Extrapolated Proximal Iterative Hard Thresholding Methods For Wavelet Frame Based Image Restoration

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]*Department of Mathematics, Shanghai Jiao Tong Univeristy, Shanghai 200240, China*

Abstract

The iterative thresholding algorithms were primarily proposed in [24] (both soft and hard) and [23, 29, 40] (soft) for solving wavelet based linear inverse problems, including image restoration problems with sparsity constraint. The analysis of iterative soft thresholding algorithms has been well studied (e.g. [28, 43]) and inspired much of the work for different applications and related minimization problems. However, iterative hard thresholding methods are less understood due to its non-convexity and non-smoothness except some studies related to sparse signal recovery [61, 51] and image restoration [66, 35, 67]. 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 Kurdyka-Łojasiewicz property theoretical framework, 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 models, 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 A is an ill-posed linear operator and ϵ is a white Gaussian noise or observation error. 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 performance for general image restoration problems.

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)

1.1. Image restoration by wavelet frame sparse approximation

The theory of frames, including wavelet frames [27, 30, 44, 58, 57, 59] and many others (e.g. curvelets [18, 19]), has been well developed for the past two decades. In recent years, wavelet frames have been widely used in image restoration and medical imaging [11, 12, 14, 15, 16, 20, 22, 24, 25, 26, 31, 33, 41, 46, 48, 49]. The basic idea of all the wavelet frame based approaches is that images can be sparsely approximated by properly designed wavelet frames. More recently, wavelet frame based approaches have been linked with variational and PDE based approaches in [9, 10, 32] and references therein, where new models and algorithms for image restoration problems have been introduced as well. In this paper, the tight wavelet frame system we used is the B-spline tight wavelet frame from [58], which was constructed by the Unitary Extension Principle [58].

We start with the balanced approach, originally used in [24, 26] 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,

$$W := [W_0^\top, W_{1,1}^\top, \ldots, W_{1,J}^\top, \ldots, W_{Q,1}^\top, W_{Q,J}^\top]^\top$$

is a multi-level wavelet tight frame transform operator (i.e $W^\top W = I$, others maybe have different definition) that converts an image to its wavelet coefficients, in which Q indicates the level of wavelet decomposition and J is the number of high-pass filters that the wavelet system used (see, e.g. [34] for more details), $W^\top$ is the transpose of W, x can be seen as a coefficient vector in the transformed domain of W, $D \succ 0$ is some weighting 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* and *synthesis* based approaches [38], although these models were developed independently in the literature. The analysis based approach aims to find an image u^* whose representation coefficients are sparsest by solving the minimization problem

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

The synthesis approach seeks for the sparsest representation coefficients x^* such that

$$x^* = \arg\min_x \frac{1}{2}\|AW^\top x - b\|_D^2 + \|\boldsymbol{\lambda}\cdot x\|_1 \tag{3}$$

and the target image is reconstructed by $u^* = W^\top x^*$. It is well known that in the case of redundant tight frame systems, these two models behave differently. The balanced model (1) unifies both popular sparse approximation models and provides balanced image quality between sparseness and regularity. When $\kappa = 0$, this model becomes the synthesis model (3) and when $\kappa \to \infty$, it becomes the analysis model (2).

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 [24][26] was proved in [15] to be a proximal forward-backward splitting algorithm [5, 28, 29, 50, 55, 62].

Furthermore, accelerated proximal gradient (APG) method was proposed in [60] for the balanced model and achieved a theoretically optimal convergence rate $O(1/k^2)$ (with strong convexity modification). More specifically, the proximal forward-backward splitting method for solving (1) is written as

$$x_{k+1} = \mathrm{Prox}_{F_1/L}(x_k - \frac{1}{L}\nabla F_2(x_k)), \tag{4}$$

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

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

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

$$(\mathcal{S}_{\boldsymbol{\gamma}}(x))_i = \begin{cases} x_i - \mathrm{sign}(x_i)\gamma_i, & \text{if } |x_i| > \gamma_i \\ 0, & \text{if } |x_i| \le \gamma_i. \end{cases} \tag{5}$$

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, 28, 29, 43, 64, 65].

1.2. Motivations and contributions

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 compressive sensing theory (see e.g. [21]). However, ℓ_0 norm based regularization still has its advantages in many applications. In [24], both iterative soft and hard thresholding algorithms are adopted and it is shown that the hard one achieved better image quality compared to the soft one. In [35, 67], wavelet frame based ℓ_0 regularization also shows better image edge preserving property compared to conventional ℓ_1 regularization. In [61], wavelet based ℓ_0 regularization is considered for MRI reconstruction. In [66], it is demonstrated that ℓ_1 regularization may fail to recover sparse solutions for some very ill-posed inverse problems and non-Gaussian noise corruption. Historically, soft and hard thresholding have been studied by Donoho and Johnstone [37] in the context of nonlinear estimator in orthogonal bases. The continuity of the soft thresholding operator may lead to less artifacts linked to Gibbs phenomena. However, the soft thresholding operator yields loss of contrast and eroded signal peaks as ℓ_1 estimator leads to bias estimation for large coefficients [39].

Thus, we are interested in replacing soft thresholding (5) by the *hard thresholding* in the numeric scheme. Define the componentwise hard thresholdding operator

$$(\mathcal{H}_{\boldsymbol{\gamma}}(x))_i = \begin{cases} \{x_i\}, & \text{if } |x_i| > \gamma_i, \\ \{0, x_i\} & \text{if } |x_i| = \gamma_i, \\ \{0\}, & \text{if } |x_i| < \gamma_i, \end{cases} \tag{6}$$

it is the solution of a proximal mapping of the ℓ_0- regularization. Since ℓ_0 function is discontinuous and non-convex, it is hard to analyze the behaviour for more complex ℓ_0 minimization problem such as

$$\min_{x\in\mathbb{R}^n} h(x) := \|\boldsymbol{\lambda}\cdot x\|_0 + f(x), \tag{7}$$

where $\|\boldsymbol{\lambda}\cdot x\|_0 = \sum_i \lambda_i |x_i|_0$ denotes the weighted number of nonzero elements in the vector x.

Analogue to the proximal forward-backward splitting method for ℓ_1 case (4), this class of problem can be solved by *iterative hard thresholding*. Note that this algorithm has already been considered in [24] in the context of wavelet frame based image super-resolution. In [6, 7], this algorithm is considered in the context of compressive sensing reconstruction [36] for $f(x) = \frac{1}{2}\|Ax - b\|^2$. Recently, ℓ_0-norm minimization problem has begun to arouse researchers' interest again. In particular, proximal forward-backward splitting algorithm has been studied in [3, 8] for solving a general nonconvex and nonsmooth problem of the form

$$\min_x g(x) + f(x), \tag{8}$$

where g is lower semi-continuous and f is smooth. Under the Kurdyka-Łojasiewicz property theoretical framework [2], it is shown in [3, 8] that the objective function value is decreasing and furthermore the sequence generated by a proximal algorithm globally converges to a critical point when the sequence is bounded. In particular, for (7), the corresponding scheme is as follows

$$x_{k+1} \in \arg\min_{x\in\mathbb{R}^n} \|\boldsymbol{\lambda}\cdot x\|_0 + \frac{L}{2}\|x - x_k + \nabla f(x_k)/L\|^2 + \frac{\mu}{2}\|x - x_k\|^2, \tag{9}$$

where L is the Lipschitz constant of ∇f, which we refer to as *proximal iterative hard thresholding* (PIHT) algorithm. For PIHT algorithm, Lu [51] has proved that the index set of zero elements of x_k changes only finitely often and the sequence x_k converges to a local minimizer of $h(x) = \lambda\|x\|_0 + f(x)$. Furthermore, if $f(x)$ is strongly convex, an estimation on the number changes of $I(x_k)$ and the iterative complexity for finding an ε-local-optimal solution has also been established in [51]. In [67] Zhang, Dong and Lu propose an adapted penalty decomposition (PD) method [52] 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 in terms of quality of recovered images. Dong and Zhang also propose in [35] the mean doubly augmented Lagrangian (MDAL) method to solve ℓ_0 analysis based model. The numerical experiments also show that using ℓ_0-norm can generate higher quality images than ℓ_1-norm based methods. However, the convergence analysis is not established for MDAL method. Besides, some other types of non-convex problems and algorithms are also of great interests to researchers, see e.g.[1, 17, 45, 47, 53, 54].

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. In particular, we propose *extrapolated proximal iterative hard thresholding* (EPIHT) algorithm and *EPIHT algorithm with line search* (EPIHT-LS) for solving the ℓ_0 regularization problem. Under the Kurdyka-Łojasiewicz property theoretical framework, 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.

The rest paper 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 the uniqueness and the convergence of the proposed iterative scheme, similar as in [51, 60], we consider the following model

$$\min_{x\in\mathbb{R}^n} H(x) := \lambda\|x\|_0 + F(x) + \frac{t}{2}\|x\|_2^2 = \lambda\|x\|_0 + G(x), \tag{10}$$

where $t \geq 0$, $F(x)$ is smooth convex function. Here we use the uniform parameter $\lambda\|x\|_0$ instead of the weighted $\|\boldsymbol{\lambda}\cdot x\|_0$ for the simplicity of notation, while all proofs in the following context can be easily extended to the weighted case. We emphasize that, the following relaxed ℓ_0 based wavelet frame balanced approach

$$\min_x \|\boldsymbol{\lambda}\cdot x\|_0 + \frac{1}{2}\|AW^\top x - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)x\|^2 + \frac{t}{2}\|x\|^2, \tag{11}$$

which will be used in section 3, is a special case of (10) when $F(x) = \frac{1}{2}\|AW^\top x - f\|_D^2 + \kappa\|(I - WW^\top)x\|^2/2$, $G(x) = F(x) + t\|x\|^2/2$.

Throughout this paper, our common **assumption** on problem (10) is: F is convex differentiable, bounded from below and ∇F is L-Lipschitz continuous.

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

$$S_\tau(x, y) = G(y) + \langle \nabla G(y), x - y\rangle + \frac{\tau}{2}\|x - y\|^2 + \|\boldsymbol{\lambda}\cdot x\|_0, \tag{12}$$

where $\tau > 0$.

We may directly use PIHT to solve the problem (10), namely

PIHT Algorithm [51]

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

$$x_{k+1} \in \arg\min_{x\in\mathbb{R}^n} S_{L+t+\sigma_k}(x, x_k) \tag{13}$$

$k = k + 1$
end(while)

The step (13) 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 (6). Note that for a general nonconvex and nonsmooth problem studied in [8]

$$\min_{x,y} s(x) + r(y) + g(x, y), \tag{14}$$

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 $g(x, y) =$

$G(x)$, (14) reduces to (10), and PALM algorithm becomes PHIT algorithm. In [51], this algorithm was also studied for ℓ_0 regularized convex cone programming and the convergence to a local minimizer was established.

The first algorithm that we propose is analogous to the extrapolation used in APG method [5, 60] with adaptive restart [56] 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 $H(y_{k+1}) > H(x_k)$

$$y_{k+1} = x_k \tag{15}$$

end(if)

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

$$x_{k+1} \in \arg\min_{x \in \mathbb{R}^n} S_{L+t+\sigma_k}(x, y_{k+1}) \tag{16}$$

$k = k + 1$

end(while)

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:

$$H(x_{k+1}) \leq S_{\tau_k}(x_{k+1}, y_{k+1}) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2, \tag{17}$$

where $\sigma_k \in (\sigma_1, \sigma_2)$ and σ_1, σ_2 are two positive constants. We present the following lemma that shows (17) 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 (17) *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

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

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

Then we propose the following extrapolated proximal iterative hard thresholding algorithm with line search (EPIHT-LS).

EPIHT-LS Algorithm

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

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

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

$$y_{k+1} = x_k \tag{21}$$

end(if)
Choose $\sigma_k \in (a, b)$ and $\tau_0^k \in [\tau_{min}, \tau_{max}]$
for $j = 0, 1, 2, \ldots,$ compute $\hat{\tau}_j = \tau_0^k/\eta^j$ and

$$\hat{x} = \arg\min_x S_{\hat{\tau}_j}(x, y_{k+1}) \tag{22}$$

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

$$\tau_k = \hat{\tau}_j \text{ and } x_{k+1} = \hat{x}$$

end(if)
end(for)
$k = k + 1$
end(while)

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

$$\tau_0^k = \min\{\tau_{max}, \max\{\tau_{min}, \frac{\langle \nabla f(x_k) - \nabla(x_{k-1}), x_k - x_{k-1}\rangle}{\|x_k - x_{k-1}\|^2}\}\}. \tag{23}$$

Remark 2. *It can be seen from the flow of 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

$$I(x) := \{i : x_i = 0\}. \tag{24}$$

Definition 3.1 (Subdifferentials [8])**.** *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*

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

Definition 3.2 (Kurdyka-Łojasiewicz property [63])**.** *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*

$$\frac{|\psi(x) - \psi(\bar{x})|^\theta}{\mathrm{dist}(0, \partial \psi(x))} \tag{25}$$

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

Lemma 3.3 (Lemma 4 in [2])**.** *Assume $f = F + G$, where F is a C^1 function with Lipschitz gradient and G is proper and lower semi-continuous function. For all $x \in \mathrm{dom}(f)$, we have $\partial f(x) = \nabla F(x) + \partial G(x)$.*

In the next, we show that the every stationary point of the ℓ_0 related minimization which is defined in (10) is a local minimizer.

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

Proof. Define $I(x) = \{i : x_i = 0\}$, by the Lemma 3.3, we know

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

Consequently, $\partial_{i \notin I(x)} f(x) = (\nabla F(x))_{i \notin I(x)} = 0$. For any Δx, by the convexity of F, we have

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

where the last inequality holds provided $\|\Delta x\|_\infty < \min\{\lambda / |(\nabla F(x))_i | i \in I(x)\}$. □

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* (10)*, 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* (10) *and $\{x_k\}$ be the sequence generated by EPIHT/EPIHT-LS algorithm. Then, it has the following descent property:*

$$H(x_k) - H(x_{k+1}) \geq \rho_1 \|x_{k+1} - y_{k+1}\|^2, \tag{26}$$

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

$$\|x_k - y_k\| \to 0, \tag{27}$$

Proof. For EPIHT algorithm, from (12), (16), (15) and Lemma 2.1, we have

$$S_{L+t+\sigma_k}(x_{k+1}, y_{k+1}) \leq S_{L+t+\sigma_k}(y_{k+1}, y_{k+1}) = H(y_{k+1}) \leq H(x_k) \tag{28}$$

Combine (28) 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 (12), (18) and (21) , we have

$$S_{\tau_k}(x_{k+1}, y_{k+1}) \leq S_{\tau_k}(y_{k+1}, y_{k+1}) = H(y_{k+1}) \leq H(x_k) \tag{29}$$

Combine (29) and (17), 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

$$H(x_{k+1}) \leq H(x_k) - \frac{\sigma_k}{2}\|x_{k+1} - y_{k+1}\|^2. \tag{30}$$

Set $\rho_1 = \frac{a}{2} > 0$, the inequality (26) holds.
Sum up k from 0 to j in (26), 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 (27). □

In the next, we bounded 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* (10) *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 $s_k \in \partial H(x_k)$

$$\|s_k\| \leq \rho_2 \|x_k - y_k\|, \tag{31}$$

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 (16) 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 (22) 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 (31) 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* (10) *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 (27), we know $y_{k_j} \to \bar{x}$ as $j \to +\infty$. From (16) and (18), we have

$$S_{\nu_k}(x_{k+1}, y_{k+1}) \leq S_{\nu_k}(\bar{x}, y_{k+1}), \tag{32}$$

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 (32), 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 [8]. □

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, [8]). *Let Ω be a compact set and let σ be a proper and lower semi-continuous function. Assume that σ is constant on Ω and satisfies the Kurdyka-Łojasiewicz property at each point of Ω. 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 Ω 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 $\mathrm{dist}(x_k, \omega(x^0)) = 0$, 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 $\Omega = \omega(x^0)$, for any $k > \ell$, we have

$$\psi'(H(x_k) - H(\bar{x}))\mathrm{dist}(0, \partial H(x_k)) \ge 1. \tag{33}$$

From Lemma 3.7, we have

$$\psi'(H(x_k) - H(\bar{x})) \ge \frac{1}{\rho_2}\|x_k - y_k\|, \tag{34}$$

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

$$\psi(H(x_k) - H(\bar{x})) - \psi(H(x_{k+1} - H(\bar{x})) \ge \psi'(H(x_k) - H(\bar{x}))(H(x_k) - H(x_{k+1})). \tag{35}$$

Define

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

From lemma 3.6, (34) and (35), 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

$$2\|x_{k+1} - y_{k+1}\| \le \|x_k - y_k\| + c_0\ \triangle_{k,k+1} \tag{36}$$

by Cauchy-Schwarz inequality. Summing (36) 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 $k > \ell$,

$$\sum_{i=\ell+1}^{k} \|x_{i+1} - y_{i+1}\| \leq \|x_\ell - y_\ell\| + C\psi(H(x_{\ell+1}) - H(\bar{x})). \tag{37}$$

Therefore,

$$\begin{aligned} \sum_{i=\ell+1}^{k} \|x_{i+1} - y_{i+1}\| &\geq \sum_{i=\ell+1}^{k} (\|x_{i+1} - x_i\| - \omega_i \|x_i - x_{i-1}\|) \\ &\geq \sum_{i=\ell+1}^{k} (1-\omega)\|x_{i+1} - x_i\| - \omega\|x_\ell - x_{\ell-1}\| \end{aligned} \tag{38}$$

Putting (37) and (38) 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* (10) *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* (10)*, $\{x_k\}$ be the sequence generated by the EPIHT/EPIHT-LS algorithm and $H(x, I) = G(x) + \mathcal{X}_{C_I}(x)$. 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 ([63])**.** *Given an index set I. Let $H(x, I) = G(x) + \mathcal{X}_{C_I}(x)$ where $G(x)$ is defined in* (10)*. 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 [63] as $G(x)$ is t-strongly convex. □

Theorem 3.13 (Convergence rate)**.** *Let $H(x)$ be the objective function defined in* (10) *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 $k > k_0$, (37) and (38), we have

$$\begin{aligned} \Delta_k &\leq C\psi(H(x_k, I(x_{k_0}))) + \frac{1}{1-w}(\Delta_{k-1} - \Delta_k) + w(\Delta_{k-2} - \Delta_{k-1}) \\ &\leq C\psi(H(x_k, I(x_{k_0}))) + \frac{1}{1-w}(\Delta_{k-2} - \Delta_k) \end{aligned} \tag{39}$$

since $0 < w < 1$ and $\Delta_{k-2} \geq \Delta_{k-1}$. From (34), lemma 3.12 and $\psi(s) = \frac{2}{t}\sqrt{s}$, we have

$$\frac{1}{t}(H(x_k, I(x_{k_0})))^{-1/2} \geq (\rho_2\|x_k - y_k\|)^{-1} \geq (\rho_2(\|x_k - x_{k-1}\| + w\|x_{k-1} - x_{k-2}\|))^{-1}. \tag{40}$$

Then, it implies

$$\psi(H(x_k, I(x_{k_0}))) = \frac{2}{t}(H(x_k, I(x_{k_0})))^{1/2} \leq \frac{2}{t}\frac{1}{t}(\rho_2(\|x_k - x_{k-1}\| + w\|x_{k-1} - x_{k-2}\|)). \tag{41}$$

Let $C_1 = C\frac{2}{t^2}\rho_2$ and $C_2 = \frac{1}{1-w}$, (39) and (41) imply

$$\Delta_k \leq C_1(\Delta_{k-2} - \Delta_k) + C_2(\Delta_{k-1} - \Delta_k) = (C_1 + C_2)(\Delta_{k-2} - \Delta_k). \tag{42}$$

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 (10), and compare them with the results of PIHT algorithm and some previous methods designed for ℓ_1-norm minimization in the literature (like FISTA, APG, etc). 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

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2 \\ &\text{s.t. } \|x\|_0 \leq r \end{aligned}$$

where $A \in R^{m\times n}$ is a data matrix, $b \in R^m$ is an observation vector, and $r \in \{1, 2, \ldots, n\}$ is some integer for controlling the sparsity of the solution. The above problem is hard to solve and one popular approach in literature is to solve regularization problem with the following form

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

where $g(x)$ is the regularization term. We consider the regularization problem

$$\min_{x\in\mathbb{R}^n} H(x) := \frac{1}{2}\|Ax-b\|^2 + \lambda_1\|x\|_0 + \frac{\lambda_2}{2}\|x\|^2 \tag{43}$$

with regularization parameters $\lambda_1, \lambda_2 > 0$ to be analogue with the Naïve elastic net model [68]

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|^2 + \lambda_1\|x\|_1 + \frac{\lambda_2}{2}\|x\|^2 \tag{44}$$

where $\lambda_1, \lambda_2 > 0$ are regularization parameters. Since $f(x) = \frac{1}{2}\|Ax-b\|^2$, and it is easy to see that $\nabla f(x) = A^\top(Ax-b)$ is Lipschitz continuous with $L \geq \lambda_{\max}(A^\top A)$, where $\lambda_{\max}(A^\top A)$ denotes the maximum eigenvalue of $A^\top A$. Our aim below is to apply the EPIHT and EPIHT-LS algorithms to solve (43), then compare the results with PIHT for solving (43) and FISTA [5] algorithm for solving (44).

In numerical experiment, the data matrix $A \in \mathbb{R}^{m\times n}$ is a Gaussian random matrix 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}. And for each pair 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 $\bar{x}$.

For parameters, we choose $\lambda_1 = 0.3, \lambda_2 = 10^{-7}, L = \lambda_{max}(A^\top A + \lambda_2 I)$ 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). And for EPIHT algorithm, we choose $\omega_k \equiv 0.8$,$\sigma_k \equiv 0.001$. For EPIHT-LS algorithm, Let $\sigma_k \equiv 0.001$, we choose τ_0^k like (23) where $\tau_{min} = 0.01 * L, \tau_{max} = 10 * L$ 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 $\mu = 0.001$ for PIHT 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 FISTA, PIHT,EPIHT, EPIHT-LS algorithms. For each algorithm and each choice of (n, s) of the solution x, we record the average runtime, the average relative error $\frac{\|x-\bar{x}\|}{\|\bar{x}\|}$ to the original signal $\bar{x}$ and the average number of iteration the algorithm needed. The numerical results are listed in Table 1, 2, 3.

One may observe from Table 2 that, compared to FISTA, the two ℓ_0-based algorithms, namely PIHT and EPIHT, can always reach solutions of higher precision. The precision of solutions between PIHT and EPIHT is similar. However, frome table 1,3, EPIHT enjoys better iteration complexity and less CPU time as well, which shows the advantage of extrapolation. And EPIHT-LS algorithms enjoys better iteration complexity than EPIHT, which shows the advantage of line search. However, the solutions reached by EPIHT-LS algorithms tends to have larger error, when the sensing matrices have larger number of columns and the test signals are less sparse. This could be due to a faster convergence to a "bad" local minimizer.

In order to show change of τ_k of EPIHT-LS algorithm, we test problem (43). And the parameters used for EPIHT-LS algorithm are the same with above. The list the result in figure 1.

Table 1: results of average number of iteration

s	n	Average number of iteration			
		FISTA	PIHT	EPIHT	EPIHT-LS
$\lfloor \frac{n}{100} \rfloor$	2500	109.8	137.4	38.1	10.8
	3000	133.2	162.0	41.2	11.6
	3500	148.4	184.7	48.1	12.1
	4000	172.4	214.4	57.6	12.9
	4500	192.1	247.3	64.6	13.4
$2\lfloor \frac{n}{100} \rfloor$	2500	145.1	175.2	47.7	15.5
	3000	182.8	211.0	57.2	19.7
	3500	239.0	254.3	68.6	24.6
	4000	355.7	323.0	86.0	27.0
	4500	602.8	385.6	95.0	28.2

Table 2: results of average relative error

s	n	Average relative error			
		FISTA	PIHT	EPIHT	EPIHT-LS
$\lfloor \frac{n}{100} \rfloor$	2500	0.3199	0.0001	0.0001	0.0001
	3000	0.3301	0.0001	0.0001	0.0001
	3500	0.3319	0.0001	0.0001	0.0001
	4000	0.3422	0.0001	0.0001	0.0001
	4500	0.3502	0.0001	0.0001	0.0001
$2\lfloor \frac{n}{100} \rfloor$	2500	0.3531	0.0001	0.0001	0.0077
	3000	0.3827	0.0001	0.0001	0.1367
	3500	0.4283	0.0540	0.0540	0.3992
	4000	0.5331	0.2890	0.2889	0.6418
	4500	0.6497	0.5580	0.5578	0.7479

Table 3: Results of average runtime

s	n	Average runtime			
		FISTA	PIHT	EPIHT	EPIHT-LS
$\lfloor \frac{n}{100} \rfloor$	2500	0.5308	0.6558	0.2631	0.2630
	3000	1.0100	1.2270	0.4236	0.4374
	3500	1.4336	1.7790	0.6162	0.5857
	4000	2.3156	2.8684	0.9798	0.8675
	4500	3.0025	3.8517	1.2692	1.1014
$2\lfloor \frac{n}{100} \rfloor$	2500	0.7076	0.8371	0.3276	0.4608
	3000	1.3859	1.5979	0.5872	0.8833
	3500	2.2971	2.4465	0.8782	1.3012
	4000	4.7908	4.3230	1.4623	1.8038
	4500	9.5030	6.0370	1.8777	2.1609

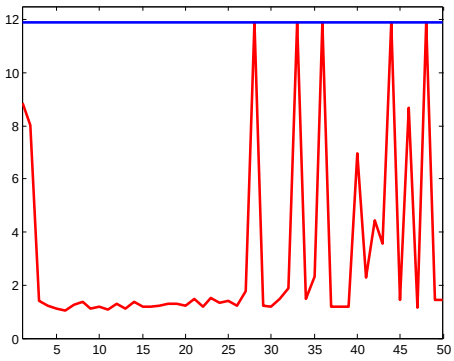

Figure 1: Horizontal axis represents the iteration number k. The red line represents τ_k of EPIHT-LS algorithm; the blue line represents global Lipschitz L.

4.2. Image Restoration

In this subsection, we conduct several numerical experiments on the image restoration problem. This part is mainly for demonstrating the improvement of the proposed ℓ_0-norm based regularization method (11) against traditional ℓ_1-norm based regularization methods, like total variation based regularization, framelet based regularization, etc.

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)

$$\boldsymbol{\lambda}_{k,j}[p] = \lambda \cdot w^{k-1},\ k = 1, \ldots, Q,\ j = 1, \ldots, J,$$

where λ is a positive constant, k 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.

Since $F(x) = \frac{1}{2}\|AW^\top x - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)x\|^2$, we have

$$\|\nabla F(x) - \nabla F(y)\| \leqslant L\|x - y\|, \forall x, y,$$

for any $L \geq \kappa + \lambda_{max}(A^*A)$. See [60] for more details. Thus for real implementation of EPIHT algorithm in all numerical experiments, 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 $\omega_{k+1} = \min(\omega_k/0.8, 0.9)$. Furthermore, we will specify the wavelet system used for generating the wavelet transform operator W and its decomposition level, as well as the choice of κ for each set of experiments conducted later. Thus, the parameters left for tuning up are the penalization weight λ and the decreasing weight w. In experiments we will manually tune up these two parameters to obtain the best restoration result.

Since the proposed objective function in (1) is discontinuous and non-convex, finding its global minimizer is computationally difficult. 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 of the ℓ_1-balanced model for both EPIHT algorithm and EPIHT-LS algorithm. Since the ℓ_1 balanced model is strongly convex, its global minimizer is unique so there will be no ambiguity. In the following subsections, when we talk about the computational time for EPIHT and EPIHT-LS algorithms, we also count in the time for computing their initial points. For effectiveness, those initial points are computed by APG.

4.2.1. CT Image Reconstruction

In this section, the proposed ℓ_0 minimization model in (11) is applied to the problem of CT image reconstruction. Note that in this case the degradation matrix A happens to be a projection matrix whose rows represent the collection different line integrals at different angles along different beams for CT scanning. For simplicity, the test is merely performed on simulated data. 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 and it utilization to projecting the phantom image to the Radon domain, we then add some Gaussian noise with variance σ to the projected data so as 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 method for wavelet tight frame based image restoration: analysis based approach (2) and balanced approach (1). Here we emphasize that both the analysis based approach and the balance approach are convex minimization models. We use the *accelerated proximal gradient (APG)* algorithm (see e.g. [60]) to solve the balanced approach, while split Bregman algorithm [16, 42] is adopted for solving the analysis based approach. For both iterative schemes that solve the balanced approach and (11), similar as in [67], 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 (11), 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 the 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 later one has slightly better PSNR values at the time when the same stopping criteria is satisfied. However, the longer processing time of EPIHT-LS is largely due to the additional function-value evaluation step as indicated in its implementation. To demonstrate this fact, we also plot the energy (i.e. value of the objective function with respect to the

	Split Bregman for (2)	APG for (1)	PIHT for (11)	EPIHT-LS for (11)
20 Projections				
PSNR	22.98	27.77	28.96	29.52
Time	39.67	70.45	135.62	172.61
	Split Bregman for (2)	APG for (1)	PIHT for (11)	EPIHT-LS for (11)
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*.

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 (11) compared to PIHT, and it also reaches a local minimizer of the objective function with slightly lower objective value. To better illustrate the effect of line-search, we also plot, for the case of 50 projections, the value of τ_k with respect to the iteration number k, see Figure 4.

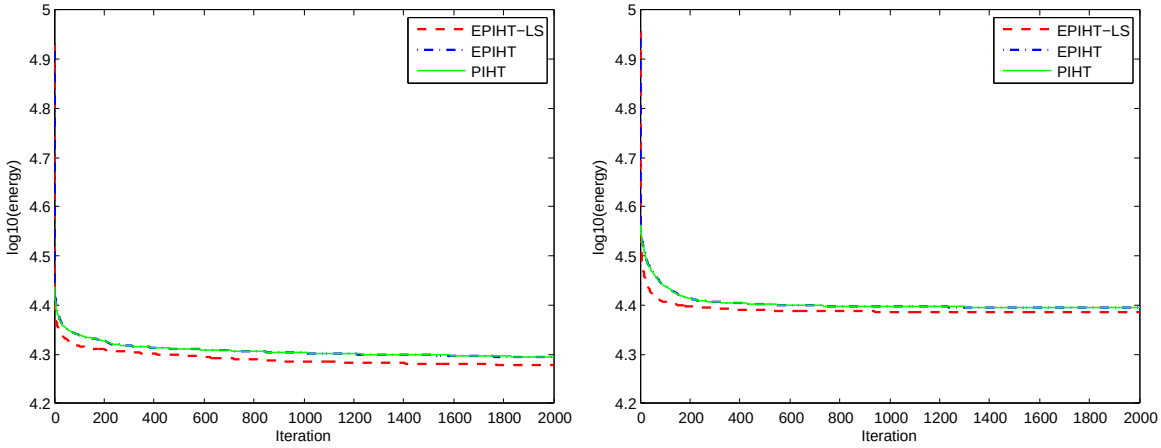

Figure 3: Energy evolving curves of CT image reconstruction, (a) for 20 projections, (b) for 50 projections.

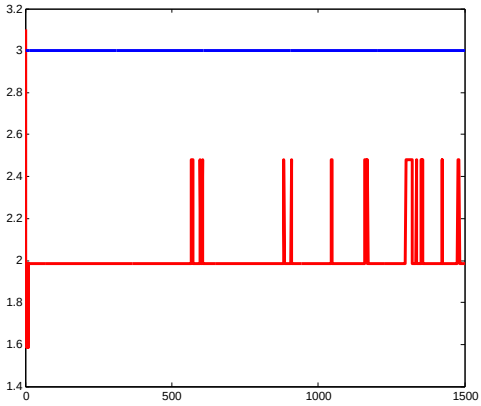

Figure 4: Horizontal axis represents the iteration number k. The red line shows the value of τ_k of EPIHT-LS algorithm; the blue line represents global Lipschitz L.

4.2.2. Image Deblurring

In this subsection, we apply the proposed model in (11) to the problem of image de-blurring, and compare the results with those of the analysis based approach in (2) and balanced approach (1). 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 (11), 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 PIHT 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

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

In our experiments, we test all three methods on twelve different images. These results are fully summarized in Table 4, 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). To better evaluate the image quality for each restoration method, we calculated the SSIM values of the processed results with respect to the original images and summarize the results in Table 5. 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 4 that, the EPIHT-LS method shows certain extent of improvement in PSNR values compared to the other two ℓ_1-norm based minimization methods. One may observe from Table 5 that the proposed ℓ_0 minimization method is consistently the best in terms of SSIM values(the closer to 1 the better), 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 as listed in Figure 5 and Figure 6 that the ℓ_0 balanced approach produces cleaner (seemingly with less artifacts) restoration results compared to the other two approaches.

5. Conclusions and perspectives

In this paper, we have studied the 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 based wavelet frame balanced approach (11). We provide a general convergence analysis for the general ℓ_0 regularization problem (10) 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 (10). 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 [51], and (ii)line search

Image	Split Bregman for (2)	APG for (1)	EPIHT-LS for (11)
boat (512)	29.67	29.70	**29.87**
bridge (512)	26.88	26.90	**26.92**
cameraman (256)	**26.97**	26.94	26.96
clock (256)	29.49	29.46	**29.82**
couple (512)	29.24	29.27	**29.58**
goldhill (512)	30.73	30.80	**31.02**
house (256)	32.47	32.42	**32.63**
jetplane (512)	32.17	32.23	**32.44**
Lena (512)	32.81	32.89	**33.15**
peppers (512)	34.13	34.09	**34.30**
pirate (512)	30.36	30.40	**30.44**
woman (512)	29.51	29.56	**29.58**

Table 4: The comparison of analysis based approach, balanced approach and EPIHT-LS for (11) in terms of PSNR values.

Image	Split Bregman for (2)	APG for (1)	EPIHT-LS for (11)
boat (512)	0.9442	0.9435	**0.9502**
bridge (512)	0.9255	0.9270	**0.9279**
cameraman (256)	0.8422	0.8380	**0.8455**
clock (256)	0.9144	0.9102	**0.9237**
couple (512)	0.9368	0.9367	**0.9453**
goldhill (512)	0.9359	0.9380	**0.9411**
house (256)	0.8590	0.8574	**0.8657**
jetplane (512)	0.9602	0.9591	**0.9648**
Lena (512)	0.9576	0.9584	**0.9633**
peppers (512)	0.9646	0.9659	**0.9705**
pirate (512)	0.9423	0.9423	**0.9463**
woman (512)	0.9419	0.9431	**0.9464**

Table 5: The comparison of analysis based approach, balanced approach and EPIHT-LS for (11) in terms of SSIM values.

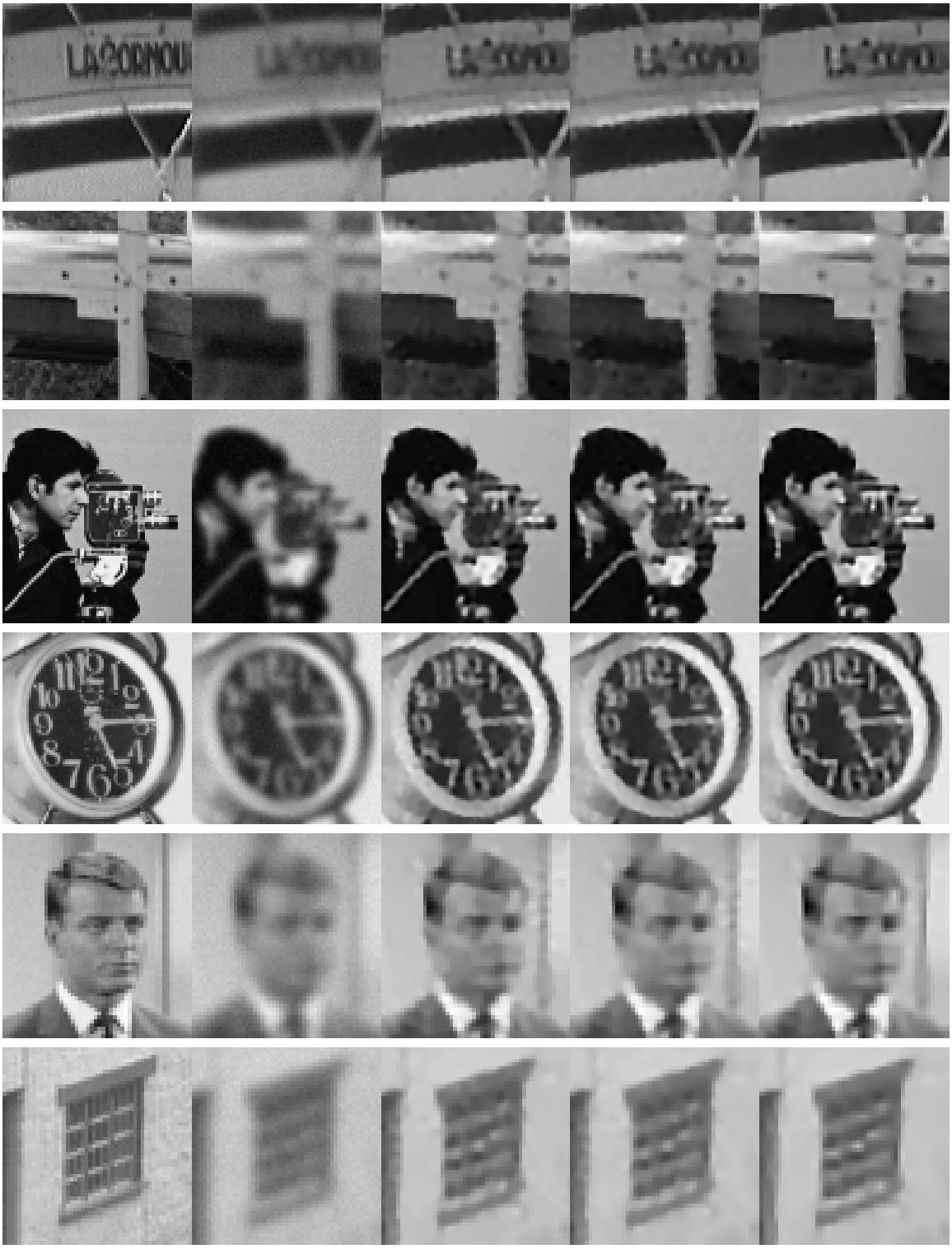

Figure 5: Zoom-in views of the processed results. From left to right: original images, noisy and blurred images, results of analysis based approach in (2), results of balanced approach (1), and results of EPIHT-LS for (11).

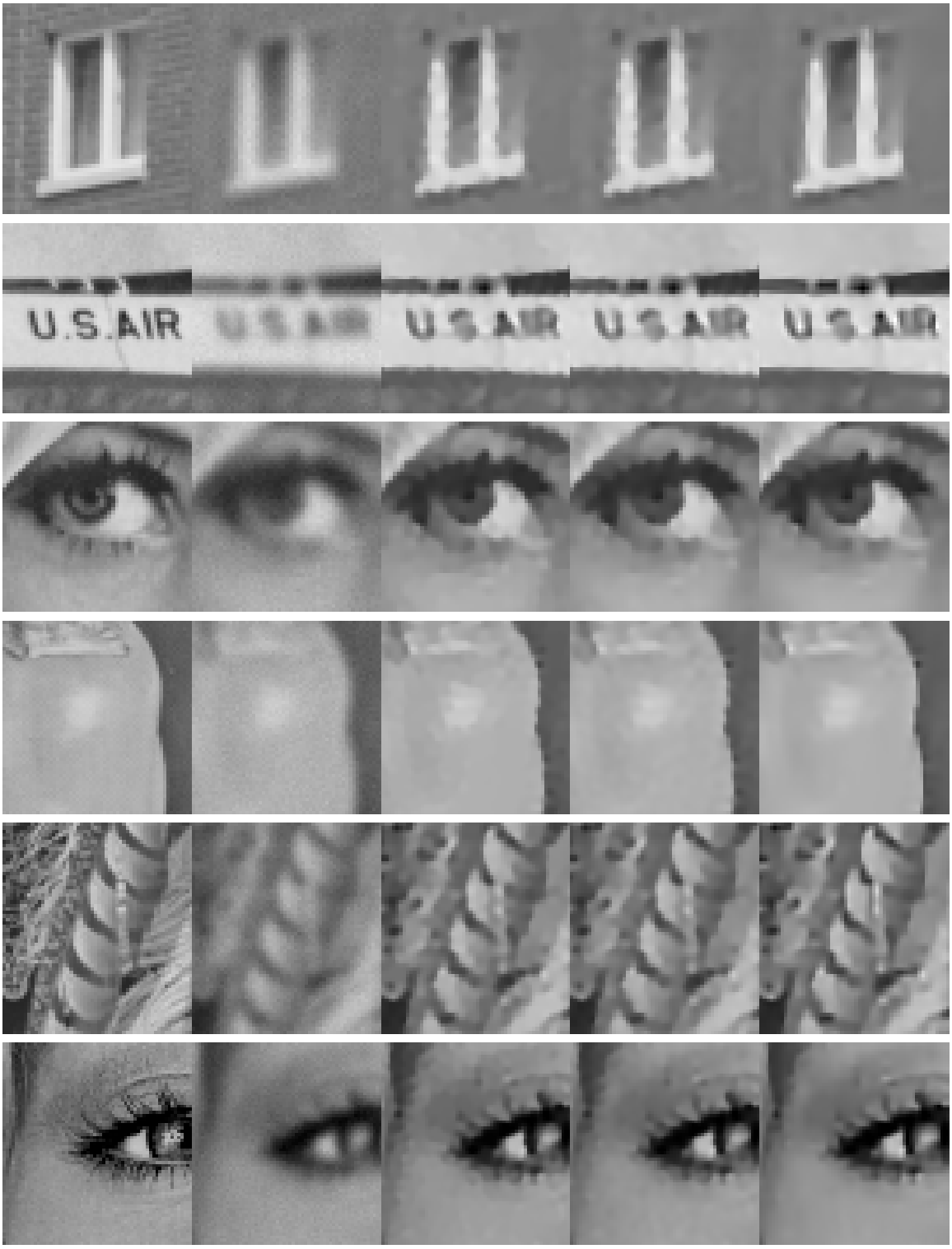

Figure 6: Zoom-in views of the processed results. From left to right: original images, noisy and blurred images, results of analysis based approach in (2), results of balanced approach (1), and results of EPIHT-LS for (11).

is helpful for identifying 'better' local solutions of the proposed ℓ_0 regularization model. However, the essence 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 and Computational Harmonic Analysis*, 27(3):265–274, 2009.

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

[9] 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.

[10] J. Cai, B. Dong, and Z. Shen. Image restorations: a wavelet frame based model for piecewise smooth functions and beyond. *Preprint*, 2014.

[11] J. Cai and Z. Shen. Framelet based deconvolution. *J. Comp. Math*, 28(3):289–308, 2010.

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

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

[17] 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.

[18] E. Candes and D. Donoho. New tight frames of curvelets and optimal representations of objects with C2 singularities. *Comm. Pure Appl. Math*, 56:219–266, 2004.

[19] E. Candes and D. Donoho. Continuous curvelet transform:: II. Discretization and frames. *Applied and Computational Harmonic Analysis*, 19(2):198–222, 2005.

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

[21] 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.

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

[23] 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 Trans. Image Processing*, 7:319–335, 1996.

[24] 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.

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

[26] 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.

[27] 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.

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

[29] 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.

[30] 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, Jan 2003.

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

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

[33] 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.

[34] 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.

[35] 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.

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

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

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

[39] 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.

[40] M. Figueiredo and R. Nowak. An em algorithm for wavelet-based image restoration. *Image Processing, IEEE Transactions on*, 12(8):906–916, Aug 2003.

[41] 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.

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

[43] 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.

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

[45] Y. Hu, C. Li, and X. Yang. Proximal gradient algorithm for group sparse optimization.

[46] 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.

[47] Y. Jiao, B. Jin, and X. Lu. A primal dual active set with continuation algorithm for the\ ell^ 0-regularized optimization problem. *arXiv preprint arXiv:1403.0515*, 2014.

[48] 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 and information processing*, 6(1):83, 2008.

[49] 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.

[50] 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.

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

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

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

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

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

[56] B. O?Donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, pages 1–18, 2013.

[57] 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.

[58] 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.

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

[60] 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.

[61] J. Trzasko, A. Manduca, and E. Borisch. Sparse mri reconstruction via multiscale l0-continuation. In *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*, pages 176–180. IEEE, 2007.

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

[63] 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.

[64] 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.

[65] 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.

[66] 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.

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

[68] 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.