization methods.

In implementation, the sparsity-promoting weight vector $\boldsymbol{\lambda}$ is set in the following manner: for any position p of the original image vector, one has (i) $\boldsymbol{\lambda}_0[p] \equiv 0$, and (ii) $\boldsymbol{\lambda}_{i,j}[p] = \lambda \cdot w^{i-1}$, $i = 1, \ldots, Q$, $j =$

s	n	Average number of iteration/Standard variance			
		FISTA	PIHT	EPIHT	EPIHT-LS
	1000	241.0/5.8429	69.5/3.6398	31.5/**0.5741**	**19.7**/0.8775
	1500	356.0/8.2048	90.9/5.1069	39.4/2.7785	**24.5/1.2530**
$\lfloor \frac{n}{100} \rfloor$	2000	468.8/12.4092	112.9/6.2949	49.1/3.4504	**29.3/1.0698**
	2500	574.6/14.6956	137.4/8.3166	58.1/4.0676	**34.6/1.5875**
	3000	683.6/18.8052	165.7/9.6504	69.3/5.8393	**40.3/1.5951**
	1000	294.7/8.2831	78.6/4.6737	37.0/2.7820	**23.4/0.8738**
	1500	441.6/9.9255	109.0/5.3367	49.5/1.4999	**30.8/1.2093**
$\lfloor \frac{2n}{100} \rfloor$	2000	591.4/15.8275	143.8/7.4423	62.5/2.5475	**39.2/1.3416**
	2500	738.3/25.3915	181.2/10.0174	78.7/3.8450	**48.5/1.9723**
	3000	908.6/41.9853	227.4/10.0060	94.6/6.5424	**60.0/2.5057**

Table 1: Results of the average and the standard variance of iterations number.

s	n	Average relative error ($\times 10^{-3}$)/Standard variance ($\times 10^{-3}$)			
		FISTA	PIHT	EPIHT	EPIHT-LS
	1000	1.0342/0.0731	0.1002/0.0184	**0.1000/0.0182**	0.1001/0.0183
	1500	1.0327/0.0628	**0.0934**/0.0192	**0.0934/0.0189**	0.0936/0.0190
$\lfloor \frac{n}{100} \rfloor$	2000	1.0636/0.0661	0.1024/0.0154	0.1014/0.0154	**0.1012/0.0153**
	2500	1.0800/0.0749	0.1009/0.0153	**0.0997**/0.0150	0.0998/**0.0147**
	3000	1.1038/0.0783	0.1002/0.0104	0.0993/**0.0100**	**0.0992/0.0100**
	1000	1.0707/0.0783	0.0988/0.0179	0.0984/0.0177	**0.0983/0.0175**
	1500	1.0993/0.0674	0.1037/0.0130	**0.1031**/0.0124	**0.1031/0.0123**
$\lfloor \frac{2n}{100} \rfloor$	2000	1.1499/0.0671	0.1015/0.0117	**0.1009/0.0111**	**0.1009/0.0110**
	2500	1.2121/0.1041	0.1048/0.0087	0.1028/0.0085	**0.1027/0.0084**
	3000	1.2813/0.1019	0.1112/0.0089	0.1090/**0.0080**	**0.1086**/0.0082

Table 2: Results of the average and the standard variance of the relative error.

s	n	Average runtime/Standard variance			
		FISTA	PIHT	EPIHT	EPIHT-LS
	1000	0.1796/0.0078	0.0549/**0.0095**	**0.0448/0.0095**	0.5525/0.0213
	1500	0.6653/0.0164	0.1764/0.0139	**0.1012/0.0116**	1.1333/0.0169
$\lfloor \frac{n}{100} \rfloor$	2000	1.6247/0.0461	0.3898/0.0223	**0.2182/0.0174**	1.9609/0.0228
	2500	2.8274/0.0930	0.6743/0.0427	**0.3581/0.0303**	3.0107/0.0467
	3000	5.2666/0.1473	1.2721/0.0764	**0.6414/0.0580**	4.3981/0.0641
	1000	0.2194/0.0100	0.0586/**0.0080**	**0.0498**/0.0120	0.5580/ 0.0168
	1500	0.8265/0.0210	0.2088/0.0112	**0.1267/0.0111**	1.1752/0.0180
$\lfloor \frac{2n}{100} \rfloor$	2000	2.0471/0.0559	0.4962/0.0267	**0.2730/0.0178**	2.0768/0.0324
	2500	3.6347/0.1302	0.8856/0.0501	**0.4796/0.0277**	3.2254/0.0533
	3000	7.0085/0.3229	1.7504/0.0764	**0.8663/0.0657**	4.8209/0.0872

Table 3: Results of the average and the standard variance of the runtime.

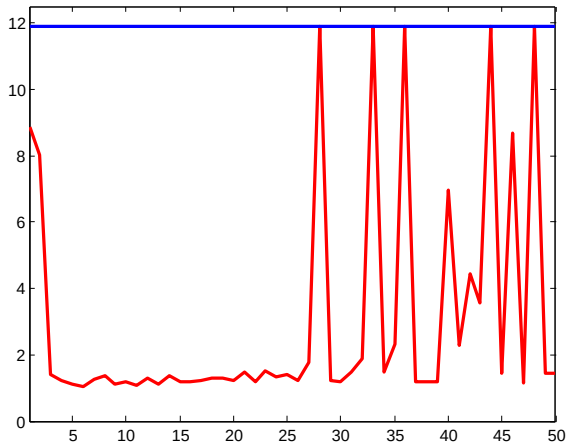


Figure 1: Horizontal axis represents the iteration number k, while the vertical axis represents τ_k. The red line represents τ_k of EPIHT-LS algorithm; the blue line represents global Lipschits constant L.

$1, \ldots, J$, where λ is a positive constant, i indicates the level of wavelet decomposition, j stands for the index of high-pass filters of the adopted wavelet tight frame system, and $w \in (0, 1)$ is a penalization-decreasing weight (i.e. the penalizing weight decreases as the level of decomposition increases). In experiments we use *isotropic* wavelet regularization for comparison since it usually leads to restoration results with better image quality.

From [71], we have

$$\|\nabla F(x) - \nabla F(y)\| \leqslant \bar{L}\|x - y\|, \forall x, y, \text{ for any } \bar{L} \geq \kappa + \lambda_{max}(A^*A).$$

Thus for EPIHT algorithm, we use $L = \kappa + \lambda_{max}(A^\top A)$ as the Lipschitz constant of ∇F. For other parameters related to the two algorithms, if not specified, we use $t = 10^{-5}$ and $\sigma_k = 0.1$ for all iterations, and the initial extrapolation weight ω_0 is set to be 0.7, and other ω_k's are set inductively by: if $H(x_k) < H(y_{k+1})$, we let $\omega_{k+1} = \max(0.8 * \omega_k, 0.1)$; otherwise, we let $\omega_{k+1} = \min(\omega_k/0.8, 0.9)$.

As our previous theoretical results imply, the proposed iterative algorithms can merely generate sequences that convergence to local minimizers. For this reason, initialization plays an important role, and it may determine the local minimizers to which the generated sequences will convergence. To guarantee the proposed algorithms will reach 'good' local minimizers, we take the solution (computed by APG) of the ℓ_1-balanced model for both EPIHT algorithm and EPIHT-LS algorithm, and we also count in the time for computing their initial points for EPIHT and EPIHT-LS algorithms.

4.2.1. CT Image Reconstruction

For simplicity, the test is merely performed on simulated data. For CT image reconstruction, degradation matrix A happens to be a projection matrix. The test image is a head phantom generated by MATLAB (version 8.1.0.604) built-in functions. After the construction of the projection matrix A, we then add some Gaussian noise with variance σ to the projected data to obtain the observed data f. In our simulations, we set $\sigma = 0.01 \cdot \|f\|_\infty$.

The results of EPIHT-LS are compared with other two popular algorithms for wavelet tight frame

based image restoration: analysis based approach

$$\min_u \frac{1}{2}\|Au-b\|_D^2 + \|\boldsymbol{\lambda}\cdot(Wu)\|_1. \tag{43}$$

and balanced approach (1). We use the *accelerated proximal gradient (APG)* algorithm (see e.g. [71]) to solve the balanced approach, while split Bregman algorithm [17, 47] is adopted for solving the analysis based approach. For both iterative schemes that solve the balanced approach and (9), similar as in [78], we use the following stopping criteria

$$\min\left\{\frac{\|x_k - x_{k-1}\|}{\max\{\|x_k\|,1\}}, \frac{\|AW^\top x_k - f\|_D}{\|f\|}\right\} < 4\times 10^{-5}.$$

As to the split Bregman algorithm for solving the analysis based approach (1), the following stopping criteria is adopted

$$\frac{\|Wu_k - x_k\|_D}{\|f\|} < 4\times 10^{-5}.$$

For both the balanced approach (1) and the proposed model in (9), we commonly set $\kappa = 2$. For all three methods involved, the preconditioning matrix D is set to be the identity matrix for simplicity. The wavelet tight frame transform W used in this simulation is the one generated by $2D$ tensor-product Haar wavelet tight frame system, and the level of wavelet decomposition is set to be 4.

We perform the test on the 128×128 Shepp-Logan phantom. In simulations one can control the number of projections, which will result in projection matrices with different row sizes. In our experiment we use 20 and 50 projections for illustration, and the results are summarized in the following Figure 2.

One may observe from Figure 2 that the ℓ_0-based algorithms achieve better reconstruction quality measured by PSNR values. We exclude the results of the EPIHT algorithm since they are very close to those of the PHIT algorithm under our experimental settings. For a comparison between PIHT and EPIHT-LS, the former one takes less processing time, while the latter one has slightly better PSNR values at the time when the same stopping criteria is satisfied. In fact, the longer processing time of EPIHT-LS is mainly due to the additional function-value evaluation step as indicated in its implementation. To demonstrate this fact, we run the experiment with 2000 steps and plot the energy (i.e. value of the objective function with respect to the current x_k) evolving curves of PIHT and EPIHT-LS under different parameter settings for displaying the actual acceleration of EPIHT-LS with respect to the number of iterations. The two plots in Figure 3 show that, at least within the first few iterations, the EPIHT-LS algorithm indeed accelerates the decreasing rate of objective function in (9) compared to PIHT, and it also reaches a local minimizer of the objective function with slightly lower objective value. In Figure 4, we also show the evolving energy curves in terms of runtime. We can see that the three methods share almost the same energy curves.

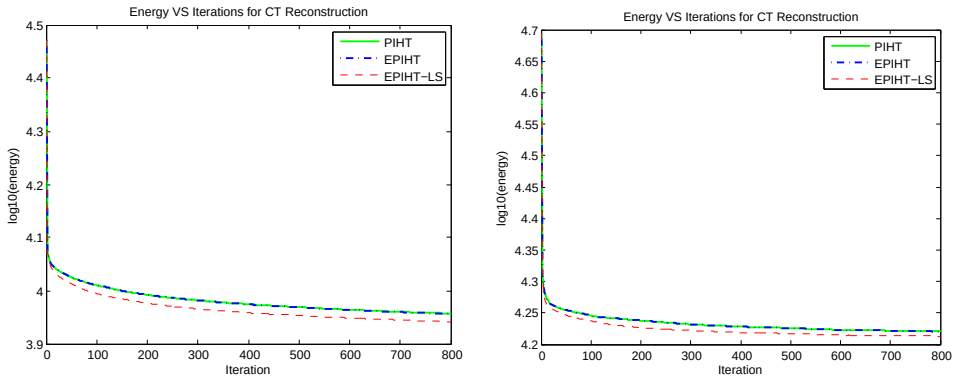
4.2.2. Image deblurring

In this subsection, we apply the proposed model (9) to the problem of image deblurring, and compare the results with those of the analysis based approach in (43), balanced approach (1) and the ℓ_0 approximation model SNC (ii) (For a fair comparison, we use wavelet tight frame basis instead of total variation)

$$\min F_\delta(u) := \frac{1}{2}||Au-b||^2 + \boldsymbol{\lambda}\cdot\phi(Wu) + \frac{t}{2}||u||^2 + \frac{\theta}{2}d^2_{[0,255]}(u)$$

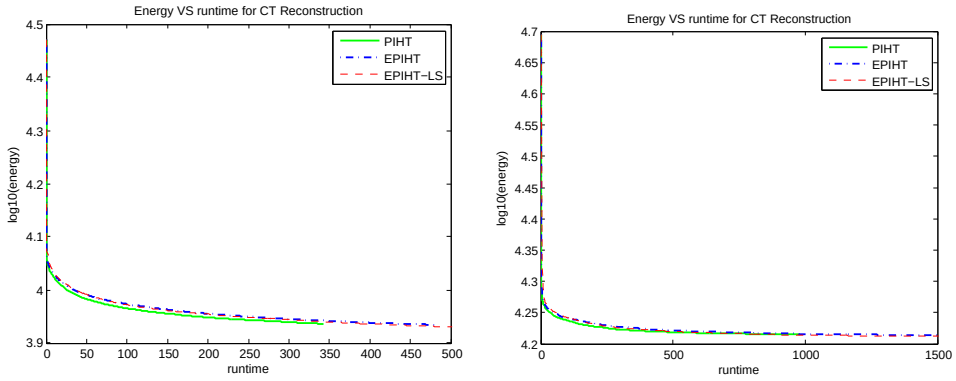
	Split Bregman for (43)	APG for (1)	PIHT for (9)	EPIHT-LS for (9)
20 Projections				
PSNR	22.98	27.77	28.96	29.52
Time	39.67	70.45	135.62	172.61
	Split Bregman for (43)	APG for (1)	PIHT for (9)	EPIHT-LS for (9)
50 Projections				
PSNR	26.17	32.34	35.15	35.88
Time	22.27	58.60	133.07	178.03

Figure 2: CT image reconstruction results. The first row shows the reconstruction results using 20 projections, and the second row shows the reconstruction results using 50 projections. The running time is listed in *seconds*.



(a) (b)

Figure 3: Energy evolving curves of CT image reconstruction, (a) for 20 projections, (b) for 50 projections.



(a) (b)

Figure 4: Energy evolving curves of CT image reconstruction, (a) for 20 projections, (b) for 50 projections.

where $\phi(z) = \frac{z^2}{2\delta^2+z^2}, z \in \mathbb{R}$ and $d_{[0,255]}(u)$ is the distance between u and interval $[0, 255]$. When $\delta \to 0^+$, $\phi(z)$ becomes ℓ_0 regularization. The above SNC (ii) is solved by Majorize-Minimize Memory Gradient (3MG) method [29]), it obtains the next iterative point by $u_{k+1} = u_k + T_k\alpha_k$ where $T_k = [d_k^1, \ldots, d_k^M]$ is some suitable search direction matrix and α_k is a multivariate step size. The parameters in SNC (ii) are $t = 10^{-5}$ $\delta = 0.1$ (seems optimal) and $\boldsymbol{\lambda}$ is chose by the same way with our algorithms. The initial point for all the algorithms and models are the same. In this simulation, the degradation matrix A is the convolution matrix of a Gaussian function (generated by 'fspecial(9,1.5)' in MATLAB), i.e. the image is blurred by Gaussian kernel. The blurred image is further corrupted by some Gaussian noise with variance σ ($\sigma = 3$ if not specified). For both balanced approach (1) and the proposed model (9), the value of κ is fixed to be 1. Same as in the case of CT image reconstruction, the iterative solver we have adopted for the analysis approach is the split Bregman algorithm, while the balanced approach is solved by the APG algorithm. Besides, the 2D tensor-product piecewise linear wavelet tight system is adopted for the generating the transforming operator W, and the level of wavelet decomposition is set to be 4 (which in most cases seemingly leads to the best reconstruction results). Moreover, we use the conditioning matrix $D := (A^\top A + 2\sigma)^{-1}$ for all 3 approaches in order to facilitate their convergence. The stopping criteria we adopt for both EPIHT-LS and APG is

$$\min\left\{\frac{\|x_k - x_{k-1}\|}{\max\{\|x_k\|, 1\}}, \frac{\|AW^\top x_k - f\|_D}{\|f\|}\right\} < 10^{-4}.$$

And for the split Bregman algorithm for solving the analysis based approach, the following stopping criteria is adopted

$$\frac{\|Wu_k - x_k\|_D}{\|f\|} < 10^{-4}.$$

The stopping criteria for 3MG algorithm is

$$\|\nabla F_\delta(u)\| \leq \min(size(u))10^{-3}$$

where u is the observed image.

In our experiments, we test all four methods on twelve different images (their initial PSNR and SSIM values are listed in Table 4). These results are fully summarized in Table 5, where the size of each image is in the small bracket following its name. The restored images are quantitatively evaluated by their

peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Furthermore, we show some zoom-in views of original images, degraded images and restored images in Figure 5 and Figure 6, so that man can evaluate the visual quality of the restoration results as well. To guarantee a fair comparison of all three methods, we have manually tuned up the parameter λ, so that best quality of the restoration images for each individual method is (approximately) achieved.

One may observe from Table 5 that, the EPIHT-LS method shows certain extent of improvement in PSNR values compared to the other two ℓ_1-norm based minimization methods and the ℓ_0 approximation model SNC(ii). The proposed ℓ_0 minimization method is consistently the best in terms of SSIM values(the closer to 1 the better) except one cases, which due to the fact that the sharpness of edges are better preserved by hard thresholding. One can also see from the zoom-in views in Figure 5 and Figure 6 that the ℓ_0 balanced approach produces cleaner (seemingly with less artifacts) restoration results compared to the other three approaches.

Image	dimesion	initial PSNR	initial SSIM
boat	512×512	26.71	0.8887
bridge	512×512	24.53	0.8485
cameraman	256×256	23.72	0.7148
clock	256×256	26.07	0.7997
couple	512×512	26.52	0.8770
goldhill	512×512	28.78	0.8967
house	256×256	28.76	0.7693
jetplane	512×512	27.70	0.9248
Lena	512×512	29.62	0.9283
peppers	512×512	31.05	0.9466
pirate	512×512	27.74	0.8969
woman	512×512	27.36	0.9122

Table 4: The initial PSNR and SSIM values of testing images.

5. Conclusions and perspectives

In this paper, we have studied two algorithms, namely extrapolated proximal iterative hard thresholding (EPIHT) and extrapolated proximal iterative hard thresholding with line-search (EPIHT-LS), for solving the ℓ_0-norm of wavelet frame approach (9) for image restoration. We provide a general convergence analysis for the general ℓ_0 regularization problem (8) by assuming $t > 0$ and $G(x)$ is convex with L-Lipschitz continuous gradient map. In particular, we have proved the global convergence of the proposed algorithms, and the limiting point of the generated sequence must be a local minimizer of the objective function $H(x)$ defined in (8). We further show in our numerical experiments that, for solving ℓ_0-norm based regularization models, the proposed EPIHT and EPIHT-LS schemes can yield better results and image quality compared to soft thresholding scheme for solving the ℓ_1-norm based models (including FISTA, APG, etc). The computational results also suggest that, in some cases (i) EPIHT algorithm is enjoys a 'faster' convergence rate compared to the PIHT algorithm proposed in [58], and (ii) line search is helpful for identifying faster local solutions of the proposed ℓ_0 regularization model. However, the essence

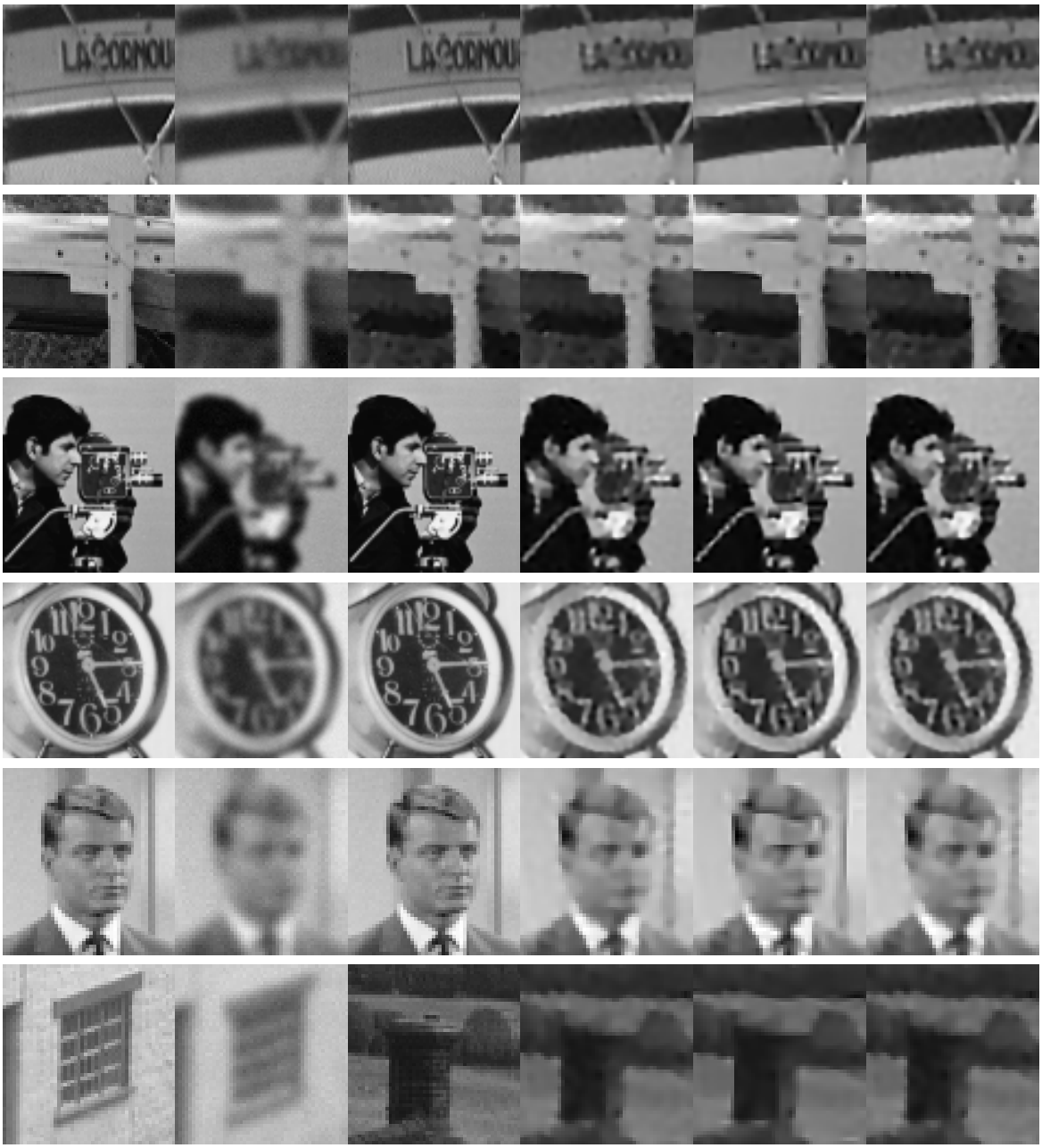


Figure 5: Zoom-in views of the processed results. From left to right: original images, noisy and blurred images, results of analysis based approach in (43), results of balanced approach (1), results of EPIHT-LS for (9), and results of 3MG [29].

Figure 6: Zoom-in views of the processed results. From left to right: original images, noisy and blurred images, results of analysis based approach in (43), results of balanced approach (1), results of EPIHT-LS for (9), and resdults of SNC(ii).

Image	Split Bregman time/PSNR/SSIM	APG time/PSNR/SSIM	EPIHT-LS time/PSNR/SSIM	3MG time/PSNR/SSIM
boat	**18.3**/29.66/0.9426	22.1/29.71/0.9435	349.3/29.87/**0.9504**	24.4/**29.80**/0.9470
bridge	**13.7**/26.85/0.9225	25.6/26.90/0.9270	139.8/26.84/0.9310	28.4/**26.97**/**0.9328**
cameraman	**7.4**/ 26.97/0.8418	8.8/26.96/0.8380	78.6/**27.00**/**0.8480**	9.4/26.99/0.8357
clock	**6.2**/29.49/0.9159	6.5/29.47/0.9102	69.4/ **29.91**/**0.9229**	7.1/29.52/ 0.9023
couple	**16.1**/29.21/0.9347	22.1/29.26/0.9367	264.8/ **29.40**/**0.9458**	24.5/29.35/0.9424
goldhill	**20.6**/30.70/0.9335	22.1/30.80/0.9380	193.3/ 30.87/**0.9423**	20.8/**30.90**/**0.9423**
house	**7.0**/32.37/0.8593	6.3/32.31/ 0.8574	43.1/ **32.45**/ **0.8638**	7.1/32.39/0.8540
jetplane	**18.4**/32.16/0.9595	22.1/32.21/0.9591	116.9/ 31.90/**0.9636**	24.2/**32.32**/0.9598
Lena	**20.5**/32.71/0.9572	22.1/32.81/0.9584	180.6/ **33.15**/**0.9618**	24.3/32.90/0.9570
peppers	20.3/33.92/0.9648	**18.3**/34.01/0.9659	94.4/ **34.30**/ **0.9705**	20.7/34.06/0.9621
pirate	**18.6**/30.34/0.9405	22.4/30.41/0.9423	267.7/30.30/**0.9471**	24.5/**30.51**/0.9470
woman	**18.2**/29.51/0.9409	21.9/29.59/0.9431	423.6/**29.68**/**0.9462**	24.5/29.63/0.9442

Table 5: The comparison of analysis based approach, balanced approach, EPIHT-LS for (9) and 3MG in terms of runtime (seconds), PSNR and SSIM values.

of these phenomenons are still not so well understood, and some issues of the proposed algorithms still needs further exploration. This is left as a future research direction.

Acknowledgements

The work of Xue Zhang, Likun Hou and Xiaoqun Zhang were partially supported by NSFC (11101277, 91330102) and 973 program (2015CB856000).

References

[1] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455), 2001.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[3] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.

[4] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988.

[5] 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.

[6] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

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

[8] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572, 2007.

[9] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of Łojasiewicz inequalities and applications. *Centre De Recerca Matemtica*, 2008.

[10] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, pages 1–36, 2013.

[11] R. I. Bot, E. R. Csetnek, and S. László. An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. *Euro Journal on Computational Optimization*, 4(1):3–25, 2016.

[12] J. Cai, R. Chan, L. Shen, and Z. Shen. Restoration of chopped and nodded images by framelets. *SIAM Journal on Scientific Computing*, 30(3):1205–1227, 2008.

[13] J. Cai, R. Chan, L. Shen, and Z. Shen. Tight frame based method for high-resolution image reconstruction. *Contemporary Applied Mathematics*, pages 1–36, 2010.

[14] J. Cai, R. H. Chan, L. Shen, and Z. Shen. Simultaneously inpainting in image and transformed domains. *Numerische Mathematik*, 112(4):509–533, 2009.

[15] J. Cai, R. H. Chan, and Z. Shen. A framelet-based image inpainting algorithm. *Applied & Computational Harmonic Analysis*, 24(2):131–149, 2008.

[16] J. Cai, B. Dong, S. Osher, and Z. Shen. Image restorations: total variation, wavelet frames and beyond. *Journal of American Mathematical Society*, 25(4):1033–1089, 2012.

[17] J. Cai, S. Osher, and Z. Shen. Split bregman methods and frame based image restoration. *Multiscale modeling and simulation*, 8(2):337–369, 2009.

[18] J. Cai and Z. Shen. Framelet based deconvolution. *Journal of Computational Mathematics*, 28(3):289–308, 2010.

[19] J. F. Cai, B. Dong, and Z. Shen. Image restoration: A wavelet frame based model for piecewise smooth functions and beyond. *Applied & Computational Harmonic Analysis*, 41(1):94–138, 2016.

[20] X. Cai, J. H. Fitschen, M. Nikolova, G. Steidl, and M. Storath. Disparity and optical flow partitioning using extended potts priors. *arXiv preprint arXiv:1405.1594*, 2014.

[21] E. Candes, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied & Computational Harmonic Analysis*, 31:59–73, 2011.

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

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

[24] A. Chambolle, R. A. DeVore, N. yong Lee, and B. J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions Image Processing*, 7:319–335, 1996.

[25] R. Chan, T. Chan, L. Shen, and Z. Shen. Wavelet algorithms for high-resolution image reconstruction. *SIAM Journal on Scientific Computing*, 24:1408–1432, 2003.

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

[27] 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.

[28] G. H. Chen and R. T. Rockafellar. Convergence rates in forward-backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

[29] E. Chouzenoux, A. Jezierska, J. C. Pesquet, and H. Talbot. A majorize-minimize subspace approach for l2-l0 image regularization. *SIAM Journal on Imaging Sciences*, 6(1):563–591, 2011.

[30] E. Chouzenoux, J. C. Pesquet, and A. Repetti. Variable metric forwardcbackward algorithm for minimizing the sum of a differentiable function and a convex function. *Journal of Optimization Theory & Applications*, 162(1):107–132, 2014.

[31] C. Chui, W. He, and J. Stöckler. Compactly supported tight and sibling frames with maximum vanishing moments. *Applied and Computational Harmonic Analysis*, 13(3):224–262, 2002.

[32] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.

[33] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

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

[35] D. Dl., G. Kerkyacharian, D. Picard, and J. Im. Density estimation by wavelet thresholding. *Annals of Statistics*, 24(2):508–539, 1970.

[36] B. Dong, H. Ji, J. Li, Z. Shen, and Y. Xu. Wavelet frame based blind image inpainting. *Applied & Computational Harmonic Analysis*, 32(2):268–279, 2011.

[37] B. Dong, Q. Jiang, and Z. Shen. Image restoration: Wavelet frame shrinkage, nonlinear evolution pdes, and beyond. *UCLA CAM Report*, 13-78, 2013.

[38] B. Dong, J. Li, and Z. Shen. X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting. *Journal of Scientific Computing*, 54 (2-3):333–349, 2013.

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

[40] B. Dong and Y. Zhang. An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration. *Journal of Scientific Computing*, 54(2-3):350–368, 2013.

[41] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[42] D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[43] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.

[44] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

[45] M. A. T. Figueiredo and R. D. Nowak. An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 12(12):906–916, 2003.

[46] H. Gao, J. Cai, Z. Shen, and H. Zhao. Robust principal component analysis-based four-dimensional computed tomography. *Physics in Medicine and Biology*, 56:3181, 2011.

[47] T. Goldstein and S. Osher. The split bregman method for ℓ_1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[48] E. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

[49] B. Han. On Dual Wavelet Tight Frames. *Applied & Computational Harmonic Analysis*, 4(4):380–413, 1997.

[50] Y. Hu, C. Li, and X. Yang. Proximal gradient algorithm for group sparse optimization. *Methods*, 67(3):294–303, 2015.

[51] X. Jia, B. Dong, Y. Lou, and S. Jiang. GPU-based iterative cone-beam CT reconstruction using tight frame regularization. *Physics in Medicine and Biology*, 56:3787–3807, 2011.

[52] Y. Jiao, B. Jin, and X. Lu. A primal dual active set with continuation algorithm for the ℓ_0-regularized optimization problem. *Applied & Computational Harmonic Analysis*, 39(3), 2014.

[53] K. K. and P. A. ω_f-stratification of subanalytic functions and the łojasiewicz inequality. *Comptes Rendus De L'académie Des Sciences:série 1, Mathématique*, 318(2):129–133, 1994.

[54] M. Li, B. Hao, and X. Feng. Iterative regularization and nonlinear inverse scale space based on translation invariant wavelet shrinkage. *International Journal of Wavelets Multiresolution & Information Processing*, 6(1):83–95, 2008.

[55] S. Liapis and G. Tziritas. Color and texture image retrieval using chromaticity histograms and wavelet frames. *Multimedia, IEEE Transactions on*, 6(5):676–686, 2004.

[56] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[57] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les Équations Aux Dérivées Partielles: paris, 1962*, pages 87–89.

[58] Z. Lu. Iterative hard thresholding methods for ℓ_0 regularized convex cone programming. *Mathematical Programming*, pages 1–30, 2012.

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

[60] Z. Lu and Y. Zhang. Iterative reweighted minimization methods for ℓ_p regularized unconstrained nonlinear programming. *Mathematical Programming*, Jan 2014.

[61] G. Marjanovic and V. Solo. On optimization and matrix completion. *Signal Processing, IEEE Transactions on*, 60(11):5714–5724, 2012.

[62] B. Mordukhovich. *Variational Analysis and Generalized Differentiation. I. Basic Theory.* Grundlehren der Mathematischen Wissenschaften, vol. 330. Springer, Berlin, 2006.

[63] J. J. Moreau. Proximité et dualité dans un espace hilbertien. In *Bull. Soc. Math. France*, pages 273–299, 1965.

[64] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[65] P. Ochs, Y. Chen, T. Brox, and T. Pock. ipiano: Inertial proximal algorithm for non-convex optimization. *Siam Journal on Imaging Sciences*, 7(2):1388–1419, 2014.

[66] B. O'Donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, pages 1–18, 2013.

[67] R. Rockafellar and W. J. B. *Variational Analysis*. Grundlehren der matematischen Wissenschaften, vol. 317, 1st edn. Springer, Berlin, 1998.

[68] A. Ron and Z. Shen. Affine systems in $L_2(\mathbb{R}^d)$ II: dual systems. *Journal of Fourier Analysis and Applications*, 3(5):617–638, 1997.

[69] 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.

[70] Z. Shen. Wavelet frames and image restorations. In *Proceedings of the International Congress of Mathematicians*, volume 4, pages 2834–2863, 2010.

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

[72] J. Trzasko, A. Manduca, and E. Borisch. Sparse MRI reconstruction via multiscale ℓ_0-continuation. In *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*, pages 176–180. IEEE, 2007.

[73] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *SIAM Journal on Optimization*, 2008.

[74] Y. Xu and W. Yin. A block coordinate descent method for multi-convex optimization with applications to nonnegative tensor factorization and completion. Technical report, DTIC Document, 2012.

[75] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.

[76] X. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011.

[77] X. Zhang, Y. Lu, and T. Chan. A novel sparsity reconstruction method from poisson data for 3d bioluminescence tomography. *Journal of Scientific Computing*, pages 1–17, 2012.

[78] Y. Zhang, B. Dong, and Z. Lu. ℓ_0 minimization for wavelet frame based image restoration. *Mathematics of Computation*, 82(282):995–1015, 2013.

[79] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.